Interplay Between 2-Oxoglutarate Oxygenases and Cancer: Studies on the Aspartyl/Asparaginyl-β-hydroxylase

---

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Inga Pfeffer

Somerville College, Oxford

Michaelmas Term 2014

Copyright of a thesis from the University of Oxford is reserved solely for the thesis author. Reproduction of this thesis in any form requires the consent of the author.

© Inga Pfeffer

2014

Interplay Between 2-Oxoglutarate Oxygenases and Cancer: Studies on the Aspartyl/Asparaginyl-*β*-hydroxylase

Inga Pfeffer
Somerville College, University of Oxford
Submitted for the degree of Doctor of Philosophy in Organic Chemistry
Michaelmas Term 2014

The ~60 human Fe(II)- and 2-oxoglutarate dependent oxygenases (2OG oxygenases) are involved in a range of biological processes, including collagen biosynthesis, hypoxic sensing, fatty acid metabolism/carnitine biosynthesis, and transcriptional and translational regulation. Not all of their functions have been assigned. However, many 2OG oxygenases are linked to cancer, and several human 2OG oxygenases are therapeutic targets.

The posttranslational 3*R*-hydroxylation of specific aspartyl/asparaginyl residues in epidermal growth factor-like domains (EGFDs) is catalysed by the aspartyl/asparaginyl-*β*-hydroxylase (AspH), a 2OG oxygenase. The physiological role of EGFD hydroxylation by AspH is unclear. However, an oxygenase domain-containing long isoform of AspH and its shorter isoform, humbug, lacking the oxygenase domain, are both overexpressed in various types of cancer, enabling the use of AspH as a tumour biomarker.

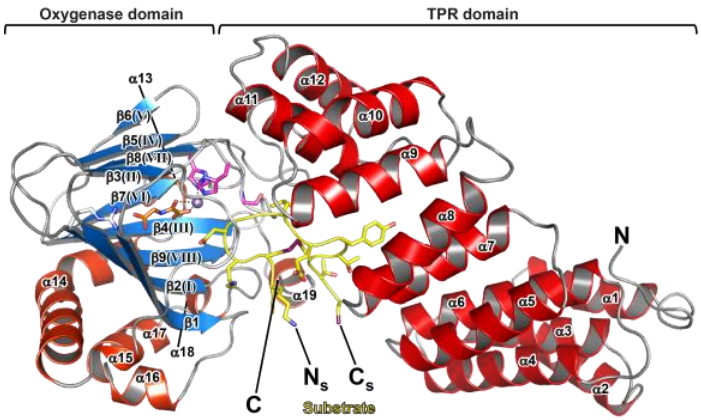


Fig. 1 | View from a crystal structure of human AspH$_{315\text{-}758}$ in complex with the first EGF-like domain of human coagulation factor X. The AspH-substrate interactions are highly specific and involve both the oxygenase and the TPR domain.

The work in this Thesis describes how structural and functional studies helped elucidate the unanticipated substrate selectivity of AspH. The results reveal a unique active site geometry (a catalytic dyad HXG…H motif instead of the HXD/E…H 'facial triad' typical of human 2OG oxygenases) and a substrate binding mode involving at least two domains of the protein. The AspH oxygenase domain alone shows no hydroxylation activity. A longer active protein including a tetratricopeptide repeat (TPR) domain is required for substrate recognition. A surprising finding was that AspH only hydroxylates EGFDs containing a non-canonical disulfide bond, suggesting roles for AspH in either quality control of EGFD folding and/or signalling in the unfolded protein response/Ca^{2+} homeostasis in the endoplasmic reticulum.

Overall, this work provides the first detailed investigation into the structure and biochemical function of human AspH and its substrate interactions. The results will form a basis for future studies towards understanding the physiological role of AspH and for the development of selective probes and inhibitors, ultimately aimed at cancer diagnosis and treatment.

Für meine Familie

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **XI**

ABBREVIATIONS ..... **XII**

1 INTRODUCTION TO 2-OXOGLUTARATE OXYGENASES AND CANCER ..... **XV**

1.1 CANCER **2**

1.1.1 CANCER CELL METABOLISM 2

1.2 2-OXOGLUTARATE OXYGENASES **3**

1.2.1 STRUCTURAL FEATURES AND MECHANISM OF 2OG OXYGENASES 4
1.2.2 SUBCLASSES OF SELECTED HUMAN 2OG OXYGENASES 6
1.2.2.1 Collagen Biosynthesis *12*
1.2.2.2 HIF Hydroxylation: Hypoxic Sensing *14*
1.2.2.3 Epigenetic Regulation *15*
1.2.2.4 Nucleic Acid Modification *16*
1.2.2.5 Ribosomal Protein Hydroxylation *17*

1.3 EPIDERMAL GROWTH FACTOR-LIKE DOMAIN HYDROXYLATION **18**

1.3.1 EPIDERMAL GROWTH FACTOR-LIKE DOMAINS 19
1.3.2 LINKS BETWEEN ASPH AND DISEASES 22
1.3.3 GENE ORGANISATION AND ASPH SPLICING ISOFORMS 23

1.4 THESIS OBJECTIVES **25**

2 FORM EVER FOLLOWS FUNCTION: STRUCTURAL AND FUNCTIONAL STUDIES OF ASPH ..... **XXVII**

2.1 ASPH *IN VITRO* CONSTRUCT DESIGN **28**

2.1.1 OXYGENASE DOMAIN PROTEIN PRODUCTION 29
2.1.2 TWO-DOMAIN CONSTRUCT PROTEIN PRODUCTION 30

2.2 ASPH *IN VITRO* ACTIVITY ASSAYS **31**

2.2.1 DEVELOPMENT OF AN MS-BASED HYDROXYLATION ACTIVITY ASSAY 31
2.2.1.1 AspH in vitro Assay Substrate *32*
2.2.1.2 MALDI-ToF-MS Hydroxylation Activity Assay *33*
2.2.2 2OG TURNOVER ASSAYS 34
2.2.2.1 [1-^{14}C]-2OG Turnover Assays *34*
2.2.2.2 Nuclear Magnetic Resonance 2OG Turnover Assays *36*

2.3 CONSTRUCT OPTIMISATION **37**

2.3.1 PROTEIN PRODUCTION AND ACTIVITY OF His_6-$AspH_{315-758}$ CONSTRUCT 38
2.3.1.1 In vitro Transcription/Translation *38*
2.3.1.2 Expression Trials for the Production of Recombinant His_6-$AspH_{315-758}$ *39*
2.3.1.3 Small-Scale Trial Purification *41*
2.3.1.4 Large-Scale Purification *42*
2.3.1.5 His_6-$AspH_{315-758}$ Hydroxylation Activity and 2OG Turnover Assays *43*

2.3.2 COMPARISON OF THE CONSTRUCTS 44
2.3.2.1 Hydroxylation Activity 44
2.3.2.2 Analytical Gel Filtration 45
2.4 ASPH CRYSTALLOGRAPHY 46
2.4.1 OXYGENASE DOMAIN STRUCTURE ANALYSIS: $ASPH_{562-758}$ 46
2.4.2 CRYSTALLOGRAPHIC STUDIES ON $ASPH_{315-758}$ 49
2.4.2.1 Crystal Structure Determination 51
2.4.2.2 Structure Analysis "Apo"-$AspH_{315-758}$ 52
2.4.2.3 Protein-Peptide Interaction Pre-Crystallisation Trial 54
2.4.3 CRYSTAL STRUCTURE OF $ASPH_{315-758}$ IN COMPLEX WITH A SUBSTRATE 55
2.4.3.1 AspH Binds Its Substrate with an "Induced Fit" Mechanism 58
2.4.3.2 Key Interactions Between AspH and Its Substrate 60
2.5 DISCUSSION 63
2.5.1 AN UNUSUAL FAMILY MEMBER WITH UNIQUE ACTIVE SITE GEOMETRY 64
2.5.2 A NON-CANONICAL DISULFIDE BOND IN THE SUBSTRATE 67
2.6 ACKNOWLEDGEMENTS FOR CHAPTER 2 69
3 ASPH REQUIRES EGFD SUBSTRATES WITH A NON-CANONICAL CYS3-4 DISULFIDE FOR ACTIVITY LXX
3.1 CHARACTERISATION OF THE HFX $EGF1_{39MER}$ SUBSTRATE 71
3.1.1 SOLUTION EVIDENCE FOR AN UNUSUAL DISULFIDE PATTERN 73
3.1.1.1 'Bottom-Up' MS/MS for hFX $EGF1_{39mer}$ Hydroxylation Site Identification 74
3.1.1.2 'Top-Down' MS/MS Studies on the hFX $EGF1_{39mer}$ with Intact Disulfides 75
3.1.1.3 Trypsin Digestion of the hFX $EGF1_{39mer}$ under Non-Reducing Conditions 78
3.2 STUDIES ON ASPH SUBSTRATE REQUIREMENTS 81
3.2.1 EXPERIMENTS WITH CANONICAL CBEGFD PROTEINS 81
3.2.1.1 Hydroxylation Assays, Trypsin Digestion and MS/MS 83
3.2.2 ASPH COSUBSTRATE AND COFACTOR DEPENDENCE 89
3.2.3 EXPERIMENTS WITH DIFFERENT BUFFER SYSTEMS 90
3.2.3.1 Optimisation of Assay and Crystallisation Conditions 91
3.2.4 DISULFIDE REQUIREMENTS 94
3.2.4.1 Characterisation and Disulfide Pattern Analyses 95
3.2.4.2 Hydroxylation Assays 99
3.3 DISCUSSION 100
3.3.1 CANONICAL VS. NON-CANONICAL DISULFIDE PATTERN 102
3.4 ACKNOWLEDGEMENTS FOR CHAPTER 3 104
4 CYCLIC PEPTIDES FOR ASPH CV
4.1 DESIGN OF A CYCLIC PEPTIDE SUBSTRATE 106
4.1.1 CYCLIC PEPTIDE SYNTHESIS 107
4.1.2 THE CYCLISED CONFORMATION IS CRITICAL FOR ACTIVITY 108
4.2 EXPLORING THE CYCLIC PEPTIDE SCAFFOLD 112
4.2.1 SUBSTRATE SCOPE AND SELECTIVITY 112

4.2.1.1 Cyclic Peptides Based on hFX EGF1$_{39mer}$ 113
4.2.1.2 Cyclic Peptides Based on Other EGFDs 120
4.2.2 Cyclic Peptide Conjugates 121
4.3 COMPETITION EXPERIMENTS WITH ASPH SUBSTRATES 127
4.4 DISCUSSION 131
4.4.1 Substrate Requirements and Disulfide Plasticity 131
4.5 ACKNOWLEDGEMENTS FOR CHAPTER 4 133
5 THE ROLE OF THE TPR DOMAIN IN SUBSTRATE RECOGNITION CXXXIV
5.1 CRYSTALLOGRAPHIC STUDIES ON ASPH-SUBSTRATE COMPLEXES 135
5.1.1 Crystal Structures of AspH$_{315-758}$ in Complex with NC-4Ser$_{39mer}$ and NC$_{26mer}$ 136
5.1.1.1 Crystallography Reveals Further Key Interactions for Binding to AspH 136
5.1.2 Complex Formation and Crystallisation of AspH with a Cyclic Peptide 139
5.1.2.1 Optimisation of the Cyclic Peptide 139
5.1.2.2 Binding Studies Using Differential Scanning Fluorimetry 142
5.1.2.3 Binding Studies Using Non-Denaturing Mass Spectrometry 143
5.1.2.4 Crystal Structure of AspH$_{315-758}$ in Complex with CP2$_{19mer}$ 152
5.2 MUTAGENESIS STUDIES 155
5.3 DISCUSSION 161
5.3.1 Disulfide Plasticity and Cooperativity 162
5.3.2 AspH Is a Highly Specific but Tolerant Enzyme 163
5.5 ACKNOWLEDGEMENTS FOR CHAPTER 5 163
6 INTERFERENCE FROM ASPARTYL SUCCINIMIDES IN MS-BASED PEPTIDE ASSAYS CLXIV
6.1 INTERFERENCE IN MS ANALYSES OF HYDROXYLATION ASSAYS 165
6.1.1 Aspartyl Succinimide Formation 166
6.2 ASU IS FORMED IN SOLUTION AND NOT DUE TO IN-SOURCE FRAGMENTATION 167
6.2.1 Analysis of Asu in Protein Standards 168
6.3 DISCUSSION 173
6.4 ACKNOWLEDGEMENTS FOR CHAPTER 6 173
7 CA^{2+} PROMOTES CANONICAL EGFD DISULFIDE FORMATION CLXXIV
7.1 THE PHYSIOLOGICAL ROLES OF CA^{2+} 175
7.1.1 The Hydroxylation Site Residue in Ca^{2+}-Binding EGFDs 176
7.1.2 Ca^{2+}-Binding of EGFDs 176
7.2 THE EFFECTS OF CA^{2+} ON HYDROXYLATION 176
7.2.1 Effects of Ca^{2+} on Hydroxylation of cbEGFD Proteins 176
7.2.2 Effects of Ca^{2+} on Hydroxylation of the Synthetic EGFD Peptides 181

7.3 DISCUSSION 185

7.4 ACKNOWLEDGEMENTS FOR CHAPTER 7 186

8 CELL-BASED STUDIES OF ASPH CLXXXVII

8.1 PROPERTIES OF THE DISTINCTIVE ASPH SPLICING ISOFORMS 188

8.1.1 PROTEOLYTIC PROCESSING OF ASPH AND SUBCELLULAR LOCALISATION 188

8.2 SUBCELLULAR LOCALISATION STUDIES ON FLAG-TAGGED ASPH 189

8.2.1 PREPARATION OF FLAG-TAGGED *ASPH* CONSTRUCTS 190
8.2.2 IMMUNOFLUORESCENCE STUDIES 192
8.2.2.1 Localisation of N- or C-FLAG-Tagged flAspH, Humbug and AspH$_{315-758}$ 192

8.3 WORK TOWARDS LONGER ASPH *IN VITRO* PROTEINS 196

8.4 DISCUSSION 197

8.4.1 HIGH SIMILARITY BETWEEN FLASPH AND JUNCTATE 199
8.4.2 ASPH HOMOLOGUES ASPHD1 AND ASPHD2 202

8.5 ACKNOWLEDGEMENTS FOR CHAPTER 8 203

9 SUMMARY, CONCLUSIONS AND FUTURE PROSPECTS CCIV

9.1 SUMMARY AND CONCLUSIONS 205

9.2 FUTURE PROSPECTS 206

10 METHODS AND MATERIALS CCVIII

10.1 GENERAL EXPERIMENTAL PROCEDURES 209

10.1.1 CHEMICALS, ANTIBIOTICS, MOLECULAR MARKERS 209
10.1.2 ASPH *IN VITRO* SUBSTRATES AND EGFD PEPTIDES 209
10.1.3 MICROBIOLOGICAL TECHNIQUES 211
10.1.4 WATER BATHS 212
10.1.5 PH MEASUREMENT 212
10.1.6 BACTERIAL GROWTH MEDIA 212
10.1.7 CENTRIFUGATION AND INCUBATION 212
10.1.8 OPTICAL DENSITY MEASUREMENTS (OD_{600}) 213
10.1.9 LYOPHILISATION 213

10.2 MOLECULAR BIOLOGY TECHNIQUES 213

10.2.1 VECTORS 213
10.2.2 PREPARATION OF OLIGONUCLEOTIDES 214
10.2.3 DNA AMPLIFICATION BY POLYMERASE CHAIN REACTIONS 214
10.2.4 SITE-DIRECTED MUTAGENESIS 214
10.2.5 *DPN*I DIGESTION OF THE PCR PRODUCTS 216
10.2.6 DNA GEL ELECTROPHORESIS 216
10.2.7 PLASMID DNA PREPARATION AND GEL EXTRACTION 217
10.2.8 DNA RESTRICTION DIGEST AND SUBCLONING 217
10.2.9 DNA LIGATION 217
10.2.10 LIGATION-INDEPENDENT CLONING (LIC) 218
10.2.11 DNA SEQUENCING 220

10.2.12 COMPETENT CELL STRAINS 220
10.2.13 COMPETENT CELL PREPARATION 220
10.2.14 TRANSFORMATIONS 221
10.2.15 GLYCEROL STOCKS 221

10.3 PROTEIN PRODUCTION 221

10.3.1 STARTER CULTURE PREPARATION AND LARGE-SCALE GROWTHS 221
10.3.2 EXPRESSION TRIALS FOR SMALL-SCALE PROTEIN PRODUCTION 222
10.3.3 LARGE-SCALE CELL LYSATE PREPARATION 222
10.3.4 SMALL-SCALE PROTEIN PURIFICATION TRIALS 223
10.3.5 VISUALISATION OF HIS-TAGGED PROTEINS USING INVISION™ HIS-TAG IN-GEL STAIN 223
10.3.6 LARGE-SCALE PROTEIN PURIFICATION 223
10.3.7 PROTEIN CONCENTRATING, BUFFER EXCHANGE AND PROTEIN CONCENTRATION DETERMINATION 225
10.3.8 SDS-PAGE ANALYSIS OF PROTEIN FRACTIONS 225
10.3.9 NATIVE PAGE 226
10.3.10 TRIS-TRICINE GEL ANALYSIS 227
10.3.11 HIS_6-TAG CLEAVAGE 227

10.4 SYNTHETIC PROCEDURES 228

10.4.1 ORGANIC SYNTHESIS 228
10.4.1.1 Chloromethylcarbonyloxysuccinimide (ClAc-OSu) 228
10.4.1.2 General Procedure for the Synthesis of Fluorescence Dye-BCN-Conjugates 228
10.4.1.3 General Procedure for Strain-Promoted Azide-Alkyne Cycloaddition 229
10.4.2 CYCLIC PEPTIDE DESIGN 229
10.4.3 SOLID-PHASE PEPTIDE SYNTHESIS 229
10.4.4 PEPTIDE CYCLISATION REACTION AND HPLC PURIFICATION 230

10.5 ANALYTICAL AND BIOPHYSICAL ASSAY METHODS 232

10.5.1 MASS SPECTROMETRY-BASED ASPH HYDROXYLATION ACTIVITY ASSAYS 232
10.5.2 MALDI-TOF-MS AND MS/MS 233
10.5.3 MS/MS DATA ACQUISITION AND PROCESSING 234
10.5.4 LC-MS ANALYSIS OF EGFD PEPTIDES AND PROTEINS 235
10.5.5 IN-SOLUTION DIGESTION OF PROTEINS WITH PROTEASES 236
10.5.6 SOLID-PHASE EXTRACTION (SPE) 237
10.5.7 DENATURING PROTEIN MASS SPECTROMETRY 237
10.5.8 NON-DENATURING (NATIVE) MASS SPECTROMETRY 238
10.5.9 DIFFERENTIAL SCANNING FLUORIMETRY 238
10.5.10 CRYSTALLISATION METHODS 239
10.5.11 STRUCTURE DETERMINATION 240
10.5.12 [1-^{14}C]-2OG TURNOVER ASSAYS 240
10.5.13 2OG TURNOVER ASSAYS BY NMR 241
10.5.14 NMR ANALYSIS OF HYDROXYLATED $CP2_{10MER}$ 241

10.6 MAMMALIAN CELL-BASED METHODS 241

10.6.1 MAMMALIAN CELL CULTURE 241
10.6.1.1 Cell Storage 242
10.6.1.2 Cell Passaging 242
10.6.2 TRANSFECTION 243
10.6.3 IMMUNOFLUORESCENCE 243
10.6.4 CDNA PREPARATION 244

11 APPENDICES CCXLV

A PROTEIN AND DNA SEQUENCES 246

B SUPPLEMENTARY ASPH ISOFORMS SEQUENCE ALIGNMENTS 246

REFERENCES 252

ACKNOWLEDGEMENTS

Many thanks to Prof Chris Schofield and Prof Ben Davis for giving me the opportunity and freedom to work on projects in very interesting research areas at the interface of chemistry, biology and medicine, for allowing me to learn so many new things, and for their mentorship and guiding me through the absolutely exciting years of my postgraduate studies. It has been a great pleasure working with you.

Thanks to both of my lab groups – the CJS group and the BGD group, past and present – for the often great working atmosphere, all the help and the fruitful discussions in the lab and outside. Many people helped me get started, introduced me to new methods and techniques, answered all my questions and were on the spot when I needed advice. I am extremely grateful and would like to thank all of you wholeheartedly.

A big thank-you to all the people that contributed to this Thesis by collaborating, providing training, materials, or fruitful discussions. Where portions of this Thesis were collaborative efforts, I have taken great care to attribute the appropriate credit to the contributors. Their contribution is stated explicitly in the text and in the acknowledgements at the end of each Chapter.

I am particularly grateful to Dr Michael McDonough and Dr Holger Kramer for their tremendous support and guidance over the past years, for being such inspiring minds and for their never-ending enthusiasm. Particular thanks also to Hanna Tarhonskaya for her great friendship and support, and for making the past years so enjoyable inside and outside the lab. In addition, I am also very grateful to my other fabulous people around – Dr Carmen Domene, Kuan-Jung 'Erika' Wu, Isabella Hüttner, Thomas Leißing, Lukas Lercher, Reida Rutte, Dr Rachael Bashford Rogers – who were always there for a good chat, brought me a coffee or treat, and helped me keep my head above the water. Many thanks to Dr Sander van Berkel for all his help and the great discussions related to synthesis or life, and to Dr Jürgen Brem for his enthusiasm and passion, to Dr Rasheduzzaman Chowdhury for always being on the spot to help with many bits and pieces, to Dr Joan Gannon for the Irish quality time with really enjoyable conversations, to Dr Martin Münzel, Dr Richard Hopkinson, Anne Makena, and Dr Anna Rydzik for the fruitful collaboration, to Dr Eleanor Bagg for my beloved lab rack, and to Dr Emily Flashman, a fabulous desk neighbour in my first year who has always been incredibly helpful. A thank-you also to my other office or lab mates, in particular Dr Shoichiro Horita for his help and the funny conversations, to Gareth Langley, Saiful Islam, Dr Armin Thalhammer, Dr Ilaria Pettinati and Dr Olly King for keeping the atmosphere as nice as possible.

I am very grateful to Mrs Val DeNewtown, Mrs Jenny Houlsby, and Mrs Anne Gough for their never-ending support not only in bureaucratic matters. Many thanks also to Prof Chris Schofield, Dr Carmen Domene, Dr Michael McDonough, Dr Holger Kramer and Dr Wilfried Hess for critical proofreading of this Thesis – for their time, consideration and advice.

Thanks also to all support staff and lab managers that kept the labs running smoothly, in particular Mrs Amy Webb. Many thanks to Mr Herminio Mansojubier for not only massively speeding up my progress by providing media but also for the enjoyable chats here and there. Many thanks also to Prof Penny Handford and her coworkers Dr Sacha Jensen and Dr Paul Taylor for their support and supply with cbEGFD proteins. Thanks to Dr Adam Hardy for initial help getting me started in the lab. Dr Holger Gerlach, Dr Leon Harrington, and Dr Mercedes Romero-Ruiz are kindly acknowledged for sharing not only their advice but also some "exotic" materials for various lab techniques. Thanks to the CRL mass spectrometry, in particular Dr James Wickens for his kind assistance and for sharing his profound knowledge with me in very fruitful discussions. Additional thanks to Ian Miller from OUCS who has also been incredibly nice and helpful.

I would also like to express my gratitude to all my other collaborators and supporters beyond the work presented in this Thesis, including the researchers at the SGC Oxford, the Wellcome Trust Centre for Cellular and Molecular Physiology (CCMP), the Clinical Neurology at the John Radcliffe Hospital (CNJR), and the whole Oxford Cancer Imaging Centre (OCIC), in particular Professor Veronique Gouverneur, who has been highly supportive. Additional thanks go to my enGENEious Conference 2012 team, to the whole Athena SWAN committee – past and present, to my YES Biotechnology Competition team members, and to my Somerville College committee fellows.

Finally, I would like to say thank-you to my family, friends, and loved ones in Germany for your patience, understanding, and encouragement throughout my DPhil. This would not have been possible without you!

ABBREVIATIONS

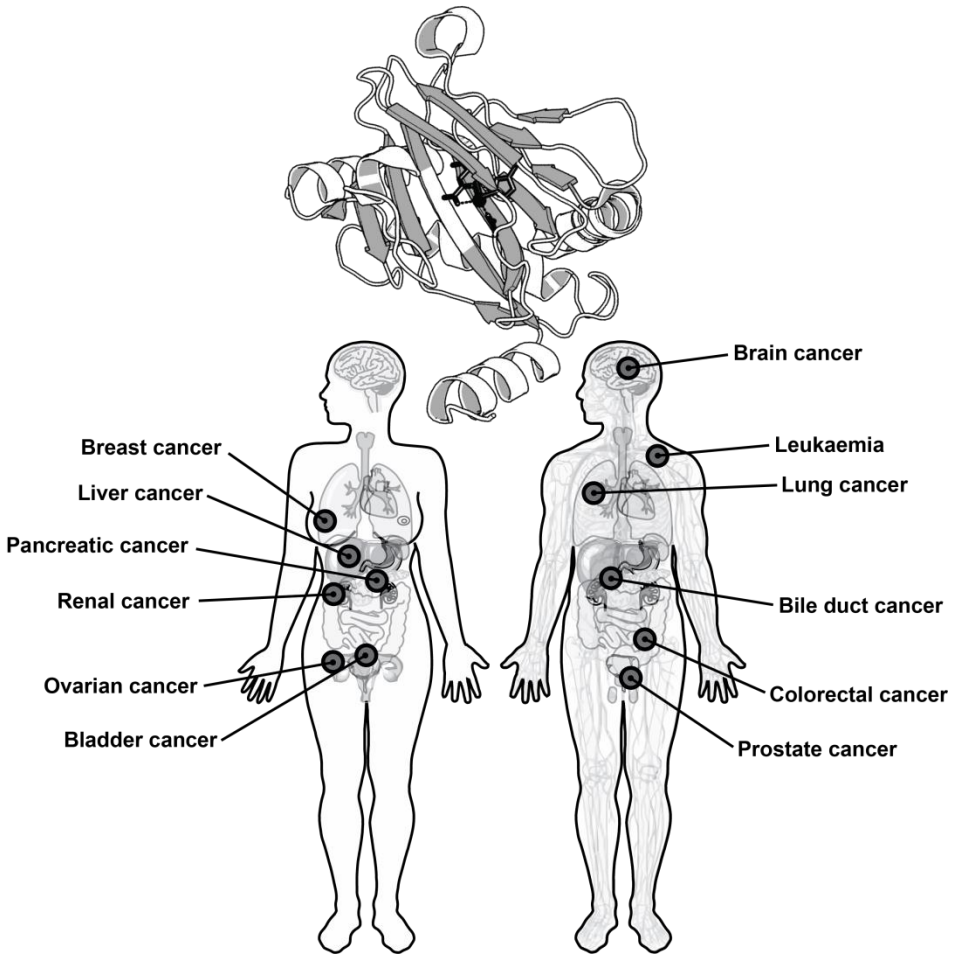
$^1O_2$	Singlet dioxygen
2,4-PDCA	2,4-Pyridine dicarboxylic acid
2HG	2-Hydroxyglutarate
2OA	2-Oxoadipate
2OG	2-Oxoglutarate
2TY	2x Tryptone/Yeast extract medium
$^3O_2$	Triplet dioxygen
αCalnexin	anti-Calnexin antibody
αFLAG	anti-FLAG-tag antibody
αPDI	anti-protein disulfide isomerase antibody
aa	Amino acids
ABH	AlkB human homologue
ACN	Acetonitrile
AlkB	Alkylated DNA repair protein B
Ambic	Ammonium bicarbonate
APS	Ammonium persulfate
Asc	Sodium ascorbate
AspH	Aspartyl/Asparaginyl-*β*-hydroxylase
$AspH_{562-758}$	His_6-tagged 2OG oxygenase domain of human AspHA (aa residues 562-758)
$AspH_{315-758}$	His_6-tagged construct of human AspHA (aa residues 315-758) containing TPR and 2OG oxygenase domains
AspHA	Aspartyl/Asparaginyl-*β*-hydroxylase isoform A
Asu	Aspartyl succinimide
Asx	Aspartic acid or asparagine residue
*β*Hya	*β*-Hydroxyaspartate
*β*Hyn	*β*-Hydroxyasparagine
BBOX	*γ*-Butyrobetaine hydroxylase
BCN	Bicyclo[6.1.0]nonyne
Bis-Tris	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
bp	Base pairs
BSA	Bovine serum albumin
C_{26mer}	Canonical 26mer peptide based on coagulation factor X EGF1
C_{39mer}	Canonical 39mer peptide based on coagulation factor X EGF1
Ca^{2+}	Calcium(II) ions
CAM	*S*-Carbamidomethyl(-cysteine)
CAPSO	3-(Cyclohexylamino)-1-propanesulfonic acid
cb	Calcium-binding
cbEGF	Calcium-binding epidermal growth factor-like domain
cDNA	Complementary DNA
CHCA	*α*-Cyano-4-hydroxy-cinnamic acid
$CHCl_3$	Chloroform
CHES	2-(Cyclohexylamino)ethanesulfonic acid
cHex	Cyclohexane
CID	Collision-induced dissociation
ClAc-OSu	Chloromethylcarbonyloxysuccinimide
CoA	Coenzyme A
CODD	*C*-terminal oxygen-dependent degradation domain
CP_{10mer}	Cyclic peptide 10mer
CP_{15mer}	Cyclic peptide 15mer (branched cyclic peptide)
CP_{19mer}	Cyclic peptide 19mer (branched cyclic peptide)
cpm	Counts per minute
CRL	Chemistry Research Laboratory
CV	Column volume
Cys^{CAM}	*S*-Carbamidomethylated cysteine
Cys^{NES}	*S*-(*N*-Ethylsuccinimido)-cysteine
δ	Chemical shift in NMR
^{D}Ala	*R*-Alanine
^{D}AzidoAla	*R*-Azidoalanine
Da	Dalton (unit of molecular mass)
DAPI	4',6'-Diamidino-2-phenylindole
DCM	Dichloromethane
dCTP	Deoxycytidine triphosphate
DFO	Deferoxamine
dGTP	Deoxyguanosine triphosphate
DHB	2,5-Dihydroxybenzoic acid
DIC	Di*iso*propylcarbodiimide
DM2OG	Dimethyl 2-oxoglutarate
DMEM	Dulbecco's modified Eagle's medium
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dpm	Disintegrations per minute
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E. coli	*Escherichia coli*
e.g.	For example (*exempli gratia*)
EDTA	Ethylenediaminetetraacetic acid
EE	Ethyl acetate
EGF	Epidermal growth factor
EGFD	Epidermal growth factor-like domain
EGFH	Epidermal growth factor hydroxylase
EGTA	Ethylene glycol-bis(2-aminoethyl-ether)-*N,N,N',N'*-tetraacetic acid
EIC	Extracted-ion chromatogram
EPPS	3-[4-(2-hydroxyethyl)piperazin-1-yl]propane-1-sulfonic acid
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
ES	Embryonic stem cells
ESI	Electrospray ionisation
et al.	*et alii* (and coworkers)
Et_2O	Diethylether
EtOH	Ethanol
FA + Tx	3.7 % paraformaldehyde and 0.1 % Triton X 100
FBS	Fetal bovine serum
FBXL11	F-box and leucine-rich repeat protein 11
Fe(II) or Fe^{II}	Ferrous iron(II) ions
FH	Fumarate hydratase
Fig.	Figure
FIH	Factor inhibiting HIF
FITC	Fluorescein *iso*thiocyanate
FLAG	3 x FLAG tag sequence
flAspH	Full-length human Aspartyl/asparaginyl-*β*-hydroxylase isoform A (aa residues 1-758)
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography

FSG	Fish skin gelatin
FWHM	Full width at half maximum (resolution)
GluC	Endoproteinase GluC (*Staphylococcus aureus* Protease V8)
Glx	Glutamic acid or glutamine residue
GSH	L-Glutathione reduced
GSSG	Glutathione disulfide (L-(–)Glutathione oxidised)
GST	Glutathione *S*-Transferase
GST-AspH$_{315\text{-}758}$	GST-tagged construct of human AspHA (aa residues 315-758, made by *ReOx*) containing TPR- and 2OG oxygenase domains
h	Hour(s)
HABA	2-(4'-Hydroxybenzeneazo)benzoic acid
HCl	Hydrochloric acid
HEK 293T	Human embryonic kidney cell line
HEPES	4-(2-Hydroxyethyl)-1-piperazine-ethanesulfonic acid
HepG2	Human liver carcinoma cell line
HeLa	Henrietta Lacks human cervical cancer cell line
HESI	Heated electrospray source ionisation
hFX	Human coagulation factor X
hFib1	Human fibrillin 1
hFX EGF1$_{39mer}$	Synthetic peptide representing the first EGF-like domain of coagulation factor X EGF1 (mixture of the canonical and non-canonical 39mer peptides resulting from air oxidation)
HIF	Hypoxia-inducible factor
His$_6$	Hexahistidine protein tag
hNotch1	Human notch 1
HPLC	High-performance liquid chromatography
HRE	Hypoxia response element
HR-ESI-MS	High resolution mass spectrometry (or spectrum) with electrospray ionisation
HRP	Horseradish peroxidase
HRV	Human rhinovirus
HSQC	Heteronuclear single-quantum correlation spectroscopy
i.d.	Internal diameter
i.e.	That is (*id est*)
IAA	Iodoacetic acid
IAM	Iodoacetamide
IC_{50}	Half-maximal inhibitory concentration
IDH	*Iso*citrate dehydrogenase
IgG	Immunoglobulin G
IMAC	Immobilised metal affinity chromatography
IP	Immunoprecipitation
IPTG	*Iso*propyl-β-D-thiogalactopyranoside
J	Coupling constant (NMR, MRS)
JARID	Jumonji (A+T)-rich interactive domain
JmjC	Jumonji-C domain
JMJD	Jumonji-C domain-containing protein
kb	Kilobases (DNA)
K_D	Dissociation constant
kDa	Kilodalton
K_M	Michaelis-Menten constant
l	Litre(s)
LB	Lysogeny broth
LC	Liquid chromatography
LIC	Ligation-independent cloning
m/z	mass-to-charge ratio
MALDI-ToF	Matrix-assisted laser desorption/ionisation-time of flight
MEFs	Mouse embryonic fibroblasts
MeOH	Methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
Mg^{2+}	Magnesium(II) ions
min	Minutes
MM2OG	Monomethyl 2-oxoglutarate
Mn(II) or Mn^{II}	Manganese(II) ions
MR	Molecular replacement
MRI	Magnetic resonance imaging
MRM	Multiple reaction monitoring
MRS	Magnetic resonance spectroscopy
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MWCO	Molecular weight cut-off
N/A	Not applicable
NAD^+/NADH	Nicotinamide adenine dinucleotide, oxidised and reduced form
$NADP^+$/NADPH	Nicotinamide adenine dinucleotide phosphate, oxidised and reduced form
NC	Non-canonical (EGF-like domain)
NC-4Ser$_{39mer}$	Non-canonical 39mer peptide based on coagulation factor X EGF1 with a single disulfide with Cys3-4 linkage, other four cysteine residues replaced by serine residues
NC$_{26mer}$	Non-canonical 26mer peptide based on Coagulation Factor X EGF1
NC$_{39mer}$	Non-canonical 39mer peptide based on coagulation factor X EGF1
ND	Not determined
NEM	*N*-Ethylmaleimide
NES	*S*-(*N*-Ethylsuccinimido)(-cysteine)
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NOG	*N*-Oxalylglycine
Ni^{2+} or Ni^{II}	Nickel(II) ions
NOBA	*N*-Oxalyl-β-alanine
NODD	*N*-Terminal oxygen-dependent degradation domain
NP40	Tergitol-type NP40 (nonyl phenoxy-polyethoxylethanol)
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at 600 nm
Orai1	Calcium release-activated calcium channel protein 1
ox.	Oxidised
PAGE	Polyacrylamide gel electrophoresis
ppm	Parts per million
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank (Research Collaboratory for Structural Bioinformatics, RCSB)
PDI	Protein disulfide isomerase
PEG	Polyethylene glycol
Pfu	*Pyrococcus furiosus*
PHD	Prolyl hydroxylase domain

pI	Isoelectric point
PTM	Posttranslational modification
pVHL	von Hippel-Lindau tumour suppressor protein
R-2HG	*R*-2-Hydroxyglutarate
red.	Reduced
RedOx Buffer	Oxido-shuffling buffer system that enables disulfide exchange
R_f	Retardation factor (thin-layer chromatography)
RNA	Ribonucleic acid
mRNA	Messenger RNA
ROS	Reactive oxygen species
rpm	Revolutions per minute
s	Seconds
S-2HG	*S*-2-Hydroxyglutarate
SA	Sinapinic acid
SAD	Single-wavelength anomalous diffraction
Sap	Saponin
SD	Standard deviation
SDH	Succinate dehydrogenase
SDS	Sodium dodecyl sulfate
SGC	Structural Genomics Consortium
SOC	Super optimal broth with catabolite repression
SOCE	Store-operated Ca^{2+} entry
SPPS	Solid-phase peptide synthesis
SR	Sarcoplasmic reticulum
STIM1	Stromal Interaction Molecule 1
SyrB2	Syringomycin halogenase
TAE	Tris-acetate-EDTA
TAPS	*N*-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid
TB	Transforming growth factor-β binding protein-like
TBST	Tris-buffered saline-Tween 20
TCA	Tricarboxylic acid
TCAI	Tricarboxylic acid cycle intermediates
TCEP	*Tris*(2-carboxyethyl)phosphine
TEMED	*N,N,N',N'*-Tetramethylethylene-diamine
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TIC	Total ion current chromatogram
TIPS	Tri*iso*propylsilane
TLC	Thin-layer chromatography
T_m	Melting temperature
TPR	Tetratricopeptide repeat
Tris	*Tris*(hydroxymethyl)aminomethane
U-2 OS	Human osteosarcoma cell line
UPLC	Ultra-performance liquid chromato-graphy
UV	Ultraviolet
v/v	Volume by volume
VEGF	Vascular endothelial growth factor
VOCC	Voltage-operated Ca^{2+} channel
w/v	Weight by volume
WHO	World Health Organisation
WT	Wildtype

Standard abbreviations for proteinogenic amino acids, nucleotides, SI units, prefixes, and the chemical elements are used throughout.

Introduction to 2-Oxoglutarate Oxygenases and Cancer



Abstract

This Chapter introduces the superfamily of 2-oxoglutarate-dependent oxygenases and describes how they are linked to cancer.

1.1 CANCER

The word cancer describes a complex class of diseases. Fundamentally, cancer is a disease of regulation of cell growth;[1-4] many cancer cells show no response to signals that normally regulate growth and division. For transformation of a normal cell into a cancer cell, genetic changes affecting cell growth and differentiation need to occur.[3] Thus, some cancer cells apparently gain near limitless replicative potential unless they are checked. The differences that define a cancer cell from a normal cell can be as simple as a single nucleotide change or more major changes such as loss of multiple nucleotides or entire chromosomes.[3] Cancer types can be very different from each other, comparing for example solid tumours with leukaemia, and may use diverse mechanisms for the cell growth and survival.[5]

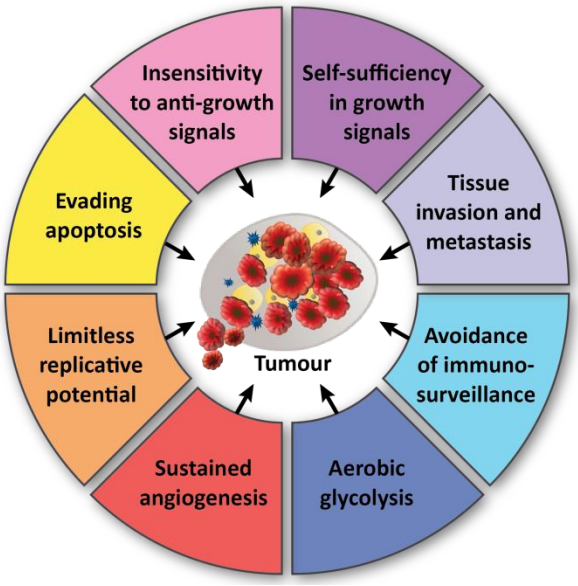


Fig. 1-1 | The hallmarks of cancer describing functional capabilities of tumours acquired during their development. Adapted from Hanahan *et al.*[1,2]

Hanahan and Weinberg have proposed hallmarks of cancer that categorise the biology of tumours (**Fig. 1-1**).[1,6] Specific characteristics of cancer cells are self-sufficiency in growth signals,[7-11] insensitivity to anti-growth signals,[12-15] evading programmed cell death (apoptosis),[16-21] tissue invasion and metastasis,[22-25] avoiding immuno-surveillance,[26-28] limitless replicative potential,[29,30] sustained angiogenesis[31] and aerobic glycolysis.[32,33]

1.1.1 Cancer Cell Metabolism

A distinct characteristic of cancer cells is their (at least often) different cell metabolism that is proposed to enable rapid proliferation and tumour growth. Cancer cells are proposed to be wasteful and to use a disproportionate share of the nutrients in their environment, compared with normal cells.[34] Normal differentiated cells rely primarily on mitochondrial oxidative phosphorylation to generate the energy needed for cellular processes.[35,36] In contrast, cancer cells employ the less efficient mechanism of aerobic glycolysis to metabolise glucose (**Fig. 1-2**), a phenomenon termed the 'Warburg effect'.[37-39]

Cancer cells benefit from their abnormal metabolism because it ensures that they have a ready supply of the building blocks needed for macromolecule synthesis, including NADPH and acetyl-CoA, in addition to ATP.[34] Furthermore, cancer cells apparently starve (and so weaken) their neighbours by claiming more than their fair share of nutrients, therefore gaining space for their

growth. Another effect of the excessive nutrient uptake is the increase of generation of reactive oxygen species.[40,41] These are proposed to promote cancer cell proliferation by inactivating growth-inhibitory phosphatase enzymes, and enhance the mutation rate by inducing DNA damage.[2]

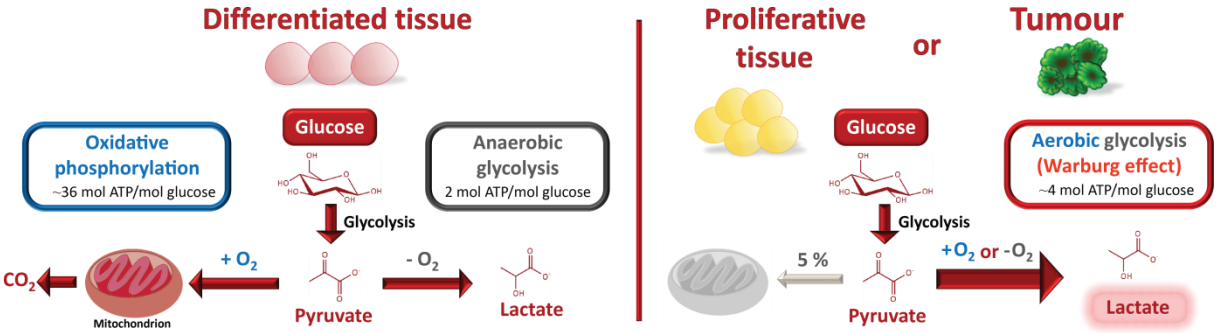


Fig. 1-2 | Overview of assigned differences between the metabolism of normal differentiated and cancer tissue. If oxygen is available, nonproliferating (differentiated) tissues first convert glucose to pyruvate via glycolysis and then completely oxidise most of the pyruvate to CO_2 via oxidative phosphorylation in the mitochondria, generating 36 ATPs per glucose.[35,36] Oxygen is essential for this process because it is the final electron acceptor. When O_2 is limited, pyruvate generated by glycolysis is metabolised to lactate (anaerobic glycolysis). This conversion to lactate allows glycolysis to continue but results in reduced ATP production (2 ATPs per glucose) compared to oxidative phosphorylation. Cancer tissues convert most pyruvate from glycolysis to lactate regardless of the availability of O_2 (aerobic glycolysis), as first observed by Warburg.[38] Diverting glucose metabolites from energy production to anabolic processes leads to accelerated cell proliferation.[38] This metabolic process is also employed by normal proliferating tissues. Nevertheless, aerobic glycolysis is much less efficient than oxidative phosphorylation and only ~4 ATPs are generated per glucose.

