New photochemical tools for time-resolved structural studies

Diana Canelhas Freitas Monteiro

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

September 2015

To my grandmother,

a bright light.

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work is explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3: A molecular understanding of aspartate α-decarboxylase activation. This chapter is based on work from the following two jointly authored publications:

Monteiro, D. C. F. *et al.* Formation of a heterooctameric complex between aspartate alpha-decarboxylase and its cognate activating factor, PanZ, is CoA-dependent. *Biochem. Biophys. Res. Commun.,* 2012, 426, 350-5. Diana C. F. Monteiro (the candidate) carried out the majority of the work and designed the experiments. Michael D. Rugen performed some preliminary ITC experiments and cloned the PanD(T57V) mutant. Dale Shepherd collected and processed the ESMS data. Shingo Nozaki and Hironori Niki performed the *in cellulo* assays.

Monteiro, D. C. F. *et al.* The Structure of the PanD/PanZ Protein Complex Reveals Negative Feedback Regulation of Pantothenate Biosynthesis by Coenzyme A. *Chem. & Biol.,* 2015, 22, 492-503. Diana C. F. Monteiro (the candidate) carried out the majority of the work and designed the experiments. Vijay Patel worked on the crystallisation and X-ray diffraction data collection of the PanD(T57V) and PanD(S25A) complexes. Christopher Bartlett cloned PanZ into the pET28a vector, expressed and purified the ^{15}N labelled PanZ and collected and processed the protein NMR spectra with the aid of Gary S. Thompson and Arnout P. Kalverda. Shingo Nozaki and Hironori Niki performed the *in celullo* assays. Thomas D. Grant, Edward E. Snell and Arwen R. Pearson collected and processed SAXS data. Arwen R. Pearson aided protein crystallisation, X-ray diffraction data collection and structure solution. Michael E. Webb performed the ITC catalysis experiment and processed the data.

Acknowledgements

First and foremost, I would like to thank my three supervisors, Dr Stuart Warriner, Dr Michael Webb and Prof. Arwen Pearson for all the patience and kindness they showed me over the last four years. The project described over the next ~200 pages was initiated after a 10 week rotation with Prof. Pearson, during which I realised that I was no longer a pure synthetic chemist but rather an applied chemist with a soft spot for crystallography. Dr Warriner and Dr Webb helped shape the initial ideas into a joint project and, along with Prof. Pearson, showed great support throughout the time of my PhD. Thank you for the long discussions and brainstorming sessions, for finding time to talk and check on the (non-)progress almost every day, for taking me on synchrotron trips and feeding me coffee and tea during the long shifts and for listening to the breakthroughs with enthusiasm before asking me about all the variables I should have thought about already.

Secondly, I have to thank all the members of the Webb, Turnbull, Warriner and Pearson groups for being good friends and fun colleagues and for making the time to go to the pub both to celebrate good results and to help making the end of a bad day at work less painful. Thank you to all past and present members of lab 1.49 for giving good advice and good hugs, teaching me a lot about biochemistry and molecular biology, for all the dinners out to the Thai on Hyde Park Corner, Akmal's (a.k.a. the toilets), Old Bar and Terrace after long days in the lab. A special thank you to Phil, who plays the best devil's advocate and with whom I had lengthy discussions about science and all other kinds of worldly issues and also Briony Yorke for teaching me quite a lot of the practical aspects of crystallography.

A big thank you to Chi Trinh for all the help, lengthy discussions and for being always fun and kind. To all the beamline scientists and facility managers for always being willing to help collect and process data as well as for keeping all the instruments running so well. To the Astbury Centre and the Wellcome Trust for giving me the opportunity to do my PhD at Leeds in such an interdisciplinary and supportive environment, always providing opportunities to grow and to meet new people, some of which who will always be inspiring role models.

Last, but definitely not least, a big thank you to my family who have always been proud and supportive of my choices and my career and have helped me grow and become who I am today.

Abstract

The current bottleneck preventing the wider application of time-resolved structural techniques is the ability to quickly and accurately trigger protein function across an ensemble of molecules either in solution or *in crystallo*. The fastest and most accurate approach to achieve protein function synchronisation is by using protein photocaging approaches. During this project, a new class of protein photocleavable crosslinking reagents was designed and synthesised. This novel crosslinking scaffold carries two α-bromoacetate groups, which were shown to react with protein surface cysteines efficiently and cleanly. The crosslinker is released from the protein by UV irradiation, by photolysis of *ortho*-nitrobenzyl groups, cleaving the surface "staple" from the cysteine residues, leaving only a small methyl carboxylate group on the cysteine side chains. The cysteine "anchors" can be easily introduced by site-directed mutagenesis. This new set of reagents were developed as a new approach to protein photocaging which decouples the protein activation step from the protein chemistry being investigated. The photocage does not target residues directly involved in function but rather aims to restrict the conformational space explored by the protein, limiting essential dynamics and preventing function. This approach opens avenues for more widely applicable protein triggering methodology, potentially allowing for time-resolved experiments to be performed in wider variety of protein classes than those investigated to date.

To test these reagents, aspartate α-decarboxylase (ADC) was chosen as a model system. This enzyme catalyses the conversion of aspartate to β-alanine, a precursor of coenzyme A. ADC is expressed as an inactive zymogen which cleaves post-translationally, yielding a catalytic pyruvoyl group. The cleavage requires an additional activating partner, PanZ. Protein homogeneity is an essential requirement for a successful time-resolved experiment and, therefore, the mechanism by which PanZ activates ADC was investigated. The crystal structure of the ADC-PanZ protein-protein complex was determined at high-resolution, as well as those of several ADC mutants. With the aid of complementary techniques (SAXS, ITC, MS, NMR, *in cellulo* studies) the molecular mechanism by which this protein-protein interaction promotes ADC activation was elucidated. The protein complex formation causes the formation of a strained, activation-competent conformation of the ADC peptide backbone at the site of cleavage. Furthermore, a novel negative feedback mechanism for pantothenate and CoA biosynthesis regulation in bacteria was discovered and proposed.

Table of Contents

VI

Glossary

[A] .... concentration of A
°C .... degrees Celsius
Å .... angstrom
AcCoA .... acetyl coenzyme A
ADC .... aspartate α-decarboxylase (mature enzyme)
aq. .... aqueous
ATP .... adenosine triphosphate
Bis-Tris propane .... 1,3-Bis[tris(hydroxymethyl)methylamino]propane
cal .... calorie
cAMP .... cyclic adenosine monophosphate
cGMP .... cyclic guanosine monophosphate
cm .... centimetres
Cm .... coumarinyl
cNMP .... cyclic nucleoside monophosphate
CoA .... coenzyme A
Da .... Dalton
DDAH .... dimethylarginine dimethylamino hydrolase
DHF .... dihydrofolate
DHFR .... dihydrofolate reductase
DLS .... Diamond Light Source
DMF .... dimethyl formamide
DMSO .... dimethyl sulfoxide
DTT .... dithiothreitol
EM .... electron microscopy
eqv. .... equivalents
ESRF .... European Synchrotron Radiation Facility
fs .... femtosecond
g .... gram
HRMS .... high-resolution mass spectrometry
IDH .... isocitrate dehydrogenase
ITC .... isothermal titration calorimetry
K .... Kelvin
kcal .... kilocalorie
K_D .... dissociation constant
kDa .... kilodalton
kpsi .... kilopound per square inch
KSCN .... potassium isocyanoacetate
L .... litre
LB .... lysogeny broth
LC-MS .... tandem liquid chromatography mass spectrometry
M .... molar
mg .... milligram
min .... minute
mL .... millilitre
mM .... millimolar
MS .... mass spectrometry
ms .... millisecond
$NADP^+$ .... nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH .... nicotinamide adenine dinucleotide phosphate (reduced form)
Ni-NTA .... nickel loaded nitrilotriacetic acid resin

X

1 Introduction

1.1 Structural studies of biomacromolecules

The last century has seen the discovery and development of biophysical techniques that have allowed for the structure of (bio)molecules to be determined with great accuracy. The rate of development of the techniques and instrumentation has been exponential; in 1912 the first X-ray diffraction studies allowed for the structures of simple metal halides to be determined (Bragg, 1913) and the first structure of a small organic molecule, hexamethylenetetramine, was solved a decade later in 1923 (Dickinson and Raymond, 1923). It took almost 50 years from these initial discoveries until the first protein structure was determined, that of myoglobin, in 1958 (Kendrew *et al.*, 1958), but since then structural biology techniques have been in constant development and there now are more than 97,000 X-ray crystallography structures published in the Protein Data Bank (PDB).

High-resolution structural information of biomacromolecules in combination with biochemical data has allowed for a much greater understanding of how macromolecular structure is related to function. From such studies structure-activity relationships can be extracted, which define the role of essential residues in protein function. For example, in the case of enzymes, the architecture of the protein explains how the substrate binds to the active site and defines which residues are responsible for the interactions between the enzyme and the substrate. The relative spatial arrangement of the functional groups of the amino-acid side chains and the substrate allows for a chemically relevant mechanism to be proposed. It also explains how the enzyme can recognise the substrate and how it is able to modulate the energy landscape of the chemical reaction allowing it to occur under physiological conditions. In the case of protein-protein and protein-DNA interactions, understanding the structure of the complexes allows for the determination of the, often large, binding interfaces and the basis for the high affinity and specificity of such interactions.

These structure-activity relationships have been vital for the understanding of how proteins function and misfunction. Protein misfunction can lead to the disregulation of biological processes and cause disease. Knowing the structures of medically relevant targets has allowed for the design of small molecules that modulate protein activity. Many drugs that are currently on the market have been designed and developed to target

specific protein binding sites, such as HIV protease inhibitors and DNA replication inhibitors (Anderson, 2003, Simmons *et al.*, 2010).

One further very interesting application of structure-function relationships is the possibility of developing new artificial protein machines that are able to mimic protein function. As proteins are highly efficient and selective, such molecular machinery can have a great impact in industry, especially in making manufacturing processes more efficient and reducing waste (Samish *et al.*, 2011, Khoury *et al.,* Yu *et al.*, 2014).

1.1.1   Why aren't high-resolution static structures enough?

Undoubtedly, structural biologists are now proficient at determining high-resolution structures of macromolecules and their complexes, and this breadth of structural information has allowed for the in-depth study of protein mechanisms. Nevertheless, even such detailed, high-resolution information is severely incomplete, as it only provides a static, ensemble averaged picture of the system. It is well established that biomacromolecules are intrinsically flexible and that dynamics are intimately connected to function (Wand, 2001, Henzler-Wildman *et al.*, 2007a).

The conformational changes that occur during protein function span a wide range of time and length scales. Some structural changes are very small, such as the breaking of a covalent or a hydrogen bond, or the rotation of an amino acid side chain. Small movements are essential during enzyme catalysis, where the protein has to first bind the substrate tightly and then adapt its conformation to accommodate the different chemical intermediates and transition states along the reaction coordinate (Henzler-Wildman *et al.*, 2007b, Klinman, 2013, Klinman and Kohen, 2013, Hanoian *et al.*, 2015).

Larger protein motions can involve single loops or helices (Malhotra *et al.*, 2013), and even whole domains. Hinge motions are involved in the opening and closing of membrane proteins (Feld and Frank, 2014, Zhou and Robinson, 2014) and large rearrangements can occur during protein-protein interactions (Boehr *et al.*, 2009).

None of these motions are isolated events, rather they are usually correlated with changes across the whole protein (Eisenmesser *et al.*, 2005). The timescales of conformational change are also intimately related to the length scale of structural change. Bond breaking or isomerisation occurs very rapidly (fs-ps), side-chain rotations in the ns time-regime and larger conformational changes on ms-s timescales (Fig. 1.1) (Gray *et al.*, 2015, Levantino *et al.*, 2015a).
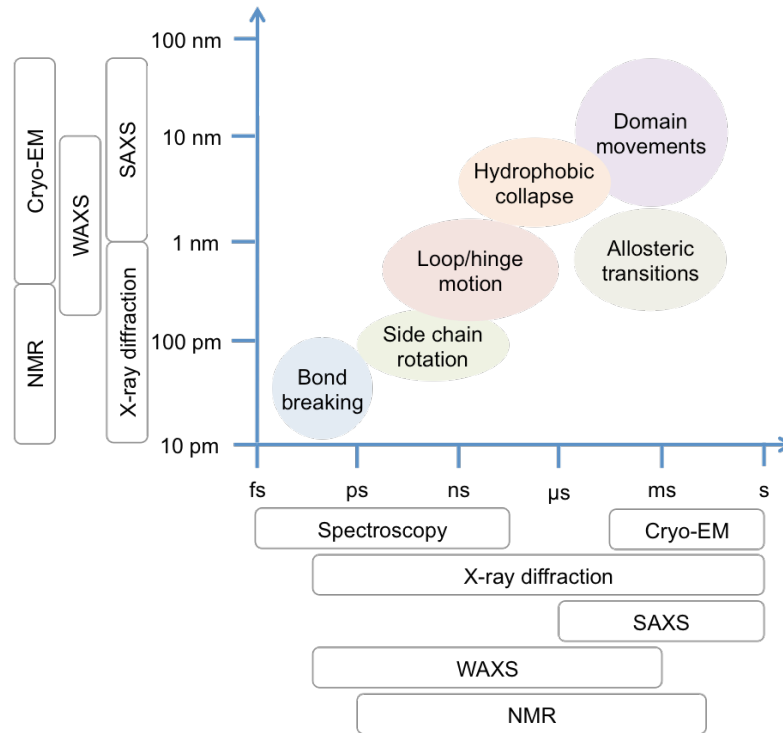
Fig. 1.1 Timescales and length-scales of protein conformational changes, showing the biophysical structural probes able to capture such transitions by both time-resolved and non-time resolved approaches.

Changes in protein conformation during function mean that structure-activity relationships also change over time and, therefore, the relationships derived from a static model are only absolutely valid for that exact conformational state of the protein. When structures for multiple conformers of a protein can be obtained, their involvement in specific stages of protein function can sometimes be rationalised with the aid of other biochemical data (Schnell *et al.*, 2004). Nevertheless, the mechanisms by which proteins interconvert between conformations are difficult to probe and the structures of short lived, transient states are particularly elusive to standard structural techniques, making the static information obtained from routine structural studies severely incomplete. To fully understand how biomacromolecules perform their function, experiments must be able to probe both the structure and dynamics of the system over time.

1.1.2 The role of protein dynamics in enzyme catalysis

As well as the understanding of protein function gained from static structures, the dynamic properties of biomacromolecules must also be investigated, adding a fourth dimension to structural studies, time. For a complete description of a system, each conformational state adopted by a protein as it functions and the energy barriers for the

3

interconversion between states have to be defined (Henzler-Wildman and Kern, 2007). The dynamics of proteins can be classified according to the energy barriers associated with each conformational transition. Higher energy barriers are associated with slower movements and wider structural changes which occur at longer timescales (µs or slower). Many aspects of biological function occur within these timescales, such as catalysis, protein-protein interactions and folding. Such processes are characterised by distinct conformational changes that can sometimes be distinguished experimentally, such as different structures observed by crystallography or the ensembles of conformations obtained by NMR or EM.

Faster timescales are associated with more local movements, such as loop motions and side chain rotations (ns-ps timescales) and even the bond breaking and making events that occur during an enzymatic reaction (fs). These faster dynamics are usually associated with a constant thermal fluctuation of the protein between structurally similar states, which contribute to the entropy of the system and are, therefore, involved in the thermodynamics of the system and vital for function. This entropic aspect can sometimes be visible in crystal structures from the analysis of temperature factors that define the level of displacement of atoms from their average position (Fraser *et al.*, 2011).

NMR can provide information about protein dynamics at both short and long timescales from relaxation dispersion experiments (Kay, 2005). Recent NMR studies have uncovered protein motions of enzymes at various timescales under steady-state conditions. Three examples have been chosen to illustrate the importance of protein dynamics for enzymatic activity.

Cyclophilin A

The timescales of specific conformational oscillations of cyclophilin A during the catalysis of *cis-trans* isomerisation of prolyl peptide bonds (Fig. 1.2), have been determined and matched to those of the catalytic step from kinetic experiments (Eisenmesser *et al.*, 2005). Residues exhibiting dynamics on timescales similar to the rate of catalysis were mutated and shown to affect substrate binding affinities and catalytic efficiency, supporting the proposal that the motions observed by NMR are directly involved in catalysis.
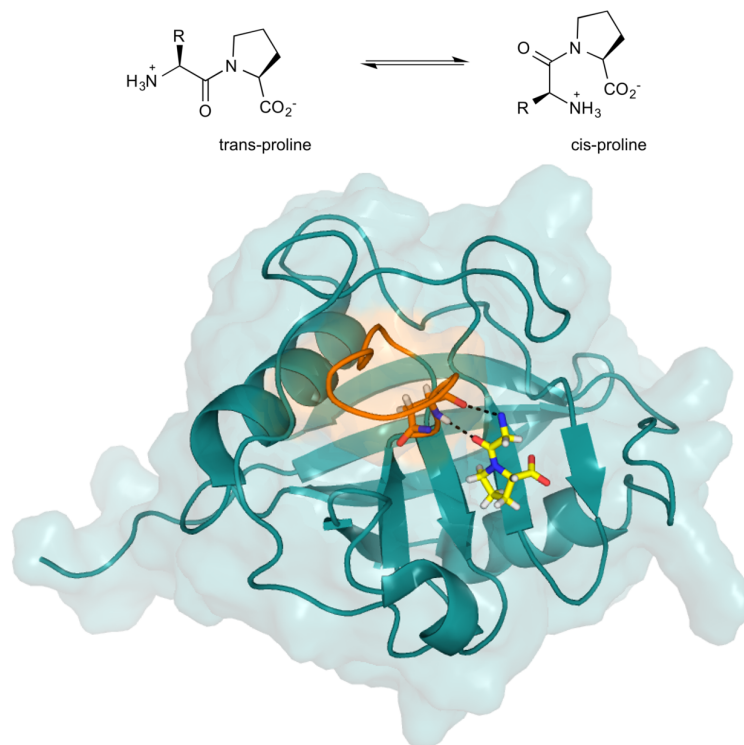
Fig. 1.2 Cyclophilin A enzymatic reaction and dynamics. (**Top**) the cis-trans isomerisation of proline containing peptides by cyclophilin A. A dipeptide is shown, with the general N-terminal amino acid side chain represented as R. (**Bottom**) cyclophilin A (PDB 4N1M), shown as a cartoon (teal) in complex with a Gly-Pro peptide (sticks with carbons coloured yellow). The [101]Ala-Ser[110] loop is shown in orange with Asn102 in sticks and the hydrogen bonds to the peptide as black dashed lines.

This enzyme is a tractable model system to be studied by molecular dynamics, as the catalytic step does not involve bond breaking or making, greatly simplifying simulation of changes at the active site. Classical mechanics modelling was used to investigate the relationship between the conformational dynamics of cyclophilin A and the enzymatic reaction (McGowan and Hamelberg, 2013). These simulations have shown that there are conformational states shared between the substrate-bound and transition-state forms and that the enzyme optimises binding to the transition state by making use of a pre-organised network of hydrogen bonds that are sampled from the substrate-bound state. Beyond the small fluctuations in the positions of active site residues that promote catalysis, larger conformational changes were also observed between the *apo* and bound structures, with changes to loops beyond the active site. One particularly interesting dynamic change is related to a loop adjacent to the active site (residues [101]Ala-Ser[110]), which becomes more flexible upon binding of the substrate, allowing for the backbone of asparagine 102 to sample more conformations and form a hydrogen bond to the substrate, promoting the turnover step (Fig. 1.2).

Dimethylarginine dimethylamino hydrolase

Dimethylarginine dimethylaminohydrolase (DDAH) catalyses the conversion of N,N-dimethylarginine to L-citrulline (Fig. 1.3). N,N-dimethylarginine is involved in the regulation of nitric oxide synthase (NOS), which, in turn, maintains vascular tension in mammals (Leiper and Nandi, 2011). The active site is covered by a dynamic loop, L1 (residues 17-27), which is unstructured in the *apo* form of the enzyme but becomes ordered upon binding of the substrate, shielding the active site (Fig. 1.3) (Murray-Rust *et al.*, 2001).
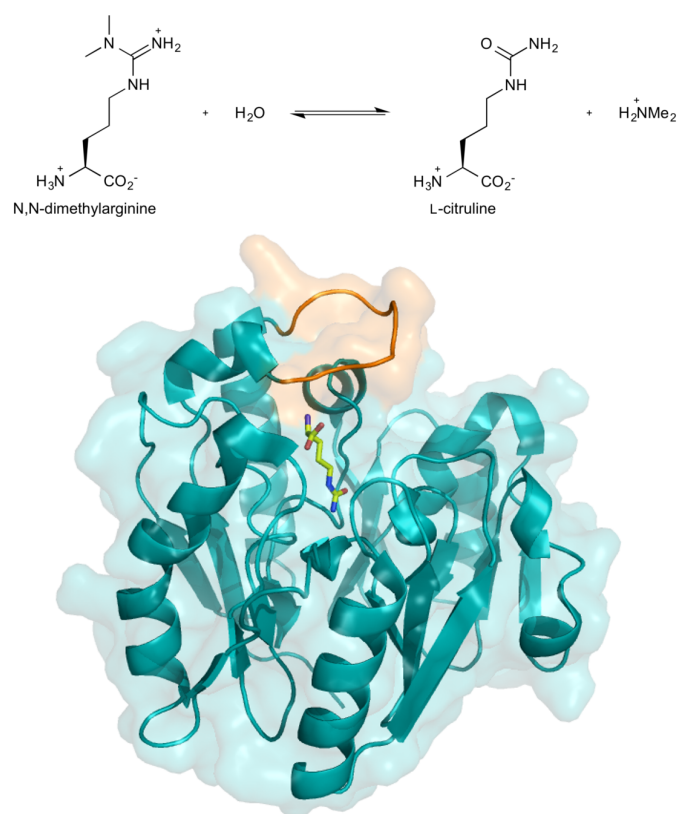
Fig. 1.3 DDAH enzymatic reaction and dynamics. (**Top**) the conversion of N,N-dimethylarginine to L-citrulline by DDAH. (**Bottom**) DDAH C254S mutant (PDB 1H70, teal cartoon) in complex with citrulline (sticks with carbons coloured yellow), The ^{17}Gly-Asp^{27} loop is shown in orange, capping the active site.

A combination of NMR, molecular dynamic simulations and fluorescence studies on the dynamics of the L1 loop and the remaining residues lining the active site have suggested that the rate limiting step of the reaction is correlated with the substrate binding step and, therefore, the conformational change of loop L1 (Rasheed *et al.*, 2014).

Dihydrofolate reductase

The dynamics of dihydrofolate reductase (DHFR) have been extensively studied, characterised and implicated in catalysis (Luk *et al.*, 2015). DHFR catalyses the reduction of dihydrofolate (DHF) to tetrahydrofolate in a hydride transfer reaction which requires NADPH as a cofactor (Fig. 1.4). Fast vibrational motions (in the fs-ps time regime) have been observed for different DHFR mutants and are proposed to promote the hydride-transfer step by sampling distances along the reaction coordinate through a network of coupled motions throughout the protein (Agarwal *et al.*, 2002). DHFR also exhibits three distinct conformations - open, closed and occluded - of the highly mobile Met20 loop (residues 9-24), which are associated with the different ligand binding states of the enzyme and, therefore, the different steps along the catalytic cycle.
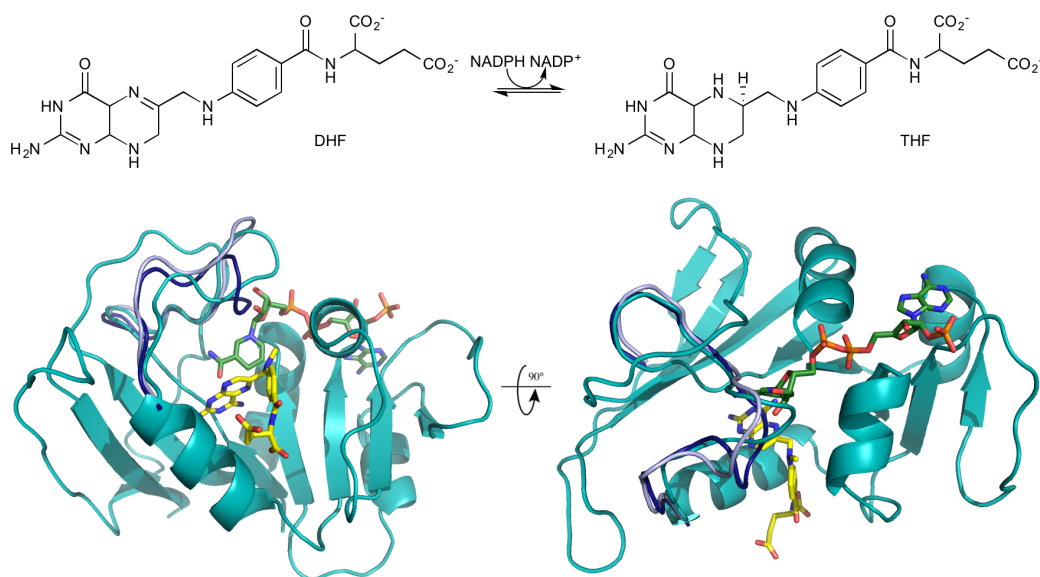
Fig. 1.4 The enzymatic activity of DHFR and different Met20 loop conformations. (**Top**) the reduction of DHF to tetrahydrofolate catalysed by DHFR. NADPH is used as the reducing agent. (**Bottom**) the closed (teal), open (light blue) and occluded (dark blue) conformations of the DHFR Met20 loop (PDB 1RX3, 1RA9 and 1RC4 respectively). Methotrexate (a transition state mimic) and NADPH are shown as sticks (carbons coloured yellow and green respectively).

The occluded and closed conformations have been previously observed by crystallography in different enzyme-ligand complexes, as shown in Fig. 1.5 (Sawaya and Kraut, 1997) and the distinct conformations have been associated with cofactor, product and substrate binding and release states. An open conformation is also visible in certain crystal forms and is thought to be associated with the interconversion between

the closed and occluded states, allowing for substrate binding and product release from the binding site.
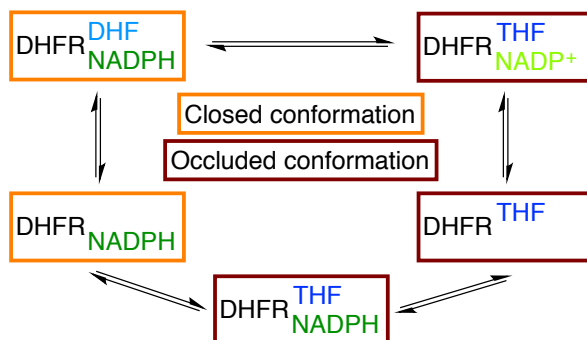
Fig. 1.5 The catalytic cycle of DHFR, showing the different enzyme complexes. The catalytic step reduces DHF tetrahydrofolate at the expense of the oxidation of NADPH to NADP$^+$, with a clear conformational change of the Met20 loop from a closed to an occluded conformation. The enzyme then releases the oxidised cofactor to bind a new molecule of NADPH, which then promotes the release of the product.

The millisecond Met20 loop motions have been experimentally correlated with rate of turnover (Bhabha *et al.*, 2011), but the hypothesis that they're directly correlated has been refuted by computational work (Adamczyk *et al.*, 2011). This is one example of the potential pitfalls of the studies described in this section, where dynamic aspects of enzymes have been *indirectly* correlated with the rates of catalytic steps. Without the direct experimental visualisation of the conformational changes over the course of the catalytic cycle, incorrect associations can be made. Dynamics of enzymes are clearly essential for efficient catalytic activity and determining the correlation between structure and dynamics is vital for obtaining a complete understanding of the system. Only by direct visualisation of the structural and chemical changes of the protein in real time can these correlations be fully founded and irrefutable. This can be done by using time-resolved techniques, where the conformations and chemical steps adopted by the protein, as it functions, can be captured in real time.

1.1.3 Time-resolved structural techniques

A typical "pump-probe" time-resolved experiment is depicted in Fig. 1.6. The ensemble of molecules in a resting state is first triggered to initiate function (t = 0). The reaction is allowed to proceed for a specific amount of time, defined as Δt. After the first defined time-delay, Δt_1, a probe pulse is applied, providing information about the exact conformational state of the system at t_1. The experiment is repeated multiple times for each time point to increase both the quality and completeness of the data. The experiment is also repeated at different time delays ($\Delta t_2, \Delta t_3 ... \Delta t_n$), providing sequential

conformational information along the reaction pathway. This information can be interpreted as a "stop-motion" type movie, which, although simplified, describes the molecular motions of the protein as it functions.
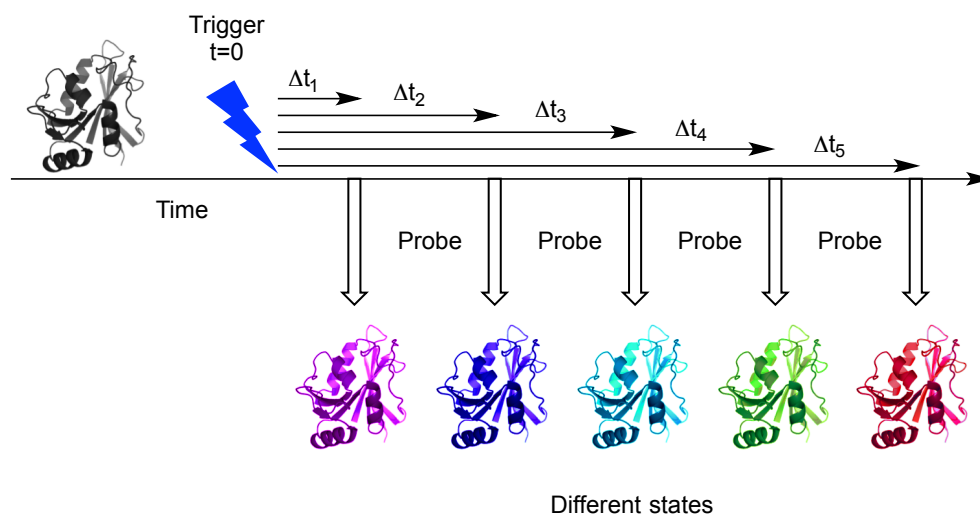
Fig. 1.6 Overview of a time-resolved experiment. Each coloured protein represents the ensemble of molecules in a distinct chemical or conformational state. At time 0, the ensemble is rapidly, and ideally uniformly triggered, starting the biological process. After a defined time delay (Δt_1) the ensemble is probed to investigate the conformational state at that Δt after triggering. The experiment is repeated multiple times to increase the signal to noise and provide complete datasets and also repeated at different time points (Δt_1-Δt_n), revealing the sequential conformations adopted during function.

A variation of this set-up can be defined as a "pump-probe-probe-probe" approach, where the biological process is initiated only once but probed multiple times at increasingly longer time delays. This approach is preferable, as it diminishes the amount of sample required for the experiment, but its feasibility is completely dependent on the experimental setup and, most importantly, on the sample itself. If the sample interacts with the probe - for example, if it becomes damaged - then multiple probing pulses must be used with caution.

Although the principle behind a trigger-probe experiment is simple, the experimental set-up is not trivial and requires a highly interdisciplinary approach. When designing a time-resolved experiment many different aspects have to be considered (Levantino *et al.*, 2015a), related to sample preparation, instrumentation and experimental requirements, as summarised in Fig. 1.7. The design and set-up of a time-resolved experiment requires a multitude of skills and relies on scientific teams that include hardware scientists and software specialists working alongside the chemists and structural biologists. The detailed intricacy of such an experiment is beyond the scope of this thesis, but some of the aspects are discussed below.
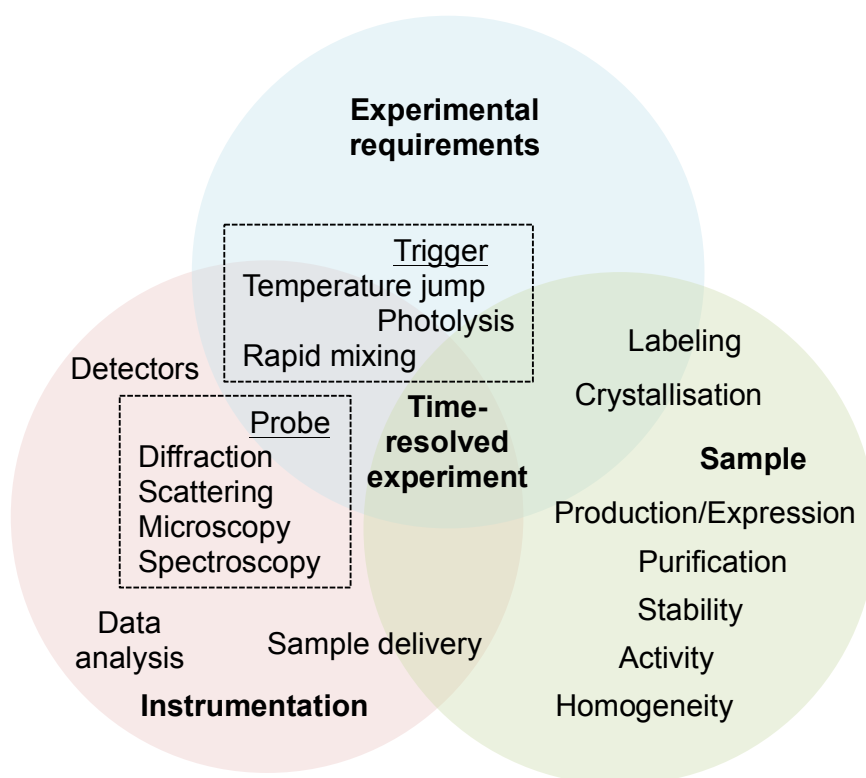
Fig. 1.7 Overview of important considerations for the design of a time-resolved experiment. The experiment has been divided into 3 main aspects: experimental requirements, sample and instrumentation, shown as 3 circles coloured blue, green and red respectively. Some detailed considerations regarding each section are shown. Where an aspect of the set-up is shared between different sections, it occupies the area shared by the circles. The possible triggering and probing techniques are grouped in dashed boxes under the corresponding header.

The sample

As in all structural studies of biomacromolecules, obtaining a suitable sample is the first concern. The protein needs to be available in sufficient quantities and, if overexpressed the protein constructs have to be stable and represent a biologically relevant state of the protein. The sample has to be homogeneous as the experiments are performed on ensembles of molecules, making a good purification protocol an essential requirement. The sample must also be active, which may require post-translational modifications or the presence of ligands or cofactors. For NMR studies, the protein has to be isotopically labelled and for diffraction experiments, it must be crystallised. Crystallisation is still an empirical process, where extensive screening may be necessary to obtain conditions that yield well-diffracting, homogeneous crystals that can also accommodate the conformational changes occurring during function.

10

Instrumentation

When collecting time-resolved data, the main consideration is the required signal to noise ratio: how to accurately collect sufficient data for each time-point to make a meaningful measurement. Recent developments in hardware, such as brighter X-ray sources, more powerful electron microscopes as well as sensitive detectors, have made time-resolved diffraction, scattering and microscopy experiments feasible (Levantino *et al.*, 2015a).

The integration of sample delivery systems within the experimental set-up can also be challenging. In many cases, the sample has to be continuously exchanged (due to radiation damage or if the reaction of interest is irreversible).

Data analysis is also non-trivial; merging of the data from repeated measurements at each time-point to give complete and meaningful datasets can be difficult, especially for diffraction and microscopy experiments. In addition, if more than one conformational state is present at any time-point, accurate deconvolution of the states is also necessary.

The experiment

The probe

Biomolecular structural techniques can be divided into distinct classes, depending on the resolution and extent of information that can be obtained from the experiments. Solution based techniques such as small-angle X-ray scattering (SAXS) and wide-angle X-ray scattering (WAXS), give a global view of the biomolecular structure. SAXS provides information about the overall shape of the molecule and is especially useful to determine overall macromolecular properties or large conformational changes (Doniach, 2001, Pollack and Doniach, 2009) whereas WAXS provides information on finer structural features, such as subtle changes in fold and secondary structure elements (Fischetti *et al.*, 2004, Chen *et al.*, 2008).

High-resolution structural techniques that give global structural information about the system are X-ray diffraction, electron microscopy (EM) and nuclear magnetic resonance (NMR). All three techniques have practical limitations:

- Crystallography requires homogeneous, well diffracting crystals, which can be difficult to obtain, especially from large proteins, biomacromolecular complexes and membrane proteins.
- EM is mainly limited by particle size and the large quantity of micrographs needed. The protein has to be sufficiently large to be distinguishable within the

micrograph, and thousands of images of the particles in different orientations have to be averaged together to improve the signal to noise ratio. The achievable resolution depends on the amount of data available and the homogeneity of the sample.

- Contrary to X-ray diffraction and EM, where the proteins are used in their native state without need for derivatisation, NMR requires the isotopic labelling of the sample with NMR active nuclei (^{15}N, ^{13}C). NMR is also limited by sample size, with smaller proteins being preferred (<60 kDa), although great advancements in isotopic labelling have recently been made to overcome this limitation (Tzeng *et al.*, 2012).

Spectroscopy is also a high-resolution technique, but only provides sparse local information. It is very sensitive and thus especially powerful for identifying fine changes in protein structures, such as changes in chemical environments of residues and bond-lengths. Proteins can be investigated in their native form or chemically modified site-specifically with probes.

The trigger

The "trigger" has to cause a change in the system that starts its function quickly and cleanly. The triggering event must be faster than the process under investigation. As mentioned previously, protein dynamics span a wide range of timescales, from fs to seconds (Fig. 1.1). Therefore, the triggering method that can be used is completely dependent on the speed of the process being investigated. Finally it must affect at least a significant fraction of the ensemble, sufficient to be detected by the probe.

There are two main ways to trigger biological processes: rapid mixing and light excitation. Rapid mixing is slower, usually allowing for ms time-resolution (West *et al.*, 2008, Konuma *et al.*, 2011), with the fastest reported continuous flow devices reaching 100 µs time-resolution (Akiyama *et al.*, 2002, Trebbin *et al.*, 2014). Solutions of the protein of interest and a binding partner (small molecule, DNA or even a second protein) are mixed to trigger function. Δt is determined from the flow rate and the distance between the mixing point and the probe pulse interaction point (Fig. 1.8, left). This technique is applicable to virtually any system and can even tolerate microcrystalline slurries for diffraction experiments (Schmidt, 2013). Recent developments in microfluidic devices have also allowed for a dramatic decrease in the volume of sample required for the experiments.

Light excitation (Fig. 1.8, right) allows for much wider range of time-resolutions to be probed (fs-ms). Short laser pulses are used to either directly initialise a photochemical reaction, through the photolysis or isomerisation of a bond or excitation of a chromophore, or to create a temperature jump (Kubelka, 2009). Unsurprisingly, proteins that are naturally photoactivatable have been the main focus of recent developments in the field of time-resolved studies, but the majority of proteins are not naturally photoactivatable. In order to be able to use light as a trigger for non-naturally photoactivatable proteins, an artificial photosensitive moiety has to be chemically introduced into the system. This is done by *photocaging*: the introduction of a photocleavable protecting group at a strategic site on the protein or its binding partner that inactivates the system until the protecting group is removed by photolysis with a light pulse.
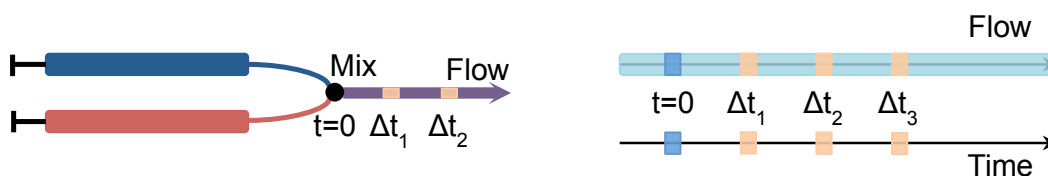
Fig. 1.8 Triggering methods for reaction initiation. Optical pump pulses are shown as blue rectangles. Probe pulses are shown as orange rectangles. (**Left**) Rapid mixing. Two solutions are mixed together at t=0. Δt is defined by the distance between the probe pulse and the point of mixing along the flow tube. The method can be used to trigger protein function by delivering binders (both in solution and as microcrystalline slurries) with different binding partners (small molecules, proteins, DNA) or by changing conditions, such as pH. (**Right**) Light activation. The diagram shows two possible experimental set-ups: proteins in a rapid flowing solution (top) or static samples (bottom). Reaction is triggered by application of a laser pulse at t=0. For the static sample, Δt is simply defined by a time delay between the two pulses, whereas in flow, Δt is defined by the distance between the probe pulse and the area of the flow tube illuminated by the triggering laser pulse.

1.1.4 Summary

Although extremely complex, time-resolved experiments that yield global structural information (from X-ray diffraction and scattering and cryo-EM data) can be performed (Shaikh *et al.*, 2014, Levantino *et al.*, 2015a), yielding fascinating insights into whole protein dynamics and protein motions during function at a variety of timescales.

For fast dynamic events (sub ms), the only systems that have been investigated with these techniques have been naturally photoactivatable proteins, meaning that the current bottleneck to making such techniques more widely applicable lies in the triggering methodology. For scattering experiments looking at relatively slow biological processes (slower than 100 μs), rapid mixing is a reasonable approach. For faster events, reaction

initiation has to be done by the use of light pulses, which require a non-trivial case-by-case development of a photocaging strategy. The following section describes the available approaches for non-natural photoactivation of proteins and the problems associated with the technology currently available, leading onto the main aim of this project.

1.2 Reaction triggering by light

1.2.1 Naturally light-activated proteins

Naturally photoactivatable proteins have been the main focus of time-resolved X-ray and scattering experiments, as these can be triggered reliably very fast both in solution and in crystals, often with extremely high yields. One such example is photoactive yellow protein (PYP), which has been extensively studied using time-resolved X-ray diffraction experiments (Schotte *et al.*, 2012, Jung *et al.*, 2013, Tenboer *et al.*, 2014). PYP is a small (14 kDa), soluble blue light receptor. It is found in photosynthetic bacteria and is thought to be responsible for signalling cascades in the organisms that cause them to move away from sources of harmful (blue) wavelengths of light (Sprenger *et al.*, 1993). In the ground state, the protein contains a trans-4-hydroxycinnamic acid cofactor, which is covalently attached to a cysteine residue (Cys69) though a thioester bond (Yamaguchi *et al.*, 2009). Upon absorption of a photon of light (400-475 nm, λmax=446 nm) (Nielsen *et al.*, 2005), the trans-alkene isomerises to a cis-alkene (Fig. 1.9), triggering a cascade of structural changes through the protein, which eventually lead to the formation of a signalling state, pB_1, which follows pB_0 but has not been observed by X-ray crystallography due to crystal packing constraints (Meyer *et al.*, 1987, Harigai *et al.*, 2003).
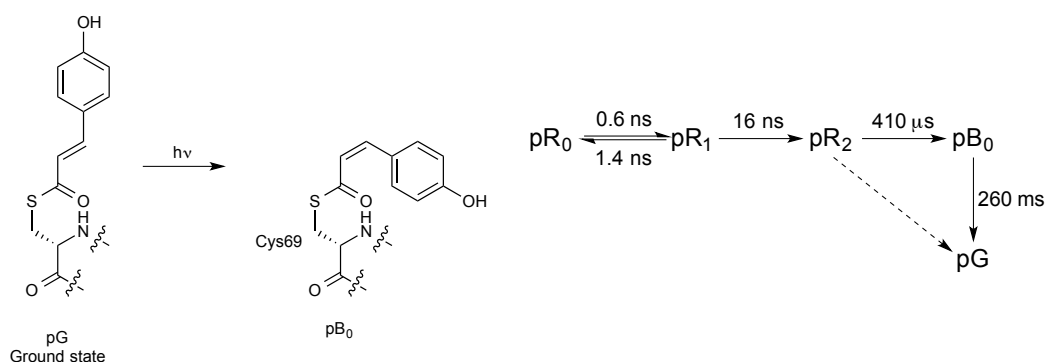
Fig. 1.9 The PYP photocycle. (**Left**) Isomerisation of the ground state (pG) of the trans-4-hydroxycinnamic acid cofactor to the cis pB_0 intermediate upon absorption of a photon of light. (**Right**) PYP photocycle intermediates structurally characterised by Schotte *et al* (Schotte *et al.*, 2012).

14

Pump-probe time-resolved Laue X-ray diffraction experiments have allowed for the differentiation of the structural intermediates of the PYP photocycle (Fig. 1.9) (Schotte *et al.*, 2012). Data were collected on timescales ranging from 100 ps to 1 s, and allowed for the visualisation of the first intermediate pR_0, a highly strained species which was too short lived to have been resolved in previous cryo-trapping (Kort *et al.*, 2004) or even ns time-resolved Laue X-ray diffraction experiments (Ihee *et al.*, 2005). The diffraction data were fitted to 4 structurally distinct intermediates, starting from a previously described kinetic model. The fraction of each intermediate at each time point was then calculated to best fit the electron density and these values evaluated to give a refined, structurally significant kinetic model. Structural changes are visible across the whole protein structure and are not limited to the vicinity of the chromophore. Recent studies using time-resolved serial crystallography at an X-ray free-electron laser have provided electron density maps with improved resolution and allowed further refinement of the relative populations of intermediates pR_1 and pR_2 at 1 μs post photoactivation (Tenboer *et al.*, 2014).

Other naturally photoactivated systems that have also been investigated by time-resolved techniques include myoglobin (Schotte *et al.*, 2003, Bourgeois *et al.*, 2006, Levantino *et al.*, 2015b), photosynthetic reaction centre (Baxter *et al.*, 2004, Wöhri *et al.*, 2010), haemoglobin (Ren *et al.*, 2012, Schotte *et al.*, 2013, Kim *et al.*, 2015), 3-isopropylmalate dehydrogenase (Hori *et al.*, 2000) and trypsin (Singer *et al.*, 1993). However, the vast majority of interesting protein targets are not naturally photoactivatable, making them even more challenging systems to work with.

1.2.2 Photocaging

A photocaging group is a small organic molecule that can be covalently attached to an essential residue within a protein or functional group on a ligand, inactivating the biological process. The photocage acts as a protecting group that can be liberated by photolysis with a light pulse of specific wavelength. There are many types of photo-removable protecting groups, but only a few have been employed in the study of biological systems so far: *ortho*-nitrobenzyls (*o*NB), coumarins (Cm) and *para*-hydroxylphenyls (*p*HP) (Corrie *et al.*, 2005, Hagen *et al.*, 2005, Mayer and Heckel, 2006, Klán *et al.*, 2013). The wavelength of maximum absorption (λ_{max}), quantum yields (Φ) and rates of cleavage vary considerably for each class of compounds (see sections 4.1 and 9.6 for a detailed description of the spectroscopic properties, advantages and disadvantages of each of the three photocaging groups).

Substrate/ligand caging

When a small molecule ligand or substrate is required for function, the photocleavable protecting group can be introduced by chemically synthesising a substrate/ligand derivative. This approach is widely applicable, but can be very challenging and has to be designed on a case-to-case basis. Firstly, the photocage has to be introduced at a site within the small molecule that inhibits function without completely inhibiting binding of the ligand to the protein, otherwise the time-resolution of the experiment becomes dependent on the post-photolysis rate of diffusion of the free ligand into its binding site. Once the photocaged molecule has been designed, the synthetic availability and ease of synthesis of the compound has to be evaluated. Even for simple chemicals, a good knowledge of synthetic chemistry is necessary to produce such molecules.
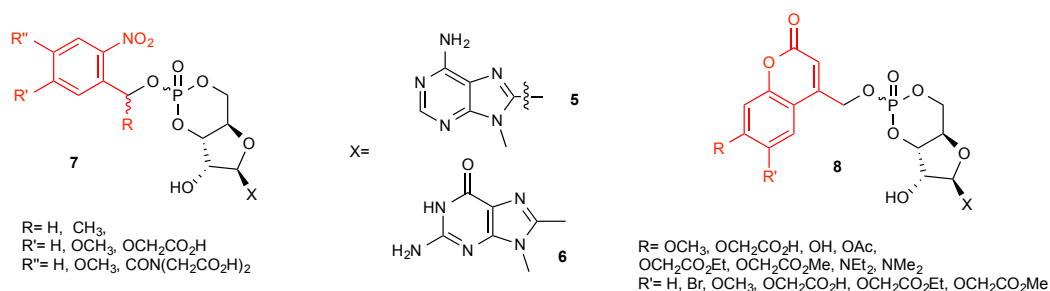
Nevertheless, photocaged biochemically or physiologically active compounds have been in use for the last 3 decades. The first biologically relevant photocaged molecule was adenosine 5'-triphosphate (ATP), bearing either a 2-nitrobenzyl group (*o*NB, Fig. 1.10, **1**) or a 1-(2-nitrophenyl)ethyl group (*o*NPE, Fig. 1.10, **2**) (Corrie *et al.*, 2005). These compounds were used to investigate the function of a human sodium-potassium pump (Kaplan *et al.*, 1978) under continuous illumination for decaging as well as in the first flash photolysis experiments to observe the interaction between myosin and actin with a ms time-resolution (McCray *et al.*, 1980, Goldman *et al.*, 1982). These initial studies, along with the flash photo-isomerisation of azobenzene derivatives (Lester and Nerbonne, 1982), were the dawn of the use of flash photolysis for the study of biological mechanisms.
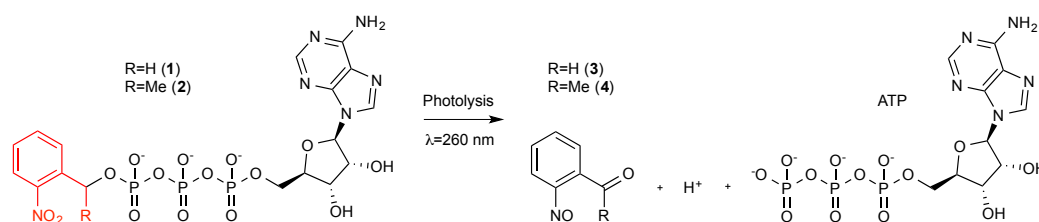
Fig. 1.10 Caged ATP with 2-nitrobenzyl and 1-(2-nitrophenylethyl) photoprotecting groups (**1** and **2** respectively) and corresponding photolysis products: 2-nitrobenzaldehyde and 1-(2-nitrophenylacetone) (**3** and **4** respectively), ATP and proton. The photocaging group is highlighted in red.

Cyclic nucleoside monophosphates (cNMPs) bearing both *o*NB or coumarinyl derivatives (Fig. 1.11, **7** and **8** respectively) have been previously used to investigate spatial- and time-dependent aspects of signalling pathways inside cells, by measuring changes in potentials across cell membranes that occur on a ms timescale after the triggering of cyclic nucleotide-gated ion channels (Corrie *et al.*, 2005). The rates of

16

cleavage of the photocaged cNMPs have not been reported for all the compounds, but where known vary between 3000 s^{-1} and 1.7 s^{-1} (Corrie and Trentham, 1993, Wang *et al.*, 2002).

R= H, CH$_3$,
R'= H, OCH$_3$, OCH$_2$CO$_2$H
R''= H, OCH$_3$, CON(CH$_2$CO$_2$H)$_2$

R= OCH$_3$, OCH$_2$CO$_2$H, OH, OAc,
OCH$_2$CO$_2$Et, OCH$_2$CO$_2$Me, NEt$_2$, NMe$_2$
R'= H, Br, OCH$_3$, OCH$_2$CO$_2$H, OCH$_2$CO$_2$Et, OCH$_2$CO$_2$Me

Fig. 1.11 Adenosine 3',5'-cyclic monophosphate (cAMP, **5**) and guanosine 3',5'-cyclic monophosphate (cGMP, **6**) protected with 2-nitrobenzyl or coumarinyl photocleavable groups (**7** and **8** respectively) used previously in time-resolved studies. The photocaging moieties are highlighted in red.

Other examples of previously used photocaged small molecules include glutamate (neuronal activity of glutamate-gated ion channels and glutamate transport) (Callaway and Yuste, 2002, Takaoka *et al.*, 2004), γ-aminobutyric acid (synaptic events of GABA receptors) (Molnár and Nadler, 2000) and hormones (regulation of gene expression) (Lin *et al.*, 2002, Link *et al.*, 2004, Shi and Koh, 2004). The majority of photocaged biologically active compounds have been used in *ex vivo*, in cell based studies or on membrane bound proteins, for example, by measuring changes in current across the membranes or along neurons or triggering apoptosis by releasing toxic proteins in cells (Goldmacher *et al.*, 1992, Hagen *et al.*, 2005). Such experiments span μs to s timescales and can be done by laser photolysis or continuous illumination. However, although the experiment triggers a local increase in concentration of a biologically active molecule, it does not require full activation or homogeneous activation as the differences in signal are measured over time. A more in-depth discussion of photocaging applications in cell biology is beyond the scope of this thesis, as this work is focused on the use of photocaged compounds for biophysical studies instead. For more detailed reviews of *in vivo* studies using photocaged compounds see (Hagen *et al.*, 2005, Mayer and Heckel, 2006, Deiters, 2009, Lee *et al.*, 2009, Yu *et al.*, 2010).

*o*NPE-caged guanosine triphosphate (GTP) was used for an early time-resolved crystallographic study of the GTPase mechanism of Ras. Ras is a GTP binding protein involved in cell signalling transmission pathways, which lead to downstream effects such as cell proliferation (Kinbara *et al.*, 2003). GTP hydrolysis leads to the inactivation of Ras and its GTPase activity is slow in the absence of a second protein GAP (GTPase-activating protein). In the absence of GAP, GTP hydrolysis is sufficiently slow that the

17

conformation of the active GTP-bound state and the intermediate GTP/GDP-bound state of Ras could be investigated. The experiment was performed by first triggering the system by the photolysis of the *o*NPE-GTP molecule using a xenon flash lamp (Schlichting *et al.*, 1989) followed by X-ray diffraction data collection using a polychromatic X-ray source at two time-points following GTP de-caging (4 min and 14 min (Schlichting *et al.*, 1990).

The first ms time-resolved crystallographic experiment of a single-turnover enzymatic reaction was performed in 1998 on isocitrate dehydrogenase (IDH), which catalyses the decarboxylation of isocitrate (**9**) to α-ketoglutarate (**11**) in the presence of NADP$^+$ and Mg^{2+}. The half-life of the enzyme-product complex is 10 ms (extendable to 50 ms by lowering the temperature or pH). The reaction was triggered by the photolysis of a 4,5-dimethoxy-*o*NPE protected NADP$^+$ molecule (**12**) using a short (0.5 ms) 450 nm light pulse (reported post-photolysis half-life of 0.05 ms) and diffraction data collected for 10 ms on a polychromatic X-ray source after a 2 ms lag-time (Stoddard *et al.*, 1998).
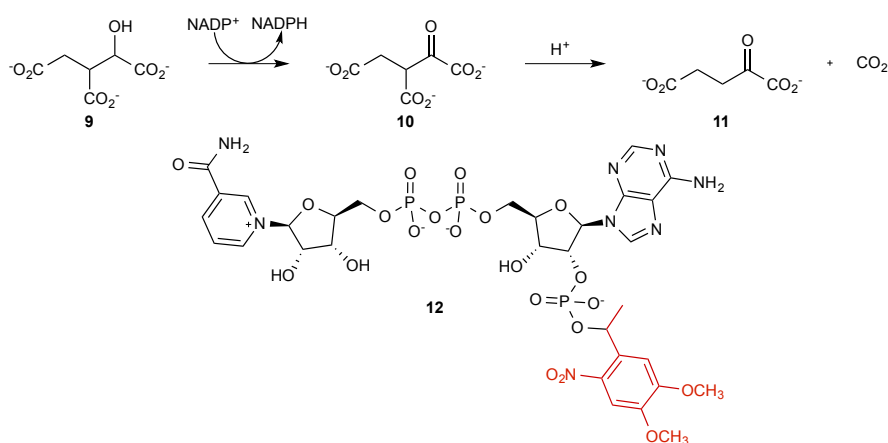
Fig. 1.12 (**Top**) Overview of steps in the catalytic dehydrogenation of isocitrate (**9**) to α-ketoglutarate (**11**) by isocitrate dehydrogenase (IDH). The first step is the oxidation of isocitrate to oxalosuccinate (**10**). (**Bottom**) the structure of the photocaged NADP$^+$ moiety (**12**) used in the time-resolved crystallographic study. The photocleavable moiety is highlighted in red.

Protein photocaging

As well as photocaging a substrate or ligand, the protein itself can also be photocaged. Site-specific incorporation of photocleavable protecting groups can be done via two distinct approaches – genetic incorporation of unnatural amino acids or using bioorthogonal reactions.

18

The site-specific[*] genetic incorporation of photocaged amino acids can be done using amber suppression, where an aminoacyl-tRNA synthetase and the corresponding tRNA are evolved to accept the desired unnatural amino acid and insert it into the protein sequence during mRNA translation in response to an amber stop codon integrated into the expression vector using site-directed mutagenesis (Fig. 1.13) (Lang and Chin, 2014). The current library of photocaged unnatural amino acids that have successfully been genetically incorporated into proteins is very limited, although it continues to slowly expand with the recent advances in the field. Efforts have mainly focused on the incorporation of amino acids caged with *o*NB derivatives. Fig. 1.14 shows all the photocaged amino acids that have been successfully genetically incorporated into proteins (Young and Schultz, 2010, Davis and Chin, 2012, Lang and Chin, 2014).
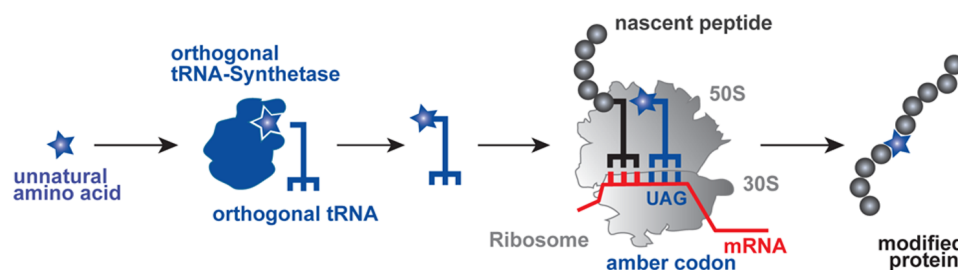

Fig. 1.13 Schematic of the genetic incorporation of unnatural amino acids (Lang and Chin, 2014). An orthogonal tRNA-synthetase/tRNA pair is evolved to accept an unnatural amino acid. During transcription, the unnatural amino acid is incorporated into the nascent protein polypeptide chain in response to an amber stop codon.
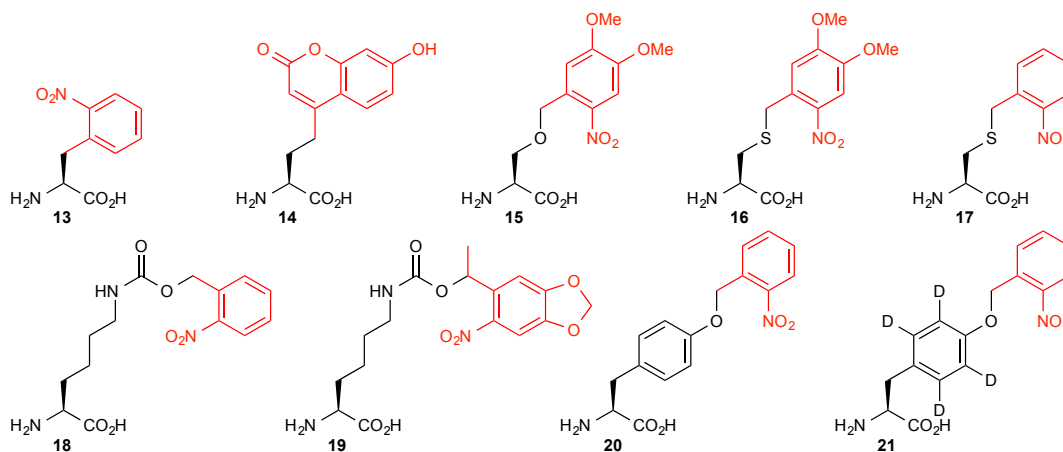
Fig. 1.14 Overview of all the photocaged protein residues introduced site-specifically by amber suppression to date. *o*NB derivatised amino acids are most common: with L-alanine (**13**), L-serine (**15**), L-cysteine (**16** and **17**), L-lysine (**18** and **19**, releasing carbamate derivatives) and L-tyrosine (**20** and **21**). Coumarin photocaging has been incorporated into L-aminobutyric acid (**14**) (Young and Schultz, 2010, Davis and Chin, 2012, Lang and Chin, 2014).

---

[*] Residue-specific incorporation of unnatural amino acids is also possible through selective pressure incorporation, leading to the substitution of all positions of a specific amino acid by an unnatural variant. This approach has not been used for photocaging and is undesirable as it causes global changes to the protein as well as the introduction of multiple chromophores which may lead to the observation of structural changes unrelated to protein function as well as decreasing the amount of available photons Johnson *et al.* (2010).

Alternatively, under specific circumstances, bioorthogonal reactions can be used to site-specifically photocage a protein residue (Fig. 1.15). Cysteines are the most nucleophilic amino acids, and can be targeted by electrophilic photocaging reagents, such as α-bromo methyl derivatives (**22**). Cysteines can also be selectively modified to dehydroalanines (**24**) post-translationally (Chalker *et al.*, 2011), and further modified by a thiol containing photocage (**25**), such as 2-nitrobenzenemethanethiol or 4-thiomethylcoumarin (Akita *et al.*, 2005).
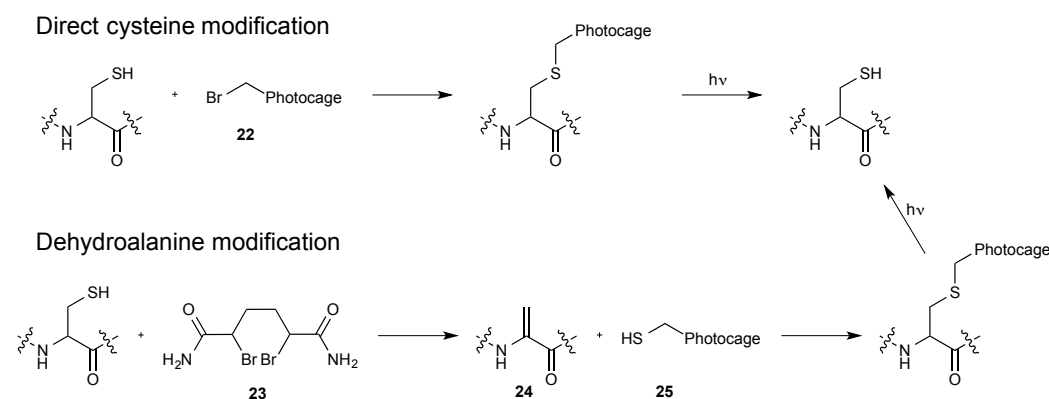
Fig. 1.15 Direct bioorthogonal modification of surface cysteine residues with photocaging reagents that can be cleaved to yield the free cysteine. The cysteine residue can be reacted with an electrophilic bromo-methyl reagent (top, **22**) or transformed to a dehydroalanine (**24**) using 1,4-dibromoadipamide (**23**) and reacted with a thiol containing reagent (bottom, **25**).

A final approach to making chemically modified proteins is native chemical ligation, where the protein is produced (semi-)synthetically (Fig. 1.16) (Bang and Kent, 2004, Durek *et al.*, 2007). The part of the protein polypeptide chain bearing the desired photocaged residue and a C-terminal thioester (**26**) is synthesised chemically (Blanco-Canosa and Dawson, 2008) and reacted with a second chemically synthesised or expressed peptide bearing an N-terminal cysteine (**I**, **27**). The thioester ligated product (**28**) then undergoes an S-N acyl shift (**II**), to yield the full protein chain carrying the unnatural amino-acid (**29**). The procedure is performed under denaturing conditions and the protein has to be folded *in vitro*, so this approach is not be suitable for a wide range of proteins.
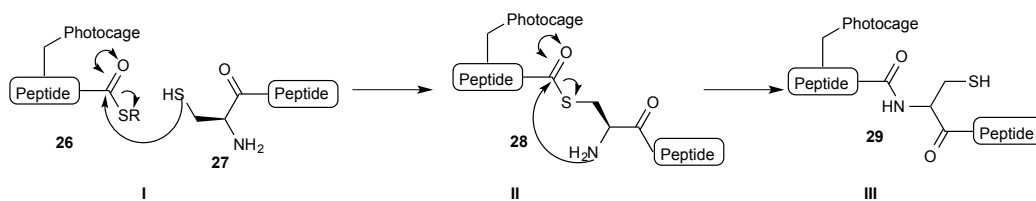
Fig. 1.16 Mechanism of native chemical ligation. A peptide carrying a C-terminal thioester group (**26**) is coupled to a second peptide bearing an N-terminal cysteine residue (**27**). The cysteine side chain attacks the thioester group (**I**), forming a thioester-ligated peptide (**28**), which undergoes an S-N acyl shift (**II**) to the desired full length peptide chain (**29**).

Several classes of proteins have been photocaged using the different methods described above, including:

- Protein kinases, by the direct cysteine chemical modification (Chang *et al.*, 1998), introduction of a photocaged thiophosphate moiety at an active site mutation T197C (Zou *et al.*, 2002) or the genetic incorporation of a photocaged lysine at the ATP binding site (Gautier *et al.*, 2011)

- Cholinesterase, by modification with a photocleavable covalent inhibitor (Loudwig *et al.*, 2003)

- Lysozyme, by the incorporation of photocaged aspartate at the active site, using *in vitro* protein expression (Mendel *et al.*, 1991)

- Ion channels, by genetic incorporation of a caged tyrosine (Tong *et al.*, 2001)

- Immunoglobulin, by residue-specific lysine derivatisation (Kossel *et al.*, 2001)

- Inteins, by the genetic incorporation of a photocaged cysteine (Ren *et al.*, 2015).

In all these examples, the modified proteins were shown to be de-activated (fully or exhibiting only very low residual activity) and activity could be recovered upon photolysis.

Summary

Photocaging of biological systems has been shown and used for the past few decades. There is enormous potential for the use of *o*NB, coumarin and *p*HP groups in the caging of proteins, substrates and ligands for efficient reaction triggering for time-resolved biophysical studies. Nevertheless, targeting protein-ligand binding-sites is challenging. The portfolio of current chemical probes for chemical modification of proteins is extensive, but residue-specific modifications with photocleavable groups that can be photolysed releasing the original amino acid side chain are limited to the chemical modification of cysteines or the genetic incorporation of the unnatural amino-acid. Labelling of ligands or substrates is much more widely applicable, but the synthesis of

the modified compounds can be very challenging, especially since many of the cofactors and ligands essential for protein function contain stereocenters and multiple functional groups.

The instrumentation and data analysis software necessary for time-resolved structural studies are continuously and rapidly evolving and have proven to be sufficiently powerful to allow for these studies to be successfully performed. Reaction triggering is clearly now the major bottleneck hindering the broader use of time-resolved techniques. Making protein photocaging technology widely applicable would relieve this bottleneck and greatly decreasing the complexity of the design of such experiments.

2   Project overview, objective and impact

The aim of this project was to design photocaging reagents that could become widely applicable for the photocaging and photo-triggering of proteins, regardless of the nature of the protein, ligands or cofactors, by decoupling the reaction-triggering chemistry from the protein processes being observed during the time-resolved experiments. In practical terms, this means moving the photocaging moiety out of the protein active site, *i.e.* neither photocaging essential amino acids involved in function nor the substrate or ligands.

Making a general reagent would decrease the complexity of the chemical synthesis currently involved in making specific photocaged substrates or ligands. Also, by making use of surface-cysteines as the target for bioconjugation of the photocaging reagents, the steps for protein modification would be simplified to the introduction of bioconjugational moieties by site-directed mutagenesis rather than requiring complex genetic incorporation of unnatural moieties.

Instead of targeting active site groups, the new photocaging reagents developed here are intended to make use of the intrinsic protein dynamics known to be involved in function. Proteins are plastic and therefore adopt multiple conformations, some of which are active and some inactive. Using crosslinking compounds, dynamic regions can be rigidified, locking the protein in an inactive state. Release of this conformational "lock" by photolysis would then allow the protein to resume its normal dynamic behaviour and resume activity.

The proposed novel approach for protein photocaging presented in this thesis is depicted in Fig. 2.1. The crosslinking reagents can be regarded as "conformational staples" which target engineered residues on the protein surface rendering the protein inactive, but retaining a substrate binding conformation. Upon cleavage of the "staple", the protein is able to relax to its active conformation and on to all the conformational states required during function.
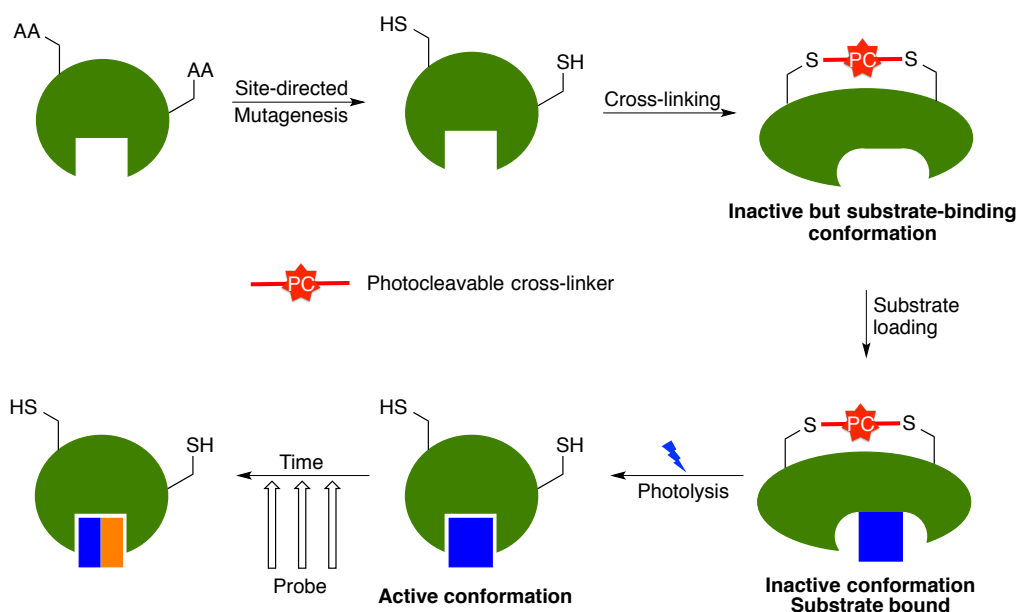
Fig. 2.1 Overview of the proposed new protein photocaging approach. Surface cysteine residues are introduced to the target protein by site-directed mutagenesis. The photocleavable crosslinker targets available cysteines specifically, "stapling" the protein in an inactive conformation. Upon photolysis, the crosslinker is cleaved and the protein allowed to resume function. Photolysis acts as the trigger which precedes the time-resolved measurements.

A second advantage of this "allosteric" approach compared to traditional active-site photocaging approaches is that it removes the interference of the photocleavage by-products with the system once it is triggered. The most widely used photocaging moieties (coumarin and ortho-nitrobenzyl groups) are of similar size to the side chains of the largest natural amino acids (tryptophan or tyrosine). The presence of such large fragments in the protein's active site post-photocleavage can alter the conformations adopted by the protein or even block protein function until the by-product has diffused out of the active site, leaving the triggering under diffusion control and abolishing all efforts for fast synchronisation of the system.

Aspartate α-decarboxylase (ADC) was chosen as the model protein for the crosslinking and cleavage studies of the photocleavable "stapling" reagents developed, as it has been well studied within the Webb group and can be overexpressed and purified easily (Webb *et al.*, 2004, Webb *et al.*, 2014). One of the main aim of this project was to make the technology available for time-resolved diffraction studies as well as solution studies, ADC has also been reproducibly crystallised within the Webb and Pearson groups, yielding consistently well diffracting crystals. The main disadvantage of using this system is that it requires a post-translational modification to become fully active. For X-ray diffraction time-resolved studies, it is important to obtain fully active, homogeneous

protein for crystallisation and so understanding protein post-translational activation became an additional essential objective of this project.

This thesis is therefore split into three main sections:

- Understanding ADC activation in order to obtain clean, homogeneous and fully active protein
- The design and synthesis of a novel class of protein photocleavable crosslinking reagents
- The crosslinking and photo-release of ADC using the developed crosslinkers.

3 A molecular understanding of aspartate α-decarboxylase activation

3.1 Introduction

3.1.1 Pantothenate biosynthesis

Pantothenate (vitamin B_5) is an essential vitamin synthesised by a common pathway in bacteria, yeast and plants. The β-alanine biosynthetic pathway for incorporation into pantothenate differs between organisms, and only that of bacterial will be discussed here. In bacteria, β-alanine (**30**) is obtained from the decarboxylation of L-aspartate (**31**) by aspartate α-decarboxylase (ADC) (Fig. 3.1).
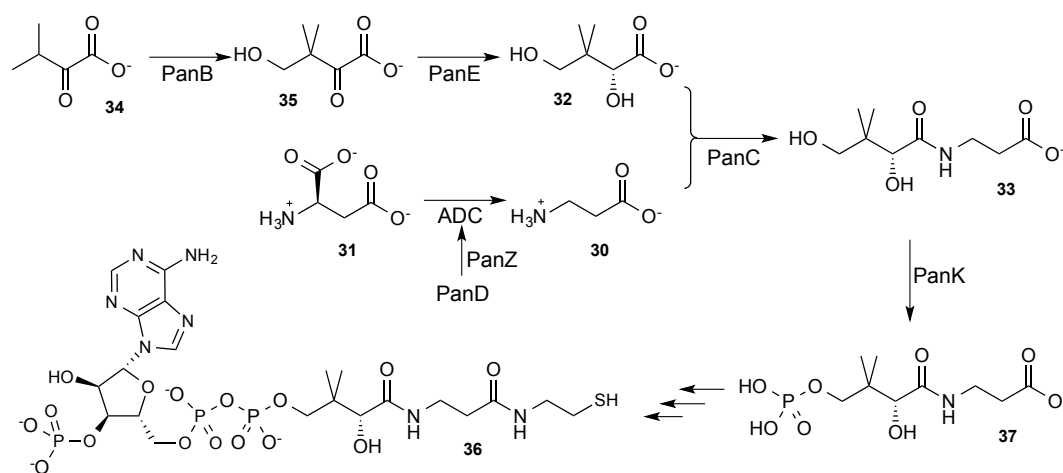
Fig. 3.1 The biosynthetic pathway of pantothenate, showing the biosynthetic intermediates and enzymes involved in each step. α-ketoisovalerate (**34**) is hydroxymethylated to ketopantoate (**35**), which is reduced to D-pantoate (**32**). L-aspartate (**31**) is decarboxylated to β-alanine (**1**) by ADC. D-pantoate (**32**) is condensed with β-alanine (**30**) to pantothenate (**33**), which is phosphorylated to phosphopantothenate (**37**). Phosphopantothenate regulates the biosynthesis of CoA (**36**) by a negative feedback mechanism. The PanD zymogen matures to its active form ADC in the presence of PanZ.

β-Alanine is then condensed with D-pantoate (**32**) (by pantothenate synthetase, PanC) to give pantothenate (**33**). α-Ketoisovalerate (**34**) is hydroxymethylated to ketopantoate (**35**) by ketopantoate hydroxymethyl transferase (PanB), which is then reduced to D-pantoate (**32**) by ketopantoate reductase (PanE). As these last two reaction steps are reversible and α-ketoisovalerate is also a precursor in L-valine biosynthesis (Maas and Vogel, 1953), the decarboxylation of L-aspartate to β-alanine becomes the first committed step towards pantothenate synthesis in bacteria, and onwards to the synthesis of coenzyme A (CoA, **36**) and its thioesters (Fig. 3.2). Pantothenate is phosphorylated to phosphopantothenate (**37**) by pantothenate kinase (PanK), which is inhibited by CoA,

creating a negative feedback loop mechanism for the regulation of CoA biosynthesis (Rock *et al.*, 2003).
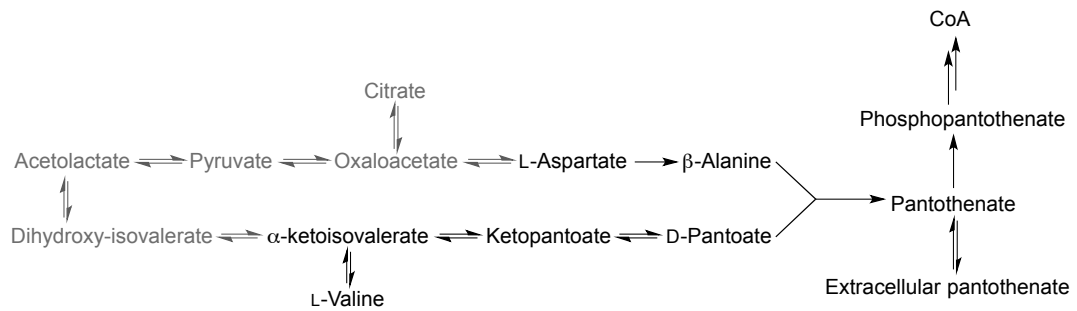
Fig. 3.2 The relationship between the biosynthetic pathways of pantothenate and L-valine. The precursors to both L-valine and L-aspartate are in equilibrium in the cell. The decarboxylation of L-aspartate to β-alanine is irreversible and, therefore, the first committed step in the biosynthesis of pantothenate and CoA.

3.1.2 PanD activation to ADC

ADC was first purified in 1979 (Williamson and Brown, 1979). It was shown to be involved in β-alanine biosynthesis in *E. coli* and contains a covalently linked, catalytically active, pyruvoyl group. Pyruvoyl-dependent enzymes (van Poelje and Snell, 1990) are a small subset of enzymes which include various decarboxylases such as histidine decarboxylase (Snell, 1986) and S-adenosylmethionine decarboxylase (Pegg, 2009). The pyruvoyl is synthesised post-translationally by the rearrangement and cleavage of the protein backbone. ADC is expressed as an inactive pro-protein, PanD, which consists of a single chain, the π-chain, and assembles into a homotetramer. The post-translational modification starts with an N-O acyl shift between Gly24 and Ser25 (Fig. 3.3). The hydroxyl group of Ser25 attacks the carbonyl carbon of Gly24 (**I**), forming an oxyoxazolidine ring (**II**), which ring-opens to an ester intermediate (**III**) (Albert *et al.*, 1998). The ester then cleaves by β-elimination, to a C-terminal glycine (Gly24) and a dehydroalanine (**IV**), which then hydrolyses to the desired N-terminal pyruvoyl group (**V**). The resulting two peptide chains are termed the β-chain ([1]Met-Gly[24]) and the α-chain ([25]Pyr-Ala[126]).

The PanD residues essential for activation of PanD to ADC have been explored by site-directed mutagenesis (Gelfman *et al.*, 1991, Schmitzberger *et al.*, 2003, Webb *et al.*, 2012, Webb *et al.*, 2014). As expected, mutation of Ser25 to alanine yields a non-activatable PanD mutant, while other mutations along the activation peptide greatly decrease the processing ability of ADC, likely due to distortions of the conformation of the loop. In the zymogen crystal structure, the carbonyl carbon of Gly24 is seen hydrogen-bonding to the fully conserved Thr57 side-chain (Schmitzberger *et al.*, 2003).

This interaction may act to polarise the carbonyl group, aiding the nucleophilic attack by Ser25, and also stabilise the oxyoxazolidine ring intermediate (**II**). Site-directed mutation of Thr57 to valine yields a completely non-activatable PanD mutant (Webb *et al.*, 2014). Thr57 is the only proximal residue besides Gly24 and Ser25 found to be essential for PanD activation.
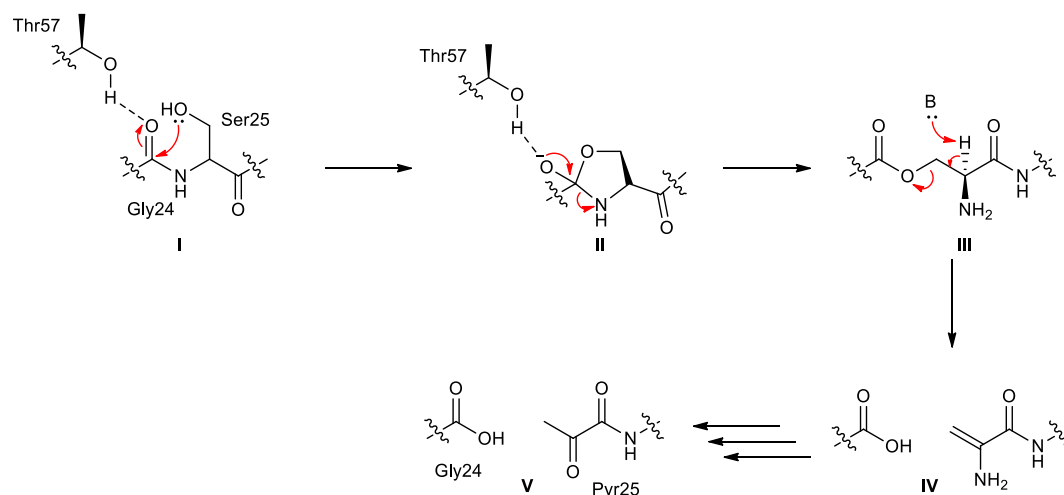

Fig. 3.3 Accepted mechanism for the activation of PanD to ADC at the start of this project (Webb *et al.*, 2004). The Ser25 hydroxyl side-chain attacks the backbone carbonyl of the adjacent residue (Gly24), which is hydrogen-bonded to Thr57 in the zymogen crystal structure (PDB 1PPY, **I**) (Schmitzberger *et al.*, 2003). The oxyoxazolidine ring opens (**II**) to the ester intermediate (**III**), which can then cleave by elimination to a dehydroalanine intermediate (**IV**). The dehydroalanine residue can then hydrolyse to the required pyruvoyl group, giving the fully active enzyme (**V**).

Other residues, such as Tyr58, His11 and Asn72 (Webb *et al.*, 2012) have also been investigated, but in all cases PanD activation was still observed (Schmitzberger *et al.*, 2003). This was surprising as it was expected that other residues would be involved as general acids or bases for the deprotonation of Ser25 (**I**), protonation of the amine during opening of the oxyoxazolidine ring (**II**) and β-elimination of the ester (**III**). At the time this study began, it was thought that the water molecules surrounding the activation site might act as general bases or that a large conformational rearrangement might be required for activation, which could involve more distal residues, not yet investigated by mutagenesis, in the mechanism.

