Studies on the Reactions of Formaldehyde with Protein N-Termini and Related Biomolecules



Tobias John, Schofield Group

Hertford College

University of Oxford

A thesis submitted for the degree of

DPhil in Chemical Biology

December 2021

To my parents
Uschi and Roland

The same thrill, the same awe and mystery, comes again and again when we look at any question deeply enough. With more knowledge comes a deeper, more wonderful mystery, luring one on to penetrate deeper still. Never concerned that the answer may prove disappointing, with pleasure and confidence we turn over each new stone to find unimagined strangeness leading on to more wonderful questions and mysteries-certainly a grand adventure!

Richard P. Feynman - What Do You Care What Other People Think? [1]

Acknowledgements

I should like to thank Prof Chris Schofield and Dr Richard Hopkinson for the interesting project which allowed me to learn at the interface between Chemistry and Biology. I feel very privileged to be able to do something for a living that genuinely excites me. I especially want to thank Chris for having me in his great group, enabling me to enjoy the everyday learning process and the inspiring and happy atmosphere in the group and Oxford. His enthusiastic supervision motivates, inspires and teaches me to be a curious and bold thinker, and it impresses me what can be achieved in a lifetime of hard work, curiosity and kindness. Thanks to Dr Ilaria Pettinati for guidance during my first year, arguably the hardest one during the phage contamination. I thank the entire mass spectrometry facility, especially Elisabete Pires, Ingvild Hvinden and Dr John Walsby-Tickle for help with mass spectrometry. Thanks to Dr Felix Dingler and Dr Christopher Millington for the mouse samples and discussions about aldehydes. I am grateful for Prof Tim Claridge and Dr Nader Amin for running a world-class NMR facility which was an incredible help for my research. I also thank Tika Malla, Dr Shuang Liu and Raphael Reinbold for the great team work during several side projects, and Tim Suits, Klemensas Simelis, Dr Jos Kamps, and Dr Roman Belle for fruitful discussions. I also want to thank Hertford College and my funding bodies EPSRC and GSK, with my GSK supervisor Vipul Patel, along with the programme administrators Nina Stamps and Akane Kawamura, who created a brilliant doctoral training opportunity. I should like to give special thanks to Laura for her continuous support and many fruitful discussions.

Declaration of Autorship

This report, and the work of which it is a record, were carried out by myself unless otherwise acknowledged.

MS/MS samples in Chapter 2 were run and analysed by Elisabete Pires and Svenja Hester. KDM4E used in Chapter 2 was provided by Eidarus Salah.
Metabolomics samples in Chapter 3 were run and analysed by Dr John-Walsby Tickle. Mouse experiments were carried out by Dr Felix Dingler, mouse sample preparation was done by Dr Christopher Millington. MS/MS samples were run and analysed by Svenja Hester.
The work described in Chapter 4 was carried out in close collaboration with Raphael Reinbold.
The work described in Chapter 5 was led by the indicated researchers (Tika R. Malla, Lennart Brewitz, Tarick El-Baba, H. T. Henry Chan, Marc A. Moesser, and Rebecca K. Walters) whilst I did essential supporting work.
A lits of human proteins with N-terminal cysteines, provided in the appendix, has been collated by Beatrice Guintoli, Thomas P. Keeley, and Norma Masson.
Keith A. Gillow, Sam Evans and John McManigle provided the LaTeXtemplate this thesis is based on.

Tobias John
Oxford, December 2021

Abstract

Formaldehyde (HCHO) is a reactive electrophile that is both a human metabolite and a pollutant. Above threshold levels HCHO is carcinogenic, likely as a consequence of its reactions with biomolecules. NMR studies have provided insight into the reactions of HCHO with amino acids and nucleobases, the latter of which may underpin HCHO-driven toxicity. HCHO is produced by the actions of demethylases in the regulation of transcription, likely causing relatively high local nuclear concentrations of HCHO. Endogenous HCHO is sufficient to cause fatal malignancies in mice deficient of HCHO protection mechanisms. The work described in this thesis explores the possibility that HCHO reacts with N-terminal residues in proteins and its biochemical consequences. The N-terminus offers the unique scenario where two nucleophiles, i.e. the N-terminal free amine and an adjacent reactive group (either the side chain of the first amino acid or the adjacent peptide bond) are in close proximity, so enabling stable ring formation upon reaction with HCHO.

Introduction. In Chapter 1, an overview of the chemistry of HCHO and its historical and modern uses are provided. The cellular detoxification mechanisms and known reactions of HCHO with biomolecules are discussed, along with biochemical reactions that release HCHO. This is followed by a description of the role of aldehydes in metabolism in prokaryotes and eukaryotes.

N-Terminal prolines. In Chapter 2, studies testing the hypothesis that histone tails might act as a HCHO sink are reported. Among the four isolated canonical histone tails, H2B shows exceptional susceptibility to reactions with HCHO. The product of the reaction of H2B with HCHO was characterized by MS and NMR, revealing formation of a stable fused 5,5-bicyclic adduct. These findings were then validated in a biological context; the results show that *in situ* generated HCHO results in the same adduct. Studies on the inhibition of 4-oxalocrotonate tautomerase (4-OT), which bears an N-terminal catalytic proline, show that HCHO can inhibit enzyme activity both *in vitro* and *in cellulo*.

N-Terminal cysteines. Competition experiments of N-terminal prolines with free cysteine reveal that cysteine reacts even more efficiently than N-terminal prolines with HCHO. Chapter 3 describes the reactions of N-terminal cysteines with HCHO. Initial studies on model peptides showed that N-terminal cysteines react with HCHO to form a stable thiazolidine; N-terminal threonine or serine residues form the analogous oxazolidine, but those adducts are less stable. Asparagine Synthetase (ASNS), which bears an N-terminal catalytic cysteine was developed as a model to investigate reactions with HCHO with N-terminal cysteine residues *in vitro* and *in cellulo*. ASNS uses its N-terminal cysteine to deamidate glutamine, forming ammonia, which travels through an intramolecular tunnel to react with β-aspartyl-AMP, forming asparagine. HCHO inhibits the N-terminal glutaminase reaction, while the asparagine synthesis activity is retained, as shown by studies with NH_4Cl rather than glutamine. Addition of HCHO leads to an increase in ASNS levels *in cellulo* and depletes intracellular asparagine.

Metampicillin. The reactions of HCHO with nucleophilic drugs are important in medicine. In Chapter 4, NMR was used to investigate the structure of metampicillin, which is produced by reaction of ampicillin with HCHO. Metampicillin, which has been used to treat biliary infections in humans and animals, was proposed to have HCHO-derived imine or hemiaminal structures. In collaboration with Raphael Reinbold, it is shown that metampicillin is a cyclic aminal.

CV19. As a consequence of the CV-19 pandemic, collaborative work was carried out on nucleophilic cysteine proteases. In Chapter 5, mass spectrometry based assays were developed to study inhibition of two proteases of SARS-CoV-2, namely M^{pro} and PLPro.

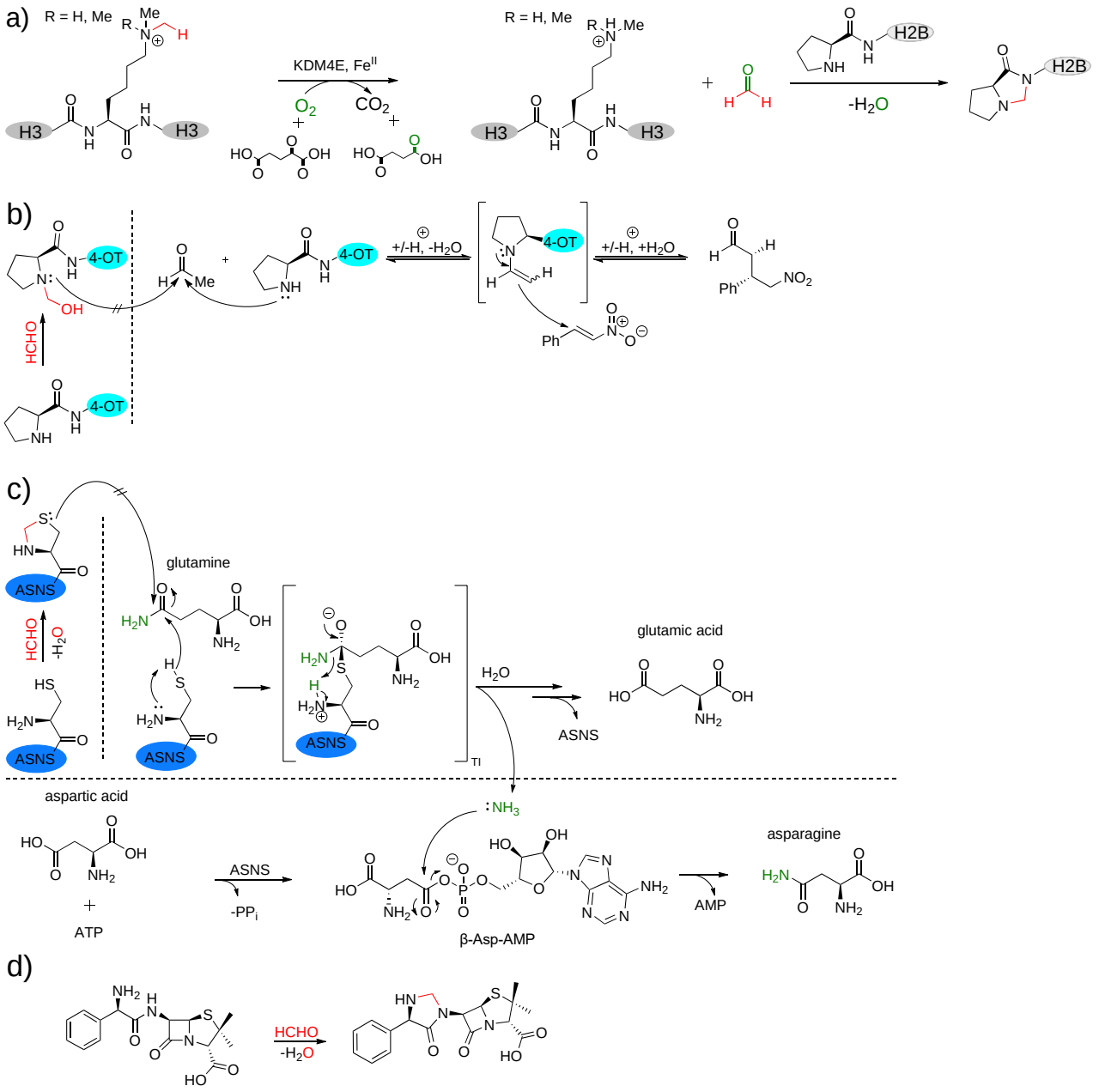


Figure 1 | Graphical Abstract. a) Demethylation of H3K9Me$_3$ produces HCHO *in situ*, which reacts with an H2B fragment present in the reaction mixture to form a fused 5,5-bicycle. **b)** 4-Oxalocrotonate tautomerase (4-OT) catalyses the Michael addition of acetaldehyde to *trans*-β-nitrostyrene via a mechanism involving an N-terminal catalytic proline; HCHO inhibits 4-OT *in vitro* and *in cellulo*. **c)** HCHO forms an N-terminal thiazolidine with the N-terminal cysteine of ASNS, inhibiting the glutaminase partial reaction. Despite the presence of HCHO, the synthase site is still intact, forming asparagine in the presence of NH_4Cl. **d)** Ampicillin reacts with HCHO to form a cyclic aminal, the correct structure for metampicillin, rather than the previously proposed imine or hemiaminal structures.

Contents

List of Abbreviations **ix**

1 Introduction **1**
1.1 Formaldehyde - Much More than the Simplest Aldehyde 2
1.2 HCHO in Biology 18
1.3 Aldehydes and Ketones in Mammals 40

2 Formaldehyde and N-terminal Prolines **54**
2.1 Abstract 56
2.2 Introduction 56
2.3 Results 62
2.4 Discussion 86
2.5 Supplementary Information 88
2.6 Methods 109

3 Formaldehyde and N-Terminal Cysteines **144**
3.1 Abstract 145
3.2 Introduction 146
3.3 Results 152
3.4 Discussion 176
3.5 Supplementary Information 179
3.6 Methods 188

4 Metampicillin **211**
4.1 Abstract 212
4.2 Introduction 212
4.3 Results 213
4.4 Discussion 230
4.5 Methods 231

5 SARS-CoV-2 Proteases **237**
5.1 Abstract 238
5.2 Introduction 238
5.3 Results 240

Bibliography **251**

Appendices **280**

List of Abbreviations

2OG 2-Oxoglutarate

2TY . . . . . 2-Tryptone/yeast extract

3CLpro . . . . . 3-Chymotrypsin-like cysteine protease

3mT . . . . . (3-Methyl)thymidine

4-OT . . . . . 4-Oxalocrotonate Tautomerase EC 5.3.2

8-oxoG 8-Oxo-7,8-dihydroguanine

20S CP . . . . . 20S Core Particle

A280 . . . . . Absorbance at 280 nm

AA-ICL . . . . AcH-derived Interstrand Crosslink

ABG . . . . . p-Aminobenzoylglutamate

ABH . . . . . AlkB Homologue

AcH . . . . . Acetaldehyde

ADH . . . . . Alcohol Dehydrogenase

ADO . . . . . Cysteamine (2-aminoethanethiol) Dioxygenase

AGA . . . . . Aspartylgucosaminidase

AGE . . . . . Advanced Glycation End-products

ALDH . . . . . Aldehyde Dehydrogenase

AlkB . . . . . Alkylation Repair Protein B

ALL . . . . . Acute Lymphoblastic Leukemia

AMC 7-Amino-4-methylcoumarin

AO . . . . . Amino Oxidase

APS . . . . . Ammonium Persulfate

ASD . . . . . Asparagine Synthetase Deficiency

ASNase . . . . . Asparaginase

ASNS . . . . . Asparagine Synthetase

ATE1 . . . . . Arginyltransferase 1

ATR . . . . . Attenuated Total Reflection

AU . . . . . Arbitrary Unit

Aze . . . . . Azetidine

BSA . . . . . Bovine Serum Albumin

CDW . . . . . Cell Dry Weight

CENP-V . . . . Human Centromere Protein V

CHCA . . . . . α-Cyano-4-Hydroxycinnamic Acid

ChIP . . . . . Chromatin Immunoprecipitation

CD . . . . . Circular Dichroism

CHD 1,3-Cyclohexanedione

CoA . . . . . Coenzyme A

COSY . . . . . Correlation Spectroscopy

CoV-2 . . . . . Coronavirus Type 2

CpG . . . . . 5'-C-phosphate-G-3'

CV . . . . . Column Volume

CYP450 . . . . Cytochrome P450

d . . . . . Doublet

Da . . . . . Dalton (mass unit)

DMEM . . . . . Dulbecco's Modified Eagle's Medium

dG . . . . . Deoxyguanine

DMGO . . . . . Dimethylglycine Oxidase

DMSO-d_6 . . . Deuterated Dimethylsulfoxide

DNA . . . . . Deoxyribonucleic Acid

dr . . . . . Diastereomeric Ratio

E. coli . . . . . *Escherichia coli*

ee . . . . . Enantiomeric Excess

EstD . . . . . *S*-Formylglutathione Hydrolase EC 3.1.2.12

EV . . . . . Empty Vector

FA . . . . . Fanconi Anemia

FAD . . . . . Flavin Adenine Dinucleotide

$FADH_2$ . . . . . Flavin Adenine Dinucleotide, reduced form

FANCD2 . . . . Fanconi Anemia Complementation group D2

Fch . . . . . Methenyltetrahydrofolate Cyclohydrolase EC 3.5.4.9

FDH . . . . . Formaldehyde Dehydrogenase

FGH . . . . . *S*-Formylglutathione Hydrolase

Fhs . . . . . Formyltetrahydrofolate Synthetase EC 6.3.4.3

FIH . . . . . Factor Inhibiting Hypoxia Inducible Factor

Fmoc . . . . . Fluorenylmethyloxycarbonyl

Fpg . . . . . Formamidopyrimidine-DNA Glycosylase

FPLC . . . . . Fast Protein Liquid Chromatography

FTO . . . . . Fat mass and Obesity-Associated Protein

GABA . . . . . Gamma-Aminobutyric Acid

Gfa . . . . . Glutathione-Dependent Formaldehyde-Activating Enzyme

GLS . . . . . Glutaminase

GSH . . . . . Glutathione

Hb . . . . . Hemoglobin

HCHO . . . . . Formaldehyde; note in this thesis HCHO normally refers to HCHO or hydrated HCHO, i.e. $H_2C(OH)_2$

H^{13}CHO . . . . ^{13}C-Labelled Formaldehyde

HDAC . . . . . Histone Deacetylase

HEPES . . . . . 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic Acid

HIV . . . . . Human Immunodeficiency Virus

HMBC . . . . . Heteronuclear Multiple-bond Correlation Spectroscopy

HMG . . . . . *S*-Hydroxymethylglutathione

HMG-CoA . . . Hydroxymethylglutaryl-CoA

HOBt 1-Hydroxybenzotriazole

HOMO . . . . . Highest Occupied Molecular Orbital

HPLC . . . . . High Performance Liquid Chromatography

Hps . . . . . Hexulose-6-phosphate Synthase EC 4.1.2.43

HRMS . . . . . High-Resolution Mass Spectrometry

HSQC . . . . . Heteronuclear Single-Quantum Correlation Spectroscopy

ICL . . . . . Interstrand Crosslink

IEC . . . . . Ion Exchange Chromatography

IPTG . . . . . Isopropyl β-*D*-Thiogalactopyranoside

IR . . . . . Infrared

JmjC . . . . . Jumonji C

kcal . . . . . Kilocalorie

KDM . . . . . Histone Lysine Demethylases

LC-MS . . . . . Liquid Chromatography-Mass Spectrometry

LUMO . . . . . Lowest Unoccupied Molecular Orbital

m . . . . . Multiplet

m6A . . . . . (6-Methyl)adenosine

MALDI . . . . . Matrix-Assisted Laser Desorption/Ionisation

MAP . . . . . Methionine aminopeptidase

Me . . . . . Methyl

MERS . . . . . Middle East Respiratory Syndrome

MF . . . . . Melamine Formaldehyde

MGF . . . . . (4*S*,9*R*)-9-(Carboxymethylcarbamoyl)-7-oxo-1,3,8-thiadiazecane-4-carboxylic acid

MGO . . . . . Methylglyoxal

M^{pro} . . . . . Main Protease

MQ . . . . . Milli-Q

MRE . . . . . Mean Residue Ellipticity

MS . . . . . Mass Spectrometry

Mtda . . . . . Methylenetetrahydrofolate Dehydrogenase EC 1.5.1.5

MTCA . . . . . 2-Methylthiazolidine-4-carboxylic acid

NAD . . . . . Nicotinamide Adenine Dinucleotide

NMR . . . . . Nuclear Magnetic Resonance

NOESY . . . . . Nuclear Overhauser Effect Spectroscopy

NOG . . . . . N-oxalylglycine

nt . . . . . Nucleotide(s)

Ntn . . . . . N-terminal nucleophile

ORF . . . . . Open Reading Frame

PA . . . . . Penicillin G Acylase

PAP . . . . . Prolyl Aminopeptidase

PCO . . . . . Plant Cysteine Oxidase

PdG . . . . . N^2-Propanoguanine

PF . . . . . Phenol Formaldhehyde

PGF . . . . . (*S*)-1-(((*R*)-2-Amino-3-(carboxymethylamino)-3-oxopropylthio)methyl)-5-oxopyrrolidine-2-carboxylic acid

Phi . . . . . 6-Phospho-3-hexuloisomerase EC 5.3.1.27

Pip . . . . . Piperidine

PLP . . . . . Pyridoxal 5'-phosphate

PLPro . . . . . Papain-like Protease

ppm . . . . . Parts per million

q . . . . . Quartet

QM/MM . . . . Quantum Mechanics/Molecular Mechanics

RIN . . . . . RNA Integrity Number

RNA . . . . . Ribonucleic Acid

RPMI . . . . . Roswell Park Memorial Institute

RuMP . . . . . Ribulose Monophosphate

s . . . . . Singlet

SAM . . . . . *S*-Adenosylmethionine

SARS . . . . . Severe Acute Respiratory Syndrome

SDS-PAGE . . . Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis

SFG . . . . . *S*-Formylglutathione

SHMT . . . . . Serine Hydroxymethyltransferase

S_N1 . . . . . Unimolecular Nucleophilic Substitution

S_NAr . . . . . Nucleophilic Aromatic Substitution

SPE-MS . . . . Solid Phase Extraction coupled to Mass Spectrometry

SPPS . . . . . Solid Phase Peptide Synthesis

SSAO . . . . . Semicarbazide-Sensitive Amine Oxidase

TEMED . . . . *N, N, N', N'*-Tetramethylethylenediamine

TFA . . . . . Trifluoroacetic acid

THF . . . . . Tetrahydrofolate

TLC . . . . . Thin Layer Chromatography

TLS . . . . . Translesion Synthesis

TOF . . . . . Time-of-Flight

TOCSY . . . . . Total Correlation Spectroscopy

Tris . . . . . Tris(hydroxymethyl)aminomethane

UF . . . . . Urea Formaldehyde

UV . . . . . Ultraviolet

Vis . . . . . Visible

1 Introduction

»Ausgeräuchert, das ist famos«, sagte er gesprächig und etwas ungereimt, indem er sich die Hände wusch und trocknete. »Ja, Methylaldehyd, das hält die stärkste Bakterie nicht aus, - H_2CO, aber es sticht in die Nase, nicht? Selbstverständlich ist strengste Sauberkeit eine Grundbedingung...«

Thomas Mann - Der Zauberberg [2]

'Fumigated it, eh? That's ripping,' he said loquaciously and rather absurdly, as he washed and dried his hands. 'Methyl aldehyde; yes, that's too much for the bacteria, no matter how strong they are. H_2CO. But it's a powerful stench. Of course, perfect sanitation is absolutely essential.'

Thomas Mann - The Magic Mountain [3]

Contents

1.1 Formaldehyde - Much More than the Simplest Aldehyde 2
1.2 HCHO in Biology 18
1.3 Aldehydes and Ketones in Mammals 40

1.1 Formaldehyde - Much More than the Simplest Aldehyde

1.1.1 History of Formaldehyde

The first scientist to describe formaldehyde (HCHO) was Alexander Michailowitsch Butlerow in 1855.[4] Its synthesis was first achieved by August Wilhelm von Hofmann in 1868 by passing methanol/air vapors over a hot platinum wire.[5] Multifaceted applications of HCHO were soon discovered. HCHO-based resins as adhesives for particle board were first used industrially in the 1940s.[6] The first detrimental health effects caused by HCHO exposure in prefabricated houses were reported in the 1960s.[7] Following these observations, permissible exposure limits were established, which are today 0.75 parts HCHO per million parts of air (0.75 ppm) in the workplace.[8] The first evidence that HCHO is carcinogenic was accrued on rodents in 1980.[9] HCHO is perceived as most frequent and best-studied indoor air pollutant, and HCHO concentration in outdoor urban environments is increasing.[7]

Industrially, HCHO is produced from the catalytic oxidation of methanol or methane.[10] HCHO is crucial for a wide range of products essential to the world's economy, which comes with constant exposure through cosmetics, furniture and flooring. Hence, exogenous HCHO is omnipresent in contemporary life.

Incidents involving acute HCHO exposure impacting public health have been reported: In the 1980s, around 99,000 lbs of urea-HCHO resin were introduced into the environment in Crown Point, Alaska by an accidental spill from a railroad tanker car, causing health problems that lasted up to 60 days (e.g. fatigue, nosebleeds, shortness of breath) in citizens living in the area.[11] In 2015, the US flooring retailer Lumber Liquidators was accused of selling laminate flooring with off-gasing HCHO well above threshold levels; this company installs more than 100 million square feet of laminate each year, resulting in many citizens having chronic exposure to HCHO.[12] A prominent example of indoor air contamination by HCHO arose from trailers provided for emergency housing after Hurricane Katrina in 2007; over 200 complaints were filed by people suffering from sinus infections and burning eyes, and HCHO levels in trailers were measured to be as high as 1.2 ppm.[13] A popular hair straightening product is the Brazilian Blowout, which contains methylene glycol (the hydrated

form of HCHO) contents as high as 10%; the product was branded HCHO-free which caused the company to be sued for failure to disclose a carcinogen and deceptive advertising. Brazilian Blowout is used in hair salons, with questionable safety measures like masks on and windows open.[12] In fact, the mortality from myeloid leukemia is associated with the duration of HCHO exposure among pathologists, funeral industry workers, and anatomists; HCHO is used to embalm bodies to prevent decomposition of tissues.[14] 'New car smell' partially arises from off-gasing volatile chemicals like HCHO which are used to build interiors of cars.[15]

1.1.2 HCHO and the Origin of Life

The RNA world hypothesis states that self-replicating RNA existed before DNA and proteins. It is therefore of interest to identify potential pathways that led from prebiotic inorganic chemistry to abiogenesis.[12] The formose reactions describes the self-condensation of HCHO resulting in mixtures of carbohydrates, including riboses, which are key building blocks for RNA. The formose reaction has hence been proposed to be involved in prebiotic chemistry and the origin of life. Both glycolaldehyde - the simplest sugar - and HCHO have been detected in outer space, providing potential entry points for abiotic organic synthesis (Figure 1.1).[16, 17] In gas-phase conditions, resembling outer space environments, the formose reaction proceeds via hydroxymethylene (the singlet carbene H–C–OH) and its tautomer HCHO in a carbonyl-ene reaction, forming glycolaldehyde.[18]

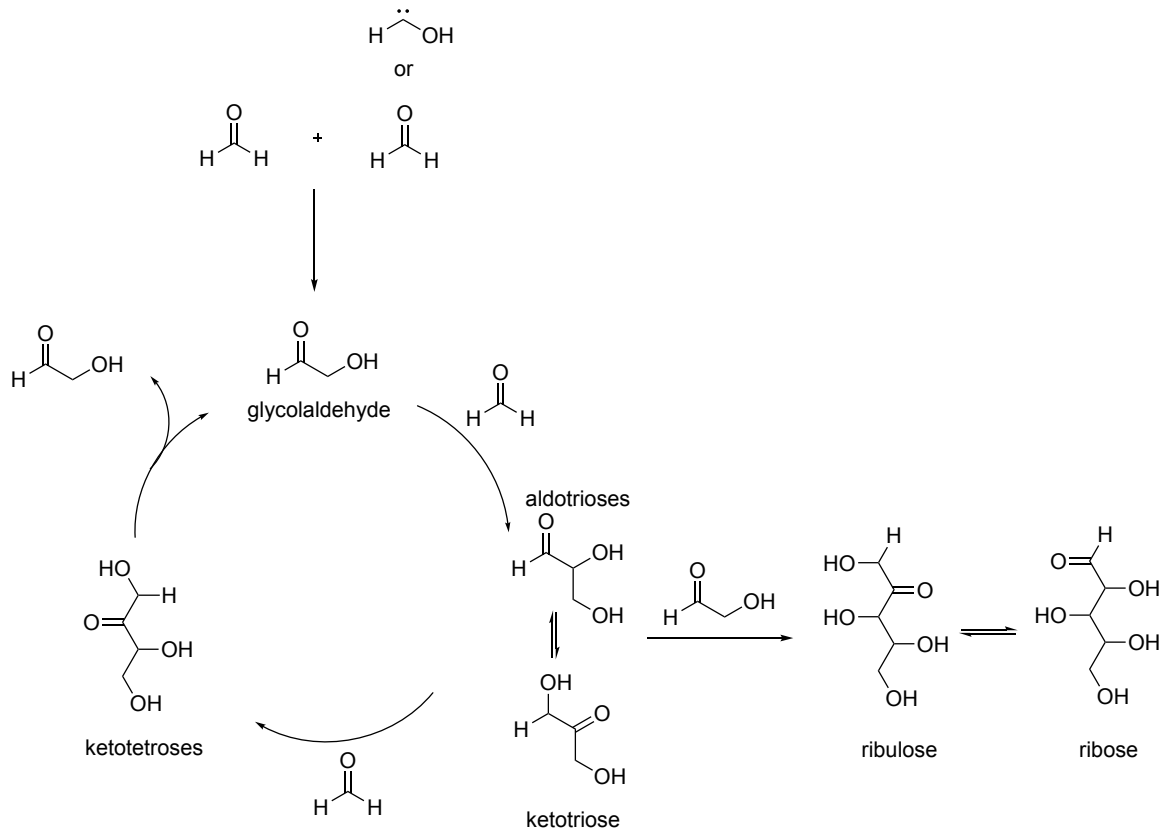


Figure 1.1 | Simplified mechanism of the formose reaction. The first step of the formose reaction in aqueous conditions forms glycolaldehyde with two HCHO molecules, which is formed by an unknown mechanism, possibly involving radiation.[19] An aldol reaction between glycolaldehyde and HCHO yields aldotrioses which isomerize to ketotrioses. Another elongation of the carbohydrate backbone with HCHO yields aldotetroses, which undergo a retro-aldol reaction to form the simplest sugar again, glycolaldehyde, hence generating an autocatalytic system. Alternatively, the ketotriose can react with glycolaldehyde to form ribulose, which in turn can isomerize to ribose, a building block of RNA. [20]

In 2008 the carbene hydroxymethylene was captured in a noble gas matrix at 12 K from glyoxylic acid via high-vacuum flash pyrolysis; surprisingly, hydroxymethylene rearranges to its thermodynamically more favored HCHO isomer by quantum mechanical tunnelling with a half-life of 2 h through a energy barrier of 30 kcal/mol (Figure 1.2), which provides an explanation why hydroxymethylene has not been detected in outer space previously.[21]

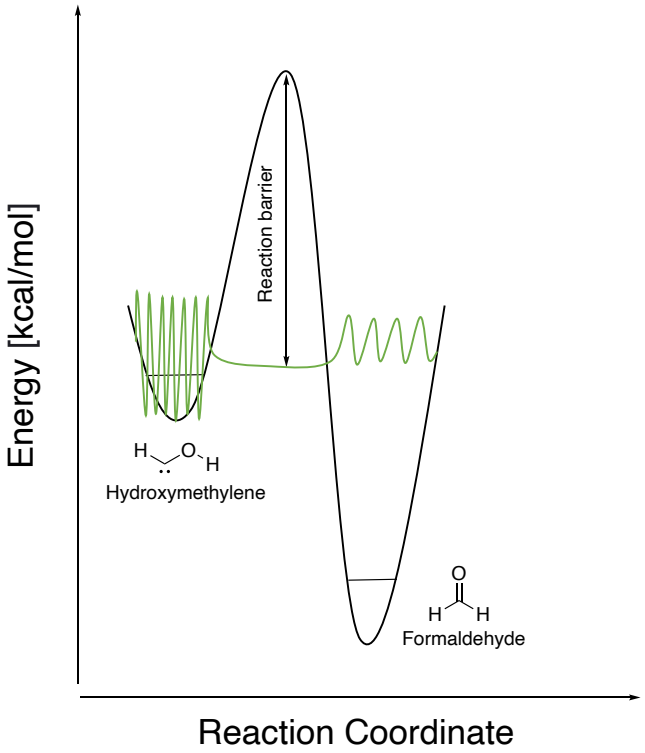


Figure 1.2 | Tunnelling of hydroxymethylene. In [21] isolated hydroxymethylene has been shown to decompose to HCHO with a reaction barrier of around 30 kcal/mol. The study was conducted at 11 K, a temperature where the thermal energy of hydroxymethylene is not sufficient to overcome the reaction barrier. Instead, the waveform of hydroxymethylene (shown in green) reaches into the other energy well by quantum tunnelling. Figure changed after [22].

1.1.3 Industrial Production of HCHO

HCHO is one of the most important feedstock chemicals due to its high reactivity and versatility; it is used in industry for paints, resins, polymers, adhesives, and resins.[23] In the wood panels industry, melamine formaldehyde (MF), urea formaldehyde (UF) and phenol formaldhehyde (PF) are the major resins used.[12] MF is further used to make tableware, and migration of both monomers melamine and HCHO into food-simulating solvents like 4% acetic acid has been shown.[24] MF resin is produced by reacting melamine and excess HCHO in aqueous solution (Figure 1.3 a); with longer reaction time and higher temperature, the hexamethylolmelamine monomers cross-link to form hydroxymethylesters (Figure 1.3 b, equation 1) or melamin reacts with methylolamine to form methylene bridges (Figure 1.3 b, equation 2); prolonged reaction times will lead to strongly cross-linked, very hard, heat resistant and insoluble resin.[25] However, if the reaction is stopped at an earlier time point, an MF foam is produced, which is used for noise cancelling and as a fire retardant (e.g. Basotect® by BASF).[26]

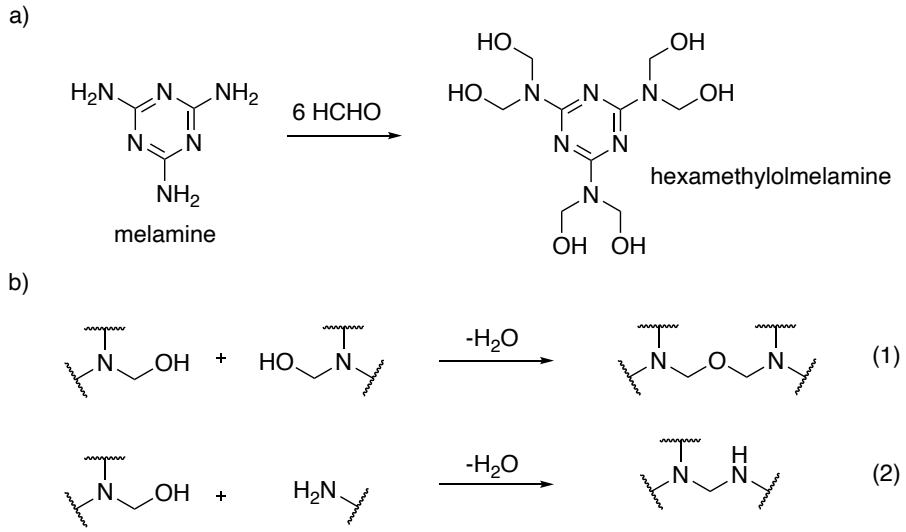


Figure 1.3 | Formation of melamine resin. a) Hexamethylolmelamine is formed after addition of HCHO to melamine. **b)** Potential ways to cross-link two hexamethylolmelamine monomers (first equation) or to cross-link one hexamethylolmelamine monomer with melamine starting material (second equation), resulting in melamine resin (duroplast) after prolonged reaction time and heating.

The global HCHO production is estimated to be 30 million tons.[27] Nowadays, the three leading industrial processes for HCHO production are the Formox process, the methanol ballast process, and the water ballast process (also known as BASF process).[23] The BASF process uses silver as a catalyst

to form HCHO via two different pathways, involving endothermic dehydrogenation and exothermic oxidation, which occur simultaneously (Figure 1.4); the reaction is achieved by passing a mixture of air and methanol over a silver catalyst with an operating temperature of 600 to 700 °C; the resulting HCHO is absorbed in water, forming formalin.[12, 28] The methanol ballast process is analogous to the BASF process, but without feeding extra water to the reactant mixture.[29] The Formox process employs molybdenum oxides as catalysts which are usually doped with iron.[30]

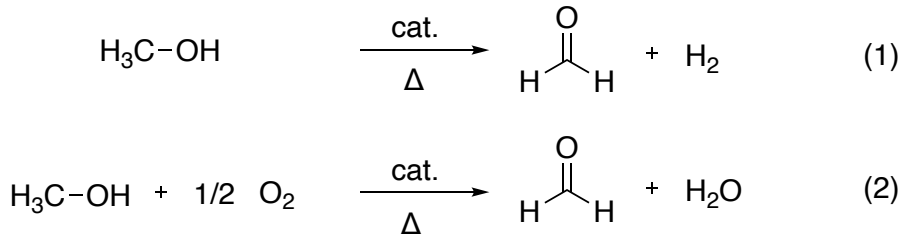


Figure 1.4 | Industrial production of HCHO. The first equation shows the endothermic dehydrogenation of methanol, forming HCHO and hydrogen. The second equation shows the exothermic oxidation, forming HCHO and water. Both reactions occur simultaneously during the BASF process.[28]

1.1.4 Chemical Properties of HCHO

The carbonyl group of aldehydes and ketones is a short (1.21 Å),[31] strong (around 179 kcal/mol),[32] and polar bond (electronegativity of C vs O, 2.55 vs 3.44, Pauling scale); both the oxygen and the carbon atom are sp^2 hybridised and are hence in the same plane. The angles on the carbon are round about 120°; perpendicular to the molecular plane are the p-orbitals of carbon and oxygen, which form the π bond; as oxygen is more electronegative than carbon, the carbonyl is polar which results in an electrophilic character of the carbon and a nucleophilic and basic character of the oxygen; the ^{1}H spectrum of formyl hydrogen atoms have a characteristic peak between δ_H 9 and 10 ppm; this is because of the π electrons which induce a local magnet field which enhances the external one; additionally, the positive partial charge of the carbonyl carbon atom has a deshielding effect.[33] Aqueous solutions of HCHO, called formalin, form the hydrate methyleneglycol and polymethyleneglycols. Commercial formalin contains 10% methanol as stabilizer to suppress polymerisation and precipitation by incorporation of terminal methoxymethyl groups (Figure 1.5 a). Anhydrous sources of HCHO are paraformaldehyde (polyoxymethylene, typically 8-100 units) and 1,3,5-trioxane ($C_3H_6O_3$), which is the cyclic trimer of HCHO (Figure 1.5 a). If commercial paraformaldehyde is suspended in H_2O and heated until a colorless solution is obtained, more than 98% of HCHO species is methyleneglycol, followed by the hydrated dimeric form of HCHO with 1%; minor species present are formic acid, methanol, non-hydrated HCHO and MeOH-capped HCHO oligomers (Figure 1.5 b).[34]

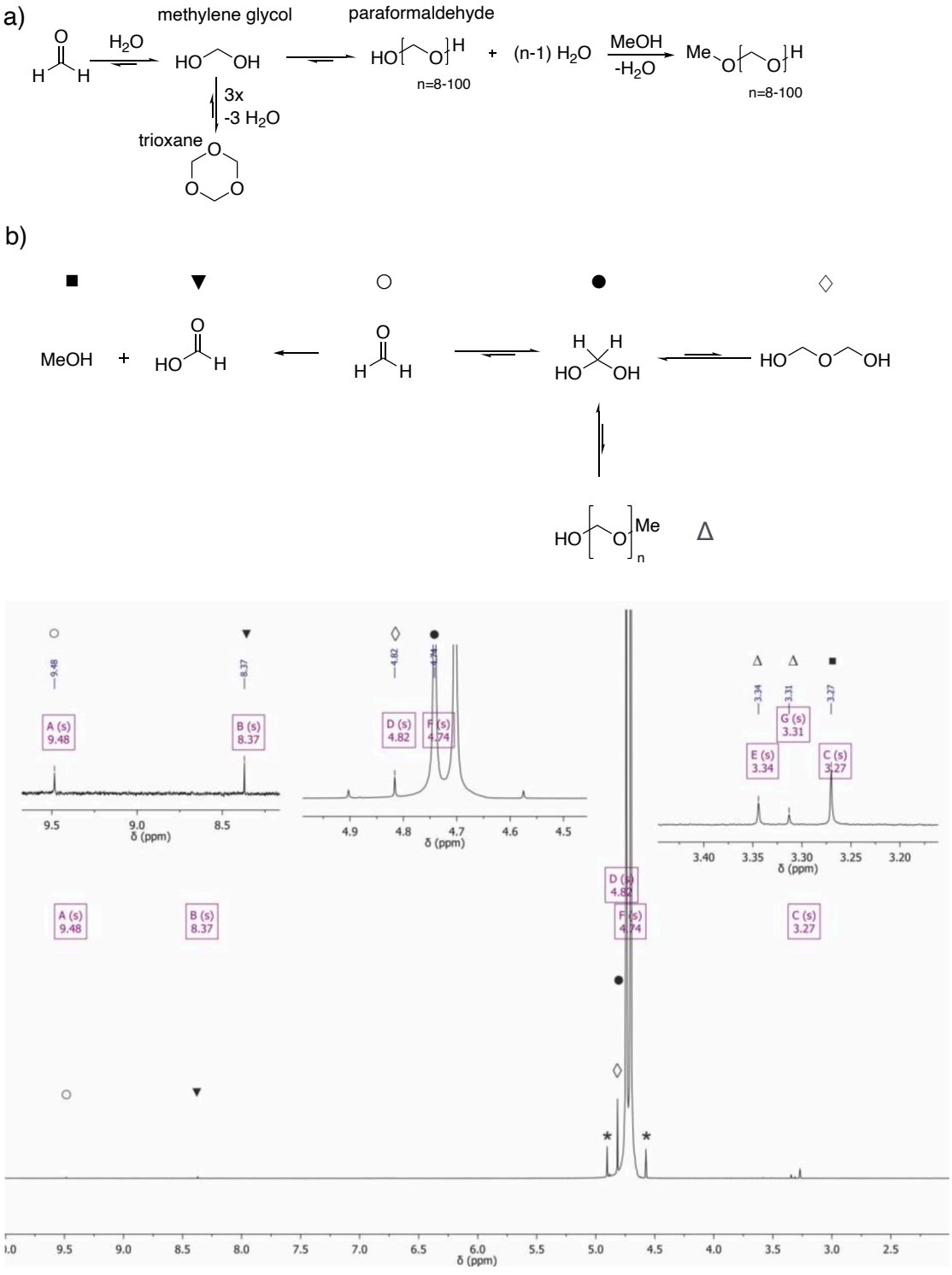


Figure 1.5 | HCHO in aqueous solution. **a)** HCHO reacts with water to form methylene glycol, which in turn forms the acetal 1,3,5-trioxane or the linear polymer paraformaldehyde. MeOH is added as stabiliser to suppress excess polymerisation, which would result in precipitation. **b)** Paraformaldehyde was dissolved in D_2O, equilibrated for 22 h, then analysed by ^{1}H NMR (700 MHz, provided by Dr Jos Kamps). The identified peaks are methanol (■, 0.2%), oligomeric forms of HCHO with methanol (Δ, <0.1%), hydrated HCHO (●, $HOCH_2OH$, 98.4%, carbon satellites are marked with asterisks), $HOCH_2OCH_2OH$ (◇, 1.3%), formic acid (▼, <0.1%), and non-hydrated HCHO (o, <0.1%). Percentages were estimated by integrating peaks to give their approximate relative abundances. Figure reproduced from [34] with permission from Dr Jos Kamps.

The Bürgi Dunitz angle describes the preferred angle of a nucleophilic reaction with a carbonyl group, whereby the π bond between the carbonyl oxygen and carbon is broken and a new σ bond between the nucleophile and carbon is formed in a concerted manner. Such requires molecular orbital overlap of the nucleophile's HOMO (Highest Occupied Molecular Orbital), usually a free electron pair, and the carbonyl group's LUMO (Lowest Unoccupied Molecular Orbital), i.e. the π^* of the carbonyl group. The p orbitals that form the carbonyl π^* are perpendicular to the carbonyl group, hence a nucleophilic attack at an angle of 90° to the carbonyl group would result in maximal orbital overlap (Figure 1.6 a). However, electrostatic repulsion of the nucleophile and the filled π orbital (Figure 1.6 b) results in preferred reaction at an angle of ∼107° (Figure 1.6 c).[35]

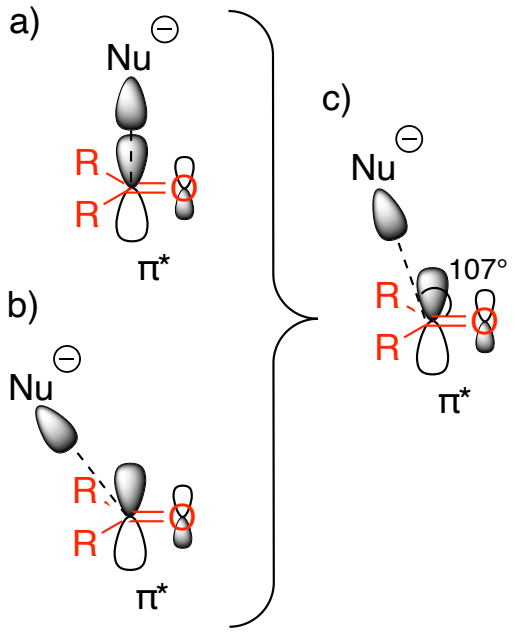


Figure 1.6 | Bürgi Dunitz angle for reaction of a nucleophile with a carbonyl compound. a) The angle for the overlap of the HOMO and LUMO is ∼90°. **b)** Repulsive coulombic forces from the filled π orbital hinder orthogonal reaction. **c)** A combination of both described effects results in a preferred reaction angle of ∼107°.

Using ^{18}O labelled H_2O, it has been demonstrated that H_2O quickly adds to ketones and aldehydes; however, the resulting hydrate, a geminal-diol, can usually not be isolated. Exceptions to this rule are HCHO, where the missing +I effect of an alkyl group and the less bulky hydrogen atoms favor addition of H_2O (as described above, see Figure 1.5). Furthermore, the presence of strongly electron-withdrawing groups adjacent to the carbonyl group (Figure 1.7 a+b), formation of 5- or 6-membered hemiacetals

(Figure 1.7 c), and ring strain (Figure 1.7 d) can destabilize the carbonyl group and promote formation of stable geminal-diols.[36] In the case of cyclopropanone and its hydrate, both the carbonyl carbon and the geminal-diol carbon exhibit bond angles of 60°; the least strained conformation of an sp^2 carbon would be 120°, and the least strained conformation of an sp^3 carbon would be 109.5°, hence the strain in the trigonal centre is relieved by hydrate formation.[37]

One prominent example that illustrates the mechanism of action of cyclopropanes *in vivo* is the natural product coprine which is present in the edible mushroom *Coprinus atramentarius*; coprine is hydrolysed *in vivo*, but not *in vitro* to glutamic acid and 1-amino-cyclopropanol; elimination of ammonia yields cyclopropanone, which rapidly forms the hydrate, which in turn reacts with the active site cysteine of aldehyde dehydrogenase (ALDH, Figure 1.7 e).[38, 39] Hence, concomitant ingestion of *Coprinus atramentarius* and ethanol can result in poisoning (Coprinus syndrome), with symptoms including facial flush, similar to the asian flush syndrome, which are due to an abnormal increase in acetaldehyde concentrations.[40]

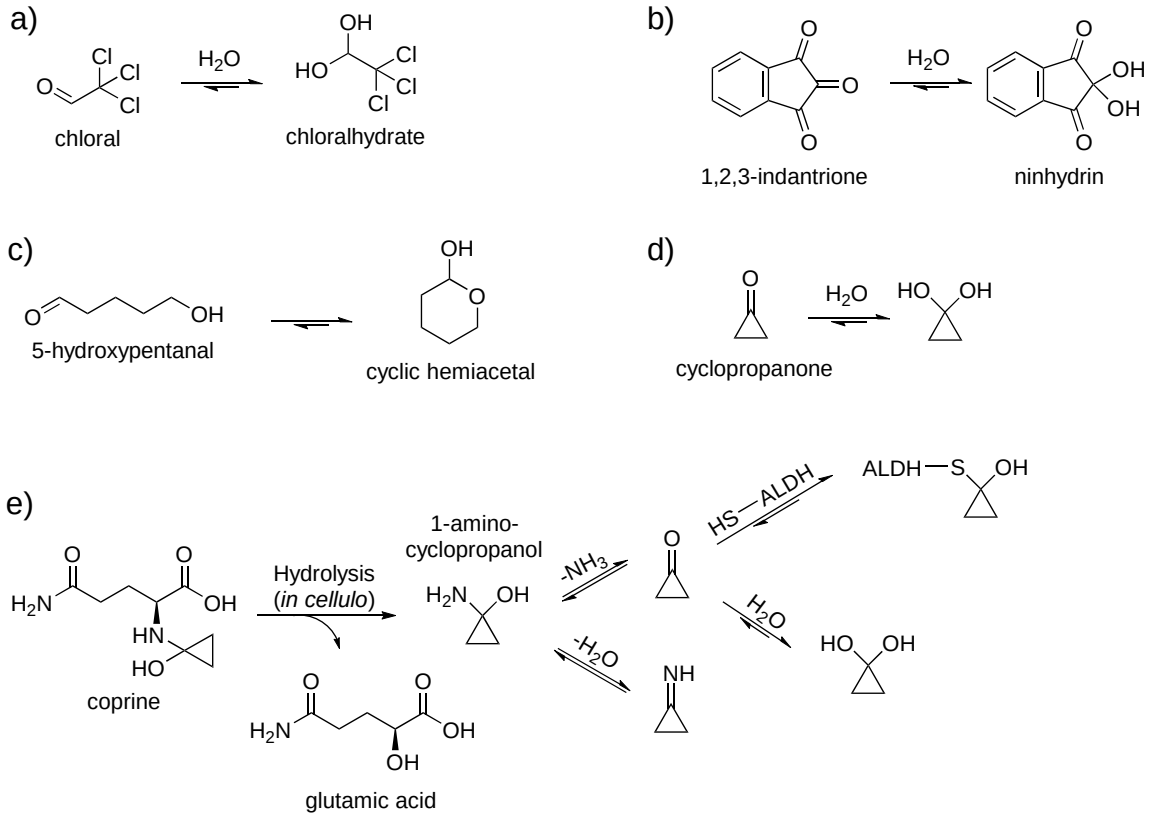


Figure 1.7 | Stable carbonyl hydrates and hemiacetals. a) Hydration facilitated by inductive charge repulsion. **b)** Hydration favored by dipole repulsion. **c)** The cyclic hemiacetal is the favored constitution, as two C-O σ bonds are energetically more favored than one C=O bond (as further exemplified by the case of some sugars). **d)** Formation of the sp^3 centre upon hydration releases 10.5° of ring strain relative to the sp^2 centre. **e)** Mechanism of action of coprine as found in *Coprinus atramentarius*: Coprine is hydrolysed *in cellulo*, releasing 1-amino-cyclopropanol, which forms cyclopropanone upon ammonia release. Cyclropropanone forms the hydrate and covalently reacts with the active site cysteine of ALDH.

As described above, HCHO is a crucial chemical feedstock for numerous industrial processes. Furthermore, HCHO is an important research chemical. Its electrophilic character has been harnessed for multiple name reactions in organic chemistry since the early 20th century (e.g. the Mannich and Eschweiler reactions),[41, 42] some of which are described here. The Cannizzaro reaction is a redox disproportionation, which is conducted in concentrated base. It is used to prepare carboxylic acids and alcohols from non-enolizable aldehydes (Figure 1.8 a).[43] The Eschweiler-Clarke reaction is used for methylation of primary and secondary amines. It employs HCHO as methylating agent and formic acid as hydride donor (Figure 1.8 b).[41] The Blanc reaction is used for the preparation of chloromethylated arenes, analogous to the Friedel-Crafts alkyation. In the Blanc reaction, the

Lewis acid $ZnCl_2$ facilitates an electrophilic aromatic substitution, and the intermediate benzyl alcohol is converted to the chlorinated product (Figure 1.8 c).[44] The Mannich reaction involves an α-CH acidic compound such as carbonyl compounds, HCHO, and a secondary/primary amine or ammonia. The amine first reacts with HCHO to form an iminium ion. Keto-enol tautomerism of the carbonyl compound forms the enol, which attacks the iminium ion, forming a β-aminocarbonyl compound (Mannich base, Figure 1.8 d).[42] Urotropin (hexamethylenetetramine) has found a wide range of applications in organic synthesis, materials and medicinal chemistry, and industrially in phenolic resins and solid fuel tablets.[45] Urotropin is produced by reaction of HCHO with NH_3,[46] forming a cage-like structure similar to adamantane. One example employing urotropin is the Duff reaction, which uses urotropin and acid to convert phenols or aromatic amines to form the formylated arene (Figure 1.8 e).[47]

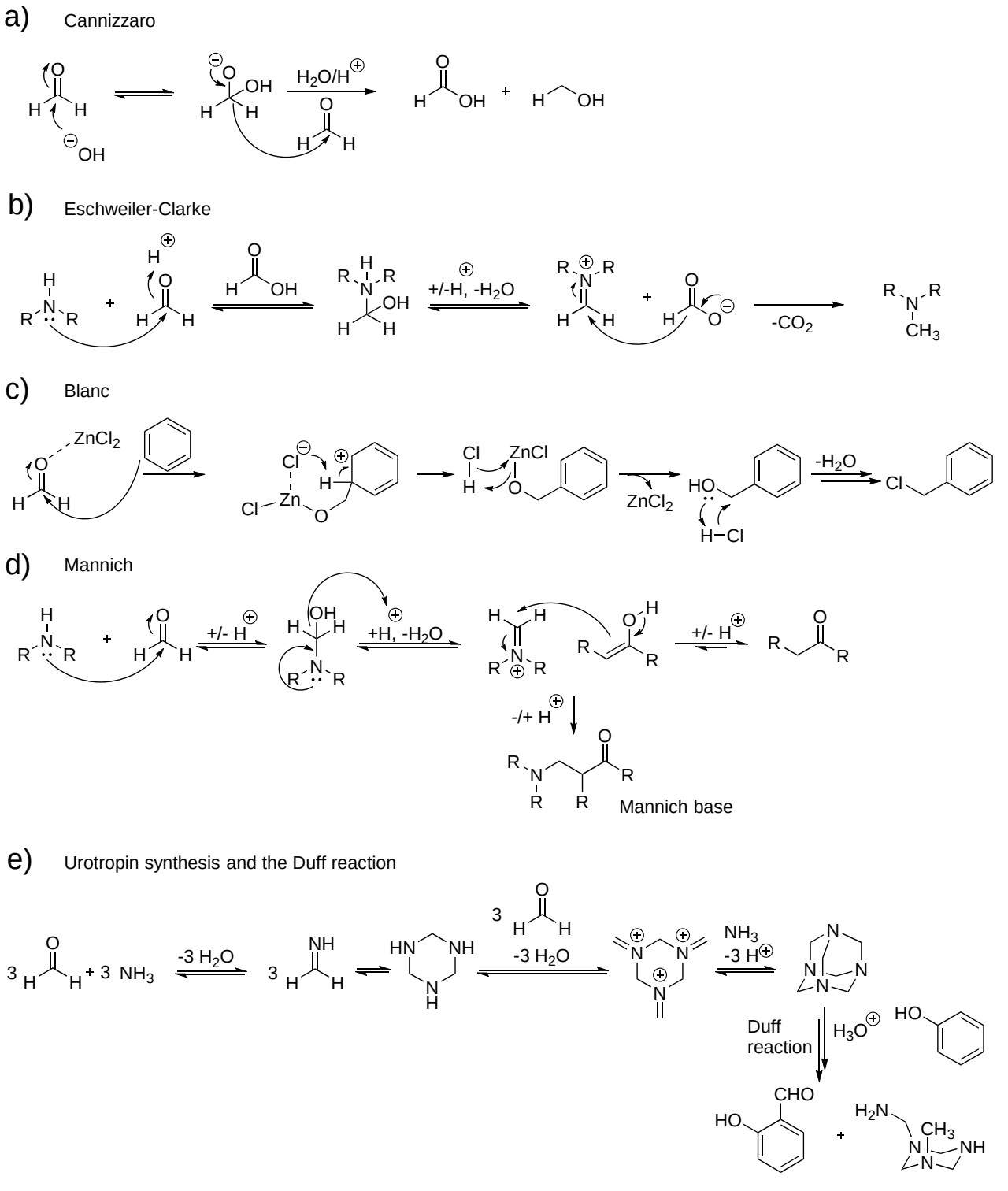


Figure 1.8 | Reactions involving HCHO. a) The Cannizzaro reaction employs two aldehydes (here exemplified with HCHO), which disproportionate in presence of concentrated base to form the corresponding alcohol and acid. **b)** Eschweiler-Clarke reaction for methylation of amines, employing a hydride transfer from formate. **c)** The Blanc reaction uses HCHO and $ZnCl_2$ to yield chloromethylated arenes. **d)** The Mannich reaction condenses HCHO with a secondary amine to form an intermediate which reacts with an enolizable carbonyl compound to form a Mannich base. **e)** Urotropin is formed by the reaction of HCHO with NH_3. Urotropin can further be used for formylations of phenol (Duff reaction).

1.1.5 Endogenous/Exogenous HCHO Exposure and its Use as Preservative and in Pharmaceuticals

Gas chromatography/mass spectrometry measurements suggest that the adult human body contains ~1.75 mg/kg HCHO,[48] corresponding to 122.5 mg for a 70 kg adult; the plasma half-life for HCHO is estimated to be around 1.5 min, which means that 41 mg/min of HCHO (122.5 mg/(2x1.5 min)) are metabolised and generated to maintain an equilibrium, wich results in around 59 g daily HCHO turnover (41 mg/min x 60 min/h x 24 h).[49, 50] A GC-MS-based method employs pentafluorobenzylhydroxylamine which reacts with HCHO to form an oxime ether, which in turn can be detected by GC-MS.[51] This GC-MS-based method has been used in a recent study to determine serum HCHO levels in mice.[52] Note, however, that it is unclear how much of this HCHO is free (as hydrate or aldehyde form) or how much is in a reacted form bound to biomolecules.

Numerous incidents have been reported where HCHO was accidentally administered to patients during surgery.[53–55] It is estimated that oral exposure to HCHO from food products accounts for about 100 mg HCHO daily, which is negligible given the calculated estimated daily endogenous HCHO turnover of 59 g.[56] Exogenous exposure of HCHO, including from food and the atmosphere likely only accounts for 1-2% of its endogenous turnover.[57] HCHO is a product of the incomplete combustion of organic compounds and is part of wood derived smoke; smoked meat is conserved by both the bactericidal effects of a HCHO layer and the antioxidative effects of phenolic substances.[33] Despite the fact that systemic exposure to HCHO below threshold levels is likely non-toxic (at least in an acute sense), sensitisation upon exposure to cosmetics containing HCHO or HCHO releasers is reported.[58–60] Cosmetics that are water based have been formulated with HCHO itself as a preservative, but modern formulations contain HCHO releasers because of the increased risk to sensitisation by HCHO itself (Figure 1.9 a).[58, 59, 61]

Prodrugs are designed to increase lipophilicity/solubility and permeability of the parent drug to enhance drug delivery. The oxymethyl group (Figure 1.9 b) is a common moiety used to provide these desired characteristics and is bioconverted by esterases or alkaline phosphatases. The enzymatic dephosporylation of Cerebyx® yields a phosphate group and the hemiaminal, which rapidly eliminates

HCHO to form the parent drug Phenytoin (Figure 1.9 c).[57] Secondary sulfonamides have been developed as increasingly useful fragments for parent drug candidates due to their bioisosteric properties with respect to carboxylic acid and amide groups. However, as such compounds often have poor physicochemical properties, the use of tertiary *N*-acyloxymethyl-sulfonamides as prodrugs is a promising way to maximise effectiveness of the parent secondary sulfonamide drug (Figure 1.9 b+d). *N*-acyloxymethyl-sulfonamide prodrugs are susceptible to spontaneous hydrolysis in an S_N1-type reaction, forming an *N*-sulfonyliminium ion which reacts with water to form the hemiaminal, which eventually releases HCHO to form the parent secondary sulfonamide drug (Figure 1.9 d).[62, 63]

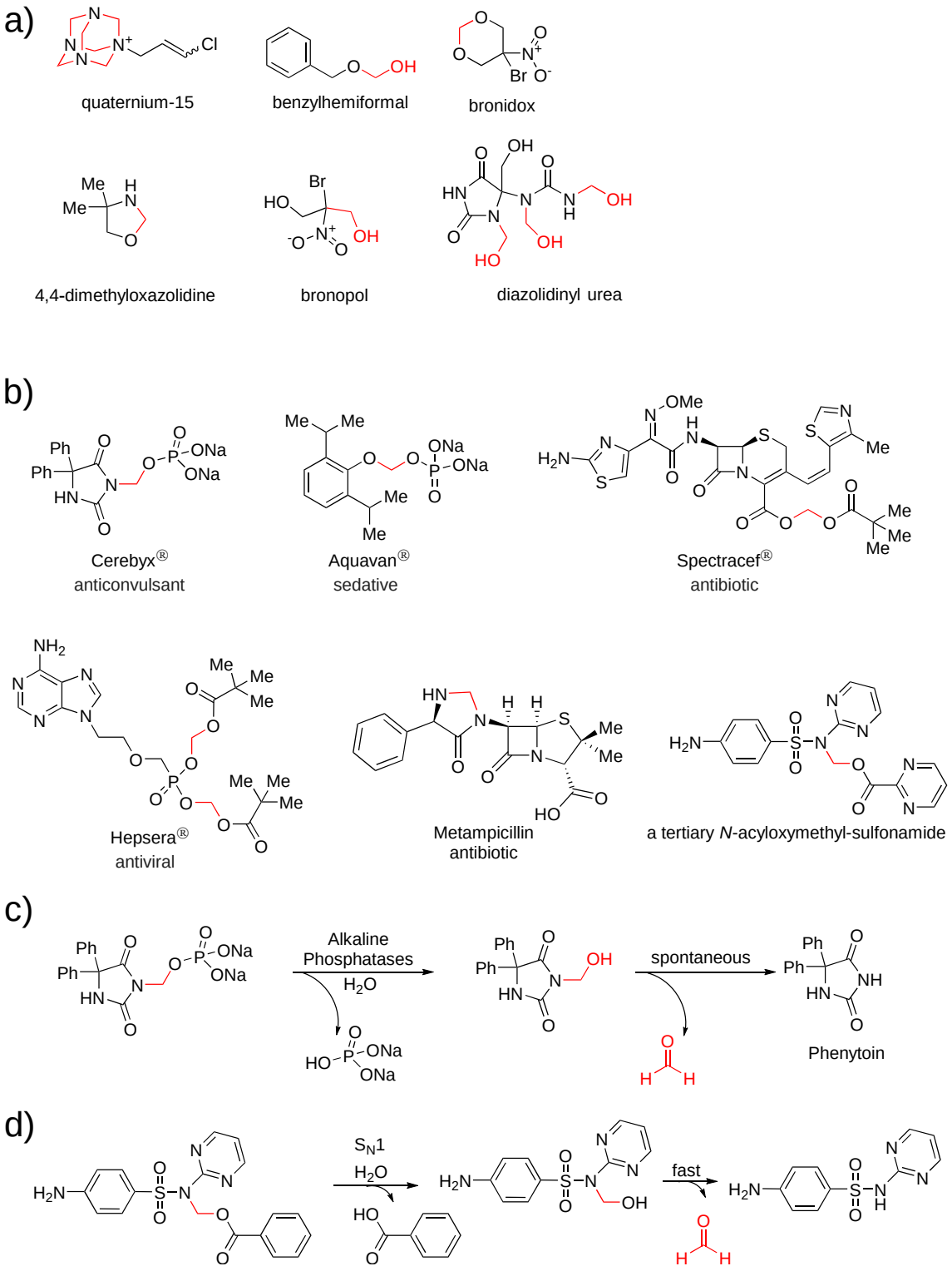


Figure 1.9 | HCHO in cosmetics and prodrugs. a) HCHO releasers permitted in cosmetics in the European Union.[64] **b)** Exemplary HCHO releasing prodrugs with oxymethyl spacers or HCHO-derived cyclic aminal moieties (metampicillin).[57, 63] Note the structure of metampicillin was reassigned as described in Chapter 4. **c)** Bioconversion of Cerebyx® by alkaline phosphatases to form HCHO and phenytoin.[57] **d)** Mechanism of HCHO release exemplified by hydrolysis of a tertiary *N*-acyloxymethyl-sulfonamide.[62]

1.2 HCHO in Biology

1.2.1 Glutathione and Cysteine Reactions with HCHO

Exposure to exogenous HCHO is toxic above threshold levels, as illustrated by the HCHO related incidents described in Chapter 1.1.1. *In vitro* studies from the 1920s onwards have shown that HCHO reacts with amino acids and proteins[65–79] and causes intra-strand cross-links in DNA.[80] To what extent this happens in cells, is unclear. It is likely that these reactions underpin the toxicity of HCHO in cells. Notably, it is proposed that endogenous HCHO production is sufficient to cause fatal malignancies in mice that are deficient of the cellular HCHO detoxification and repair system (*Adh5*$^{-/-}$ and *Fancd2*$^{-/-}$), details of which are provided in Chapter 1.3.[81] Recent studies have investigated the reactions of HCHO with biomolecules such as amino acids,[82] nucleic acids,[83], peptides,[40, 84] and cross-links between DNA and amino acids[85] employing modern NMR techniques, which are detailed in the following.

Among the proteinogenic amino acids, cysteine has been reported to undergo the most efficient reaction with HCHO, to first form a hemithioacetal, which is assigned as kinetic product. This then reacts to form the thermodynamic product thioproline (also known as timonacic, thiazolidine-4-carboxylic acid or thiaproline)[86] (Figure 1.10 a); no intermediate has been observed in the conversion of the hemithioacetal to thioproline.

Among the tested amino acids, thioproline is the most stable and biologically relevant identified molecule resulting from the reaction with HCHO. Addition of an excess of the HCHO scavenger 1,3-cyclohexanedione (1,3-CHD) did not result in re-formation of cysteine throughout the analysis time of 16 h, suggesting the thiazolidine is stable. Addition of H^{13}CHO resulted in slow incorporation (>48 h) of H^{13}CHO into thioproline, implying a slow dynamic equilibrium of thioproline and cysteine (Figure 1.10 a).

Reacting cysteine with other biologically relevant carbonyl compounds (e.g. acetaldehyde, acetone, glyoxal or methylglyoxal) produces analogous thiazolidines, which are, however, rapidly displaced to form thioproline if reacted with excess HCHO.[82] However, when thioproline is treated with an excess of these aldehydes and ketones, no displacement was observed in aqueous conditions (the pH was not

adjusted in these reactions), suggesting the HCHO-cysteine-derived thiazolidine is the most stable (Figure 1.10 b).[82]

It is generally accepted that a major mechanism for the intracellular detoxification of HCHO is via its reaction with glutathione (GSH) to form *S*-hydroxymethylglutathione (HMG); the reaction occurs via its cysteinyl-thiol, further highlighting the importance of thiols in HCHO relevant biology; apart from the well established *S*-hydroxymethyl species, a bicyclic 2:1 HCHO:GSH adduct was discovered, named $BiGF_2$ using co-crystallization with CBR1, a reductase containing an endogenous GSH binding site.[87] More detailed NMR characterisation revealed further monocyclic structures (MGF and PGF, Figure 1.10 c). However, HMG is favored over the less reactive cyclic structures, and the latter might only have biological relevance when HCHO levels are elevated and HMG clearance is dysfunctional.[84] It is likely that HMG is detoxified to *S*-formylglutathione before cyclisation can occur. Other HCHO-derived moieties like hemiaminals on N-terminal prolines are to the best of our knowledge not prone to enzymatic detoxification, hence cyclisations with adjacent nucleophiles like amide bonds are possible.

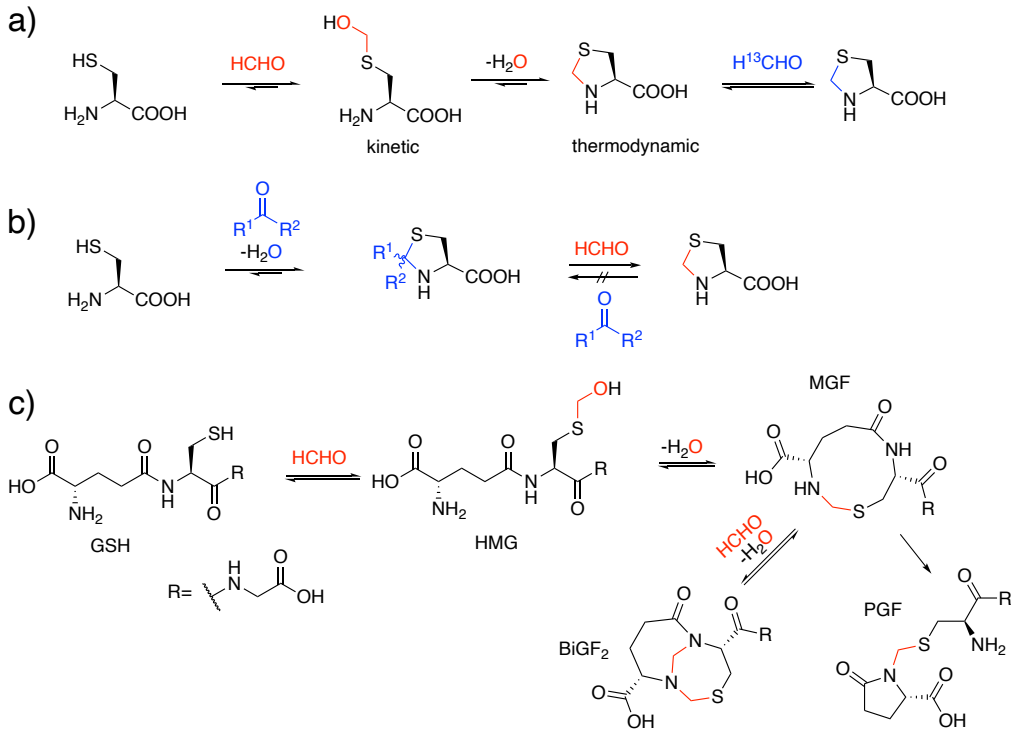


Figure 1.10 | Reaction of glutathione and cysteine with HCHO. a) Upon reaction with HCHO, cysteine rapidly forms a hemithioacetal, which cyclises to form a thiazolidine, which is in slow dynamic equilibrium with cysteine, although no loss of HCHO-derived groups upon treatment with 1,3-CHD was observed.[82] **b)** Incorporation of other biologically relevant carbonyl compounds into cysteine to form the analogous thiazolodine was tested for acetaldehyde (R^1=H, R^2=Me), acetone (R^1, R^2=Me), glyoxal (R^1=H, R^2=C(O)H) or methylglyoxal (R^1=H, R^2=C(O)Me). These thiazolidines were displaced by addition of excess HCHO to form the HCHO-derived thiazolidine. If thioproline was reacted with excess acetaldehyde, acetone, glyoxal or methylglyoxal, no displacement was observed.[82] **c)** Glutathione (GSH) forms HMG upon reaction with HCHO, which was observed *in vitro* to form MGF and PGF, and upon another addition of HCHO, $BiGF_2$.[84, 87]

1.2.2 How HCHO Reacts with Amino Acids

HCHO reacts with nucleophiles to form three characterised types of general adducts (Figure 1.11): (i) methylol (hydroxymethyl) groups which undergo condensation to form (ii) Schiff bases (a sub-class of imines, $R_1R_2C{=}NR'$; ($R' \neq H$)), and (iii) methylene bridges if two appropriately positioned neighboring nucleophiles are present; the methylol adducts exchange their hydroxy group with the oxygen atom in water through the equilibrium with the Schiff base.[88, 89]

Studies on the reactions of single amino acids with excess HCHO in water have revealed complexity even with these seemingly simple building blocks.[82] Histidine rapidly forms the the cyclised N^{π}- and N^{α}-bridged compound, along with hydroxymethylation of the imidazole ring and Pictet-Spengler type *C*-N^{α}-cyclisation (Figure 1.11 a). The secondary alcohol of threonine forms an oxazolidine upon reaction with HCHO, which is only observed in presence of base in case of the primary alcohol of serine (Figure 1.11 b), which may be due to the Thorpe-Ingold effect. Tryptophan undergoes hydroxymethylation of the indole ring nitrogen and *C*-N^{α}-cyclisation under acidic conditions, analogous to histidine (Figure 1.11 c). Upon reaction with HCHO, asparagine and glutamine react to form 6- and 7-membered lactams, respectively (Figure 1.11 d+e); additionally, only asparagine was observed to form *N*-hydroxymethylated lactams (Figure 1.11 d). Reaction with HCHO and arginine led to a bicyclic adduct containing two HCHO-derived units (Figure 1.11 f).

Among the amino acids with aliphatic side chains, no reaction (at least to give a discrete product) between HCHO and the N^{α} amino group was observed under the conditions used in this study, i.e. a 10-fold excess HCHO. However, taking the nucleophilicity of the N^{α} groups into account, these reactions likely occur and are potentially not observable by NMR, even though the resulting hemiaminals or imines are likely intermediates *en route* to the cyclised products as observed with non-aliphatic amino acids (Figure 1.11 a-f). Indeed, these (transient) N^{α}-reaction products can be reduced to form stable methyl groups using $NaCNBH_3$; treating alanine with HCHO and $NaCNBH_3$ resulted in formation of *N*-mono- and *N*-di-methylated products (Figure 1.11 g).

Interestingly, thioproline was the only HCHO-derived amino acid reaction product that did not reform

the parent amino acid upon addition of 40-fold excess the HCHO scavenger 1,3-CHD (relative to HCHO) within 3 days. Notably, under the conditions used in that study, reactions with lysine and HCHO did not result in substantial amounts of the hydroxymethyl-HCHO adduct on the N^{ϵ}-amine. The reaction products of cysteine and histidine with HCHO have been confirmed in human cells and mice that were treated with HCHO/HCHO precursors, highlighting the potential for these two amino acids as alternative routes for HCHO metabolism.[86] Given the fast and reversible reaction of HCHO with cysteine to form thioproline, thioproline has the potential to act as a reservoir for HCHO.[86] Overall, the work to date suggests the potential for enormous complexity in the reactions of HCHO with biologically relevant nucleophiles. Such complexity is also possible with respect to other biologically relevant electrophiles, e.g. acetaldehyde and α-oxoaldehydes. However, it should be noted that biology may have evolved to limit the reactivity of such functional groups, as exemplified by the predominantly cyclic nature of many biologically relevant sugars, e.g. glucose. α-oxoaldehydes like methylglyoxal and glyoxal are more electrophilic due to the oxo-group α to the carbonyl group. In contrast, the carboxylic acid α to the carbonyl group in glyoxylic acid makes the carbonyl group in pyruvate less electrophilic. The bulkiness of the alkyl groups in ketones like acetone lower their reactivity; acetaldehyde, having only one alkyl group, is more reactive than ketones. However, HCHO, having no bulky alkyl groups, is the most reactive biologically relevant aldehyde.

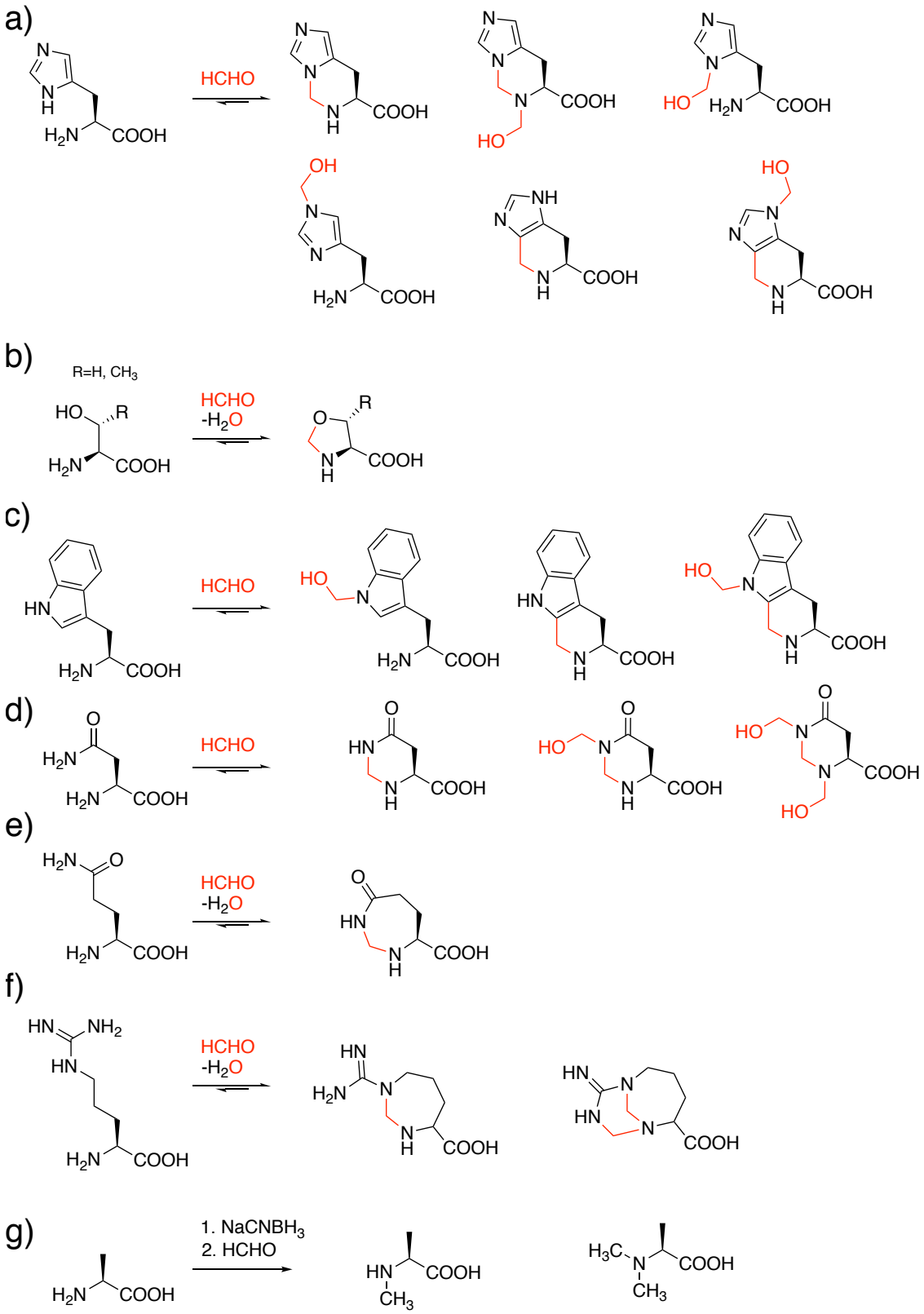


Figure 1.11 | How HCHO reacts with amino acids. **a)** Products of the reactions of HCHO with histidine. **b)** Products of the reactions of HCHO with hydroxyl group containing amino acids threonine and serine. **c)** Products of the reactions of HCHO with tryptophan. **d+e)** Products of the reaction of HCHO with sidechain amide-containing asparagine (d) and glutamine (e). **f)** Products of the reaction of HCHO with arginine. **g)** Products of the reductive methylation of alanine.[82]

1.2.3 How HCHO Reacts with Nucleobases

In vivo, the carcinogenic properties of HCHO are proposed to be principally due to its reactions with DNA, giving rise to bone marrow failure, kidney and liver dysfunction, as well as leukemia if cellular HCHO detoxification mechanisms are inoperative.[81] Hence, elucidating the cellular nature of the HCHO-DNA adducts that underpin HCHO-driven toxicity may be a key to understanding HCHO-derived carcinogenicity. Pioneering work by McGhee and von Hippel found hemiaminal adducts resulting from reaction of the canonical nucleobases with HCHO and revealed the potential of HCHO adducts to interfere with Watson-Crick hydrogen bonding. Hydroxymethylated adenine can form hydrogen bonds with thymine, but in a less stable manner than unmodified adenine. A hydroxymethyl as found on the exocyclic amine of adenine or cytosine group can exist in two isomeric forms, i.e. in case of adenine, syn to N_1/anti to N_7, therefore blocking canonical Watson-Crick base-pairing positions, or anti to N_1/syn to N_7. It was observed that the rotational isomer which disrupts Watson-Crick hydrogen bonding (i.e. syn to N_1/anti to N_7) is the preferred one (Figure 1.12 a).[90, 91] The DNA-HCHO adducts have been more recently characterised by modern NMR techniques and showed to form hemiaminal adducts on the endo- (in case of thymidine) and exocyclic nitrogens (adenosine, guanosine, cytidine) of the individual nucleobases. The preferred formation of the endocyclic HCHO adduct on thymidine and analogously, uridine and inosine bases, may imply that their hydrogen bonding capacity is directly altered by reaction with HCHO. In this study, no HCHO-derived methylene cross-links were assigned, but it is conceivable that conformationally constrained nucleobases in cellular DNA form cross-linked adducts.[83]

The methylol groups as observed by reaction of nucleobases with HCHO can also be formed by oxidation of a methyl group to give a hemiaminal, as occurs during catalysis by at least one demethylase acting on histones and nucleic acids (Figure 1.12 b+c). (3-Methyl)thymidine (3mT) and (6-methyl)adenosine (m6A) are oxidised by the human nucleic acid demethylase fat mass and obesity-associated protein (FTO) to form the *N*-hydroxymethyl adducts, which degrade to form the free base and HCHO. While (6-hydroxymethyl)adenosine only slowly fragments to the free base, (3-hydroxymethyl)thymidine is very

fast to form thymidine, suggesting base-specific accumulation of *N*-hydroxymethyl adducts in nucleic acids at least *in vitro*; similarly, lysine demethylases (KDMs) demethylate N^{ϵ}-methylated lysine residues that first form unstable methylol adducts which in turn rapidly fragment to form the demethylated lysine and HCHO (Figure 1.12 c).[83] HCHO is routinely used for stabilising native protein–DNA interactions in cells by cross-linking DNA with proteins (e.g. chromatin immunoprecipitation, ChIP) [92] or for stabilisation and preservation of tissue architecture and cell morphology for immunofluorescence.[93]

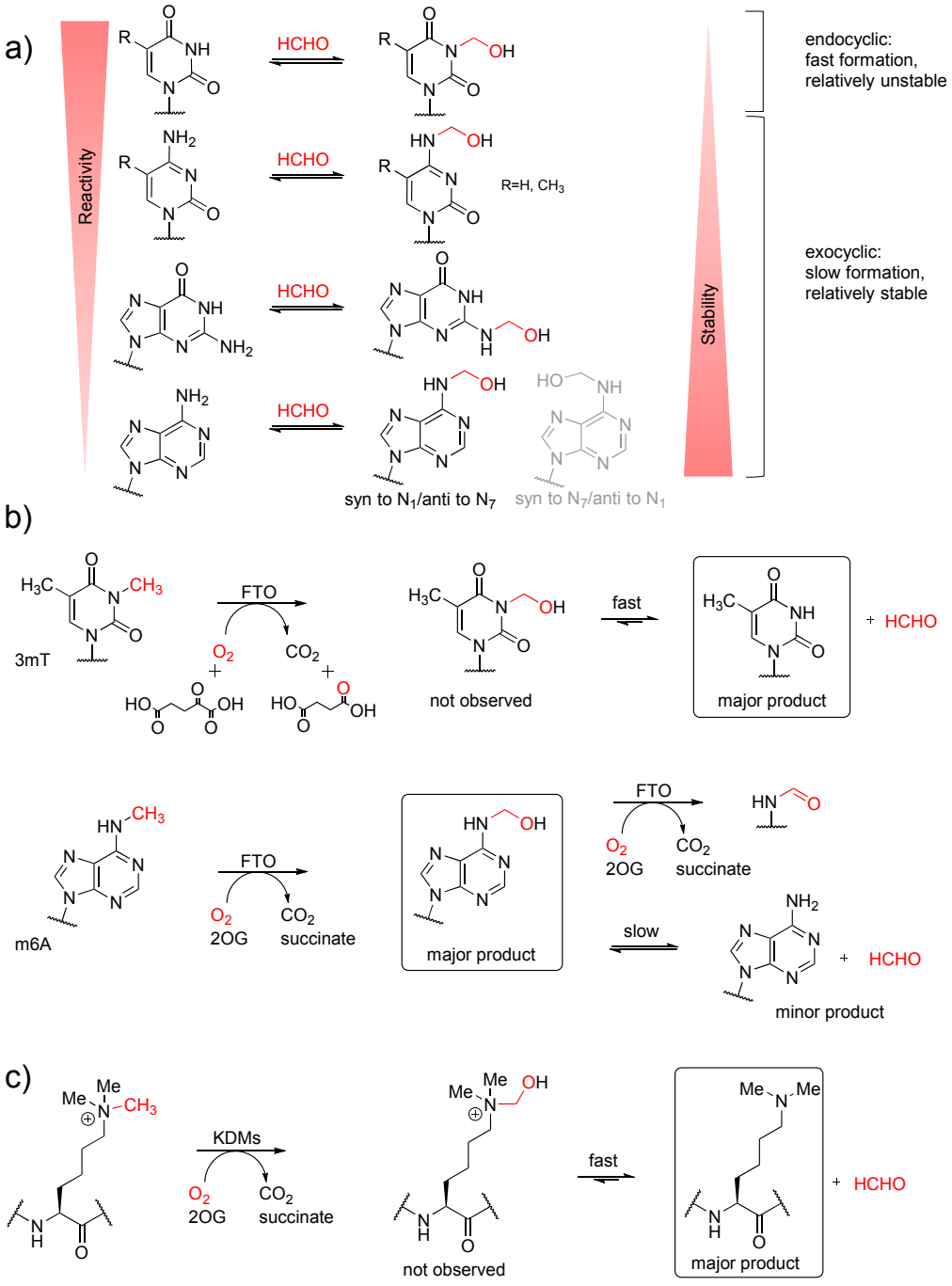


Figure 1.12 | HCHO-derived DNA adducts. a) Stability and reactivity of nucleotides reacting with HCHO. The grey rotational isomer of (6-hydroxymethyl)deoxyadenosine monophosphate is not formed due to steric hindrance.[83] **b)** *N*-Methylated nucleobases were oxidised with FTO to form the *N*-hydroxymethyl adducts. In the case of (3-hydroxymethyl)thymidine the nascent product degrades fast and in case of (6-hydroxymethyl)adenosine slowly to form the free base and HCHO.[83] Note there is evidence that the major FTO product can be oxidised to the *N*-formyl product.[94] **c)** KDM-catalysed demethylation of N^{ϵ}-methylated lysine to form an unstable hemiaminal, which rapidly degrades to form demethylated lysine and HCHO.[83]

1.2.4 HCHO as Disinfectant

The preserving properties of HCHO were first discovered by Ferdinand Blum in 1893. Blum realized that his fingers dried and hardened upon experimenting with formalin, and soon successfully fixed (i.e. preserved from decay resulting from putrefaction or autolysis) a mouse in a 10% HCHO solution; within a few years, several medical schools across Europe used HCHO for preservation purposes; HCHO has insecticidal, fungicidal and bactericidal properties, hence it prevents the growth of putrefactive organisms; by 1910 HCHO has replaced heavy metal salts like arsenic that were previously used for human cadaver embalming completely; nowadays, over 100 years later, little has changed in embalming techniques using HCHO.[95]

The travelling exhibition *Body Worlds* by Gunther von Hagens allows the public to experience human anatomy showing common diseases, ailments and medical advances, intending to enhance healthy living and promote health awareness; the exhibition has reached around 40 million visitors in 45 cities around the world; von Hagens uses his patented plastination technique, the first step of which is injecting HCHO into arteries to preserve the bodies to be exhibited.[96] In hatcheries, HCHO fumigation is used to disinfect hatching eggs from microbial contamination; however, if executed improperly, HCHO can damage the dormant embryo. The two leading methods for HCHO fumigation are heating a suspension of paraformaldehyde and water or addition of formalin to potassium permanganate (2:1, v/w). The latter process produces heat by oxidation of HCHO, which is sufficient to vaporize the water containing the remaining HCHO.[97, 98]

1.2.5 HCHO in Medicine

The first evidence of HCHO inactivation of toxins to render its atoxic product (toxoid) was accrued a century ago in 1921; these early studies demonstrated that formalinised diphtheria toxin protein injection results in immunity of guinea-pigs against diphtheria.[99] It was recognised that it is crucial to destroy the lethality of a toxoid without changing significantly its antigenic value. Even in these early studies, it was already appreciated that the amount of HCHO relative to toxin required to result in the desired immunogenic effect is dependent on the amino-nitrogen content of the toxin; the work

of Glenny et al. in 1921 is one of the first reports pointing out that for vaccinations it is desirable to understand the detailed chemistry underlying toxin antigen inactivation by HCHO.[100]

Today HCHO treatment remains a key method for chemical inactivation of pathogens; in addition to toxins, HCHO has been used to inactivate viruses like the poliovirus[101] and bacteria like *Bordetella pertussis*.[102]

Novel inactivated whole virus vaccines against enterovirus 71,[103] coxsackievirus[104] or coronaviruses such as SARS-CoV-1[105] have been treated with HCHO, underlining further potential even with future vaccine development. However, the HCHO-inactivated respiratory syncytial virus (RSV) vaccine has increased the severity of the virus, enhancing pathology to subsequent infection.[106] It was suggested that the increase in severity was caused by using suboptimal concentrations of HCHO, implicating that HCHO-inactivated RSV might not recapitulate the antigenic epitopes of infectious RSV.[107, 108] Overall, these observations suggest that it is crucial to understand the detailed biology and chemistry of the mechanism involved upon HCHO treatment.

Further outbreaks in France and the United Kingdom of the foot-and-mouth disease virus (FMDV) in 1981 and Venezuelan equine encephalitis pandemic in Central America in 1969-1972 have arisen due to improper HCHO inactivation of the vaccines, which as a result carried residual infectivity.[109] One possible explanation for these incidents is that it is crucial when inactivating viruses that their RNA or DNA are rendered non-infectious; it was found that HCHO may cross-link the RNA of FMDV to its capsid proteins, which might leave some viral RNA infectious.[110]

The use of HCHO itself has been criticised amongst anti-vaccination protestors due to its general toxicity/carcinogenicity.[111] Indeed, the manufacturing process of formalinised vaccines results in residual free HCHO in the finished vaccine (0.00008–0.02%), but it is estimated that endogenous HCHO is around 100-fold higher than vaccine-introduced exogenous HCHO, hence residual HCHO in vaccines does not represent an unsafe constituent in vaccines.[112] Further studies on the mechanisms of HCHO inactivation on poliovirus (PV) vaccines conclude that HCHO treatment leads to a decreased propensity of the virus to bind to the human receptors, reduces the ability of PV to convert to the necessary intermediate particle 135S required for cell entry, and renders PV RNA non-infectious.[101]

Whilst progress has been made to elucidate HCHO reaction products on proteins and peptides, the complexity and heterogeneity of these adducts gives rise to difficulties regarding batch-to-batch consistency in vaccine products.[88] Notably, for cross-links to occur in peptides or proteins, the reactive moieties need to be in close proximity (distances of α-carbon atoms 3.8-10.9 Å) and there is conflicting evidence whether or not solvent accessibility plays a role for conversion of reactive residues.[113, 114] The reported HCHO-induced peptide or protein modifications are summarised in Figure 1.13. 4-imidazolidinone formation upon reaction of HCHO with the N-terminal amine and the adjacent peptide bond nitrogen has been reported (Figure 1.13 a), as well as cross-links between the guanidinium group of arginine and the primary amine of lysine (Figure 1.13 b); no intermolecular cross-links between peptides were found in this study, neither were reactions between two primary amines to form a methylene bridge (Figure 1.13 c); reaction conditions are used in this study are based on the inactivation of diphtheria toxin for vaccine production, i.e. 50-fold excess HCHO, pH 7.2 and 35 °C for 48 h.[89] Cross-linking reactions between lysine and the ortho-position of tyrosine were also reported (Figure 1.13 d), as well cross-linking of lysine and tryptophan (Figure 1.13 e).[114] Lysine residue pairs in peptides reacted with HCHO to form formylated/methylated and α,β-unsaturated aldehyde/methylated products (Figure 1.13 f), the mechanisms of which are given in Figure 1.14, again highlighting the potential for great complexity of the reactions of HCHO with biomolecules.[88]

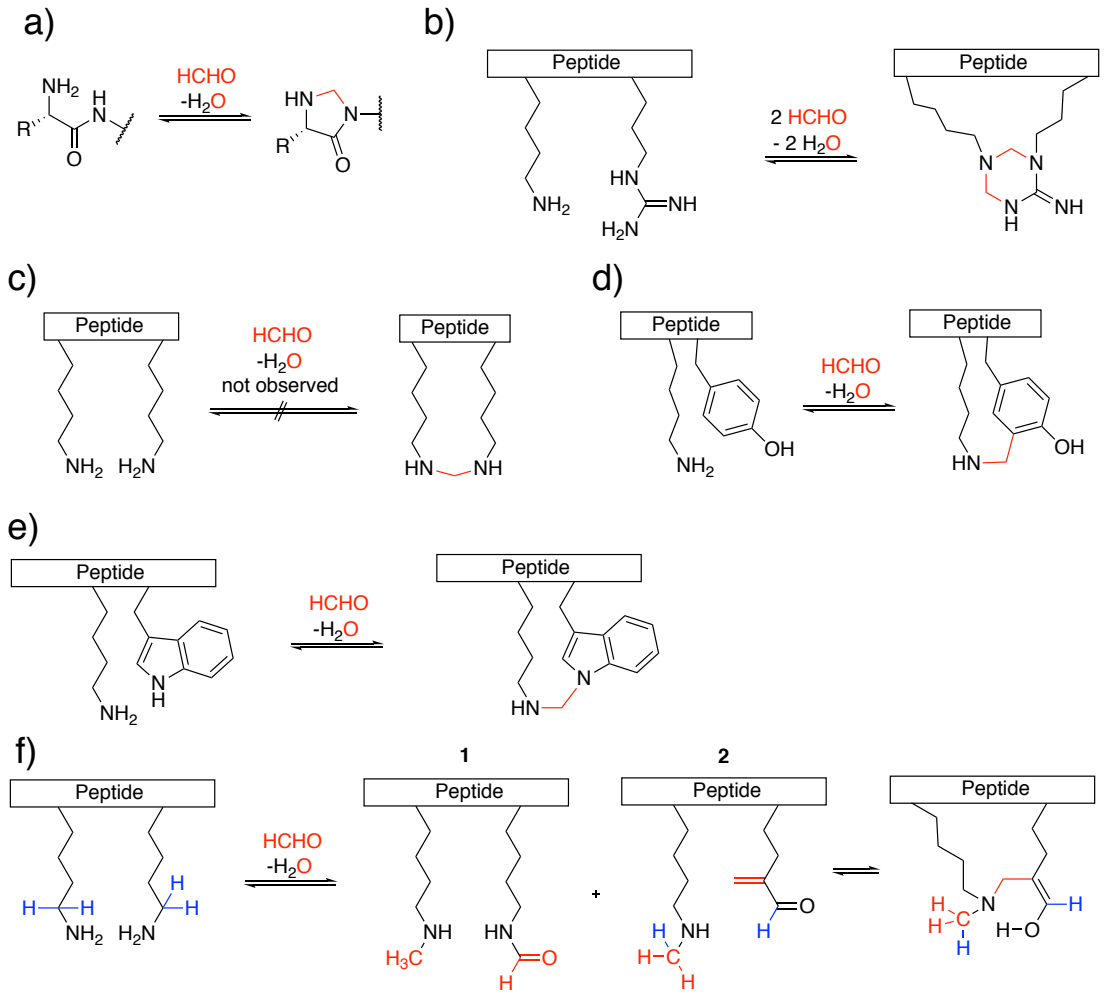


Figure 1.13 | Reported HCHO-derived peptide adducts. a) N-terminal amines form 4-imidazolidinones upon reaction with HCHO.[89] **b)** The guanidinium group of arginine forms two methylene bridges with the ϵ-amino group of lysine.[89] **c)** Cross-linking between two lysine amino groups was not observed.[89] **d)** The ϵ-amino group of lysine cross-links tyrosine on the ortho-position at its phenol ring in an S_NAr-type reaction.[114] **e)** The indol nitrogen of tryptophan cross-links with the amino group of lysine.[114] **f)** Summary of products found on Ac-GDVEKGAK-OH upon treatment with HCHO. Mechanisms for the formation of structures **1** and **2** are given in Figure 1.14.[88]

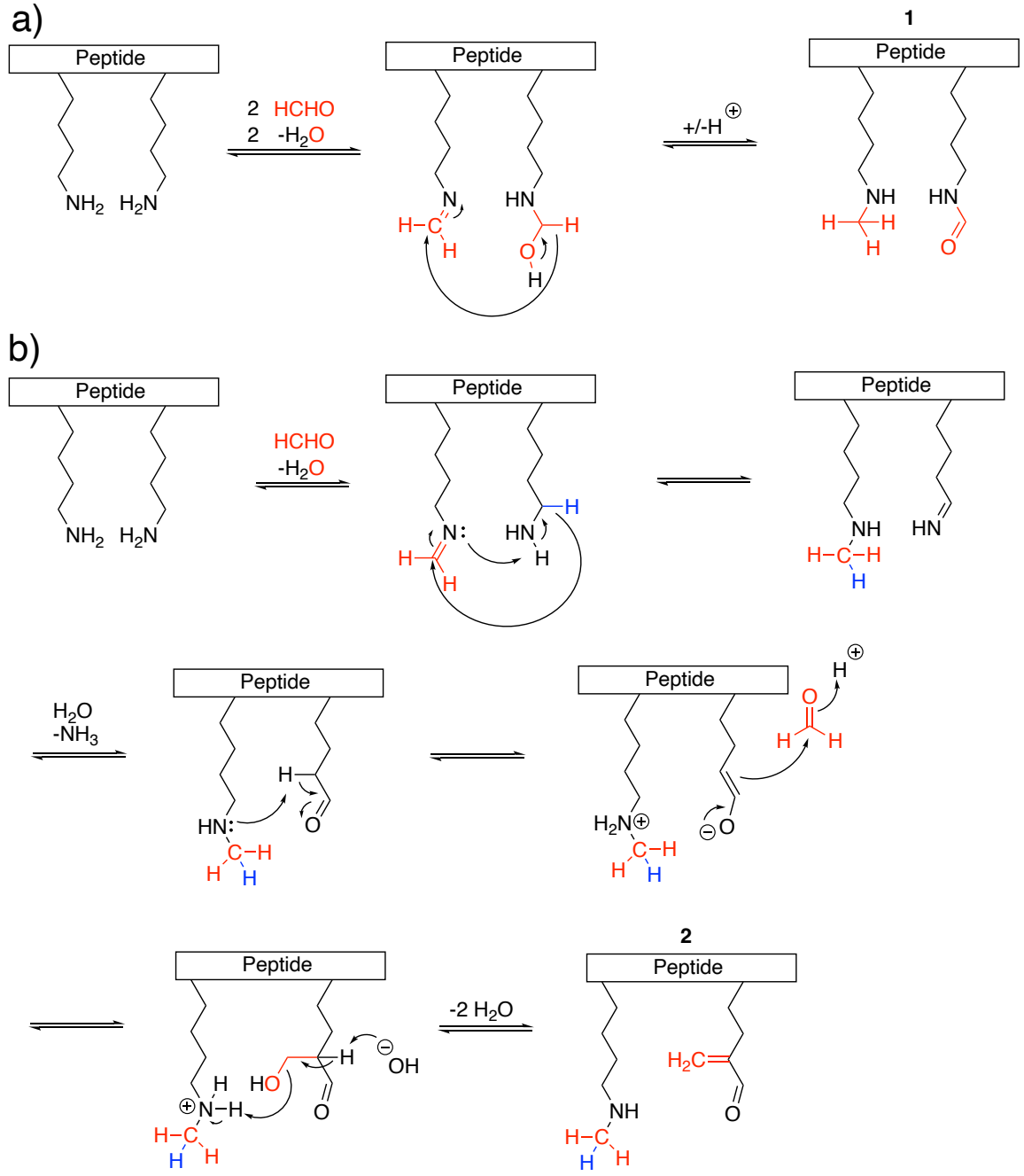


Figure 1.14 | Proposed mechanism for formation of structures 1 and 2 shown in Figure 1.13. **a)** Upon reaction of two HCHO molecules with two adjacent ϵ-amino groups of lysine in the peptide Ac-GDVEKGAK-OH, the methylol group transfers a hydride ion to the imine, resulting in formation of **1**. The imine nitrogen is proposed to function as base, deprotonating the methylol hydrogen. **b)** Upon reaction of only one molecule of HCHO with the ϵ-amino group of lysine, imine formation on the non-reacted lysine leads to a concerted hydride transfer to the HCHO-derived imine, forming the *N*-methylated adduct. Subsequent hydrolysis of the resulting imine leads to formation of an aldehyde, which tautomerises to form an enolate, which in turn reacts with another HCHO molecule. This adduct undergoes a β-elimination to form **2**.[88] Related hydride shifts upon reaction of HCHO with the amino acid ornithine have been reported elsewhere.[82]

1.2.6 Enzymatic Reactions that Release HCHO

In addition to exogenous sources of HCHO, HCHO is produced endogenously as a by-product of various biochemical pathways. Hence, cells are likely to be constantly exposed to endogenously exposed HCHO. One of the most dramatically elevated levels of HCHO with up to 500 µM (being observed by comparing the amplitude of FANCD2 foci by immunofluorescence in differentiating cells versus HCHO-treated cells; the role of FANCD2 is discussed in Chapter 1.3) is found in blood progenitor cells during transcriptional reprogramming, possibly as a result of nucleosome demethylation events.[115] This and other observations also suggest that there may be a gradient of HCHO, with very high local concentrations of HCHO present on some occasions in the nucleus.[83]

As described before (Figure 1.12 b), FTO is an RNA demethylase acting both in the nucleus and cytoplasm resulting in HCHO formation.[116] In eukaryotes, additional nuclear HCHO formation arises from histone lysine demethylases (KDMs), which catalyse the demethylation of methylated lysine residues. KDMs are divided into two main classes, defined by their catalytic mechanism: the FAD-dependent amino oxidase (AO), and the Fe(II) and 2-oxoglutarate-dependent hydroxylases (JmjC domain-containing proteins). Both classes form the unstable hydroxymethylated intermediate, followed by elimination of HCHO (JmjC catalysis is shown in Figure 1.15 a).[117] The detailed mechanism of JmjC catalysis involves the coordination of the Fe(II) centre to normally three protein-derived ligands; following bidentate 2OG binding to the active site Fe(II), binding of the substrate promotes oxygen binding. Oxidative decarboxylation of 2OG results in succinate and carbon dioxide production, and in formation of a reactive Fe(IV)=O (ferryl) intermediate. The latter abstracts a hydrogen atom from the substrate, forming a substrate radical and an Fe(III)-OH species; subsequently, the substrate radical and a hydroxyl radical derived from the Fe(III)-OH complex form the hydroxylated product, completing the catalytic cycle upon product dissociation (Figure 1.15 b).[118]

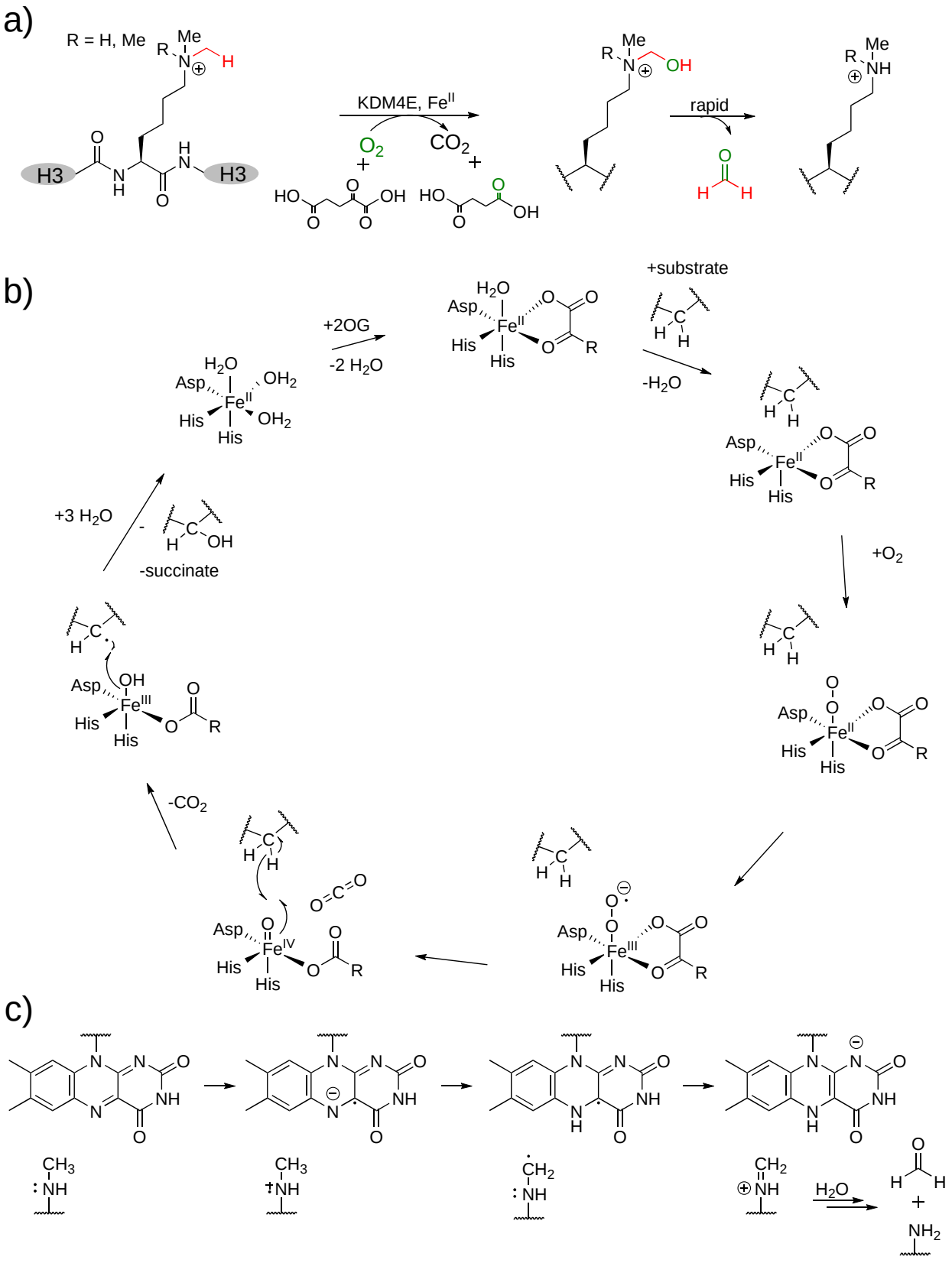


Figure 1.15 | Outline mechanisms for KDM catalysis. a) Reaction scheme for an oxidative demethylation as catalysed by the JmjC oxygenase KDM4E, forming a hydroxymethyl intermediate which eliminates HCHO.[117] **b)** General catalytic mechanism for the Fe(II)/2OG oxygenases. R=CH_2CH_2COOH.[118] **c)** Reaction scheme for a demethylation as catalysed by the FAD-dependent amino oxidase LSD1.[119]

Another class of non-redox reactions releasing HCHO involves serine hydroxymethyltransferases (SHMT), which reversibly convert serine to glycine, forming HCHO as a by-product; first, an ϵ-amino group of SHMT reacts with pyridoxal 5'-phosphate (PLP) to form a Schiff base (internal aldimine). Serine reacts with this species, releasing the lysine amino group to form another Schiff base (external aldimine). Upon HCHO release, a quinonoid intermediate is formed which is stabilised by resonance with the π-electron system of PLP. The free amine reacts with the Schiff base to regenerate the internal aldimine, releasing glycine (Figure 1.16 a).[120] DMGO (dimethylglycine oxidase) catalyses the conversion of dimethylglycine to the mono-demethylated product sarcosine using flavin adenine dinucleotide (FAD) as a cofactor, releasing HCHO as a byproduct which is coupled to the formation of 5,10-methylenetetrahydrofolate (5,10-CH_2-THF, Figure 1.16 b).[121] If enzymatic HCHO release is not directly coupled to THF, it is likely that HCHO is scavenged by cellular nucleophiles like glutathione and cysteine. SSAO (semicarbazide-sensitive amine oxidase) catalyses the deamination of primary amines using a quinonoid cofactor, the regeneration of which releases ammonia and hydrogen peroxide; first, the cofactor reacts with the amine to form a Schiff base (intermediate 1, Figure 1.16 c). Proton abstraction and concomitant reduction of the cofactor results in a Schiff base intermediate (intermediate 2, Figure 1.16 c), which is hydrolysed to form HCHO and an aminoquinol form. To regenerate the cofactor, two electrons are transferred from the aminoquinol to O_2, forming an iminoquinone and H_2O_2, which is hydrolysed to release NH_4^+, regenerating the cofactor (Figure 1.16 c).[122, 123] Notably, elevated SSAO-catalysed deamination of methylamine in transgenic mice has been shown to result in increased formation of HCHO-protein adducts and pulmonary inflammation; the inflammation process may be a result of the HCHO-derived reactions.[124]

The cytochrome P450 family is involved in the demethylation of various plant secondary metabolites via the hydroxymethylated adduct, eventually releasing HCHO and the demethylated product; examples of these reactions include nicotine demethylation to form nornicotine,[125] tropine demethylation to form nortropine,[126] caffeine demethylation to form paraxanthine,[127] and theobromine demethylation to form 7-methylxanthine (Figure 1.16 d).[128] An average espresso contains ∼60 mg of caffeine and has the potential to form 0.31 mmol of HCHO upon absorption and demethylation in the liver.[129]

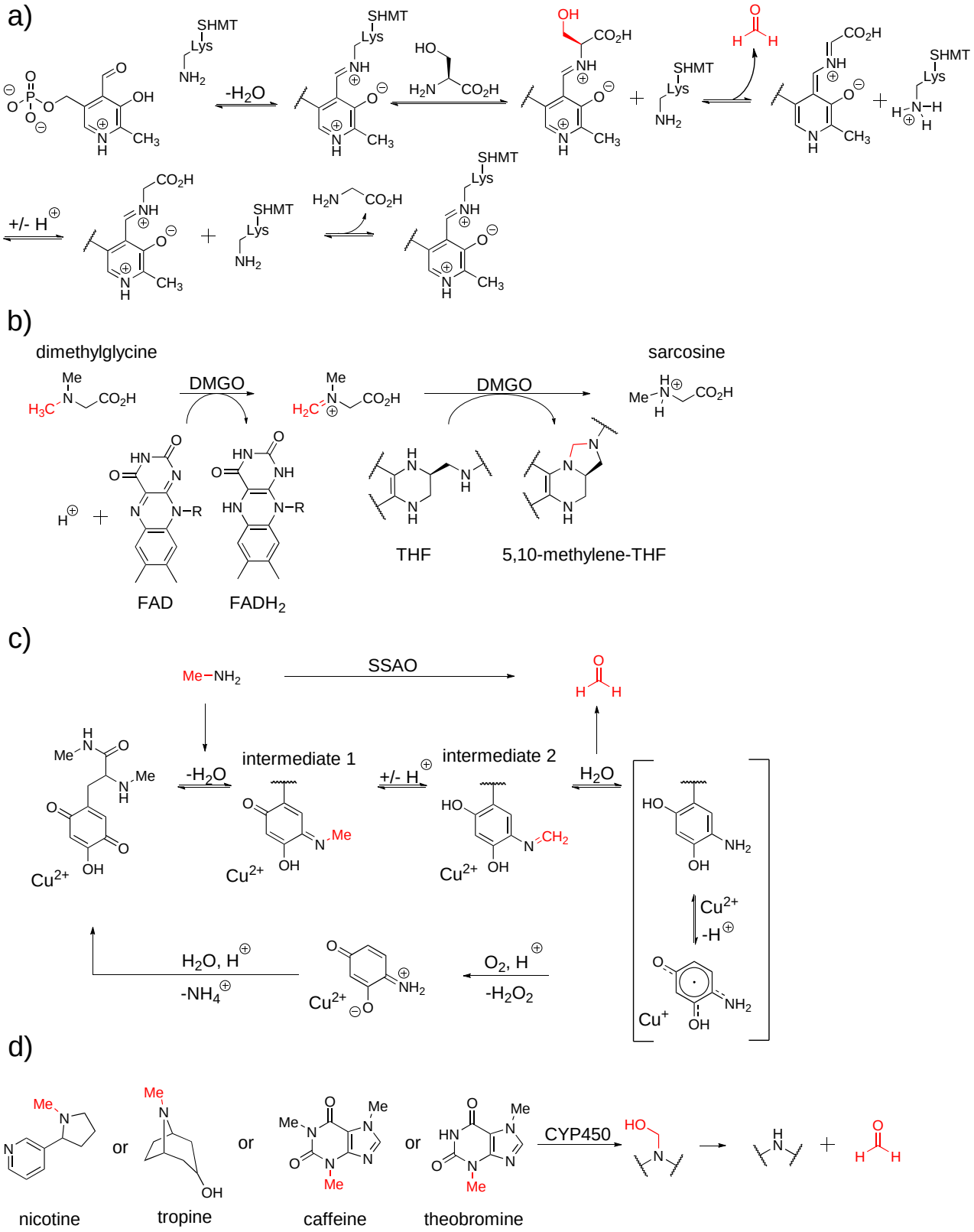


Figure 1.16 | Overview of some reported biochemical reactions with HCHO as by-product. Reactions involved in demethylation events include SHMT (a),[120] DMGO (b),[121] SSAO (c),[122, 123] and CYP450 (d).[125–128]

1.2.7 HCHO in Bacteria[130]

Due to the cytotoxic nature of HCHO, microorganisms have evolved several metabolic strategies to cope with the intracellular accumulation of this electrophile. Cellular metabolism of methylated compounds leads to HCHO as a byproduct.[121, 122, 125–128] HCHO readily forms *N*-hydroxymethyl adducts with the endo- and exocyclic nitrogens of DNA/RNA (Figure 1.12 a) or nucleophilic groups on proteins (Figures 1.11, 1.13, 1.14); such *N*-methylol adducts are susceptible to condensation to form an imine, which is, in turn, prone to react with another amine group to form a stable methylene bridge cross-link.[83, 131, 132] Similarly, thiol groups can react with HCHO to form hemithioacetals which can react with another amine group to form a thiazolidine (Figure 1.10 a). Protein-protein, DNA-DNA, and DNA-protein cross-links have been described.[89, 133–136] Concomitantly, DNA and protein damage resulting from HCHO exposure have been described.[137–139]

In bacteria, HCHO is detoxified via three different mechanisms, i.e. thiol-, ribulose monophosphate (RuMP)-, or pterin-dependent processes. Each of the mechanisms shares the principle that HCHO first reacts with a nucleophile to form a less reactive adduct. Of all the detoxification pathways, the thiol-dependent one is the best studied (Figure 1.17 a). The thiol-dependent HCHO detoxification pathway is shared by plants, animals, and some bacteria and employs the non-enzymatic reaction of HCHO with glutathione (GSH). In certain bacteria, e.g. *Paracoccus denitrificans*, this reaction is reported to be catalysed by the glutathione-dependent HCHO-activating enzyme Gfa.[140] GSH is a tripeptide with a γ peptide bond between the carboxyl group of the glutamate side chain and cysteine, followed by a glycine residue. The reaction of HCHO and GSH results in *S*-hydroxymethylglutathione (HMG), which is further oxidized by the NAD^+-dependent enzyme AdhC to give *S*-formylglutathione (SFG).[141] HMG can react further to cyclic structures, but their biological relevance is unclear (Figure 1.10 c).[84] SFG is subsequently hydrolysed by EstD to yield formate and GSH.[142] However, not all bacteria use glutathione in the thiol-dependent pathway; HCHO can also be detoxified using bacillithiol or mycothiol (Figure 1.17 b).[143, 144]

In the pterin-dependent pathway (Figure 1.17 c), the two adjacent reactive secondary amines in

tetrahydrofolate (THF) react with HCHO to form a methylene bridge between N5 and N10, forming 5,10-CH_2-THF. Subsequently, 5,10-CH_2-THF is oxidised by the $NADP^+$-dependent dehydrogenase MtdA to form 5,10-methenyl-THF. The cyclohydrolase Fch then hydrolyses the latter to yield 10-formyl-THF. Ultimately, the THF ligase Fhs hydrolyses 10-formyl-THF to form THF and formate.[145, 146] However, several methylotrophic proteobacteria and methanogenic archaea employ the analogous tetrahydromethanopterin (THMP) in lieu of THF.[145, 147]

The RuMP pathway employs a different HCHO detoxification strategy via reaction of HCHO with sugar phosphates. The main difference between the RuMP pathway, the thiol-dependent and the pterin-dependent pathways is that the RuMP pathway follows an assimilation of the HCHO adduct into carbon metabolism, whereas the thiol-dependent and the pterin-dependent pathways detoxify the adduct, resulting in the formation of formate. The RuMP-pathway employs a hexulose-6-phosphate synthase (Hps) catalysed condensation reaction of HCHO with ribulose-5-phosphate, yielding *D*-arabino-3-hexulose- 6-phosphate; subsequent isomerization by 6-phospho-3-hexuloisomerase (Phi) results in fructose-6-phosphate, which is used for glycolytic pathways (Figure 1.17 d).[148]

However, several organisms including *Pseudomonas aeruginosa*, *Pseudomonas putida*, and *Burkholderia fungorum* detoxify HCHO independent of pterin, RuMP or thiols; instead, they employ a one-step mechanism using NAD^+ and a HCHO dehydrogenase (EC 1.2.1.46), yielding formate and NADH (Figure 1.17 e).[149–151]

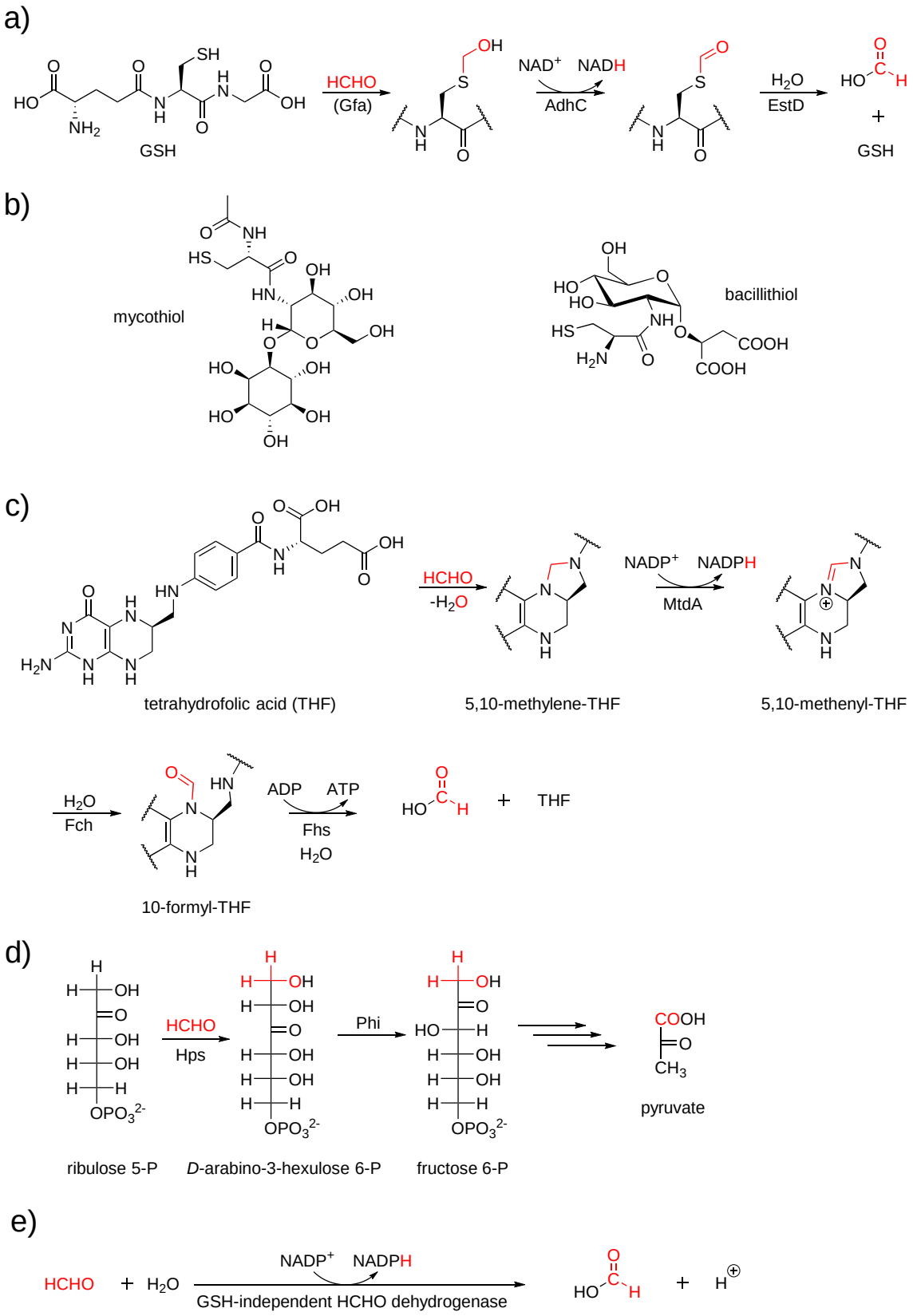


Figure 1.17 | Bacterial pathways for HCHO detoxification. **a)** Thiol-dependent pathway for *Neisseria meningitidis*, illustrated with glutathione (GSH). **b)** The structures of alternative thiol-compounds mycothiol and bacillithiol are given. **c)** Pterin-dependent pathway, illustrated by tetrahydrofolate (THF). **d)** Ribulose monophosphate (RuMP) pathway, showing assimilation of HCHO into 1C metabolism. **e)** NAD^+-dependent HCHO dehydrogenase pathway. This figure was adapted from [130].

