# Biophysical studies on enzymes related to antibiotics




A thesis submitted to the Board of the Faculty of Mathematical, Physical & Life Sciences in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Jos J. A. G. Kamps

St. Peter's College

Hillary term 2020








# Biophysical studies on enzymes related to antibiotics

Jos Johannes Antonius Geertruida Kamps DPhil thesis

St. Peter's College, University of Oxford Hillary term 2020

## Abstract

The work described in this thesis concerns two research areas: (i) time resolved (tr) crystallography studies on two enzymes, oxacillinase 10 (OXA-10) and isopenicillin N synthase (IPNS); and (ii) the reactivity of formaldehyde towards amino acids. OXA-10 is a class D serine β-lactamase that confers bacterial resistance against clinically relevant β-lactam antibiotics. IPNS is a non-heme Fe dependent enzyme that catalyses the formation of isopenicillin N (IPN) from L-δ-(α-aminoadipoyl)-L-cysteinyl D-valine (ACV), which is an important step in the biosynthesis of penicillins and cephalosporins. Formaldehyde is an environmental pollutant and a metabolite in many biochemical reactions. Formaldehyde is toxic and considered carcinogenic above threshold levels, however its reactivity with biomolecules is not well defined.

Chapter 1 provides an overview of the literature regarding the biosynthesis of pencillins and cephalosporins and emphasizes the β-lactam ring formation step catalysed by IPNS, discussing the proposed catalytic mechanism in more detail. The role of β-lactamases, and in particular OXA-10, in antibiotic resistance is also discussed.

Chapter 2 provides an introduction to tr-crystallography, with an emphasis on the use of X-ray free electron lasers (XFELs). A brief description of the physics behind XFELs is given, as well as the application of XFELs in macromolecular crystallography. Preparation, characterisation, and sample delivery of microcrystal slurries for serial crystallography is introduced.

Chapter 3 describes methods for the preparations of OXA-10, and discusses the results of (tr-)crystallographic studies on OXA-10. Different crystallisation conditions for OXA-10 were explored and probed for diffraction using either synchrotron or XFEL radiation. A combination of fixed target and drop-on-demand sample delivery systems were used for the serial crystallography studies on OXA-10.

Chapter 4 describes the preparation of ACV, IPNS, and a $^{19}$F labelled IPNS S55C variant. The chapter covers the use of (tr-)crystallographic studies on IPNS. IPNS was co-crystallised to give IPNS • Fe • ACV microcrystals, which were exposed to $O_2$ in order to initiate the reaction, and probed for diffraction at defined time points using XFEL radiation. Diffraction data and X-ray emission spectroscopy data were recorded in parallel. The results demonstrated a conformational change in an α-helix and β-strand region and suggested that Fe leaves the active site of IPNS after IPN formation *in crystallo*. To monitor the conformational changes in an α-helix and β-strand region of IPNS a $^{19}$F labelled IPNS S55C variant was produced and used for $^{19}$F NMR experiments. The results provided evidence for similar conformational changes of the α-helix and β-strand region in solution, consistent with the time resolved crystallographic studies. Non-denaturing mass spectrometry and $^{19}$F NMR studies were used to investigate the loss of Fe at the IPNS active site after IPN formation, as observed *in crystallo*. The results suggest that IPN induced loss of Fe occurs *in crystallo* but not in solution.

Chapter 5 summarises the work conducted on OXA-10 and IPNS, and discusses lessons drawn from the experimental work described.

Chapter 6 provides an overview of the literature concerning the exposure of humans to exogenous and endogenous sources of formaldehyde and summarizes the effects of formaldehyde on health. Previous biochemical and chemical studies on the reactivity of formaldehyde with biomolecules (e. g. DNA, proteins, peptides, nucleobases, amino acids) are discussed.

Chapter 7 describes studies on the reactions of formaldehyde with amino acids as monitored using NMR spectroscopy. The products of the reactions were characterised, and their relative stabilities investigated. Competition assays using different amino acids were carried out to establish their relative reactivities. Different biologically relevant carbonyl containing molecules were tested and their reactivities compared with that of formaldehyde. The results demonstrate that formaldehyde was more reactive and forms more stable adducts with amino acids than the other tested compounds.

**Dedication**

Opgedragen aan mijn ouders

# Acknowledgements

I would like to take the opportunity to sincerely thank my supervisors Prof Tim Claridge and Prof Chris Schofield for their guidance throughout my DPhil. I am especially grateful for the scientific freedom I have been given by both, exploring various projects in different fields of science. I would also like to thank Tim for his guidance and productive scientific discussions, and Chris for improving my scientific and writing skills during my stay in Oxford.

I would like to express my gratitude towards my friend and colleague Dr Patrick Rabe, who has had a major impact on this work. I greatly enjoyed working together with him and have gained a great deal of respect towards him.

For proofreading my thesis I would like to thank Dr Patrick Rabe, Dr Jürgen Brem, Dr Richard Hopkinson, Dr Jan Kern, Dr Ian Clifton, Dr Pierre Aller, Aistė Skorupskaitė, Pauline Lang, Raphael Reinbold, Tobias John, James Linyard, Felix Urbitsch, and Marie Wong.

Much of the work in this thesis is product of collaborative work, and thus I would like to thank all the people I have had the privilege to have collaborated with. In particular I would like to thank Dr Allen Orville and Dr Jan Kern, who played an important role in all of the XFEL experiments and coordinating these large collaborative projects. During these experiments the data processing and operating of the various pieces of equipment was done by members of the XFEL hub, Dr Agata Butryn and Dr Pierre Aller, and members of the NBNL including, Dr Vittal Yachandra, Dr Junko Yano, Dr Uwe Bergman, Dr Cindy Pham, Dr Aaron Brewster, Dr Kyle Sutherlin, Dr Asmit Bhowmick, Dr and Iris Young, Dr In-sik Kim, Dr Joshua Heinemann, Dr Philipp Simon, and Isabel Bogacz. I would also like to thank the other members of XFEL proposals 2017, LS00, LS34, PCS110, PCS115, LU50, A8017, and A8088 for making these experiments so enjoyable. I would like to thank the XFEL hub, Dr Allen Orville, Dr Agata Butryn, and Dr Pierre Aller and the beamline scientists at Diamond Light Source for their support during the fixed target experiments.

I would also like to thank Denis Shutin for collaborating on the non-denaturing MS experiments, and Elisabete Pires for advising with the protein digestion experiments.

I am very thankful for the people who taught me the ropes in NMR spectroscopy, including Prof Tim Claridge, Dr Barbara Odell, Dr Nick Rees, and Dr Nadar Amin. I am also thankful to Dr Thomas Leissing, Dr Jürgen Brem, Dr Ian Clifton, and Dr Michael McDonough for introducing me to macromolecular crystallography and teaching me tips and tricks. I am also thankful for working together with Dr Richard Hopkinson, Dr Wei Wu, Dr Jia Chen, and Roman Belle. I am grateful for the members of NMR and MS, and the TDWC and CJS group, past and present, for the good times in- and outside of the lab.

I would also like to thank the SBM CDT and their partners for funding, and my fellow SBM CDT members for their friendship and support.

Finally, I would like to thank my parents and sister for their continued love and support during my "time abroad".

## Author contributions

All experiments described in this thesis were carried out by myself unless otherwise stated. Contributions from collaborators are mentioned in the text with their names in parentheses.

All images obtained from external sources are used with prior permission, and appropriately cited.

# Table of contents

Abbreviations ..... xiii

Chapter 1 ..... 1

1.1 A brief history of antibiotics ..... 2

1.1.1 Early work on antibiotics ..... 2

1.1.2 Discovery of the β-lactam antibiotics ..... 2

1.2 Biosynthesis and fermentation based production of β-lactams ..... 6

1.2.1 ACV synthase ..... 6

1.2.2 Isopenicillin N amidohydrolase and acyltransferase ..... 8

1.2.3 Isopenicillin N epimerase ..... 8

1.2.4 Cephalosporins ..... 9

1.3 Isopenicillin-N synthase ..... 9

1.3.1 Discovery of ACV and IPNS ..... 9

1.3.2 ACV analogues and the mechanism of IPNS ..... 10

1.3.3 Spectroscopic studies on IPNS ..... 13

1.3.4 Crystallography of IPNS ..... 15

1.3.5 Proposed mechanism of IPNS ..... 16

1.4 Antibiotic resistance and β-lactamases ..... 17

1.4.1 Resistance through alteration of penicillin-binding proteins ..... 18

1.4.2 Resistance through cell wall permeability ..... 18

1.4.3 Resistance through enzymatic inactivation ..... 18

1.4.4 Class D serine β-lactamases and OXA-10 ..... 19

1.4.5 Hydrolysis mechanism of serine β-lactamases ..... 20

1.4.6 Carbapenem hydrolysis by oxacillinases ..... 20

1.5 Aims ..... 21

1.6 References ..... 22

Chapter 2 ..... 28

2.1 Developments in synchrotron based crystallography ..... 29

2.2 Time-resolved crystallography ..... 30

2.2.1 Observing transitions in macromolecular crystallography ..... 30

2.2.2 Reaction initiation in crystals ..... 31

2.2.3 Laue method ..... 32

2.2.4 Cryo-crystallography and modified systems ..... 32

2.2.5 Serial synchrotron crystallography ..... 33

2.3 X-ray Free Electron Laser radiation ..... 33

2.3.1 Free Electron Laser physics ..... 33
2.3.2 Radiation damage ..... 36
2.3.3 Application of coherent properties of FEL radiation ..... 38
2.4 The use of XFELs in protein crystallography ..... 39
2.5 Sample and sample delivery for XFEL studies ..... 42
2.5.1 Sample characteristics ..... 42
2.5.2 Sample delivery ..... 43
2.5.2 Drop-on-demand sample delivery system ..... 45
2.6 X-ray emission spectroscopy ..... 46
2.7 Experimental setup ..... 48
2.8 References ..... 49
Chapter 3 ..... 59
3.1 Introduction ..... 60
3.1.1 Hydrolysis of β-lactams by OXA-10 ..... 60
3.1.2 Time resolved SFX using OXA-10 ..... 62
3.2 Preparation of recombinant OXA-10 ..... 64
3.3 Time resolved X-ray crystallography ..... 65
3.3.1 Crystallisation of OXA-10 ..... 65
3.3.2 XFEL experiment on OXA-10 at LCLS ..... 71
3.3.3 Data acquisition and processing ..... 72
3.3.4 Search for different crystallisation conditions ..... 75
3.3.5 Fixed target experiment at DLS ..... 77
3.3.6 A room temperature fixed target structure of OXA-10 ..... 79
3.3.7 NMR studies on viscosity and stability of β-lactam compounds ..... 80
3.3.8 XFEL experiment on OXA-10 at SACLA ..... 83
3.4 Discussion ..... 85
3.5 Future work ..... 88
3.6 Conclusions ..... 89
3.7 Protein production and purification ..... 90
3.7.1 General methods ..... 90
3.7.2 Expression and purification of OXA-10 ..... 91
3.8 Crystallography ..... 93
3.8.1 Crystallisation of OXA-10 ..... 93
3.8.2 Seed stock preparation ..... 94
3.8.3 Data collection ..... 94
3.9 Diffusion experiments by NMR ..... 95

3.10 References ..... 96
Chapter 4 ..... 98
4.1 Introduction ..... 99
4.1.1 Crystallisation of isopenicillin N synthase ..... 99
4.1.2 Time resolved SFX using isopenicillin N synthase ..... 100
4.2 Preparation of isopenicillin N and substrate ..... 101
4.2.1 Substrates from a commercial source ..... 101
4.2.2 Solution phase synthesis of ACV ..... 102
4.2.3 Solid phase peptide synthesis of ACV ..... 104
4.2.4 Preparation of deuterated ACV analogues ..... 105
4.2.5 Preparation of IPN ..... 109
4.2.6 Production of recombinant IPNS ..... 109
4.3 Time resolved X-ray crystallography ..... 111
4.3.1 IPNS sample preparation at LCLS trip 1 ..... 111
4.3.2 Data acquisition ..... 112
4.3.3 X-ray emission spectroscopy data collection ..... 115
4.3.4 IPNS anaerobic room temperature structure ..... 116
4.3.5 IPNS sample preparation at LCLS trip 2 ..... 117
4.3.6 Structure of IPNS incubated with $O_2$ ..... 119
4.3.7 Structure of IPNS in complex with IPN ..... 124
4.3.8 Structure of IPNS exposed to nitric oxide ..... 128
4.4 Conformational changes in isopenicillin N synthase ..... 130
4.4.1 Structural rearrangements in IPNS ..... 130
4.4.2 Protein observed $^{19}F$ NMR spectroscopy ..... 130
4.5 Mutation and fluorine labelling of IPNS ..... 133
4.5.1 Expression and fluorine labelling of IPNS S55C variant ..... 133
4.5.2 Investigating labelling site and efficiency ..... 136
4.5.3 Circular dichroism analyses ..... 138
4.5.4 Structural evidence of IPNS S55C BTFA ..... 139
4.6 Helix rearrangement experiments by NMR ..... 141
4.6.1 Titration experiments of IPNS variant by $^{19}F$ NMR ..... 141
4.6.2 Nitric oxide exposure of IPNS variant complexes ..... 143
4.6.3 Binding of vanadium(IV) oxide to IPNS* ..... 146
4.7 Binding of isopenicillin N to IPNS ..... 148
4.7.1 Introduction ..... 148
4.7.2 Non-denaturing mass spectrometry ..... 148

4.7.3 Binding of IPN to IPNS variant by $^{19}F$ NMR spectroscopy ..... 151
4.8 Discussion ..... 153
4.8.1 IPNS time resolved crystallography ..... 153
4.8.2 Solution state studies on IPNS ..... 155
4.9 Future work ..... 158
4.10 Conclusions ..... 161
4.11 Materials and methods ..... 163
4.11.1 General procedures ..... 163
4.11.2 Preparation of ACV using solution phase synthesis ..... 164
4.11.3 Preparation of deuterated valine ..... 170
4.11.4 Preparation of ACV using solid phase peptide synthesis ..... 178
4.11.5 Purification by preparative-HPLC ..... 179
4.12 Protein production and purification ..... 180
4.12.1 General methods ..... 180
4.12.2 Expression and purification of IPNS ..... 181
4.12.3 Expression and purification of IPNS (hexahistidine affinity tag) ..... 182
4.12.4 Fluorine labelling of IPNS S55C ..... 183
4.12.5 Dialysis of IPNS ..... 183
4.12.6 Trypsin digestion ..... 184
4.12.7 Circular dichroism experiments ..... 184
4.13 Non-denaturing mass spectrometry ..... 184
4.14 Nuclear Magnetic Resonance experiments ..... 185
4.14.1 Sample preparation for NMR spectroscopy ..... 185
4.15 Crystallography ..... 186
4.15.1 General crystallisation method ..... 186
4.15.2 Batch crystallisation of IPNS microcrystals ..... 186
4.15.3 Seed stock preparation ..... 187
4.15.4 Data collection ..... 187
4.16 References ..... 188
Chapter 5 ..... 191
5.1 Crystallisation of OXA-10 ..... 192
5.2 Time resolved crystallography on IPNS ..... 193
5.3 Future perspective ..... 194
5.4 References ..... 195
Chapter 6 ..... 197
6.1 Formaldehyde as a chemical ..... 198

6.2 Exogenous formaldehyde sources ..... 198
6.2.1 Formaldehyde from consumption and inhalation ..... 198
6.2.2 Formaldehyde in pharmaceuticals ..... 199
6.2.3 Occupational and consumer exposure to formaldehyde ..... 200
6.3 Endogenous formaldehyde sources ..... 201
6.3.1 Formaldehyde inside humans ..... 201
6.3.2 Biochemical processes that release formaldehyde ..... 202
6.3.3 Metabolism of formaldehyde ..... 204
6.4 Effects of formaldehyde on health ..... 206
6.4.1 Toxic effects of formaldehyde ..... 206
6.4.2 Formaldehyde and carcinogenicity ..... 207
6.4.3 Health effects related to endogenous formaldehyde ..... 208
6.5 Formaldehyde chemistry ..... 209
6.5.1 Reactions of formaldehyde with biological molecules ..... 210
6.6 Aims ..... 213
6.7 References ..... 213
Chapter 7 ..... 221
7.1 Introduction ..... 222
7.1.1 Background ..... 222
7.1.2 Approach ..... 222
7.2 HCHO in aqueous solution ..... 223
7.3 Amino acids and formaldehyde – product characterisation ..... 225
7.3.1 Experimental conditions ..... 225
7.3.2 Cysteine and HCHO ..... 225
7.3.3 Homocysteine and HCHO ..... 227
7.3.4 Penicillamine and HCHO ..... 228
7.3.5 Serine and formaldehyde ..... 229
7.3.6 Threonine and HCHO ..... 231
7.3.7 *allo*-Threonine and HCHO ..... 232
7.3.8 Homoserine and HCHO ..... 233
7.3.9 Asparagine and HCHO ..... 234
7.3.10 Glutamine and HCHO ..... 235
7.3.11 Histidine and formaldehyde ..... 236
7.3.12 Tryptophan and HCHO ..... 239
7.3.13 Ornithine and HCHO ..... 240
7.3.14 Lysine and HCHO ..... 245

7.3.15 Arginine and HCHO ..... 246
7.3.16 5-Hydroxy lysine and HCHO ..... 247
7.3.17 Other amino acids and HCHO ..... 249
7.4 pH and concentration dependence studies ..... 250
7.4.1 Experimental conditions ..... 250
7.4.2 Cysteine and homocysteine pH-controlled studies ..... 250
7.4.3 Serine, threonine, *allo*-threonine, and homoserine pH-controlled studies ..... 251
7.4.4 Asparagine and glutamine pH-controlled studies ..... 253
7.4.5 Histidine and tryptophan pH-controlled studies ..... 253
7.4.6 Ornithine, lysine and arginine under pH-controlled conditions ..... 254
7.4.7 Oxazolidine formation pH effects ..... 254
7.4.8 Cellular concentrations ..... 257
7.5 Assessing product stability ..... 259
7.6 Amino acid competition experiments ..... 263
7.6.1 Competition between cysteine and different amino acids for reaction with HCHO ..... 263
7.6.2 Competition experiments of thiol containing amino acids ..... 264
7.6.3 Competition experiments of alcohol containing amino acids ..... 265
7.6.4 Competition experiments with a range of amino acids ..... 265
7.6.5 A spectator amino acid and the rate of reaction of HCHO ..... 267
7.7 Comparison of the reactivity of Cys with HCHO using different electrophiles ..... 269
7.8 Cysteine containing peptides ..... 274
7.9 Reductive $N^{\alpha}$-Methylation ..... 275
7.10 Discussion ..... 277
7.11 Conclusions ..... 283
7.12 Methods ..... 284
7.12.1 General methods ..... 284
7.12.2 Experimental methods ..... 284
7.12.3 Product characterisation ..... 287
7.13 References ..... 302

# Abbreviations

| | |
|---|---|
| 1D | 1-dimensional |
| 2D | 2-dimensional |
| 2OG | 2-oxoglutarate |
| 2TY | 2 × tryptone/yeast extract |
| 5hLys | 5-hydroxy lysine |
| 5,10-$CH_2$-THF | 5,10-methylene-tetrahydrofolate |
| Ac | acyl |
| αACV | L-α-(aminoadipoyl)-L-cysteinyl D-valine |
| ACA | L-α-(aminoadipoyl)-L-cysteinyl D-alanine |
| ACdV | L-δ-(α-aminoadipoyl)-L-cysteinyl D-3-[$^2$H]-valine |
| ACG | L-α-(aminoadipoyl)-L-cysteinyl glycine |
| ACV (bold) | L-δ-(α-aminoadipoyl)-L-cysteinyl D-valine |
| ACV (not-bold) | L-(α-aminoadipoyl)-L-cysteinyl D-valine |
| AC(PHB)V | (*S*)-(4-methoxy)benzyl L-δ-(α-aminoadipoyl)-L-cysteinyl D-valine |
| AdCV | L-δ-(α-aminoadipoyl)-L-3,3-[$^2H_2$]-cysteinyl D-valine |
| ADE | acoustic droplet ejector |
| ADH | aldehyde dehydrogenase |
| $BiGF_2$ | (5*R*,10*S*)-5-(carboxymethylcarbamoyl)-7-oxo-3-thia-1,6-diazabicyclo[4.4.1]undecane-10-carboxylic acid |
| BlaC | *M. tuberculosis* β-lactamase |
| Boc | *tert*-butoxycarbonyl |
| $Boc_2O$ | *tert*-butoxycarbonyl anhydride |
| BTFA | 1-bromo-3,3,3-trifluoro acetone |
| Bzh | benzhydryl |
| CD | circular dichroism |
| CDI | coherent diffraction imaging |
| CV | column volume |
| CoA | coenzyme A |
| Da | Dalton (mass unit) |
| DCH | 1,3-diketocyclohexane |
| DHP-I | dehydropeptidase I |
| DIC | *N*,*N*′-diisopropylcarbodiimide |
| DIPEA | diisopropyl ethylamine |
| DLS | diamond light source |
| DMAP | *N*,*N*-dimethyl 4-aminopyridine |
| DMDM | 1,3-dimethylol-5,5-dimethyl (hydantoin) |
| DMF | *N*,*N*-dimethylformamide |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| DOD | drop-on-demand |
| DOSY | diffusion NMR spectroscopy |
| DSBH | double-stranded β-helix |
| DTT | dithiothreitol |
| EDCI | 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide |
| EDTA | ethylenediaminetetraacetic acid |
| EEDQ | *N*-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline |
| EM | electron microscopy |
| EPR | electron paramagnetic resonance |
| eq. | equivalents |
| ESI | electrospray ionisation |
| ESRF | European synchrotron radiation facility |
| Et | ethyl |
| $Et_2O$ | diethylether |
| EtOAc | ethyl acetate |

| | |
|---|---|
| EXAFS | extended X-ray absorption fine structure |
| FA | fanconi anemia |
| FAD | flavin adenine dinucleotide |
| FEL | free electron laser |
| Fmoc | fluorenylmethyloxycarbonyl |
| FPLC | fast protein liquid chromatography |
| fs | femtosecond ($10^{-15}$ s) |
| GPCR | G-protein-coupled receptors |
| GSH | glutathione |
| HATU | 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate |
| Hcy | homocysteine |
| HEPES | 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid |
| HMBC | heteronuclear multi-bond correlation spectroscopy |
| HMG | (*S*)-hydroxymethylglutathione |
| HOBt | 1-hydroxybenzotriazole |
| HPLC | high performance liquid chromatography |
| HRMS | high resolution mass spectrometry |
| HSQC | heteronuclear single-quantum correlation spectroscopy |
| IARC | international Agency for Research on Cancer |
| idx | indexed hit rate |
| IPN | isopenicillin N |
| IPNS | isopenicillin N synthase |
| IPNS* | isopenicillin N synthase S55C variant labelled with a fluorine tag |
| IPTG | isopropyl β-D-thiogalactopyranoside |
| IR | infrared |
| KDM | lysine demethylase |
| KIE | kinetic isotope effect |
| laser | light amplified stimulated emission of radiation |
| LBNL | Lawrence Berkeley national laboratory |
| LCP | lipidic cubic phase |
| LC-MS | liquid chromatography-mass spectrometry |
| MBL | metallo β-lactamase |
| Me | methyl |
| MES | 2-morpholin-4-ylethanesulfonic acid |
| MOPS | 3-Morpholinopropane-1-sulfonic acid |
| MS | mass spectrometry |
| MX | macromolecular crystallography |
| NAD | nicotinamide adenine dinucleotide |
| NMR | nuclear magnetic resonance |
| NO | nitric oxide |
| NOESY | nuclear Overhauser effect spectroscopy |
| OEC | oxygen-evolving complex |
| OXA | oxacillinase |
| PBP | penicillin binding protein |
| PDB | protein data bank |
| PEG | polyethylene glycol |
| Ph | phenyl |
| PMB | *para*-methoxybenzyl |
| ppm | parts per million |
| ps | picosecond |
| psBBOX | *Pseudomonas* γ-butyrobetaine hydroxylase |
| PS II | photosystem II |
| PYP | photoactive yellow protein |
| RMSD | root-mean-square deviation |
| RNA | ribonucleic acid |

| | |
|---|---|
| rt | room temperature |
| SAM | (*S*)-adenosylmethionine |
| SASE | self-amplified spontaneous emission |
| SBL | serine β-lactamase |
| SDS-PAGE | sodium dodecylsulfate polyacrylamide gel electrophoresis |
| SFG | (*S*)-formylglutathione |
| SF-ROX | serial femtosecond rotation crystallography |
| SFX | serial femtosecond X-ray crystallography |
| SHMT | serine hydroxymethyltransferase |
| $SiO_2$ | silicon dioxide |
| SPPS | solid phase peptide synthesis |
| SSAO | semicarbazide-sensitive amine oxidase |
| Su | succinimide |
| TCEP | 3,3',3''-phosphanetriyltripropanoic acid |
| TEMED | tetramethylenediamine |
| TFA | 2,2,2-trifluoro acetic acid |
| TIPS | triisopropyl silane |
| TLC | thin layer chromatography |
| THF (bold) | tetrahydrofolate |
| THF (not-bold) | tetrahydrofuran |
| $^t$Bu | *tert*-butyl |
| tr | time resolved |
| Tris | tris(hydroxymethyl)aminomethane |
| Trt | triphenylmethyl |
| UV | ultraviolet |
| VIS | visible |
| wt | wild-type |
| XANES | X-ray absorption near edge structure |
| XES | X-ray emission spectroscopy |
| XFEL | X-ray free electron laser |

Standard abbreviations for amino acids (single and triple letter), SI units/prefixes, and chemical elements are used throughout. Unusual amino acids have been added to the abbreviations list.

# Chapter 1

## Biosynthesis of β-lactam antibiotics and β-lactamase resistance

# 1.1 A brief history of antibiotics

## 1.1.1 Early work on antibiotics

Infectious diseases have played an import role in the life expectancy of humans.[1] Improvements in sanitation and awareness of possible health threats, followed by the introduction of antibiotics with other medicinal methodologies such as vaccines, have had a profound effect on the mortality rate of humans worldwide.[2] Up to 10 years of additional life expectancy has been attributed to the clinical use of antibiotics to treat infectious diseases.[3,4]

Arsphenamine (more commonly known as Salvarsan, Figure 1.1) was the first chemotherapeutic drug to be used for the treatment of infectious diseases; it was discovered by Paul Ehrlich and co-workers in 1910.[5] Salvarsan was effective for treatment of syphilis until the 1940s; however, it was toxic to patients.[6] Sulfonamidochrysoidine (Prontosil, Figure 1.1), was discovered when investigating the effect of dyes on microorganisms, and showed selective toxicity against bacteria.[7,8] Both Prontosil and Salvarsan are prodrugs that are converted to an active form *in vivo*. Sulphonamide antibiotics, such as Prontosil, inhibit bacterial folate biosynthesis by acting as a competitive for *p*-aminobenzoic acid, so preventing biosynthesis of nucleic acids (DNA / RNA) in bacteria.[9] Humans obtain folate through their diet instead of metabolism, and so Prontosil is selectively toxic to bacteria.

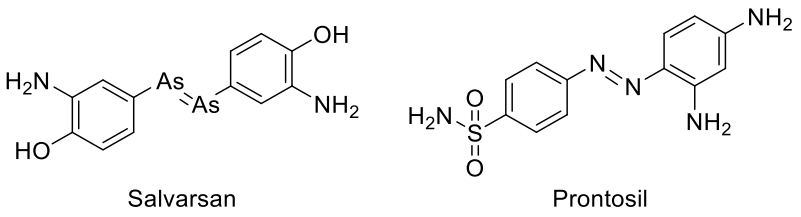



**Figure 1.1**. Structures of Salvarsan (arsphenamine) and Prontosil (sulfonamidochrysoidine) two examples of the early clinically used antibiotics.[10]

## 1.1.2 Discovery of the β-lactam antibiotics

### 1.1.2.1 Penicillins

The discovery of the penicillins, one of the most important drugs of the last century, revolutionised antimicrobial treatment.[11,12] The discovery of the activity of the penicillins is sometimes credited to Alexander Fleming, however, reports prior to his work described the disruption of bacterial growth in

presence of certain strains of fungi.[13] The work by Fleming and co-workers was remarkable because it proposed that the activity was due to a metabolite toxic to bacteria.[10] Fleming observed that the growth of *Stephylococcus aureus* on an old culture plate which was left open, was inhibited by the presence of a fungus.[14] Fleming reported on the fungus, *Penicillium notatum*, and compared its apparent antimicrobial activity to other fungi. The apparently instable nature of penicillin, and the difficulty to extract it from the fungus, prevented its therapeutic development by Fleming.[10] Subsequent efforts led by Howard Florey and Ernst Chain at the University of Oxford led to the partial purification of penicillin and demonstrated its antimicrobial activity in mice infected with *Clostridium septique*, *Streptomonas pyrogenes*, or *Stephylococcus aureus*.[15] Treatment of a human patient with partially purified penicillin led to rapid improvement. Unfortunately, insufficient amounts of penicillin were available to completely cure the patient.[16] The effectiveness of penicillin was also tested on soldiers in North Africa during World War II.[17] These studies were the basis for a wartime collaboration between the United States and the United Kingdom (1941-1942), which resulted in the large scale production of highly purified penicillins. *Penicillium chrysogenum* obtained from mouldy cantaloupe at a grocery market, turned out to be the most efficient fungus for the production of penicillin at the time, and was thus championed as the main strain for the production of penicillin.[10] The fermentation and purification processes of penicillin were optimised. The chemical structure of penicillin was debated in the early stages of its development, but unambiguously demonstrated by crystallographic analysis by Dorothy Hodgkin (Figure 1.2).[18] Work in the late 1940s showed that certain penicillins with hydrophobic side chain residues had improved activity, and could be obtained by the addition of hydrophobic acyl groups to the fermentation broth (e. g. penicillin G, Figure 1.2).[19]

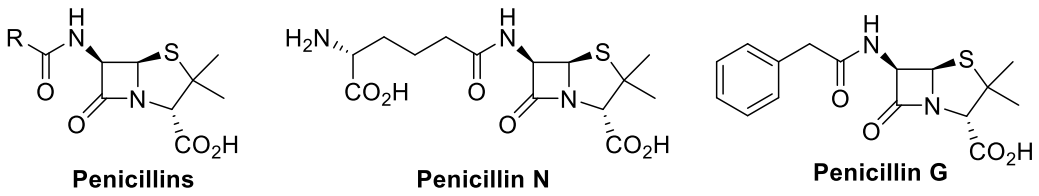



**Figure 1.2**. General structure of the penicillins and two exemplary penicillins. Penicillin N is a natural product produced by the strain discovered by Fleming, and penicillin G is obtained from certain fungal strains (e. g. *P. Chrysogenum*) by the addition of phenylacetic acid.

### 1.1.2.2 Cephalosporins

In search for new antibiotic producing organisms, Brotzu isolated a fungus, *Cephalosporium acremonium*, near the seaside in Italy.[20] *C. acremonium* was observed to inhibit the growth of both gram positive and gram negative bacteria. In the early 1950s work by Newton and Abraham led to the isolation of cephalosporin C (Figure 1.3) produced by *C. acremonium*.[21] The fused β-lactam thiazine ring system of the cephalosporin was found to be more stable in aqueous solutions than the corresponding penicillins, and in general had a broader antibacterial spectrum.[22,23] Similarly to the penicillins, the α-aminoadipoyl (Aad) side chain of cephalosporins could be exchanged for a hydrophobic residue, though this had to be achieved by modification of isolated cephalosporin C. Later on penicillins and cephalosporins were discovered to have a common precursor (Section **1.2**)

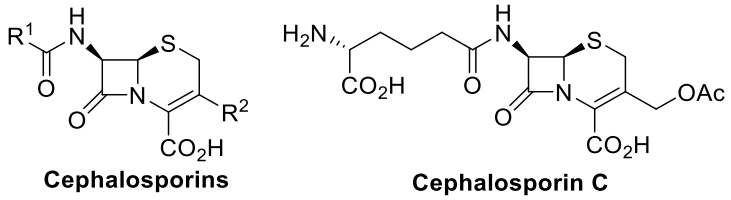



**Figure 1.3**. General structure of cephalosporins and the structure of cephalosporin C.

### 1.1.2.3 Monocyclic β-bactams

The discovery of monocyclic β-lactams started with the discovery of the nocardicins from *Nocardia uniformis* by Aoki *et al* in 1976 (Figure 1.4).[24] The nocardicins, like all monocyclic β-lactams, differ from penicillins and cephalosporins by containing a β-lactam ring that is not fused with a second ring.[25] Only moderate antibacterial activity is observed for most of the naturally occurring monocyclic β-lactams. Over time clinically useful synthetic derivatives such as aztreonam (Figure 1.4) have been developed with improved antibiotic properties.[26]

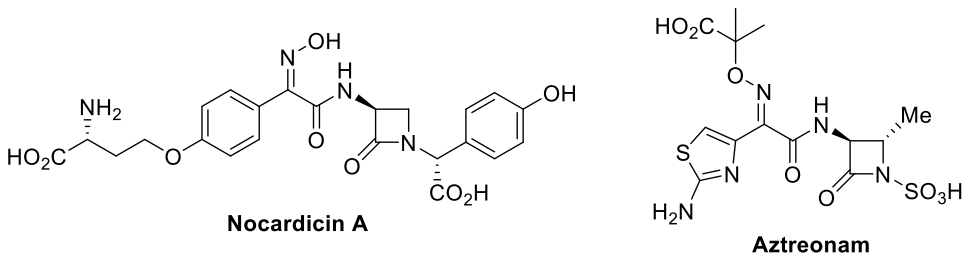



**Figure 1.4**. Structures of Nocardicin A and Aztreonam. Nocardicin A is obtained from the fermentation broth of *N. uniformis*, and Aztreonam is obtained by synthesis.[25]

### 1.1.2.4 Carbapenems

A clinically important class of β-lactams, the carbapenems, is defined as a β-lactam and pyrroline fused ring system.[27] The carbapenem story began with the discovery of olivanic acid from *Streptomyces clavuligerus* (Figure 1.5).[28] This was followed by the isolation of thienamycin from *Streptomyces cattleya* (Figure 1.5).[29] Carbapenems exhibited beneficial antimicrobial properties, such as broad spectrum activity, and greater stability against β-lactamases (at the time of their development) than cephalosporins and penicillins.[27] Development of more soluble, base stable, and potent analogues resulted in the development of ertapenem (Figure 1.5) and other carbapenems. The instalment of a 1-β methyl group onto the pyrroline ring helped to protect against hydrolysis by dehydropeptidase I (DHP-I).[30] Due to their broad spectrum and resistance to β-lactamases, and the recent concerns in global rise of resistance against common antibiotics, carbapenems are considered "drugs of last resort" and are of great clinical relevance.[31,32] A drawback in carbapenem treatments is that few oral formulations are available.

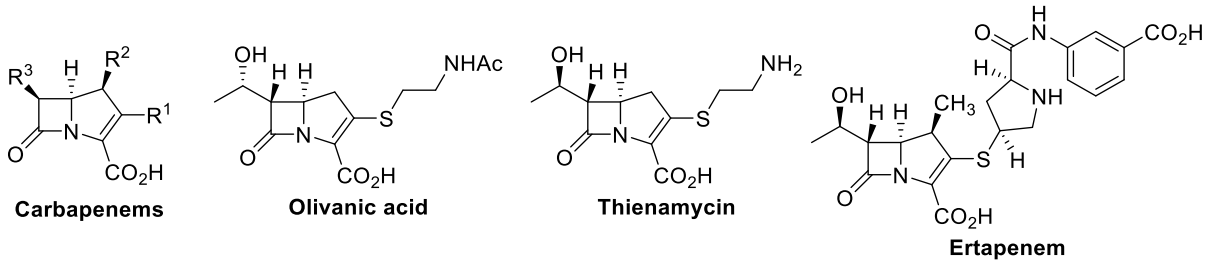



**Figure 1.5**. General structure of the carbapenems and three exemplary carbapenems. Olivanic acid and thienamycin were isolated from *S. clavuligerus*. Ertapenem is a synthetic analogue based on the carbapenem structure.[27]

### 1.1.2.5 Mechanism of action for β-lactam based antibiotics

The targets of β-lactam antibiotics are nucleophilic serine employing peptidoglycan transpeptidase penicillin binding proteins (PBPs).[33] PBPs are involved in the bacterial cell wall biosynthesis and catalyse the polymerisation and cross-linking of peptidoglycan strands. Binding of β-lactam antibiotics to the PBP active site, is at least in part by mimicking the D-Ala-D-Ala motif in peptidoglycans (Figure 1.6). At the active site a Ser residue is irreversibly acylated through a nucleophilic attack of an active site Ser residue to the β-lactam ring, forming an acyl-enzyme complex (Figure 1.6).[33,34] The β-lactam acylated PBPs are unable to bind and/or catalyse crosslinking of peptidoglycan substrates, and so the cell wall biosynthesis is interrupted. Ultimately, the interrupted cell wall biosynthesis leads to lysis of the bacterial cell.[35,36] PBPs are

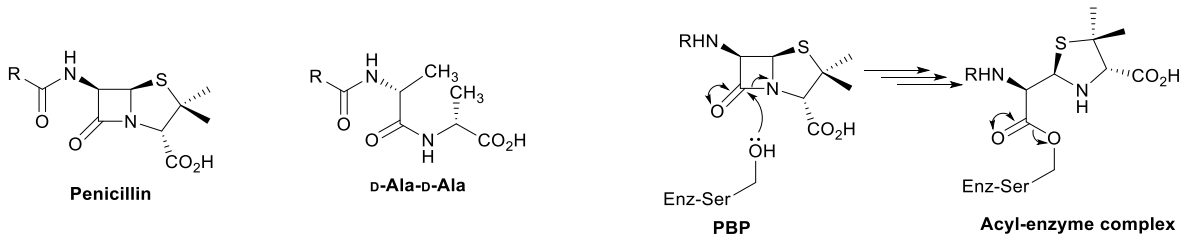



**Figure 1.6**. Structure of the penicillins and the D-Ala-D-Ala motif present on the terminus of peptidoglycans, and a mechanism for acylation of penicillin-binding proteins (PBPs).[33]

specifically produced by bacteria, and do not have close human homologues, helping to achieve selective toxicity for bacteria over humans. Currently, β-lactam based antibiotics are clinically the most prescribed antibiotics (65%).[37] The production of monocyclic β-lactams and carbapenems for clinical purposes, relies on syntheticroutes.[26,38,39] However, the production of cephalosporins and penicillins, which comprise almost 90% of the prescribed β-lactam based antibiotics,[37] relies on fermentation.[39] Thus, the biosynthetic/fermentation pathway by which cephalosporins and penicillins are prepared is important.

## 1.2 Biosynthesis and fermentation based production of β-lactams

### 1.2.1 ACV synthase

The direct biosynthetic pathway for the production of penicillins and cephalosporins originates with the assembly for their common precursor tripeptide L-δ-(α-aminoadipoyl)-L-cysteinyl D-valine (**LLD-ACV**, or **ACV**, Section **1.4.1**) from its constituent amino acids; L-Aad, L-Cys, and L-Val.[40,41] The first evidence for this step came from cell free extracts from *C. acremonium* which demonstrated to synthesize **ACV** from its three constituent amino acids. It was unclear at the time if **ACV** was prepared by a single or multiple enzymes. Conclusive evidence for the biosynthetic preparation of **ACV** came from the extraction and purification of a multidomain, multimodular, non-ribosomal peptide synthetase, ACV synthethase (ACVS), from dry mycelium (Figure 1.7).[42] Incubations of isolated ACVS with L-Cys, L-Aad, and L-Val, and subsequently with isotopically enriched L-3-[$^2$H]-Val or $H_2O$, suggested the epimerisation occurs during the catalysis (Figure 1.7).[43] The source of L-Aad for the formation of **ACV** in fungi such as *P. chrysogenum*, is Lys biosynthesis. Aad is produced as an intermediate in a pathway shared by the biosynthesis of penicillins and Lys in fungi. Disruption of the biosynthetic pathway for Lys directly correlates with a decrease in penicillin

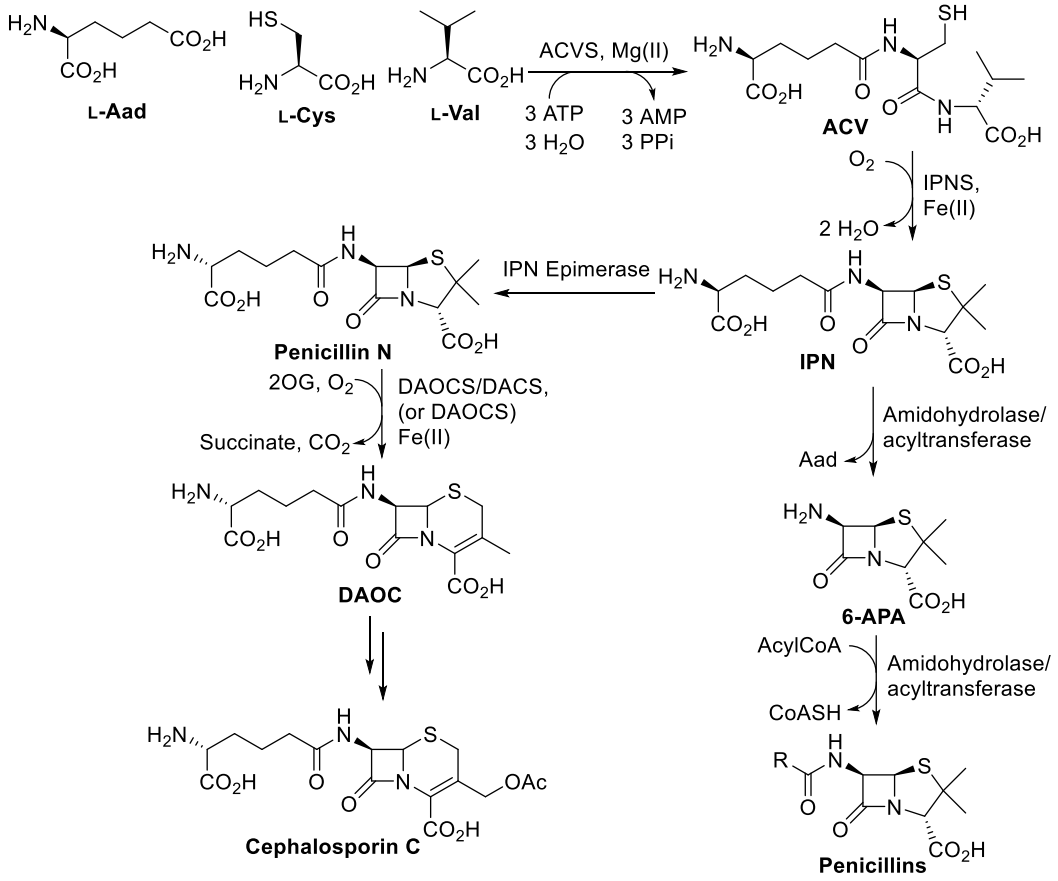



**Figure 1.7**. Biosynthetic pathway for the production of penicillins and cephalosporins.[46] ACV synthase (ACVS) catalyses the formation of **ACV** and consumes adenosine triphosphate (ATP) and $H_2O$ to form adenosine monophosphate (AMP) and pyrophosphate (PPi); isopenicillin N synthase (IPNS) catalyses the formation of **IPN** from **ACV** and consumes $O_2$ to produce 2 molecules of $H_2O$; an amidohydrolase/acyltransferase removes the 2-aminoadypoyl (Aad) sidechain to give 6-aminopenicillanic acid (**6-APA**), and can transfer acylated coenzyme A (CoA) to form acylated penicillins; IPN epimerase epimerises L-Aad to D-Aad; the bifunctional enzyme deacetocycephalosporin C synthase / deacetylcephalosporin C synthase (DAOCS/DACS) catalyses the formation of deacetoxycephalosporin C (DAOC) using 2-oxoglutarate (2OG), and is involved in the subsequent oxidation of the exocyclic methyl group (not shown), before an acyltransferase (not shown) catalyses the formation of cephalosporin C.[46]

formation.[44] The tripeptide **ACV** is subsequently converted into the β-lactam and thiazolidine containing isopenicillin N (**IPN**), catalysed by isopenicillin N synthase (IPNS, Figure 1.7), as will be discussed in depth later on (Section **1.3**).[45–47]

### 1.2.2 Isopenicillin N amidohydrolase and acyltransferase

After formation of **IPN**, amidohydrolases / acyl transferases can then remove the Aad side chain from **IPN**, producing 6-amino penicillanic acid (**6-APA**).[48] In addition to deacylating **IPN**, some acyl transferases are able to exchange the Aad chain for different hydrophobic groups (Figure 1.7).[39,49] Acyl transferase activity was observed in fermentation of some fungi, including *P. chrysogenum*, in broths supplemented with substituted acetic acids, which resulted in the formation of penicillins with different side chains.[50] Typically, the residues were hydrophobic and some of the newly formed penicillins displayed improved antimicrobial activity compared to penicillin N.[19] Isotopic labelling studies demonstrated exchange of $^{35}S$ between penicillin G and V when incubating both compounds with *P. chrysogenum*,[51,52] which was confirmed by studies using partially purified isopenicillin N acyl transferase (AT) enzyme prepared from cell free *P. chrysogenum* extracts.[53] Studies with fully purified AT obtained from *P. chrysogenum*, showed that the Aad sidechain in **IPN** is exchanged for a benzyl group (forming penicillin G) derived from phenylacetyl coenzyme A (CoA).[54] In presence of acyl-CoA substrates the AT enzyme acylates **6-APA** (Figure 1.7). When sufficient **IPN** was available, but in absence of any appropriate acyl-CoA, formation of **6-APA** was observed. The affinity of the enzyme for **IPN** was significantly lower in absence of a phenylacetyl CoA, suggesting **6-APA** remains bound to the enzyme during catalysis. Crystallographic studies demonstrated that the AT enzyme is unable to accommodate the larger cephalosporin ring, and thus are unable to exchange side chains on cephalosporins, separating the penicillin and cephalosporin pathways.[55]

### 1.2.3 Isopenicillin N epimerase

The ability to exchange the Aad side chain in **IPN** with non-polar groups occurs in some fungi, including *P. chrysogenum* and *Aspergillus nidulans*, but not by other fungi such as *C. acremonium*.[54,56] Some fungi (including *C. acremonium*) are able to epimerise the L-Aad side chain of **IPN** to the D-Aad sidechain at penicillin N, which is the substrate for the cephalosporin forming ring expandase enzyme (Figure 1.7). **IPN** is not a substrate for the ring expansion enzyme. Thus the formation of penicillin N by epimerisation branches the biosynthetic preparation of penicillins and cephalosporins.[46] In prokaryotes the epimerisation of **IPN** is catalysed by a pyridoxal dependent epimerase. The epimerisation mechanism is likely to involve imine or

enamine formation, as proposed for related epimerases.[46,57] In fungi the epimerisation of **IPN** is proposed to involve 3 enzymatic steps, which are encoded by two genes cefD1 and cefD2.[58]

### 1.2.4 Cephalosporins

In fungi, penicillin N undergoes ring expansion to form deacetoxycephalosporin C (DAOC, Figure 1.7).[59,60] Ring expansion is catalysed by a 2-oxoglutarate (2OG) dependent bifunctional oxygenase, deacetoxycephalosporin C synthase and deacetylcephalosporin C synthase (DAOCS/DACS). In prokaryotes such as *S. clavuligerus*, DAOC is prepared in two sequential steps by two separate enzymes; DAOCS and DAOC hydroxylase.[61] In fungi DAOC is hydroxylated by DAOCS/DACS and can then be acylated to give cephalosporin C (Figure 1.7).[59] Notably, **IPN** is not a substrate for ring expansion by DAOCS/DACS, separating the biosynthetic pathways of penicillins and cephalosporins.

## 1.3 Isopenicillin-N synthase

### 1.3.1 Discovery of ACV and IPNS

IPNS facilitates the formation of the penicillin core with its fused β-lactam and thiazolidine ring system (Figure 1.7). As discussed above, after formation, **IPN** can be functionalised to form different penicillins or cephalosporins. Thus the oxidation catalysed by IPNS is a key reaction in preparation of cephalosporin and penicillin based antibiotics. What follows here is a brief discussion of the discovery of IPNS and the work done to elucidate the catalytic mechanism of IPNS.

The discovery of a common precursor peptide that was shared between the penicillins and cephalosporins started with the incorporation of isotopically labelled Val and Cys into penicillins and cephalosporins.[62,63] Small amounts of δ-(α-aminoadipoyl)-cysteinyl valine (stereochemistry not defined) were isolated from mycelium of *P. chrysogenum*[64,65] and *C. acremonium*[66]. Cell free systems of the fungi were developed in order to overcome mycelium permeability associated difficulties that hindered further studies on the conversion of synthetic **ACV** analogues.[67] Incubation studies with the LLD, LLL, and DLD diastereomers of δ-(α-aminoadipoyl)-cysteinyl [4,4'-$^{3}$H]-valine using cell free systems demonstrated that the LLD-diastereomer was the precursor for penicillin N.[68] Cell free incubation with **ACV** suggested **IPN** was

formed first, and subsequently converted into penicillin N.[69,70] These studies demonstrated that the conversion required the presence of Fe(II), $O_2$, a reducing agent such as ascorbate, and a dithiol reducing agent such as dithiothreitol, to ensure high activity (Figure 1.8). IPNS was first purified by HPLC from *C. acremonium*.[71,72] Later on the gene encoding for IPNS was cloned into *Escherichia coli*, expressed and characterised,[73,74] and has since been isolated from multiple sources, with a range of sequence similarities.[75–83]

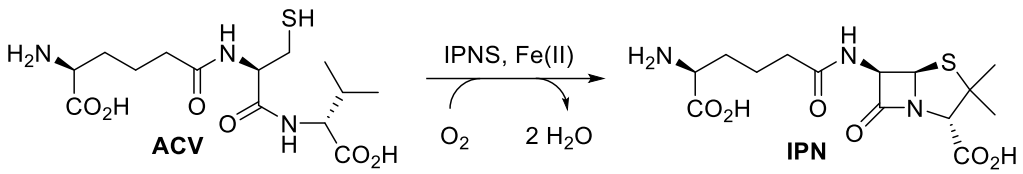



**Figure 1.8**. The reaction catalysed by isopenicillin N synthase (IPNS) converting **ACV** into isopenicillin N (**IPN**) consuming $O_2$ and producing $H_2O$.[45,46]

## 1.3.2 ACV analogues and the mechanism of IPNS

### 1.3.2.1 α-Amino adipic acid analogues

The discovery of **ACV** as the precursor for penicillins and the availability of cell free systems and subsequently, recombinant IPNS, allowed for mechanistic investigations. Initial attempts focused on the use of substrate analogues[84] in cell free systems, leading to the hypothesis of a two-step mechanism in which the formation of a mono cyclic β-lactam was initially formed, followed by a second cyclisation to form the thiazolidine ring.[84,85] Despite extensive investigations, no evidence was found for any enzyme-free intermediate(s). The Aad sidechain requires a 6 carbon chain to be accepted as a substrate of IPNS, as shown by turnover studies using **ACV** analogues (**4** and **5**, Figure 1.9) containing Aad sidechains of varying lengths.[86] However, both L and D stereochemistry of the 2-aminoadipoyl side chain **1** and an analogue containing adipic acid **2** were substrates for the IPNS catalysed reaction, but not amine **6** (Figure 1.9). Interestingly, some **ACV** analogues containing hydrophobic residues (e. g. benzyl group, **3**) on the Aad position were still turned over to their penicillin products, albeit as poor substrates.[84]

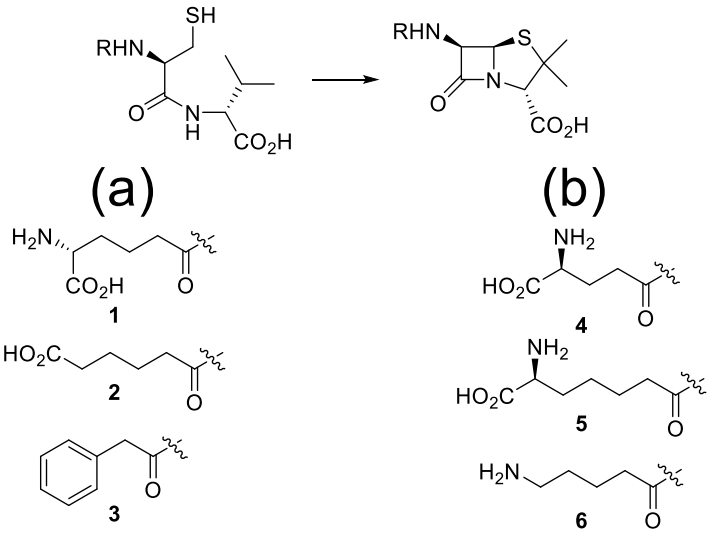



**Figure 1.9**. An overview of some of the alternative substrates tested for conversion by IPNS.[84] Column (a) shows substrate analogues that were converted into their corresponding penicillin products by IPNS, while column (b) shows substrate analogues that were not converted into their corresponding penicillin products by IPNS.

### 1.3.2.2 Cysteine analogues

A limited number of modifications on the **ACV** Cys gave tripeptides that were still accepted as substrates for IPNS turnover.[84] Cell free incubation studies using stereoselectively deuterated **ACV 7** and **8** as substrates, demonstrated that IPNS performs a stereospecific cyclisation of the substrate (Figure 1.10).[87] In addition, specifically deuterated substrates (i. e. L-δ-(α-aminoadipoyl)-L-3,3-[$^2H_2$]-cysteinyl D-valine (**AdCV**) and L-δ-(α-aminoadipoyl)-L-cysteinyl D-3-[$^2$H]-valine (**ACdV**, Figure 1.10) were mixed with non-labelled substrate (1 : 1) and incubated with IPNS.[85] After derivatisation and work up, the ratios of the starting peptides provided evidence for the formation of the β-lactam ring, before the formation of the thiazolidine ring.[85] Substitution of **ACV** Cys by homocysteine (Hcy) giving tripeptide Aad-Hcy-Val gave γ-lactams upon incubation with IPNS.[88] The presence of a thiol in the substrate remains mandatory for catalysis by IPNS.

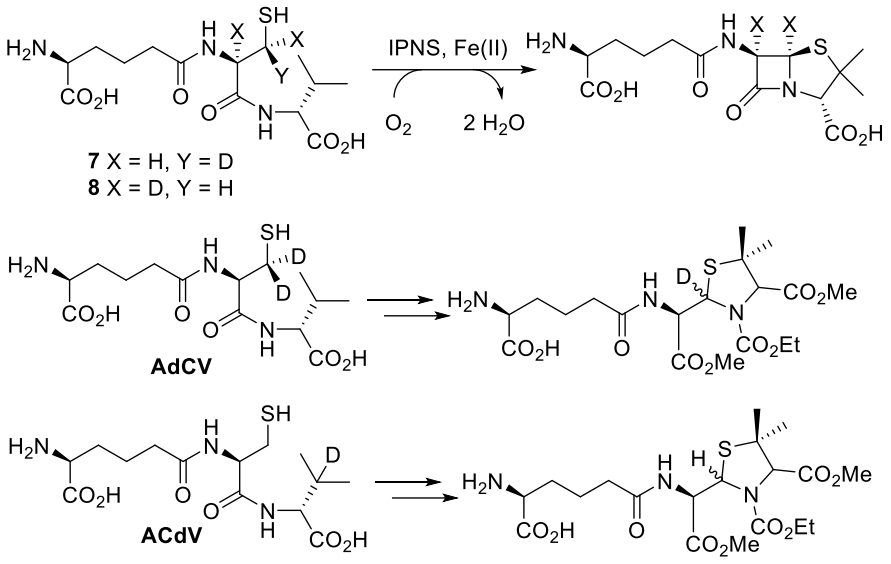



**Figure 1.10**. Selectively deuterated **ACV** substrate analogues used to probe the mechanism of IPNS.[84] Note that the products of the reactions of **AdCV** and **ACdV** were obtained after quenching the IPNS catalysed reaction and derivatisation. The ratios of the remaining protonated starting material and deuterated starting material were compared.

### 1.3.2.3 Valine analogues

Substrate analogues containing unnatural amino acids with unsaturated aliphatic side chains or hydrophobic groups at the Val position of **ACV** have extensively been tested.[84] For example incubation studies with aminobutyrate containing **ACV** analogues and selectively deuterated aminobutyrate containing **ACV** analogues **9** and **10** suggested the possibility for a free-radical mechanism in which inversion of stereochemistry would be possible (Figure 1.11).[89,90] To test for a possible free-radical mechanism in the formation of penicillins, incubation studies of cyclopropyl Ala containing **ACV** analogue **11** were performed (Figure 1.11).[91] The products identified after incubation were suggestive of a free-radical mechanism.

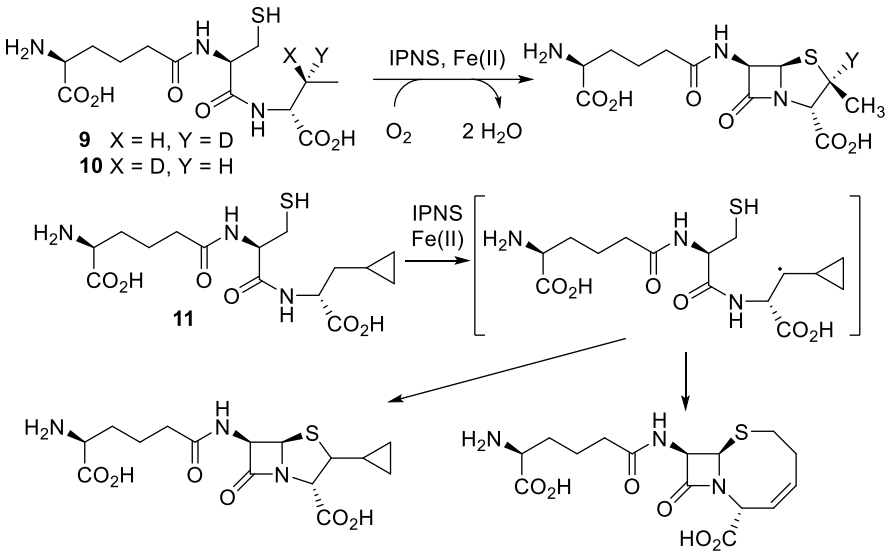



**Figure 1.11**. Analogues of **ACV** substrate that have been incubated with IPNS in order to probe the catalytic mechanism.[84]

### 1.3.3 Spectroscopic studies on IPNS

Mössbauer spectroscopy on IPNS with $^{57}Fe$ enriched enzyme shows that iron binds as a high-spin Fe(II) to the active site.[92] A decrease in the isomer shift was observed upon exposure of the IPNS : Fe complex to **ACV** under anaerobic conditions, suggesting substrate binding perturbs the Fe centre. Nitric oxide (NO) was used as an $O_2$ surrogate and was observed to bind to the Fe(II) in the IPNS : Fe complex. Electron paramagnetic resonance (EPR) analysis of the IPNS : Fe : NO complex was in agreement with binding of the **ACV** thiol in proximity of Fe. UV-VIS spectroscopic analysis of the IPNS : Fe : ACV : NO complex gave a maximum absorbance at 508 nm. Using Fe edge extended X-ray absorption fine structure (EXAFS) techniques on the anaerobic complex provided further evidence for the chelation of a thiol to Fe(II) upon exposure of the IPNS : Fe complex to **ACV** substrate.[93] Furthermore, the data were consistent with the hypothesis that without **ACV** substrate, the Fe(II) was only coordinated by N/O containing ligands. Mutagenesis work in combination with additional EXAFS experiments, confirmed that thiol chelation to the Fe(II) centre derived from **ACV** substrate, and not from endogenous Cys present in wt-IPNS.[94] Binding of the **ACV** thiol to Fe(II) centre promoted binding of $O_2$, as a consequence of charge donation from the thiolate to Fe.[95,96] Formation of a Fe(III) superoxide becomes energetically favourable.

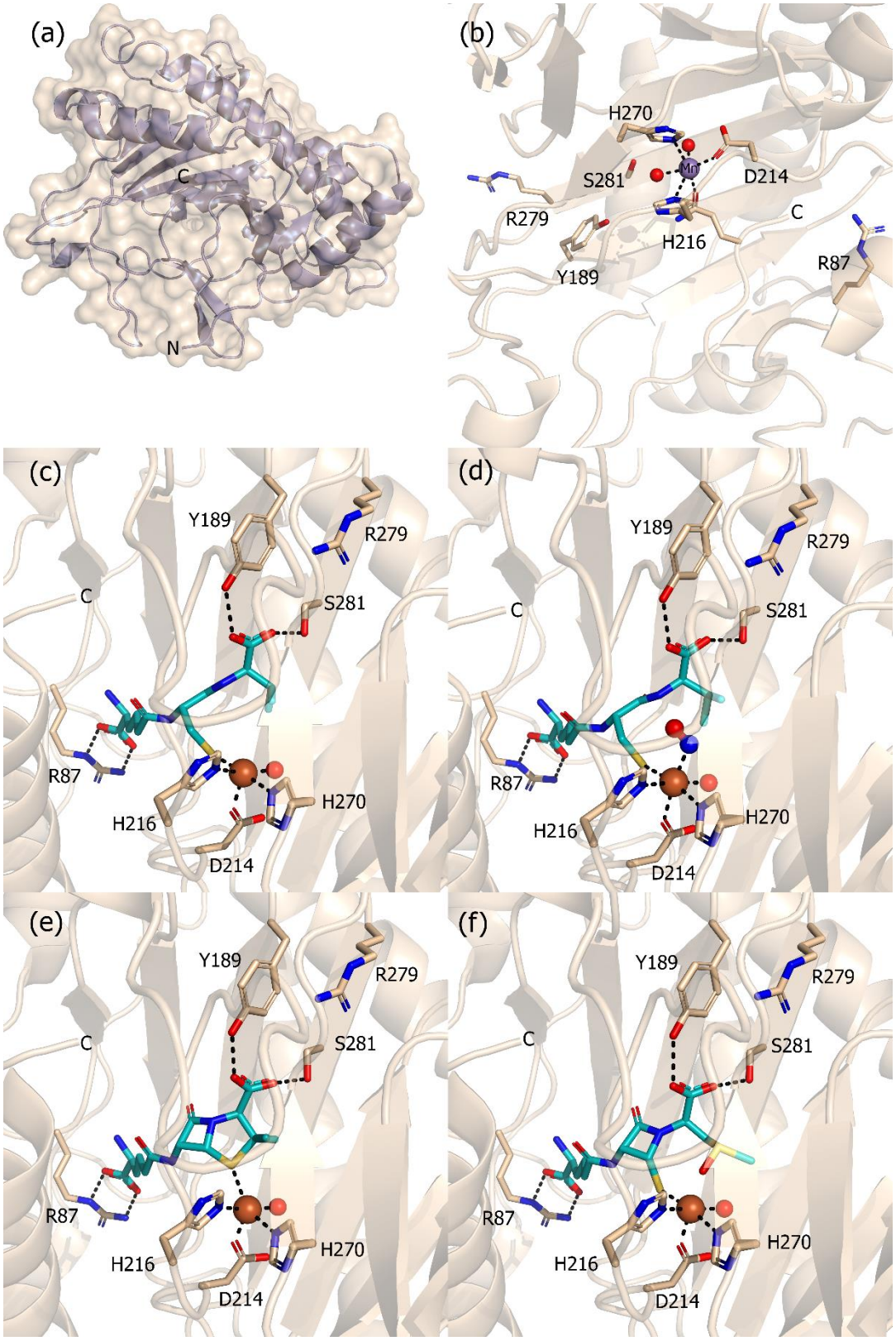



**Figure 1.12**. Views from crystal structures of IPNS. (a) The overall fold of IPNS : Mn(II); view of the active site of (b) IPNS : Mn(II); (c) IPNS : Fe(II) : ACV; (d) IPNS : Fe(II) : ACV : NO; (e) IPNS : Fe(II) : IPN, note that **ACV** (not shown) is also modelled in the active site (30%); (f) IPNS : Fe(II) : ACmC after exposure to hyperbaric $O_2$. PDB accession codes in order of display: 1IPS,[97] 1BK0,[98] 1BLZ,[98] 1QJE,[99] 1QJF[99].

### 1.3.4 Crystallography of IPNS

The first crystals obtained from *apo*-IPNS derived from recombinant *C. acremonium* IPNS, and diffracted to 3.5 Å.[100] Subsequently, Roach *et al.* solved the first structure of IPNS from *A. nidulans* in complex with Mn(II) to 2.5 Å (Figure 1.12 (a) and (b)).[97,101] The overall fold of the resting state structure of IPNS consists of a distorted double-stranded β-helix (DSBH, also known as "jelly roll", Figure 1.12 (a)) core. At the active site, a single metal is coordinated by Asp216, His214, His270 (also referred to as the "facial triad", Figure 1.12 (b)), and Gln330. Subsequently, anaerobic structures of IPNS in complex with **ACV** and Fe(II), show that upon binding of **ACV**, Gln330 and a $H_2O$ were displaced from the Fe centre (Figure 1.12 (c)).[98,102] This leads to the formation of a five-coordinate Fe site, which had a vacant site available for $O_2$ binding. Displacement of Gln330 is accompanied by a rearrangement of the *C*-terminal region (Asn326 – Thr331). The **ACV** thiolate was chelating to the Fe(II) centre in the IPNS : Fe : ACV complex. The **ACV** Aad sidechain made a salt bridge with Arg87. The **ACV** Val carboxylic acid is positioned to bind to Ser281 and Tyr189, and *via* an $H_2O$ mediated interaction with Arg279. A hydrophobic pocket, created by residues Thr221, Leu223, Leu231, Val272, and Pro283, surrounds the **ACV** Val isopropyl group. The residues that made up this hydrophobic pocket are important for substrate recognition and catalysis.[103] A complex of IPNS : Fe : ACV : NO was obtained by exposing IPNS • Fe • ACV to NO (Figure 1.12 (d)).[98] NO was used as a surrogate for $O_2$, and was observed to bind *via* its nitrogen *trans* to Asp216 at the vacant site of Fe(II), adjacent to the **ACV** thiolate. Consequently, the **ACV** Val isopropyl group moves away from the Fe centre to accommodate the NO ligand. No further changes in the overall protein structure were observed. Attempts to observe turnover of **ACV** substrate by IPNS *in crystallo* by exposing IPNS • Fe • ACV to $O_2$ under hyperbaric pressure (40 bar), resulted in partial formation of **IPN** (Figure 1.12 (e)).[99] Residual electron density suggested that a fraction of the crystal contained **ACV** (~30%, not illustrated in Figure 1.12(e)). Extended exposure times IPNS • Fe • ACV crystals to $O_2$ under hyperbaric pressure resulted in low occupancy, suggesting dissociation of the **IPN** product. So far attempts to crystallize IPNS in presence of **IPN** have not yet been reported. This could be due to the limited stability of **IPN** in aqueous solutions over longer periods of time. IPNS was successfully been crystallized with a cyclobutanone analogue of **IPN**.[104] Further evidence for the formation of the β-lactam ring prior to thiazolidine ring formation, came from crystals of IPNS in complex with Fe(II) and substrate analogue

δ-(α-aminoadipoyl)-cysteinyl *S*-methyl cysteine (**ACmC**).[99] This substrate analogue was designed to interrupt the catalytic cycle after formation of the β-lactam ring, which it did after exposure to $O_2$ under hyperbaric pressure. Additional electron density at the *S*-methyl Cys suggested oxidation of the sulfur, which was rationalized by reaction of the thio-ether with the Fe(IV) oxo species present after formation of the β-lactam ring (Figure 1.12 (f)). Direct crystallographic evidence for the formation of the Fe(IV) oxo species was not provided. Kinetic and spectroscopic studies have been conducted on IPNS using natural isotope **ACV** and deuterium labelled substrates; L-δ-(α-aminoadipoyl)-L-3,3-[$^2H_2$]-cysteinyl D-valine (**AdCV**) and L-δ-(α-aminoadipoyl)-L-cysteinyl D-2,3,4,4,4,4',4',4'-[$^2H_8$]-valine (**ACd$_8$V**).[105] The strong KIE of the deuterium labelled substrates was leveraged to prolong the lifetimes of the Fe(III) superoxide and Fe(IV) oxo intermediates, allowing for characterisation by Mössbauer and UV-VIS spectroscopy. Based on these combined crystallographic, spectroscopic, computational, and substrate analogue observations on the reaction of IPNS have led to the proposed mechanism for its catalytic turnover of **ACV** into **IPN** (Figure 1.13).

### 1.3.5 Proposed mechanism of IPNS

In the currently proposed IPNS mechanism, **ACV** binds to the Fe(II) at the active site, displacing Gln330 and a Fe-ligated $H_2O$ (Figure 1.13 **A** to **B**). Chelation of the **ACV** thiol activates Fe(II) for binding of $O_2$, creating a Fe(III) superoxide (Figure 1.13 **B**). The formed Fe(III) superoxide abstracts the pro-(*S*) cysteinyl β-hydrogen (Figure 1.13 **D**). Inner sphere electron transfer results in the formation of an Fe(II) chelating a thiolaldehyde (Figure 1.13 **E**). The so formed Fe(II) hydroperoxo intermediate then abstracts a proton from the cysteinyl valine amide. Nucleophilic attack of the amide nitrogen onto the thioaldehyde, results in the formation of the monocyclic β-lactam intermediate linked to a ferryl Fe(IV) oxo species (Figure 1.13 **F**). The latter then abstracts the **ACV** Val β-hydrogen in a radical mechanism to form a ferric Fe(III) hydroxyl intermediate (Figure 1.13 **G**). The Val radical recombines with an electron on sulfur forming the thiazolidine ring and completing the synthesis of **IPN** and restoring the original Fe(II) oxidation state (Figure 1.13 **H**).[45,46]

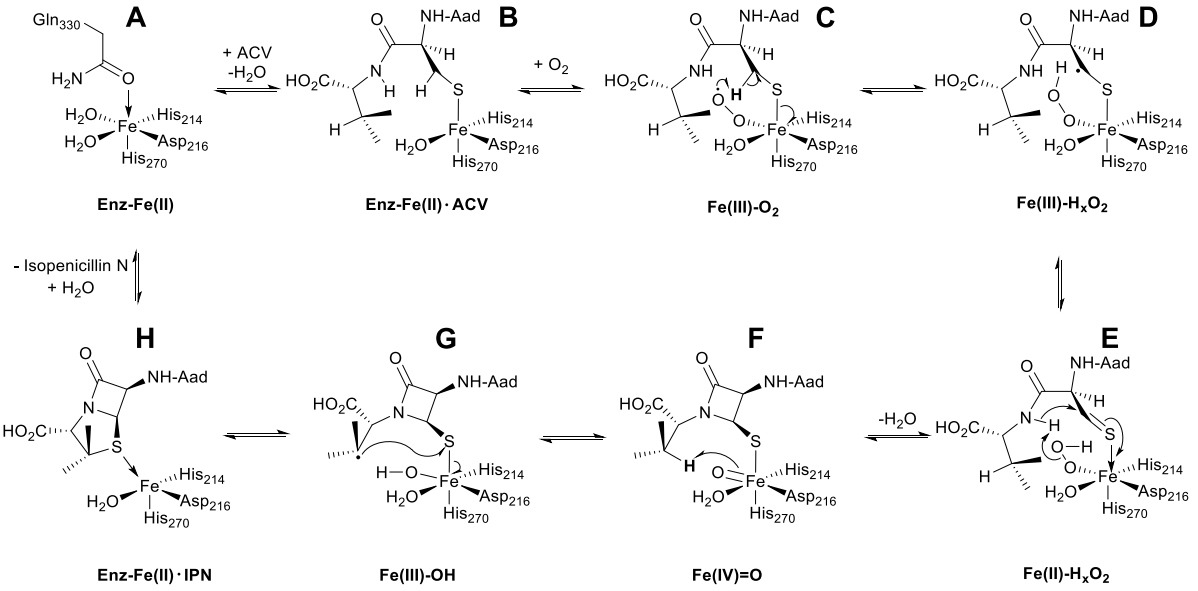



**Figure 1.13.** The proposed mechanism for the IPNS catalysed turnover of **ACV**, which is based on substrate analogue, crystallographic, computational, and kinetic studies.[45,46] *This scheme is an adaptation from Rabe et al.*[45]

## 1.4 Antibiotic resistance and β-lactamases

Likely due to the natural production of β-lactam products by fungi, bacteria can evolve mechanisms to counteract β-lactams. Microorganisms in environments apparently devoid of antibiotics, contain multiple genes that encode for mechanisms to combat β-lactams, suggesting their origin precede the use of antibiotics by mankind as a way to treat bacterial infections.[106] Genome studies estimate the origin of certain β-lactamases to be around 2 billion years ago.[106,107] The first observation of resistance against β-lactam antibiotics was reported a year before the clinical introduction of penicillin as an antibiotic.[108]

In recent years the effectiveness of β-lactam antibiotics has increasingly been threatened by the development of antibiotic resistance.[109] Of particular interest are carbapenem resistant bacterial strains. Carbapenems are often considered the last line defence against multi resistant strains. Global initiatives and national policies have been adjusted to address antibiotic resistance in pathogenic bacterial strains.[110] A brief description of how resistance can be manifested in microorganisms and in particular how β-lactamases confer resistance is given below.

### 1.4.1 Resistance through alteration of penicillin-binding proteins

Modification of the antibiotic target, i. e. PBPs for β-lactam antibiotics, either through mutation or by obtaining genes encoding for a variant of the target, is an important resistance mechanism.[111] Methicillin-resistant *Staphylococcus aureus* (MRSA) have acquired the *mec* gene, which encodes for PBP2, which displays less affinity to β-lactam antibiotics.[112] Obtaining resistance through production of PBP variants is sometimes referred to as intrinsic resistance to distinguish it from other mechanisms.[111]

### 1.4.2 Resistance through cell wall permeability

The permeability of the cell wall can be changed by altered expression or structural modification of membrane proteins (porins), prohibiting β-lactam antibiotics from reaching their PBP targets in the periplasm.[113] A decrease in *oprD* expression in *Pseudomonas aeruginosa* is associated with imipenem resistance.[114,115] In combination with overexpression of efflux pumps, such as MexAB-OprM, which actively remove small molecules including β-lactam antibiotics, a barrier is created.[116,117] Typically, a combination of resistance mechanisms manifests resistance in bacterial strains.

### 1.4.3 Resistance through enzymatic inactivation

Enzymatic hydrolysis of β-lactam antibiotics is considered the most important mechanism for resistance.[118,119] The first report of a β-lactam antibiotic resistance describes enzymatic inactivation of penicillin by a serine β-lactamase (SBL).[108] In 1965, a different class of enzymes, i.e. one utilising Zn(II) ions to activate a $H_2O$ for the hydrolysis of the β-lactam ring was discovered; metallo β-lactamases (MBLs).[120] Mechanistically MBLs and SBLs are two distinctly different enzymes. Because of low sequence similarity between MBLs and SBLs, they are thought to have different evolutionary origins.[106] Based on sequence similarity, SBLs are divided into Ambler classes A, C, and D,[106,121] and MBLs are grouped as class B. The SBLs share structural features and highly conserved residues with PBPs.[122] Class A SBLs are most common, and facilitated resistance against penicillins and cephalosporins. More recently, extended spectrum SBLs CTX-M and KPC, have been reported at an increased frequency. Class C SBLs, AmpC

enzymes, are generally encoded in the chromosomes, and are insensitive to clavulanic acid and their preferred substrates are cephalosporins.[119]

### 1.4.4 Class D serine β-lactamases and OXA-10

Class D SBLs, or oxacillinases (OXAs) were among the first β-lactamases to be identified.[123,124] Initially OXA enzymes were uncommon and their transfer was plasmid mediated, with penicillins as their preferred substrates.[125] OXA enzymes were found to hydrolyse oxacillins more rapidly than penicillins. Over time the OXA enzymes have evolved to confer carbapenem resistance, and certain strains (e. g. *Acinetobacter baumannii*) were discovered which have chromosomal encoded OXA enzymes.[123] Related *Acinetobacter* species were able to transfer their OXA enzymes, enhancing carbapenem resistance. In general, OXA enzymes are resistant to SBL inhibitors; their activity is affected by the presence of halide ions at high concentrations.[126] Sequence alignment of multiple OXA enzymes identified several conserved motives, including Ser-X-X-Lys (Ser67 for OXA-10) which is part of the active site.[127] An active site Lys is carbamylated, and stabilized by a nearby Trp;[128] the carbamylated Lys acts as a general acid and base during hydrolysis (Figure 1.14).[129] The evolution of OXA enzymes to mediate carbapenem resistance, causes concerns for the clinical efficacy of carbapenems.

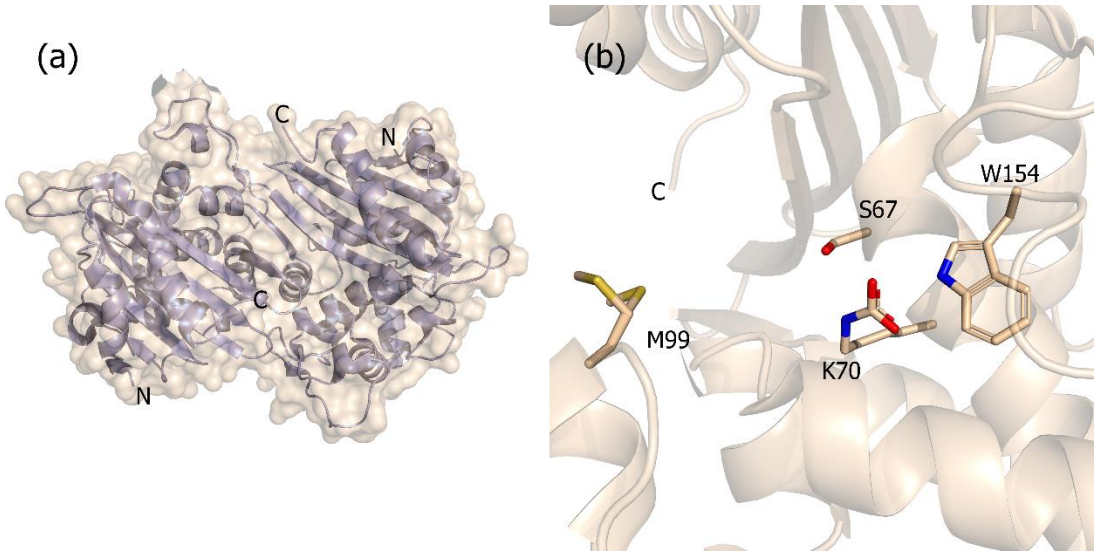



**Figure 1.14.** Views from a crystal structure of OXA-10. (a) The overall fold of OXA-10 (two of the four molecules in the unit cell are displayed); (b) view of the active site showing the nucleophilic Ser67 which is used for hydrolysis, and carbamylated Lys70. PDB accession code 1K55.[129]

### 1.4.5 Hydrolysis mechanism of serine β-lactamases

Certain steps of the mechanism by which SBLs hydrolyse β-lactam antibiotics are similar to the inhibition of PBPs by β-lactam antibiotics.[130] After binding of the β-lactam compound nucleophilic attack of an active site Ser opens the β-lactam ring creating an acyl-enzyme complex. For PBPs, hydrolysis of the acyl-enzyme complex is slow, leading to inhibition (Figure 1.15).[131] By contrast, hydrolysis of the SBLs acyl-enzyme complex by a hydrolytic $H_2O$ molecule is rapid, releasing the hydrolysed β-lactam (Figure 1.15).[132] Activation of hydrolytic $H_2O$ is achieved by active site residues, such as Ser, Glu, Tyr and Lys in class A and C SBLs.[133,134] In OXA enzymes a carbamylated Lys is used in general acid/base catalysis (Figure 1.14).[129]

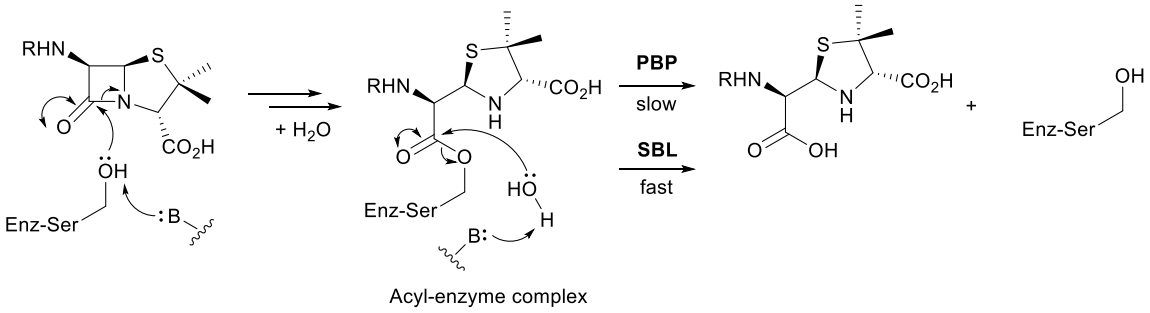



**Figure 1.15.** Outline of proposed mechanism for formation of the acyl-enzyme complex for penicillin-binding proteins (PBPs) and serine β-lactamases (SBLs). After formation of the acyl-enzyme complex, hydrolytic $H_2O$ facilitates cleavage of the antibiotic from the active site. Note that for PBPs the rate of hydrolysis of the acyl-enzyme complex is slower than for SBLs.[130]

### 1.4.6 Carbapenem hydrolysis by oxacillinases

The serine β-lactamase catalysed hydrolysis of carbapenems, similarly to that of penicillins and cephalosporins, starts with opening of the β-lactam ring to form an acyl-enzyme complex (Figure 1.16). Hydrolysis of the acyl-enzyme complex liberates the hydrolysed product.[135,136] Tautomerization between $\Delta^1$ pyrroline and $\Delta^2$ pyrroline leads to loss of the stereochemistry at C2 centre, and the formation of two diastereomers of the $\Delta^1$ pyrroline product. Recently, an alternative mechanism for the hydrolysis of the acyl-enzyme complex by OXA enzymes (i. e. OXA-10, OXA-23, OXA-48), one mediated by the hydroxy ethyl group present on carbapenems was reported (Figure 1.16).[137] The β-lactone products formed through this mechanism are able to inhibit SBLs, although generally displaying lower affinity compared to their corresponding β-lactams. Notably, the formation of β-lactones was demonstrated for carbapenems containing a 1-β-methyl group (i. e. ertapenem, meropenem, biopenem, and doripenem).

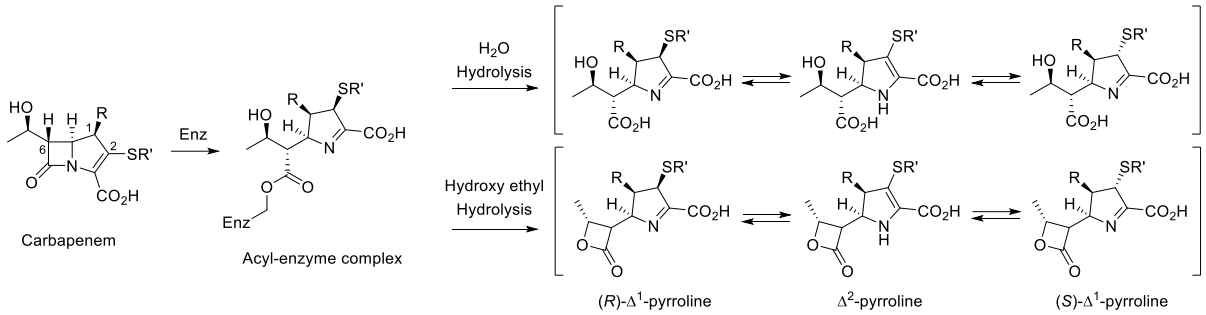



**Figure 1.16.** Overview of the major products formed during degradation of carbapenems by class D serine β-lactamases (SBLs). After formation of the acyl-enzyme complex, hydrolysis of the substrate results in formation of the inactive hydrolysed β-lactam, alternatively reactions to give β-lactone can occur.[136,137]

## 1.5 Aims

The work in this thesis related to OXA-10 is described in Chapter 3, and the work related to IPNS is described in Chapter 4. The broad aim of the work on OXA-10 was to investigate the hydrolysis mechanism of carbapenems by OXA-10, using time-resolved crystallography. The aim of the work on IPNS described in Chapter 4 was to study proposed intermediates (in particular Fe(III) superoxide and Fe(IV) oxo, Figure 1.13 **C** and **F**) crystallographically. Some of the crystallographic observations led to the investigation of structural rearrangements during catalysis in solution, and the investigation of the binding of Fe to the IPNS active site in presence of the **IPN** product.

## 1.6 References


1. Finch, C. E. Inflammatory Exposure and Historical Changes in Human Life-Spans. *Science.* **305**, 1736–1739 (2004).
2. Kyriazis, M. Ageing Throughout History : The Evolution of Human Lifespan. *J. Mol. Evol.* **88**, 57–65 (2020).
3. McDermott, W. & Rogers, D. E. Social ramifications of control of microbial disease. *Johns Hopkins Med. J.* **151**, 302–312 (1982).
4. Hollis, A. & Ahmed, Z. Preserving Antibiotics, Rationally. *N. Engl. J. Med.* **369**, 2474–2476 (2013).
5. Jones, H. W. Report of a Series of Cases of Syphilis Treated by Ehrlich's Arsenobenzol at the Walter Reed General Hospital, District of Columbia. *Bost. Med. Surg. J.* **164**, 381–383 (1911).
6. Schwartz, R. S. Paul Ehrlich's Magic Bullets. *N. Engl. J. Med.* **350**, 1079–1080 (2004).
7. Iyer, H. V. History Revisited—Prontosil Red. *J. Emerg. Med.* **35**, 209–210 (2008).
8. Laub, G. R. Discovery of the sulfa drugs. *South. Med. J.* **79**, 782 (1986).
9. Kalkut, G. Sulfonamides and Trimethoprim. *Cancer Invest.* **16**, 612–615 (1998).
10. Zaffiri, L., Gardner, J. & Toledo-Pereyra, L. H. History of Antibiotics. From Salvarsan to Cephalosporins. *J. Investig. Surg.* **25**, 67–77 (2012).
11. Kardos, N. & Demain, A. L. Penicillin: the medicine with the greatest impact on therapeutic outcomes. *Appl. Microbiol. Biotechnol.* **92**, 677–687 (2011).
12. Hare, R. New Light on the History of Penicillin. *Med. Hist.* **26**, 1–24 (1982).
13. Kong, K.-F., Schneper, L. & Mathee, K. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS* **118**, 1–36 (2010).
14. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. Influenze. *Br. J. Exp. Pathol.* **10**, 226–236 (1929).
15. Chain, E. *et al.* Penicillin as a chemotherpeutic agent. *Lancet* **236**, 226–228 (1940).
16. Abraham, E. P. *et al.* Further Observations on Penicillin. *Lancet* **238**, 177–189 (1941).
17. Quinn, R. Rethinking antibiotic research and development: World War II and the penicillin collaborative. *Am. J. Public Health* **103**, 426–434 (2013).
18. Crowfoot, D., Bunn, C. W., Rogers-Low, B. W. & Turner-Jones, A. The X-ray Crystallographic Investigation of the Structure of Penicillin. in *Chemistry of Penicillin* 310–366 (Princeton University Press, 1949).
19. Behrens, O. K. *et al.* Biosynthesis I Biological precursors for benzylpenicillin (Penicillin G). *J. Biol. Chem.* **175**, 751–764 (1948).
20. Sader, H. S. & Jones, R. N. Historical overview of the cephalosporin spectrum: Four generations of structural evolution. *Antimicrob. Newsl.* **8**, 75–82 (1992).
21. Newton, G. G. & Abraham, E. P. Isolation of cephalosporin C, a penicillin-like antibiotic containing D-α-aminoadipic acid. *Biochem. J.* **62**, 651–658 (1956).
22. Abraham, E. P. & Newton, G. G. F. The structure of cephalosporin C. *Biochem. J.* **79**, 377–393 (1961).
23. Hodgkin, D. C. & Maslen, E. N. The x-ray analysis of the structure of cephalosporin C. *Biochem. J.* **79**, 393–402 (1961).
24. Aoki, H. *et al.* Nocardicin a, a new monocyclic β-lactam antibiotic I. Discovery, isolation and characterization. *J. Antibiot. (Tokyo).* **29**, 492–500 (1976).
25. Tahlan, K. & Jensen, S. E. Origins of the β-lactam rings in natural products. *J. Antibiot. (Tokyo).* **66**, 401–410 (2013).
26. Sykes, R. B. & Bonner, D. P. Aztreonam: The first monobactam. *Am. J. Med.* **78**, 2–10 (1985).
27. Papp-Wallace, K. M., Endimiani, A., Taracila, M. A. & Bonomo, R. A. Carbapenems: Past, present, and future. *Antimicrob. Agents Chemother.* **55**, 4943–4960 (2011).

28. Butterworth, D., Cole, M., Hanscomb, G. & Rolinson, G. N. Olivanic acids, a family of β-lactam antibiotics with β-lactamase inhibitory properties produced by Streptomyces species. I. Detection, properties and fermentation studies. *J. Antibiot. (Tokyo).* **32**, 287–294 (1979).

29. Kahan, J. S. *et al.* Thienamycin, a new β-lactam antibiotic i. discovery, taxonomy, isolation and physical properties. *J. Antibiot. (Tokyo).* **32**, 1–12 (1979).

30. Fukasawa, M. *et al.* Stability of meropenem and effect of 1β-methyl substitution on its stability in the presence of renal dehydropeptidase I. *Antimicrob. Agents Chemother.* **36**, 1577–1579 (1992).

31. Paterson, D. L. Recommendation for treatment of severe infections caused by Enterobacteriaceae producing extended-spectrum β-lactamases (ESBLs). *Clin. Microbiol. Infect.* **6**, 460–463 (2000).

32. Paterson, D. L. & Bonomo, R. A. Clinical Update Extended-Spectrum Beta-Lactamases : a Clinical Update. *Clin. Microbiol. Rev.* **18**, 657–686 (2005).

33. Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. The penicillin-binding proteins: Structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**, 234–258 (2008).

34. McDonough, M. A. *et al.* Structures of two kinetic intermediates reveal species specificity of penicillin-binding proteins. *J. Mol. Biol.* **322**, 111–122 (2002).

35. Cho, H., Uehara, T. & Bernhardt, T. G. Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* **159**, 1300–1311 (2014).

36. Wong, F. & Amir, A. Mechanics and Dynamics of Bacterial Cell Lysis. *Biophys. J.* **116**, 2378–2389 (2019).

37. Bush, K. & Bradford, P. A. β-Lactams and β-Lactamase Inhibitors: An Overview. *Cold Spring Harb Perspect Med* 1–22 (2016).

38. Sasaki, A. & Sunagawa, M. Recent advances in carbapenem chemistry: (Review). *Chem. Heterocycl. Compd.* **34**, 1249–1265 (1998).

39. Hamed, R. B. *et al.* The enzymes of β-lactam biosynthesis. *Nat. Prod. Rep.* **30**, 21–107 (2013).

40. Baldwin, J. E. & Abraham, E. The biosynthesis of penicillins and cephalosporins. *Nat. Prod. Rep.* **5**, 129 (1988).

41. Byford, M. F., Baldwin, J. E., Shiau, C.-Y. & Schofield, C. J. The Mechanism of ACV Synthetase. *Chem. Rev.* **97**, 2631–2650 (1997).

42. Van Liempt, H., Von Dohren, H. & Kleinkauf, H. δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase from Aspergillus nidulans. The first enzyme in penicillin biosynthesis is a multifunctional peptide synthetase. *J. Biol. Chem.* **264**, 3680–3684 (1989).

43. Baldwin, J. E. *et al.* Exchange of the valine 2-H in the biosynthesis of L-δ-(α-aminoadipoyl)-L-Cysteinyl-D-valine. *Tetrahedron* **49**, 3221–3226 (1993).

44. Casqueiro, J., Gutiérrez, S., Bañuelos, O., Hijarrubia, M. J. & Martín, J. F. Gene targeting in Penicillium chrysogenum: Disruption of the lys2 gene leads to penicillin overproduction. *J. Bacteriol.* **181**, 1181–1188 (1999).

45. Rabe, P., Kamps, J. J. A. G., Schofield, C. J. & Lohans, C. T. Roles of 2-oxoglutarate oxygenases and isopenicillin N synthase in β-lactam biosynthesis. *Nat. Prod. Rep.* **35**, 735–756 (2018).

46. Hamed, R. B. *et al.* The enzymes of β-lactam biosynthesis. *Nat. Prod. Rep.* **30**, 21–107 (2013).

47. Schofield, C. J. *et al.* Proteins of the penicillin biosynthesis pathway. *Curr. Opin. Struct. Biol.* **7**, 857–864 (1997).

48. Whiteman, P. A. *et al.* Acyl coenzyme A: 6-aminopenicillanic acid acyltransferase from Penicillium chrysogenum and Aspergillus nidulans. *FEBS Lett.* **262**, 342–344 (1990).

49. Aplin, R. T., Baldwin, J. E., Cole, S. C. J., Sutherland, J. D. & Tobin, M. B. On the production of α,β-heterodimeric acyl-coenzyme A: isopenicillin N-acyltransferase of Penicillium chrysogenum. Studies using a recombinant source. *FEBS Lett.* **319**, 166–170 (1993).

50. Behrens, O. K. *et al.* Biosynthesis of penicillins; biological precursors for benzylpenicillin (penicillin G). *J. Biol. Chem.* **175**, 793–809 (1948).

51. Peterson, W. H. & Wideburg, N. E. Enzymatic interconversion of penicillins G and V. *Int. Abstr. Biol. Sci., Suppl.* 136 (1958).

52. Peterson, W. H. & Wideburg, N. E. Enzymatic interconversion of penicillins G and V. *Proc. Intern. Congr. Biochem. 4th Vienna* **15**, 136–137 (1960).

53. Pruess, D. L. & Johnson, M. J. Penicillin acyltransferase in Penicillium chrysogenum. *J. Bacteriol.* **94**, 1502–1508 (1967).

54. Alvarez, E. *et al.* The isopenicillin-N acyltransferase of Penicillium chrysogenum has isopenicillin-N amidohydrolase, 6-aminopenicillanic acid acyltransferase and penicillin amidase activities, all of which are encoded by the single penDE gene. *Eur. J. Biochem.* **215**, 323–332 (1993).

55. Bokhove, M. *et al.* Structures of an Isopenicillin N Converting Ntn-Hydrolase Reveal Different Catalytic Roles for the Active Site Residues of Precursor and Mature Enzyme. *Structure* **18**, 301–308 (2010).

56. Batchelor, F. R., Doyle, F. P., Nayler, J. H. C. & Rolinson, G. N. Synthesis of Penicillin: 6-Aminopenicillanic Acid in Penicillin Fermentations. *Nature* **183**, 257–258 (1959).

57. Watanabe, A., Yoshimura, T., Mikami, B. & Esaki, N. Tyrosine 265 of alanine racemase serves as a base abstracting α-hydrogen from L-alanine: The counterpart residue to lysine 39 specific to D-alanine. *J. Biochem.* **126**, 781–786 (1999).

58. Ullán, R. V. *et al.* A novel epimerization system in fungal secondary metabolism involved in the conversion of isopenicillin N into penicillin N in Acremonium chrysogenum. *J. Biol. Chem.* **277**, 46216–46225 (2002).

59. Baldwin, J. E. *et al.* Purification and initial characterization of an enzyme with deacetoxycephalosporin C synthetase and hydroxylase activities. *Biochem. J.* **245**, 831–841 (1987).

60. Dotzlaf, J. E. & Yeh, W. K. Copurification and characterization of deacetoxycephalosporin C synthetase/hydroxylase from Cephalosporium acremonium. *J. Bacteriol.* **169**, 1611–1618 (1987).

61. Baker, B. J., Dotzlaf, J. E. & Yeh, W.-K. Deacetoxycephalosporin C hydroxylase of Streptomyces clavuligerus. Purification, characterization, bifunctionality, and evolutionary implication. *J. Biol. Chem.* **266**, 5087–5093 (1991).

62. Abraham, E. P., Newton, G. G. F. & Hale, C. W. Purification and some properties of cephalosporin N, a new penicillin. *Biochem. J.* **58**, 94–102 (1954).

63. Arnstein, H. R. V. & Clubb, M. E. The biosynthesis of penicillin. 8. Investigation of cyclic cysteinylvaline peptides as precursors. *Biochem. J.* **68**, 528–535 (1958).

64. Arnstein, H. R. V. & Morris, D. The structure of a peptide, containing α-aminoadipic acid, cystine and valine, present in the mycelium of Penicillium chrysogenum. *Biochem. J.* **76**, 357–361 (1960).

65. Flynn, E. H., McCormick, M. H., Stamper, M. C., DeValeria, H. & Godzbski, C. W. A New Natural Penicillin from Penicillium Chrysogenum. *J. Am. Chem. Soc.* **84**, 4594–4595 (1962).

66. Loder, P. B. & Abraham, E. P. Isolation and nature of intracellular peptides from a cephalosporin C-producing Cephalosporium sp. *Biochem. J.* **123**, 471–476 (1971).

67. Fawcett, P. A., Loder, P. B., Duncan, M. J., Beesley, T. J. & Abraham, E. P. Formation and properties of protoplasts from antibiotic producing strains of Penicillium chrysogenum and Cephalosporium acremonium. *J. Gen. Microbiol.* **79**, 293–309 (1973).

68. Fawcett, B. P. A. *et al.* Synthesis of delta-(alpha-Aminoadipyl ) cysteinylvallne and its Role in Penicillin Biosynthesis. *Biochem. J.* **157**, 651–660 (1976).

69. Konomi, T. *et al.* Cell-free conversion of δ-(l-α-aminoadipyl)-l-cysteinyl-d-valine into an antibiotic with the properties of isopenicillin N in Cephalosporium acremonium. *Biochem. J.* **184**, 427–430 (1979).

70. O'Sullivan, J., Bleaney, R. C., Huddleston, J. A. & Abraham, E. P. Incorporation of 3H from δ-(L-α-amino (4,5-3H)adipyl)-L-cysteinyl-D-(4,4-$^{3}$H)valine into isopenicillin N. *Biochem. J.* **184**, 421–426 (1979).

71. Hollander, I. J., Shen, Y. Q., Heim, J., Demain, A. L. & Wolfe, S. A pure enzyme catalyzing penicillin biosynthesis. *Science.* **224**, 610–612 (1984).

72. Pang, C. P. *et al.* Purification of isopenicillin N synthetase. *Biochem. J.* **222**, 789–795 (1984).

73. Samson, S. M. *et al.* Isolation, sequence determination and expression in Escherichia coli of the isopenicillin N synthetase gene from Cephalosporium acremonium. *Nature* **318**, 191–194 (1985).

74. Carr, L. G., Skatrud, P. L., Scheetz, M. E., Queener, S. W. & Ingolia, T. D. Cloning and expression of the isopenicillin N synthetase gene from Penicillium chrysogenum. *Gene* **48**, 257–266 (1986).

75. Ramos, F. R., Lopez-Nieto, M. J. & Martin, J. F. Isopenicillin N synthetase of Penicillium chrysogenum, an enzyme that converts δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine to isopenicillin N. *Antimicrob. Agents Chemother.* **27**, 380–387 (1985).

76. Ramón, D., Carramolino, L., Patiño, C., Sánchez, F. & Peñalva, M. A. Cloning and characterization of the isopenicillin N synthetase gene mediating the formation of the β-lactam ring in Aspergillus nidulans. *Gene* **57**, 171–181 (1987).

77. Weigel, B. J. *et al.* Cloning and expression in Escherichia coli of isopenicillin N synthetase genes from Streptomyces lipmanii and Aspergillus nidulans. *J. Bacteriol.* **170**, 3817–3826 (1988).

78. Shiffman, D., Mevarech, M., Jensen, S. E., Cohen, G. & Aharonowitz, Y. Cloning and comparative sequence analysis of the gene coding for isopenicillin N synthase in Streptomyces. *MGG Mol. Gen. Genet.* **214**, 562–569 (1988).

79. Leskiw, B. K. *et al.* Cloning and nucleotide sequence determination of the isopenicillin N synthetase gene from Streptomyces clavuligerus. *Gene* **62**, 187–196 (1988).

80. Diez, B. *et al.* Two genes involved in Penicillium biosynthesis are linked in a 5.1 kb. *Mol. Gen. Genet. MGG* **218**, 572–576 (1989).

81. Baldwin, J. E., Blackburn, J. M., Schofield, C. J. & Sutherland, J. D. High level expression in Escherichia coli of a fungal gene under the control of strong promoters. *FEMS Microbiol. Lett.* **68**, 45–51 (1990).

82. Baldwin, J. E., Blackburn, J. M., Sutherland, J. D. & Wright, M. C. High-level soluble expression of isopenicillin N synthase isozymes in E. coli. *Tetrahedron* **47**, 5991–6002 (1991).

83. Patino, C., Sanchez, F. & Penalva, M. A. Low level expression in Escherichia coli of a fungal gene under the control of strong promoters. *FEMS Microbiol. Lett.* **58**, 139–143 (1989).

84. Baldwin, J. E. & Bradley, M. Isopenicillin N synthase: mechanistic studies. *Chem. Rev.* **90**, 1079–1088 (1990).

85. Baldwin, J. E., Adlington, R. M., Moroney, S. E., Field, L. D. & Ting, H. Stepwise Ring Closure in Penicillin Biosynthesis. Initial β-Lactam Formation. *J. Chem. Soc. Chem. Commun.* 984–986 (1984).

86. Baldwin, J. E. *et al.* Penicillin biosynthesis: active site mapping with aminoadipoylcysteinylvaline variants. *J. Chem. Soc., Chem. Commun.* **51**, 1225–1227 (1984).

87. Baldwin, J. E., Adlington, R. M., Robinson, N. G. & Ting, H.-H. Stereospecificity of β-lactam formation in penicillin biosynthesis. *J. Chem. Soc., Chem. Commun.* 409–411 (1986) doi:10.1039/C39860000409.

88. Baldwin, J. E., Blackburn, J. M., Sako, M. & Schofield, C. J. Formation of a Novel Bicyclic γ-Lactam with lsopenicillin N Synthase. *J. Chem. Soc. Chem. Commun.* **2**, 970–972 (1989).

89. Baldwin, J. E. *et al.* Penicillin biosynthesis. Dual pathways from a modified substrate. *J. Chem. Soc. Chem. Commun.* 1317–1319 (1983).

90. Baldwin, J. E. *et al.* Penicillin biosynthesis. On the stereochemistry of carbon-sulphur bond formation with modified substrates. *J. Chem. Soc. Chem. Commun.* **3**, 1319–1320 (1983).

91. Baldwin, J. E., Adlington, R. M., Domayne-Hayman, B. P., Knight, G. & Ting, H. H. Use of the cyclopropylcarbinyl test to detect a radical-like intermediate in penicillin biosynthesis. *J. Chem. Soc. Chem. Commun.* 1661–1663 (1987).

92. Chen, V. J. *et al.* Spectroscopic studies of isopenicillin N synthase. A mononuclear nonheme Fe2+ oxidase with metal coordination sites for small molecules and substrate. *J. Biol. Chem.* **264**, 21677–21681 (1989).

93. Scott, R. a *et al.* X-ray absorption spectroscopic studies of the high-spin iron(II) active site of isopenicillin N synthase: evidence for Fe-S interaction in the enzyme-substrate complex. *Biochemistry* **31**, 4596–4601 (1992).

94. Orville, a M. *et al.* Thiolate ligation of the active site $Fe^{2+}$ of isopenicillin N synthase derives from substrate rather than endogenous cysteine: spectroscopic studies of site-specific Cys --> Ser mutated enzymes. *Biochemistry* **31**, 4602–4612 (1992).

95. Kovacs, J. A. & Brines, L. M. Understanding how the thiolate sulfur contributes to the function of the non-heme iron enzyme superoxide reductase. *Acc. Chem. Res.* **40**, 501–509 (2007).

96. Brown, C. D., Neidig, M. L., Neibergall, M. B., Lipscomb, J. D. & Solomon, E. I. VTVH-MCD and DFT studies of thiolate bonding to $\{FeNO\}^7$/ $\{FeO_2\}^8$ complexes of isopenicillin N synthase: Substrate determination of oxidase versus oxygenase activity in nonheme Fe enzymes. *J. Am. Chem. Soc.* **129**, 7427–7438 (2007).

97. Roach, P. L. *et al.* Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* vol. 375 700–704 (1995).

98. Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–830 (1997).

99. Burzlaff, N. I. *et al.* The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401**, 721–724 (1999).

100. Fujishima, Y. *et al.* Crystallization and Preliminary X-ray Diffraction Studies on a Recombinant Isopenicillin N Synthase from Cephalosporium acremonium. *J. Mol. Biol.* **242**, 712–714 (1994).

101. Roach, P. L., Schofield, C. J., Baldwin, J. E., Clifton, I. J. & Hajdu, J. Crystallization and Preliminary-X-Ray Diffraction Studies on Recombinant Isopenicillin-N Synthase from Aspergillus-Nidulans. *Protein Sci.* **4**, 1007–1009 (1995).

102. Roach, P. L. *et al.* Anaerobic crystallisation of an isopenicillin N synthase.Fe(II).substrate complex demonstrated by X-ray studies. *Eur. J. Biochem.* **242**, 736–740 (1996).

103. McNeill, L. A. *et al.* Terminally Truncated Isopenicillin N Synthase Generates a Dithioester Product: Evidence for a Thioaldehyde Intermediate during Catalysis and a New Mode of Reaction for Non-Heme Iron Oxidases. *Chem. - A Eur. J.* **23**, 12815–12824 (2017).

104. Stewart, A. C., Clifton, I. J., Adlington, R. M., Baldwin, J. E. & Rutledge, P. J. A cyclobutanone analogue mimics penicillin in binding to isopenicillin N synthase. *ChemBioChem* **8**, 2003–2007 (2007).

105. Tamanaha, E. *et al.* Spectroscopic Evidence for the Two C–H-Cleaving Intermediates of Aspergillus nidulans Isopenicillin N Synthase. *J. Am. Chem. Soc.* **138**, 8862–8874 (2016).

106. Hall, B. G. & Barlow, M. Evolution of the serine β-lactamases: Past, present and future. *Drug Resist. Updat.* **7**, 111–123 (2004).

107. Hall, B. G. & Barlow, M. Structure-based phylogenies of the serine β-lactamases. *J. Mol. Evol.* **57**, 255–260 (2003).

108. Abraham, E. P. & Chain, E. An Enzyme from Bacteria able to Destroy Penicillin. *Nature* **146**, 837–837 (1940).

109. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).

110. Laxminarayan, R. *et al.* Review The Lancet Infectious Diseases Commission on antimicrobial resistance : 6 years later. *Lancet Infect. Dis.* **3099**, 1–10 (2020).

111. Mulligan, M. E. *et al.* Methicillin-resistant Staphylococcus aureus: A consensus review of the microbiology, pathogenesis, and epidemiology with implications for prevention and management. *Am. J. Med.* **94**, 313–328 (1993).

112. Utsui, Y. & Yokota, T. Role of an altered penicillin-binding protein in methicillin- and cephem-resistant Staphylococcus aureus. *Antimicrob. Agents Chemother.* **28**, 397–403 (1985).

113. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta - Proteins Proteomics* **1794**, 808–816 (2009).

114. Quale, J., Bratu, S., Gupta, J. & Landman, D. Interplay of efflux system, ampC, and oprD expression in carbapenem resistance of Pseudomonas aeruginosa clinical isolates. *Antimicrob. Agents Chemother.* **50**, 1633–1641 (2006).

115. Lee, J. Y. & Ko, K. S. OprD mutations and inactivation, expression of efflux pumps and AmpC, and metallo-β-lactamases in carbapenem-resistant Pseudomonas aeruginosa isolates from South Korea. *Int. J. Antimicrob. Agents* **40**, 168–172 (2012).

116. Hancock, R. E. W. The bacterial outer membrane as a drug barrier. *Trends Microbiol.* **5**, 37–42 (1997).

117. Nikaido, H. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiol. Mol. Biol. Rev.* **67**, 29–47 (1996).

118. Babic, M., Hujer, A. M. & Bonomo, R. A. What's new in antibiotic resistance? Focus on beta-lactamases. *Drug Resist. Updat.* **9**, 142–156 (2006).

119. Bush, K. Past and Present Perspectives on β-Lactamases. *Antimicrob. Agents Chemother.* **62**, 1–20 (2018).

120. Sabath, L. D. & Abraham, E. P. Zinc as a cofactor for cephalosporinase from Bacillus cereus 569. *Biochem. J.* **98**, 10–12 (1966).

121. Ambler, R. P. The structure of β-lactamases. *Philos. Trans. R. Soc. London. B, Biol. Sci.* **289**, 321–331 (1980).

122. Medeiros, A. a. Evolution and Dissemination of β-Lactamases Accelerated by Generations of β-Lactam Antibiotics. *Clin. Infect. Dis.* **24**, S19–S45 (1997).

123. Evans, B. A. & Amyes, S. G. B. OXA β-lactamases. *Clin. Microbiol. Rev.* **27**, 241–263 (2014).

124. Walther-Rasmussen, J. & Høiby, N. OXA-type carbapenemases. *J. Antimicrob. Chemother.* **57**, 373–383 (2006).

125. Bush, K., Jacoby, G. A. & Medeiros, A. A. A functional classification scheme for β-lactamases and its correlation with molecular structure. *Antimicrob. Agents Chemother.* **39**, 1211–1233 (1995).

126. Stojanoski, V. *et al.* Structural basis for different substrate profiles of two closely related class D β-lactamases and their inhibition by halogens. *Biochemistry* **54**, 3370–3380 (2015).

127. Couture, F., Lachapelle, J. & Levesque, R. C. Phylogeny of LCR-1 and OXA-5 with class A and class D beta-lactamases. *Mol. Microbiol.* **6**, 1693–1705 (1992).

128. Baurin, S. *et al.* Critical role of tryptophan 154 for the activity and stability of class D β-lactamases. *Biochemistry* **48**, 11252–11263 (2009).

129. Golemi, D., Maveyraud, L., Vakulenko, S., Samama, J. P. & Mobashery, S. Critical involvement of a carbamylated lysine in catalytic function of class D β-lactamases. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14280–14285 (2001).

130. Lamotte-Brasseur, J. *et al.* The structures and catalytic mechanisms of active-site serine β-lactamases. *Biotechnol. Genet. Eng. Rev.* **12**, 189–230 (1994).

131. Lu, W. P., Kincaid, E., Sun, Y. & Bauer, M. D. Kinetics of β-lactam interactions with penicillin-susceptible and -resistant penicillin-binding protein 2x proteins from Streptococcus pneumoniae. Involvement of acylation and deacylation β-lactam resistance. *J. Biol. Chem.* **276**, 31494–31501 (2001).

132. Hata, M. *et al.* Substrate deacylation mechanisms of serine-β-lactamases. *Biol. Pharm. Bull.* **29**, 2151–2159 (2006).

133. Lietz, E. J., Truher, H., Kahn, D., Hokenson, M. J. & Fink, A. L. Lysine-73 is involved in the acylation and deacylation of β-lactamase. *Biochemistry* **39**, 4971–4981 (2000).

134. Goldberg, S. D., Iannuccilli, W., Nguyen, T., Ju, J. & Cornish, V. W. Identification of residues critical for catalysis in a class C β-lactamase by combinatorial scanning mutagenesis. *Protein Sci.* **12**, 1633–1645 (2003).

135. Jeon, J. H. *et al.* Structural basis for carbapenem-hydrolyzing mechanisms of carbapenemases conferring antibiotic resistance. *Int. J. Mol. Sci.* **16**, 9654–9692 (2015).

136. Lohans, C. T. *et al.* Mechanistic Insights into β-Lactamase-Catalysed Carbapenem Degradation Through Product Characterisation. *Sci. Rep.* **9**, 1–9 (2019).

137. Lohans, C. T. *et al.* A New Mechanism for β-Lactamases: Class D Enzymes Degrade 1β-Methyl Carbapenems through Lactone Formation. *Angew. Chemie - Int. Ed.* **57**, 1282–1285 (2018).

# Chapter 2

## Serial crystallography and X-ray free electron lasers

## 2.1 Developments in synchrotron based crystallography

Currently, several different techniques are used to study the 3D structure of macromolecules, such as DNA and proteins. These techniques include NMR spectroscopy,[1,2] mass spectrometry (MS),[3] circular dichroism (CD),[4] cryo electron microscopy (cryo-EM),[5] small angle X-ray scattering,[6,7] and X-ray crystallography[8,9] Currently, the most widely used technique to obtain structural information on macromolecules, as evidenced by the annual number of deposited structures to the protein data bank (PDB), is X-ray crystallography (9662 structures, 89% of the total number of deposited structures in 2019).[10–12] The atomic resolution obtained from crystals remains unparalleled. X-ray crystallography experiments are performed using crystalline materials, e. g. proteins, and electromagnetic radiation. Laboratory X-ray sources can be used to obtain structural information. However, in recent years laboratory X-ray sources have been more frequently used for screening conditions and obtaining preliminary data. Currently, the predominant X-ray sources for experiments are synchrotrons, accounting for 85% of the structures deposited annually to the PDB.[13]

Development of synchrotrons and betatrons in the 1940s led to the observation of synchrotron radiation.[14] However, most of the research application using synchrotrons were in fields of high-energy and particle physics. In the late 1960s, this changed with the development of electron storage rings, the basis for modern synchrotrons, which allow a continuously circulating current at a fixed energy level for extended periods of time (up to several hours), providing X-ray beams at rates up to 50 Hz.[15] These synchrotrons provided parasitic synchrotron radiation, and are considered the first generation synchrotron light sources. Increased interest in the application of synchrotron radiation, led to the construction of dedicated synchrotron radiation sources, referred to as second generation light sources. Later, the introduction of insertion devices further improved the brightness of the X-ray sources.[15] These third generation synchrotrons are currently commonly used light sources, providing X-ray beams with an averaged brilliance up to $10^{21}$ ph/s/mrad$^2$/mm$^2$/0.1%BW.[16] The shortest operating pulse duration is on the order of $10^1$ – $10^2$ ps, restricted to the length of the electron bunch in the storage rings.[16,17] The X-ray beam can be provided at a rate from a few Hz to 100s of MHz frequencies using pulse picking schemes.[16,18–21] Large area detectors used in typical macromolecular crystallography experiments are able to operate between 100s Hz to 2 kHz

frequencies.[22–24] The introduction of X-ray free electron lasers (XFELs) capable of operating at MHz frequencies,[25] lead to the development of the AGIPD detector, capable of running at 4.5 MHz[26].[27] Currently, a next generation of synchrotron light sources is under development, which attempts to gain brilliance by reducing emittance (spatial size and the angular spread) of the electron beam, using double triple bend achromat lattice architecture.[16,28–30]

Alongside improvements of the radiation sources, introduction of robotics for sample handling,[31–33] improved workflow,[34] automated data processing,[35–37] developments in detector technology,[22,38,39] and remote control operation of the beamline,[34,40] have resulted in an efficient strategy for data collection which has increased the throughput of samples.[34,41–43] A prime example of the extent of automation is the standardised sample holder (SPINE) that has been introduced and widely used to enhance robotic sample handling in accuracy, reliability, and ultimately sample throughput.[31] The magnetic crystal mount can be fitted with different types of pins, and contains a digital ID number for tracking. Nowadays, many of the storage and transport equipment are tailored to the use of the standard SPINE sample holders. In addition, many aspects of the crystallisation of macromolecules have undergone streamlining and automation with commercially available initial and optimisation screens,[44–46] the introduction of efficient liquid-handling robots,[47,48] and automated imaging and harvesting systems that are able to inspect and fish crystal from trays.[48–50] Thus, the overall process for macromolecular crystallography is streamlined and optimised, greatly reducing the number of researchers required per experiment and the time required per experiment.

## 2.2 Time-resolved crystallography

### 2.2.1 Observing transitions in macromolecular crystallography

Traditionally, X-ray crystallography has been used to capture space and time averaged static images of macromolecules.[51,52] Myoglobin, haemoglobin, and lysozyme mark the first biological macromolecules to be structurally characterised by X-ray crystallography providing still images.[53–56] Static images of proteins or enzyme inhibitor/substrate complexes provide evidence for interactions of specific residues with small molecules and enhance the understanding of the involvement of active site residues in mechanistic operations.[10,57,58] Diffusion of small molecule inhibitors or substrates into crystals, and indeed to a certain

extent, structural rearrangements inside protein crystals (reaching equilibrium) are possible within crystals, due to the large solvent content (27 – 78%).[59,60] This means that, to a certain extent, a protein crystal resembles a highly concentrated solution.[51,59] Biological processes in these crystals frequently proceed *via* approximately the same mechanism, and often times with retention of the crystal lattice.[51,61] However, the information obtained from proteins give a time averaged static image and are limited in detail about atomic rearrangements. Thus, from a time-resolved standpoint, these features appear to be at odds with the notion of capturing transient species using crystallography.[51,62,63] However, many different processes occur during biochemical reactions involving enzymes and proteins. The time scale at which these processes such as protein refolding, photocycling, and chemical bond formation, occur, span a large range of time domains; from femtoseconds (fs, $10^{-15}$ s)[64,65] all the way to many seconds or even days[66]. Thus, by choosing the appropriate time scale, it is possible to observe structural changes using short X-ray exposures.

### 2.2.2 Reaction initiation in crystals

In order to observe transient species, a reaction has to be initiated synchronously throughout the whole crystal, which can be challenging. Different strategies have been employed in order to trigger reactions in crystals such as light illumination (VIS – X-rays), and diffusion of small molecules.[52,63,67] Light mediated triggering can be applied to protein systems that are inherently sensitive to light activation.[68–72] Additionally, systems can be modified or designed to become light sensitive through the use of caged substrates[73–76] or introduction of photo dissociable groups on the proteins[77]. Diffusion of small molecules is particularly suited for enzymatic reactions, allowing inhibitors, substrates, or products to diffuse into the active site.[78–80] The diffusion approach is relatively slow and depends on the crystal morphology and diffusion properties of the small molecule.[81–83] Typically, soaking of protein crystals with small molecules is performed at or close to ambient temperature.[67] Note, that the rate of diffusion of the small molecules has to outrun the catalysis rate in order to achieve homogenous initiation of the reaction throughout the crystal.[84] In order to out speed the rate of catalysis for a fast enzyme, crystal dimensions of <10 μm are required.[84–86]

### 2.2.3 Laue method

If the lifetime of the intermediate of interest is sufficiently long, data can be collected through a routine monochromatic X-ray experiment. However, in order to observe transient species with lifetimes shorter than the data collection time, a different approach has to be used. An example of an approach to probe sub-microsecond transitions, is Laue diffraction.[87–89] In contrast to typical monochromatic X-ray experiments, the Laue method relies on employing a polychromatic light source and a stationary crystal obtaining a full data set much faster.[87] Often times, these experiments rely on laser illumination (pump) to initiate the reaction, before exposure to X-ray radiation (probe), and are performed at room temperature.[77,82,90] In principle, these pump-probe experiments can achieve a time resolution on the order of $10^1$ – $10^2$ ps.[91] However, studies using the Laue approach suffer from experimental and analytical challenges.[87,88,92] Due to the nature of the collected data, deriving from a single shot, of a single crystal, it is sensitive to imperfections in the crystal lattice. Overlapping spots are inherent to the Laue method,[87] but can be mitigated by the use of "pink beam" Laue.[88,92] Additionally, uneven sampling through space leads to fewer reflections typically at lower resolution.[87]

### 2.2.4 Cryo-crystallography and modified systems

Alternative, strategies have been developed to extend the lifetimes of intermediates by chemical or physical trapping.[89,93] Note that these strategies generally rely on perturbing the system under investigation, facilitating observation of intermediates, and are thus are considered not real-time-resolved or pseudo time-resolved.[63,94] Flash-cooling of crystal samples is typically employed to reduce the effects of radiation damage, but has also been employed to kinetically trap intermediates.[95,96] The intermediate is allowed to build up under (near) ambient conditions, and is subsequently cryo-cooled (trap freezing).[51,67] Alternatively, the crystal is first cryo-cooled, before a structural rearrangement is initiated (freeze trapping).[94] Both trap freezing and freeze trapping are distinctly different and are employed depending on the nature and extent of the structural change(s) under investigation.[51] Both strategies rely on a reduction in rate constants, preventing the progression of the reaction at a particular (meta) stable state.[51,67,89] Additionally, chemical modification of the protein system, e. g. by mutagenesis of the protein,[97] or the use of a substrate analogue,[98]

can facilitate the stabilisation of an intermediate.[51,63] Crystallisation under non-productive conditions can also aid structural characterisation of transient species.[51,99] However, these techniques perturb the system under investigation, and thus are not considered time-resolved.[63]

### 2.2.5 Serial synchrotron crystallography

More recently, successes with XFELs (Section 2.3) led to development in the field of serial crystallography.[100,101] The approach uses microcrystal slurries and a vastly different sample delivery method (Section 2.5) to collect diffraction data at ambient temperature. The concept of serial crystallography experiments was subsequently applied at synchrotrons; serial synchrotron crystallography (SSX).[102] SSX approaches for time-resolved crystallography experiments have certain advantages over single crystal approaches; radiation damage mitigation by the use of multiple crystals,[103,104] and consequently the use of smaller crystals improving homogenous reaction initiation and synchronisation[86]. The SSX approach provides a different strategy for data acquisition, allowing for mixing and pump-probe experiments, complimenting pre-existing strategies for time-resolved crystallography.[102] Although microcrystal experiments have been fielded before, using synchrotron based microfocus beamlines,[105–107] these beamlines mostly focused on structural determination of samples previously inaccessible due to their small crystal size.[108] Compared to synchrotron based microfocus beamlines, XFELs provide radiation damage free data using a more focused beam, with higher brilliance.

## 2.3 X-ray Free Electron Laser radiation

### 2.3.1 Free Electron Laser physics

Successes in physics in the 1960s resulted in development of laser (Light Amplified Stimulated Emission of Radiation) technology for infrared (IR) and visible light.[109,110] Extending this technology to X-ray wavelengths was challenging and took decades of innovation. An important step towards this end was the concept of obtaining coherent monochromatic radiation from electrons moving through a periodic magnetic array (undulator), as proposed by Motz in 1951.[111] Decades later in the 1970s, this concept was applied and led to the theoretical conceptualisation and realisation of free-electron lasers (FELs) operating

in the IR regime.[112,113] The undulator technology used in FEL lasers, was later on incorporated into third generation synchrotrons as a method to increase brilliance of the X-ray beam.[114,115] Technological improvements of FELs in the 1980s and 1990s led to the development of single-pass linear accelerators capable of generating short pulsed photon beams with wavelengths in the soft X-ray regime.[116–119] A few years later, this was followed by the operation of FELs in the hard X-ray regime, XFELs,[120,121] resulting in a gain in popularity of FELs from a broader scientific community.[116] The linear design of the XFELs, although able to source multiple end-stations, limits the number of parallel experiments that can be conducted at any point in time compared to synchrotrons.[122,123] The high quality relativistic electron beam that is required for the lasing process necessitates a linear accelerator design for XFELs.[124] These XFELs provide an X-ray source with a peak brightness 9 – 10 orders higher than from synchrotron light sources (Figure 2.1), a property attributed to the extended undulator arrays used to generate X-rays.[125,126] The relativistic electron beam oscillates within the undulator magnetic field generating radiation with a wavelength, given by[117,127]:

$$\lambda = \frac{\lambda_U(1+K^2)}{2\gamma^2} \quad (2.1),$$

where $\lambda$ is the wavelength of the emitted radiation, $\gamma$ the electron energy, $\lambda_U$ is the undulator period, and $K$ the undulator strength parameter. The latter is given by[117]:

$$K = \frac{eB_U\lambda_U}{2\pi mc^2} \quad (2.2),$$

in which $e$ is the electron charge, $m$ the rest mass of an electron, $c$ the speed of light, $B_U$ the magnetic field. For undulators with $K << 1$, the applied oscillating magnetic field experienced by the relativistic electron beam results in an increase in transverse and decrease in forward speed with increase in undulator magnetic field. Thus, changes in the magnetic field applied, alter the speed of the electron beam and enable tuning of the emitted radiation.[127] For undulators with $K >> 1$, also referred to as wigglers, the applied magnetic field results in emittance of partial coherent radiation at a wider opening angle.[16] Undulators at XFEL radiation sources provide high peak brilliance (Figure 2.1), and pulses of short duration (fs time range), with high transverse and moderate longitudinal coherence.[125] At synchrotron radiation sources undulators provide high average brightness, and transverse coherence. The difference is due to the length of the

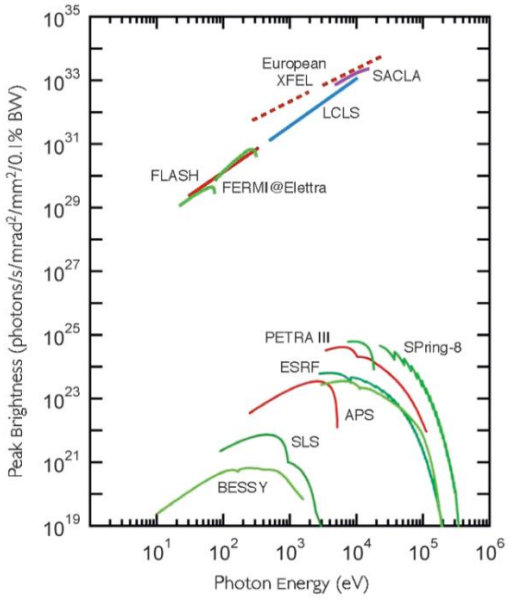



**Figure 2.1:** Comparison of peak brightness and photon energies available at different X-ray sources, including synchrotrons and XFELs. Reprinted with permission from MRS bulletin (Cambridge University Press)[125].

undulator; the extended undulator length used for XFELs gives rise to self-amplified spontaneous emission (SASE),[128,129] a process in which the electron bunch interacts with previously emitted photons as these propagate together through the oscillating magnetic field. The interaction between the electrons and photons leads to electrons gaining and losing energy depending on their relative phase, resulting into the formation of microbunches (Figure 2.2).[124,130] A sufficiently straight undulator trajectory is required for electron bunches and photons to remain overlapped during the SASE process.[128,129] As the electrons reach a close proximity to one another, charge effects prevent further bunching, leading to saturation of the SASE process. Importantly, the SASE process can occur at any wavelength by varying the energy of the electron beam.[124] Due to the random fluctuations of the initial electron beam, the electrons organise in different, uncorrelated microbunches.[117,126] Consequently, an energy spread is observed within the complete bunch, affecting longitudinal coherence of the electron beam. Additionally, changes in the initial electron beam give rise to fluctuations in the emitted radiation; average wavelength, photon energy, pulse energy.[131] To address the intrinsic noise of the SASE process, and improve the stability and photon energy of the pulse, the longitudinal coherence, and provide control over the pulse structure, seeding schemes have been developed.[132–134]

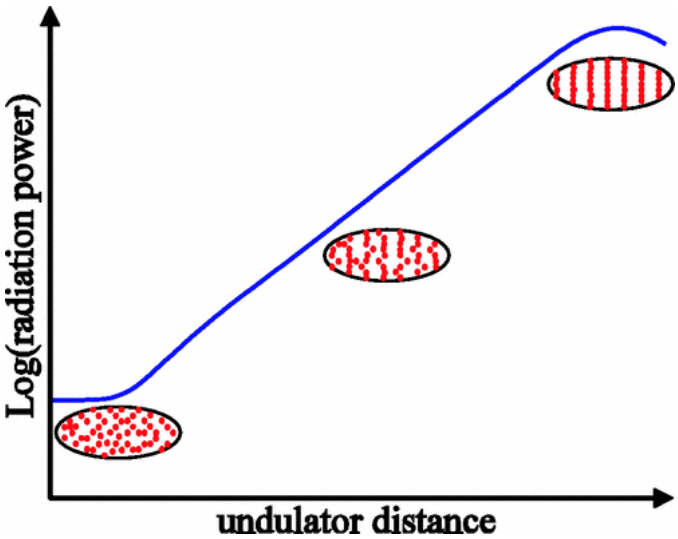



**Figure 2.2:** The self organising process of electrons into microbunches and the growth of radiation power as a function of undulator distance. Reprinted with permission from Physical reiew special topics – accelerators and beams (APS Physics).[126]

Certain techniques are sensitive to changes in beam properties (e. g. changes in photon energy (absorption spectroscopy[135]), or coherence (coherent diffraction imaging, CDI, Section **2.3.3**[136]) and thus proper diagnostics of the X-ray source are important[137]. The SASE process is able to produce short pulses, ranging between 100 fs and 100 attosecond ($10^{-18}$ s).[138] Control over pulse duration can be achieved by the changing the initial electron bunch length. In comparison, the shortest pulses achievable using synchrotron sources are on the order of picoseconds (ps, $10^{-12}$).[139] A typical pulse contains $10^{12}$ to $10^{13}$ photons with a repetition rate of 60 – 120 Hz using a normal cavity FEL design, resulting in an overall flux comparable to what is achieved at synchrotrons per second.[116,120,121,140,141] An increased repetition rate, up to MHz, can be achieved when using superconducting cavity FEL design.[25,116] Depending on tuning of the source, an energy density between $10^{16}$ and $10^{21}$ W cm$^{-2}$ can be reached.[25,120,141]

### 2.3.2 Radiation damage

The energy deposition into the sample is orders of magnitudes higher compared to synchrotrons. As mentioned above, the advances in synchrotron radiation resulted in light sources that were capable of damaging the crystals, necessitating the use of cryo-cooling to reduce radiation damage.[95,96,142,143] Thus, it is unsurprising that the combination of high flux and high energy photons of the XFEL source, result in severe radiation damage to the sample.[144] The radiation damage caused by X-ray radiation can be divided

into global and local effects. Global effects include processes such as thermal motion and photoionisation, which affect all atoms in a statistically predicable fashion. This rationalisation is due the relative abundance and similar X-ray cross section of nitrogen, oxygen, and carbon atoms inside macromolecular crystals. Local damage is less well understood, as it is element specific and dependent on the local environment. Heavy elements, such as sulfur and transition metal ions, are particularly susceptible to local damage due to their significantly increased X-ray cross section.