Post-translational rearrangements leading to backbone cleavage have been observed in other systems, such as the processing of inteins (Fig. 3.4, **A**) (Paulus, 2000), activation of N-terminal nucleophile (Ntn) hydrolases (Kim *et al.*, 2006, Buller *et al.*, 2012) and the autoproteolysis of SEA domains (Fig. 3.4, **B**) (Levitin *et al.*, 2005, Johannson *et al.*, 2009). In most protein systems, post-translational cleavage or rearrangement of the backbone is auto-catalytic (Paulus, 2000), with the exception of a small subset of

proteins, for which the rate of modification is accelerated by accessory proteins (McDonald *et al.*, 1998, Trip *et al.*, 2011).
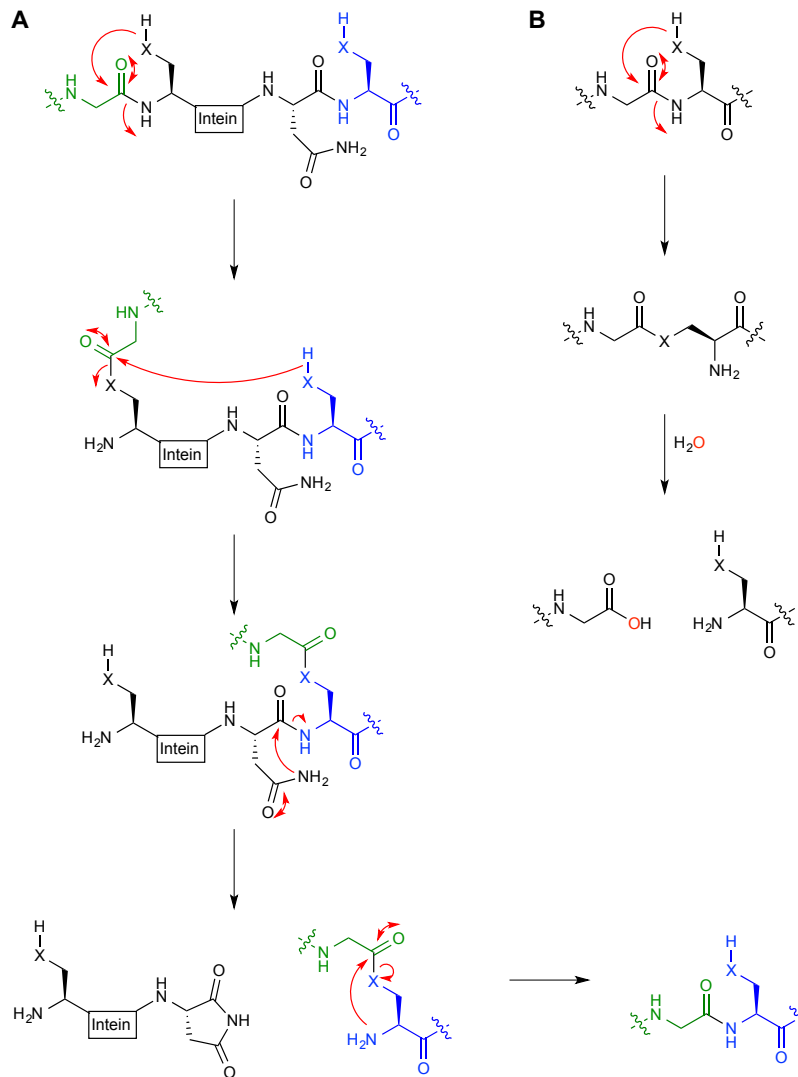
Fig. 3.4 General mechanism of post-translational cleavage of Inteins (**A**) and other systems, such as Ntn hydrolases and SEA domains (**B**). X can be oxygen (Ser or Thr residues) or sulphur (cysteine residue).

ADC is such a case, requiring an accessory protein is required for its full maturation. Overexpressed *E. coli* PanD activates slowly *in vitro* at 37 °C. The protein has to be incubated overnight at this temperature before significant levels of protein cleavage are observed (Nozaki *et al.*, 2012). Furthermore, thermal activation of PanD leads to the formation of both the desired pyruvoyl cofactor, as well as a N-terminal serine by-product, from the direct hydrolysis of the ester intermediate (Fig. 3.3, **II**) (Ramjee *et al.*, 1997). The presence of this by-product is clearly visible in the electron density of the published ADC crystal structure (PDB 1AW8) (Albert *et al.*, 1998) and it has also been commonly found in other structures solved during the course of this investigation (i.e.

PDB 4ANO, Yorke, B.A. & Monteiro, D.C.F, unpublished results), but which will not be discussed in this thesis. Nevertheless, the formation of the serine by-product is a major problem with regard to the production of homogeneous and fully activated enzyme for biophysical studies and even more so for time-resolved structural studies, where the signal-to-noise ratio as well as the interpretation of the data can be challenging. Therefore, understanding the molecular basis of activation of PanD to ADC will allow for both a more in-depth understanding of the system as well as providing the means for the design of a robust protocol for the isolation of the fully active enzyme.

The slow thermal activation suggests that an activating factor that accelerates PanD maturation to ADC (with clean pyruvoyl group formation) must be present *in vivo* (Ramjee *et al.*, 1997). Despite this, for decades the post-translational rearrangement of ADC, and all other pyruvoyl-dependent enzymes, has been presumed to be autocatalytic. However, recently the accessory proteins for the activation of pyruvoyl-dependent histidine decarboxylases, HdcB (Trip *et al.*, 2011), as well as for the activation of PanD, PanZ (Nozaki *et al.*, 2012, Stuecker *et al.*, 2012a, Stuecker *et al.*, 2012b) have been discovered.

3.1.3   PanZ: the activating factor for PanD maturation *in vivo*

The gene *panZ* was first identified in 1990, in a region of the *E. coli* genome containing the genes for branched amino-acid transporters. The gene is located downstream of a region coding for proteins involved in branched chain amino acid transport but its deletion leads to pantothenate auxotrophy (Adams *et al.*, 1990). PanZ, also known as YhhK, was subsequently found to be indirectly linked to the production of β-alanine, as mutations to the gene (in *S. typhimurium*) (Primerano and Burns, 1983) or its knockout (in *E. coli*) (Nozaki *et al.*, 2012) cause β-alanine auxotrophy (Fig. 3.5). The mechanism of this remained unclear as all the enzymes involved in the pantothenate biosynthetic pathway (Fig. 3.1 PanB, PanE, PanC and ADC) had been overexpressed and purified and shown to catalyse each of the steps of the pathway without the need for further accessory proteins (Webb *et al.*, 2004).

	None	L-Aspartate	D-Pantoate	β-Alanine	D-Pantothenate	

wild-type
Δ*panZ*
Δ*panZ*, vector
Δ*panZ*, His-PanZ
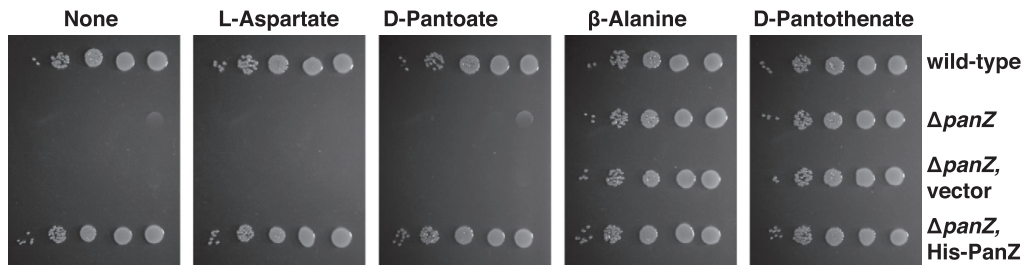
Fig. 3.5 Growth defects of Δ*panZ E. coli* cells and growth recovery with hexaHis-tagged PanZ complementation, adapted from (Nozaki *et al.*, 2012). Cell cultures were grown overnight in M9 glucose medium containing the indicated additives and show that Δ*panZ* cells are β-alanine auxotrophs.

A possible role for PanZ was proposed to be in PanD maturation. To test this hypothesis, Nozaki *et al.* investigated the role of PanZ in PanD activation *in vivo* by examination of the extent of PanD cleavage in Δ*panZ* cells with and without complementation with PanZ (Fig. 3.6) (Nozaki *et al.*, 2012). Δ*panZ* cells showed no PanD activation, but complementation of the PanZ knockouts with a plasmid harbouring hexaHis-tagged PanZ recovered PanD cleavage, indicating that PanZ is the accessory protein necessary for efficient PanD activation *in vivo* (Fig. 3.6).
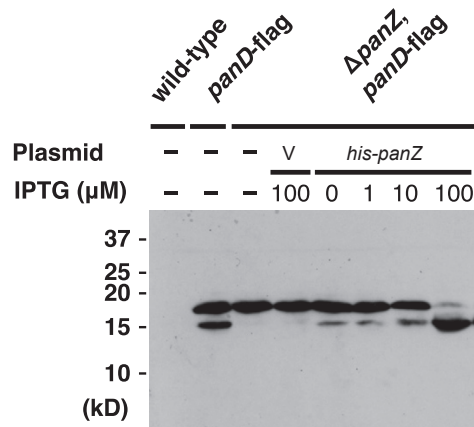
Fig. 3.6 SDS-PAGE analysis of the activation of PanD *in vivo* showing the requirement of PanZ for cleavage (adapted from (Nozaki *et al.*, 2012). PanD is detected by imunoblot analysis using an anti-FLAG antibody. PanD expressed in Δ*panZ* cells is not activated. Cleavage is restored with the expression of PanZ by induction with IPTG.

The PanZ sequence was annotated as a putative acyl-transferase as it carries a CoA binding motif. The 3D structure of PanZ was determined by NMR by J. Cort *et al.* (unpublished data, PDB 2K5T), confirming the binding of CoA. Nevertheless, no evidence of PanZ acetyltransferase activity could be observed using a radiometric assay with [1-^{14}C]acetyl coenzyme A and site-directed mutagenesis of PanD-lysine residues did not affect PanD maturation (Stuecker *et al.*, 2012a).

3.1.4   β-Alanine biosynthesis and CoA regulation

A possible role of PanD in CoA biosynthesis regulation was first proposed in 1982 by Cronan and co-workers (Cronan *et al.*, 1982). Supplementation of *E. coli* cultures with β-alanine led to increased cellular levels of CoA. Pantothenate supplementation did not have as strong an effect on CoA biosynthesis as β-alanine, and therefore β-alanine supply was proposed to regulate CoA production. Nevertheless, a mechanism for such a regulatory role of ADC was not identified. The only regulatory step for CoA biosynthesis that has been so far identified and characterised is the phosphorylation of pantothenate (Fig. 3.1, **33**) to phosphopantothenate (Fig. 3.1, **37**) by pantothenate kinase, which is allosterically regulated by CoA (Yun *et al.*, 2000, Rock *et al.*, 2003). No other regulatory mechanisms involved in the production of earlier metabolites on the route to pantothenate had been identified when this project began.

3.2   Results and Discussion

The main aim of this project was to elucidate the mechanism by which PanZ promotes PanD maturation *in vivo*, as ADC was the test system chosen for the testing and application of the novel photocleavable crosslinkers. Obtaining a cleanly activated, homogeneous enzyme preparation is as important as adopting the correct caging strategy or collecting and interpreting the data in time-resolved studies.

The PanD.PanZ interaction was investigated using a range of biophysical techniques, including X-ray crystallography and isothermal titration calorimetry (ITC). The results obtained allowed the proposal of a revised mechanism of activation of PanD to ADC. They also provided insight into a new mode of regulation of CoA biosynthesis in bacteria and the observation of previously unreported X-ray-induced protein chemistry. The methods used, observations and results are described in the following sections.

3.2.1   Protein overexpression and purification

ADC is a tetrameric enzyme, each 13.8 kDa protomer consisting of a six-stranded, double-ψ β-barrel. The overall complex is 50 Å in diameter and doughnut shaped, with a large central pore open to solvent. The four active sites are located at the protomer-protomer interfaces. As discussed in the introduction to this chapter, the PanD(T57V) site-directed mutant does not undergo the post-translational modification that leads to activation to ADC. This mutant was therefore chosen for the biophysical characterisation of the PanD.PanZ protein complex to allow for a clear investigation of the protein-protein interaction with no interference from the activation reaction.

PanD(T57V) is expressed as the intact π-chain and does not undergo *in vivo* activation or thermal cleavage. The structure of this site-directed PanD mutant has recently been reported, showing the intact and unmodified protein backbone in the activation region (Webb *et al.*, 2014).

PanD(T57V) and PanZ were overexpressed and purified as described in sections 8.5 and 8.6. Details regarding the optimisation of protein overexpression, SDS-PAGE analysis of the purified proteins and the size-exclusion chromatography traces can be found in section 9.1. In summary, N-terminally hexaHis-tagged PanD(T57V) was overexpressed in auto-induction media from previously transformed *E. coli* MG1655 Δ*panZ* Δ*panD* (DE3) cells and purified by nickel affinity chromatography and size exclusion chromatography. C-Terminally hexaHis-tagged PanZ (Nozaki *et al.*, 2012) was initially overexpressed from *E. coli* BL21 Gold (DE3) cells using arabinose induction in the presence of pantothenate. The requirement for pantothenate during overexpression was later rationalised as overexpression of PanZ leads to β-alanine auxotrophy (see section 3.2.6 for further discussion). N-Terminally and C-terminally hexaHis-tagged PanZ (Cort, 2009) was subcloned by Chris Bartlett (Wellcome Trust rotation student) into the pET28a vector for overexpression by auto-induction in *E. coli* MG1655 Δ*panZ* Δ*panD* (DE3) cells. The N-terminally tagged variant was found to be unable to interact with PanD. The inhibition of the interaction by the N-terminal hexaHis-tag was later rationalised, as the PanZ Met1 residue is involved in the protein-protein interface (see section 3.2.5 for details). Therefore, only the C-terminally tagged PanZ was used during this project.

3.2.2 Calorimetry studies of PanZ.Coa and PanD(T57V).PanZ

The binding affinity of PanZ for CoA was investigated by ITC. CoA was titrated into two different PanZ samples, from different purification runs (Fig. 3.7). The titrations show that ~60% of the protein is purified with CoA already bound and that the K_D for the interaction is approximately 6.2 μM. The ability of PanZ to bind CoA is consistent with both the published structure of PanZ by NMR (PDB 2K5T), which shows a PanZ.CoA complex, and the initial annotation of the PanZ sequence as a putative acyl transferase. Dialysis of the protein against *SEC buffer A* allowed for partial removal of CoA, as seen by the decrease in the incompetent fraction for a CoA titration into PanZ (Fig. 3.7 C). Full removal of CoA from the sample by longer dialysis times caused precipitation of the protein.
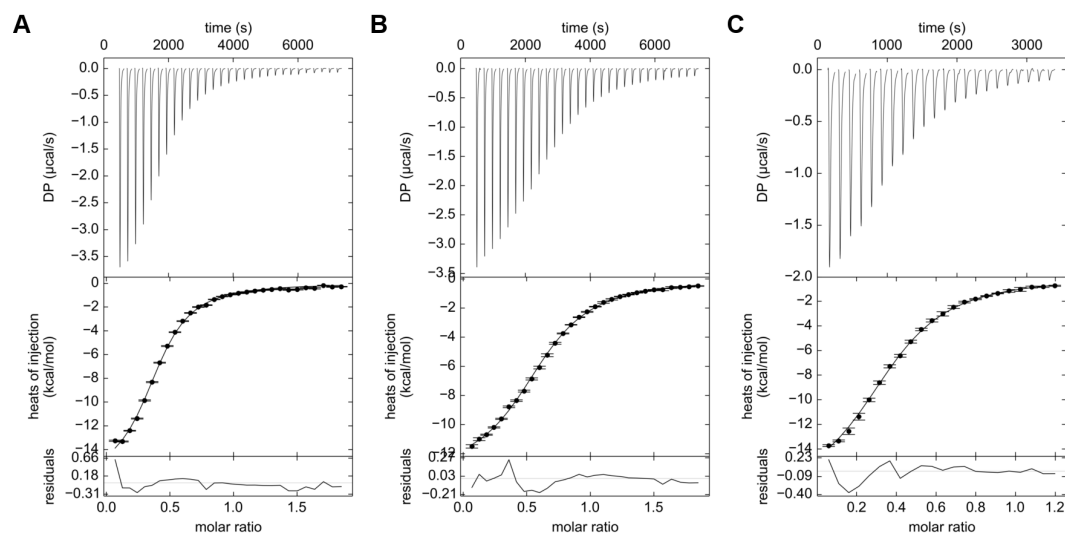
Fig. 3.7 Titration of CoA into PanZ. All samples were in 50 mM Tris (pH 7.5), 100 mM NaCl and 0.1 mM DTT (*SEC buffer A*). (**A**) and (**B**) titration of 800 μM CoA into two different 100 μM PanZ samples, showing 0.60 and 0.62 equivalents of CoA co-purified with the protein and K_D values of 6.2 ± 0.5 μM and 5.2 ± 0.2 μM respectively. (**C**) titration of 400 μM CoA into dialysed 57 μM PanZ showing a K_D of 9.40 ± 0.4 μM.

The PanD.PanZ interaction was next investigated by ITC. It was expected that the PanD.PanZ interaction would compete with CoA binding to PanZ as this would be consistent with a model in which ADC activation is negative feedback regulated by the concentration of CoA in the cell. In this model, at high CoA concentrations CoA would compete with PanD for PanZ binding, lowering the amount of fully matured ADC in the cell. In turn, less β-alanine (and subsequently CoA) would be produced, down-regulating the concentration of CoA in the cell. But, contrary to expectation, the titration of PanD(T57V) into PanZ showed a binding coefficient equal to the fraction of PanZ.CoA rather than *apo*-PanZ (PanD protomer to PanZ monomer, Fig. 3.8, **A**). Saturation of PanZ with CoA showed a clear upwards shift in the binding stoichiometry (Fig. 3.8, **B**), suggesting that the PanD(T57V).PanZ interaction is CoA dependent, rather than competitive (Monteiro *et al.*, 2012). The binding affinity between the two proteins in the presence of CoA was found to be 240 nM.

The protein-protein interaction was also investigated by ITC in the presence of acetyl CoA (AcCoA) instead of CoA, in order to improve the consistency of the titrations. Thermodynamic parameters from ITC varied slightly between runs and it was hypothesised that CoA may form disulphide-linked dimers in solution, disrupting the protein-protein interaction. Acetyl CoA is unable to form such dimers, and was therefore used in place of CoA in all subsequent experiments. The subsequent

PanD(T57V).PanZ titrations performed in the presence of AcCoA were more accurate, reproducible and showed a binding stoichiometry closer to 1 (Fig. 3.9).
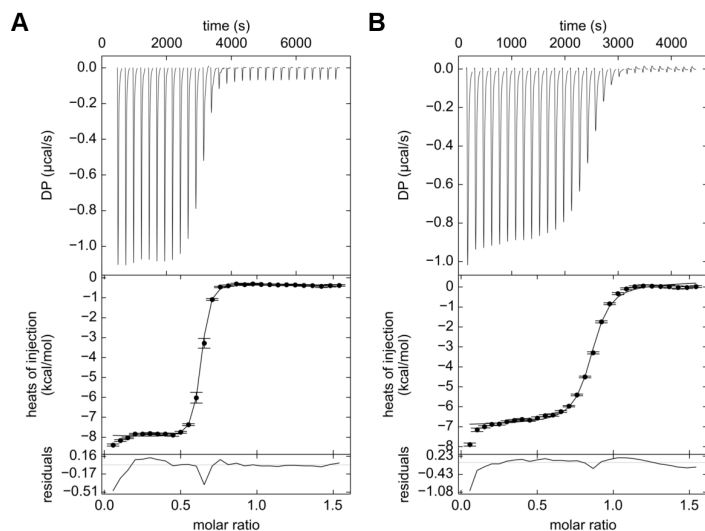
Fig. 3.8 ITC traces for the titration of (**A**) 400 μM PanD(T57V) (by protomer) into 60 μM PanZ without excess CoA and (**B**) 394 μM PanD(T57V) into 59 μM PanZ in the presence of 100 μM CoA. All samples were in 50 mM Tris (pH 7.5), 100 mM NaCl and 0.1 mM DTT (*SEC buffer A*). In the presence of excess CoA, the binding stoichiometry increases, indicating that the interaction is mediated by CoA. The K_D for the protein-protein interaction in the presence of CoA was calculated to be 241 ± 38 nM.
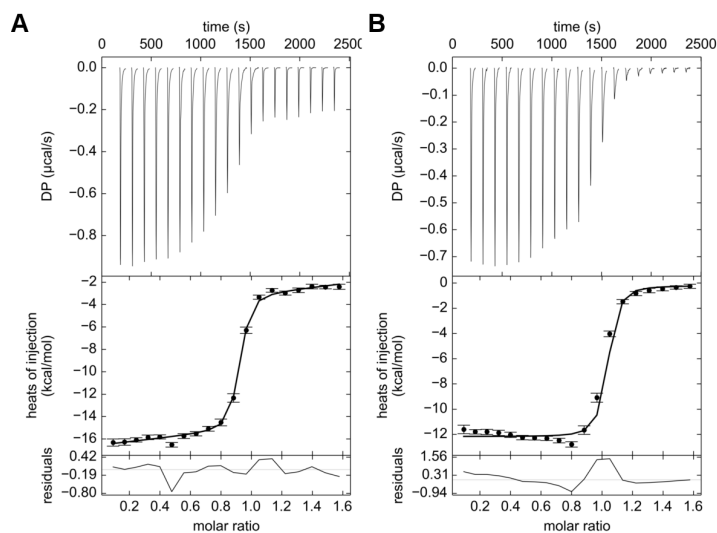
Fig. 3.9 Pairwise titration of PanD(T57V) into PanZ in the presence of 394 μM AcCoA (excess). All samples were in 50 mM Tris (pH 7.5), 100 mM NaCl and 0.1 mM DTT (*SEC buffer A*). (**A**) 263 μM PanD(T57V) (protomer concentration) titrated into 35 μM PanZ. (**B**) 263 μM PanZ titrated into 35 μM PanD(T57V) (protomer concentration). The K_D for the protein-protein interaction was found to be 35 ± 9 nM from globally fitting both titrations.
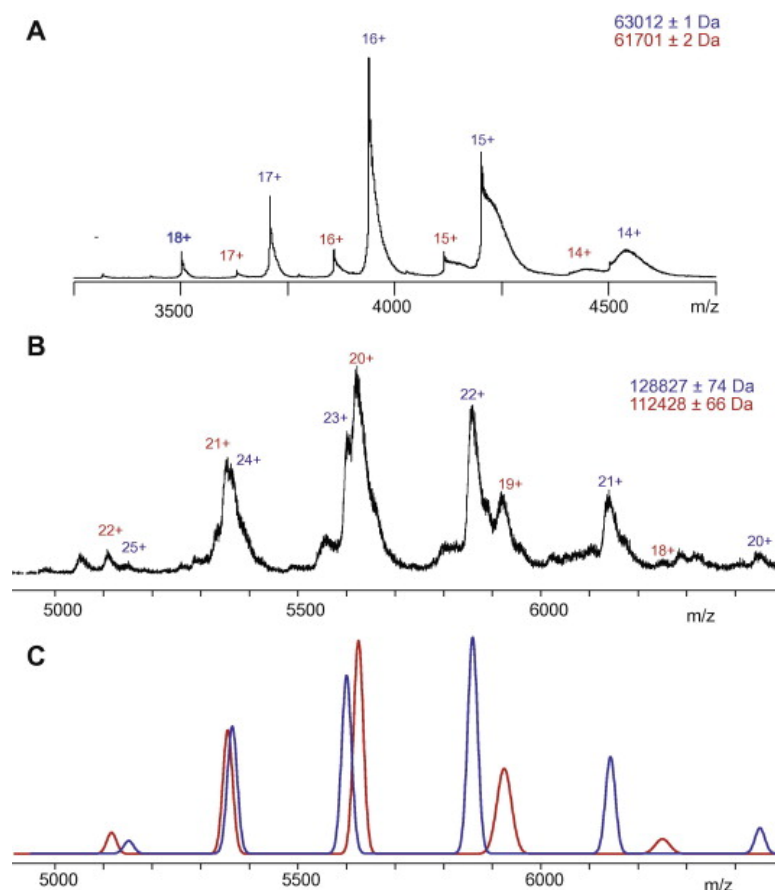
3.2.3 The PanD(T57V).PanZ interaction is CoA dependent

To confirm the stoichiometry of the PanD(T57V).PanZ complex and unambiguously determine the dependence upon CoA for interaction, both mass-spectrometry and NMR experiments were conducted.

Electrospray mass spectrometry (ESMS) was used to observe three protein complexes in the gas phase: the PanZ.CoA complex, the PanD(T57V) homotetramer and the PanD(T57V).PanZ.CoA complex (Monteiro *et al.*, 2012). Dr Dale Shepherd collected the mass spectrometry data and assisted with interpretation of the spectra. The PanZ sample containing 1 eqv. of CoA showed three species corresponding to *apo*-PanZ, PanZ.CoA and PanZ.2CoA (15.614, 16.381 and 17.148 kDa respectively). This revealed that the protein retained sufficient tertiary structure to bind CoA in the gas phase. The ternary species was rationalised as a possible CoA dimer, from disulphide formation. The PanD(T57V) sample revealed two tetrameric species corresponding to the PanD(T57V) homotetramer and a smaller heterotetrameric species, thought to correspond to partial proteolysis of PanD(T57V) (63.0 and 61.7 kDa respectively). This hypothesis was confirmed by MS-MS to dissociate the heterotetramer into its constituents: the full protein (15.753 kDa) and a truncated polypeptide (14.442 kDa).

Finally, the PanD(T57V).PanZ heterooctamer was investigated in the presence of excess CoA. Once again, two species were visible in the spectrum, corresponding to the PanD(T57V) tetramer bound to 3 and 4 equivalents of PanZ.CoA (128.8 kDa and 112.43 kDa respectively). The 4:3 heteroheptameric species was thought to be correlated with the contamination of the PanD(T57V) sample with the truncated polypeptide. A simulated spectrum of the 4:4 and 4:3 PanD(T57V).PanZ.CoA species was compared to the experimental data and found to be in good agreement. The data supported the hypothesis that the PanD(T57V).PanZ complex only forms in the presence of CoA (Monteiro *et al.*, 2012).

Fig. 3.10 ESMS studies of the PanD(T57V).PanZ interaction. (**A**) PanD(T57V) sample, showing two species corresponding to the expected homotetramer (63.012 kDa, blue) and a smaller heterotetramer (61.701 kDa, red). (**B**) PanD(T57V).PanZ.CoA sample in the presence of excess CoA, showing two species corresponding to PanD(T57V) with 4 PanZ.CoA (blue) and 3 PanZ.CoA (red) species bound (128.827 kDa and 112.428 kDa respectively). (**C**) simulated spectrum of the PanD(T57V).4PanZ.4CoA and PanD(T57V).3PanZ.3CoA complexes, matching the experimentally obtained spectrum in (**B**).

To confirm the absolute requirement of CoA for the PanD(T57V).PanZ protein-protein interaction, an NMR experiment was devised. As the PanD(T57V).PanZ.CoA complex is ~120 kDa, it was expected that, upon binding of the two proteins, the slow tumbling rate of the resulting complex would cause severe broadening of the NMR signals, effectively making the spectrum disappear. Chris Bartlett (Wellcome trust PhD rotation student) overexpressed ^{15}N-labelled PanZ from minimal media containing ^{15}NH$_4$Cl, purified the protein by nickel-affinity chromatography and SEC as described previously (Monteiro *et al.*, 2015) and recorded four different ^{1}H-^{15}N HSQC NMR spectra, shown in Fig. 3.11: (**A**) a spectrum of the sample obtained from the purification of the protein, which shows multiple species in solution, confirming the presence of *apo* and CoA-bound PanZ fractions; (**B**) a spectrum of PanZ in the presence of excess CoA, showing a single species corresponding to PanZ.AcCoA; (**C**) a spectrum of the *apo-*

37

PanZ/PanZ.AcCoA sample mixed with 1 eqv. of PanD(T57V) (PanD protomer to PanZ monomer), showing a different spectrum to that of PanZ.AcCoA, therefore corresponding to the *apo*-PanZ fraction of the sample; and (**D**) a spectrum of AcCoA-saturated PanZ mixed with 1 eqv. of PanD(T57V) (PanD protomer to PanZ monomer), showing obliteration of the PanZ protein signals and full complex formation.
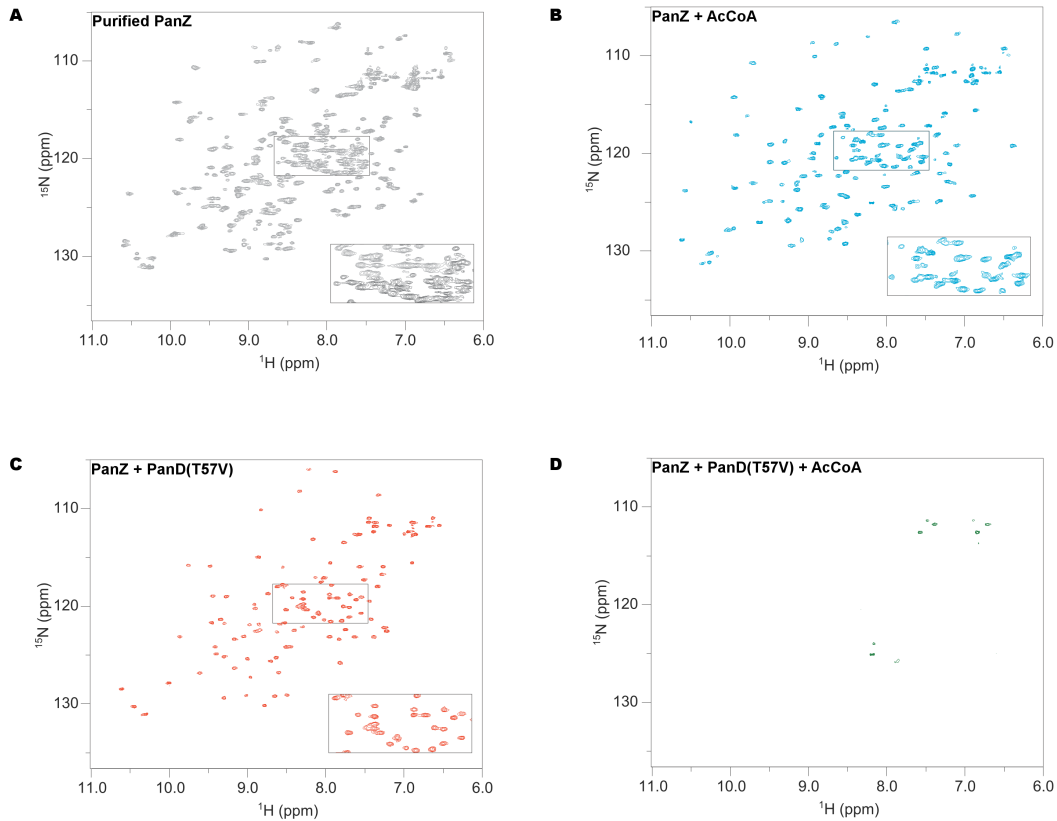
Fig. 3.11 NMR spectra of ^{15}N-labeled PanZ (0.2 mM in *SEC buffer A*). Purified protein (grey) showed multiple species in solution, PanZ.AcCoA and *apo*.PanZ. Addition of AcCoA simplifies the spectrum (blue) to a single species, PanZ.AcCoA. Addition of PanD(T57V) to purified PanZ also simplifies the spectrum (red), but to a different species, *apo*.PanZ. Addition of PanD(T57V) to PanZ.AcCoA causes loss of all signals, indicating full complex formation (green).

The addition of PanD(T57V) to the *apo*.PanZ/PanZ.AcCoA mixture yields a different spectrum to that of PanZ.AcCoA, meaning that only the PanZ.AcCoA fraction is capable of binding to PanD(T57V), causing the disappearance of PanZ.AcCoA signals and leaving a spectrum that corresponds to the unbound *apo*.PanZ fraction. Full spectrum disappearance, and, therefore, complete complexation of PanZ to PanD(T57V) is only observed when PanD(T57V) is added to a AcCoA saturated PanZ sample. This experiment confirmed that full PanD(T57V).PanZ protein complex formation only occurs in the presence of excess AcCoA, proving that the PanD(T57V).PanZ interaction is CoA dependent.

The CoA dependence of PanD binding to PanZ and, therefore, of the activation of PanD to ADC, opened two further questions beyond the understanding of the mechanism of PanD activation by PanZ: what is the role of CoA in the formation of the protein-protein interaction, and whether PanD.PanZ complexation has an effect on CoA regulation *in vivo*. Alongside investigating the mechanism of PanD maturation by PanZ, the role of this protein-protein interaction in the regulation of CoA biosynthesis was also investigated.

3.2.4   The PanD.PanZ protein complex structure

Initial crystallisation screen

PanD(T57V) and PanZ were mixed in a 1:1 (protomer to monomer) ratio in the presence of 2 equivalents of CoA[*]. The protein complex was concentrated to ~4 mg mL^{-1}. A 96 condition sparse matrix crystallisation screen (Crystal Screen I and II, Hampton Research) was set-up and very small crystals were obtained in 200 mM Li$_2$SO$_4$, 100 mM Tris pH 8.5 and 30% *w/v* PEG 4000 (18 °C). The crystallisation conditions were optimised and larger bipyramidal crystals (50 μm) were obtained in 100 mM Tris pH 7.8-8.0, 180-220 mM Li$_2$SO$_4$ and 20% *w/v* PEG 4000 (Fig. 3.12, **A**).

The crystals were cryo-protected by soaking for a few seconds in mother liquor containing 20% *w/v* glycerol, mounted on cryo-loops or meshes and flash cooled in liquid nitrogen. The crystals did not diffract on the in-house source. Preliminary data were collected at beamline I03 at Diamond Light Source (DLS) at 100 K with maximum resolution varying between 9.5 and 4.6 Å (see section 9.2). Data reduction in Xia2 (the automated integration, scaling and merging pipeline at DLS) (Winter, 2010) showed the protein crystallised in space group *I422* (a=b=84.5 Å, c=154 Å, α=β=γ=90°).

Phasing of the data was attempted using both Molrep (Vagin and Teplyakov, 1997) and Phaser (McCoy *et al.*, 2007), using an ADC protomer as a search model (modified PDB 1PPY) (Schmitzberger *et al.*, 2003). In multiple datasets, apparent solutions for the tetramer could be found, some clearly incorrect, as the imposed 4-fold crystallographic symmetry did not yield the ADC tetramer. In all cases, no reasonable electron density was available for the PanZ monomers.

---

[*] The initial crystallisation trials were performed concomitantly with the initial ITC binding assays and, therefore, excess CoA was used instead of AcCoA. The first batch of optimised crystals were obtained with excess CoA, but all the following crystallisation trays were set with excess AcCoA, as this coincided with the confirmation of PanD(T57V).PanZ binding in the presence of AcCoA by ITC.

As an alternative to phasing by molecular replacement, experimental phasing by anomalous diffraction was attempted, with both heavy metal and selenium derivatisation. Firstly, heavy metals were incorporated by soaking the crystals in mother liquor containing Os, W, Pt and Hg salts for a few minutes followed by back soaking for a few seconds. Data were collected at beamline I24 (DLS), but the anomalous signal obtained was insufficient for phasing. Nevertheless, three higher resolution datasets, to 2.5-2.7 Å, were obtained from Hg(OAc)$_2$-derivatised crystals. The crystal symmetry appeared to be $P4_22_12$ with a much larger unit cell (a=b=176.1 Å, c=161.3 Å, α=β=γ=90°), containing 2 PanD(T57V).PanZ.AcCoA[*] octamers in the asymmetric unit cell (the unit cell contents were calculated using the Matthews coefficient calculator in CCP4) (Kantardjieff and Rupp, 2003). Data scaling and merging proved difficult, with bad statistics from Aimless (Evans and Murshudov, 2013). Closer analysis of the automated data reduction outputs from FastDP and Xia2 (Winter, 2010) showed space group ambiguity, with different data reduction strategies outputting varying crystal symmetries ($P422$ and $I422$). Data reduction in alternative space groups was attempted, followed by similar phasing strategies as previously described for the *apo*-crystals, but without success. Data were also collected for crystals derivatised with other metal salts ((NH$_4$)$_{10}$H$_2$(W$_2$O$_7$)$_6$ and K$_2$OsCl$_6$). These had lower diffracting power and had, once again, ambiguous crystal symmetries ($C222_1$ or $P4_22_12$ for tungstate and $P4_22_12$ or $I422$ for osmium).

Following the unsuccessful structure solution by heavy metal derivatisation, selenomethionine (SeMet) labelled PanZ was produced. SeMetPanZ was overexpressed in SeMet supplemented M9 minimal media using methionine auxotrophic *E. coli* B834 (DE3) cells (see sections 9.1.3 and 8.5.3) (Sreenath *et al.*, 2005) and the PanD(T57V).PanZ(SeMet).CoA complex was prepared similarly to the PanD(T57V).PanZ.CoA complex used previously. Unfortunately, the protein complex precipitated during crystallisation attempts using the established conditions.

Second crystallisation screen

A new 96 condition sparse matrix crystallisation screen (Crystal Screen I and II, Hampton Research) of the PanD(T57V).PanZ.AcCoA complex at a concentration of 8 mg mL^{-1} yielded small square plates (Fig. 3.12, **B**) in 100 mM sodium cacodylate

---

[*] The crystals used for these experiments were obtained during the second batch of optimised crystallisation, which coincided with the AcCoA ITC experiments. Therefore, the crystals were grown in excess AcCoA, rather than CoA as used in the previous batch.

pH 6.5, 1.4 M sodium acetate. Crystals were cryo-protected and cryo-cooled as before and initial data were collected at 100 K at beamline I24 (DLS), to a maximum resolution of 2.3 Å (see section 9.2). Indexing was ambiguous, with multiple possible solutions (*I4*, *C222*, *P4$_2$22* and *P422*) obtained from automatic data reduction with Xia2 (Winter, 2010). All the crystals were heavily twinned, with twin fractions calculated by Pointless (Evans, 2006) varying between 0.371 and 0.5. Space group ambiguity may be an artefact of twinning, where the twin axis can be misassigned as a crystal symmetry axis. Reducing the data in the correct space group is essential for correct structure solution. Data from the two highest resolution datasets, initially thought to be in space group *P4$_2$22* were integrated again in *P1* using XDS and scaled and merged in different space groups (*P4*, *C222*, *P222*, *P2$_1$*, *P1*) in Aimless (Evans and Murshudov, 2013). The best merging statistics (R$_{merge}$ and R$_{p.i.m.}$) were obtained in space group *P2$_1$* (a=b=84.8 Å, c=82.4 Å, α=β=γ=90°). Molecular replacement of the PanD protomer using Molrep (Vagin and Teplyakov, 1997) or Phaser (McCoy *et al.*, 2007) in *P2$_1$* and *P1* was successful, but once again the PanZ monomer could not be placed. The phased data were analysed with Zanuda (Lebedev and Isupov, 2014) which suggested a further two possible space groups (*P42$_1$2* and *P2$_1$2$_1$2*).
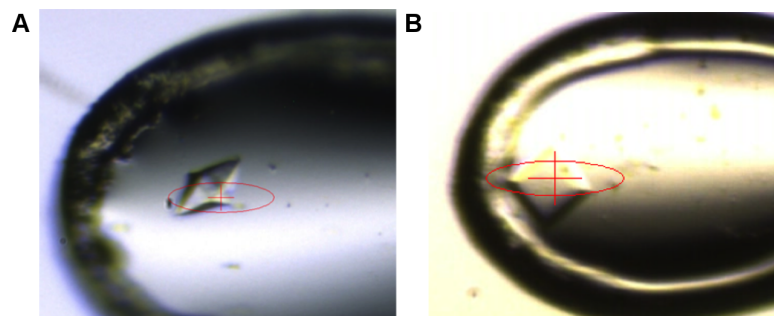
Fig. 3.12 Cryo-cooled PanD(T57V).PanZ.AcCoA crystals in nylon loops, aligned to the synchrotron beam. In both cases, the beam is 70 μm wide (red ellipse). (**A**) bipyramidal crystals obtained from 100 mM Tris pH 8.0, 200 mM Li$_2$SO$_4$ and 20% *w/v* PEG 4000. (**B**) square plates obtained from 100 mM sodium cacodylate, pH 6.5, 1.4 M sodium acetate.

All other datasets collected from the further optimised crystallisation conditions suffered from similar pathologies as described. Although several methods were attempted to obtain correct indexing, no unambiguous solution for the whole protein complex showing defined electron density for PanD(T57V) as well as PanZ was obtained. It was suggested that some of the crystal pathologies observed could be associated with the 4-fold rotational symmetry of the PanD(T57V).PanZ.AcCoA complex, as described in the next section.

Twinning and structure solution

Although twinning is a relatively common occurrence in protein crystallography and can cause difficulties during indexing and phasing, the fact that some of the datasets collected during the course of this project showed reasonable solutions for the ADC complex but no interpretable electron density for PanZ could also indicate crystal lattice perturbations. As the ADC tetramer has a 4-fold axis that could be associated with a crystallographic symmetry axis, if there were even small changes in the conformation of the PanZ molecules surrounding the tetramer, this could have resulted in the poorly-defined electron density observed for PanZ. A pictorial representation of possible lattice deformations is shown in Fig. 3.13.

ADC tetramer with flexible PanZ   Perfect crystal   Lattice deformations
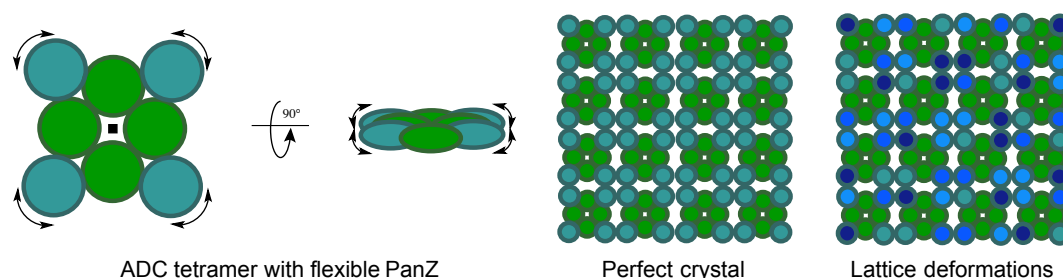
Fig. 3.13 Potential source for PanD(T57V).PanZ.AcCoA crystal lattice deformations. The PanD(T57V) tetramer is shown in green and different positions/conformations of PanZ in different shades of blue. The PanD(T57V) has a four-fold axis which may be associated with crystal symmetry. In a perfect crystal, PanZ would adopt the same conformation through the lattice. If PanZ adopts different conformations, lattice deformations may arise.

To overcome these problems, three approaches were suggested. Firstly, avoid cryo-cooling-induced conformational changes of the relative protein orientations (*i.e.* collect the data at room temperature, RT). If PanZ is mobile, collecting the data at RT could allow for a single orientation to be captured, whereas cryo-cooling may lock the protein in several different conformational landscape energy minima, "blurring" the electron density for PanZ. A second option was to attempt lattice tightening by dehydration. Decreasing the water content in the crystal may restrict the flexibility of the proteins, at the risk of destroying the lattice completely if the crystal is too dehydrated. Successful lattice tightening can be translated into an increase in diffraction resolution or changes in crystal symmetry (Bowler *et al.*, 2015). To achieve this, the crystals are mounted under a humidifying air stream and diffraction quality is monitored over a slow decrease in water content in the crystal. This is achieved by slowly decreasing the relative humidity of the air stream surrounding the crystal in 0.5 or 1% steps. The last approach is to screen for further crystallisation conditions that may yield an untwinned, well-diffracting crystal.

A further seven 96-well sparse matrix commercial crystallisation screens were set-up [Crystal Screen HT 1 and 2, Index, Salt RX (Hampton Research), Morpheus (Molecular Dimensions) and Wizard I and II (Emerald Biosciences)], using PanD(T57V).PanZ.AcCoA at a concentration of 10 mg mL^{-1} with 2 eqv. of AcCoA (per PanZ monomer) to help stabilise the complex. From these, more than 120 potential crystallisation hits were obtained, and well-formed crystals from 5 conditions were chosen for initial screening and data collection at ID29 at ESRF. The data were to be collected at RT, to avoid cryo-cooling, and dehumidification was also attempted using the HC1 instrument (Sanchez-Weatherby et al., 2009).

From all the crystals screened, one diffracted extremely well (crystallised in 200 mM potassium thiocyanate, 100 mM Bis-Tris propane pH 6.5, 20% w/v PEG 3350) and data were collected at RT, at 99.5% relative humidity, to 1.6 Å resolution (see section 9.2). Data were integrated in space group *I4* (a=b=86.4 Å, c=81.0 Å, α=β=γ=90°) with Xia2 (Winter, 2010) and scaled and merged using Aimless (Evans and Murshudov, 2013) with good statistics (Table 1). Molecular replacement using Molrep (Vagin and Teplyakov, 1997) successfully placed one PanD protomer (PDB 4AZD, with a contrast of 14.68) followed by one PanZ protein in the asymmetric unit cell (PDB 2K5T, with a contrast of 5.19). The structure was refined using Refmac5 (Murshudov et al., 2011) and manually rebuilt in Coot (Emsley et al., 2010).

The new crystallisation conditions were then optimised to give square pyramidal crystals up to 300 μm in size, by varying the protein to mother liquor ratios. 10 μL hanging drops containing 4:1 protein:mother liquor ratios gave the largest crystals (Fig. 3.14). A second dataset from these crystals was collected in-house at RT, to 1.7 Å resolution (see section 9.2). The data were integrated using HKL3000 (Otwinowski and Minor, 1997), scaled and merged using Aimless (Evans and Murshudov, 2013) and phased by molecular replacement using Molrep (Vagin and Teplyakov, 1997) using the structure obtained from the synchrotron dataset as a search model (Table 1). The in-house structure was refined using Refmac5 (Murshudov et al., 2011) and manually rebuilt in Coot (Emsley et al., 2010) and found to be isostructural to the synchrotron structure.

	PanD(T57V).PanZ.AcCoA	PanD(T57V).PanZ.AcCoA
Data collection	synchrotron	in-house
Space group	*I4*	*I4*
Cell dimensions		
a, b, c (Å)	86.4, 86.4, 81.0	86.3, 86.3, 80.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	59.1-1.61 (1.64-1.61)	33.7-1.70 (1.75-1.70)
R_{merge} (%)	5.8 (49.8)	3.7 (39.8)
$R_{p.i.m.}$ (%)	4.4 (41.2)	3.6 (36.6)
$<I> / sI$	11.8 (1.6)	13.3 (2.2)
Completeness (%)	97.4 (80.1)	99.7 (99.7)
Redundancy	4.1 (2.6)	2.4 (2.3)
Absorbed dose (kGy)[*]	70	20
Refinement		
Resolution (Å)	59.1-1.61 (1.65-1.61)	29.52-1.70 (1.75-1.70)
No. reflections	35579 (2180)	30609 (2242)
R_{work} (%)	12.9 (24.8)	14.4 (32.7)
R_{free} (%)	15.0 (26.2)	17.5 (36.2)
No. atoms		
Protein	2090	2064
Ligand/ion	53	65
Water	126	145
B-factors		
Protein (main chain) (Å^2)	22.27	20.35
Protein (side chain) (Å^2)	28.91	25.99
Ligand/ion (Å^2)	28.19	20.35
Water (Å^2)	35.55	32.45
R.m.s. deviations		
Bond lengths (Å)	0.029	0.023
Bond angles (°)	3.290	2.33
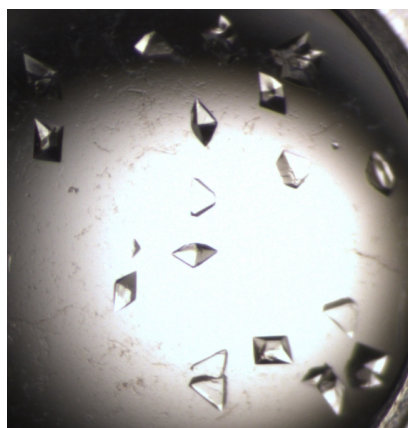
Fig. 3.14 Optimised crystallisation conditions for the PanD(T57V).PanZ.AcCoA complex gave large (300 μm in the longest dimension), clean, square-pyramidal crystals.

---

[*] Calculated using RADDOSE-3D (Zeldin *et al.*, 2013)

### 3.2.5 Structural analysis of the in-house PanD(T57V).PanZ.AcCoA complex

Overall architecture of the protein-protein complex

As expected from the ITC binding studies, the PanD(T57V).PanZ.AcCoA complex is a heterooctamer. The complex is cross-shaped, with four PanZ protomers bound adjacent to the ADC loops that are cleaved post-translationally to form the active site. The AcCoA binding site is also adjacent to the protein-protein interface, consistent with the previously observed CoA-dependent complex formation.

A 1:1 mixture of PanD(T57V) and PanZ was concentrated to ~10 mg mL^{-1}, 2 eqv. of AcCoA were added and SAXS data collected at varying dilutions on beamline 4-2 of the Stanford Synchrotron Radiation Light Source (SSRL) by Dr Thomas Grant, Dr Edward Snell and Prof. Arwen Pearson to confirm that the heterooctameric structure of the PanD.PanZ.AcCoA complex obtained from X-ray diffraction corresponded to the true architecture of the complex in solution. The data were analysed with the help of Dr Thomas Grant (Monteiro *et al.*, 2015). The experimental solution scattering data matched the crystallographic heterooctamer structure both at high and low scattering angles (Fig. 3.16, **A**). Fig. 3.16 shows the plot of the change in calculated scattering of the complex (I) with scattering angle (q) compared to the experimental data. Small deviations from the experimental data could be accounted for by including a small amount (32%) of a heterooctamer dimer, consistent with the packing in the crystal (Fig. 3.16, **B**) along with the eight unresolved hexaHis-tags (one for each protomer) with each amino acid modelled as a sphere (Fig. 3.16, **C**). The modelled purification tags and calculated SAXS envelope are shown in Fig. 3.17.
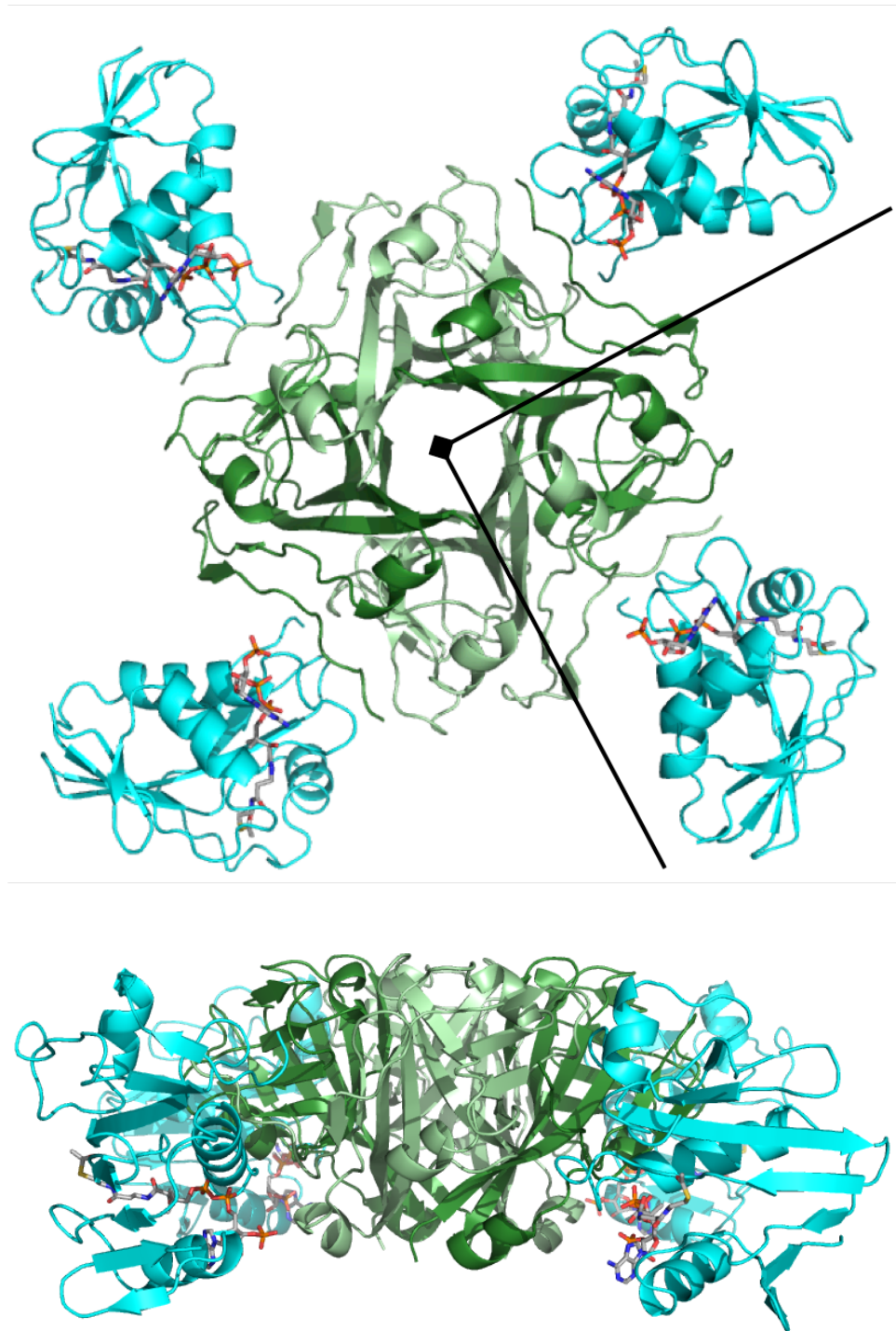
Fig. 3.15 Architecture of the PanD(T57V).PanZ.AcCoA complex. PanD(T57V) protomers are shown in light and dark green, PanZ in blue and AcCoA in sticks coloured by atom colours (carbons in grey). PanZ binds next to the PanD(T57V) protomer interface, at the PanD processing (and active) site. AcCoA is found adjacent to the protein-protein interface. The a and b axes of the unit cell are represented as black lines along with the 4-fold rotation axis as a black square.
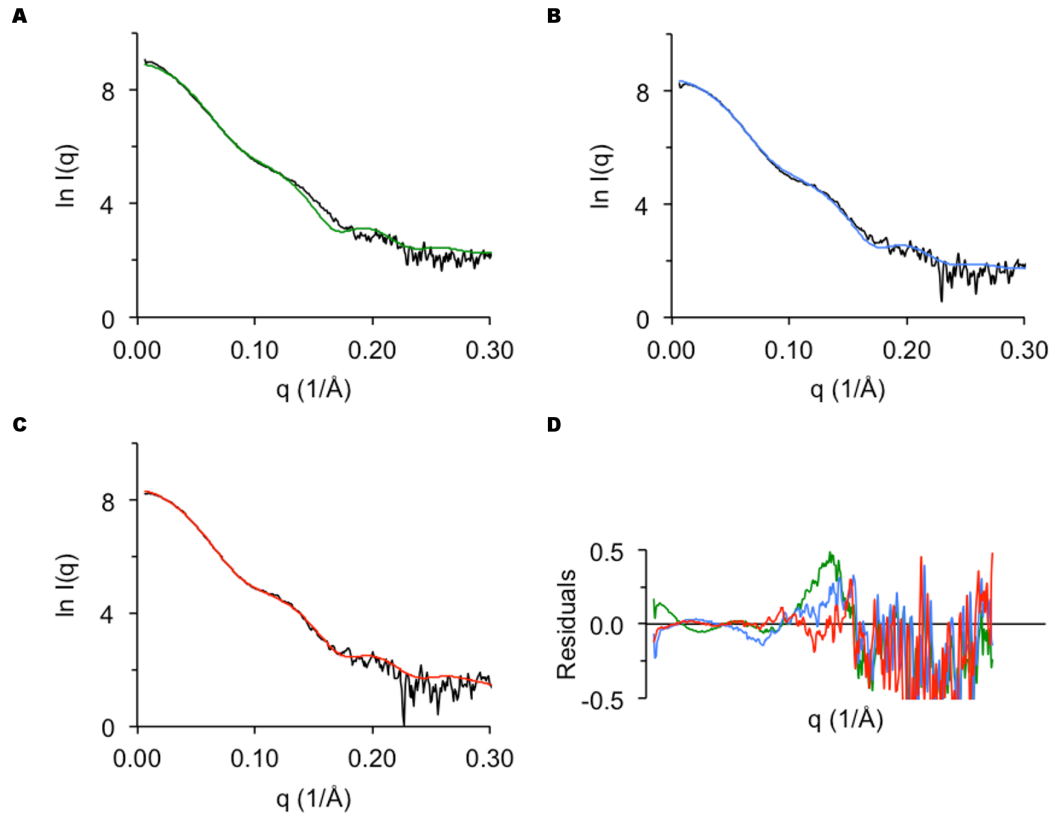
Fig. 3.16 Plot of ln q vs q (Å^{-1}) of experimental and calculated SAXS data for the PanD(T57V).PanZ.AcCoA complex. The experimental data are shown in black. (**A**), (**B**) and (**C**) show the calculated scattering of the crystallographic heterooctamer (green) with 32% of octamer dimers (blue) and also the modelled 8 crystallographically-unresolved hexaHis-tags (red) fitted against the experimental data. (**D**) shows the residuals of the fits.
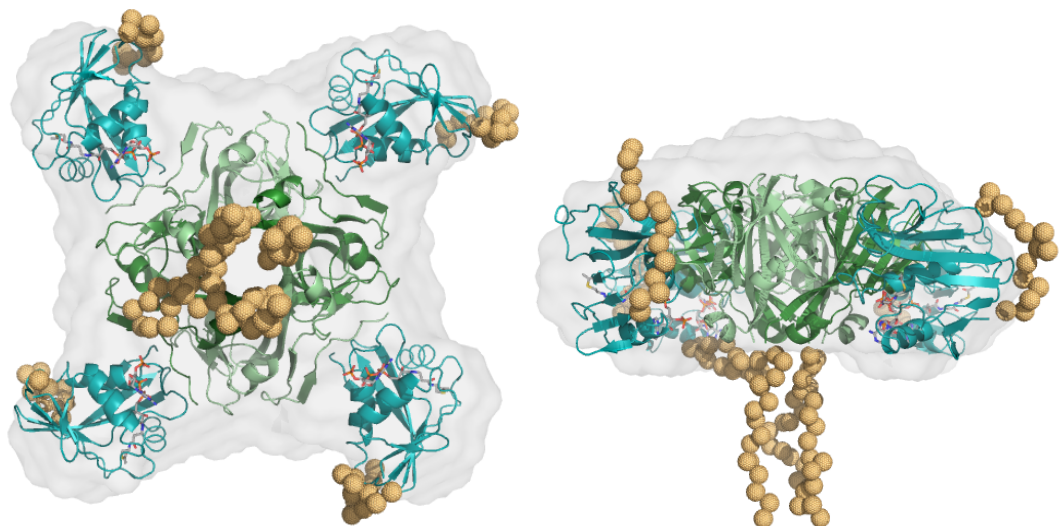
Fig. 3.17 Calculated SAXS envelope overlaid with the PanD(T57V).PanZ.AcCoA complex structure obtained by X-ray crystallography. PanD(T57V) is shown in green, PanZ in blue, CoA as sticks (coloured by atom and with carbons in grey) and the modelled crystallographically unresolved hexaHis tags as orange spheres. The purification tags were modelled as single spheres for each residue of the peptide chain.

Acetyl coenzyme A binding places PanZ in a PanD(T57V)-binding competent conformation

The solution structure of PanZ was previously elucidated by NMR spectroscopy in the presence of excess CoA by Cort *et al.* (Northeast Structural Genomics Consortium, PDB 2K5T, Fig. 3.18). The structure of PanZ in the PanD(T57V).PanZ.AcCoA complex was found to be very similar to that of CoA-bound PanZ in in solution. The Pan [26]Tyr-Val[33] loop is the most dynamic region in the solution data and it also corresponds to a region with high B-factors in the X-ray structure. As the AcCoA binding site is directly involved in the protein-protein interface and AcCoA is required for binding (as shown by ITC, ESMS and NMR in sections 3.2.2 and 3.2.3), it is likely that binding of CoA (or its thioesters) places PanZ in a PanD-binding competent conformation. The acetyl group of AcCoA points away from the protein-protein interface, supporting the findings of Stuecker *et al.* that PanZ does not act as an acetyl transferase during PanD activation (Stuecker *et al.*, 2012a).
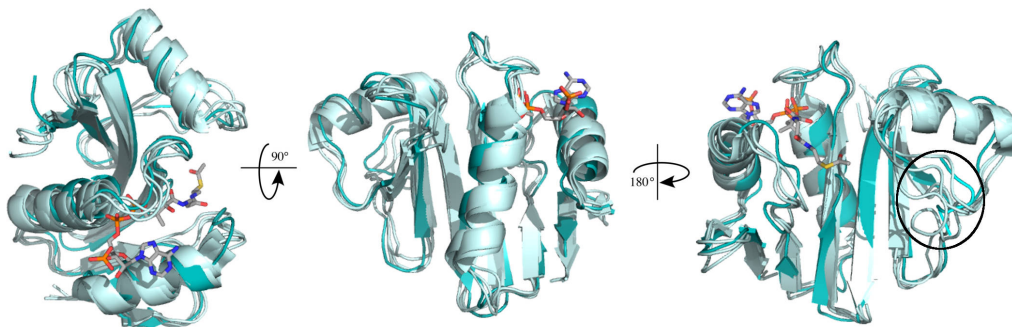
Fig. 3.18 Overlay of the structure of PanZ in complex with PanD(T57V) and AcCoA (dark blue) with the previously reported solution structure of PanZ in the presence of CoA (light blue, PDB 2K5T, 3 lowest energy conformers shown). The [26]Tyr-Val[33] loop (black circle, right) is poorly resolved in the X-ray electron density and also adopts multiple conformations in the NMR structure ensemble.

Binding of AcCoA to PanZ appears to stabilise the PanZ [66]Leu-Gly[76] loop (Fig. 3.19) through several hydrogen bonds (Leu66, Val68, Gly76, Gln79), electrostatic interactions (Arg74 to the pyridoxal phosphate), metal chelation (Mg^{2+} to Thr72 and the pyrophosphate) and cation-π interactions (Arg74 to the adenosine ring). This loop then forms part of the interface with PanD, crossing above the ADC C-terminal tail.
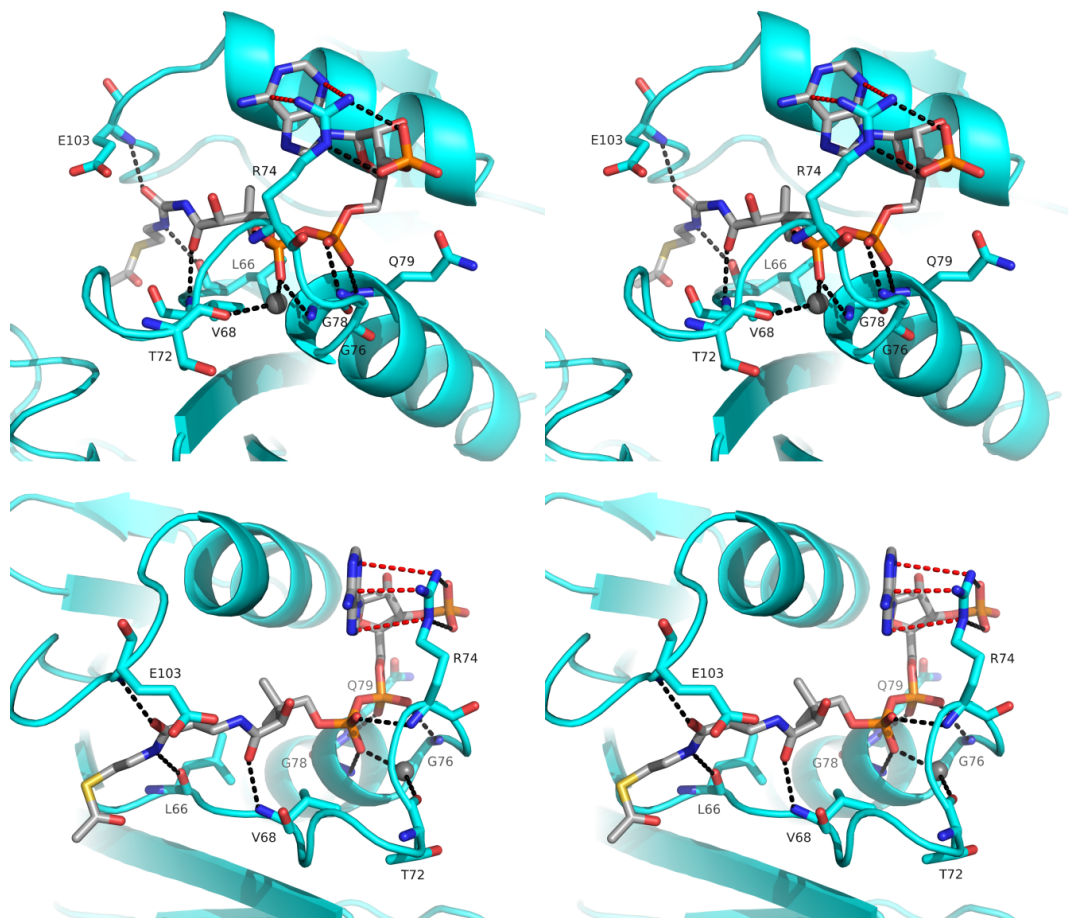
Fig. 3.19 Stereo images of the AcCoA binding site in PanZ, showing the interactions between the residues, the CoA moiety and Mg^{2+}. Mg^{2+} is shown as a grey sphere and AcCoA in sticks coloured by atom (carbons in grey). PanZ is shown in blue, with residues interacting with AcCoA as sticks, coloured by atom (carbons in blue). Hydrogen bonds and electrostatic interactions are shown as black lines and cation-π interactions as red lines.

The PanD(T57V) C-terminal tail is an integral part of the protein-protein interaction

The PanZ ^{66}Leu-Gly^{76} loop, ^{43}Arg-Leu^{48} loop and N-terminus sandwich the PanD(T57V) C-terminal tail (Fig. 3.20), bridging the protein-protein interface, through electrostatic interactions (PanZ-Arg75 to PanD(T57V)-Glu40), direct hydrogen bonds (PanZ-Thr72 to PanD(T57V)-Val123, PanZ-Glu46 to PanD(T57V)-Tyr22 and PanZ-Asn45 to PanD(T57V)-Gly23), van der Waals interactions (PanZ-Phe44 to PanD(T57V)-Pro122) and several water mediated hydrogen-bonding interactions. The PanZ-Asn45 to PanD(T57V)-Gly23 hydrogen bond is of considerable importance and will be discussed in the section dedicated to the mechanism of PanD post-translational modification (section 3.2.7).
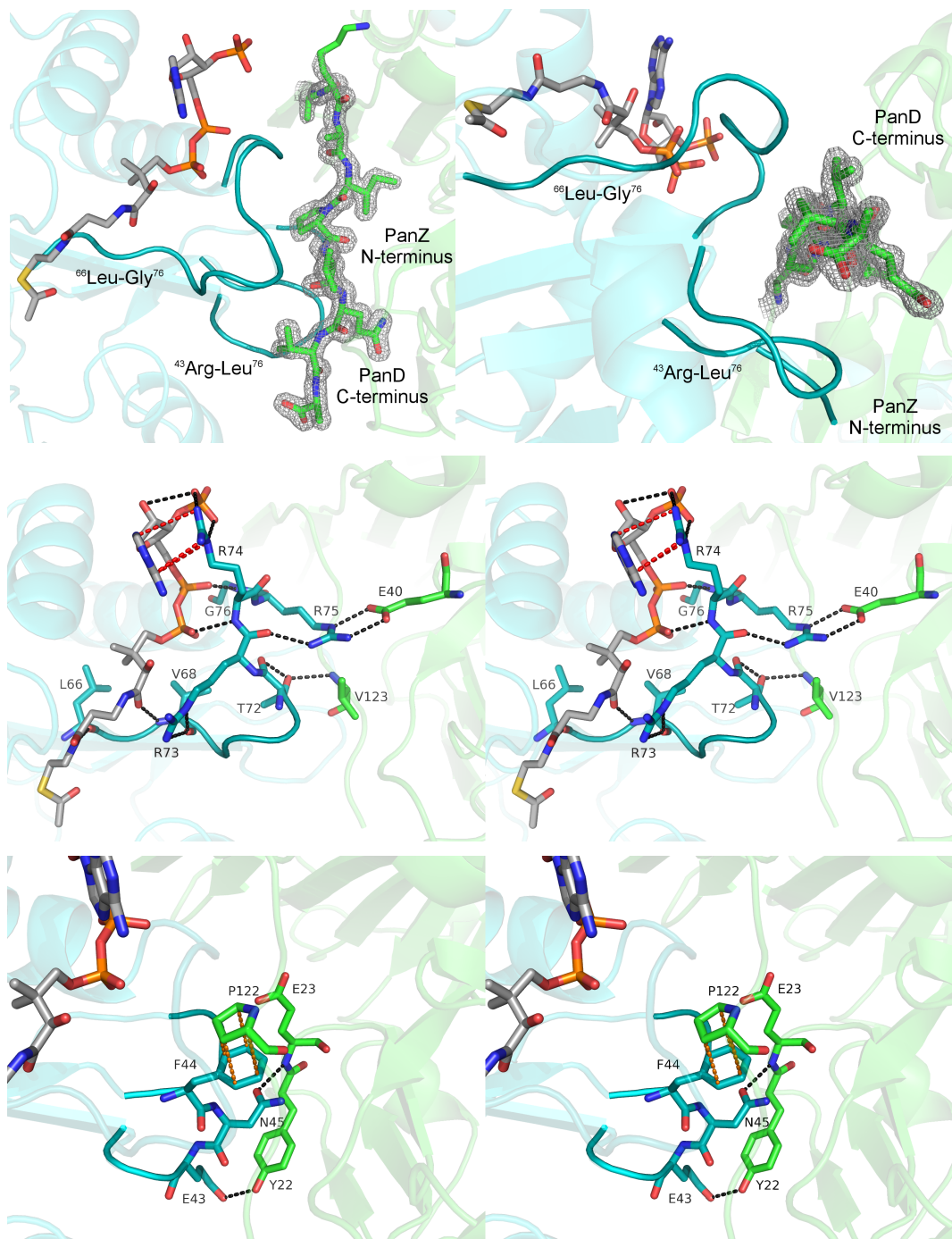
Fig. 3.20 Stereo-views of the PanD(T57V).PanZ.AcCoA interaction. PanD(T57V) is shown in green, PanZ in blue and AcCoA as sticks coloured by atom (carbons grey). Interacting residues from PanD(T57V) and PanZ are shown as sticks coloured by atom (carbons green and blue respectively). Hydrogen bonds as black dotted lines, cation-π interactions as red dotted lines and van der Waals interactions as orange dotted lines. (**Top**) the PanZ [66]Leu-Gly[76] and [43]Arg-Leu[48] loops and N-terminus sandwich the PanD C-terminal tail above and below. (**Middle**) The main interactions of AcCoA with the PanZ [66]Leu-Gly[76] loop, organising it to bridge the protein-protein interface. (**Bottom**) The interactions between the PanZ [43]Arg-Leu[48] loop and N-terminus and PanD.

On binding of PanZ, the PanD(T57V) C-terminal tail (^{120}Ala-Ala^{126}) becomes an integral part of the protein-protein interface. This was surprising as the hydrophobic C-terminal chain of ADC is intrinsically disordered and is unstructured in the majority of published ADC crystal structures, with the exception of that of the ADC(N72A) site-directed mutant, where it becomes ordered due to a crystal packing interaction (Webb *et al.*, 2012). Therefore, the importance of the C-terminus of ADC for binding of PanZ was investigated by ITC. A truncated mutant, PanD(T57V/K119STOP) was generated by site-directed mutagenesis, overexpressed in auto-induction media and purified as previously described for the PanD(T57V) mutant (see sections 9.1 and 8 for further details).

Titration of PanD(T57V/K119STOP) against PanZ.AcCoA (Fig. 3.21, **B**) showed no measurable binding compared to the titration of PanD(T57V) against PanZ.AcCoA (Fig. 3.21, **A**), confirming that the PanD C-terminal tail is essential for the protein-protein interaction. Interestingly, the alignment of ADC protein sequences across different bacteria shows that the C-terminal peptide is fully conserved across all PanZ producing bacteria (Fig. 3.22), further supporting its role as an essential component in PanD.PanZ protein complex formation.
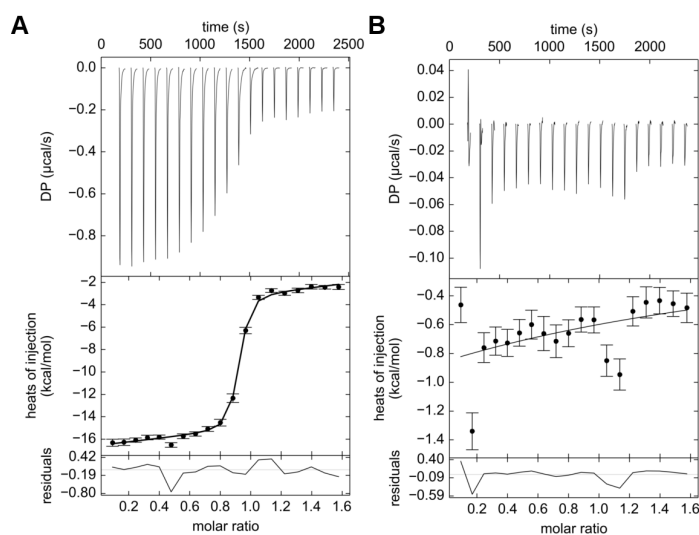
**Fig. 3.21** ITC titrations of (**A**) 263 μM PanD(T57V) into 35 μM PanZ.AcCoA in the presence of excess AcCoA (400 μM) and (**B**) 35 μM PanD(T57V/K119STOP) intro 263 μM PanZ.AcCoA in the presence of excess AcCoA (394 μM). All samples were in 50 mM Tris (pH 7.5), 100 mM NaCl and 0.1 mM DTT (*SEC buffer A*). Truncation of the PanD C-terminal tail shows no measurable binding to PanZ compared to the full-length protein.

```
P. aeruginosa    110 NQLSHTSEAIPIQVA                   126
E. coli          110 NEMKRTAKAIPVQVA                   126
E. carotovora    110 NELQRQAKAIPVQVA                   126
Y. pestis        110 NQLQRKAKAVPVQVA                   126
C. glutamicum    110 NRIVALGNDLAEALPGSGLLTS-RSI        136
S. coelicolor    110 NRIVGLGADASEPVPGSDQERSPQAVSA      139
M. tuberculosis  110 NKPIDMGHDPAFVPENAGELLDPRLGVG      139
T. thermophilus  110 NRILEVRKG                         120
L. pneumophila   111 NRLKEIRPERIGVKSRIPYPA             133
C. crescentus    110 NLIKKAA                           118
B. subtilis      110 NQIIEMLGAEKAGTIL                  127
S. aureus        110 NVIIEMIHEKENTIVL                  127
H. pylori        110 NEILEK                            116
```

Fig. 3.22 PanD sequence alignment. The conserved C-terminal region across all PanZ producing bacteria (*P. aeruginosa*, *E. coli*, *E. carotovora* and *Y. pestis*) is highlighted in yellow.

3.2.6 The ADC.PanZ complex is involved in CoA biosynthesis regulation

The data obtained to this point in the project suggest a positive feedback role of PanZ in regulating CoA biosynthesis, as the PanD.PanZ interaction is CoA dependent. At low CoA concentrations, PanZ exists as *apo*-PanZ, which is not able to bind to PanD. PanD activation to ADC, does not occur, halting β-alanine and, consequentially, CoA production. At high CoA concentrations, PanZ exists mainly as the PanZ.CoA species, which is able to bind to and promote PanD activation to ADC. Therefore, at high CoA concentrations ADC should be mainly activated and able to catalyse the conversion of L-aspartate to β-alanine, which, in turn, would lead to an uncontrolled increase in the cellular pools of pantothenate and CoA (Fig. 3.23).
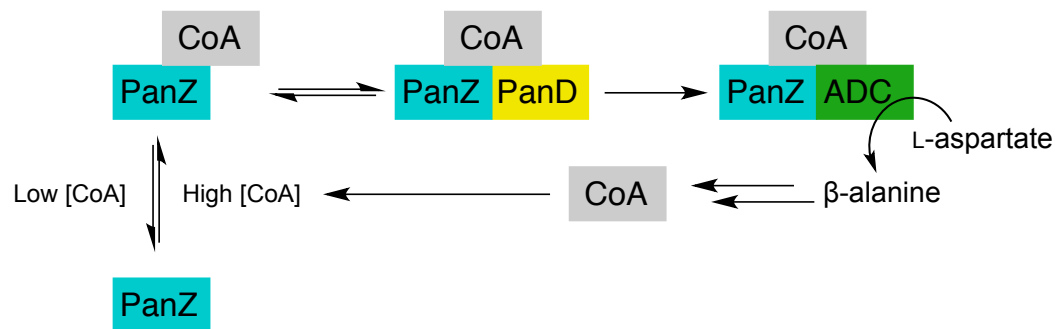
Fig. 3.23 PanZ-mediated, positive feedback loop of CoA biosynthesis suggested by ITC, NMR and X-ray crystallography data. At high CoA concentrations, PanZ mainly exists as PanZ.CoA which is able to bind to the PanD zymogen, activating it to ADC. ADC then catalyses the conversion of L-aspartate to β-alanine, a precursor to CoA.

As CoA biosynthesis is negatively feedback regulated downstream of aspartate by the interaction of CoA with pantothenate kinase, which phosphorylates pantothenate, (Rock *et al.*, 2003), β-alanine would be continuously produced and excess pantothenate excreted out of the cell (Jackowski and Rock, 1981), while still keeping the concentration of CoA roughly constant. This mechanism for the production of CoA is both unexpected and unreasonable in a biological context, as it would lead to the exhaustion of the cellular pools of aspartate. Therefore, further investigations were

52

carried out in order to determine whether the PanD.PanZ interaction could act to negatively regulate CoA biosynthesis, potentially via a mechanism unrelated to the PanZ-mediated PanD activation.

As PanZ is a CoA sensor (it binds CoA) it was hypothesised that it could perform a secondary function, beyond the acceleration of PanD activation, in regulating CoA biosynthesis. As the structures of PanD(T57V) (PDB 4AZD) (Webb *et al.*, 2014) and ADC (PDB 1PPY) (Schmitzberger *et al.*, 2003) are similar, PanZ.CoA could possibly also bind to the activated enzyme and, since the protein-protein interaction surface is adjacent to the active site, potentially act as an inhibitor of ADC. Under these circumstances, then PanZ would perform two distinct functions: activation and inhibition of ADC. As the PanD.PanZ interaction is CoA dependent, then this secondary function would lead to a negative feedback loop mechanism of CoA biosynthesis. To explore this hypothesis, the binding affinity of PanZ.CoA to the activated enzyme, WT ADC, was investigated by ITC, the structure of the WT ADC.PanZ.AcCoA complex solved by X-ray crystallography and the effect of PanZ.AcCoA upon ADC catalysis assessed both *in vitro* and *in vivo*.

Protein purification

WT PanD was overexpressed in auto-induction media using the same protocol as for PanD(T57V) and previously transformed *E. coli* C41 (DE3) cells harbouring a pRSETa plasmid encoding the protein. Using WT *E. coli* cells allows for the overexpressed PanD protein to be efficiently activated to ADC by heating the collected cell pellet at 37 °C overnight prior to storage at -80 °C (see section 9.1.6). This is hypothesised to be due to the presence of endogenous PanZ, which can activate the zymogen. Therefore, WT PanD was obtained from non-heated cell pellets and WT ADC from heated cell pellets. The purification was carried out as previously described for PanD(T57V) and the extent of cleaved WT ADC analysed by tris-tricine SDS-PAGE (see sections 9.1.5, and 8).

WT PanD.PanZ.AcCoA and WT ADC.PanZ.AcCoA crystal structures

The non-activated protein was crystallised with PanZ.AcCoA (WT PanD.PanZ.AcCoA complex) using the same crystallisation conditions as for the PanD(T57V).PanZ.AcCoA complex. The crystals were cryo-protected by soaking for a few seconds in mother liquor containing increasing concentrations of glycerol (5%, 10% and 20% *w/v*) and cryo-cooled in liquid nitrogen. Data were collected at beamline I04-1 (DLS), to 1.2 Å

resolution. Data were integrated with Xia2 (Winter, 2010), scaled and merged using Aimless (Evans and Murshudov, 2013) and did not appear to be twinned, suggesting that the previous twinning problems were either not caused by cryo-cooling but rather are related to the crystal form, or that the affinity of PanZ for the WT protein is tighter, limiting the conformational flexibility of the complex. Data were phased by molecular replacement using Molrep (Vagin and Teplyakov, 1997) with the PanD(T57V).PanZ.AcCoA structure as a model. The structure was refined using Refmac5 (Murshudov *et al.*, 2011) and rebuilt manually using Coot (Emsley *et al.*, 2010) (Table 2).

The WT PanD.PanZ.AcCoA complex was very similar to the PanD(T57V).PanZ.AcCoA complex. PanZ binds ADC across the active site (Fig. 3.24), supporting, from a structural point of view, the hypothesised inhibitory effect of PanZ on ADC catalysis. As expected, WT PanD is seen to be completely activated, confirming that PanZ.AcCoA is able to efficiently promote the maturation of the zymogen to the active enzyme, though, surprisingly, residue 25 shows a tetrahedral moiety, rather than the expected planar geometry of the catalytic pyruvoyl group. The density was best modelled as a hydrated pyruvoyl (Fig. 3.24), which could arise from changes in the electrostatic environment around the active site upon binding of PanZ or be due to the crystallisation buffer composition.

To determine whether the tetrahedral centre formed upon the interaction of the two proteins or if it was in fact correlated with the activation process, thermally-activated WT ADC was also complexed with PanZ in the presence of 2 eqv. of AcCoA (per PanZ monomer) and crystals grown under the same conditions as for the WT PanD.PanZ.CoA complex. The crystals were handled, data collected and reduced and the structure solved as described for the WT PanD.PanZ.AcCoA complex. Once again, the active site showed clean cleavage of the peptide backbone at residue Gly24, with the presence of the tetrahedral moiety at residue 25, showing that this is associated simply with the formation of the protein complex, rather than with the activation process. This structure also confirms that thermal activation of PanD to ADC can be done *in cellulo*, without the formation of serine hydrolysis by-products as described in section 3.1.2 and visible in the first ADC crystal structure (PDB 1AW8) (Albert *et al.*, 1998). In practical terms, the fastest and easiest protocol for the production of fully active ADC from PanD overexpression is by expressing the protein in WT *E. coli* cells and then heating the cell pellets at 37 °C overnight, rather than by incubation of PanD and PanZ.AcCoA.