Tumours often have low oxygen levels (hypoxia), as they outgrow their blood supplies. Animals react to chronic conditions of limiting oxygen by the hypoxic response mediated by the hypoxia-inducible factor (HIF), an α,β-heterodimeric transcription factor.[42-46] Under hypoxic conditions, HIFs accumulate and trigger the transcription of a wide variety of genes, many of which support cell survival, as described above.[42,47] These include genes that promote glycolysis and suppress oxidative phosphorylation.[42] Mutations that promote HIF production or which compromise HIF degradation, such as those leading to loss of the tumour suppressor von Hippel-Lindau protein (pVHL), can induce the hypoxic response.[46]

1.2 2-Oxoglutarate Oxygenases

The superfamily of 2-oxoglutarate (2OG)- and ferrous iron (Fe(II))-dependent non-haem oxygenases (2OG oxygenases) play an important roles in the catalysis of many biological modifications.[48-51] 2OG oxygenases are present in most organisms, and the range of different types of oxidative reactions they are able to catalyse is diverse (**Fig. 1-3**). In plants and microorganisms, the types of reactions catalysed by 2OG oxygenases include hydroxylations,[52-54] desaturations,[52] ring closures,[55] epimerisations,[56] and halogenations.[42, 48-50, 54, 57-59] However, in animals only hydroxylation and demethylation via hydroxylation are known 2OG oxygenase catalysed reactions.[60]

Over 60 2OG oxygenases have been identified as being encoded for by the human genome, and not all of their functions have been assigned thus far.[48] They are involved in a wide range of biological processes, including collagen biosynthesis, fatty acid metabolism/carnitine biosynthesis, hypoxic sensing, and transcriptional and translational regulation.[48] Their substrates include proteins, nucleic acids and small molecules. Several human 2OG oxygenases are therapeutic targets, including those involved in carnitine biosynthesis and the HIF-regulating prolyl- and asparaginyl- hydroxylases.

Using molecular oxygen, 2OG oxygenases couple the two-electron oxidation of the prime substrate to the oxidative decarboxylation of 2OG to give succinate and carbon dioxide. All 2OG oxygenases identified thus far use Fe(II) as a cofactor, and some require ascorbate for optimal activity.[51,61]

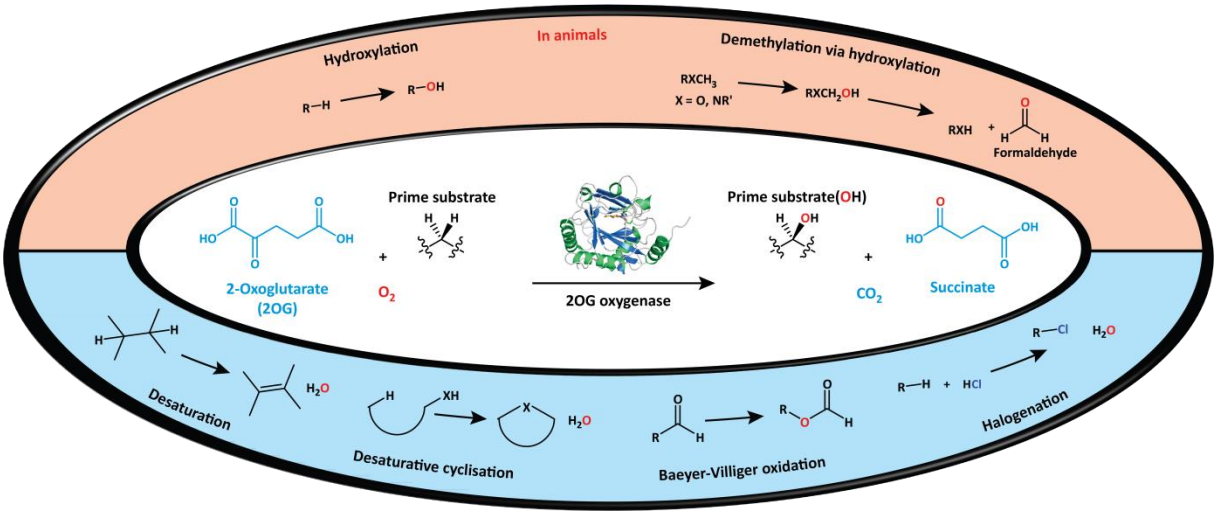


Fig. 1-3 | 2OG oxygenases catalyse a wide range of oxidative reactions on unactivated C-H bonds. In animals only hydroxylation and demethylation via hydroxylation are known 2OG oxygenase catalysed recations.

1.2.1 Structural Features and Mechanism of 2OG Oxygenases

The catalytic domains of all structurally characterised 2OG oxygenases share a double-stranded β-helix (DSBH) core fold comprised of eight antiparallel β-strands (**Fig. 1-4**), the so-called 'jelly-roll', double Greek key motif, cupin or Jumonji C (JmjC) fold.[50, 62-64] The eight strands form two four-stranded antiparallel β-sheets (minor and major β-sheet) building the core. The DSBH supports a highly but not universally conserved Fe^{2+}-binding motif in form of a triad of two histidines and one carboxylate, either aspartate or less frequently glutamate (HXD/E…H triad motif).[62,63] This catalytic or 'facial triad' is located on strands II and VII of the DSBH (**Fig. 1-4**). The 2OG-binding residues are less well conserved and characteristic of different subfamilies. A highly conserved basic residue, either arginine or lysine, coordinates the 2OG C-5-carboxylate and is located either on strand VIII, IV, or I of the DSBH.[64] Additional secondary structure elements surrounding the DSBH are common among most 2OG oxygenases. These include extra β-strands *N*-terminal to the DSBH, *N*- and *C*-terminal

helices, and a loop insert between the fourth and fifth β-strand. The substrate-binding residues vary but often include residues located on strands I and II and extra β-strands *N*-terminal to the DSBH.[50,62]

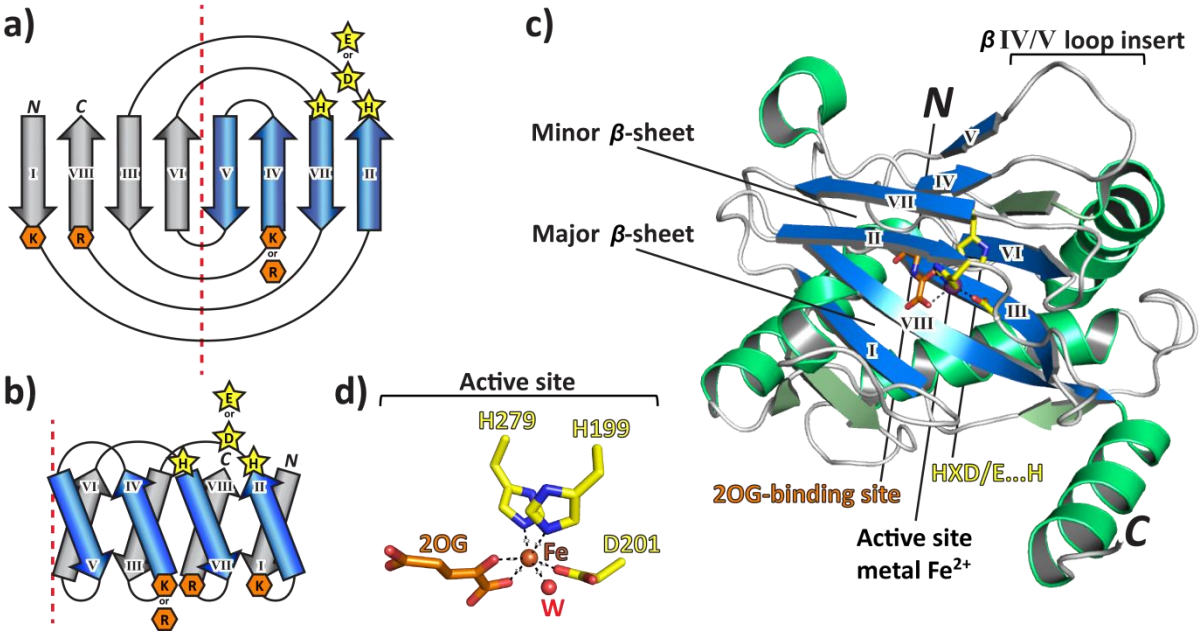


Fig. 1-4 | The jelly-roll core of 2OG oxygenases. **a)** Schematic of the double Greek key motif; **b)** folded to the double-stranded β-helix (DSBH). The DSBH core strands (I-VIII) are numbered. The HXD/E...H motif catalytic triad is located on strands II and VII of the DSBH and is indicated by the yellow stars. Orange hexagons indicate a highly conserved basic 2OG-binding residue, arginine or lysine, located either on strand VIII, IV, or less commonly I. **c)** Ribbons representation of a typical 2OG oxygenase (PDB ID: 3HQR[65]) with the DSBH core strands in blue. **d)** The active site of a typical 2OG oxygenase (PDB ID: 3P3N[66]). The HXD/E...H catalytic triad, 2OG, and a water (W) molecule coordinate the Fe(II) in octahedral geometry.

A consensus mechanism for 2OG oxygenases has been proposed based on structural, spectroscopic and quantum mechanics studies.[50,51, 67-71] In the resting state, the Fe(II) is octahedrally coordinated by the side chains of the catalytic triad and usually three water molecules (**Fig. 1-5**). The ferrous active site Fe(II) (electron configuration [Ar]3d^6) has a paramagnetic high-spin electron configuration ((t_{2g})^4(e_g)^2). Two of these water molecules are displaced by the cosubstrate 2OG,[68,69] which coordinates the Fe^{2+} centre in a bidentate manner with its 2-oxo group and the C-1-carboxylate.[52] The coordination position of the carboxylate can thereby vary between family members, as in the cases of the HIF hydroxylases factor inhibiting HIF (FIH) and prolyl hydroxylase 2 (PHD2, **Fig. 1-5**).[42, 50, 62]

The prime substrate has been proposed to bind prior to dioxygen[72] and to weaken the binding of the remaining water molecule to the metal.[53] This water molecule is then displaced by dioxygen in the catalytic cycle. End-on binding of oxygen results in initial formation of an Fe(III)-$O_2^{\bullet-}$ superoxo species, which then attacks the electrophilic carbon of the 2OG 2-oxo group to give a six-membered cyclic peroxide.[73] A rearrangement under heterolytic O-O bond cleavage then leads to formation of a highly reactive oxidising Fe(IV) intermediate (high-spin Fe(IV)=O species) and to decarboxylation of 2OG to give succinate.[67,74] The Fe(IV)=O species then abstracts a hydrogen atom

from the prime substrate, which leads to a free radical and an Fe(III)-OH species. The free radical subsequently abstracts the hydroxyl group from the metal to form the hydroxylated substrate and ferrous Fe(II) complexed with succinate.

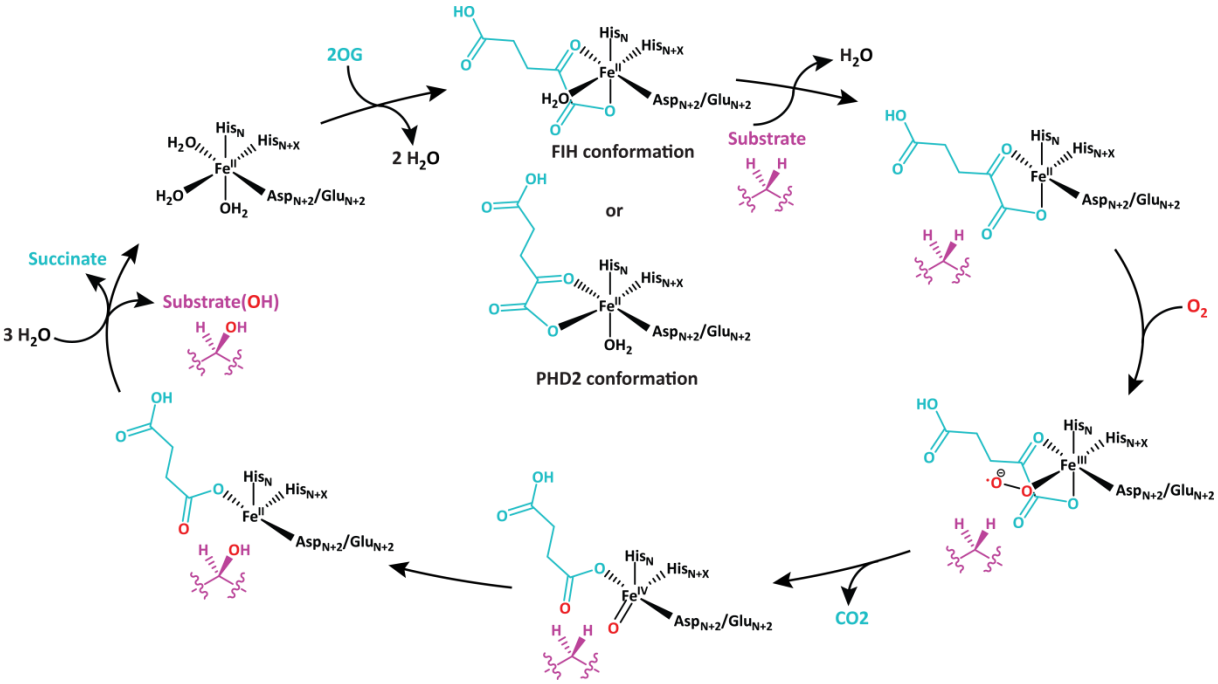


Fig. 1-5 | Consensus mechanism of 2OG oxygenases. The active site Fe(II) is octahedrally coordinated by the conserved HxD/E…H catalytic triad and three water molecules in the resting state. The cosubstrate 2OG binds to the Fe(II) by displacing two water molecules. Binding of the prime substrate in the active site weakens the coordination of the remaining water molecule to the metal, facilitating oxygen binding. The highly reactive Fe(IV)=O intermediate is subsequently formed concomitant to the decarboxylation of 2OG to give succinate. The Fe(IV)=O enables insertion of an oxygen atom into the C-H bond of the prime substrate via a radical mechanism. Structural studies have revealed that 2OG can bind to the Fe(II) centre in different conformations, as shown here for the 2OG oxygenases factor inhibiting HIF (FIH) and prolyl hydroxylase 2 (PHD2).

Some 2OG oxygenases such as the HIF hydroxylase PHD2 have been shown to react slowly with oxygen, consistent with their role oxygen sensors in the hypoxic response.[75,76] Similarly, for the histone demethylase JMJD2E a slow reaction with oxygen has been reported.[77] Furthermore, hypoxia attenuated the activity of histone demethylases in macrophages, which suggested that epigenetic gene regulation may directly depend on the availability of oxygen.[78,79] Many 2OG oxygenases could have roles as oxygen sensors, and hence signalling enzymes, to enable cells to adapt their gene expression in response to changes in the surrounding conditions.

1.2.2 Subclasses of Selected Human 2OG Oxygenases

In contrast to the structural conservation of the DSBH fold, the primary amino acid sequences of the different 2OG oxygenases are less well conserved (less than 20 %) and often contain multiple domains. The identification and classification of 2OG oxygenases by bioinformatics analyses can

therefore be challenging. The superfamily can be classified into subfamilies (**Fig. 1-6**) based on their sequence similarity, structural similarity, substrate specificity, their physiological roles, or biochemical reactions. Many of the human members of the superfamily have been linked to cancer and other diseases such as anaemia, ischemic or inflammatory diseases, and represent potential therapeutic targets (**Table 1-1**).[42, 48, 80-81]

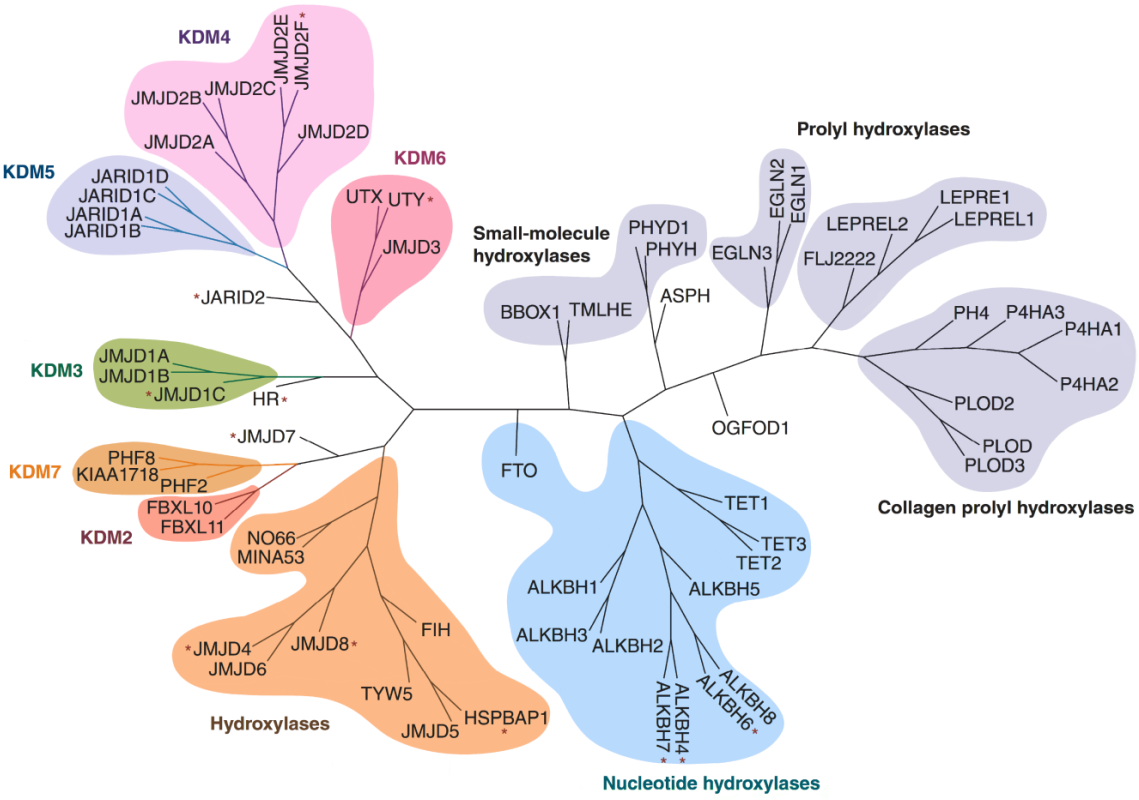


Fig. 1-6 | Phylogenetic tree of human 2OG oxygenases. The figure showing evolutionary subgrouping of the superfamily was generated on the basis of sequence alignments. Adapted from Johansson *et al.*[82] (SGC Oxford).

Abbreviations in Table 1-1 (on the following pages): PLOD: procollagen-lysine, 2-oxoglutarate 5-dioxygenase; ER: endoplasmic reticulum; LEPRE: leucine- and proline-enriched proteoglycan (leprecan); P4HA: procollagen-proline, 2-oxoglutarate 4-dioxygenase; JMJD: Jumonji domain-containing protein; HIF: hypoxia inducible factor; EGLN: Egl nine homologue; PHD: HIF prolyl-4-hydroxylase; FIH: factor inhibiting HIF-1/HIF asparagine hydroxylase; ARD: ankyrin repeat domain; KDM: lysine-specific demethylase; FBXL: F-box and leucine-rich repeat protein; H3K: histone 3 lysine; me1: monomethyl; me2: dimethyl; me3: trimethyl; JARID: Junonji, AT-rich interactive domain-containing protein; UTX: ubiquitously transcribed tetratricopeptide repeat protein on the human X chromosome; UTY: ubiquitously transcribed tetratricopeptide repeat protein on the human Y chromosome; PHF: plant homeodomain finger protein; JHDM: JmjC domain-containing histone demethylation protein; eRF1: eukaryotic translation termination factor 1; NFATc: nuclear factor of activated T-cells, cytoplasmic; U2AF65: splicing factor U2 auxiliary factor 65 kDa subunit; LUC7L2: putative RNA-binding protein Luc7-like 2; HSPBAP: 27 kDa heat shock protein-associated protein; PASS: protein associated with small stress protein; MINA53: MYC proto-oncogene-induced nuclear antigen (53 kDa); RPL27a: 60S ribosomal protein L27a; NO66: nucleolar protein 66; RPL8: 60S ribosomal protein L8; OGFOD: 2-oxoglutarate and iron-dependent oxygenase domain-containing protein; RPS23: 40S ribosomal protein 23; AlkBH: Alkylated DNA repair protein alkB homologue; 1mA: 1-methyladenosine; 3mC: 3-methylcytidine; m6A: N^6-methyladenosine; tRNA: transfer RNA; FTO: fat mass and obesity associated protein; 3mU: 3-methyluridine; 3mT: 3-methylthymidine; TET: ten-eleven translocation 1 gene protein; 5mC: 5-methylcytosine; BBOX: gamma-butyrobetaine hydroxylase; TMLHE: trimethyllysine hydroxylase; EGF-like: epidermal growth factor-like; EGFH: EGF-like domain hydroxylase; AspH: aspartyl/asparaginyl-β-hydroxylase; AspHD: aspartyl/asparaginyl-β-hydroxylase domain-containing protein.

Table 1-1 | Physiological roles and subcellular localisation of selected human 2OG oxygenases and their known links to cancer. (*Localisation proposed by similarity)

Physiological role	Class	Member	Reaction catalysed	Subcellular localisation	Links to cancer
Collagen biosynthesis	PLODs	PLOD1	Procollagen Lys-5*R* hydroxylation	ER	Mutations associated with breast and colorectal cancer[83]
		PLOD2	Procollagen Lys-5*R* hydroxylation	ER	Promotes breast cancer[84] and sarcoma metastasis,[85] overexpressed in cervical carcinoma,[86] use as prognostic marker in hepatocellular carcinoma,[87] glioblastoma,[88] oral[89] and bladder carcinoma[90]
		PLOD3	Procollagen Lys-5*R* hydroxylation	ER	Use as biomarker for gastric[91] and pancreatic cancer[92]
	C-P3H	LEPRE1	Procollagen Pro-3*S* hydroxylation	ER	Upregulated in invasive bladder cancer[93]
		LEPREL1	Procollagen Pro-3*S* hydroxylation	ER	Overexpressed in thyroid carcinoma[94]
		LEPREL2	Procollagen Pro-3*S* hydroxylation	ER	–
	C-P4H	P4HA1	Procollagen Pro-4*R* hydroxylation	ER	Essential role for breast cancer metastasis,[84,95-96] prostate cancer[97]
		P4HA2	Procollagen Pro-4*R* hydroxylation	ER	Promotes breast cancer progression and metastasis,[98,99,96] use as prognostic marker for oral squamous cell carcinoma,[100,101] thyroid cancer,[102] glioblastoma[103]
		P4HA3	Procollagen Pro-4*R* hydroxylation	ER	–
HIF hydroxylation	PHD	PHD1 (EGLN2)	HIF-α Pro-4*R* hydroxylation	Nucleus	Reported tumour suppressor activity,[104] but oestrogen-inducible in breast carcinoma and inhibition reduced tumourigenesis,[105] associated with poor prognosis in breast cancer,[106] upregulated also in ovarian[107] and renal clear cell carcinoma;[108] use as prognostic marker in non-small cell lung cancer[109]
		PHD2 (EGLN1)	HIF-α Pro-4*R* hydroxylation	Cytoplasm, nucleus	Tumour suppressor activity in various cancers,[110,111] associated with increased survival in breast cancer,[106] use as prognostic marker in gastric cancer;[112] overexpression in head and neck squamous cell carcinoma is associated with tumour aggressiveness[113]
		PHD3 (EGLN3)	HIF-α Pro-4*R* hydroxylation	Cytoplasm, nucleus	Tumour suppressor activity,[114,115] proposed regulator of tumour growth, apoptosis and angiogenesis in pancreatic[116] and breast cancer and associated with good prognosis;[106] use as serological marker for renal cell carcinoma[117] and favourable prognostic marker in gastric cancer[118]
	JmjC	FIH (HIF1AN)	HIF-α/ARD Asn-3*S* hydroxylation	Cytoplasm	Suggested tumour suppressor role;[119] micro-RNA-31 impedes *HIF1AN* expression in colorectal[120] and head and neck squamous carcinoma;[121] the *HIF1AN* gene is often deleted in glioblastoma,[122] leading to constitutive activation of HIF-1α; prognostic marker in invasive breast cancer[123,124]

Physiological role	Class	Member	Reaction catalysed	Subcellular localisation	Links to cancer
Epigenetic regulation	KDM2/JmjC	FBXL11 (KDM2A)	H3K36me1/me2, p65 N^{ε}-Lys demethylation	Nucleus	Linked to invasion and migration in breast cancer[125]
		FBXL10 (KDM2B)	H3K36me1/me2, H3K4me3 N^{ε}-Lys demethylation	Nucleus	Proposed oncogene in myeloid leukaemia[126], promotes pancreatic cancer[127]
	KDM3/JmjC	JMJD1A (KDM3A)	H3K9me2/me1 N^{ε}-Lys demethylation	Cytoplasm, nucleus	Inhibition reduced suppressed tumour growth by downregulating angiogenesis and macrophage infiltration,[128] overexpressed in prostate,[129] colorectal,[130] hepatocellular[131] and renal cell carcinoma[132] and used as prognostic marker
		JMJD1B (KDM3B)	H3K9me2/me1 N^{ε}-Lys demethylation	Nucleus	Proposed tumour suppressor role,[133,134] the *KDM3B* gene is commonly deleted in myeloid leukaemia[135] and breast cancer[136]
		JMJD1C (JHDM2C)	Unknown	Nucleus*	Proposed tumour suppressor[137]
	KDM4/JmjC	JMJD2A (KDM4A)	H3K9me2/me3, H3K36me2/me3, H1.4K26me2/me3 N^{ε}-Lys demethylation	Nucleus	Proposed oncogenic role[138-141] in breast[142] and prostate[143,144] cancers, promotes colon cancer by regulation of apoptosis and proliferation,[145] use as prognostic marker for gastric[146] and bladder[147] cancer
		JMJD2B (KDM4B)	H3K9me2/me3, H3K36me2/me3, H1.4K26me2/me3 N^{ε}-Lys demethylation	Nucleus	Prostate cancer target,[148,143] role in carcinogenesis of bladder and lung cancer,[149] breast,[150-153] colorectal cancer,[154] promotes gastric cancer invasion and metastasis[155,156]
		JMJD2C (KDM4C)	H3K9me2/me3, H3K36me2/me3, H1.4K26me2/me3 N^{ε}-Lys demethylation	Nucleus	Candidate oncogene[138,157] that promotes expression of further oncogenes,[158] coactivator of HIF-1 in breast cancer progression,[159] overexpressed in oesophageal squamous cell carcinomas[160], breast cancer,[161] and used as prognostic marker[162,163]
		JMJD2D (KDM4D)	H3K9me2/me3, H1.4K26me2/me3 N^{ε}-Lys demethylation	Nucleus	Linked to androgen receptor activation in prostate cancer,[144] can stimulate cell proliferation and survival[164,165]
		JMJD2E (KDM4E)	H3K9me2/me3, H1.4K26me2/me3 N^{ε}-Lys demethylation	Nucleus	–
		JMJD2F (KDM4F)	Unknown	Nucleus*	–
	KDM5/JmjC	JARID1A (KDM5A)	H3K4me1/me2/me3 N^{ε}-Lys demethylation	Nucleus	As fusion-oncoprotein in leukaemia,[166,167] linked to development of cancer drug resistance,[168] regulates cellular senescence,[169] prognostic marker for better prognosis in breast cancer[170]
		JARID1B (KDM5B)	H3K4me1/me2/me3 N^{ε}-Lys demethylation	Nucleus	Role in regulation of tumour-suppressive cellular senescence in colorectal cancer,[171] retinoblastoma,[169] suppression of angiogenesis and metastasis in breast cancer,[172-174] favourable prognostic marker in melanoma and breast cancer[175]
		JARID1C (KDM5C)	H3K4me1/me2/me3 N^{ε}-Lys demethylation	Nucleus	Suppression of tumour growth through interaction with pVHL,[176] inactivating mutation in renal carcinoma[177]

Physiological role	Class	Member	Reaction catalysed	Subcellular localisation	Links to cancer
Epigenetic regulation	KDM5/JmjC	JARID1D (KDM5D)	H3K4me1/me2/me3 N^{ε}-Lys demethylation	Nucleus	Frequently deleted in prostate cancer[178]
		JARID2	Unknown	Nucleus*	Frequently deleted in acute myeloid leukaemia[179,180]
	KDM6/JmjC	UTX (KDM6A)	H3K27me2/me3 N^{ε}-Lys demethylation	Nucleus	Regulates cell proliferation and invasiveness in breast cancer cells and is associated with poor prognosis,[181] inactivating somatic mutations in tumours,[182] prognostic marker for breast cancer[183]
		UTY	H3K27me2/me3 N^{ε}-Lys demethylation	Nucleus	Expressed in prostate cancer[184]
		JMJD3 (KDM6B)	H3K27me2/me3 N^{ε}-Lys demethylation	Nucleus	Proposed role as tumour suppressor,[185] loss of KDM6B in pancreatic cancer associated with enhanced aggressiveness,[186] over-expressed in various tumours,[187-192] promotes tumour cell invasion and metastasis and knockdown inhibited invasion;[193] use as diagnostic biomarker in renal cell carcinoma[194]
	KDM7/JmjC	PHF2	H3K9me2, N^{ε}-Lys demethylation	Nucleus	Overexpressed in oesophageal squamous cell[195] and breast carcinoma[196]
		PHF8	H3K9me2/me1, H4K20me1, N^{ε}-Lys demethylation	Nucleus	Oncogenic role in oesophageal squamous cell carcinoma and used as diagnostic marker,[195] target in leukaemia treatment,[197] linked to increased cell proliferation, migration and invasion in prostate cancer[198]
		JHDM1D (KDM7A)	H3K36me1/me2 N^{ε}-Lys demethylation	Nucleus	Increased *JHDM1D* expression suppressed tumour growth by down-regulating angiogenesis[199]
Protein hydroxylation	JmjC	JMJD4	eRF1 Lys-4 hydroxylation	Nucleus	–
		JMJD5 (KDM8)	H3K36me2/me1 N^{ε}-Lys demethylation, NFATc1 hydroxylation	Nucleus	Regulates cancer cell proliferation,[200] overexpressed in leukaemia and breast cancer cells[201]
		JMJD6	U2AF65, LUC7L2, H3R2me2, H4R3me2/me1, H2A/H2B, H3/H4 Lys-5 hydroxylation, Arg-demethylation	Nucleus	Overexpression in breast cancer,[202] lung adenocarcinoma,[203] prognostic marker linked to poor prognosis[202] and angiogenesis,[204] promotes colon carcinogenesis[205]
		JMJD7	Unknown	Unknown	–
		JMJD8	Unknown	Unknown	–
		HSPBAP1 (PASS1)	Unknown	Cytoplasm*	Linked to a chromosomal aberration found in renal carcinoma[206]
Ribosome hydroxylation	JmjC	MINA53 (MINA)	RPL27a His-3*S*-hydroxylation, H3K9me3 demethylation	Nucleus	Therapeutic target oncogene for lung cancer,[207-209] associated with increased cell proliferation[210] and also overexpressed in colon cancer,[211] cholangiocarcinoma,[212] hepatocellular,[213] gastric,[214] pancreatic,[215] renal,[216] oesophageal squamous cell carcinoma,[217] neuroblastoma[218] and lymphoma,[219] prognostic marker[217,220]
		NO66	RPL8, H3K4, H3K36 His-3*S* hydroxylation	Nucleus	Overexpressed in non-small cell lung cancer[207]
	OGFOD	OGFOD1	RPS23 Pro-3*S* hydroxylation	Cytoplasm, nucleus	Potential use as serological biomarker for chronic lymphocytic leukaemia[221]

Physiological role	Class	Member	Reaction catalysed	Subcellular localisation	Links to cancer
Nucleic acid modification	AlkBH/JmjC	AlkBH1	DNA demethylation 1mA, 3mC	Nucleus, cytoplasm, mitochondria	Linked to non-small cell lung cancer[222]
		AlkBH2	DNA demethylation 1mA, 3mC	Nucleus	Contributes to bladder cancer progression,[223] associated with risk of meningioma[224] and mediator to drug resistance in glioblastoma,[225] down-regulated in gastric cancer[226]
		AlkBH3	DNA demethylation 1mA, 3mC	Nucleus, cytoplasm	Contributes to cancer cell survival and angiogenesis in non-small cell lung,[227] urothelial[228] and pancreatic cancer,[229] knockdown[230] or inhibition[231] reduced prostate cancer growth
		AlkBH4	Unknown	Nucleus, cytoplasm	–
		AlkBH5	m6A DNA demethylation	Nucleus	Induced by hypoxia[232]
		AlkBH6	Unknown	Nucleus, cytoplasm	–
		AlkBH7	Unknown	Nucleus, cytoplasm	Role in programmed necrosis as potential target in apoptosis-resistant cancer cells[233]
		AlkBH8	tRNA hydroxylation	Cytoplasm	Contributes to bladder cancer progression[234]
		FTO	3mU RNA, 3mT DNA, m6A RNA demethylation	Nucleus	Linked to breast cancer[235,236] and colorectal cancer risk;[237] besides cancer strongly associated with obesity,[238] Alzheimer's disease,[239] cardiovascular disease[240]
	TET/JmjC	TET1	5mC hydroxylation	Nucleus	Oncogenic role in acute myeloid leukaemia,[241-243] but down-regulated in hepatocellular,[244] gastric[245] and breast carcinoma[246] and used as prognostic marker, suppresses tumour development and invasion in prostate and breast cancer tissues[247]
		TET2	5mC hydroxylation	Nucleus	Mutations promote tumourigenesis in myeloid malignancies,[243, 248-250] down-regulated in oral squamous cell carcinoma[251]
		TET3	5mC hydroxylation	Nucleus	Mutations in myeloid malignancies[243]
Fatty acid metabolism	BBOX1/ TMLHE	TMLHE	N^{ϵ}-Lys-me3 3*S* hydroxylation	Mitochondrion matrix	–
		BBOX1 (GBBH)	*γ*-Butyrobetaine 3*S* hydroxylation	Cytoplasm	Potentially linked to several cancer types[139-140]
EGF-like domain hydroxylation	EGFH	AspH	EGF-like domains Asp/Asn-3*R* hydroxylation	ER	Full-length AspH and shorter isoform humbug overexpressed in hepato-cellular,[252] gastric,[253] pancreatic,[254] kidney, prostate, ovarian, breast,[255] lung,[256] colorectal,[257] bile duct carcinoma[258] and glioblastoma;[259] *AspH* is highly upregulated in hypoxia,[260] associated with malignant transformation,[261] worse clinical outcome and poorer prognosis for survival;[262] mice knockout linked to increased tumour growth;[263] used as cancer biomarker[262,264]
		AspHD1	Unknown	Unknown	–
		AspHD2	Unknown	Unknown	–

Functional subclasses of human 2OG oxygenases include (i) the HIF hydroxylases that are important in hypoxic signalling, (ii) the JmjC histone demethylases that have roles in epigenetic regulation, (iii) DNA/RNA hydroxylases and demethylases that have roles in nucleic acid repair mechanisms, (iv) small molecule hydroxylases involved in carnitine biosynthesis/fatty acid metabolism,[265] and (v) protein hydroxylases such as the collagen prolyl and lysyl hydroxylases (biosynthesis of collagen), ribosomal protein hydroxylases, and the epidermal growth factor-like domain hydroxylase.[48] Protein hydroxylation catalysed by 2OG oxygenases enables further modifications and is important for protein-protein interactions.[80,266] 2OG oxygenases are involved in all steps of gene expression, and hence offer various points for potential therapeutic intervention (**Fig. 1-7**). The following sections briefly describe some major human 2OG oxygenase subfamilies defined by their physiological role and briefly discuss their links to cancer. **Section 1-3** introduces the 2OG oxygenase responsible for epidermal growth factor-like domain hydroxylation, which the research presented in this Thesis focused on.

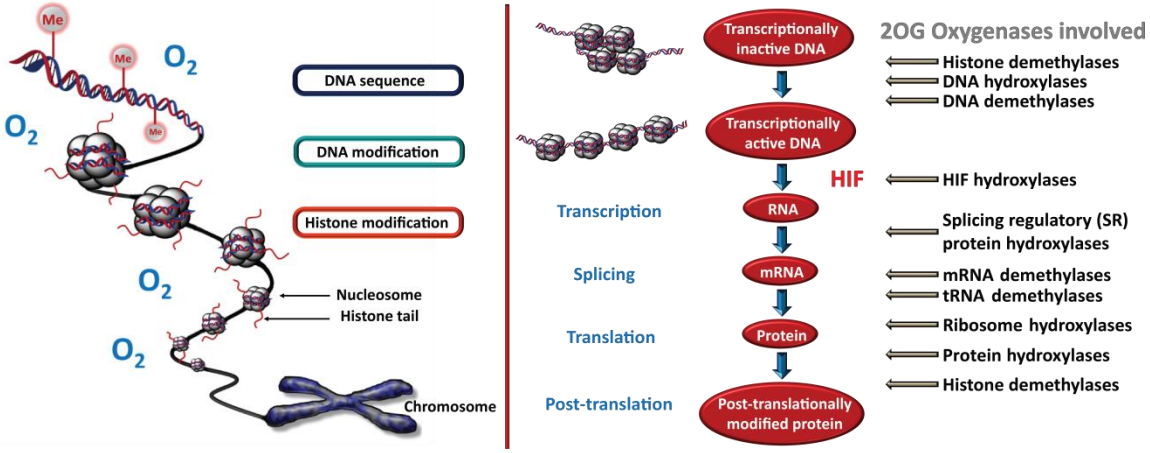


Fig. 1-7 | 2OG oxygenases are involved in all steps of gene expression. Some interaction points (e.g. HIF hydroxylases) seem more important than others for targeting diseases. Adapted from Qiu[267] and Loenarz *et al.*[80]

1.2.2.1 Collagen Biosynthesis

Collagens are the most abundant proteins in mammals and the main component of connective tissues.[268] They form highly organised supramolecular assemblies in the extracellular matrix (ECM) as structural support of cells within tissues, for the connection of tissues within an organ, and for the attachment or migration of cells.[269] The biosynthesis of collagens includes posttranslational hydroxylation of prolyl and lysyl residues in procollagen, the precursor form, (**Fig. 1-8**). Procollagen prolyl-4-hydroxylase (C-P4H) was the first reported human 2OG oxygenase.[270] C-P4H is a heterotetramer of two subunits α and β ($\alpha_2\beta_2$),[271] in which the α-subunit is the catalytically active 2OG oxygenase (P4HA1-3),[272] and the β-subunit is identical to protein disulfide isomerase (PDI).[273,274] The enzyme is localised to the endoplasmic reticulum (ER),[275] and its activity is strongly dependent on ascorbate (vitamin C).[276,277] Prolyl-4-hydroxylation is the most frequent collagen modification and has

been reported to stabilise the collagen triple helix under physiological conditions.[278] Prolyl-3-hydroxylation catalysed by procollagen prolyl-3-hydroxylases or leprecans (C-P3Hs) can occur after prolyl-4-hydroxylation on a neighbouring proline, but is less common. Lysyl-5-hydroxylation in collagens is catalysed by procollagen lysyl-5-hydroxylases (PLODs) and also affects collagen structure. It can serve as an anchor point for subsequent glycosylation and build interstrand cross-links that stabilise the collagen structure.[279-281]

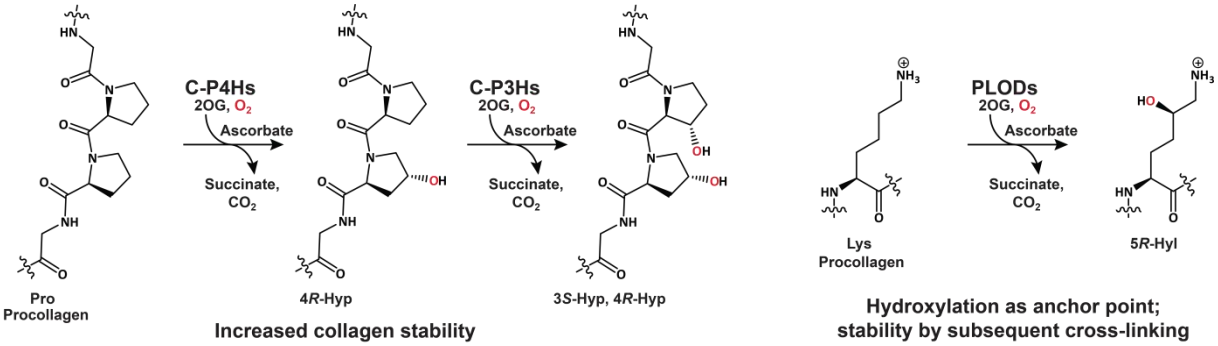


Fig. 1-8 | Roles of 2OG oxygenases in collagen biosynthesis. 2OG oxygenase-catalysed prolyl and lysyl hydroxylations of procollagen are important for the structural stability of the collagen triple helix. Hyp: hydroxyproline; Hyl: hydroxylysine.

The ECM composition and structure plays a critical role for processes such as general growth and wound healing but also for the development of cancer. ECM proteins like collagens and integrins are linked to cancer progression and metastasis, as the establishment of an invasive and metastatic tumour phenotype relies on adhesive cell-cell or cell-matrix interactions.[282-284] Increased collagen deposition (**Fig. 1-9**) has been linked to tumour growth and invasiveness.[285-287] The PLODs and collagen prolyl-4-hydroxylase subunits α-1 and 2 (P4HA1, P4HA2) have been reported to promote cancer progression by regulating collagen deposition[95-99] and to even have essential roles for breast cancer metastasis.[84] In addition, their expression in cancerous tissues is used as a prognostic and diagnostic marker (see **Table 1-1**).

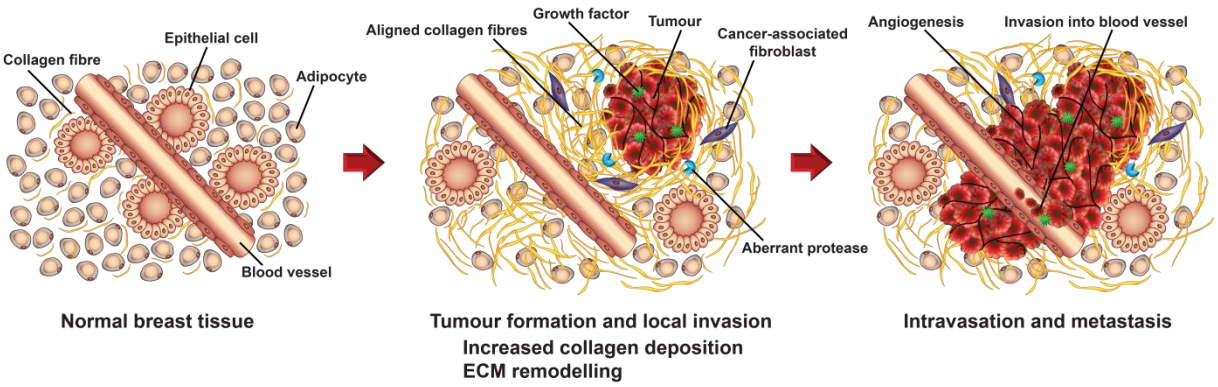


Fig. 1-9 | Links between collagen biosynthesis and cancer. Lump and tumour formation, e.g. in breast cancer: Increased collagen deposition and extracellular matrix (ECM) remodelling are required for tumourigenesis and invasiveness. ECM remodelling includes upregulation of growth factors and proteases such as matrix metalloproteinases (MMPs) that degrade the membrane surrounding a tumour. Aligned collagen fibres within the interstitial matrix facilitate local invasion and subsequent intravasation and metastasis.

1.2.2.2 HIF Hydroxylation: Hypoxic Sensing

As described in **Section 1.1.1**, most tumours are hypoxic. Studies on the hypoxia sensing mechanisms have demonstrated that posttranslational prolyl and asparaginyl hydroxylation of HIF play an important role in gene expression and hypoxic signalling.[42,47,288] HIF is a heterodimeric transcription factor consisting of a regulatory α-subunit and a constitutively expressed β-subunit.[289] HIF-α is mainly regulated by three oxygen-dependent prolyl hydroxylase isozymes (PHD1-3, Egl nine homologues (EGLN)) and the asparaginyl hydroxylase factor inhibiting HIF (FIH, gene *HIF1AN*), which are members of the 2OG oxygenase family (**Fig. 1-10**).[290]

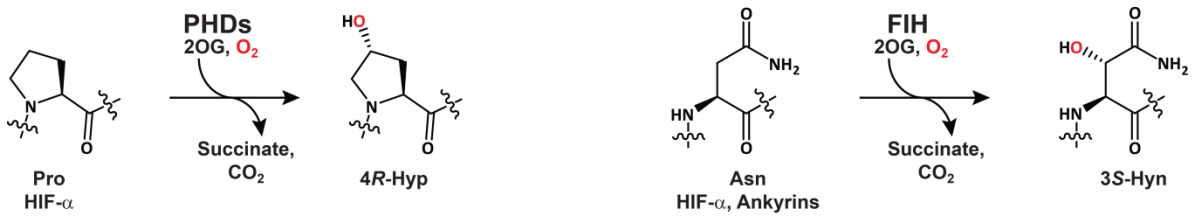


Fig. 1-10 | Prolyl and asparaginyl hydroxylases of HIF-α have important functions in hypoxic sensing. Hyp: hydroxyproline; Hyn: hydroxyasparagine.