1.2.8 HCHO Sensing in Bacteria[152, 153]

HCHO sensing in bacteria is not fully understood, but to date three different families of regulators have been described. All three pathways follow the same principle: HCHO detoxification is achieved by HCHO-responsive transcription factors which are responsible for downstream HCHO metabolism. However, all three transcription factors have conserved cysteines in common, the mutation of which results in loss of the HCHO detoxification response. One of the most well-studied detoxification systems is the *frmRAB* operon in *E. coli*, providing first insights into the exact mechanism of action of the HCHO-responsive transcription factors and how biological systems may sense HCHO. FrmR is a transcriptional repressor that senses HCHO by forming a HCHO-derived inter-subunit methylene bridge between Cys35 and Pro2 (Figure 1.18 a). This modification dramatically alters the surface charge of FrmR, resulting in dissociation of FrmR from the promoter of *frm*AB. The dissociation of FrmR results in the biosynthesis of the enzymes FrmA and FrmB, which are part of the GSH-dependent pathway for HCHO detoxification in *E. coli*.

The FrmA- and FrmB-dependent detoxification of HCHO in *E. coli* is analogous to the pathway described in Figure 1.17 a for *Neisseria meningitidis*: FrmA is a glutathione-dependent HCHO dehydrogenase, forming *S*-formyl glutathione from *S*-(hydroxymethyl) glutathione. FrmB is a *S*-formylglutathione hydrolase, hydrolising *S*-formyl glutathione to form glutathione (GSH) and formate (Figure 1.18 b). Another example of HCHO-responsive transcription factors is HxlR in *Bacillus subtilis*, which binds to the *hxlAB* operon upon HCHO-induced cross-linking of Cys11 and Lys13, resulting in expression of Hps and Phi (see Figure 1.17 d);[154, 155] the HxlR transcription factor coupled to a fluorescent protein has been harnessed to generate a HCHO sensor allowing temporal-spatial visualization of HCHO in mouse brain and mammalian cells.[155]

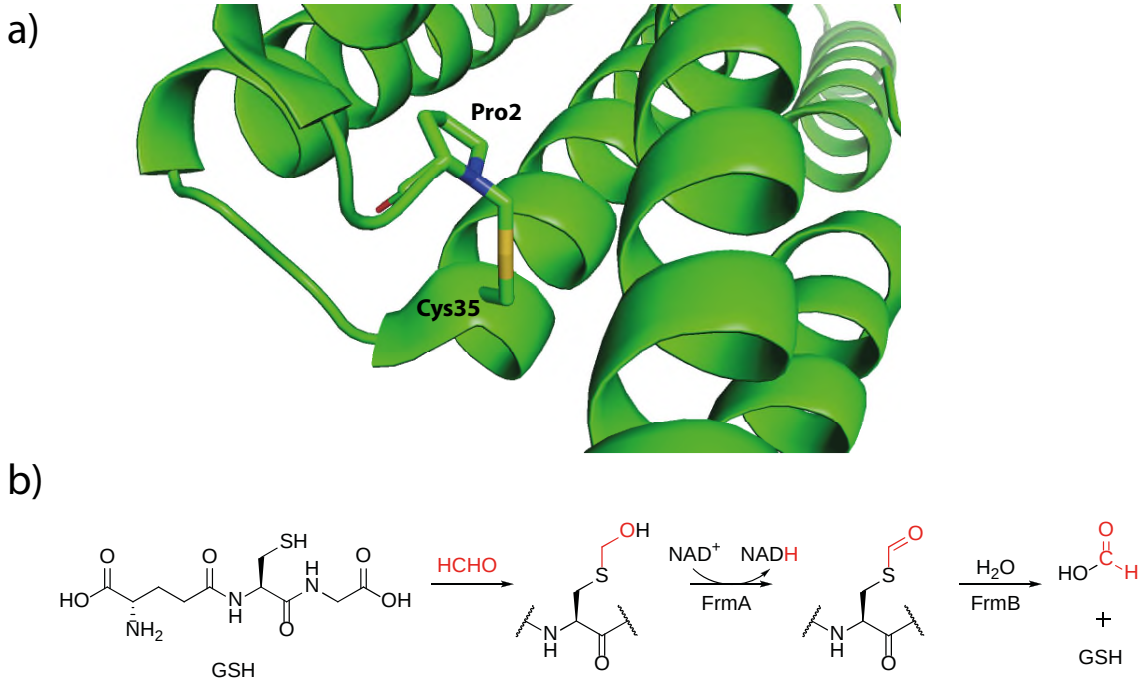


Figure 1.18 | The *frmRAB* operon in *E. coli*. a) View from a crystal structure of HCHO-induced FrmR (PDB 5LBM). The transcription factor FrmR reacts with HCHO to form a methylene bridge between Cys35 and Pro2 (shown as sticks), so promoting transcription of *frmAB*. **b)** GSH reacts with HCHO to form *S*-formylglutathione, which is converted to *S*-formyl glutathione by FrmA. FrmB then hydrolyses *S*-formyl glutathione, regenerating GSH and forming formic acid.[152]

1.3 Aldehydes and Ketones in Mammals

1.3.1 HCHO and Acetaldehyde

In humans, HCHO concentration in blood ranges from 20-100 μM.[48, 156, 157] These high values suggest that HCHO is produced continually, and one of the constant sources of HCHO is 1C metabolism; the mitochondrial folate pathway employs interconversion of serine to glycine via serine hydroxymethyltransferase 1 and 2 (SHMT1 and SHMT2), forming HCHO,[158] which in turn reacts with THF to form 5,10-methylene-THF. THF and 5,10-methylene-THF are key metabolites and their concentrations can be as high as 20 μM in rats.[159] Notably, 5,10-methylene-THF has been shown to degrade *in vitro*,[160] and along with THF and dihydrofolate (DHF) is prone to oxidative decomposition to release HCHO *in cellulo* (Figure 1.19 a); apart from HCHO release, oxidative degradation of THF results in release of its other two components, a pteridine moiety and p-aminobenzoylgluamate (ABG,

Figure 1.19 b).[161] It is proposed that 5,10-methylene-THF is more stable to degradation than THF, indicating a non-enzymatic feedback mechanism: High HCHO levels lead to 5,10-methylene-THF formation resulting from reaction of THF and HCHO, which both depletes the cellular HCHO pool and stabilises 5,10-methylene-THF; however, in low HCHO conditions, unreacted THF degrades to release HCHO which can be incorporated into 1C metabolism.[146]

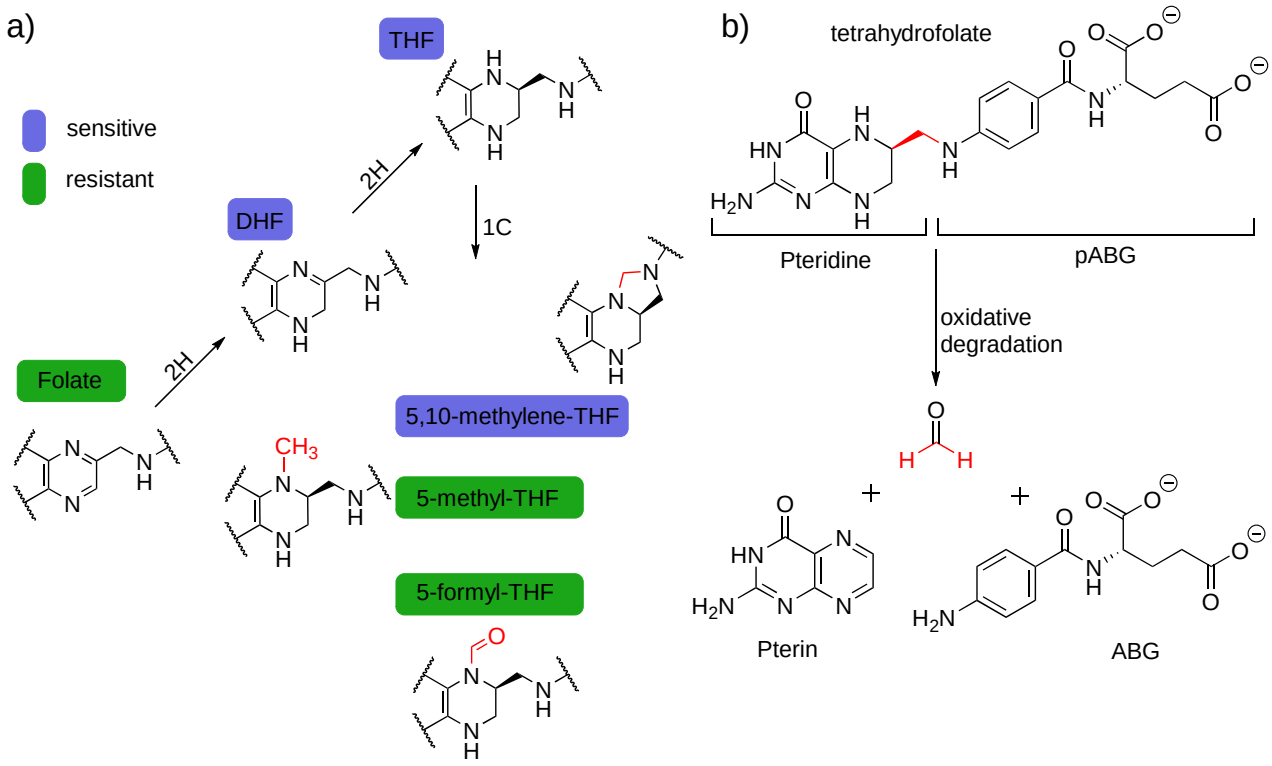


Figure 1.19 | Oxidation of THF and its derivatives. a) DHF, THF and 5,10-methylene-THF are most prone to oxidation (blue), and hence HCHO release, whilst folate, 5-methyl-THF and 5-formyl-THF are resistant (green). **b)** Suggested mechanism of THF decomposition, releasing HCHO, a pterin moiety and p-aminobenzoylgluamate (ABG). Figure changed after [161] and [146].

Cellular HCHO is detoxified by glutathione (GSH) which reacts probably non-enzymatically with HCHO via its cysteine thiol to form *S*-hydroxymethylglutathione (HMG); HMG is oxidised to *S*-formylglutathione by ADH (ADH5 in humans, located in the cytosol),[162] which is further metabolized by *S*-formylglutathione hydrolase (FGH) to recover GSH and release formate (Figure 1.20).[163] This pathway represents the "tier 1" protection mechanism, preventing HCHO to cross-link DNA.[161]

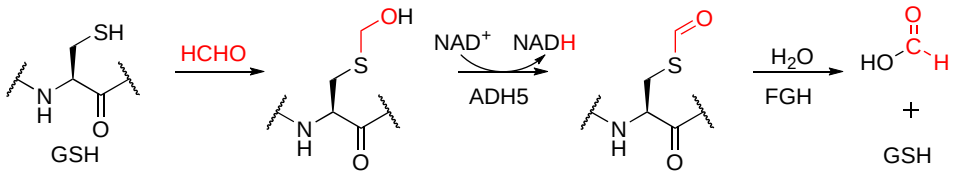


Figure 1.20 | HCHO detoxification in humans. GSH reacts with HCHO to form *S*-formylglutathione, which is turned over to *S*-formyl glutathione by ADH5. Hydrolysis of *S*-formyl glutathione by FGH yields GSH and formic acid.[163]

HCHO-derived formate has been shown to be a meaningful source of 1C units for nucleotide synthesis in a ADH5-dependent manner; hence, detoxification of genotoxic HCHO results in a benign 1C unit that drives nucleotide metabolism.[161] Turnover of mitochondrial HCHO to formate is catalysed by ALDH2.[164] Whilst conversion of HCHO to formate is relatively quick, metabolism of formate to CO_2 is THF-dependent and realtively slow due to low THF concentrations in the liver.[165] Formate reacts with THF to form 10-CHO-THF, and a 10-CHO-THF dehydrogenase (ALDH1L1 in the cytosol, ALDH1L2 in mitochondria) catalyses the oxidation of 10-CHO-THF to CO_2 in mammals (Figure 1.21).[129]

Methanol poisoning is an example of formate toxicity in humans: Methanol itself is not very toxic, can be excreted by kidneys and lungs, and is quickly metabolised by ADH1 in the liver to HCHO, and then to formic acid via the GSH-dependent pathway described above. Formic acid can lead to acidosis, an abnormal decrease of blood pH, hence impeding mitochondrial oxidative phosphorylation by inhibition of cytochrome oxidase.[165] The selective toxicity of formate to the visual nerve may arise from a susceptibility of the prelaminar region of the optic nerve head to a decrease in available energy caused by inhibition of cytochrome oxidase by formate.[166]

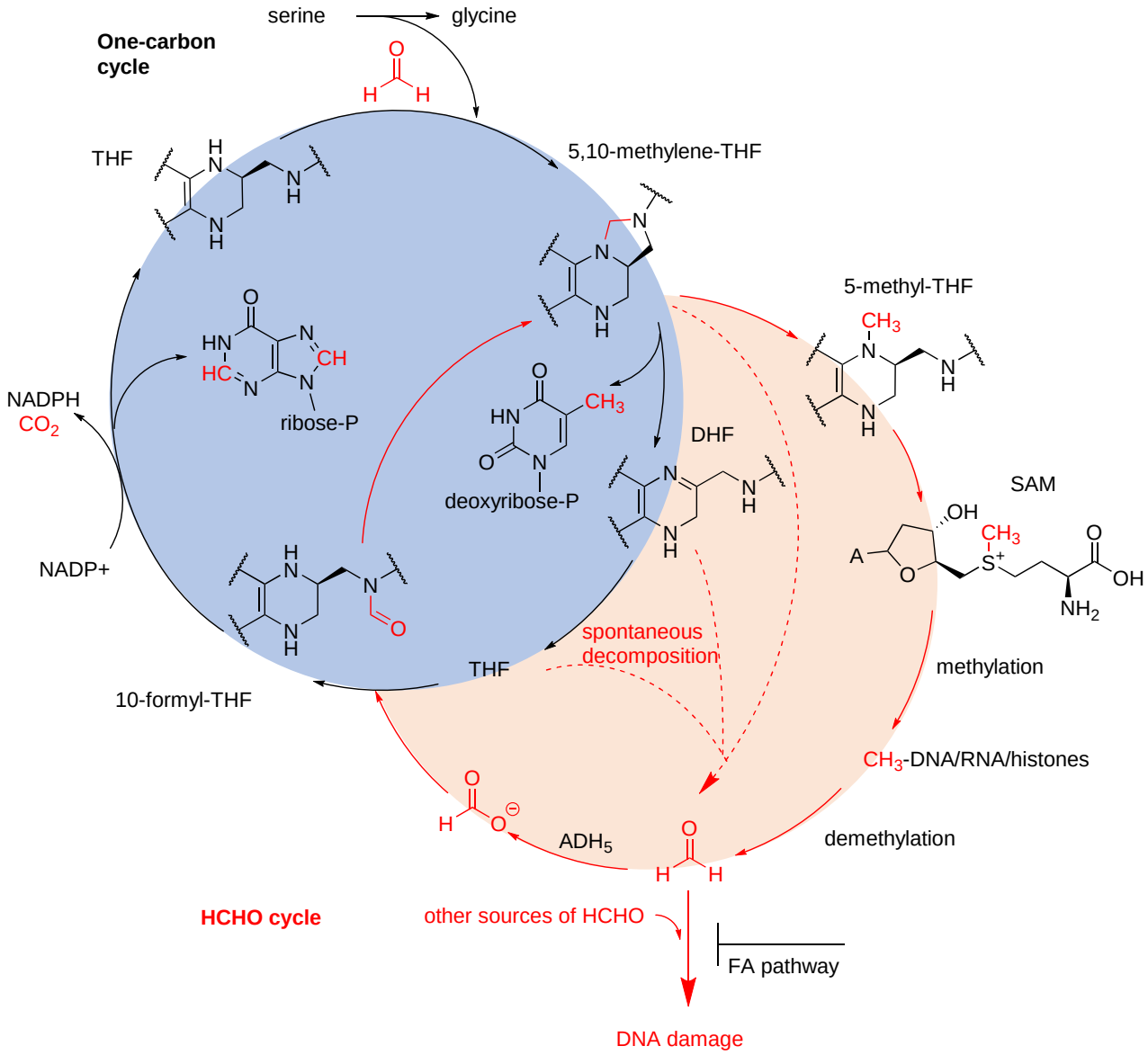


Figure 1.21 | Circuitous route for HCHO-derived 1C units. Suggested model for the mammalian HCHO cycle, in which both formate- and THF-derived 1C units are used for nucleotide synthesis. The spontaneous decomposition of folate derivatives, demethylation of DNA/RNA/histones, and conversion of serine to glycine increase endogenous HCHO. This figure was changed after [161].

In *Adh5* $^{-/-}$ mice, endogenous HCHO adducts have been shown to accumulate on DNA; these lesions can be repaired by the Fanconi anemia (FA) DNA cross-link repair pathway (mediated by FANCD2, "tier 2" protection), which is deficient in 1 in 160,000 individuals, leading to FA, which increases the risk of cancer and results in dysfunctional blood cell production; *Adh5* $^{-/-}$ *Fancd* $^{-/-}$ mice develop bone marrow failure caused by endogenous HCHO, meaning that mammalian cells produce sufficient HCHO to cause genotoxic damage.[81]

Similarly, "tier 1" protection against acetaldehyde (CH_3CHO) is mediated by ALDH2. Most studies investigated CH_3CHO toxicity as a result of ethanol intake, as ethanol is metabolised to CH_3CHO by alcohol dehydrogenase (ADH) and subsequently CH_3CHO is converted to acetyl-CoA by aldehyde dehydrogenase 2 (ALDH2). Loss of ALDH2 leads to DNA damage; again, the FA pathway represents "tier 2" protection against these CH_3CHO derived lesions, and is further complemented by homologous recombination and NHEJ; if these lesions are not repaired by the FA pathway, functionally impaired hematopoietic stem cells are the result.[167] Around 500 million people cannot clear CH_3CHO as they carry a polymorphism in *ALDH2* (asian alcohol flush),[168] leading to higher risk of esophagal cancer and aversive reaction upon ethanol consumption.[169] Further, ADH1B and ADH1C variants are prevalent among the Asian population.[170]; the coding variant of *ADH1B* common in Asian populations oxidises ethanol $\sim$70-fold faster, leading to the aversive effects of transiently elevated CH_3CHO; carriers of the *ADH1B* variant have been found to be less susceptible to alcoholism.[171] Endogenous CH_3CHO toxicity was reported on mouse ovary during antral follicular development with concomitant increased steroidogenesis, indicating CH_3CHO is a by-product of steroid biosynthesis.[172] CH_3CHO derived DNA interstrand cross-links (ICL, Figure 1.22 a) can be repaired by the FA pathway which uses excision to form a strand with an adduct and a broken strand; the first is removed by excision repair and the latter is repaired by homologous recombination (Figure 1.22 b).[173] Crotonaldehyde, produced by the aldol condensation of acetaldehyde, has been shown to be oxidised by microsomal (ALDH3A2) subcellular fractions, but not by cytosolic (ALDH1) and poorly by mitochondrial (ALDH2) fractions.[174]

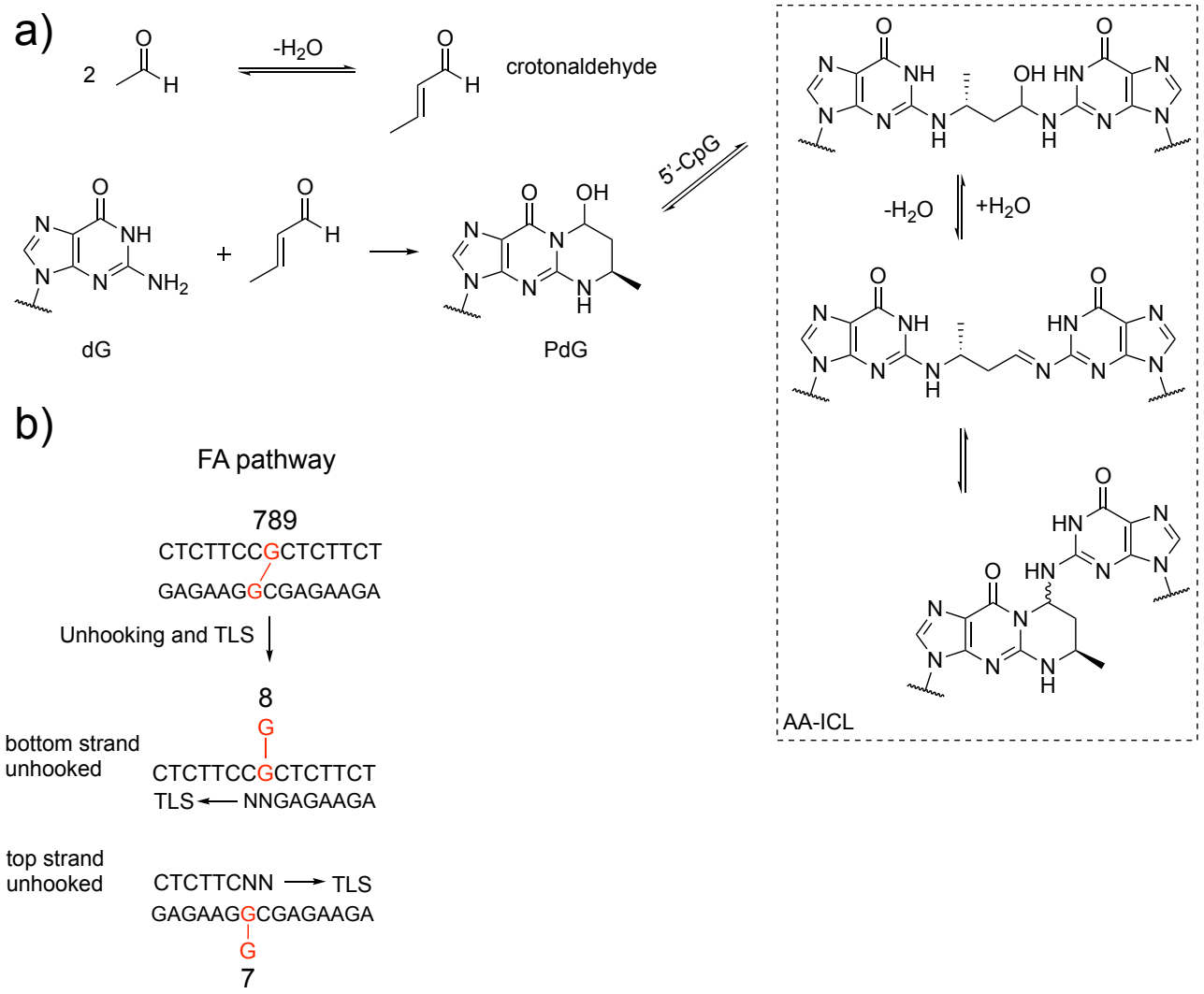


Figure 1.22 | CH_3CHO derived DNA cross-links. a) Proposed mechanism of CH_3CHO-derived interstrand cross-link (ICL) formation (AA-ICL). First, two acetaldehyde molecules react to form the aldol condensation product crotonaldehyde. The exocyclic amine of deoxyguanine (dG) reacts with crotonaldehyde to form N^2-propanoguanine (PdG). The exocyclic amine of 5'-CpG then reacts with PdG to form AA-ICL, which is in equilibrium between three forms. **b)** ICL repair by the FA pathway shows how the adduct is bypassed by translesion synthesis (TLS, left). This figure was modified from that reported in [173]. Note HCHO cannot react via aldol type chemistry to form a Michael acceptor.

The complexity in cross-linking reaction is further highlighted by the potential of HCHO to react with CH_3CHO to form prop-2-enal, also known as acrolein (Figure 1.23 a). The industrial synthesis of acrolein from HCHO and CH_3CHO was first described in 1937.[175] Acrolein is of toxicologic interest due to its high reactivity with protein thiol groups.[176] Acrolein is formed as metabolic by-product upon degradation of threonine by myeloperoxidases,[177] lipid peroxidation of polyunsaturated fatty acids,[178] catabolism of polyamines such as spermidine,[179] metabolism of anti-cancer drugs like

cyclophosphamide (Figure 1.23 c).[180] Acrolein is further formed from burning of fat-containing foods and is found in cigarette smoke.[181, 182] The toxicity in cigarette smoke is mainly due to α,β-unsaturated aldehydes such as acrolein.[183] Acrolein has been shown be detoxified via the glutathione-dependent pathway (Figure 1.23 b).[176] Furthermore, acrolein has been shown to be oxidised to the corresponding acid by various members of the ALDH familiy, but was turned over most efficiently by ALDH2; however high concentrations of acrolein above 10 µM resulted in ALDH2 inhibtion, whereas ALDH3A1 was not inactivated even at acrolein concentrations as high as 20 mM, indicating different scopes for acrolein detoxification within the ALDH enzyme family.[184]

Heating up food rich in carbohydrates results in formation of the carcinogen and neurotoxin acrylamide. The proposed major reactants leading to acrylamide formation are asparagine and sugars. The α-amino group of asparagine is proposed to react with the carbonyl group of sugars, and upon elimination of water, a Schiff base is formed. The Schiff base undergoes decarboxyation to form an intermediate that reacts with water to form 3-aminopropionamide, which eliminates ammonia to form acrylamide (Figure 1.23 d).[185] GSH has been proposed to be one of the major reactants of acrylamide; reaction of GSH with acrylamide results in formation of the corresponding Michael addition product (Figure 1.23 d).[186]

Paracetamol (also known as *N*-acetyl-para-aminophenol or acetaminophen) has been reported to react with GSH, thereby depleting GSH levels, which results in higher prevalence of reactive oxygen and reactive nitrogen species. These reactive species lead to necrotic changes in hepatocytes. Acetamoniphen overdosing accounts for half of all cases of acute liver failure in the United States and the United Kingdom.[187]

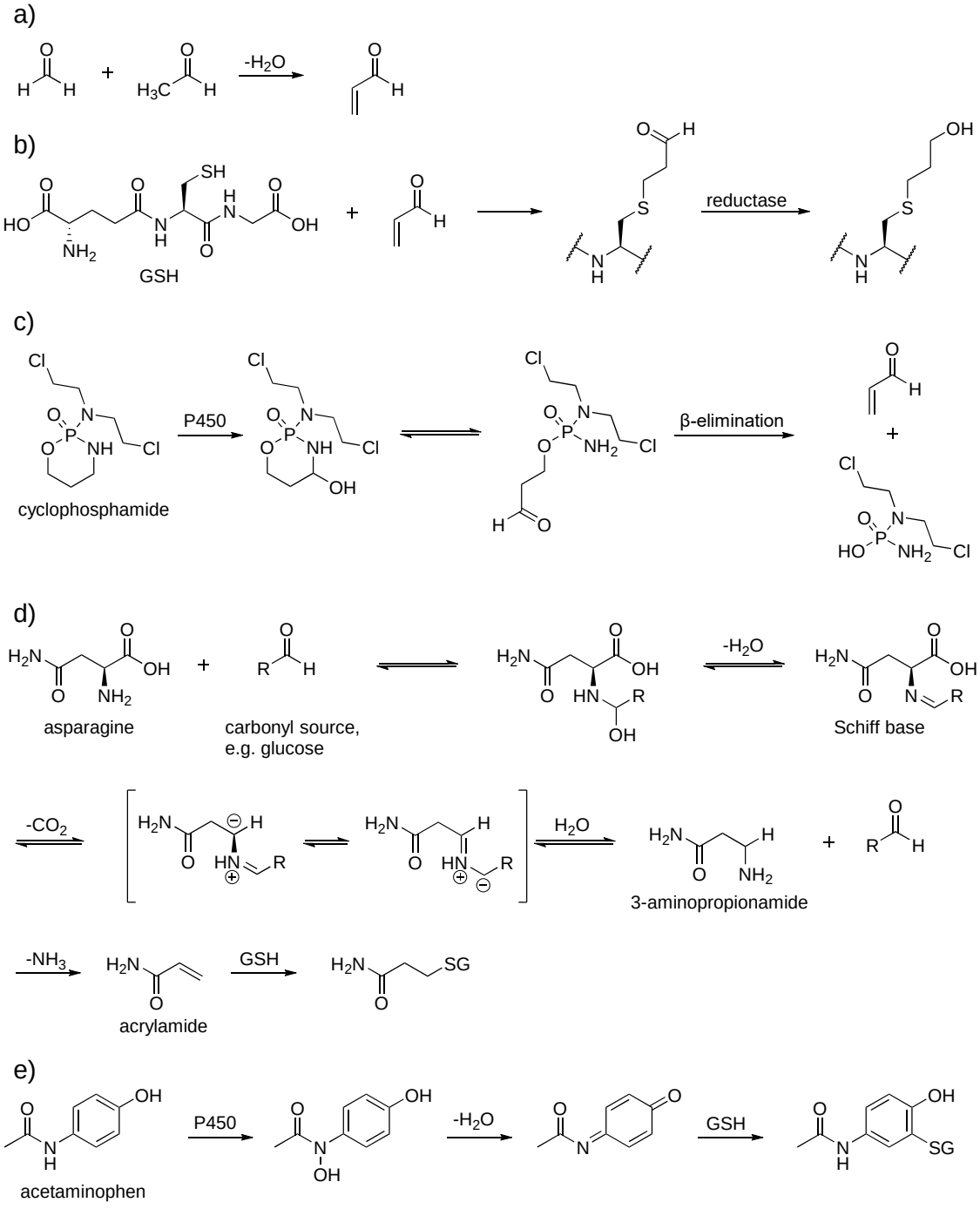


Figure 1.23 | Michael acceptors that react with GSH. a) Reaction of HCHO and CH_3CHO to yield acrolein. The first described industrial synthesis of acrolein used equimolecular amounts of vapours of 30% HCHO and CH_3CHO over argillaceous earth at 260-300 °C.[175] **b)** Glutathione (GSH) reacts with acrolein to form the corresponding Michael addition product which was found to be reduced *in cellulo*.[176] **c)** P450-mediated oxidation of the anti-cancer prodrug cyclophosphamide results in the hydroxylated intermediate, which undergoes β-elimination to form the toxic metabolites acrolein and phosphoramide mustard.[188] **d)** Proposed mechanism of acrylamide formation from asparagine and a carbonyl source, followed by reaction of acrylamide with GSH.[185, 186] **e)** P450-mediated *N*-hydroxylation of acetaminophen, followed by elimination of water and reaction with GSH.[189, 190]

1.3.2 Ketosis and Diabetes

Ketone bodies are crucial energetic substrates derived from fatty acids and are major glucose-sparing units under carbohydrate-limiting circumstances such as fasting; upon β-oxidation of fatty acids in hepatocytes, ultimately resulting in acetyl-CoA, carbohydrate-depleted conditions result in low oxaloacetate levels which would otherwise combine with acetyl-CoA in the first step of the Krebs cycle; this makes acetyl-CoA available for hepatic ketogenesis, ultimately resulting in the ketone bodies acetoacetate, acetone, and β-hydroxybutyrate (Figure 1.24 a), the latter of which is not a ketone; these three relatively stable carbonyl compounds are then released from the liver and distributed to extrahepatic tissues, where they are oxidised to acetyl-CoA which enters the Krebs cycle (Figure 1.24 b); levels of these ketone bodies are elevated in diabetes mellitus.[191]

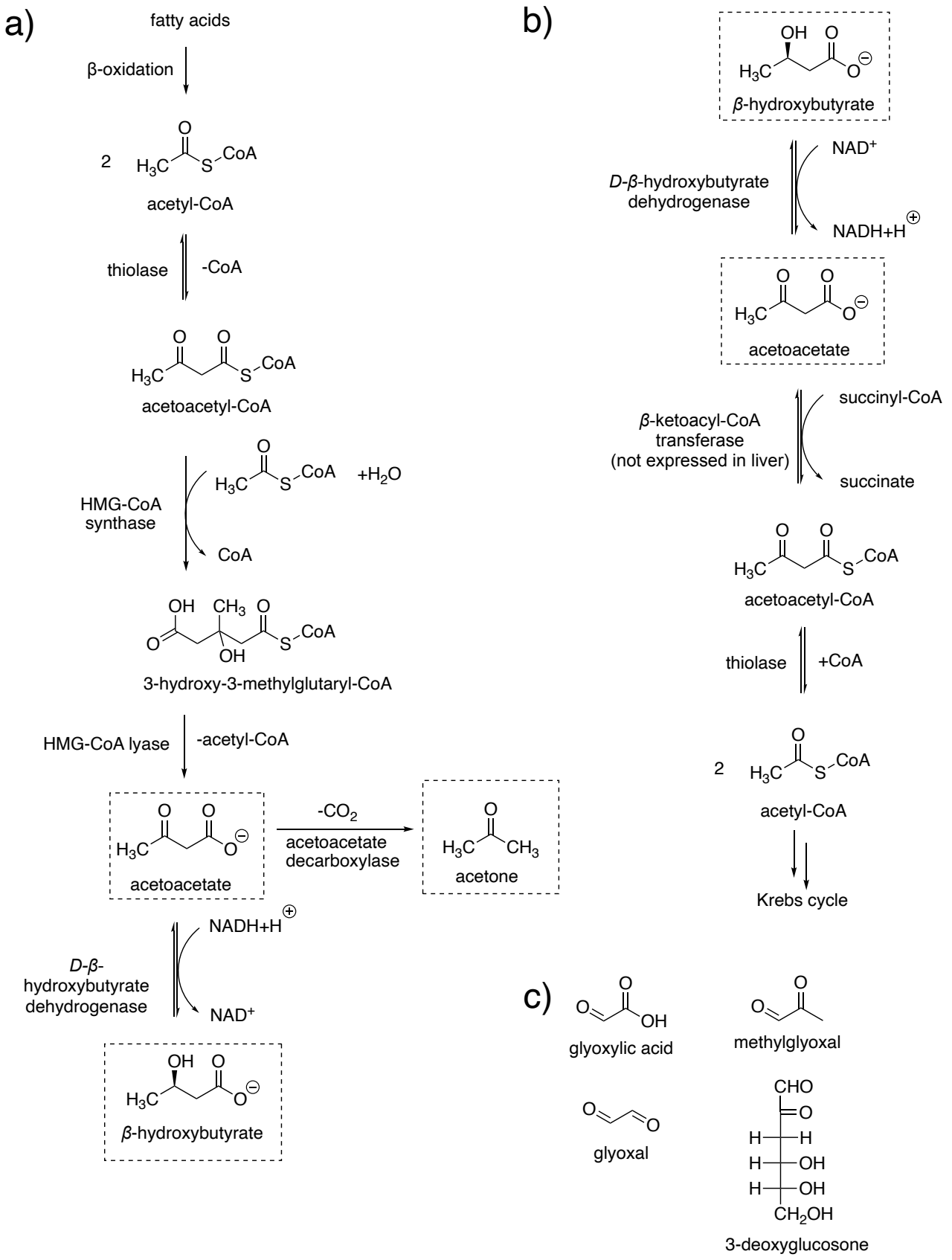


Figure 1.24 | Ketone body metabolism and reactive aldehydes. a) Hepatic ketogenesis. Changed after [191, 192]. **b)** Ketone body oxidation. Ketone bodies are highlighted in boxes. Changed after [191, 192]. **c)** Reactive α-oxoaldehydes involved in AGE formation.[193]

α-Oxoaldehydes are elevated in diabetes mellitus, which are reactive glycating compounds resulting in formation of advanced glycation end-products (AGE) on proteins and peptides; methylglyoxal (MGO), glyoxal, and 3-deoxyglucosone (Figure 1.24 c) are examples of α-oxoaldehydes and are degradation products of glucose and involved in non-enzymatic glycosylation (glycation) of proteins which in turn are involved in the development of diabetic pathophysiology.[193] As opposed to glucose, MGO can cross the cell membrane without cofactors or the action of transporters.[194, 195] Further, among the α-oxoaldehydes, MGO has the highest endogenous flux (3 mmol per day for a healthy human adult), making it a primary concern.[196] In aqueous solution, MGO is present in its unhydrated, monohydrated and dihydrated form in an approximate ratio of 1:78:28.[197] In contrast, glyoxal exists as ∼0.005% unhydrated, 0.5% monohydrated, 1–2% dimeric, and the remainder dihydrated form.[198, 199] The high proportion of unhydrated form of MGO relative to glyoxal is proposed to make MGO the more reactive α-oxoaldehydes.[200] The most important source of cytotoxic MGO is glycolysis, where MGO can be formed via decomposition of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate.[201] MGO is detoxified via the glyoxalase system, where MGO first reacts non-enzymatically with GSH to form the hemithioacetal, which is converted by glyoxalase I to the thioester *S*-lactoyl-glutathione. This thioester is then hydrolysed to lactate by gloyxalase II, recovering GSH (Figure 1.25 a). MGO is reported to react with lysine and arginine residues on histones to form AGEs, which can be deglycated by the aciton of DJ-1 (Figure 1.25 b). Under physiologically relevant conditions, glucose has been shown to form MGO, glyoxal and 3-deoxyglucosone over time.[193] Glyoxal is detoxified analogous to MGO (Figure 1.25 a) via the glyoxalase pathway. 3-Deoxyclucosone has been shown to be oxidised by ALDH1A1 to 2-keto-3-deoxygluconate, which is then excreted via the urine.[202]

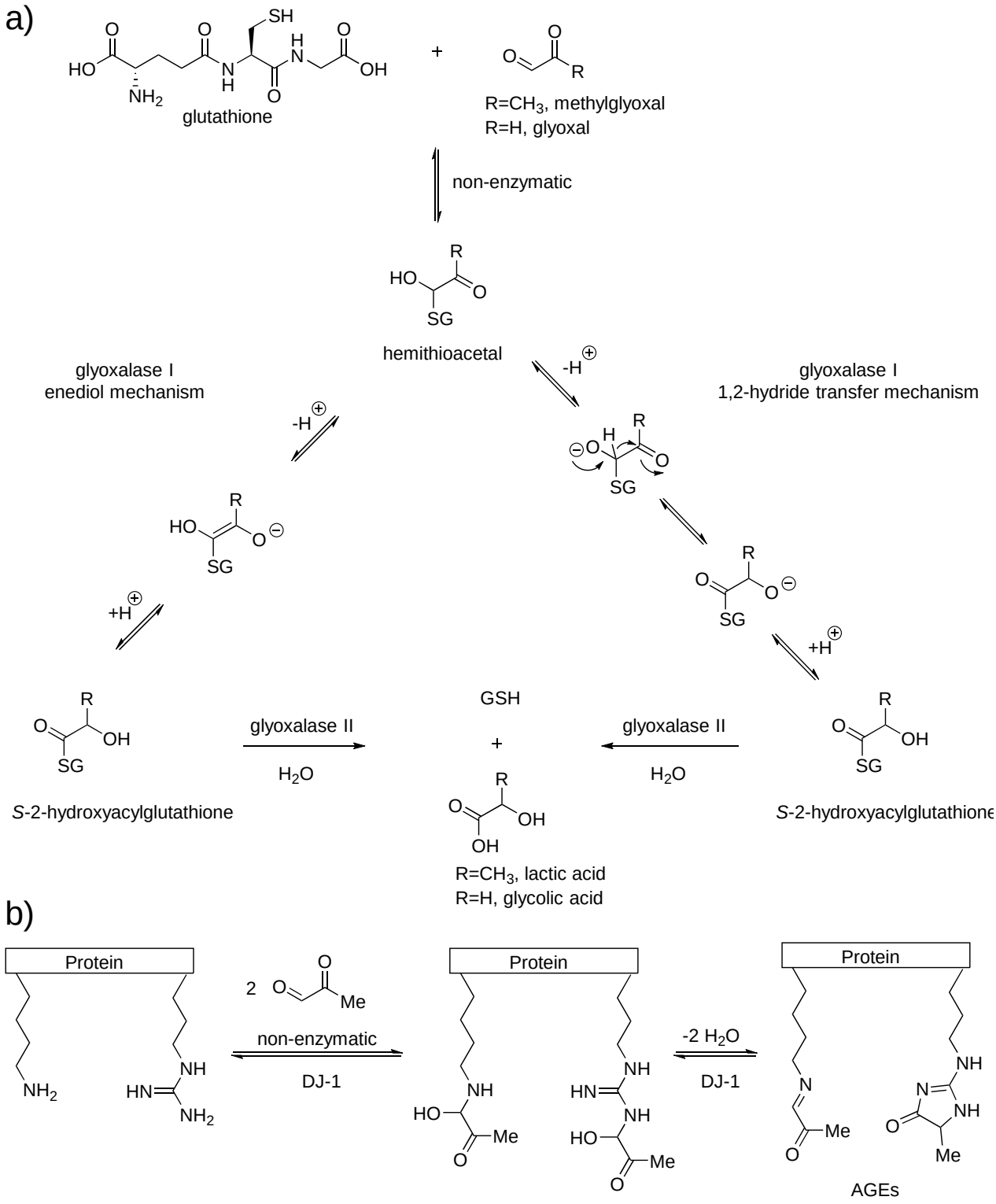


Figure 1.25 | Detoxification of MGO/glyoxal and MGO-induced AGEs. a) After non-enzymatic reaction of GSH and α-oxoaldehydes like MGO or glyoxal to a hemithioacetal, glyoxalase I catalyses the conversion to an *S*-hydroxyacylglutathione. The reaction mechanism for glyoxalase I is controversial and is proposed to proceed either via an enediol mechanism (left) or an 1,2-hydride transfer mechanism (right). The thioester is then hydrolysed by glyoxalase II to form the α-hydroxyacid and GSH.[203, 204] **b)** Reported reactions of MGO with histones resulting in AGE formation. The oncogenic protein DJ-1 is a histone deglycase that reverses MGO-induced damage.[205]

Another marker metabolite of diabetes is glycine-derived glyoxylic acid,[206] levels of which are altered three years before a formal diagnosis of diabetes in plasma; glyoxylic acid is also involved in the pathology of AGE related diabetes (Figure 1.24 c).[207] Hemoglobin (Hb) is a heterotetrameric (two α and two β subunits bound non-covalently) protein in red blood cells that binds oxygen, hence Hb is responsible for carrying oxygen from lungs to the rest of the body. Glucose reacts in its linear aldehyde form with the N-terminal amino group of the Hb β-chain (Figure 1.26) and approximately with 15 other lysine residues in Hb, resulting in so-called HbA1c. The relatively long half life of Hb of $\sim$120 days reflects the average level of blood glucose during this period; as glucose levels are typically elevated in persons with diabetes, higher HbA1c levels can be detected by a range of chromatographic or chemical methods, thus enabling diagnosis of diabetes.[208, 209]

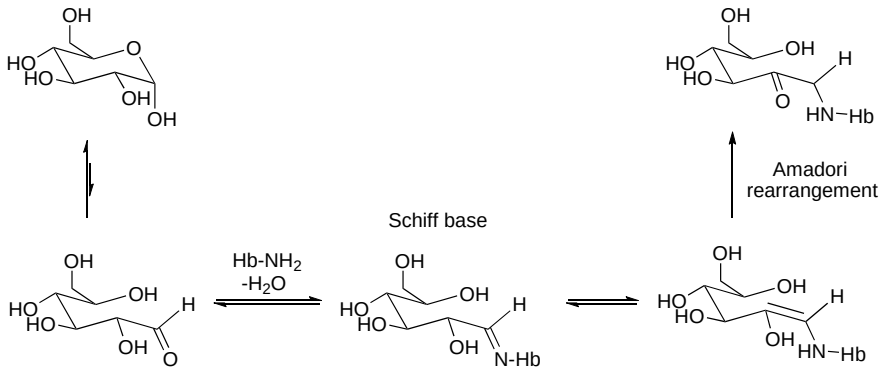


Figure 1.26 | Glycation of hemoglobin. Glucose reacts in its linear aldehyde form with the N-terminus of hemoglobin (Hb) to form a Schiff base, which undergoes two tautomerisation reactions (enaminol formation and keto-enol tautomerisation) to irreversibly form the Amadori product.[210, 211]

Despite the prevalence of an array of different endogenous aldeyhdes and ketones, HCHO - the simplest aldehyde - is prone to reactions with nucleophiles due to its highly electrophilic nature. Even with simplest biomolecules it reacts with a perplexing complexity and the extent of its involvement in healthy biology remains an open question. Excess amounts of HCHO relative to any given protein or other biomolecule are conceivable, especially with respect to the potential of high local nuclear HCHO concentrations. The notion of high local nuclear HCHO levels led to the hypothesis that histones may bear certain amino acid sequences that scavenge HCHO. Indeed, as described in Chapter 2, N-terminal prolines as found on histone H2B are particularly susceptible to reactions with HCHO. Competition

experiments showed that cysteine reacts even more efficiently with HCHO than N-terminal prolines. Chapter 3 describes the susceptibility of asparagine synthetase, which bears an N-terminal catalytic cysteine, to inhibition by HCHO and other biologically relevant aldehydes *in vitro* and *in cellulo*. The work described herein provides insight into the potential biochemical consequences of HCHO reacting with protein N-termini, where the free N-terminal amine enables ring formation with HCHO and any nucleophilic group in close proximity. Studying the fundamentals of non-enzymatic post-translational modifications is crucial for understanding biochemical consequences of these processes. By extension, this work also highlights the potential for reactions underlying the non-enzymatic metabolome, where reactive aldehydes might play a key role. In Chapter 4, the structure of metampicillin which is formed by reacting ampicillin with HCHO is revised to be a cyclic aminal rather than the previously published imine or hemiaminal. These structural insights may be relevant to understanding why metampicillin is selectively accumulated in the bile compared to its parent drug ampicillin.[212, 213]

1.3.3 Aims

The objectives of the work described in this thesis were to study the potential biochemical consequences of reactions of HCHO with both N-terminal proline- (Chapter 2) and N-terminal cysteines-residues (Chapter 3). Initial *in vitro* findings to elucidate the structures of the adducts used NMR and mass spectrometry as a primary analytical tool. These studied were followed by investigations on how stable the formed HCHO adducts are in presence of HCHO scavengers. Model enzymes bearing an N-terminal catalytic proline or an N-terminal catalytic cysteine were then employed to study HCHO dependent enzyme inhibition *in vitro*. The data obtained *in vitro* were then validated by studies in cells, aiming to understand the potential for the *in vitro* findings to be relevant *in cellulo*. Chapter 4 provides detailed NMR studies on the formation of metampicillin, increasing the understanding of the intricacies in the reactions of HCHO with nucleophilic drugs. In Chapter 5, the nucleophilic cysteine proteases of SARS-CoV-2, namely M^{pro} and PLPro, are studied. Mass spectrometry based assays were developed to study inhibition of these proteases.

2

Formaldehyde and N-terminal Prolines

Laboratorium - im Sinne des Mittelalters, weitläufige unbehülfliche Apparate zu phantastischen Zwecken.
[...]
Es leuchtet! seht! – Nun läßt sich wirklich hoffen,
Daß, wenn wir aus viel hundert Stoffen
Durch Mischung – denn auf Mischung kommt es an –
Den Menschenstoff gemächlich komponieren,
In einen Kolben verlutieren
Und ihn gehörig kohobieren,
So ist das Werk im stillen abgetan.
Zum Herd gewendet.
Es wird! die Masse regt sich klarer!
Die Überzeugung wahrer, wahrer:
Was man an der Natur Geheimnisvolles pries,
Das wagen wir verständig zu probieren,
Und was sie sonst organisieren ließ
Das lassen wir kristallisieren.

Johann Wolfgang von Goethe - Faust II [214]

Laboratory - After the manner of the Middle Ages;
extensive, ponderous apparatus for fantastic purposes.
[...]
It brightens, -see! Sure, now, my hopes increase
That if, from many hundred substances,
Through mixture - since on mixture all depends-
The human substance gently be compounded,
And by a closed retort surrounded,
Distilled, and fed, and slowly founded,
Then in success the secret labor ends.
(Again turning towards the furnace.)
'T will be! the mass is working clearer
Conviction gathers, truer, nearer!
The mystery which [...] in Nature lies
We dare to test, by knowledge led;
And that which she was wont to organize
We crystallize, instead.

Johann Wolfgang von Goethe - Faust II [215]

Contents

2.1 Abstract **56**
2.2 Introduction **56**
2.3 Results **62**
2.4 Discussion **86**
2.5 Supplementary Information **88**
2.6 Methods **109**

2.1 Abstract

As described in Chapter 1.2.6, there is indirect evidence that chromatin will be subject to locally high HCHO concentrations in the nucleus. It was hypothesised that histones may play a role in HCHO scavenging and/or that HCHO may contribute to the regulation of transcription. HCHO is produced during demethylase-catalysed demethylation of histones and local surges during transcriptional reprogramming are reported. In this Chapter, two proteins containing N-terminal proline residues, histone H2B and 4-oxalocrotonate tautomerase (4-OT), are studied with regard to their reactions with HCHO. The studies reveal both proteins react with HCHO on their N-terminal proline residues. With H2B, HCHO reacts with proline to form of a 5,5-bicyclic adduct that is stable to competition with other biologically relevant carbonyl compounds and thiol-containing HCHO scavengers. With 4-OT, pre-incubation with HCHO induced inhibition of catalytic activity *in vitro* and *in cellulo*, which is proposed to be due to formation of hemiaminals on the catalytic N-terminal proline of 4-OT. The HCHO-derived modification of the N-terminal proline of 4-OT was confirmed by MS/MS analyses of 4-OT overexpressed in *E. coli* cells that were treated with HCHO. Collectively, the results imply that N-terminal proline residues on proteins can react with HCHO to form adducts of varying stabilities. These studies therefore expand the understanding of HCHO's biologically relevant chemistry and suggest novel regulatory functions for HCHO-derived protein adducts in cells.

2.2 Introduction

Most organisms are likely constantly exposed to HCHO from exogenous or endogenous sources (Chapters 1.1.1 and 1.2.6). HCHO sensing has therefore evolved to enable detoxification responses. One example of a HCHO sensing/detoxication system is the *frmRAB* operon (Chapter 1.2.8). In *E. coli* the transcription factor FrmR acts as a HCHO sensor forming a HCHO-derived methylene bridge that links the spatially adjacent N-terminal Pro2 and Cys35 with HCHO, enabling DNA binding and consequently a HCHO detoxification response. Binding of the regulator FrmR to DNA results in transcription of *FrmA* and *FrmB*. Upon reaction of HCHO with glutathione (GSH) to

form *S*-hydroxymethylglutathione (HMG), FrmA converts HMG to *S*-formylglutathione, which is hydrolysed to GSH and formate by FrmB.[152] The assignment of FrmR as HCHO sensing transcription factor and the conservation of the Pro2 and Cys35 residues of FrmR in *Salmonella enterica* and *E. coli* raises the question of whether N-terminal proline residues in general have evolutionary roles in HCHO sensing. To the best of our knowledge, in the whole of biology only two enzymes with catalytic N-terminal prolines have been characterised, one in the *E. coli* Fpg protein, which is a DNA glycosylase/AP lyase involved in removing base lesions, e.g. the mutagenic C8-oxo-guanine (8-oxoG, Figure 2.1 a).[216] The other enzyme with an N-terminal catalytic proline has been identified in *Pseudomonas putida* mt-2: 4-oxalocrotonate tautomerase (4-OT) uses its N-terminal proline to tautomerise the non-conjugated (4*E*)-2-oxo-hexenedioate (4-oxalocrotonate) to the conjugated (3*E*)-2-oxo-hexenedioate (Figure 2.1 b).[217] Replacing the N-terminal proline of 4-OT with alanine still results in isomerization of the substrate, but at lower rates; interestingly, the P1A mutant was alkylated by the substrate, forming an imine intermediate, a reaction that was not observed with wt-4-OT.[218] Thus, it seems that the N-terminal proline of 4-OT has been evolutionary conserved to avoid nucleophilic reactions (possibly including Schiff base formation) with its substrate.

Some microorganisms use aromatic hydrocarbons as the only source of energy and carbon; one of the pathways is the catechol meta-fission pathway; this pathway uses xylenes, benzene, and toluene to yield catechol or derivatives, which upon ring fission are further catabolised inter alia by 4-OT, to ultimately yield acetaldehyde and pyruvate (Figure 2.2).[219]

4-OT displays promiscuous activity of asymmetric Michael-type additions, catalysing the reaction of *trans*-β-nitrostyrene with linear aldehydes to yield γ-nitroaldehydes (*dr* values of up to 93:7 and *ee* values of up to 81%), which are useful precursors for GABA-analogues like the marketed pharmaceutical phenibut (Figure 2.1 c).[220, 221] Interestingly, substituting the catalytic proline of 4-OT with (2*S*,4*R*)-4-fluoroproline, (2*S*,4*S*)-4-fluoroproline, (2*S*)-3,4-dehydroproline, or (4*R*)-1,3-thiazolidine-4-carboxylic acid resulted in loss of catalytic activity, indicating that even minute structural modifications on the catalytic proline are not tolerated.[222]

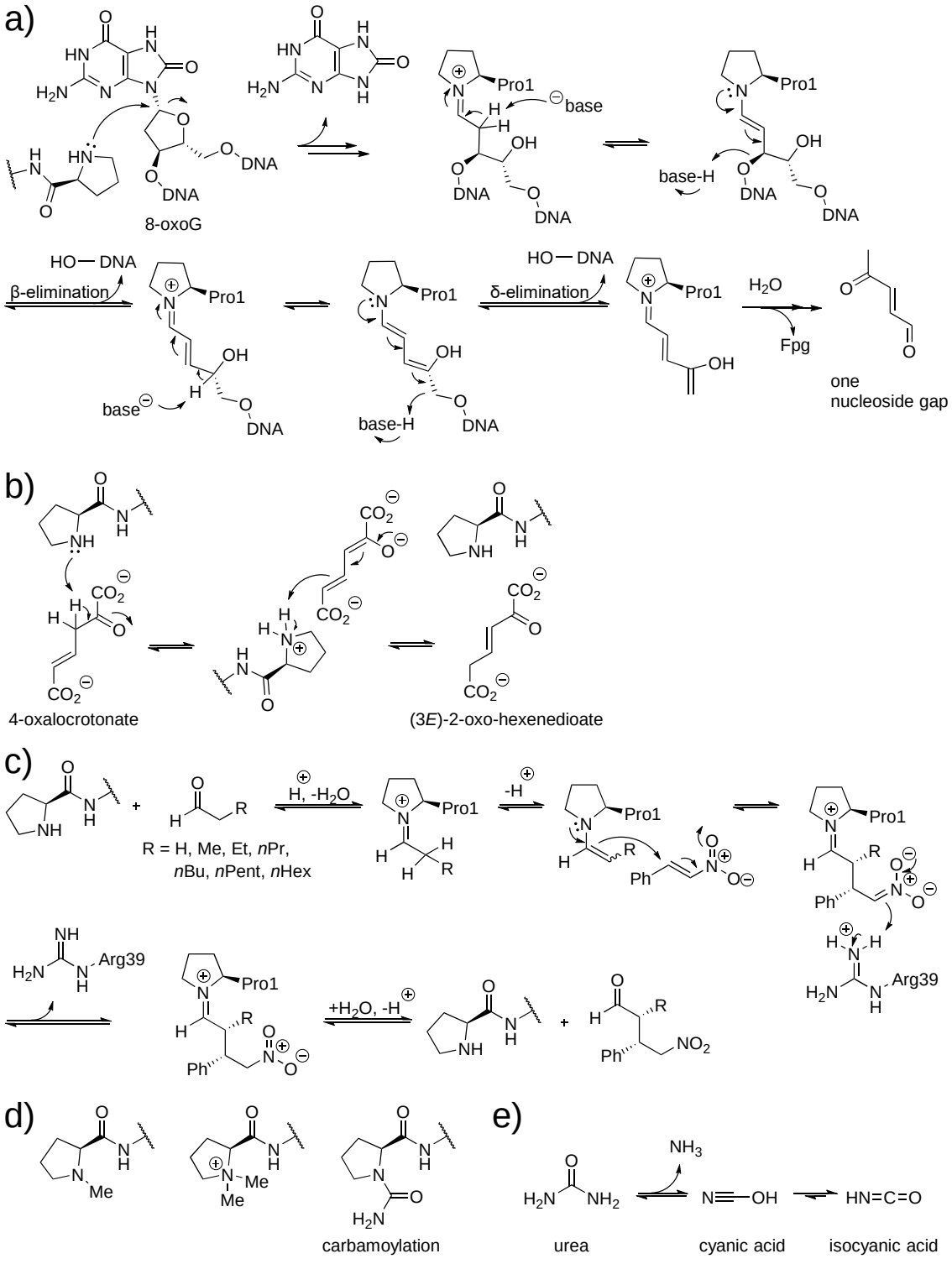


Figure 2.1 | Biochemical reactions catalysed by N-terminal prolines and PTMs of N-terminal prolines. **a)** The Fpg catalysed excision mechanism of 8-oxoG.[223] **b)** Tautomerisation of 4-oxalocrotonate as catalysed by 4-OT.[218] **c)** Promicuous catalysis of 4-OT reacting linear aldehydes with *trans*-β-nitrostyrene employing its N-terminal catalytic proline.[220] **d)** PTMs found on N-terminal prolines.[224–226] **e)** Urea degradation resulting in formation of isocyanic acid, which is reported to react with protein amino groups, resulting in carbamoylation as shown in Figure 2.1 d.[227]

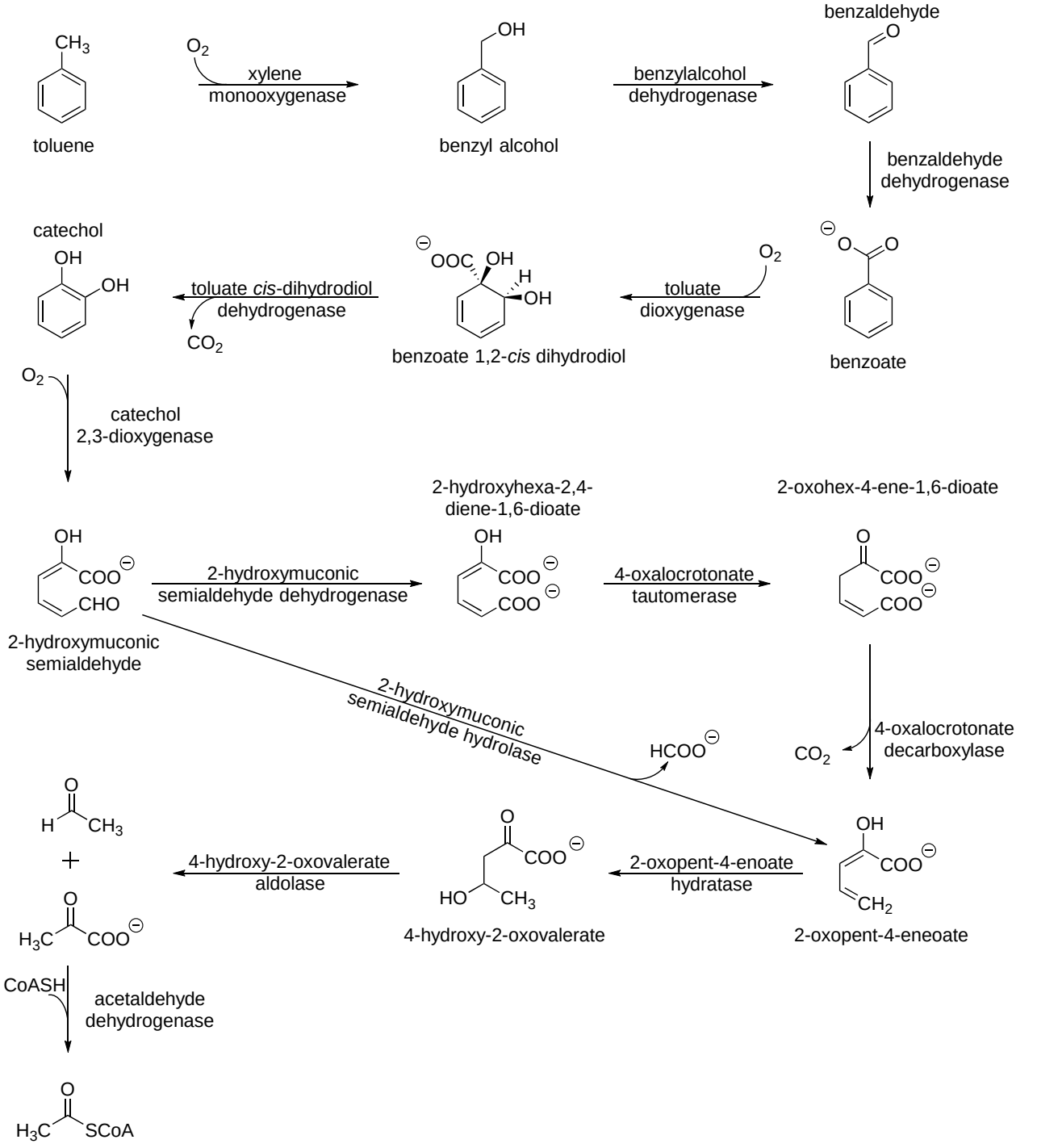


Figure 2.2 | Toluene catabolism by *P. putida* mt-2. *P. putida* mt-2 carries a transmissible plasmid which encodes genes that enable metabolism of aromatic compounds like toluene.[228, 229]

Reported post-translation modifications of N-terminal prolines include *N*-methylation in *Drosophila melanogaster*, where heat shock or arsenite treatment led to an increase in methylation of the N-terminal proline of H2B in cultured *Drosophila melaogaster* cells;[224] *N*-dimethylation on tryptic H2B peptides has been found on the N-terminal proline of gonads from *Asterias rubens*;[225] crystallographic evidence suggestes carbamoylation at the N-termial proline in HIV protease, likely resulting from residual isocyanic acid in urea which has been used for purification of the protein (Figure 2.1 d+e).[226] In renal diseases such as uremia, elevated urea leads to carbamoylation (not be confused with carbamylation, which is the reaction of amino groups with CO_2)[230] of the amino group of single amino acids and lysyl residues on proteins, thereby contributing to toxicity in uremia, e.g. by hampering protein hormone activity, protein biosynthesis (by lowering the amount of free amino acids), or enzymatic activity.[231] Prolyl aminopeptidases (PAPs) are enzymes that remove N-terminal prolines and are conserved across kingdoms; in *Arabidopsis thaliana*, PAP2 cleaves mitochondrial N-terminal targeting peptides containing N-terminal prolines; these peptides are accrued after mitochondrial protein import and signal peptide cleavage, and would be toxic if not hydrolysed.[232] Whilst reactions of HCHO with biomolecules such as amino acids[82] and nucleobases[83] have been defined, the reactions of HCHO with proteins are barely studied,[89, 114] though elucidating the resulting adducts on proteins is crucial for vaccine development.[88] Of particular interest are spatially close nucleophilic amino acids enabling cross-linking, and N-terminal amino acids bearing nucleophilic side chains which enable ring formation with HCHO and the free N-terminus.

Although 4-imidazolidinones (resulting from the reaction of HCHO with an N-terminal primary amine and the adjacent peptide bond nitrogen, Figure 1.13 a) have been reported on peptides *in vitro*,[89] the data provided in this Chapter suggest that the primary N-terminal amine and the adjacent peptide bond are relatively slow to react with HCHO, leaving doubts about the biological significance of this reaction.

Among the reported 178 human proteins bearing N-terminal prolines,[233] those in close proximity to nucleosomes are of particular interest, as a HCHO gradient with high local nuclear concentrations is proposed due to histone demethylation events. In eukaryotic nucleosomes, DNA is wound $\sim$1.65 times

around an octamer of basic histone proteins (two copies of each H2A, H2B, H3, and H4); the modification of these structures enables regulation of transcription: RNA Polymerase II transcribes DNA into RNA after the ordered disassembly of the nucleosome. The cellular identity can be regulated by PTMs on histone amino acids (methylation, phosphorylation, acetylation, ubiquitylation, sumoylation, etc.), sometimes so-called epigenetic changes, which regulate chromatin structure. Most histone PTMs are reversible, with opposing enzymes catalysing the addition (writers) or removal (erasers) of these modifications; histone PTMs regulate transcription either by directly affecting histone-DNA interactions and hence chromatin structure, or by acting as scaffolds where reader proteins bind to.[234] Whilst e.g. acetylation of the ϵ-amino group of histone lysines neutralises its positive charge, resulting in weaker binding of acetyl-lysine to the negatively charged DNA, other modifications, such as methylations on lysine (me1, me2 or me3) or arginine (mono-, symmetrically or asymmetrically di-methylaion) residues do not alter the charge of histone residues; *S*-adenosylmethionine (SAM) is used as methyl group donor which is enzymatically transferred to the ϵ-amino group of histone lysines or to the guanidinium group of arginine.[235] Further, methylation may affect gene expression in an activating (e.g. H3K4me3) or in a repressing (e.g. H3K27me3) manner.[236]

A recent study shows that transcriptional reprogramming in hematopoitetic stem cells with concomitant demethylation events results in an overload of nuclear HCHO as high as 500 µM, resulting in genome instability in absence of DNA crosslink repair (Fanconi anemia, FA).[115] Hence, it was proposed that histone tails might act as HCHO sinks with certain residues being particularly susceptible to reactions with HCHO. The results on histone tails show that N-terminal proline residues, being the only secondary amine among the proteinogenic amino acids, show enhanced reactivity towards HCHO relative to primary amines. The reaction of N-terminal prolines with HCHO results in formation of a fused 5,5-bicycle, bridging the proline nitrogen and the adjacent peptide bond nitrogen. H2B and 4-oxalocrotonate tautomerase (4-OT) were developed as model proteins, both bearing an N-terminal proline (catalytic in case of 4-OT), to show the biological significance of the reaction of HCHO with N-terminal prolines.

2.3 Results

Initial studies focused on determining the reactivity of histone tails with HCHO. Peptides containing the first 15 residues of histone N-terminal tails (H2A, H2B, H3, and H4) were synthesised and mixed individually with HCHO at different concentrations (equimolar to 10^4-fold excess) at pH 7.4. The reaction mixtures were then analysed by MALDI MS periodically over 24 hours. Interestingly, all peptides reacted with HCHO. The mass shifts observed (+12 Da, +24 Da and +36 Da relative to starting peptide) suggested formation of multiple -CH2- methylene groups (Figures 2.3 a, 2.15, 2.16); however, the H2A, H3 and H4 peptides only reacted significantly when mixed with a 10^4-fold excess of HCHO, whereas the H2B peptide reacted more significantly at lower HCHO concentrations but only formed an adduct (or adducts) with one +12 Da mass shift (Figure 2.3 b). No evidence for cross-linking was observed from these MALDI MS experiments.

MALDI MS experiments were then conducted on the H2B peptide and HCHO at 37 °C (Figure 2.3 c), and across a physiologically relevant pH range at ambient temperature (pH 5.6, 8.2; Figure 2.3 d). These studies revealed more significant formation of the +12 Da adduct(s) at 37 °C and under alkaline conditions. MS/MS analyses suggested only one +12 Da adduct is formed and revealed that the modification occurs on the N-terminal proline (Figure 2.3 e+f). This finding was supported by MALDI MS analyses with an H2B P1A variant peptide, which reacted much less proficiently with HCHO (Figure Figure 2.4 d).

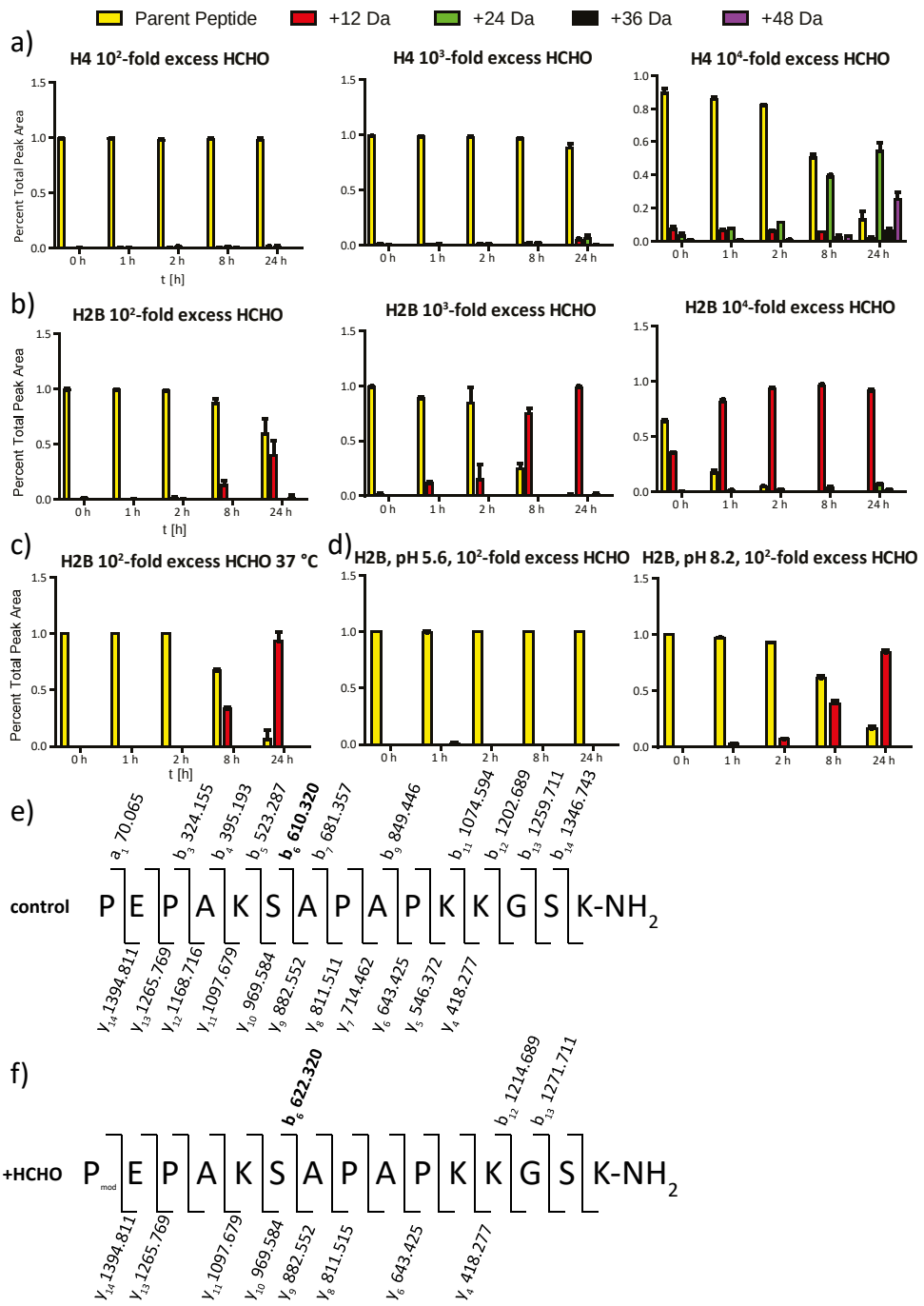


Figure 2.3 | H2B reacts more efficiently than H4 with HCHO. a) Reaction of H4 15mer (NH_2-SGRGKGGKGLGKGGA-NH_2) with HCHO (10^2-, 10^3-, and 10^4-fold excess), pH 7.4, ambient temperature. **b)** Reaction of H2B 15mer with HCHO (10^2-, 10^3-, and 10^4-fold excess HCHO), pH 7.4, ambient temperature. **c)** Reaction of H2B 15mer with 10^2-fold excess HCHO, pH 7.4, 37 °C. **d)** Reaction of H2B 15mer with 10^2-fold excess HCHO, pH 5.6 or 8.2, ambient temperature. HCHO adducts are shown in red (+12 Da), green (+24 Da), black (+36 Da) and purple (+48 Da), parent peptide (no reaction) is shown in yellow. Reactions for a-d were monitored by MALDI MS after addition of HCHO (0 h), after 1 h, 2 h, 8 h, and 24 h. Errors: SD of the mean (n=3, technical repeats). **e+f)** Top-down MALDI MS/MS results summary with H2B 15mer without (e) and with (f) HCHO. MS/MS analyses shows the H2B peptide (C-terminal amide) and the generated a-, b- (N-terminal) and y-ions (C-terminal) obtained upon fragmentation, indicating that the +12 Da shift occurs at the N-terminus (b_6, in bold).

Studies were then conducted on a truncated H2B pentapeptide (NH-PEPAK-NH_2), which reacted with ^{13}C-labelled HCHO to form an adduct with a +13 Da mass shift (Figure 2.4 c). When compared to untreated peptide, chemical shift changes in the ^{1}H resonances corresponding to the N-terminal prolyl α, γ, and δ hydrogens were observed (Figure 2.4 a). Two new ^{1}H resonances were also identified and were assigned to the hydrogens of a methylene bridge (δ_H 4.3 ppm and δ_H 4.5 ppm, Figure 2.4 b). HMBC correlations from the ^{1}H resonances at δ_H 2.59 ppm and δ_H 3.14 ppm, corresponding to the N-terminal prolyl δ hydrogens, to the methylene bridge ^{13}C resonance (at δ_C 66 ppm) indicate modification on the N-terminal prolyl secondary amine (Figure 2.4 b). This proposal importantly implies that HCHO reacts more efficiently with secondary amines (i.e. the prolyl amine) rather than with the primary amines of alanine and serine present on H2A, H3 and H4 N-termini, at least to give detectable adducts. It is likely that the second position of the peptide (glutamate in case of H2B) only plays a minor role with regards to the reactivity of N-terminal prolines; although no experimental evidence was accrued, an alanine at the second position is not expected to change the reactivity of the N-terminal proline. Equally, replacing the N-terminal proline with the secondary amine sarcosine is expected to result in reactivity that is comparable to the reactivity of an N-terminal proline.

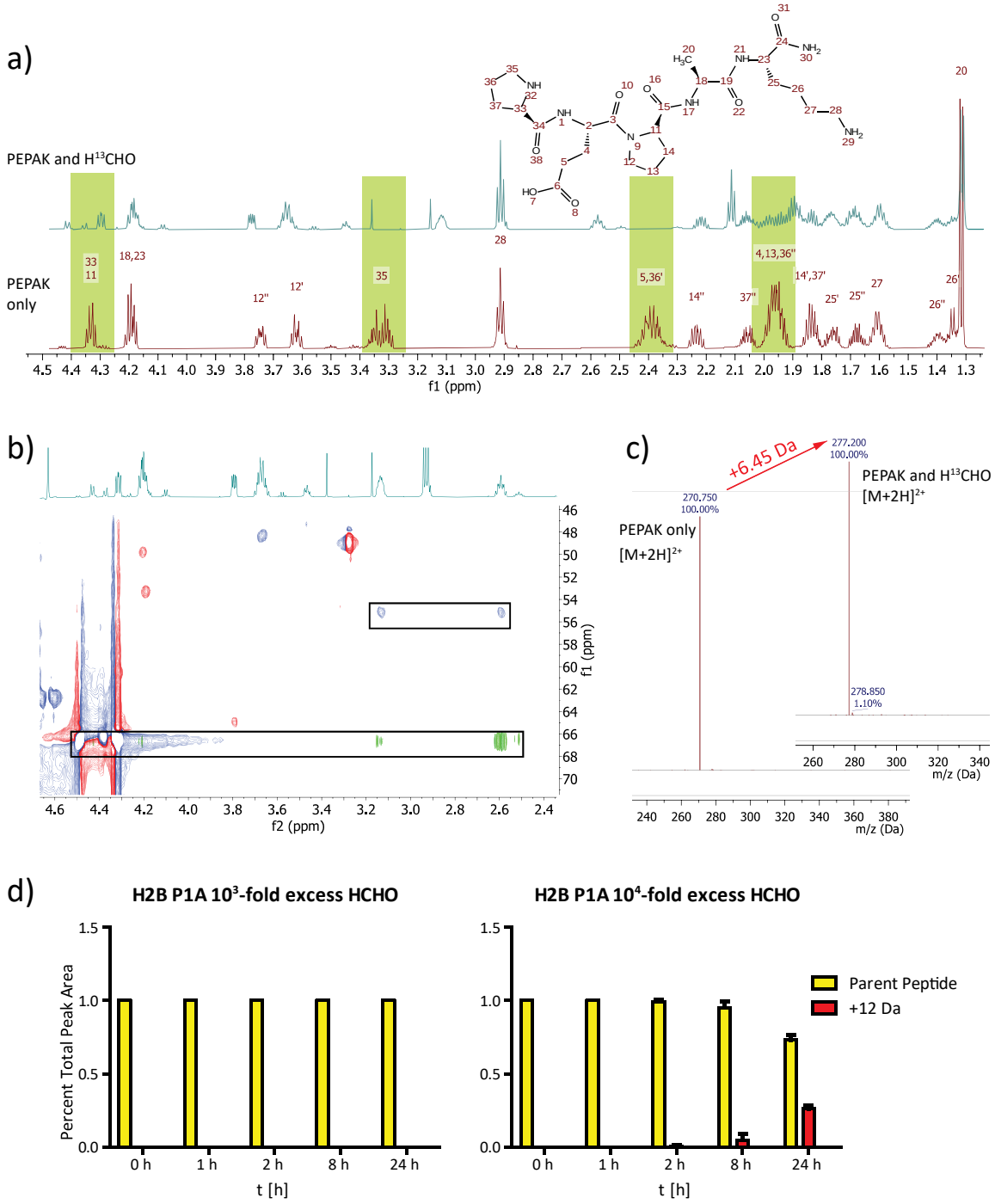


Figure 2.4 | The N-terminal proline of H2B reacts with HCHO. a) ^{1}H NMR (700 MHz) of the H2B pentapeptide (H2B 1-5 NH-PEPAK-NH_2) before (bottom panel) and after addition of H^{13}CHO (top panel, prepared *in situ*). Chemical shift changes are highlighted with green bars. **b)** Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC spectra of the H2B pentapeptide upon *in situ* reaction with H^{13}CHO. Signals of interest are highlighted with a black rectangle. **c)** LC-MS spectra of the H2B pentapeptide before and after reaction with H^{13}CHO (*in situ*), showing a mass increase of 12.9 Da. **d)** H2B P1A 15mer was reacted with a 10^3- and 10^4-fold excess HCHO at pH 7.4 and ambient temperature. The reaction was monitored over 24 h by MALDI MS. Errors: SD of the mean (n=3, technical repeats). H2B P1A 1-15 (NH_2-AEPAKSAPAPKKGSK-NH_2).

To further assign the structure of the prolyl adduct on the H2B peptide, studies were conducted on a dipeptide that mimics the first two N-terminal residues of histone H2B (NH-PE-OH, **compound 2**). This dipeptide was reacted with a low molar excess of HCHO (10-fold) at pH 10 for 12 hours at ambient temperature, which resulted in formation of an analogous methylene-containing adduct (as determined by NMR analysis). This adduct was then isolated from the mixture using HPLC before being redissolved in D_2O and subjected to full NMR characterisation. These analyses revealed the methylene bridge to be connected via the prolyl amine and the amido nitrogen of the prolyl-glutamate peptide bond, forming a 5,5-bicycle (Figure 2.5 a, **2a**; Figure 2.18). This assignment was supported by the results of an experiment monitoring reductive methylation of **2** with HCHO and $NaBH_3CN$. In this experiment, an α-amino methylated proline residue was obtained, suggesting that methylene bridge formation occurs via prior formation of an N-terminal iminium ion (Figure 2.29).

Studies were then conducted on dipeptides containing either a 4-membered azetidine (**1**) or 6-membered piperidine (**3**), or alanine (**4**) in place of proline. The structures of any HCHO-derived adducts with these dipeptides were then elucidated by 2D NMR analyses after their isolation by HPLC. Fused bicyclic adducts were observed with both the piperidine-containing **3** (**3a**, Figures 2.5 a, 2.17) and the azetidine-containing **1** (**1a**, Figures 2.5 a, 2.19); however, the 4,5-bicycle derived from **1a** was unstable and degraded to form a hemiaminal (**1b**, Figures 2.5 a, 2.20). Alanine-containing **4** formed an adduct that contained two HCHO-derived methylene groups – one cyclic aminal and one N-hydroxymethyl group connected off the N-terminus (**4a**, Figures 2.5 a, 2.21).

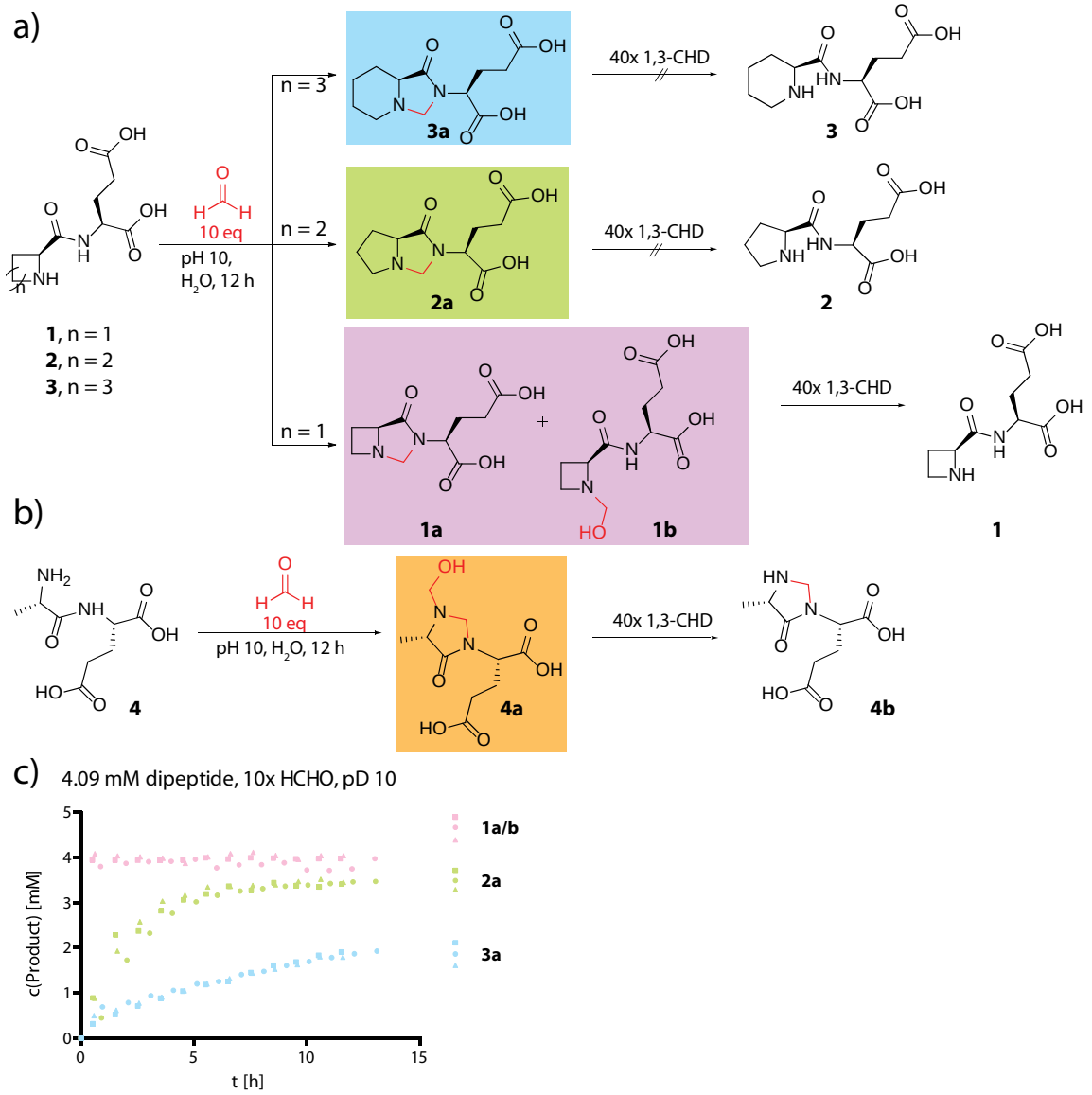


Figure 2.5 | HCHO reacts to form a methylene bridge between the N-terminal nitrogen and the adjacent peptide bond nitrogen. **a)** PipGlu (n=3, **3**), ProGlu (n=2, **2**) and AzeGlu (n=1, **1**) reactions with HCHO. **b)** AlaGlu (**4**) reaction with HCHO. Reactions were carried out with a 10-fold excess HCHO at ambient temperature for 12 h. Substructures that result from the reaction(s) with HCHO are highlighted in red (**1a+b**, **2a**, **3a**, **4a+b**). Adding a 40-fold excess of 1,3-cyclohexandione (1,3-CHD) to the dipeptide-HCHO adducts (**1a+b**, **2a**, **3a**, **4a**) results in removal of the HCHO-derived methylene bridge in **1a+b** and in removal of the hemiaminal in **4a**; by contrast the aminal bridges of **4a**, **2a** and **3a** are stable to treatment with 1,3-CHD. **c)** NMR time-course (400 MHz) monitoring the reaction of a 10-fold excess HCHO with 4.09 mM dipeptides (PipGlu, **3**; ProGlu, **2**; AzeGlu **1**) at pD 10 (n=3 independent replicates). Note **4** did not react to give an observed product under these conditions (Figure 2.27d).

To assess the stability of the dipeptide-derived HCHO adducts (**1a+b**, **2a**, **3a**, **4a**), competition experiments were performed. The dipeptides were each reacted with a 10-fold excess HCHO at pD 10

overnight before a 40-fold excess of the HCHO scavenger 1,3-cyclohexanedione (1,3-CHD) was added. The mixtures were then monitored by 2D NMR. **2a** and **3a** were stable in the mixtures throughout the analysis period (22 hours, Figures 2.5 a, 2.22, 2.23); however, **1a+b** underwent degradation, which correlated with loss of the HCHO-derived methylene bridges of **1a+b** (Figures 2.5 a, 2.24). The methylene bridge of **4a** was stable over the analysis period but the hemiaminal group of **4a** disappeared, leading to formation of **4b** (Figures 2.5 a, 2.25).

Similar 1,3-CHD competition studies with the H2B pentapeptide-derived adduct revealed no degradation of its methylene bridge (Figure 2.26 a, 1.). The stability of **2a** was then further tested in competition experiments with cysteine and glutathione, which are biologically relevant HCHO scavengers. In these studies, **2a** was stable in the presence of either scavenger, even when the scavengers were present at a 40-fold excess (Figure 2.26 a, 3.-5.). **2a** was also stable across a physiological pH range of pH 5.9, 7.4 and 8.5 (pD 6.3, 7.8 and 8.9, Figure 2.26 a, 6.-8.). The kinetics of the dipeptide reactions with HCHO were then profiled. Each dipeptide was reacted with a 10-fold excess HCHO at pD 10 and the progress of each reaction was monitored over time by ^{1}H NMR (Figures 2.5 c, 2.27). **1** reacted most proficiently (the reaction was too fast to determine the reaction rate exactly, but the formation rate of **1a** is at least 1.734 µM/s), followed by **2** (**2a** formation rate 0.235 µM/s) and **3** (**3a** formation rate 0.054 µM/s); however, **4** did not react under these conditions (Figure 2.27 d). Taken together, these results indicate that **2a** is chemically stable under physiologically relevant conditions, which suggests adducts of this type can persist in cells and can potentially have biological functions.

The reactions of **2** with other biologically relevant carbonyl compounds were then profiled. Samples were prepared containing **2** and a 100-fold excesses of either pyruvate, 2OG, acetone, glyoxylic acid or acetaldehyde (CH_3CHO); the samples were then left to react for 12 hours before HPLC purification and analysis by ^{1}H NMR. No reactivity was observed with any carbonyl except for acetaldehyde at 37 °C and pH 10 (Figure 2.6 a+c). LC-MS analyses suggest that the acetaldehyde-derived adduct degrades during isolation (Figure 2.6 b). Therefore, these experiments suggest that N-terminal prolines selectively react with HCHO over some biologically relevant carbonyl compounds, or at least to form stable adducts.

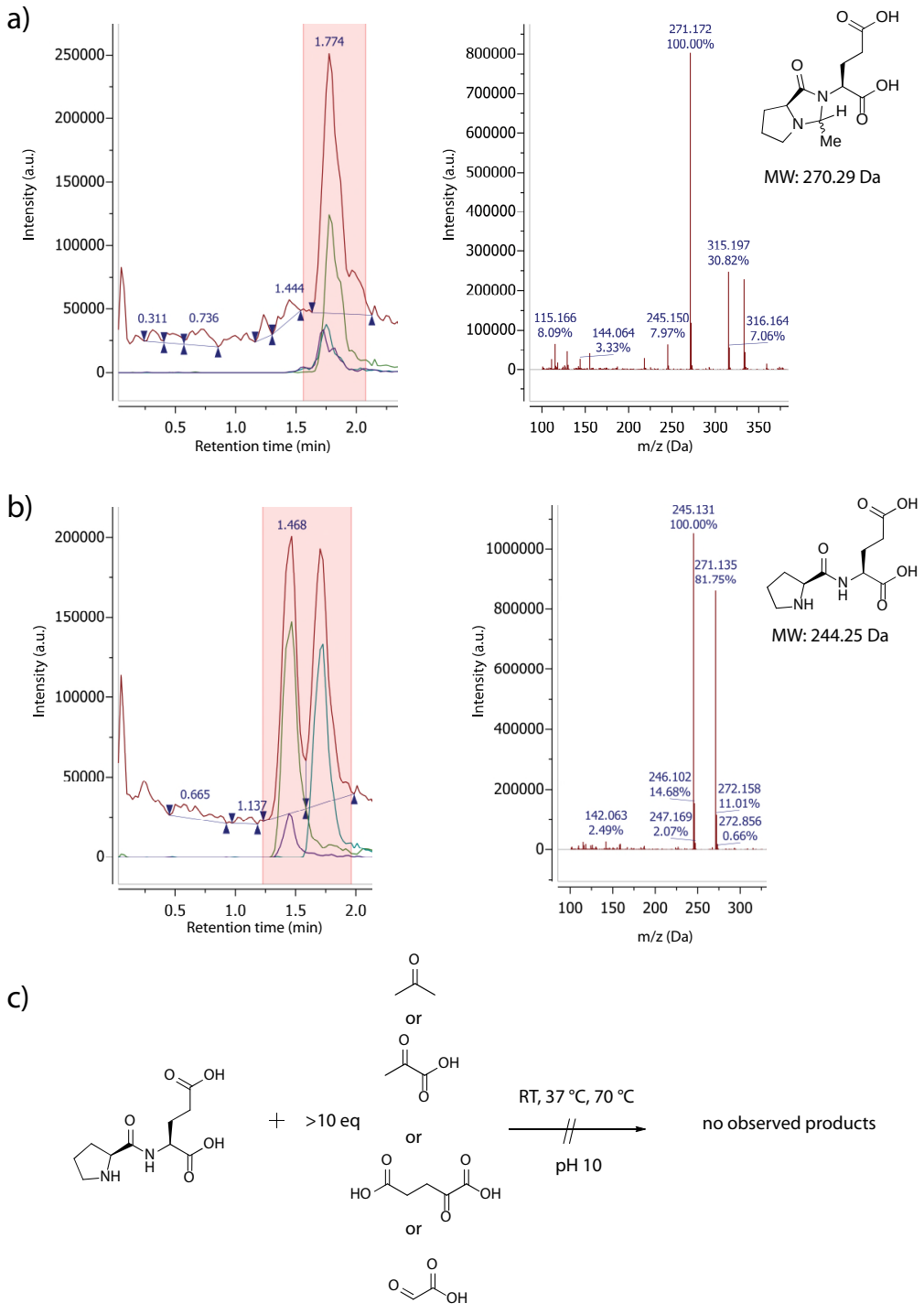


Figure 2.6 | Reactions of the H2B dipeptide with biologically relevant carbonyl compounds. **a)** LC-MS trace of the HPLC fraction containing the product generated by the reaction of a 100-fold excess acetaldehyde with **2** at 37 °C (left graph) showing full conversion of **2** to form an acetaldehyde adduct (m/z=271 Da, right graph). **b)** LC-MS trace (left graph) upon lyophilization reveals the adduct is not stable, partially forming the starting material (**2**) again (m/z=245 Da, right graph). **c)** Excess acetone, pyruvate, 2OG or glyoxylic acid were added to **2** at different temperatures (ambient temperature, 37 °C, and 70 °C), but no conversion was seen as determined by LC-MS (data not shown), suggesting these carbonyl compounds do not react with N-terminal prolines.