Lastly, it was important to determine whether the hydrated pyruvoyl appears only at cryo-temperatures or if it was a genuine, stable feature of the complex, even at RT. A second dataset for the WT ADC.PanZ.AcCoA complex from a crystal from the same crystallisation batch was collected at RT on the in-house X-ray source. Data were integrated with XDS (Kabsch, 2010) and scaled, merged, phased and refined as described for the cryo-temperature synchrotron dataset (Table 2). The electron density at the pyruvoyl centre was once again clearly tetrahedral, confirming that it was the electrostatic environment surrounding the ADC active site in the crystal that stabilises the hydrated form of the pyruvoyl preferentially to the α-ketoamide.
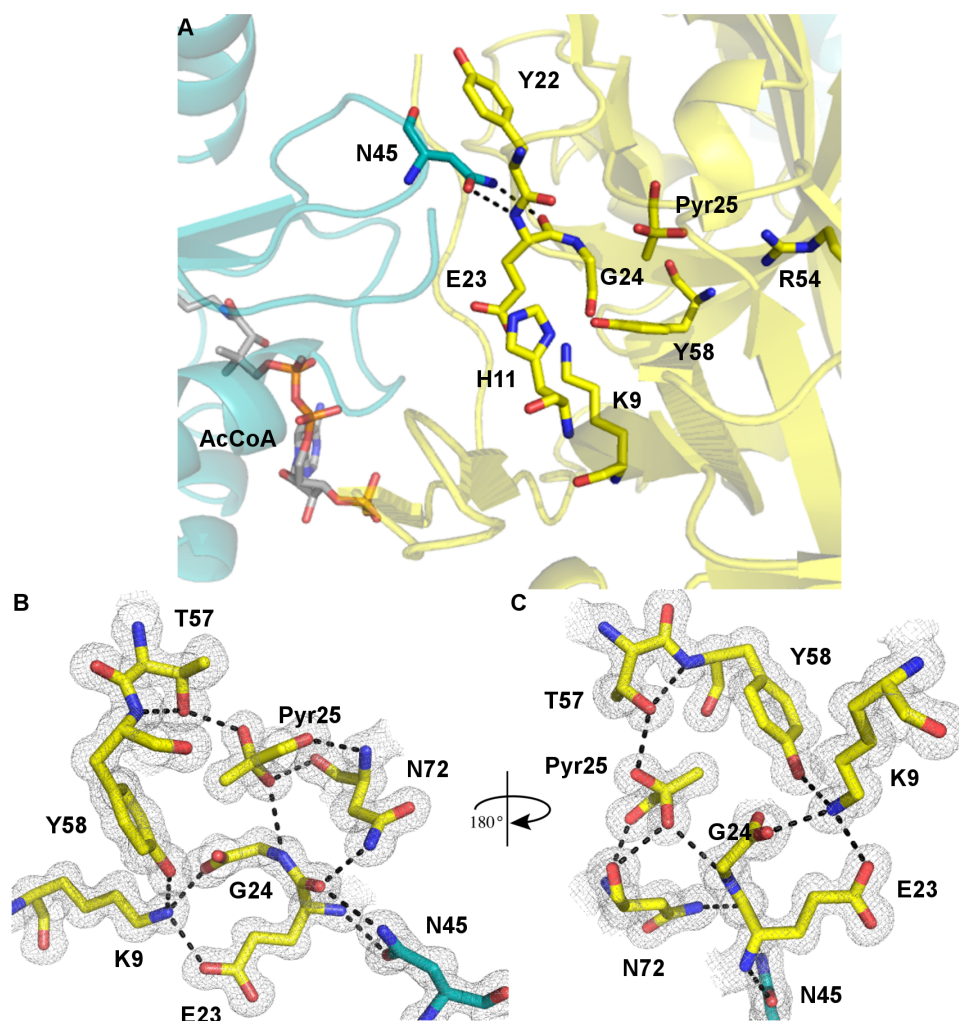
Fig. 3.24 The WT ADC.PanZ.AcCoA complex, showing the interaction of the two proteins at the active site. WT ADC is shown in yellow, PanZ in blue and CoA as grey sticks. 2Fo-Fc map contoured at 1 σ. (**A**) PanZ.AcCoA binds adjacent to the ADC active site. The main residues in the enzyme active site are shown as sticks (Lys9, His11, Tyr22, Glu23, Glu24, Pyr25, Arg54, Tyr58) and the hydrogen bond between ADC-Glu23 and PanZ-Asn45 is shown as dashes. (**B**) and (**C**) the ADC active site showing the pyruvoyl in a hydrated form. The electron density shows a clear tetrahedral moiety. The hydrated pyruvoyl (Pyr25) is stabilised by a network of hydrogen bonds to adjacent residues. One hydroxyl group is a H-bond acceptor from Thr57 and the second hydroxyl group a H-bond donor to Asn72-CO backbone and a H-bond acceptor from Gly24-NH backbone.

Table 2 WT PanD.PanZ.AcCoA X-ray diffraction data reduction and refinement statistics

	WT PanD. PanZ.AcCoA	WT ADC. PanZ.AcCoA	WT ADC. PanZ.AcCoA
Data collection	synchrotron	synchrotron	in-house
Temperature	100 K	100 K	RT
Space group	*I4*	*I4*	*I4*
Cell dimensions			
a, b, c (Å)	85.9, 85.9, 80.1	86.6, 86.6, 79.9	86.4, 86.4, 81.1
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	29.28-1.16	42.78-1.25	43.19-1.85
	(1.18-1.16)	(1.27-1.25)	(1.89-1.85)
R_{merge} (%)	7.7 (56.7)	9.2 (76.3)	15.4 (77.1)
$R_{p.i.m.}$ (%)	5.4 (39.5)	5.4 (47.4)	12.3 (55.5)
$<I>$ / sI	9.0 (2.1)	7.4 (7.0)	5.8 (1.6)
Completeness (%)	99.6 (93.7)	99.8 (99.9)	100.0 (100.0)
Redundancy	5.1 (4.4)	3.6 (3.4)	4.4 (4.4)
Refinement			
Resolution (Å)	60.72 (1.16)	42.78 (1.25)	43.18 (1.85)
No. reflections	99822 (6688)	75141 (5473)	24079 (1752)
R_{work} (%)	10.8 (19.1)	11.4 (20.9)	14.91 (28.9)
R_{free} (%)	11.5 (20.2)	14.4 (23.9)	18.89 (25.1)
No. atoms			
Protein	2126	2118	2129
Ligand/ion	112	92	58
Water	290	269	124
B-factors			
Protein (main chain) (Å^2)	15.11	14.75	26.20
Protein (side chain) (Å^2)	21.50	19.48	31.17
Ligand/ion (Å^2)	21.74	22.63	31.33
Water (Å^2)	33.14	32.10	35.42
R.m.s. deviations			
Bond lengths (Å)	0.034	0.034	0.019
Bond angles (°)	2.86	2.84	2.15

PanZ inhibits ADC catalysis

The active site of ADC is composed of residues from two adjacent subunits. The enzyme contains a number of fully conserved residues, the vast majority of which are located close to the active site (Webb *et al.*, 2004). The catalytic mechanism of ADC starts with the condensation of L-aspartate with the pyruvoyl group, forming an iminium intermediate (Fig. 3.25, **II**). The β-carboxylate forms a salt-bridge to Arg54 and the α-carboxylate makes an ionic interaction with Lys9, which, in turn, is hydrogen-bonded to His11 and Tyr58. The C-terminus of the β-chain of ADC (^{24}Gly-Tyr^{22}) is mobile and can adopt different conformations, visible in the structure of the enzyme (PDB 1AW8)

(Albert *et al.*, 1998) and also in the crystal structures of ADC with different ligands bound to the active site (Yorke, Monteiro *et al.*, unpublished results)[*]. Once bound, L-aspartate undergoes decarboxylation at the α-carboxylate (Williamson and Brown, 1979) to an enolate (**III**), which is then protonated by Tyr58 (Saldanha *et al.*, 2001) to the β-alanine iminium intermediate (**V**). The product is then released by hydrolysis, regenerating the pyruvoyl group (**VI**).
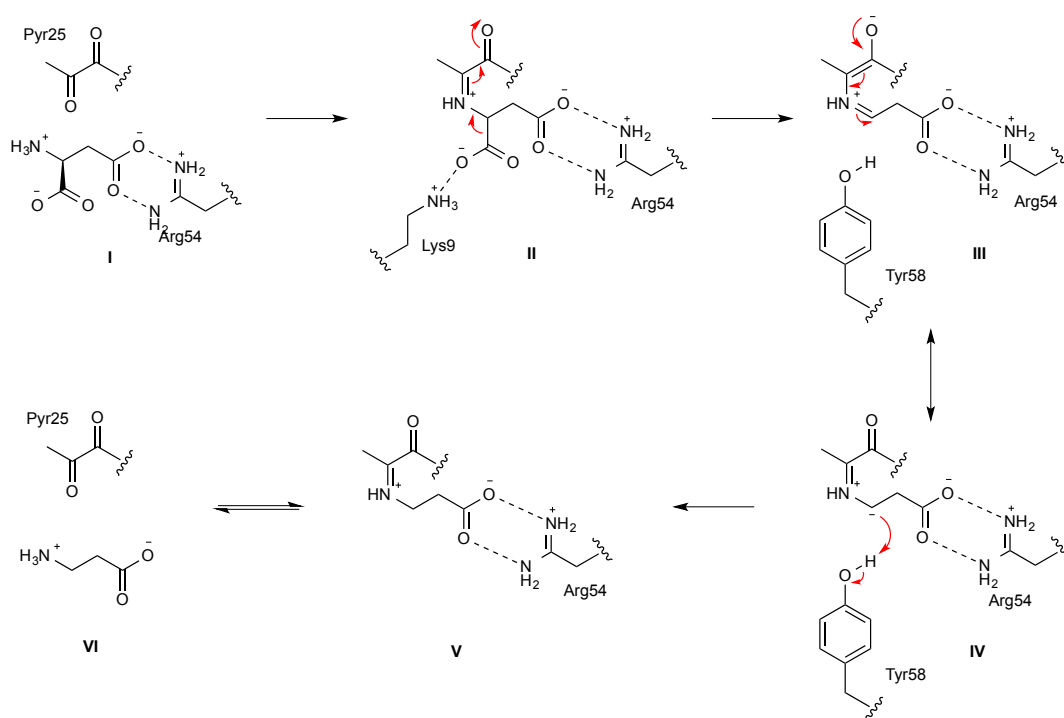
Fig. 3.25 Mechanism of L-asparate decarboxylation by ADC (Webb *et al.*, 2004). L-Aspartate binds at the ADC active site, forming a salt bridge to Arg54 (**I**), promoting the formation of the iminium intermediate (**II**). The α-carboxylate is cleaved to give the enolate intermediate (**III**), which is protonated by Tyr58 (**IV**) (Saldanha *et al.*, 2001) to the β-alanine iminium intermediate (**V**). The product is then released by hydrolysis (**VI**).

In vitro assays

The assay for ADC catalysis is difficult due to the lack of a suitable chromophore in the enzyme, substrate or product. One way to assay the turnover rate is to monitor the heat of reaction. Heat evolution can monitored directly by ITC (Todd and Gomez, 2001) and is directly proportional to enthalpy. Enthalpic changes during catalysis are mainly dominated by the cleavage of the substrate to release CO_2 (Fig. 3.25, **II**) and also from the protonation of the enolate intermediate (Fig. 3.25, **IV**). This method has been previously shown to be a good method to assay a range of enzymatic reactions,

---

[*] Crystal structures of ADC with different ligands bound at the active site show a large degree of flexibility for the C-terminus of the β-chain. These structures will not be discussed in this thesis but are currently being prepared for publication.

including those of DHFR, trypsin, creatine phosphokinase and urease (Todd and Gomez, 2001).

The activity of ADC in the presence of different PanZ concentrations was investigated by ITC using a multiple injection approach. The titrations and data fitting were performed by Dr M. E. Webb as part of this project. Twenty-five aliquots of L-aspartate were injected into ADC, allowing just enough time for the system to equilibrate between the injections. The titration was repeated in the presence of 0.5 and 2 equivalents of PanZ.AcCoA. Fig. 3.26 shows the data fitted from the titrations. Upon addition of 0.5 eqv. of PanZ.AcCoA ADC activity is reduced to 50% and with 2 eqv. of PanZ.AcCoA only 10% activity is recovered. This assay clearly shows that, not only does PanZ.AcCoA bind to activated ADC, but complex formation results in ADC inhibition.
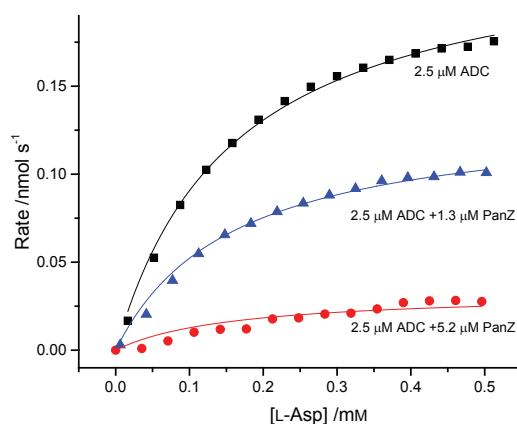

Fig. 3.26 Relative rates of turnover of WT ADC (2.5 μM, protomer concentration) at varying L-aspartate concentrations in the presence of 0, 0.5 and 1.0 eqv. of PanZ.AcCoA (per ADC protomer). In the presence of 0.5 eqv. of PanZ.AcCoA (1.3 μM), the relative rate of catalysis is approximately halved, whereas in the presence of 2 eqv. of PanZ.AcCoA, only 10% of activity is recovered.

The concentrations of ADC and PanZ.AcCoA used in this ITC assay were considerably above the binding affinity between the proteins. To investigate PanZ inhibition of ADC further, a second assay was set up where ADC turnover over the course of several hours was monitored by ^{1}H NMR spectroscopy. As the enzyme has a slow turnover rate and the substrate and product are simple organic molecules, a ^{1}H NMR spectrum of the reaction mixture can be recorded and the percentage turnover extrapolated from the ratio between the aspartate Cα/Cβ and β-alanine CH_2 signals.

WT ADC (100 nM) was mixed with varying concentrations of PanZ (200 nM, 1 μM and 10 μM) with 2 equivalents of AcCoA (400 nM, 2 μM and 20 μM respectively).

0.5 mM L-aspartate (~3×K_M) was added to the protein mixtures and the samples incubated at 37 °C. Aliquots were taken at 2, 4 and 24 h intervals, mixed with D_2O and ^{1}H NMR spectra collected immediately with a water suppression pulse sequence. Three experiments were run with excess AcCoA added (100 μM) and three without. Two sets of spectra for two ADC turnover time-courses in the presence of low PanZ concentrations are shown in Fig 3.27. The spectra show a decrease in L-aspartate concentration and an increase in β-alanine concentration, meaning that ADC is able to catalyse the reaction and that the percentage turnover can be estimated from the NMR integrals. The experiments showed that at low PanZ concentrations (200 nM and 1 μM), ADC is not inhibited, and the majority of the substrate had been consumed after 24 hours. Nevertheless, at 10 μM PanZ, the enzyme is almost completely inhibited, with only 5% turnover occurring after 24 hours in the presence of excess AcCoA (Fig. 3.28). The concentration dependence of the inhibition of ADC by PanZ may indicate a possible competitive mode of inhibition. Further experiments using different substrate concentrations would have to be performed to support this hypothesis.
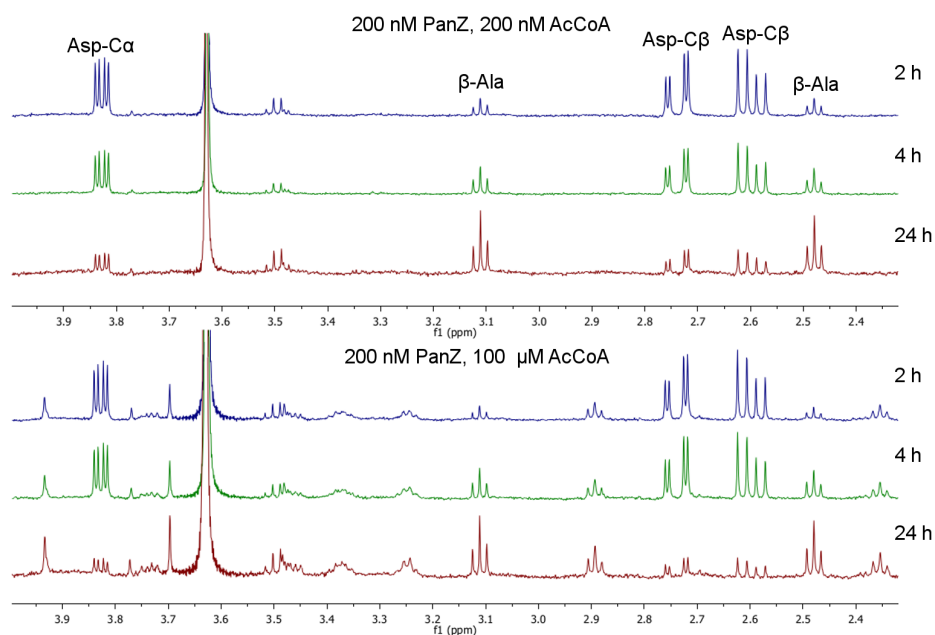
Fig. 3.27 Representative traces of ^{1}H NMR experiments to follow the conversion of L-aspartate to β-alanine by ADC, in the presence of different PanZ concentrations. Spectra were recorded in 10% D_2O with a water suppression pulse sequence. The traces correspond to two experiments with and without excess AcCoA added. Both experiments were run with 100 nM WT ADC, 200 nM PanZ and 0.5 mM L-aspartate in 50 mM phosphate buffer (pH 7.4), 100 mM NaCl. The signals corresponding to L-aspartate and β-alanine are labelled. The excess AcCoA complicates the spectrum but does not obscure the amino acid signals.
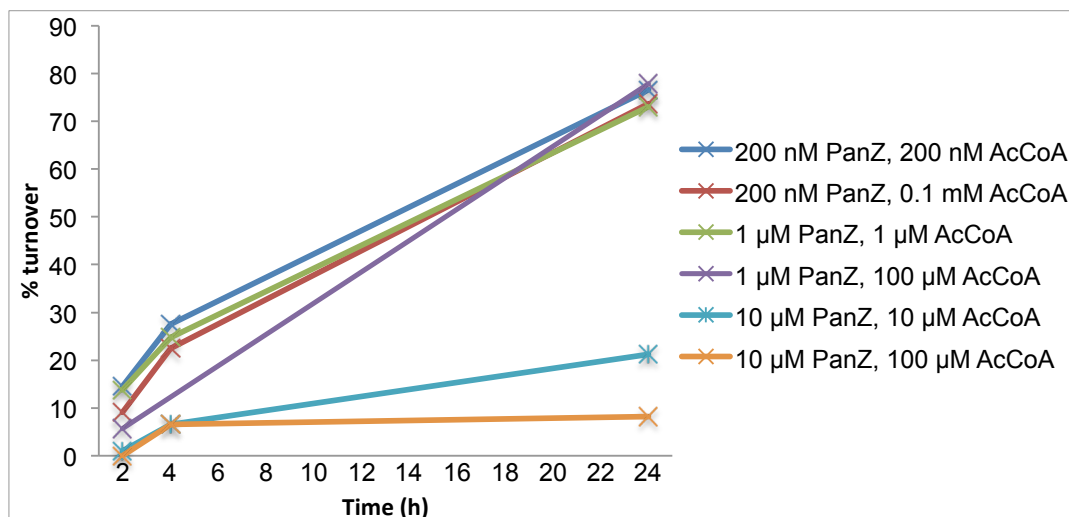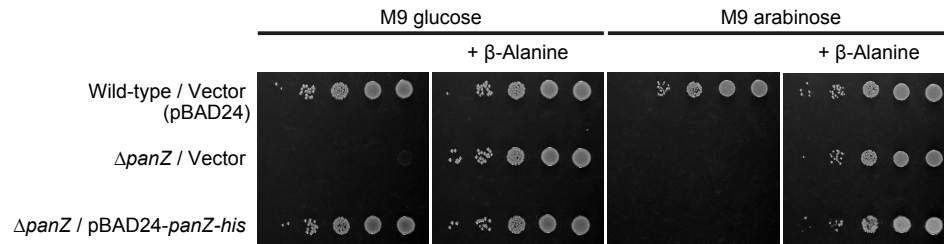
Fig. 3.28 The percentage inhibition of WT ADC by PanZ calculated from ^{1}H NMR spectra. WT ADC is able to catalyse the conversion of L-aspartate to β-alanine in the presence of 1 μM and 10 μM PanZ, but is inhibited in the presence of 10 μM PanZ. All experiments were carried out with 100 nM WT ADC in 50 mM phosphate buffer (pH 7.4), 100 mM NaCl and 0.5 mM L-aspartate. Spectra were recorded in 10% D_2O with a water suppression pulse sequence.

In cellulo assays

The effect of PanZ upon ADC catalysis was also investigated *in vivo*. This work was performed by Dr S. Nozaki as part of this project. Δ*panZ* knockout cells were transformed with a pBAD24.PanZ-his plasmid and grown in M9 minimal media agar plates with and without β-alanine supplementation. Δ*panZ* cells are β-alanine auxotrophs and, therefore, cannot grow without β-alanine supplementation, as PanD cannot be activated to ADC *in vivo* in the absence of PanZ and β-alanine is essential for bacterial production of pantothenate and CoA. Uninduced Δ*panZ* cells carrying the pBAD24.PanZ-his plasmid recovered growth in the absence of β-alanine supplementation, due to leaky overexpression of PanZ which complements the knockout strain. The experiment was repeated in M9 minimal media containing arabinose, which triggers overexpression of PanZ. When PanZ is overexpressed, the cells became β-alanine auxotrophs once again, showing that higher concentrations of PanZ also act to inhibit ADC catalysis *in vivo*.

60

|	M9 glucose		M9 arabinose	
		+ β-Alanine		+ β-Alanine
Wild-type / Vector (pBAD24)				
Δ*panZ* / Vector				
Δ*panZ* / pBAD24-*panZ-his*				

Fig. 3.29 *In cellulo* investigation of the effect of PanZ on ADC catalysis. Serial dilutions of *E. coli* cells were grown in M9 minimal media with and without β-alanine supplementation. Δ*panZ* knockout cells are β-alanine auxotrophs and do not grow in the absence of β-alanine. When complemented with a pBAD24.PanZ-his plasmid, leaky expression of PanZ recovers cell growth in the absence of β-alanine. Under overexpression conditions (agar with arabinose), cells containing the pBAD24.PanZ-his plasmid become β-alanine auxotrophs once again.

Conclusion

The inhibitory effect of PanZ on ADC catalysis, observed both *in vitro* and *in cellulo*, indicates a role for PanZ in negatively regulating CoA production *in vivo*. In fact, this regulatory mechanism may be the primary function of PanZ. These data suggest a model where PanZ acts primarily as a CoA sensor, binding tightly to ADC when the CoA concentration in the cell is high, inhibiting ADC catalysis and slowing down the production of β-alanine. As CoA is depleted, the ADC.PanZ.CoA complex dissociates, releasing ADC and restoring β-alanine production. Higher concentrations of β-alanine then lead to increased concentrations of CoA, eventually promoting once again the complexation of ADC with PanZ.CoA, therefore creating a negative feedback loop mechanism for CoA biosynthesis (Fig. 3.30). Two further mechanisms related to this feedback loop act to control the concentrations of β-alanine, L-aspartate and CoA. First, at high L-asparate concentrations, PanZ.CoA must compete for ADC binding and pantothenate will still be produced, any excess of which can be excreted from the cell (Jackowski and Rock, 1981). Second, ADC undergoes inactivation after ~300 turnovers, which also allows for a more gentle equilibration of the cellular pools of these metabolites (Webb *et al.*, 2004, Konst *et al.*, 2009).

The X-ray structures give two possible mechanisms for the inhibition of ADC by PanZ: PanZ occludes the binding site, preventing the formation of the substrate-enzyme complex, and appears to promote the stabilisation of the pyruvoyl group in a hydrated form, which is unsuitable for catalysis. By binding at the ADC active site for inhibition, PanZ may have evolved a secondary function in accelerating PanD activation in some organisms. As this activation step is irreversible, PanZ has only to bind once and transiently to ADC to catalyse enzyme activation, and this can thus still occur even at low CoA concentrations.
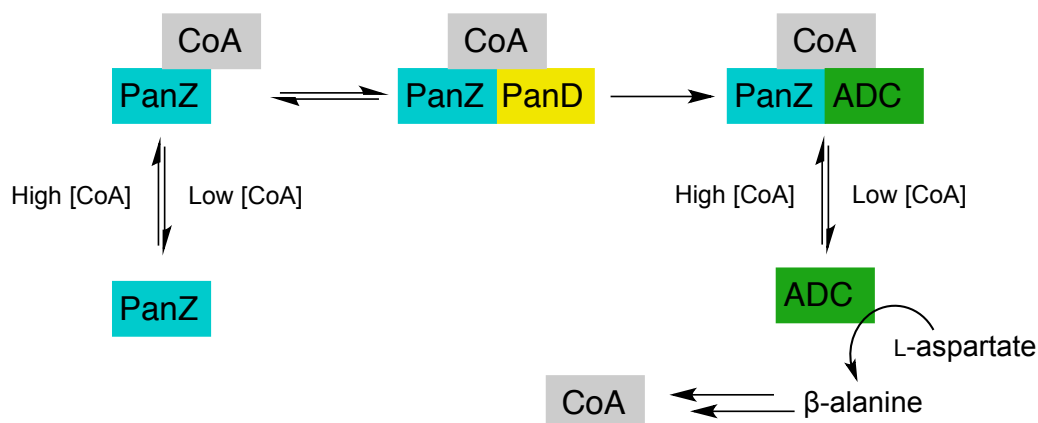
Fig. 3.30 Proposed mechanism for the regulation of CoA biosynthesis *in vivo*. PanZ binding to CoA is dependent on CoA concentration in the cell. PanZ.CoA is able to bind to PanD, activating it to ADC in a irreversible step. Even at low CoA concentrations, PanD activation should still occur, although the binding of PanZ.CoA to PanD/ADC would be only transient. PanZ.CoA is able to stay bound to ADC, inhibiting catalysis, until the concentration of CoA drops, causing the complex to dissociate and allowing ADC to resume catalysis. Increased levels of β-alanine would increase the levels of CoA once more, promoting the association of the proteins and generating a negative feedback loop mechanism.

### *3.2.7* **How PanZ accelerates PanD activation**

A close comparison of the structure of PanD(T57V).PanZ.AcCoA complex with the previously published structure of the PanD zymogen (PDB 1PPY) (Schmitzberger *et al.*, 2003) gave insight into the mechanism of PanD activation by PanZ. Both structures are isostructural throughout for PanD, with the exception of the PanD [17]His-Tyr[22] loop in the free zymogen, which precedes the region of activation PanD-[23]Glu-Cys[26]. The loop is unstructured in the free PanD zymogen (Fig. 3.31, yellow, Fig. 3.32 **A**). Upon binding to PanZ, the loop is shifted 6.4 Å upwards, adopting a beta-sheet conformation, as seen in the PanD(T57V).PanZ.AcCoA complex (Fig. 3.31, green). From the structural overlays, the position of the PanD-[17]His-Tyr[22] loop in the free zymogen is incompatible with the position of PanZ-Asn45 in the complex. PanD-Tyr22 is fully conserved (sequence alignment shown in section 9.7.1) and has not, so far, been shown to be involved in ADC catalysis. Its conservation may, therefore, indicate a function in promoting PanD post-translational cleavage, and this was indeed observed in the PanD(T57V)-PanZ.AcCoA structure. The aromatic ring of Tyr22 packs against the side chain of PanZ-Asn45 and the Cγ carbon of PanD-Glu19 (Fig. 3.33). These interactions promote the formation of the β-sheet at PanD residues [17]His-Gly[19].
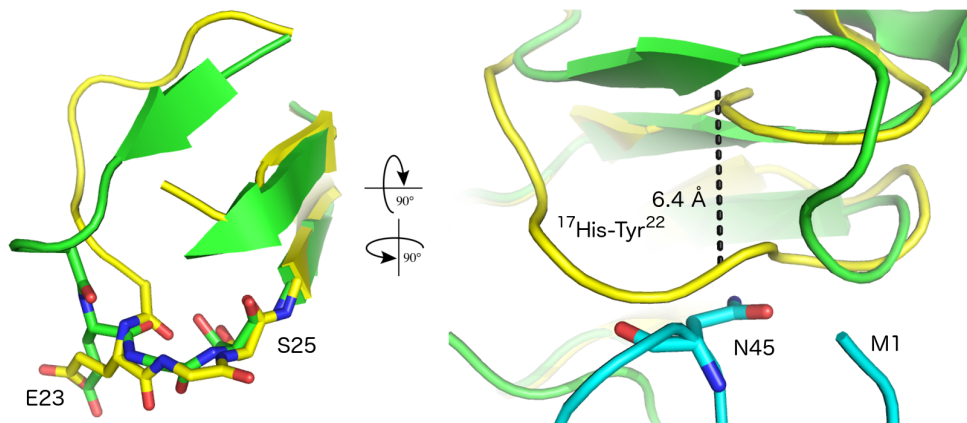
Fig. 3.31 Overlay of the PanD(T57V).PanZ.AcCoA complex structure (PDB 4CRZ, green) (Monteiro *et al.*, 2015) with the major conformer of the PanD zymogen structure (PDB 1PPY) (Schmitzberger *et al.*, 2003), focusing on the processing loop. The ^{23}Glu-Ser^{25} region is similarly structured, whereas the preceding region corresponding to ^{17}His-Tyr^{22} is shifted 6.4 Å upwards upon binding of PanZ (shown in blue).
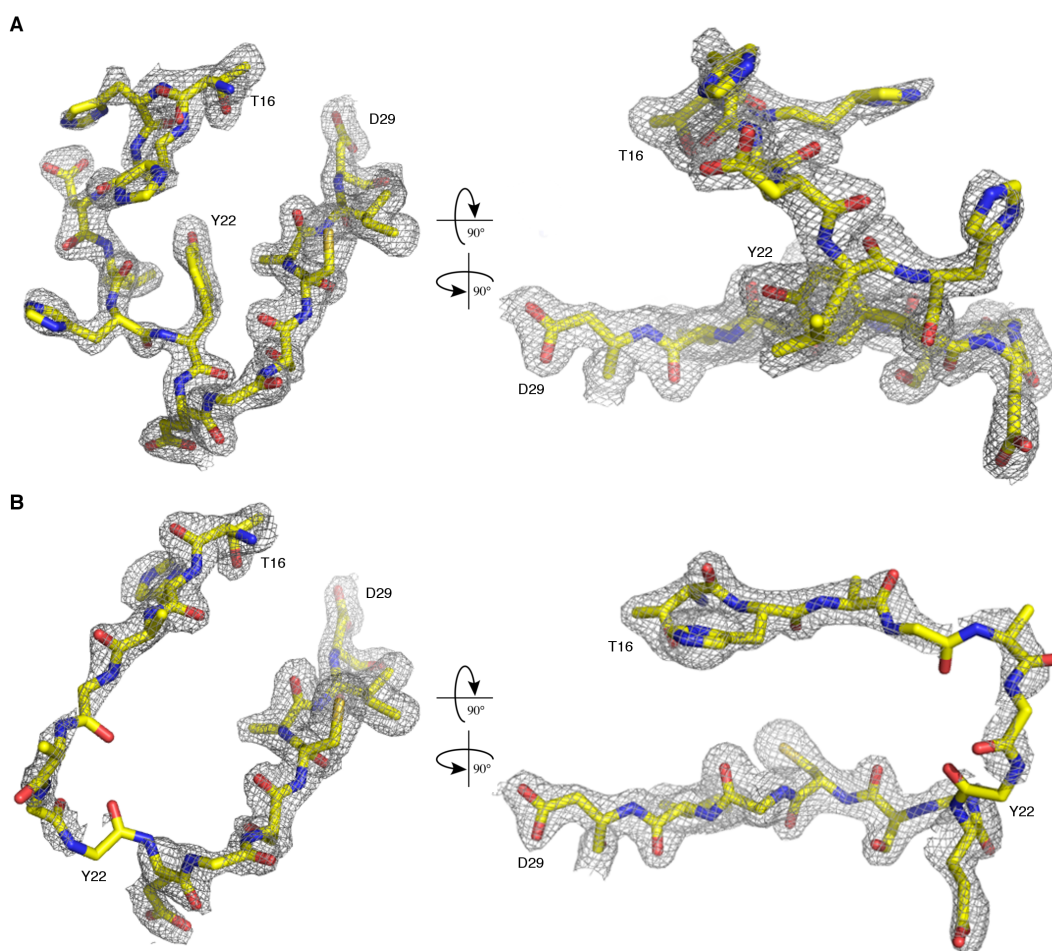
Fig. 3.32 Conformations of the mobile loop of the PanD zymogen in the previously reported X-ray structure (PDB 1PPY) (Schmitzberger *et al.*, 2003). (**A**) The free, unstructured chain, well resolved in the electron density. (**B**) The second structure corresponds to the higher energy conformation of the chain, noticeable by the poorly defined electron density.
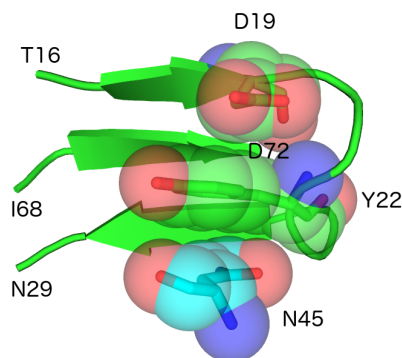
Fig. 3.33 Packing of residues in PanD upon binding of PanZ. PanD(T57V) (Monteiro *et al.*, 2015) is shown in green and PanZ in blue. The PanD processing conformation is maintained by the formation of two parallel and an anti-parallel β-sheets (residues [16]Thr-Asp[19], [68]Ile-Asn[72] and [29]Asn-Tyr[22]). PanZ-Asn45 stacks against the fully conserved PanD-Tyr22, which, in turn, interacts with PanD-Glu19, promoting the formation of the β-sheet and the reorganisation of the PanD-[23]Glu-Cys[26] activation peptide.

As the unstructured conformation of the PanD-[17]His-Tyr[22] loop is preferentially adopted by the free, non-processed PanD zymogen, it is clearly an incompetent conformation for activation of PanD to ADC. The structured conformation, consisting of three β-sheets, is adopted by both the PanD(T57V)-PanZ.AcCoA as well as the free fully activated enzyme (PDB 1AW8) (Albert *et al.*, 1998), indicating that adoption of this conformation is essential for enzyme maturation. Interestingly, this activation-competent PanD conformation of the PanD-[17]His-Tyr[22] loop can also be observed in a small percentage of the uncleaved WT zymogen structure in one of the two proteins present in the asymmetric unit (Schmitzberger *et al.*, 2003). The electron density from which this minor conformation was modelled is badly defined (Fig. 3.32, **B**), an indication that the conformation is unstable. As the enzyme activates slowly *in vitro* (Ramjee *et al.*, 1997) it is hypothesised that, once the zymogen is expressed, the PanD-[17]His-Cys[26] loop can adopt multiple conformations in the absence of PanZ and that the activation-competent conformation, although unfavoured, is still thermally accessible, which would account for many of the early reports of PanZ-independent PanD activation.

The conformational changes within the PanD-[17]His-Tyr[22] loop, along with the direct interaction of PanZ-Asn45 with the backbone amide of PanD(T57V)-Glu23 (Fig. 3.20), cause subtle, but crucial downstream conformational changes to the PanD-[23]Glu-Cys[26] processing loop. The PanZ-Asn45 residue had been previously reported to be essential for PanZ to promote PanD maturation (Nozaki *et al.*, 2012). These changes are responsible for promoting the N-O acyl shift required for post-translational cleavage of PanD to ADC.

The conformation of unstructured PanD-[17]His-Tyr[22] loop seen in the free zymogen clearly shows why this conformation is activation-incompetent. The first step in the N-O acyl shift mechanism is the nucleophilic attack of the PanD-Ser25 hydroxyl on the adjacent PanD-Gly24 carbonyl carbon. The relative spatial arrangement of the reactive centres in the major conformation seen in the zymogen structure is clearly incompatible with the required nucleophilic attack as the PanD-Ser25(OH) and PanD-Gly24(CO) are 4.2 Å apart and parallel to each other (Fig. 3.34, left). Upon binding of PanZ, the carbonyl group of PanD-Gly24 rotates more than 90° upwards, hydrogen bonding to PanD-Tyr58[*] instead of PanD-Thr57 (Fig. 3.34, right)[†]. The hydroxyl of PanD-Ser25 forms a hydrogen bond to the backbone carbonyl of PanD-Glu23. In the new conformation, the distance between the PanD-Ser25 hydroxyl and the carbonyl of PanD-Gly24 is reduced to 3.2 Å and the reaction centres are placed in an almost perpendicular position relative to each other. This new arrangement of the reaction centres is much more compatible with the conformation required for orbital overlap for the nucleophilic attack.
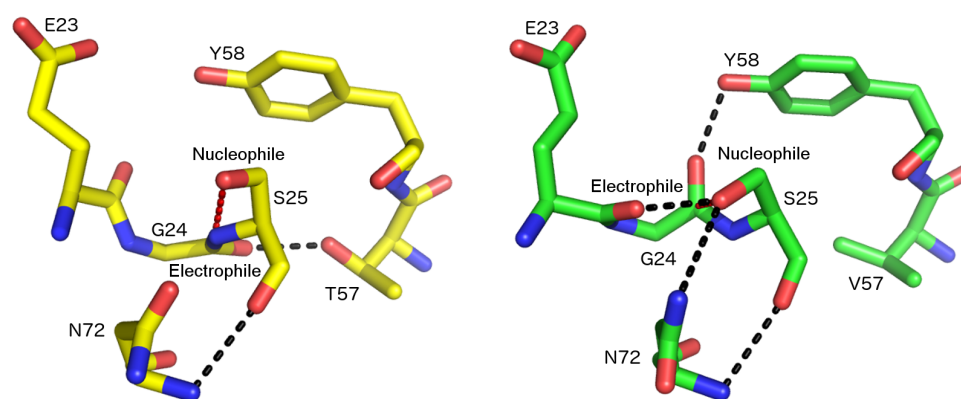
Fig. 3.34 Conformations of the free WT ADC zymogen (yellow) (Schmitzberger *et al.*, 2003) and PanD(T57V) (Monteiro *et al.*, 2015) in complex with PanZ (green). In the WTADC zymogen, Gly24 hydrogen bonds to Thr57. The carbonyl group of Gly24 and the hydroxyl group of Ser25 are parallel to each other and 4.2 Å apart, in a conformation incompatible with nucleophilic attack. Upon binding of PanZ, Gly24 rotates upwards, hydrogen bonding to Tyr58 and is placed perpendicular to the hydroxyl group of Ser25. The two groups are now 3.1 Å apart in a conformation more compatible with nucleophilic attack.

---

[*] Previous studies have shown that mutation of Tyr58 to phenylalanine does not abolish ADC activation. The hydrophilic environment surrounding the activation peptide can likely provide a new hydrogen-bonding partner to Gly24 even in the absence of Tyr58.

[†] The orientation of the side chain of Asn72 differs between the two models. As N and O atoms are indistinguishable in the electron density, the correct assignment of rotamers has to be based on chemical observations. In 4CRZ structure, Ans72 has been built with a rotamer that satisfies the hydrogen-bonding requirements of Ser25.

The ^{24}Gly-Ser^{25} amide bond is severely distorted in the PanD(T57V).PanZ.AcCoA complex, with the orientation of the N-H and C=O bonds diverging from planarity. The distortion is also visible in the conformation adopted by PanD-Ser25, which occupies a Ramachandran disallowed region (Fig. 3.35). These distortions and deviations suggest that the activation loop of PanD is held in a high-energy conformation when binding to PanZ.
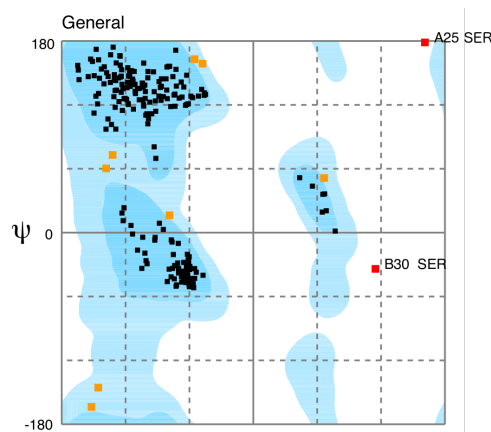

Fig. 3.35 Ramachandran plot of PanD(T57V).PanZ.AcCoA (Monteiro *et al.*, 2015), showing two outliers: Ser25, located in the activation loop of PanD(T57V) and Ser30, located in a mobile loop in PanZ, with poorly-defined electron density. Plot made using RAMPAGE (Lovell *et al.*, 2003).

Such conformationally-induced post-translational modifications have already been described for other systems, such as Ntn-hydrolases, where active-site peptides can explore multiple conformations, only one of which can undergo autoproteolysis (Buller *et al.*, 2012). This class of enzymes uses cis-autoprocessing to cleave an internal peptide bond which yields a catalytically active N-terminal amine (Dodson and Wlodawer, 1998). The N-O acyl shift can use threonine, serine or cysteine residues, but the specific geometries of the reactive conformations of each nucleophile are still unknown. The mechanisms and structural constraints of auto-catalytic post-translational modifications are difficult to study due to the difficulty of trapping a biologically relevant cleavage precursor. Recently, Buller *et al.* were able to elucidate the structure of an unprocessed Ntn-hydrolase. The slow, but activation competent mutant T282C, where the threonine nucleophile is mutated to a cysteine nucleophile, was trapped in two conformations at the cleavage site. From the geometrical analysis of the arrangement of the residues, one of the conformations was suggested to be competent for autoproteolysis whereas the other was incompetent. This flexibility of the cleavage peptide suggests that the protein can explore multiple conformations and that cleavage is initiated when the correct

geometry is adopted. The reactive conformation is favoured as it relieves steric clashes and torsion strains.

3.2.8 The PanD(S25A).PanZ.AcCoA complex

To verify that the observed distorted conformation of the PanD(T57V).PanZ.AcCoA activation peptide was caused by PanZ binding to PanD(T57V) rather than being an effect of the site-directed mutagenesis, the interaction of a second non-activatable PanD mutant, PanD(S25A), with PanZ.AcCoA was investigated. Binding of the two proteins was first investigated by ITC: pairwise titrations of PanD(S25A) into PanZ.AcCoA and PanZ.AcCoA into PanD(S25A) showed a 4-fold decrease in binding affinity of the proteins compared to the binding of PanD(T57V) to PanZ.AcCoA (K_D=157 nM compared to 35 nM for PanD(T57V), Fig. 3.36). The difference in binding affinity was mainly due to a difference in the entropic contribution to binding (ΔS=-26.6 cal mol^{-1} K^{-1} for PanD(S25A) compared to -5.8 cal mol^{-1} K^{-1} for PanD(T57V)), rather than enthalpic contributions (ΔH=-10.2 kcal for PanD(S25A) compared to -9.3 kcal mol^{-1} for PanD(T57V)). The activating loop may be more flexible in PanD(S25A), suffering a higher entropy change upon binding to PanZ.AcCoA.
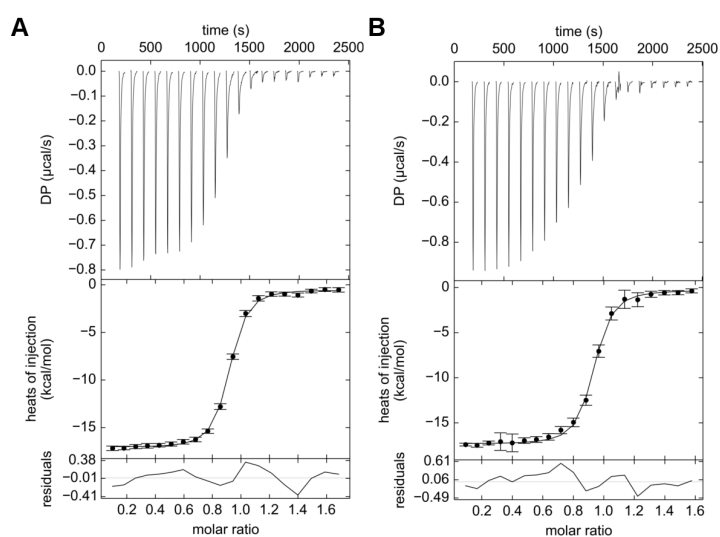
Fig. 3.36 Pairwise titration of PanD(S25A) into PanZ.AcCoA and PanZ.AcCoA into PanD(S25A). (**A**) titration of 257 μM PanD(S25A) into 32 μM PanZ in the presence of 394 μM AcCoA. (**B**) titration of 263 μM PanZ.AcCoA into 35 μM PanD(S25A) in the presence of 400 μM AcCoA.

The PanD(S25A).PanZ.AcCoA complex was crystallised using the previously optimised protocol. Smaller, square pyramidal crystals (50 to 100 μm) were obtained from the same crystallisation conditions used for the previous complexes and data were

collected in-house at RT to 2.1 Å resolution. The data were processed and the structure solved in the same way as the PanD(T57V).PanZ.AcCoA in-house structure (Table 3).

Table 3 PanD(S25A).PanZ.AcCoA in-house X-ray diffraction data reduction and refinement statistics

	PanD(S25A).PanZ.AcCoA
Data collection	in-house
Space group	*I4*
Cell dimensions	
a, b, c (Å)	86.3, 86.3, 80.8
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	33.7-2.10 (2.15-2.10)
R_{merge} (%)	18.1 (62.1)
$R_{p.i.m.}$ (%)	15.5 (51.7)
$<I>$ / sI	5.7 (1.8)
Completeness (%)	99.5 (99.8)
Redundancy	3.2 (3.1)
Refinement	
Resolution (Å)	33.7-2.10 (2.15-2.10)
No. reflections	16351 (1238)
R_{work} (%)	17.27 (24.5)
R_{free} (%)	23.7 (32.2)
No. atoms	
Protein	2021
Ligand/ion	62
Water	62
B-factors	
Protein (main chain) (Å^2)	24.29
Protein (side chain) (Å^2)	28.57
Ligand/ion (Å^2)	28.47
Water (Å^2)	26.86
R.m.s. deviations	
Bond lengths (Å)	0.019
Bond angles (°)	2.09

The complex was once again identical to the previously solved structures, with only small deviations in the activation peptide region. Strain is again observed as a Ramachandran disallowed conformation, this time of Gly24 instead of Ser25 (Fig. 3.37). PanZ-Asn45 binds to the backbone of PanD(S25A)-Glu23 as in the PanD(T57V).PanZ.AcCoA structure, causing the [17]His-Tyr[22] loop to adopt the activation-competent conformation. Due to the smaller size of alanine compared to serine a different conformation of residues 24 and 25 is seen. The carbonyl carbon of Gly24 is rotated away from Tyr58 and the methyl group of Ala25 points towards Thr57 (Fig. 3.38). The previously reported structure of the PanD(S25A) mutant also shows a different conformation for the activation loop to that adopted by the PanD(S25A)

mutant when in complex with PanZ.CoA, confirming that the strained conformation adopted by the PanD.PanZ.AcCoA complexes is promoted by PanZ binding and that the high-energy conformation observed is not caused by the site-directed mutation of Thr57 or Ser25.
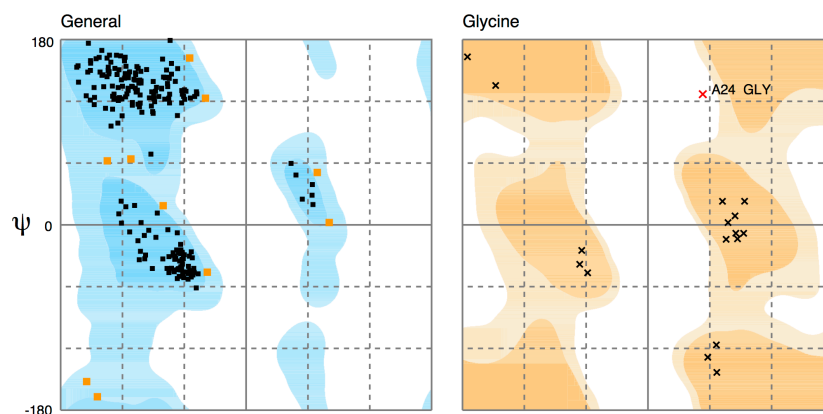
Fig. 3.37 Ramachandran plot of PanD(S25A).PanZ.AcCoA, showing Gly24, located in the activation loop of PanD(S25A), as an outlier.
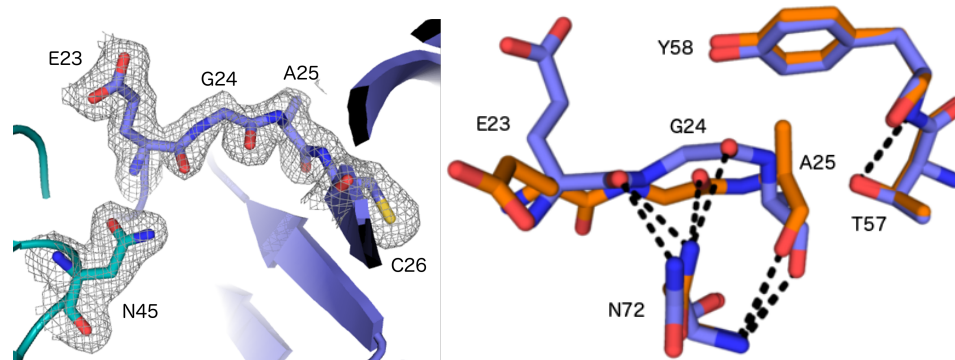
Fig. 3.38 Comparison of the PanD(S25A).PanZ.AcCoA complex to the previously reported structure of PanD(S25A) (PDB 1PQE (Schmitzberger *et al.*, 2003)). In the PanD(S25A).PanZ.AcCoA complex, PanD(S25A) is shown in dark blue and PanZ in teal. 1PQE is shown in orange (**Left**) The PanD(S25A).PanZ.AcCoA activation loop showing the $2F_o$-F_c map contoured at 1 σ. (**Right**) detail of the activation loop, showing the structural differences that occur upon PanZ binding to S25A, causing the strained conformation, with hydrogen-bonds in black dashed lines.

3.2.9 PanZ promotes PanD activation by "mechano-chemistry"

The Ramachandran disallowed conformation adopted by Ser25 in the PanD(T57V).PanZ.AcCoA complex suggests that binding of PanZ to PanD(T57V) promotes strain and distorts the activation region of ADC. This strain was captured more clearly in the PanD(T57V).PanZ.AcCoA dataset collected with the synchrotron X-ray source. The in-house and synchrotron structures are virtually identical, with the exception of the activation site. The in-house collected dataset showed the [24]Gly-Ser[25] peptide intact, as expected, as PanD(T57V) cannot undergo activation or cleavage by

69

PanZ or thermally (Webb *et al.*, 2014). Surprisingly, the synchrotron dataset showed a clear break in the electron density between the two residues (Fig. 3.39). With no prior indication of post-translational cleavage of this ADC mutant in any biochemical or X-ray structural data collected either in-house or with a synchrotron source, the cleavage must be promoted due to the complexation of PanD(T57V) with PanZ.AcCoA.
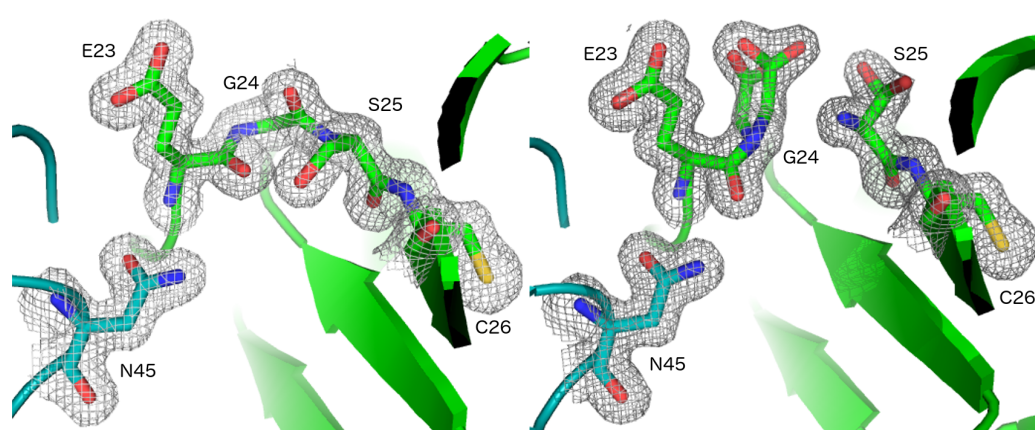
**Fig. 3.39** The activation peptide in the PanD(T57V).PanZ.AcCoA complexes. PanD(T57V) is shown in green, PanZ in teal and the $2F_o$-F_c map as a grey mesh contoured at 1 σ. (**Left**) in-house dataset, showing the intact peptide chain; (**Right**) synchrotron dataset showing the cleaved chain. In the synchrotron dataset, the electron density map shows two alternate conformations for both the terminal methyl carboxylate in Gly24 and the hydroxyl side-chain of Ser25.

To distinguish whether the cleavage was caused by X-ray dose-dependent fragmentation of the polypeptide backbone during data collection or whether complexation of PanD(T57V) with PanZ could lead to the formation of some ester intermediate (Fig. 3.3, **III**) which could be more easily hydrolysed during data collection, both the PanD(T57V).PanZ.AcCoA and PanD(S25A).PanZ.AcCoA complexes in solution were treated with hydroxylamine. Hydroxylamine is a good nucleophile and reacts readily with esters. Solutions of both complexes and individual proteins with and without AcCoA were incubated at 37 °C overnight with added hydroxylamine. The samples were boiled at 100 °C for 10 min, SDS-PAGE loading buffer was added and the samples boiled for a further 10 min. The solutions were loaded onto 10% Tris-tricine SDS-PAGE gels. All proteins ran intact through the gel, showing single bands for each of the proteins at the expected molecular weight (~15 kDa), confirming that PanZ binding to PanD(T57V) or PanD(S25A) does not generate a unstable or easily hydrolysable species and that peptide bond cleavage must have occurred during X-ray irradiation of the complex in the crystal.

The crystal structure shows clear formation of a C-terminal carboxylate at Gly24 and a Ser25 residue (both visible in two distinct conformations). The cleavage did not

promote pyruvoyl formation and, as PanD(T57V).PanZ.AcCoA is not easily hydrolysable, a different cleavage mechanism must occur in the X-ray beam.
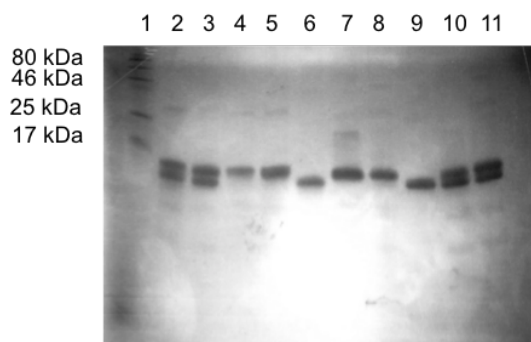
Fig. 3.40 SDS-PAGE analysis of the treatment of PanD(T57V) and PanD(S25A) and corresponding complexes with hydroxylamine. Lane 1: protein ladder; then, from left to right, lanes 2-11: PanD(T57V).PanZ.AcCoA with HONH$_2$, PanD(S25A).PanZ.AcCoA with HONH$_2$, PanD(T57V) with AcCoA and HONH$_2$, PanD(S25A) with AcCoA and HONH$_2$, PanZ.AcCoA with HONH$_2$, PanD(T57V) with AcCoA, PanD(S25A) with AcCoA, PanZ.AcCoA, PanD(T57V) with PanZ, PanD(S25A) with PanZ.

X-ray radiation can interact in different ways with a (protein) crystal (Shmueli, 2010). First, the X-ray photons can collide elastically with the atoms, the resulting coherent scattering giving rise to the diffraction pattern. The X-rays can also collide inelastically, where some of the energy is transferred to electrons, resulting in an excited state which then relaxes back to the ground state releasing a lower energy photon. And finally, the X-ray photon can be completely absorbed by the atom, causing the expulsion of a photoelectron from the inner electron shells. This expulsion of photoelectrons is the major source of radiation damage during X-ray diffraction experiments.

Photoelectrons cause a cascade of ionization events (estimated at ~500 events per 12 keV photoelectron), as well as thermal heating, disruption of the crystal lattice and bond scissions (O'Neill *et al.*, 2002). Radiation damage manifests globally as a reduction in diffraction intensity and resolution, increase in unit cell size, worse data reduction statistics (R$_{merge}$) and increased disorder (Wilson *B* value) with increasing X-ray dose. It also manifests locally as site-specific chemical changes which may lead to biologically irrelevant artefacts that can make interpretation of the biological structure difficult (Garman, 2010). Damage is considerably more pronounced at RT compared to cryo-temperatures, as ionised species can diffuse easily through the crystal. With the recent great increases in synchrotron source brilliance, radiation damage at cryo-temperatures has also become more pronounced, as more photoelectrons are produced and these can still diffuse through the crystal at 100 K.

Specific structural damage within the biomacromolecule occurs by the cleavage of covalent bonds. Certain bonds are more susceptible to radiation damage, and, therefore, specific functional groups will undergo chemical changes at different rates (Weik *et al.*, 2000). At cryo-temperatures, damage is dominated by the tunnelling of electrons through the protein backbone until an electron-affinic site is found. Disulphide bonds break first (Weik *et al.*, 2002), followed by decarboxylation of glutamate and aspartate residues, tyrosine dehydroxylation and cleavage of the methyl group in methionines.

As the protein crystal is composed of 20% to 80% mother liquor by volume, water molecules account for a large percentage of photoionization events, producing •OH radicals as well as solvated electrons, H• radicals and other species (Ward, 1988). At cryo-temperatures, radicals are unable to diffuse through the crystal. But, at RT, the ionised species can diffuse and interact with the protein, spreading the radiation-induced changes quickly.
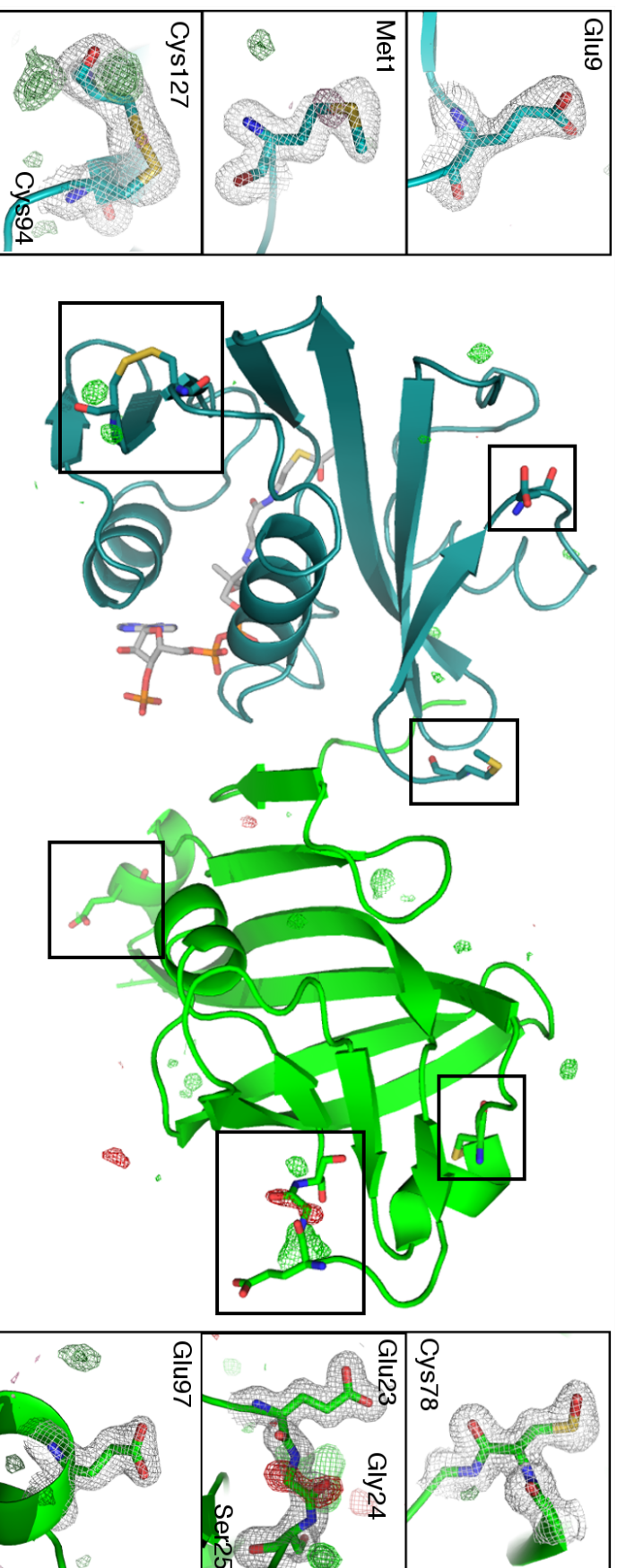
The PanD(T57V).PanZ.AcCoA structure obtained from synchrotron data collection did not show any evidence of global radiation damage. The data collection strategy was designed carefully to maximise the crystal life-time, but even during careful data collection at RT, radiation damage is common. A careful analysis of the structure and electron density (both F_o-F_c and $2F_o$-F_c) maps showed only minor cleavage of the PanZ-Cys127-PanZ-Cys94 disulphide bond and partial cleavage of PanZ-Met1. Neither decarboxylation of glutamates or aspartates nor oxidation of the AcCoA phosphate groups was observed. One clear radiation-induced chemical change was the full oxidation of the PanD(T57V)-Cys78 (Fig. 3.41). As electrons are strongly reducing and •OH radicals oxidizing (Ravelli and McSweeney, 2000), these observations suggest that the main source of radiation-induced chemical modifications to the proteins during this RT synchrotron data collection was caused by •OH radical chemistry.

As the crystal tolerated the full data collection and no "typical" radiation induced damage is visible, the cleavage of the [24]Gly-Ser[25] amide bond, clearly visible as negative difference electron density when the intact chain is modelled, is a completely novel radiation-induced chemical event, specific to this protein complex system. To date, no instances of radiation-induced cleavage of protein backbone amide bonds have been reported. Amide bonds are not good electron sinks or easily susceptible to radical chemistry. As discussed previously, binding of PanZ to PanD(T57V) causes a conformational change in the cleavage peptide, placing Gly24 and Ser25 in the correct geometry for nucleophilic attack and also in a strained conformation, reflected in the

Ramachandran disallowed conformation of Ser25. The cleavage of the amide bond in the synchrotron dataset confirms the high-energy state of the new conformation of the peptide chain, as a ground state amide bond would not undergo radiation-induced cleavage. The chemistry is likely promoted by a hydroxyl radical attack on the carbonyl carbon of the highly distorted, high-energy amide bond, causing cleavage of the bond to the N-terminal serine and C-terminal glycine residues.

The amount of radiation dose absorbed by a crystal during X-ray data collection is measured in Grays (1 Gy=1 J Kg^{-1}). The dose absorbed by the crystals during in-house and synchrotron data collection was calculated with RADDOSE 3D (Zeldin *et al.*, 2013), and found to be 20 kGy and 70 kGy respectively. Dose is the only significant difference between the two datasets collected and so this cleavage event seems to be dose-dependent. This observation is in accordance with previous studies (Southworth-Davies *et al.*, 2007), that show that X-ray radiation-induced chemical and structural changes to protein crystals during data collection is dose-dependent.

From the structural information and radiation-induced chemistry described here, a revised mechanism for the activation of PanD can now be proposed. First, the PanD activation loop, [23]Glu-Cys[26] (Fig. 3.42, **I**), undergoes a conformational rearrangement upon binding of PanZ. This forces the loop into a high-energy, "activating" conformation: the Gly24-Ser25 amide bond is severely distorted and susceptible to nucleophilic attack and the hydroxyl group of Ser25 is placed close proximity and in the correct conformation for nucleophilic attack upon the Gly24 carbonyl (**II**). Tyr58 hydrogen bonds to the carbonyl of Gly24, aiding the formation of the oxyoxazolidine ring intermediate (**III**). Thr57 can then act both as proton donor and acceptor during the ring opening and β-elimination steps (**III**, **IV**). This new proposed role of Thr57 as the hydrogen donor/acceptor, couples the final two steps leading to the dehydroalanine intermediate, avoiding the possible direct hydrolysis of the ester intermediate, commonly seen as a side reaction during the thermal activation of PanD. The conformation observed in the PanD(T57V).PanZ.AcCoA structure also shows a clear hydrogen bond between the Ser25 hydroxyl and the backbone carbonyl of Glu23, which acts as a proton shuffling residue to aid the deprotonation of Ser25.

Fig. 3.41 The asymmetric unit cell of the PanD(T57V).PanZ.AcCoA structure. PanD(T57V) is shown in green, PanZ in teal and AcCoA as grey sticks. The $2F_o$-F_c map is contoured at ± 3 σ (light green and red) and ± 4 σ (dark green and red). 4 sites typically prone to radiation damage are highlighted (Cys127-Cys94, Glu97, Met1 and Glu9) showing no or minimal damage. Cys78 is fully oxidised. The ADC processing loop is modelled intact to show the clear difference density map indicating full cleavage.

74
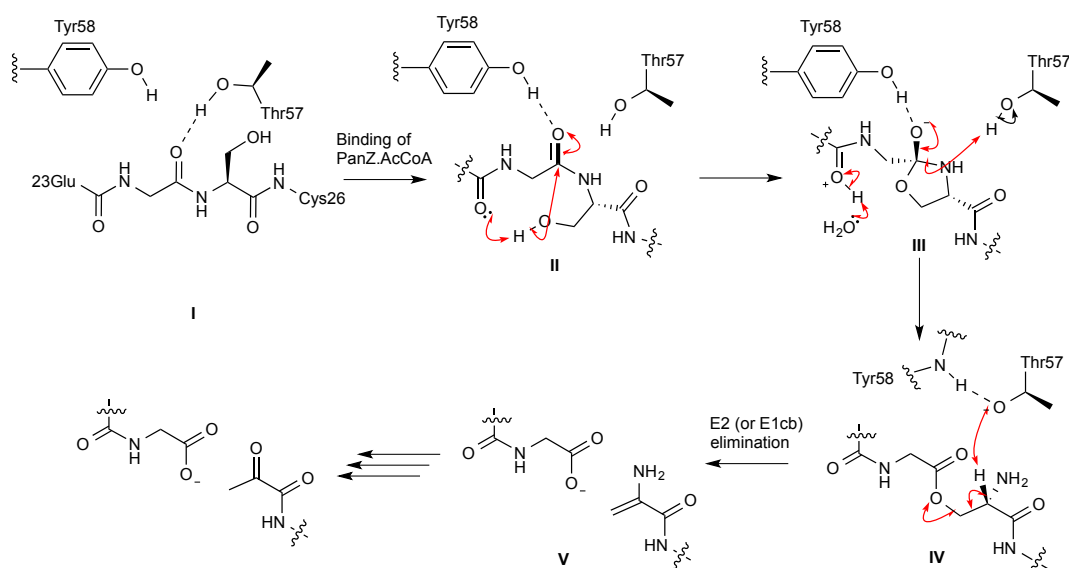
Fig. 3.42 Proposed revised mechanism of PanD activation to ADC in PanZ producing bacteria. The mechanism can be generalised to all PanD homologues, if the first conformational change is caused by folding of the protein backbone after translation. The PanD backbone adopts the reactive conformation upon binding of PanZ (**II**) which allows for the nucleophilic attack of Ser25 on the distorted Gly 24 carbonyl. Tyr58 stabilises the ring intermediate (**III**), which collapses to the ester intermediate (**IV**) with protonation from the Thr57 hydroxyl. Ester cleavage occurs by β-elimination mediated by Thr57 to a dehydroalanine (**V**), which then hydrolyses to the pyruvoyl group.

Conformational strain leading to amide distortions, which, in turn, translate into a loss of resonance stabilisation of the amide bond, making it susceptible to attack, have been suggested in recent structural and theoretical studies of the Muc1 SEA domain (Johansson *et al.*, 2008, Sandberg *et al.*, 2008). Like ADC, the SEA domain undergoes autoproteolysis by an N-O acyl shift mechanism, employing a serine as the nucleophile. The uncleaved protein chain cannot be easily accommodated in the folded protein. As a consequence of folding, the cleavage peptide is distorted, lowering the free energy barrier of activation and placing the residues in the correct geometry for the initial nucleophilic attack. Mutations to the core of the protein that lower the melting temperature of the protein also affect the rate of cleavage of the peptide.

3.3   Conclusion

This project was initiated parallel to the photochemical work to understand the molecular basis of the maturation of PanD to ADC and the involvement of PanZ as an activation catalyst. Using ITC, the two proteins were shown to interact tightly and to form a heterooctameric protein complex. The binding interaction was also shown to be

CoA dependent, firstly from the ITC data and then later supported by both native mass spectrometry and NMR.

Novel X-ray structures of WT ADC in complex with PanZ as well as two PanD mutants (PanD(T57V) and PanD(S25A)) in complex with PanZ, confirming the requirement of CoA for the interaction, were determined. The crystallographic studies were non-trivial and required extensive screening and the careful design of diffraction experiments to overcome the crystal pathologies encountered (low resolution, twinning and lattice deformations). With the introduction of cryo-cooled techniques for crystal handling and improvement of crystal life-times during data collection, RT data collection has fallen into misuse. It is a more time-consuming and labour intensive way of collecting data at synchrotrons but, with the current advancements in hardware and software, it is a technique that is gaining some revival. For this project, it was resorted to as an attempt at improving diffraction quality but the data also captured much more interesting aspects of the system. A clear, radiation-induced backbone cleavage of the non-processing mutant PanD(T57V) in complex with PanZ.AcCoA allowed for the proposal of a new cleavage mechanism of PanD to ADC, where the ^{23}Glu-Ser^{25} loop region adopts a high-energy, strained, cleavage-competent conformation upon the binding of PanZ. The distortion of the ^{24}Gly-Ser^{25} amide bond, as well as the tightening of the loop and placement of the nucleophilic Ser25-OH group in a better position for attack on the Gly24 carbonyl group explain, from a chemical point of view, how this N→O acyl shift is efficiently promoted *in vivo*.

Crystallisation of PanD with PanZ.AcCoA showed complete activation of the enzyme, which indicates that PanZ.CoA is able to efficiently catalyse the maturation of PanD to ADC, without the formation of hydrolysed side-products. The complex was structurally similar in the PanD(T57V), PanD(S25A) and WT ADC complexes: PanZ binds to the same interface in all cases, across the active site. PanZ binds to ADC, regulating its activity, as shown by the decrease of catalytic activity of ADC in the presence of PanZ *in vitro* (by ITC and ^{1}H NMR assays) as well as a β-alanine auxotrophic phenotype in *E. coli* cells overexpressing PanZ, and this has led to a proposal for a new mode of regulation of the CoA biosynthetic pathway.

As both the PanD.PanZ.AcCoA and ADC.PanZ.AcCoA structures were isostructural they show that PanD can be efficiently activated to ADC either thermally, by heating the expression *E. coli* cell pellet containing the overexpressed zymogen at 37 °C prior to

purification, or by incubation with PanZ.CoA. For simplicity and ease, thermal activation was chosen as the preferred method for the preparation of fully active ADC.

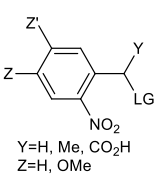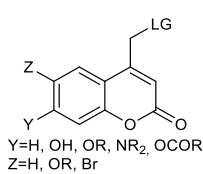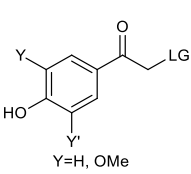# 4 Synthesis of photocleavable crosslinkers

4.1 Introduction

Photocaging of bioactive molecules is a non-trivial process and, although the first studies date back to the 1970s (Kaplan *et al.*, 1978) and numerous papers regarding the synthesis, photolysis and use of these compounds are available, the information about the spectroscopic and chemical properties of caged biomolecules is scarce and incomplete. This lack of information is mainly due to the nature of the work. Most photocaged compounds have been used in cellular studies where fast time-resolution is not a requirement and the photocages are released using continuous, low-power illumination (Hagen *et al.*, 2005). The described rates and yields of cleavage of these compounds are difficult to compare as, usually, the percentage of cleavage is quoted without a reference to the power of the illumination source used. Furthermore, similar photocaging groups have been photolysed at varying wavelengths in different studies. With the scarcity of comparable information, in most cases, the relative efficiency of different photocaging approaches can only be evaluated qualitatively.

Nevertheless, some studies about the effect of substituents, leaving groups and cleavage conditions on the efficiency of photolysis and product release of photocages have been previously reported (Corrie *et al.*, 2005, Klán *et al.*, 2013). Most of the information available is qualitative, where different scaffolds are compared under dissimilar conditions. In cases where both rate and yield of photocleavage information is available, the values are usually calculated from the fragmentation of the compounds following a short (ns) high-power pulse of light. In many studies though, the yield of cleavage has only been determined using continuous, low-power irradiation, and the time necessary for release of the compound is quoted instead. This information is useful for compounds employed in cellular studies, where the processes being investigated are slow (seconds to minutes) but not for fast time-resolved studies, where a well-defined triggering time is necessary.

The efficiency of cleavage is defined as the percentage of release of compound after the photolysing light pulse and is influenced by both the extinction coefficient (ε, proportion of photons absorbed) and quantum yield (φ, fraction of molecules that cleave following absorption of a photon) of the photocaged system. The quantum yield is dominated by the nature of the photocaging scaffold (e.g. *ortho*-nitrobenzyl, coumarin or *para*-hydroxyphenyl, Table 4) as this determines the de-caging mechanism. Higher

extinction coefficients improve the photolysis efficiency, as the compound is able to absorb light more strongly, although this must be balanced with the sample thickness. Although the time-resolved X-ray diffraction experiments for single turnover systems require a significant percentage of the molecules in the ensemble to be triggered simultaneously and this can be promoted by a higher ε, compounds with very high ε will absorb light so strongly that laser penetration through the sample will be very low. The extinction coefficient can be modulated by the addition of substituents to the chromophoric centre, where electron-donating groups tend to lead to an increase in the value of ε (see section 4.1.1) (Corrie *et al.*, 2005, Klán *et al.*, 2013). Even with these general guidelines, the exact spectroscopic and chemical properties of a photocaged compound can only be fully determined once the compound is synthesised. Nevertheless, a small overview of the aspects to consider when designing a photocleavable scaffold is given below. For a more extensive description of some of the trends observed for *ortho*-nitrobenzyl (*o*NB), coumarin (Cm) and *para*-hydroxylphenyl (*p*HP) photoprotecting group scaffolds see section 9.6.[*]

Table 4 General properties of photocaging groups

group	*o*-nitrobenzyl	coumarin	*p*-hydroxyphenyl
λ_{max} (nm)	254-345	320-390	280-304
ε_{max} ($M^{-1}cm^{-1}$)	600-27000	6000-20000	9000-15000
φ[†]	0.64-0.01	0.02-0.30	0.9-0.03
k (s^{-1})	$10 - 3\times10^4$	$1\times10^8 - 2\times10^9$	$1\times10^7 - 2\times10^9$
solubility (H_2O)	Poor-Medium	Poor-Medium	Good

The focus of this project was on the design and synthesis of novel and stable photocleavable crosslinking reagents that can be used to crosslink a protein and also that can release two cysteine residues upon photolysis. The choice of photocage was mainly based on the synthetic availability of the general designed scaffold, but, nevertheless, it is important to understand how altering photocaging groups affects their performance during photolysis.

---

[*] For a non-exhaustive list of examples of caged compounds to date, please see section 9.5.

[†] The range of values is quoted as a trend relative to the range of λ_{max}. φ tends to decrease with increasing λ_{max} for *o*NB and *p*-hydroxyphenyl groups and tends to increases with λ_{max} for coumarins.

4.1.1 Photocaging scaffolds

The main considerations guiding the design of new photocaging strategies are:

<u>Wavelength of photolysis</u>

Proteins and nucleic acids have absorption maxima at wavelengths of 280 nm and 260 nm respectively. Ideally, a photocaged compound will absorb light at longer wavelengths, to avoid filtering of the laser light by the macromolecule which lowers the number of photons available to be absorbed by the photocage and may also lead to UV-induced chemical damage to the protein (Bensasson *et al.*, 1993). Especially for time-resolved experiments, the laser power and pulse length used has to be optimised to obtain rapid and uniform decaging across all the molecules in the sample and avoid radiation damage.

<u>Rate of photolysis</u>

The timescales of the processes that can be observed during a time-resolved experiment are completely dependent on how quickly the system is triggered. The rate of decaging has to be faster than that of the reaction of interest. The timescales achievable are completely dependent on the photocaging moiety used, as the chemical mechanism of decaging dictates how fast the product can be released (Table 4). *o*NB photocages are the slowest (ms-µs, see section 9.6.1), followed by coumarins and *p*HP groups (ns, see sections 9.6.2 and 9.6.3).

<u>Yield of compound release</u>

Ideally, all the product molecules should be released from the photocaged derivatives upon irradiation. A low decaging yield can be caused by low penetration of the triggering laser through the sample, which is especially problematic for dense samples (such as crystals) or when targeting highly absorbent samples (with very high ε). Another cause for partial decaging is a low quantum yield of the photocage, which is a property of the compound itself. Nevertheless, it is still possible to obtain useable data from fractional triggering of the sample, as long as the event is well synchronised and sufficient molecules are released to obtained a good signal-to-noise ratio, by accounting for the unreacted fraction of the sample during data analysis, as long as the level of excitation can be determined (Ren *et al.*, 1999).

Photocaging groups are aromatic and, in most cases, poorly soluble in aqueous buffers. A small percentage of organic solvent (dimethyl sulfoxide or dimethyl formamide preferentially) may have to be used to aid solubility, but the concentration of organic solvent tolerated is dependent on the protein and the sample conditions. Furthermore, the photocaged compounds must be stable in the specific buffer used for at least the duration of the experiment, but preferentially for several days to aid sample preparation.

Synthetic availability

Taking all of the above considerations into account, in principal optimal photocaged compounds for each specific experiment can be designed, but it is also important to ensure that the compound can be synthesised easily and cleanly, and in sufficient quantity for characterisation and use in the time-resolved experiment.

4.2 Photocleavable crosslinker design and rationale

Beyond the aspects related to the photocaging moiety, several other details must be considered when designing the photocleavable crosslinker scaffolds. The proposed compounds aimed to fulfil as many of the following requirements as possible:

- Bioconjugation reactivity: two cysteine-reactive moieties to selectively and cleanly crosslink to engineered protein surface cysteine residues;
- Length: variable crosslinking length through an extendable linker to afford a high degree of control over conformational locking;
- Photolysis: clean cleavage of the crosslinker and release of the conformational lock;
- Photolysis products: cleavage should yield a minimally modified protein, releasing two free cysteines where possible.

Two general scaffolds that could accommodate these requirements were designed (Fig. 4.1). In the first design, both caged cysteines are attached to a single photolabile group. The cage may release each cysteine sequentially to yield the fully deprotected protein, although such compounds have not been described in literature and, therefore, the exact mechanism of cleavage of such a scaffold can only be extrapolated. The second design resembles a more typical photocaging strategy, where each cysteine residue is released by the cleavage of one photocage. In this case, two photolabile moieties are linked through an inert and extendable linker and each carry one cysteine-reactive moiety.
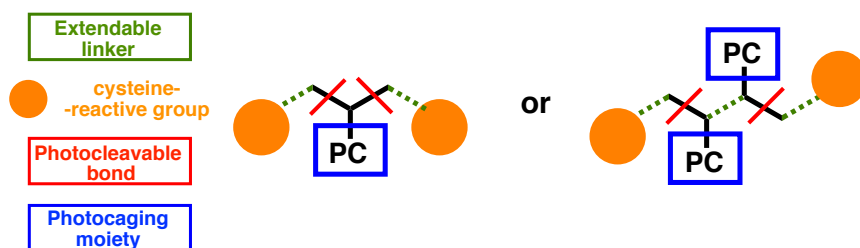
Fig. 4.1 Cartoon representation of two possible photocleavable crosslinker scaffolds that fulfil the specifications described above. In both cases, the chromophore is represented as a blue box, the cysteine-bioconjugating moiety as an orange circle, the possible extendable linker regions as dashed, green lines and the photocleavable bonds crossed with a red line.

4.3 Alkyl bis-α-bromo-*o*NB crosslinkers
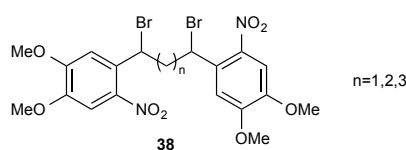
Fig. 4.2 General structure of the previously published alkyl bis-α-bromo-*o*NB crosslinkers, showing the different components: 4,5-dimethoxy-*o*NB chromophore (blue), extendable alkyl linker (green), cysteine-reactive secondary bromides (orange) and the photocleavable bond in red.

At the start of this project, only one family of compounds had been previously reported in the literature that matched this desired set of characteristics (Omran and Specht, 2009b, Omran and Specht, 2009a). The reported scaffold contained two 4,5-dimethoxy-*o*NB (DM*o*NB) groups connected by an extendable alkyl linker (**38**, Fig. 4.2). The DM*o*NB C_α is brominated, which should selectively react with surface cysteine residues. The compounds cleave upon UV irradiation at 364 nm, releasing the free cysteines, and the expected mechanism of cleavage is shown in Fig. 4.3 (Il'ichev *et al.*, 2004). Crosslinker **38** is first bioconjugated to the protein via two surface cysteines (**39**). The "stapled" protein can then be irradiated with UV light, causing the excitation of the photocage (**40**), which decays to the *aci*-nitro intermediate **41**. Intermolecular rearrangement leads to a cyclic intermediate (**42**) which can cleave, releasing a cysteine reside to give the nitrobenzyl ketone by-product (**43**).

This family of bis-α-bromo-*o*NB compounds (Fig. 4.2) was synthesised first. The described synthetic route varied depending on the length of the linker group and initial work was carried out on the shorter (n=1 and n=2) crosslinkers. The shortest crosslinkers were investigated first (n=1), as these would yield the tightest control over conformational locking, with reduced flexibility compared to the longer scaffolds (n=3).
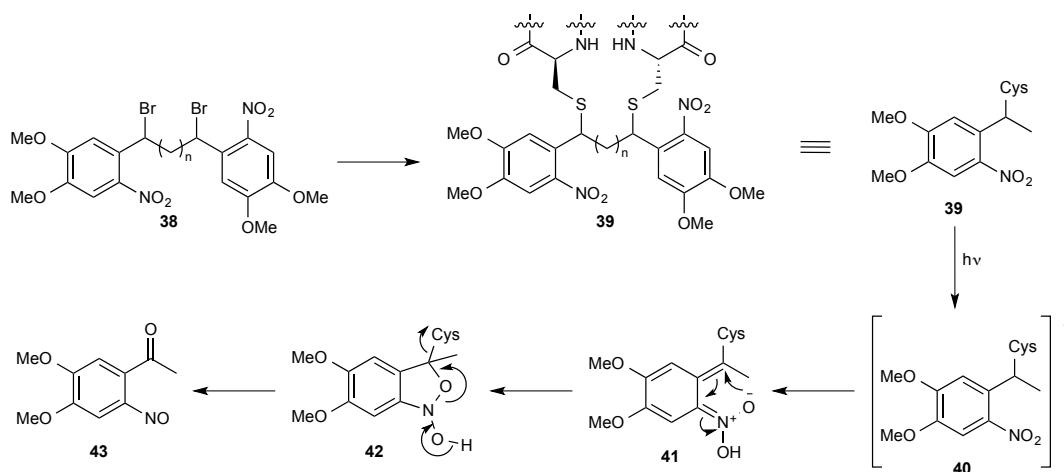
**Fig. 4.3** Expected reaction and cleavage mechanism for the crosslinkers. The di-bromo crosslinker (**38**) reacts with cysteine residues on the protein surface, stapling the protein (**39**). Upon UV irradiation, the compound is placed in a singlet excited state (**40**), which decays to an *aci*-nitro intermediate (**41**). This intermediate then cyclises to a benzisoxazoline ring (**42**), which opens, releasing the cysteine residues and yielding the nitroso by-product (**43**).