Much of the radiation damage is caused by photoabsorption and inelastic scattering of X-rays.[144,145] Absorption of photon energy by atoms leads to the release of inner core electrons (photoelectrons), resulting in ionisation.[146] The excited state of the ion can decay *via* Auger decay or X-ray fluorescence.[144] In the Auger decay process, the core hole is filled by an outer shell electron, concomitant with the removal of a second electron. The second electron carries away an energy proportional to the energy difference between the core and outer shell level involved in the Auger process. The Auger process can take between 0.1 – 10 fs depending on the size of the element, and is generally faster for heavier atoms.[147] Thus, heavy atoms (e. g. sulfur, and transition metals) often times undergo ionisation steps faster than neighbouring atoms and form "hot-spots" for ionisation.[148] For the X-ray fluorescence process, the core hole is filled by an outer shell electron, concomitant with the release of an X-ray fluorescent photon. The energy of the fluorescent photon is proportional to the difference between the core and outer shell level involved in the fluorescent process. Depending on the type of atom, either process occurs; lighter atoms favour Auger decay, whereas heavier atoms favour X-ray fluorescence.[144] The extremely high intensity of the XFEL pulse leads to multiple photoabsorption processes, resulting in a highly charged state.[149,150] The final charge of the ion depends on the size of the atom.

After emission, the photoelectron or Auger electron collide with neighbouring atoms in creating a cascade, leading to the release of secondary electrons; secondary ionisation.[151] Depending on the energy of the electron, thermalisation can take up to 50 fs to complete. Ionisation of the sample leads to bond breakage and creation of a plasma-like state. Subsequently, formation and diffusion of radicals (e. g. sulphide or hydroxide ions) leads to chemical modifications of the protein structure, visible in the obtained electron density maps.[96,152] Simulations demonstrated that ions can move up to 0.5 nm in less than 100 fs. Eventually,

the temperature within the sample rises to 500,000 K, and the highly charged atoms start to repel, leading to a Coulomb explosion and destruction of the sample itself.[153] Note, that in conventional crystallography photoionisation is sparse, and that mobility of ions and by extension secondary radiation effects can be mitigated by cryo-cooling of the sample.[154,155]

Modelling studies have suggested a maximum XFEL pulse duration of 10 fs in order to outrun the effects of radiation damage.[153] However, experimental evidence shows strong diffraction patterns of crystals probed with pulses of up to 300 fs.[156] This unusual tolerance to radiation is rationalised by the loss of crystalline order before termination of the pulse.[146,157] During irradiation, atomic disorder within the crystal lattice increases due to atomic and radical mobility and Coulomb repulsion.[144] The loss of periodicity over time consequently reduces Bragg diffraction, ultimately "turning off" completely before the end of the pulse. Thus, the collected diffraction data is self-gating to provide an image of the macromolecule at the start of the pulse.[146,157] The pulse length appears to be shorter than the incident pulse duration. However, heavy atoms, such as sulfur and transition metals, will ionise at a higher rate than the global average ionisation.[148] Experimental studies have demonstrated that radiation damage to Fe-S clusters and disulfide bonds can be observed for moderate and relatively long XFELs pulses.[158–160] Despite these concerns, multiple studies have reported on damage free structures, by comparing bond lengths obtained using electron density maps to different spectroscopic experiments.[161–163] In principle, this "diffraction before destruction"[164] feature eliminates radiation damage,[153,165] and enables room temperature data collection[166].

### 2.3.3 Application of coherent properties of FEL radiation

The highly coherent properties of XFEL radiation has attracted much attention and led to the development of and application in experimental techniques, including CDI,[136,167] X-ray Cross Correlation Analysis (XCCA),[168] and X-ray photon correlation spectroscopy (XPCS)[169]. The partial coherence of synchrotron radiation leads to experimental limitations in these approaches[170], and thus is less suited for these techniques. For example, CDI uses the coherent properties of XFELs in diffraction experiments to enable imaging of 2D and 3D non-crystalline specimens, allowing for diffraction limited imaging.[167] Within a diffraction experiment, typically, three aspects are required to obtain a structure from the Fourier transform

of the reflection data; wavelength, phase, and amplitudes. Wavelength is predetermined by the experimental setup, and the amplitudes can be derived from the intensity spots obtained during the experiment. Phase information is typically lost in the process and gives rise to the "phase problem".[171] The idea of the CDI approach to imaging of non-crystalline specimens is often associated with the observation that Bragg diffraction leads to undersampling of the diffracted intensities relative to Shannon's theorem.[167,172] In the 1980s, with improved computing technologies, iterative feedback algorithms were developed that allow phase information to be extracted from adequately sampled intensity data.[173–176] Thus, in principle, both amplitudes and phase information could be obtained from diffracted intensities of non-crystalline specimens when sampled above the Nyquist frequency.[174,177] Importantly, all image recovery algorithms relate the image of the object to the scattered image on the detector via Fourier transformation, which is only valid if the optical field is fully coherent.[178] Predictive calculations suggested that the coherent patch of light should be twice the size of the object,[179] however more recent modelling and experimental work demonstrated a more stringent dependency on coherence.[178,180] Additionally, by avoiding the use of lenses, the final image obtained using CDI is aberration free, and thus, in theory, resolution is limited to diffraction and dose.[167]

## 2.4 The use of XFELs in protein crystallography

With the introduction of the first XFEL facilities operating in the hard X-ray regime, new opportunities for macromolecular crystallography were created.[181] Due to the short pulses available at XFEL facilities, outrunning most of the radiation damage, avoiding the need for cryo-cooling, and enabling data collection under physiologically relevant temperatures and atmospheric pressure.[153,165] The highly focused beam allowed for data acquisition on submicron sized crystals, enabling fast mixing and diffusion experiments.[86] Additionally, the fs pulse pushed the shortest possible probing time into the fs regime.

Photoactive yellow protein (PYP) is a member of the Per-Arnt-Sim domain superfamily, and is involved in signal transduction.[182,183] PYP is a cytoplasmic protein that binds the chromophore *p*-coumaric acid. The chromophore is linked via Cys69 and undergoes a reversible *trans* to *cis* isomerisation around its thiol-ester bond when excited with blue light.[184] The structure of PYP has been studied extensively, and prior studies

have employed the Laue method, which has proven to be a powerful method.[185,186] However, the Laue method requires sufficiently large crystals and of high quality; in contrast the serial crystallography approach used at XFELs allows for averaging and exclusion of outliers. Synchronised initiation of the reaction can be challenging in larger crystals. Additionally, the time resolution was limited to 100 ps. With the shorter fs pulses available at XFELs, time resolved serial femtosecond crystallography (tr-SFX) pump-probe studies on PYP demonstrated how, after illumination with blue light, a *trans* to *cis* isomerisation occurs inside the protein.[187,188] Time stamps between 140 fs and 3 ps were recorded demonstrating in great detail the photo induced structural change in the chromophore. A microcrystal slurry sample was provided using a liquid jet sample delivery system (Section **2.5.2**). The results were compared to reported studies employing the Laue method, and demonstrated comparable quality data.[185–187]

Photosystem II (PS II) is a protein complex that catalyses the formation of $O_2$ from $H_2O$ by absorption of visible light.[189] The PS II system consists of 19 subunits with ~60 different cofactors, and is bound to the thylakoid membranes in plants, algae, and cyanobacteria.[189,190] Light is absorbed by the $P_{680}$ reaction centre of PS II and leads to charge separation. During this process a radical cation $P_{680}^{\bullet+}$, and quinone radical anion $Q_B^{\bullet-}$ are formed. The oxygen-evolving complex (OEC), containing a $Mn_4CaO_5$ cluster, is oxidised by the $P_{680}^{\bullet+}$ radical cations via Tyr161 ($Y_Z$). After four consecutive oxidations ($S_{0\text{-}3}$), $O_2$ is produced in the final step ($S_4$) of the five-step cycle ($S_{0\text{-}4}$).[191] The $S_{0\text{-}3}$ intermediates are relatively stable, with a lifespan on the order of seconds. The highly reactive $S_4$ state is transient and not yet conclusively characterised.

The first structural studies on PS II identified the $Mn_4CaO_5$ cluster as the catalytic site involved in the formation of the O-O bond.[191,192] The resolution of structures obtained in the initial studies (3.8 – 2.9 Å) precluded unambiguous assignment of the individual atoms.[191–193] In a 1.9 Å resolution structure of PS II,[194] the electron densities of the different atoms were sufficiently separated to allow assignment of the four Mn(III/IV) and single Ca(II) metal ions that comprised the distorted cubane structure. A combination of synchrotron based structural studies and various spectroscopic studies were used to characterise the structure and provide evidence for the mechanism of the OEC in PS II.[192,193,195–198] However, there were discrepancies between the models obtained by spectroscopic and X-ray crystallographic studies in the $Mn_4Ca$ cluster. Bond distances between the Mn and O atoms within the OEC complex were elongated in

crystallographic analyses compared to distances determined by X-ray absorption near-edge structure (XANES) and Extended X-ray absorption fine structure (EXAFS) studies. These differences were largely attributed to radiation damage occurring during data collection.[198] Radiation damage was recognised as an important limitation to obtaining high resolution structures,[96] and has become an increasing concern with the introduction of more powerful synchrotron radiation sources. Localised damage around the $Mn_4Ca$ cluster occurs before the onset of global radiation damage (Section **2.3.2**).[96,199] Dose characteristics for PS II have been demonstrated with XANES and EXAFS studies.[200] Attempts to mitigate radiation damage by cryo-cooling samples[155] failed to suppress radiation damage at the $Mn_4Ca$ cluster in OEC.[198]

With the introduction of XFELs, new efforts towards studying PS II intermediates have been undertaken.[162,201–203] Stable intermediates ($S_{0-3}$) of the catalytic cycle have been characterised, notably demonstrating bond distances in agreement with spectroscopic studies, suggesting no observable radiation damage.[162] Data collection was performed at room temperature, and the different excited states were obtained by photo excitation of microcrystals. Additionally the diffraction data demonstrated a structural rearrangement of the $Mn_4Ca$ cluster and insertion of an additional oxygen atom.[201,202] Although this event had been observed in a reported study, and computationally predicted, the chemical nature of the $H_2O$ derived oxygen atom was different than reported. In certain studies, diffraction data on PS II was supported by simultaneously collected X-ray emission spectroscopy (XES) data.[201,204,205]

G-protein-coupled receptors (GPCRs) are a large family of membrane proteins involved in signal transduction pathways in humans,[206] and constitute an important class of drug targets.[207] Rhodopsin is a light sensitive GPCR that, together with its ligand retinal, enables vision in low light situations.[208] Structural determination of GPCRs, including rhodopsin, has been particularly challenging due to their low expression levels, unstable nature in solution, and inability to crystallise.[209,210] In order to crystallise membrane proteins, the extended trans membrane hydrophobic regions have to be shielded from solvent.[211] Shielding can be achieved by the use of detergents, which promotes micelle formation, and helps solubilise the protein and was proven beneficial for crystallisation.[212,213] More recently, membrane mimicking mesophase substances, such as lipidic cubic phase (LCP), have been introduced to help facilitate crystallisation of membrane proteins.[211,214] Although these techniques have been proven to be useful, crystallisation of membrane

proteins remains challenging as many of the GPCRs produce micron-sized crystals, which are too small to study using synchrotron radiation. The introduction of XFELs has provided a method to obtain high resolution structural data, using a highly focused beam, while circumventing radiation damage.[209,210]

Observing catalytic turnover within enzyme crystals can be challenging. As mentioned above (Section **2.2.2**), initiation of an enzymatic reaction can be achieved by diffusion of a substrate/ligand into the active site. Indeed, a protein crystal largely consists of solvent,[59,215] allowing small molecules to diffuse through the lattice.[84,86] An early example is the diffusion of heavy metals into protein crystals, in order to solve the phase problem.[216] However, successful initiation of an enzymatic reaction *in crystallo* is dependent on the diffusion rate and the catalytic turnover rate; diffusion has to be sufficiently fast to outpace catalytic turnover to ensure the reaction is initiated throughout the crystal lattice.[84,86] Computational and experimental studies suggest that to outrun catalytic turnover for faster enzymes, a crystal size of <10 μm is required.[84–86] In a study on *M. tuberculosis* β-lactamase (BlaC), a class C serine β-lactamase (Section **1.4**), a mix-and-inject approach using liquid jets (Section **2.5.2**) was employed.[139,217,218] A substrate, ceftriaxone, was mixed with BlaC microcrystals (5 × 2 × 2 μm) and then probed for diffraction after different time delays (30, 100, 500, 2000 ms).[217] After 30 ms full occupancy for the ceftriaxone substrate was observed at the active site. High occupancy of the covalent acyl-enzyme complex was observed after 500 ms. The mix-and-inject experiment demonstrates that for the smaller BlaC crystals, diffusion of the substrate was not rate-limiting.

## 2.5 Sample and sample delivery for XFEL studies

### 2.5.1 Sample characteristics

Crystals probed by the XFEL beam, are destroyed by the measurement, which requires a serial crystallography approach for data acquisition. The "exact" number of crystals required per dataset depends on the size of the unit cell, the diffraction quality, and crystal symmetry. Nevertheless, a vastly different approach for sample preparation is required for XFEL experiments[219] than for traditional crystallography, where a single large (>100 μm) crystal is used for data collection. In order to collect data in an efficient manner, samples are supplied continuously, which is different from the well-established conventional single

crystal approach employed in most diffraction experiments currently (Section 2.1). The sample requirements have resulted in a shift towards microcrystal slurries as a preferred sample form.[156,220–222] Although microcrystal slurries can be obtained from initial screening conditions, optimised conditions for macrocrystals provide a starting point for the development of microcrystal slurries. In certain cases, it is possible to obtain microcrystals from macrocrystals by physical crushing of the crystals.[220] However, this could lead to a degeneracy of the sample quality.[220] Optimisation of the crystallisation conditions, e. g. precipitant to enzyme ratio, can promote the formation of smaller crystals. By seeding these optimised conditions with the crushed macrocrystals, the number of nucleation sites is increased, resulting in the formation of a homogenous microcrystal slurry. Often times, attempts are made to prepare the microcrystal slurries using batch methods in order to obtain sufficient quantities. Multiple rounds of seeding using the obtained microcrystal slurry, can in some cases improve the quality of the microcrystal slurry.[220] Determining the quality of the microcrystal slurry can be done in different ways. Optical microscopy can be used to inspect the crystal size and density, however depending on the size of the crystals this approach can be challenging.[223] Alternatively transmission electron microscopy, intrinsic UV fluorescence, or dynamic light scattering could be used to investigate average size and density of the microcrystal slurry.[222,223] Depending on the size of the crystals, their diffraction quality can be assessed in serials crystallography experiments using synchrotron radiation.

### 2.5.2 Sample delivery

Common sample delivery methods for SFX experiments include: liquid jet systems,[224,225] viscous extruder (e. g lipidic cubic phase (LCP), hydroxycellulose, grease),[226] serial fs rotation crystallography (SF-ROX),[161,163,227] fixed target,[228,229] and acoustic droplet ejection (ADE, Section **2.5.2**) with solid support or on a moving conveyor belt[230–232]. Liquid jet systems produce a continuous stream of microcrystal slurry, typically at high speeds (10 – 100 μL · $min^{-1}$) and containing small crystals (<50 μm).[219] These jet systems allow for fast mixing experiments and can be combined with high data collection rates (up to 1 MHz).[27] In general, liquid jet systems have a relatively low hit rate (compared to other sample delivery systems) and thus, can have high sample consumption rates (10 – 60 μL · $min^{-1}$).[219] The narrow capillaries used to deliver

crystals to the X-ray beam, place a low upper limit to the crystal dimensions. Viscous extruder systems use greasy matrices, such as LCP as a medium to deliver crystals to an X-ray source.[226,233] Membrane proteins are often difficult to crystallise (Section **2.4**) and can benefit from crystallisation in LCP.[211] Solubility and stability of the membrane proteins is enhanced by the matrix of LCP, which mimics a trans membrane interface[234]. Thus grease extruders are often used in combination with membrane proteins crystallised in LCP.[233] In addition, LCP crystallisation conditions generally tend to favour smaller crystals on average, which again benefit from XFELs due to the higher brilliance. Viscous extruders offer a low sample consumption rate (<1 μL · min$^{-1}$),[219] but are limited to samples compatible with the matrices, and ill-suited for mixing experiments due to the high viscosity of the samples used. The use of goniometers in SF-ROX is an adaptation of well-established rotation crystallography methods employed at synchrotrons. Instead of a continuously replenishing source of microcrystals, a single larger crystal is used. The crystal is rotated and displaced in order to provide multiple unexposed sample volumes.[161,163,235] Energy and photon density of the beam are tuned (e. g. by defocusing) such that the sample is not destroyed at the interaction region.[219] Release of photoelectrons as a consequence of radiation damage renders a portion of the sample unusable.[219] The size of this area is usually empirically established.[161] This approach is particularly efficient as it requires only a few crystals, but also provides control over orientation of the crystal, avoiding the need to acquire substantial number of images as a result of random orientation of crystals in different serial methods (e. g. liquid jets, fixed target, and viscous extruder). Fixed target methods for sample delivery rely on crystals from a crystal slurry to be located at predictable locations on a grid.[228] This allows for precise control over the possible location of crystals and consequently reduces sample consumption (~1 μL · min$^{-1}$).[219] The crystals are typically captured inside small funnel-shaped wells present on a membrane, putting constraints on crystal dimensions. A systematic orientation of crystals within the wells is observed at higher concentrations of crystals, which could affect completeness and acquisition time of the structure. Recent developments have led to an approach that allows for mixing/diffusion of substrates into crystals on fixed target chips introduced through liquid application methods.[236,237] The ADE approach uses a high frequency transducer to create small droplets (pL to nL size, 20 – 200 μm diameter) containing crystals. Droplets can be dispensed onto a fixed mount,[238] or alternatively onto a conveyor belt and moved to the X-ray interaction region[232]. Droplets are dispensed on demand to the X-ray source to minimise sample consumption.

Alternatively, a tape drive system dispensing a continuous stream of microcrystal slurry can be used.[239] The tape-drive system was designed to enable mix-and-diffuse type experiments with high repetition rates.

Although the different sample delivery methods discussed above were put in context for use at XFELs, these methods can be extended to use at synchrotron facilities. Indeed, many of the examples discussed above have been used at synchrotrons.

### 2.5.2 Drop-on-demand sample delivery system

Building on earlier concepts, a more recently developed system for sample delivery in SFX, uses an ADE in combination with a conveyer belt, to carry crystal containing droplets to the X-ray beam. The system will be referred to as the drop-on-demand (DOD, also known as drop-on-tape) sample delivery system.[230] The sample consumption for the DOD system is relatively conservative (5 – 30 μL · $min^{-1}$), and is suitable for a range of crystal sizes (5 – 200 μm). It accommodates a versatile "reaction region", which enables chemical or physical initiation of reactions, e. g. oxygenation[230] or laser irradiation[204]. Alternatively, the reaction region can be equipped with a second acoustic dispenser to enable drop-on-drop mixing. The drop-on-drop approach could be used to add ligands or substrates to the microcrystals and allow for mix-and-diffuse type experiments. The sample is provided to the ADE dispenser *via* a capillary (200 – 300 μm inner diameter), putting an upper limit to the crystal size. Although SFX is not inherently limited in crystal size, sample availability or geometric constraints of the setup can limit the crystal size. In addition to X-ray diffraction data, the DOD system can be fitted with a von Hamos setup which is used to collect X-ray emission spectroscopy (XES, Section **2.6**) data at a 90° angle, allowing for simultaneous data collection. XES data provides detailed information on the electronics of transition state metals in the active site of enzymes (Figure 2.3). Collecting XES data requires a customised setup using high resolution analyser crystals, to collect a large solid angle of emitted photons. The instrumentation required for XES experiments is nonstandard and thus the technique is less commonly used. It does however exhibit particular advantages over analogous techniques (e. g. X-ray absorption spectroscopy); a single photon energy is sufficient to obtain a spectrum, and the technique is insensitive to fluctuations in photon energy as long as it is above the absorption edge of the element under investigation.

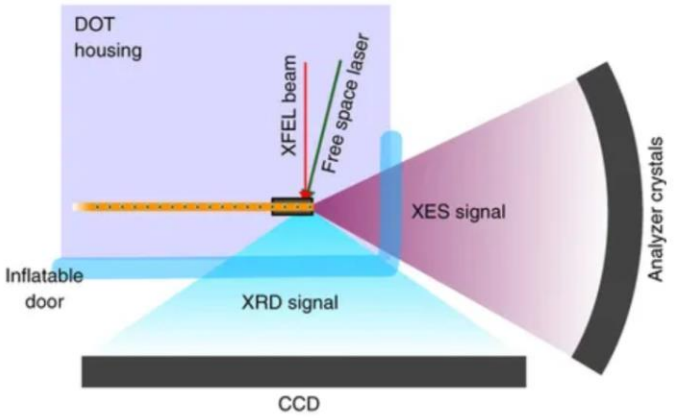



**Figure 2.3:** Schematic representation of the instrumentation used to collect X-ray emission spectroscopy (XES) and X-ray diffraction data, reprinted with permission from Nature Methods (Nature publishing group)[230]. A schematic overview of XES and diffraction data collection using the drop on demand sample delivery system. Note, diffraction data is collected in-line with the XFEL beam, while XES data is collected at a 90° angle.

## 2.6 X-ray emission spectroscopy[240]

XES is a powerful tool to study the occupied electron orbitals of transition metal complexes, and provides information on the effective spin state of transition metal ions at enzyme active sites.[240] Interaction of a transition metal complex with hard X-rays can result in photoionisation, creating a 1s core hole (Section **2.2.2**).[240] The excited state decays by filling the core hole with an electron from an outer shell; 2p to 1s (Kα), 3p to 1s ($K\beta_{1,3}$), or 3d to 1s ($K\beta_{2,5}$). Possible transitions are dictated by selection rules and occur with the emission of a photon. The energy of the emitted photon is characteristic for the transition occurring. The strongest observed fluorescence lines, Kα, derive from 2p to 1s transitions. Transitions from higher orbitals give rise to weaker fluorescence lines; 8 fold lower for $K\beta_{1,3}$ and 50 – 100 fold lower for $K\beta_{2,5}$ compared to Kα lines. The fluorescence lines are sensitive to changes in the number of unpaired electrons in the valance orbital. A change from high spin to low spin for example results in a substantial change in peak shape and chemical shift, while addition or removal of a single electron (change in oxidation state) constitute a more subtle change. This sensitivity is also dependent on the fluorescence line being observed; changes in the number of unpaired electrons are more pronounced in the Kβ lines than Kα lines, due to the larger overlap of 3p and 3d wavefunction compared to the 2p and 3d wavefunction.

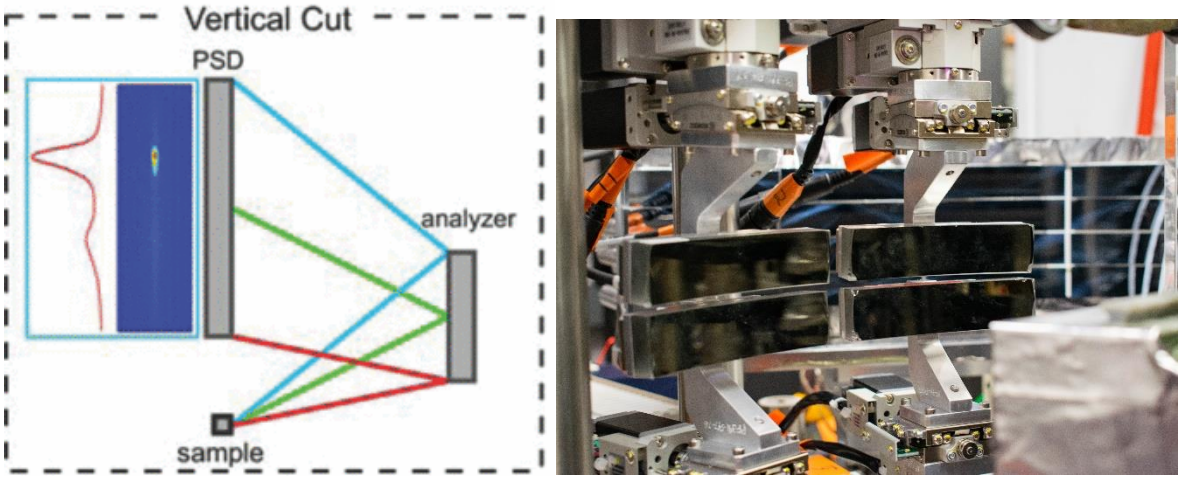



**Figure 2.4:** Schematic representation of the von Hamos setup and picture image of the analyser crystal setup used to collect X-ray emission spectroscopy (XES), reprinted (schematic picture) with permission from AIP Review of Scientific Instruments (AIP publishing)[241]. Left: schematic representation of XES data collection, emitted from the sample, reflected by the crystals onto the photo sensitive detector. Right: picture image of the analyser crystal setup.

Different experimental setups have successfully been used to record XES data. Focusing of the signal is typically done using spherically bent analyser crystals, positioned along the Roland circle.[242] Depending on the manufacturing process, either Johann[243] or Johansson[244] type analyser crystals are used to refocus a large solid angle onto the detector. This particular setup is commonly used and provides a high signal to noise and operates by scanning along the different photon energies. However, due to the nature of the XFEL experiment, it is desirable to collect the full energy spectrum in a single shot.[241,245] A dispersive setup, using analyser crystals in a von Hamos geometry enables collecting a 2D energy plane (Figure 2.4). Thus, the DOD system is equipped with analyser crystals in a von Hamos geometry to enable a spectral collection in shot-by-shot energy dispersive fashion.[230] Similar to the electron density data, XES measurements benefit from the "probe-before-destroy" feature of the fs pulses used at XFEL facilities, by eliminating the X-ray damage.[245]

## 2.7 Experimental setup

The XFEL experiments carried out on OXA-10 and IPNS were done in collaboration with Dr Allen Orville (XFEL hub, DLS, Didcot) and the group of Dr Jan Kern, Dr Junko Yano, and Dr Vittal Yachandra (Lawrence Berkeley National Laboratory (LBNL), Berkeley). The DOD strategy was used as the preferred sample delivery system for both enzymes. For mixing experiments the reaction region of the DOD system was equipped with a PolyPico droplet dispenser[246] that allowed for mixing of substrate with the OXA-10 microcrystals, while $O_2$ incubation of IPNS was achieved by equipping the reaction region of the DOD system with an oxygenation chamber. Incubation experiments of IPNS with $O_2$ provided Fe Kα XES data, in addition to X-ray diffraction data, which was recorded simultaneously at a 90° angle (Figure 2.4). Data processing during the XFEL experiments was carried out by members of the XFEL hub and LBNL lab. Although different sample delivery systems were considered, the DOD system was championed for these experiments on the following basis: (i) the relatively conservative sample consumption; (ii) the possibility to remain under atmospheric pressure of (inert) gas (i. e. He); (iii) the ability to control the initiation of the reaction by exposure to $O_2$ in a controlled fashion; (iv) the possibility to field a different reaction triggering mechanism (e. g. drop-on-drop mixing for OXA-10 and $O_2$ exposure for IPNS); and (v) the possibility to simultaneously collect XES data supporting the interpretation of the diffraction data.

## 2.8 References


1. Göbl, C., Madl, T., Simon, B. & Sattler, M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **80**, 26–63 (2014).
2. Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. & Allain, F. H. T. Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **58**, 1–61 (2011).
3. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602 (2019).
4. Wallace, B. . & Janes, R. W. Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics. *Curr. Opin. Chem. Biol.* **5**, 567–571 (2001).
5. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 (2015).
6. Li, J. *et al.* Application of the small-angle X-ray scattering technique for structural analysis studies: A review. *J. Mol. Struct.* **1165**, 391–400 (2018).
7. Tuukkanen, A. T., Spilotros, A. & Svergun, D. I. Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons. *IUCrJ* **4**, 518–528 (2017).
8. Dauter, Z. Current state and prospects of macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1–11 (2006).
9. Dobson, C. M. Biophysical Techniques in Structural Biology. *Annu. Rev. Biochem.* **88**, 25–33 (2019).
10. Hasnain, S. Impact and influence of crystallography across the sciences. *IUCrJ* **3**, 389–390 (2016).
11. Berman, H. M. *et al.* Federating Structural Models and Data: Outcomes from A Workshop on Archiving Integrative Structures. *Structure* **27**, 1745–1759 (2019).
12. Beale, J. H. Macromolecular X-ray crystallography: Soon to be a road less travelled? *Acta Crystallogr. Sect. D Struct. Biol.* **76**, 400–405 (2020).
13. Skarzynski, T. Collecting data in the home laboratory: evolution of X-ray sources, detectors and working practices. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1283–1288 (2013).
14. Elder, F. R., Gurewitsch, A. M. & Langmuir, R. V. Radiation from Electrons in a Synchrotron. *Phys. Rev.* **71**, 829–830 (1947).
15. Robinson, A. L. History of Synchrotron Radiation. *Synchrotron Radiat. News* **28**, 4–9 (2015).
16. Couprie, M. E. & Filhol, J. M. X radiation sources based on accelerators. *Comptes Rendus Phys.* **9**, 487–506 (2008).
17. Zholents, A., Heimann, P., Zolotorev, M. & Byrd, J. Generation of subpicosecond X-ray pulses using RF orbit deflection. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **425**, 385–389 (1999).
18. Sun, C., Portmann, G., Hertlein, M., Kirz, J. & Robin, D. S. Pseudo-single-bunch with adjustable frequency: A new operation mode for synchrotron light sources. *Phys. Rev. Lett.* **109**, 1–5 (2012).
19. Holldack, K. *et al.* Single bunch X-ray pulses on demand from a multi-bunch synchrotron radiation source. *Nat. Commun.* **5**, 1–7 (2014).
20. Jankowiak, A. & Wüstefeld, G. Low-α Operation of BESSY II and Future Plans for an Alternating Bunch Length Scheme BESSY$^{VSR}$. *Synchrotron Radiat. News* **26**, 22–24 (2013).
21. Silly, M. G. *et al.* Pump-probe experiments at the TEMPO beamline using the low-α operation mode of Synchrotron SOLEIL. *J. Synchrotron Radiat.* **24**, 886–897 (2017).
22. Leonarski, F. *et al.* Fast and accurate data collection for macromolecular crystallography using the JUNGFRAU detector. *Nat. Methods* **15**, 799–804 (2018).
23. Casanas, A. *et al.* EIGER detector: Application in macromolecular crystallography: *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 1036–1048 (2016).

24. Leonarski, F. *et al.* JUNGFRAU detector for brighter X-ray sources: Solutions for IT and data science challenges in macromolecular crystallography. *Struct. Dyn.* **7**, (2020).
25. Tschentscher, T. *et al.* Photon Beam Transport and Scientific Instruments at the European XFEL. *Appl. Sci.* **7**, 592 (2017).
26. Wunderer, C. B. *et al.* Detector developments at DESY. *J. Synchrotron Radiat.* **23**, 111–117 (2016).
27. Grünbein, M. L. *et al.* Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nat. Commun.* **9**, 1–9 (2018).
28. Eriksson, M., Van Der Veen, J. F. & Quitmann, C. Diffraction-limited storage rings-A window to the science of tomorrow. *J. Synchrotron Radiat.* **21**, 837–842 (2014).
29. Yabashi, M. & Tanaka, H. The next ten years of X-ray science. *Nat. Photonics* **11**, 12–14 (2017).
30. Androula Alekou *et al.* Double Triple Bend Achromat for Next Generation 3 GeV Light Sources. *Napac 2016* 1237 (2016).
31. Cipriani, F. *et al.* Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1251–1259 (2006).
32. Nurizzo, D. *et al.* RoboDiff: Combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 966–975 (2016).
33. Bowler, M. W. *et al.* MASSIF-1: A beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J. Synchrotron Radiat.* **22**, 1540–1547 (2015).
34. Grimes, J. M. *et al.* Where is crystallography going? *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 152–166 (2018).
35. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
36. Powell, H. R. X-ray data processing. *Biosci. Rep.* **37**, 1–14 (2017).
37. Winter, G. & McAuley, K. E. Automated data collection for macromolecular crystallography. *Methods* **55**, 81–93 (2011).
38. Owen, R. L., Juanhuix, J. & Fuchs, M. Current advances in synchrotron radiation instrumentation for MX experiments. *Arch. Biochem. Biophys.* **602**, 21–31 (2016).
39. Andrä, M. *et al.* Development of low-energy X-ray detectors using LGAD sensors. *J. Synchrotron Radiat.* **26**, 1226–1237 (2019).
40. Gabadinho, J. *et al.* MxCuBE: A synchrotron beamline control environment customized for macromolecular crystallography experiments. *J. Synchrotron Radiat.* **17**, 700–707 (2010).
41. Leslie, A. G. W. *et al.* Automation of the collection and processing of X-ray diffraction data - A generic approach. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1924–1928 (2002).
42. Garman, E. F. Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules. *Science.* **343**, 1102–1108 (2014).
43. Dauter, Z. & Wlodawer, A. Progress in protein crystallography. *Protein Pept. Lett.* **23**, 201–210 (2016).
44. Shaw Stewart, P. & Mueller-Dieckmann, J. Automation in biological crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70**, 686–696 (2014).
45. Shaw Stewart, P. D. & Baldock, P. F. M. Practical experimental design techniques for automatic and manual protein crystallization. *J. Cryst. Growth* **196**, 665–673 (1999).
46. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 414–423 (1994).
47. Shah, A. K. *et al.* On increasing protein-crystallization throughput for X-ray diffraction studies. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 123–129 (2005).
48. Hiraki, M. *et al.* Development of an automated large-scale protein-crystallization and monitoring system for high-throughput protein-structure analyses. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1058–1065 (2006).
49. Luft, J. R. *et al.* Macromolecular crystallization in a high throughput laboratory - The search phase. *J. Cryst. Growth* **232**, 591–595 (2001).

50. Cipriani, F. *et al.* CrystalDirect: A new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 1393–1399 (2012).

51. Moffat, K. Time-Resolved Crystallography. *Acta Crystallogr. Sect. A Found. Crystallogr.* **54**, 833–841 (1998).

52. Schlichting, I. & Goody, R. S. Triggering methods in crystallographic enzyme kinetics. in *Methods* vol. 277 467–490 (1997).

53. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662–666 (1958).

54. Perutz, M. *et al.* Structure of haemoglobin. *Nature* **185**, 416–422 (1960).

55. Johnson, L. N. & Phillips, D. C. Structure of Some Crystalline Lysozyme-Inhibitor Complexes Determined by X-Ray Analysis At 6 Å Resolution. *Nature* **206**, 761–763 (1965).

56. Blake, C. *et al.* The three-dimensional structure of hen eggwhite lysozyme. *Nature* **206**, 757–761 (1965).

57. Palmer, R. A. & Niwa, H. X-ray crystallographic studies of protein – ligand interactions. 973–979 (2003).

58. Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995–1014 (2014).

59. Matthews, B. W. X-ray Crystallographic Studies of Proteins. *Annu. Rev. Phys. Chem.* **27**, 493–493 (1976).

60. Weichenberger, C. X. & Rupp, B. Ten years of probabilistic estimates of biocrystal solvent content: New insights via nonparametric kernel density estimate. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1579–1588 (2014).

61. Mozzarelli, A. & Rossi, G. L. Protein function in the crystal. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 343–365 (1996).

62. Moffat, K. Laue diffraction and time-resolved crystallography: A personal history. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **377**, (2019).

63. Wilmouth, R. C., Clifton, I. J. & Neutze, R. Recent successes in time-resolved protein crystallography. *Nat. Prod. Rep.* **17**, 527–533 (2000).

64. Martin, J. L. & Vos, M. H. Femtosecond biology. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 199–222 (1992).

65. Moffat, K. Femtosecond structural photobiology. *Science.* **361**, 127–128 (2018).

66. Zegers, I. *et al.* Hydrolysis of a slow cyclic thiophosphate substrate of RNase T1 analyzed by time-resolved crystallography. *Nat. Struct. Biol.* **5**, 280–283 (1998).

67. Bourgeois, D. & Weik, M. *Kinetic protein crystallography: A tool to watch proteins in action*. *Crystallography Reviews* vol. 15 (2009).

68. Jung, A., Reinstein, J., Domratcheva, T., Shoeman, R. L. & Schlichting, I. Crystal Structures of the AppA BLUF Domain Photoreceptor Provide Insights into Blue Light-mediated Signal Transduction. *J. Mol. Biol.* **362**, 717–732 (2006).

69. Möglich, A. & Moffat, K. Structural Basis for Light-dependent Signaling in the Dimeric LOV Domain of the Photosensor YtvA. *J. Mol. Biol.* **373**, 112–126 (2007).

70. Ihee, H. *et al.* Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7145–7150 (2005).

71. Van Thor, J. J., Georgiev, G. Y., Towrie, M. & Sage, J. T. Ultrafast and low barrier motions in the photoreactions of the green fluorescent protein. *J. Biol. Chem.* **280**, 33652–33659 (2005).

72. Bourgeois, D. *et al.* Extended subnanosecond structural dynamics of myoglobin revealed by Laue crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4924–4929 (2006).

73. Josts, I. *et al.* Photocage-initiated time-resolved solution X-ray scattering investigation of protein dimerization. *IUCrJ* **5**, 667–672 (2018).

74. Hadfield, A. & Hajdu, J. On the photochemical release of phosphate from 3,5-dinitrophenyl phosphate in a protein crystal. *J. Mol. Biol.* **236**, 995–1000 (1994).

75. Ursby, T. *et al.* Cryophotolysis of caged compounds: A technique for trapping intermediate states in protein crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 607–614 (2002).

76. Adams, S. R. & Tsien, R. Y. Controlling Cell Chemistry with Caged Compounds. *Annu. Rev. Physiol.* **55**, 755–784 (1993).

77. Šrajer, V. *et al.* Photolysis of the carbon monoxide complex of myoglobin: Nanosecond time- resolved crystallography. *Science.* **274**, 1726–1729 (1996).

78. Wang, Y. & Guo, H. C. Crystallographic Snapshot of a Productive Glycosylasparaginase-Substrate Complex. *J. Mol. Biol.* **366**, 82–92 (2007).

79. Berthold, C. L. *et al.* Crystallographic Snapshots of Oxalyl-CoA Decarboxylase Give Insights into Catalysis by Nonoxidative ThDP-Dependent Decarboxylases. *Structure* **15**, 853–861 (2007).

80. Cavazza, C. *et al.* Flexibility of thiamine diphosphate revealed by kinetic crystallographic studies of the reaction of pyruvate-ferredoxin oxidoreductase with pyruvate. *Structure* **14**, 217–224 (2006).

81. Hajdu, J., Acharya, K. R., Stuart, D. I., Barford, D. & Johnson, L. N. Catalysis in enzyme crystals. *Trends Biochem. Sci.* **13**, 104–109 (1988).

82. Hajdu, J. *et al.* Analyzing protein functions in four dimensions. *Nat. Struct. Biol.* **7**, 1006–1012 (2000).

83. O'Hara, P., Goodwin, P. & Stoddard, B. L. Direct measurement of diffusion rates in enzyme crystals by video absorbance spectroscopy. *J. Appl. Crystallogr.* **28**, 829–834 (1995).

84. Schmidt, M. Reaction initiation in enzyme crystals by diffusion of substrate. *Crystals* **10**, 1–15 (2020).

85. Geremia, S., Campagnolo, M., Demitri, N. & Johnson, L. N. Simulation of diffusion time of small molecules in protein crystals. *Structure* **14**, 393–400 (2006).

86. Schmidt, M. Mix and Inject: Reaction Initiation by Diffusion for Time-Resolved Macromolecular Crystallography. *Adv. Condens. Matter Phys.* **2013**, 1–10 (2013).

87. Moffat, K. Laue diffraction. in *Methods in Enzymology* vol. 277 433–447 (1997).

88. Ren, Z. *et al.* Laue crystallography: Coming of age. *J. Synchrotron Radiat.* **6**, 891–917 (1999).

89. Stoddard, B. L. New results using Laue diffraction and time-resolved crystallography. *Curr. Opin. Struct. Biol.* **8**, 612–618 (1998).

90. Genick, U. K. *et al.* Structure of a protein photocycle intermediate by millisecond time- resolved crystallography. *Science.* **275**, 1471–1475 (1997).

91. Oang, K. Y. *et al.* Sub-100-ps structural dynamics of horse heart myoglobin probed by time-resolved X-ray solution scattering. *Chem. Phys.* **442**, 137–142 (2014).

92. Bourgeois, D., Schotte, F., Brunori, M. & Vallone, B. Time-resolved methods in biophysics. 6. Time-resolved Laue crystallography as a tool to investigate photo-activated protein dynamics. *Photochem. Photobiol. Sci.* **6**, 1047–1056 (2007).

93. Moffat, K. & Henderson, R. Freeze trapping of reaction intermediates. *Curr. Opin. Struct. Biol.* **5**, 656–663 (1995).

94. Bourgeois, D. & Royant, A. Advances in kinetic protein crystallography. *Curr. Opin. Struct. Biol.* **15**, 538–547 (2005).

95. Rodgers, D. W. Cryocrystallography. *Structure* **2**, 1135–1140 (1994).

96. Garman, E. F. Radiation damage in macromolecular crystallography: What is it and why should we care? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 339–351 (2010).

97. Mesecar, A. D., Stoddard, B. L. & Koshland, D. E. Orbital steering in the catalytic power of enzymes: Small structural changes with large catalytic consequences. *Science.* **277**, 202–206 (1997).

98. Burzlaff, N. I. *et al.* The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401**, 721–724 (1999).

99. Verschueren, K. H. G., Seljee, F., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* **363**, 693–698 (1993).

100. Standfuss, J. & Spence, J. Serial crystallography at synchrotrons and X-ray lasers. *IUCrJ* **4**, 100–101 (2017).

101. Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. A Bright Future for Serial Femtosecond Crystallography with XFELs. *Trends in Biochemical Sciences* vol. 42 749–762 (2017).

102. Diederichs, K. & Wang, M. Serial Synchrotron X-Ray Crystallography (SSX). in *Comprehensive Medicinal Chemistry II* vol. 3 239–272 (2017).

103. de la Mora, E. *et al.* Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4142–4151 (2020).

104. Ebrahim, A. *et al.* Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 151–159 (2019).

105. Sanchez-Weatherby, J. *et al.* VMXi: a fully automated, fully remote, high-flux in situ macromolecular crystallography beamline. *J. Synchrotron Radiat.* **26**, 291–301 (2019).

106. Miller, M. S. *et al.* Getting the most out of your crystals: Data collection at the new high-flux, Microfocus MX beamlines at NSLS-II. *Molecules* **24**, (2019).

107. Flot, D. *et al.* The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).

108. Smith, J. L., Fischetti, R. F. & Yamamoto, M. Micro-crystallography comes of age. *Curr. Opin. Struct. Biol.* **22**, 602–612 (2012).

109. Zinth, W., Laubereau, A. & Kaiser, W. The long journey to the laser and its rapid development after 1960. *Eur. Phys. J. H* **36**, 153–181 (2011).

110. Hecht, J. Short history of laser development. *Opt. Eng.* **49**, 091002 (2010).

111. Motz, H. Applications of the radiation from fast electron beams. *J. Appl. Phys.* **22**, 527–535 (1951).

112. Madey, J. M. J. Stimulated emission of bremsstrahlung in a periodic magnetic field. *J. Appl. Phys.* **42**, 1906–1913 (1971).

113. Deacon, D. A. G. *et al.* First Operation of a Free-Electron Laser. *Phys. Rev. Lett.* **38**, 892–894 (1977).

114. Yamamoto, S., Shioya, T., Sasaki, S. & Kitamura, H. Construction of insertion devices for elliptically polarized synchrotron radiation. *Rev. Sci. Instrum.* **60**, 1834–1837 (1989).

115. Elleaume, P. Insertion devices for the new generation of synchrotron sources: A review. *Rev. Sci. Instrum.* **63**, 321–326 (1992).

116. Pellegrini, C. The history of X-ray free-electron lasers. *Eur. Phys. J. H* **37**, 659–708 (2012).

117. Geloni, G., Huang, Z. & Pellegrini, C. CHAPTER 1. The Physics and Status of X-ray Free-electron Lasers. in 1–44.

118. Tiedtke, K. *et al.* The soft x-ray free-electron laser FLASH at DESY: Beamlines, diagnostics and end-stations. *New J. Phys.* **11**, (2009).

119. Shintake, T. *et al.* A compact free-electron laser for generating coherent radiation in the extreme ultraviolet region. *Nat. Photonics* **2**, 555–559 (2008).

120. Emma, P. *et al.* First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photonics* **4**, 641–647 (2010).

121. Ishikawa, T. *et al.* A compact X-ray free-electron laser emitting in the sub-ångström region. *Nat. Photonics* **6**, 540–544 (2012).

122. Feng, Y. *et al.* Demonstration of simultaneous experiments using thin crystal multiplexing at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 626–633 (2015).

123. Zhu, D. *et al.* Performance of a beam-multiplexing diamond crystal monochromator at the Linac Coherent Light Source. *Rev. Sci. Instrum.* **85**, (2014).

124. Milton, S. V. *et al.* Exponential gain and saturation of a self-amplified spontaneous emission free-electron laser. *Science.* **292**, 2037–2041 (2001).

125. Assoufid, L. & Graafsma, H. Next-generation materials for future synchrotron and free-electron laser sources. *MRS Bull.* **42**, 418–423 (2017).

126. Huang, Z. & Kim, K. J. Review of x-ray free-electron laser theory. *Phys. Rev. Spec. Top. - Accel. Beams* **10**, 1–26 (2007).

127. Pellegrini, C., Marinelli, A. & Reiche, S. The physics of x-ray free-electron lasers. *Rev. Mod. Phys.* **88**, 1–55

(2016).

128. McNeil, B. W. J. & Thompson, N. R. X-ray free-electron lasers. *Nat. Photonics* **4**, 814–821 (2010).

129. Bonifacio, R., De Salvo, L., Pierini, P., Piovella, N. & Pellegrini, C. Spectrum, temporal structure, and fluctuations in a high-gain free-electron laser starting from noise. *Phys. Rev. Lett.* **73**, 70–73 (1994).

130. Kondratenko, A. M. & Saldin, E. L. Generation of Coherent Radiation By a Relativistic Electron Beam in an Ondulator. *Part. Accel. Print* **10**, 207–216 (1980).

131. Inubushi, Y. *et al.* Determination of the Pulse Duration of an X-Ray Free Electron Laser Using Highly Resolved Single-Shot Spectra. *Phys. Rev. Lett.* **109**, 6–9 (2012).

132. Amann, J. *et al.* Demonstration of self-seeding in a hard-X-ray free-electron laser. *Nat. Photonics* **6**, 693–698 (2012).

133. Giannessi, L. Seeding and Harmonic Generation in Free-Electron Lasers. in *Synchrotron Light Sources and Free-Electron Lasers* 119–147 (Springer International Publishing, 2020).

134. Geloni, G. Self-Seeded Free-Electron Lasers. in *Synchrotron Light Sources and Free-Electron Lasers* 191–223 (Springer International Publishing, 2020).

135. Patterson, B. D. & Abela, R. Novel opportunities for time-resolved absorption spectroscopy at the X-ray free electron laser. *Phys. Chem. Chem. Phys.* **12**, 5647 (2010).

136. Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. Beyond crystallography: Diffractive imaging using coherent X-ray light sources. *Science.* **348**, 530–535 (2015).

137. Koch, A. *et al.* Operation of photon diagnostic imagers for beam commissioning at the European XFEL. *J. Synchrotron Radiat.* **26**, 1489–1495 (2019).

138. Ding, Y., Huang, Z., Ratner, D., Bucksbaum, P. & Merdji, H. Generation of attosecond x-ray pulses with a multicycle two-color enhanced self-amplified spontaneous emission scheme. *Phys. Rev. Spec. Top. - Accel. Beams* **12**, 2–7 (2009).

139. Orville, A. M. Entering an era of dynamic structural biology. *BMC Biol.* **16**, 1–4 (2018).

140. Kang, H. S. *et al.* Hard X-ray free-electron laser with femtosecond-scale timing jitter. *Nat. Photonics* **11**, 708–713 (2017).

141. Milne, C. J. *et al.* SwissFEL: The Swiss X-ray Free Electron Laser. *Appl. Sci.* **7**, 1–57 (2017).

142. Garman, E. Cool data: Quantity AND quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 1641–1653 (1999).

143. Pflugrath, J. W. Practical macromolecular cryocrystallography. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **71**, 622–642 (2015).

144. Nass, K. Radiation damage in protein crystallography at X-ray free-electron lasers. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 211–218 (2019).

145. Garman, E. F. & Weik, M. X-ray radiation damage to biological samples: recent progress. *J. Synchrotron Radiat.* **26**, 907–911 (2019).

146. Caleman, C. *et al.* Ultrafast self-gating Bragg diffraction of exploding nanocrystals in an X-ray laser. *Opt. Express* **23**, 1213 (2015).

147. Campbell, J. L. & Papp, T. Widths of the atomic *K-N*7 levels. *At. Data Nucl. Data Tables* **77**, 1–56 (2001).

148. Caleman, C., Jares Junior, F., Grånäs, O. & Martin, A. V. A Perspective on Molecular Structure and Bonding-Breaking in Radiation Damage in Serial Femtosecond Crystallography. *Crystals* **10**, 585 (2020).

149. Hau-Riege, S. P. Nonequilibrium electron dynamics in materials driven by high-intensity x-ray pulses. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **87**, 4–7 (2013).

150. Aquila, A. *et al.* Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt. Express* **20**, 2706 (2012).

151. Ziaja, B., Szöke, A., Van der Spoel, D. & Hajdu, J. Space-time evolution of electron cascades in diamond. *Phys. Rev. B - Condens. Matter Mater. Phys.* **66**, 241161–241169 (2002).

152. Garrison, W. M. Reaction Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins. *Chem. Rev.* **87**, 381–398 (1987).

153. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).

154. Warkentin, M. *et al.* Global radiation damage: Temperature dependence, time dependence and how to outrun it. *J. Synchrotron Radiat.* **20**, 7–13 (2013).

155. Garman, E. F. & Owen, R. L. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 32–47 (2006).

156. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).

157. Barty, A. *et al.* Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photonics* **6**, 35–40 (2012).

158. Lomb, L. *et al.* Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys. Rev. B - Condens. Matter Mater. Phys.* **84**, 1–6 (2011).

159. Nass, K. *et al.* Indications of radiation damage in ferredoxin microcrystals using high-intensity X-FEL beams. *J. Synchrotron Radiat.* **22**, 225–238 (2015).

160. Nass, K. *et al.* Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses. *Nat. Commun.* **11**, 1–9 (2020).

161. Hirata, K. *et al.* Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. *Nat. Methods* **11**, 734–736 (2014).

162. Kern, J. *et al.* Room temperature femtosecond X-ray diffraction of photosystem II microcrystals. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9721–9726 (2012).

163. Halsted, T. P. *et al.* Catalytically important damage-free structures of a copper nitrite reductase obtained by femtosecond X-ray laser and room-temperature neutron crystallography. *IUCrJ* **6**, 761–772 (2019).

164. Chapman, H. N., Caleman, C. & Timneanu, N. Diffraction before destruction. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, (2014).

165. Chapman, H. N. *et al.* Femtosecond time-delay X-ray holography. *Nature* **448**, 676–679 (2007).

166. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16247–16252 (2011).

167. Marchesini, S. *et al.* Coherent X-ray diffractive imaging: applications and limitations. *Opt. Express* **11**, 2344 (2003).

168. Kurta, R. P., Altarelli, M. & Vartanyants, I. A. X-ray cross-correlation analysis of disordered ensembles of particles: Potentials and limitations. *Adv. Condens. Matter Phys.* **2013**, (2013).

169. Grübel, G., Stephenson, G. B., Gutt, C., Sinn, H. & Tschentscher, T. XPCS at the European X-ray free electron laser facility. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms* **262**, 357–367 (2007).

170. Sinha, S. & Tolan, M. Effects of partial coherence on the scattering of x rays by matter. *Phys. Rev. B - Condens. Matter Mater. Phys.* **57**, 2740–2758 (1998).

171. Taylor, G. The phase problem. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 1881–1890 (2003).

172. Sayre, D. Some implications of a theorem due to Shannon. *Acta Crystallogr.* **5**, 843–843 (1952).

173. Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758 (1982).

174. Miao, J., Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A* **15**, 1662 (1998).

175. Gerchberg, R. W. & Saxton, W. O. Gerchberg—Saxton algorithm: experimental realisation and modification for the problem of formation of multimode laser beams. *Optik (Stuttg).* **35**, 237–246 (1972).

176. Bates, R. H. T. Uniqueness of solutions to two-dimensional fourier phase problems for localized and positive images. *Comput. Vision, Graph. Image Process.* **25**, 205–217 (1984).

177. Miao, J. & Sayre, D. On possible extensions of X-ray crystallography through diffraction-pattern oversampling. *Acta Crystallogr. Sect. A Found. Crystallogr.* **56**, 596–605 (2000).

178. Nugent, K. A. *Coherent methods in the X-ray sciences. Advances in Physics* vol. 59 (2010).

179. Spence, J. C. H., Weierstall, U. & Howells, M. Coherence and sampling requirements for diffractive imaging. *Ultramicroscopy* **101**, 149–152 (2004).

180. Williams, G. J., Quiney, H. M., Peele, A. G. & Nugent, K. A. Coherent diffractive imaging and partial coherence. *Phys. Rev. B - Condens. Matter Mater. Phys.* **75**, 1–7 (2007).

181. Liu, H. & Lee, W. The XFEL protein crystallography: Developments and perspectives. *Int. J. Mol. Sci.* **20**, (2019).

182. Krekic, S. *et al.* Spectrokinetic characterization of photoactive yellow protein films for integrated optical applications. *Eur. Biophys. J.* **48**, 465–473 (2019).

183. Pellequer, J. L., Wager-Smith, K. A., Kay, S. A. & Getzoff, E. D. Photoactive yellow protein: A structural prototype for the three-dimensional fold of the PAS domain superfamily. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5884–5890 (1998).

184. Baca, M. *et al.* Complete Chemical Structure of Photoactive Yellow Protein: Novel Thioester-Linked 4-Hydroxycinnamyl Chromophore and Photocycle Chemistry. *Biochemistry* **33**, 14369–14377 (1994).

185. Kim, T. W. *et al.* Protein structural dynamics of photoactive yellow protein in solution revealed by pump-probe X-ray solution scattering. *J. Am. Chem. Soc.* **134**, 3145–3153 (2012).

186. Schotte, F. *et al.* Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19256–19261 (2012).

187. Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science.* **346**, 1242–1246 (2014).

188. Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science.* **352**, 725–729 (2016).

189. Barber, J. Photosystem II: An enzyme of global significance. *Biochem. Soc. Trans.* **34**, 619–631 (2006).

190. Nelson, N. & Ben-Shem, A. The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell Biol.* **5**, 971–982 (2004).

191. Shen, J.-R. The Structure of Photosystem II and the Mechanism of Water Oxidation in Photosynthesis. *Annu. Rev. Plant Biol.* **66**, 23–48 (2015).

192. Zouni, A. *et al.* Crystal structure of photosystem II from Synechococcus elongatus at 3.8 Å resolution. *Nature* **409**, 739–743 (2001).

193. Guskov, A. *et al.* Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.* **16**, 334–342 (2009).

194. Umena, Y., Kawakami, K., Shen, J. R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9Å. *Nature* **473**, 55–60 (2011).

195. Yeagle, G. J., Gilchrist, M. L., McCarrick, R. M. & Britt, R. D. Multifrequency pulsed electron paramagnetic resonance study of the S 2 state of the photosystem II manganese cluster. *Inorg. Chem.* **47**, 1803–1814 (2008).

196. Berthomieu, C. & Hienerwadel, R. Fourier transform infrared (FTIR) spectroscopy. *Photosynth. Res.* **101**, 157–170 (2009).

197. Haumann, M. *et al.* Biochemistry: Photosynthetic $O_2$ formation tracked by time-resolved x-ray experiments. *Science.* **310**, 1019–1021 (2005).

198. Yano, J. & Yachandra, V. K. Where water is oxidized to dioxygen: Structure of the photosynthetic $Mn_4Ca$ cluster from X-ray spectroscopy. *Inorg. Chem.* **47**, 1711–1726 (2008).

199. Ravelli, R. B. & Garman, E. F. Radiation damage in macromolecular cryocrystallography. *Curr. Opin. Struct. Biol.* **16**, 624–629 (2006).

200. Yano, J. *et al.* X-ray damage to the Mn4Ca complex in single crystals of photosystem II: A case study for metalloprotein crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12047–12052 (2005).

201. Kern, J. *et al.* Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563**, 421–425 (2018).

202. Ibrahim, M. *et al.* Untangling the sequence of events during the $S_2 \rightarrow S_3$ transition in photosystem II and implications for the water oxidation mechanism . *Proc. Natl. Acad. Sci.* **117**, 202000529 (2020).

203. Suga, M. *et al.* Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL. *Nature* **543**, 131–135 (2017).

204. Fransson, T. *et al.* X-ray Emission Spectroscopy as an in Situ Diagnostic Tool for X-ray Crystallography of Metalloproteins Using an X-ray Free-Electron Laser. *Biochemistry* **57**, 4629–4637 (2018).

205. Kern, J. *et al.* Simultaneous Femtosecond X-ray Spectroscopy and Diffraction of Photosystem II at Room Temperature. *Science.* **340**, 491–495 (2013).

206. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).

207. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).

208. Zhou, X. E., Melcher, K. & Xu, H. E. Structure and activation of rhodopsin. *Acta Pharmacol. Sin.* **33**, 291–299 (2012).

209. Zhou, X. E. *et al.* X-ray laser diffraction for structure determination of the rhodopsin-arrestin complex. *Sci. Data* **3**, 1–13 (2016).

210. Liu, W. *et al.* Serial Femtosecond Crystallography of G Protein-Coupled Receptors. *Science.* **342**, 1521–1524 (2013).

211. Zabara, A. *et al.* Lipidic Cubic Phase-Induced Membrane Protein Crystallization: Interplay between Lipid Molecular Structure, Mesophase Structure and Properties, and Crystallogenesis. *Cryst. Growth Des.* **17**, 5667–5674 (2017).

212. Deisenhofer, J. & Michel, H. The Photosynthetic Reaction Center from the Purple Bacterium Rhodopseudomonas viridis. *Science.* **245**, 1463–1473 (1989).

213. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. Biophys. Acta - Biomembr.* **1666**, 105–117 (2004).

214. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009).

215. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).

216. Pike, A. C. W., Garman, E. F., Krojer, T., Von Delft, F. & Carpenter, E. P. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 303–318 (2016).

217. Olmos, J. L. *et al.* Enzyme intermediates captured 'on the fly' by mix-and-inject serial crystallography. *BMC Biol.* **16**, 1–15 (2018).

218. Schmidt, M. Time-resolved macromolecular crystallography at pulsed X-ray sources. *Int. J. Mol. Sci.* **20**, (2019).

219. Martiel, I., Müller-Werkmeister, H. M. & Cohen, A. E. Strategies for sample delivery for femtosecond crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 160–177 (2019).

220. Dods, R. *et al.* From Macrocrystals to Microcrystals: A Strategy for Membrane Protein Serial Crystallography. *Structure* **25**, 1461-1468.e2 (2017).

221. Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science.* **337**, 362–364 (2012).

222. Lee, D. B. *et al.* Supersaturation-controlled microcrystallization and visualization analysis for serial femtosecond crystallography. *Sci. Rep.* **8**, 1–10 (2018).

223. Darmanin, C. *et al.* Protein crystal screening and characterization for serial femtosecond nanocrystallography. *Sci. Rep.* **6**, 1–8 (2016).

224. DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D. Appl. Phys.* **41**, (2008).

225. Calvey, G. D., Katz, A. M., Schaffer, C. B. & Pollack, L. Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers. *Struct. Dyn.* **3**, (2016).

226. Liu, W., Ishchenko, A. & Cherezov, V. Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nat. Protoc.* **9**, 2123–2134 (2014).

227. Schlichting, I. Serial femtosecond crystallography: The first five years. *IUCrJ* **2**, 246–255 (2015).

228. Hunter, M. S. *et al.* Fixed-target protein serial microcrystallography with an X-ray free electron laser. *Sci. Rep.* **4**, 1–5 (2014).

229. Oghbaey, S. *et al.* Fixed target combined with spectral mapping: Approaching 100% hit rates for serial crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 944–955 (2016).

230. Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat. Methods* **14**, 443–449 (2017).

231. Roessler, C. G. *et al.* Acoustic Injectors for Drop-On-Demand Serial Femtosecond Crystallography Resource Acoustic Injectors for Drop-On-Demand Serial Femtosecond Crystallography. *Structure* 631–640 (2016).

232. Roessler, C. G. *et al.* Acoustic methods for high-throughput protein crystal mounting at next-generation macromolecular crystallographic beamlines. *J. Synchrotron Radiat.* **20**, 805–808 (2013).

233. Liu, W., Wacker, D., Wang, C., Abola, E. & Cherezov, V. Femtosecond crystallography of membrane proteins in the lipidic cubic phase. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, (2014).

234. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14532–14535 (1996).

235. Halsted, T. P. *et al.* An unprecedented dioxygen species revealed by serial femtosecond rotation crystallography in copper nitrite reductase. *IUCrJ* **5**, 22–31 (2018).

236. Mehrabi, P. *et al.* The HARE chip for efficient time-resolved serial synchrotron crystallography. *J. Synchrotron Radiat.* **27**, 1–11 (2020).

237. Mehrabi, P. *et al.* Liquid application method for time-resolved analyses by serial synchrotron crystallography. *Nat. Methods* **16**, 979–982 (2019).

238. Soares, A. S. *et al.* Acoustically Mounted Microcrystals Yield High-Resolution X-ray Structures. *Biochemistry* **50**, 4399–4401 (2011).

239. Beyerlein, K. R. *et al.* Mix-and-diffuse serial synchrotron crystallography. *IUCrJ* **4**, 769–777 (2017).

240. Glatzel, P. & Bergmann, U. High resolution 1s core hole X-ray spectroscopy in 3d transition metal complexes - Electronic and structural information. *Coord. Chem. Rev.* **249**, 65–95 (2005).

241. Alonso-Mori, R. *et al.* A multi-crystal wavelength dispersive x-ray spectrometer. *Rev. Sci. Instrum.* **83**, 073114 (2012).

242. Wittry, D. B. & Barbi, N. C. X-ray crystal spectrometers and monochromators in microanalysis. *Microsc. Microanal.* **7**, 124–141 (2001).

243. Johann, H. H. Die Erzeugung lichtstarker Röntgenspektren mit Hilfe von Konkavkristallen. *Zeitschrift für Phys.* **69**, 185–206 (1931).

244. Johansson, T. Über ein neuartiges, genau fokussierendes Röntgenspektrometer. *Zeitschrift für Phys.* **82**, 507–528 (1933).

245. Alonso-Mori, R. *et al.* Energy-dispersive X-ray emission spectroscopy using an X-ray free-electron laser in a shot-by-shot mode. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19103–19107 (2012).

246. Davy, B. *et al.* Reducing sample consumption for serial crystallography using acoustic drop ejection. *J. Synchrotron Radiat.* **26**, 1820–1825 (2019).

# Chapter 3

## Time-resolved crystallography on OXA-10

# 3.1 Introduction

The work described in this chapter focuses on time resolved (tr) serial crystallography on oxacillinase-10 (OXA-10) using synchrotron and X-ray free electron laser (XFEL) radiation. The aim of the studies on OXA-10, a clinically relevant β-lactamase (Section **1.4**), was to investigate the carbapenem hydrolysis mechanism. It was envisaged that the initial focus would be on the acylation mechanism and formation of the acyl-enzyme complex, while future efforts would investigate the deacylation and formation of carbapenem derived β-lactone products. Below is given a brief description of the specific research aims of the work.

### 3.1.1 Hydrolysis of β-lactams by OXA-10

The recent development of bacterial resistance against a range of antibiotics, including penicillins, cephalosporins, and importantly carbapenems, represents a major threat to human health (Section **1.4**).[1,2] Much of the resistance against β-lactam antibiotics is manifested in the production of β-lactamases,[3] which catalytically hydrolyse the β-lactam ring.[1] The β-lactamases are divided into several classes; A, C, and D, which are all serine β-lactamases (SBLs), and class B, which comprises metallo β-lactamases (MBLs).[4] The oxacillinase (OXA) variants are class D SBLs and are characterised by a carbamylated Lys at the active site (Figure 3.1),[5] which acts as a general acid / base during hydrolysis,[6,7] and confer resistance to penicillins, cephalosporins, and, in some cases, carbapenems.[8–10] Carbapenems are considered antibiotics of last resort, and thus are of great clinical importance. Hydrolysis of carbapenems results in the formation of the ring opened $\Delta^1$ pyrroline product (Figure 3.2).[11] Tautomerization between the $\Delta^1$ pyrroline and $\Delta^2$ pyrroline forms leads to loss of the stereochemistry at C2 centre, leading to the formation of two diastereomers of the $\Delta^1$ pyrroline product form. Recently an alternative mechanism for the hydrolysis of the acyl-enzyme intermediate, forming a carbapenem derived β-lactone, was described for some OXA variants (including OXA-10, OXA-23, and OXA-48).[7,12] For some carbapenems, containing a C1 β-methyl group (i. e. ertapenem, meropenem, biopenem, and doripenem) hydrolysis of the acyl-enzyme intermediate can be mediated by the hydroxyethyl sidechain of the carbapenem, instead of a hydrolytic $H_2O$, resulting in the formation of β-lactone products (Figure 3.2).[12]

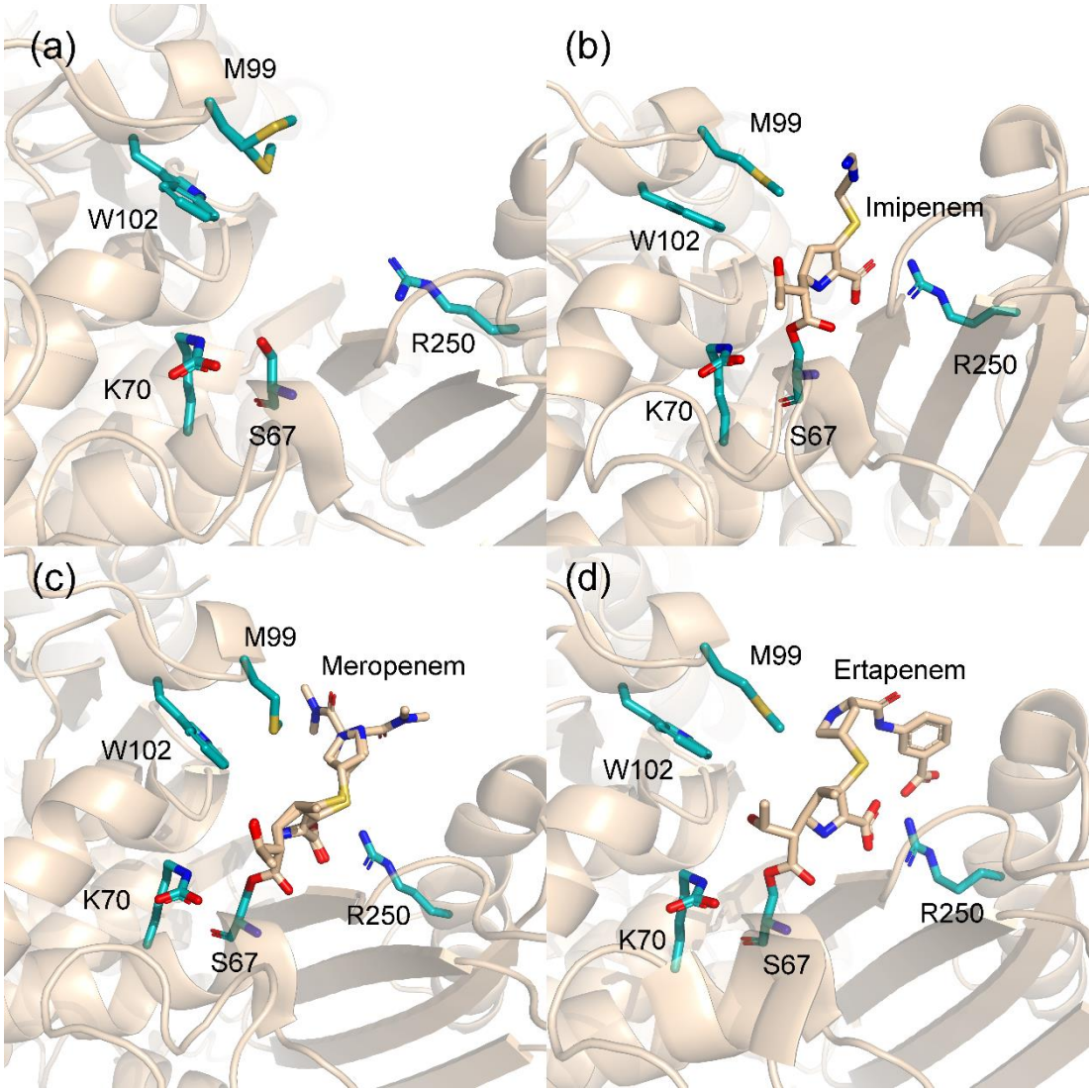



**Figure 3.1.** Views from structures of OXA-10 with different carbapenems soaked. View from structures of: (a) *apo*-OXA-10 showing Ser67 which is used for hydrolysis, and carbamylated Lys70; (b) OXA-10 in complex with imipenem; (c) OXA-10 in complex with meropenem; (d) OXA-10 in complex with ertapenem. PDB accession codes: 1K55,[5] 6SKP (imipenem), 6SKQ (meropenem), 6SKR (ertapenem).

The proposed mechanism for hydrolysis of carbapenems was based on extensive NMR experiments, UV-Vis spectroscopy and crystallography of various compounds. Within both the well-established hydrolysis mechanism and the recently reported β-lactone forming mechanism, there are unanswered questions, for which tr-SFX experiments using carbapenems and OXA enzymes, might improve understanding of the carbapenem hydrolysis mechanism by providing crystallographic evidence for intermediates. For instance, the electronic and steric effects on the C2 position of carbapenems and their impact on tautomerization between $\Delta^1$ pyrroline and $\Delta^2$ pyrroline are currently poorly defined. In addition,

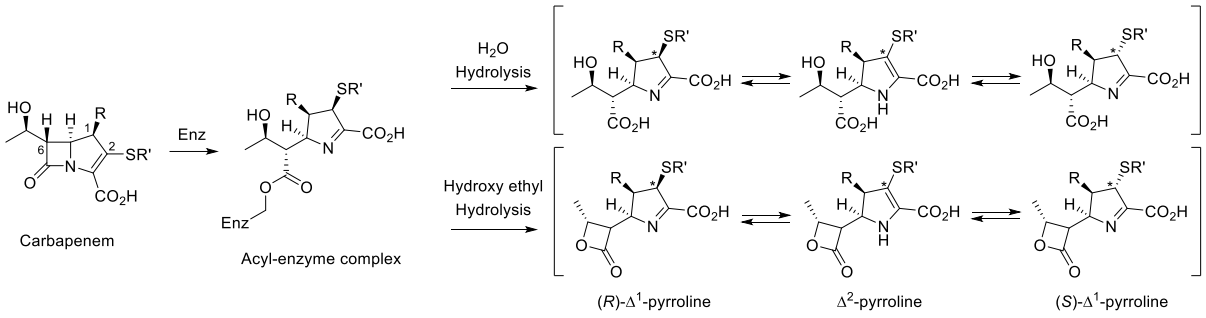



**Figure 3.2:** Carbapenem hydrolysis reaction as catalysed by serine β-lactamases (SBLs) forming general hydrolysis product, and an alternative hydrolysis pathway for C1 β-methyl carbapenems resulting in the formation of β-lactones, as observed for class D SBLs. After formation of the $\Delta^1$ pyrroline, tautomerization to $\Delta^2$ pyrroline results in loss of stereochemistry at C2 (indicated with *). The proposed mechanisms are based on NMR spectroscopy and X-ray crystallography studies.[7,12]

tr-SFX experiments could improve the understanding of the recently discovered hydrolysis mechanism that leads to the formation of β-lactone products.

### 3.1.2 Time resolved SFX using OXA-10

OXA-10 was considered for tr-SFX experiments as it has been shown to produce crystals diffracting to high resolution (<2.0 Å) using synchrotron radiation and large amounts of sample can be obtained in high purity.[13,14]. The OXA-10 active site is solvent exposed, facilitating soaking of ligands and inhibitors, as has been demonstrated with a range of compounds[14] including carbapenems imipenem, meropenem and ertapenem (Figure 3.3). Ertapenem was considered as preferred substrate, because of its clinical relevance as an antibiotic of last resort, high solubility (>100 mM), and the ability to form β-lactone products.[12] In addition to ertapenem, imipenem, ceftazidime, and clavulanic acid (Figure 3.3) were used as alternative substrates to give a broader understanding of β-lactam hydrolysis. In the envisaged SFX experiments, OXA-10 crystals would be mixed with ertapenem, requiring the substrate/inhibitor to diffuse into the lattice. To ensure that diffusion of the substrate/inhibitor through the crystal lattice is faster than the catalytic turnover, small microcrystals (<5 μm) were required.[15,16] Data collection on such crystals benefits from a focused X-ray beam with high brilliance such as provided by XFELs.

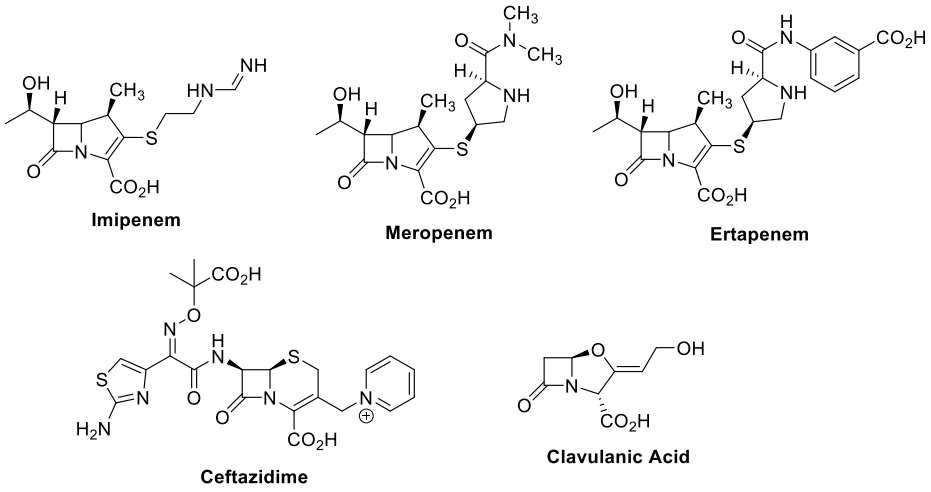



**Figure 3.3:** Structures of ertapenem, meropenem, ceftazidime, imipenem, and clavulanic acid.

## 3.2 Preparation of recombinant OXA-10

*A plasmid encoding for OXA-10 and protocols for expression and purification were provided by Dr Christopher Lohans (University of Oxford).*

Production of the recombinant form of *Pseudomonas aeruginosa* OXA-10 was done in *Escherichia coli* BL21 (DE3) cells, according to reported protocols using the pET22b expression vector.[17] In brief, recombinant OXA-10 in *E. coli* cells (BL21 DE3, kanamycin resistant) transformed with the appropriate plasmid were grown in 2TY media (24 × 600 mL, 14.4 L) for protein production. Cells were grown to an $OD_{600}$ of 0.6 before cooling to 18 °C, expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were incubated at 18 °C for 4 h. The cells were suspended in buffer and lysed by sonication. Subsequently, the protein was purified by SP Sepharose anion exchange chromatography, followed by Superdex™ 75 size exclusion column chromatography. The protein purity was assessed by SDS-PAGE analysis (Figure 3.4). A total of 690 mg of pure OXA-10 was obtained.

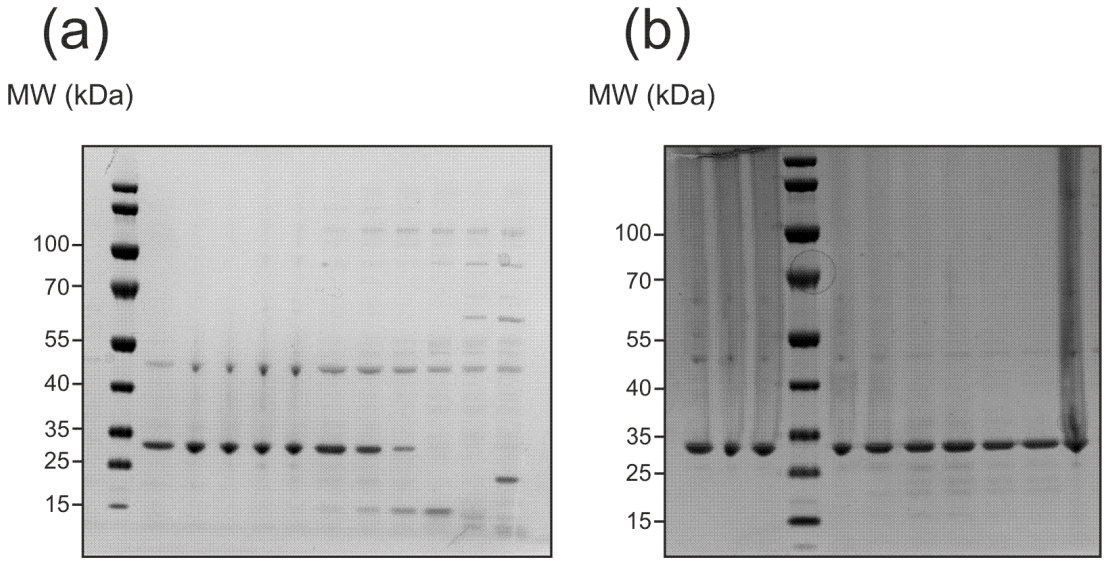



**Figure 3.4:** SDS-PAGE of *P. aeruginosa* OXA-10. (a) SDS-PAGE of OXA-10 fractions after SP sepharose anion exchange chromatography step; (b) SDS-PAGE of OXA-10 fractions after Superdex™ chromatography step. Molecular weight of OXA-10 is 28.6 kDa.

## 3.3 Time resolved X-ray crystallography

### 3.3.1 Crystallisation of OXA-10

Initial crystallisation conditions for OXA-10 were found by using the PACT premier broad screen. The conditions were further optimised (magnesium(II) formate (0 – 350 mM), and PEG 3350 (15 – 26% (v/v)) at room temperature (E8 conditions). Crystals of OXA-10 showed a predominantly needle morphology and appeared within 5 hours. Single macrocrystals (Figure 3.5 (a)) were obtained, cryo-cooled and analysed for diffraction using DLS macromolecular crystallography (MX) beamlines.

A structure of *apo*-OXA-10 was determined to 1.53 Å resolution, initial phases were obtained by molecular replacement using a reported *apo*-OXA-10 structure (PDB accession code: 5MMY)[14] as a search model. Compared to the reported *apo*-OXA-10 structures, the overall fold is similar with an RMSD of 0.42 Å over the peptide backbone (Figure 3.5 (b)). The active site Ser67 is observed in two conformations, while Lys70 is carbamylated, which is typical for catalytically active OXA enzymes.

**Table 3.1 Data collection and refinement statistics of initial refinements on OXA-10 and OXA-10 in complex with ertapenem**

| **Data set** | **OXA-10** | **OXA-10 • ertapenem** |
|---|---|---|
| Data collection | | |
| Beamline (Wavelength, Å**)** | DLS I03 (0.9795) | DLS I03 (0.9795) |
| Detector | PILATUS 6M-F | PILATUS 6M-F |
| Data processing | autoPROC 1.0.5 | Xia2 dials |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell dimensions | | |
| *a,b,c* (Å**)** | 48.8, 94.4, 126.6 | 48.8, 94.2, 125.8 |
| α,β,γ (°) | 90, 90, 90 | 90, 90, 90 |
| No of molecules/ASU | 2 | 2 |
| No. of reflections | 87297 (8614)[a] | 38961 (3838)[a] |
| Resolutions (Å) | 31.92 – 1.54 (1.60 – 1.54)[a] | 53.7 – 2.08 (2.15 – 2.08)[a] |
| $R_{merge}$ | | |
| I/δI | 1.34 | 1.33 |
| Completeness (%) | 99.9 (100)[a] | 97.5 (100)[a] |
| Multiplicity | | |
| Wilson B value (Å$^2$) | 21.4 | 28.7 |
| **Refinement** | | |
| $R_{work}/R_{free}$ | 0.1858 / 0.1982 | 0.2647 / 0.2907 |
| No. atoms | | |
| - Enzyme | 4052 | 3649 |
| - $H_2O$ | 405 | 163 |
| Avg. B-factors | | |
| - Enzyme | 29.7 | 49.1 |
| - $H_2O$ | 40.0 | 50.3 |
| R.m.s. deviations | | |
| - Bond lengths (Å) | 0.012 | 0.021 |
| - Bond angles (°) | 1.21 | 1.05 |

[a] Highest resolution shell in parentheses

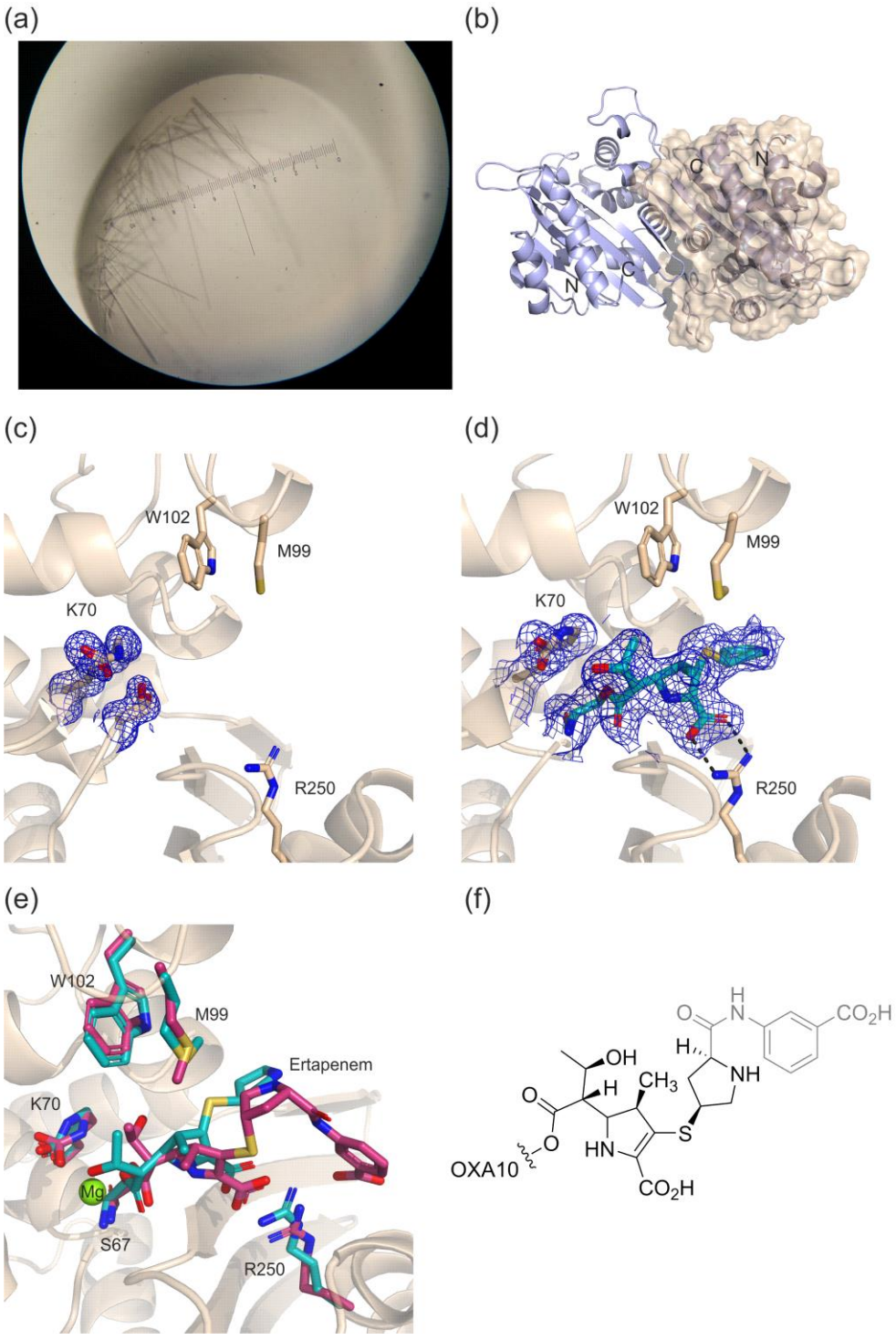



**Figure 3.5:** Crystals and structure of OXA-10 under optimised conditions. (a) Optical microscope image of OXA-10 crystals after optimisation (18% PEG 3350, 200 mM magnesium formate); (b) a view from the crystal structure showing the overall fold of *apo*-OXA-10 obtained from macrocrystals produced using E8 conditions; (c) view of the active site of *apo*-OXA-10 obtained from macrocrystals produced using E8 conditions; (d) structure of OXA-10 macro crystal obtained using E8 conditions, soaked with ertapenem (30 min); (e) OXA-10 structure soaked with ertapenem (cyan) overlaid with reported structure of OXA-10 with ertapenem (pink, PDB accession code: 6SKR); (f) structure of ertapenem at the active site of OXA-10, in black is highlighted what was modelled for ertapenem soak, and in grey is added what was modelled in the reported structure of ertapenem.

The macrocrystals of OXA-10 that were obtained, were used for soaking experiments with a range of antibiotics including: thiacephalosporin, cefminox, cefotetan, temocillin, cefnetazole, and ertapenem (Figure 3.6). Stock solutions of the antibiotics were prepared (100 mM stock, in well solution) and added to the wells containing OXA-10 macrocrystals (1 : 1). The crystals were soaked (1 – 60 min), cryo-cooled and data were collected at MX beamlines at DLS. The estimated solvent content in the *apo* OXA-10 crystal is 54%.[18–20] The high solvent content typically encountered in protein crystals, facilitates soaking of ligands into matured crystals. Analysis of reported ligand bound OXA-10 structures,[21] although obtained using different crystallisation conditions, show that the solvent exposed active site of OXA-10 is accessible to small molecule ligands through soaking of matured crystals. This strategy was considered particularly advantageous to quickly investigate different ligands for binding to OXA-10. A structure of OXA-10 in complex with ertapenem was determined to 2.1 Å resolution by molecular replacement using a published structure (PDB accession code: 5MMY)[14] as a search model (Figure 3.5 (c)). The unit cell dimension along the *c*-axis decreased by 0.8 Å in the new structure compared to the *apo* OXA-10. Changes in unit cell dimensions have previously been associated with binding of a ligand or a reaction taking place inside crystals.[22,23] After initial refinements the overall fold of the OXA-10 structure soaked with ertapenem was determined to be similar to the *apo*-OXA-10 structure with an RMSD of 0.24 Å over the peptide backbone. Inspection of the active site revealed additional electron density, confirming the acylation of Ser67 by ertapenem (Figure 3.5 (c)). Ertapenem also forms a salt bridge between its carboxylate group and Arg250 of OXA-10. The hydrolysed β-lactam ring, pyrroline ring, and pyrrolidine rings can be accounted for with clear electron density. In contrast, the amidobenzoic acid chain could not be modelled with sufficient electron density, likely due to mobility in this region. The electron density shows that carbapenem derived pyrroline ring displays (*S*)-$\Delta^1$ stereochemistry, rather than $\Delta^2$ pyrroline form observed in a deposited structure (PDB accession code: 6SKR, Figure 3.5 (d)). The overall fold of the ertapenem soaked OXA-10 structure and the deposited ertapenem bound OXA-10 however, are similar, with an RMSD of 0.39 Å over the peptide backbone. Active site residues Try102, Met99, Arg250, and Lys70, show similar conformations in both structures, and show Arg250 interacting with the pyrroline carboxylate. The reported structure of ertapenem bound OXA-10 contains a Mg(II) ion in the active site, in close proximity to the carbamylated Lys70 (2.9 Å) and the ester bond between Ser67 and ertapenem (3.0 Å). Consequently, the orientation of

ertapenem in the reported structure with Mg(II) and the obtained structure are different (Figure 3.5 (e)). The carbapenem C6 hydroxyethyl sidechain in the reported structure is oriented away from the Mg(II), forcing the pyrroline ring and further side chain to adopt an altered conformation compared to the obtained structure. The resolution of the ertapenem bound OXA-10 structure was lower than for the *apo* OXA-10 structure, possibly reflecting degradation of the crystals during the soaking process. Initial attempts with higher concentration of ertapenem (500 mM) were unsuccessful, resulting in crystals that did not show diffraction. Soaking of the OXA-10 crystals with ertapenem (100 mM) beyond 30 min similarly resulted in loss of diffraction. A more systematic approach with different concentrations could potentially improve the quality of the crystals and thus improve resolution. Alternatively, co-crystallisation of OXA-10 with ertapenem under optimised conditions might provide a higher resolution structure. However, as discussed above for the hydrolysis of ertapenem by OXA-10, tautomerisation and epimerisation of the hydrolysed β-lactam product can occur, complicating matters. Although details about the conditions used to obtain the reported structure are unknown, based on the proposed mechanism, it could be that the structure was obtained by a longer period of equilibration of ertapenem with the enzyme (e. g. extended soaking times or co-crystallisation), compared to the ertapenem bound OXA-10 structure reported in this study. Co-crystallisation conditions of ertapenem with OXA-10 would not be suitable for the planned tr crystallography experiments as the enzyme is highly active. The envisaged experiments relied on soaking of the substrate into the active site to initiate the reaction. Thus it was considered important to investigate the suitability of the OXA-10 crystals under the optimised conditions, for soaking with small molecules, in particular ertapenem.

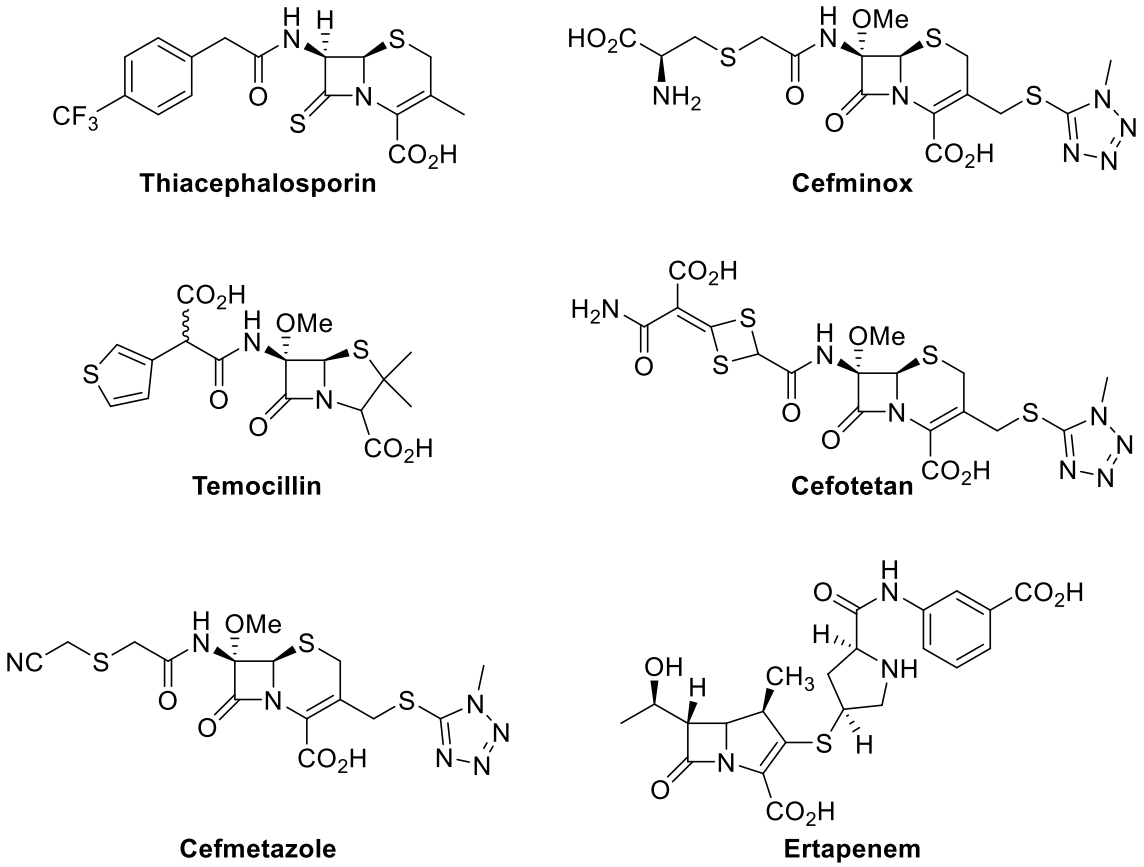



**Figure 3.6:** Structures of thiacephalosporin, cefminox, cefotetan, temocillin, cefmetazole, and ertapenem.

The optimised E8 OXA-10 macrocrystals were used to prepare seed stocks. The seed stocks were added to the crystallisation conditions to promote nucleation in small test batches (20 μL) in an attempt to optimise the production of microcrystalline slurries. Crystallisation of OXA-10 in small batches (20 μL) was optimised for the enzyme : precipitant ratio (2 : 1 to 1 : 20), and subsequently for the addition of seeds stocks (0 – 10% (v/v)). The enzyme : precipitant ratio had a great influence on the average crystal size (Figure 3.7 (a)). The quality and density of the seed stocks was found to be important as repeats of the same experiment with freshly prepared, dilute seed stocks did not yield the same quality (e. g. density, size) of crystal slurry. Therefore, each seed stock was tested on a small scale prior to use. Addition of seed stocks to the crystallisation conditions provided nucleation points, improving homogeneity between crystals and reducing their average size (50 – 100 μm in length). Microcrystal optimisation was subsequently moved to larger batch optimisation (100 μL, a 96 well plate). The average crystal size could to some extent be influenced by shaking of the crystal slurries in batch plates (500 – 1000 rpm, plate shaker). Shaking speeds exceeding 1000 rpm usually resulted in significant quantities of precipitation over time (24 h, Figure 3.7 (b),

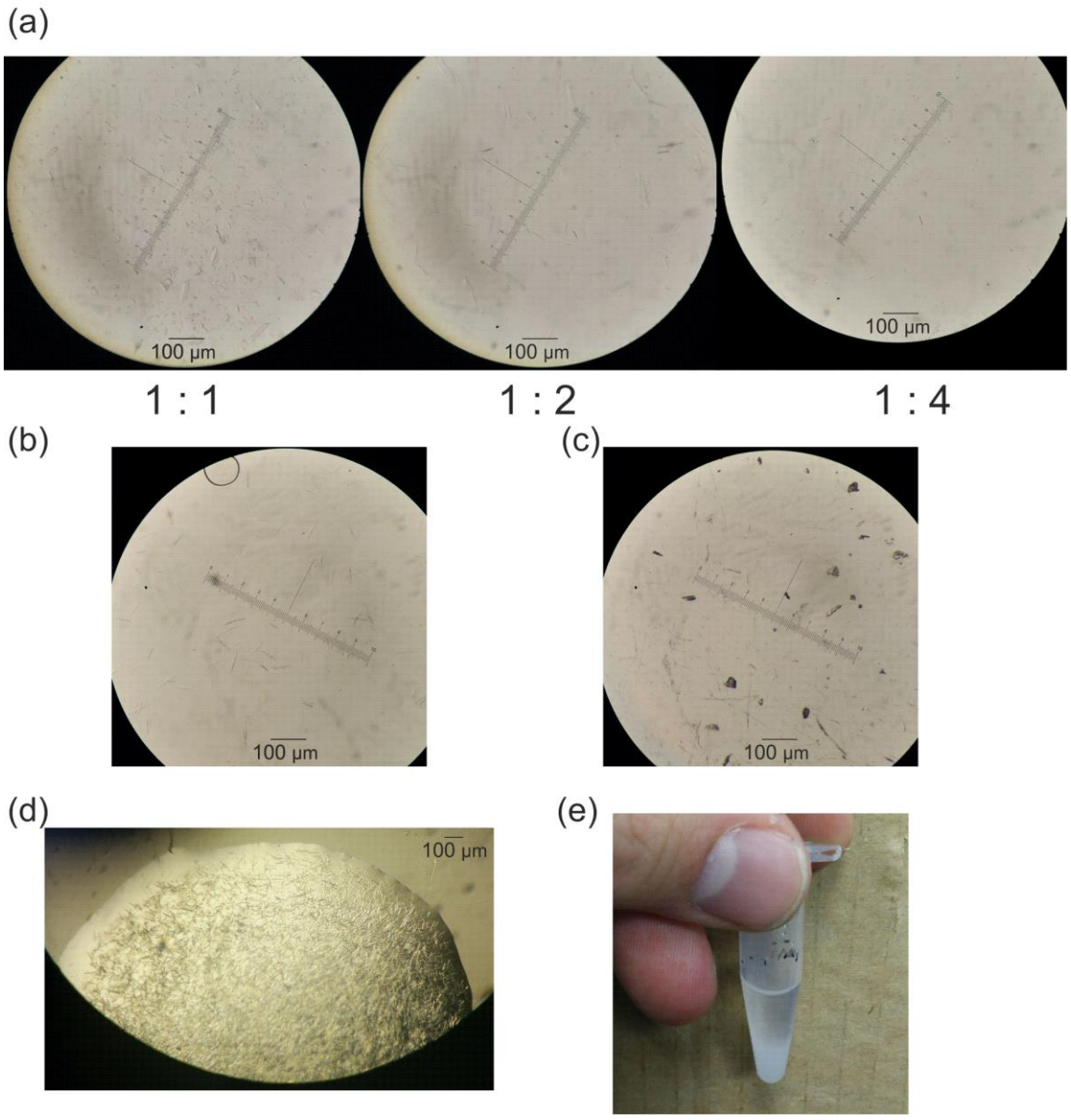



**Figure 3.7:** Optical microscope images of OXA-10 microcrystals and a picture of settled crystal slurry. Optical microscope image of (a) OXA-10 microcrystals obtained using different enzyme : precipitant ratios (1 : 1 – 4); (b) OXA-10 microcrystals obtained using E8 conditions shaken at ~900 rpm; (c) OXA-10 microcrystals obtained using E8 conditions shaken at ~1200 rpm; (d) concentrated OXA-10 microcrystal batch; (e) Eppendorf containing settled microcrystal slurry of OXA-10.

(c)). The slurries of OXA-10 microcrystals were concentrated by removing supernatant after pelleting of the microcrystals over 6 – 8 hours (Figure 3.7 (d)), until dense crystal slurries were obtained (Figure 3.7 (c)). The E8 OXA-10 microcrystals were tested for diffraction using microfocus beamline i24 at DLS. Testing of E8 OXA-10 microcrystals for diffraction was done at i24 DLS using a microfocus X-ray beam. During the experiment the microcrystal slurries were provided by use of a viscous extruder. OXA-10 microcrystal

slurries were prepared and mixed with lipidic cubic phase (LCP) and hydroxycellulose. Mixing of the crystal slurry with the greasy matrices was done using two connected syringes and moving the crystal slurry from one side to the other. Once mixed with a viscous matrix, the OXA-10 microcrystals were tested for diffraction using i24 microfocus beam at DLS. The OXA-10 microcrystal sample in LCP matrix appeared of comparable quality to the batch before mixing with matrix (albeit at lower concentration) based on visual inspection, however, no diffraction data could be obtained. The E8 conditions for crystallisation of OXA-10 were pushed forward for XFEL experiments.

### 3.3.2 XFEL experiment on OXA-10 at LCLS

In preparation of the XFEL experiment at LCLS, OXA-10 microcrystals were obtained on site using the optimised conditions (Section **3.3.1**) in a 96 well plate (100 μL per well). Microcrystals appeared within 3 hours and were left for an additional 24 hours. The slurries were allowed to settle over 8 hours and concentrated by removing the supernatant. After re-suspending, the OXA-10 microcrystals were analysed using an optical microscope to determine crystal dimensions and density (30 – 50 μm, by 5 – 8 μm; $7.35 \cdot 10^7$) (Figure 3.8 (a)). The crystal density was determined using a counting chamber grid (Figure 3.8 (b)). The dense crystal slurry was passed through a short piece of tubing (300 μm inner diameter) and loaded into a syringe (1 mL). The syringe was placed in a syringe pump and connected to tubing feeding to the DOD sample delivery system (Figure 3.8 (c)). To prevent the crystals from settling into a solid pellet, which could cause clogging of the system, the sample was rocked using a shaker (~30 s amplitudes, Figure 3.8 (c)). The beam was operated at 30 Hz, using 40 fs pulses at 9.5 keV. Data acquisition and collection for these serial crystallography experiments is different from traditional macrocrystals.

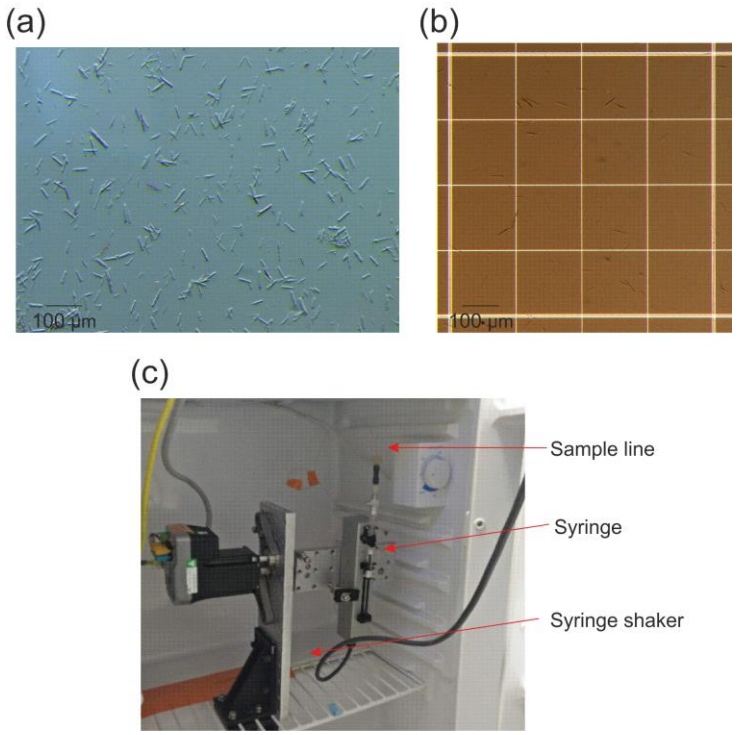



**Figure 3.8:** Optical microscope images of OXA-10 microcrystals obtained using E8 conditions and the crystal slurry sample in syringe connected to the DOD sample delivery system. (a) Optical microscope image of a dense OXA-10 microcrystal slurry; (b) optical microscope image of diluted (100 fold) OXA-10 microcrystal sample on a counting chamber (single quadrant shown) used to determine crystal density; (c) a fridge used to store syringes containing microcrystal slurries connected to a line (300 µm inner diameter) feeding the DOD sample delivery system.

### 3.3.3 Data acquisition and processing

*The DOD sample delivery system was operated by Dr Jan Kern, Dr Cindy Pham, Dr In-sik Kim, Dr Joshua Heinemann, Dr Philipp Simon, and Isabel Bogacz, LBNL, Berkeley. Data processing was performed by Dr Pierre Aller and Dr Agata Butryn from DLS, Didcot; Dr Aaron Brewster, Dr Kyle Sutherlin, Dr Asmit Bhowmick, and Dr Iris Young, LBNL, Berkeley.*

Acquisition of X-ray diffraction data for each SFX experiment was divided over different blocks, called runs. Each run contained an arbitrary number of images recorded over a fixed period of time, typically 5 minutes. The collected data was processed using the cctbx-xfel pipline[24] and provided immediate feedback about sample quality (e. g. resolution, hit rate, indexing rate; this was referred to as online data processing), and was processed and corrected for misalignments (e. g. detector distance/geometry) to eventually be merged to a final diffraction data set.

The DOD sample delivery system provides droplets of microcrystal samples on demand at the same rate as the data collection rate (i. e. 30 Hz).[25] The droplet hit rate was defined as the fraction of images derived from droplets that were hit by the X-ray beam out of the total number of images collected (Figure 3.9 (a)). Recorded images were analysed for diffraction patterns. The fraction of images that contain indexable diffraction data out of the total number of images is referred to as the indexing hit rate (idx). These statistical analyses of the sample were usually swiftly provided (~10 min after run completion) and guided the team during data acquisition.

When data collection was completed the corresponding runs were grouped together per sample and corrected for misalignments, before merging together to produce a final diffraction data set. Before any data acquisition the detector distance was accurately determined using silver behenate (Figure 3.9 (b)). Refinement of the detector geometry and distance allowed for compensation in deviations (Figure 3.9 (c)). The optimised detector geometries determined per run were used in an ensemble refinement to give the final data file. The experiments carried out using the DOD sample delivery system, used an ADE to dispense nL droplets on a Kapton belt. Due to the nature of this experimental setup, the Kapton belt casts a shadow on the detector, and thus the diffraction pattern during data acquisition (Figure 3.9 (d)). Consequently, parts of the diffraction pattern showed reduced intensity. The relative reduction in background in the shaded area compared to the rest of the image was used to compensate for the loss in intensity using a mask. In addition, a mask was generated to remove the area occupied by the beam stop (Figure 3.9 (d)).

The OXA-10 microcrystal slurry was supplied in a syringe and dispensed at 30 Hz by the DOD system. The belt speed was set so that the droplets would be on the belt for 500 ms in total before being probed by the XFEL beam. At first, data was collected to obtain a resting state structure of OXA-10, showing diffraction (~2 Å) (Figure 3.9 (a)). However, only a few of the diffraction images were indexed (~3% on average over multiple runs, Figure 3.9 (a)), suggesting that multiple lattices were hit during data acquisition. Overlaying spots prevented indexing of the images. Different batches of OXA-10 with varying crystal density were tested, resulting in the same outcome. The results of these combined experiments suggest that the problem with multiple lattices caused by defects in the crystals, is likely inherent to the crystallisation conditions. This led to the exploration of different crystallisation conditions.

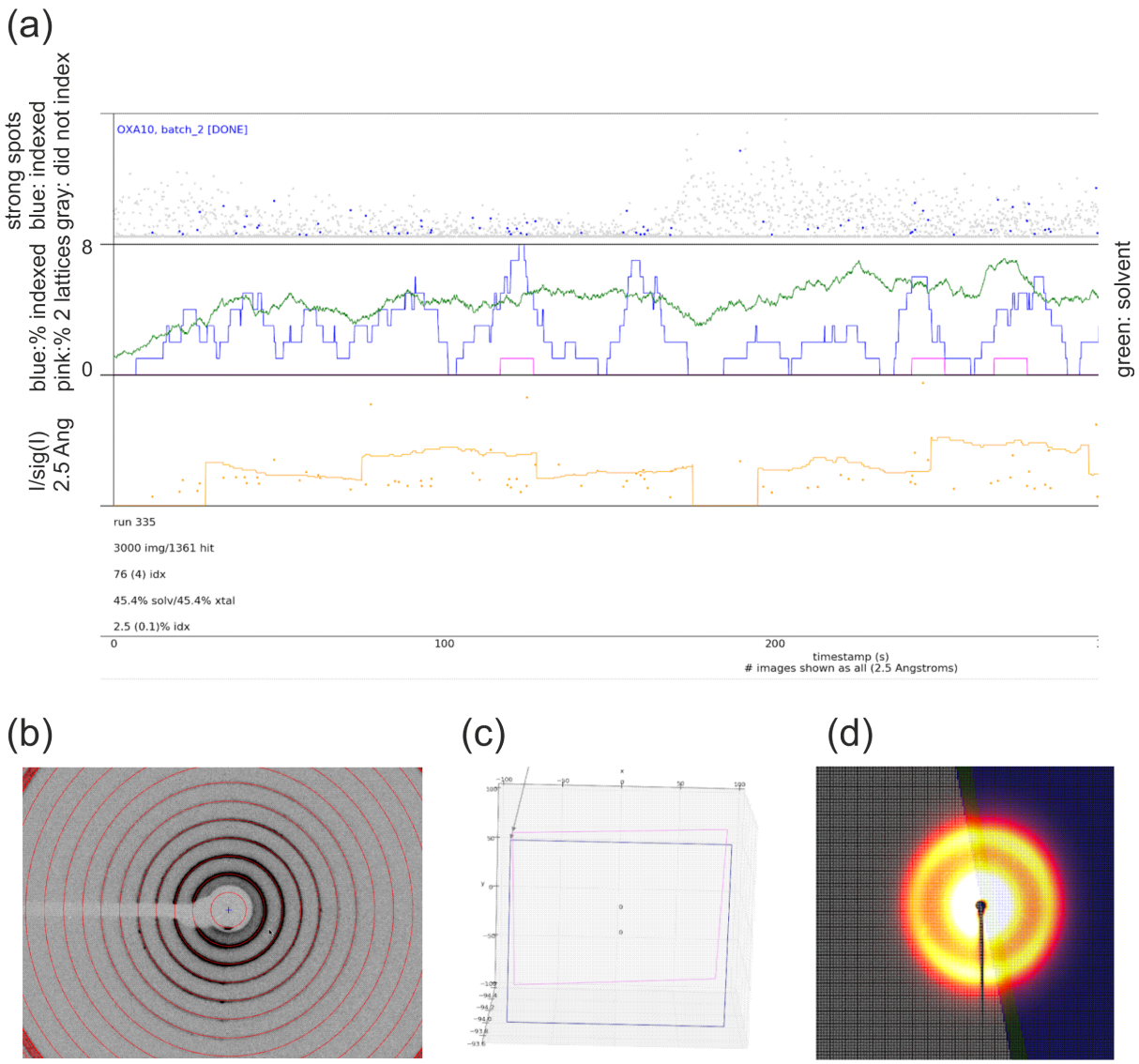



**Figure 3.9:** Overview of different aspects of the processing of serial crystallographic data obtain from microcrystals. (a) Statistical analyses of data collected on OXA-10 microcrystals grown under optimised E8 conditions. Data was collected in separate runs. Top panel: Blue spots were indexed spots, grey spots were not indexed; middle top panel: green line represents images identified to containing solvent (%), blue line represents the number of images containing spots that are indexed (%), pink line represents images containing multiple lattices (%); middle bottom: spot intensity; bottom panel: summary of data with run numbers. (b) Silver behenate diffraction pattern used to calibrate the detector distance; (c) detector geometry before (blue) and after (pink) geometry refinement; (d) position of the shadow (highlighted in black) caused by the Kapton belt on the X-ray diffraction detector.

### 3.3.4 Search for different crystallisation conditions

Although the OXA-10 microcrystals grown under the optimised E8 conditions (Section **3.3.2**) demonstrated diffraction, the low number of images that could be integrated prevented data collection and further use of these conditions. It is possible that the OXA-10 microcrystals that were produced suffered from defects in the crystal lattice, which is difficult to observe under an optical microscope. It was suggested that the rapid crystallisation (<3 hours) might have had a negative effect on the quality of the crystals. In order to address this problem, different crystallisation conditions for OXA-10 were explored. Crystallisation trials were performed with commercially available broad screens (Molecular Dimensions) including the JCSG+, PACT premier, and SaltRX. Many of the tested conditions yielding OXA-10 crystals, contained considerable amounts of PEG. At a later stage it was found that the high viscosity caused by PEG slows down diffusion rates for small molecules, which might become problematic for diffusion of substrates into the OXA-10 active site. Conditions were selected based on the absence of particular compounds such as certain buffers or anions (i. e. HEPES, MOPS or halide ions,[14,26]), as these have been reported to bind to the OXA-10 active site. In addition, established crystallisation conditions for OXA-10 at 4 °C (*provided by Dr Jürgen Brem, University of Oxford*), as well as published conditions were selected for optimisation.[13,27–29] The crystallisation conditions obtained for OXA-10 from the broad screens (Table 3.2, entries 2 – 7) were tested for diffraction at the MX beamlines of DLS. Selection criteria for batch optimisation were based on resolution of the obtained OXA-10 macrocrystals (≤ 2.1 Å). Optimisation was done similar to the procedure described above (Section **3.3.2**); after precipitant selection (Table 3.2), the enzyme : precipitant : seed stock ratios were optimised (Table 3.2). The optimised conditions were tested for diffraction using microfocus beamline i24 at DLS.

**Table 3.2 Crystallisation conditions for OXA-10.**

| Entry | Condition name | Optimisation ranges | Resolution (Å)[a] | Optimised ratio enz : precip : seed |
|---|---|---|---|---|
| 1 | E8 | $Mg(HCO_2)_2$ (0 – 350 mM), PEG3350 (16 – 25%) | 1.53 | 1 : 4 : 0.5 |
| 2 | A8 | $NH_4HCO_2$ (0 – 350 mM), PEG3350 (16 – 25%) | 1.45 | 1 : 6 : 1 |
| 3 | A10 | $KHCO_2$ (0 – 350 mM) PEG3350 (16 – 25%) | 1.9 | 1 : 8 : 0.5 |
| 4 | C4 | HEPES (0.1 M, pH 7.0), PEG6000 (5 – 20%) | 2.1 | 1 : 2 : 0.33 |
| 5 | C12 | PEG1000 (10%) | 2.3 | - |
| 6 | G4 | KSCN (0.2 M), Bistris (0.1 M, pH 7.5), PEG3350 (16 – 25%) | 2.0 | 2 : 1 : 0.33 |
| 7 | H6 | $NH_4OAc$ (0.1 M), Bistris (0.1 M, pH 5.5), PEG10,000 (5 – 20% | 2.0 | 1 : 4 : 0.25 |
| 8 | 144A[29] | NaCl (0.2 M), NaOAc (0.1 M, pH 5.0), PEG6000 (20%) | 1.5 | - |
| 9 | 144B[13] | $(NH_4)_2SO_4$ (1.8 M), MES (0.1 M, pH 6.5) | 1.7 [b] | - |
| 10 | 144C[28] | $(NH_4)_2SO_4$ (2.0 M), Tris (0.1 M, pH 8.1), $NaHCO_3$ (20 mM) | 1.6 [b] | - |
| 11 | 144D[27] | $(NH_4)_2SO_4$ (2.0 M), Tris (0.1 M, pH 8.5) / HEPES (0.1 M pH 7.5) | 1.6 | 1 : 9 : 1 |
| 12 | 4 °C | $Mg(HCO_2)_2$ (200 mM), PEG3350 (20%) | 1.8[b] | 1 : 4 : 0.125 |

[a] Highest resolution obtained from macrocrystals under cryo cooled conditions.
[b] Single crystals were obtained after addition of 10% glycerol

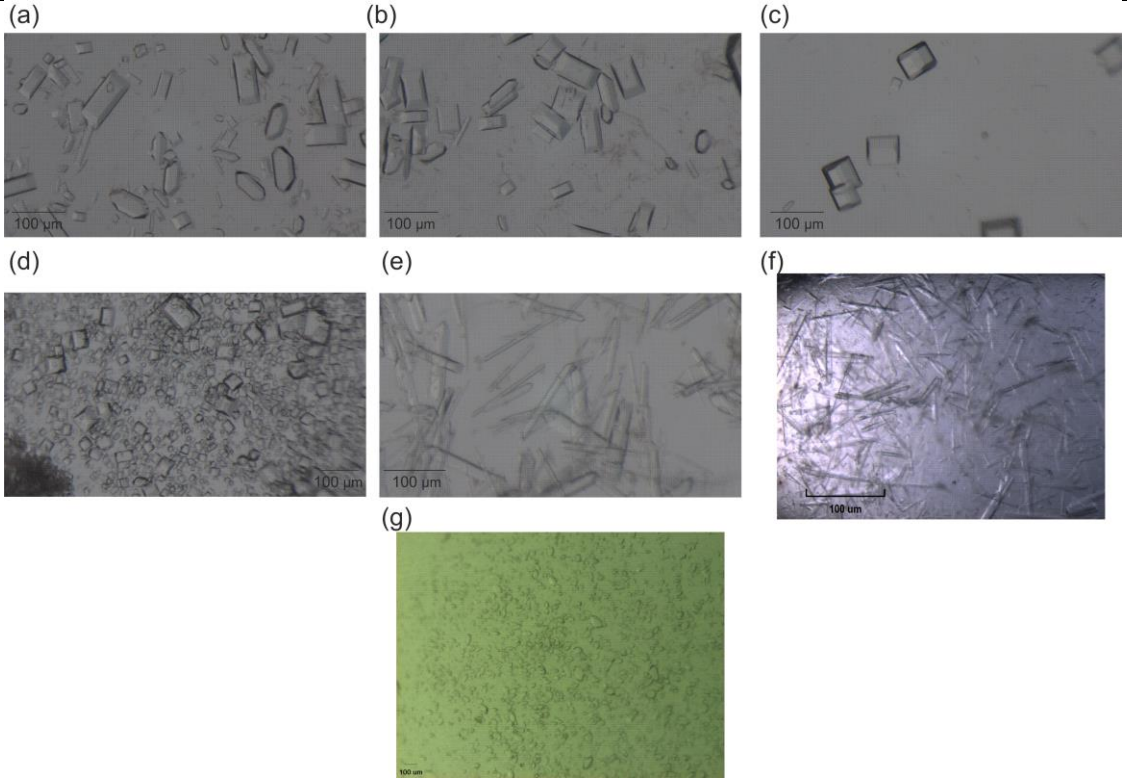



**Figure 3.10:** Optical microscope images of OXA-10 microcrystals. (a) Microcrystals of OXA-10 using optimised A8 conditions; (b) optimised A10 conditions; (c) optimised C4 conditions; (d) optimised G4 conditions; (e) optimised H6 conditions; (f) optimised 144D conditions; (g) optimised 4 °C conditions.

### 3.3.5 Fixed target experiment at DLS

*Data acquisition and processing were performed by Dr Pierre Aller, Dr Agata Butryn, and beamline scientists at DLS, Didcot.*

In order to test the different conditions that gave OXA-10 microcrystals, fixed target experiments using the microfocus beam at i24 at DLS were performed. The fixed target approach developed by Owen *et al.*,[30,31] uses a chip containing 25,600 conically shaped wells. For sample preparation (*in collaboration with Dr John Beale*), the crystal slurry was transferred into a humidified tent (>80%) to avoid the crystals from drying out. Crystal slurry was pipetted on top of the chip and a mild vacuum was applied to remove the solvent and trap the crystals in conically shaped wells. The chip was then sealed with a Mylar film and transferred to the beamline. The chip was aligned and data was subsequently collected. Online data processing provided direct feedback on sample quality including statistics on: hit rate, indexing rate, unit cell distribution, and resolution.

The E8 conditions developed for XFEL were further evaluated using the fixed target experiments at DLS. Two batches of E8 microcrystals were prepared, one with seed stock and a second batch without the addition of seed stocks, to investigate if the seeding had any effect on the quality of the crystals. The average size of OXA-10 crystals (~100 μm) was larger than the size of the crystals during the XFEL experiment. Samples created from both batches showed diffraction (2.9 Å). However, both samples suffered from a low number of indexed images (3.4 % and 1.5%), confirming that the problem with multiple lattices may have been inherent to the crystallisation conditions. The A8 conditions were tested using the fixed target approach and diffracted to 2.2 Å. Fortunately, the OXA-10 microcrystals grown using the optimised A8 conditions resulted in data with reasonable indexing rates (17 – 28%). Processing of the data suggested that the images were indexed with two different unit cell parameters (Figure 3.11 (c)). The H6 conditions were also tested, using the optimised conditions with and without seed stocks added. Crystals in both samples exhibited the same cubic morphology, but the crystals in the non-seeded conditions were larger than in the seeded conditions, judged by visual inspection of both OXA-10 H6 microcrystal slurries. Both non-seeded and seeded H6 OXA-10 micro crystal batches showed diffraction (3.8 Å, and 2.3 Å respectively). The hit rates for the non-seeded batch (2.9%) were considerably lower than for the seeded batch (9.6%). OXA-10

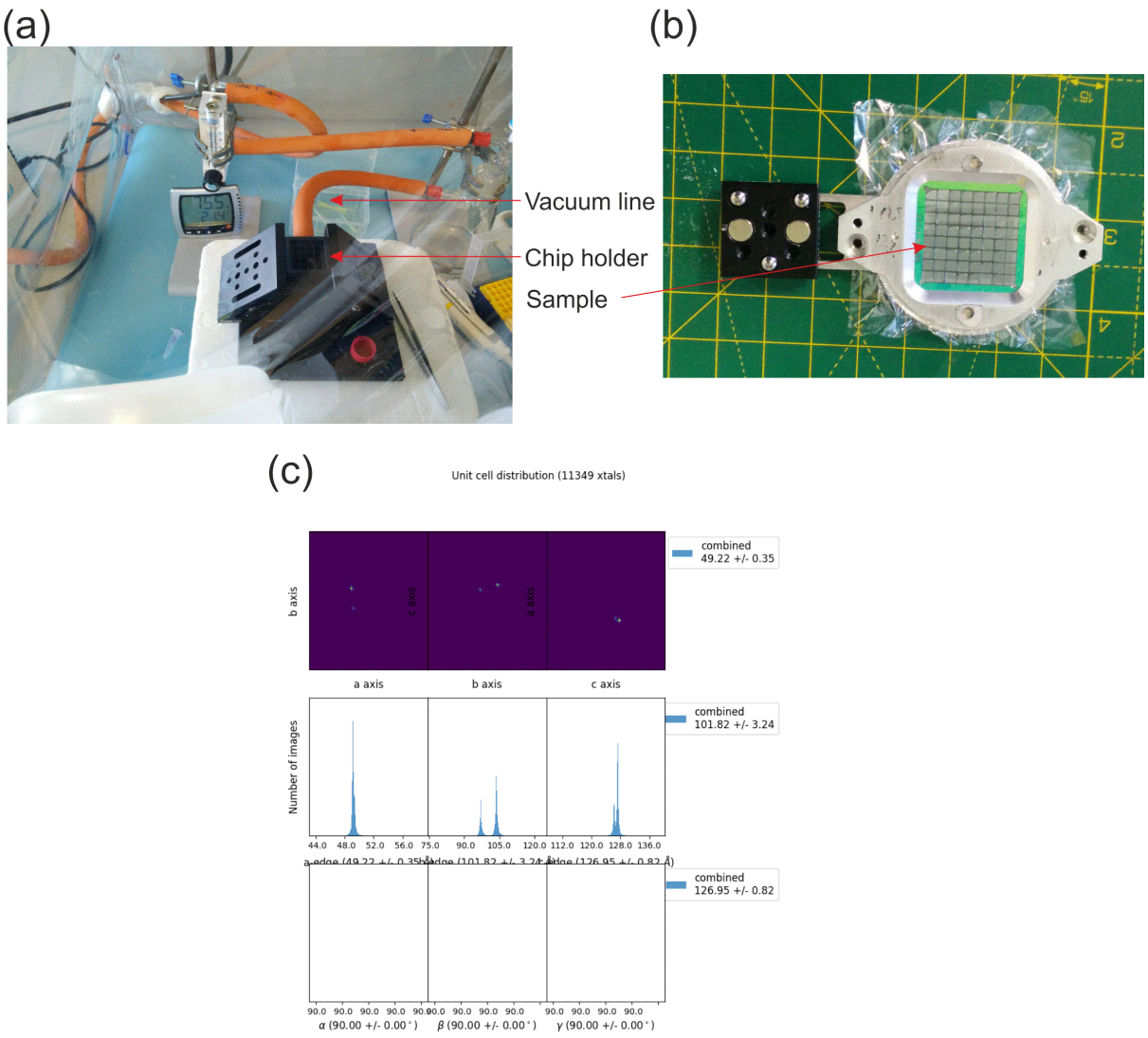



**Figure 3.11:** Fixed target experimental setup for preparation of a chip and OXA-10 data analysis. (a) Chip holder connected to vacuum line inside humidifying chamber; (b) chip containing OXA-10 microcrystals, sealed with Mylar; (c) analysis of the unit cell parameters of OXA-10 microcrystals produced using A8 conditions.

microcrystals produced using C4 conditions, showed diffraction (2.3 Å). The C4 OXA-10 microcrystal sample showed reasonable hit and indexing rates (17 – 30%), with a single set of unit cell parameters. In agreement with prior observations, the results of the fixed target experiment suggest that the E8 crystallisation conditions were unsuitable for OXA-10, presumably due to problems during crystal growth. The results obtained using the A8 conditions for the crystallisation of OXA-10 microcrystals, suggested a reasonable hit and indexing rate. However, the obtained images were indexed into different unit cells. Although data analyses were able to distinguish between the two different unit cells, data collection might require more time and could cause complications when increasing the complexity of the experiment

(e. g. soaking of antibiotics). Both the H6 and C4 conditions for the crystallisation of OXA-10 provided diffracting samples (2.3 Å), with reasonable hit rates (9.6% and 17 – 30% respectively). Thus, these conditions were selected for future experiments.