It was then tested whether HCHO released during demethylase catalysis can lead to modification of H2B. MALDI MS analyses were conducted on samples containing the 15mer H2B peptide and the demethylase KDM4E (KDM4E was prepared by Eidarus Salah), which catalyses the demethylation of methylated lysine residues on histone H3. Analysis of a trimethylated H3 peptide substrate ($H3K9Me_3$) in the mixture revealed KDM4E-dependent demethylation; in addition, the H2B peptide in the mixture was also observed to undergo modification, forming a +12 Da adduct (Figures 2.7 a+b, 2.28 b+c). Repeating the experiment in the presence of N-oxalylglycine, a reported KDM inhibitor,[237] ablated demethylation of H3 peptide and also prevented formation of the +12 Da adduct (Figure 2.7 c+d). This result implies that the +12 Da modification of the H2B peptide is induced by KDM4E-catalysed production of HCHO. MS/MS studies on recombinant H2B protein were then conducted, for which recombinant H2B was reacted with a 100-fold excess of HCHO for 24 h before undergoing digestion and analysis using an Orbitrap Elite LC-MS/MS mass spectrometer. The MS/MS spectra revealed a +12 Da modification at the N-terminus of the protein (Figure 2.7 e).

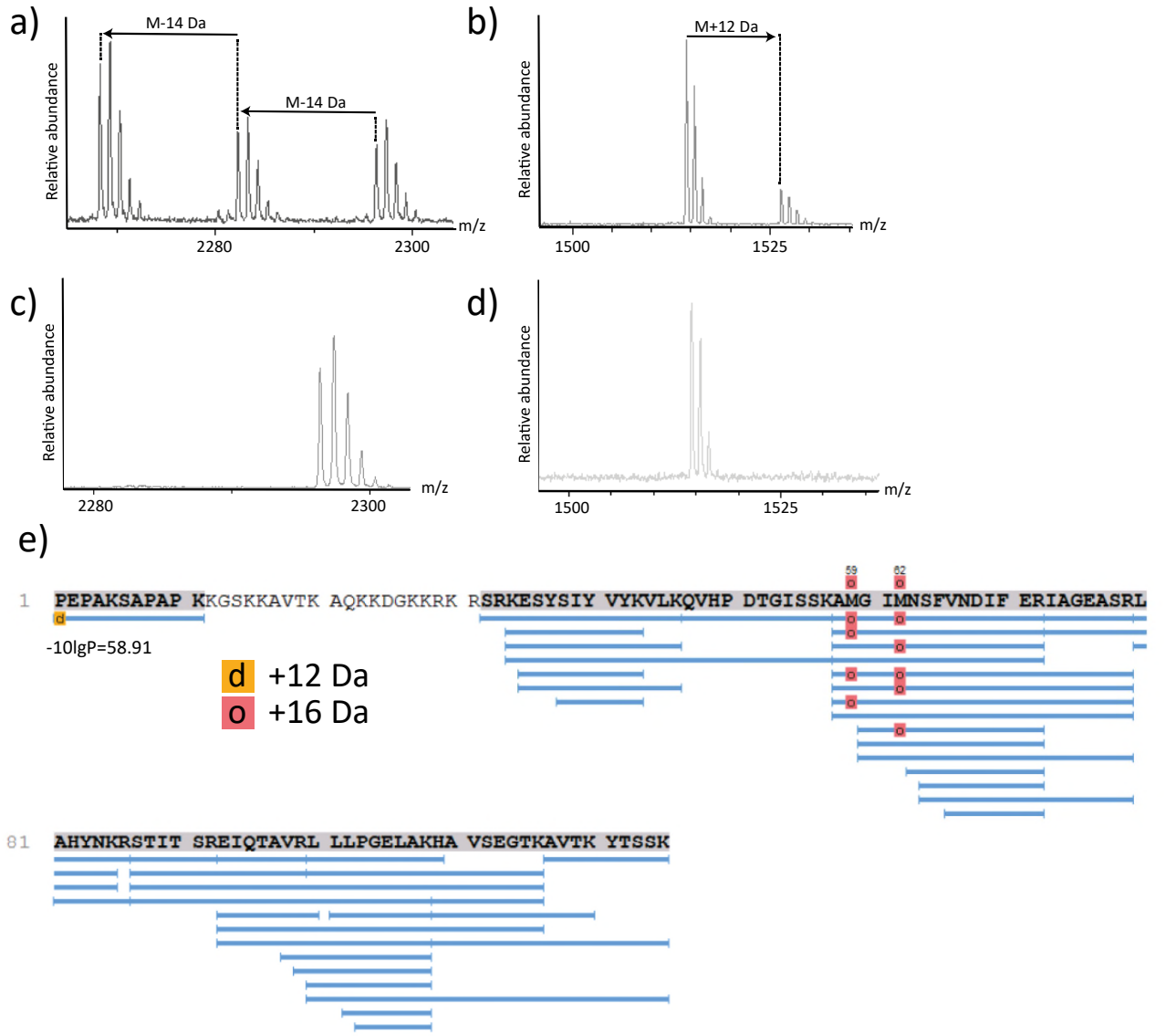


Figure 2.7 | *In situ* generated HCHO reacts with H2B and reaction of recombinant H2B with HCHO. a+b) Demethylation of H3K9Me$_3$ 21mer (2296 Da) to form the dimethylated (2282 Da, M-14 Da) and monomethylated (2268 Da, M-28 Da) products in the presence of 2OG and KDM4E after 24 h incubation, indicative of HCHO formation (a). With H2B present in the reaction mixture of (a), the H2B 15mer reacts with *in situ* generated HCHO (24 h) to form the described +12 Da HCHO-derived adduct (Na^+ adducts of parent peptide [1514 Da] and HCHO-adduct [1526 Da, M+12 Da] are shown) in presence of 2OG, the co-substrate of KDM4E (b). **c+d)** In presence of NOG, a broad spectrum inhibitor of 2OG-dependent enzymes, no demethylation is observed after 24 h (c), suggesting no HCHO is formed under these conditions. Without enzymatic formation of HCHO, no reaction of H2B is observed after 24 h (d). **e)** Recombinant H2B protein was reacted with a 100-fold excess HCHO for 24 h and then subjected to trypsin digestion, followed by LC-MS/MS analyses. The results show that the N-terminal proline bears a +12 Da shift with a –10logP>20. MS/MS samples were run and analysed by Elisabete Pires. KDM4E was prepared by Eidarus Salah.

Studies then investigated as to whether **2** can remove HCHO-derived methylene bridges from thioproline, metampicillin or *S*-(hydroxymethyl)glutathione (HMG) and whether **2** can compete with cysteine or glutathione (GSH) for reaction with HCHO. These studies are different to those monitoring adduct stability of **2a** (see above, Figures 2.5 a, 2.23, 2.26) where the adduct was formed prior to addition of the scavenger. Addition of **2** to pre-reacted samples of cysteine, metampicillin or glutathione with HCHO did not reveal formation of **2a** (Figures 2.8 a-c, 2.30). NMR analyses of the reaction of HCHO with **2** competing with cysteine or GSH revealed no significant formation of **2a**, while formation of cysteine- or glutathione-derived HCHO adducts were observed (Figure 2.31). Therefore, whilst compound **2a** is chemically stable, the thiol groups of cysteine and glutathione appear to react more efficiently with HCHO.[82] These results suggest that high local levels of HCHO might be required for N-terminal proline residue to react with HCHO *in cellulo*; however, high local HCHO concentrations are reported during transcriptional reprogramming and may also occur in other contexts.[115] Given that the bacterial transcription factor FrmR induces the HCHO detoxification response upon reaction of its N-terminal proline with HCHO, it is likely that surges of HCHO are relevant in cells.[152]

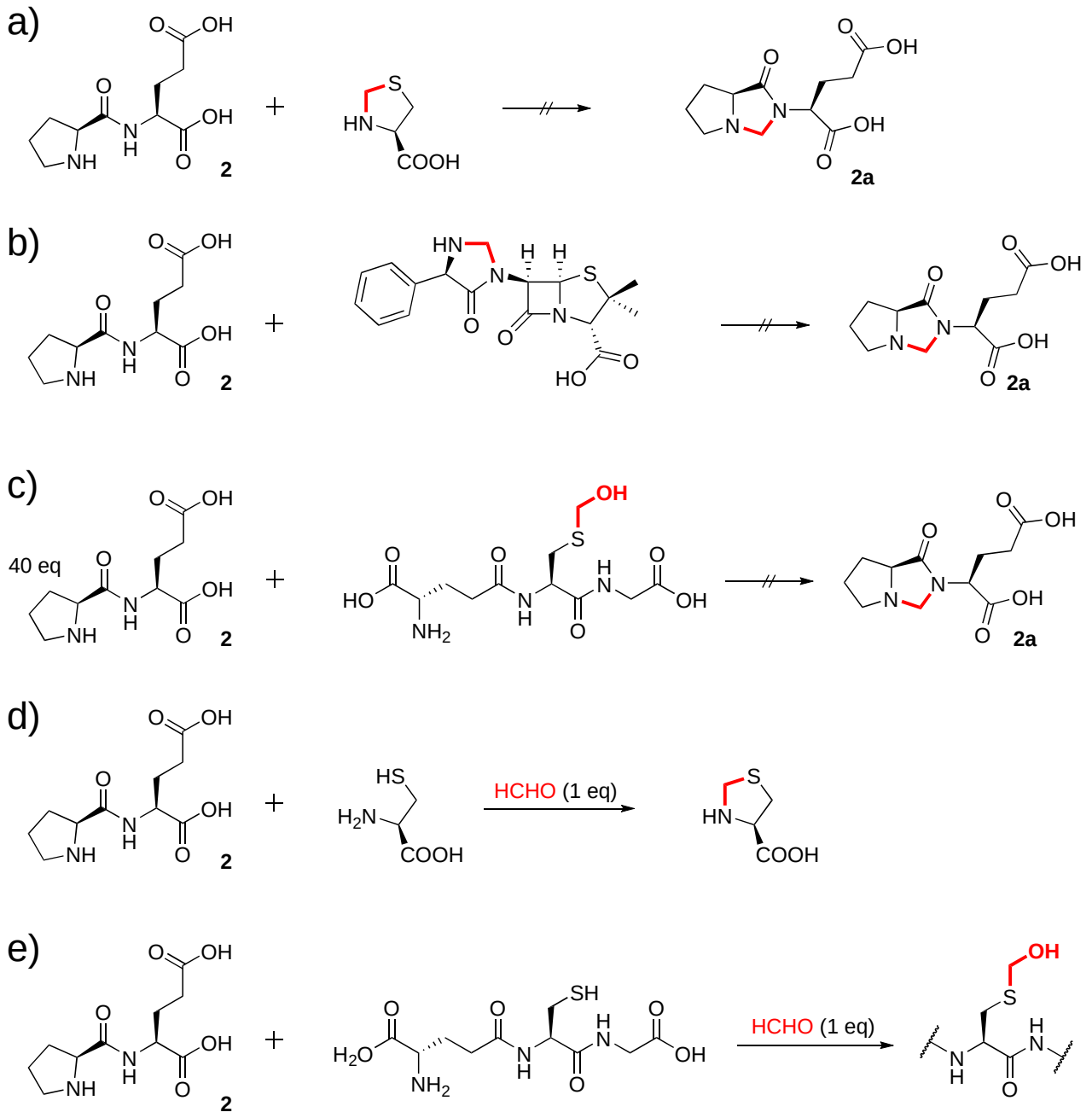


Figure 2.8 | Thiol groups react faster with HCHO than compound 2. a+b) Equimolar amounts of **2** does not remove HCHO-derived methylene bridges from thioproline (a) or metampicillin (b). **c)** A 40-fold excess of **2** does not remove the HCHO-derived methylene bridge of HMG. **d+e)** Cysteine (d) and glutathione (e) react with HCHO to form thioproline and *S*-(hydroxymethyl)glutathione (HMG), respectively, in the presence of **2**.

To prevent yeast, mould and microbial growth in cosmetic products, preservatives are used. One class of preservatives comprises HCHO-releasing agents (or HCHO donors) that degrade over time to release HCHO. It was determined whether HCHO-releasing agents used in the formulation of ointments, creams and baby wipes[238] can induce adduct formation. HCHO-releasing agents are reported to be efficient antimicrobial preservatives; however, health concerns have been raised over their effects on skin irritation and their links to cancer and asthma initiation.[239, 240] Allergic contact dermatitis induced by HCHO is often chronic, likely because of the fact that HCHO is one of the most common indoor pollutants and hence difficult to avoid.[58] The reactivities of the HCHO donors diazolidinyl urea (Figure 2.9 e) and imidazolidinyl urea (Figure 2.9 f) were tested in MALDI MS experiments with the H2B 15mer peptide. A +12 Da shift on the H2B peptide was observed in the samples after 8 hours incubation with 100-fold excess HCHO, which suggests that they can initiate adduct formation via HCHO release. Diazolidinyl urea showed slightly higher reactivity with the H2B peptide than imidazolidinyl urea (Figure 2.9 a+b), which correlates with their reportedly different HCHO release rates in water.[241] MS/MS analyses confirmed that the modification occurs at the N-terminal proline (Figure 2.9 c+d), as observed with HCHO itself (Figure 2.3 f).

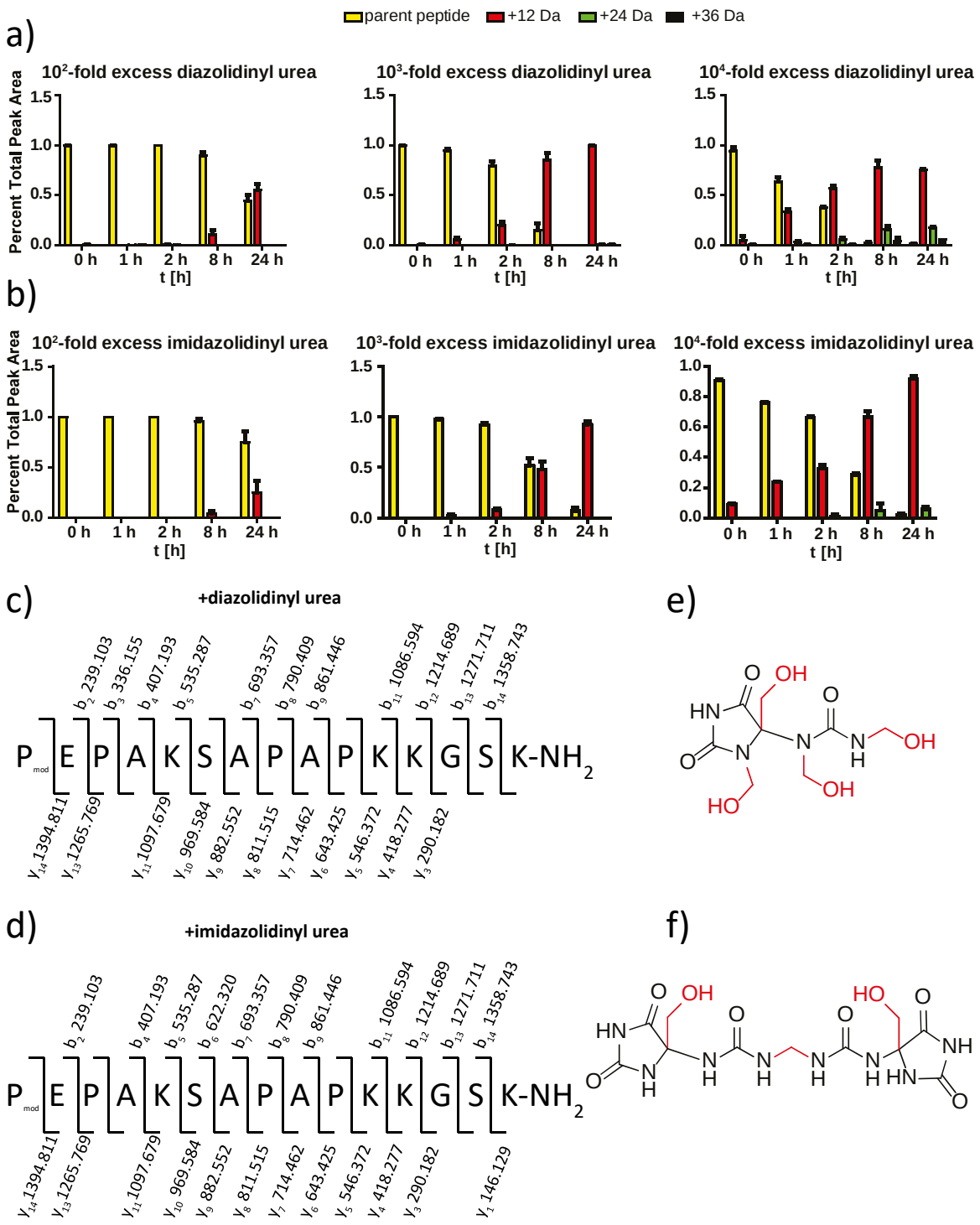


Figure 2.9 | The HCHO donors diazolidinyl urea and imidazolidinyl urea result in formation of an N-terminal HCHO adduct with the H2B 15mer. a+b) Diazolidinyl urea (a) or imidazolidinyl urea (b) were reacted (10- to 10^4-fold excess) with the H2B 15mer peptide (pH 7.4, ambient temperature). Reactions were monitored by MALDI MS over 24 h. Errors: SD of the mean (n=3, technical repeats). **c+d)** MALDI MS/MS results summary based on reaction of H2B 15mer after treatment with 1000-fold excess diazolidinyl urea (c) or imidazolidinyl urea (d) for 12 h. **e)** Structure of diazolidinyl urea as revised by Lehmann et al.[242] **f)** Structure of imidazolidinyl urea as revised by Lehmann et al.[242]

Given that N-terminal amino acids and their modifications are often crucial for protein stability,[243] it was then assessed as to whether HCHO affects the turnover rate of H2B protein in HEK293T cells. For these experiments, cycloheximide (CHX) was used to inhibit protein synthesis, which allowed for analysis of protein degradation without having the bias of *de novo* protein synthesis. While HCHO treatment consistently led to higher H2B protein levels after prolonged CHX treatment (>10 hours, Figure 2.10 a-c), the same was also true for H3 and H2A (Figure 2.10 c). RNA-Seq analyses also revealed an upregulation of H2B in HEK293T cells treated with HCHO (0.37-fold less in untreated cells, q=0.043) (Figure 2.10 e), but also in H2A (0.33-fold less in untreated cells, q=0.043) and H3 (0.42-fold less in untreated cells, q=0.034); RNA-Seq raw data were obtained from the Oxford Genomics Centre. Another protein bearing an N-terminal proline, LTA4H, which is not related to histones, was then tested. These studies revealed no evidence for a HCHO-dependent effect on LTA4H stability (Figure 2.10 c+d). Taken together, the cellular results indicate that the stability of histone H2B and LTA4H are not selectively affected by reactions with HCHO on their N-terminal prolines, at least under the tested conditions. However, there may be other HCHO-mediated mechanisms promoting their biosynthesis and/or stability.

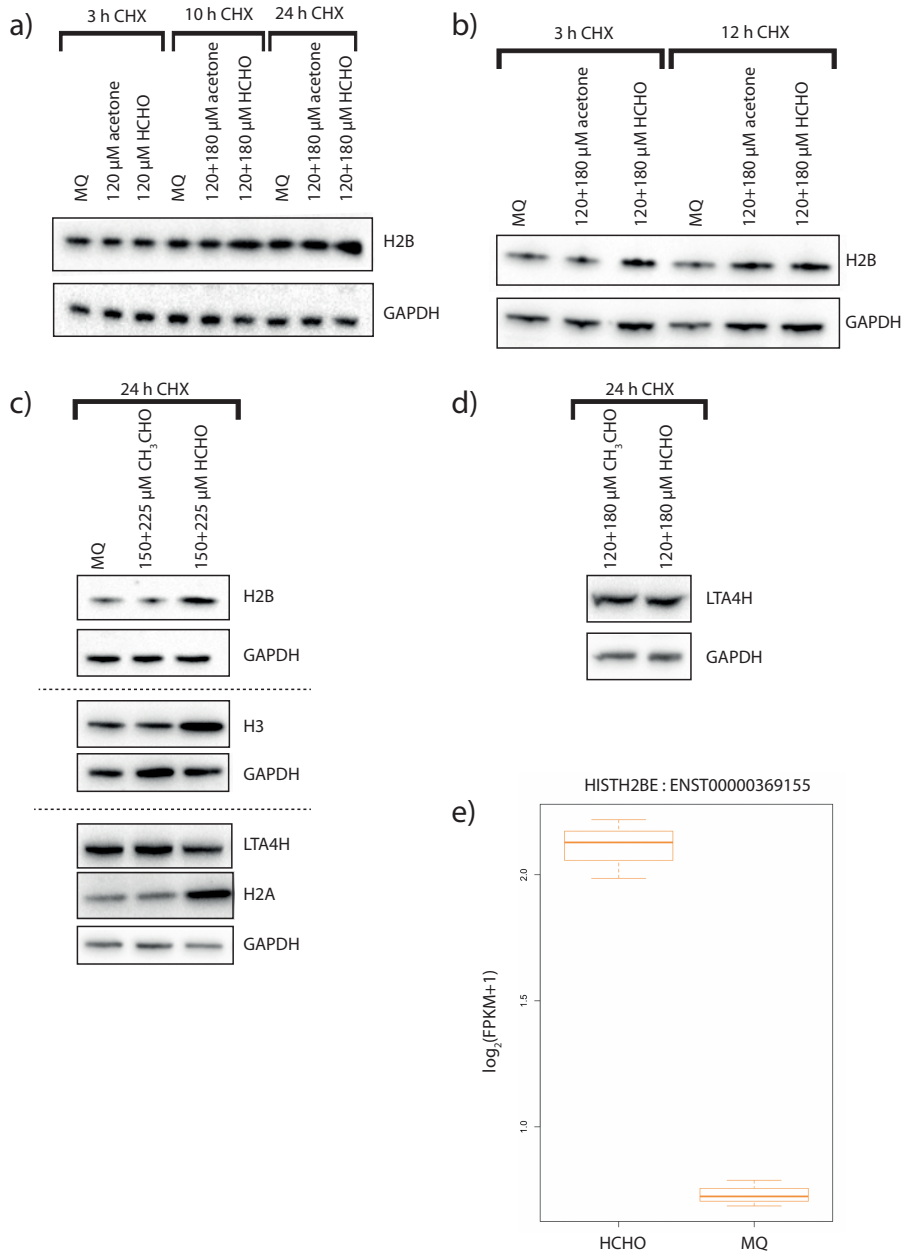


Figure 2.10 | Stability of proteins with an N-terminal proline in the presence of HCHO and RNA-Seq analysis. a) HEK293T cells were treated with 120 µM HCHO (or acetone, or MQ water control; 100 mM stock solutions of carbonyl compounds in MQ water) for 14 h. The cells were then treated with 100 µM CHX (from a 100 mM DMSO stock solution) and were harvested 3 h later. Additionally, HEK293T cells were treated with 120 µM HCHO (or acetone, or MQ water control) for 14 h, followed by addition of CHX. 7.5 h after the CHX treatment, a further treatment of 180 µM carbonyl compound was added. The cells were then harvested 10 h and 24 h after the CHX addition. **b)** HEK293T cells were treated with MQ water/acetone/HCHO overnight, then treated with both the next dose of MQ water/acetone/HCHO and CHX; the cells were then harvested 3 h or 12 h after the second dose. **c+d)** HEK293T cells were treated with MQ water/acetone/HCHO overnight and later treated with both the next dose of MQ water/acetone/HCHO and with CHX; cells were harvested 24 h after the second dose. The final concentration of CHX was 100 µM in all experiments. **e)** H2B transcript levels in HEK293T cells treated with HCHO (150+225 µM) or untreated cells. Errors: SD of the mean (n=3). RNA-Seq raw data were obtained from the Oxford Genomics Centre.

Having identified HCHO-mediated reactions on histone H2B, efforts were then focused on studying 4-oxalocrotonate tautomerase (4-OT) from *Pseudomonas putida*. 4-OT contains an N-terminal proline residue and catalyses the tautomerisation of the non-conjugated 4*E*-2-oxo-hexenedioate (4-oxalocrotonate) to the conjugated 3*E*-2-oxo-hexenedioate (Figure 2.1 b).[217] 4-OT activity was investigated by employing NMR- and absorbance-based assays that monitor the promiscuous 4-OT-catalysed Michael addition of *trans*-β-nitrostyrene and acetaldehyde (CH_3CHO, Figure 2.11 a).[220, 221] Importantly, the N-terminal proline of 4-OT is proposed to react the CH_3CHO to form an enamine intermediate, which then reacts with *trans*-β-nitrostyrene to form a Michael addition product (Figure 2.1 c);[221] the N-terminal proline is therefore essential for 4-OT activity.

Under the standard conditions, 4-OT completely converted CH_3CHO and *trans*-β-nitrostyrene to the product 4-nitro-3-phenylbutanal within 14 min, as determined by ^{1}H NMR analysis (Figure 2.11 b); however, when 4-OT was pre-incubated with a 200-fold excess of HCHO, a significant reduction in the rate of reaction of observed (Figures 2.11 c+d).

The kinetic data underlying the 4-OT catalysed Michael addition of CH_3CHO and *trans*-β-nitrostyrene were then determined by absorbance (K_m=281 µM ± 42 for *trans*-β-nitrostyrene, K_m=74 mM ± 4.8 for CH_3CHO, Figure 2.33 a-d). It was then studied whether or not the substrate CH_3CHO competes with HCHO for reaction at the N-terminal proline. Samples were prepared containing 4-OT incubated with a fixed 200-fold molar excess of HCHO and a range of CH_3CHO concentrations; depletion of *trans*-β-nitrostyrene was then monitored over time by absorbance. With increasing CH_3CHO concentrations, HCHO-dependent inhibition was less significant, which indicates that CH_3CHO and HCHO compete at the same site (Figure 2.33 e).

Given that HCHO forms the previously described 5,5-bicycle over time, it was hypothesised that longer pre-incubation times of 4-OT with HCHO correlate with higher inhibition. Interestingly, absorbance assays show only a very mild increase in inhibition with increasing pre-incubation time of HCHO with 4-OT (Figure 2.11 e), although MALDI MS data show an increase of a +24 Da species over time, suggesting the formation of two methylene bridges (Figure 2.34 a). Circular dichroism (CD) spectra show no alterations in the secondary structure of 4-OT when preincubated with HCHO (Figure 2.34 b). The +12 Da modification on the N-terminal proline of 4-OT was confirmed by MS/MS (Figure 2.12). Given that increasing the pre-incubation time increases the formation of the +24 Da species, but only mildly increases inhibition, it was hypothesised that the inhibitory species is possibly not the 5,5-bicycle, but a hemiaminal (Figure 2.29). Hemiaminals are intermediates in imine/iminium formation (Figure 2.29, Figure 2.35 a) and are more dynamic, which suggests they are more likely to be affected by the CH_3CHO concentration. Using HCHO and $NaCNBH_3$, the formation of an N-terminal *N*-methyl-prolyl, resulting in a +14 Da adduct (Figure 2.35), was observed to form much faster as compared to the formation of the +12 Da aminal methylene bridge (-N-CH_2-N-) with HCHO only (Figure 2.3 b); this importantly implies that hemiaminal formation is likely the biologically relevant species.

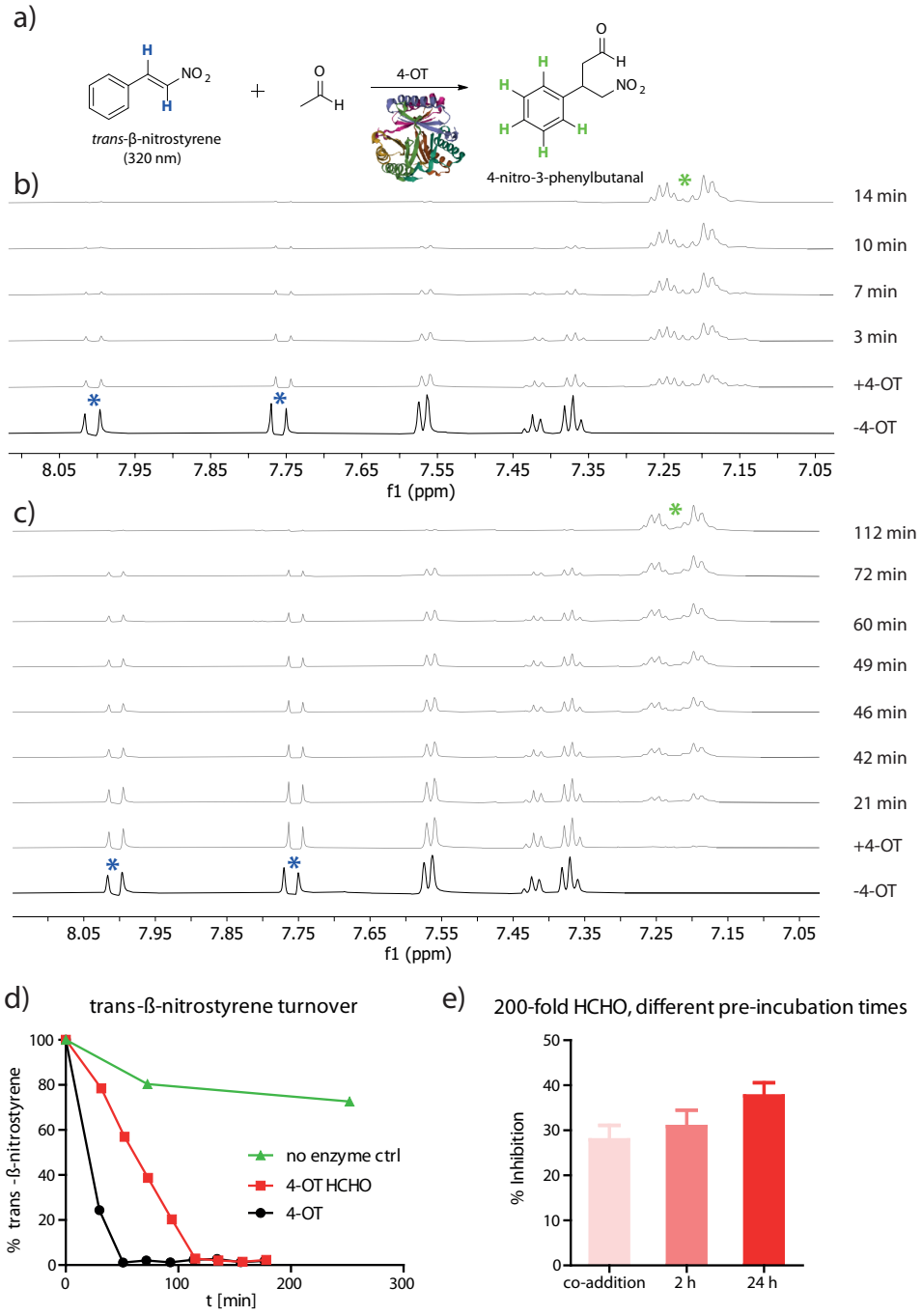
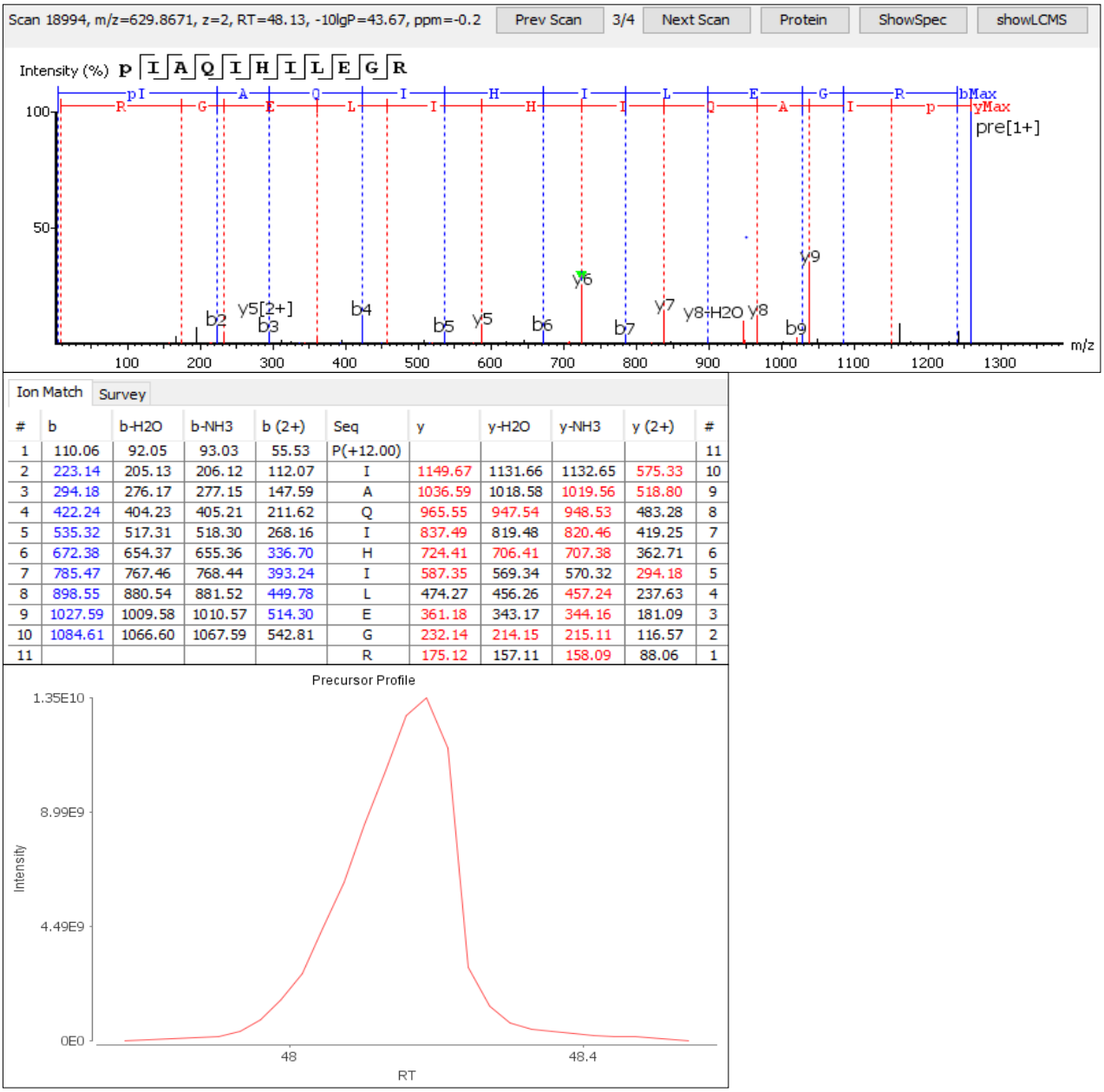


Figure 2.11 | HCHO inhibits 4-OT *in vitro*. a) Scheme showing the 4-OT catalysed (pdb 4X19) Michael addition of CH_3CHO to *trans*-β-nitrostyrene, resulting in 4-nitro-3-phenylbutanal. Double bond hydrogen atoms (blue) and aromatic product hydrogen atoms (green) were used to monitor the reaction by NMR. **b)** ^{1}H NMR (700 MHz) showing complete conversion to 4-nitro-3-phenylbutanal (green asterisk) upon addition of 4-OT (14 min). **c)** Complete product formation is only observed after about 2 h with 4-OT preincubated for 17 h with a 200-fold excess of HCHO (^{1}H NMR (700 MHz)). **d)** Independent replicate of (b+c), showing evidence for slow non-enzymatic depletion of *trans*-β-nitrostyrene (^{1}H NMR (700 MHz)). The NMR signals highlighted with a blue and green asterisk in (b) and (c) were used to generate graph. **e)** Assay monitoring the absorbance of *trans*-β-nitrostyrene at 320 nm (pH 7.4), showing a mild increase in inhibition correlates with prolonged pre-incubation time of 4-OT with HCHO. Errors: SD of the mean (n=2). Similar results were obtained at pH 7.8 (Figure 2.33 f).



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	110.06	92.05	93.03	55.53	P(+12.00)					11
2	223.14	205.13	206.12	112.07	I	1149.67	1131.66	1132.65	575.33	10
3	294.18	276.17	277.15	147.59	A	1036.59	1018.58	1019.56	518.80	9
4	422.24	404.23	405.21	211.62	Q	965.55	947.54	948.53	483.28	8
5	535.32	517.31	518.30	268.16	I	837.49	819.48	820.46	419.25	7
6	672.38	654.37	655.36	336.70	H	724.41	706.41	707.38	362.71	6
7	785.47	767.46	768.44	393.24	I	587.35	569.34	570.32	294.18	5
8	898.55	880.54	881.52	449.78	L	474.27	456.26	457.24	237.63	4
9	1027.59	1009.58	1010.57	514.30	E	361.18	343.17	344.16	181.09	3
10	1084.61	1066.60	1067.59	542.81	G	232.14	214.15	215.11	116.57	2
11					R	175.12	157.11	158.09	88.06	1

Figure 2.12 | Recombinant 4-OT reacts with HCHO to form an N-terminal +12 Da adduct. 4-OT was incubated with HCHO (200-fold excess, 37 °C, 16 h in 20 mM potassium phosphate, pH 7.4) before being subjected to trypsin digestion and LUMOS LC-MS/MS analyses. LUMOS LC-MS/MS was run and analysed by Svenja S. Hester.

It was then investigated whether HCHO-dependent inhibition of 4-OT can occur in bacterial cells. BL21(DE3) cells overexpressing 4-OT (Figure 2.13 a-c)[244] were treated with *trans*-β-nitrostyrene and CH_3CHO; the 4-OT catalysed reaction was then monitored by the absorbance signal of *trans*-β-nitrostyrene. The cells depleted *trans*-β-nitrostyrene within 270 min (Figure 2.13 a), whereas very slow depletion of *trans*-β-nitrostyrene was observed in the control cells containing empty vector (EV), or in a sample without cells. The low-level depletion observed in these controls is likely due to chemical degradation and/or reaction of *trans*-β-nitrostyrene with cellular nucleophiles. Strikingly, however, HCHO was observed to inhibit 4-OT activity in the cells in a dose-dependent manner (Figure 2.13 a+b). The amounts of HCHO resulting in cellular 4-OT inhibition did not (350 μM HCHO) or only mildly (700 μM HCHO) hamper cellular proliferation (Figure 2.13 d)

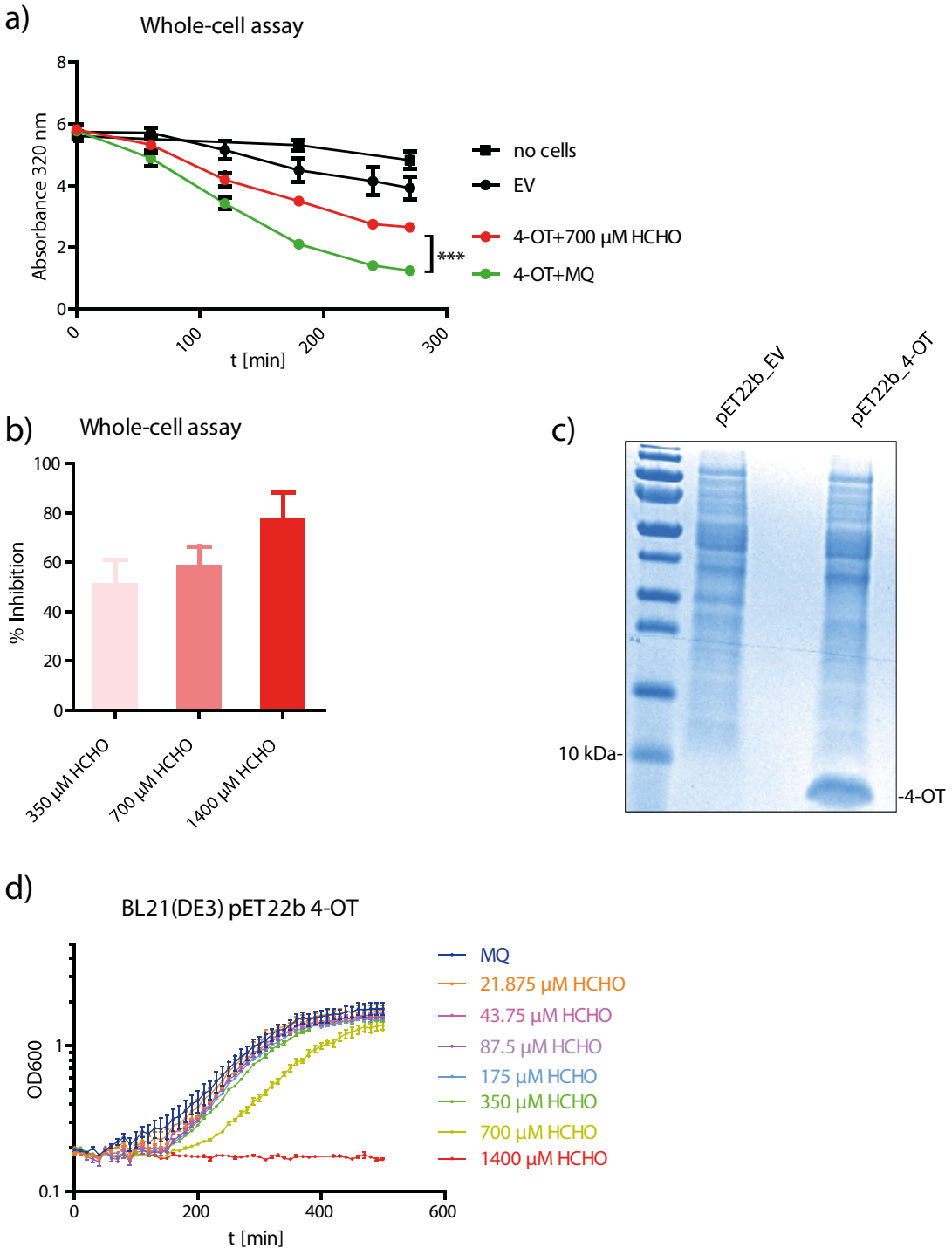
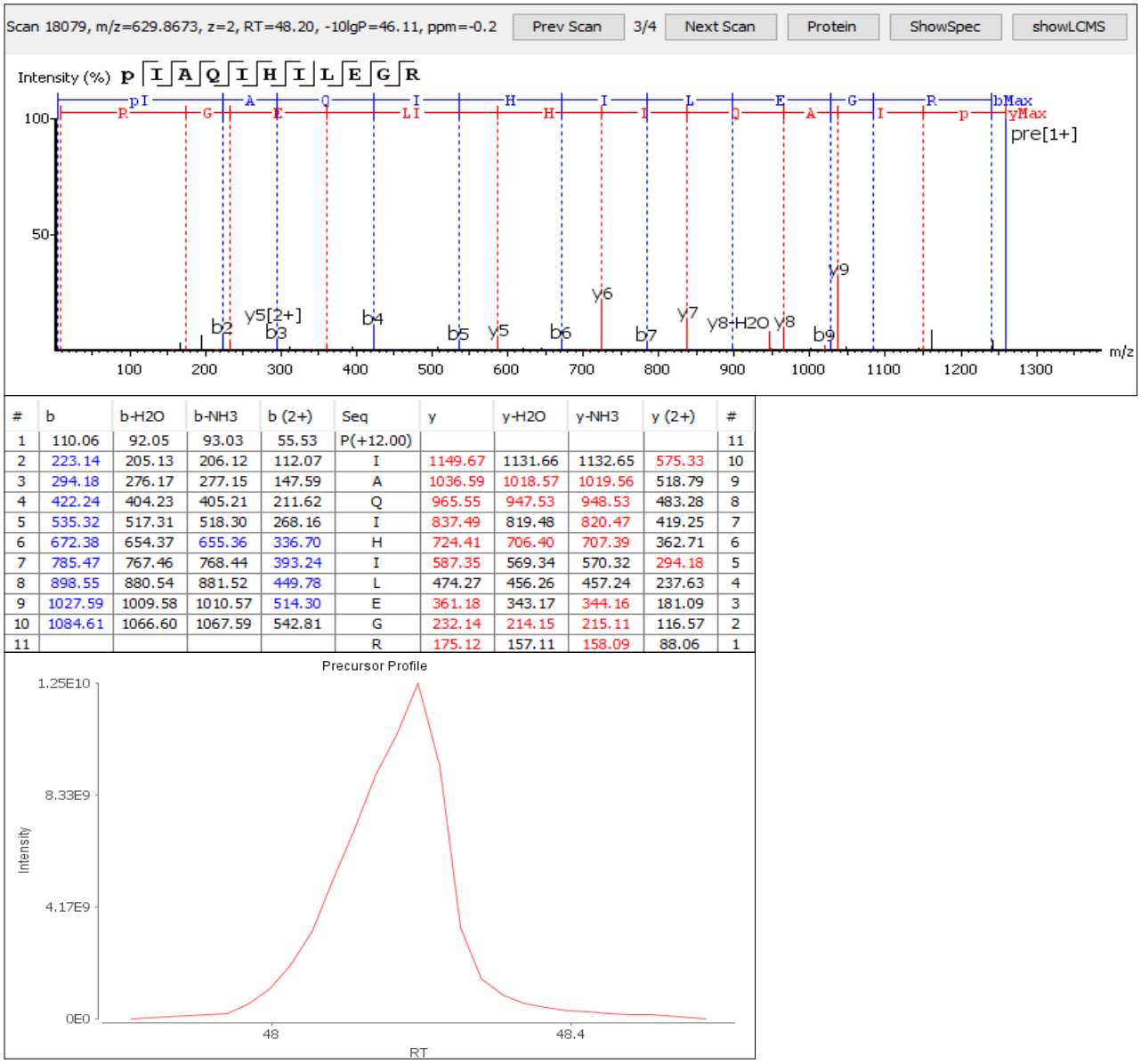


Figure 2.13 | HCHO inhibits 4-OT catalysis in *E. coli* cells. a) CH_3CHO and *trans*-β-nitrostyrene were added to BL21(DE3) cells transformed with either pET-22b-4-OT (Milli-Q (MQ) water control in green, +HCHO in red), pET-22b-EV (emtpy vector, black circle) or buffer only (black square). The depletion of *trans*-β-nitrostyrene was monitored by absorbance at 320 nm. Errors: SD of the mean (n=3, each replicate in duplicate, statistical analysis: 1way ANOVA). **b)** The same assay as in (a) was used to determine the percentage inhibition of 4-OT with different amounts of HCHO. Errors: SEM (n=3, each replicate in duplicate). **c)** SDS-PAGE gel of whole cell lysates confirming the overexpression of 4-OT at the beginning of the whole-cell assay. **d)** Growth curve showing the effects of HCHO to BL21(DE3) cell growth. Errors: SEM (n=3 technical repeats).

In bacteria overexpressing 4-OT, HCHO treatment resulted in formation of the +12 Da modification on the N-terminal proline of 4-OT, as observed by in-gel digestion (Figure 2.36) of 4-OT followed by MS/MS (Figure 2.14). Overall, these studies suggest that HCHO can regulate the activity of 4-OT in bacteria by reacting with its N-terminal proline to form hemiaminals, which might react over time with the adjacent peptide bond nitrogen to form stable aminals.



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	110.06	92.05	93.03	55.53	P(+12.00)					11
2	223.14	205.13	206.12	112.07	I	1149.67	1131.66	1132.65	575.33	10
3	294.18	276.17	277.15	147.59	A	1036.59	1018.57	1019.56	518.79	9
4	422.24	404.23	405.21	211.62	Q	965.55	947.53	948.53	483.28	8
5	535.32	517.31	518.30	268.16	I	837.49	819.48	820.47	419.25	7
6	672.38	654.37	655.36	336.70	H	724.41	706.40	707.39	362.71	6
7	785.47	767.46	768.44	393.24	I	587.35	569.34	570.32	294.18	5
8	898.55	880.54	881.52	449.78	L	474.27	456.26	457.24	237.63	4
9	1027.59	1009.58	1010.57	514.30	E	361.18	343.17	344.16	181.09	3
10	1084.61	1066.60	1067.59	542.81	G	232.14	214.15	215.11	116.57	2
11					R	175.12	157.11	158.09	88.06	1

Figure 2.14 | HCHO reacts with the N-terminal proline of 4-OT *in cellulo*. *E. coli* cells overexpressing 4-OT were treated for 4.5 h with 700 µM HCHO. Proteins were resolved by tricine SDS-PAGE (Figure 2.36), bands corresponding to 4-OT were excised and subjected to trypsin digestion. The spectra obtained by LUMOS LC-MS/MS confirm the presence of the N-terminal +12 Da adduct. LUMOS LC-MS/MS was run and analysed by Svenja S. Hester.

2.4 Discussion

As cells routinely produce HCHO, it is conceivable that stable HCHO adducts with regulatory or sensing roles form over time, with the potential for HCHO being directly involved in regulating chromatin structure and function. A drawback in HCHO research is the lack of established quantification methods for biological samples.[245] Especially subcellular levels of HCHO may be vastly different than generally appreciated. Chromatin is highly dynamic in its methylation state, which probably leads to high local concentrations of HCHO in the nucleus.[115]

The studies on the reactions of HCHO on histone tails reveal that H2B reacts very efficiently with HCHO. The high reactivity is based on its N-terminal proline which forms a 5,5-bicyclic adduct, suggesting any N-terminal proline could be a potential HCHO reservoir. The adduct is chemically very stable as no evidence of instability in presence of excess of various other HCHO scavengers was accrued. Other carbonyl compounds did not exhibit similar reactivity, indicating that HCHO is an exceptional aldehyde with respect to its reactions with biomacromolecules. However, in the presence of GSH or cysteine competing with the N-terminal proline of the H2B derived dipeptide for HCHO, the thiol groups of the small molecules exhibit greater reactivity. These results indicate that cellular HCHO detoxification mechanisms based on GSH must likely be overwhelmed for reactions of HCHO with N-terminal prolines to occur (assuming equal accessibility of the small molecules and proteins to HCHO, which might not always be the case). This result is in line with the finding that the bacterial transcription factor FrmR switches on the HCHO detoxification response upon cross-linking of HCHO with its N-terminal proline and an internal cysteine residue.[152] Nevertheless, it is shown that enzymatically produced HCHO reacts *in situ* with the N-terminal proline of H2B, further underlining the high reactivity of proline towards HCHO even in presence of other biomolecules. Evidence was also presented that the HCHO adduct is formed on recombinant H2B protein. Overall, these findings illustrate the potential of HCHO to form stable adducts that might be involved in downstream biological responses.

Studies with 4-OT were employed to investigate reactions of HCHO with N-terminal prolines. 4-OT

uses its N-terminal catalytic proline in catalysis, and the presented studies indicate that reaction of 4-OT with HCHO inhibits 4-OT, but likely the hemiaminal *en route* to the 5,5-bicycle is the inhibitory species. To the best of our knowledge, this is the first study to show the potential of reversible-covalent species to interfere with biological processes. This greatly highlights the intriguing complexity of the deceptively simple molecule HCHO. Overall, this study reveals that the N-terminal proline of H2B and in extension any N-terminal proline containing protein, 207 of which were found in humans,[233] has the potential to act as sink for HCHO, with consequent potential roles for such proteins in HCHO sensing.

Given that HCHO is used to fumigate rooms for disinfectant purposes,[97, 98] it is of interest to determine the exact mechanism of toxicity, although direct HCHO induced chemical modification of DNA likely plays a role.[83] The catalytic N-terminal proline of the *E. coli* DNA repair enzyme Fpg will likely be susceptible to inhibition by HCHO, which raises the question whether or not inactivation of this DNA repair pathway plays a role in the toxicity of HCHO.

2.5 Supplementary Information

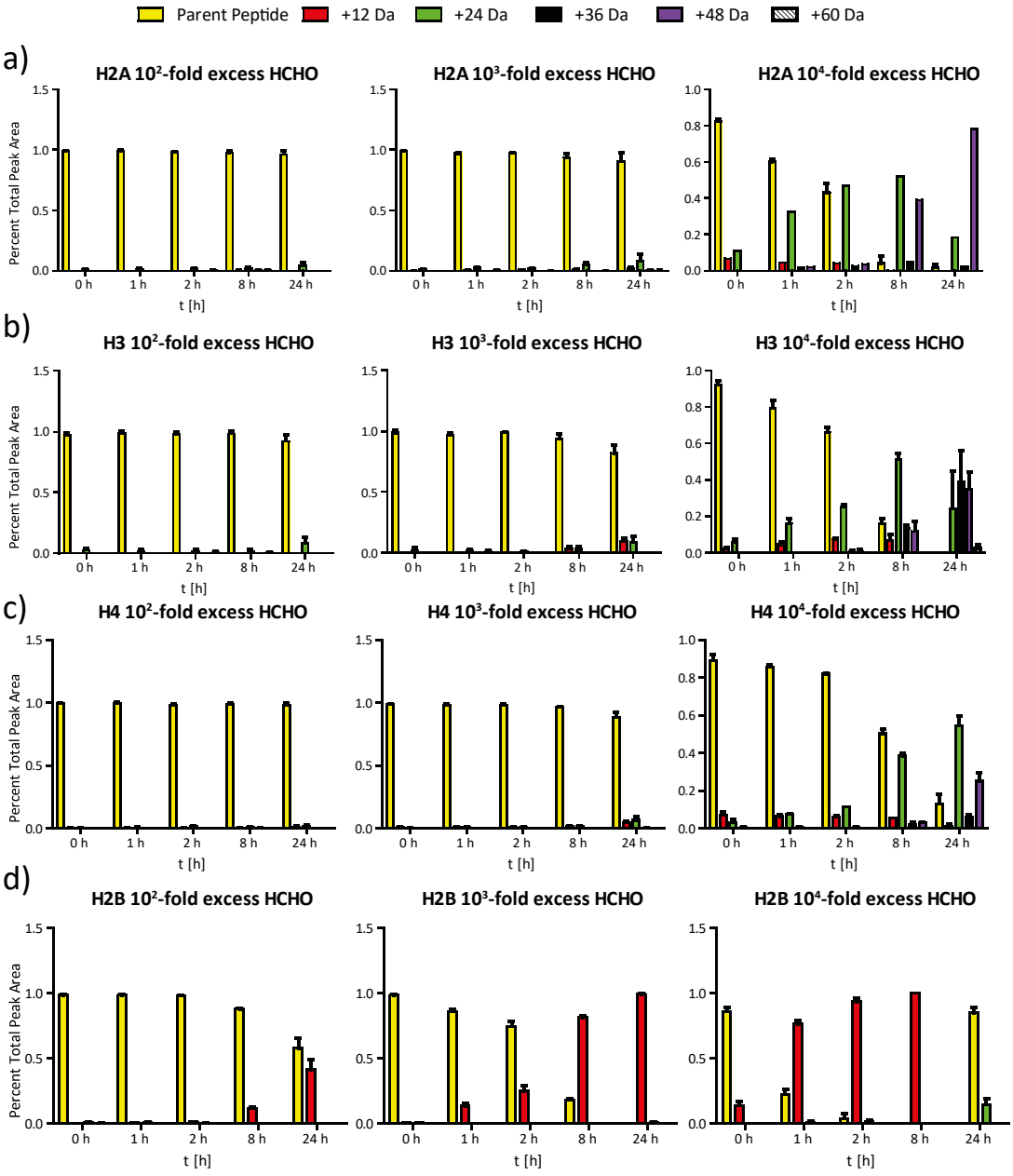


Figure 2.15 | Reaction of histone peptides with formaldehyde (HCHO). Different concentrations of HCHO (10^2- to 10^4-fold excess) were reacted with H2A (a), H3 (b), H4 (c), and H2B (d) (pH 7.4 in 50 mM potassium phosphate, ambient temperature) and monitored by MALDI MS. Errors: SD of the mean (n=3, technical repeats). No adduct formation was observed for any of the peptides with a 10-fold excess or equimolar amounts of HCHO (data not shown). H2A 1-15 (NH_2-SGRGKQGGKARAKAK-NH_2); H3 1-15 (NH_2-ARTKQTARKSTGGKA-NH_2); H4 1-15 (NH_2-SGRGKGGKGLGKGGA-NH_2); H2B 1-15 (NH-PEPAKSAPAPKKGSK-NH_2). Note the +12 Da adduct observed with the H2B peptide corresponds to addition of an N-terminal methylene group, resulting in a fused 5,5-bicycle (Figure 2.5 a).

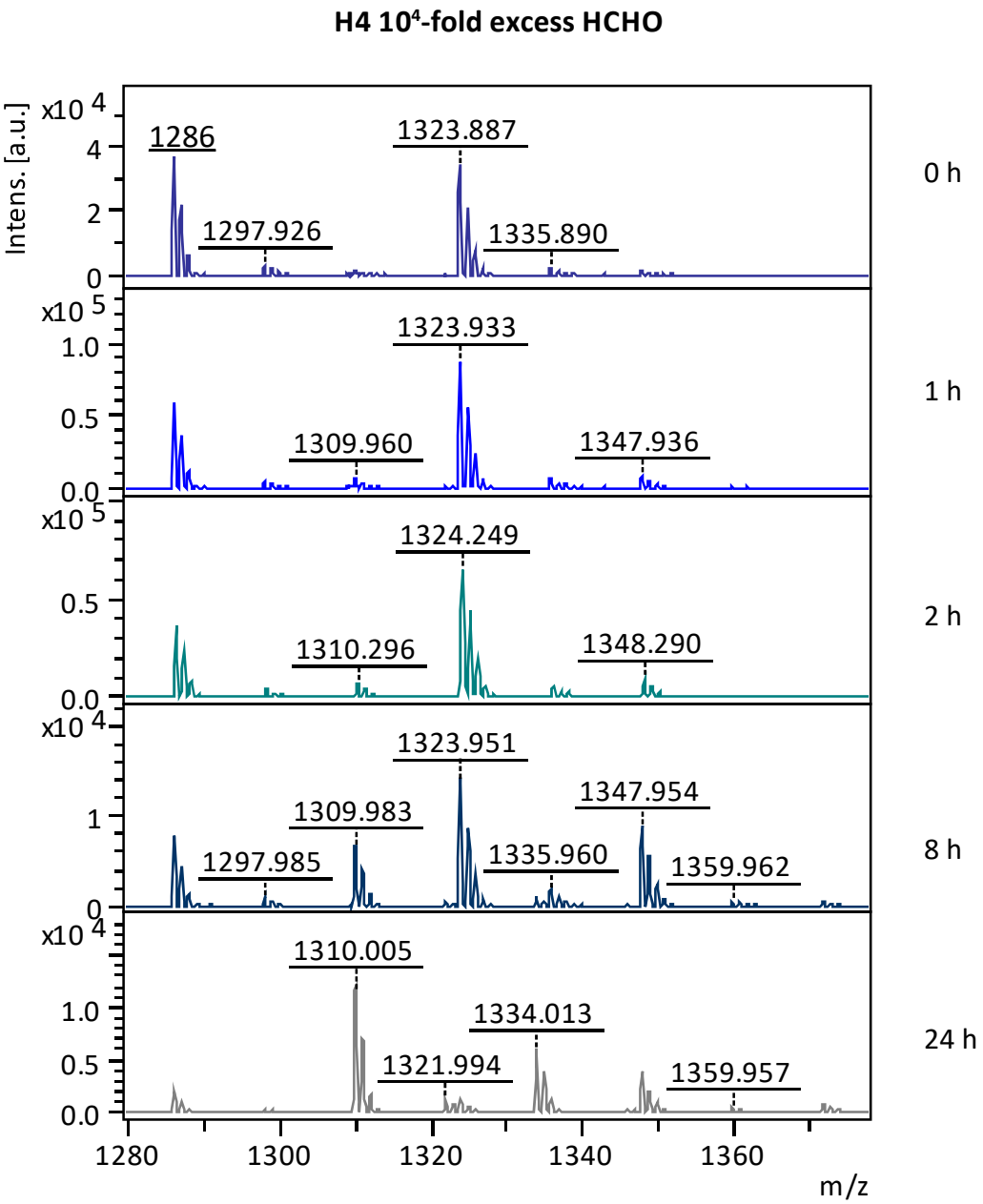


Figure 2.16 | Raw mass spectrometric data obtained from the reaction of HCHO with the H4 15mer peptide. The reaction mixture (2.5 µM peptide concentration, 25 mM HCHO, 50 mM potassium phosphate at pH 7.4) was analysed directly after addition of HCHO (0 h), after 1 h, 2 h, 8 h and 24 h by MALDI MS. Peaks show parent peptide $[M+H]^+$: 1286; parent peptide $[M+K]^+$: 1324 Da; +12 Da HCHO adduct $[M+H]^+$: 1298 Da; +12 Da HCHO adduct $[M+K]^+$: 1336 Da; +24 Da HCHO adduct $[M+H]^+$: 1310 Da; +24 Da HCHO adduct $[M+K]^+$: 1348 Da; +36 Da HCHO adduct $[M+H]^+$: 1322 Da; +36 Da HCHO adduct $[M+K]^+$: 1360 Da; +48 Da HCHO adduct $[M+H]^+$: 1334 Da; +48 Da HCHO adduct $[M+K]^+$: 1372 Da. Areas under curve were used to generate the plot in Figure 2.15, with a 10^4-fold excess of HCHO. For all other plots based on MALDI MS, data processing was executed similarly. H4 1-15 (NH_2-SGRGKGGKGLGKGGA-NH_2).

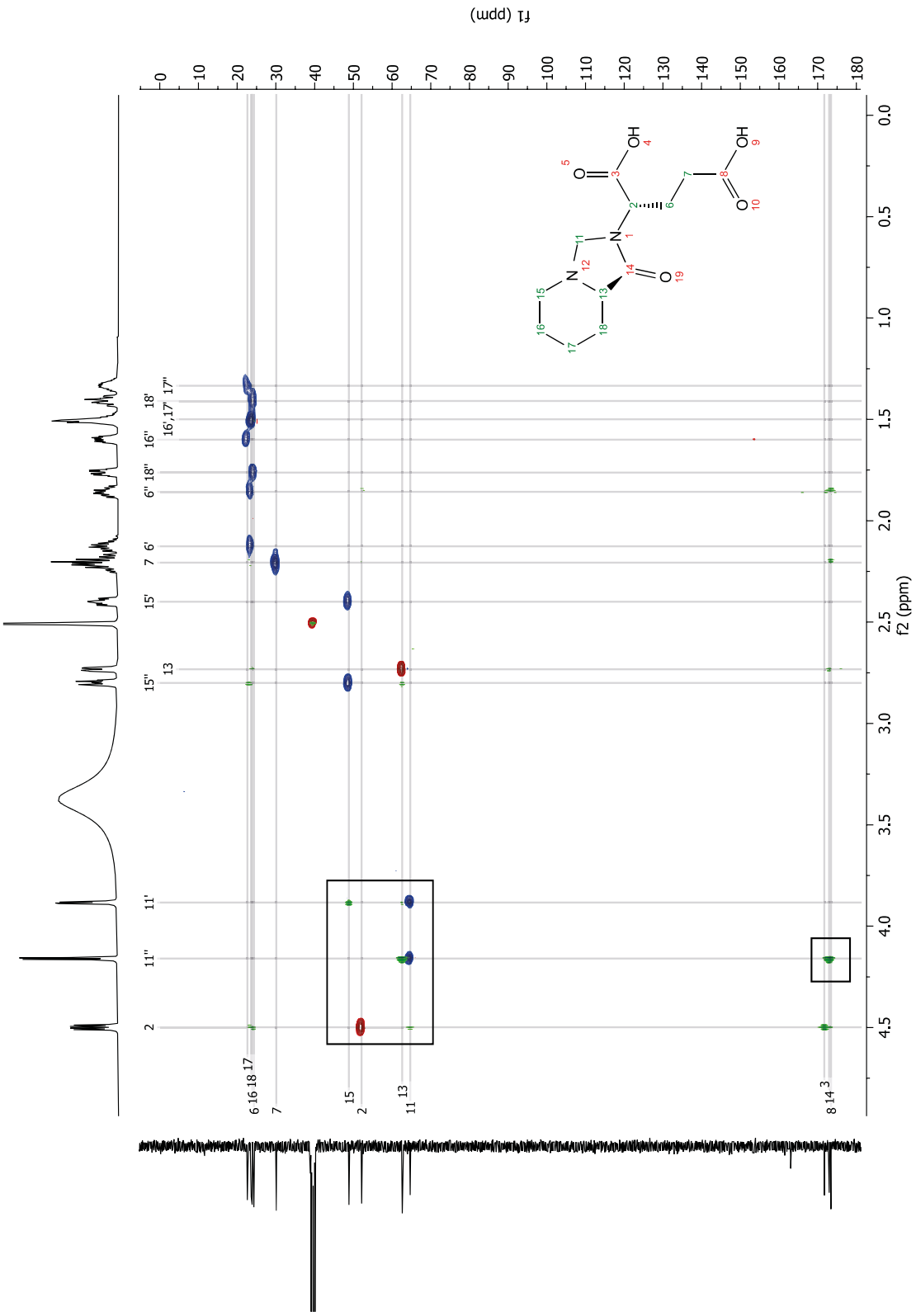


Figure 2.17 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC for the HPLC purified PipGlu-derived 3a (700 MHz). Important interactions are highlighted in a rectangle. The x-axis shows the ^{1}H spectrum and the y-axis shows the ^{13}C spectrum.

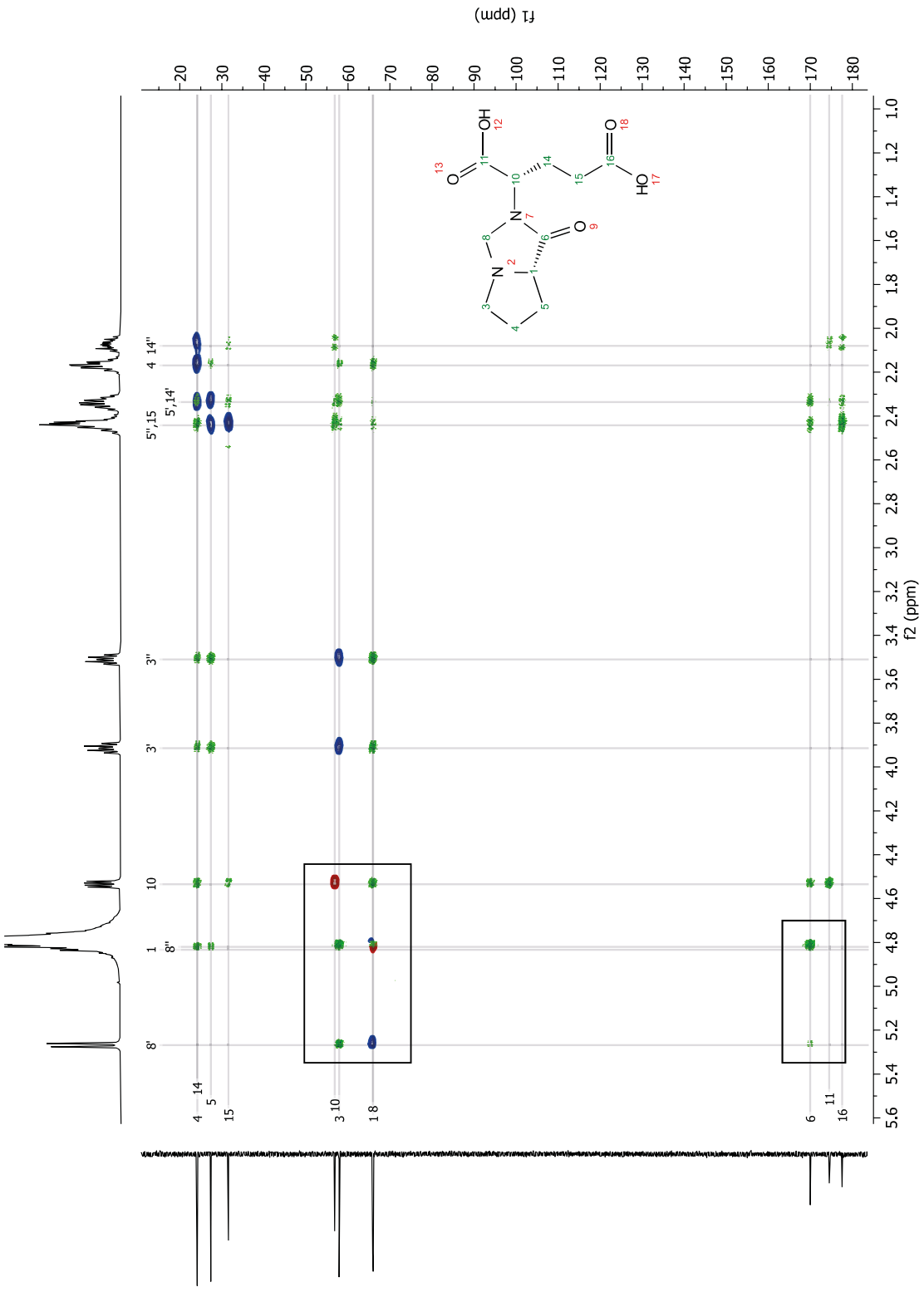


Figure 2.18 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC for the HPLC purified ProGlu-derived 2a (600 MHz). Important interactions are highlighted in a rectangle.

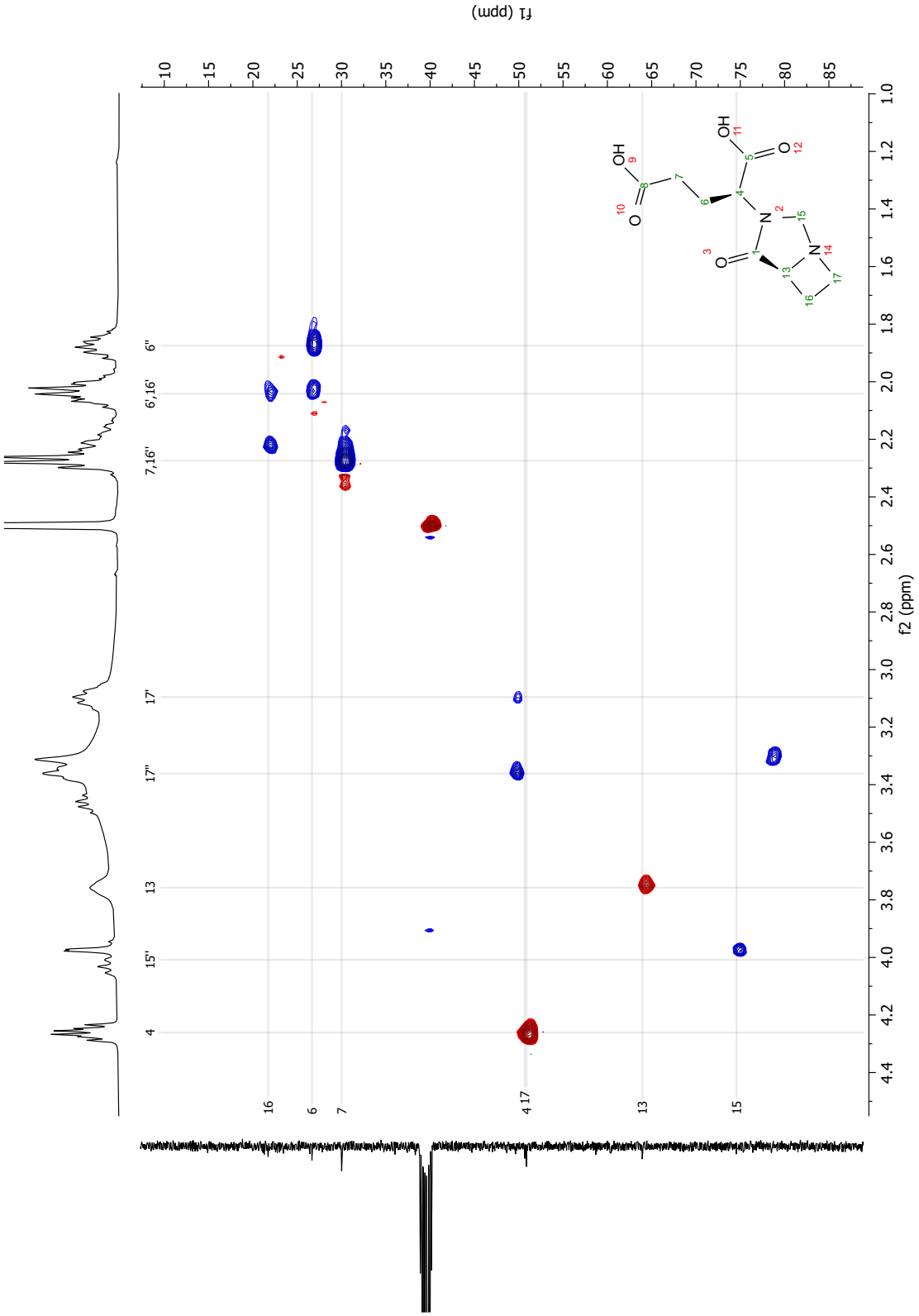


Figure 2.19 | 2D ^{1}H-^{13}C HSQC for the HPLC purified AzeGlu-derived 1a (400 MHz). Note the peak at δ_H 3.98 and δ_C 74.6 ppm was tentatively assigned to a methylene bridge; this signal was not observed when the sample was measured again at a later time-point (6 days later, see Figure 2.20), indicating the methylene bridge is relatively unstable in a 4,5-bicyclic ring system.

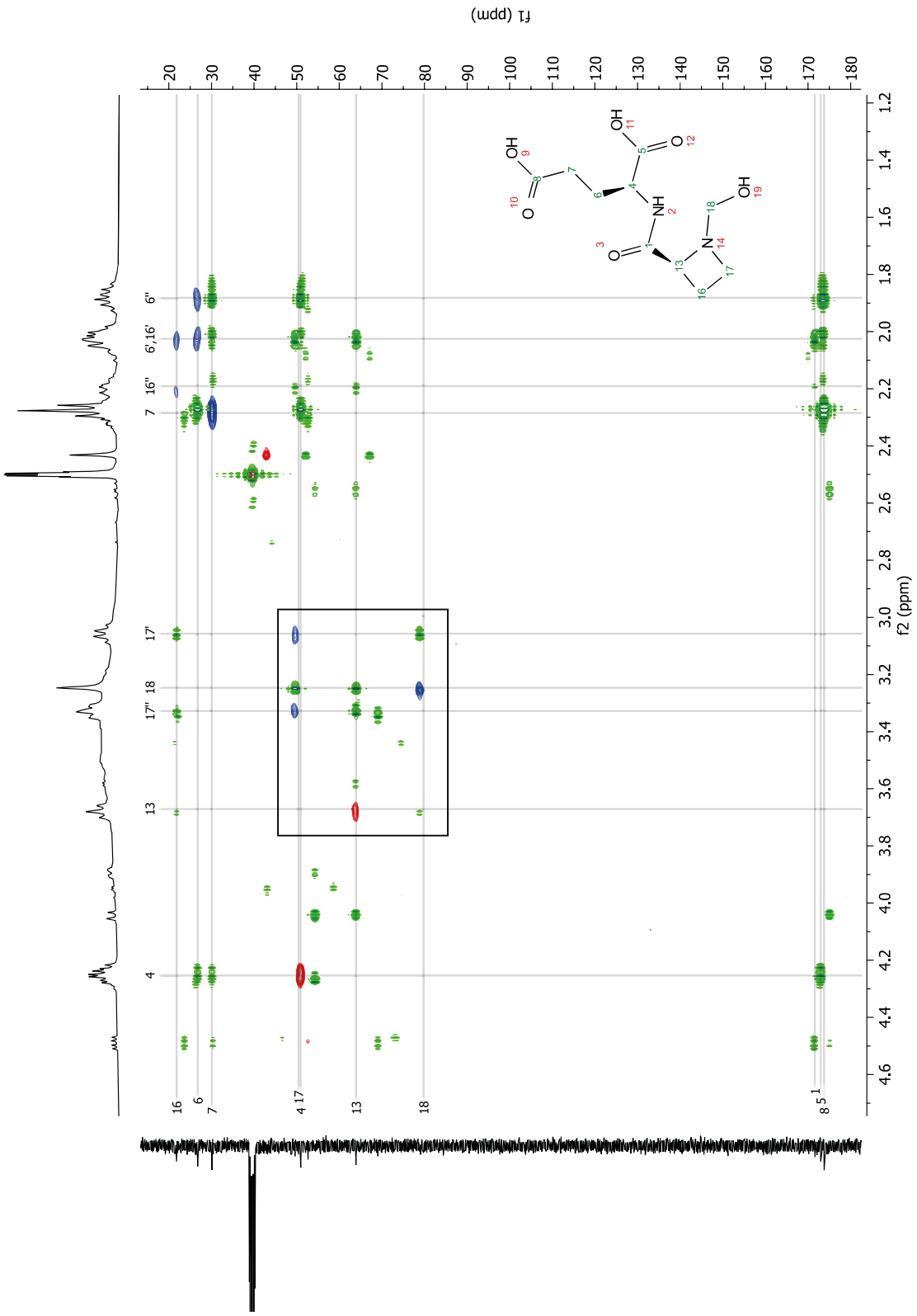


Figure 2.20 | Overlay of ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC for the HPLC purified AzeGlu-derived 1a/b 6 days later (400 MHz). The signals from Figure 2.19 at δ_H 3.98 and δ_C 74.6 ppm have disappeared and another signal at δ_H 3.25 and δ_C 79.8 ppm appears, which is also present in Figure 2.19 in trace amounts.

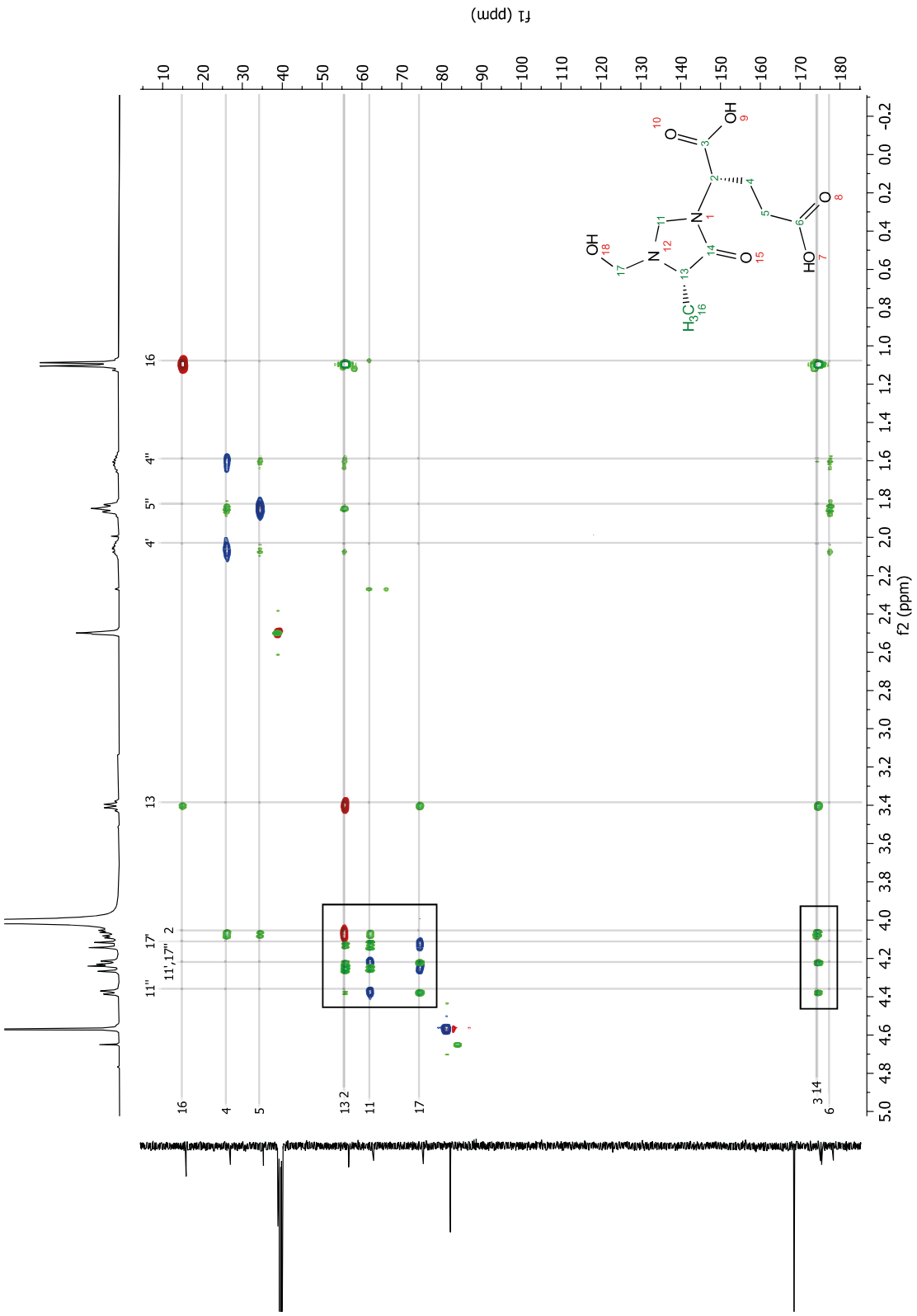


Figure 2.21 | Overlay of ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC for the HPLC purified AlaGlu-derived 4a (400 MHz). The formation of a product with two methylenes is observed upon reaction of **4** with HCHO, one assigned to a hemiaminal at δ_C 75.4, δ_H 4.13 and δ_H 4.24 ppm; the other one was assigned to an aminal at δ_C 63.0, δ_H 4.38 and δ_H 4.24 ppm. Important interactions are highlighted in a rectangle.

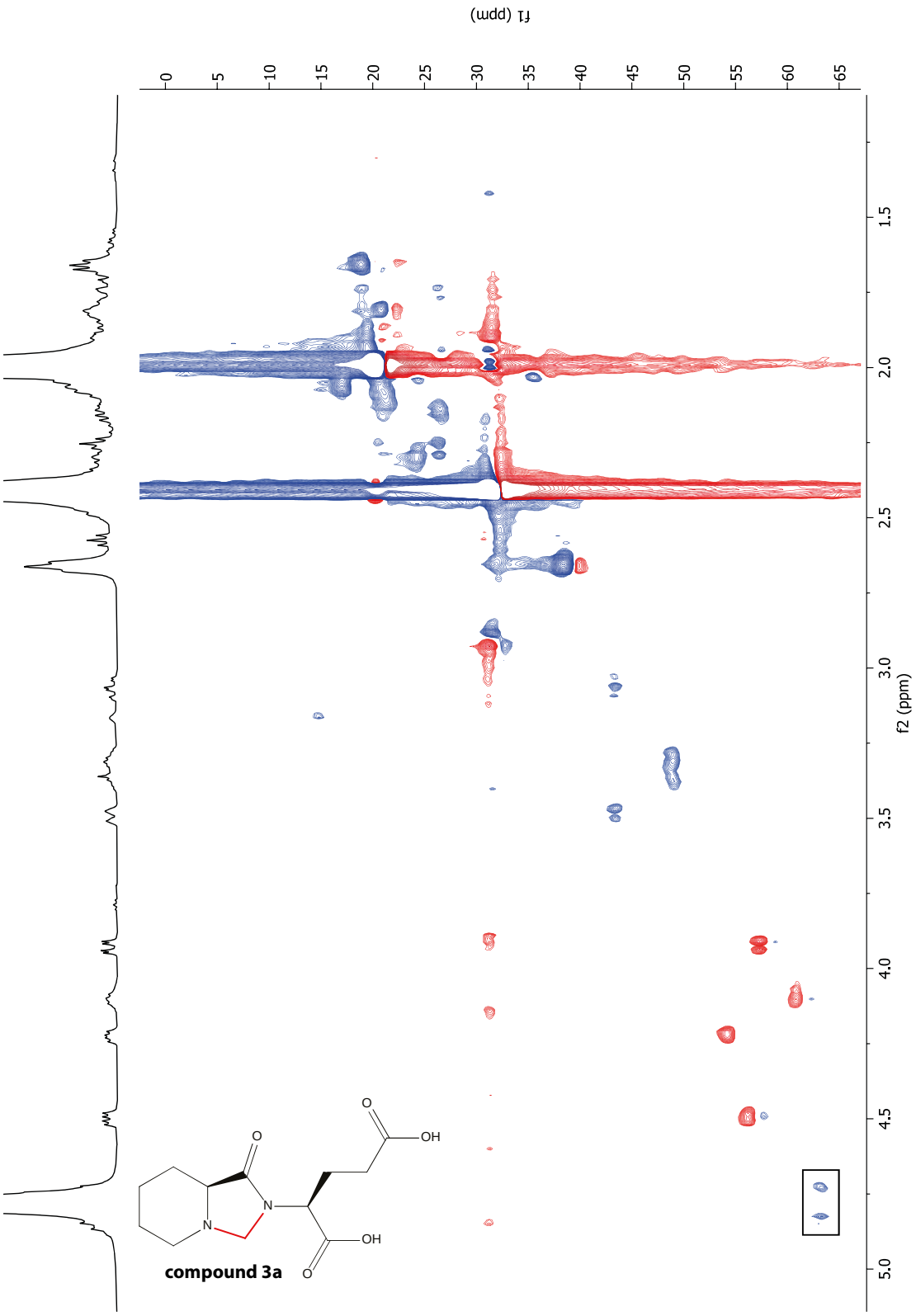


Figure 2.22 | 2D ^{1}H-^{13}C HSQC showing addition of a 40-fold excess of 1,3-cyclohexanedione to *in situ* generated PipGlu-derived 3a (400 MHz). 3 was reacted with a 10-fold excess HCHO *in situ* overnight, followed by addition of a 40-fold excess of 1,3-cyclohexanedione. The methylene bridge marked in a rectangle is in alignment with the methylene bridge of HPLC purified **3a** at δ_C 64.7 ppm, suggesting the 6,5-bicyclic adduct is stable.

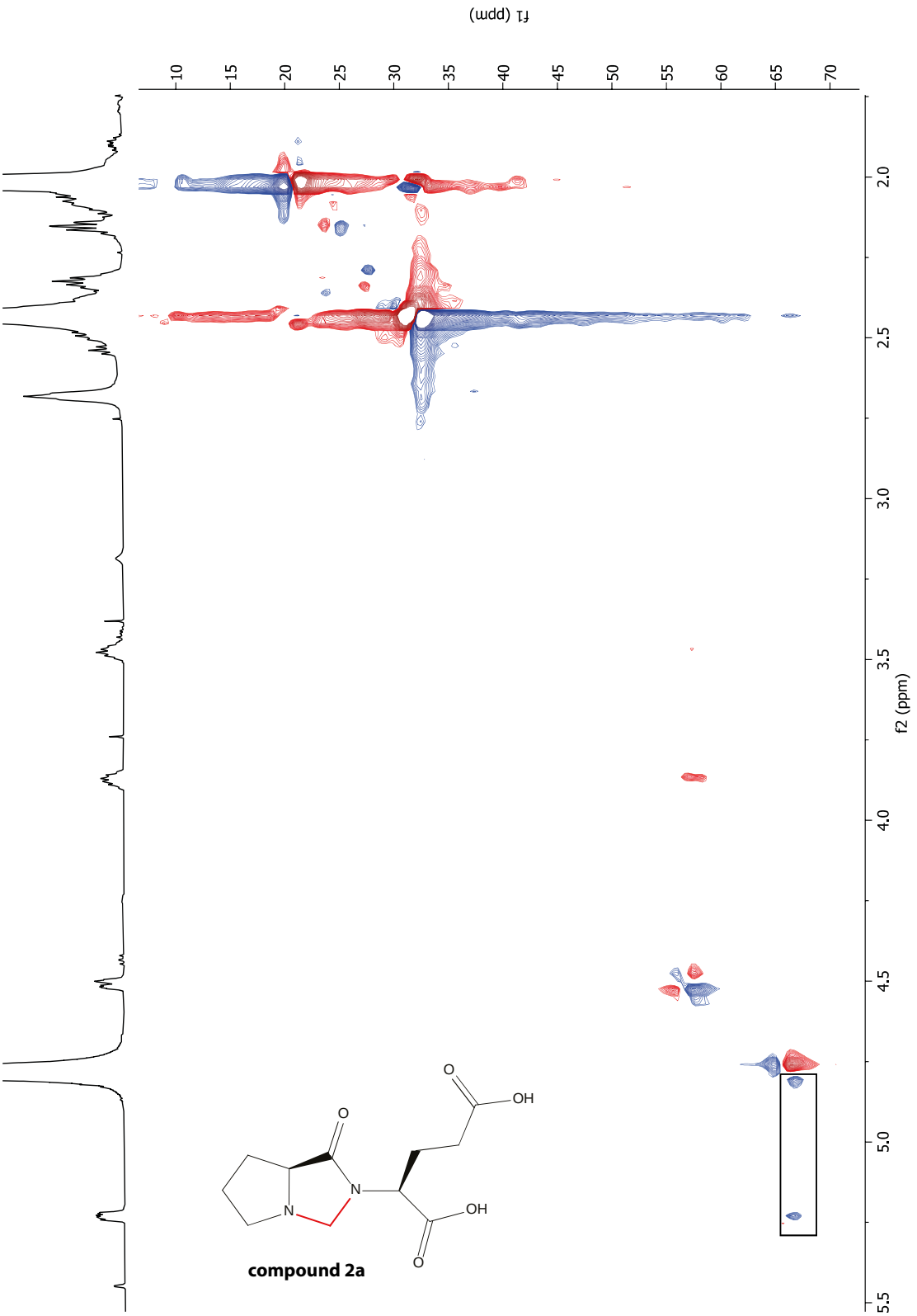


Figure 2.23 | 2D ^{1}H-^{13}C HSQC showing addition of a 40-fold excess of 1,3-cyclohexanedione to *in situ* generated ProGlu-derived 2a (600 MHz). **2** was reacted with a 10-fold excess HCHO *in situ* overnight, followed by addition of a 40-fold excess of 1,3-cyclohexanedione. The methylene bridge marked in a rectangle is in alignment with the methylene bridge of HPLC purified **2a** at δ_C 66.7 ppm, suggesting the 5,5-bicyclic adduct is stable.

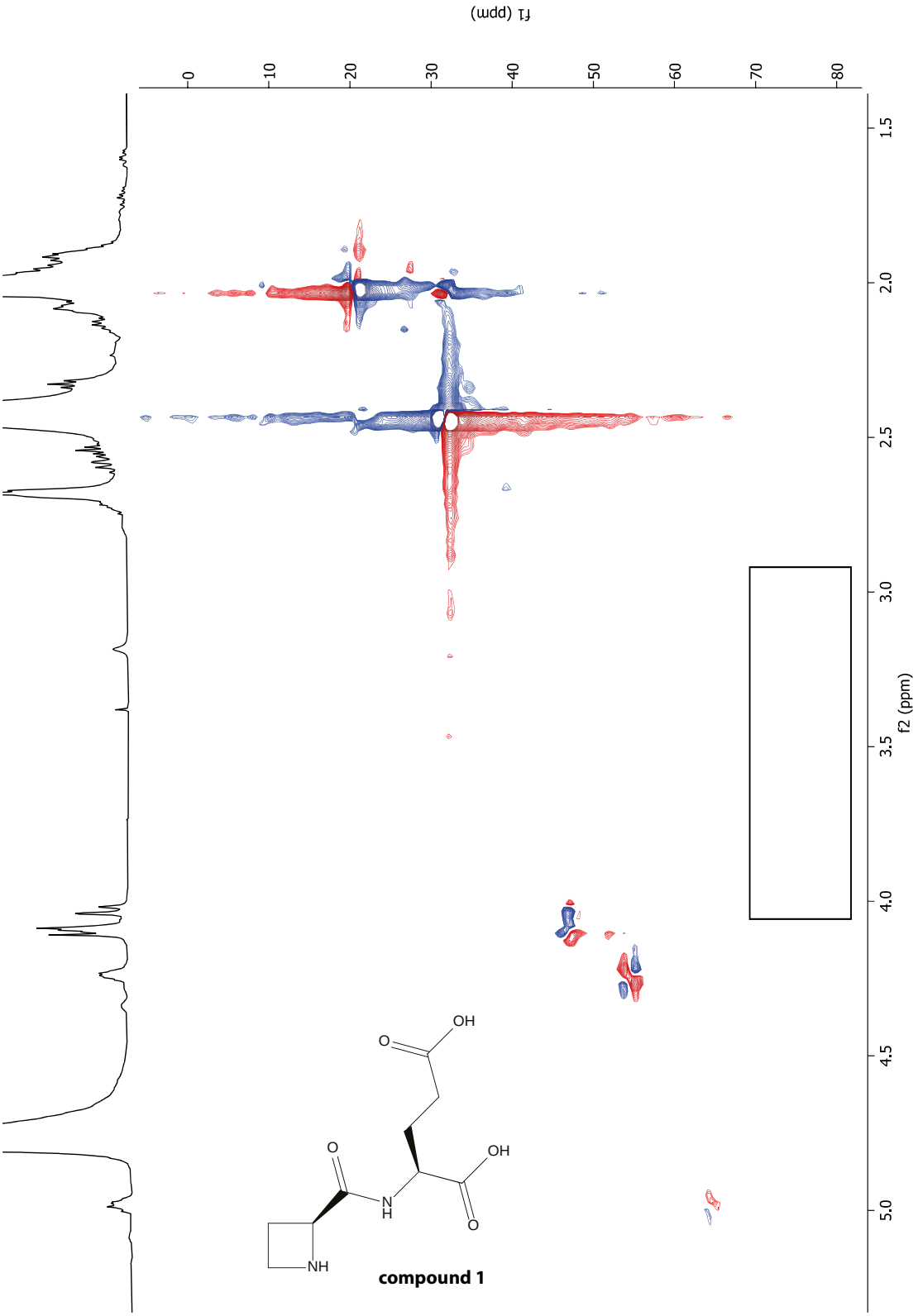


Figure 2.24 | 2D ^{1}H-^{13}C HSQC showing addition of a 40-fold excess of 1,3-cyclohexanedione to *in situ* generated AzeGlu-derived 1a/b (600 MHz). 1 was reacted with a 10-fold excess HCHO *in situ* overnight, followed by addition of a 40-fold excess of 1,3-cyclohexanedione. The aminal or hemiaminal methylene bridge of **1a/b** is expected to appear between δ_C 75 and δ_C 80 ppm (marked in a rectangle); no signal was detected in that area, suggesting the 4,5-bicyclic adduct is not stable.

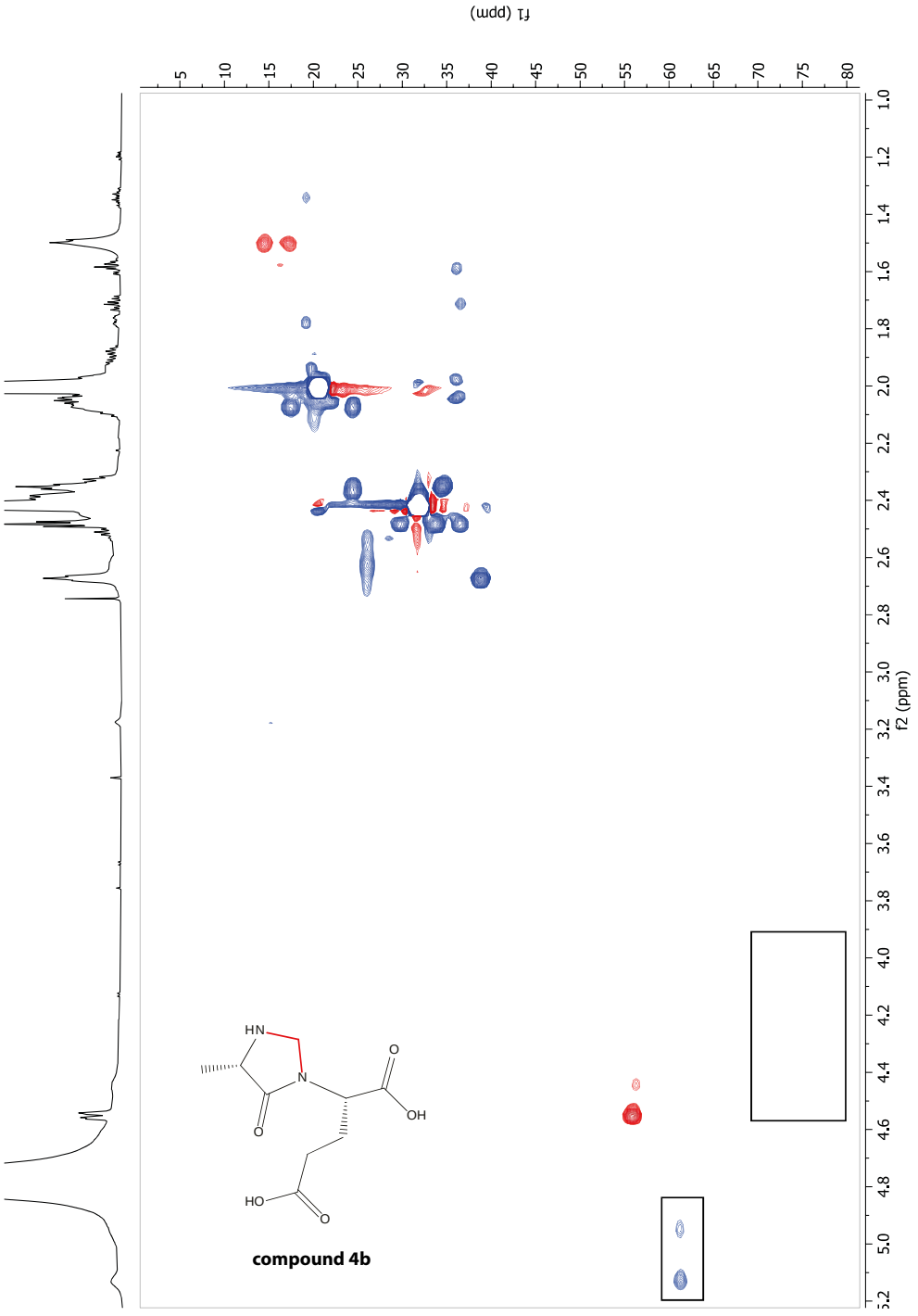


Figure 2.25 | 2D ^{1}H-^{13}C HSQC showing addition of a 40-fold excess of 1,3-cyclohexanedione to *in situ* generated AlaGlu-derived 4a (700 MHz). 4 was reacted with a 10-fold excess HCHO *in situ* for 3 weeks to promote formation of **4a**, followed by addition of a 40-fold excess of 1,3-cyclohexanedione. The hemiaminal methylene bridge of **4a** is expected to appear at δ_C 75.4 ppm; no signal is detected in that area (empty rectangle), suggesting the hemiaminal methylene bridge is not stable. However, the aminal bridge signal (marked in a rectangle) is in alignment with the methylene bridge of HPLC purified **4a** at δ_C 63.0 ppm. This suggests the HCHO-derived aminal, but not the hemiaminal methylene bridge is stable in presence of excess 1,3-cyclohexandione, resulting in formation of **4b**.

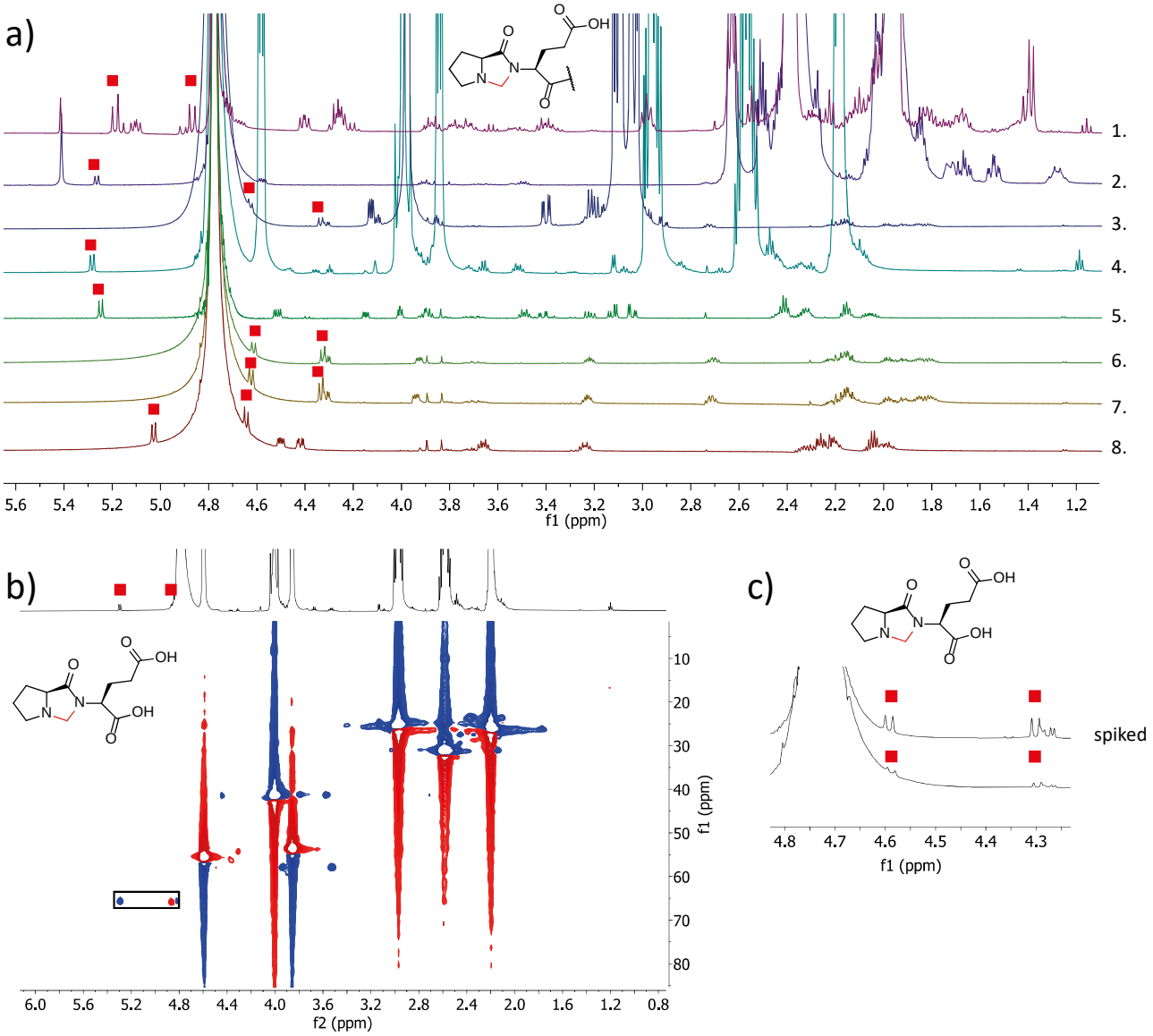


Figure 2.26 | Evidence that the methylene bridge between the proline nitrogen and the amide bond nitrogen is chemically stable in the presence of HCHO scavengers. a) ^{1}H NMR of HPLC isolated PEPAK HCHO adduct (1., 400 MHz) or HPLC isolated ProGlu-derived **2a** (2.-8.) in presence of 40-fold excess competing nucleophiles (1,3-cyclohexanedione (1.+2.), cysteine (3.), and GSH (4.), 600 MHz), at 75 °C with 1 equivalent cysteine (5., 600 MHz) or at different physiological pH conditions (pD 8.9 (6.), pD 7.8 (7.), pD 6.3 (8.), 700 MHz). **b)** All results in (a) were confirmed by 2D ^{1}H-^{13}C HSQC measurements (the methylene bridge is marked in a rectangle), here exemplified by **2a** in presence of a 40-fold excess of GSH (600 MHz). **c)** Competition reactions were spiked with HPLC purified **2a** to ensure correct assignment, here exemplified by spiking the mixture of **2a** in presence of a 40-fold excess of cysteine (lower panel) with HPLC purified **2a** (upper panel) (600 MHz). Methylene bridge signals of **2a**/PEPAK HCHO adduct are marked in red throughout the figure.