4.3.1 Synthesis of 1,3-dibromo crosslinkers
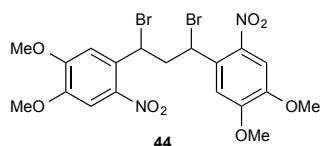
The synthetic route to the 1,3-dibromo-*o*NB crosslinker (**44**) is shown in Fig. 4.4 (Omran and Specht, 2009a). Claisen condensation of ethyl 3,4-dimethoxy benzoate (**45**) with 3,4-dimethoxy acetophenone (**46**) yields the first intermediate, a 1,3-diketone (**47**). The reported reaction conditions used benzene at reflux temperature. To avoid the use of benzene, the reaction was attempted in toluene instead. The 1,3-diketone (**47**) was obtained in moderate yields with longer reaction times (36 h) and at higher temperatures (110 °C). 1,3-Diketone **47** had to be reduced to the corresponding 1,3-diol (**48**) prior to nitration, as the 1,3-unsaturated system leads to deactivation of the ring towards electrophilic aromatic substitution. The 1,3-diol **48** was obtained quantitatively using a standard sodium borohydride reduction protocol but the following nitration step to compound **49** was unsuccessful. To avoid over-nitration of the highly activated ring (due to the two methoxy substituents), the reaction was performed at low temperature (-50 °C). LC-MS analysis of the reaction showed the presence of tetra-nitrated species, even with low-temperature quenching of the reaction mixture, possibly from the formation of unstable alkyl-nitrates from the alcohols as the crude products decomposed during flash chromatography.
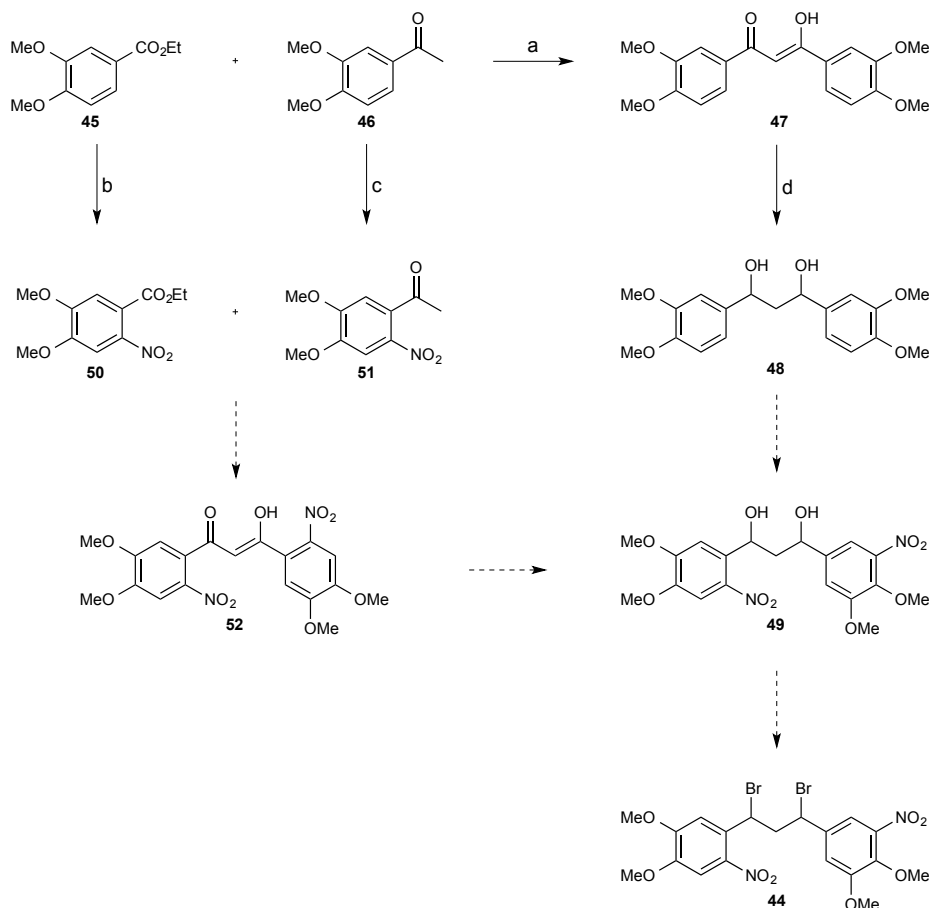
Fig. 4.4 Proposed synthetic route to 1,3-bis(bromo)-1,3-bis(*o*NB) crosslinker (**44**). a) NaH (1.1 eqv.), dry PhMe, 110 °C, 36 h, 33%; b) HNO₃, AcOH, 0 °C-RT, 24 h, 60%; c) Ac₂O, HNO₃, 0 °C-RT, 3 h, 69%; d) NaBH₄ (3 eqv.), CH₂Cl₂, MeOH, RT, 40 min, 99%.

To overcome the nitration problems, a second synthetic route was designed, starting from ethyl 2-nitro-4,5-dimethoxy benzoate (**50**) and 2-nitro-4,5-dimethoxy acetophenone (**51**). The precursors were easily nitrated using HNO₃/AcOH (Knesl *et al.*, 2006) and HNO₃/Ac₂O (Dyer and Turnbull, 1999) mixtures respectively, but the Claisen condensation step to 1,3-diketone **52** was unsuccessful using the previously employed conditions, with both starting materials recovered and no evidence of conversion visible by either TLC or LC-MS.

In the interest of time, and because it was important to investigate early on whether this class of compounds had the desired reactivity for protein conjugation, the short crosslinkers were abandoned at this stage and instead the longer crosslinkers (**53**, n=3) were synthesised. In contrast to the shorter 1,3-crosslinkers, the nitration step for the longer 1,5-crosslinkers could be performed on the 1,5-diketone intermediate instead of the 1,5-diol, and it was therefore expected that the previously described synthetic problems would not be encountered.

84

4.3.2 Synthesis of 1,5-dibromo crosslinkers
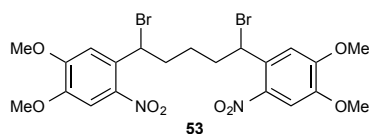
The synthetic route to the 1,5-crosslinker (**53**) is shown in Fig. 4.5 and is applicable to all compounds of this scaffold with linker lengths of 3 or more carbons. The 1,5-diketone homodimer (**54**) was synthesised from the Friedel Crafts acylation of vetrarole (**55**) with glutaryl chloride (**56**) in moderate yields. The work-up of the reaction proved challenging, with the formation of emulsions from aluminium complexes. Quenching of the reaction using Rochelle salts did not improve the work-up but quenching into a 2 M HCl/ice slurry with very slow warming to RT (overnight stirring) gave clear layers which could be better extracted, although without improvement of the isolated yield. The 1,5-diketone (**54**) was nitrated to compound (**57**) using a TFA/HNO$_3$ mixture at 0 °C, reduced quantitatively to the 1,5-diol (**58**) using a standard sodium borohydride reduction protocol and easily brominated using acetyl bromide with catalytic HBr to give the desired 1,5-dibromo crosslinker (**53**) as a 1:1 mixture of diastereomers.
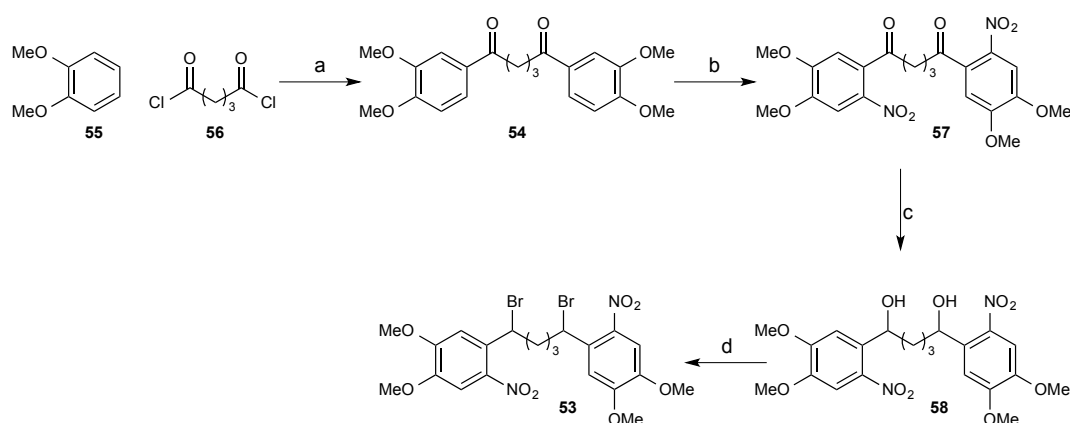
Fig. 4.5 Synthetic route to 1,5-bis(bromo)-1,5-bis(oNB) crosslinker (**53**). a) AlCl$_3$ (2.4 eqv.), -10 °C – 0 °C, 5 h, 22%; b) HNO$_3$ (2.6 eqv.), TFA (8 eqv.), CH$_2$Cl$_2$, -15 °C - 0 °C, 2 h, 53%; c) NaBH$_4$ (4 eqv.), CH$_2$Cl$_2$, MeOH, RT, 45 min, 99%; d) acetyl bromide (3 eqv.), HBr (20 mol%), 0 °C-RT, 1.5 h, 99%.

4.3.3 Reactivity studies

Omran *et al.* performed the crosslinking experiments, in the presence of 20 eqv. of N-benzoylcysteine methyl ester as a surrogate for protein surface cysteines, in 50% EtOH in phosphate buffer (pH 7.3). This protocol was considered unsuitable for protein modification for several reasons: firstly, the high percentage of ethanol will probably not be compatible with most proteins. The high equivalents of protein would also be a

problem, as protein expression yields may not be sufficient and separating the modified and unmodified proteins would be very difficult.

The crosslinking reactions were therefore performed with two eqv. of cysteine or other thiol containing small molecule (Fig. 4.6) and followed by LC-MS. Crosslinker consumption was monitored using the UV trace for each run. An overview of the experiments performed is summarised in Table 5.
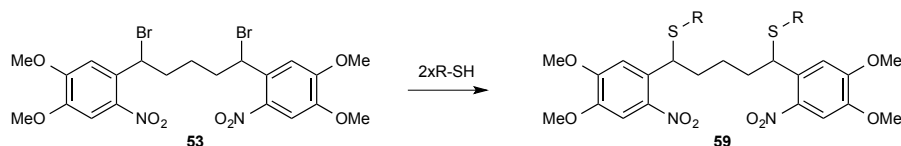
Fig. 4.6 Schematic of derivatisation of the 1,5-dibromo crosslinker (**53**) with thiols, to give 1,5-dithiols (**59**).

In all cases, the crosslinker was still visible after 4 h – none of the reactions went to completion. Cysteine was used first as the nucleophile, followed by cysteine ethyl ester as it was thought that cysteine might not be fully soluble if in its zwitterionic form. No change was visible when performing the reaction in the same buffer/EtOH conditions described in the literature, using only 2 eqv. of nucleophile. The following reactions were attempted with benzyl mercaptan (BnSH), so the reactions could be followed better using UV absorption. Tetrabutylammonium iodide (TBAI) was used as an additive, to see if some exchange of the bromides to iodides could occur and accelerate the reaction, but no change was observed between conditions with and without TBAI. Some single crosslinking was visible in most cases, occurring faster at higher pH (pH 8.0 rather than 7.4), as expected for a nucleophilic substitution. Higher concentrations also increased the rate of reaction and the compounds tolerated DMSO:buffer mixtures.

As none of the conditions attempted led to complete consumption of the crosslinkers or clean formation of the doubly substituted product, it was concluded that these compounds bearing secondary bromides were not sufficiently reactive for protein bioconjugation. Nevertheless, differences were noticed when exchanging the organic solvent and pH, indicating better reactions in DMSO rather than EtOH and at pH 8.0 compared to pH 7.4, which should be taken into account when performing future crosslinking experiments.

Table 5 Investigation of crosslinking ability of 1,5-dibromo crosslinkers (CL, **53**)

Solvent	Thiol	[thiol] (mM)	[CL] (mM)	Additive	pH	Conclusions
1:1 EtOH:buffer	Cys	2	1	-	7.4	Cysteine and CysOEt are sparingly soluble in 50% ethanol
1:1 EtOH:buffer	CysOEt	2	1	-	7.4	
1:1 EtOH:buffer	CysOEt	2	1	TBAI	7.4	
CH$_2$Cl$_2$	BnSH	2	1	-	-	Crosslinking does not occur in the absence of buffer.
CH$_2$Cl$_2$	BnSH	2	1	TBAI	-	
1:4 CH$_2$Cl$_2$:EtOH	BnSH	2	1	-	-	
1:4 CH$_2$Cl$_2$:EtOH	BnSH	2	1	TBAI	-	
1:1 DMSO:buffer	BnSH	2	1	-	7.4	The crosslinking is very slow with 1 mM crosslinker. Addition of TBAI has no effect.
1:1 DMSO:buffer	BnSH	2	1	TBAI	7.4	
1:1 DMSO:buffer	BnSH	5	2.5	-	7.4	Reaction is faster at pH 8.0. Changing the concentration of DMSO from 50% to 33% has no effect on rate of reaction.
1:1 DMSO:buffer	BnSH	5	2.5	-	8.0	
1:2 DMSO:buffer	BnSH	5	2.5	-	7.4	
1:2 DMSO:buffer	BnSH	5	2.5	-	8.0	

4.4 α,α-Di(bromoacetyl) toluene crosslinkers

In order to improve the reactivity of the crosslinkers towards thiol containing compounds and, therefore, cysteines, a more reactive bioconjugation moiety had to be engineered into the scaffold. α-Bromoacetates are not as reactive as α-iodoacetates, but still react readily and exclusively with sulfhydryl groups when used in equimolar concentrations (Hermanson, 2008a). Furthermore, as carboxylates are good leaving groups, photolytic cleavage of the crosslinkers to the cysteine methyl carboxylates should be fast. Therefore, the newly designed scaffold carries two α-bromoacetate groups, both attached to the same photocleavable moiety (Fig. 4.7). Such compounds have not been previously described in the literature and the possible hydrolysis of the ci-esters in buffer was of concern in the design of the crosslinker. Nevertheless, cyclic acetals bearing benzyl groups have been previously used as linkers for protein affinity purification. 6-Membered ring cyclic acetals were shown to be stable in buffer, even in the presence of TFA 1% (Lee *et al.*, 2015).
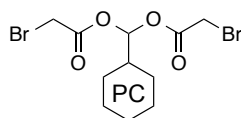
Fig. 4.7 General scaffold design for the α,α-di(bromoacetyl) toluene photocleavable crosslinkers. One photocaging moiety is attached to two α-bromoacetates which can react readily with cysteines and cleave sequentially to methyl carboxylates.

As these photocleavable di-bromoacetyl compounds are completely novel, the cleavage mechanism of the crosslinkers can only be speculated. In the case of *o*NB based crosslinkers, the first moiety is likely to be released by a typical *o*NB photolysis mechanism as shown in Fig. 4.3 (Klán *et al.*, 2013). The α,α-di(bromoacetyl)-*ortho*-nitrotoluene (**60**) is reacted with two cysteines to give the crosslinked moiety (**61**). Upon absorption of a photon, the nitrobenzyl moiety cleaves, releasing the first cysteine methyl carboxylate group (**62**) and forming a nitroso-anhydride intermediate (**63**). This nitroso intermediate is no longer photocleavable and, therefore, the anhydride will have to fragment, possibly by hydrolysis to 2-nitrobenzoic acid (**64**), to release the second methyl carboxylate (Fig. 4.8). Therefore, although the crosslinking reagent is symmetrical, the cleavage mechanism is expected to occur in two distinct steps, which are expected to have very different rates as the hydrolysis step is likely to be slow (c.f. $k \sim 3 \times 10^{-3}$ s^{-1} for acetic anhydride in H$_2$O) (Gold, 1948) in comparison to the first photolysis reaction, which is expected to occur on ms timescales (see section 9.6.1).
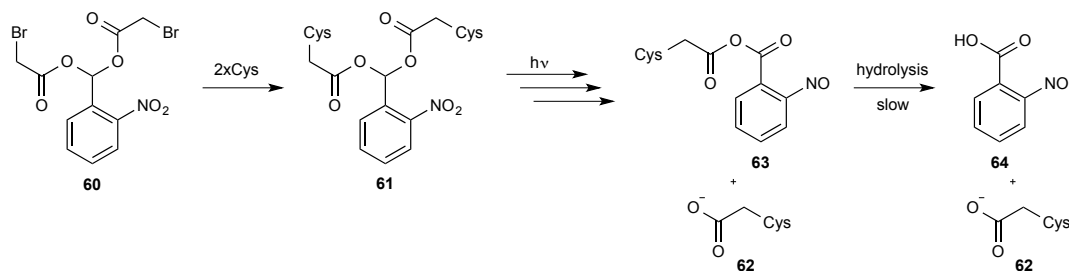
Fig. 4.8 α,α-Di(bromoacetyl)-*ortho*-nitrotoluene (**60**) reaction and cleavage. The crosslinker (**60**) reacts with two surface cysteines to yield the crosslinked protein (**61**). Photolysis of the chromophore leads to the release of one of the cysteine anchors as a cysteine methyl carboxylate (**62**) and yields a 2-nitrosobenzyl anhydride group (**63**), which is still attached to the second cysteine residue. The anhydride intermediate then slowly hydrolyses to 2-nitroso benzoic acid (**64**), releasing the second cysteine residue.

In the case of coumarinyl crosslinkers, a different cleavage profile is expected (Fig. 4.9). The 4-(α,α-dibromo)methyl coumarin (**65**) is reacted with the two surface cysteine residues to yield the crosslinked protein (**66**). Upon absorption of a photon, the coumarin is promoted to an excited state (**67**), which undergoes heterolytic cleavage releasing the cysteine methylcarboxylate (**62**). The resulting carbocation (**68**) is

quenched by the buffer, giving an alcohol intermediate (**69**). This species should then still be able to absorb a second photon (**70**), releasing the second cysteine methyl carboxylate residue (**62**) in a similar fashion to the first. Diol **72** would probably be short lived, collapsing to aldehyde **73**. Although the photolysis events would have to be sequential, coumarins have higher absorption cross-sections and generally much faster cleavage rates (10s of ns) compared to *o*NB derivatives. A second possible mechanism involves the direct fragmentation of alcohol intermediate **67** to diol **64**.
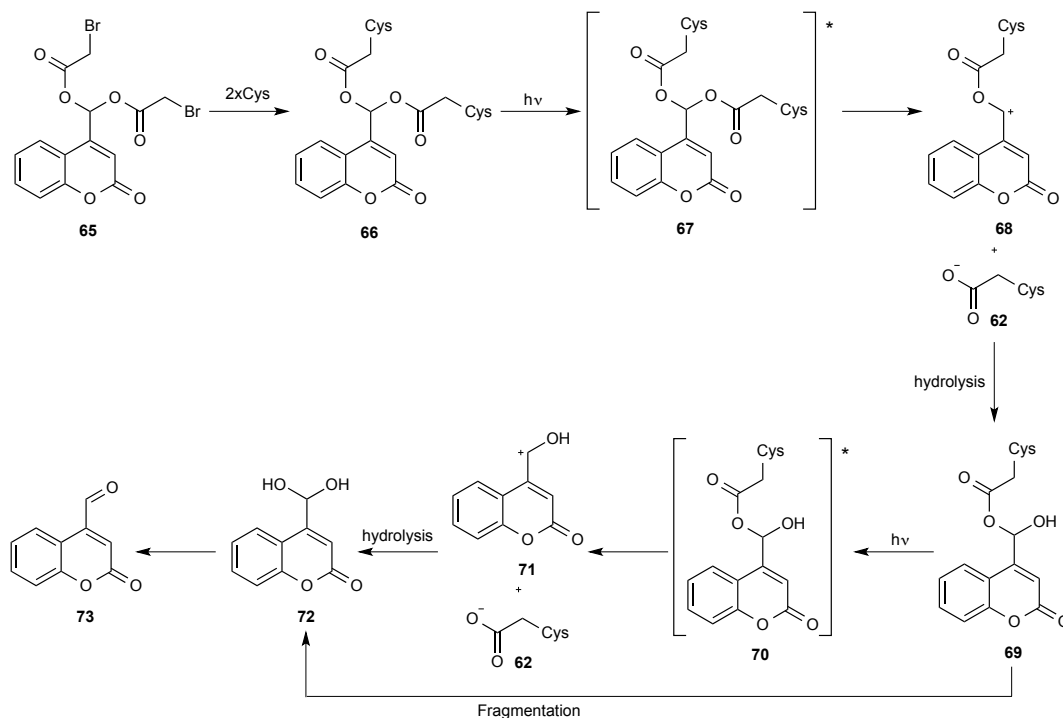
Fig. 4.9 7-(α,α-Di(bromoacetyl))-methylcoumarin reaction (**65**) and cleavage. The crosslinker (**65**) reacts with two surface cysteines to yield the crosslinked protein (**66**). Photolysis of the chromophore leads to the release of one of the cysteine anchors as a cysteine methyl carboxylate (**62**) and yields a coumarinyl carbocation (**68**), which hydrolyses to an alcohol (**69**) in buffer. This intermediate is still attached to the second cysteine residue but is able to absorb a second photon, undergoing a similar cleavage mechanism to release the second cysteine residue and another coumarinyl cation intermediate (**71**). The carbocation hydrolyses to a diol (**72**), which quickly interconverts to the coumarinyl aldehyde (**73**).

In order to simplify the initial synthesis and to allow for rapid testing of this novel class of crosslinkers for reactivity towards cysteines and stability in aqueous solutions, a non-cleavable, benzyl crosslinker was synthesised first.

4.4.1 Synthesis

The non-cleavable α,α-(dibromoacetyl) toluene crosslinker (**74**) was synthesised from benzaldehyde (**75**) and bromoacetic anhydride (Fig. 4.10). Previously reported procedures for the synthesis of diacetyl toluene require heating benzaldehyde in neat

acetic anhydride. As bromoacetic anhydride is a solid and considerably more expensive than acetic anhydride, reactions in solvent (CCl$_4$ and MeCN) with silica-supported sulphuric acid as a catalyst were investigated first, but no conversion was observed (Table 6). Bromoacetic anhydride melts at 40 °C, and therefore the reaction was attempted in neat anhydride. It was found that high temperatures gave better yields, and the dibromoacetyl toluene compound was obtained in excellent yield from overnight heating at 85 °C. The product was a slow-crystallising, colourless oil, which could be triturated in cold hexane to remove any traces of anhydride and bromoacetic acid. Since the ^{1}H NMR spectra of a mixture of the starting materials and of the product are very similar, product formation was confirmed by IR, where the C=O stretch frequencies shift from 1832 cm^{-1} for bromoacetic anhydride to 1763 cm^{-1} for the product.
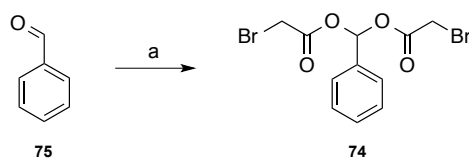
Fig. 4.10 Synthesis of the α,α-di(bromoacetyl) toluene non-photocleavable crosslinker. a) bromoacetic anhydride (1.5 eqv.), 85 ° C, 16 h, 75%.

Table 6 Attempted reaction conditions for the synthesis of di-(bromoacetyl) toluene

Solvent	[PhCO] (mM)	(BrAc)$_2$O (eqv.)	T (°C)	catalyst	Time (h)	Conclusions	Ref.
CCl$_4$	720	1	RT	H$_2$SO$_4$/SiO$_2$	48	Silica supported sulphuric acid is not an efficient catalyst	(Olah and Mehrotra,
MeCN	1150	1	60	H$_2$SO$_4$/SiO$_2$	20		(Kavala and Patel, 2005)
Neat	-	1.5	60	-	16	High temperatures and longer reaction times are required	
Neat	-	1.5	85	-	3		(Rahman and Jahng, 2007)
Neat	-	1.5	85	-	16		

4.4.2 Reactivity and stability studies

The α,α-di(bromoacetyl) toluene crosslinker (**74**) was first investigated for reactivity towards thiols. The reaction was followed by LC-MS, so a dilution series of the crosslinker in MeCN was carried out first to check the detection limits of the spectrometer. Good UV and MS traces were obtained down to 5 mM crosslinker concentration in MeCN (Fig. 4.11).

90
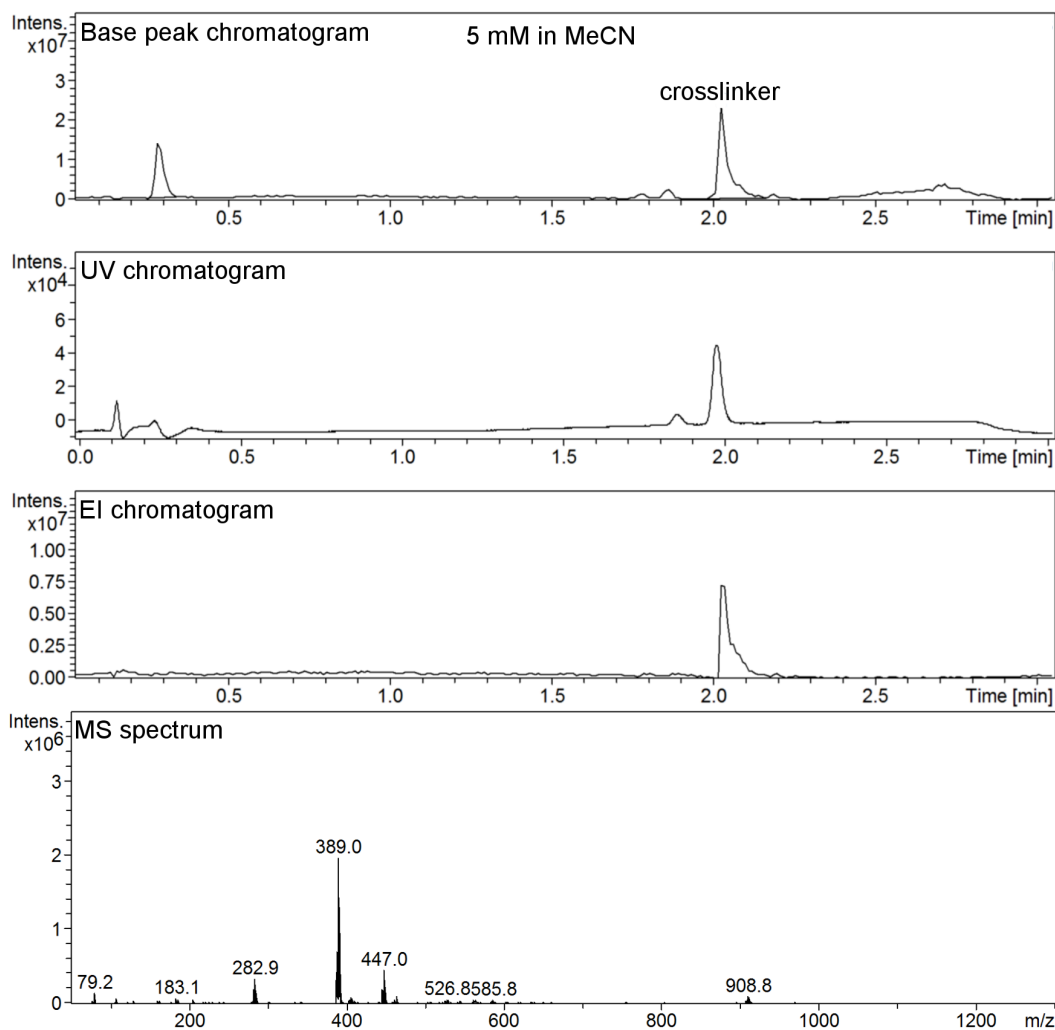
Fig. 4.11 LC-MS run of a 5 mM solution of the α,α-(dibromoacetyl)toluene crosslinker (**74**) in MeCN. From top to bottom: base-peak chromatogram showing the counts on the ion detector during the LC run, showing one major peak at 2.05 min retention time; UV chromatogram during the LC run showing one main UV peak corresponding to the main peak on the chromatogram; extracted ion chromatogram for the expected mass for the crosslinker, matching both the UV and MS peaks; m/z spectrum of the major peak, showing a mass of 389.0, corresponding to the crosslinker adduct with Na^+.

The compound was dissolved to a concentration of 5 mM in buffer (50 mM potassium phosphate pH 8.0, 100 mM NaCl, 5 mM TCEP) containing a varying concentration of MeCN (10-35%) and found to be insoluble up to 30% MeCN, with the appearance of heavy oil emulsions. The same procedure was used to check for solubility in buffer/DMSO mixtures instead, where white suspensions were visible up to 15% DMSO, but with no oily droplets formed.

Crosslinking experiments in 35% MeCN/buffer or 35% DMSO/buffer mixtures with 2 eqv. of mercaptopropionic acid (**76**) were set up and the reaction followed by LC-MS. The reaction was slow in MeCN with both crosslinker and mono-substituted product

(**77**) still visible after 30 min. The LC trace also showed several impurities, including the formation of an additional heavier by-product. The reaction did not proceed to completion, even with prolonged reaction times.
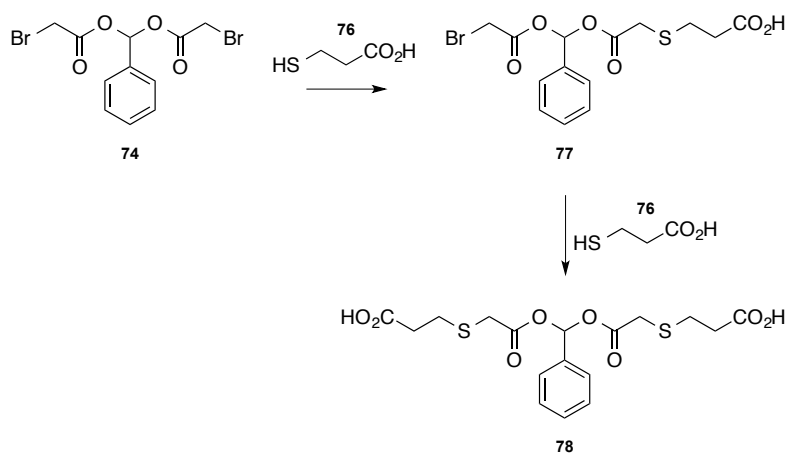
Fig. 4.12 The reaction of the α,α-(dibromoacetyl)toluene crosslinker (**74**) with mercaptopropionic acid (**76**), through the mono-substituted intermediate (**77**) to the expected di-substituted product (**78**).

The reaction was much faster and cleaner in 35% DMSO/buffer, reaching completion within 30 min with no observable mono-substituted intermediate (**77**). Only two species were visible after this time, corresponding to the desired di-substituted product (**78**) and a small amount of the same heavy by-product observed in the reaction in MeCN.

Although the reactions were run in high percentages of DMSO, this was mainly due to the high concentrations needed for observation of the compounds by LC-MS. In practice, the reactions with proteins will be carried out at much lower concentrations (down to ~10 μM), which should allow for a much lower percentage of organic solvent to be used. Therefore, it was concluded that the new crosslinkers bearing α-acetyl bromides were sufficiently reactive and compatible with protein bioconjugation.
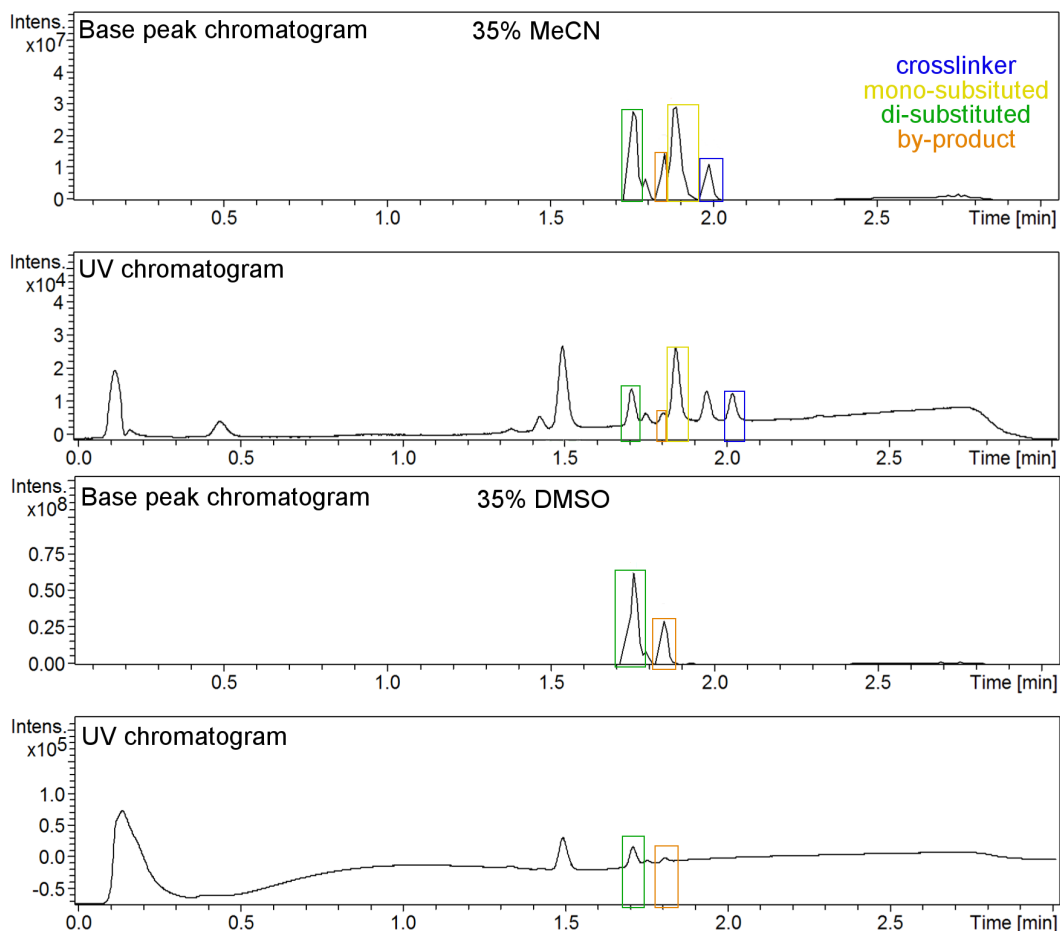
Fig. 4.13 LC-MS traces of the reaction of α,α-(dibromoacetyl)toluene (**74**) with mercaptopropionic acid (**76**), in 35% MeCN and 35% DMSO in buffer. Both the UV and base peak chromatograms are shown. The starting material (blue, **74**), mono-substituted intermediate (yellow, **77**), product (green, **78**) and heavy by-product (orange) are highlighted. The reaction in DMSO is much cleaner and faster than that in MeCN.

For further characterisation of the crosslinker, α,α-(dibromoacetyl)toluene (**74**) was derivatised with ethane thiol on a larger scale, in MeCN in the presence of triethylamine (Fig. 4.14).
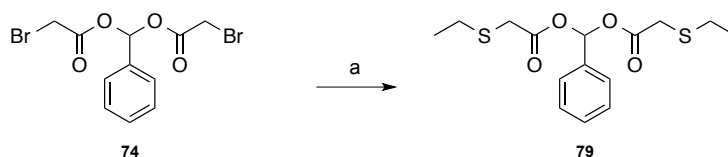
Fig. 4.14 Reaction of α,α-(dibromoacetyl)toluene (**74**) with ethane thiol. a) EtSH (20 eqv.), NEt$_3$ (4 eqv.), MeCN, 3 h, RT, 33%.

The stability of both the crosslinker (**74**) and α,α-di(ethanethiol) toluene (**79**) in buffer at different pH was investigated next, using ^{1}H NMR spectroscopy. The compounds were dissolved in 50% d$_6$-DMSO in buffer at varying pH (50 mM potassium phosphate pH 3.0-8.0, 100 mM NaCl) and the ^{1}H NMR spectrum recorded with a water

93

suppression pulse sequence at different time intervals. The α-proton signal of the crosslinker was compared to the PhC2/C6 proton signal of the aldehyde to give a ratio of the two species and, therefore, an approximate measure of hydrolysis. The α-proton signal overlays with the PhC3/C5 aldehyde proton signals so the integral was calculated as the difference between the total signal for both protons and that of the PhC2/C6 proton.

The crosslinker (**74**) was found to be unstable across the whole pH range, with full hydrolysis to the aldehyde within 1 h at pH 7.0-8.0. Hydrolysis was slower at lower pH, but 30% hydrolysis was still observed within 2.5 h at pH 3.0. The decomposition of the di(bromoacetate) mainly occurs from base-catalysed hydrolysis, but the compound is still susceptible to acid-catalysed hydrolysis.
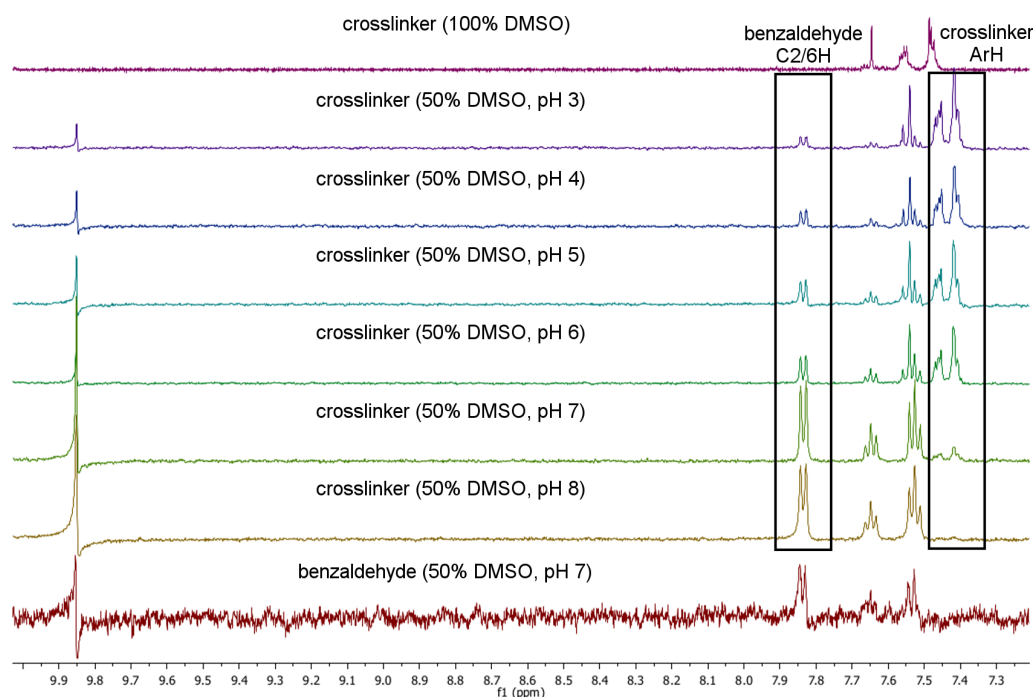
Fig. 4.15 ^{1}H NMR spectra of the α,α-(dibromoacetyl) toluene crosslinker (**74**) in 50% d_6-DMSO in buffer at different pH (50 mM potassium phosphate pH 3.0-8.0, 100 mM NaCl). ^{1}H NMR spectra were recorded between 1h and 2.5 h after sample preparation. From top to bottom: 100% d_6-DMSO; pH 3.0 at 160 min, pH 4.0 at 140 min, pH 5.0 at 120 min, pH 6.0 at 100 min, pH 7.0 at 80 min, pH 8.0 at 60 min, and benzaldehyde in at pH 7.0.

The thiol derivative (**79**) was more stable, hydrolysing to give benzaldehyde only after 16 h at pH 8.0 and not showing any hydrolysis at pH 5.0 and below (Fig. 4.16). The presence of the α-bromides promotes faster hydrolysis rates and allows for acid-catalysed hydrolysis of the di(acetate) core. Although the bioconjugated crosslinker was still not stable above pH 7.0, hindering its wider application, especially in crystallographic studies where the pH of the crystallisation conditions can be an

94

essential factor in obtaining well-diffracting crystals, it was nevertheless deemed important to investigate the synthetic availability of the photocleavable derivatives.
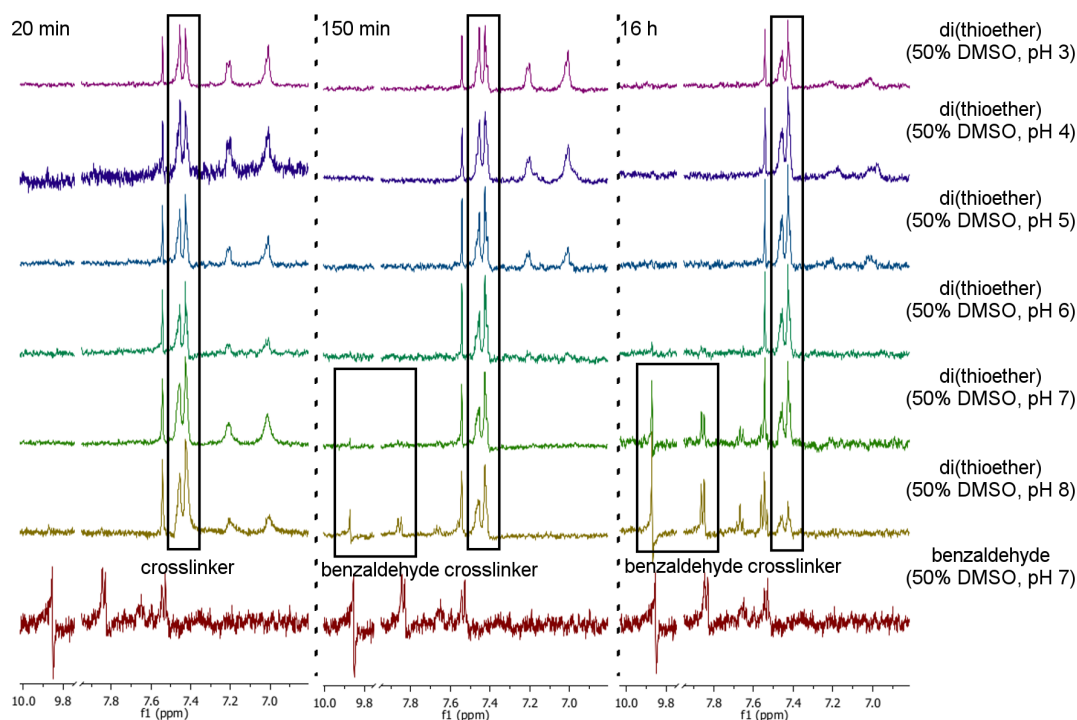
Fig. 4.16 ^{1}H NMR spectra of the stability studies of the α,α-(di(ethylthio)acetyl) toluene derivative (**79**) in 50% d$_6$-DMSO in buffer at different pH. From left to right: spectra at 20 min, 150 min and 16 h. From top to bottom: pH 3.0, pH 4.0, pH 5.0, pH 6.0, pH 7.0, pH 8.0 and benzaldehyde at pH 7.0.

4.4.3 Synthesis of photocleavable derivatives

Although the synthesis of the toluene crosslinker was simple and high yielding, attempts at synthesising the corresponding 7-methoxy coumarinyl and 4,5-dimethoxy-2-nitrobenzyl crosslinkers (**80** and **81** respectively) were unsuccessful, mainly due to solubility issues. Both aldehyde precursors (4-formyl-7-methoxycoumarin, **82**, and 4,5-dimethoxy-2-nitrobenzaldehyde, **83**) were synthesised as described in sections 4.5.3 and 4.5.4, respectively, and subjected to the same reaction conditions used for the synthesis of the toluene crosslinker (**74**), by heating in bromoacetic anhydride.
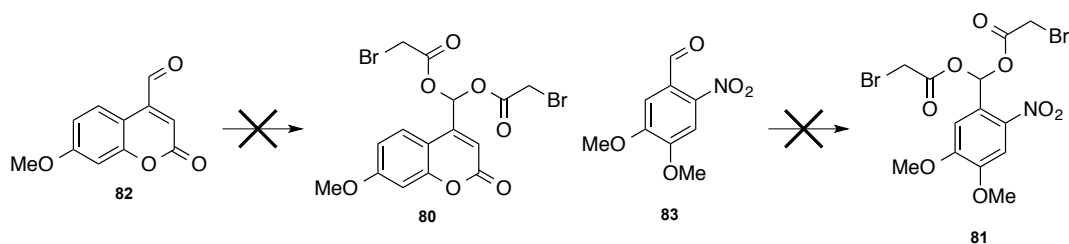
Fig. 4.17 Attempted synthesis of the coumarinyl and *o*NB crosslinkers (**80** and **81** respectively).

4-formyl-7-methoxycoumarin (**82**) was sparingly soluble in bromoacetic anhydride, with solutions only obtained after several hours of heating. Overnight heating at 85 °C showed no consumption of the starting materials and heating at 120 °C led only to the decomposition of the anhydride, with no consumption of the coumarin. Several reaction conditions were attempted, as shown in Table 7, including acid and base catalysis and microwave irradiation. Heating the coumarin in 10 equivalents of bromoacetic anhydride at 85 °C led to the formation of product **80** (as seen by LC-MS), but the compound decomposed during purification by flash chromatography. Trituration of the crude mixture in hexane and distillation of the excess anhydride did not yield pure products, and the crude mixture decomposed back to the starting material over time, indicating that the product was unstable. Simultaneously, a similar approach was taken to obtain the *o*NB crosslinker (**81**) from 4,5-dimethoxy-2-nitrobenzaldehyde (**83**) although also unsuccessfully. The reaction conditions attempted are shown in Table 8.

Table 7 Reaction conditions attempted for the synthesis of the 7-methyl-α,α-(dibromoacetyl) coumarin (**80**)

Solvent	$(BrAc)_2O$ (eqv.)	T (°C)	catalyst	Time (h)	Notes	Ref.
CH_2Cl_2	3	RT	$Cu(OTf)_2$	16	Reaction does not proceed in solvents in the presence of a strong Lewis acid.	(Lata Chandra *et al.*, 2000)
MeCN	3	RT	$Cu(OTf)_2$	16		
Neat	1.5	85	$BrCH_2COBr$	16	Reaction may require more $(BrAc)_2O$ eqv.	
Neat	1.5	85	BrAcNa	16		
Neat	2.5	85	BrAcNa	16		
Neat	1.5	85	-	16	Longer reaction times and more $(BrAc)_2O$ eqv. lead to product formation. Product does not tolerate flash chromatography.	(Rahman and Jahng, 2007)
Neat	4	85	-	4		
Neat	10	85	-	24		
Neat	10	100	-	16	Starting materials decompose above 85 °C.	
Neat	10	80+140	-	0.5+0.5	Microwave irradiation. Product is visible by LC-MS but does not tolerate flash chromatography.	

Solvent	(BrAc)$_2$O (eqv.)	T (°C)	catalyst	Time (h)	Notes	Ref.
Neat	1.5	85	BrCH$_2$COBr	16	The reaction does not proceed with up to 3 eqv. of (BrAc)$_2$O – the starting material is even less soluble than the *o*NB derivative.	(Rahman and Jahng, 2007)
Neat	1.5	85	-	18		
Neat	3	85	-	16		
MeCN	3	RT	Cu(OTf)2	16	Reaction does not proceed in solvent with strong Lewis acid catalysis.	(Lata Chandra *et al.*, 2000)
Neat	10	80+120	-	1	Microwave irradiation. Product visible by LC-MS but starting materials mainly decomposed.	

The synthesis of the photocleavable derivatives required considerable optimisation. Their apparent instability to flash chromatography indicated an inherent instability of the scaffolds and, along with the instability of the non-cleavable compound in aqueous solutions at pH 7.0 and above, the di(acetyl) scaffold was considered not suitable for further development. Since ester functional groups are found in a number of biological small molecules (such as acetyl coenzyme A), it was proposed that the instability of this scaffold was due to the gem-di(acetyl) group rather than generally caused by the introduction of the (bromo)acetate bioconjugating groups.

A singly-bromoacetylated toluene derivative was synthesised to check for stability. Benzyl-1-methyl-1-bromoacetate (**83**) was easily synthesised from α-methyl benzyl alcohol (**84**) and bromoacetyl bromide in the presence of NaHCO$_3$ as a base. **83** was reacted with ethanethiol in MeCN with NEt$_3$ to give the thiol derivative.
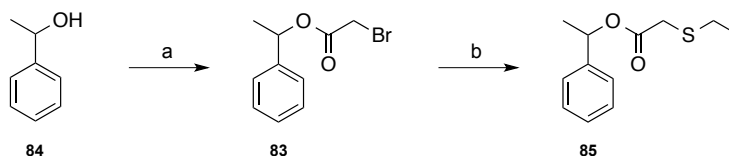
Fig. 4.18 Synthesis of benzyl 1-methyl-1-bromoacetate (**83**) and benzyl 1-methyl-1-(ethylthio)acetate (**85**). a) bromoacetyl bromide, NaHCO3 (3 eqv.), MeCN, 0 °C-RT, 2 h, 82%; b) EtSH (10 eqv.), NEt$_3$ (2 eqv.), MeCN, RT, 3 h, 94%.

Both compounds were subjected to stability studies similar to those carried out with the dibromoacetyl crosslinker (**74**). 5 mM solutions of each compound in 50% d$_6$-DMSO/buffer mixtures at varying pH (50 mM potassium phosphate pH 3.0-8.0, 100 mM NaCl) were prepared and ^{1}H NMR spectra collected with a water suppression pulse sequence at 1 h and 16 h intervals. The percentage hydrolysis was accounted for

using the ratio of the integral of the signal corresponding to the methyl group, as it was easily distinguishable from the remainder chemical shifts.

For the bromoacetyl compound (**83**), the signal shifts from 1.45 ppm to 1.31 ppm for the hydrolysis of the acetate group at the carbonyl carbon (**84**). A shift of the same signal to 1.46 ppm was suspected to correspond to hydrolysis of the α-bromide instead (**86**).
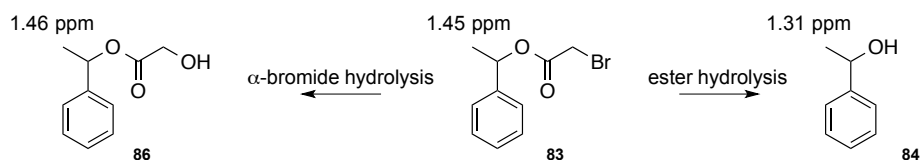
Fig. 4.19 Hydrolysis of benzyl 1-methyl-1-bromoacetate (**83**), showing the changes in chemical shifts of the methyl group, depending on the hydrolysed moiety. Hydrolysis at the ester carbonyl causes a shift from 1.45 ppm to 1.31 ppm (**84**) and at the α-bromide group a shift from 1.45 ppm to 1.46 ppm (**86**).

The single bromoacetate (**83**) compound is much more stable than the dibromoacetyl compound (**74**), where no hydrolysis of the acetate group is visible after 16 h in solution at pH up to 6.0 and 22-50% hydrolysis at pH 7.0-8.0, whereas the dibromoacetyl compound hydrolysis completely within 1 h at pH 7.0-8.0 and also undergoes rapid hydrolysis at lower pH (Fig. 4.20). The α-bromide group is also partially hydrolysed over 16 h (25-50% across pH 3.0-8.0). Since the bioconjugation reaction should occur much faster than the hydrolysis of the conjugating group (Hermanson, 2008a), this was not considered detrimental. The ethanethiol derivative (**85**) was found to be very stable, with no hydrolysis observable by ^{1}H NMR spectroscopy after 16 h in solution (Fig. 4.21). Therefore, the α-bromide group activates the ester moiety towards base-catalysed hydrolysis, which is no longer observed once the compound is derivatised to the α-thiother.

From these studies, it was concluded that the α-bromoacetyl bromide moieties were appropriate for bioconjugation to cysteine residues, with better reactivity profiles in DMSO/buffer mixtures compared to MeCN/buffer mixtures. Although the dibromoacetyl compound (**74**) was unstable, the single bromoacetate was considerably more stable and completely stable when derivatised to the thioether. A new scaffold was designed bearing α-bromoacetate groups as bioconjugation moieties, with a spacer group linker in-between, as described in the next section.

98
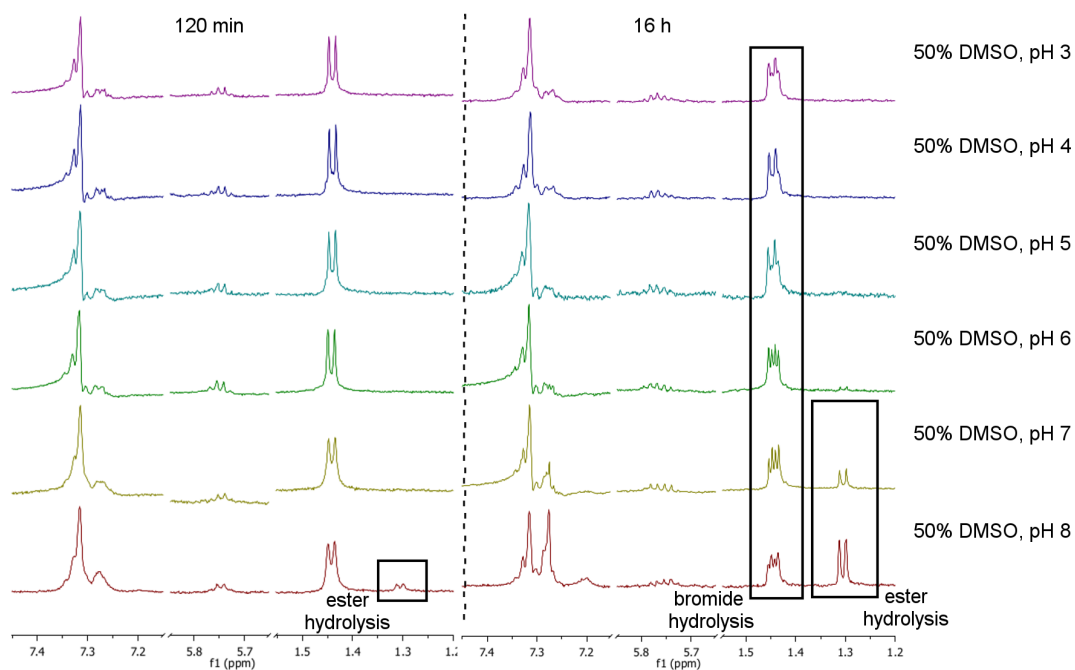
Fig. 4.20 ^{1}H NMR spectra of the stability studies of benzyl 1-methyl-1-bromoacetate (**83**) in 50% d$_6$-DMSO in buffer at different pH. From left to right: spectra at 120 min and 16 h. From top to bottom: pH 3.0, pH 4.0, pH 5.0, pH 6.0, pH 7.0 and pH 8.0.
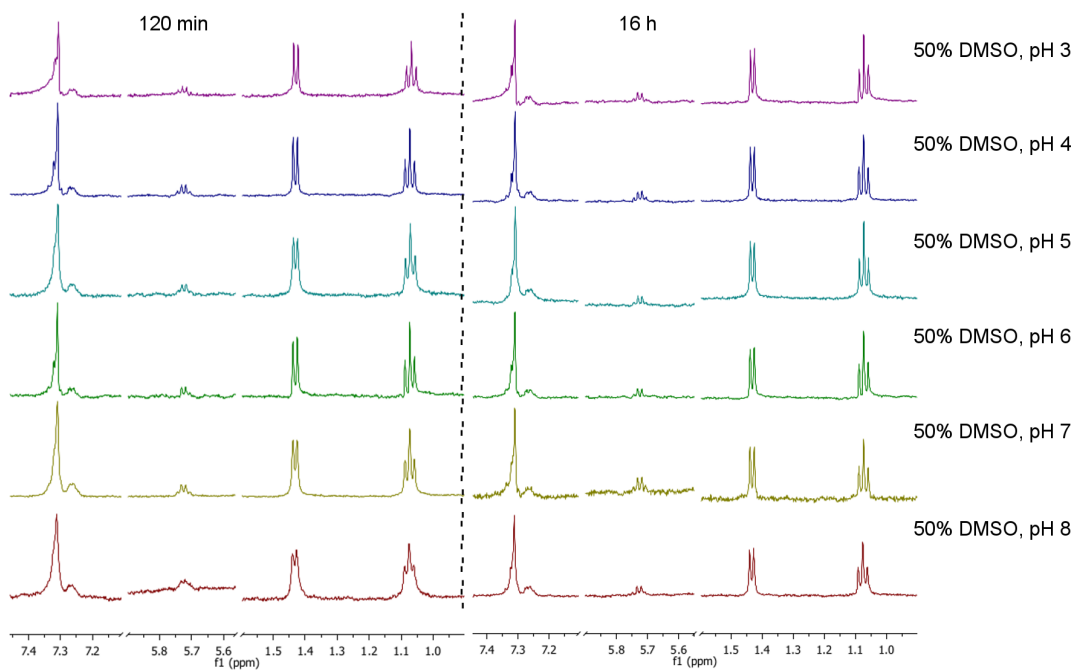
Fig. 4.21 ^{1}H NMR spectra of the stability studies of benzyl 1-methyl-1-theylthioacetate (**85**) in 50% d$_6$-DMSO in buffer at different pH. From left to right: spectra at 120 min and 16 h. From top to bottom: pH 3.0, pH 4.0, pH 5.0, pH 6.0, pH 7.0 and pH 8.0.

4.5 Third generation crosslinkers

4.5.1 Design and rationale

In order to obtain a cysteine-reactive and stable crosslinker, two α-bromo acetyl moieties had to be engineered at two distinct cleavage sites into the new scaffold design. Therefore, the new compound contained two photocages, connected by an extendable linker chain, as shown in Fig. 4.22. For simplicity of the synthesis, the linker chain was designed as an alkene moiety as two symmetrical photocaging groups could be joined through a metathesis step.
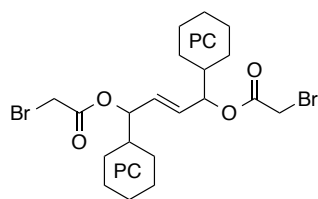
Fig. 4.22 General scaffold design for the α-bromo acetyl photocleavable crosslinkers. Two photocages are linked through an extendable alkene chain and each is flanked by an α-bromo acetyl moiety for attachment to cysteines and easy release as a carboxylate.

4.5.2 Synthesis of bis-phenyl non-cleavable crosslinker

As done for the previously designed scaffolds, a non-cleavable compound, using a phenyl ring in place of the photocaging group, was synthesised first to test for the synthetic route, stability of the compound in buffer and reactivity towards thiols. The synthetic route is shown in Fig. 4.23. The racemic 1-phenylprop-2-en-1-ol (**87**) was synthesised from benzaldehyde (**75**) via a Grignard reaction with vinylmagnesium bromide (Cornil *et al.*, 2014) and resolved kinetically using a non-specific lipase from *Candida antarctica B* (Štambaský *et al.*, 2008). The chirality and enantiomeric excess of the resolved S-alcohol (**(S)-87**) was confirmed by reaction with both R and S Mosher's acyl chlorides to give the diastereomeric (S,R)- and (S,S)-Mosher's esters (Fig. 4.24, **88** and **89**) (Mallavadhani and Rao, 1994). The ^{1}H NMR spectra of the two diastereomers obtained was analysed and compared to that of the parent secondary alcohol to give $\Delta\delta^{S\text{-}R}$ – differences in chemical shifts between the diastereomers, especially noticeable for the alkenyl protons – which were used to assign the relative spatial arrangement of the substituents, and, therefore, the chirality of the parent alcohol (**(S)-87**). Mosher's esters preferentially adopt conformations as shown in Fig. 4.24 and the benzene ring on the MTP moiety shields the closest group, causing a lower chemical shift. As is clearly visible in insert **a** from Fig. 4.25, the chemical shifts for the alkenyl

100

protons are shifted downfield for the S-ester (**89**) compared to the R-ester (**88**), which means that the alkene group is shielded in the R-ester (**88**), indicating that the alcohol is the S-enantiomer, as shown in Fig. 4.24.

Fig. 4.23 Synthesis of the alkenyl bis-toluene non-cleavable crosslinker (**94**). a) Vinylmagnesium bromide (1.2 eqv.), dry THF, 0 °C, 4 h, 81%; b) Novozyme 435 (26 mg/mmol), isopropenyl acetate (4.2 eqv.), dry PhMe, 4 Å molecular sieves, 40 °C, 16 h, 72%; c) KOH (1.05 eqv.), MeOH, 50 °C, 2 h, 90%; d) Pyridine (2 eqv.), bromoacetyl bromide (2 eqv.), dry CH₂Cl₂, 0 °C, 2 h, 83%; e) i) Hoveyda-Grubbs II catalyst (10 mol%), dry CH₂Cl₂, 45 °C, 40 h, ii) CNCH₂CO₂K (22 mol%), MeOH, RT, 1 h, 45%; f) Pyridine (2 eqv.), bromoacetyl bromide (2 eqv.), dry CH₂Cl₂, 0 °C, 2 h, 82%; N-acetyl cysteine (1.1 eqv.), NEt₃ (2.2 eqv.), MeCN, RT, 1 h, 86%.

Fig. 4.24 Pictorial representation of the method for the confirmation of the chirality of (S)-1-phenylprop-2-en-1-ol (**(S)-87**) by Mosher's ester derivatisation and ¹H NMR analysis.

101

protons are shifted downfield for the S-ester (**89**) compared to the R-ester (**88**), which means that the alkene group is shielded in the R-ester (**88**), indicating that the alcohol is the S-enantiomer, as shown in Fig. 4.24.

Fig. 4.23 Synthesis of the alkenyl bis-toluene non-cleavable crosslinker (**94**). a) Vinylmagnesium bromide (1.2 eqv.), dry THF, 0 °C, 4 h, 81%; b) Novozyme 435 (26 mg/mmol), isopropenyl acetate (4.2 eqv.), dry PhMe, 4 Å molecular sieves, 40 °C, 16 h, 72%; c) KOH (1.05 eqv.), MeOH, 50 °C, 2 h, 90%; d) Pyridine (2 eqv.), bromoacetyl bromide (2 eqv.), dry CH_2Cl_2, 0 °C, 2 h, 83%; e) i) Hoveyda-Grubbs II catalyst (10 mol%), dry CH_2Cl_2, 45 °C, 40 h, ii) $CNCH_2CO_2K$ (22 mol%), MeOH, RT, 1 h, 45%; f) Pyridine (2 eqv.), bromoacetyl bromide (2 eqv.), dry CH_2Cl_2, 0 °C, 2 h, 82%; N-acetyl cysteine (1.1 eqv.), NEt_3 (2.2 eqv.), MeCN, RT, 1 h, 86%.

Fig. 4.24 Pictorial representation of the method for the confirmation of the chirality of (S)-1-phenylprop-2-en-1-ol (**(S)-87**) by Mosher's ester derivatisation and ^{1}H NMR analysis.

Fig. 4.25 ^{1}H NMR spectra of Mosher's esters. The (S,S)-diastereomer (**89**) is shown in red (top) and the (S,R)-diastereomer (**88**) in blue (bottom). The main spectra show the compounds after flash chromatography. Insert **a** shows the overlaid spectra and **b** the (R,R)-diastereomer prior to flash chromatography, which was used to calculate the enantiomeric purity of the alcohol (62% *ee*).

The enantiomeric excess of the S-alcohol (**(S)-87**) was determined using the ^{1}H NMR spectrum of the crude (S,R)-diastereomer compound, prior to flash chromatography. Two signals for the alkenyl CH are clearly visible, corresponding to the two different (S,R) and (R,R) diastereomers. From the integrals a 62% *ee* was calculated.

Upon confirmation of the chirality of the (S)-alcohol (**(S)-87**), the (R)-acetate (**(R)-90**) was hydrolysed using KOH in MeOH with heating and the (R)-alcohol obtained (**(R)-87**) used to test the bromoacetylation step (Fig. 4.26). Initial reactions of the (R)-alcohol (**(R)-87**) with bromoacetyl bromide and sodium hydrogen carbonate in MeCN yielded an unexpected major product **91**, which was thought to arise from the rearrangement of the allyl bromoacetate. This problem was overcome by using pyridine in CH$_2$Cl$_2$ instead (Sawant and Jennings, 2006).[*] For further characterisation, the (R)-bromoacetate (**(R)-92**) was also reacted with N-acetyl cysteine in MeCN in the presence of triethylamine to give the cysteine derivative **93**. From the ^{1}H NMR spectrum only one diastereomer was visible, confirming that the (R)-alcohol (**(R)-87**), and, therefore, the (R)-α-bromoacetate (**(R)-92**), had high enantiomeric purity (>90% ee).
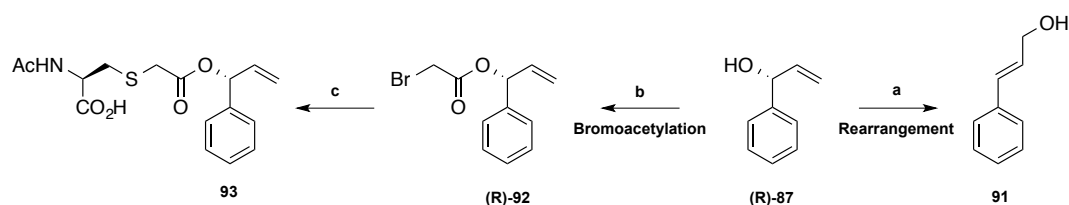
Fig. 4.26 Reaction of the (R)-allyl alcohol (**(R)-87**) with bromoacetyl bromide (2 eqv.) and the two products observed by ^{1}H NMR using different reaction conditions. a) NaHCO$_3$ in MeCN yielded the rearranged product **91.** b) Pyridine (2 eqv.) in CH$_2$Cl$_2$ yielded the desired bromoacetate (**R)-92** in 83% yield. c) reaction of allyl bromoacetate (**R)-92** with N-acetyl cysteine (1.1 eqv.), NEt$_3$ (2.2 eqv.) in MeCN, 86% yield.

The allyl-bromoacetate (**R)-92** was subjected to similar stability studies as described previously, and found to be stable across the same pH range (3.0-8.0) with no apparent hydrolysis over 20 h. The (S)-allyl alcohol (**(S)-87**) was then derivatised to the (S)-allyl bromoacetate (**(S)-92**) using the same conditions as for the synthesis of the (R)-allyl bromoacetate (**(R)-92**). The first metathesis experiment was carried out with the (R)-allyl alcohol (**(R)-87**), followed by the (S)-bromoacetate (**(S)-92**). A summary of the conditions attempted is shown in Table 9. Initial conditions for the metathesis of the allyl alcohol compound showed no conversion in the presence of 2nd Generation Grubbs catalyst (GII). Metathesis of the bromoacetate was attempted instead, as it was thought

---

[*] It was also important to note that yields improved considerably when using freshly purchased bromoacetyl bromide, which hydrolyses over time to bromoacetic acid and HBr.

that the ester group may allow for further coordination to the Ru centre. Some conversion was visible in the presence of GII. 1st Generation Grubbs catalyst (GI) did not promote any metathesis, with no product visible by TLC or LC-MS. 2nd generation Hoveyda-Grubbs catalyst gave the highest yield for the dibromoacetate product (**94**) and cleaner reactions with longer reaction times. HGII catalyst is known to be more efficient at coupling electron deficient alkenes (Chatterjee *et al.*, 2003). The catalyst was easily removed by treatment of the reaction mixture with potassium isocyanoacetate, which complexes with Ru metal and can be easily removed by flash chromatography, giving clean products as off-white solids.

Table 9 Metathesis reaction conditions attempted for the synthesis of the non-cleavable crosslinker

SM	[SM]* (mM)	Catalyst	[cat.] (mol%)	T (°C)	Time (h)	Yield (%)	Ru chelation	Notes
R-OH	200	GII	5	40	16	0	-	Alcohol does not undergo cross-metathesis but α,α-bromoacetate does. Flash chromatography not suitable for Ru removal.
S-AcBr	78	GII	5	40	16	35	-	
R-AcBr	78	GII	5	40	16	20	-	
R-AcBr	400	GII	10	40	16	40	-	Higher GII catalyst loading needed.
S-AcBr	400	GII	5	40	40	-	THMP[†]	
S-AcBr	400	GI	5	40	16	-	THMP	GI is not a suitable catalyst
S-AcBr	200	HGII	5+5	40	40	45	KICAc[‡]	HGII gives highest yield. KICAc chelation protocol is best for reaction clean-up.

^{1}H NMR stability studies of the (S,S)-di(bromoacetate) crosslinker (**94**) were attempted, but the compound was very insoluble in 50% d_6-DMSO in buffer at the concentrations necessary for clear signal detection using the water-suppression pulse sequence. Nevertheless, the compound was taken through to protein modification studies (see section 5.3). Concomitantly, the synthesis of cleavable crosslinkers carrying *o*NB or coumarinyl moieties was initiated, as described in the following sections.

---

* Reactions were run in dry and degassed CH_2Cl_2.

† Tris(hydroxymethyl)phosphine. 10×catalyst loading in NEt_3. Stirred at RT. (Maynard and Grubbs, 1999)

‡ Potassium 2-isocyanoacetate. 4×catalyst loading as a solution in MeOH. Stirred at RT. (Galan *et al.*, 2007)

4.5.3 Synthesis of *o*NB crosslinker

A route to the *o*NB crosslinker (**95**) based on the synthesis of the benzyl version was developed by C. Forde (MChem student with the Warriner group) and then repeated as part of this project to provide material for protein crosslinking studies. The *o*NB photocleavable crosslinker was synthesised as shown in Fig. 4.27.
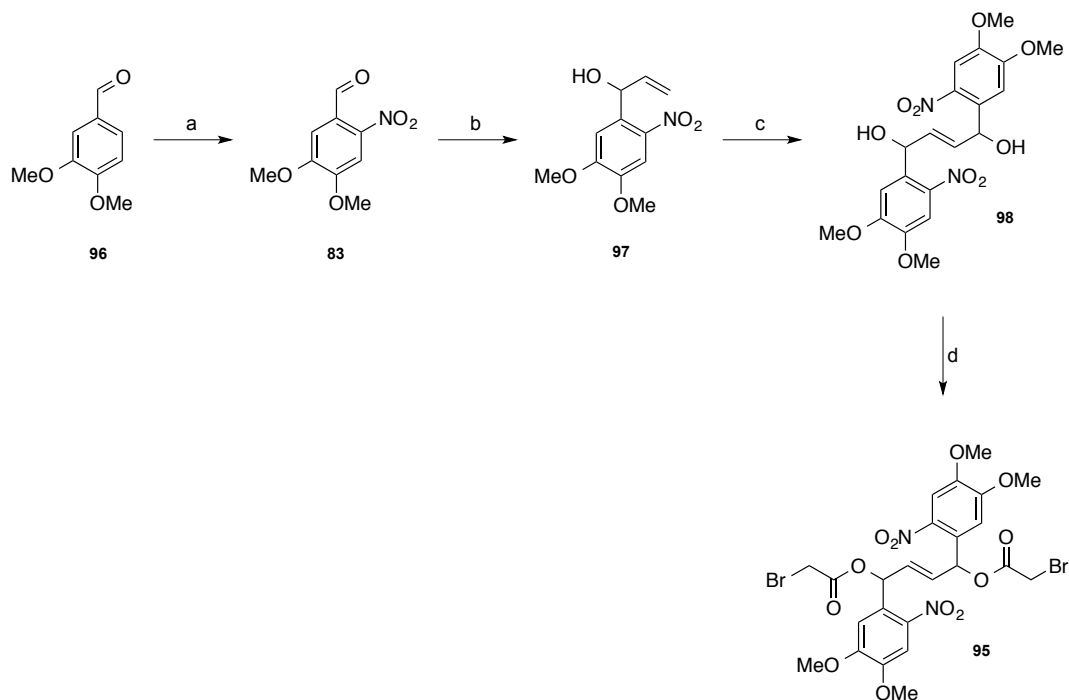
Fig. 4.27 Overall synthesis of the alkenyl bis-*o*NB photocleavable crosslinker (**95**). a) HNO_3 (15 eqv.), 0 °C, 3 h, 85% yield; b) Vinylmagnesium bromide (1.5 eqv.), dry THF, -78 °C-RT, 16 h, 70%; c) i) Hoveyda-Grubbs II catalyst (5 mol%), dry CH_2Cl_2, 45 °C, 16 h, ii) $CNCH_2CO_2K$ (25 mol%), MeOH, RT, 1 h, 93%; d) bromoacetyl bromide (2 eqv.), pyridine (2 eqv.), CH_2Cl_2, 0 °C, 1 h, 26%, 9:1 mixture of diastereomers.

The parent 4,5-dimethoxy-2-nitrobenzaldehyde (**83**) was easily obtained from the nitration of commercially available 3,4-dimethoxybenzaldehyde (**96**) in neat nitric acid. The aldehyde was reacted with vinylmagnesium bromide at low temperature to give the racemic allylic alcohol (**97**). C. Forde attempted a variety of strategies for the kinetic resolution of the racemic *o*NB allylic alcohol, including enzymatic resolution (Chen and Xiang, 2011) and alternative routes using Ir-catalysed reactions (Lyothier *et al.*, 2006, Roggen and Carreira, 2011). However, the kinetic resolution attempts were unsuccessful and the synthesis was continued with the racemic mixture. The 1,4-diol scaffold (**98**) was obtained in very good yield as a mixture of diastereomers from the metathesis of the racemic alcohol in the presence of 2nd generation Hoveyda-Grubbs catalyst. This compound was found to be extremely insoluble and was isolated by gravity filtration of the metathesis reaction mixture (after treatment with potassium

isocyanoacetate in methanol for the removal of Ru contaminants). Bromoacetylation to the final crosslinker (**95**) was carried out as previously described, in a lower yield than expected, due to contamination of some of the fractions from flash chromatography containing the desired product with an unidentified, coloured impurity. The identity of the final product was confirmed by both HRMS and NMR. The ^{1}H NMR spectrum showed a 9:1 mixture of the two diastereomers. As the two compounds show the same retention value by thin layer chromatography in hexane:ethyl acetate mixtures and no separation by LC-MS, the apparent separation of the diastereomers by flash chromatography was surprising.

Due to the very low solubility of the final compound, no stability studies of the crosslinker in solution were attempted, but the crosslinker was taken forward to test for protein crosslinking and cleavage as described in section 5.3.

4.5.4 Synthesis of bis-coumarinyl scaffold

The synthesis of the coumarinyl crosslinker (**99**) was attempted simultaneously with that of the *o*NB crosslinker (**95**). A similar synthetic route to that previously described was followed (Fig. 4.28).

The parent 4-methyl-7-methoxycoumarin (**100**) was first synthesised in large scale using a Pechman condensation of 3-methoxyphenol (**101**) and the ethylacetoacetate (**102**) in sulphuric acid. Oxidation to the 4-formyl-7-methoxycoumarin (**82**) was initially performed following a common protocol for the allylic oxidation of 4-methylcoumarins using selenium dioxide (Wang *et al.*, 2007, Huynh *et al.*, 2012), but the reactions were found to be difficult to scale-up, mainly due to the large quantities of toxic selenium contaminated waste which required bleaching due to stench. The product yields also varied, possibly due to the age of the oxidant.

A second route to the 4-formyl-7-methoxycoumarin (**82**) was attempted. 4-chloromethyl-7-methoxycoumarin (**103**) was synthesised similarly to the 4-methyl-7-methoxycoumarin (**100**), from 3-methoxyphenol (**101**) and ethyl 4-chloroacetoactetate (**104**). Direct hydrolysis of the 4-chloromethylcoumarin (**103**) to the 4-hydroxymethylcoumarin (**105**) in water was attempted, but the procedure found to be unreliable in large scale due to the large volume of water needed (very dilute reaction mixtures are necessary) and the formation of hydrochloric acid, which causes the reaction to stall and slow down, especially when working at larger scales. Neutralisation of the acid (up to pH 8) with sodium bicarbonate led to the acceleration of the reaction

but also the rapid formation of a by-product from the rearrangement of the coumarin under basic conditions. Therefore, the 4-chloromethylcoumarin (**103**) was instead substituted to the coumarinyl acetate (**106**), which was subsequently hydrolysed in EtOH/HCl to the 4-hydroxymethylcoumarin (**105**) followed by oxidation to the 4-formylcoumarin (**82**) in reasonable yield using pyridinium chlorochromate (PCC).

Although this route increased the number of steps to the final crosslinker and required the use of a chromium oxidant (toxic), the yields were consistent and the amount of toxic waste produced was reduced considerably. 4-Formyl-7-methoxycoumarin (**82**) was reacted with vinylmagnesium bromide in dry THF to give the allyl alcohol (**107**), which was taken through to the 1,4-methylcoumarin-1,4-diol (**108**) using the same conditions as described previously to obtain the 1,4-*o*NB-1,4-diol scaffold (Fig. 4.27, **98**). The metathesis product was found to be very insoluble, and was once again recovered by gravity filtration of the reaction mixture after treatment with potassium isocyanoacetate in MeOH for the removal of Ru contaminants.

The bromoacetylation of the 1,4-methylcoumarin-1,4-diol (**108**) to the final coumarinyl crosslinker (**99**) was attempted using the standard conditions employed for all other α-bromoacetyl compound synthesised previously, but without success, as the 1,4-methylcoumarin-1,4-diol (**108**) was even less soluble than the 1,4-(2-nitrobenzyl)-1,4-diol (**98**). Several conditions were attempted, including DMF/CH_2Cl_2 mixtures (Tang *et al.*, 2012), PTSA catalysis (Dhakane *et al.*, 2014) and neat bromoacetyl bromide, but in all cases heavy emulsions appeared during work-up and the crudes could not be purified by flash chromatography. HPLC was also attempted, but with only 3% yield.

Coumarinyl allyl alcohol **107** was converted to the corresponding bromoacetate (**109**) and cross-metathesis attempted using the same conditions as described previously for the benzyl bromoacetate (**94**), but no product was formed. Therefore, the coumarinyl crosslinker was abandoned.
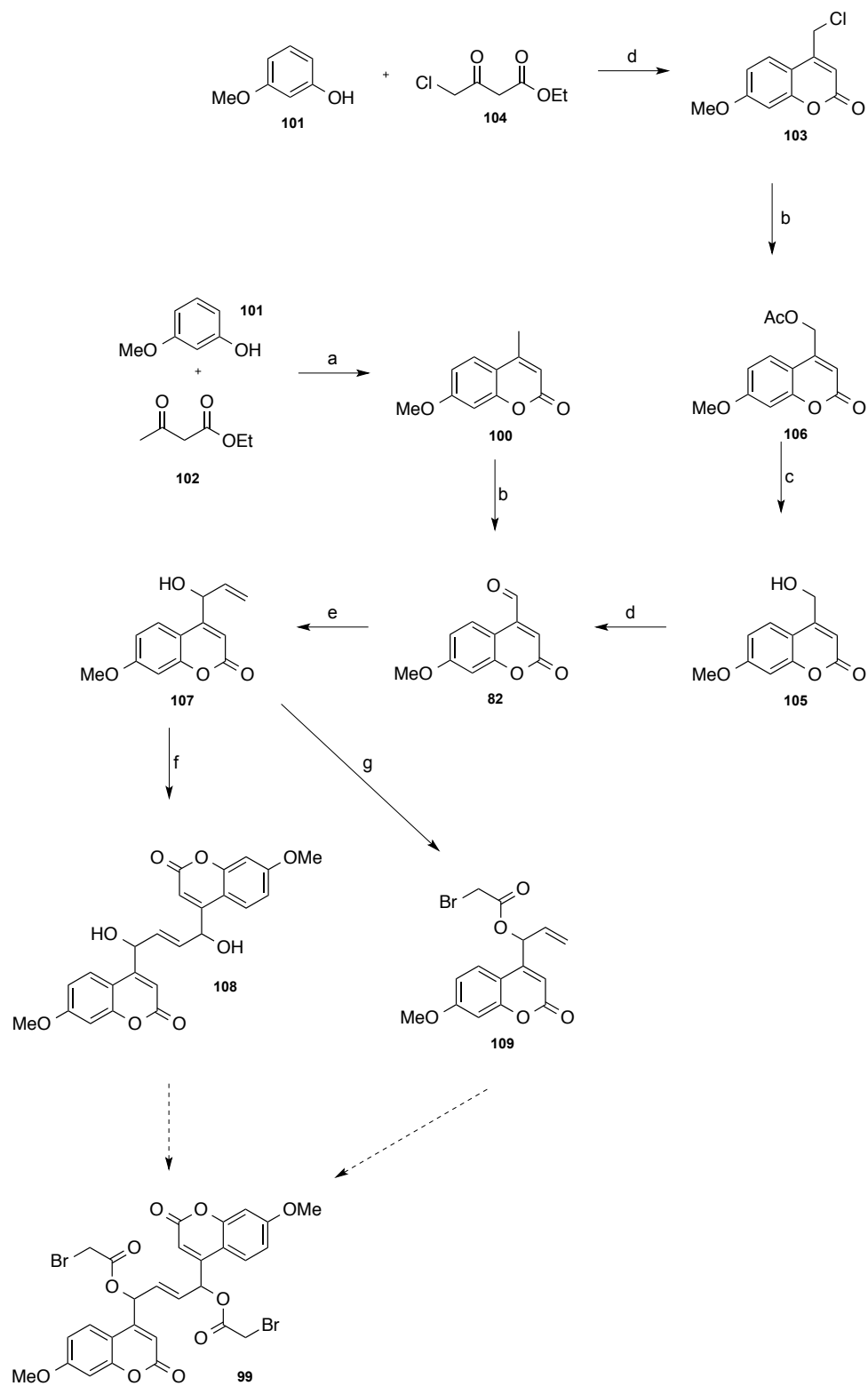
Fig. 4.28 Overall synthetic route to the alkenyl bis-coumarinyl photocleavable crosslinker scaffold **99**. a) H_2SO_4, 0 °C, 3 h, 60%; b) NaOAc (3 eqv.), AcOH, 120 °C, 20 h, 64%; c) 3:1 EtOH:HCl, 78 °C, 45 min, 98%; d) PCC (1 eqv.), $MgSO_4$, dry CH_2Cl_2, RT, 16 h, 49%; e) Vinylmagnesium bromide (1.2 eqv.), dry THF, 0 °C, 3 h, 34%; f) i) Hoveyda-Grubbs II catalyst (5 mol%), dry CH_2Cl_2, 45 °C, 16 h, ii) $CNCH_2CO_2K$ (25 mol%), MeOH, RT, 1 h, quantitative; g) bromoacetyl bromide (2 eqv.), pyridine (2 eqv.), CH_2Cl_2, 0 °C, 1 h, 38%.

4.6 Conclusion

During the course of this project, three different photocleavable crosslinker scaffolds were investigated. The first scaffold had been previously described by Omran *et al.* and was composed of two cleavable *o*NB moieties capable of releasing two cysteine residues upon UV irradiation (Omran and Specht, 2009a, Omran and Specht, 2009b). The crosslinker was designed to be bioconjugated by reaction of cysteine amino acids on the protein surface with secondary bromides at the benzylic positions of the *o*NB groups. One of the described compounds, bearing a propyl linker between the two chromophores (**53**), was synthesised according to the literature and tested for reactivity towards thiols in solution. The secondary bromides were found to be insufficiently reactive towards thiol containing molecules and, therefore, not useful for the clean modification of proteins by crosslinking.

To overcome the lack of reactivity, a novel crosslinker carrying two α-bromoacetate bioconjugating moieties to be cleaved by UV irradiation of one single chromophore was designed (Fig. 4.7). Both acetates would cleave from the benzylic position, making the overall molecule an acetal derivative. Although cyclic acetals have been previously described as suitable groups in linkers for protein affinity purification tags (Lee *et al.*, 2015), the α,α-dibromoacetate toluene crosslinker synthesised here (**74**, section 4.4) was found to be susceptible to hydrolysis in buffer, especially at high pH (8.0). This observation along with difficulties encountered in the synthesis and isolation of both the *o*NB and coumarinyl derivatives (**81** and **80** respectively) rendered these compounds unsuitable for further development. Nevertheless, the α-bromoacetates were tested for reactivity towards mercaptopropionic acid in buffer and found to be suitable moieties for bioconjugation. This knowledge was employed in the redesign of the molecules, yielding the third generation crosslinkers described.