### 3.3.6 A room temperature fixed target structure of OXA-10

A fixed target room temperature structure of *apo*-OXA-10 was determined to 2.3 Å resolution by molecular replacement using a published structure (PDB accession code: 5MMY)[14] as a search model (Figure 3.12 (a)). The unit cell of OXA-10 contains two molecules of enzyme, as observed with the structure of OXA-10 obtained from macrocrystals using E8 conditions. The overall folds of OXA-10 in the macrocrystal structure and room temperature fixed target structure after initial refinements are similar, with an RMSD of 0.29 Å over the peptide backbone. As expected, Lys70 at the active site of OXA-10 was carbamylated (Lys70, Figure 3.12 (b)).

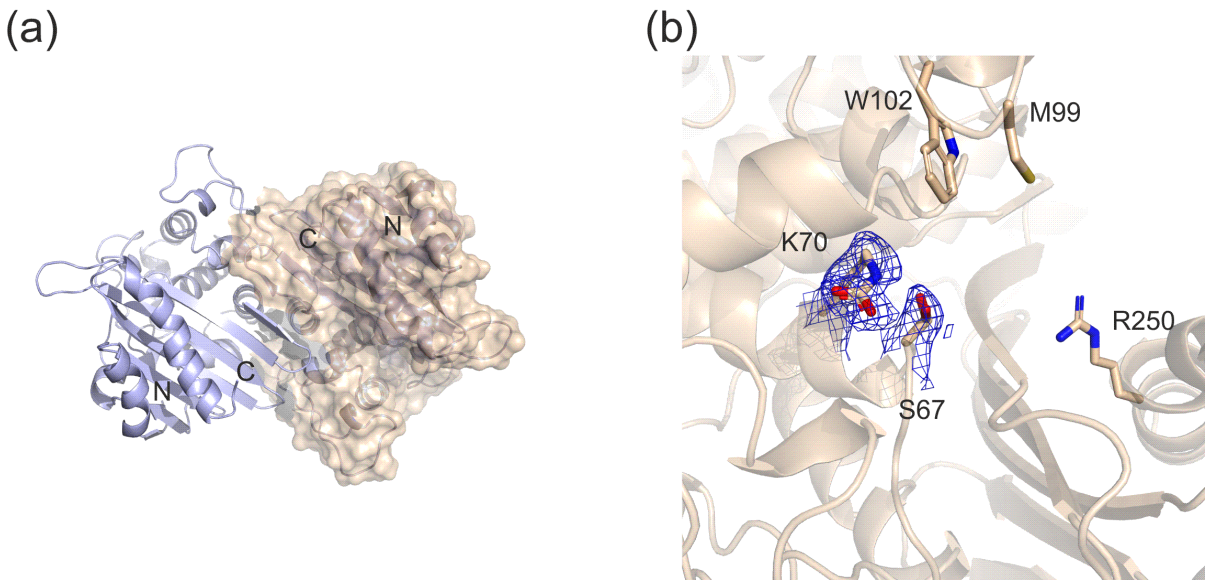



**Figure 3.12:** Fixed target structure of *apo*-OXA-10 form microcrystals using H6 microcrystals.

| **Table 3.3 Data collection and refinement statistics of OXA-10.** | |
|---|---|
| **Data set** | **OXA-10** |
| Data colleection | |
| Beamline (Wavelength, Å**)** | DLS i24 |
| Detector | Pilatus3 6M |
| Data processing | Dials |
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| *a,b,c* (Å**)** | 49.16 103.81 127.39 |
| α,β,γ (°) | 90, 90, 90 |
| No of molecules/ASU | 2 |
| No. of reflections | 29754 (2921)[a] |
| Resolutions (Å) | 80.47 – 2.3 (2.38 – 2.3)[a] |
| $R_{split}$ | 39.5 (51.6)[a] |
| I/δI | 31.13 (4.80)[a] |
| CC1/2 | 73.4 (40.3)[a] |
| Completeness (%) | 100 (100)[a] |
| Multiplicity | 34.43 (11.16)[a] |
| Wilson B value (Å$^2$) | 32.39 |
| **Refinement** | |
| $R_{work}/R_{free}$ | 0.2524 / 0.2717 |
| No. atoms | |
| - Enzyme | 3753 |
| - $H_2O$ | 111 |
| Avg. B-factors | |
| - Enzyme | 42.58 |
| - $H_2O$ | 42.18 |
| R.m.s. deviations | |
| - Bond lengths (Å) | 0.002 |
| - Bond angles (°) | 0.51 |
| [a] Highest resolution shell in parentheses | |

### 3.3.7 NMR studies on viscosity and stability of β-lactam compounds

*Stability assays were done in collaboration with Pauline Lang (University of Oxford)*

As part of the tr-SFX experiments with β-lactamases, including OXA-10, incubation of microcrystals with antibiotics and inhibitors was planned. The list of substrates included: ertapenem, imipenem, ceftazidime, and clavulanic acid. Sample delivery during the XFEL experiment was planned to be done using the DOD system equipped with a PolyPico dispenser.[32] Microcrystal slurries of the β-lactamases were dispensed using the ADE system, while the PolyPico system dispenses pL droplets of concentrated solutions of the β-lactam containing compounds (i. e. ertapenem, imipenem, ceftazidime, and clavulanic acid). The PolyPico system is capable of dispensing droplets at a high repetition rate (up to 37,000 Hz) and should enable for the addition of multiple droplets to the microcrystal droplets. As the droplets launched from the PolyPico hit the microcrystal containing droplets, turbulence is caused and facilitates diffusion of the compound throughout the droplet. Cartridges of ~1.5 mL were supplying the PolyPico system and would contain compound for an extended period of time.

To investigate the stability of the compounds under these conditions, $^1$H NMR time course experiments were run using ertapenem, imipenem, ceftazidime, and clavulanic acid (Figure 3.13). Ceftazidime (100 mM) and clavulanic acid (100 mM) were readily soluble in $D_2O$ and did not show appreciable decomposition at room temperature under the tested conditions. Clavulanic acid was supplied as a solid in cellulose (1 : 1) and after dissolving in water and filtering, produced an acidic solution (pH 5 – 6). Ertapenem (100 mM) was similarly soluble in $D_2O$ and showed minor decomposition (~8%) over 6 hours at room temperature. Imipenem (50 mM) was less soluble in $D_2O$ and showed considerable decomposition over the 6 hour time period (39%). These results suggest that dispensing of imipenem might lead to the incubation of carbapenem hydrolysis product to the β-lactamase sample, instead of the intact β-lactam. Thus, to avoid incubation with hydrolysed products, the cartridge containing compound should be exchanged frequently, especially for imipenem.

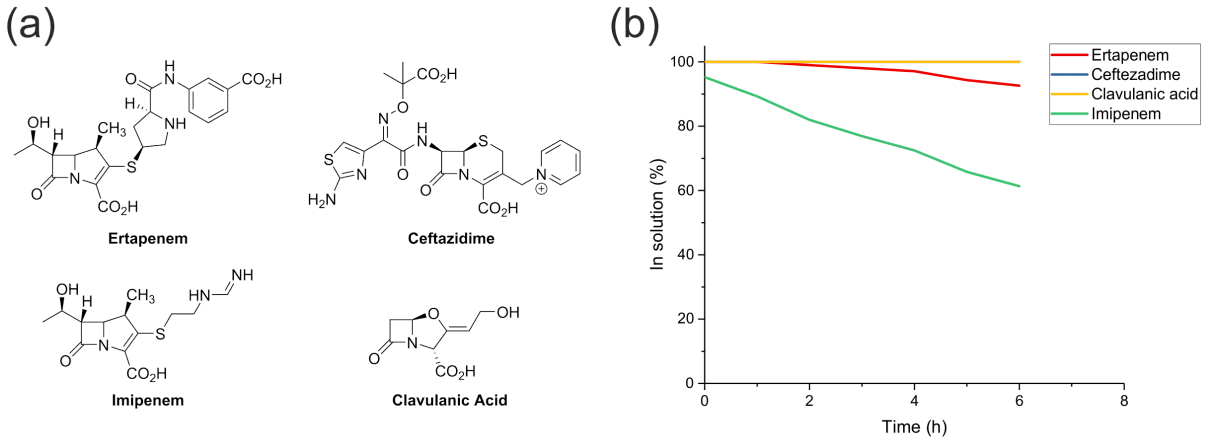



**Figure 3.13:** Structures of ertapenem, ceftezadime, clavulanic acid and imipenem and plot of their stability followed by $^1$H NMR (600 MHz). (a) Structures of ertapenem, imipenem, ceftazidime, and clavulanic acid. Conditions were as follows: Ertapenem (100 mM), ceftazidime (100 mM), clavulanic acid (100 mM), or imipenem (50 mM) were dissolved in $D_2O$ and their stability was followed by $^1$H NMR at room temperature (25 °C).

From experiments run in parallel (i. e. with IPNS) it became apparent that viscosity of the crystallisation conditions, might affect the diffusion of the compounds in the droplet. Thus, the diffusion properties of the different β-lactam compounds were investigated under various conditions. The tested compounds included ertapenem, imipenem, ceftazidime, and clavulanic acid. From these compounds ertapenem is clinically the most relevant. Ertapenem is a carbapenem antibiotic, a class of β-lactam antibiotics less prone to resistance. Additionally, the solubility and stability of ertapenem in water are favourable. Thus, further investigations focused on the diffusion properties of ertapenem. As expected, the diffusion of ertapenem was temperature and concentration dependent (Table 3.4, entry 1 – 3). Upon addition of $(NH_4)_2SO_4$ (1 M and 2 M) a linear decrease in diffusion coefficient of ertapenem was observed (3.32 to 2.64 and 2.13 · $10^{-10}$ $m^2 \cdot s^{-1}$, 21% and 36%, Table 3.4, entry 4 and 5). A similar change was observed with the addition of 5 – 10% (w/v) of PEG3350 (2.59 and 2.24 ·$10^{-10}$ $m^2 \cdot s^{-1}$ respectively, 23% and 33%, Table 3.4, entry 7 and 8). Further increases in PEG3350 content (20% (w/v)) led to further reduction in diffusion coefficient (1.78 · $10^{-10}$ $m^2 \cdot s^{-1}$, Table 3.4, 47% entry 9). The use of longer chain PEG10,000 (10% (w/v)) resulted in a reduction with the diffusion coefficient of ertapenem (3.32 to 2.54 · $10^{-10}$ $m^2/s$, 24%, Table 3.4, entry 10), however not as stark as the reduction observed for shorter chained PEG3350 at the same concentration (2.24 · $10^{-10}$ $m^2$ s, Table 3.4, entry 8). Diffusion coefficient for ceftazidime (3.65 ·$10^{-10}$ $m^2 \cdot s^{-1}$, Table 3.4, entry 11) was found to be comparable to the diffusion coefficient of ertapenem (3.32 ·$10^{-10}$ $m^2 \cdot s^{-1}$, Table 3.4, entry 2) under the same conditions. Clavulanic acid and imipenem displayed higher diffusion coefficients than ertapenem (5.59 and 4.42 · $10^{-10}$ $m^2 \cdot s^{-1}$ respectively, Table 3.4, entry 12 and 13).

**Table 3.4 NMR experiments on ertapenem in crystallisation buffer**

| Entry | Compound | Conditions[a] | Diffusion coefficient ($\cdot 10^{-10}$ $m^2 \cdot s^{-1}$)[b] |
|---|---|---|---|
| 1 | Ertapenem[c] | $D_2O$ | 4.60 |
| 2 | Ertapenem | $D_2O$ | 3.32 |
| 3 | Ertapenem[d] | $D_2O$ | 2.87 |
| 4 | Ertapenem | $D_2O$ + $(NH_4)_2SO_4$ (1 M) | 2.64 |
| 5 | Ertapenem | $D_2O$ + $(NH_4)_2SO_4$ (2 M) | 2.13 |
| 6 | Ertapenem | $D_2O$ + PEG3350 (2.5 v/v%) | 2.82 |
| 7 | Ertapenem | $D_2O$ + PEG3350 (5.0 v/v%) | 2.59 |
| 8 | Ertapenem | $D_2O$ + PEG3350 (10 v/v%) | 2.24 |
| 9 | Ertapenem | $D_2O$ + PEG3350 (20 v/v%) | 1.78 |
| 10 | Ertapenem | $D_2O$ + PEG10,000 (10 v/v%) | 2.54 |
| 11 | Ceftazidime | $D_2O$ | 3.65 |
| 12 | Clavulanic acid | $D_2O$ | 5.59 |
| 13 | Imipenem | $D_2O$ | 4.42 |

a) Standard conditions: compound (10 mM) in $D_2O$ (>98% $^2H$), recorded at 298 K (25 °C)

b) Derived from line fitting using non-overlapping resonances (Section 3.9)

c) 37 °C

d) Ertapenem at 50 mM

### 3.3.8 XFEL experiment on OXA-10 at SACLA

Based on the experiments described above, OXA-10 (200 mg), crystallisation conditions A8, C4, H6, G4, 4 °C, and 144D and their corresponding seed stocks were used for XFEL experiments performed at SACLA. On site optimisation of the conditions led to the successful production of OXA-10 microcrystals using A8, H6, and 144D conditions. The A8 condition provided a dense crystal slurry of average sized OXA-10 microcrystals (24 μm average, 53 μm maximum size, at $1.9 \cdot 10^8$ crystals per mL). Samples prepared with the H6 conditions suffered from low concentration, and limited sample volume. OXA-10 microcrystals obtained using the 144D conditions were produced from hanging drop vapour diffusion rather than the optimised batch conditions. Attempts to grow crystals in batch did not result in dense crystal slurries, even after optimisation on site. A limited sample volume was available (370 μL, 30 – 50 μm in size, $7.35 \cdot 10^7$ crystals per mL). Attempts to grow C4, and H6 conditions did eventually afford crystals, however,

a low crystal density and insufficient quantity made the samples unsuitable for XFEL studies. OXA-10 microcrystals could not be obtained using the optimised G4 and 4 °C conditions, even after optimisation. This might have been caused by thawing of the seed stocks derived from microcrystals grown using the 4 °C conditions during transport.

Data acquisition on the A8 OXA-10 microcrystals showed low resolution diffraction (>2.8 Å), and a low hit rate (~1%). A total of 298 images were collected, which is an insufficient number for a useful data set. However, analysis of the obtained images showed a similar distribution in unit cells as was observed during fixed target synchrotron experiments. Data acquisition on OXA-10 microcrystals obtained from the 144D conditions similarly suffered from a low number of indexed images. Interestingly, when soaking the crystal slurry with ertapenem, a change in unit cell dimensions was observed (Figure 3.14). Only a limited number of images were recorded before the experiment ran out of sample. It must be noted that the temperature near the DOD sample delivery system reached 36 – 37 °C, and that cooling of the sample line was not always sufficient. This could have affected the quality of the tested microcrystal slurries.

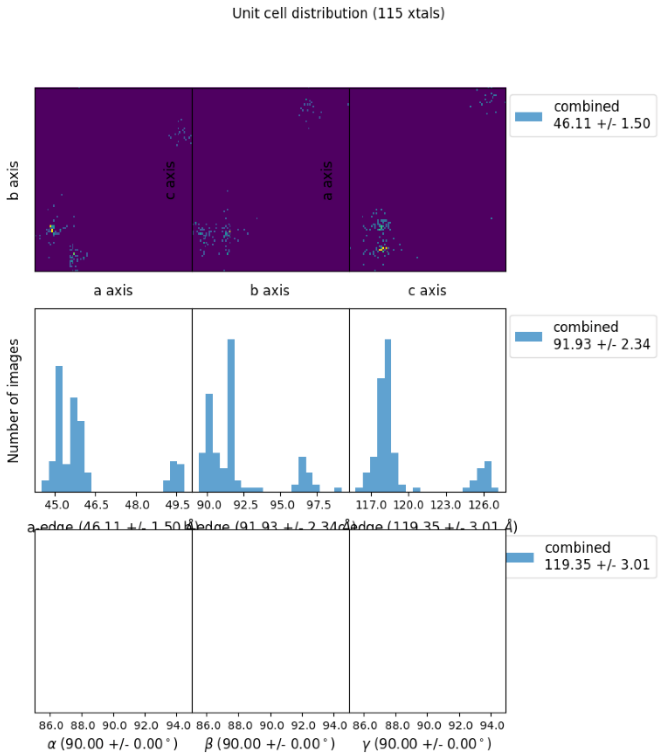



**Figure 3.14:** Plotted analyses of the distribution of unit cell parameters after incubation of 144D OXA-10 microcrystal slurries with ertapenem. *Image was prepared by Dr Pierre Aller and Dr Agata Butryn from DLS, Didcot.*

## 3.4 Discussion

This chapter describes work on the production and purification of OXA-10 and its subsequent crystallisation under a range of conditions. Macrocrystals of OXA-10 were obtained using optimised E8 conditions and probed for diffraction (1.53 Å) using synchrotron radiation. Soaking of the E8 OXA-10 macrocrystals with ertapenem demonstrating that an acyl enzyme complex could be obtained. Subsequently, these conditions were optimised for the production of E8 OXA-10 microcrystal slurries and tested for diffraction at DLS in a viscous extruder experiments, using LCP and hydroxycellulose as matrices. Visual inspection of the crystal slurries after mixing with either of the matrices suggested that the samples were of comparable quality to the microcrystal slurries, however, diffraction of the crystals was not observed. It was unclear to what extent the matrices affected the diffraction quality of the OXA-10 microcrystal slurries. The microcrystals of OXA-10 obtained using E8 conditions showed reasonable diffraction (~2 Å) during SFX experiments at LCLS. However, a low number of diffraction patterns could successfully be integrated from the data that was collected, suggesting that multiple lattices were obtained per image. Thorough testing of the E8 conditions was repeated during fixed target experiments at DLS with E8 OXA-10 microcrystals from different batches and at varying concentrations. Again, diffraction of the microcrystals was observed and multiple lattices were hit, suggesting that the problem was inherent to the conditions used for the production of the E8 OXA-10 microcrystals. It is possible that the crystals suffered from deformations such as twinning or polymorphism, in which case different domains of the crystal are joined together resulting in superimposed diffraction patterns. These overlapping diffraction patterns can lead to indexing difficulties. In response to these observations, different crystallisation conditions for OXA-10 were explored.

After identifying various crystallisation conditions from broad screens including JCSG+, PACT premier, and SaltRX, conditions were selected and optimised. Microcrystals of OXA-10 obtained in optimised A8 conditions were also tested for diffraction using fixed target experiments at DLS. Diffraction to 2.2 Å was observed, with a reasonable hit rate (17 – 28%); however, the images obtained could be indexed with two different unit cell dimensions, suggesting the sample was non-isomorphous. Although it is possible to index images from both isoforms separately, this was considered less suited for future mixing experiments. The A8 OXA-10 microcrystal slurries, alongside other conditions discussed below, were carried forward for

XFEL experiments. At SACLA a DOD system was used equipped with a PolyPico to dispense β-lactam containing compounds to the microcrystal slurry droplets. Data collection on A8 OXA-10 microcrystal slurries was complicated by low resolution diffraction (>2.8 Å) and a low indexed hit rate (~1%). This was in stark contrast with the fixed target experiments. During the experiment at SACLA, the OXA-10 microcrystal slurries were exposed to warm temperatures (36 – 37 °C) while in the lines and notably while on the Kapton tape of the DOD system. This could have affected the quality of the crystals during the experiment.

Microcrystal slurries of OXA-10 obtained from optimised H6 and C4 conditions (Table 3.2) were tested for diffraction using fixed target experiments at DLS, showed reasonable diffraction (2.3 Å) and hit rates (9.6, and 17 – 30% respectively). A room temperature fixed target structure of OXA-10 using the H6 conditions was obtained (2.3 Å, Figure 3.12). The structure showed high similarity to reported structures obtained from cryo-cooled macrocrystals, with an RMSD of 0.29 Å over the peptide backbone.[5,6] The H6 and C4 conditions, as well as G4 and 4 °C conditions (Table 3.2) were similarly carried forward for XFEL experiments. Preparation of the microcrystal slurry samples immediately prior and during the beamtime failed to provide sufficient quantities of slurries, even after onsite optimisation of the conditions. Future experiments using H6, C4, G4, and 4 °C conditions for the production of OXA-10 microcrystals, will consider the possibility to ship microcrystal slurries. In order to do so, the quality of the microcrystals should be evaluated over time.

The reported crystallisation conditions, referred to as 144A – D (Table 3.2),[13,27–29] were used to obtain OXA-10 macrocrystals, and tested for diffraction. Optimisation for the production of OXA-10 microcrystal slurries in batch was completed for the 144D conditions. Limited testing was done on the slurries obtained under 144D conditions, however these conditions were carried forward for XFEL experiments. The optimised 144D batch conditions did not robustly produce microcrystal slurries during the sample preparation of the SACLA experiment and thus, hanging drop vapour diffusion technique was applied to obtain sufficient quantities of 144D OXA-10 microcrystal slurries. During the SFX XFEL experiment at SACLA, data collection on 144D OXA-10 microcrystals demonstrated reasonable diffraction, but suffered from a low number of integrated images, similar to A8 OXA-10 microcrystal slurries. It is possible that the quality of the 144D OXA-10 microcrystals might have been affected by the

relatively warm environment (36 – 37 °C) near the DOD system. Incubation experiments of 144D OXA-10 microcrystals with ertapenem gave a limited number of integrated images (115), that demonstrated a change in the unit cell parameters (Figure 3.14). In different systems (e. g. IPNS, BlaC, riboswitch) a change in unit cell parameters has been associated with structural changes, implying that a binding / a reaction could be taking place.[22,23]

Experiments on IPNS performed in parallel, raised concerns about diffusion properties of small molecules in the crystals which were investigated by NMR spectroscopy. The diffusion properties of ertapenem, imipenem, ceftazidime, and clavulanic acid were investigated under different conditions using NMR. The results suggest that the diffusion coefficient of these β-lactam based molecules could be significantly influenced by the crystallisation conditions employed, in particular by high concentrations of PEG. Currently, it is unclear to what extent the crystallisation conditions affect acylation of OXA-10 *in crystallo*. It could be that certain crystallisation conditions will be avoided in the future.

## 3.5 Future work

Finding conditions for the crystallisation of OXA-10 was relatively straightforward; the challenging aspect was selecting conditions that would provide crystals that diffract to high resolutions (<2 Å), using working conditions that would not bind/inhibit the active site (e. g. in absence of HEPES, MOPS, and high concentrations of halides). Solution state NMR experiments (Section **3.3.7**) demonstrated that the diffusion properties of ertapenem, imipenem, ceftazidime, and clavulanic acid were significantly affected by the presence of PEG. Currently, it is unclear how this effect on the diffusion of molecules, i. e. ertapenem, imipenem, ceftazidime, and clavulanic acid, in solution translate to their diffusion through the OXA-10 crystal lattice, and consequently affect catalytic activity in crystals. Additional systematic kinetic experiments on the rate of acylation of OXA-10 by the β-lactam compounds (i. e. ertapenem, imipenem, ceftazidime, and clavulanic acid) under various conditions could improve the understanding of the influence of the crystallisation conditions on the catalytic activity of OXA-10. Future experiments might further restrict crystallisation conditions (e. g. lowering PEG concentrations).

Some of the optimised conditions (i. e. C4, H6, and 144D, Table 3.2) for OXA-10 have shown promising initial results during fixed target experiments at DLS, diffracting to reasonable resolution with good hit rates. These initial results demonstrate the value of synchrotron based serial crystallography and will be explored further in future experiments. The SFX experiments at SACLA were troubled by crystallisation difficulties and relatively high temperatures (36 – 37 °C) during the experiment that might have affected the quality of the OXA-10 microcrystals. In order to avoid the challenges of crystallisation onsite, preparing and shipping microcrystal slurries prior to the experiment should be considered. However, the quality of the crystals should be evaluated over time to ensure high resolution diffraction during the XFEL experiments.

## 3.6 Conclusions

The work described in this chapter concerns the production and purification of OXA-10, and its crystallisation under various conditions. The different crystallisation conditions of OXA-10 were explored and probed for diffraction. The initial E8 conditions gave macrocrystals that diffracted to 1.53 Å and which were amenable to soaking with ertapenem. However, when probing the corresponding E8 microcrystals by XFEL radiation or synchrotron radiation, a low number of images could be integrated due to multiple overlapping lattices. Diluting the crystal slurry, and slurries from different batches all suffered from this issue, suggesting the multiple lattices were inherent to the crystallisation conditions. Different conditions (Table 3.2) for crystallisation were explored, tested, and optimised to produce microcrystal slurries. Testing of OXA-10 microcrystals obtained using H6 and C4 conditions using fixed target experiments at DLS demonstrated diffraction of both conditions to (2.3 Å) with a reasonable indexed hit rates (9.6, and 17 – 30% respectively). During these experiments A8 OXA-10 microcrystals were also tested which showed diffraction to 2.2 Å, with promising indexed hit rates (17 – 28%). However, these A8 OXA-10 microcrystals were non-isomorphous, indexing with two different unit cell parameters within the same space group. Although indexing was possible, it was considered less favourable for future mixing experiments. During XFEL experiments OXA-10 crystallisation was attempted under different conditions (A8, C4, H6, G4, 4 °C, and 144D). OXA-10 microcrystals obtained under A8, and 144D conditions were tested for diffraction; surprisingly, A8 OXA-10 microcrystals demonstrated low resolution diffraction (>2.8 Å) with a low number of indexed images (~1%), while 144D OXA-10 microcrystals suffered from a similarly low number of indexed images. Soaking of the 144D OXA-10 microcrystals with ertapenem resulted in a change in unit cell parameters, however only few images were collected before the sample ran out. It is possible that the quality of the crystals suffered from the higher temperatures in the hutch (36 – 37 °C).

Solution state NMR experiments on the diffusion of ertapenem, clavulanic acid, ceftazidime, and imipenem suggested that their diffusion properties are affected by the crystallisation conditions used. At present it is unclear to what extend the conditions affect the turnover kinetics in crystals. The results of these experiments might restrict the conditions used in future crystallisation trials.

## 3.7 Protein production and purification

### 3.7.1 General methods

Growth medium 2TY was prepared with tryptone (16 g · $L^{-1}$), yeast extract (10 g · $L^{-1}$), and sodium chloride (5 g/L). All media were sterilised prior to growth using an autoclave.

SDS-PAGE were prepared according to the recipe in table 3.5 using the Mini-PROTEAN Tetra Cell system (Bio-Rad). During preparation of the running gel, the buffer was covered with isopropyl alcohol to prevent it from drying out whilst setting. Tetramethylethylenediamine (TEMED) was added last to both the running and the stacking gel.

**Table 3.5 Reagents for the preparation of the stacking and running gel (4 SDS-PAGE).**

| Reagent | Running gel | Stacking gel |
|---|---|---|
| Tris(hydroxymethyl)aminomethane | 4.00 mL (1.5 M, pH 8.8) | 1.25 mL (0.5 M, pH 8.5) |
| $H_2O$ | 5.06 mL | 3.05 mL |
| Ammonium persulfate (10 w/v%) | 267 µL | 50 µL |
| Bis-acrylamide (30 w/v%) | 6.67 mL | 670 µL |
| Sodium dodecyl sulfate (10 w/v%) | 160 µL | 50 µL |
| Tetramethylethylenediamine | 13.3 µL | 5.0 µL |

Loading (2 ×) and running (10 ×) buffers were prepared according to the recipes below (Table 3.6). Protein samples (8 µL) were mixed with loading buffer (8 µL), and heated to 95 °C for 10 minutes. Subsequently, the samples (10 µL) were loaded onto an SDS-PAGE. Electrophoresis was run at constant voltage (180 V, 50 minutes). Gels were then stained with SDS-PAGE staining solution (10 min), and de-stained with $H_2O$.

**Table 3.6 Components of SDS-PAGE loading and running buffers, and staining solution**

| Solution | Reagent | Amount |
|---|---|---|
| Loading buffer (2 ×)[a] | Sodium dodecyl sulfate | 6.00 g |
| | Dithiothreitol | 4.65 g |
| | Tris HCl (1 M, pH 6.8) | 3.00 mL |
| | Bromophenol blue | 30.0 mg |
| | Glycerol | 23.5 mL |
| Running buffer (10 ×)[b] | Sodium dodecyl sulfate | 10.0 g |
| | Tris HCl | 30.0 g |
| | Glycine | 144 g |
| Staining solution[c] | Coomasie brilliant blue G250 | 400 mg |
| | Perchloric acid (70%) | 34 mL |
| | Ethanol | 50 mL |

a) Components were dissolved in $H_2O$ to give a total of 150 mL.
b) Components were dissolved in $H_2O$ to give a total of 1.00 L.
c) Components were dissolved in $H_2O$ to give a total of 900 mL.

### 3.7.2 Expression and purification of OXA-10

*Recombinant P. aeruginosa OXA-10 plasmid was obtained from Dr Christopher Lohans (University of Oxford) and protocols for expression and purification were obtained from Dr Jürgen Brem (University of Oxford). The stock originates from a plasmid provided by Dr Moreno Galleni, University of Liege, Belgium.*

#### 3.7.2.1 Bacterial growth

BL21(DE3) competent cells (50 μL, *obtained from Dr Patrick Rabe, University of Oxford*) for transformation were thawed on ice in presence of the OXA-10 producing plasmid (1 μL, *obtained from Dr Christopher Lohans, University of Oxford*) for 30 minutes. The cells were then heat shocked (45 s) at 42 °C, and subsequently cooled on ice (5 min). Super optimal broth with catabolite repression (SOC) outgrowth medium (300 μL) was added to the mixture and the cells were grown for 1 h at 37 °C. After incubation, the cells (200 μL) were dispensed on an 2TY-agar plate containing kanamycin (25 μg · $mL^{-1}$ final concentration), and incubated overnight at 37 °C. A single colony was selected, picked and used to inoculate, and grown in 2TY media (100 mL) overnight at 37 °C. The starter culture (650 μL) was mixed with glycerol (450 μL) and stored at –80 °C.

Starter cultures were grown in 2TY medium (100 mL, in 500 mL Erlenmeyer flasks), supplemented with kanamycin (25 μg · $mL^{-1}$ final concentration), and inoculated from a glycerol stock, at 37 °C for 16 h.

Large scale growths in 2TY medium (24 × 600 mL, in 2 L Erlenmeyer flasks), supplemented with kanamycin (25 μg · $mL^{-1}$ final concentration), were inoculated with starter culture (1 : 1000 dilution), and continued to grow until $OD_{600}$ 1.6 (~6 h). The cultures were cooled to 18 °C and inoculated with isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM final concentration), and incubated for an additional 4 h. The cells were pelleted by centrifugation (8000 rpm, 10 minutes) at 4 °C and stored at –80 °C.

#### 3.7.2.2 Purification of OXA-10

The cells were thawed and suspended in lysis buffer A (50 mM MES, pH 6.0), supplemented with DNAse I (10 μg / mL), and phenylmethylsulfonyl fluoride (10 μg · $mL^{-1}$), at 4 °C. The cells were then sonicated at 4 °C (9.9 s sonication, 9.9 s break, 60% amplitude, 10 min, Sonics Vibra-Cell). Debris was removed by

centrifugation (20,000 rpm, 25 minutes) and the supernatant was collected. After filtration through a sterile filter (0.45 μm), the supernatant was loaded onto an anion exchange column (SP Sepharose, 100 mL), preequilibrated with buffer A (3 column volumes). The protein was eluted with a gradient of buffer B (50 mM MES, pH 6.0; 1 M NaCl). Fractions containing protein were identified by UV trace analysis and their purity checked by SDS-PAGE. The selected fractions were combined, concentrated, and further purified using size-exclusion chromatography (Superdex™ 75 column, 300 mL, GE Healthcare preequilibrated with buffer C (MES 50 mM, pH 6.0, NaCl, 100 mM, TCEP, 10 mM, 1 column volume)). Fractions containing protein were identified by UV trace and their purity checked by SDS-PAGE. The fractions containing pure protein (>95%) of interest were selected and concentrated to 8.5 mg · $mL^{-1}$, aliquoted and stored at –80 °C.

# 3.8 Crystallography

## 3.8.1 Crystallisation of OXA-10

### 3.8.1.1 Broad screens

Crystals of OXA-10 (60, 8.5 mg · $mL^{-1}$) were obtained using vapor diffusion technique at 25 °C or 4 °C in Intelli-Plate 96-3 (Art Robbins Instruments). Commercial crystallisation broad screens, including JCSG+, PACT premier, and SaltRX (Molecular Dimensions), or custom optimisation conditions (prepared manually) were used without further treatment. The crystallisation plates were prepared using a Rigaku Phoenix RE drop setter (Art Robbins Instruments). When stated, crystals were transferred to buffer solution supplemented with glycerol as a cryo-protectant. Single crystals were obtained and cryo cooled in $N_2$ (l).

### 3.8.1.2 Sitting drop plates

Crystallisation of OXA-10 (2 – 4 μL, 8.5 mg · $mL^{-1}$) in 24 well sitting drop plates (Art Robbins Instruments) at 25 °C or 4 °C, were prepared manually. The surrounding well was filled with buffer (200 or 400 μL) and a droplet (6 μL total volume) was created in different ratios. When stated, crystals were transferred to buffer solution supplemented with glycerol as a cryo-protectant. Single crystals were obtained and cryo cooled in $N_2$ (l).

### 3.8.1.3 Hanging drop plates

Crystallisation of OXA-10 (1.3 – 2.7 μL, 8.5 mg · $mL^{-1}$) using hanging drop vapour diffusion was done using a 24 well hanging drop plate. Droplets (4 μL total volume) of protein and precipitant solution ($(NH_4)_2SO_4$ (2.0 M), Tris (0.1 M, pH 8.5) /HEPES (0.1 M pH 7.5)) were placed over a well containing precipitant solution (500 μL).

#### 3.8.1.4 Soaking of substrate

Stock solutions were prepared by dissolving ertapenem in crystallisation solution (50 or 100 mM). Droplets containing OXA-10 crystals were diluted with a range of concentrations of ertapenem stock solutions (6 : 1 – 1 : 1). After incubation, single crystals were obtained and cryo cooled in $N_2$ (l).

#### 3.8.1.5 Sitting drop plates

Microcrystal slurries of OXA-10 were obtained using a batch plate (96 wells, 100 μL per well). Optimised precipitant solutions were prepared separately and stored at rt.

### 3.8.2 Seed stock preparation

Stocks were prepared from dense crystal slurries (Section **3.8.2**) or from single crystals (grown according to reported procedure[33,34]). Suspensions containing crystals (100 μL) were transferred to seed bead kit (plastic, Hampton research) cooled to 4 °C. The solution was vortexed (30 s) and then cooled to 4 °C (30 s), the steps were repeated 10 times. The seed stocks were then aliquoted (10 μL) and stored at –80 °C.

### 3.8.3 Data collection

Synchrotron crystallographic data sets were collected using MX beamlines at Diamond Light Source (Didcot, UK). XFEL crystallographic data sets were collected using the Linac Coherent Light Source (LCLS, California, United States) or SPring-8 Angstrom Compact free electron LAser (SACLA, Japan). Phases were solved by molecular replacement using Phaser/CCP4 (IPNS accession code 1BLZ)[35]. Refinements were done using Phenix,[36] and iterative fitting using Coot[37] software packages. RMSD values were determined using PyMol 2.3. Calculation of expected protein molecular weight was done using an online bioinformatic tool, ExPASy (web.expasy.org/compute_pi/, accessed 2016 – 2020). The solvent content within unit cells were estimated using MATTPROB using an online tool (ruppweb.org/mattprob/default.html, accessed May 2020).[18–20]

## 3.9 Diffusion experiments by NMR

All compounds for the diffusion NMR spectroscopy (DOSY) experiments, including ertapenem, meropenem, clavulanic acid (1 : 1 in cellulose), ceftazidime, PEG3350 / PEG10,000, ammonium sulfate, were obtained from commercial sources and used without further purification. Stock solutions were prepared in $D_2O$ (>99% $^2H$) without buffering, and diluted appropriately with $D_2O$ to give the desired concentration. All diffusion experiments were carried out in 5 mm NMR tubes (Norell) at 298 K on a Bruker AVIII 500 equipped with a BBO probe unless stated otherwise. Spectra were collected using the LED experiment using bipolar gradients (ledbpgp2s) with a squared gradient ramp of 16 steps between 5 and 95%, a gradient pulse length of 1.5 ms and a diffusion time of 100 ms. For each step 8 scans were recorded. Suitable peaks were selected for integration and fitted with equation (3.1) using Bruker Topspin 3.6.1. The calculated results for different peaks were averaged to give the final diffusion coefficient.

$$I_G = I_{G=0} e^{-(\gamma \delta G)^2 D\left(\Delta - \frac{\delta}{3}\right)} \quad (3.1)$$

Where $I$ = measured signal intensity, $I_{G=0}$ = signal intensity for g = 0, $\gamma$ = magnetogyric ratio, $D$ = diffusion coefficient, $G$ = gradient strength, $\Delta$ = diffusion time, $\delta$ = gradient pulse length.

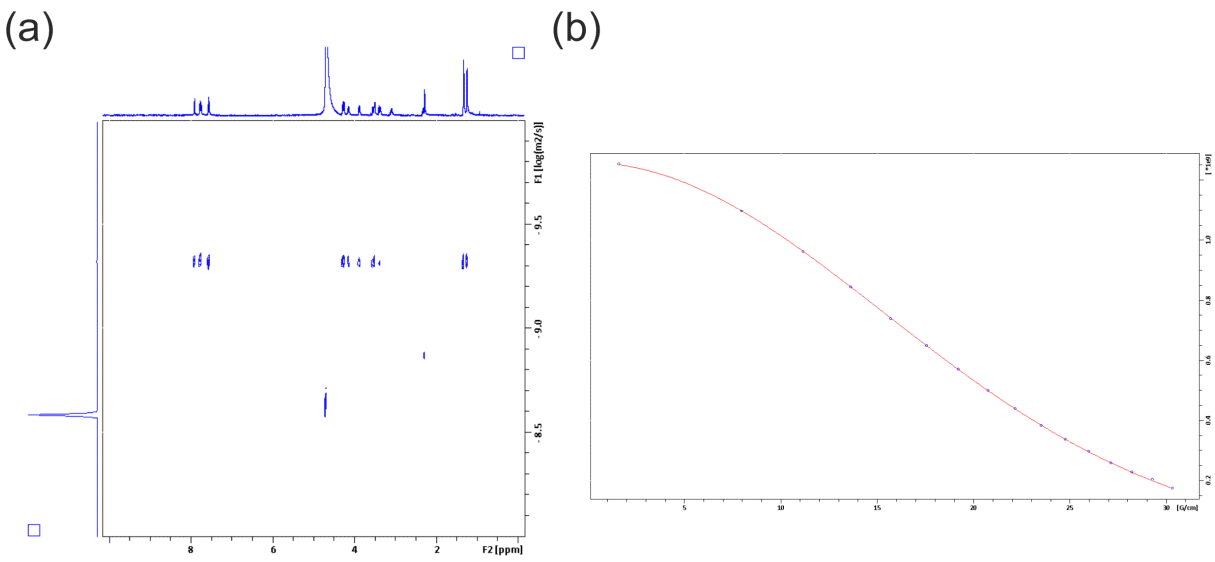



**Figure 3.15:** Exemplary diffusion ordered spectroscopy (DOSY) data and point fitting. (a) DOSY spectrum of ertapenem in $D_2O$ at 310 K; (b) curve fitting of a single integral region using equation (3.1). Data were processed using Bruker Topsin 3.6.1.

## 3.10 References


1. Bush, K. & Bradford, P. A. β-Lactams and β-Lactamase Inhibitors: An Overview. *Cold Spring Harb Perspect Med* 1–22 (2016).
2. Llarrull, L. I., Testero, S. A., Fisher, J. F. & Mobashery, S. The future of the β-lactams. *Curr. Opin. Microbiol.* **13**, 551–557 (2010).
3. Bush, K. Proliferation and significance of clinically relevant β-lactamases. *Ann. N. Y. Acad. Sci.* **1277**, 84–90 (2013).
4. Drawz, S. M. & Bonomo, R. A. Three decades of β-lactamase inhibitors. *Clin. Microbiol. Rev.* **23**, 160–201 (2010).
5. Maveyraud, L. *et al.* Insights into class D β-lactamases are revealed by the crystal structure of the OXA10 enzyme from Pseudomonas aeruginosa. *Structure* **8**, 1289–1298 (2000).
6. Golemi, D. *et al.* The first structural and mechanistic insights for class D β-lactamases: Evidence for a novel catalytic process for turnover of β-lactam antibiotics. *J. Am. Chem. Soc.* **122**, 6132–6133 (2000).
7. Lohans, C. T. *et al.* Mechanistic Insights into β-Lactamase-Catalysed Carbapenem Degradation Through Product Characterisation. *Sci. Rep.* **9**, 1–9 (2019).
8. Evans, B. A. & Amyes, S. G. B. OXA β-lactamases. *Clin. Microbiol. Rev.* **27**, 241–263 (2014).
9. Antunes, N. T. *et al.* Class D β-lactamases: Are they all carbapenemases? *Antimicrob. Agents Chemother.* **58**, 2119–2125 (2014).
10. Queenan, A. M. & Bush, K. Carbapenemases: The versatile β-lactamases. *Clin. Microbiol. Rev.* **20**, 440–458 (2007).
11. Jeon, J. H. *et al.* Structural basis for carbapenem-hydrolyzing mechanisms of carbapenemases conferring antibiotic resistance. *Int. J. Mol. Sci.* **16**, 9654–9692 (2015).
12. Lohans, C. T. *et al.* A New Mechanism for β-Lactamases: Class D Enzymes Degrade 1β-Methyl Carbapenems through Lactone Formation. *Angew. Chemie - Int. Ed.* **57**, 1282–1285 (2018).
13. King, D. T., King, A. M., Lal, S. M., Wright, G. D. & Strynadka, N. C. J. Molecular Mechanism of Avibactam-Mediated β-Lactamase Inhibition. *ACS Infect. Dis.* **1**, 175–184 (2016).
14. Lohans, C. T. *et al.* 13C-Carbamylation as a mechanistic probe for the inhibition of class D β-lactamases by avibactam and halide ions. *Org. Biomol. Chem.* **15**, 6024–6032 (2017).
15. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).
16. Schmidt, M. Time-resolved macromolecular crystallography at pulsed X-ray sources. *Int. J. Mol. Sci.* **20**, (2019).
17. Baurin, S. *et al.* Critical role of tryptophan 154 for the activity and stability of class D β-lactamases. *Biochemistry* **48**, 11252–11263 (2009).
18. Kantardjieff, K. A. & Rupp, B. Proceedings of 10th International Conference on Formal Methods in Computer-Aided Design, FMCAD 2010, Lugano, Switzerland, October 20-23. *Protein Sci.* **12**, 1865–1871 (2003).
19. Weichenberger, C. X. & Rupp, B. Ten years of probabilistic estimates of biocrystal solvent content: New insights via nonparametric kernel density estimate. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1579–1588 (2014).
20. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
21. Brem, J. *et al.* Rhodanine hydrolysis leads to potent thioenolate mediated metallo-β 2-lactamase inhibition. *Nat. Chem.* **6**, 1084–1090 (2014).
22. Olmos, J. L. *et al.* Enzyme intermediates captured 'on the fly' by mix-and-inject serial crystallography. *BMC Biol.* **16**, 1–15 (2018).
23. Stagno, J. R. *et al.* Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017).
24. Sauter, N. K., Hattne, J., Grosse-Kunstleve, R. W. & Echols, N. New Python-based methods for data processing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1274–1282 (2013).
25. Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat. Methods* **14**, 443–449 (2017).
26. Stojanoski, V. *et al.* Structural Basis for Different Substrate Profiles of Two Closely Related Class D β-Lactamases and Their Inhibition by Halogens. *Biochemistry* **54**, 3370–3380 (2015).
27. Golemi, D., Maveyraud, L., Vakulenko, S., Samama, J. P. & Mobashery, S. Critical involvement of a carbamylated lysine in catalytic function of class D β-lactamases. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14280–14285 (2001).
28. McKinney, D. C. *et al.* 4,5-Disubstituted 6-Aryloxy-1,3-dihydrobenzo[c][1,2]oxaboroles Are Broad-Spectrum Serine β-Lactamase Inhibitors. *ACS Infect. Dis.* **1**, 310–316 (2016).
29. Brem, J. *et al.* Structural basis of metallo-β-lactamase, serine-β-lactamase and penicillin-binding protein inhibition by cyclic boronates. *Nat. Commun.* **7**, 12406 (2016).
30. Sherrell, D. A. *et al.* A modular and compact portable mini-endstation for high-precision, high-speed fixed

target serial crystallography at FEL and synchrotron sources. *J. Synchrotron Radiat.* **22**, 1372–1378 (2015).
31. Owen, R. L. *et al.* Low-dose fixed-target serial synchrotron crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 373–378 (2017).
32. Davy, B. *et al.* Reducing sample consumption for serial crystallography using acoustic drop ejection. *J. Synchrotron Radiat.* **26**, 1820–1825 (2019).
33. Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–30 (1997).
34. Roach, P. L. *et al.* Anaerobic crystallisation of an isopenicillin N synthase.Fe(II).substrate complex demonstrated by X-ray studies. *Eur. J. Biochem.* **242**, 736–740 (1996).
35. Byford, M. F., Baldwin, J. E., Shiau, C.-Y. & Schofield, C. J. The Mechanism of ACV Synthetase. *Chem. Rev.* **97**, 2631–2650 (1997).
36. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

# Chapter 4

## Time-resolved crystallography on Isopenicillin N Synthase

# 4.1 Introduction

The work in this chapter describes crystallographic, NMR spectroscopic, and mass spectrometric (MS) studies on isopenicillin N synthase (IPNS). Time resolved (tr) serial femtosecond (fs, $10^{-15}$ s) crystallography (SFX) experiments were carried out to investigate the catalytic cycle of IPNS and study proposed intermediates; in particular the Fe(III) superoxide and Fe(IV) oxo species (Section **1.3.4**). During the tr crystallographic studies a structural rearrangement of α-helix 3 and β-strand 13 in IPNS as a consequence of $O_2$ exposure was observed. To investigate the rearrangement in solution, NMR studies on IPNS were performed. Investigation of the **IPN** product bound IPNS structure manifested the partially reduced occupancy of Fe at the active site. In order to investigate potential influence of the presence of **IPN** on dissociation of Fe from the active site after catalytic turnover, solution state NMR studies and non-denaturing MS studies were carried out.

### 4.1.1 Crystallisation of isopenicillin N synthase

The first structure of IPNS (*Aspergillus nidulans*) was in complex with Mn(II) and was reported by Roach *et al.* to 2.5 Å resolution (Section **1.3.4**).[1,2] An anaerobic structure of IPNS in complex with Fe(II) and its tripeptide substrate L-δ-(α-aminoadipoyl)-L-cysteinyl D-valine (**ACV**) (1.3 Å) was presented in 1997.[3] The crystallisation conditions used in these studies gave rise to three different crystal forms: a trigonal crystal form belonging to the $P3_121$ space group (hexagonal morphology); an orthorhombic crystal form belonging to the $P2_12_12_1$ space group (plate morphology); and another orthorhombic crystal form belonging to the $P2_12_12_1$ space group (needle morphology).[4] Studies on exposing the orthorhombic crystal form to $O_2$ under hyperbaric pressure (60 bar), led to partial turnover of the **ACV** substrate to isopenicillin N (**IPN**) in crystals (Figure 1.12, Section **1.3.4**).[5] The electron density maps suggested a mixture of **IPN** and **ACV** in the active site; prolonged exposure (>320 min) led to low occupancy for **IPN** and **ACV**, which was rationalized by the **IPN** product leaving the active site (Figure 1.12 (e)). Exposure of IPNS crystals in complex with a substrate analogue, δ-(α-aminoadipoyl)-cysteinyl-*S*-methyl cysteine (**ACmC**), to $O_2$ under pressure (20 bar), was claimed to interrupt the catalytic cycle at the monocyclic β-lactam intermediate (Figure 1.12 (f)).[5] Alongside the formation of the β-lactam ring, the sulfur of the *S*-methylated Cys was oxidised, which was

rationalised *via* reaction of a high valent Fe(IV) ferryl species. However, direct crystallographic evidence for the formation of a ferryl intermediate has been lacking.

### 4.1.2 Time resolved SFX using isopenicillin N synthase

Part of the work discussed in this chapter concerns tr serial crystallography using XFELs and / or synchrotron radiation to study the catalytic cycle of IPNS. The reaction catalysed by IPNS is unique in that it catalyses the formation of a β-lactam and thiazolidine ring consecutively *via* a 4 electron oxidation of **ACV** substrate, involving two high valent Fe intermediates (i.e. Fe(III) superoxide and Fe(IV) oxo). High valent Fe intermediates like those involved in IPNS catalysis, are proposed as intermediates in many redox reactions. The XFEL approach for IPNS differs from previous *in crystallo* turnover attempts, as tr-SFX can be used to collect X-ray diffraction and X-ray emission spectroscopy (XES) data simultaneously at defined time points at physiological temperature, using pure $O_2$. To support the tr-SFX data, cryo-cooled IPNS substrate were obtained using synchrotron radiation.

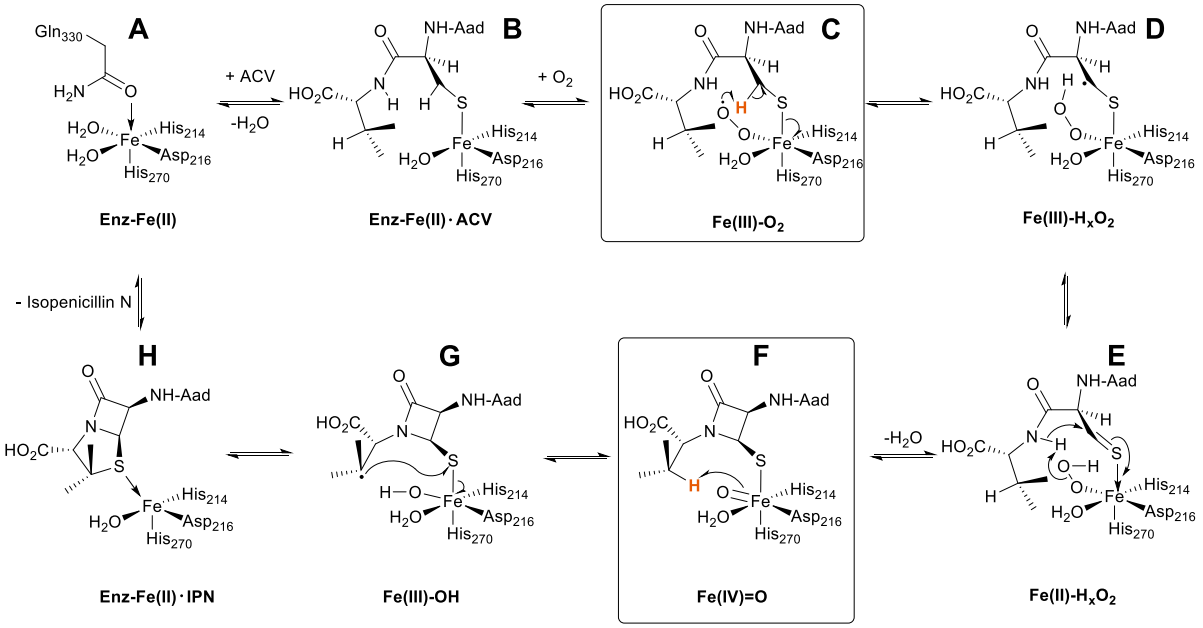



**Figure 4.1.** The proposed mechanism for the IPNS catalysed turnover of **ACV**, which is based on substrate analogue, crystallographic, computational, and kinetic studies.[6,7] Abstraction of protons (orange) of intermediates C and F have been proposed to exhibit a strong kinetic isotope effect (KIE).[8] *This scheme is an adaptation from one in Rabe et al.*[7]

IPNS was considered an attractive candidate for tr-SFX studies at XFEL, as it can be produced in large amounts,[9] and it co-crystallises robustly with Fe(II) and **ACV** to produce crystals that diffract to high resolution. Moreover, there is literature precedent for substrate turnover experiments with IPNS, suggesting IPNS is susceptible to reactions in crystalline form (Section **1.3.4**). The IPNS • Fe • ACV microcrystals (Section **4.3.1**) are relatively small in dimension, a feature that is proposed to minimise $O_2$ gradients throughout the crystal lattice. In addition, isotopically labelled L-δ-(α-aminoadipoyl)-L-2,2-[$^2H_2$]-cysteinyl D-valine (**AdCV**), and L-δ-(α-aminoadipoyl)-L-cysteinyl D-[$^2$H]-valine (**ACdV**) were prepared. These two substrates show a strong kinetic isotope effect (KIE),[10,11] which it was envisaged might be exploited in order to slow down the catalytic cycle significantly at particular stages, facilitating the characterisation of intermediates; i. e. Fe(III) superoxide and Fe(IV) oxo intermediates (Figure 4.1 C and F). The preparation of the **ACV** substrate, isotopically labelled **ACV** and IPNS for the time resolved crystallography experiments outlined above, are discussed below.

## 4.2 Preparation of isopenicillin N and substrate

### 4.2.1 Substrates from a commercial source

Initially, the LLD-ACV substrate was obtained from a commercial supplier (GL Biochem, Shanghai, China). After producing recombinant IPNS (Section **4.2.6**), the conditions for turnover were optimised (IPNS (20 – 50 μM), iron (II) (50 – 200 μM), sodium ascorbate (200 – 1500 μM), ACV/substrate analogues (1 – 2 mM), and TCEP/DTT (1 – 2 mM)). However, under the tested conditions, the ACV substrate failed to convert fully to the **IPN** product (31% conversion, as observed by MS and $^1$H NMR, Figure 4.2, (a)). After thorough investigation, the structure of the purchased substrate was shown to be LLD-α-(aminoadipoyl)cysteinyl valine (**αACV**) instead of LLD-δ-(aminoadipoyl)cysteinyl valine (**ACV**) (Figure 4.2, (b)). The tripeptide was purchased together with analogues including LLD-aminoadipoyl cysteinyl glycine, and LLD-aminoadipoyl cysteinyl *O*-methyl serine. All of these tripeptides contained the α-(α-aminoadipoyl) regioisomers. It was decided to prepare the correct tripeptides synthetically (Section **4.2.2**).

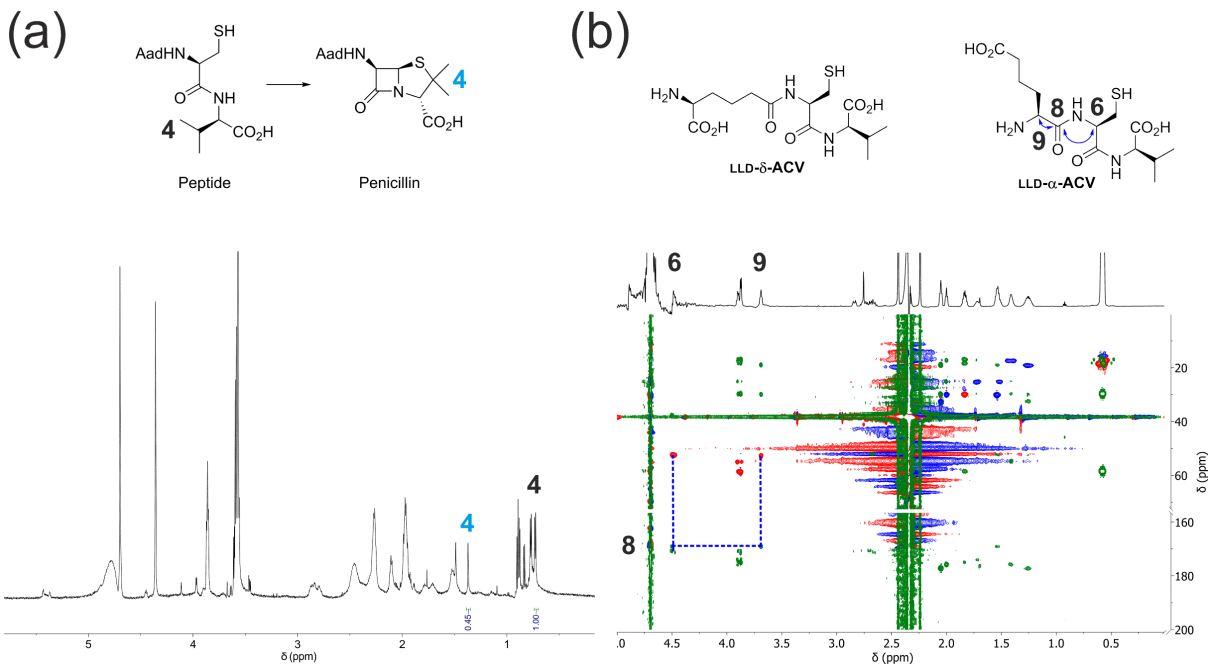



**Figure 4.2:** $^{1}$H NMR spectrum from incubation of ACV (GL Biochem, Shanghai, China) with IPNS, and HSQC and HMBC analyses of tripeptide ACV. (a) $^{1}$H NMR of ACV incubated with IPNS, Fe, ascorbate, and TCEP. Resonances used for integration and to estimate conversion are labelled. (b) Structures of two regioisomers of LLD-α-aminoadipoyl cysteinyl valine (ACV): δ-(α-aminoadipoyl)cysteinyl valine, and α-(α-aminoadipoyl)cysteinyl valine; HSQC (red/blue) and HMBC (green) overlay of a solution containing commercially supplied ACV. Coupling of Cys C-$H_{\alpha}$ (○) to Aad main chain C=O carboxylic acid which in turn couples to Aad C-$H_{\alpha}$ (●), suggesting a connectivity pattern consistent with α-LLD-ACV.

### 4.2.2 Solution phase synthesis of ACV

The solution phase synthesis of the substrate of IPNS, LLD-**ACV**, has been reported.[12–16] An adaptation of a route reported by Baldwin and coworkers[15] was used for the preparation of **ACV** with a final acid mediated deprotection step (Figure 4.3). Thus, commercially available L-α-aminoadipic acid (Aad) was efficiently protected on its amino group (85%) to give *N*-Boc protected **1** (Figure 4.3 (1)). Diacid **1** was protected using PMB-Cl, to give a mixture of α- and/or δ-protected Aad esters, which were separated by column chromatography to yield the desired C-1 ester protected **2** (37%). Cys was *S*-protected by stirring with diphenylmethanol in TFA to afford *S*-benzhydryl protected Cys **3** efficiently (91%) (Figure 4.3 (2)).[17] *N*-Boc protected D-Val **4** was protected using diazodiphenylmethane, which was freshly prepared from benzophenone hydrazone using HgO (64%).[18] Alternatively, protection of D-Val was achieved through reaction of *N*-Boc protected D-Val **4** and benzophenone hydrazone using $MnO_2$ to give **4a** (60%), as

reported.[19] Subsequent deprotection of the *N*-Boc group of **4a** was achieved using *p*-toluenesulfonic acid (quantitative) (Figure 4.3 (3)) to give salt **5**.

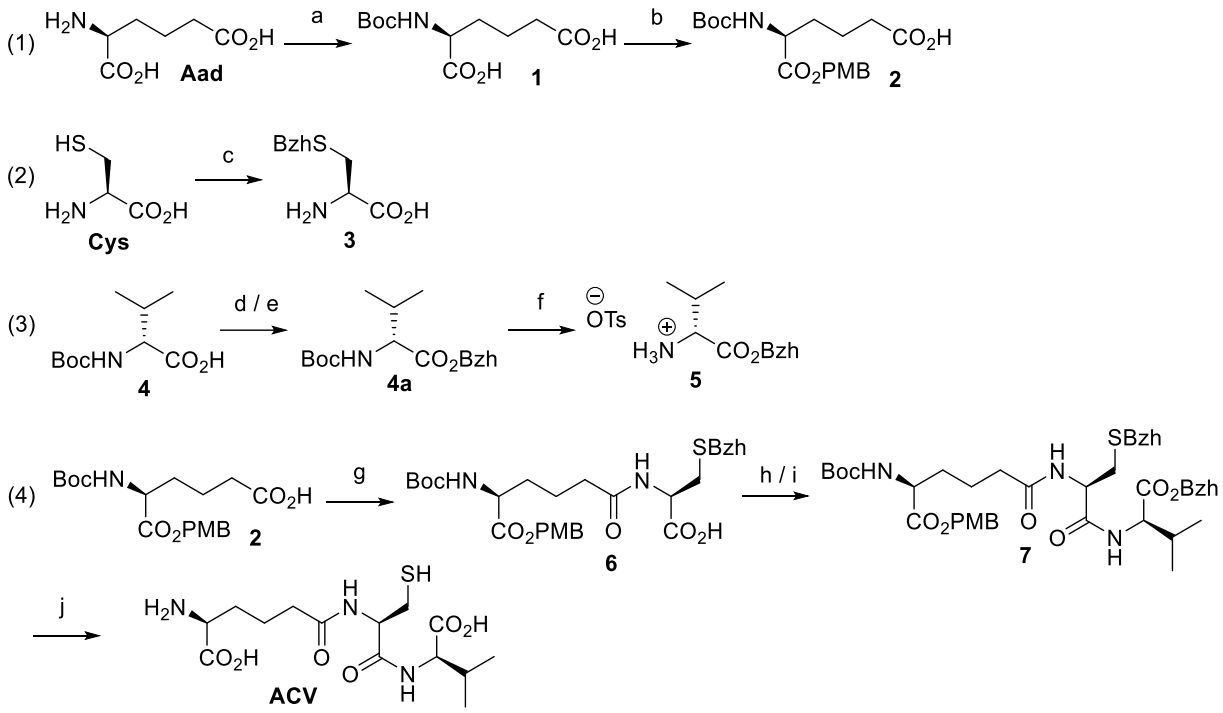



**Figure 4.3:** The solution phase synthesis of **ACV**. *Reagents and conditions:* (a) $Boc_2O$, NaOH, $^tBuOH$ / $H_2O$ (1 : 1), rt, 48 h, 85%; (b) PMB-Cl, $Et_3N$, DMF, 35 °C, 24 h, 37%; (c) diphenylmethanol, TFA, rt, 2 h, 91%; (d) $Ph_2CN_2$, $CH_2Cl_2$, rt, 1 h, 64%; (e) $MnO_2$, $KH_2PO_4$, benzophenone hydrazine, $CH_2Cl_2$, rt, 16 h, 60%; (f) *p*-toluenesulfonic acid, THF, 1 h, quant; (g) i) isobutyl chloroformate, $Et_3N$, –15 °C, THF, ii) **3**, THF / $H_2O$ (1 : 1), –15 °C to rt, 24 h, 33%; (h) $Et_3N$, EEDQ, $Na_2SO_4$, **5**, $CH_2Cl_2$, rt, 16 h; (i) $Et_3N$, HATU, **5**, $CH_2Cl_2$, rt, 16 h, 35%; (j) anisole, TFA, rt, 6 h, 60%.

Bisprotected Aad **2** was activated using isobutyl chloroformate; the active ester was then coupled with *S*-protected Cys **3** to afford dipeptide **6** (Figure 4.3 (4)). Fully protected ACV **7** was obtained by coupling of dipeptide **6** with benzhydryl ester **5**, using either EEDQ or HATU. The procedure using HATU gave a more consistent yield (~35%) and was used for larger scale reactions. Subsequent TFA-mediated deprotection of **7** afforded **ACV**. The crude **ACV** was treated with TCEP before purification by preparative-HPLC with an acetonitrile gradient in $H_2O$ (both containing TFA (0.1% (v/v)), followed by a second purification by preparative-HPLC with an acetonitrile gradient in $H_2O$ (containing ammonium bicarbonate (10 mM)). The **ACV** product was obtained by lyophilisation (60% from **7**). Attempts to purify **ACV** in one step by HPLC using ammonium bicarbonate (10 mM) solution suffered from poor separation. Disulfide formation after the HPLC purification using aqueous TFA was limited, but was substantial when

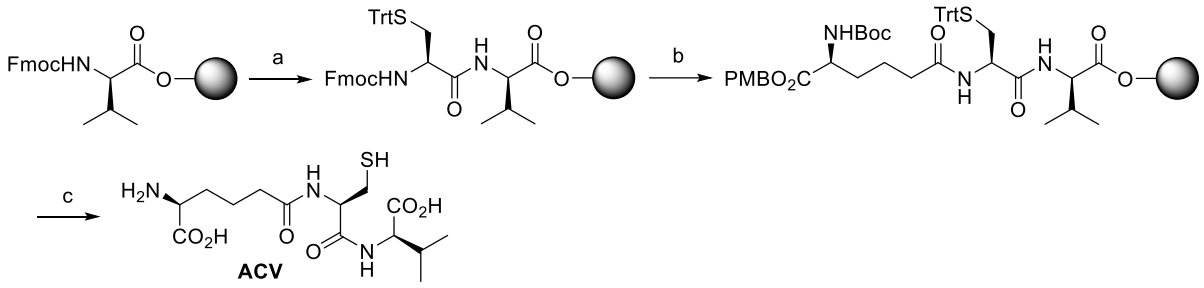



**Figure 4.4:** The solid phase peptide synthesis of **ACV**. *Reagents and conditions:* (a) DIC, ethyl cyanohydroxyiminoacetate (Oxyma), Fmoc-Cys(Trt)-OH, DMF, mw, 10 min; (b) DIC, oxyma, Boc-Aad(OH)-OPMB, mw, 20 min; (c) TFA, $H_2O$, TIPS, 2,2'-(ethylenedioxy)diethanethiol, rt, 6 h.

purifying with ammonium bicarbonate as observed by $^1$H NMR; therefore, fractions containing the **ACV** peptide, were immediately frozen after collection to limit disulfide bond formation.

### 4.2.3 Solid phase peptide synthesis of ACV

To enable a more efficient preparation of **ACV** and its analogues, a route using solid phase peptide synthesis (SPPS) was explored (Figure 4.4). Standard *N*$^a$-Fmoc chemistry[20] was used for the preparation of **ACV**, starting from commercially available Fmoc-D-Val linked to the Wang resin. Amino acid coupling was typically mediated by DIC and Oxyma and assisted by microwave heating.[21,22] Fmoc-Cys(Trt)-OH is commercially available and was used in excess (5 equivalents) to couple with the resin linked D-Val, produced by piperidine mediated deprotection of the resin linked Fmoc-D-Val. The diprotected Aad **2** was in excess (1.5 – 2 equivalents) in the final coupling step and allowed to react for a prolonged period of time (20 rather than 10 minutes). The tripeptide was cleaved from the resin and deprotected using a cleavage mixture (75% (v/v) TFA, 2.5% $H_2O$, 2.5% TIPS, 2,2'-(ethylenedioxy)diethanethiol 20%). Initial attempts to cleave the tripeptide from the resin, suffered from low yields of **ACV**, due to production of a by-product, **AC(PHB)V** (Figure 4.5). The structure of **AC(PHB)V** was assigned by NMR. Thus, through bond correlations by HMBC showed coupling of the benzylic methylene to the Cys β-methylene and aromatic CH-5, suggesting a structure consistent with **AC(PHB)V** (Figure 4.5 (b)). The *S*-linked *p*-methoxybenzyl group on Cys residues arise from a reaction of a thiol, typically Cys at positions 1, 2, or 3, with the Wang resin under acidic conditions, and is a known challenge when using Wang resin in SPPS.[23] **AC(PHB)V** was converted to the intended **ACV** peptide by mixing with excess 2,2'-(ethylenedioxy)diethanethiol and TFA.

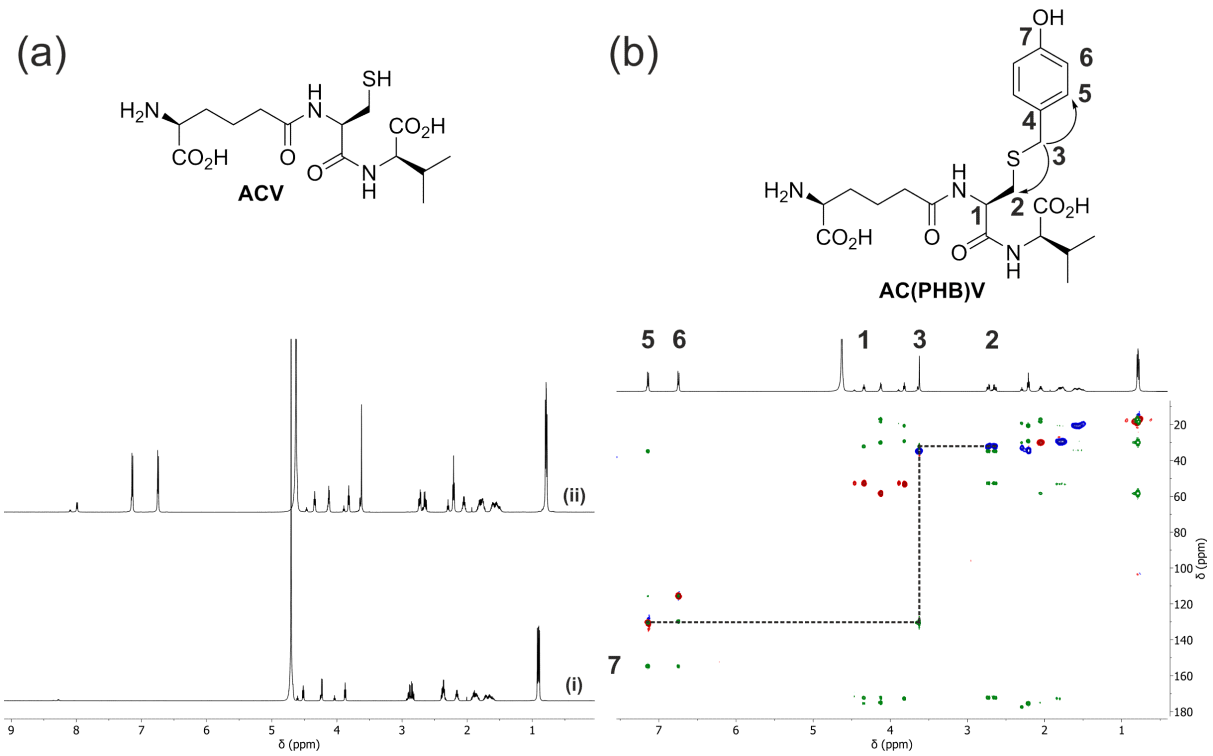



**Figure 4.5:** Identification of the side product encountered in SPPS route to **ACV**. (a) (i) $^{1}$H NMR of **ACV** after prep HPLC purification, (ii) $^{1}$H NMR of the **AC(PHB)V** side product obtained after preparative HPLC purification; (b) HSQC (blue/red) and HMBC (green) analysis of **AC(PHB)V**.

The side reaction was suppressed by increasing the content of 2,2'-(ethylenedioxy)diethanethiol (20% (v/v)) in the cleavage mixture. Purification of the final tripeptide was achieved by preparative-HPLC (Section **3.2.2**). The SPPS approach was used to obtain deuterated **ACV** analogues **AdCV** and **ACdV**.

### 4.2.4 Preparation of deuterated ACV analogues

Two steps in the catalytic cycle of IPNS show a strong KIE (Section **1.3.4**),[10,11] which could be exploited to facilitate the crystallographic characterisation of certain intermediates. The preparation of selectively deuterated **ACV** substrates was thus carried out. The sites selected for deuterium incorporation were the Cys β-methylene and the Val C3 methine, i.e. peptides **AdCV** and **ACdV**, respectively (Figure 4.6). Preparation of **AdCV** was done using SPPS and commercially available Fmoc-3,3-[$^{2}H_2$]-L-Cys(Trt)-OH (Cambridge Isotopes, *in collaboration with Dr Patrick Rabe, University of Oxford*), using the SPPS method described above (Section **3.2.3**), following repeated synthesis of unlabelled substrate to have robust and reliable conditions. After piperidine mediated Fmoc deprotection of Fmoc-D-Val coupled on Wang resin,

coupling of Fmoc-3,3-[$^2H_2$]-L-Cys(Trt)-OH in a slight excess (1.5 equivalents) for a prolonged period of time (20 rather than 10 minutes) was achieved. After piperidine deprotection of the Fmoc group from the deuterated Cys connected to the resin, coupling of protected Aad **2** was achieved. Finally, the peptide was deprotected and cleaved from the resin using a cleavage mixture (75% (v/v) TFA, 2.5% $H_2O$, 2.5% TIPS, 20% 2,2'-(ethylenedioxy)diethanethiol).

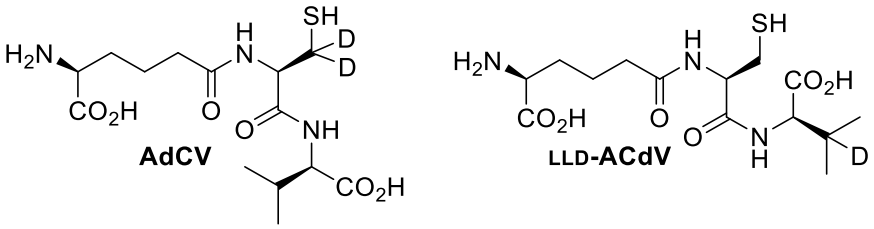



**Figure 4.6:** Selectively deuterated tripeptides **AdCV** and **ACdV**.

Preparation of **ACdV** by SPPS first required synthesis of Fmoc-3-[$^2$H]-Val-OH. A modified route to that reported by Baldwin and co-workers was used (Figure 4.7).[24] The route started from sodium α-ketoisovalerate, which was reacted under basic conditions in $D_2O$. Discontinuation of sodium α-ketoisovalerate as a commercial precursor by multiple chemical suppliers, led to alteration of the route. Thus, Grignard reaction of diethyl glyoxal **11** with $^i$PrMgCl as reported[25], was used to prepare ketone **12** (75%), which was used to obtain **13** under basic conditions. Deuterium incorporation (>98%) and ester hydrolysis of **12** to give **13** (74%), were achieved by stirring in $D_2O$ for 3 hours under basic conditions; a high level of deuterium incorporation (>98%) was confirmed by $^1$H NMR. Reductive amination of **13** with ammonium chloride and sodium cyanoborohydride, with crystallisation of the product from EtOH, afforded racemic deuterated Val **14** (31%).[24] Racemic Val **14** was acylated with acetic anhydride to give **16** in quantitative yield (Figure 4.7). Subsequent treatment with hog kidney acylase resulted in selective deacylation of the *N*$^a$-acyl group of the L-enantiomer of **15**.[24,26] The non-reacted *N*-acylated D-Val **15** was separated using a cation exchange column to give 80% of the theoretical maximum yield of the desired D-Val **15** enantiomer. Deprotection using HCl, followed by Fmoc protection in presence of sodium carbonate gave **D-16** which was achieved in a moderate yield (63%). Manual coupling of Fmoc-Val **D-16** to Wang resin was achieved using DIC, HOBt, and DMAP.[27] After SPPS and purification by

preparative-HPLC as described above (Section **3.2.2**), **LLD-ACdV** was obtained as a single diastereomer, with a high level (>98%) of deuterium incorporation (as analysed by $^1$H NMR).

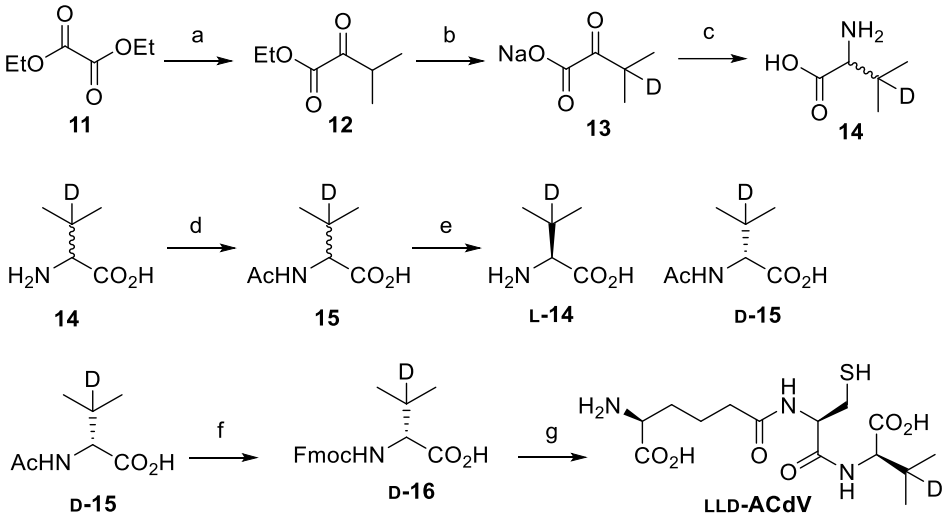



**Figure 4.7:** Route for the preparation of racemic Fmoc-[3d]-Val-OH **12**. *Reagents and conditions:* (a) $^i$PrMgBr, $Et_2O$ / $CH_2Cl_2$ (1 : 1), –10 °C, 2 h, 75%; (b) NaOD, $D_2O$, rt, 3 h, 74%; (c) $NH_4Cl$, $NaBH_3CN$, MeOH, 0 °C to rt, overnight, 31%; (d) $Ac_2O$, AcOH, 100 °C, 1 h, quantitative; (e) i) Hog kidney acylase, $H_2O$, rt, 21 h, 40%; (f) HCl, $H_2O$, reflux, 3 h, ii) Fmoc-OSu, $Na_2CO_3$, dioxane / $H_2O$ (1 : 1), 63%, overnight; (g) SPPS, see Figure 4.4; (h) Fmoc-OSu, $Na_2CO_3$, $H_2O$ 1,4-dioxane (1 : 1), rt, 24 h (63%); SPPS as described in Section **3.2.2**.

In an alternative route, racemic **14** was Fmoc protected to give **16** (Figure 4.8) and used in SPPS to afford a mixture of two diastereomers, **LLL-ACdV** and **LLD-ACdV** (Figure 4.8); these were separated by preparative-HPLC (*in collaboration with Dr Patrick Rabe, University of Oxford*) after synthesizing unlabelled **LLL-ACV** by SPPS). This route was found to be convenient for large scale preparation of **LLD-ACdV**, by shortening the initially developed route by 3 steps and avoiding enzymatic deacylation resolution on a large scale (Figure 4.8). For large scale preparation of **ACdV**, the $D_2O$ layer after deuterium incorporation was extracted to separate the double Grignard product **12b** from the crude solid before performing the reductive amination. Ethanol extraction after the reductive amination step separated most of the boronate salt from deuterated valine **14**. The scaled up route was a modified version of the protocol described above (Figure 4.8), which avoided large scale purifications until the step for the preparation of **16**.

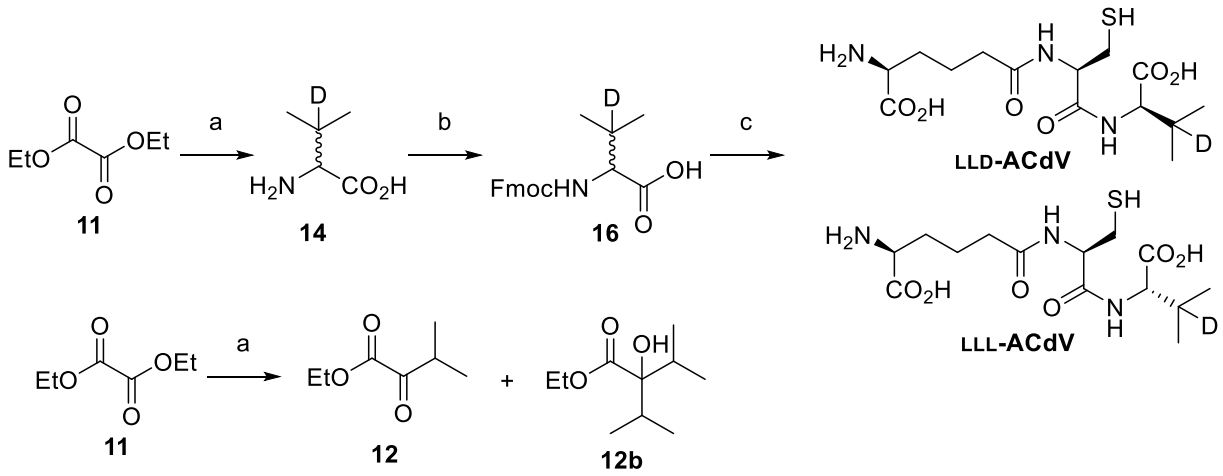



**Figure 4.8:** Route for the preparation of racemice Fmoc protected deuterated valine and the two corresponding diastereomeric ACV peptides thereof. *Reagents and conditions:* (a) i) $^{i}$PrMgBr, $Et_2O$ / $CH_2Cl_2$ (1 : 1), –10 °C, 2 h, ii) NaOD, $D_2O$, rt, 3 h, iii) $NH_4Cl$, $NaBH_3CN$, MeOH, 0 °C to rt, overnight; (b) Fmoc-OSu, $NaHCO_3$, 1,4-dioxane / $H_2O$ (1 : 1), 20-22%; (c) SPPS as described in Section 3.**2.2**.

For the time resolved crystallography studies a substantial amounts of **ACV** (~500 mg), **AdCV** (~100 mg), and **ACdV** (~80 mg) were prepared and purified (*prepared in collaboration with Dr Patrick Rabe, University of Oxford*). The purities of the **ACV**, **AdCV**, and **ACdV** were analysed by $^{1}$H NMR (Figure 4.9). Note the lack of Cys-β protons in the spectrum of **AdCV** and of a Val-β proton for **ACdV**, indicating a high level of $^{2}$H incorporation as detected by $^{1}$H NMR (Figure 4.9).

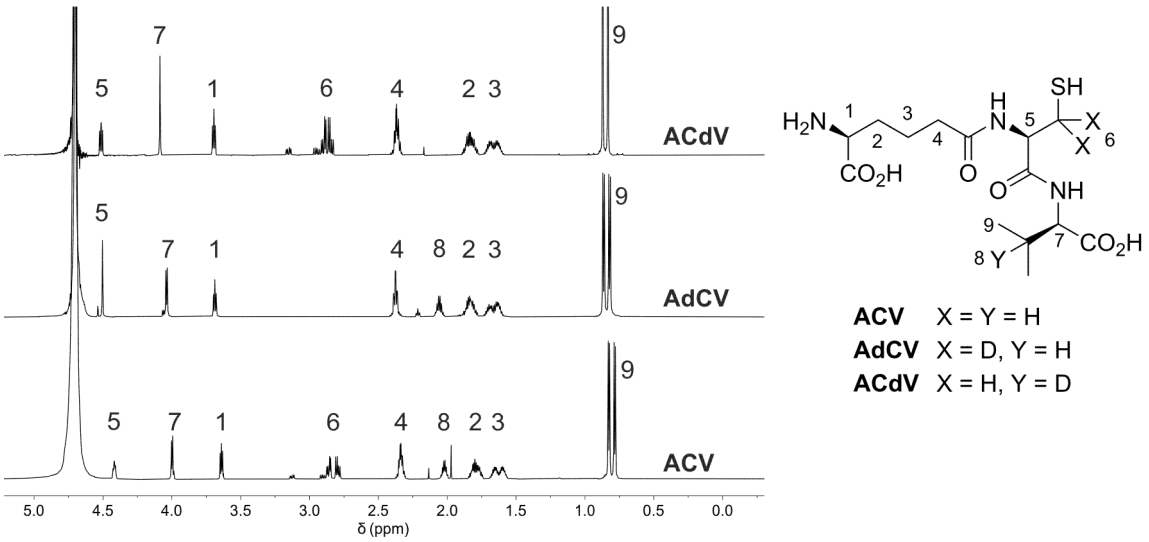



**Figure 4.9:** $^{1}$H NMR (700 MHz) analyses of **ACV**, **AdCV**, and **ACdV** recorded in $D_2O$. Resonances in the spectra are assigned to correspond with the structure.

### 4.2.5 Preparation of IPN

In order to undertake binding studies (Section **4.7**) of the **IPN** product to IPNS, **IPN** was prepared based on reported procedures employing deprotection by hydrogenation in the final step (Figure 4.10).[28–31] Thus, the α-amino group of Aad was protected with *p*-nitrobenzyl chloroformate, affording **17** in moderate yield (68%). Subsequent protection of the diacid **17** with PNB-Br gave a statistical mixture of α- and / or δ-PNB protected acids, with the desired product **18** being obtained in moderate yield (43%).[28,30,31] The commercially available 6-aminopenicillanic acid (**6-APA**) core was protected on its carboxylic acid using PNB-Br; the crude product was treated with *p*-toluenesulfonic acid to obtain **19** as a salt (57%).[29] EEDQ mediated coupling of ester **18** to protected **6-APA** core **19**, gave **20** (54%). Hydrogenation in presence of 10% Pd/C afforded crude **IPN** (79%), which was further purified by preparative-HPLC to give >95% purity as observed by $^1$H NMR (700 MHz).

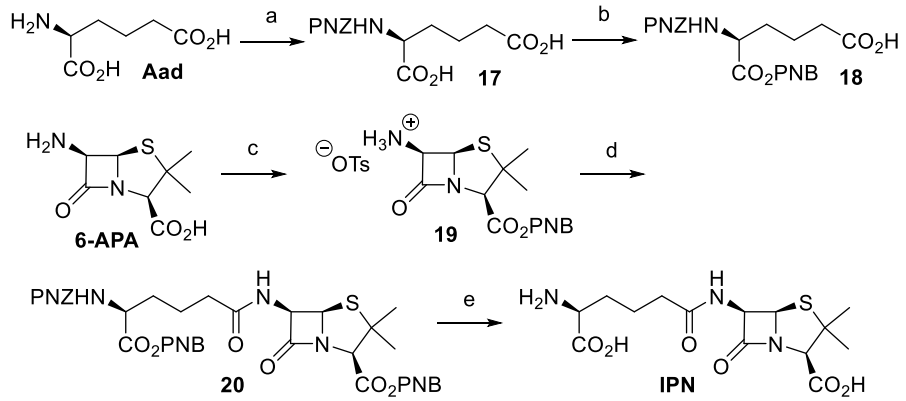



**Figure 4.10:** Route for the preparation of **IPN**. *Reagents and conditions:* (a) *p*-nitrobenzyol chloroformate, NaOH, 1,4-dioxane, rt, 30 min, 68%; (b) dicyclohexylamine, PNB-Br, DMF, 80 °C, 40 min, 43%; (c) i) $Et_3N$, *p*-nitrobenzyl bromide, ii) *p*-toluenesulfonic acid, acetone, rt, 30 min, 54%; (d) EEDQ, **18**, $Et_3N$, $CH_2Cl_2$, 20 h, 54%; (e) $H_2$ (g), Pd/C, THF / $H_2O$ (1 : 1), rt, 2 h, 79%.

### 4.2.6 Production of recombinant IPNS

*Glycerol stock stocks of A. nidulans IPNS and expression and purification protocols were provided by Dr Jürgen Brem (University of Oxford). Purification of IPNS was done in conjunction with Dr Patrick Rabe (University of Oxford).*

Expression of the gene encoding for *A. nidulans* wt IPNS was done using *E. coli* (NM554) cells, as reported using the pJB703 expression vector (Figure 4.11).[9] This vector places expression of the IPNS gene under a single *lac* promotor (P*lac*, Figure 4.11 (d)), producing high levels of soluble protein, without the use of IPTG.

In brief, recombinant IPNS in *E. coli* cells (NM554, chloramphenicol resistant) were transformed with the appropriate plasmid and grown in 2TY media (24 × 600 mL, 14.4 L) for production. Cells were grown for 16 hours, pelleted by centrifugation. The cells were suspended in buffer and lysed by sonication. Purification of the protein was achieved using a Q Sepharose FF anion exchange column, followed by a Superdex™ 75 size exclusion column, and a MonoQ anion exchange column. The purity of the protein was assessed by SDS-PAGE analysis (Figure 4.11).

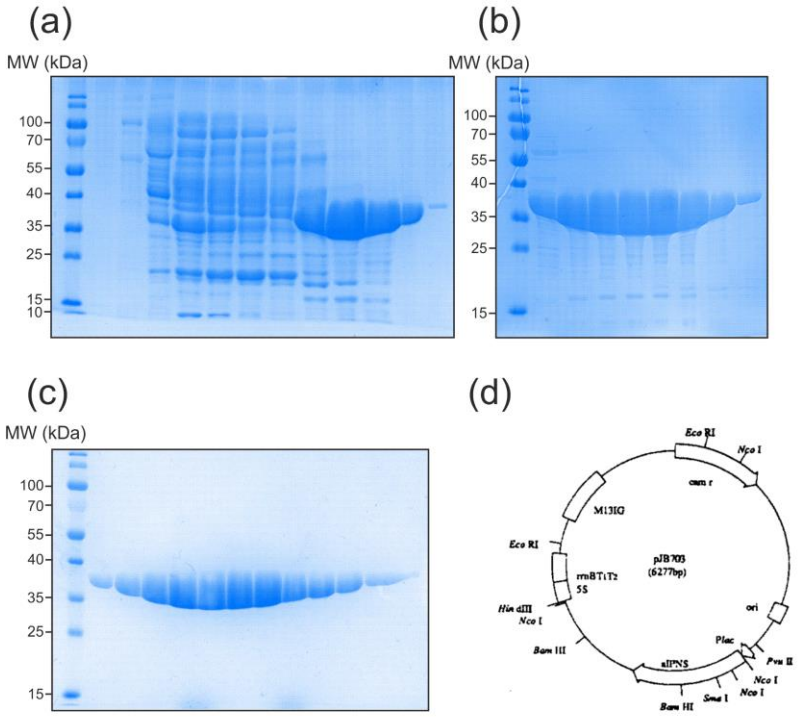



**Figure 4.11:** SDS-PAGE and plasmid map of vector pJB703 used for the expression of *A. nidulans* IPNS.[9] (a) SDS-PAGE of IPNS fractions after Q sepharose FF anion exchange chromatography step; (b) SDS-PAGE of IPNS fractions after Superdex™ 75 size exclusion chromatography step; (c) SDS-PAGE of IPNS fractions after MonoQ anion exchange chromatography step; (d) plasmid map of vector pJB703[9] used for the expression of *A. nidulans* IPNS. Molecular weight of IPNS is 37.5 kDa. Image (d) was reprinted with permission from Tetrahedron (Elsevier).

## 4.3 Time resolved X-ray crystallography

### 4.3.1 IPNS sample preparation at LCLS trip 1

*Crystallisation of IPNS microcrystals were performed together with Dr Thomas Leissing (University of Oxford) and the enzyme was prepared in collaboration with Dr Patrick Rabe (University of Oxford). The protocol for the production of IPNS • Fe • ACV microcrystal slurries were obtained from Dr Patrick Rabe (University of Oxford).*

Orthorhombic IPNS • Fe • ACV microcrystals with a needle morphology were crystallised according to optimised protocols (*provided by Dr Patrick Rabe, University of Oxford*), with stoichiometric amounts of Fe (1 equivalent), while shaking anaerobically in a 96 well plate (100 μL per well) in batches. Microcrystals appeared over 12 hours and continued to grow for 24 – 48 hours (Figure 4.12 (a)). The crystals were harvested by combining the well solutions into a single Eppendorf vial

. The crystals settled at the bottom of the tube upon standing (6 h) (Figure 4.12 (b)), and were concentrated by removing supernatant and re-suspending the pellet. Crystal density and size were determined (~1 mL, 2.75 – 3 · $10^8$ crystals · $mL^{-1}$, 50 – 90 μm) using an optical microscope (Section **3.2.2**). The dense crystal slurry was passed through a short piece of tubing (300 μm) and loaded in a gas-tight Hamilton syringe fitted with a gas lock, immediately prior to data collection. The syringe was then transferred into an anaerobic sandwich box and carried to the experiment hutch, to ensure the sample remains anaerobic (Figure 4.12 (c)). The syringe was connected to a tube and transferred to the DOD sample delivery system anaerobically, where it is dispensed onto the Kapton belt in a He environment. To prevent crystals from settling and creating a solid pellet, the syringe was rocked using a shaker (~ 20 s amplitudes). The sample was consumed at a rate of 6 μL · $min^{-1}$, creating ~3 nL droplets. The beam was operated with 40 fs pulses, at 9.5 keV, and data was collected at 30 Hz. The setup was enclosed in an anaerobic bag, filled with He during data acquisition (Figure 4.12 (d)).

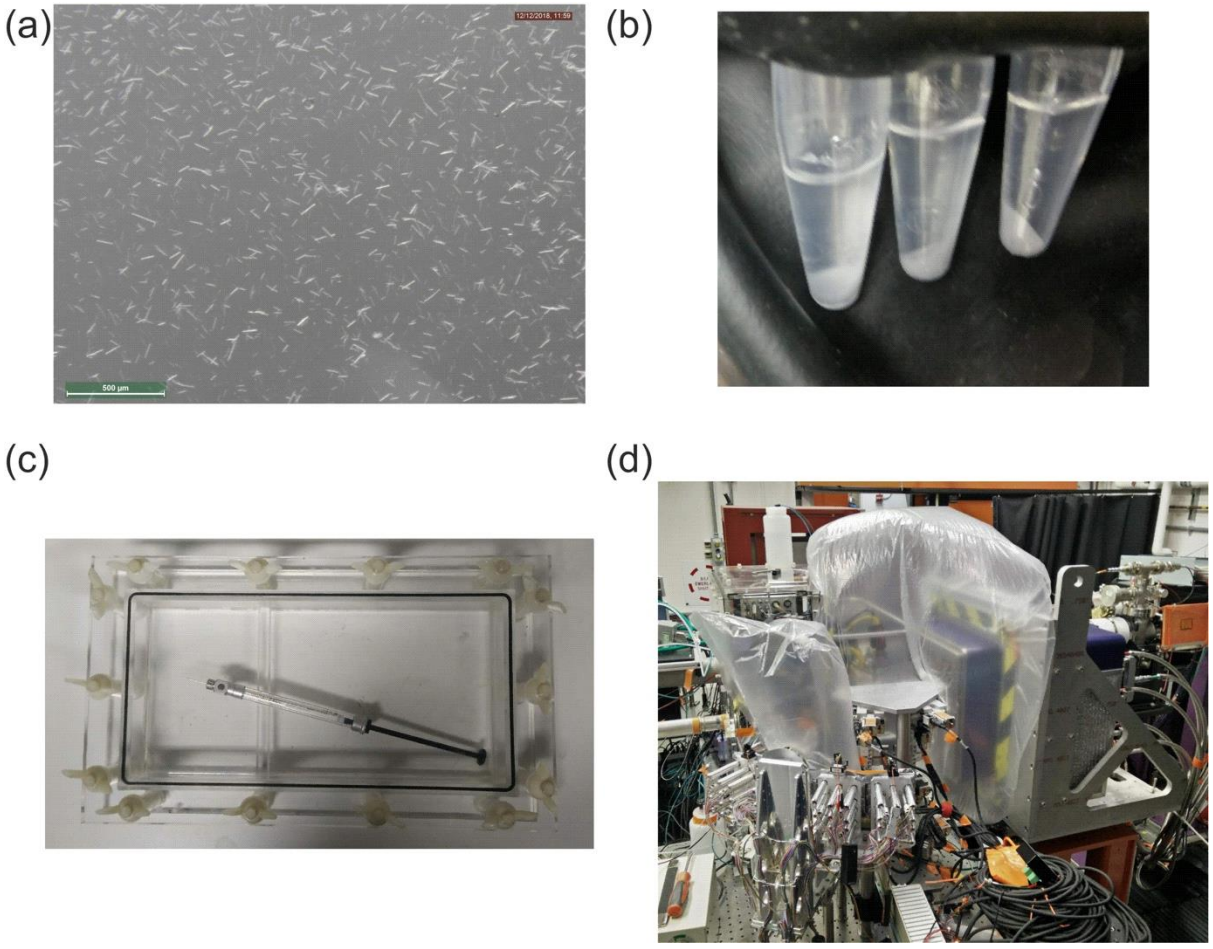



**Figure 4.12:** Prepared sample and experimental setup for IPNS crystals at Linac Coherent Light Source (LCLS, California, United States). (a) Visual inspection of matured IPNS • Fe • ACV crystals using an optical microscope; (b) pelleted crystals inside the glovebox, prior to concentrating and resuspending; (c) syringe containing IPNS microcrystal slurry in anaerobic sandwich box; (d) Experimental setup, with the sample delivery system to the left, crystals for the von Hamos spectrometer in the front, and the detector, enclosed in "the bag" filled with He gas.

### 4.3.2 Data acquisition

*The DOD sample delivery system was operated by Dr Jan Kern, Dr Cindy Pham, Dr In-sik Kim, Dr Joshua Heinemann, Dr Philipp Simon, and Isabel Bogacz, LBNL, Berkeley. Data processing was performed by Dr Pierre Aller and Dr Agata Butryn from DLS, Didcot; Dr Aaron Brewster, Dr Kyle Sutherlin, Dr Asmit Bhowmick, and Dr Iris Young, LBNL, Berkeley.*

An anaerobic dataset was collected for IPNS • Fe • ACV microcrystals to ensure the sample was not compromised during the handling or data collection. Once this had been confirmed, the IPNS • Fe • ACV

microcrystals were exposed to $O_2$ by allowing the droplets of crystal slurry to pass through an oxygen chamber (>95% $O_2$), initiating the reaction (Figure 4.13, oxygen chamber is placed at the "reaction region").[32] The belt speed was adjusted to allow the droplets to incubate for 6 and 3 s on the tape, before being probed by the X-ray beam. Unfortunately, for both the 6 and 3 s exposures, it seems likely that the dry atmosphere caused the droplets to dehydrate allowing salt crystals to form. Consequently, the overall quality of the data was poor. A complete data set was collected on crystals with a total incubation time of 500 ms. The remaining sample was incubated for 2.5 s, suffering less from dehydration than compared to previous attempts at 6 and 3 s. Diffraction to the edges of the detector (1.4 Å) was observed, however for the final data set the resolution was cut at 2.4 Å due to low completeness in the higher resolution bins. Simultaneously, Kα fluorescence spectroscopy data were collected for each of the time points, however no real time feedback was available for this experiment.

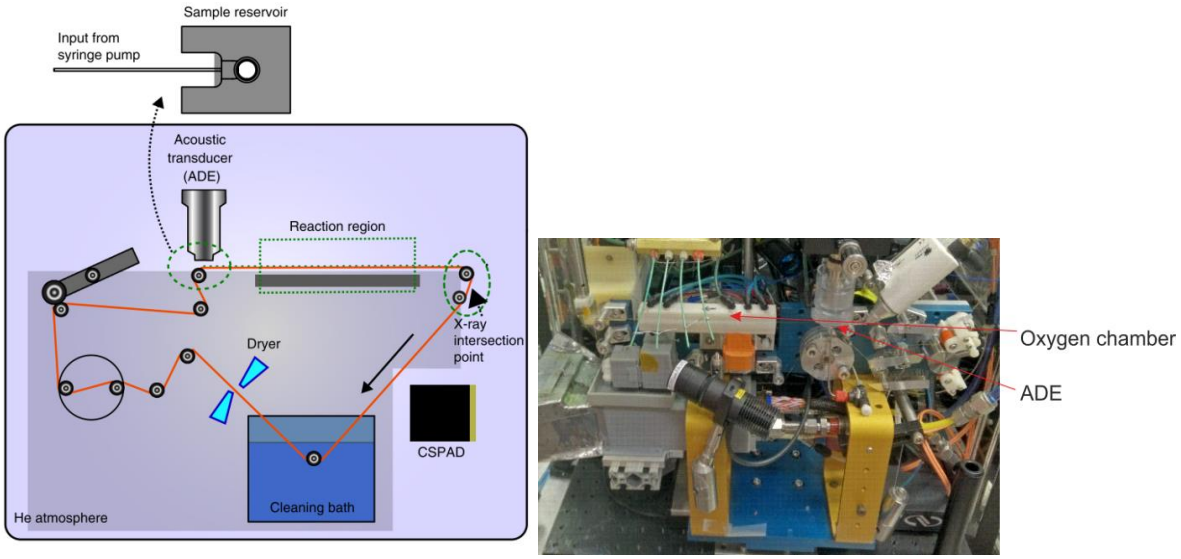



**Figure 4.13:** Schematic representation of the drop on demand sample delivery system, reprinted with permission from Nature Methods (Nature publishing group).[32] (a) Microcrystal slurry is continuously supplied to the sample reservoir through a capillary attached to an anaerobic syringe. The acoustic droplet ejector (ADE) dispenses droplets containing micro crystals at a rate of 30 Hz onto the polyimide belt. The conveyor belt passes the droplets through the reaction region (oxygen chamber, containing >95% $O_2$), and presents the droplets to the X-ray beam at the X-ray intersection point. The tape is washed and dried *in situ* before transferring new droplets. (b) Picture of the oxygen chamber.

Online data processing of the data collected on IPNS • Fe • ACV microcrystals indicated high hit (Figure 4.14 (a), 85.6 – 96%) and indexing hit (idx, Figure 4.14 (a), 19.6 – 25.7%) rates. Indexing hit rates vary depending on the sample, a hit rate of 10 – 35% being considered reasonable.[32] Indexing hit rates for data collection during LCLS trip 2 (discussed below, Section **4.3.5**) were higher, with a hit rate of up to 54%, presumably due to higher crystal density in the slurry. Only the indexed images were used in the final dataset. The data collected for each sample was obtained from many different crystals. These crystals had different dimensions and varied in their unit cell parameters (Figure 4.14 (b)). The measured cell parameters were refined, along with the detector distance and geometry, to minimise deviations and enable data from different runs to be merged.

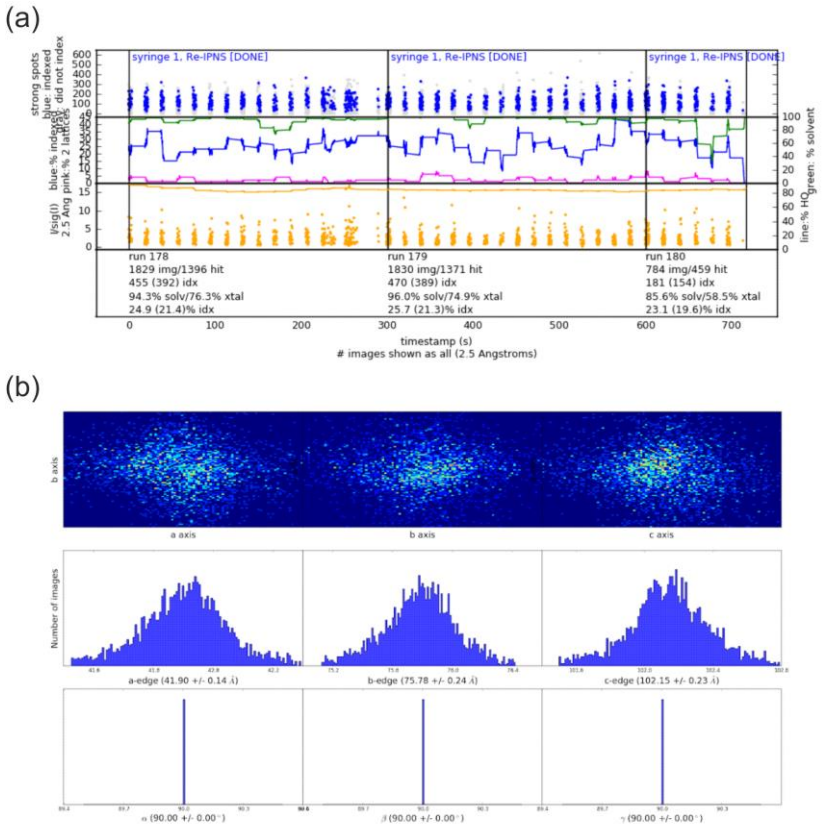



**Figure 4.14:** Overview of different aspects of the processing of serial crystallographic data obtain from microcrystals. (a) Statistical analyses of data collected on IPNS • Fe • ACV microcrystals in resting state. Data were collected in separate runs, indicated by the box. Top panel: Blue spots were indexed, grey spots were not indexed; middle top panel: green line represents images identified to containing solvent, blue line represents the number of images containing spots that are indexed, pink line represents images containing multiple lattices; middle bottom: spot intensity; bottom panel: summary of data with run numbers. (b) Analyses of unit cell constants of IPNS • Fe • ACV microcrystals in resting state. *Image was prepared by Dr Pierre Aller and Dr Agata Butryn from DLS, Didcot.*

### 4.3.3 X-ray emission spectroscopy data collection

*XES data acquisition and processing were done by Dr Jan Kern, Dr Kyle Sutherland, Dr Cindy Pham, LBNL, Berkeley.*

Emission spectroscopy on Fe (Kα) was collected for all samples. The results from the data sets under anaerobic, 500 ms incubation with $O_2$, and 2.5 s incubation with $O_2$ samples, were compared (Figure 4.15). The XES spectrum depicts two peaks, $K\alpha_1$ (right) and $K\alpha_2$ (left), and is normalized for $K\alpha_2$. The spectra are then analysed for changes in width (and thus height) in $K\alpha_1$, which are due to changes in the effective number of 3d electrons in Fe. IPNS • Fe • ACV microcrystals under anaerobic conditions were determined to contain Fe(II) (Figure 4.15, blue line). Exposure of the IPNS • Fe • ACV microcrystals to $O_2$ (500 ms) led to a change in $K\alpha_1$ line, implying a change in oxidation state to Fe(III). Further incubation of IPNS • Fe • ACV microcrystals to $O_2$ (2.5 s) appeared to change the oxidation state to Fe(II). These results were used to support interpretation of the diffraction data collected from the different time points.

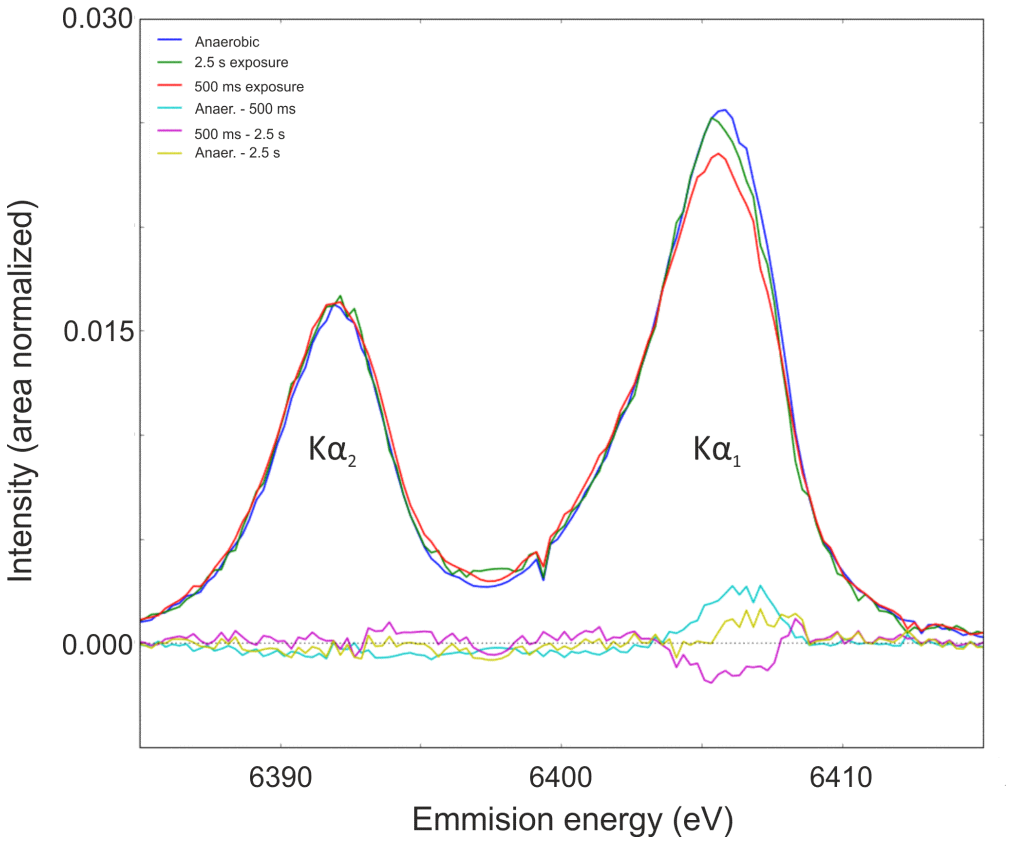



**Figure 4.15:** Kα lines derived from Fe present in IPNS • Fe • ACV microcrystals. The $K\alpha_1$ (right peak) is sensitive to changes in the effective number of 3d electrons. Kα spectra of IPNS • Fe • ACV anaerobic (blue), IPNS • Fe • ACV exposed to $O_2$ (500 ms, red), IPNS • Fe • ACV exposed to $O_2$ (2.5 s, green). Subtracting of the spectra was used to obtain the differences between the differently $O_2$ exposed states. *Data were analysed by Dr Cindy Pham, Dr Kyle Sutherland and Dr Jan Kern.*

### 4.3.4 IPNS anaerobic room temperature structure

For the resting state XFEL structure of IPNS a total of 7415 images (36% of total images) divided over 10 runs were merged, providing a structure complete to a resolution of 1.4 Å (Figure 4.16 (a)). The structure was solved by molecular replacement (*Dr Patrick Rabe, University of Oxford*, PDB accession code 1BLZ[3]). The refined room temperature structure of IPNS shows clear density corresponding to the substrate and metal in the active site (Figure 4.16 (a) and (b)). The conformation of the substrate inside the active site and overall protein fold are similar to reported IPNS cryo-structures, with an RMSD of 0.19 Å (Figure 4.16 (c), 1BLZ[3]). Possibly as a consequence of the structure of IPNS being obtained at room temperature, it exhibits higher B-factors (27.9 and 35.5 Å$^2$ average for the enzyme and **ACV** respectively) compared to a reported structure obtained under cryo-cooled conditions (PDB accession code 1BLZ[3], 9.9 and 9.7 Å$^2$ average for the enzyme and **ACV** respectively[33]). In agreement with previous reports, the Fe is coordinated hexagonally by His214, Asp216, His270, a $H_2O$ and the cysteinyl thiolate of **ACV**, leaving one vacant site available. The Fe bound water occupies the site *trans* to His214. The aminoadipoyl carboxylate is positioned to form a salt bridge with Arg87, anchoring the aminoadipoyl chain in a fixed position. The **ACV** valine carboxylate is oriented away from Fe and is forming hydrogen bonds with Ser281 and Trp189, and a water mediated hydrogen bond with Arg279. These results confirm it is possible to obtain XFEL structural information about IPNS in complex with **ACV** and iron, from anaerobically grown IPNS microcrystals at a high resolution.

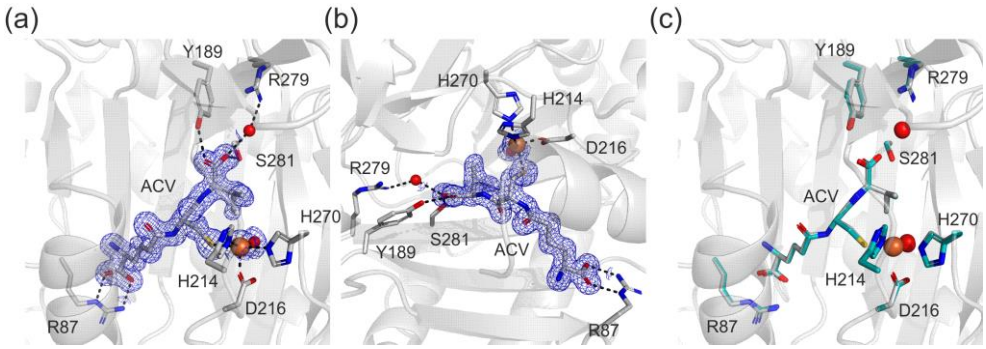



**Figure 4.16:** Views from an XFEL crystal structure obtained from IPNS • Fe • ACV microcrystals under anaerobic conditions. Crystal structure of IPNS in complex with Fe(II) and ACV. (a) View of the active site of IPNS showing Fe (orange), the active site residues and **ACV**; alternative view of the active site of IPNS showing Fe, the active site residues and **ACV**; (c) view of the active site of IPNS showing the **ACV** substrate and residues interacting with the substrate (grey), compared with the reported structure obtained under cryo-cooled conditions (teal, 1BK0).[34] The 2mFo-Fc maps (blue) are contoured to 1 σ. *Refinement and data analysis were done by Dr Patrick Rabe (University of Oxford).*

| **Table 4.1 Data collection and refinement statistics of IPNS • Fe • ACV under anaerobic conditions.** | |
|---|---|
| **Data set** | **IPNS • Fe • ACV** |
| Data colleection | |
| Beamline (Wavelength, Å**)** | LCLS (1.265) |
| Detector | Rayonix 340 |
| Data processing | cctbx.xfel |
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| *a,b,c* (Å**)** | 41.8, 75.9, 101.9 |
| α,β,γ (°) | 90, 90, 90 |
| No of molecules/ASU | 1 |
| No. of reflections | 60277 (2602)[a] |
| Resolutions (Å) | 19.93 – 1.4 (1.45 – 1.4)[a] |
| $R_{split}$ | 24.5 (143)[a] |
| I/δI | 20.2 (0.21)[a] |
| CC1/2 | 94.3 (--4.0)[a] |
| Completeness (%) | 100 (99.9)[a] |
| Multiplicity | 41.0 (8.5)[a] |
| Wilson B value (Å$^2$) | 20.23 |
| **Refinement** | |
| $R_{work}/R_{free}$ | 0.1761 / 0.2149 |
| No. atoms | |
| - Enzyme | 2685 |
| - Ligand | 30 |
| - $H_2O$ | 190 |
| Avg. B-factors | |
| - Enzyme | 27.9 |
| - Ligand | 35.5 |
| - $H_2O$ | 38.9 |
| R.m.s. deviations | |
| - Bond lengths (Å) | 0.009 |
| - Bond angles (°) | 0.96 |
| [a] Highest resolution shell in parentheses | |

### 4.3.5 IPNS sample preparation at LCLS trip 2

*Preparation of IPNS microcrystals was performed in collaboration with Dr Patrick Rabe (University of Oxford). The DOD sample delivery system was operated by Dr Jan Kern, Dr Cindy Pham, Dr In-sik Kim, Dr Joshua Heinemann, Dr Philipp Simon, and Isabel Bogacz, LBNL, Berkeley. Data processing was performed by Dr Pierre Aller and Dr Agata Butryn from DLS, Didcot; Dr Aaron Brewster, Dr Kyle Sutherlin, Dr Asmit Bhowmick, and Dr Iris Young, LBNL, Berkeley.*

After optimisation at the Stanford Synchrotron Radiation Lightsource (SSRL), microcrystals of IPNS were crystallised as indicated above, while shaking in a 96 well plate (100 µL per well) in batches under anaerobic conditions. Crystals of IPNS • Fe • ACV, and IPNS • Fe • AdCV were successfully obtained (Figure 4.17 (a) and (b)). Attempts to crystallise IPNS with **ACdV** were limited in success and were not available in sufficient quantities for data collection. Crystals of IPNS • Fe • ACV appeared in 6 – 12 hours and continued growing for 24 – 48 hours. The mature crystals were harvested and concentrated by settling as discussed above (Section **3.3.7**). The final crystal slurry was analysed (3 batches, 1 mL per batch, 0.4 – 1.6 · $10^8$ crystals · $mL^{-1}$, 60 – 90 µm). The crystal slurry was supplemented with glycerol (10 v/v%) to

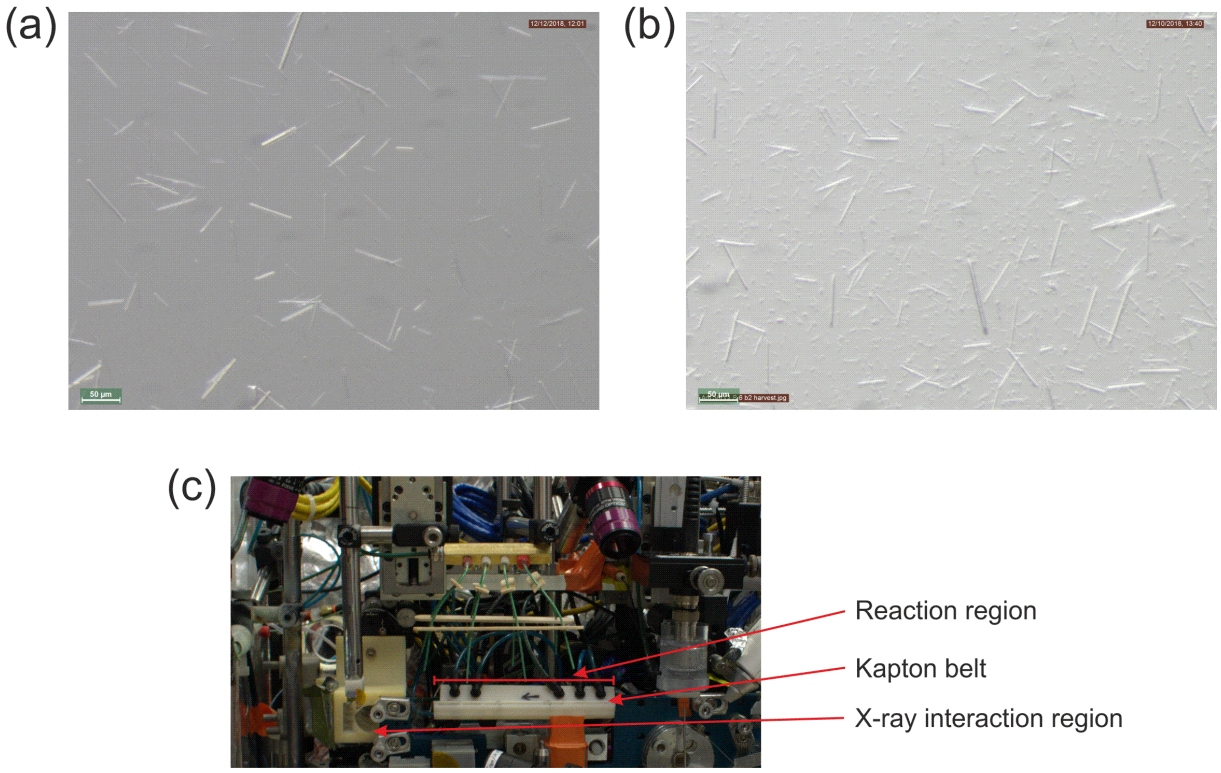



**Figure 4.17:** Optical microscope images of IPNS crystals with **ACV** and **AdCV**, and image of drop on demand setup. (a) Optical microscope image of IPNS • Fe • ACV microcrystals; (b) optical microscope image of IPNS • Fe • AdCV microcrystals; (c) image of the drop on demand sample delivery system: in white is the oxygen chamber, the green tubes surrounding the oxygen chamber are used to increase local humidity.

prevent drying out of the crystals inside the bag. Visual inspection using an optical microscope suggested that the droplets diluted with glycerol (10%) took longer to dry out (~30% longer). Additionally, the He inside the closed bag, was circulated and passed through a water bath in order to humidify the environment. Tubing was used to dispense humidified He around the tape (Figure 4.17 (c), green pieces of tubing). Samples of IPNS • Fe • ACV and IPNS • Fe • AdCV were consumed at a rate of 7 μL · $min^{-1}$, creating 4 nL droplets. The beam was operated with 35 fs pulses, at 9.5 keV, and data collection was done at 30 Hz. Diffraction to a resolution of 1.4 Å was observed for all datasets, with indexed hit rates up to 54%. Approximately 70 – 80 μL of slurry were consumed per data set, which contained 10,000 – 17,000 merged images, and were collected efficiently in 15 – 20 minutes. The $K\alpha_1$ fluorescence spectroscopy data collected in parallel for each of the datasets was inconclusive due to low signal to noise (potentially due to the setup which contained only a single usable germanium crystal in the von Hamos setup). The following total incubation times were used during the data collection with crystals of IPNS • Fe • ACV: anaerobic, 400 ms,

500 ms, 800 ms, 1.6 s, 3 s, 6 s, 8 s. The following total incubation times were used during the data collection with crystals of IPNS • Fe • AdCV: 1.6 s, 3 s, 6 s, 8 s, 1 min. A method called in-well mixing uses a buffer line to dilute the crystal slurry in the well which is supplied with anaerobic sample. The well is situated immediately before the ADE. By diluting the anaerobic crystal slurry with oxygen saturated buffer (1 : 1 ratio), a time point of approximately 1 min is achieved. The belt speed was set to a default speed, incubating the droplets for 500 ms on the belt.

### 4.3.6 Structure of IPNS incubated with $O_2$

After the initial successes obtaining resting state and $O_2$ exposed crystal structures as describe above, new time points were selected for probing. Crystals were prepared anaerobically and diluted with glycerol (10%, under anaerobic conditions) to slow dehydration of the droplets on the belt. However, the addition of glycerol affects the solubility and diffusion of $O_2$ into the droplets,[35] likely indirectly slowing down the rate of the reaction. In the experimental setup employed, after mixing with glycerol under anaerobic conditions, the sample was transferred into a Hamilton syringe and transported inside an anaerobic sandwich box. Immediately prior to the tr-SFX experiment, the syringe was removed from the sandwich box and connected to the DOD sample delivery system. The sample is then transferred through a capillary under

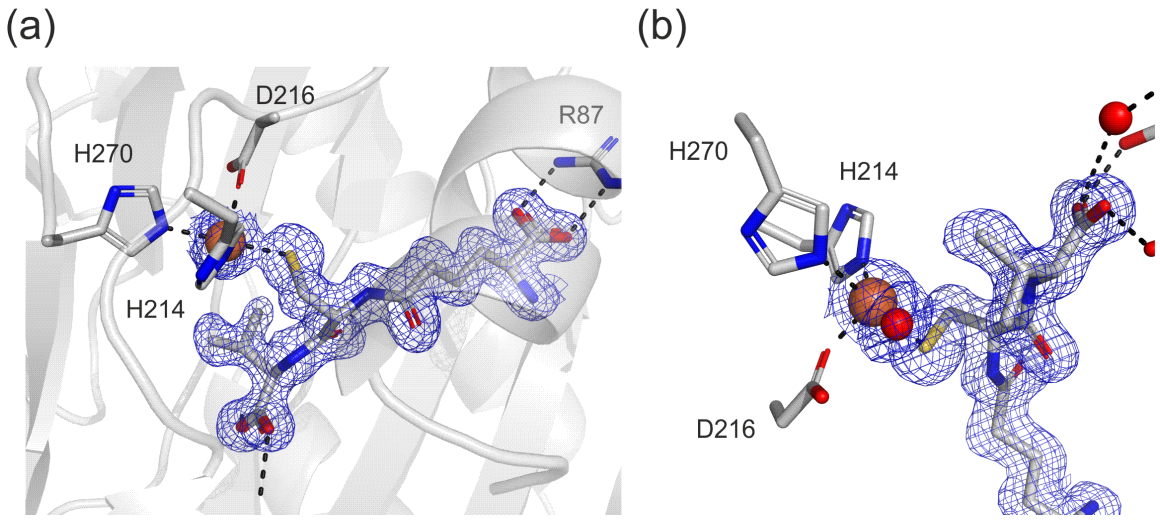



**Figure 4.18:** View from a crystal structure of IPNS • Fe • ACV obtained after exposing microcrystals to $O_2$ for 800 ms. (a) View of the active site of IPNS showing Fe (orange sphere), the active site residues and **ACV**; (b) view of **ACV** bound to Fe metal via the cysteine thiolate. The 2mFo-Fc maps (blue) are contoured to 1 σ. *Refinements and data analysis was done by Dr Patrick Rabe (University of Oxford).*

anaerobic conditions, and dispensed on the Kapton tape of the DOD system in a He environment. Exposure of the IPNS • Fe • ACV microcrystals to $O_2$ is achieved by passing through the $O_2$ incubation chamber, which is supplied by pure $O_2$ from a cylinder. The electron density maps obtained from IPNS microcrystals exposed to $O_2$ for 800 ms, show substrate and Fe bound in the active site (Figure 4.18). The density observed for the **ACV** substrate is clear with minimal changes compared to the anaerobic structure, with RMSD of 0.06 Å over the backbone peptide.