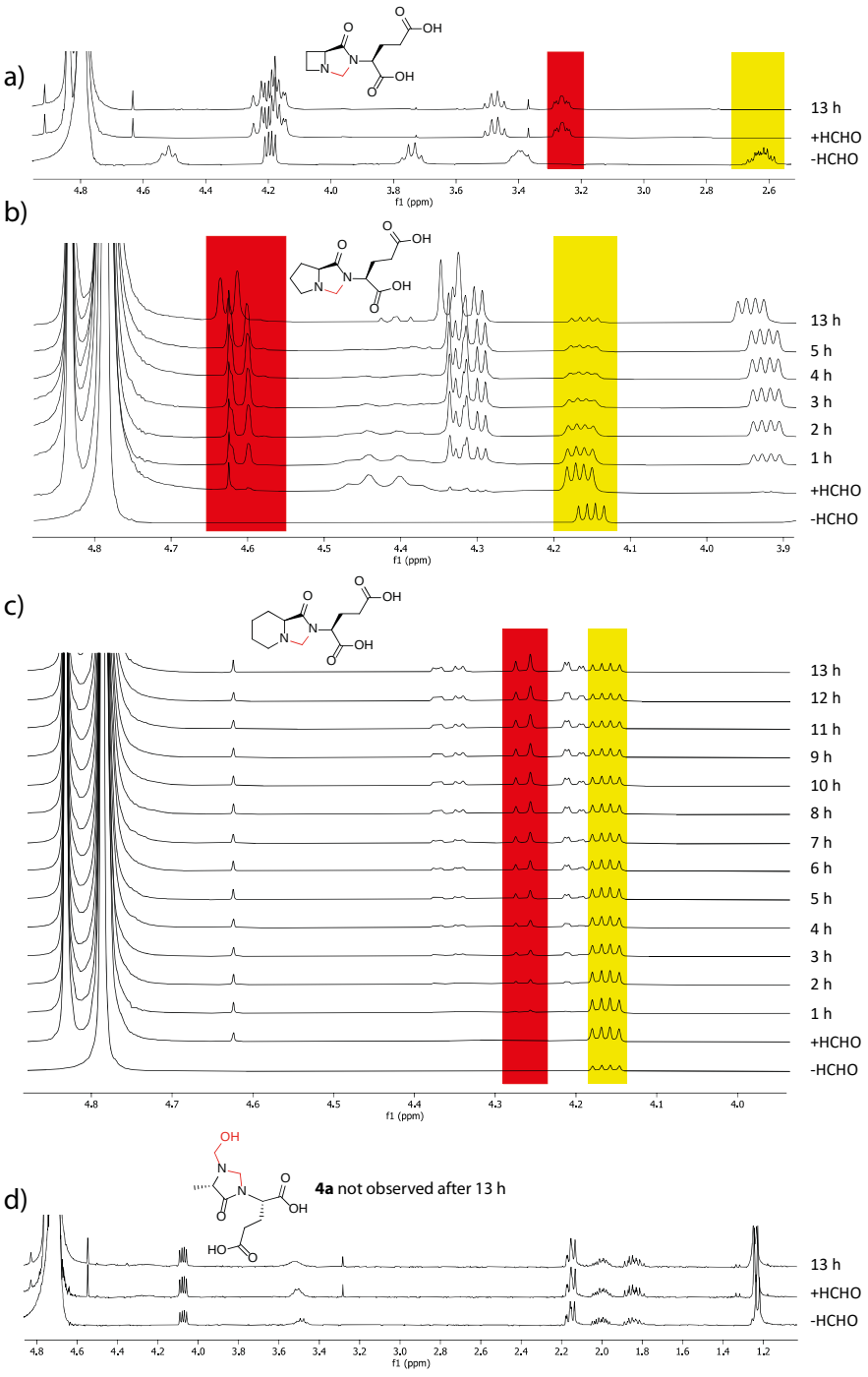


Figure 2.27 | ^{1}H NMR showing time-course analyses of dipeptides upon addition of a 10-fold excess HCHO at pD 10 (400 MHz). a) Reaction of **1** with a 10-fold excess of HCHO (pD 10) showing full conversion directly after addition of HCHO. **b)** Reaction of **2** with a 10-fold excess of HCHO (pD 10). **c)** Reaction of **3** with a 10-fold excess of HCHO (pD 10). **d)** Reaction of **4** with a 10-fold excess of HCHO (pD 10) indicating that **4** is extremely slow to react under these conditions. The first panel shows dipeptide only (-HCHO), the second panel shows the spectrum taken directly after addition of HCHO (+HCHO). All reactions were then monitored each hour for 13 h. The yellow rectangle shows peaks used for quantification of starting material, the red rectangle shows peaks used for quantification of products. The expected products are shown above each time-course. Data was used to generate the plot in Figure 2.5 c.

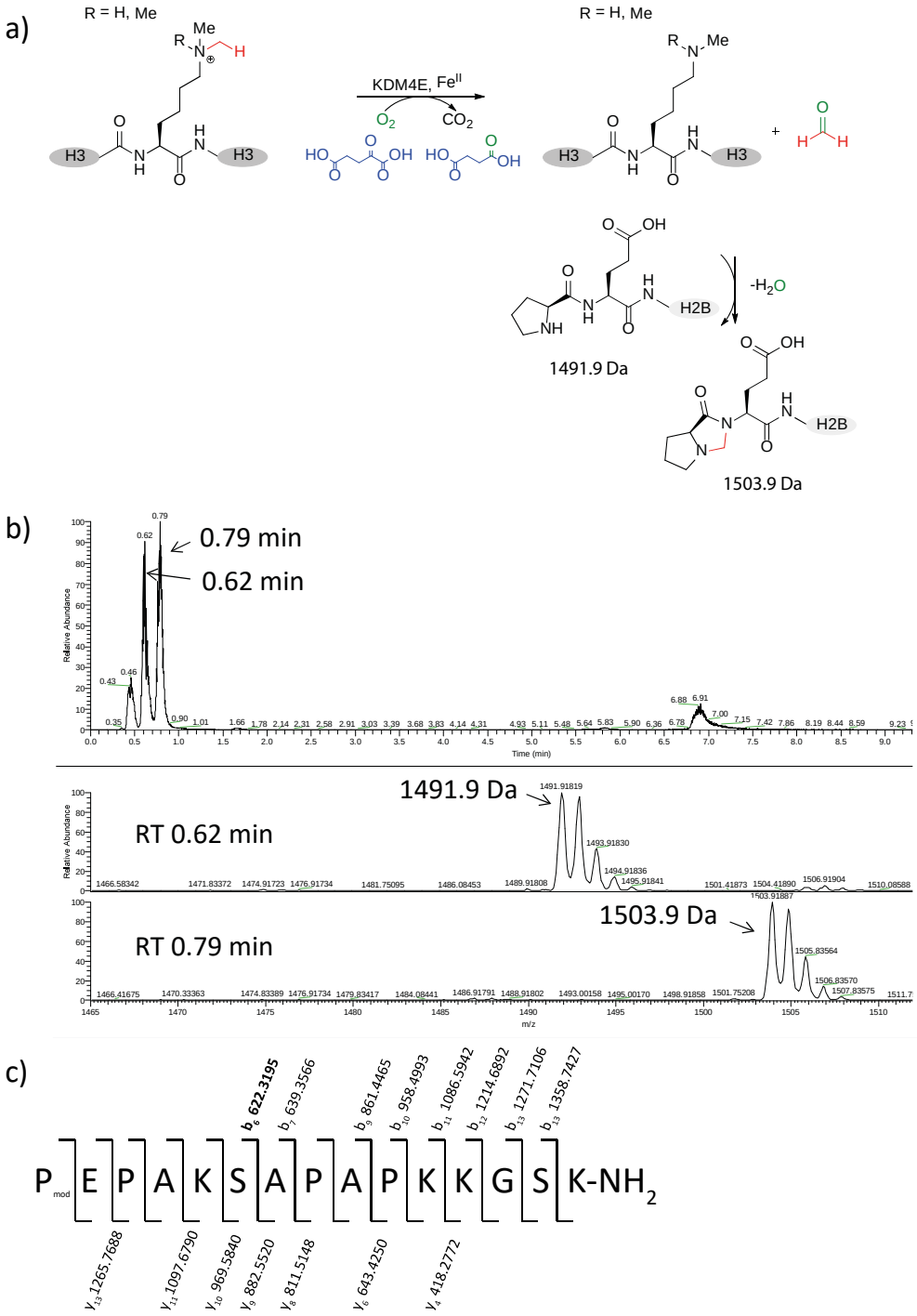


Figure 2.28 | H2B reacts with *in situ* generated HCHO. a) H3K9Me$_3$ is demethylated by KDM4E, resulting in formation of HCHO. HCHO reacts with the N-terminus of the H2B 15mer peptide. **b)** MS/MS results after 24 h incubation of KDM4E with H3K9Me$_3$ which enzymatically generates HCHO. Elution profile showing retention time of unreacted H2B 15mer at 0.62 min (1491.9 Da, unreacted H2B 15mer) and the modified H2B at 0.79 min (1503.9 Da, resembling M+12 Da). Sample was run and analysed by Elisabete Pires. **c)** MS/MS results summary showing that the observed +12 Da shift occurs on the N-terminus of H2B 15mer peptide (b_6 in bold). Samples were run and analysed by Elisabete Pires. H3K9Me$_3$ 1-21 (NH_2-ARTKQTARK(Me_3)STGGKAPRKQLA-NH_2).

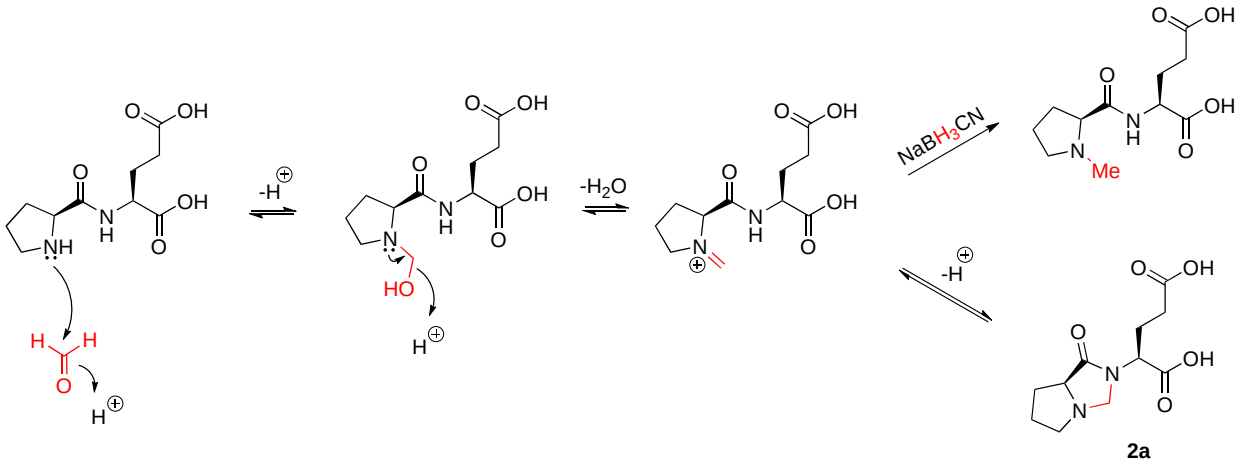


Figure 2.29 | Outline mechanism for methylene bridge formation to give a bicycle. Nucleophilic attack of the N-terminal nitrogen on HCHO results in hemiaminal and subsequent iminium formation. A methylene bridge is formed via 5-endo-trig cyclisation (lower path), resulting in **2a**. The iminium species was reduced with $NaBH_3CN$ to yield an *N*-methylated proline residue (upper path). Substructures that result from the reaction with HCHO and $NaBH_3CN$ are highlighted in red.

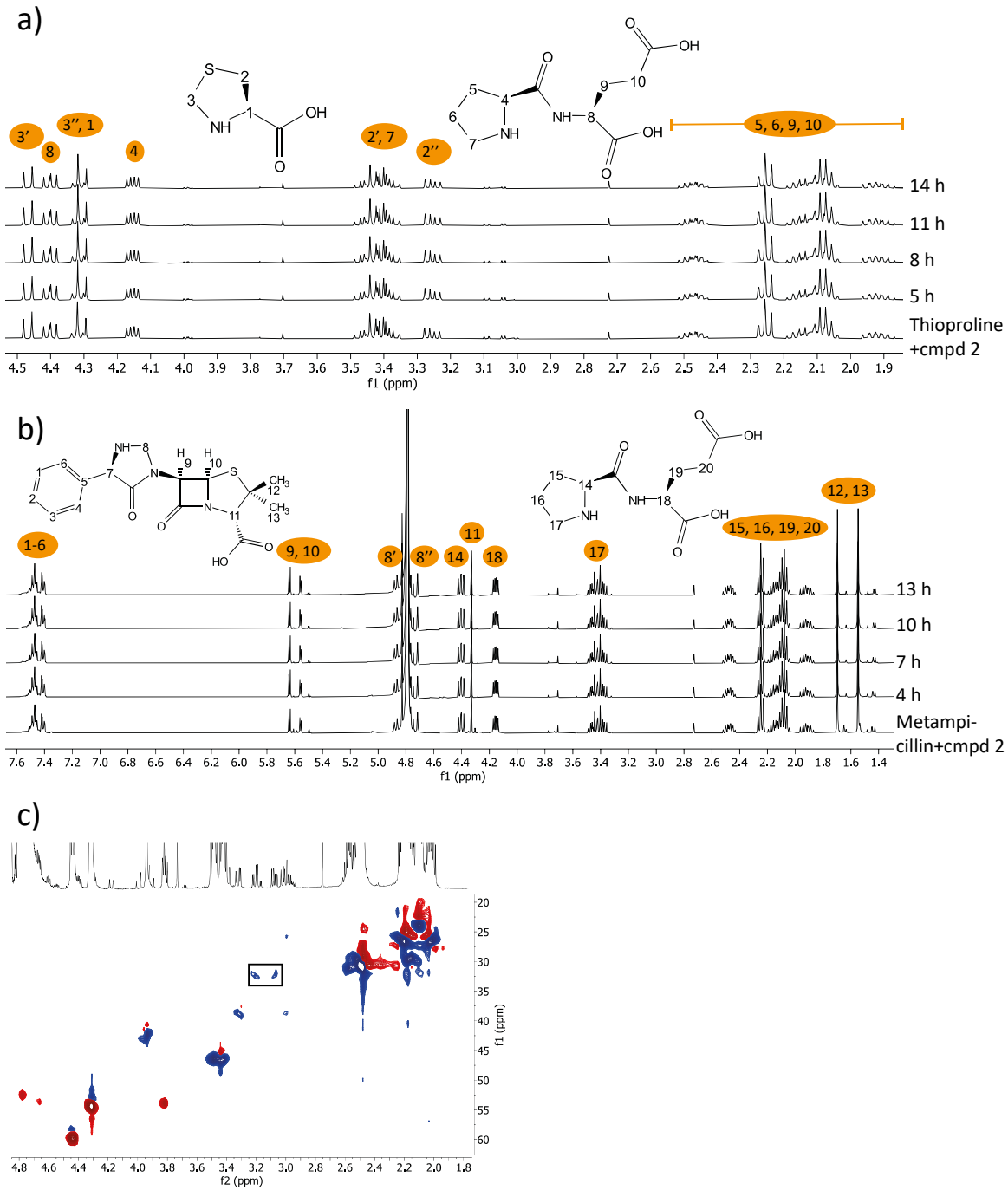


Figure 2.30 | Compound 2 does not react with molecules bearing HCHO-derived methylene bridges. **a)** HCHO was reacted with cysteine to generate thioproline *in situ* which was incubated with 1 equivalent of **2** and monitored over 14 h by ^{1}H NMR (400 MHz) showing the HCHO-derived methylene bridge on thioproline is not removed by **2**. **b)** HCHO was reacted with ampicillin to generate metampicillin *in situ* which was incubated with 1 equivalent of **2** and monitored over 13 h by ^{1}H NMR (400 MHz) showing the HCHO-derived methylene bridge on metampicillin is not removed by **2**. **c)** HCHO was reacted with glutathione (GSH) to generate *S*-(hydroxymethyl)glutathione (HMG) *in situ*, to which 40 equivalents of **2** were added, followed by monitoring the reaction by 2D ^{1}H-^{13}C HSQC (600 MHz). The HCHO-derived methylene bridge on HMG (marked with a rectangle) is not removed by an excess of **2** under the tested conditions.

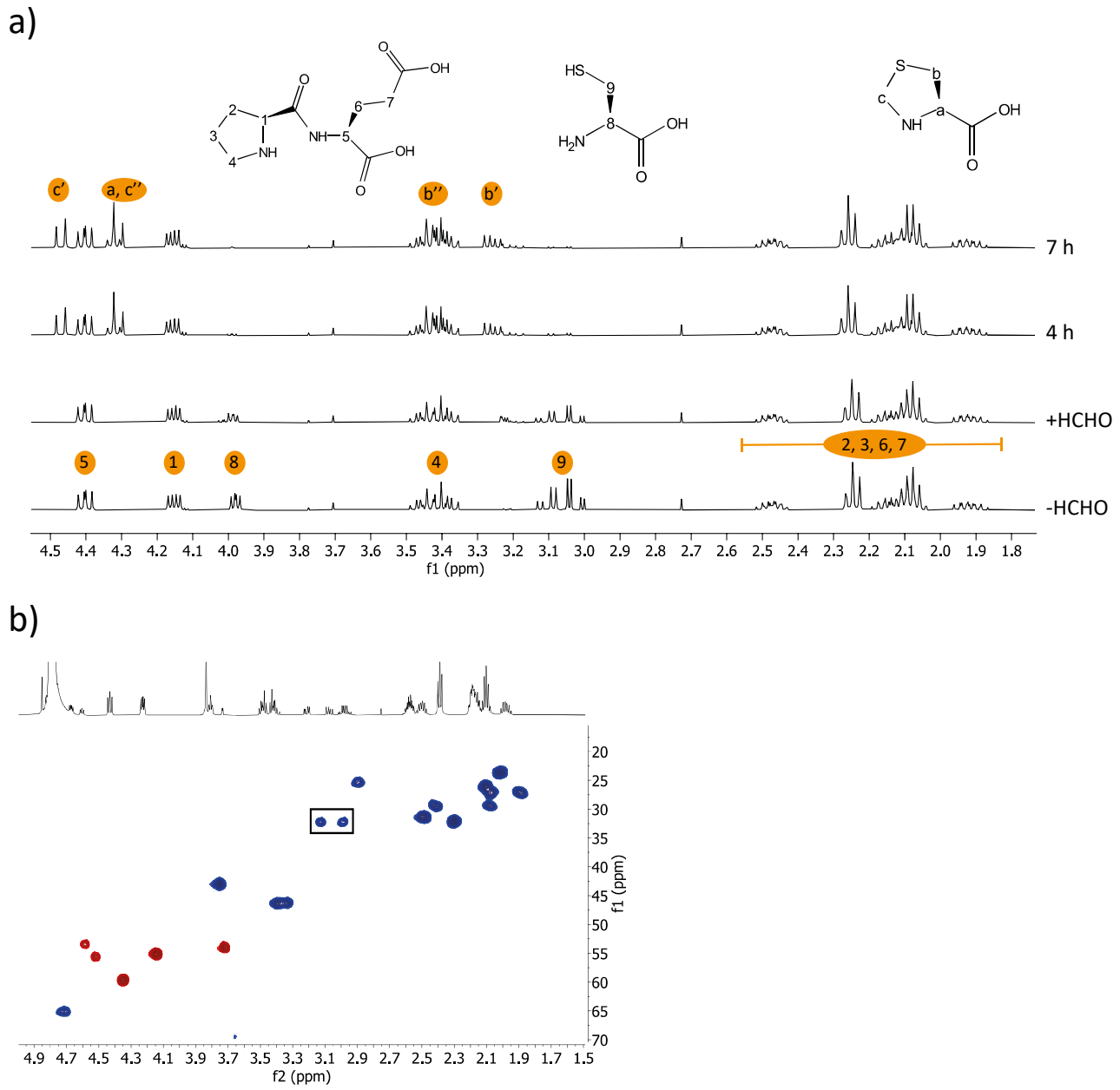


Figure 2.31 | Cysteine-derived thiol groups react more readily with HCHO than compound 2. a) 1 equivalent HCHO was added to cysteine and **2**. Monitoring of the reaction over 7 h by ^{1}H NMR (400 MHz) shows cysteine reacts with HCHO to form thioproline. **b)** 2D ^{1}H-^{13}C HSQC (600 MHz) analysis of the reaction of GSH and **2** in the presence of 1 equivalent HCHO after 10 h. The rectangle shows the HCHO-derived hydroxymethyl adduct on *S*-(hydroxymethyl)glutathione (HMG).

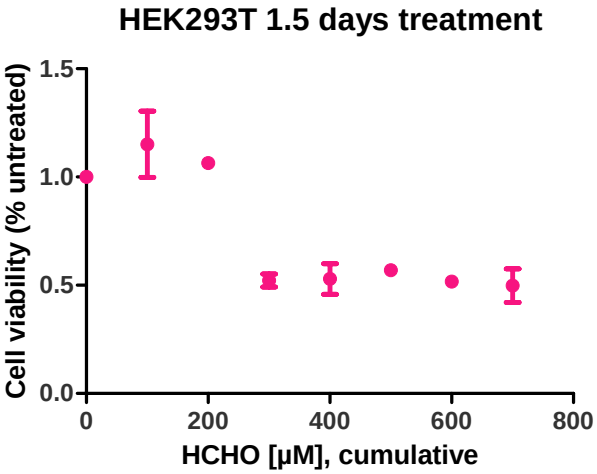


Figure 2.32 | HEK293T cell viability with two cumulative HCHO treatments. 10,000 cells were seeded in a 96-well plate. After the cells were attached (>8 h), the cells were treated with the first dose of HCHO, followed by the second dose of HCHO 24 h after that. The second dose was 1.5-fold higher than the first dose. E.g. the data point at 200 µM HCHO means addition of 80 µM HCHO, followed by addition of 120 µM HCHO. 14 h after the second treatment, viable cells were estimated using CellTiter 96 AQueous One Solution Reagent (Promega, G358C). Errors: SD of the mean (n=3, technical repeats).

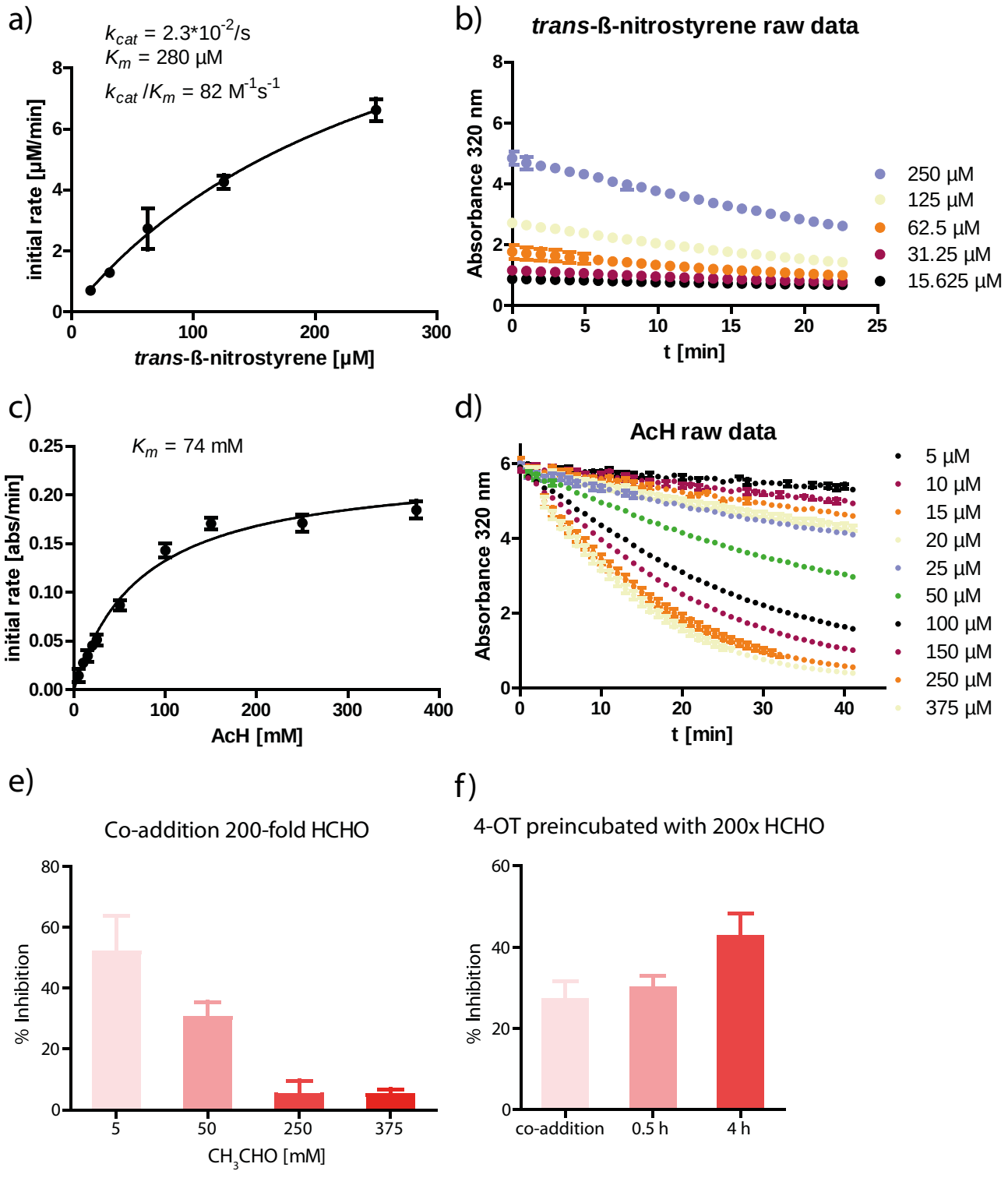


Figure 2.33 | Kinetic parameters of 4-OT. a) The K_m value of *trans*-β-nitrostyrene was determined to be 281 µM ± 42. **b)** Raw data used to determine the K_m value of *trans*-β-nitrostyrene. **c)** For acetaldehyde (CH_3CHO), the kinetic parameters of 4-OT are K_m=74 mM ± 4.8. **d)** Raw data used to determine the K_m value of CH_3CHO. **e)** A 200-fold molar excess of HCHO was added to 4-OT, immediately followed (no pre-incubation) by addition of the indicated amounts of CH_3CHO and 200 µM *trans*-β-nitrostyrene at pH 7.4. **f)** Assay monitoring the absorbance of *trans*-β-nitrostyrene at 320 nm (pH 7.8), showing an increase in inhibition correlates with prolonged pre-incubation time of 4-OT with HCHO. Errors: SD of the mean (n=2).

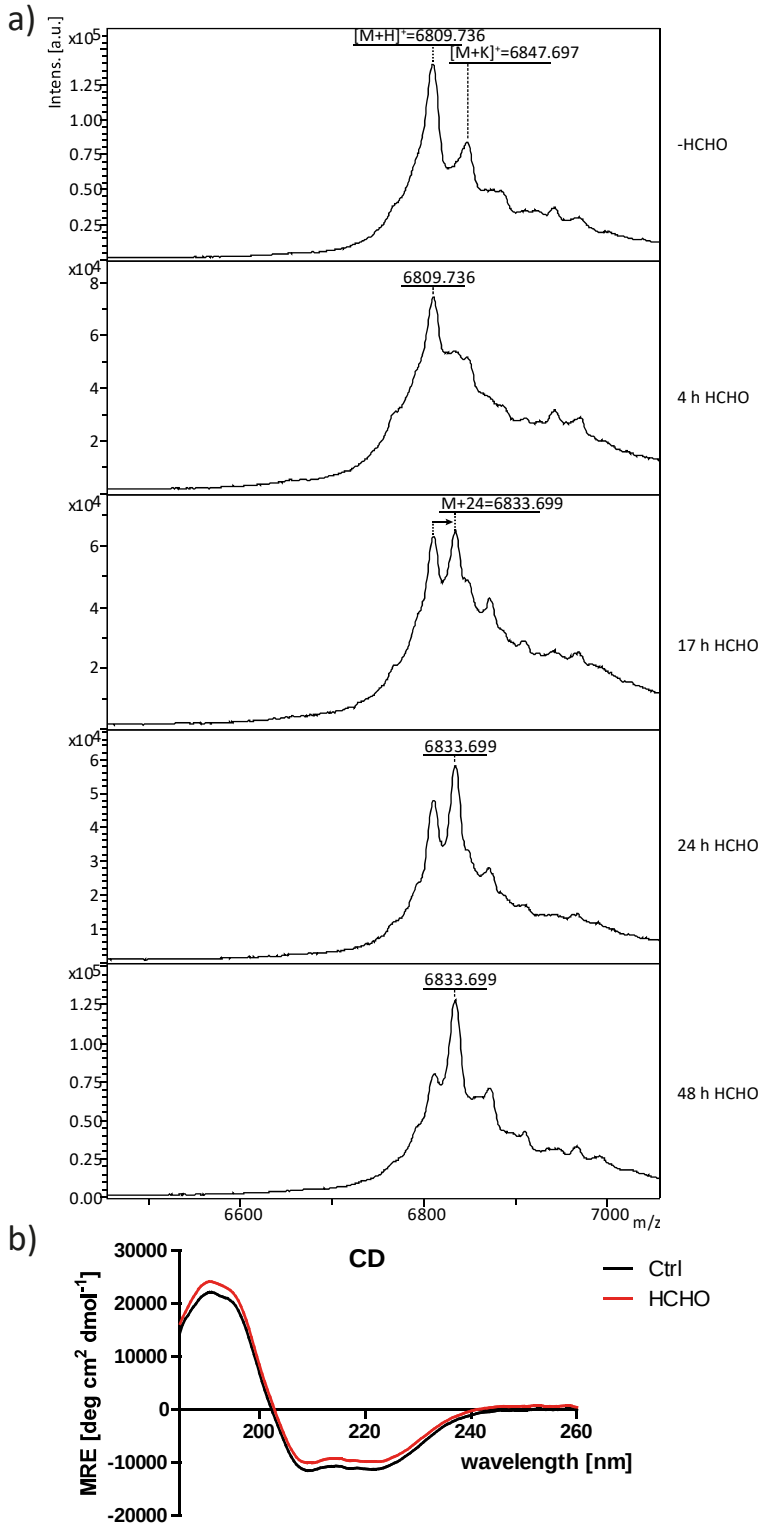


Figure 2.34 | Evidence that HCHO covalently reacts with 4-OT without altering its secondary structure. **a)** MALDI MS results showing a +24 Da shift of 4-OT after incubation (17 h) with a 200-fold excess of HCHO at pH 7.8. Similar results were obtained at pH 7.4 (data not shown). **b)** Circular dichroism spectra of 4-OT pre-incubated with a 200-fold excess of HCHO for 48 h, compared with 4-OT treated with a buffer only control.

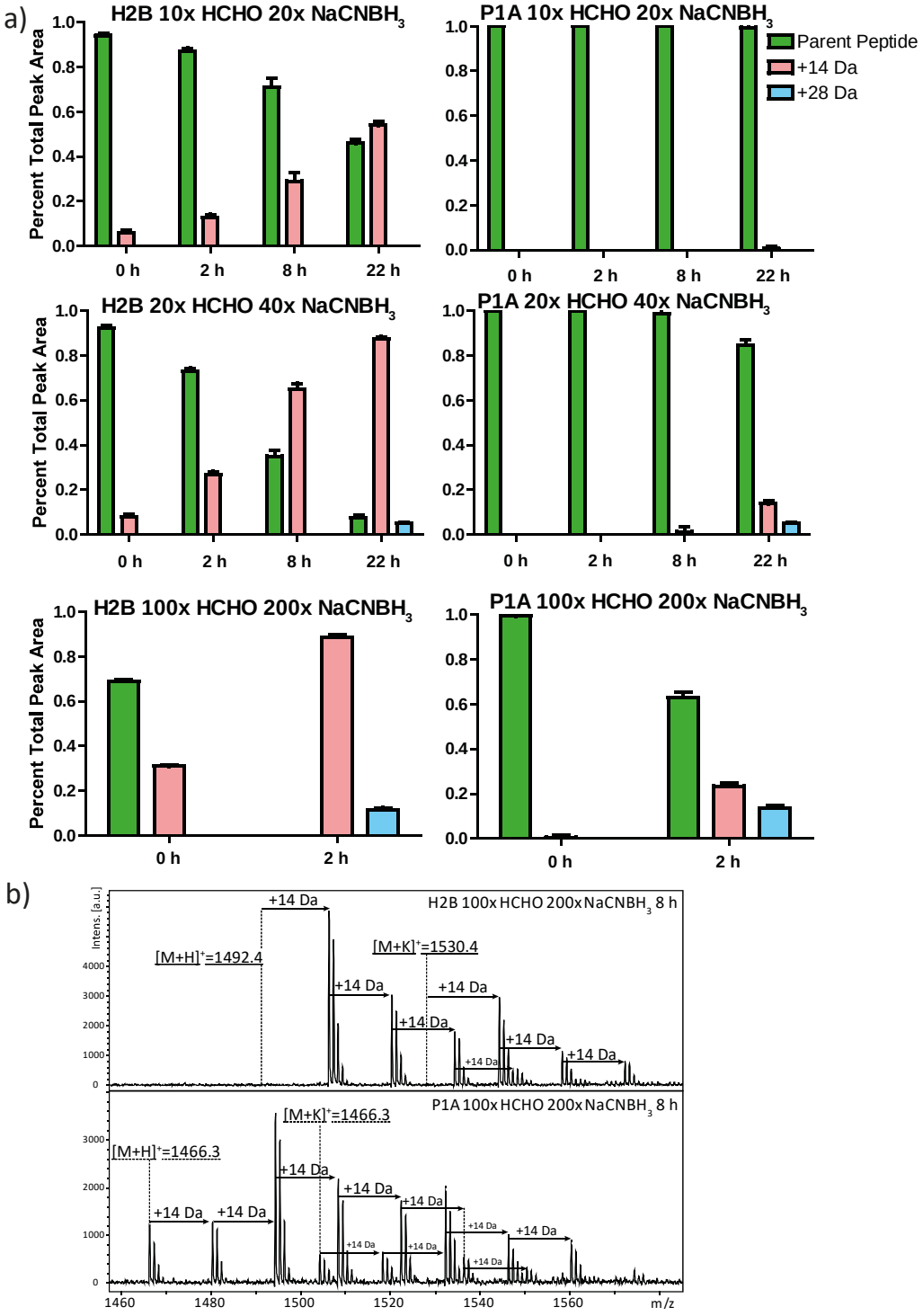


Figure 2.35 | Reductive methylation of the N-terminal proline of H2B. a) The H2B 15mer peptide or its P1A analogue were reacted with the indicated excess of HCHO and $NaCNBH_3$ in 20 mM potassium phosphate, pH 7.4. The reaction was monitored over time by MALDI MS. Errors: SD of the mean (n=3, technical repeats). **b)** Reaction of the H2B 15mer or its P1A analogue using a 100-fold excess of HCHO and a 200-fold excess of $NaCNBH_3$ (relative to peptide) results in abundant multiple +14 Da adducts after 8 h.

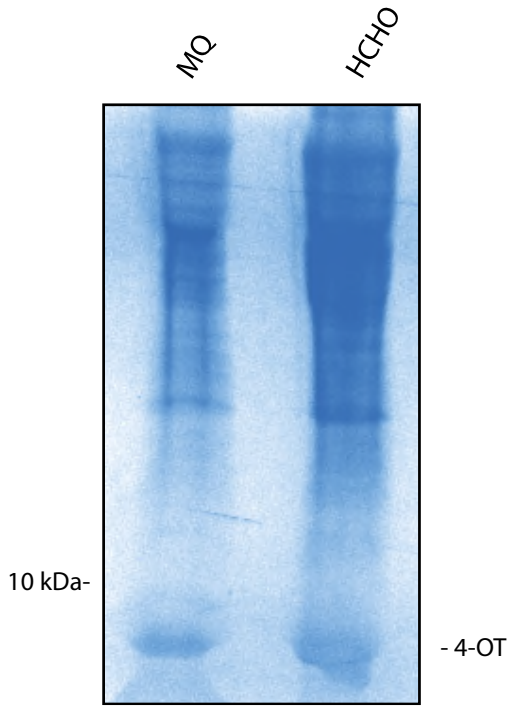


Figure 2.36 | In-gel trypsin digestion of 4-OT. Following re-inoculation using an overnight culture, 4-OT was biosynthesised for 24 h in BL21(DE3) (37 °C and shaking at 200 rpm, no IPTG added). After 24 h, 700 µM HCHO or MQ water were added and incubated (270 min, 37 °C and shaking at 200 rpm). Cells were centrifuged, sonicated and the supernatants containing the proteins were resolved by tricine SDS-PAGE. The bands corresponding to 4-OT were excised and subjected to trypsin digestion (Figure 2.14).

2.6 Methods

Paraformaldehyde, DCDO, $H^{13}CHO$, acetaldehyde, NaOD, DCl, tricine, 1-Boc-*L*-azetidine-2-carboxylic acid, *N*-methylmorpholine, *L*-glutathione (reduced), 1-Boc-*L*-proline-2-carboxylic acid, 1-Boc-*L*-piperidine-2-carboxylic acid, 1-Boc-*L*-alanine-2-carboxylic acid, iodoacetamide acid and 1,3-cyclohexanedione were from Sigma-Aldrich. Further chemicals were H-*L*-Glu(OtBu)-OtBu*HCl (Iris Biotech), dithiothreitol (Apollo Scientific), *L*-cysteine (Chem-Impex), EDC hydrochloride (Fluorochem), diazolidinyl urea (MP Biomedicals), imidazolidinyl urea (Alfa Aesar), cycloheximide (Cayman Chemical, CAY-14126-1), *trans*-β-nitrostyrene (Acros Organics 386530250).

2.6.1 Characterisation methods

NMR spectra (HSQC, HMBC, COSY, ^{1}H, ^{13}C) were obtained using either Bruker Avance AV700, AV600, AV500, or AV400 spectrometers. To reference spectra to the solvent peak, the following

shifts were used: δ_H = D_2O 4.79 ppm; δ_H = D_6-DMSO 2.5 ppm, δ_C = D_6-DMSO 39.52 ppm; δ_H = $CDCl_3$ 7.26 ppm, δ_C = $CDCl_3$ 77.16 ppm; δ_H MeOD = 3.31 ppm, δ_C = MeOD 49 ppm. For reporting multiplicities, the following abbreviations are used: m (multiplet), q (quartet), t (triplet), d (doublet), dd (doublet of doublets), s (singlet). For ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC spectra, the color codes are as follows: red HSQC signals denote CH or CH_3, blue HSQC signals denote CH_2 coupling, and green signals denote HMBC coupling. A Waters LCT Premier ESI mass spectrometer was used to obtain high resolution electron spray ionisation mass spectra (HRMS). An LC-MS Agilent Technologies 1200 series instrument was used to obtain electrospray ionisation mass spectra.

2.6.2 Solid-phase peptide synthesis

Peptides were synthesised on a 0.1 mmol scale as C-terminal amides from the C- to N-terminus on a rink amide-MBHA resin (100–200 mesh, 0.6–0.8 mmol/g loading, AGTC Bioproducts) using a LibertyBlue microwave peptide synthesiser (CEM) as described.[246] *N*-β-Fmoc protected amino acids (CS Bio, Novabiochem, Sigma-Aldrich, TCI, Alfa Aesar, Merck or AGTC Bioproducts) and, if required, acid-labile protecting groups were employed. Coupling and deprotection steps used the instrument's standard methods and were microwave assisted.

Amino acid solutions (0.2 M), dissolved in DMF (peptide synthesis grade, AGTC Bioproducts), were prepared. *N*,*N*'-Diisopropylcarbodiimide (TCI Europe) was used for coupling with Oxyma Pure. For Fmoc deprotection, 20% (v/v) piperidine in DMF was used. Upon synthesis of the peptide including final deprotection, the resin was washed three times with CH_2Cl_2 and dried in air. Acid labile protecting groups were removed and cleavage from the resin was conducted using 5 ml of a deprotection mixture (1,3-dimethoxybenzene (2.5%), triisopropylsilane (2.5%), Milli-Q (MQ) water (2.5%) in CF_3CO_2H (92.5%)) for 4 h at ambient temperature. Upon filtration of the resulting mixture, the peptide was precipitated with ice-cold Et_2O (45 ml). The solid was pelleted (4255 g, 10 min, 4.0 °C), the liquid was decanted off and the solid was dried on air. Before purification, peptides were dissolved in H_2O and lyophilised to remove CF_3CO_2H. Peptides were dissolved in H_2O and, if necessary, with MeCN to aid solubility. Upon filtration (0.45 µm), HPLC purification

was conducted using a Shimadzu HPLC purification system (composed of DGU-20A, 2 LC-20AR, CBM-20A, SPD-20A, and FRC-10A units) equipped with a NX-C18 LC column (250 × 21.2 mm, 110 Å; Phenomenex Gemini). The mobile phase employed a gradient of MeCN + 0.1% (v/v) formic acid in H_2O + 0.1% (v/v) formic acid. Upon HPLC purification, MALDI MS or LC-MS were used to analyse fractions. Fractions containing the product were pooled according to purity and lyophilised. The purified peptide was dissolved in H_2O and its concentration was determined using quantitative ^{1}H NMR (700 MHz) spectroscopy:[247] 16 µl of the D_2O peptide solution were added to 143 µl D_2O and 1 µl of 1 mg/ml 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP; Apollo Scientific). The peptide concentration was calculated using the following equation: $M_x/M_y=(I_x/I_y)x(N_y/N_x)$, with M_x/M_y representing the molar ratio of TSP and the peptide, I_x/I_y the signal intensities corresponding to the methyl protons of TSP and the peptide (typically averaged over the Ala, Val, Thr, and Leu residues using Global Spectrum Deconvolution), and N_y/N_x number of nuclei responsible for the intensity. For this study, the following N-terminal peptides were made as C-terminal amides, corresponding to the human sequences: H2B 1-15 (NH-PEPAKSAPAPKKGSK-NH_2); H2B P1A 1-15 (NH_2-AEPAKSAPAPKKGSK-NH_2); H2B 1-5 (NH-PEPAK-NH_2); H2A 1-15 (NH_2-SGRGKQGGKARAKAK-NH_2); H3 1-15 (NH_2-ARTKQTARKSTGGKA-NH_2); H3K9Me_3 1-21 (NH_2-ARTKQTARK(Me_3)STGGKAPRKQLA-NH_2); H4 1-15 (NH_2-SGRGKGGKGLGKGGA-NH_2).

2.6.3 Synthesis of dipeptides

Boc- and tert-butyl-protected dipeptides

L-Glutamic acid di-tert-butyl ester hydrochloride (1.24 mmol) and the Boc-protected amino acid (Boc-*L*-alanine, Boc-*L*-proline, Boc-*L*-azetidine, Boc-*L*-piperidine; 1.24 mmol) were dissolved in 10 ml CH_2Cl_2 at 0 °C. *N*-Methylmorpholine (1.24 mmol) and EDAC.Cl (1.24 mmol) were added and the mixture was left to warm to ambient temperature and stirred for 1 h. Upon completion of the reaction, the mixture was quenched with 1 M HCl (aq) to pH 7, the organic layer was separated and washed with 1 M NaCl, dried over $MgSO_4$ and the solvent removed. Flash column chromatography (0-20% MeOH in CH_2Cl_2 over 15 column volumes (CV) in a 10 g SiO_2 column) afforded the target

compound (291 mg, 51%, 0.64 mmol for the *L*-Pro-*L*-Glu derivative; 213 mg, 36%, 0.45 mmol for the *L*-Piperidine-*L*-Glu derivative; 300 mg, 56%, 0.7 mmol for the *L*-Ala-*L*-Glu derivative; 312 mg, 57%, 0.71 mmol for the *L*-Azetidine-*L*-Glu derivative).

Boc-*L*-Ala-*L*-Glu(di-tBu-ester) ^{1}H NMR (400 MHz, $CDCl_3$) δ_H = 6.72 (d, J=7.9, 1H, 2), 5.16 – 4.99 (m, 1H, 22), 4.47 – 4.39 (m, 1H, 4), 4.21 – 4.06 (m, 1H, 21), 2.36 – 2.14 (m, 2H, 7), 2.15 – 2.00 (m, 1H, 6”), 1.98 – 1.80 (m, 1H, 6’), 1.44 (s, 9H, 27, 28, 29), 1.42 (s, 9H, 11, 12, 13), 1.41 (s, 9H, 18, 19, 20), 1.33 (d, J=7.1, 3H, 23); ^{13}C NMR (101 MHz, $CDCl_3$) δ_C = (C1), 172.3 (C8), 170.9 (C5), 155.5 (C24), 82.4 (C17), 80.8 (C10), 80.1 (C26), 53.5 (C4), 52.3 (C21), 31.5 (C7), 28.4 (C18, C19, C20), 28.2 (C27, C28, C29), 28.1 (C11, C12, C13), 27.7 (C6), 18.5 (C23); ESI-MS, $[M+H]^+$: 431.28 (calculated), 453.20 (observed, obtained as Na^+ adduct); HRMS (ESI+) $(M+H)^+$: 431.27518 (calculated for $C_{21}H_{39}N_2O_7$), 431.27505 (observed).

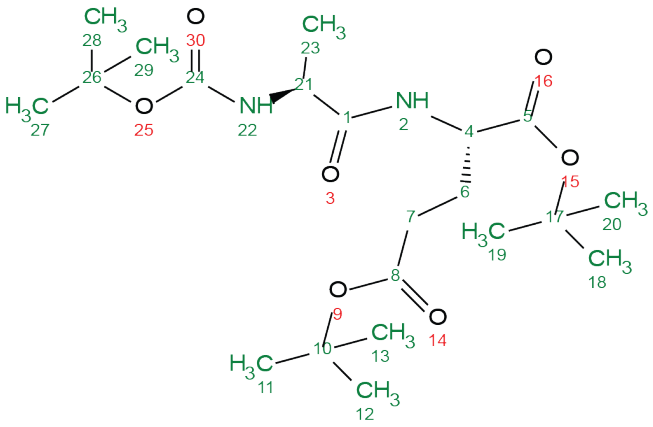


Figure 2.37 | *L*-Ala-*L*-Glu dipeptide with Boc and tert-butyl protection.

Boc-*L*-Pro-*L*-Glu(di-tBu-ester) ^{1}H NMR (400 MHz, $CDCl_3$) δ_H = 6.66 (s, 1H, 2), 4.43 (s, 1H, 4), 4.23 (d, J=29.8, 1H, 21), 3.40 (d, J=55.4, 2H, 24), 2.35 – 2.14 (m, 3H, 7, 22”), 2.14 – 2.02 (m, 1H, 6”), 1.92 – 1.80 (m, 2H, 6’, 22’, 23), 1.45 (s, 18H, 11, 12, 13, 29, 30, 31), 1.42 (s, 9H, 18, 19, 20); ^{13}C NMR (101 MHz, $CDCl_3$) δ_C = 172.2 (C1, C5, C8), 171.0 (C26), 82.3 (C17), 80.8 (C28), 77.4 (C10), 60.1 (C21), 52.0 (C4), 47.1 (C24), 31.5 (C7), 31.2 (C22), 28.4 (C11, C12, C13), 28.2 (C18, C19, C20), 28.1 (C29, C30, C31), 24.8 (C6), 23.9 (C23); ESI-MS, $[M+H]^+$: 457.29 (calculated), 479.20 (observed, obtained as Na^+ adduct); HRMS (ESI+) $(M+H)^+$: 457.29083 (calculated for $C_{23}H_{41}N_2O_7$), 457.29023 (observed).

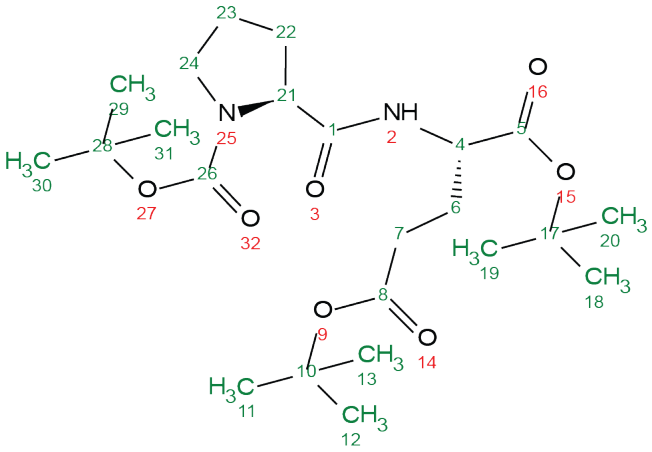


Figure 2.38 | *L*-Pro-*L*-Glu dipeptide with Boc and tert-butyl protection.

Boc-*L*-Piperidine-*L*-Glu(di-tBu-ester) ^{1}H NMR (500 MHz, $CDCl_3$) δ_H = 6.66 (d, J=66.1, 1H, 15), 4.73 (d, J=35.6, 1H, 4), 4.45 (d, J=7.2, 1H, 17), 4.06 (d, J=82.7, 1H, 2'), 2.81 (s, 1H, 2"), 2.36 – 2.23 (m, 1H, 26), 2.20 (dd, J=16.1, 5.7, 1H), 2.16 – 2.05 (m, 1H, 25"), 1.93 – 1.81 (m, 1H, 25'), 1.68 – 1.50 (m, 4H, 1, 5", 6"), 1.48 (s, 8H, 10, 11, 12), 1.46 (s, 2H, 6'), 1.45 (s, 8H, 31, 32, 33), 1.44 (s, 1H, 5'), 1.43 (s, 9H, 21, 22, 23); ^{13}C NMR (126 MHz, $CDCl_3$) δ_C = 172.1 (C18), 171.2 (C14), 170.9 (C27), 156.1 (C7), 82.4 (C20), 80.8 (C30), 80.7 (C9), 54.5 (C4), 52.3 (C17), 42.7 (C2), 31.7 (C26), 28.5 (C31, C32, C33), 28.2 (C21, C22, C23), 28.1 (C10, C11, C12), 27.7 (C25), 25.6 (C5), 25.0 (C1), 20.6 (C6); ESI-MS, $[M+H]^+$: 470.3 (calculated), 493.2 (observed, obtained as Na^+ adduct); HRMS (ESI+) $(M+H)^+$: 471.30648 (calculated for $C_{24}H_{43}N_2O_7$), 471.30673 (observed).

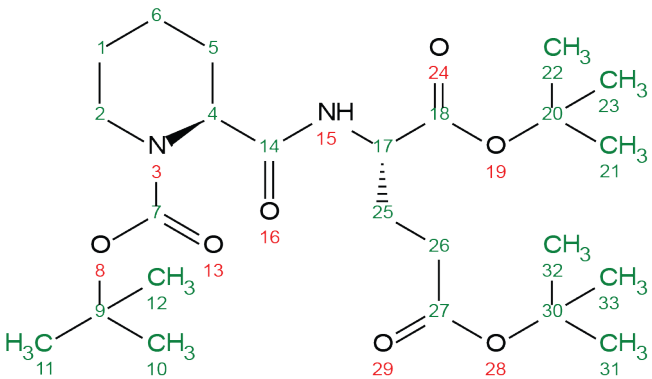


Figure 2.39 | *L*-Piperidine-*L*-Glu dipeptide with Boc and tert-butyl protection.

Boc-*L*-Azetidine-*L*-Glu(di-tBu-ester) ^{1}H NMR (500 MHz, $CDCl_3$) δ_H = 4.62 (t, J=7.9, 1H, 13), 4.47 (td, J=7.9, 5.1, 1H, 3), 3.90 (q, J=8.3, 1H, 15"), 3.80 (td, J=8.3, 6.3, 1H, 15'), 2.41 (s, 2H, 14),

2.34 – 2.20 (m, 2H, 9), 2.18 – 2.07 (m, 1H, 4”), 1.97 – 1.87 (m, 1H, 4’), 1.46 (s, 9H, 18, 19, 20), 1.45 (s, 9H, 22, 23, 24), 1.43 (s, 9H, 28, 29, 30); ^{13}C NMR (126 MHz, $CDCl_3$) δ_C = 172.1 (C1), 171.5 (C5), 170.7 (C10), 157.2 (C25), 82.2 (C17), 81.1 (C21), 80.7 (C27), 62.3 (C13), 52.0 (C3), 47.1 (C15), 31.5 (C9), 28.5 (C14), 28.3 (C18, C19, C20), 28.2 (C28, C29, C30), 28.1 (C22, C23, C24), 27.9 (C4); ESI-MS, $[M+H]^+$: 443.28 (calculated), 465.2 (observed, obtained as Na^+ adduct); HRMS (ESI+) $(M+H)^+$: 443.27518 (calculated for $C_{22}H_{39}N_2O_7$), 443.27518 (observed).

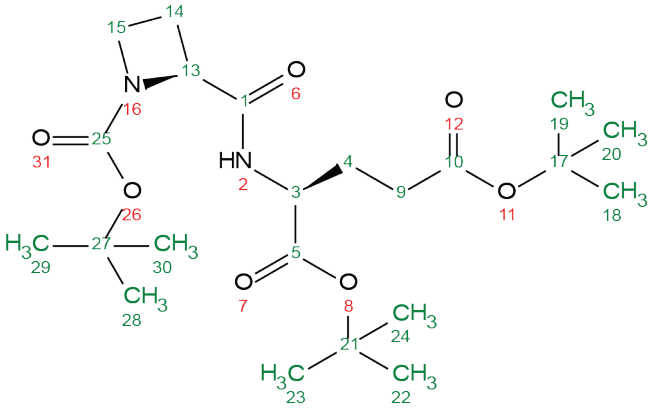


Figure 2.40 | *L*-Azetidine-*L*-Glu dipeptide with Boc and tert-butyl protection.

Deprotection of dipeptides

The deprotection procedure was adapted from [248]. The protected dipeptide was dissolved in CH_2Cl_2 and a 50-fold molar excess of CF_3CO_2H was added. The mixture was stirred at ambient temperature for 2 h. Upon completion of the reaction, the mixture was evaporated to dryness and the oily residue was evaporated three times with toluene. The oily residue was extracted with H_2O/ethyl acetate and the organic phase was washed with 1 M HCl. The aqueous layer was concentrated and purified by HPLC using a ACE5 C18 column, 100 x 21.2 mm employing a 2% isocratic elution at 2% MeCN in H_2O, 0.1% formic acid over 12 min to afford the target compounds (118 mg, 76%, 0.48 mmol for *L*-proline-*L*Glu 133 mg (**2**), 81%, 0.58 mmol for *L*-azetidine-*L*-Glu (**1**); 102 mg, 88%, 0.39 mmol for *L*-piperidine-*L*-Glu (**3**); 116 mg, 76%, 0.53 mmol for *L*-alanine-*L*-Glu (**4**)).

***L*-Ala-*L*-Glu (compound 4)** ^{1}H NMR (400 MHz, MeOD) δ_H = 4.48 (m, 1H, 4), 3.99 (q, 1H, 13), 2.46 (m, 2H, 7), 2.24 (m, 1H, 6’), 1.97 (m, 1H, 6”), 1.54 (dd, 3H, 15); ^{13}C NMR (101 MHz,

MeOD) δ_C = 176.3 (8), 174.8 (5), 171.2 (1), 53.0 (4), 50.1 (13), 31.0 (7), 27.7 (6), 17.5 (15); ESI-MS, $[M+H]^+$: 219.10 (calculated), 219.15 (observed); HRMS (ESI+) $(M+H)^+$: 219.09755 (calculated for $C_8H_{15}N_2O_5$), 219.09781 (observed).

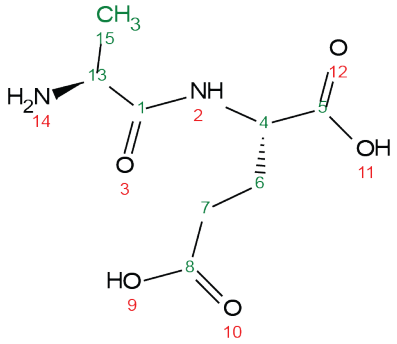


Figure 2.41 | *L*-Ala-*L*-Glu dipeptide (4).

***L*-Piperidine-*L*-Glu (compound 3)** ^{1}H NMR (400 MHz, MeOD) δ_H = 4.49 (dd, J=9.2, 5.0, 1H, 10), 3.83 (dd, J=11.6, 3.3, 1H, 4), 3.44 – 3.36 (m, 1H, 2”), 3.08 – 2.96 (m, 1H, 2’), 2.43 (dd, J=8.0, 6.9, 2H, 15), 2.33 – 2.13 (m, 2H, 5’, 14”), 2.06 – 1.83 (m, 3H, 1”, 6”, 14’), 1.81 – 1.59 (m, 3H, 1’, 5”, 6’); ^{13}C NMR (101 MHz, MeOD) δ_C = 176.3 C(16), 174.3 (C11), 170.2 (C7), 59.0 (C4), 53.0 (C10), 44.9 C(2), 31.1 (C15), 28.5 (C5), 27.7 (C14), 23.0 (C1), 22.8 (C6); ESI-MS, $[M+H]^+$: 259.13 (calculated), 259.15 (observed); HRMS (ESI+) $(M+H)^+$: 259.12885 (calculated for $C_{11}H_{19}N_2O_5$), 259.12887 (observed).

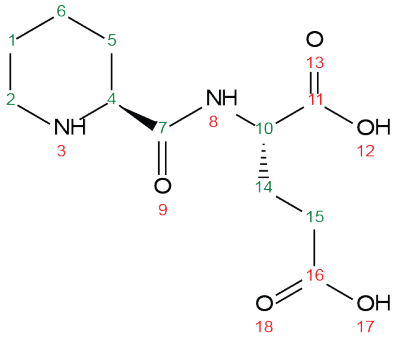


Figure 2.42 | *L*-Piperidine-*L*-Glu dipeptide (3).

***L*-Pro-*L*-Glu (compound 2)** ^{1}H NMR (400 MHz, MeOD) δ_H = 4.48 (dd, J=9.2, 5.0, 1H, 4), 4.36 (dd, J=8.6, 6.3, 1H, 13), 3.47 – 3.33 (m, 2H, 16), 2.52 – 2.38 (m, 3H, 7, 14’), 2.30 – 2.17 (m, 1H, 6’), 2.16 – 1.91 (m, 4H, 6”, 14”, 15); ^{13}C NMR (101 MHz, MeOD) δ_C = 176.2 (C8), 174.3 (C5), 170.0 (C1), 61.0 (C13), 53.3 (C4), 47.5 (C16), 31.1 (C7), 30.9 (C14), 27.6 (C6), 24.9 (C15);

ESI-MS, $[M+H]^+$: 245.11 (calculated), 245.00 (observed); HRMS (ESI+) $(M+H)^+$: 245.11320 (calculated for $C_{10}H_{17}N_2O_5$), 245.11315 (observed).

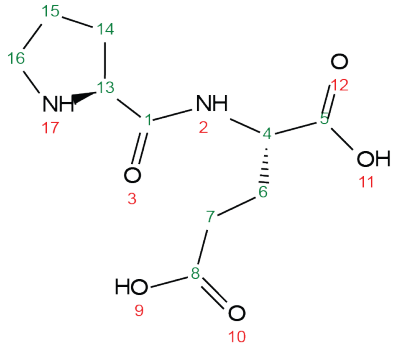


Figure 2.43 | *L*-Pro-*L*-Glu dipeptide (2).

***L*-Azetidine-*L*-Glu (compound 1)** ^{1}H NMR (400 MHz, MeOD) δ_H = 5.04 (t, J=8.4, 1H, 4), 4.51 (dd, J=9.1, 5.0, 1H, 8), 4.12 (q, J=9.2, 1H, 2'), 4.02 – 3.92 (m, 1H, 2"), 2.94 – 2.77 (m, 1H, 1'), 2.69 – 2.51 (m, 1H, 1"), 2.43 (t, J=7.5, 2H, 11), 2.32 – 2.13 (m, 1H, 10'), 2.03 – 1.89 (m, 1H, 10"); ^{13}C NMR (101 MHz, MeOD) δ_C = 174.6 (C12), 173.1 (C9), 169.2 (C5), 59.9 (C4), 53.3 (C8), 45.2 (C2), 30.9 (C11), 27.5 (C10), 24.6 (C1); ESI-MS, $[M+H]^+$: 231.10 (calculated), 231.15 (observed); HRMS (ESI+) $(M+H)^+$: 231.09755 (calculated for $C_9H_{15}N_2O_5$), 231.09762 (observed).

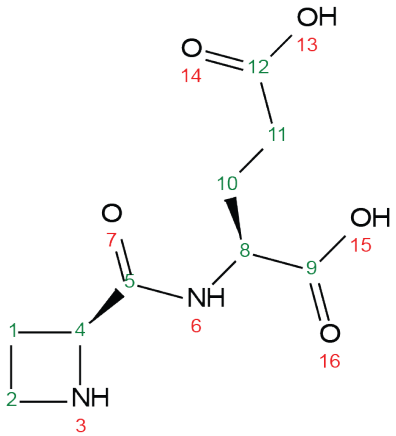


Figure 2.44 | *L*-Azetidine-*L*-Glu dipeptide (1).

Dipeptide reactions with HCHO

The deprotected dipeptide was then used to perform reactions with HCHO (12 h, ambient temperature, H_2O). In general, to obtain HCHO in solution, paraformaldehyde was resuspended in a microwave

vial in the respective solvent (here H_2O) and heated at 95 °C for 3 h in a Biotage Initiator+ under stirring to obtain a colourless solution. 10-fold molar excess of aqueous HCHO (410 µmol) was added to the dipeptide (41 µmol) and NaOD was added to obtain a reaction mixture at pH 10 (pH 12 for compound 4). The solution was neutralized after 12 h and subjected to HPLC purification (isocratic elution, 2% MeCN in H_2O, 0.1% formic acid, 10 min, column ACE5 C18, 100 x 21.2 mm). The product was concentrated to give a colourless solid. Yields were 3 mg, 28.5%, 11.7 µmol for *L*-proline-*L*-Glu-HCHO-adduct (**2a**); 7 mg, 63%, 25.9 µmol for *L*-piperidine-*L*-Glu-HCHO-adduct (**3a**); 6.5 mg, 26%, 21.1 µmol for *L*-alanine-*L*-Glu-HCHO adduct (**4a**); 5 mg for *L*-azetidine-*L*-Glu-HCHO-adduct which was present as 50% hemiaminal (9.6 µmol, 23% yield, (**1b**)), and 50% aminal (10.3 µmol, 25% yield, (**1a**)), the latter of which degraded gradually.

***L*-Ala-*L*-Glu HCHO (compound 4a)** ^{1}H NMR (400 MHz, DMSO) δ_H = 4.42 – 4.36 (m, 1H, 11”), 4.28 – 4.20 (m, 2H, 11’, 17”), 4.13 (d, J=10.8, 1H, 17’), 4.07 (dd, J=11.5, 4.2, 1H, 2), 3.40 (q, J=6.8, 1H, 13), 2.12 – 1.98 (m, 1H, 4’), 1.90 – 1.80 (m, 2H, 5”), 1.67 – 1.55 (m, 1H, 4”), 1.10 (d, J=6.9, 3H, 16); ^{13}C NMR (101 MHz, DMSO) δ_C = 178.4 (C6), 175.4 (C3), 175.1 (C14), 75.4 (C17), 63.0 (C11), 56.70 (C13), 56.5 (C2), 35.3 (C5), 27.0 (C4), 15.9 (C16); ESI-MS, $[M+H]^+$: 261.10 (calculated) 231.10 (calculated for aminal-methylene bridge only, without hemiaminal), 231.10 (observed; hemiaminal group not stable under LC-MS conditions); HRMS (ESI-) $(M-H)^-$: 229.0830 (calculated for $C_9H_{13}N_2O_5^-$), 229.0829 (observed); hemiaminal not stable under HRMS conditions. Note that a few drops of D_2O were added to the compound to aid solubility in DMSO for NMR measurements.

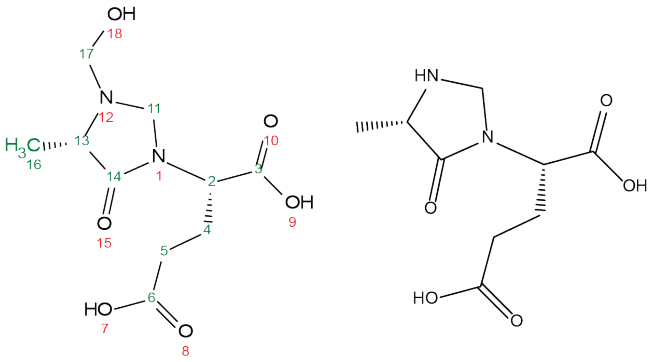


Figure 2.45 | *L*-Ala-*L*-Glu adduct formed by reaction with HCHO. As observed in NMR (left, containing hemiaminal and aminal methylene bridge (**4a**)) and a proposed structure as seen in LC-MS (right, only with aminal methylene bridge (**4b**)) due to instability of the hemiaminal.

***L*-Piperidine-*L*-Glu HCHO (compound 3a)** ^{1}H NMR (700 MHz, DMSO) δ_H = 4.49 (dd, J=11.0, 4.8, 1H, 2), 4.15 (d, J=5.6, 1H, 11”), 3.88 (dd, J=5.6, 2.0, 1H, 11’), 2.85 – 2.76 (m, 1H, 15”), 2.74 – 2.69 (m, 1H, 13), 2.42 – 2.35 (m, 1H, 15’), 2.27 – 2.15 (m, 2H, 7), 2.14 – 2.06 (m, 1H, 6’), 1.90 – 1.80 (m, 1H, 6”), 1.79 – 1.73 (m, 1H, 18”), 1.64 – 1.56 (m, 1H, 16”), 1.55 – 1.45 (m, 2H, 16’, 17’), 1.44 – 1.36 (m, 1H, 18’), 1.36 – 1.28 (m, 1H, 17”); ^{13}C NMR (101 MHz, DMSO) δ_C = 173.4 (C8), 173.0 (C14), 171.7 (C3), 64.7 (C11), 62.6 (C13), 52.1 (C2), 48.9 (C15), 30.1 (C7), 24.3 (C6), 23.9 (C16), 23.6 (C18), 22.6 (C17); ESI-MS, $[M+H]^+$: 271.12 (calculated), 271.15 (observed); HRMS (ESI+) $(M+H)^+$: 271.1288 (calculated for $C_{12}H_{19}N_2O_5$), 271.1288 (observed).

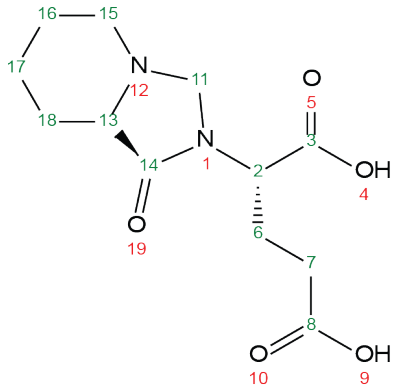


Figure 2.46 | *L*-Pip-*L*-Glu adduct formed by reaction with HCHO (3a).

***L*-Pro-*L*-Glu HCHO (compound 2a)** ^{1}H NMR (600 MHz, D_2O) δ_H = 5.27 (dd, J=8.9, 0.9, 1H, 8’), 4.84 – 4.83 (m, 4H, 1), 4.82 – 4.81 (m, 1H, 8”), 4.53 (dd, J=10.8, 4.8, 1H, 10), 3.96 – 3.84 (m, 1H, 3’),

3.56 – 3.45 (m, 1H, 3”), 2.47 – 2.41 (m, 3H, 5”, 15), 2.37 – 2.31 (m, 2H, 5’, 14’), 2.20 – 2.13 (m, 2H, 4), 2.11 – 2.03 (m, 1H, 14”); ^{13}C NMR (151 MHz, D_2O) δ_C = 177.6 (C16), 174.5 (C11), 170.0 (C6), 66.0 (C1), 65.9 (C8), 57.9 (C3), 56.8 (C10), 31.6 (C15), 27.4 (C5), 24.1 (C4), 24.1 (C14); ^{13}C chemical shifts were referenced using the solvent lock signal; ESI-MS, $[M+H]^+$: 257.11 (calculated), 257.15 (observed); HRMS (ESI+) $(M+H)^+$: 257.1132 (calculated for $C_{11}H_{17}N_2O_5^+$), 257.1133 (observed).

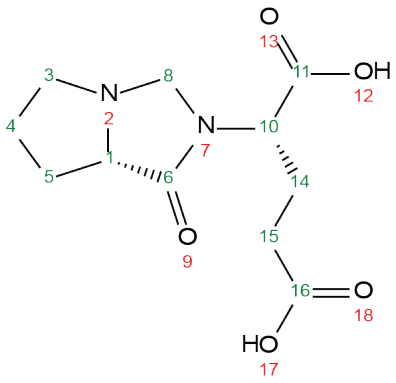


Figure 2.47 | *L*-Pro-*L*-Glu adduct formed by reaction with HCHO (2a).

***L*-Azetidine-*L*-Glu HCHO - Aminal (compound 1a)** ^{1}H NMR (400 MHz, DMSO) δ_H = 4.33 – 4.21 (m, 1H, 4), 4.09 – 3.91 (m, 1H, 15”), 3.76 (s, 1H, 13), 3.36 (q, J=7.1, 6.4, 1H, 17”), 3.31 (s, 1H), 3.14 – 3.03 (m, 1H, 17’), 2.32 – 2.12 (m, 3H, 7, 16”), 2.10 – 1.94 (m, 2H, 6’, 16’), 1.92 – 1.78 (m, 1H, 6”); ^{13}C NMR (101 MHz, DMSO) δ_C = 173.7 (C8), 172.9 (C5), 163.0 (C1), 74.6 (C15), 63.9 (C13), 50.9 (C4), 50.7 (C17), 30.0 (C7), 26.7 (C6), 21.7 (C16); ESI-MS, $[M+H]^+$: 243.10 (calculated), 243.0 (observed); HRMS (ESI+) $(M+H)^+$: 243.0981 (calculated for $C_{10}H_{15}N_2O_5^+$), 243.0984 (observed).

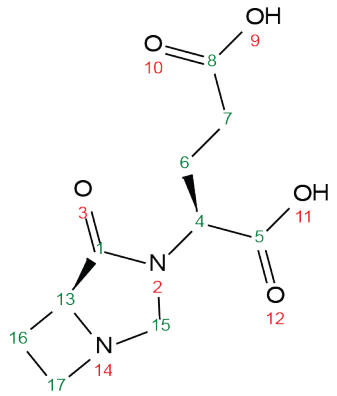


Figure 2.48 | *L*-Azetidine-*L*-Glu adduct formed by reaction with HCHO. This unstable aminal form (**1a**) gradually degrades to a hemiaminal (see below).

***L*-Azetidine-*L*-Glu HCHO - Hemiaminal (compound 1b)** ^{1}H NMR (400 MHz, DMSO) δ_H = 4.31 – 4.18 (m, 1H, 4), 3.67 (q, J=7.4, 6.6, 1H, 13), 3.36 – 3.30 (m, 1H, 17”), 3.25 (s, 1H, 18), 3.06 (q, J=8.0, 1H, 17’), 2.34 – 2.23 (m, 2H, 7), 2.23 – 2.12 (m, 1H, 16”), 2.11 – 1.96 (m, 2H, 6’, 16’), 1.93 – 1.78 (m, 1H, 6”); ^{13}C NMR (101 MHz, DMSO) δ_C = 173.8 (C8), 173.0 (C5), 171.6 (C1), 79.8 (C18), 63.9 (C13), 50.9 (C4), 50.4 (C17), 30.1 (C7), 26.8 (C6), 21.8 (C16); ESI-MS, $[M+H]^+$: 261.10 (calculated), 243.0 (observed; hemiaminal group not stable under LC-MS conditions); HRMS (ESI-) $(M-H)^-$: 259.0936 (calculated for $C_{10}H_{15}N_2O_6^-$), 259.0928 (observed).

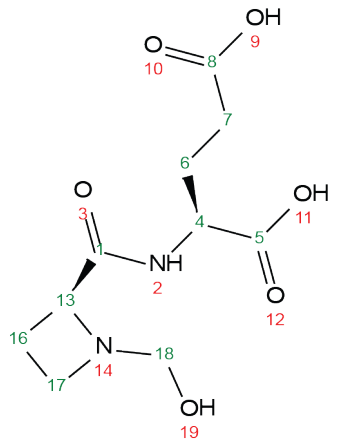


Figure 2.49 | *L*-Azetidine-*L*-Glu adduct formed by reaction with HCHO as its hemiaminal form (1b).

Reductive methylation

The procedure for reductive methylation was changed after [89]. *L*-Pro-*L*-Glu dipeptide (82 µmol) was dissolved in H_2O, HCHO (10 equivalents, 820 µmol) and $NaBH_3CN$ (3 equivalents, 246 µmol) were added. The mixture was basified to pH 9 with 1 M NaOH and stirred for 3 h at 40 °C. Upon completion of the reaction, the mixture was washed with Et_2O and the $NaBH_3CN$ in the aqueous layer was quenched with NH_4Cl. The aqueous phase was purified by flash column chromatography with a solvent gradient of 0-50% MeCN (0.5% formic acid) in H_2O (0.5% formic acid) over 15 column volumes in a 12 g Biotage SNAP KP-C18-HS. The solvent was removed *in vacuo* to give a colourless solid, which necessitated further purification by HPLC. HPLC purification was done with H_2O and MeCN (with 0.1% CF_3CO_2H) using an elution gradient of 1-10% over 15 minutes,

then 10-50% over 8 minutes (Sunfire C18 5 µm 10x150 mm column). The product was concentrated to give a colourless solid. Yield: 16 mg, 76%, 62 µmol.

***L*-Pro-*L*-Glu - reductive methylation** ^{1}H NMR (400 MHz, D_2O) δ_H = 4.50 (dd, J=9.0, 5.3, 1H, 4), 4.23 (dd, J=9.2, 7.6, 1H, 13), 3.86 – 3.72 (m, 1H, 16'), 3.31 – 3.19 (m, 1H, 16"), 2.96 (s, 3H, 18), 2.70 – 2.57 (m, 1H, 14"), 2.53 (t, J=7.2, 2H, 7), 2.32 – 2.20 (m, 2H, 6', 15'), 2.19 – 2.11 (m, 1H, 14'), 2.11 – 1.99 (m, 1H, 6", 15"); ^{13}C NMR (101 MHz, D_2O) δ_C = 177.0 (C8), 174.5 (C5), 168.5 (C1), 68.6 (C13), 56.5 (C16), 52.6 (C4), 40.3 (C18), 30.1 (C7), 29.1 (C14), 25.6 (C6), 22.5 (C15); ESI-MS, $[M+H]^+$: 259.13 (calculated), 259.15 (observed).

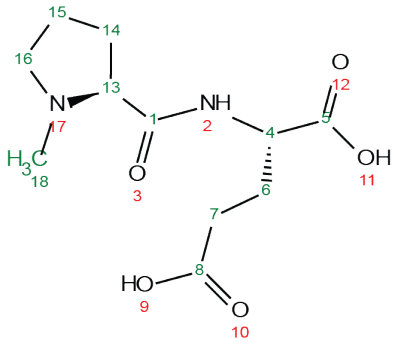


Figure 2.50 | *L*-Pro-*L*-Glu after reductive methylation.

2.6.4 NMR of PEPAK pentapeptide and HCHO

To have another approach in addition to MS/MS to show that the reaction with HCHO takes place on the N-terminus of H2B, 1 mM H2B pentapeptide (24 µl of 22 mM stock in H_2O) was mixed with 25 mM H^{13}CHO (126 µl of a 100 mM stock in H_2O) and D_2O (510 µl total reaction mixture volume after adjusting the pD to 9). The mixture was left for 24 h at ambient temperature and analysed by NMR, whilst comparing the results with unreacted pentapeptide (Figure 2.4 a-c).

NH-PEPAK-NH_2 Pentapeptide This peptide was prepared on solid phase as described in Chapter 2.6.2. ^{1}H NMR (600 MHz, D_2O) δ_H = 4.77 – 4.75 (m, 1H, 2), 4.48 – 4.43 (m, 2H, 11, 33), 4.36 – 4.27 (m, 2H, 18, 23), 3.90 – 3.83 (m, 1H, 12"), 3.77 – 3.70 (m, 1H, 12'), 3.51 – 3.38 (m, 2H, 35', 35"), 3.03 (q, J=6.9, 6.2, 2H, 28), 2.61 – 2.46 (m, 3H, 5, 36'), 2.40 – 2.32 (m, 1H, 14'), 2.24 – 2.16 (m,

1H, 37”), 2.13 – 2.03 (m, 5H, 4, 13, 36”), 2.01 – 1.92 (m, 2H, 14”, 37’), 1.91 – 1.84 (m, 1H, 25”), 1.83 – 1.77 (m, 1H, 25’), 1.76 – 1.69 (m, 2H, 27), 1.56 – 1.50 (m, 1H, 26”), 1.49 – 1.46 (m, 1H, 26’), 1.44 (d, J=7.2, 3H, 20); ^{13}C NMR (151 MHz, D_2O) δ_C = 177.3 (C6), 176.4 (C24), 175.0 (C19), 173.7 (C15), 171.2 (C34), 169.6 (C3), 60.4 (C33), 59.5 (C11), 53.2 (C23), 51.3 (C2), 49.6 (C18), 48.0 (C12), 46.5 (C35), 39.2 (C28), 30.4 (C25), 29.8 (C5), 29.7 (C36), 29.4 (C14), 26.2 (C27), 25.5 (C37), 24.7 (C4), 23.7 (C13), 22.0 (C26), 16.4 (C20); ESI-MS, $[M+H]^{2+}$: 270.66 (calculated), 270.75 (observed); HRMS (ESI+) (M+H)+: 540.3140 (calculated for $C_{24}H_{42}N_7O_7^+$), 540.3139 (observed).

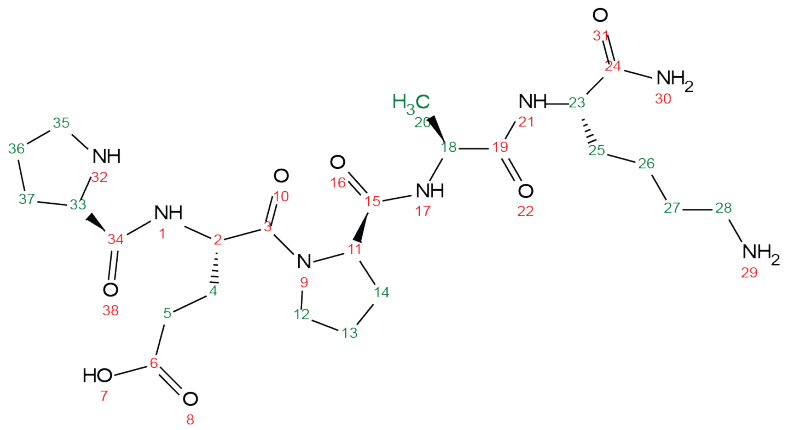


Figure 2.51 | Pentapeptide resembling H2B N-terminus as synthesized on peptide synthesizer.

NH-PEPAK-NH_2 pentapeptide HCHO adduct

300 µl of a 22 mM stock of the pentapeptide (6.78 umol, 3.6 mg) were reacted with a 100-fold excess of HCHO (678 µmol, 678 µl of a 1 M stock solution) at pH 9 for 12 h. The mixture was then HPLC purified using an ACE5 C18 column, 100 x 21.2 mm employing an 2%-40% MeCN gradient in H_2O, 0.1% formic acid over 30 min. The product was concentrated to give a colourless solid. Yields were 1.4 mg, 2.55 µmol, 38%.

^{1}H NMR (600 MHz, D_2O) δ_H = 5.2 (d, J=8.95, 1H, 32’), 5.2 – 5.1 (m, 1H, 2), 4.9 (d, J=8.99, 1H, 32”), 4.7 (dd, J=9.30, 4.54, 1H, 34), 4.4 (dd, J=8.39, 6.04, 1H, 11), 4.3 – 4.3 (m, 3H, 18, 23), 3.9 – 3.7 (m, 3H, 12, 36’), 3.4 – 3.3 (m, 1H, 36”), 3.0 (q, J=7.73, 3H, 28), 2.5 – 2.4 (m, 2H, 5, 38”), 2.4 – 2.3 (m, 2H, 14’, 38’), 2.2 – 2.2 (m, 1H, 4’), 2.1 – 2.0 (m, 4H, 4”, 13, 37), 2.0 – 1.9 (m, 1H, 14”), 1.9 – 1.8 (m, 1H, 25”), 1.8 – 1.8 (m, 1H, 25’), 1.7 – 1.7 (m, 2H, 27), 1.5 – 1.5 (m, 1H, 26), 1.4 (d, J=7.22,

4H, 20); ^{13}C NMR (151 MHz, D_2O) δ_C = 177.7 (C6), 176.5 (C19), 175.0 (C24), 170.7 (C35), 168.7 (C3), 168.1 (C15), 65.6 (C34), 65.6 (C32), 60.3 (C11), 57.5 (C36), 53.2 (C23), 52.8 (C2), 49.6 (C18), 48.4 (C12), 39.2 (C28), 30.8 (C5), 30.4 (C25), 29.5 (C14), 27.3 (C38), 26.2 (C27), 24.6 (C13), 24.3 (C37), 24.0 (C4), 22.1 (C26), 16.3 (C20); ESI-MS, $[M+H]^{2+}$: 276.66 (calculated), 276.80 (observed); HRMS (ESI+) $(M+H)^+$: 552.31402 (calculated for $C_{25}H_{42}N_7O_7^+$), 552.31393 (observed).

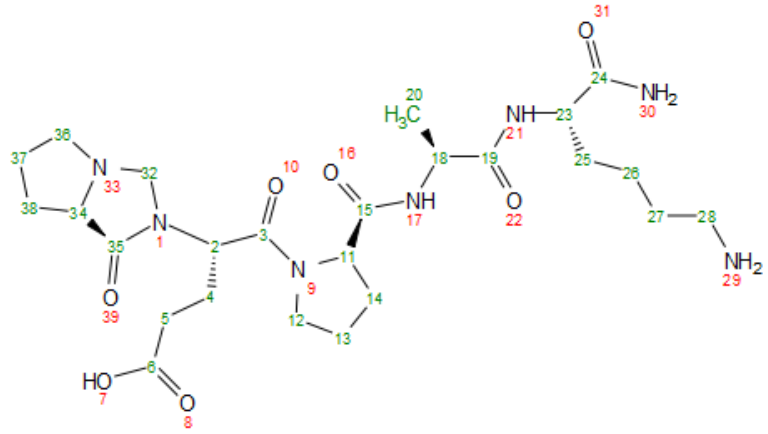


Figure 2.52 | Pentapeptide resulting from the reaction with HCHO.

NMR time course and stability – dipeptides

The dipeptide (23.3 µmol) was dissolved in D_2O; the pD was adjusted to 10 (corresponding to pH 9.6) using NaOD and DCl (total volume: 1 ml, 23.3 mM stock). HCHO (30.0 mg, 1.0 mmol) was suspended with D_2O in a sealed microwave vial and heated until a clear colourless solution was obtained; subsequently the pD was adjusted with NaOD and DCl to 10 (total volume: 3 ml, 1/3 M stock). For time course measurements, the dipeptide stock (100 µL, 2.33 µmol) at pD 10 and D_2O (400 µl) were mixed and recorded. Subsequently, HCHO (70 µl, 23.3 µmol, 10 equivalents) at pD 10 was added and progress was monitored over the indicated time. To see if the *in situ* formed HCHO adducts are chemically stable, 40-fold excess with respect to starting material of the HCHO scavenger 1,3-cyclohexanedione (93.2 µl, 93.2 µM of a 1 M stock in D_2O) was added. Additionally, stability studies of the HPLC purified **2a** and the corresponding pentapeptide PEPAK-HCHO adduct were conducted. To achieve a 40-fold excess of HCHO scavengers (46.8 µmol) with respect to peptide starting material, the appropriate volumes of GSH (186.4 µl of a 250 mM stock), cysteine (46.6 µl

of a 1 M stock) or 1,3-cyclohexanedione (46.6 µl of a 1 M stock) were added to the peptide-HCHO adducts (50 µl, 1.17 µmol, 23.3 mM stock in D_2O) in 400 µl D_2O, thoroughly mixed and recorded. To test for stability of the adduct in different pH conditions, HPLC purified **2a** (45 µl of a 23.3 mM stock) was dissolved in 405 µl 50 mM potassium phosphate buffer (D_2O) at pH 5.9, 7.4 and 8.5 (pD 6.3, 7.8 and 8.9), thoroughly mixed and recorded (Figure 2.26 a, 6.-7.).

NMR competition experiments

For equimolar competitions (Figures 2.30, 2.31), 400 µl of D_2O were mixed with component 1 (50 µl, 1.17 µmol, 23.3 mM stock) and component 2 (50 µl, 1.17 µmol, 23.3 mM stock) and recorded. If indicated, equimolar amounts of HCHO were added (6 µl of a 0.2 M stock were used). To test whether **2a** is stable in higher temperatures, the mixture in the NMR tube was heated for 15 min at 75 °C (Figure 2.26 a, 5.).

Reaction of compound 2 with other carbonyl compounds

To assess whether **2** reacts with carbonyl compounds other than HCHO, 81 µmol **2** was mixed with up to 100 equivalents carbonyl compound at pH 10 and reacted overnight at ambient temperature, 37 °C and 75 °C. The progress of the reaction was tested by LC-MS; as only acetaldehyde showed reactivity, only the ProGlu-acetaldehyde adduct was HPLC purified. However, upon drying the sample, and despite multiple attempts, conversion to the starting material (**2**) was always observed.

2.6.5 MALDI MS assays

MALDI MS measurements were carried out using a MALDI Autoflex Speed, Bruker machine. For preparation of CHCA (α-cyano-4-hydroxycinnamic acid) matrix, CHCA was dissolved in MQ water/MeCN (1:1), 0.1% CF_3CO_2H (1 mg CHCA per 100 µl mixture) and sonicated. The sample (1 µl) was then spotted with 1 µl of saturated CHCA solution on a ground steel target (Bruker) and analysed. HCHO titration experiments were conducted at pH 5.6, 7.4, or 8.2 in 5 µl potassium phosphate buffer (50 mM), 2.5 µM peptide concentration (2.5 µl of a 10 µM stock in MQ water), 2.5 µM/25 µM/250 µM/2.5 mM/25 mM HCHO or urea compound (from equimolar to 10^4-fold

excess; 2.5 µl of a 10 µM/100 µM/1 mM/10 mM stock in MQ water). All reactions were performed in 10 µl final reaction volume in a 384 well plate (Microplate, 384 well, PP, V-bottom, 781280, Greiner) at ambient temperature or if indicated at 37 °C, which was sealed (StarSeal Sealing Tape Polyolefin Film, E2796-9793, StarLab). Samples were taken after 0 h (directly after adding HCHO), 1 h, 2 h, 8 h and 24 h. For the reductive methylation reactions, 2.5 µl 20 mM potassium phosphate buffer, pH 7.4, 2.5 µl peptide (10 µM stock in MQ water), 2.5 µl HCHO (100 µM/200 µM/1 mM stock in MQ water), and 2.5 µl $NaCNBH_3$ (200 µM/400 µM/2 mM µM stock in buffer) were used. Exemplary raw data are shown in Figure 2.16.

MALDI top-down MS/MS

For MALDI top-down MS/MS (Figures 2.3 e+f, 2.9 c+d), H2B peptide (2.5 µM final concentration) was reacted with a 10^3 excess HCHO for 12 h in MQ water (10 µl reaction volume) at ambient temperature, then spotted on a ground steel plate with CHCA (for preparation of CHCA see Chapter 2.6.5). Control conditions without HCHO were: H2B 15mer (100 µM) in MQ water (10 µl reaction volume), then spotted on a ground steel plate with CHCA. Note MQ water was chosen because peptides fragmented better than with the previously used potassium phosphate buffer.

2.6.6 KDM4E catalysed *in situ* formation of HCHO – reaction with H2B 15mer

The reaction shown in Table 2.1 was conducted at ambient temperature and was monitored at 0 h (directly after starting the reaction), 1 h, 2 h, 8 h and 24 h (only the 24 h time point is shown in Figure 2.7 a-d) using CHCA matrix and MALDI MS as described above in Chapter 2.6.5. KDM4E was prepared by Eidarus Salah.

Table 2.1: KDM4E generated HCHO reaction with H2B 15mer.

Component	Stock	Volume [µl]	Final Conc.
KDM4E	88 µM in 150 mM NaCl and 50 mM HEPES	2	8.8 µM
Sodium Ascorbate	10 mM	2	1 mM

continued . . .

. . . continued

Buffer	**Components**	**Final Concentration**	
Ammonium Ferrous Sulphate	1 mM in 400 µM HCl	2	100 µM
H2B 1-15 peptide	8 mM	3	1200 µM
HEPES	100 mM, pH 7.4	8	∼40 mM
2OG **or** NOG control	8 mM	2	800 µM
H3K9Me$_3$ 1-21 peptide	10 mM	1	500 µM
		20 µl total volume	

Along with MALDI MS, the reaction after 24 h described in Table 2.1 was subjected to Orbitrab Elite LC-MS/MS, which was performed by Elisabete Pires (Figure 2.28). Peptide analyses were performed using a Thermo Scientific Dionex Ultimate 3000 chromatography system coupled directly to a Orbitrap Elite mass spectrometer (Thermo Scientific, San Jose, CA) with a HESI II electrospray ionisation source. A 5 µL partial loop injection was used for all analyses and the chromatographic separation was performed using a Chromolith FastGradient RP-18, 2.0 × 50 mm, 1.6 µm particle size column. The LC flow rate was 0.300 ml/min. The peptides were separated with a linear gradient of 5-95% Buffer B (MeCN, 0.1% formic acid) / Buffer A (MQ water, 0.1% formic acid) during 10 minutes. Analysis was performed in positive ion mode using a scan ranges from 410-2000 and resolution set to 15,000. (AGC target 1e6, maximum injection time 250 ms) and subsequent HCD MS/MS spectra (AGC target 5e4, maximum injection time 100 ms) of the parent mass m/z 491.920 and 503.92 peaks were acquired in the FTMS. HCD fragmentation was performed at 35% of normalized collision energy. All spectra were manually checked and validated or disqualified. The analysis was performed with Xcalibur software.

2.6.7 Tissue culture

Cells were from American Type Culture Collection (ATCC; HEK293T ref. CRL-3216). The authentication from ATCC was relied upon, none of the cell lines were further authenticated. Cell lines were regularly tested to be mycoplasma negative (MycoAlert Mycoplasma Detection Kit ref. no LT07-218). Cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM, D6546, Sigma)

supplemented with 10% FBS (Gibco) and 1% GlutaMAX (2mM final concentration, Gibco) in a 37 °C incubator at 5% CO_2. For harvesting the cells, the medium was aspirated and cells were washed twice with phosphate buffered saline (PBS, Sigma, D8537). Upon aspirating PBS, radioimmunoprecipitation assay buffer (1x RIPA [Sigma, R0278], supplemented with protease inhibitor [cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Roche]), was used to scrape the cells. The cell suspension in the reaction tube was either frozen at -20 °C or incubated for 45 min on ice and vortexed several times during that time interval. After centrifuging (16,000 g, 15 min, 4 °C), the supernatant was used for analyses.

2.6.8 SDS-PAGE

Polyacrylamide gel electrophoresis was performed using NuPAGE 4-12% Bis-Tris Protein Gels (Life Tech) in 1x MOPS running buffer (20x NuPAGE MOPS SDS Running Buffer, NP0001). SDS-PAGE was run at 170 V for 40 min (Mini Gel Tank, Life Technologies). The PageRuler Prestained Protein Ladder (Thermo Scientific, reference 26617) was used to compare protein sizes.

2.6.9 Western blot

Proteins resolved by SDS-PAGE were transferred onto a nitrocellulose membrane (Amersham Protran Premium 0.2 NC 300 mm, GE Healthcare) using a current of 350 mA for 60-90 min or 85 mA overnight (Mini Protean Tetra Cell, Bio-Rad) in transfer buffer containing 10% methanol (20x NuPAGE Transfer Buffer, Invitrogen). Membranes were blocked using 5% milk (w/v) in PBS 1‰ tween (PBST; Tween-20, Sigma-Aldrich) at ambient temperature for at least 30 min. The membranes were then incubated with the appropriate primary antibody (1:1000) in 5% milk (w/v) PBST at 4 °C overnight on a shaker. After three 10-min washes with PBST, the blots were incubated with the appropriate secondary antibody (in 5% milk PBS-Tween in a 1:5000 dilution) for 1 h at ambient temperature followed by three 10-min washes with PBST and detection using GE Healthcare Amersham ECL Prime Western Blotting Detection Reagent (RPN2236) and Bio-Rad Universal Hood iii. Primary antibodies used were GAPDH Loading Control Monoclonal Antibody (GA1R, Invitrogen, MA5-15738), H2B (ab52484), H2A (EPR17470), H3 (ab10799), leukotriene A4 hydrolase (ab133512).

Secondary antibodies used were anti-mouse IgG, HRP-linked Antibody (Cell Signaling, reference 7076), anti-Rabbit IgG (H+L) HRP Conjugate (Promega reference W4011).

2.6.10 RNA-Seq analysis

HEK293T cells ($0.9*10^6$) were seeded into 6 cm dishes as biological triplicates. The next day, cells were treated with 150 µM HCHO (or MQ water control), followed by another cumulative treatment of 225 µM HCHO (or MQ water control) the day after that. Cells were harvested 4 h after the second treatment. For harvesting, cells were washed twice with ice-cold PBS (Sigma, D8537); RNA was isolated using the PureLink™ RNA Mini Kit (Invitrogen, 12183020) following the manufacturer's recommendations on "purification from plant and animal cells", "monolayer cells", "syringe homogenization", and "on-column DNase treatment" (PureLink DNase Set, Invitrogen, 12185010). Samples were sent for RNA-Seq PolyA to the Oxford Genomics Centre.

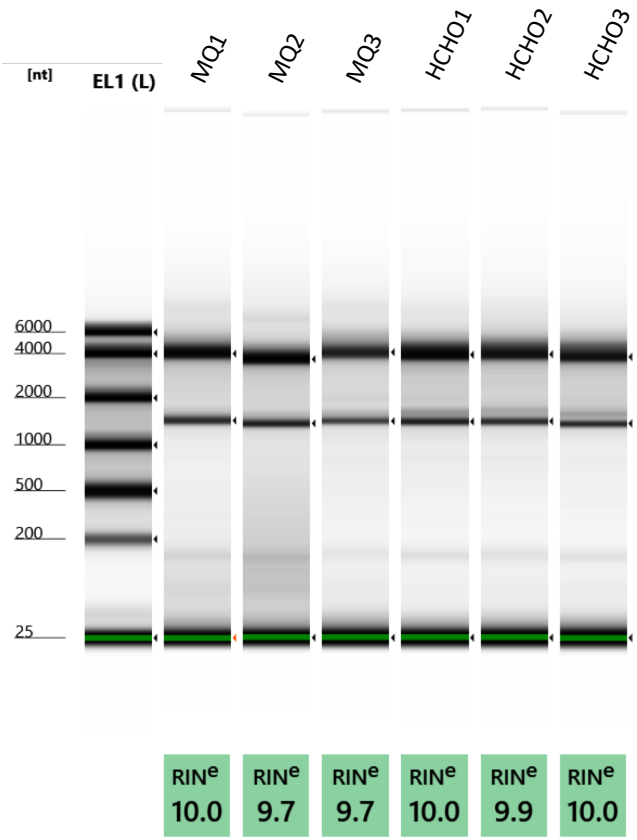


Figure 2.53 | Gel electrophoresis showing isolated RNA. Ribosomal S18 and S28 RNA are shown for all six samples, along with RNA integrity number (RIN). Gel was produced and RIN estimated by the Oxford Genomics centre.

For transcript-level expression analysis, reads were mapped for each sample to the reference genome (`ftp.ensembl.org/pub/release-77/gtf/homo_sapiens/Homo_sapiens.GRCh38.77.gtf.gz`) using HISAT[249] and the resulting SAM files were sorted and converted to BAM files using SAMtools.[250] The resulting BAM files were merged with Stringtie[251] and transcript abundances were estimated and table counts created for Ballgown[252] using the differential expression analysis protocol in [253].

2.6.11 Cloning and biosynthesis of 4-OT

The open reading frame of 4-OT from *Pseudomonas putida* was purchased from GenScript, was codon-optimized for *E. coli*, and provided in a plasmid cloning vector. The sequence of the purchased gene is given below and is flanked by NdeI on the 5' end and by BamHI on the 3' end.

5'-CATATGCCGATTGCGCAGATCCACATCCTGGAAGGTCGTAGCGATGAACAGAA
AGAAACCCTGATTCGTGAAGTTAGCGAGGCGATTAGCCGCAGCCTGGATGCGCCGCT
GACCAGCGTGCGTGTTATCATTACCGAGATGGCGAAGGGTCATTTTGGTATCGGTGGT
GAACTGGCGAGCAAAGTGCGTCGTTAAGGATCC-3'

The DNA encoding for 4-OT was subcloned into a pET-22b(+) vector using NdeI and BamHI restriction sites and transformed into XL10-Gold cells (NEB, C3040H). Clones were sequence verified (Eurofins) and the expression plasmid was transformed into the host strain (BL21(DE3) competent *E. coli* (NEB, C2527H)). Several colonies were combined to generate a glycerol stock. 2TY (+50 µg/ml ampicillin) was inoculated with cells from a glycerol stock overnight. The next day, 4x600 ml (+50 µg/ml ampicillin) were inoculated with 1:100 (v/v) overnight culture and left at 37 °C, 180 rpm overnight (no IPTG added, see expression trials in Figure 2.54). Cell harvest and purification are changed after [222]. The next day, cells were centrifuged (4250 g, 30 min, 4 °C), washed with 10 mM NaH_2PO_4, pH 8 (=buffer A) (the 2.4 l medium were divided into six centrifuge tubes and each tube was washed with 50 ml) and centrifuged again using the same conditions. The pellet (19.74 g from 2.4 l culture) was then frozen directly in -80 °C. 5 g pellet were resuspended in 25 ml buffer A and sonicated (10 min, 9.9 s on, 9.9 s off, 38% amplitude). After centrifugation (48384 g, 30 min, 4 °C),

the supernatant was made 1.6 M in $(NH_4)_2SO_4$ by adding the solid salt and gradually dissolving it overnight at 4 °C on a tube roller. After centrifugation (48384 g, 30 min, 4 °C) and filtration (0.45 µm), the cleared lysate was desalted using a HiPrep 26/10 column (GE Healthcare, Vienna, Austria) on an Äkta pure FPLC system (GE Healthcare), which was conditioned with Buffer A. The lysate was then loaded as 2x15 ml with at a flow of 2 ml/min. Buffer A was then used to elute the proteins at a flowrate of 2 ml/min, whilst monitoring UV absorption at 280 nm and collecting relevant fractions (Figure 2.55 a). Next, the protein eluate (without concentration) was loaded onto a 11 ml DEAE Fast Flow ion exchange column (GE Healthcare; self-packed in a XK 16 column from GE Healthcare) previously equilibrated with Buffer A. After loading at 2 ml/min, the column was left for 15 min to allow protein binding to the column. Non-binding proteins were removed by washing the column with Buffer A, as indicated by a stable A280 trace. This was followed by gradient elution for 7 CV from Buffer A to Buffer B (10 mM NaH_2PO_4, 90 mM Na_2SO_4, pH 8) with a flow rate of 2 ml/min. Fractions containing 4-OT were collected in 5 ml tubes using a fraction collector (A205, 4-OT does not contain aromatic amino acids), which was confirmed by tris-tricine-SDS-PAGE (Figure 2.55 b+c, Table 2.2); the mass was determined by solid phase extraction coupled to mass spectrometry (SPE-MS) to be 6811.51 Da (calculated: 6810.84 Da). UV-absorption at 205 nm (Nanodrop) and BCA assay (Thermo Fisher) were used to determine the yield of 4-OT-containing fractions, which were subsequently aliquoted, rapidly frozen and stored at -80 °C. The yield was 5 ml of 5.4 mg/ml (BCA) for 5 g pellet.

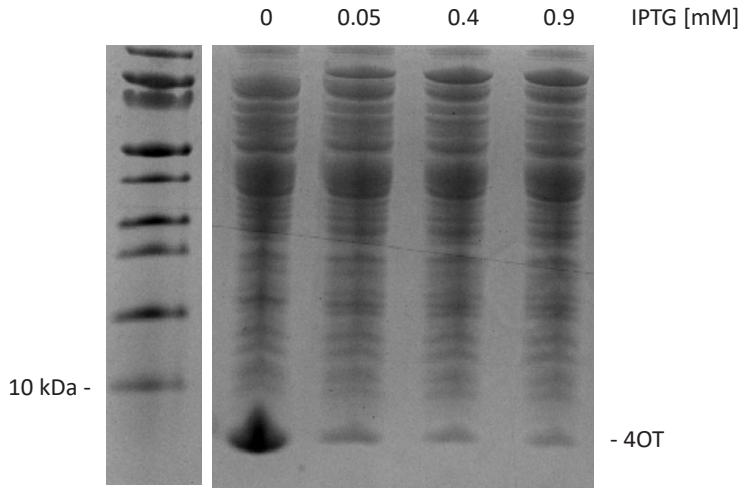


Figure 2.54 | Expression trial of 4-OT in 2TY, 37 °C overnight. Cells were pelleted, sonicated, centrifuged, and the supernatant was loaded onto a tris-tricine gel (see Table 2.2).

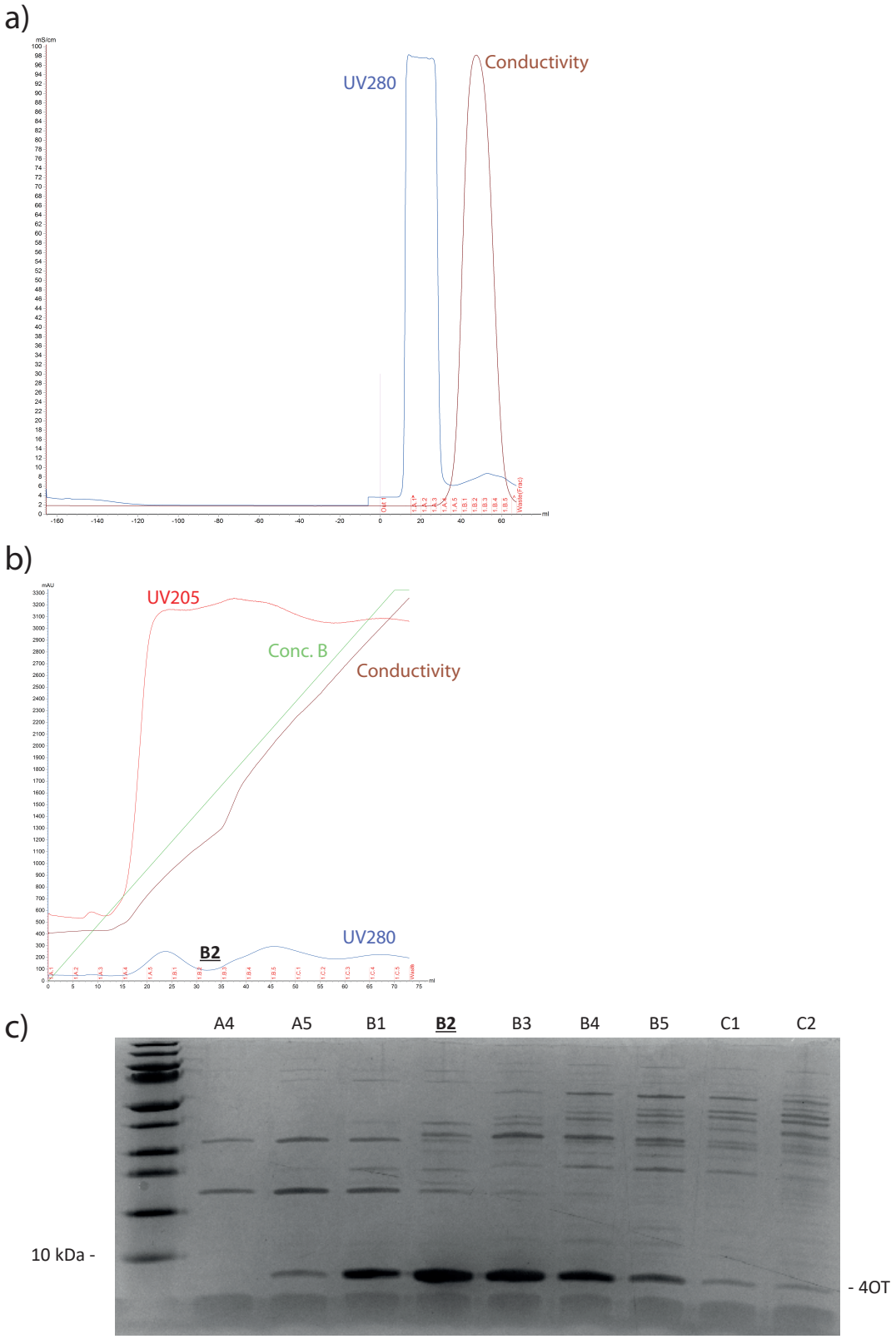


Figure 2.55 | Äkta chromatograms and fractions obtained after Ion Exchange Chromatography. **a)** Chromatogram for desalting column. **b)** Chromatogram for anion exchange column. **c)** Tris-tricine gel (see Table 2.2) showing the fractions containing 4-OT. The bold/underlined fraction B2 was used for further studies.

Table 2.2: Tris-tricine SDS-PAGE gel and buffers (as of `phiellab.com/attachments/TrisTricine.pdf`).

Component	Volume
Separating Gel (10%)	
H_2O	1.75 ml
3 M Tris-HCl pH 8.45/10 mM SDS	2.5 ml
30% Bis-acrylamide 29:1	2.5 ml
Glycerol	0.75 ml
30% APS	7 µl
TEMED	7 µl
Stacking Gel (4%)	
H_2O	1.95 ml
3 M Tris-HCl pH 8.45/10 mM SDS	0.775 ml
30% Bis-acrylamide 29:1	0.4 ml
30% APS	7 µl
TEMED	7 µl
10x Anode Buffer: 2 M tris, pH 8.8	
10x Catode Running Buffer: 1 M tris/1M tricine, pH 8.3, 1% (w/v) SDS	

2.6.12 4-OT NMR and absorbance assays

The Michael addition of acetaldehyde to *trans*-β-nitrostyrene was followed by ^{1}H NMR (700 MHz, Table 2.3) and absorbance measurements (Table 2.4).

Table 2.3: 4-OT NMR assay in 10% DMSO-D_6.