A crosslinker carrying two benzyl groups linked by an alkene spacer and one α-bromoacetate moiety at the benzylic position for protein bioconjugation was designed and synthesised (Fig. 4.22, section 4.5.2). The compound was used to test the synthetic route and an enantiomerically enriched crosslinker was obtained (**94**). The equivalent *o*NB compound was also successfully synthesised (**95**), although as a mixture of diastereomers as the kinetic resolution strategies employed to resolve the first chiral intermediate in the synthetic route (**97**) were unsuccessful. Nevertheless, both compounds were taken through to protein modification studies as described in the next

section. Unfortunately, attempts to synthesise the faster-cleaving, coumarinyl crosslinker were unsuccessful, mainly due to the insoluble nature of the compound.

5 Protein crosslinking and photocleavage studies using the third generation crosslinking reagents

## 5.1 Protein crosslinking in biological investigations

Protein crosslinking has been used to investigate diverse aspects of protein activity. Its most common use is to target protein-protein interactions, where crosslinkers target both proteins and the efficiency of crosslinking is distance-dependent. Crosslinking between the proteins occurs only when the proteins are in a complex, even if just transiently, creating a covalent link that then allows for isolation and characterisation of the interacting partners (Trakselis *et al.*, 2005, Tang and Bruce, 2009). Information on both the identity of the binding partners and the binding interface can be obtained. The chemical crosslinkers can carry a variety of functional groups, which can target specific amino acid side chains, such as sulfhydryls, carboxylates or amines, and have variable lengths and even photo-reactive groups, where crosslinking only occurs in response to UV irradiation (Hermanson, 2008b, Hermanson, 2008c). Chemical crosslinking can also offer information about protein structure, by providing distance constraints between specific residues in the protein sequence, which allow for the modelling of the 3D shape of the protein, supporting protein fold prediction (Mouradov *et al.*, 2008, Leitner *et al.*, 2010).

Chemical crosslinking has also been used to probe protein dynamics. Membrane transporters are known to adopt very distinct conformations during function and X-ray crystallographic structures of such proteins in "inward-open" and "outward-open" states have been previously published (Forrest *et al.*, 2011, Waight *et al.*, 2013, Zhou *et al.*, 2014). Crosslinking of residues that are in close proximity in one conformation but distal in a different conformation has been shown to block or hinder transport activity for nucleoside transporters (Valdés *et al.*, 2012), serotonin transporters (Tao *et al.*, 2009) and glutamate transporters (Jiang *et al.*, 2011). All these studies have looked at activity of the transporters, rather than the dynamic or structural changes that occur as a result of crosslinking. Dutta *et al.* have used crosslinking to show that the blocking of the interconversion between the "inward-open" and "outward-open" conformations of an *E. coli* multidrug transporter, EmrE, leads to protein inactivation (Dutta *et al.*, 2014). By using a hetero-bifunctional crosslinker capable of reacting with a cysteine residue (S107C, introduced by site-directed mutagenesis) and a lysine residue 7.3 Å away, the transporter conformation was locked (determined by NMR), preventing the

interconversion between the inner and outward open conformations and inhibiting transport of ethidium (in-cell assay).

5.2 Aspartate α-decarboxylase as a test system

The final stage of this project was to test the crosslinking and photolysis properties of the novel crosslinkers described in section 4.5. The test system chosen for the study was aspartate α-decarboxylase (ADC, see section 3). The C-terminus of the β-chain of ADC, residues [22]Tyr-Glu[24], is highly flexible and can be seen in two distinct conformations in the crystal structure of the WT enzyme (PDB 1AW8) (Albert *et al.*, 1998). The flexibility of this region is a known characteristic of ADC and has been seen in several structures of the enzyme bound to different ligands, including glutamate, serine, cysteic acid and 1-methyl-L-aspartate (unpublished results). The loop flanks the active site of the enzyme, shielding the substrate from the bulk solvent. It was hypothesised that the mobility of the loop is related to substrate binding and product release and that during decarboxylation, Gly24 caps the active site to exclude water molecules.
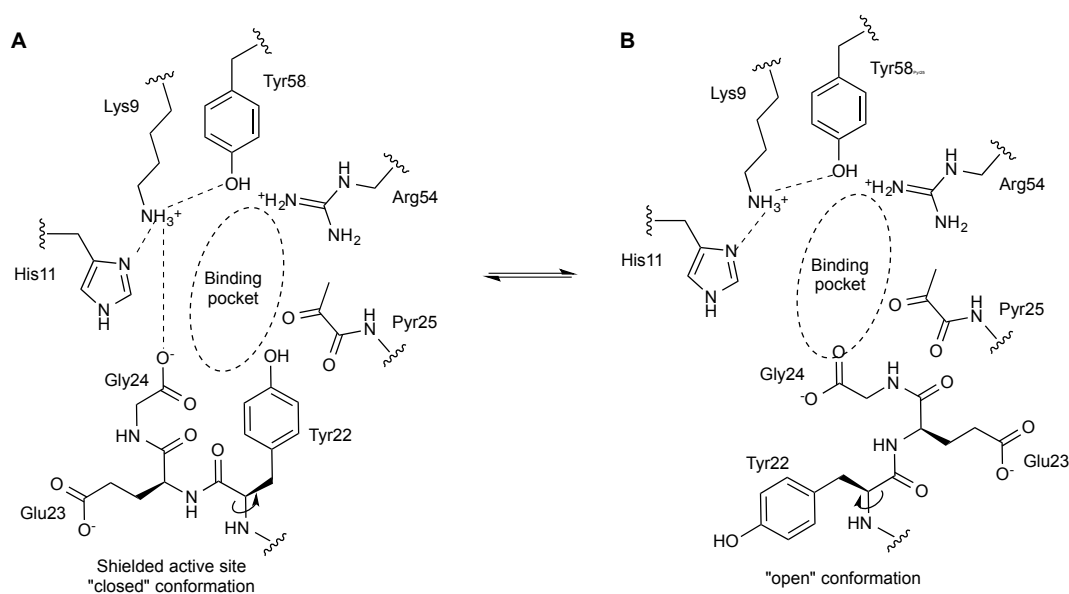
Fig. 5.1 Schematic representation of the two conformations of the ADC [22]Tyr-Gly[24] loop. (**A**) "closed" conformation. The residues close across the binding pocket opening, shielding it from water. Rotation around the Cα-N bond of Tyr22 gives the "open" conformation, shown in (**B**), where the binding pocket is exposed.

Locking of the loop into a non-active conformation has the potential to slow or block ADC turnover. Three initial pairs of sites for the introduction of cysteine residues for crosslinking were chosen. Each pair of cysteines should react with one crosslinker molecule, effectively stapling the loop into a fixed conformation. The three pairs also afforded different crosslinking lengths. Alongside the introduction of the cysteine

112

"anchors", two native cysteine residues (Cys26 and Cys78) were also mutated to avoid side reactions with the crosslinker.

5.2.1 Site-directed mutagenesis

This work was carried out in parallel with the structural studies. Three site-directed PanD mutants were generated by mutagenesis, each containing two residues mutated to cysteine and two cysteines removed:

- PanD H17C/E23C/C26S/C78A
- PanD D19C/E23C/C26S/C78A
- PanD E23C/C26S/A62C/C78A

Glu23 was chosen as the common cysteine "anchor" for all the PanD mutants as it adopts two very distinct conformations in the WT enzyme structure (PBD 1AW8). His17, Asp19 and Ala62 were then chosen to form the cysteine anchor pairs with Glu23. The side-chains of all three residues are located at the surface of the protein, rather than towards the core of the fold, and are also in close proximity to Glu23, spanning approximate crosslinking distances of 6 to 16 Å, depending on the cysteine pair and the conformation of the ^{21}His-Glu^{24} loop (Fig. 5.2).
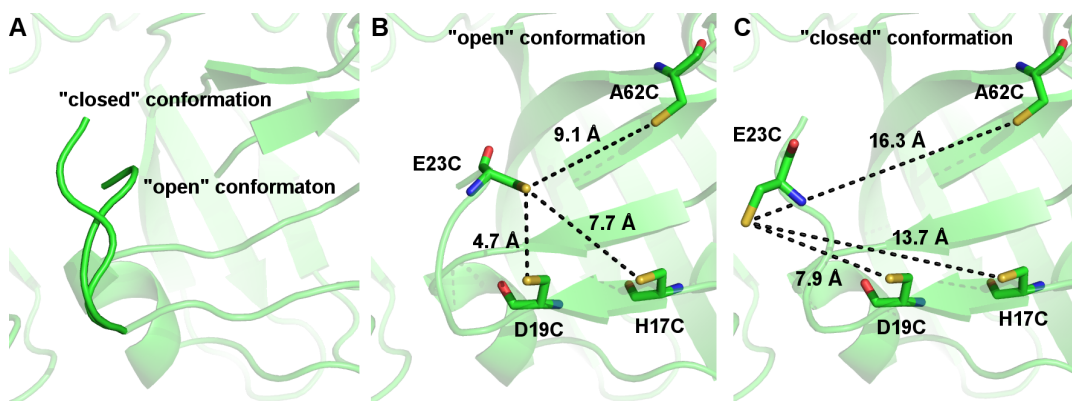
Fig. 5.2 The proposed sites for the introduction of cysteine anchors for crosslinking of ADC. ADC is shown as a cartoon in green. The mutations have been simulated in Pymol starting from PDB 1AW8 (Albert *et al.*, 1998); the figures do not represent a solved crystal structure of the mutant. (**A**) the two different conformations of loop ^{21}His-Glu^{24} an "open" conformation where the residues do not shield the active site, and a "closed" conformation, where the active site is shielded from solvent. The "closed" conformation is expected to be the active conformation, as it is adopted in ADC-ligand complexes (Yorke & Monteiro, unpublished results) (**B**) the approximate crosslinking distances between the E23C/H17C, E23C/D19C and E23C/A62C cysteine pairs in the open state. (**C**) the approximate crosslinking distances between the E23C/H17C, E23C/D19C and E23C/A62C cysteine pairs in the closed state.

The tetramutants were generated in parallel following a standard site-directed mutagenesis protocol (described in detail in section 8.4.3). The mutagenesis primers

were designed following standard primer design guidelines as described in section 8.4.1. Firstly, single PanD mutants (E23C, H17C, D19C, C26S, A62C and C78A) were generated from mutagenesis of the pRSETA-ADC-WT vector plasmid (Saldanha *et al.*, 2001). Following PCR, the mixtures were checked by agarose gel electrophoresis for the presence of large DNA fragments, which confirm a successful amplification of the plasmid, and transformed into *E. coli* XL10 cells. The plasmids were extracted from minicultured colonies using a miniprep protocol and sequenced to verify whether the desired mutation was successfully introduced. If the sequence was correct, the plasmid was then transformed into *E. coli* MG1655 Δ*panD* Δ*panZ* (DE3) cells ready for overexpression of the protein (see section 8.4).

Gel electrophoresis of the initial PCRs showed high molecular weight species for the PanD E23C, H17C and C78A experiments (Fig. 5.3). Although the PCR reactions to generate the PanD D19C and C26S mutants were unsuccessful in the first trial, they were successfully upon repetition of the reaction. The PanD A62C mutant was obtained by addition of 1% *v/v* DMSO to the PCR mixture, indicating that the primer may be prone to secondary structure formation in solution (Fig. 5.4, lane 8).
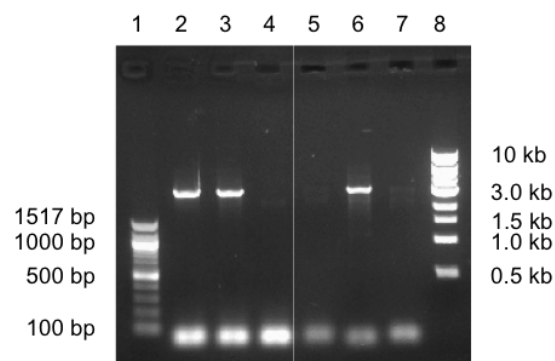
Fig. 5.3 Agarose electrophoresis gel of ADC site-directed mutagenesis PCR experiments starting from the pRSETA-ADC-WT vector plasmid. Lane 1 and lane 8: 100 base pairs (bp) and 1 kilobase pairs (kb) molecular weight markers. Lanes 2-7: PanD E23C, PanD H17C, PanD D19C, PanD C26S, PanD C78A, PanD A62C.

PanD H17C/E23C, D19C/E23C, E23C/A62C and C26S/C78A double mutants were generated from the initial PanD E23C, C78A, C26S and D19C mutant plasmids using the same protocol as for the single mutations. As the E23C point mutation is very close to the H17C and D19C mutations, the primers overlap both codons and thus had to carry both mutations. Once again, mutations for the introduction of the A62C were more successful in the presence of 1% *v/v* DMSO, as visible from the comparison of lanes 9 and 4 in Fig. 5.4 (with and without DMSO respectively).
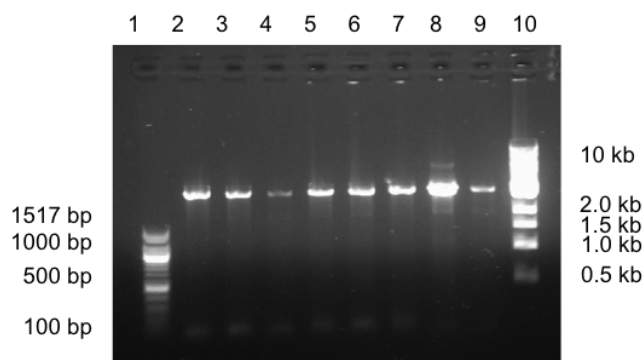
Fig. 5.4 Agarose electrophoresis gel of ADC site-directed mutagenesis PCR experiments. Lane 1 and lane 10: 100 bp and 1 kbp molecular weight markers; lanes 2-4: PanD E23C plasmid with D19C/E23C, H17C/E23C and A62C primers; lane 5: PanD C78A plasmid with C26S primers; lane 6: PanD C26S plasmid with C78A primers; lane 7: PanD D19C plasmid with E23C/D19C primers; lane 8: pRSETA-ADC-WT vector plasmid with A62C primers; lane 9: PanD E23C plasmid with A62C primer in the presence of 1% *v/v* DMSO.

The mutagenesis reactions were chosen to target all possible combinations of plasmids and primers. In certain cases, DNA sequencing showed errors and new colonies from the *E. coli* XL10 cell transformations had to be picked, minicultured, miniprepped and sequenced before the next round of mutagenesis. Therefore, the choice of parent plasmids for further mutations to the triple and quadruple mutants was dependent solely on how quickly the plasmid with the correct sequence was isolated. The triple and quadruple mutants were obtained using the same site-directed mutagenesis approach as for the double mutants and the exact sequence of mutagenesis steps is shown in Fig. 5.5.

Although gel electrophoresis is an important step to determine whether the DNA product of the PCR is of the correct size, it is not very sensitive to very low quantities of DNA. In order to streamline this multiple mutagenesis process, all the PCR products were subjected to digestion with Dpn1 and transformed into *E. coli* XL10 cells. In a few cases, colonies were obtained even when no clear plasmid DNA band was visible by electrophoresis.
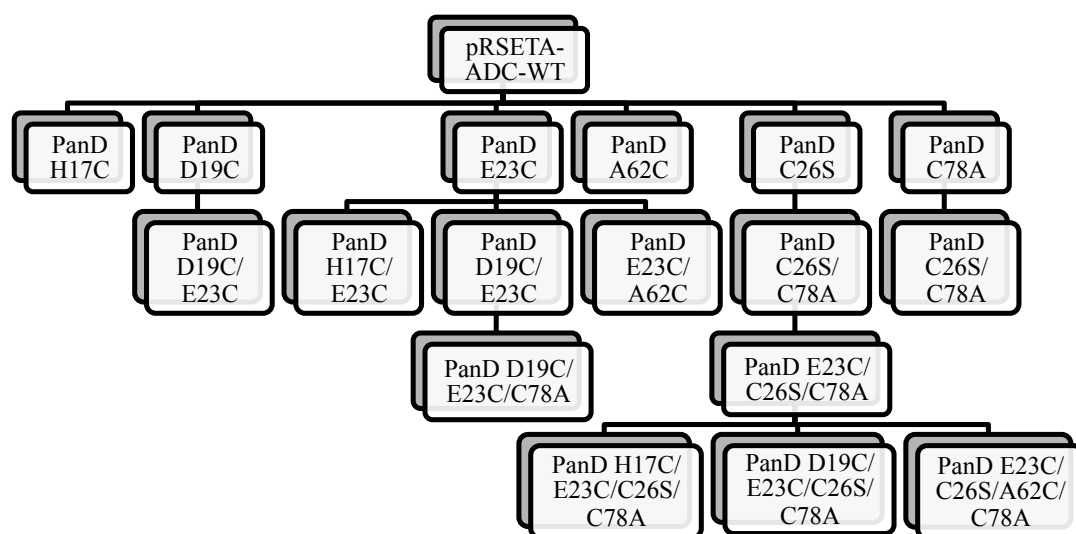
Fig. 5.5 Strategy for site-directed mutagenesis of the pRSETA-ADC-WT plasmid to generate the ADC tetramutants, showing the parent plasmids and the mutations generated at each PCR step.

5.2.2 Protein overexpression and purification

Initial overexpression of the three quadruple mutants PanD H17C/E23C/C26S/C78A, PanD D19C/E23C/C26S/C78A and PanD E23C/C26S/A62C/C78A was attempted using similar conditions to WT PanD described in section 9.1. *E. coli* MG1655 *ΔpanD ΔpanZ* (DE3) cells harbouring one of the plasmids coding for the mutants were minicultured overnight in LB media containing ampicillin. 1 mL of the miniculture was used to inoculate 1 L of auto-induction media and the cells grown for 20 h at 37 °C. The cells were collected by centrifugation and either frozen at -80 °C for storage or incubated for 24 h at 37 °C prior to freezing for activation of PanD to ADC.

The ADC tetramutants from pellets harvested from 1 L of thermally activated overexpression cell culture were purified first and expression levels and protein molecular weight were checked by SDS-PAGE (Fig. 5.6). All three mutant proteins were obtained predominantly as a single polypeptide chain of ~15 kDa, corresponding to the molecular weight of the intact zymogen protein chain, rather than the two smaller molecular weight chains (~11 kDa and 4 kDa) expected for the fully activated protein. This suggested that the mutations introduced to the protein may have prevented post-translational processing of the PanD zymogen to its active form, ADC.

116
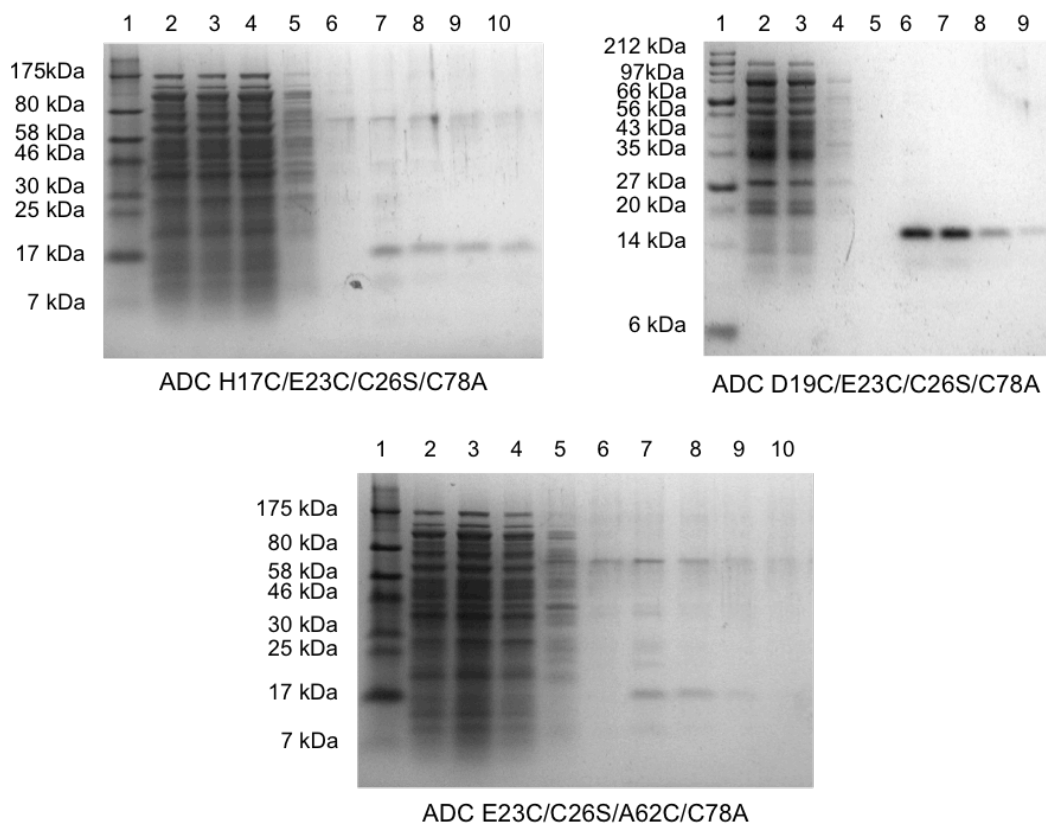
Fig. 5.6 Overexpression of ADC tetramutants and purification by Ni-NTA. Tris-tricine SDS-PAGE gels, showing a band at ~16 kDa in the elution fractions. ADC H17C/E23C/C26S/C78A: Lane 1: protein ladder; lanes 2-4: flow-through; lane 5: 50 mL wash fraction; lanes 6-10: 5×5 mL elution fractions. ADC D19C/E23C/C26S/C78A: Lane 1: protein ladder; lanes 2-3: flow-through; lane 4: 50 mL wash fraction; lanes 5-9: 5×5 mL elution fractions. ADC E23C/C26S/A62C/C78A: Lane 1: protein ladder; lanes 2-4: flow-through; lane 5: 50 mL wash fraction; lanes 6-10: 5×5 mL elution fractions.

The proteins were subjected to further purification by size-exclusion chromatography. Interestingly, the spectroscopy traces for absorption at $\lambda=280$ nm showed a different profile to that usually seen for the PanD(T57V) and WT ADC proteins (see Fig. 9.2 and Fig. 9.10). Both PanD(T57V) and WT ADC show a single peak with a minor shoulder which is thought to be an oligomer of the tetramer. For these mutants, the shoulder peak to the main protein peak carried a much higher percentage of the purified protein. This could indicate either a concentration dependent oligomerisation of the protein or an artefact that occurs due to the site-directed mutagenesis.
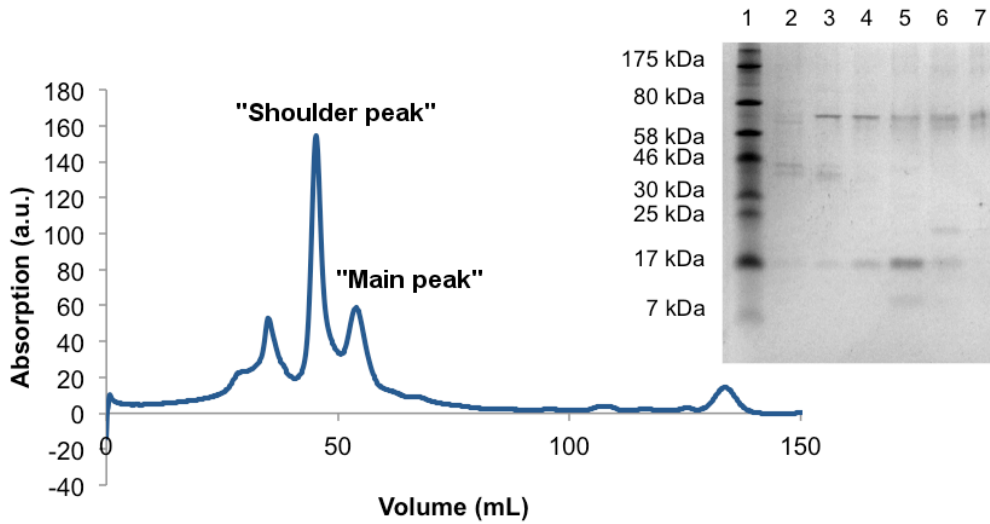
Fig. 5.7 ADC H17C/E23C/C26S/C78A SEC purification 280 nm absorbance trace, showing three protein peaks, and SDS-PAGE analysis of the elution fractions. Lane1: protein ladder; lanes 2-3: early peak (29-38 mL retention volume), unknown impurity; lanes 4-5: fractions from 39-48 mL retention volume, typically the PanD "shoulder peak" from a potential oligomer; lanes 6-7: fractions from 49-58 mL retention volume, typically the PanD "Main peak" from the PanD tetramer.
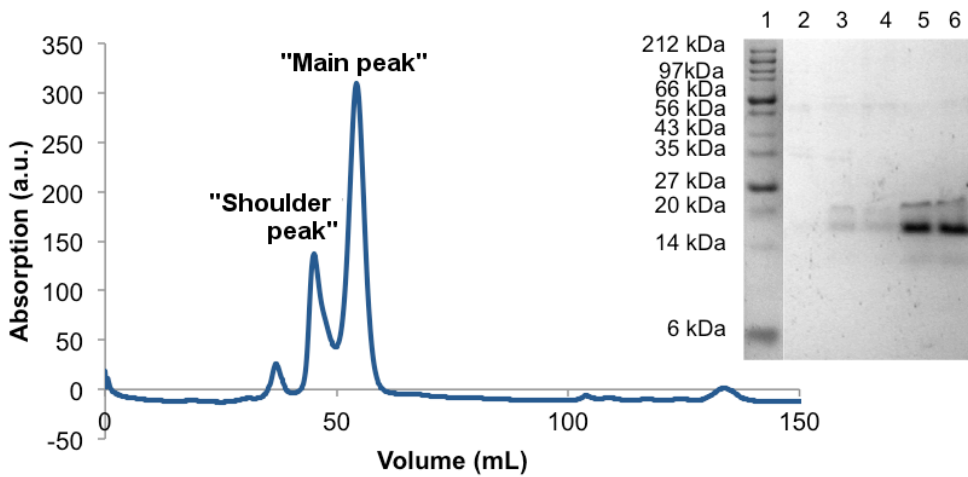
Fig. 5.8 ADC D19C/E23C/C26S/C78A SEC purification 280 nm absorbance trace, showing three protein peaks, and SDS-PAGE analysis of the elution fractions. Lane1: protein ladder; lanes 2-3: fractions from 39-49 mL retention volume, typically the PanD "shoulder peak" from a potential oligomer; lanes 4-6: fractions from 49-60 mL retention volume, typically the PanD "Main peak" from the PanD tetramer.

118
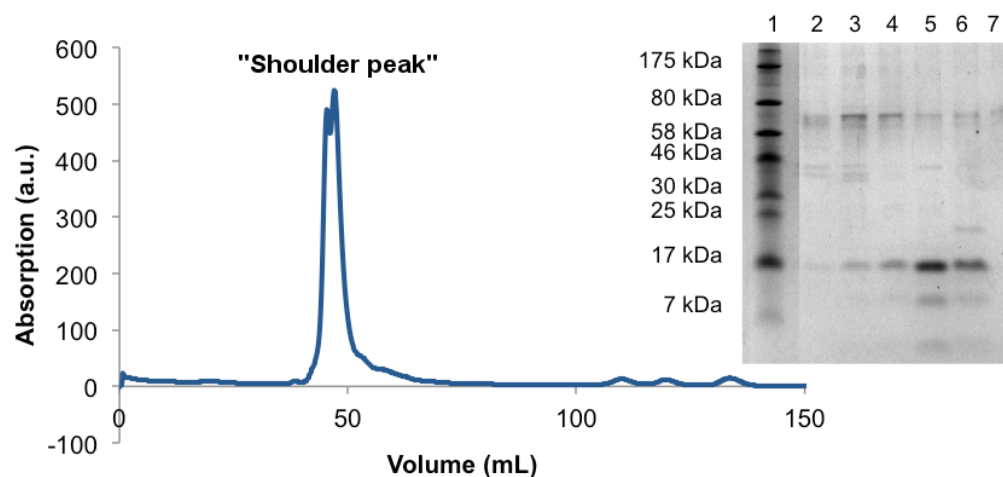
Fig. 5.9 ADC E23C/C26S/A62C/C78A SEC purification 280 nm absorbance trace, showing two protein peaks, and SDS-PAGE analysis of the elution fractions. Lane1: protein ladder; lanes 2-4: fractions from 39-45 mL retention volume; lanes 4-7: fractions from 45-50 mL retention volume. Both peaks correspond to the typical PanD "shoulder peak", with no protein visibly eluting at 50-60 mL retention volume, typically the PanD "main peak".

The identities of the purified mutants were confirmed by high-resolution mass-spectrometry (HRMS). As the Tris-tricine SDS-PAGE gels of the SEC fractions showed the same species across both the "main" and "shoulder" peaks for all the proteins, the fractions were combined and characterised by LC-HRMS system in 50 mM Tris pH 7.4, 100 mM NaCl and 0.5 mM TCEP. The LC column separates all the buffer components, allowing for clean HRMS spectra to be collected without prior manual buffer exchange. One further advantage of the LC system is that it also allows different protein species to be separated prior to MS, making it possible to obtain clear mass spectra for each of the possible ADC chains: the π-chain (~15 kDa, prior to post-translational cleavage) and the α and β-chains (~4 kDa and ~11 kDa, after post-translational cleavage).

Although Tris-tricine SDS-PAGE analysis shows very low levels of protein cleavage, the instrument is sufficiently sensitive to detect the different chains. All three chains for the ADC H17C/E23C/C26S/C78A mutant were identifiable. The π-chain and the β-chains each showed one single mass (Fig. 5.10, **A** and **B**). The resolution of the spectra was sufficiently high to show that both chains were in an oxidised form, suggesting the presence of a disulphide bond between the Cys17 and Cys23. The spectra for the α-chain was composed of two distinct species (Fig. 5.10, **C**), one corresponding to the post-translationally introduced pyruvoyl group and the other to the serine by-product that occurs from the hydrolysis of the ester intermediate during post-translational processing. Assuming similar ionisation strengths of the Ser25 and Pyr25-containing α-chains, the two species appear to be in a 1:1 ratio. This observation, along with the low

levels of cleavage seen on the tris-tricine SDS-PAGE of the purified protein indicates inefficient processing of the mutant, possibly associated with a change in conformation of the activation loop.

Similar results were obtained for the ADC D19C/E23C/C26S/C78A mutant. The spectra for the π and β-chains (Fig. 5.11 **A** and **B**) chains shows a 1:1 mix between the oxidised and reduced forms of the protein, indicating the possible formation of a disulphide bond between D19C and E23C in roughly 50% of the protein. The β-chain also shows approximately 50% contamination of the cleaved fraction with the Ser25 by-product, indicating inefficient post-translational processing of this mutant (Fig. 5.11, **C**).

The HRMS spectra for the ADC E23C/C26S/A62C/C78A mutant showed only the π and α-chains (Fig. 5.12, **A** and **B**). The β-chain was not visible from the LC-HRMS run, possibly due to very low protein concentration. Nevertheless, the two traces obtained confirmed that the correct mutant was generated and also that the protein was overexpressed in its reduced form, with no apparent disulphide between Cys23 and Cys62. From the expected positioning of the two newly introduced cysteine residues, this mutant should have the longest distance between the two cysteine anchors (Fig. 5.2), which explains why the disulphide is not formed. Contrary to the two previous tetramutants, the α-chain shows a single species corresponding to the hydrolysis at Ser25, with no indication of pyruvoyl formation from cleavage of the protein.

Although mass spectra for the α and β-chains for the mutants were obtained, according to the SDS-PAGE analysis the level of post-translational cleavage of the proteins was minimal. The large percentage of contamination by serine cleavage by-products also indicates that the mutants did not activate efficiently, especially the A62C mutant, which does not show any evidence of pyruvoyl formation. Although probably unsuitable for activity studies, the mutants were carried forward for further overexpression trials as the inactive constructs could still be used to test for crosslinking and photocleavage using the third generation crosslinkers described in section 4.5.
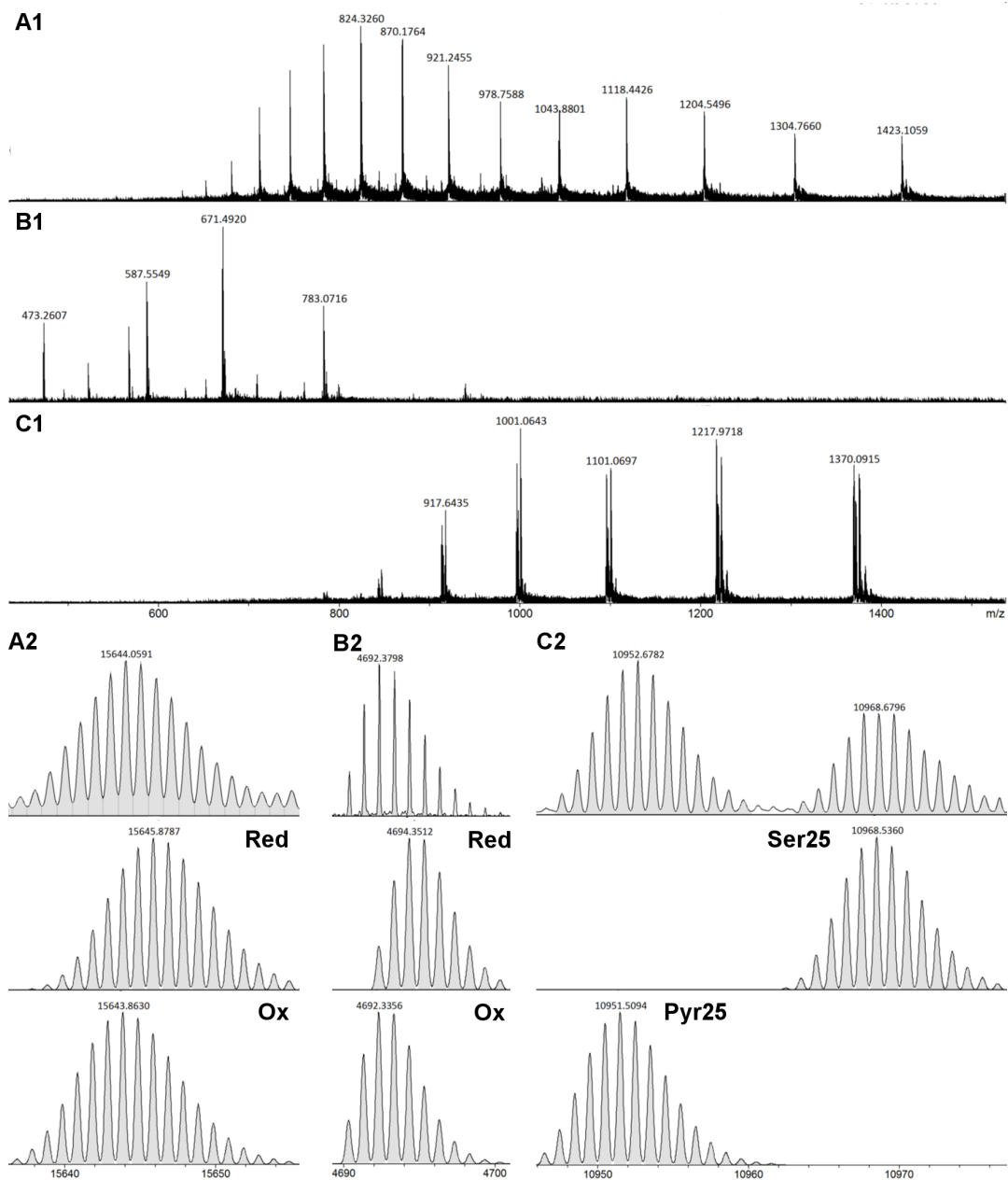
Fig. 5.10 HRMS of ADC H17C/E23C/C26S/C78A. The raw spectra for each protein chain are shown in **A1**, **B1** and **C1** and the corresponding deconvoluted reconstructions in **A2**, **B2** and **C2** along with simulated deconvoluted spectra underneath for comparison. (**A**) π-Chain mainly in its oxidised form (15.644 kDa). (**B**) β-Chain, mainly in its oxidised form (4.692 kDa). (**C**) α-Chain, showing both Pyr25 and Ser25 (10.953 and 10.969 kDa respectively).
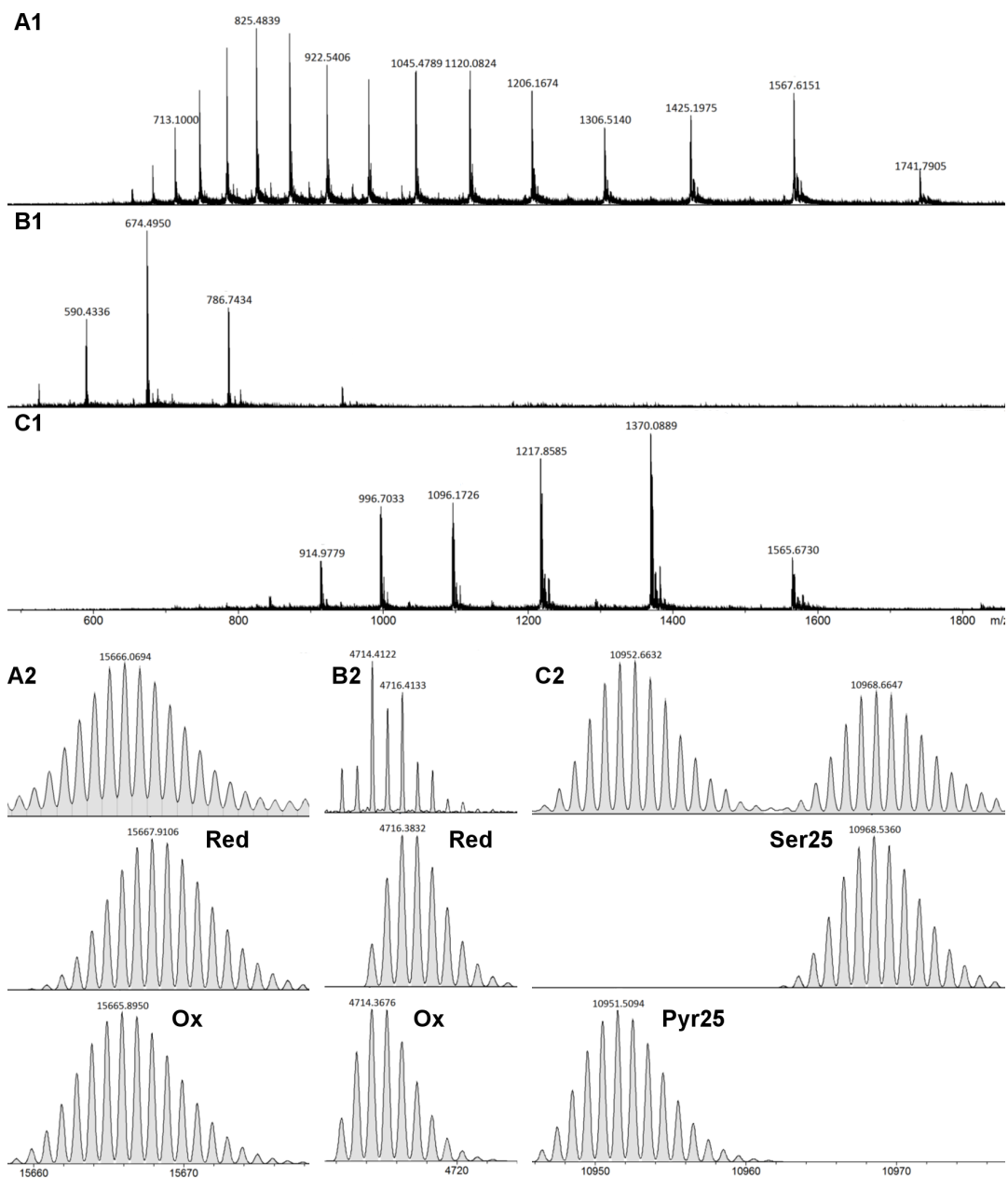
Fig. 5.11 HRMS of ADC D19C/E23C/C26S/C78A. The raw spectra for each chain are shown in **A1**, **B1** and **C1** and the corresponding deconvoluted reconstructions in **A2**, **B2** and **C2** with the simulated deconvoluted spectra underneath for comparison. (**A**) π-Chain mainly in its oxidised form (15.666 kDa). (**B**) β-Chain, 50% oxidised and 50% reduced forms (4.714 and 4.716 kDa respectively). (**C**) α-Chain, showing both Pyr25 and Ser25 (10.953 and 10.969 kDa respectively).
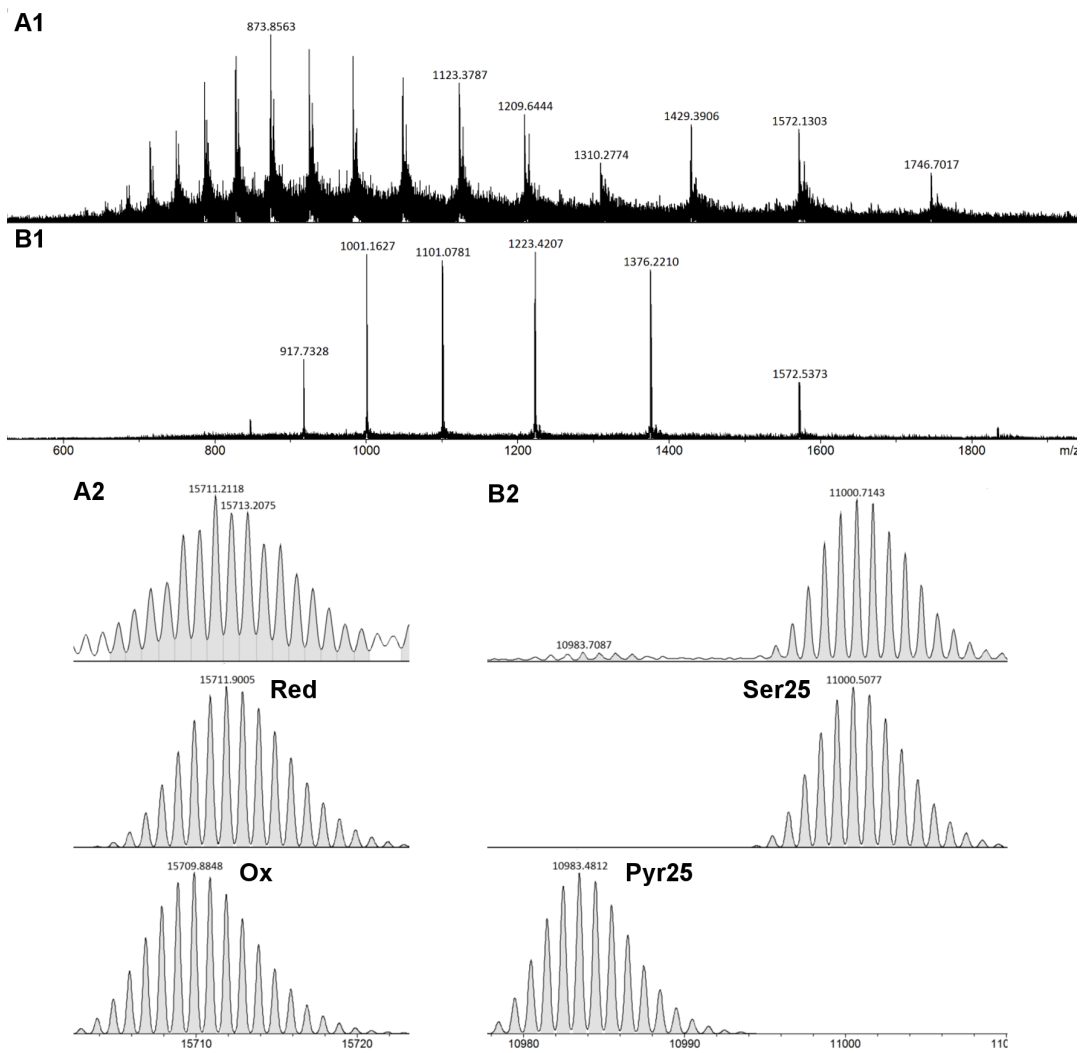
Fig. 5.12 HRMS of ADC A62C/E23C/C26S/C78A. The raw spectra for each chain are shown in **A1** and **B1** and the corresponding deconvoluted reconstructions in **A2** and **B2** with the simulated deconvoluted spectra underneath for comparison. (**A**) π-Chain mainly in its reduced form (15.711 kDa). (**B**) α-Chain, showing only Ser25 (11.001 kDa).

The yields for the ADC H17C/E23C/C26S/C78A and ADC E23C/C26S/A62C/C78A proteins were very poor under the initial overexpression conditions attempted (<0.5 mg L^{-1}), whereas the ADC D19C/E23C/C26S/C78A mutant expressed moderately better (0.9 mg L^{-1}). To investigate if the low protein yield for the two mutants was a consequence of poor overexpression or of protein degradation or unfolding during the activation of the pellet at 37 °C, both proteins were purified from non-heated *E. coli* cell pellets. The yield for the PanD H17C/E23C/C26S/C78A protein increased marginally (Fig. 5.13, left), indicating that heating the *E. coli* cell pellet may have caused aggregation or proteolysis of t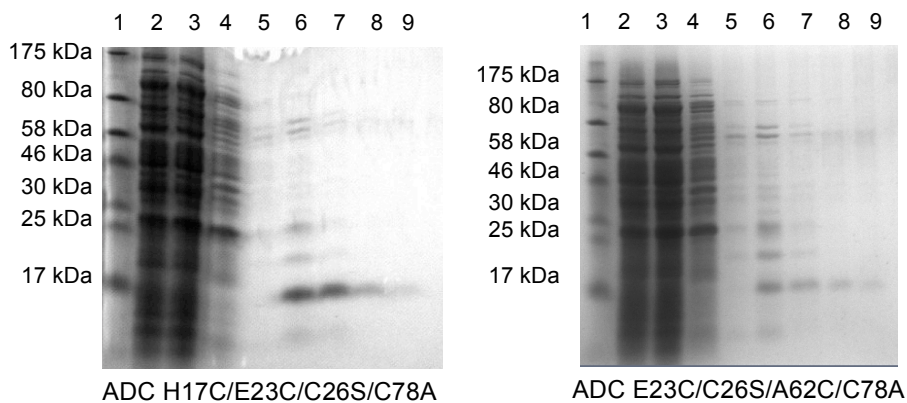he overexpressed protein. Unfortunately, no such improvement was visible for the PanD E23C/C26S/A62C/C78A mutant, which still expressed very poorly (Fig. 5.13, right).

ADC H17C/E23C/C26S/C78A

ADC E23C/C26S/A62C/C78A

Fig. 5.13 Overexpression of ADC tetramutants and purification by Ni-NTA. Tris-tricine SDS-PAGE gels, showing a band at ~16 kDa in the elution fractions. ADC H17C/E23C/C26S/C78A: Lane 1: protein ladder; lanes 2-3: flow-through; lane 4: 50 mL wash fraction; lanes 5-9: 5×5 mL elution fractions. ADC E23C/C26S/A62C/C78A: Lane 1: protein ladder; lanes 2-3: flow-through; lane 4: 50 mL wash fraction; lanes 5-9: 5×5 mL elution fractions.

Due to time constraints, only one more overexpression trial was attempted to try to increase protein yields. All three tetramutants were overexpressed in auto-induction media at 20 °C for 24 hours. The cells were incubated at a lower temperature than before to attempt to improve folded protein yields by avoiding possible aggregation and proteolysis of the misfolded proteins in the cells. The PanD H17C/E23C/C26S/C78A mutant expressed very well (10 mg L^{-1}) under these conditions and a small improvement in yield for the PanD E23C/C26S/A62C/C78A (2 mg L^{-1}) mutant was also visible. Interestingly, the yield for the PanD D19C/E23C/C26S/C78A protein was poorer than before.
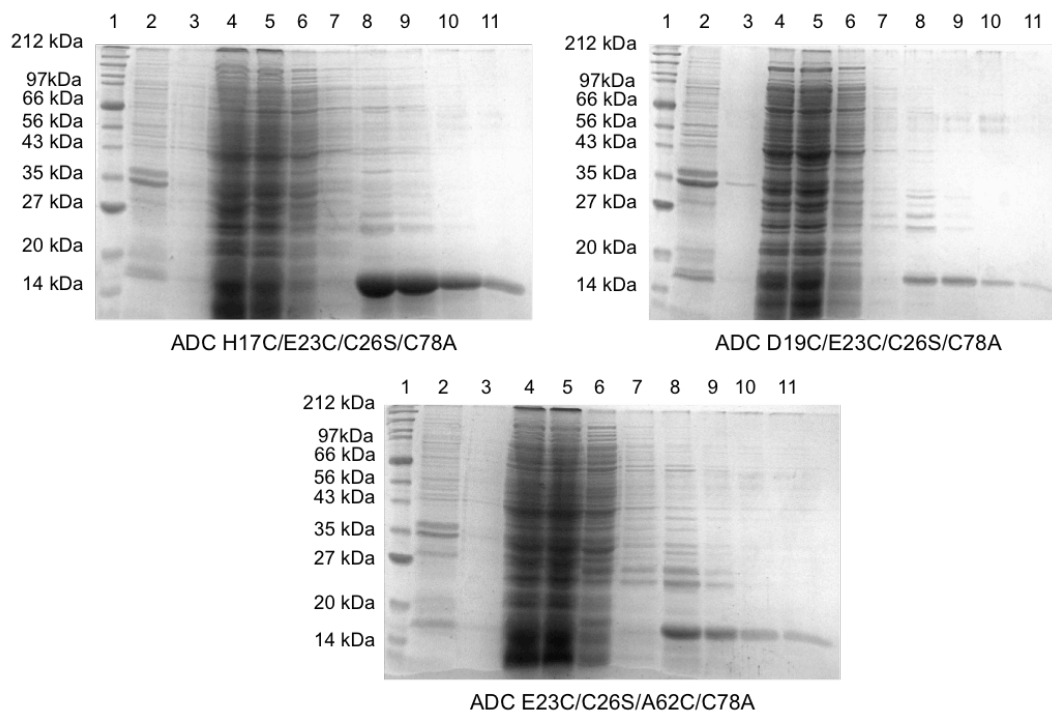
Fig. 5.14 Overexpression of ADC tetramutants and purification by Ni-NTA. Tris-glycine SDS-PAGE gels, showing a band at ~16 kDa in the elution fractions. ADC H17C/E23C/C26S/C78A: Lane 1: protein ladder; lanes 2: overexpression cell pellet; lane 3: overexpression cell pellet (1:100 dilution); lanes 4-6: flow-through; lane 7: 50 mL wash fraction; lanes 8-11: 4×5 mL elution fractions. ADC D19C/E23C/C26S/C78A: Lane 1: protein ladder; lanes 2: overexpression cell pellet; lane 3: overexpression cell pellet (1:100 dilution); lanes 4-6: flow-through; lane 7: 50 mL wash fraction; lanes 8-11: 4×5 mL elution fractions. ADC E23C/C26S/A62C/C78A: Lane 1: protein ladder; lanes 2: overexpression cell pellet; lane 3: overexpression cell pellet (1:100 dilution); lanes 4-6: flow-through; lane 7: 50 mL wash fraction; lanes 8-11: 4×5 mL elution fractions.

The conditions for the overexpression of all three tetramutant proteins differ. In two of the cases, lower temperatures (20 °C) seemed to result in higher yields, possibly due to better protein folding within the cells. In the case of ADC D19C/E23C/C26S/C78A the protein overexpressed in much higher yields at higher temperatures (37 °C) and, therefore, overexpression conditions have to be evaluated on a case-by-case basis. Even though none of the tetramutants express as well as WT ADC or any of the single ADC mutants described previously in this thesis, overexpression under these conditions yields sufficient protein for the testing of the crosslinking compounds.

Mutations to the mobile loop also have detrimental effects on the protein post-translational modification. Contrary to WT PanD, heating of the overexpression pellets led to very low levels of cleavage and high amounts of contamination of the cleavage products with the serine hydrolysis by-product instead of the desired catalytic pyruvoyl.

5.2.3 I*n vitro* activation of ADC H17C/E23C/C26S/C78A with PanZ

Although enzyme activation did not occur cleanly by heating the overexpression *E. coli* cell pellets, the activation of ADC H17C/E23C/C26S/C78A was attempted *in vitro* in the presence of PanZ after protein purification. The protein was incubated at 37 °C for 72 h and the extent of cleavage monitored at 16 h, 48 h and 72 h (Fig. 5.15). ADC H17C/E23C/C26S/C78A was incubated in the presence of PanZ and CoA as well as TCEP, to check if disulphide formation between Cys17 and Cys23 had any effect in protein activation. Protein cleavage was only visible in samples containing both the ADC mutant protein and 1.1 eqv. of PanZ but, even at prolonged incubation times, only a fraction of the protein underwent cleavage. Higher amounts of cleavage were observed in samples containing TCEP, indicating that the constraining of the loop by the disulphide bond hinders activation. Addition of AcCoA had no effect, probably due to the partial loading of PanZ with CoA from overexpression. The low amount of protein cleavage obtained from these studies indicates that this PanD mutant cannot be efficiently matured post-translationally. This is not optimal for the activity studies required to evaluate the extent of inhibition caused by crosslinking of the protein. The protein could potentially be fully activated by incubation with more equivalents of PanZ.AcCoA, but this could cause problems during purification as the proteins may not dissociate fully and contamination may occur. One other option would be to attempt the activation by incubation of the purified protein at 60 °C overnight (Ramjee *et al.*, 1997), but as thermal activation of the purified WT protein leads to contamination with hydrolysis by-products and the HRMS of the β-chain of the H17C/E23C/C26S/C78A tetramutant already shows close to ~50% contamination (Fig. 5.10), this method may yield low levels of catalytically competent enzyme. Both strategies would require both optimisation and accurate quantification of enzyme activation. Unfortunately, due to time constraints, this activation study has not yet been carried out. Nevertheless, the PanD(H17C/E23C/C26S/C78A) mutant could still be used to investigate the protein modification ability of the crosslinkers described in section 4.5 and was used to show that the crosslinkers could efficiently staple and release proteins as described in the following sections.

12.5 µM ADC	+	+	-	+	+	-	+	+	-	+	+	-
13.8 µM PanZ	+	-	+	+	-	+	+	-	+	+	-	+
17.5 µM AcCoA	+	+	+	-	-	-	+	+	+	-	-	-
5 mM TCEP	+	+	+	+	+	+	-	-	-	-	-	-
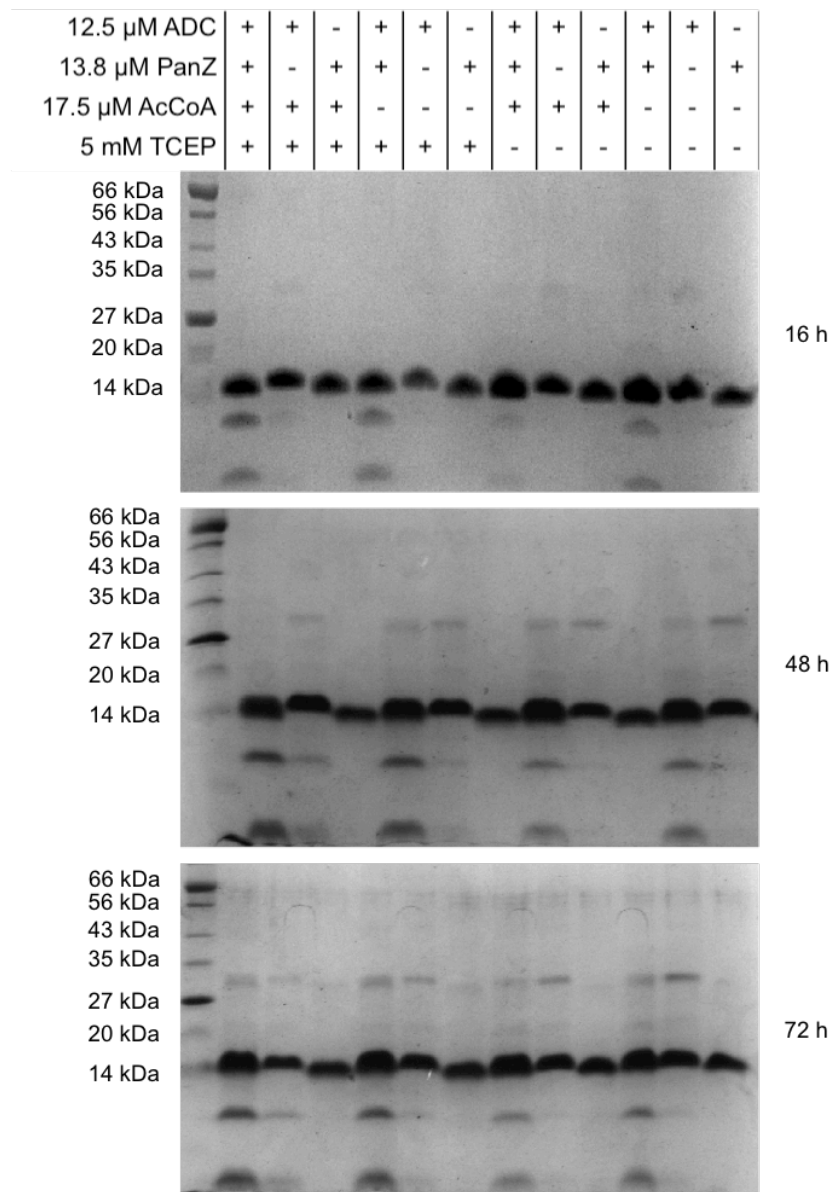
**Fig. 5.15** The cleavage of ADC H17C/E23C/C26S/C78A by PanZ.AcCoA. 12.5 µM ADC in 50 mM potassium phosphate buffer (pH 7.4, 100 mM NaCl, 0.5 mM TCEP) was incubated at 37 °C for 72 in the presence of PanZ, AcCoA and TCEP (as described for each lane) and samples were analysed at 16 h, 48 h and 72 h incubation periods. Partial cleavage is visible in the presence of PanZ, with increasing levels up to 48 h.

127

5.3 Protein modification studies

5.3.1 Protein crosslinking studies

Initial protein modification experiments were performed on ADC H17C/E23C/C26S/C78A using (S)-1-phenylprop-2-enyl bromoacetate (**(S)-92**) to test for the ease of protein modification using allyl bromoacetates. The protein was first concentrated to ~1.1 mg/mL in phosphate buffer (50 mM pH 8.0, 100 mM NaCl, 0.5 mM TCEP). As the HRMS of the sample showed that the protein was 2 Da lighter than expected (Fig. 5.10), indicating a possible disulphide between the two surface cysteines (Cys17-Cys23), TCEP was added to a final concentration of 5.5 mM and the protein incubated at 37 °C for 30 min to fully reduce the disulphide. A solution of the bromoacetate **(S)-92** in DMF was added to the protein sample to give a final concentration of 1 mM bromoacetate (~20 eqv.) in 10% *v/v* DMF in buffer. The mixture was incubated at 37 °C for 1 h. The HRMS spectrum (Fig. 5.16, **A**) showed one single protein species present in solution, with increased molecular weight corresponding to the labelling of the protein with two allyl benzyl ester groups (15.994 kDa compared to 15.646 kDa for the reduced protein). This indicated complete and clean labelling of both cysteines with the allyl bromoacetate, proving that allyl bromoacetate groups are efficient labelling agents under these conditions.

Following the successful labelling of ADC H17C/E23C/C26S/C78A with allyl bromoacetate **(S)-92**, protein crosslinking using the non-cleavable crosslinker (**94**) was attempted next following the same experimental procedure. The protein was first incubated with 5 mM TCEP for 30 min to reduce the disulphide bond and the crosslinker was added as a solution in DMF to give a final concentration of 0.5 mM crosslinker (~10 eqv.) in 10% *v/v* DMF in buffer. Once again, analysis by HRMS (Fig. 5.16, **B**) showed a single protein species in solution, with a molecular weight corresponding to that of the crosslinked protein (15.966 kDa). The spectrum showed neither evidence of labelled but not crosslinked protein nor of double protein labelling, proving that the bioconjugation reaction with the di-bromoacetate crosslinker (**50**) was clean and comparable to that with the single benzyl-bromoacetate species (**(S)-47**). Finally, the cleavable *o*NB crosslinker (**95**) was tested for its crosslinking ability under the same conditions as before. Reaction of the reduced protein with 0.5 mM *o*NB crosslinker (~10 eqv.) for 1 h at 37 °C also led to clean and complete protein crosslinking, as shown by the single protein species visible by HRMS in Fig. 5.16 (**C**) (molecular weight of 16.176 kDa).
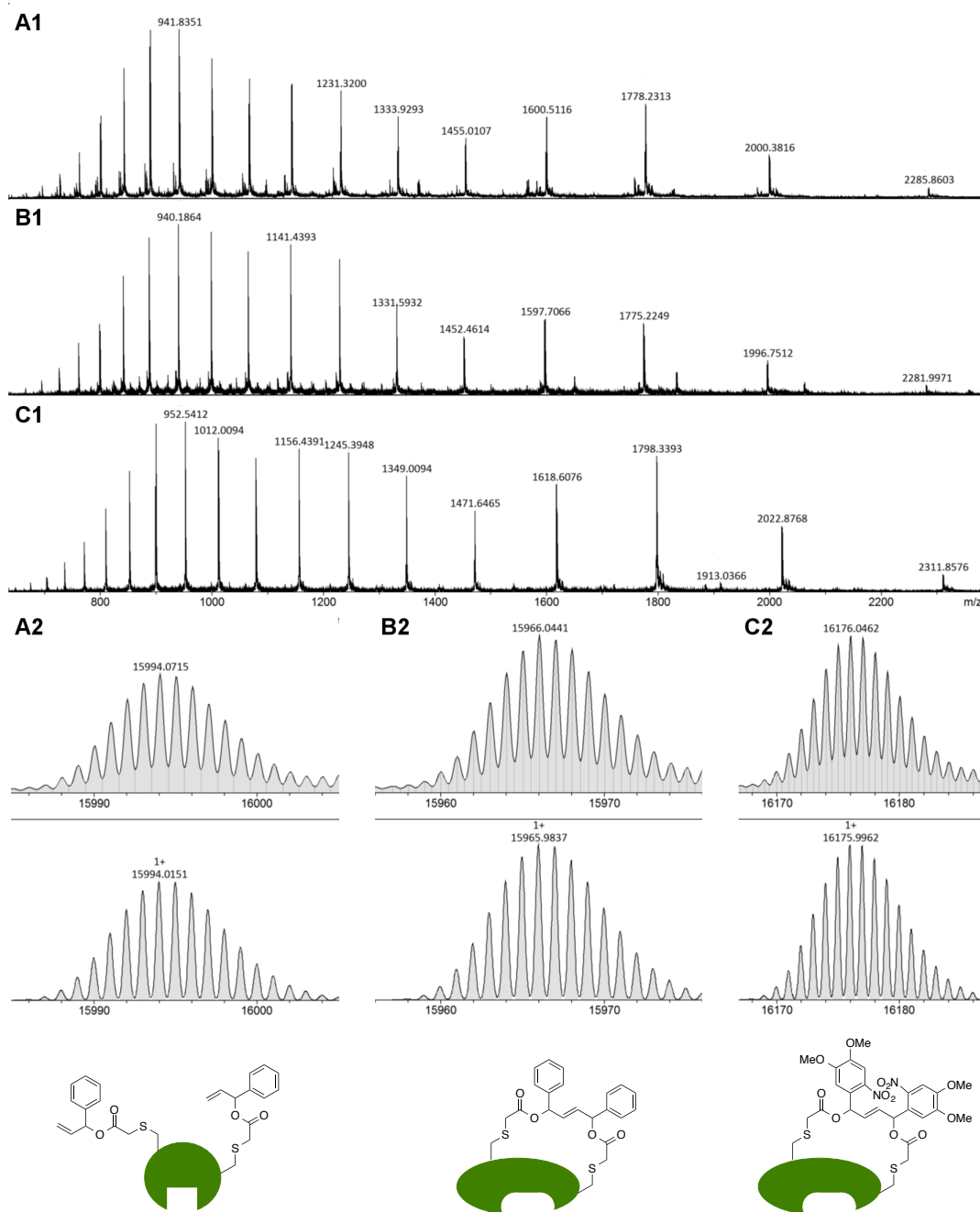
Fig. 5.16 ADC H17C/E23C/C26S/C78A protein modification results. The raw LC-HRMS spectra for each modified protein species are shown in **A1**, **B1** and **C1** and the corresponding deconvoluted reconstructions in **A2**, **B2** and **C2** along with the simulated deconvoluted reconstructions underneath for comparison. A pictorial representation of the different protein modifications is also provided under each reconstruction. (**A**) modification with 20 eqv. of the allyl benzyl bromoacetate compound (**(S)-92**). (**B**) modification with 10 eqv. of the benzyl non-cleavable crosslinker (**94**). (**C**) modification with 10 eqv. of the *o*NB cleavable crosslinker (**95**). The reactions were carried out by incubation of the protein in potassium phosphate buffer (50 mM pH 7.4, 100 mM NaCl, 5 mM TCEP) at 37 °C for 1 h.

5.3.2   Cleavable *o*NB-crosslinker studies

Freshly *o*NB-crosslinked ADC H17C/E23C/C26S/C78A was first buffer exchanged from the 10% DMF/buffer mixture into phosphate buffer (50 mM pH 7.4, 100 mM NaCl, 0.5 mM TCEP) to remove the excess crosslinker and DMF. The sample was placed in a short path quartz cuvette (5 mm) and exposed to UV light using a mercury lamp (UVP PenRay® 11SC-1, 1 mW, λ=254 nm). Cleavage of the crosslinker was monitored by HRMS at time intervals corresponding to 1, 6, 16, 31 and 46 min total irradiation time. The HRMS spectra showed the intact protein (1 min irradiation) cleaving cleanly to the expected methyl carboxylate doubly labelled protein. Assuming similar ionisation strengths of the different protein species, the relative intensities of the deconvoluted MS peaks were used to estimate the ratio of crosslinked to cleaved protein at each time point. 50% cleavage occurred in roughly 10 min irradiation time. The rate of cleavage appears to slow down over time, only reaching completion after 46 min of irradiation, probably due to the formation of strongly absorbing *o*NB ketone by-products which can act as UV filters (Klán *et al.*, 2013).

If the two photocleavable groups were independent of each other, the slow rate of photolysis should allow for the observation of an intermediate species corresponding to the cleavage of only one side of the staple, with formation of the corresponding nitroso species, and, therefore, without full release of the reagent from the protein. Since no cleavage intermediates were visible, once the first *o*NB group is cleaved off the protein, the crosslinker may fragment, releasing the second methyl carboxylate group in a "dark" reaction without requiring the second *o*NB group to undergo photolysis to release the protein. Of course, this photocleavage-fragmentation coupled mechanism may be promoted by the low intensity of the UV lamp used and a different mechanism may occur at higher light intensities, where a double photocleavage event may dominate instead, promoted by the higher number of photons available to be absorbed by the *o*NB moieties.
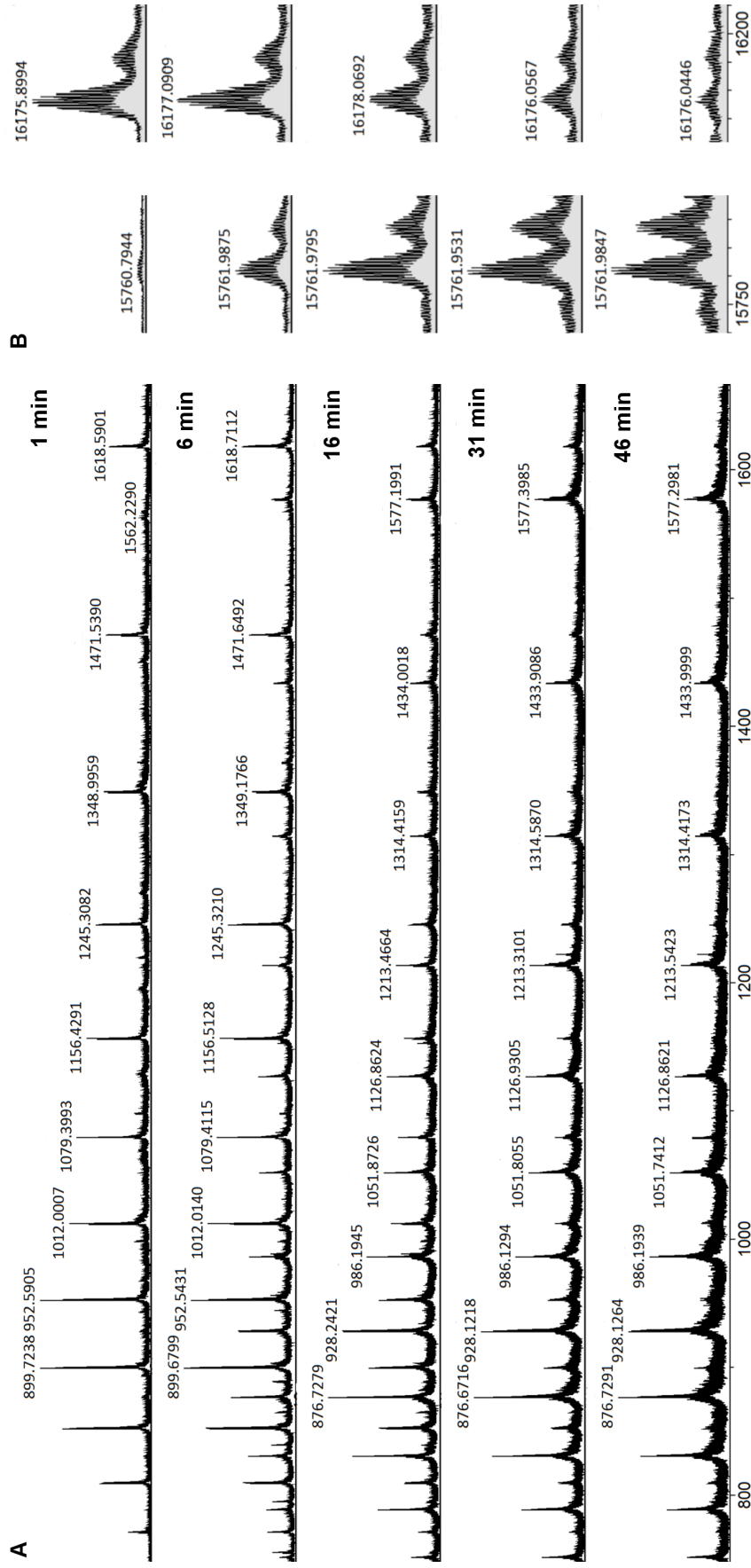
Fig. 5.17 Time-course of the cleavage of crosslinked ADC H17C/E23C/C26S/C78A (*o*NB crosslinker **95**) under UV irradiation. A 64 μM solution of the crosslinked protein in phosphate buffer (50 mM pH 7.4, 100 mM NaCl, 0.5 mM TCEP) in a 5mm path length quartz cuvette placed at a distance of 1 cm from the lamp was continuously irradiated using a Hg lamp (UVP Pen-Ray®, 1 mW at sample position). The amount of cleavage was quantified over time by LC-HRMS by comparing the intensities of the deconvoluted peaks for the crosslinked and cleaved species. (**A**) raw MS spectra at different time-points (1, 6, 16, 31 and 46 min). (**B**) deconvoluted spectra showing the peak for the crosslinked protein on the right (16.176 kDa) and for the cleaved protein on the left (15.762 kDa).

5.4 Conclusion

Although the ADC H17C/E23C/C26S/C78A mutant could not be post-translationally processed to the active enzyme and, therefore, assays to establish whether crosslinking can inhibit protein activity could not be carried out, these studies have shown that that the third generation crosslinkers performed as expected from the design stage. These reagents bearing two phenyl or *o*NB moieties linked through an alkene chain were able to crosslink the protein rapidly and cleanly, without any visible side-products (such as doubly-labelled species). The crosslinking ability of the compounds did not seem to change with increased size of the *o*NB group compared to the non-cleavable phenyl derivative and α-bromoacetates were shown to be good protein labelling reagents. The cleavable *o*NB crosslinker was shown to be cleanly photocleaved by UV irradiation of the sample, even using a low intensity lamp. Although cleavage by UV irradiation of the crosslinked protein leaves a methyl carboxylate group attached to the engineered surface cysteines, it is hypothesised that the dynamics of the protein will dominate post-photolysis and that the small size of this appended group should not influence the behaviour of the free protein. Nevertheless, this assumption will only be verified once the compounds are tested on a fully functioning enzyme construct.

6 Discussion

The last few decades have seen great technological advancements in structural biology techniques, such as X-ray diffraction and scattering, electron microscopy, spectroscopy and NMR. These techniques have allowed for the study of proteins and biomacromolecules with unprecedented structural detail. But even when atomic-resolution is possible, the information obtained is severely incomplete as data on the properties of the system are very scarce. Some efforts have been made to obtain this dynamic information, but the majority of the studies have only provided information on the dynamics of the systems under equilibrium conditions and, although this information can be indirectly correlated with function, the actual states adopted by the protein along the reaction coordinate are unknown.

Very recently, especially since the commission of ultra-bright X-ray sources (free electron lasers as well as synchrotrons), structural biologists have taken a renewed interest in time-resolved structural studies. Such experiments allow for the direct coupling of structural information with function, by observing the states adopted by proteins, in real-time, during function. This approach bridges the gap between functional and structural studies and is capable of finally delivering a much more comprehensive and accurate picture of the involvement of protein motions in function, clarifying the role of protein dynamics. Nevertheless, these techniques are still in their infancy, suffering from challenges mainly related to the synchronisation of protein activity across the ensemble of molecules that make up the sample.

In order to perform a time-resolved measurement, the molecules constituting the sample (as a crystal or solution) have to be triggered, initiating function. Triggering has to be homogeneous and synchronised across the ensemble and is done by changing the sample environment, either by rapid mixing or using light pulses (Fig. 1.8). The main disadvantage of rapid mixing is the time-resolutions that can be achieved, which are defined by the diffusion limit. The fastest mixing currently achievable is in the region of 100s of microseconds. On the other hand, light triggering is only limited by the length of the pulse necessary to deliver sufficient photons to trigger the reaction. It can, therefore, span a much wider time-resolution, down to femtoseconds. However, in order to use a laser-based activation of a protein, a photochemical moiety directly involved in protein function has to be accessible. Due to this limitation, it is unsurprising that the fastest and highest impact time-resolved structural studies have focused on naturally photoactivatable systems, where a naturally occurring chromophore can be activated

using light (section 1.2.1). Nevertheless, the vast majority of proteins are not naturally photoactivatable, requiring the introduction of an artificial chromophore.

Photocaging can be used to provide a method for light-induced protein function of non-naturally photoactivatable systems. This approach focuses on the introduction of a photolabile chemical protecting group at a strategic site on the protein, ligand or substrate, which prevents function until removed. Photolysis of the protecting group is promoted by the absorption of a photon of light of a specific wavelength. Current approaches for the introduction of such chemistry rely either on the full chemical synthesis of a substrate or ligand derivative, carrying the desired photocage, or the direct photocaging of protein residues (section 1.2.2). For photocaged ligands and substrates, the compound has to be carefully designed so that the location of the photocaging protecting group hinders function without completely abolishing binding of the ligand to the protein. Once designed, the compound also has to be synthesised, which can be a non-trivial process, in addition to which the spectroscopic and cleavage properties of the new ligand cannot be absolutely predicted beforehand. Alternatively, protein photocaging can be performed by the introduction of an unnatural amino acid during protein expression, but the current library of photocaged amino acids available for genetic incorporation is very limited, mainly focusing on *o*NB based photocages (Fig. 1.14). Although this technology is developing rapidly, the time required to diversify this library of genetically incorporable photocaged amino acids is completely unpredictable.

With the limitations of the photocaging technology currently available, there is scope for the development of new photocaging strategies that are more widely applicable and less system-dependent. Such advancements would liberate the current bottleneck for time-resolved structural studies of single turnover, non-naturally photoactivatable proteins.

During the course of this project, a new class of photocaging reagents was developed. The reagents were designed to work as protein crosslinkers that could staple across two surface cysteines which can be introduced using standard site-directed mutagenesis. In order to efficiently photocage the protein, the cysteine mutations have to be carefully chosen to make use of the intrinsic dynamic properties of the protein. By surface "stapling", the protein is confined to a smaller conformational space and is unable to access states known to be essential for function. Upon photolysis of the "staple", the

protein is liberated and allowed to resume its normal dynamic behaviour, triggering function.

Three main crosslinking scaffolds were investigated, synthesised and tested for protein bioconjugating activity and stability and, when a suitable compound was found, for cleavage under UV irradiation. The first synthesised crosslinker (section 4.3) contained two *o*NB photolabile moieties connected by an extendable alkyl chain and two secondary bromides at the benzylic position for protein bioconjugation (Omran and Specht, 2009a, Omran and Specht, 2009b). Despite some synthetic difficulties, one compound of this class was synthesised (**53**) but found not to be sufficiently reactive towards thiols to be useful for cysteine bioconjugation. Therefore, α-bromoacetates were proposed as an alternative to the secondary bromide moieties, as these are much more electrophilic and known to react with cysteines (Hermanson, 2008a) and a second scaffold was designed to contain two cleavable α-bromoacetate groups (section 4.4). An initial non-cleavable surrogate was synthesised (**74**) in a one-pot reaction and tested for reactivity and stability. The crosslinker was found to react quickly and cleanly with thiols, supporting the use of α-bromoacetates as bioconjugation moieties but the proximity of the two acetyl groups rendered the compound susceptible to hydrolysis, especially at higher pH. Although the rate of hydrolysis of the gem-α-thioacetate (**79**) was slower than that of the gem-α-bromoacetate (**74**), all attempts at synthesising the corresponding *o*NB (**81**) and coumarinyl (**80**) derivatives were unsuccessful and to date, no alternate routes have been found in the literature. A combination of the synthetic difficulties with the hydrolytic propensity of the gem-α-diacetate scaffold led to the decision that this class of compound was not suitable for further development.

A third generation crosslinker was then designed, bearing two chromophores, each linked to a single α-bromoacetate group, to improve stability of the compounds in buffer. Once again, a benzyl non-photocleavable surrogate (**94**) was synthesised first to test the synthetic route and the stability of the compounds. As the crosslinker contained two chiral centres at the benzylic positions, a chiral resolution step was introduced to attempt to synthesise an enantiomerically pure crosslinker. Once the first racemic intermediate (allyl alcohol **87**) was synthesised, the two enantiomers were separated using an enzymatic chiral resolution. Although the S-alcohol **(S)-87** product from the kinetic resolution was not as enantiomerically enriched as expected, it was taken forward and the crosslinker was obtained as a single diastereomer, as seen from the presence of a single species by ^{1}H NMR. As the compound was purified by flash

chromatography, it is possible that the minor diastereomer was separated during purification.

The same kinetic resolution approach was applied to the *o*NB allyl alcohol (**97**), but with no success. Testing of other enzymes for selectivity towards *o*NB allyl alcohols could potentially yield a suitable enzyme, but since the lipase employed (lipase PS from *B. cepacia*) is one of the most promiscuous lipases and has been shown to be active towards bulky, aromatic secondary alcohols (Chen and Xiang, 2011), the probability of finding a more active enzyme was low. There are numerous non-enzymatic methods for the synthesis of enantiomerically enriched allylic alcohols. Some examples of kinetic resolution approaches include the Sharpless epoxidation using titanium catalysts (Johnson *et al.*, 1993), selective acylation using Fu's catalyst with chiral iron-DMAP complexes (Ruble *et al.*, 1997, Bellemin-Laponnaz *et al.*, 2000) and selective oxidation using ruthenium catalysts (Hashiguchi *et al.*, 1997, Nishibayashi *et al.*, 1999). For this project, the methodology developed by the Carreira group was attempted, using iridium catalysts for the kinetic resolution of secondary alcohols, where one enantiomer reacts preferentially with a primary alcohol to give an enantiomerically pure ether (Roggen and Carreira, 2011). Although the method had been shown to be applicable to bulky aromatic secondary alcohols, the catalyst could not distinguish between the *o*NB alcohol enantiomers.

One alternative approach would be the direct stereoselective synthesis of the allyl alcohol, by either the stereoselective addition of vinyl groups to the allyl ketone using rhodium or iridium catalysts (Skucas *et al.*, 2007) or the stereoselective reduction of ketones (Hashiguchi *et al.*, 1995, Corey and Helal, 1998). Another possible solution would be to purchase a library of different commercially available enzymes and metal catalysts and develop a high-throughput screening method for efficient kinetic resolution conditions. This would require the design of a suitably broad set of reaction conditions that could be run in parallel (possibly using a 96-well plate set-up) coupled with a chiral HPLC protocol that could quantify the efficiency of each resolution reaction easily. Such screens are not currently commercially available and, therefore, this was not a suitable method within the timescale of this project.

Albeit inadvertently, some separation of the *o*NB crosslinker diastereomers was achieved during purification by flash chromatography in the last step of the synthesis of the *o*NB crosslinker (**95**). The ^{1}H NMR spectrum of the purified product showed a mixture of diastereomers, in a 9:1 ratio. The identity of the major and minor species

could not be obtained from spectroscopic analysis alone, and so chiral HPLC methods for the separation of the three possible stereoisomers of the diol precursor **98** (RR, SS and the meso compound) are currently being screened. If a successful method is found, then the purified di-α-bromoacetate crosslinker (**95**) will be hydrolysed back to the corresponding mixture of alcohols (**98**) and analysed by chiral HPLC. Once the identity of the major and minor species are know, the ^{1}H NMR shifts can be assigned to each specific diastereomer. Under such conditions, the meso compound should be separable. Although obtaining an enantiomerically pure crosslinker would be preferable, it has been hypothesised that the differences in conformation of the stapled protein caused by the different enantiomers will be minor and that upon photolysis and release of strain, the recovery of dynamic flexibility will dominate the conformations adopted by the protein. This hypothesis could be tested using molecular dynamics simulations.

Aspartate α-decarboxylase was chosen as the model system to test the new crosslinkers as the C-terminal loop of the β-chain (residues ^{17}His-Glu^{24}) is highly mobile and can adopt two distinct conformations (Albert *et al.*, 1998), only one of which is thought to be active (Yorke & Monteiro, unpublished results). Locking the loop in a conformation resembling the inactive form should abolish (or at least greatly reduce) its catalytic activity. The mutations were designed from inspection of the crystal structure (PDB 1AW8), which contains the two different conformations. Three cysteine pairs were chosen to target different crosslinking lengths and each ADC cysteine double mutant also carried two further point mutations to remove two naturally occurring cysteines (Cys26 and Cys78 mutated to serine and alanine respectively).

The first hurdle encountered was that the ADC mutants did not express well under standard ADC overexpression conditions. After different expression trials, the yield for the H17C/E23C mutant was increased considerably and that for the D19C/E23C mutant moderately. In both cases, the protein did not activate efficiently thermally, showing very low levels of cleavage of the zymogen and also high levels of contamination of Ser25 by-products from hydrolysis of the ester intermediate (Fig. 3.42). The yield for the E23C/A62C variant was very poor under all overexpression conditions attempted and the very small percentage of protein observed to cleave thermally post-translationally showed only the hydrolysed by-product with no pyruvoyl cofactor formed, indicating that the A62C mutation was possibly detrimental for both overexpression and maturation of the protein.

It would be important to investigate whether the lower protein overexpression yield and maturation inefficiency observed were caused by the additive effects from the combination of mutations from any specific point mutations. Each of the single, double and triple ADC mutants should be expressed individually using the standard protocol described for WT ADC (including maturation of the protein by thermal activation in the cell pellet) and the levels of expression and efficiency of maturation (percentage cleaved and ratio between pyruvoyl cofactor formation and hydrolysis to the serine by-product) evaluated. Whereas Cys78 is on the protein surface (close to the central pore) and thus its mutation is not expected to have any major structural impact on the protein, Cys26 is buried and located next to the active site. Its mutation to serine could be a major cause of the low overexpression yields and inefficient protein maturation, as from the sequence conservation of ADC this residue tends to be larger and non-polar (cysteine, valine or isoleucine, see section 9.7.1).