Under normoxic conditions, the PHDs catalyse hydroxylation of prolyl residues of HIF-α (either Pro402 or Pro564) within the oxygen-dependent degradation domain (ODD), which is recognised by a ubiquitin ligase complex (VCB complex) containing the tumour suppressor protein von Hippel-Lindau (pVHL).[45, 291-292] It is then marked for degradation by proteasomes by adding a chain of ubiquitin (Ub) proteins, thus suppressing the HIF-mediated hypoxic response (**Fig. 1-11**).[293] Under normoxia, HIF-α is also hydroxylated at Asn803 within its *C*-terminal activation domain (CAD) in a reaction catalysed by FIH.[294-297] This hydroxyl group prevents binding of HIF-α to the p300/cyclic adenosine monophosphate response element-binding protein (CBP) transcriptional coactivator complex, thus inhibiting HIF-*α*-mediated transcription.[298-300]

Under hypoxic conditions, the activity of the HIF hydroxylases is impaired. HIF-α is then translocated to the nucleus where it forms a heterodimer with HIF-β and binds to hypoxia response elements (HREs) of certain genes, enabling their transcription (**Fig. 1-11**).[47] HIF has about 200 target genes, including vascular endothelial growth factor (VEGF), glucose transporter 1 (GLUT1), carbonic anhydrase IX (CAIX), transforming growth factor α (TGF-α), erythropoietin (EPO), and a wide variety of other targets of therapeutic interest.[42, 301] HIF-α isoforms (HIF-1α, HIF-2α) are highly upregulated in many cancers, and linked to an aggressive tumour phenotype.[302-304] In general, HIF target genes expression promotes and supports processes essential for tumour growth such as angiogenesis/vascularisation, growth and immortalisation, pH control, metabolic reprogramming and glycolysis, metastasis and invasion.[301,302] Modulation of the HIF-mediated hypoxic response is of (potential) use in cancer treatment as well as for treatment of a wide range of other diseases.[305-307]

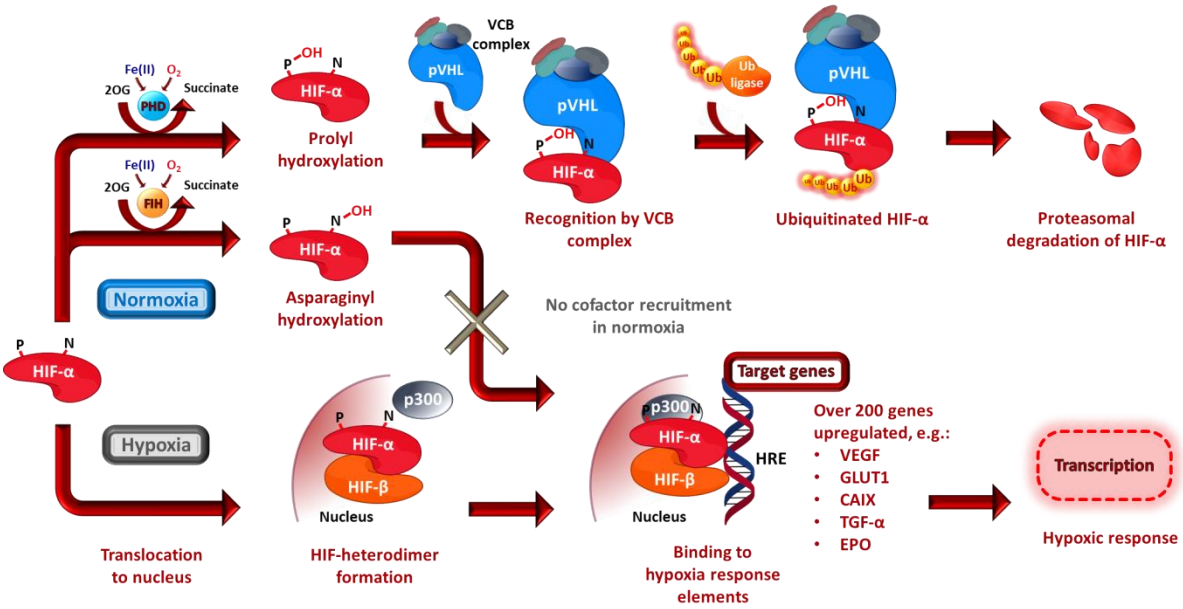


Fig. 1-11 | Schematic regulation of HIF function. Under normoxic conditions, HIF-α is hydroxylated by the PHDs and ultimately degraded by proteasomes. Asparaginyl hydroxylation catalysed by FIH prevents p300/CBP cofactor recruitment. Under hypoxic conditions, HIF-α translocates to the nucleus where it forms a heterodimer with HIF-β and binds to hypoxia response elements (HREs) of its target genes, enabling transcription.

Tumour suppressor activity has been reported for all the HIF hydroxylases (see **Table 1-1**). The chromosomal region containing *HIF1AN* is often deleted in glioblastoma.[122] The HIF hydroxylases are also used as prognostic markers in various cancers. PHD2 (EGLN1) is the primary regulator of HIF-1α and is associated with increased survival in many cancers including breast carcinoma.[106] In contrast, PHD1 (EGLN2) has been reported to be oestrogen-inducible in breast carcinoma, and its inhibition diminished tumourigenesis.[105] PHD1 and PHD3 (EGLN3) seem to have additional HIF-independent roles in cell proliferation and apoptosis, and have thus been suggested as potential cancer targets.[308] Overall, both activation and inhibition of the HIF hydroxylases have therapeutic potential for the treatment of cancer or other diseases.[46, 309-311]

1.2.2.3 Epigenetic Regulation

2OG oxygenases also have distinct roles in epigenetic regulation of gene expression by histone modification. Methylation of lysyl and arginyl residues in histones was long believed to be irreversible and can signal for both gene activation and gene repression, depending on the position of the

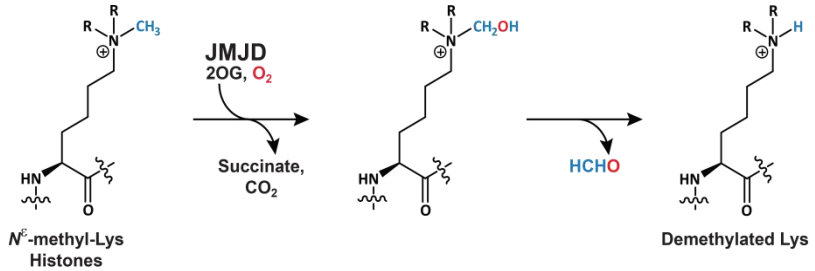


Fig. 1-12 | JmjC-domain containing protein (JMJD)-catalysed oxidative demethylation of N^{ε}-methylated lysyl residues in histones.

methylated residue and the methylation state.[312] Studies in 2006 identified the first 2OG oxygenases from the jumonji domain-containing (JmjC) subfamily capable of demethylating N^{ε}-methy-lated lysyl and arginyl residues in histone tails.[138, 313-315] The demethylation reaction proceeds via methyl hydroxylation with concomitant release of formaldehyde (**Fig. 1-12**). Different subfamilies of human JmjC histone lysine-specific demethylases (KDMs) have selectivities for particular histone residues and methylation states, partly due to additional non-catalytic domains such as bromo-, chromo- and tudordomains. For example, members of the JmjC-domain containing 2 (JMJD2) subfamily are selective for demethylation of the tri- and di- N^{ε}-methylation states of specific lysines on histone H3,[316] whereas other subfamilies, such as the F-box and leucine-rich repeat proteins (FBXL) and plant homeodomain finger protein (PHF) subfamilies are selective for the di- and mono-N^{ε}-methylated states and do not accept the trimethylated state.[317,315]

Many histone demethylases are overexpressed in a variety of cancers, and some, such as members of the KDM4 subfamily, have even been assigned oncogenic roles (see **Table 1-1**). Many histone demethylases are involved in regulation of cell proliferation, apoptosis and cancer progression. Inhibition of some family members has been reported to reduce tumour growth.[128] On the other hand, a number of demethylases, such as KDM5 subfamily members, have been found to be frequently deleted in some tumour types and thus linked to tumour suppressor activities (see **Table 1-1**). The histone demethylases as epigenetic modulators are interesting targets for drug discovery.[318,319]

1.2.2.4 Nucleic Acid Modification

2OG oxygenases are also involved in nucleic acid modifications. They have roles in nucleic acid repair, cell differentiation and metabolic regulation. In *Escherichia coli* (*E. coli*), the 2OG oxygenase AlkB catalyses repair of toxic and mutagenic *N*-alkylated single-stranded DNA and RNA caused by alkylating agents.[320] AlkB can demethylate a number of alkylated bases and therefore rescue the nucleic acid strands from their *N*-methylated toxic states.[321,322] Eight human AlkB homologues (AlkBH1-8) have been proposed. AlkBH1-5 have been reported to have DNA demethylation activity similar to that of AlkB and to be involved in repair mechanisms of toxic and mutagenic DNA lesions.[320, 323-325] The nucleic acid demethylation reaction is proposed to occur via hydroxylation of the methyl group, followed by release of formaldehyde (**Fig. 1-13**).

Elevated levels of some of the human AlkB homologues have been observed in cancers and linked to increased cancer cell survival and tumour progression (**Table 1-1**). AlkBH2 and 3 have also been proposed to mediate drug resistance in some tumour types that are treated with alkylating reagents.[326,225]

Besides the AlkB homologues, another 2OG oxygenase, the fat mass and obesity associated protein (FTO), also demethylates DNA and RNA substrates.[238, 327] FTO is not only linked to obesity but

also associated with Alzheimer's[239] and cardiovascular disease[240] and an increased breast and colorectal cancer risk.[235-237]

2OG-dependent enzymes of the ten-eleven translocation 1 gene (TET) protein subfamily catalyse hydroxylation of the epigenetic marker 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), which can also be further oxidised to the mutagenic 5-formylcytosine and 5-carboxy-cytosine.[328-330] There are many links between TET1-3 and cancers (see **Table 1-1**). Highly upregulated TET1 has been reported to play an oncogenic role in acute myeloid leukaemia,[241] but it was found to be down-regulated and tumour-suppressing in other cancer types.[244-247] Furthermore, *TET1-3* mutations have been reported to promote tumourigenesis in myeloid malignancies.[243] Both monitoring of *TET1-3* expression and levels of 5mC hydroxylation are being used a prognostic markers for cancer progression.

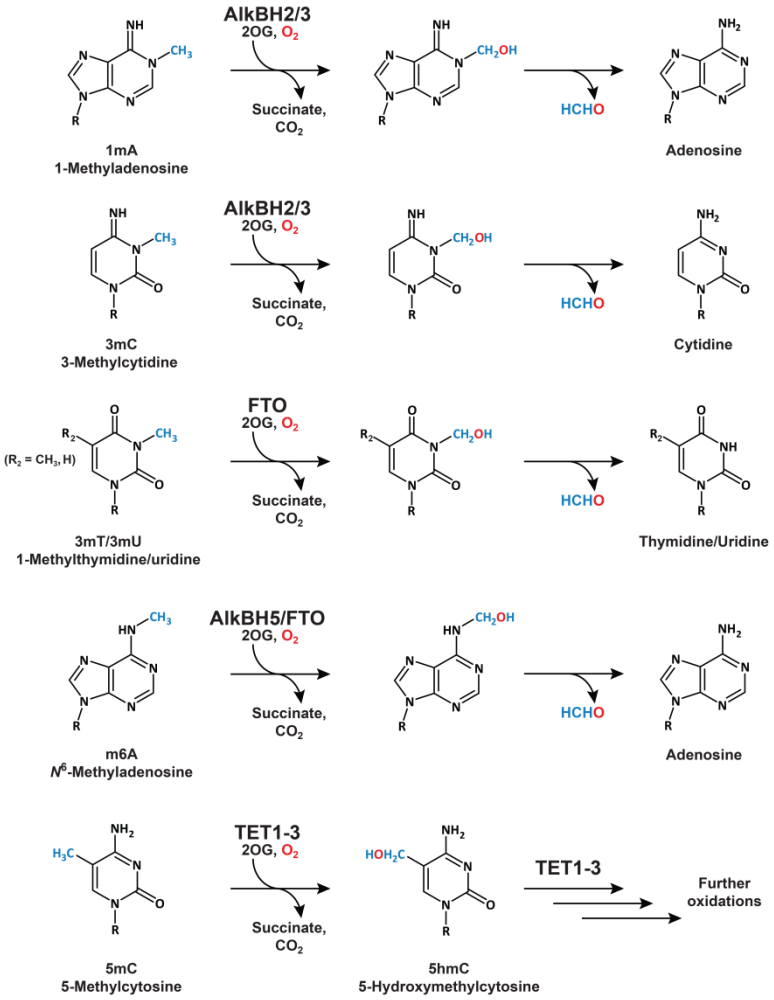


Fig. 1-13 | Examples of nucleic acid modifications catalysed by 2OG oxygenases.

1.2.2.5 Ribosomal Protein Hydroxylation

Recently, it was found that ribosome hydroxylation catalysed by 2OG oxygenases occurs in organisms ranging from prokaryotes to humans.[331] 2OG oxygenases may therefore also be involved in translational regulation. The human 2OG oxygenases Myc-induced nuclear antigen (MINA53, a 53 kDa protein) and nucleolar protein 66 (NO66) catalyse hydroxylation of certain histidyl residues in the ribosomal proteins RPL27a and RPL8 respectively (**Fig. 1-14**).[331] In addition, MINA53 and NO66 have also been reported to be histone demethylases;[332,333] however, these activities were not observed by other groups and are hence controversial.[331] The functions of ribosome hydroxylations catalysed by MINA53 and NO66 are not known to date and require further investigation.

The human 2OG- and Fe(II)-dependent oxygenase domain-containing protein 1 (OGFOD1) was recently reported to hydroxlyate a conserved prolyl residue in the small ribosomal subunit protein RPS23 (**Fig. 1-14**).[334] A role in translational regulation as a modulator of translation accuracy has therefore been proposed for OGFOD1.[335]

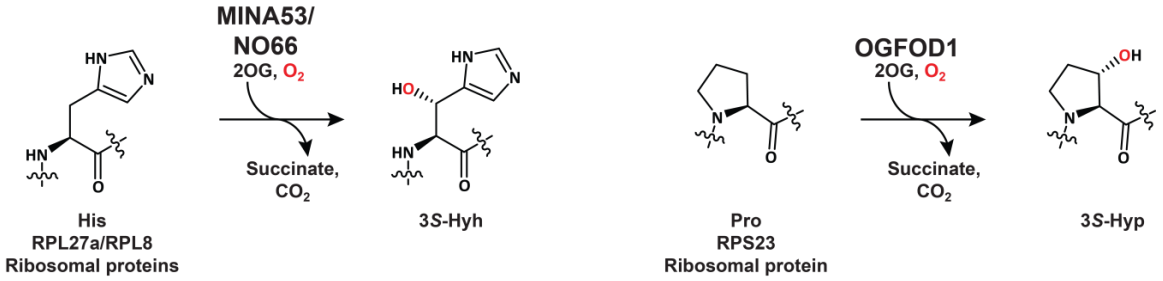


Fig. 1-14 | 2OG-oxygenase-catalysed ribosomal protein hydroxylation on histidyl or prolyl residues. Hyh: hydroxyhistidine; Hyp: hydroxyroline.

MINA53 is highly expressed in many cancers and linked to increased cell proliferation and oncogenesis (see **Table 1-1**). It has been reported that *MINA53* is a direct target gene of the oncogenic transcription factor and cell proliferation mediator c-Myc.[210] Overexpression of NO66 in non-small cell lung cancer has also been reported.[207] High *MINA53* expression correlates with poor prognosis for survival and hence also serves as a prognostic marker. The links between the enzymatic activities of MINA53 and NO66 and cancer yet need to be elucidated, but both proteins are potential therapeutic targets for cancer treatment.

1.3 EPIDERMAL GROWTH FACTOR-LIKE DOMAIN HYDROXYLATION

The discovery that certain aspartyl and asparaginyl residues in extracellular epidermal growth factor-like domains (EGFDs) are hydroxylated[336,337] was a landmark in posttranslational modification research, because it demonstrated that proteins other than collagen/collagen-like domains are directly and extensively modified by reactions involving atmospheric oxygen. EGFD β-hydroxylation is catalysed by the aspartyl/asparaginyl-β-hydroxylase (AspH, or in the literature also known as AAH, EGFH, BAH), an Fe^{2+}- and 2OG-dependent oxygenase.[338-340] AspH was first isolated in 1989 from rat[338] and bovine liver microsomes,[341] and later from human osteosarcoma cDNA.[342] However, a comprehensive functional and structural characterisation of human AspH is lacking thus far, and the function of EGFD hydroxylation hence remains unclear to date.

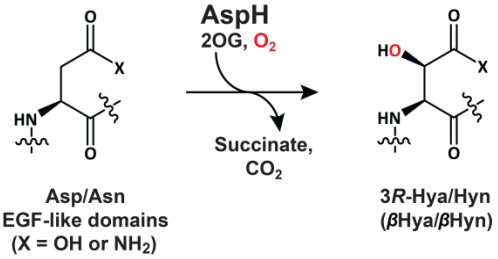


Fig. 1-15 | Reaction catalysed by AspH. A specific aspartyl or asparaginyl residue in EGF-like domains is hydroxylated on the β-carbon. Hya: hydroxyaspartate; Hyn: hydroxyasparagine.

AspH catalyses the 3*R*-hydroxylation of Asp/Asn residues involved in calcium (Ca^{2+})-binding in a subtype of EGFDs. Hydroxylation by AspH gives the *erythro* stereoisomers (2*S*,3*R*)-hydroxyaspartate (*β*Hya) or (2*S*,3*R*)-hydroxyasparagine (*β*Hyn) respectively (**Fig. 1-15**).[343] The HIF hydroxylase FIH is the only other eukaryotic asparaginyl hydroxylase thus far identified, but it shows no sequence or significant secondary structure similarity to AspH. Furthermore, in contrast to AspH, FIH catalyses formation of the *threo* isomer ((2*S*,3*S*)-hydroxyasparagine, see **Fig. 1-10**).[296]

1.3.1 Epidermal Growth Factor-like Domains

The epidermal growth factor-like domain (EGFD) is an evolutionarily conserved protein domain of about forty amino acids length, defined by six cysteine residues that form disulfide cross-links in characteristic manner. The epidermal growth factor (EGF) precursor is the archetype of all EGFDs.[344,345] This membrane spanning protein contains nine EGFDs, but only the domain closest to the membrane is proteolytically cleaved and released as active EGF (53 amino acid residues).[346,347] The disulfide bonds have been found to be linked in a Cys1-3, 2-4, and 5-6 pattern (**Fig. 1-16**). This disulfide pattern is commonly accepted as being canonical and believed to be required for biological activity.[344,348] The site of EGFD hydroxylation is positioned between the third and the fourth cysteine (abbreviated as C_3 and C_4) of the domains in the consensus sequence C_3X(**D**/**N**)XXXX(Y/F)XC_4XC_5 (with X being any proteinogenic amino acid).[338]

EGFDs usually occur in repeats, and a subtype comprising about half of the identified EGFDs in humans contains the consensus sequence (D/N)X(D/N)(Q/E/D)C_1…C_3X(D/N)XXXX(Y/F)XC_4 for Ca^{2+}-binding (cbEGFDs). This Ca^{2+}-binding consensus sequence is always accompanied by the consensus for Asp/Asn hydroxylation, whereas there are a few cases of EGFDs that are hydroxylated or contain the hydroxylation consensus sequence but not the *N*-terminal Ca^{2+}-binding residues (e.g. protein S EGF1, protein Z EGF1, and several domains in the notch ligand delta-like 1).

The specific canonical disulfide pattern is known to endow EGFDs with a typical fold with a core of two adjacent antiparallel *β*-strands. This fold is rigidified upon Ca^{2+} coordination.[349] Only "correctly" folded EGFDs with canonical disulfides and presence of the core *β*-strands are believed to be able to bind Ca^{2+}.[350] The crystal structure of the first EGFD of human coagulation factor IX in complex with Ca^{2+} (hFIX EGF1, PDB 1EDM,[351] **Fig. 1-16**) is representative of the typical EGFD fold. The Ca^{2+} ion is coordinated by seven ligands in geometry of a pentagonal bipyramid.[351] The Ca^{2+}-binding residues in a cbEGFD are the side chains of the amino acids corresponding to Asp47, Gln50, and Asp64/*β*Hya64 as well as the backbone carbonyls of Gly48 and Asp65.[352] The highly conserved Asp49, which is part of the Ca^{2+}-binding consensus sequence, does not coordinate the Ca^{2+} ion. The seventh ligand is assumed to be a solvent molecule or to be donated from a neighbouring domain in intact proteins within EGFD repeats. The hydroxyl group of the *β*Hya/*β*Hyn residue is likely not a direct Ca^{2+}

ligand; the side chain 3*R*-position is orientated away from the Ca^{2+}, as observed in the available X-ray and NMR structures (see **Fig. 1-16**). No significant effects of hydroxylation on Ca^{2+}-binding[353] nor on clotting[339] or anticoagulant cofactor activity[354] were observed when fully *β*-hydroxylated and unhydroxylated EGFDs were compared.

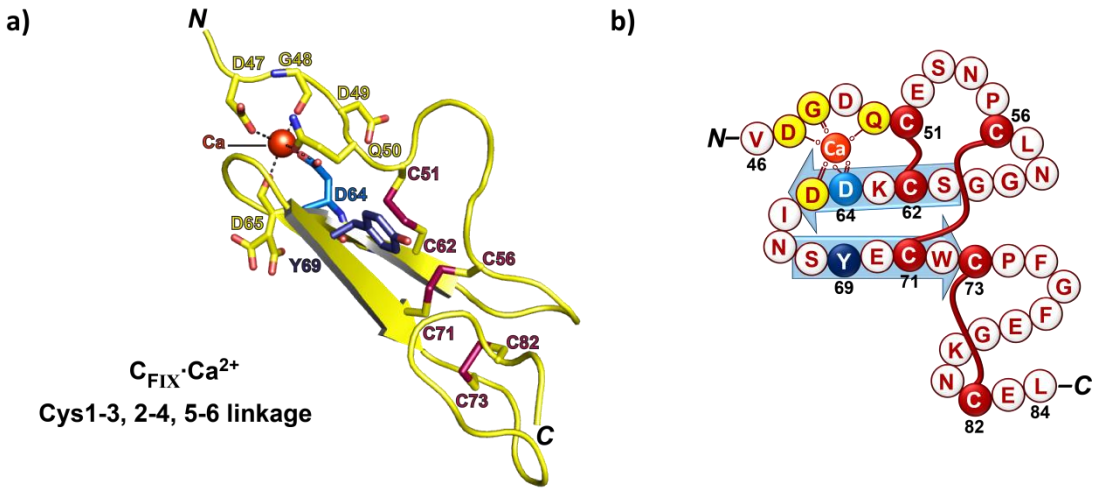


Fig. 1-16 | Structure of a typical EGF-like domain, the first EGFD of human coagulation factor IX (hFIX EGF1). **a)** X-ray crystal structure of hFIX EGF1 in complex with Ca^{2+} (resolution: 1.50 Å; PDB: 1EDM[351]; side chain of Asp65 found in two conformations) showing the typical fold of Ca^{2+}-binding EGFDs with canonical disulfide linkage in Cys1-3, 2-4, 5-6 pattern (C_{FIX}). C_{FIX}·Ca^{2+}: canonical EGFD form of hFIX EGF1 in complex with Ca^{2+}. **b)** Corresponding schematic. The hydroxylation site Asp64 is marked in bright-blue, whereas the remaining Ca^{2+}-binding residues are highlighted in yellow, and the conserved Tyr residue on the adjacent *β*-sheet is coloured dark-blue.

EGFDs are found in a wide range of cell surface-bound and secreted proteins with a diversity of functions in metazoans.[355] Secreted proteins containing EGFDs with the consensus sequence for Asp/Asn hydroxylation include coagulation factors (VII, IX, X, XII), proteins of the anti-coagulant (vitamin K-dependent proteins C, S, Z) and the complement system (complement C1r and C1s subcomponents), uromodulin, fibrillins and latent transforming growth factor β-binding proteins (LTBPs), which are major protein families of the extracellular matrix (**Fig. 1-17**). Many cell membrane-bound proteins such as the EGF precursor, thrombomodulin, the low-density lipoprotein receptor (LDLR), notch as well as notch ligands, delta-like and jagged, also contain several EGFDs with the hydroxylation consensus sequence.[355] Frequently, only partial hydroxylation was found when EGFDs were extracted from plasma and analysed for their *β*Hya/*β*Hyn content by amino acid analysis.[336] For example, the first EGFDs of both human and bovine coagulation factor X (FX) have been reported to be nearly fully hydroxylated, whereas only approximately 30 %-60 % hydroxylation was found in the corresponding EGFDs of coagulation factor IX (FIX).[356]

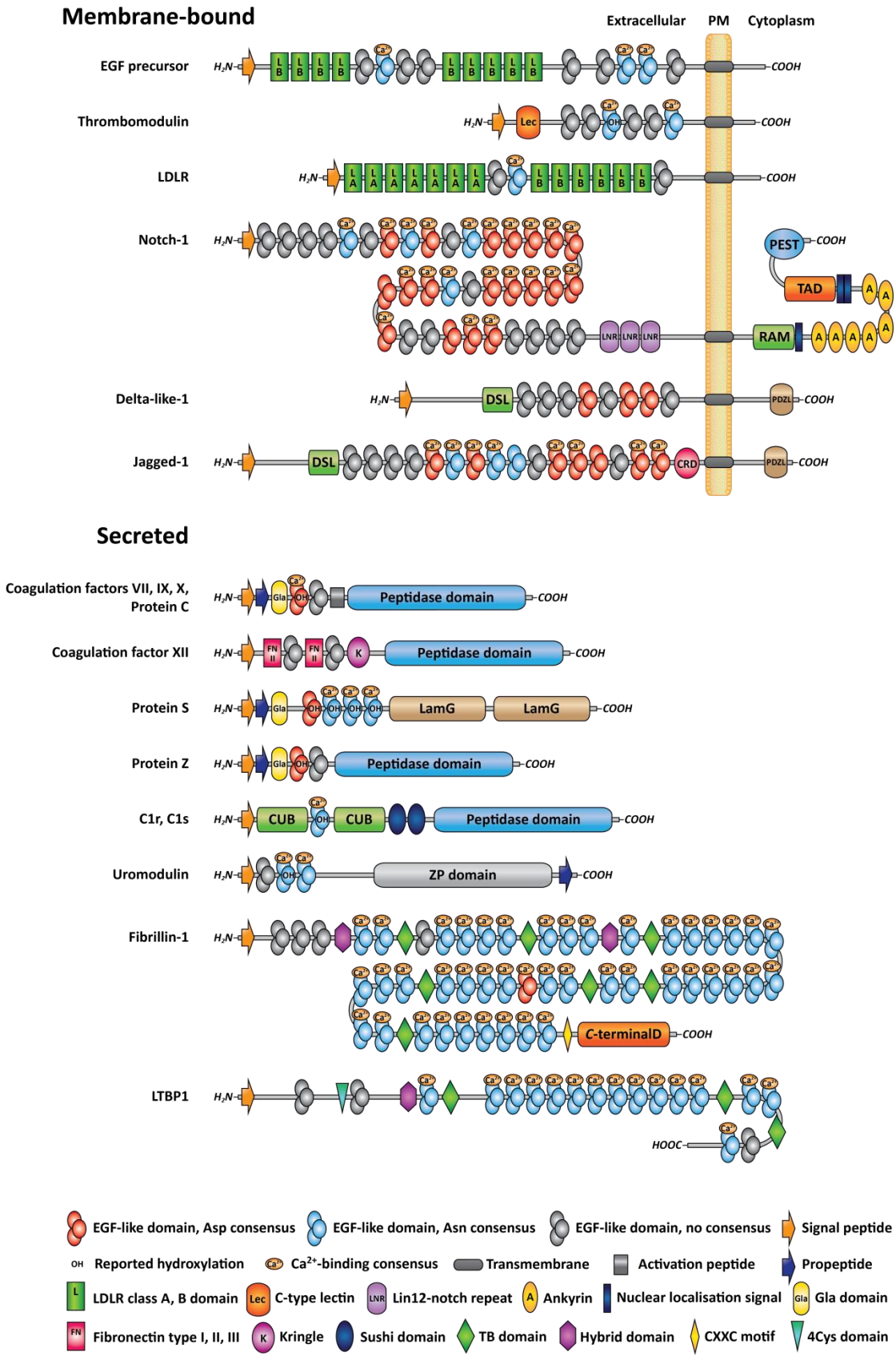


Fig. 1-17 | Domain architecture of selected EGFD proteins. EGFDs usually occur in repeats. PM: plasma membrane; LDLR: low-density lipoprotein receptor; RAM: recombination signal-binding protein 1 for J-κ association module; TAD: trans-activating domain; PEST: proline, glutamic acid, serine/threonine-rich motif; DSL: delta/serrate/LAG-2 domain; PDZL: PDZ domain; CRD: cysteine-rich domain; Gla: ; LamG: laminine G domain; CUB: CUB domain; ZP domain: *zona pellucida*-like domain; LTBP1: latent transforming growth factor β-binding protein 1; TB domain: transforming growth factor β-binding domain. For references to the reported hydroxylation see **Chapter 5**, **Fig. 5-4**.

The exact roles of EGFD modules in proteins are not fully understood. It has been suggested that EGFDs could be important for protein-protein interactions.[355] EGFD repeats are often found adjacent to other distinct domains or repetitive motifs such as likewise disulfide-rich transforming growth factor β-binding (TB domains) or γ-carboxyglutamic acid (Gla)-containing domains. In addition, EGFDs in the vitamin K-dependent coagulation proteins have been reported to be required for proper folding of the adjacent Gla domains.[357] The largest numbers of EGFDs, however, are found in fibrillins and notch proteins (e.g. 47 EGFDs in human fibrillin-1, 36 EGFDs in human notch-1). Most of the EGFDs in these proteins contain the consensus for hydroxylation as well as Ca^{2+}-binding (see **Fig. 1-17**), but in most cases it is not known whether they are hydroxylated *in vivo* and to which extent.

1.3.2 Links Between AspH and Diseases

An absence of posttranslational EGFD Asp/Asn β-hydroxylation in mice leads to developmental defects similar to deficiencies observed by interruption of Notch signalling and causes increased tumour growth incidence.[263] The distinct phenotype of the knockout mice includes a shortened snout compared to the normal breed, syndactyly and cleft palate.[263]

An enzymatically active AspH form and its shorter isoform humbug are strongly expressed in various types of cancer including hepatocellular, pancreatic, prostate, lung, renal, brain, ovarian, breast, colorectal and bile duct carcinoma as well as leukaemia (**Fig. 1-18**).[252, 255-258] In contrast, normal tissues only exhibit low levels of AspH, enabling the use of AspH as a tumour biomarker.[264] Furthermore, *ASPH* is very strongly upregulated by hypoxia.[260] Injection of stably transfected clones overexpressing murine AspH into nude mice led to the formation of tumours.[261] Moreover, high expression of AspH in an enzymatically active form is associated with malignant transformation,[261,252] worse clinical outcome and poorer prognosis for survival.[264,253] The role of AspH in cancers, and if inhibition of its enzymatic activity can be a potential therapeutic intervention, needs to be further investigated, but it is already being used as a biomarker for cancerous tissue in histopathological studies.[262] Furthermore, the US companies Panacea Pharma and its Canadian spin-off Panacea Global, Inc. offer cancer diagnostics based on detection of AspH using an antibody.[358,359]

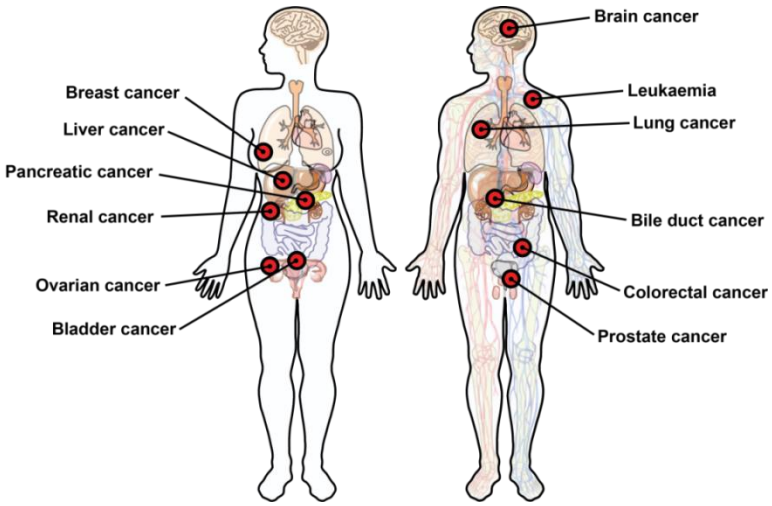


Fig. 1-18 | High overexpression of AspH has been found in various cancers.

Recently, *ASPH* mutations have also been linked to facial dysmorphism, lens dislocation and anterior-segment abnormalities in humans with Traboulsi syndrome.[360] The authors suggested that the phenotype observed in humans corresponds well to the developmental defects found in the mouse knockout model.

1.3.3 Gene Organisation and AspH Splicing Isoforms

The human *ASPH* (National Center for Biotechnology Information (NCBI) database[361] gene ID: 444, A*β*H-J-J locus) expression is regulated by two promoters and its messenger RNA transcript undergoes extensive alternative splicing,[362,363] resulting in twelve AspH isoforms (AspHA-L, NCBI database, **Fig. 1-19**). Most of these isoforms are *C*-terminally truncated, lacking the catalytic domain.

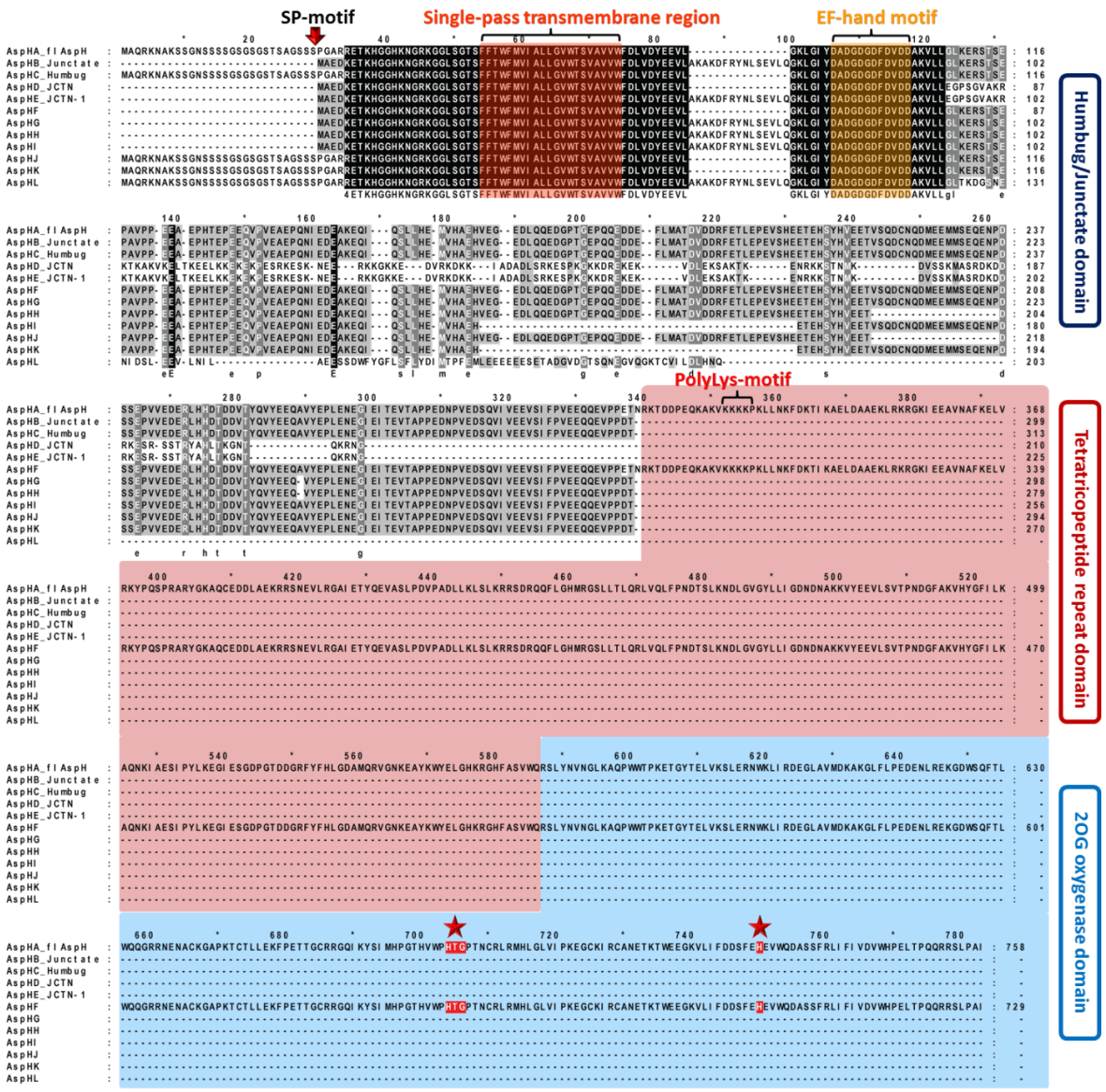


Fig. 1-19 | Sequence alignment (performed using Clustalω[364]) of the twelve human AspH splicing isoforms listed in the NCBI database[361] (AspH gene ID: 444). AspHA is the longest isoform, thus referred to as full-length AspH (flAspH), and contains 758 amino acids. It has a humbug/junctate domain comprising the short cytoplasmic *N*-terminus, a single-pass transmembrane region (TM), a Ca^{2+}-binding EF-hand motif, and highly negatively charged part within the ER lumen, a mainly helical TPR domain, and the *C*-terminal oxygenase domain harbouring the catalytic function. JCTN = Junctin; JCTN-1 = cardiac Junctin. The HxG…H catalytic dyad is marked with red stars, and other distinctive motifs are indicated on the alignment.

The transcript encoding for AspHA represents the longest isoform consisting of 758 amino acids and is thus accepted as canonical full-length AspH (flAspH, **Fig. 1-20**). This isoform has been used as basis for preparation of constructs in this Thesis. Other distinct isoforms are junctate[365] (AspHB), humbug[366] (AspHC), and junctin[367,368] (AspHD, JCTN, and cardiac Junctin-1 AspHE) but all lack the catalytic domain. AspH isoforms G-I are junctate-like, while isoforms J-K are humbug-like. *AspH* is thought to play an important role in Ca^{2+} homeostasis, since the encoded isoforms junctate and junctin have been shown to regulate Ca^{2+} release from the endoplasmic or sarcoplasmic reticulum respectively.[369,370] Further information on the AspH isoforms can be found in **Chapter 8** and **Appendix B**.

A sequence alignment (**Fig. 1-20**) shows that all isoforms contain a type II single-pass transmembrane region spanning the ER membrane. In addition, a conserved Ca^{2+}-binding EF-hand motif has been identified within an *N*-terminal luminal humbug/junctate domain.[371] Only AspHA and AspHF have additional *C*-terminal tetratrico-peptide repeat (TPR) and oxygenase domains. TPR domains alone have no catalytic activity. Some other 2OG oxygenases including human collagen prolyl hydroxylases (both C-P4Hs and C-P3Hs), UTX and UTY of the KDM6 subfamily of histone demethylases also contain TPR repeats.

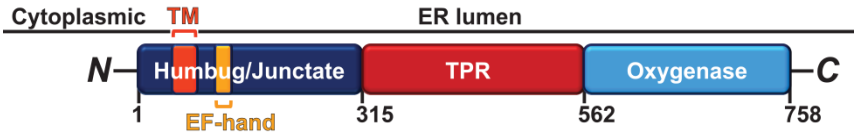


Fig. 1-20 | Proposed domain architecture of full-length AspH (AspH isoform A) as ER transmembrane protein. TM: transmembrane; TPR: tetratricopepide repeat.

Remarkably, bioinformatics analyses have suggested that the active site of AspH differs significantly from those of other 2OG oxygenases: AspH lacks the carboxylate-containing residue of the highly conserved HXD/E…H 'facial triad' Fe(II)-binding motif,[372] which is typical of 2OG oxygenases (see **Section 1.2.1**). AspH possesses a His-Xxx-Gly…His (HXG…H) catalytic dyad motif instead; the aspartate or glutamate residue is not present. This suggests that AspH contains two Fe(II)-binding residues only. A sequence alignment of human, bovine, murine and fruit fly AspH shows high sequence similarity between the oxygenase domains in the different species and conservation of the predicted HXG…H Fe(II)-binding motif (**Fig. 1-21**).

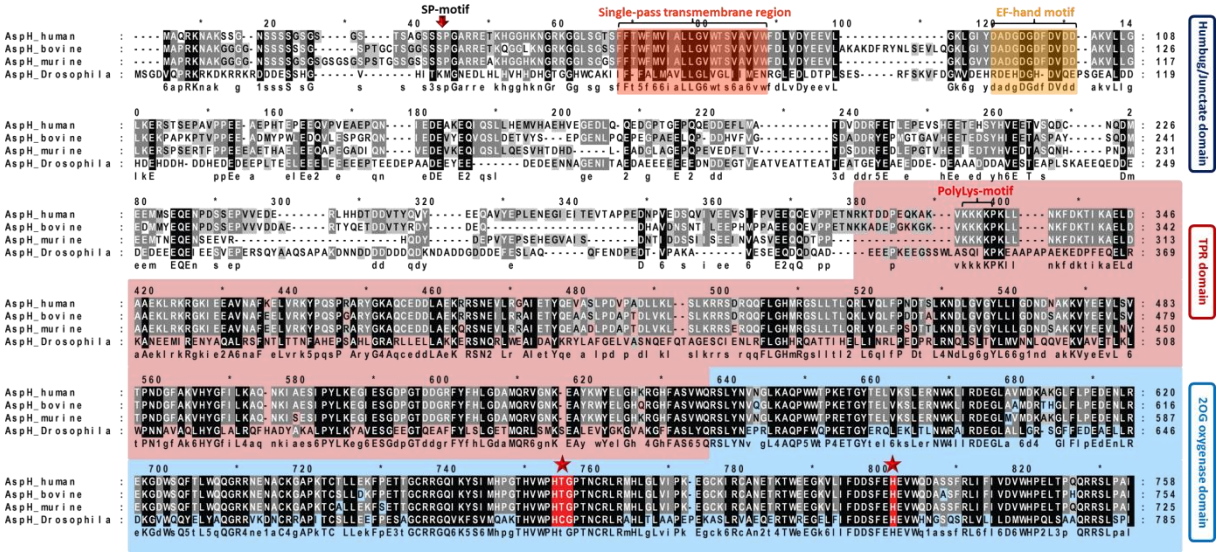


Fig. 1-21 | Sequence alignment of human (*Homo sapiens*, gene ID: 444), bovine (*Bos taurus*, gene ID: 286771), murine (*Mus musculus*, gene ID: 65973), and fruit fly (*Drosophila melanogaster*, gene ID 36778), showing high sequence similarity within the oxygenase domain and conservation of the HxG…H motif in all species.

1.4 THESIS OBJECTIVES

The introduction to this Thesis aims to outline how members of the 2OG oxygenase superfamily are linked to cancer and where therapeutic potential can be exploited.

One member of the human 2OG oxygenases 'sticks out': The predicted unusual active site metal coordination motif (a catalytic dyad HXG…H motif instead of the typical HXD/E…H 'facial triad') makes AspH a particularly interesting 2OG oxygenase for structural and mechanistic studies. Furthermore, despite the efforts of researchers over the last decades since the discovery of AspH in 1989, the physiological role of AspH-catalysed EGFD hydroxylation remains unclear to date. The described strong links between enzymatically active AspH and cancer (**Section 1.3.2**) further rationalise functional studies on the role of EGFD hydroxylation. The phenotype in the reported mouse knockout and the disease links demonstrate that EGFD hydroxylation is of biological significance. The main focus of the work described in this Thesis was to provide an *in vitro* characterisation of human AspH to gain a better understanding of its potential physiological role (**Fig. 1-22**).

Specific aims of the research described herein:

1) The establishment of assays to test the enzymatic activity of human AspH and evaluation of the enzyme-substrate interactions.
2) Studies on structure-function relationships of human AspH by performing crystallographic studies with suitable protein constructs. Investigation of the predicted unusual active site metal coordination and comparison to other members of the 2OG oxygenase superfamily.