When exposing the IPNS microcrystals to $O_2$ for 1600 ms, new density was observed at the Fe centre, opposite to the metal ligating Asp216 (Figure 4.18). Molecular $O_2$ was modelled into the new electron density in the position *trans* to Asp216 in an "end-on" ($\eta^1$) fashion to form an octahedral complex (Figure 4.19 (a) and (b)). Binding of $O_2$ to Fe happens at approximately the same coordination site as reported for nitric oxide (NO) in the structure of IPNS • Fe • ACV • NO,[3] however, the Fe-O-O angle (164°) is different from the Fe-N-O angle (120°) and the displacement of the isopropyl group of **ACV** is different (Figure 4.19 (c)). The new density was modelled with $O_2$ at an occupancy of 80%. Analyses of the B-factors of both O-atoms in the bound $O_2$ are similar (40.8 and 41.5 $Å^2$ for Fe-**O**-O, and Fe-O-**O** respectively), but are higher than the average for the enzyme (30.2 $Å^2$). The appearance of $O_2$ correlated with a loss in electron density for the **ACV** Val isopropyl group, which was modelled at 20%. The majority (80%) of the observed electron density was refined with the missing isopropyl methyl groups (because of potential delocalization during the catalysis process) at the position of the **ACV** Val. The density observed for Cys and Aad part of **ACV** remained clear. Comparing the IPNS • Fe • ACV structure obtained under anaerobic conditions and exposed to $O_2$ for 1600 ms showed an overall similar fold with an RMSD of 0.10 Å over the backbone peptide. Additional electron density was observed for a region covering α-helix 3 (residues 49 – 65) and three residues situated on β-strand 13 (residues 280 – 282) (Figure 4.19 (d), (e) and (f)). This suggests that the residues present on α-helix 3 and β-strand 13 are present in two different conformation (Figure 4.19 (e) and (f), arrows). In parallel XES data was collected on all samples, however, analysis of the data suggested too little was collected to provide conclusive evidence for a changes in the oxidation state of Fe.

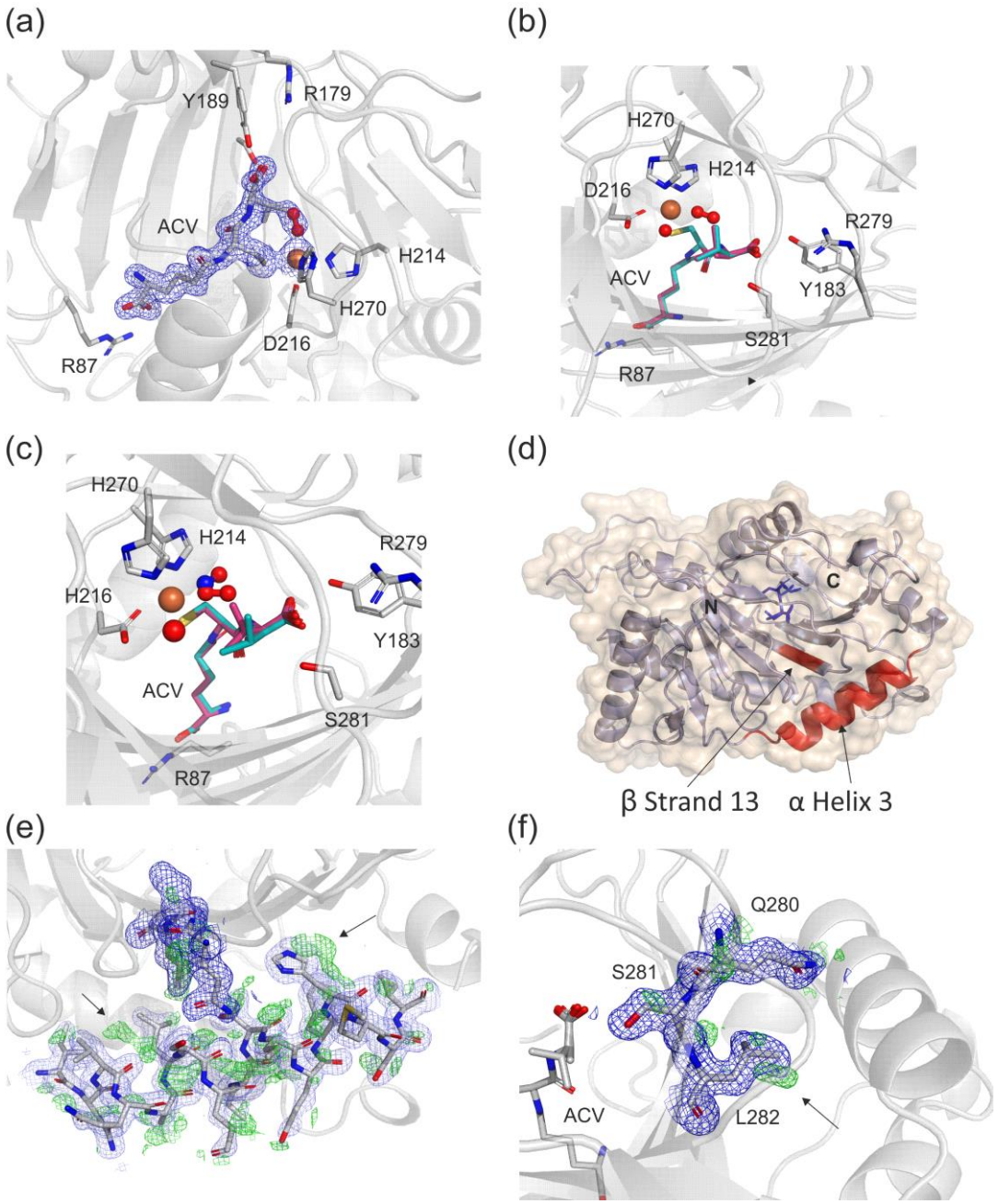



**Figure 4.19:** View from a crystal structure of IPNS • Fe • ACV obtained after exposing microcrystals to $O_2$ for 1600 ms. (a)View from the active site of IPNS showing Fe, the active site residues and **ACV**; (b) view of **ACV** bound to Fe metal via the cysteine thiolate, and additional density at the metal centre which is modelled as $O_2$; (c) overlay of a structure from IPNS • Fe • ACV microcrystals exposed to $O_2$ for 1600 ms (pink), and a structure from IPNS • Fe • ACV macrocrystals exposed to nitric oxide (NO) (teal, PDB accession code 1BLZ). (d) view of α-helix 3 (residues 49 – 65) and β-strand 13 (residues 280 – 282); (e) view of residues 49 – 65 on α-helix 3; (f) view of residues 280 – 282 present on β-strand 13. The 2mFo-Fc maps (blue) are contoured to 1 σ, and mFo-Fc OMIT maps (green) are contoured to 2.5 σ. *Refinements and data analysis were done by Dr Patrick Rabe (University of Oxford).*

| **Table 4.2 Data collection and refinement statistics of IPNS • Fe • ACV incubated with $O_2$.** | | |
|---|---|---|
| **Data set** | **IPNS • Fe • ACV $O_2$ incubation (800 ms)** | **IPNS • Fe • ACV $O_2$ incubation (1600 ms)** |
| Data colleection | | |
| Beamline (Wavelength, Å**)** | LCLS (1.265) | LCLS (1.265) |
| Detector | Rayonix 340 | Rayonix 340 |
| Data processing | cctbx.xfel | cctbx.xfel |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell dimensions | | |
| *a,b,c* (Å**)** | 41.9, 75.9, 101.9 | 41.9, 75.8, 101.8 |
| α,β,γ (°) | 90, 90, 90 | 90, 90, 90 |
| No of molecules/ASU | 1 | 1 |
| No. of reflections | 60581 (2428)[a] | 59409 (2474)[a] |
| Resolutions (Å) | 20.29 – 1.4 (1.45 – 1.4)[a] | 19.93 – 1.41 (1.46 – 1.41)[a] |
| $R_{split}$ | 17.0 (142)[a] | 18.6 (138)[a] |
| I/δI | 26.5 (0.22)[a] | 24.3 (0.14)[a] |
| CC1/2 | 99.7 (--6.6)[a] | 96.0 (0.0)[a] |
| Completeness (%) | 100 (100)[a] | 100 (100)[a] |
| Multiplicity | 81.0 (12.2) | 62.6 (9.8)[a] |
| Wilson B value (Å²) | 20.3 | 21.5 |
| **Refinement** | | |
| $R_{work}/R_{free}$ | 0.1631 / 0.1985 | 0.1748 / 0.2116 |
| No. atoms | | |
| - Enzyme | 2670 | 2642 |
| - Ligand | 30 | 45 |
| - $H_2O$ | 192 | 166 |
| Avg. B-factors | | |
| - Enzyme | 29.6 | 30.2 |
| - Ligand | 35.2 | 39.8 |
| - $H_2O$ | 41.6 | 42.1 |
| R.m.s. deviations | | |
| - Bond lengths (Å) | 0.008 | 0.005 |
| - Bond angles (°) | 0.89 | 0.70 |
| [a] Highest resolution shell in parentheses | | |

Exposure of the IPNS • Fe • ACV microcrystals to $O_2$ for 3 s resulted in an increased intensity of the electron density at α-helix 3 and β-strand 13, suggesting a higher occupancy for the alternative conformation of α-helix 3 and β-strand 13, compared to the structure of IPNS • Fe • ACV exposed to $O_2$ for 1.6 s. The structure of IPNS • Fe • ACV exposed to $O_2$ for 3 s, is currently under refinement (*done by Dr Patrick Rabe, University of Oxford*) and being modelled with an additional alternative conformation for α-helix 3 and β-strand 13, albeit at low occupancy (~15 – 25%). The time dependence of the appearance of the additional electron density at α-helix 3 and β-strand 13 implies a connection to the $O_2$ content at the active site. However, the occupancies modelled for $O_2$ at the active site, do not match the (low) occupancies for the alternative conformation of α-helix 3 and β-strand 13. Instead, the alternative conformation of α-helix 3 and β-strand 13 always contribute to a minor fraction of the model, and are present together with the conformation of α-helix 3 and β-strand 13 observed under anaerobic conditions.

The exposure of IPNS • Fe • ACV microcrystals to $O_2$ coincides with a change in unit cell dimensions (Figure 4.20). Although a distribution of unit cell dimensions is typically obtained for each structure

(Section **3.3.3**), variations between different samples are generally minor (Figure 4.20, blue and red bars). Along the b-axis, where the difference in length is largest, a change of ~0.3 Å (~0.4%) is observed, between the anaerobic (Figure 4.20, blue and red bars) and $O_2$ exposed (Figure 4.20, 3 s, green bars) crystals. A small yet measurable change is observed. At present it is unclear if the observed change in unit cell dimension is connected to the additional alternative conformation for α-helix 3 and β-strand 13, as opposed to a change in experimental conditions (e. g. associated with dehydration). Note, these changes were observed in samples obtained from identical crystal slurry batches; the only difference being in the exposure time to $O_2$.

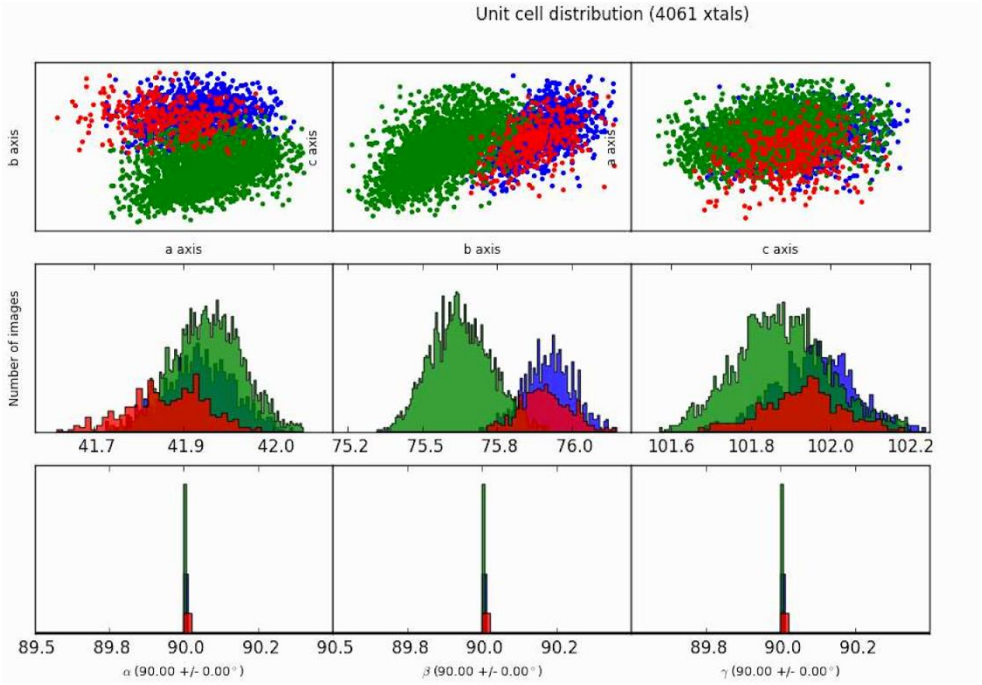



**Figure 4.20:** Analysis of unit cell parameters in IPNS • Fe • ACV microcrystals exposed to $O_2$ for different durations. Comparison of two anaerobic samples from the same IPNS • Fe • ACV microcrystal batch (blue and red bars), and a sample of IPNS • Fe • ACV microcrystals (green bars) exposed to $O_2$ (3 s). *Image was prepared by Dr Pierre Aller and Dr Agata Butryn from DLS, Didcot.*

### 4.3.7 Structure of IPNS in complex with IPN

For the XFEL experiments it was considered important to show complete turnover of the IPNS • Fe • ACV complex to give a product complex or product released structure. Even though crystallisation of IPNS was first reported more than two decades ago, to date no high occupancy IPNS • Fe • IPN structure has been reported. Attempts to turn over the substrate *in crystallo* by exposure to hyperbaric $O_2$ (40 bar, 320 min, PDB accession code 1QJE) by Baldwin and co-workers,[5] were in part successful and gave partial occupancies of **IPN** in the active site.[5] However, analysis of the electron density at the active site suggests. that the crystals still contain 30% IPNS • Fe • ACV. Attempts with prolonged exposure to hyperbaric $O_2$ (40 bar, up to 640 min) were unsuccessful; low electron density at the active site was observed, suggesting slow release of the **IPN** product. Given the need for microcrystals for XFEL work, attempts to turn over substrate in the IPNS • Fe • ACV complex in microcrystal slurries were carried out. IPNS was crystallised under anaerobic conditions as discussed above (*by Dr Patrick Rabe, University of Oxford,* Section **4.3.1**). The IPNS • Fe • ACV microcrystal slurry was mixed with LCP under anaerobic conditions, and subsequently exposed to air (12 h), followed by crystallographic analysis using XFEL radiation (SACLA, Japan). After refinements (*carried out by Dr Thomas Leissing and Dr Patrick Rabe*), analyses of the electron density maps showed complete turnover of the **ACV** substrate to the **IPN** product (Figure 4.21), which unlike the reported structure, does not appear to dissociate from the active site over time. Subsequent turnover IPNS • Fe • IPN (cryo-)structures and fixed target room temperature structures (notably, all obtained using crystals showing needle morphology) analysed using synchrotron radiation, similarly showed 100% occupancy for **IPN**, even after 12 weeks of exposure to $O_2$. These observations suggest that in the particular experimental setup, i. e. using microcrystal slurries of crystals with the needle morphology, the product of the catalytic reaction does not leave the active site. This peculiarity of these microcrystals was exploited to enable the determination of an **IPN** product bound IPNS structure. The crystals used in the present study have the same space group ($P2_12_12_1$) as the structure reported by Baldwin and co-workers,[4] but exhibit a different morphology (needle morphology for microcrystals, plate morphology for reported structure). The overall fold of the protein in the reported structure and the structure obtained by turnover of the IPNS • Fe • ACV complex, show great similarity with an RMSD of 0.29 Å over the peptide backbone. The calculated solvent content in the crystals in the present study (43%)

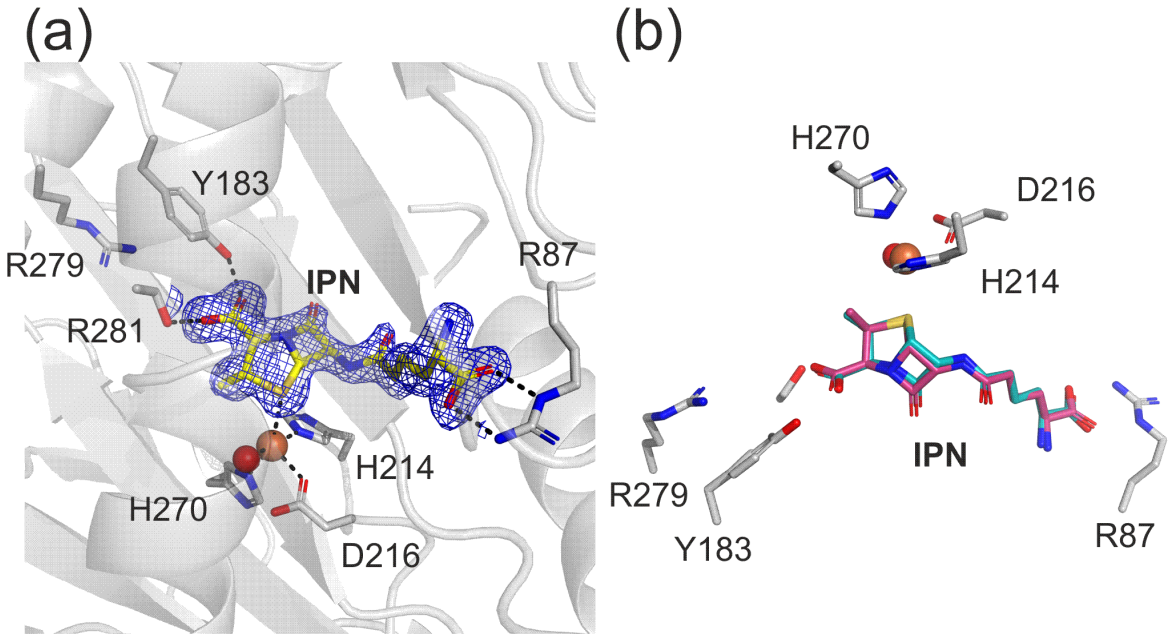



**Figure 4.21:** View from a crystal structure of IPNS • Fe • IPN obtained by turnover of IPNS • Fe • ACV microcrystals by exposure to $O_2$ from air over 12 h. (a) View from the active site of IPNS, showing **IPN** bound; (b) view from the active site of IPNS, showing **IPN** (pink), overlaid with **IPN** from a reported structure (PDB accession code 1QJE)[5]. The 2mFo-Fc maps (blue) are contoured to 1 σ. *Refinements and data analysis was done by Dr Thomas Leissing and Dr Patrick Rabe (University of Oxford).*

| **Table 4.3 Data collection and refinement statistics of IPNS • Fe • IPN obtained by turnover of IPNS • Fe • ACV by exposure to $O_2$ for 12 h.** | |
|---|---|
| **Data set** | **IPNS • Fe • ACV** |
| Data colleection | |
| Beamline (Wavelength, Å**)** | SACLA (1.137) |
| Detector | MPCCD |
| Data processing | DIALS |
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| *a,b,c* (Å**)** | 42.0, 75.6, 101.3 |
| α,β,γ (°) | 90, 90, 90 |
| No of molecules/ASU | 1 |
| No. of reflections | 43794 (4319)[a] |
| Resolutions (Å) | 42.37 – 1.6 (1.657 – 1.6)[a] |
| $R_{split}$ | 0.128 (0.519)[a] |
| I/δI | 43.7 (2.81)[a] |
| Completeness (%) | 99.9 (100)[a] |
| Multiplicity | 110.6 (37.3)[a] |
| Wilson B value (Å$^2$) | 20.64 |
| **Refinement** | |
| $R_{work}/R_{free}$ | 0.1757 / 0.2062 |
| No. atoms | |
| - Enzyme | 2613 |
| - Ligand | 30 |
| - $H_2O$ | 138 |
| Avg. B-factors | |
| - Enzyme | 27.8 |
| - Ligand | 65.3 |
| - $H_2O$ | 39.5 |
| R.m.s. deviations | |
| - Bond lengths (Å) | 0.013 |
| - Bond angles (°) | 1.15 |
| [a] Highest resolution shell in parentheses | |

is similar to the calculated content in the reported structure (46% calculated, 39% reported).[4,36–38] The unit cell parameters of the structure in the current study deviate from the reported structure (*a* = 42.0 *b* = 75.6 *c* = 101.3 versus *a* = 46.7 *b* = 71.9 *c* = 101.7 for the current structure and reported structure (PDB accession code 1QJE)[5] respectively). Interestingly, the occupancy for Fe at the active site was reduced to 50% in the structure obtained from turnover of the IPNS • Fe • ACV complex. A similar observation was made in subsequent IPNS • Fe • IPN structures. A structure recorded 12 weeks after exposure of IPNS • Fe • ACV to $O_2$, led to complete absence of Fe at the active site. Notably, the occupancy for **IPN** remained 100%. These observations suggest that **IPN** is unable to leave the active site, but Fe is able to. It could be that the crystal lattice contributes to the dissociation of molecules from the active site. Interestingly, the additional electron density observed in structures obtained from IPNS • Fe • ACV microcrystals exposed to $O_2$ for 800 and 1600 ms and in the IPNS • Fe • ACV • NO structure, were not observed.

The B-factors of the active site residues from the different structures, i. e. the IPNS • Fe • ACV anaerobic, exposed to $O_2$ (800 and 1600 ms), and the IPNS • Fe • IPN structure, were compared. Compared to the anaerobic structure, the B-factors for the **ACV** substrate increase (Table 4.4), suggesting an increase in the mobility of the substrate. This result is in agreement with the observation that the substrate undergoes a conformational change as observed in the 800 and 1600 ms $O_2$ exposure structures. In the IPNS • Fe • IPN structure, the **IPN** product exhibits relatively high B-factors (Table 4.4), suggesting mobility of the atoms, possibly reflecting an unstable enzyme-product complex. Of the residues in the active site engaging in interactions with the **ACV** substrate, Tyr189, Arg279, and Ser281 experience a slight increase in B-factors with prolonged $O_2$ exposure times. Interestingly, these residues exhibit a lower B-factor in the **IPN** product structure (Table 4.4), suggesting a more ridged conformation after catalysis is completed. The average B-factors for α-helix 3 and β-strand 13 similarly increase with prolonged $O_2$ exposure, suggesting increased flexibility. In all the structures the average B-factors for α-helix 3 is higher than the average B-factor of the enzyme (Table 4.4). These observations together with the additional electron density observed near α-helix 3 and β-strand 13, suggest a possible second conformation and a higher degree of flexibility for α-helix 3.

**Table 4.4 B-factor overview**

| **Residue** | **Anaerobic (Å²)** | **800 ms[a] (Å²)** | **1600 ms[a] (Å²)** | **IPN (Å²)** |
|---|---|---|---|---|
| **ACV/IPN** | 20.7 | 25.3 | 33.8 | 59.6[b] |
| Aad | 20.4 | 24.6 | 32.9 | 53.0[b] |
| Cys | 19.3 | 23.0 | 34.8 | 78.2[b] |
| Val | 22.3 | 27.7 | 34.3 | 53.9[b] |
| Arg87 | 20.1 | 22.3 | 24.3 | 16.7 |
| Tyr189 | 19.9 | 23.2 | 27.2 | 18.1 |
| Arg279 | 19.9 | 21.3 | 23.1 | 15.3 |
| Ser281 | 17.6 | 20.3 | 23.4 | 12.0 |
| α-helix 3 | 30.7 | 33.2 | 35.7 | 28.9 |
| β-strand 13 | 19.9 | 22.2 | 24.5 | 19.1 |
| Enzyme avg. | 27.9 | 29.6 | 30.2 | 27.8 |

a) Time is referring to exposure of the IPNS • Fe • ACV to $O_2$
b) Average B-factor for the atoms in **IPN** that derive from the corresponding amino acid in **ACV**

### 4.3.8 Structure of IPNS exposed to nitric oxide

Crystallographic studies on IPNS in complex with L-δ-(α-aminoadipoyl)-L-cysteinyl D-valine (**ACV**) and Fe(II) using IPNS • Fe • ACV microcrystals (Section **4.3.6**), led to the observation of additional electron density around α-helix 3 (residue 49 – 65) and β-strand 13 (residue 280 – 282) in IPNS (Figure 4.22 (a), highlighted in red). The additional electron density was modelled with an alternative conformation (Figure 4.22 (b) and (c)) for residues on α-helix 3 and β-strand 13 during refinements (*performed by Dr Patrick Rabe, University of Oxford*). The occupancy for the alternative conformation of α-helix 3, correlates with $O_2$ exposure time; extending the $O_2$ exposure time, results in additional electron density and a higher occupancy of the alternative conformation. However, the occupancy for $O_2$ at the active site does not match the occupancy of the alternative conformation of α-helix 3. To investigate the alternative conformation for α-helix 3 and β-strand 13, a structure of IPNS • Fe • ACV • NO was obtained by exposure of IPNS • Fe • ACV crystals to NO gas (Figure 4.22, *refined and analysed by Dr Patrick Rabe, University of Oxford*). Crystals of IPNS • Fe • ACV were prepared under anaerobic conditions. Crystals were transferred with the crystallisation solution by pipetting into an Eppendorf vial and being kept under a positive $N_2$ pressure while being transported into a fume hood. Inside the fume hood the sample was put under a positive NO (1,000 ppm in $N_2$) pressure, and purged for 1 hour. The crystals were subsequently fished and cryo-cooled before data collection. Macrocrystals of IPNS • Fe • ACV have been exposed to NO before,[3] however, the raw data were not available to enable detailed reanalysis. Comparison of the overall fold of the reported structure (PDB accession code: 1BLZ) and the obtained cryo structure of IPNS • Fe • ACV • NO shows high similarity, with an RMSD of 0.36 Å. Detailed analyses of the Fe-bound NO, show a higher B-factors for both atoms (22.4 and 28.4 $Å^2$, for the N and O-atom respectively) compared to the average for the protein (19.5 $Å^2$). The structure of IPNS • Fe • ACV • NO further demonstrates that the **ACV** substrate bound to Fe via the cysteinyl thiolate. The Fe centre is hexagonally coordinated by a triad of residues (i. e. His214, Asp216, and His270), $H_2O$ and NO. In the IPNS • Fe • ACV • NO further structure, the **ACV** Val isopropyl group is oriented away from the Fe to accommodate for the presence of NO at the Fe centre. The Fe-N-O angle (130°) is greater than in the reported structure by Baldwin and co-workers (120°, Figure 4.22 (c))[3]. The residues at α-helix 3 and β-strand 13 were modelled in the conformation observed in the anaerobic structure (50% occupancy), and in the newly observed alternative conformation (50%

occupancy, Figure 4.22 (b) and (c), teal for anaerobic, pink for alternative conformation). The additional electron density observed in the $O_2$ exposed structures have not been commented on before.[5]

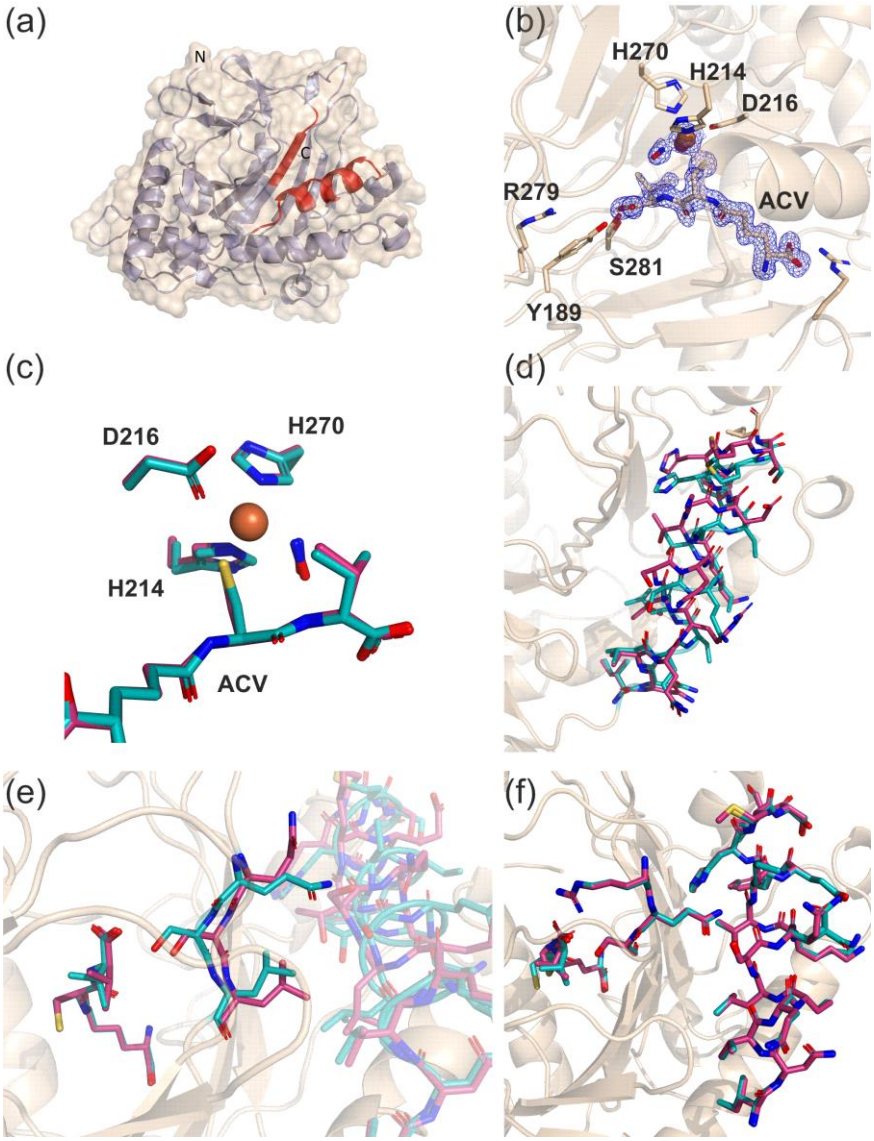



**Figure 4.22:** Views from structures of IPNS obtained under anaerobic and $O_2$ exposed conditions. (a) Highlighted in red are α-helix 3 (residues 49 – 65) and residues on β-strand 13 (residues 280 – 282), which display alternative conformations after NO exposure; (b) view from the active site of IPNS • Fe • ACV • NO (2mFo-Fc maps contoured to 1σ); (c) view from the active site of IPNS • Fe • ACV • NO (teal) overlaid with the reported IPNS • Fe • ACV • NO structure (PDB accession code: 1BLZ)[3]. (d) the residues on α-helix 3 in their conformation under anaerobic conditions (teal) and their alternative conformation after NO exposure (pink) are both modelled at 50% occupancy; (e) three residues on β-strand 13 and **ACV** are highlighted and their conformations under anaerobic conditions (teal) and their alternative conformation after NO exposure (pink) shown (both modelled at 50% occupancy); (f) overlay of the structure of IPNS • Fe • ACV under anaerobic conditions (teal) and IPNS • Fe • IPN (pink) obtained through turn-over in IPNS • Fe • ACV microcrystals. *Data analysis and refinements were performed by Dr Patrick Rabe (University of Oxford).*

## 4.4 Conformational changes in isopenicillin N synthase

### 4.4.1 Structural rearrangements in IPNS

Crystallographic studies discussed above (Section **4.3**) on IPNS in complex with **ACV** and Fe(II) using IPNS • Fe • ACV microcrystals, as well as IPNS • Fe • ACV • NO, led to the observation of additional electron density around α-helix 3 (residue 49 – 65) and β-strand 13 (residue 280 – 282) in IPNS (Figure 4.22 (a), highlighted in red). The additional electron density was interpreted as an alternative conformation for residues on α-helix 3 and β-strand 13 as a result of exposure to $O_2$ or NO (Figure 4.22 (b) and (c)). A correlation between occupancy for the alternative conformation of α-helix 3 and β-strand 13 was found; longer $O_2$ exposure led to an increase of the density correlating to the alternative conformation. However, occupancy of $O_2$ and NO modelled at the active site did not reflect the occupancy of the alternative conformation of α-helix 3 and β-strand 13. Interestingly, a structure of IPNS • Fe • IPN (Figure 4.21 and 4.22 (f), Section **4.3.7**) obtained by exposing IPNS • Fe • ACV microcrystals to $O_2$ (>12 h), did not exhibit the additional electron density observed in the early $O_2$ exposure time points (Figure 4.22 (d)). Instead, the conformation of α-helix 3 matches the conformation of IPNS • Fe • ACV structure acquired under anaerobic conditions. The additional electron density observed in the $O_2$ exposed structures have not been commented on before.[5] To better understand and investigate whether the mobility of the α-helix 3 region on IPNS as a consequence of $O_2$ / NO exposure occurs in solution, $^{19}$F NMR studies with a labelled variant of IPNS were carried out. For these experiments NO was used as an $O_2$ surrogate to avoid substrate turnover.

### 4.4.2 Protein observed $^{19}$F NMR spectroscopy

#### 4.4.2.1 Solution state structural analysis by NMR

$^{19}$F NMR spectroscopy has been used in a biological context for studies of inhibitor binding, drug mechanism of action and fragment screening.[39–42] Many of these applications rely on a fluorine atom present on the ligand to obtain structural information. Covalent labelling of proteins with $^{19}$F has been exploited to monitor conformational changes upon ligand binding.[41] Notably, structural evidence for the mobility of a loop in New Delhi metallo-β-lactamase (NDM-1) was supported by $^{19}$F NMR studies in solution.[43] Similarly, $^{19}$F-labelled γ-butyrobetaine hydroxylase from *Pseudomonas* (psBBOX), which is difficult to

crystallize, was used to study structural mobility of a loop.[44] The information obtained in these studies, including changes in secondary structures using $^{19}F$ NMR spectroscopy, would often not have been possible with alternative methods such as UV or circular dichroism (CD) spectroscopy. Solution state structural evidence can be obtained using traditional NMR spectroscopy methods, requiring $^{15}N$, $^{13}C$, and / or $^{2}H$ labelling of the entire protein.[45,46] These labelling strategies are technically challenging on large proteins, require expensive isotopically labelled material, and involve long data acquisition times, limiting their utility. However, to a certain extent, structural information can also be obtained by $^{19}F$ NMR.[42,47,48] Biological samples generally lack the presence of endogenous fluorine, which makes $^{19}F$ NMR spectroscopy well suited for studies on them. The high natural abundance of an NMR observable isotope ($^{19}F$, 100%), with spin quantum number *I* of ½, makes $^{19}F$ the second most sensitive natural abundance nucleus for NMR spectroscopy (83% relative to $^{1}H$).[41,47] The spectral width of $^{19}F$ NMR spans a wide range (>400 ppm), much larger than a typical $^{1}H$ NMR spectra range.[47] Covalently bound fluorine atoms exhibit a large intrinsic sensitivity to changes in their micro-environment, e. g. changes in orientation. The enhanced sensitivity of $^{19}F$ NMR spectroscopy is attributed to the anisotropic distribution of the electrons in the three 2-p orbitals of fluorine.[49] The enhanced sensitivity of $^{19}F$ and its large spectral range typically allows for well resolved 1D NMR spectra, and the sensitivity towards changes in environment makes it ideally suited for monitoring dynamic structural changes in proteins in solution.[43,44,50,51]

#### 4.2.2.2 Approaches for $^{19}F$ labelling of proteins

Various methodologies are available to install modified fluorinated amino acids into proteins. Such methods can be divided into two categories: biosynthetic incorporation,[52] and chemical modification of the intact protein[53]. Advances in synthetically prepared fluorinated amino acids and their commercial availability, have popularised their use in the biosynthetic incorporation strategy in recent years, e. g. auxotrophic work.[54,55] However, these methods occasionally suffer from issues such as partial labelling, perturbation of the protein fold, solution stability problems, or multiple labelling sites which complicate spectral assignments.[52] Some of these drawbacks can be addressed using site specific labelling. An example of such a method uses orthogonal amber codon suppression to achieve single site labelling with unnatural amino acids.[56,57]

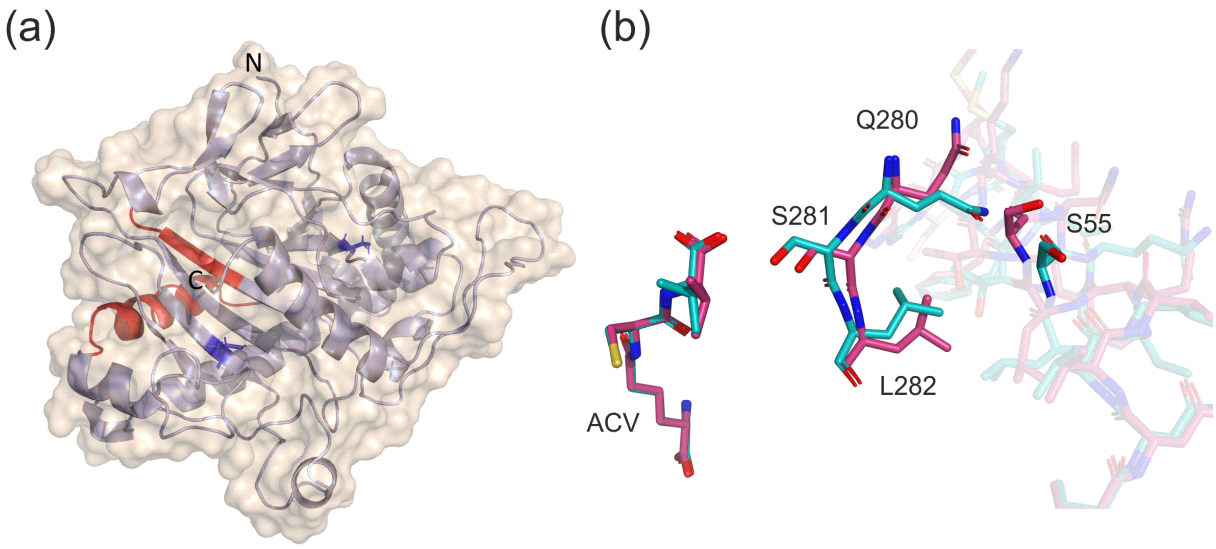



**Figure 4.23:** Views from structures of IPNS showing alternative conformations as a result of NO exposure. (a) View of the overall fold of wt-IPNS, highlighted are Cys104 and Cys253 (blue), and α-helix 3 (residues 49 – 65) and β-strand 13 (residues 280 – 282) (red); (b) view of residues 280 – 282 present on β-strand 13 and the **ACV** substrate, displayed in the conformation observed under anaerobic conditions (teal) and the additional alternative conformation observed after NO exposure (pink). Notably, the conformation of Ser55 changes considerably as a result of NO exposure.

However, the biosynthetic methods can suffer from low expression levels or toxic effects resulting from the fluorinated compound supplemented in the broth.[52] Thus, *in vitro* chemical modification of an exposed Cys residue at the desired location has been developed as an approach to site selectively introduce a $^{19}F$ label. Wt-IPNS contains two Cys residues, however neither are solvent exposed according to reported crystal structures,[3] and both located remotely from α-helix 3 (Figure 4.22 (c)). Ser55 is located on α-helix 3 (Figure 4.23), in the region close to Gln280 and Leu282 (located on β-stand 13 between the active site and the α-helix 3) and is solvent exposed. The conformation of Ser55 changes considerably as a consequence of NO (and by extrapolation $O_2$) exposure (Figure 4.23 (b)). On the basis of (i) the location of Ser55 on α-helix 3, (ii) the observed conformational change Ser55 undergoes upon exposure to NO, and (iii) the solvent exposed nature of the residue, an S55C variant of IPNS was produced and chemically labelled using reported methodology[43,44].

# 4.5 Mutation and fluorine labelling of IPNS

## 4.5.1 Expression and fluorine labelling of IPNS S55C variant

*A glycerol stock of IPNS pCOLD construct in Escherichia coli BL21(DE3) along with production, expression and purification protocol were obtained from Dr Ilaria Pettinati (University of Oxford). A glycerol stock of IPNS S55C construct in BL21(DE3) was obtained from James Linyard (University of Oxford).*

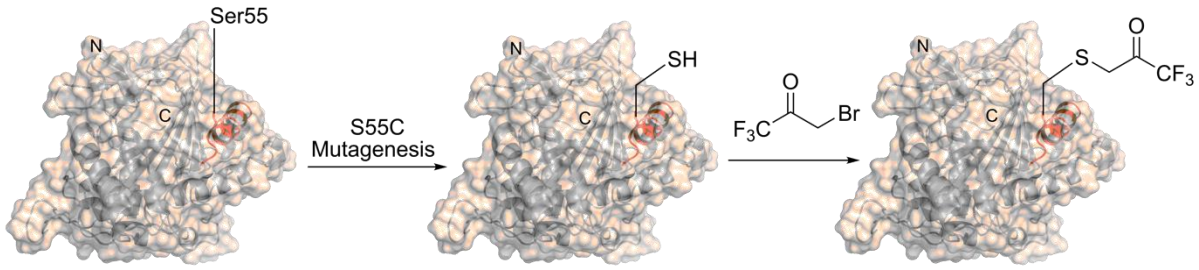



**Figure 4.24:** Schematic approach for labelling of IPNS. Ser55 is positioned on α-helix 3 (red), which is suggested to change conformation based on time resolved crystallography experiments. IPNS S55C variant is obtained by mutagenesis, and labelled using 3-bromo-1,1,1-trifluoro acetone (BTFA) (Tris buffer, 50 mM, pH 8.5; TCEP 2 equivalents). *The structure used in this illustration was generated from the structure of IPNS in complex with Fe(II) and* ***ACV****, and was refined and analysed by Dr Patrick Rabe (University of Oxford).*

Recombinant codon-optimised IPNS from *Aspergillus nidulans* and IPNS S55C mutant for use in $^{19}$F labelling, $^{19}$F NMR binding studies, and NMR / LC-MS based assays, were expressed using *E. coli* expression system. The gene encoding for IPNS was inserted into a pCOLD vector (*by Dr Ilaria Pettinati, University of Oxford*) and expressed with an *N*-terminal hexahistidine affinity tag in BL21 (DE3) *E. coli* cells. IPNS residue Ser55 was mutated to Cys (S55C, by *James Linyard, University of Oxford*), and expression was done according to protocol (*provided by Dr Illaria Pettinati, University of Oxford*). In brief, for both wt-IPNS and IPNS S55C in *E. coli* cells (BL21 DE3, kanamycin resistant) transformed with appropriate plasmids were grown in 2TY media. Cells were grown to an $OD_{600}$ of 0.6 before cooling to 18 °C, expression was induced by addition of isopropyl β-D-1-thiogalactopyranosided (IPTG). Cells were incubated at 18 °C for 12 h, before harvesting by centrifugation. The cells were suspended in buffer and lysed by sonication. wt-IPNS protein was purified by FPLC using a HisTrap™ column, followed by size exclusion chromatography. The hexahistidine affinity tag was removed using the 3C protease, and the tag and protease were manually separated using a HisTrap™ column. For IPNS S55C purification, the cells were lysed by sonication, and the protein purified by FPLC using a HisTrap™ column, followed by cleavage of the hexahistidine affinity

tag using 3C protease. The tag and protease were removed manual using a HisTrap™ column. The protein was further purified by a gel filtration column to give IPNS S55C. Protein purity was assessed by SDS-PAGE gel (Figure 4.25) and its identity confirmed by denaturing MS (Figure 4.26).

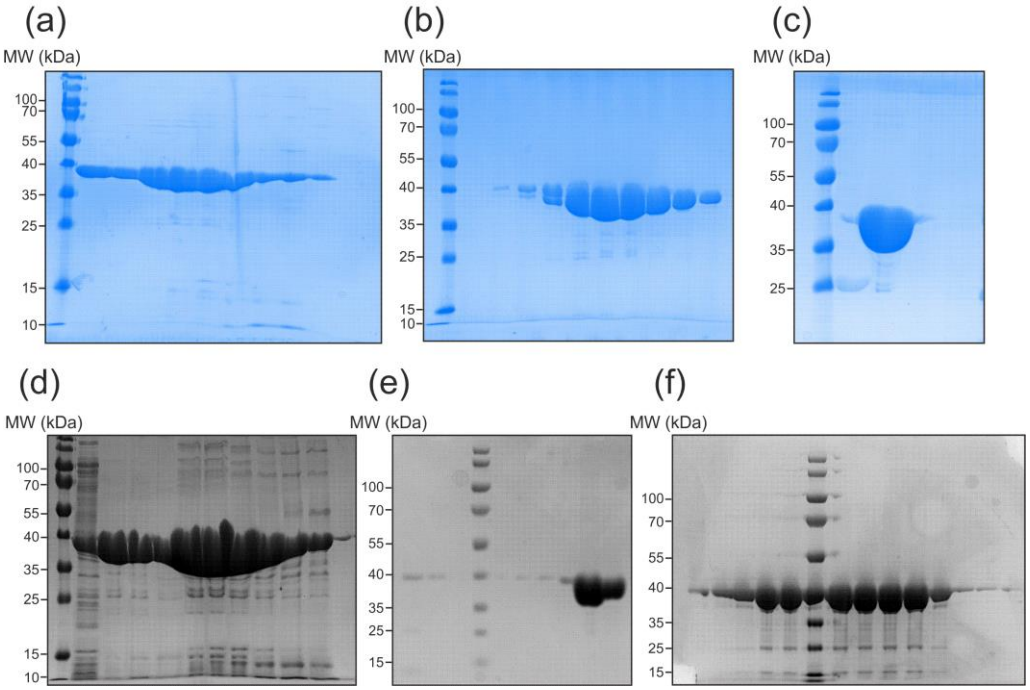



**Figure 4.25:** SDS-PAGE gel of IPNS S55C (pCOLD construct) at different purification stages. (a) wt-IPNS after the initial HisTrap™ column; (b) wt-IPNS after gel filtration column; (c) wt-IPNS after His-tag cleavage 3C protease and purification using a manual HisTrap™ column; (d) IPNS S55C after initial HisTrap™ column; (e) IPNS S55C after His-tag cleavage 3C protease and purification using a manual HisTrap™ column; (f) IPNS S55C after gel filtration column. The molecular weight is indicated in kDa.

To assess the purity and confirm the incorporation of the mutation of IPNS S55C variant, denaturing mass spectra were recorded. The IPNS S55C variant was subsequently labelled, by exposure of IPNS S55C to the sulfhydryl selective reagent 3-bromo-1,1,1-trifluoro acetone (BTFA).[43,58] When wt-IPNS (pJB703 construct,[9] m/z = 37378, Section **3.2.5**), which contains two Cys residues (i. e. Cys104 and Cys253), was exposed to BTFA (30 equivalents, Tris buffer (50 mM, pH 8.5)) and analysed by denaturing LC-MS, no acylation was observed (Figure 4.26 (a) and (b)). By contrast, treatment of IPNS S55C (pCOLD construct, m/z = 37768 Da, Figure 4.26 (c)) with BTFA under the same conditions (30 equivalents, Tris buffer (50 mM, pH 8.5)), resulted in a mass increase of +128 Da (m/z = 37896 Da, Figure 4.26 (d)). The mass of the labelled IPNS S55C corresponds to 37896, which is 18 Da higher than the mass of a $CF_3COCH_2$ group

(110 Da), suggesting the trifluoro ketone may be present in its hydrated form as reported before.[43] The enzyme was subjected to dialysis with EDTA and 1,10-phenanthroline in order to remove any residual metal and stored at –80 °C. Overall yield of the expression of IPNS S55C, labelling with BTFA, and dialysis was 8 mg · $L^{-1}$. Substrate turnover assays showed both the IPNS S55C variant and the BTFA labelled IPNS S55C variant were catalytically active. In addition, turnover assays in absence of supplied Fe demonstrated that the BTFA labelled IPNS S55C was *apo*. Note, further studies on the kinetic parameters of the IPNS S55C variant and BTFA labelled variant are required to assess their similarity to wt-IPNS.

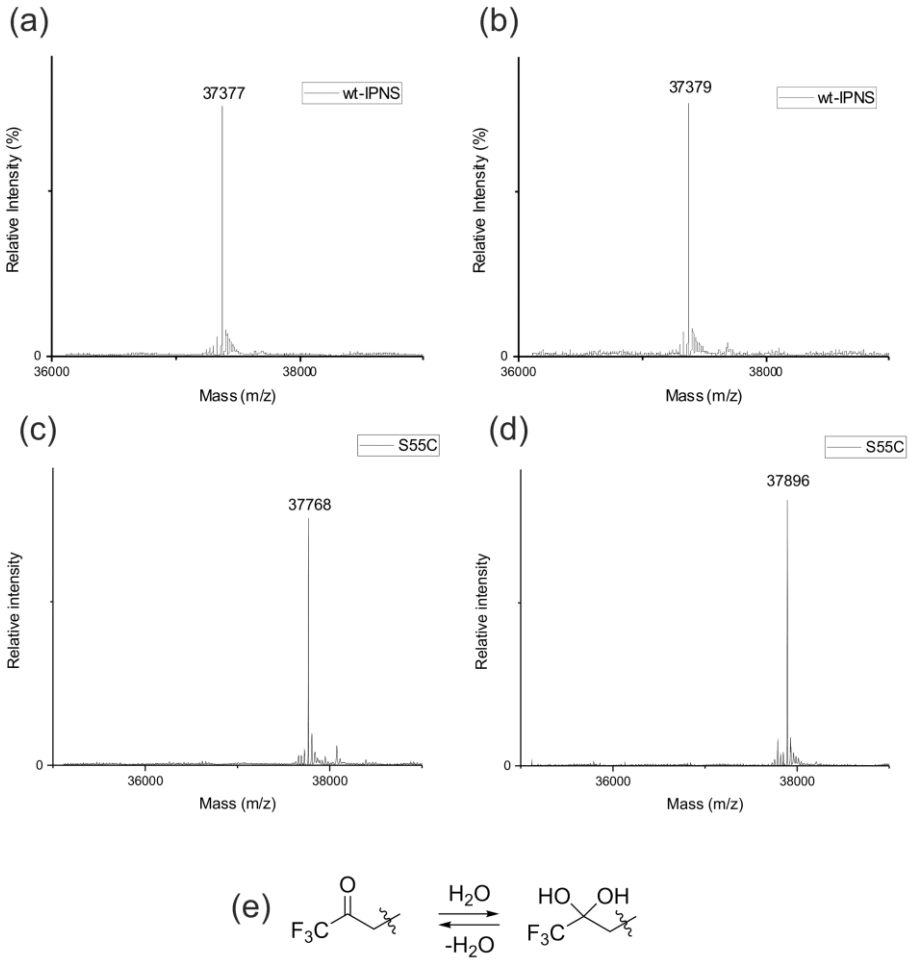



**Figure 4.26:** Denaturing MS spectra of IPNS S55C variant before and after labelling with 3-bromo-1,1,1-trifluoro acetone (BTFA). (a) Wt-IPNS (pJB703 construct) reference spectrum; (b) wt-IPNS (pJB703 construct) after incubation with BTFA (30 equivalents, 3 hours); (c) IPNS S55C variant (m/z = 37768 Da); (d) IPNS S55C variant after incubation with BTFA (30 equivalents, 3 hours) (m/z = 37896); (e) the trifluoro ketone label in equilibrium with its ketone and dihydroxy form.

### 4.5.2 Investigating labelling site and efficiency

*The trypsin digestion of BTFA labelled IPNS S55C was run according to a protocol provided by Elisabete Pires (University of Oxford), who performed the LC-MS MS assay, and guided me through data analysis.*

To investigate the location of the fluorine label, IPNS S55C BTFA (IPNS*) was fragmented by porcine trypsin digestion, then analysed by LC-MS MS. The digestion conditions gave rise to the formation of 225 unique peptide fragments, covering 96% of the protein sequence, as identified by LC-MS analysis (Figure 4.27). The fragmentation studies confirmed instalment of the BTFA label at residue 55 in the IPNS* sequence, suggesting complete labelling of this Cys residue (Figure 4.27 (b)). Analyses of the fragments containing Cys104 revealed that some had a mass consistent with the BTFA modification. This observation was surprising as no labelling was observed by denaturing MS using wt-IPNS (under identical conditions), and only a single BTFA label was observed by denaturing MS using IPNS S55C (Figure 4.26) after incubation with BTFA (Section **4.3.1**). Detailed analyses of the relative abundance of the fragments containing the BTFA modification at Cys104, suggest only a minor fraction (<2%) of the protein contains a labelled Cys104 site (Figure 4.27 (c)). It is possible that the low levels of apparent Cys104 reaction occur on the intact protein but at too low a level to detect using denaturing LC-MS.

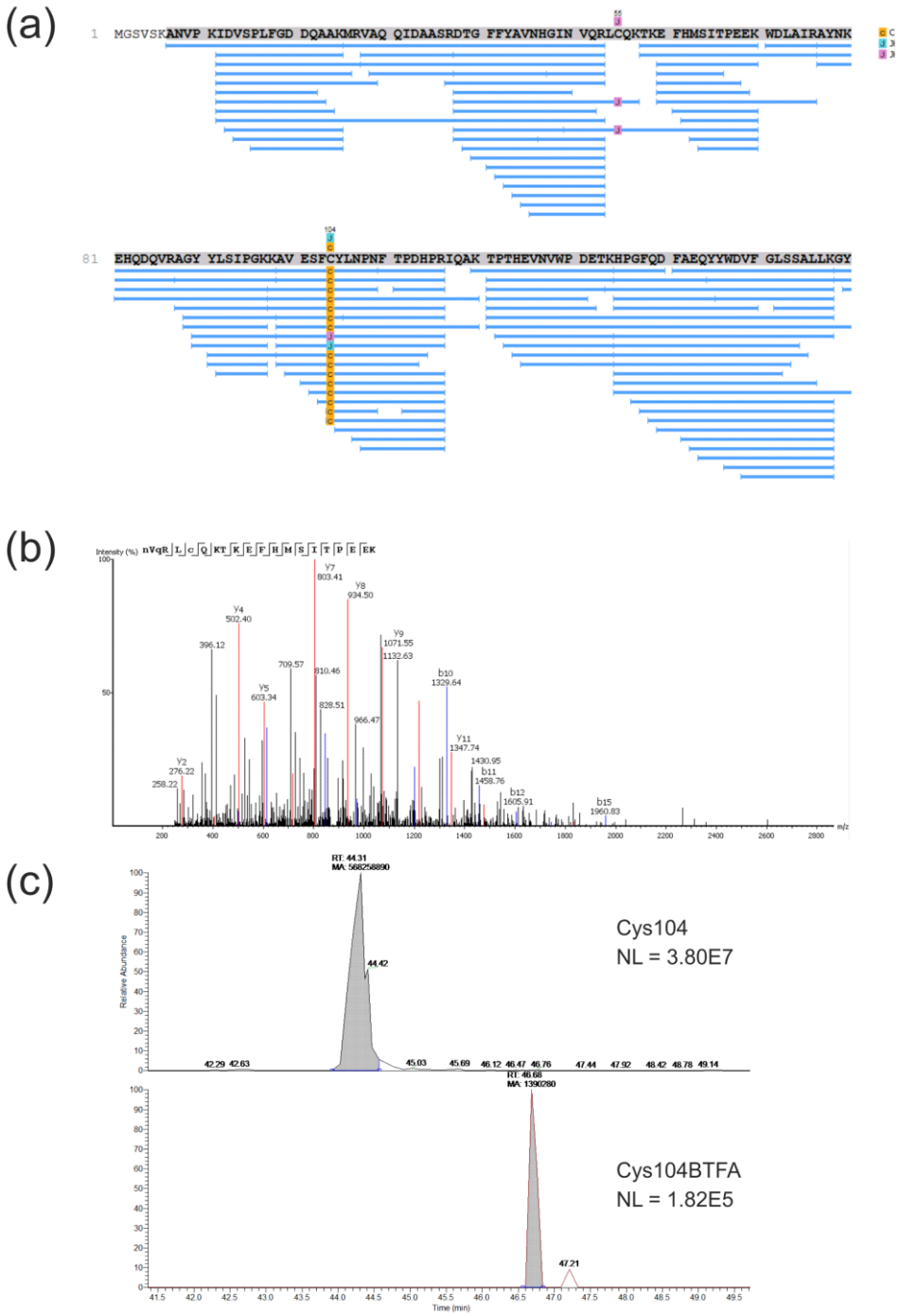



**Figure 4.27:** Nano LC-MS analyses of IPNS S55C BTFA (IPNS*) digested by porcine trypsin. Labelled IPNS* was subjected to trypsin digestion, followed by LC-MS analyses. The data were analysed using BSI PEAKS studio 8.5. (a) Part of the sequence (residues 1 – 130) are aligned with fragments identified in the raw data, showing C55 and its modification. Highlighted are modifications including carbamidomethylation (orange c), BTFA labelling (teal j), and hydrated BTFA labelling (pink j). Note that the data suggests two fragments (out of 21) that have Cys104 labelled with BTFA; (b) mass spectrum of one of the fragments containing BTFA labelled Cys55 (2675 Da); (c) Comparison of relative intensity of a peptide fragment of Cys104 not modified by BTFA (top) and a peptide fragment with BTFA modification (bottom). Their relative intensities ($3.80 \cdot 10^7$ and $1.82 \cdot 10^5$) differ by two orders of magnitude.

### 4.5.3 Circular dichroism analyses

To evaluate the effect of labelling of the IPNS S55C variant with BTFA on the secondary structure, CD analyses were performed. The CD spectra of wt-IPNS (JB703 construct, red line, Figure 4.28), and wt-IPNS (pCOLD construct, blue line, Figure 4.28), were recorded and compared to the spectra collected on IPNS S55C variant, and IPNS* (Figure 4.28). The predicted α-helix and β-strand contents as observed by CD for both the unlabelled and labelled variant, correlate well to the contents found for both wt-IPNS batches. These results suggest that BTFA labelling does not substantially affect the secondary structure content and overall fold of IPNS in solution.

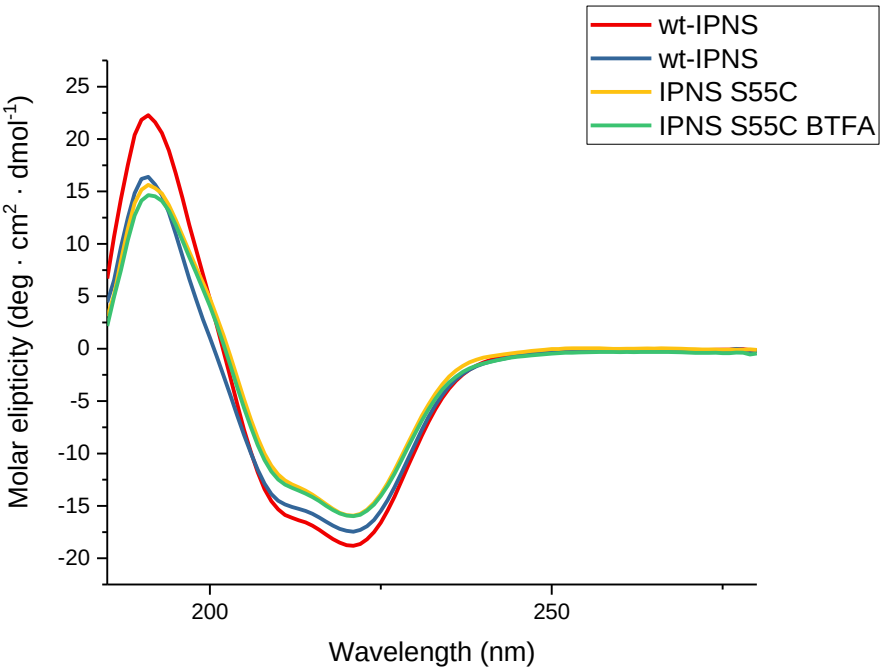



**Figure 4.28:** Circular dichroism (CD) spectra of wt-IPNS and IPNS variants. The wt-IPNS (pJB703 construct,[9] red), wt-IPNS (pCOLD construct, blue), IPNS S55C variant (pCOLD construct, yellow), and labelled IPNS S55C BTFA (IPNS*, derived from IPNS S55C, green) are shown.

### 4.5.4 Structural evidence of IPNS S55C BTFA

To provide further evidence for the position of the BTFA label on IPNS*, and to compare the structural features of IPNS* with wt-IPNS, IPNS* • Fe • ACV crystals were grown (*by Mr James Linyard, University of Oxford*) and diffraction data collected (*by Dr Patrick Rabe, University of Oxford*). A structure of IPNS* • Fe • ACV was determined to 1.3 Å, and after refinements (*by Dr Patrick Rabe, University of Oxford*) the overall fold of IPNS* was found to be similar to wt-IPNS (Figure 4.29 (a)), which is in agreement with the CD data. Comparison of the conformation of α-helix 3 (residues 49 – 65) and β-strand 13 (residues 280 – 282) in IPNS* and wt-IPNS both under anaerobic conditions, shows a similar structure with **ACV** and the residues present on β-strand 13 in identical conformations (Figure 4.29 (b) and (c)). Comparing residue 55 with neighbouring residues in IPNS* and wt-IPNS, reveals that their conformations are similar, with an RMSD of 0.23 Å over the peptide backbone (Figure 4.29 (d)). The BTFA label is oriented to the bulk solvent and away from β-strand 13 (Figure 4.29 (e)). Based on the electron density map around residue Cys104, no evidence could be found for acylation (Figure 4.29 (f)). However, it cannot be excluded that low levels of acylation (<2%) are part of the crystal structure. The anaerobic IPNS* • Fe • ACV structure is similar to the wt-IPNS • Fe • ACV structure, confirming that IPNS* is a suitable model system to investigate conformational rearrangement in solution.

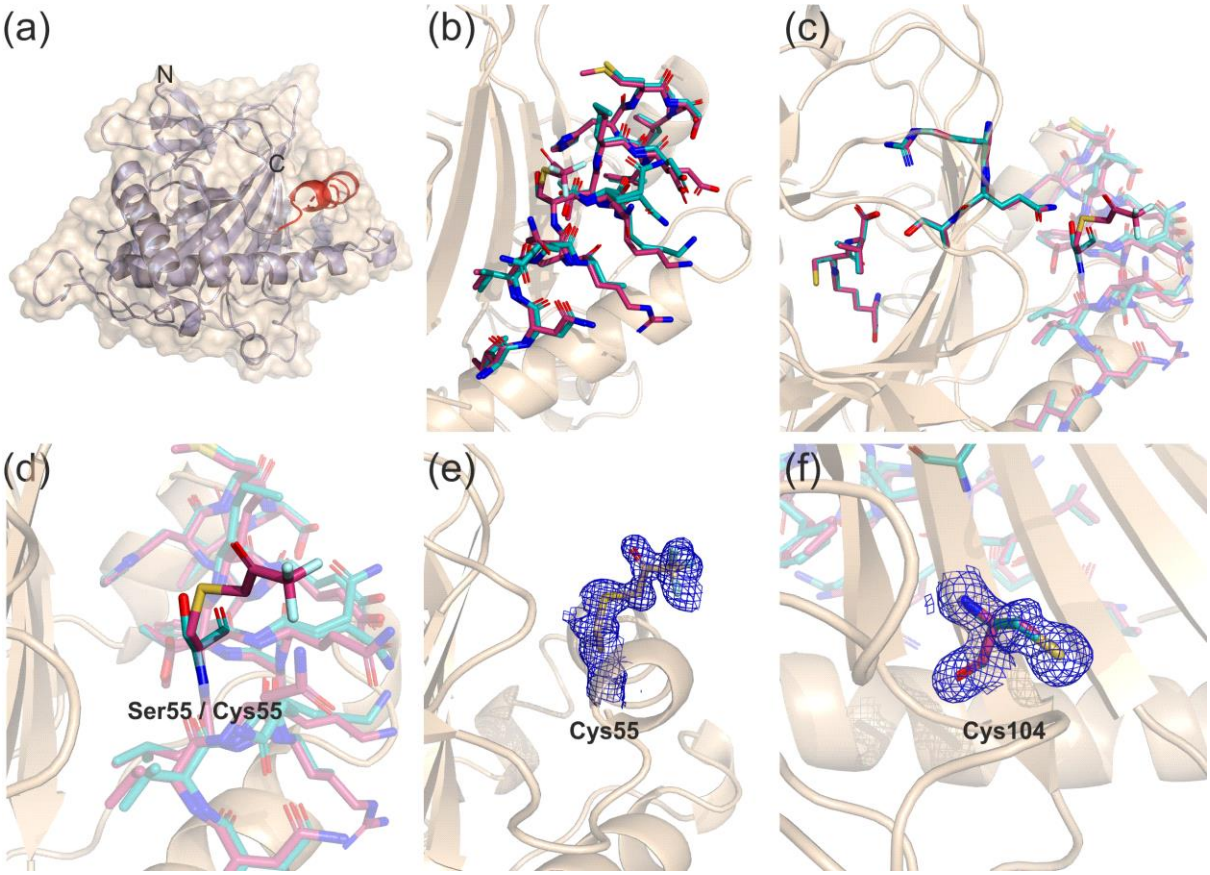



**Figure 4.29:** Views from structures of IPNS S55C BTFA (IPNS*) compared to wt-IPNS both under anaerobic conditions. (a) Overall view of the structure of IPNS* under anaerobic conditions, with α-helix 3 highlighted in red; (b) view of the α-helix 3 (residues 49 – 65) of IPNS* (pink) superimposed on wt-IPNS (teal) both under anaerobic conditions; (c) superimposed structures of IPNS* (pink) and wt-IPNS (teal), with **ACV**, residues at β-strand 13 (residues 280 – 282), and Ser55 (wt-IPNS) / Cys55 BTFA (IPNS*) highlighted; (d) superimposed structures of IPNS* (pink) and wt-IPNS (teal), with Ser55 (wt-IPNS) and Cys55 BTFA (IPNS*) highlighted; (e) view from the IPNS* structure highlighting Cys55, with 2mFo-Fc maps contoured to 1σ; (f) view from structure of IPNS* (pink) superimposed on wt-IPNS (teal), with Cys104 from the IPNS* structure contoured with 2mFo-Fc maps to 1σ. *Crystals were grown by Mr James Linyard, and refinements were performed by Dr Patrick Rabe (University of Oxford).*

## 4.6 Helix rearrangement experiments by NMR

### 4.6.1 Titration experiments of IPNS variant by $^{19}$F NMR

The peak assigned to IPNS* ($\delta_F$ –84.88 ppm) in the $^{19}$F NMR spectrum, appeared to be a broader resonance than the TFA ($\delta_F$ –76.55 ppm) standard, which is in agreement with observations from nuclei located on large protein molecules (Figure 4.30).[39] In addition, a significantly broad resonance ($\delta_F$ –83.1 to –83.8 ppm) was observed. This resonance appears to remain constant in chemical shift and relative intensity to the IPNS* resonance throughout the experiments.

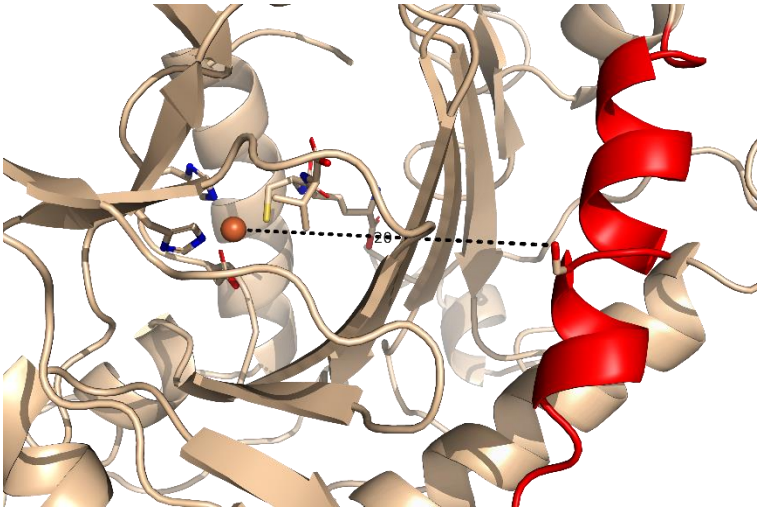



**Figure 4.30:** View of the structure of IPNS • Fe • ACV under anaerobic conditions. Highlighted in red is α-helix 3 (residues 49 – 65). Shown in the structure is the distance from Fe(II) in the active site to Ser55, present on α-helix 3 (~20 Å away from Fe(II)). *Structure of anaerobic IPNS as refined by Dr Patrick Rabe (University of Oxford) was used to create this image.*

Titration of Fe(II) to IPNS* under anaerobic conditions (IPNS*, 150 μM; in Tris-$d_{11}$, 25 mM, pH 7.5; TFA, 100 μM) using a glovebox (<1 ppm $O_2$), led to the observation of a new resonance ($\delta_F$ –84.72 ppm) by $^{19}$F NMR, which was assigned as the IPNS* : Fe(II) complex (Figure 4.31). Formation of the complex was completed with the addition of stoichiometric amounts of Fe(II) (1 equivalent with respect to IPNS*). Notably, the $^{19}$F label introduced at residue 55 on the IPNS S55C variant, is at a considerable distance from the active site (~20 Å, determined by crystallography, Figure 4.30), illustrating the remarkable sensitivity of $^{19}$F labelling to distant chemical, conformational, or solvation changes, as reported in other cases.[43,44] The resonance assigned to the IPNS* : Fe complex appears as a new peak, suggesting slow exchange with the *apo*-IPNS* peak, relative to the NMR time scale.[41] The IPNS* : Fe resonance ($\delta_F$ –84.72 ppm) appears to be flanked by a shoulder ($\delta_F$ –84.87 ppm). The intensity of this shoulder relative to the IPNS* resonance

appeared to be constant throughout the titrations carried out. Further investigation is required to elucidate the origin of this additional resonance ($\delta_F$ –84.87 ppm) and the broad resonance observed (between $\delta_F$ --83.1 and --83.8 ppm). Their constant appearance with respect to the IPNS* : Fe complex may reflect a different conformation or (partially) unfolded protein.

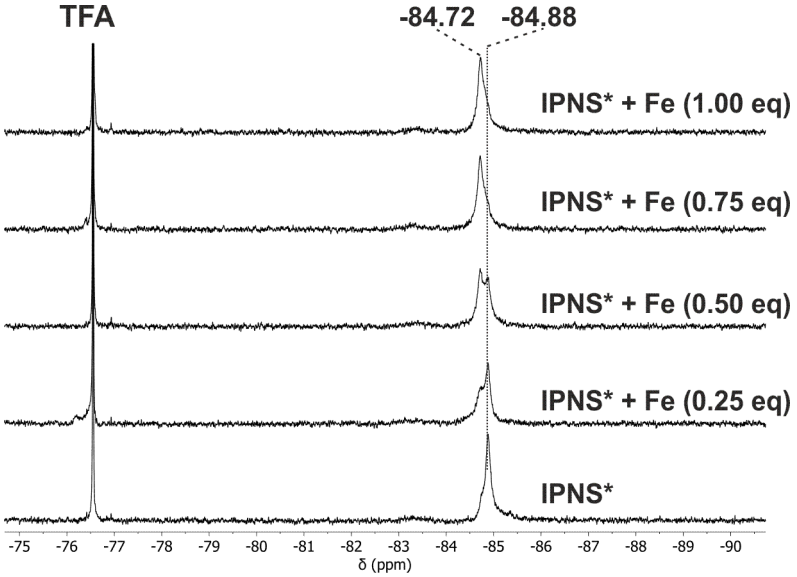



**Figure 4.31:** $^{19}F$ NMR titration of IPNS* with Fe(II). A solution of IPNS* (150 μM) was titrated with $FeSO_4$, with TFA (100 μM) as a standard for referencing.

Titration of **ACV** to the IPNS* : Fe complex under anaerobic conditions, resulted in the observation of a new resonance ($\delta_F$ –85.12 ppm) which was assigned to the IPNS* : Fe : ACV complex (Figure 4.32). Complete conversion to the IPNS* : Fe : ACV complex ($\delta_F$ --85.12 ppm) required an excess of **ACV** (64 equivalents) and revealed a broad resonance ($\delta_F$ –84.87 ppm). Addition of more Fe(II) or **ACV** did not change the appearance of the broad peak ($\delta_F$ --84.87 ppm). In the absence of **ACV** and Fe(II), this peak ($\delta_F$ –84.87 ppm) may represent a different conformational isomer of IPNS*. Further work will have to be carried out, analysing the behaviour of the resonance by varying the temperature or changing the buffer, to investigate the origin of this broad resonance.

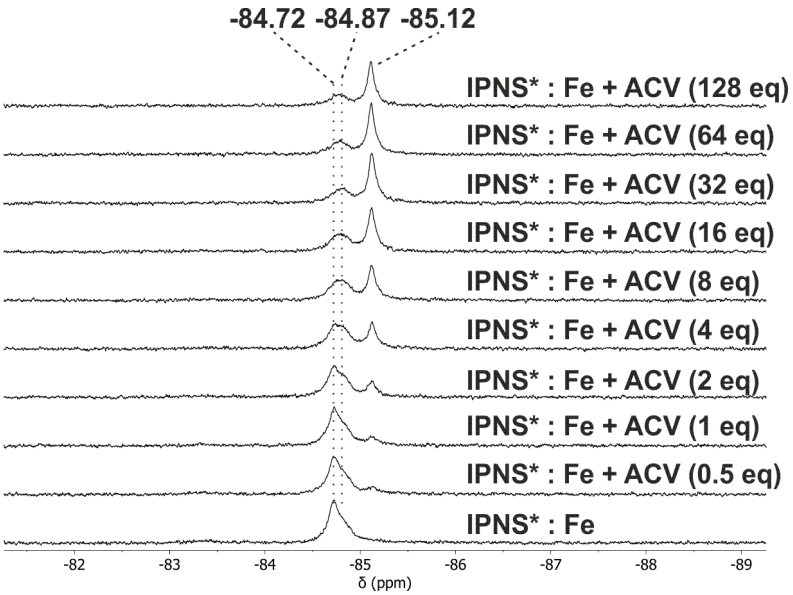



**Figure 4.32:** $^{19}F$ NMR titration of **ACV** to the IPNS* : Fe complex. A solution of IPNS* (150 μM) and $FeSO_4$ (1 equivalent) was titrated with **ACV**, with TFA (100 μM) as a standard for referencing.

### 4.6.2 Nitric oxide exposure of IPNS variant complexes

The IPNS* : Fe : ACV complex (IPNS* at 150 μM; Fe(II), 5 equivalents; and **ACV**, 20 equivalents) was prepared in a Young's tap NMR tube (5 mm) under anaerobic conditions. The Young's tap NMR tube was then connected to a glass apparatus (Figure 4.33 (a)). One compartment (the left hand side compartment, Figure 4.33 (a)) of the glass apparatus was filled with gaseous $N_2$, and the right compartment, including the NMR tube, was put under a mild vacuum. On the right side, gaseous NO was generated *in situ* by addition of sulfuric acid to sodium nitrite in presence of potassium iodide (Figure 4.33 (a)).[59] After NO had formed, addition of sulfuric acid was ceased, and the valve connecting the two halves was opened, to allow NO to equilibrate (10 min) into the NMR tube. The Young's tap NMR tube was closed, inverted, and allowed to equilibrate for an additional 20 minutes, before $^{19}F$ NMR spectral data were collected. A relatively large chemical shift change ($\Delta\delta_F$ 0.31 ppm) in the assigned IPNS : Fe : ACV complex resonance ($\delta_F$ –85.12 ppm) was observed. Colourimetric pH measurements of the solution after exposure to *in situ* generated NO, suggested the acidity of the solution had dropped to pH 2. The resonance ($\delta_F$ –85.43 ppm, Figure 4.33 (b)) that was identified by NMR was therefore likely to derive from soluble unfolded protein, rather than a properly folded protein. Interestingly, the broad signal ($\delta_F$ --84.82 ppm) disappeared, possibly due to precipitation from solution, or due to complete unfolding of the protein, which gives rise to a signal

overlapping with the unfolded IPNS* resonance ($\delta_F$ –85.43 ppm). Further investigation is required to determine the origin of the broad resonance at $\delta_F$ –84.87 ppm.

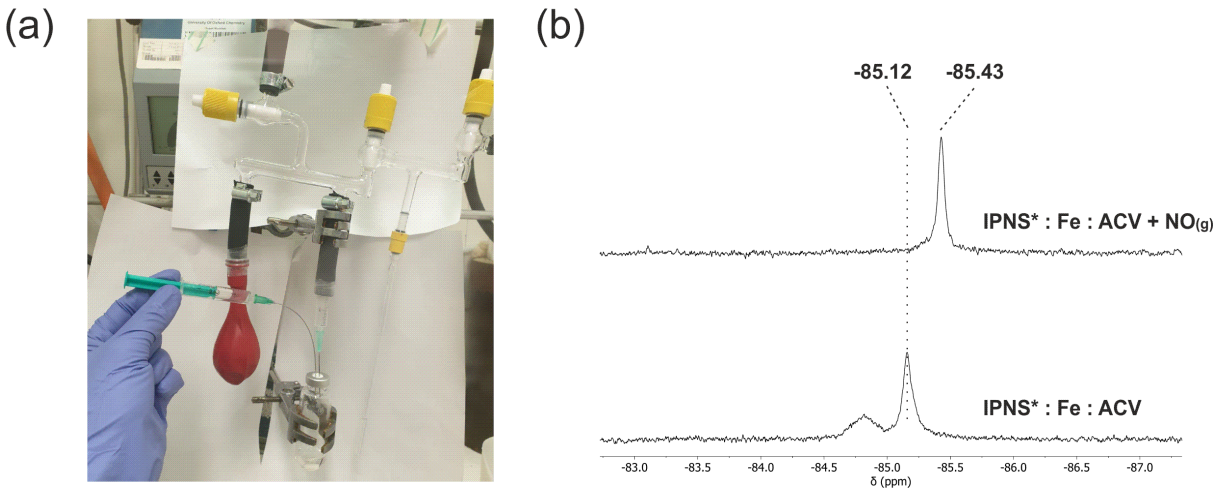



**Figure 4.33:** Picture of the nitric oxide (NO) exposure setup and $^{19}$F NMR of an IPNS* complex with Fe(II), and **ACV**, exposed to *in situ* generated NO. (a) Image of the setup used to expose the complex to *in situ* prepared NO gas. The NMR tube was connected to the glass apparatus and put under a mild vacuum (using tubing on the right connector). In the left compartment, under an inert atmosphere NO gas was generated from a reaction between sodium nitrite, in presence of potassium iodide (vial), and sulfuric acid (syringe), and stored in the balloon.[59] When sufficient NO was produced, the valve between the left and right compartment was opened, and NO gas flowed from left to right. (b) $^{19}$F NMR spectra of IPNS* : Fe(II) : ACV (IPNS, 150 μM; $FeSO_4$, 5 equivalents; **ACV**, 20 equivalents) before and after exposed to *in situ* generated NO gas.

Exposure of the IPNS* : Fe : ACV complex was repeated with dilute NO (1000 ppm in $N_2$ gas, BOC). The IPNS* : Fe : ACV complex (IPNS*, 150 μM; Fe(II), 1 equivalent; **ACV**, 128 equivalents) was prepared in a Young's tap NMR tube (5 mm) anaerobically. The IPNS* : Fe : ACV complex sample was then connected to the glass apparatus (Figure 4.34 (a)) that was fitted with a vacuum line, a line supplying dilute NO gas, and a balloon. A mild vacuum was created inside the Young's tap fitted NMR tube (5 mm) and the right half of the glass setup, while the left half was put under positive NO pressure, filling the balloon. The tube was subsequently filled with NO gas from the balloon by opening the valve that separates the two compartments (Figure 4.34 (a)). The sample was allowed to equilibrate for 10 minutes before disconnecting from the glass apparatus, inverted in order to mix with the gas in the headspace with the solution, and allowed to equilibrate for 20 minutes (Figure 4.34 (b)). Upon visual inspection of the sample, the colour of the solution appears to have changed from pale yellow to pink, which is indicative of a sulfur to Fe charge transfer in a IPNS : Fe : ACV : NO complex, as observed previously.[60,61] The $^{19}$F NMR spectrum of the

solutions demonstrated a change in chemical shift ($\Delta\delta_F$ 0.06 ppm) of the IPNS* : Fe : ACV complex resonance (to $\delta_F$ –85.18 ppm) compared to the peak before NO exposure, coinciding with sharpening of the resonance ($\delta_F$ –85.18 ppm). When exposing an IPNS* : Fe complex to NO (Figure 4.34 (b)), no substantial difference in resonance or peak shape was observed ($\delta_F$ –84.72 ppm) (Figure 4.34 (d)). Importantly, colourimetric pH measurements of both IPNS* and NO containing solution, suggested the pH had not changed by the exposure to the dilute NO gas. Additionally, IPNS : Fe is known to bind NO.[60,61] These combined results suggest that the change in chemical shift is a consequence of NO interacting/binding with/to the IPNS* : Fe : ACV complex, rather than indirectly affecting the chemical shift of the resonance, e. g. by changing pH or through electronic effects. These results also suggest the presence of **ACV** in the active site is required for the change in chemical shift to occur. This is consistent with the hypothesis based on crystallographic data which suggests that the isopropyl group of the **ACV** Val adopts an alternative conformation upon binding of $O_2$ / NO, resulting at least in partial change of conformation of Ser281, and consequently changes in the conformations of α-helix 3 and β-strand 13. Interestingly, crystallographic evidence suggests two conformations for α-helix 3 and β-strand 13 exist in a 1 : 1 ratio when NO (100% occupancy) is bound to Fe at the active site. The observations by $^{19}F$ NMR demonstrate a single peak, perhaps representing an average resonance of both conformations. Notably, the broad resonance ($\delta_F$ –83.1 to –83.8 ppm) has disappeared. This could be due to precipitation or overlap with the IPNS : Fe : ACV complex resonance after exposure to NO.

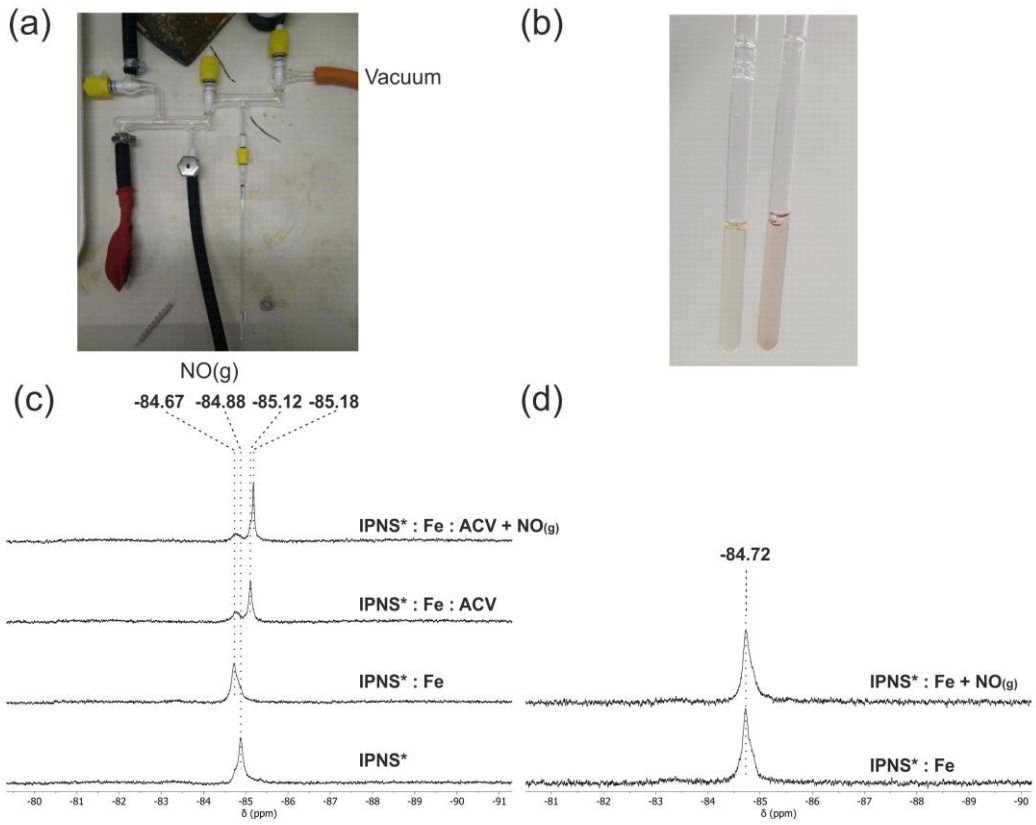



**Figure 4.34:** $^{19}F$ NMR of IPNS* complexes with Fe(II) and **ACV** exposed to nitric oxide (NO, 1000 ppm in $N_2$). (a) Schematics of the setup used to add NO gas to the sample itself; (b) image of a sample of IPNS* : Fe(II) complex (left), and IPNS* : Fe(II) : ACV complex, both after exposure to NO (1000 ppm in $N_2$) gas; (c) A solution of IPNS* (150 μM), $FeSO_4$ (1 equivalent) and **ACV** (128 equivalents), was exposed to NO (1000 ppm in $N_2$ gas) for 1 hour; (d) A solution of IPNS* (150 μM) and $FeSO_4$ (1 equivalent), was exposed to NO (1000 ppm in $N_2$ gas) for 1 hour. Both solutions contain TFA (100 μM) as a standard for referencing.

### 4.6.3 Binding of vanadium(IV) oxide to IPNS*

The binding of vanadium(IV) oxide (V(IV)O) to IPNS* was then investigated by $^{19}F$ NMR. In crystallography studies V(IV)O has been employed as an analogue of the Fe(IV)=O intermediate in 2OG oxygenase catalysis.[62–64] Crystallographic studies of IPNS with V(IV)O, and **ACV** at the active site, exhibited an alternative conformation for α-helix 3. Although it cannot presently be ruled out that the loop movement in the IPNS • V(IV)O • ACV structure may be due to contaminating residual Fe in the IPNS active site, this observation suggests that $O_2$ binding / interaction away from the active site might influence the conformation of α-helix 3 in IPNS. This was thus investigated by $^{19}F$ NMR spectroscopy.

Titration of V(IV)O to *apo*-IPNS* (150 μM) under anaerobic conditions gave rise to the formation of a new resonance ($\delta_F$ --84.66 ppm, Figure 4.35 (a)) with a similar chemical shift as the IPNS* : Fe complex ($\delta_F$ --84.66 ppm, Figure 4.31). Unlike with IPNS* : Fe formation, no evidence for a shoulder or additional new resonances was observed after IPNS* : VO complex formation. This could be due to partial precipitation of IPNS* from solution. A slight excess of V(IV)O (1.5 equivalents) with respect to IPNS* was required to complete IPNS* : VO complex formation. Subsequently, **ACV** was titrated into the complex, giving a new resonance, which was assigned to IPNS* : VO : ACV ($\delta_F$ --85.21 ppm, Figure 4.35 (b)). An excess of **ACV** (108 equivalents) was required to complete formation of the IPNS* : VO : ACV complex. Exposure to $O_2$ from air (1 hour) did not result in a noticeable change in chemical shift of the IPNS* : VO : ACV resonance.

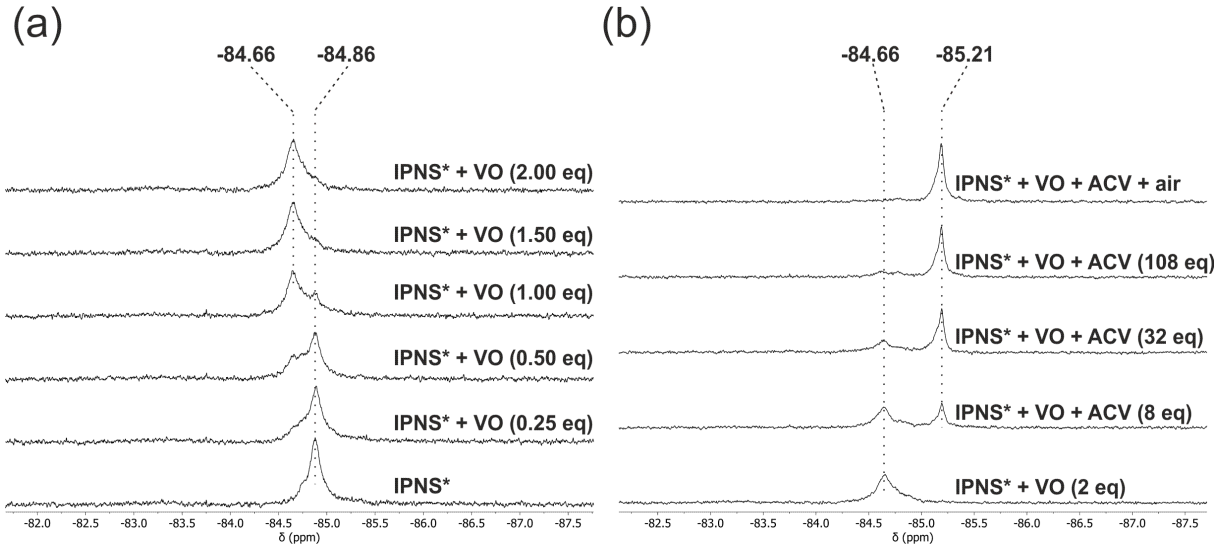



**Figure 4.35:** $^{19}F$ NMR of a titration of V(IV)O followed by **ACV** to IPNS*. (a) A solution of IPNS* (150 μM) is titrated with V(IV)O; (b) a solution of IPNS* (150 μM) and V(IV)O (2 equivalents) is titrated with **ACV**, followed by exposure to air for 1 hour.

# 4.7 Binding of isopenicillin N to IPNS

## 4.7.1 Introduction

Analyses of the electron density of the IPNS • Fe • IPN structure (Section **4.3.7**), obtained through turnover of IPNS • Fe • ACV complex by exposure to $O_2$ from air, clearly showed **IPN** bound to the active site at high occupancy. Interestingly, the occupancy for Fe was observed to be reduced (50%), while the occupancy for **IPN** remained was apparently 100%. Subsequent IPNS • Fe • IPN structures obtained by turnover of IPNS • Fe • ACV crystals similarly showed a reduced occupancy for Fe, suggesting a further decrease in Fe occupancy with prolonged $O_2$ exposure. Several possible explanations for this observation were considered: (i) it might be that Fe(II) leaves the active site after catalysis; (ii) upon completion of the catalytic cycle, steric constraints in the (crystalline) active site lead to the release of Fe; (iii) After completing the catalytic cycle, the Fe may be oxidised and in doing so lose binding affinity to the active site, slowly being released (this is less likely since one would expect higher Fe oxidations states to bind more tightly. In order to investigate this observation in solution, binding studies of Fe(II) to IPNS in the presence and absence of **IPN** were conducted using non-denaturing mass spectrometry (MS) and $^{19}F$ NMR spectroscopy. The studies specifically investigated the binding of Fe(II) to IPNS in presence of **IPN**, in order to verify if it would be possible that Fe leaves the active site of IPNS after catalytic turnover.

## 4.7.2 Non-denaturing mass spectrometry

*These experiments were carried out in collaboration with Denis Shutin, and Dr Patrick Rabe (University of Oxford).*

Recombinant IPNS from *A. nidulans* was used for non-denaturing ESI mass spectrometry (MS), and was produced using a pBJ703 vector construct in NM554 *E. coli* cells.[9] Purification of the protein was done by FPLC, using Q-sepharose FF and Superdex 75 chromatography,[1] as described in Section **3.2.7**. The enzyme was buffer exchanged to ammonium acetate three times, to ensure all residual salts were removed.

The active site of IPNS binds Fe(II), **ACV**, and **IPN**. In order to investigate the binding events of these components separately, non-denaturing MS was used. The mass spectra of *apo*-IPNS manifested a m/z of 3400.12 Da (11 charge state), which corresponds to a deconvoluted mass of 37401 Da (Figure 4.36 (a)), which is ~121 Da lower than anticipated from calculations of the theoretical mass (37522 Da) derived from

the amino acid sequence, suggesting *C*-terminal Thr331 residue was removed. Upon titration of Fe(II) acetate, a mass shift corresponding to binding of a single Fe (56 ± 1 Da) to IPNS was observed (Figure 4.36 (a)). Additionally, a second mass shift was observed, which corresponded to binding of Fe and acetate (115 ± 1 Da, 59 Da acetate, 56 Da Fe). Mono metal ion complex formation was apparently complete with the addition of stoichiometric amounts of Fe(II). Addition of an excess of Fe(II) (5 equivalents) did not result in a further mass shift. Similarly, a titration of Zn(II) acetate, resulted in chelation of Zn (64 ± 1 Da) to IPNS (Figure 4.36 (b)).

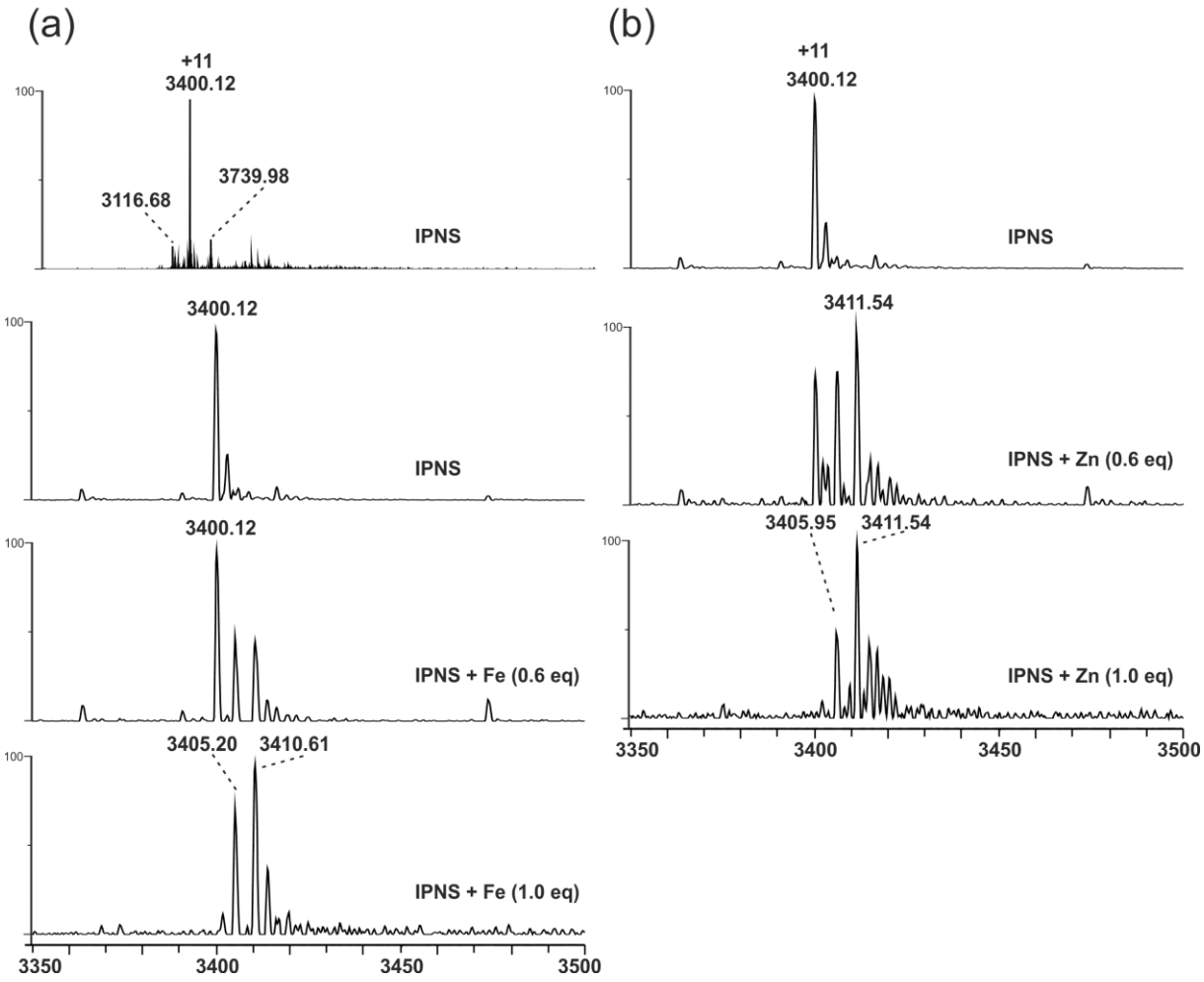



**Figure 4.36:** Non-denaturing mass spectra of a titration of Fe(II) and Zn(II) acetate to *apo*-IPNS. (a) Titration of *apo*-IPNS (18 μM) with Fe(II) acetate; (b) titration of *apo*-IPNS (18 μM) with Zn(II) acetate. The spectra showed the 11+ charge state of IPNS.

Binding of **IPN** to IPNS was then investigated by non-denaturing MS. Addition of an excess of **IPN** (5 equivalents, Figure 4.37 (a)) was required to observe a mass shift corresponding to the IPNS : IPN bound state (360 ± 1 Da, ~1 : 5 ratio to *apo*-IPNS). *Apo*-IPNS peaks remained observable, even when a larger excess of **IPN** (25 equivalents) was used (Figure 4.37 (a)), suggesting binding affinity of **IPN** towards IPNS is lower than for Fe(II) or Zn(II), under the tested conditions. Interestingly, an additional mass shift (722 ± 1 Da) corresponding to a mass of a IPNS : IPN (1 : 2) complex was observed (Figure 4.37 (a)). Note that given the reactivity of the β-lactam of the penicillin ring it is the possibility that (partial) protein acylation occurs. The relative intensities suggest that IPNS (56%) is most abundant, followed by IPNS : IPN (1 : 1) (23%), and IPNS : IPN (1 : 2) (21%, Figure 4.37 (a)). Upon addition of Fe(II) acetate (1 equivalent) to this solution, all aforementioned IPNS protein complexes exhibit a mass shift (~55 ± 1 Da) (Figure 4.37 (a)). In addition to the mass shift as a result of Fe(II) binding, the relative abundance of the species changed: IPNS : Fe (67%), IPNS : Fe : IPN (1 : 1 : 1) (25%), and IPNS : Fe : IPN (1 : 1 : 2) (8%). A similar result was obtained using Zn(II) acetate, showing a mass shift corresponding to binding of Zn(II) to the IPNS complexes (Figure 4.37 (b)). These combined results suggest that binding of Fe(II) and Zn(II) to IPNS is stronger than **IPN**, and that the presence of **IPN** does not obstruct the metal ion from binding to IPNS.

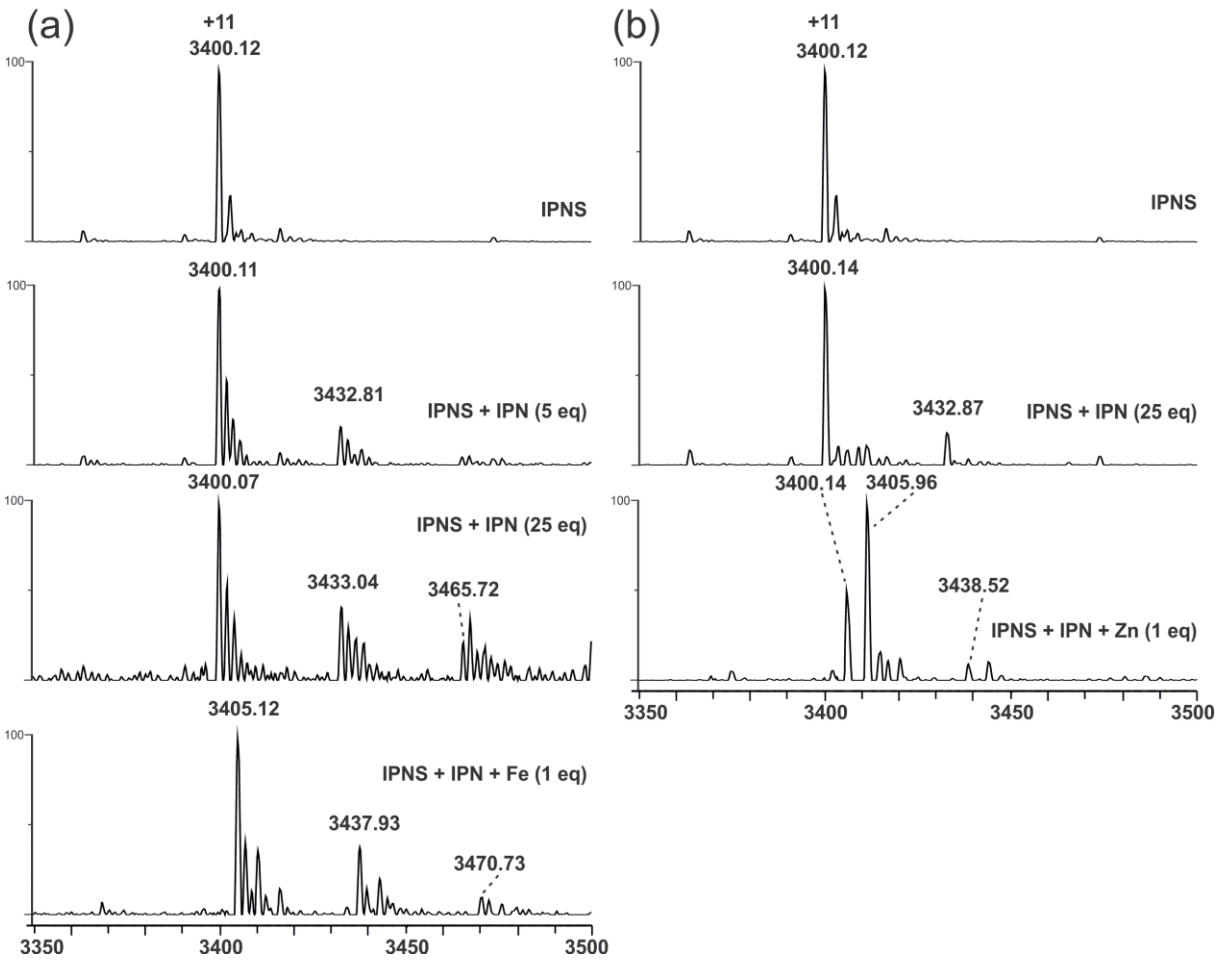



**Figure 4.37:** Non-denaturing mass spectra of a titration of Fe(II) acetate and Zn(II) acetate to *apo*-IPNS. (a) Addition of Fe(II) acetate to *apo*-IPNS (18 μM) in presence of IPN (25 equivalents); (b) addition of Zn(II) acetate (1 equivalent) to *apo*-IPNS (18 μM) in presence of **IPN** (25 equivalents). The spectra show the 11+ charge state of IPNS.

### 4.7.3 Binding of IPN to IPNS variant by $^{19}$F NMR spectroscopy

The binding of **IPN** to IPNS was also investigated by NMR. The $^{19}$F labelled variant IPNS, IPNS*, was used to compare binding of Fe(II) in presence and in absence of **IPN**. Titration of **IPN** to IPNS* did not result in the formation of a new resonance, but instead led to signal broadening, suggesting weak binding to IPNS*. An excess of **IPN** (10 equivalents), led to moderate signal broadening (Figure 4.38 (a)). Titration of Fe(II) to the IPNS (150 μM) and **IPN** (10 equivalents) mixture (Figure 4.38 (b)) resulted in a similar observation to that observed with the titration of Fe(II) to IPNS (150 μM) (Figure 4.31). These combined results suggest the binding of Fe(II) to IPNS* is stronger than binding of **IPN** to IPNS*, which is in agreement with the observations by non-denaturing MS.

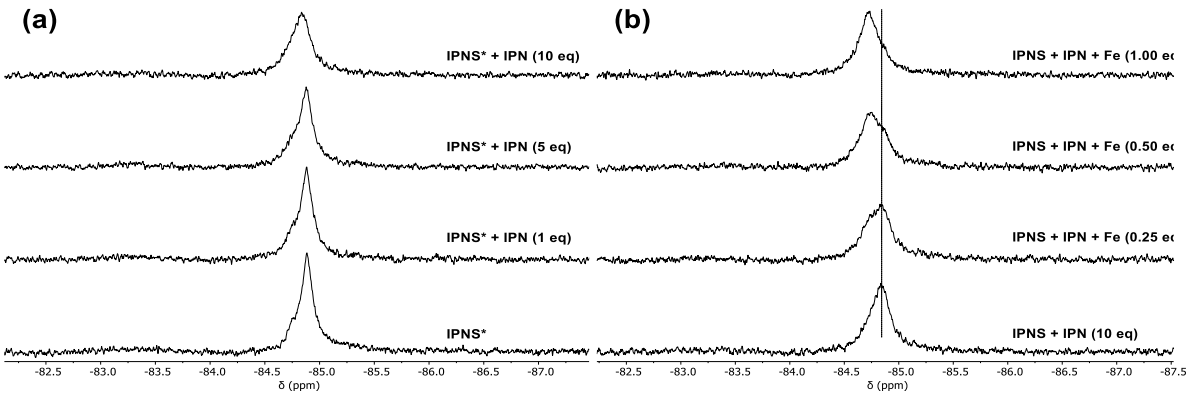



**Figure 4.38:** $^{19}F$ NMR of a titration of **IPN** and Fe(II) to IPNS*. (a) A solution of IPNS* (150 μM) is titrated with **IPN**; (b) a solution of IPNS* (150 μM) and **IPN** (10 equivalents) is titrated with Fe(II)$SO_4$.

The combined MS and NMR studies support the hypothesis that the reduced Fe occupancy following exposure of the IPNS • Fe • ACV microcrystals to $O_2$ is due to crystal lattice constraints that hinder release of **IPN** from the IPNS active site with a consequent 'pressure' to force Fe out of the active site. This proposal is also consistent with weaker binding of the **IPN** product than the substrate **ACV** to IPNS, consistent with the efficient catalysis by IPNS observed in a cellular context.

## 4.8 Discussion

### 4.8.1 IPNS time resolved crystallography

The results in this chapter demonstrate that IPNS • Fe • ACV microcrystals can be grown anaerobically using batch methods and could be used in serial crystallography / SFX experiments. The IPNS • Fe • ACV microcrystals were exposed to $O_2$, and the progress of the reaction was monitored by tr-SFX using an XFEL radiation source. The anaerobic room temperature structure of IPNS • Fe • ACV showed an almost identical fold to the reported IPNS • Fe • ACV structure obtained under cryo conditions from macrocrystals, with an anticipated overall increase in B-factors due to the higher temperature at which the data were collected. The **ACV** substrate adopts a similar conformation in both the room temperature structure and in the cryo structure. Incubation of the IPNS • Fe • ACV microcrystals with $O_2$ (1600 ms) resulted in the observation of additional density near the Fe centre at the active site, which was modelled as an Fe(III) superoxide complex in agreement with the proposed IPNS mechanism (Figure 1.13 and Figure 4.1). The isopropyl group of the **ACV** Val was observed to move away slightly from the Fe centre when compared to the anaerobic structure to accommodate for the $\eta^1$-bonding of $O_2$ to the metal centre. The observations differ from the IPNS • Fe • ACV • NO complex in which the isopropyl group of the **ACV** Val was found to move away further in comparison to the IPNS • Fe • ACV • $O_2$ structure, to accommodate for binding of NO. Unexpected additional electron density for α-helix 3 and β-strand 13 were observed in the structure of the IPNS • Fe • ACV microcrystals exposed to $O_2$ (1600 ms). The additional electron density implies an $O_2$ induced alternative conformation for α-helix 3 and β-strand 13. Although qualitative comparison of the electron density maps at different time points (anaerobic, 800 ms, and 1600 ms) show an increased electron density for the alternative conformation of α-helix 3 and β-strand 13 with an increase in $O_2$ incubation time, the occupancies of the alternative conformation of α-helix 3 and β-strand 13 are not reflected by the occupancy of $O_2$ at the active site. Moreover, the conformation for α-helix 3 and β-strand 13 observed under anaerobic conditions remains the major conformation at the time points (800 ms, 1600 ms) thus far analysed. This alternative helix conformation has not been observed in the reported complex structures (anaerobic, $O_2$ and NO exposed). Further incubation of the IPNS • Fe • ACV microcrystals with $O_2$ coincided with a change in unit cell parameters. In SFX a unit cell distribution is common, as data is collected from different crystals, even when data is collected on different crystal batches.

The changes observed in unit cell parameters upon $O_2$ incubation were small (~0.4%), but clearly present. At present there is no evidence connecting the change in unit cell parameters to the observed additional conformation in α-helix 3 and β-strand 13. It is possible that it relates to dehydration due to the prolonged residence time of the droplets on the Kapton belt.

The first complete IPNS • Fe • IPN product structure was obtained by turnover of IPNS • Fe • ACV microcrystals by exposure to air. The overall fold of the IPNS was highly similar to the reported partial IPNS • Fe • IPN product structure.[5] Previous attempts to obtain an **IPN** complex with extended exposure to $O_2$ were unsuccessful, resulting in low electron density at the active site, suggesting dissociation of **IPN** from the active site within 640 min. Interestingly, although the occupancy for **IPN** remained 100%, the occupancy for Fe was reduced to 50%. A subsequent structure of IPNS • Fe • IPN showed complete absence of Fe. Although in these circumstance the inability of IPNS to release **IPN** after formation was advantageous, allowing for an IPNS • Fe • IPN product structure to be obtained, it is arguable to what extend these structures represent product formation in solution.

Alongside the diffraction data, XES data was collected on the samples. XES data acquired on an initial time point (500 ms $O_2$ exposure), suggested that the oxidation state (predominantly) changed from Fe(II) to Fe(III). This observation is consistent with the proposed oxidation of Fe because of $O_2$ binding. Unfortunately, the XES data collected on the time points of the second LCLS trip (i. e. time points 800 ms and 1600 ms) did not contain sufficient data to provide conclusive evidence to determine the oxidation state of Fe. This could be due to a combination of shorter XES data acquisition times compared to the previous experiment or be due to fewer germanium crystals (i. e. 1 versus 4 usable crystals) available for the von Hamos setup.

Overall, the progression of the IPNS reaction has been observed *in crystallo*; however, the time points at which the intermediates have been observed differ from previously reported observations in solution.[10] The formation of Fe(III) superoxide species was proposed to form within 2 ms from mixing with $O_2$ according to stop flow fluorescence with non-deuterated substrate (i. e. **ACV**).[8] Fe(IV) oxo intermediate was reported to accumulate around 25 ms in stop flow fluorescence experiments. The combination of diffraction and

XES data suggests the Fe(III) superoxide species was the major species after 500 ms in the microcrystalline state. It is possible that these differences between the solution and crystalline state reflect the rates of passive diffusion of $O_2$ into the droplets containing microcrystals slowing the onset rate of reaction.

### 4.8.2 Solution state studies on IPNS

The results of substrate turnover experiments in IPNS • Fe • ACV microcrystal experiments using XFEL radiation, supported by synchrotron cryo-structures discussed above (Section **4.3**), showed conformational flexibility in α-helix 3 (residues 49 – 65) and β-strand 13 (residues 280 – 282) in IPNS (Figure 4.19). After exposure of IPNS • Fe • ACV microcrystals to $O_2$ for 800 or 1600 ms, additional electron density was observed for α-helix 3 and β-strand 13. A longer $O_2$ exposure time correlated with a higher intensity and occupancy for the alternative conformation of α-helix 3 and β-strand 13, however the alternative conformations always consisted a fraction (<25%) of the crystals. Interestingly, structural data collected on the anaerobic IPNS • Fe • ACV complex, and IPNS • Fe • IPN complex obtained by exposing IPNS • Fe • ACV microcrystals to $O_2$, both showed a single conformation for α-helix 3 and β-strand 13. Structural data collected on NO exposed cryo-cooled IPNS • Fe • ACV macrocrystals showed a similar conformational change in α-helix 3 and β-strand 13; two conformations for α-helix 3 and β-strand 13 were modelled in a 1 : 1 ratio. This phenomenon was investigated in solution by $^{19}F$ labelling of an IPNS S55C variant, as described in this chapter.

The $^{19}F$ label was installed remote from the active site (~20 Å) on α-helix 3 (residue 55). Titration of Fe(II) to IPNS* (150 μM) under anaerobic conditions was performed, the results illustrate the exceptional sensitivity of fluorine to chemical changes in its microenvironment. Binding of Fe(II) was completed with the addition of stoichiometric amounts (1 equivalent). Titration of **ACV** with the IPNS* : Fe(II) complex was completed with the addition of an excess of **ACV** (64 equivalents). Exposure of the IPNS* : Fe(II) : ACV complex to NO (1000 ppm in $N_2$), resulted in a change in chemical shift ($\Delta\delta_F$ 0.06 ppm) of the previously assigned resonance to the IPNS* : Fe(II) : ACV complex, and a visual colour change as reported before,[60,61] suggesting interaction (binding) of NO with the IPNS* : Fe(II) : ACV complex. Importantly, exposure of the IPNS* : Fe(II) complex to NO (1000 ppm in $N_2$) under otherwise identical conditions, did

not result in a change in chemical shift. These combined results suggest that: (i) binding of **ACV** to the Fe(II) at the IPNS* active site is causes the conformational change at α-helix 3 and β-strand 13 to occur; and (ii) the change in chemical shift observed for exposing IPNS* : Fe(II) : ACV is not due to indirect effects (e. g. small changes in pH or the presence of soluble NO). *In crystallo*, the NO exposed IPNS • Fe • ACV crystals show the presence of two conformations for α-helix 3 and β-strand 13. In solution a single resonance is observed, suggesting the resonance represents a time average of the two conformations. These solution state studies, combined with the crystallographic results (Section **3.3.12**), support the hypothesis that to accommodate binding of $O_2$ / NO to Fe(II) in the active site, the **ACV** substrate, and in particular the Val residue, must adopt an alternative conformation to that in the IPNS • Fe • ACV complex. Consequently, the interaction of **ACV** with Ser281 on β-strand 13 is changed, requiring the two neighbouring residues (i. e. Gln280 and Leu282), at least in part, to adopt an alternative conformation. The change in conformation of the residues present on β-strand 13 (residues 280 – 282) alter their interaction with α-helix 3 (residues 49 – 65), allowing an alternative conformation to be adopted. Once the catalytic cycle is completed and **IPN** has formed, no $O_2$ is left at the Fe(II) in the IPNS active site, removing the steric constraints that resulted in the conformational change in Ser281 and its neighbouring residues Gln280 and Leu282, and adopting the same conformation as observed for the anaerobic resting state structure.

The structural data gathered on $O_2$ exposed IPNS • V(IV)O • ACV crystals, suggests a similar conformational change in α-helix 3 and β-strand 13 as observed with IPNS • Fe • ACV crystals under the same conditions. Notably, V(IV)O is fully coordinated when bound to the facial triad, **ACV**, and $H_2O$ at the IPNS active site, and does not allow $O_2$ to bind while bound. These results suggest that $O_2$ binding / interaction with the IPNS • V(IV)O • ACV complex could happen remote from the active site, and contradict the hypothesis that binding of $O_2$ / NO at the active site is the only mechanism by which α-helix 3 and β-strand 13 adopt alternative conformations. However, it could not be ruled out that the IPNS • V(IV)O • ACV crystals used in the $O_2$ exposure studies were likely contaminated with residual Fe(II), and would therefore show alternative conformations for α-helix 3 and β-strand 13. Structural studies on IPNS • V(IV)O • ACV crystals will have to be repeated with completely *apo*-IPNS to fully address this question. In the solution state $^{19}F$ NMR studies described in this chapter, titration of V(IV)O to IPNS* (150 μM) was completed with the addition of an excess of V(IV)O (2 equivalents). Subsequent titration of

**ACV** to form IPNS* : V(IV)O : ACV was completed with the addition of an excess of **ACV** (108 equivalents). Exposing the IPNS* : V(IV)O : ACV complex to $O_2$ from air did not result in any observable change in chemical shift, implying no change in conformation for α-helix 3 or β-strand 13 had occurred. These results imply that it is unlikely that remote $O_2$ binding site present on an IPNS* : V(IV)O : ACV complex (and by extrapolation IPNS* : Fe(II) : ACV) causes the conformational change in α-helix 3 and β-strand 13 as observed in crystals. These solution state data would be in agreement with the hypothesis that binding of $O_2$ / NO to Fe(II) results in a conformational change in α-helix 3 and β-strand 13 due to an alternative conformation of **ACV** at the active site. Repeating the IPNS • V(IV)O • ACV crystal experiments is important to fully address this question.

Crystallographic turnover studies using IPNS • Fe • ACV microcrystals gave the first high occupancy IPNS • Fe • IPN structure. After refinement (*performed by Dr Thomas Leissing and Dr Patrick Rabe, University of Oxford*), the occupancy for Fe was set to 50%, suggesting partial dissociation of the metal from the active site. To investigate whether this is part of the catalytic cycle or an anomality *in crystallo*, solution state NMR studies and non-denaturing MS studies were carried out. Titration studies of Fe(II) to IPNS with analysis by non-denaturing MS suggested a strong affinity of Fe(II) to bind to IPNS, as stoichiometric amounts of metal saturated the IPNS : Fe complex. By contrast, synthetically obtained **IPN** had to be used in excess (25 equivalents) to observe partial binding of **IPN** to IPNS. A fraction of the total IPNS was found to bind **IPN**, suggesting a lower binding affinity of **IPN** to IPNS compared to Fe(II) to IPNS. Interestingly, the excess of **IPN** gave rise to the formation of a IPNS : IPN (1 : 2) complex. Addition of Fe(II) (1 equivalent) to the IPNS : IPN mixture resulted in a general mass shift corresponding to the binding of Fe in all complex masses observed. These results suggest that binding of Fe(II) to IPNS is not affected by the presence of **IPN** in solution, which is in line with the general mechanistic proposal for all oxygenases. Alternatively, $^{19}F$ labelled IPNS* was used to test for binding of Fe(II) in presence of **IPN**. Titration of Fe(II) to IPNS*, resulted in the observation of a new resonance by $^{19}F$ NMR, assigned as the IPNS* : Fe(II) complex. Formation of this complex was completed by the addition of stoichiometric amounts of Fe(II). Titration of **IPN** to IPNS* resulted in broadening of the IPNS* signal by $^{19}F$ NMR, suggesting weak interaction. A titration of Fe(II) to IPNS* with an excess of **IPN** (10 equivalents) in solution, suggested IPNS* : Fe(II) complex formation. The presence of **IPN** broadened the resonances, but did not affect chemical shift of

the resonances. Again, binding of Fe(II) did not seem to be affected by **IPN** in solution, which is in agreement with the observations by non-denaturing MS. These findings are supported by recent crystallographic observations on IPNS • IPN crystals, obtained by exposing IPNS • Fe • ACV crystals to air for 12 weeks (*unpublished data Dr Patrick Rabe, Unviersity of Oxford*). Analysis of the electron density maps suggest no metal is present at the active site. The combined results suggest that Fe does not dissociate from the IPNS active site after formation of **IPN** in typical solution state catalysis. It is possible that after formation of **IPN**, Fe slowly dissociated from the active site of IPNS, likely as a result of steric compression. Binding of **ACV** substrate together with Fe(II) to the active site appears to be less sterically demanding, perhaps due to the flexible nature of the substrate. These results agree with general observations on enzymatic catalysis, in which products do not bind strongly to avoid inhibition. Note that previous attempts to obtain an IPNS • Fe • IPN structure were unsuccessful, likely in part due to release of product from the active site. The overall fold of IPNS in both the reported structure (PDB accession code 1JQE[5]) and structure in the present work are similar. Although the unit cell contains a considerable volume of $H_2O$; coupled with an apparent lack of blocking of movement by symmetry related molecules this suggests conformational changes in the secondary structure are possible in the crystals. It is thus unclear exactly why **IPN** is prevented from dissociating from the active site.

## 4.9 Future work

The data collected on IPNS • Fe • ACV microcrystals incubated with $O_2$, show the potential for room temperature time resolved serial crystallography using XFEL radiation. The focus of the work has been on acquiring data from early time points, the results of which have resulted in a number of unexpected observations. Many of the experimental results were obtained with fully protonated **ACV**. Future work should focus on exploring later time points (>10 s). However, these $O_2$ incubations might pose challenges, such as dehydration of the droplets due to the extended time on the belt, or desynchronization of the catalytic cycle at later incubations[65]. The deuterated substrate analogues **AdCV** and **ACdV** were designed to leverage the two strong KIEs (Figure 4.1 and Figure 1.13) to stall the catalytic cycle at particular intermediate points. Future experiments should focus on the use of the deuterated substrates **AdCV** and

**ACdV**. The use of liquid jets as a sample delivery system could be considered. The increased mixing of the sample might accelerate the turnover, leading to shorter time points, and perhaps an easier approach to detect later intermediates. However, the use of liquid jets requires more sample than for the DOD system. Experiments with IPNS • Fe • AdCV and IPNS • Fe • ACdV microcrystals are also limited by the availability of expensive deuterated substrates. Mutagenesis studies on the residues on β-strand 13 (residues 280 – 282) could provide evidence for the extent to which each of the residues (Gln280, Ser281, and Leu282) are involved in the rearrangement of α-helix 3. Similarly, residues on α-helix 3 (residues 49 – 65) could be probed by mutagenesis for their involvement. Complementary substrate analogue work, with less bulky substrate analogues, such as L-δ-(aminoadipoyl)-L-cysteinyl D-alanine (**ACA**) could provide evidence linking the binding of $O_2$ to Fe at the active site to alternative conformation.

The data presented in this chapter also describe a solution state approach to investigating a conformational change in α-helix 3 and β-strand 13 located on wt-IPNS. IPNS S55C variant was successfully produced and labelled with BTFA. Spectral data collected on IPNS* showed high sensitivity for the $^{19}F$ label at residue 55 on α-helix 3 to metal binding and **ACV** binding at the active site. Exposure of IPNS* : Fe(II) : ACV to dilute NO (1000 ppm, in $N_2$), resulted in an observable change in the $^{19}F$ chemical shift. It would be interesting to see if it is possible to titrate NO into the complex, to investigate the effect of sub-stoichiometric amounts of NO on the change in chemical shift, e. g. would two resonances appear (i. e. in slow exchange on the NMR time scale) or would a gradual change in chemical shift be observed (i. e. in fast exchange on the NMR time scale). The solution state $^{19}F$ NMR experiments suggest that for a change in conformation to occur, **ACV** has to be present at the active site, as IPNS* : Fe(II) did not result in a change in chemical shift upon exposure to NO. It would be interesting to see what the effect is when different substrate analogues are complexed (i. e. IPNS* : Fe(II) : analogue), and exposed to NO. For instance, if LLD-δ-(2-aminoadipoyl)cysteinyl glycine (**ACG**) or LLD-δ-(2-aminoadipoyl)cysteinyl alanine (**ACA**) are complexed with IPNS* and Fe(II), does exposure to NO result in a change in chemical shift? These two substrates would lack the bulky isopropyl group that occupies space around the Fe at the active site. Currently, crystallographic data on the anaerobic structure (Figure 4.29), suggests the conformation of Cys55 is similar to the conformation of Ser55. IPNS* • Fe • ACV crystals could be exposed to NO and analysed, to provide evidence for the alternative conformation of α-helix 3 in IPNS*. Although turnover

assays with IPNS* suggest that the enzyme is catalytically active, assessment of the kinetic parameters of IPNS* and a comparison to wt-IPNS are necessary to provide further evidence for the validity of IPNS* as an appropriate model system for wt-IPNS. Additionally, the $^{19}F$ label on IPNS* provides a useful tool for investigating binding of different metals or substrate analogues to IPNS.

The solution $^{19}F$ NMR spectroscopy and non-denaturing MS studies presented in this chapter demonstrate that binding of Fe(II) to IPNS does not appear to be affected by the presence of **IPN**. This observation suggests that loss of Fe for the IPNS active site, after **IPN** formation in a typical catalytic cycle is unlikely during normal solution catalysis. Currently, it remains unclear what causes the loss of Fe at the IPNS active site in the IPNS • Fe • IPN structure. Importantly, previous attempts to obtain an IPNS • Fe • IPN structure by exposure to hyperbaric $O_2$ (40 bar, 640 min) were unsuccessful,[34] leading to loss of **ACV** derived electron density at the active site due to slow product release. Future studies could involve investigating the different crystal form used in the reported study to investigate the effect of crystal packing on product release.

## 4.10 Conclusions

The results described in this chapter concern the production and purification of IPNS, **ACV**, **AdCV**, and **ACdV**, and their crystallisation. Microcrystal slurries of IPNS • Fe • ACV were successfully prepared under anaerobic conditions and subsequently exposed to $O_2$ to initiate a reaction *in crystallo*. A room temperature structure of IPNS obtained from IPNS • Fe • ACV microcrystals showed great similarity to the reported macrocrystal structure obtained under cryo conditions.[34] Incubation of IPNS • Fe • ACV microcrystal slurries with $O_2$ resulted in additional electron density at the active site and suggests $O_2$ binds in a $\eta^1$ fashion to Fe. Consequently, the isopropyl group of Val in **ACV** in proximity of the metal centre became distorted. In addition, extra electron density around α-helix 3 (residues 49 – 65) and β-strand 13 (residues 280 – 282) was observed in structures obtained from crystal slurries exposed to $O_2$. This additional electron density around α-helix 3 and β-strand 13 was rationalised as an alternative conformation for the residues on α-helix 3 and β-strand 13. Ser281 located on β-strand 13 is interacting with the **ACV** Val carboxylate, while Gln280 and Leu282 are involved in interactions with residues on α-helix 3 (Leu54, Ser55, Thr58, Lys59). A longer exposure to $O_2$ appears to correlate with more of the additional electron density around α-helix 3 and β-strand 13, suggesting a higher occupancy for the alternative conformation. Further investigation is required to demonstrate if the mobility in α-helix 3 and β-strand 13 are also present in solution, and to correlate their appearance with $O_2$ exposure. In addition, the tr-SFX experiments showed it was possible to observe a change in Kα spectra upon exposing IPNS • Fe • ACV microcrystals to $O_2$ (500 ms). The change was rationalised as oxidation of Fe(II) to Fe(III). At present there is insufficient data to provide conclusive evidence for any oxidation states during the experiments of the second visit (800 and 1600 ms $O_2$ exposure).