As Glu23 is directly involved in the interaction between PanZ.CoA and PanD, and the mobile region is in close proximity to the site of cleavage, targeting the loop by site-directed mutagenesis might always have a detrimental effect on protein maturation. From inspection of the available crystal structures for the enzyme, no other highly dynamic regions can be identified and, therefore, ADC may be a better candidate for standard substrate photocaging approaches. Photocaged L-aspartate derivatives have been previously synthesised by Dr B. A. Yorke and shown to bind to the ADC active site (unpublished results). This work is still on-going, in collaboration with Dr M. V. Sans (Hamburg Centre for Ultrafast Imaging).

Nevertheless, these crosslinking compounds should be suitable for a wide range of enzymes and other proteins that show dynamic properties that could be harnessed for protein inactivation. Dihydrofolate reductase (DHFR, section 1.1.2) is an obvious candidate, as the dynamics of the Met20 loop have been extensively studied and the involvement of its three distinct conformations (open, closed and occluded) during the different steps involved in catalysis have been thoroughly characterised. Targeting the Met20 loop for conformational locking should have the desired effect of decreasing the catalytic ability of DHFR. The design of the cysteine anchors could be initially done by inspection of the available crystal structures but complementing the study with MD simulations would allow for a more informed and careful design of the crosslinking sites. Although the dynamic properties of this enzyme have been thoroughly investigated there is still scope to further study the system as two main questions remain

unanswered: what is the transition state conformation and how does DHFR exchange between the closed and occluded conformations. The main conformations have been obtained from structural studies of DHFR constructs in complex with substrate or cofactor analogues, leaving the exact mechanisms of conformer interchange and the corresponding transient structures still unresolved (Venkitakrishnan *et al.*, 2004).

Although activity studies could not be carried out due to the problems discussed above, the crosslinkers were still shown to perform as expected from the design stage. The reagents were successfully used to crosslink an ADC mutant and the bioconjugation reaction proceeded cleanly and quickly, with no double labelling or hydrolysis visible by high-resolution mass spectrometry. Once the protein was crosslinked, low-power irradiation with UV light (λ_{max}=254 nm), using a 1 mW mercury lamp, showed that the crosslinker cleaved cleanly, leaving the expected small methyl carboxylate groups on the cysteine anchors. The rate of cleavage decreased over time, possibly due to the absorption from the nitrobenzyl ketone by-products.

The quantum yield could not be estimated from this experiment as quantifying the amount of photons absorbed by the sample is non-trivial. The most efficient way to obtain a value for the quantum yield of these compounds will be performing cleavage reactions of both a previously characterised compound, such as *o*NB-ATP (Fig. 1.10), and the novel crosslinkers in parallel and under the same conditions (lamp power, irradiation time and optical density) and extrapolating the value for the quantum yield by comparison of the two experiments.

The rate of cleavage of the nitrobenzyl upon absorption will be best estimated using pump-probe spectroscopy, where the formation of the nitroso product could be monitored in real time using either by UV-Vis or IR (Klán *et al.*, 2013). These experiments can also help elucidate the mechanism of cleavage of the crosslinker, as they could allow for the detection of cleavage intermediates. These measurements are also non-trivial, reflected in the lack of information regarding the kinetics of cleavage of photocaged compounds (sections 1.2.2 and 9.6).

The crosslinker synthesised and characterised in this thesis is expected to follow similar cleavage rates to previously characterised *o*NB photocaged compounds (Corrie *et al.*, 2005), on the order of milliseconds, although the exact rate of cleavage may be faster (or slower) due to the influence of the allylic group. To obtain higher time resolutions, a different photocaging moiety will have to be used. The synthesis of a coumarinyl version of this crosslinker was also attempted, but unfortunately due to the highly

insoluble nature of the di-coumarin diol scaffold (**108**), the final bromoacetylation step could not be performed. Grubbs metathesis of the allyl α-bromoacetate intermediate (**109**) also failed, with no observable consumption of the starting material. To date, no synthetic solution to these issues has been found and thus the development of a coumarin-based crosslinker would depend on the design of a new di-coumarinyl scaffold (**110**), such as the one shown in Fig. 6.1. Once again, solubility problems are expected as the scaffold is highly conjugated and the nature of the vinyl coumarin may have a detrimental effect on the metathesis step. Nevertheless, the synthesis to the first intermediate (**111**) from the condensation of 2-hydroxy acetophenone (**112**) with 3-methylacryloyl chloride (**113**) has been previously described in literature (Königs *et al.*, 2008) and, therefore, the metathesis reaction could be quickly investigated. The high conjugation of this scaffold may confer very interesting, although completely unknown, spectroscopic properties worth investigating.
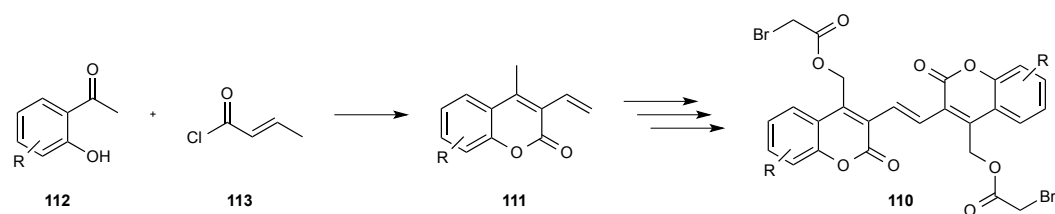
Fig. 6.1 A 4th generation photocleavable crosslinker scaffold (**110**). The design is based on the knowledge gained throughout this thesis and attempts to circumvent the problems encountered during the synthesis of the coumarinyl crosslinker. The first step to the 3-alkenylcoumarin (**111**) from the condensation of 2-hydroxy acetophenone (**112**) with 3-methylacryloyl chloride (**113**) has been reported previously (Königs *et al.*, 2008). The alkenylcoumarin could be used to test the metathesis step before any further investment of time in the synthesis of the final crosslinker.

This thesis describes initial steps towards the design and development of new photocleavable crosslinkers. There is still great scope for the development of novel photocaging reagents and, also very importantly, for a comprehensive characterisation of such compounds. Furthermore, there is also great interest from structural biologists in obtaining functional and well-characterised compounds for time-resolved experiments. Such work is best carried out in a highly interdisciplinary environment, as a collaborative effort between chemists, structural biologists and spectroscopists will allow for a targeted approach, from the compound design stage to the characterisation and testing of the compounds under relevant experimental conditions.

7 Conclusions

During this project, a new class of photocleavable crosslinking reagents, that allow for the stapling across two protein surface cysteine residues, was designed, synthesised and tested. These reagents aimed to decouple the triggering step in time-resolved experiments from the chemistry and function being observed. Three different scaffolds were investigated, each providing information about the stability, synthetic availability and reactivity of functional groups and photocaging moieties. An *ortho*-nitrobenzyl crosslinker carrying bromoacetate groups as protein bioconjugation moieties was synthesised and found to be stable and able to staple across engineered surface cysteine mutations on aspartate α-decarboxylase. The crosslinked protein was subjected to low power UV irradiation, which showed that the crosslinker could be cleaved to release the protein. Experiments to obtain both the quantum yield and rates of decaging under pump-probe conditions using brighter UV sources are currently being designed and will be carried out in collaboration with the Hamburg Centre for Ultrafast Imaging.

The ADC mutants generated for testing the crosslinking and photocleavage of these novel reagents could not be fully and cleanly activated, which prevented the investigation of crosslinking on enzymatic activity. Nevertheless, the mutants will be tested for activation by thermal activation at high temperatures and *in cellulo* in WT *E. coli* cell lines and the extent of cleavage and pyruvoyl formation quantified by mass spectrometry. Once this information is obtained, activity assays will be carried out if the protein is found to be sufficiently activated. Current efforts are also focusing on targeting a different enzyme, DHFR, for de-activation studies by crosslinking.

Alongside the development of these new photocleavable crosslinkers, the molecular basis of activation of ADC was investigated, as obtaining fully active, homogeneous protein is an important step in time-resolved experiments as it greatly decreases the levels of heterogeneity during data reduction. The function of PanZ as the activating factor for ADC was investigated using several biophysical techniques, such as ITC, NMR, MS and SAXS and the ADC.PanZ.AcCoA crystal structure was solved at high resolution (1.5 Å) along with two complexes of PanZ with PanD mutants. The activation step was found to be promoted by the formation of a strained and activation-competent conformation of PanD upon binding to PanZ. This mechanism, coined "mechano-chemistry" is supported by the observation of an unprecedented backbone cleavage event caused by X-ray radiolysis. The CoA dependence of the interaction also allowed for a novel mode of regulation of CoA production in bacteria to

be proposed, imposing a negative feedback loop mechanism where the formation of the enzyme which produces an intermediate (β-alanine) is regulated by a final metabolite. This mechanism was confirmed by both *in vitro* catalytic assays as well as *in cellulo* studies.

Although the activity studies have not been carried out yet, the designed photocleavable crosslinkers were shown to react as expected. It will be important to fully characterise the cleavage kinetics of these compounds to establish the timescales the crosslinkers could provide for time-resolved studies. Nevertheless, there is still scope for the further development of new scaffolds, as well as a more thorough characterisation of the kinetics of the currently available photocaging approaches.

8   Experimental

8.1   General information and instrumentation

The following procedures were used for all the experimental work, unless otherwise stated. All reagents and consumables were purchased from known general suppliers, including Sigma Aldrich, Alfa Aesar, Merck, Iris Biochem, Fisher Scientific, Acros Organics, IDT, GE Healthcare, Hampton Research, Molecular Dimensions, Mitegen and Millipore. All water used was filtered to 15 MΩ, except for water used for PCR work and crystallisation experiments, which was filtered to 18.2 MΩ. Sterile work was performed adjacent to a blue flame or in a flow hood. All glassware, plastic-ware and media contaminated with genetically modified bacteria were disposed of by bleaching using Virkon or Presept for 24 h or autoclaved (120 °C for 20 min) prior to disposal as aqueous waste or by incineration. NMR data were collected using a Bruker Avance 500, Bruker DRX500 or Bruker DPX300 spectrometer and analysed using MestReNova (Mestrelab Research). Chemical shifts are reported in parts per million (ppm) downfield from tetramethyl silane (TMS) and coupling constants (J) are given in Hertz (Hz). Infra-red (IR) spectra were recorded using a Bruker alpha with platinum ATR spectrometer. High-resolution mass spectrometry was carried out on a Bruker maXis Impact spectrometer using electrospray ionisation. Small molecules were analysed as solutions in ultrapure solvents. Protein MS was carried out in the original buffers, as the samples were purified and denatured by liquid chromatography (LC, 0-95% MeCN in H_2O over 7 min). The raw mass spectrum was deconvoluted using a maximum entropy algorithm, part of the Data Analysis software (Bruker). Routine low-resolution liquid chromatography mass spectrometry (LC-MS and LC-MSMS) was performed on a Phenomenex Luna C18 50×2 mm column using a gradient of 5-90% MeCN in H_2O over 1.8 min using electrospray ionisation. Results are quoted as mass-charge ratios (m/z). Analytical thin layer chromatography (TLC) was performed using silica gel pre-coated plates (Merck) and visualised using UV irradiation or standard staining agents (potassium permanganate, dinitrophenylhydrazine, bromocresol green, *p*-anisaldehyde and phosphomolybdic acid). Flash column chromatography was carried out on silica gel 60 (230-400 mesh, Merck) or on Florisil® (<200 mesh, Sigma). Solvents were removed under reduced pressure using a Buchi rotary evaporator at diaphragm pump pressure (typically 40-450 mbar). Traces of solvents were removed under high vacuum. High performance liquid chromatography (HPLC) analysis was carried out on an Agilent HPLC system, using a Thermo Electron Corporation

Hyperprep HS C18 column (18 μm, 250×4.6 mm) and diode array as a detector. A gradient of water and MeCN (5-95%) was used as solvent at a flow rate of 1 mL min^{-1}.

8.2 Stock solution recipes

The recipe components were dissolved in 15 MΩ H_2O to the final volume stated. All stock solutions, with the exception of *1000×metals, 25×salts* and *20×NPS,* were sterilised by filtration through a 0.22 μm membrane.

25×salts (1L): 88.7 g Na_2HPO_4, 85 g KH_2PO_4, 67 g NH_4Cl and 17.8 g Na_2SO_4.

1000×metals (100 mL): 1.35 g $FeCl_3.6H_2O$, 438 mg $CaCl_2.6H_2O$, 198 mg $MnCl_2.4H_2O$, 288 mg $ZnSO_4.7H_2O$, 47.6 mg $CoCl_2.6H_2O$, 34.1 mg $CuCl_2.2H_2O$, 48.4 mg $Na_2MoO_4.2H_2O$, 34.6 mg Na_2SeO_3, 12.4 mg H_3BO_3, 52.6 mg $NiSO_4.6H_2O$ and 500 μL HCl (37% in H_2O).

20×NPS (200 mL): 13.2 g $(NH_4)_2SO_4$, 27.2 g KH_2PO_4, 29.2 g Na_2HPO_4

20×NPS (^{15}N) (200 mL): 13.2 g ($^{15}NH_4)_2SO_4$, 27.2 g KH_2PO_4, 29.2 g Na_2HPO_4

50×amino acids (50 mL): 500 mg of each: sodium L-glutamate, L-lysine hydrochloride, L-arginine hydrochloride, L-histidine hydrochloride, L-aspartic acid, L-alanine, L-proline, L-glycine, L-threonine, L-serine, L-glutamine, L-asparagine, L-valine, L-leucine, L-isoleucine, L-phenylalanine and L-tryptophan.

1000×vitamins (no B_{12}) (20 mL): 400 μL 10 mM nicotinic acid, 400 μL pyridoxine hydrochloride, 400 μL 10 mM thiamine hydrochloride, 400 μL 10 mM *para*-aminobenzoic acid, 400 μL 10 mM pantothenate, 100 μL 1 mM folic acid and 100 μL 1 mM riboflavin.

1000×vitamins (with B_{12}) (20 mL): same as for *1000×vitamins (no B_{12})* with 800 μL 5 mM vitamin B_{12}.

50×5052 (50 mL): 12.5 g glycerol, 36.5 mL H_2O, 1.25 g glucose and 5 g lactose.

M9 salt solution (1 L): 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g, NH_4Cl.

10% Ampicillin: 1 g ampicillin sodium salt in 10 mL of H_2O. Split into 1 mL aliquots and stored at -20 °C.

5% Kanamycin: 5 g kanamycin sulphate in 10 mL of H_2O. Split into 1 mL aliquots and stored at -20 °C.

10×PCR buffer: 100 mM Tris (pH 8.3), 500 mM KCl, 15 mM $MgCl_2$. Stored in 1 mL aliquots at -20 °C.

Pwo polymerase storage buffer: 50 mM Tris (pH 7.9), 50 mM KCI, 0.1 mM EDTA, 1 mM DTT, 50% glycerol.

TAE buffer (1L): 4.84 g Tris, 5.71 mL glacial acetic acid and 292 mg EDTA.

6×DNA loading buffer: 5 mM Tris (pH 8.0), 0.15 % *w/v* Orange Gelb, 2.5% *w/v* Ficoll 400, 10 mM EDTA.

8.3 Preparation of media for cell growth and cloning

8.3.1 LB media

LB media was obtained as a freeze-dried powder. A 25 g L^{-1} suspension of LB in H$_2$O was autoclaved at 120 °C for 20 min.

8.3.2 LB-agar media

A suspension of 25 g L^{-1} LB powder and 15 g L^{-1} of Agar powder in H$_2$O was autoclaved at 120 °C for 20 min.

8.3.3 Auto-induction media

10 g of Tryptone, 5g of yeast extract, 40 mL of *25×salts solution*, 10 mL of 50% *v/v* aqueous glycerol, 1 mL of 2 M MgCl$_2$ aqueous solution and 200 μL of *1000×metals solution* were mixed in a 2 L conical flask. H$_2$O was added up to 1L and the suspension autoclaved at 120 °C for 20 min. To the cooled sterilised solution was added 1 mL of 50% *w/v* aqueous glucose (filtered through a 0.22 μm filter) and 10 mL of 20% *w/v* aqueous lactose (filtered through a 0.22 μm filter).

8.3.4 SeMet minimal media A

For 100 mL final volume. 92 mL of H$_2$O, 10 μL of *1000×metals* and 1 mL 1 M aqueous MgSO$_4$ were mixed thoroughly, 5 mL of *20×NPS* were added and the solution autoclaved at 120 °C for 20 min. To the sterilised solution were added 1.25 mL of 40% *w/v* aqueous glucose, 1 mL of *50×amino acids*, 400 μL of 25 mg mL^{-1} methionine, 100 μL of *1000×vitamins (with B$_{12}$)* and 100 μL of 10% *w/v* aqueous ampicillin.

8.3.5 SeMet minimal media B

For 500 mL final volume. 450 mL of H$_2$O, 500 μL of 1 M MgSO$_4$ and 50 μL of *1000×metals* were mixed thoroughly, 25 mL of *20×NPS* were added and the solution autoclaved at 120 °C for 20 min. 10 mL of *50×5052*, 10 mL of *50×amino acids*, 200 μL

of 25 mg mL^{-1} methionine, 2.5 mL 25 mg mL^{-1} selenomethionine, 500 µL 10% *w/v* aqueous ampicillin and 500 µL *1000×vitamins (no B$_{12}$)* were added.

8.3.6 SeMet wash buffer

For 100 mL final volume. 0.10 g of NaH$_2$PO$_4$, 0.58 g Na$_2$HPO$_4$, 2.92 g NaCl and 20 g ethylene glycol were mixed and water added up to 100 mL.

8.3.7 ^{15}N minimal media

For 500 mL final volume. 450 mL of H$_2$O, 500 µL of 1 M MgSO$_4$ and 50 µL of *1000×metals* were mixed thoroughly, 25 mL of *20×NPS (^{15}N)* were added and the solution autoclaved at 120 °C for 20 min. To the sterilised solution were added 10 mL of *50×5052*, 10 mL of *50×amino acids*, 100 µL of *1000×vitamins (with B$_{12}$)*, 100 µL of 10% *w/v* aqueous ampicillin.

8.4 Site-directed mutagenesis, cloning and sequencing

The following protocols were used to generate site-directed mutants of PanD (H17C, D19C, E23C, C26S, A62C, C78A, H17C/E23C, D19C/E23C, E23C/A62C, C26S/C78A,   D19C/E23C/C78A,   D19C/E23C/C26S,   E23C/C26S/C78A, H17C/E23C/C26S/C78A,   D19C/E23C/C26S/C78A,   E23C/C26S/A62C/C78A, T57V/K119STOP). All ADC mutants were generated from the pRSETA-ADC-WT vector plasmid (Saldanha *et al.*, 2001), with the exception of the T57V/K119STOP mutant which was generated from the pRSETA-ADC(T57V) plasmid (Webb *et al.*, 2014).

8.4.1 Primer design

The overlapping site-directed mutagenesis primers for QuikChange (Papworth *et al.*, 1996) were designed manually. They were 25-45 bases long (the average number of bases was 26). The codon carrying the desired mutation was placed roughly in the middle of the primer, with 10-15 bases corresponding to the plasmid sequence on each side. The exact sequence was designed in order to obtain a melting temperature (T$_m$) ≥78 °C, calculated using

$$T_\text{m} = 81.5 + 0.41(\%\text{GC}) - (675/N) - \%\text{mismatch}$$

Where %GC is the percentage of G or C bases, N is the total number of bases in the primer and %mismatch the number of mismatched bases compared to the initial

sequence. The sequence was also optimised to have a GC content of approximately 40% and terminate on one or more C or G bases.

8.4.2   Primers used for genetic manipulation

Mutant	Sequence
H17C	5'-CGTGAAAGTGACTTGTGCGGACCTGC-3' 3'-CGCAAGAACCTTCATAGTGCAGGTCC-5'
D19C	5'-GTGACTCATGCGTGCCTGCACTATGAA-3' 3'-CAATGGCGCAAGAACCTTCATAGTGC-5'
E23C	5'-CTCSTGCGGACCTGCACTCTTGTGGTT-3' 3'-CAATGGCGCAAGAACCACAATAGTGC-5'
C26S	5'-CACTATGAAGGTTCTTCCGCCATTGAC-3' 3'-CCTGGTCAATGGCGGAAGAACCTTCA-5'
A62C	5'-CCACTTATGCCATCGCGTGTGAACGC-3' 3'-CTCGAACCGCGTTCACACGCGATGGC-5'
C78A	5'-CGGCGGCCCACGCAGCCAGTGTCGGC-3' 3'-GCCGACACTGGCTGCGTGGGCCGCCG-5'
H17C (E23C)	5'-CGTGAAAGTGACTTGTGCGGACCTGC-3' 3'-CGGAAGAACCACAATAGTGCAGGTCC-5'
D19C (E23C)	5'-GTGACTCATGCGTGCCTGCACTATTGT-3' 3'-CAATGGCGCAAGAACCACAATAGTGC-5'
E23C (C26S)	5'-CTCATGCGGACCTGCACTATTTGTGTT-3' 3'-CAATGGCGGAAGAACCACAATAGTGC-5'
H17C (E23C/C26S)	5'-CGTGAAAGTGACTTGTGCGGACCTGC-3' 3'-CGGAAGAACCACAATAGTGCAGGTCC-5'
D19C (E23C/C26S)	5'-GTGACTCATGCGTGCCTGCACTATTGT-3' 3'-CAATGGCGGAAGAACCACAATAGTGC-5'
K119STOP	5'-GAAACGTACCGCGTAAGCGAT-3' 3'-GTACCGGAATCGCTTACGCGG-5'

8.4.3   Mutagenesis protocol

Site-directed mutants of ADC were obtained using the standard QuikChange mutagenesis protocol published by Agilent Technologies (Papworth *et al.*, 1996). All primers were dissolved in 18 MΩ water to a final concentration of 200 μM. Initial mutations were performed starting from the expression pRSETA plasmid encoding a N-terminally hexaHis-tagged WT PanD (Saldanha *et al.*, 2001) or a N-terminally hexaHis-tagged PanD(T57V) (Webb *et al.*, 2014).

For each site-directed mutagenesis reaction, the PCR mix was prepared by mixing 20 ng of the template plasmid (as a solution in 10 mM Tris-Cl, pH 8.5, from miniprep using the Qiagen QIAprep Spin Miniprep Kit), 0.4 μL of 25 nM dNTPs (an equimolar

solution of dGTP, dCTP, dTTP, dATP), 5 μL of 10×PCR buffer, 0.2 μL forward primer (200 μM solution), 0.2 μL reverse primer (200 μM solution) and 18 MΩ H$_2$O to 49 μL followed by 1 μL of 3 U μL^{-1} Pwo polymerase. PCR was performed using the following protocol: 30 s at 95 °C, 16×(30 s at 95 °C, 60 s at 55 °C, 300 s at 68 °C) and 300 s at 68 °C.

8.4.4   Agarose gel electrophoresis

The following protocol is for a standard 1% agarose gel. 0.4 g of agarose was suspended in 40 mL of *TAE buffer*. The suspension was heated (microwave) to obtain a solution. *TAE* buffer was added to compensate for any evaporation followed by 1.5 μL of ethidium bromide. The solution was inverted twice, poured into a casting mould and the desired comb inserted. Once the gel was set, it was transferred to the electrophoresis running tank, submerged in TAE buffer and the DNA samples (5-10 μL of 5:1 *v/v* mixtures of the desired DNA solution and *6×DNA loading buffer*) were loaded and the gel run at 100 V for 20 min.

8.4.5   Transformation of chemically competent *E. coli* cells

This protocol was used for the transformation of all *E. coli* strains used, XL10, BL21 Gold, BL21 Star, C41 (DE3), MG1655 Δ*panZ* Δ*panD* (DE3) and B834 (DE3).

1-5 μL of plasmid was mixed with 50 μL (or 10 μL when transforming XL10 cells) of chemically competent cells in a sterile 1.5 mL tube on ice. The cells were incubated for 10 min to allow diffusion of the plasmid, then at 42 °C for 45 s before re-incubation on ice for a further 10 min for plasmid uptake. 1 mL of LB media was added and the cells incubated at 37 °C for 45 min. 100 μL of culture were used to inoculate sterilised agar plates, containing the desired antibiotic for selection (100 μg mL^{-1} of ampicillin or 50 μg mL^{-1} of Kanamycin). The remaining 900 μL of cell culture were centrifuged at 13,000*g* for 30 seconds, 800 μL of supernatant removed, the pelleted cells resuspended and used to inoculate a second agar plate. The plates were incubated at 37 °C for 16 hours.

8.4.6   Miniculture of *E. coli* cells

A single colony from transformation of *E. coli* cells strains harbouring the desired plasmid for overexpression purposes was picked and incubated in 5 mL of LB media containing the desired antibiotic for selection (100 μg mL^{-1} of ampicillin or 50 μg mL^{-1} of Kanamycin) for 16 hours at 37 °C with shaking. The miniculture was used to

inoculate 1L overexpression media as well as to make glycerol cell stocks (1:1 mixture with 80% *v/v* sterilised glycerol (in H_2O), flash frozen and stored at -80 °C for later use).

8.4.7   Sequencing of plasmids

A single colony of *E. coli* XL10 cells transformed with the desired plasmid was minicultured as described above. The cells were harvested by centrifugation at 4,500 rpm for 10 min and the plasmid harvested with a QIAprep miniprep kit (QIAGEN®), using the manufacturer's protocol and buffers (Sambrook and Russell, 2001). 150 ng of T7 sequencing primer was added to 10 ng of plasmid in 50 µL of 18 MΩ H_2O. Sequencing was performed by GATC® and analysed using Bioedit.

8.5   Protein overexpression

N-Terminally hexaHis-tagged WT ADC, PanD(T57V) and PanD(S25A) from vectors pRSETA-ADC-WT (Saldanha *et al.*, 2001), pRSETA-ADC(T57V) (Webb *et al.*, 2014) and pRSETA-ADC(S25A) (Schmitzberger *et al.*, 2003) respectively. C-Terminally hexaHis-tagged PanZ was overexpressed from vector pBAD24-PanZ (Nozaki *et al.*, 2012) or pET28a-PanZ (Monteiro *et al.*, 2015). All other PanD mutants were expressed from pRSETA vectors encoding the desired proteins generated as described in 8.4.

8.5.1   Overexpression in auto-induction media

All N-terminal hexaHis-tagged WT PanD was overexpressed in *E. coli* C41 (DE3) cells harbouring a pRSETA-ADC-WT plasmid (Saldanha *et al.*, 2001). N-Terminally hexaHis-tagged PanD mutants were overexpressed in *E. coli* MG1655 *ΔpanZ ΔpanD* (DE3) cells harbouring pRSETA plasmids coding for the desired proteins. C-Terminally hexaHis-tagged PanZ was overexpressed in *ΔpanZ ΔpanD E. coli* C41 cells harbouring a pET28a plasmid. Overexpression was done by auto-induction of T7 RNA polymerase (Studier, 2005). 1 L of sterilised auto-induction media with antibiotic (100 µg mL^{-1} ampicillin for PanD or 50 µg mL^{-1} kanamycin for PanZ) in a 2 L conical flask was inoculated with 1 mL of overnight miniculture. The culture was incubated for 20 h at 37 °C with shaking (200 rpm). The cells were collected by centrifugation (10,000*g* for 10 min) and were stored at –80 °C prior to lysis for protein purification. For activated ADC, the cell pellets were incubated at 37 °C overnight prior to storage at -80 °C.

8.5.2 Overexpression in LB media by arabinose induction

C-Terminally hexaHis-tagged PanZ was overexpressed by induction of the arabinose promoter of the pBAD24 plasmid coding for the protein in *E. coli* BL21 Gold cells. 1 L of sterilised LB media with 100 μg mL^{-1} ampicillin and 250 mM calcium pantothenate in a 2 L conical flask was inoculated with 1 mL of overnight miniculture. The culture was incubated at 37 °C with shaking (200 rpm) and overexpression induced by addition of 5 g of arabinose (0.5% *w/v*) at OD$_{600}$=0.6. The cells were incubated for a further 16 hours. The cells were collected by centrifugation (10,000g for 10 min) and were stored at -80 °C prior to lysis for protein purification.

8.5.3 Overexpression of SeMet PanZ

C-Terminally hexaHis-tagged PanZ was overexpressed in *E. coli* B834 (DE3) cells harbouring a pBAD24 plasmid coding for the desired protein following a previously described protocol (Sreenath *et al.*, 2005). In summary, 3 mL of overnight miniculture in *SeMet minimal media A* was used to inoculate 100 mL of *SeMet minimal media A*. The cell culture was incubated at 37 °C, 200 rpm for 18 h. Following incubation, 20 mL of the cell culture were used to inoculate 480 mL of *SeMet minimal media B* and the cultures incubated at 37 °C, 200 rpm for 24 h. The cells were collected by centrifugation (10,000g for 10 min), resuspended in *SeMet wash buffer* to wash any remaining SeMet away, collected by centrifugation once again (10,000g for 10 min) and then stored at -80 °C prior to lysis for protein purification.

8.6 Protein purification

8.6.1 Buffer recipes and preparation

Buffers used for Ni-NTA chromatography were used directly without sterilisation or filtration. Size exclusion chromatography buffers were filtered through a 0.22 μm membrane under reduced pressure.

Ni-NTA stock buffer: 500 mM K$_2$HPO$_4$, 3000 mM NaCl, pH 7.4 (15.0 g KH$_2$PO$_4$, 67.8 g K$_2$HPO$_4$, 175.3 g NaCl).

Ni-NTA lysis buffer: 50 mM K$_2$HPO$_4$, 300 mM NaCl, 10 mM imidazole, pH 7.4 (100 mL *Ni-NTA stock buffer*, 681 mg imidazole).

Ni-NTA wash buffer: 50 mM K$_2$HPO$_4$, 300 mM NaCl, 50 mM imidazole, pH 7.4 (100 mL *Ni-NTA stock buffer*, 3.40 g imidazole).

Ni-NTA elution buffer: 50 mM K_2HPO_4, 300 mM NaCl, 250 mM imidazole, pH 7.4 (100 mL *Ni-NTA stock buffer*, 17.02 g imidazole).

SEC buffer A: 50 mM Tris, 100 mM NaCl, 0.1 mM DTT, pH 7.5 (4.93 g Tris, 2.92 g Tris-HCl, 11.70 g NaCl, 30.9 mg DTT).

SEC buffer B: 50 mM KH2PO4, 100 mM NaCl, 0.5 mM TCEP, pH 7.5 (1.24 g KH_2PO_4, 7.11 g K_2HPO_4, 5.84 g NaCl, 143 mg TCEP).

8.6.2 Purification of hexaHis-tagged proteins by Ni-NTA chromatography

E. coli cells collected from 1 L of overexpression cultures were resuspended in Ni-NTA lysis buffer (50 mM K_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 7.4) with protease inhibitor cocktail (cOmplete EDTA-free, Roche) according to the manufacturer's protocol at 4 °C. The cell slurry was lysed mechanically using a Constant Systems cell disrupter (20 kpsi). DNase (a few crystals) was added to the cell lysate, which was cleared of cell debris by centrifugation (30,000 *g*, 45 min). The Ni-NTA agarose resin (Qiagen, 10 mL) was resuspended in 20 mL of H_2O, washed under gravity with 30 mL of H_2O and equilibrated with 50 mL of Ni-NTA lysis buffer (50 mM K_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 7.4). The clear lysate was loaded onto the Ni-NTA agarose column and low affinity binders removed by washing the resin with 50 mL of Ni-NTA wash buffer (50 mM K_2HPO_4, 300 mM NaCl, 50 mM imidazole, pH 7.4). The protein was eluted in 5×5 mL fractions of Ni-NTA elution buffer (50 mM K_2HPO_4, 300 mM NaCl, 250 mM imidazole, pH 7.4). The protein size, overexpression level and purity were assessed by SDS-PAGE.

8.6.3 Ni-NTA column regeneration

Following protein elution, the Ni-NTA loaded agarose resin was regenerated to remove any contaminants. The resin was washed with 30 mL of 0.1% *w/v* sodium dodecyl sulphate (SDS) followed by 30 mL of 0.1 M NaOH. Ni^{2+} was removed by chelation with 30 mL of 10 mM ethylenediaminetetraacetic acid (EDTA), the agarose resin washed with 100 mL of H_2O and regenerated with 20 mL of 500 mM Ni_2SO_4. Excess, unbound Ni^{2+} was removed by washing with 100 mL of H_2O and the column stored in 20% *v/v* EtOH.

8.6.4 Size-exclusion chromatography

Elution fractions from Ni-affinity purification containing protein were combined and concentrated to 1 mL (Amicon centrifugal concentrator 10 kDa MWCO, 4,000 *g*). The

concentrated protein was loaded onto a Superdex® 75 column (HiLoad® 16/60 or 26/60, GE Healthcare) and eluted isocratically with *SEC buffer A or B* (ÄKTA purifier, GE Healthcare) in 5 mL (26/60 column) or 3 mL (16/60 column) fractions and stored at 4 °C. Protein elution was monitored by absorption at 280 nm and the chromatogram used to assess purity.

8.6.5 Determination of protein concentration

Protein concentration was determined using UV absorbance. 1.5 μL of sample were loaded onto the Nanodrop instrument and the absorbance measured at 280 nm. Protein concentration was calculated using PanD ε_{280} 15470 M^{-1} cm^{-1} or PanZ ε_{280} 26470 M^{-1} cm^{-1}, obtained empirically from the protein sequences using the ProtParam online platform (Gasteiger *et al.*, 2005).

8.7 SDS-PAGE analysis

8.7.1 Buffer recipes

5×SDS-PAGE Tris-glycine running buffer: 15.1 g Tris base, 94 g glycine, 5 g SDS.

SDS-PAGE Tris-tricine gel buffer: 3 M Tris, 0.3% SDS, pH 8.45 (the pH was adjusted by addition of 4 M NaOH).

SDS-PAGE Tris-tricine cathode buffer: 0.1 M Tris, 0.1 M tricine, 0.1% SDS, pH 8.23 (the pH was adjusted by addition of 4 M NaOH).

SDS-PAGE Tris-tricine anode buffer: 0.2 M Tris, pH 8.90 (the pH was adjusted by addition of 4 M NaOH).

2×SDS-PAGE loading buffer: 50 mM Tris hydrochloride, pH 6.8 (the pH was adjusted by addition of 4 M NaOH), 100 mM dithiothreitol, 2% *w/v* SDS, 0.1% *w/v* bromophenol blue, 10% *w/v* glycerol.

Coomassie blue stain: 1.0 g Coomassie blue R-250, 400 mL MeOH, 500 mL H_2O, 100 mL glacial acetic acid.

Coomassie destain: 400 mL MeOH, 500 mL H_2O, 100 mL glacial acetic acid.

8.7.2 Sample preparation

Purified protein solutions

30 μL of protein in buffer was mixed with *2×SDS-PAGE loading buffer*. The solutions were heated at 100 °C for 10 min (to ensure full denaturation of the protein).

E. coli cells

1 mL of cell culture was centrifuged at 13,000 rpm, the supernatant removed, the cell pellet resuspended in Ni-NTA lysis buffer. 30 μL of the suspension was mixed with *2×SDS-PAGE loading buffer* and heated at 100 °C for 10 min (to ensure full denaturation of the protein).

8.7.3 Tris-glycine PAGE

The 12% separating gel was prepared according the table below. All components except for the ammonium persulphate (APS) and tetramethylethylenediamine (TEMED) were mixed thoroughly by inversion. Prior to casting, APS was added followed by TEMED and the solution mixed carefully between the additions. The mixture was poured and a layer of H_2O-saturated *n*-butanol was added to ensure a levelled surface. The stacking gel mixture was prepared following the same protocol as for the separating gel. Once the separating gel was set, the *n*-butanol was blotted off, the stacking gel solution added and the comb inserted. The gel was loaded with the desired protein samples in 1×SDS-PAGE loading buffer and run in 1×SDS-PAGE Tris glycine running buffer at constant voltage (180 V).

Tris-glycine PAGE	Separating gel (12%)	Stacking gel
1.5 M Tris pH 8.8	3.75 mL	-
1.0 M Tris pH 6.8	-	625 μL
30 % acrylamide (29:1 acrylamide:bis-acrylamide	6.00 mL	660 μL
10% SDS	150 μL	50 μL
Water	5.10 mL	3.66 mL
15% APS	200 μL	100 μL
TEMED	10 μL	10 μL

8.7.4 Tris-tricine PAGE

The 10% separating and stacking gels were prepared according the table below. All components except for the ammonium persulphate (APS) and tetramethylethylenediamine (TEMED) were mixed thoroughly by vortexing and the solution allowed to set for 20 min prior to casting, APS was added followed by TEMED and the solution mixed carefully between the additions. The separating gel was poured followed by the stacking gel, the comb inserted and the gels allowed to set simultaneously. Gentle addition of the stacking gel solution was essential to minimise disturbances to the separating gel. The gel was loaded with the desired protein samples in 1×SDS-PAGE loading buffer and the *SDS-PAGE Tris-tricine cathode* and *SDS-*

PAGE Tris-tricine anode buffers were added. The gel was run at constant current of 30 mA until the samples entered the separating gel and then at 60 mA.

Tris-tricine PAGE	Separating gel (10%)	Stacking gel
Glycerol	2 mL	-
Tris-tricine gel buffer	5.00 mL	3.10 mL
30 % acrylamide (29:1 acrylamide:bis-acrylamide	5.08 mL	1.66 mL
Water	2.92 mL	7.74 mL
15% APS	100 μL	100 μL
TEMED	10 μL	10 μL

8.7.5   Visualisation of SDS-PAGE gels

Coomassie blue stain (50 mL) was added to the gel, the solution warmed in a microwave (15 s, full 750 W) and the gel incubated at RT for 1.5 h with rocking. The stain was removed, *Coomassie destain* was added, the gel incubated for 3 h with rocking, washed with H_2O and photographed (Gel Doc XR system, Bio-Rad).

8.8   Binding assays

All binding assays were performed using isothermal titration calorimetry (ITC). Experiments were performed using a Microcal iTC200 (GE) or Microcal VP-ITC (GE) thermostated at 25 °C. For the microcal iTC200, the ligand sample was loaded into the sample cell (200 μL) and the titration sample into the sample syringe (40 μL). Each titration series consisted of a sacrificial 0.4 μL injection followed by 19 injections of 2 μL. For the VP-ITC, the ligand sample was loaded into the sample cell (1.5 mL) and the titration sample into the sample syringe (300 μL). Each titration series consisted of a sacrificial 2 μL injection followed by 29 injections of 10 μL each. The titration sample concentration was ~8× that of the ligand sample. The proteins were expressed and purified as described in 8.5 and 8.6, and used in *SEC buffer A*. The titration data were analysed using Origin 6.5. After baseline subtraction, data were fitted to a single site-binding model. For global fitting, data were integrated using NITPIC (Keller *et al.*, 2012) before global fitting to a one-site binding model in SEDPHAT (Houtman *et al.*, 2007). For titrations of purified PanZ (a mixture of *apo*-PanZ and PanZ.CoA) against CoA and against PanD(T57V) without excess CoA or AcCoA, binding constants were re-calculated accounting for the binding incompetent fractions of PanZ in each case. A summary of all the titrations and binding constants obtained can be found in appendix 9.4.

8.9 Protein crystallisation

8.9.1 Initial crystallisation screens

The PanD(T57V)-PanZ.CoA protein complex was prepared from solutions of the proteins (overexpressed from the pRSETA-ADC(T57V) (Webb *et al.*, 2014) and the pBAD24-PanZ (Nozaki *et al.*, 2012) vectors and purified as described in section 8.6) in *SEC buffer A* (50 mM Tris, 100 mM NaCl, 0.1 mM DTT, pH 7.5). The proteins were mixed in a 1:1 ratio (protomer to monomer), concentrated to 3.5 mg mL^{-1} (Amicon centrifugal concentrator 10 kDa MWCO, 4,500 g) and 1 PanZ equivalent of CoA was added. Sparse matrix screens using Crystal Screen HT 1 and Crystal Screen HT 2 (96 conditions, Hampton Research) were set up manually in a MRC 96-well plate (Molecular Dimensions) by the sitting drop diffusion method with 1.0 μL protein:1.0 μL mother liquor sitting drops at 18 °C. Small bipyramidal crystals were obtained in 100 mM Tris (pH 7.8-8.0), 180-220 mM Li$_2$SO$_4$ and 20% *v/v* PEG 4000. A second crystallisation screen was set up similarly but using the PanD(T57V-PanZ.AcCoA complex instead. Small square plates were obtained in 100 mM sodium cacodylate, pH 6.5, 1.4 M sodium acetate and small bipyramidal crystals were obtained in 100 mM Tris (pH 7.8), 200 mM Li$_2$SO$_4$ and 20% *v/v* PEG 4000.

8.9.2 Further crystallisation screens

The PanD(T57V)-PanZ.AcCoA protein complex was prepared from solutions of the proteins overexpressed from the pRSETA-ADC(T57V) (Webb *et al.*, 2014) and the pET28a-PanZ (Nozaki *et al.*, 2012) vectors and purified as described in section 8.6 in *SEC buffer A* (50 mM Tris, 100 mM NaCl, 0.1 mM DTT, pH 7.5). The proteins were mixed in a 10:11 PanD(T57V):PanZ ratio (protomer to monomer), concentrated to ~9 mg mL^{-1} (Amicon centrifugal concentrator 10 kDa MWCO, 4,500 g) and 2 equivalent of AcCoA were added. Sparse matrix screens using Crystal Screen HT 1 and 2, Index, Salt RX (Hampton Research), Morpheus (Molecular Dimensions) and Wizard I and II (Emerald Biosciences) were set-up using the Oryx 6 Douglas crystallisation robot (Douglas instruments) in MRC 96-well plates (Molecular Dimensions) using 1.0 μL protein:1.0 μL mother liquor at 18 °C. More than 120 crystallisation hits were obtained, with various salts and buffers. The largest crystals were observed generally in conditions containing high molecular weight PEG precipitants, simple salts of mono- or divalent cations and Tris or Bis-Tris buffers.

8.9.3 Optimisation of initial crystallisation conditions

Optimisation of crystallisation conditions was carried out by varying salt, buffer and precipitant concentrations and buffer pH. Crystals were grown by the hanging drop diffusion method, in 2 μL protein:2 μL mother liquor drops in 24-well plates.

The PanD(T57V)-PanZ.CoA complex was prepared as described in section 8.9.1 and crystallised in 200 mM Li_2SO_4, 100 mM Tris pH 8.5 and 30% v/v PEG 4000. The PanD(T57V)-PanZ.AcCoA complex was prepared as described in section 8.9.2 and crystallised in 100 mM sodium cacodylate, pH 6.5, 1.4 M sodium acetate and in 200 mM KSCN, 100 mM Bis-Tris propane pH 6.5, 20% v/v PEG 3350. The PanD(S25A)-PanZ.AcCoA was prepared similarly to the PanD(T57V)-PanZ.AcCoA complex and crystallised in 200 mM KSCN, 100 mM Bis-Tris propane pH 6.5, 20% v/v PEG 3350.

8.9.4 Heavy-metal derivatisation

PanD(T57V).PanZ.CoA crystals obtained as described in section 8.9.3 were derivatised with heavy-metals. The crystals were harvested from the crystallisation drops using CryoLoops (Hampton Research) or MicroMeshes™ (Mitegen) and transferred to a 2 μL drop of mother liquor on a glass cover slip containing 10 μM of a heavy-metal salt. $Hg(OAc)_2$, $(NH_4)_{10}H_2(W_2O_7)_6$, K_2OsCl_6, $(H_2NCH_2CH_2NH_2)PtCl_2$, $Pb(CH_3CO_2)_4$ and $K_2Pt(CN)_4$ were screened. The drops were sealed in a 24-well crystallisation plate to avoid dehydration. The crystals were allowed to soak for a few minutes and transferred to new mother liquor drops without the heavy-metal additives to attempt to remove any non-specifically bound metal derivatives.

8.9.5 X-ray Diffraction experiments

Evaluation of crystal quality

Crystal size and quality were evaluated by light microscopy and confirmed by X-ray diffraction. Diffraction patterns were collected using either the in-house source (MicroMax-007 HF microfocus rotating anode generator, 30 mA, Raxis4++), at both RT and 100 K, or at Diamond Light Source (beamlines I24, I03, I04-1 and I02) at 100 K.

Synchrotron cryo-temperature data collection

Small bipyramidal crystals for the PanD(T57V).PanZ.CoA complex were obtained in 200 mM Li_2SO_4, 100 mM Tris pH 8.5 and 30% v/v PEG 4000. Some crystals were

derivatised as described in 8.9.4. Bipyramidal crystals for the PanD(T57V).PanZ.AcCoA, WT ADC.PanZ.AcCoA and WT PanD.PanZ.AcCoA complexes were obtained in 200 mM KSCN, 100 mM Bis-Tris propane pH 6.5, 20% *v/v* PEG 3350. Square plates for the PanD(T57V).PanZ.AcCoA complex were obtained from 100 mM sodium cacodylate, pH 6.5, 1.4 M sodium acetate. The crystals were harvested and mounted on CryoLoops (Hampton Research) or MicroMeshes™ (Mitegen) on goniometer bases. The crystals were cryo-protected stepwise in mother liquor containing 5%, 10% and 20% *v/v* glycerol by soaking for a few seconds in each condition and flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K under a cryo-stream of dry nitrogen at beamlines I03, I24 and I04-1 at DLS. For each crystal, 3 diffraction patterns were first obtained with an attenuated X-ray beam and used to evaluate crystal quality, index the crystal lattice and devise a data collection strategy using EDNA (Incardona *et al.*, 2009). A summary of the data collection strategy details for the PanD(T57V).PanZ.CoA and PanD(T57V).PanZ.AcCoA complex described in section 3.2 can be found in section 9.2.

Data for the WT PanD.PanZ.AcCoA complex were collected at beamline I03 at DLS at λ= 0.9763 Å. 900 frames of 0.2° oscillation, 0.1 s exposure and 30% transmission were collected to a maximum resolution of 1.16 Å. The data were integrated in space group *I4* (a=b=85.9 Å, c=80.1 Å, $\alpha=\beta=\gamma=90°$) using XDS (Kabsch, 2010).

Data for the WT ADC.PanZ.AcCoA complex were collected at beamline I04-1 at DLS at λ= 0.9174 Å. 1000 frames of 0.2° oscillation, 0.2 s exposure and 43% transmission were collected to a maximum resolution of 1.25 Å. The data were integrated in space group *I4* (a=b=85.9 Å, c=80.1 Å, $\alpha=\beta=\gamma=90°$) using XDS (Kabsch, 2010).

Synchrotron RT data collection

Bipyramidal crystals for PanD(T57V).PanZ.AcCoA complex were obtained in 200 mM KSCN, 100 mM Bis-Tris propane pH 6.5, 20% *v/v* PEG 3350. The crystals were mounted on CryoLoops (Hampton Research) or MicroMeshes™ (Mitegen) on goniometer bases. Data for the PanD(T57V)-PanZ.AcCoA were collected at RT at beamline ID29 at ESRF. The crystals were placed under a humidifying air stream (HC1 instrument (Sanchez-Weatherby *et al.*, 2009)) and kept at 95% relative humidity throughout the experiment. 3 diffraction patterns were first obtained and used to evaluate crystal quality, index the crystal lattice and devise a data collection strategy using EDNA Mxv1. 2320 frames of 0.05° oscillation, 0.04 s exposure and 0.12% transmission were collected to a maximum resolution of 1.64 Å. The data were

integrated in space group *I4* (a=b=86.4 Å, c=81.0 Å, α=β=γ=90°) using Xia2 (Winter, 2010).

RT in-house data collection

Bipyramidal crystals for the PanD(T57V).PanZ.AcCoA, WT ADC.PanZ.AcCoA and PanD(S25A).PanZ.AcCoA complexes were obtained in 200 mM KSCN, 100 mM Bis-Tris propane pH 6.5, 20% *v/v* PEG 3350. The crystals were harvested and mounted on CryoLoops (Hampton Research) or MicroMeshes™ (Mitegen) on goniometer bases. The mounted crystals were inserted into plastic sleeves containing a 2 μL plug of mother liquor and data collected at RT using the in-house source (Micro-Max-007 HF microfocus rotating anode generator, 30 mA and a R-axis 4++ detector) at λ=1.5418 Å. 2 initial diffraction patterns were collected and used to evaluate crystal quality, index the crystal lattice and devise a data collection strategy using HKL3000 (Otwinowski and Minor, 1997). 60 frames of 1° oscillation, 30 s exposure and full transmission were collected to a maximum resolution of 1.70 Å and 2.10 Å resolution respectively. The PanD(T57V)-PanZ.AcCoA data were integrated in space group *I4* (a=b=86.4 Å, c=81.0 Å, α=β=γ=90°) and PanD(S25A)-PanZ.AcCoA in space group *I4* (a=b=86.3 Å, c=80.8 Å, α=β=γ=90°) using HKL3000 (Otwinowski and Minor, 1997). The WT ADC.PanZ.AcCoA data were integrated in space group *I4* (a=b=85.6 Å, c=79.9 Å, α=β=γ=90°) using XDS (Kabsch, 2010).

Structure solution and refinement

All datasets were scaled and merged in Aimless (Evans and Murshudov, 2013). The synchrotron dataset for PanD(T57V).PanZ.AcCoA collected at the ESRF on beamline ID29 (PDB 4CRY) was solved by molecular replacement of the PanD(T57V) protomer (PDB 4AZD, (Webb *et al.*, 2014) first, followed by the lowest energy conformation of the NMR structure of PanZ (PDB 2K5T) using Molrep (Vagin and Teplyakov, 1997). The model was subjected to cycles of refinement using Refmac5 (Murshudov *et al.*, 2011) and manual rebuilding using Coot (Emsley *et al.*, 2010). All subsequent complex structures of PanD(T57V).PanZ.AcCoA and PanD(S25A).PanZ.AcCoA from in-house datasets and WT ADC.PanZ.AcCoA and WT PanD.PanZ.AcCoA from in-house and synchrotron datasets were solved by molecular replacement of the PDB 4CRY model using Molrep and were manually rebuilt and refined using Coot and Refmac5 respectively. Isotropic refinement parameters were used for the PanD(T57V).PanZ.AcCoA, PanD(S25A).PanZ.AcCoA and RT WT ADC.PanZ.AcCoA

158

structures. Mixed isotropic/anisotropic refinement parameters were used for the cryo WT ADC.PanZ.AcCoA and WT PanD.PanZ.AcCoA structures, where all protein atoms were refined anisotropically, with the exception of the PanZ [26]Tyr-Val[33] loop which is mobile. Waters were inspected manually and refined isotropically if considered mobile. All ligands were also refined isotropically with the exception of AcCoA.

8.10 Complementation assays

The *in cellulo* assays were performed by Shingo Nozaki and Hironori Niki. The full protocol has been described previously (Monteiro *et al.*, 2015).

8.11 SAXS

The SAXS data were collected by Dr Thomas Grant, Dr Edward Snell and Prof. Arwen Pearson and processed by Dr Thomas Grant according to the previously published protocol (Monteiro *et al.*, 2015).

8.12 ^{1}H NMR ADC activity assay

Activated WT ADC and PanZ were overexpressed and purified as described in sections 8.5.1 and 8.6. A mixture of 1:2 WT ADC and PanZ (final concentrations of 15.8 and 31.7 μM respectively) in *SEC buffer A* with 50 μM AcCoA was diluted into *SEC buffer B* to a final concentration of 100 nM WT ADC and 200 nM PanZ (and consequently 320 nM AcCoA). The mixture was split into 6 aliquots, supplemented with AcCoA, 0.5 mM L-aspartate and PanZ, as described in Table 10, and incubated at 37 °C. ^{1}H NMR samples were prepared by mixing 630 μL of the enzymatic reaction with 70 μL of D_2O (10% *v/v* D_2O in H_2O final concentration) and the 1D ^{1}H NMR spectrum recorded immediately using a water suppression pulse sequence. The reaction progression was monitored at 2 h, 4 h and 24 h after the start of incubation. Substrate turnover was monitored by analysis of the ratio of the integrals of the L-aspartate CαH and the β-alanine CH_2 signals. The integrals were calculated using Mestrenova (Mestrelab research software). The experiment was repeated with a 250 nM WT ADC and 500 nM PanZ solution supplemented with AcCoA (100 μM) and 5 mM L-aspartate.

**Table 10** Sample composition for ^{1}H NMR assay of WT ADC catalysis in the presence of PanZ

	1	2	3	4	5	6
[PanZ] (μM)	0.20	0.20	1.00	1.00	10.0	10.0
[AcCoA] (μM)	0.32	100	0.32	100	0.32	100

8.13 Synthesis of photocleavable crosslinkers
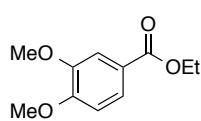
Ethyl 3,4-dimethoxybenzoate (45). (Choshi *et al.*, 1992) 3,4-Dimethoxybenzoic acid (15.0 g, 82 mmol) was dissolved in EtOH (100 mL), conc. H_2SO_4 (2.60 mL) was added and the solution heated under reflux for 20 h. The solution was cooled to RT, diluted with H_2O (200 mL), extracted with EtOAc (3×120 mL), the organic layer quenched with sat. aq. $NaHCO_3$ (150 mL), washed with brine (100 mL), dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure, yielding the product as a clear liquid (16.1 g, 93%), which was used without further purification. ^1H NMR (CDCl$_3$, 300 MHz): 7.69 (1H, dd, *J* 1.8, 8.1, ArC6H), 7.55 (1H, d, *J* 1.8, ArC2H), 6.89 (1H, d, *J* 8.1, ArC5H), 4.36 (2H, q, *J* 7.2, CH$_2$), 3.94 (6H, s, OCH$_3$), 1.39 (3H, t, *J* 7.2, CH$_3$). ^{13}C NMR (CDCl$_3$, 75 MHz): 166.4 (C=O), 152.8 (ArC4), 148.5 (ArC3), 123.4 (ArC1), 123.0 (ArC6), 111.9 (ArC2), 110.1 (ArC5), 60.8 (CH$_2$), 55.9 (OCH$_3$), 14.4 (CH$_3$). HRMS: 233.0775 [M+Na]$^+$ ([C$_{11}$H$_{14}$O$_4$Na]$^+$ requires 233.0784), 211.0967 [M+H]$^+$ ([C$_{11}$H$_{15}$O$_4$]$^+$ requires 211.0965). IR: 1712 cm^{-1} (C=O).
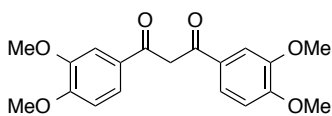
**1,3-Bis(3,4-dimethoxyphenyl)-1,3-propanedione (47).** Prepared using a modified procedure to previously described (Omran and Specht, 2009a). NaH (1.75g, 44 mmol) was washed in dry hexane (2×35 mL) and dry PhMe (35 mL) and suspended in dry PhMe (160 mL) under a N_2 atmosphere. 3,4-Dimethoxyacetophenone (7.02 g, 39 mmol) was added and the mixture stirred for 15 min. Ethyl 3,4-dimethoxybenzoate (7.74 mL, 43 mmol) was added slowly, the mixture stirred for a further 30 min and heated under reflux for 36 h. The solution was cooled to RT, quenched into H_2O (150 mL), the aqueous layer extracted with EtOAc (3×100 mL), the organic extracts combined and washed with brine (100 mL), dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure to yield an orange solid. The crude was purified by flash chromatography (150 g SiO$_2$, with a solvent gradient of 30-65% EtOAc in hexane) yielding the product as a bright yellow solid (4.38 g, 33%). ^1H NMR (CDCl$_3$, 300 MHz): 7.59 (2H, dd, *J* 2.0, 8.5, ArC6H), 7.54 (2H, d, *J* 2.0, ArC2H), 6.91 (2H, d, *J* 8.5 Hz, ArC5H), 6.74 (1H, s, CH$_2$), 3.97 (6H, s, OCH$_3$), 3.94 (6H, s, OCH$_3$). ^{13}C NMR (CDCl$_3$, 75 MHz): 184.5 (C=O), 152.7 (ArC4), 149.1 (ArC3), 128.4 (ArC1), 121.0 (ArC6), 110.6 (ArC2), 109.9 (ArC5), 91.8 (C2), 56.0 (OCH$_3$), 55.9 (OCH$_3$). HRMS: 367.1157 [M+Na]$^+$ ([C$_{19}$H$_{20}$O$_6$Na]$^+$ requires 367.1152), 345.1329 [M+H]$^+$ ([C$_{19}$H$_{21}$O$_6$]$^+$ requires 345.1333). IR: 1736 cm^{-1} (C=O), 1599 cm^{-1} (C=C).
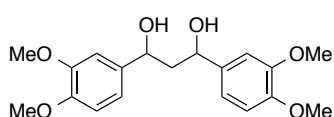
**1,3-Bis(3,4-dimethoxyphenyl)-1,3-propanediol (48).** (Omran and Specht, 2009a) 1,3-Bis(3,4-dimethoxyphenyl)-1,3-propanedione (2.55 g, 7.4 mmol) was dissolved in CH_2Cl_2 (37.5 mL) and MeOH (300 mL) was added followed by NaBH$_4$ (855 mg, 22.6 mmol). The mixture was stirred at RT for 40 min and the solvent removed under reduced pressure, leaving a yellow solid. The solid was redissolved in CH_2Cl_2 (80 mL), washed with H_2O (2×50 mL) and brine (50 mL) and the organic layer dried over Na_2SO_4. The solvent was removed under reduced pressure yielding the product as a yellow foam (2.58 g, 99%, mixture of diastereomers, 2:1 meso:anti), which was used without further purification. ^{1}H NMR (CDCl$_3$, 500 MHz): 6.92 (0.68H, d, J 1.7, ArC2H), 6.88 (1.32H, d, J 1.4, ArC2H'), 6.87-6.83 (2H, m, ArC6H and ArC6H'), 6.81 (0.68H, d, J 6.8, ArC5H), 6.79 (1.32H, d, J 8.1, ArC5H'), 4.91 (0.68H, dd, J 10.2, 2.4, CH), 4.88 (1.32H, t, J 5.8, CH'), 3.86 (6H, s, OCH$_3$ and OCH$_3$'), 3.85 (6H, s, OCH$_3$ and OCH$_3$'), 3.92 (0.68H, br s, OH), 3.51 (1.32H, br s, OH'), 2.18 (0.66H, dt, J 14.5, 10.2, C2H), 2.12 (0.68H, t, J 5.8, CH$_2$), 1.91 (0.66H, dt, J 14.5, 2.7, C2H). ^{13}C NMR (CDCl$_3$, 75 MHz): 184.5 (C=O), 152.7 (ArC4), 149.1 (ArC3), 128.4 (ArC1), 121.0 (ArC6), 110.6 (ArC2), 109.9 (ArC5), 74.8 (CH), 71.5 (CH'), 56.0 (OCH$_3$), 55.9 (OCH$_3$), 47.7 (CH$_2$) 46.6 (CH$_2$'). HRMS: 347.1487 [M-H][-] ([C$_{19}$H$_{23}$O$_6$][-] requires 347.1500), 371.1474 [M+Na][+] ([C$_{19}$H$_{24}$O$_6$Na][+] requires 371.1465). HPLC: 0.32 min. IR: 3469 cm[-1] (br, OH), 1138 cm[-1] (C-OH).
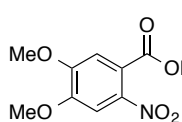
Ethyl 4,5-dimethoxy-2-nitrobenzoate (50). (Knesl *et al.*, 2006) Nitric acid (70%, 6.9 mL) was cooled to 0 °C and a solution of ethyl 3,4-dimethoxybenzoate (5.17 mL, 28.6 mmol) in AcOH (28.6 mL) was added dropwise. The mixture was stirred at 0 °C for 30 min, slowly warmed up to RT and stirred for a further 24 h. The solution was poured into an ice/H$_2$O slurry (150 mL), extracted with EtOAc (2×50 mL), the organic extracts combined and washed with sat. aq. NaHCO$_3$ (3×50 mL) and brine (2×75 mL), dried over Na$_2$SO$_4$, decolourised by addition of charcoal, filtered and the solvent removed under reduced pressure. The crude product was dried in a desiccator to remove any traces of solvent, yielding the product as an off-white amorphous solid (3.95 g, 60%), which was used without further purification. ^{1}H NMR (CDCl$_3$, 300 MHz): 7.60 (1H, s, ArC3H), 7.08 (1H, s, ArC6H), 4.38 (2H, q, J 7.2, CH$_2$), 3.99 (3H, s, OCH$_3$), 3.98 (3H, s, OCH$_3$), 1.36 (3H, t, J 7.2, CH$_3$). ^{13}C NMR (CDCl$_3$, 75 MHz): 165.8 (C=O), 152.5 (ArC5), 150.3 (ArC4), 142.0 (ArC2), 122.0 (ArC1), 110.8 (ArC6), 1067.0 (ArC3), 62.4 (CH$_2$), 56.6

(OCH₃), 56.6 (OCH₃), 13.8 (CH₃). HRMS: 278.0647 $[M+Na]^+$ ($[C_{11}H_{13}NO_6Na]^+$ requires 278.0635). IR: 1710 cm⁻¹ (C=O), 1519 cm⁻¹ (NO₂).
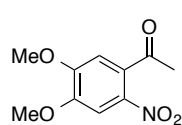
**4,5-Dimethoxy-2-nitroacetophenone (51).** (Dyer and Turnbull, 1999) Acetic anhydride (10 mL) was dissolved in nitric acid (70%, 200 mL), cooled to 0 °C and a solution of 3,4-dimethoxyacetophenone (11.5 g, 63.8 mmol) in acetic anhydride (40 mL) was added dropwise. The mixture was allowed to warm up to RT and stirred for a further 3 h. The solution was poured into H₂O (1.5 L), further diluted to 2 L and cooled in an ice bath to yield a heavy yellow precipitate. The product was filtered under reduced pressure, washed with water and dried in a desiccator (9.89 g, 69%), which was used without further purification. ¹H NMR (CDCl₃, 300 MHz): 7.62 (1H, s, ArC3H), 6.80 (1H, s, ArC6H), 4.02 (3H, s, OCH₃), 3.99 (3H, s, OCH₃), 2.50 (3H, s, CH₃). ¹³C NMR (CDCl₃, 75 MHz): 202.4 (CO), 154.0 (ArC5), 149.6 (ArC4), 138.0 (ArC2), 132.8 (ArC1), 110.8 (ArC6), 107.0 (ArC3), 56.7 (OCH₃), 56.6 (OCH₃), 30.4 (CH₃). HRMS: 226.0709 $[M+H]^+$ ($[C_{10}H_{12}NO_5]^+$ requires 226.0710), 248.0529 $[M+Na]^+$ ($[C_{10}H_{11}NO_5Na]^+$ requires 248.0529). IR: 1733 cm⁻¹ (C=O), 1494 cm⁻¹ (NO₂), 1372 cm⁻¹ (NO₂).
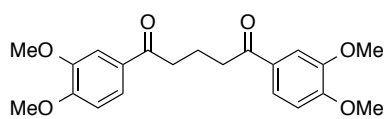
1,5-Bis(3,4-dimethoxyphenyl)-1,5-pentanedione (54). (Omran and Specht, 2009a) Aluminium chloride (3.20 g, 24.0 mmol) was suspended in CH₂Cl₂ (20 mL), the mixture cooled to -10 °C and 1,2-dimethoxybenzene (2.80 mL, 22.0 mmol) was added dropwise over 10 min. Pentanedioyl dichloride (1.26 mL, 10.0 mmol) was added at a rate of 20 µL min⁻¹, the solution warmed up to 0 °C and stirred at this temperature for 5 h. The mixture was quenched into 2 M aq. HCl (100 mL) in ice (200 g) and allowed to warm up to RT. CH₂Cl₂ (100 mL) was added added to give two clear layers. The aqueous layer was further extracted with 200 mL of CH₂Cl₂, the organic extracts combined and washed with H₂O (100 mL), 10% *w/v* aq. K₂CO₃ (3×100 mL) and brine (100 mL), dried over Na₂SO₄, filtered and the solvent removed under reduced pressure. The resulting red oil was purified by flash chromatography (100 g SiO₂, with a gradient of 0-20% Et₂O in CH₂Cl₂), yielding the product as an off-white solid (820 mg, 22%). ¹H NMR (CDCl₃, 300 MHz): 7.63 (2H, dd, *J* 1.1, 8.2, ArC6H), 7.56 (2H, d, *J* 1.1, ArC2H), 6.89 (2H, d, *J* 8.2, ArC5H), 3.95 (12H, s, OCH₃), 3.07 (4H, t, *J* 6.9, <u>CH₂</u>CH₂<u>CH₂</u>), 2.19 (2H, qn, *J* 6.9, CH₂<u>CH₂</u>CH₂). ¹³C NMR (CDCl₃, 75 MHz): 198.6 (C=O), 153.3 (ArC4), 149.0 (ArC3), 130.1 (ArC1), 122.7 (ArC6), 110.2 (ArC2), 110.4 (ArC5), 56.1 (OCH₃), 56.0 (OCH₃), 37.4

($\underline{CH_2}CH_2\underline{CH_2}$), 19.5 ($CH_2\underline{CH_2}CH_2$). HRMS: 395.1459 [M+Na]$^+$ ([$C_{21}H_{24}O_6Na$]$^+$ requires 395.1471), 373.1659 [M+H]$^+$ ([$C_{21}H_{25}O_6$]$^+$ requires 373.1651). IR: 1661 cm^{-1} (C=O).
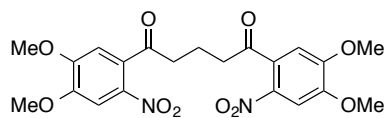
1,5-Bis(4,5-dimethoxy-2-nitrophenyl)-1,5-pentanedione (57). Prepared using a modified procedure to previously described (Omran and Specht, 2009a). HNO$_3$ (70%, 220 µL, 3.48 mmol) was dissolved in CH$_2$Cl$_2$ (2.6 mL) at -15 °C. TFA (810 µL, 10.5 mmol) was added dropwise at -15 °C and the mixture stirred for 45 min at this temperature. 1,5-Bis(3,4-dimethoxyphenyl)-1,5-pentanedione (458 mg, 1.31 mmol) in CH$_2$Cl$_2$ (1.4 mL) was added dropwise and the mixture warmed up to 0 °C and stirred for a further 2 h. The reaction was quenched by careful addition of sat. aq. NaHCO$_3$ (14 mL), the mixture extracted with CH$_2$Cl$_2$ (14 mL), the organic extract washed with brine, dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The crude was purified by flash chromatography (75 g SiO$_2$, 10% Et$_2$O in CH$_2$Cl$_2$), yielding the product as a yellow solid (307 mg, 53%). ^{1}H NMR (CDCl$_3$, 300 MHz): 7.59 (2H, s, ArC3H), 6.85 (2H, s, ArC6H), 4.00 (6H, s, OCH$_3$), 3.97 (6H, s, OCH$_3$), 2.96 (4H, t, *J* 6.6, $\underline{CH_2}CH_2\underline{CH_2}$), 2.25 (2H, p, *J* 6.6, $CH_2\underline{CH_2}CH_2$). ^{13}C NMR (CDCl$_3$, 75 MHz): 202.3 (C=O), 154.3 (ArC5), 149.4 (ArC4), 137.9 (ArC2), 138.0 (ArC1), 108.7 (ArC6), 106.8 (ArC3), 56.8 (OCH$_3$), 56.6 (OCH$_3$), 41.4 ($\underline{CH_2}CH_2\underline{CH_2}$), 17.4 ($CH_2\underline{CH_2}CH_2$). HRMS: 485.1168 [M+Na]$^+$ ([$C_{21}H_{22}N_2O_{10}Na$]$^+$ requires 485.1172). IR: 1694 cm^{-1} (C=O).
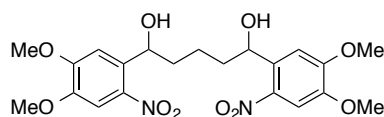
1,5-Bis(4,5-dimethoxy-2-nitrophenyl)-1,5-pentanediol (58). (Omran and Specht, 2009a) 1,5-Bis(4,5-dimethoxy-2-nitrophenyl)-1,5-pentanedione (300 mg, 0.68 mmol) was suspended in a mixture of CH$_2$Cl$_2$:MeOH (1:1, 7 mL), and NaBH$_4$ (104 mg, 2.74 mmol) was added in one portion. The mixture was stirred for 45 min, evaporated to dryness and redissolved in CH$_2$Cl$_2$ (12 mL). The solution was washed with H$_2$O (2×7 mL) and brine (7 mL), dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure to yield the product as a flaky, yellow solid (298 mg, 99%)*, which was used without further purification. ^{1}H NMR (CDCl$_3$, 300 MHz): 7.58 (2H, s, ArC3H), 7.29 (2H, s, ArC6H), 5.45 (2H, m, CH), 4.00 (6H, s, OCH$_3$), 3.94 (6H, s, OCH$_3$), 2.46 (1H, d, *J* 3.9, OH), 2.35 (1H, d, *J* 3.9, OH), 1.84 (6H,

---

* Only one set of signals was visible by NMR: either the diastereomers have the same NMR spectrum or the reduction was stereoselective.

m, CH$_2$). ^{13}C NMR (CDCl$_3$, 75 MHz): 153.7 (ArC5), 147.7 (ArC4), 139.6 (ArC2), 136.1 (ArC1), 109.0 (ArC6), 107.7 (ArC3), 69.4 (CH), 56.4 (OCH$_3$), 56.4 (OCH$_3$), 37.7 (CH$_2$CH$_2$CH$_2$), 23.0 (CH$_2$CH$_2$CH$_2$). HRMS: 489.1489 [M+Na]$^+$ ([C$_{21}$H$_{26}$N$_2$O$_{10}$Na]$^+$ requires 489.1485). IR: 3292 cm^{-1} (br, OH), 1507 cm^{-1} (NO$_2$), 1324 cm^{-1} (NO$_2$).
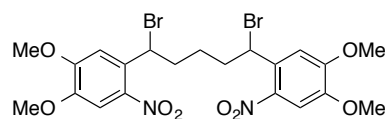

1,5-Bis(4,5-dimethoxy-2-nitrophenyl)-1,5-dibromopentane (53). (Omran and Specht, 2009a) 1,5-Bis(4,5-dimethoxy-2-nitrophenyl)-1,5-pentanediol (45 mg, 0.10 mmol) was dissolved in acetyl bromide (289 µL, 0.29 mmol) at 0 °C and HBr (48% aq. solution, 20 µL, 0.02 mmol) was added. The solution was allowed to warm up to RT and stirred for 1.5 h. The mixture was quenched by addition of sat. aq. NaHCO$_3$ (5 mL), extracted with CH$_2$Cl$_2$ (3×7 mL), the organic extracts combined and washed with brine (7 mL), dried over Na$_2$SO$_4$, filtered and concentrated under reduced pressure. The crude was filtered through a short bed of Florisil® and the solvent removed under reduced pressure, yielding the product as a light yellow foam (55 mg, 99%, 1:1 mixture of diastereomers). ^{1}H NMR (300 MHz, CDCl$_3$): 7.49 (2H, s, ArC3H), 7.46 (2H, s, ArC3H'), 7.23 (2H, s, ArC6H), 7.22 (2H, s, ArC6H'), 5.89 (2H, t, *J* 7.2, CH), 5.85 (2H, t, *J* 5.4, CH'), 4.02 (6H, s, OCH$_3$), 4.01 (6H, s, OCH$_3$'), 3.95 (6H, s, OCH$_3$), 3.94 (6H, s, OCH$_3$'), 2.2 (8H, m, CH$_2$CH$_2$CH$_2$/CH$_2$CH$_2$CH$_2$'), 1.80 (4H, m, CH$_2$CH$_2$CH$_2$/CH$_2$CH$_2$CH$_2$). ^{13}C NMR (CDCl$_3$, 75 MHz): 153.4 (ArC5), 153.3 (ArC5'), 148.6 (ArC4/ArC4'), 139.9 (ArC2/ArC2'), 131.4 (ArC1), 131.2 (ArC1'), 111.7 (ArC6), 111.6 (ArC6'), 107.4 (ArC3/ArC3'), 56.5 (OCH$_3$/OCH$_3$'), 56.4 (OCH$_3$/OCH$_3$'), 48.5 (CH), 48.2 (CH'), 39.5 (CH$_2$CH$_2$CH$_2$), 39.2 (CH$_2$CH$_2$CH$_2$'), 26.4 (CH$_2$CH$_2$CH$_2$), 26.3 (CH$_2$CH$_2$CH$_2$'). HRMS: 511.0712 [M-Br]$^+$ ([C$_{21}$H$_{24}$^{79}Br$_1$N$_2$O$_8$]$^+$ requires 511.0716), 628.9533 [M+K]$^+$ ([C$_{21}$H$_{24}$^{79}Br$_2$N$_2$O$_8$K]$^+$ requires 628.9536). IR: 1495 cm^{-1} (NO$_2$), 1345 cm^{-1} (NO$_2$).
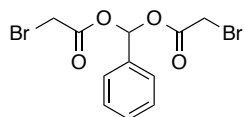
α,α-Di(bromoacetyl) toluene (74). Benzaldehyde (195 µL, 1.92 mmol) and bromoacetic anhydride (750 mg, 2.89 mmol) were mixed and stirred at 85 °C for 16 h under inert atmosphere. The mixture was quenched into sat. aq. NaHCO$_3$ (50 mL), the aqueous layer extracted with CH$_2$Cl$_2$ (3×50 mL), the organic extracts combined, washed with brine (50 mL), filtered and the solvent removed under reduced pressure. The light yellow oil was allowed to slowly crystallise at RT, cooled to 0 °C and triturated with Hexane, to give a low melting, colourless, crystalline solid (525 mg, 75%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.71 (1H, s, CH), 7.54 (2H, dd, *J* 6.1, 7.6, PhC3H/PhC5H), 7.45-7.43 (3H, m,

PhC2H/PhC4H/PhC6H), 3.90 (4H, s, CH$_2$). ^{13}C NMR (75 MHz, CDCl$_3$): 165.0 (C=O), 133.8 (PhC1), 130.4 (PhC4), 128.8 (PhC2/PhC6), 126.7 (PhC3), 91.8 (CH), 25.0 (CH$_2$). HRMS: 381.9282 [M+NH$_4$]$^+$ ([C$_{11}$H$_{14}$^{81}Br$_2$O$_4$N]$^+$ requires 381.9290), 386.8836 [M+Na]$^+$ ([C$_{11}$H$_{10}$^{79}Br^{81}BrO$_4$Na]$^+$ requires 386.8843), 402.8574 [M+K]$^+$ ([C$_{11}$H$_{10}$^{79}Br^{81}Br O$_4$K]$^+$ requires 402.8583). IR: 1763 cm^{-1} (C=O).
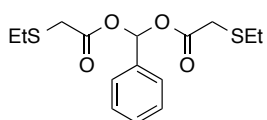
α,α-Di((ethylthio)acetyl) toluene (79). Benzyl 1,1-di(bromoacetate) (488 mg, 1.3 mmol) was dissolved in MeCN (6.3 mL) and NEt$_3$ (725 μL, 5.2 mmol) was added followed by EtSH (1.97 mL, 26.6 mmol). The mixture was stirred at RT for 3 h and the solvent removed under reduced pressure. The crude material was redissolved in CH$_2$Cl$_2$ and evaporated twice to remove all traces of EtSH before purification by flash chromatography (10% EtOAc in hexane) yielding the product as a colourless liquid (140 mg, 33%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.75 (1H, s, CH), 7.58-7.55 (2H, m, PhC3H/PhC5H), 7.45-7.43 (3H, m, PhC2H/PhC4H/PhC6H), 3.31 (4H, s, COCH$_2$), 2.66 (4H, q, J 7.4, CH$_2$CH$_3$), 1.37 (6H, t, J 7.4, CH$_2$CH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 168.3 (C=O), 134.8, 130.0, 128.7 (PhC2/PhC4/PhC6), 126.8 (PhC3/PhC5), 90.7 (CH), 33.0 (COCH$_2$), 26.5 (CH$_2$CH$_3$), 14.1 (CH$_2$CH$_3$). HRMS: 346.1163 [M+NH$_4$]$^+$ ([C$_{15}$H$_{24}$O$_4$S$_2$N]$^+$ requires 346.1147), 679.1588 [2M+Na]$^+$ ([C$_{30}$H$_{40}$O$_8$S$_4$Na]$^+$ requires 679.1504). IR: 1746 cm^{-1} (C=O).
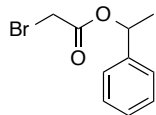
Benzyl 1-methyl-1-bromoacetate (83). NaHCO$_3$ (5.2 g, 62 mmol) was suspended in MeCN (100 mL) and cooled to 0 °C. 1-Phenyl ethanol (2.5 mL, 20.7 mmol) was added followed by bromoacetyl bromide (4.5 mL, 52 mmol), the mixture stirred at 0 °C for 30 min, warmed up to RT and stirred for a further 2 h. The suspension was filtered, the solvent removed under reduced pressure and the residue purified by flash chromatography (12% EtOAc in hexane) yielding the product as a colourless liquid (4.11 g, 82%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.40-7.32 (5H, m, PhCH), 5.96 (1H, q, J 6.6, CH), 3.86 (2H, s, CH$_2$), 1.61 (3H, d, J 6.6, CH$_3$); ^{13}C NMR (75 MHz, CDCl$_3$): 166.4 (C=O), 140.7, 128.5, 128.2 (PhC2/PhC4/PhC6), 126.2 (PhC3/PhC5), 74.4 (CH), 26.2 (CH$_2$), 21.9 (CH$_3$). HRMS: 264.9834 [M+Na]$^+$ ([C$_{10}$H$_{11}$^{79}BrO$_2$Na]$^+$ requires 264.9840). IR: 1731 cm^{-1} (C=O).
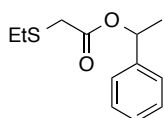
Benzyl 1-methyl-1-(ethylthio)acetate (85). (1.0 g, 4.1 mmol) was dissolved in MeCN (20 mL) and EtSH (3.1 mL, 41.1 mmol) was added followed by NEt$_3$ (1.14 mL, 8.2 mmol) were added. The solution was stirred at RT for 3 h and the solvent removed under reduced pressure. The crude

material was redissolved in CH$_2$Cl$_2$ and evaporated twice to remove all traces of EtSH. The crude material was purified by flash chromatography (10% EtOAc in hexane) to yield the product as a colourless liquid (853 mg, 94%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.32-7.21 (5H, m, PhC_H_), 5.84 (1H, q, _J_ 6.6, C_H_), 3.22 (2H, s, COC_H$_2$_), 2.54 (2H, q, _J_ 7.4, C_H$_2$_CH$_3$), 1.50 (3H, d, _J_ 6.6, CHC_H$_3$_), 1.16 (3H, t, _J_ 7.4, CH$_2$C_H$_3$_). ^{13}C NMR (75 MHz, CDCl$_3$): 168.8 (C=O), 141.3, 128.5, 128.0 (PhC2/PhC4/PhC6), 126.1 (PhC3/PhC5), 73.3 (CH), 33.5 (COCH$_2$), 26.5 (_C_H$_2$CH$_3$), 22.1 (CH_C_H$_3$), 14.1 (CH$_2$_C_H$_3$). HRMS: 225.0945 [M+H]$^+$ ([C$_{12}$H$_{17}$O$_2$S]$^+$ requires 225.0949), 242.1211 [M+NH$_4$]$^+$ ([C$_{12}$H$_{20}$NO$_2$S]$^+$ requires 242.1215), 247.0766 [M+Na]$^+$ ([C$_{12}$H$_{16}$O$_2$SNa]$^+$ requires 247.0769), 471.1638 [2M+Na]$^+$ ([C$_{24}$H$_{32}$O$_4$S2Na]$^+$ requires 471.1640). IR: 1724 cm^{-1} (C=O).
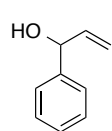
1-Phenylprop-2-en-1-ol (87). (Štambaský _et al._, 2008) Benzaldehyde (4.86 mL, 47.8 mmol) was dissolved in dry THF (160 mL) and cooled to 0 °C. Vinylmagnesium bromide (82 mL, 57.4 mmol) was added dropwise and the mixture stirred at this temperature for 4 h. The suspension was quenched with sat. aq. NH$_4$Cl (250 mL), extracted with EtOAc (2×250 mL), the extracts combined, washed with brine (200 mL), dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The crude material was purified by flash chromatography (10-15% EtOAC in hexane) to give the product as a light yellow liquid (5.2 g, 81%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.39-7.34 (4H, m, PhC_H_), 7.30-7.27 (1H, m, PhC_H_), 6.06 (1H, ddd, _J_ 6.1, 10.3, 17.0, _H_C=CH$_2$), 5.36 (1H, dt, J 1.3, 17.0, HC=C_H$_a$_CH$_b$), 5.21 (1H, m, CH), 5.20 (1H, dt, _J_ 1.3, 10.3, HC=CH$_a$_H$_b$_), 1.93 (1H, d, _J_ 3.7, OH); ^{13}C NMR (75 MHz, CDCl$_3$): 140.2 (_H_C=CH$_2$), 128.6, 128.0, 127.8, 126.2 (Ph), 115.1 (HC=_C_H$_2$), 75.4 (CH); HRMS: 291.1347 [M+Na]$^+$ ([C$_{18}$H$_{20}$O$_2$Na]$^+$ requires 291.1361). IR: 1642 cm^{-1} (C=C), 3357 cm^{-1} (br, OH).