Component	Stock	Volume [µl]	Final Conc.
Potassium phosphate, H_2O	20 mM, pH 7.8	275	
Acetaldehyde	75 mM in buffer	200	20 mM
Trans-β-nitrostyrene	3 mM in DMSO-D_6	75	0.3 mM
4-OT	198.25 µM	200	52.87 µM
		750 µl total volume	10% DMSO-D_6

For the NMR assay shown in Table 2.3, 50 µl 4-OT (793 µM stock) were preincubated at 37 °C for the indicated times with 150 µl of potassium phosphate buffer, 20 mM, pH 7.8 (control condition) or 150 µl of 52.87 mM HCHO in potassium phosphate buffer, resulting in 200 µl of 198.25 µM 4-OT (with a 200-fold molar excess HCHO relative to 4-OT). All components apart from 4-OT (buffer, acetaldehyde, *trans*-β-nitrostyrene) were mixed thoroughly in an NMR tube and a ^{1}H spectrum was recorded (700 MHz). Upon addition of the enzyme, the progress of the reaction was followed over time (298 K).

Table 2.4: 4-OT absorbance assay in 10% DMSO.

Component	**Stock**	**Volume [µl]**	**Final Conc.**
Potassium phosphate	20 mM, pH 7.8	0, 50 or 100	
Acetaldehyde	75 mM in buffer	50	18.75 mM
trans-β-nitrostyrene	0.8 mM in 40% DMSO/60% buffer	50	0.2 mM
4-OT	40 µM in buffer	50	10 µM
		200 µl total volume	10% DMSO

For the absorbance assay shown in Table 2.4, 50 µl 4-OT (40 µM stock, diluted in potassium phosphate buffer, 20 mM, pH 7.8 or pH 7.4) was preincubated at 37 °C for the indicated times with 50 µl potassium phosphate buffer (control condition) or 50 µl 8 mM HCHO stock (prepared in buffer, resulting in a 200-fold molar excess HCHO relative to 4-OT) in a 96-well plate (Corning Clear Flat Bottom, reference 3370). At the end of the indicated incubation times, 0.8 µl of the pre-incubated enzyme was spotted with 0.8 µl CHCA matrix on a ground steel target (Bruker) and analysed by MALDI MS (Figure 2.34 a). This was followed by addition of the two substrates acetaldehyde and *trans*-β-nitrostyrene (in that specific order) into the 96-well plate, and after thorough mixing by pipetting up and down five times, the absorbance spectra were recorded on a ClariostarPlus or Pherastar (BMG Labtech) at ambient temperature (Figures 2.33 f, 2.11 e). For experiments involving higher acetaldehyde concentrations (Figure 2.33 e), the acetaldehyde stocks were adjusted accordingly, while keeping the added volume of acetaldehyde (50 µl) constant. For control experiments without enzyme,

100 µl potassium phosphate was added to 50 µl acetaldehyde and 50 µl *trans*-β-nitrostyrene. Each condition was performed in technical triplicates and in two independent duplicates.

Table 2.5: Kinetic parameters for *trans*-β-nitrostyrene.

Component	**Stock**	**Volume [µl]**	**Final Conc.**
Trans-β-nitrostyrene	4-fold in 40%DMSO/60%buffer	50	15.625-500 µM
Potassium phosphate	20 mM, pH 7.4	50	0.2 mM
Acetaldehyde	200 mM in buffer	50	50 mM
4-OT	40 µM in buffer	50	10 µM
		200 µl total volume	10% DMSO

Table 2.6: Kinetic parameters for acetaldehyde.

Component	**Stock**	**Volume [µl]**	**Final Conc.**
Trans-β-nitrostyrene	1.5 mM in 40%DMSO/60%buffer	50	375 µM
Potassium phosphate	20 mM, pH 7.4	50	0.2 mM
Acetaldehyde	4-fold in buffer	50	5-375 mM
4-OT	40 µM in buffer	50	10 µM
		200 µl total volume	10% DMSO

The kinetic parameters of 4-OT for both substrates *trans*-β-nitrostyrene (Table 2.5) and acetaldehyde (Table 2.6) were determined (Figure 2.33 a-d), and the conditions for acetaldehyde were derived from [254]. Each condition was performed in technical triplicates and in two independent duplicates. The different substrate stocks were all prepared as 4-fold stocks. The experiments were performed in independent duplicates using Greiner UV-star plates (ref. 655801) on a ClariostarPlus (BMG Labtech) at ambient temperature. The slopes of the initial reaction rate were fitted with the Michaelis-Menten equation (Et= 10 µM) using non-linear regression (GraphPad Prism 5) to obtain the K_m and k_{cat} values.

2.6.13 Circular Dichroism (CD) measurements

CD analyses were performed using a Chirascan CD spectrometer (Applied Photophysics) equipped with a Peltier temperature-controlled cell holder. The spectra were obtained ranging from 260-185 nm in 0.5 nm intervals in triplicates at 23 °C. Upon subtracting background signal for buffer only, the spectra were smoothed and averaged using the Savitzky–Golay filter (Window size 4). Upon normalisation of data to the protein concentration, the mean residue ellipticity (MRE) was calculated with the following formula: MRE=θ/(10*l*N*C) [deg*cm^2*dmol^{-1}]; θ: degree of ellipticity, l: path length (0.1 cm), N: number of amino acids, C: Concentration (mol/L). For CD measurements, 100 µl 4-OT (in 20 mM potassium phosphate buffer, pH 7.8, 40 µM) were incubated with 100 µl of 8 mM HCHO (prepared in potassium phosphate buffer, resulting in a 200-fold excess of HCHO) for 48 h at 37 °C (control: 100 µl phosphate buffer), resulting in 20 µM 4-OT (0.136 mg/ml).

2.6.14 Whole-cell assays

2TY (+50 µg/ml ampicillin) was inoculated with BL21(DE3) cells (NEB, C2527H) containing the pET22b-4-OT or a pET22b-EV (empty vector) construct from a glycerol stock overnight. The next morning, an overnight culture was inoculated with 1:100 (v/v) starter culture and left at 37 °C, 180 rpm overnight (no IPTG added). The next morning, samples for SDS-PAGE were withdrawn, supplemented with 2x Laemmli buffer and boiled for 5 min at 95 °C (Figure 2.13 c). Cells were harvested (5000 g, 10 min, ambient temperature) and the cell pellets were resuspended in 20 mM potassium phosphate buffer, pH 7.4 to 5 g CDW/l (CDW= cell dry weight). 60 ml of cell suspension was added to 250 ml Schott Duran glass bottles. HCHO was added to the indicated final concentration from a 100 mM stock in MQ water, and the corresponding volume of MQ water only was added as control. After adding acetaldehyde to 20 mM final concentration (2 M stock in MQ water) and *trans*-β-nitrostyrene to 2 mM final concentration (200 mM stock in EtOH), 800 µl were withdrawn from each condition (corresponding to time point 0). The assay was then conducted at 28 °C with shaking at 150 rpm, each condition in duplicate. Withdrawn samples were centrifuged (13000 g, 4 °C, 5 min) and 50 µl supernatant was then mixed with 150 µl potassium phosphate buffer (20 mM, pH 7.4). Absorbance at

320 nm was measured immediately using Pherastar (BMG Labtech) in a 96-well plate (Flat bottom, Costar, reference 3370) in 5 instrumental replicates.

At the end of the whole-cell assay, the leftover of the duplicate pET22b-4-OT reaction mixture was pooled (100 ml at 2 mM *trans*-β-nitrostyrene at the beginning of the assay, 200 µmol) and centrifuged (5000 g, 10 min, ambient temperature). The supernatant was then extracted and purified after Narancic et al.[244] In brief, the supernatant was extracted with 4x40 ml ethyl acetate. The combined organic layer was dried over $NaSO_4$, filtered and concentrated *in vacuo*. The crude product was purified using 40% EtOAc in cyclohexane over 24 CV (Biotage Sfär Silica D, 5 g) to yield a yellowish oil. Yield: 9.2 mg, 24%, 47.7 µmol. NMR data of 4-nitro-phenylbutanal are in agreement with literature values.[244] The stereochemistry of 4-nitro-phenylbutanal was not determined.

4-Nitro-3-phenylbutanal ^{1}H NMR (400 MHz, $CDCl_3$) δ_H = 9.71 (t, J=1.2, 1H, 11), 7.36 – 7.22 (m, 5H, 4, 5, 6, 7, 8), 4.65 (qd, J=12.5, 7.4, 2H, 9), 4.08 (p, J=7.3, 1H, 2), 3.03 – 2.88 (m, 2H, 1); ^{13}C NMR (101 MHz, $CDCl_3$) δ_C = 198.9 (C10), 138.3 (C3), 129.4 (C5, C7), 128.3 (C6), 127.5 (C4, C8), 79.6 (C1), 46.6 (C9), 38.1 (C2). ESI-MS, [M+H]-: 192.07 (calculated), 192.2 (observed).

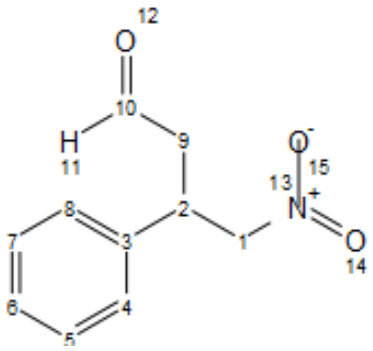


Figure 2.56 | 4-Nitro-3-phenylbutanal.

2.6.15 Growth curve

Multiple colonies from a freshly streaked ampicillin plate with BL21(DE3) cells (NEB, C2527H) containing the pET22b-4-OT construct were resuspended in sterile PBS (Sigma, D8537) to an OD600 of 0.1, which was then diluted 1/10 in PBS. 20 µl of the resulting cell suspension was then added to 180 µl of 2TY (without antibiotic) containing the respective amounts of HCHO or

an equivalent volume of sterile MQ water in a sterile 96-well plate (Flat bottom, Costar, reference 3595), resulting in the final HCHO concentration indicated. The assay was conducted using Pherastar (BMG Labtech) at 37 °C at an absorbance of 600 nm, with shaking before each cycle (double orbital movement) at 300 rpm for 30 s (Figure 2.13 d). The plate was covered with a Star Seal advanced Polyolefin film (E2796-9795) to prevent evaporation.

2.6.16 Immunoprecipitation

Immunoprecipitations were performed using Dynabeads™ Protein G for Immunoprecipitation (Invitrogen), adapted from the manufacturer's recommendations. In brief, the Dynabeads were vortexed in the vial for 30 s to resuspend the magnetic beads. The bead suspension (50 µl) was transferred to a 1.5 ml Eppendorf tube, which was placed on a magnet and the supernatant removed. The tube was removed from the magnet and the antibody for the target protein (7.5 µg), diluted in 200 µl PBS (0.02% Tween 20) was added to the magnetic beads and incubated with rotation (10 min, ambient temperature). The tube was placed on the magnet and the supernatant removed. The tube was removed from the magnet and the beads were washed with 200 µl PBS (0.02% Tween 20) by gentle pipetting. The tube was placed on the magnet and the supernatant removed. 2 mg of RIPA lysed sample containing the antigen diluted with 200 µl PBS (0.02% Tween 20) was added to the bead-antibody complex which was resuspended by gentle pipetting. This mixture was incubated with rotation for 2 h at ambient temperature, which enables the antigen to bind to the bead-antibody complex. The tube was then placed on the magnet and the supernatant was transferred to a clean tube for analysing binding efficiency of the antigen. The bead-antibody-antigen complex was washed 3 times using 200 µl PBS (without Tween 20); the supernatant was removed with aid of the magnet after each wash. The bead-antibody-antigen complex was resuspended using 100 µl PBS (without Tween 20) and transferred to a new tube to prevent co-elution of proteins attached to the tube. The tube was placed on the magnet to remove the supernatant. 20 µl 50 mM glycine pH 2.8 and 10 µl of 2x NuPAGE LDS Sample buffer containing dithiothreitol (DTT, dissolved in 100 µl LDS Sample buffer 4x buffer to 0.04 M, which was then diluted 1:1 with MQ water to result in 2x sample buffer) were used to resuspend the

bead-antibody-antigen complex. The antibody-antigen complex was released from the magnetic beads by heating the sample for 10 min at 70 °C. The tube was placed on a magnet and the supernatant was used for downstream analyses, namely western blot and tryptic digestions.

2.6.17 Trypsin digestion of proteins

This procedure for trypsin digestion was changed after [255]. If recombinant H2B (Histone H2B Human, Recombinant, M2505S, NEB) was used, 15 µg of H2B were treated with a 100-fold excess of HCHO in MQ water for 24 h at ambient temperature. If the protein was immunoprecipitated from HEK293T cells, the whole target antigen elution was used after confirming the presence of the antigen by western blot. First, the sample was reduced with dithiothreitol (DTT; 2 µl of 85 mM DTT in 50 mM ammonium bicarbonate) for 40 min at 56 °C. The sample was then alkylated with iodoacetamide (IAA; 7 µl of 55 mM iodoacetamide) for 30 min in the dark at ambient temperature. To remove excess IAA, 3 µl of 85 mM DTT in 50 mM ammonium bicarbonate were added and the mixture was incubated for 10 min in the dark at ambient temperature. The next acetone precipitation step was only performed if proteins had been immunoprecipitated: 6 volumes of ice-cold acetone were used to precipitate the sample, whilst vortexing several times, followed by storage overnight at -20 °C. The sample was then centrifuged at (15000 g, 10 min, 4 °C), the supernatant was removed, the pellet was dried (5 min) and resolubilized with 27 µl 50 mM ammonium bicarbonate. Digestion was performed using Pierce Trypsin Protease, MS Grade. The lyophilized trypsin pellet (20 µg) was reconstituted using 20 µl of 50 mM acetic acid and 60 µl of 50 mM ammonium bicarbonate, resulting in a 0.25 µg/µl stock solution. The trypsin stock solution (3 µl) was added to the sample and incubated overnight at 37 °C. To promote digestion in folded parts of the protein, another digestion step was performed in 80% acetonitrile; thus, 1 µl of trypsin stock solution and 124 µl acetonitrile were added to the sample, followed by incubation at 37 °C for 3 hours. Digestion was stopped by adding 5% formic acid and the sample dried using vacuum centrifugation (Eppendorf Concentrator Plus, >2 h, 30 °C, 1400 rpm with rotor F-45-48-11, V-AQ mode). For immunoprecipitated samples, the pellet was resolubilized in 20 µl 0.1% formic acid, 98% MQ water, 2% acetonitrile and analysed by Orbitrap Elite LC-MS/MS. For recombinant H2B, the

sample peptides were resolubilized in 20 µl MQ water containing 0.1% formic acid. Samples were purified using ZipTip (OMIX C18 pipette tips, 10 µL tip, 2 - 10 µL elution volume). The ZipTip was equilibrated using 10 µl of buffer B (60% acetonitrile, 40% MQ water, 0.1% fomic acid), which was aspirated and dispensed twice; to finish the equilibration, 10 µl of buffer A (98% MQ water, 2% acetonitrile, 0.1% formic acid) was used, which was aspirated and dispensed twice. To bind the peptides, 10 µl of sample was aspirated and dispensed within the sample tube 10 times. After dispensing the last time, 10 µl of buffer A were used to wash the peptides bound to the ZipTip, followed by dispensing into waste; this washing step is repeated 4 times. Peptides were eluted into a new tube using 20 µl of buffer B. Samples were then dried using vacuum centrifugation (Eppendorf Concentrator Plus, as described above) and resuspended in 20 µl 0.1% formic acid in MQ water for Orbitrap Elite LC-MS/MS analyses.

2.6.18 Orbitrap Elite LC-MS/MS for tryptic proteins

LC-MS/MS samples were run and analysed by Elisabete Pires. Resulting tryptic peptides were performed on a NanoAcquity-UPLC system (Waters) connected to an Orbitrap Elite mass spectrometer (Thermo Fischer Scientific) possessing an EASY-Spray nano-electrospray ion source (Thermo Fischer Scientific). The peptides were trapped on a packed guard column (75 µm i.d. x 20 mm, Acclaim Pepmap100 C18, 3µm, 120 Å) using solvent A (0.1% formic acid in water) at a pressure of 140 bar. The peptides were separated on an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 15 mm, RSLC C18, 3 µm, 100 Å) using a 68 min linear gradient ranging from 3 to 97% of solvent B (0.1% formic acid in MeCN). The flow was set to 300 nl/min and the column was heated to 40 °C. The nESI source was operated at a needle voltage of 1600 V, and the ion transfer tube temperature was set to 275 °C. The separated peptides were electrosprayed directly into the mass spectrometer operating in a data-dependent mode using a CID based method. Full scan MS spectra (scan range 350-1500 m/z, resolution 120000, AGC target 1e6, maximum injection time 250 ms) and subsequent CID MS/MS spectra (AGC target 5e4, maximum injection time 100 ms) of the 10 most intense peaks were acquired in the Ion Trap. CID fragmentation was performed at 35% of normalized collision energy and the signal intensity threshold was kept at 500 counts. The CID method used performs beam-type CID

fragmentation of the peptides.

The analysis was performed with Peaks 8.5. The raw MS file was searched against the respective protein sequence. Trypsin with a maximum number of 3 missed cleavages and one unspecific end was selected as the protease. Carbamidomethylation (cysteine) was set as a fixed modification, oxidation (methionine), deamination (asparagine, glutamine) and +12 Da on the N-terminal proline were set as variable modifications. The precursor mass tolerance was set as 15 ppm. The fragment mass tolerances for CID were set to 0.8 Da. All spectra were manually validated. For all peptides present at -10lgP> 20, spectra were manually checked and validated or disqualified.

2.6.19 In-gel digestion of proteins

Proteins resolved by SDS-PAGE were excised with a scalpel, cut into 1-2 mm pieces and placed in a 1.5 ml reaction tube. Wash solution (200 µl; 50% MeOH, 5% acetic acid in MQ water) was added to the gel pieces and incubated overnight at ambient temperature while shaking. The wash solution was removed, followed by addition of 200 µl wash solution and rinsing for 2-3 h at ambient temperature. The wash solution was removed upon centrifugation. Acetonitrile (200 µl) was added to dehydrate the gel pieces for 5 min at ambient temperature, followed by centrifugation and removal of acetonitrile. The dehydration step was repeated using again 200 µl acetonitrile (5 min, ambient temperature, removal of acetonitrile). Dithiothreitol (DTT; 30 µl of 10 mM DTT in 100 mM NH_4HCO_3) was used to reduce the sample (30 min, ambient temperature), followed by removal of the solution. Iodoacetamide (IAA; 30 µl of 50 mM IAA in MQ water) was used to alkylate the sample (30 min, ambient temperature), followed by removal of the solution. Acetonitrile (200 µl, 5 min, ambient temperature) was used to dehydrate the gel pieces, resulting in a white color of the gel pieces. Upon removal of acetonitrile, the sample was rehydrated (200 µl of 100 mM NH_4HCO_3, 10 min, ambient temperature), followed by removal of the solution. Acetonitrile (200 µl, 5 min, ambient temperature) was used to dehydrate the gel pieces, followed by removal of the solution. The lyophilized trypsin pellet (20 µg, Pierce Trypsin Protease, MS Grade) was reconstituted using 1 ml of ice-cold 50 mM NH_4HCO_3, resulting in a 20 ng/µl stock solution (10x stock). Upon dilution of the 10x trypsin stock using 50 mM NH_4HCO_3 to 1x, 30 µl of

the 1x trypsin stock were added to the sample and incubated (10 min on ice, occasional gentle mixing). Upon centrifugation, excess trypsin solution was removed. NH_4HCO_3 (5 µl of a 50 mM solution) was added to the sample to carry out the trypsin digestion (37 °C overnight). NH_4HCO_3 (50 µl of a 50 mM solution) was added to the sample and incubated (10 min, ambient temperature, occasional vortex mixing). The supernatant was collected in a fresh 1.5 ml collection tube. Extraction buffer 1 (50 µl of 50% acetonitrile, 5% formic acid in MQ water) was added to the tube containing the gel pieces and incubated (10 min, ambient temperature, occasional vortex mixing), followed by addition of the supernatant to the collection tube. Extraction buffer 2 (50 µl of 85% acetonitrile, 5% formic acid in MQ water) was added to the tube containing the gel pieces and incubated (10 min, ambient temperature, occasional vortex mixing), followed by addition of the supernatant to the collection tube. The sample in the collection tube was dried (Eppendorf Concentrator Plus, >2 h, 30 °C, 1400 rpm with rotor F-45-48-11, V-AQ mode) and resuspended in 20 µl buffer A (98% MQ water, 2% acetonitrile, 0.1% formic acid), followed by analysis using LUMOS LC-MS/MS.

2.6.20 LUMOS MS/MS samples

LUMOS MS/MS samples were ran and analysed by Svenja S. Hester. Dried peptides were reconstituted in 5% DMSO and 5% formic acid and were analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS) using Ultimate 3000 UHPLC (ThermoFisher Scientific) connected to an Orbitrap Fusion Lumos Tribrid (ThermoFisher Scientific). Briefly, peptides were loaded onto a trap column (PepMapC18; 300 µm x 5mm, 5µm particle size, Thermo Fischer) and separated on a 50 cm-long EasySpray column (ES803, Thermo Fischer) with a gradient of 2-35% acetonitrile in 5% DMSO, 0.1% formic acid at a 250 nl/min flow rate over 60 min. Eluted peptides were then analysed on an Orbitrap Fusion Lumos Tribrid platform (instrument control software v3.3). Data were acquired in data-dependent mode, with the advance peak detection (APD) enabled. Survey scans were acquired in the Orbitrap at 120 k resolution over a m/z range 400 -1500, AGC target of 4e5 and S-lens RF of 30. Precursor ions were isolated in the Quad (1.6 isolation window), fragmented

in the HCD cell with 30% CE), and analysed in the Orbitrap at 30K resolution with an AGC target of 5e4 and 54 ms maximum injection time and a 7sec dynamic exclusion list.

3

Formaldehyde and N-Terminal Cysteines

Without knowing and respecting enzymes, better still loving them, answers to the most basic questions of growth, development, and disease will remain beyond reach.

Arthur Kornberg - For the Love of Enzymes [256]

Contents

3.1 Abstract **145**
3.2 Introduction **146**
3.3 Results **152**
3.4 Discussion **176**
3.5 Supplementary Information **179**
3.6 Methods **188**

3.1 Abstract

Asparagine has been reported to be low in alcoholics and diabetes type 2 patients. However, the molecular mechanism for this observation is unknown. Of the common amino acids, cysteine reacts most efficiently with HCHO, resulting in formation of the cyclic molecule thioproline *in vitro* and *in cellulo*. This observation raises the question as to whether the analogous reaction with proteins bearing an N-terminal cysteine residue is biologically relevant. Studies with peptides show that the adducts formed by reaction of HCHO with N-terminal serines and N-terminal threonines are less stable than the analogous adduct formed with N-terminal cysteines. This chapter reports studies with the N-terminal nucleophilic cysteine enzyme asparagine synthetase (ASNS) that reveal the potential for its activity to be modified by HCHO. While the N-terminal glutaminase domain of ASNS is inhibited by HCHO, the activity at its C-terminal synthetase domain is not affected by HCHO, indicating selective partial inhibition of only one of the two ASNS domains. Crucially, alcohol-derived acetaldehyde and glucose-derived α-oxoaldehydes, which are elevated in alcoholics and diabetes type 2 patients, respectively, inhibit ASNS *in vitro*. HCHO elevates ASNS protein levels *in cellulo*. Metabolomic analyses show depletion of intracellular asparagine upon aldehyde treatment in HEK293T, LN-18 cell lines, and in *Aldh2*$^{-/-}$ mice. These results provide a potential molecular explanation for low asparagine levels in alcoholics and diabetes type 2 patients.

3.2 Introduction

One of the main detoxification routes of HCHO is via reaction with glutathione, the first step of which is hemithioacetal formation by the reaction of the cysteinyl residue of glutathione (GSH) with HCHO.[87] Free cysteine has been shown to react with HCHO in human cells and in *E. coli* to form heterocyclic thioproline; of the common amino acids, cysteine is apparently one of the fastest to react with HCHO and the most stable HCHO adduct;[82, 86, 257] other biologically relevant aldehydes have been shown to react with cysteine to form the analogous thiazolidines (Figure 3.1 c).[82] Thioproline incorporation into proteins instead of proline is proposed to contribute to the toxicity of HCHO.[257]

All known HCHO sensors (transcription factors) have highly conserved cysteine residues which are crucial in enabling the HCHO detoxification response;[152, 258] substitution of these cysteines leads to a non-responsive transcription factor in all cases studied. The detailed mechanism of how cysteine residues sense HCHO is not known in most cases, however reaction at the cysteinyl thiol is likely.[130] In *E. coli* FrmR acts as a HCHO sensor forming a HCHO-derived methylene bridge that links N-terminal Pro2 and Cys35 with HCHO, a reaction that promotes DNA binding.[152] These examples highlight the importance of cysteine moieties in HCHO relevant biology.

The identity of the N-terminal residues of proteins regulates the ubiquitin-depdendent degradation signal of short-lived proteins (the N-end rule).[259] The cysteine branch of the N-degron pathway starts with removal of the N-terminal methionine, as catalysed by the ubiquitous methionine aminopeptidase (MAP), followed by oxidation of the N-terminal cysteine by cysteamine (2-aminoethanethiol) dioxygenase (ADO) in humans/plant cysteine oxidases (PCOs) in plants.[260] The so generated cysteine sulfinic acid is a substrate for argininyltransferases; upon addition of an N-terminal arginine, the proteins are targeted for degradation (Figure 3.1 a).[261] According to a recent proteomic N-terminome profiling study, there are 95 human proteins bearing an N-terminal cysteine.[262] There is thus the potential for HCHO to regulate N-terminal cysteine proteins by reaction of HCHO and cysteine to form an N-terminal thioproline.

Prokaryotes employ two structurally different asparagine synthetase (ASNS) enzymes to biosynthesise

asparagine; ASNS A uses ammonia, whereas ASNS B uses ammonia or glutamine as a nitrogen source.[263] Whereas both ASNS A and ASNS B families are reported to be present in prokaryotes, mammalian cells utilize only one form of the ASNS enzyme, and the enzyme mechanism corresponds to prokaryotic ASNS B employing glutamine as a nitrogen donor[263] In the work described in this chapter, ASNS B was developed as a model for the reaction of HCHO with an enzyme bearing an N-terminal catalytic cysteine. As typical for Ntn-hydrolases, ASNS catalysis begins with the deprotonation of the N-terminal nucleophile by its own α-amino group; ASNS uses its N-terminal cysteine to deamidate glutamine, forming ammonia, which travels through an intramolecular tunnel to react with β-aspartyl-AMP, forming asparagine (Figure 3.1 b).[264]

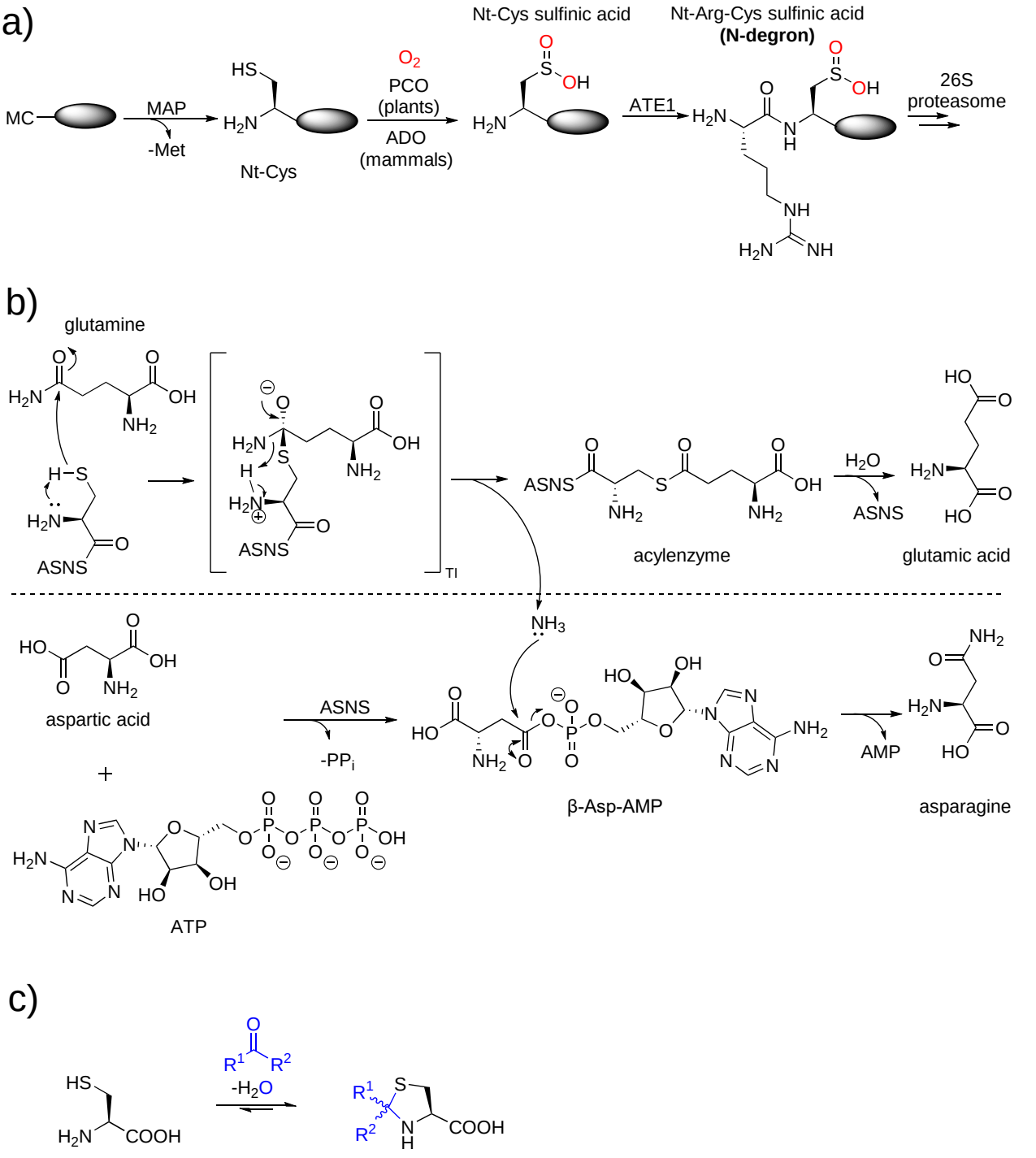


Figure 3.1 | N-Terminal cysteines involved in biology and the reactions of cysteine with carbonyl compounds. **a)** Oxidation of N-terminal cysteines and subsequent proteasomal degradation. Methionine aminopeptidase (MAP), cysteamine (2-aminoethanethiol) dioxygenase (ADO), plant cysteine oxidases (PCOs), arginyltransferase 1 (ATE1). The grey oval denotes any protein. The figure is modified from [243, 260]. **b)** ASNS catalysed formation of the non-essential amino acid asparagine involving deamidation of glutamine and concomitant ammonia release. Ammonia then travels through an intramolecular tunnel into the synthetase domain, where it reacts with β-aspartyl-AMP to form asparagine.[264] **c)** Reactive carbonyl compounds may interfere with biological processes involving free N-terminal cysteines. Formaldehyde (R^1=H, R^2=H), acetaldehyde (R^1=H, R^2=Me), acetone (R^1, R^2=Me), glyoxal (R^1=H, R^2=C(O)H) or methylglyoxal (R^1=H, R^2=C(O)Me).[82]

ASNS is part of the Ntn-hydrolase (N-terminal nucleophile) superfamily, which is a diverse set of enzymes comprising an N-terminal catalytic serine, threonine, or cysteine residue; all Ntn-hydrolases catalyse the cleavage of amide bonds.[265] Some examples of Ntn-hydrolases are the 20S core particle (CP) of eukaryotic cells (N-terminal threonine),[266] the *E. coli* penicillin G acylases (N-terminal serine),[267] the *Bacillus subtilis* glutamine 5-phosphoribosyl-1-pyrophosphate amidotransferases (N-terminal cysteine),[268] the *E. coli* glucosamine 6-phosphate synthase (N-terminal cysteine),[269] the human aspartylglucosaminidase (N-terminal threonine),[270] and the *Ochrobactrum anthropi L*-aminopeptidase-*D*-Ala-esterase/amidase (N-terminal serine) (Figure 3.2).[271]

Given that free cysteine reacts efficiently with HCHO,[82] it was hypothesised that N-terminal cysteines may be prone to reaction with HCHO, resulting in thiazolidine formation, rendering the N-terminus catalytically inactive. Asparagine levels are decreased in alcoholics[272] and individuals with diabetes type 2,[273] which might be due to inhibition of ASNS by acetaldehyde and α-oxoaldehydes, respectively (Figure 3.1 c).

Asparagine synthetase deficiency (ASD) is a rare neurological disorder which results in intellectual disability, microcephaly, and epileptic-like seizures; asparagine appears to cross the blood-brain barrier inefficiently, hence local *de novo* synthesis is required in the brain.[274] Several mutations in the *ASNS* gene give rise to ASD.[275]

A crystal structure of human ASNS has provided insight into the structural basis of ASD, although a detailed understanding on how the clinical variants alter the mechanism of ASNS is lacking.[276] The *ASNS* gene is expressed at low levels in acute lymphoblastic leukemia (ALL) cells, which makes these malignant cells auxotrophic for the non-essential amino acid asparagine;[277] this is used therapeutically by using the enzyme asparaginase (ASNase), which converts asparagine to aspartate and ammonia to deplete blood asparagine levels, which leads to apoptosis of ALL cells relying on exogenous asparagine. The use of ASNase was the first successful approach where cancer has been treated by exploiting a metabolic feature of cancer cells.[277]

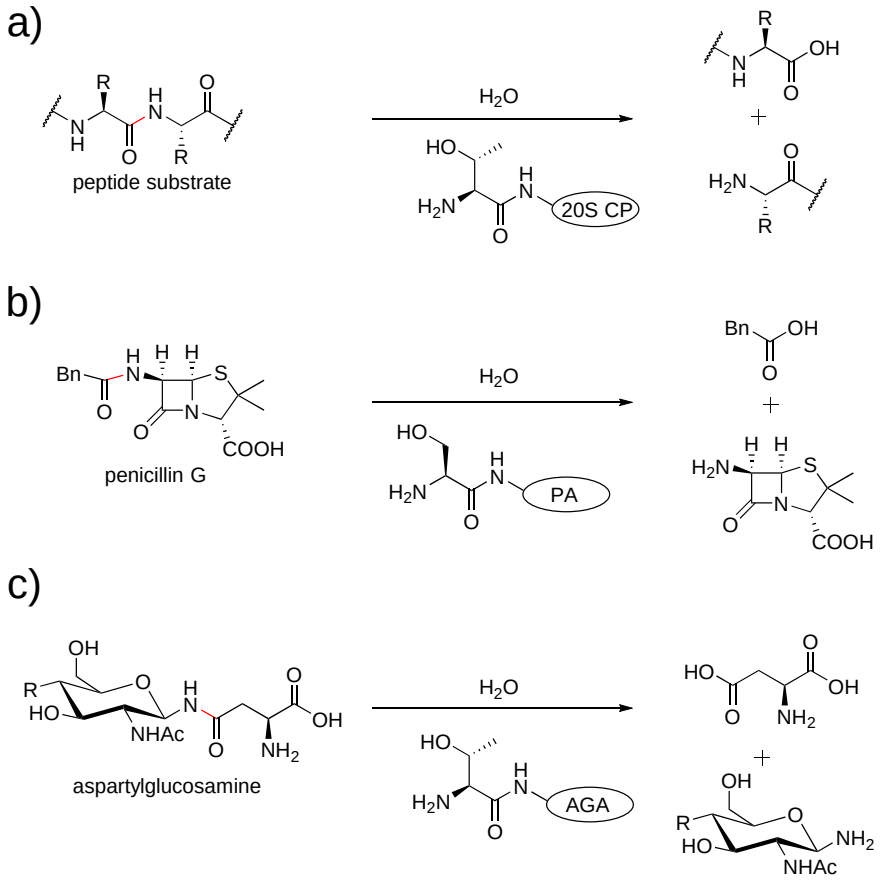


Figure 3.2 | Examples of Ntn-hydrolase reactions. a) Hydrolysis of a peptide substrate by the 20S core particle (CP), catalysed by its N-terminal threonine.[266] **b)** Hydrolysis of penicillin G by penicillin G acylase (PA), catalysed by its N-terminal serine.[267] **c)** The enzyme aspartylglucosaminidase (AGA) cleaves the amide bond between asparagine and a glycan using its N-terminal threonine.[270] Scissile amide bonds are shown in red.

The concentration of asparagine in human blood is reported to be ∼0.05-0.1 mM.[278] Glutamine is the amino acid with the highest concentration in blood, i.e. 0.5-0.8 mM; in cell culture medium glutamine is supplemented at 2 mM, typically in the form of *L*-alanyl-*L*-glutamine (GlutaMAX) due to the chemical instability of glutamine resulting in pyroglutamate formation.[278–280] In Dulbecco's Modified Eagle Medium (DMEM), which is the most broadly used basal medium for supporting the growth of many adherent mammalian cell types, there is no asparagine; in Roswell Park Memorial Institute (RPMI) medium, which is used for growth of human leukemia cells which have low levels of ASNS, asparagine is supplementd to 0.38 mM.[278] Glutamine is metabolised extensively, mainly by glutaminase (GLS), whereas asparagine has been proposed to be a 'dead-end' metabolite.[281] Intracellular asparagine has been shown to act as exchange factor for the extracellular amino acids

serine, arginine, and histidine; this antiport activates nucleotide and protein synthesis. This observation highlights the importance of asparagine in cancer cell growth and might explain the efficacy of ASNase in ALL cells.[281] Asparagine is not catabolised in mammalia due to an absence of cytosolic ASNase activity and is conditionally essential under glutamine-deprived conditions.[282] The following work explores HCHO-derived non-enzymatic post-translational modifications on N-terminal cysteines in the context of Ntn hydrolases, and more specifically, asparagine biosynthesis. The results extend recent findings of HCHO participating in the non-enzymatic metabolome - where HCHO has been shown to react with the amino acid cysteine to form thioproline[82, 86, 257] - to the protein level.

3.3 Results

Ntn hydrolases employ an N-terminal serine, threonine or cysteine residue in their catalysis. It was hypothesised that reaction of an N-terminal nucleophile with HCHO will render the catalytic moieties unavailable for catalysis. To assess this idea, SerGlu (**compound 1**), ThrGlu (**compound 2**) and CysGlu (**compound 3**) dipeptides were synthesized and reacted with a 10-fold excess HCHO at pD 9.4 for 12 h and subsequently analysed by NMR. All three compounds formed an N-terminal 5-membered ring with an additional hemiaminal off the N-terminal nitrogen *in situ* (**1a, 2a, 3a**, Figures 3.3 a+b, 3.17, 3.18, 3.19). The HCHO-derived adducts on **1a** and **2a** could not be isolated by HPLC and were not stable under LC-MS conditions; the thiazolidine, but not the hemiaminal of **3a** was stable under HPLC conditions (Figure 3.20), resulting in formation of **3b** (Figure 3.3 b).

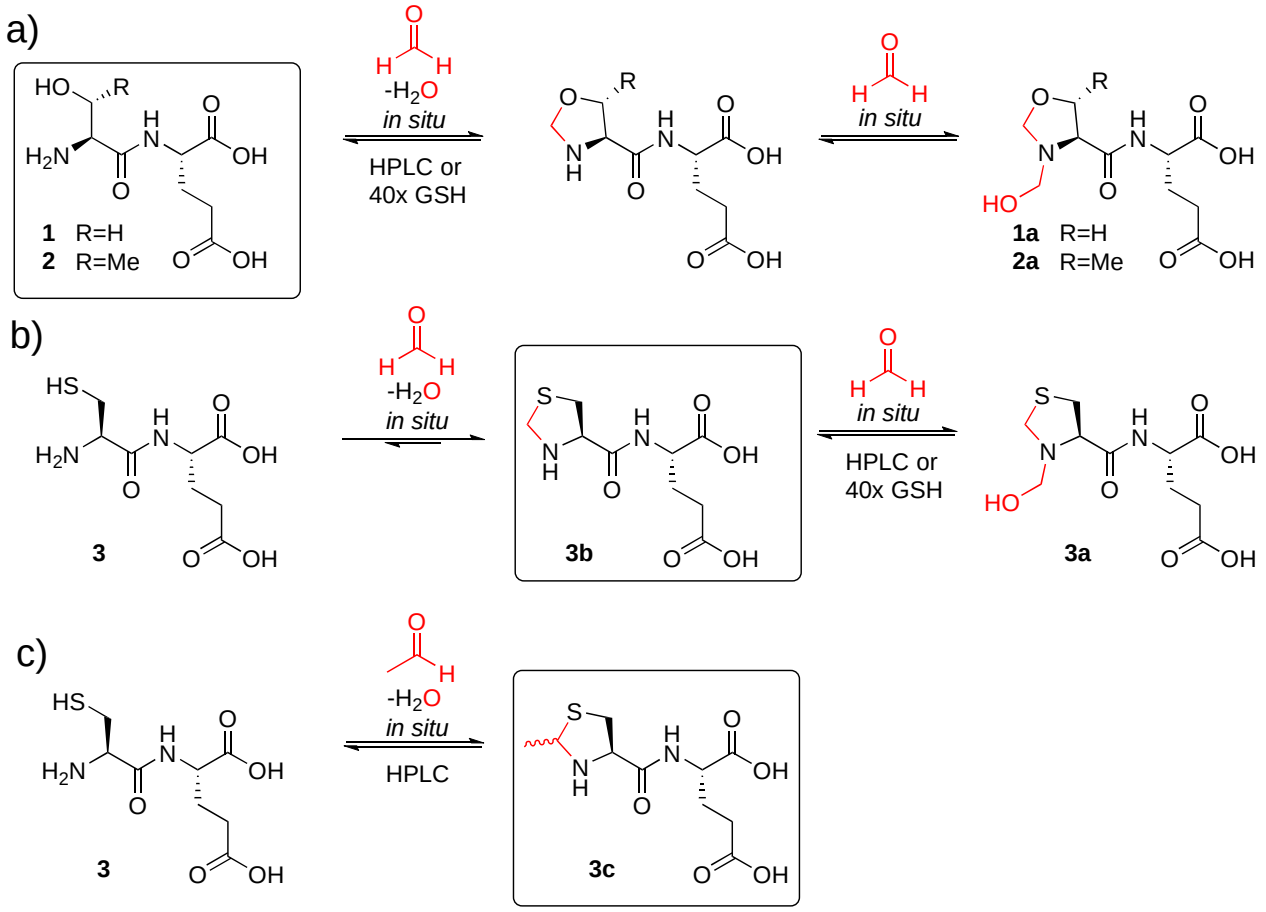


Figure 3.3 | N-Terminal cysteines form stable thiazolidines. a) 1 and **2** were reacted with HCHO *in situ* to form the N-terminal oxazolidines **1a** and **2a**, which are unstable under HPLC conditions or upon addition of a 40-fold excess of glutathione (GSH; 40:1 GSH:dipeptide), resulting in formation of **1** and **2**. **b) 3** was reacted with HCHO *in situ* to form the N-terminal thiazolidine with hemiaminal **3a**; the hemiaminal is not stable, whereas the thiazolidine is stable under HPLC conditions and upon addition of 40-fold excess GSH, resulting in formation of **3b**. **c) 3** was reacted with acetaldehyde *in situ* to form the N-terminal thiazolidine **3c**, which was found to release acetaldehyde over time after HPLC purification. All reactions were carried out using a 10-fold excess of aldehyde compound at pD 9.4 for 12 h. Boxed structures are the main species found after HPLC purification.

However, all the HCHO-derived adducts in **1a** and **2a** were readily removed by a 40-fold excess of the biologically relevant HCHO scavenger glutathione (GSH), while the thiazolidine but not the hemiaminal of **3a** was found to be stable in presence of 40-fold excess GSH, resulting in formation of **3b** (Figure 3.4). **3** was also reacted with acetaldehyde (CH_3CHO) under the same conditions and subjected to HPLC purification, resulting in formation of **3c**. Whilst the acetaldehyde adduct **3c** was stable under HPLC conditions, the NMR always contained a small fraction of starting material (both acetaldehyde [δ_H 1.35 ppm] and **3** [δ_H 3.16 ppm]), despite multiple attempts (Figures 3.3 c, 3.21). This may indicate that **3c** is not as stable as its HCHO counterpart **3b**.

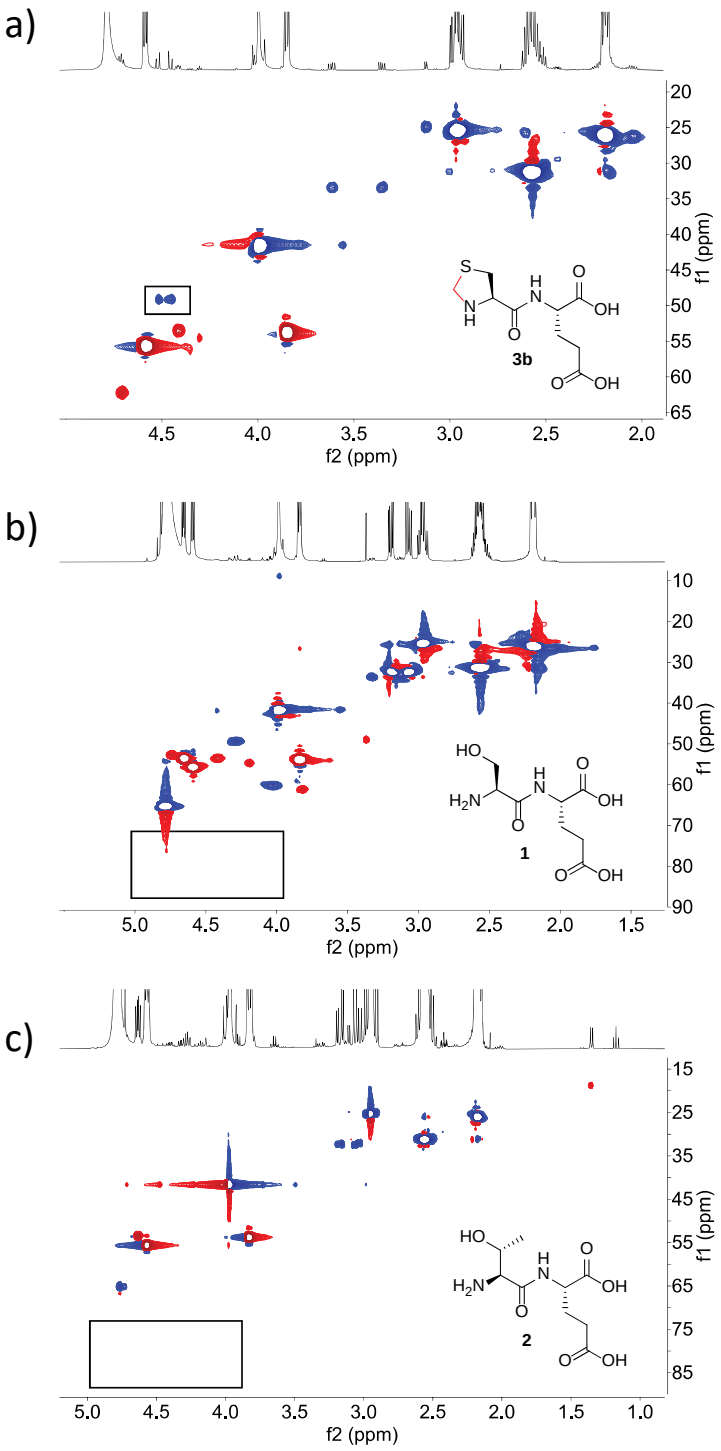


Figure 3.4 | Compound 3b is stable in presence of a 40-fold excess of GSH, whilst 1a and 2a are not. **a)** HPLC purified **3b** with 40-fold excess glutathione (GSH, 600 MHz). The HCHO derived methylene signals are highlighted in a rectangle and are consistent with the HCHO-derived methylene signal of purified **3b** at δ_C 49.5 ppm. **b+c) 1a** (b, 600 MHz) and **2a** (c, 400 MHz) were generated *in situ* employing a 10-fold excess of HCHO, followed by addition of a 40-fold excess GSH (relative to dipeptide). The rectangle indicates the region where HCHO derived oxazolidine methylene bridge signals should appear (δ_C 84.2 ppm for **1a** and δ_C 83.6 ppm for **2a**). The absence of signals indicates **1a** and **2a** are not stable in presence of excess GSH, resulting in formation of **1** and **2**.

Asparagine synthetase (ASNS) was chosen as model Ntn-hydrolase bearing an N-terminal catalytic cysteine to investigate reactions of HCHO and N-terminal cysteines at the protein level. To investigate the reactivity of its N-terminal cysteine, a peptide corresponding to the first seven N-terminal residues of *E. coli* ASNS B was synthesised. 10- and 100-fold excess HCHO were titrated onto the peptide and monitored over time at pH 7.4 in 50 mM potassium phosphate by MALDI MS. The observed +12 Da shift suggests thiazolidine formation and occurs very rapidly; almost full turnover is observed after 1 h in 10-fold excess HCHO (Figures 3.5 a, 3.22).

Full-length ASNS B was then heterologously biosynthesized in *E. coli* (Figure 3.41) with the aim of developing an *in vitro* assay using ^{1}H NMR (700 MHz). ASNS B uses its N-terminal cysteine to deamidate glutamine, forming ammonia, which travels through an intramolecular tunnel to react with β-aspartyl-AMP, forming asparagine (Figure 3.1 b).[264] For the NMR assay, ASNS B was preincubated with HCHO or Milli-Q (MQ) water control (20 min, 37 °C), followed by addition of the preincubation mixture to the substrates (glutamine, aspartic acid, Mg-ATP). The resulting reaction mixture was then incubated (37 °C, 4 h), quenched with CF_3CO_2H and analysed by ^{1}H NMR (700 MHz). As reported (using HPLC measurements employing amino acids that were converted to UV-active phenylisothiocyanate derivatives), the glutaminase partial reaction which forms glutamate and ammonia is coupled to the synthetase reaction, forming asparagine from aspartate (Figure 3.5 b, panel 2, +ASNS);[283] this reaction does not occur under the same conditions without ASNS B (Figure 3.5 b, panel 1, -ASNS). By use of NH_4Cl rather than glutamine (Q) as source of ammonia, ASNS B was observed to convert aspartate to asparagine (Figure 3.5 b, panel 3, -Q, NH_4Cl), as reported.[283] Almost complete inhibition of the glutaminase partial reaction can be seen when ASNS B is preincubated with a 100-fold molar excess of HCHO (20 min, 37 °C). This is likely a consequence of the reaction of HCHO with the N-terminal cysteine residue of ASNS B. NH_3, which is necessary for the synthetase reaction, is not released, resulting in a lack of asparagine formation (Figure 3.5 b, panel 5, 100x HCHO). However, when NH_4Cl is present in the reaction mixture, the synthetase partial reaction of ASNS B still takes place using a HCHO concentration which inhibited glutaminse activity (Figure 3.5 b, panel 9, 10^3x HCHO rescue). This observation suggests that only the N-terminal reaction of ASNS

B is inhibited by HCHO. The 'rescue' with NH_4Cl does not occur at 10^4-fold molar excess HCHO, where the enzyme seems to be completely inactive, i.e. both the glutaminase and the synthetase partial reactions are inhibited by HCHO (Figure 3.5 b, panel 10, 10^4x HCHO rescue). Inhibition of the synthetase domain at 10^4-fold molar excess HCHO may be due to reaction of HCHO with residues other than the N-terminal cysteine.

An endpoint assay screen of an array of biologically relevant carbonyl compounds (acetaldehyde, acetone, methylglyoxal, and pyruvate; see Figures 3.6 and 3.7 for structures of carbonyl compounds) was then tested for their ability to inhibit ASNS B. The results suggest methylglyoxal (MGO) and acetaldehyde (Figure 3.24, panels 3 and 5, respectively), but not pyruvate or acetone modulate ASNS B activity (Figure 3.24, panels 2 and 4, respectively). Given that α-oxoaldehydes like MGO are more electrophilic due to the oxo-group α to the carbonyl group, the results showing that MGO inhibits ASNS B are in line with the expected high reactivity of MGO. In contrast, the carboxylic acid α to the carbonyl group in pyruvate makes the carbonyl group in pyruvate less electrophilic, which is in agreement with pyruvate not acting as inhibitor for ASNS B. The bulkiness of the alkyl groups in ketones like acetone reduce their reactivity; acetaldehyde, having only one alkyl group, is more reactive and exhibits, in contrast to acetone, inhibition against ASNS B.

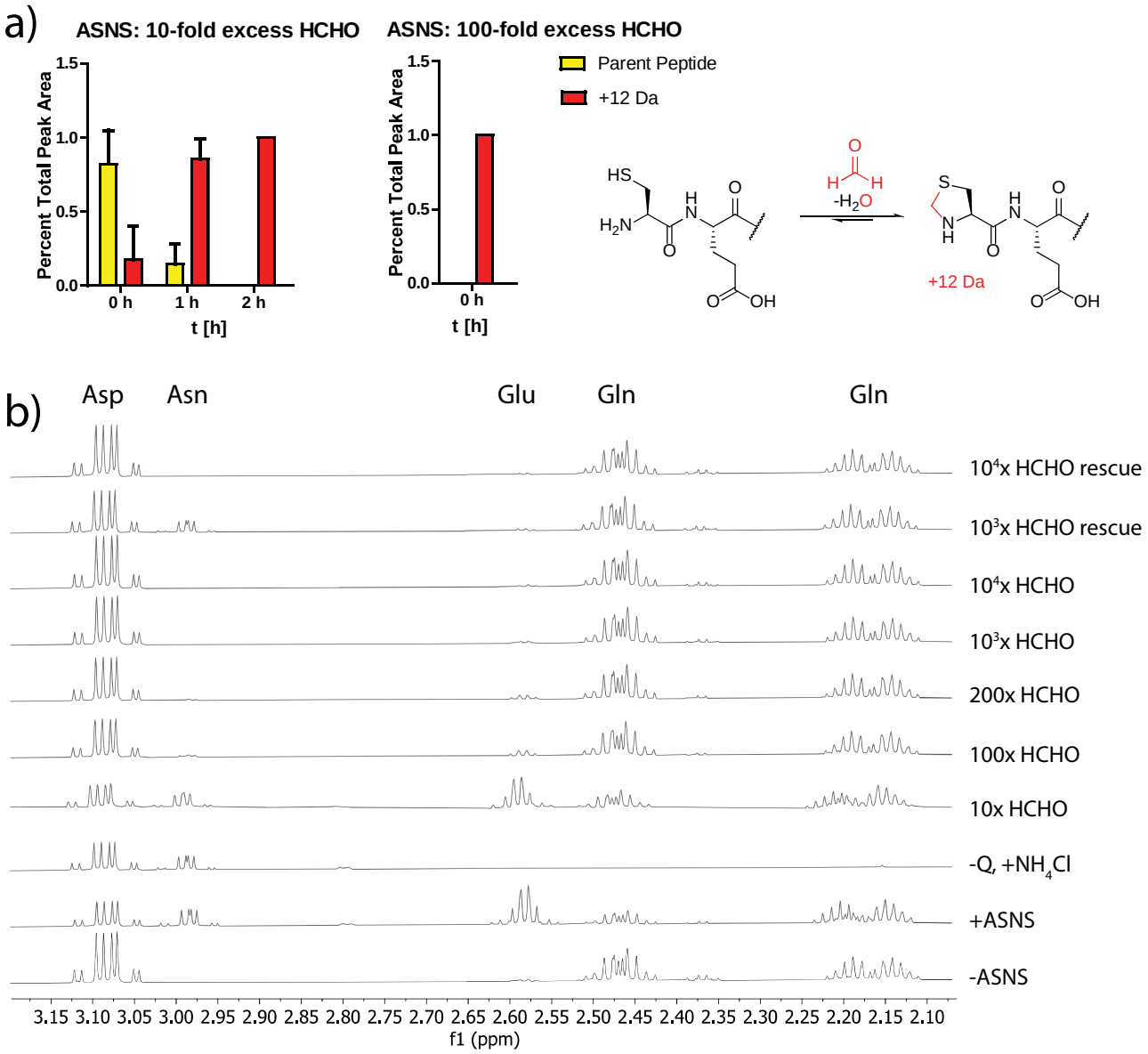


Figure 3.5 | HCHO selectively inhibits the ASNS B glutaminase reaction. a) Reaction of ASNS B 7mer peptide (NH_2-CSIFGVF-NH_2) with HCHO (10-fold and 100-fold excess HCHO) at pH 7.4 and 50 mM potassium phosphate. HCHO adducts are shown in red (+12 Da); the parent peptide (not reacted) is shown in yellow. Reactions were monitored directly upon addition of HCHO (0 h), and at the indicated times. The integrated area of the peaks were used to generate graphs (compare with Figure 3.22). Errors: SEM (n=3, technical repeats). **b)** ^{1}H NMR (700 MHz) showing the conversion of glutamine (Q, δ_H 2.42-2.51 and 2.10-2.22 ppm) to glutamate (δ_H 2.55-2.61 ppm) and aspartate (δ_H 3.03-3.13 ppm) to asparagine (δ_H 2.97-3.00 ppm) in presence of ASNS B (panel 2, +ASNS). The glutaminase reaction, which is dependent on the N-terminal cysteine of ASNS B, is inhibited by HCHO. However, the synthetase reactivity of ASNS B is still functional in presence of NH_4Cl (panel 9, 10^3x HCHO rescue), but not if 10^4-fold excess HCHO were used to preincubate the enzyme (panel 10, 10^4x HCHO rescue). The experiment was conducted as independent replicate (Figure 3.23).

Next, protein observed MS was performed to investigate if any of the tested carbonyl compounds react covalently with ASNS B. The observed mass of ASNS B (63594 Da) is consistent with its calculated mass (calculated: 63595 Da; Figure 3.6). In agreement with the turnover results obtained from the NMR assay showing inhibition by HCHO (Figure 3.5 b, panels 4-8), HCHO very efficiently reacts with ASNS B as shown by MS under likely denaturing conditions. Multiple adducts on ASNS B are observed by MS using a 100-fold excess HCHO. Only minor mass shifts were observed at a 10-fold excess HCHO, still showing a major peak of unreacted ASNS B at 63594 Da. No remaining protein without HCHO adducts was observed at a 1000-fold excess HCHO (Figure 3.7).

In line with the NMR assay results which indicated no inhibition of ASNS B for pyruvate and acetone (Figure 3.24, panels 2 and 4, respectively), pyruvate and acetone do not form covalent adducts with ASNS B at 1000-fold molar excess (Figure 3.6). The ethanol degradation product acetaldehyde[172] and the diabetes marker metabolites glyoxylic acid, glyoxal, methylglyoxal,[193] as well as the endogenous Michael acceptor acrolein[177] reacted with ASNS B to form covalent adducts (Figure 3.7). In the case of glyoxylic acid and methylglyoxal, up to four adducts are formed, while acetaldehyde apparently reacts with ASNS B to only form one adduct.

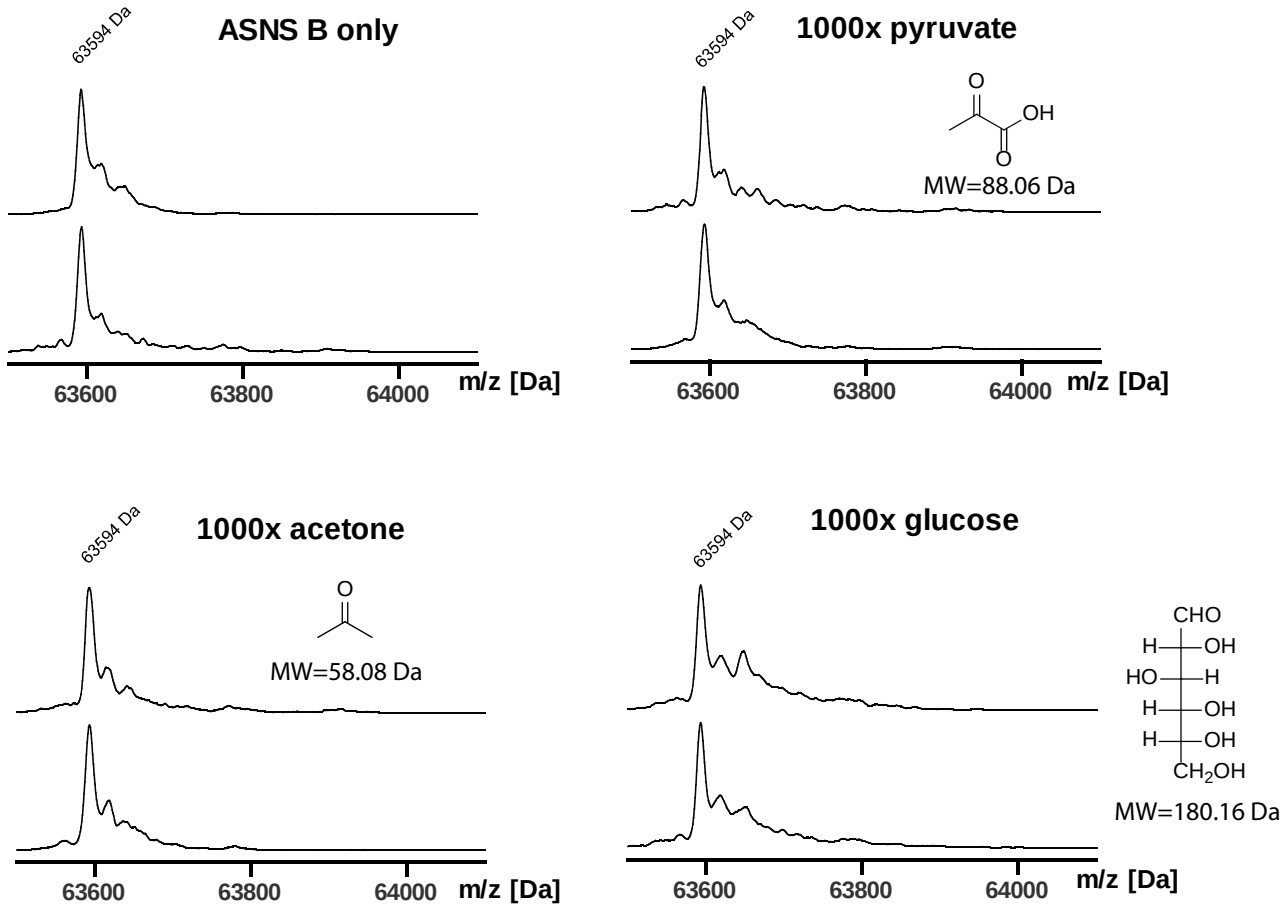


Figure 3.6 | Protein observed MS – ASNS B and non-reactive carbonyl compounds. ASNS B was preincubated with MQ water (ASNS B only) or the indicated molar excess of carbonyl compound for 20 min at 37 °C, then subjected to analysis by solid phase extraction coupled to mass spectrometry (SPE-MS), performed in technical duplicate.

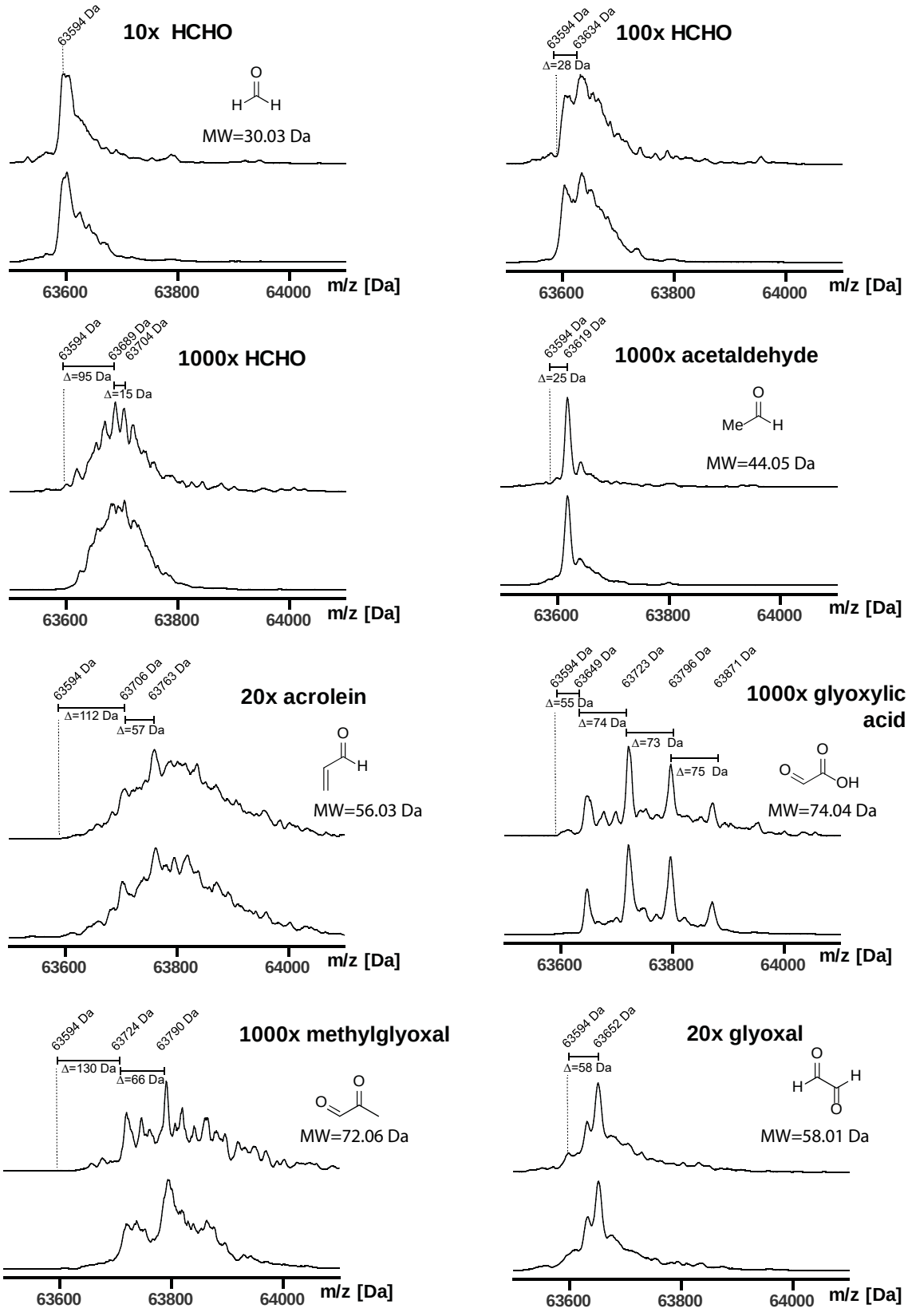


Figure 3.7 | Protein observed MS – ASNS B and reactive carbonyl compounds. ASNS B was preincubated with MQ water (ASNS B only) or the indicated molar excess of carbonyl compound for 20 min at 37 °C, then subjected to analysis by SPE-MS, performed in technical duplicate.

A set of other biologically relevant carbonyl compounds was then tested for inhibition of ASNS B activity using the ^{1}H NMR assay described before (Figures 3.5 b, 3.24). The ASNS B catalysed conversion of glutamine to glutamic acid and aspartic acid to asparagine was monitored over ∼130 min (Figure 3.8). After preincubating ASNS B for 45 min at 37 °C with or without carbonyl compounds, the preincubation mix was added to the substrates and the reaction was monitored by ^{1}H NMR (700 MHz) at 37 °C. Preincubating ASNS B with a 100-fold molar excess glucose, acetone, glyoxylic acid or a 500-fold excess pyruvate does not modulate ASNS B activity relative to the MQ water control (Figure 3.8 d). Crucially, a 100-fold excess of both acetaldehyde and the α-oxoaldehydes methylglyoxal significantly inhibit ASNS B, as indicated by the slow biosynthesis of asparagine (Figure 3.8 d). A 20-fold molar excess HCHO (Figure 3.8 c+d) or glyoxal is sufficient to almost completely inhibit ASNS B catalysis (Figure 3.8 d). No ASNS B catalysed asparagine formation is observed in presence of 20-fold molar excess of the Michael acceptor acrolein (Figure 3.8 d).

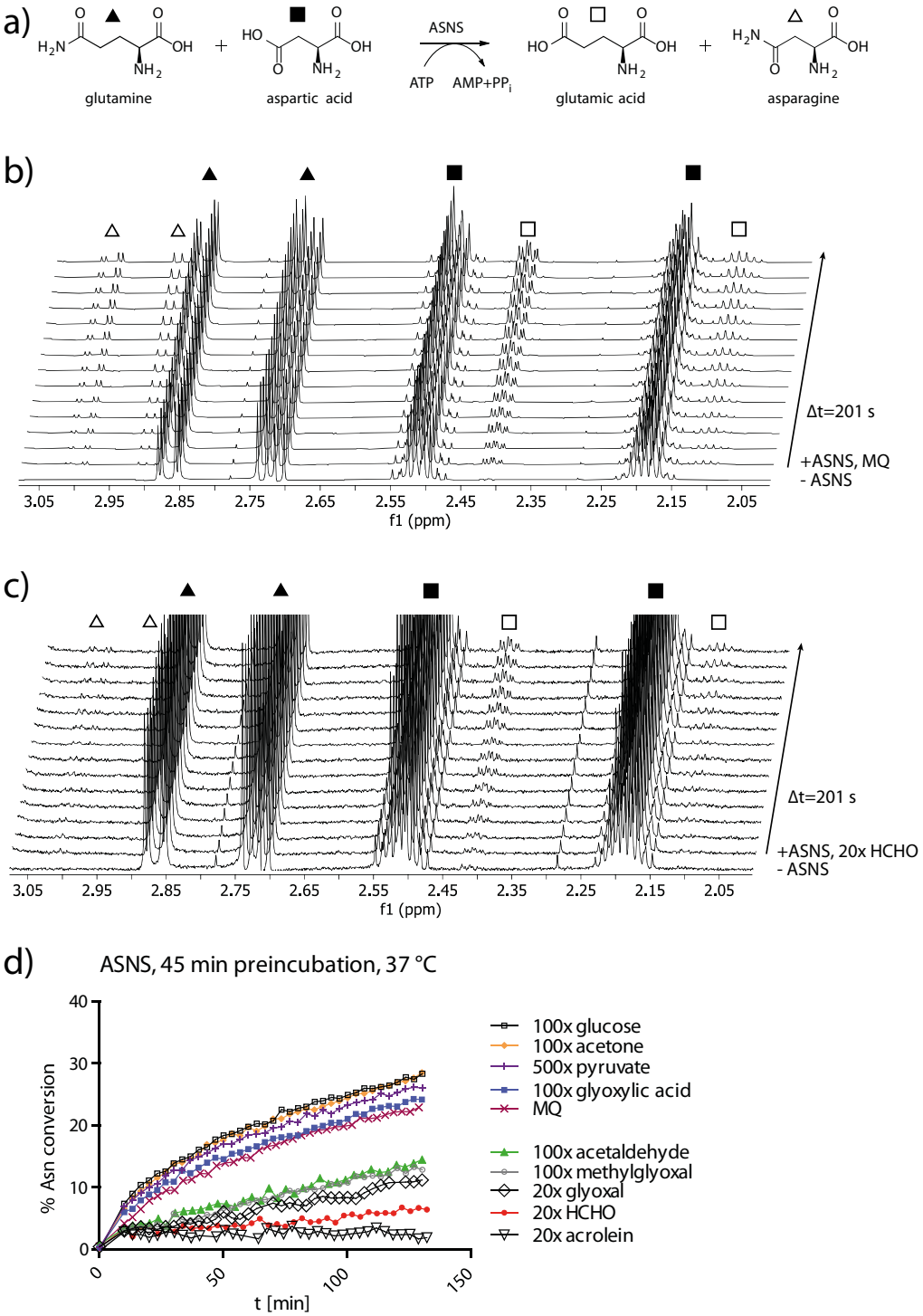


Figure 3.8 | ASNS B is inhibited by several aldehydes. a) Outline scheme for ASNS B catalysis. **b+c)** Exemplary data set (MQ water control, (b); 20-fold excess HCHO, (c)) showing the ASNS B catalysed conversion of glutamine (▲) and aspartic acid (■) to glutamic acid (□) and asparagine (Δ) (^{1}H NMR (700 MHz)). The first panel shows substrates only without addition of ASNS B (-ASNS). The second acquisition was started 10 min after mixing the assay components with ASNS B (+ASNS). Spectra are shown at 201 s intervals (acquisition time for each spectrum was 179 s). Note the position of the peaks is different to Figure 3.5 b, because CF_3CO_2H was used to quench ASNS B for the endpoint assay in Figure 3.5 b. **d)** ASNS B was preincubated with the indicated molar excess of carbonyl compounds (or MQ water control) for 45 min at 37 °C and added to the assay components, followed by monitoring the reaction by ^{1}H NMR (700 MHz) (n=2 independent repeats, repeat is shown in Figure 3.25).

The N-end rule describes the influence of an N-terminal amino acid in a given protein to its half-life.[243] The thiazolidine that forms on N-terminal cysteines upon reaction with HCHO might be sufficient structural modification to influence protein stability. It is therefore of interest to test the half-life of ASNS in presence and absence of HCHO. The protein synthesis inhibitor cycloheximide (CHX) is used to assay for protein stability, enabling analysis without confounding contributions of *de novo* biosynthesis. No change in ASNS stability was observed using CHX following HCHO treatment (Figure 3.27).

If the partial inhibition of ASNS B by HCHO translates into cells, then cellular asparagine levels should inversely correlate with HCHO levels. Testing this hypothesis in human cells is possible because Dulbecco's Modified Eagle Medium (DMEM) used in this study does not contain asparagine and human asparagine biosynthesis is only achieved through ASNS.[277] To test this hypothesis *in cellulo*, HEK293T cells were treated with a non-lethal dose of HCHO, i.e. <250 µM (Figure 3.26 c), MQ water control or acetone as a carbonyl compound stress control. Asparagine (Asn) levels do not change with increasing HCHO, as well as GSH and cysteine levels, while thioproline levels increase (Figure 3.9), as reported.[86] Thioproline results from the reaction of free cysteine with HCHO. The unchanged GSH and cysteine levels suggest both molecules are reconstituted in cells upon reaction with HCHO. While glutathione-dependent detoxification of HCHO is well studied,[87] the enzyme capable of regenerating cysteine from thioproline remains to be elucidated.

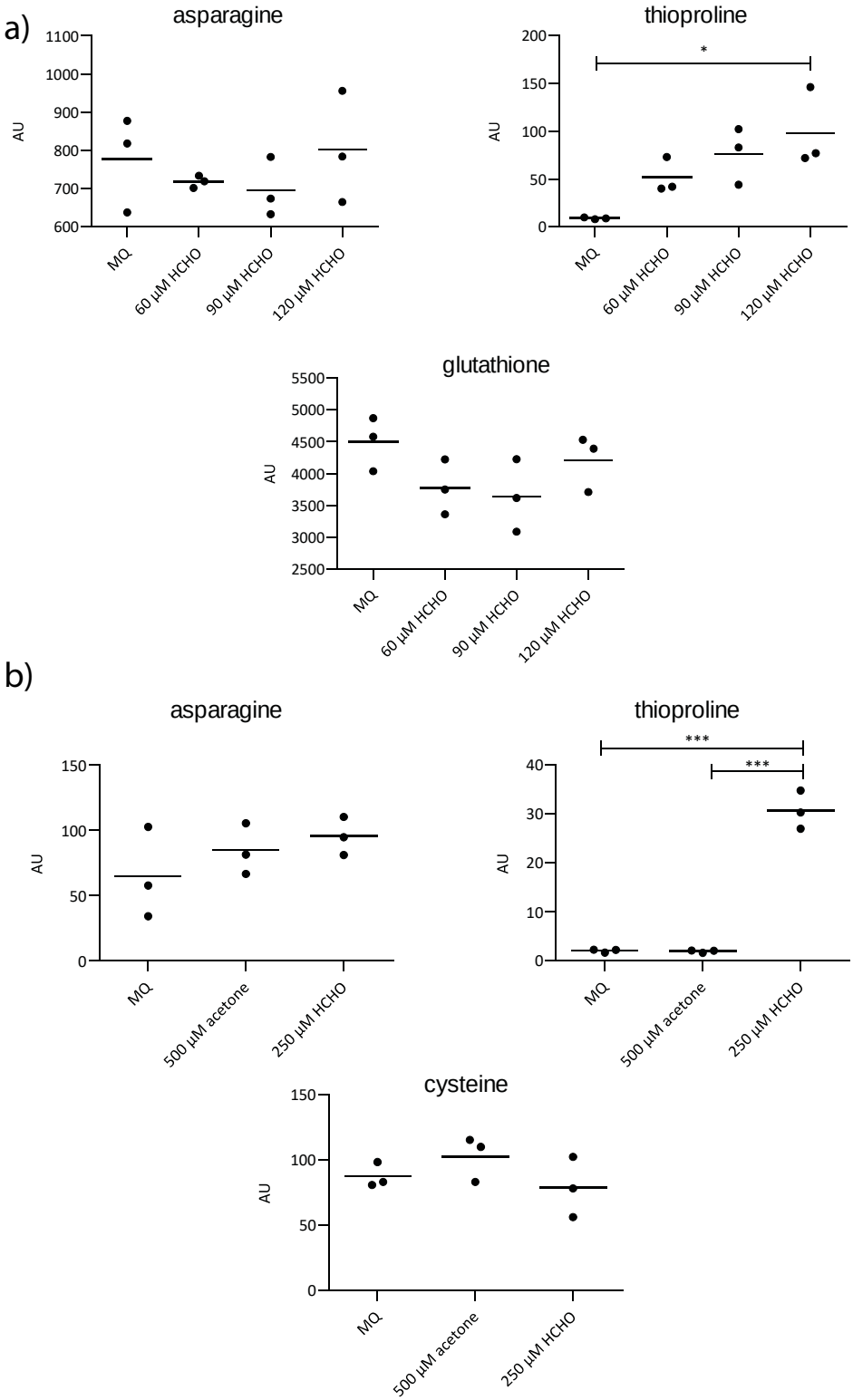


Figure 3.9 | Metabolomics analyses of HEK293T treated with HCHO. a) HEK293T cells were treated with the indicated amounts of HCHO or MQ water for 14 h. **b)** HEK293T were treated with two cumulative doses that were 24 h apart. The cells were harvested 16 h after the second treatment. Each indicated dose represents the total cumulative dose of HCHO. Each consecutive dose was 1.5-fold the preceding dose. Individual data points and mean are shown, statistical significance was assessed using 1way ANOVA; ***P $\leq$ 0.001, *P $\leq$ 0.05, ns: not significant. AU=Arbitrary Unit. Metabolomics samples were prepared by myself, but run and analysed by Dr John-Walsby Tickle.

In the glioblastoma cell line LN-18, addition of HCHO again apparently correlated with elevated thioproline levels (Figure 3.10 a). Crucially, significant depletion of asparagine levels, as well as cysteine and glutathione (GSH) levels were observed upon HCHO treatment (Figure 3.10 a). The observed difference between HEK293T and LN-18 cells might be due to less efficient regeneration of cysteine and GSH from the HCHO adducts thioproline and *S*-hydroxymethylglutathione, respectively, in case of LN-18. This depletion of cysteine and GSH might lead to higher intracellular free HCHO levels, which might inhibit ASNS, ultimately leading to less asparagine biosynthesis.

It is of interest to test if asparagine level decrease upon HCHO treatment in samples that are deficient in GSH dependent detoxification. GSH levels can be depleted by intake of food high in acrylamide or medication such as paracetamol (Figure 1.23 d+e).[186, 189] Further, the rare disease glutathione synthetase deficiency leads to low levels of glutathione, caused by an inability to biosynthesise GSH.[284] In collaboration with Dr Christopher Millington and Dr Felix Dingler (KJ Patel laboratory), wildtype or Adh5$^{-/-}$ mouse livers were tested for their metabolites and ASNS levels. Adh5$^{-/-}$ mice are deficient in metabolising *S*-formylglutathione, which results from reaction of GSH and HCHO. Hence, Adh5$^{-/-}$ mice cannot detoxify HCHO via the GSH-dependent pathway. MeOH, which is metabolised to HCHO,[165] was used in this experiment instead of HCHO. Thioproline levels are significantly elevated in Adh5$^{-/-}$ mice, highlighting the importance of cysteine acting as sink for HCHO (Figure 3.10 b). However, asparagine levels are unaltered among the samples, indicating that thioproline formation alone is sufficient to protect cells from excess HCHO. ASNS protein levels in the Adh5$^{-/-}$ mice are mildly elevated relative to the wildtype mice (Figure 3.10 c), which might indicate a metabolic feedback mechanism in response to elevated HCHO levels.

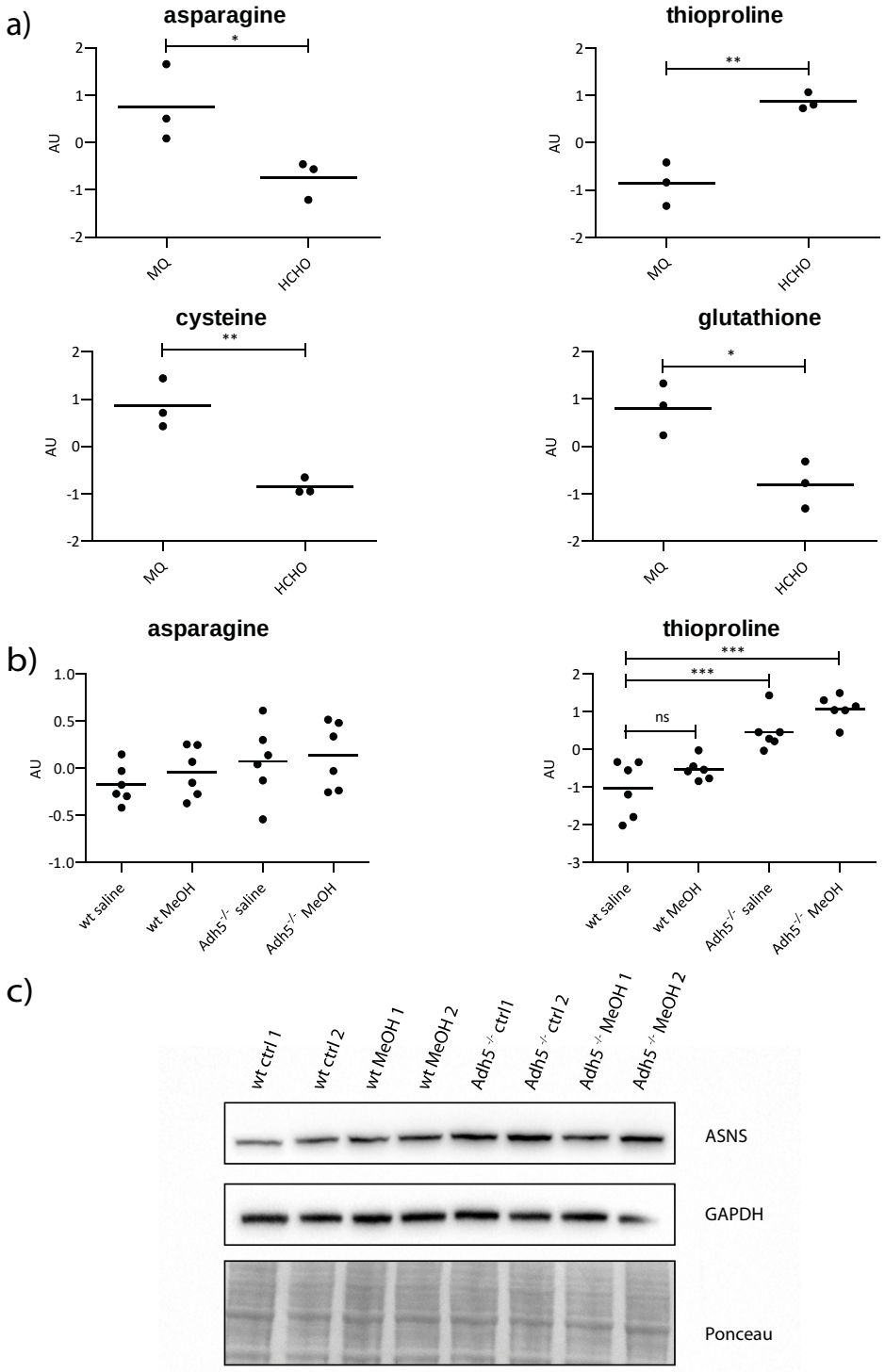


Figure 3.10 | Metabolomics and western blot analyses of LN-18 and mice. a) LN-18 cells were treated with 180 µM HCHO for 24 h, followed by a second treatment of 270 µM and harvest 26 h after the second treatment. **b+c)** Liver metabolite (b, n=6) and protein (c, n=2) levels of wildtype (wt) or Adh5$^{-/-}$ mice that were treated with saline (control) or MeOH for 6 h. An Origene ASNS antibody (TA503444) was used to obtain signals for the mouse samples. Mouse experiments and liver homogenisation were conducted by Dr Felix Dingler and Dr Christopher Millington. Individual data points and mean are shown, statistical significance was assessed using unpaired t-test (a) and 1way ANOVA (b); ***P $\leq$ 0.001, **P $\leq$ 0.01, *P $\leq$ 0.05, ns: not significant. Metabolomics samples were prepared by myself (analysis of (a) was done by myself), but run and analysed (samples in (b) only) by Dr John-Walsby Tickle.

Other than too efficient HCHO detoxification by GSH and cysteine, unchanged asparagine levels might be due to several reasons: First, cells might be able compensate for the lack of functional ASNS enzyme by increasing its biosynthesis, as seemed to be the case in Adh5$^{-/-}$ mice, albeit mildly (Figure 3.10 c). In fact, treating HEK293T cells with HCHO leads to an increase in ASNS protein levels (Figure 3.11). This elevation in ASNS levels was only observed after acute exposure (i.e. less than 16 h, Figure 3.11); if the HCHO exposure was prolonged, the effect of elevated ASNS levels was less pronounced (Figure 3.26 a); under prolonged HCHO exposure *ASNS* mRNA levels are also mildly increased, but not in a statistically significant manner (Figure 3.26 b).

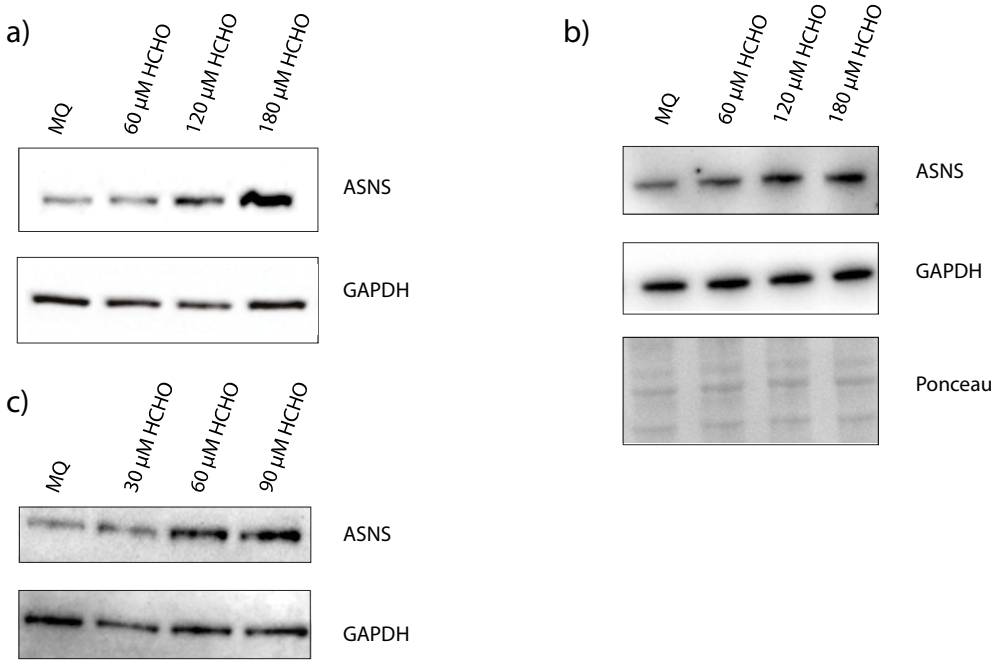


Figure 3.11 | HCHO upregulates ASNS biosynthesis *in cellulo*. a-c) HEK293T cells were treated with the indicated amounts of HCHO or MQ water control for 16 h (a+c) or 9 h (b), followed by analysis of ASNS and GAPDH protein levels by western blot.

To test whether increased ASNS biosynthesis compensates for a lack of asparagine following HCHO treatment, the protein synthesis inhibitor cycloheximide (CHX) was used. In presence of CHX, a mild but not significant apparent difference in asparagine levels was observed in the HCHO samples, whilst HCHO treatment again correlated with elevated thioproline levels (Figure 3.12 a). A potential explanation of these results is the fact that asparagine is proposed to be a 'dead-end' metabolite;[281] hence, asparagine is only used for protein biosynthesis and with CHX present, the cellular asparagine

pool cannot be depleted for protein biosynthesis.

It was then tested whether endogenous ammonium could be used by the synthetase domain of ASNS to catalyse the reaction of aspartate and ammonium, resulting in formation of asparagine. In analogy to the NMR assay (Figure 3.5 b, panel 9), it was hypothesised that ASNS which is only partially inhibited by HCHO on the N-terminal glutaminase domain can still form asparagine in presence of ammonia and aspartate. Levels of endogenous ammonia in cells can be relatively high; the amido-nitrogen of glutamine is released by cellular glutaminase activity, forming glutamate and releasing ammonium; further, spontaneous chemical decomposition of glutamine to form pyroglutamate and ammonia is significant, with an average decomposition kinetics of around 13 days at 37 °C; a reduction of growth rate and proliferation is reported at around 1-5 mM ammonium salts;[285] hence, the available ammonium present in cells might be enough for partially inactive ASNS bearing the N-terminal thiazolidine to produce sufficient asparagine in the synthetase domain from ammonium and β-aspartyl-AMP (as shown in Figure 3.5 b, panel 9). To address this possibility, HEK293T cells were cultured in 5 mM $^{15}NH_4Cl$ with and without HCHO (Figure 3.12 b). Indeed, ^{15}N-Asn is observed (Figure 3.12 b, left), but also ^{15}N-Gln (Figure 3.12 b, right). Glutamine is biosynthesised from glutamate and ammonia,[286] hence the results cannot distinguish whether ^{15}N-Asn originates from free $^{15}NH_4Cl$ or from ^{15}N-Gln that was biosynthesised from glutamate and $^{15}NH_4Cl$.

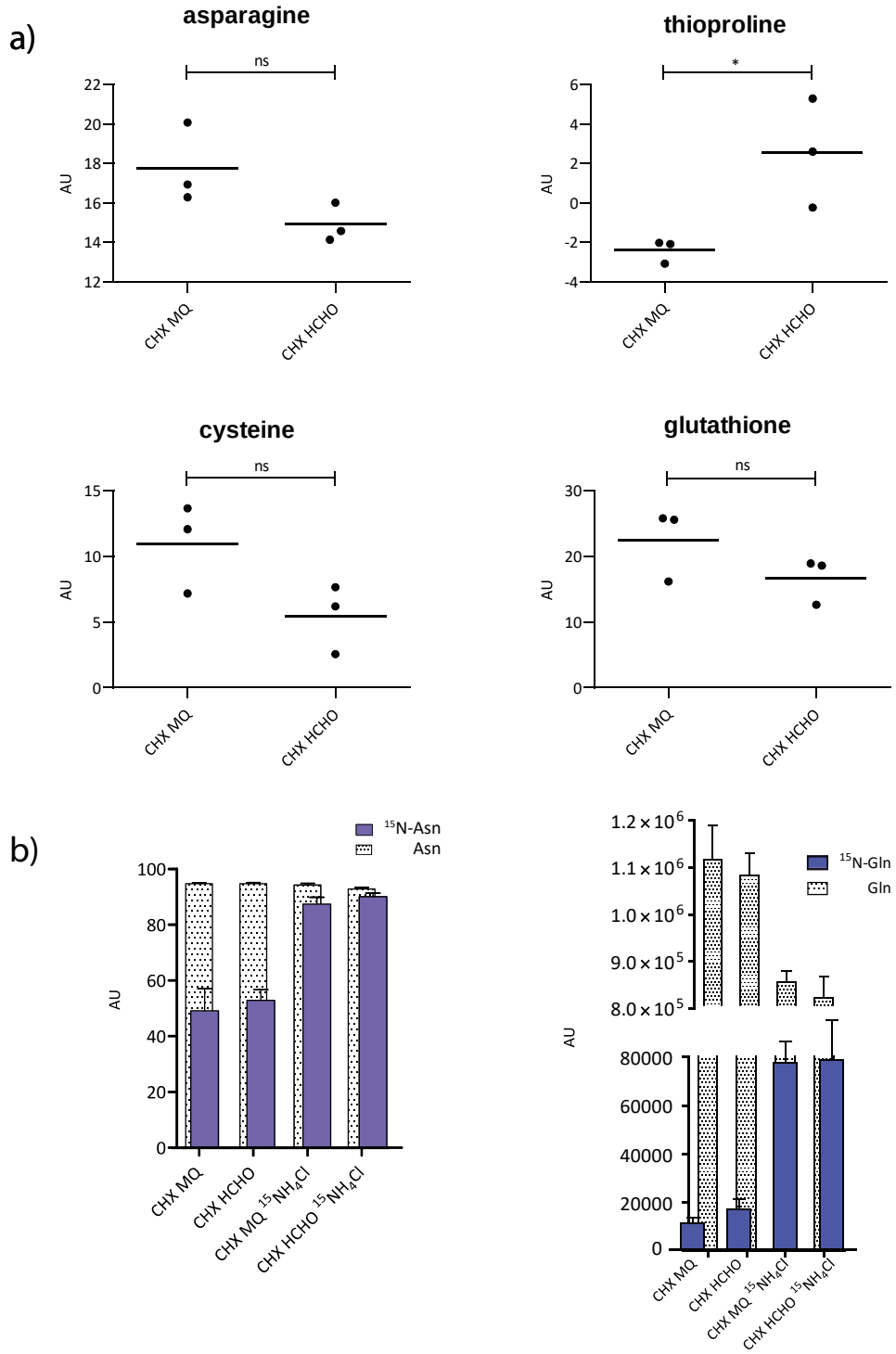


Figure 3.12 | Metabolomics analyses of HEK293T treated with cycloheximide and $^{15}NH_4Cl$. a) HEK293T cells were treated with cycloheximide (CHX, 10 µM), and additionally either HCHO (150 µM) or a MQ water control for 16.5 h. **b)** HEK293T cells were treated with cycloheximide (CHX, 10 µM), and additionally either HCHO (150 µM) or MQ water control for 16.5 h. In addition to CHX and HCHO, cells were also treated with 5 mM $^{15}NH_4Cl$ for 16.5 h. Note the ^{15}N-Asn and ^{15}N-Gln peaks in samples that were not treated with $^{15}NH_4Cl$ result from isotopic peaks of Asn and Gln. Individual data points and mean are shown (a); statistical significance was assessed using an unpaired t-test (a); *P $\leq$ 0.05, ns: not significant. Errors: SD of the mean (n=3). Metabolomics samples were prepared by myself, but run and analysed by Dr John-Walsby Tickle.

Given that fetal bovine serum (FBS) used in growth media for human cells contains aspargine,[287] it was hypothesised that asparagine present in FBS might bias the previously obtained results. Dialysed FBS is obtained by processing the raw serum through a dialysis system that has a 10 kDa cutoff filter, thereby removing - among other biomolecules - many proteins, peptides, and amino acids. Indeed, when comparing the previously used growth medium (DMEM, supplemented with GlutaMAX and FBS) with growth medium that was prepared with dialysed FBS (DMEM, supplemented with dialysed FBS), a ∼40-fold change in asparagine content based on original abundance was observed (Figure 3.13 a, normalised units are shown). In addition, it was hypothesised that when cells are seeded without glutamine/GlutaMAX, further asparagine starvation would occur, as asparagine is biosynthesised from glutamine and aspartic acid. Crucially, in presence of dialysed FBS, treating HEK293T cells with HCHO or acetaldehyde leads to very significant and extremely significant asparagine depletion, respectively (Figure 3.13 b). As observed before, thioproline levels increase upon HCHO treatment (Figure 3.13 b). Interestingly, acetaldehyde treatment leads to elevated cysteine-derived 2-methylthiazolidine-4-carboxylic acid (MTCA) (Figure 3.13 b), as described before.[288] No depletion of cysteine or glutathione was observed. Unrelated amino acids like arginine were not affected by the aldehyde treatments (Figure 3.13 b).

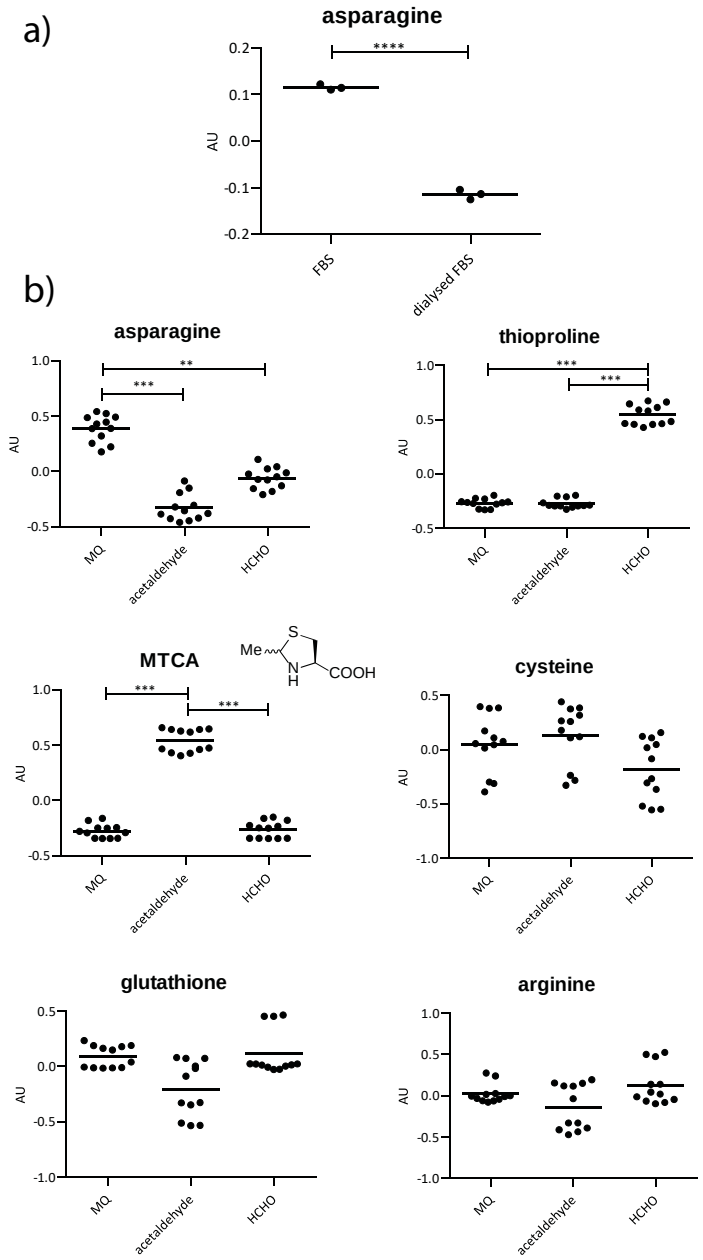


Figure 3.13 | Metabolomics analyses of HEK293T grown in presence of dialysed FBS. a) Asparagine levels in previously used growth medium (DMEM, FBS, GlutaMAX) compared to DMEM supplemented with dialysed FBS. Instrumental triplicates are shown. **b)** HEK293T cells were seeded in growth medium (DMEM, dialysed FBS, no glutamine/GlutaMAX). 8 h later, MQ water, 200 µM HCHO or 2 mM acetaldehyde were added; additionally, all experimental groups were supplemented with 2 mM glutamine. 15 h later, the same amount of aldehydes or MQ water was added, followed by harvest 8 h later. Individual data points of biological quadruplicates ran as instrumental triplicates (12 data points per group) and mean are shown, statistical significance was assessed based on the average of the instrumental triplicates using 1way ANOVA; ****P $\leq$ 0.0001, ***P $\leq$ 0.001, **P $\leq$ 0.01, *P $\leq$ 0.05. Metabolomics samples were prepared and analysed by myself, but run by Dr John-Walsby Tickle. The structure of the cysteine-acetaldehyde adduct is given, 2-methylthiazolidine-4-carboxylic acid (MTCA).

These results were reproduced in an independent experiment with HEK293T cells. HCHO, acetaldehyde and methylglyoxal led to depletion of asparagine (Figure 3.14). Thioproline is elevated only in cells treated with HCHO, but also in cells treated with methylglyoxal, consistent with the observation that commercial methylglyoxal contains trace amounts of HCHO (Figure 3.14).[289] The cysteine-acetaldehyde adduct MTCA is only elevated in cells treated with acetaldeyde (Figure 3.14). Asparagine and MTCA levels in EtOH-treated cells were not altered relative to cells treated with MQ water, consistent with high expression levels of alcohol dehydrogenase in liver tissue, but not in kidney tissues like HEK293T used in this study.[290] The GSH-aldehyde adducts could not be detected under our conditions, possibly due to its low stability and fast turnover in cells. The cysteine-methylglyoxal adduct was not detected, likely due the instability of the cysteine-methylglyoxal adduct. Some unrelated amino acids like alanine, arginine, and histidine were observed to be depleted in a similar way as asparagine in the aldehyde-treated samples (Figure 3.14), which might be consistent with the role of asparagine as amino acid exchange factor.[281]

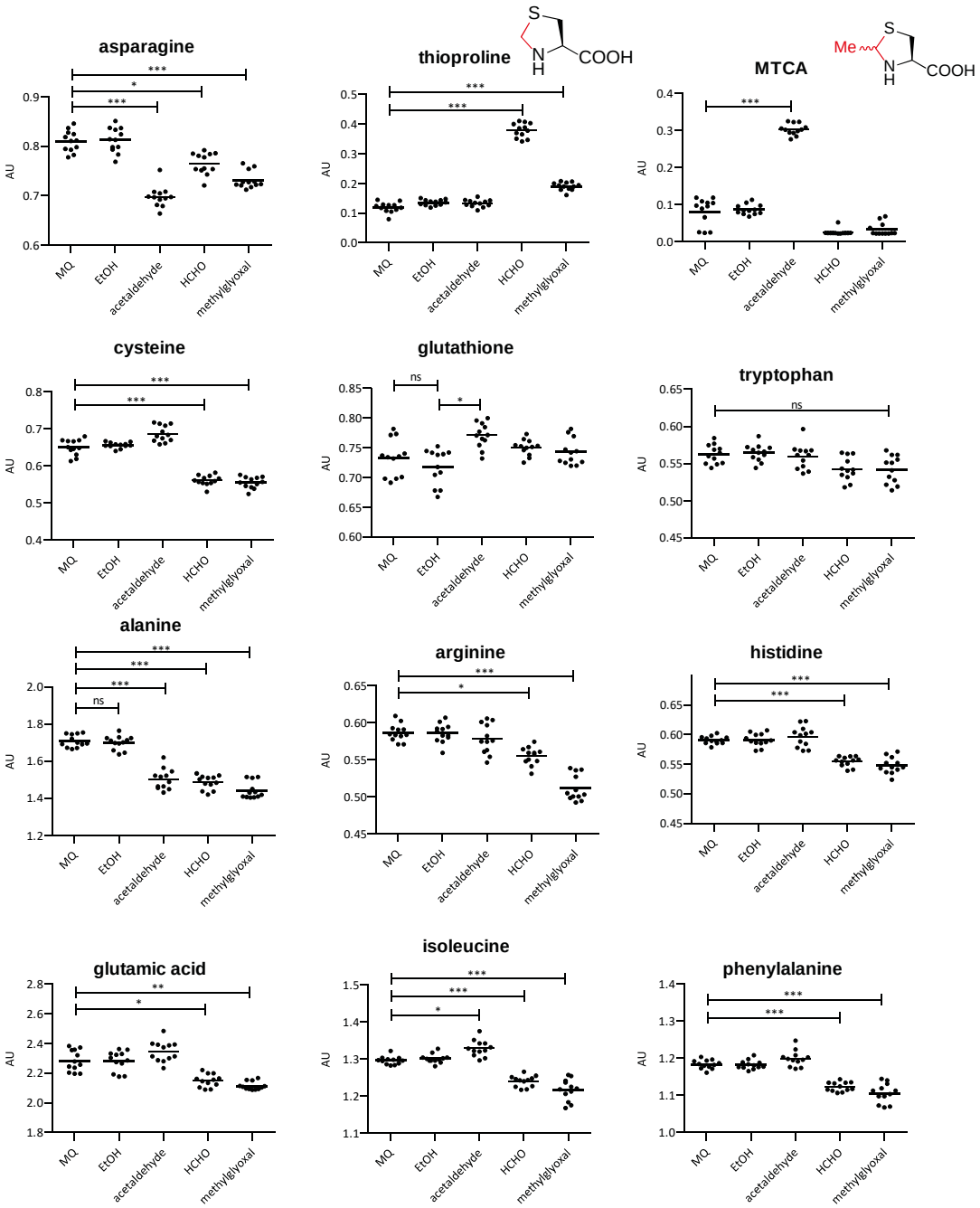


Figure 3.14 | Metabolomics analyses of HEK293T grown in presence of dialysed FBS. HEK293T cells were seeded in growth medium (DMEM, dialysed FBS, no glutamine/GlutaMAX). 7.5 h later, MQ water, 17 mM ethanol (EtOH), 2 mM acetaldehyde, 150 µM HCHO or 750 µM methylglyoxal were added. All experimental groups were supplemented with 2 mM glutamine. 15.5 h later, the same amount of aldehydes or MQ water was added (with the exception of methylglyoxal, where 250 µM were added). Cells were harvested 8 h later. The structures of the cysteine-acetaldehyde adduct, 2-methylthiazolidine-4-carboxylic acid (MTCA), and the cysteine-HCHO adduct thioproline are given. Individual data points of biological quadruplicates ran as instrumental triplicates (12 data points per group) and means are shown. Statistical significance was assessed based on the average of the instrumental triplicates using 1way ANOVA; ****P $\leq$ 0.0001, ***P $\leq$ 0.001, **P $\leq$ 0.01, *P $\leq$ 0.05. Metabolomics samples were prepared and analysed by myself, but run by Dr John-Walsby Tickle.