3) Studies on AspH-substrate complexes by crystallographic and mass spectrometry-based methods, in particular on how AspH recognises and binds its substrates.
4) Exploration of the substrate selectivity and scope of human AspH *in vitro*, and the development of alternative cyclic peptide-based substrates.
5) Determination of the role of the TPR domain in AspH-catalysed EGFD hydroxylation.
6) Investigation of the effects of Ca^{2+} ions on hydroxylation of Ca^{2+}-binding EGFDs.
7) Studies on the subcellular localisation of catalytically functional AspH constructs and linking the insights gained from structural and functional studies to a cellular context.
8) Exploration of approaches to study interactions between AspH or other 2OG oxygenases and small molecules to assist the development of selective inhibitors/activators in future studies.

A primary interest of the work described in this Thesis was to link the biochemical and structural results with the biological function of AspH-catalysed EGFD hydroxylation. Ultimately, the findings will help to pave the way for the development of new methods for cancer treatment.

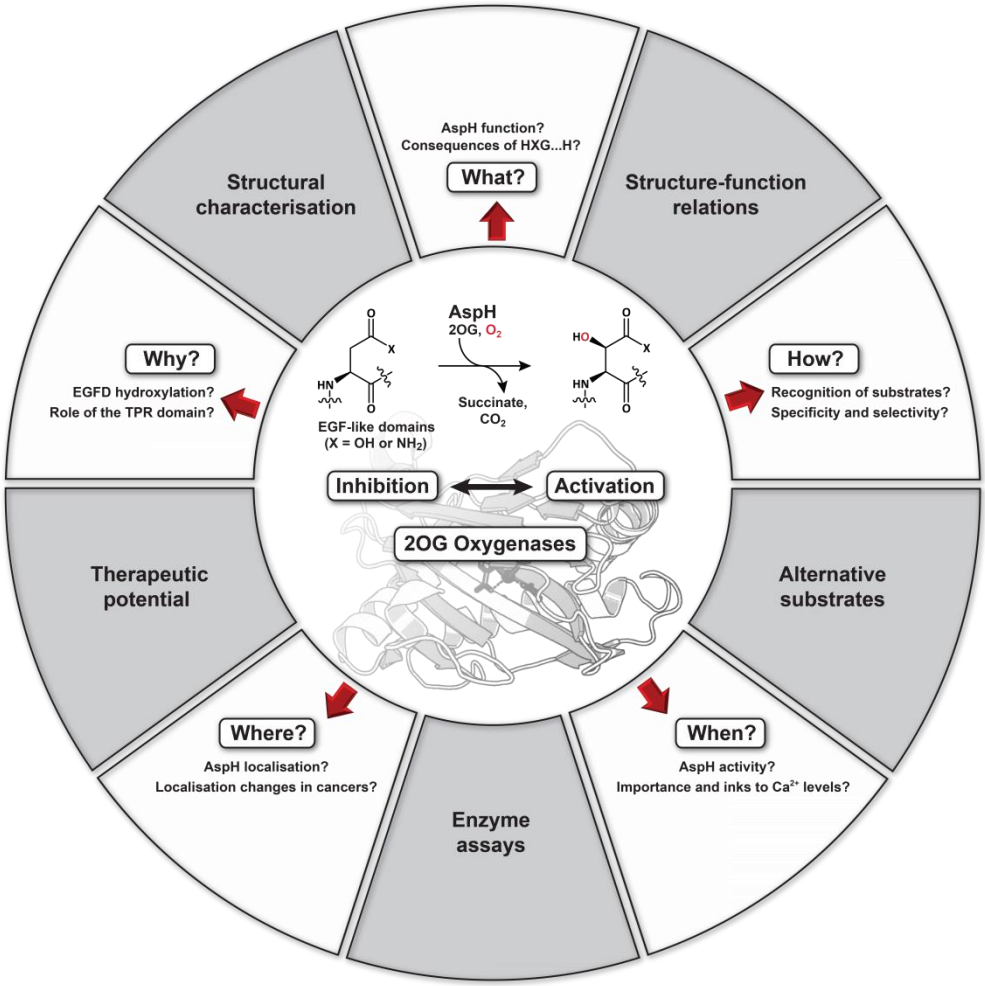
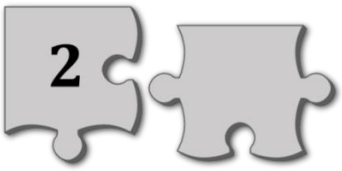
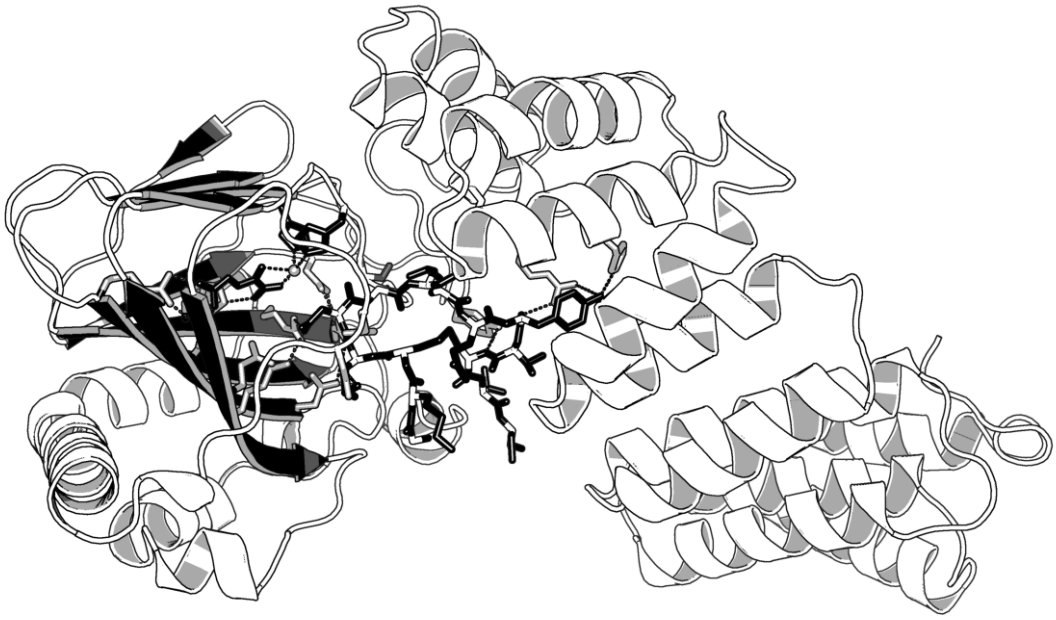


Fig. 1-22 | Outline of the aims of the research described in this Thesis. The work is focused on an *in vitro* characterisation of human AspH to gain a better understanding of the potential physiological role of the unusual 2OG oxygenase.



Form Ever Follows Function:
Structural and Functional Studies of AspH



Abstract

This Chapter covers the aspects that led to the selection of an AspH construct for structural work, the development of assays to test for enzymatic activity, and the first crystallographic studies on AspH.

2.1 ASPH *IN VITRO* CONSTRUCT DESIGN

Finding a suitable construct for crystallographic studies poses a common bottleneck in structural biology.[373,374] The protein not only needs to be soluble and producible in sufficient quantities but should also fold properly and have a high probability to crystallise. Expression constructs should therefore be rationally designed and often need to be optimised. Dividing a protein into individual domains can help predict truncations for constructs with high probability to fold and crystallise. Frequently, flexible regions like the *N*- and *C*-termini, internal loops or even entire domains of the protein need to be removed to obtain a crystallisable sample. Different experimental and computational methods have proven useful for this approach. Experimental information on compact, globular domains can be gained from limited proteolysis of the proteins and subsequent mass spectrometry analysis or *N*-terminal sequencing for identification of the fragments.[375-377] Domain boundaries can also be assigned based on computational analysis. Sequence alignments (e.g. using common software such as Clustalω,[364] T-Coffee,[378] or BLAST[379] for database searching) can identify sequence conservation patterns among family members and help predict secondary structure and unfolded or disordered regions (e.g. using DALI,[380] DISOPRED2,[381] comparative modelling using 3D-Jury[382]).

The motivation for obtaining a crystal structure is often to aid the study of structure-function relationships. A crystal structure captures static states of the protein. The development of methods to assay the activity of the protein constructs is therefore a complementary tool to validate the construct design and resulting structures. Activity such as enzyme catalysis or protein-ligand interactions is often facilitated by a discrete domain.[373] Thus, the expression of a full-length construct is not always necessary to address research problems.

The initial work on AspH was focused on the oxygenase domain harbouring the catalytic function, as has been done for other 2OG oxygenases embedded in multi-domain proteins (e.g. JmjC histone demethylases such as JMJD2A,[383] PHF8,[384] UTX[385]). Sequence alignment analyses (see **Fig. 1-20**) of the twelve AspH isoforms indicated several domains, the *N*-terminal humbug/junctate domain (amino acid (aa) residues 1-314), the TPR domain (mainly helical structure, aa 315-561) and the 2OG oxygenase domain (aa 562-758). In addition, enzymatically active 52 kDa and 56 kDa AspH species had originally been isolated from bovine liver, which were proposed to be products of proteolytic cleavage from the *C*-terminal (TPR and oxygenase) domains of the full-length enzyme.[386,387] Similar to a limited proteolysis approach,[375-377] this suggested a stable construct corresponding to amino acid residues 315-758 of human AspH. Moreover, recombinant production of a bovine 52 kDa protein construct in *E. coli* has also been reported.[387]

Initial constructs for the work on AspH were cloned by researchers at *ReOx Ltd*. Dr Kirsty S. Hewitson and Dr Luke A. McNeill provided a construct encoding for human $AspH_{315-758}$ with an

N-terminal glutathione *S*-transferase (GST)-tag coding sequence (GST-AspH$_{315-758}$) in the pGEX-6P-1 vector. Obtaining a soluble, enzymatically active *N*-terminally hexahistidine (His$_6$)-tagged protein was not successful (personal communication, Dr Kirsty S. Hewitson).

Dr Grazyna Kochan at the Structural Genomics Consortium (SGC) Oxford produced an oxygenase domain construct of human AspH comprising amino acid residues 562-758 with an *N*-terminal His$_6$-tag coding sequence (His$_6$-AspH$_{562-758}$) in the pNIC28-Bsa4 vector. This was the longest soluble protein that the SGC researchers were able to obtain with their high-throughput construct design approach[373,388] (personal communication, Dr Grazyna Kochan). Enzymatic activity of this construct was not assessed by the SGC.

The GST-AspH$_{315-758}$ and His$_6$-AspH$_{562-758}$ were produced for the initial studies in this Thesis. The plasmids containing these expression constructs were gifts from Dr Kirsty S. Hewitson and Dr Grazyna Kochan. Established protocols for expression and purification were followed as described below. To date, no integrated molecular analysis of genomes and their expression complementary DNA clone (I.M.A.G.E.[389] cDNA clone) for flAspH is available. The sequence encoding for flAspH was therefore later cloned as described in **Chapter 8**.

2.1.1 Oxygenase Domain Protein Production

Recombinant His$_6$-AspH$_{562-758}$ (**Fig. 2-1a**) was produced adapting the SGC protocol, as described in **Chapter 10**. Optimal construct expression conditions in *E. coli* were established by Dr Grazyna Kochan. The *E. coli* cells (BL21 (DE3)) were grown for 16 h at 18 °C post induction (addition of *iso*propyl-*β*-D-thiogalacto-pyranoside (IPTG) to a final concentration of 0.1 mM). Purification using Ni^{2+}-affinity and subsequent size-exclusion chromatography achieved protein of high purity, as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (**Fig. 2-1b** and **c**).

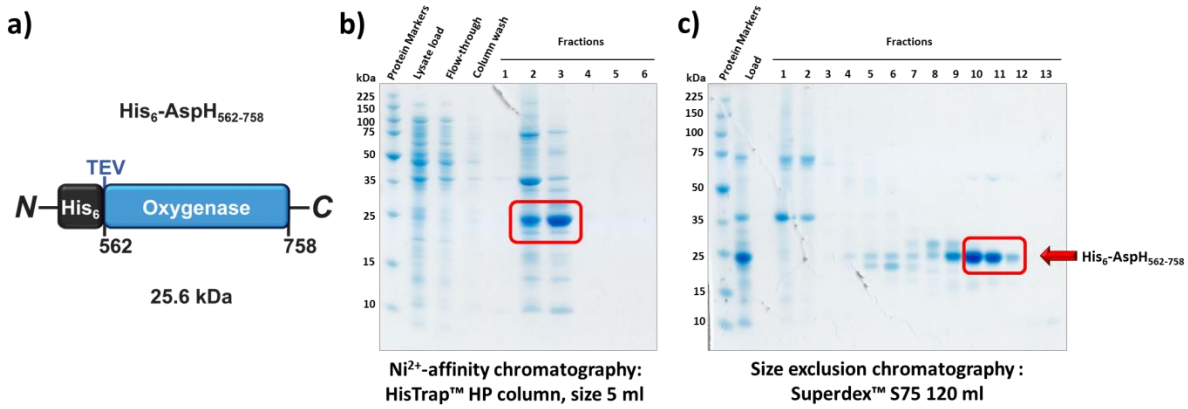


Fig. 2-1 | Protein production of His$_6$-AspH$_{562-758}$. **a)** Schematic of the His$_6$-tagged oxygenase domain (His$_6$-AspH$_{562-758}$). The construct was cloned into the pNIC-28-Bsa4 vector. SDS-PAGE gels with Coomassie staining depicting a typical purification: **b)** 10 %-12 % SDS-PAGE gel after Ni^{2+}-affinity purification; **c)** 10 %-12 % SDS-PAGE gel after size exclusion chromatography (Superdex™ 75 120 ml).

The correct mass of the protein (25,601 Da, the peak of highest intensity in the deconvoluted mass spectrum) was confirmed using intact protein electrospray ionisation mass spectrometry (ESI-MS) under denaturing conditions (**Fig. 2-2**). The deconvoluted mass spectrum shows two major peaks. The second peak at 25,779 Da likely corresponds to the protein with gluconoylated His_6-tag (a mass addition of +178 Da), which is a common and well-documented covalent modification to *N*-terminally His_6-tagged proteins produced in *E. coli*.[390,391]

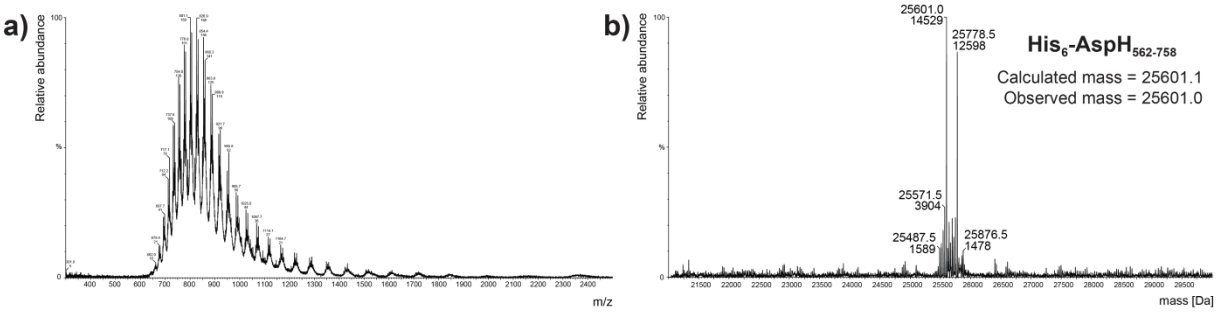


Fig. 2-2 | ESI-MS analysis of purified His_6-$AspH_{562-758}$. **a)** Raw spectrum (charge state envelope); **b)** deconvoluted mass spectrum. The expected mass of the protein (25,601 Da) was confirmed. The peak at 25,779 Da corresponds to the protein with gluconoylated His_6-tag[390,391] (a mass addition of +178 Da).

2.1.2 Two-Domain Construct Protein Production

Recombinant GST-$AspH_{315-758}$ (**Fig. 2-3a**) comprising 675 amino acids was produced following the *ReOx Ltd* protocol established by Dr Kirsty Hewitson, as described in detail in **Chapter 10**. The *E. coli* cells (BL21 (DE3)) were grown for 4 h at 37 °C post induction (addition of 1.0 mM IPTG). The protein was purified using GST-affinity and subsequent size-exclusion chromatography. SDS-PAGE analysis, however, showed four major bands, the top band around 75 kDa corresponding to the expected protein mass) with the lower three having lower molecular weights than that expected for GST-$AspH_{315-758}$ (**Fig. 2-3b** and **c**).

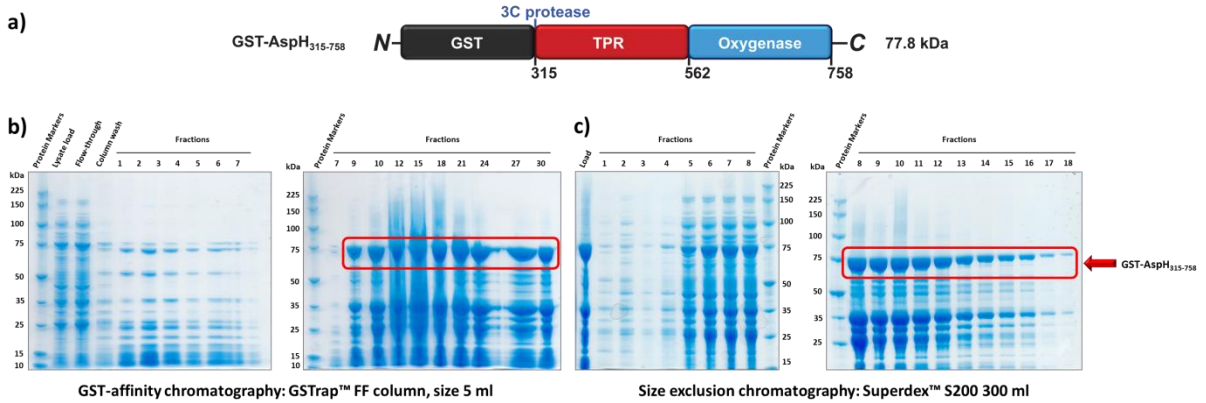


Fig. 2-3 | Protein production of GST-$AspH_{315-758}$. **a)** Schematic of the GST-tagged protein with both the TPR and oxygenase domains (GST-$AspH_{315-758}$). The construct was cloned into the pGEX-6P-1 vector.

SDS-PAGE gels with Coomassie staining depicting a typical purification: **b)** 10 %-12 % SDS-PAGE gels after GST-affinity purification; **c)** 10 %-12 % SDS-PAGE gels after size exclusion chromatography (Superdex™ 200 300 ml).

The difficulties in the isolation of this protein had already been observed (personal communicaton, Dr Kirsty Hewitson). The band at around 25 kDa could have been the result of proteolytic cleavage of the GST-tag. Purifying the protein in the presence of protease inhibitors did not prevent the presence of multiple bands observed on the SDS-PAGE gels.

The protein sample obtained from size-exclusion chromatography fractions could not be characterised by protein mass spectrometry. Investigation of the multiple bands by western blot analysis performed by Dr Kirsty Hewitson at *ReOx Ltd* using an anti-GST antibody (*New England Biolabs, NEB*) showed that several bands were recognised by the anti-GST antibody (**Fig. 2-4**). In contrast, His_6-tagged control proteins were not recognised. These results suggested that GST-$AspH_{315-758}$ protein degradation was occurring. Furthermore, the researchers performed *N*-terminal sequencing, which showed that all bands had the same *N*-terminal sequence corresponding to that of the GST fusion protein.

Despite these observed difficulties, the GST-$AspH_{315-758}$ resultant after size-exclusion chromatography was used for activity assays and comparison with the oxygenase domain construct.

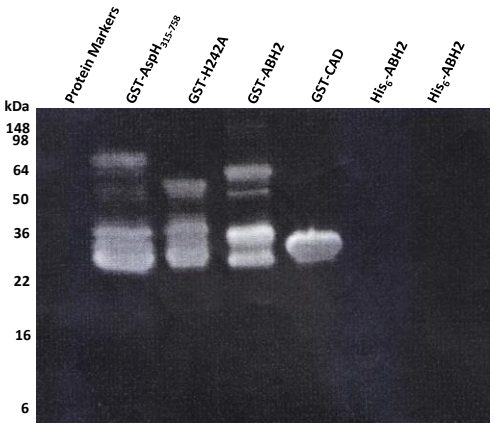


Fig. 2-4 | Western blot detection of GST-$AspH_{315-758}$ and control proteins carried out by probing with an anti-GST antibody (*NEB*). This experiment was performed by Dr Kirsty Hewitson at *ReOx Ltd*. Blot produced from a 15 % SDS-PAGE gel. All GST-tagged proteins were recognised by the antibody, whereas the His-tagged proteins served as negative controls.

2.2 AspH *IN VITRO* Activity Assays

2.2.1 Development of an MS-Based Hydroxylation Activity Assay

A number of different assay methods exist for probing the activity of 2OG oxygenases.[318] To be able to study structure-function relationships, a hydroxylation activity assay was developed. Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-ToF-MS) or liquid chromatography mass spectrometry (LC-MS) are well-suited mild ionisation methods for the

detection of peptides or other biomolecules.[392] Since EGFD peptides have been reported as substrates of AspH,[340-341, 356] a hydroxylation assay based on detection of the hydroxylated EGFD peptide product by mass spectrometry methods was developed. This gives direct quantitative information of hydroxylase activity, compared with other reported indirect methods for measuring AspH activity by cosubstrate 2OG turnover.[340-341, 393]

2.2.1.1 AspH *in vitro* Assay Substrate

The *in vitro* substrate chosen for the initial studies was a synthetic 39mer peptide representing the first EGFD of human coagulation factor X (hFX $EGF1_{39mer}$, **Fig. 2-5**), which was reported to be nearly fully hydroxylated when isolated from plasma.[356] In addition, *in vitro* activity of bovine AspH (bAspH) was demonstrated using synthetic peptides corresponding to the first EGFDs of bovine FX[341] and FIX[340].

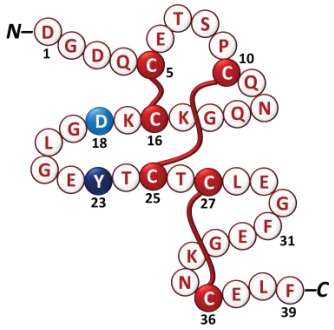


Fig. 2-5 | Schematic of the first EGFD of human coagulation factor X (hFX $EGF1_{39mer}$) shown with canonical disulfide links in a Cys1-3, 2-4, 5-6 pattern.

The linear peptide corresponding to the hFX $EGF1_{39mer}$ (LP_{39mer}, amino acids (aa) 86-124, reduced cysteines) precursor was prepared by solid-phase peptide synthesis (SPPS) and purified by *Peptide Synthetics* (Hampshire, UK). Different methods for the formation of the three disulfides were tested, for activity with recombinantly made GST-$AspH_{315-758}$ (personal communication, Dr Kirsty Hewitson). The synthetic procedures were established by *Peptide Synthetics*, and full details were hence not disclosed. Briefly, in the first tested method the disulfide links were formed by use of three orthogonal protecting groups for the cysteines enabling subsequent regioselective oxidation to give the hFX $EGF1_{39mer}$ with the specific canonical EGFD disulfide pattern (Cys1-3, 2-4, 5-6), as in a reported procedure.[394] However, the GST-$AspH_{315-758}$ did not show any activity when tested with the synthesised material.

A resynthesis of the LP_{39mer} peptide was then carried out using the same protecting group on all the cysteines. Following deprotection, purification and reduction, the three disulfides in the hFX $EGF1_{39mer}$ were then allowed to form by air oxidation of the free cysteines, as in a reported

procedure.[395] The air-mediated oxidation method relies on spontaneous folding *in vitro* to a native conformation predetermined by the amino acid sequence.[396] This was the same method as used for the synthesis of the bovine FX[341] and FIX[340] peptides that had shown activity with bovine AspH.[395] When tested with the GST-AspH$_{315\text{-}758}$, some activity was observed with this peptide material, as determined by a MALDI-ToF-MS-based hydroxylation assay (see **Section 2.2.1.2**). This peptide material was then used for subsequent hydroxylation activity assays.

2.2.1.2 MALDI-ToF-MS Hydroxylation Activity Assay

For MALDI-ToF-MS analysis, the peptide analyte (or other biomolecules such as oligonucleotides, small proteins, sugars or large organic molecules) is mixed and cocrystallised with a matrix on a MALDI target plate and ionised by laser excitation.[392,397] The high laser energy triggers desorption of the sample and the matrix material, and the analyte molecules are ionised by protonation or deprotonation. From the MALDI ionisation source, the charged molecules are then electrostatically directed into the time-of-flight (ToF) mass analyser, in which they are separated by their mass-to-charge ratio (m/z).[398,399] Common MALDI matrices are organic acids of low molecular weight that have strong optical absorption in either the UV or IR range to ensure protonation of the analyte, facile evaporation and efficient absorption of the laser irradiation respectively.[400] Some advantages of MALDI-ToF-MS are the comparably small amounts of sample required, the relatively fast and simple analysis, and that mainly the singly charged peptide analyte species are observed. This can reduce the complexity of spectra significantly.

For the hydroxylation assay using MALDI-ToF-MS analysis, recombinant AspH protein was incubated with the hFX EGF1$_{39mer}$ peptide in the presence of Fe(II), 2OG, O_2 and ascorbate (Asc) for a duration of typically 30-60 minutes (min) at 37 °C prior to analysis by MALDI-ToF-MS (see details in **Chapter 10**). The results were compared to a control mixture of the same reagents without the enzyme.

When the AspH oxygenase domain construct, His$_6$-AspH$_{562\text{-}758}$, was tested in the assay, no hydroxylation was observed. In contrast, the GST-AspH$_{315\text{-}758}$ showed activity, indicated by the detection of a peptide with a mass difference of +16 Da (**Fig. 2-6**), compared to the control, which showed no such mass shift. This result indicated that the TPR domain has a role in enzymatic catalysis. After 60 min incubation approximately 40 % of the peptide was hydroxylated.

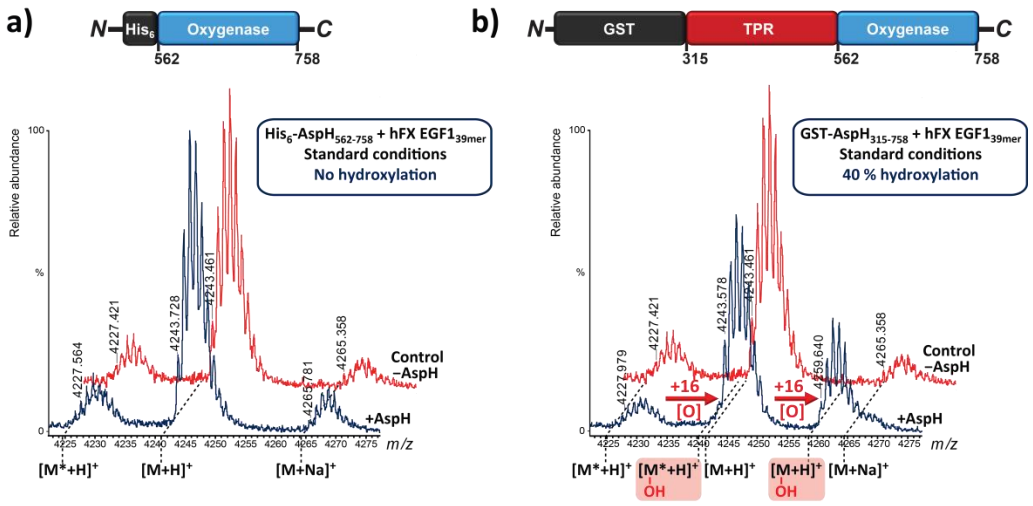


Fig. 2-6 | MALDI-ToF-MS analysis of the hydroxylation activity assay incubated for 60 min using His_6-$AspH_{562-758}$ or GST-$AspH_{315-758}$. **a)** Spectra of the hFX $EGF1_{39mer}$ peptide incubated with His_6-$AspH_{562-758}$ (SGC construct). Only the peptide molecule peak at m/z 4,243.7 for the monoisotopic mass of the singly charged species $[M+H]^+$ of the hFX $EGF1_{39mer}$ is observed, but no substrate hydroxylation. **b)** Spectra of the peptide incubated with GST-$AspH_{315-758}$ (*ReOx Ltd* construct) showing activity with the substrate. Hydroxylation is indicated by the additional peptide molecule peak $[M+H]^+$ + a mass difference of +16 Da (m/z 4,259.6) for the monoisotopic mass of the hydroxylated species $[M(OH)+H]^+$ of the hFX $EGF1(OH)_{39mer}$, corresponding to insertion of an oxygen atom during the reaction. 'Standard conditions' refers to the reaction buffer system (50 mM HEPES, pH 7.5) and presence of all cofactors/cosubstrates (Fe(II), Asc, 2OG, O_2). $[M^*+H]^+$ stands for a derivatised peptide species that also became hydroxylated. This species was further investigated in **Chapter 6**.

Relatively high laser energies were required to ionise the hFX $EGF1_{39mer}$ peptide in the MALDI-ToF-MS measurements. Isotope-resolved mass spectra were therefore only attained using the *Bruker Daltonics* Ultraflex I™ MALDI-ToF/ToF mass spectrometer for the measurements but could typically not be achieved using the *Waters* Micromass® MALDI micro MX™ mass spectrometer. The best results for the hFX $EGF1_{39mer}$ peptide were often obtained when 2,5-dihydroxybenzoic acid[401-403] (DHB) was used as MALDI matrix instead of the more common α-cyano-4-hydroxycinnamic acid[404] (CHCA).

2.2.2 2OG Turnover Assays

2.2.2.1 [1-^{14}C]-2OG Turnover Assays

His_6-$AspH_{562-758}$ and GST-$AspH_{315-758}$ were then tested in a radiochemical [1-^{14}C]-2OG turnover assay;[338, 405,393] this assay quantifies decarboxylation (turnover) of 2OG. This assay was conducted to test whether the observed activity was affected by the TPR domain. The reaction is followed by measuring the amount of carbon dioxide released specifically from 2OG in this reaction. [1-^{14}C]-labelled 2OG is added to the examined reaction mixtures, and the ^{14}C-carbon dioxide ($^{14}CO_2$) released by 2OG turnover is captured in solution by basic hyamine hydroxide (in a tube contained within the reaction vessel) as a bicarbonate adduct (scheme in **Fig. 2-7a**). This is then detected by liquid scintillation counting of the β-emission.

Substrate uncoupled turnover of 2OG is a common phenomenon observed for many 2OG oxygenases.[45] It has also been discovered that some 2OG oxygenases have stoichiometric ascorbate dependence when undergoing uncoupled turnover. This has been attributed to oxidation of the ferrous iron to an Fe(IV) intermediate, which deactivates the enzyme.[406] Ascorbate may reduce the Fe(IV) state.[407] Uncoupled 2OG turnover can be useful as a test for 2OG oxygenase activity of recombinantly produced enzymes even when substrates are unknown.

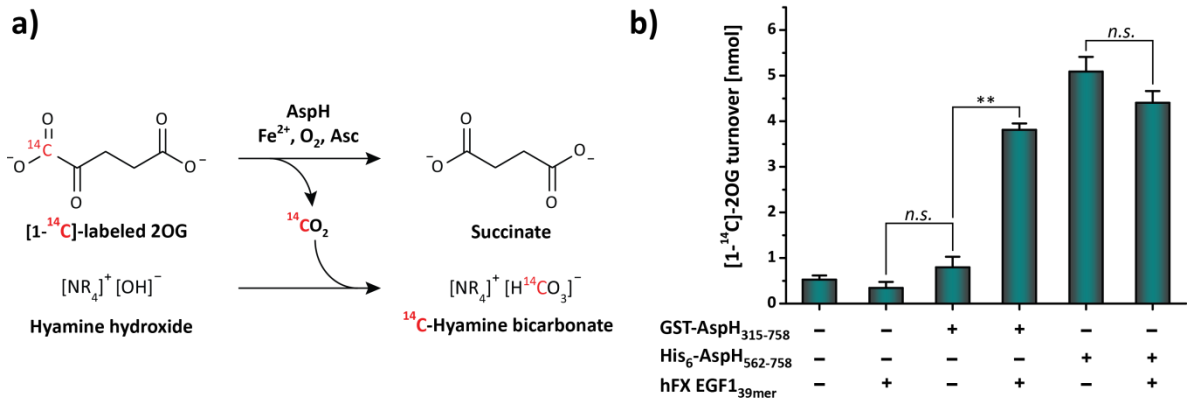


Fig. 2-7 | [1-^{14}C]-2OG turnover assay for AspH. **a)** Reaction scheme with release of $^{14}CO_2$, which is captured in solution as bicarbonate adduct by hyamine hydroxide, which is then detected by liquid scintillation counting of the β-emission. Hyamine hydroxide is abbreviated as $[NR_4]^+[OH]^-$. **b)** Results from the coupled (+ substrate) and uncoupled (- substrate) assays for His_6-$AspH_{562-758}$ and GST-$AspH_{315-758}$ in comparison to controls without enzyme (- enzyme). Significant turnover by the GST-$AspH_{315-758}$ was only stimulated in the presence of the hFX $EGF1_{39mer}$ peptide substrate (in the absence of the substrate: $p = 0.343$; in the presence of the substrate: $p = 0.005$). His_6-$AspH_{562-758}$ catalysed high 2OG turnover, which was independent of the presence of the hFX $EGF1_{39mer}$ substrate (no significant difference in the presence or absence of the substrate: $p = 0.368$). Data (background subtracted) represent the mean of triplicate experiments with error bars showing the standard deviation ($n = 3$; mean ± SD; $^{**}p < 0.01$; *n.s.*: not significant).

His_6-$AspH_{562-758}$ was observed to catalyse a high 2OG turnover, which was independent of the presence of the hFX $EGF1_{39mer}$ substrate (**Fig. 2-7b**). In contrast, significant turnover by the two-domain construct GST-$AspH_{315-758}$ was only stimulated in presence of the hFX $EGF1_{39mer}$ peptide substrate. These results support the proposal that 2OG decarboxylation is catalysed by the oxygenase domain. However, the results suggest that the TPR domain may be required for substrate hydroxylation and could thus be involved in substrate recognition, similar to the PHD zinc finger domain in 2OG oxygenases of the KDM7 subfamily of histone demethylases such as PHF8[408] (**Fig. 2-8**).

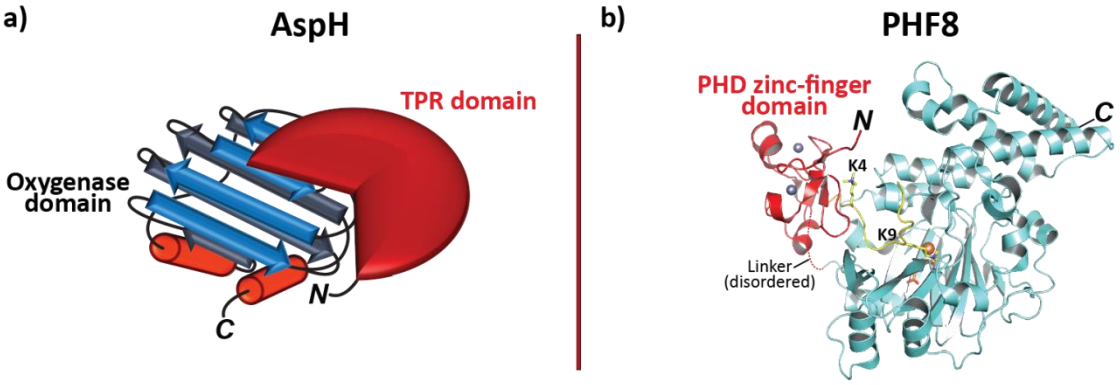


Fig. 2-8 | Functional analogy of non-enzymatic domains for substrate recognition. **a)** The TPR domain (red) in AspH, and **b)** the PHD zinc-finger domain (red) of PHF8 (PDB: 3KV4[408]); histone 3 substrate peptide in yellow.

2.2.2.2 Nuclear Magnetic Resonance 2OG Turnover Assays

The radiochemical 2OG turnover assay has disadvantages. For example, radioactive waste is generated, the assay is time-consuming and prone to imprecision owing to the technical complexity of the assay set-up, and a high background signal in the absence of enzyme. An alternative 2OG turnover assay using nuclear magnetic resonance (NMR) was therefore also developed for assaying AspH, using a modified version of a reported method.[409] 2OG turnover is detected by monitoring the relative intensities of the signals of succinate formation (observed as evolving singlet resonance at the chemical shift δ_H 2.32 ppm) and simultaneous 2OG depletion (decreasing intensity of triplet signal at δ_H 2.35 ppm) over time. The signals were assigned to the aliphatic protons of succinate or 2OG respectively by analogy to reference samples.

Time course data for substrate uncoupled 2OG turnover by GST-AspH$_{315\text{-}758}$ and His$_6$-AspH$_{562\text{-}758}$ were acquired by preparing samples containing enzyme, 2OG, ascorbate and Fe(II), and analysing the reaction by ^{1}H NMR using a 700 MHz NMR spectrometer (experiment conducted with assistance from Dr Richard Hopkinson who operated the spectrometer). In order to minimise the time delay between mixing and NMR analysis, the samples were pipetted into 2 mm NMR tubes, then transferred to the NMR spectrometer immediately after preparation. The ^{1}H NMR spectra after 2 h (shown in **Fig. 2-9**) support the results observed in the radiochemical 2OG turnover assay. For His$_6$-AspH$_{562\text{-}758}$ a singlet signal demonstrating succinate formation was clearly recognised at the appropriate chemical shift. GST-AspH$_{315\text{-}758}$ only shows very little succinate formation, which is in agreement with the findings from the radiochemical assay that 2OG turnover is only stimulated in presence of the substrate. However, compared with the 2OG turnover in the radiochemical assay, the signal for succinate formation is really low for the assay time of 2 h, which questions the significance of the results. This could be due to the limited oxygen availability in the NMR tube. Furthermore, the NMR assay requires substantial amounts of the reagents and measurement time. This assay was therefore not pursued.

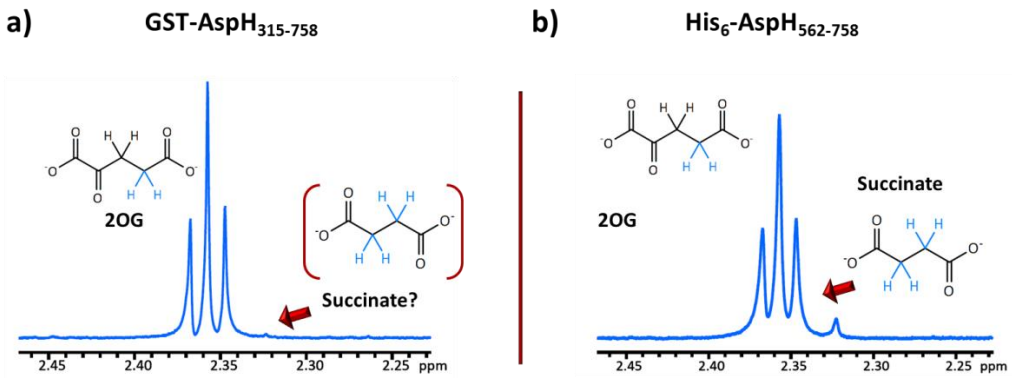


Fig. 2-9 | Substrate-uncoupled NMR 2OG turnover assay. **a)** GST-AspH$_{315-758}$, **b)** His$_6$-AspH$_{562-758}$. The assay was conducted in absence of peptide substrate. Shown spectra were obtained after 2 h incubation under NMR assay conditions (in presence of enzyme, Fe(II), Asc, 2OG, buffer, as stated in **Chapter 10**). The shown triplet represents the protons of 2OG; the singlet peak that of succinate.

2.3 CONSTRUCT OPTIMISATION

Use of GST-AspH$_{315-758}$ exhibited substantial disadvantages. Complications during purification had already been observed at *ReOx Ltd*. The purified protein seemed to degrade in a manner not prevented by addition of protease inhibitors, and mass spectrometry was inconclusive (see **Section 2.1.2**). Nevertheless, the protein showed activity with the hFX EGF1$_{39mer}$ substrate. The relatively large GST-tag (~26 kDa) is often thought to help solubilising the GST fusion partner protein. However, due to its size and solubilising properties, the tag can alter the folding of the fusion protein and complicate crystallographic studies by preventing crystallisation. Furthermore, protein of high purity is required for crystallographic studies. A 3C protease tag cleavage site is present, but precipitation of the protein after tag cleavage is a possible risk.

A His$_6$-tagged protein comprising both the TPR and the oxygenase domain was generated to obtain a construct more amenable to sufficient purification and crystallographic studies than the GST-tagged protein was. The sequence encoding for AspH$_{315-758}$ was subcloned from the pGEX-6P-1 into the pET-28a(+) vector between the *BamH*I and *Not*I restriction sites (**Fig. 2-10**), and its incorporation into this vector was confirmed by DNA sequencing.

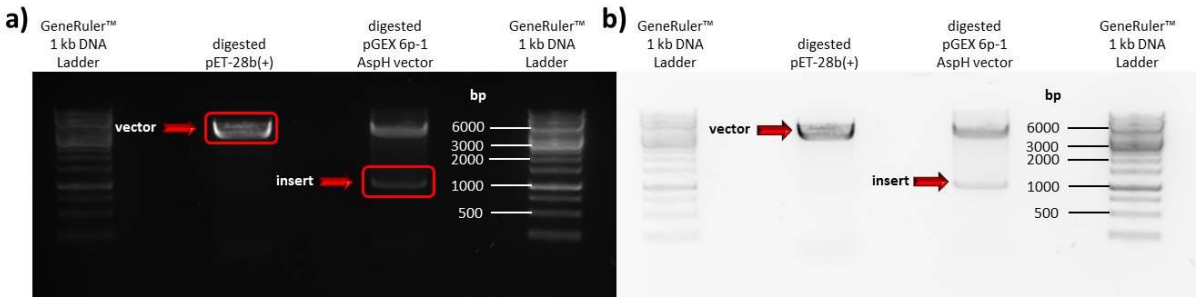


Fig. 2-10 | Agarose gel (1 %) stained with SYBR® Safe DNA Gel Stain. Images of the linearised DNA fragment products after double digestion of the plasmids with the *BamH*I and *Not*I restriction enzymes, representing subcloning of the sequence encoding for AspH$_{315-758}$ from the pGEX-6P-1 into the pET-28a(+) vector. **a)** Gel under UV light, **b)** inverted colours for clarity.

2.3.1 Protein Production and Activity of His_6-$AspH_{315-758}$ Construct

2.3.1.1 *In vitro* Transcription/Translation

His_6-$AspH_{315-758}$ protein production was investigated on an analytical scale by *in vitro* transcription/translation (IVTT).[410-412] IVTT has advantages over conventional expression in cells (e.g. *E. coli*), for example when an overexpressed product is toxic to the host cell, when the product is insoluble forms inclusion bodies, or when the protein is rapidly degraded by intracellular proteases.[411] The cell-free expression system is usually composed of extracts from rabbit reticulocytes, *E. coli*, or wheat germ, which contain all the macromolecular components required for translation of exogenous RNA, i.e. 70S or 80S ribosomes, tRNAs, aminoacyl-tRNA synthetases, inititation, elongation and termination factors.[413] The extracts are then supplemented with amino acids, energy sources such as ATP and GTP, cofactors such as Mg^{2+}, and energy regenerating systems to facilitate translation.[414] DNA can be used as a template, which is first transcribed into mRNA and then translated. Modified amino acids such as radioactive [^{35}S]-methionine can be incorporated in the protein, enabling analysis of the IVTT results by autoradiography.

Analytical IVTT to investigate His_6-$AspH_{315-758}$ production was conducted together with Dr Holger Gerlach (Hagan Bayley Group, Department of Chemistry, University of Oxford) using an *E. coli* S30 Extract System for Circular DNA IVTT kit (*Promega*) with incorporation of [^{35}S]-methionine. Following the IVTT reactions, an SDS-PAGE gel was run to analyse the protein production. Subsequently, the gel was dried and then autoradiographed (exposed to an X-ray film for 16 h) prior to film processing (**Fig. 2-11**).