Kinetic resolution of 1-phenylprop-2-en-1-ol. (Štambaský _et al._, 2008) 1-Phenylprop-2-en-1-ol (5.0 g, 37.3 mmol) was dissolved in dry PhMe (320 mL) and Novozyme 435 (985 mg) was added followed by 4 Å molecular sieve powder (5.0 g) and isopropenyl acetate (17.3 mL, 157 mmol). The suspension was stirred vigorously at 40 °C for 16 h, filtered and the solvent removed under reduced pressure. The residue was purified by flash chromatography (0-15% EtOAc in hexane) yielding (S)-1-phenylprop-2-en-1-ol (1.67 g, 33%, 66% _ee_) and (R)-1-phenylprop-2-enyl acetate (2.58 g, 39%) as colourless liquids.
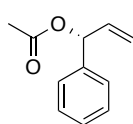
(R)-1-Phenylprop-2-enyl acetate ((R)-90). ^{1}H NMR (500 MHz, CDCl$_3$): 7.40-7.27 (4H, m, PhCH), 7.26-7.21 (1H, m, PhCH), 6.26 (1H, dt, J 5.9, 1.3, CH), 6.01 (1H, ddd, J 5.9, 10.5, 17.1, H$\underline{C}$=CH$_2$), 5.29 (1H, dt, J 17.1, 1.3, HC=C$\underline{H}_a$H$_b$), 5.25 (1H, dt, J 10.5, 1.3, HC=CH$_a\underline{H}_b$), 2.21 (3H, s, CH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 169.4 (C=O), 138.9, 128.6, 128.2, 127.1 (Ph), 136.3 (H$\underline{C}$=CH$_2$), 117.0 (HC=$\underline{C}$H$_2$), 76.2 (CH), 21.2 (CH$_3$). IR: 1494 cm^{-1} (C=O)
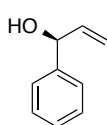
(S)-1-Phenylprop-2-en-1-ol ((S)-87). ^{1}H NMR (500 MHz, CDCl$_3$): 7.39-7.34 (4H, m, PhCH), 7.30-7.27 (1H, m, PhCH), 6.06 (1H, ddd, J 6.1, 10.3, 17.0, H$\underline{C}$=CH$_2$), 5.36 (1H, dt, J 17.0, 1.3, HC=C$\underline{H}_a$H$_b$), 5.21 (1H, m, CH), 5.20 (1H, dt, J 10.3, 1.3, HC=CH$_a\underline{H}_b$), 1.93 (1H, d, J 3.7, OH); ^{13}C NMR (75 MHz, CDCl$_3$): 140.2 (H$\underline{C}$=CH$_2$), 128.6, 128.0, 127.8, 126.2 (Ph), 115.1 (HC=$\underline{C}$H$_2$), 75.4 (CH). HRMS: 133.0652 [M-H]$^-$ ([C$_9$H$_9$O]$^-$ requires 133.0653). IR: 1642 cm^{-1} (C=C), 3352 cm^{-1} (br, OH).
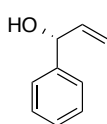
(R)-1-Phenylprop-2-en-1-ol ((R)-87). (Štambaský *et al.*, 2008) (R)-1-Phenylprop-2-enyl acetate (2.58 g, 14.6 mmol) was cooled to 0 °C and a solution of KOH (858 mg, 15.3 mmol) in MeOH (12 mL) was added dropwise. The mixture was heated to 50 °C and stirred for 2 h. The solution was diluted with brine (25 mL) and extracted with EtOAc (3×40 mL), the organic extracts combined, dried over Na$_2$SO$_4$, filtered and the solvent removed to give the product as a yellow liquid (1.76 g, 90%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.39-7.34 (4H, m, PhCH), 7.30-7.27 (1H, m, PhCH), 6.06 (1H, ddd, J 6.1, 10.3, 17.0, H$\underline{C}$=CH$_2$), 5.36 (1H, dt, J 17.0, 1.3, HC=C$\underline{H}_a$H$_b$), 5.21 (1H, m, CH), 5.20 (1H, dt, J 10.3, 1.3, HC=CH$_a\underline{H}_b$), 1.93 (1H, d, J 3.7, OH). ^{13}C NMR (75 MHz, CDCl$_3$): 140.2 (H$\underline{C}$=CH$_2$), 128.6, 128.0, 127.8, 126.2 (Ph), 115.1 (HC=$\underline{C}$H$_2$), 75.4 (CH). IR: 1642 cm^{-1} (C=C), 3350 cm^{-1} (br, OH).

General procedure for the synthesis of Mosher's esters
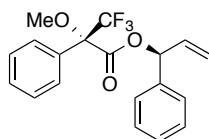
(S)-1-Phenylprop-2-en-1-ol S-MTPA ester (89). (Mallavadhani and Rao, 1994) (S)-1-Phenylprop-2-en-1-ol (5.97 μL, 44.5 μmol) was dissolved in dry CH$_2$Cl$_2$ (1.5 mL) and cooled to 0 °C. DMAP (10.9 mg, 89 μmol) and Et$_3$N (8.68 μL, 62.3 μmol) were added followed by (S)-MTPACl (10 μL, 53.4 μmol). The mixture was stirred for 10 min, quenched with sat. aq. NH$_4$Cl (1.5 mL) and diluted with Et$_2$O (2 mL). The organic layer was washed with water (1.5 mL) and brine (1.5 mL), dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The crude was purified by flash chromatography (0-5% EtOAc in hexane), yielding the product as a colourless oil. ^{1}H NMR (500 MHz,

CDCl$_3$): 7.43-7.31 (10H, m, PhC$\underline{H}$), 6.49 (1H, dt, J 5.9, 1.3, CH), 5.96 (1H, ddd, J 5.9, 10.5, 17.0, $\underline{H}$C=CH$_2$), 5.29 (1H, dt, J 17.0, 1.3, HC=C$\underline{H}_a$H$_b$), 5.26 (1H, dt, J 10.5, 1.3, HC=CH$_a\underline{H}_b$), 3.47 (3H, s, OCH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 165.6 (C=O), 137.6 (CF$_3$), 134.9 ($\underline{H}$C=CH$_2$), 132.3, 129.5, 128.7, 128.3, 127.5, 127.4, 125.3 (Ph), 117.8 (HC=$\underline{C}$H$_2$), 78.3 (CH), 77.2 ($\underline{C}$CF$_3$), 55.4 (CH$_3$). HRMS: 368.1473 [M+NH$_4$]$^+$ ([C$_{19}$H$_{21}$NF$_3$O$_3$]$^+$ requires 368.1473), 373.1028 [M+Na]$^+$ ([C$_{19}$H$_{17}$F$_3$O$_3$Na]$^+$ requires 373.1028), 723.2158 [2M+Na]$^+$ ([C$_{38}$H$_{34}$F$_6$O$_6$Na]$^+$ requires 723.2157). IR: 1746 cm^{-1} (C=O), 1411 cm^{-1} (C=C).
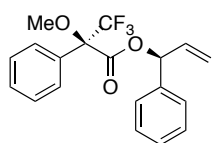
(S)-1-Phenylprop-2-en-1-ol R-MTPA ester (88). ^{1}H NMR (500 MHz, CDCl$_3$): 7.42-7.24 (10H, m, PhC$\underline{H}$), 6.45 (1H, dt, J 6.3, 1.2, CH), 6.04 (1H, ddd, J 6.3, 10.4, 17.1, $\underline{H}$C=CH$_2$), 5.39 (1H, dt, J 17.1, 1.2, HC=C$\underline{H}_a$H$_b$), 5.31 (1H, dt, J 10.4, 1.2, HC=CH$_a\underline{H}_b$), 3.54 (3H, s, OCH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 165.5 (C=O), 137.6 (CF$_3$), 135.2 ($\underline{H}$C=CH$_2$), 132.2, 129.5, 128.5, 128.5, 128.3, 127.4, 127.1 (Ph), 118.4 (HC=$\underline{C}$H$_2$), 78.6 (CH), 77.2 ($\underline{C}$CF$_3$), 55.5 (CH$_3$). HRMS: 368.1455 [M+NH$_4$]$^+$ ([C$_{19}$H$_{21}$NF$_3$O$_3$]$^+$ requires 368.1473), 373.1009 [M+Na]$^+$ ([C$_{19}$H$_{17}$F$_3$O$_3$Na]$^+$ requires 373.1028), 723.2148 [2M+Na]$^+$ ([C$_{38}$H$_{34}$F$_6$O$_6$Na]$^+$ requires 723.2157). IR: 1746 cm^{-1} (C=O), 1411 cm^{-1} (C=C).
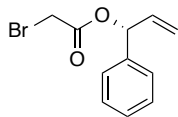
(R)-1-Phenylprop-2-enyl bromoacetate ((R)-92). (R)-1-Phenylprop-2-en-1-ol (600 mg, 4.47 mmol) was dissolved in dry CH$_2$Cl$_2$ (45 mL) and the solution cooled to 0 °C. Pyridine (720 μL, 8.94 mmol) was added followed by bromoacteyl bromide (779 μL, 8.94 mmol), forming a white suspension. The mixture was stirred at 0 °C for 2 h and quenched with sat. aq. NH$_4$Cl (100 mL). The aqueous layer was extracted with CH$_2$Cl$_2$ (3×100 mL), the organic extracts combined, washed with sat. aq. CuSO$_4$ (100 mL), dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The crude material was purified by flash chromatography (0-8% EtOAc in hexane), yielding the product as a colourless liquid (936 mg, 82%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.38-7.36 (4H, m, PhC$\underline{H}$), 7.35-7.31 (1H, m, PhC$\underline{H}$), 6.30 (1H, dt, J 6.0, 1.2, CH), 6.02 (1H, ddd, J 6.0, 10.5, 17.1, $\underline{H}$C=CH$_2$), 5.36 (1H, dt, J 17.1, 1.2, HC=C$\underline{H}_a$H$_b$), 5.30 (1H, dt, J 10.5, 1.2, HC=CH$_a\underline{H}_b$), 3.89 (1H, d, J 12.4, CH$_a\underline{H}_b$Br), 3.86 (1H, d, J 12.4, C$\underline{H}_a$H$_b$Br). ^{13}C NMR (75 MHz, CDCl$_3$): 166.2 (C=O), 138.2 ($\underline{H}$C=CH$_2$), 135.1 (PhC1), 128.7, 128.5, 127.2 (PhC2/PhC3/PhC4), 117.8 (HC=$\underline{C}$H$_2$), 78.1 (CH), 26.2 (CH$_2$); HRMS: 276.9834 [M+Na]$^+$ ([C$_{11}$H$_{11}$^{79}BrO$_2$Na]$^+$ requires 276.9834), 278.9812 [M+Na]$^+$ ([C$_{11}$H$_{11}$^{81}BrO$_2$Na]$^+$ requires 278.9814). IR: 1734 cm^{-1} (C=O), 1644 cm^{-1} (C=C).

168
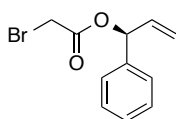
(S)-1-Phenylprop-2-enyl bromoacetate ((S)-92). (S)-1-Phenylprop-2-en-1-ol (3.93 g, 29.3 mmol) was dissolved in dry CH_2Cl_2 (30 mL) and the solution cooled to 0 °C. Pyridine (4.72 mL, 58.6 mmol) was added followed by bromoacetyl bromide (5.10 mL, 58.6 mmol), forming a yellow suspension. The mixture was stirred at 0 °C for 2 h and quenched in sat. aq. NH_4Cl (100 mL). The aqueous layer was extracted with CH_2Cl_2 (3×100 mL), the organic extracts combined, washed with sat. aq. $CuSO_4$ (100 mL) and brine (100 mL), dried over Na_2SO_4, filtered and the solvent removed under reduced pressure. The crude was purified by flash chromatography (0-8% EtOAc in hexane), yielding the product as a yellow liquid (6.18 g, 83%). ^1H NMR (500 MHz, $CDCl_3$): 7.38-7.36 (4H, m, PhC$\underline{H}$), 7.35-7.32 (1H, m, PhC$\underline{H}$), 6.29 (1H, dt, J 6.0, 1.2, CH), 6.02 (1H, ddd, J 6.0, 10.5, 17.1, $\underline{H}C=CH_2$), 5.36 (1H, dt, J 17.1, 1.2, HC=C$\underline{H_a}H_b$), 5.30 (1H, dt, J 10.5, 1.2, HC=CH$_a\underline{H_b}$), 3.89 (1H, d, J 12.4, C$\underline{H_a}H_b$Br), 3.86 (1H, d, J 12.4, CH$_a\underline{H_b}$Br). ^{13}C NMR (75 MHz, $CDCl_3$): 166.2 (C=O), 138.0 ($\underline{H}C=CH_2$), 135.4 (PhC1), 128.7, 128.5, 127.2 (PhC2/C3/C4), 117.8 (HC=$\underline{C}H_2$), 78.1 (CH), 26.0 (CH_2). HRMS: 276.9831 [M+Na]$^+$ ([$C_{11}H_{11}{}^{79}BrO_2Na$]$^+$ requires 276.9834). IR: 1733 (C=O).
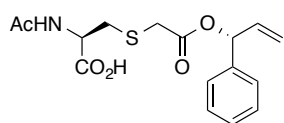
N-acetyl-S-[(R)-1-phenylprop-2-enylacetate]-L-cysteine (93). (R)-1-Phenylprop-2-enyl bromoacetate (200 mg, 0.78 mmol) was dissolved in MeCN (4 mL) and N-acetyl-L-cysteine (140 mg, 0.86 mmol) was added followed by NEt$_3$ (240 μL, 1.72 mmol). The mixture was stirred for 1h, quenched into aq. 1M HCl (25 mL), extracted with EtOAc (3×30 mL), the organic extracts combined, dried over Na_2SO_4, filtered and the solvent removed under reduced pressure. The crude was purified by flash chromatography (1% AcOH and 5% MeOH in CH_2Cl_2) yielding the product as a colourless oil (225 mg, 86%). ^1H NMR (500 MHz, $CDCl_3$): 7.38-7.36 (4H, m, PhC$\underline{H}$), 7.34-7.32 (1H, m, PhC$\underline{H}$), 6.91 (1H, d, J 6.7, NH), 6.27 (1H, d, J 6.1, CH), 6.02 (1H, ddd, J 6.1, 10.4, 17.0, C$\underline{H}$=CH$_2$), 5.33 (1H, dt, J 17.0, 1.2, HC=C$\underline{H_a}H_b$) 5.29 (1H, dt, J 10.4, 1.2, HC=CH$_a\underline{H_b}$), 4.74 (1H, ddd, J 4.7, 6.0, 6.7, HNC$\underline{H}$), 3.40 (1H, d, J 15.7, C$\underline{H_a}H_b$), 3.35 (1H, d, J 15.7, CH$_a\underline{H_b}$), 3.14 (1H, dd, J 6.0, 14.2, Cys-C$\underline{H_a}H_b$), 3.05 (1H, dd, J 4.7, 14.2, Cys-CH$_a\underline{H_b}$), 2.04 (3H, s, CH$_3$). ^{13}C NMR (75 MHz, $CDCl_3$): 176.5 (CO$_2$H), 173.4 (CONH), 170.1 ($\underline{C}O_2CH$), 138.1, 128.7, 128.5, 127.2 (Ph), 135.5 ($\underline{H}C=CH_2$), 117.8 (HC=$\underline{C}H_2$), 77.9 (CH), 52.4 (cys-CH), 34.9 (COCH$_2$), 34.5 (cys-CH$_2$), 22.8 (CH$_3$). HRMS: 338.10568 [M+H]$^+$ ([$C_{16}H_{20}NO_5S$]$^+$ requires 338.1062), 360.08761 [M+N]$^+$ ([$C_{16}H_{19}NO_5SNa$]$^+$ requires 360.088), 382.06967 [M-H+2Na]$^+$ ([$C_{16}H_{18}NO_5SNa_2$]$^+$

requires 382.0701), 697.18562 [2M+Na]$^+$ ([C$_{32}$H$_{38}$NO$_{10}$S$_2$Na]$^+$ requires 683.1834). IR: 1713 cm^{-1} (C=O), 1627 cm^{-1} (C=O), 3306-2929 cm^{-1} (br, OH, NH, CH).
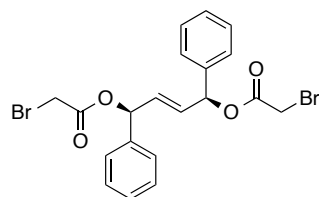
2-Butene-1,4-diol di(bromoacetate) (94). (S)-1-Phenylprop-2-enyl bromoacetate (301 μL, 1.60 mmol) was dissolved in dry CH$_2$Cl$_2$ (7 mL) and the solution flushed with N$_2$. Hoveyda-Grubbs II catalyst (50 mg, 0.08 mmol, 5 mol%) was added, the mixture heated to 45 °C and stirred at this temperature for 16 h. The solution was cooled to RT, a second aliquot of Hoyveda-Grubbs II catalyst was added (50 mg, 0.08 mmol, 5 mol%), the solution re-heated to 45 °C and stirred for a further 24 h. The mixture was cooled to RT, a solution of potassium 2-isocyanoacetate (45 mg, 22 mol%) in MeOH (1 mL) was added and the mixture stirred for 1 h. The solvent was removed under reduced pressure and the crude purified by flash chromatography (5-10% EtOAc in hexane), yielding the product as an off-white solid (172 mg, 45%). (500 MHz, CDCl$_3$): 7.33-7.21 (10H, m, PhCH), 6.26 (2H, dd, J 1.4, 3.1, CH), 5.93 (2H, dd, J 1.4, 3.1, HC=CH), 3.71 (4H, s, CH$_2$). ^{13}C NMR (75 MHz, CDCl$_3$): 166.1 (C=O), 137.6 (PhC1), 130.6 (HC=CH), 128.8, 128.7, 128.3 (PhC2/PhC3/PhC4), 76.8 (CH), 25.9 (CH$_2$). HRMS: 504.9481 [M+Na]$^+$ ([C$_{20}$H$_{18}$O$_4$^{79}Br^{81}BrNa]$^+$ requires 504.9444). IR: 1741 cm^{-1} (C=O), 1675 cm^{-1} (C=C), 1601 cm^{-1}, 1584 cm^{-1}, 1492 cm^{-1} (ArC=C), 3029 cm^{-1}, 2982-2908 cm^{-1} (C-H).
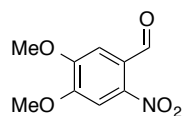
4,5-Dimethoxy-2-nitrobenzaldehyde (83). Conc. HNO$_3$ (70%, 80 mL, 903 mmol) was cooled to 0 °C and 3,4-dimethoxybenzaldehyde (10.0 g, 60.2 mmol) was added portionwise. The solution was allowed to slowly warm up to RT over 3 h, poured into a water/ice slurry (600 mL), the precipitate collected by filtration and washed with ice cold EtOH, to give the product as a yellow solid (10.9 g, 85%). ^{1}H NMR (500 MHz, CDCl$_3$): 10.46 (s, 1H, COH), 7.62 (s, 1H, PhC3H), 7.43 (s, 1H, PhC6H), 4.04 (s, 3H, OCH$_3$), 4.03 (s, 3H, OCH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 187.7 (C=O), 153.3 (PhC5), 152.5 (PhC4), 143.9 (PhC2), 125.6 (PhC1), 109.8 (PhC6), 107.2 (PhC3), 56.8 (OCH$_3$), 56.8 (OCH$_3$). HRMS: 212.0554 [M+Na]$^+$ ([C$_9$H$_{10}$NO$_5$]$^+$ requires 212.0554). IR: 1756 cm^{-1} (C=O), 2905 cm^{-1}, 2839 cm^{-1}, 3095-3064 cm^{-1} (C-H), 1509 cm^{-1}, 1333 cm^{-1} (NO$_2$).
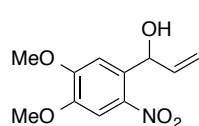
1-(4,5-Dimethoxy-2-nitrobenzyl)-prop-2-en-1-ol (97). 4,5-Dimethoxy-2-nitrobenzaldehyde (2.0 g, 9.47 mmol) was dissolved in dry THF (50 mL) and cooled to -78 °C. Vinylmagnesium bromide (1M solution in THF, 14.2 mL, 14.2 mmol) was added slowly and the solution

stirred and allowed to warm up to RT slowly over 16 h. The solution was poured slowly into sat. aq. NH$_4$Cl/ice slurry (100 mL), extracted with EtOAc (3×100 mL), the extracts combined and washed with brine (100 mL), dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The crude mixture was purified by flash chromatography (20% EtOAc in hexane) yielding the product as a yellow fluffy solid (1.58 g, 70%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.58 (1H, s, PhC3H), 7.21 (1H, s, PhC6H), 6.08 (1H, ddd, J 17.2, 10.5, 4.7, C<u>H</u>=CH$_2$), 5.93 (1H, d, J 4.7, CH), 5.43 (1H, dt, J 17.2, 1.3, HC=C<u>H</u>$_a$H$_b$), 5.25 (1H, dt, J 10.5, 1.3, HC=CH$_a$<u>H</u>$_b$), 3.99 (3H, s, OCH$_3$), 3.95 (3H, s, OCH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 153.6 (PhC5), 148.0 (PhC4), 140.1 (PhC2), 138.1 (H<u>C</u>=CH$_2$), 133.1 (PhC1), 115.8 (HC=<u>C</u>H$_2$), 109.8 (PhC6), 107.9 (PhC3), 69.8 (CH), 56.4 (OCH$_3$), 56.4 (OCH$_3$). HRMS: 262.0864 [M+H]$^+$ ([C$_{11}$H$_{14}$NO$_5$]$^+$ requires 240.0867), 262.0689 [M+Na]$^+$ ([C$_{11}$H$_{13}$NO$_5$Na]$^+$ requires 262.0686), 501.1486 [2M+Na]$^+$ ([C$_{22}$H$_{26}$N$_2$NaO$_{10}$]$^+$ requires 501.1480). IR: 3533 cm^{-1} (OH), 3110-2840 cm^{-1} (C-H), 1613 cm^{-1}, 1502 cm^{-1} (C=C).
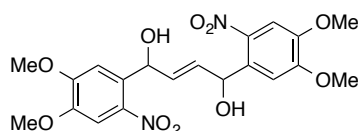
1,4-Di(4,5-dimethoxy-2-nitrobenzyl)-but-2-en-1,4-diol (98). 1-(4,5-Dimethoxy-2-nitrobenzyl)-prop-2-en-1-ol (500 mg, 2.1 mmol) was dissolved in dry CH$_2$Cl$_2$ (7 mL), the solution flushed with N$_2$ and Hoveyda-Grubbs II catalyst (65.5 mg, 0.10 mmol, 5 mol%) was added. The mixture was heated at 45 °C for 16 h. The solution was cooled to RT, a solution of potassium isocyanoacetate (65 mg, 25 mol%) in MeOH (2.3 mL) was added and the mixture stirred for 1 h, forming a heavy precipitate. The precipitate was collected by filtration and washed with methanol, yielding the product as an off-white solid (440 mg, 93%, 45:55 mixture of diastereomers). ^{1}H NMR (500 MHz, d$_6$-DMSO): 7.54 (1.8H, s, PhC3H), 7,53 (2.2H, s, PhC3H'), 7.22 (1.8H, s, PhC6H), 7.21 (2.2H, s, PhC6H'), 5.86-5.83 (8H, m, C2H, C2H', C1H, OH), 5.72 (4H, br s, C1H', OH'), 3.86 (5.4H, s, OMe), 3.86 (5.4H, s, OMe), 3.85 (13.2H, s, OMe', OMe'). ^{13}C NMR (75 MHz, d$_6$-DMSO): 152.9 (PhC5), 152.9 (PhC5'), 147.2 (PhC4/PhC4'), 139.5 (PhC2/PhC2'), 134.3 (C2), 134.2 (C2'), 131.3 (PhC1), 131.3 (PhC1'), 109.6 (PhC6), 109.4 (PhC6'), 107.4 (PhC3), 107.4 (PhC3'), 67.1 (C1), 67.0 (C1'), 56.0 (CH$_3$/CH$_3$'), 55.9 (CH$_3$/CH$_3$'). HRMS: 473.1182 [M+Na]$^+$ ([C$_{20}$H$_{22}$N$_2$NaO$_{10}$]$^+$ requires 473.1167), 923.2450 [2M+Na]$^+$ ([C$_{40}$H$_{44}$N$_4$NaO$_{20}$]$^+$ requires 923.2442). IR: 3512 cm^{-1}, 3470 cm^{-1} (OH), 1514 cm^{-1}, 1327 cm^{-1} (NO$_2$).
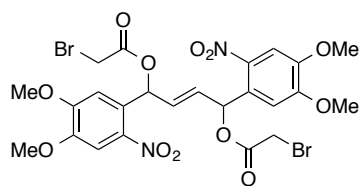
1,4-Di-*O*-bromoacetyl-1,4-di(4,5-dimethoxy-2-nitrobenzyl)-but-2-en-1,4-diol (95). 1,4-Di(4,5-dimethoxy-2-nitrobenzyl)-but-2-en-1,4-diol (100 mg, 0.22 mmol) was suspended in CH_2Cl_2, the suspension cooled to 0 °C and bromoacetyl bromide (38 μL, 0.44 mmol) was added followed by pyridine (36 μL, 0.44 mmol). The solution was stirred at 0 °C for 1 h, quenched into H_2O (5 mL), extracted with CH_2Cl_2 (3×5 mL), the extracts combined and washed with sat. aq. $CuSO_4$ (2×5 mL) and the solvent removed under reduced pressure. The crude product was purified by flash chromatography (20%-33% EtOAc in hexane), yielding the product as a light brown powder (57 mg, 37%, 9:1 ratio of diastereomers). ^1H NMR (500 MHz, CDCl$_3$): 7.63 (2H, s, PhC3H'), 7.58 (2H, s, PhC3H), 7.05 (4H, dd, *J* 2.7, 1.0, C1H/C1H'), 7.04 (4H, s, PhC6H/PhC6H'), 6.09 (2H, dd, *J* 1.0, 2.7, C2H'), 6.06 (2H, dd, *J* 1.0, 2.7, C2H), 4.01 (s, 6H, OCH$_3$'), 4.00 (s, 6H, OCH$_3$), 3.96 (s, 6H, OCH$_3$'), 3.94 (s, 6H, OCH$_3$), 3.92 (d, 2H, *J* 11.9 C$\underline{H}_aH_bBr$), 3.88 (d, 2H, *J* 11.9 C$\underline{H}_aH_bBr'$), 3.84 (d, 2H, *J* 11.9 CH$_a\underline{H}_bBr$), 3.80 (d, 2H, *J* 11.9 CH$_a\underline{H}_bBr'$). ^{13}C NMR (75 MHz, CDCl$_3$): 165.5 (C=O), 153.5 (PhC5), 148.6 (PhC4), 140.3 (PhC2), 130.0 (C2), 127.7 (PhC1), 109.7 (PhC6), 107.8 (PhC3), 71.8 (C1), 56.7 (CH$_3$), 56.4 (CH$_3$), 25.5 (CH$_2$). HRMS: 710.0043 [M+NH$_4$]$^+$ ([C$_{24}$H$_{28}^{79}$Br^{81}BrN$_3$O$_{12}$]$^+$ 710.0014), 708.0060 [M+NH$_4$]$^+$ ([C$_{24}$H$_{28}^{79}$Br^{79}BrN$_3$O$_{12}$]$^+$ requires 708.0035), 730.9324 [M+K]$^+$ ([C$_{24}$H$_{24}^{79}$Br^{81}BrN$_2$O$_{12}$K]$^+$ requires 730.9308), 728.9345 [M+K]$^+$ ([C$_{24}$H$_{24}^{79}$Br^{79}BrN$_2$O$_{12}$K]$^+$ requires 728.9328). IR: 2844 (C-H), 1743 cm^{-1} (C=O), 1515 cm^{-1}, 1330 cm^{-1} (NO$_2$).
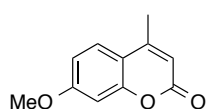
7-Methoxy-4-methylcoumarin (100). (Fournier *et al.*, 2013) Conc. H_2SO_4 (120 mL) was cooled to 0 °C and 3-methoxyphenol (17.5 mL, 161 mmol) was added dropwise with vigorous stirring. The solution was stirred for 10 min and ethyl acetoacetate (22.4 mL, 177 mmol) was added dropwise, keeping the temperature below 6 °C. The mixture was stirred at 0 °C for 3 h, diluted into a H_2O/ice slurry (600 mL), filtered under suction, dried on the filter and on high vacuum overnight, leaving the product as a white solid (37.8 g, quantitative). ^1H NMR (500 MHz, CDCl$_3$): 7.50 (1H, d, *J* 8.7, CmC5H), 6.87 (1H, dd, *J* 2.5, 8.7, CmC6H), 6.83 (1H, d, *J* 2.5, CmC8H), 6.14 (1H, br s, CmC3H), 3.88 (3H, s, OCH$_3$), 2.40 (3H, d, *J* 1.2, CH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 162.7 (CmC7), 161.4 (CmC2), 155.3 (Cm8a), 152.7 (CmC4), 125.6 (CmC5), 113.6, 112.3, 111.9 (CmC3/CmC6/CmC4a), 100.8 (CmC8), 55.8 (OCH$_3$). HRMS: 213.0517 [M+Na]$^+$ ([C$_{11}$H$_{10}$O$_3$Na]$^+$ requires 213.0522),

381.1332 [2M+H]$^+$ ([C$_{22}$H$_{21}$O$_6$]$^+$ requires 381.1333), 403.1151 [2M+Na]$^+$ ([C$_{22}$H$_{20}$O$_6$Na]$^+$ requires 403.1152). IR: 1727 cm^{-1} (C=O), 1606 cm^{-1} (C=C).
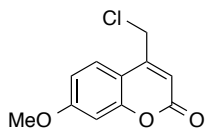
4-Chloromethyl-7-methoxycoumarin (103). (Fournier *et al.*, 2013) Conc. H$_2$SO$_4$ (130 mL) was cooled to 0 °C and 3-methoxyphenol (17.5 mL, 161 mmol) was added dropwise with vigorous stirring. The solution was stirred for 10 min and ethyl 4-chloro acetoacetate (23.9 mL, 177 mmol) was added dropwise, keeping the temperature below 6 °C. The mixture was stirred at 0 °C for 3 hours. The mixture was diluted into a H$_2$O/ice slurry (600 mL), filtered under suction, dried on the filter and on high vacuum overnight, leaving the product as a white flaky solid (21.1 g, 60%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.57 (1H, d, *J* 8.9, CmC5H), 6.90 (1H, dd, *J* 2.5, 8.9, CmC6H), 6.85 (1H, d, *J* 2.5, CmC8H), 6.40 (1H, br s, CmC3H), 4.63 (2H, d, *J* 2.5, CH$_2$), 3.89 (3H, s, OCH$_3$). ^{13}C NMR (75 MHz, CDCl3): 163.0 (CmC7), 160.8 (CmC2), 155.8 (Cm8a), 149.6 (CmC4), 125.2 (CmC5), 112.7, 112.6 (CmC3/CmC6), 110.8 (CmC4a), 101.0 (CmC8), 55.8 (CH$_2$), 41.3 (OCH$_3$). HRMS: 225.0314 [M+H]$^+$ ([C$_{11}$H$_{10}$^{35}ClO$_3$]$^+$ requires 225.0313), 242.0573 [M+NH4]$^+$ ([C$_{11}$H$_{13}$^{35}ClNO$_3$]$^+$ requires 242.0578), 247.0130 [M+Na]$^+$ ([C$_{11}$H$_9$^{35}ClNaO$_3$]$^+$ requires 247.0132). IR: 1720 cm^{-1} (C=O), 1608 cm^{-1} (C=C).
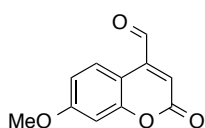
4-Formyl-7-methoxycoumarin

From SeO$_2$ oxidation (82). (Huynh *et al.*, 2012) 7-Methoxy-4-methylcoumarin (4.5 g, 23.6 mmol) was suspended in xylenes (60 mL) and the mixture heated under reflux. Once the starting material was in solution, SeO$_2$ (3.94 g, 35.5 mmol) was added portionwise and the dark solution heated under reflux for 16 h. The solution was filtered hot to remove the black selenium and the solvent removed under reduced pressure. The dark yellow crude product was purified by flash chromatography (20-100% EtOAc in hexane) yielding the product as dark yellow fluffy solid (2.48 g, 51%). ^{1}H NMR (500 MHz, CDCl$_3$): 10.07 (1H, s, COH), 8.50 (1H, d, *J* 9.0, CmC5H), 6.92 (1H, dd, *J* 2.5, 9.0, CmC6H), 6.87 (1H, d, *J* 2.5, CmC8H), 6.71 (1H, s, CmC3H), 3.90 (3H, s, OCH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$): 191.7 (COH), 163.4 (CmC7), 160.7 (CmC2), 156.5 (Cm8a), 143.8 (CmC4), 127.3 (CmC5), 122.1 (CmC3), 113.3 (CmC6), 108.2 (CmC4a), 101.1 (CmC8), 55.8 (OCH$_3$). HRMS: 205.0488 [M+H]$^+$ ([C$_{11}$H$_9$O$_4$]$^+$ requires 205.0495), 227.0311 [M+Na]$^+$ ([C$_{11}$H$_8$NaO$_4$]$^+$ requires 227.0315), 409.0921 [2M+H]$^+$ ([C$_{22}$H$_{17}$O$_8$]$^+$ requires 409.0918). IR: 1722 cm^{-1} (C=O), 1702 cm^{-1} (C=O), 1607 cm^{-1} (C=C), 1592 cm^{-1} (C=C).

From PCC oxidation. Prepared from a modified protocol to that previously described (Pisani *et al.*, 2013). 4-Dydroxymethyl-7-methoxycoumarin (2.2 g, 10.7 mmol) was suspended in dry CH_2Cl_2 (200 mL) and stirred for 15 min. $MgSO_4$ (4.6 g) was added followed by PCC (4.6 g, 10.7 mmol) and the mixture stirred at RT for 16 h. The dark solution was filtered through celite and the solvent removed under reduced pressure. The dark brown residue was purified by flash chromatography (25-60% EtOAc in hexane) yielding the product as a yellow flaky solid (1.06 g, 49%). ^{1}H NMR (500 MHz, CDCl$_3$): 10.07 (1H, s, COH), 8.50 (1H, d, *J* 9.0, CmC5H), 6.92 (1H, dd, *J* 2.5, 9.0, CmC6H), 6.87 (1H, d, *J* 2.5, CmC8H), 6.71 (1H, s, CmC3H), 3.90 (3H, s, OCH$_3$). HRMS: 205.0489 [M+H]$^+$ ([C$_{11}$H$_9$O$_4$]$^+$ requires 205.0495), 409.0916 [2M+H]$^+$ ([C$_{22}$H$_{17}$O$_8$]$^+$ requires 409.0918). IR: 1722 cm^{-1} (C=O), 1703 cm^{-1} (C=O), 1608 cm^{-1} (C=C), 1594 cm^{-1} (C=C).
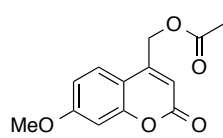
4-Acetoxymethyl-7-methoxycoumarin (107). 4-Chloromethyl-7-methoxycoumarin (5.5 g, 22.3 mmol) was dissolved in AcOH (60 mL) and sodium acetate (5.5 g, 67.0 mmol) was added. The solution was heated under reflux 20 h. The solution was diluted into a H$_2$O/ice slurry (600 mL), the precipitate collected under suction, dried on the filter and then under high vacuum overnight. The product was obtained as a white solid (3.30 g, 64%, 93% purity) and used without further purification. ^{1}H NMR (500 MHz, CDCl$_3$): 7.41 (1H, d, *J* 9.6, CmC5H), 6.87 (1H, dd, *J* 3.5, 9.6, CmC6H), 6.86 (1H, d, *J* 3.5, CmC8H), 6.34 (1H, s, CmC3H), 5.26 (2H, s, CH$_2$), 3.88 (3H, s, OCH$_3$), 2.21 (3H, s, COC$\underline{H}_3$). ^{13}C NMR (75 MHz, CDCl3): 170.1 ($\underline{C}$OCH$_3$), 162.9 (CmC7), 160.9 (CmC2), 155.6 (CmC8a), 149.1 (CmC4), 124.4 (CmC5), 112.7 (CmC6), 110.6 (CmC4a), 110.0 (CmC3), 101.2 (CmC8), 61.2 (CH$_2$), 55.8 (COC$\underline{H}_3$), 20.7 (OCH$_3$). HRMS: 249.0759 [M+H]$^+$ ([C$_{13}$H$_{13}$O$_5$]$^+$ requires 249.0758), 266.1019 [M+NH$_4$]$^+$ ([C$_{13}$H$_{16}$NO$_5$]$^+$ requires 266.1023), 271.0576 [M+Na]$^+$ ([C$_{13}$H$_{12}$NaO$_5$]$^+$ requires 271.0577). IR: 1711 cm^{-1} (C=O), 1605 cm^{-1} (C=C).
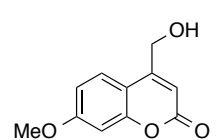
4-Hydroxymethyl-7-methoxycoumarin (106). 4-Acetoxymethyl-7-methoxycoumarin (2.80 g, 12.0 mmol) was dissolved in 3:1 EtOH:conc. HCl (43 mL) and the solution heated under reflux for 45 min. The mixture was allowed to cool to RT and diluted into a H$_2$O/ice slurry (400 mL). The white precipitate formed was collected under suction, dried on the filter and in high vacuum overnight, yielding the product as a white flaky solid (2.4 g, 98%, 93% purity). ^{1}H NMR (500 MHz, CDCl$_3$): 7.43 (1H, d, *J* 9.5, CmC5H), 6.90-6.80 (2H,

m, CmC6H/CmC8H), 6.47 (1H, t, J 1.5, CmC3H), 4.88 (2H, d, J 1.5, CH$_2$), 3.88 (3H, s, OCH$_3$). HRMS: 207.0650 [M+H]$^+$ ([C$_{11}$H$_{11}$O$_4$]$^+$ requires 207.0652). IR: 3380 cm^{-1} (OH), 1682 cm^{-1} (C=O), 1605 cm^{-1} (C=C), 1085 (C-O).
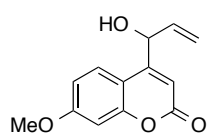
4-(1-Hydroxy-2-propenyl)-7-methoxycoumarin (108). 4-Formyl-7-methoxycoumarin (1.5 g, 7.3 mmol) was suspended in dry THF (24 mL) and cooled to 0 °C. Vinylmagnesium bromide (0.7 M in THF, 12.5 mL, 8.8 mmol) was added dropwise and the mixture stirred at 0 °C for 3 hours. The solution was quenched with sat. aq. NH$_4$Cl (100 mL), the aqueous solution extracted with EtOAc (3×100 mL), the extracts combined and washed with brine (100 mL), dried over Na$_2$SO$_4$, filtered and the solvent removed under reduced pressure. The crude was purified by flash chromatography (33-50% EtOAc in hexane) yielding the product as a yellow gum (569 mg, 34%). ^{1}H NMR (500 MHz, CDCl$_3$): 7.61 (1H, d, J 9.6, CmC5H), 6.83 (2H, m, CmC6H/CmC8H), 6.48 (1H, s, CmC3H), 6.07 (1H, ddd, J 6.1, 10.3, 17.1, C$\underline{H}$=CH$_2$), 5.52 (1H, d, J 17.1, HC=CH$\underline{_a}$H$_b$), 5.47 (1H, t, J 6.1, CH), 5.38 (1H, d, J 10.3, HC=CH$_a$$\underline{H_b}$), 3.81 (3H, s, OCH$_3$), 2.05 (1H, s, OH). ^{13}C NMR (75 MHz, CDCl$_3$): 162.5 (CmC7), 161.8 (CmC2), 155.8, 155.6 (CmC8a/CmC4), 137.0 ($\underline{C}$H=CH$_2$), 125.9 (CmC5), 118.6 (CH=$\underline{C}$H$_2$), 112.4 (CmC6), 110.9 (CmC4a), 109.4 (CmC3), 101.0 (CmC8), 71.0 (CH), 55.7 (OCH$_3$). HRMS: 233.0813 [M+H]$^+$ ([C$_{13}$H$_{13}$O$_4$]$^+$ requires 233.0809), 255.0620 [M+Na]$^+$ ([C$_{13}$H$_{12}$O$_4$Na]$^+$ requires 255.0628), 465.1566 [2M+H]$^+$ ([C$_{26}$H$_{25}$O$_8$]$^+$ requires 465.1544). IR: 3459 cm^{-1} (OH), 1704 cm^{-1} (C=O), 1607 cm^{-1} (C=C).

1,4-Di(7-methoxy-4-methylcoumarinyl)-but-2-en-1,4-diol (109). 4-(1-Hydroxy-2-propenyl)-7-methoxycoumarin (1.0 g, 4.3 mmol) was suspended in dry CH$_2$Cl$_2$ (30 mL), the solution flushed with N$_2$ and Hoveyda-Grubbs II catalyst (135 mg, 0.22 mmol, 5 mol%) was added. The suspension was heated at 45 °C for 36 h and cooled to RT. A solution of potassium isocyanoacetate (135 mg, 20 mol%) in MeOH (3.5 mL) was added and the mixture stirred for 1 h, forming a heavy precipitate. The precipitate was collected by filtration and washed with methanol, yielding the product as an off-white solid (650 mg, 80%, 95% purity, 3:2 mixture of diastereomers). ^{1}H NMR (500 MHz, d$_6$-DMSO): 7.75 (1H, d, J 8.9, CmC5H), 7.69 (1H, d, J 8.9, CmC5H'), 6.98 (1H, d, J 2.1, CmC8H), 6.95 (1H, d, J 2.1, CmC8H'), 6.76 (1H, dd, J 2.1, 8.9, CmC6H), 6.70 (1H, dd, J 2.1, 8.9, CmC6H'), 6.33 (1H, s, CmC3H'), 6.31 (1H, s, CmC3H), 6.06 (1H, br s,

CH=CH), 6.04 (1H, br s, CH=CH'), 6.01 (2H, m, OH), 5.48 (2H, br s, CH/CH'), 3.85 (3H, s, OCH$_3$), 3.84 (3H, s, OCH$_3$'). ^{13}C NMR (75 MHz, d$_6$-DMSO): 162.0, 162.0 (CmC7), 160.6 (CmC2), 157.5 (CmC8a), 155.2, 155.2 (CmC4), 132.4, 132.2 (HC=CH), 126.9, 126.8 (CmC5), 111.6, 111.5 (CmC6), 110.5, 110.4 (CmC4a), 108.3, 108.1 (CmC3), 100.8, 100.7 (CmC8), 68.6, 68.5 (CH), 55.8 (OCH$_3$). HRMS: 437.1253 [M+H]$^+$ ([C$_{24}$H$_{21}$O$_8$]$^+$ requires 437.1231), 873.2361 [2M+H]$^+$ ([C$_{48}$H$_{41}$O$_{16}$]$^+$ requires 873.2390). IR: 1694 cm-1 (C=O), 1611 cm-1 (C=C).

9 Appendix

9.1 PanD and PanZ protein overexpression optimisation and purification

9.1.1 PanD(T57V)

HexaHis-tagged PanD(T57V) was overexpressed in auto-induction media (0.01% *w/v* ampicillin) from previously transformed *E. coli* MG1655 Δ*panZ* Δ*panD* (DE3) glycerol stocks containing the desired plasmid. The cells were disrupted mechanically in the presence of EDTA-free protease inhibitor cocktail to prevent proteolysis of the C-terminus and centrifuged to remove cell debris. The protein was first purified by nickel affinity chromatography (Ni-NTA) under gravity as described in section 8.6.3. Protein expression levels and purity were assessed by SDS-PAGE (Fig. 9.1), which showed one main overexpressed protein band with the expected molecular weight (~15 kDa).

Fig. 9.1 SDS-PAGE gel of PanD(T57V) after Ni-NTA purification. Lane 1: protein ladder, lanes 2-5: flow-through; lane 6: 50 mL wash fraction; lanes 7-11: 5×5 mL elution fractions. The protein was overexpressed as single band of the expected approximate size (~15 kDa).

The protein was further purified by size-exclusion chromatography (SEC) and eluted into 50 mM Tris buffer, pH 7.4, containing 100 mM NaCl and 0.1 mM DTT (SEC buffer A, section 8.6.4). The protein elutes as a single peak with a shoulder, which is thought to be an oligomer of the tetramer.

Fig. 9.2 SEC trace for the purification of PanD(T57V) on a Superdex® 75 column (HiLoad® 16/60), showing a single peak (120-146 mL) with a shoulder (100-120 mL) for the elution of purified PanD(T57V) protein. The shoulder is thought to be an oligomer of the tetramer.

9.1.2 PanZ

Arabinose-inducible pBAD24 vector encoding C-terminally hexaHis-tagged PanZ (Nozaki *et al.*, 2012) was transformed into *E. coli* BL21 Gold (DE3) cells following a standard heat-shock transformation protocol (see section 8.4.5). Initial overexpression in LB growth media (20 h, 37 °C, induced using 0.5% *w/v* arabinose at OD_{600}=0.4) with and without pantothenate supplementation was poor. Overexpression trials were set up to investigate the effect of both pantothenate supplementation and concentration of arabinose on induction, as described in Table 11. In all cases, expression was induced at OD_{600}=0.4-0.6. Overexpression levels were quantified over time (1.5 h, 2.5 h and 16 h after induction) by SDS-PAGE (Fig. 9.3).

Higher concentrations of arabinose (0.5%-1%) showed better overexpression with contamination of a high molecular weight impurity. Supplementation with up to 250 mM pantothenate improved expression further at 0.5% arabinose and appeared to reduce the amount of high molecular weight contaminants produced. The improvement of PanZ expression in the presence of pantothenate was later rationalised as because the overexpression of PanZ leads to β-alanine auxotrophy in *E. coli* cells, which, in turn, leads to cell death. Addition of pantothenate to the growth medium decreases the toxic effect of PanZ overexpression. It was also noted that PanZ overexpressed in the presence of pantothenate ran more slowly on SDS-PAGE than PanZ overexpressed without pantothenate supplementation. This could indicate a higher percentage of CoA-bound PanZ, as CoA.PanZ is 800 Da heavier than *apo*.PanZ and carries 3 negatively charged phosphate groups.

178

Table 11 PanZ expression trial conditions, related to the SDS-PAGE gels in **Fig. 9.3**.

	E. coli strain	[arabinose] (*w/v* %)	[pantothenate] (µM)	Culture size (mL)
1	BL21 Gold	0.5	1000	1000
2	BL21 Gold	0.5	0	1000
3	C41 (DE3)	0.5	0	100
4	BL21 Star (DE3)	0.5	0	100
5	BL21 Gold (DE3)	0.05	0	100
6	BL21 Gold (DE3)	0.1	0	100
7	BL21 Gold (DE3)	0.5	0	100
8	BL21 Gold (DE3)	1.0	0	100
9	BL21 Gold (DE3)	0.5	5	20
10	BL21 Gold (DE3)	0.5	10	20
11	BL21 Gold (DE3)	0.5	100	20
12	BL21 Gold (DE3)	0.5	250	20

Fig. 9.3 Quantification of protein expression during PanZ overexpression trials by SDS-PAGE. (**Top**) lane 1 – protein ladder; lanes 2 & 3 – trial 1 at 0 h and 16 h; lanes 4 & 5 – trial 2 at 0 h and 16 h; lanes 6-25: trials 3-7 at 0 h, 1.5 h, 2.5 h and 16 h. (**Bottom**) lane 1 – protein ladder; lanes 2-21: trials 8-12 at 0h, 1.5 h, 2.5 h, and 16 h.

PanZ was successfully overexpressed from 1L overnight culture of *E. coli* BL21 Gold (DE3) cells in LB growth media (37 °C), supplemented with 250 mM pantothenate and induced at OD_{600}=0.4-0.6 with 0.5% *w/v* arabinose. The cells were collected, mechanically disrupted and the protein purified by Ni-NTA chromatography as described for PanD(T57V). SDS-PAGE of the fractions showed good overexpression and the expected molecular weight for hexaHis-tagged PanZ (Fig. 9.4).

Fig. 9.4 SDS-PAGE gel of PanZ purification by Ni-NTA overexpressed from the pBAD24 plasmid. Lane 1: protein ladder; lanes 2-4: flow-through; lane 5: 50 mL wash fraction; lanes 6-10: 5×5 mL elution fractions.

Later in the project, Chris Bartlett (Wellcome Trust rotation PhD student) subcloned C-terminally hexaHis-tagged PanZ into the pET28a vector from the pBAD24 plasmid, to allow for overexpression by auto-induction of the T7 RNA polymerase in *E. coli* MG1655 Δ*panZ* Δ*panD* (DE3) cells. The protein overexpressed well using auto-induction and was purified by Ni-NTA chromatography (Fig. 9.5) and SEC and eluted into 50 mM Tris buffer, pH 7.4 containing 100 mM NaCl and 0.1 mM DTT (SEC buffer A, Fig. 9.6).

Fig. 9.5 SDS-PAGE gel of PanZ purification by Ni-NTA overexpressed from the pET28a plasmid. Lane 1: protein ladder; lanes 2-3: flow-through; lane 4: 50 mL wash fraction; lanes 5-9: 5×5 mL elution fractions.

180

Fig. 9.6 SEC trace for the purification of PanZ on a Superdex® 75 column (HiLoad® 16/60), showing a single peak for the elution of PanZ between 170-212 mL.

9.1.3   SeMet PanZ

SeMet PanZ was overexpressed from *E. coli* B384 (DE3) cells harbouring a pBAD24 plasmid coding for C-terminally hexaHis-tagged PanZ, following a previously published autoinduction protocol (see section 8.5.3) (Sreenath *et al.*, 2005). The protein was purified by Ni-NTA and SEC (Fig. 9.7) as described for PanD(T57V).

Fig. 9.7 SEC trace for the purification of SeMet PanZ on a Superdex® 75 column (HiLoad® 16/60), showing a single peak for the elution of the protein between 180-210 mL.

9.1.4   PanD(T57V/K119STOP)

PanD(T57V/K119STOP) was generated using a standard site-directed mutagenesis protocol as described in section 8.4 using the template vector pRSETA-ADC(T57V) (Webb *et al.*, 2014). The pRSETA-ADC(T57V/K119STOP) plasmid was directly transformed into *E. coli* XL10 cells and a colony minicultured overnight in LB media

containing 0.01 % *w/v* ampicillin. The plasmid was miniprepped and sent for sequencing. Upon confirmation of the correct mutation, the plasmid was transformed into *E. coli* MG1655 *ΔpanZ ΔpanD* (DE3) cells for overexpression. The protein was overexpressed by auto-induction of the T7 polymerase and purified by Ni-NTA and SEC, similarly to PanD(T57V) (Fig. 9.8).

Fig. 9.8 SEC trace for the purification of PanD(T57V/K119STOP) on a Superdex® 75 column (HiLoad® 16/60), showing a single peak (120-146 mL) for the elution of the protein.

9.1.5 Not-activated WT PanD

WT PanD (not activated enzyme) was overexpressed from previously transformed *E. coli* C41 (DE3) cells harbouring a pRSETA-ADC-WT plasmid (Saldanha *et al.*, 2001) by auto-induction of the T7 RNA polymerase (37 °C, 18 h). The cells were collected and stored at -80 °C or used directly. The protein was purified by Ni-NTA and SEC as described for PanD(T57V) (Fig. 9.9).

Fig. 9.9 SEC trace for the purification of WT PanD on a Superdex® 75 column (HiLoad® 26/60), showing a single peak (50-60 mL) with a shoulder peak (40-50 mL) for the elution of the protein.

9.1.6 Activated WT ADC

WT PanD (not activated enzyme) was overexpressed from previously transformed *E. coli* C41 (DE3) cells harbouring a pRSETA-ADC-WT plasmid (Saldanha *et al.*, 2001) as described in the previous section. The cell pellet was collected by centrifugation and heated at 37 °C for 24 h for activation of the PanD zymogen to ADC. After activation, the cell pellet was stored at -80 °C or used directly for purification of WT ADC by Ni-NTA and SEC (Fig. 9.10) as described for PanD(T57V). The protein purity, yield and extent of activation were determined by SDS-PAGE (Fig. 9.11).

Fig. 9.10 SEC trace for the purification of WT ADC on a Superdex® 75 column (HiLoad® 26/60), showing a single peak (50-60 mL) with a shoulder peak (40-50 mL) for the elution of the protein.

183

Fig. 9.11 Tris-tricine SDS-PAGE gel of WT ADC purification by Ni-NTA and SEC. Lane 1: protein ladder; lanes 2-4: flow-through; lane 5: 50 mL wash fraction; lanes 6-10: 5×5 mL elution fractions; lanes 11-15: SEC fractions.

The fractions arising from the main SEC peak and the shoulder peak showed that both are cleaved ADC. It had been previously proposed that the early peak corresponded to the non-activated PanD protein and the later peak to the activated protein, but the SDS-PAGE analysis shows similar bands across all the elution fractions. The shoulder is instead thought to be an oligomer of tetramers.

9.2 X-ray diffraction data collections

PanD(T57V).PanZ.CoA crystals from 200 mM Li$_2$SO$_4$, 100 mM Tris pH 8.5 and 30% *v/v* PEG 4000

Beamline I03, Diamond Light source, 27/02/2012

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium		native
2	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	9.51 Å (Xia2 3dii)	native
3	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	8.33 Å (Xia2 3dii)	native
4	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	7.98 Å (Xia2 3dii)	native
5	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	5.47 Å (Xia2 3dii)	native
6	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	4.62 Å (Fast DP)	native
7	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	4.77 Å (Fast DP)	native
8	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium	4.94 Å (Fast DP)	native
9	1°	180	180°	726.9 mm	0.9763 Å	1.00 s	100%	100 K	medium		native
10	1°	360	360°	476.0 mm	1.6531 Å	1.00 s	100%	100 K	small	9.68 Å (Xia2 3dii)	For S SAD

Heavy-metal derivatised PanD(T57V).PanZ.CoA crystals from 200 mM Li$_2$SO$_4$, 100 mM Tris pH 8.5 and 30% v/v PEG 4000

Beamline I24, Diamond Light source, 05/03/2012

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
11	1°	360	360°	476.0 mm	1.6531 Å	1.00 s	100%	100 K	small	12.50 Å (Xia2 3dii)	For S SAD
12	1°	360	360°	476.0 mm	1.6531 Å	1.00 s	100%	100 K	small	7.77 Å (Xia2 3dii)	For S SAD
1	0.15°	1000	150°	476.8 mm	1.0740 Å	0.50 s	9.84%	100 K	10×10 μm	4.66 Å (Xia2 3dii)	(NH$_4$)$_{10}$H$_2$(W$_2$O$_7$)$_6$
2	0.15°	900	135°	561.3 mm	0.9997 Å	0.10 s	10.04%	100 K	10×10 μm	3.67 Å (Xia2 3dii)	K$_2$OsCl$_6$
3	0.15°	900	135°	561.3 mm	0.9997 Å	0.10 s	10.04%	100 K	10×10 μm	6.79 Å (Xia2 3dii)	K$_2$OsCl$_6$
4	0.20°	500	100°	533.4 mm	1.0093 Å	0.10 s	9.49%	100 K	10×10 μm		Hg(OAc)$_2$
5	0.20°	2000	400°	533.4 mm	1.0093 Å	0.10 s	9.49%	100 K	10×10 μm	3.28 Å (Xia2 3dii)	Hg(OAc)$_2$
6	0.20°	1000	200°	533.4 mm	1.0093 Å	0.10 s	9.49%	100 K	10×10 μm	4.26 Å (Xia2 3dii)	Hg(OAc)$_2$
7	0.20°	1000	200°	490.8 mm	1.0081 Å	0.10 s	9.49%	100 K	10×10 μm	4.21 Å (Xia2 3dii)	Hg(OAc)$_2$
8	0.2°	1000	200°	455.9 mm	1.0735 Å	0.10 s	10.43%	100 K	10×10 μm	5.06 Å (Xia2 3dii)	(NH$_4$)$_{10}$H$_2$(W$_2$O$_7$)$_6$

PanD(T57V).PanZ.AcCoA crystals from 100 mM sodium cacodylate pH 6.5, 1.4 M sodium acetate

Beamline I24, Diamond Light Source 23/06/2013

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	0.20°	900	180°	505.9 mm	0.9778 Å	0.20 s	51.75%	100 K	10×10 μm	2.68 Å (Fast DP)	native
2	0.20°	900	180°	505.9 mm	0.9778 Å	0.30 s	78.24%	100 K	10×10 μm	2.69 Å (Xia2 3da)	native
3	0.15°	900	135°	505.9 mm	0.9778 Å	0.30 s	78.24%	100 K	10×10 μm	2.33 Å (Xia2 3dii)	native
4	0.15°	500	75°	505.9 mm	0.9778 Å	0.30 s	78.24%	100 K	10×10 μm	2.45 Å (Xia2 3dii)	native
5	0.20°	750	150°	505.9 mm	0.9778 Å	0.30 s	78.24%	100 K	10×10 μm	2.60 Å (Xia2 3da)	native
6	0.20°	725	145°	505.9 mm	0.9778 Å	0.20 s	61.01%	100 K	10×10 μm	2.59 Å (Xia2 3da)	native

PanD(T57V).PanZ.AcCoA crystals from 100 mM Bis-Tris propane pH 6.8, 200 mM KSCN, 20% *v/v* PEG 3350

Beamline ID29, ESRF 6/11/2013

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	0.05°	2320	116°	339.6 mm	0.9762 Å	0.04 s	0.12%	294 K	50×30 μm	1.61 Å (Fast DP)	native

In-house source, Leeds 3/12/13

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	1.00°	60	60°	110 mm	1.5418 Å	30 s	100%	293 K	-	1.70 Å	native

PanD(S25A).PanZ.AcCoA crystals from 100 mM Bis-Tris propane pH 6.8, 200 mM KSCN, 20% v/v PEG 3350

In-house source, Leeds 3/12/13

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	1.00°	60	60°	110 mm	1.5418 Å	30 s	100%	293 K	-	2.1 Å	native

WT ADC.PanZ.AcCoA crystals from 100 mM Bis-Tris propane pH 6.8, 200 mM KSCN, 20% v/v PEG 3350

Beamline I03, Diamond Light Source 06/10/2014

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	0.20°	900	180°	176.4 mm	0.9763 Å	0.1 s	30%	100 K	50×20 μm	1.16 Å (FastDP)	native

WT ADC.PanZ.AcCoA crystals from 100 mM Bis-Tris propane pH 6.8, 200 mM KSCN, 20% *v/v* PEG 3350

Beamline I04-1, Diamond Light Source 22/01/2015

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	0.20°	1000	200°	255.1 mm	0.9174 Å	0.2 s	43.48%	100 K	30×30 μm	1.20 Å (DIALS)	native

In-house source, Leeds 17/09/2014

dataset	oscillation	images	oscillation (total)	distance (detector)	λ	exposure	transmission	T	aperture	max resolution (pipeline)	comments
1	1.00°	180	180°	100.0 mm	1.5418 Å	20 s	100.0%	293K		1.80 Å (Fast DP)	native

9.3 Calculation of twinning fractions

Twinning is detected during scaling and merging of the data by using the "L-test" (Padilla and Yeates, 2003). This statistical approach analyses the intensity distribution of reflections. The intensities of two unrelated reflections that are close in reciprocal space ($I(h_1)$ and $I(h_2)$) are analysed to give a quantity L, defined as:

$$L \equiv \frac{I(h_1) - I(h_2)}{I(h_1) + I(h_2)}$$

where the range of L is -1 to 1. Each pair of reflections should have similar expected intensity values when crystals are twinned. As dividing the difference of intensities of the spots by their sum leads to normalisation of the values of L, no further normalisation is required, for example, across resolution shells. The probability distribution of L, $N(|L|)$ is then plotted. For untwinned, acentric data:

$$N(|L|) = |L|$$

which would generate a straight line intersecting the origin. For a perfect twin:

$$N(|L|) = \frac{|L|(3 - L^2)}{2}$$

giving a different plot of the cumulative distribution of L. Plotting the calculated values for a dataset against the theoretical curves gives an indication of the presence of twinning as well as of the twin fraction, α (Fig. 9.12).

Fig. 9.12 Plot of the cumulative distribution of |L| (adapted from (Padilla and Yeates, 2003)). The prediction for acentric, non-twinned data is shown as a bold line; for perfect twinned, acentric data as a dashed line and the statistics for experimental data from a highly twinned dataset as open circles.

190

The twin fraction, α, is defined as:

$$I_{h,k,l} = \alpha I_{h,k,l}(\text{crystal1}) + (1 - \alpha)I_{h,k,l}(\text{crystal2})$$

and can also be calculated using the H-test (Yeates, 1997). In this case, two reflections related by a potential twin law, $I(h_1)$ and $I(h_2)$, are analysed to give the quantity H:

$$H = \frac{|I(h_1) - I(h_2)|}{I(h_1) + I(h_2)}$$

The probability distribution of H, $P(H)$, is then plotted and compared against predicted distributions for different twinning fractions to give α for the experimental dataset (Fig. 9.13), where:

$$P(H) = 0 \quad \text{for} \quad H < 0$$

$$P(H) = \frac{H}{1 - 2\alpha} \quad \text{for} \quad 0 \leq H \leq 1 - 2\alpha$$

$$P(H) = 1 \quad \text{for} \quad H > 1 - 2\alpha$$
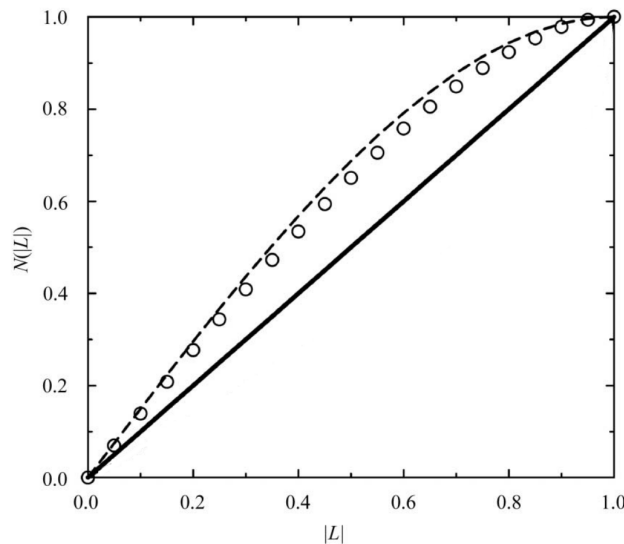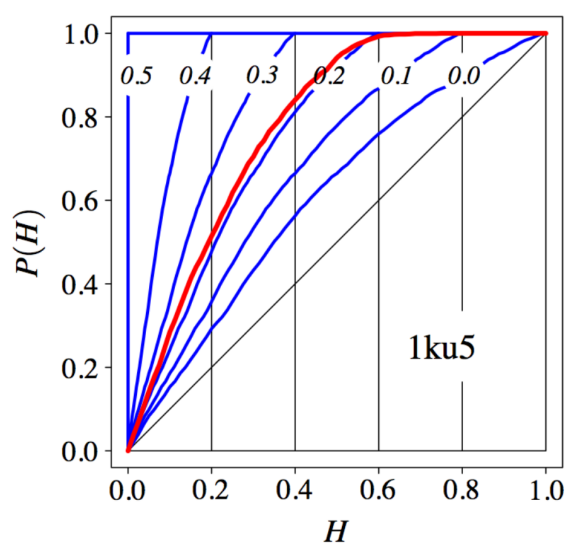
Fig. 9.13 Probabilty distribution of H (adapted from (Lebedev, 2013)). The predicted cumulative probability distribution of H for different twin fractions, α=0.0-0.5 (untwinned to perfectly twinned) are shown as blue lines. The calculated cumulative distribution of H for PDB 1KU5 (Li *et al.*, 2003) is shown in red, with an α≈0.2.

9.4 ITC binding experiments

	Ligand	[Ligand] (μM)	Titrant	[Titrant] (μM)	Additive	[Additive] (μM)	K$_d$ (nM)	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})	ΔG (kcal mol^{-1})	Notes
1	PanZ	100	CoA	800	-	-	6241±511	-16.52±0.41	-31.58	-7.10	0.60 incompetent fraction
2	PanZ	57	CoA	400	-	-	9397±440	-13.34±0.70	-21.73	-6.86	0.38 incompetent fraction
3	PanZ	100	CoA	800	-	-	5212±183	-17.89±0.12	-35.82	-7.21	0.62 incompetent fraction
4	PanZ	60	PanD(T57V)	400	-	-	69±12	-7.91±0.04	6.23	-9.77	0.39 incompetent fraction
5	PanZ	59	PanD(T57V)	394	CoA	100	241±38	-6.92±0.07	7.08	-9.03	No incompetent fraction
6	PanZ	35	PanD(T57V)	263	AcCoA	400	35.1±22.5	-11.97±0.43	-6.04	-10.17	Both pairwise titrations fitted globally
7	PanD(T57V)	35	PanZ	263	AcCoA	400					
8	PanZ	32	PanD(S25A)	257	AcCoA	394	128.8±10.6	-17.36±0.06	-26.7	-9.40	Both pairwise titrations fitted globally
9	PanD(S25A)	35	PanZ	263	AcCoA	400					
10	PanD(T57V/K119STOP)	35	PanZ	263	AcCoA	394	-	-	-	-	No measurable binding

9.5 A selection of photocaged compounds previously used in biological and physiological studies
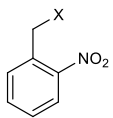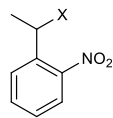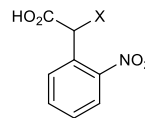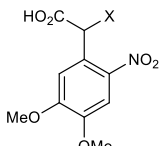
Photocage	Leaving group	λ_{max} (nm)	ε_{max} (M^{-1}cm^{-1})	φ	k (s^{-1})	ref
7-alkoxycoumarin	General	325	4000-12000	0,1	1×10^8	(Corrie et al., 2005)
7-methoxycoumarin	cAMP	325	13300	0,12	4.2×10^8	(Schade et al., 1999)
7-(acyloxy)coumarin	cAMP	313	7650	0.056		(Furuta and Iwamura, 1998)
7-(propyloxy)coumarin	cAMP	313	6500	0.054		(Furuta and Iwamura, 1998)
7-hydroxycoumarin	cAMP	326	16800	0.062		(Furuta and Iwamura, 1998)
7-hydroxycoumarin	OAc	325	11600	0.025		(Furuta et al., 1999)
6,7-dialkoxycoumarin	General	345	1000-12000	0,1		(Corrie et al., 2005)
6,7-dimethoxycoumarin	cAMP	349	11000	0.04		(Eckardt et al., 2002)
6-bromo-7-hydroxycoumarin	General	375	13000-19000	0,1		(Corrie et al., 2005)
6-bromo-7- hydroxycoumarin	cAMP	374	16300	0.11		(Furuta et al., 2004)
6-bromo-7- hydroxycoumarin	OAc	397	15900	0.065		(Furuta et al., 1999)
7-dialkylaminocoumarin	General	395	16000-20000	0,3	1×10^9	(Corrie et al., 2005)
7-diethylaminocoumarin	cGMP	402	18600	0.21	1×10^9	(Hagen et al., 2001)
7-dimethylaminocoumarin	ATP	385	15300	0.086	1.6×10^9	(Schönleber et al., 2002)
2-nitrobenzene	Pyvalate	254		0.13		
2-nitrobenzene	ATP	260	26600	0.19		(Kaplan et al., 1978)
1-ethyl-2-nitrobenzene	Pyvlate	254		0.64		
1-ethyl-2-nitrobenzene	ATP	260	26600	0.63	86	(Walker et al., 1988)
1-ethyl-2-nitrobenzene	γ-Glu				9	(Wieboldt et al., 1994)
1-ethyl-2-nitrobenzene	N-Glu			0.06	2.2×10^3	(Wieboldt et al., 1994)
(2-nitrobenzene)ethanoic acid	γ-Glu	262	5100	0.14	33×10^3	(Wieboldt et al., 1994)
(2-nitrobenzene)ethanoic acid	α-Glu			0.14	8.7×10^3	(Wieboldt et al., 1994)
(2-nitrobenzene)ethanoic acid	N-Glu			0.044	410	(Wieboldt et al., 1994)
3,4-dimethoxy-6-nitrobenzene	α-Glu	345	5900	0.006		(Wilcox et al., 1990)

3,4-dimethoxy-1-ethyl-6-nitrobenzene	ATP	249	17200	0.07	18	(Wootton and Trentham, 1989)
para-hydroxyphenyl	ATP	286	14600	0.3	1×10^7	(Givens *et al.*, 1997)
para-hydroxyphenyl	γ-Glu			0.08	7×10^7	(Givens *et al.*, 1997)
para-hydroxyphenyl	O-GABA	282	14454	0.21	1.9×10^8	(Corrie and Trentham, 1992)
4-hydroxy-3-methoxyphenyl	γ-Glu	279	9310	0.035		(Conrad *et al.*, 2000)
4-hydroxy-3-methoxyphenyl	O-GABA	279	9332	0.04	26.1×10^8	(Corrie and Trentham, 1992)
3,5-dimethoxy-4-hydroxyphenyl	γ-Glu	304	11730	0.035	2×10^7	(Conrad *et al.*, 2000)
3,5-dimethoxy-4-hydroxyphenyl	O-GABA	303	0.03	0.03	2.2×10^7	(Corrie and Trentham, 1992)

9.6 Photocage scaffolds, properties and cleavage mechanism

9.6.1 *ortho*-Nitrobenzyl

Table 12 Overview and general properties of *ortho*-nitrobenzyl photocleavable groups

λ_{max} (nm)	254	254	262	345
ε_{max} (M^{-1}cm^{-1})	ca. 27000	ca. 27000	ca. 5000	ca. 600
φ	0.1-0.2	0.1-0.64	0.04-0.14	0.01
k (s^{-1})	10 - 200	10 - 1000	9×10^3 - 3×10^4	N/A
solubility (H$_2$O)	Poor	Poor	Good	Poor

ortho-Nitrobenzyl photocleavable protecting groups are one of the most used and investigated photocages to date. *o*NB groups absorb mainly in the λ=280 nm region and the rate of release of product varies between 10 ms-100 µs depending on the exact chromophore being used as well as external factors such as pH. The cleavage mechanism (Fig. 9.14) has two distinct stages. The first is the light step, which involves the absorption of a photon and the excitation of the chromophore, followed by the dark step which then causes release of the product. Upon absorption of a photon, the nitrobenzyl group is placed in an excited singlet (I) state which can cross to a longer lived triplet state (Schmierer *et al.*, 2010). Both excited states can then decay to the *aci*-nitro intermediate (II) which cyclises to a benzisoxazoline ring (IV). The ring intermediate then opens to an alcohol (V), which, in turn, hydrolyses to release the photoprotected moiety, leaving a nitroso photolysis by-product (VI). The rate of *o*NB cleavage is greatly dependent on pH due to the equilibrium established between the protonated and deprotonated *aci*-nitro species formed after excitation of the chromophore (II and III respectively) as only the protonated form (II) is able to ring close to the benzisoxazoline intermediate (IV). The rate of *aci*-nitro decay decreases decreases rapidly with increasing pH.
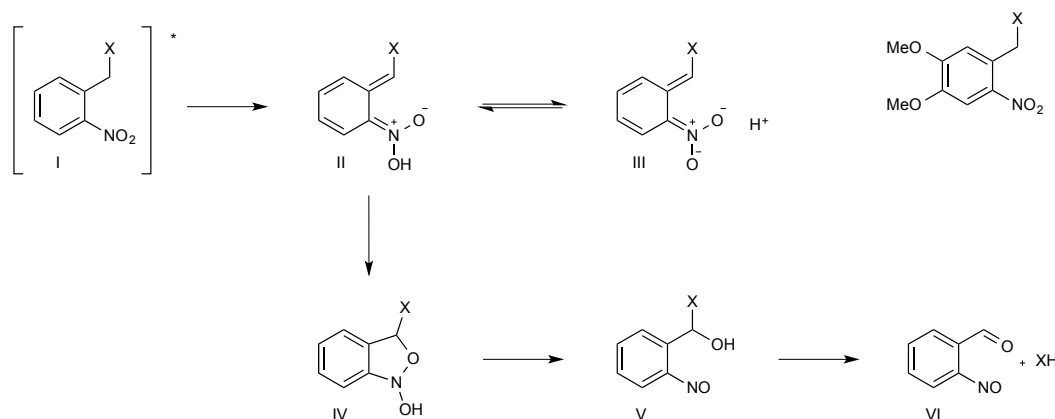
Fig. 9.14 Mechanism of photo-cleavage of *ortho*-nitrobenzyl protecting groups.

The rate determining step for the release of the product depends on the leaving group. At neutral conditions, the rate determining step in the release of poor leaving groups, such as alcohols (Il'ichev *et al.*, 2004), is the hydrolysis of the ring-opened intermediate (V), whereas for good leaving groups, such as phosphates, the rate of release has been found to match that of the decay of the *aci*-intermediate (II) instead (Walker *et al.*, 1988). The rate of cleavage of these compounds has been found to be greatly dependent upon pH, solvent, substitutions of both the ring and the benzylic position and the leaving group. Although the exact effect of changing each of these parameters is difficult to predict (Holmes, 1997, Corrie *et al.*, 2005), some general characteristics can be determined and are summarised below.

Electron-donating ring substituents
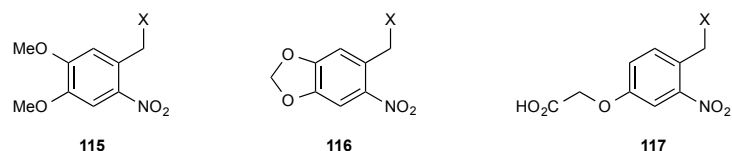
Fig. 9.15 *o*NB scaffolds with varying electron-donating groups. **115**: 4,5-Dimethoxy-2-nitrobenzyl; **116**: 4,5-Methylenedioxy-2-nitrobenzyl; **117**: 4-(2-Ethanoic acid)oxy-2-nitrobenzyl.

Modification of the aromatic ring causes changes to the absorption spectrum of the compounds. The introduction of electron-donating groups in the aromatic ring increases the absorption at higher wavelengths. Veratryl and methylenedioxy based nitrobenzyl photocages (compounds **115** and **116**) have been given considerable attention as these have large absorption cross-sections at $\lambda > 350$ nm. But, the introduction of strong electron-donating groups (such as amines) *para* to the nitro group have been reported to significantly lower the decaging efficiency of these compounds (Corrie *et al.*, 2005).

196

One other advantage of introducing polar functional groups on the ring, such as carboxylic acids (compound **117**) (Allan *et al.*, 1998), is the potential to improve solubility of the compounds.

Substitution at the benzylic carbon

Substitution of the benzylic position mainly affects the quantum yield of photolysis. Furthermore, ring opening of the benzisoxazoline intermediate leads to the formation of a carbonyl at the benzylic centre. In the simplest of nitrobenzyl cages, the by-product of photolysis is an aldehyde (Fig. 9.16), which is unreactive when at low concentrations, but reactive at higher concentrations (Engels and Schlaeger, 1977). The aldehyde by-product is also coloured, and can act as a light filter. Introduction of a methyl substituent at the benzylic position leads to the formation of an unreactive ketone photolysis by-product instead (Fig. 9.16).
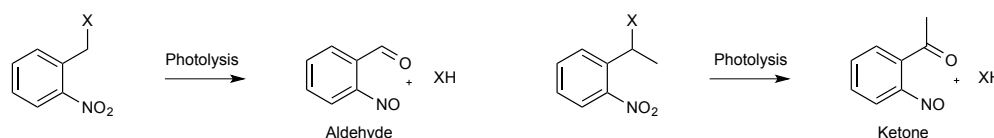
Fig. 9.16 Photolysis of 2-methyl nitrobenzyl and 2-ethyl nitrobenzyl photoprotecting groups and corresponding products.

Substitutions at this position have also been shown to increase the rates of release of product. The introduction of a methyl group has been shown to cause a 7-fold decrease in the photolysis half-life of compound **119** vs compound **118** (Fig. 9.17) in PBS buffer (Holmes, 1997) and a 5-fold increase in quantum yield of compound **121** vs compound **120** (Fig. 9.17) in solution (Reichmanis *et al.*, 1985). This effect has also been observed for the addition of strongly electron-withdrawing groups (such as CF_3) (Specht and Goeldner, 2004), although further studies on other electron-withdrawing substituents (in particular halogens) did not show such radical improvements in efficiency of the compounds
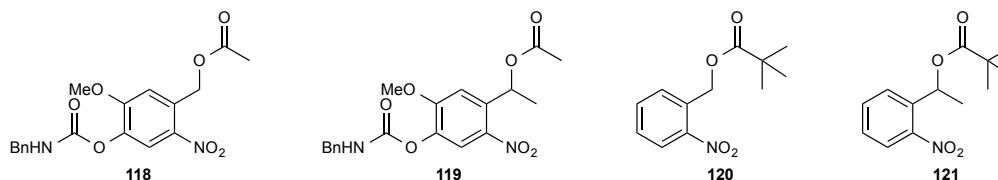
Fig. 9.17 Examples of *o*NB photocaged acetate (Holmes, 1997) and *tert*-butylacetate (Reichmanis *et al.*, 1985).

<u>Leaving groups</u>

In general, good leaving groups have increased rates of release. The most common way of protecting compounds with nitrobenzyl photocages is to directly attach the compound to the benzylic carbon. This approach has been used for the release of different groups, such as thiols, carboxylic acids, histidines and phosphates (Corrie *et al.*, 2005, Mayer and Heckel, 2006, Klán *et al.*, 2013).

Poorer leaving groups, such as alcohols and amines, are usually attached as the corresponding carbonates or carbamates, increasing the efficiency of cleavage. It is important to note that the product of photolysis of these compounds is still the carbonate or carbamate derivative of the desired product, which has to undergo post-photolytic fragmentation to the amine or carboxylate (Fig. 9.18). The fragmentation step is slow, with a rate of roughly 10^{-3} s^{-1} (Pocker *et al.*, 1978).
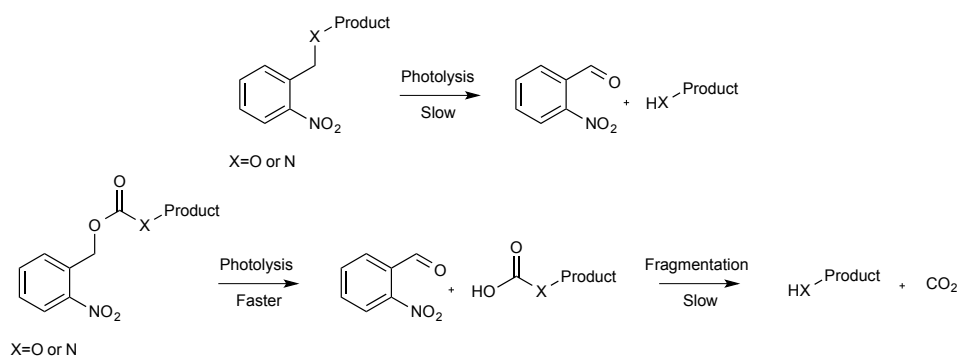
Fig. 9.18 Release of alcohols or amines from *o*NB protected compounds. Alcohols and amines can be directly coupled to the benzylic carbon, but the rate of photolysis of the photocage is slow. The compounds can also be photocaged through the corresponding carbonates or cabamates, which increases the rate of photolysis but the photolytic products have to undergo fragmentation to the desired alcohol or amine.

9.6.2 Coumarinyl

Table 13 Overview and general properties of coumarinyl photocleavable groups

λ_{max} (nm)	320	345	375	390
ε_{max} ($M^{-1}cm^{-1}$)	6000-12000	1000-12000	13000-19000	16000-20000
φ	0.02-0,15	0.04-0,1	0.02-0,11	0.09-0,3
k (s^{-1})	1×10^{8} - 4×10^{8}	N/A	N/A	1×10^{9} - 2×10^{9}
solubility (H_2O)	Variable	Variable	Good	Medium

The use of (coumarin-4-yl)-methyl cages in a biological context was first described in 1984 for the release of phosphate esters (Givens and Matuszewski, 1984). The phosphate moiety was efficiently released from the 7-methoxycoumarinyl-4-methyl derivative by illumination at $\lambda=360$ nm, which is higher than the oNB systems discussed previously. The good absorption cross-sections at longer wavelengths made this photoprotection group centre-stage for later developments. The release rates of coumarins are much faster than oNBs, reaching ns timescales for the release of strong acid conjugate bases such as methyl sulfonates (Geißler et $al.$, 2005). One last advantage of coumarins is the large extinction coefficients reported for this class of compounds.

Upon absorption of a photon, the coumarin relaxes to an excited singlet state (Fig. 9.19, II), which can either decay back to the ground state (II→I) or undergo heterolytic bond cleavage (II→III). The cleavage event is very fast, reaching a rate of 2×10^{10} s^{-1} (determined for coumarinyl protected phosphate esters) (Schade et $al.$, 1999) and the ion pair formed can then either recombine to the starting material (III→I) or dissociate (III→IV) and react with a nucleophile in the medium (V) to complete the cleavage pathway. In this last step, the recombination reaction is the dominant pathway, making the ion-pair dissociation the rate determining step in coumarin photolysis (Schmidt et $al.$, 2007). The choice of buffer also affects the rate of separation of the tight ion pair (III), altering the ratio of the recombination and separation pathways, and therefore, the quantum yield.
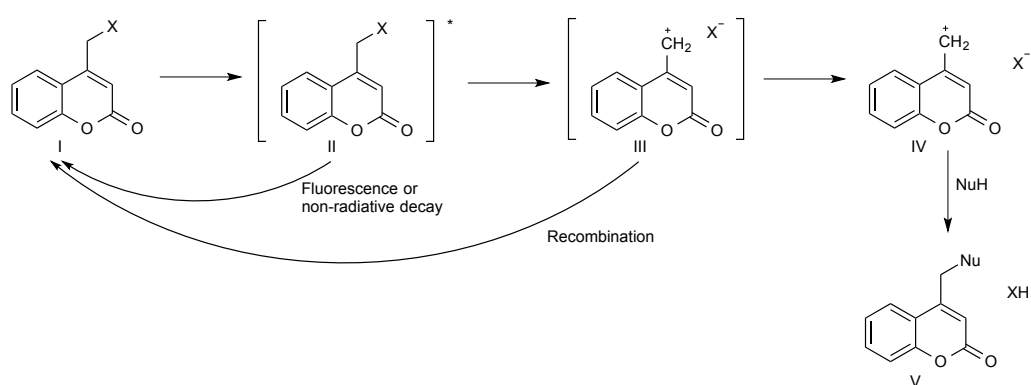
Fig. 9.19 Mechanism of coumarin photocleavage, Upon absorption of a photon, the coumarin (I) is placed in an excited singlet or triplet state (II). This state can decay back to ground state by fluorescence or non-radiatively or cleave heterolytically giving a tightly bound ion pair (III). The ion pair can recombine to the initial photocaged compound (I) or separate, causing the release of the product (IV). A nucleophilic species (usually water) then quenches the carbocation (V).

Substitution to the ring
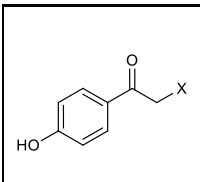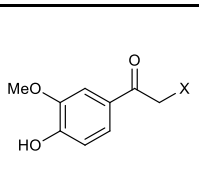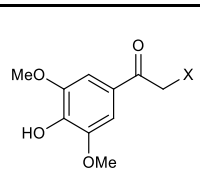
The introduction of substituents in the coumarin ring mainly alters the absorption wavelength of the cage and can help with solubility issues of the compounds. The parent coumarin has an absorption maximum at λ~310 nm and introduction of hydroxyl or methoxy moieties at C7 increases the maximum absorption wavelength by 15-20 nm. C7 substitution with free or alkyl amines causes a significant shift in absorption to 350-395 nm. Addition of methoxy groups at C7 and C6 further increase the absorption wavelength to 345 nm and a combination of C6-Br with C7-OH to 375 nm, whereas C6-Br/C7-OMe has its maximum absorption at λ=330 nm. Therefore, the exact effect of ring substitutions can't be easily predicted. Table 13 shows a few examples of coumarinyl compounds and their maximum absoptions. The introduction of electron donating groups has an effect on the efficiency of cleavage, as it facilitates heterolysis and hinders ion pair recombination (Schmidt *et al.*, 2007).

Leaving group

Just as described for *o*NB photocages, the leaving group has a significant effect on the efficiency of product release during photolysis of coumarins. Good leaving groups, such as phosphates and sulphates, cleave efficiently, whereas poorer leaving groups, such as alcohols or thiols, are resistant to heterolysis. Such groups are usually linked through a carbonate linker to the coumarin but have to undergo decarboxylation before complete release of the desired product (Klán *et al.*, 2013).

9.6.3 *para*-Hydroxyphenyl

Table 14 Overview and general properties of *para*-hydroxyphenyl photocleavable groups

λ_{max} (nm)	285	280	304
ε_{max} (M^{-1}cm^{-1})	ca. 15000	ca. 9000	ca. 12000
φ	0.2-0.9	0.03-0.04	0.03
k (s^{-1})	1×10^7 - 2×10^9	ca. 2×10^9	ca. 2×10^7
Solubility (H$_2$O)	Good	Good	Good

para-Hydroxylphenyl (*p*HP) photoprotecting groups are much more recent than oNB or Cm photocages. These have found applications in neurobiology and enzyme catalysis

investigations (Klán *et al.*, 2013). These groups present good solubilities and high quantum yields but have maximum absorption wavelengths in the 280 nm region (Klíčová *et al.*, 2012), which is much lower than the preferred illumination region of biological samples.