In mice treated with either saline or EtOH, EtOH resulted in decreased asparagine levels, especially in *Aldh2*$^{-/-}$ mice (Figure 3.15). Given that Aldh2 converts acetaldehyde to acetic acid,[170] the *Aldh2*$^{-/-}$ mice likely accumulate acetaldehyde, which might inhibit ASNS, resulting in lower asparagine levels. Interestingly, the cysteine-acetaldehyde adduct was not detected in any of these mouse samples.

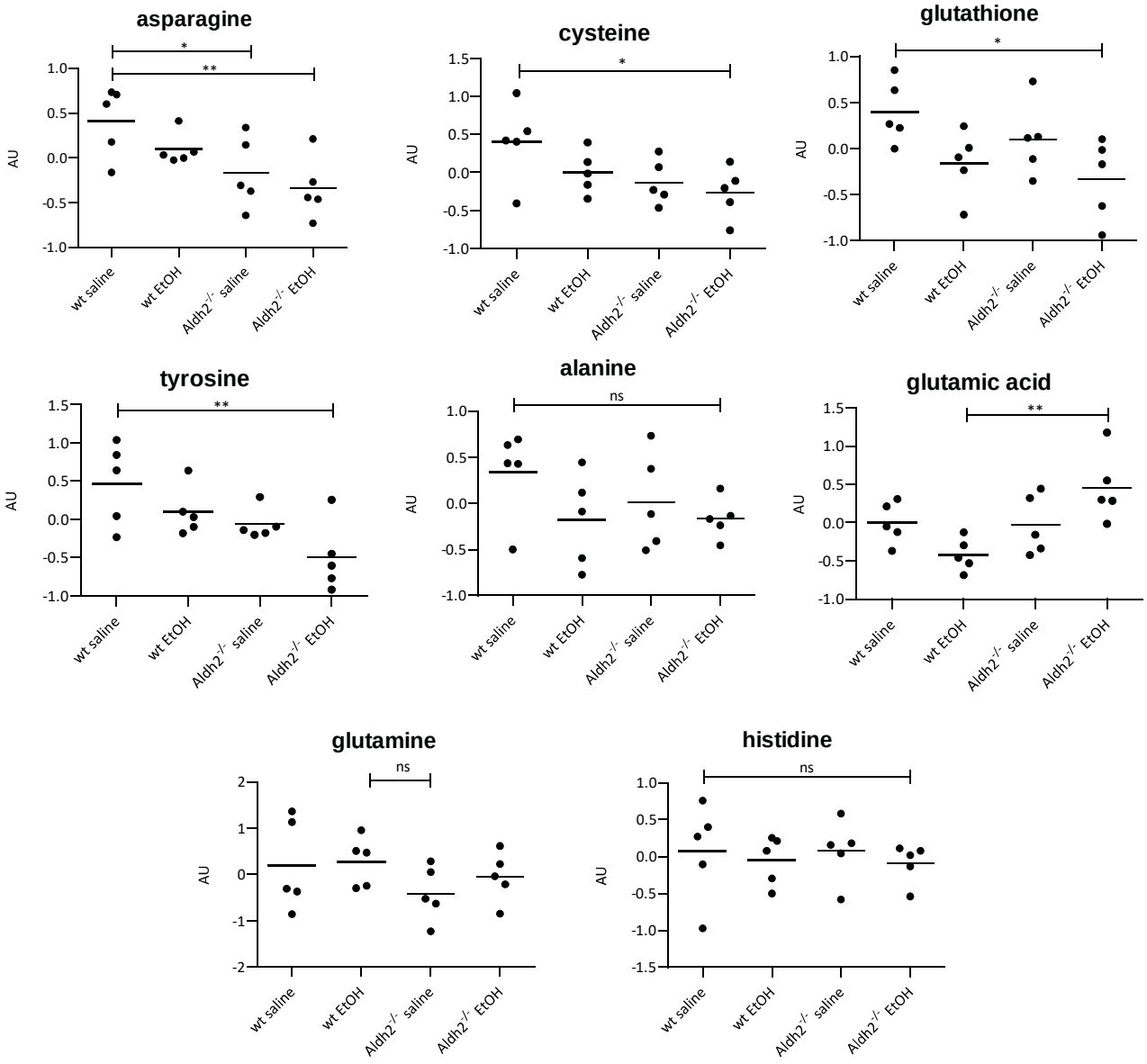


Figure 3.15 | Metabolomics analyses of mouse livers. Mice were treated with either saline as a control or EtOH. Individual data points of biological replicates (n=5) and means are shown. Statistical significance was assessed using 1way ANOVA; ****P $\leq$ 0.0001, ***P $\leq$ 0.001, **P $\leq$ 0.01, *P $\leq$ 0.05. Metabolomics samples were prepared and analysed by myself, but run by Dr John-Walsby Tickle. The mouse experiment was conducted by Dr Felix Dingler and liver samples were homogenised by Dr Christopher Millington.

3.4 Discussion

As cells routinely produce HCHO, it is conceivable that stable HCHO adducts with regulatory or sensing roles form over time, with the potential for HCHO being directly involved in regulating chromatin structure and function. A drawback in HCHO research is the lack of established quantification methods for biological samples.[245] Especially subcellular levels of HCHO may be vastly different than generally appreciated. Chromatin is highly dynamic in its methylation state, which probably leads to high local concentrations of HCHO in the nucleus.[115] Cysteine is likely one of the most important biomolecules in HCHO relevant biology.[82, 86, 257] The studies described here on the reactions of HCHO and the model enzyme ASNS B which bears an N-terminal cysteine show selective inhibition of its glutaminase domain, while the synthetase domain is still intact (Figure 3.5 b). Crucially, acetaldehyde and the α-oxoaldehydes methylglyoxal and glyoxal modulate ASNS B activity (Figure 3.8 d). As these aldehyde compounds are both elevated in alcoholics and type 2 diabetes patients, respectively, the results presented here provide a potential molecular mechanism for the observation that these two groups have low asparagine levels.[272, 273]

Metabolomics analyses provided insights in how human cell lines and mice respond to HCHO. Thioproline levels were consistently elevated (Figures 3.9, 3.10, 3.12, 3.14). This observation highlights the efficiency of detoxification systems that have evolved around toxic endogenous aldehydes like HCHO. Crucially, aldehyde treatment leads to depletion of asparagine levels (Figures 3.10 a, 3.13 b, 3.14, 3.15). Acetaldehyde treatment resulted in formation of the cysteine-acetaldehyde adduct MTCA. Reactions of cysteine with aldehydes are therefore likely a major pathway of aldehyde detoxification. In fact, cysteine has been described to alleviate hangover symptoms;[291] the presented data showing that cysteine scavenges toxic acetaldehyde to form MTCA might provide an explanation for hangover alleviation by cysteine. Asparagus, which is high in asparagine, has been reported to alleviate hangover symptoms[292] and the use of aspartic acid (the substrate for ASNS) and asparagine itself is patented (EP0631778A1) to prevent alcohol toxicity. The presented data may provide the molecular mechanism for the positive effects of cysteine, aspartic acid and asparagine against hangovers.

Furthermore, cells react to HCHO with an apparent increase in ASNS levels (Figure 3.11), suggesting a minor decrease in asparagine availability caused by ASNS inhibition can be sensed and responded to. Acetaldehyde levels in blood are reported to be relatively low, i.e. <0.5 µM.[293] However, blood acetaldehyde levels in persons that are heterozygotes for inactive ALDH2 are reported to increase to ∼30 µM (∼80 µM in saliva) after consuming one beer, whereas blood acetaldehyde levels of homozygotes for active ALDH2 stay at baseline levels (∼40 µM in saliva).[294] In diseases like glutathione synthetase deficiency, or upon administration of high doses of paracetamol or intake of food high in acrylamide, GSH levels are low,[186, 189, 284] which might make cells more susceptible to reactions with endogenous aldehydes. In fact, alcoholics and diabetes type 2 patients, who have high levels of acetaldehyde and α-oxoaldehydes, respectively, have low levels of asparagine.[272, 273]

Given that ASNS levels are low in acute lymphoblastic leukemia (ALL),[277] it is conceivable to develop molecules that release reactive aldehydes upon binding to ASNS; in combination with ASNase, this could be a potent treatment against this form of cancer (Figure 3.16).

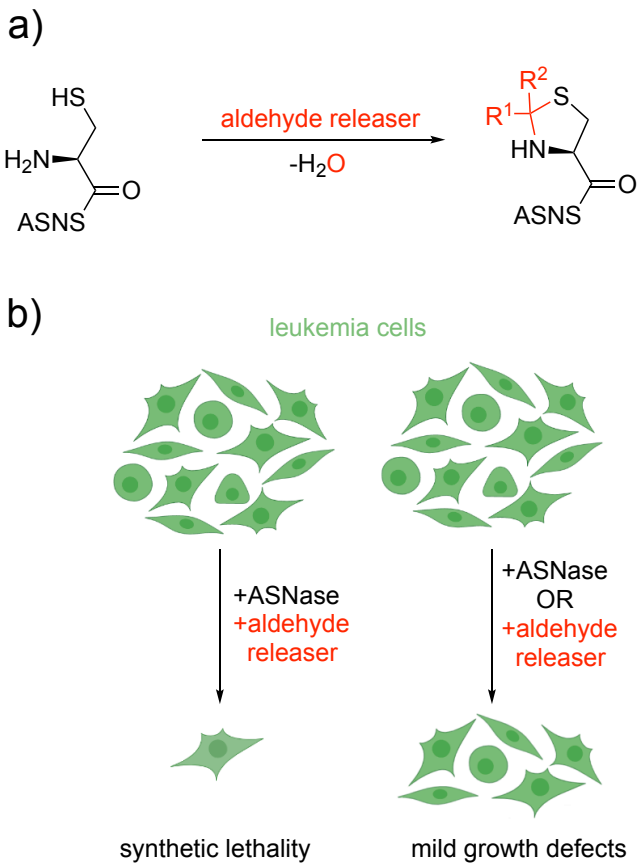


Figure 3.16 | Proposed combination therapy. a) A molecule could be developed that binds to ASNS and releases aldehyde moieties over time, which could result in a high local concentration of aldehydes resulting in ASNS inhibition. **b)** As ASNS levels are low in acute lymphoblastic leukemia (ALL), these cells are particularly vulnerable to asparagine depletion by asparaginase (ASNase).[277] In combination with aldehyde releasers, a synthetic lethality could hamper ALL cell growth even more.

3.5 Supplementary Information

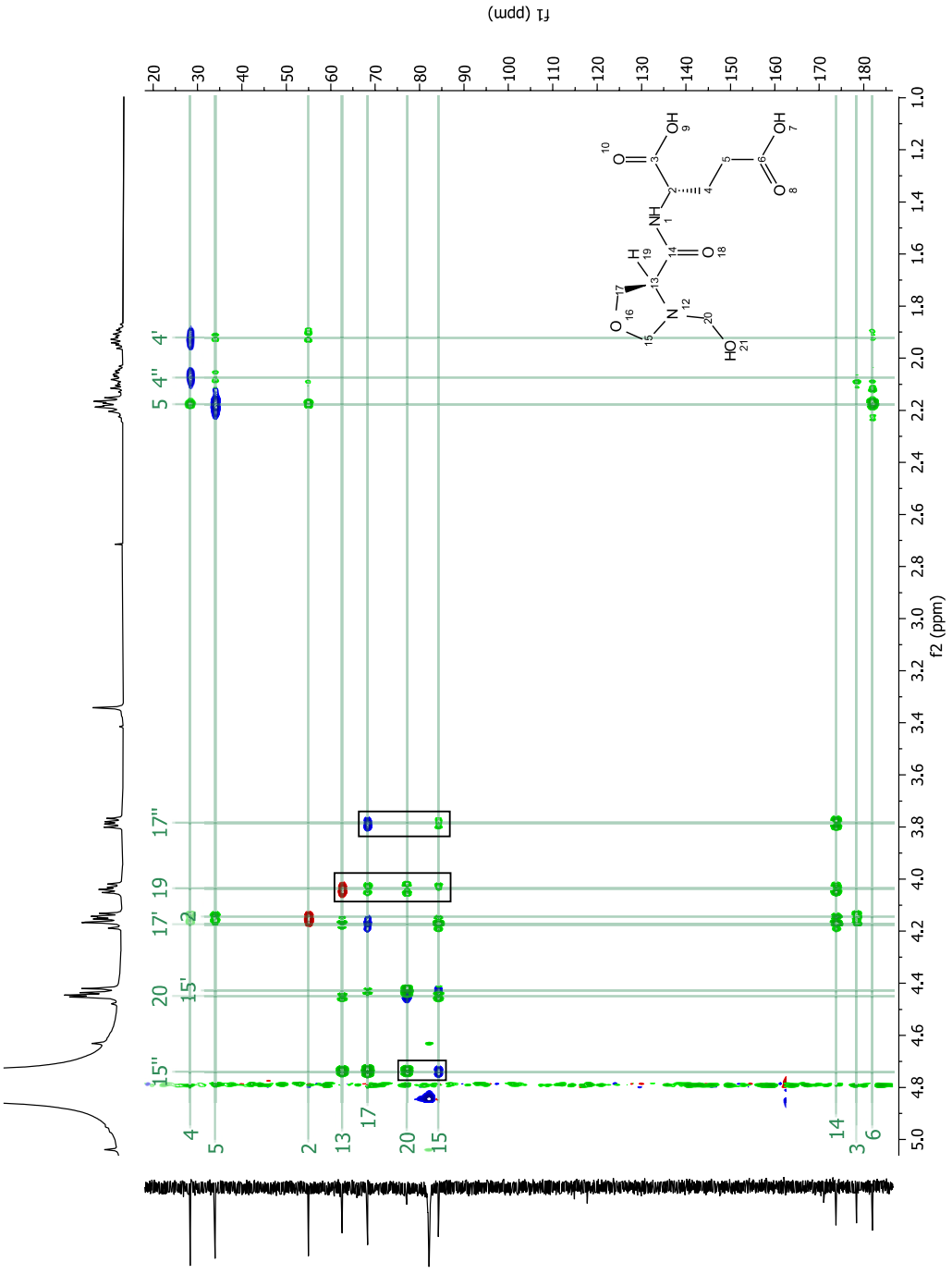


Figure 3.17 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC spectra for *in situ* generated SerGlu-derived 1a (400 MHz). Important interactions are highlighted in a rectangle. The x-axis shows the ^{1}H spectrum and the y-axis shows the ^{13}C spectrum.

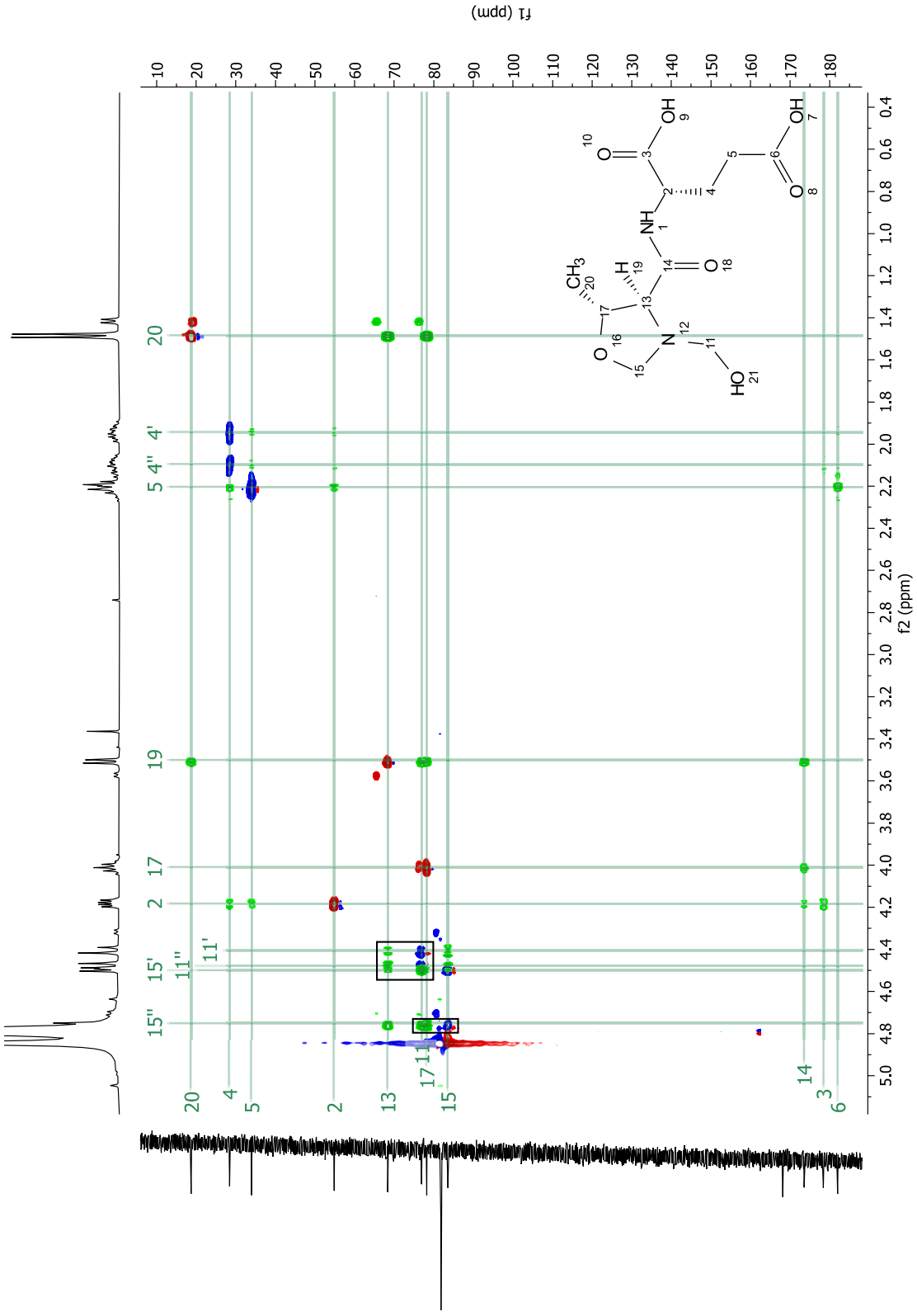


Figure 3.18 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC spectra for *in situ* generated ThrGlu-derived 2a (400 MHz). Important interactions are highlighted in a rectangle.

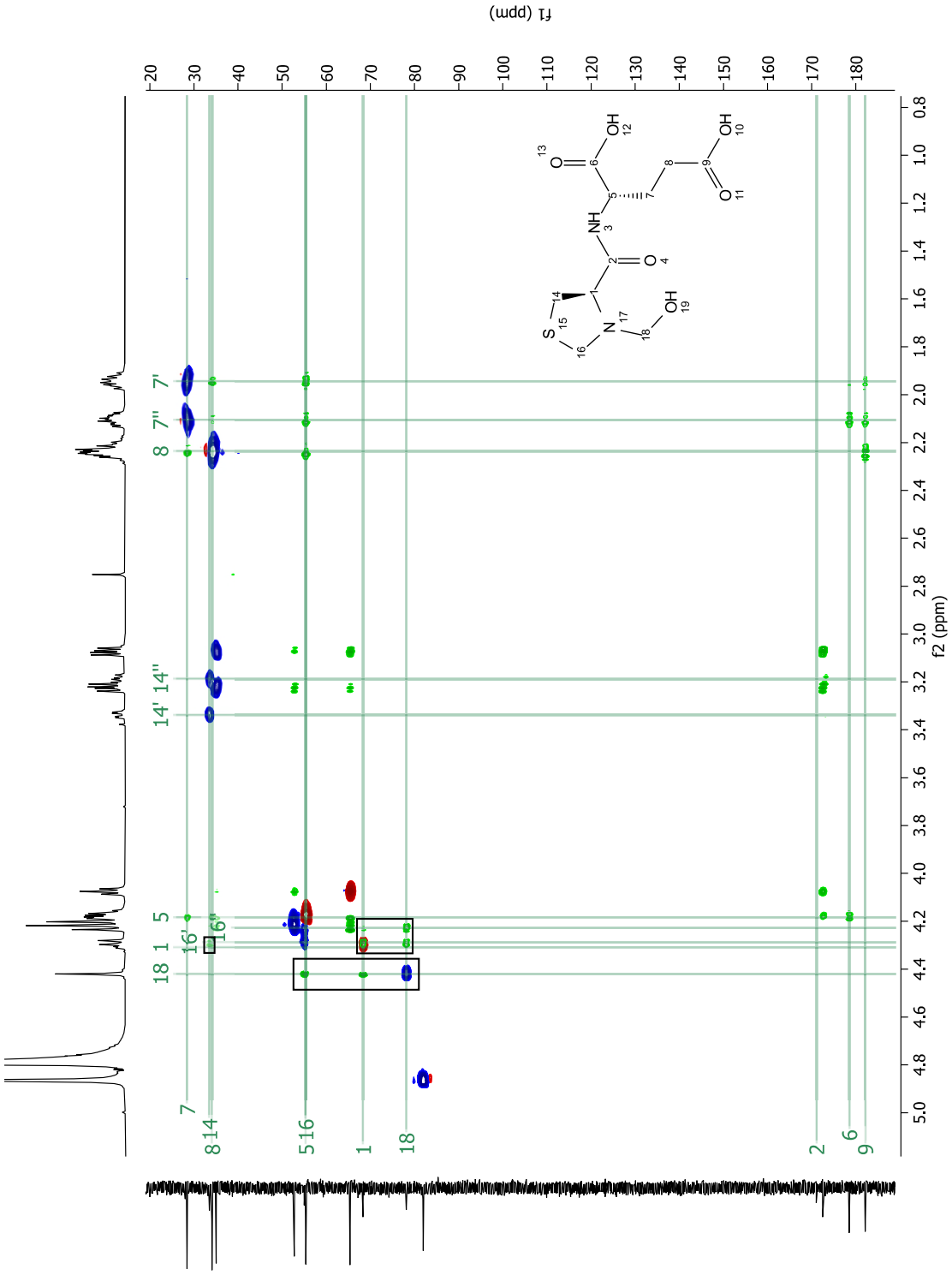


Figure 3.19 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC spectra for *in situ* generated CysGlu-derived 3a (600 MHz). Important interactions are highlighted in a rectangle. Spectrum is taken 3 days after addition of HCHO to promote formation of the hemiaminal.

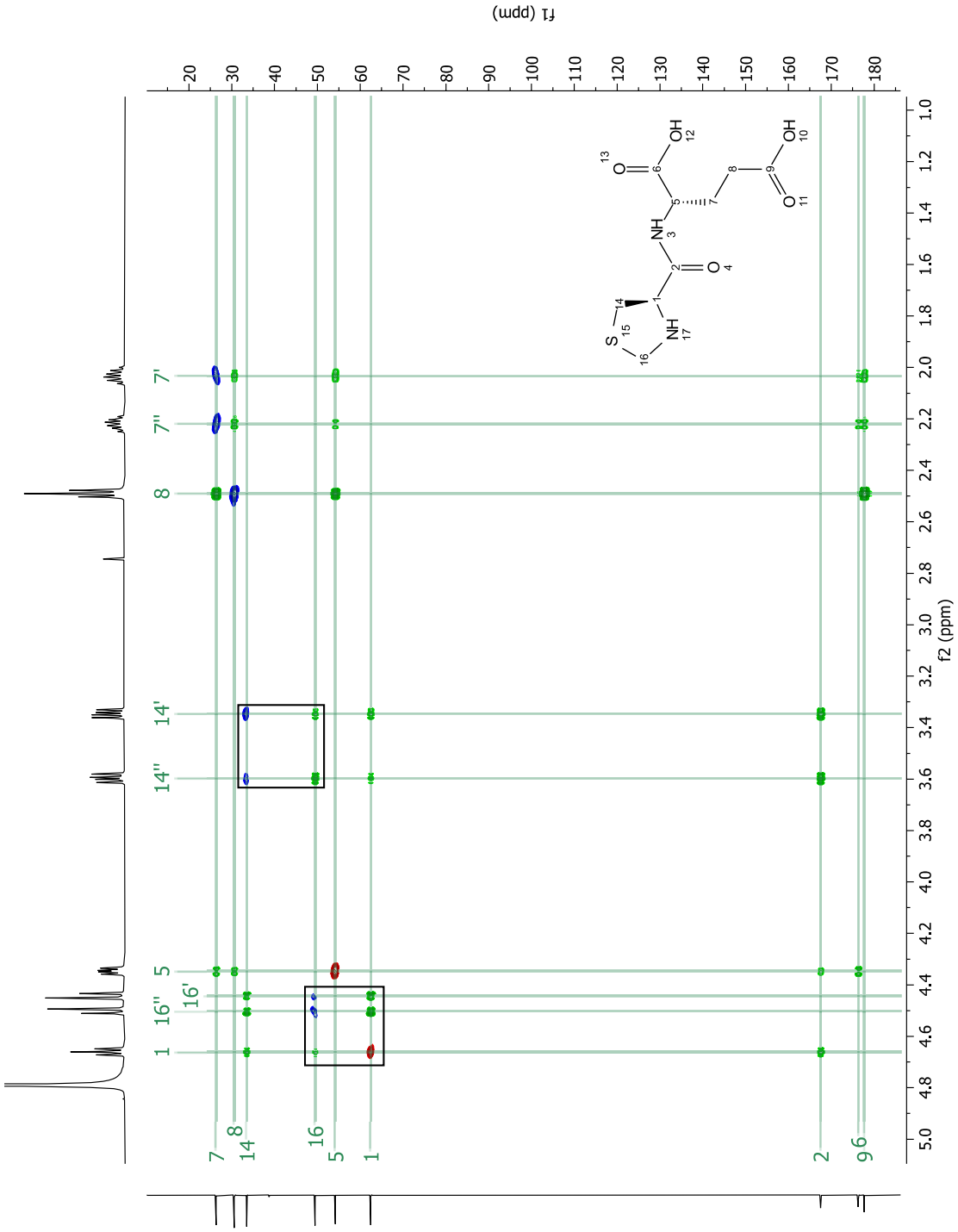


Figure 3.20 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC spectra for HPLC purified CysGlu-derived 3b (600 MHz). Important interactions are highlighted in a rectangle.

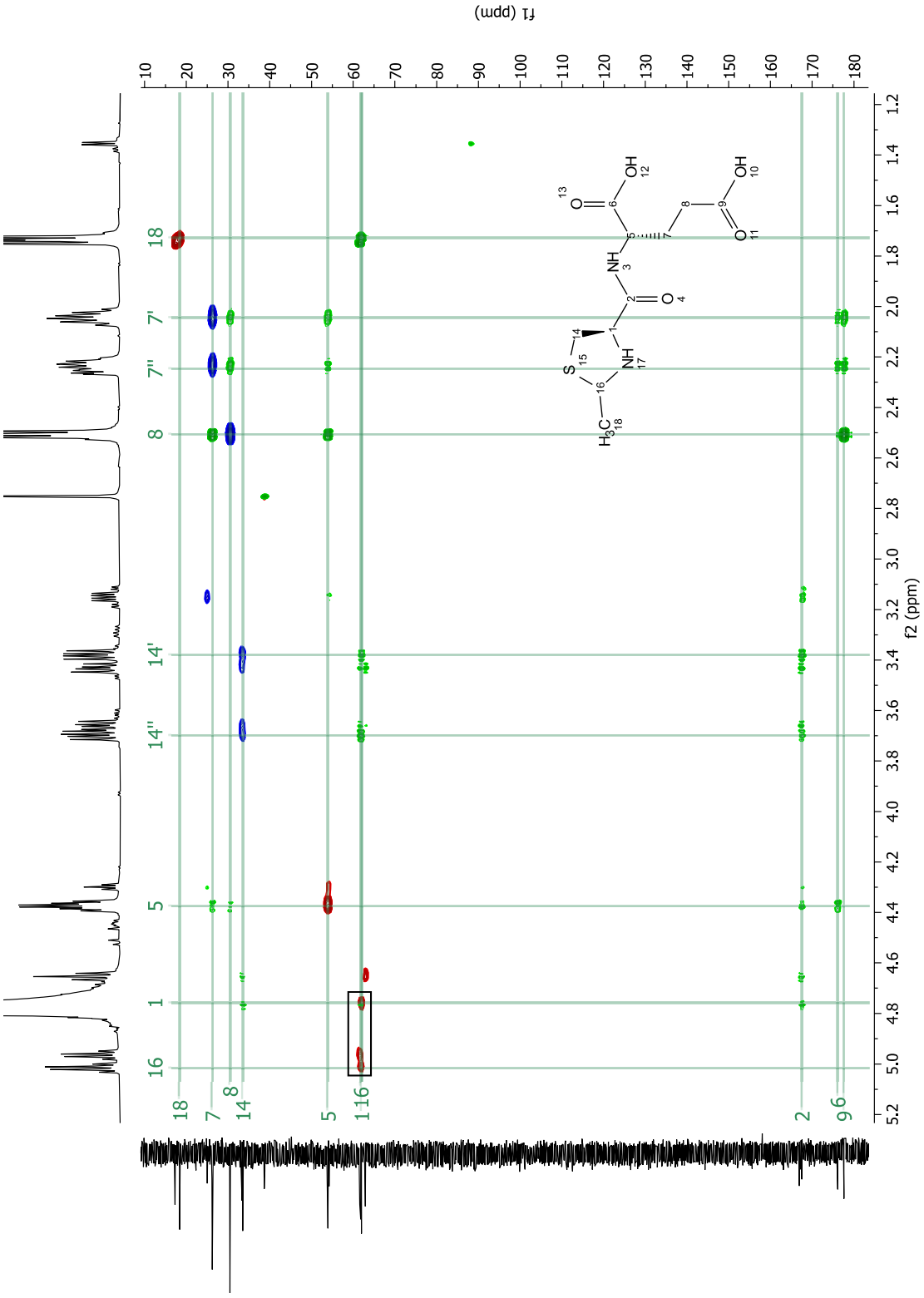


Figure 3.21 | Overlay of 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMBC spectra for HPLC purified CysGlu-derived 3c (600 MHz). Important interactions are highlighted in a rectangle.

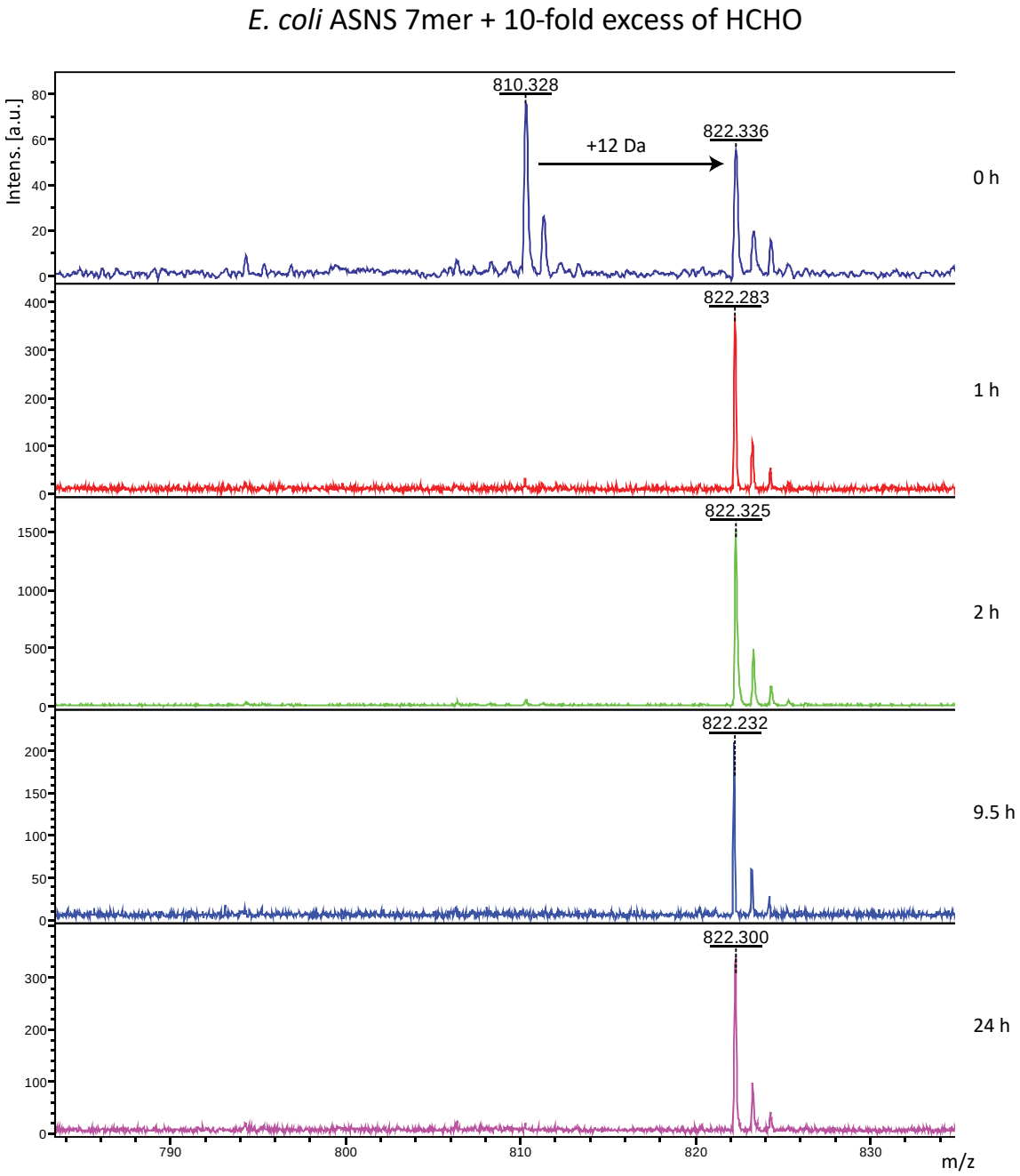


Figure 3.22 | Exemplary MALDI MS data obtained from addition of HCHO to the N-terminal ASNS B 7mer peptide. The reaction mixture (2.5 µM peptide concentration, 25 µM HCHO, 50 mM potassium phosphate at pH 7.4) was analysed directly after addition of HCHO (0 h), after 1 h, 2 h, 9.5 h and 24 h by MALDI MS. Only the K^+ adducts of the peptide substrate and product were observed. Peaks show parent peptide $[M+K]^+$: 810 Da; +12 Da HCHO adduct $[M+K]^+$: 822 Da. Peak areas were used to generate the graph in Figure 3.5 a. The 7mer peptide (NH_2-CSIFGVF-NH_2) corresponds to the first 7 residues of *E. coli* ASNS B.

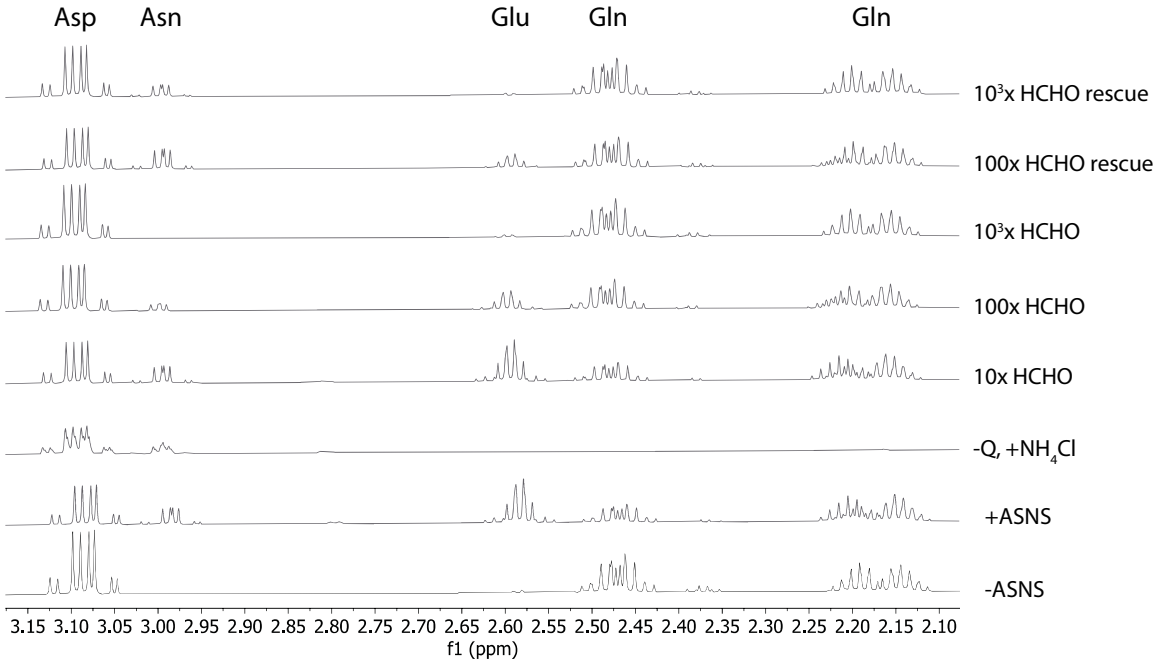


Figure 3.23 | HCHO selectively inhibits the ASNS B glutaminase reaction - independent replicate. ^{1}H NMR (700 MHz) showing the independent replicate of Figure 3.5 b.

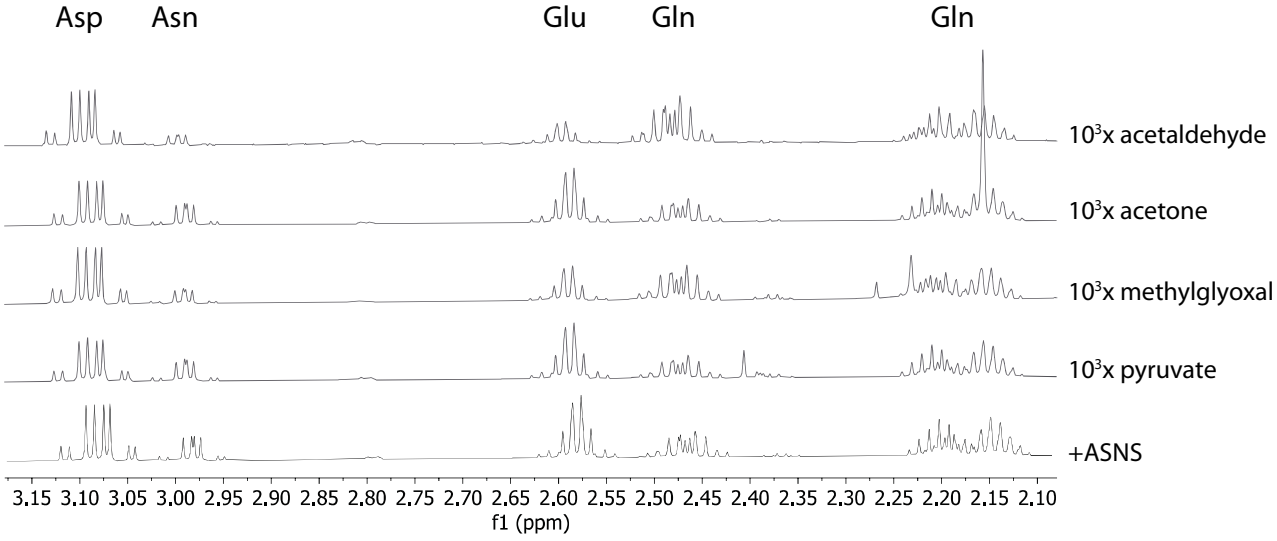


Figure 3.24 | Endpoint assay results for screening of carbonyl compounds as potential ASNS B inhibitors. ^{1}H NMR (700 MHz) showing biologically relevant carbonyl compounds (1000-fold molar excess) that were tested for their ability to inhibit ASNS B. ASNS B was preincubated with the indicated carbonyl compounds for 20 min at 37 °C, which was followed by addition of the preincubation mix to the substrates. After 4 h at 37 °C, the reaction was quenched with CF_3CO_2H and analysed by ^{1}H NMR.

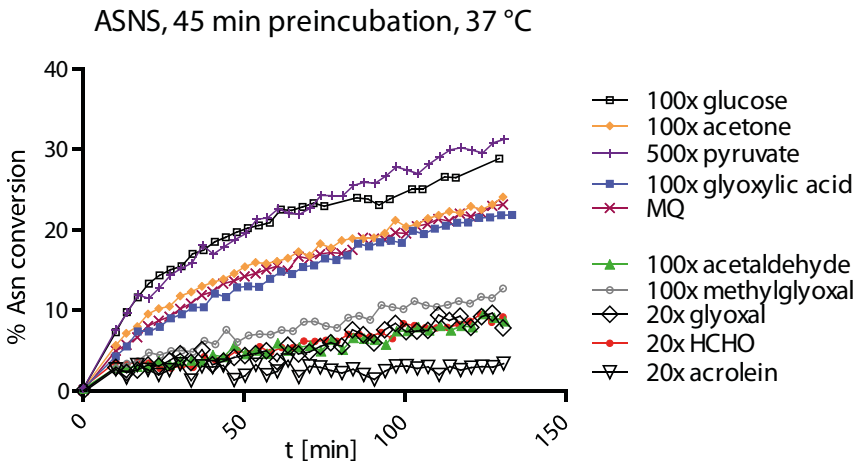


Figure 3.25 | ASNS B is inhibited by several aldehydes. ^{1}H NMR (700 MHz) showing the independent replicate of Figure 3.8 d.

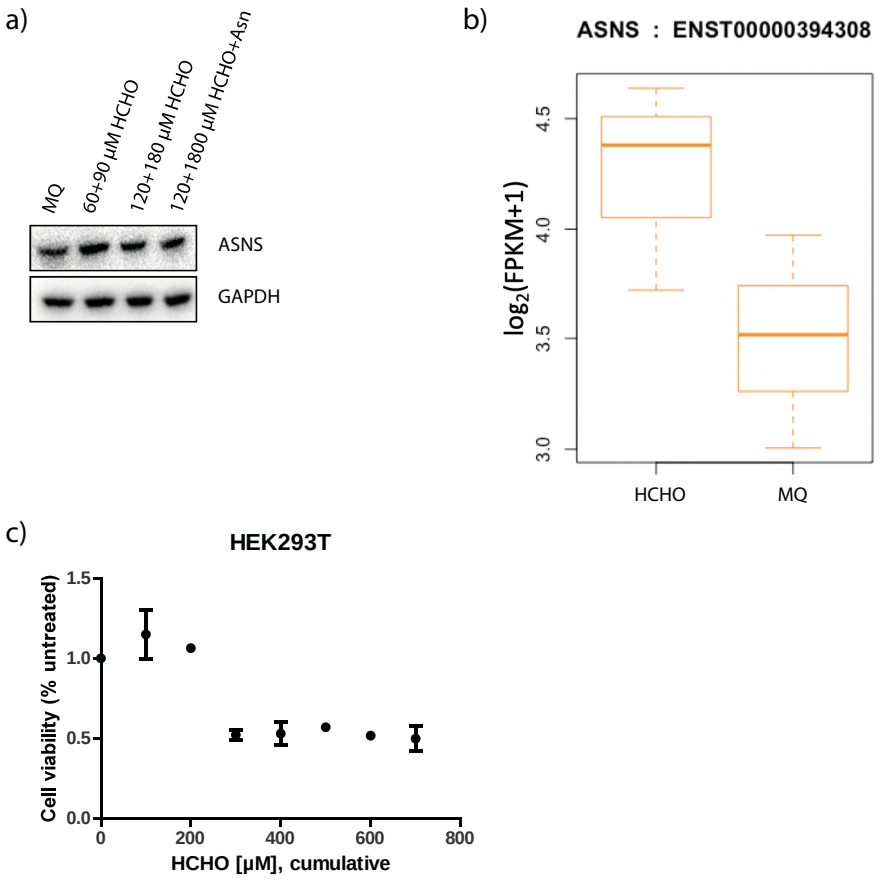


Figure 3.26 | Western blot and RNA-Seq studies. a) HEK293T cells were treated with a first dose of HCHO, followed by the second dose 19 h after that, with cell harvest 7 h after the second treatment. The addition of 5 mM asparagine did not alter ASNS levels (rightmost lane). **b)** The ASNS mRNA levels in HEK293T cells, either treated with HCHO (150+225 µM) or MQ water. Gene ID MSTRG.24406, fc=0.58, p=0.298, q=0.594. Errors: SD of the mean (n=3). Raw data were obtained from the Oxford Genomics Centre. **c)** HEK293T cells were treated with two cumulative doses (i.e. without exchanging medium between the two treatments) of HCHO which were 24 h apart. 14 h after the second treatment, viable cells were estimated using CellTiter 96 AQueous One Solution Reagent. Errors: SEM (n=3, technical repeats).

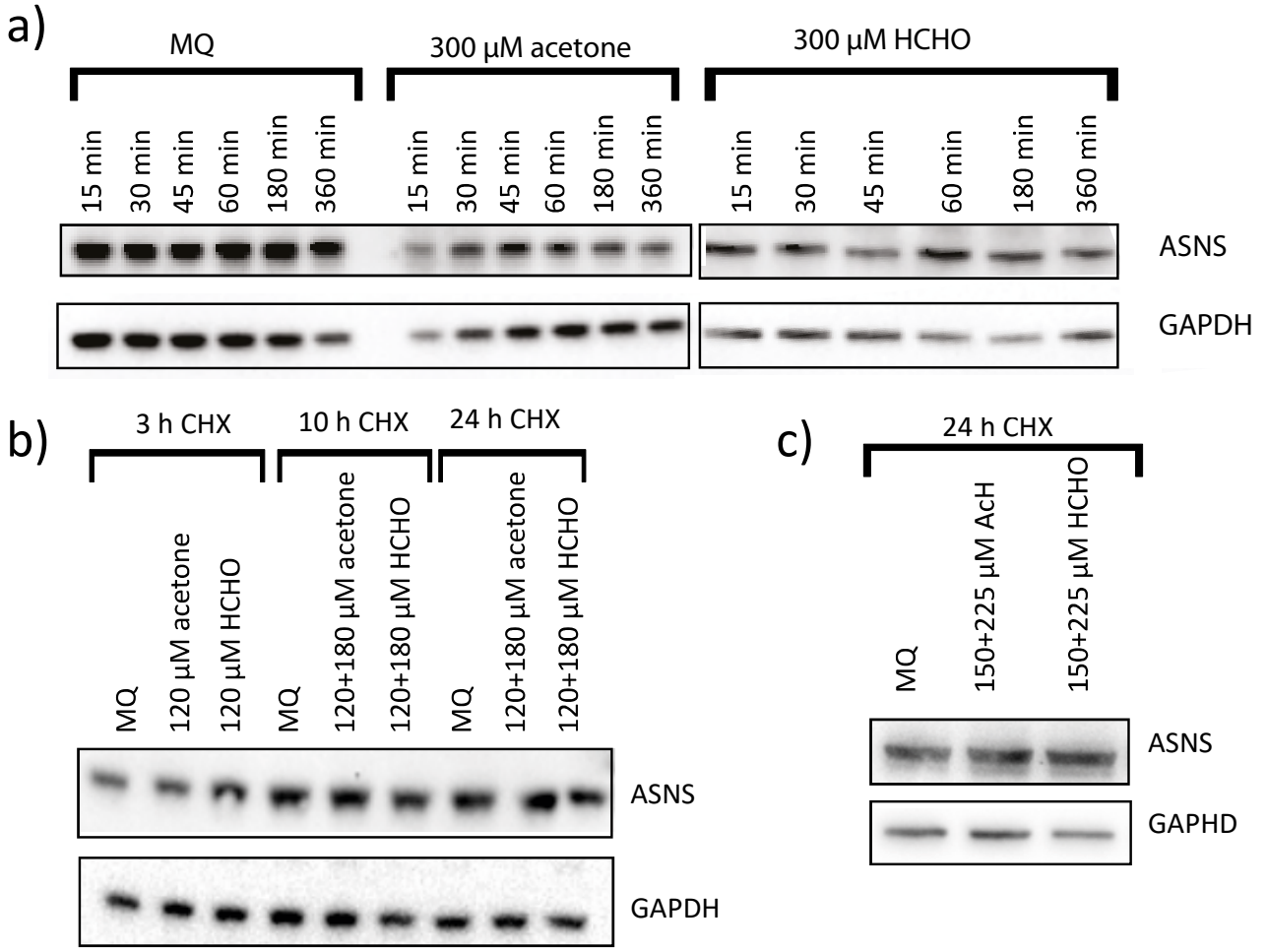


Figure 3.27 | Stability studies on ASNS in cells treated with cycloheximide (CHX). a) HEK293T cells were treated with the indicated amounts of HCHO, acetone, or MQ water control for 14 h. Cells were then treated with CHX and harvested at the indicated timepoints. **b)** HEK293T cells were treated with 120 µM HCHO (or acetone, or MQ water control; 100 mM stock solutions of carbonyl compounds in MQ water) for 14 h. The cells were then treated with 100 µM CHX and were harvested 3 h later. Additionally, HEK293T cells were treated with 120 µM HCHO (or acetone, or MQ water control) for 14 h, followed by addition of CHX. 7.5 h after the CHX treatment, a further treatment of 180 µM carbonyl compound was added. The cells were then harvested 10 h and 24 h after the CHX addition. **c)** HEK293T cells were treated with MQ water/acetone/HCHO overnight and subsequently treated with a second dose of MQ water/acetone/HCHO along with CHX, and harvested after 24 h. The final CHX concentration was 100 µM in all experiments (from a 100 mM DMSO stock).

3.6 Methods

3.6.1 Synthesis of dipeptides

The synthesis of dipeptides was performed as described in Chapter 2.6.3. Starting materials used were Boc-Cys(Trt)-OH (Insight Biomtechnology Ltd), Boc-O-tert-butyl-serine (Alfa Aesar), and

Boc-Thr(tBu)-OH (Fluorochem Ltd).

Boc-*L*-Cys(Trt)-*L*-Glu(di-tBu-ester) ^{1}H NMR (400 MHz, $CDCl_3$) δ_H = 7.45 – 7.38 (m, 6H, 35, 37, 42, 44, 47, 49), 7.33 – 7.25 (m, 6H, 34, 38, 41, 45, 46, 50), 7.25 – 7.17 (m, 3H, 36, 43, 48), 6.61 (d, J = 7.9 Hz, 1H, 6), 4.80 (d, J = 7.7 Hz, 1H, 1), 4.40 (td, J = 7.9, 4.9 Hz, 1H, 8), 3.85 (brs, 1H, 2), 2.81 – 2.69 (appdd, m, 1H, 4'), 2.49 (dd, J = 12.8, 5.1 Hz, 1H, 4"), 2.35 – 2.19 (m, 2H, 11), 2.18 – 2.05 (m, 1H, 10'), 1.92 – 1.79 (m, 1H, 10"), 1.43 (s, 9H, 29, 30, 31), 1.42 (s, 9H, 18, 19, 20), 1.40 (s, 9H, 22, 23, 24); ^{13}C NMR (101 MHz, $CDCl_3$) δ_C = 172.2 (C12), 170.6 (C9), 170.3 (C3), 155.4 (C25), 144.5 (C33, C39, C40), 129.7 (C35, C37, C42, C44, C47, C49), 128.2 (C34, C38, C41, C45, C46, C50), 127.0 (C36, C43, C48), 82.4 (C17), 80.6 (C21), 77.4 (C28), 67.3 (C32), 53.7 (C2), 52.3 (C8), 33.7 (C4), 31.4 (C11), 28.4 (C18, C19, C20), 28.2 (C29, C30, C31), 28.1 (C22, C23, C24), 28.0 (C10); HRMS (ESI+) $(M+H)^+$: 727.34115 (calculated for $C_{42}H_{51}N_2O_7S_1$), 727.33818 (observed). Yield: 644 mg, 0.91 mmol, 73%.

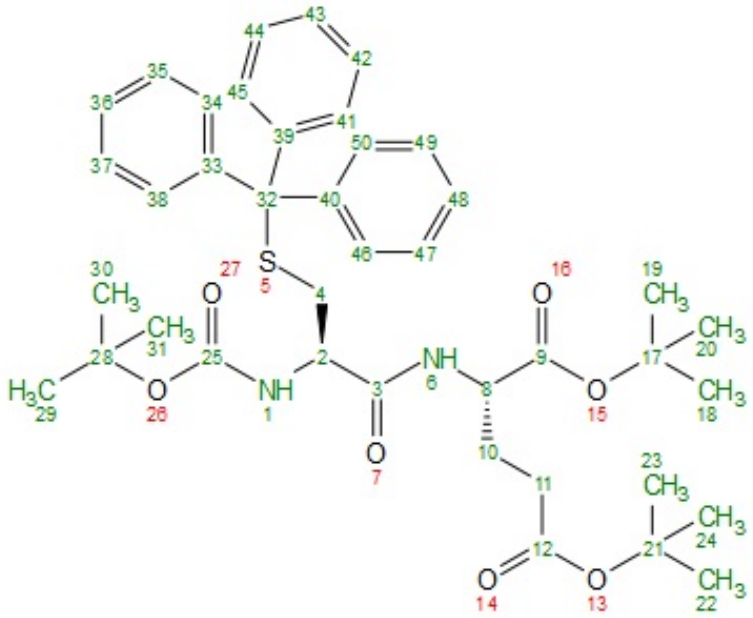


Figure 3.28 | *L*-Cys-*L*-Glu dipeptide with Boc, trityl and tert-butyl protection.

Boc-*L*-Ser(tBu)-*L*-Glu(di-tBu-ester) ^{1}H NMR (400 MHz, $CDCl_3$) δ_H = 7.17 (brs, 1H, 5), 5.41 (brs, 1H, 1), 4.53 – 4.44 (m, 1H, 6), 4.15 (brs, 1H, 2), 3.81 (dd, J = 8.8, 3.4 Hz, 1H, 4'), 3.38 (appdd, t, J = 7.6 Hz, 1H, 4"), 2.36 – 2.19 (m, 2H, 11), 2.19 – 2.08 (m, 1H, 10'), 1.92 – 1.80 (m, 1H, 10"), 1.45 (s, 18H, 21, 22, 23, 33, 34, 35), 1.43 (s, 9H, 25, 26, 27), 1.19 (s, 9H, 29, 30, 31); ^{13}C NMR (101

MHz, $CDCl_3$) δ_C = 172.2 (C12), 170.9 (C8), 170.6 (C3), 155.5 (C17), 82.3 (C32), 80.6 (C24), 80.2 (C20), 74.0 (C28), 62.0 (C4), 54.6 (C2), 52.2 (C6), 31.4 (C11), 28.5 (C33, C34, C35), 28.2 (C21, C22, C23), 28.1 (C10), 27.5 (C25, C26, C27), 27.1 (C29, C30, C31); HRMS (ESI+) $(M+H)^+$: 503.3327 (calculated for $C_{25}H_{47}N_2O_8$), 503.3326 (observed). Yield: 399 mg, 0.79 mmol, 64%.

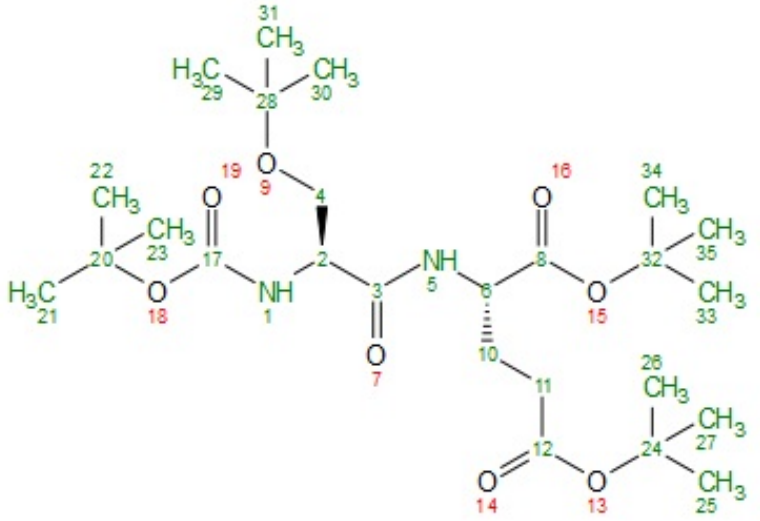


Figure 3.29 | *L*-Ser-*L*-Glu dipeptide with Boc and tert-butyl protection.

Boc-*L*-Thr(tBu)-*L*-Glu(di-tBu-ester) ^{1}H NMR (600 MHz, $CDCl_3$) δ_H = 7.55 (d, J = 7.6 Hz, 1H, 5), 5.59 (d, J = 5.7 Hz, 1H, 1), 4.42 (td, J = 7.6, 5.3 Hz, 1H, 6), 4.14 – 4.09 (m, 1H, 4), 4.12 – 4.08 (m, 1H, 2), 2.34 – 2.18 (m, 2H, 11', 11"), 2.13 – 2.04 (m, 1H, 10'), 1.97 – 1.89 (m, 1H, 10"), 1.44 (dd, J = 16.3, 8.3 Hz, 27H, 27, 28, 29), 1.26 (s, 9H, 19, 20, 21), 1.08 (d, J = 6.2 Hz, 3H, 17); ^{13}C NMR (151 MHz, $CDCl_3$) δ_C = 172.0 (C12), 170.6 (C8), 170.0 (C3), 155.7 (C30), 82.2 (C22), 80.7 (C26), 79.7 (C33), 75.3 (C18), 67.1 (C4), 58.6 (C2), 52.6 (C6), 31.6 (C11), 28.5 (C23, C24, C25), 28.2 (C34, C35, C36), 28.1 (C19, C20, C21), 27.8 (C10), 27.0 (C27, C28, C29), 17.4 (C17); HRMS (ESI+) $(M+H)^+$: 517.3483 (calculated for $C_{26}H_{49}N_2O_8$), 517.3484 (observed). Yield: 481 mg, 0.93 mmol, 83%.

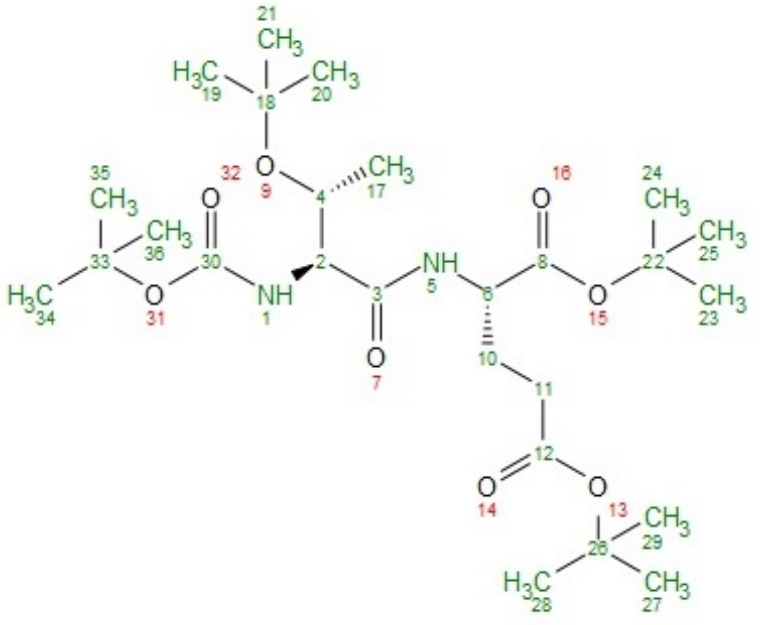


Figure 3.30 | *L*-Thr-*L*-Glu dipeptide with Boc and tert-butyl protection.

Deprotected dipeptides

Dipeptides were deprotected as described in Chapter 2.6.3. Only for the protected CysGlu dipeptide, the amount of CF_3CO_2H was increased to 70 equivalents at 40 °C along with 6 equivalents triisopropylsilane to scavenge the cleaved trityl group.

***L*-Cys-*L*-Glu (3)** ^{1}H NMR (400 MHz, D_2O) δ_H = 4.43 (dd, J = 8.9, 5.3 Hz, 1H, 8), 4.28 (t, J = 5.4 Hz, 1H, 2), 3.12 (t, J = 5.6 Hz, 2H, 4), 2.56 – 2.48 (m, 2H, 11), 2.31 – 2.17 (m, 1H, 10”), 2.10 – 1.97 (m, 1H, 10’); ^{13}C NMR (101 MHz, D_2O) δ_C = 177.2 (C12), 175.3 (C9), 167.9 (C3), 54.1 (C2), 53.0 (C8), 30.2 (C11), 25.7 (C4), 25.0 (C10); ESI-MS, $[M+H]^+$: 251.07 (calculated), 251.10 (observed); HRMS (ESI+) $(M+H)^+$: 251.06962 (calculated for $C_8H_{15}N_2O_5S_1$), 251.06977 (observed). Yield: 25 mg, 0.1 mmol, 11%.

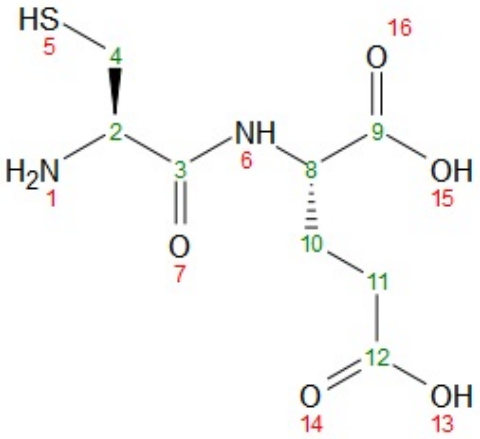


Figure 3.31 | *L*-Cys-*L*-Glu dipeptide (3).

***L*-Ser-*L*-Glu (1)** ^{1}H NMR (400 MHz, D_2O) δ_H = 4.49 (dd, J = 8.8, 5.3 Hz, 1H, 6), 4.18 (dd, J = 5.8, 4.1 Hz, 1H, 2), 4.05 (dd, J = 12.4, 4.1 Hz, 1H, 4'), 3.98 (dd, J = 12.5, 5.8 Hz, 1H, 4"), 2.56 – 2.45 (m, 2H, 11), 2.32 – 2.18 (m, 1H, 10'), 2.12 – 1.97 (m, 1H, 10"); ^{13}C NMR (101 MHz, D_2O) δ_C = 177.1 (C12), 174.9 (C8), 167.7 (C3), 60.1 (C4), 54.5 (C2), 52.6 (C6), 30.0 (C11), 25.9 (C10); ESI-MS, $[M+H]^+$: 235.09 (calculated), 235.15 (observed); HRMS (ESI-) $(M-H)^-$: 233.0779 (calculated for $C_8H_{13}N_2O_6$), 233.0778 (observed). Yield: 39.2 mg, 0.17 mmol, 21%.

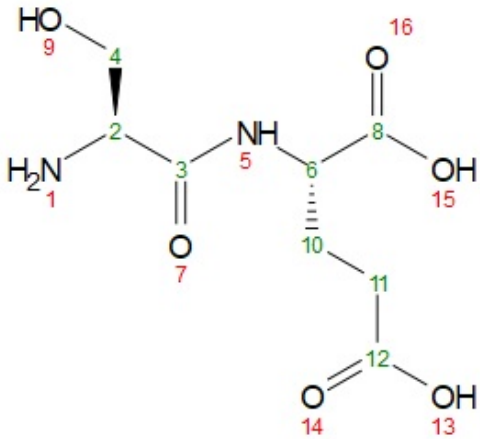


Figure 3.32 | *L*-Ser-*L*-Glu dipeptide (1).

***L*-Thr-*L*-Glu (2)** ^{1}H NMR (400 MHz, D_2O) δ_H = 4.40 (dd, J = 8.7, 5.5 Hz, 1H, 6), 4.09 (p, J = 6.4 Hz, 1H, 4), 3.81 (d, J = 6.4 Hz, 1H, 2), 2.44 (t, J = 7.2, 1.2 Hz, 2H, 11), 2.23 – 2.10 (m, 1H, 10"),

1.96 (ddt, J = 14.2, 8.7, 7.1 Hz, 1H, 10'), 1.27 (d, J = 6.5 Hz, 3H, 17); ^{13}C NMR (101 MHz, D_2O) δ_C = 177.1 (C12), 174.7 (C8), 167.9 (C3), 66.2 (C4), 58.6 (C2), 52.5 (C6), 30.0 (C11), 25.8 (C10), 18.7 (C17); ESI-MS, $[M+H]^+$: 249.11 (calculated), 249.16 (observed); HRMS (ESI+) $(M+H)^+$: 249.1081 (calculated for $C_9H_{17}N_2O_6$), 249.1082 (observed). Yield: 117 mg, 0.47 mmol, 51%.

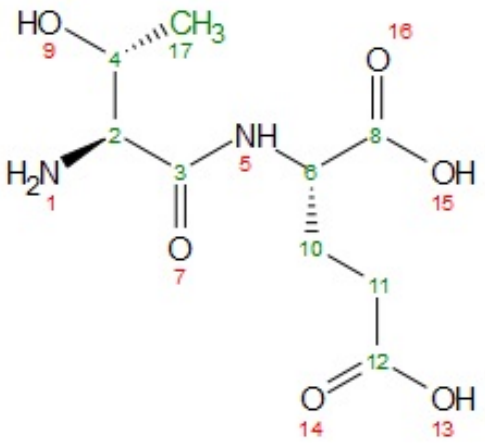


Figure 3.33 | *L*-Thr-*L*-Glu dipeptide (2).

Deprotected dipeptides with HCHO and acetaldehyde

HCHO reactions were performed as described in Chapter 2.6.3. Note that among dipeptides **1**, **2** and **3**, the HCHO adduct **3b** adduct was the only one to be stable to HPLC purification. The acetaldehyde adduct **3c** was prepared the same way as the HCHO adduct (10-fold excess, pH 9, overnight).

***L*-Cys-*L*-Glu HCHO adduct *in situ* (3a)** ^{1}H NMR (600 MHz, D_2O) δ_H = 4.42 (s, 2H, 18), 4.31 (d, J = 3.1 Hz, 1H, 1), 4.29 (d, J = 7.5 Hz, 1H, 16'), 4.23 (d, J = 10.0 Hz, 1H, 16"), 4.18 (m, 1H, 5), 3.34 (dd, J = 11.1, 3.1 Hz, 1H, 14'), 3.19 (dd, J = 11.1, 7.4 Hz, 1H, 14"), 2.24 (m, 2H, 8), 2.11 (m, 1H, 7"), 1.94 (m, 1H, 7'); ^{13}C NMR (101 MHz, D_2O) δ_C = 182.1 (C9), 178.5 (C6), 171.1 (C2), 78.1 (C18), 68.3 (C1), 55.3 (C5), 55.3 (C16), 34.1 (C8), 33.5 (C14), 28.4 (C7); ESI-MS, $[M+H]^+$: **3a** not found, **3b** is present: 263.07 (calculated), 264.10 (observed).

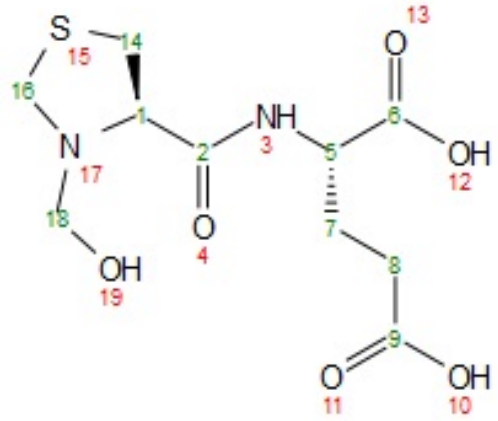


Figure 3.34 | *L*-Cys-*L*-Glu dipeptide (3a) adduct formed by reaction with HCHO *in situ*.

***L*-Cys-*L*-Glu HCHO adduct (3b)** ^{1}H NMR (600 MHz, D_2O) δ_H = 4.66 (t, J = 7.1 Hz, 1H, 1), 4.50 (d, J = 10.2 Hz, 1H, 16”), 4.44 (d, J = 10.8 Hz, 1H, 16’), 4.35 (dd, J = 8.7, 5.2 Hz, 1H, 5), 3.60 (dd, J = 12.1, 7.5 Hz, 1H, 14”), 3.35 (dd, J = 12.1, 6.8 Hz, 1H, 14’), 2.49 (t, J = 7.5 Hz, 2H, 8), 2.26 – 2.18 (m, 1H, 7”), 2.08 – 1.99 (m, 1H, 7’); ^{13}C NMR (151 MHz, D_2O) δ_C = 177.7 (C9), 176.3 (C6), 167.5 (C2), 62.4 (C1), 54.2 (C5), 49.5 (C16), 33.5 (C14), 30.6 (C8), 26.3 (C7); ESI-MS, $[M+H]^+$: 263.07 (calculated), 263.10 (observed); HRMS (ESI+) $(M+H)^+$: 263.06962 (calculated for $C_9H_{15}N_2O_5S_1$), 263.06971 (observed). Yield from 20 mg starting material: 12 mg, 46 µmol, 57%.

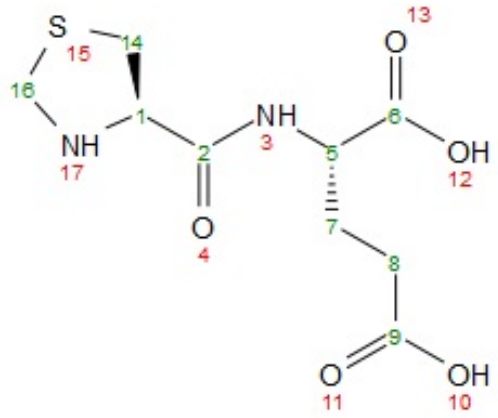


Figure 3.35 | HPLC purified *L*-Cys-*L*-Glu dipeptide adduct (3b) formed by reaction with HCHO.

***L*-Cys-*L*-Glu acetaldehyde (3c)** ^{1}H NMR (600 MHz, D_2O) δ_H = 5.02 (q, J = 6.5 Hz, 1H, 16, A, 4.97 (q, J = 6.5 Hz, 1H, 16, B, 4.76 (d, J = 7.6 Hz, 1H, 1, A, 4.66 (d, J = 7.6 Hz, 1H, 1, B), 4.37

(m, 1H, 5), 3.70 (dd, J = 12.3, 7.6 Hz, 1H, 14", A), 3.65 (dd, J = 12.3, 7.6 Hz, 1H, 14", B), 3.38 (dd, J = 12.3, 7.6 Hz, 1H, 14', A), 3.43 (dd, J = 12.3, 7.6 Hz, 1H, 14', B), 2.51 (t, J = 7.4 Hz, 2H, 8), 2.25 (m, 1H, 7"), 2.04 (m, 1H, 7'), 1.73 (dd, J = 6.5, 1.0 Hz, 3H, 18, A), 1.75 (dd, J = 6.5, 1.0 Hz, 3H, 18, B); ^{13}C NMR (151 MHz, D_2O) δ_C = 177.6 (C9), 176.1 (C6), 167.5 (C2), 62.0 (C1, A), 62.7 (C1, B), 61.8 (C16, A), 61.5 (C16, B) 53.9 (C5), 33.5 (C14, A+B), 30.5 (C8), 26.2 (C7), 18.4 (C18, A), 17.3 (C18, B); ESI-MS, $[M+H]^+$: 277.09 (calculated), 277.106 (observed); HRMS (ESI+) $(M+H)^+$: 277.08527 (calculated for $C_{10}H_{17}N_2O_5S_1$), 277.08541 (observed). Yield from 20 mg starting material: 10 mg, 36 µmol, 45%; diastereoisomers **A**: 56%, **B**: 44%.

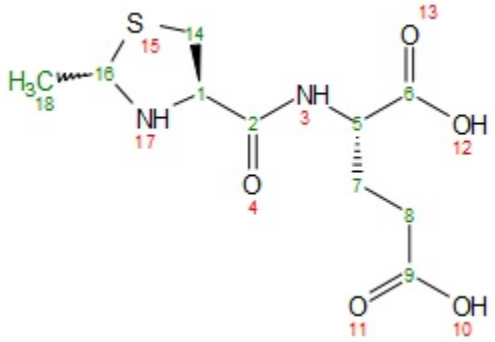


Figure 3.36 | *L*-Cys-*L*-Glu adduct (3c, diastereoisomers A and B) formed by reaction with acetaldehyde.

***L*-Ser-*L*-Glu HCHO adduct *in situ* (1a)** ^{1}H NMR (400 MHz, D_2O) δ_H = 4.74 (m, 1H, 15"), 4.45 (d, J = 3.1 Hz, 2H, 20), 4.43 (d, J = 6.2 Hz, 1H, 15'), 4.17 (d, 1H, 17'), 4.14 (dd, J = 5.1, 3.4 Hz, 1H, 2), 4.04 (dd, J = 8.3, 4.9 Hz, 1H, 19), 3.78 (dd, J = 8.4, 4.9 Hz, 1H, 17"), 2.18 (m, 2H, 5), 2.07 (m, 1H, 4"), 1.92 (m, 1H, 4'); ^{13}C NMR (101 MHz, D_2O) δ_C = 181.9 (C6), 178.4 (C3), 173.8 (C14), 84.2 (C15), 77.1 (C20), 68.3 (C17), 62.6 (C13), 55.0 (C2), 34.0 (C5), 28.4 (C4); ESI-MS, $[M+H]^+$: not found, only starting material (**1**) found.

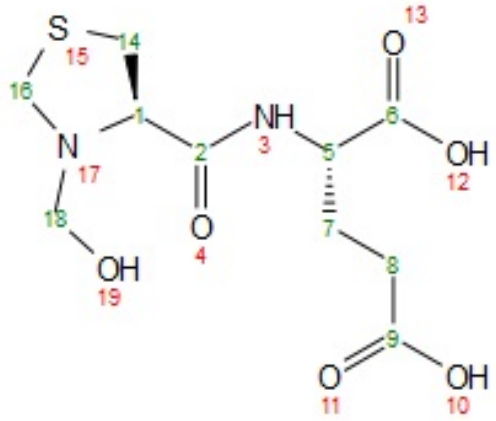


Figure 3.37 | *L*-Ser-*L*-Glu dipeptide adduct (1a) formed by reaction with HCHO *in situ*.

***L*-Thr-*L*-Glu HCHO adduct *in situ* (2a)** ^{1}H NMR (400 MHz, D_2O) δ_H = 4.75 (s, 1H, 15”), 4.50 (d, J = 4.3 Hz, 1H, 15’), 4.48 (d, J = 8.3 Hz, 1H, 11”), 4.40 (d, J = 10.5 Hz, 1H, 11’), 4.18 (dd, J = 8.6, 4.4 Hz, 1H, 2), 4.01 (m, 1H, 17), 3.51 (d, J = 6.6 Hz, 1H, 19), 2.20 (dt, J = 8.3, 6.4 Hz, 2H, 5), 2.10 (m, 1H, 4”), 1.94 (m, 1H, 4’), 1.49 (d, J = 6.1 Hz, 3H, 20); ^{13}C NMR (101 MHz, D_2O) δ_C = 182.0 (C6), 178.4 (C3), 173.5 (C14), 83.6 (C15), 78.2 (C17), 76.9 (C11), 68.4 (C13), 54.9 (C2), 34.0 (C5), 28.4 (C4), 18.7 (C20); ESI-MS, $[M+H]^+$: not found, only starting material (**2**) found.

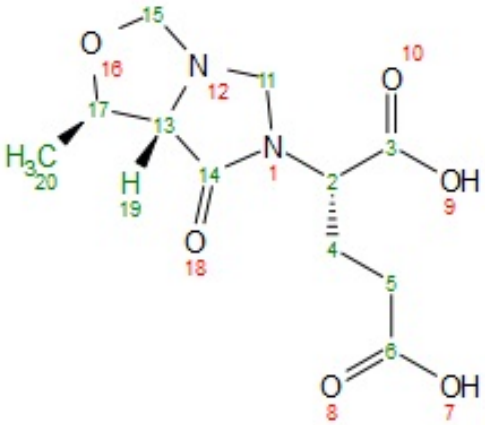


Figure 3.38 | *L*-Thr-*L*-Glu dipeptide adduct (2a) formed by reaction with HCHO *in situ*.

3.6.2 Stability of compounds 1a, 2a and 3b

A 40-fold molar excess GSH (from a 250 mM stock in D_2O) was added to *in situ* generated solutions **1a** or **2a** (10-fold excess HCHO, overnight) or to HPLC purified **3b**. The reactions were monitored by NMR (400 or 600 MHz, see Figure 3.4).

3.6.3 Peptide synthesis

Peptide synthesis was performed as described in Chapter 2.6.2. For this chapter, the following peptide was synthesised: *E. coli* ASNS 7mer (NH_2-CSIFGVF-NH_2).

3.6.4 MALDI MS assays

MALDI MS assays were conducted as described in Chapter 2.6.5 at pH 7.4 potassium phosphate buffer (50 mM). No cross-links or further adducts were observed when titrating HCHO onto NH_2-CSIFGVF-NH_2 (Figures 3.5 a, 3.22), but the peptide was partially present as disulfide-bridged dimer (data not shown).

3.6.5 Cloning and biosynthesis trials of *E. coli* ASNS

Gene amplification of *ASNS B*

The open reading frame (ORF) of *E. coli* ASNS B was from Thermo Fisher (GeneArt) and was codon-optimised for *E. coli* expression. The sequence of the purchased gene is given below.

5'-ATGTGCAGCATTTTTGGCGTGTTCGATATCAAAACCGATGCAGTTGAACTGCGTA
AAAAAGCACTGGAACTGAGCCGTCTGATGCGTCATCGTGGTCCGGATTGGAGCGGTAT
TTATGCAAGCGATAATGCAATTCTGGCACATGAACGTCTGAGCATTGTTGATGTTAAT
GCCGGTGCACAGCCGCTGTATAATCAGCAGAAAACCCATGTTCTGGCAGTGAATGGTG
AAATCTATAATCATCAGGCACTGCGTGCAGAATATGGTGATCGTTATCAGTTTCAGAC
CGGTAGCGATTGTGAAGTTATTCTGGCCCTGTATCAAGAAAAAGGTCCGGAATTTCTG
GATGATCTGCAGGGTATGTTTGCATTTGCACTGTATGATAGCGAGAAAGATGCATATC
TGATTGGTCGTGATCATCTGGGTATTATTCCTCTGTATATGGGCTATGATGAACATGGT

CAGCTGTATGTTGCCAGCGAAATGAAAGCCCTGGTTCCGGTTTGTCGTACCATTAAAG
AATTTCCGGCAGGTAGCTATCTGTGGTCACAGGATGGCGAAATTCGTAGCTATTATCA
TCGTGATTGGTTCGATTATGACGCCGTGAAAGATAATGTGACCGATAAAAATGAACTG
CGCCAGGCACTGGAAGATAGCGTTAAAAGCCATCTGATGAGTGATGTTCCGTATGGTG
TTCTGCTGAGCGGTGGTCTGGATAGCAGCATTATTAGCGCAATTACCAAAAAATACGC
AGCCCGTCGTGTTGAAGATCAAGAACGTAGCGAAGCATGGTGGCCTCAGCTGCATAGC
TTTGCAGTTGGTCTGCCTGGTAGTCCGGATCTGAAAGCAGCACAAGAAGTTGCAAACC
ATCTGGGCACCGTTCATCATGAAATTCATTTTACCGTTCAAGAAGGTCTGGACGCAATT
CGTGATGTGATTTATCACATTGAAACCTATGATGTGACCACCATTCGTGCAAGCACCC
CGATGTATCTGATGTCACGTAAAATCAAAGCCATGGGCATTAAAATGGTTCTGAGTGG
TGAAGGTAGTGATGAAGTTTTTGGTGGCTATCTGTATTTTCACAAAGCACCGAATGCA
AAAGAACTGCATGAAGAAACCGTTCGTAAACTGCTGGCACTGCACATGTATGACTGTG
CACGTGCCAATAAAGCAATGAGCGCATGGGGTGTTGAAGCACGTGTTCCGTTTCTGGA
TAAAAAGTTCCTGGATGTTGCCATGCGTATTAATCCGCAGGATAAAATGTGTGGCAAT
GGCAAAATGGAAAAACATATCCTGCGTGAATGCTTCGAGGCATATCTGCCTGCAAGCG
TTGCATGGCGTCAGAAAGAACAGTTTAGTGATGGTGTTGGTTATAGCTGGATTGATAC
CCTGAAAGAAGTGGCAGCACAGCAGGTTAGCGATCAGCAGCTGGAAACCGCACGTTT
TCGTTTTCCGTATAATACCCCGACCAGCAAAGAAGCGTATCTGTATCGCGAAATTTTTG
AAGAACTGTTTCCGCTGCCGAGCGCAGCAGAATGTGTTCCTGGTGGTCCGAGCGTTGC
CTGTAGCAGCGCAAAAGCAATTGAATGGGATGAAGCCTTCAAAAAAATGGATGATCC
GAGCGGTCGTGCCGTTGGTGTTCATCAGAGCGCCTATAAATGA-3’

The *ASNS B* gene was amplified by PCR (Tables 3.2, 3.3), introducing an NdeI restriction site at the 5’ end and an XhoI restriction site at the 3’ end of the gene with an additional CCCGGG adjacent to the restriction sites to increase efficiency of the restriction enzymes (forward primer 5’-CCCGGGCATATGTGCAGCATTTT-3’, reverse primer 5’-CCCGGGCTCGAGTTTATAGGCGCTCTG-3’, Merck, Table 3.1). The black part of the primers in Table 3.1 is part of the ORF of *E. coli* ASNS B. After restriction digest with NdeI and XhoI (overnight at 37 °C; New England Biolabs) of both the

gene insert and pET22b, the digestion mixtures were purified (GeneJet Gel Extraction Kit, Thermo Fisher Scientific). T4 DNA ligase was used to ligate the fragments (2 days at ambient temperature; New England Biolabs). Upon transformation of the ligation mixture into XL10-Gold cells (NEB, C3040H), the plasmid DNA of individual clones was isolated (Thermo Scientific GeneJet Plasmid Miniprep Kit) and the DNA sequence verified by sequencing (Eurofins).

Table 3.1: Primers used for PCR.

Primers
fw 5'-CCCGGGCATATGCATATGTGCAGCATTTT-3'
rv 5'-CCCGGGCTCGAGTTTATAGGCGCTCTG-3'

Table 3.2: PCR for Plasmid DNA or Gene Blocks.

Components	**Volume [µl]**	**Final Concentration**
DNA template	50 ng	
5x Phusion HF Buffer	10	
Cloned Pfu DNA Polymerase	1	
dNTPs	1 µl from 10 mM (each) stock	200 µM
Forward primer	1 µl from 10 µM stock	200 nM
Reverse primer	1 µl from 10 µM stock	200 nM
H_2O	Add to 50 µl	
Total Volume	50	

Table 3.3: Plasmid DNA/Gene Block PCR Settings.

	Temperature	**Time**	**Cycles**
Initial denaturation	95 °C	2 min	
Denaturation	95 °C	30 s	
Annealing	5 °C below Tm of primers	30 s	
Elongation	72 °C	1 min per kb	30
Final elongation	72 °C	10 min	

Restriction digest

Restriction digests were conducted overnight at 37 °C (Table 3.4). Digested DNA (gene insert and pET22b) was seperated on 1% agarose containing SYBR Safe DNA Gel Stain (invitrogen, S33102) and purified with GeneJet Gel Extraction Kit (Thermo Fisher Scientific).

Table 3.4: Restriction Digest.

Components	**Volume [µl]**
DNA	2 µg
10x CutSmart Buffer	5
NdeI	2
XhoI	2
H_2O	Add to 50 µl
Total Volume	50

DNA ligation

DNA ligation was carried out as depicted in Table 3.5.

Table 3.5: Ligation.

Components	**Volume [µl]**
Vector	20-50 ng
Insert	molar ratio 1:3 (vector:insert)
10x ligase buffer	2
T4 DNA Ligase	1
H_2O	Add to 20 µl
Total Volume	20

The ligation mixture was incubated for 2 days at ambient temperature.

Transformation

For generating bacterial strains containing the desired plasmid, 35 µl of XL10-Gold cells (NEB, C3040H) were thawed on ice and supplemented with 5 µl of the corresponding ligation mixture in

a pre-cooled Falcon tube. Transformation reactions were incubated for 20 min on ice. Afterwards the mixture was incubated at 42 °C for 30 s and incubated on ice for 2 min. Then, 230 µl of 2TY were added to the suspension and the cells were incubated at 37 °C (180 rpm, 60 min). Finally, the cells were spread on a prewarmed agar plate containing the appropriate antibiotic (50 µg/ml) and incubated overnight at 37 °C or over three days at ambient temperature.

Plasmid purification and verification

Upon transformation of XL10-Gold cells, single colonies were picked from the agar plates and used to inoculate 5 ml 2TY containing 50 µg/ml antibiotic. The cells were grown overnight at 37 °C and 180 rpm. Cells were centrifuged at 3000 g for 3 min. After removing the supernatant, the DNA was purified using the Thermo Scientific GeneJet Plasmid Miniprep Kit following the manufacturer's instructions. The correct sequence of the amplified genes and generated constructs was confirmed by colony PCR (see Tables 3.6 and 3.7) and DNA sequencing (eurofins).

Colony PCR

10 µl of bacterial overnight culture were supplemented with 10 µl H_2O, boiled for 10 min at 95 °C and pop-spinned. 5 µl of supernatant were used to conduct the PCR.

Table 3.6: Colony PCR.

Components	**Volume [µl]**	**Final Concentration**
Supernatant	5	
2x BioMixRed (BioLine)	10	
Forward primer	1 µl from 10 mM stock	500 nM
Reverse primer	1 µl from 10 mM stock	500 nM
H_2O	Add to 20 µl	
Total Volume	20	

Table 3.7: Colony PCR Settings.

	Temperature	**Time**	**Cycles**
Initial denaturation	95 °C	5 min	
Denaturation	95 °C	30 s	
Annealing	5 °C below Tm of primers	30 s	
Elongation	72 °C	2.5 min	20
Final elongation	72 °C	5 min	

Biosynthesis trials

The expression plasmid was transformed into the host strain (BL21(DE3) competent *E. coli* (NEB, C2527H)) and plated on TB-agar containing 50 µg/ml of the appropriate antibiotic. Several colonies were combined to inoculate 2TY medium (+50 µg/ml antibiotic), which was used to make a glycerol stock after 5 h of growth (800 µl of cell suspension and 200 µl glycerol, subsequently frozen on dry ice). The glycerol stock was used to inoculate a starter culture overnight at 37 °C and 180 rpm (+50 µg/ml antibiotic). The next day, the main culture (containing 50 µg/ml antibiotic in 2TY) was inoculated with 1:100 (v/v) of starter culture and grown until OD600 of 0.6 at 37 °C and 180 rpm. IPTG was added to the appropriate concentration and growth continued at the expression temperature at 180 rpm overnight. The cells were collected by centrifugation (3000 g, 10 min, 4 °C), the supernatant removed and stored at -80 °C. As soon as the cells were deep frozen, 2-4 ml HisTrap A buffer (20 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 7.5) containing a tiny amount of DNase (Sigma Aldrich) were used to resuspend the pellet by vortexing whilst keeping everything on ice during the entire procedure. The cells were disrupted by sonication (10 s on, 10 s off, 2 min on in total, 40% amplitude). 1 ml of lysate was then centrifuged (10 min, 4 °C, 18000 g). The supernatant was transferred onto Profinity IMAC resins (approximately 150 µl for 1 ml lysate) for His-tagged proteins (Bio-Rad) which were washed previously three times with 600 µl HisTrap A buffer (2 min at 2650 g; last time of wash: equal amounts of HisTrap A [v/v] were used to resuspend beads). The protein-bead slurry was rotated for 30 min at 25 rpm in 4 °C (Stuart Tube Rotator with Free Microcentrifuge Tube Holder,

120 VAC). The bead-bound proteins were washed three times with 200 µl HisTrap A buffer (0.5 g, 4 °C, 2 min). Proteins were eluted with 200 µl HisTrap B buffer (20 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.5) and rotated (5 min, 25 rpm, 4 °C). Upon centrifugation (0.5 g, 4 °C, 2 min), the supernatant was used to analyse protein content via SDS-PAGE (Figure 3.39). For large scale biosynthesis, it was decided to use 18 °C at 0.01 mM.

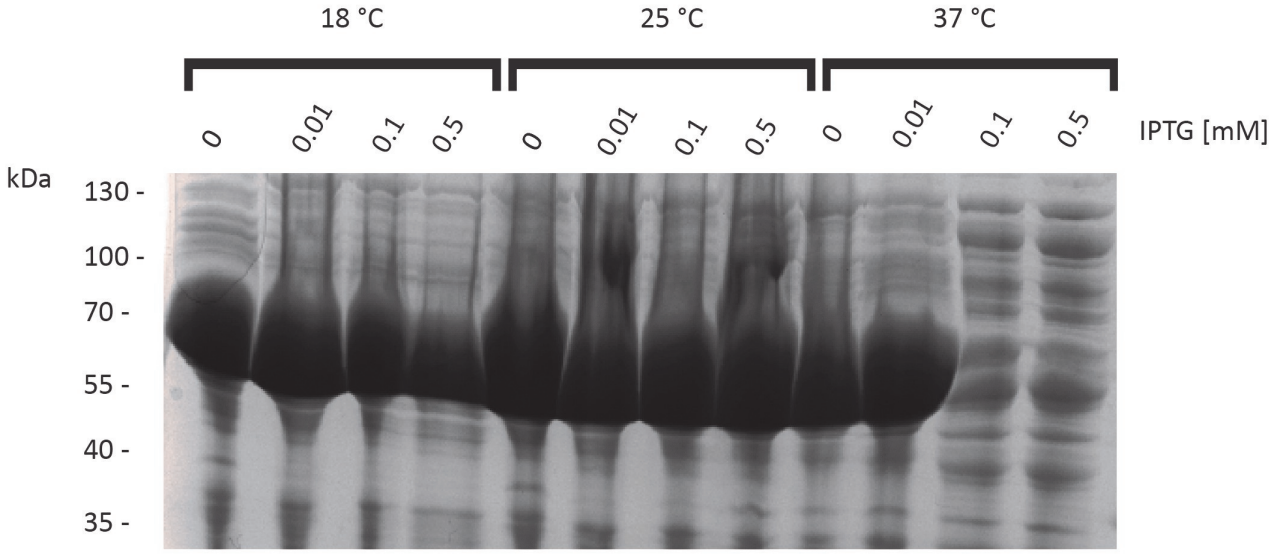


Figure 3.39 | SDS-PAGE analysis of biosynthesis trials for *E. coli* ASNS B. Trials were done at 18 °C, 25 °C and 37 °C overnight at 0, 0.01, 0.1 and 0.5 mM IPTG.

3.6.6 Biosynthesis and purification of *E. coli* ASNS B

A glycerol stock of BL21(DE3) competent *E. coli* cells (NEB, C2527H) transformed with the expression plasmid pET22b ASNS B was used to inoculate a starter culture that was grown overnight at 37 °C and 180 rpm (+50 µg/ml ampicillin). The next day, 6 l of main culture (containing 50 µg/ml ampicillin in 2TY) were inoculated with 1:100 (v/v) of starter culture and grown until OD600 of 0.85 at 37 °C and 150 rpm. Protein biosynthesis was induced with 0.01 mM IPTG and continued at 18 °C and 150 rpm overnight. Three drops of liquid polypropylene were used to remove foam to increase aeration. Cells were pelleted (9820 g, 8 min, 4 °C) to yield 70 g pellet which was frozen at -80 °C. The pellet (17.5 g) was resuspended in 70 ml anion exchange buffer A (Table 3.8, +one protease inhibitor tablet [Sigma FAST EDTA free], +DNase I [Sigma]) at 4 °C. Cells were disrupted by 10 min sonication (Sonics Vibra-Cell) at 60% amplitude (10 s on, 10 s off, 5 min total sonication time). Cellular debris was

then removed by centrifugation (63988 g, 30 min, 4 °C). The cell lysate was filtered using 0.45 µm filters (Filtropur S, Sarstedt). The filtrate was loaded onto a Q-Sepharose (8ml) column employing the following salt gradient to purify *E. coli* ASNS B: The sample was applied at 2 ml/min, the column was then washed with anion exchange buffer A for 5 column volumes (CV); 15% anion exchange buffer B (Table 3.8) for 2 CV was followed by a linear gradient from 15% to 100% over 6 CV, and 6 CV of 100% buffer B. Fractions 16-55 were then pooled and concentrated to 2 ml (Amicon Ultra-15, Centrifugal Filters, Ultracel 50 k, 4 °C, 3100 g; Figure 3.40).

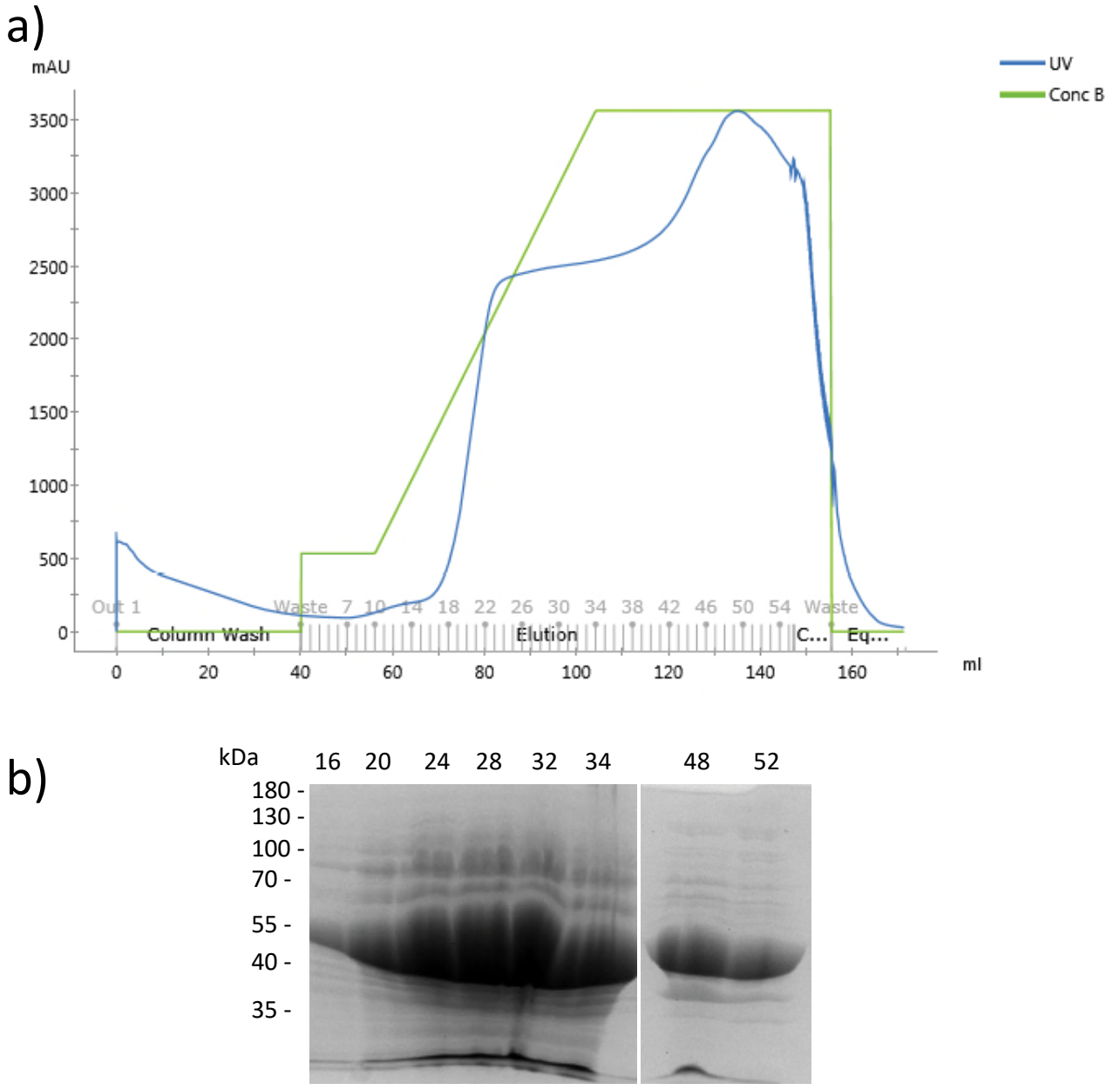


Figure 3.40 | Anion Exchange elution profile of *E. coli* ASNS B. a) Elution profile showing UV trace and concentration of anion exchange buffer B. **b)** SDS-PAGE analysis of indicated fractions. Fractions 21-46 were pooled and concentrated.

Subsequently, the concentrated protein was applied to a size exclusion column (S200 300 ml) using a flowrate of 0.5 ml/min. Proteins were eluted using 1.1 CV size exclusion buffer, and fractions 21-46 were pooled and concentrated (Amicon Ultra-15, Centrifugal Filters, Ultracel 50 k, 4 °C, 3100 g) to yield 1.2 ml of 91.1 mg/ml ASNS B (18.2 mg from 1 l of culture) (Figure 3.41). The concentrated protein was rapidly frozen and stored at -80 °C.

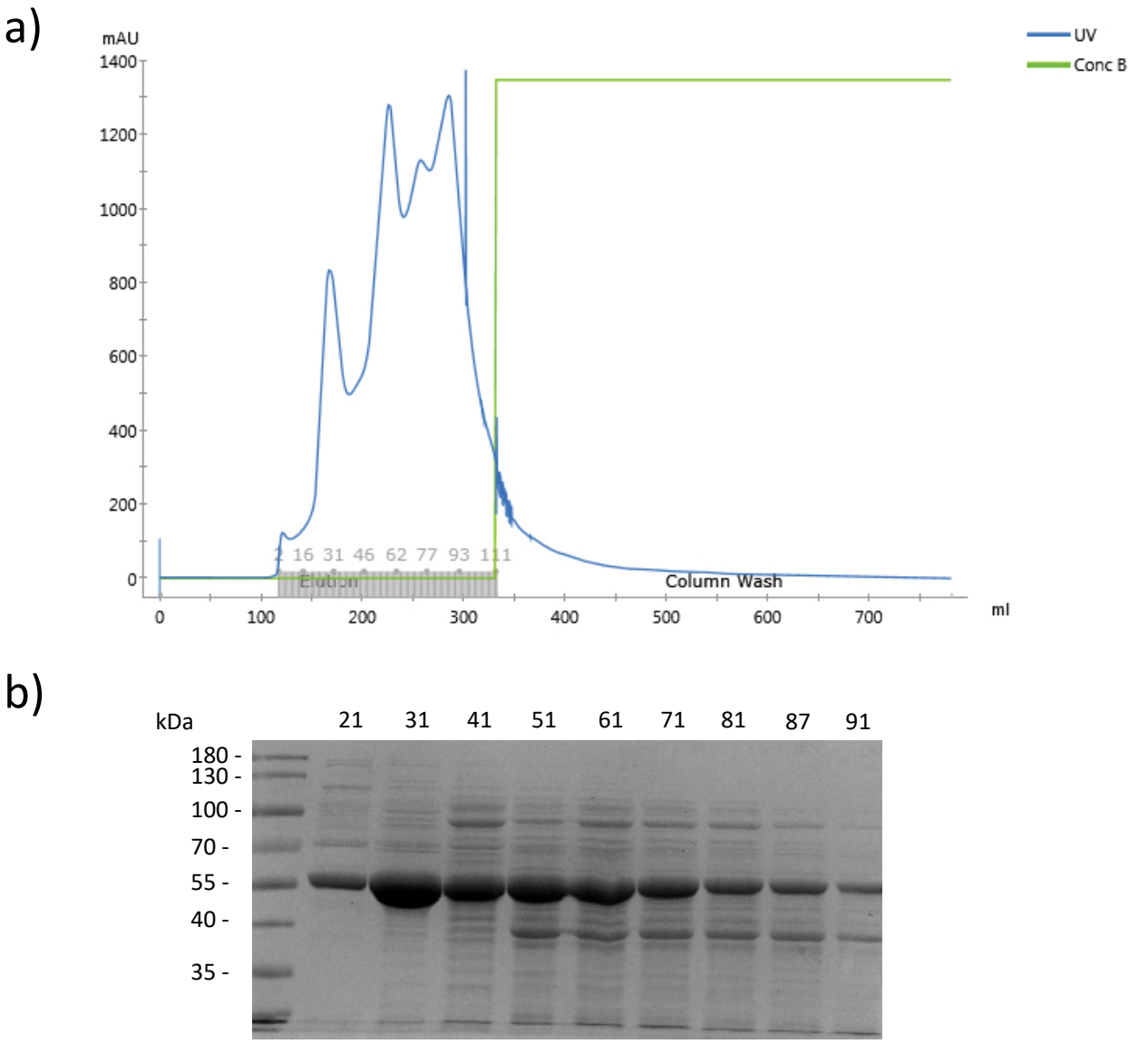


Figure 3.41 | Size Exclusion Chromatography of *E. coli* ASNS B. a) Elution profile showing UV trace. **b)** SDS-PAGE analysis of indicated fractions.

SPE-MS analysis suggests a mass of 63594.09 Da (calculated 63594.67 Da), which corresponds to the mass of the protein being cleaved off its N-terminal methionine, but including the non-ORF C-terminal tail sequence LEHHHHHH (Figure 3.42). The isoelectric point of ASNS B (without N-terminal methionine), but including the non-ORF C-terminal tail sequence LEHHHHHH is pI=5.72.

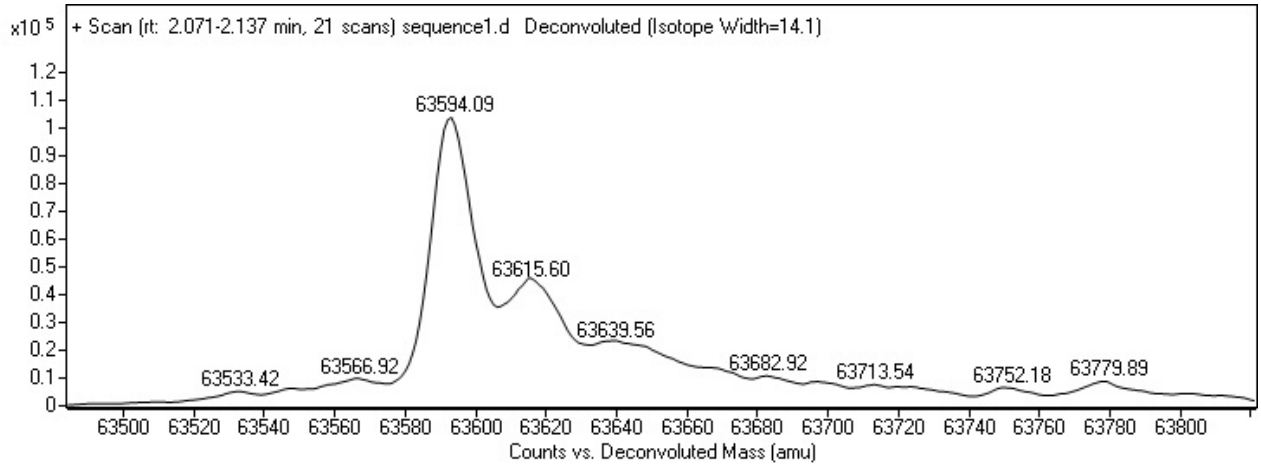


Figure 3.42 | SPE-MS data obtained from purified *E. coli* ASNS B. The calculated mass 63594.7 Da matches the observed experimental mass of 63594.1 Da.

Table 3.8: Buffers for purification of *E. coli* ASNS B.

Buffer	**Components**	**Final Concentration**
Anion Exchange A (filtered, degassed) pH 6.5	BisTris NaCl	50 mM 0 mM
Anion Exchange B (filtered, degassed) pH 6.5	BisTris NaCl	50 mM 1 M
Size Exclusion buffer (filtered, degassed) pH 7.5	BisTris NaCl	50 mM 150 mM

3.6.7 ASNS B NMR assays

Table 3.9: ASNS B assay components, exemplified for a 100-fold excess carbonyl compound.

Component	Stock	Volume [µl]	Final Conc.
ASNS B	15.7 µM (1 mg/ml)	10	314 nM
carbonyl compound	1 mM	15.7	31.4 µM
ATP	50 mM	10	1 mM
Gln (optional)	50 mM	10	1 mM
Asp	50 mM	10	1 mM
$MgCl_2$	100 mM	50	10 mM
NH_4Cl (optional)	50 mM	50	5 mM
50 mM potassium phospate pD 8		add to 500 µl	

ASNS B was diluted from the frozen stock solution (91.1 mg/ml) with size exclusion chromatography (SEC) buffer (50 mM BisTris pH 7.5, 150 mM NaCl) to 1 mg/ml (15.7 µM). If ASNS B had been preincubated with carbonyl compounds, 15.7 µl the indicated molar excess of carbonyl compound was added to 10 µl of 1 mg/ml ASNS B and incubated at 37 °C for 20 min (45 min for the detailed kinetic inhibition studies, with the NMR probe temperature at 37 °C during the kinetic study, Figure 3.8 b-d). The preincubated enzyme/carbonyl compound mixture was then added to the assay mixture containing ATP, Gln (Gln was not added in the "exogenous NH_3" experiment, e.g. Figure 3.5 b, panel 3), Asp, $MgCl_2$, and NH_4Cl (optional for "exogenous NH_3" and "rescue" experiment, Figure 3.5 b, panel 3+9+10). Assay mixtures were incubated in an Eppendorf tube at 37 °C for 4 h, quenched with 10 µl CF_3CO_2H, and frozen until analysis by ^1H NMR (700 MHz). All other component stock solutions (ATP, Gln, Asp, NH_4Cl) were prepared in 50 mM potassium phospate buffer in D_2O (pH 7.6 corresponding to pD 8); only $MgCl_2$ was prepared in D_2O due to solubility issues in potassium phosphate buffer; the pH of all these components was then adjusted to pH 7.6 (pD 8) using NaOD and DCl. Carbonyl compounds were prepared in MQ water without pH adjustment (glyoxylic acid was prepared in 50 mM potassium phosphate, pD 8 for the kinetic inhibition studies, Figure 3.8 d); the stock solutions of the carbonyl compounds were prepared in such that 15.7 µl of carbonyl compound solution was added

to the ASNS B solution, independent of the excess carbonyl compound used. For controls without ASNS B, 10 µl SEC buffer was preincubated with 15.7 µl MQ water; to assess enzyme activity without carbonyl compounds present, the enzyme was preincubated with 15.7 µl MQ water.