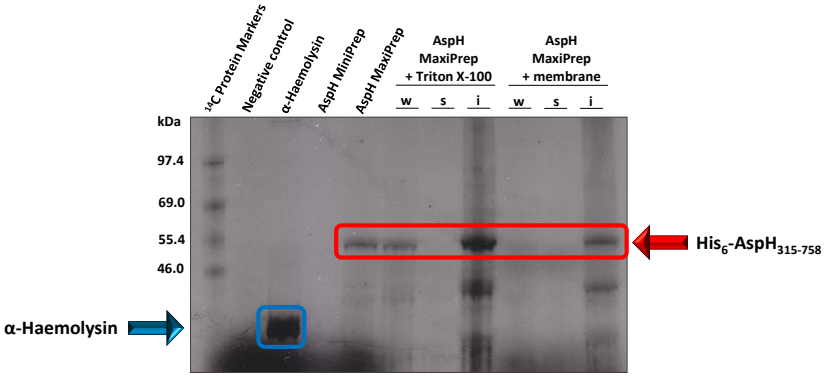


Fig. 2-11 | *In vitro* transcription/translation of His_6-$AspH_{315-758}$ (54.5 kDa, marked with the red arrow) using [^{35}S]-methionine. The reactions were analysed by SDS-PAGE, and the dried gel was then autoradiographed. The negative control contained no DNA; pure α-haemolysin (33.2 kDa, marked with the blue arrow) was used as a positive control. AspH plasmid DNA was purified using either *Qiagen* QIAprep Miniprep Kit or QIAprep Maxiprep Kit. Additionally, samples with added Triton X-100 (0.5 %) or rabbit red blood cell membrane were analysed as whole sample (w), soluble (s) and insoluble (i) fractions.

The results indicated that His_6-$AspH_{315-758}$ had been synthesised using the plasmid DNA template (**Fig. 2-11**). IVTT requires a very high purity of the plasmid DNA template. The protein was

produced from transfection-grade plasmid DNA isolated using a procedure with an *iso*propanol precipitation step (MaxiPrep, QIAfilter Plasmid Maxi Kit, *QIAGEN*). No protein band was observed in an IVTT sample for which plasmid DNA from a simpler isolation procedure had been used (MiniPrep, GeneJET Plasmid Miniprep Kit, *Fermentas*). Because full-length AspH is a transmembrane protein, and there was the chance that the $AspH_{315-758}$ construct was insoluble without the GST-tag, the effects of Triton X-100 as solubilising adjuvant and the addition of rabbit red blood cell membranes (provided by Dr Holger Gerlach) was assessed. Such membranes could assist assembly of a transmembrane protein or associated domains because of its affinity to the membranes. Samples of these mixtures were centrifuged, and the soluble and insoluble fractions were separately analysed on the gel. A band corresponding to His_6-$AspH_{315-758}$ was observed in the whole sample and in the resuspended pellet in the Triton X-100 supplemented sample and only found in the resuspended pellet in the case of the sample with added membranes. These results suggested that the His_6-$AspH_{315-758}$ may be insoluble without the solubilising GST-tag.

2.3.1.2 Expression Trials for the Production of Recombinant His_6-$AspH_{315-758}$

The His_6-$AspH_{315-758}$ pET-28a(+) plasmid DNA for was transformed into *E. coli* BL21 (DE3) cells, and small-scale construct expression trials were then conducted. The expression levels were first assessed at temperatures of 18 °C, 28 °C, and 37 °C, cell growth durations of 4 h and 16 h after induction when the OD_{600} reached 0.6-0.8 with 0.2 mM, 0.5 mM, or 1.0 mM IPTG respectively and analysed by SDS-PAGE (**Fig. 2-12**).

Bands corresponding to the soluble His_6-$AspH_{315-758}$ (54.5 kDa) were apparent but the band intensities, and hence protein production levels, were low. The highest levels of soluble protein were obtained after induction with 1.0 mM IPTG at an optical density measured at 600 nm (OD_{600}) of 0.6-0.8 with further growth for 16 h at 28 °C. Protein bands in the range of the expected molecular mass were also observed in the uninduced cells, which could be a sign of leaky expression. The sequence encoding for human $AspH_{315-758}$ had been cloned into a prokaryotic expression vector, and codon usage analysis of the sequence (not shown) indicated several rare codons, potentially complicating expression in *E. coli*. Nevertheless, the expression trials suggested that soluble protein could be produced. In further trials, expression was optimised (**Fig. 2-13**).

The yield of soluble recombinant His_6-$AspH_{315-758}$ was substantially increased by induction of the expression during late log phase, as also observed for other proteins.[415] A reduced temperature and a low IPTG concentration can contribute to the higher yield. Induction with 0.1 mM IPTG at a high cell density (optical density measured at 600 nm (OD_{600}) 1.1-1.3) with further growth for 16 h at 18 °C enabled optimal production of the His_6-$AspH_{315-758}$ from *E. coli* BL21 (DE3). In addition, these conditions are cost-effective due to much lower use of the costly IPTG.

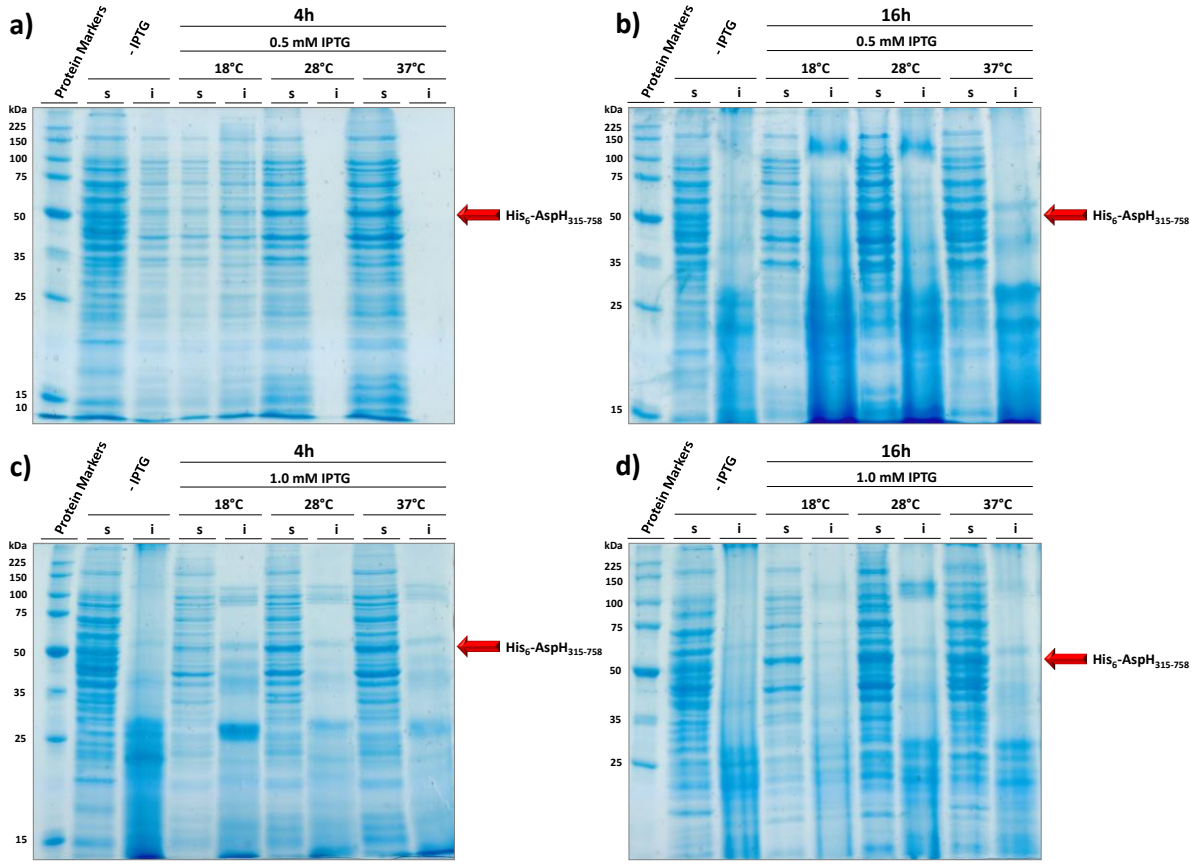


Fig. 2-12 | SDS-PAGE gel (12 %) analysis of protein production from expression trials for the His_6-$AspH_{315-758}$ construct cloned into pET-28a(+). Gels stained with Coomassie Blue. Shown are soluble (s) and insoluble (i) fractions for the different conditions as indicated (temperatures 18 °C, 28 °C, or 37 °C, induction of the expression at OD_{600} 0.6-0.8). **a)** Cell growth duration of 4h after induction with 0.5 mM IPTG; **b)** 16 h cell growth after induction with 0.5 mM IPTG; **c)** 4 h cell growth after induction with 1.0 mM IPTG; **d)** 16 h after induction with 1.0 mM IPTG. Additionally, samples of uninduced cells (-IPTG) grown at 37 °C for 16 h were analysed.

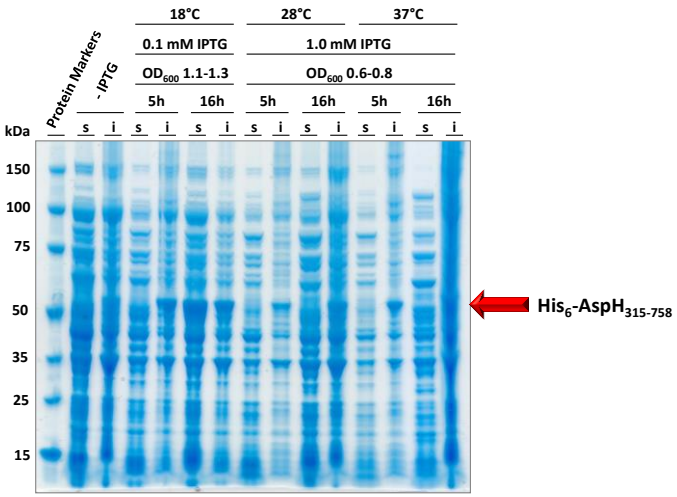


Fig. 2-13 | SDS-PAGE gel (4 %-12 %) analysis of protein production from expression trials for the His_6-$AspH_{315-758}$ construct cloned into pET-28a(+). Gel stained with Coomassie Blue. Shown are soluble (s) and insoluble (i) fractions for the different conditions as indicated (temperatures 18 °C, 28 °C, or 37 °C).

2.3.1.3 Small-Scale Trial Purification

Protein produced at 28 °C and 37 °C (1.2 l cultures of *E. coli* BL21 (DE3) grown at either 28 °C for 16 h or 37 °C for 5 h) was subjected to small-scale trial purifications (**Fig. 2-14**). The cleared lysate from each cell pellet was used for manual small-scale Ni^{2+}-affinity chromatography (1 ml HisTrap™ HP, *GE Healthcare*); the protein fractions were analysed by SDS-PAGE and with InVision™ His-Tag In-Gel Stain (*Invitrogen*). This stain allows for fast visualisation of His-tag fusion protein bands under UV light, which can be performed directly in the polyacrylamide gel following electrophoresis. The method is sensitive and highly specific (comparable to western blotting), and the same gel can be stained with Coomassie afterwards.

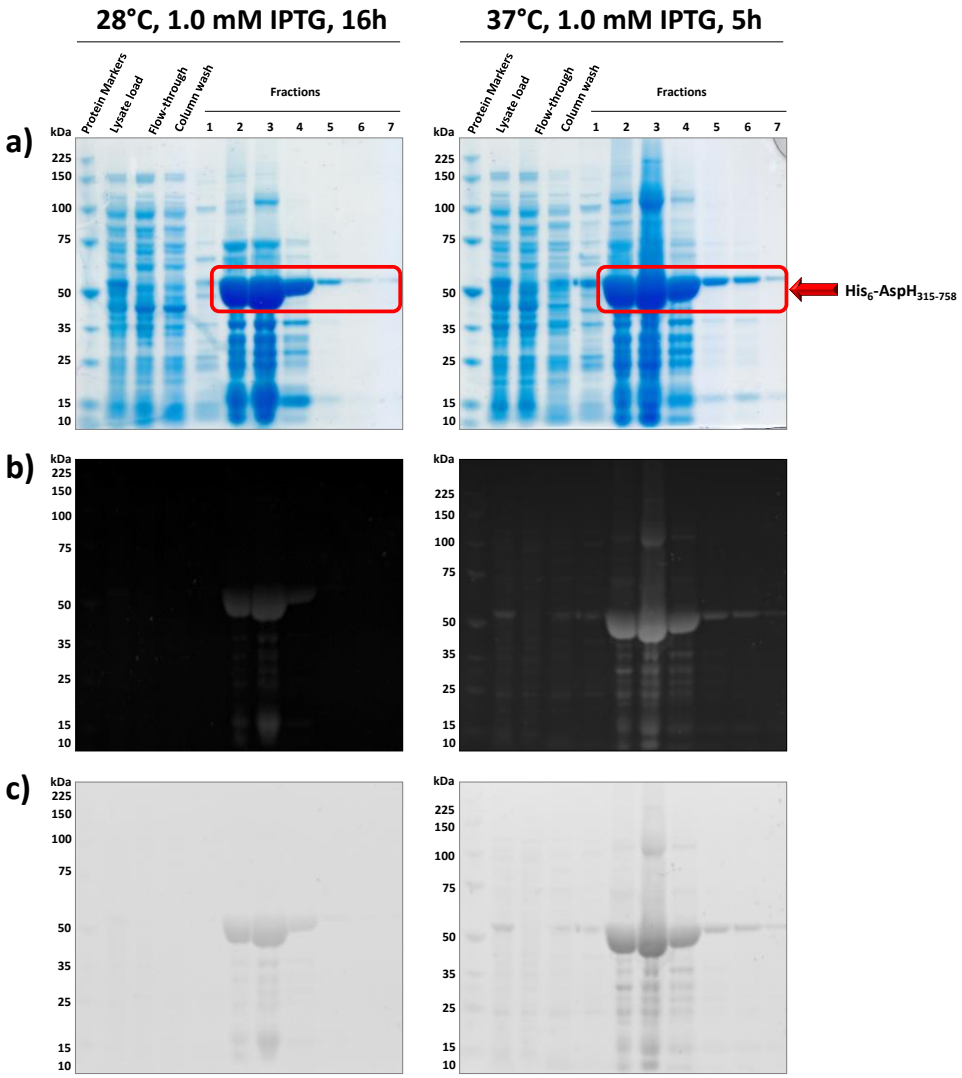


Fig. 2-14 | His_6-$AspH_{315-758}$ (54.5 kDa) small-scale affinity purification trials using HisTrap™ HP, column size 1 ml. **a)** 4 %-12 % SDS-PAGE gels of protein fractions from growth for 16 h after induction with 1.0 mM IPTG at 28 °C and 37 °C respectively. Gels stained with Coomassie Blue for total protein detection. **b)** Staining with InVision™ His-Tag In-Gel Stain (*Invitrogen*) for visualisation of His-tagged protein bands. Images taken under UV light (324 nm); **c)** colours inverted for clarification. Perfect Protein Markers™ (*Novagen*) consist of His-tagged proteins, and therefore also served as a positive control for the detection.

Protein at the expected mass was clearly enriched up in the trials (**Fig. 2-14**), at visually comparable levels in the samples from 28 °C and 37 °C. Staining with InVision™ His-Tag In-Gel Stain provided further evidence that these bands corresponded to His-tagged protein. The small-scale chromatography was performed as a single step a gradient of the elution buffer (elution directly with 100 % elution buffer, 500 mM imidazole, 50 mM HEPES, 500 mM NaCl).

2.3.1.4 Large-Scale Purification

The small-scale purification trials had shown that the His_6-$AspH_{315-758}$ was amenable to affinity purification. The production of His_6-$AspH_{315-758}$ was then scaled up, and the protein was purified using Ni^{2+}-affinity chromatography (5 ml HisTrap™ HP, *GE Healthcare*) with gradient elution as first step of the large-scale protein purification. Protein was obtained in high yield, but the purity after nickel affinity purification alone was insufficient for further studies (**Fig. 2-15a**). Therefore, size-exclusion chromatography (HiLoad™ 16/60 Superdex™ 75 120 ml, *GE Healthcare*) was subsequently performed to yield protein of high purity (**Fig. 2-15b**). The yields achieved were approximately 4.7 mg/litre culture grown at 28 °C for 16 h, 2.5 mg/litre culture grown at 37 °C for 5 h (at both temperatures induction with 1.0 mM IPTG at OD_{600} 0.6-0.8; increased yields of up to 8.3 mg/litre culture were obtained from growth at 18 °C for 16 h after induction with 0.1 mM IPTG at OD_{600} 1.1-1.3).

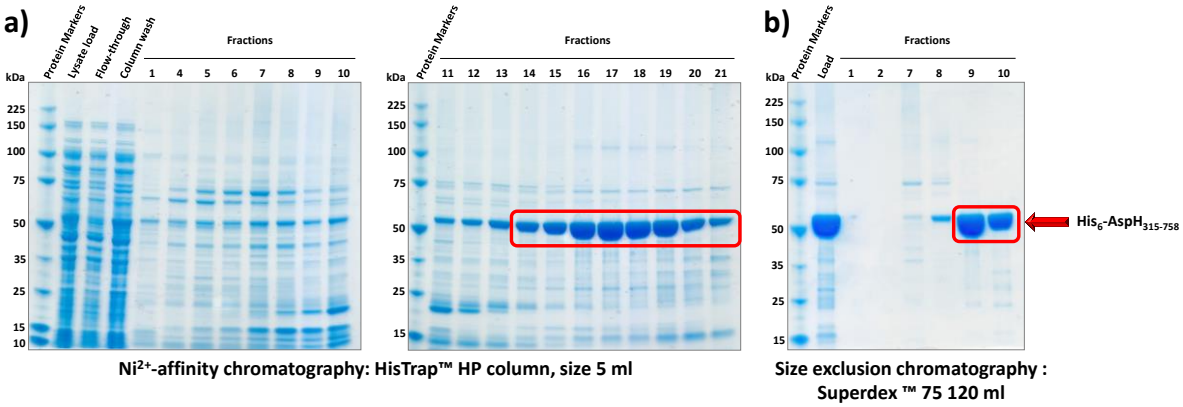


Fig. 2-15 | Protein production of His_6-$AspH_{315-758}$. SDS-PAGE gels (4 %-12 %) with Coomassie staining depicting a typical purification. **a)** After affinity purification using HisTrap™ HP, column size 5 ml; **b)** after size-exclusion purification using HiLoad™ 16/60 Superdex™ 75 120 ml. Red circled fractions were pooled and concentrated.

The achieved purity of His_6-$AspH_{315-758}$ was sufficient for further functional and crystallographic studies. The correct mass (54,519.1 Da) of the purified His_6-$AspH_{315-758}$ was confirmed using intact protein ESI-MS analysis (**Fig. 2-16**).

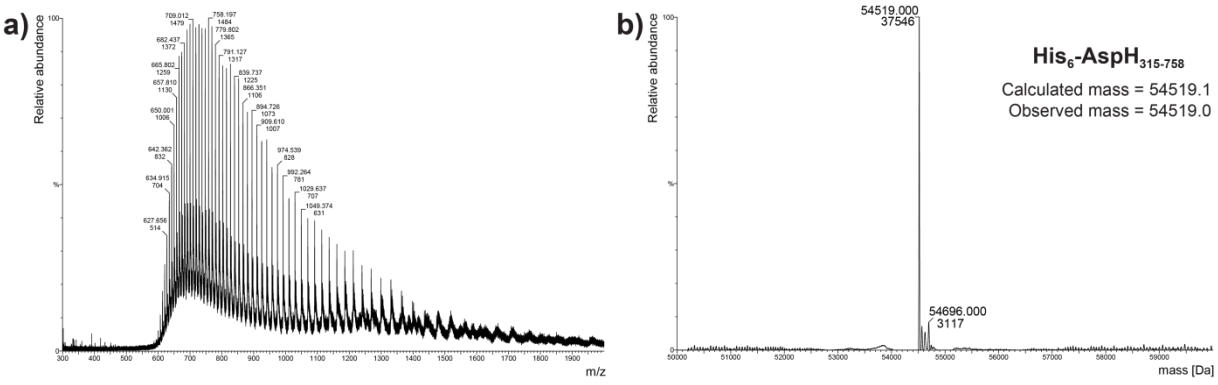


Fig. 2-16 | ESI-MS of purified His_6-$AspH_{315-758}$. **a)** Raw spectrum (charge state envelope); **b)** deconvoluted mass spectrum.

2.3.1.5 His_6-$AspH_{315-758}$ Hydroxylation Activity and 2OG Turnover Assays

The activity of His_6-$AspH_{315-758}$ was investigated using the established MALDI-ToF-MS hydroxylation activity and the radiochemical [1-^{14}C]-2OG turnover assays. His_6-$AspH_{315-758}$ produced from expression at all temperatures (18 °C, 28 °C, and 37 °C) showed indistinguishable activity in the assays. Approximately 40 % of the hFX $EGF1_{39mer}$ peptide was hydroxylated after 60 min incubation at 37 °C (**Fig. 2-17a**), the same level of hydroxylation that was observed with the GST-$AspH_{315-758}$. The results from the 2OG turnover assay indicate that His_6-$AspH_{315-758}$ has a higher activity than GST-$AspH_{315-758}$. His_6-$AspH_{315-758}$ catalysed high 2OG turnover, which was further stimulated in the presence of the peptide substrate (**Fig. 2-17b**).

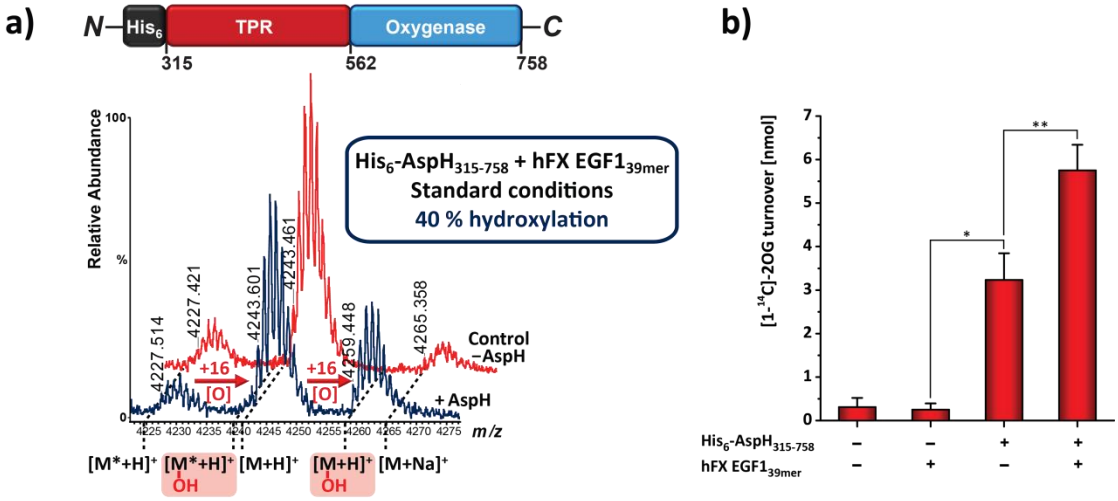


Fig. 2-17 | His_6-$AspH_{315-758}$ activity analysis. **a)** MALDI-ToF-MS spectra of the hydroxylation activity assay conducted for 60 min showing activity with the hFX $EGF1_{39mer}$ substrate (peak at m/z 4,259.6 for the monoisotopic mass of the hydroxylated peptide $[M(OH) + H]^+$). Standard conditions refer to the reaction buffer system (50 mM HEPES, pH 7.5) and presence of all cofactors/cosubstrates (Fe(II), Asc, 2OG, O_2). $[M^* + H]^+$ stands for a derivatised peptide species that also gets hydroxylated (further investigated in **Chapter 6**). **b)** Results from the [1-^{14}C]-2OG turnover assay conducted in the presence of all cofactors/cosubstrates (Fe(II), Asc, 2OG, O_2). His_6-$AspH_{315-758}$ catalysed high 2OG turnover, which was further stimulated in the presence of the hFX $EGF1_{39mer}$ peptide substrate (without the substrate: $p = 0.014$; in the presence of the substrate: $p = 0.007$). Data (background subtracted) represent the mean of triplicate experiments, with error bars showing SD ($n = 3$; mean ± SD; *$p < 0.05$; **$p < 0.01$).

2.3.2 Comparison of the Constructs

In summary, both two-domain constructs (GST-AspH$_{315-758}$ and His$_6$-AspH$_{315-758}$) comprising both the TPR and the oxygenase domain show activity with the hFX EGF1$_{39mer}$ peptide substrate. The oxygenase domain alone (His$_6$-AspH$_{562-758}$) is inactive in the presence of the hFX EGF1$_{39mer}$ substrate but able to catalyse uncoupled 2OG turnover. The His$_6$-AspH$_{315-758}$ construct appears to be best-suited for further structural and biochemical studies due to its high purity and activity (**Fig. 2-18**).

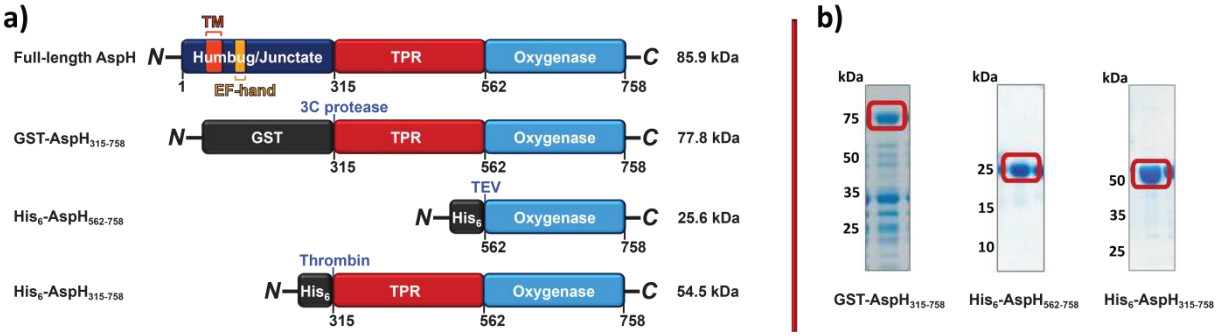


Fig. 2-18 | Overview of the AspH constructs. **a)** Schematics of the domain architectures. **b)** Details from SDS-PAGE gels after purification showing degradation of GST-AspH$_{315-758}$ running as four major bands, whereas His$_6$-AspH$_{562-758}$ and His$_6$-AspH$_{315-758}$ could be obtained in high purity.

2.3.2.1 Hydroxylation Activity

The reactions with the His$_6$-AspH$_{315-758}$ and GST-AspH$_{315-758}$ were monitored by time course assays (**Fig. 2-19**). The reaction catalysed by the His$_6$-tagged protein proceeded at a significantly faster rate than the reaction with the GST-tagged protein. In both cases, however, only approximately 40 % of the peptide was hydroxylated. The reaction appeared to stop at the 40 % hydroxylation level and was not further increased with a longer time incubation time or adding additional enzyme.

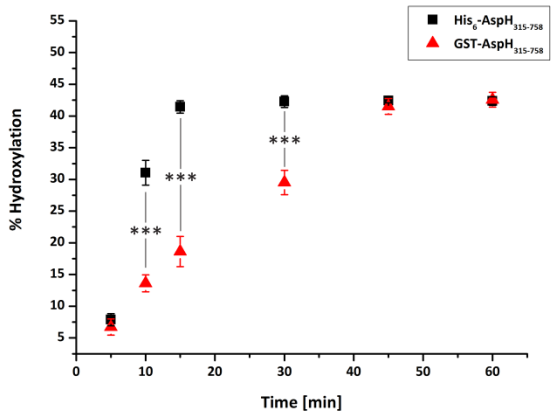


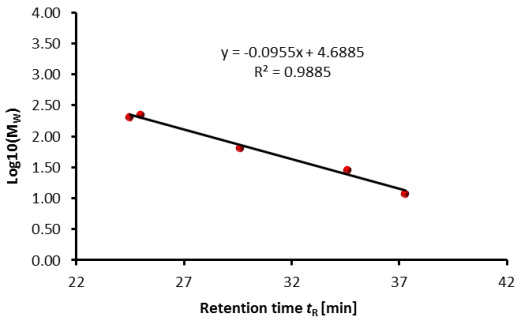
Fig. 2-19 | Time course MALDI-ToF-MS hydroxylation activity assay comparison of His$_6$-AspH$_{315-758}$ and GST-AspH$_{315-758}$. The His$_6$-tagged protein appears more active, as the reaction proceeds faster than with the GST-tagged protein (10 min: p = 0.0003; 15 min: p = 0.0005; 30 min: p = 0.0005). Only approximately 40 % of the hFX EGF1$_{39mer}$ peptide is hydroxylated by both proteins. Data represent the mean of duplicate experiments, with error bars showing SD (n = 2; mean ± SD; ***p < 0.001).

2.3.2.2 Analytical Gel Filtration

The GST-AspH$_{315-758}$, His$_6$-AspH$_{315-758}$, and His$_6$-AspH$_{562-758}$ were further analysed by analytical gel filtration, a technique that enables separation of molecules based on size and shape. A Superdex™ 200 column (30 ml, optimum separation range 10-600 kDa, *GE Healthcare*) with a high specification Sepharose resin was used for analysis of the molecular mass and the oligomerisation state of the different AspH proteins in solution. Elution of the protein(s) from the column is monitored by its absorbance at 280 nm (A_{280}). A calibration curve is constructed using protein standards, and the molecular weight can be determined from their retention time (t_R) on the column.

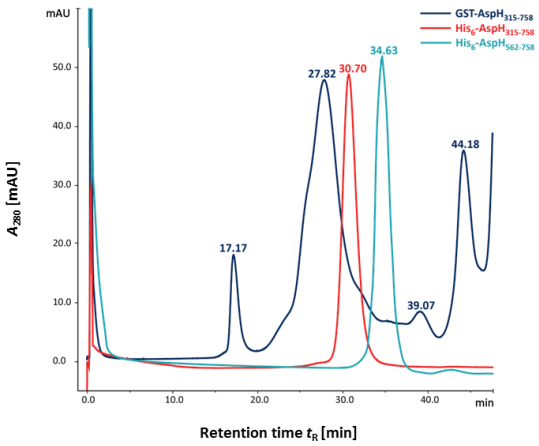
The His$_6$-AspH$_{315-758}$ and His$_6$-AspH$_{562-758}$ proteins both eluted as single peaks corresponding to the monomers (**Fig. 2-20**), indicated by comparison of the calculated molecular masses with the expected values. The GST-AspH$_{315-758}$ on the other hand eluted as four major peaks; none of which corresponded to the expected mass. This result, however, concurred with the observed four bands on SDS-PAGE gels.

a)



Protein standard	M_W [kDa]	$Log_{10}(M_W)$	t_R [min]
β-Amylase	223	2.35	25.0
Myosin	205	2.31	24.5
Albumin	66	1.82	29.6
Carbonic anhydrase	29	1.46	34.6
Cytochrome C	12	1.08	37.3

b)



AspH protein	Peak at t_R [min]	Calculated M_W [kDa]	Expected M_W [kDa]
His$_6$-AspH$_{315-758}$	30.7	57.1	54.5
His$_6$-AspH$_{562-758}$	34.63	24.1	25.6
GST-AspH$_{315-758}$	17.17 27.82 39.07 44.18	-* 107.6 (9.1)* -*	77.9

Fig. 2-20 | Analytical gel filtration of different AspH proteins using Superdex™ 200 (30 ml, *GE Healthcare*). **a)** Calibration curve of molecular masses (in kDa) vs. retention time t_R (in min) constructed from the given protein standards eluted from the column. **b)** Chromatogram showing His$_6$-AspH$_{315-758}$ and His$_6$-AspH$_{562-758}$ eluting as sharp single peaks corresponding to the monomers, whereas GST-AspH$_{315-758}$ eluted as four major peaks; none of which corresponded to the expected mass (*: no accurate determination possible due to the separation range of the Superdex™ 200: 10-600 kDa). The peak eluted after 17.17 min (within the void volume of the column) indicates that protein aggregates may have formed in the GST-AspH$_{315-758}$ sample.

Compared to the GST-AspH$_{315-758}$, the His$_6$-AspH$_{315-758}$ has several advantages, as it can be purified and characterised easily, shows higher activity, and is stable in storage buffers without added NaCl, which can be of advantage for crystallisation (in order to avoid potential formation of salt crystals). The more active His$_6$-AspH$_{315-758}$ was therefore used to optimise the assay conditions (described in **Chapter 3**). All experiments from here on were conducted using the His$_6$-AspH$_{315-758}$, unless stated otherwise.

2.4 AspH Crystallography

*The AspH$_{562-758}$ structure that is analysed in **Section 2.4.1** was produced by the researchers at the SGC Oxford. Dr Michael A. McDonough harvested crystals, collected and processed the data, solved and refined all AspH$_{315-758}$ structures.*

2OG oxygenases have emerged as an important family of human enzymes, and crystal structures have been determined for representatives of most subfamilies.[62] However, at the start of the crystallographic studies in this Thesis no structural characterisation of AspH had been reported.

The AspH activity assay results suggested that the TPR domain has a role in substrate recognition. To investigate the observations, crystallographic studies have been initiated to gain a better understanding of how the TPR domain is involved.

2.4.1 Oxygenase Domain Structure Analysis: AspH$_{562-758}$

In collaboration with Prof Udo Oppermann's group at the SGC Oxford, the first AspH X-ray crystal structure of the oxygenase domain construct, His$_6$-AspH$_{562-758}$, was determined. Dr Grazyna Kochan produced the His$_6$-AspH$_{562-758}$ with selenomethionine (SeMet) incorporated in place of methionine residues.[416] The incorporation of Se atoms can enable one to overcome the phase problem.[417] The three-dimensional structure of the protein can then be elucidated using single-wavelength anomalous diffraction (SAD).[418-420]

The His$_6$-AspH$_{562-758}$ construct in complex with Ni(II) replacing the endogenous active site Fe(II) and L-malate (L-Mal, **Fig. 2-21**) was crystallised by Dr Grazyna Kochan (SGC Oxford). Crystals were grown by vapour diffusion in 150 nl sitting drops at 20°C. The protein (9.5 mg/ml) was mixed with $NiSO_4$ (10 mM final concentration) and the 2OG cosubstrate analogue *N*-oxalylglycine (NOG, 1 mM final concentration) prior to crystallisation to obtain a catalytically inert complex. The precipitant contained 50 mM sodium malate, 20 % PEG3350, and 30 % dextran sulfate (sodium salt). The drops were prepared by mixing 100 nl of the protein mixture with 50 nl of the precipitant (protein-to-well ratio 1:2). Crystals were harvested and cryo-cooled in liquid nitrogen with 25 % glycerol as cryoprotectant. Data of 2.05 Å resolution (**Table 2-1**) were collected at Diamond Light

Source synchrotron MX beamline I04 and were trigonal symmetry (space group $P3_121$). The protein X-ray crystal structure was solved by Dr Tobias Krojer (SGC Oxford). Although the protein was crystallised in presence of NOG to replace the 2OG cosubstrate, the ligand observed bound in the active site turned out to be L-malate. This ligand was present in the precipitant in large excess over NOG (50 mM vs. 1 mM). As a TCA cycle intermediate, L-malate is a 2OG cosubstrate mimic, and the structure shows that it binds to the protein in the 2OG-binding pocket.

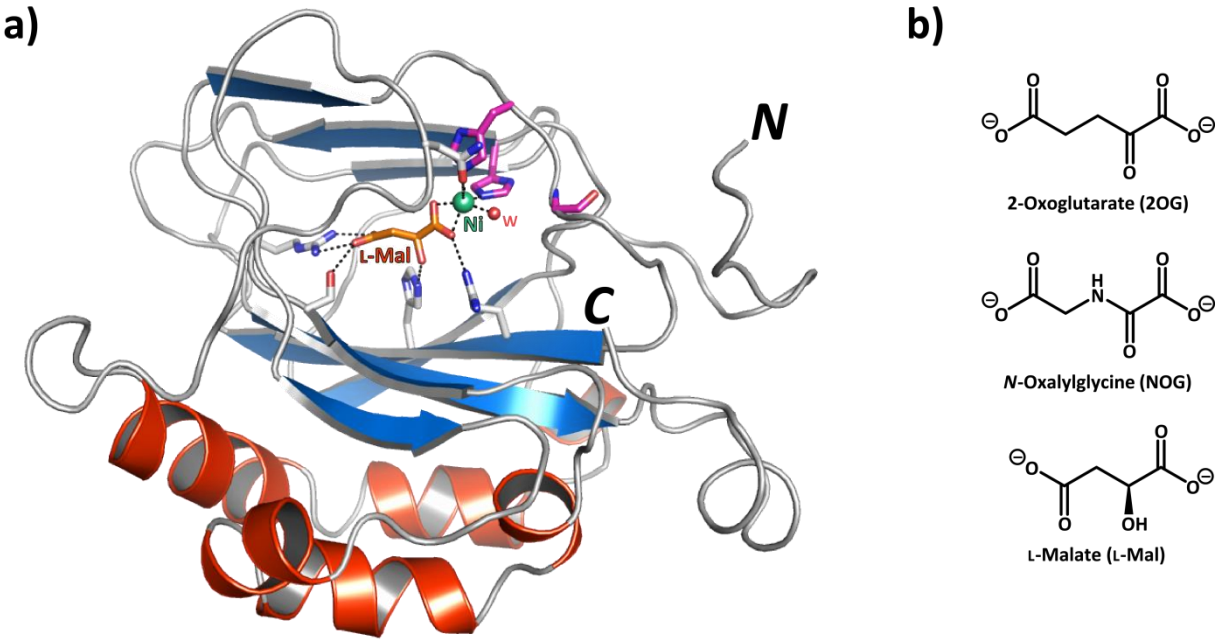


Fig. 2-21 | View from a crystal structure of human His_6-$AspH_{562-758}$ in complex with with Ni(II) and L-malate replacing Fe(II) and 2OG (resolution: 2.05 Å; structure determined by the SGC Oxford). **a)** Overview of the crystal structure; **b)** skeletal formulas of 2OG, its analogue *N*-oxalylglycine (NOG) and the mimic L-malate (L-Mal). W: water molecule.

Table 2-1 | Crystallisation conditions and data collection for the $AspH_{562-758}$ SAD (SeMet) structure obtained by the SGC Oxford (refinement data still in progress).

	$AspH_{562-758}$·Ni^{II}·L-Malate
Data collection	
Space group	$P3_121$
Cell dimensions	
a, *b*, *c* [Å]	133.22, 133.22, 44.64
α, β, γ [°]	90.0, 90.0, 120.0
Resolution [Å]	19.87-2.05 (2.16-2.05)*
R_{sym} or R_{merge}	0.089 (0.819)*
I / σI	11.5 (2.3)*
Completeness [%]	99.9 (100.0)*
Redundancy	7.4 (7.5)*
Symmetry	Trigonal
Method	Vapour diffusion, sitting drop (150 nl), protein-to-well ratio 1:2
Temp. [K]	293
Precip. cond.	9.5 mg/ml $AspH_{562-758}$, 20 mM $NiSO_4$, 1 mM NOG, 50 mM sodium malate, 20 % PEG3350, 30 % dextran sulfate
Cryoprotectant	25 % glycerol
Diffraction Source	Synchrotron
Type	Diamond Beamline I04
Wavelength List [Å]	0.9763
Site	Diamond Light Source

*Values in parentheses are for highest-resolution shell.

The $AspH_{562-758}$ structure reveals the core DSBH fold, as present in other 2OG oxygenases, with the active site located at one end of the DSBH. Notably, the *N*- and *C*-termini of the protein are situated close to the active site. However, in contrast to other 2OG-dependent hydroxylases, the active site metal is ligated by the side chains of only two residues donated by the protein (His679, His725) located on strands II and VII of the DSBH (**Fig. 2-22**). The L-malate C-1-carboxylate coordinates the metal in bidentate manner. In addition, a water molecule and the side chain of Asn618 coordinate the metal. Hydrophobic residues including Trp625, Met670, Ile7102, Trp711, Val727, Ile737, and Ile739 line the 2OG-binding pocket. In addition, several polar and charged residues of the protein form hydrogen bonds and electrostatic interactions. The C-4-carboxylate of L-malate is positioned to form a salt bridge with the side chain of Arg735 (a basic residue is always present in this region in 2OG oxygenases, see **Chapter 1**) and a hydrogen bond to the hydroxyl group of Ser668. The side chains of His690 and Arg688 also form hydrogen bonds to the L-malate hydroxyl group and the C-1-carboxylate. The AspH catalytic domain structure shows that this domain alone is able to bind divalent metal and 2OG-like compounds, consistent with the observed 2OG turnover.

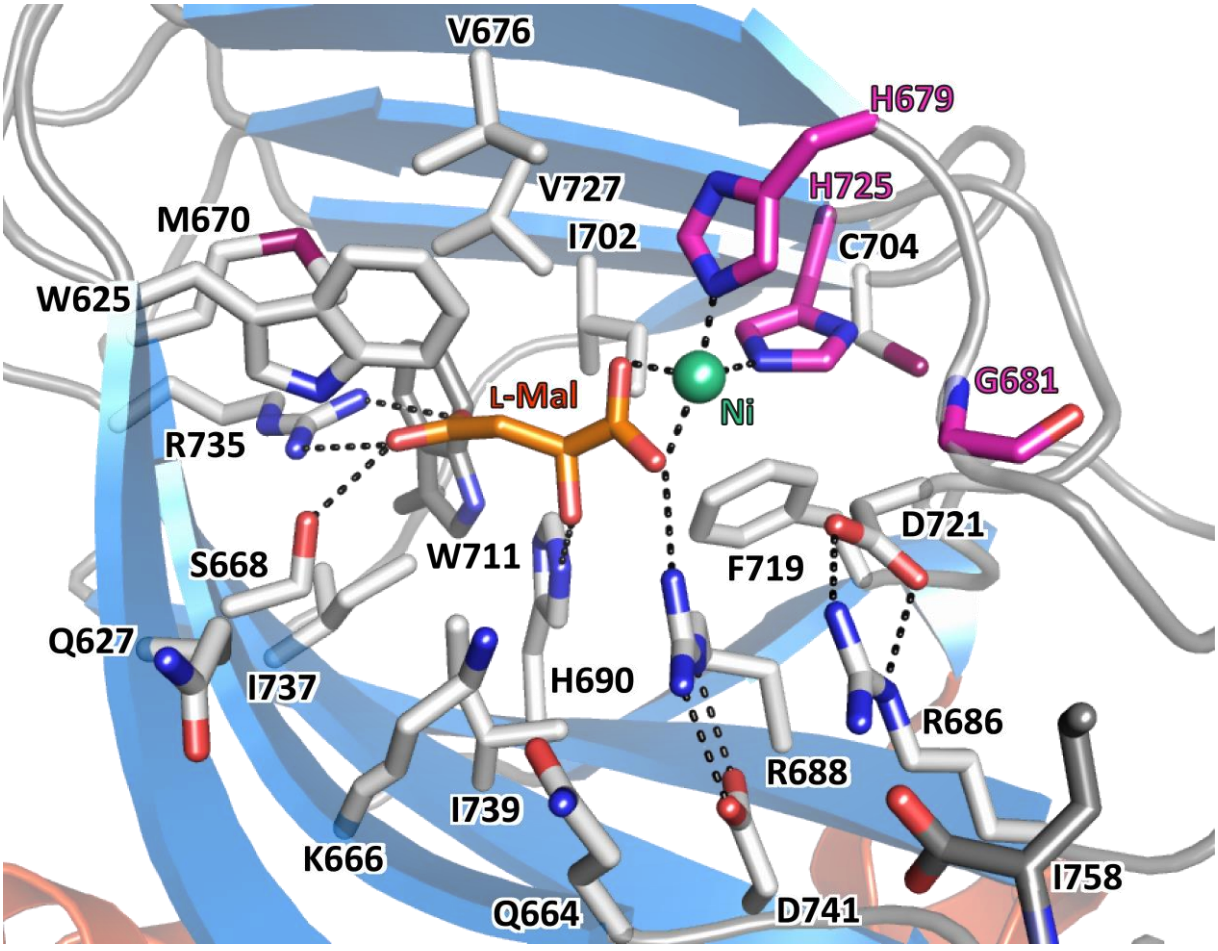


Fig. 2-22 | Close-up view of the $AspH_{562-758}$·Ni^{II}·L-Malate crystal structure active site. The HXG...H catalytic dyad is shown in magenta. The water molecule and loop residue N618 additionally coordinating the active site metal are omitted for clarity.

The structure analysis showed that a loop comprising residues 616-619 was positioned over the active site in a way that would appear to block primary substrate binding (**Fig. 2-23**), "closing" over it. Several residues that could act as potential Fe(II)-binding surrogates surround the active site. The side chain of Asn618 appears to interact directly with the active site metal. Glu617 sticks in the active site potentially mimicking the Asp/Asn substrate, and its side chain could potentially also coordinate the metal. Although AspH lacks the carboxylate of the highly conserved catalytic triad, an aspartic acid residue (Asp721) is located close to the active site metal. Asp721 forms a salt bridge with Arg686; however, the proximity of the Asp721 side chain carboxylate to the active site centre could enable coordination of the metal upon a slight conformational change.