With respect to the $^{19}F$ NMR studies of IPNS* discussed in this chapter, the following observations were made: (i) exposure of IPNS* : Fe(II) : ACV to NO gas resulted in a change in the $^{19}F$ chemical shift; (ii) exposure of IPNS* : Fe(II) to NO gas did not result in an observable change in $^{19}F$ chemical shift; (iii) the exposure of IPNS* : V(IV)O : ACV to $O_2$ from air did not result in an observable change in chemical shift. These combined observations thus provide supporting evidence that movement of α-helix 3 and β-strand 13 occurs in solution. Importantly, these observations are in agreement with $O_2$ / NO binding at the Fe(II) centre and inducing a conformational change in the **ACV** substrate, that alters the interaction

with neighbouring residues, in particular Ser281 on β-strand 13. This in turn changes the interaction of the neighbouring residues of Ser281 on β-strand 13 (i. e. Gln280 and Leu282) with α-helix 3 (Figure 4.22). The fact that no change in chemical shift is observed when exposing IPNS* : V(IV)O : ACV to $O_2$ from air, is in agreement with these observations, and does not favour a hypothetical remote binding site for $O_2$ / NO in the IPNS* structure that could induce a conformational change in α-helix 3 and β-strand 13 as observed in the crystallographic experiments. Investigations on the binding of Fe to IPNS in presence and absence of **IPN**, suggest that **IPN** does not affect binding of Fe(II) to IPNS in solution. The combined non-denaturing MS and $^{19}F$ NMR results suggest that the observed lower occupancy of Fe at the IPNS active site occurs in crystals, but not in solution.

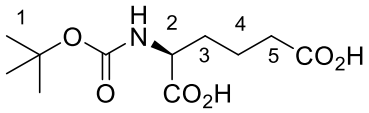
# 4.11 Materials and methods

## 4.11.1 General procedures

Reactions were carried out in oven dried round-bottomed flasks with magnetic stirring in a fume-hood. Anhydrous $CH_2Cl_2$, $Et_2O$, and THF were obtained freshly from an mBraun SPS 800 solvent purification system (alumina column purification). Chemicals and additional anhydrous solvents were obtained from commercial sources and used without further purifications, unless stated otherwise. Reaction progress was monitored by thin layer chromatography (TLC) using Merck silica gel 60 F254 coated on aluminum foil or using a liquid chromatography − mass spectrometry (LC−MS) system, Agilent Technologies 1260 Infinity Series, equipped with a 6120 Quadrupole mass spectrometer and optionally run over a Merck Chromolith Performance C18 (100 × 4.2 mm) HPLC column. Chromatographic purification of synthetic compounds was performed using a Biotage Isolera flash purification system, fitted with Biotage prepacked SNAP KP-Sil or SNAP-ULTRA columns and analytical grade solvents. NMR spectra were recorded using a Bruker AVANCE AV400, Bruker AVII 500 equipped with a TXI H/F/C probe, Bruker AVIII HD 500 equipped with a BBFO probe, Bruker AVII 500 equipped with a He CryoProbe, Bruker AVIII HD 600 equipped with a BB-F/H $N_2$ Prodigy CryoProbe or a Bruker AVIII 700 equipped with a TCI H/C/N He CryoProbe, at their respective frequencies. NMR samples were prepared in 5 mm tubes, and recorded at 298 K, unless stated otherwise. Coupling constants (*J*) are reported to the nearest 0.5 Hz. Chemical shifts ($\delta$) are reported in parts per million (ppm) and referenced to their respective solvent resonance. Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, m = multiplet, dd = doublet of doublet, br s = broad singlet. Infrared (IR) spectra were recorded using a Bruker Tensor 27 instrument, and reported in wavenumbers ($cm^{-1}$). Optical activity, $[\alpha]_D^T$ in $10^{-1}$ deg $cm^2$ $g^{-1}$, values were obtained using a PerkinElmer 341 polarimeter. Samples were prepared in analytical grade solvents, with concentrations reported in g per 100 mL.

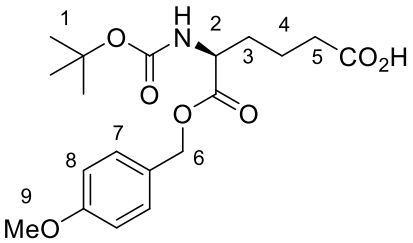
## 4.11.2 Preparation of ACV using solution phase synthesis

### 4.11.2.1 (*S*)-*N*-Boc-α-aminoadipic acid (1)[66]




*An adaptation of the procedure reported by De Cat et al was used.*[66] To a suspension of α-aminoadipic acid (10.0 g, 62.1 mmol, 1 eq.) in $H_2O$ / $^t$BuOH (1 : 1, 200 mL) was added NaOH (6.20 g, 155 mol, 2.5 eq.) portionwise at rt, obtaining a clear solution. $Boc_2O$ (13.3 g, 60.9 mmol, 0.98 eq.) was then added portionwise; the resulting solution was stirred for 48 h. The reaction mixture was diluted with EtOAc (200 mL) then acidified (pH 3) with HCl (1 M, ~160 mL). The layers were separated, and the aqueous layer was extracted with EtOAc (3 × 200 mL). The combined organic layers were washed with brine (1 × 400 mL), dried over $Na_2SO_4$, filtered and concentrated. The desired compound was obtained as clear colourless oil (13.8 g, 52.7 mmol, 85%), and used without further purification. $^1$H NMR (500 MHz, $D_2O$) δ: 4.00 (br s, 1H, $H_2$), 2.33 (t, $J$ = 7.0 Hz, 2H, $H_5$), 1.82 – 1.71 (m, 1H, $H_3$), 1.67 – 1.52 (m, 3H, $H_{3'}$ and $H_4$), 1.33 (s, 9H, $H_1$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 178.1, 176.6, 157.8, 81.5, 53.5, 33.0, 29.9, 27. 6, 20.6. $m/z$ (ESI$^+$) 284 (100%, $[M+Na]^+$). HR ESI-MS: 284.1105 ($[M+Na]^+$, $C_{11}H_{19}O_6NNa$, calc. 284.1105. IR (film) $\nu_{max}$ (cm$^{-1}$): 2935, 1710, 1530, 1409, 1248, 1164. $[\alpha]_D^{25}$ --9.3 ° ($c$ = 1.43, $CH_3OH$). The analytical data are in agreement with the reported values.[66]

### 4.11.2.2 (*S*)-1-*para*-Methoxybenzyl-*N*-Boc-α-aminoadipic acid (2)[15]

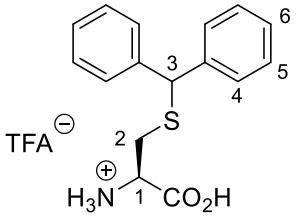



*An adaptation of the procedure reported by Baldwin et al was used.*[15] To a solution of compound **1** (13.8 g, 52.7 mmol, 1 eq.) in DMF (50 mL) under a $N_2$ atmosphere was added $Et_3N$ (8.00 mL, 57.4 mmol, 1.09 eq., anhydrous); the solution was then warmed to 35 °C. Subsequently, *para*-methoxybenzyl chloride (8.20 mL, 58.0 mmol, 1.1 eq.) was added *via* a syringe; the resultant mixture was stirred for 48 h. The reaction mixture was diluted with EtOAc (200 mL) and $H_2O$ / saturated $NH_4Cl$ (1 : 1, 100 mL). The layers were separated and the organic layer was washed with $H_2O$ / saturated $NH_4Cl$ (1 : 1, 6 × 75 mL), followed by brine (1 × 100 mL). The combined organic extracts were dried over $Na_2SO_4$, filtered and concentrated. The crude oil was purified by flash column chromatography ($SiO_2$, EtOAc in $CH_2Cl_2$, 0 – 30%): the desired compound was obtained as a clear colourless oil (7.44 g, 19.5 mmol, 37%). $^1$H NMR (500 MHz, $CD_3OD$) δ: 7.30 (d, $J$ = 8.5 Hz, 2H, $H_7$), 6.90 (d, $J$ = 8.5 Hz, 2H,

$H_8$), 5.13 (d, $J$ = 12.0 Hz, 1H, $H_6$), 5.05 (d, $J$ = 12.0 Hz, 1H, $H_{6'}$), 4.18 – 4.08 (m, 1H, $H_2$), 2.29 (t, $J$ = 7.0 Hz, 2H, $H_5$), 1.88 – 1.75 (m, 1H, $H_3$), 1.79 – 1.60 (m, 3H, $H_{3'}$ and $H_4$), 1.44 (s, 9H, $H_1$). $^{13}$C NMR (126 MHz, $CD_3OD$) δ: 175.5, 172.8, 159.8, 156.7, 129.8, 127.9, 113.6, 79.3, 66.4, 54.4, 53.6, 32.8, 30.6, 27.4, 21.0. $m/z$ (ESI$^-$ ) 404 (100%, $[M+Na]^+$). HR ESI-MS: 404.1679 ($[M+Na]^+$, $C_{19}H_{27}O_7NNa$, calc. 404.1680. IR (film) $\nu_{max}$ (cm$^{-1}$): 2937, 1710, 1515, 1409, 1248, 1165. $[\alpha]_D^{25}$ – 2.9 ° ($c$ = 2.0, $CHCl_3$).

#### 4.11.2.3 (*S*)-Benzhydryl-L-cysteine TFA salt (3)[15,67]

*An adaptation of the procedure reported by Baldwin et al. and Elkins et al was used.*[15,67] To a solution of *L*-cysteine (14.00 g, 115.6 mmol, 1 eq.) dissolved in TFA (10 mL), diphenylmethanol (23.42 g, 127.1 mmol, 1.1 eq.) was added portionwise at rt. The resulting solution was stirred for 1 h. Afterwards, TFA was removed under reduced pressure. The residual white solid was washed with $Et_2O$ (1 × 100 mL), filtered and dried under reduced pressure. The desired compound was obtained as a white solid (30.08 g, 104.7 mmol, 91%). $^1$H NMR (500 MHz, $CD_3OD$) δ: 8.29 – 8.23 (m, 4H, $H_4$), 8.14 – 8.08 (m, 4H, $H_5$), 8.04 – 8.00 (m, 2H, $H_6$), 6.21 (s, 1H, $H_3$), 4.39 (dd, $J$ = 8.0, 4.0 Hz, 1H, $H_1$), 3.64 (dd, $J$ = 14.0, 8.0 Hz, 1H, $H_2$), 3.53 (dd, $J$ = 14.0, 8.0 Hz, 1H, $H_{2'}$). $^{13}$C NMR (126 MHz, $CD_3OD$) δ: 170.6, 142.5, 142.3, 129.6, 129.5, 129.3, 129.2, 128.3, 54.0, 54.0, 33.7. $^{19}$F NMR (471 MHz, $CD_3OD$) δ: –76.55. $m/z$ (ESI$^+$) 288 (100%, $[M+H]^+$). HR ESI-MS: 288.1051 ($[M+H]^+$) $C_{16}H_{18}O_2NS$, calc. 288.1053. IR (film) $\nu_{max}$ (cm$^{-1}$): 2361, 1608, 1417, 700.

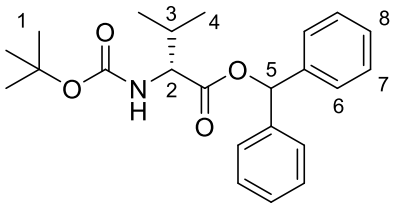



#### 4.11.2.4 Diazodiphenylmethane[18]

*An adaptation of the procedure reported by Miller was used.*[18] To a suspension of HgO (13.0 g, 60.0 mmol, 2.4 eq.) and $Na_2SO_4$ (5.80 g, 40.8 mmol, 1.6 eq.) in $Et_2O$ (100 mL), was added benzophenone hydrazine (4.90 g, 25.0 mmol, 1 eq.) at rt. A solution of KOH (2 mL, sat. solution in EtOH) was added to the suspension. The reaction mixture was shielded from direct light by covering the round bottom flask with aluminium foil. Subsequently, benzophenone hydrazine (5.00 g, 25.5 mmol, 1 eq.) was added portionwise. After stirring for 90 min, the deep purple / brown suspension was filtered and the solids were washed with $Et_2O$ (2 × 50 mL). The solvents were removed by rotary evaporation, applying mild pressure reduction and minimal

heat (30 °C), while protecting the compound from direct exposure to light using aluminium foil. The purple oil crystallised on cooling and was stored at –20 °C to avoid decomposition. The product decomposes slowly at room temperature in the dark, and more rapidly on exposure to light, and was thus used immediately without further purification.

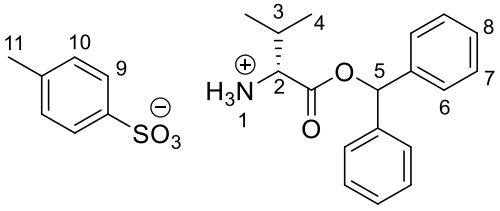
#### 4.11.2.5 Benzhydryl *N*-Boc-D-valinate (4)[15]




*An adaptation of the procedure reported by Baldwin et al was used.*[15] A solution of *N*-Boc-D-valine (2.00 g, 9.20 mmol, 1 eq.) in $CH_2Cl_2$ (10 mL) was prepared and stirred vigorously at rt. Pellets of $Ph_2CN_2$ (1.97 g, 10.1 mmol 1.1 eq.) were added until the pink colour of the solution persisted. The resulting solution was stirred for an additional 1 h. The solvent was removed under reduced pressure, and the crude solid was purified using flash column chromatography ($SiO_2$, EtOAc in *c*-hexane, 0 – 25%): the desir ed compound was obtained as a white solid (2.27 g, 5.90 mmol, 64%).

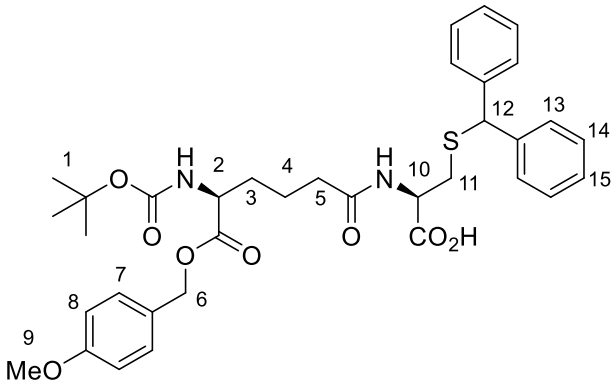
*Alternatively* ***4*** *was prepared according to a procedure described by Squitieri et al.*[19] To a solution of *N*-Boc-D-valine (2.00 g, 9.20 mmol, 1 eq.), $MnO_2$ (6.40 g, 73.6 mmol, 8 eq.), and $KH_2PO_4$ (10.0 g, 73.6 mmol, 8 eq.) in $CH_2Cl_2$ (90 mL) was added a solution of benzophenone hydrazone (2.17 g, 11.0 mmol, 1.2 eq.) in $CH_2Cl_2$ (50 mL) dropwise over 5 min at rt. After 16 h, the reaction mixture was filtered through a plug of silica and washed with EtOAc (3 × 40 mL). The resultant filtrate was concentrated under reduced pressure, and the crude product purified by flash column chromatography ($SiO_2$, EtOAc in *c*-hexane, 0 – 25%): the desired compound was obtained as a white solid (2.11 g, 5.50 mmol, 60%). $^1$H NMR (400 MHz, $CDCl_3$) δ: 7.27 (d, $J$ = 4.0 Hz, 2H, $H_6$), 7.25 – 7.20 (m, 2H, $H_7$), 5.17 – 5.03 (m, 1H, $H_5$), 4.34 (d, $J$ = 4.5 Hz, 1H, $H_2$), 2.10 – 2.02 (m, 1H, $H_3$), 1.43 (s, 9H, $H_1$), 0.89 (d, $J$ = 7.0 Hz, 3H, $H_4$), 0.75 (d, $J$ = 7.0 Hz, 3H, $H_4$'). $^{13}$C NMR (126 MHz, $CDCl_3$) δ: 171.6, 155.7, 139.8, 139.6, 128.5, 128.1, 128.0, 127.4, 127.0, 79.8, 77.8, 58.6, 31.3, 28.3, 19.1, 17.2. $m/z$ (ESI$^+$) 406 (100%, $[M+Na]^+$). HR ESI-MS: 406.1990 ($[M+Na]^+$), $C_{23}H_{29}O_4NSNa$, calc. 406.1989. IR (film) $\nu_{max}$ (cm$^{-1}$): 2969, 2361, 1714, 1497, 1366, 1156, 699.

#### 4.11.2.6 *para*-Toluenesulfonyl benzhydryl D-valinate (5)



*Was prepared according to the procedure as described by Baldwin et al.*[15] To a solution of benzhydryl *N*-Boc-D-valinate (1.00 g, 2.60 mmol, 1 eq.) in THF (10 mL), was added *para*-toluenesulfonic acid (496 mg, 2.60 mmol, 1 eq.) at rt. The solution was stirred for 1 h, concentrated under reduced pressure. This process was repeated thrice. The desired compound was obtained as clear and colourless oil (1.18 g, 2.60 mmol, apparent quantitative yield), and was used without further purification. $^{1}$H NMR (500 MHz, $CDCl_3$) δ: 8.10 (br s, 3H, $H_1$), 7.67 (d, $J$ = 8.0 Hz), 2H, $H_9$), 7.34 – 7.20 (m, 10H, $H_6$ and $H_7$ and $H_8$), 7.05 (d, $J$ = 8.0 Hz, 2H, $H_{10}$), 6.92 (s, 1H, $H_5$), 4.38 (d, $J$ = 5.0 Hz, 1H, $H_2$), 2.31 (s, 3H, $H_{11}$), 2.19 – 2.10 (m, 1H, $H_3$), 0.94 (d, $J$ = 7.0 Hz, 3H, $H_4$), 0.79 (d, $J$ = 7.0 Hz, 3H, $H_{4'}$). $^{13}$C NMR (126 MHz, $CDCl_3$) δ: 173.4, 141.0, 140.1, 139.8, 130.4, 129.5, 129.0, 127.9, 126.2, 76.9, 58.0, 29.2, 21.3, 19.0, 17.4. $m/z$ (ESI$^+$) 284 (100%, $[M+H]^+$). The analytical data are in agreement with the reported values.[68]

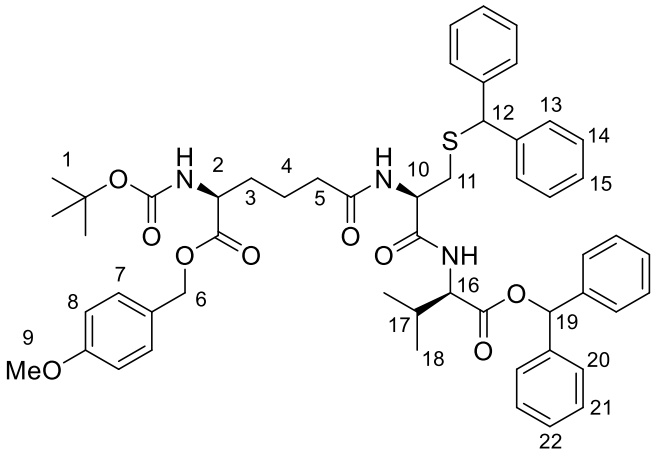
#### 4.11.2.7 (*S*)-Benzhydryl-*N*-((*S*)-5-((tert-butoxycarbonyl)amino)-6-((4-methoxybenzyl)oxy)-6-oxohexanoyl)-L-cysteine (6)[15]



*Was prepared according to the procedure described by Baldwin et al.*[15] A solution of **2** (1.04 g, 2.73 mmol, 1 eq.), and $Et_3N$ (380 µL, 2.73 mmol, 1 eq.) in THF (8 mL) under $N_2$ atmosphere, was cooled to –15 °C. Subsequently, $^{i}$BuOCOCl (370 µL, 2.84 mmol, 1.05 eq.) was added and the resulting solution was stirred for 30 minutes. Afterwards, cysteine **3** (784 mg, 2.82 mmol, 1.1 eq.), and $Et_3N$ (1.14 mL, 8.18 mmol, 3 eq.) were added and the solution was stirred for 24 h, then allowed to warm to rt. Subsequently, HCl (16.4 mL, 1 M), $H_2O$ (40 mL), and EtOAc (50 mL) were added. The layers were separated and the aqueous layer was extracted with EtOAc (3 × 50 mL). The combined organic extracts were washed with brine (50 mL), dried over $Na_2SO_4$, filtered and concentrated. The crude oil was purified by flash column chromatography ($SiO_2$, MeOH in $CH_2Cl_2$

with 0.1% (v/v) AcOH, 0 – 10%): the desired compound was obtained as a clear colourless oil (1.47 g, 2.27 mmol, 83%). $^1$H NMR (500 MHz, $CDCl_3$) δ: 7.41 (d, *J* = 7.5 Hz, 4H, $H_{14}$), 7.34 – 7.26 (m, 4H, $H_{13}$), 7.25 (d, *J* = 8.0 Hz, 2H, $H_7$), 6.88 (d, *J* = 8.0 Hz, 2H, $H_{15}$), 6.48 (d, *J* = 7.5 Hz, 2H, $H_8$), 5.23 (s, 1H, $H_{12}$), 5.08 (q, *J* = 8.0 Hz, 2H, $H_6$), 4.70 (d, *J* = 6.5 Hz, 1H, $H_{10}$), 4.36 – 4.31 (m, 1H, $H_2$), 3.81 (s, 3H, $H_9$), 2.94 – 2.75 (m, 2H, $H_{11}$), 2.35 – 2.14 (m, 2H, $H_5$), 1.84 (d, *J* = 12.0 Hz, 1H, $H_3$), 1.69 (d, *J* = 12.8 Hz, 3H, $H_{3'}$ and $H_4$), 1.43 (s, 9H, $H_1$). $^{13}$C NMR (126 MHz, $CDCl_3$) δ: 173.3, 172.6, 159.8, 155.8, 140.8, 140.6, 130.2, 128.8, 128.7, 128.4, 127.4, 114.0, 80.4, 67.1, 55.3, 54.3, 52.9, 51.5, 35.0, 34.0, 31.8, 28.3, 21.2. *m*/*z* (ESI$^+$) 673 (100%, $[M+Na]^+$). HR ESI-MS: 651.2731 ($[M+H]^+$, $C_{35}H_{43}O_8N_2S$, calc. 651.2735. IR (film) $\nu_{max}$ ($cm^{-1}$): 2938, 2340, 1725, 1710, 1514, 1249, 1169. $[\alpha]_D^{25}$ – 6.0 ° (*c* = 1.2, $CHCl_3$).

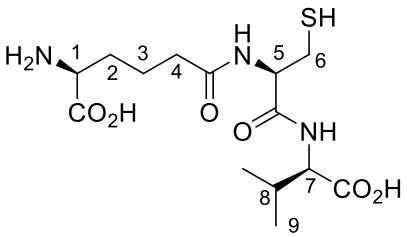
#### 4.11.2.8 *para*-Methoxybenzyl (6*S*,12*R*,15*R*)-12-((benzhydrylthio)methyl)-15-isopropyl-6-(((4-methoxybenzyl)oxy)carbonyl)-2,2-dimethyl-4,10,13-trioxo-3-oxa-5,11,14-triazahexadecan-16-oate (7)[15]




*Was prepared according to the procedure described by Baldwin et al.*[15] To a solution of dipeptide **6** (660 mg, 1.02 mmol, 1 eq.), and $Et_3N$ (144 μL, 2.03 mmol, 2 eq.) in MeCN (15 mL), was added HATU (405 mg, 1.07 mmol, 1.05 eq.), and left stirring for 30 min at rt. Subsequently, valine **5** (463 mg, 1.015 mmol, 1 eq.) was added and the resulting solution was stirred for 25 h. The solution was diluted with EtOAc (25 mL) and $NH_4Cl$ (25 mL, sat.). The layers were separated and the organic layer was washed with $NaHCO_3$ (1 × 25 mL, sat.), and brine (1 × 25 mL), dried over $Na_2SO_4$, filtered and evaporated. The crude oil was purified by flash column chromatography ($SiO_2$, EtOAc in *c*-hexane, 0 – 40%): The desired compound was obtained as a clear colourless oil (309 mg, 0.355 mmol, 35%). $^1$H NMR (500 MHz, $CDCl_3$) δ: 7.47 – 6.86 (m, 22H, $H_7$, $H_{13}$, $H_{14}$, $H_{15}$, $H_{20}$, $H_{21}$ and $H_{22}$), 6.72 (d, *J* = 8.5 Hz, 2H, $H_8$), 6.24 (d, *J* = 7.5 Hz, 1H, $H_{19}$), 5.30 (s, 1H, $H_{12}$), 5.16 – 5.05 (m, 2H, $H_6$), 4.60 (dd, *J* = 8.5, 4.5 Hz, 1H, $H_{10}$), 4.47 (q, *J* = 7.0 Hz, 1H, $H_{16}$), 4.30 (br s, 1H, $H_2$), 3.81 (s, 3H, $H_9$), 2.83 (dd, *J* = 14.0, 7.0 Hz, 1H, $H_{11}$), 2.67 (dd, *J* = 14.0, 7.0 Hz, 1H, $H_{11'}$), 2.28 – 2.09 (m, 2H, $H_5$), 1.82 (d, *J* = 11.7 Hz, 1H, $H_3$), 1.72

– 1.56 (m, 3H, $H_{3'}$ and $H_4$), 1.44 (s, 9H, $H_1$), 0.83 (d, $J$ = 7.0 Hz, 3H, $H_{18}$) 0.80 (d, $J$ = 7.0 Hz, 3H, $H_{18'}$). $^{13}$C NMR (126 MHz, $CDCl_3$) δ: 172.6 172.5, 170.5, 170.2, 159.8, 155.5, 141.0, 141.0, 139.6, 139.4, 130.2, 128.8, 128.6, 128.5, 128.4, 128.4, 128.2, 128.1, 127.5, 127.41, 127.0, 114.0, 79.9, 78.0, 66.9, 57.3, 55.3, 54.5, 53.2, 52.1, 35.4, 34.3, 31.9, 31.2, 28.4, 21.3, 19.1, 17.3. $m/z$ (ESI$^+$) 916 (100%, $[M+Na]^+$). HR ESI-MS: 938.4006 ($[M+Na]^+$), $C_{53}H_{61}O_9N_3SNa$, calc. 938.4020. IR (film) $\nu_{max}$ ($cm^{-1}$): 2934, 2340, 1725, 1711, 1585, 1249, 1167. $[\alpha]_D^{25}$ --2.0 ° ($c$ = 0.9, $CHCl_3$).

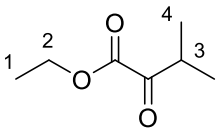
#### 4.11.2.9 L-δ-(α-Aminoadipoyl)-L-cysteinyl-D-valine (ACV)[15]




*Was prepared according to the procedure as described by Baldwin et al.*[15] To a round-bottomed flask charged with protected tripeptide **7** (10 mg, 0.011 mmol, 1 eq.) and anisole (200 μL), was added TFA (2 mL). The resulting solution was stirred at rt for 3 h, after which time it was diluted with $CH_2Cl_2$ (5 mL) and concentrated under reduced pressure. The crude oil was purified by preparative-HPLC: C8 column, MeCN in $H_2O$ with TFA (0.01% (v/v)), followed by purification on the same column using MeCN in $H_2O$ with $NaHCO_3$ (10 mM). Fractions containing the desired product, as confirmed by LC-MS, were lyophilised, to afford the desired compound as a white fluffy powder (3.2 mg, 0.0067 mmol, 60%). $^1$H NMR (500 MHz, $D_2O$) δ: 4.47 (dd, $J$ = 7.0, 5.5 Hz, 1H, $H_5$), 4.03 (d, $J$ = 5.5 Hz, 1H, $H_7$), 3.65 (t, $J$ = 6.0 Hz, 1H, $H_1$), 2.83 (dq, $J$ = 14.0, 6.5 Hz, 2H, $H_6$), 2.33 (t, $J$ = 7.0 Hz, 2H, $H_4$), 2.10 – 1.97 (m, 1H, $H_8$), 1.88 – 1.72 (m, 2H, $H_2$), 1.72 – 1.53 (m, 2H, $H_3$), 0.83 (d, $J$ = 7.0 Hz, 3H, $H_9$), 0.80 (d, 7.0 Hz, 3H, $H_{9'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 177.7, 176.0, 174.4, 171.4, 60.4, 55.8, 54.3, 34.7, 30.5, 29.7, 25.5, 20.8, 18.8, 17.2. $m/z$ (ESI$^+$) 364 (100%, $[M+H]^+$). HR ESI-MS: 364.1536 ($[M+H]^+$, $C_{14}H_{26}O_6N_3S$, calc. 364.1537. IR (film) $\nu_{max}$ ($cm^{-1}$): 2958, 2366, 1634, 1541, 1457, 1404. $[\alpha]_D^{25}$ – 19.8 ° ($c$ = 0.35, $H_2O$). The analytical data are in agreement with the reported values.[69]

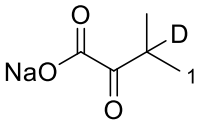
### 4.11.3 Preparation of deuterated valine

#### 4.11.3.1 Ethyl 3-methyl-2-oxobutanoate (12)[25]



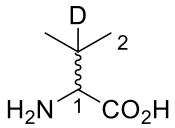

*Was prepared according to the procedure described by Fizet.*[25] A two neck flask equipped with a dripping funnel was charged with diethyloxalate (1.36 mL, 10.0 mmol, 1 eq.) and put under a $N_2$ atmosphere. $Et_2O$ / $CH_2Cl_2$ (1 : 1, 20 mL) was added and the resulting solution was cooled to –10 °C. To this solution was added $^{i}$PrMgBr (2.0 M in THF, 6.35 mL, 12.7 mmol, 1.27 eq.) dropwise using a syringe over 10 minutes. The resulting solution was stirred for an additional 30 min. The mixture was diluted with HCl (aq., 1 M, 12 mL), and the layers were separated. The aqueous layer was washed with $Et_2O$ (2 × 30 mL). The combined organic extracts were dried over $Na_2SO_4$, filtered and evaporated. The crude oil was purified by flash column chromatography ($SiO_2$, EtOAc in c-hexane, 0 – 20%): the desired compound was obtained as a clear colourless oil (1.08 g, 7.47 mmol, 75%). $^1$H NMR (600 MHz, $CDCl_3$) δ: 4.26 (q, *J* = 7.0 Hz, 2H, $H_2$), 3.19 (hept, *J* = 7.0 Hz, 1H, $H_3$), 1.30 (t, *J* = 7.0 Hz, 3H, $H_1$), 1.10 (d, *J* = 7.0 Hz, 6H, $H_4$). $^{13}$C NMR (151 MHz, $CDCl_3$) δ: 198.3, 161.9, 62.2, 37.0, 17.2, 14.0. *m/z* (ESI$^+$) 145 (100%, $[M+H]^+$). The analytical data are in agreement with the reported values.[70]

#### 4.11.3.2 Sodium 3-deutero-3-methyl-2-oxobutanoate (13)[24]



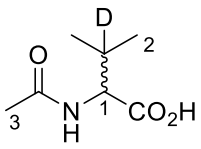

*An adaptation of the procedure reported by Wan and Baldwin was used.*[24] To an emulsion of crude **11** (1.08 g, 7.47 mmol, 1 eq.) diluted with $D_2O$ (50 mL), was added NaOD (30%, 2.5 mL, pH ~12), while stirring vigorously under $N_2$ atmosphere at rt. The resulting emulsion was stirred for 2.5 h, and then extracted with $Et_2O$ (3 × 40 mL). The aqueous solution was acidified with DCl (aq. in $D_2O$, 35%, pH 5), and concentrated. The crude oil was distilled (62 – 70 °C at 20 mbar), affording the desired compound as a clear colourless oil (654 mg, 5.54 mmol, 74%). $^1$H NMR (600 MHz, $D_2O$) δ: 1.06 (s, 6H, $H_1$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 211.5, 181.1, 36.6, 16.34. *m/z* (ESI$^+$) 140 (100%, $[M+Na]^+$). The analytical data are in agreement with the reported values.[71]

#### 4.11.3.3 Racemic 3-deutero valine (14):[24]



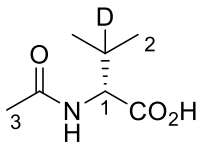

*An adaptation of the procedure reported by Wan and Baldwin was used.*[24] To a suspension of the sodium salt of **13** (654 mg, 5.54 mmol, 1 eq.) in MeOH (40 mL), was added $NH_4Cl$ (603 mg, 11.1 mmol, 2 eq.), and allowed to stir for 15 min at 4 °C. Subsequently, $NaBH_3CN$ (379 mg, 6.00 mmol, 1.09 eq.) was added and the reaction was allowed to warm to rt, and left stirring overnight. Afterwards, HCl (conc., ~2 mL) was added dropwise while vigorously stirring. The solvent was removed under reduced pressure. The crude white solid was suspended in EtOH (40 mL), transferred to a falcon tube (50 mL) and vortexed (2 × 30 s). The tubes were centrifuged (2 min at 4000 rpm), and the liquid was decanted and collected. The EtOH layer was concentrated and dried under reduced pressure. The dry solid was dissolved in EtOH (5 mL) and aniline (3 mL). After three days the desired compound was collected in crystalline form (201 mg, 1.72 mmol, 31%). $^1$H NMR (600 MHz, $D_2O$) δ: 3.55 (d, $J$ = 6.0 Hz, 1H, $H_1$), 1.04 (s, 3H, $H_2$), 0.88 (s, 3H, $H_{2'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 173.9, 60.4, 29.4, 28.6, 18.1, 16.8. $m/z$ (ESI$^+$) 119 (100%, $[M+H]^+$). The analytical data are in agreement with the reported values.[24]

#### 4.11.3.4 Racemic *N*-acyl valine (15)[24]



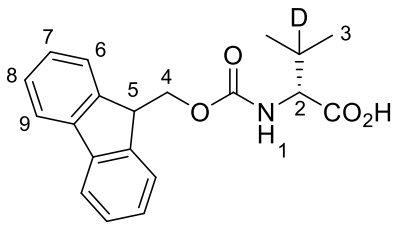

*Was prepared according to the procedure reported by Wan and Baldwin.*[24] A solution of racemic valine **14** (74.0 mg, 0.626 mmol, 1 eq.), and acetic anhydride (500 µL, 5.30 mmol, 8.5 eq.) in acetic acid (4 mL), was heated to 100 °C. After 1 h, the solution was cooled to rt, and the solvent was removed under reduced pressure. The desired compound was obtained as a clear colourless oil (100 mg, 0.626 mmol, apparent quantitative yield) and was used in the next step without further purification. $^1$H NMR (600 MHz, $D_2O$) δ: 3.55 (d, $J$ = 6.0 Hz, 1H, $H_1$), 1.99 (s, 3H, $H_3$), 1.04 (s, 3H, $H_2$), 0.88 (s, 3H, $H_{2'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 173.4, 170.0, 57.2, 29.8, 22.5, 17.9, 16.4. $m/z$ (ESI$^+$) 184 (100%, $[M+Na]^+$). HR ESI-MS: 159.0879 ($[M-H]^-$) $C_{16}H_{18}O_2NS$, calc. 159.0885. IR (film) $\nu_{max}$ (cm$^{-1}$): 2963, 1597, 1397.

#### 4.11.3.5 D-*N*-Acyl valine (D-15)[24]




*An adaptation of the procedure reported by Wan and Baldwin.*[24] The *N*-acylated valine **16** (100 mg, 0.626 mmol, 1 eq.) was dissolved in phosphate buffer (35 mL, 50 mM, pH 7.1, prepared in $D_2O$ (>98%)). Hog porcine kidney acylase (8.75 mg, Simga-Aldrich) was added and the solution was stirred for 22 h at 37 °C. The solution was shortly heated using a heat gun, filtered (0.45 μm filter) and concentrated under reduced pressure. The crude oil was re-dissolved in $H_2O$ (500 μL), and purified using a cation exchange column (Dowex 50W, preequilibrated with HCl (1 M)): the product eluted when washed with $H_2O$, and was lyophilised affording title compound as a white solid (40.2 mg, 0.251 mmol, 40%). $^1$H NMR (600 MHz, $D_2O$) δ: 3.55 (s, 1H, $H_1$), 1.99 (s, 3H, $H_3$) 1.04 (s, 3H, $H_2$), 0.88 (s, 3H, $H_{2'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 173.4, 170.0, 57.2, 29.8, 22.5, 17.9, 16.4. $m/z$ (ESI$^+$) 161 (100%, $[M+H]^+$). HR ESI-MS: 159.0879 ($[M-H]^-$) $C_{16}H_{18}O_2NS$, calc. 159.0885. IR (film) $\nu_{max}$ (cm$^{-1}$): 2963, 1597, 1397. $[\alpha]_D^{25}$ + 5.4 ° ($c$ = 0.2, $H_2O$).

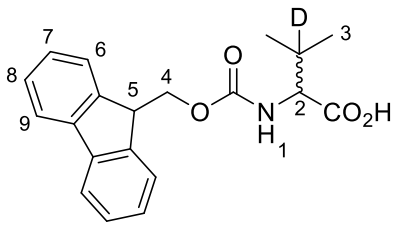
#### 4.11.3.6 Fmoc-D-3-[$^2$H]-valine (D-16)[24]




*An adaptation of the procedure reported by Wan and Baldwin.*[24] The *N*-acylated valine **D-16** (40.2 mg, 0.251 mmol, 1 eq.) was dissolved in HCl (aq., 35%, 4 mL) in $H_2O$ (6 mL). The resulting solution was refluxed for 3 h. The solvent was removed and the crude product was re-dissolved in $H_2O$ / 1,4-dioxane (1 : 1, 10 mL). To this solution was added Fmoc-OSu (93.0 mg, 0.276 mmol, 1.1 eq.), and $Na_2CO_3$ (29.0 mg, 0.276 mmol, 1.1 eq.). The resulting solution was stirred overnight at rt. EtOAc (50 mL) and HCl (aq., 1 M, 20 mL) were added, and the layers were separated. The aqueous layer was extracted with EtOAc (2 × 50 mL). The combined organic extracts were dried over $Na_2SO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography ($SiO_2$, EtOAc with 1% (v/v) HCOOH in *c*-hexane, 0 – 50%): the desired compound was obtained as a white solid (54 mg, 0.159 mmol, 63%). $^1$H NMR (500 MHz, $CDCl_3$) δ: 7.79 (d, $J$ = 7.5 Hz, 2H, $H_9$), 7.62 (d, $J$ = 7.5 Hz, 2H, $H_6$), 7.45 – 7.42 (m, 2H, $H_7$), 7.36 – 7.32 (m, 2H, $H_8$), 5.25 (d, $J$ = 9.0 Hz, 1H, $H_1$), 4.45 (d, $J$ = 7.0 Hz, 2H, $H_4$), 4.39 (d, $J$ = 9.0 Hz, 1H, $H_2$), 4.27 (t, $J$ = 7.0 Hz, 1H, $H_5$), 1.02 (s, 3H, $H_3$), 0.98 (s, 3H, $H_{3'}$). $^{13}$C NMR (126 MHz, $CDCl_3$) δ: 176.9, 156.4, 143.7, 141.3, 127.8, 127.1, 125.1, 120.0, 67.2, 58.9,

47.2, 31.1, 19.1, 17.5. *m/z* (ESI+) 362 (100%, $[M+Na]^+$). HR ESI-MS: 341.1608 ($[M+H]^+$, $C_{20}H_{21}DO_4N$, calc. 341.1606. IR (film) $\nu_{max}$ ($cm^{-1}$): 2929, 1584, 1482, 1442. $[\alpha]_D^{25}$ – 5.2 ° (*c* = 1.0, $CHCl_3$).

#### 4.11.3.7 Racemic *N*-Fmoc 3-deutero valine (16)



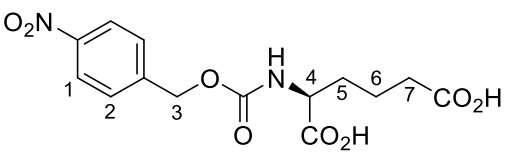

To a solution of crude valine **14** (201 mg, 1.72 mmol, 1 eq.) in $H_2O$ / MeCN (1 : 1, 25 mL), was added $Na_2CO_3$ (200 mg, 1.89 mmol, 1.1 eq.) and Fmoc-OSu (638 mg, 1.89 mmol, 1.1 eq.), and the resulting solution was left stirring for 30 h at rt. The reaction was quenched by the addition of HCl (aq., 2 M, 2 mL), and the aqueous layer was extracted with $Et_2O$ (3 × 20 mL). The combined organic extracts were dried over $Na_2SO_4$, filtered and concentrated. The crude pale yellow oil (200 mg) was purified by column chromatography ($SiO_2$, EtOAc with 1 v/v% HCOOH in *c*-hexane, 0 – 30%): The desired compound was obtained as a white solid (338 mg, 0.990 mmol, 58%). $^1$H NMR (500 MHz, $CDCl_3$) δ: 7.79 (d, *J* = 7.5 Hz, 2H, $H_9$), 7.62 (d, *J* = 7.5 Hz, 2H, $H_6$), 7.44 – 7.41 (m, 2H, $H_7$), 7.36 – 7.33 (m, 2H, $H_8$), 5.25 (d, *J* = 9.0 Hz, 1H, $H_1$), 4.45 (d, *J* = 7.0 Hz, 2H, $H_4$), 4.39 (d, *J* = 9.0 Hz, 1H, $H_2$), 4.27 (t, *J* = 7.0 Hz, 1H, $H_5$), 1.02 (s, 3H, $H_3$), 0.98 (s, 3H, $H_{3'}$). $^{13}$C NMR (126 MHz, $CDCl_3$) δ: 176.9, 156.4, 143.7, 141.3, 127.8, 127.1, 125.1, 120.0, 67.2, 58.9, 47.2, 31.1, 19.1, 17.5. *m/z* (ESI+) 362 (100%, $[M+Na]^+$). HR ESI-MS: 341.1608 ($[M+H]^+$, $C_{20}H_{21}DO_4N$, calc. 341.1606. IR (film) $\nu_{max}$ ($cm^{-1}$): 2929, 2384, 1584, 1482, 1442.

#### 4.11.3.8 Large scale preparation of *N*-Fmoc 3-deutero valine (16)

To a solution of diethyloxalate (20.5 mL, 131 mmol, 1 eq.) in $Et_2O$ / $CH_2Cl_2$ (1 : 1, 100 mL), cooled to --10 °C, was added $^i$PrMgBr (2.0 M in THF, 83.2 mL, 166 mmol, 1.27 eq.) dropwise over 1 h. After stirring for an additional 30 min, the reaction mixture was diluted with HCl (aq., 1 M, 200 mL). The aqueous layer was extracted with $Et_2O$ (2 × 200 mL). The combined organic extracts were washed with brine (100 mL), dried over $Na_2SO_4$, filtered and concentrated under reduced pressure. The crude oil (24.6 g) was carried through without further purification. To an emulsion of the crude oil in $D_2O$ (200 mL), was added NaOH (30%, 10 mL, pH ~12) and stirred vigorously under a $N_2$ atmosphere at rt. After stirring for 2.5 h,

the emulsion was extracted with $Et_2O$ (3 × 150 mL). The aqueous layer was then acidified with DCl (aq., 35%, to pH 9), and concentrated under reduced pressure. The crude white solid (~30 g) was dissolved in MeOH (40 mL), and $NH_4Cl$ (20 g, 381 mmol, 2 eq.) was added and allowed to stir for 15 min at 4 °C. Subsequently, $NaBH_3CN$ (13.0 g, 207 mmol, 1.09 eq.) was added and the reaction was stirred overnight and allowed to warm to rt. Afterwards HCl (aq., ~20 mL) was added dropwise while vigorously stirring. The solvent was removed under reduced pressure. The crude white solid was suspended with EtOH (150 mL), divided over 4 falcon tubes (50 mL) and vortexed (2 × 30 s). The tubes were centrifuged (2 min at 4000 rpm), and the liquid was decanted and collected. The solid was suspended with EtOH (150 mL), vortexed and centrifuged. The combined EtOH layers were concentrated, affording a crude white solid (~14 g). A portion of the crude white solid (~5 g) was dissolved in $H_2O$ / MeCN (1 : 1, 50 mL), and to it was added $Na_2CO_3$ (4.00 g, 37.7 mmol, 1.1 eq.) and Fmoc-OSu (12.4 g, 36.8 mmol, 1.1 eq.), and the resulting solution was stirred for 30 h at rt. The reaction was quenched by the addition of HCl (2 M, 100 mL), and the aqueous layer was extracted with $Et_2O$ (3 × 100 mL). The combined organic extracts were dried over $Na_2SO_4$, filtered and then concentrated. The crude pale yellow oil (18 g) was purified by column chromatography (2 g of crude per column (25 g), $SiO_2$, EtOAc with 1% (v/v) HCOOH in *c*-hexane, 0 – 30%): the desired compound was obtained as a white solid (362 mg, 1.06, ~22% over 4 steps). Characterisation data was identical to the compound isolated in Section **3.7.3.6**.

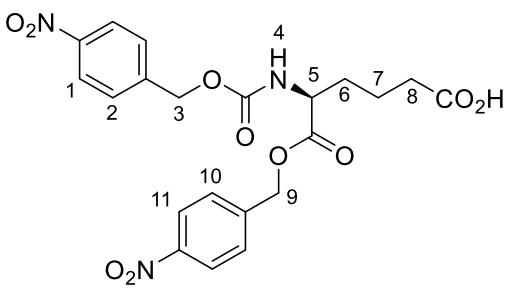
#### 4.11.2.1 *N*-4-nitrobenzyloxycarbonyl (S)-α-aminoadipic acid (17):




*Was prepared according to the procedure reported by Tarhonskaya et al.*[28] To a solution of L-α-aminoadipic acid (203 mg, 1.26 mmol, 1 eq.) and NaOH (116 mg, 2.90 mmol, 2.3 eq.) in $H_2O$ (15 mL) was added *para*-nitrobenzyl chloroformate (380 mg, 1.76 mmol, 1.4 eq.) in 1,4-dioxane (10 mL) dropwise over 5 min at 0 °C. The reaction mixture was allowed to warm to rt and stirred for an hour. The mixture was filtered and diluted with EtOAc (25 mL), and the layers separated. The aqueous layer was acidified with HCl (1 M) to pH 3 and extracted with EtOAc (3 × 25 mL). The combined organic extracts were washed with brine (50 mL), dried over $MgSO_4$, filtered and evaporated. The desired compound was obtained as a white solid (292 mg, 0.857 mmol, 68%). $^1$H NMR (600 MHz, DMSO-$d_6$) δ: 8.22 (d, $J$ = 8.5

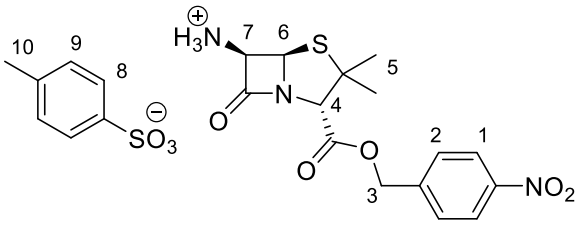
Hz, 2H, $H_1$), 7.62 (d, $J$ = 8.5 Hz, 2H, $H_2$), 5.20 (s, 2H, $H_3$), 4.03 – 3.90 (m, 1H, $H_4$), 2.22 (t, $J$ = 7.5 Hz, 2H, $H_7$), 1.81 – 1.72 (m, 1H, $H_5$), 1.71 – 1.51 (m, 3H, $H_{5'}$ and $H_6$). $^{13}$C NMR (151 MHz, DMSO-$d_6$) δ: 174.6, 174.1, 156.3, 147.4 145.5, 128.5, 124.7, 64.7, 54.1, 33.6, 30.7, 21.7. $m/z$ (ESI$^+$) 341 (100%, $[M+H]^+$). These results are in agreement with the reported data.[28]

#### 4.11.2.2 4-Nitrobenzyl-*N*-4-nitrobenzyloxycarbonyl (S)-α-aminoadipic acid (18):



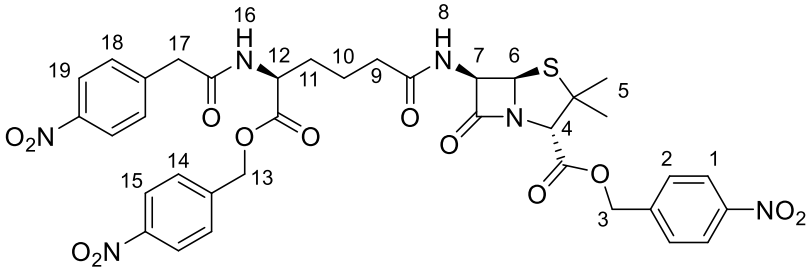

*Was prepared according to the procedure reported by Tarhonskaya et al.*[28] To a solution of diacid **16** (192 mg, 0.565 mmol, 1 eq.) in DMF (5 mL) heated to 80 °C, was added dicyclohexylamine (103 mg, 0.565 mmol, 1 eq.). After stirring for 5 minutes, *para*-nitrobenzylbromide (122 mg, 0.565 mmol, 1 eq.) dissolved in DMF (2 mL) was added. A precipitate formed and the resulting mixture was stirred for an additional 40 min, before cooling down to rt. The mixture was diluted with EtOAc (50 mL) and filtered. The filtrate was washed with $H_2O$ (1 × 25 mL) and brine (1 × 25 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude brown oil was purified by flash column chromatography ($SiO_2$, EtOAc with 1% HCOOH in c-hexane, 20 – 80%): the desired compound was obtained as a clear colourless oil (113 mg, 0.238 mmol, 43%). $^1$H NMR (600 MHz, $CDCl_3$) δ: 8.15 – 8.08 (m, 4H, $H_1$ and $H_{11}$), 7.47 – 7.38 (m, 4H, $H_2$ and $H_{10}$), 5.50 (d, $J$ = 8.0 Hz, 1H, $H_4$), 5.31 – 5.17 (m, 4H, $H_3$ and $H_9$), 4.49 – 4.43 (m, 1H, $H_5$), 2.46 – 2.34 (m, 2H, $H_8$), 1.99 – 1.91 (m, 1H, $H_6$), 1.84 – 1.75 (m, 1H, $H_{6'}$), 1.74 – 1.62 (m, 2H, $H_7$). $^{13}$C NMR (151 MHz, $CDCl_3$) δ: 178.4, 171.8, 155.6, 147.9, 147.7, 143.6, 142.3, 128.6, 128.1, 123.9, 123.8, 65.8, 65.6, 53.8, 36.9, 31.4, 20.3. $m/z$ (ESI$^+$) 476 (100%, $[M+H]^+$). These results are in agreement with the reported data.[31]

#### 4.11.2.3 4-Nitrobenzyl 6-aminopenicillanate tosylate salt (19)



*Was prepared according to the procedures reported by Tarhonskaya et al. and Huffman.*[28,31] *Reaction performed Dr Patrick Rabe (University of Oxford).* To a solution of 6-aminopenicillanic acid (2.00 g, 9.25 mmol, 1 eq.) in acetone (4 mL), was added $Et_3N$ (1.40 mL, 10.1 mmol, 1 eq.) at 0 °C. The reaction was stirred for 30 min, followed by dropwise addition of 4-nitrobenzyl bromide (2.20 g, 10.1 mmol, 1 eq.) in acetone (9 mL). The mixture was stirred for 4 hours, after which it was poured into diethyl ether (30 mL, anhydrous). The suspension was filtered, and the filtrate was washed with aqueous $NaHCO_3$ (30 mL, sat.) and $H_2O$ (30 mL). The organic layer was then treated with *para*-toluenesulfonic acid monohydrate (1.9 g, 10.1 mmol, 1 eq.) in acetone (7 mL), resulting in the formation of desired compound as white precipitate (2.78 g, 5.30 mmol, 57%). $^1$H NMR (600 MHz, DMSO-$d_6$) δ: 8.27 (d, $J$ = 9.0 Hz, 2H, $H_1$), 7.72 (d, $J$ = 9.0 Hz, 2H, $H_8$), 7.48 (d, $J$ = 8.5 Hz, 2H, $H_2$), 7.12 (d, $J$ = 8.5 Hz, 2H, $H_9$), 5.56 (d, $J$ = 4.5 Hz, 1H, $H_7$), 5.38 (d, $J$ = 4.5 Hz, 2H, $H_3$), 5.13 (d, $J$ = 4.5 Hz, 1H, $H_6$), 2.29 (s, 3H, $H_{10}$), 1.65 (s, 3H, $H_5$), 1.44 (s, 3H, $H_{5'}$). $^{13}$C NMR (151 MHz, DMSO-$d_6$) δ: 169.1, 167.3, 147.9, 146.3, 143.2, 138.0, 129.8, 128.5, 126.0, 124.1, 70.2, 66.2, 65.7, 64.7, 57.7, 31.3, 27.1, 21.2. $m/z$ (ESI$^+$) 352 (100%, $[M+H]^+$). These results are in agreement with the reported data.[28]

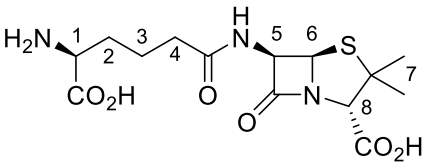
#### 4.11.2.4 N6-δ-(α-4-nitrobenzyl-N-4-nitrobenzyl α-aminoadipoyl)-6-aminopenicillan 4-nitrobenzyl ester (20):



*Was prepared according to the procedure reported by Tarhonskaya and Huffman.*[28,31] *Reaction performed Dr Patrick Rabe (University of Oxford).* A solution of tosylate salt **19** (1.45 g, 1.78 mmol, 1 eq.) in EtOAc (50 mL), was washed with $NaHCO_3$ (1 × 50 mL, sat.), $H_2O$ (1 × 50 mL) and brine (1 × 50 mL), and then dried over $MgSO_4$, filtered and evaporated to afford free base (890 mg, 2.75 mmol, 98%).

To a solution of **18** (989 mg, 2.08 mmol, 1 eq.) and 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline (514 mg, 2.08 mmol, 1.0 eq.) in $CH_2Cl_2$ (30 mL) was added benzylester **19b** (675 mg, 2.08 mmol, 1 eq.) at rt. The mixture was stirred for 20 h, after which the solvent was removed under reduced pressure. The crude solid was dissolved in EtOAc (30 mL). The organic layer was washed with HCl (1 × 30 mL, 1 M), $NaHCO_3$ (1 × 30 mL, sat.), $H_2O$ (1 × 30 mL), and brine (1 × 30 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography ($SiO_2$, EtOAc in $CH_2Cl_2$ 0 – 50%): the desired compound was obtained as a colourless oil (900 mg, 1.13 mmol, 54%). $^1$H NMR (600 MHz, $CDCl_3$) δ: 8.28 – 8.17 (m, 6H, $H_1$, $H_{15}$ and $H_{19}$), 7.57 – 7.48 (m, 6H, $H_2$, $H_{14}$ and $H_{18}$), 6.17 (d, $J$ = 9.0 Hz, 1H, $H_8$), 5.69 (dd, $J$ = 9.0, 4.5 Hz, 1H, $H_7$), 5.56 (d, $J$ = 8.0 Hz, 1H, $H_{16}$), 5.53 (d, $J$ = 4.5 Hz, 1H, $H_6$), 5.28 – 5.23 (m, 4H, $H_3$ and $H_{13}$), 5.21 (s, 2H, $H_{17}$), 4.48 (s, 1H, $H_4$) 4.45 – 4.40 (m, 1H, $H_{12}$), 2.38 – 2.31 (m, 1H, $H_9$), 2.27 – 2.21 (m, 1H, $H_{9'}$), 1.97 – 1.88 (m, 1H, $H_{11}$), 1.81 – 1.68 (m, 3H, $H_{11'}$ and $H_{10}$), 1.63 (s, 3H, $H_5$), 1.43 (s, 3H, $H_{5'}$). $^{13}$C NMR (151 MHz, $CDCl_3$) δ: 173.8, 171.7, 171.2, 167.2, 155.6, 148.1, 147.9, 147.7, 143.5, 142.2, 141.6, 129.1, 128.6, 128.1, 124.0, 123.9, 123.8, 70.5, 68.1, 66.0, 65.8, 65.6, 58.6, 53.6, 34.8, 31.6, 31.5, 27.1, 20.9. $m/z$ (ESI$^+$) 809 (100%, $[M+H]^+$). These results are in agreement with the reported data.[28]

#### 4.11.2.5 Isopenicillin N (IPN):




*An adaptation of the procedure reported by Tarhonskaya and Huffman was used.*[28,31] A suspension of protected IPN **19** (100 mg, 0.124 mmol, 1 eq.), $NaHCO_3$ (10.9 mg, 0.12 mmol, 1 eq.), and 10% Pd/C (100 mg, 1 mass eq.) in $H_2O$ / THF (1 : 1, 20 mL) was stirred at rt. The flask was evacuated and purged with $H_2$ thrice, before stirring for 2.5 h. The suspension was filtered over Celite and the filtrate was concentrated. Further purification of the crude oil by preparative HPLC (1% acetonitrile in $H_2O$ supplemented with $NH_4HCO_3$ (10 mM) for 5 min, then 1 – 18% over 17 min) afforded the desired compound as a white solid (35 mg, 0.098 mmol, 79%). $^1$H NMR (700 MHz, $D_2O$) δ: 5.49 (d, $J$ = 4.0 Hz, 1H, $H_5$), 5.41 (d, $J$ = 4.0 Hz, 1H, $H_6$), 4.17 (s, 1H, $H_8$), 3.66 (dd, $J$ = 7.0, 5.5 Hz, 1H, $H_1$), 2.34 (t, $J$ = 7.5 Hz, 2H, $H_4$), 1.88 – 1.76 (m, 2H, $H_2$), 1.70 – 1.59 (m, 2H, $H_3$), 1.57 (s, 3H, $H_7$), 1.45 (s, 3H, $H_{7'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ 175.9,

174.9, 174.6, 174.5, 73.2, 66.5, 64.4, 57.9, 54.5, 34.5, 30.3, 29.9, 26.5, 20.9. *m/z* (ESI$^+$) 360 (100%, [M+H]$^+$). These results are in agreement with the reported data.[28]

### 4.11.4 Preparation of ACV using solid phase peptide synthesis

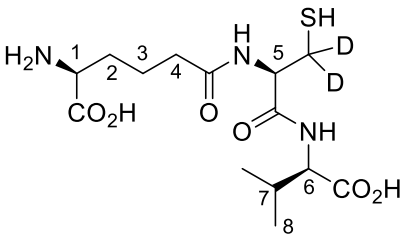
Solid phase peptide synthesis (SPPS) was carried out using a Liberty Blue microwave assisted solid phase peptide synthesizer (CEM corporation). Peptide synthesis on Wang resin (Matrix Innovations) was done sequentially starting from their *C*-termini, using standard Fmoc SPPS protocol.[20,72] When required (for instance with **ACdV**), coupling of the first peptide to the resin was done manually. To a suspension of Wang resin (1.00 g, 1.00 mmol / g, 1 eq.) in $CH_2Cl_2$ / DMF (9 : 1, 10 mL), were added diisopropylcarbodiimide (DIC, 252 mg, 2.00 mmol, 2 eq.), HOBt (270 mg, 2.00 mmol, 2 eq.), DMAP (12.0 mg, 0.10 mmol, 0.1 eq.), and Fmoc-3-[$^2$H]-Val-OH (680 mg, 2.00 mmol, 2 eq.). The solution was stirred slowly for 20 h, at rt. Subsequently, $Ac_2O$ (200 μL, 2.10 mmol, 2.1 eq.) and pyridine (200 μL, 2.40 mmol, 2.4 eq.) were added, and the suspension was stirred for 30 min at rt. The suspension was filtered and the resin washed with DMF, $CH_2Cl_2$ (40 mL), and MeOH (40 mL) (× 3), and dried before using in coupling steps. The coupling steps were carried out with DIC and oxyma in DMF in a microwave vessel at 90 °C, for 10 or 20 min. Afterwards, the peptide was treated with a solution of piperidine in DMF (20% (v/v)) for 10 min to remove the Fmoc group, after which the next coupling step took place. Protected Fmoc-Cys(Trt)-OH (Novabiochem, typically 5 eq. per coupling), and Fmoc-3,3-[$^2H_2$]-Cys(Trt)-OH (Cambridge Isotopes, 1.5 eq. per coupling), were obtained from commercial suppliers, while Boc-Aad(OH)OPMB **2** (typically 1.5 – 2.0 eq. per coupling) was prepared (Section **3.7.2.2**). After all couplings were completed, the peptide was cleaved from the resin, using a TFA based cleavage mixture (75% (v/v) TFA, 2.5% $H_2O$, 2.5% TIPS, 2,2'-(ethylenedioxy)diethanethiol 20%), for 6 h at rt. The resin was then removed by filtration, and the solvent removed ($CH_2Cl_2$ was added to co-evaporate TFA). The peptide was then precipitated from cold $Et_2O$ (40 mL, –80 °C). After centrifuging (4000 rpm, 2 min, 4 °C), the ether layer was decanted and evaporated, then the precipitation process was repeated. The combined precipitates were checked by LC-MS for the presence of the desired peptide, which was subsequently purified by preparative-HPLC.

### 4.11.5 Purification by preparative-HPLC

#### 4.11.5.1 Purification of tripeptides by preparative-HPLC

In the optimised procedure the sample was supplemented with TCEP (1.0 eq.) before purification by preparative-HPLC. Purification of the peptides by preparative-HPLC was done using a Shimadzu HPLC system, equipped with LC20AR pumps, a LC SIL 10A autosampler, a LC SPD-20A UV detector, a LC CBM-20A controller unit, and a LC FRC10A fraction collector. Peptides were purified using a SunFire™ prep column (C18, 5 μm, 150 mm length, 10 mm diameter), typically in two stages. First, samples were eluted by a gradient of acetonitrile in $H_2O$ (2% (v/v) for 5 min, then 2 – 30% over 18 minutes), both supplemented with TFA (0.1% (v/v)). Fractions containing the desired peptide were quickly frozen in $N_2$ (l) and lyophilised. Then, samples were eluted by a gradient of acetonitrile in $H_2O$ (1% for 5 min, then 1 – 18% over 18 min), in which $H_2O$ was supplemented with $NH_4HCO_3$ (10 mM). The fractions were immediately frozen in $N_2$ (l) and lyophilised. **IPN** was purified using a SunFire™ prep column (C18, 5 μm, 150 mm length, 10 mm diameter). The samples were eluted by a gradient of acetonitrile in $H_2O$ (1% for 5 min, then 1 – 18% over 18 min), in which $H_2O$ was supplemented with $NH_4HCO_3$ (10 mM). The fractions containing **IPN**, were immediately frozen in $N_2$ (l) and lyophilised. Fractions were checked for peptides or **IPN** by LC-MS. Peptides and **IPN** were then analysed by $^1$H NMR to check for purity, and used for crystallography, NMR spectroscopy, and non-denaturing MS studies as an ammonium salt.

#### 4.11.5.2 Characterisation of L-δ-(α-aminoadipoyl)-L-3,3-[$^2H_2$]-cysteinyl-D-valine (AdCV):




The **AdCV** peptide was prepared using SPPS as described above (Section **3.7.4**), using Fmoc-L-3,3-[$^2H_2$]-cysteine(Trt)-OH (Cambridge Isotopes). $^1$H NMR (500 MHz, $D_2O$) δ: 4.46 (s, 1H, $H_5$), 4.02 (d, $J$ = 6.0 Hz, 1H, $H_6$), 3.65 (dd, $J$ = 6.5, 5.5 Hz, 1H, $H_1$), 2.33 (t, $J$ = 7.0 Hz, 2H, $H_4$), 2.10 – 1.96 (m, 1H, $H_7$), 1.89 – 1.72 (m, 2H, $H_2$), 1.72 – 1.52 (m, 2H, $H_3$), 0.83 (d, $J$ = 7.0 Hz, 3H, $H_8$), 0.80 (d, $J$ = 7.0 Hz, 3H, $H_{8'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 177.9, 176.0, 174.4, 171.4, 60.6, 55.7, 54.4, 34.7, 30.6, 29.8, 25.4, 20.8, 18.9, 17.2. $m/z$ (ESI$^+$) 366 (100%, $[M+H]^+$). HR ESI-MS:

366.1652 ([M+H]$^+$, $C_{14}H_{24}D_2O_6N_3S$, calc. 366.1662. IR (film) $\nu_{max}$ (cm$^{-1}$): 2961, 1628, 1560, 1457, 1397. $[\alpha]_D^{25}$ – 19.5 ° ($c$ = 0.52, $H_2O$).

#### 4.11.5.3 Characterisation of L-δ-(α-aminoadipoyl)-L-cysteinyl-D-3-[$^2$H]-valine (ACdV):

The **ACdV** peptide was prepared using SPPS as described above (Section **4.7.4**), using either synthetically prepared Fmoc-3-[$^2$H]-Val-OH or Fmoc-L-3-[$^2$H]-Val-OH (Section **4.7.3.8**) coupled to Wang resin (Section **4.7.4**). $^1$H NMR (500 MHz, $D_2O$) δ: 4.48 (dd, $J$ = 7.0, 5.5 Hz, 1H, $H_5$), 4.03 (s, 1H, $H_6$), 3.65 (t, $J$ = 6.0 Hz, 1H, $H_1$), 2.83 (dq, $J$ = 14.0, 6.5 Hz, 2H, $H_6$), 2.33 (t, $J$ = 7.5 Hz, 2H, $H_4$), 1.86 – 1.72 (m, 2H, $H_2$), 1.70 – 1.53 (m, 2H, $H_3$), 0.83 (s, 3H, $H_8$), 0.80 (s, 3H, $H_{8'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 178.0, 176.0, 174.3, 171.6, 59.8, 55.7, 54.3, 34.7, 30.5, 29.7, 25.5, 20.8, 18.6, 17.1. $m/z$ (ESI$^+$) 365 (100%, [M+H]$^+$). HR ESI-MS: 365.1597 ([M+H]$^+$, $C_{14}H_{24}D_2O_6N_3S$, calc. 365.1600. IR (film) $\nu_{max}$ (cm$^{-1}$): 2960, 1632, 1456, 1399. $[\alpha]_D^{25}$ – 19.3 ° ($c$ = 0.37, $H_2O$).

## 4.12 Protein production and purification

### 4.12.1 General methods

Growth medium 2TY was prepared with tryptone (16 g · L$^{-1}$), yeast extract (10 g · L$^{-1}$), and sodium chloride (5 g/L). All media were sterilised prior to growth using an autoclave.

SDS-PAGE were prepared according to the recipe in table 4.5 using the Mini-PROTEAN Tetra Cell system (Bio-Rad). During preparation of the running gel, the buffer was covered with isopropyl alcohol to prevent it from drying out whilst setting. Tetramethylethylenediamine (TEMED) was added last to both the running and the stacking gel.

**Table 4.5 Reagents for the preparation of the stacking and running gel (4 SDS-PAGE).**

| Reagent | Running gel | Stacking gel |
|---|---|---|
| Tris(hydroxymethyl)aminomethane | 4.00 mL (1.5 M, pH 8.8) | 1.25 mL (0.5 M, pH 8.5) |
| $H_2O$ | 5.06 mL | 3.05 mL |
| Ammonium persulfate (10 w/v%) | 267 µL | 50 µL |
| Bis-acrylamide (30 w/v%) | 6.67 mL | 670 µL |
| Sodium dodecyl sulfate (10 w/v%) | 160 µL | 50 µL |
| Tetramethylethylenediamine | 13.3 µL | 5.0 µL |

Loading (2 ×) and running (10 ×) buffers were prepared according to the recipes below (Table 4.6). Protein samples (8 μL) were mixed with loading buffer (8 μL), and heated to 95 °C for 10 minutes. Subsequently, the samples (10 μL) were loaded onto an SDS-PAGE. Electrophoresis was run at constant voltage (180 V, 50 minutes). Gels were then stained with SDS-PAGE staining solution (10 min), and de-stained with $H_2O$.

**Table 4.6 Components of SDS-PAGE loading and running buffers, and staining solution**

| Solution | Reagent | Amount |
|---|---|---|
| Loading buffer (2 ×)[a] | Sodium dodecyl sulfate | 6.00 g |
| | Dithiothreitol | 4.65 g |
| | Tris HCl (1 M, pH 6.8) | 3.00 mL |
| | Bromophenol blue | 30.0 mg |
| | Glycerol | 23.5 mL |
| Running buffer (10 ×)[b] | Sodium dodecyl sulfate | 10.0 g |
| | Tris HCl | 30.0 g |
| | Glycine | 144 g |
| Staining solution[c] | Coomasie brilliant blue G250 | 400 mg |
| | Perchloric acid (70%) | 34 mL |
| | Ethanol | 50 mL |

a) Components were dissolved in $H_2O$ to give a total of 150 mL.
b) Components were dissolved in $H_2O$ to give a total of 1.00 L.
c) Components were dissolved in $H_2O$ to give a total of 900 mL.

### 4.12.2 Expression and purification of IPNS

*Recombinant A. nidulans IPNS (pJB703 construct) glycerol stock and protocols for expression and purification were obtained from Dr Jürgen Brem (University of Oxford).*

Starter cultures were grown in 2TY medium (100 mL, in 500 mL Erlenmeyer flasks), supplemented with chloramphenicol (25 μg / mL final concentration), and inoculated from a glycerol stock, at 37 °C for 16 h.

Large scale growths in 2TY medium (24 × 600 mL, in 2 L Erlenmeyer flasks), supplemented with chloramphenicol (25 μg · $mL^{-1}$ final concentration), were inoculated with starter culture (6 mL / flask), and incubated at 37 °C for 16 h. The cells were pelleted by centrifugation (8000 rpm, 10 minutes) at 4 °C, and stored at –80 °C.

The cells were thawed and suspended in lysis buffer (Tris 50 mM, pH 7.5, buffer A), supplemented with DNAse I (10 μg · $mL^{-1}$), lysozyme (1 mg · $mL^{-1}$), and phenylmehtylsulfonyl fluoride (10 μg · $mL^{-1}$), at 4 °C. The cells were then lysed by sonication at 4 °C (9.9 s sonication, 9.9 s break, 60% amplitude, 12 min), cell debris was removed by centrifugation (22000 rpm, 20 minutes) and the supernatant collected. After filtration through a sterile filter (0.45 μm), the supernatant was loaded onto an anion exchange column (Q Sepharose FF, 50 mL), preequilibrated with buffer A (3 column volumes). The protein was eluted with a gradient of buffer B (Tris 50 mM, pH 7.5; 1 M NaCl). Fractions containing IPNS were identified by UV

trace and their purity was checked by SDS-PAGE. The selected fractions were combined, concentrated, and further purified by size-exclusion chromatography (Superdex™ 75 column, 300 mL, GE Healthcare, preequilibrated with buffer C (Tris, 50 mM, pH 7.5; NaCl 200 mM, 1 column volume)). Fractions containing protein were identified by UV trace analysis and their purity checked by SDS-PAGE. The selected fractions were combined, concentrated, and further purified by anion column chromatography (Mono Q column), preequilibrated with buffer A (2 column volumes). Fractions containing protein were identified by UV trace and their purity checked by SDS-PAGE. The fractions containing pure IPNS were selected and concentrated. In order to obtain *apo*-IPNS, the enzyme was dialysed as described in section **4.10.3**. The protein was concentrated to 50 mg · $mL^{-1}$, aliquoted and stored at --80 °C.

### 4.12.3 Expression and purification of IPNS (hexahistidine affinity tag)

*A glycerol stock of pCOLD IPNS construct in Escherichia coli BL21(DE3) along with production, expression- and purification procedures were obtained from Dr Ilaria Pettinati. A cryo-glycerol stock of IPNS S55C construct in BL21(DE3) as well as expression and purification protocols were obtained from James Linyard (University of Oxford).*

Starter cultures were grown in 2TY medium (100 mL, in 500 mL Erlenmeyer flasks), supplemented with ampicillin (50 μg · $mL^{-1}$ final concentration), and inoculated from a glycerol freeze, at 37 °C for 16 hours.

Large scale growths in 2TY medium (24 × 600 mL, in 2 L Erlenmeyer flasks), supplemented with ampicillin (50 μg · $mL^{-1}$ final concentration), were inoculated with starter culture (1 : 1000 dilution), and incubated at 37 °C until the appropriate $OD_{600}$ (0.60) was obtained. The incubation temperature was reduced to 18 °C, and the cultures were induced isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM final concentration), and left incubating for an additional 16 h. The cells were pelleted by centrifugation (8000 rpm, 10 minutes) at 4 °C and transferred to a sealable plastic bag and stored at --80 °C.

The cells were thawed and resuspended in lysis buffer A (HEPES 25 mM, pH 7.5; NaCl 500 mM; imidazole 5 mM), supplemented with DNAse I (10 μg · $mL^{-1}$), lysozyme (1 mg · $mL^{-1}$), and phenylmethylsulfonyl fluoride (PMSF, 10 μg · $mL^{-1}$), at 4 °C. The cells were then subjected to sonication at 4 °C (9.9 s sonication, 9.9 s break, 60% amplitude, 12 min, Sonics Vibra-Cell), and debris was removed by centrifugation

(22000 rpm, 30 minutes) and the supernatant was collected. After filtration through a sterile filter (0.4 μm), the supernatant was loaded onto a HisTrap column (HisTrap HP GE Healthcare, 5 mL), preequilibrated with buffer A (5 column volumes). The protein was washed with buffer A (20 – 30 column volumes) and eluted with a gradient of buffer B (HEPES 25 mM, pH 7.5; NaCl 500 mM; imidazole 500 mM). Fractions containing protein were identified by UV trace and their purity checked by SDS-PAGE gel. The selected fractions were combined, concentrated, and purified by size-exclusion chromatography (Superdex™ 75 column, 300 mL, GE Healthcare, preequilibrated with buffer C (HEPES, 25 mM, pH 7.5; NaCl 500 mM, 1 column volumes)). Fractions containing protein were identified by UV trace and their purity checked by SDS-PAGE gel. The fractions containing pure protein (>95%) of interest were selected and concentrated. The protein was incubated with 3C protease (1 : 100 w/w%) over night at 4 °C and subsequently purified using a HisTrap column (HisTrap HP GE Healthcare, 5 mL), preequilibrated with buffer A. Highly purified of fractions were selected and concentrated. In order to obtain *apo*-IPNS, the enzyme was dialysed (Section **4.10.3**). The protein was concentrated to 50 mg · $mL^{-1}$ aliquoted and stored at --80 °C.

### 4.12.4 Fluorine labelling of IPNS S55C

*An adaptation of the protocol reported by Rydzik.*[43]

After purification (as described in Section **4.12.3**), IPNS S55C (350 μM) was incubated with tris-(2-carboxyetheyl)phosphine (TCEP, 1 equivalent), 3-bromo-1,1,1-trifluoroacetone (BTFA, 30 equivalents), in buffer (Tris 25 mM, pH 8.5; NaCl 100 mM), for 90 minutes at 4 °C. Alkylation was confirmed by denaturing MS and the enzyme was buffer exchanged (Tris 25 mM, pH 8.5; NaCl 100 mM) using a PD-10 desalting column (GE Healthcare). Subsequently, the enzyme was dialysed as described below (Section **4.12.5**).

### 4.12.5 Dialysis of IPNS

IPNS was transferred to a dialysis cassette (1 – 3 mL, 10K, Thermo Scientific, Slide-A-Lyzer®) and dialysed against buffer A (1 : 250 sample : solution ratio, Tris 25 mM, pH 8.0; EDTA 50 mM; 1,10-phenantroline 10 mM; NaCl 100 mM, 3 × 3 hours), followed by dialysis against buffer B (1 : 250 sample : solution ratio,

Tris 25 mM, pH 8.0; NaCl 100 mM, 1 × 12 h and 2 × 3 h). The protein was concentrated to 50 mg · mL$^{-1}$, aliquoted and stored at --80 °C. Overall yield was 8 mg · L$^{-1}$ was obtained.

### 4.12.6 Trypsin digestion

*According to a protocol provided by Elisabete Pires (University of Oxford).*

Labelled mutant IPNS* (15 μL, 1 mg / mL) was reduced with dithiothreitol (DTT, 2 μL, 85 mM), in ammonium bicarbonate (10 mM), for 40 minutes at 56 °C. Subsequently, the protein was alkylated with iodoacetamide (7 μL, 55 mM), for 30 minutes in the dark, at rt. The solution was treated with DTT (3 μL, 85 mM), for 10 minutes in the dark, at rt. Trypsin (3 μL, 1 : 20 w/w) was added and allowed to react for 16 h, at 37 °C. The sample was subsequently diluted with acetonitrile (120 μL) and trypsin (1 μL, 1 : 100 w/w) was added, and the resulting solution was allowed to react for 3 h, at 37 °C. Digestion was quenched with the addition of formic acid (7.5 μL), vacuum centrifuged to dryness and redissolved water (18 μL) supplemented with formic acid (0.1 % V/V). The resulting solution was aspirated into a ZipTip column, equilibrated with water / acetonitrile (98 : 2, 0.1% formic acid). The fragments were eluted using water / acetonitrile (40 : 60, 0.1% formic acid), vacuum centrifuged to dryness and redissolved in water / acetonitrile (98 : 2, 0.1% formic acid), and measured by nanoLC.

### 4.12.7 Circular dichroism experiments

Enzyme samples were buffer exchanged into phosphate buffer (50 mM, pH 8.0, to avoid halide ions), diluted (0.1 mg · mL$^{-1}$ final concentration) and transferred into a cuvette (0.1 cm path length). CD spectra were recorded using a Chirascan circular dichroism spectrometer (Applied Photophysics), equipped with a temperature-controlled cuvette holder, at 23 °C. Spectra were recorded form 180 – 300 nm (0.5 nm intervals) and performed in triplicate, giving an averaged final result.

## 4.13 Non-denaturing mass spectrometry

IPNS was buffer exchanged into ammonium acetate (1 M) using a Bio-Spin column (Bio-Rad) two times, followed by an overnight dialysis with ammonium acetate (1 M) and diluted to 15 – 20 μM (concentration

verified by Nanodrop). Non-denaturing mass spectrometry data were acquired using an Orbitrap Extended Mass Range (EMR) prototype[73]. All data were recorded in positive ion mode, from a static nanospray source (nsp), using a gold-plated capillary prepared in house. Nitrogen was used as a collision gas (the pressure of around $1 \cdot 10^{-9}$ mbar recorder in the orbitrap). Capillary temperature was set to 30 °C, and a spray voltage of 1.4 kV. The nitrogen was used as a collision gas with no HCD voltage (0 V) applied. The acquired data was processed and analysed using Thermo Xcalibur 4.13.

## 4.14 Nuclear Magnetic Resonance experiments

Experiments with $^{19}F$ labelled IPNS* (IPNS S55C BTFA) were recorded using a Bruker AVIII HD 600 equipped with a BB-F/H $N_2$ CryoProbe, in 3 mm MATCH NMR tubes (Cortectnet) 5 mm regular or Young valve NMR tube (Norell), at 298 K, unless stated otherwise. $^{1}H$ NMR data were referenced to the respective solvent signals. $^{19}F$ NMR data were processed with 3 Hz Lorentzian line broadening and referenced to 2,2,2-trifluoro acetic acid (100 μM, at $\delta_F$ --76.55 ppm). Data were processed with MestReNova 14.1 and Bruker TOPSPIN 3.6.1.

### 4.14.1 Sample preparation for NMR spectroscopy

Samples were prepared under anaerobic conditions (<0.2 ppm $O_2$) in a glovebox (Belle Technology). Large quantities of solutions (>2 mL) were deoxygenated by purging with argon (30 minutes). Solid materials ($FeSO_4$, ACV) and NMR tubes were left for 16 h inside the glovebox to further deoxygenate. Samples of *apo*-IPNS* were transferred to the glovebox without purging immediately prior to use for experiments. Appropriate stock solutions of $FeSO_4$ (100 mM), ACV (10 mM), and TFA (10 mM) were prepared. The samples were prepared in Tris buffer (50 mM, pH 7.5). Aliquots of *apo*-IPNS* (50 mg/mL, 1.35 mM, 50 μL) were used. Total sample volume was 450 μL and contained 10% (v/v) $D_2O$. Samples were transferred into a 5 mm Young valve NMR tube and recorded.

# 4.15 Crystallography

## 4.15.1 General crystallisation method

Samples for IPNS crystallisation were prepared under anaerobic conditions (<0.2 ppm $O_2$) in a glovebox (Belle Technology). Solutions (>2 mL) were deoxygenated by purging with argon (30 minutes). Solid materials ($FeSO_4$, **ACV** / substrate analogues), and 96-well plates (Cryschem) were left 16 h inside the glovebox to further deoxygenate. Samples of IPNS (50 mg · $mL^{-1}$) were transferred to the glovebox and purged, immediately prior to use for experiments.

## 4.15.2 Batch crystallisation of IPNS microcrystals

*Protocol obtained from Dr Patrick Rabe (University of Oxford)*

Crystallisation of IPNS in batch form was done using a batch plate (96 wells, 100 μL per well). A typical batch, consisting of ~30 wells (~3 mL, total) was prepared as follows. To a solution of IPNS (200 μL, 50 mg · $mL^{-1}$) was added $FeSO_4$ (2.67 μL, 100 mM, 1.00 eq.), followed by **ACV** (2.0 mg, in Tris-buffer 40 μL, 25 mM, pH 8.5). Precipitant solution was prepared using $Li_2SO_4$ (1.995 mL, 2.7 M), $H_2O$ (405 μL), and Tris (300 μL, 1 M, pH 8.5). Enzyme solution was added to the precipitant solution and supplemented with concentrated seed stocks (10 μL, prepared from IPNS • Fe • ACV crystals). The resulting solution was immediately dispensed over the 96-well plate (100 μL per well), sealed using a clear seal (Starlab), and shaken using a plate shaker (750 rpm, SciQuip). Typical $Li_2SO_4$ optimisation was done to improve crystal density and sample quality.

Samples were harvested after 48 h, combined, and the crystals were allowed to settle. The batch was concentrated (threefold, ~1 mL) by removing supernatant. The sample was transferred to an anaerobic Hamilton syringe (1 mL) equipped with a gas lock. The anaerobic syringe was transported in an anaerobic sandwich box from the glovebox to the beamline.

### 4.15.3 Seed stock preparation

Stocks were prepared from dense crystal slurries (Section **4.15.2**) or from single crystals grown according to the reported procedure[3,4]. Suspensions containing crystals (100 μL) were transferred to seed bead kit (plastic, Hampton research) cooled to 4 °C. The solution was vortexed (30 s) and then cooled to 4 °C (30 s), the steps were repeated 10 times. The seed stocks were then aliquoted (10 μL) and stored at –80 °C.

### 4.15.4 Data collection

Synchrotron crystallographic data sets were collected using MX beamlines at Diamond Light Source (Didcot, UK). XFEL crystallographic data sets were collected using the Linac Coherent Light Source (LCLS, California, United States) or SPring-8 Angstrom Compact free electron LAser (SACLA, Japan). Phases were solved by molecular replacement using Phaser/CCP4 (IPNS accession code 1BLZ)[33]. Refinements were done using Phenix,[74] and iterative fitting using Coot[75] software packages. RMSD values were determined using PyMol 2.3.

## 4.16 References


1. Roach, P. L., Schofield, C. J., Baldwin, J. E., Clifton, I. J. & Hajdu, J. Crystallization and Preliminary-X-Ray Diffraction Studies on Recombinant Isopenicillin-N Synthase from Aspergillus-Nidulans. *Protein Sci.* **4**, 1007–1009 (1995).
2. Roach, P. L. *et al.* Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* vol. 375 700–704 (1995).
3. Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–30 (1997).
4. Roach, P. L. *et al.* Anaerobic crystallisation of an isopenicillin N synthase.Fe(II).substrate complex demonstrated by X-ray studies. *Eur. J. Biochem.* **242**, 736–740 (1996).
5. Burzlaff, N. I. *et al.* The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401**, 721–724 (1999).
6. Hamed, R. B. *et al.* The enzymes of β-lactam biosynthesis. *Nat. Prod. Rep.* **30**, 21–107 (2013).
7. Rabe, P., Kamps, J. J. A. G., Schofield, C. J. & Lohans, C. T. Roles of 2-oxoglutarate oxygenases and isopenicillin N synthase in β-lactam biosynthesis. *Nat. Prod. Rep.* **35**, 735–756 (2018).
8. Tamanaha, E. *et al.* Spectroscopic evidence for the two C-H-cleaving intermediates of Aspergillus nidulans isopenicillin N synthase. 1–32.
9. Baldwin, J. E., Blackburn, J. M., Sutherland, J. D. & Wright, M. C. High-level soluble expression of isopenicillin N synthase isozymes in E. coli. *Tetrahedron* **47**, 5991–6002 (1991).
10. Tamanaha, E. *et al.* Spectroscopic Evidence for the Two C–H-Cleaving Intermediates of Aspergillus nidulans Isopenicillin N Synthase. *J. Am. Chem. Soc.* **138**, 8862–8874 (2016).
11. Baldwin, J. E., Adlington, R. M., Moroney, S. E., Field, L. D. & Ting, H. Stepwise Ring Closure in Penicillin Biosynthesis. Initial β-Lactam Formation. *J. Chem. Soc. Chem. Commun.* 984–986 (1984).
12. Baldwin, J. E. *et al.* Synthesis of δ-(L-α-Aminoadipoyl)-L-cysteinyh-valine t and some Carbon-13 and Nitrogen-15 Labelled lsotopomers. *J. Chem. Soc. Perkin Trans. 1* 2253–2257 (1981).
13. Fawcett, B. P. A. *et al.* Synthesis of delta-(alpha-Aminoadipyl ) cysteinylvallne and its Role in Penicillin Biosynthesis. *Biochem. J.* **157**, 651–660 (1976).
14. Thomson, G. A., Scott, A. I. & Baxter, R. L. Synthesis of 6-(L-α-Aminoadipyl)-L-cysteinyl-D-valine and δ-(L-α-Aminoadipyl)-L-cysteinyl-D-Valylglycine. *J. Chem. Soc. Perkin Translocat. 1* 941–944 (1983).
15. Baldwin, J. E., Bradley, M., Adlington, R. M., Norris, W. J. & Turner, N. J. Further evidence for the involvement of a monocyclic β-lactam in the enzymatic conversion of δ-L-α-aminoadipoyl-L-cysteinyl-D-valine into isopenicillin N. *Tetrahedron* **47**, 457–480 (1991).
16. Wolfe, S. & Jokinen, M. G. Total synthesis of 6-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV), a biosynthetic precursor of penicillins and cephalosporins. *Can. J. Chem.* **57**, 1388–1396 (1979).
17. Bradley, M., Adlington, R. M. & Turner, N. J. Further evidence for the involvement of a monocyclic β-lactam in the enzymatic conversion of δ-L-α-aminoadipoyl-L-cysteinyl-D-valine into isopenicillin N. *Tetrahedron* **47**, 457–480 (1991).
18. Miller, J. B. Preparation of Crystalline Diphenyldiazomethane. *J. Org. Chem.* **24**, 560–561 (1958).
19. Squitieri, R. A., Shearn-nance, G. P., Hein, J. E. & Shaw, J. T. Synthesis of Esters by in Situ Formation and Trapping of Diazoalkanes. 1–7 (2016) doi:10.1021/acs.joc.6b00408.
20. Behrendt, R., White, P. & Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **22**, 4–27 (2016).
21. Subirós-Funosas, R., Prohens, R., Barbas, R., El-Faham, A. & Albericio, F. Oxyma: An efficient additive for peptide synthesis to replace the benzotriazole-based HOBt and HOAt with a lower risk of explosion. *Chem. - A Eur. J.* **15**, 9394–9403 (2009).
22. El-Faham, A. & Albericio, F. Peptide coupling reagents, more than a letter soup. *Chem. Rev.* **111**, 6557–6602 (2011).
23. Stathopoulos, P., Papas, S., Pappas, C. & Mousis, V. Side reactions in the SPPS of Cys-containing peptides. *Amino Acids* **44**, 1357–1363 (2013).
24. Baldwin, J. E. & Wan, T. S. Penicillin biosynthesis: Retention of configuration at C-3 of valine during its incorporation into the arnstein tripeptide. *Tetrahedron* **37**, 1589–1595 (1981).
25. Fizet, C. Synthese de la dimethyl-4,4-dihydrofuranedione-2,3. *Helv. Chim. Acta* **65**, 2024–2028 (1982).
26. Chenault, H. K., Dahmer, J. & Whitesides, G. M. Kinetic Resolution of Unnatural and Rarely Occurring Amino Acids: Enantioselective Hydrolysis of A-Acyl Amino Acids Catalyzed by Acylase I. *J. Am. Chem. Soc.* **111**, 6354–6364 (1989).
27. Schmuck, C., Heil, M., Scheiber, J. & Baumann, K. Charge interactions do the job: A combined statistical and combinatorial approach to finding artificial receptors for binding tetrapeptides in water. *Angew. Chemie - Int. Ed.* **44**, 7208–7212 (2005).
28. Tarhonskaya, H. *et al.* Studies on Deacetoxycephalosporin C Synthase Support a Consensus Mechanism for 2 - Oxoglutarate Dependent Oxygenases. *Biochemistry* **53**, 2483–2493 (2014).

29. Josephine, H. R., Kumar, I. & Pratt, R. F. The Perfect Penicillin ? Inhibition of a Bacterial DD-Peptidase by. *J. Am. Chem. Soc.* **126**, 8122–8123 (2004).
30. Baldwin, J. E., Herchent, S. R. & Singh, P. D. Syntheses of Penicillin N, [6α-3H]Penicillin N and [10-14C,6α-3H]Penicillin N. *Biochem. J.* **186**, 881–887 (1980).
31. Huffman, G. W. Improved preparation of penicillin N and isopenicillin N. *Org. Prep. Proced. Int.* **20**, 497–503 (1988).
32. Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat. Methods* **14**, 443–449 (2017).
33. Byford, M. F., Baldwin, J. E., Shiau, C.-Y. & Schofield, C. J. The Mechanism of ACV Synthetase. *Chem. Rev.* **97**, 2631–2650 (1997).
34. Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–830 (1997).
35. Nogami, H. & Kato, Y. Diffusion Coefficient of Oxygen in Propylene Glycol-Water Mixtures and Glycerol-Water Mixtures. *YAKUGAKU ZASSHI* **82**, 120–126 (1962).
36. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
37. Kantardjieff, K. A. & Rupp, B. Proceedings of 10th International Conference on Formal Methods in Computer-Aided Design, FMCAD 2010, Lugano, Switzerland, October 20-23. *Protein Sci.* **12**, 1865–1871 (2003).
38. Weichenberger, C. X. & Rupp, B. Ten years of probabilistic estimates of biocrystal solvent content: New insights via nonparametric kernel density estimate. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1579–1588 (2014).
39. Dalvit, C. & Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **62**, 2218–2244 (2019).
40. Dalvit, C. Ligand- and substrate-based 19F NMR screening: Principles and applications to drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **51**, 243–271 (2007).
41. Arntson, K. E. & Pomerantz, W. C. K. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. *J. Med. Chem.* **59**, 5158–5171 (2016).
42. Gee, C. T. *et al.* Protein-observed 19F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat. Protoc.* **11**, 1414–1427 (2016).
43. Rydzik, A. M. *et al.* Monitoring Conformational Changes in the NDM-1 Metallo-β-lactamase by 19 F NMR Spectroscopy. *Angew. Chemie Int. Ed.* **53**, 3129–3133 (2014).
44. Lesniak, R. *et al.* 19F NMR studies on γ-Butyrobetaine Hydroxylase provide mechanistic insights and suggest a dual inhibition mode. *Chem. Commun.* **55**, (2019).
45. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science (80-. ).* **274**, 1531–1534 (1996).
46. Wider, G. NMR techniques used with very large biological macromolecules in solution. *Methods Enzymol.* **394**, 382–398 (2005).
47. Chen, H., Viel, S., Ziarelli, F. & Peng, L. F NMR: A valuable tool for studying biological events. *Chem. Soc. Rev.* **42**, 7971–7982 (2013).
48. Gee, C. T., Koleski, E. J. & Pomerantz, W. C. K. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed 19 F NMR Spectroscopy . *Angew. Chemie* **127**, 3806–3810 (2015).
49. Rastinejad, F., Evilia, C. & Lu, P. Studies of nucleic acids and their protein interactions by 19F NMR. *Methods Enzymol.* **261**, 560–575 (1995).
50. Bann, J. G., Pinkner, J., Hultgren, S. J. & Frieden, C. Real-time and equilibrium 19F-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 709–714 (2002).
51. Hoeltzli, S. D. & Frieden, C. Refolding of [6-19F]tryptophan-labeled Escherichia coli dihydrofolate reductase in the presence of ligand: A stopped-flow NMR spectroscopy study. *Biochemistry* **37**, 387–398 (1998).
52. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The Preparation of 19F-Labeled Proteins for NMR Studies. *Methods Enzymol.* **380**, 400–415 (2004).
53. Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of 19F NMR to studies of protein structure and dynamics. *Prog. Nucl. Magn. Reson. Spectrosc.* **62**, 1–33 (2012).
54. Kim, H.-W., Perez, J. A., Ferguson, S. J. & Campbell, I. D. The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. *FEBS Lett.* **272**, 34–36 (1990).
55. Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of 19F-Tryptophan for protein NMR spectroscopy. *Chem. Commun.* **48**, 10681–10683 (2012).
56. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science (80-. ).* **292**, 498–500 (2001).

57. Jackson, J. C., Hammill, J. T. & Mehl, R. A. Site-specific incorporation of a 19F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **129**, 1160–1166 (2007).
58. Brown, W. E. & Seamon, K. B. Quantitation and characterization of the trifluoroacetonyl derivative of cysteine: A useful NMR probe. *Anal. Biochem.* **87**, 211–222 (1978).
59. Johnston, H. L. & Giauque, W. F. The heat capacity of nitric oxide from 14 °K. To the boiling point and the heat of vaporization. Vapor pressures of solid and liquid phases. The entropy from spectroscopic data. *J. Am. Chem. Soc.* **51**, 3194–3214 (1929).
60. Chen, V. J. *et al.* Spectroscopic studies of isopenicillin N synthase. A mononuclear nonheme Fe2+ oxidase with metal coordination sites for small molecules and substrate. *J. Biol. Chem.* **264**, 21677–21681 (1989).
61. Orville, a M. *et al.* Thiolate ligation of the active site Fe2+ of isopenicillin N synthase derives from substrate rather than endogenous cysteine: spectroscopic studies of site-specific Cys --> Ser mutated enzymes. *Biochemistry* **31**, 4602–4612 (1992).
62. Martinie, R. J. *et al.* Vanadyl as a Stable Structural Mimic of Reactive Ferryl Intermediates in Mononuclear Nonheme-Iron Enzymes. *Inorg. Chem.* **56**, 13382–13389 (2017).
63. Mitchell, A. J. *et al.* Visualizing the reaction cycle in an Iron(II)- and 2-(Oxo)-glutarate-dependent hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
64. Dunham, N. P. *et al.* Two Distinct Mechanisms for C–C Desaturation by Iron(II)- and 2-(Oxo)glutarate-Dependent Oxygenases: Importance of α-Heteroatom Assistance. *J. Am. Chem. Soc.* **140**, 7116–7126 (2018).
65. Moffat, K. Time-Resolved Crystallography. *Acta Crystallogr. Sect. A Found. Crystallogr.* **54**, 833–841 (1998).
66. De Cat, I. *et al.* Controlling the position of functional groups at the liquid/solid interface: Impact of molecular symmetry and chirality. *J. Am. Chem. Soc.* **133**, 20942–20950 (2011).
67. LONG, A. J. *et al.* Structural studies on the reaction of isopenicillin N synthase with the substrate analogue delta-(l-alpha-aminoadipoyl)-l-cysteinyl-d-alpha-aminobutyrate. *Biochem. J.* **372**, 687–693 (2003).
68. Baldwin, J. E. *et al.* Cephalosporin biosynthesis: A branched pathway sensitive to an isotope effect. *Tetrahedron* **47**, 9881–9900 (1991).
69. Baldwin, J. E., Adlington, M., Bird, W., Schofield, J. & Field, A. Studies on the Exchange of Valine-Oxygen During The Biosynthesis of 6-(L-α-Aminoadipoyl)-L-cysteinyl-D-valine. *Tetrahedron* **48**, 1099–1108 (1992).
70. Lebarillier, L., Outurquin, F. & Paulmier, C. Synthesis and reactivity of α and β-chloro-α-phenylselanyl esters. *Tetrahedron* **56**, 7495–7502 (2000).
71. Arakawa, K., Cao, Z., Suzuki, N. & Kinashi, H. Isolation, structural elucidation, and biosynthesis of 15-norlankamycin derivatives produced by a type-II thioesterase disruptant of Streptomyces rochei. *Tetrahedron* **67**, 5199–5205 (2011).
72. Pedersen, S. L., Tofteng, a. P., Malik, L. & Jensen, K. J. Microwave heating in solid-phase peptide synthesis. *Chem. Soc. Rev.* **41**, 1826 (2012).
73. Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. R. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* **9**, 1084–1086 (2012).
74. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
75. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

# Chapter 5

## Summary

## 5.1 Crystallisation of OXA-10

The work described in this thesis concerning oxacillinase-10 (OXA-10, Chapter 3) describes its expression, purification and crystallisation for use in time resolved crystallographic studies. During broad screening, a variety of crystallisation conditions were found for OXA-10. However, it proved challenging to establish suitable conditions that give high-resolution diffraction (<2 Å) and which avoid the use of buffer compounds / ions that could inhibit / bind to the active site (HEPES, MOPS, or halide ions at high concentrations). The initial conditions found for the crystallisation of OXA-10 (referred to as the E8 conditions) were optimised to produce microcrystal slurries. Serial synchrotron and X-ray free electron laser (XFEL) crystallography experiments at diamond light source (DLS, Didcot) and Linac Coherent Light Source (LCLS, California United States), using a viscous extruder, fixed target, or drop-on-demand (DOD) sample delivery system, consistently showed that multiple lattices were being hit at a range of concentrations and with samples from different batches. This outcome could be a result of crystal growth artefacts such as twinning or polymorphism. The problems with the crystal sample were believed to be inherent to the conditions used and led to the development of different crystallisation conditions for the production of OXA-10 microcrystals.

Fixed target experiments with microcrystal slurries of OXA-10 obtained using the A8, H6, and C4 conditions showed reasonable diffraction (2.2 – 2.3 Å) with a reasonable indexed hit rate (17 – 30%). The A8 conditions appeared to give rise to non-isomorphous samples, which were considered less suitable for future mixing experiments. A room temperature serial crystallographic structure of OXA-10 was obtained using the H6 conditions. This structure was very similar to reported OXA-10 structures obtained from macrocrystals under cryo-cooled conditions (RMSD of 0.29 Å, PDB accession code 5MMY)[1]. Several conditions, i. e. A8, C4, H6, G4, 144D, and 4 °C, for the production of OXA-10 microcrystals were carried forward for XFEL experiments. Of these conditions, some (H6, G4, and 4 °C) gave rise to difficulties during production, lacking sufficient quantities of sample to acquire data during the beamtime. Eventually the A8 crystal form was tested during the serial femto second crystallography (SFX) XFEL method at SPring-8 Angstrom Compact free electron Laser (SACLA, Hyogo, Japan). However, somewhat surprisingly, the sample suffered from low resolution diffraction (<2.8 Å) and a low number of indexed images (~1%). OXA-10 microcrystals obtained using the 144D conditions similarly suffered from a low

number of indexed images (~1%). Incubating 144D OXA-10 microcrystals with ertapenem resulted in a change in unit cell parameters, which in previous SFX experiments had been associated with binding of a ligand / reaction *in crystallo*.

Overall the results have demonstrated that it is possible to use OXA-10 for serial crystallography work. The initial results on incubation of OXA-10 with ertapenem look promising. Future experimental efforts should focus on improving the robustness of crystallisation conditions and on establishing kinetics of the acylation reaction *in crystallo*. Future crystallisation conditions suitable for time resolved work might be restricted as solution state NMR studies demonstrated that diffusion of possible ligands (i. e. ertapenem, imipenem, ceftazidime, and clavulanic acid) are significantly affected by compound used for particular crystallisation conditions (e. g. PEG). A major challenge was encountered with the E8 OXA-10 microcrystals. Although diffraction of the microcrystal slurries was observed, a low number of indexed images was obtained, precluding structural determination. It is possible that the more accessible fixed target experiments performed at synchrotrons would be better suited to investigate the performance of the different microcrystal slurries, so guiding choice of crystal form for future time resolved work.

## 5.2 Time resolved crystallography on IPNS

The work done on IPNS (Chapter 3) describes the production, purification, and crystallisation of IPNS with L-δ-(aminoadipoyl)-L-cysteinyl D-valine (**ACV**), and its deuterated analogues L-δ-(α-aminoadipoyl)-L-2,2-[$^2H_2$]-cysteinyl D-valine (**AdCV**), and L-δ-(α-aminoadipoyl)-L-cysteinyl D-[$^2$H]-valine (**ACdV**). IPNS was successfully co-crystallised with **ACV** to give IPNS • Fe • ACV microcrystal slurries under anaerobic conditions and used to obtain a room temperature serial structure. The room temperature structure showed high similarity with the reported structure obtained from macrocrystals under cryo-cooled conditions. IPNS • Fe • ACV microcrystals were successfully incubated with $O_2$ to initiate the enzymatic reaction. Extra electron density was observed at the active site suggesting that $O_2$ binds to the Fe centre, which is in agreement with the proposed mechanism. This was further supported by Kα X-ray emission spectroscopy (XES) on Fe collected simultaneously with diffraction data,

which demonstrated a change in oxidation state from Fe(II) to Fe(III). A previously unreported alternative conformation for α-helix 3 (residues 49 – 65) and β-strand 13 (residues 280 – 282) was observed.

This alternative conformation for α-helix 3 and β-strand 13 was further investigated in solution (Chapter 4) using $^{19}$F NMR spectroscopy. The IPNS S55C variant was expressed, purified, and labelled with 1-bromo 3,3,3-trifluoro acetone (BTFA), a sulfhydryl selective alkylating reagent.[2,3] BTFA labelled IPNS S55C (IPNS*) was used in $^{19}$F NMR studies and provided supporting evidence for the alternative conformation of α-helix 3, using nitric oxide (NO) as an $O_2$ surrogate. The solution state $^{19}$F NMR observations were consistent with a mechanism in which binding of NO (and by extrapolation $O_2$) results in a change in conformation of the **ACV** substrate, altering the interaction with residues on β-strand 13 which in turn changes interactions with α-helix 3. In addition, $^{19}$F NMR studies, in combination with non-denaturing mass spectrometry (MS) studies, were used to investigate the binding of Fe to the IPNS active site in presence of isopenicillin N (**IPN**). The results demonstrate that the loss of Fe at the IPNS active site observed in IPNS • Fe • IPN crystals does not occur in solution. The combined studies on IPNS demonstrate that time resolved SFX on IPNS • Fe • ACV microcrystals, supported by complementary techniques (e. g. NMR spectroscopy and MS) is a feasible strategy to investigate the dynamics during catalysis by IPNS.

## 5.3 Future perspective

The crystallographic studies described in this thesis focused on the use of tr-SFX to investigate catalysis by OXA-10 and IPNS. The experiments were designed to leverage aspects of XFEL capability, e. g. the use of small crystals to reduce the diffusion time of small molecules into the crystal lattice for OXA-10, and (in the future) to photolabile intermediates (e. g. Fe(IV) oxo) in the IPNS catalytic cycle. For experiments planned with IPNS the DOD sample delivery system provides the opportunity to collect diffraction data and XES data simultaneously, facilitating the interpretation of electron density maps. The tr-SFX experiments for OXA-10 and IPNS were performed in conjunction with synchrotron based experiments which provided supporting data. Given the experience gained in the studies described in this thesis, it is

proposed that a combination of both synchrotron cryo and room temperature fixed target methods and XFEL based experiments should be pursued as a matter of course.

## 5.4 References

1. Lohans, C. T. *et al.* 13C-Carbamylation as a mechanistic probe for the inhibition of class D β-lactamases by avibactam and halide ions. *Org. Biomol. Chem.* **15**, 6024–6032 (2017).

2. Brown, W. E. & Seamon, K. B. Quantitation and characterization of the trifluoroacetonyl derivative of cysteine: A useful NMR probe. *Anal. Biochem.* **87**, 211–222 (1978).

3. Rydzik, A. M. *et al.* Monitoring Conformational Changes in the NDM-1 Metallo- b-lactamase by 19 F NMR Spectroscopy. *Angew. Chemie* **126**, 3193–3197 (2014).

# Part B

# Chapter 6

## Formaldehyde chemistry in biology

## 6.1 Formaldehyde as a chemical

Formaldehyde (HCHO) was first described by the Russian scientist Alexander Michailowitsch Butlerow in 1855 and is chemically the simplest aldehyde.[1] The preparation of HCHO by dehydration of methanol was described by German chemist August Wilhelm von Hofmann in 1867.[2] HCHO is now a widely used chemical with a vast number of applications ranging from industrial to consumer products; the estimated global production of HCHO is 20 million tons per year.[3] HCHO is found as a by-product in cigarette smoke, fires, and car exhausts. Furniture, carpet, and plywood also contain small amounts of HCHO. HCHO has a long history as a fixative used in pathology and embalming. HCHO is also proposed as one of the first chemicals present on this planet, and as such is suspected to be involved in the origin of life.[1,4] HCHO continues to play an important role in everyday life. It is also produced in many biochemical processes, in health and disease. What follows below is an overview of HCHO exposure sources and the impact of HCHO on health. Finally, the chemistry of HCHO in biological settings is discussed.

## 6.2 Exogenous formaldehyde sources

### 6.2.1 Formaldehyde from consumption and inhalation

In daily life, people are exposed to HCHO from a variety of sources. Such exposure occurs through the digestive system, absorption through skin, and inhalation of air. Levels of HCHO near the ground in coastal, ocean, and mountain areas range from 0.05 – 14 $\mu g \cdot m^{-3}$.[1,5] On average the HCHO exposure in large cities is higher than rural areas.[6] Sources for the production of HCHO include (forest) fires,[7] industry, power plants,[1,6] and transport exhaust[8].[1] HCHO levels vary depending on the season, humidity, and the age of surrounding buildings.[9] Indoor levels of HCHO are generally higher than outdoor levels. The estimated indoor levels of HCHO are between 10 – 90 $\mu g \cdot m^{-3}$ (mean reported level ~30 $\mu g \cdot m^{-3}$)[10–12] and on average are considered to be higher in public places (e. g. libraries / shopping centres).[12] Indoor sources of HCHO include wood products,[13] furniture, carpets, cooking fumes,[14] and cigarette smoke[15] (including e-cigarettes[16,17]).[1,5] HCHO is also present in food and food products including fish, meat, vegetables, fruits, and dairy products. The average intake of HCHO from these sources is estimated around 1.5 – 14 mg per day for adults.[5] Consumption of methanol, either as a compound or through hydrolysable esters, indirectly increases the level of HCHO as a metabolic precursor to HCHO.[18] Similarly, dealkylation of compounds

in food can increase the levels of HCHO.[5] Occupational exposure may result in a higher HCHO exposure dose (Section **6.2.3**).

### 6.2.2 Formaldehyde in pharmaceuticals

The concept of prodrugs involves the reversible derivatization of a parent drug to overcome barriers that otherwise limit drug availability / delivery. The parent drug is functionalised to improve particular physical aspects (e. g. solubility), in order to promote permeation of biological membranes. Drugs that are functionalised as a prodrug often contain a "linker" group that connects the parent drug to an appropriately selected moiety. Some drugs, including Cerebyx®, Viread ®, and Spectracef®, contain an oxymethyl linker group (Figure 6.1).[5] In some cases, the additional moiety can be removed inside humans. Cerebyx® contains an oxymethyl linker attached to a phosphate to improve its solubility. Cerebyx® is dephosphorylated in presence of alkaline phosphatase, producing a hydroxymethyl adduct of phenytoin. At physiological pH, the hydroxymethyl adduct is liberated forming phenytoin and HCHO (Figure 6.1). In such a mechanism, stoichiometric amounts of HCHO with respect to the drug molecule, are generated. Given the stoichiometry of the amount of HCHO released and the amount of drug prescribed, it is possible to calculate the amount of HCHO released at a given moment. However, not all drugs are administered on a daily basis, which leads to a lower impact of the HCHO that is released. Even when these drugs are administered on a daily basis, for instance Cereyx®, the amount of HCHO released is a fraction of the total daily doses of HCHO administered.[5]

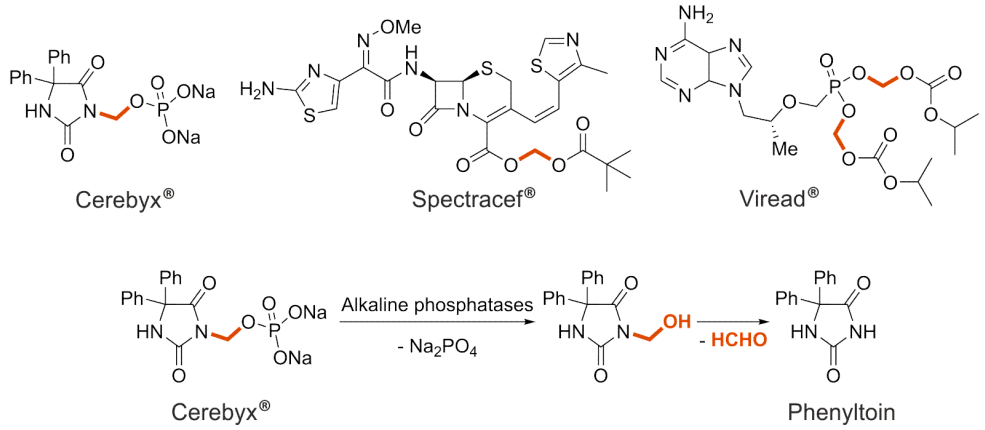



**Figure 6.1:** Structures and mechanism of exemplary HCHO releasing prodrugs. The HCHO producing structural features are highlighted in orange.[5]

Toxin derived vaccines are used to protect against microbial infections. For the preparation of vaccines, the toxins must be inactivated. Treatment of a toxin with HCHO is a common method for this and was introduced about a century ago.[19,20] An important factor in the quality of the vaccine is the detoxifying process; the HCHO concentration, temperature, and composition of the matrix used in the preparation (typically containing Gly or Lys) are all relevant.[21] During the preparation of vaccines, the concentration of HCHO is diluted, and the vaccine tested for quality control.[22] As with prodrugs, the amount of HCHO that is administered *via* vaccines is a fraction of the total daily amount. Vaccines are administered infrequently, typically as a single dose, and thus chronical exposure to HCHO from vaccines is considered to be of limited concern. Although HCHO mediated detoxification is an important step in the preparation of toxoid based vaccines, the nature of the modifications introduced by HCHO and their locations within microbes have not been defined.[23]

### 6.2.3 Occupational and consumer exposure to formaldehyde

HCHO sensitisation as a result of occupational exposure is well-established, and includes medical personnel, biological researchers, goods manufacturers (including leather, clothes, furniture, paper), hairdressers, embalmers, and metalworkers.[24] For example, a hair straightening procedure called "progressive blow-dry" relies on applying HCHO containing products (3 – 11% (v/v) HCHO solutions) with subsequent heating of the HCHO derived products.[25] Consequently, the levels of HCHO in the working environment frequently exceed the short and long term exposure limits. Sensitisation to HCHO of workers in industry where HCHO is frequently used is widely reported.[24,26,27]

Water based cosmetic products including nail polish / hardener,[28] and skincare products,[29] and detergents are often supplemented with HCHO.[30–32] Water based cosmetic products, as well as detergents have in the past commonly been supplemented with HCHO as preservative.[24,33] During the 1940s and 1950s the use of HCHO as preservative increased because it was effective and cheap. Due to an increase in sensitisation to HCHO, most products currently contain HCHO releasers as a substitute.[30,34] Common examples of HCHO releasing formulations[30] include: diazolidinylurea,[35] quaternium-15,[36] 1,3-dimethylol-5,5-dimethylhydantoin (DMDM hydantoin),[37] and bronopol[38] (Figure 6.2). In contact with water these

compounds release HCHO. Contact dermatosis can also be observed with the HCHO releasing compounds, suggesting sensitisation to HCHO.[39] Consumer sensitisation to HCHO was initially observed for clothing.[24] Resins containing HCHO were introduced in the 1920s to reduce wrinkling of clothing.[40] During the 1950s and 1960s, an increase in contact allergy was reported, whereas the 1970s saw fewer cases.[41] This change is correlates with to the use of different textile finishes, which relied less on the use of HCHO. Determination of a safe HCHO concentration is difficult due to many false positives and negatives in the patch test, which is the standard test method used for HCHO sensitisation.[30,42]

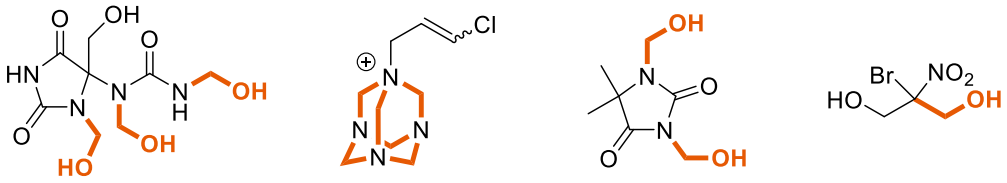



**Figure 6.2:** Structures of commonly used HCHO releasers which can substitute for HCHO as preservative. The HCHO producing structural features are highlighted in orange.[30]

Wood based furniture and constructions often rely on HCHO containing adhesives.[13] Most commonly used adhesives are urea-HCHO resins, due to their low costs, rapid curing, bond strength, and colourless glue line. Chemical adhesives were introduced in the early 20th century to replace bio-based adhesives.[1] Cost effectiveness and stability under humid conditions were more favourable for the chemical adhesive.[13] Depending on the product requirements, the adhesive can be modified with additives, such as phenol, melamine, or resorcinol, to improve its properties. Reduction of the use of HCHO in wood based composites was achieved in the 1960s, and intensified in the 1980s with the discovery of the carcinogenic properties of HCHO in animal models.[1]

## 6.3 Endogenous formaldehyde sources

### 6.3.1 Formaldehyde inside humans

It is suggested that HCHO is present in many human body fluids including urine, blood, intercellular tissue, and within cells.[5] Several methods using a combination of $^{14}C$ labelled HCHO and / or 5,5-dimethyl-cyclohexane-1,3-dione (dimedone) have been developed to determine the concentration of HCHO in blood

of rats, monkeys and humans.[43–46] Reported concentrations for HCHO in human blood vary between 0.4 – 2.77 μg $mL^{-1}$ (20 – 100 μM).

Apart from radiolabelling, fluorescent probes have been developed to image HCHO fluxes in epileptic brains during metabolic and epigenetic processes in animals.[46] Studies on HCHO levels in blood of rhesus monkeys (6 ppm), F-344 rats (15 ppm), and humans (2 ppm) have shown that under controlled exposure conditions, no increase in blood HCHO levels was observed *via* inhalation.[44,47] Ingestion of a significant quantity of HCHO similarly does not appear to change the HCHO levels in blood.[48] However, following ingestion of HCHO, the level of formate in blood is increased, suggesting fast metabolism of HCHO. HCHO has an estimated half-life of 1.5 min, and is mostly excreted as formate.[49] The total content of HCHO in humans has been determined to be between 1.75 – 2.6 mg $kg^{-1}$, and HCHO is considered to be present in all aqueous tissue.[50,51] This leads to an estimated rate of 22 – 41 mg · $min^{-1}$ of HCHO being formed and metabolised to maintain a stable equilibrium.[5,50–52] Per day (24 h) this rate equates to 31 – 58 g of HCHO. In perspective, the estimated amount of exogenous HCHO obtained from ingestion or inhaling accounts for 1 – 2% of the estimated total HCHO being metabolised on a daily basis.[5,52]

### 6.3.2 Biochemical processes that release formaldehyde

Reactions inside the human body produce HCHO as a metabolite or use HCHO in catabolism. For example, Ser can reversibly be converted to Gly and HCHO by pyridoxal phosphate dependent serine hydroxymethyltransferase (SHMT, Figure 6.3 **A**).[53] HCHO produced in this reaction is condensed with tetrahydrofolate (**THF**) forming 5,10-methylenetetrahydrofolate (**5,10-$CH_2$-THF**) which serves as a one-carbon (1C) carrier as part of the folate (vitamin $B_9$) cycle.[54,55] First Ser forms an imine with pyridoxal phosphate, and in presence of **THF** HCHO is transferred *via* a retro aldol reaction (Figure 6.3 **A**).[56–58] SHMTs contribute considerably to the available amount of formate and indirectly to the amount of HCHO in mice.[59]

Dimethylglycine dehydrogenase (DMDGH) is a flavoprotein which catalyses the oxidative demethylation of dimethylglycine (Figure 6.3 **B**).[60,61] DMDGH is found in mitochondria of liver cells and is part of the

choline degradation.[60,61] During DMDGH catalysis, dimethylglycine is demethylated to produce Gly and HCHO. The released HCHO reacts with tetrahydrofolate to form **5,10-$CH_2$-THF** (Figure 6.3 **B**).

Semicarbazide-sensitive amine oxidase (SSAO) catalyses the oxidative deamination of primary amines and is related to monoamine oxidases (Figure 6.3 **C**). SSAO is distinctive due to its inhibition by semicarbazide, but it is not inhibited by typical acetylenic monoamine oxidase inhibitors (e. g. clorgyline, L-deprenyland pargyline).[62–64] Methylamine is a metabolite produced by bacteria from the breakdown of sarcosine, epinephrine, and creatinine.[62–64] Methylamine is also a major metabolite in nicotine metabolism.[65] Deamination of endogenous methylamine by SSAO results in the formation of $NH_3$, $H_2O_2$, and HCHO (Figure 6.3 **C**).

Fatty acid metabolism in mammalian liver by lipid peroxidation of meat produces different carbonyl products.[66–69] More specifically, the peroxidation of arachidonic-, linolinic-, linoleic-, and oleic acid have all been demonstrated to produce aldehydes including malonaldehyde, 4-hydroxy-2-hexenal, 4-hydroxy-2-nonenal, acrolein, 2-pentenal, 2-hexenal, 2-heptenal, 2-octenal, acetaldehyde, and HCHO.[67,69,70] HCHO was reported to contribute to a large portion of the aldehyde products formed.[67]

Cytochrome P450 enzymes catalyse a diverse set of oxidative reactions including aromatic and aliphatic hydroxylation, oxidation of sulphur / nitrogen / alkene, deamination, dehalogenation, desulphurization, and de alkylation. The oxidative dealkylation of methyl groups, e. g. of those in ephedrine or benzphetamine,[71,72] by cytochrome P450 results in the formation of HCHO (Figure 6.3 **D**).[73]

Methylated residues on histone proteins (Lys or Arg) are epigenetic marks that regulate gene expression (Figure 6.3 **E**).[74] These methyl marks on Lys residues are dynamic and can be removed by Lys demethylases (KDMs) or flavin adenine dinucleotide (FAD) dependent amino oxidases.[75] Although enzymes for the methylation of Arg have been identified,[76] the pathway for specific Arg demethylation has not been reported.[77] Some KDMs have been shown to also recognize methylated Arg residues.[78] Nucleic acids, i.e. DNA and RNA, are also subject to methylation. Studies of the demethylation of 5-methyl cytosine by methyl-CpG-binding domain from SG9 insect cells,[79] and *Escherichia coli* DNA repair enzyme AlkB,[80] have both been shown to form HCHO during demethylation. However, whether equivalent reactions occur in humans is still debated.

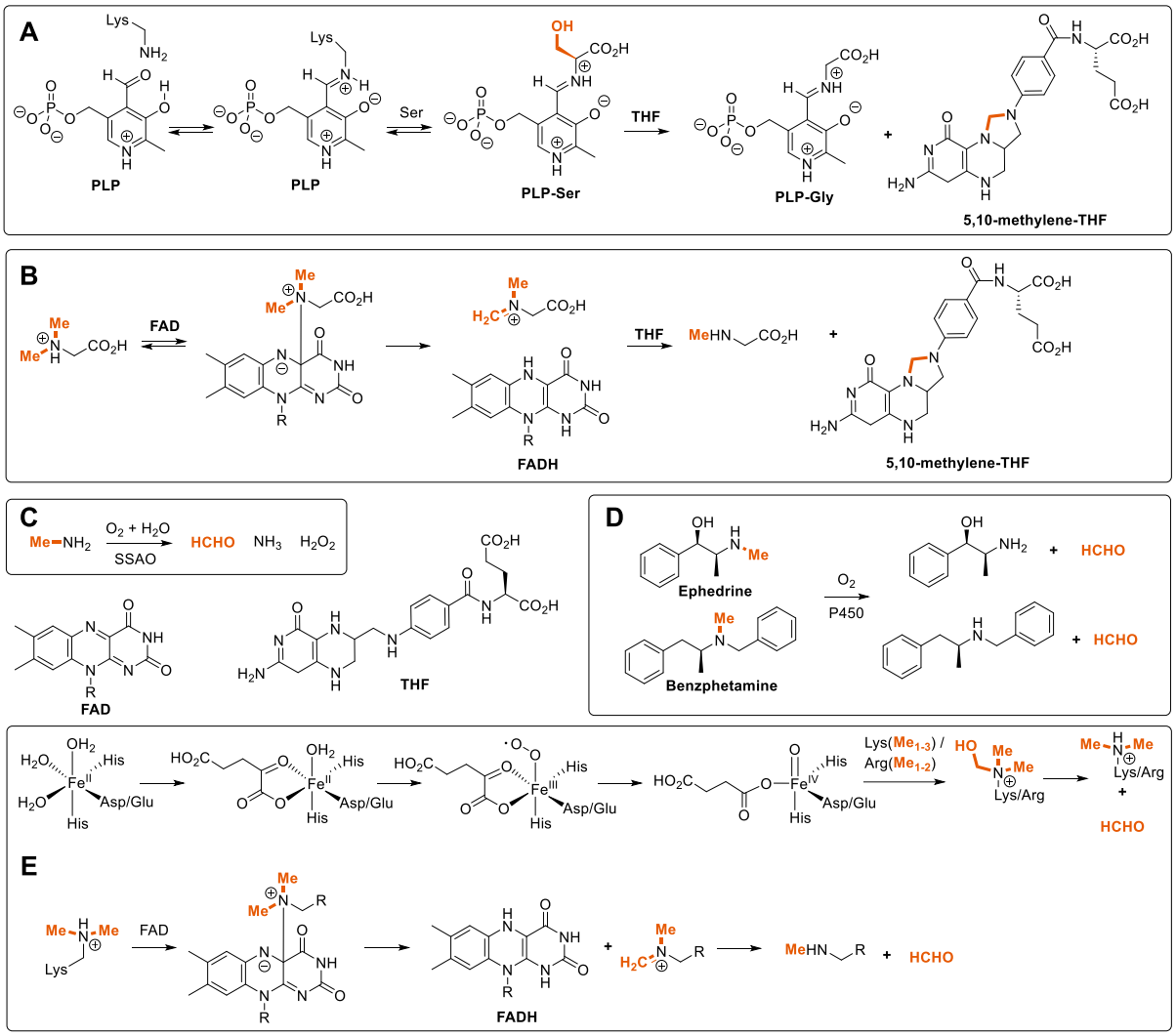



**Figure 6.3:** Overview of some of the biochemical reactions that release HCHO as a 'by-product'. The reactions include enzymes such as serine hydroxymethyl transferases (**A**), dimethylglycine dehydrogenases (**B**), semicarbazide-sensitive oxidases (**C**), cytochrome P450 (**D**), and histone demethylases (**E**). The HCHO producing structural features are highlighted in orange.