3.6.8 ASNS B Protein observed MS

For protein observed MS, ASNS B was diluted with SEC buffer (Table 3.8) to 1 mg/ml, analogously to the NMR assay described above. 30 µl of 1 mg/ml enzyme stock was then reacted with 47.1 µl of carbonyl compounds (0.1, 1 or 10 mM stock in MQ water) at 37 °C for 20 min, which was then subjected to analysis by SPE-MS (Figures 3.6, 3.7).

3.6.9 Tissue culture

Tissue culture experiments were performed as described in Chapter 2. Dialysed FBS was from Thermo Fisher (A3382001). All downstream analyses, including SDS-PAGE (Chapter 2.6.8), western blot (Chapter 2.6.9), RNA-Seq (Chapter 2.6.10), immunoprecipitation (Chapter 2.6.16), trypsin digestion (Chapter 2.6.17), and MS/MS analyses (Chapter 2.6.20) are equal to procedures described in Chapter 2. Cells used in this chapter were from American Type Culture Collection (ATCC; LN-18, ref. CRL-2610; HEK293T, ref. CRL-3216). The authentication from ATCC was relied upon, none of the cell lines were further authenticated. Primary antibodies used in this chapter were GAPDH Loading Control Monoclonal Antibody (GA1R), abcam Asparagine Synthetase Antibody (ab226413; used for immunoprecipitation only), Origene ASNS antibody (TA503444; used for mouse samples only), Cell Signaling Asparagine Synthetase Antibody (ref. no. 20843). Secondary antibodies used were anti-mouse IgG, HRP-linked Antibody (Cell Signaling, reference 7076), anti-Rabbit IgG (H+L) HRP Conjugate (Promega reference W4011).

3.6.10 Metabolomics

For cell harvesting, the medium was removed by decanting, cells were washed twice with PBS (Sigma, D8537) (residual PBS was wiped off), snap frozen with N_2 (l) and 200 µl of 80% MeOH 20% MQ

water were used to scrape cells. The cells were then centrifuged at (18000 g, 25 min, 4 °C). For normalizing each experimental condition, dsDNA concentration of the supernatant was measured using Nanodrop One (Thermo Scientific). The supernatant was passed through centrifugal filters (pre-washed with MQ water; Amicon Ultra Centrifugal Filters, 10 kDa Ultracel, 0.5 ml; 18000 g for 25 min at 4 °C). The filtrate of each experimental condition was diluted with 80% MeOH 20% MQ water in a way that the dsDNA concentration of all samples was the same (based on the lowest dsDNA concentration obtained). Next, the AccQ-Tag Ultra Derivatization Kit (Waters, SKU: 186003836) was used to derivatize amino acids following the manufacturer's recommendations. All metabolomics samples (Figures 3.9, 3.10, 3.12) were prepared by myself, but ran and analysed by Dr John Walsby-Tickle (see individual metabolomics Figures for details on contributions).

4

Metampicillin

His audience was twenty-five [...] scientists who looked so disheveled, dressed in such bad, worn, pathetic clothing you swore they were begging on the streets. [...] So what if they looked homeless. They were [...] free. It wasn't a 'career,' it wasn't ball gowns and glamour, it was pure, pure science, it was What's the [...] problem? *How do we solve the problem?*

Chandler Burr - The Emperor of Scent [295]

Contents

4.1 Abstract **212**
4.2 Introduction **212**
4.3 Results **213**
4.4 Discussion **230**
4.5 Methods **231**

The work described in this chapter was published in [296] and was carried out in close collaboration with Raphael Reinbold. All the figures in this thesis are either identical or slightly modified to the published figures. The text was modified wherever appropriate.

4.1 Abstract

Metampicillin is a penicillin antibiotic that is prepared by the reaction of ampicillin with formaldehyde (HCHO). Despite being used for treatment of infections in animals and humans, the structure of metampicillin has been contentious. NMR studies reveal that metampicillin contains a HCHO-derived cyclic aminal. NMR time-course experiments with excess HCHO reveal formation of a product with an additional exocyclic hydroxymethyl group on the cyclic aminal. The exocyclic hydroxymethyl group is readily removed on reaction with the HCHO scavenger 1,3-cyclohexanedione, whereas the cyclic aminal methylene is more stable.

4.2 Introduction

Metampicillin is a β-lactam antibiotic that is approved for use in veterinary[297] and human medicine.[298] It is synthesised by the reaction of HCHO with ampicillin (**1**).[299] Metampicillin is considered a prodrug form of **1**, which is released after acid-catalysed fragmentation in the stomach. Whilst in most infections the clinical efficacy of metampicillin is comparable to that of **1**, it is reported to accumulate in bile, leading to greater activity against biliary infections.[212, 213] Further, the stability of metampicillin in serum is greater than in aqueous acidic conditions.[300] Upon parenteral administration of metampicillin, biliary concentations are 300 times higher than concentrations of **1** injected under the same conditions, implying that the liver favourably secretes metampicillin over **1**.[212]

Given the clinical use of metampicillin, it is remarkable that its structure is contentious. The presence of the core ampicillin-derived scaffold of **1** is accepted; however, the structure of its HCHO-derived component(s) is unequivocally assigned (Figure 4.1).[212, 297, 299, 300] Both imine (**4**)[301] and hemiaminal (**5**)[302] structures for metampicillin have been proposed, and there are other possibilities (**2, 3**; Figure 4.1). Studies on the stability of the HCHO-derived component(s) of metampicillin are also lacking, though ampicillin is reported to form 3-phenyl-6-ethyl-pyrazin-2-one with HCHO under acidic conditions (**6**, Figure 4.1).[245]

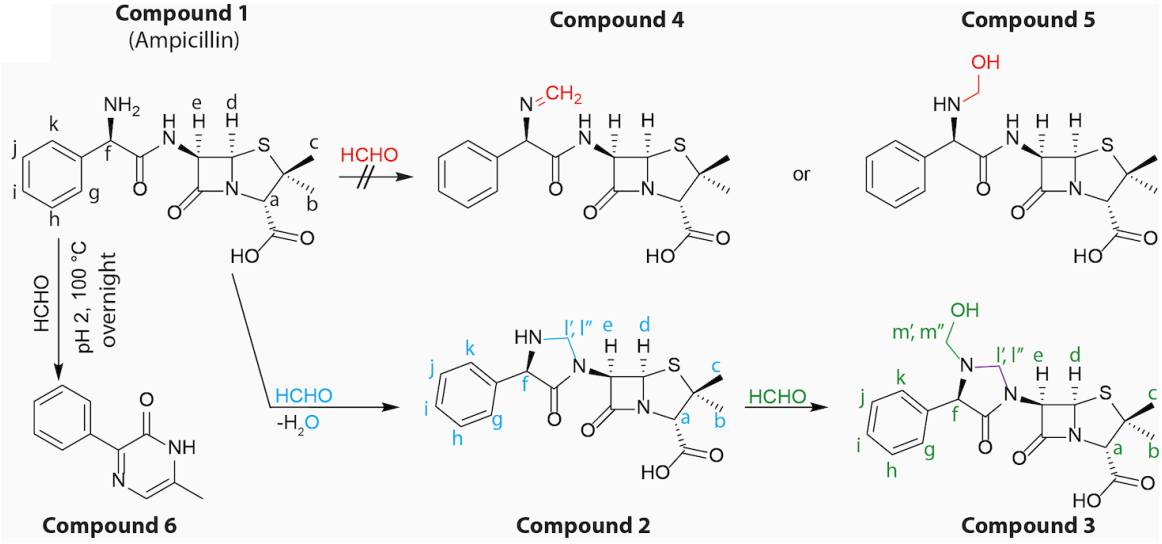


Figure 4.1 | Reactions of ampicillin with HCHO. Reported structures for metampicillin (**4**[301] and **5**[302]) and the revised metampicillin structure (**2**). **2** is observed upon HPLC purification/lyophilisation, whereas **3** is observed in solution with excess HCHO, and could not be isolated. **6** is formed by reaction of **1** with HCHO under prolonged acidic conditions (100 °C) as reported.[245] These results were obtained in close collaboration with Raphael Reinbold.[245]

This chapter describes NMR studies on the stability and structure of metampicillin in water. The presence of the core scaffold of **1** in metampicillin is confirmed and co-existing cyclic aminal and hemiaminal products in solution are identified. The results reveal that the cyclic aminal methylene (**2**) is more stable than the hemiaminal (**3**), indicating **2** represents the major component of solid metampicillin (**2**).

4.3 Results

First, the structure of metampicillin after reaction from ampicillin with HCHO and purification was determined. Metampicillin was prepared following a reported procedure, reacting **1** with an excess of HCHO in water.[302] Hence, **1** was dissolved in water and reacted with an excess of HCHO (10-fold) for 2 h at ambient temperature. Following reverse-phase HPLC of the reaction mixture (Figure 4.17), the major product was lyophilised to yield a white solid, which was analysed by ^{1}H NMR (Figure 4.2).

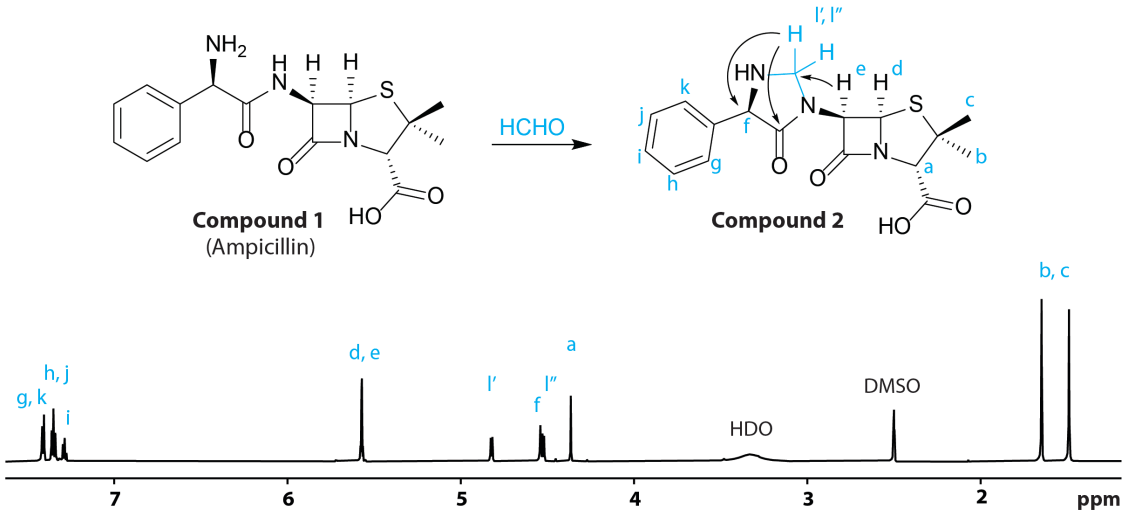


Figure 4.2 | ^{1}H NMR (700 MHz, D_6-DMSO) spectrum showing HPLC purified metampicillin (2). Metampicillin (**2**) prepared from ampicillin (**1**). Arrows show ^{1}H-^{13}C-HMBC correlations of **2**, indicating a cyclic aminal formation (see Figure 4.4). These results were obtained in close collaboration with Raphael Reinbold.

The β-lactam ring in metampicillin was confirmed by NMR analysis of the solid in D_6-DMSO (Figure 4.2, C<u>H</u>S and COC<u>H</u>NCO, δ_H 5.62–5.53 ppm; see Chapter 4.5.2 for full assignment and numbering). The described penicillin-derived and the phenyl group ^{1}H resonances at δ_H 7.44-7.39 ppm, δ_H 7.38–7.33 ppm, and δ_H 7.31–7.26 (Figure 4.3, **2**) have similar chemical shifts compared to **1** (Figure 4.3, **1**), confirming the presence of the ampicillin-derived scaffold. Using 2D ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC, the penicillin-derived hydrogen atoms at δ_H 4.36, δ_H 1.65, and δ_H 1.49 ppm were identified in **2** (Figure 4.4). Two new resonances at δ_H 4.82 and δ_H 4.52 ppm in **2** were assigned to the HCHO-derived hydrogen atoms (Figure 4.4), as similar resonances in the spectrum of **1** are not present.

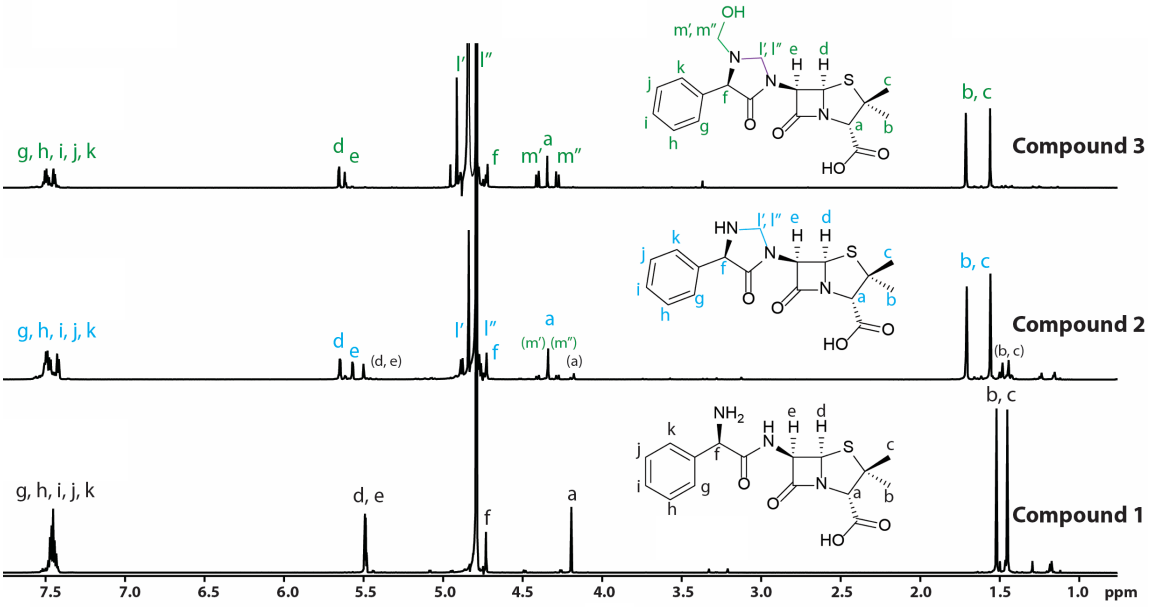


Figure 4.3 | Formation of cyclic aminal 2 and hemiaminal 3 from 1. ^{1}H (700 MHz) spectra of **1**, **2** and **3** in D_2O. Adding a twofold excess of HCHO to **1** leads to formation of **2** (second panel); trace amounts of **1** and **3** are present under these conditions (in brackets). Addition of an excess of HCHO results in further reaction to **3** (upper panel). 2D NMR was used to assign structures **2** and **3** (Figures 4.4, 4.6, 4.7). These results were obtained in close collaboration with Raphael Reinbold.

Using 2D ^{1}H-^{13}C-HSQC (Figure 4.4), these new resonances at δ_H 4.82 and δ_H 4.52 ppm were assigned to be attached to the carbon atom at δ_C 63.0 ppm, indicating formation of a cyclic aminal methylene group, Schiff base or hemiaminal. To distinguish between formation of a Schiff base (**4**), hemiaminal (**5**) or a cyclic aminal methylene bridge (**2**), ^{1}H-^{13}C-HMBC analysis were conducted. Correlations at δ_H 4.82 ppm with δ_C 60.9 ppm, and of δ_H 5.57 ppm with δ_C 63.0 ppm (Figure 4.4) indicate formation of a cyclic aminal methylene bridge (Figures 4.1, 4.3, **2**). In summary, these results assign the structure of metampicillin as containing a cyclic aminal, a structure which has not been previously proposed for metampicillin; notably, no other HCHO-derived compound was detectable other than **2** upon HPLC purification and lyophilisation (Figures 4.1, 4.2, 4.4).

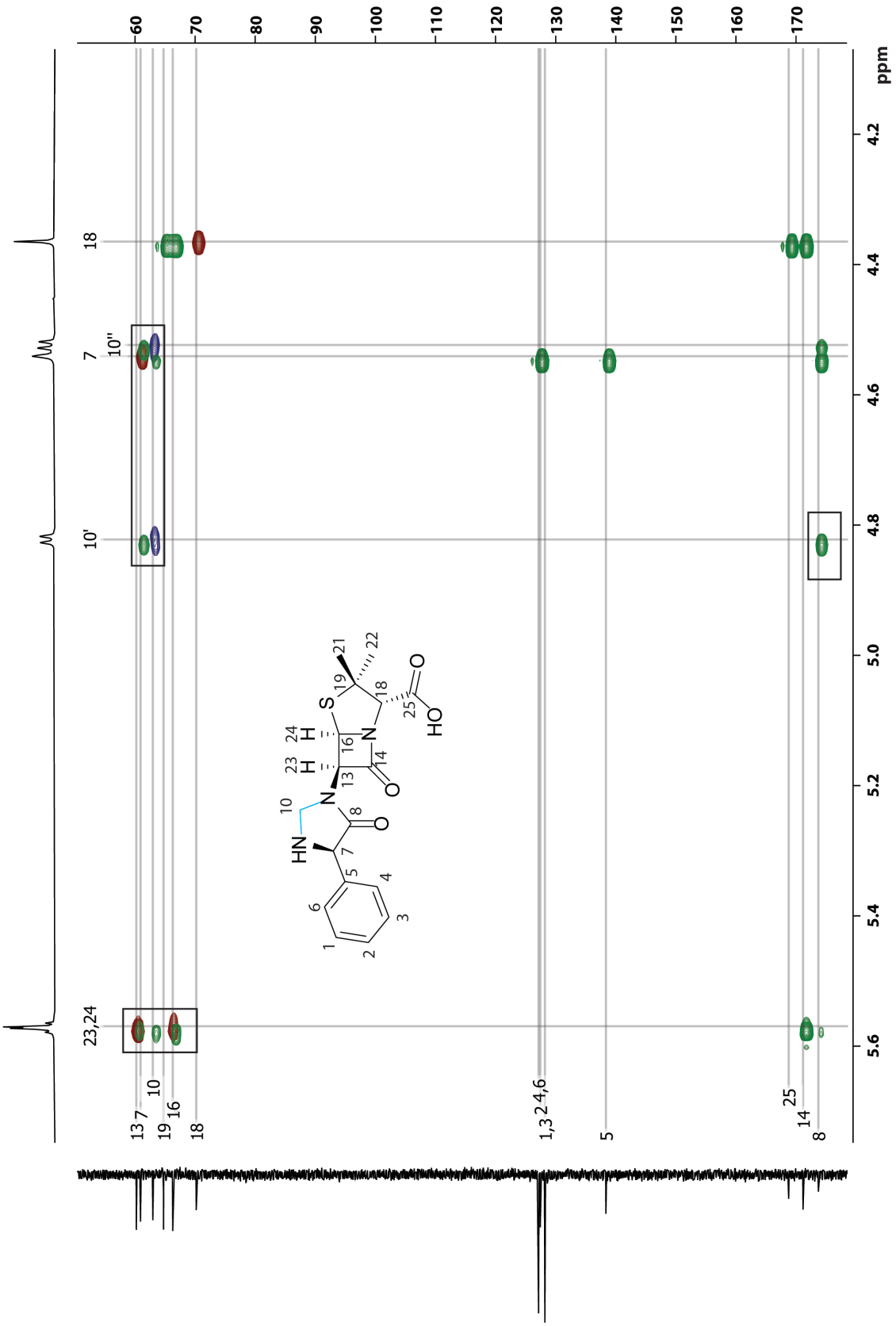


Figure 4.4 | ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC NMR analysis of HPLC-purified 2 in D_6-DMSO. The y-axis shows the ^{13}C spectrum and the x-axis the ^{1}H spectrum (700 MHz). Interactions marked with boxes are crucial for structure elucidation of **2**. These results were obtained in close collaboration with Raphael Reinbold.

Next, NMR time-course experiments were carried out on the reaction of ampicillin (**1**) in water. A 5-fold excess of HCHO was added to **1** and monitored in D_2O over 11 h at ambient temperature (Figure 4.5). Formation of the cyclic aminal methylene bridge at δ_H 4.88 and δ_H 4.77 ppm was detected after the first time-point (0.5 h) and signal intensities of these peaks reached a maximum after 3 h. Further, the formation of novel ^1H resonances (δ_H 4.41 and δ_H 4.28 ppm), provisionally assigned to a hemiaminal, was observed, the signal of which increased in intensity, too. Interestingly, the cyclic aminal methylene bridge of **2** and **3** (marked in blue and purple, respectively) differ slightly in their chemical shifts (Figure 4.5, 4.10); the cyclic aminal methylene signals in **2** at δ_H 4.88 and 4.77 ppm (blue) correspond to the cyclic aminal methylene signal in **3** at δ_H 4.89 and 4.73 ppm (purple, Figure 4.5).

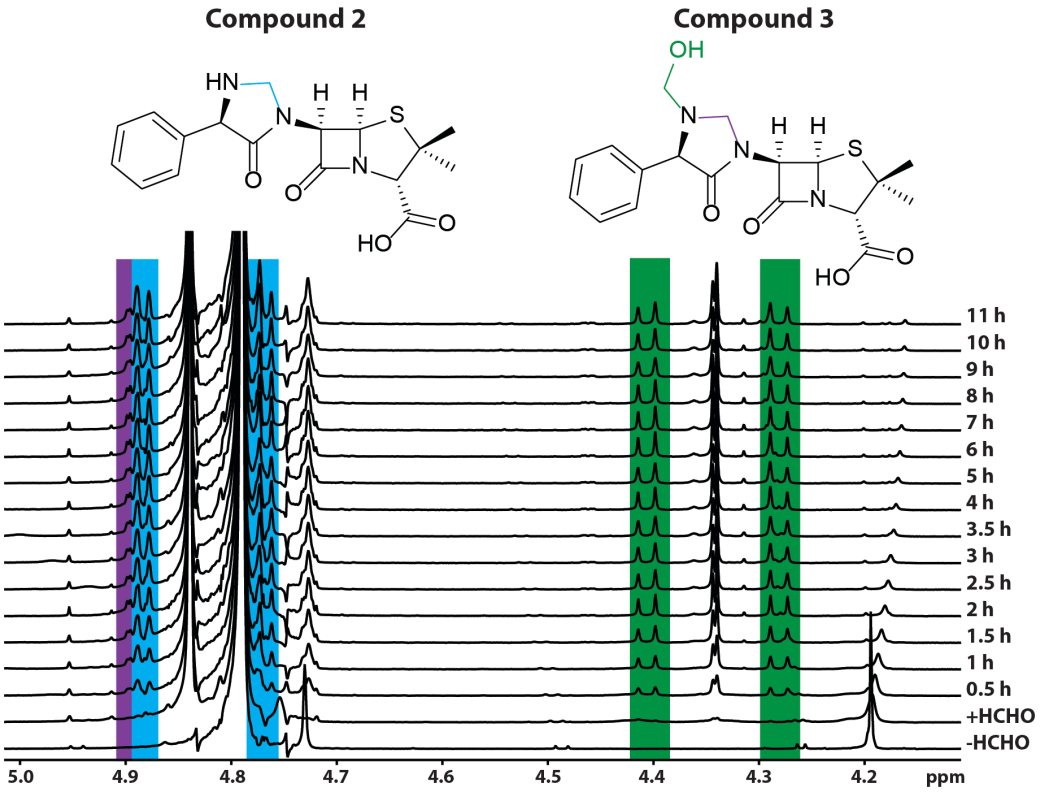


Figure 4.5 | Reaction progress of 1 upon addition of a 5-fold excess HCHO. ^1H spectrum (700 MHz) showing the formation of a cyclic aminal methylene (**2**, blue) and a novel species with an additional exocyclic hemiaminal group (**3**, green) in D_2O. The cyclic aminal methylene of **3** is shown in purple. These results were obtained in close collaboration with Raphael Reinbold.

The tentatively assigned species **3** was then characterised by 2D NMR employing 10-fold excess HCHO reaction conditions which resulted in formation of **3** as the major product overnight. The ^{1}H-^{1}H-COSY spectrum confirms that the two HCHO-derived moieties on **3** are present as methylene groups (Figure 4.6).

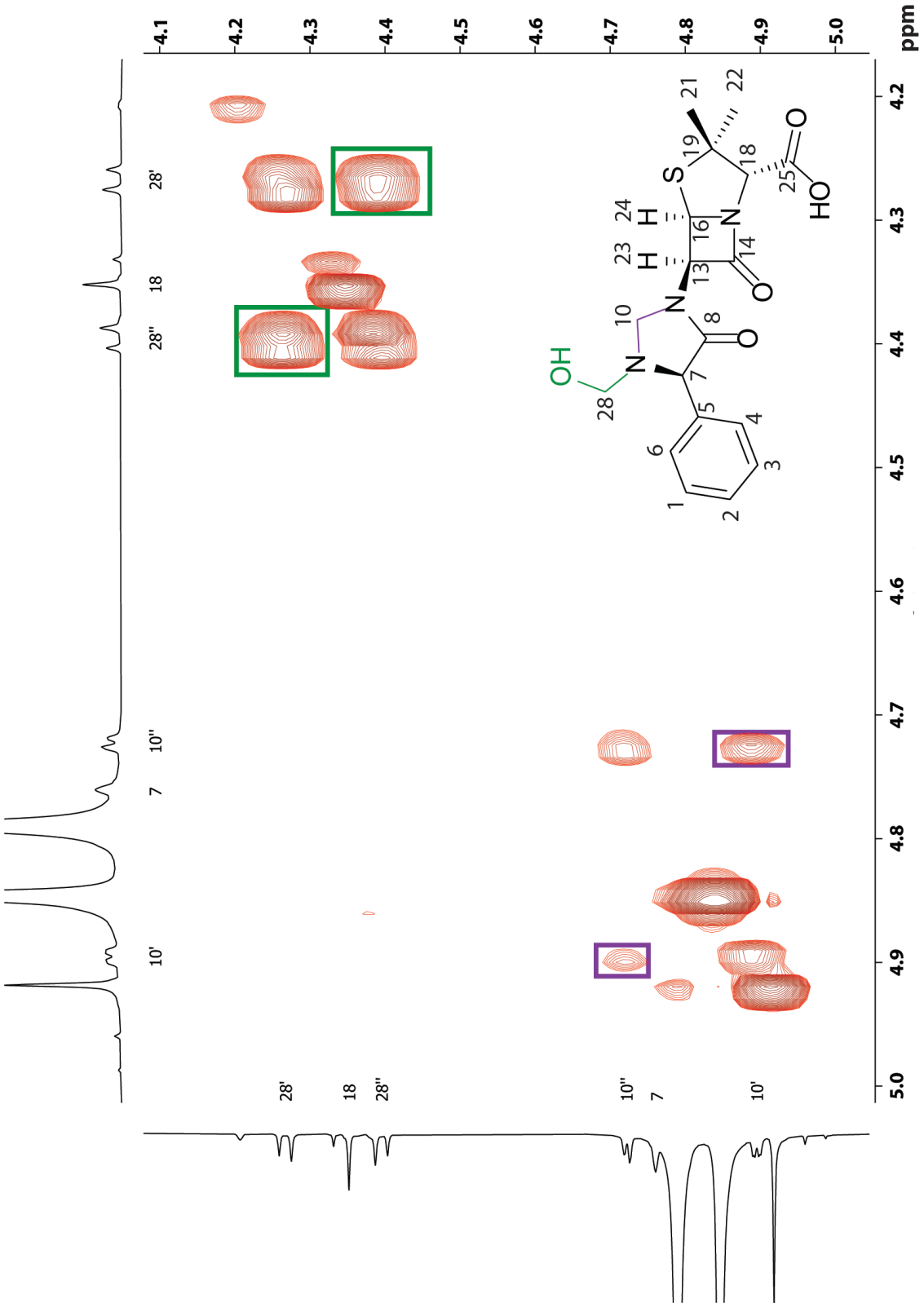


Figure 4.6 | ^{1}H-^{1}H-COSY analysis of the *in situ* formed 3. A 10-fold excess HCHO was used to prepare **3** from **1** *in situ* in D_2O. The cyclic aminal methylene signals are boxed in purple and the hydroxymethyl signals are boxed in green. These results were obtained in close collaboration with Raphael Reinbold.

^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC analyses confirmed the assignment of the cyclic aminal (δ_H 4.90 and 4.72 ppm) and, crucially, enabled assignment of the novel HCHO-derived hemiaminal (**3**, δ_H 4.40 and 4.27 ppm). Importantly, C-10 (cyclic aminal methylene bridge, δ_C 64.3 ppm) and not C-28 (hydroxymethyl methylene, δ_C 71.4 ppm) correlates with hyrogen 23 (δ_H 5.62 ppm), suggesting C-10 forms the cyclic aminal methylene bridge, whereas C-28 is part of an exocyclic hydroxymethyl (Figure 4.7). The ^{1}H-^{13}C-HMBC analysis was also useful to identify whether or not the two HCHO-derived species co-exist and are both connected to the side-chain amine. The results revealed correlations between the hemiaminal (δ_H 4.40 and 4.27 ppm) and both the cyclic aminal methylene bridge (δ_C 64.3 ppm) and the carbon atom connected with the side chain α-amino group (δ_C 64.7 ppm), indeed suggesting co-existence of the two HCHO-derived groups (**3**) and connection to the side-chain amine (**3**, Figures 4.1, 4.3, 4.7).

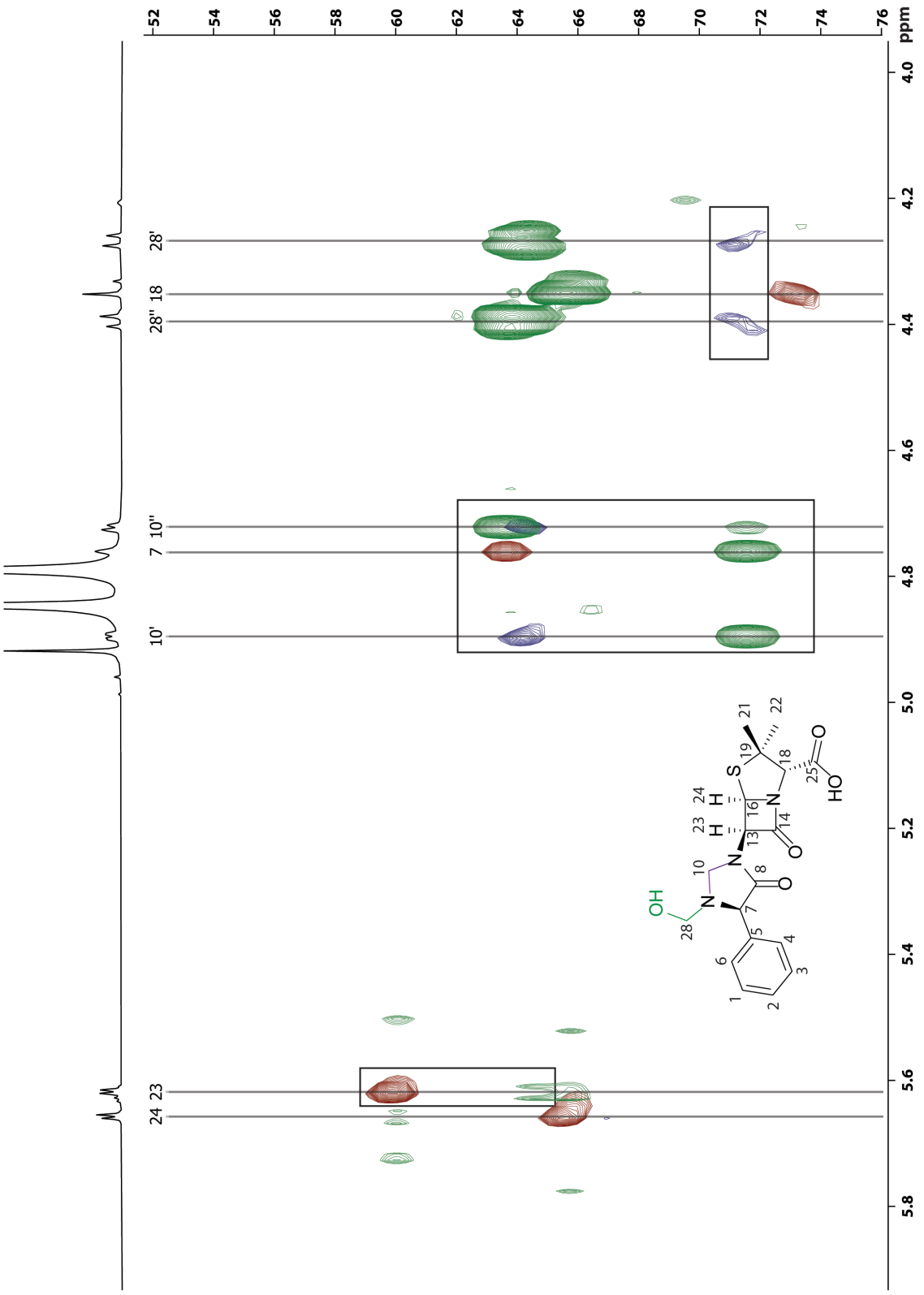


Figure 4.7 | ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC overlay of the methylene regions of 3 (700 MHz). **3** was prepared from reacting a 10-fold excess HCHO with **1** in D_2O. Interactions marked with boxes are crucial for structure elucidation of **3**. These results were obtained in close collaboration with Raphael Reinbold.

To investigate if the cyclic aminal methylene bridge and hemiaminal form at the same time or whether one of the HCHO-derived groups forms first, different HCHO concentrations were added to **1**. Upon addition of equimolar amounts of HCHO to **1**, δ_H resonances corresponding to the methylene bridge of **2** were detected after 0.5 h and increased over time, whereas the hemiaminal signals of **3** remained at low levels (Figure 4.8). These results indicate that the methylene bridge of **2** is the first of the two observed HCHO-derived adducts to form.

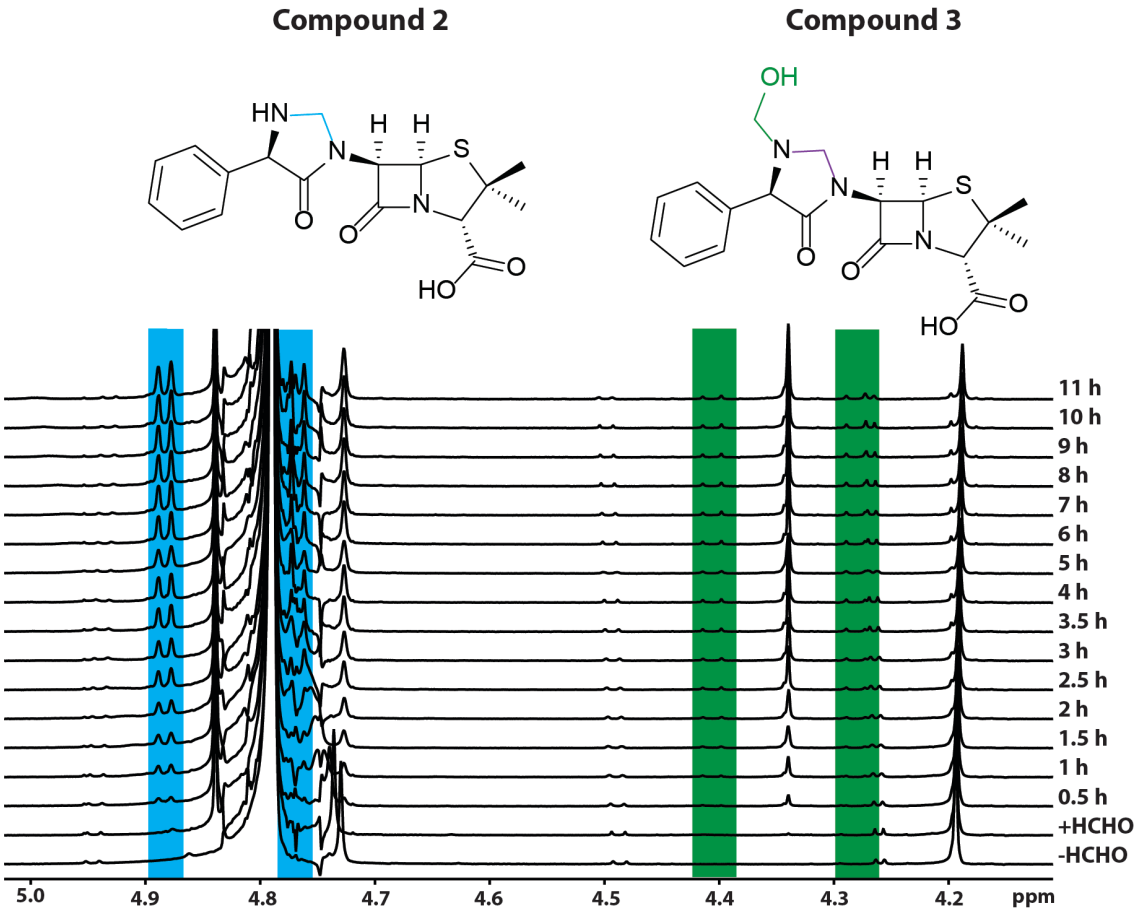


Figure 4.8 | Time-course studies on reaction of 1 with equimolar amounts of HCHO. Reaction of **1** with equimolar HCHO with monitoring by ^{1}H NMR (700 MHz) in D_2O. Under these HCHO-limiting conditions the time-course reveals formation of **2** (blue) as the major product and only trace amounts of **3** (green). The cyclic aminal hydrogen atoms of **3** were not detected, probably as a result of their low intensities. These results were obtained in close collaboration with Raphael Reinbold.

Upon increasing the amount of HCHO to a twofold excess (Figure 4.9 a), methylene bridge resonances of **2** and additional methylene bridge and hemiaminal resonances of **3** were detected. Notably, the methylene bridge of **2** is formed faster and is more abundant than the hemiaminal

resonance of **3** (Figure 4.9 a). Further reaction of HCHO with **2** to form **3** is observed at higher HCHO concentrations (Figure 4.9 b). This observation is illustrated by two regions in the spectra, namely the C<u>H</u>CO_2H (Figure 4.9 b) and the methyl group regions (Figure 4.11).

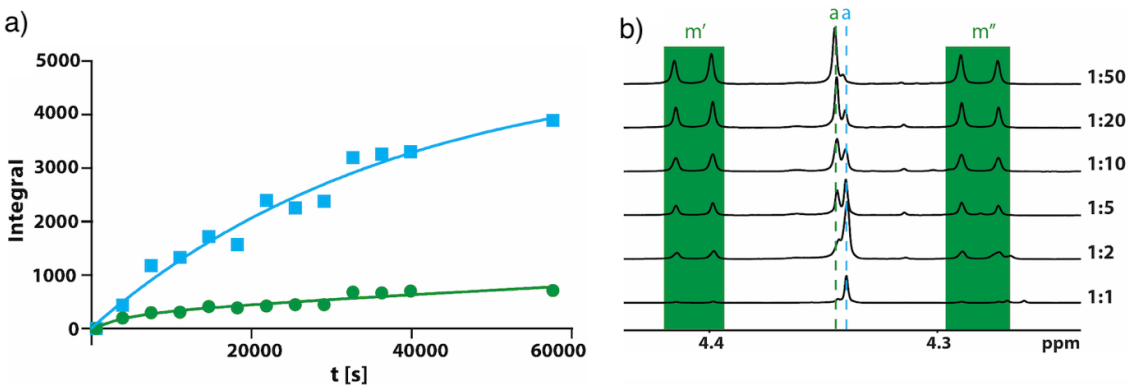


Figure 4.9 | Time-course studies on formation of 2 and 3. a) Plot of integrals showing the cyclic aminal resonance of **2** (blue) and the hemiaminal resonance of **3** (green), as seen upon addition of a twofold excess of HCHO to **1**. **b)** ^{1}H NMR spectra showing both hemiaminal protons (green, "m") of **3** upon addition of different amounts of HCHO to **1** after 4 h. The C<u>H</u>CO_2H hydrogens of **2** (higher field singlet, blue "a") and **3** (lower field singlet, green "a") result in two overlapping singlets at δ_H 4.34 ppm. Addition of further HCHO results in increased **3** and decreased **2**. These results were obtained in close collaboration with Raphael Reinbold.

The shifts of the C<u>H</u>CO_2H hydrogens are assigned to **2** (higher field singlet, blue "a") and **3** (lower field singlet, green "a") at δ_H 4.34 ppm (Figures 4.9 b, 4.10); the more HCHO is added to **1**, the more the lower field singlet corresponding to C<u>H</u>CO_2H in **3** is observed (Figures 4.9 b, 4.10). Note the decrease in the cyclic aminal signal in **2** (two doublets in blue) corresponds with an increase in the cyclic aminal signal in **3** (purple) with increasing HCHO concentrations (Figure 4.10).

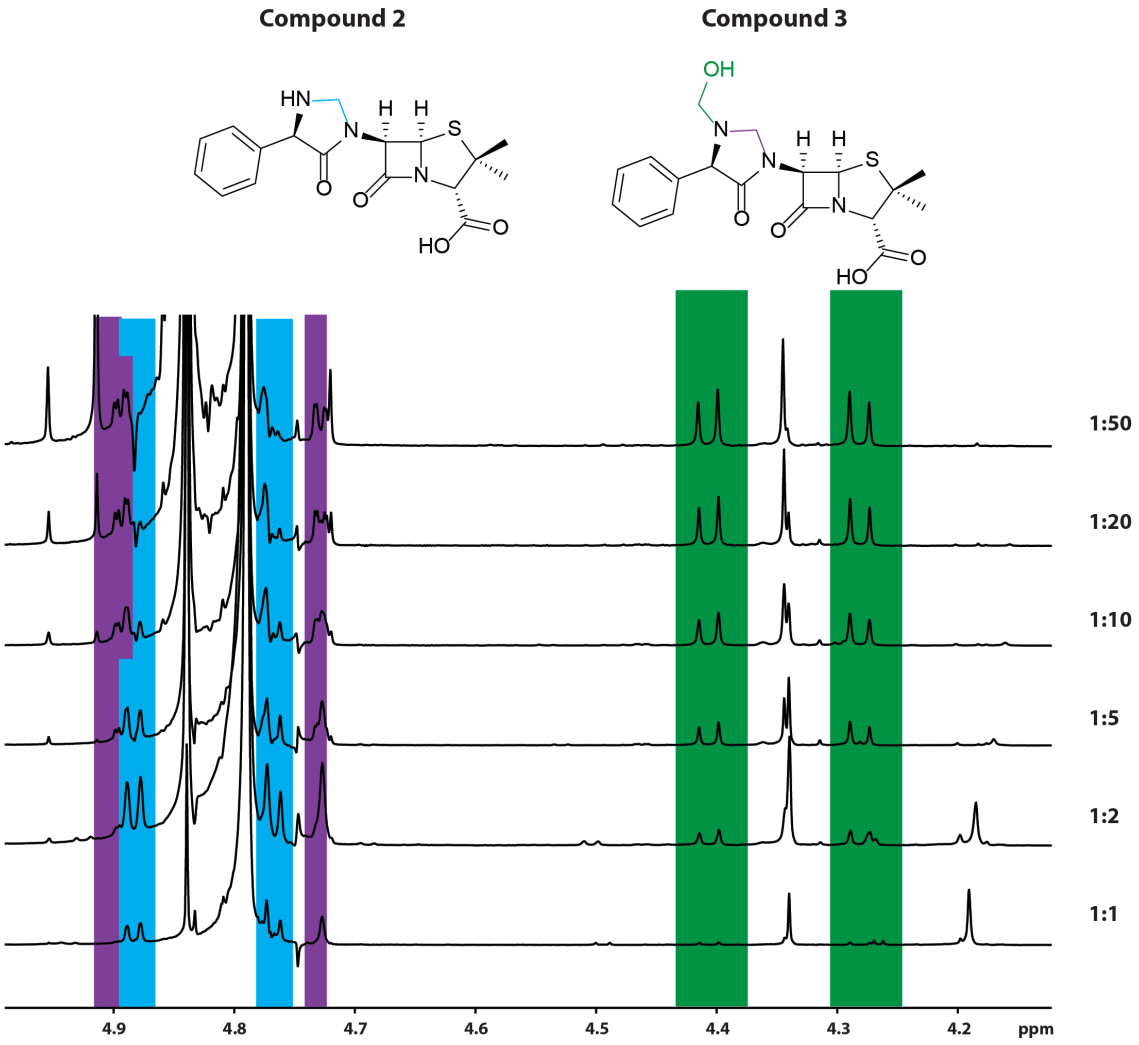


Figure 4.10 | Spectra of the cyclic aminal and hemiaminal regions at different HCHO concentrations. ^{1}H NMR (700 MHz) after mixing different amounts of HCHO with **1** (4 h, D_2O). The hydroxymethyl signal of **3** is shaded in green, whereas the cyclic aminal signals of **2** and **3** are shaded in blue and purple, respectively. These results were obtained in close collaboration with Raphael Reinbold.

Additionally, analysis of the methyl group region illustrates the further reaction of HCHO with **2** resulting in formation of **3**. The two signals at δ_H 1.45 and 1.51 ppm are assigned to **1**, and the two overlapping signals at δ_H 1.708 and 1.713 ppm correspond to **2** (higher field singlet) and **3** (lower field singlet), respectively. Increasing the amount of HCHO favours pushing the equilibrium towards **3** (Figure 4.11).

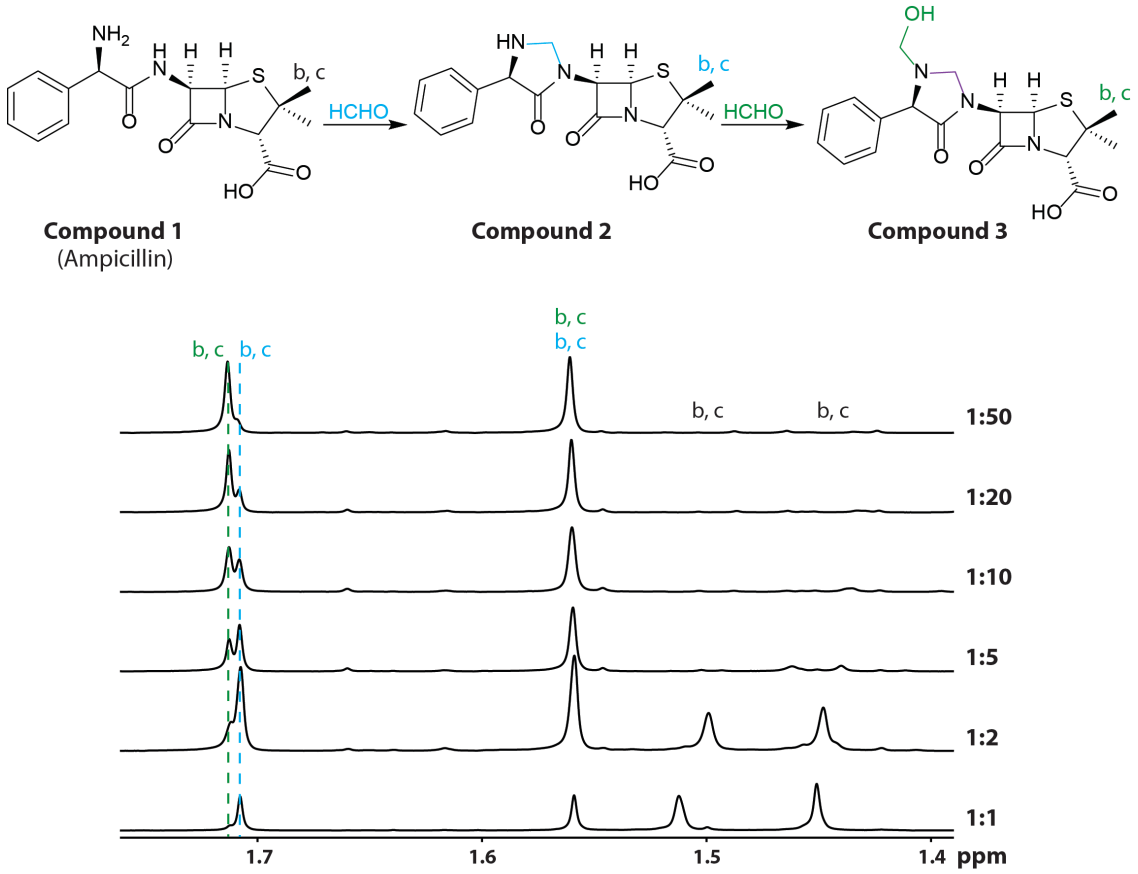


Figure 4.11 | Methyl group region showing reaction of 1 with different amounts of HCHO. ^{1}H NMR (700 MHz) spectra after 4 h incubation of **1** with different amounts of HCHO in D_2O. The signal at δ_H 1.56 ppm is assigned due to the overlapping resonances of one methyl group from **2** and one methyl group from **3**. These results were obtained in close collaboration with Raphael Reinbold.

To test whether the side chain α-amino group is more nucleophilic in basic conditions, leading to faster formation of the HCHO-derived species **2** and **3**, the time-course studies were repeated under alkaline conditions (pD 9). The results revealed faster formation of **3** at pD 9 compared to pD 7.4, as seen by the rapid shift of the methyl group signals. This observation suggests higher reactivity of the side-chain amine under basic conditions, rather than neutral conditions at pD 7.4 (Figure 4.12).

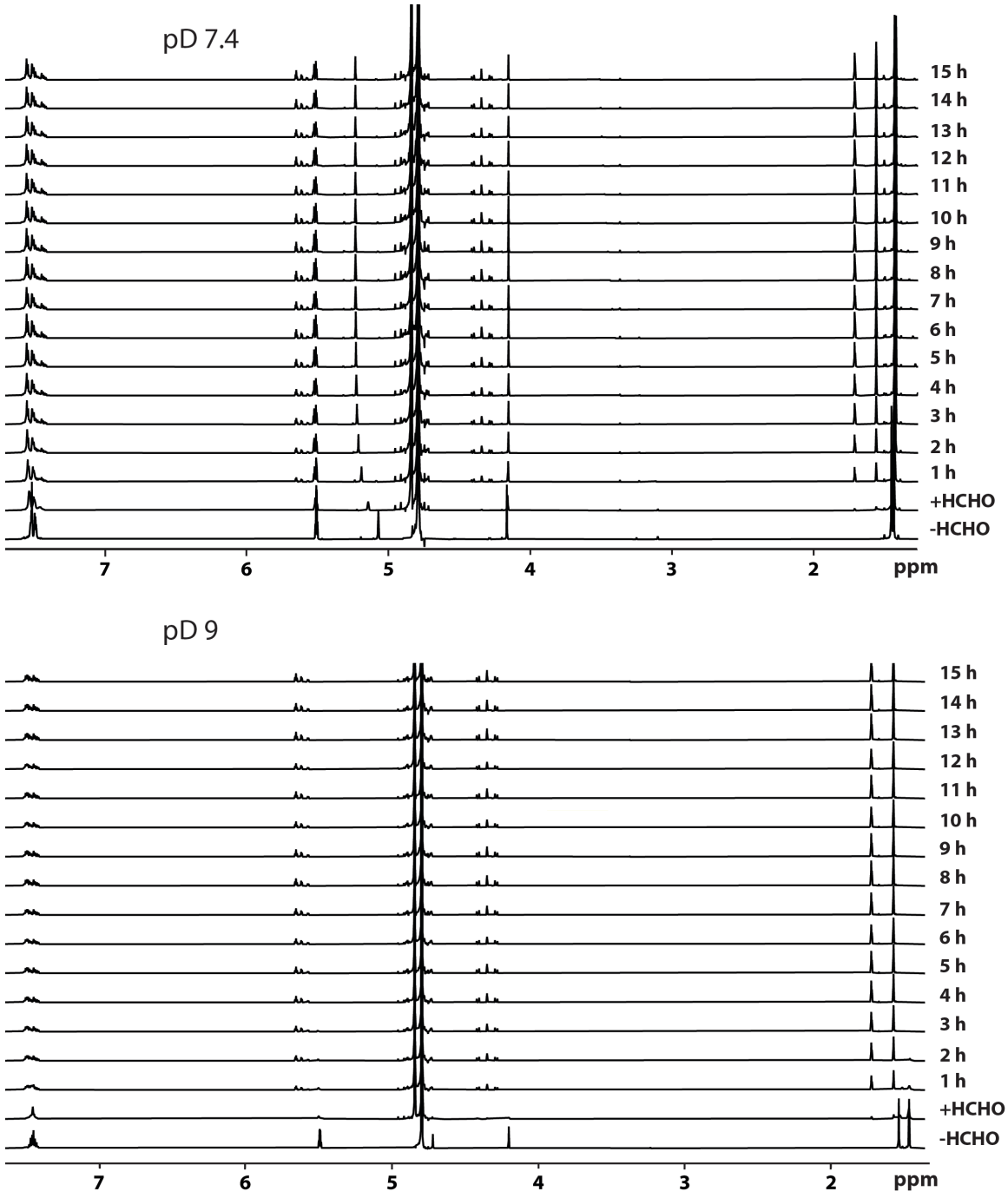


Figure 4.12 | Reaction of 1 with HCHO at different pH values. ^{1}H NMR (700 MHz) monitoring the reaction of **1** with 10-fold excess HCHO at pD 7.4 and pD 9, showing that the reaction is faster to proceed in basic conditions, best illustrated by the shift of the methyl group region (compare with Figure 4.11). These results were obtained in close collaboration with Raphael Reinbold.

To investigate the stability of the HCHO-derived groups in **2** and **3**, the HCHO scavenger 1,3-cyclohexanedione (CHD) was used, which reacts with HCHO to form dimeric adducts (Figure 4.13).[303]

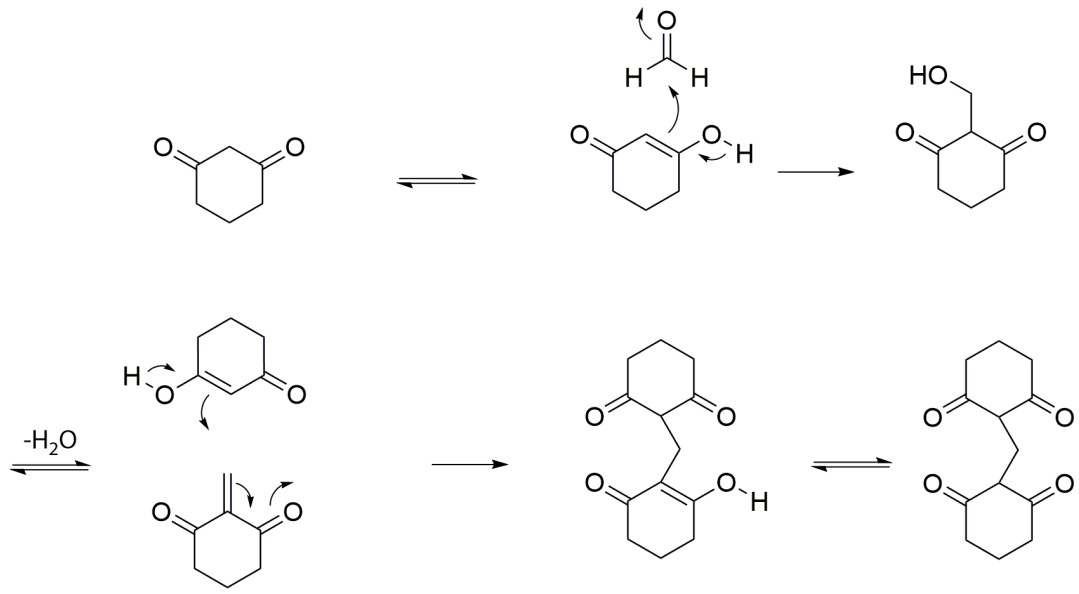


Figure 4.13 | Reaction mechanism of HCHO sequestration by 1,3-cyclohexanedione. The enol form of 1,3-cyclohexanedione reacts with HCHO to form a hemiaminal, which dehydrates to form an electron-poor alkene intermediate. This in turn can react with the enol form of an unreacted 1,3-cyclohexanedione molecule to form a stable methylene bridged dimer.[303]

For the stability studies, a solution of ampicillin (**1**) was reacted with a 2-fold excess of HCHO for 24 h at ambient temperature. Next, different amounts of CHD were added and the resultant mixtures were monitored by ^{1}H NMR and ^{1}H-^{13}C-HSQC. Upon addition of 1 or 2 equivalents of CHD, the hemiaminal group of **3** was removed to result in **2** (Figure 4.14 a), whereas the methylene bridge of **2** remained. However, upon addition of 4 equivalents CHD, the methylene bridge of **2** was observed to decrease in intensity, too (Figure 4.14 b). Addition of 20 equivalents CHD resultes in complete removal of the metylene bridge in **2** after 10 min (Figures 4.14 c and 4.15). In summary, these results indicate that the cyclic aminal in **2** is relatively stable, whereas the hemiaminal in **3** is labile.

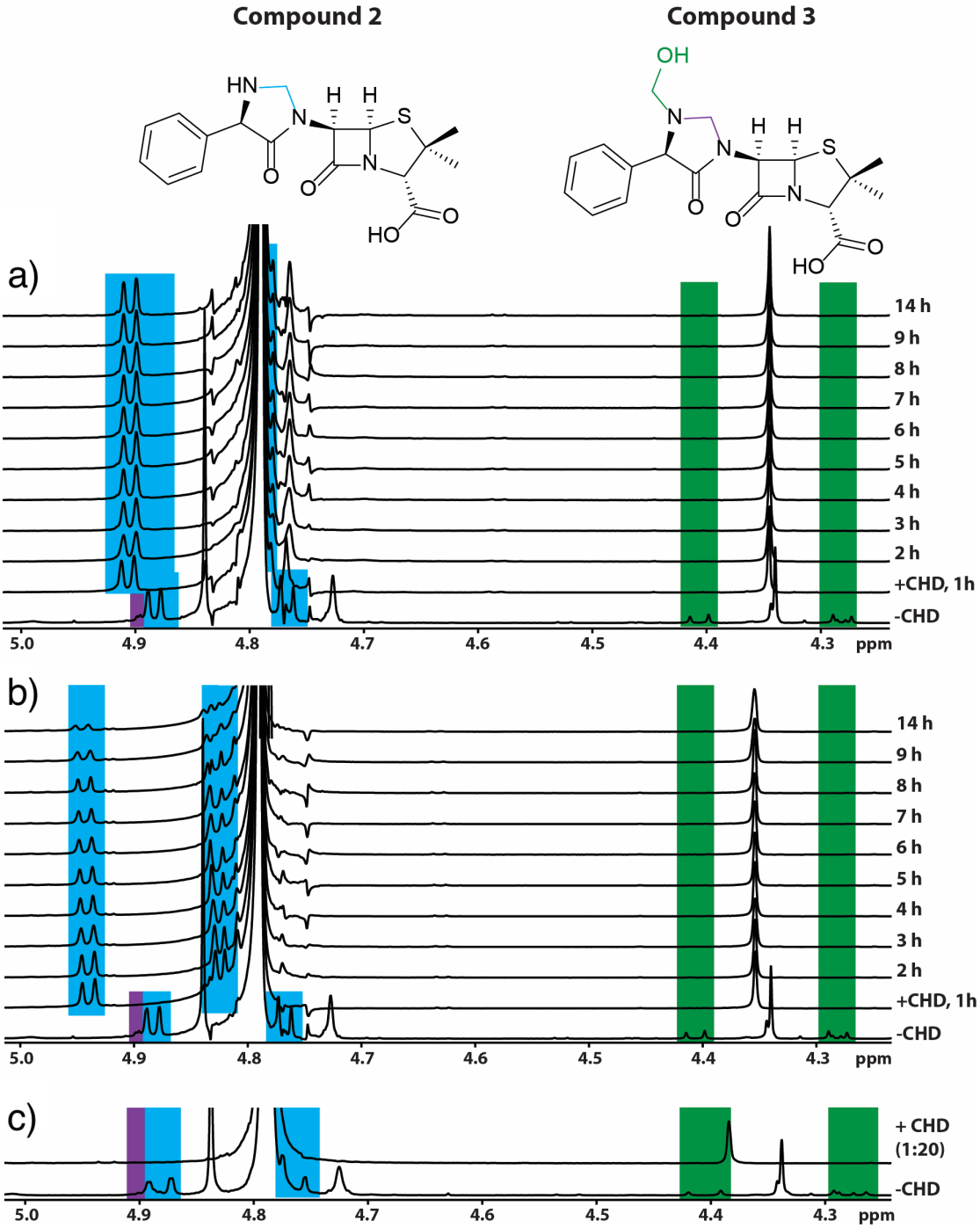


Figure 4.14 | Time-course studies on the reactions of 2 and 3 upon addition of CHD. 1 was reacted in D_2O with a twofold excess HCHO overnight to result in **2** and **3**, followed by 1,3-cyclohexanedione (CHD) addition and monitoring by ^{1}H NMR (700 MHz). **a)** A twofold excess of CHD (with respect to HCHO) was added and the reaction followed by ^{1}H NMR, showing loss of the hydroxymethyl group (green), resulting in formation of **2**. **b)** A 4-fold excess of CHD leads to loss of the cyclic aminal methylene group over time, resulting in formation of **1**. **c)** ^{1}H NMR (400 MHz) showing that increasing CHD to a 20-fold excess results in immediate loss of all HCHO-derived groups (Figure 4.15). The cyclic aminal signals of **2** and **3** are shaded in blue and purple, respectively. These results were obtained in close collaboration with Raphael Reinbold.

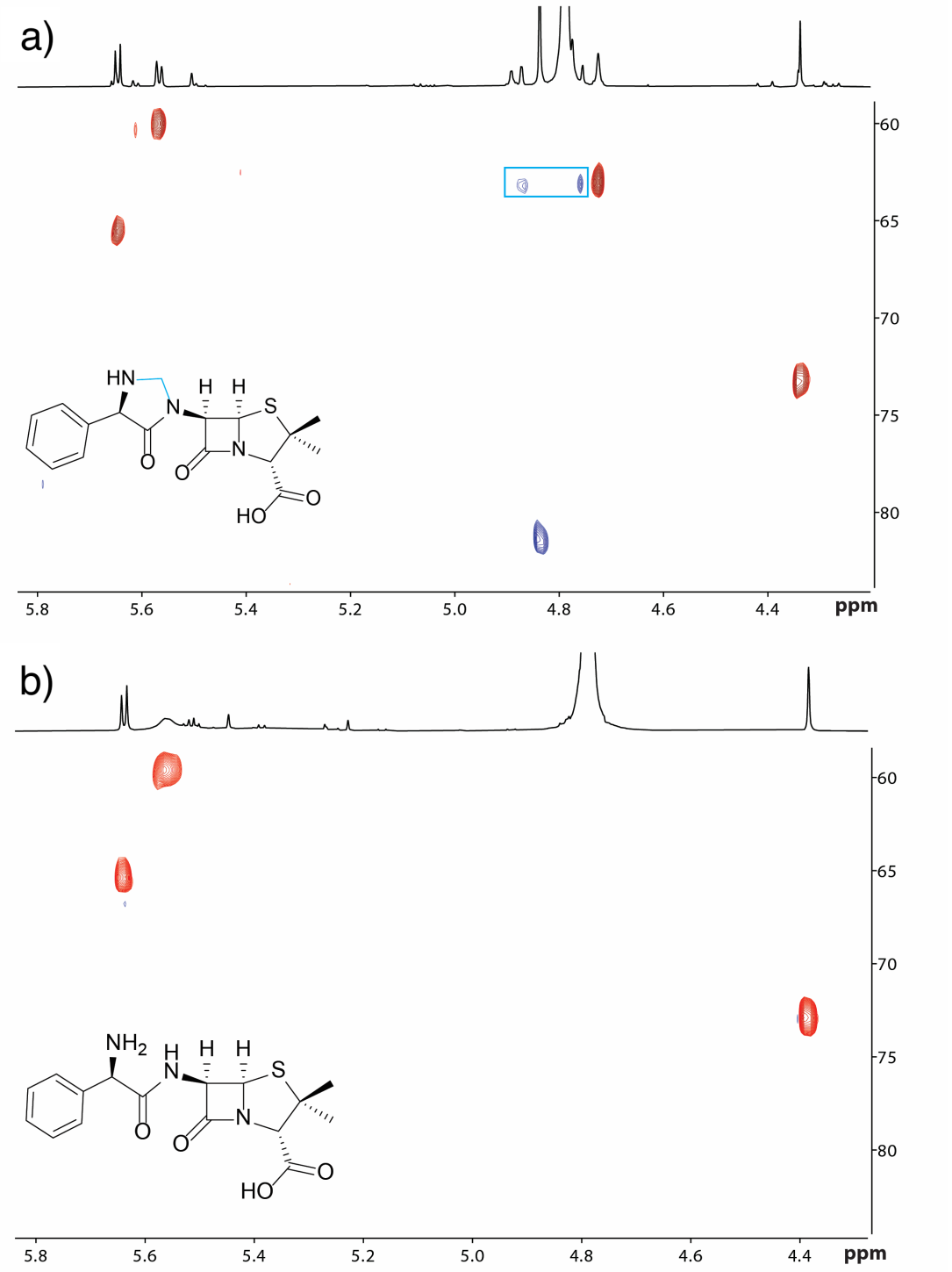


Figure 4.15 | Stability of the methylene bridge of 2 upon reaction with CHD. a) ^{1}H-^{13}C-HSQC (400 MHz, D_2O) confirming the presence of the methylene bridge in **2** upon reaction of **1** with a twofold excess of HCHO overnight. **b)** ^{1}H-^{13}C-HSQC (400 MHz, D_2O) confirming the loss of the methylene group of **2** upon addition of a 20-fold excess of CHD. These results were obtained in close collaboration with Raphael Reinbold.

4.4 Discussion

The NMR studies on the reaction of ampicillin (**1**) in water show the formation of two major products. **2**, which likely represents the major component of solid metampicillin, contains a HCHO-derived cyclic aminal methylene bridge which is first to form under limiting HCHO conditions. **3** is formed after reaction of an additional HCHO molecule which forms a hemiaminal off the side-chain amino group. No evidence was accrued for the previously reported structures of metampicillin, i.e. **4** and **5**,[301, 302] although it is likely these are intermediates *en route* to **2** and **3** (Figure 4.16). Stability studies on **2** and **3** using CHD show that the cyclic aminal is more stable than the hemiaminal. Only **2** is observed upon lyophilization of a mixture of **1** with excess HCHO (Figure 4.2). Taken together, these results reveal the potential for complexity in the reactions of HCHO with drugs bearing multiple nucleophilic groups in close proximity, enabling ring formation with HCHO.

The detailed elucidation of the structure of metampicillin could be relevant to an understanding why metampicillin accumulates in bile compared to ampicillin (**1**),[212] and why metampicillin (**2**) has different physicochemical properties compared to ampicillin.[82–84] These findings might be extended to identify pro-drug-like HCHO derivatives of existing small molecule therapeutics, with the aim to target specific tissues and organs.[304] These pro-drugs might have enhanced stability and bioavailability, as illustrated by the use of acyloxymethyl groups with therapeutics.[63] Further, these findings may be extended to use other reactive carbonyl compounds to target drugs in a tissue-specific manner. Although HCHO is the simplest aldehyde, this work further illustrates the potential intricacies of the reactions of HCHO with nucleophilic, biologically relevant molecules, which often occur in a reversible manner. The reactivity of HCHO could be harnessed in case of ampicillin (**1**) to use HCHO as a protecting group of the side-chain amine and the amide nitrogen, which can be removed by addition of HCHO scavengers such as CHD, cysteine or GSH.

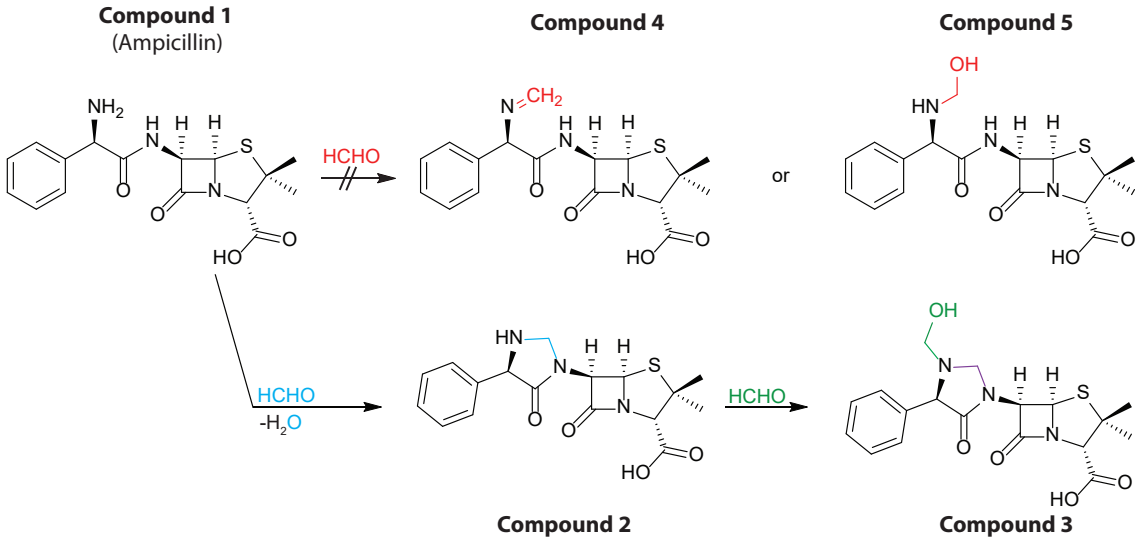


Figure 4.16 | Summary of previously reported metampicillin strucutres and structures identified in this work. Reaction of **1** with HCHO was reported to result in **4** and **5**. However, **2** was identified as major component of metampicillin. Formation of **3** results from additional reaction of **2** with HCHO *in situ*.

4.5 Methods

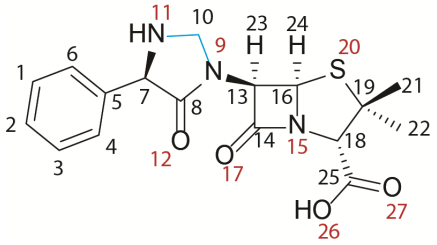
Reagents

Paraformaldehyde, NaOD, DCl and 1,3-cyclohexanedione were from Sigma-Aldrich, and sodium ampicillin was from Apollo Scientific.

4.5.1 Characterisation methods

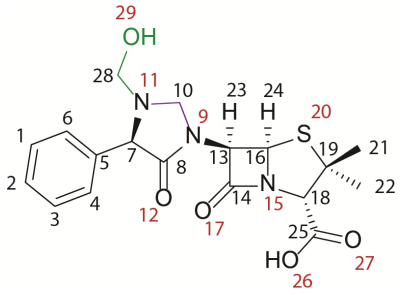
NMR spectra (^{1}H, ^{13}C, ^{1}H-^{13}C-HSQC, ^{1}H-^{13}C-HMBC) were recorded on either Bruker Avance AV400, AV600 or AV700. Edlock was used to reference spectra to the solvent peak (δ_H D_6-DMSO = 2.5 ppm, δ_C D_6-DMSO = 39.52 ppm; δ_H = D_2O 4.79 ppm). For 2D NMR spectra, the color code is the following: blue denotes CH_2, red denotes CH or CH_3 coupling (^{1}H-^{13}C-HSQC); green denotes ^{1}H-^{13}C-HMBC coupling. A Waters LCT Premier ESI spectrometer was used to measure high-resolution mass spectra (HRMS). An Agilent single quadrupole machine was used to obtain mass spectra. A Stuart automatic melting point SMP40 was used to obtain melting points with a heating rate of 20 °C per minute. A Bruker Tensor 27 FT-IR spectrometer was used to obtain infrared (IR) spectra using attenuated total reflection at ambient temperature.

4.5.2 Characterisation of HPLC purified metampicillin (2)



Metampicillin (**2**) was obtained as colorless solid using the procedure described below in Chapter 4.5.4 (12 mg, 33 µmol, 33%). ^{1}H NMR (700 MHz, D_6-DMSO) δ_H = 7.44 – 7.39 (m, 2H, 4, 6), 7.38 – 7.33 (m, 2H, 1, 3), 7.31 – 7.26 (m, 1H, 2), 5.62 – 5.53 (m, 2H, 23, 24), 4.82, 4.52 (ABq, J = 6.7 Hz, 2H, 10', 10"), 4.54 (s, 1H, 7), 4.36 (s, 1H, 18), 1.65 (s, 3H, 21 or 22), 1.49 (s, 3H, 21 or 22); ^{13}C NMR (101 MHz, D_6-DMSO) δ_C = 173.7 (C8), 171.2 (C14), 168.8 (C25), 138.4 (C5), 128.2 (C1, C3), 127.4 (C2), 127.2 (C4, C6), 70.2 (C18), 66.3 (C16), 64.7 (C19), 63.0 (C10), 60.9 (C7), 60.2 (C13), 30.5 (C21 or C22), 26.7 (C21 or C22); ESI-MS, $[M+H]^+$: 362.12 (calculated), 362.100 (observed); HRMS (ESI+) $[M+H]^+$: 362.1169 (calculated for $C_{17}H_{20}N_3O_4S$), 362.1169 (observed); IR $\tilde{\nu}$ 1768.84 cm^{-1}, 1737.55 $^{-1}$ (CON) (Figure 4.18); melting point: 195 °C (degradation).

4.5.3 Characterisation of *in situ* reaction product (3)



^{1}H NMR (700 MHz, D_2O) δ_H = 7.51 – 7.43 (m, 5H, 1, 2, 3, 4, 6), 5.66 (d, J = 3.9 Hz, 1H, 24), 5.62 (d, J = 3.9 Hz, 1H, 23), 4.90, 4.72 (ABq, J = 5.3, 2H, 10', 10''), 4.76 (s, 1H, 7), 4.40, 4.27 (ABq, J = 11.4, 2H, 28', 28''), 4.35 (s, 1H, 18), 1.72 (s, 3H, 21 or 22), 1.57 (s, 3H, 21 or 22); ^{13}C NMR (101 MHz, D_2O) δ_C = 174.0 (C8), 171.9 (C14), 171.9 (C25), 135.5 (C5), 129.0 (C1-4, C6), 71.4 (C28), 73.2 (C18), 65.6 (C16), 65.4 (C19), 64.3 (C10), 63.7 (C7), 60.0 (C13), 30.1 (C21 or C22), 26.7 (C21 or C22).

4.5.4 Synthesis of metampicillin

The synthesis of metampicillin was modified after [302]. 0.1 mmol sodium ampicillin (**1**, 37 mg) was reacted with 1 mmol (10 equivalents) aqueous HCHO (1 ml of a 1 M solution) and stirred at ambient temperatrure for 2 h. The resultant mixture was subjected to HPLC purification (2% MeCN (v/v) in H_2O, 0.1% formic acid (v/v) to 60% (v/v) over 12 min, reverse-phase column, Hichrom ACE5 C18, 100 x 21.2 mm). The product was lyophilised to yield a colourless solid, which was not soluble in D_2O.

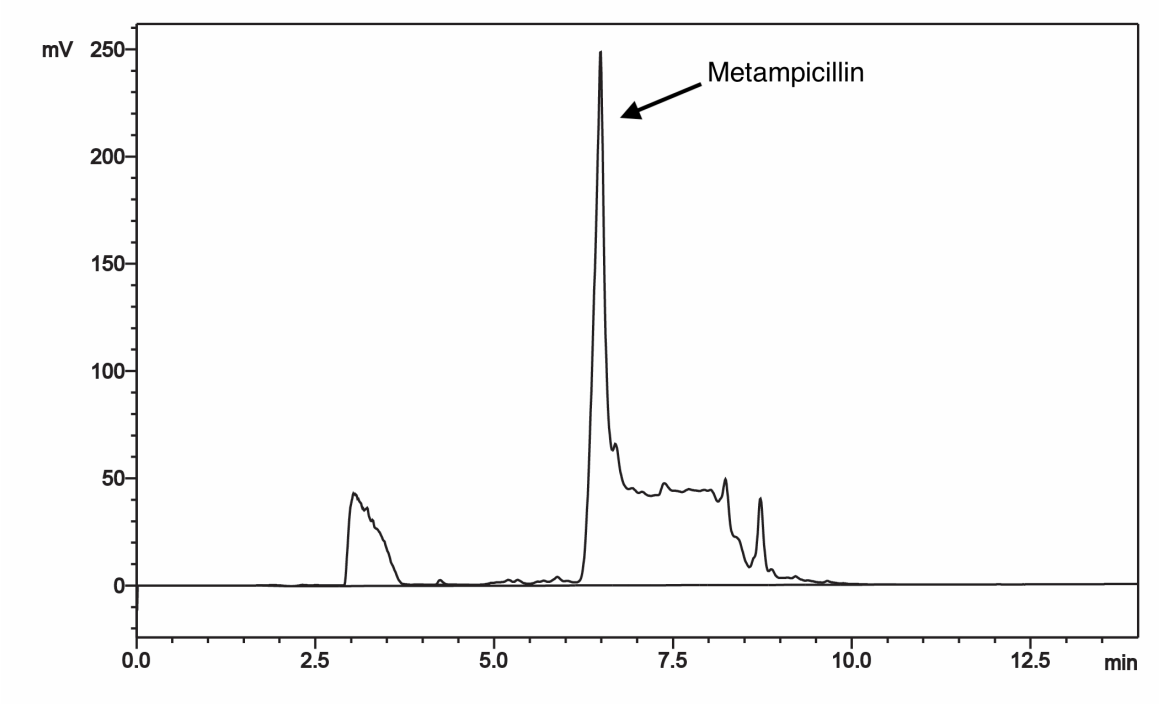


Figure 4.17 | HPLC trace showing the mixture obtained upon reaction of ampicillin (1) with a 10-fold excess HCHO. Fractions containing metampicillin were identified by LC-MS, combined and lyophilised.

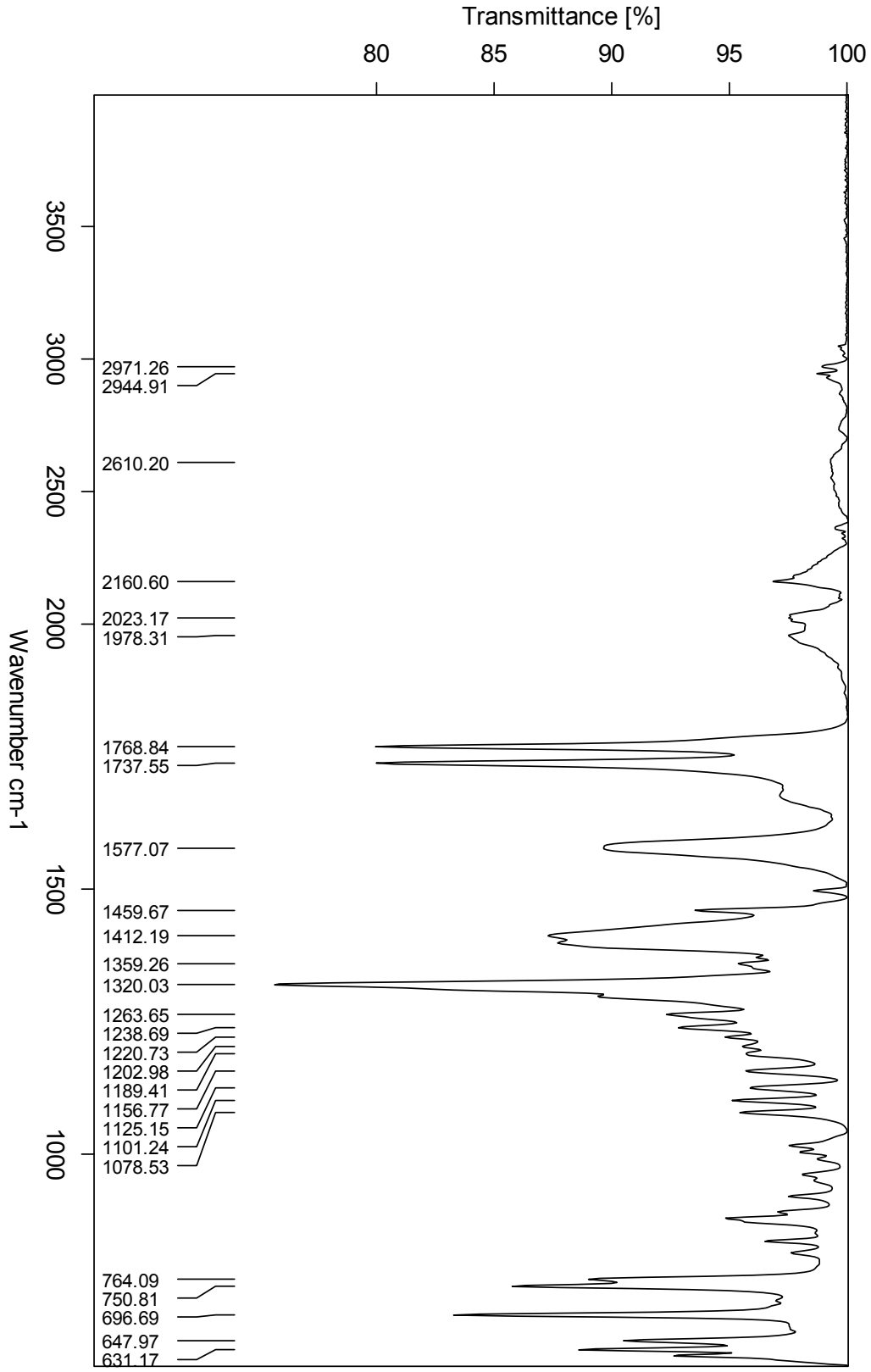


Figure 4.18 | IR analysis of HPLC purified 2. $\tilde{v}$ 1769 cm^{-1}, C=O lactam; 1738 cm^{-1}, C=O amide.

4.5.5 *In situ* studies for 2D NMR

The ampicillin sodium salt (**1**, 18.5 mg, 0.05 mmol) was reacted with 10 equivalents HCHO (500 µl of a 1 M stock in D_2O, 0.5 mmol) overnight at ambient temperature (no pD ajusted) (Figures 4.6, 4.7).

4.5.6 Time-course studies

The ampicillin (**1**) sodium salt (8.7 mg, 23.3 µmol) was dissolved in D_2O and the pD was adjusted to 9 or 7.4 (equivalent to pH 8.6 and 7, respectively) using DCl and NaOD, resulting in a total volume of 1 ml. For preparation of 0.33 M aqueous HCHO, paraformaldehyde was suspended in D_2O and heated until dissolved; subsequently, the pD was adjusted to 9 and 7.4. For time-course studies, 400 µl of D_2O was mixed with the ampicillin stock solution (100 µl, 2.33 µmol) and ^{1}H NMR spectra were recorded. Next, the HCHO stock with the respective pD (10 equivalents, 23.3 µmol) was added and the progress of the reaction was followed by ^{1}H NMR (700 MHz, Figure 4.12). Additional time-course studies were executed as described above, but without adjusting the pD: ampicillin (2.33 µmol) and different amounts of HCHO from a 0.67 M stock in D_2O were added to a final volume of 675 µl to D_2O (Figures 4.9, 4.5, 4.10, 4.8, 4.11); GraphPad Prism (Version 5.04) was used to plot ^{1}H integrals, and non-linear regression was used for curve fitting (Figure 4.9 a).

4.5.7 Stability studies

HCHO (7 µl of a 0.67 M stock in D_2O, 4.66 µmol, 2 equivalents) was added to **1** (100 µl of a 23.3 mM stock in D_2O, 2.33 µmol) and adjusted to a final volume of 675 µl with D_2O (no pD adjustment), and the resultant mixture was reacted overnight at ambient temperature. Subsequently, 600 µl (2.06 µmol ampicillin, 4.12 µmol HCHO) of this mixture was transferred to an NMR tube, followed by recording ^{1}H-^{13}C-HSQC (Figure 4.15 a) and ^{1}H spectra. Next, an excess of CHD was added (1:2, 1:4, 1:20, relative to HCHO) and the intensity of the HCHO-derived moieties in **2** and **3** was monitored by ^{1}H-^{13}C-HSQC (Figure 4.15 b) and ^{1}H NMR (Figure 4.14).

5

SARS-CoV-2 Proteases

I would not be one of those who will foolishly drive a nail into mere lath and plastering; such a deed would keep me awake nights. Give me a hammer, and let me feel for the furring. Do not depend on the putty. Drive a nail home and clinch it so faithfully that you can wake up in the night and think of your work with satisfaction [...]. Every nail driven driven should be another rivet in the machine of the universe, you carrying on the work. Rather than love, than money, than fame, give me truth.

Henry David Thoreau - Walden [305]

Contents

5.1 Abstract **238**
5.2 Introduction **238**
5.3 Results **240**

The work described in this chapter was done during the first pandemic induced lockdown in the United Kingdom between April and July 2020 where I was among a few lucky volunteers who were granted full access to the Chemistry Research Laboratory to work on SARS-CoV-2. Whilst I did not lead the studies conducted, I did essential supporting work, including by the design and synthesis of peptide substrates, as well as carrying out mass spectrometry analysis, supporting work described in publications [246], [306], [307], and [308].

5.1 Abstract

SARS-CoV-2 is the etiological agent that has caused the outbreak of the coronavirus pandemic in 2019. SARS-CoV-2 is a positive strand RNA virus that encodes two functional proteases, namely M^{pro} and PLPro, both of which are essential for viral replication. As nucleophilic cysteine proteases, these proteases might be amenable to inhibition by anti-viral chemotherapy. This was tested in the Schofield group using high-throughput solid phase extraction coupled to mass spectrometry (SPE-MS). M^{pro} has a strong tendency to dimerize in solution; this was harnessed to investigate on allosteric inhibitors that disrupt M^{pro} dimerization and hence modulate substrate cleavage, in an effort led by the group of Prof Carol V. Robinson. The combined inhibition results suggest β-lactams are potential candidates for covalent M^{pro} inhibition and M^{pro} dimerization is amenable for inhibition by non-covalent inhibitors. Using molecular dynamics and combined quantum mechanics/molecular mechanics (QM/MM), the groups of Prof Duarte and Prof Garrett Morris studied the detailed mechanism of substrate binding to M^{pro}, leading to crucial mechanistic insights that will help the development of future peptidomimetic inhibitors.

5.2 Introduction

Coronaviruses (CoVs) are part of the *Coronaviridae* family, which are positive-sensed, enveloped, single-stranded RNA viruses; their name arises from their crown-like appearance under the electron microscope.[309] Recent examples of CoV outbreaks are the epidemic Middle East Respiratory Syndrome CoV (MERS-CoV) in 2013 and the pandemic SARS-CoV in 2003; both diseases arise from zoonotic transmission, which is also true for SARS-CoV-2.[310] SARS-CoV-2 belongs to the genus *Betacoronavirus* and contains a 30-kb RNA genome encoding polyprotein precursors pp1a and pp1ab - the latter of which is generated by a -1 frameshift - various accessory proteins, and four structural proteins (spike, envelope, membrane, and nucleocapsid).[311] Upon injection of SARS-CoV-2 RNA into the host cell and translation, M^{pro} (also known as 3CLpro) is released by an autocleavage mechanism

which is not well understood; next, cleavage of the polyproteins by M^{pro} and PLPro into indivual proteins occurs, a process which is essential for viral replication.[312] Importantly, M^{pro} specifically cleaves after glutamine residues, a feature which has not been reported for any human protease, an observation that might help to enable less pan-protease side effects for M^{pro} inhibitors.[313] Upon non-covalent binding of substrate to M^{pro}, a scissile peptide bond reacts via His-41 enabled Cys-His dyad with Cys-145 to form a hydrolytically labile thioester (Figure 5.1).[246]

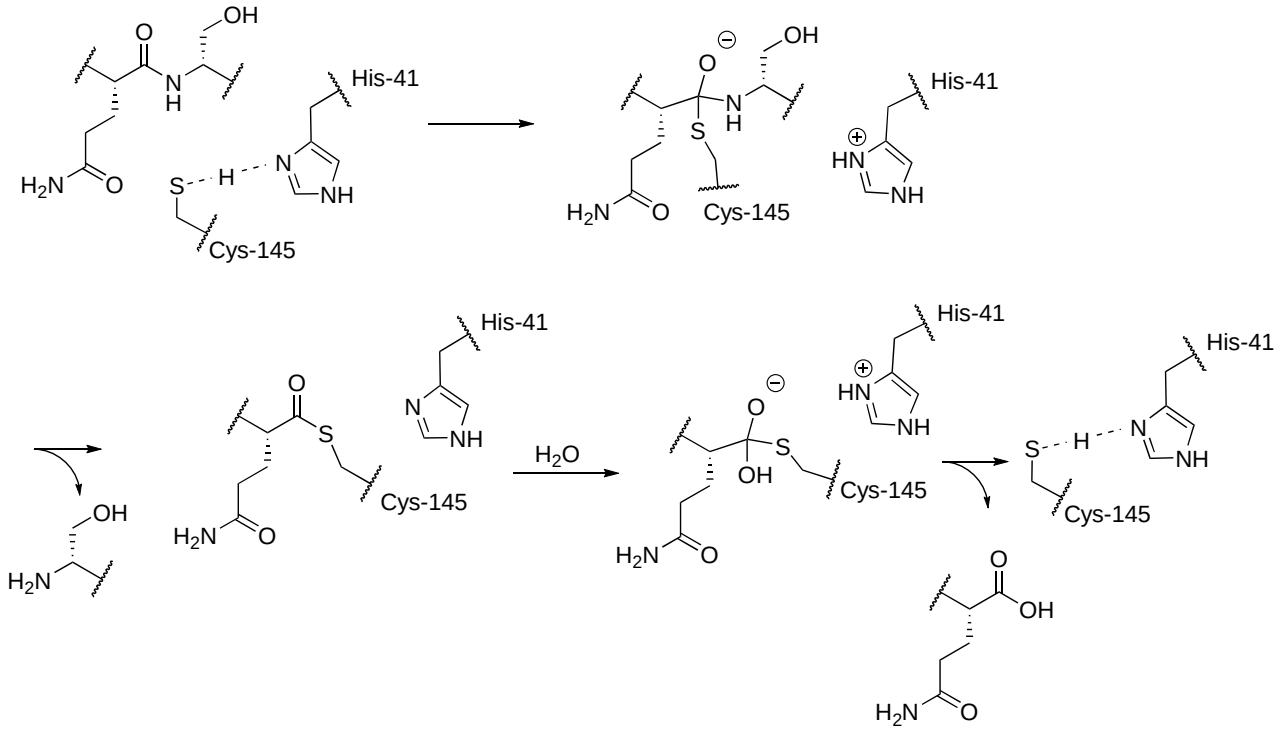


Figure 5.1 | Outline proteolytic mechanism of SARS-CoV-2 M^{pro} catalysis. Nucleophilic attack by Cys-145 on the peptide bond carbonyl results in a tetrahedral intermediate. The cleavage of the scissile peptide bond is facilitated by His-41, resulting in a thioester, which in turn is hydrolysed. The figure was modified from [314].

5.3 Results

M^{pro} is essential for processing SARS-CoV-2 viral polyproteins, resulting in biologically functional viral proteins, which then exert vital functions for the viral replication in its host.[315] As a nucleophilic cysteine protease, M^{pro} represents an attractive target for anti-viral chemotherapy. To screen for M^{pro} inhibitors, a peptide resembling the nsp4/5 cleavage site of the SARS-CoV-2 polyprotein was used; the M^{pro}-catalysed depletion of the peptide substrate was monitored by solid phase extraction coupled to mass spectrometry (SPE-MS).

I designed the peptide (NH_2-TSAVLQ'SGFRK-NH_2, apostrophe denotes the cleavage site) in a way that it is long enough for detection by SPE-MS, whilst not exceeding 30% of basic amino acids within the peptide substrate sequence; too basic sequences are typically not detected, most likely due to poor retention on the SPE C4-cartridge; SPE-MS was chosen over fluorescence based methods as both substrate depletion and product formation can be monitored simultaneously by SPE-MS, as well as covalent modification of the enzyme itself. The latter was a particularly important feature for the studies (conducted by Tika Malla, Schofield group), as potentially covalently reacting β-lactams were one of the main compound class screened in the assay. SPE-MS is a widely used technique for high-throughput screens of enzymes,[316–318] and was used to screen a set of available β-lactam compounds.

A penicillin V sulfone C3 benzyl ester (BI-41) (Figure 5.2 a+b) was found to exhibit an IC_{50} of 1.6 μM, suggesting β-lactams are useful scaffolds for acylating agents targeting M^{pro}. Some related penicillin esters were not observed to react with M^{pro} (Figure 5.2 c). Covalent modification data suggest that BI-41 reacts only a single time with M^{pro} (Figure 5.2 b).

N3 is an established compound that is known to only react with the active site cysteine of M^{pro}, whilst leaving the 11 non-active site cysteines unmodified.[319] Adding BI-41 to M^{pro} pre-incubated with N3 does not result in additional peaks, suggesting BI-41 only reacts with active site Cys-145 (Figure 5.2 d). The inhibitor data imply that β-lactam compounds represent scaffolds that can be used for *S*-acylation-mediated inhibition of thiol proteases. Such acylation reactions are in contrast with the established *S*-alkylating agents, and have been little explored.

More generally, the presented data show that SPE-MS is a useful technique to study inhibition by both product formation and substrate depletion, with the additional benefit of gaining insights for covalent M^{pro} modification from the same experiment. This study should give a starting point for lead optimization chemistry using β-lactams as scaffolds, with the ultimate goal of finding a clinically useful inhibitor to deactivate SARS-CoV-2 M^{pro}.

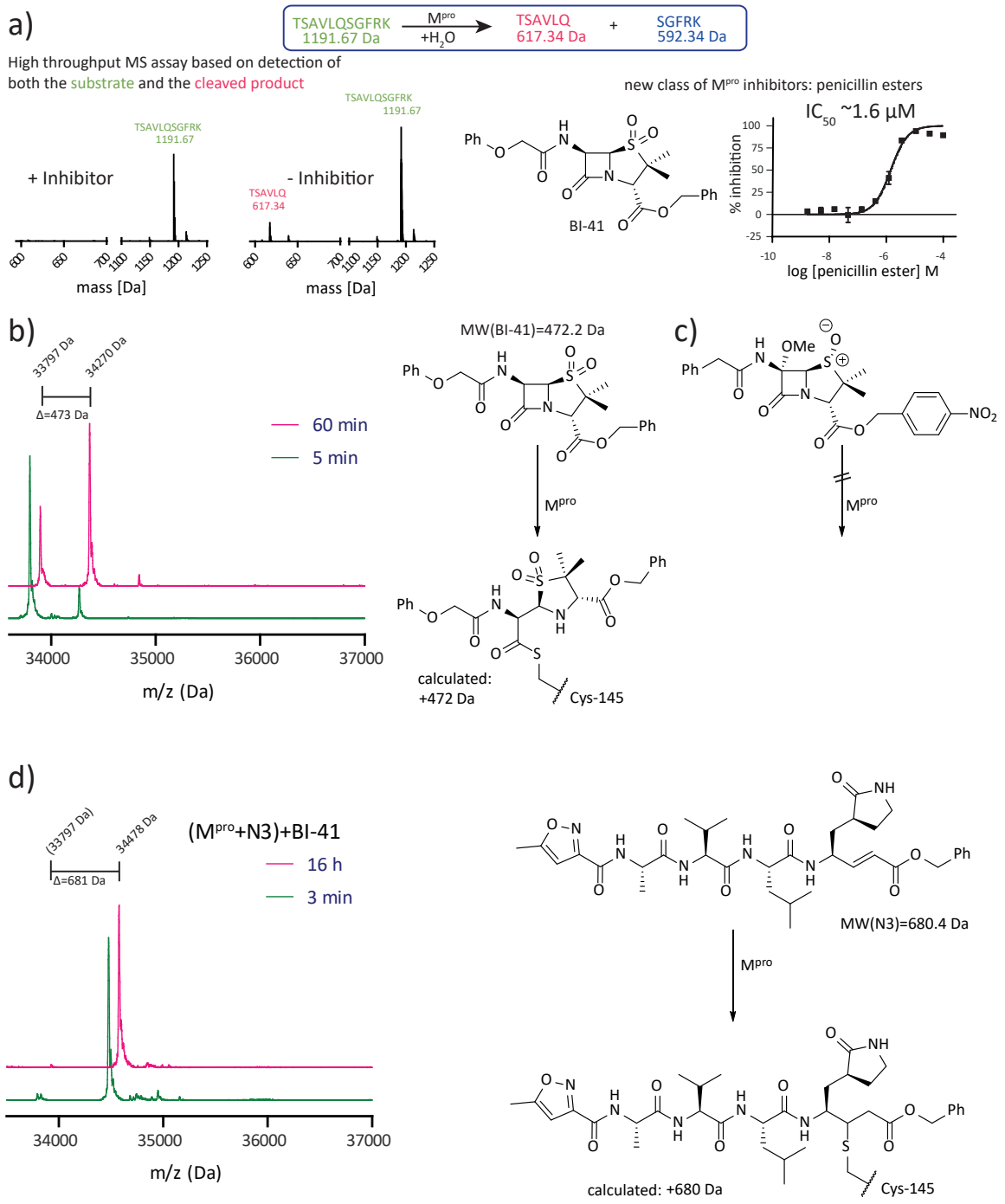


Figure 5.2 | Summary of selected M^{pro} SPE-MS work. a) An SPE-MS assay using a peptide substrate (green), resulting in two daughter peptides (red and blue) was employed to screen for covalent inhibitors of M^{pro}, including penicillin esters like BI-41. **b)** A 20-fold excess of BI-41 was incubated with M^{pro} for 5 min or 60 min, indicating BI-41 reacts one single time with M^{pro} (mass of wild-type M^{pro}: 33797 Da). **c)** Related penicillin esters do not exhibit reactivity with M^{pro}. **d)** M^{pro} preincubated with N3 results in modification of Cys-145;[319] upon addition of BI-41 to that mixture, no additional peaks are observed after 3 min and 16 h, suggesting BI-41 only reacts with Cys-145. SPE purification is denaturing, hence only monomeric M^{pro} is observed. The figure was modified from [246], MS data were generated by Tika Malla using the peptide substrate NH_2-TSAVLQ'SGFRK-NH_2 synthesised by me, synthesis of penicillin esters by Dr Lennart Brewitz.

M^{pro} forms a homodimer in solution,[320] an oligomeric form which is likely the optimal state for catalysis. Non-covalent allosteric inhibition, aiming to disrupt M^{pro} dimerization, is also a potential strategy to inhibit M^{pro}. Without reactive warheads, this compound class typically exhibits higher stability and less toxic side effects caused by off-target reactivity relative to covalent inhibitors.[306] The small molecule x1187 was found to bind to the dimer interface as observed by previous crystallographic screens (PDB 5RFA, Figure 5.3 a). x1187 was tested for its potential to destabilize the dimer by non-denaturing mass spectrometry (conducted by Tarick El-Baba, Carol Robinson group, using the peptide substrate NH_2-TSAVLQ'SGFRK-NH_2 synthesised by me). Without inhibitor present, M^{pro} forms a homodimer (deconvoluted mass of 67591±0.5 Da, in agreement with the calculated mass of dimeric M^{pro} (67592 Da, Figure 5.3 b (1), lowest panel). However, upon addition of a 20-fold molar excess of x1187, up to 15% monomeric M^{pro} is observed (Figure 5.3 b (1), upper panel). Next, x1187 was assessed for its potential to modulate enzymatic activity of M^{pro} using the previously used peptide substrate NH_2-TSAVLQ'SGFRK-NH_2 synthesised by me. Without inhibitor, the dimer M^{pro}-substrate signal is fully depleted after 10 min, suggesting full turnover of the substrate NH_2-TSAVLQ'SGFRK-NH_2 (Figure 5.3 b (2)), with a half life of about 150 s; notably, no signal was obtained for monomeric M^{pro}-substrate, suggesting no binding event of the peptide substrate to M^{pro} monomer (data not shown). Upon addition of x1187, the half life of the dimer M^{pro}-substrate complex increases to about 400 s (Figure 5.3 b (3)), suggesting x1187 indeed inhibits enzymatic activity.

In summary, the results obtained with x1187 may imply that the presented mass spectrometric approach is useful to gain insights into the mode of action of non-covalent inhibitors, whether it is by disrupting the dimerization or reversible non-competitive binding. This study should promote further research into the development of related non-covalent inhibitors that will be useful for treatment of CV-19 and viral diseases in general.

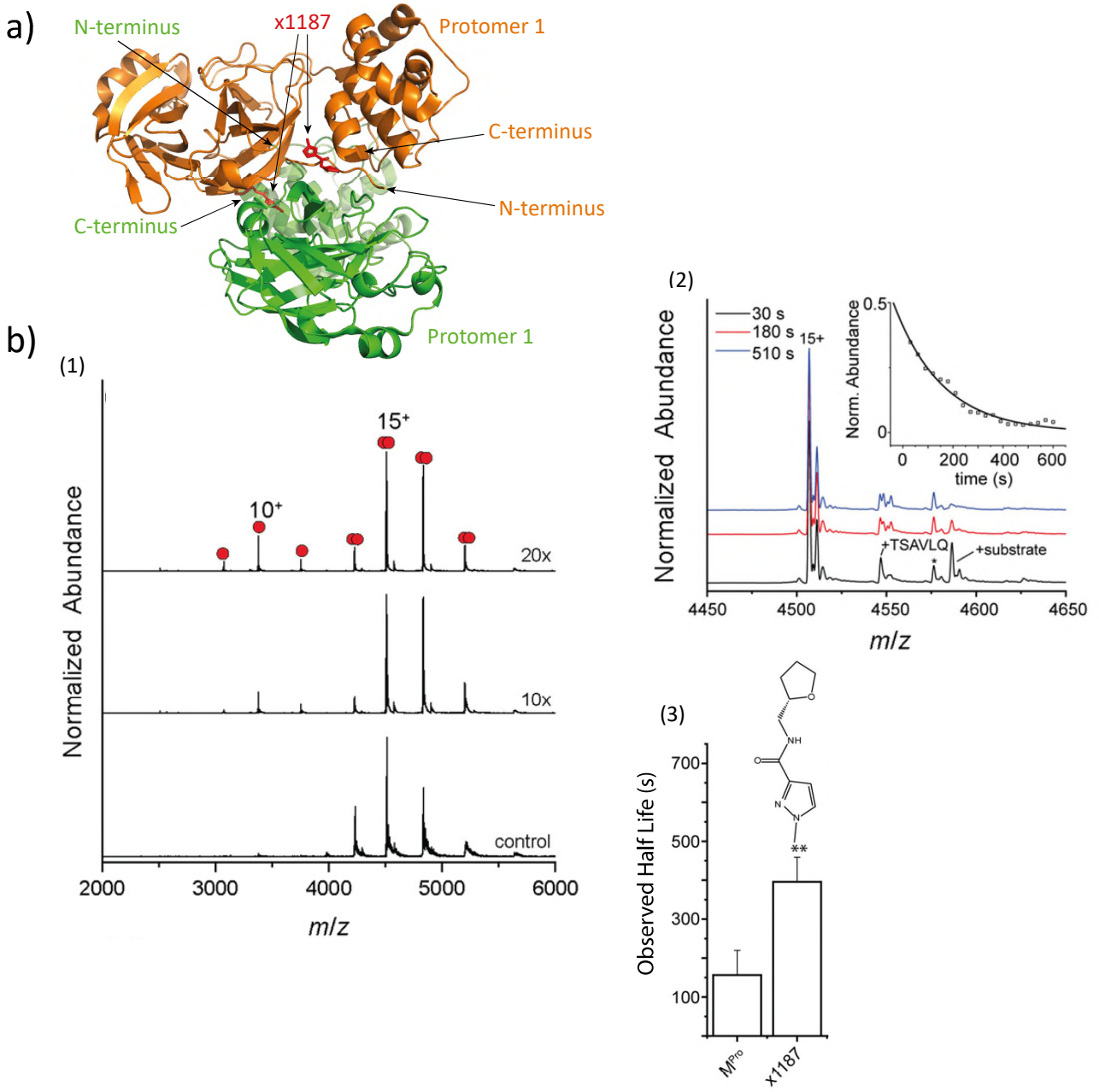


Figure 5.3 | Summary of CV19 non-denaturing MS work. **a)** View from a crystal structure of the M^{pro} dimer (PDB 5RFA), showing both monomers (green and orange) and the non-covalent inhibitor x1187 bound to the dimer interface. **b)** (1) Non-denaturing mass spectra of 5 µM M^{pro} upon addition of different molar equivalents of x1187 (10-fold and 20-fold), showing monomer formation (red circle); in DMSO control conditions (lowest panel), M^{pro} is dimeric (two overlapping red circles). (2) Three time points showing the substrate (NH_2-TSAVLQ'SGFRK-NH_2) cleavage reaction, indicating free dimeric M^{pro} (15+ state), dimeric M^{pro} with bound intact substrate, and dimeric M^{pro} with bound cleaved substrate. The inset shows the relative abundance of dimeric M^{pro}-substrate complex versus time. (3) Half-lives of dimeric M^{pro}-substrate complex in absence or presence of x1187, indicating x1187 inhibits M^{pro} activity. Figures are copied/modified from [306], data were generated by Dr Tarick El-Baba using the peptide substrate NH_2-TSAVLQ'SGFRK-NH_2 synthesised by me.