The $AspH_{562\text{-}758}$·Ni^{II}·L-malate structure obtained by the SGC reveals interesting details about the AspH active site geometry. However, analysis of the structure suggested that the catalytic domain is in an inactive state with its active site blocked, which could explain the lack of hydroxylation activity in assays with the hFX $EGF1_{39mer}$ *in vitro* peptide substrate. To further investigate the role of the TPR domain in substrate recognition, crystallographic studies with the longer His_6-$AspH_{315\text{-}758}$ construct were pursued.

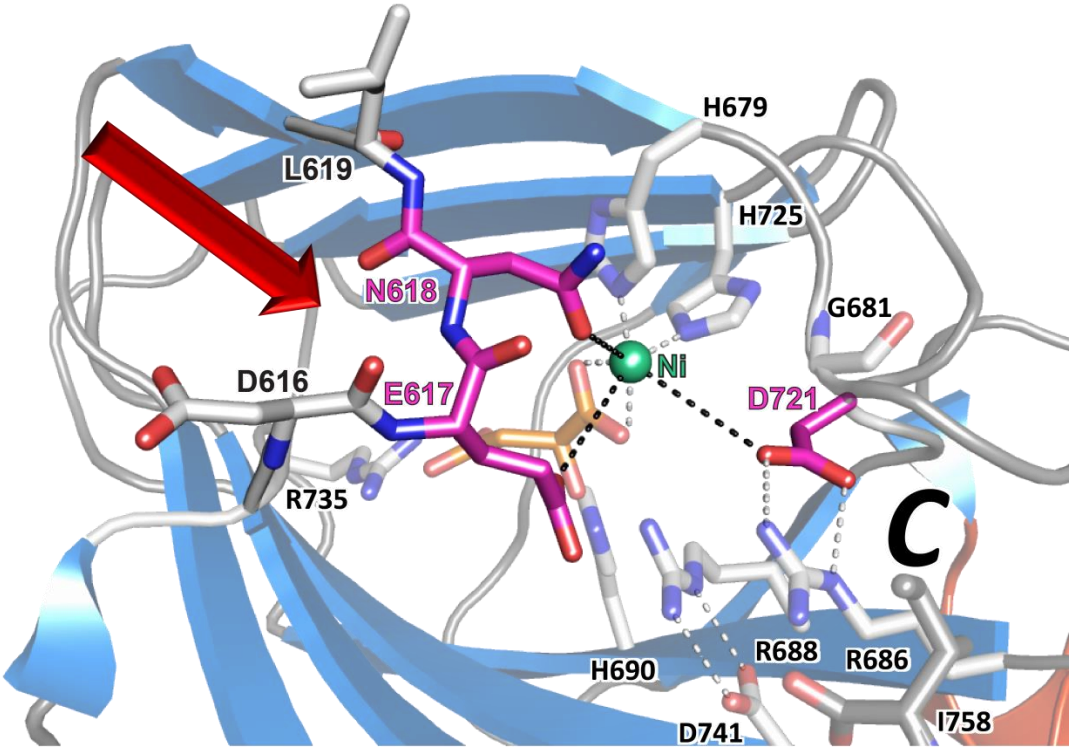


Fig. 2-23 | $AspH_{562\text{-}758}$·Ni^{II}·L-malate crystal structure close-up view, showing the loop residues that obstruct the active site (residues 616-619; indicated with a red arrow) and potential Fe(II)-binding surrogates close to the active site.

2.4.2 Crystallographic Studies on $AspH_{315\text{-}758}$

Since His_6-$AspH_{315\text{-}758}$ had never been crystallised before, various commercially available crystallisation conditions were screened using high-throughput methods (**Fig. 2-24**). Initial screens

focused on crystallisation of the protein without substrate ("apo"). For the trials, His_6-$AspH_{315-758}$ (18 mg/ml) was mixed with 1 mM $MnCl_2 \cdot 4H_2O$ and 2 mM NOG, substituting for Fe(II) and 2OG respectively, to form a stable catalytically inactive complex. Vapour diffusion crystallisation trials (96-well plates) were set up with sitting drops of 100-200 nl protein and various commercially available sparse matrix crystallisation screens (JCSG, PACT, INDEX, PEG/Ion, Structure, Salt; *Hampton Research* Molecular Dimensions) using Phoenix RE crystallisation robot (*Rigaku*). The plates were then incubated at two temperatures, 20 °C or 4 °C.

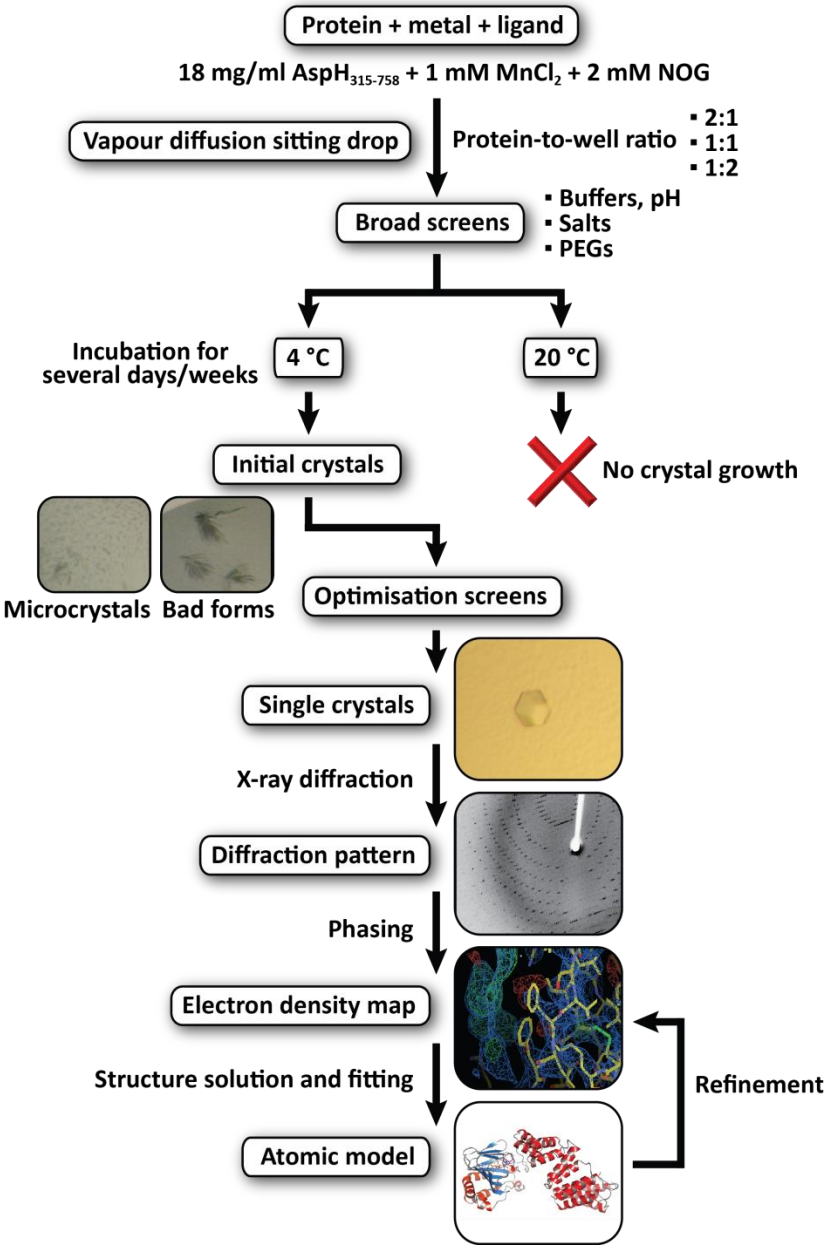


Fig. 2-24 | Flowchart of $AspH_{315-758}$ crystallisation, optimisation, and crystal structure determination.

While no crystals were obtained from the plates incubated at 20 °C, crystals formed in various different well conditions within a few days in plates incubated at 4 °C. However, mostly poor crystal forms such as 'haystack' or small clusters were obtained. Crystal growth under selected well

conditions was optimised (variation of the salt and polyethylene glycol (PEG) concentrations) in order to obtain larger single crystals with rhombohedral morphology (**Fig. 2-25**). Crystals were harvested using nylon loops and cryo-cooled using 25 % glycerol added to the well solution as cryoprotectant (performed by Dr Michael A. McDonough). The best crystals diffracted to a resolution of 2.0 Å, and the diffraction pattern confirmed that these were protein and not salt crystals. $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG$ crystals were found to have orthorhombic symmetry, space group is $P2_12_12_1$.

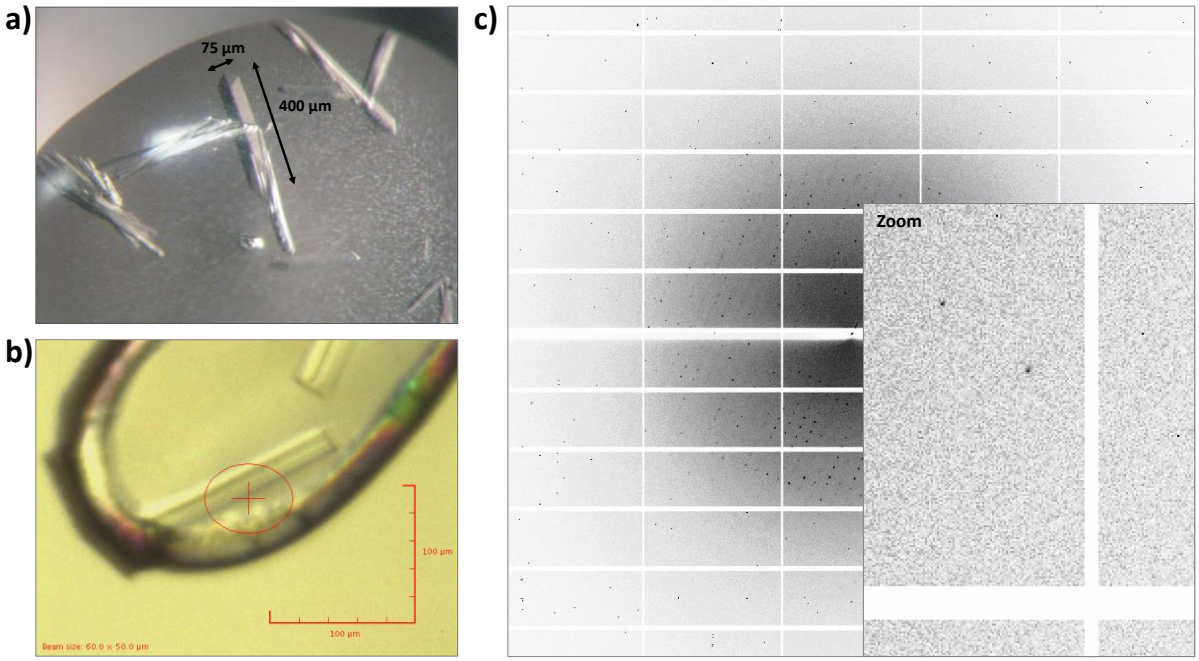


Fig. 2-25 | $AspH_{315\text{-}758}$ crystal growth and diffraction. **a)** Orthorhombic crystals of $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG$, orthorhombic space group $P2_12_12_1$, crystallised at 4 °C in a sitting drop from an optimisation screen with precipitant containing 100 mM bicine, pH 9.0, 100 mM sodium chloride, 27 % v/v PEG 550 monomethyl ether. Long axis of the large crystal = 400 µm. **b)** Crystals harvested using a nylon loop. **c)** A typical diffraction pattern of the protein crystals.

2.4.2.1 Crystal Structure Determination

The $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG$ ("apo") crystal structure was determined by Dr Michael A. McDonough. Data were indexed, integrated and scaled, and the structure of $AspH_{315\text{-}758}$ including both oxygenase and TPR domains was solved by the molecular replacement (MR) method[421] (data collection, phasing, and refinement statistics for the $AspH_{315\text{-}758}$ structures determined by molecular replacement are given in **Table 2-2** below). The coordinates of the AspH oxygenase domain (SGC data, current PDB entry: 3RCQ) were first used to successfully identify rotation and translation solutions for the oxygenase domain. Rotation and translation solutions for the TPR domain were determined using the *C*-terminal TPR domain of peroxisomal targeting signal 1 receptor (PEX5) as described in detail in **Chapter 10**. Iterative rounds of model building and fitting in COOT[422] and refinement in PHENIX[423] were performed until the decreasing R and R_{free} no longer converged and the final model was obtained.

2.4.2.2 Structure Analysis "Apo"-AspH$_{315-758}$

The overall conformation of the oxygenase domain in the AspH$_{315-758}$ crystal structure (**Fig. 2-26**), including metal and 2OG binding sites, is very similar to that of the AspH$_{562-758}$ structure with the exception of the loop obstructing the active site (residues 616-619). The active site is harboured in the barrel-like core of the DSBH composed of eight anti-parallel strands (comprising β2, β9, β4, β7 building the major sheet, and β3, β8, β5, β6 building the minor sheet). The DSBH core is flanked by eight helices in the "apo" structure (α13-α19, including α15a that is not observed in all structures). The TPR domain consists of six repeats comprising twelve helices, each connected as helix-loop-helix motifs (α1-α12) and is positioned adjacent to the active site opening of the oxygenase domain.

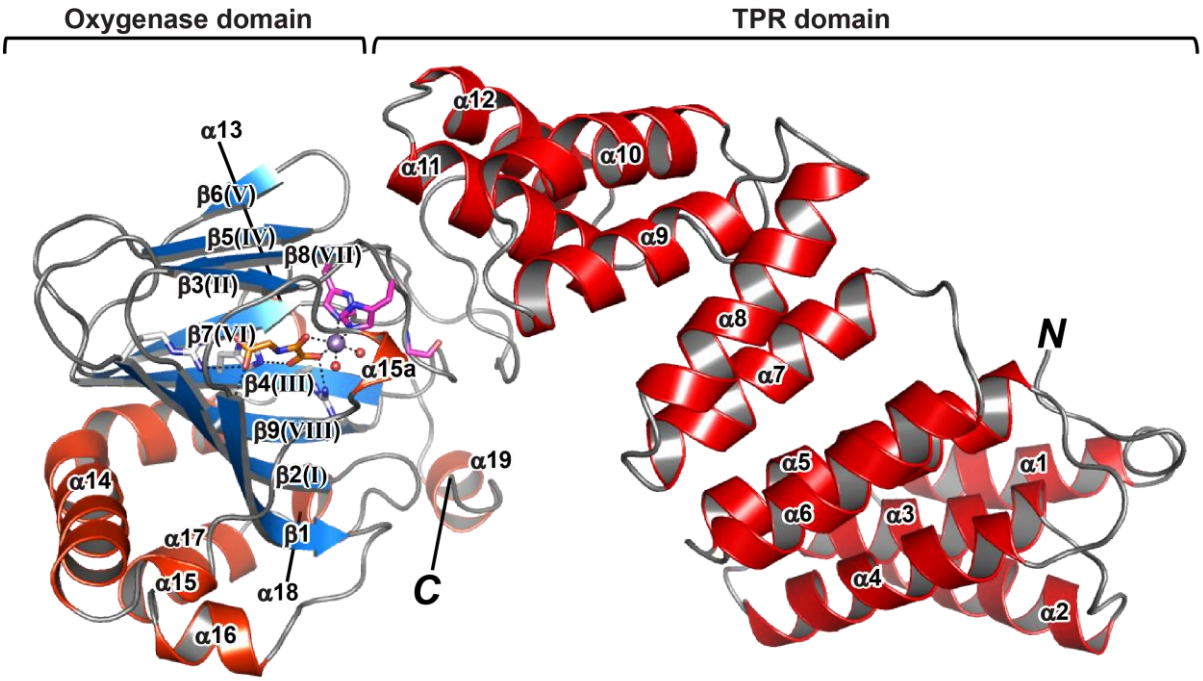


Fig. 2-26 | View from a crystal structure of human His$_6$-AspH$_{315-758}$ in complex with with Mn(II) and NOG substituting for Fe(II) and 2OG (resolution: 2.14 Å).

The active site metal is octahedrally coordinated by the side chains of the catalytic dyad histidines, His679 and His725, the C-1-carboxylate of NOG and two water molecules (**Fig. 2-27**). The C-5-carboxylate of NOG forms a salt bridge with the Arg735 side chain and a hydrogen bond to the Ser668 side chain in the cosubstrate pocket, similar to the L-malate C-4-carboxylate in the AspH$_{562-758}$·Ni^{II}·L-malate structure. His690 and Arg688 bind the NOG C-1-carboxylate. Arg688 additionally forms a salt bridge with Asp741, stabilising its position. The AspH *C*-terminus (Ile758) is positioned in close proximity to the active site.

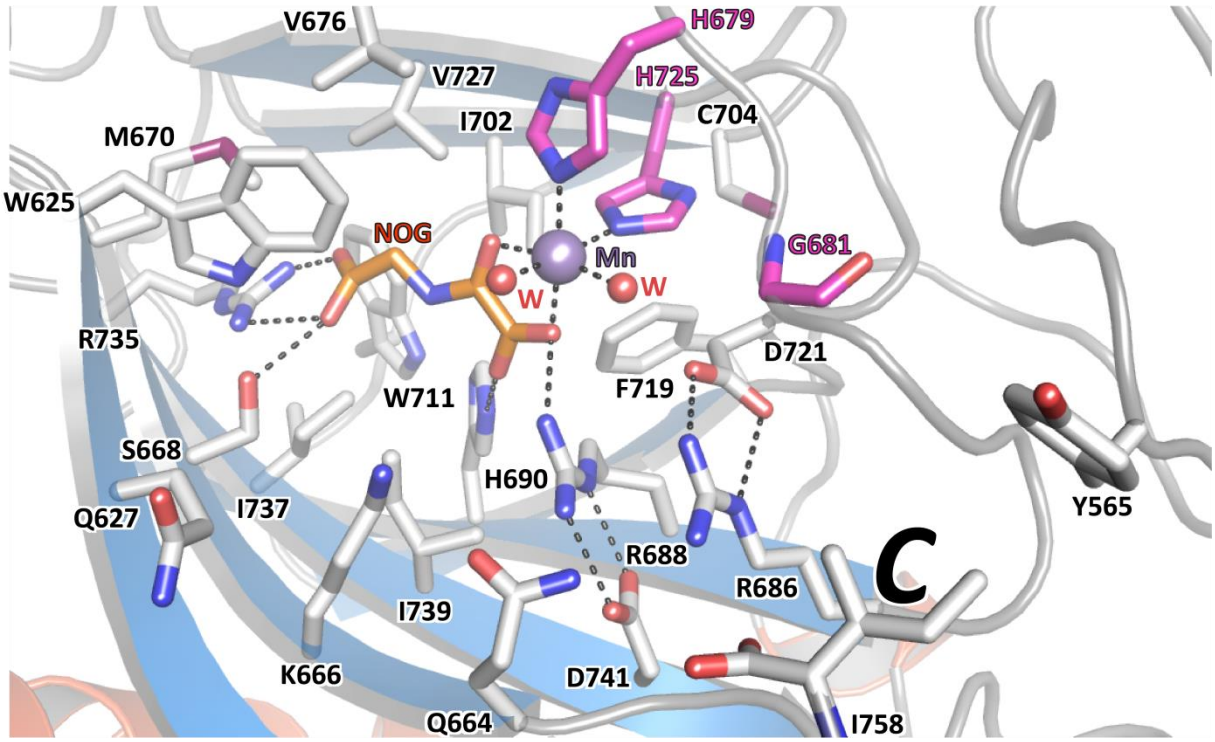


Fig. 2-27 | Close-up view of the $AspH_{315-758}$·Mn^{II}·NOG crystal structure active site. The HXG...H catalytic dyad is shown in magenta. The active site metal is octahedrally coordinated by the side chains of the catalytic dyad histidines, the C-1-carboxylate of NOG and two water molecules.

Interestingly, the loop comprising residues 616-619 that obstructs the active site in the $AspH_{562-758}$ structure is observed in a different conformation in the $AspH_{315-758}$ "apo" structure (**Fig. 2-28**; superimposition in **Fig. 2-36b**). In contrast to the $AspH_{562-758}$ structure, the loop residues Glu617 and Asn618 are not orientated towards the active site but are directed away from the active site metal, indicating a degree of flexibility in this loop region.

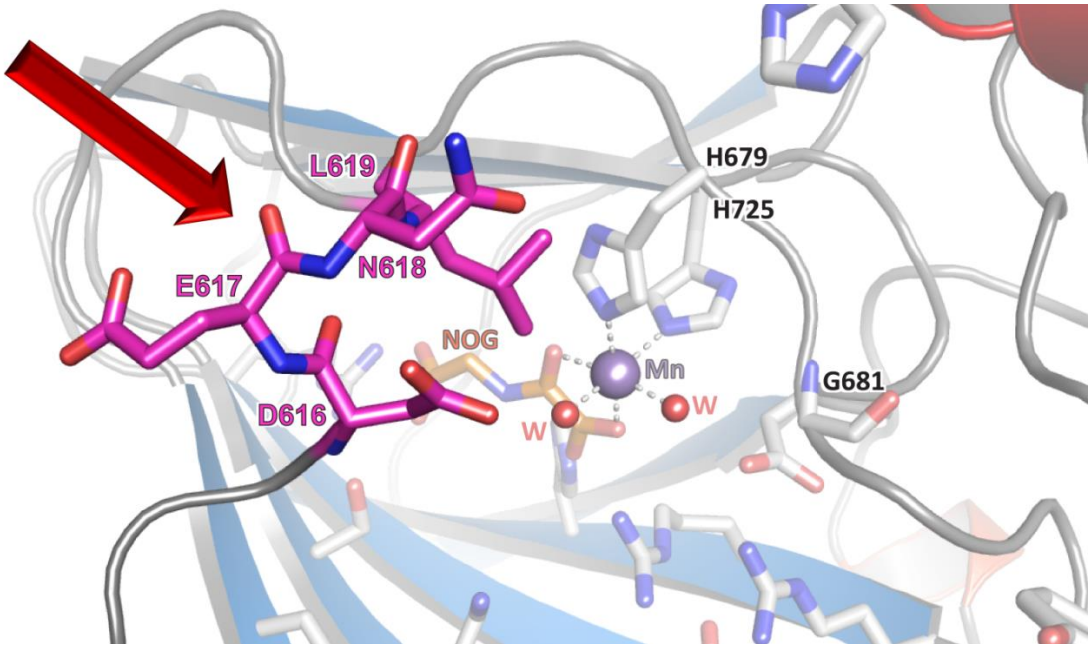


Fig. 2-28 | Close-up view of the $AspH_{315-758}$·Mn^{II}·NOG crystal structure showing the flexible loop region (magenta). The side chains of loop residues Glu617 and Asn618 stick out of the active site (indicated with the red arrow).

Analysis of the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG$ structure suggested that the "open space" observed between the catalytic domain's active site and the TPR domain could allow binding of a 40-residue EGFD substrate with TPR domain involvement. To test this hypothesis, co-crystallisation experiments of the His_6-$AspH_{315-758}$ construct with the hFX $EGF1_{39mer}$ substrate were pursued.

2.4.2.3 Protein-Peptide Interaction Pre-Crystallisation Trial

For successful crystallisation, a homogenous protein-peptide complex is normally preferred, and unbound peptide in the wells could hinder the crystallisation process. Since the hFX $EGF1_{39mer}$ peptide material is expensive, and relatively large amounts are required for crystallisation trials, $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}$hFX $EGF1_{39mer}$ complex formation was assessed on a small scale prior to the co-crystallisation trials (**Fig. 2-29**).

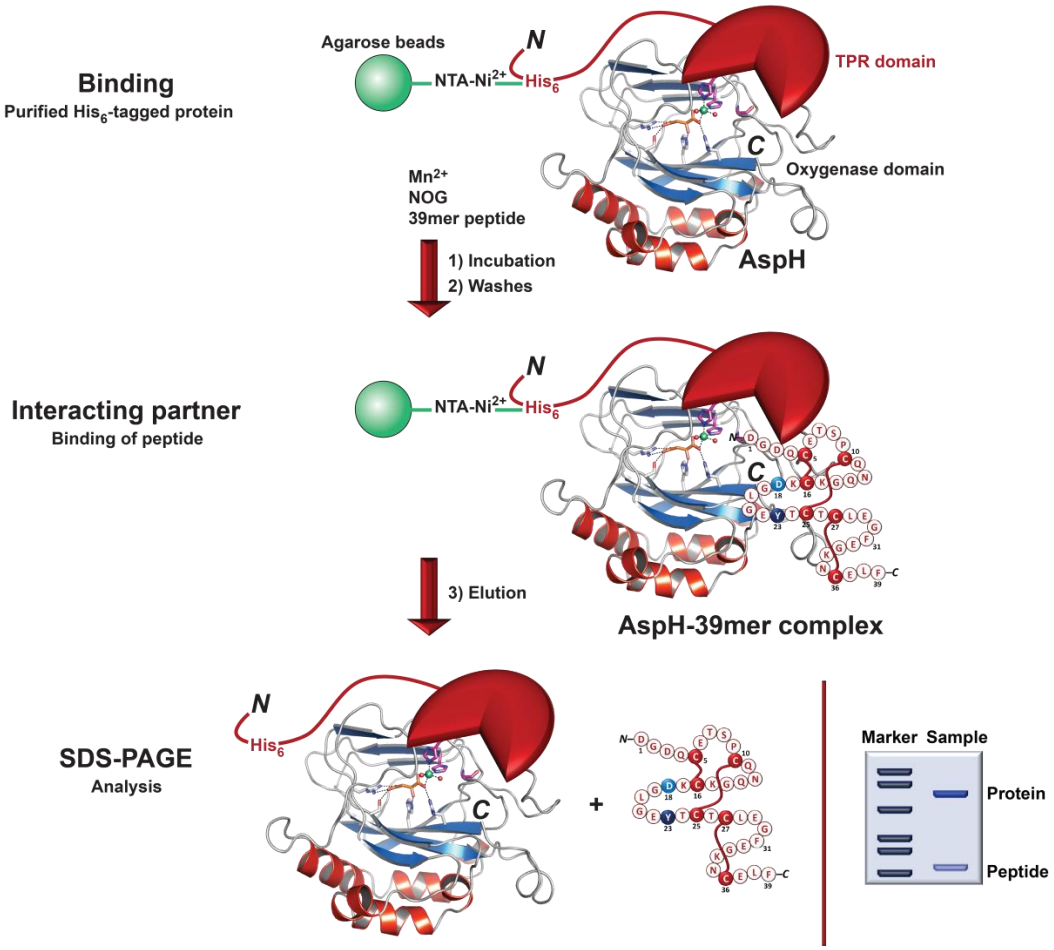


Fig. 2-29 | Schematic protein-peptide interaction study by binding of purified His_6-tagged AspH protein to Ni^{2+}-NTA agarose beads, followed by incubation with the hFX $EGF1_{39mer}$ *in vitro* peptide substrate in presence of $MnCl_2{\cdot}4\,H_2O$ and NOG, and elution with subsequent SDS-PAGE analysis. (AspH oxygenase domain figure made from the SGC data (current PDB entry: 3RCQ); the relative orientation of the oxygenase and the TPRs domain may change upon substrate binding, so the TPR domain is therefore shown as schematic).

Purified His_6-$AspH_{315-758}$ (54.5 kDa) was incubated at 4 °C for 16 h with the hFX $EGF1_{39mer}$ peptide substrate (4.2 kDa, AspH-to-substrate ratio 1:2.2) in presence of $MnCl_2 \cdot 4\ H_2O$, NOG and immobilised on Ni^{2+}-nitrilotriacetic acid (NTA) agarose beads. The $AspH_{315-758} \cdot Mn^{II} \cdot NOG \cdot$ hFX $EGF1_{39mer}$ complex (58.7 kDa) and can then be isolated from the mixture. Peptide bound to the enzyme elutes together with the protein and can be analysed by SDS-PAGE, which should result in two bands on the gel. Unbound peptide would be present in the flow-through.

However, due to the small size of the peptide (4.2 kDa), no band for the peptide was detected in conventional Tris-glycine SDS-PAGE analysis, as the peptide likely ran off the gel (**Fig. 2-30a**, on a 10 % NuPAGE® Novex precast Bis-Tris gel (*Invitrogen*) run in 2-(*N*-morpholino)-ethanesulfonic acid (MES) buffer). The same samples were then analysed by SDS-PAGE using a Tris-tricine gel system.[424] Tris-tricine SDS-PAGE offers high resolution especially for the small protein and peptide range[425] in contrast to the high-acrylamide Laemmli[426] gels, which have a stacking limit that is too high for peptides.[424,425] The Tris-tricine SDS-PAGE analysis showed a band smear in the molecular mass range of the peptide in the first eluate fraction with the protein (**Fig. 2-30b**), which was not found in the control load of purified His_6-$AspH_{315-758}$ alone. Although the peptide band was a smear rather than a distinct band, the results from this protein-peptide interaction experiment were encouraging to proceed with co-crystallisation trials.

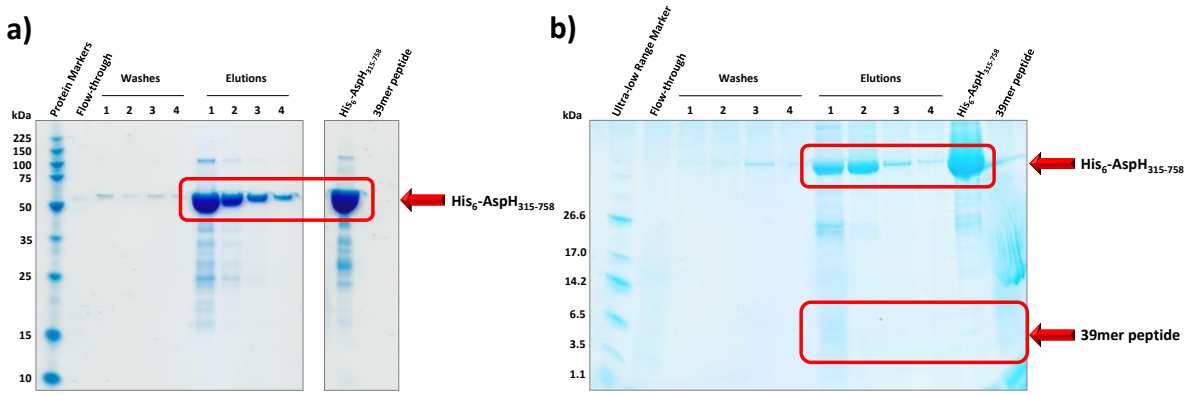


Fig. 2-30 | SDS-PAGE gel analysis of the protein-peptide interaction trial samples. Coomassie staining. **a)** Samples analysed on a 10 % NuPAGE® Novex Bis-Tris Tris-glycine SDS-PAGE gel (run with MES buffer). **b)** The same samples analysed on a 10-20 % Mini-PROTEAN® Tris-tricine gel (Bio-RAD). A band smear in the molecular mass range of the hFX $EGF1_{39mer}$ peptide observed in the first eluate fraction with the protein was not found in the control load of purified His_6-$AspH_{315-758}$ alone, showing that a stable protein-peptide complex was isolated in the interaction trial.

2.4.3 Crystal Structure of $AspH_{315-758}$ in Complex with a Substrate

Co-crystallisation experiments of His_6-$AspH_{315-758}$ (18 mg/ml) with 1 mM $MnCl_2 \cdot 4\ H_2O$, 2 mM NOG and the synthetic hFX $EGF1_{39mer}$ peptide were then pursued. An enzyme to substrate ratio of 1:2.2 was used for the trials to ensure that the substrate is present in slight excess but not so much as to use up disproportionate amounts of the expensive material. The co-crystallisation conditions for the

AspH- hFX $EGF1_{39mer}$ substrate complex could be completely different from the conditions for $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG$. Therefore, trials with a number of commercially available sparse matrix crystallisation screens (Structure, PACT, INDEX, JCSG, PEG/Ion, Salt, MIDAS; *Hampton Research* Molecular Dimensions) were set up. Although no crystals of $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG$ were obtained from crystallisation trials incubated at 20 °C, the enzyme-substrate complex formation could be affected by temperature. Crystallisation plates were therefore set up at both 20 °C and 4 °C again.

As with the apo-$AspH_{315-758}$, no sign of crystal formation was observed in the plates incubated at 20 °C. Microcrystal growth slowly started about one week after setting up in some trials incubated at 4 °C. The drops and crystal growth were monitored (manually). After several weeks, crystals of various shapes and appearance had appeared, with some having long needle shapes, chunky polygonal rods, or spheroidal forms. Proteins can often undergo conformational changes upon binding of a substrate, which can alter its shape and crystal packing, resulting in a different space group. The protein-substrate complex may crystallise under completely different conditions than $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG$ without the substrate, resulting in visibly different crystal morphology.

Several crystals with different morphology were harvested by cryo-cooling in liquid nitrogen as described earlier, and data were collected using in-house X-ray radiation to determine the space group and unit cell size. Some crystals diffracted to about 2.9 Å resolution and were indexed as monoclinic, space group $P2_1$ (**Fig. 2-31**). Crystals having identical morphology were only observed in a few conditions (PACT Premier screen well conditions containing 0.1 M Bis-Tris propane, pH 8.5, 20 % w/v PEG 3350 and either 0.2 M sodium bromide (well H2), 0.2 M potassium thiocyanate (well H4), or 0.2 M sodium nitrate (well H5)). The monoclinic crystals turned out to be twinned, and a structure solution could not be obtained using the collected data.

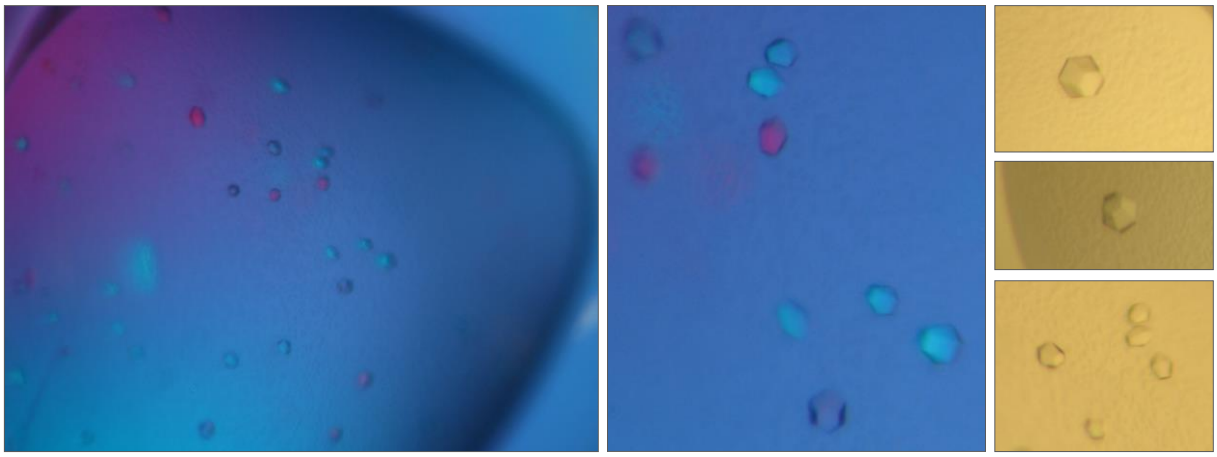


Fig. 2-31 | Morphology of the monoclinic twinned crystal form, space group $P2_1$. The crystals grew in trials with His_6-$AspH_{315-758}$ (18 mg/ml, i.e. 330 µM), 1 mM $MnCl_2{\cdot}4\,H_2O$, 2 mM NOG and 726 µM hFX $EGF1_{39mer}$ peptide, incubated at 4 °C in a sitting drop with precipitant containing 100 mM bis-tris-propane, pH 8.5, 200 mM KSCN, 20 % v/v PEG 3350. Crystals were photographed under normal and polarised light.

Most crystals from the substrate co-crystallisation trials had the same morphology (**Fig. 2-32**) as observed for the "apo" $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG$. The protein-substrate complex crystals were also orthorhombic, space group $P2_12_12_1$, but interestingly, the unit cell size differed significantly (see **Table 2-2**). Data to 2.09 Å resolution were obtained, and the crystal structure was determined by molecular replacement (work performed by Dr Michael A. McDonough).

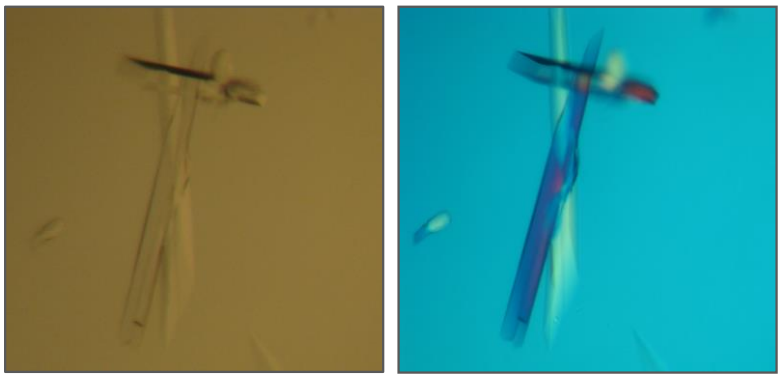


Fig. 2-32 | Morphology of $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}$hFX EGF1 complex crystals, orthorhombic space group $P2_12_12_1$, crystallised at 4 °C in a sitting drop with precipitant containing 100 mM sodium acetate trihydrate, pH 7.0, 12 % PEG 3,350. Crystals photographed under normal and polarised light.

Table 2-2 | Crystallisation conditions and data collection for the $AspH_{315-758}$ structures determined by molecular replacement (work still in progress; refinements performed by Dr Michael A. McDonough).

	$AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG$ ("apo")	**$AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}$hFX $EGF1_{39mer}$**
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* [Å]	48.73, 70.24, 170.42	50.25, 91.02, 122.67
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution [Å]	2.14	2.09
Completeness [%]	99.0	99.0
Symmetry	Orthorhombic	Orthorhombic
Crystal size	400 x 100 µm	200 x 100 µm
Method	Vapour diffusion, sitting drop (300 nl), protein-to-well ratio 1:2	Vapour diffusion, sitting drop (300 nl), protein-to-well ratio 1:2
Temp. [K]	277	277
Precip. cond.	Opti. screen: PACT, E1-well3 18 mg/ml His_6-$AspH_{315-758}$, 1 mM $MnCl_2$, 2 mM NOG, 100 mM bicine, pH 9.0, 100 mM sodium chloride, 27 % v/v PEG 550 MME	PACT, F7-well3 18 mg/ml His_6-$AspH_{315-758}$, 1 mM $MnCl_2$, 2 mM NOG, 100 mM Bis Tris Propane, pH 6.5, 200 mM sodium acetate, 20 % w/v PEG 3350, 726 µM hFX $EGF1_{39mer}$
AspH:Substrate ratio	N/A	1:2.2 (330 µM:726 µM)
Cryoprotectant	25 % glycerol	25 % glycerol
Diffraction Source	Synchrotron	Synchrotron
Site	Diamond Light Source	Diamond Light Source
Type	Diamond Beamline I02	Diamond Beamline I04
Wavelength List [Å]	0.979	0.9795
Detector	Pilatus 6M	Pilatus 6M-F

The crystal structure showed $AspH_{315-758}$ in complex with Mn(II), NOG, and the hFX $EGF1_{39mer}$ peptide substrate (**Fig. 2-33**). The structure reveals that the TPR domain is directly involved in substrate binding. Electron density for fourteen residues (-$GKC_3K\underline{\mathbf{D}}GLGEYTC_4TC_5$-) of the hFX $EGF1_{39mer}$ substrate was apparent between the AspH active site region and the TPR domain; the

other residues of the substrate were disordered. The enzyme-substrate interactions are described in more detail below.

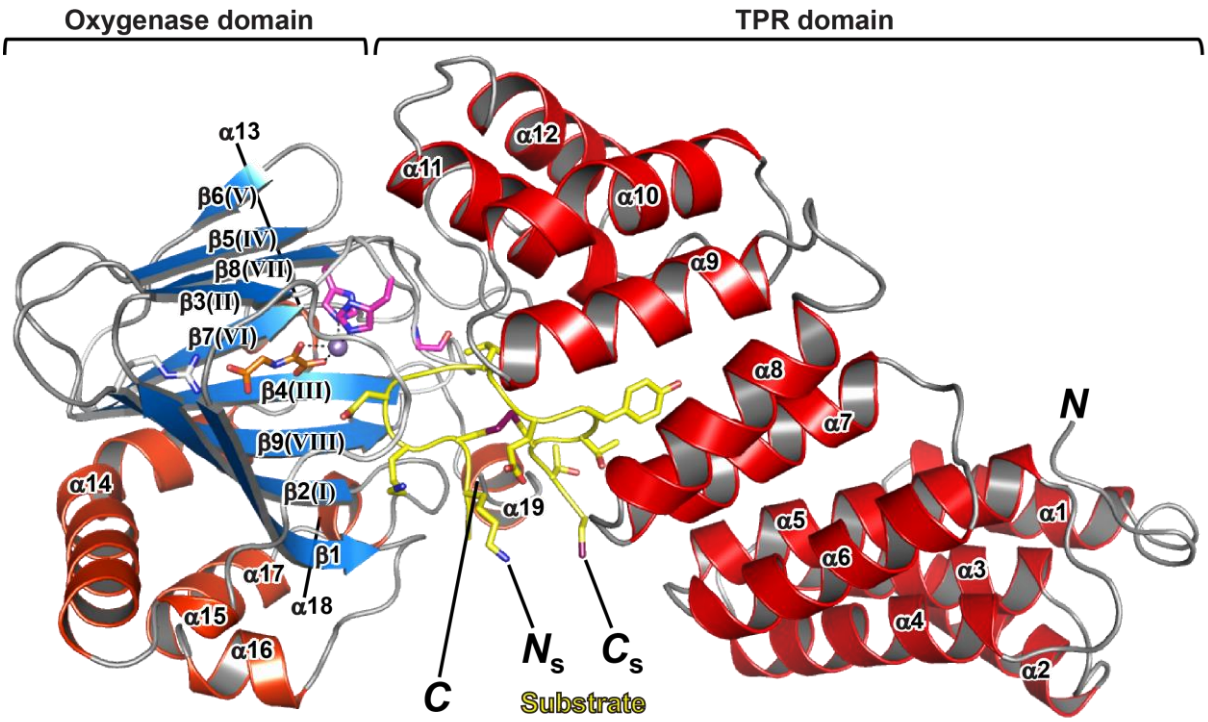


Fig. 2-33 | View from a crystal structure (resolution: 2.09 Å) of human His_6-$AspH_{315-758}$ in complex with with Mn(II), NOG and the hFX $EGF1_{39mer}$ peptide substrate (yellow).

2.4.3.1 AspH Binds Its Substrate with an "Induced Fit" Mechanism

Major conformational changes within the protein upon substrate binding are observed by comparison of the structure of "apo"-$AspH_{315-758}$ without substrate with the $AspH_{315-758}$- hFX $EGF1_{39mer}$ substrate complex structure (**Fig. 2-34a**). The two domains undergo a conformational change (about 15 Å movement) upon substrate binding, thereby clamping the substrate within a pocket. The conformational changes demonstrate that AspH binds its substrate with an "induced fit" that likely strengthens the binding while positioning the substrate for hydroxylation.

A surface representation of the $AspH_{315-758}$-hFX $EGF1_{39mer}$ substrate complex shows that the substrate is clamped in the pocket between the oxygenase and the TPR domain (**Fig. 2-34b**). The *N*- and *C*-terminal extensions of the peptide substrate exit the pocket and are disordered, likely due to a lack of protein substrate interactions and thus a higher degree of flexibility of these parts.