### 6.3.3 Metabolism of formaldehyde

The metabolism of HCHO relies on reactions with nucleophiles, in particular in humans with glutathione (GSH, Figure 6.4 **A**).[81] This mechanism is shared by plants, animals, and some bacteria. The Cys thiol in GSH reacts with HCHO to give *S*-hydroxymethylglutathione (**HMG**), which is oxidised by NAD-dependent alcohol dehydrogenases (ADH5 in humans).[82,83] The product, *S*-formylglutathione (**SFG**), undergoes hydrolysis catalysed by *S*-formylglutathione hydrolase, to give formate (Figure 6.4). Formate can then enter C1 metabolism and react to give 5-formyl-**THF**, which can be processed further into nucleic

acids and amino acids (Figure 6.4).[84,85] GSH and ADH5 are also involved in detoxification mechanism of nitric oxide (NO), in which *S*-nitrosoglutathione is formed. As a result, ADH5 is sometimes referred to as nitrosglutathione reductase.[86] The non-enzymatic reaction of HCHO with GSH in aqueous solutions can result in the formation of multiple (bi)cyclic products (Figure 6.4 **B**).[87] The bicyclic product (5*R*,10*S*)-5-(carboxymethylcarbamoyl)-7-oxo-3-thia-1,6- diazabicyclo[4.4.1]undecane-10-carboxylic acid (**BiGF$_2$**) was observed by crystallography in complex with human carbonyl reductase 1.[88]

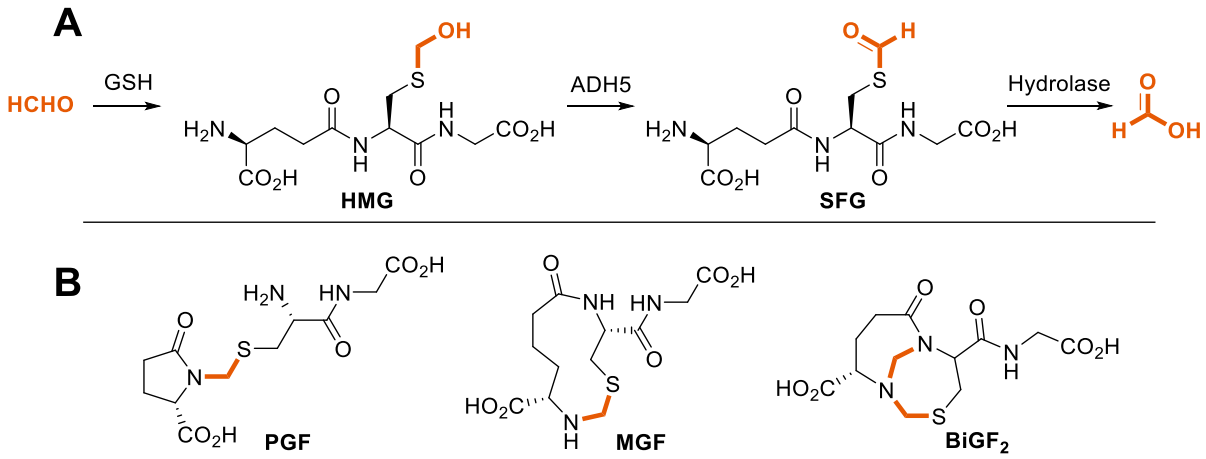



**Figure 6.4:** Detoxification pathway of HCHO involving glutathione (GSH). **A** The reaction of HCHO with GSH results in the formation of *S*-hydroxymethyl GSH (**HMG**), and is subsequently oxidised to give *S*-formyl GSH (**SFG**) by for instance aldehyde dehydrogenase 5 (ADH5), and releases formic acid after hydrolysis in the final step catalysed by hydrolase enzymes;[81] **B** alternative products formed in the reaction of GSH with HCHO, including **PGF**, **MGF**, and **BiGF$_2$**.[87] The HCHO derived structural features are highlighted in orange.

DMDGH and SHMT are involved in C1 metabolism (Figure 6.5 **A**) and store HCHO as **5,10-CH$_2$-THF** (Figure 6.5). The latter can be reduced to form 5-methyl-**THF** and methylate homocysteine (Hcy), and is involved in the biosynthesis of *S*-adenosylmethionine (SAM, Figure 6.5 **B**). **5,10-CH$_2$-THF** is also a substrate for thymidylate synthase (Figure 6.5 **C**). Alternatively, **5,10-CH$_2$-THF** can be oxidized to form 5-formyl-**THF** which is used in the biosynthesis of adenosine monophosphate (Figure 6.5 **D**). HCHO can be excreted as $CO_2$ and exhaled,[44] or excreted *via* urine or feces.[5] Dysregulation of the C1 pathway is related to different diseases, including Fanconi anemia (FA, Section **6.4.3**).

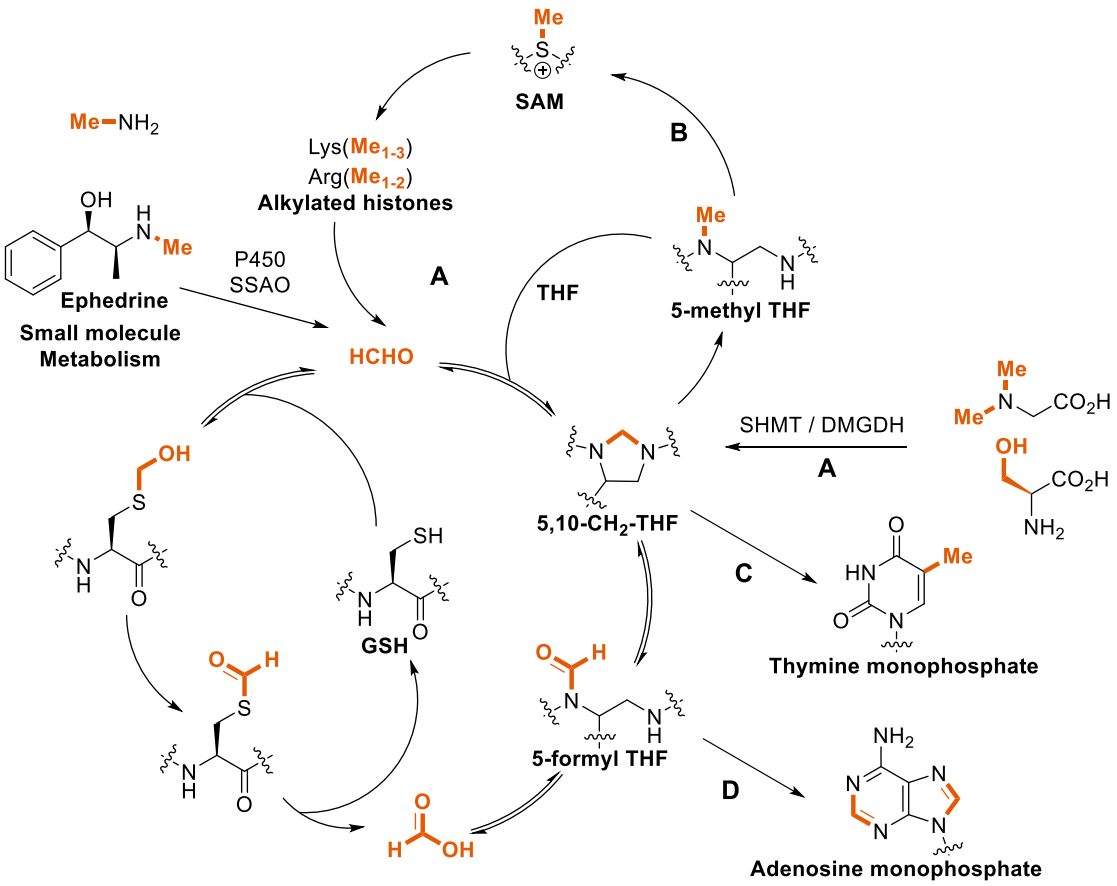



**Figure 6.5:** A selection of reactions from the C1 metabolism that involve HCHO and their relationship to each other. Highlighted are: **A** Serine hydroxymethyl transferase (SHMT) and dimethylglycine dehydrogenase (DMGDH); **B** methyl transfer from 5-methyl tetrahydrofolate (5-methyl-**THF**); **C** incorporation of HCHO into thymine monophosphate; **D** incorporation of HCHO into adenosine monophosphate.[81] The HCHO derived structural features are highlighted in orange.

## 6.4 Effects of formaldehyde on health

### 6.4.1 Toxic effects of formaldehyde

In humans, acute exposure to high levels of HCHO can lead to irritation of the eyes, nose and throat, and can eventually lead to death.[89] Reversible irritation of eye, nose, and throat can occur at air concentrations between 0.5 – 3.7 mg $m^{-3}$.[90] The human odour threshold for HCHO is reported to be 0.03 mg $m^{-3}$.[1] Studies on irritation by HCHO report a threshold level of 0.5 – 1 ppm, with irritation of eyes being most sensitive.[91] The skin sensitisation properties of HCHO are well-known.[30] Allergic reactions to HCHO through contact with cosmetics or clothing have long been reported.[24] Whether HCHO is directly linked to asthma, or whether airway irritation resembles the symptoms of asthma is still debated.[92] Although extensively studied,

no clear evidence that HCHO induces asthma has been presented.[93] Long term exposure results in irritation of eyes and upper and lower airways. Sensory irritation is the predominant effect of long term HCHO exposure, which will progress to cytotoxic irritation. These effects are time independent, and are concentration dependent.[1]

### 6.4.2 Formaldehyde and carcinogenicity

In 1980 Swenberg *et al.* found evidence that long term exposure of F-344 rats to airborne HCHO (15 ppm, 18 months) caused squamous cell carcinomas in the nasal cavity in rats.[94] This discovery was followed up by a study using rats and mice exposed to a range of HCHO concentrations (2 – 15 ppm, 24 months 6 h/day 5 days per week); the results demonstrated that HCHO induced lesions in the nasal cavity.[95] Casanova and co-workers have published a series of studies showing the airborne uptake of HCHO in F-344 rats and rhesus monkeys.[96–99] Nasal cancer was induced in rats at higher HCHO concentrations (10 – 15 ppm), and results in an increase in DNA-protein crosslinking, which was confirmed using radiolabelled $H^{14}CHO$. At HCHO concentrations exceeding 10 ppm, the major detoxification pathway, which relies on GSH, in rats is proposed to be saturated, explaining the increased incidence of DNA-protein crosslinking. Note that at all HCHO concentrations (0.3 – 15 ppm) DNA-protein crosslinking was observed in respiratory mucosa. Investigation of bone marrow (femur) showed no evidence for the incorporation of labelled HCHO. Exposure of rats, monkeys, and humans to HCHO did not result in a measurable difference in the concentration of HCHO in their blood,[44,47] and thus toxicity induced remotely from deposition are considered unlikely by many authors.[1,100–103] The proposed mechanism of action in rats leading to nasal tumour development involves repeated cytotoxicity.[104]

A large cohort study by Hauptmann *et al.* investigated the mortality of workers in HCHO using industries, which showed an increased incidence of nasopharyngeal cancer.[105] The results of this study form the basis for the classification of HCHO as a human carcinogen by the International Agency for Research on Cancer (IARC).[1,3,4] However, the evaluation by the IARC has been challenged by different authors.[100,106] Internal comparison analysis and different categorisation by Marsh and Youk[106] on the data used in the study by Hauptman *et al.*[105] did not show any statistically significant increase in nasopharyngeal cancer as a result of

HCHO exposure. Additionally, some of the workers in the cohort studies were also exposed to other suspected risk factors (metal dust or sulphuric acid mists) that could cause respiratory system cancer.[107] The data reported by Hauptman *et al.*[105] are inconsistent with results from related studies.[108] A review of the literature by different authors suggests that it is debatable whether exposure to HCHO increases the risk of nasopharyngeal cancer in humans.[102,109] Recent studies have shown that airborne HCHO does induce DNA modifications correlated to the HCHO exposure concentration in rats.[110] However, no DNA modifications were observed when moving to different species.[111]

Two large cohort studies by Hauptman *et al.* and Pinkerton *et al.* studied the correlation between leukaemia and HCHO exposure.[112,113] The IARC evaluated the results of the epidemiologic studies and concluded that the evidence was strong, but insufficient to link HCHO exposure to leukaemia in humans.[1] However, the results of a third study are inconsistent with the findings by Hauptman *et al.* and Pinkerton *et al.*.[108] The evaluation of the statistical data of these studies identifying a link between HCHO and leukaemia have been questioned[1,4,100,110] and lack support of a biological/biochemical mode of action.[100] A more recent study by Hauptmann *et al.*[114] correlating HCHO exposure to leukaemia under embalmers, was criticised for a lack of correlation to HCHO concentrations.[4,110] Indirect mechanisms for the induction of leukaemia have been proposed to damage hematopoietic stem/progenitor or primitive pluripotent stem cells which then migrate to the bone marrow to form the initiating leukaemia cells, as an explanation.[115,116] Extensive investigations over the past three decades have demonstrated that HCHO does not reach internal organs.[111] Thus, it appears unlikely, at least below threshold levels, that HCHO alone causes leukaemia,[1,4,100,110,111] and some have suggested re-evaluation of the current classification of HCHO as a carcinogen by IARC.[102,111]

### 6.4.3 Health effects related to endogenous formaldehyde

The genotoxic effects associated with HCHO have been proposed to result from DNA-protein crosslinks and DNA adducts.[117,118] Both endogenous and exogenous HCHO are believed to cause DNA-protein crosslinks and adducts, making it difficult to distinguish the source of the effect.[96,119] Apart from forming DNA-protein crosslinks, *in vitro* studies have shown that HCHO is able to form inter DNA strand crosslinks, and hydroxymethyl adducts on nucleobases.[120–128] Studies of model systems have demonstrated

that HCHO induced adducts are able to cause mutations in multiple genes.[129–131] In situations where the DNA repair mechanism is compromised, these HCHO-induced modifications can play a role in increasing the risk of cancer.[117] Fanconi anemia (FA) is a rare genetic disease in humans that is marked by genomic instability, bone marrow failure, and a cancer predisposition.[132,133] Defective repair of endogenously produced inter strand DNA crosslinks as result of reactive aldehydes are the main cause of FA.[134] These reactive aldehydes include HCHO, acetaldehyde, acrolein, malondialdehyde, and crotonaldehyde. Endogenous HCHO has been shown to be an important source of DNA damage.[117,135] These combined results suggest a strong link between disrupted aldehyde metabolism and DNA repair in FA. HCHO can act as a genotoxin or as a C1 metabolite.[81,136]

## 6.5 Formaldehyde chemistry

HCHO is an electrophile, and thus is prone to reaction with nucleophiles in aqueous solutions. On its own, HCHO in $H_2O$ rapidly reacts to give hydrated HCHO (Figure 6.6).[137] Cyclic oligomeric HCHO products such as meta-HCHO can also be formed (Figure 6.6). Para-HCHO is poorly soluble in $H_2O$ but can be cracked to produce individual molecules of hydrated HCHO by heating. In presence of amines, HCHO readily forms imines (Schiff base[138]), which in turn are electrophilic. These HCHO adducts are hydrolytically cleaved and in equilibrium with free HCHO. Schiff bases formed in presence of a reducing agent (e. g. $NaBH_3CN$) are readily reduced, forming methylated amines.[139] HCHO derived imines under basic conditions can be reduced in presence of formic acid releasing $CO_2$ in the Eschweiler-Clarke reduction / methylation.[140–142] Alternatively, HCHO can undergo disproportionation reactions with itself or other aldehydes; Cannizzaro reaction.[143,144] The Cannizzaro reaction involves aldehydes that lack α-hydrogens, preventing them from undergoing aldol condensation. These reactions occur most efficiently in basic aqueous solutions. HCHO in acidic solutions is also capable of disproportionation.

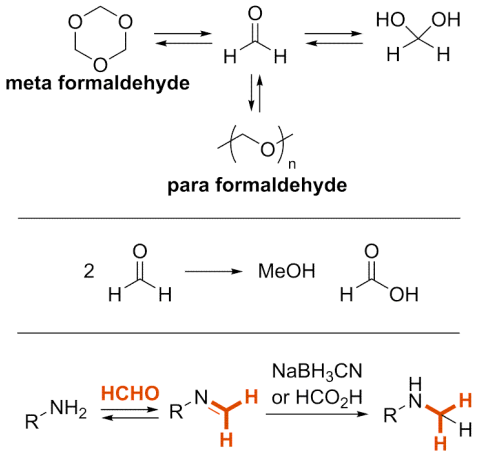



**Figure 6.6:** General chemistry involving HCHO showing the hydrated state, and Cannizzaro and Eschweiler-Clarke reactions. In water, HCHO is mainly found in its hydrated form, but also in oligomeric forms. The HCHO derived structural features are highlighted in orange.

### 6.5.1 Reactions of formaldehyde with biological molecules

Due to its highly reactive nature HCHO, has been found to react with a range of different nucleophiles. Studies on the *in vitro* reactivity of HCHO with macromolecules have shown that it reacts with DNA, RNA, and proteins. Amongst the products observed after HCHO exposure are DNA adducts, protein adducts, protein modifications, inter DNA strand cross-links, and protein-DNA cross-links. HCHO reacts with exocyclic amine groups present on cytosine, guanine, and adenine.[123,125] The reactions of HCHO with thymine and uracil occur on the endocyclic base *N*, however generally at slower rate.[123,126] The hydroxymethyl adducts that are formed have varying stabilities.[123,125,126] Unstructured polymers of the nucleotides show similar reactivity profile to their constituent monomers; hydroxymethylation of the poly nucleotides takes place at the same rate.[126,127] Double stranded DNA reacts with HCHO in a reversible way with HCHO, and changes the melting point of the DNA complex, suggesting alterations in the structure. The internal hydrogen bonding between the two DNA strands is affected by HCHO.[127,128] Nucleosides and *tert*-butyloxycarbonyl protected amino acids are able to covalently link when reacting with HCHO; adenosine, guanine, cytosine were found to react with Lys, Cys, His, and Trp.[145] Conversely, thymine, Arg, Tyr, Asn, and Gln were not able to react in presence of HCHO.

The reactions of HCHO with proteins result in hydroxymethyl and other modifications.[146] Cross-linking of amine groups, both intra- and inter-molecularly, has been observed.[146,147] In particular, Arg, Trp, Tyr, and His residues are able to react with HCHO to form cross-links.[147–149] When in presence of free primary / secondary amines, and amides, HCHO preferentially reacts with amines.[150] Condensation of Arg on proteins was limited to free amines; no reactions between Arg on proteins, HCHO and free amides were observed.[150] More recent structural and biochemical studies on the diphtheria toxin showed that all primary amino groups (Lys side chains) present on the protein were accessible to HCHO, and a third of them was able to produce cross-links.[23] The resultant protein was catalytically inactive, but still recognized by antigens. Studies on histone proteins exposed to HCHO suggested that unmodified Lys on the unstructured tails, and in the globular domain are able to condense with HCHO.[151] Post-translational modifications on Lys residues prohibited reactions.

Crosslinking of DNA with proteins is suggested to be a major DNA alteration induced by HCHO.[145,152] Lys, Cys, His, and Trp residues present on proteins can be involved in crosslinking with guanine, cytosine and adenosine nucleotides present on the DNA.[145,152] Histone proteins, which contain many Lys residues, cross-link efficiently with DNA in presence of HCHO.[145] In order to map out modifications of each amino acid with HCHO, model peptide systems were allowed to react with HCHO in presence of different amino acids. These studies identified peptides containing Tyr to readily crosslink with free Gly in the presence of HCHO.[153]

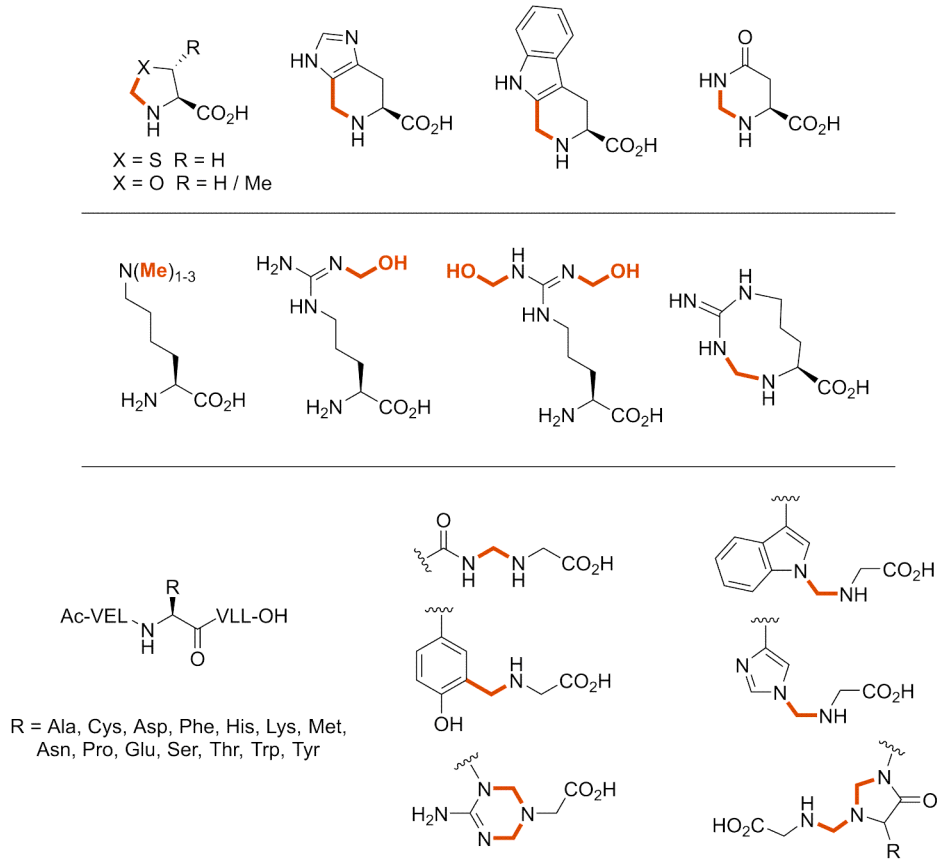



**Figure 6.7:** Products of reactions between HCHO and free amino acids or in peptides that have been identified in previous studies.[153–157] The HCHO derived structural features are highlighted in orange.

Following pioneering work in the 1920s, numerous studies have reported on the reaction of HCHO with amino acids.[147,149,150,158–168] Many products have been identified after purification, typically in context of that particular amino acid. More recent studies have been carried out and identified Lys and Arg as reaction partners of HCHO; Arg was suggested to be a carrier molecule of HCHO in biological systems.[155] However, a limited number of studies on amino acids has been reported using modern techniques. Additionally, few studies have considered the relative rates of reaction and stability between amino acids and HCHO.

## 6.6 Aims

The broad aims of the work related to HCHO which is described in Chapter 7 of this thesis were to increase the understanding of the reactivity of HCHO with amino acids, including (i) the identification of products of the reaction between HCHO and amino acids; (ii) investigating the relative rates of formation of the different products, and their relative stabilities; (iii) investigating the effects of concentration and pH on the HCHO derived adduct formation; and comparing the reactivity of HCHO with that of other aldehydes.

## 6.7 References


1. Salthammer, T., Mentese, S. & Marutzky, R. Formaldehyde in the indoor environment. *Chem. Rev.* **110**, 2536–2572 (2010).
2. Hofmann, A. W. Contributions to the history of methylic aldehyde. *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **16**, 156–159 (1867).
3. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Chemical agents and related occupations. *IARC Monogr. Eval. Carcinog. Risks Hum.* **100**, 9–562 (2012).
4. Swenberg, J. A. *et al.* Formaldehyde carcinogenicity research: 30 years and counting for mode of action, epidemiology, and cancer risk assessment. *Toxicol. Pathol.* **41**, 181–189 (2013).
5. Dhareshwar, S. S. & Stella, V. J. Your prodrug releases formaldehyde: Should you be concerned? No! *J. Pharm. Sci.* **97**, 4184–4193 (2008).
6. Tang, X. *et al.* Formaldehyde in China: Production, consumption, exposure levels, and health effects. *Environ. Int.* **35**, 1210–1224 (2009).
7. Na, K. & Cocker, D. R. Fine organic particle, formaldehyde, acetaldehyde concentrations under and after the influence of fire activity in the atmosphere of Riverside, California. *Environ. Res.* **108**, 7–14 (2008).
8. Kleinman, M. T. & Mautz, W. J. *The Effects of Exercise on Dose and Dose Distribution of Inhaled Automotive Pollutants. Research Report 45* (1991).
9. WHO. *WHO Guidelines for indoor air quality*. (2010) doi:10.1016/B978-0-12-386454-3.00911-8.
10. Gilbert, N. L. *et al.* Housing characteristics and indoor concentrations of nitrogen dioxide and formaldehyde in Quebec City, Canada. *Environ. Res.* **102**, 1–8 (2006).
11. Dingle, P. & Franklin, P. Formaldehyde levels and the factors affecting these levels in homes in Perth, Western Australia. *Indoor Built Environ.* **11**, 111–116 (2002).
12. Marchand, C., Bulliot, B., Le Calvé, S. & Mirabel, P. Aldehyde measurements in indoor environments in Strasbourg (France). *Atmos. Environ.* **40**, 1336–1345 (2006).
13. Solt, P. *et al.* Technological performance of formaldehyde-free adhesive alternatives for particleboard industry. *Int. J. Adhes. Adhes.* **94**, 99–131 (2019).
14. Peng, C. Y., Lan, C. H., Lin, P. C. & Kuo, Y. C. Effects of cooking method, cooking oil, and food type on aldehyde emissions in cooking oil fumes. *J. Hazard. Mater.* **324**, 160–167 (2017).
15. Baker, R. R. The generation of formaldehyde in cigarettes-Overview and recent experiments. *Food Chem. Toxicol.* **44**, 1799–1822 (2006).
16. Gillman, I. G., Kistler, K. A., Stewart, E. W. & Paolantonio, A. R. Effect of variable power levels on the yield of total aerosol mass and formation of aldehydes in e-cigarette aerosols. *Regul. Toxicol. Pharmacol.* **75**, 58–65 (2016).
17. Salamanca, J. C. *et al.* E-cigarettes can emit formaldehyde at high levels under conditions that have been reported to be non-averse to users. *Sci. Rep.* **8**, 6–11 (2018).

18. d'Alessandro, A. *et al.* Formate in serum and urine after controlled methanol exposure at the threshold limit value. *Environ. Health Perspect.* **102**, 178–181 (1994).

19. Glenny, A. T. & Hopkins, B. E. Diphtheria Toxoid As an Immunizing Agent. *Br. J. Exp. Pathol.* **4**, 283–288 (1923).

20. Sabbaghi, A., Miri, S. M., Keshavarz, M., Zargar, M. & Ghaemi, A. Inactivation methods for whole influenza vaccine production. *Rev. Med. Virol.* **29**, 1–11 (2019).

21. Metz, B., Jiskoot, W., Hennink, W. E., Crommelin, D. J. A. & Kersten, G. F. A. Physicochemical and immunochemical techniques predict the quality of diphtheria toxoid vaccines. *Vaccine* **22**, 156–167 (2003).

22. Ogasawara, N., Misaki, M. & Adachi, S. Method for the determination of residual formaldehyde in an inactivated Zika vaccine formulated with aluminum hydroxide gel. *Biologicals* **58**, 73–75 (2019).

23. Metz, B. *et al.* Identification of Formaldehyde-Induced Modifications in Diphtheria Toxin. *J. Pharm. Sci.* **109**, 543–557 (2020).

24. Thyssen, J. P., Johansen, J. D. & Menné, T. Contact allergy epidemics and their controls. *Contact Dermatitis* **56**, 185–195 (2007).

25. Pexe, M. E. *et al.* Hairdressers are exposed to high concentrations of formaldehyde during the hair straightening procedure. *Environ. Sci. Pollut. Res.* **26**, 27319–27329 (2019).

26. Helander, I. Contact Urticaria From Leather Containing Formaldehyde. *Arch. Dermatol.* **113**, 1443 (1977).

27. Lindskov, R. Contact urticaria to formaldehyde. *Contact Dermatitis* **8**, 333–334 (1982).

28. Young, A. S. *et al.* Phthalate and Organophosphate Plasticizers in Nail Polish: Evaluation of Labels and Ingredients. *Environ. Sci. Technol.* **52**, 12841–12850 (2018).

29. Hauksson, I. *et al.* Skincare products containing low concentrations of formaldehyde detected by the chromotropic acid method cannot be safely used in formaldehyde-allergic patients. *Br. J. Dermatol.* **174**, 371–379 (2016).

30. De Groot, A. C., Flyvholm, M. A., Lensen, G., Menné, T. & Coenraads, P. J. Formaldehyde-releasers: Relationship to formaldehyde contact allergy. Contact allergy to formaldehyde and inventory of formaldehyde-releasers. *Contact Dermatitis* **61**, 63–85 (2009).

31. Malinauskiene, L., Blaziene, A., Chomiciene, A. & Isaksson, M. Formaldehyde may be found in cosmetic products even when unlabelled. *Open Med.* **10**, 323–328 (2015).

32. Miller, H. E. & Taussig, L. R. Cosmetics. *J. Am. Med. Assoc.* **84**, 1999–2002 (1925).

33. Gholami, A., Mohsenikia, A. & Masoum, S. Determination of Very Low Level of Free Formaldehyde in Liquid Detergents and Cosmetic Products Using Photoluminescence Method. *J. Anal. Methods Chem.* **2016**, (2016).

34. Fasth, I. M., Ulrich, N. H. & Johansen, J. D. Ten-year trends in contact allergy to formaldehyde and formaldehyde-releasers. *Contact Dermatitis* **79**, 263–269 (2018).

35. Lehmann, S. V., Hoeck, U., Breinholdt, J., Olsen, C. E. & Kreilgaard, B. Characterization and chemistry of imidazolidinyl urea and diazolidinyl urea. *Contact Dermatitis* **54**, 50–58 (2006).

36. De Groot, A. C., Blok, J. & Coenraads, P. J. Relationship between formaldehyde and quaternium-15 contact allergy. Influence of strength of patch test reactions. *Contact Dermatitis* **63**, 187–191 (2010).

37. Asada, A., Doi, T., Takeda, A. & Kajimura, K. Quantification of 1,3-dimethylol-5,5-dimethylhydantoin and its decomposition products in cosmetics by high-performance liquid chromatography. *J. Pharm. Biomed. Anal.* **67–68**, 163–168 (2012).

38. Sanyal, A. K., Basu, M. & Banerjee, A. B. Rapid ultraviolet spectrophotometric determination of bronopol: Application to raw material analysis and kinetic studies of bronopol degradation. *J. Pharm. Biomed. Anal.* **14**, 1447–1453 (1996).

39. Aalto-Korte, K., Kuuliala, O., Suuronen, K. & Alanko, K. Occupational contact allergy to formaldehyde and formaldehyde releasers. *Contact Dermatitis* **59**, 280–289 (2008).

40. Belsito, D. V. & Skinner, S. M. Allergic contact dermatitis from formaldehyde resins in permanent press clothing: An underdiagnosed cause of generalized dermatitis. *J. Am. Acad. Dermatol.* **27**, 962–968 (1992).

41. O'quinn, S. E. & Kennedy, C. B. Contact Dermatitis Due to Formaldehyde in Clothing Textiles. *J. Am. Med.*

*Assoc.* **194**, 593–596 (1965).

42. Trattner, A., Johansen, J. D. & Menné, T. Formaldehyde concentration in diagnostic patch testing: Comparison of 1% with 2%. *Contact Dermatitis* **38**, 9–13 (1998).

43. Szarvas, T. *et al.* Determination of Endogenous Formaldehyde Level in Human Blood and Urine by Dimedone-$^{14}$C Radiometric method. *J. Radioanal. Nucl. Chem.* **106**, 357–367 (1986).

44. Heck, H. A. *et al.* Formaldehyde (CH2O) Concentrations in the Blood of Humans and Fischer-344 Rats Exposed to $CH_2O$ Under Controlled Conditions. *Am. Ind. Hyg. Assoc. J.* **46**, 1–3 (1985).

45. Luo, W., Li, H., Zhang, Y. & Ang, C. Y. Determination of formaldehyde in blood plasma by high-performance liquid chromatography with fluorescence detection. *J. Chromatogr. B. Biomed. Sci. Appl.* **753**, 253–257 (2001).

46. Chen, J. *et al.* Imaging of formaldehyde fluxes in epileptic brains with a two-photon fluorescence probe. *Chem. Commun.* **56**, 3871–3874 (2020).

47. Casanova, M., Heck, H. A., Everitt, J. I., Harrington, W. W. & Popp, J. A. Formaldehyde concentrations in the blood of rhesus monkeys after inhalation exposure. *Food Chem. Toxicol.* **26**, 715–716 (1988).

48. Eells, J. T. Formaldehyde Poisoning. *J. Am. Med. Assoc.* **246**, 1237–1238 (1981).

49. McMartin, K. E., Martin-Amat, G., Noker, P. E. & Tephly, T. R. Lack of a role for formaldehyde in methanol poisoning in the monkey. *Biochem. Pharmacol.* **28**, 645–649 (1979).

50. Owen, B. A., Dudney, C. S., Tan, E. L. & Easterly, C. E. Formaldehyde in drinking water: Comparative hazard evaluation and an approach to regulation. *Regul. Toxicol. Pharmacol.* **11**, 220–236 (1990).

51. Cascieri, T. C. & Clary, J. J. Formaldehyde-oral toxicity assessment. *Comments Toxicol.* **4**, 295–304 (1992).

52. Endogenous formaldehyde turnover in humans compared with exogenous contribution from food sources. *EFSA J.* **12**, 3550 (2014).

53. Schirch, V. & Szebenyi, D. M. E. Serine hydroxymethyltransferase revisited. *Curr. Opin. Chem. Biol.* **9**, 482–487 (2005).

54. Blom, H. J. & Smulders, Y. Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. *J. Inherit. Metab. Dis.* **34**, 75–81 (2011).

55. Zheng, Y. & Cantley, L. C. Toward a better understanding of folate metabolism in health and disease. *J. Exp. Med.* **216**, 253–266 (2019).

56. Scarsdale, J. N., Radaev, S., Kazanina, G., Schirch, V. & Wright, H. T. Crystal structure at 2.4 Å resolution of E. coli serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. *J. Mol. Biol.* **296**, 155–168 (2000).

57. Chiba, Y. *et al.* Mechanism for folate-independent aldolase reaction catalyzed by serine hydroxymethyltransferase. *FEBS J.* **279**, 504–514 (2012).

58. Fernandes, H. S., Ramos, M. J. & Cerqueira, N. M. F. S. A. Catalytic Mechanism of the Serine Hydroxymethyltransferase: A Computational ONIOM QM/MM Study. *ACS Catal.* **8**, 10096–10110 (2018).

59. Meiser, J. *et al.* Serine one-carbon catabolism with formate overflow. *Sci. Adv.* **2**, 1–11 (2016).

60. Porter, D. H., Cook, R. J. & Wagner, C. Enzymatic properties of dimethylglycine dehydrogenase and sarcosine dehydrogenase from rat liver. *Arch. Biochem. Biophys.* **243**, 396–407 (1985).

61. Luka, Z., Pakhomova, S., Loukachevitch, L. V., Newcomer, M. E. & Wagner, C. Folate in demethylation: The crystal structure of the rat dimethylglycine dehydrogenase complexed with tetrahydrofolate. *Biochem. Biophys. Res. Commun.* **449**, 392–398 (2014).

62. Zeisel, S. H., Wishnok, J. S. & Blusztajn, J. K. Formation of methylamines from ingested choline and lecithin. *Pharmacology* **225**, 320–324 (1953).

63. Dar, M. S., Morselli, P. L. & Bowman, E. R. The enzymatic systems involved in the mammalian metabolism of methylamine. *Gen. Pharmacol. Vasc. Syst.* **16**, 557–560 (1985).

64. Schayer, R. W., Smiley, R. L. & Kaplan, E. H. The metabolism of epinephrine containing isotopic carbon. *J. Biol. Chem.* **198**, 545–551 (1952).

65. Yu, P. H. Increase of formation of methylamine and formaldehyde in vivo after administration of nicotine

and the potential cytotoxicity. *Neurochem. Res.* **23**, 1205–1210 (1998).

66. Domínguez, R. *et al.* A comprehensive review on lipid oxidation in meat and meat products. *Antioxidants* **8**, 1–31 (2019).

67. Tamura, H., Kitta, K. & Shibamoto, T. Formation of Reactive Aldehydes from Fatty Acids in a $Fe^{2+}/H_2O_2$ Oxidation System. *J. Agric. Food Chem.* **39**, 439–442 (1991).

68. Lorenzo, J. M. Influence of the type of fiber coating and extraction time on foal dry-cured loin volatile compounds extracted by solid-phase microextraction (SPME). *Meat Sci.* **96**, 179–186 (2014).

69. Yasuhara, A. & Shibamoto, T. Quantitative Analysis of Volatile Aldehydes Formed from Various Kinds of Fish Flesh during Heat Treatment. *J. Agric. Food Chem.* **43**, 94–97 (1995).

70. Ayala, A., Muñoz, M. F. & Argüelles, S. Lipid Peroxidation: Production, Metabolism, and Signaling Mechanisms of Malondialdehyde and 4-Hydroxy-2-Nonenal. *Oxid. Med. Cell. Longev.* **2014**, 1–31 (2014).

71. Dann, R. E., Feller, D. R. & Snell, J. F. The microsomal N-demethylation of the stereoisomers of ephedrine. *Eur. J. Pharmacol.* **16**, 233–236 (1971).

72. Zhang, H., Im, S.-C. & Waskell, L. Cytochrome b 5 Increases the Rate of Product Formation by Cytochrome P450 2B4 and Competes with Cytochrome P450 Reductase for a Binding Site on Cytochrome P450 2B4. *J. Biol. Chem.* **282**, 29766–29776 (2007).

73. Karki, S. B. & Dinnocenzo, J. P. On the mechanism of amine oxidations by p450. *Xenobiotica* **25**, 711–724 (1995).

74. Jambhekar, A., Dhall, A. & Shi, Y. Roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **20**, 625–641 (2019).

75. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16**, 525–534 (2012).

76. Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* **13**, 37–50 (2013).

77. Blanc, R. S. & Richard, S. Arginine Methylation: The Coming of Age. *Mol. Cell* **65**, 8–24 (2017).

78. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7**, (2016).

79. Hamm, S. *et al.* On the mechanism of demethylation of 5-methylcytosine in DNA. *Bioorganic Med. Chem. Lett.* **18**, 1046–1049 (2008).

80. Roy, T. W. & Bhagwat, A. S. Kinetic studies of Escherichia coli AlkB using a new fluorescence-based assay for DNA demethylation. *Nucleic Acids Res.* **35**, e147–e147 (2007).

81. Hopkinson, R. J. & Schofield, C. J. Deciphering Functions of Intracellular Formaldehyde: Linking Cancer and Aldehyde Metabolism. *Biochemistry* **57**, 904–906 (2018).

82. Fernández, M. R., Biosca, J. A. & Parés, X. *S*-nitrosoglutathione reductase activity of human and yeast glutathione-dependent formaldehyde dehydrogenase and its nuclear and cytoplasmic localisation. *Cell. Mol. Life Sci.* **60**, 1013–1018 (2003).

83. Uotila, L. & Koivusalo, M. Formaldehyde Dehydrogenase from Human Liver. *J. Biol. Chem.* **249**, 7653–7663 (1974).

84. Thrasher, J. D. & Kilburn, K. H. Embryo toxicity and teratogenicity of formaldehyde. *Arch. Environ. Health* **56**, 300–311 (2001).

85. Restani, P. & Galli, C. L. Oral toxicity of formaldehyde and its derivatives. *Crit. Rev. Toxicol.* **21**, 315–328 (1991).

86. Barnett, S. D. & Buxton, I. L. O. The role of S-nitrosoglutathione reductase (GSNOR) in human disease and therapy. *Crit. Rev. Biochem. Mol. Biol.* **52**, 340–354 (2017).

87. Hopkinson, R. J., Barlow, P. S., Schofield, C. J. & Claridge, T. D. W. Studies on the reaction of glutathione and formaldehyde using NMR. *Org. Biomol. Chem.* **8**, 4915–4920 (2010).

88. Bateman, R. & Shokat, K. M. Glutathione traps formaldehyde by formation of a bicyclo [4,4,1] undecane adduct. *Org. Biomol. Chem.* **5**, 3363–3367 (2007).

89. Liteplo, R. G., Beauchamp, R., Meek, M. E. & Chénier, R. Concise International Chemical Assessment Document 40: Formaldehyde. *IPCS Concise Int. Chem. Assess. Doc.* (2002).

90. Kotzias, D. *et al. Critical Appraisal of the Setting and Implementation of Indoor Exposure Limits in the EU*. (2005).

91. Arts, J. H. E., Muijser, H., Kuper, C. F. & Woutersen, R. A. Setting an indoor air exposure limit for formaldehyde: Factors of concern. *Regul. Toxicol. Pharmacol.* **52**, 189–194 (2008).

92. Ezratty, V. *et al.* Effect of formaldehyde on asthmatic response to enhaled allergen challenge. *Environ. Health Perspect.* **115**, 210–214 (2007).

93. Arts, J. H. E., Muijser, H., Kuper, C. F. & Woutersen, R. A. Setting an indoor air exposure limit for formaldehyde: Factors of concern. *Regul. Toxicol. Pharmacol.* **52**, 189–194 (2008).

94. Swenberg, J. A., Kerns, W. D., Mitchell, R. I., Gralia, E. J. & Pavkov, K. L. Induction of Squamous Cell Carcinomas of the Rat Nasal Cavity by Inhalation Exposure to Formaldehyde Vapor. *Cancer Res.* **40**, 3398–3402 (1980).

95. Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J. & Swenberg, J. a. Carcinogenicity of Formaldehyde in Rats and Mice after Long-Term Inhalation Exposure Carcinogenicity of Formaldehyde in Rats and Mice after Long-Term Inhalation Exposure. *Cancer Res.* **43**, 4382–4392 (1983).

96. Casanova, M., Deyo, D. F. & Heck, H. A. Covalent binding of inhaled formaldehyde to dna in the nasal mucosa of fischer 344 rats: Analysis of formaldehyde and dna by high-performance liquid chromatography and provisional pharmacokinetic interpretation. *Toxicol. Sci.* **12**, 397–417 (1989).

97. Casanova, M. & Heck, H. A. Further studies of the metabolic incorporation and covalent binding of inhaled [$^{3}$H]- and [$^{14}$C]formaldehyde in Fischer-344 rats: Effects of glutathione depletion. *Toxicol. Appl. Pharmacol.* **89**, 105–121 (1987).

98. Casanova, M. *et al.* Covalent binding of inhaled formaldehyde to DNA in the respiratory tract of rhesus monkeys: Pharmacokinetics, rat-to-monkey interspecies scaling, and extrapolation to man. *Toxicol. Sci.* **17**, 409–428 (1991).

99. Casanova-Schmitz, M., Starr, T. B. & Heck, H. d. A. Differentiation between metabolic incorporation and covalent binding in the labeling of macromolecules in the rat nasal mucosa and bone marrow by inhaled [$^{14}$C]- and [$^{3}$H]formaldehyde. *Toxicol. Appl. Pharmacol.* **76**, 26–44 (1984).

100. Casanova, M. *et al.* Re: Mortality From Lymphohematopoietic Malignancies Among Workers in Formaldehyde Industries. *JNCI J. Natl. Cancer Inst.* **96**, 966–967 (2004).

101. Cole, P. & Axten, C. Formaldehyde and leukemia: An improbable causal relationship. *Regul. Toxicol. Pharmacol.* **40**, 107–112 (2004).

102. Duhayon, S., Hoet, P., Maele-Fabry, G. & Lison, D. Carcinogenic potential of formaldehyde in occupational settings: A critical assessment and possible impact on occupational exposure levels. *Int. Arch. Occup. Environ. Health* **81**, 695–710 (2008).

103. Rhomberg, L. R., Bailey, L. A., Goodman, J. E., Hamade, A. K. & Mayfield, D. Is exposure to formaldehyde in air causally associated with leukemia?-A hypothesis-based weight-of-evidence analysis. *Crit. Rev. Toxicol.* **41**, 555–621 (2011).

104. Bolt, H. M. Experimental toxicology of formaldehyde. *J. Cancer Res. Clin. Oncol.* **113**, 305–309 (1987).

105. Hauptmann, M., Lubin, J. H., Stewart, P. A., Hayes, R. B. & Blair, A. Mortality from solid cancers among workers in formaldehyde industries. *Am. J. Epidemiol.* **159**, 1117–1130 (2004).

106. Marsh, G. M. & Youk, A. O. Reevaluation of mortality risks from nasopharyngeal cancer in the formaldehyde cohort study of the National Cancer Institute. *Regul. Toxicol. Pharmacol.* **42**, 275–283 (2005).

107. Marsh, G. M., Youk, A. O., Buchanich, J. M., Erdal, S. & Esmen, N. A. Work in the metal industry and nasopharyngeal cancer mortality among formaldehyde-exposed workers. *Regul. Toxicol. Pharmacol.* **48**, 308–319 (2007).

108. Coggon, D., Harris, E. C., Poole, J. & Palmer, K. T. Extended Follow-Up of a Cohort of British Chemical Workers Exposed to Formaldehyde. *CancerSpectrum Knowl. Environ.* **95**, 1608–1615 (2003).

109. Bachand, A. M., Mundt, K. A., Mundt, D. J. & Montgomery, R. R. Epidemiological studies of formaldehyde exposure and risk of leukemia and nasopharyngeal cancer: a meta-analysis. *Crit. Rev. Toxicol.* **40**, 85–100 (2010).

110. Lu, K., Collins, L. B., Ru, H., Bermudez, E. & Swenberg, J. A. Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. *Toxicol. Sci.* **116**, 441–451 (2010).

111. Nielsen, G. D., Larsen, S. T. & Wolkoff, P. Re-evaluation of the WHO (2010) formaldehyde indoor air quality guideline for cancer risk assessment. *Arch. Toxicol.* **91**, 35–61 (2017).

112. Hauptmann, M., Lubin, J. H., Stewart, P. A., Hayes, R. B. & Blair, A. Mortality From Lymphohematopoietic Malignancies Among Workers in Formaldehyde Industries. *J. Natl. Cencer Inst.* **95**, 1615–1623 (2003).

113. Pinkerton, L. E., Hein, M. J. & Stayner, L. T. Mortality among a cohort of garment workers exposed to formaldehyde: An update. *Occup. Environ. Med.* **61**, 193–200 (2004).

114. Hauptmann, M. *et al.* Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. *J. Natl. Cancer Inst.* **101**, 1696–1708 (2009).

115. Zhang, L., Steinmaus, C., Eastmond, D. A., Xin, X. K. & Smith, M. T. Formaldehyde exposure and leukemia: A new meta-analysis and potential mechanisms. *Mutat. Res. - Rev. Mutat. Res.* **681**, 150–168 (2009).

116. Zhang, L. *et al.* Formaldehyde and Leukemia: Epidemiology, Potential Mechanisms, and Implications for Risk Assessment Luoping. *Environ. Mol. Mutagen.* **51**, 181–191 (2010).

117. Voulgaridou, G. P., Anestopoulos, I., Franco, R., Panayiotidis, M. I. & Pappa, A. DNA damage induced by endogenous aldehydes: Current state of knowledge. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **711**, 13–27 (2011).

118. Barker, S., Weinfeld, M. & Murray, D. DNA-protein crosslinks: Their induction, repair, and biological consequences. *Mutat. Res. - Rev. Mutat. Res.* **589**, 111–135 (2005).

119. Swenberg, J. A. *et al.* Endogenous versus exogenous DNA adducts: Their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol. Sci.* **120**, 130–145 (2011).

120. Zhong, W. & Hee, S. Q. Quantitation of normal and formaldehyde-modified deoxynucleosides by high-performance liquid chromatography/UV detection. *Biomed. Chromatogr.* **18**, 462–469 (2004).

121. Zhong, W. & Que Hee, S. S. Formaldehyde-induced DNA adducts as biomarkers of in vitro human nasal epithelial cell exposure to formaldehyde. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **563**, 13–24 (2004).

122. Zhong, W. & Que Hee, S. S. Comparison of UV, fluorescence, and electrochemical detectors for the analysis of formaldehyde-induced DNA adducts. *J. Anal. Toxicol.* **29**, 182–187 (2005).

123. Shishodia, S. *et al.* NMR analyses on N-hydroxymethylated nucleobases-implications for formaldehyde toxicity and nucleic acid demethylases. *Org. Biomol. Chem.* **16**, 4021–4032 (2018).

124. Moeller, B. C. *et al.* Determination of N 2-hydroxymethyl-dG adducts in the nasal epithelium and bone marrow of nonhuman primates following $^{13}CD_2$-formaldehyde inhalation exposure. *Chem. Res. Toxicol.* **24**, 162–164 (2011).

125. McGhee, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. I. Reaction with Exocyclic Amino Groups of DNA Bases. *Biochemistry* **14**, 1281–1296 (1975).

126. McGhee, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. II. Reaction with Endocyclic Imino Groups of DNA Bases. *Biochemistry* **14**, 1297–1303 (1975).

127. McGhee, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. 3. Equilibrium Denaturation of DNA and Synthetic Polynucleotides. *Biochemistry* **16**, 3267–3276 (1977).

128. McGhee, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. 4. Mechanism of the Initial Reaction of Formaldehyde with DNA. *Biochemistry* **16**, 3276–3293 (1977).

129. Speit, G., Schmid, O., Neuss, S. & Schütz, P. Genotoxic effects of formaldehyde in the human lung cell line A549 and in primary human nasal epithelial cells. *Environ. Mol. Mutagen.* **49**, 300–307 (2008).

130. Speit, G., Neuss, S., Schütz, P., Fröhler-Keller, M. & Schmid, O. The genotoxic potential of glutaraldehyde in mammalian cells in vitro in comparison with formaldehyde. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **649**, 146–154 (2008).

131. Speit, G., Neuss, S. & Schmid, O. The human lung cell line A549 does not develop adaptive protection against the DNA-damaging action of formaldehyde. *Environ. Mol. Mutagen.* **405**, NA-NA (2009).

132. Duxin, J. P. & Walter, J. C. What is the DNA repair defect underlying Fanconi anemia? *Curr. Opin. Cell Biol.*

**37**, 49–60 (2015).

133. Müller, L. U. W. & Williams, D. A. Finding the needle in the hay stack: Hematopoietic stem cells in Fanconi anemia. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **668**, 141–149 (2009).

134. Voulgaridou, G.-P., Anestopoulos, I., Franco, R., Panayiotidis, M. I. & Pappa, A. DNA damage induced by endogenous aldehydes: current state of knowledge. *Mutat. Res.* **711**, 13–27 (2011).

135. Pontel, L. B. *et al.* Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Mol. Cell* **60**, 177–188 (2015).

136. Burgos-Barragan, G. *et al.* Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism. *Nature* **548**, 549–554 (2017).

137. Winkelman, J. G. M., Voorwinde, O. K., Ottens, M., Beenackers, A. A. C. M. & Janssen, L. P. B. M. Kinetics and chemical equilibrium of the hydration of formaldehyde. *Chem. Eng. Sci.* **57**, 4067–4076 (2002).

138. Schiff, H. Mittheilungen aus dem Universitätslaboratorium in Pisa: Eine neue Reihe organischer Basen. *Justus Liebigs Ann. Chem.* **131**, 118–119 (1864).

139. Gidley, M. J. & Sanders, J. K. Reductive methylation of proteins with sodium cyanoborohydride. *Biochem. J.* **203**, 331–334 (1982).

140. Eschweiler, W. Ersatz von an Stickstoff gebundenen Wasserstoffatomen durch die Methylgruppe mit Hülfe von Formaldehyd. *Berichte der Dtsch. Chem. Gesellschaft* **38**, 880–882 (1905).

141. Clarke, H. T., Gillespie, B. & Weisshaus, S. Z. The Action of Formaldehyde on Amines and Amino Acids. *J. Am. Chem. Soc.* **55**, 4571–4587 (1933).

142. Pine, S. H. & Sanchez, B. L. The Formic Acid-Formaldehyde Methylation of Amines. *J. Org. Chem.* **36**, 829–832 (1971).

143. Cannizzaro, S. Ueber den der Benzoesaure entsprechenden Alkohol. *Liebigs Ann. der Chemie und Pharm.* **88**, 129–130 (1853).

144. Swain, C. G., Powell, A. L., Sheppard, W. A. & Morgan, C. R. Mechanism of the Cannizzaro Reaction. *J. Am. Chem. Soc.* **101**, 3576–3583 (1979).

145. Lu, K. *et al.* Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino Acids and Deoxynucleosides and Their Oligomers. *J. Am. Chem. Soc.* **132**, 3388–3399 (2010).

146. Fraenkel-Conrat, H., Cooper, M. & Olcott, H. S. The Reaction of Formaldehyde with Proteins. *J. Am. Chem. Soc.* **67**, 950–954 (1945).

147. Fraenkel-Conrat, H. & Olcott, H. S. Reaction of Formaldehyde with Proteins; II. Participation of the Guanidyl Groups and Evidence of Crosslinking. *J. Am. Chem. Soc.* **68**, 34–37 (1946).

148. Fraenkel-Conrat, H., Brandon, A. & Olcott, H. S. The reaction of formaldehyde with proteins; IV. participation of indole groups gramicidin. *J. Biol. Chem.* **168**, 99–118 (1947).

149. Fraenkel-Conrat, H. & S.Olcott, H. Reaction of formaldehyde with proteins; VI. cross-linking of amino groups with phenol, imidazole, or indole groups. *J. Biol. Chem* **174**, 827–843 (1948).

150. Fraenkel-Conrat, H. & Olcott, H. S. The Reaction of Formaldehyde with Proteins; V. Cross-linking between Amino and Primary Amide or Guanidyl Groups. *J. Am. Chem. Soc.* **70**, 2673–2684 (1948).

151. Lu, K., Boysen, G., Gao, L., Collins, L. B. & Swenberg, J. A. Formaldehyde-induced histone modifications in vitro. *Chem. Res. Toxicol.* **21**, 1586–1593 (2008).

152. Neuss, S., Holzmann, K. & Speit, G. Gene expression changes in primary human nasal epithelial cells exposed to formaldehyde in vitro. *Toxicol. Lett.* **198**, 289–295 (2010).

153. Metz, B. *et al.* Identification of Formaldehyde-induced Modifications in Proteins. *J. Biol. Chem.* **279**, 6235–6243 (2004).

154. Kallen, R. G. The mechanism of reactions involving Schiff base intermediates. Thiazolidine formation from L-cysteine and formaldehyde. *J. Am. Chem. Soc.* **93**, 6236–6248 (1971).

155. Trézl, L. *et al.* Antagonistic reactions of arginine and lysine against formaldehyde and their relation to cell proliferation, apoptosis, folate cycle and photosynthesis. *Mol. Cell. Biochem.* **244**, 167–176 (2003).

156. Mason, R. P., Sanders, J. K. M. & Gidley, M. J. NMR Visualization of Free Asparagine in Potato Tissue

Using Adduct Formation with [$^{13}$C]Formaldehyde. *Phytochemistry* **25**, 1567–1571 (1986).

157. Remelli, M., Pulidori, F., Guerrini, R. & Bertolasi, V. Synthesis of spinacine and spinacine derivatives: crystal and molecular structures of $N^{\pi}$-hydroxymethyl spinacine and $N^{\alpha}$-methyl spinaceamine. *J. Chem. Crystallogr.* **27**, 507–513 (1991).

158. Bowes, J. H. & Cater, C. W. The reaction of glutaraldehyde with proteins and other biological materials. *J. R. Microsc. Soc.* **85**, 193–200 (1966).

159. Marini, A. Spectral Detection of the reaction of Formaldehyde with Histidine Residues of α-Chymotrypsin*. *J. Biol. Chem.* **242**, 5736–5743 (1967).

160. Guthe, K. F. & Arbor, A. The Formaldehyde-Hemoglobin Reaction. *J. Biol. Chem.* **234**, 3169–3174 (1969).

161. Myers, J. S. & Hardman, J. K. Formaldehyde-induced in the a Subunit Escherichia coli Tryptophan Synthetase ". *J. Biol. Chem.* **246**, 3863–3869 (1971).

162. Brown, J. H. The Formal titration of Bacteriological Media. *J. Bacteriol.* **8**, 245–267 (1922).

163. Wadsworth, A. & Pangborn, M. C. Formaldehyde Reaction with Amino Acids. *J. Biol. Chem.* **116**, 423–436 (1936).

164. Smith, A. K., Handler, P. & Mrgudich, J. N. The Reaction of Formaldehyde with Amino Acids: X-ray Diffraction Patterns. *J. Phys. Chem.* **44**, 874–880 (1940).

165. Middlebrook, W. R. & Phillips, H. The Action of Formaldehyde on the Cystine Disulphide Linkages in Wool. *Biochem. J.* **36**, 294–302 (1942).

166. French, D. & Edsall, J. T. *Advances in Protein Chemistry: The Reactions of Formaldehyde with Amino Acids and Proteins.* (Academic Press Inc., 1945).

167. Middlebrook, W. R. The irreversible combination of formaldehyde with proteins. *Biochem. J.* **44**, 17–23 (1949).

168. Carpenter, D. C. The reaction of formaldehyde with D-Threonine. *J. Franklin Inst.* **255**, 209–214 (1953).

# Chapter 7

## Formaldehyde and its reactions with amino acids

# 7.1 Introduction

## 7.1.1 Background

From a chemical perspective, formaldehyde (HCHO) is the simplest aldehyde. There is evidence that HCHO is present in body fluids such as urine, blood, intercellular tissue, as well as in cells (Chapter 6).[1] Humans are exposed to exogenous sources of HCHO, e. g. *via* inhalation (Section **6.2**), and produce HCHO as an endogenous metabolite *via* enzyme catalysed reactions (Section **6.3**). Above threshold levels, exogenous HCHO is toxic for humans, and has been associated with cancer in animal studies. Recent work in a mouse model has revealed a strong correlation between compromised HCHO metabolism and organ dysfunction, bone marrow failure, leukaemia, and liver cancer.[2] Elevated HCHO levels were found under these conditions. HCHO is a highly potent electrophile that can react with a diverse set of nucleophiles in a biological setting. Indeed, various studies from the 1920s onwards concerning the reactivity of amino acids/peptides with HCHO support its potential for reaction with biomolecules[3,4,13–17,5–12] (Section **6.5**). HCHO is a potent cross-linking agent of biomacromolecules, causing intra-strand DNA and protein-DNA cross-links *in vitro*, which are associated with genotoxic effects.[18,19] It remains unclear to what extend these reactions occur in cells, though it is probable that these or related reactions are linked to its toxic and carcinogenic effects. The endogenous production of HCHO suggests some of these adducts or products formed could play regulatory or sensing roles. Recent reports on the reaction of HCHO with Arg, Lys,[20] short peptides[21] and nucleic acids[19,22] have used modern techniques to investigate the reactivity of HCHO with biological building blocks; these studies precede the work presented in this chapter.

## 7.1.2 Approach

To better understand how HCHO affects cellular functions in health and disease, it is important to define its reactivity towards biomolecules. Systematic studies using model systems could provide comprehensive insight into the reactivity of HCHO. The work described in this chapter employed an NMR spectroscopy based approach to investigate the reactions of HCHO with amino acids. The aims of the work included: (i) to characterise the products of the reactions between amino acids and HCHO; (ii) to evaluate their relative stabilities and rate of formation; (iii) to investigate the effects of concentration and pH on their formation and stabilities; and (iv) to compare the reactivity of HCHO to other biologically relevant ketones.

## 7.2 HCHO in aqueous solution

The electrophilic character of HCHO means it exists in different forms in solution. Thus, it was anticipated that a number of species, including oligomeric forms and disproportionated products, would exist in the formalin solutions used in this study.[23–26] The HCHO solutions used here were initially analysed for their composition under the different conditions used. The formalin solutions used were prepared from solid commercial paraformaldehyde, which was suspended in $D_2O$ and then heated (85 – 95 °C) until a clear solution was obtained. The subsequent formalin solution was left at room temperature for at least 16 hours to reach equilibrium. The major species detected by $^1$H NMR in solution was hydrated HCHO, which was identified by its characteristic singlet resonance at $\delta_H$ 4.74 ppm (Figure 7.1, ●). On the basis of $^1$H NMR analysis, this species accounted for over 98% of all identified species in solution. The hydrated dimeric species of HCHO (Figure 7.1, $\delta_H$ 4.82, ◊) was the second most abundant, comprising about 1% of all identified species in solution. The aldehyde form of HCHO accounted for a minor fraction of the total amount of identified species (Figure 7.1), with a characteristic singlet at $\delta_H$ 9.48 ppm (Figure 7.1, o). These observations are in agreement with previous reports.[23–26] Two additional low intensity resonances in the solution were assigned to formic acid (Figure 7.1, $\delta_H$ 8.37 ppm, ▼) and methanol (Figure 7.1, $\delta_H$ 3.27 ppm, ▪) respectively. These two species are likely to form through Cannizzaro-type disproportionation reactions.[27–29] Integrating the peak areas of the methanol and formic acid resonances indicates methanol was more abundant in solution than formic acid; this observation is probably due to trace methanol present in the paraformaldehyde used to prepare the solution. The methanol present in solution has the potential to react with HCHO to cap HCHO-oligomers. These oligomers are identified by low intensity singlets above the methanol resonance (Figure 7.1, Δ), and their abundance is dependent on the methanol concentration in solution.[25]

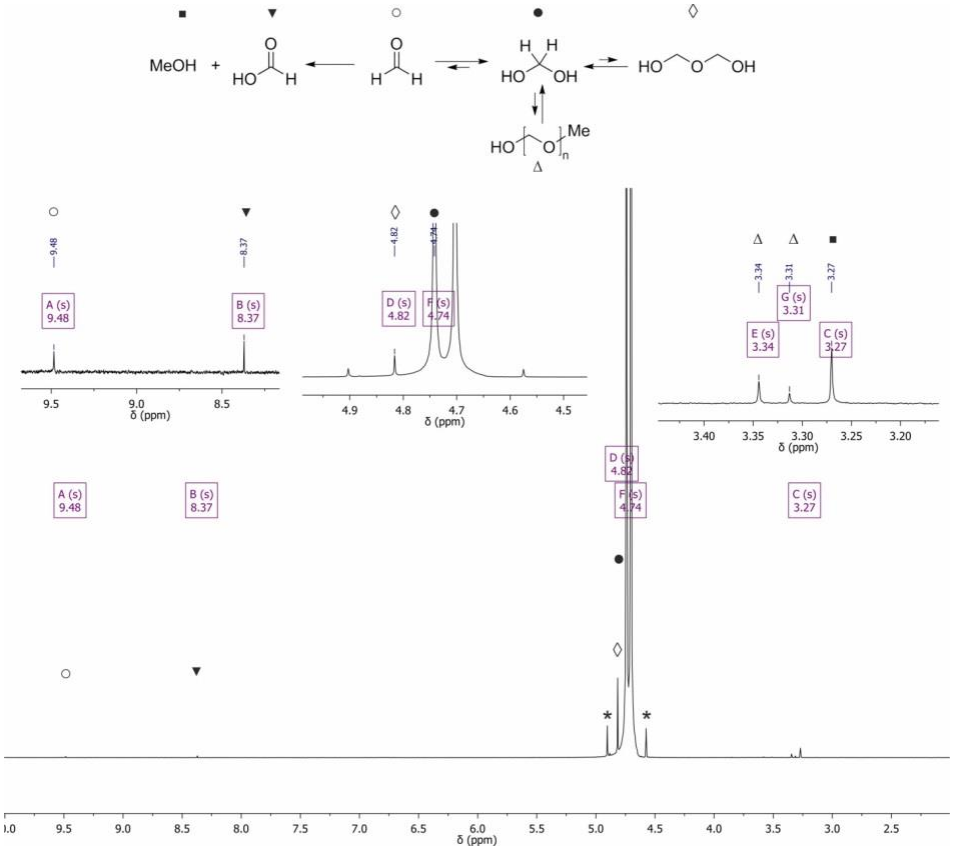



**Figure 7.1:** $^1$H NMR (700 MHz) of a solution of HCHO (75 mM) in $D_2O$. The spectrum is consistent with the presence of: non-hydrated HCHO (o, <0.1%), formic acid (▼, <0.1%), $HOCH_2OCH_2OH$ (◊, 1.3%), hydrated HCHO ($HOCH_2OH$, ~98.4%, ● carbon satellites are marked with *), methanol (▪, ~0.2%), and oligomeric forms of HCHO with methanol (Δ, <0.1%). Approximate percentages were determined by integrating peaks to establish their relative abundances.

## 7.3 Amino acids and formaldehyde – product characterisation

### 7.3.1 Experimental conditions

For the initial investigations on the reactions of amino acids with HCHO, common proteinogenic amino acids were used. This selection was subsequently broadened by the addition of post-translationally modified lysines and common non-proteinogenic amino acids. Thus, the final set of tested amino acids comprised: alanine (Ala), cysteine (Cys), serine (Ser), threonine (Thr), D/L-homocysteine (Hcy), penicillamine (Pen), homoserine (Hse), *allo*-threonine (*allo*-Thr), lysine (Lys), ornithine (Orn), arginine (Arg), histidine (His), tryptophan (Trp), tyrosine (Tyr), asparagine (Asn), glutamine (Gln), $N^{\varepsilon}$-methyllysine (Lys(Me)), $N^{\varepsilon},N^{\varepsilon}$ -dimethyllysine ($Lys(Me)_2$), 5-hydroxylysine (5hLys), and proline (Pro). The standard reaction conditions were: amino acid (4.34 mM), HCHO (10 equivalents) in $D_2O$, at 298 K, monitored over a maximum of 48 hours by $^1H$ NMR (500, 600, 700 MHz). Acid (DCl in $D_2O$) or base (NaOD, in $D_2O$) were added as specified. All amino acids were used in their (L)/(2S)-stereochemical forms, unless stated otherwise.

### 7.3.2 Cysteine and HCHO

The first amino acid to be studied in detail for reaction with HCHO was Cys; this was because of the importance of the reactions of thiols with HCHO in cells.[30] At extended time points the reaction between Cys (4.34 mM) and HCHO (2 equivalents) produced thiazolidine **1** as the apparent 'thermodynamic' product (Figure 7.2), in agreement with previous studies.[31] The structure of product **1** was assigned by through-bond correlations between the HCHO-derived methylene ($\delta_H$ 4.33 and 4.21 ppm) and the β-$CH_2$ ($\delta_H$ 3.29 and 3.18 ppm) as well as by correlations to the α-CH ($\delta_H$ 4.29 ppm, Figure 7.3 (a)). These observations are consistent with previous spectra reported on thiazolidine **1**.[32]

At earlier time points (10 – 30 min), hemithioacetal **2** was observed as an apparent 'kinetic' product. The observation of an HMBC correlation between the HCHO-derived hemithioacetal methylene ($\delta_H$ 4.62 ppm) and the β-$CH_2$ ($\delta_H$ 3.15 ppm) was used to assign the structure as **2** (Figure 7.3 (b)). The lifetime of the kinetic product **2** could be extended by the addition of acid (DCl, 2 equivalents), possibly due to increased protonation of the Cys $N^{\alpha}$-amino group. Elevated temperatures (308 K, 318 K) or the addition of base

(NaOD, 1 equivalent) increased the rate of thiazolidine **1** formation. Potential loss of stereochemical purity under the basic or acidic conditions employed was investigated by quantitative $^{1}H$ NMR experiments in $D_2O$; no incorporation of $^{2}H$ at the α-position was observed, suggesting that the stereochemistry of Cys and His is maintained (Appendix Figure 1).

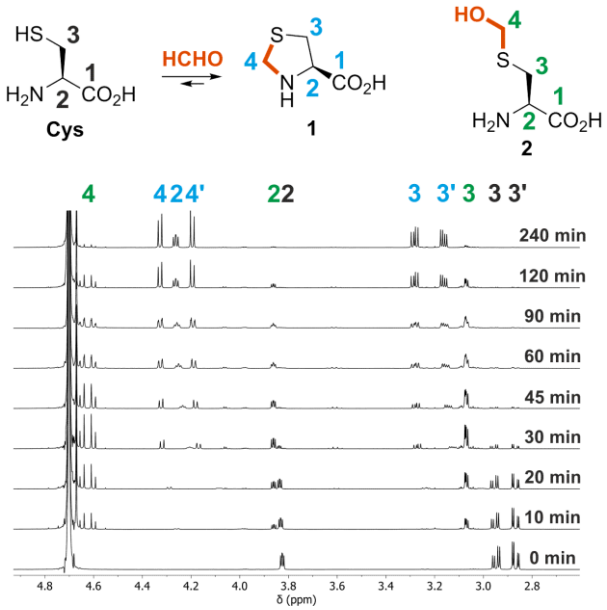



**Figure 7.2:** The reaction of cysteine (4.34 mM) and HCHO (2 equivalents), forming thiazolidine **1** and hemithioacetal **2**, (followed over 4 hours by $^{1}H$ NMR at 700 MHz).

In the presence of $O_2$ or an equivalent oxidising reagent, Cys is prone to oxidation to its dimer cystine.[33] The cystine product can be monitored by the formation of a new set of β-$CH_2$ resonances ($\delta_H$ 3.24 and 3.04 ppm). Cystine formation during the reaction of Cys with HCHO was not observed under standard conditions, likely due to the rapid hydroxymethylation of the Cys-thiol, preventing dimerization.

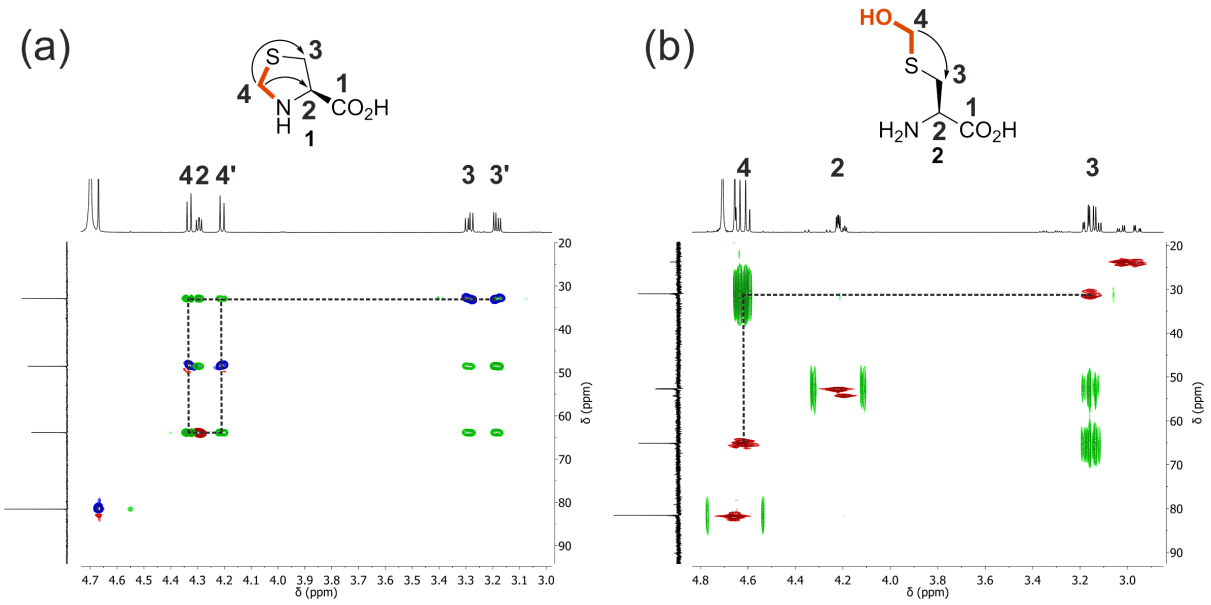



**Figure 7.3:** HSQC (red/blue) and HMBC (green) correlations of thiazolidine **1** and hydroxymethyl adduct **2**, formed in reactions between cysteine and HCHO under standard (a) and acidic (b) conditions. (a) The reaction between cysteine and HCHO after 6 hours, with the complementary $^{1}H$ and $^{13}C$ NMR spectra plotted on the respective axes; (b) the reaction between cysteine and HCHO in presence of acid (DCl, 4 equivalent) after 4 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.3 Homocysteine and HCHO

When reacting Hcy with HCHO under standard conditions, a cyclic product, thiazinane **3**, was formed as the apparent thermodynamic product (Figure 7.4 (a)), in an analogous manner to the reaction with Cys. The structure of thiazinane **3** was assigned by through-bond HMBC correlations from the HCHO-derived $CH_2$ ($\delta_H$ 4.16 and 4.11 ppm) to the α-CH ($\delta_H$ 3.52 ppm) and γ-$CH_2$ ($\delta_H$ 2.85 and 2.69 ppm) (Figure 7.4 (a)). These observations are in agreement with previously reported data for the reaction of Hcy with HCHO.[34,35]

Similar to the observations with Cys, an apparent kinetic product could also be identified in the reaction between Hcy and HCHO. Hydroxymethylation of the Hcy thiol resulted in the formation of hemithioacetal **4**. The structure of **4** was confirmed by through-bond correlations between the hemithioacetal-$CH_2$ ($\delta_H$ 4.67 ppm) and γ-$CH_2$ ($\delta_H$ 2.75 ppm), and the lack of coupling to the α-CH ($\delta_H$ 4.15 ppm). The chemical shifts of the HCHO-derived hemithioacetal and bridging $CH_2$ are comparable to the shifts corresponding to the equivalent positions in **1** and **2**.

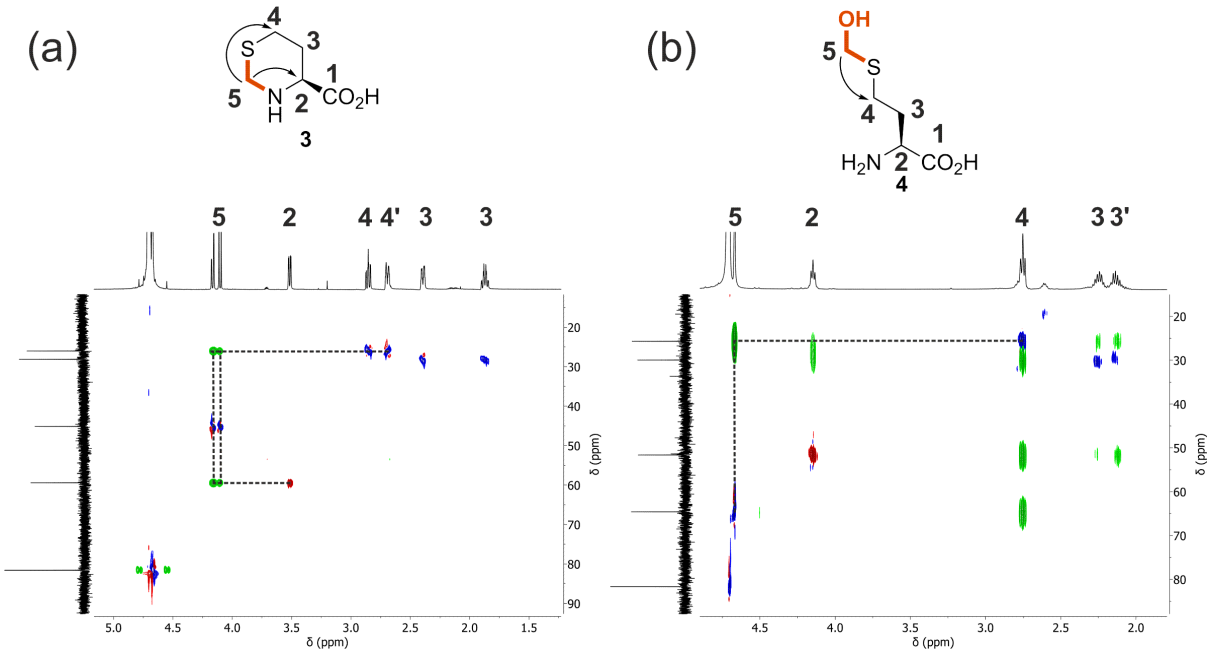



**Figure 7.4:** HSQC (red/blue) and HMBC (green) correlations of thiazinane **3** and hydroxymethyl adduct **4**, formed by a reaction of homocysteine and HCHO under standard (a) and acidic (b) conditions. (a) The reaction between homocysteine and HCHO after 6 hours, with the complementary $^{1}H$ and $^{13}C$ NMR spectra plotted on the respective axes; (b) reaction between homocysteine and HCHO in presence of acid (DCl, 4 equivalents) after 4 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

In presence of $O_2$ or other oxidising conditions, Hcy is prone to disulfide formation. Disulfide formation was not observed during the reaction, i. e. under the standard conditions employed.

### 7.3.4 Penicillamine and HCHO

Penicillamine (Pen), a Cys analogue dimethylated at the β-position, reacted to produce thiazolidine **5** as the thermodynamic product, and hydroxymethylated **6** as the kinetic product, in an analogous manner to the products observed with Cys. Structural assignment for thiazolidine **5** was achieved by through-bond HMBC correlations (Figure 7.5 (a)) between methylene-5 ($\delta_H$ 4.44 and 4.36 ppm) and both the α-CH ($\delta_H$ 3.91 ppm) and the quartenary carbon-3 ($\delta_C$ 52.9 ppm), which is in agreement with literature reports.[36] HMBC couplings between methylene-5 ($\delta_H$ 4.87 and 4.75 ppm) and the quaternary carbon-3 ($\delta_C$ 46.2 ppm), and a lack of coupling to the α-CH ($\delta_H$ 4.07 ppm), supported the assignment of **6** as *S*-hydroxymethylated Pen.

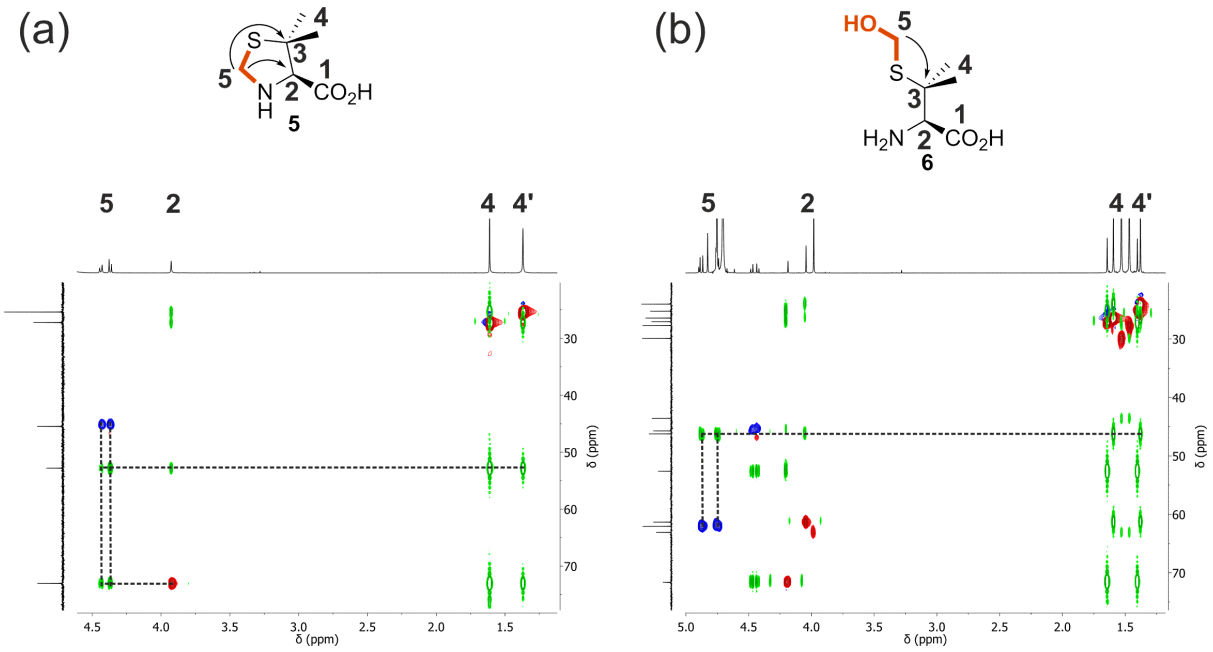



**Figure 7.5:** HSQC (red/blue) and HMBC (green) correlations of thiazolidine **5** (a) and hydroxymethyl adduct **6** (b), formed by a reaction of penicillamine and HCHO. (a) Reaction between homocysteine and HCHO after 3 hours, with the complementary $^{1}H$ and $^{13}C$ NMR spectra plotted on the respective axes; (b) reaction between penicillamine and HCHO in presence of acid (DCl, 2 equivalents) after 3 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

The chemical shifts of products **5** and **6** are in agreement with the products of the reaction between Cys / Hcy and HCHO. Under aerobic conditions, Pen is prone to disulfide formation. During the reaction between Pen and HCHO under standard aerobic conditions, no disulfide formation was observed.

### 7.3.5 Serine and formaldehyde

In contrast to work with the thiol-containing amino acids, the reaction of Ser with HCHO under standard conditions did not yield a clearly defined product. Addition of base (NaOD, 1 equivalent) to the Ser-HCHO mixture revealed new broad $^{1}H$ resonances (Figure 7.6). The observation of broadened signals ($\delta_H$ 4.49, 3.99, 3.73 and 3.54 ppm) implied the formation of a product or products that may be in dynamic exchange. These signals were altered by changing the temperature of the mixture. At 283 K, the new resonances were well defined, and enabled structural assignment of **7**. Through-bond HMBC correlations from methylene-4 ($\delta_H$ 4.49 and 3.99 ppm) to α-CH ($\delta_H$ 3.54 ppm) and β-methylene ($\delta_H$ 3.73 and 3.38 ppm) enabled assignment of the structure as oxazolidine **7** (Figure 7.7).

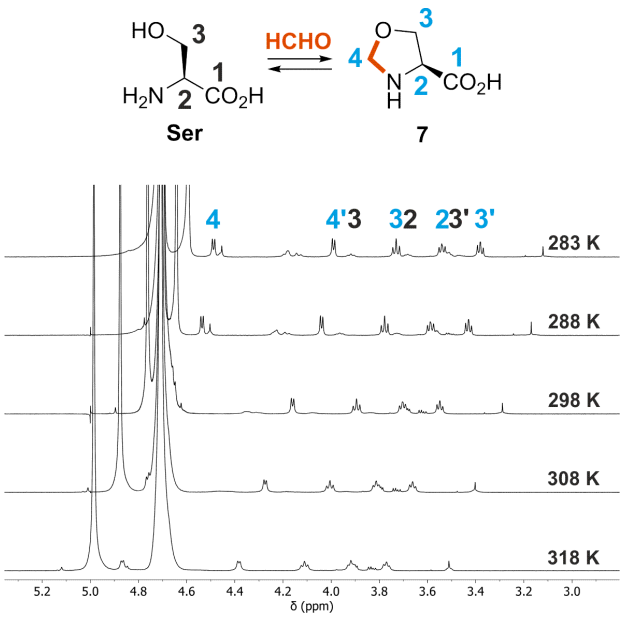



**Figure 7.6:** $^1$H NMR (600 MHz) analyses of the equilibrium between Ser (4.34 mM) and HCHO (10 equivalents), and their condensed product, oxazolidine **7**. The spectra were recorded at different temperatures, which appears to alter the rate of exchange. The equilibrium shifts towards the oxazolidine at lower temperatures.

Oxazolidine **7** formed slowly relative to its structural analogue thiazolidine **1**. The difference in rates between thiazolidine and oxazolidine formation has previously been observed,[37,38] and correlates with the difference in $pK_a$ of the serinyl alcohol group (predicted 13.4[39]) and cysteinyl thiol group (~8.4[40]). Deprotonation of the side chain alcohol might therefore be, at least in part, rate-limiting. Additional broadened resonances ($\delta_H$ 4.47 and 4.20 ppm) were observed in the spectra of the Ser-HCHO mixture; however, the lack of couplings observed in the HMBC spectra precluded assignment of these resonances.

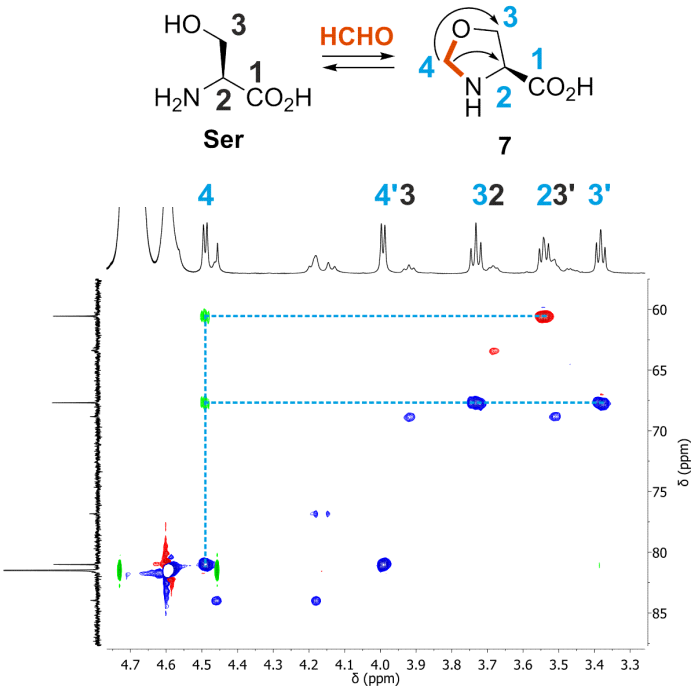



**Figure 7.7:** HSQC (red/blue) and HMBC (green) correlations of oxazolidine **7**, formed by a reaction of serine and HCHO under standard conditions with base (NaOD, 1 equivalent), cooled to 283 K, after one hour. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.6 Threonine and HCHO

In contrast to Ser, the reaction of Thr under standard conditions with HCHO, produced defined resonances ($\delta_H$ 5.01 and 4.64 ppm, Figure 7.8). Through-bond HMBC correlations between the HCHO-derived bridging methylene ($\delta_H$ 5.01 and 4.64 ppm), and both C3 ($\delta_H$ 3.71 ppm) and C2 ($\delta_H$ 4.16 ppm), enabled assignment of the product as oxazolidine **8**. The chemical shifts of the resonances assigned to the C5-methylene bridge ($\delta_H$ 5.01 and 4.64 ppm, $\delta_C$ 77.3 ppm), were comparable with the shifts assigned to the Ser-derived **7**, but were considerably more deshielded in comparison with the equivalent methylene resonances for thiazolidines **1**, **3**, and **5** (Figures 7.2 – 7.5). Formation of oxazolidine **8** was observed under the standard conditions; however, addition of base (NaOD, 1 equivalent) increased the extent of formation of **8**. The subtle difference between the threonyl secondary alcohol and the serinyl primary alcohol apparently promotes the formation of **8**. The methyl group present in Thr could potentially give rise to the Thorpe-Ingold effect,[41] facilitating the formation of the methylated oxazolidine ring. Alternatively, the methyl group in Thr provides an

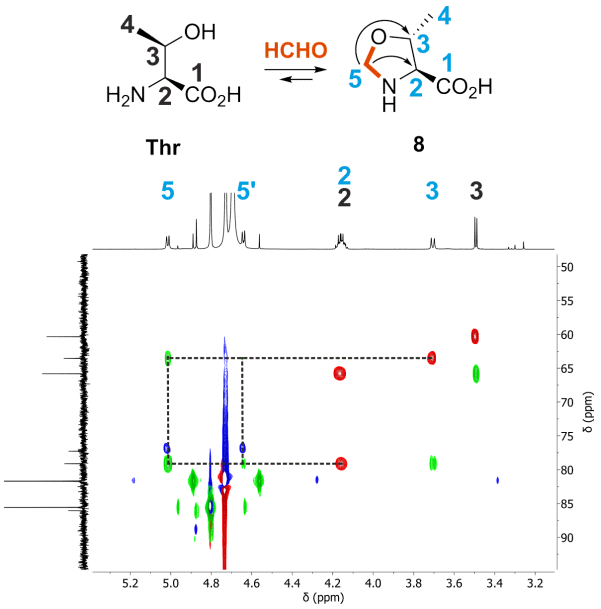



**Figure 7.8:** HSQC (red/blue) and HMBC (green) correlations of oxazolidine **8**, formed by a reaction of threonine and HCHO under standard conditions, recorded after 2 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

inductive effect which might affect the nucleophilicity of alcohol on Thr. A difference in oxazolidine product stability might also favour the formation of **8** over **7** under the standard conditions.

### 7.3.7 *allo*-Threonine and HCHO

The epimer of Thr, i.e. *allo*-Thr, also reacted with HCHO under standard conditions to form an oxazolidine product, i.e. **9** (Figure 7.9). The bridging methylene ($\delta_H$ 5.02 and 4.51 ppm, $\delta_C$ 76.7 ppm) was observed to couple to the α-CH ($\delta_H$ 4.09 ppm) and C3 ($\delta_H$ 4.39 ppm, and $\delta_C$ 76.8 ppm) in HMBC spectra, suggesting a cyclic oxazolidine structure analogous to that formed with Thr (i.e. **8**). Formation of an oxazolidine product under standard conditions was less efficient for *allo*-Thr compared to Thr; this may reflect the *cis*-substitution of the C2 carboxylate and C3 methyl in **9**, compared to **8**, where they are in a *trans*-configuration. The addition of base (NaOD, 1 equivalent) increased the amount of cyclic product formed.

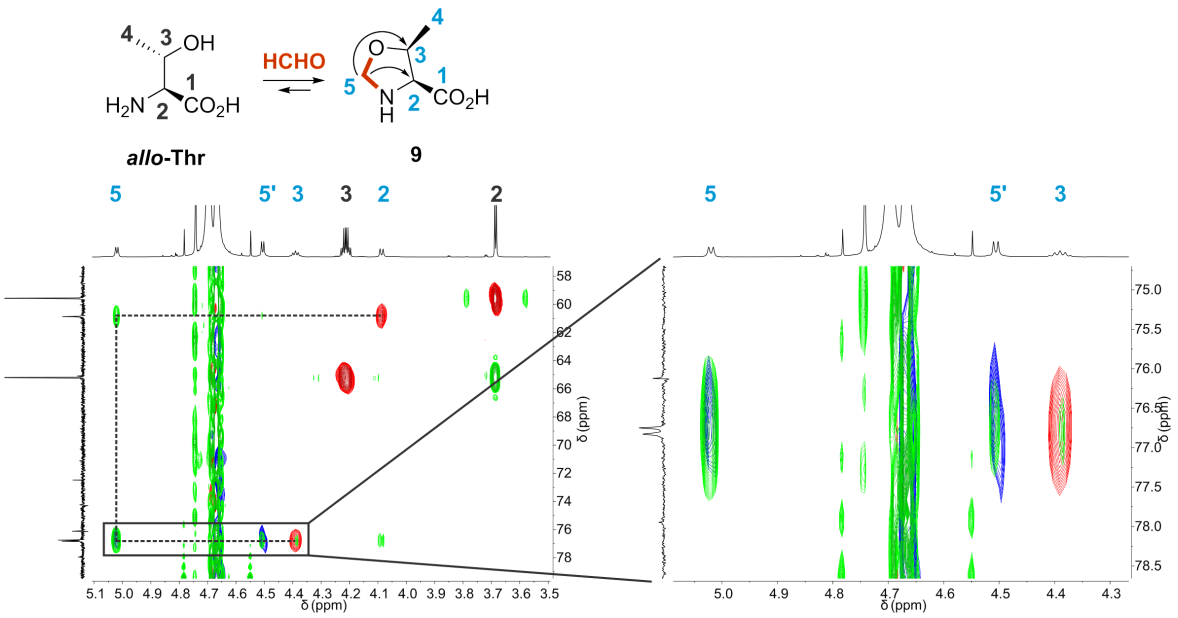



**Figure 7.9:** HSQC (red/blue) and HMBC (green) correlations of oxazolidine **9**, formed by a reaction of *allo*-threonine and HCHO under standard conditions, recorded after 1 hour. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.8 Homoserine and HCHO

Under standard conditions, Hse reacted with HCHO to form oxazinane **10** (Figure 7.10). The structure of oxazinane **10** was assigned using through-bond HMBC correlations. The familiar deshielded bridging methylene ($\delta_H$ 4.86 and 4.41 ppm, $\delta_C$ 75.3 ppm) was observed to couple to both the α-CH ($\delta_H$ 3.76 ppm), and C4 ($\delta_H$ 4.02, and 3.62 ppm, $\delta_C$ 66.9 ppm). The chemical shifts were comparable to those of resonances corresponding to oxazolidines **7**, **8**, and **9**, suggesting a cyclic oxazinane structure. When compared to the oxazolidine analogues (**8**, **9**), oxazinane **10** is more efficiently formed, possibly reflecting an increased stability of the oxazinane compared to the oxazolidine ring structure. Addition of base (NaOD, 1 equivalent) increased the formation of oxazinane **10**.

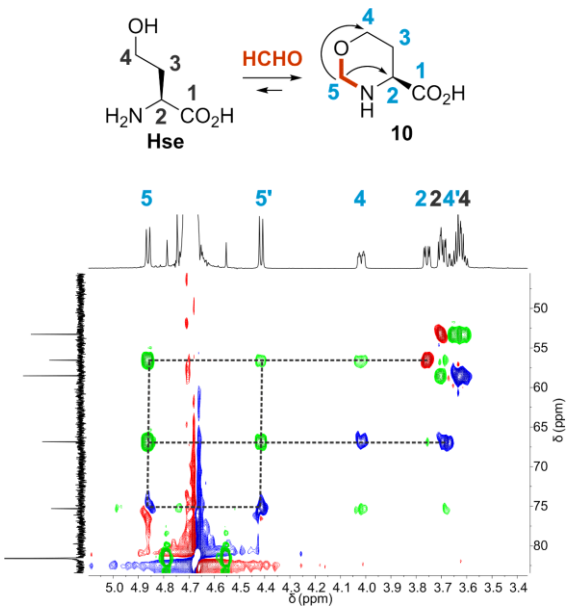



**Figure 7.10:** HSQC (red/blue) and HMBC (green) correlations of oxazinane **10**, formed by a reaction of homoserine and HCHO under standard conditions, after 1 hour. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.9 Asparagine and HCHO

Exposing Asn to HCHO under standard conditions gave new resonances ($\delta_H$ 4.55 and 4.43 ppm). Through-bond HMBC correlations enabled assignment of the structure as lactam **11**. Correlations from the HCHO-derived methylene ($\delta_H$ 4.55 and 4.43 ppm) to both the α-CH ($\delta_H$ 4.05 ppm) and the lactam carbonyl ($\delta_C$ 172.1 ppm), as well as couplings from methylene-3 ($\delta_H$ 2.86 and 2.64 ppm) to the lactam carbonyl, confirmed formation of lactam **11**. These results are in agreement with previous reports on the reaction of Asn with HCHO in potato tissue.[42] The rate of lactam **11** formation was significantly slower than the reactions of HCHO with the thiol and alcohol containing amino acids; lactam **11** was formed over 48 hours. The reaction was accelerated by the addition of base (NaOD, 1 equivalent), increasing the temperature, or increasing the concentration. When performing the same reaction at higher Asn (43.4 mM) and HCHO (10 equivalents) concentrations, HCHO adducts **12** and **13** were also formed. Modification of the amide nitrogen appears likely to occur prior to modification of the α-amino group. Through bond correlations of the hydroxymethyl group in **12** to the amide carbonyl ($\delta_C$ 171.2 ppm), and bridging methylene-5 ($\delta_H$ 4.30 and 4.21 ppm), and the deshielding of the modification itself ($\delta_C$ 66.9 ppm), suggested hemiaminal formation on the amide nitrogen of **12** (Figure 7.11 (b)).

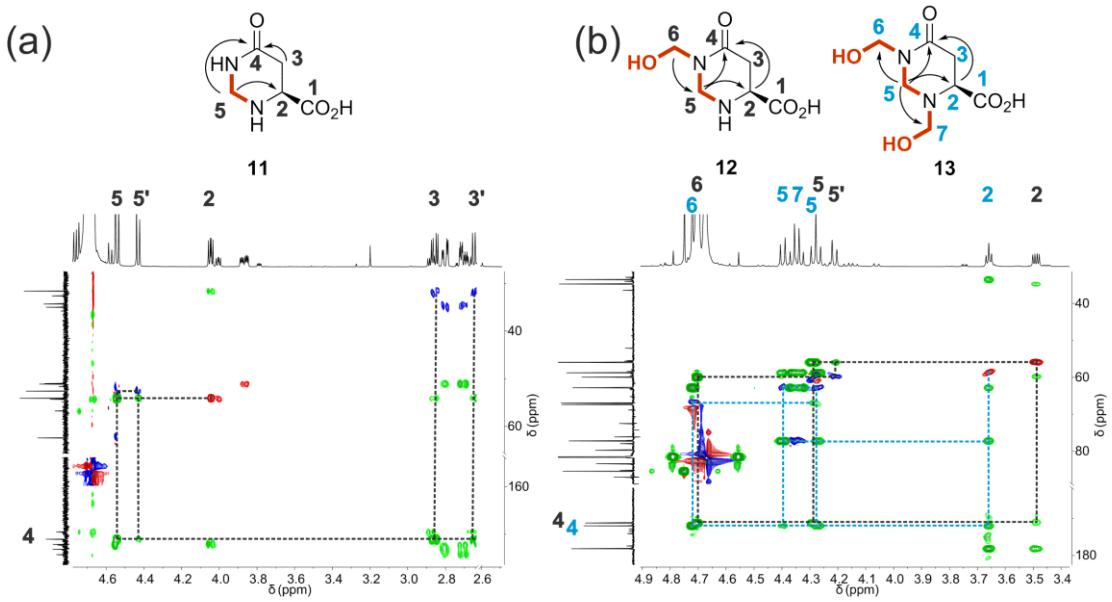



**Figure 7.11:** HSQC (red/blue) and HMBC (green) correlations of lactams **11**, **12**, and **13**, formed by a reaction of asparagine and HCHO under standard conditions after 12 hours (a), and with concentrated asparagine (43.4 mM) after 48 hours (b). Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Hydroxymethylation on the α-amino group as observed in lactam **13**, was assigned using correlations from a new resonance ($\delta_H$ 4.35 ppm, $\delta_C$ 77.2 ppm) to α-CH ($\delta_H$ 3.66 ppm), and the bridging methylene-5 ($\delta_H$ 4.40 and 4.27 ppm). Both lactams **12** and **13** species accrued over 48 hours at a slower rate than the formation of lactam **11**.

### 7.3.10 Glutamine and HCHO

Under standard conditions, Gln also reacted with HCHO, giving new resonances ($\delta_H$ 4.59 and 4.51 ppm). Chemical shift analysis and through-bond HMBC correlations suggested formation of the lactam **14**. HMBC correlations between the bridging methylene ($\delta_H$ 4.59 and 4.51 ppm) and both the amide carbonyl ($\delta_C$ 180.3 ppm) and α-CH, suggest formation of lactam **14**. These results are in agreement with the previously reported shifts for lactam **14**.[42] Compared to the rate of lactam **11** formation, lactam **14** formed less efficiently, reaching 63% conversion within 48 hours. Although the addition of base (NaOD, 1 equivalent) or increasing the concentration of the reaction, did accelerate the formation of lactam **14**, no further lactam adducts were detected.

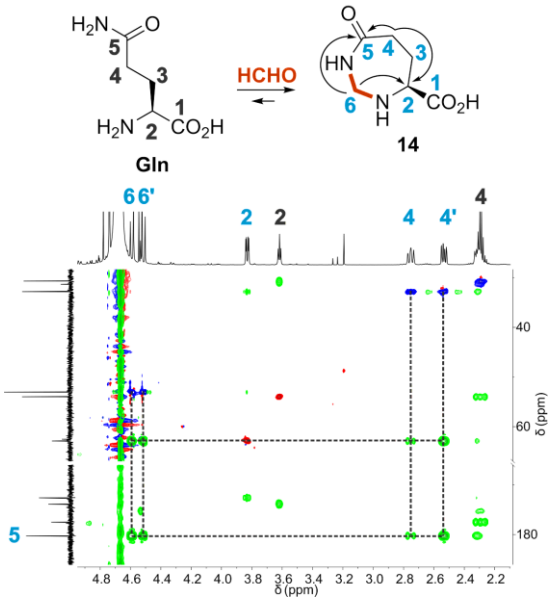



**Figure 7.12:** HSQC (red/blue) and HMBC (green) correlations of amide **14**, formed by a reaction of glutamine and HCHO under standard conditions, recorded after 48 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.11 Histidine and formaldehyde

Exposing His to HCHO under standard conditions resulted in the rapid formation of a new product within 20 minutes (Figure 7.13). Through-bond HMBC correlations and chemical shift comparison suggested the product to be the cyclic imidazole **15**. The new deshielded doublets ($\delta_H$ 5.31 and 4.96 ppm) coupled to a quaternary carbon in the aromatic region (C4, $\delta_C$ 128.7 ppm) and to C6 ($\delta_H$ 8.42 ppm), as well as the α-CH ($\delta_H$ 3.61 ppm). Further hydroxymethylation of this structure was observed when exposing His at high concentrations (43.4 mM) to HCHO (10 equivalents), forming hemiaminal **16** (Figure 7.13). HMBC correlations from the new hemiaminal ($\delta_H$ 4.42 ppm) to the bridging methylene ($\delta_H$ 5.20 ppm) and α-CH ($\delta_H$ 3.94 ppm), implied the modification occurs on the α-amino group. Interestingly, no modification of the τ-nitrogen of the imidazole ring of His was observed.

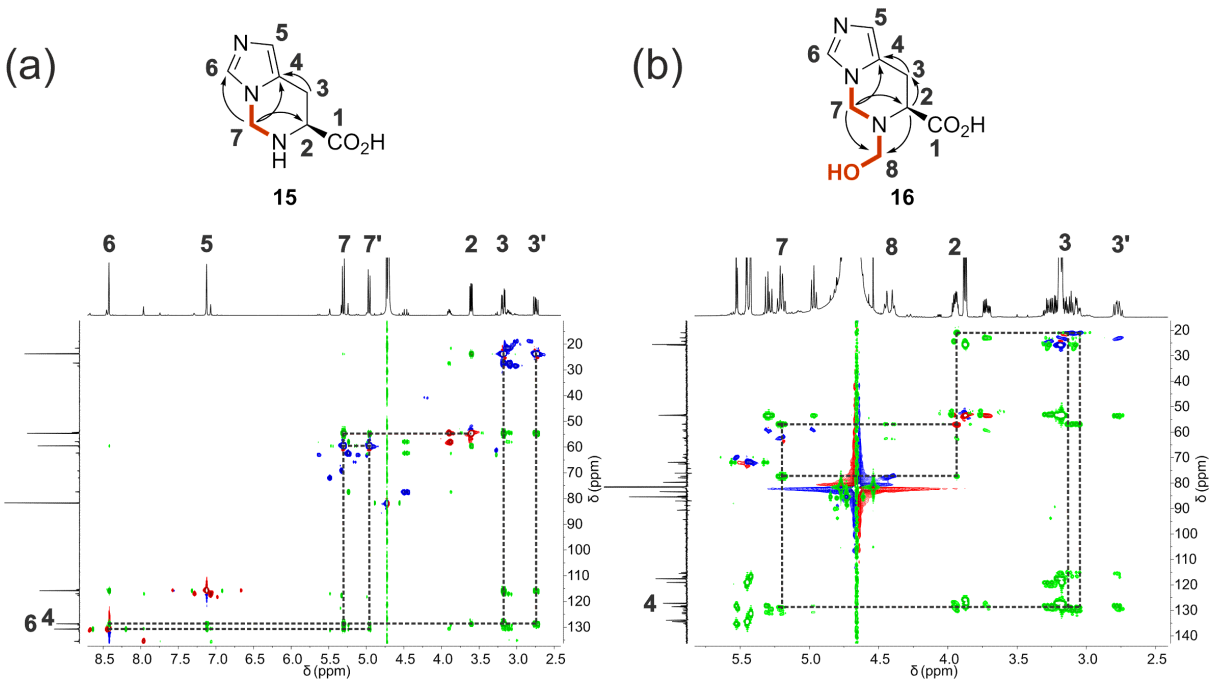



**Figure 7.13:** HSQC (red/blue) and HMBC (green) correlations of imidazoles **15** and **16** formed by reaction of histidine and HCHO (a) under standard conditions after 48 hours, and (b) with more concentrated histidine (43.4 mM) after 48 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Higher concentrations of His gave rise to additional products, i.e. modified histidines **17** and **18**. Through-bond correlations from the new resonances ($\delta_H$ 5.53, 5.52, 5.46, and 5.45 ppm) were observed to the imidazole ring. Product **18**, which was the more abundant product, gave clear correlations to the two protiated centres at C5 ($\delta_H$ 7.37 and 7.22 ppm), and C6 ($\delta_H$ 8.68 and 8.65 ppm). His **17**, which was less abundant, showed correlations between the new hemiaminal ($\delta_H$ 5.53 and 5.52) and both the C6 hydrogens ($\delta_H$ 8.94 ppm and 8.73 ppm) and the quaternary C4 ($\delta_C$ 128.5 and 131.3 ppm). Steric constraints imposed by the hydroxymethyl group could potentially play a role, in producing two identical sets of connectivity.

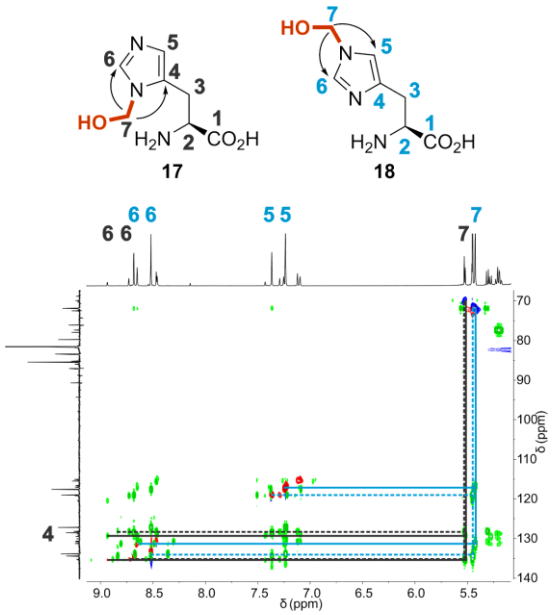



**Figure 7.14:** HSQC (red/blue) and HMBC (green) correlations of products **17** and **18** formed by a reaction of histidine (43.4 mM) and HCHO under concentrated conditions, recorded after 5 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Further incubation of the high concentration mixture of His (43.4 mM) with HCHO (>12 hours) led to formation of imidazoles **19** and **20**, likely via a Pictet-Spengler-type reaction.[43] Through-bond correlations in **19** coupled new methylene resonances ($\delta_H$ 4.20 ppm) to the α-CH ($\delta_H$ 4.00 ppm) and to two quaternary resonances ($\delta_C$ 123.5 and 136.7 ppm, C4 and C6 respectively). Moreover, the carbon-13 chemical shift for the bridging methylene was considerably more shielded ($\delta_C$ 40.8 ppm) compared to those found for **15** and **16** ($\delta_C$ 59.5 and 62.3 ppm respectively). These combined data suggest *C*,*N*-cyclisation occurs, forming **19** during. Similar chemical shifts and correlations were observed for **20** (Figure 7.15), with an additional hemiaminal ($\delta_H$ 5.35 ppm) correlating to CH-5 ($\delta_H$ 7.71 ppm) and quaternary C4 ($\delta_C$ 122.6 ppm) on the imidazole ring. Chemical shifts found for the *C*,*N*-cyclised products **19** and **20** are in agreement with prior reports.[44] When compared to the *N*,*N*-cyclised products, the *C*,*N*-cyclised products formed less efficiently.

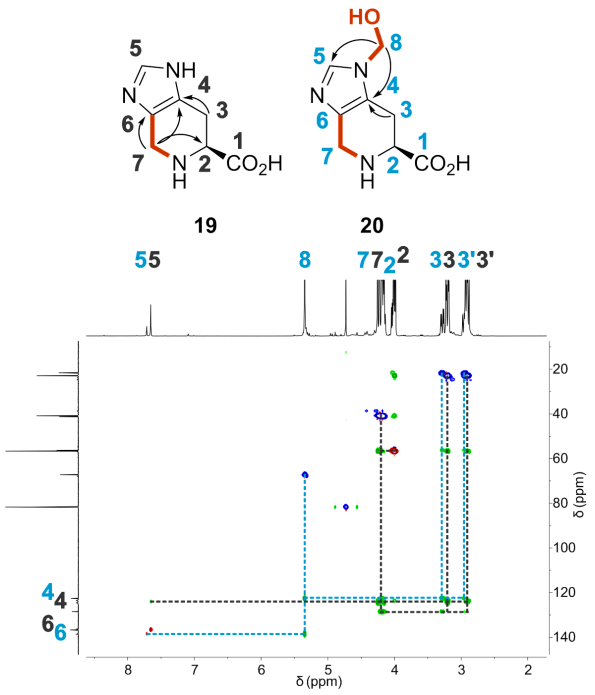



**Figure 7.15:** HSQC (red/blue) and HMBC (green) correlations of amine **19** and **20** formed by a reaction of glutamine (43.4 mM) and HCHO under more concentrated conditions, recorded after 21 days. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.12 Tryptophan and HCHO

When Trp was reacted with HCHO under standard conditions, new resonances were observed within 1 hour. A new singlet resonance ($\delta_H$ 5.57 ppm) was coupled to quaternary C6 ($\delta_C$ 136.0 ppm) and CH-5 ($\delta_H$ 7.25 ppm) resonances, suggesting hydroxymethylation of the indole nitrogen; the product was therefore assigned as Trp adduct **21** (Figure 7.16 (a)). Alongside **21**, two other products were observed. New resonances ($\delta_H$ 4.46 and 4.38 ppm) coupled to two quaternary carbons, C4 ($\delta_C$ 109.8 ppm) and C5 ($\delta_C$ 126.3 ppm) resonances, as well as to the α-CH ($\delta_H$ 3.98 ppm); these resonances were therefore assigned to the cyclic Trp **22** (Figure 7.16 (b)). Cyclisation of Trp is probably achieved via a Pictet-Spengler-type reaction. Further hydroxymethylation of **22** to give **23** was observed, as evidenced by the observation of a new AB system ($\delta_H$ 5.50 ppm). This AB system likely reflects hindrance of free rotation of the hydroxymethyl group in **23**. This is not observed in the case of **21**, as rotational entropy allows for indistinguishable coupling, whereas with **23** the imidazole ring system is locked to the α-amine, causing inequivalent coupling.

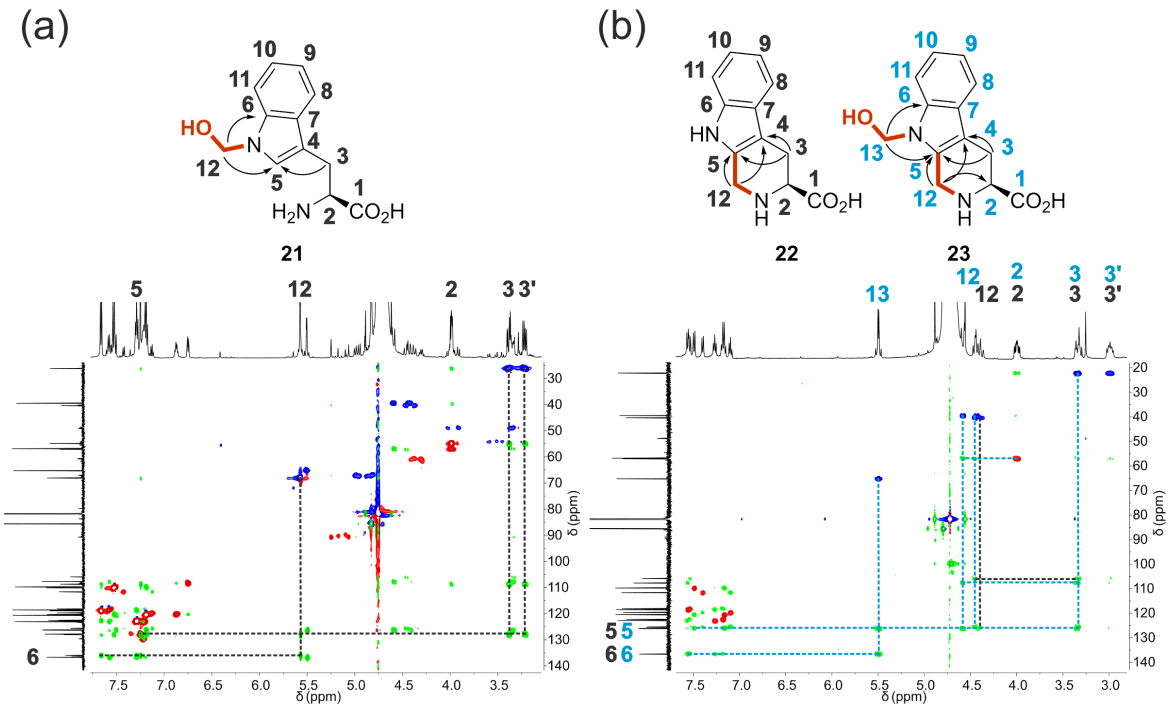



**Figure 7.16:** HSQC (red/blue) and HMBC (green) correlations of indoles **15**, **16** and **17** formed by a reaction of tryptophan and HCHO (a) under standard conditions after 2 hours, and (b) with more concentrated tryptophan (43.4 mM) after 7 days. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.13 Ornithine and HCHO

Reactions between Orn and HCHO were only observed in presence of added base (NaOD, 1 equivalent); upon addition of base, rapid formation of a single species was observed (Figure 7.17). Through-bond couplings from the bridging methylene ($\delta_H$ 4.14 and 3.87 ppm, $\delta_C$ 62.7 ppm) to the α-CH ($\delta_H$ 3.37 ppm) and δ-$CH_2$ ($\delta_H$ 2.97 and 2.65 ppm) suggested a cyclic structure. The two new deshielded methylenes ($\delta_H$ 4.81 and 4.49 ppm, and $\delta_H$ 4.75 and 4.37 ppm) both coupled to the bridging methylene. Methylene-8 ($\delta_H$ 4.81 and 4.49 ppm) was observed to couple to the δ-methylene, methylene-7 ($\delta_H$ 4.75 and 4.36 ppm), and to the α-CH. The carbon-13 chemical shifts for methylenes-7 and 8 ($\delta_C$ 84.9 and 86.2 ppm respectively) are significantly deshielded, in agreement with suggested structure **24**. Moreover, the bridging methylene-6 ($\delta_C$ 62.7 ppm) is in a similar chemical shift region as the bridging methylenes in the His derivatives **17** and **18**. The combined data is in agreement with the proposed structure **24**.

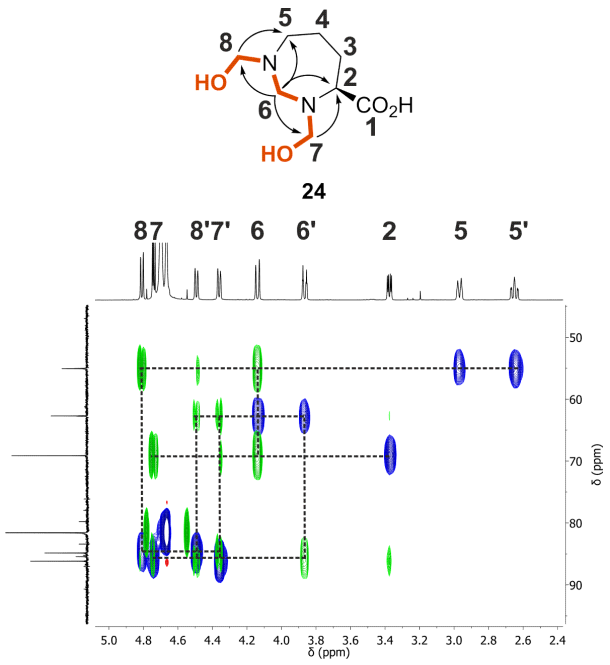



**Figure 7.17:** HSQC (red/blue) and HMBC (green) correlations of diazepane **24** formed by a reaction of ornithine and HCHO under alkaline standard conditions (NaOD, 1 equivalent), after 5 minutes. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Prolonged incubation of the solution containing **24** (>6 hours) gave rise to the formation of what appeared to be 4 new species. Further investigation showed these were two structurally related compounds and their corresponding formyl conformational isomers (rotamers). The two methylene-6 resonances ($\delta_H$ 5.14 and 4.33 ppm, and 61.2 ppm; $\delta_H$ 4.75 and 4.62 ppm, and $\delta_C$ 65.9 ppm) coupled both to the α-CH resonances ($\delta_H$ 4.28 and 4.27 ppm), and to their respective δ-methylene resonances ($\delta_H$ 3.36 and 2.97 ppm; $\delta_H$ 3.21 and 2.86 ppm). The two methyl groups ($\delta_H$ 2.80 and 2.61 ppm) were clearly identified by HSQC, and were found to couple to their respective δ-methylene resonances. The shielded formyl peaks were readily identified, and exhibited couplings to the bridging methylene resonances and the α-CH resonances. The carbon-13 chemical shifts observed for the methyl groups ($\delta_C$ 39.0 and 40.3 ppm), and formyl groups ($\delta_C$ 167.2 and 166.7 ppm, respectively), are in the expected chemical shift range. These combined data enabled assignment of the four sets of observed species to the same product, diazepane **25**.

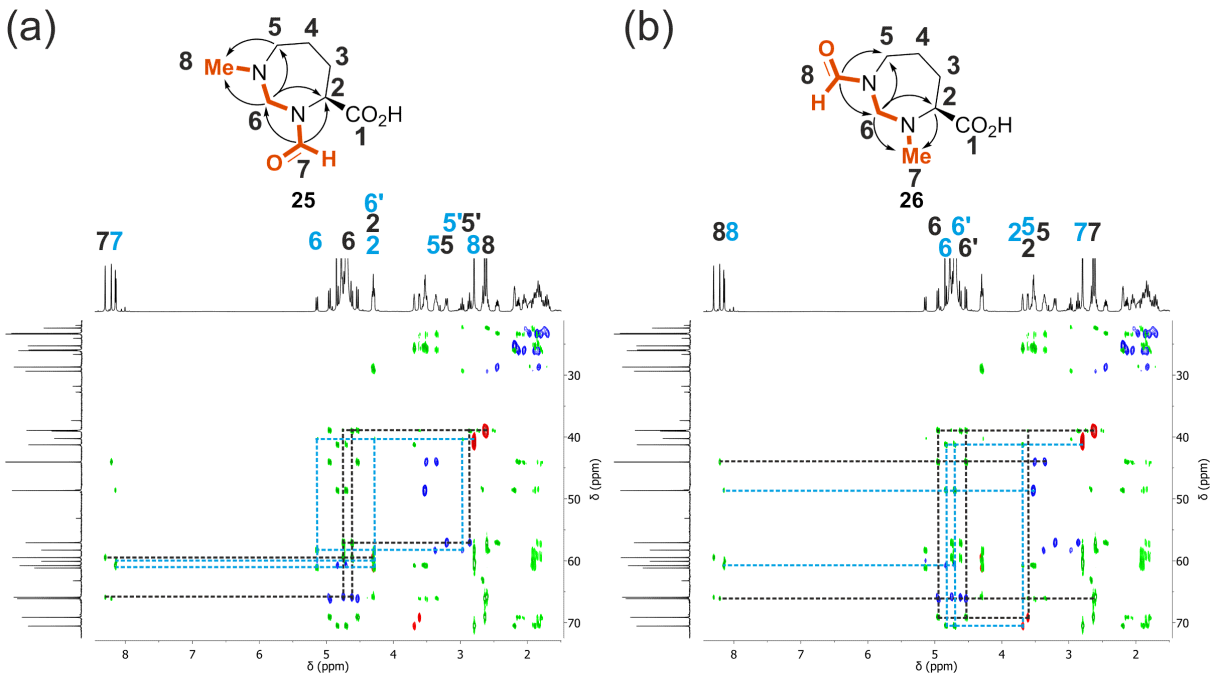



**Figure 7.18:** HSQC (red/blue) and HMBC (green) correlations of diazepanes **25** and **26** formed by a reaction of ornithine and HCHO under alkaline standard conditions (NaOD, 1 equivalent), after 48 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

The remaining two sets of signals, were similarly found to derive from two formyl conformational isomers of the same product (Figure 7.18 (b)), a regioisomer of substituted diazepane **25**. Similar through bond correlations from their bridging methylene resonances ($\delta_H$ 5.14 and 4.29 ppm; $\delta_H$ 4.74 and 4.62 ppm) to the α-CH resonances ($\delta_H$ 4.30 and 4.30 ppm) and δ-methylene resonances ($\delta_H$ 3.53 ppm; $\delta_H$ 3.51 and 3.36 ppm) were observed. By contrast with the correlations observed for **25**, these two species exhibited couplings from their methyl resonances ($\delta_H$ 2.80 and 2.61 ppm) to their respective bridging methylene resonances and α-CH resonances ($\delta_H$ 3.69 and 3.62 ppm). These data suggest the two species correspond to diazepane **26**. The carbon-13 chemical shifts of either α-CH or δ-methylene differ considerably, depending on the presence of a methyl or a formyl group on their neighbouring amine. Notably, no double formylated or methylated products were observed in the mixture, suggesting intramolecular disproportionation yields **25** and **26**. Product formation is likely achieved *via* an internal hydride transfer, as suggested by Dale and Sigvartsen in their study investigating cyclisation of simpler structurally related diamines.[45]

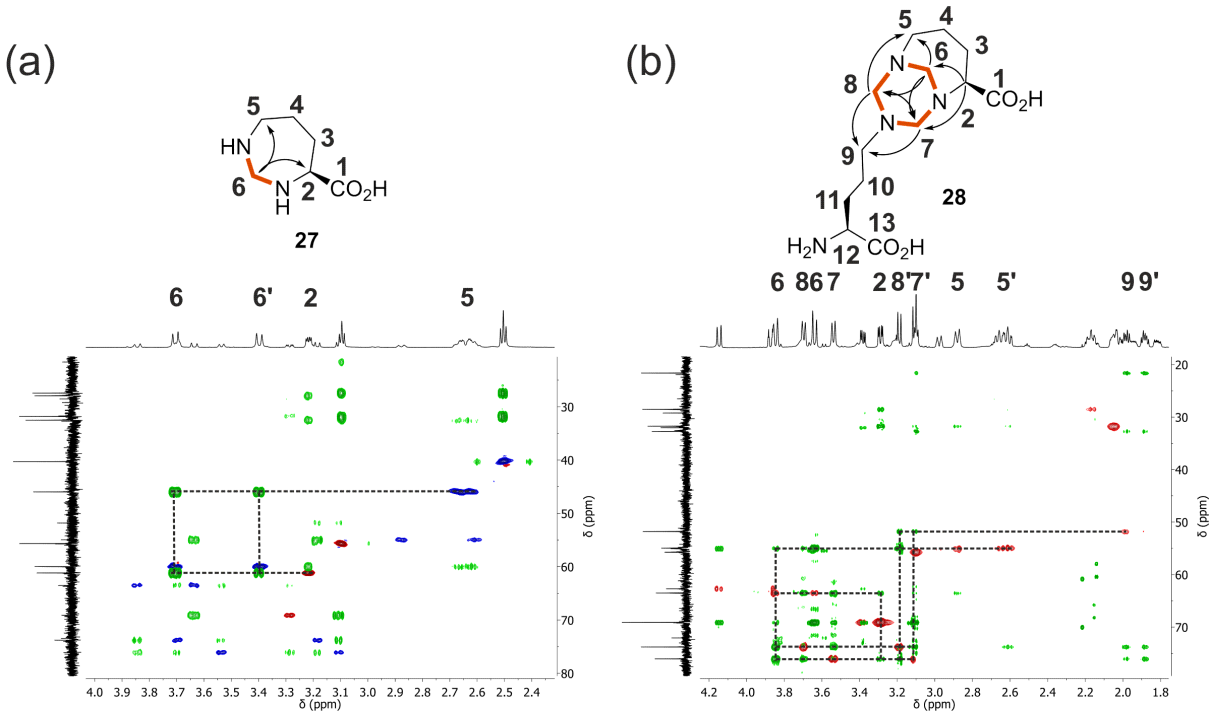



**Figure 7.19:** HSQC (red/blue) and HMBC (green) correlations of diazepanes **27** and **28** formed by a reaction of ornithine and HCHO (a) at low concentrations of ornithine (4.34 mM) and stoichiometric amounts of HCHO, after 10 minutes, and (b) at high concentrations of ornithine (43.4 mM) and stoichiometric amounts of HCHO, after 10 minutes. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Limiting the excess of HCHO to stoichiometric amounts with respect to Orn (4.34 mM), resulted in the detection of a new diazepane structure **27** (56%) alongside a minor product **28** and remaining Orn (Figure 7.19 (a)). Similar to diazepanes **24** – **26**, HMBC correlations from bridging methylene-6 resonances ($\delta_H$ 3.71 and 3.40 ppm) to the α-CH ($\delta_H$ 3.22 ppm) resonance and δ-methylene resonance ($\delta_H$ 2.65 ppm), enabled assignment of the structure as diazepane **27**, which is a possible intermediate in the formation of **24**. Performing the reaction between Orn (43.4 mM) and HCHO (1 equivalent) under more concentrated conditions allowed for the detection of dimeric aminal **28** (27%) as well as diazepane **27** (31%) and unreacted Orn (Figure 7.19 (b)). Through-bond correlations from bridging methylene-6 ($\delta_H$ 3.85 and 3.54 ppm) to the α-CH ($\delta_H$ 3.29 ppm) and methylene-5 ($\delta_H$ 2.88 and 2.62 ppm), confirmed the familiar cyclic structure also present in **24** – **27**. The two remaining bridging methylenes-7 ($\delta_H$ 3.54 and 3.10 ppm) and 8 ($\delta_H$ 3.72 and 3.19 ppm) exhibited correlations to each other and methylene-6. Couplings from methylene-7 to the α-CH, and couplings from methylene-8 to methylene-5 suggested functionalisation on the two amines. Methylene-7 and 8 both coupled to a third resonance ($\delta_H$ 1.98 and 1.89 ppm), which was

assigned to the δ-methylene of a second Orn molecule. These collective data suggest the structure of the minor product is the dimeric aminal **28**. The difference in diastereotopic character of the sidechain methylenes, e. g. more splitting observed for methylenes 3 – 5 than for their respective counterparts 9 – 11, supports the assigned structure. Assignments for the bridging methylenes 6 – 8 are made by through-bond correlations. Surprisingly, chemical shifts (both $\delta_H$ and $\delta_C$) for methylenes-6 and 8 are more similar to each other, than to methylene-7.

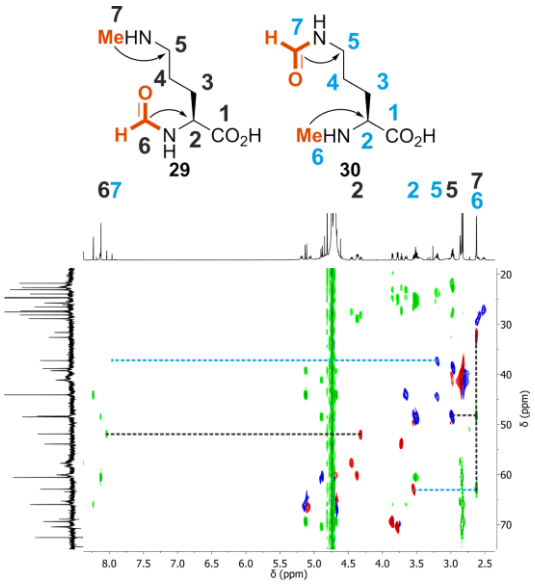



**Figure 7.20:** HSQC (red/blue) and HMBC (green) correlations of ornithine analogues **29** and **30** formed by a reaction of ornithine and HCHO under alkaline standard conditions (NaOD, 1 equivalent), after 7 days. On the two axes are projected the complementary $^{1}H$ and $^{13}C$ NMR spectra.