In another study conducted by H. T. Henry Chan, Marc A. Moesser, Rebecca K. Walters, Tika R. Malla and others, different classical molecular mechanics and quantum mechanical techniques, supported by SPE-MS-based kinetic data were used to investigate the molecular details involving recognition of the 11 natural M^{pro} substrates, which I designed and synthesised (Figure 5.4).[308] The S1 and S2 pockets (accommodating substrate residues P1 and P2, respectively) are major contributors to substrate binding by their characteristic hydrophilic character, major energy contribution to substrate binding, and conserved hydrogen bonds in S1 crucial for substrate recognition. Further, molecular dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling suggest a neutral catalytic dyad involving Cys-145 and His-41 is thermodynamically favoured in M^{pro} bound to an uncleaved substrate (Figure 5.1). Interestingly, the C-terminal part of substrates (P' side) seems to be bound more weakly than the N-terminal part (P-side); this may be a result of the P' side needing to act as leaving group after acyl-enzyme complex formation (Figure 5.1). Conversely, this might also suggest the P-side is more involved in substrate recognition, and hence the development of S-site-binding peptidomimetic inhibitors might seem a fruitful avenue for designing M^{pro} inhibitors. The modelling studies led to the design and synthesis of simple peptides that were shown to be M^{pro} inhibitors.

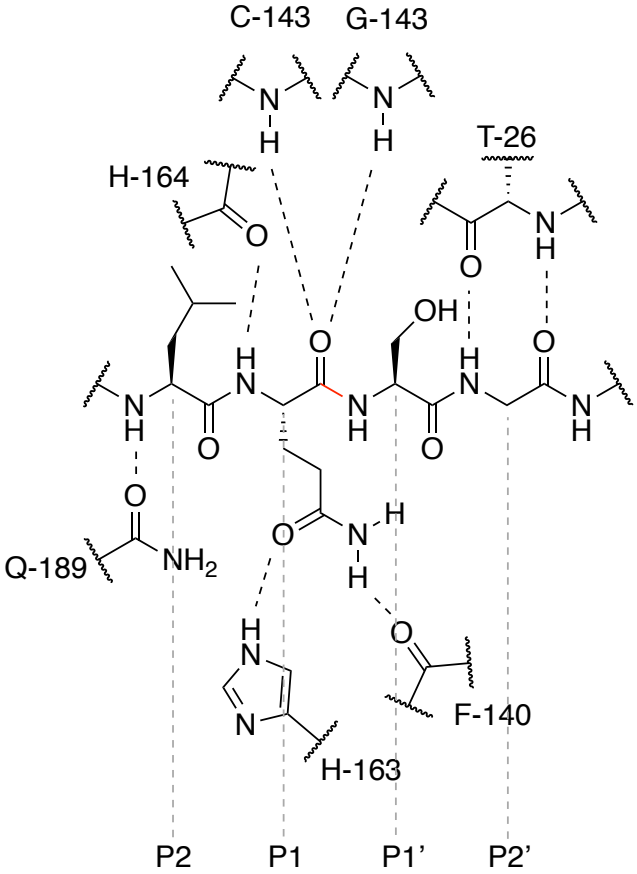


Figure 5.4 | Summary of M^{pro} modelling work. Hydrogen bonds between peptide substrate and M^{pro} residues, both involving side chain and backbone interactions. The scissile peptide bond is shown in red. The figure is modified from [308]; relevant kinetic data were generated by H. T. Henry Chan, Marc A. Moesser, Rebecca K. Walters, and Tika Malla using the 11 peptide substrates synthesised by me.

Besides studying M^{pro}, the second protease of SARS-CoV-2 was studied, namely PLPro. The recognition sequence consensus of PLPro is LXGG'XX, with X being any amino acid and the apostrophe representing the cleavage site.[321] In addition to cleaving viral proteins, PLPro also modulates the host's immune response by cleaving isopeptide bonds between ubiquitin (or the related interferon-stimulated gene (ISG)) and the N$^{\epsilon}$-amine of a substrate lysine (Figure 5.5 a+b).[322] With the knowledge gained by the SPE-MS assay developed for M^{pro}, it was envisioned that an SPE-MS based assay, and further a fluorescent assay could be deployed for PLPro. The fluorescent substrate NH_2-RLRGG'[AMC] (7-amino-4-methylcoumarin (AMC), bound via its amino group to the C-terminal glycine residue of the peptide), was used to study the PLPro-mediated cleavage of the scissile anilide bond (Figure 5.5 c). The release of the fluorogenic AMC enables determination of cleavage rates.[323] The fluorescence

detection assay shows that NH_2-RLRGG'[AMC] is turned over by PLPro (Figure 5.6 a). The PLPro-mediated cleavage of NH_2-RLRGG'[AMC], resulting in formation of free 7-amino-4-methylcoumarin (AMC), was confirmed by SPE-MS (Figure 5.6 b). Several non-fluorescent peptides, including NH_2-ELNGG'AVTRY-OH (Figure 5.6 c), NH_2-ELMRELNGG'AYTRYVDNN-NH_2, Ac-ALKGG'KIV-NH_2 and Ac-RELNGG'AYTRY-NH_2 (data not shown) were not cleaved under the same conditions that led to cleavage of the fluorescent substrate (Figure 5.6 b). Our group has continued to develop an SPE-MS based assay for PLPro. Preliminary SPE-MS data suggest that PLPro preferably cleaves N^{ϵ}-amine linked substrates rather than linear substrates under the tested conditions, indicating fluorescence based assays using substrates like NH_2-RLRGG'[AMC] that have been used in an array of different studies[324–326] might not be the most representative *in vitro* substrates for PLPro.

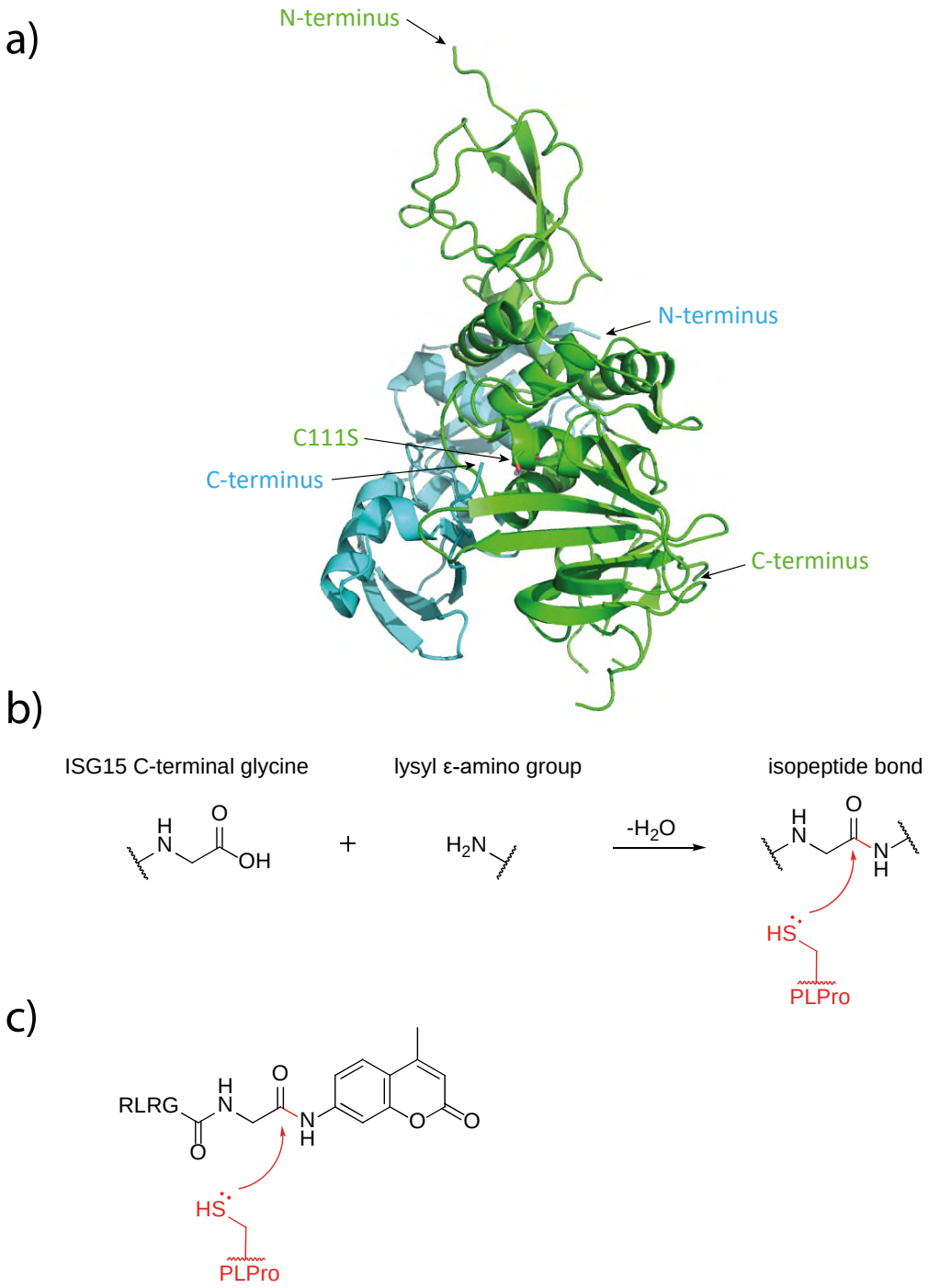


Figure 5.5 | ISG15 bound to PLPro C111S. a) View from a crystal structure of PLPro (green; the mutated active site serine is labelled and shown in sticks) bound to ISG15 (cyan), highlighting the close proximity of the C-terminus of ISG15 to the active site of PLPro at position 111 (PDB 6YVA). An isopeptide bond between the C-terminus of ISG15 and an N^{ϵ}-amine of a substrate lysine is a reported cleavage site for PLPro.[322] **b)** Conjugation of ISG15 via its C-terminal glycine and a substrate lysyl ϵ-amino group, resulting in an isopeptide bond, is a substrate for PLPro.[327] **c)** Fluorescent substrate containing the RLRGG motif of ISG15 bound via its C-terminal carboxyl group to the amino group of 7-amino-4-methylcoumarin (AMC), which has been used for inhibition studies on PLPro in the presented work. Scissile bonds and the active site cysteine of PLPro are shown in red.

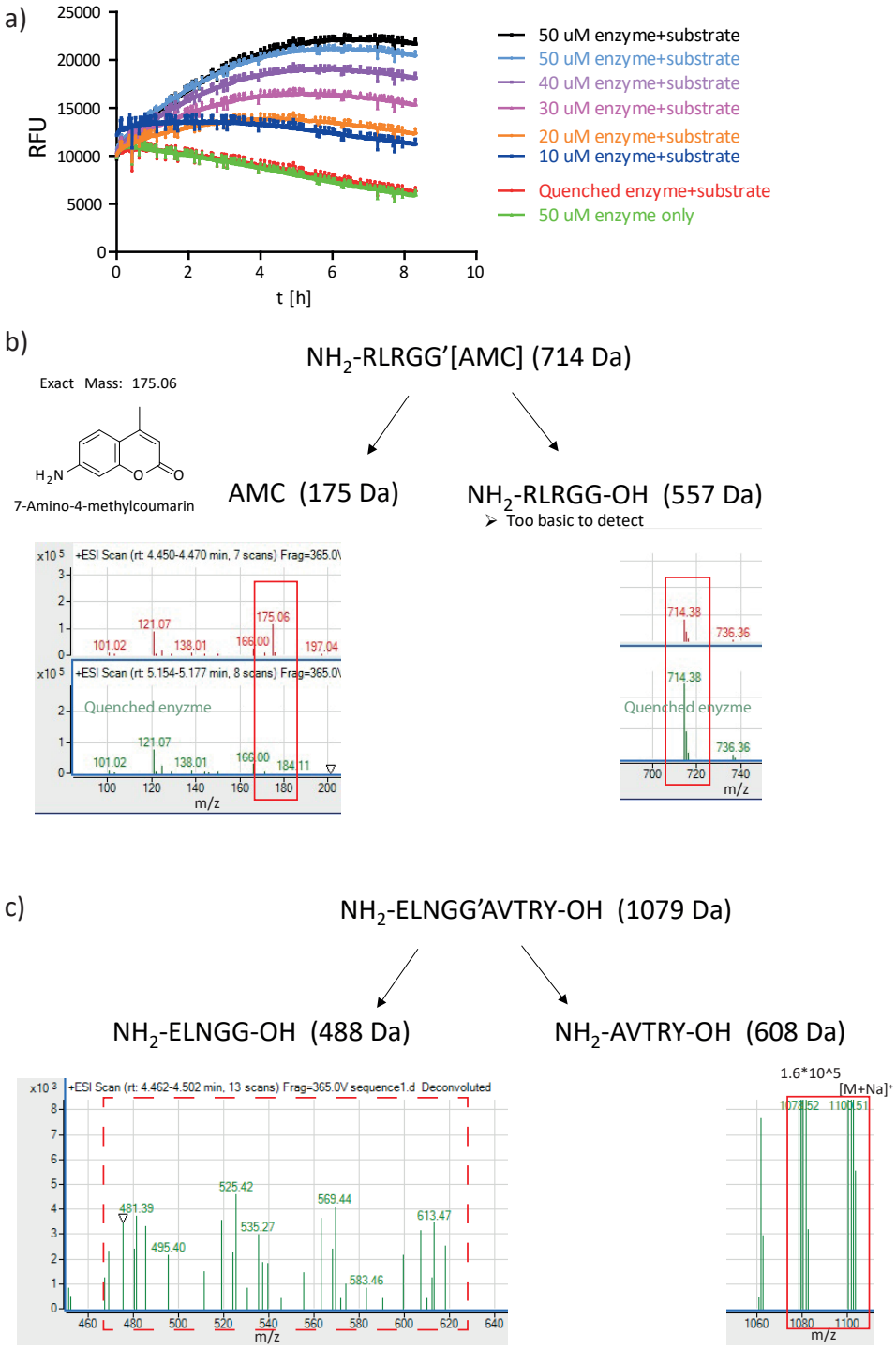


Figure 5.6 | Summary of CV19 PLPro work. **a)** Fluorescence assay results showing cleavage of NH_2-RLRGG[AMC]. **b)** SPE-MS data confirming the cleavage of the fluorescent substrate, but not with quenched enzyme (10% (v/v) of 10% AcOH, lower panel in green). Note the peptide fragment NH_2-RLRGG-OH is too basic to detect under SPE-MS conditions. **c)** SPE-MS data showing that no cleavage was observed for the non-fluorescent substrate NH_2-ELNGG'AVTRY-OH, as indicated by a high abundance of intact peptide ($1.6*10^5$), while none of the proposed cleavage products was observed. Assay conditions were: 50 mM Tris pH 7.9, 200 mM NaCl, 2 mM DTT, >4 h incubation at 37 °C, 200 µM PLPro, 2.5 µM peptide (1 µM for the fluorescence assay).

Pfizer's nirmatrelvir, which is an orally bioavailable M^{pro} inhibitor, is currently available under emergency use authorization by the Food and Drug Administration for patients running risk for progression to severe cases of SARS-CoV-2.[328] Nirmatrelvir inhibits M^{pro} with an IC_{50} of 19.2 nM.[329]

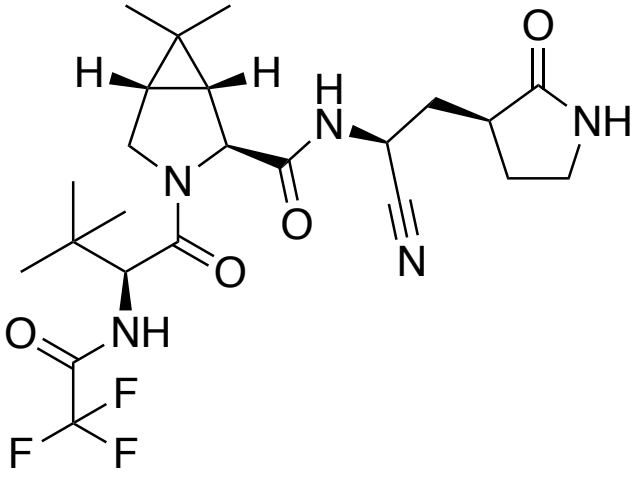


Figure 5.7 | Structure of nirmatrelvir.

Structure-activity studies of various scaffolds, including the covalent inhibitor ebselen, with M^{pro} are ongoing in the Schofield group.[330] Further, the SPE-MS assay that was developed for M^{pro} has been employed for studies on PLPro inhibition.[331]

Bibliography

[1] R. P. Feynman and R. Leighton, *"What do you care what other people think?": further adventures of a curious character*. WW Norton & Company, 2001.

[2] T. Mann, *Der Zauberberg. Berlin: S*, 1926.

[3] T. Mann, H. T. Lowe-Porter, *et al.*, *The magic mountain*. Modern library, 1927.

[4] H. M. Leicester, "Alexander Mikhailovich Butlerov," *Journal of chemical education*, vol. 17, no. 5, p. 203, 1940. DOI: 10.1021/ed017p203.

[5] A. Hofmann, "Zur Kenntniss des Methylaldehyds," *Berichte der deutschen chemischen Gesellschaft*, vol. 2, no. 1, pp. 152–159, 1869. DOI: 10.1002/jlac.18681450312.

[6] R. Köhler and W. Enzensberger, "Melaminharze in der Holztechnik," *Holz als Roh-und Werkstoff*, vol. 10, no. 2, pp. 51–56, 1952. DOI: 10.1007/BF02605555.

[7] T. Salthammer, S. Mentese, and R. Marutzky, "Formaldehyde in the indoor environment," *Chemical reviews*, vol. 110, no. 4, pp. 2536–2572, 2010. DOI: 10.1021/cr800399g.

[8] C. D. Cortez, F. C. Bato, T. J. G. Bautista, J. M. G. Cantor, C. L. Gandionco III, and S. P. Reyes, "Development of formaldehyde detector," *International Journal of Information and Electronics Engineering*, vol. 5, no. 5, p. 385, 2015. DOI: 10.7763/IJIEE.2015.V5.564.

[9] J. A. Swenberg, W. D. Kerns, R. I. Mitchell, E. J. Gralla, and K. L. Pavkov, "Induction of squamous cell carcinomas of the rat nasal cavity by inhalation exposure to formaldehyde vapor," *Cancer research*, vol. 40, no. 9, pp. 3398–3402, 1980.

[10] M. Bohnet, *Ullmann's encyclopedia of industrial chemistry*. Wiley-Vch, 2003.

[11] R. E. Madison, A. Broughton, and J. D. Thrasher, "Immunologic biomarkers associated with an acute exposure to exothermic byproducts of a ureaformaldehyde spill," *Environmental health perspectives*, vol. 94, pp. 219–223, 1991. DOI: 10.1289/ehp.94-1567939.

[12] L. Zhang, *Formaldehyde: Exposure, Toxicity and Health Effects*. Royal Society of Chemistry, 2018, vol. 37.

[13] C. Babington, "Fema slow to test toxicity of trailers," *USA Today*, vol. 21, 2007.

[14] M. Hauptmann *et al.*, "Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde," *Journal of the National Cancer Institute*, vol. 101, no. 24, pp. 1696–1708, 2009. DOI: 10.1093/jnci/djp416.

[15] A. Reddam and D. C. Volz, "Inhalation of two Prop 65-listed chemicals within vehicles may be associated with increased cancer risk," *Environment International*, vol. 149, p. 106 402, 2021. DOI: 10.1016/j.envint.2021.106402.

[16] J. M. Hollis, F. J. Lovas, and P. R. Jewell, "Interstellar glycolaldehyde: The first sugar," *The Astrophysical Journal Letters*, vol. 540, no. 2, p. L107, 2000.

[17] L. E. Snyder, D. Buhl, B. Zuckerman, and P. Palmer, "Microwave detection of interstellar formaldehyde," *Physical Review Letters*, vol. 22, no. 13, p. 679, 1969. DOI: 10.1103/PhysRevLett.22.679.

[18] A. K. Eckhardt, M. M. Linden, R. C. Wende, B. Bernhardt, and P. R. Schreiner, "Gas-phase sugar formation using hydroxymethylene as the reactive formaldehyde isomer," *Nature Chemistry*, vol. 10, no. 11, pp. 1141–1147, 2018. DOI: 10.1038/s41557-018-0128-2.

[19] C. Appayee and R. Breslow, "Deuterium studies reveal a new mechanism for the formose reaction involving hydride shifts," *Journal of the American Chemical Society*, vol. 136, no. 10, pp. 3720–3723, 2014. DOI: 10.1021/ja410886c.

[20] D. Kopetzki and M. Antonietti, "Hydrothermal formose reaction," *New Journal of Chemistry*, vol. 35, no. 9, pp. 1787–1794, 2011. DOI: 10.1039/C1NJ20191C.

[21] P. R. Schreiner *et al.*, "Capture of hydroxymethylene and its fast disappearance through tunnelling," *Nature*, vol. 453, no. 7197, pp. 906–909, 2008. DOI: 10.1038/nature07010.

[22] M. Räsänen, "Cool it, baby," *Nature*, vol. 453, no. 7197, pp. 862–863, 2008. DOI: 10.1038/453862a.

[23] A. M. Bahmanpour, A. Hoadley, and A. Tanksale, "Critical review and exergy analysis of formaldehyde production processes," *Reviews in Chemical Engineering*, vol. 30, no. 6, pp. 583–604, 2014. DOI: 10.1515/revce-2014-0022.

[24] T. Sugita, H. Ishiwata, and K. Yoshihira, "Release of formaldehyde and melamine from tableware made of melamine—formaldehyde resin," *Food Additives & Contaminants*, vol. 7, no. 1, pp. 21–27, 1990. DOI: 10.1080/02652039009373815.

[25] Römpp, "Thieme Römpp Online. Lexikon," 2017.

[26] D. Wang, X. Zhang, S. Luo, S. Li, *et al.*, "Preparation and property analysis of melamine formaldehyde foam," *Adv Mater Phys Chem*, vol. 2, no. 4, pp. 63–67, 2012. DOI: 10.4236/ampc.2012.24B018.

[27] L. E. Heim, H. Konnerth, and M. H. Prechtl, "Future perspectives for formaldehyde: Pathways for reductive synthesis and energy storage," *Green Chemistry*, vol. 19, no. 10, pp. 2347–2355, 2017. DOI: 10.1039/C6GC03093A.

[28] G. J. Millar and M. Collins, "Industrial production of formaldehyde using polycrystalline silver catalyst," *Industrial & Engineering Chemistry Research*, vol. 56, no. 33, pp. 9247–9265, 2017. DOI: 10.1021/acs.iecr.7b02388.

[29] M. Qian, M. Liauw, and G. Emig, "Formaldehyde synthesis from methanol over silver catalysts," *Applied Catalysis A: General*, vol. 238, no. 2, pp. 211–222, 2003. DOI: 10.1016/S0926-860X(02)00340-X.

[30] M. Rellán-Piñeiro and N. López, "The active molybdenum oxide phase in the methanol oxidation to formaldehyde (Formox process): A DFT study," *ChemSusChem*, vol. 8, no. 13, pp. 2231–2239, 2015. DOI: 10.1002/cssc.201500315.

[31] D. R. Lide, *CRC handbook of chemistry and physics*. CRC press, 2004, vol. 85.

[32] K. B. Wiberg and Y. Wang, "A comparison of some properties of C= O and C= S bonds," *Arkivoc*, vol. 45, p. 56, 2011. DOI: 10.3998/ark.5550190.0012.506.

[33] K. P. C. Vollhardt and N. E. Schore, *Organische chemie*. John Wiley & Sons, 2020.

[34] J. J. Kamps, "Biophysical studies on enzymes related to antibiotics," Ph.D. dissertation, University of Oxford, 2020.

[35] J. Dunitz, J. Lehn, G. Wipff, *et al.*, "Stereochemistry of reaction paths at carbonyl centres," *Tetrahedron*, vol. 30, no. 12, pp. 1563–1572, 1974. DOI: 10.1016/S0040-4020(01)90678-7.

[36] W. Reusch, "Professor emeritus (michigan state u. virtual)," *Textbook of Organic Chemistry*,

[37] K. B. Wiberg, K. M. Morgan, and H. Maltz, "Thermochemistry of carbonyl reactions. 6. a study of hydration equilibria," *Journal of the American Chemical Society*, vol. 116, no. 24, pp. 11 067–11 077, 1994. DOI: 10.1021/ja00103a024.

[38] J. S. Wiseman and R. H. Abeles, "Mechanism of inhibition of aldehyde dehydrogenase by cyclopropanone hydrate and the mushroom toxin coprine," *Biochemistry*, vol. 18, no. 3, pp. 427–435, 1979. DOI: 10.1021/bi00570a006.

[39] H. Marchner and O. Tottmar, "Studies in vitro on the inactivation of mitochondrial rat-liver aldehyde dehydrogenase by the alcohol-sensitizing compounds cyanamide, 1-aminocyclopropanol and disulfiram," *Biochemical pharmacology*, vol. 32, no. 14, pp. 2181–2188, 1983. DOI: 10.1016/0006-2952(83)90224-1.

[40] D. Michelot, "Poisoning by Coprinus atramentarius," *Natural toxins*, vol. 1, no. 2, pp. 73–80, 1992. DOI: 10.1002/nt.2620010203.

[41] W. Eschweiler, "Ersatz von an Stickstoff gebundenen Wasserstoffatomen durch die Methylgruppe mit Hülfe von Formaldehyd," *Berichte der deutschen chemischen Gesellschaft*, vol. 38, no. 1, pp. 880–882, 1905. DOI: 10.1002/cber.190503801154.

[42] C. Mannich and W. Krösche, "Ueber ein Kondensationsprodukt aus Formaldehyd, Ammoniak und Antipyrin," *Archiv der Pharmazie*, vol. 250, no. 1, pp. 647–667, 1912. DOI: 10.1002/ardp.19122500151.

[43] S. Cannizzaro, "Ueber den der Benzoësäure entsprechenden Alkohol," *Justus Liebigs Annalen der Chemie*, vol. 88, no. 1, pp. 129–130, 1853. DOI: 10.1002/jlac.18530880114.

[44] G. L. Blanc, "Preparation of aromatic chloromethylenic derivatives," *Bull. soc. chim. France*, vol. 33, p. 313, 1923.

[45] A. M. Kirillov, "Hexamethylenetetramine: An old new building block for design of coordination polymers," *Coordination Chemistry Reviews*, vol. 255, no. 15-16, pp. 1603–1622, 2011. DOI: 10.1016/j.ccr.2011.01.023.

[46] A. Butlerow, "Ueber ein neues Methylenderivat," *Justus Liebigs Annalen der Chemie*, vol. 115, no. 3, pp. 322–327, 1860.

[47] J. Duff and E. Bills, "273. reactions between hexamethylenetetramine and phenolic compounds. Part I. A new method for the preparation of 3-and 5-aldehydosalicylic acids," *Journal of the Chemical Society (Resumed)*, pp. 1987–1988, 1932. DOI: 10.1039/JR9320001987.

[48] H. d. Heck, M. Casanova-Schmitz, P. B. Dodd, E. N. Schachter, T. J. Witek, and T. Tosun, "Formaldehyde (CH_2O) concentrations in the blood of humans and Fischer-344 rats exposed to (CH_2O) under controlled conditions," *American Industrial Hygiene Association Journal*, vol. 46, no. 1, pp. 1–3, 1985. DOI: 10.1080/15298668591394275.

[49] T. C. Cascieri and J. J. Clary, "Formaldehyde– oral toxicity assessment," *Comments on Toxicology*, vol. 4, no. 4, pp. 295–304, 1992.

[50] B. Owen, C. Dudney, E. Tan, and C. Easterly, "Formaldehyde in drinking water: Comparative hazard evaluation and an approach to regulation," *Regulatory toxicology and pharmacology*, vol. 11, no. 3, pp. 220–236, 1990. DOI: 10.1016/0273-2300(90)90023-5.

[51] B. Cancho, F. Ventura, and M. T. Galceran, "Determination of aldehydes in drinking water using pentafluorobenzylhydroxylamine derivatization and solid-phase microextraction," *Journal of Chromatography A*, vol. 943, no. 1, pp. 1–13, 2002. DOI: 10.1016/S0021-9673(01)01437-6.

[52] F. A. Dingler *et al.*, "Two aldehyde clearance systems are essential to prevent lethal formaldehyde accumulation in mice and humans," *Molecular cell*, vol. 80, no. 6, pp. 996–1012, 2020. DOI: 10.1016/j.molcel.2020.10.012.

[53] R. Dandriyal, K. Y. Giri, S. Alam, and A. P. Singh, "Accidental intraoral formalin injection: A rare case report," *Clinics and practice*, vol. 4, no. 3, pp. 60–63, 2014. DOI: 10.4081/cp.2014.686.

[54] P. C. Swami, R. Raval, M. Kaur, and J. Kaur, "Accidental intraoral injection of formalin during extraction: Case report," *British Journal of Oral and Maxillofacial Surgery*, vol. 54, no. 3, pp. 351–352, 2016. DOI: 10.1016/j.bjoms.2015.09.016.

[55] A. Smedra-Kaźmirska, L. Zydek, M. Barzdo, W. Machała, and J. Berent, "Accidental intravenous injection of formaldehyde," *Anestezjologia intensywna terapia*, vol. 41, no. 3, pp. 163–165, 2009.

[56] E. F. S. Authority, "Endogenous formaldehyde turnover in humans compared with exogenous contribution from food sources," *EFSA Journal*, vol. 12, no. 2, p. 3550, 2014. DOI: 10.2903/j.efsa.2014.3550.

[57] S. S. Dhareshwar and V. J. Stella, "Your prodrug releases formaldehyde: Should you be concerned? No!" *Journal of pharmaceutical sciences*, vol. 97, no. 10, pp. 4184–4193, 2008. DOI: 10.1002/jps.21319.

[58] A. C. De Groot, M.-A. Flyvholm, G. Lensen, T. Menné, and P.-J. Coenraads, "Formaldehyde-releasers: Relationship to formaldehyde contact allergy. contact allergy to formaldehyde and inventory of formaldehyde-releasers," *Contact dermatitis*, vol. 61, no. 2, pp. 63–85, 2009. DOI: 10.1111/j.1600-0536.2009.01582.x.

[59] I. M. Fasth, N. H. Ulrich, and J. D. Johansen, "Ten-year trends in contact allergy to formaldehyde and formaldehyde-releasers," *Contact dermatitis*, vol. 79, no. 5, pp. 263–269, 2018. DOI: 10.1111/cod.13052.

[60] K. Aalto-Korte, O. Kuuliala, K. u. Suuronen, and K. Alanko, "Occupational contact allergy to formaldehyde and formaldehyde releasers," *Contact Dermatitis*, vol. 59, no. 5, pp. 280–289, 2008. DOI: 10.1111/j.1600-0536.2008.01422.x.

[61] J. P. Thyssen, J. D. Johansen, and T. Menné, "Contact allergy epidemics and their controls," *Contact Dermatitis*, vol. 56, no. 4, pp. 185–195, 2007. DOI: 10.1111/j.1600-0536.2006.01058.x.

[62] T. Calheiros, J. Iley, F. Lopes, and R. Moreira, "Acyloxymethyl as a drug protecting group. synthesis and reactivity of N-acyloxymethylsulfonamide prodrugs," *Bioorganic & Medicinal Chemistry Letters*, vol. 5, no. 9, pp. 937–940, 1995. DOI: 10.1016/0960-894X(95)00146-K.

[63] F. Lopes, R. Moreira, and J. Iley, "Acyloxymethyl as a drug protecting group. part 6: N-acyloxymethyl-and N-[(aminocarbonyloxy) methyl] sulfonamides as prodrugs of agents containing a secondary sulfonamide group," *Bioorganic & medicinal chemistry*, vol. 8, no. 4, pp. 707–716, 2000. DOI: 10.1016/S0968-0896(00)00015-8.

[64] A. C. De Groot, I. R. White, M.-A. Flyvholm, G. Lensen, and P.-J. Coenraads, "Formaldehyde-releasers in cosmetics: Relationship to formaldehyde contact allergy: Part 1. characterization, frequency and relevance of sensitization, and frequency of use in cosmetics," *Contact Dermatitis*, vol. 62, no. 1, pp. 2–17, 2010. DOI: 10.1111/j.1600-0536.2009.01615.x.

[65] J. H. Brown, "The formol titration of bacteriological media," *Journal of Bacteriology*, vol. 8, no. 3, p. 245, 1923. DOI: 10.1128/jb.8.3.245-267.1923.

[66] H. Clarke, H. Gillespie, and S. Weisshaus, "The action of formaldehyde on amines and amino acids," *Journal of the American Chemical Society*, vol. 55, no. 11, pp. 4571–4587, 1933. DOI: 10.1021/ja01338a041.

[67] A. Wadsworth and M. Pangborn, "Formaldehyde reaction with amino acids," *J. Biol. Chem*, vol. 116, pp. 423–436, 1936.

[68] A. K. Smith, P. Handler, and J. Mrgudich, "The reaction of formaldehyde with amino acids: X-ray diffraction patterns," *The Journal of Physical Chemistry*, vol. 44, no. 7, pp. 874–880, 1940. DOI: 10.1021/j150403a006.

[69] W. R. Middlebrook and H. Phillips, "The action of formaldehyde on the cystine disulphide linkages in wool: The subdivision of the combined cystine into two fractions differing in their reactivity towards formaldehyde," *Biochemical Journal*, vol. 36, no. 3-4, pp. 294–302, 1942. DOI: 10.1042/bj0360294.

[70] D. French and J. T. Edsall, "The reactions of formaldehyde with amino acids and proteins," in *Advances in protein chemistry*, vol. 2, Elsevier, 1945, pp. 277–335. DOI: 10.1016/S0065-3233(08)60627-0.

[71] H. Fraenkel-Conrat and H. S. Olcott, "Reaction of formaldehyde with proteins. II. Participation of the guanidyl groups and evidence of crosslinking," *Journal of the American Chemical Society*, vol. 68, no. 1, pp. 34–37, 1946. DOI: 10.1021/ja01205a011.

[72] H. Fraenkel-Conrat and H. S. Olcott, "The reaction of formaldehyde with proteins. V. Cross-linking between amino and primary amide or guanidyl groups," *Journal of the American Chemical Society*, vol. 70, no. 8, pp. 2673–2684, 1948. DOI: 10.1021/ja01188a018.

[73] W. Middlebrook, "The irreversible combination of formaldehyde with proteins," *Biochemical Journal*, vol. 44, no. 1, pp. 17–23, 1949. DOI: 10.1042/bj0440017.

[74] K. F. Guthe, "The formaldehyde-hemoglobin reaction," *Journal of Biological Chemistry*, vol. 234, no. 12, pp. 3169–3173, 1959. DOI: 10.1016/S0021-9258(18)69642-5.

[75] J. Myers and J. Hardman, "Formaldehyde-induced cross-linkages in the subunit of the Escherichia coli tryptophan synthetase," *J. Biol. Chem*, vol. 246, pp. 3863–3869, 1971. DOI: 10.1016/S0021-9258(18)62114-3.

[76] K. Takahashi, "The reactions of phenylglyoxal and related reagents with amino acids," *The Journal of Biochemistry*, vol. 81, no. 2, pp. 395–402, 1977. DOI: 10.1093/oxfordjournals.jbchem.a131471.

[77] Y. Kitamoto and H. Maeda, "Reevaluation of the reaction of formaldehyde at low concentration with amino acids," *The Journal of Biochemistry*, vol. 87, no. 5, pp. 1519–1530, 1980. DOI: 10.1093/oxfordjournals.jbchem.a132893.

[78] D. Tome, N. Naulet, and A. Kozlowski, "^{13}C-NMR characterization of formaldehyde bonds in model mixtures and proteins containing lysine," *International Journal of Peptide and Protein Research*, vol. 25, no. 3, pp. 258–266, 1985. DOI: 10.1111/j.1399-3011.1985.tb02173.x.

[79] N. S. Joshi, L. R. Whitaker, and M. B. Francis, "A three-component Mannich-type reaction for selective tyrosine bioconjugation," *Journal of the American Chemical Society*, vol. 126, no. 49, pp. 15 942–15 943, 2004. DOI: 10.1021/ja0439017.

[80] J. D. McGhee and P. H. Von Hippel, "Formaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA," *Biochemistry*, vol. 16, no. 15, pp. 3276–3293, 1977. DOI: 10.1021/bi00634a002.

[81] L. B. Pontel *et al.*, "Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen," *Molecular cell*, vol. 60, no. 1, pp. 177–188, 2015. DOI: 10.1016/j.molcel.2015.08.020.

[82] J. J. Kamps, R. J. Hopkinson, C. J. Schofield, and T. D. Claridge, "How formaldehyde reacts with amino acids," *Communications Chemistry*, vol. 2, no. 1, pp. 1–14, 2019. DOI: 10.1038/s42004-019-0224-2.

[83] S. Shishodia *et al.*, "NMR analyses on N-hydroxymethylated nucleobases–implications for formaldehyde toxicity and nucleic acid demethylases," *Organic & biomolecular chemistry*, vol. 16, no. 21, pp. 4021–4032, 2018. DOI: 10.1039/C8OB00734A.

[84] R. J. Hopkinson, P. S. Barlow, C. J. Schofield, and T. D. Claridge, "Studies on the reaction of glutathione and formaldehyde using NMR," *Organic & biomolecular chemistry*, vol. 8, no. 21, pp. 4915–4920, 2010. DOI: 10.1039/C0OB00208A.

[85] K. Lu *et al.*, "Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers," *Journal of the American Chemical Society*, vol. 132, no. 10, pp. 3388–3399, 2010. DOI: 10.1021/ja908282f.

[86] M. Pietzke *et al.*, "Amino acid dependent formaldehyde metabolism in mammals," *Communications Chemistry*, vol. 3, no. 1, pp. 1–10, 2020. DOI: 10.1038/s42004-020-0324-z.

[87] R. Bateman, D. Rauh, and K. M. Shokat, "Glutathione traps formaldehyde by formation of a bicyclo [4.4.1] undecane adduct," *Organic & biomolecular chemistry*, vol. 5, no. 20, pp. 3363–3367, 2007. DOI: 10.1039/B707602A.

[88] T. J. Michiels *et al.*, "Novel formaldehyde-induced modifications of lysine residue pairs in peptides and proteins: Identification and relevance to vaccine development," *Molecular Pharmaceutics*, vol. 17, no. 11, pp. 4375–4385, 2020. DOI: 10.1021/acs.molpharmaceut.0c00851.

[89] B. Metz *et al.*, "Identification of formaldehyde-induced modifications in proteins reactions with model peptides," *Journal of Biological Chemistry*, vol. 279, no. 8, pp. 6235–6243, 2004. DOI: 10.1074/jbc.M310752200.

[90] J. D. McGhee and P. H. Von Hippel, "Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases," *Biochemistry*, vol. 14, no. 6, pp. 1281–1296, 1975. DOI: 10.1021/bi00677a029.

[91] J. D. McGhee and P. H. Von Hippel, "Formaldehyde as a probe of DNA structure. II. Reaction with endocyclic imino groups of DNA bases," *Biochemistry*, vol. 14, no. 6, pp. 1297–1303, 1975. DOI: 10.1021/bi00677a030.

[92] J. Kennedy-Darling and L. M. Smith, "Measuring the formaldehyde protein–DNA cross-link reversal rate," *Analytical chemistry*, vol. 86, no. 12, pp. 5678–5681, 2014. DOI: 10.1021/ac501354y.

[93] J. G. Donaldson, "Immunofluorescence staining," *Current protocols in cell biology*, vol. 69, no. 1, pp. 4–3, 2015. DOI: 10.1002/0471143030.cb0403s69.

[94] Y. Fu *et al.*, "FTO-mediated formation of N^6-hydroxymethyladenosine and N^6-formyladenosine in mammalian RNA," *Nature communications*, vol. 4, no. 1, pp. 1–8, 2013. DOI: 10.1038/ncomms2822.

[95] J. H. Bedino, "Embalming chemistry: Glutaraldehyde versus formaldehyde," *Champion Expanding Encyclopedia of Mortuary Practices*, vol. 649, pp. 2614–2632, 2003.

[96] S. L. Jagger, M. M. Dubek, and E. Pedretti, "'It's a personal thing': Visitors' responses to body worlds," *Museum Management and Curatorship*, vol. 27, no. 4, pp. 357–374, 2012. DOI: 10.1080/09647775.2012.720185.

[97] S. Cadirci *et al.*, "Disinfection of hatching eggs by formaldehyde fumigation–a review," *Arch. Geflügelk*, vol. 73, no. 2, pp. 116–123, 2009.

[98] N. Ackland, M. Hinton, and K. Denmeade, "Controlled formaldehyde fumigation system," *Applied and Environmental Microbiology*, vol. 39, no. 3, pp. 480–487, 1980. DOI: 10.1128/aem.39.3.480-487.1980.

[99] A. Glenny and H. Südmersen, "Notes on the production of immunity to diphtheria toxin," *Epidemiology & Infection*, vol. 20, no. 2, pp. 176–220, 1921. DOI: 10.1017/S0022172400033945.

[100] A. Glenny, B. E. Hopkins, and C. Pope, "Further notes on modification of diphtheria toxin by formaldehyde," *The Journal of Pathology and Bacteriology*, vol. 27, no. 3, pp. 261–270, 1924. DOI: 10.1002/path.1700270309.

[101] T. Wilton, G. Dunn, D. Eastwood, P. D. Minor, and J. Martin, "Effect of formaldehyde inactivation on poliovirus," *Journal of virology*, vol. 88, no. 20, pp. 11 955–11 964, 2014. DOI: 10.1128/JVI.01809-14.

[102] W. Dias *et al.*, "An improved whole cell pertussis vaccine with reduced content of endotoxin," *Human vaccines & immunotherapeutics*, vol. 9, no. 2, pp. 339–348, 2013. DOI: 10.4161/hv.22847.

[103] Z. Reed and M. J. Cardosa, "Status of research and development of vaccines for enterovirus 71," *Vaccine*, vol. 34, no. 26, pp. 2967–2970, 2016. DOI: 10.1016/j.vaccine.2016.02.077.

[104] V. Stone *et al.*, "A hexavalent Coxsackievirus B vaccine is highly immunogenic and has a strong protective capacity in mice and nonhuman primates," *Science Advances*, vol. 6, no. 19, eaaz2433, 2020. DOI: 10.1126/sciadv.aaz2433.

[105] M. E. Darnell *et al.*, "Severe acute respiratory syndrome coronavirus infection in vaccinated ferrets," *The Journal of infectious diseases*, vol. 196, no. 9, pp. 1329–1338, 2007. DOI: 10.1086/522431.

[106] P. J. Openshaw, F. J. Culley, and W. Olszewska, "Immunopathogenesis of vaccine-enhanced RSV disease," *Vaccine*, vol. 20, S27–S31, 2001. DOI: 10.1016/S0264-410X(01)00301-2.

[107] A. M. Killikelly, M. Kanekiyo, and B. S. Graham, "Pre-fusion F is absent on the surface of formalin-inactivated respiratory syncytial virus," *Scientific reports*, vol. 6, p. 34 108, 2016. DOI: 10.1038/srep34108.

[108] W. Zhang *et al.*, "The optimal concentration of formaldehyde is key to stabilizing the pre-fusion conformation of respiratory syncytial virus fusion protein," *Viruses*, vol. 11, no. 7, p. 628, 2019. DOI: 10.3390/v11070628.

[109] F. Brown, "Review of accidents caused by incomplete inactivation of viruses," *Developments in biological standardization*, vol. 81, pp. 103–107, 1993.

[110] T. Twomey, J. Newman, T. Burrage, P. Piatti, J. Lubroth, and F. Brown, "Structure and immunogenicity of experimental foot-and-mouth disease and poliomyelitis vaccines," *Vaccine*, vol. 13, no. 16, pp. 1603–1610, 1995. DOI: 10.1016/0264-410X(95)00079-G.

[111] A. Kata, "A postmodern Pandora's box: Anti-vaccination misinformation on the internet," *Vaccine*, vol. 28, no. 7, pp. 1709–1716, 2010. DOI: 10.1016/j.vaccine.2009.12.022.

[112] R. J. Mitkus, M. A. Hess, and S. L. Schwartz, "Pharmacokinetic modeling as an approach to assessing the safety of residual formaldehyde in infant vaccines," *Vaccine*, vol. 31, no. 25, pp. 2738–2743, 2013. DOI: 10.1016/j.vaccine.2013.03.071.

[113] J. Toews, J. C. Rogalski, and J. Kast, "Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications," *Analytica chimica acta*, vol. 676, no. 1-2, pp. 60–67, 2010. DOI: 10.1016/j.aca.2010.07.040.

[114] B. Metz *et al.*, "Identification of formaldehyde-induced modifications in diphtheria toxin," *Journal of Pharmaceutical Sciences*, vol. 109, no. 1, pp. 543–557, 2020. DOI: 10.1016/j.xphs.2019.10.047.

[115] X. Shen *et al.*, "A surge of DNA damage links transcriptional reprogramming and hematopoietic deficit in Fanconi anemia," *Molecular cell*, vol. 80, no. 6, pp. 1013–1024, 2020. DOI: 10.1016/j.molcel.2020.11.040.

[116] J. Wei *et al.*, "Differential m6a, m6am, and m1a demethylation mediated by FTO in the cell nucleus and cytoplasm," *Molecular cell*, vol. 71, no. 6, pp. 973–985, 2018. DOI: 10.1016/j.molcel.2018.08.011.

[117] L. J. Walport *et al.*, "Mechanistic and structural studies of KDM-catalysed demethylation of histone 1 isotype 4 at lysine 26," *FEBS letters*, vol. 592, no. 19, pp. 3264–3273, 2018. DOI: 10.1002/1873-3468.13231.

[118] E. Flashman, L. M. Hoffart, R. B. Hamed, J. M. Bollinger Jr, C. Krebs, and C. J. Schofield, "Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen," *The FEBS journal*, vol. 277, no. 19, pp. 4089–4099, 2010. DOI: 10.1111/j.1742-4658.2010.07804.x.

[119] X. Kong *et al.*, "Catalytic mechanism investigation of lysine-specific demethylase 1 (LSD1): A computational study," *PLoS One*, vol. 6, no. 9, e25444, 2011. DOI: 10.1371/journal.pone.0025444.

[120] J. Liang, Q. Han, Y. Tan, H. Ding, and J. Li, "Current advances on structure-function relationships of pyridoxal 5'-phosphate-dependent enzymes," *Frontiers in molecular biosciences*, vol. 6, p. 4, 2019. DOI: 10.3389/fmolb.2019.00004.

[121] T. Tralau, P. Lafite, C. Levy, J. P. Combe, N. S. Scrutton, and D. Leys, "An internal reaction chamber in dimethylglycine oxidase provides efficient protection from exposure to toxic formaldehyde," *Journal of Biological Chemistry*, vol. 284, no. 26, pp. 17 826–17 834, 2009. DOI: 10.1074/jbc.M109.006262.

[122] M. Mure, "Tyrosine-derived quinone cofactors," *Accounts of chemical research*, vol. 37, no. 2, pp. 131–139, 2004. DOI: 10.1021/ar9703342.

[123] J. S. Foot *et al.*, "PXS-4681A, a potent and selective mechanism-based inhibitor of SSAO/VAP-1 with anti-inflammatory effects in vivo," *Journal of Pharmacology and Experimental Therapeutics*, vol. 347, no. 2, pp. 365–374, 2013. DOI: 10.1124/jpet.113.207613.

[124] H. Y. Peter *et al.*, "Involvement of semicarbazide-sensitive amine oxidase-mediated deamination in lipopolysaccharide-induced pulmonary inflammation," *The American journal of pathology*, vol. 168, no. 3, pp. 718–726, 2006. DOI: 10.2353/ajpath.2006.050970.

[125] H. Yamanaka *et al.*, "CYP2A6 and CYP2B6 are involved in nornicotine formation from nicotine in humans: Interindividual differences in these contributions," *Drug metabolism and disposition*, vol. 33, no. 12, pp. 1811–1818, 2005. DOI: 10.1124/dmd.105.006254.

[126] R. A. Kwiecień, R. Molinié, P. Paneth, V. Silvestre, J. Lebreton, and R. J. Robins, "Elucidation of the mechanism of N-demethylation catalyzed by cytochrome P450 monooxygenase is facilitated by exploiting nitrogen-15 heavy isotope effects," *Archives of biochemistry and biophysics*, vol. 510, no. 1, pp. 35–41, 2011. DOI: 10.1016/j.abb.2011.03.007.

[127] N. L. Benowitz, P. Jacob III, H. Mayan, and C. Denaro, "Sympathomimetic effects of paraxanthine and caffeine in humans," *Clinical Pharmacology & Therapeutics*, vol. 58, no. 6, pp. 684–691, 1995. DOI: 10.1016/0009-9236(95)90025-X.

[128] J. Tao *et al.*, "Theoretical study on the N-demethylation mechanism of theobromine catalyzed by P450 isoenzyme 1A2," *Journal of Molecular Graphics and Modelling*, vol. 61, pp. 123–132, 2015. DOI: 10.1016/j.jmgm.2015.06.017.

[129] M. Pietzke, J. Meiser, and A. Vazquez, "Formate metabolism in health and disease," *Molecular metabolism*, vol. 33, pp. 23–37, 2020. DOI: 10.1016/j.molmet.2019.05.012.

[130] N. H. Chen, K. Y. Djoko, F. J. Veyrier, and A. G. McEwan, "Formaldehyde stress responses in bacterial pathogens," *Frontiers in microbiology*, vol. 7, p. 257, 2016. DOI: 10.3389/fmicb.2016.00257.

[131] M. Y. Feldman, "Reactions of nucleic acids and nucleodroteins with formaldehyde," *Progress in nucleic acid research and molecular biology*, vol. 13, pp. 1–49, 1973. DOI: 10.1016/S0079-6603(08)60099-9.

[132] C. C. Conaway, J. Whysner, L. K. Verna, and G. M. Williams, "Formaldehyde mechanistic data and risk assessment: Endogenous protection from DNA adduct formation," *Pharmacology & therapeutics*, vol. 71, no. 1-2, pp. 29–55, 1996. DOI: 10.1016/0163-7258(96)00061-7.

[133] K. Lu, L. B. Collins, H. Ru, E. Bermudez, and J. A. Swenberg, "Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia," *Toxicological sciences*, vol. 116, no. 2, pp. 441–451, 2010. DOI: 10.1093/toxsci/kfq061.

[134] M. J. Solomon and A. Varshavsky, "Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures," *Proceedings of the National Academy of Sciences*, vol. 82, no. 19, pp. 6470–6474, 1985. DOI: 10.1073/pnas.82.19.6470.

[135] C. Loshon, P. Genest, B. Setlow, and P. Setlow, "Formaldehyde kills spores of bacillus subtilis by DNA damage and small, acid-soluble spore proteins of the $\alpha\alpha\alpha\alpha\alpha\alpha$/$\beta\beta\beta\beta\beta\beta\beta$-type protect spores against this DNA damage," *Journal of applied microbiology*, vol. 87, no. 1, pp. 8–14, 1999. DOI: 10.1046/j.1365-2672.1999.00783.x.

[136] H. d. Heck and M. Casanova, "The implausibility of leukemia induction by formaldehyde: A critical review of the biological evidence on distant-site toxicity," *Regulatory Toxicology and Pharmacology*, vol. 40, no. 2, pp. 92–106, 2004. DOI: 10.1016/j.yrtph.2004.05.001.

[137] R. Yu *et al.*, "Formation, accumulation, and hydrolysis of endogenous and exogenous formaldehyde-induced DNA damage," *Toxicological sciences*, vol. 146, no. 1, pp. 170–182, 2015. DOI: 10.1093/toxsci/kfv079.

[138] M. Casanova-Schmitz, T. B. Starr, and H. d. Heck, "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," *Toxicology and applied pharmacology*, vol. 76, no. 1, pp. 26–44, 1984. DOI: 10.1016/0041-008X(84)90026-7.

[139] M. Casanova, K. T. Morgan, E. A. Gross, O. R. Moss, and H. d. Heck, "DNA—protein cross-links and cell replication at specific sites in the nose of F344 rats exposed subchronically to formaldehyde," *Toxicological Sciences*, vol. 23, no. 4, pp. 525–536, 1994. DOI: 10.1006/faat.1994.1137.

[140] M. Goenrich, S. Bartoschek, C. H. Hagemeier, C. Griesinger, and J. A. Vorholt, "A glutathione-dependent formaldehyde-activating enzyme (Gfa) from paracoccus denitrificans detected and purified via two-dimensional proton exchange NMR spectroscopy," *Journal of biological chemistry*, vol. 277, no. 5, pp. 3069–3072, 2002. DOI: 10.1074/jbc.C100579200.

[141] L. Uotila and M. Koivusalo, "Purification and properties of S-formylglutathione hydrolase from human liver," *Journal of Biological Chemistry*, vol. 249, no. 23, pp. 7664–7672, 1974. DOI: 10.1016/S0021-9258(19)81288-7.

[142] L. Uotila and M. Koivusalo, "Formaldehyde dehydrogenase from human liver purification, properties, and evidence for the formation of glutathione thiol esters by the enzyme," *Journal of Biological Chemistry*, vol. 249, no. 23, pp. 7653–7663, 1974. DOI: 10.1016/S0021-9258(19)81287-5.

[143] G. L. Newton *et al.*, "Bacillithiol is an antioxidant thiol produced in Bacilli," *Nature chemical biology*, vol. 5, no. 9, pp. 625–627, 2009. DOI: 10.1038/nchembio.189.

[144] S. Sakuda, Z.-Y. Zhou, and Y. Yamada, "Structure of a novel disulfide of 2-(N-acetylcysteinyl)amido-2-deoxy-α-D-glucopyran-osyl-myo-inositol produced by Streptomyces sp," *Bioscience, biotechnology, and biochemistry*, vol. 58, no. 7, pp. 1347–1348, 1994. DOI: 10.1271/bbb.58.1347.

[145] J. A. Vorholt, "Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria," *Archives of microbiology*, vol. 178, no. 4, pp. 239–249, 2002. DOI: 10.1007/s00203-002-0450-2.

[146] X. Chen, S. Y. Chothia, J. Basran, and R. J. Hopkinson, "Formaldehyde regulates tetrahydrofolate stability and thymidylate synthase catalysis," *Chemical Communications*, 2021. DOI: 10.1039/D1CC01425K.

[147] B. E. H. Maden, "Tetrahydrofolate and tetrahydromethanopterin compared: Functionally distinct carriers in C1 metabolism," *Biochemical Journal*, vol. 350, no. 3, pp. 609–629, 2000. DOI: 10.1042/bj3500609.

[148] N. Kato, H. Yurimoto, and R. K. Thauer, "The physiological role of the ribulose monophosphate pathway in bacteria and archaea," *Bioscience, biotechnology, and biochemistry*, vol. 70, no. 1, pp. 10–21, 2006. DOI: 10.1271/bbb.70.10.

[149] Y. Liao *et al.*, "Structure of formaldehyde dehydrogenase from Pseudomonas aeruginosa: The binary complex with the cofactor NAD+," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 69, no. 9, pp. 967–972, 2013. DOI: 10.1107/S174430911302160X.

[150] C. J. Marx, J. A. Miller, L. Chistoserdova, and M. E. Lidstrom, "Multiple formaldehyde oxidation/detoxification pathways in Burkholderia fungorum LB400," *Journal of bacteriology*, vol. 186, no. 7, pp. 2173–2178, 2004. DOI: 10.1128/JB.186.7.2173-2178.2004.

[151] M. Ando, T. Yoshimoto, S. Ogushi, K. Rikitake, S. Shibata, and D. Tsuru, "Formaldehyde dehydrogenase from pseudomonas putida," *The Journal of Biochemistry*, vol. 85, no. 5, pp. 1165–1172, 1979.

[152] K. J. Denby *et al.*, "The mechanism of a formaldehyde-sensing transcriptional regulator," *Scientific reports*, vol. 6, p. 38 879, 2016. DOI: 10.1038/srep38879.

[153] D. Osman *et al.*, "The effectors and sensory sites of formaldehyde-responsive regulator FrmR and metal-sensing variant," *Journal of Biological Chemistry*, vol. 291, no. 37, pp. 19 502–19 516, 2016. DOI: 10.1074/jbc.M116.745174.

[154] H. Yurimoto, R. Hirai, N. Matsuno, H. Yasueda, N. Kato, and Y. Sakai, "HxlR, a member of the DUF24 protein family, is a DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible hxlAB operon in Bacillus subtilis," *Molecular microbiology*, vol. 57, no. 2, pp. 511–519, 2005. DOI: 10.1111/j.1365-2958.2005.04702.x.

[155] R. Zhu *et al.*, "Genetically encoded formaldehyde sensors inspired by a protein intra-helical crosslinking reaction," *Nature communications*, vol. 12, no. 1, pp. 1–13, 2021. DOI: 10.1038/s41467-020-20754-4.

[156] W. Luo, C. Y. Ang, and H. C. Thompson Jr, "Rapid method for the determination of ampicillin residues in animal muscle tissues by high-performance liquid chromatography with fluorescence detection," *Journal of Chromatography B: Biomedical Sciences and Applications*, vol. 694, no. 2, pp. 401–407, 1997. DOI: 10.1016/S0378-4347(97)00171-0.

[157] K. Nagy, F. Pollreisz, Z. Takats, and K. Vekey, "Atmospheric pressure chemical ionization mass spectrometry of aldehydes in biological matrices," *Rapid communications in mass spectrometry*, vol. 18, no. 20, pp. 2473–2478, 2004. DOI: 10.1002/rcm.1648.

[158] A. S. Tibbetts and D. R. Appling, "Compartmentalization of mammalian folate-mediated one-carbon metabolism," *Annual review of nutrition*, vol. 30, pp. 57–81, 2010. DOI: 10.1146/annurev.nutr.012809.104810.

[159] D. W. Horne, "Neither methionine nor nitrous oxide inactivation of methionine synthase affect the concentration of 5, 10-methylenetetrahydrofolate in rat liver," *The Journal of nutrition*, vol. 133, no. 2, pp. 476–478, 2003. DOI: 10.1093/jn/133.2.476.

[160] M. Osborn, P. Talbert, and F. Huennekens, "The structure of "active formaldehyde"(N^5, N^{10}-methylene tetrahydrofolic acid)," *Journal of the American Chemical Society*, vol. 82, no. 18, pp. 4921–4927, 1960. DOI: 10.1021/ja01503a043.

[161] G. Burgos-Barragan *et al.*, "Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism," *Nature*, vol. 548, no. 7669, p. 549, 2017. DOI: 10.1038/nature23481.

[162] P. C. Sanghani, C. L. Stone, B. D. Ray, E. V. Pindel, T. D. Hurley, and W. F. Bosron, "Kinetic mechanism of human glutathione-dependent formaldehyde dehydrogenase," *Biochemistry*, vol. 39, no. 35, pp. 10 720–10 729, 2000. DOI: 10.1021/bi9929711.

[163] J. Nakamura, D. W. Holley, T. Kawamoto, and S. J. Bultman, "The failure of two major formaldehyde catabolism enzymes (ADH5 and ALDH2) leads to partial synthetic lethality in C57BL/6 mice," *Genes and Environment*, vol. 42, pp. 1–6, 2020. DOI: 10.1186/s41021-020-00160-4.

[164] Y. L. Dorokhov, A. V. Shindyapina, E. V. Sheshukova, and T. V. Komarova, "Metabolic methanol: Molecular pathways and physiological roles," *Physiological reviews*, vol. 95, no. 2, pp. 603–644, 2015. DOI: 10.1152/physrev.00034.2014.

[165] A. Galvez-Ruiz, S. M. Elkhamary, N. Asghar, and T. M. Bosley, "Visual and neurologic sequelae of methanol poisoning in Saudi Arabia," *Saudi medical journal*, vol. 36, no. 5, p. 568, 2015. DOI: 10.15537/smj.2015.5.11142.

[166] A. Sadun, "Acquired mitochondrial impairment as a cause of optic nerve disease," *Transactions of the American Ophthalmological Society*, vol. 96, p. 881, 1998.

[167] J. I. Garaycoechea *et al.*, "Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells," *Nature*, vol. 553, no. 7687, pp. 171–177, 2018. DOI: 10.1038/nature25154.

[168] C.-L. Lai *et al.*, "Dominance of the inactive asian variant over activity and protein contents of mitochondrial aldehyde dehydrogenase 2 in human liver," *Alcoholism: Clinical and Experimental Research*, vol. 38, no. 1, pp. 44–50, 2014. DOI: 10.1111/acer.12215.

[169] A. Yokoyama and T. Omori, "Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers," *Japanese Journal of Clinical Oncology*, vol. 33, no. 3, pp. 111–121, 2003. DOI: 10.1093/jjco/hyg026.

[170] M. Y. Eng, S. E. Luczak, and T. L. Wall, "ALDH2, ADH1B, and ADH1C genotypes in Asians: A literature review," *Alcohol Research & Health*, vol. 30, no. 1, p. 22, 2007.

[171] L. J. Bierut *et al.*, "ADH1B is associated with alcohol dependence and alcohol consumption in populations of european and african ancestry," *Molecular psychiatry*, vol. 17, no. 4, pp. 445–450, 2012. DOI: 10.1038/mp.2011.124.

[172] T. Kawai *et al.*, "Endogenous acetaldehyde toxicity during antral follicular development in the mouse ovary," *Reproductive Toxicology*, vol. 33, no. 3, pp. 322–330, 2012. DOI: 10.1016/j.reprotox.2012.01.001.

[173] M. R. Hodskinson *et al.*, "Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms," *Nature*, pp. 1–6, 2020. DOI: 10.1038/s41586-020-2059-5.

[174] H. Niknahad *et al.*, "Modulating carbonyl cytotoxicity in intact rat hepatocytes by inhibiting carbonyl-metabolizing enzymes. I. Aliphatic alkenals," *Chemico-biological interactions*, vol. 143, pp. 107–117, 2003. DOI: 10.1016/S0009-2797(02)00185-0.

[175] Roessler Deutsche Gold- und Sil (DE); Degussa (DE), Patent GB-513772-A, 1937.

[176] S. Horiyama *et al.*, "Intracellular metabolism of α, β-unsaturated carbonyl compounds, acrolein, crotonaldehyde and methyl vinyl ketone, active toxicants in cigarette smoke: Participation of glutathione conjugation ability and aldehyde–ketone sensitive reductase activity," *Chemical and Pharmaceutical Bulletin*, vol. 64, no. 6, pp. 585–593, 2016. DOI: 10.1248/cpb.c15-00986.

[177] M. M. Anderson, S. L. Hazen, F. F. Hsu, J. W. Heinecke, *et al.*, "Human neutrophils employ the myeloperoxidase-hydrogen peroxide-chloride system to convert hydroxy-amino acids into glycolaldehyde, 2-hydroxypropanal, and acrolein. a mechanism for the generation of highly reactive alpha-hydroxy and alpha, beta-unsaturated aldehydes by phagocytes at sites of inflammation," *The Journal of clinical investigation*, vol. 99, no. 3, pp. 424–432, 1997. DOI: 10.1172/JCI119176.

[178] K. Uchida, M. Kanematsu, Y. Morimitsu, T. Osawa, N. Noguchi, and E. Niki, "Acrolein is a product of lipid peroxidation reaction: Formation of free acrolein and its conjugate with lysine residues in oxidized low density lipoproteins," *Journal of Biological Chemistry*, vol. 273, no. 26, pp. 16 058–16 066, 1998. DOI: 10.1074/jbc.273.26.16058.

[179] G. Houen, K. Bock, A. L. Jensen, *et al.*, "HPLC and NMR investigation of the serum amine oxidase catalyzed oxidation of polyamines," *Acta Chemica Scandinavica*, vol. 48, pp. 52–52, 1994. DOI: 10.3891/acta.chem.scand.48-0052.

[180] J. Cannon, C. A. Links, and L. R. Cos, "Cyclophosphamide-associated carcinomaof urothelium: Modalities for prevention," *Urology*, vol. 38, no. 5, pp. 413–416, 1991. DOI: 10.1016/0090-4295(91)80228-Y.

[181] K. Abraham, S. Andres, R. Palavinskas, K. Berg, K. E. Appel, and A. Lampen, "Toxicology and risk assessment of acrolein in food," *Molecular nutrition & food research*, vol. 55, no. 9, pp. 1277–1290, 2011. DOI: 10.1002/mnfr.201100481.

[182] T. Watanabe and D. M. Aviado, "Functional and biochemical effects on the lung following inhalation of cigarette smoke and constituents. II. Skatole, acrolein, and acetaldehyde," *Toxicology and Applied Pharmacology*, vol. 30, no. 2, pp. 201–209, 1974. DOI: 10.1016/0041-008X(74)90091-X.

[183] Y. Hoshino, T. Mio, S. Nagai, H. Miki, I. Ito, and T. Izumi, "Cytotoxic effects of cigarette smoke extract on an alveolar type II cell-derived cell line," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 281, no. 2, pp. L509–L516, 2001. DOI: 10.1152/ajplung.2001.281.2.L509.

[184] B. Yoval-Sánchez and J. S. Rodrıguez-Zavala, "Differences in susceptibility to inactivation of human aldehyde dehydrogenases by lipid peroxidation byproducts," *Chemical research in toxicology*, vol. 25, no. 3, pp. 722–729, 2012. DOI: 10.1021/tx2005184.

[185] D. Taeymans *et al.*, "A review of acrylamide: An industry perspective on research, analysis, formation, and control," *Critical reviews in food science and nutrition*, vol. 44, no. 5, pp. 323–347, 2004. DOI: 10.1080/10408690490478082.

[186] G. C. Tong, W. K. Cornwell, and G. E. Means, "Reactions of acrylamide with glutathione and serum albumin," *Toxicology letters*, vol. 147, no. 2, pp. 127–131, 2004. DOI: 10.1016/j.toxlet.2003.10.021.

[187] J. A. Hinson, D. W. Roberts, and L. P. James, "Mechanisms of acetaminophen-induced liver necrosis," *Adverse drug reactions*, pp. 369–405, 2010. DOI: 10.1007/978-3-642-00663-0_12.

[188] S. Steinbrecht *et al.*, "Synthesis of cyclophosphamide metabolites by a peroxygenase from Marasmius rotula for toxicological studies on human cancer cells," *AMB Express*, vol. 10, no. 1, pp. 1–13, 2020. DOI: 10.1186/s13568-020-01064-w.

[189] A. Vliegenthart *et al.*, "Circulating acetaminophen metabolites are toxicokinetic biomarkers of acute liver injury," *Clinical Pharmacology & Therapeutics*, vol. 101, no. 4, pp. 531–540, 2017. DOI: 10.1002/cpt.541.

[190] S. Mehta, "Metabolism of paracetamol,(acetaminophen), acetanilide and phenacetin," *Pharma X Change info Archived from the original on*, vol. 28, 2012.

[191] A. A. Kadir, K. Clarke, and R. D. Evans, "Cardiac ketone body metabolism," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1866, no. 6, p. 165 739, 2020. DOI: 10.1016/j.bbadis.2020.165739.

[192] M. M. Cox and D. L. Nelson, *Lehninger principles of biochemistry*. Wh Freeman New York, 2008, vol. 5.

[193] P. J. Thornalley, A. Langborg, and H. S. Minhas, "Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose," *Biochemical Journal*, vol. 344, no. 1, pp. 109–116, 1999. DOI: 10.1042/bj3440109.

[194] P. J. Thornalley, "Modification of the glyoxalase system in human red blood cells by glucose in vitro," *Biochemical Journal*, vol. 254, no. 3, pp. 751–755, 1988. DOI: 10.1042/bj2540751.

[195] Y. Yang *et al.*, "Prolonged exposure to methylglyoxal causes disruption of vascular K_{ATP} channel by mRNA instability," *American Journal of Physiology-Cell Physiology*, vol. 303, no. 10, pp. C1045–C1054, 2012. DOI: 10.1152/ajpcell.00020.2012.

[196] N. Rabbani and P. J. Thornalley, "Glyoxalase 1 modulation in obesity and diabetes," *Antioxidants & redox signaling*, vol. 30, no. 3, pp. 354–374, 2019. DOI: 10.1089/ars.2017.7424.

[197] P. J. Thornalley, A. Yurek-George, and O. K. Argirov, "Kinetics and mechanism of the reaction of aminoguanidine with the α-oxoaldehydes glyoxal, methylglyoxal, and 3-deoxyglucosone under physiological conditions," *Biochemical pharmacology*, vol. 60, no. 1, pp. 55–65, 2000. DOI: 10.1016/S0006-2952(00)00287-2.

[198] E. A. Betterton and M. R. Hoffmann, "Henry's law constants of some environmentally important aldehydes," *Environmental science & technology*, vol. 22, no. 12, pp. 1415–1418, 1988. DOI: 10.1021/es00177a004.

[199] E. B. Whipple, "Structure of glyoxal in water," *Journal of the American Chemical Society*, vol. 92, no. 24, pp. 7183–7186, 1970. DOI: 10.1021/ja00727a027.

[200] C. Nigro *et al.*, "Dicarbonyl stress at the crossroads of healthy and unhealthy aging," *Cells*, vol. 8, no. 7, p. 749, 2019. DOI: 10.3390/cells8070749.

[201] L. B. Sullivan, D. Y. Gui, and M. G. Vander Heiden, "Altered metabolite levels in cancer: Implications for tumour biology and cancer therapy," *Nature Reviews Cancer*, vol. 16, no. 11, pp. 680–693, 2016. DOI: 10.1038/nrc.2016.85.

[202] F. Collard, D. Vertommen, J. Fortpied, G. Duester, and E. Van Schaftingen, "Identification of 3-deoxyglucosone dehydrogenase as aldehyde dehydrogenase 1A1 (retinaldehyde dehydrogenase 1)," *Biochimie*, vol. 89, no. 3, pp. 369–373, 2007. DOI: 10.1016/j.biochi.2006.11.005.

[203] C.-I. Lin, R. M. McCarty, and H.-w. Liu, "The enzymology of organic transformations: A survey of name reactions in biological systems," *Angewandte Chemie International Edition*, vol. 56, no. 13, pp. 3446–3489, 2017. DOI: 10.1002/anie.201603291.

[204] P. B. L. Pun and M. P. Murphy, "Pathological significance of mitochondrial glycation," *International journal of cell biology*, vol. 2012, 2012. DOI: 10.1155/2012/843505.

[205] Q. Zheng *et al.*, "Reversible histone glycation is associated with disease-related changes in chromatin architecture," *Nature communications*, vol. 10, no. 1, pp. 1–12, 2019. DOI: 10.1038/s41467-019-09192-z.

[206] S. Ratner, V. Nocito, and D. Green, "Glycine oxidase," *Journal of Biological Chemistry*, vol. 152, no. 1, pp. 119–133, 1944. DOI: 10.1016/S0021-9258(18)72097-8.

[207] V. J. Nikiforova *et al.*, "Glyoxylate, a new marker metabolite of type 2 diabetes," *Journal of diabetes research*, vol. 2014, 2014. DOI: 10.1155/2014/685204.

[208] M. A. Troester, A. B. Lindstrom, L. L. Kupper, S. Waidyanatha, and S. M. Rappaport, "Stability of hemoglobin and albumin adducts of benzene oxide and 1, 4-benzoquinone after administration of benzene to F344 rats," *Toxicological Sciences*, vol. 54, no. 1, pp. 88–94, 2000. DOI: 10.1093/toxsci/54.1.88.

[209] C. Weykamp, "HbA1c: A review of analytical and clinical aspects," *Annals of laboratory medicine*, vol. 33, no. 6, pp. 393–400, 2013. DOI: 10.3343/alm.2013.33.6.393.

[210] M. Amadori, "Atti reale accad. nazl," *Lincei,(6)*, vol. 13, p. 72, 1931.

[211] S. Hörber, P. Achenbach, E. Schleicher, and A. Peter, "Harmonization of immunoassays for biomarkers in diabetes mellitus," *Biotechnology advances*, vol. 39, p. 107 359, 2020. DOI: 10.1016/j.biotechadv.2019.02.015.

[212] H. Chardon and J. Duval, "Quelques aspects de la pharmacocinétique de la métampicilline chez l'homme: Son élimination biliaire comparée à celle de l'ampicilline," *Médecine et Maladies Infectieuses*, vol. 4, no. 3, pp. 109–118, 1974. DOI: 10.1016/S0399-077X(74)80103-4.

[213] J. Brogard, M. Pinget, M. Adloff, M. Dorner, and J. Lavillaureix, "Experimental and clinical pharmacology of metampicillin in the biliary tract," *Journal of Antimicrobial Chemotherapy*, vol. 2, no. 4, pp. 363–371, 1976. DOI: 10.1093/jac/2.4.363.

[214] J. W. Von Goethe, *Faust: zweiter teil*. OR Reisland, 1896.

[215] J. W. Von Goethe, *Faust: A tragedy by Joh. Wolfg. von Goethe, translated, in the original metres, by Bayard Taylor. II*. Strakan & Company, 1871.

[216] O. M. Sidorkina and J. Laval, "Role of the N-terminal proline residue in the catalytic activities of the Escherichia coli Fpg protein," *Journal of Biological Chemistry*, vol. 275, no. 14, pp. 9924–9929, 2000. DOI: 10.1074/jbc.275.14.9924.

[217] E. Zandvoort, B.-J. Baas, W. J. Quax, and G. J. Poelarends, "Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: Enamine formation and aldolase activity," *ChemBioChem*, vol. 12, no. 4, pp. 602–609, 2011. DOI: 10.1002/cbic.201000633.

[218] A. Brik, P. E. Dawson, and E. Keinan, "The product of the natural reaction catalyzed by 4-oxalocrotonate tautomerase becomes an affinity label of its mutant," *Bioorganic & medicinal chemistry*, vol. 10, no. 12, pp. 3891–3897, 2002. DOI: 10.1016/S0968-0896(02)00385-1.

[219] C. P. Whitman, "The 4-oxalocrotonate tautomerase family of enzymes: How nature makes new enzymes using a β–α–β structural motif," *Archives of biochemistry and biophysics*, vol. 402, no. 1, pp. 1–13, 2002. DOI: 10.1016/S0003-9861(02)00052-8.

[220] Y. Miao, E. M. Geertsema, P. G. Tepper, E. Zandvoort, and G. J. Poelarends, "Promiscuous catalysis of asymmetric michael-type additions of linear aldehydes to β-nitrostyrene by the proline-based enzyme 4-oxalocrotonate tautomerase," *ChemBioChem*, vol. 14, no. 2, pp. 191–194, 2013. DOI: 10.1002/cbic.201200676.

[221] E. M. Geertsema, Y. Miao, P. G. Tepper, P. de Haan, E. Zandvoort, and G. J. Poelarends, "Biocatalytic michael-type additions of acetaldehyde to nitroolefins with the proline-based enzyme 4-oxalocrotonate tautomerase yielding enantioenriched γ-nitroaldehydes," *Chemistry–A European Journal*, vol. 19, no. 43, pp. 14 407–14 410, 2013. DOI: 10.1002/chem.201302351.

[222] M. S. Lukesch, T. Pavkov-Keller, K. Gruber, K. Zangger, and B. Wiltschi, "Substituting the catalytic proline of 4-oxalocrotonate tautomerase with non-canonical analogues reveals a finely tuned catalytic system," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019. DOI: 10.1038/s41598-019-39484-9.

[223] S. Boiteux, F. Coste, and B. Castaing, "Repair of 8-oxo-7, 8-dihydroguanine in prokaryotic and eukaryotic cells: Properties and biological roles of the Fpg and OGG1 DNA N-glycosylases," *Free Radical Biology and Medicine*, vol. 107, pp. 179–201, 2017. DOI: 10.1016/j.freeradbiomed.2016.11.042.

[224] R. Desrosiers and R. Tanguay, "Methylation of Drosophila histones at proline, lysine, and arginine residues during heat shock," *Journal of Biological Chemistry*, vol. 263, no. 10, pp. 4686–4692, 1988. DOI: 10.1016/S0021-9258(18)68837-4.

[225] A. Martinage, G. Briand, A. Van Dorsselaer, C. H. Turner, and P. Sautiere, "Primary structure of histone H2B from gonads of the starfish Asterias rubens: Identification of an N-dimethylproline residue at the amino-terminal," *European journal of biochemistry*, vol. 147, no. 2, pp. 351–359, 1985. DOI: 10.1111/j.1432-1033.1985.tb08757.x.

[226] F. M. Olajuyigbe, N. Demitri, J. O. Ajele, E. Maurizio, L. Randaccio, and S. Geremia, "Carbamylation of N-terminal proline," *ACS medicinal chemistry letters*, vol. 1, no. 6, pp. 254–257, 2010. DOI: 10.1021/ml100046d.

[227] S. Sun, J.-Y. Zhou, W. Yang, and H. Zhang, "Inhibition of protein carbamylation in urea solution using ammonium-containing buffers," *Analytical biochemistry*, vol. 446, pp. 76–81, 2014. DOI: 10.1016/j.ab.2013.10.024.

[228] P. M. Lavanchy, "Microbial community metabolic concurrence involved in toluene degradation: Effect of oxygen availability on catabolic gene expression of aerobic and anaerobic toluene degrading bacteria," Ph.D. dissertation, 2009.

[229] R. Parales, J. Parales, D. Pelletier, and J. Ditty, "Diversity of microbial toluene degradation pathways," *Advances in applied microbiology*, vol. 64, pp. 1–73, 2008. DOI: 10.1016/S0065-2164(08)00401-2.

[230] S. Delanghe, J. R. Delanghe, R. Speeckaert, W. Van Biesen, and M. M. Speeckaert, "Mechanisms and consequences of carbamoylation," *Nature Reviews Nephrology*, vol. 13, no. 9, pp. 580–593, 2017. DOI: 10.1038/nrneph.2017.103.

[231] L. M. Kraus and A. P. Kraus Jr, "Carbamoylation of amino acids and proteins in uremia," *Kidney international*, vol. 59, S102–S107, 2001. DOI: 10.1046/j.1523-1755.2001.59780102.x.

[232] A. S. Ghifari, P. F. Teixeira, B. Kmiec, A. Pružinská, E. Glaser, and M. W. Murcha, "A mitochondrial prolyl aminopeptidase PAP2 releases N-terminal proline and regulates proline homeostasis during stress response," *The Plant Journal*, vol. 104, no. 5, pp. 1182–1194, 2020. DOI: 10.1111/tpj.14987.

[233] J. Yeom, S. Ju, Y. Choi, E. Paek, and C. Lee, "Comprehensive analysis of human protein N-termini enables assessment of various protein forms," *Scientific reports*, vol. 7, no. 1, pp. 1–13, 2017. DOI: 10.1038/s41598-017-06314-9.

[234] S. Venkatesh and J. L. Workman, "Histone exchange, chromatin structure and the regulation of transcription," *Nature reviews Molecular cell biology*, vol. 16, no. 3, pp. 178–189, 2015. DOI: 10.1038/nrm3941.

[235] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Cell research*, vol. 21, no. 3, pp. 381–395, 2011. DOI: 10.1038/cr.2011.22.

[236] E. L. Greer and Y. Shi, "Histone methylation: A dynamic mark in health, disease and inheritance," *Nature Reviews Genetics*, vol. 13, no. 5, pp. 343–357, 2012. DOI: 10.1038/nrg3173.

[237] C. J. Cunliffe, T. J. Franklin, N. J. Hales, and G. B. Hill, "Novel inhibitors of prolyl 4-hydroxylase. 3. inhibition by the substrate analog N-oxaloglycine and its derivatives," *Journal of medicinal chemistry*, vol. 35, no. 14, pp. 2652–2658, 1992. DOI: 10.1021/jm00092a016.

[238] Y. L. Liou, M. E. Ericson, and E. M. Warshaw, "Formaldehyde release from baby wipes: Analysis using the chromotropic acid method," *Dermatitis*, vol. 30, no. 3, pp. 207–212, 2019. DOI: 10.1097/DER.0000000000000478.

[239] A. Duong, C. Steinmaus, C. M. McHale, C. P. Vaughan, and L. Zhang, "Reproductive and developmental toxicity of formaldehyde: A systematic review," *Mutation Research/Reviews in Mutation Research*, vol. 728, no. 3, pp. 118–138, 2011. DOI: 10.1016/j.mrrev.2011.07.003.

[240] K.-H. Kim, S. A. Jahan, and J.-T. Lee, "Exposure to formaldehyde and its potential human health hazards," *Journal of Environmental Science and Health, Part C*, vol. 29, no. 4, pp. 277–299, 2011. DOI: 10.1080/10590501.2011.629972.

[241] C. Lv, J. Hou, W. Xie, and H. Cheng, "Investigation on formaldehyde release from preservatives in cosmetics," *International journal of cosmetic science*, vol. 37, no. 5, pp. 474–478, 2015. DOI: 10.1111/ics.12212.

[242] S. V. Lehmann, U. Hoeck, J. Breinholdt, C. E. Olsen, and B. Kreilgaard, "Characterization and chemistry of imidazolidinyl urea and diazolidinyl urea," *Contact dermatitis*, vol. 54, no. 1, pp. 50–58, 2006. DOI: 10.1111/j.0105-1873.2006.00735.x.

[243] D. J. Gibbs, J. Bacardit, A. Bachmair, and M. J. Holdsworth, "The eukaryotic N-end rule pathway: Conserved mechanisms and diverse functions," *Trends in Cell Biology*, vol. 24, no. 10, pp. 603–611, 2014. DOI: 10.1016/j.tcb.2014.05.001.

[244] T. Narancic *et al.*, "Highly efficient michael-type addition of acetaldehyde to β-nitrostyrenes by whole resting cells of Escherichia coli expressing 4-oxalocrotonate tautomerase," *Bioresource technology*, vol. 142, pp. 462–468, 2013. DOI: 10.1016/j.biortech.2013.05.074.

[245] R. Reinbold *et al.*, "Formaldehyde quantification using ampicillin is not selective," *Scientific reports*, vol. 9, no. 1, pp. 1–7, 2019. DOI: 10.1038/s41598-019-54610-3.

[246] T. R. Malla *et al.*, "Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors," *Chemical Communications*, vol. 57, no. 12, pp. 1430–1433, 2021. DOI: 10.1039/D0CC06870E.

[247] S. K. Bharti and R. Roy, "Quantitative 1H NMR spectroscopy," *TrAC Trends in Analytical Chemistry*, vol. 35, pp. 5–26, 2012. DOI: 10.1016/j.trac.2012.02.007.

[248] V. Poláčková, P. Čmelová, R. Górová, and R. Šebesta, "Peptide-catalyzed stereoselective michael addition of aldehydes and ketones to heterocyclic nitroalkenes," *Monatshefte für Chemie-Chemical Monthly*, vol. 149, no. 4, pp. 729–736, 2018. DOI: 10.1007/s00706-017-2126-2.

[249] D. Kim, B. Langmead, and S. L. Salzberg, "HISAT: A fast spliced aligner with low memory requirements," *Nature methods*, vol. 12, no. 4, pp. 357–360, 2015. DOI: 10.1038/nmeth.3317.

[250] H. Li *et al.*, "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009. DOI: 10.1093/bioinformatics/btp352.

[251] M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, and S. L. Salzberg, "Stringtie enables improved reconstruction of a transcriptome from RNA-seq reads," *Nature biotechnology*, vol. 33, no. 3, pp. 290–295, 2015. DOI: 10.1038/nbt.3122.

[252] A. C. Frazee, G. Pertea, A. E. Jaffe, B. Langmead, S. L. Salzberg, and J. T. Leek, "Ballgown bridges the gap between transcriptome assembly and expression analysis," *Nature biotechnology*, vol. 33, no. 3, pp. 243–246, 2015. DOI: 10.1038/nbt.3172.

[253] M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg, "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown," *Nature protocols*, vol. 11, no. 9, p. 1650, 2016. DOI: 10.1038/nprot.2016.095.

[254] E. Zandvoort, E. M. Geertsema, B.-J. Baas, W. J. Quax, and G. J. Poelarends, "Bridging between organocatalysis and biocatalysis: Asymmetric addition of acetaldehyde to β-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase," *Angewandte Chemie*, vol. 124, no. 5, pp. 1266–1269, 2012. DOI: 10.1002/ange.201107404.

[255] T. Liu *et al.*, "Biochemical and biophysical analyses of hypoxia sensing prolyl hydroxylases from dictyostelium discoideum and toxoplasma gondii," *Journal of Biological Chemistry*, vol. 295, no. 49, pp. 16 545–16 561, 2020. DOI: 10.1074/jbc.RA120.013998.

[256] A. Kornberg, *For the love of enzymes: the odyssey of a biochemist*. Harvard University Press, 1991.

[257] J. A. Patterson *et al.*, "Thioproline formation as a driver of formaldehyde toxicity in Escherichia coli," *Biochemical Journal*, vol. 477, no. 9, pp. 1745–1757, 2020. DOI: 10.1042/BCJ20200198.

[258] N. T. T. Huyen *et al.*, "Genome-wide responses to carbonyl electrophiles in Bacillus subtilis: Control of the thiol-dependent formaldehyde dehydrogenase AdhA and cysteine proteinase YraA by the MerR-family regulator YraB (AdhR)," *Molecular microbiology*, vol. 71, no. 4, pp. 876–894, 2009. DOI: 10.1111/j.1365-2958.2008.06568.x.

[259] R.-G. Hu, J. Sheng, X. Qi, Z. Xu, T. T. Takahashi, and A. Varshavsky, "The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators," *Nature*, vol. 437, no. 7061, pp. 981–986, 2005. DOI: 10.1038/nature04027.

[260] M. D. White *et al.*, "Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets," *Nature communications*, vol. 8, no. 1, pp. 1–9, 2017. DOI: 10.1038/ncomms14690.

[261] N. Masson *et al.*, "Conserved N-terminal cysteine dioxygenases transduce responses to hypoxia in animals and plants," *Science*, vol. 365, no. 6448, pp. 65–69, 2019. DOI: 10.1126/science.aaw0112.

[262] P. Kaushal, Y. Kwon, S. Ju, and C. Lee, "An SDS-PAGE based proteomic approach for N-terminome profiling," *Analyst*, vol. 144, no. 23, pp. 7001–7009, 2019. DOI: 10.1039/C9AN01616C.

[263] R. Manhas *et al.*, "Identification and functional characterization of a novel bacterial type asparagine synthetase a: A tRNA synthetase paralog from Leishmania donovani," *Journal of Biological Chemistry*, vol. 289, no. 17, pp. 12 096–12 108, 2014. DOI: 10.1074/jbc.M114.554642.

[264] N. Richards and S. M. Schuster, "Mechanistic issues in asparagine synthetase catalysis," *Advances in enzymology and related areas of molecular biology*, vol. 72, pp. 145–198, 1998. DOI: 10.1002/9780470123188.ch5.

[265] C. Oinonen and J. Rouvinen, "Structural comparison of Ntn-hydrolases," *Protein Science*, vol. 9, no. 12, pp. 2329–2337, 2000. DOI: 10.1110/ps.9.12.2329.

[266] E. M. Huber, W. Heinemeyer, X. Li, C. S. Arendt, M. Hochstrasser, and M. Groll, "A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome," *Nature communications*, vol. 7, no. 1, pp. 1–10, 2016. DOI: 10.1038/ncomms10900.

[267] H. J. Duggleby, S. P. Tolley, C. P. Hill, E. J. Dodson, G. Dodson, and P. C. Moody, "Penicillin acylase has a single-amino-acid catalytic centre," *Nature*, vol. 373, no. 6511, pp. 264–268, 1995. DOI: 10.1038/373264a0.

[268] J. L. Smith *et al.*, "Structure of the allosteric regulatory enzyme of purine biosynthesis," *Science*, vol. 264, no. 5164, pp. 1427–1433, 1994. DOI: 10.1126/science.8197456.

[269] M. N. Isupov *et al.*, "Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: Evidence from the 1.8 Å crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase," *Structure*, vol. 4, no. 7, pp. 801–810, 1996. DOI: 10.1016/S0969-2126(96)00087-1.

[270] C. Oinonen, R. Tikkanen, J. Rouvinen, and L. Peltonen, "Three-dimensional structure of human lysosomal aspartylglucosaminidase," *Nature structural biology*, vol. 2, no. 12, pp. 1102–1108, 1995. DOI: 10.1038/nsb1295-1102.

[271] C. Bompard-Gilles *et al.*, "A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-ala-esterase/amidase from Ochrobactrum anthropi," *Structure*, vol. 8, no. 2, pp. 153–162, 2000. DOI: 10.1016/S0969-2126(00)00091-5.

[272] O. Kärkkäinen *et al.*, "Changes in circulating metabolome precede alcohol-related diseases in middle-aged men: A prospective population-based study with a 30-year follow-up," *Alcoholism: Clinical and Experimental Research*, vol. 44, no. 12, pp. 2457–2467, 2020. DOI: 10.1111/acer.14485.

[273] F. Ottosson, E. Smith, O. Melander, and C. Fernandez, "Altered asparagine and glutamate homeostasis precede coronary artery disease and type 2 diabetes," *The Journal of Clinical Endocrinology & Metabolism*, vol. 103, no. 8, pp. 3060–3069, 2018. DOI: 10.1210/jc.2018-00546.

[274] E. K. Ruzzo *et al.*, "Deficiency of asparagine synthetase causes congenital microcephaly and a progressive form of encephalopathy," *Neuron*, vol. 80, no. 2, pp. 429–441, 2013. DOI: 10.1016/j.neuron.2013.08.013.

[275] C. L. Lomelino, J. T. Andring, R. McKenna, and M. S. Kilberg, "Asparagine synthetase: Function, structure, and role in disease," *Journal of Biological Chemistry*, vol. 292, no. 49, pp. 19 952–19 958, 2017. DOI: 10.1074/jbc.R117.819060.

[276] W. Zhu *et al.*, "High-resolution crystal structure of human asparagine synthetase enables analysis of inhibitor binding and selectivity," *Communications biology*, vol. 2, no. 1, pp. 1–14, 2019. DOI: 10.1038/s42003-019-0587-z.

[277] M. Chiu, G. Taurino, M. G. Bianchi, M. S. Kilberg, and O. Bussolati, "Asparagine synthetase in cancer: Beyond acute lymphoblastic leukemia," *Frontiers in oncology*, vol. 9, p. 1480, 2020. DOI: 10.3389/fonc.2019.01480.

[278] L. D. Stegink *et al.*, "Plasma amino acid concentrations and amino acid ratios in normal adults and adults heterozygous for phenylketonuria ingesting a hamburger and milk shake meal," *The American journal of clinical nutrition*, vol. 53, no. 3, pp. 670–675, 1991. DOI: 10.1093/ajcn/53.3.670.

[279] P. Chen and S. W. Harcum, "Effects of elevated ammonium on glycosylation gene expression in CHO cells," *Metabolic Engineering*, vol. 8, no. 2, pp. 123–132, 2006.

[280] J. Van der Valk *et al.*, "Optimization of chemically defined cell culture media–replacing fetal bovine serum in mammalian in vitro methods," *Toxicology in vitro*, vol. 24, no. 4, pp. 1053–1063, 2010. DOI: 10.1016/j.tiv.2010.03.016.

[281] A. S. Krall, S. Xu, T. G. Graeber, D. Braas, and H. R. Christofk, "Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor," *Nature communications*, vol. 7, no. 1, pp. 1–13, 2016. DOI: 10.1038/ncomms11457.

[282] N. N. Pavlova *et al.*, "As extracellular glutamine levels decline, asparagine becomes an essential amino acid," *Cell metabolism*, vol. 27, no. 2, pp. 428–438, 2018. DOI: 10.1016/j.cmet.2017.12.006.

[283] S. Boehlein, N. Richards, and S. Schuster, "Glutamine-dependent nitrogen transfer in Escherichia coli asparagine synthetase B. searching for the catalytic triad," *Journal of Biological Chemistry*, vol. 269, no. 10, pp. 7450–7457, 1994. DOI: 10.1016/S0021-9258(17)37307-6.

[284] P. S. Atwal, C. R. Medina, L. C. Burrage, and V. R. Sutton, "Nineteen-year follow-up of a patient with severe glutathione synthetase deficiency," *Journal of human genetics*, vol. 61, no. 7, pp. 669–672, 2016. DOI: 10.1038/jhg.2016.20.

[285] M. Schneider, I. W. Marison, and U. Von Stockar, "The importance of ammonia in mammalian cell culture," *Journal of biotechnology*, vol. 46, no. 3, pp. 161–185, 1996. DOI: 10.1016/0168-1656(95)00196-4.

[286] F. M. Issoglio *et al.*, "Exploring the catalytic mechanism of human glutamine synthetase by computer simulations," *Biochemistry*, vol. 55, no. 42, pp. 5907–5916, 2016. DOI: 10.1021/acs.biochem.6b00822.

[287] Y. Elhassan, G. Wu, A. Leanez, R. Tasca, A. Watson, and M. Westhusin, "Amino acid concentrations in fluids from the bovine oviduct and uterus and in KSOM-based culture media," *Theriogenology*, vol. 55, no. 9, pp. 1907–1918, 2001. DOI: 10.1016/S0093-691X(01)00532-5.

[288] Z. Li, X. Zhang, Y. Xue, J. Zhang, M. Li, and C. Luo, "Cellular and molecular evidence of acetaldehyde elimination and intracellular environment antioxidation by L-cysteine," *Journal of Chemistry*, vol. 2018, 2018. DOI: 10.1155/2018/6864574.

[289] N. Rabbani and P. J. Thornalley, "Measurement of methylglyoxal by stable isotopic dilution analysis LC-MS/MS with corroborative prediction in physiological samples," *Nature protocols*, vol. 9, no. 8, pp. 1969–1979, 2014. DOI: 10.1038/nprot.2014.129.

[290] L. Di, A. Balesano, S. Jordan, and S. M. Shi, "The role of alcohol dehydrogenase in drug metabolism: Beyond ethanol oxidation," *The AAPS Journal*, vol. 23, no. 1, pp. 1–21, 2021. DOI: 10.1208/s12248-020-00536-y.

[291] C. P. Eriksson *et al.*, "L-cysteine containing vitamin supplement which prevents or alleviates alcohol-related hangover symptoms: Nausea, headache, stress and anxiety," *Alcohol and Alcoholism*, vol. 55, no. 6, pp. 660–666, 2020. DOI: 10.1093/alcalc/agaa082.

[292] B.-Y. Kim *et al.*, "Effects of asparagus officinalis extracts on liver cell toxicity and ethanol metabolism," *Journal of food science*, vol. 74, no. 7, H204–H208, 2009. DOI: 10.1111/j.1750-3841.2009.01263.x.

[293] C. P. Eriksson *et al.*, "Measurement of acetaldehyde: What levels occur naturally and in response to alcohol?" In *Novartis Foundation Symposium*, Chichester; New York; John Wiley; 1999, vol. 285, 2007, p. 247. DOI: 10.1002/9780470511848.ch18.

[294] A. Yokoyama *et al.*, "Salivary acetaldehyde concentration according to alcoholic beverage consumed and aldehyde dehydrogenase-2 genotype," *Alcoholism: Clinical and Experimental Research*, vol. 32, no. 9, pp. 1607–1614, 2008. DOI: 10.1111/j.1530-0277.2008.00739.x.

[295] C. Burr, *The emperor of scent: A true story of perfume and obsession*. Random House Incorporated, 2004.

[296] R. Reinbold, T. John, P. Spingardi, A. Kawamura, C. J. Schofield, and R. J. Hopkinson, "Metampicillin is a cyclic aminal produced by reaction of ampicillin with formaldehyde," *Scientific reports*, vol. 10, no. 1, pp. 1–5, 2020. DOI: 10.1038/s41598-020-74990-1.

[297] EMA, *European medicines agency, european surveillance of veterinary antimicrobial consumption. sales of veterinary antimicrobial agents in 30 european countries in 2016.(ema/275982/2018)*, 2018.

[298] W. H. Organization *et al.*, "Critically important antimicrobials for human medicine: Ranking of antimicrobial agents for risk management of antimicrobial resistance due to non-human use," 2017.

[299] R. Sutherland, S. Elson, and E. Croydon, "Metampicillin," *Chemotherapy*, vol. 17, no. 3, pp. 145–160, 1972. DOI: 10.1159/000220849.

[300] M. L. Grayson *et al.*, *Kucers' The Use of Antibiotics Sixth Edition: A Clinical Review of Antibacterial, Antifungal and Antiviral Drugs*. CRC Press, 2010.

[301] E. Ferrero, "Méthode pour l'étude de la résistance des penicillines à la dégradation produite par la flore intestinale," *Proc 5th Int Congr Chemother. Vienna, Verlag der Wiener Medizinischen Akademie*, vol. 1, p. 1, 1967.

[302] G. E. Stanley, *N-methylol-alpha-aminobenzyl-penicillin*, US Patent 3,198,788, Aug. 1965.

[303] D. Spencer and T. Henshall, "The kinetics and mechanism of the reaction of formaldehyde with dimedone. part I," *Journal of the American Chemical Society*, vol. 77, no. 7, pp. 1943–1948, 1955.

[304] R. A. Bragg *et al.*, "Aortic binding of AZD5248: Mechanistic insight and reactivity assays to support lead optimzation," *Chemical research in toxicology*, vol. 28, no. 10, pp. 1991–1999, 2015. DOI: 10.1021/acs.chemrestox.5b00236.

[305] H. D. Thoreau, *The Writings of Henry David Thoreau: Walden Or, Life in the Woods*. Houghton, Mifflin & Company, 1893.

[306] T. J. El-Baba *et al.*, "Allosteric inhibition of the SARS-CoV-2 main protease: Insights from mass spectrometry based assays," *Angewandte Chemie International Edition*, vol. 59, no. 52, pp. 23 544–23 548, 2020. DOI: 10.1002/anie.202010316.

[307] H. Achdout *et al.*, "Covid moonshot: Open science discovery of SARS-CoV-2 main protease inhibitors by combining crowdsourcing, high-throughput experiments, computational simulations, and machine learning," *bioRxiv*, 2020. DOI: 10.1101/2020.10.29.339317.

[308] H. T. H. Chan *et al.*, "Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding," *Chemical Science*, vol. 12, pp. 13 686–13 703, 2021. DOI: 10.1039/D1SC03628A.

[309] S. Su *et al.*, "Epidemiology, genetic recombination, and pathogenesis of coronaviruses," *Trends in microbiology*, vol. 24, no. 6, pp. 490–502, 2016. DOI: 10.1016/j.tim.2016.03.003.

[310] Z.-W. Ye, S. Yuan, K.-S. Yuen, S.-Y. Fung, C.-P. Chan, and D.-Y. Jin, "Zoonotic origins of human coronaviruses," *International journal of biological sciences*, vol. 16, no. 10, p. 1686, 2020. DOI: 10.7150/ijbs.45472.

[311] Y. Chen, Q. Liu, and D. Guo, "Emerging coronaviruses: Genome structure, replication, and pathogenesis," *Journal of medical virology*, vol. 92, no. 4, pp. 418–423, 2020. DOI: 10.1002/jmv.25681.

[312] R. Hilgenfeld, "From SARS to MERS: Crystallographic studies on coronaviral proteases enable antiviral drug design," *The FEBS journal*, vol. 281, no. 18, pp. 4085–4096, 2014. DOI: 10.1111/febs.12936.

[313] L. Zhang *et al.*, "α-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: Structure-based design, synthesis, and activity assessment," *Journal of medicinal chemistry*, vol. 63, no. 9, pp. 4562–4578, 2020. DOI: 10.1021/acs.jmedchem.9b01828.

[314] K. Świderek and V. Moliner, "Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods," *Chemical Science*, vol. 11, no. 39, pp. 10 626–10 630, 2020. DOI: 10.1039/D0SC02823A.

[315] J. Ziebuhr, E. J. Snijder, and A. E. Gorbalenya, "Virus-encoded proteinases and proteolytic processing in the nidovirales," *Microbiology*, vol. 81, no. 4, pp. 853–879, 2000. DOI: 10.1099/0022-1317-81-4-853.

[316] J. Meng *et al.*, "Screening of HIV-1 protease using a combination of an ultra-high-throughput fluorescent-based assay and rapidfire mass spectrometry," *Journal of biomolecular screening*, vol. 20, no. 5, pp. 606–615, 2015. DOI: 10.1177/1087057115570838.

[317] L. Brewitz, A. Tumber, I. Pfeffer, M. A. McDonough, and C. J. Schofield, "Aspartate/asparagine-β-hydroxylase: A high-throughput mass spectrometric assay for discovery of small molecule inhibitors," *Scientific reports*, vol. 10, no. 1, pp. 1–14, 2020. DOI: 10.1038/s41598-020-65123-9.

[318] L. Brewitz, A. Tumber, X. Zhang, and C. J. Schofield, "Small-molecule active pharmaceutical ingredients of approved cancer therapeutics inhibit human aspartate/asparagine-β-hydroxylase," *Bioorganic & medicinal chemistry*, vol. 28, no. 20, p. 115 675, 2020. DOI: 10.1016/j.bmc.2020.115675.

[319] Z. Jin *et al.*, "Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors," *Nature*, vol. 582, no. 7811, pp. 289–293, 2020. DOI: 10.1038/s41586-020-2223-y.

[320] L. Zhang *et al.*, "Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors," *Science*, vol. 368, no. 6489, pp. 409–412, 2020. DOI: 10.1126/science.abb3405.

[321] J. Osipiuk *et al.*, "Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors," *Nature communications*, vol. 12, no. 1, pp. 1–9, 2021. DOI: 10.1038/s41467-021-21060-3.

[322] T. Klemm *et al.*, "Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2," *The EMBO journal*, vol. 39, no. 18, e106275, 2020. DOI: 10.15252/embj.2020106275.

[323] J. L. Harris, B. J. Backes, F. Leonetti, S. Mahrus, J. A. Ellman, and C. S. Craik, "Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries," *Proceedings of the National Academy of Sciences*, vol. 97, no. 14, pp. 7754–7759, 2000. DOI: 10.1073/pnas.140132697.

[324] Z. Chen *et al.*, "Ginkgolic acid and anacardic acid are specific covalent inhibitors of SARS-CoV-2 cysteine proteases," *Cell & bioscience*, vol. 11, no. 1, pp. 1–8, 2021. DOI: 10.1186/s13578-021-00564-x.

[325] B. T. Freitas *et al.*, "Characterization and noncovalent inhibition of the deubiquitinase and deisgylase activity of SARS-CoV-2 papain-like protease," *ACS infectious diseases*, vol. 6, no. 8, pp. 2099–2109, 2020. DOI: 10.1021/acsinfecdis.0c00168.

[326] Y. Zhao *et al.*, "High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors," *Protein & cell*, pp. 1–12, 2021. DOI: 10.1007/s13238-021-00836-9.

[327] D. J. Lenschow, "Antiviral properties of ISG15," *Viruses*, vol. 2, no. 10, pp. 2154–2168, 2010. DOI: 10.3390/v2102154.

[328] J. Hammond *et al.*, "Oral nirmatrelvir for high-risk, nonhospitalized adults with COVID-19," *New England Journal of Medicine*, 2022. DOI: 10.1056/NEJMoa2118542.

[329] L. D. Saravolatz, S. Depcinski, and M. Sharma, "Molnupiravir and nirmatrelvir-ritonavir: Oral COVID antiviral drugs," *Clinical Infectious Diseases*, 2022. DOI: 10.1093/cid/ciac180.

[330] S. T. Thun-Hohenstein *et al.*, "Structure-activity studies reveal scope for optimisation of ebselen-type inhibition of SARS-CoV-2 main protease," *ChemMedChem*, vol. 17, no. 4, e202100582, 2022. DOI: 10.1002/cmdc.202100582.

[331] L. Brewitz *et al.*, "Mass spectrometric assays reveal discrepancies in inhibition profiles for the SARS-CoV-2 papain-like protease," *ChemMedChem*, 2022. DOI: 10.1002/cmdc.202200016.

[332] V. Hugo, *Les misérables*. C. Lassalle, 1863.

Appendices

Table 1: Examples of human proteins with N-terminal prolines, retrieved from [233]. For a full list, see [233].

UniProt entry	**Gene name**
A0A087WWY3	FLNA
E9PAV3	NACA
E9PB61	ALYREF
G3V2D9	APEX1
J3KN32	RCOR1
O00458	IFRD1
O00479	HMGN4
O14602	EIF1AY
O14908	GIPC1
O14980	XPO1
O15304	SIVA1
O43747	AP1G1
O43913	ORC5
O43920	NDUFS5
O60814	HIST1H2BK
O60828	PQBP1
O75083	WDR1
O75369	FLNB
O75694	NUP155
O75821	EIF3G
O75886	STAM2
O94973	AP2A2
O95059	RPP14
O95400	CD2BP2
O95478	NSA2
O96000	NDUFB10
P01106	MYC
P04075	ALDOA
P04818	TYMS
P05114	HMGN1
P05198	EIF2S1
P08238	HSP90AB1
P09211	GSTP1
P09960	LTA4H
Q16740	CLPP

Table 2: Human proteins with N-terminal cysteines, collated by Beatrice Guintoli, Thomas P. Keeley, and Norma Masson.

UniProt entry	**Gene name**
Q93063	EXT2
Q8TD08	MAPK15
Q6ZW05	PTCHD4
Q8IVV7	GID4
Q6ICL3	TANGO2
P43489	TNFRSF4
Q6JQN1	ACAD10
Q9H3Z7	ABHD16B
Q8NFK1	GJC3
O94921	CDK14
Q96HD9	ACY3
Q9BV47	DUSP26
Q6ZRH7	CATSPERG
P49901	SMCP
Q8WWZ1	IL1F10
O75608	LYPLA1
O00198	HRK
Q06210	GFPT1
P78549	NTHL1
Q6EMK4	VASN
Q02556	IRF8
Q53SF7	COBLL1
Q92551	IP6K1
P26012	ITGB8
O14775	GNB5
Q6UVY6	MOXD1
P29460	IL12B
Q9ULL4	PLXNB3
P09326	CD48
Q00987	MDM2
P42338	PIK3CB
Q96RS0	TGS1
P60660	MYL6
Q9BZR8	BCL2L14
Q9UK05	BMP-9
B2RXH4	BTBD18
Q96A28	SLAMF9
P19022	CDH2

Le [...] riche a cent distractions brillantes et grossières, les courses de chevaux, la chasse, les chiens, le tabac, le jeu, les bons repas, et le reste ; occupations des bas côtés de l'âme aux dépens des côtés hauts et délicats. Le [...] pauvre se donne de la peine pour avoir son pain ; il mange ; quand il a mangé, il n'a plus que la rêverie. Il va aux spectacles gratis que Dieu donne ; il regarde le ciel, l'espace, les astres, les fleurs, les enfants, l'humanité dans laquelle il souffre, la création dans laquelle il rayonne.

Victor Hugo - Les Misérables [332]