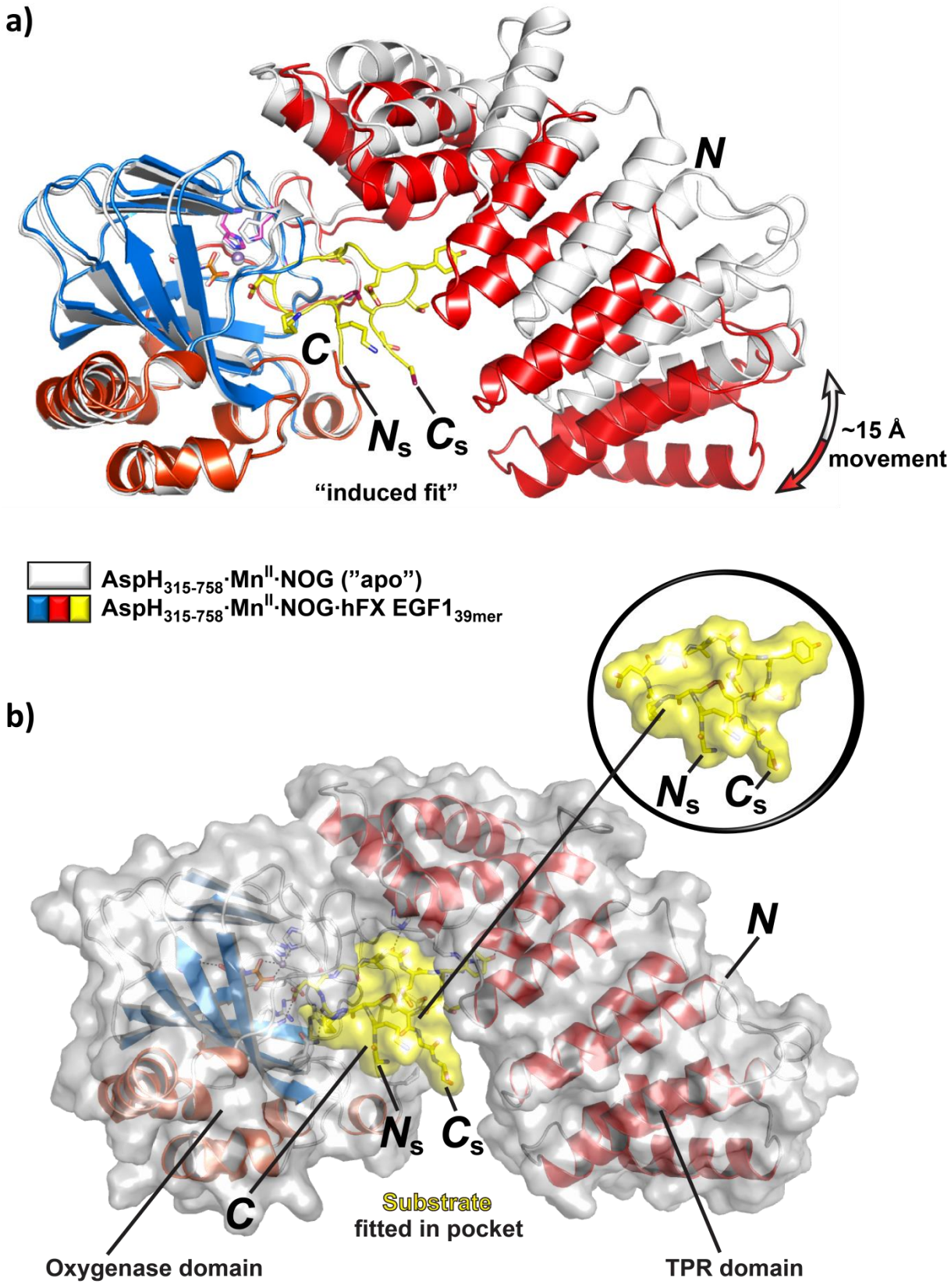


Fig. 2-34 | Evidence for "induced fit" binding of the AspH substrate. **a)** Superimposition of the oxgenase domains of "apo"-$AspH_{315-758}$ structure without substrate (light grey) with the $AspH_{315-758}$-hFX $EGF1_{39mer}$ substrate complex structure (coloured) shows that the two domains undergo a conformational change (about 15 Å movement) upon substrate binding. **b)** Surface representation of the $AspH_{315-758}$·Mn^{II}·NOG·hFX $EGF1_{39mer}$ complex. The substrate is clamped between the oxygenase and the TPR domain. N_S/C_S: *N*- and *C*-termini of the substrate.

2.4.3.2 Key Interactions Between AspH and Its Substrate

The hydroxylation site of the hFX EGF1$_{39mer}$ peptide substrate (Asp18$_{39mer}$) is positioned in the active site with the 3*R*-hydrogen oriented towards the metal (**Fig. 2-35**). The side chains of Arg688 and Lys666 interact with the Asp18$_{39mer}$ carboxylate. The side chains of Arg686 and Gln664 bind the peptide backbone of the substrate, and the side chain of Arg686 is also held in position by a salt bridge with Asp721, and Arg688 is stabilised by a salt bridge with Asp741.

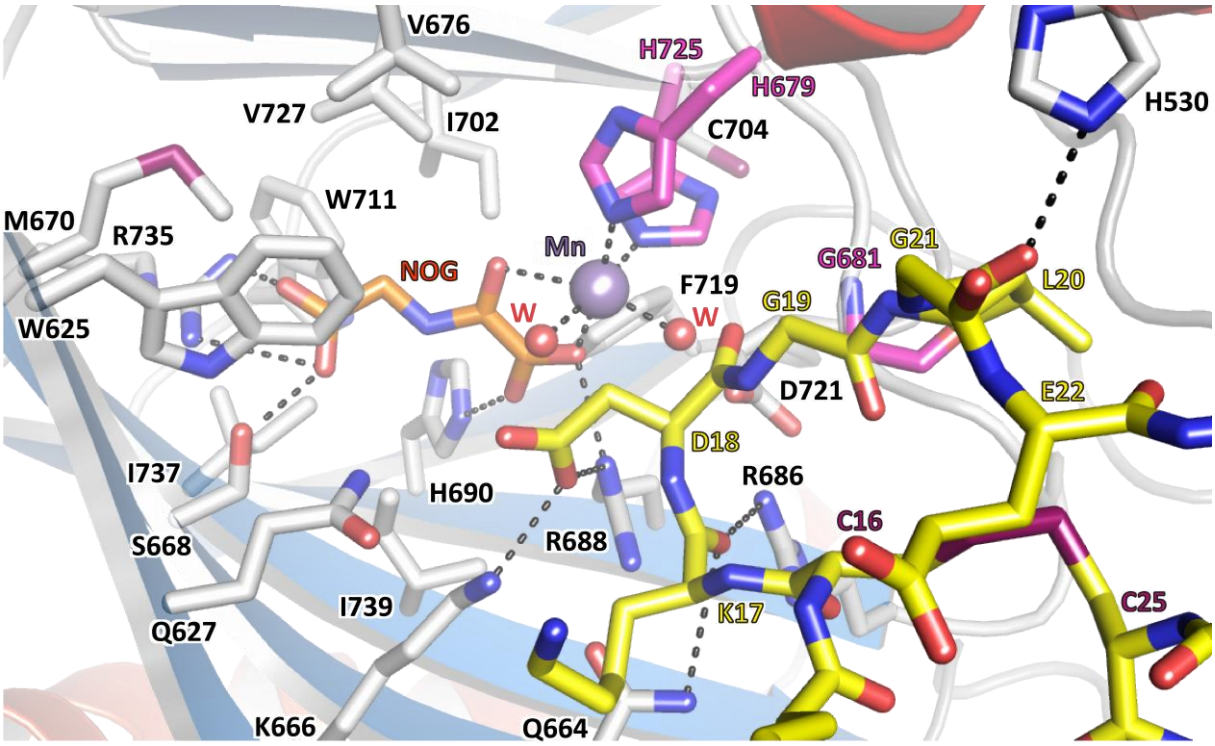


Fig. 2-35 | Close-up view of the AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ complex active site. The HXG…H catalytic dyad is shown in magenta, the substrate in yellow. The hydroxylation site of the peptide substrate (Asp18$_{39mer}$) is positioned in the active site with the 3*R*-hydrogen oriented towards the metal.

The flexible loop comprising residues 616-619 first identified in the "apo"-AspH$_{315-758}$·and the AspH$_{562-758}$ structures has a different conformation in the AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ complex structure. The side chain of loop residue Asp616 forms an electrostatic interaction with the side chain of the substrate residue Lys17$_{39mer}$, and the side chain of loop residue Glu617 forms a hydrogen bond with the amide nitrogen of the Asp18$_{39mer}$ backbone (**Fig. 2-36a**). Superimposition of the different AspH crystal structures reveals different conformations for several loop residues, further supporting the proposed flexibility of this loop region (**Fig. 2-36b**). Glu617 is positioned in the active site mimicking the substrate Asp18$_{39mer}$ residue in the AspH$_{562-758}$·Ni^{II}·NOG·L-Malate structure, whereas this residue is directed away from the active site in the "apo"-AspH$_{315-758}$ structure. In the AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ structure the loop folds over the active site to bind the substrate backbone and the Glu617 side chain forms a hydrogen bond with the Asp18$_{39mer}$ backbone.

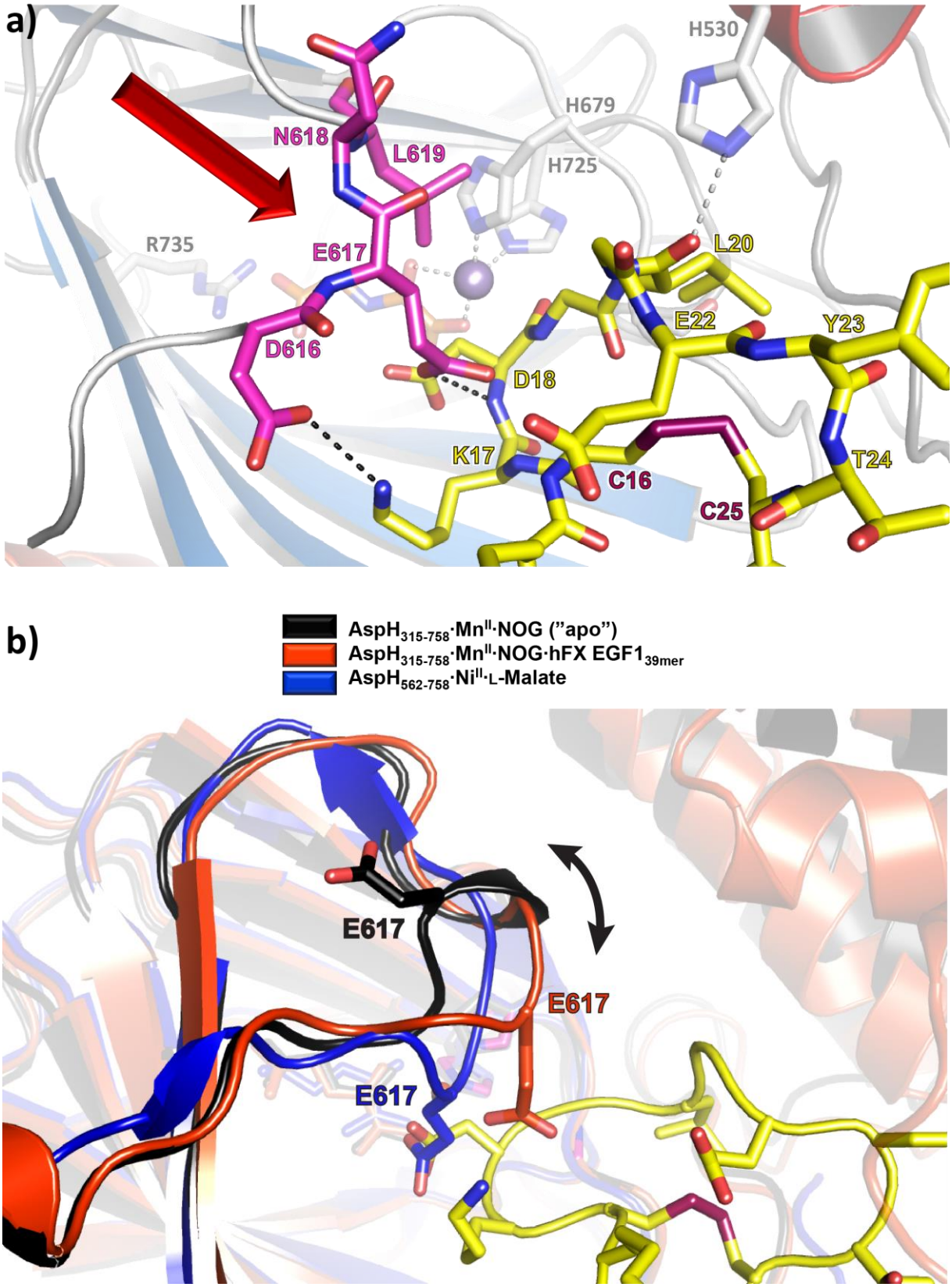


Fig. 2-36 | Flexibility of the loop region comprising AspH residues 616-619. **a)** Close-up of the $AspH_{315-758}$-hFX $EGF1_{39mer}$ complex active showing important interactions between the loop (magenta) and the substrate. **b)** Superimposition of the AspH structures shows three different conformations in the loop region (aa 616-619; only the side chain of the representative loop residue Glu617 is shown, others are omitted for clarity).

The crystal structure of the $AspH_{315-758}$·Mn^{II}·NOG·hFX $EGF1_{39mer}$ complex also revealed important interactions between the substrate and the TPR domain. The conserved residue $Tyr23_{39mer}$, which is part of the proposed hydroxylation consensus sequence (-C_3X(D/N)XXXX(Y/F)XC_4XC_5-),[338] is positioned in a hydrophobic pocket of the TPR domain (**Fig. 2-37**). The substrate binding pocket is lined by hydrophobic residues including Phe527, Ile497, Phe489, Phe496, Val492, Val462, Leu465 and Leu466. Additional electrostatic interactions further stabilise the substrate binding; the side chains of His493 and Arg526 form hydrogen bonds with the $Tyr23_{39mer}$ hydroxyl group and backbone carbonyl, and the side chains of Phe496 and Arg526 also π-stack with the $Tyr23_{39mer}$ side chain. The side chains of His530 and Tyr565 interact with the backbone carbonyl of $Leu27_{39mer}$ and the backbone nitrogen of $Cys25_{39mer}$ respectively.

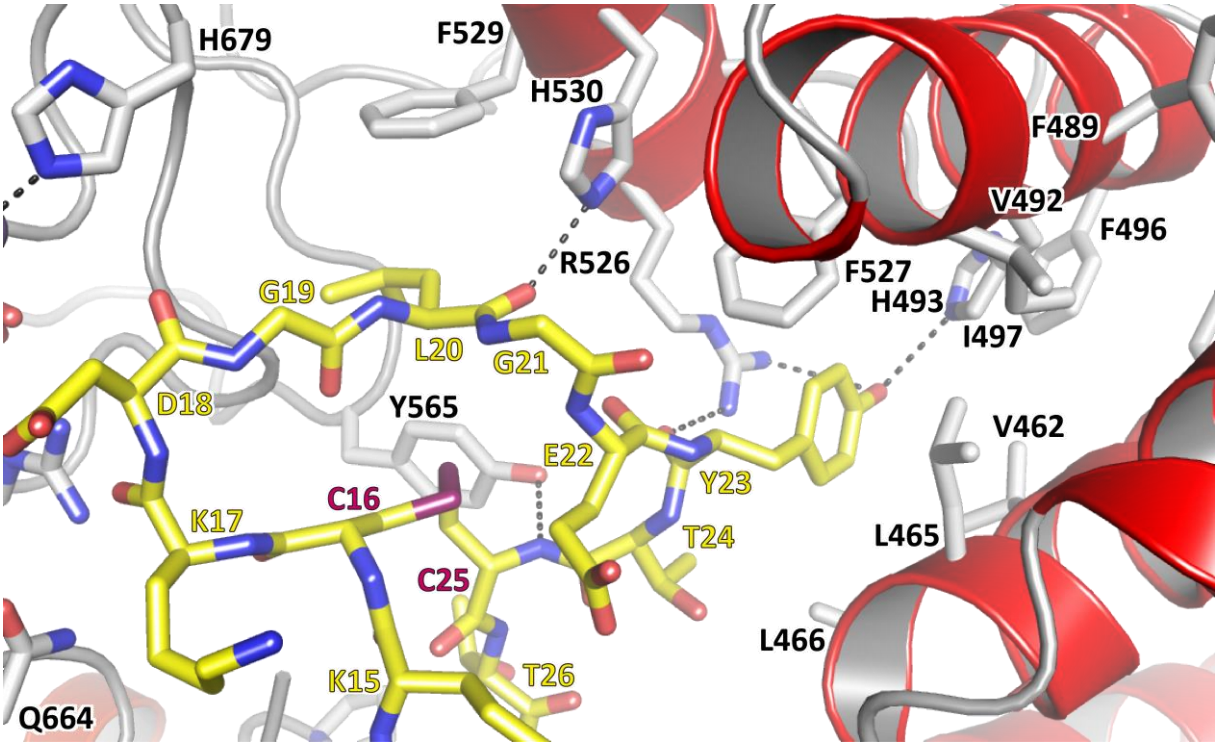


Fig. 2-37 | Substrate binding in a hydrophobic pocket within the TPR domain. The highly conserved $Tyr23_{39mer}$ interacts with several hydrophobic residues of the TPR domain. In addition, substrate binding is supported by electrostatic interactions between the side chains of several AspH residues and the substrate backbone.

Interestingly, analysis of the hFX $EGF1_{39mer}$ substrate electron density maps revealed clear evidence for a disulfide bond between $Cys16_{39mer}$ and $Cys25_{39mer}$. This corresponds to a Cys3-4 EGFD linkage and suggested the presence of a non-canonical disulfide pattern in AspH substrates, i.e. not the expected Cys1-3, 2-4, 5-6 pattern found in deposited crystal and NMR structures of isolated EGFDs. Moreover, the adjacent anti-parallel β-strands typical of EGFDs were not observed in the $AspH_{315-758}$·Mn^{II}·NOG·hFX $EGF1_{39mer}$ crystal structure. The AspH *C*-terminal residue, Ile758, is remarkably close to the active site and the observed disulfide bond in the substrate (**Fig. 2-38**), raising the question if the *C*-terminus could also have a role in substrate recognition and/or catalysis.

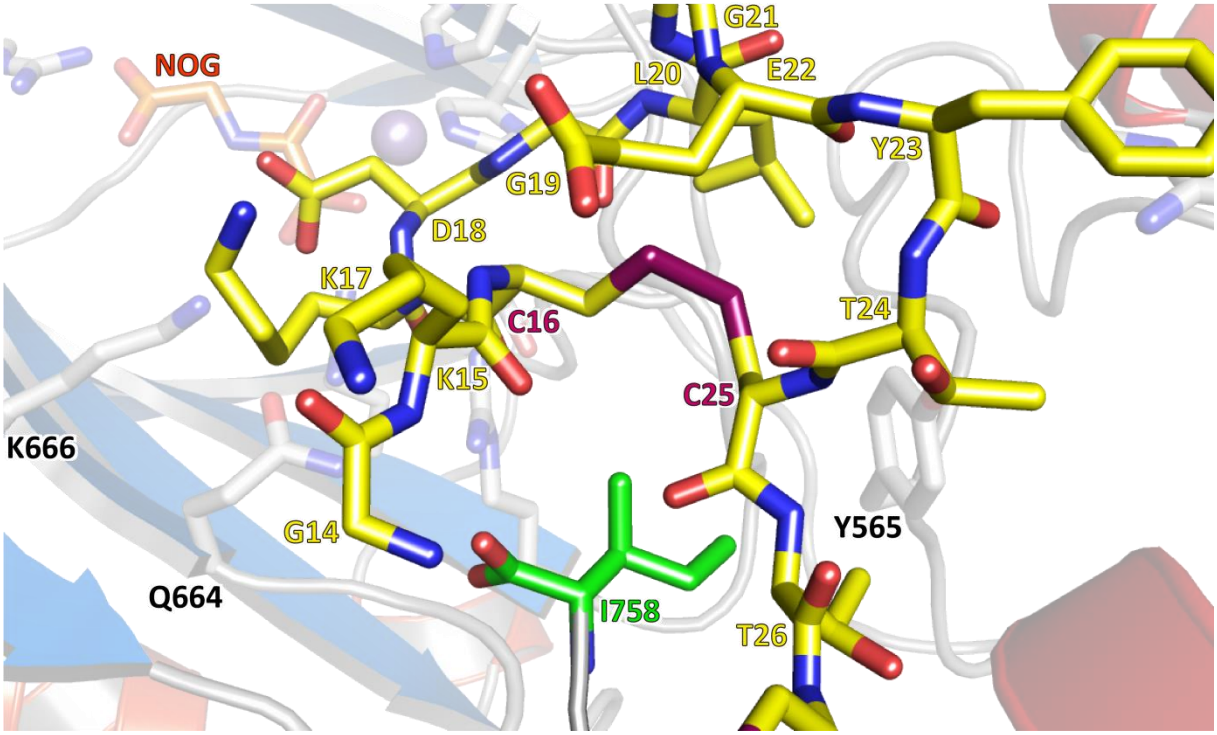


Fig. 2-38 | Close-up view of the observed substrate disulfide bond in the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ complex crystal structure. A non-canonical disulfide bond between$Cys16_{39mer}$ and $Cys25_{39mer}$ was found, which corresponds to Cys3-4 linkage and differs from the expected canonical Cys1-3, 2-4, 5-6 pattern. The AspH *C*-terminus (I758, green) is in close proximity to the active site and the substrate disulfide, which suggest a role for substrate recognition and/or catalysis.

2.5 DISCUSSION

The crystallographic studies on AspH presented here provide insights into a novel substrate conformation and reveal important movements between the oxygenase and TPR domains that explain why the TPR domain is required for substrate binding and hydroxylation.

Although only fourteen residues of the hFX $EGF1_{39mer}$ substrate were apparent in the electron density maps, the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ structure appeared to capture key interactions between AspH and its EGFD substrates. Moreover, the active site metal binding geometry is apparently widely identical in all obtained AspH crystal structures (except from the mobile loop). Several features of the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ complex structure (**Fig. 2-39**) indicate highly specific interactions including a number of protein-protein contacts including: i) the interaction of the side chain of AspH residue Tyr565 with the backbone amide of $Cys25_{39mer}$; ii) the close proximity of AspH Ile758 to the substrate disulfide bond; iii) the interactions of the flexible loop residues with the substrate. These findings suggest that the unexpected Cys3-4 disulfide bond found in the substrate is not a crystallographic 'artefact'.

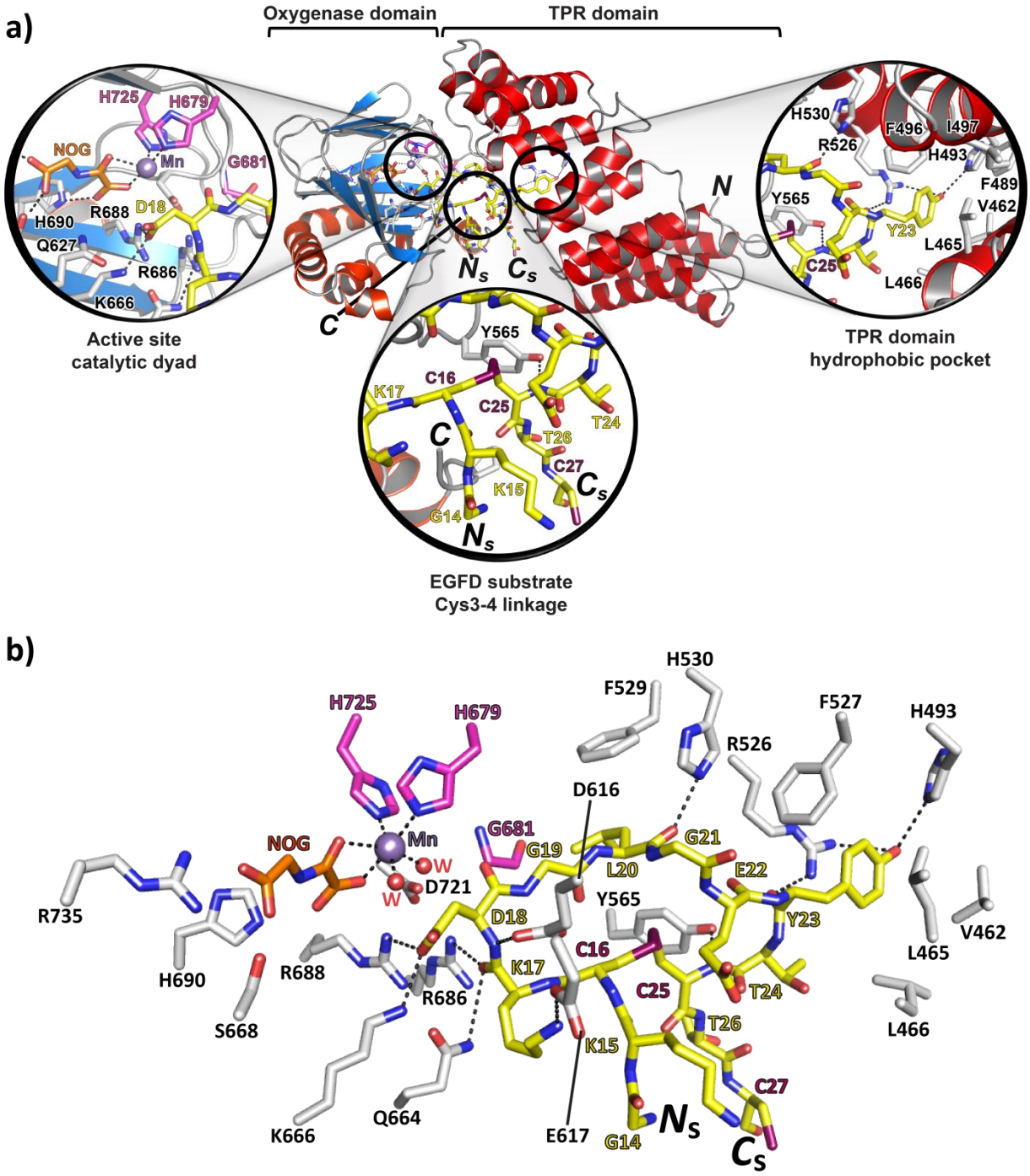


Fig. 2-39 | Summary of key features and interactions in the obtained $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ crystal structure. **a)** AspH has an unprecedented active site geometry with its catalytic dyad; the TPR domain is essential for substrate recognition, and a non-canonical Cys3-4 disulfide was observed in the substrate. **b)** Sticks representation of AspH residues interacting with the EGFD substrate.

2.5.1 An Unusual Family Member with Unique Active Site Geometry

As described in **Chapter 1**, the active site of 2OG oxygenases typically features a 'facial triad' comprised of one carboxylate (either Asp or Glu) and two histidine residues ligating to the Fe(II), as e.g. in FIH (**Fig. 2-40a**, PDB: 3P3N[66]). The active site metal is octahedrally coordinated. Cosubstrate 2OG serves as a bidentate ligand (C-1-carboxylate and C-2-carbonyl), and the sixth coordination site

is occupied by a water molecule, which is displaced by oxygen in the catalytic cycle. AspH lacks the highly conserved carboxylate ligand of the 'facial triad'; this residue is replaced by a glycine (HXG…H motif). Only two Fe(II)-binding ligands (His679 and His725, catalytic dyad) are donated by the protein (**Fig. 2-40b**). All 2OG oxygenases thus far identified that have only two protein ligands coordinating the Fe(II) centre are prokaryotic halogenases that contain a HXA…H motif such as the representative SyrB2[427] (**Fig. 2-40c**, PDB: 2FCT[427]). The halogenases have a halide ion and a water molecule occupying the remaining ligand positions. In contrast, the equivalent halide coorination site in AspH is occupied by a water molecule. The AspH structure presents the first example of a mammalian member of the 2OG oxygenase superfamily with such an Fe(II)-coordination mode.

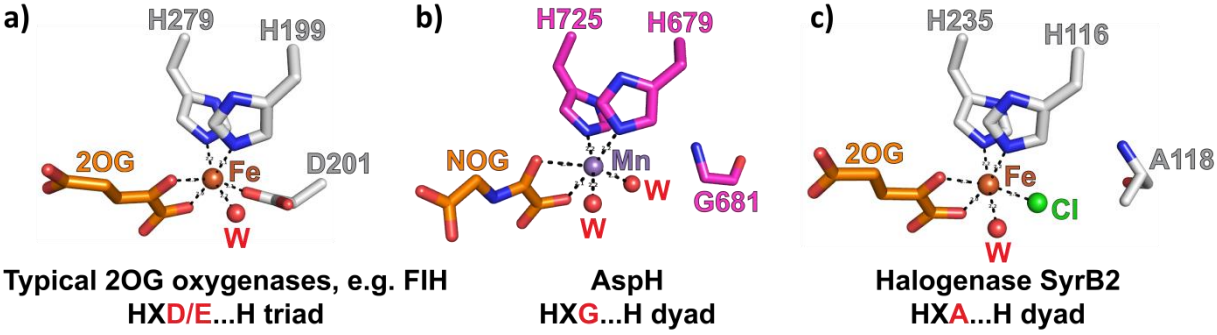


Fig. 2-40 | Structural representation of 2OG oxygenase active site geometries. **a)** The highly conserved catalytic triad (HXD/E…H) typical for 2OG oxygenases, represented by the active site of the HIF asparaginyl hydroxylase FIH (PDB: 3P3N[66]). **b)** AspH active site with HXG…H, showing coordination of the metal by His679 and His725, bidentate metal coordination by NOG and two water (W) molecules. **c)** The active site of 2OG-dependent prokaryotic halogenases, represented by the syringomycin E halogenase SyrB2 (PDB: 2FCT[427]), showing metal coordination by a catalytic dyad (HXA…H), the 2OG C-1-carboxylate and a halide.

The unusual Fe(II)-coordination mode in AspH with the HXG…H motif and two water molecules presents several potential binding sites for oxygen; detailed mechanistic studies will be required to elucidate the exact catalytic mechanism. Based on the consensus 2OG oxygenase mechanism, a catalytic cycle involving the catalytic dyad can be proposed for AspH (**Fig. 2-41**).

Variants of FIH demonstrated that the active site carboxylate ligand is not required for Fe(II)-binding and hydroxylation, as the FIH-D201G variant retained hydroxylase activity, while the D201A mutant lost the ability to bind the substrate.[428] None of the FIH variants showed halogenase activity in this study.[428]

Superimposition of the AspH and SyrB2 structures reveals that there is about the same amount of space (6.0 Å measured from the metal center to C-α of AspH Gly681 or the SyrB2 Ala118 methyl group C-β respectively) for a chloride ion to coordinate the Fe(II) (**Fig. 2-42**). For SyrB2 the positioning of the substrate was found to be a determinant favouring chlorination over hydroxylation, and slight repositioning could alter the reaction.[429] In the AspH crystal structure, the substrate $Asp18_{39mer}$ C-β hydroxylation site is closer to the water molecule often found in the same

position in other 2OG oxygenases than to the position of the chloride ion in syrB2 (2.8 Å compared to 3.4 Å). For AspH, no halogenation of any substrates in presence of chloride or bromide ions has been observed in the studies thus far.

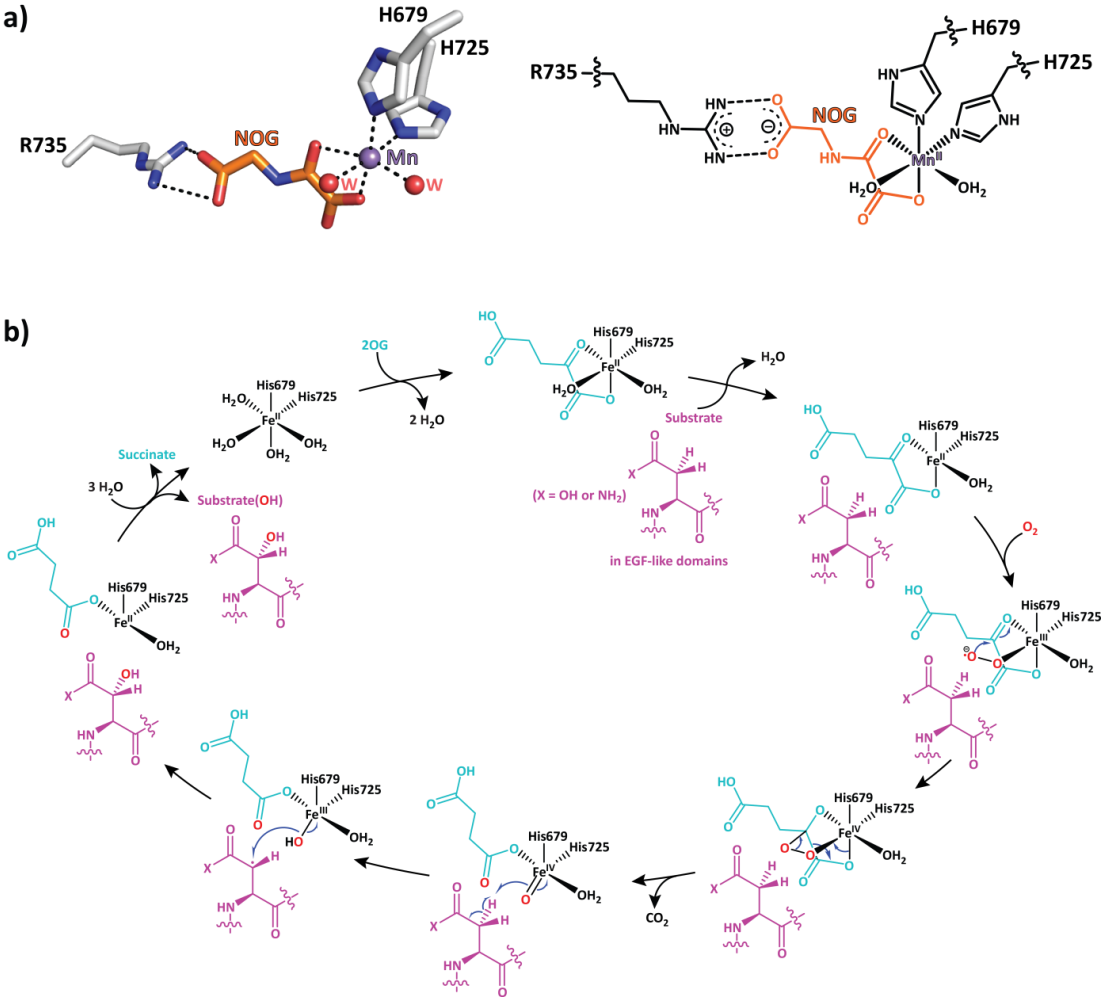


Fig. 2-41 | The unique AspH active site geometry. **a)** Structural and skeletal formal representation of the AspH active site. **b)** Proposed mechanism involving the catalytic dyad, based on the consensus 2OG oxygenase mechanism (see **Chapter 1**).

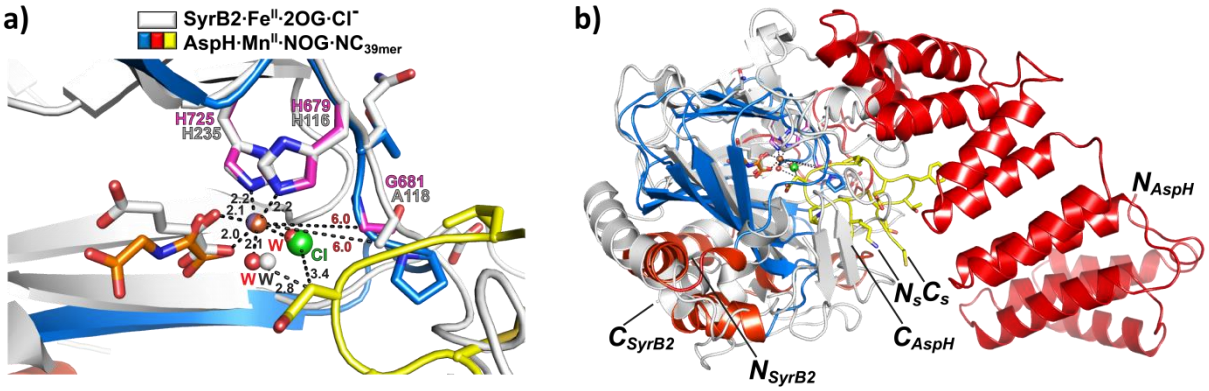


Fig. 2-42 | Superimposition of a structure for the halogenase SyrB2 (light grey, PDB: 2FCT[427]) with that of AspH$_{315\text{-}758}$ (coloured) in complex with the hFX EGF1$_{39mer}$ substrate (yellow). **a)** Active site close-up. Dashed lines indicate the distance between the metal centre and the AspH Gly681 C-α or the SyrB2 Ala118 methyl group C-β respectively, potentially allowing space for a halide ion. The Asp18$_{39mer}$ C-β of the AspH substrate, however, is farther away from the chloride ion than from the water molecule, which is predicted to be displaced by oxygen during catalysis. **b)** Overall, AspH and SyrB2 show no sequence or strong secondary structure similarity.

As described in **Chapter 1**, *AspH* knockout studies in mice resulted in phenotypes with developmental defects (**Fig. 2-43a**).[263] The observed similar abnormalities in humans with Traboulsi syndrome (**Fig. 2-43b**) were found to originate from *AspH* mutations affecting the catalytic activity.[360] The found Arg735Trp variant likely disrupts 2OG-binding as a result of loss of the critical salt bridge interaction and potential clashes between 2OG and the bulky Trp side chain (**Fig. 2-43c** and **d**).

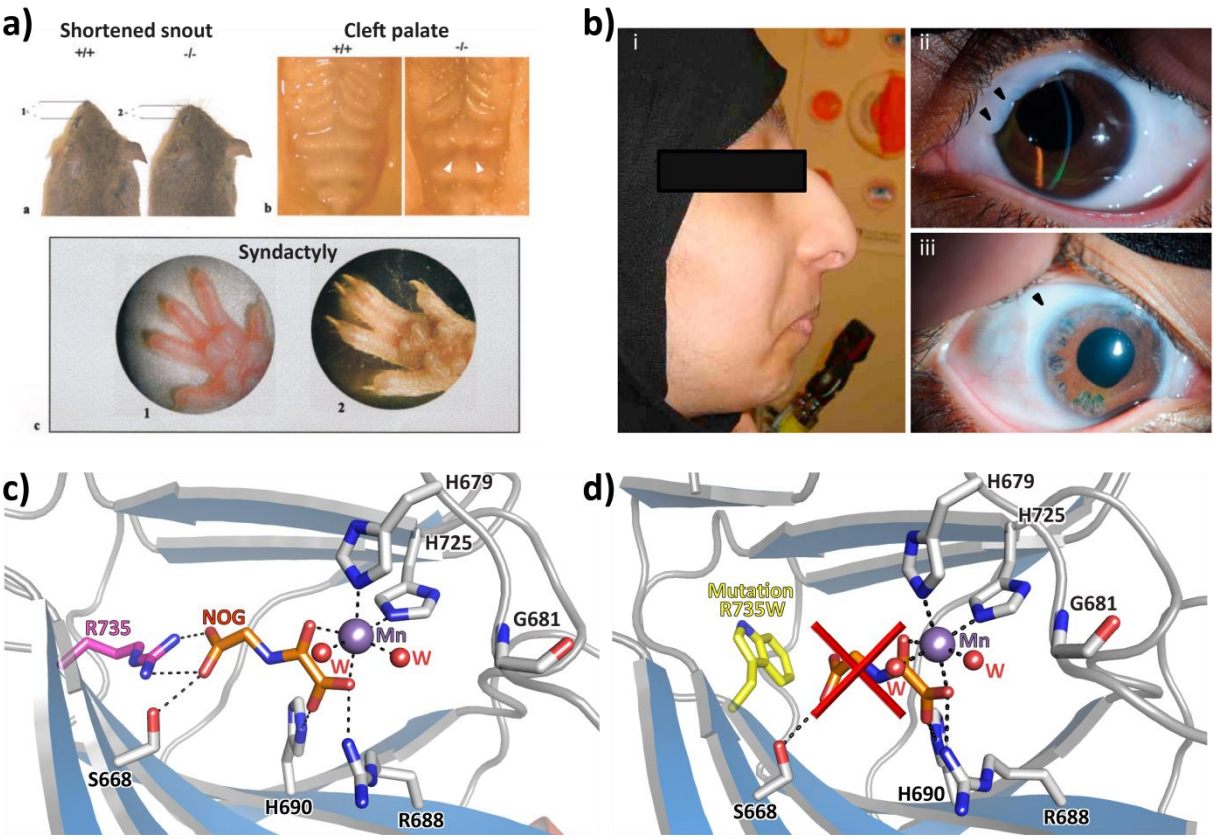


Fig. 2-43 | Links of impaired AspH activity to potential biological significance. **a)** Observed phenotypes in form of developmental defects in *AspH* mouse knockout study (adapted from Dinchuk *et al.* with permission).[263] **b)** Observed phenotypes in humans with Traboulsi syndrome: facial dysmorphism, lens dislocation and anterior-segment abnormalities (reproduced from Patel *et al.* with permission).[360] **c)** Close-up of the "apo"-AspH$_{315\text{-}758}$ structure 2OG-binding pocket showing the salt bridge between Arg735 and the NOG C-5-carboxylate. **d)** Model of the R735W mutation found in humans with Traboulsi syndrome.

2.5.2 A Non-Canonical Disulfide Bond in the Substrate

The observation of a Cys3-4 disulfide bond in the hFX EGF1$_{39mer}$ substrate in the AspH$_{315\text{-}758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ structure raised the question if this was linked to the physiological function of AspH. At this stage it was unclear if this disulfide bond was already present in the peptide material before incubation with AspH, or if disulfide shuffling had occurred, possibly induced by AspH.

Some proteins localised to the endoplasmic reticulum (ER) such as protein disulfide isomerase (PDI) have chaperone-like activities.[430,431] PDI aids protein folding by post-translational

formation and breaking of disulfide bonds between cysteine residues to the preferred arrangement.[430,431] PDI contains four thioredoxin domains, which are characterised by two vicinal cysteines in a CXXC motif and enable reduction and disulfide exchange in a substrate.[432,433] However, localisation of all the cysteine residues in the $AspH_{315\text{-}758}$ structure suggested that they are too distant and buried from the substrate Cys3-4 disulfide to be directly involved in disulfide rearrangement (**Fig. 2-44**). Interestingly though, AspH Cys641 and Cys648 form an intramolecular disulfide bond in all crystal structures.

To investigate the non-canonical Cys3-4 disulfide observed in the HFX $EGF1_{39mer}$ peptide substrate in the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}$hFX $EGF1_{39mer}$ structure, the peptide material was further characterised and detailed biochemical analyses on AspH substrate requirements were conducted in the following **Chapters**.