The groups cleave via an interesting mechanism, similar to a Favorskii rearrangement. The absorption of a photon leads to the excitation of the *p*HP moiety to a triplet state which then rearranges to a cyclopropanone intermediate (**II**) with release of the product. The cyclopropanone then hydrolyses to *para*-hydroxybenzoic acid (**III**, Fig. 9.20). The rate of cyclopropanone formation, and consequently of product release, is very fast, occurring in the ns timescale (Conrad *et al.*, 2000), and is dependent mainly on the leaving group.
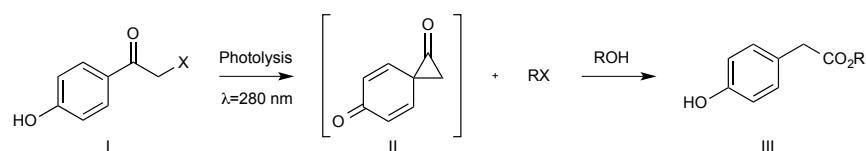
Fig. 9.20 Cleavage mechanism of *p*HP protected molecules. Absorption of a photon leads to the formation of a cyclopropanone intermediate (also called the Favorskii intermediate) and decaging of the biomolecule. The intermediate then hydrolyses to a carboxylate (Klán *et al.*, 2013).

The leaving group

*p*HPs have been mainly investigated for the release of strong conjugate bases, such as sulfonates, phosphates and carboxylates and lower quantum yields and rates of release have been reported for poorer leaving groups (Givens *et al.*, 2008). Increasing the pKa of the corresponding acid of the leaving group causes a decrease in the quantum yield as well as a decrease in the rate of product release (Klán *et al.*, 2013).

Subsition on the chromophore

Alterations to the chromophore have effects on both the maximum absorption wavelength and quantum efficiency. The introduction of meta-electron donating groups causes an increase of the absorption wavelength, but lowers the quantum efficiencies of the compounds whereas electron-accepting groups increase the quantum yield.

9.6.4   Summary

There are three main scaffolds for photocages currently used in biological applications, each having different spectroscopic properties as well as cleavage rates. *o*NB photocages are the slowest (10^{-3}-10^{-4} s), followed by coumarins and *para-*

hydroxyphenyls (10^{-6}-10^{-8} s). The cleavage rates depend on both the photocage scaffold and the protected substrate, as well as the buffer composition and pH.

*o*NB photocages are the slowest and the benzaldehyde photolysis by-products are reactive and can act as light filters, however *o*NB photocaged compounds are easily synthesised. The disadvantages can be overcome by the introduction of a substituent at the benzylic site, but this leads to the creation of a new chiral centre, making the synthesis and purification of the compounds more challenging.

Solubility is a problem for the use of both coumarins and *o*NBs. *p*HPs are much more soluble and synthetically available but have very short maximum absorption wavelengths, the major disadvantage of using these protecting groups.Therefore, choosing a photocage scaffold is not a straightforward process. Faster release rates usually come with either solubility problems or bad spectroscopic properties. Although the scaffold will have relatively predictable absorption spectra, the cleavage rates and quantum yields cannot be predicted and can only be completely defined once the compound is synthesised and characterised experimentally.

9.7 Protein sequence alignments

9.7.1 Full PanD protein sequences from different organisms

Semi-conserved residues are marked as . and conserved residues as *. The conserved C-terminal region across all PanZ producing bacteria is highlighted in red. The alignment was performed using ClustalW2 (McWilliam *et al.*, 2013).

```
P. aeruginosa     1  MHAIMLKAKLHRAEVTHAVLDYEGSCAIDGDWLDLSGIREYEQIQIYNVDNGERFTTYA   59
E. coli           1  MIRTMLQGKLHRVKVTHADLHYEGSCAIDQDFLDAAGILENEAIDIWNVTNGKRFSTYA   59
E. carotovora     1  MIRTMLQGKLHRVKVTQADLHYEGSCAIDQDFMDAAGILEYEAIDIYNVDNGQRFSTYA   59
Y. pestis         1  MIRTMLQGKLHRVKVTQADLHYEGSCAIDQDFLEAAGILEYEAIDIYNVDNGQRFSTYA   59
C. glutamicum     1  MLRTILGSKIHRATVTQADLDYVGSVTIDADLVHAAGLIEGEKVAIVDITNGARLETYV   59
S. coelicolor     1  MLRTLIKSKIHRATVTQADLHYVGSVTIDADLLDAADLLPGELVHIVDVTNGARLETYV   59
M. tuberculosis   1  MLRTMLKSKIHRATVTCADLHYVGSVTIDADLMDAADLLEGEQVTIVDIDNGARLVTYA   59
T. thermophilus   1  MKRVMFHAKIHRATVTQADLHYVGSVTVDQDLLDAAGILPFEQVDIYDITNGARLTTYA   59
L. pneumophila    1 MAYRKMLKSKIHRACVTQADLDYEGSITISPELLKVANILPYEAVNVWNITAGTRFETYA   60
C. crescentus     1  MLLTMLKAKLHRATVTQADLDYEGSIAIDRDLLDASGILPNEQVDVLNITNGARFTTYA   59
B. subtilis       1  MFRTMMRAKLHRATVTEANLNYVGSITIDEDLMDAVNIVENEKVQIVNNNNGARLETYV   59
S. aureus         1  MIRTMMNAKIHRARVTESNLNYVGSITIDSDILEAVDILPNEKVAIVNNNNGARFETYV   59
H. pylori         1  MTFEMLYSKIHRATITDANLNYVGSITIDEDLAKLAKLREGMKVEIVDVNNGERFSTYV   59
                        .   *.**   .* .  * * **  ..   .        .   ..      * *  **

P. aeruginosa    60  IRAENGSKMISVNGAAAHKAKVGDRVIICAYAHYSEAELASHKPRMLYMAPGNQLSHTS  118
E. coli          60  IAAERGSRIISVNGAAAHCASVGDIVIIASFVTMPDEEARTWRPNVAYFEGDNEMKRTA  118
E. carotovora    60  IAGERGSRIISVNGAAARCACVGDKLIICSYVQMSDEQARSHSPKVAYFSGENELQRQA  118
Y. pestis        60  IAAERGSRIISVNGAAARCACVGDKLIICSYVQMSYAAARLHHPKVAYFEGENQLQRKA  118
C. glutamicum    60  IVGDAGTGNICINGAAAHLINPGDLVIIMSYLQATDAEAKAYEPKIVHVDADNRIVALG  118
S. coelicolor    60  IEGERGSSVIGINGAAAHLVHPGDLVILISYAQVTDAEARSLRPRVVHVDGDNRIVGLG  118
M. tuberculosis  60  ITGERGSSVIGINGAAAHLVHPGDLVILIAYATMDDRARTYQPRIVFVDAYNKPIDMG   118
T. thermophilus  60  LPGERGSSVIGINGAAAHLVKPGDLVILVAYGVFDEEEARNLKPTVVLVDERNRILEVR  118
L. pneumophila   61  ITGEKGSTDICVNGAAAHLVTPGDLVIIASFTQILEEDCAAHEPTVVFVDQFNRLKEIR  119
C. crescentus    60  IEAPRGSKVIGVNGAAARLVQKNDLVIVVTYCQMPAEEARNYAPTVVLLDEGNLIKKAA  118
B. subtilis      60  IKGERGSSVVCLNGAAARLVQPGDKVIIICYGLVAEENIHKQEPKIAVLDDDNQIIEML  118
S. aureus        60  IAGERGSGKICLNGAASRLVEVGDVVIIMTYAQLNEEEIKNHAPKVAVMNEDNVIIEMI  118
H. pylori        60  ILGKK-KGEICVNGAAARKVAIGDVVIILAYASMNEDEINAHKPSIVLVDEKNEILEK   116
                        .        . .****..     .* .*. ..          * .      *

P. aeruginosa   119  E<mark>AIPIQVA</mark>                            126
E. coli         119  K<mark>AIPVQVA</mark>                            126
E. carotovora   119  K<mark>AIPVQVA</mark>                            126
Y. pestis       119  K<mark>AVPVQVA</mark>                            126
C. glutamicum   119  NDLAEALPGSGLLTS-RSI                  136
S. coelicolor   119  ADASEPVPGSDQERSPQAVSA                139
M. tuberculosis 119  HDPAFVPENAGELLDPRLGVG                139
T. thermophilus 119  KG                                   120
L. pneumophila  120  PERIGVKSRIPYPA                       133
C. crescentus   119                                       118
B. subtilis     119  GAEKAGTIL                            127
S. aureus       119  HEKENTIVL                            127
H. pylori       116                                       116
```

9.7.2 DNA sequence alignments for the different PanD constructs

```
WT                 ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCgtGGATC
H17C               ATGCGGGGTT CTCATcatCA TCATCATcaT GGTCTGGTTC CGCgTGGATc
D19C               ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
E23C               ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
C26S               ATGCGGGGTT CTCATCATCA TcatcaTCAT GGTCTGGTTC CGCGTGGATC
A62C               ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCgtGGATC
C78A               ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCgtGGATC
H17CE23C           ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCgtGGATC
H17CC26S           ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
H17CC78A           ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCgtGGATC
D19CE23C           ATGCGGGGTT CTCATCATCA TCATCATCAT gGTCTGGTTC CGCgtGGATC
D19CC26S           ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
E23CA62C           ATGCGGGGtT CTCATCATCA TCATCATCAT ggtCTgGtTC CGCgtgGATC
E23CC78A           ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
C26SC78A           ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
H17CE23CC78A       ATGCGGGGTT CTCATCATCA TCATCATCAT gGTCTGGTTC CGCgtgGATC
D19CE23CC78A       ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
E23CA62CC78A       ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
E23CC26SC78A       ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
H17CE23CC26SC78A   ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC
D19CE23CC26SC78A   ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCgtGGATC
E23CC26SA62CC78A   ATGCGGGGTT CTCATCATCA TCATCATCAT GGTCTGGTTC CGCGTGGATC

WT                 CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtG AAAGTGACTC
H17C               catgATTCGC ACGATGctGC AGGGCAAACt CCACCGcgtg AAAGTGACTT
D19C               CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
E23C               CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
C26S               CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
A62C               CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
C78A               CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtG AAAGTGACTC
H17CE23C           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtG AAAGTGACTT
H17CC26S           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTT
H17CC78A           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtG AAAGTGACTT
D19CE23C           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtg AAAGTGACTC
D19CC26S           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
E23CA62C           CATgATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtg AAAGTGACTC
E23CC78A           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
C26SC78A           CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
H17CE23CC78A       CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtg AAAGTGACTT
D19CE23CC78A       CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtG AAAGTGACTC
E23CA62CC78A       CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
E23CC26SC78A       CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC
H17CE23CC26SC78A   CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTT
D19CE23CC26SC78A   CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCgtG AAAGTGACTC
E23CC26SA62CC78A   CATGATTCGC ACGATGCTGC AGGGCAAACT CCACCGCGTG AAAGTGACTC

WT                 ATGCGGACCT GCACTATGAA GGTTCTTGCG CCATTGACCA GGATTTTCTT
H17C               GTGCGgACCT GCACTATGAA GGTTCTTGCG CCATtgACCA GGATTTTCtT
D19C               ATGCGTGCCT GCACTATGAA GGTTCTTGCG CCATTGACCA GGATTTTCTT
E23C               ATGCGGACCT GCACTATTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
C26S               ATGCGGACCT GCACTATGAA GGTTCTTCCG CCATTGACCA GGATTTTCTT
A62C               ATGCGGACCT GCACTATGAA GGTTCTTGCG CCATTGACCA GGATTTTCTT
C78A               ATGCGGACCT GCACTATGAA GGTTCTTGCG CCATTGACCA GGATTTTCTT
H17CE23C           GTGCGGACCT GCACTATTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
H17CC26S           GTGCGGACCT GCACTATGAA GGTTCTTCCG CCATTGACCA GGATTTTCTT
H17CC78A           GTGCGGACCT GCACTATGAA GGTTCTTGCG CCATTGACCA GGATTTTCTT
D19CE23C           ATGCGTGCCT GCACTATTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
D19CC26S           ATGCGTGCCT GCACTATGAA GGTTCTTCCG CCATTGACCA GGATTTTCTT
E23CA62C           ATGCGGACct gCACTAttgt GGTTCTTGCG CCATtgACCA GGATTTTCTT
E23CC78A           ATGCGGACCT GCACTATTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
C26SC78A           ATGCGGACCT GCACTATGAA GGTTCTTCCG CCATTGACCA GGATTTTCTT
H17CE23CC78A       GTGCGGACCT GCACtaTTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
D19CE23CC78A       ATGCGTGCCT GCACTATTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
E23CA62CC78A       ATGCGGACCT GCACTATTGT GGTTCTTGCG CCATTGACCA GGATTTTCTT
E23CC26SC78A       ATGCGGACCT GCACTATTGT GGTTCTTCCG CCATTGACCA GGATTTTCTT
H17CE23CC26SC78A   GTGCGGACCT GCACTATTGT GGTTCTTCCG CCATTGACCA GGATTTTCTT
D19CE23CC26SC78A   ATGCGTGCCT GCACTATTGT GGTTCTTCCG CCATTGACCA GGATTTTCTT
E23CC26SA62CC78A   ATGCGGACCT GCACTATTGT GGTTCTTCCG CCATTGACCA GGATTTTCTT

WT                 GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
H17C               GACGCAgCCG gtATTCTCGA AAACGAAGCC ATtgATATCT GGAATGTCac
D19C               GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
E23C               GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
C26S               GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
A62C               GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
```

```
C78A                    GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
H17CE23C                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
H17CC26S                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
H17CC78A                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
D19CE23C                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
D19CC26S                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
E23CA62C                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
E23CC78A                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
C26SC78A                GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
H17CE23CC78A            GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
D19CE23CC78A            GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
E23CA62CC78A            GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
E23CC26SC78A            GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
H17CE23CC26SC78A        GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
D19CE23CC26SC78A        GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC
E23CC26SA62CC78A        GACGCAGCCG GTATTCTCGA AAACGAAGCC ATTGATATCT GGAATGTCAC

WT                      CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
H17C                    CAACGGCAAG CGTTTCTCCA CttatgCCAT CGCGGCAGAA CGCGgttCGA
D19C                    CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
E23C                    CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
C26S                    CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
A62C                    CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGTGTGAA CGCGGTTCGA
C78A                    CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
H17CE23C                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
H17CC26S                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
H17CC78A                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
D19CE23C                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
D19CC26S                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
E23CA62C                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGTGTGAA CGCGGTTCGA
E23CC78A                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
C26SC78A                CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
H17CE23CC78A            CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
D19CE23CC78A            CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
E23CA62CC78A            CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGTGTGAA CGCGGTTCGA
E23CC26SC78A            CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
H17CE23CC26SC78A        CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
D19CE23CC26SC78A        CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGGCAGAA CGCGGTTCGA
E23CC26SA62CC78A        CAACGGCAAG CGTTTCTCCA CTTATGCCAT CGCGTGTGAA CGCGGTTCGA

WT                      GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
H17C                    GAATTATTTC TGTTAACGGT GCGGCGGCCC ActGCGCCag tGTCGGCGAT
D19C                    GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
E23C                    GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
C26S                    GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
A62C                    GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
C78A                    GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
H17CE23C                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
H17CC26S                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
H17CC78A                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
D19CE23C                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
D19CC26S                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
E23CA62C                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACTGCGCCAG TGTCGGCGAT
E23CC78A                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
C26SC78A                GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
H17CE23CC78A            GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
D19CE23CC78A            GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
E23CA62CC78A            GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
E23CC26SC78A            GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
H17CE23CC26SC78A        GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
D19CE23CC26SC78A        GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT
E23CC26SA62CC78A        GAATTATTTC TGTTAACGGT GCGGCGGCCC ACGCAGCCAG TGTCGGCGAT

WT                      ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
H17C                    ATTGTCATCA TCGCCAGCtt cGTTACCATG CCAGATGAAG AAGCTCGCAC
D19C                    ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
E23C                    ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
C26S                    ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAg AAGCTCGCAC
A62C                    ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
C78A                    ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
H17CE23C                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
H17CC26S                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
H17CC78A                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
D19CE23C                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
D19CC26S                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
E23CA62C                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
E23CC78A                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
C26SC78A                ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
H17CE23CC78A            ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
D19CE23CC78A            ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
```

205

```
E23CA62CC78A           ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
E23CC26SC78A           ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
H17CE23CC26SC78A       ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
D19CE23CC26SC78A       ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC
E23CC26SA62CC78A       ATTGTCATCA TCGCCAGCTT CGTTACCATG CCAGATGAAG AAGCTCGCAC


WT                     CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
H17C                   CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAagcGTA
D19C                   CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
E23C                   CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
C26S                   CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
A62C                   CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
C78A                   CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
H17CE23C               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
H17CC26S               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
H17CC78A               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
D19CE23C               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
D19CC26S               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
E23CA62C               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
E23CC78A               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
C26SC78A               CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
H17CE23CC78A           CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
D19CE23CC78A           CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
E23CA62CC78A           CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
E23CC26SC78A           CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
H17CE23CC26SC78A       CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
D19CE23CC26SC78A       CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA
E23CC26SA62CC78A       CTGGCGACCC AACGTCGCCT ATTTTGAAGG CGACAATGAA ATGAAACGTA


WT                     CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
H17C                   CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
D19C                   CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
E23C                   CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
C26S                   CCGCGAAAGC GATTCCgGtA CaGGTTGCTT GA
A62C                   CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
C78A                   CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
H17CE23C               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
H17CC26S               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
H17CC78A               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
D19CE23C               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
D19CC26S               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
E23CA62C               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
E23CC78A               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
C26SC78A               CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
H17CE23CC78A           CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
D19CE23CC78A           CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
E23CA62CC78A           CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
E23CC26SC78A           CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
H17CE23CC26SC78A       CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
D19CE23CC26SC78A       CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
E23CC26SA62CC78A       CCGCGAAAGC GATTCCGGTA CAGGTTGCTT GA
```

9.7.3 Protein sequence alignment for the different PanD constructs

```
WT                  -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYE GSCAIDQDFL 33
H17C                -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTCADLHYE GSCAIDQDFL 33
D19C                -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHACLHYE GSCAIDQDFL 33
E23C                -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYC GSCAIDQDFL 33
C26S                -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYE GSSAIDQDFL 33
A62C                -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYE GSCAIDQDFL 33
C78A                -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYE GSCAIDQDFL 33
H17CE23C            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTCADLHYC GSCAIDQDFL 33
H17CC26S            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTCADLHYE GSSAIDQDFL 33
H17CC78A            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTCADLHYE GSCAIDQDFL 33
D19CE23C            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHACLHYC GSCAIDQDFL 33
D19CC26S            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHACLHYE GSSAIDQDFL 33
E23CA62C            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYC GSCAIDQDFL 33
E23CC78A            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYC GSCAIDQDFL 33
C26SC78A            -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYE GSSAIDQDFL 33
H17CE23CC78A        -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTCADLHYC GSCAIDQDFL 33
D19CE23CC78A        -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHACLHYC GSCAIDQDFL 33
E23CA62CC78A        -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYC GSCAIDQDFL 33
E23CC26SC78A        -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYC GSSAIDQDFL 33
H17CE23CC26SC78A    -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTCADLHYC GSSAIDQDFL 33
D19CE23CC26SC78A    -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHACLHYC GSSAIDQDFL 33
E23CC26SA62CC78A    -17   MRGSHHHHHH GLVPRGSMIR TMLQGKLHRV KVTHADLHYC GSSAIDQDFL 33


WT                   34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
H17C                 34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
D19C                 34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
E23C                 34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
C26S                 34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
A62C                 34   DAAGILENEA IDIWNVTNGK RFSTYAIACE RGSRIISVNG AAAHCASVGD 83
C78A                 34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
H17CE23C             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
H17CC26S             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
H17CC78A             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
D19CE23C             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
D19CC26S             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHCASVGD 83
E23CA62C             34   DAAGILENEA IDIWNVTNGK RFSTYAIACE RGSRIISVNG AAAHCASVGD 83
E23CC78A             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
C26SC78A             34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
H17CE23CC78A         34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
D19CE23CC78A         34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
E23CA62CC78A         34   DAAGILENEA IDIWNVTNGK RFSTYAIACE RGSRIISVNG AAAHAASVGD 83
E23CC26SC78A         34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
H17CE23CC26SC78A     34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
D19CE23CC26SC78A     34   DAAGILENEA IDIWNVTNGK RFSTYAIAAE RGSRIISVNG AAAHAASVGD 83
E23CC26SA62CC78A     34   DAAGILENEA IDIWNVTNGK RFSTYAIACE RGSRIISVNG AAAHAASVGD 83


WT                   84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
H17C                 84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
D19C                 84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
E23C                 84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
C26S                 84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
A62C                 84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
C78A                 84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
H17CE23C             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
H17CC26S             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
H17CC78A             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
D19CE23C             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
D19CC26S             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
E23CA62C             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
E23CC78A             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
C26SC78A             84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
H17CE23CC78A         84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
D19CE23CC78A         84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
E23CA62CC78A         84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
E23CC26SC78A         84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
H17CE23CC26SC78A     84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
D19CE23CC26SC78A     84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
E23CC26SA62CC78A     84   IVIIASFVTM PDEEARTWRP NVAYFEGDNE MKRTAKAIPV QVA 126
```

10 References

ADAMCZYK, A. J., CAO, J., KAMERLIN, S. C. L. & WARSHEL, A. 2011. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. U.S.A.,* 108**,** 14115-14120.

ADAMS, M. D., WAGNER, L. M., GRADDIS, T. J., LANDICK, R., ANTONUCCI, T. K., GIBSON, A. L. & OXENDER, D. L. 1990. Nucleotide sequence and genetic characterization reveal six essential genes for the LIV-I and LS transport systems of Escherichia coli. *J. Biol. Chem.,* 265**,** 11436-11443.

AGARWAL, P. K., BILLETER, S. R., RAJAGOPALAN, P. T. R., BENKOVIC, S. J. & HAMMES-SCHIFFER, S. 2002. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.,* 99**,** 2794-2799.

AKITA, S., UMEZAWA, N. & HIGUCHI, T. 2005. On-bead fuorescence assay for serine/threonine kinases. *Org. Lett.,* 7**,** 5565-5568.

AKIYAMA, S., TAKAHASHI, S., KIMURA, T., ISHIMORI, K., MORISHIMA, I., NISHIKAWA, Y. & FUJISAWA, T. 2002. Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U.S.A.,* 99**,** 1329-1334.

ALBERT, A., DHANARAJ, V., GENSCHEL, U., KHAN, G., RAMJEE, M. K., PULIDO, R., SIBANDA, B. L., VON DELFT, F., WITTY, M., BLUNDELL, T. L., SMITH, A. G. & ABELL, C. 1998. Crystal structure of aspartate decarboxylase at 2.2 Å resolution provides evidence for an ester in protein self-processing. *Nat. Struct. Mol. Biol.,* 5**,** 289-293.

ALLAN, A. C., WARD, J. L., BEALE, M. H. & TREWAVAS, A. J. 1998. Caged plant growth regulators. *Methods Enzymol.,* 291**,** 474-483.

ANDERSON, A. C. 2003. The process of structure-based drug design. *Chem. & Biol.,* 10**,** 787-97.

BANG, D. & KENT, S. B. H. 2004. A one-pot total synthesis of crambin. *Angew. Chem. Int. Ed. ,* 43**,** 2534-2538.

BAXTER, R. H. G., PONOMARENKO, N., ŠRAJER, V., PAHL, R., MOFFAT, K. & NORRIS, J. R. 2004. Time-resolved crystallographic studies of light-induced structural changes in the photosynthetic reaction center. *Proc. Natl. Acad. Sci. U.S.A.,* 101**,** 5982-5987.

BELLEMIN-LAPONNAZ, S., TWEDDELL, J., RUBLE, J. C., BREITLING, F. M. & FU, G. C. 2000. The kinetic resolution of allylic alcohols by a non-enzymatic acylation catalyst; application to natural product synthesis. *Chem. Comm.***,** 1009-1010.

BENSASSON, R. V., LAND, E. J. & TRUSCOTT, T. G. 1993. *Excited states and free radicals in biology and medicine: Contributions from flash photolysis and pulse radiolysis*, Oxford University Press.

BHABHA, G., LEE, J., EKIERT, D. C., GAM, J., WILSON, I. A., DYSON, H. J., BENKOVIC, S. J. & WRIGHT, P. E. 2011. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science,* 332**,** 234-238.

BLANCO-CANOSA, J. B. & DAWSON, P. E. 2008. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew. Chem. Int. Ed. ,* 47**,** 6851-6855.

BOEHR, D. D., NUSSINOV, R. & WRIGHT, P. E. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.,* 5**,** 789-796.

BOURGEOIS, D., VALLONE, B., ARCOVITO, A., SCIARA, G., SCHOTTE, F., ANFINRUD, P. A. & BRUNORI, M. 2006. Extended subnanosecond structural dynamics of myoglobin revealed by Laue crystallography. *Proc. Natl. Acad. Sci. U.S.A.,* 103**,** 4924-9.

BOWLER, M. W., MUELLER, U., WEISS, M. S., SANCHEZ-WEATHERBY, J., SORENSEN, T. L. M., THUNNISSEN, M. M. G. M., URSBY, T., GOBBO, A., RUSSI, S., BOWLER, M. G., BROCKHAUSER, S., SVENSSON, O. & CIPRIANI, F. 2015. Automation and experience of controlled crystal dehydration: results from the european synchrotron HC1 collaboration. *Cryst. Growth Des.,* 15**,** 1043-1054.

BRAGG, W. L. 1913. The structure of some crystals as indicated by their diffraction of X-rays. *Proc. Math. Phys. Eng. Sci.,* 89**,** 248-277.

BULLER, A. R., FREEMAN, M. F., WRIGHT, N. T., SCHILDBACK, J. F. & TOWNSEND, C. A. 2012. Insights into cis-autoproteolysis reveal a reactive state formed through conformational rearrangement. *Proc. Natl Acad. Sci. U.S.A.,* 109**,** 2308-2313.

CALLAWAY, E. M. & YUSTE, R. 2002. Stimulating neurons with light. *Curr. Op. Neurobiol.,* 12**,** 587-592.

CHALKER, J. M., GUNNOO, S. B., BOUTUREIRA, O., GERSTBERGER, S. C., FERNANDEZ-GONZALEZ, M., BERNARDES, G. J. L., GRIFFIN, L., HAILU, H., SCHOFIELD, C. J. & DAVIS, B. G. 2011. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.,* 2**,** 1666-1676.

CHANG, C.-Y., FERNANDEZ, T., PANCHAL, R. & BAYLEY, H. 1998. Caged catalytic subunit of cAMP-dependent protein kinase. *J. Amer. Chem. Soc.,* 120**,** 7661-7662.

CHATTERJEE, A. K., CHOI, T.-L., SANDERS, D. P. & GRUBBS, R. H. 2003. A general model for selectivity in olefin cross metathesis. *J. Amer. Chem. Soc.,* 125**,** 11360-11370.

CHEN, B., SYSOEVA, TATYANA A., CHOWDHURY, S. & NIXON, B. T. 2008. Regulation and action of the bacterial enhancer-binding protein AAA+ domains. *Biochem. Soc. Trans.,* 36**,** 89-93.

CHEN, P. & XIANG, P. 2011. Kinetic resolution of allylic alcohols via stereoselective acylation catalyzed by lipase PS-30. *Tet. Lett.,* 52**,** 5758-5760.

CHOSHI, T., HORIMOTO, S., WANG, C. Y., NAGASE, H., ICHIKAWA, M., SUGINO, E. & HIBINO, S. 1992. Synthesis of dibenzoylmethane derivatives and inhibition of mutagenicity in Salmonella typhimurium. *Dhem. Pharm. Bull.,* 40**,** 1047-1049.

CONRAD, P. G., GIVENS, R. S., WEBER, J. F. W. & KANDLER, K. 2000. new phototriggers:1 extending the p-hydroxyphenacyl $\pi - \pi *$ absorption range. *Org. Lett.,* 2**,** 1545-1547.

COREY, E. J. & HELAL, C. J. 1998. Reduction of carbonyl compounds with chiral oxazaborolidine catalysts: a new paradigm for enantioselective catalysis and a powerful new synthetic method. *Angew. Chem. Int. Ed. ,* 37**,** 1986-2012.

CORNIL, J., GONNARD, L., GUÉRINOT, A., REYMOND, S. & COSSY, J. 2014. Lewis acid catalyzed synthesis of cyclic carbonates, precursors of 1,2- and 1,3-diols. *Eur. J. Org. Chem.,* 2014**,** 4958-4962.

CORRIE, J. E. T., FURUTA, T., GIVENS, R., YOUSEF, A. L. & GOELDNER, M. 2005. Photoremovable protecting groups used for the caging of biomolecules. *Dynamic Studies in Biology.* Wiley-VCH Verlag GmbH & Co. KGaA.

CORRIE, J. E. T. & TRENTHAM, D. R. 1992. Synthetic, mechanistic and photochemical studies of phosphate esters of substituted benzoins. *J. Chem. Soc. Perkin Trans. 1***,** 2409-2417.

CORRIE, J. E. T. & TRENTHAM, D. R. 1993. *Biological applications of photochemical switches*, John Wiley & Sons.

CORT, J. R., YEE, A., ARROWSMITH, C.H., KENNEDY, M.A. 2009. *RE: Cofactor optimization.*

CRONAN, J. E., JR., LITTEL, K. J. & JACKOWSKI, S. 1982. Genetic and biochemical analyses of pantothenate biosynthesis in Escherichia coli and Salmonella typhimurium. *J. Bacteriol.,* 149**,** 916-22.

DAVIS, L. & CHIN, J. W. 2012. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell. Biol.,* 13**,** 168-182.

DEITERS, A. 2009. Light activation as a method of regulating and studying gene expression. *Curr. Op. Chem. Biol.,* 13**,** 678-686.

DHAKANE, V., CHAVAN, H., THAKARE, V., ADSUL, L., SHRINGARE, S. & BANDGAR, B. 2014. Novel ibuprofen prodrugs with improved pharmacokinetics and non-ulcerogenic potential. *Med. Chem. Res.,* 23**,** 503-517.

DICKINSON, R. G. & RAYMOND, A. L. 1923. The crystal structure of hexamethylene-tetramine. *J. Amer. Chem. Soc.,* 45**,** 22-29.

DODSON, G. & WLODAWER, A. 1998. Catalytic triads and their relatives. *Trends Biochem. Sci.,* 23**,** 347-52.

DONIACH, S. 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem. Rev.,* 101**,** 1763-1778.

DUREK, T., TORBEEV, V. Y. & KENT, S. B. H. 2007. Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme. *Proc. Natl. Acad. Sci. U.S.A.,* 104**,** 4846-4851.

DUTTA, S., MORRISON, E. A. & HENZLER-WILDMAN, K. A. 2014. Blocking dynamics of the SMR transporter EmrE impairs efflux activity. *Biophys. J.,* 107**,** 613-20.

DYER, R. G. & TURNBULL, K. D. 1999. Hydrolytic stabilization of protected p-hydroxybenzyl halides designed as latent quinone methide precursors. *J. Org. Chem.,* 64**,** 7988-7995.

ECKARDT, T., HAGEN, V., SCHADE, B., SCHMIDT, R., SCHWEITZER, C. & BENDIG, J. 2002. Deactivation behavior and excited-state Properties of (coumarin-4-yl)methyl derivatives. 2. photocleavage of selected (coumarin-4-yl)methyl-caged adenosine cyclic 3',5'-monophosphates with fluorescence enhancement. *J. Org. Chem.,* 67**,** 703-710.

EISENMESSER, E. Z., MILLET, O., LABEIKOVSKY, W., KORZHNEV, D. M., WOLF-WATZ, M., BOSCO, D. A., SKALICKY, J. J., KAY, L. E. & KERN, D. 2005. Intrinsic dynamics of an enzyme underlies catalysis. *Nature,* 438**,** 117-121.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Cryst.,* D66**,** 486-501.

ENGELS, J. & SCHLAEGER, E. J. 1977. Synthesis, structure, and reactivity of adenosine cyclic 3',5'-phosphate-benzyltriesters. *J. Med. Chem.,* 20**,** 907-911.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Cryst.,* D62**,** 72-82.

EVANS, P. R. & MURSHUDOV, G. N. 2013. How good are my data and what is the resolution? *Acta Cryst.,* D69**,** 1204-14.

FELD, G. K. & FRANK, M. 2014. Enabling membrane protein structure and dynamics with X-ray free electron lasers. *Curr. Op. Struct. Biol.,* 27**,** 69-78.

FISCHETTI, R. F., RODI, D. J., GORE, D. B. & MAKOWSKI, L. 2004. Wide-angle X-ray solution scattering as a Probe of ligand-Induced conformational changes in proteins. *Chem. & Biol.,* 11**,** 1431-1443.

FORREST, L. R., KRÄMER, R. & ZIEGLER, C. 2011. The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta,* 1807**,** 167-188.

FOURNIER, L., AUJARD, I., LE SAUX, T., MAURIN, S., BEAUPIERRE, S., BAUDIN, J.-B. & JULLIEN, L. 2013. Coumarinylmethyl caging groups with redshifted absorption. *Chem. Eur. J.,* 19**,** 17494-17507.

FRASER, J. S., VAN DEN BEDEM, H., SAMELSON, A. J., LANG, P. T., HOLTON, J. M., ECHOLS, N. & ALBER, T. 2011. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.,* 108**,** 16247-16252.

FURUTA, T. & IWAMURA, M. 1998. New caged groups: 7-substituted coumarinylmethyl phosphate esters. *Methods Enzymol.,* 291**,** 50-63.

FURUTA, T., TAKEUCHI, H., ISOZAKI, M., TAKAHASHI, Y., KANEHARA, M., SUGIMOTO, M., WATANABE, T., NOGUCHI, K., DORE, T. M., KURAHASHI, T., IWAMURA, M. & TSIEN, R. Y. 2004. Bhc-cNMPs as either water-soluble or membrane-permeant photoreleasable cyclic nucleotides for both one- and two-photon excitation. *ChemBioChem,* 5**,** 1119-1128.

FURUTA, T., WANG, S. S.-H., DANTZKER, J. L., DORE, T. M., BYBEE, W. J., CALLAWAY, E. M., DENK, W. & TSIEN, R. Y. 1999. Brominated 7-hydroxycoumarin-4-ylmethyls: Photolabile protecting groups with biologically useful cross-sections for two photon photolysis. *Proc. Natl. Acad. Sci. U.S.A.,* 96**,** 1193-1200.

GALAN, B. R., KALBARCZYK, K. P., SZCZEPANKIEWICZ, S., KEISTER, J. B. & DIVER, S. T. 2007. A rapid and simple cleanup procedure for metathesis reactions. *Org. Lett.,* 9**,** 1203-1206.

GARMAN, E. F. 2010. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Cryst.,* D66**,** 339-51.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S. E., WILKINS, M., APPEL, R. & BAIROCH, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. *In:* WALKER, J. (ed.) *The Proteomics Protocols Handbook.* Humana Press.

GAUTIER, A., DEITERS, A. & CHIN, J. W. 2011. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Amer. Chem. Soc.,* 133**,** 2124-2127.

GEISSLE, D., ANTONENKO, Y. N., SCHMIDT, R., KELLER, S., KRYLOVA, O. O., WIESNER, B., BENDIG, J., POHL, P. & HAGEN, V. 2005. (Coumarin-4-yl)methyl esters as highly efficient, ultrafast phototriggers for protons and their application to acidifying membrane surfaces. *Angew. Chem. Int. Ed. ,* 44**,** 1195-1198.

GELFMAN, C. M., COPELAND, W. C. & ROBERTUS, J. D. 1991. Site-directed alteration of four active-site residues of a pyruvoyl-dependent histidine decarboxylase. *Biochemistry,* 30**,** 1057-62.

GIVENS, R. S., HEGER, D., HELLRUNG, B., KAMDZHILOV, Y., MAC, M., CONRAD, P. G., COPE, E., LEE, J. I., MATA-SEGREDA, J. F., SCHOWEN, R. L. & WIRZ, J. 2008. The photo-favorskii reaction of p-hydroxyphenacyl compounds Is initiated by water-assisted, adiabatic extrusion of a triplet biradical. *J. Amer. Chem. Soc.,* 130**,** 3307-3309.

GIVENS, R. S., JUNG, A., PARK, C.-H., WEBER, J. & BARTLETT, W. 1997. New photoactivated protecting groups. 7. p-hydroxyphenacyl: a phototrigger for excitatory amino acids and peptides1. *J. Amer. Chem. Soc.,* 119**,** 8369-8370.

GIVENS, R. S. & MATUSZEWSKI, B. 1984. Photochemistry of phosphate esters: an efficient method for the generation of electrophiles. *J. Amer. Chem. Soc.,* 106**,** 6860-6861.

GOLD, V. 1948. The hydrolysis of acetic anhydride. *Transact. Faraday Soc.,* 44**,** 506-518.

GOLDMACHER, V. S., SENTER, P. D., LAMBERT, J. M. & BLATTLER, W. A. 1992. Photoactivation of toxin conjugates. *Bioconjugate Chemistry,* 3**,** 104-107.

GOLDMAN, Y. E., HIBBERD, M. G., MCCRAY, J. A. & TRENTHAM, D. R. 1982. Relaxation of muscle fibres by photolysis of caged ATP. *Nature,* 300**,** 701-705.

GRAY, A., HARLEN, O. G., HARRIS, S. A., KHALID, S., LEUNG, Y. M., LONSDALE, R., MULHOLLAND, A. J., PEARSON, A. R., READ, D. J. & RICHARDSON, R. A. 2015. In pursuit of an accurate spatial and temporal model of biomolecules at the atomistic level: a perspective on computer simulation. *Acta Cryst.,* D71**,** 162-172.

HAGEN, V., BENDIG, J., FRINGS, S., ECKARDT, T., HELM, S., REUTER, D. & KAUPP, U. B. 2001. Highly efficient and ultrafast phototriggers for cAMP and cGMP by using long-wavelength UV/Vis-activation. *Angew. Chem. Int. Ed. ,* 40**,** 1045-1048.

HAGEN, V., BENNDORF, K., KAUPP, U. B., PAVLOS, C. M., XU, H., TOSCANO, J. P., HESS, G. P., GILLESPIE, D. C., KIM, G. & KANDLER, K. 2005. Control of cellular activity. *Dynamic Studies in Biology.* Wiley-VCH Verlag GmbH & Co. KGaA.

HANOIAN, P., LIU, C. T., HAMMES-SCHIFFER, S. & BENKOVIC, S. 2015. Perspectives on electrostatics and conformational motions in enzyme catalysis. *Acc. Chem. Res.,* 48**,** 482-489.

HARIGAI, M., IMAMOTO, Y., KAMIKUBO, H., YAMAZAKI, Y. & KATAOKA, M. 2003. Role of an N-terminal loop in the secondary structural change of photoactive yellow protein. *Biochemistry,* 42**,** 13893-900.

HASHIGUCHI, S., FUJII, A., HAACK, K.-J., MATSUMURA, K., IKARIYA, T. & NOYORI, R. 1997. Kinetic resolution of racemic secondary alcohols by RuII-catalyzed hydrogen transfer. *Angew. Chem. Int. Ed. ,* 36**,** 288-290.

HASHIGUCHI, S., FUJII, A., TAKEHARA, J., IKARIYA, T. & NOYORI, R. 1995. Asymmetric transfer hydrogenation of aromatic ketones catalyzed by chiral ruthenium(II) complexes. *J. Amer. Chem. Soc.,* 117**,** 7562-7563.

HENZLER-WILDMAN, K. & KERN, D. 2007. Dynamic personalities of proteins. *Nature,* 450**,** 964-972.

HENZLER-WILDMAN, K. A., LEI, M., THAI, V., KERNS, S. J., KARPLUS, M. & KERN, D. 2007a. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature,* 450**,** 913-916.

HENZLER-WILDMAN, K. A., THAI, V., LEI, M., OTT, M., WOLF-WATZ, M., FENN, T., POZHARSKI, E., WILSON, M. A., PETSKO, G. A., KARPLUS, M., HUBNER, C. G. & KERN, D. 2007b. Intrinsic motions along an enzymatic reaction trajectory. *Nature,* 450**,** 838-844.

HERMANSON, G. T. 2008a. Chapter 1 - Functional targets. *In:* HERMANSON, G. T. (ed.) *Bioconjugate Techniques (Second Edition).* New York: Academic Press.

HERMANSON, G. T. 2008b. Chapter 4 - Homobifunctional crosslinkers. *In:* HERMANSON, G. T. (ed.) *Bioconjugate Techniques (Second Edition).* New York: Academic Press.

HERMANSON, G. T. 2008c. Chapter 5 - Heterobifunctional crosslinkers. *In:* HERMANSON, G. T. (ed.) *Bioconjugate Techniques (Second Edition).* New York: Academic Press.

HOLMES, C. P. 1997. Model studies for new o-nitrobenzyl photolabile linkers: substituent effects on the rates of photochemical cleavage. *J. Org. Chem.,* 62**,** 2370-2380.

HORI, T., MORIYAMA, H., KAWAGUCHI, J., HAYASHI-IWASAKI, Y., OSHIMA, T. & TANAKA, N. 2000. The initial step of the thermal unfolding of 3-

isopropylmalate dehydrogenase detected by the temperature-jump Laue method. *Protein Eng.,* 13**,** 527-33.

HOUTMAN, J. C., BROWN, P. H., BOWDEN, B., YAMAGUCHI, H., APPELLA, E., SAMELSON, L. E. & SCHUCK, P. 2007. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Prot. Sci.,* 16**,** 30-42.

HUYNH, T. H. V., ABRAHAMSEN, B., MADSEN, K. K., GONZALEZ-FRANQUESA, A., JENSEN, A. A. & BUNCH, L. 2012. Design, synthesis and pharmacological characterization of coumarin-based fluorescent analogs of excitatory amino acid transporter subtype 1 selective inhibitors, UCPH-101 and UCPH-102. *Bioorg. Med. Chem.,* 20**,** 6831-6839.

IHEE, H., RAJAGOPAL, S., SRAJER, V., PAHL, R., ANDERSON, S., SCHMIDT, M., SCHOTTE, F., ANFINRUD, P. A., WULFF, M. & MOFFAT, K. 2005. Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc. Natl. Acad. Sci. U.S.A.,* 102**,** 7145-50.

IL'ICHEV, Y. V., SCHWÖRER, M. A. & WIRZ, J. 2004. Photochemical reaction mechanisms of 2-nitrobenzyl compounds: methyl ethers and caged ATP. *J. Amer. Chem. Soc.,* 126**,** 4581-4595.

INCARDONA, M.-F., BOURENKOV, G. P., LEVIK, K., PIERITZ, R. A., POPOV, A. N. & SVENSSON, O. 2009. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synch. Rad.,* 16**,** 872-879.

JACKOWSKI, S. & ROCK, C. O. 1981. Regulation of coenzyme A biosynthesis. *J. Bacteriol.,* 148**,** 926-932.

JIANG, J., SHRIVASTAVA, I. H., WATTS, S. D., BAHAR, I. & AMARA, S. G. 2011. Large collective motions regulate the functional properties of glutamate transporter trimers. *Proc. Natl. Acad. Sci. U.S.A.,* 108**,** 15141-15146.

JOHANNSON, D. G., WALLIN, G., SANDBERG, A., MACAO, B., ÅQVIST, J. & HÄRD, T. 2009. Protein autoproteolysis: conformation strain linked to the rate of peptide cleavage by the pH dependence of the N-O acyl shift. *J. Amer. Chem. Soc.,* 131**,** 9475-9477.

JOHANSSON, D. G. A., MACAO, B., SANDBERG, A. & HÄRD, T. 2008. SEA domain autoproteolysis accelerated by conformational Strain: mechanistic Aspects. *J. Mol. Biol.,* 377**,** 1130-1143.

JOHNSON, J. A., LU, Y. Y., VAN DEVENTER, J. A. & TIRRELL, D. A. 2010. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Op. Chem. Biol.,* 14**,** 774-780.

JOHNSON, R. A., SHARPLESS, K. B. & OJIMA, I. 1993. Catalytic asymmetric synthesis. *VCH, New York***,** 227-272.

JUNG, Y. O., LEE, J. H., KIM, J., SCHMIDT, M., MOFFAT, K., ŠRAJER, V. & IHEE, H. 2013. Volume-conserving trans–cis isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography. *Nat. Chem.,* 5**,** 212-220.

KABSCH, W. 2010. XDS. *Acta Cryst.,* D66**,** 125-132.

KANTARDJIEFF, K. A. & RUPP, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.,* 12**,** 1865-1871.

KAPLAN, J. H., FORBUSH, B. & HOFFMAN, J. F. 1978. Rapid photolytic release of adenosine 5'-triphosphate from a protected analog: utilization by the sodium:potassium pump of human red blood cell ghosts. *Biochemistry,* 17**,** 1929-1935.

KAVALA, V. & PATEL, BHISMA K. 2005. Reinvestigation of the mechanism of gem-diacylation: chemoselective conversion of aldehydes to various gem-diacylates and their cleavage under acidic and basic conditions. *Eur. J. Org. Chem.,* 2005**,** 441-451.

KAY, L. E. 2005. NMR studies of protein structure and dynamics. *J. Magn. Reson.,* 173**,** 193-207.

KELLER, S., VARGAS, C., ZHAO, H., PISZCZEK, G., BRAUTIGAM, C. A. & SCHUCK, P. 2012. High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.,* 84**,** 5066-73.

KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R. G., WYCKOFF, H. & PHILLIPS, D. C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature,* 181**,** 662-6.

KHOURY, G. A., SMADBECK, J., KIESLICH, C. A. & FLOUDAS, C. A. 2013. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechn.,* 32**,** 99-109.

KIM, J. G., KIM, T. W., KIM, J. & IHEE, H. 2015. Protein structural dynamics revealed by time-resolved X-ray solution scattering. *Acc. Chem. Res.*

KIM, J. K., YANG, I. S., SHIN, H. Y., CHO, K. J., RYU, E. K., KIM, S. H., PARK, S. S. & KIM, K. H. 2006. Insight into autoproteolytic activation from the structure of cephalosporin acylase: a protein with two proteolytic chemistries. *Proc. Natl Acad. Sci. U.S.A.,* 103**,** 1732-1737.

KINBARA, K., GOLDFINGER, L. E., HANSEN, M., CHOU, F.-L. & GINSBERG, M. H. 2003. Ras GTPases: integrins' friends or foes? *Nat. Rev. Mol. Cell. Biol.,* 4**,** 767-778.

KLÁN, P., ŠOLOMEK, T., BOCHET, C. G., BLANC, A., GIVENS, R., RUBINA, M., POPIK, V., KOSTIKOV, A. & WIRZ, J. 2013. Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy. *Chem. Rev.,* 113**,** 119-191.

KLÍČOVÁ, Ĺ., ŠEBEJ, P., ŠOLOMEK, T., HELLRUNG, B., SLAVÍČEK, P., KLÁN, P., HEGER, D. & WIRZ, J. 2012. Adiabatic triplet state tautomerization of p-hydroxyacetophenone in aqueous solution. *J. Phys. Chem. A,* 116**,** 2935-2944.

KLINMAN, J. P. 2013. Importance of protein dynamics during enzymatic C–H bond cleavage catalysis. *Biochemistry,* 52**,** 2068-2077.

KLINMAN, J. P. & KOHEN, A. 2013. Hydrogen tunneling links protein dynamics to enzyme catalysis. *Annu. Rev. Biochem.,* 82**,** 471-496.

KNESL, P., RÖSELING, D. & JORDIS, U. 2006. Improved synthesis of substituted 6,7-dihydroxy-4-quinazolineamines: tandutinib, erlotinib and gefitinib. *Molecules,* 11**,** 286.

KÖNIGS, P., NEUMANN, O., HACKELÖER, K., KATAEVA, O. & WALDVOGEL, S. R. 2008. Versatile one-pot synthesis of 3-alkenylcoumarins. *Eur. J. Org. Chem.,* 2008**,** 343-349.

KONST, P. M., FRANSSEN, M. C. R., SCOTT, E. L. & SANDERS, J. P. M. 2009. A study on the applicability of l-aspartate [small alpha]-decarboxylase in the biobased production of nitrogen containing chemicals. *Green Chem.,* 11**,** 1646-1652.

KONUMA, T., KIMURA, T., MATSUMOTO, S., GOTO, Y., FUJISAWA, T., FERSHT, A. R. & TAKAHASHI, S. 2011. Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. *J. Mol. Biol.,* 405**,** 1284-1294.

KORT, R., HELLINGWERF, K. J. & RAVELLI, R. B. 2004. Initial events in the photocycle of photoactive yellow protein. *J. Biol. Chem.,* 279**,** 26417-24.

KOSSEL, A. H., CAMBRIDGE, S. B., WAGNER, U. & BONHOEFFER, T. 2001. A caged Ab reveals an immediate/instructive effect of BDNF during hippocampal synaptic potentiation. *Proc. Natl. Acad. Sci. U.S.A.,* 98**,** 14702-14707.

KUBELKA, J. 2009. Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem. Photobiol. Sci.,* 8**,** 499-512.

LANG, K. & CHIN, J. W. 2014. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.,* 114**,** 4764-4806.

LATA CHANDRA, K., SARAVANAN, P. & INGH, V. K. 2000. An efficient method for diacetylation of aldehydes. *Synlett,* 2000**,** 359-360.

LEBEDEV, A. A. 2013. Twinning and other pathologies.

LEBEDEV, A. A. & ISUPOV, M. N. 2014. Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Cryst.,* D70**,** 2430-2443.

LEE, H.-M., LARSON, D. R. & LAWRENCE, D. S. 2009. Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS Chem. Biol.,* 4**,** 409-427.

LEE, S., WANG, W., LEE, Y. & SAMPSON, N. S. 2015. Cyclic acetals as cleavable linkers for affinity capture. *Org. Biomol. Chem.,* 13**,** 8445-8452.

LEIPER, J. & NANDI, M. 2011. The therapeutic potential of targeting endogenous inhibitors of nitric oxide synthesis. *Nat. Rev. Drug Discov.,* 10**,** 277-91.

LEITNER, A., WALZTHOENI, T., KAHRAMAN, A., HERZOG, F., RINNER, O., BECK, M. & AEBERSOLD, R. 2010. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics,* 9**,** 1634-49.

LESTER, H. A. & NERBONNE, J. M. 1982. Physiological and pharmacological manipulations with light flashes. *Annu. Rev. Biophys. Bioeng.,* 11**,** 151-175.

LEVANTINO, M., YORKE, B. A., MONTEIRO, D. C. F., CAMMARATA, M. & PEARSON, ARWEN R. 2015a. Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. *Curr. Op. Struct. Biol.,* in press.

LEVANTINO, M., SCHIRÒ, G., LEMKE, H. T., COTTONE, G., GLOWNIA, J. M., ZHU, D., CHOLLET, M., IHEE, H., CUPANE, A. & CAMMARATA, M. 2015b. Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat. Commun.,* 6.

LEVITIN, F., STERN, O., WEISS, M., GIL-HENN, C., ZIV, R., PROLOCIMER, Z., SMORODINSKY, N. I., RUBINSTEIN, D. B. & WRESCHNER, D. H. 2005. The MUC1 SEA module is a self-cleaving domain. *J. Biol. Chem.,* 280**,** 33374-33386.

LI, T., SUN, F., JI, X., FENG, Y. & RAO, Z. 2003. Structure based hyperthermostability of archaeal histone HPhA from Pyrococcus horikoshii. *J. Mol. Biol.,* 325**,** 1031-7.

LIN, W., ALBANESE, C., PESTELL, R. G. & LAWRENCE, D. S. 2002. Spatially discrete, light-driven protein expression. *Chem. & Biol.,* 9**,** 1347-1353.

LINK, K. H., CRUZ, F. G., YE, H.-F., O'REILLY, K. E., DOWDELL, S. & KOH, J. T. 2004. Photo-caged agonists of the nuclear receptors RARγ and TRβ provide unique time-dependent gene expression profiles for light-activated gene patterning. *Bioorg. Med. Chem.,* 12**,** 5949-5959.

LOUDWIG, S., NICOLET, Y., MASSON, P., FONTECILLA-CAMPS, J. C., BON, S., NACHON, F. & GOELDNER, M. 2003. Photoreversible Inhibition of Cholinesterases: Catalytic Serine-Labeled Caged Butyrylcholinesterase. *ChemBioChem,* 4**,** 762-767.

LOVELL, S. C., DAVIS, I. W., ARENDALL, W. B., 3RD, DE BAKKER, P. I., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S. & RICHARDSON, D. C. 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins,* 50**,** 437-50.

LUK, L. Y. P., LOVERIDGE, E. J. & ALLEMANN, R. K. 2015. Protein motions and dynamic effects in enzyme catalysis. *Phys. Chem. Chem. Phys.,* Advance article.

LYOTHIER, I., DEFIEBER, C. & CARREIRA, E. M. 2006. Iridium-catalyzed enantioselective synthesis of allylic alcohols: silanolates as hydroxide equivalents. *Angew. Chem. Int. Ed. ,* 45**,** 6204-6207.

MAAS, W. K. & VOGEL, H. J. 1953. α-Ketoisovaleric acid, a precursor of pantothenic acid in escherichia coli. *J. Bacteriol.,* 65**,** 388-393.

MALHOTRA, K., SATHAPPA, M., LANDIN, J. S., JOHNSON, A. E. & ALDER, N. N. 2013. Structural changes in the mitochondrial Tim23 channel are coupled to the proton-motive force. *Nat. Struct. Mol. Biol.,* 20**,** 965-972.

MALLAVADHANI, U. V. & RAO, Y. R. 1994. Enantioselective synthesis of cinnamyl-1-phenyl-2-propenyl ether: A metabolite of marine green algal species caulerpa racemosa. *Tet. Asymm.,* 5**,** 23-26.

MAYER, G. & HECKEL, A. 2006. Biologically active molecules with a "light switch". *Angew. Chem. Int. Ed. ,* 45**,** 4900-4921.

MAYNARD, H. D. & GRUBBS, R. H. 1999. Purification technique for the removal of ruthenium from olefin metathesis reaction products. *Tet. Lett.,* 40**,** 4137-4140.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J. Appl. Cryst.,* 40**,** 658-674.

MCCRAY, J. A., HERBETTE, L., KIHARA, T. & TRENTHAM, D. R. 1980. A new approach to time-resolved studies of ATP-requiring biological systems; laser flash photolysis of caged ATP. *Proc. Natl. Acad. Sci. U.S.A.,* 77**,** 7237-7241.

MCDONALD, J. P., FRANK, E. G., LEVINE, A. S. & WOODGATE, R. 1998. Intermolecular cleavage by UmuD-like mutagenesis proteins. *Proc. Natl Acad. Sci. U.S.A.,* 95**,** 1478-1483.

MCGOWAN, L. C. & HAMELBERG, D. 2013. Conformational plasticity of an enzyme during catalysis: intricate coupling between cyclophilin A dynamics and substrate turnover. *Biophys J,* 104**,** 216-26.

MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nuc. Acids Res.,* 41**,** W597-W600.

MENDEL, D., ELLMAN, J. A. & SCHULTZ, P. G. 1991. Construction of a light-activated protein by unnatural amino acid mutagenesis. *J. Amer. Chem. Soc.,* 113**,** 2758-2760.

MEYER, T. E., YAKALI, E., CUSANOVICH, M. A. & TOLLIN, G. 1987. Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry,* 26**,** 418-423.

MOLNÁR, P. & NADLER, J. V. 2000. γ-Aminobutyrate, α-carboxy-2-nitrobenzyl ester selectively blocks inhibitory synaptic transmission in rat dentate gyrus. *Eur. J. Pharm.,* 391**,** 255-262.

MONTEIRO, DIANA C. F., PATEL, V., BARTLETT, CHRISTOPHER P., NOZAKI, S., GRANT, THOMAS D., GOWDY, JAMES A., THOMPSON, GARY S., KALVERDA, ARNOUT P., SNELL, EDWARD H., NIKI, H., PEARSON, ARWEN R. & WEBB, MICHAEL E. 2015. The structure of the PanD/PanZ protein complex reveals negative feedback regulation of pantothenate biosynthesis by coenzyme A. *Chem. & Biol.,* 22**,** 492-503.

MONTEIRO, D. C. F., RUGEN, M. D., SHEPHERD, D., NOZAKI, S., NIKI, H. & WEBB, M. E. 2012. Formation of a heterooctameric complex between aspartate alpha-decarboxylase and its cognate activating factor, PanZ, is CoA-dependent. *Biochem. Biophys. Res. Commun.,* 426**,** 350-5.

MOURADOV, D., KING, G., ROSS, I. L., FORWOOD, J. K., HUME, D. A., SINZ, A., MARTIN, J. L., KOBE, B. & HUBER, T. 2008. Protein structure determination using a combination of cross-linking, mass spectrometry, and molecular modeling. *Methods Mol. Biol.,* 426**,** 459-74.

MURRAY-RUST, J., LEIPER, J., MCALISTER, M., PHELAN, J., TILLEY, S., SANTA MARIA, J., VALLANCE, P. & MCDONALD, N. 2001. Structural insights into the hydrolysis of cellular nitric oxide synthase inhibitors by dimethylarginine dimethylaminohydrolase. *Nat. Struct. Biol.,* 8**,** 679-83.

MURSHUDOV, G. N., SKUBAK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.,* D67**,** 355-367.

NIELSEN, I. B., BOYÉ-PÉRONNE, S., EL GHAZALY, M. O. A., KRISTENSEN, M. B., BRØNDSTED NIELSEN, S. & ANDERSEN, L. H. 2005. Absorption spectra of photoactive yellow protein chromophores in vacuum. *Biophys. J.,* 89**,** 2597-2604.

NISHIBAYASHI, Y., TAKEI, I., UEMURA, S. & HIDAI, M. 1999. Extremely high enantioselective redox reaction of ketones and alcohols catalyzed by RuCl2(PPh3)(oxazolinylferrocenylphosphine). *Organometallics,* 18**,** 2291-2293.

NOZAKI, S., WEBB, M. E. & NIKI, H. 2012. An activator for pyruvoyl-dependent l-aspartate α-decarboxylase is conserved in a small group of the γ-proteobacteria including Escherichia coli. *MicrobiologyOpen,* 1**,** 298-310.

O'NEILL, P., STEVENS, D. L. & GARMAN, E. 2002. Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical prespective. *J. Synch. Rad.,* 9**,** 329-332.

OLAH, G. A. & MEHROTRA, A. K. 1982. Catalysis by solid superacids; 161. improved nafion-H catalyzed preparation of 1,1-diacetates from aldehydes. *Synthesis,* 1982**,** 962-963.

OMRAN, Z. & SPECHT, A. 2009a. Short-length dimethoxynitrophenyl photo-cleavable crosslinkers, synthesis and photolysis. *J. Photochem. Photobiol. A,* 208**,** 125-130.

OMRAN, Z. & SPECHT, A. 2009b. Synthesis and photochemical properties of photo-cleavable crosslinkers. *Tet. Lett.,* 50**,** 2434-2436.

OTWINOWSKI, Z. & MINOR, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.,* 276**,** 307-326.

PADILLA, J. E. & YEATES, T. O. 2003. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Cryst.,* D59**,** 1124-30.

PAPWORTH, C., BAUER, J. C., BRAMAN, J. & WRIGHT, D. A. 1996. Site-directed mutagenesis in one day with >80% efficiency. *Strategies,* 9**,** 3-4.

PAULUS, H. 2000. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.,* 69**,** 447-496.

PEGG, A. E. 2009. S-Adenosylmethionine decarboxylase. *Essays Biochem.,* 46**,** 25-46.

PISANI, L., CATTO, M., NICOLOTTI, O., GROSSI, G., DI BRACCIO, M., SOTO-OTERO, R., MENDEZ-ALVAREZ, E., STEFANACHI, A., GADALETA, D. & CAROTTI, A. 2013. Fine molecular tuning at position 4 of 2H-chromen-2-one derivatives in the search of potent and selective monoamine oxidase B inhibitors. *Eur. J. Med. Chem.,* 70**,** 723-739.

POCKER, Y., DAVISON, B. L. & DEITS, T. L. 1978. Decarboxylation of monosubstituted derivatives of carbonic acid. Comparative studies of water- and acid-catalyzed decarboxylation of sodium alkyl carbonates in water and water-d2. *J. Amer. Chem. Soc.,* 100**,** 3564-3567.

POLLACK, L. & DONIACH, S. 2009. Time-resolved X-ray scattering and RNA folding. *Methods Enzymol.,* 469**,** 253-68.

PRIMERANO, D. A. & BURNS, R. O. 1983. Role of acetohydroxy acid isomeroreductase in biosynthesis of pantothenic acid in Salmonella typhimurium. *J. Bacteriol.,* 153**,** 259-69.

RAHMAN, M. A. F. M. & JAHNG, Y. 2007. Preparation of geminal diacylates (acylals) of aldehydes – scope and reactivity of aldehydes with acid anhydrides. *Eur. J. Org. Chem.,* 2007**,** 379-383.

RAMJEE, M. K., GENSCHEL, U., ABELL, C. & SMITH, A. G. 1997. Escherichia coli L-aspartate alpha-decarboxylase: preprotein processing and observation of reaction intermediates by electrospray mass spectrometry. *Biochem. J.,* 323**,** 661-669.

RASHEED, M., RICHTER, C., CHISTY, L. T., KIRKPATRICK, J., BLACKLEDGE, M., WEBB, M. R. & DRISCOLL, P. C. 2014. Ligand-dependent dynamics of the active-site lid in bacterial dimethylarginine dimethylaminohydrolase. *Biochemistry,* 53**,** 1092-1104.

RAVELLI, R. B. G. & MCSWEENEY, S. M. 2000. The 'fingerprint' that X-rays can leave on structures. *Structure,* 8**,** 315-328.

REICHMANIS, E., SMITH, B. C. & GOODEN, R. 1985. o-Nitrobenzyl photochemistry: Solution vs. solid-state behavior. *J. Polym. Sci.,* 23**,** 1-8.

REN, W., JI, A. & AI, H.-W. 2015. Light activation of protein splicing with a photocaged fast Intein. *J. Amer. Chem. Soc.,* 137**,** 2155-2158.

REN, Z., BOURGEOIS, D., HELLIWELL, J. R., MOFFAT, K., SRAJER, V. & STODDARD, B. L. 1999. Laue crystallography: coming of age. *J. Synch. Rad.,* 6**,** 891-917.

REN, Z., SRAJER, V., KNAPP, J. E. & ROYER, W. E., JR. 2012. Cooperative macromolecular device revealed by meta-analysis of static and time-resolved structures. *Proc. Natl. Acad. Sci. U.S.A.,* 109**,** 107-12.

ROCK, C. O., PARK, H.-W. & JACKOWSKI, S. 2003. Role of feedback regulation of pantothenate kinase (CoaA) in control of coenzyme A levels in Escherichia coli. *J. Bacteriol.,* 185**,** 3410-3415.

ROGGEN, M. & CARREIRA, E. M. 2011. Enantioselective allylic etherification: selective coupling of two unactivated alcohols. *Angew. Chem. Int. Ed. ,* 50**,** 5568-5571.

RUBLE, J. C., LATHAM, H. A. & FU, G. C. 1997. Effective kinetic resolution of secondary alcohols with a planar−chiral analogue of 4-(dimethylamino)pyridine. use of the Fe(C5Ph5) group in asymmetric catalysis. *J. Amer. Chem. Soc.,* 119**,** 1492-1493.

SALDANHA, S. A., BIRCH, L. M., WEBB, M. E., NABBS, B. K., VON DELFT, F., SMITH, A. G. & ABELL, C. 2001. Identification of Tyr58 as the proton donor in the aspartate alpha-decarboxylase reaction. *Chem. Comm.,* 18**,** 1760-1761.

SAMBROOK, J. & RUSSELL, D. 2001. *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press.

SAMISH, I., MACDERMAID, C. M., PEREZ-AGUILAR, J. M. & SAVEN, J. G. 2011. Theoretical and computational protein design. *Annu. Rev. Phys. Chem.,* 62**,** 129-149.

SANCHEZ-WEATHERBY, J., BOWLER, M. W., HUET, J., GOBBO, A., FELISAZ, F., LAVAULT, B., MOYA, R., KADLEC, J., RAVELLI, R. B. G. &

CIPRIANI, F. 2009. Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Cryst.,* D65**,** 1237-1246.

SANDBERG, A., JOHANSSON, D. G. A., MACAO, B. & HÄRD, T. 2008. SEA domain autoproteolysis accelerated by conformational strain: energetic aspects. *J. Mol. Biol.,* 377**,** 1117-1129.

SAWANT, K. B. & JENNINGS, M. P. 2006. Efficient total syntheses and structural verification of both diospongins A and B via a common δ -lactone intermediate. *J. Org. Chem.,* 71**,** 7911-7914.

SAWAYA, M. R. & KRAUT, J. 1997. Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry,* 36**,** 586-603.

SCHADE, B., HAGEN, V., SCHMIDT, R., HERBRICH, R., KRAUSE, E., ECKARDT, T. & BENDIG, J. 1999. Deactivation behavior and excited-state properties of (coumarin-4-yl)methyl derivatives. 1. photocleavage of (7-methoxycoumarin-4-yl)methyl-caged acids with fluorescence enhancement. *J. Org. Chem.,* 64**,** 9109-9117.

SCHLICHTING, I., ALMO, S. C., RAPP, G., WILSON, K., PETRATOS, K., LENTFER, A., WITTINGHOFER, A., KABSCH, W., PAI, E. F., PETSKO, G. A. & GOODY, R. S. 1990. Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature,* 345**,** 309-315.

SCHLICHTING, I., RAPP, G., JOHN, J., WITTINGHOFER, A., PAI, E. F. & GOODY, R. S. 1989. Biochemical and crystallographic characterization of a complex of c-Ha-ras p21 and caged GTP with flash photolysis. *Proc. Natl. Acad. Sci. U.S.A.,* 86**,** 7687-90.

SCHMIDT, M. 2013. Mix and inject: reaction initiation by diffusion for time-resolved macromolecular crystallography. *Adv. Cond. Matter Phys.,* 2013**,** 10.

SCHMIDT, R., GEISSLER, D., HAGEN, V. & BENDIG, J. 2007. Mechanism of photocleavage of (coumarin-4-yl)methyl esters. *J. Phys. Chem. A,* 111**,** 5768-5774.

SCHMIERER, T., LAIMGRUBER, S., HAISER, K., KIEWISCH, K., NEUGEBAUER, J. & GILCH, P. 2010. Femtosecond spectroscopy on the photochemistry of ortho-nitrotoluene. *Phys. Chem. Chem. Phys.,* 12**,** 15653-15664.

SCHMITZBERGER, F., KILKENNY, M. L., LOBLEY, C. M., WEBB, M. E., VINKOVIC, M., MATAK-VINKOVIC, D., WITTY, M., CHIRGADZE, D. Y., SMITH, A. G., ABELL, C. & BLUNDELL, T. L. 2003. Structural constraints on protein self-processing in L-aspartate-alpha-decarboxylase. *EMBO J.,* 22**,** 6193-204.

SCHNELL, J. R., DYSON, H. J. & WRIGHT, P. E. 2004. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu. Rev. Biophys. Biomol. Struct.,* 33**,** 119-140.

SCHÖNLEBER, R. O., BENDIG, J., HAGEN, V. & GIESE, B. 2002. Rapid photolytic release of cytidine $5'$ -diphosphate from a coumarin derivative: a new tool for the investigation of ribonucleotide reductases. *Bioorg. Med. Chem.,* 10**,** 97-101.

SCHOTTE, F., CHO, H. S., KAILA, V. R., KAMIKUBO, H., DASHDORJ, N., HENRY, E. R., GRABER, T. J., HENNING, R., WULFF, M., HUMMER, G., KATAOKA, M. & ANFINRUD, P. A. 2012. Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proc. Natl. Acad. Sci. U.S.A.,* 109**,** 19256-61.

SCHOTTE, F., CHO, H. S., SOMAN, J., WULFF, M., OLSON, J. S. & ANFINRUD, P. A. 2013. Real-time tracking of CO migration and binding in the α and β

subunits of human hemoglobin via 150-ps time-resolved Laue crystallography. *Chem. Phys.,* 422**,** 98-106.

SCHOTTE, F., LIM, M., JACKSON, T. A., SMIRNOV, A. V., SOMAN, J., OLSON, J. S., PHILLIPS , G. N., WULFF, M. & ANFINRUD, P. A. 2003. Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. *Science,* 300**,** 1944-1947.

SHAIKH, T. R., YASSIN, A. S., LU, Z., BARNARD, D., MENG, X., LU, T.-M., WAGENKNECHT, T. & AGRAWAL, R. K. 2014. Initial bridges between two ribosomal subunits are formed within 9.4 milliseconds, as studied by time-resolved cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.,* 111**,** 9822-9827.

SHI, Y. & KOH, J. T. 2004. Light-activated transcription and repression by using photocaged SERMs. *ChemBioChem,* 5**,** 788-796.

SHMUELI, U. 2010. Reciprocal space. *Int. Tables Cryst.,* B.

SIMMONS, K. J., CHOPRA, I. & FISHWICK, C. W. G. 2010. Structure-based discovery of antibacterial drugs. *Nat. Rev. Micro.,* 8**,** 501-510.

SINGER, P., SMALAS, A., CARTY, R., MANGEL, W. & SWEET, R. 1993. The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science,* 259**,** 669-673.

SKUCAS, E., NGAI, M.-Y., KOMANDURI, V. & KRISCHE, M. J. 2007. Enantiomerically enriched allylic alcohols and allylic amines via C–C bond-forming hydrogenation: asymmetric carbonyl and imine vinylation. *Acc. Chem. Res.,* 40**,** 1394-1401.

SNELL, E. E. 1986. Pyruvoyl-dependent histidine decarboxylase from Lactobacillus 30a: purification and properties. *Methods Enzymol.,* 122**,** 128-35.

SOUTHWORTH-DAVIES, R. J., MEDINA, M. A., CARMICHAEL, I. & GARMAN, E. F. 2007. Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure,* 15**,** 1531-1541.

SPECHT, A. & GOELDNER, M. 2004. 1-(o-Nitrophenyl)-2,2,2-trifluoroethyl ether derivatives as stable and efficient photoremovable alcohol-protecting groups. *Angew. Chem. Int. Ed. ,* 43**,** 2008-2012.

SPRENGER, W. W., HOFF, W. D., ARMITAGE, J. P. & HELLINGWERF, K. J. 1993. The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J. Bacteriol.,* 175**,** 3096-104.

SREENATH, H. K., BINGMAN, C. A., BUCHAN, B. W., SEDER, K. D., BURNS, B. T., GEETHA, H. V., JEON, W. B., VOJTIK, F. C., ACETI, D. J., FREDERICK, R. O., PHILLIPS JR, G. N. & FOX, B. G. 2005. Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Prot. Expres. Purif.,* 40**,** 256-267.

ŠTAMBASKÝ, J., MALKOV, A. V. & KOČOVSKÝ, P. 2008. Synthesis of enantiopure 1-arylprop-2-en-1-ols and their tert-butyl carbonates. *J. Org. Chem.,* 73**,** 9148-9150.

STODDARD, B. L., COHEN, B. E., BRUBAKER, M., MESECAR, A. D. & KOSHLAND, D. E. 1998. Millisecond Laue structures of an enzyme-product complex using photocaged substrate analogs. *Nat. Struct. Mol. Biol.,* 5**,** 891-897.

STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Prot. Exp. Purif.,* 41**,** 207-234.

STUECKER, T. N., HODGE, K. M. & ESCALANTE-SEMERENA, J. C. 2012a. The missing link in coenzyme A biosynthesis: PanM (formerly YhhK), a yeast GCN5 acetyltransferase homologue triggers aspartate decarboxylase (PanD) maturation in Salmonella enterica. *Mol. Microbiol.,* 84**,** 608-619.

STUECKER, T. N., TUCKER, A. C. & ESCALANTE-SEMERENA, J. C. 2012b. PanM, an acetyl-coenzyme A sensor required for maturation of L-aspartate decarboxylase (PanD). *MBio,* 3.

TAKAOKA, K., TATSU, Y., YUMOTO, N., NAKAJIMA, T. & SHIMAMOTO, K. 2004. Synthesis of carbamate-type caged derivatives of a novel glutamate transporter blocker. *Bioorg. Med. Chem.,* 12**,** 3687-3694.

TANG, L., FAN, T. M., BORST, L. B. & CHENG, J. 2012. Synthesis and biological response of size-specific, monodisperse drug–silica nanoconjugates. *ACS Nano,* 6**,** 3954-3966.

TANG, X. & BRUCE, J. E. 2009. Chemical cross-linking for protein-protein interaction studies. *Methods Mol. Biol.,* 492**,** 283-93.

TAO, Z., ZHANG, Y. W., AGYIRI, A. & RUDNICK, G. 2009. Ligand effects on cross-linking support a conformational mechanism for serotonin transport. *J. Biol. Chem.,* 284**,** 33807-14.

TENBOER, J., BASU, S., ZATSEPIN, N., PANDE, K., MILATHIANAKI, D., FRANK, M., HUNTER, M., BOUTET, S., WILLIAMS, G. J., KOGLIN, J. E., OBERTHUER, D., HEYMANN, M., KUPITZ, C., CONRAD, C., COE, J., ROY-CHOWDHURY, S., WEIERSTALL, U., JAMES, D., WANG, D., GRANT, T., BARTY, A., YEFANOV, O., SCALES, J., GATI, C., SEURING, C., SRAJER, V., HENNING, R., SCHWANDER, P., FROMME, R., OURMAZD, A., MOFFAT, K., VAN THOR, J. J., SPENCE, J. C. H., FROMME, P., CHAPMAN, H. N. & SCHMIDT, M. 2014. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science,* 346**,** 1242-1246.

TODD, M. J. & GOMEZ, J. 2001. Enzyme Kknetics determined using calorimetry: a general assay for enzyme activity? *Anal. Biochem.,* 296**,** 179-187.

TONG, Y., BRANDT, G. S., LI, M., SHAPOVALOV, G., SLIMKO, E., KARSCHIN, A., DOUGHERTY, D. A. & LESTER, H. A. 2001. Tyrosine decaging leads to substantial membrane trafficking during modulation of an inward rectifier potassium channel. *J. Gen. Physiol.,* 117**,** 103-118.

TRAKSELIS, M. A., ALLEY, S. C. & ISHMAEL, F. T. 2005. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug. Chem.,* 16**,** 741-50.

TREBBIN, M., KRUGER, K., DEPONTE, D., ROTH, S. V., CHAPMAN, H. N. & FORSTER, S. 2014. Microfluidic liquid jet system with compatibility for atmospheric and high-vacuum conditions. *Lab on a Chip,* 14**,** 1733-1745.

TRIP, H., MULDER, N. L., RATTRAY, F. P. & LOLKEMA, J. S. 2011. HdcB, a novel enzyme catalysing maturation of pyruvoyl-dependent histidine decarboxylase. *Mol. Microbiol.,* 79**,** 861-871.

TZENG, S. R., PAI, M. T. & KALODIMOS, C. G. 2012. NMR studies of large protein systems. *Methods Mol. Biol.,* 831**,** 133-40.

VAGIN, A. & TEPLYAKOV, A. 1997. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.,* 30**,** 1022-1025.

VALDÉS, R., SHINDE, U. & LANDFEAR, S. M. 2012. Cysteine cross-linking defines the extracellular gate for the Leishmania donovani nucleoside transporter 1.1 (LdNT1.1). *J. Biolog. Chem.,* 287**,** 44036-44045.

VAN POELJE, P. D. & SNELL, E. E. 1990. Pyruvoyl-dependent enzymes. *Annu. Rev. Biochem.,* 59**,** 29-59.

VENKITAKRISHNAN, R. P., ZABOROWSKI, E., MCELHENY, D., BENKOVIC, S. J., DYSON, H. J. & WRIGHT, P. E. 2004. Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry,* 43**,** 16046-16055.

WAIGHT, A. B., PEDERSEN, B. P., SCHLESSINGER, A., BONOMI, M., CHAU, B. H., ROE-ZURZ, Z., RISENMAY, A. J., SALI, A. & STROUD, R. M. 2013. Structural basis for alternating access of a eukaryotic calcium/proton exchanger. *Nature,* 499**,** 107-110.

WALKER, J. W., REID, G. P., MCCRAY, J. A. & TRENTHAM, D. R. 1988. Photolabile 1-(2-nitrophenyl)ethyl phosphate esters of adenine nucleotide analogs. Synthesis and mechanism of photolysis. *J. Amer. Chem. Soc.,* 110**,** 7170-7177.

WAND, A. J. 2001. Dynamic activation of protein function: A view emerging from NMR spectroscopy. *Nat. Struct. Mol. Biol.,* 8**,** 926-931.

WANG, L., CORRIE, J. E. T. & WOOTTON, J. F. 2002. Photolabile precursors of cyclic nucleotides with high aqueous solubility and stability. *J. Org. Chem.,* 67**,** 3474-3478.

WANG, T., ZHAO, Y., SHI, M. & WU, F. 2007. The synthesis of novel coumarin dyes and the study of their photoreaction properties. *Dyes and Pigments,* 75**,** 104-110.

WARD, J. F. 1988. DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *In:* WALDO, E. C. & KIVIE, M. (eds.) *Prog. Nucleic Acid Res. Mol. Biol.*: Academic Press.

WEBB, M. E., LOBLEY, C. M. C., SOLIMAN, F., KILKENNY, M. L., SMITH, A. G., BLUNDELL, T. L. & ABELL, C. 2012. Structure of Escherichia coli aspartate [alpha]-decarboxylase Asn72Ala: probing the role of Asn72 in pyruvoyl cofactor formation. *Acta Cryst.,* F68**,** 414-417.

WEBB, M. E., SMITH, A. G. & ABELL, C. 2004. Biosynthesis of pantothenate. *Nat. Prod. Rep.,* 21**,** 695-721.

WEBB, M. E., YORKE, B. A., KERSHAW, T., LOVELOCK, S., LOBLEY, C. M. C., KILKENNY, M. L., SMITH, A. G., BLUNDELL, T. L., PEARSON, A. R. & ABELL, C. 2014. Threonine 57 is required for the post-translational activation of E. coli aspartate a-decarboxylase. *Acta Cryst.,* D70**,** 1166-1172.

WEIK, M., BERGES, J., RAVES, M. L., GROS, P., MCSWEENEY, S., SILMAN, I., SUSSMAN, J. L., HOUEE-LEVIN, C. & RAVELLI, R. B. G. 2002. Evidence for the formation of disulfide radicals in protein crystals upon X-ray irradiationPresented at the `Second International Workshop on Radiation Damage to Crystalline Biological Samples' held at Advanced Photon Source, Chicago, USA, in December 2001. *J. Synch. Rad.,* 9**,** 342-346.

WEIK, M., RAVELLI, R. B. G., KRYGER, G., MCSWEENEY, S., RAVES, M. L., HAREL, M., GROS, P., SILMAN, I., KROON, J. & SUSSMAN, J. L. 2000. Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl Acad. Sci. U.S.A.,* 97**,** 623-628.

WEST, J. M., XIA, J., TSURUTA, H., GUO, W., O'DAY, E. M. & KANTROWITZ, E. R. 2008. Time evolution of the quaternary structure of Escherichia coli aspartate transcarbamoylase upon reaction with the natural substrates and a slow, tight-binding inhibitor. *J. Mol. Biol.,* 384**,** 206-18.

WIEBOLDT, R., GEE, K. R., NIU, L., RAMESH, D., CARPENTER, B. K. & HESS, G. P. 1994. Photolabile precursors of glutamate: synthesis, photochemical properties, and activation of glutamate receptors on a microsecond time scale. *Proc. Natl. Acad. Sci. U.S.A.,* 91**,** 8752-8756.

WILCOX, M., VIOLA, R. W., JOHNSON, K. W., BILLINGTON, A. P., CARPENTER, B. K., MCCRAY, J. A., GUZIKOWSKI, A. P. & HESS, G. P. 1990. Synthesis of photolabile precursors of amino acid neurotransmitters. *J. Org. Chem.,* 55**,** 1585-1589.

WILLIAMSON, J. M. & BROWN, G. M. 1979. Purification and properties of L-Aspartate-alpha-decarboxylase, an enzyme that catalyzes the formation of beta-alanine in Escherichia coli. *J. Biol. Chem,* 254**,** 8074-8082.

WINTER, G. 2010. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.,* 43**,** 186-190.

WÖHRI, A. B., KATONA, G., JOHANSSON, L. C., FRITZ, E., MALMERBERG, E., ANDERSSON, M., VINCENT, J., EKLUND, M., CAMMARATA, M., WULFF, M., DAVIDSSON, J., GROENHOF, G. & NEUTZE, R. 2010. Light-induced structural changes in a photosynthetic reaction center caught by laue diffraction. *Science,* 328**,** 630-633.

WOOTTON, J. W. & TRENTHAM, D. R. 1989. "Caged" compounds to probe the dynamics of cellular processes: synthesis and properties of some novel photosensitive p-2-nitrobenzyl esters of nucleotides. *NATO ASI Series, Series C,* 272**,** 277-296.

YAMAGUCHI, S., KAMIKUBO, H., KURIHARA, K., KUROKI, R., NIIMURA, N., SHIMIZU, N., YAMAZAKI, Y. & KATAOKA, M. 2009. Low-barrier hydrogen bond in photoactive yellow protein. *Proc. Natl. Acad. Sci. U.S.A.,* 106**,** 440-444.

YEATES, T. O. 1997. Detecting and overcoming crystal twinning. *Methods Enzymol.,* 276**,** 344-358.

YOUNG, T. S. & SCHULTZ, P. G. 2010. Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J. Biolog. Chem.,* 285**,** 11039-11044.

YU, F., CANGELOSI, V. M., ZASTROW, M. L., TEGONI, M., PLEGARIA, J. S., TEBO, A. G., MOCNY, C. S., RUCKTHONG, L., QAYYUM, H. & PECORARO, V. L. 2014. Protein design: toward functional metalloenzymes. *Chem. Rev.,* 114**,** 3495-3578.

YU, H., LI, J., WU, D., QIU, Z. & ZHANG, Y. 2010. Chemistry and biological applications of photo-labile organic molecules. *Chem. Soc. Rev.,* 39**,** 464-473.

YUN, M., PARK, C.-G., KIM, J.-Y., ROCK, C. O., JACKOWSKI, S. & PARK, H.-W. 2000. Structural basis for the feedback regulation of Escherichia coli pantothenate kinase by coenzyme A. *J. Biol. Chem.,* 275**,** 28093-28099.

ZELDIN, O. B., GERSTEL, M. & GARMAN, E. F. 2013. RADDOSE-3D: time- and space-resolved modelling of dose in macromolecular crystallography. *J. Appl. Cryst.,* 46**,** 1225-1230.

ZHOU, M. & ROBINSON, C. V. 2014. Flexible membrane proteins: functional dynamics captured by mass spectrometry. *Curr. Op. Struct. Biol.,* 28**,** 122-130.

ZHOU, X., LEVIN, E. J., PAN, Y., MCCOY, J. G., SHARMA, R., KLOSS, B., BRUNI, R., QUICK, M. & ZHOU, M. 2014. Structural basis of the alternating-access mechanism in a bile acid transporter. *Nature,* 505**,** 569-573.

ZOU, K., CHELEY, S., GIVENS, R. S. & BAYLEY, H. 2002. Catalytic subunit of protein kinase A caged at the activating phosphothreonine. *J. Amer. Chem. Soc.,* 124**,** 8220-8229.