After allowing Orn to react with HCHO under the standard alkaline conditions described above, a portion of diazepanes **25** and **26** lost their bridging methylene, resulting in the formation of linear versions of the disproportionated products **29** and **30**. Correlations from the formyl ($\delta_H$ 8.04 ppm) to the α-CH resonance ($\delta_H$ 4.32 ppm), and correlations from a methyl resonance ($\delta_H$ 2.62 ppm) to the δ-methylene resonances ($\delta_H$ 3.51 and 2.97 ppm), suggested one of the products to be **29**. In a similar fashion, through-bond couplings from the second formyl ($\delta_H$ 7.96 ppm) to δ-methylene resonances ($\delta_H$ 3.20 ppm), and a methyl resonance ($\delta_H$ 2.62 ppm) to the α-CH resonance ($\delta_H$ 3.54 ppm), enabled assignment of this product as **30**. Although these species were encountered after 2 days of reaction, full characterisation of these linear products required sufficient product to be accumulated, which took several days. These collective

observations on the reaction of Orn with HCHO illustrate the dynamic nature and potential for complexity of the reactions between amino acids and HCHO.

### 7.3.14 Lysine and HCHO

Similar to Orn, Lys was only observed to react with HCHO under alkaline conditions (pD 9). These conditions promoted the formation of low levels (<5%) of $N^{\varepsilon}$-methylated Lys **31** over 48 hours. HMBC correlations between the methyl resonance ($\delta_H$ 2.56 ppm) and the ε-methylene ($\delta_H$ 2.92 ppm), enabled assignment of the structure as **31**. Interestingly, a third resonance ($\delta_H$ 2.85 ppm), likely to be derived from a ε-methylene, was identified by HSQC. The low intensity of the signal and broad nature of the resonances in the mixture obscured full assignment of the third structure. By contrast with Orn, cyclisation of Lys to form diazocanes was not observed, suggesting the chain length is detrimental for the formation of such species. The observations regarding the chain length are in agreement with literature reports on diaminoalkanes.[45] The rate of reaction for Lys was also found to be lower than for Orn; 48 hours were required for the accumulation of a minor amount of product. In contrast, all of Orn reacted within minutes to give cyclic diazepane **24**. No evidence for further $N^{\varepsilon}$-methylation, nor of $N^{\alpha}$-methylation, was accrued. Moreover, reactions of $N^{\varepsilon}$-methyl Lys and $N^{\varepsilon},N^{\varepsilon}$-dimethyl Lys under identical conditions did not result in further methylation.

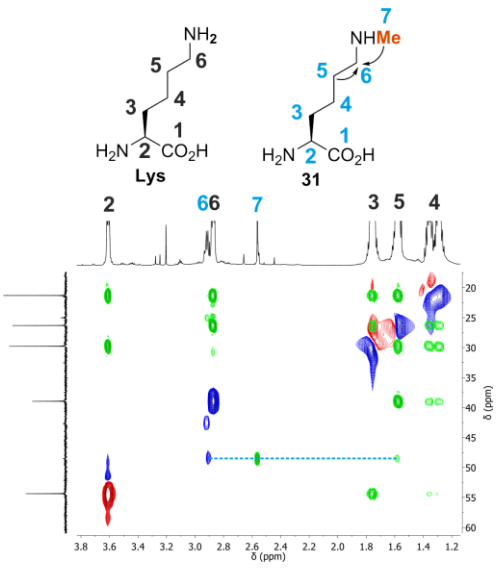



**Figure 7.21:** HSQC (red/blue) and HMBC (green) correlations of methylated lysine **31** formed by a reaction of lysine and HCHO under alkaline standard conditions (pD 9), after 48 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

### 7.3.15 Arginine and HCHO

Dissolving Arg under standard conditions resulted in moderately alkaline (pH 9) solutions. Arg reacted with HCHO under these conditions to form new cyclic products (Figure 7.22). Through-bond correlations of the major product after 24 hours, coupled two new HCHO-derived methylene resonances ($\delta_H$ 4.37 and 4.02 ppm; $\delta_H$ 4.28 and 4.19 ppm) to the α-CH resonance ($\delta_H$ 3.06 ppm) and the guanidino-C ($\delta_C$ 157.0 ppm), suggesting a cyclic structure. The bridging methylene-7 ($\delta_H$ 4.28 and 4.19 ppm) exhibited through-bond couplings to the δ-methylene resonances ($\delta_H$ 3.65 and 2.98 ppm), in contrast to methylene-8, which lacked any further couplings. These observation suggested formation of the 1,6,8-triazabicyclodecane **32**. Initial time points (1 – 6 hours) indicated the formation of an additional transient species. HMBC analysis suggested the transient resonances ($\delta_H$ 4.35 and 4.11 ppm) correspond to bridging methylene-7 (Figure 7.22), and couple to the α-CH ($\delta_H$ 3.17 ppm), the δ-methylene ($\delta_H$ 3.32 ppm), and the guanidino-C ($\delta_C$ 156.1 ppm) groups. Therefore, the transient species was assigned as diazepane **33**.

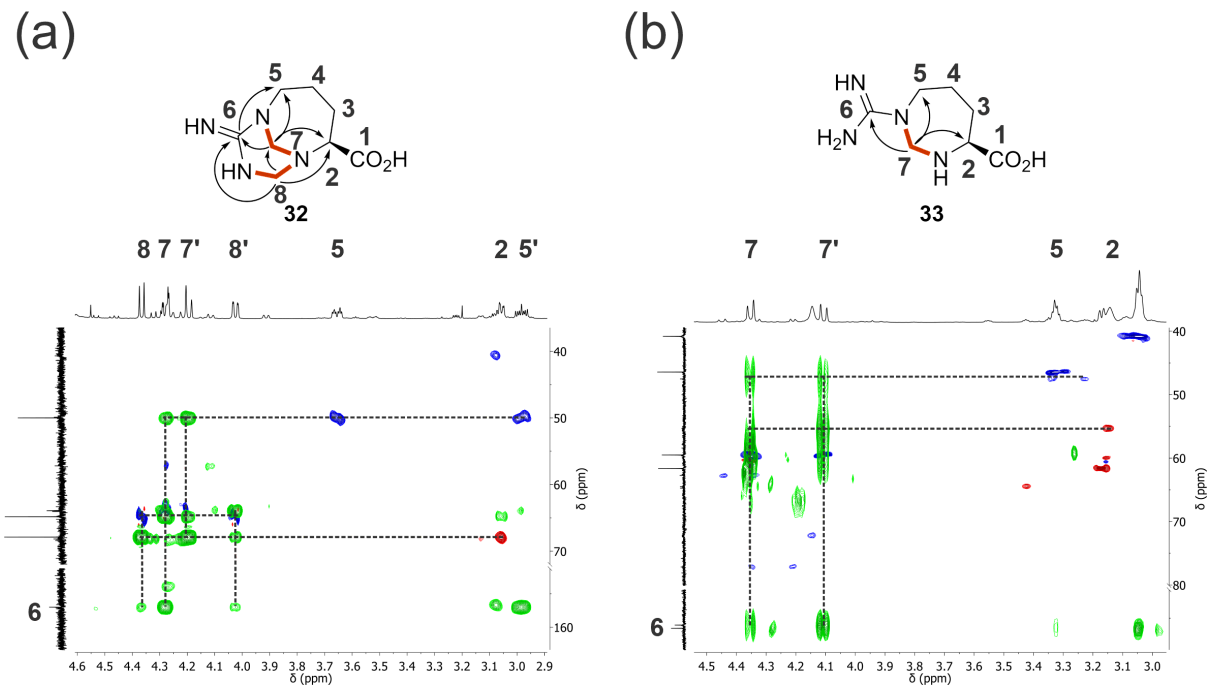



**Figure 7.22:** HSQC (red/blue) and HMBC (green) correlations of triazabicyclodecane **32** and diazepane **33** formed by a reaction of arginine and HCHO (a) under standard conditions after 24 hours, and (b) under more concentrated arginine (43.4 mM) conditions with a slight excess of HCHO (3 equivalents) after 4 hours. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Earlier observations with Orn and Lys that reveal the importance of sidechain length for cyclisation, together with the observation that **32** is replaced by **33** over time, suggests diazepane **33** could well be an intermediate *en route* to triazabicyclodecane **32**. Performing the reaction at higher Arg (43.4 mM) concentrations led to the formation of additional trace level signals. These might correspond to acyclic hydroxymethylated Arg products, as suggested previously.[20]

### 7.3.16 5-Hydroxy lysine and HCHO

Post-translationally modified amino acid 5-hydroxylysine (5hLys), is produced in collagen and other proteins.[46] Exposing 5hLys to HCHO under standard conditions, or in presence of acid (DCl, 1 equivalent), did not result in any observable product. Under alkaline conditions (NaOD, 1 equivalent), 5hLys reacted with HCHO, resulting in the formation of new broad resonances, which suggests the formation of a

dynamic species. The broad nature of the new resonances precluded full assignment of any new products. However, a tentative assignment based on chemical shift analysis and correlations from the diastereotopic ε-methylene resonance, suggest the possible formation of oxazolidine **34**. The reaction mixture was cooled (278 K, Figure 7.23) to slow down the exchange, facilitating characterisation.

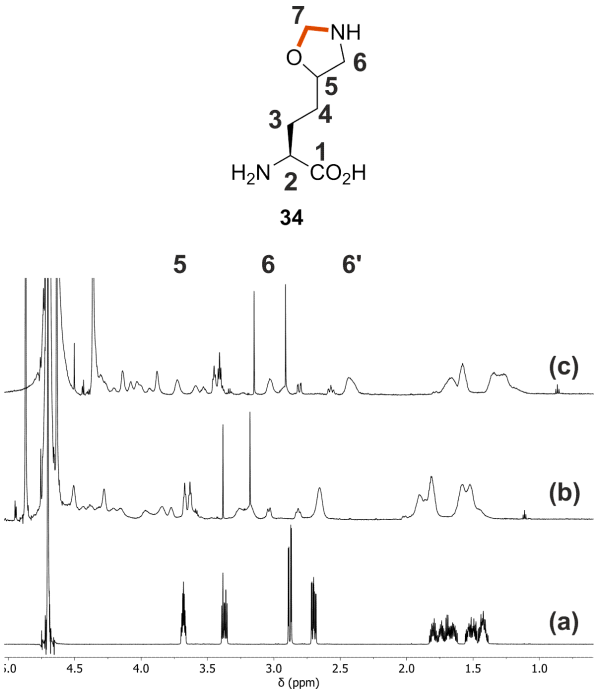



**Figure 7.23:** $^1$H NMR analyses of the equilibrium between 5-hydroxylysine (4.34 mM) and tentatively assigned oxazolidine **34**, in presence of: (a) base (NaOD, 1 equivalent) at 298 K; (b) base (NaOD, 1 equivalent) and HCHO (10 equivalents) at 298 K; (c) base (NaOD, 1 equivalent) and HCHO (10 equivalents) at 278 K. Each mixture was allowed to equilibrate for at least 20 minutes.

COSY correlations showed the resonance pair assigned as methylene-6 ($\delta_H$ 3.03 and 2.45 ppm), coupled to a new single resonance ($\delta_H$ 3.72 ppm), which is tentatively assigned as methine-5. Moreover, the diastereotopic nature of the methylene-6 resonance is in good agreement with prior observations of oxazolidines **7** – **10** and thiazolidines **1** and **3**. Both the tentatively assigned methylene-6 and methine-5 resonances show through-bond couplings to a deshielded carbon resonance ($\delta_C$ 83 ppm), indicative of an oxazolidine methylene bridge, as observed with **7** – **10**. The proposed structure, oxazolidine **34**, is in agreement with experiments carried out in Ser / Thr / *allo*-Thr, and fits with chemical shift analyses.

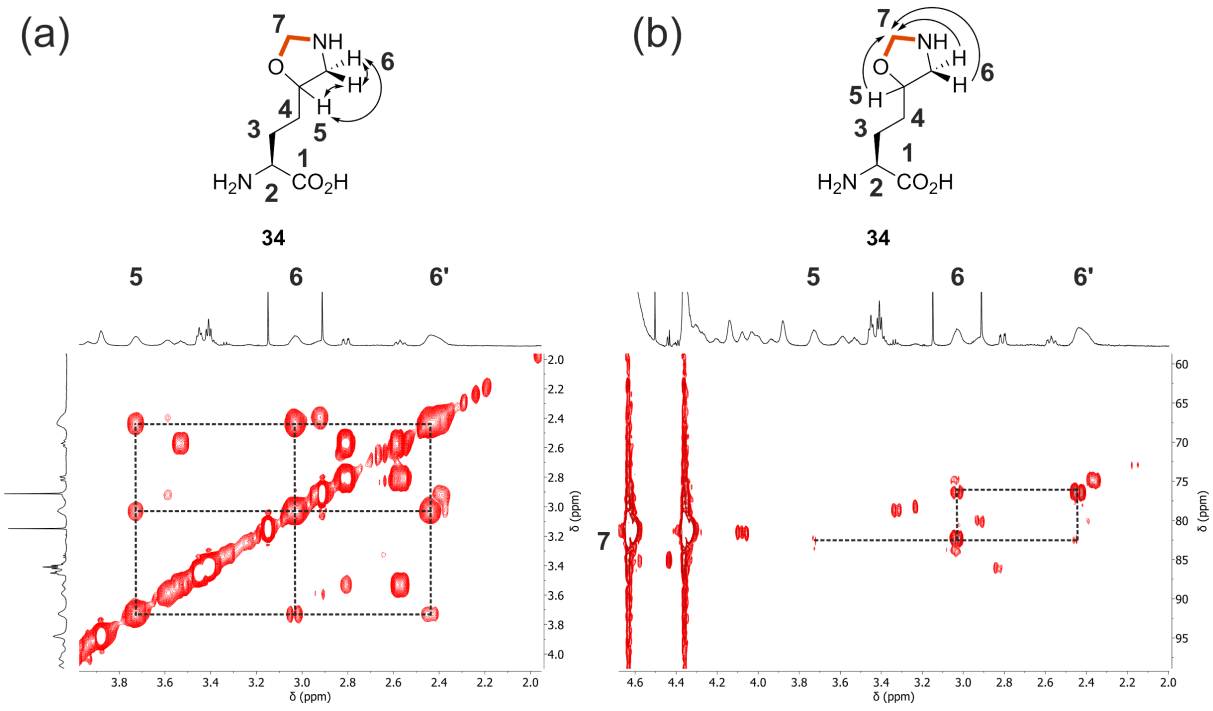



**Figure 7.24:** COSY (a) and HMBC (b) correlations of tentatively assigned oxazolidine **34** formed in a reaction between 5-hydroxylysine and HCHO under alkaline standard conditions (NaOD, 1 equivalent), after 2 hours. On the axis of the COSY, two $^{1}H$ spectra are projected (a), while the HMBC contains a $^{1}H$ spectrum, but lacks a $^{13}C$ spectrum due to low intensity signals.

### 7.3.17 Other amino acids and HCHO

Reactions between Tyr and HCHO under standard conditions, did not result in the formation of new detectable products. Solubility issues prevented the use of Tyr at higher concentrations. However, previous reports on model peptides containing Tyr in presence of Gly and an excess of HCHO suggest it reacts to form Tyr-$CH_2$-Gly adducts.[21] HCHO-derived crosslinking between Tyr and Gly has been observed crystallographically. Experiments with Ala and Pro under standard conditions, did not result in the formation of new detectable products. In general, the results of these experiments seem to imply that reactions under certain conditions can take place between HCHO and the α-amino group on amino acids, but require the involvement of a nucleophilic chain for the formation of stable, readily detectable products.

## 7.4 pH and concentration dependence studies

### 7.4.1 Experimental conditions

Investigations then focused on reactions between amino acids and HCHO under pH-controlled conditions in an attempt to represent cellular conditions more closely. Each of these reactions was carried out at a specific pH, with a fixed amino acid concentrations (4.34 mM or 43.4 mM), in presence of an excess of HCHO (10 equivalents), and monitored up to 48 hours after HCHO addition. The experiments were carried out in $D_2O$ (>99.9%), with amino acid and HCHO solutions pre-adjusted to one of three observed pH values: pH 5.6, 7.1, and 8.6. The pD value can be calculated from the observed pH reading using the following equation:[47]

$$pD = pH^* + 0.41 \quad (5.1)$$

Herein pH* represents the observed pH value, and pD the actual value, hence the off-set of the pH values by 0.4.

### 7.4.2 Cysteine and homocysteine pH-controlled studies

As a starting point, the thiol containing amino acids, including Cys and Hcy were investigated. Reactions between Cys (4.34 and 43.4 mM) and HCHO at any of the three studied pH values resulted in rapid formation of thiazolidine **1** (Table 7.1, entry 1). Hydroxymethylated **2** was only observed with the lower concentrations of Cys (4.34 mM) at pD 6 and 7.5 (Figure 7.25). Efficiency of thiazolidine **1** formation increased with increased concentration and under more alkaline conditions. Reactions between Hcy (4.34 and 43.4 mM) and HCHO similarly led to rapid thiazinane **3** formation (Table 7.1, entry 2). At lower Hcy (4.34 mM) concentrations, the reaction with HCHO was comparable to that of Cys with HCHO. Thiazinane **3** formation was more efficient with increased Hcy concentrations, and under more alkaline conditions.

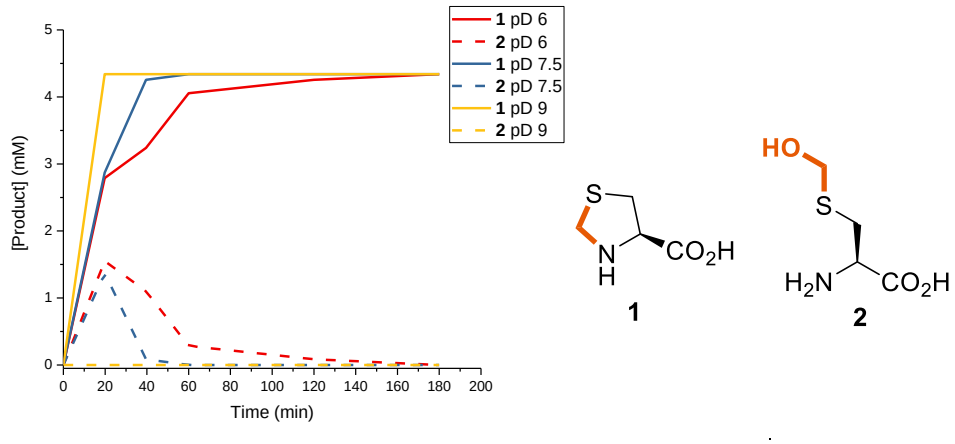



**Figure 7.25:** Time course of the reaction between cysteine (4.34 mM) and HCHO (10 equivalents) performed at different pH values (pD 6 (red lines), pD 7.5 (blue lines), pD 9 (yellow lines)), showing the formation of thiazolidine **1** (solid lines) and hydroxymethylated cysteine **2** (dashed lines).

### 7.4.3 Serine, threonine, *allo*-threonine, and homoserine pH-controlled studies

As observed previously, the reaction between Ser and HCHO is less efficient than the reactions of thiol containing amino acids. At the lower Ser (4.34 mM) concentration, oxazolidine **7** was only observed at pD 9 (Table 7.1, entry 3, 8.3%, percentage of Ser converted to oxazolidine **7**). Although monitored for 48 hours, no further oxazolidine **7** formation was observed after 20 minutes after the addition of HCHO, suggesting rapid establishment of the equilibrium between oxazolidine **7** and unreacted Ser. Similarly to observations with Cys and Hcy, the efficiency of oxazolidine **7** formation was increased at higher concentrations of Ser, and at higher pH. In contrast to the reaction with Ser, the reaction between Thr and HCHO, resulting in oxazolidine **8** formation, was observed under all tested conditions, and to a higher extent of conversion (Table 7.1, entry 4, 10.7 – 57.4 %). Oxazolidine formation was more efficient at higher concentrations and at higher pH. Formation of oxazolidine **9** in the reaction between *allo*-Thr and HCHO was rapid (Table 7.1, entry 5) but was less efficient than formation of oxazolidine **8**. Formation of a *cis*-substituted oxazolidine ring in **9**, as opposed to the *trans*-substituted ring in **8**, might cause steric hindrance, favouring formation of **8** relative to **9**. The reaction between Hse and HCHO resulted in formation of oxazinane **10** (Table 7.1, entry 6). Compared to Ser, Thr, and *allo*-Thr, cyclisation with Hse was more efficient under all tested conditions (38.7 – 86.0 %). Oxazinane formation is suggested to be more efficient than oxazolidine

formation. As observed before, higher concentrations and more alkaline conditions resulted in more efficient cyclisation with Hse than for other alcohol containing amino acids.

**Table 7.1 Overview of the products and their distribution after 48 hours as a function of pH.**

| | | Low concentration (4.34 mM) | | | High concentration (43.4 mM) | | |
|---|---|---|---|---|---|---|---|
| | **pD** | **6.0** | **7.5** | **9.0** | **6.0** | **7.5** | **9.0** |
| **Entry** | **Composition** | **%** | **%** | **%** | **%** | **%** | **%** |
| 1 | **Cys** | 0.0[a] | 0.0[b] | 0.0[c] | 0.0[c] | 0.0[c] | 0.0[c] |
| | **1** | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | **2** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | Hcy | 0.0[d] | 0.0[a] | 0.0[b] | 0.0 [c] | 0.0 [c] | 0.0 [c] |
| | **3** | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | **4** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | Ser | 100.0 | 100.0 | 91.7[d] | 92.6 | 86.2 | 77.5 |
| | **7** | 0.0 | 0.0 | 8.3 | 7.4 | 13.8 | 22.5 |
| 4 | Thr | 89.3[e] | 87.0[e] | 75.5[e] | 55.6[e] | 52.6[e] | 42.6[e] |
| | **8** | 10.7 | 13.0 | 24.5 | 44.4 | 47.4 | 57.4 |
| 5 | *allo*-Thr | 96.2[d] | 93.5[d] | 80.0[d] | 77.5 | 74.6 | 66.2 |
| | **9** | 3.8 | 6.5 | 20.0 | 22.5 | 25.4 | 33.8 |
| 6 | Hse | 61.3[d] | 54.5[d] | 42.7[d] | 18.0[d] | 16.0[d] | 14.0[d] |
| | **10** | 38.7 | 45.5 | 57.3 | 82.0 | 84.0 | 86.0 |
| 7 | Asn | 25.0 | 23.9 | 18.2 | 4.6 | 0.5 | 2.6 |
| | **11** | 70.9 | 70.2 | 72.4 | 95.4 | 93.1 | 43.9 |
| | **12** | 4.1 | 6.0 | 20.4 | 0.0 | 6.4 | 47.4 |
| | **13** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.1 |
| 8 | Gln | 82.6 | 81.3 | 70.0 | 41.3[f] | 38.5[f] | 39.4[g] |
| | **14** | 17.4 | 18.7 | 30.0 | 58.7 | 61.5 | 60.6 |
| 9 | His | 58.8 | 40.2 | 39.7 | 43.3 | 0.7 | 0.0 |
| | **15** | 21.2 | 42.6 | 42.1 | 13.2 | 22.0 | 17.1 |
| | **16** | 3.9 | 4.6 | 5.2 | 12.8 | 30.9 | 35.7 |
| | **17** | 1.9 | 3.8 | 4.3 | 6.6 | 0.0 | 0.0 |
| | **18** | 14.2 | 4.4 | 4.7 | 22.4 | 35.5 | 31.2 |
| | **19** | 0.0 | 4.4 | 4.0 | 1.7 | 8.9 | 11.9 |
| | **20** | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 4.1 |
| 10 | Trp | 11.9 | 10.2 | 6.9 | 0.0 | 0.0 | 3.2 |
| | **21** | 66.2 | 67.8 | 49.5 | 0.0 | 0.0 | 6.9 |
| | **22** | 15.2 | 12.5 | 19.3 | 56.1 | 55.4 | 15.4 |
| | **23** | 6.6 | 9.5 | 24.3 | 43.9 | 44.6 | 74.5 |
| 11 | Orn | 100.0 | 100.0 | 71.1 | 88.4 | 87.4 | 65.5 |
| | **24** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | **25** | 0.0 | 0.0 | 13.3 | 5.7 | 6.3 | 16.9 |
| | **26** | 0.0 | 0.0 | 15.6 | 5.9 | 6.2 | 17.2 |
| | **27** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | **28** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | **29** | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 |
| | **30** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 |
| 12 | Lys | 100.0 | 100.0 | 100.0 | 89.7 | 88.1 | 75.0 |
| | **31** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.3 |
| | **31b** | 0.0 | 0.0 | 0.0 | 10.3 | 11.9 | 20.7 |
| 13 | Arg | 100.0 | 95.2 | 75.2 | 98.0 | 96.2 | 87.7 |
| | **32** | 0.0 | 4.8 | 23.3 | 2.0 | 4.8 | 13.3 |
| | **33** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Overview of the products observed for each amino acid at the stated concentrations in presence of HCHO (10 equiv.) after a maximum of 48 hours, displayed as percentage conversions.
a) Equilibrium was reached after 45 minutes
b) Equilibrium was reached after 25 minutes
c) Equilibrium was reached after 5 minutes
d) Equilibrium was reached after 1 hour
e) Equilibrium was reached after 20 min
f) Reaction was monitored for 40 hours
g) Reaction was monitored for 44 hours

### 7.4.4 Asparagine and glutamine pH-controlled studies

The reactions between Asn and HCHO, under pH controlled conditions, resulted in formation of lactams **11**, **12**, and **13** (Table 7.1, entry 7). Formation of lactam **11** was observed under all tested conditions (43.9 – 93.1 %). In contrast to the thiol- and alcohol-containing amino acids discussed above (Cys, Hcy, Pen, Ser, Thr, *allo*-Thr, Hse), formation of the lactams was much slower; only lactam **11** was observed after 48 hours. Additional hydroxymethylation, resulting in the formation of lactam **12**, was observed under most conditions, and was favoured under more alkaline conditions. Formation of lactam **13**, was only observed at higher Asn (43.4 mM) concentrations and under alkaline conditions (pD 9). When exposing Gln to HCHO under the pH controlled conditions, gradual formation of lactam **14** was observed, albeit less efficiently than with Asn (Table 7.1, entry 8, 17.4 – 61.5 %).

### 7.4.5 Histidine and tryptophan pH-controlled studies

When exposing His to HCHO under the pH-controlled conditions, rapid formation of *N*,*N*-cyclised product **15** was observed (Table 7.1, entry 9). Interestingly, hydroxymethylated His **18** was most efficiently formed from His (4.34 and 43.4 mM) at pD 6 (Table 7.1, entry 9, 14.2%, and 22.4% respectively). At lower His (4.34 mM) concentrations and at pD 7.5 and 9, *N,N*-cyclisation was promoted, forming **15** (Table 7.1, entry 9, ~42%). At the higher His (43.4 mM) concentrations, and at pD 7.5 or 9, clear formation of **16** was observed (Table 7.1, entry 9, 30.9 – 35.7%). Additionally, *C,N*-cyclisation was dominant (4.0 – 11.9%), at pD 7.5 and 9, at both low and high concentrations of His. Hydroxymethylation of Trp, resulting in the formation of **21**, was observed at the lower Trp (4.34 mM) concentrations for any of the pD conditions (Table 7.1, entry 10, 49.5 – 67.8%). Some Pictet-Spengler product **22** (Table 7.1, entry 10, 12.5 – 19.3%) and **23** (6.6 – 24.3%) were observed under the lower Trp (4.34 mM) concentrations. At the higher Trp (43.4 mM) concentrations, *C,N*-cyclisation was promoted, resulting in a higher formation of products **22** and **23** (up to 56.1% and 74.5% respectively).

### 7.4.6 Ornithine, lysine and arginine under pH-controlled conditions

Exposing Orn (4.34 mM) to HCHO at pD 9 resulted in the formation of diazepanes **25** and **26** (Table 7.1, entry 11). At higher Orn (43.4 mM) concentrations, formation of **25** and **26** was observed across the different pH ranges, however production of **25**/**26** was favoured at pD 9 (5 – 6 % at pD 6 and 7.5; 17% at pD 9). Trace amounts of the linear methylated and formylated products were observed (Table 7.1, entry 11, 0.1 – 0.3%). Interestingly, formation of **24** was not observed under these conditions. The reactions between Lys (43.4 mM) and HCHO resulted in trace amounts of Lys **31** at higher concentrations and pD 9, after 48 hours. As alluded to earlier (Section **7.3.14**), evidence for the formation of another species was accrued, potentially corresponding to either a hemiaminal or an imine. However, the broad nature of the resonances precluded full assignment. Studies with Lys(Me) and $Lys(Me)_2$ under identical conditions did not lead to evidence for further methylation. As with Lys, Arg remained largely unreactive at both concentrations and across the pH range, only providing evidence for the formation of **32**. The reaction of Arg with HCHO resulted in the formation of **32** after 48 hours. Formation of **32** was most efficient under more alkaline conditions (pD 9, up to 23.3%). Initial time points (up to 6 hours, pD 9) showed evidence for the formation of **33**.

### 7.4.7 Oxazolidine formation pH effects

To elaborate on the results found with oxazolidines and oxazinanes, and to further investigate the effect of pH on the reactions between HCHO and Ser, Thr, Hse, and 5hLys, new experiments at different pH values were then carried out. The amino acids (4.34 mM) were individually allowed to react with HCHO (10 equivalents) at different pHs (pD 6.4, 7.0, 8.0, 9.0, 9.8, 11.0, and 12.0, in 25 mM phosphate buffer), and the mixtures were equilibrated for at least 20 minutes before acquisition. Formation of oxazolidine **7** from Ser was observed to be most significant at pD 12, resulting in more >50% conversion (Figure 7.26 (a)). In contrast, a similar conversion of Thr to oxazolidine **8**, was achieved at pD 8 (Figure 7.26 (b)). These results were confirmed when combining both Thr and Ser (2.5 mM each) in a single solution, and allowing them to react with an excess of HCHO (20 equivalents) – this experiment resulted in a similar observation.

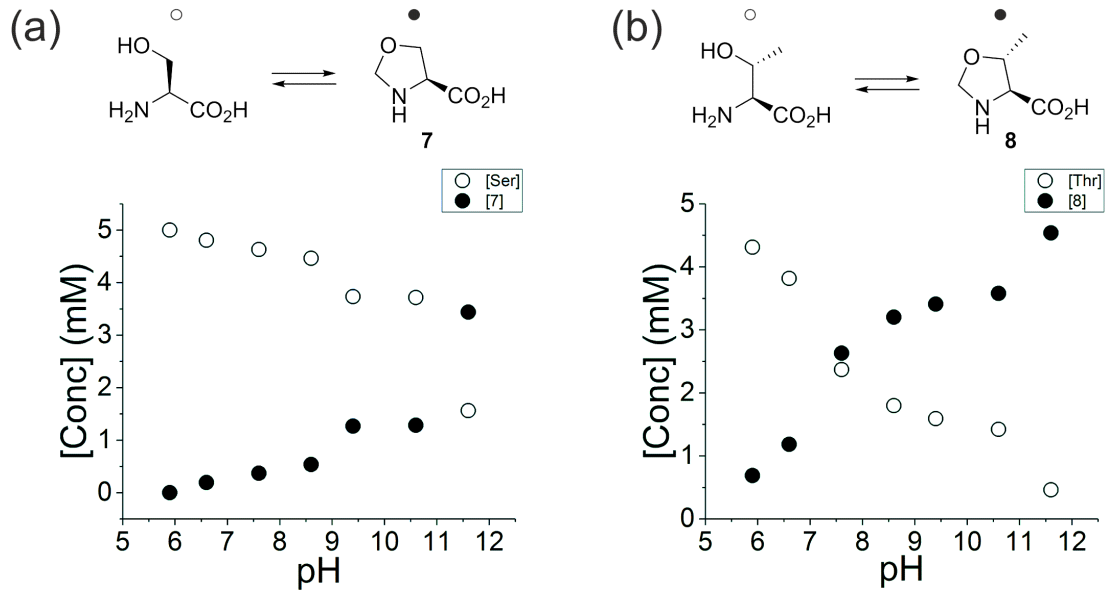



**Figure 7.26:** Reaction of serine with oxazolidine **7** and threonine with oxazolidine **8** plotted as a function of pH. At each pH, a mixture of serine or threonine with HCHO was allowed to equilibrate for at least 20 minutes. The concentration of each species was determined from relative $^1$H NMR peak integrals and the starting concentration of the amino acid. (a) Serine (5 mM, ○) in equilibrium with oxazolidine **7** (●); (b) Threonine (5 mM) in equilibrium with oxazolidine **8** (●).

The reaction between Hse and HCHO to form oxazolidine **10** reached 50% conversion at neutral pH (Figure 7.27 (a)). These observations suggest oxazinane formation is more favourable than oxazolidine formation under neutral pH conditions. When reacting 5hLys with HCHO, resulting in the formation of tentatively assigned oxazolidine **34**, more alkaline conditions are required to achieve 50% conversion to **34** (Figure 7.27 (b)). Although tentatively assigned, the observations in these pH dependent studies correlate well with results obtained with Ser, and are therefore in line with oxazolidine formation being favoured at higher pH.

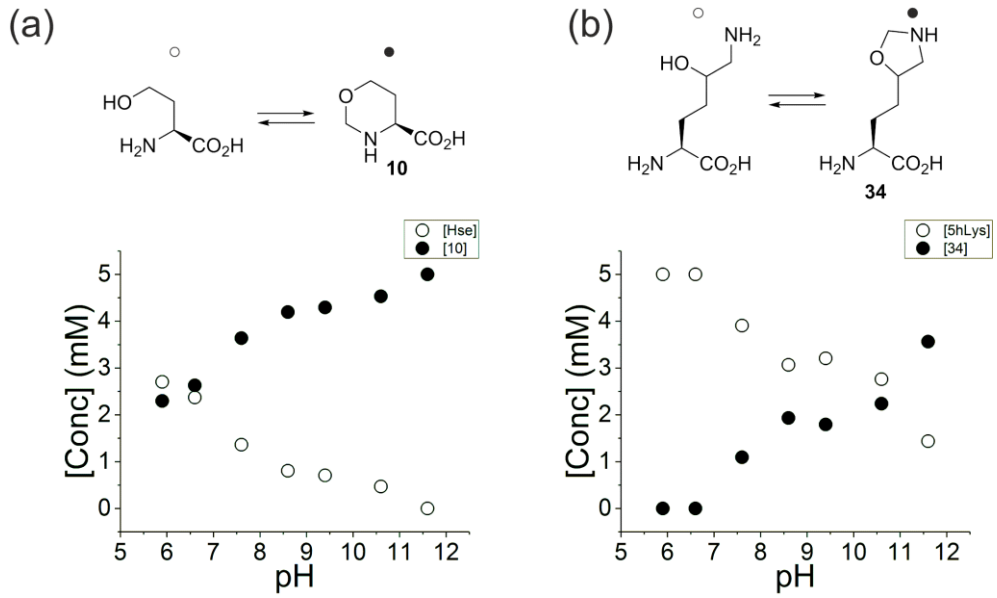



**Figure 7.27:** Equilibria of homoserine with oxazolidine **10** and 5-hydroxylysine with oxazolidine **34** plotted as a function of pH. At each pH, a mixture of serine or threonine with HCHO was allowed to equilibrate for at least 20 minutes. The concentration of each species was determined from relative $^1$H NMR peak integrals and the starting concentration of the amino acid. (a) Homoserine (5 mM, ○) in equilibrium with oxazolidine **10** (●); (b) 5-Hydroxylysine (5 mM) in equilibrium with oxazolidine **34** (●).

A similar pH-dependent reaction profile was observed for Thr and Ser (both at 2.5 mM) when both were present in the same mixture with HCHO in excess (20 equivalents). These observations suggest an HCHO concentration is saturating for reaction with Thr / Ser. The significant difference in reactivity between Thr and Ser was described earlier (Section **7.3.6**) and could derive from a kinetic Thorpe-Ingold effect caused by the steric bulk of the Thr methyl group. Other explanations such as $pK_a$ or electronic effects cannot be excluded. A mixture containing Thr and *allo*-Thr (both at 2.5 mM) in presence of HCHO (20 equivalents) showed an identical pH-dependent reactivity profile. Cyclisation to form oxazolidines seemed to be favoured Thr over *allo*-Thr under the tested conditions.

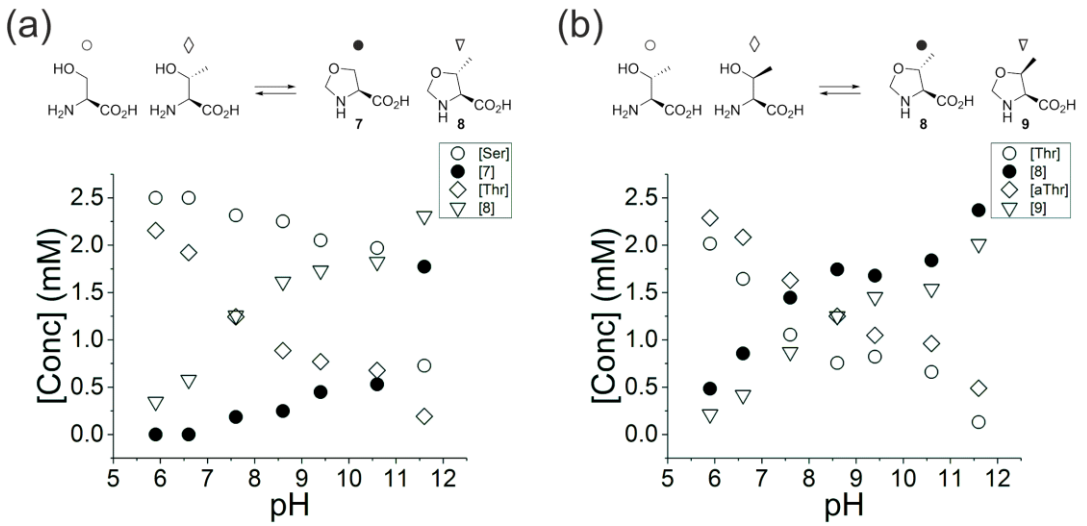



**Figure 7.28:** Mixtures of serine and threonine, and threonine and *allo*-threonine in equilibrium with their respective oxazolidine products, plotted as function of pH. At each pH, a mixture of threonine with serine or *allo*-threonine in presence of an excess of HCHO was allowed to equilibrate for at least 20 minutes. The concentration of each species was determined from relative $^{1}$H NMR peak integrals and the starting concentration of the amino acid. (a) Serine (2.5 mM, ○) and threonine (2.5 mM, ◊) in equilibrium with oxazolidines **7** (●) and **8** (▼); (b) Threonine (2.5 mM, ○) and *allo*-threonine (2.5 mM, ◊) in equilibrium with oxazolidines **8** (●) and **9** (▼).

### 7.4.8 Cellular concentrations

Levels of HCHO in food products[48] or from the air[49,50] have been studied. Studies on the *in vivo* concentrations of HCHO[48] have led to rough whole body estimates and in some cases to tissue or fluid specific concentrations.[1,51] However, determining local concentrations within tissues, let alone cellular and subcellular concentrations, remains challenging. It is arguable that effective molarity and local constraints play a role in artificially increasing concentrations on a local level, reaching concentrations covered in this study. Nevertheless, to increase the biological relevance of this study, attention moved to performing reactions with HCHO at reported *in vivo* concentrations. Although estimates for the exact concentration vary, generally ppm levels (μM) of HCHO are reported as the lower limit. Therefore, experiments with Cys or His were performed at μM concentrations, with a lower amino acid to HCHO ratio. The reaction between Cys (25 μM) and HCHO (30 μM, 1.16 equivalents), resulted in the formation of thiazolidine **1** (Figure 7.29 (a)), despite the lower concentration and limited excess. The reaction between His and HCHO however, did not result in any observable product formation under these conditions. This could be as a result of the lower HCHO excess or due to the lower concentrations at which this reaction was carried out.

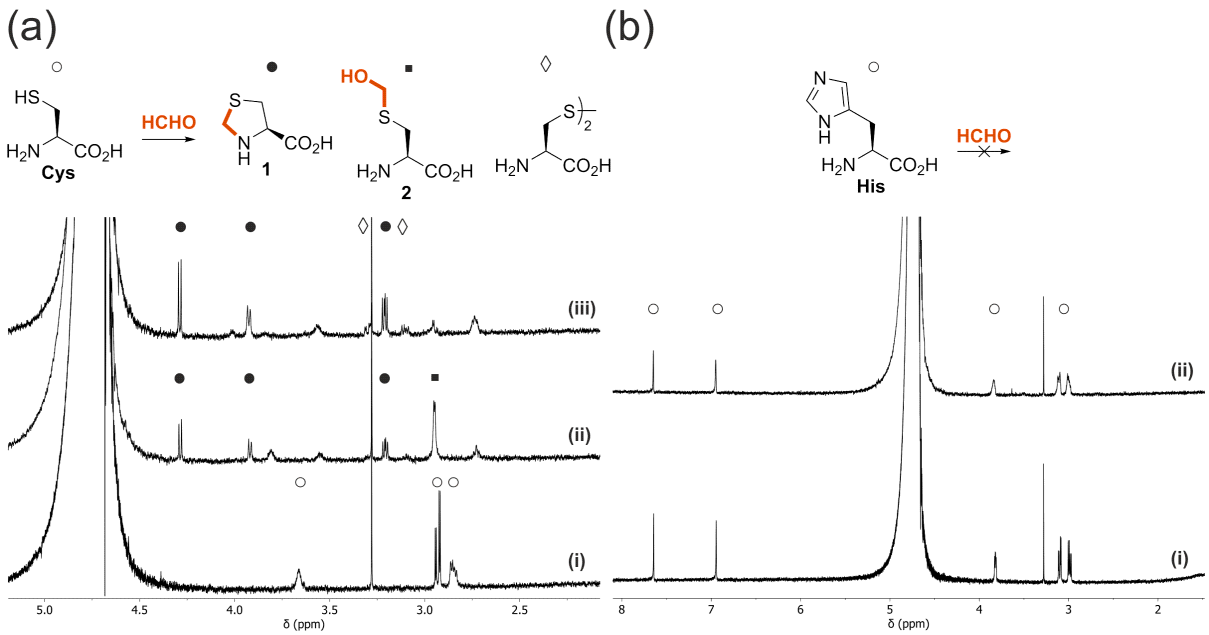



**Figure 7.29:** $^{1}$H NMR spectra of the reactions of cysteine and histidine with HCHO at μM concentrations. (a) Time course of the reaction between cysteine (25 μM) and HCHO (30 μM), (i) cysteine (25 μM) before addition of HCHO, (ii) 30 minutes after addition of HCHO (30 μM), (iii) 10 hours after addition of HCHO; (b) Reaction between histidine (25 μM) and HCHO (30 μM), 1) before and 2) 48 hours after addition of HCHO, (i) histidine (25 μM) before addition of HCHO, (ii) 48 hours after addition of HCHO.

## 7.5 Assessing product stability

The stabilities of the products observed from the reactions between amino acids and HCHO were then analysed. The excess of HCHO was removed from the mixtures, after which the lifetimes of the adducts were monitored. Removal of HCHO was achieved by either: (i) treating the mixtures with 1,3-diketocyclohexane (DCH, 40 equivalents with respect to the amino acid), a more soluble analogue of dimedone;[52,53] or (ii) lyophilisation of the mixture, followed by redissolving the residue in $D_2O$. Removal of HCHO by lyophilisation was often achieved using multiple cycles of lyophilising and redissolving in $D_2O$. Solutions containing amino acids (4.34 mM) and HCHO (10 equivalents) were allowed to react to completion or equilibrium (3 – 24 hours), before HCHO removal. When either method was applied to a solution containing both **15** and **16**, rapid re-formation of His was observed (within 1 hour), suggesting fast degradation of the two adducts (Figure 7.30).

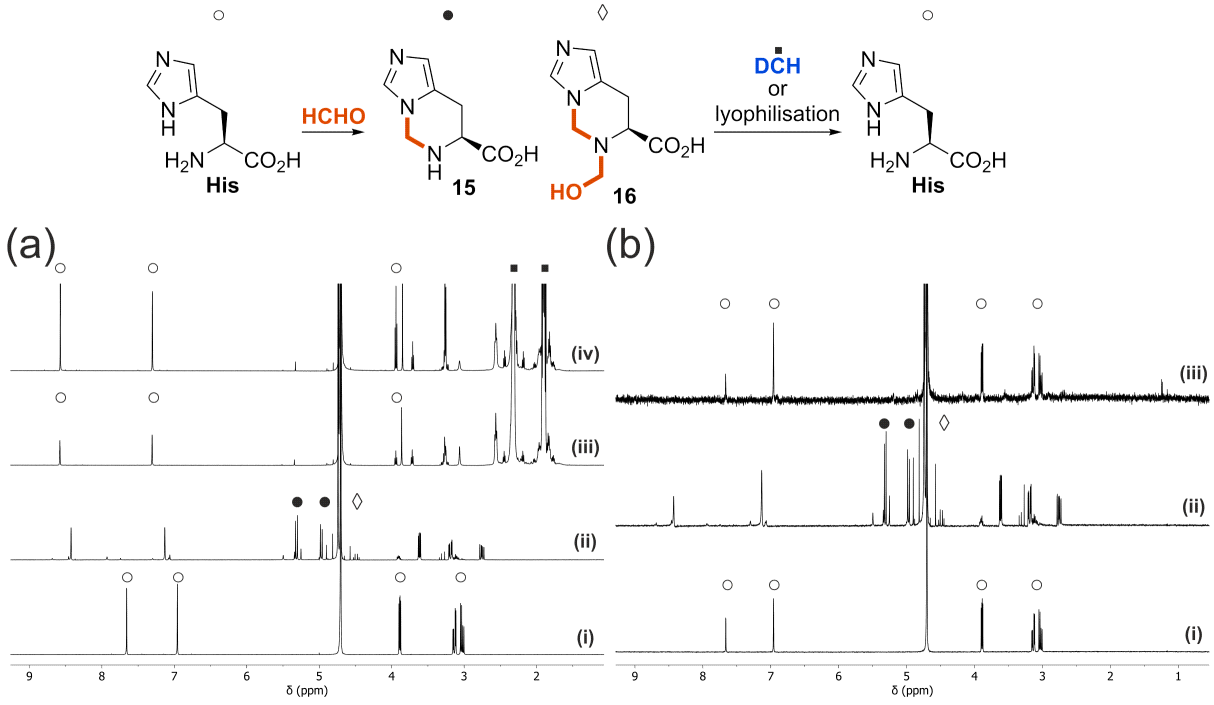



**Figure 7.30:** Stability studies on histidine-derived **15** and **16** using DCH and lyophilisation. Histidine (4.34 mM) was allowed to react with HCHO (10 equivalents) for 10 hours. Subsequently, the excess of HCHO is removed using (a) diketocyclohexane (DCH), or (b) lyophilisations cycles. (a) DCH stability assay, (i) Histidine (4.34 mM), (ii) 10 hours after addition of HCHO (10 equivalents), (iii) 16 hours after addition of DCH (40 equivalents), (iv) spiked with histidine (1 equivalent); (b) lyophilisation stability assay, (i) histidine (4.34 mM), (ii) 10 hours after addition of HCHO (10 equivalents), (iii) after 5 lyophilisation cycles.

When Ser was reacted with an excess of HCHO in presence of base (NaOD 1 equivalent), an equilibrium between oxazolidine **7** and the parent amino acid was established (Appendix Figure 2). Addition of DCH to the mixture resulted in fast (1 hour) conversion to the parent amino acid Ser. Similarly, when subjecting Thr-derived **8** to DCH, rapid formation of Thr was observed (<4 hours, Appendix Figure 3). Removal of HCHO was also successfully achieved using lyophilisation (7 cycles), resulting in complete conversion to Thr (Appendix Figure 4). In comparison, treatment of Arg-derived **32** with DCH resulted in partial formation of Arg at a slower rate (72 hours, Appendix Figure 5). Lyophilisation (8 cycles) of a solution containing Arg-derived **32** similarly led to only partial removal of the HCHO derived bridging methylenes (Appendix Figure 6). Applying either of the methods (for 72 hours or 7 lyophilisation cycles) to a mixture containing both lactams **11** and **12** revealed slow and incomplete formation of Asn, suggesting these adducts (**11**, **12**) are relatively stable compared to oxazolidines **7** and **8** (Appendix Figures 7 and 8).

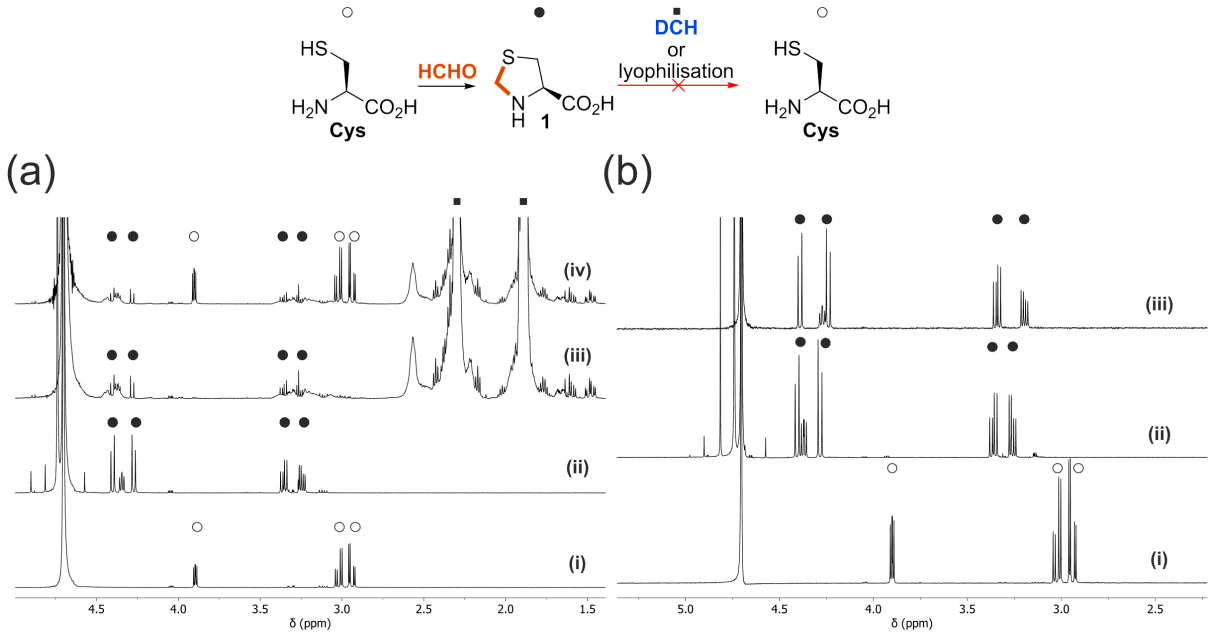



**Figure 7.31:** Stability studies on thiazolidine **1** using DCH and lyophilisation. Cysteine (4.34 mM) is allowed to react with HCHO (10 equivalents) for 10 hours. Subsequently, the excess of HCHO is removed using (a) diketocyclohexane (DCH), or (b) lyophilisations cycles. (a) DCH stability assay, (i) cysteine (4.34 mM), (ii) 10 hours after addition of HCHO (10 equivalents), (iii) 16 hours after addition of DCH (40 equivalents), (iv) spiked with cysteine (1 equivalent), note the appearance of new peaks suggesting no cysteine being present in the sample prior to spiking; (b) lyophilisation stability assay, (i) cysteine (4.34 mM), (ii) 10 hours after addition of HCHO (10 equivalents), (iii) after 5 lyophilisation cycles.

Treatment of mixtures containing Trp-derived **22** and **23** with DCH resulted in loss of the hydroxymethyl adduct in **23** to give **22** over 48 hours, but did not result in loss of the *C,N*-bridging methylenes (Appendix Figure 9). Treatment of thiazolidine **1** with DCH or lyophilisation cycles did not result in formation of Cys during the analysis, suggesting unusual stability of **1** (Figure 7.31). To investigate whether any exchange of HCHO in solution occurs with the methylene bridge in thiazolidine **1**, and as an alternative to the two described stability methods, $^{13}$C-labelled HCHO (H$^{13}$CHO) was added to the solution containing thiazolidine **1**. Over time, an increase in the intensity of the $^{13}$C satellites of the *N,S*-bridging methylene in thiazolidine **1** in the $^{1}$H NMR (Figure 7.32 (a)) confirmed incorporation of the $^{13}$C isotope label. The slow incorporation (over 48 hours) of the $^{13}$C-label into **1** shows Cys and thiazolidine **1** are in slow dynamic exchange. When Trp derived **21** and **23** were exposed to H$^{13}$CHO (16 hours), incorporation of the label was observed in the hydroxymethyl groups, but not in the *C,N*-bridging methylene.

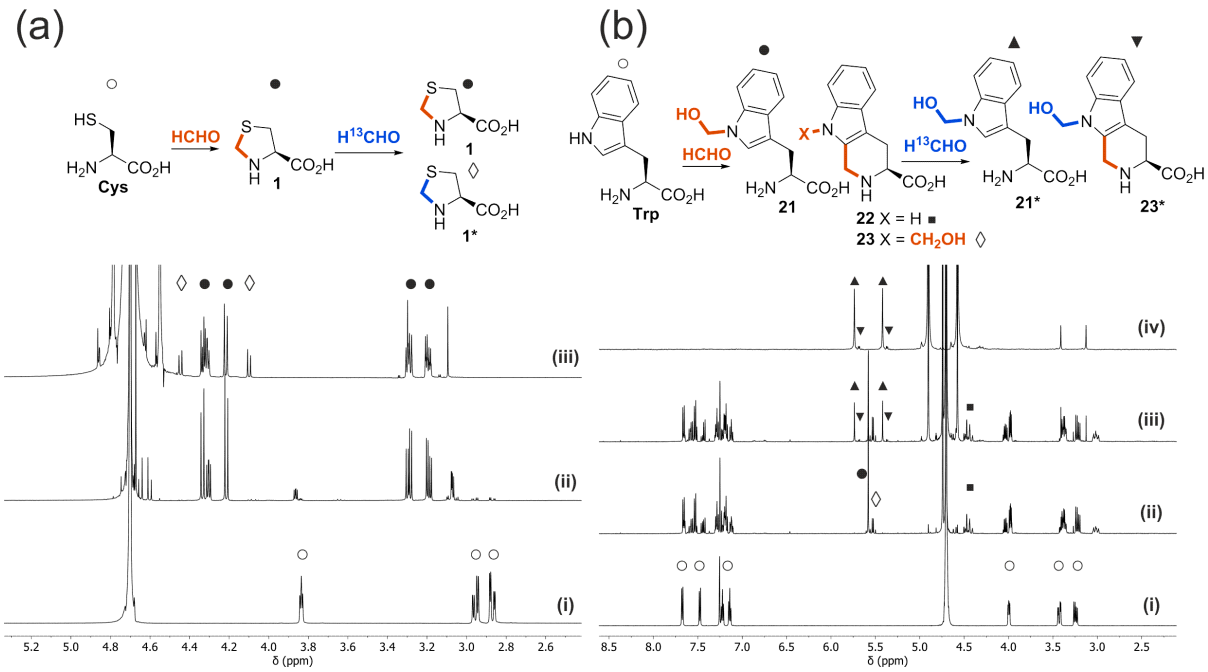



**Figure 7.32:** Stability studies on thiazolidine **1** and tryptophan-derived **21**, **22**, and **23** using $^{13}$C-labelled H$^{13}$CHO. Cysteine (a) or tryptophan (b) (either at 4.34 mM) were allowed to react with HCHO for up to 24 hours. Subsequently, the H$^{13}$CHO was added and analysed for exchange with $^{13}$C. (a) Cysteine with H$^{13}$CHO, (i) cysteine (4.34 mM), (ii) 3 hours after addition of HCHO (2 equivalents), (iii) 48 hours after addition of H$^{13}$CHO (10 equivalents); (b) Tryptophan with H$^{13}$CHO, (i) tryptophan (4.34 mM), (ii) 24 hours after addition of HCHO (5 equivalents), (iii) 16 hours after addition of H$^{13}$CHO (5 equivalents), (iv) 1D HSQC analysis of the sample, showing carbon-13 satellites.

The lack of incorporation of the isotope label demonstrated little, or no, retro Pictet-Spengler reaction occurs, suggesting these HCHO derived *C,N*-bridges are amongst the most stable HCHO-derived adducts. By extrapolation, these results suggest that the *C,N*-bridging methylenes in His-derived **19** and **20**, proposed to be formed using a similar mechanism, are of similar stability. Although not investigated in detail, it is probable that the formyl and methyl modifications found in Orn-derived **25**, **26**, **28** and **29**, and Lys **31** are similarly stable adducts.

## 7.6 Amino acid competition experiments

### 7.6.1 Competition between cysteine and different amino acids for reaction with HCHO

The results presented thus far support the proposal that Cys manifests particular reactivity towards HCHO, resulting in rapid and stable thiazolidine formation. Thus, the initial competition experiments focused on investigating the effect of the addition of amino acids to reaction mixtures containing Cys and HCHO. In these competition experiments, a mixture of Cys and either Thr, Hse, Asn, His, Trp, Orn, Lys, or Arg, were allowed to react with stoichiometric amounts of HCHO (1 equivalent, 1 : 1 : 1 ratio).

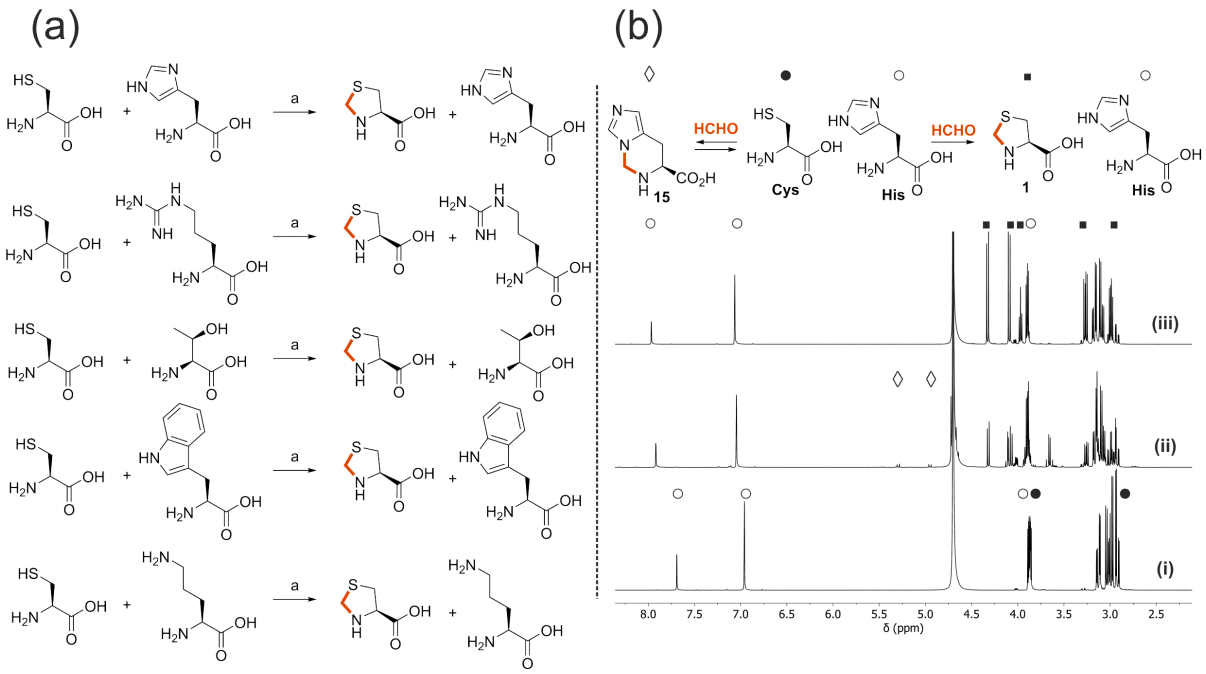



**Figure 7.33:** Competition assays between cysteine and individual amino acids for reaction with HCHO. (a) Overview of the individual amino acids (4.34 mM) used to compete with cysteine (4.34 mM) for reaction with HCHO (1 equivalent); (b) $^{1}$H NMR time course of a competition between cysteine and histidine, (i) cysteine (4.34 mM) and histidine (1 equivalent), (ii) 40 minutes after addition of HCHO (1 equivalent), (iii) 4 hours after HCHO addition.

In all cases, the Cys-derived thiazolidine **1** was the exclusively observed product; **1** was formed within 12 hours following addition of HCHO (Appendix Figures 10 – 17). Early time points showed the formation of hydroxymethyl adduct **2**, which over time was converted into thiazolidine **1**. With His, the initial time points (<1 hour) showed the formation of *N,N*-cyclised **15**, which eventually reverted to give His alongside thiazolidine **1** (Figure 7.33 (b)) Competition experiments with Thr, gave trace amounts of oxazolidine **8** after addition of HCHO, which rapidly reverted back to free HCHO and free amino acid.

### 7.6.2 Competition experiments of thiol containing amino acids

Competition experiments with the three thiol-containing amino acids Cys, Hcy and Pen, were then conducted. Stoichiometric amounts of HCHO were added to a mixture of Cys, Hcy and Pen (4.34 mM each, 1 : 1 : 1 ratio), and allowed to react for 6 hours (Figure 7.34). The major product in the mixture at the end of the analysis was Pen-derived thiazolidine **5** (55% of total product formed). The Cys-derived thiazolidine **1** was the next most abundant product (30%), while the Hcy-derived thiazinane **3** (15%) was the least abundant product. Under the tested conditions, thiazolidine ring formation is more efficient and/or more stable than thiazinane ring formation.

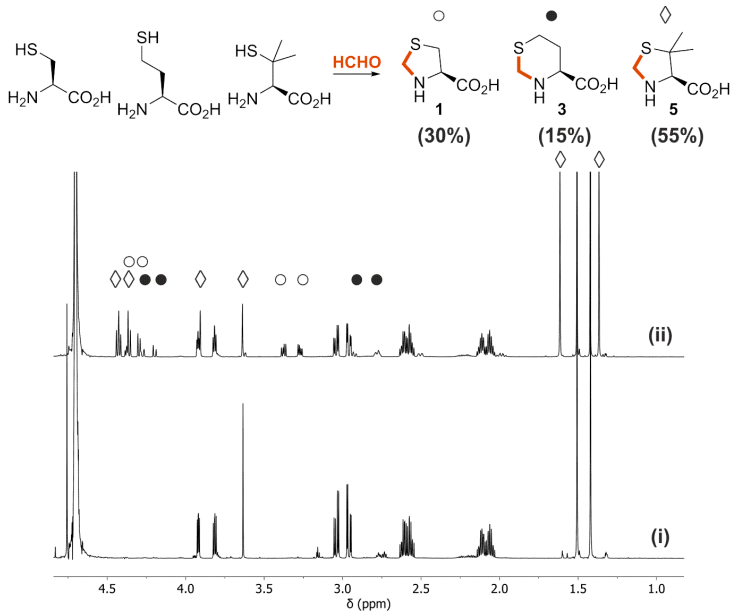



**Figure 7.34:** Competition assays between cysteine, homocysteine, and penicillamine for reaction with HCHO. A mixture of cysteine (4.34 mM), homocysteine (1 equivalent), and penicillamine (1 equivalent) to which HCHO (1 equivalent) is added. (i) mixture of thiol containing amino acids, (ii) mixture 48 hours after addition of HCHO.

The observation that thiazolidine **5** is more abundant than **1** correlated with observations of the relative stability and rate of formation of oxazolidines **7** and **8** (Section **7.6.3**). These results seem to be in line with the potential for the kinetic Thorpe-Ingold effect.

### 7.6.3 Competition experiments of alcohol containing amino acids

Using a similar set up as for the thiol-containing amino acids Ser, Thr, *allo*-Thr and Hse were allowed to react with stoichiometric amounts of HCHO (1 : 1 : 1 : 1 : 1 ratio). This mixture however, did not reveal a significant amount of product formation. Moving away from the limiting HCHO conditions (10 equivalents), resulted in observed formation of oxazolidines **8** (35%, percentage of total product) and **9** (11%), and oxazinane **10** (54%). When using a larger excess of HCHO (100 equivalents), more of the individual amino acids were converted into cyclic products, giving **8** (39%), **9** (18%), and **10** (43%). Notably, Ser still did not react with HCHO to form oxazolidine **7** under the tested conditions.

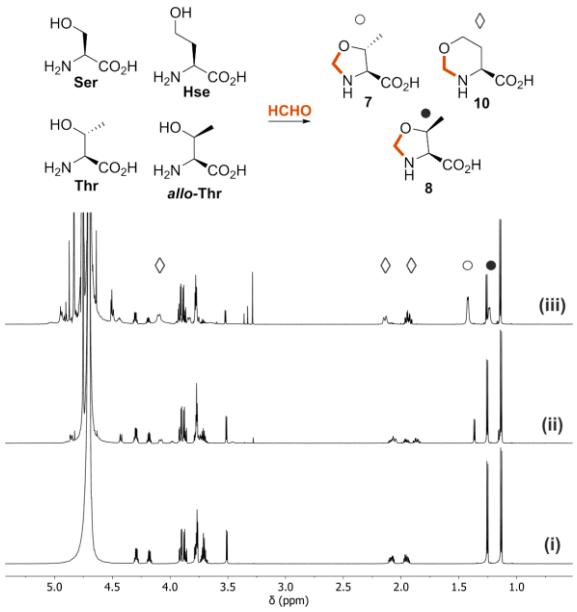



**Figure 7.34:** Competition assays between serine, homoserine, threonine, and *allo*-threonine for reaction with HCHO. A mixture of serine (4.34 mM), threonine (1 equivalent), *allo*-threonine (1 equivalent), and homoserine (1 equivalent) to which HCHO is added. (i) mixture of alcohol containing amino acids 1 hour after addition of HCHO (1 equivalent), (ii) mixture 3 hours after the addition of HCHO (10 equivalents total), (iii) mixture 3 hours after addition of HCHO (100 equivalents total). Note the lack of reactivity of serine towards HCHO under the tested conditions.

### 7.6.4 Competition experiments with a range of amino acids

To reflect the diversity and complexity of a biological sample more closely, competition experiments with a larger number of amino acids were conducted. A mixture of amino acids, containing Ser, Thr, *allo*-Thr, Hse, His, Trp, Asn, Orn, Arg, and Lys, all present at the same concentration, was prepared, and allowed

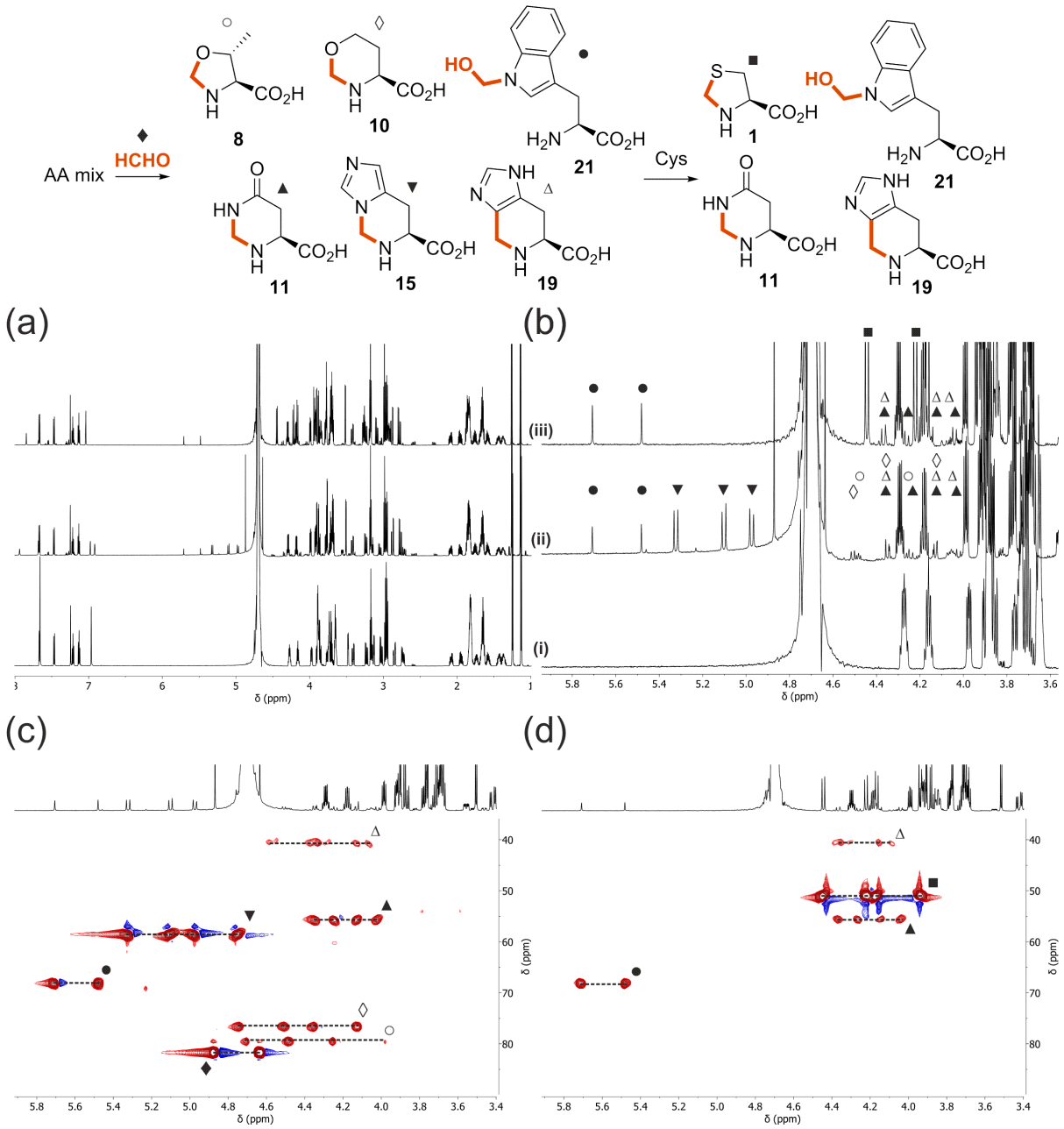



**Figure 7.35:** Competition assays using a mixture of amino acids in presence of limiting amount of H$^{13}$CHO, followed by cysteine addition. A mixture containing serine, threonine, *allo*-threonine, homoserine, histidine, tryptophan, asparagine, ornithine, arginine, and lysine (all at 4.34 mM) was exposed to H$^{13}$CHO (4.34 mM, 1 equivalent). Subsequently, cysteine was added and allowed to react. (a) & (b) (i) Mixture of amino acids before H$^{13}$CHO addition, (ii) 36 hours after addition of H$^{13}$CHO (1 equivalent), (iii) 36 hours after addition of cysteine (1 equivalent); (c) HSQC of the reaction mixture 36 hours after addition of H$^{13}$CHO; (d) HSQC of the reaction mixture 36 hours after the addition of cysteine. On the top axis is projected the complementary $^{1}$H spectrum in (c) and (d). Note that the HSQC was run without $^{13}$C decoupling.

to react with HCHO (1 equivalent) for 36 hours (Figure 7.35 (a) and (b)). His-derived **15** was found to be the predominant adduct (29%), followed by Asn-derived lactam **11** (10%). Other products including oxazolidine **8** (7%) and oxazinane **10** (9%), Trp-derived **21** (7%), and His-derived **19** (4%), were also observed, while some HCHO remained unreacted (Figure 7.35 (c)). Subsequently, Cys (1 equivalent) was added to the mixture and allowed to react, resulting in the disappearance of some (**8**, **10** and **17**) of the adducts from the mixture, correlating with the formation of thiazolidine **1** (Figure 7.35). Adducts **11**, **19**, and **21** persisted in solution, revealing their kinetic stability (Figure 7.35 (c) and (d)). Notably, these adducts were not observed when Cys was in direct competition with Trp, His, or Asn (Section **7.5.1**).

### 7.6.5 A spectator amino acid and the rate of reaction of HCHO

Some of the amino acids that have been tested in these studies did not react to form a distinct adduct. It is, however, possible that these amino acids are in fast dynamic exchange with HCHO, as evidenced by the reaction of Ser with HCHO. These apparent equilibria might affect the rate of reaction of other amino acids present in complex mixtures. To investigate this, Cys (5 mM) and HCHO (1 equivalent) were mixed in presence of "spectator" amino acids and monitored over time. Ala, Asn, Thr or His (10 equivalents) were used as spectator amino acids in these experiments. To prevent changes in rates as a result of small differences in pH, the mixtures were prepared in phosphate buffer (50 mM, pD 7.5). The addition of Ala or Asn as a spectator amino acid to the reaction mixture of Cys and HCHO did not result in the formation of any observable products. With Thr or His, formation of oxazolidine **8** or *N,N*-bridged **15** were observed, respectively. The formation of Thr-derived **8** did not affect the initial formation rate of **1**. By contrast, the formation of His-derived **15** resulted in a slower rate of thiazolidine **1** formation.

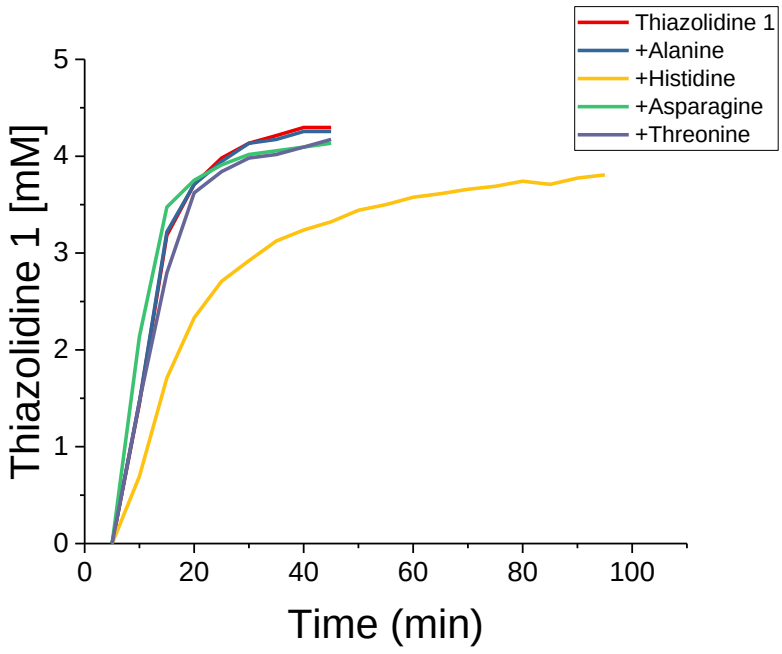



**Figure 7.36:** Plot of the rates of formation of thiazolidine **1** from a reaction between cysteine and HCHO in presence and absence of different amino acids. Reaction between cysteine and HCHO in absence of a spectator amino acid (red); cysteine (4.34 mM) in presence of: alanine (10 equivalents, blue); histidine (10 equivalents, blue); asparagine (10 equivalents, blue); and threonine (10 equivalents, blue). Concentrations of thiazolidine **1** were determined by peak integration from the $^{1}$H NMR time course, and referenced to the original cysteine concentration.

## 7.7 Comparison of the reactivity of Cys with HCHO using different electrophiles

The focus of the studies then moved to investigating the relative reactivity of HCHO compared to other biologically relevant electrophiles. As Cys was established as the most potent nucleophile within the tested set of amino acids, it was used to check for reactivity from other biologically relevant electrophiles. These electrophiles included: acetaldehyde (AcH), acetone, glyoxal and methylglyoxal. From these reactions between Cys and electrophiles a set of thiazolidine analogous to **1** were formed. These carbonyl compounds were found in hydrated and non-hydrated form while in solution. Interestingly, their level of hydration was less than for HCHO. Acetaldehyde was mainly (~70%) in the non-hydrated form under the tested conditions. Reaction of Cys (4.34 mM) with acetaldehyde resulted in the formation of thiazolidines **35a** and **35b** (Figure 7.37 (a)). Through-bond correlations from CH-4 ($\delta_H$ 4.82 and 4.74 ppm) to both methylene-3 ($\delta_H$ 3.40 and 3.19 ppm; 3.34 and 3.24 ppm) and αCH ($\delta_H$ 4.37 and 4.29 ppm), assigned the thiazolidine structure. The formation of thiazolidines **35a** and **35b** was completed after 1 hour. Formation of the *trans*-substituted ring, as confirmed by $^1$H-$^1$H NOESY, was favoured (1.3 : 1). Under the tested conditions in $D_2O$, the unhydrated form of acetone was observed exclusively. The reaction between Cys and acetone resulted in the formation of **36**, as confirmed by HMBC correlations from the quaternary C ($\delta_C$ 73.2 ppm) to both methylene-3 ($\delta_H$ 3.57 and 3.41 ppm) and α-CH ($\delta_H$ 4.52 ppm) (Figure 7.37 (b)). The conversion to thiazolidine **36** was incomplete (33%), even after addition of base (NaOD, 1 equivalent, 56%).

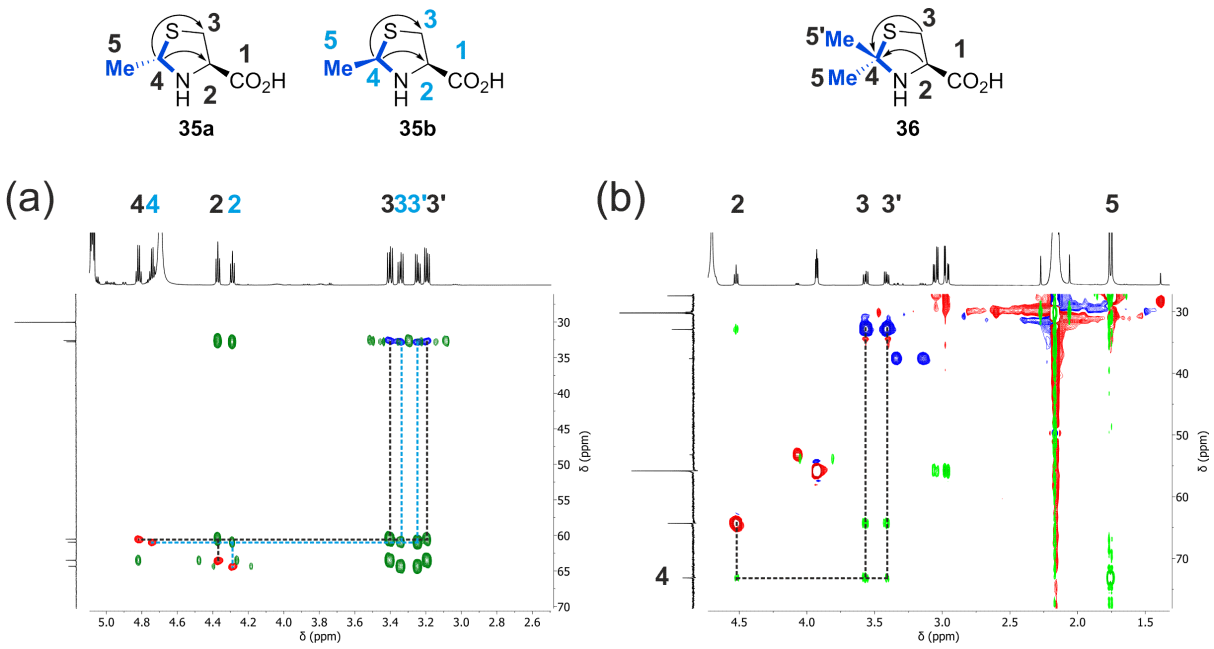



**Figure 7.37:** HSQC (red/blue) and HMBC (green) correlations of thiazolidines **35a**, **35b** and **36** formed by a reaction of cysteine and acetaldehyde (a) and acetone (b). Cysteine (4.34 mM) with: (a) acetaldehyde (10 equivalents) forming thiazolidines **35a** and **35b**, and (b) acetone (10 equivalents) forming thiazolidine **36**. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

Glyoxal and methylglyoxal were found to be mainly in their aldehyde forms under the experimental conditions (43.4 mM of aldehyde, in unbuffered $D_2O$), as observed by $^{1}H$ NMR. The reaction between Cys (4.34 mM) and glyoxal (10 equivalents) resulted in the formation of thiazolidines **37a** and **37b** favouring the *trans*-substituted ring (1.7 : 1, **37a** : **37b**, respectively). HMBC correlations from both **37a** and **37b** demonstrated couplings from their respective methine-4 resonances ($\delta_H$ 5.19 and 5.15 ppm) to both methylene-3 resonances ($\delta_H$ 3.33 and 3.24 ppm; $\delta_H$ 3.30 and 3.21 ppm) and α-CH resonances ($\delta_H$ 4.39 and 4.46 ppm) (Figure 7.38 (a)). No bisthiazolidine formation, resulting from double condensation of Cys with glyoxal, was observed under the tested conditions. Even when Cys was in excess (2 equivalents), no bisthiazolidine product was observed. The reaction between Cys (4.34 mM) and methylglyoxal (10 equivalents) resulted in the formation of thiazolidines **38a** and **38b**, favouring the *trans*-substituted ring (1.8 : 1) (Figure 7.38 (b)). Through-bond correlations from methine-4 resonances ($\delta_H$ 5.65 and 5.72 ppm) to both α-CH resonances ($\delta_H$ 4.45 and 4.55 ppm) and methylene-3 ($\delta_H$ 3.42 and 2.99 ppm; $\delta_H$ 3.38 and 3.18 ppm) within **38a** and **38b** respectively, were used to assign the products.

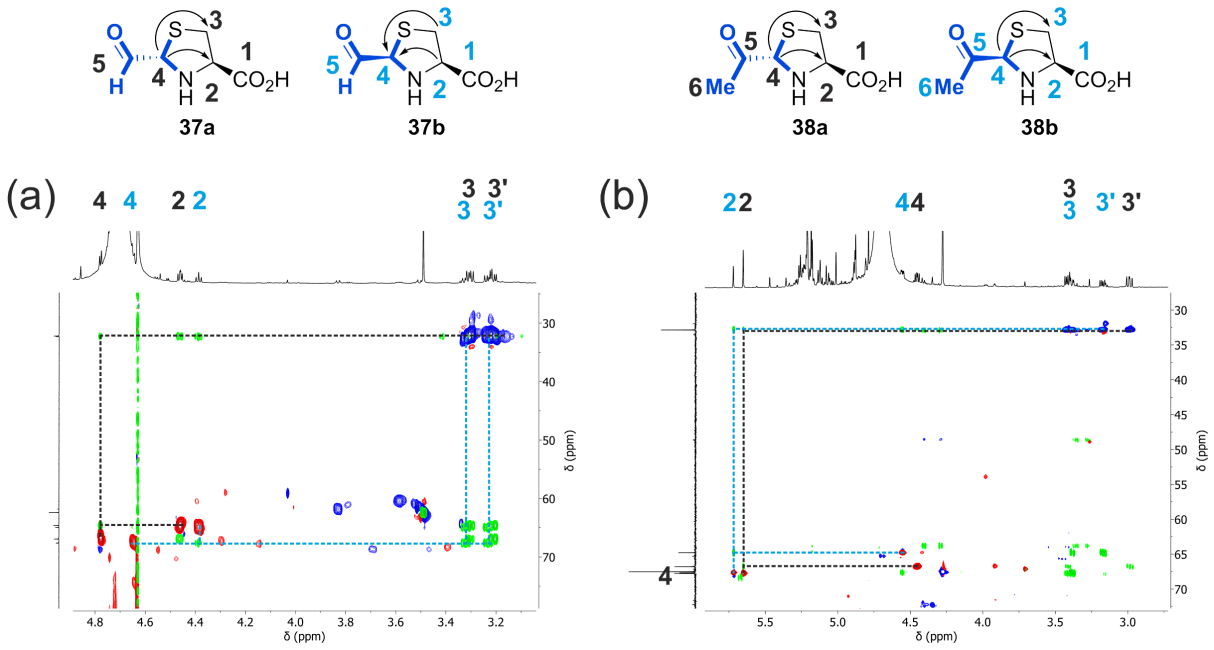



**Figure 7.38:** HSQC (red/blue) and HMBC (green) correlations of thiazolidines **37a**, **37b**, **38a** and **38b** formed by a reaction of cysteine and (a) glyoxal and(b) methylglyoxal. Cysteine (4.34 mM) with: (a) glyoxal (10 equivalents) forming thiazolidines **37a** and **37b**, and (b) methylglyoxal (10 equivalents) forming thiazolidines **38a** and **38b**. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

The reactivity of these different electrophiles was then compared to that of HCHO. A solution of Cys (4.34 mM) was treated with each of the electrophiles (10 equivalents) and allowed to react to completion. HCHO (100 equivalents) was then added and the reaction was monitored over time by $^{1}H$ NMR. The extent to which the modifications were displaced by HCHO were tracked over time to establish relative stabilities and reactivities. The reverse reaction was also studied, i.e. initial formation of thiazolidine **1** from Cys (4.34 mM) and HCHO (10 equivalents), followed by the exposure to each of the different electrophiles (100 equivalent). Displacement of acetone adduct **36** by HCHO was observed to be fastest of such reactions to be studied, resulting in full conversion to thiazolidine **1** within 5 minutes from addition of HCHO (Figure 7.39 (a)). Notably, the reverse reaction did not result in the observation of displacement of HCHO from thiazolidine **1**. Glyoxal-derived products **37a** and **37b** were displaced by HCHO to result in the formation of **1**, within an hour (Figure 7.39 (c)). The reverse set up did not result in any displacement of the HCHO-derived modification in thiazolidine **1**.

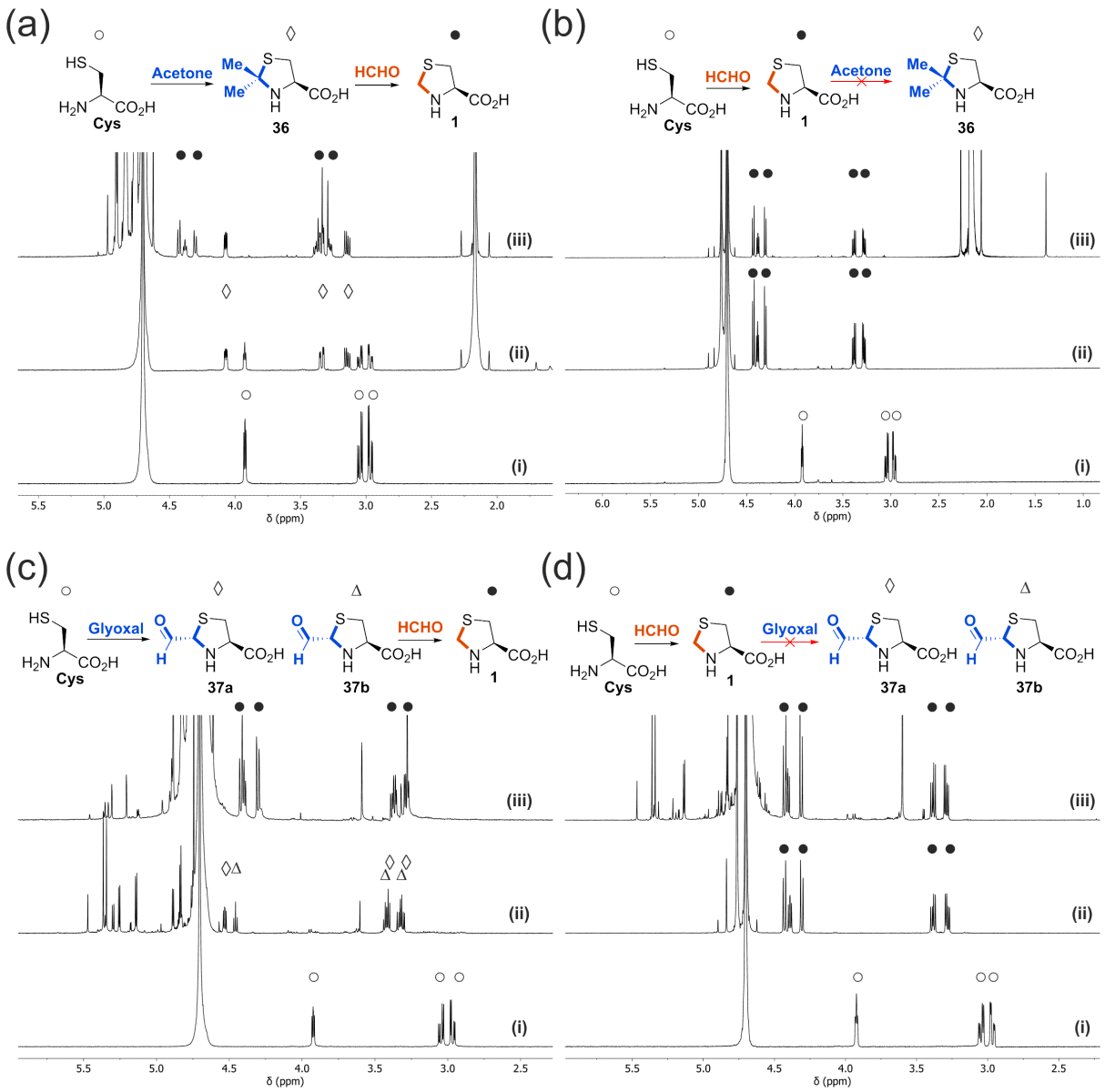



**Figure 7.39:** Electrophile competition assays reacting cysteine with acetone, glyoxal and HCHO. (a) (i) Cysteine (4.34 mM), (ii) was allowed to react with acetone (10 equivalents), (iii) followed by the addition of HCHO (100 equivalents); (b) (i) cysteine (4.34 mM), (ii) was allowed to react with HCHO (10 equivalents), (iii) followed by addition of acetone (100 equivalents); (c) (i) cysteine (4.34 mM), (ii) was allowed to react with glyoxal (10 equivalents), (iii) followed by addition of HCHO (100 equivalents); (d) (i) cysteine (4.34 mM), (ii) was allowed to react with HCHO (10 equivalents), (iii) followed by addition of glyoxal (100 equivalents).

Four hours after the addition of HCHO, methylglyoxal-derived products **38a** and **38b** were completely converted to thiazolidine **1** (Figure 7.40(a)). The reverse reaction, starting with thiazolidine **1** did not result in the conversion to thiazolidines **38a** and **38b**. Displacement of acetaldehyde-derived modifications in **35a** and **35b** was completed 5 hours after the addition of HCHO (Figure 7.40 (c)). Again, no displacement of

HCHO-derived modification in thiazolidine **1** was observed in the reverse set up. The combined results suggest that the reactions of HCHO with Cys, and by extrapolation with other amino acids, form the most stable products from the set of biologically relevant electrophiles tested.

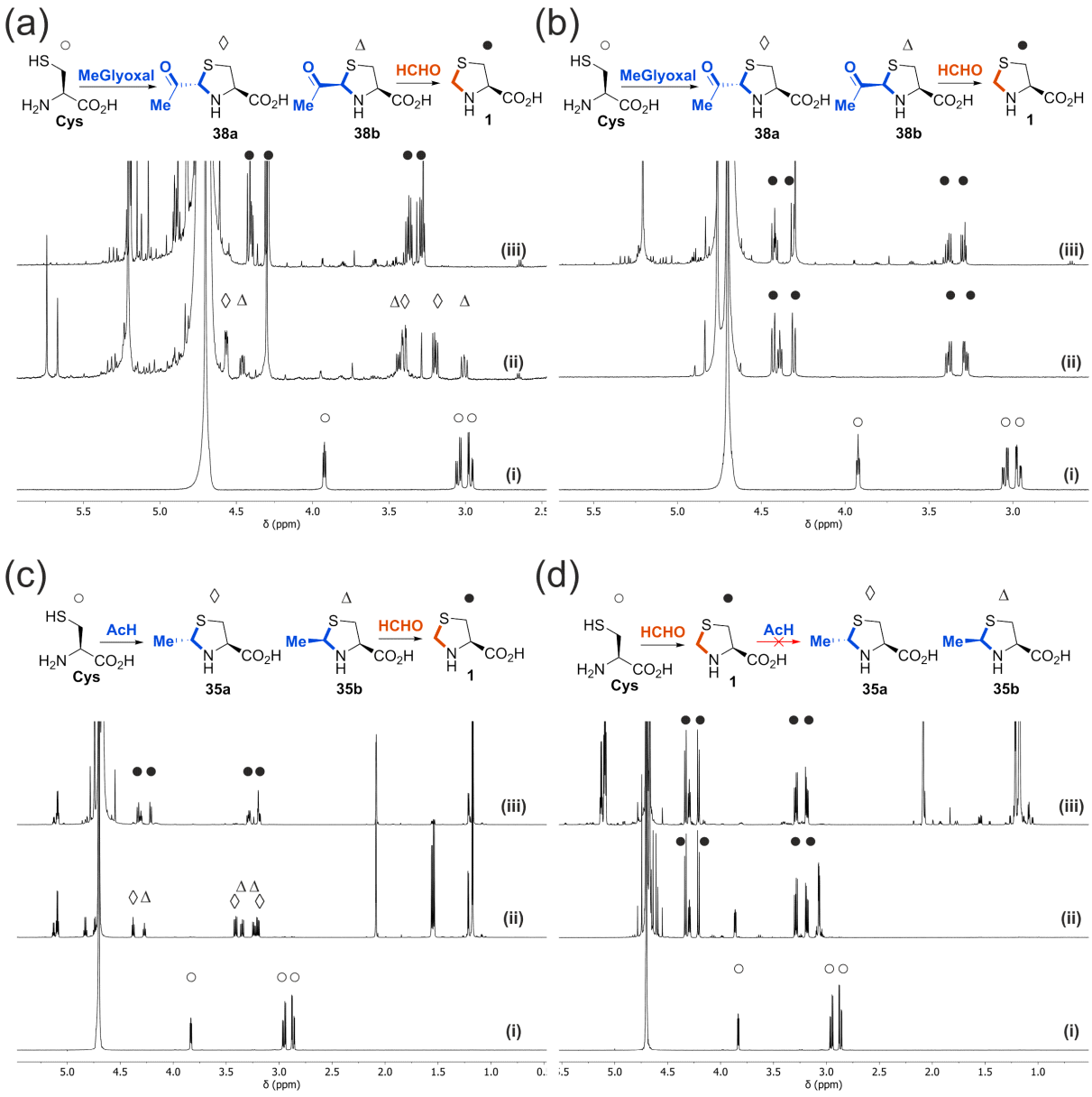



**Figure 7.40:** Electrophile competition assays reacting cysteine with methylglyoxal, acetaldehyde, or HCHO. (a) (i) Cysteine (4.34 mM), (ii) was allowed to react with methylglyoxal (10 equivalents), (iii) followed by the addition of HCHO (100 equivalents); (b) (i) cysteine (4.34 mM), (ii) was allowed to react with HCHO (10 equivalents), (iii) followed by addition of methylglyoxal (100 equivalents); (c) (i) cysteine (4.34 mM), (ii) was allowed to react with acetaldehyde (10 equivalents), (iii) followed by addition of HCHO (100 equivalents); (d) (i) cysteine (4.34 mM), (ii) was allowed to react with HCHO (10 equivalents), (iii) followed by addition of acetaldehyde (100 equivalents).

## 7.8 Cysteine containing peptides

The results described so far imply that the reactivity of HCHO is most efficient when reacting with thiol-containing amino acids, initially giving an *S*-hydroxymethyl adduct. An $N^a$-amino group is required for the rapid formation of the stable thiazolidine **1** in the case of Cys. To investigate whether similar modifications occur in peptides and proteins, two short peptides were exposed to HCHO. Glutathione (4.34 mM), a trimeric peptide that involved in HCHO detoxification pathways,[54] and the structurally related tripeptide L-δ-(2-aminoadipoyl)-L-cysteinyl-D-valine (**ACV**, 4.34 mM)), were separately exposed to an excess of HCHO (Figure 7.41). Under the neutral tested conditions, both glutathione and **ACV** reacted to form *S*-hydroxymethyl adducts with HCHO. These observations are in agreement with previous studies with glutathione.[55]

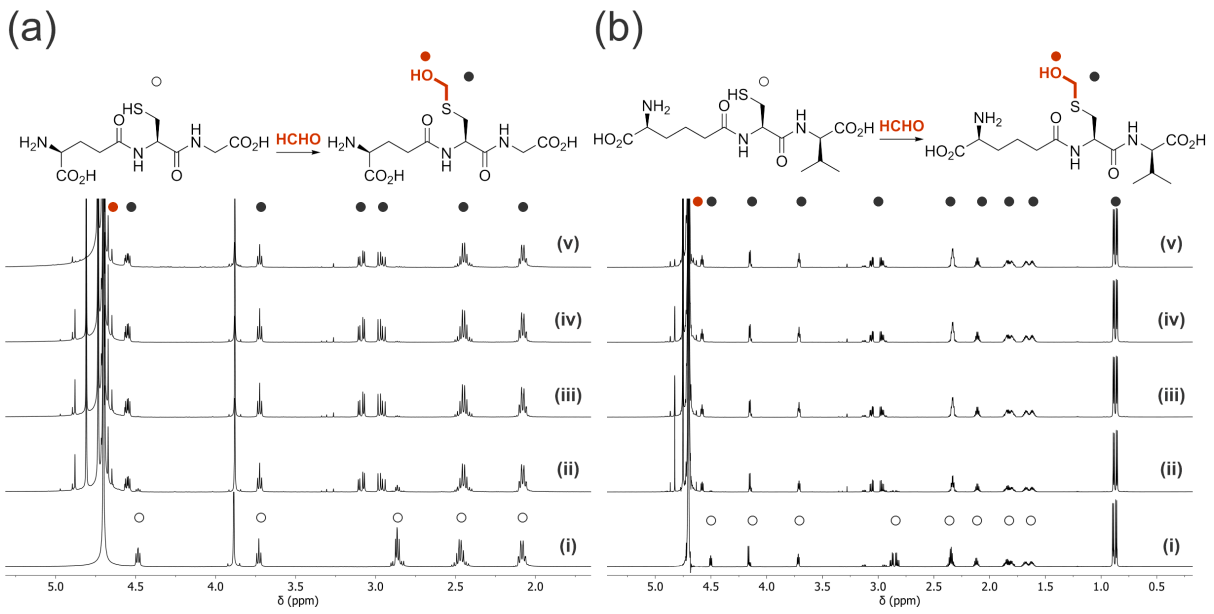



**Figure 7.41:** Reactions between HCHO and the peptides glutathione and L-δ-(2-aminoadipoyl)-L-cysteinyl-D-valine (**ACV**). (a) (i) Glutathione (4.34 mM), (ii) 5 minutes after addition of HCHO (10 equivalents), (iii) 15 mintues, (iv) 25 minutes, (v) 48 hours; (b) (i) **ACV** (4.34 mM), (ii) 5 minutes after addition of HCHO (10 equivalents), (iii) 15 mintues, (iv) 25 minutes, (v) 48 hours.

## 7.9 Reductive $N^{\alpha}$-Methylation

During these studies, no direct evidence for the formation of modifications on α-amino group was obtained with any of the amino acids. Given the observed products from reactions between HCHO and amino acids, these modifications, including imines and hemiaminals, are likely intermediates *en route* to the observed cyclic products. The intermediates are presumably transient and/or unstable, precluding direct observation in our NMR experiments. To investigate these intermediates, established *N*-methylation procedures were used to then "trap" these transient intermediates. To a mixture of either Cys, Ala, or His (4.34 mM) with an excess of sodium cyanoborohydride ($NaCNBH_3$, 5 equivalents), stoichiometric amounts of HCHO (1 equivalent) were added. With Ala and His, both mono- and di-$N^{\alpha}$-methylated products (**39a**/**b** and **40a**/**b**) were observed, within an hour (Figure 7.42 (a) and (b)). No methylation or modification at the imidazole ring in His was observed, suggesting imine formation does not occur under the tested conditions. Exposing Cys to these reductive conditions, resulted in the formation of $N^{\alpha}$-methylated thiazolidine **41** (Figure 7.41 (c)). It is unclear whether cyclisation or methylation occurs first.

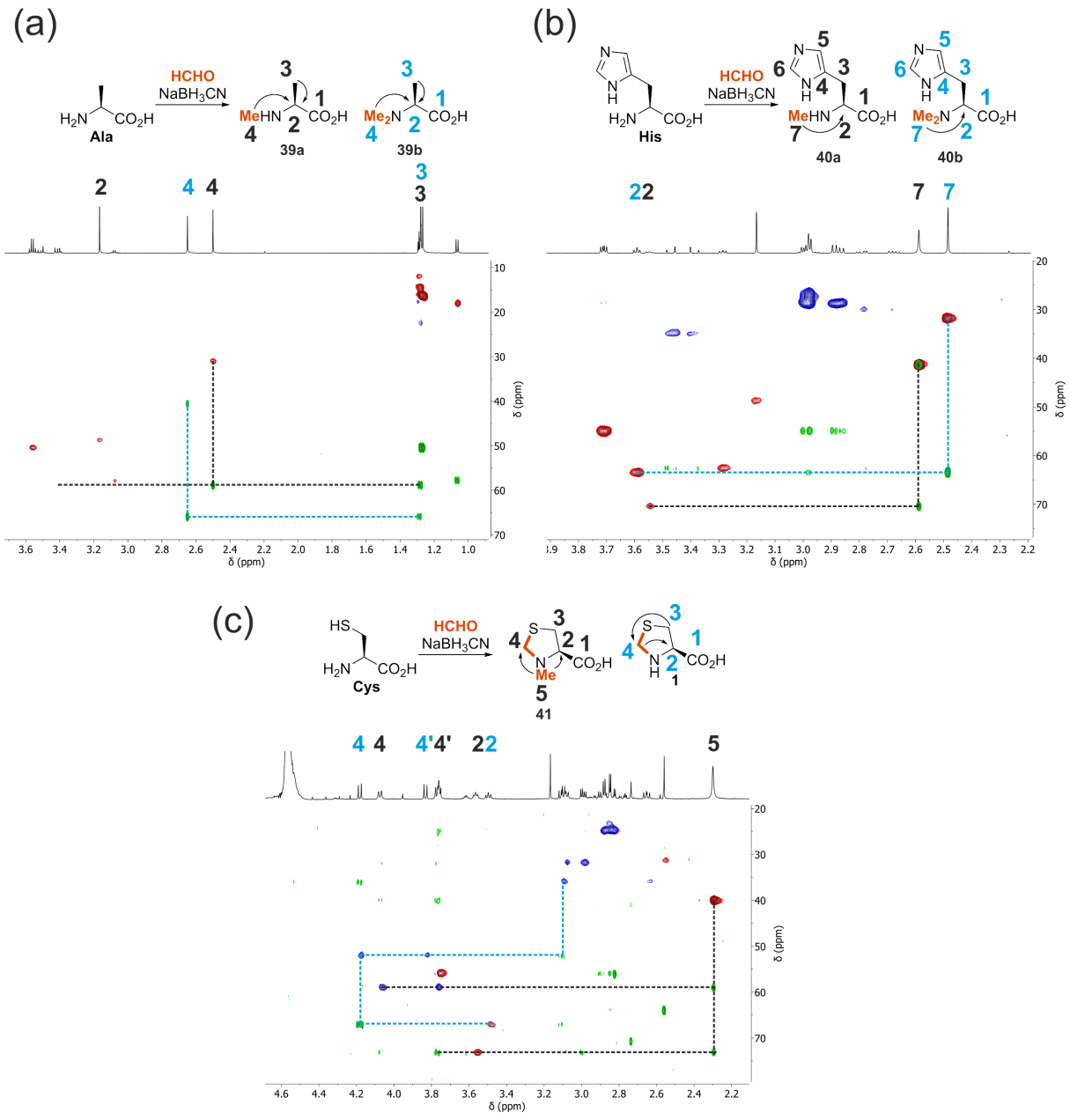



**Figure 7.41:** HSQC (red/blue) and HMBC (green) correlations of mixtures containing alanine, histidine or cysteine with $NaBH_3CN$ and HCHO, and analysed after 6 hours. (a) A solution of alanine (4.34 mM) and $NaBH_3CN$ (5 equivalents) was allowed to react with HCHO (1.1 equivalent), forming mono- and di-*N*-methylated alanine, (b) a solution of histidine (4.34 mM) and $NaBH_3CN$ (5 equivalents) was allowed to react with HCHO (1.1 equivalent), forming mono- and di-*N*-methylated histidine, (c) a solution of cysteine (4.34 mM) and $NaBH_3CN$ (5 equivalents) was allowed to react with HCHO (1.1 equivalent), forming thiazolidine **1** and *N*-methylated thiazolidine **1**. Complementary $^{1}H$ and $^{13}C$ NMR spectra are plotted on the respective axes.

## 7.10 Discussion

The experimental work described in this chapter concerned the reactions of HCHO and related electrophiles with common proteinogenic and other biologically relevant amino acids. The experiments employed a systematic approach with analysis principally by NMR. The investigations identified numerous products including previously described and undescribed structures. They clearly reveal the potential for complexity, even in reactions between the simplest aldehyde, HCHO, and simple biologically relevant small molecules. This level of complexity is illustrated by the reaction of Orn with HCHO (Figure 7.42).

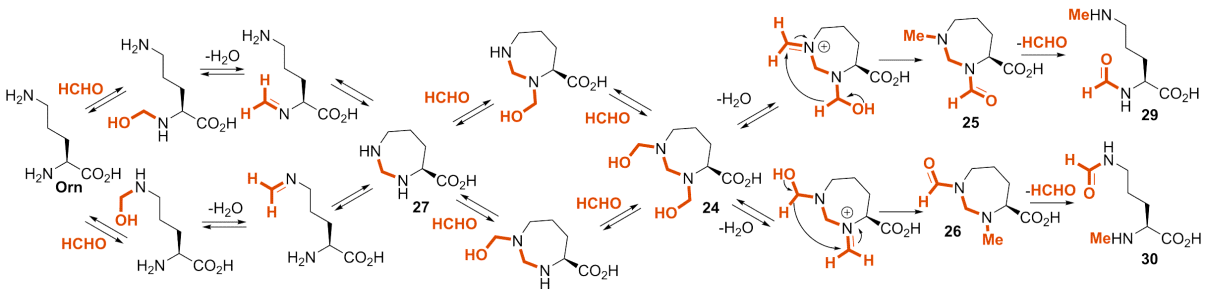



**Figure 7.42:** Reaction of Orn with HCHO in the presence of base results in the formation of several cyclic products **24** – **27** and linear products **29** and **30**. Shown here are potential mechanistic pathways for the formation of these products.

Of the investigated common proteinogenic amino acids, Cys appears to react most efficiently with HCHO, initially to form an *S*-hydroxymethyl adduct **2**, and then a stable thiazolidine product **1** (Figure 7.43 (a)). In peptides, such as glutathione or structurally related tripeptide **ACV**, the Cys residue rapidly reacts to form a thiohemiacetal (Figure 7.43 (b)). Formation of the hemithioacetal in glutathione is part of the HCHO detoxification process.[54] Apart from this species and the thiazolidine product, additional cyclic products have been reported in the reaction between HCHO and glutathione.[55,56] The biological significance of these different states and cyclic ring structures is not defined but could have implications for HCHO toxicity. The results also reveal subtle differences in the reactions of amino acid side chains, as observed with Thr/*allo*-Thr, the products and adducts different stabilities. Disulfide formation was not observed when Cys or one of the Cys-containing peptides were treated with HCHO. Under aerobic conditions, these results suggest the potential for HCHO to directly influence redox levels in cells by influencing the extent of disulfide bond formation. The data reveals the potential for complexity with these simple and dynamic systems.

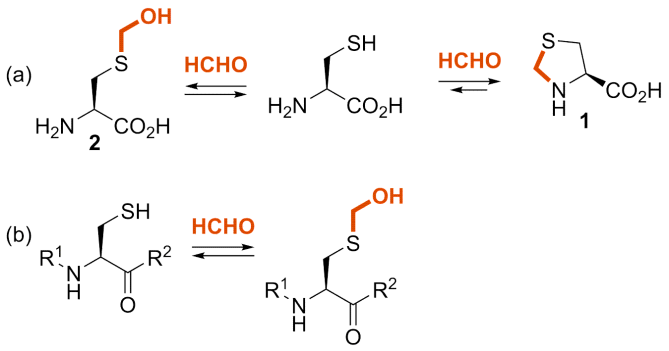



**Figure 7.43:** Reaction of HCHO with Cys, Cys-containing tripeptide glutathione, or L-δ-(2-aminoadipoyl)-L-cysteinyl-D-valine (**ACV**). (a) The reversible reaction of Cys with HCHO, showing the formation of kinetically more stable hydroxymethylated **2** and the thermodynamically more favourable **1**; (b) The reversible reaction of tripeptide glutathione ($R^1$ = γ-L-Glutamyl, $R^2$ = Gly), or **ACV** ($R^1$ = δ-adipoyl, $R^2$ = D-valine) with HCHO, resulting in the formation of *S*-hydroxymethyl adducts of the two peptides.

During these studies, no direct evidence for $N^{\alpha}$-modification, such as imine or hemiaminal formation, was observed under the tested conditions. This is consistent with previous NMR studies reporting on the mechanism of $N^{\varepsilon}$-methyllysine demethylation by lysine demethylases.[53] A hemiaminal is proposed to be formed as a consequence of catalytic oxidation of the methyl group, followed by rapid fragmentation. Although no direct evidence was found for the formation of hemiaminals or imines, indirect evidence, such as broadening of resonances with Lys, Ser, Thr, *allo*-Thr, and Hse, suggest these species do exist. These hemiaminal intermediates have indeed been observed in other studies.[22,45,55] Formation of thiazolidine **1**, in the reaction between Cys and HCHO, likely proceeds via an unobserved $N^{\alpha}$-modified Schiff base intermediate, and not via the observed thiohemiaminal.[31] A reaction of HCHO with amino acids (i.e. Cys, Ala, His) in presence of $NaCNBH_3$ resulted in the observation of $N^{\alpha}$-methylated Cys, Ala and His. These derive from reduction of unobserved hemiaminals and imines formed at the $N^{\alpha}$-amino group, suggesting these species do indeed exist.

Tris buffer reacts with HCHO in aqueous solutions.[57] Detailed reports on the products have not been reported,[58] though monocyclic and bicyclic oxazolidine structures have been proposed as possible products of the reaction by Tris with HCHO (Figure 7.44 (a)).[57,59] Hydroxymethylated Tris **42** is a stable and observable product,[22] in contrast the cyclic oxazolidine structures **43** and **44** have so far not been detected. Tris, and other Good's type buffers, are capable of releasing HCHO when in presence of •OH radicals.[60] Additionally, Meldonium, a competitive substrate / inhibitor of γ-butyrobetaine hydroxylase, forms a stabile

hemiaminal after oxidations at the enzyme active site (Figure 7.44 (b)).[61] The hemiaminal is in equilibrium with a cyclic imidazole structure. As proteins, DNA, nucleic acids, and many more biomolecules contain multiple nucleophilic sites, these molecules could sequester HCHO and act as reservoirs. Indeed HCHO is an established tool in studying chromatin complexes.[57]

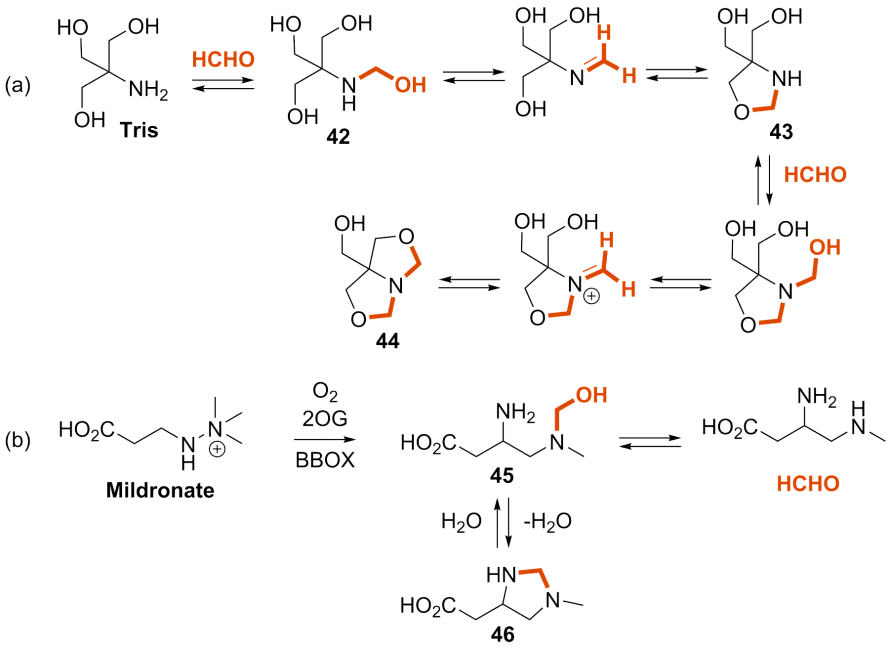



**Figure 7.44:** The reaction of tris(hydroxymethyl)aminomethane (Tris) with HCHO and the oxidation of Meldonium (Mildronate) by γ-butyrobetaine dioxygenase (BBOX) to form a hydroxymethylated product. (a) The reaction of Tris with HCHO, forming up to two oxazolidine rings in the final product, is used to scavenge HCHO in biological samples; (b) the oxidation of Mildronate by BBOX results in C-3 oxidation, followed by Stevens rearrangement to form a *N*-hydroxymethylated butyric acid **45** in the active site, which can cyclise to form an imidazole ring, or release HCHO. Other products are produced.

A range of carbonyl compounds, including HCHO, acetaldehyde, acetone, glyoxal and methylglyoxal, were investigated for reaction with Cys. From this selection, HCHO appears to be most reactive, resulting in the formation of thiazolidine **1**, as a stable product (Figure 7.45). These observations place HCHO, with its small size and electrophilic character and its presence as side product in numerous biochemical reactions, in a unique position for reaction with biomolecules.

A caveat on the work discussed in this chapter is that an excess of HCHO was used. In some cases, prolonged reaction times were also required to result in the formation of adducts. Reports on the *in vivo* concentrations of HCHO differ, reflecting the difficulty in its analysis.[1,48,51] This is especially true for local

levels of HCHO. It may well be that localised concentrations of HCHO, at cellular or subcellular level are higher than generally perceived. Chromatin, for example, is highly dynamic in its methylation level; HCHO is a side product of demethylation.[53] Depending on the local structural context, additional constraints on stereoelectronics or concentrations may alter the rates of reaction or stability of products. Catalytic processes perhaps involving enzymes (or other catalysts) might contribute in a context-dependent manner. Notably, the enhanced reactivity of HCHO towards biological nucleophiles is observed, despite HCHO largely being available its less reactive hydrated form (>98%). The aldehyde form, which is normally presumed to be the reactive species, is present as a minor fraction (<0.1%).

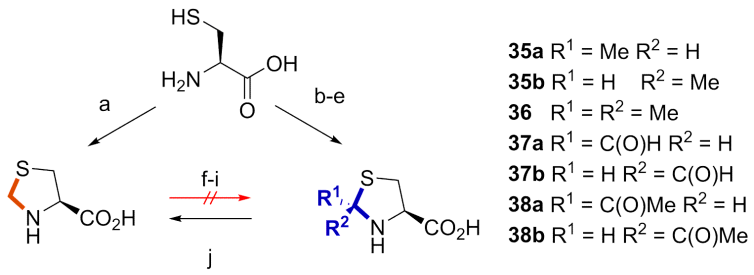



**Figure 7.45:** Electrophile competition assays in which Cys is reacted with different biologically-relevant carbonyl compounds. Reactions between Cys (4.34 mM) and an electrophile (10 equivalents) are completed first, and subsequently a larger excess of a second electrophile (100 equivalents) is added. (a) HCHO (10 equivalents); (b) acetaldehyde (10 equivalents); (c) acetone (10 equivalents); (d) glyoxal (10 equivalents); (e) methylglyoxal (10 equivalents); (f) acetaldehyde (100 equivalents); (g) acetone (100 equivalents); (h) glyoxal (100 equivalents); (i) methylglyoxal (100 equivalents); (j) HCHO (100 equivalents).

The stability of the adducts formed in these reactions was evaluated by removing free HCHO from the solution using DCH as an HCHO scavenger, repeated lyophilisation cycles, or exchange with labelled H$^{13}$CHO. A range of stabilities was observed, showing rapid exchange in oxazolidine systems. Bridging methylenes in *N,N*-cyclised products, hemiaminals, or thiohemiaminals have been shown to be reversible modifications. Some of the modifications encountered with Orn and Lys, such as methylation and formylation, are likely near-irreversible. The dynamic nature of most of the adducts renders their observation in cells challenging. Using isolated systems is therefore important for the understanding of the reactions of HCHO with the complex mixture of biological molecules in cells.

Chromatin, with its basic elements such as histones and nucleobases, has substantial potential to react with HCHO. Indeed in this study, it is shown that HCHO has the potential to result in methylation of amino

acids, in particular Lys, and by extrapolation proteins including histones. Prior studies on nucleic acids and DNA have shown the potential of HCHO to form relatively stable *N*-hydroxymethyl adducts, with significant half-lives, suggesting these could accumulate over time.[22,62,63] Moreover, Mannich-type chemistry at chromatin can lead to the crosslinking of proteins and DNA.[64,65] This is of current interest as multiple enzymes have been identified to regulate modifications, including methylation, of histone proteins and nucleic acids. This could mean HCHO alone has the potential to regulate chromatin structure and function.

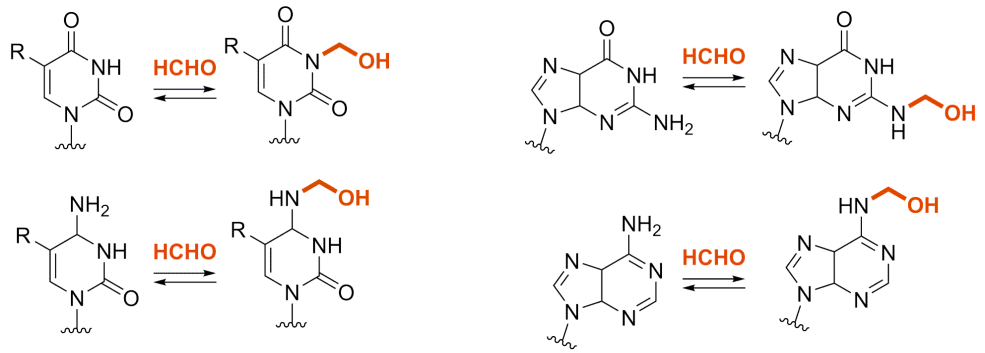



**Figure 7.46:** Overview of modifications observed in reactions of nucleotides with HCHO. R = H, $CH_3$ or $CH_2OH$.[22]

In a broader context, the ability of HCHO to react at proteins or interaction regions or enzymatic active sites reveals the potential for it to alter non-covalent interactions and regulate enzymatic activity. By reacting with interaction regions, non-covalent interactions are altered (e. g. masking interaction regions or establishing covalent interactions). These potential nucleophilic sites also adjust the available levels of HCHO.

The ability of HCHO to modify chromatin biochemistry by reacting with basic amino acids (Orn, His, Arg, Lys) to give cyclic products, and hemiaminal adducts, and occasionally *N*-methylated and formylated products, is of significant interest. Such species may be relevant to the long term health risks associated with environmental exposure to high levels of HCHO, or perhaps even exposure via HCHO-producing pro-drugs.[61,66,67] It is also possible that such modifications occur in healthy cells, due to locally high HCHO levels as a result of release in biochemical reactions (e.g. increased demethylase activity). The reactions described in this chapter would be of potential interest for any early stage evolutionary biochemistry studies as they describe the fundamental systems in current cellular biology in reaction with one another.

Moreover, some of the amino acids with *N*-containing side chains (e. g. Arg, Orn, His, Trp) were able to react with HCHO to give cyclic structures. Pictet-Spengler-type chemistry with His and Trp resulted in the formation of *C*,*N*-cyclised products, which seem to be irreversible products. The complex ring structures formed with Orn and Arg are also particularly notable.

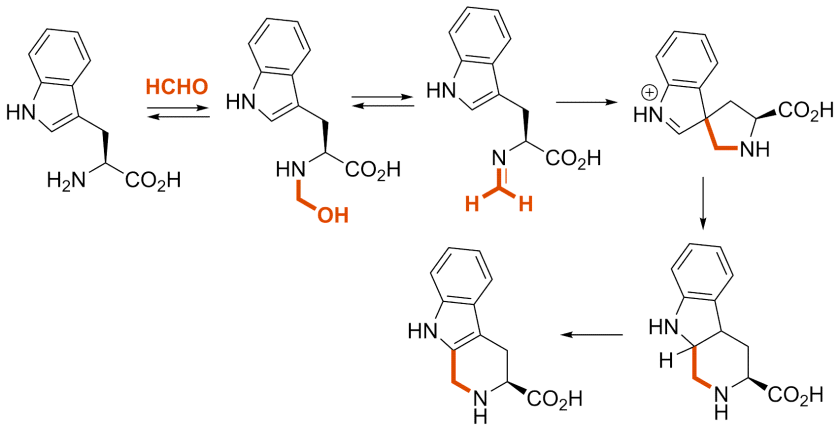



**Figure 7.47:** Pictet-Spengler reaction of tryptophan with HCHO resulting in the formation of **22**.

The possibility of the reactions of amino acids (and by implication peptides and proteins) with HCHO also have the potential to be affected by other biomolecules. For example, collagen Lys C-5 hydroxylation allows for oxazolidine formation. Given Jumonji Domain Containing 6 (JMJD6) is reported to be a histone Lys hydroxylase, such reactions have potential to be involved in the regulation of transcription.

## 7.11 Conclusions

Overall the results described in this chapter reveal that the reactions between HCHO and amino acids can result in the formation of complex mixtures of dynamic products. Together with previous studies, these reactions reveal the enormous potential for complexity in the reactions of HCHO in biochemistry. Nucleophilic sites present on proteins or DNA form potential sites for modification by HCHO. Currently the reactions of HCHO with histone tails, which have a high density of nucleophilic residues, are being investigated. Although these combined data suggest chromatin would be capable to react with HCHO, its involvement in regulation of normal healthy biology is presently an open question. Deregulation of HCHO metabolism has been linked to the development of genetic diseases.[68] HCHO is used as a pain initiator in animal models. The response is likely to be at least in part triggered via reaction of a Cys residue in a transient receptor potential sensory protein.[69] HCHO makes its way into everyday life as it is present in food[70] (e. g. spinacine),[71] as a preservative,[72] in pro-drugs[66] or drug metabolites (e. g. Mildronate[61]) and in natural products (e. g. tubulysin)[73]. Thus, the reactions outcomes described in this chapter may have societal relevance. Much of the work described in this chapter has been reported in a recent publication.[74]

## 7.12 Methods

### 7.12.1 General methods

Alanine monohydrate, arginine, asparagine, 1,3-cyclohexadione, cysteine, dimedone, 5-hydroxylysine hydrochloride salt, homocysteine, D/L-homoserine, lysine hydrochloride salt, *N*ε-methyllysine hydrochloride salt, *N*ε,*N*ε-dimethyllysine hydrochloride salt, ornithine hydrochloride salt, proline, serine, threonine, *allo*-threonine, tryptophan, and tyrosine, were from commercial suppliers and used without purification. All amino acids were used in their (L)/(2S)-stereochemical form, unless stated otherwise. Samples were prepared in $D_2O$ (purity 99.9 atom % D), unless stated otherwise. Starting materials and products formed during the reactions were monitored by NMR and characterized using a combination of 1D and 2D techniques. Samples were recorded at 298 K, in a 5 mm tube, unless stated otherwise, using a Bruker AVII 500 equipped with a TXI H/F/C probe, Bruker AVIII HD 500 equipped with a BBFO probe, Bruker AVII 500 equipped with a He CryoProbe, Bruker AVIII HD 600 equipped with a BB-F/H $N_2$ CryoProbe or a Bruker AVIII HD 700 equipped with a TCI H/C/N He CryoProbe, at their respective frequencies.

### 7.12.2 Experimental methods

#### 7.12.2.1 Standard conditions reaction setup

The amino acids (0.0247 mmol) were weighted and dissolved in $D_2O$ (1 mL) to give a 25 mM amino acid stock solution. HCHO (30.0 mg, 1.0 mmol) was suspended in $D_2O$ (3 mL) and heated with a hairdryer until a clear colourless solution was obtained to give a 335 mM stock solution of HCHO. In the cases of acetaldehyde, acetone, glyoxal or methylglyoxal an equimolar solution was created by addition of an appropriate volume of one of the desired aldehydes to $D_2O$ (3 mL). A typical sample consisted of the amino acid stock solution (100 μL, 2.47 μmol, 1 equivalent) and $D_2O$ (400 μL); the solution was thoroughly mixed and a $^1H$ NMR spectrum was recorded. In some cases, DCl (1 M) or NaOD (1 M) was added; the solutions were mixed and $^1H$ NMR spectrum was recorded. Subsequently HCHO (7-70 μL, 2.47-24.7 μmol 1-10 equivalents) was added and the progress of the reaction was monitored over time. Variable temperature studies required the samples to be adjusted to the appropriate temperature before the addition of

electrophile. Competition experiments were conducted using amino acid stock solution (100 μL), with another amino acid stock solution (100 μL) and $D_2O$ (300 μL), with a limiting amount of HCHO (1 equivalent).

#### 7.12.2.2 Adjusted pH reactions

For pH-adjusted reactions, an appropriate stock solution was diluted to a minimum volume of 4 mL. The pH was then adjusted to: 5.60, 7.10 or 8.60, corresponding to pD 6.00, 7.50 or 9.00, using NaOD and DCl (0.1 M and 0.01 M). A HCHO stock solution was prepared by suspending paraformaldehyde (30.0 mg, 1 mmol) in $D_2O$ and subsequently adjusting the pH with NaOD and DCl (0.1 M and 0.01 M) to the desired pH (total volume eventually 3 mL). Amino acid stock solution (400 μL, 2.47 μmol) was mixed with $D_2O$ (100 μL); and NMR spectra were recorded. Subsequently, HCHO (7-70 μL, 2.46-24.6 μmol, 1-10 equivalents) was added and reaction progress was monitored over time (maximum 48 hours).

#### 7.12.2.3 pH dependent equilibria experiments

To a buffered solution (144 μL, 25 mM phosphate buffer, at pH = 5.9, 6.6, 7.6, 8.6, 9.4, 10.6, 11.6), an amino acid (16 μL, 50 mM stock solution; in the case of 2 amino acids: 2 × 8 μL was used) was added. Subsequently, HCHO (2.38 uL, 3.35 M, 10 equivalents) was then added. The mixture was allowed to equilibrate in a 3 mm tube, for at least 20 minutes before recording a $^1$H NMR spectrum.

#### 7.12.2.4 Stability assays using 1,3-cyclohexadione

Upon completion of a standard reaction, as monitored by $^1$H NMR, a solution of 1,3-cyclohexadione (22.4 mg, 0.2 mmol, in 180 μL, 1.1 M; 85 uL, 40 equivalents) was added to the reaction mixture. Formation of the parent amino acid formation was followed by NMR over time.

#### 7.12.2.5 Lyophilisation to test stability of HCHO adducts

Upon completion of a standard reaction, as monitored by $^{1}H$ NMR, the solution was frozen in liquid $N_2$, then freeze dried overnight. The mixture was then redissolved in $D_2O$ and allowed to equilibrate for 1 hour, then analyzed by $^{1}H$ NMR. The sample was then put through a cycle of lyophilisation and redissolving several times (5-8).

#### 7.12.2.6 Using $^{13}C$ labeled HCHO to test for exchange of HCHO adducts

A reaction mixture with amino acid was stirred under standard conditions with unlabelled HCHO (5 equivalents), until the species of interest was observed. Subsequently, labelled $H^{13}CHO$ (5 equivalents) was added to the solution and the progress of the reaction was monitored over time.

#### 7.12.2.7 Competition assays

A stock solution of one amino acid (25 mM) was prepared and added (100 μL) to $D_2O$ (300 μL). Cysteine (100 μL, or another amino acid) was then added, to give an equimolar solution (50 mM, 1 : 1). This was followed by the addition of HCHO (1 equivalent, 3.35 M stock solution). Reactions were monitored for 16 hours until all HCHO was consumed, as observed by $^{1}H$ NMR.

For competition assays with more than 4 amino acids, higher stock solution concentrations (200 mM) were used.

#### 7.12.2.8 Spectator amino acid experiments

A solution of cysteine (5 mM) and a 'spectator' amino acid (25 mM), in phosphate buffer (50 mM, pH 7.10, in >95% $D_2O$) was made and HCHO (1 equivalent with respect to cysteine, 3.35 M stock solution) was added. The reaction was followed by $^{1}H$ NMR.

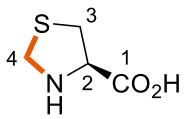
#### 7.12.2.9 Reductive amination

A mixture of amino acid (5 mM) and $NaBH_3CN$ (5 equivalents) in $D_2O$ was prepared. Subsequently, HCHO (1.1 equivalents, 3.35 M stock solution) was added. The reaction was followed and upon completion analysed by $^1H$ NMR.

### 7.12.3 Product characterisation

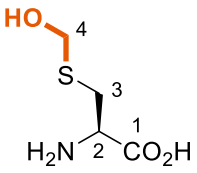
#### 7.12.3.1 Cysteine products

**(*S*)-4-Carboxy-1,3-thiazolidine 1**:




The named compound was observed when cysteine (4.34 mM) was reacted with HCHO (10 equivalents) and allowed to react for 1 hour. $^1H$ NMR (700 MHz, $D_2O$) δ: 4.33 (d, 10.0 Hz, 1H, $H_4$) 4.29 (dd, $J$ = 7.5, 6.0 Hz, 1H, $H_2$), 4.21 (d, $J$ = 10.0 Hz, 2H, $H_{4'}$), 3.29 (dd, 1H, 12.0, 7.5 Hz, $H_3$), 3.18 (dd, 1H, 12.0, 7.5 Hz, $H_{3'}$). $^{13}C$ NMR (176 MHz, $D_2O$) δ: 172.0, 63.9, 48.6, 32.9. ESI-MS calculated for $C_4H_8O_2NS$ $[M+1]^+$ 134.02703, found 134.02704. These assignments are in agreement with previously reported data.[75]

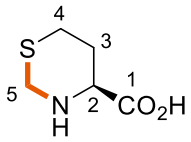
**(*S*)-Hydroxymethyl-L-cysteine 2:**




The named compound was observed when cysteine (4.34 mM) was reacted with HCHO (2 equivalents), and allowed to react for 30 minutes, in presence of acid (DCl, 2 equivalents). $^1H$ NMR (700 MHz, $D_2O$) δ: 4.64 (d, $J$ = 12.0 Hz, 2H, $H_4$), 4.60 (d, $J$ = 12.0 Hz, 2H, $H_{4'}$), 4.22 (dd, $J$ = 7.0, 4.0 Hz, 1H, $H_2$), 3.15 (dd, $J$ = 15.5, 4.0 Hz, 2H, $H_3$). $^{13}C$ NMR (176 MHz, $D_2O$) δ: 170.4, 65.1, 52.7, 31.0.

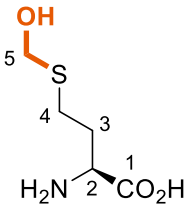
### 7.12.3.2 Homocysteine products

**(*S*/*R*)-4-Carboxy-1,3-thiazanane 3:**




The named compound was observed when homocysteine (4.34 mM) was reacted with HCHO (10 equivalents) and allowed to react for 1 hour. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.16 (d, $J$ = 13.0 Hz, 1H, $H_{5}$), 4.10 (d, $J$ = 13.0 Hz, 1H, $H_{5'}$), 3.52 (dd, $J$ = 12.0, 3.0 Hz, 1H, $H_{2}$), 2.85 (dt, $J$ = 14.0, 3.0 Hz, 1H, $H_{4}$), 2.69 (dt, $J$ = 14.0, 3.0 Hz, 1H, $H_{4'}$), 2.39 (dq, $J$ = 15.0, 3.5 Hz, 1H, $H_{3}$), 1.87 (dq, $J$ = 15.0, 3.5 Hz, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 170.0, 59.4, 45.2, 28.1, 26.0. These data are in agreement with those reported.[34]

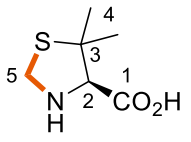
**(*S*/*R*)-Hydroxymethyl-L/D-homocysteine 4:**




The named compound was observed when homocysteine (4.34 mM) was reacted with HCHO (2 equivalents), and allowed to react for 30 minutes, in presence of (DCl, 2 equivalents). $^{1}$H NMR (500 MHz, $D_2O$) δ: 4.67 (s, 2H, $H_{5}$), 4.15 (t, $J$ = 6.5 Hz, 1H, $H_{2}$), 2.75 (t, $J$ = 7.5 Hz, 2H, $H_{4}$), 2.28 – 2.21 (dt, $J$ = 14.5, 7.5 Hz, 1H, $H_{3}$), 2.14 (dt, $J$ = 14.5, 7.5 Hz, 1H, $H_{3'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 171.6, 64.6, 51.6, 30.0, 25.7.

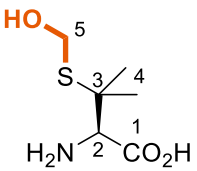
### 7.12.3.3 Penicillamine products

**(*R*)-5,5-Dimethylthiazolidine-4-carboxylic acid 5:**




The named compound was observed when penicillamine (4.34 mM) was reacted with HCHO (10 equivalents) and allowed to react for 30 minutes. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.43 (d, $J$ = 10.5 Hz, 1H, $H_{6}$), 4.36 (d, $J$ = 10.5 Hz, 1H, $H_{6'}$), 3.91 (s, 1H, $H_{2}$), 1.61 (s, 3H, $H_{4}$), 1.37 (s, 3H, $H_{4'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 170.0, 73.2, 52.9, 27.3, 25.4. These data are in agreement with those reported.[76]

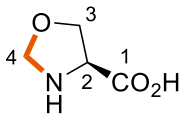
**(*S*)-Hydroxymethyl-L-penicillamine 6:**




The named compound was observed when penicillamine (4.34 mM) was reacted with HCHO (10 equivalents) and allowed to react for 30 minutes. $^1$H NMR (700 MHz, $D_2O$) δ: 4.87 (d, *J* = 13.0 Hz, 1H, $H_6$), 4.75 (d, *J* = 13.0 Hz, 1H, $H_{6'}$), 4.07 (s, 1H, $H_2$), 1.59 (s, 3H, $H_4$), 1.38 (s, 3H, $H_{4'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 169.8, 62.0, 61.3, 46.2, 26.4, 24.0.

### 7.12.3.4 Serine product

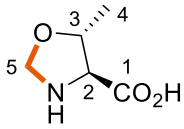
**(*S*)-4-Carboxy-1,3-oxazolidine 7:**




The named compound was observed when serine (4.34 mM) was reacted with HCHO (10 equivalents) and allowed to react for 20 minutes, in presence of base (NaOD, 1 equivalent). $^1$H NMR (600 MHz, $D_2O$) δ: 4.49 (d, *J* = 6.0 Hz, 1H, $H_4$), 3.99 (d, *J* = 6.0 Hz, 1H, $H_{4'}$), 3.73 (t, *J* = 7.0 Hz, 1H, $H_3$), 3.54 (t, *J* = 7.0, Hz, 1H, $H_2$), 3.38 (t, *J* = 7.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 178.2, 81.0, 67.7, 60.6. These data are in agreement with those reported.[77]

### 7.12.3.5 Threonine product

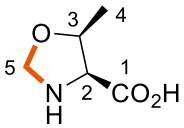
**(*4S,5R*)-5-methyloxazolidine-4-carboxylic acid 8:**




The named compound was observed when threonine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 20 minutes. $^1$H NMR (700 MHz, $D_2O$) δ: 5.01 (d, *J* = 6.0 Hz, 1H, $H_5$), 4.64 (d, *J* = 6.0 Hz, 1H, $H_{5'}$), 4.17 – 4.13 (m, 1H, $H_2$), 3.72 (d, 6.0 Hz, 1H, $H_3$), 1.43 (d, *J* = 6.0 Hz, 3H, $H_4$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 177.8, 79.8, 76.8, 67.9, 18.8. These data are in agreement with those reported.[77]

#### 7.12.3.6 *allo*-Threonine product

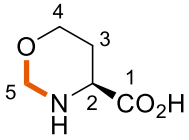
**(*4S,5S*)-5-methyloxazolidine-4-carboxylic acid 9:**




The named compound was observed when *allo*-threonine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 20 minutes. $^{1}$H NMR (700 MHz, $D_2O$) δ: 5.02 (d, *J* = 6.0 Hz, 1H, $H_5$), 4.51 (d, *J* = 6.0 Hz, 1H, $H_{5'}$), 4.39 (t, *J* = 6.5 Hz, 1H, $H_2$), 4.09 (d, *J* = 7.0 Hz, 1H, $H_3$), 1.20 (d, *J* = 6.0 Hz, 3H, $H_4$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 170.6, 76.8, 76.7, 60.9, 14.5.

#### 7.12.3.7 Homoserine product

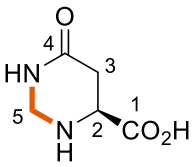
**(*S*)-4-Carboxy-1,3-oxazinane 10:**




The named compound was observed when homoserine (4.34 mM) was reacted with HCHO (10 equivalents), and allowed to react for 20 minutes. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.86 (d, *J* = 9.5 Hz, 1H, $H_5$), 4.41 (d, *J* = 9.5 Hz, 1H, $H_{5'}$), 4.02 (dt, *J* = 11.5, 2.5 Hz, 1H, $H_4$), 3.76 (dd, *J* = 12.0, 4.0 Hz, 1H, $H_2$), 3.63 – 3.61 (m, 1H, $H_{4'}$), 2.02 – 1.99 (m, 1H, $H_3$), 1.89 – 1.84 (m, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 174.3, 75.3, 66.9, 56.5, 26.2.

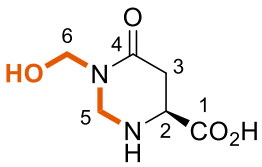
#### 7.12.3.8 Asparagine products

**(*S*)-6-Oxohexahydropyrimidine-4-carboxylic acid 11:**



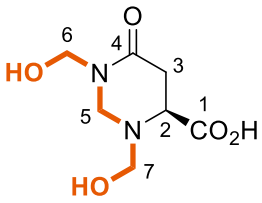

The named compound was observed when asparagine (4.34 mM) was reacted with HCHO (10 equivalents), for 12 hours. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.55 (d, *J* = 11.5 Hz, 1H, $H_5$), 4.43 (d, *J* = 11.5, Hz, 1H, $H_{5'}$), 4.05 (dd, *J* = 10.0, 6.5 Hz, 1H, $H_2$), 2.86 (dd, *J* = 17.5, 10.0 Hz, 1H, $H_3$), 2.64 (dd, *J* = 17.5, 10.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 172.1, 171.0, 54.1, 52.6, 31.6. ESI-MS calculated for $C_5H_9O_3N_2$ $[M+1]^+$ 145.06077, found 145.06088. These data are in agreement with those reported.[42]

**(*S*)-1-(Hydroxymethyl)-6-oxohexahydropyrimidine-4-carboxylic acid 12:**




The named compound was observed when asparagine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 24 hours, in presence of base (NaOD, 1 equivalent). $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.70 (s, 2H, $H_6$), 4.30 (d, $J$ = 12.0 Hz, 1H, $H_5$), 4.21 (d, $J$ = 12.0 Hz, 1H, $H_{5'}$), 3.49 (dd, $J$ = 11.0, 5.5 Hz, 1H, $H_2$), 2.61 – 2.51 (m, 1H, $H_3$), 2.27 (dd, $J$ = 17.5, 11.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 178.2, 171.2, 66.9, 59.9, 55.9, 34.7. ESI-MS calculated for $C_6H_{11}O_4N_2$ $[M+1]^+$ 175.07133, found 175.07137.

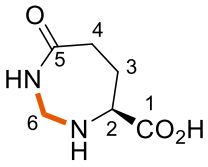
**(*S*)-1,3-Bis(hydroxymethyl)-6-oxohexahydropyrimidine-4-carboxylic acid 13:**




The named compound was observed when asparagine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 24 hours, in presence of base (NaOD, 1 equivalent). $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.72 (s, 2H, $H_6$), 4.40 (d, $J$ = 12.0 Hz, 1H, $H_5$), 4.37 (d, $J$ = 11.0 Hz, 2H, $H_7$), 4.33 (d, $J$ = 11.0 Hz, 2H, $H_{7'}$), 4.27 (d, $J$ = 12.0 Hz, 1H, $H_{5'}$), 3.66 (t, $J$ = 7.0 Hz, 1H, $H_2$), 2.61 – 2.51 (m, 2H, $H_3$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 178.2, 172.0, 77.2, 67.4, 62.8, 58.8, 33.5. ESI-MS calculated for $C_7H_{13}O_5N_2$ $[M+1]^+$ 205.08190, found 205.08169.

**7.12.3.9 Glutamine product**

**(*S*)-7-Oxo-1,3-diazepane-4-carboxylic acid 14:**

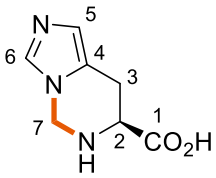



The named compound was observed when glutamine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 24 hours. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.59 (d, $J$ = 14.5 Hz, 1H, $H_6$), 4.51 (d, $J$ = 14.5 Hz, 1H, $H_{6'}$), 3.83 (dd, $J$ = 11.0, 3.5 Hz, 1H, $H_2$), 2.75 (dd, $J$ = 15.5, 8.5 Hz, 1H, $H_4$), 2.53 (dd, $J$ = 15.5, 8.5 Hz, 1H, $H_{4'}$), 2.33 – 2.29 (m, 1H, $H_3$), 1.92 – 1.86 (m, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 180.3, 172.6, 62.9, 53.0, 32.8, 23.6. ESI-MS

calculated for $C_7H_{13}O_5N_2$ $[M+1]^+$ 159.07642, found 159.07646. These assignments are in agreement with previously published data.[42]

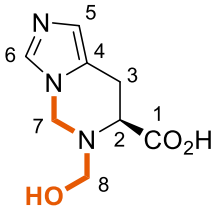
### 7.12.3.10 Histidine products

**(*S*)-5,6,7,8-Tetrahydroimidazo[1,5-c]pyrimidine-7-carboxylic acid 15:**



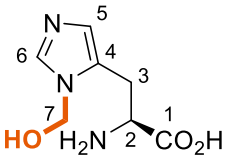
The named compound was observed when histidine (4.34 mM) was reacted with HCHO (10 equivalents), and allowed to react for 20 minutes. $^1$H NMR (500 MHz, $D_2O$) δ: 8.42 (s, 1H, $H_6$), 7.12 (s, 1H, $H_5$), 5.31 (d, $J$ = 12.5 Hz, 1H, $H_7$), 4.96 (d, $J$ = 12.5 Hz, 1H, $H_{7'}$), 3.61 (dd, $J$ = 11.0, 5.0 Hz, 1H, $H_2$), 3.18 (dd, $J$ = 16.5, 5.0, Hz, 1H, $H_3$), 2.74 (dd, $J$ = 16.5, 5.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 178.1, 130.7, 128.7, 115.8, 59.5, 54.6, 23.8.

**(*S*)-5-(Hydroxymethyl)-4,5,6,7-tetrahydro-3H-imidazo[4,5-c]pyridine-6-carboxylic acid 16:**



The named compound was observed when histidine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 1 day. $^1$H NMR (700 MHz, $D_2O$) δ: 8.46 (s, 1H, $H_6$), 7.10 (s, 1H, $H_5$) 5.22 (d, $J$ = 13.0, 1H, $H_7$), 5.18 (d, $J$ = 13.0, 1H, $H_{7'}$), 4.44 (d, 11.0 Hz, 1H, $H_8$), 4.40 (d, 11.0 Hz, 1H, $H_{8'}$), 3.94 (m, 1H, $H_2$), 3.14 – 3.04 (m, 2H, $H_3$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 176.1, 130.6, 128.9, 115.1, 77.4, 62.3, 56.9, 23.0.

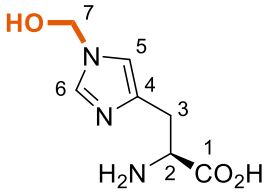
***N*$^{\pi}$-(Hydroxymethyl)-*L*-histidine 17:**



The named compound was observed when histidine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 1 day. $^1$H NMR (700 MHz, $D_2O$) δ (two stereoisomers, A/B): 8.94 (s, 1H, $H_{A6}$) 8.73 (s, 1H, $H_{B6}$), 7.43 (s, 1H, $H_{A5}$), 7.29 (s, 1H, $H_{B5}$), 5.53 (s, 2H, $H_{B7}$) 5.52 (s, 2H, $H_{A7}$), 3.96 – 3.92 (m, 1H, $H_{A2/B2}$), 3.29 – 3.22 (m, 2H, $H_{B3}$), 3.12 – 3.05 (m, 2H, $H_{A3}$). $^{13}$C NMR (176 MHz, $D_2O$) δ (both isomers (A/B)): 176.0 (2, A/B)*, 135.5 (A),

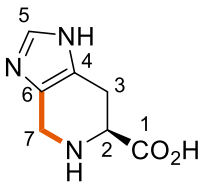
135.2 (B), 129.3 (A), 120.4 (A), 128.2 (B), 119.2 (B), 70.3 (A), 69.9 (B), 53.0 (2, A/B)*, 25.0 (2, A/B)*. * = couplings found using HMBC/HSQC, chemical shift estimate is being reported.

### *N*$^{\tau}$-(Hydroxymethyl)-*L*-histidine 18:



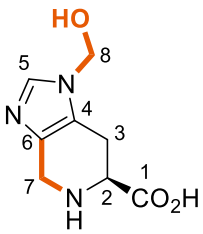

The named compound was observed when histidine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 1 day. $^1$H NMR (700 MHz, $D_2O$) δ (two stereoisomers, A/B): 8.68 (s, 1H, $H_{B6}$) 8.65 (s, 1H, $H_{A6}$), 7.37 (s, 1H, $H_{B5}$), 7.22 (s, 1H, $H_{A5}$), 5.45 (s, 2H, $H_{B7}$) 5.43 (s, 2H, $H_{A7}$), 3.90 – 3.83 (m, 1H, $H_{A2/B2}$), 3.22 – 3.14 (m, 2H, $H_{B3}$). $^{13}$C NMR (176 MHz, $D_2O$) δ (both isomers (A/B)): 176.0 (2, A/B)*, 134.4 (B) 131.3 (A), 129.5 (A7), 128.5 (B), 119.0 (B), 116.7 (A), 72.0 (A), 71.9 (B), 53.0 (2, A/B)*, 24.4 (B), 21.0 (A). * = couplings found using HMBC/HSQC, chemical shift estimate is being reported.

### (*S*)-4,5,6,7-Tetrahydro-3H-imidazo[4,5-c]pyridine-6-carboxylic acid 19:




The named compound was observed when histidine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 1 day. $^1$H NMR (500 MHz, $D_2O$) δ: 7.66 (s, 1H, $H_5$), 4.22 (d, $J$ = 15.0 Hz, 1H, $H_7$), 4.18 (d, $J$ = 15.0 Hz, 1H, $H_{7'}$), 4.00 (dd, $J$ = 11.0, 5.5 Hz, 1H, $H_2$), 3.21 (dd, $J$ = 16.5, 5.0 Hz, 1H, $H_3$), 2.91 (dd, $J$ = 16.5, 5.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 173.2, 136.7, 124.6, 123.5, 56.7, 40.8, 22.9. These assignments are in agreement with previously published data.[44]

### (S)-1-(Hydroxymethyl)-4,5,6,7-tetrahydro-1H-imidazo[4,5-c]pyridine-6-carboxylic acid 20:

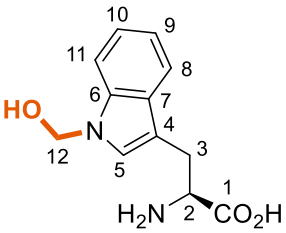



The named compound was observed when histidine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 1 day. $^1$H NMR (500 MHz, $D_2O$) δ: 7.71 (s, 1H, $H_5$), 5.35 (s, 2H, $H_8$), 4.24 - 4.20 (m, 2H, $H_7$), 4.04 (dd, $J$ = 10.5, 5.5 Hz,

1H, $H_{2}$), 3.32 – 3.25 (m, 1H, $H_{3}$), 2.99 – 2.92 (m, 1H, $H_{3'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 172.8, 138.6, 128.5, 122.6, 67.2, 56.3, 41.2, 21.6. These assignments are in agreement with previously published data.[44]

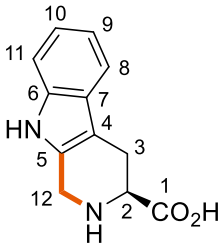
### 7.12.3.11 Tryptophan products

**1-(Hydroxymethyl)-*L*-tryptophan 21:**



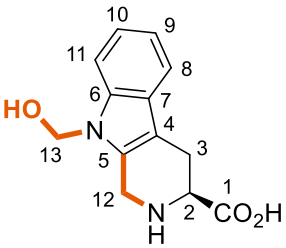

The named compound was observed when tryptophan (4.34 mM) was reacted with HCHO (10 equivalents), and allowed to react for 2 hours. $^{1}$H NMR (600 MHz, $D_2O$) δ: 7.66 (d, $J$ = 8.0 Hz, 1H, $H_{8}$), 7.53 (d, $J$ = 8.5 Hz, 1H, $H_{11}$), 7.30 – 7.27 (m, 1H, $H_{10}$), 7.25 (s, 1H, $H_{5}$), 7.21 – 7.18 (m, 1H, $H_{9}$), 5.57 (s, 1H, $H_{12}$), 3.99 (dd, $J$ = 9.0, 5.0 Hz, 1H, $H_{2}$), 3.39 (dd, $J$ = 15.0, 5.0 Hz 1H, $H_{3}$), 3.22 (dd, $J$ = 15.0, 5.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 174.3, 136.0, 128.0, 127.8, 122.7, 120.3, 118.9, 110.1, 108.7, 68.1, 54.9, 26.2.

**(*S*)-2,3,4,9-Tetrahydro-1H-pyrido[3,4-b]indole-3-carboxylic acid 22:**




The named compound was observed when tryptophan (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 2 days. $^{1}$H NMR (500 MHz, $D_2O$) δ: 7.56 (d, $J$ = 8.0 Hz, 1H, $H_{8}$), 7.49 (d, $J$ = 8.0 Hz, 1H, $H_{11}$), 7.20 – 7.16 (m, 1H, $H_{10}$), 7.12 – 7.08 (m, 1H, $H_{9}$), 4.46 (d, $J$ = 15.0 Hz, 1H, $H_{12}$), 4.38 (d, $J$ = 15.0 Hz, 1H, $H_{12'}$), 3.98 (dd, $J$ = 9.5, 4.0 Hz, 1H, $H_{2}$), 3.34 (dd, $J$ = 16.5, 5.5 Hz, 1H, $H_{3}$), 2.99 (dd, $J$ = 16.5, 5.5 Hz, 1H, $H_{3'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 173.7, 136.7, 126.3, 125.6, 123.0, 119.7, 118.2, 109.8, 105.9, 56.8, 39.5, 22.2.

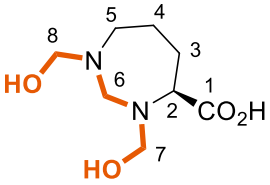
**(*S*)-9-(Hydroxymethyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole-3-carboxylic acid 23:**




The named compound was observed when tryptophan (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 2 days. $^{1}$H NMR (500 MHz, $D_2O$) δ: 7.54 (d, *J* = 8.0 Hz, 1H, $H_8$), 7.40 (d, *J* = 8.0 Hz, 1H, $H_{11}$), 7.20 – 7.16 (m, 1H, $H_{10}$), 7.12 – 7.08 (m, 1H, $H_9$), 5.50 (dd, *J* = 37.0, 11.5 Hz, 2H, $H_{13}$), 4.58 (d, *J* = 10.5 Hz, 1H, $H_{12}$), 4.43 (d, *J* = 10.5 Hz, 1H, $H_{12'}$), 4.01 (dd, *J* = 9.5, 4.0 Hz, 1H, $H_2$), 3.34 (dd, *J* = 16.5, 5.5 Hz, 1H, $H_3$), 2.99 (dd, *J* = 16.5, 5.5 Hz, 1H, $H_{3'}$). $^{13}$C NMR (126 MHz, $D_2O$) δ: 173.9, 136.5, 126.2, 125.7, 122.5, 120.6, 118.5, 111.6, 107.7, 65.25, 57.1, 40.39 22.28.

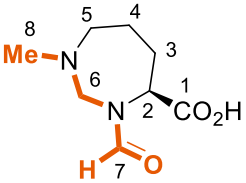
### 7.12.3.12 Ornithine products

**(*S*)-1,3-Bis(hydroxymethyl)-1,3-diazepane-4-carboxylic acid 24:**




The named compound was observed when ornithine (HCl salt, 43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 3 minutes, in presence of base (NaOD, 2 equivalents). $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.81 (d, *J* = 11.0 Hz, 1H, $H_8$), 4.75 (d, *J* = 11.0 Hz, 1H, $H_7$), 4.49 (d, *J* = 11.0 Hz, 1H, $H_{8'}$), 4.36 (d, *J* = 11.0 Hz, 1H, $H_{7'}$), 4.14 (d, *J* = 14.5 Hz, 1H, $H_6$), 3.87 (d, *J* = 14.5, 1H, $H_{6'}$), 3.37 (dd, *J* = 12.0, 4.5 Hz, 1H, $H_2$), 2.97 (dd, *J* = 14.5, 2.0 Hz, 1H, $H_5$), 2.65 (dd, *J* = 14.5, 2.0 Hz, 1H, $H_{5'}$), 2.21 – 2.13 (m, 1H, $H_3$), 2.05 – 1.98 (m, 1H, $H_4$), 1.84 – 1.78 (m, 1H, $H_{4'}$), 1.73 – 1.65 (m, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 183.0, 86.2, 84.9, 69.1, 62.7, 55.1, 32.0, 29.2.

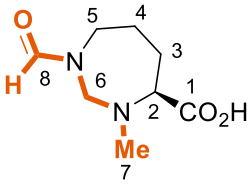
**(*S*)-3-Formyl-1-methyl-1,3-diazepane-4-carboxylic acid 25:**




The named compound was observed when ornithine (HCl salt, 43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 2 hours, in presence of base (NaOD, 2 equivalents). $^{1}$H NMR (600 MHz, $D_2O$) δ (two isomers, C and D): 8.30 (s, 1H, $H_{C7}$), 8.15 (s, 1H, $H_{D7}$), 5.14 (d, *J* = 13.0 Hz, 1H, $H_{D6}$), 4.77 – 4.73 (m, 1H, $H_{C6}$), 4.62 (d,

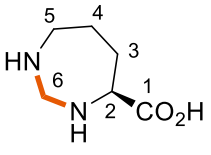
*J* = 14.0 Hz, 1H, $H_{C6'}$), 4.34 – 4.26 (m, 3H, $H_{C2/D2/D6'}$), 3.42 – 3.32 (m, 1H, $H_{D5}$), 3.21 (t, *J* = 13.0 Hz, 1H, $H_{C5}$), 2.97 (t, *J* = 13.0 Hz, 1H, $H_{D5'}$), 2.86 (t, *J* = 13.0 Hz, 1H, $H_{C5'}$), 2.80 (s, 3H, $H_{D8}$), 2.61 (s, 4H, $H_{C8/D3}$), 2.48 – 2.42 (m, 1H, $H_{C3}$), 2.05 – 2.00 (m, 1H, $H_{D4}$), 1.93 – 1.76 (m, 4H, $H_{C3'/D3'/C4}$), 1.76 – 1.65 (m, 1H, $H_{D4'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ (both isomers, C and D): 177.9 (D1), 177.9 (C1), 167.2 (C8), 166.7 (D8), 65.9 (C6), 61.2 (D2), 60.1 (D6), 59.5 (C2), 58.3 (D5), 57.1 (C5), 40.3 (D8), 39.0 (C8), 29.4 (D3), 28.7 (C3), 23.4 (C4), 22.4 (D4).

**(*S*)-1-Formyl-3-methyl-1,3-diazepane-4-carboxylic acid 26:**



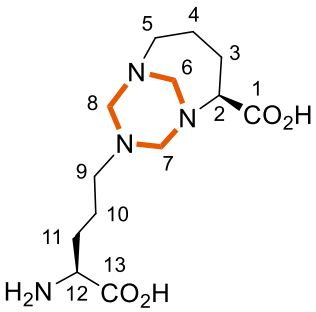

The named compound was observed when ornithine (HCl salt, 43.4 mM) was reacted with HCHO (10 equivalents) and allowed to react for 2 hours, in presence of base (NaOD, 2 equivalents). $^1$H NMR (600 MHz, $D_2O$) δ (two isomers, A and B): 8.21 (s, 1H, $H_{A8}$), 8.14 (s, 1H, $H_{B8}$), 4.96 (d, *J* = 14.0 Hz, 1H, $H_{A6}$), 4.86 – 4.81 (m, 1H, $H_{B6}$), 4.70 (s, 1H, $H_{B6'}$), 4.54 (d, *J* = 14.0 Hz, 1H, $H_{A6'}$), 3.72 – 3.66 (m, 1H, $H_{B2}$), 3.63 – 3.59 (m, 1H, $H_{A2}$), 3.56 – 3.48 (m, 2H, $H_{A5/B5}$), 3.42 – 3.32 (m, 1H, $H_{A5'}$), 3.04 – 3.00 (m, 1H, $H_{B5'}$), 2.80 (s, 3H, $H_{B7}$), 2.64 (s, 3H, $H_{A7}$), 2.13 – 2.04 (m, 2H, $H_{A3}$ and $H_{B3}$), 1.93 – 1.76 (m, 6H, $H_{A3}$, $H_{B3}$, $H_{A4}$ and $H_{B4}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 175.5 (A1), 173.9 (B1), 167.6 (B8), 166.8 (A8), 70.5 (B2), 69.1 (A2), 66.1 (A6), 60.8 (B6), 48.6 (B5), 44.1 (A5), 41.3 (B7), 38.9 (A7), 25.9 (2, A5 and B5), 23.3 (2, A6 and B6).

**(S)-1,3-Diazepane-4-carboxylic acid 27:**



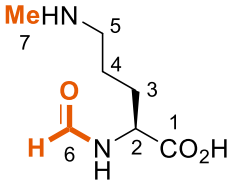

The named compound was observed when ornithine (HCl salt, 4.34 mM) was reacted with HCHO (1 equivalent), and allowed to react for 10 minutes, in presence of base (NaOD, 2 equivalents). $^1$H NMR (700 MHz, $D_2O$) δ: 3.71 (d, *J* = 13.5 Hz, 1H, $H_6$), 3.40 (d, *J* = 13.5 Hz, 1H, $H_{6'}$), 3.22 (dd, *J* = 9.0, 4.0 Hz, 1H, $H_2$), 2.70 – 2.59 (m, 2H, $H_5$), 1.66 – 1.51 (m, 2H, $H_3$), 1.50 – 1.40 (m, 2H, $H_4$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 182.2, 61.2, 60.0, 46.0, 32.5, 27.9.

**(*1R,2S,6R*)-8-((*S*)-4-Amino-4-carboxybutyl)-1,6,8-triazabicyclo[4.3.1]decane-2-carboxylic acid 28:**



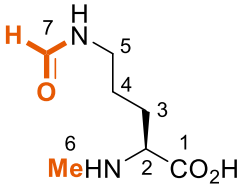
The named compound was observed when ornithine (HCl salt, 43.4 mM) was reacted with HCHO (1 equivalent), and allowed to react for 20 minutes, in presence of base (NaOD, 2 equivalents). $^1$H NMR (700 MHz, $D_2O$) δ: 3.85 (d, *J* = 14.5, 1H, $H_6$), 3.72 (d, *J* = 11.5, 1H, $H_8$), 3.64 (d, *J* = 14.5, 1H, $H_{6'}$) 3.54 (d, *J* = 11.5 Hz, 1H, $H_7$), 3.29 (dd, *J* = 12.0, 4.0 Hz, 1H, $H_2$), 3.19 (d, *J* = 11.5 Hz, 1H, $H_{8'}$), 3.10 (m, *J* = 11.5 Hz, 2H, $H_{7'}$ and $H_{12}$), 2.91 – 2.85 (m, 1H, $H_5$) 2.65 – 2.59 (m, 1H, $H_{5'}$), 2.20 – 2.12 (m, 2H, $H_3$), 1.99 – 195 (m, 1H, $H_9$), 1.91 – 1.86 (m, 1H, $H_{9'}$), 1.69 – 1.62 (m, 4H, $H_3$ and $H_{11}$), 1.55 – 1.46 (m, 2H, $H_4$ and/or $H_{10}$), 1.35 – 1.28 (m, 2H, $H_4$ and/or $H_{10}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 183.6, 183.5, 76.1, 73.8, 69.1, 63.5, 55.7, 55.0, 51.8, 32.8, 31.7, 28.5, 21.6.

**(*S*)-2-Formamido-5-(methylamino)pentanoic acid 29:**



The named compound was observed when ornithine (HCl salt, 43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 2 days, in presence of base (NaOD, 2 equivalents). $^1$H NMR (700 MHz, $D_2O$) δ: 8.04 (s, 1H, $H_6$), 4.32 (dd, *J* = 8.0, 5.0, 1H, $H_2$), 3.53 – 3.49 (m, 1H, $H_5$) 3.01 – 2.96 (m, 1H, $H_{5'}$), 2.62 (s, 3H, $H_7$), 1.89 – 1.84 (m, 1H, $H_3$), 1.75 – 1.68 (m, 1H, $H_{3'}$), 1.70 – 1.63 (m, 2H, $H_4$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 176.2, 163.9, 51.90, 48.4, 32.6, 28.1, 21.8.

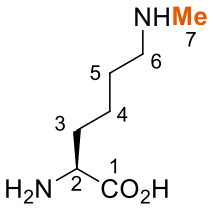
**(S)-5-Formamido-2-(methylamino)pentanoic acid 30:**



The named compound was observed when ornithine (HCl salt, 43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 2 days, in presence of base (NaOD, 2 equivalents). $^1$H NMR (700 MHz, $D_2O$) δ: 7.96 (s, 1H, $H_7$), 3.56 – 3.53 (m, 1H, $H_2$), 3.20 – 3.17 (m, 2H, $H_5$), 2.62 (s, 3H, $H_6$), 1.85 – 1.75 (m, 2H, $H_3$), 1.57 – 1.44 (m, 2H, $H_4$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 173.2, 164.3, 63.0, 37.3, 31.6, 29.1, 24.0.

### 7.12.3.13 Lysine product

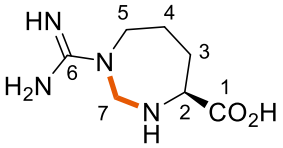
**$N^{\varepsilon}$-Methyl-*L*-lysine 31:**




The named compound was observed when lysine (HCl salt, 43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 48 hours, at pD 9. $^{1}$H NMR (700 MHz, $D_2O$) δ: 3.61 (dd, $J$ = 8.0, 3.0 Hz, 1H, $H_2$), 2.92 (t, $J$ = 10.0 Hz, 2H, $H_6$), 2.56 (s, 3H, $H_7$), 1.80 – 1.71 (m, 2H, $H_3$), 1.62 – 1.55 (m, 2H, $H_5$), 1.40 – 1.24 (m, 2H, $H_4$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 174.5, 54.4, 48.5, 32.6, 29.8, 26.3, 21.3. These assignments are in agreement with previously reported data.[20]

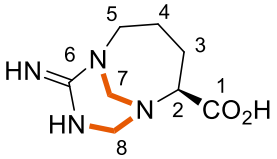
### 7.12.3.14 Arginine products

**(S)-1-Carbamimidoyl-1,3-diazepane-4-carboxylic acid 32:**




The named compound was observed when arginine (43.4 mM) was reacted with HCHO (3 equivalents), and allowed to react for 4 hours. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.35 (d, $J$ = 14.5 Hz, 1H, $H_7$) 4.11 (d, $J$ = 14.5 Hz, 1H, $H_{7'}$), 3.35 – 3.28 (m, 2H, $H_5$), 3.20 – 3.14 (m, 2H, $H_2$), 1.98 – 1.91 (m, 1H, $H_3$), 1.82 – 1.77 (m, 1H, $H_4$), 1.70 – 1.63 (m, 1H, $H_{4'}$), 1.62 – 1.56 (m, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 180.7, 156.1, 61.6, 59.5, 46.4, 31.6, 25.0.

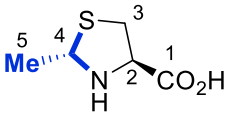
**(*2S*)-7-Imino-1,6,8-triazabicyclo[4.3.1]decane-2-carboxylic acid 33:**




The named compound was observed when arginine (4.34 mM) was reacted with HCHO (10 equivalents), and allowed to react for 24 hours. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.37 (d, $J$ = 12.0 Hz, 1H, $H_7$), 4.28 (d, $J$ = 14.0 Hz, 1H, $H_8$), 4.19 (d, $J$ = 14.0 Hz, 1H, $H_{8'}$), 4.02 (d, $J$ = 12.0, H z, 1H, $H_{7'}$), 3.65 (dt, $J$ = 15.0, 4.5 Hz, 1H, $H_5$), 3.08 – 3.03 (m, 1H, $H_2$), 3.01 – 2.96 (m, 1H, $H_{5'}$), 1.95 – 1.87 (m, 2H, $H_3$ and $H_4$), 1.83 – 1.78 (m, 2H, $H_3$ and $H_4$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 181.1, 157.0, 67.9, 64.8, 63.9, 50.0, 28.8, 28.2. ESI-MS calculated for $C_8H_{15}O_2N_4$ $[M+1]^+$ 199.11895, found 199.11892.

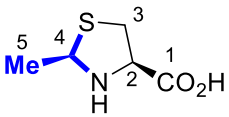
### 7.12.3.15 Cysteine and acetaldehyde products

**(2*S*,4*R*)-2-Methylthiazolidine-4-carboxylic acid 35a:**




The named compound was observed when cysteine (43.4 mM) was reacted with acetaldehyde (10 equivalents), and allowed to react for 1 hours, in presence of and base (NaOD, 1 e quivalent). $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.82 (q, *J* = 6.5 Hz, 1H, $H_4$), 4.37 (dd, *J* = 7.5, 6.5 Hz, 1H, $H_2$), 3.40 (dd, *J* = 12.0, 7.5 Hz, 1H, $H_3$), 3.19 (dd, *J* = 12.0, 7.5 Hz, 1H, $H_{3'}$), 1.53 (d, *J* = 6.5 Hz, 3H, $H_5$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 171.8, 63.5, 60.5, 32.6, 18.3.

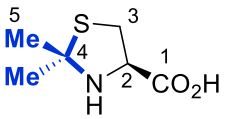
**(2*R*,4*R*)-2-Methylthiazolidine-4-carboxylic acid 35b:**




The named compound was observed when cysteine (43.4 mM) was reacted with acetaldehyde (10 equivalents) and base (NaOD, 1 equivalent) after 1 hour. $^{1}$H NMR (700 MHz, $D_2O$) δ: 4.74 (q, *J* = 6.5 Hz, 1H, $H_4$), 4.29 (dd, *J* = 7.5, 6.5 Hz, 1H, $H_2$), 3.34 (dd, *J* = 12.0, 7.5 Hz, 1H, $H_3$), 3.24 (dd, *J* = 12.0, 7.5 Hz, 1H, $H_{3'}$), 1.55 (d, *J* = 6.5 Hz, 3H, $H_5$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 171.8, 64.3, 61.0, 32.8, 17.3.

### 7.12.3.16 Cysteine and acetone product

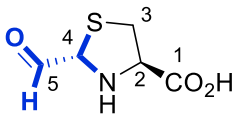
**(*R*)-2,2-dimethylthiazolidine-4-carboxylic acid 36:**




The named compound was observed when cysteine (43.4 mM) was reacted with acetone (10 equivalents), and allowed to react for 3 hours, in presence of base (NaOD, 1 equivalent). $^{1}$H NMR (600 MHz, $D_2O$) δ: 4.52 (dd, *J* = 8.0, 4.0 Hz, 1H, $H_2$), 3.57 (dd, *J* = 12.5, 8.0 Hz, 1H, $H_3$), 3.41 (dd, *J* = 12.5, 8.0 Hz, 1H, $H_{3'}$), 1.77 (s, 3H, $H_5$), 1.75 (s, 3H, $H_{5'}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 172.3, 73.2, 64.3, 32.9, 27.5, 26.9.

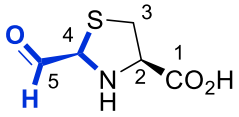
### 7.12.3.17 Cysteine and glyoxal products

**(2*S*,4*R*)-2-(2-Oxoacetyl)thiazolidine-4-carboxylic acid 37a:**



The named compound was observed when cysteine (43.4 mM) was reacted with glyoxal (10 equivalents), and allowed to react for 1 hour. $^{1}$H NMR (700 MHz, $D_2O$) δ: 5.15 (d, *J* = 5.0 Hz, 1H, $H_5$), 4.79 – 4.76 (m, 1H, $H_4$) 4.46 (dd, *J* = 7.0, 5.0 Hz, 1H, $H_2$), 3.30 (dd, *J* = 11.5, 7.0 Hz, 1H, $H_3$), 3.21 (dd, J = 11.5, 7.0 Hz, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 171.3, 88.6, 66.9, 64.6, 32.2.

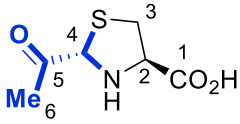
**(2*R*,4*R*)-2-(2-Oxoacetyl)thiazolidine-4-carboxylic acid 37b:**



The named compound was observed when cysteine (43.4 mM) was reacted with HCHO (10 equivalents), and allowed to react for 1 hour. $^{1}$H NMR (700 MHz, $D_2O$) δ: 5.19 (d, *J* = 6.0 Hz, 1H, $H_5$), 4.65 – 4.63 (m, 1H, $H_4$), 4.39 (dd, *J* = 7.0, 5.0 Hz, 1H, $H_2$), 3.33 (d, *J* = 7.0 Hz, 1H, $H_3$), 3.25 – 3.23 (m, 1H, $H_{3'}$). $^{13}$C NMR (176 MHz, $D_2O$) δ: 171.1, 88.3, 67.6, 64.9, 32.3.

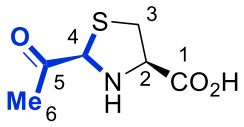
### 7.12.3.18 Cysteine and methylglyoxal products

**(2*S*,4*R*)-2-Acetylthiazolidine-4-carboxylic acid 38a:**



The named compound was observed when cysteine (43.4 mM) was reacted with methylglyoxal (10 equivalents), and allowed to react for 1 hour. $^{1}$H NMR (600 MHz, $D_2O$) δ: 5.65 (s, 1H, $H_4$), 4.45 (dd, *J* = 7.5, 3.5 Hz, 1H, $H_2$), 3.44 – 3.41 (m, 1H, $H_3$), 2.99 (dd, *J* = 12.5, 10.0 Hz, 1H, $H_{3'}$), 2.34 (s, 3H, $H_6$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 200.4, 170.8, 67.8, 66.8, 32.9, 25.1.

**(2*R*,4*R*)-2-Acetylthiazolidine-4-carboxylic acid 38b:**




The named compound was observed when cysteine (43.4 mM) was reacted with methylglyoxal (10 equivalents), and allowed to react for 1 hour. $^{1}$H NMR (600 MHz, $D_2O$) δ: 5.72 (s, 1H, $H_4$), 4.55 (dd, $J$ = 7.5, 3.5 Hz, 1H, $H_2$), 3.40 – 3.37 (m, 1H, $H_3$), 3.18 (dd, $J$ = 12.5, 7.0 Hz, 1H, $H_{3'}$), 2.31 (s, 3H, $H_{11}$). $^{13}$C NMR (151 MHz, $D_2O$) δ: 200.7, 171.6, 67.7, 64.8, 32.9, 24.9.

## 7.13 References


1. Szarvas, T. *et al.* Determination of Endogenous Formaldehyde Level in Human Blood and Urine by Dimedone-$^{14}$C Radiometric method. *J. Radioanal. Nucl. Chem.* **106**, 357–367 (1986).

2. Pontel, L. B. *et al.* Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Mol. Cell* **60**, 177–188 (2015).

3. *Brown, J. H. The Formal titration of Bacteriological Media. *J. Bacteriol.* **8**, 245–267 (1922).

4. Wadsworth, A. & Pangborn, M. C. Formaldehyde Reaction with Amino Acids. *J. Biol. Chem.* **116**, 423–436 (1936).

5. Smith, A. K., Handler, P. & Mrgudich, J. N. The Reaction of Formaldehyde with Amino Acids: X-ray Diffraction Patterns. *J. Phys. Chem.* **44**, 874–880 (1940).

6. Middlebrook, W. R. & Phillips, H. The Action of Formaldehyde on the Cystine Disulphide Linkages in Wool. *Biochem. J.* **36**, 294–302 (1942).

7. French, D. & Edsall, J. T. *Advances in Protein Chemistry: The Reactions of Formaldehyde with Amino Acids and Proteins.* (Academic Press Inc., 1945).

8. Fraenkel-Conrat, H. & Olcott, H. S. Reaction of Formaldehyde with Proteins; II. Participation of the Guanidyl Groups and Evidence of Crosslinking. *J. Am. Chem. Soc.* **68**, 34–37 (1946).

9. Fraenkel-Conrat, H. & Olcott, H. S. The Reaction of Formaldehyde with Proteins; V. Cross-linking between Amino and Primary Amide or Guanidyl Groups. *J. Am. Chem. Soc.* **70**, 2673–2684 (1948).

10. Middlebrook, W. R. The irreversible combination of formaldehyde with proteins. *Biochem. J.* **44**, 17–23 (1949).

11. Guthe, K. F. & Arbor, A. The Formaldehyde-Hemoglobin Reaction. *J. Biol. Chem.* **234**, 3169–3174 (1969).

12. Myers, J. S. & Hardman, J. K. Formaldehyde-induced in the a Subunit Escherichia coli Tryptophan Synthetase ". *J. Biol. Chem.* **246**, 3863–3869 (1971).

13. Takahashi, K. The Reactions of Phenylglyoxal and related reagents with amino acids. *J. Biochem.* **81**, 395–402 (1977).

14. Kitamoto, Y. & Maeda, H. Reevaluation of the reaction of formaldehyde at low concentration with amino acids. *J. Biochem.* **87**, 1519–1530 (1980).

15. Tome, D., Naulet, N. & Kozlowski, A. $^{13}$C-n.m.r. characterization of formaldehyde bonds in model mixtures and proteins containing lysine. *Int. J. Pept. Protein Res.* **25**, 258–266 (1985).

16. Joshi, N. S., Whitaker, L. R. & Francis, M. B. A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J. Am. Chem. Soc.* **126**, 15942–15943 (2004).

17. Clarke, H. T., Gillespie, B. & Weisshaus, S. Z. The Action of Formaldehyde on Amines and Amino Acids. *J. Am. Chem. Soc.* **55**, 4571–4587 (1933).

18. McGhee, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. 4. Mechanism of the Initial Reaction of Formaldehyde with DNA. *Biochemistry* **16**, 3276–3293 (1977).

19. Lu, K. *et al.* Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. *J. Am. Chem. Soc.* **132**, 3388–3399 (2010).

20. Trézl, L. *et al.* Antagonistic reactions of arginine and lysine against formaldehyde and their relation to cell proliferation, apoptosis, folate cycle and photosynthesis. *Mol. Cell. Biochem.* **244**, 167–176 (2003).

21. Metz, B. *et al.* Identification of Formaldehyde-induced Modifications in Proteins. *J. Biol. Chem.* **279**, 6235–6243 (2004).

22. Shishodia, S. *et al.* NMR analyses on N-hydroxymethylated nucleobases-implications for formaldehyde toxicity and nucleic acid demethylases. *Org. Biomol. Chem.* **16**, 4021–4032 (2018).

23. Ott, M., Fischer, H. H., Maiwald, M., Albert, K. & Hasse, H. Kinetics of oligomerization reactions in formaldehyde solutions: NMR experiments up to 373 K and thermodynamically consistent model. *Chem. Eng. Process. Process Intensif.* **44**, 653–660 (2005).

24. Hahnenstein, I., Albert, M., Hasse, H., Kreiter, C. G. & Maurer, G. NMR Spectroscopic and Densimetric Study of Reaction Kinetics of Formaldehyde Polymer Formation in Water, Deuterium Oxide, and Methanol. *Ind. Eng. Chem. Res.* **34**, 440–450 (1995).

25. Schmitz, N., Burger, J. & Hasse, H. Reaction Kinetics of the Formation of Poly(oxymethylene) Dimethyl Ethers from Formaldehyde and Methanol in Aqueous Solutions. *Ind. Eng. Chem. Res.* **54**, 12553–12560 (2015).

26. Rivlin, M., Eliav, U. & Navon, G. NMR Studies of the Equilibria and Reaction Rates in Aqueous Solutions of Formaldehyde. *J. Phys. Chem. B* **119**, 4479–4487 (2015).

27. Cannizzaro, S. Ueber den der Benzoesaure entsprechenden Alkohol. *Liebigs Ann. der Chemie und Pharm.* **88**, 129–130 (1853).

28. Loew, O. Ueber die Condensation des Formaldehyds unter verschiedenen Bedingungen. *Berichte der Dtsch. Chem. Gesellschaft* **21**, 270–275 (1888).

29. Swain, C. G., Powell, A. L., Sheppard, W. A. & Morgan, C. R. Mechanism of the Cannizzaro Reaction. 3576–3583 (1979).

30. Hopkinson, R. J. *et al.* Studies on the Glutathione-Dependent Formaldehyde-Activating Enzyme from Paracoccus denitrificans. *PLoS One* **10**, e0145085 (2015).

31. Kallen, R. G. The mechanism of reactions involving Schiff base intermediates. Thiazolidine formation from L-cysteine and formaldehyde. *J. Am. Chem. Soc.* **93**, 6236–6248 (1971).

32. Pellegrini, N. *et al.* New Dipeptides Containing Thiazolidine-4-carboxylic Acid Derivatives: Synthesis and Characterization Using NMR Techniques and X-Ray Data. *Chem. Pharm. Bull.* **47**, 950–955 (1999).

33. Li, Z., Gentry, Z., Murphy, B. & Vannieuwenhze, M. S. Scalable Synthesis of Orthogonally Protected β-Methyllanthionines by Indium(III)-Mediated Ring Opening of Aziridines. *Org. Lett.* **21**, 2200–2203 (2019).

34. Barve, A. *et al.* Differences in heterocycle basicity distinguish homocysteine from cysteine using aldehyde-bearing fluorophores. *Chem. Commun.* **50**, 8219–8222 (2014).

35. Lewis, J. N., Inloes, R. L. & Hes, J. Synthetic Sulfur-Containing Amino Acids. Inhibition of Transport Systems in S37 Ascites Tumor Cells. *J. Med. Chem.* **21**, 1070–1073 (1978).

36. Smalley, P. S. & Bell, R. A. Amino-acid zwitterion equilibria: vibrational and nuclear magnetic resonance studies of methyl-substituted thiazolidine-4-carboxylic acids. *Can. J. Chem.* **64**, 1215–1219 (1986).

37. Saiz, C., Escalante, M. & Furlan, R. L. E. Dithioacetal exchange : a new reversible reaction for Dynamic Combinatorial Chemistry. **4**, 2–5 (2009).

38. Canle, M., Lawley, A., McManus, E. C. & More O'Ferrall, R. A. Rate and equilibrium constants for oxazolidine and thiazolidine ring-opening reactions. *Pure Appl. Chem.* **68**, 813–818 (1996).

39. ACD/I-Lab, 5.0.0.184, Advanced Chemistry Development, Inc., Toronto, On, Canada, www.acdlabs.com, . No Title. (2019).

40. Amino acids and acidity values, http://academics.keene.edu/rblatchly/Chem220/hand/npaa/aawpka.htm, (accessed 1 May 2019). *Amino acids and acidity values, http://academics.keene.edu/rblatchly/Chem220/hand/npaa/aawpka.htm, (accessed 1 May 2019).*

41. Beesley, R. M., Ingold, C. K. & Thorpe, J. F. The Formation and Stability of spiro-Compounds. Part I. spiro-Compounds from cyclo-Hexane. *J. Chem. Soc. Trans.* **107**, 1080–1106 (1915).

42. Mason, R. P., Sanders, J. K. M. & Gidley, M. J. NMR Visualization of Free Asparagine in Potato Tissue Using Adduct Formation with [$^{13}$C]Formaldehyde. *Phytochemistry* **25**, 1567–1571 (1986).

43. Stöckigt, J., Antonchick, A. P., Wu, F. & Waldmann, H. The Pictet – Spengler Reaction in Nature and in Organic Chemistry. *Angew. Chemie - Int. Ed.* **50**, 8538–8564 (2011).

44. Remelli, M., Pulidori, F., Guerrini, R. & Bertolasi, V. Synthesis of spinacine and spinacine derivatives: crystal and molecular structures of Nπ-hydroxymethyl spinacine and Nα-methyl spinaceamine. *J. Chem. Crystallogr.* **27**, 507–513 (1991).

45. Dale, J. & Sigvartsen, T. Condensation of Alkanediamines with formaldehyde; intra molecular disproportionation of N-hydroxymethyl groups into N-methyl and N-formal groups. *Acta Chem. Scand.* **45**, 1064–1070 (1991).

46. Markolovic, S., Wilkins, S. E. & Schofield, C. J. Protein Hydroxylation Catalyzed by 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20712–20722 (2015).

47. Covington, A. K., Robinson, R. A., Bates, R. G. & Bureau, N. Use of the Glass Electrode in Deuterium Oxide and the Relation between the Standardized pD ( paD ) Scale and the Operational pH in Heavy Water. *Anal. Chem.* **40**, 700–706 (1968).

48. Endogenous formaldehyde turnover in humans compared with exogenous contribution from food sources. *EFSA J.* **12**, 3550 (2014).

49. Chapter 5.8 Formaldehyde. in *Air Quality Guidelines - Second Edition* vol. 3 1–25 (2001).

50. Salthammer, T. Formaldehyde sources , formaldehyde concentrations and air exchange rates in European housings. *Build. Environ.* **150**, 219–232 (2019).

51. Luo, W., Li, H., Zhang, Y. & Ang, C. Y. Determination of formaldehyde in blood plasma by high-performance liquid chromatography with fluorescence detection. *J. Chromatogr. B. Biomed. Sci. Appl.* **753**, 253–257 (2001).

52. Vorländer, D. & Kalkow, F. Die Formaldehydverbindung des Hydroresorcins. *Justus Liebigs Ann. Chem.* **307**, 356–374 (1899).

53. Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. W. & Schofield, C. J. Monitoring the activity of 2-oxoglutarate dependent histone demethylases by NMR spectroscopy: Direct observation of formaldehyde. *ChemBioChem* **11**, 506–510 (2010).

54. Harms, N., Ras, J., Reijnders, W. N. M., Van Spanning, R. J. M. & Stouthamer, A. H. S-formylglutathione hydrolase of Paracoccus denitrificans is homologous to human esterase D: A universal pathway for formaldehyde detoxification? *J. Bacteriol.* **178**, 6296–6299 (1996).

55. Hopkinson, R. J., Barlow, P. S., Schofield, C. J. & Claridge, T. D. W. Studies on the reaction of glutathione and formaldehyde using NMR. *Org. Biomol. Chem.* **8**, 4915–4920 (2010).

56. Bateman, R. & Shokat, K. M. Glutathione traps formaldehyde by formation of a bicyclo [4,4,1] undecane adduct. *Org. Biomol. Chem.* **5**, 3363–3367 (2007).

57. Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. Formaldehyde Crosslinking : A Tool for the Study of Chromatin Complexes. *J. Biol. Chem.* **290**, 26404–26411 (2015).

58. Sutherland, B. W., Toews, J. & Kast, J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein – protein interactions. *J. Mass Spectrom.* **43**, 699–715 (2008).

59. Wu, C. *et al.* Sequence-Specific Capture of Protein-DNA Complexes for Mass Spectrometric Protein Identification. *PLoS One* **6**, e26217 (2011).

60. Shiraishi, H., Kataoka, M., Morita, Y. & Umemoto, J. Interactions of Hydroxyl Radicals with Tris ( Hydroxymethyl ) Aminomethane and Good ' s Buffers Containing Hydroxymethyl or Hydroxyethyl Residues Produce Formaldehyde. *Free Radic. Res. Commun.* **19**, 315–321 (1993).

61. Henry, L., Leung, I. K. H., Claridge, T. D. W. & Schofield, C. J. γ-Butyrobetaine hydroxylase catalyses a Stevens type rearrangement. *Bioorganic Med. Chem. Lett.* **22**, 4975–4978 (2012).

62. McGhee, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. I. Reaction with Exocyclic Amino Groups of DNA Bases. *Biochemistry* **14**, 1281–1296 (1975).

63. Lu, K. *et al.* Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino Acids and Deoxynucleosides and Their Oligomers. *J. Am. Chem. Soc.* **132**, 3388–3399 (2010).

64. Mannich, C. & Krösche, W. Ueber ein Kondensationsprodukt aus Formaldehyd, Ammoniak und Antipyrin. *Arch. Pharm.* **250**, 647–667 (1912).

65. Stingele, J. *et al.* Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol. Cell* **64**, 688–703 (2016).

66. Friede, H. A., Campoli-Richards, D. M. & Goa, K. L. Sultamicillin. *Drugs* **37**, 491–522 (1989).

67. Zhang, L., Steinmaus, C., Eastmond, D. A., Xin, X. K. & Smith, M. T. Formaldehyde exposure and leukemia: A new meta-analysis and potential mechanisms. *Mutat. Res. - Rev. Mutat. Res.* **681**, 150–168 (2009).

68. Voulgaridou, G. P., Anestopoulos, I., Franco, R., Panayiotidis, M. I. & Pappa, A. DNA damage induced by endogenous aldehydes: Current state of knowledge. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **711**, 13–27 (2011).

69. Macpherson, L. J. *et al.* An Ion Channel Essential for Sensing Chemical Damage. *J. Neurosci.* **27**, 11412–11415 (2007).

70. Abdulla, S. A., Zailina, H. & Abu Bakar, F. Health Risk Assessment of Adults Consuming Commercial Fish Contaminated with Formaldehyde. *Food Public Heal.* **3**, 52–58 (2013).

71. Restani, P., Campagne, P., Fiecchi, A., Resmini, P. & Galli, C. L. Identification of spinacine as a the principal reaction product of γ-casein with formaldehyde in cheese. *Food Chem. Toxicol.* **26**, 441–446 (1988).

72. Halla, N. *et al.* Cosmetics Preservation : A Review on Present Strategies. *Molecules* **23**, 1–41 (2018).

73. Kaur, G. *et al.* Biological evaluation of tubulysin A : a potential anticancer and antiangiogenic natural product. *Biochem. J.* **242**, 235–242 (2006).

74. Kamps, J. J. A. G., Hopkinson, R. J., Schofield, C. J. & Claridge, T. D. W. How formaldehyde reacts with amino acids. *Commun. Chem.* **2**, 1–14 (2019).

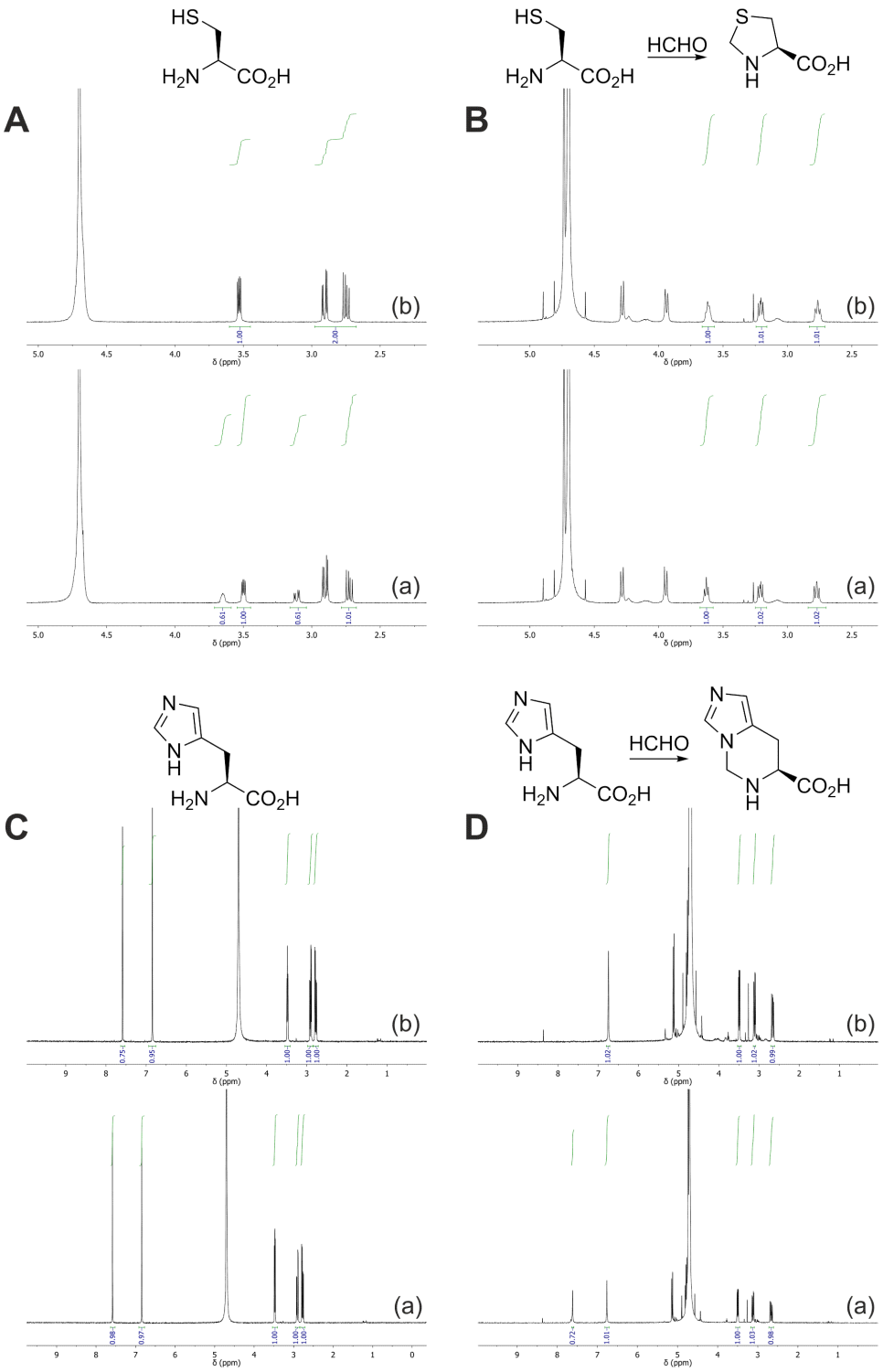
75. Jagtap, R. M., Thorat, S. H., Gonnade, R. G., Khan, A. A. & Pardeshi, S. K. X-ray crystal structures and anti-breast cancer property of 3-tert-butoxycarbonyl-2-arylthiazolidine-4-carboxylic acids. *New J. Chem.* **42**, 1078–1086 (2018).

76. Ikunaka, M., Matsumoto, J. & Nishimoto, Y. A concise synthesis of ( 2 S , 3 S ) -BocAHPBA and (R)-BocDMTA , chiral building blocks for peptide-mimetic HIV protease inhibitors. *Tetrahedron* **13**, 1201–1208 (2002).

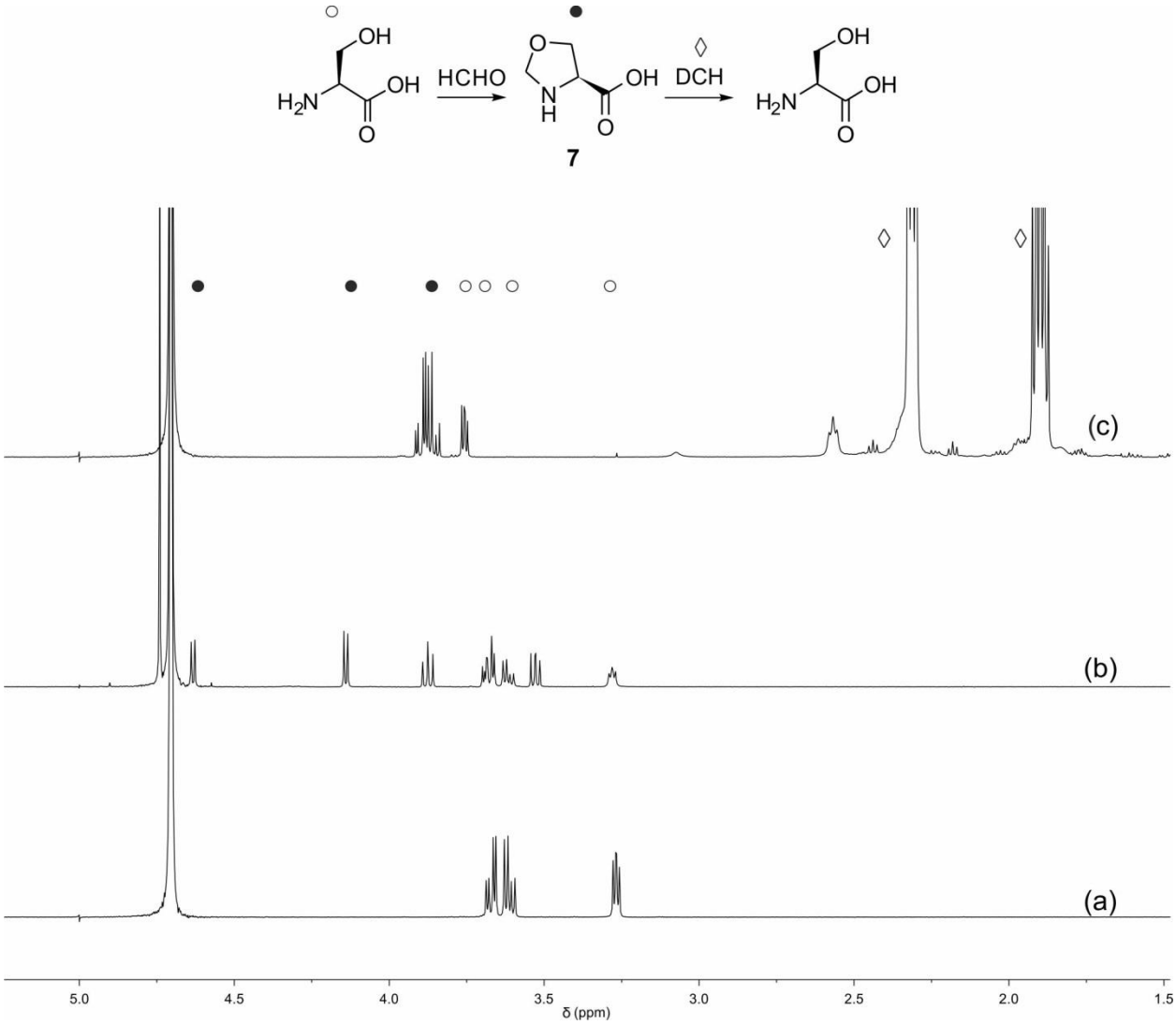
77. Vishnumaya, M. R. & Singh, V. K. Highly Efficient Small Organic Molecules for Enantioselective Direct Aldol Reaction in Organic and Aqueous Media. *J. Org. Chem.* **74**, 4289–4297 (2009).

# Appendix 1

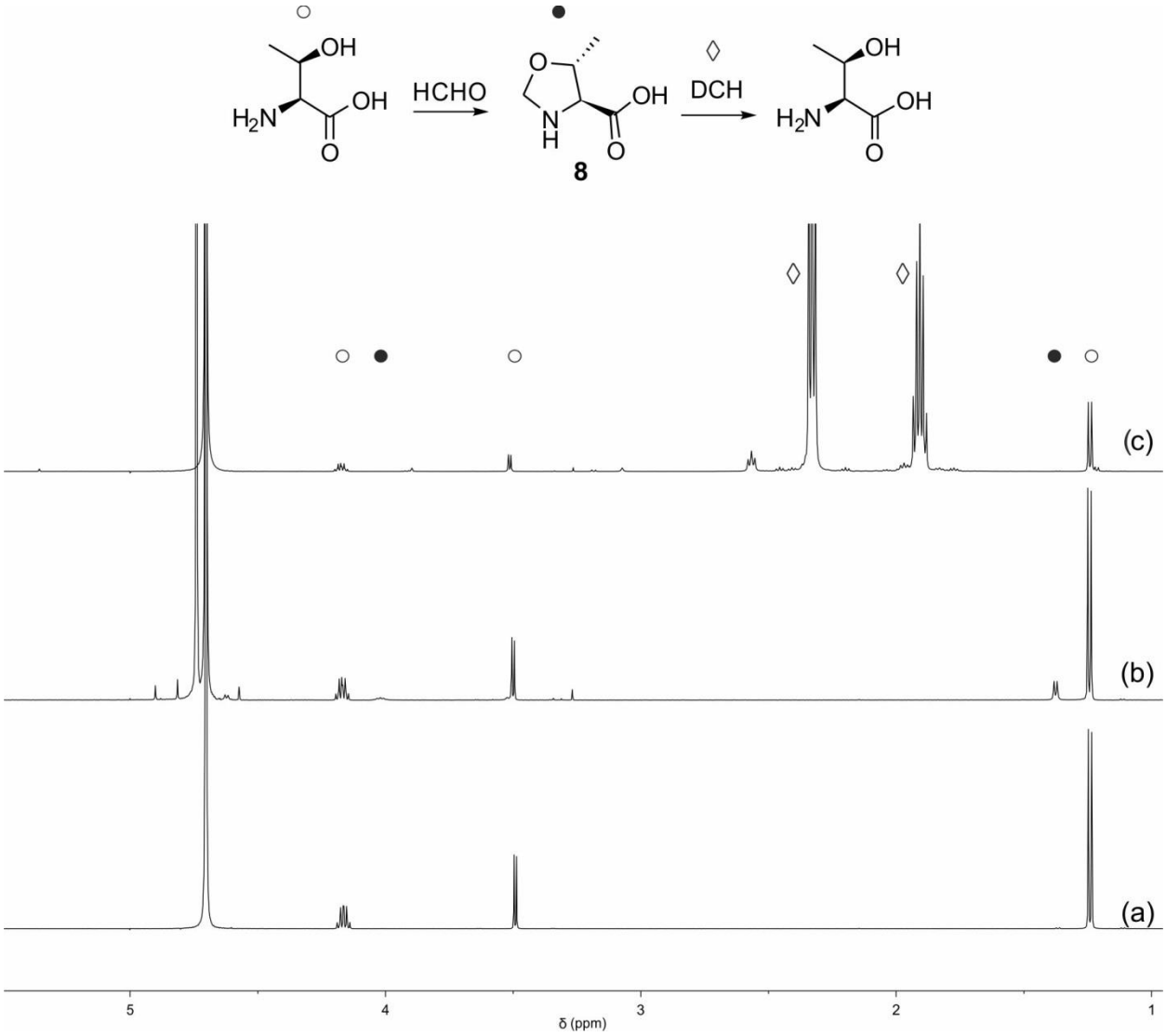
## Supplementary figures



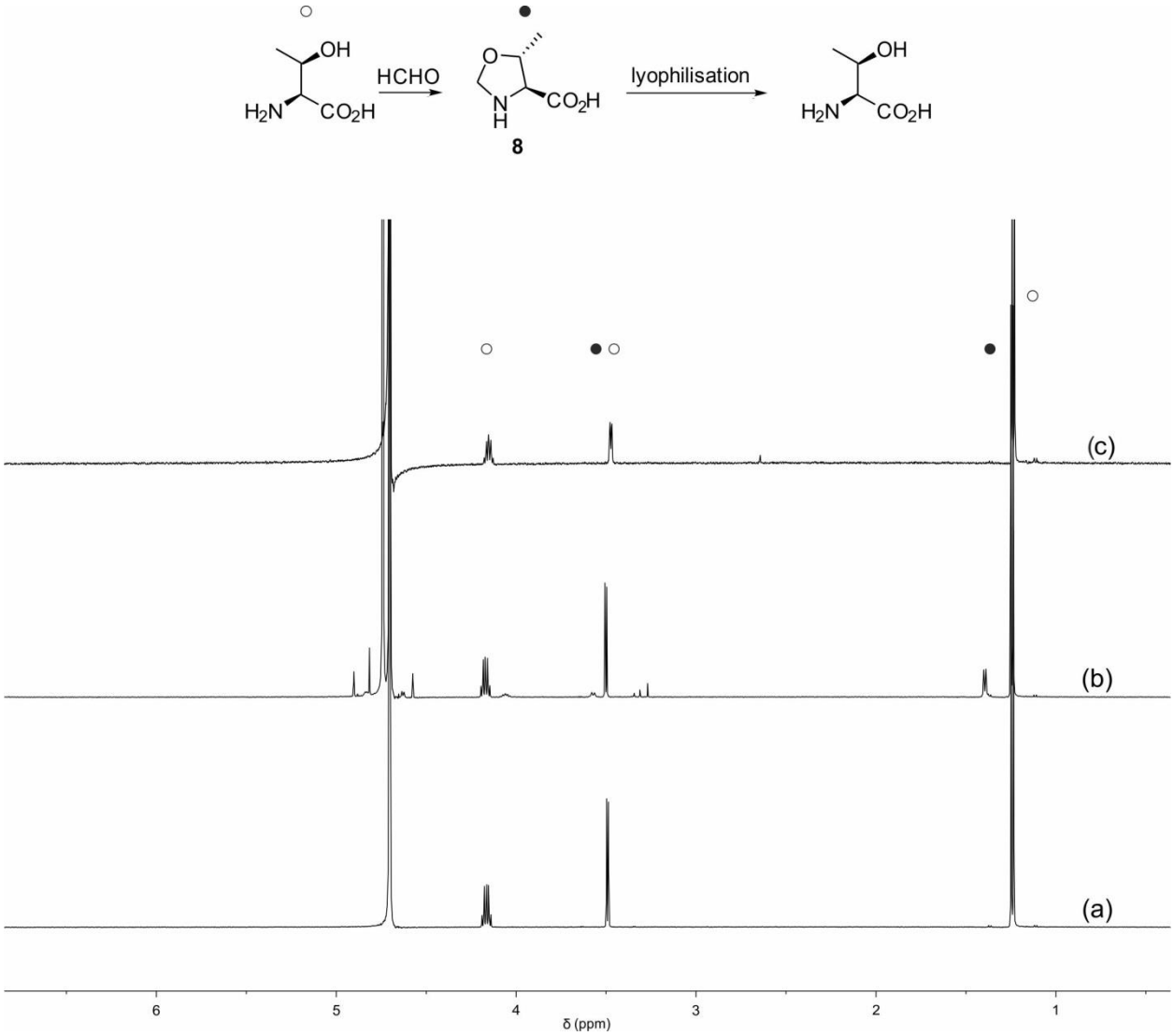

**Appendix Figure 1:** Quantitative $^{1}H$ NMR analyses measuring $^{2}H$ incorporation at C-α to investigate possible loss of chirality. Note the lack of evidence for C-α deuterium incorporation, suggesting no chirality loss during the experiment. **A**) Cysteine (5 mM) in the presence of base (NaOD, 5 mM,): (a) 15 minutes; (b) 48 hours. **B**) Cysteine (5 mM) in the presence of base (NaOD, 5 mM,) and HCHO (5 mM) over time: (a) 15 minutes; (b) 48 hours. **C**) Histidine (5 mM) in the presence of base (NaOD, 5 mM,) over time: (a) 15 minutes; (b) 48 hours. **D**) Cysteine (5 mM) in the presence of base (NaOD, 5 mM) and HCHO (5 mM): (a) 15 minutes; (b) 48 hours. Note these results show the reaction of HCHO (and other reactive carbonyl compounds) have the potential to alter thiol/disulfide equilibria.



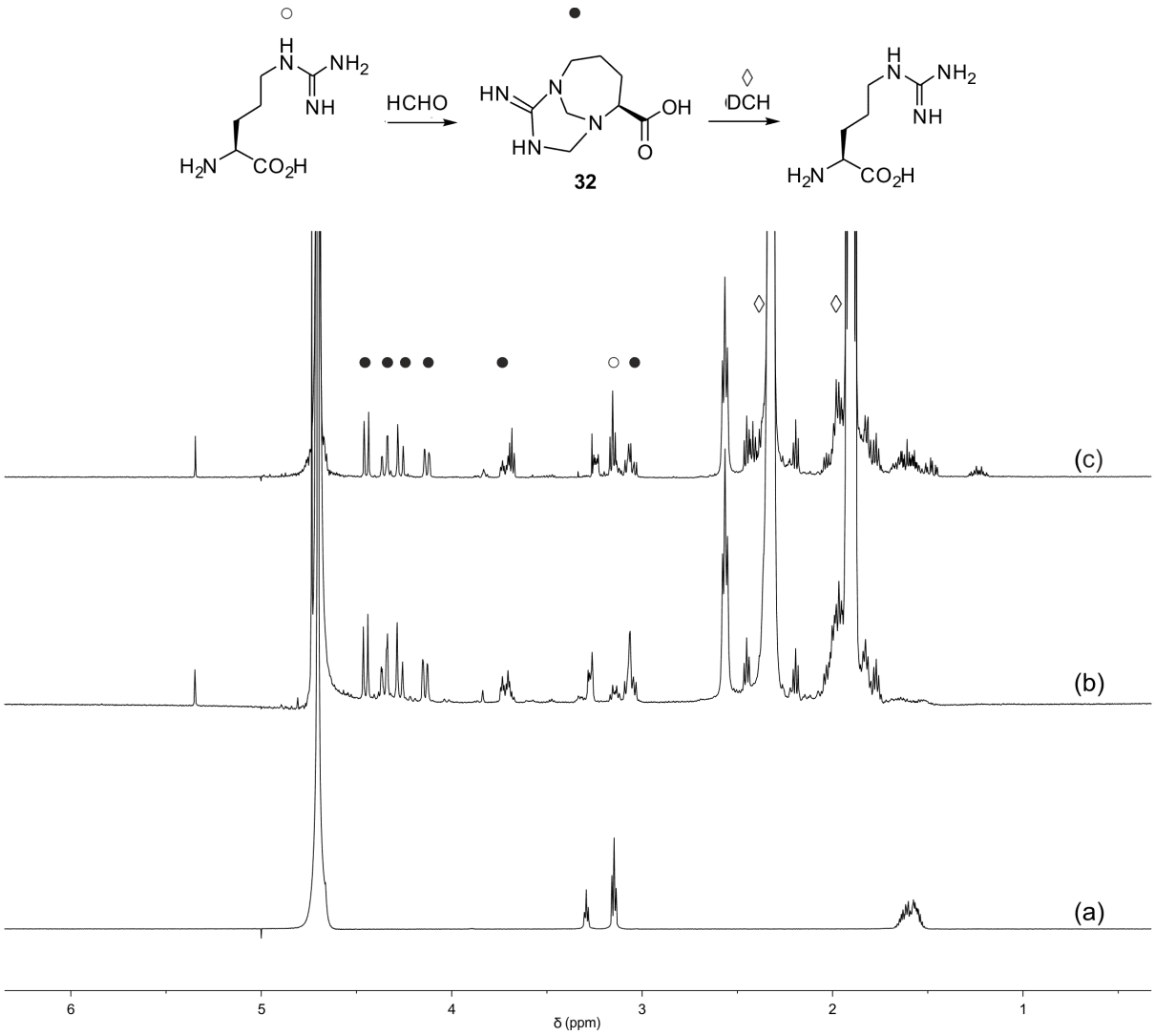

**Appendix Figure 2:** Diketone stability assay with serine derived **7**. Reversibility assay of **7** (●) product of a reaction of HCHO with serine (○), using an excess of 1,3-cyclohexadione (DCH, ◊): (a) serine (4.34 mM) and base (NaOD 1 equiv.); (b) 1 hour after the addition of HCHO; (c) 1 hour after the addition of 1,3-cyclohexadione. Note, other spectra were used in the assignment of the structures of the shown products.



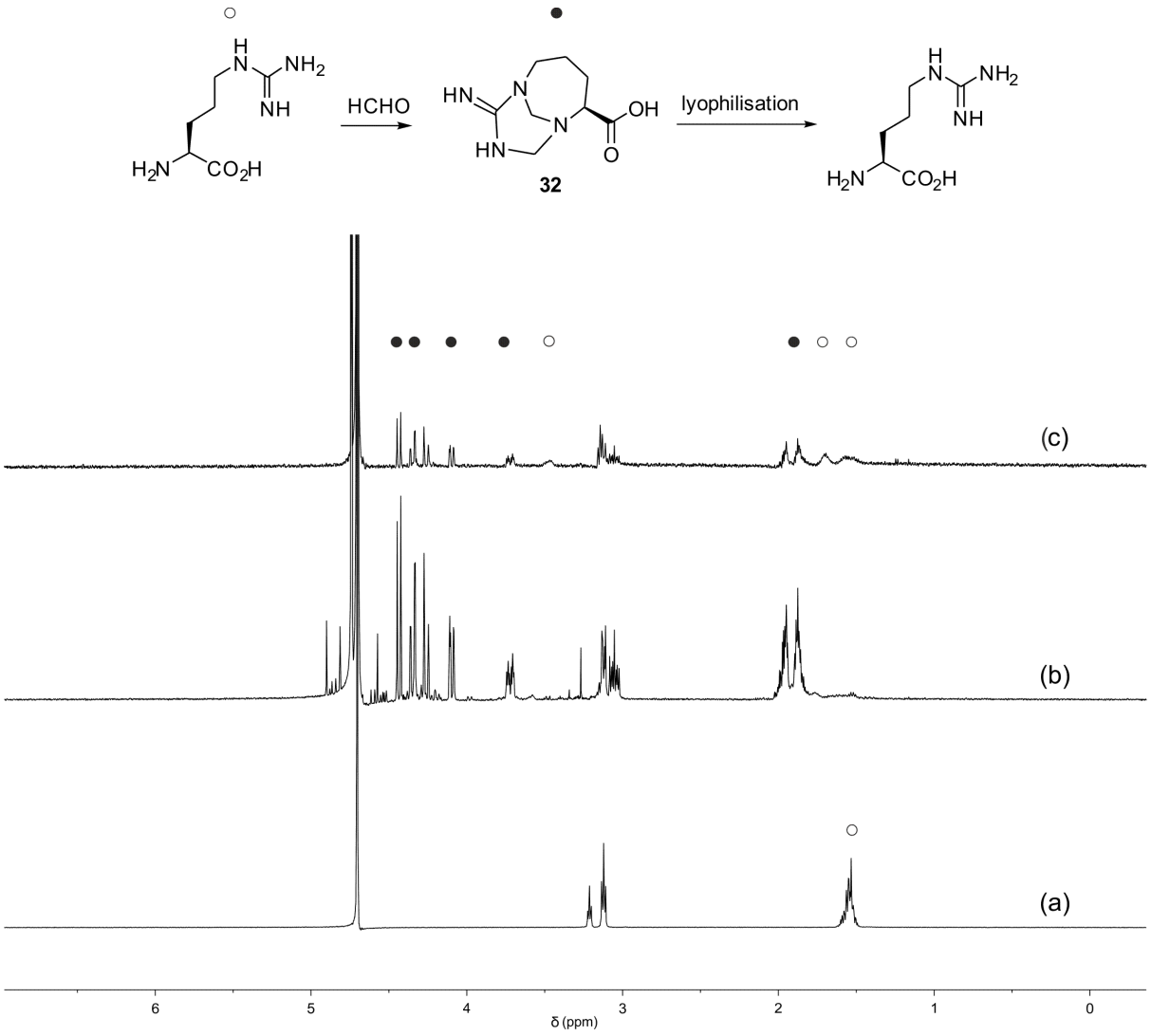

**Appendix Figure 3:** Diketone stability assay with threonine derived **8**. Reversibility assay of **8** (●) product of a reaction of HCHO with threonine (○), using an excess of 1,3-cyclohexadione (DCH, ◊): (a) threonine (4.34 mM); (b) 1 hour after the addition of HCHO (10 equiv.); (c) 4 hours after the addition of 1,3-cyclohexadione (40 equiv.). Note, other spectra were used in the assignment of the structures of the shown products.



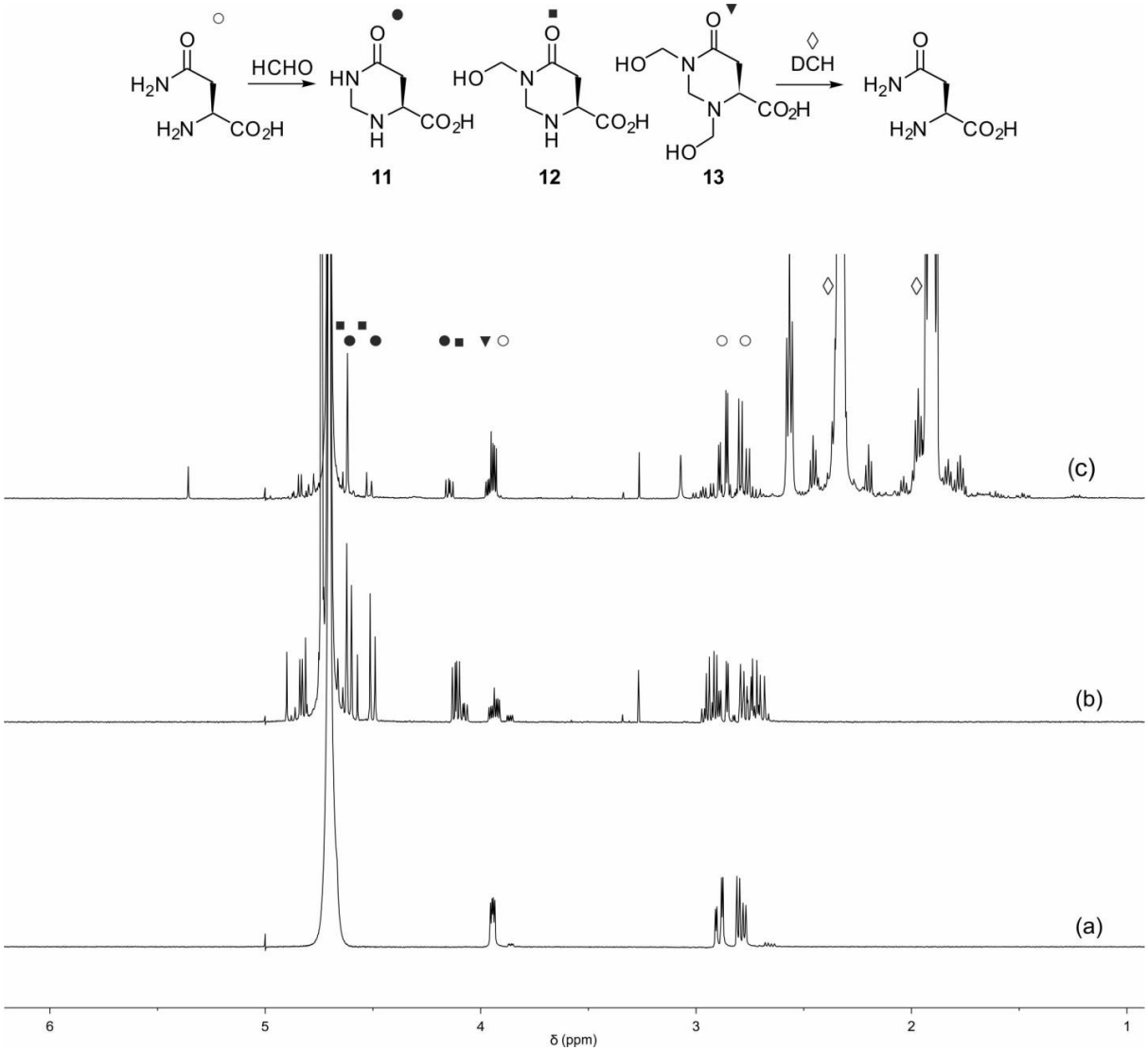

**Appendix Figure 4:** Lyophilisation stability assay with threonine derived **8**. Reversibility assay of **8** (●) product of a reaction of HCHO with threonine (○), using multiple cycles of lyophilisation: (a) threonine (4.34 mM); (b) 1 hour after the addition of HCHO (10 equiv.); (c) after 7 lyophilisation cycles. Note, other spectra were used in the assignment of the structures of the shown products.



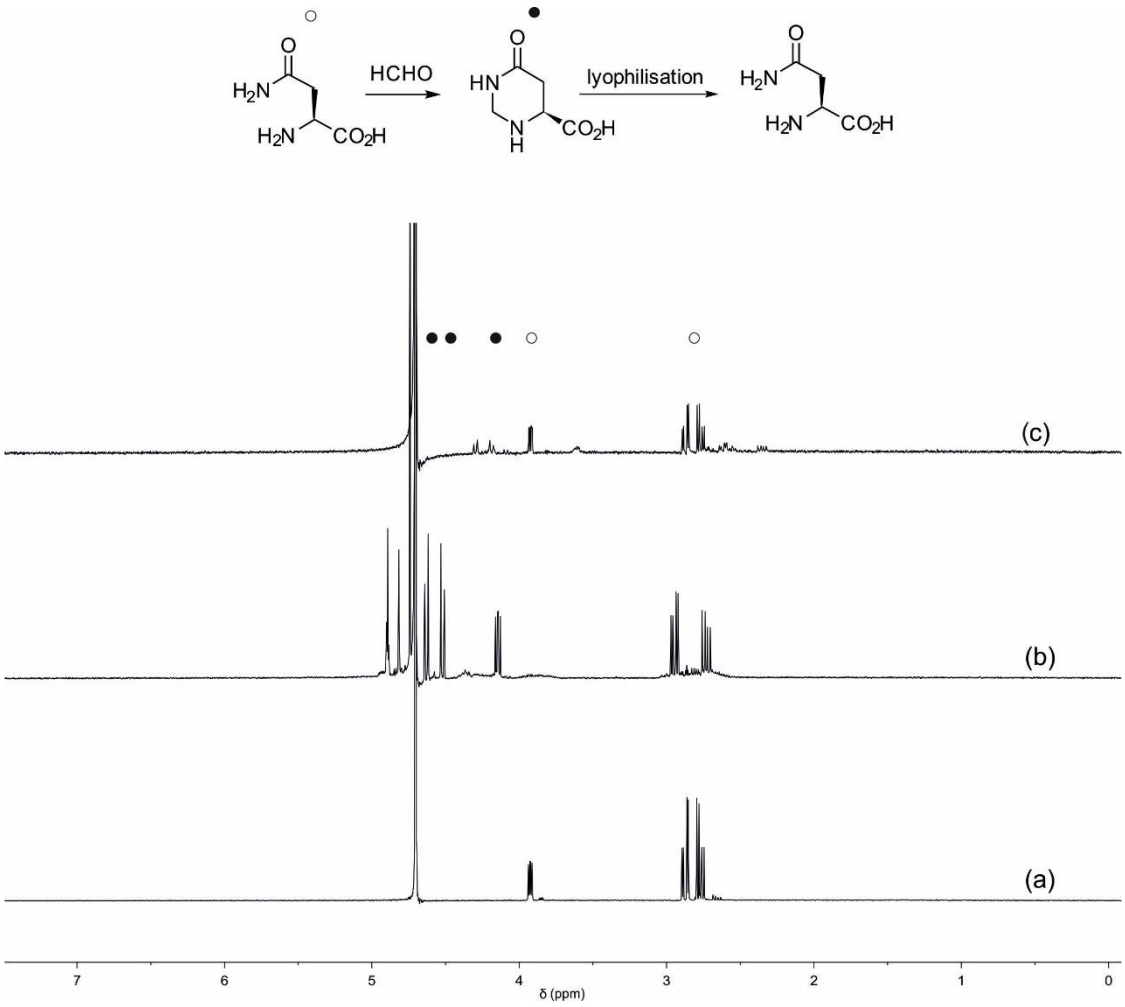

**Appendix Figure 5:** Diketone stability assay with arginine derived **24**. Reversibility assay of **24** (●) product of a reaction of HCHO with arginine (○), using an excess of 1,3-cyclohexadione (DCH, ◊): (a) arginine (4.34 mM); (b) 48 hours after the addition of HCHO (10 equiv.); (c) 72 hours after the addition of 1,3-cyclohexadione (40 equiv.). Note, other spectra were used in the assignment of the structures of the shown products.



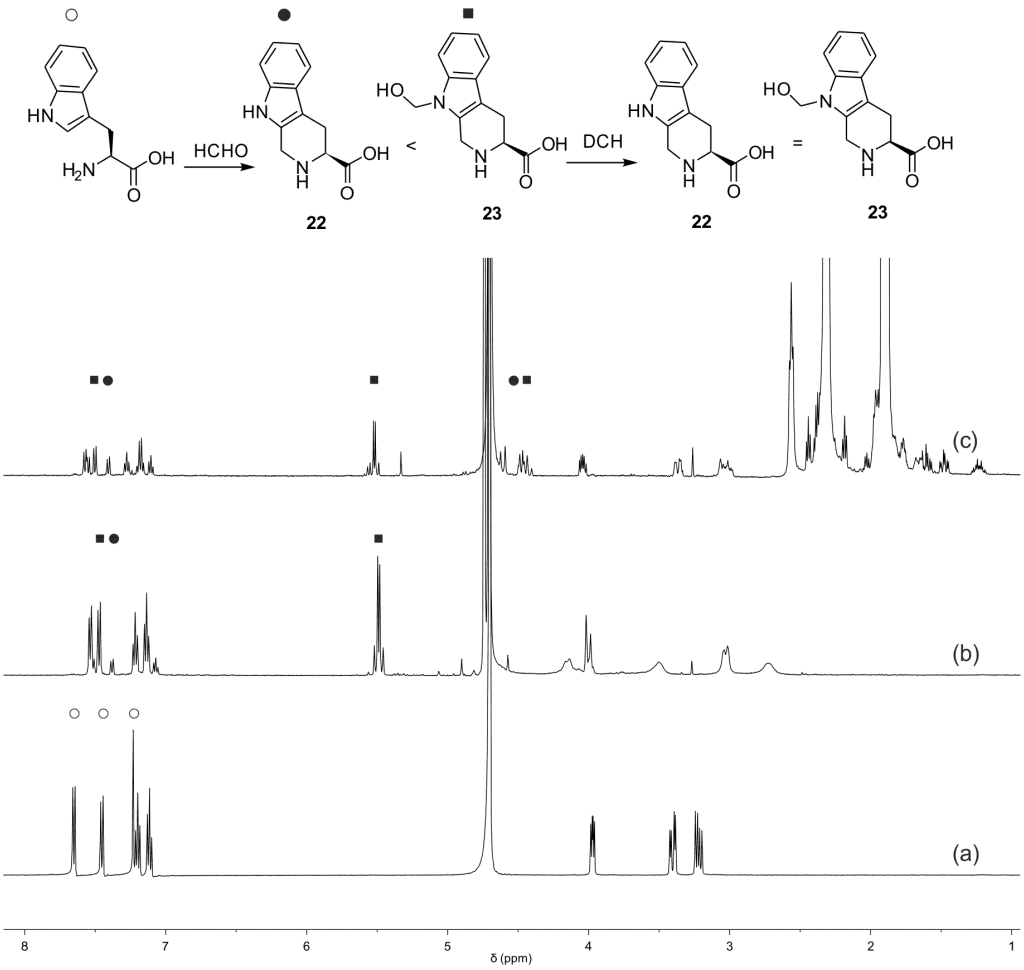

**Appendix Figure 6:** Lyophilisiation stability assay with arginine derived 24. Reversibility assay of 24 (●) product of a reaction of HCHO with arginine (○), using multiple cycles of lyophilisation: (a) arginine (4.34 mM); (b) 48 hour after the addition of HCHO (10 equiv.); (c) after 8 lyophilisation cycles. Note, other spectra were used in the assignment of the structures of the shown products.



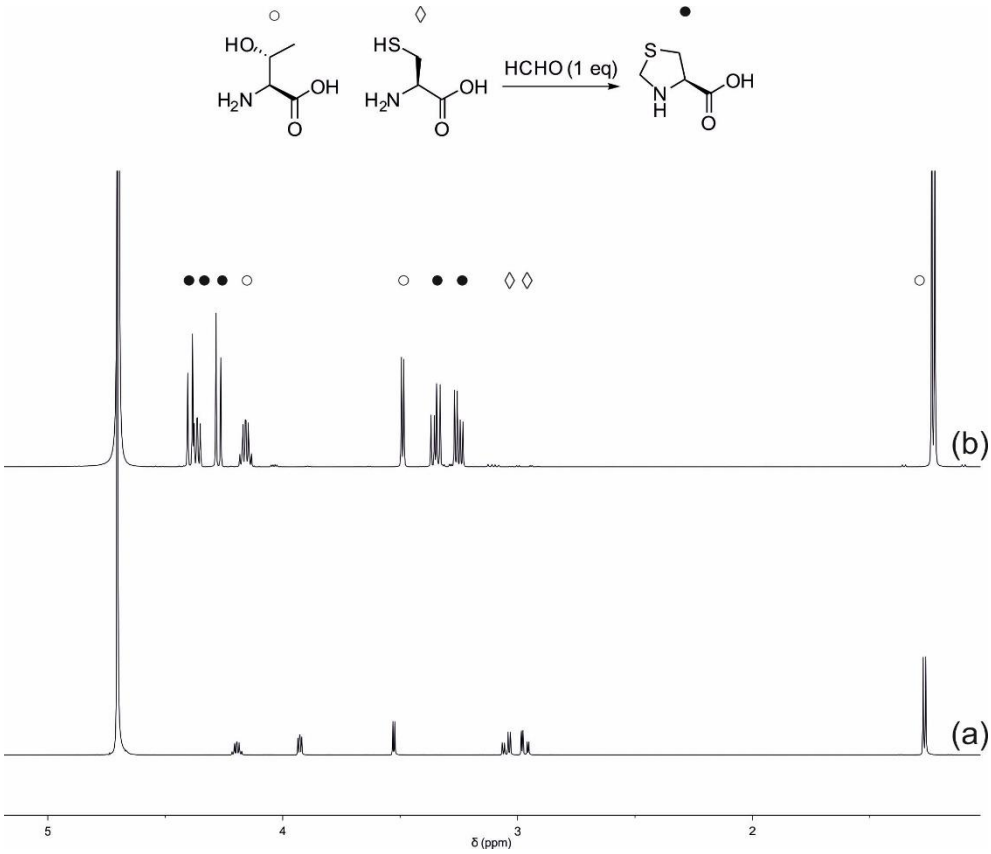

**Appendix Figure 7:** Diketone stability assay with asparagine derived **11**, **12** and **13**. Reversibility assay of **11** (●), **12** (▪) and **13** (▼) products of a reaction of HCHO with Asn (○), using an excess of 1,3-cyclohexadione (DCH, ◊): (a) asparagine (4.34 mM); (b) 48 hours after the addition of HCHO (10 equiv.); (c) 72 hours after the addition of 1,3-cyclohexadione (40 equiv.). Note, other spectra were used in the assignment of the structures of the shown products.



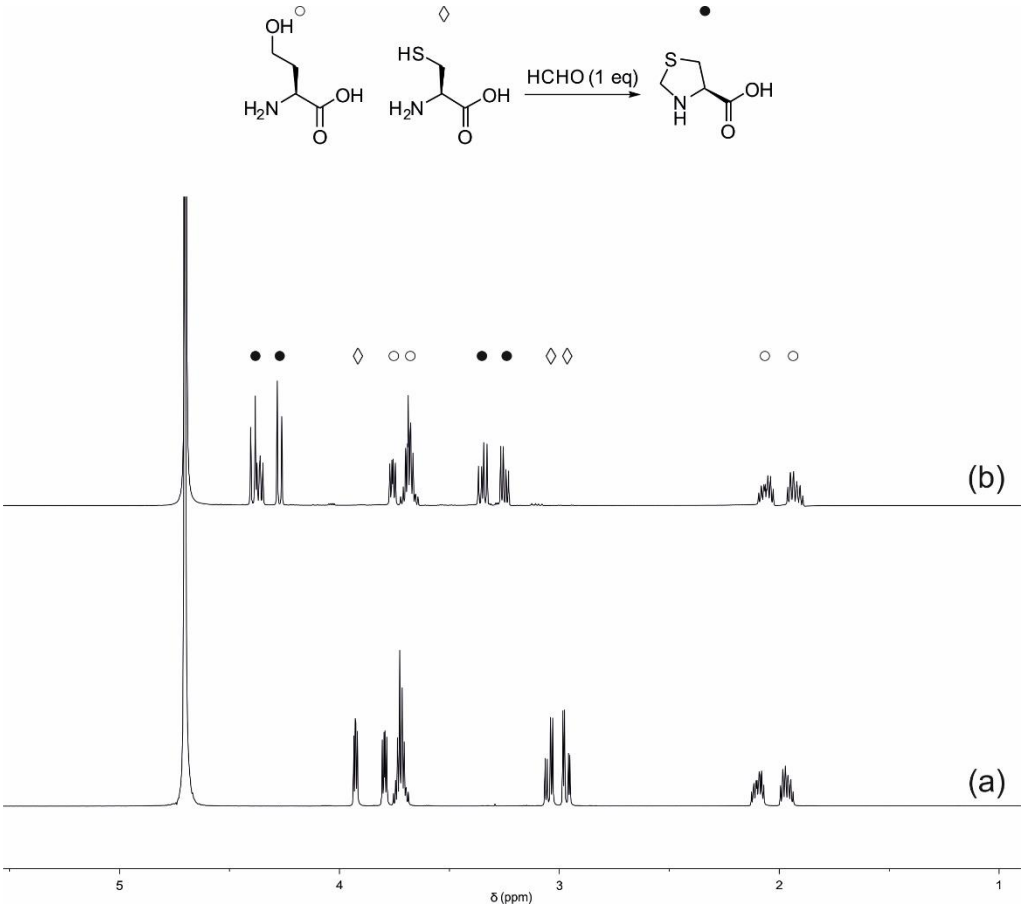

**Appendix Figure 8:** Diketone stability assay with asparagine derived **11**. Reversibility assay of **11** (●) product of a reaction of HCHO with asparagine (○), using multiple cycles of lyophilisation: (a) asparagine (4.34 mM); (b) 48 hour after the addition of HCHO (10 equiv.); (c) after 8 lyophilisation cycles. Note, other spectra were used in the assignment of the structures of the shown products.



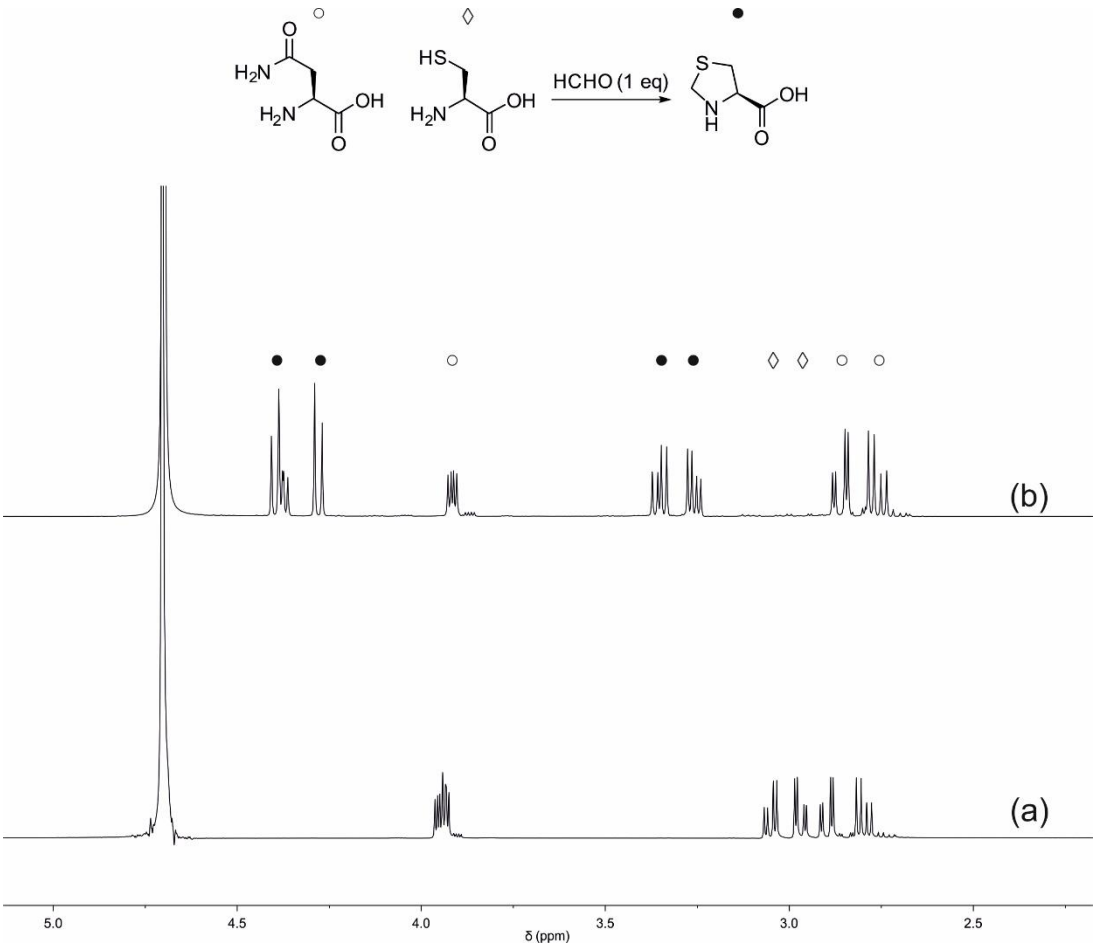

**Appendix Figure 9:** Diketone stability assay with tryptophan derived **22** and **23**. Reversibility assay of **22** (●) and **23** (▪) product of a reaction of HCHO with tryptophan (○), using an excess of 1,3-cyclohexadione (DCH, ◊): (a) tryptophan (4.34 mM); (b) 48 hour after the addition of HCHO (10 equiv.); (c) 48 hours after DCH treatment. Note, other spectra were used in the assignment of the structures of the shown products.



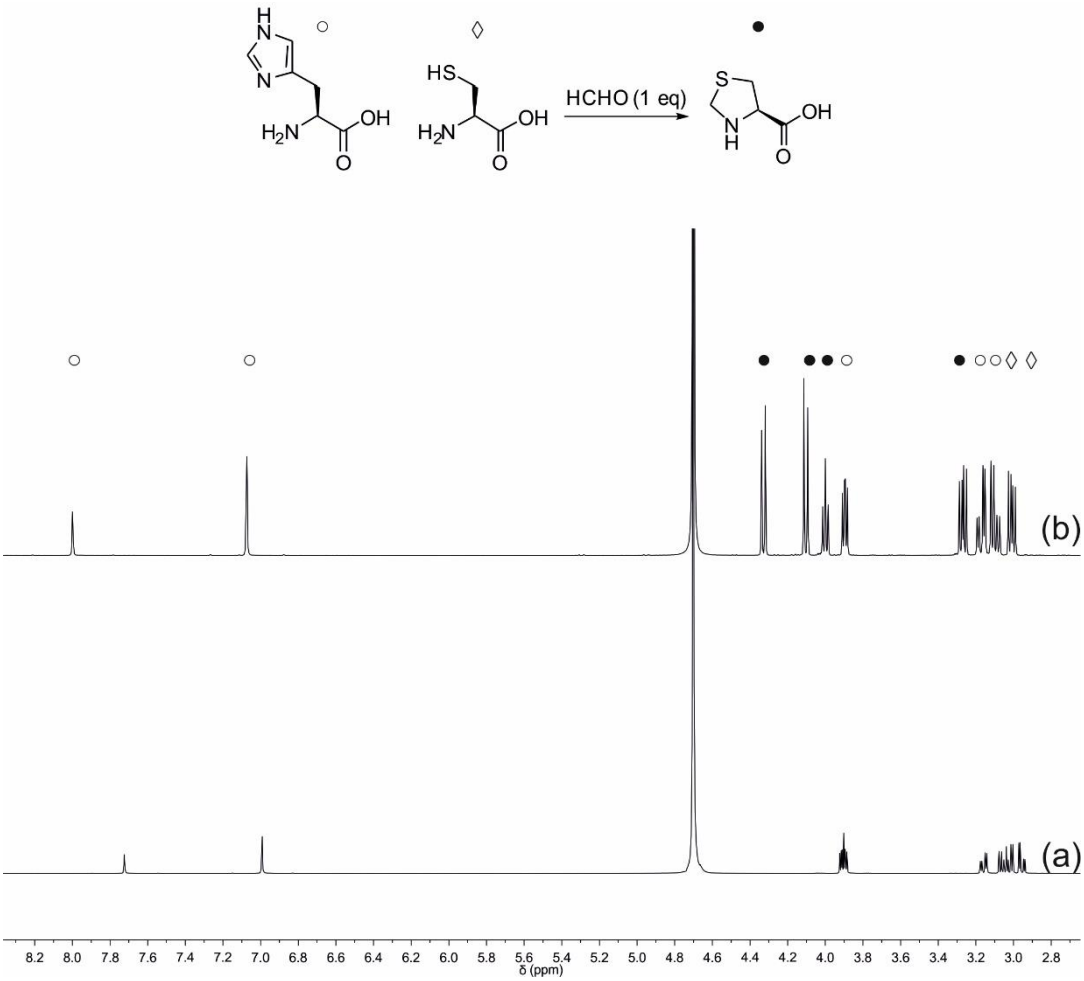

**Appendix Figure 10:** Cysteine and threonine competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and threonine (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●). (a) before addition of HCHO; (b) 14 hours after addition.



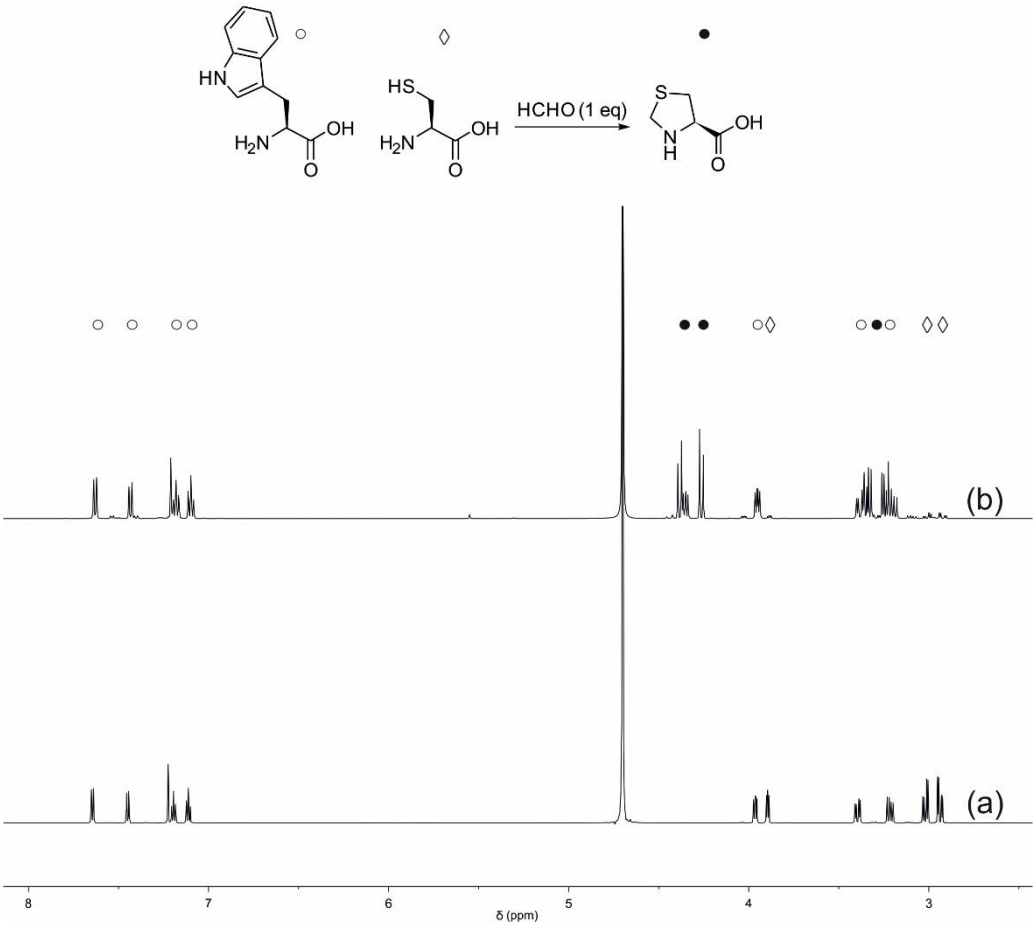

**Appendix Figure 11:** Cysteine and homoserine competition experiment. $^1$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and homoserine (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●). (a) before addition of HCHO; (b) 14 hours after addition.



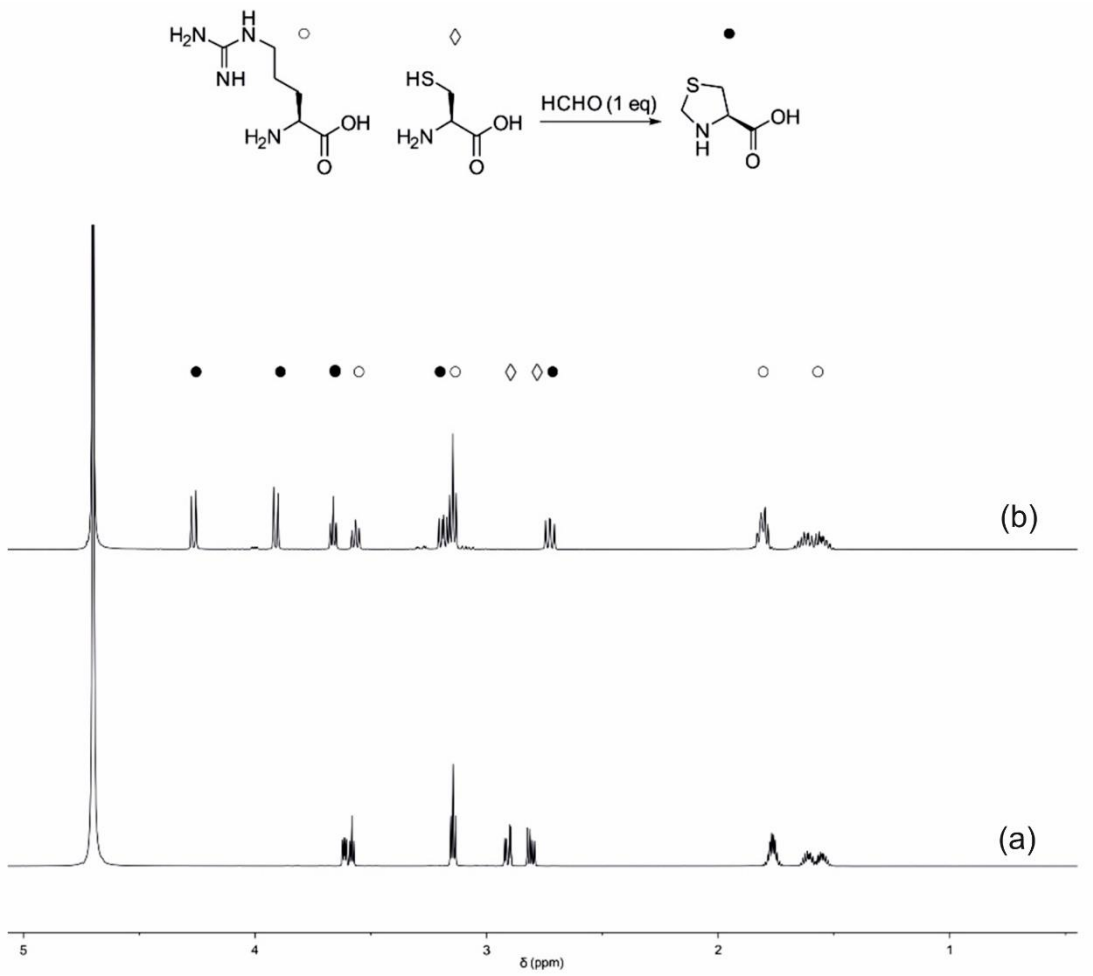

**Appendix Figure 12:** Cysteine and asparagine competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and asparagine (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●): (a) before addition of HCHO; (b) 14 hours after.



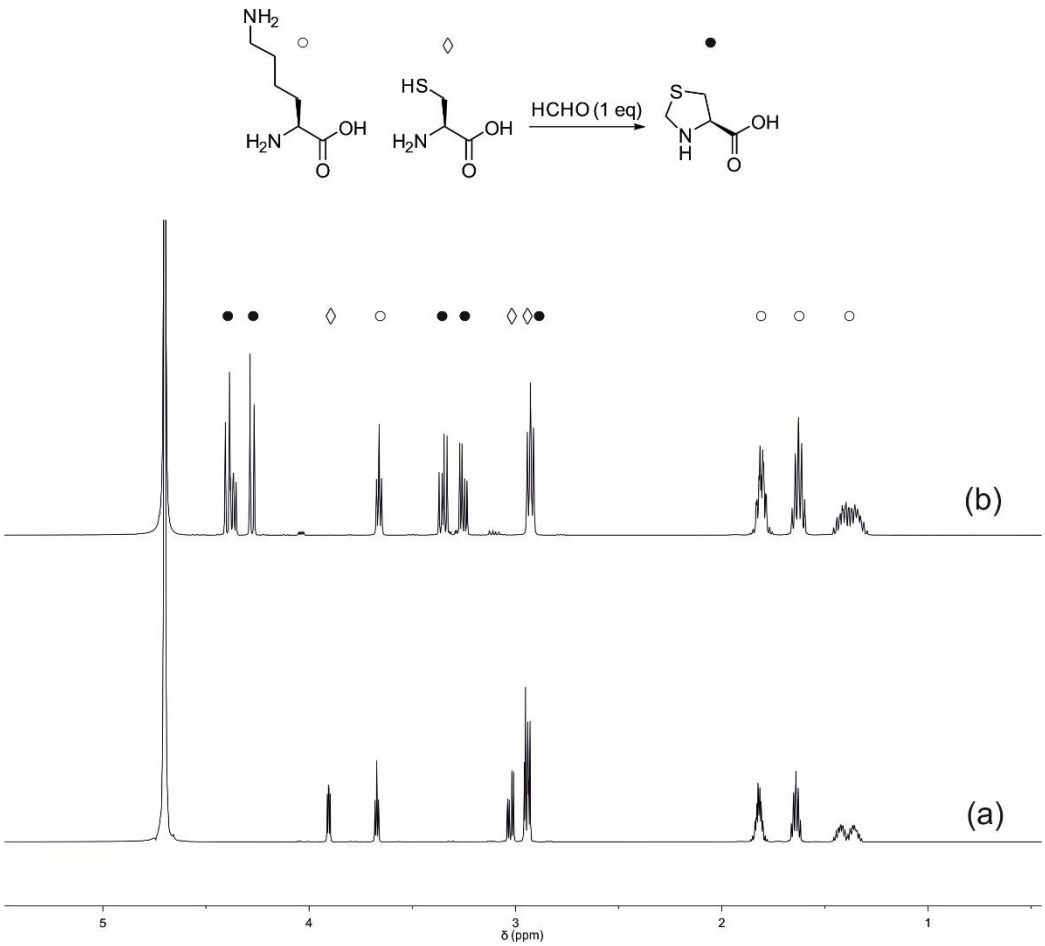

**Appendix Figure 13:** Cysteine and histidine competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and histidine (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●): (a) before addition of HCHO; (b) 14 hours after addition.



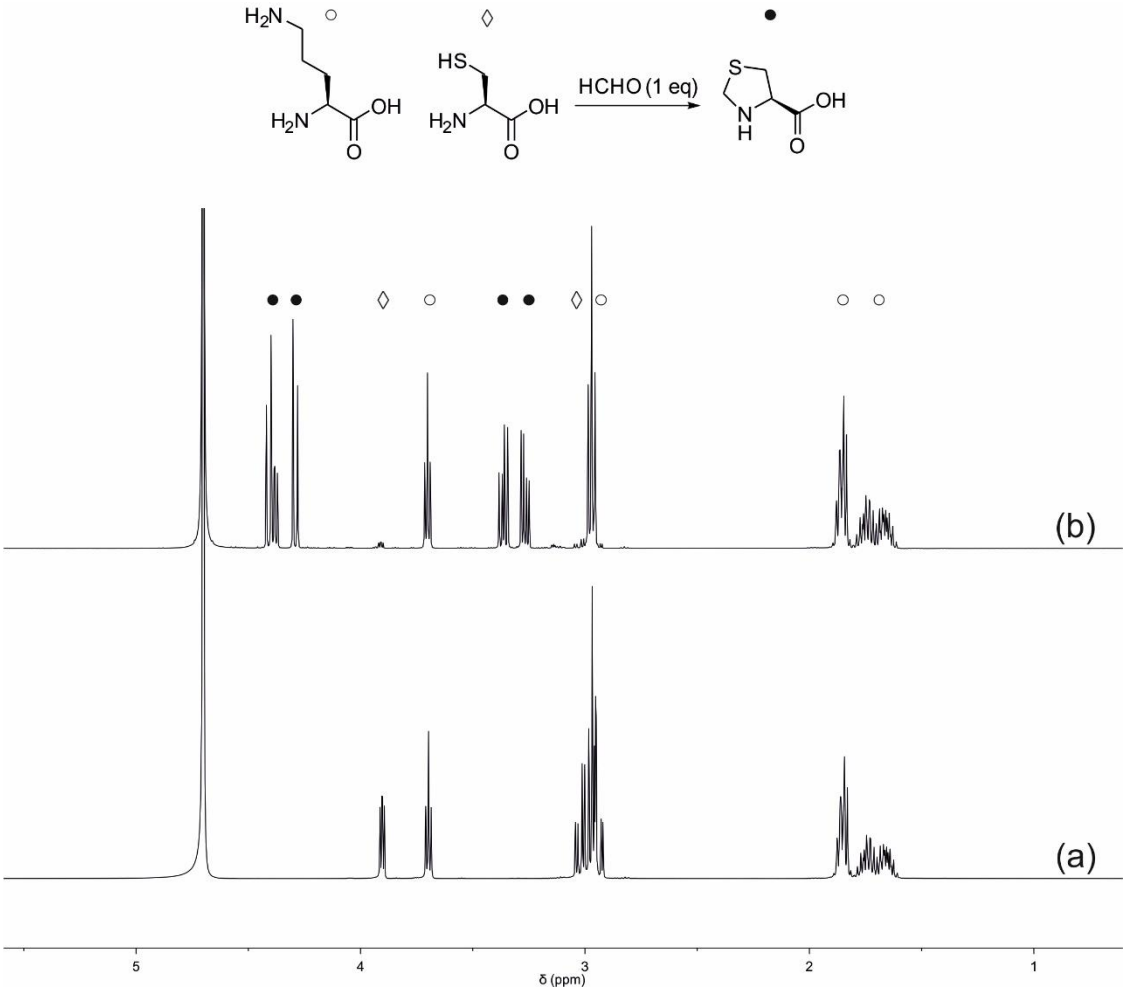

**Appendix Figure 14:** Cysteine and tryptophan competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and tryptophan (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●): (a) before addition of HCHO; (b) 14 hours after addition.




**Appendix Figure 15:** Cysteine and arginine competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and arginine (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●): (a) before addition of HCHO; (b) 14 hours after addition.




**Appendix Figure 16:** Cysteine and lysine competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ◊) and lysine (1 equiv., ○) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●): (a) before addition of HCHO; (b) 14 hours after addition.




**Appendix Figure 17:** Cysteine and ornithine competition experiment. $^{1}$H NMR analysis of the competition assay between cysteine (4.34 mM, ○) and ornithine (1 equiv., ◊) and HCHO (1 equiv.) showing formation of thiazolidine **1** (●): (a) before addition of HCHO; (b) 14 hours after addition.