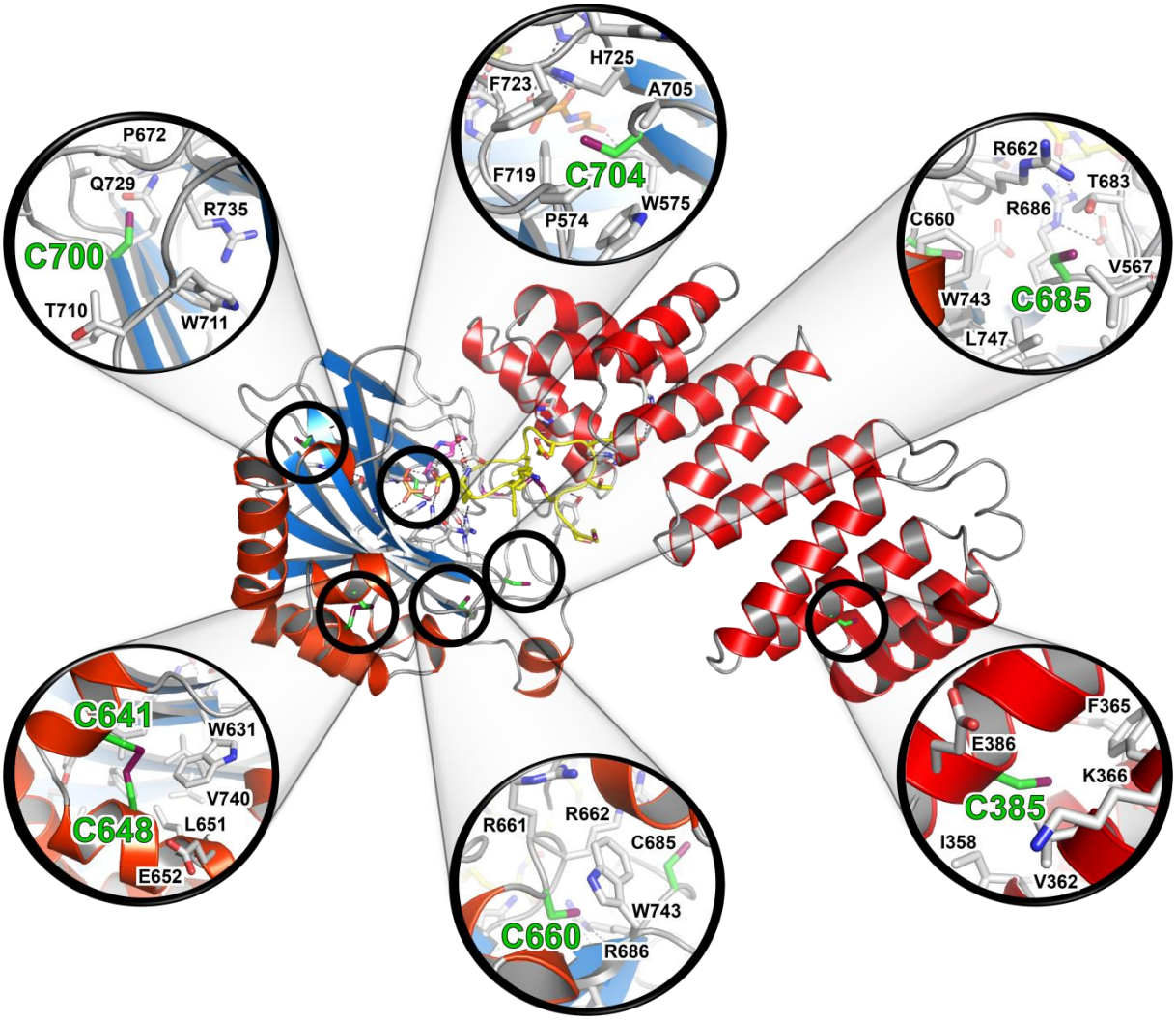
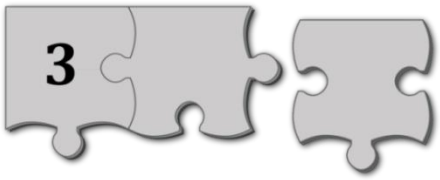


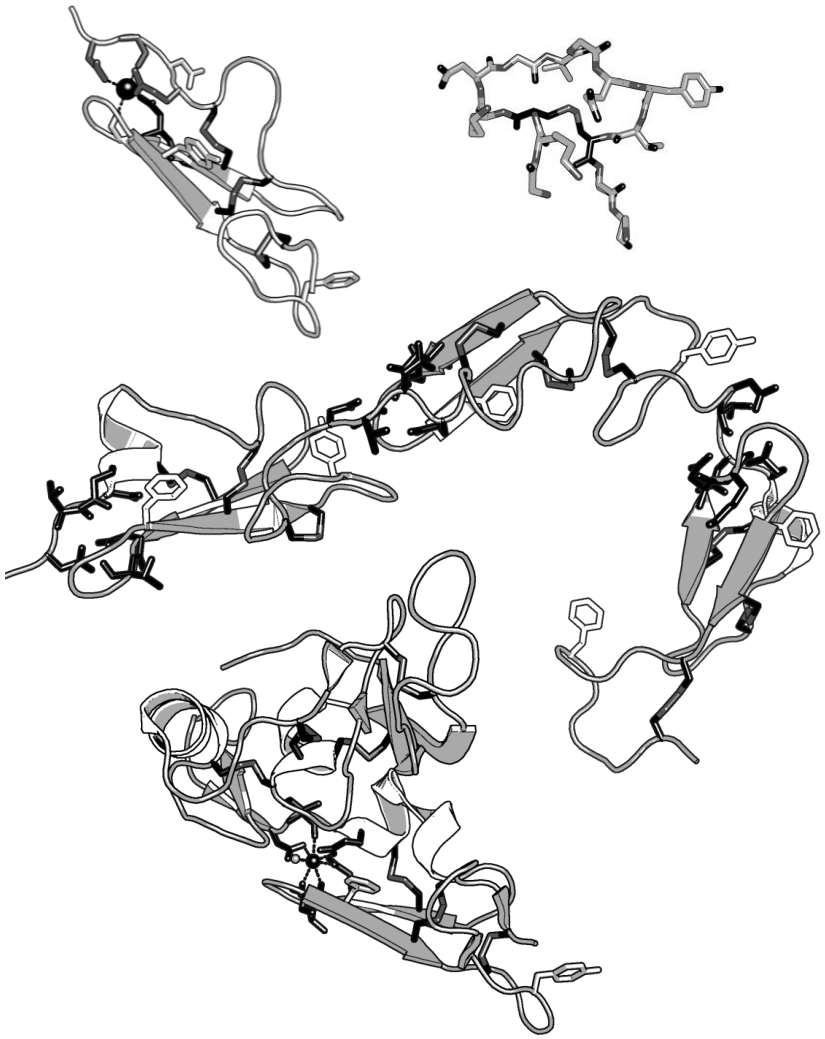
Fig. 2-44 | Position of all $AspH_{315\text{-}758}$ cysteine residues (highlighted in green) as observed in the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}$hFX $EGF1_{39mer}$ structure.

2.6 ACKNOWLEDGEMENTS FOR CHAPTER 2

Dr Michael A. McDonough harvested crystals, collected and processed the data, solved and refined the $AspH_{315-758}$ structures and is kindly acknowledged for these contributions. The SGC researchers Dr T. Krojer, Dr G. Kochan, Prof U. Oppermann are acknowledged for providing the His_{6}-$AspH_{562-758}$ plasmid and for $AspH_{562-758}$·Ni^{II}·L-Malate crystal structure determination. *ReOx Ltd* researchers Dr K. Hewitson and Dr L. McNeill are acknowledged for providing the GST-$AspH_{315-758}$ plasmid and initial work on AspH. Many thanks also go to Dr R.J. Hopkinson for operating the NMR spectrometer, Dr A.P. Hardy for training in the radiochemical 2OG turnover assays, Dr H. Gerlach for providing reagents for IVTT and assistance in performing the assay, Dr J. Brem for running protein standards with analytical gel filtration chromatography for construction of the calibration curve, Dr L. Harrington for providing reagents for Tris-tricine SDS-PAGE, the *Peptide Synthetics* (Hampshire, UK) team for the hFX $EGF1_{39mer}$ syntheses, and the researchers and staff at Diamond Light Source, Didcot, for kind technical assistance.



AspH Requires EGFD Substrates with a Non-Canonical Cys3-4 Disulfide for Activity



Abstract

This Chapter describes how the unexpected findings arising from the first AspH-substrate crystal structure were investigated and how the AspH substrate selectivity was characterised *in vitro*.

3.1 CHARACTERISATION OF THE HFX EGF1$_{39MER}$ SUBSTRATE

Disulfide bonds can stabilise the folded structures of proteins and are often critical for functionally relevant folding.[434-436] Since a non-canonical Cys3-4 disulfide bond was observed in the hFX EGF1$_{39mer}$ substrate in the AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ structure, it was important to characterise the hFX EGF1$_{39mer}$ substrate further by investigating its disulfide pattern in solution. For a monomeric oxidised EGFD with its six cysteine residues, 15 different disulfide patterns are theoretically possible (**Fig. 3-1**).

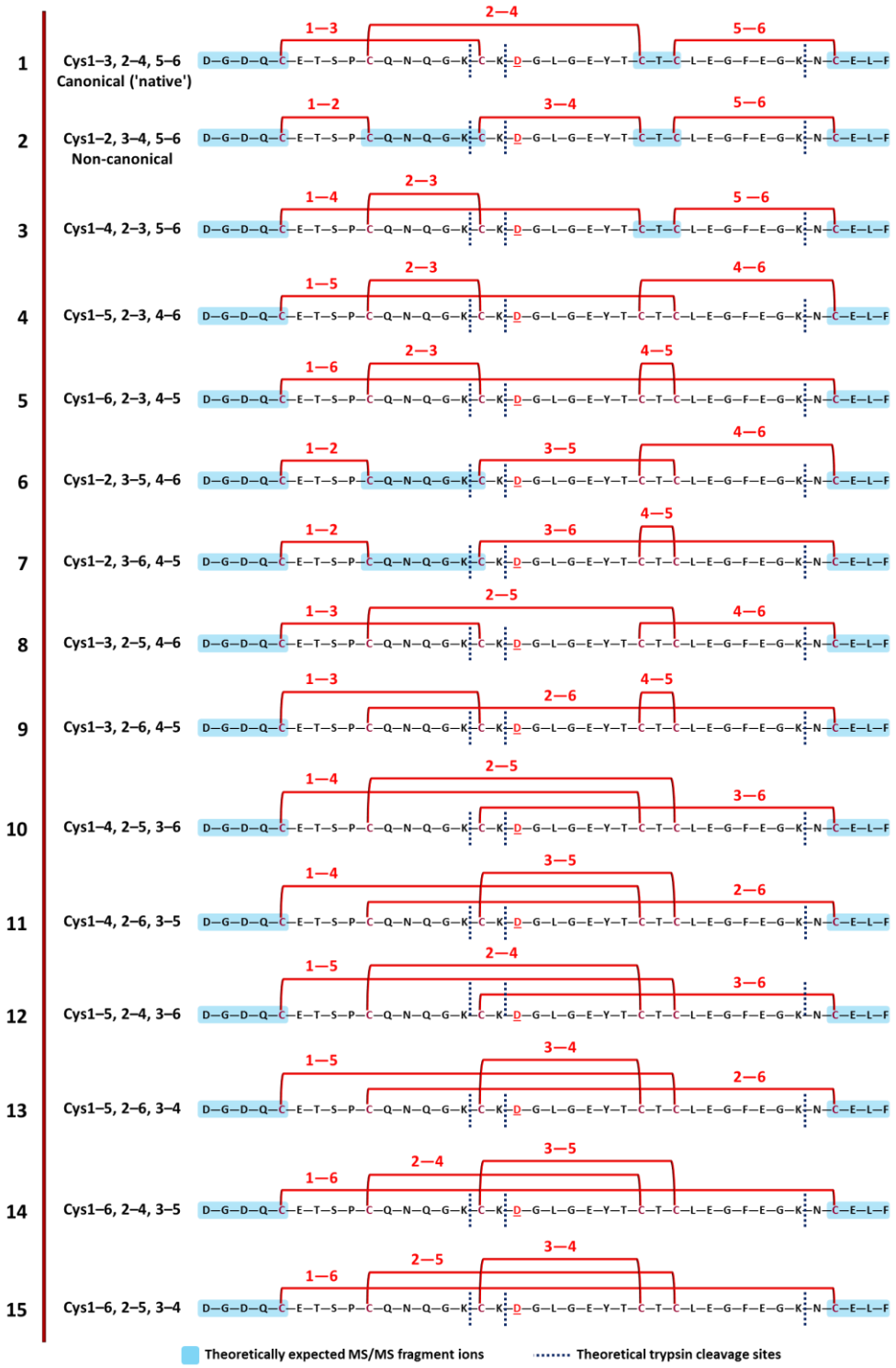


Fig. 3-1 | Schematic of the 15 theoretically possible disulfide bond arrangements in the hFX EGF1$_{39mer}$.

LC-MS analyses of the hFX EGF1$_{39mer}$ material using a C_8 column (*Phenomenex* Aeris™ WIDEPORE 3.6 µm XB-C8) on a Dionex™ UltiMate 3000 HPLC (*Thermo Scientific*) and a Q Exactive™ mass spectrometer equipped with an Orbitrap mass analyser (see **Chapter 10** for details) showed a broad, unsymmetrical peak in the UV absorbance trace (retention time: t_R = 4.41 min; **Fig. 3-2a**).

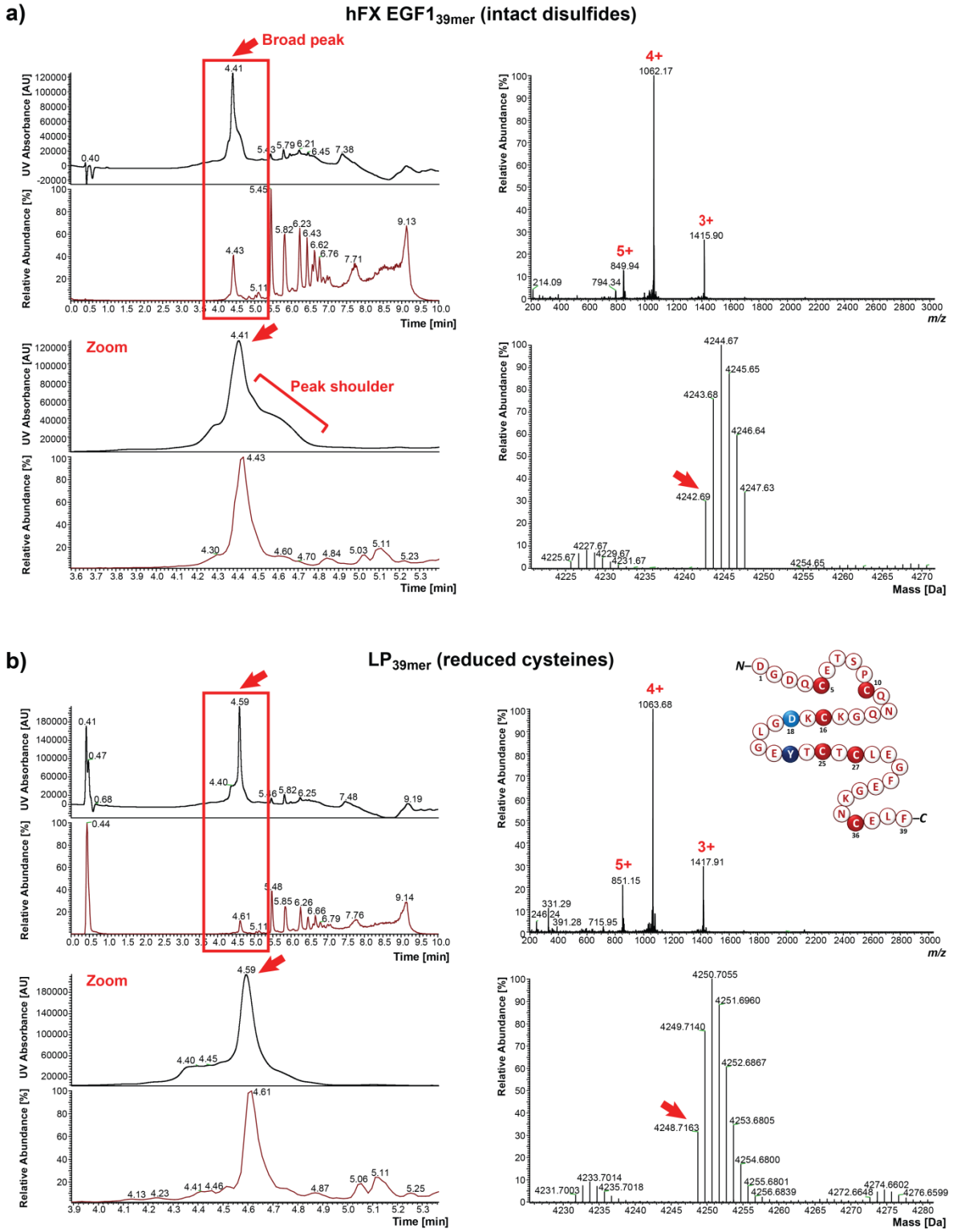


Fig. 3-2 | LC-MS analyses of **a)** the hFX EGF1$_{39mer}$ material (intact disulfides) in H_2O and **b)** the corresponding linear peptide, LP$_{39mer}$ (cysteines reduced using *tris*(2-carboxyethyl)phosphine (TCEP) in 50 mM citrate buffer, pH 3.0, which remained in the sample until analysis) performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (*Thermo Scientific*), coupled to a Dionex™ UltiMate® 3000 HPLC. Represented as UV absorbance and MS traces, and the mass spectra corresponding to the indicated peak (most abundant charge states and deconvoluted mass spectra, monoisotopic mass indicated).

The MS scan analysis revealed that the entire broad peak (including the peak shoulder) corresponded to one or several peptide species with the same molecular weight, indicating that several disulfide isoforms may have co-eluted under the broad peak (**Fig. 3-2a**; M_W = 4,242.69 Da, monoisotopic peak; consistent with the calculated mass of the peptide with six oxidised cysteines, i.e. three disulfides). An analysis of the corresponding linear peptide, LP_{39mer}, which was prepared using the reducing reagent *tris*(2-carboxyethyl)phosphine (TCEP) for reduction of the disulfides, resulted in a much sharper peak with a longer retention time, compared to the oxidised form of the peptide (t_R = 4.59 min; **Fig. 3-2b**; M_W = 4,248.72 Da, monoisotopic peak; consistent with the calculated mass of the peptide with reduced cysteines).

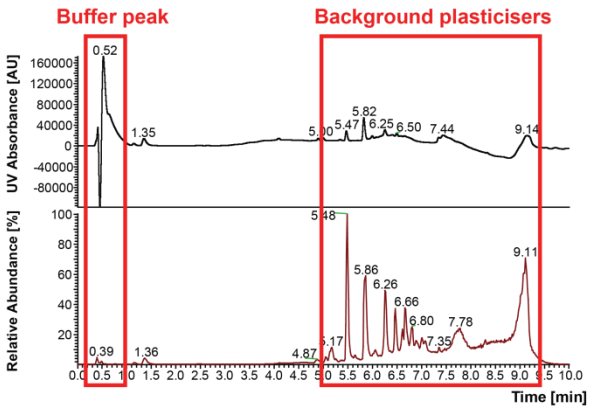


Fig. 3-3 | LC-MS analysis of a buffer control sample (50 mM citrate buffer, pH 3.0) showing background peaks likely resulting from contaminants such as plasticisers found in all samples.

A control sample (50 mM citrate buffer, pH 3.0) run under the same LC-MS conditions allowed the identification of background peaks in the MS traces (**Fig. 3-3**). These peaks likely resulted from contaminants such as plasticisers, which were observed in all samples.

3.1.1 Solution Evidence for an Unusual Disulfide Pattern

Tandem mass spectrometry (MS/MS) peptide fragmentation approaches were then employed for further characterisation of the hFX $EGF1_{39mer}$ material (**Fig. 3-4**). Fragmentation of ions in MS/MS analyses provides information about the structure and composition of the molecule and can thus be used for peptide/protein sequencing.[437-439]

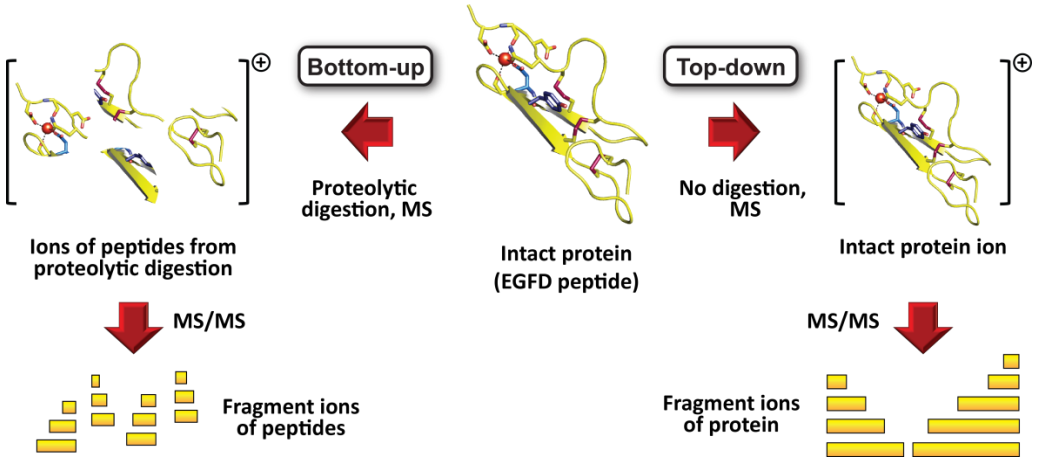


Fig. 3-4 | Scheme of 'bottom-up' and 'top-down' proteomics approaches. Both approaches were used for the characterisation of EGFD peptides and proteins in this work.

The 'bottom-up' approach[440,441] (**Fig. 3-4**), in which the EGFD peptides or proteins were first digested using a protease and then analysed using MS and MS/MS, was used to identify the hydroxylation site. Similar to common 'top-down' proteomics methods,[442-446] another approach subjecting the intact EGFD peptide or protein (including intact disulfides) to MS and MS/MS analysis was used to characterise the disulfide patterns. MS and MS/MS measurements were performed using a *Bruker Daltonics* Ultraflex I™ MALDI-ToF/ToF mass spectrometer or a *Bruker Daltonics* amaZon™ Ion Trap LC-MS/MS system (see **Chapter 10**).

Peptide sequencing using MS/MS and comparison of the observed with the calculated masses for the peptide fragment ions can reveal the site of a posttranslational modification. MS/MS fragmentation of peptides readily occurs along the peptide backbone, resulting in the most commonly observed b- and y-ion series (**Fig. 3-5a**; the charge is retained on the *N*- or the *C*-terminus respectively).[437] Other sequence-specific fragment ion series or satellite ions (**Fig. 3-5b**) that may also be useful for peptide sequencing can be observed under certain conditions.[437, 447]

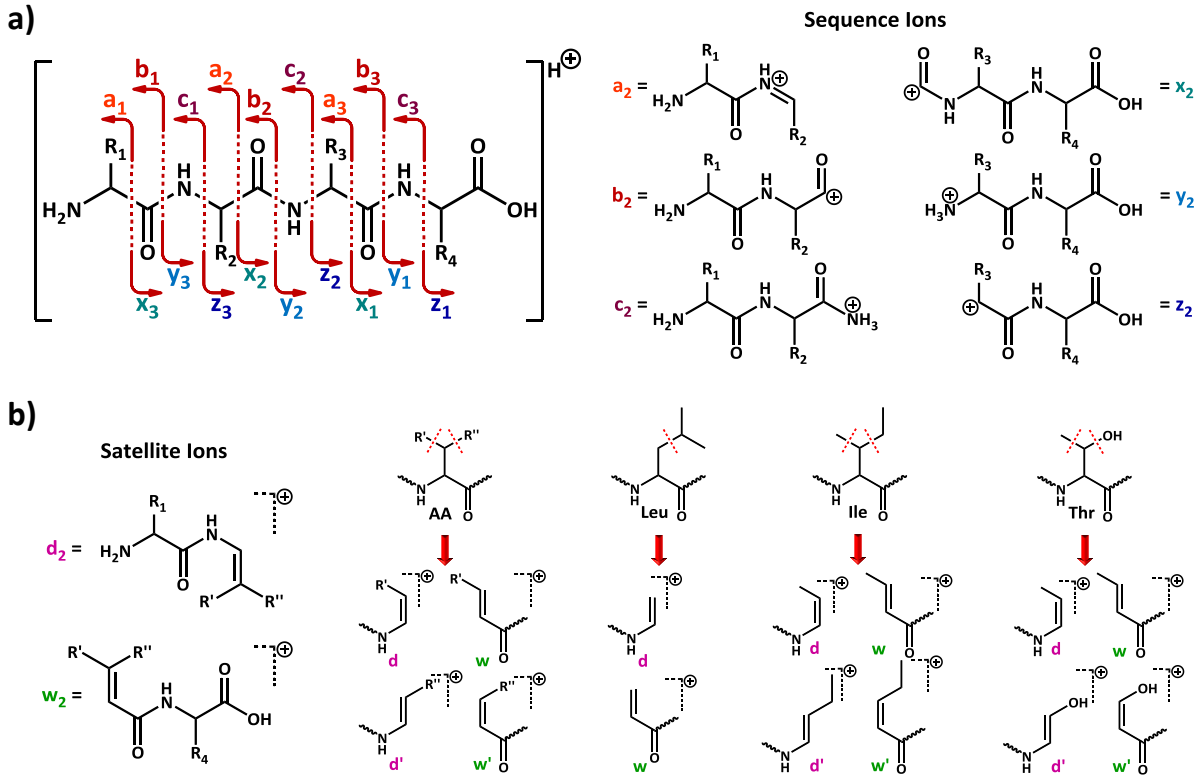


Fig. 3-5 | Standard nomenclature and schematic of the formation of common ions during tandem mass spectrometry (MS/MS) fragmentation of a peptide chain. **a)** Formation of sequence ions. Primarily observed ions are b- and y-type ions derived from fragmenting along the peptide bond. **b)** Formation of satellite bonds derived from partial cleavage of side chains of certain amino acids (AA: any amino acid with different R' and R'').

3.1.1.1 'Bottom-Up' MS/MS for hFX EGF1$_{39mer}$ Hydroxylation Site Identification

The 'bottom-up' MS/MS approach was used to define the hydroxylation site in the hFX EGF1$_{39mer}$ peptide. The peptide was hydroxylated using AspH$_{315-758}$ under the standard hydroxylation assay

conditions. In a standard procedure (see **Chapter 10**), the disulfides were then reduced using the reducing reagent dithiothreitol (DTT), subsequently alkylated using iodoacetamide and digested with trypsin. LC-MS/MS fragmentation data of the hydroxylated hFX EGF1$_{39mer}$ peptide (hFX EGF1(OH)$_{39mer}$) supported that the modification (+16 Da) occurred at Asp18$_{39mer}$ (**Fig. 3-6**). This was consistent with the reported hydroxylation site.[341,356]

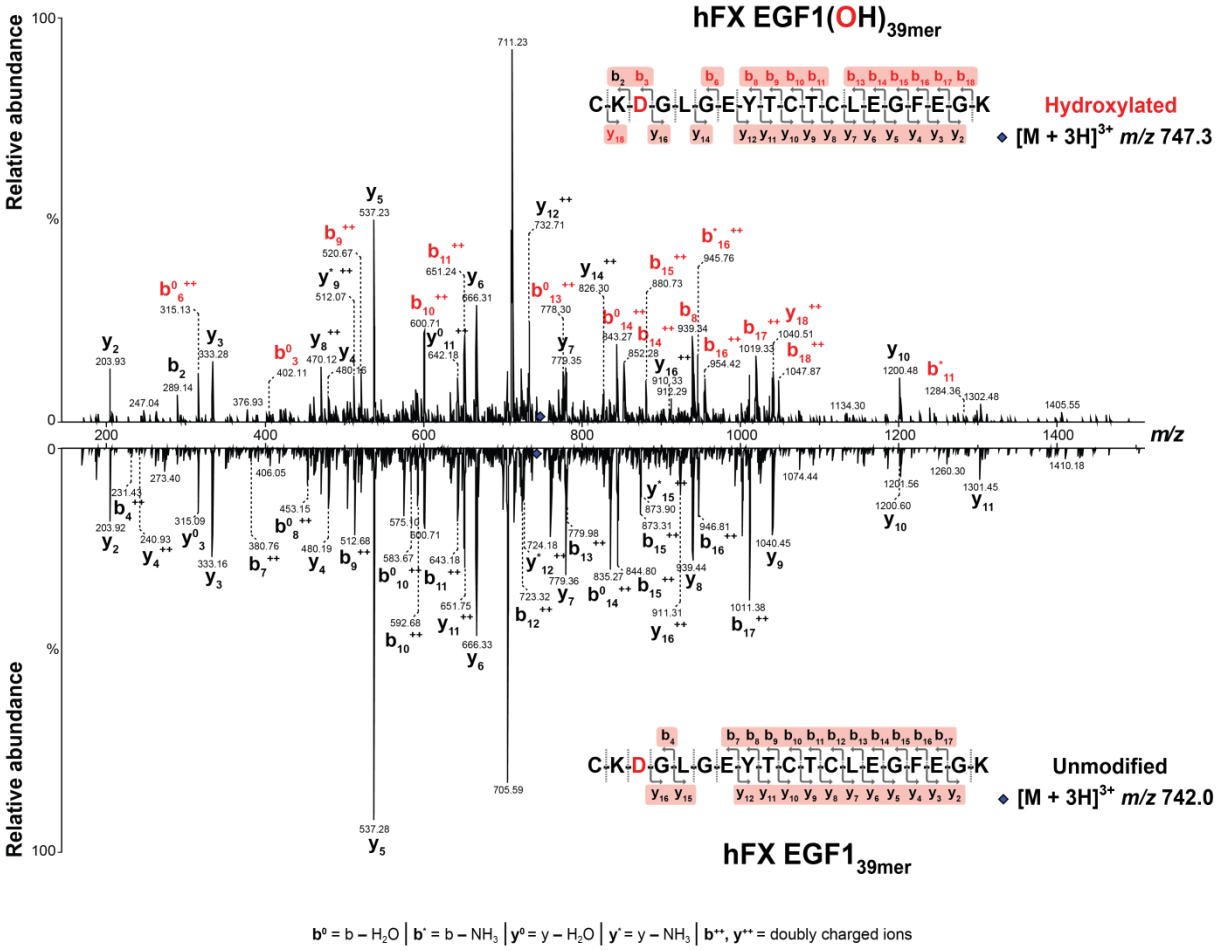


Fig. 3-6 | LC-MS/MS fragmentation of the hydroxylated (by AspH hydroxylation assay) and the unmodified hFX EGF1$_{39mer}$ digested with trypsin ('bottom-up' approach; precursor ions from residues 16-34, *S*-carbamidomethylated [CKD(OH)GLGEYTCTCLEGFEGK + 3H]$^{3+}$ *m/z* 747.3 corresponding to the monoisotopic mass of 2,238.9 Da for the hydroxylated and [CKDGLGEYTCTCLEGFEGK + 3H]$^{3+}$ *m/z* 742.0 corresponding to the monoisotopic mass of 2,222.9 Da for the unmodified peptide). Cysteine residues were reduced using dithiothreitol (DTT) and *S*-carbamidomethylated using iodoacetamide (IAM) prior to digestion. The data support modification at Asp18$_{39mer}$. Ions coloured in red have the modification (+ 16 Da, relative to the unmodified amino acid in the control sample), black coloured ions are unmodified. The data were obtained using a *Bruker Daltonics* amaZon™ Ion Trap LC-MS/MS system.

3.1.1.2 'Top-Down' MS/MS Studies on the hFX EGF1$_{39mer}$ with Intact Disulfides

In a 'top-down' approach, MS/MS fragmentation of the hFX EGF1$_{39mer}$ material with intact disulfides was performed to investigate the disulfide pattern, similar to reported methods.[448-451] Peptide bonds within intra-chain disulfide loops are generally resistant to fragmentation in the gas phase.[449,452] In disulfide-bonded peptides, fragmentation hence typically preferentially occurs in the peptide parts surrounding the disulfide loops.[449]

The commercially obtained purified hFX $EGF1_{39mer}$ material, dissolved in water without any buffer or exposure to AspH, was spotted onto a MALDI target plate (mixed with CHCA MALDI matrix) and then analysed using a *Bruker Daltonics* Ultraflex I™ MALDI-ToF/ToF mass spectrometer. The results suggested that a peptide isoform with non-canonical disulfides in Cys1-2, 3-4, 5-6 was already present in the material dissolved in water (in absence of AspH; **Fig. 3-7**). Characteristic fragmentation ions (e.g. y_{24}-y_{29}) were observed that would not be expected in EGFDs with the crossing disulfides of the canonical Cys1-3, 2-4 linkage. A comparison with the 'top-down' MS/MS data of the LP_{39mer} peptide (reduced cysteines, **Fig. 3-8**) supported the proposal that the observed fragmentation pattern was characteristic of the underlying disulfide linkages.

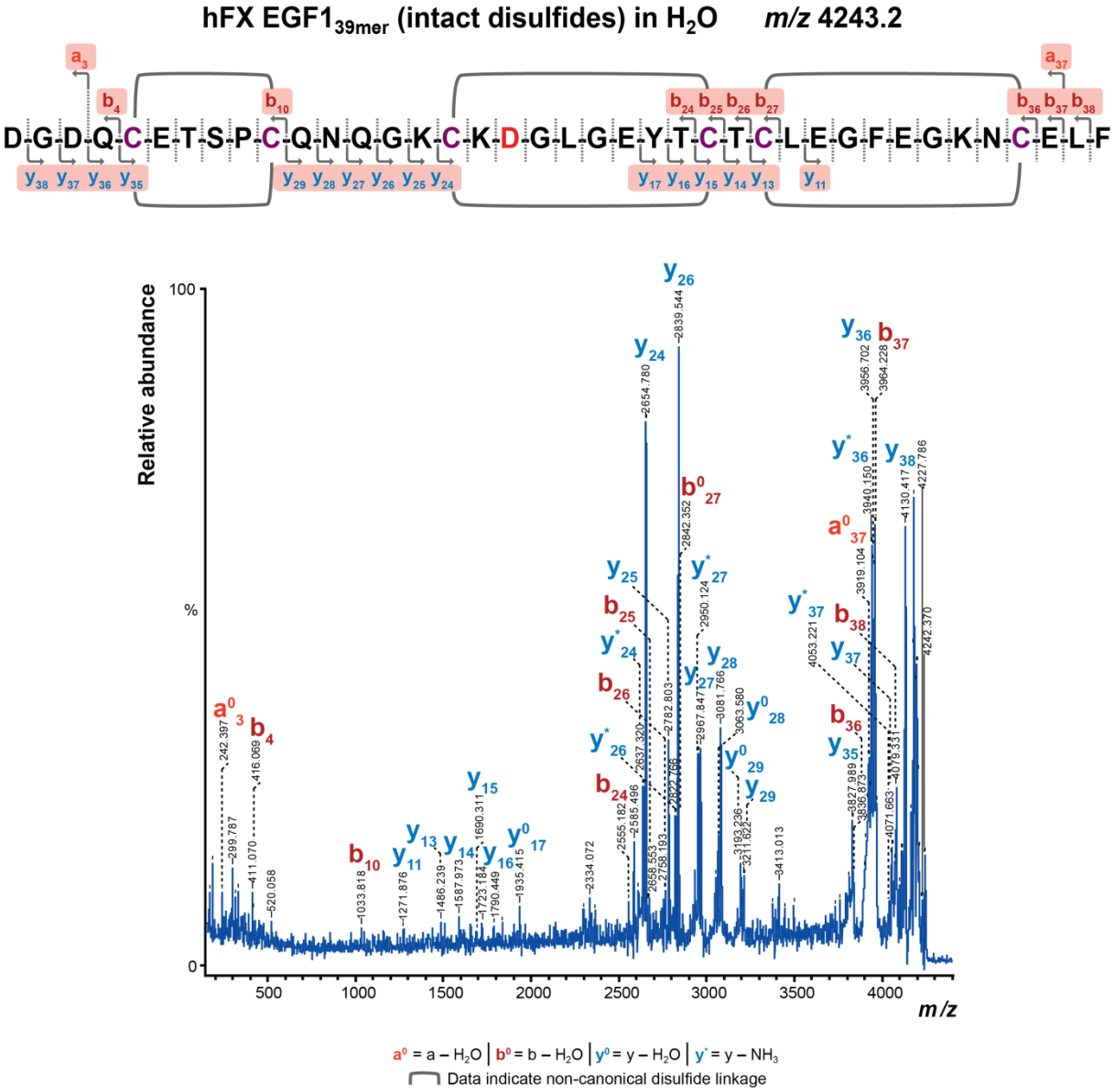


Fig. 3-7 | 'Top-down' approach MALDI-ToF-ToF MS/MS fragmentation data of the hFX $EGF1_{39mer}$ peptide material with intact disulfides in H_2O ($[M+H]^+$ *m/z* 4,243.2). A characteristic fragmentation pattern indicated that a peptide isoform with a non-canonical Cys1-2, 3-4, 5-6 pattern was already present in solution in absence of AspH.

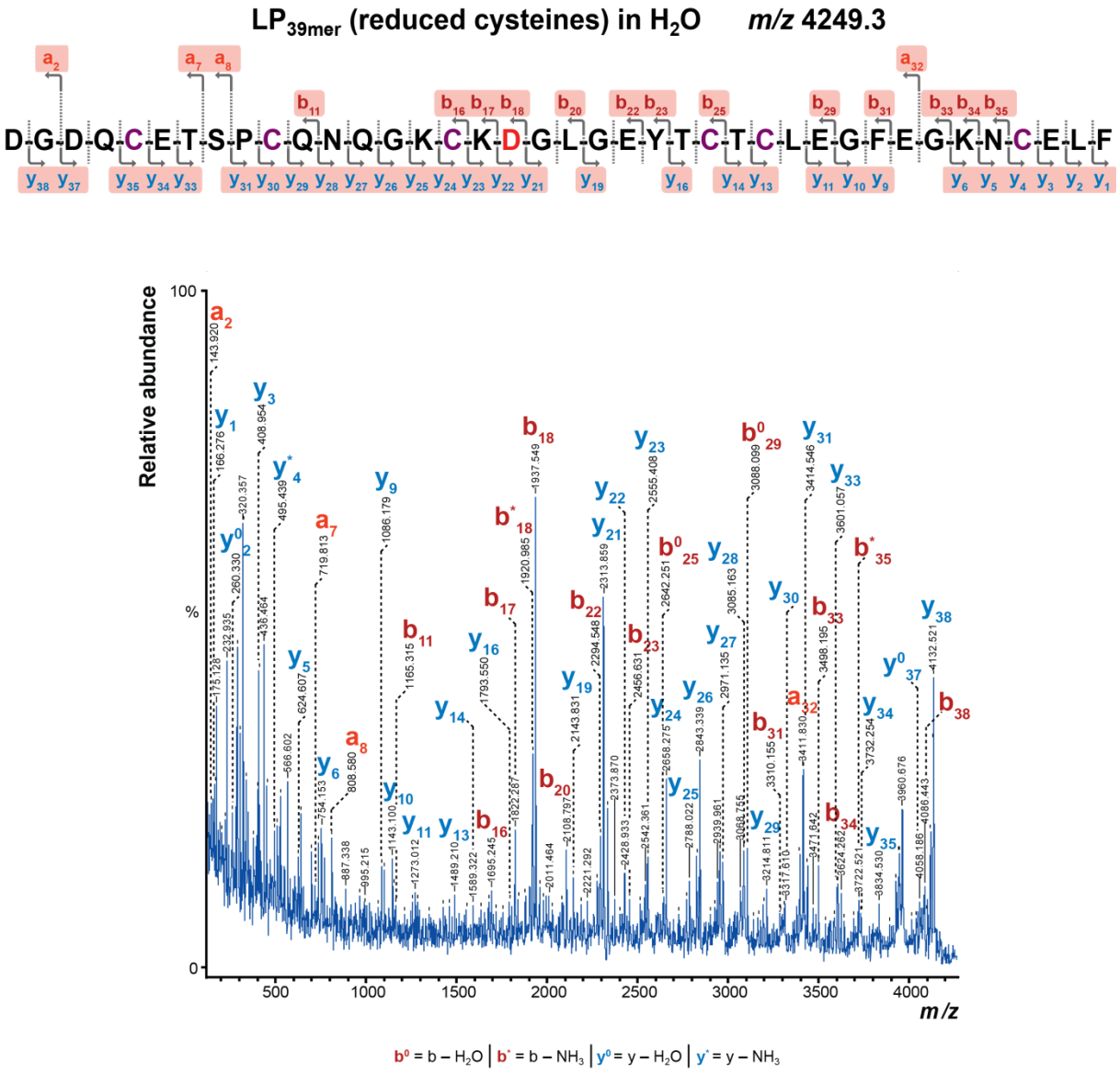


Fig. 3-8 | 'Top-down' approach MALDI-ToF-ToF MS/MS fragmentation data of the LP_{39mer} peptide (reduced cysteines) in H_2O ($[M+H]^+$ *m/z* 4,249.3).

Since the 'top-down' MS/MS fragment ions expected for the hFX $EGF1_{39mer}$ with canonical Cys1-3, 2-4, 5-6 disulfide linkage would be the same as the found ions (**Fig. 3-9**), except the y_{24}-y_{29} ions characteristic of the non-canonical Cys1-2 linkage, the observed fragmentation pattern (**Fig. 3-7**) was possibly the result from a mixture of several disulfide isoforms.

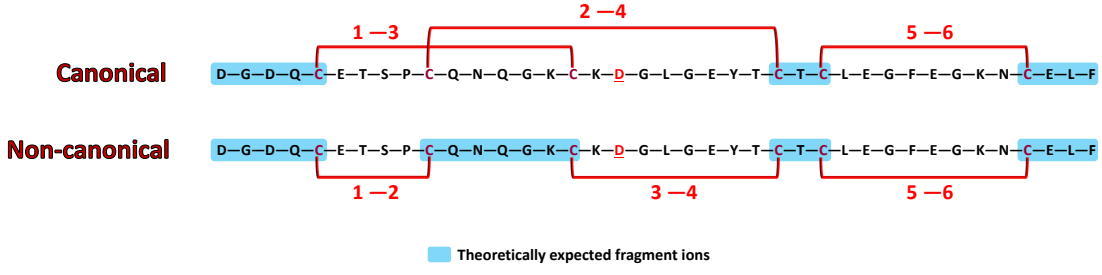


Fig. 3-9 | Schematic of the hFX $EGF1_{39mer}$ peptide with the canonical ("native") Cys1-3, 2-4, 5-6 and the non-canonical Cys1-2, 3-4, 5-6 disulfide linkages. The theoretically expected fragment ions in MS/MS studies on the peptide with intact disulfides are shaded in blue.

3.1.1.3 Trypsin Digestion of the hFX EGF1$_{39mer}$ under Non-Reducing Conditions

The hFX EGF1$_{39mer}$ material was subjected to trypsin digestion under non-reducing conditions (i.e.: with intact disulfides) without reduction and alkylation steps (**Fig. 3-10**) to further investigate its composition.

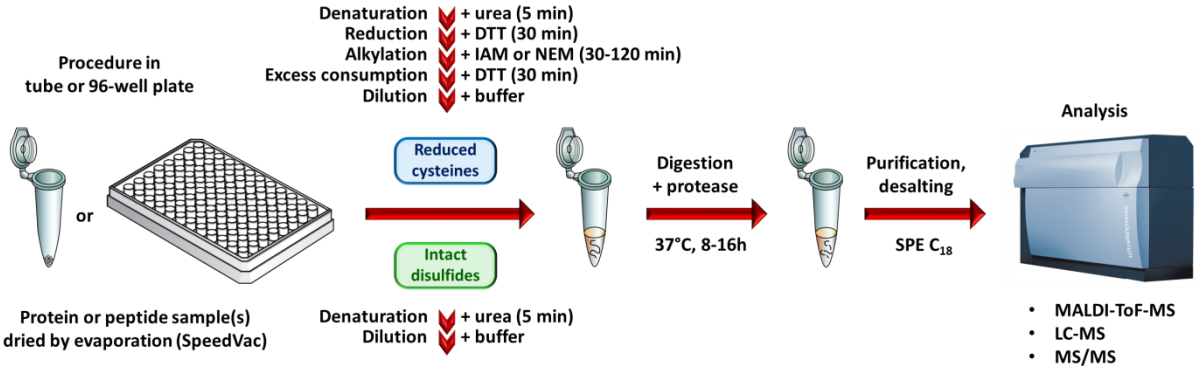


Fig. 3-10 | Flowchart of the in-solution digestion procedures under either reducing (cysteines are first reduced using a reducing reagent such as dithiothreitol (DTT) and subsequently alkylated using an alkylating reagent, iodoacetamide (IAM) or *N*-ethylmaleimide (NEM)) or non-reducing conditions (intact disulfides; no reduction and alkylation steps). The procedure was either performed in individual tubes or in 96-well plates in case of numerous samples. SPE: solid-phase extraction.

Trypsin is a relatively selective protease and usually cleaves to the *C*-terminal side of Lys and Arg residues.[453] However, there are cases where trypsin activity is blocked, for example when a Pro residue follows the Lys/Arg or when the Lys is positioned in a CKD-motif.[454] The hFX EGF1$_{39mer}$ contains three Lys residues, Lys15, Lys17, and Lys34. Digestion of the non-canonical isoform with Cys1-2, 3-4, 5-6 disulfide linkages, as observed in the the 'top-down' MS/MS fragmentation data, was expected to give two peptides as a result of cleavage *C*-terminal to Lys15 (**Fig. 3-11**). A cyclised 15mer with Cys 1-2 disulfide and a 24mer connected by the Cys3-4 and Cys5-6 disulfides were expected. In contrast, the hFX EGF1$_{39mer}$ isoform with canonical Cys1-3, 2-4, 5-6 disulfide linkages was expected to result in a 39mer peptide connected by the disulfides or to be resistant to proteolytic cleavage due to the limited accessibility of the cleavage sites in this folded state. Every cleavage was expected to add a mass of +18 Da to the peptide ion mass for the addition of water to the cleaved peptide bond.

The hFX EGF1$_{39mer}$ material (denatured in 6 M urea) was digested with trypsin for 16 h at 37 °C (in 100 mM phosphate buffer, pH 6.6) and desalted by solid-phase extraction prior to MALDI-ToF-MS analysis.

Consistent with the assumption that Lys15 in the non-canonical hFX EGF1$_{39mer}$ isoform with Cys1-2, 3-4, 5-6 disulfide linkages was readily accessible, the expected resulting 24mer (peptide species **8** and **7**) and 15mer peptides (peptide species **5**) were observed in the MALDI-ToF-MS data (**Fig. 3-12**). In addition, 39mer peptide species (peptides **4** and **3**) were observed, which indicated

that only limited trypsinolysis of these peptide species was possible, as for example expected for the canonical hFX EGF1$_{39mer}$ isoform with Cys1-3, 2-4, 5-6 disulfide linkages.

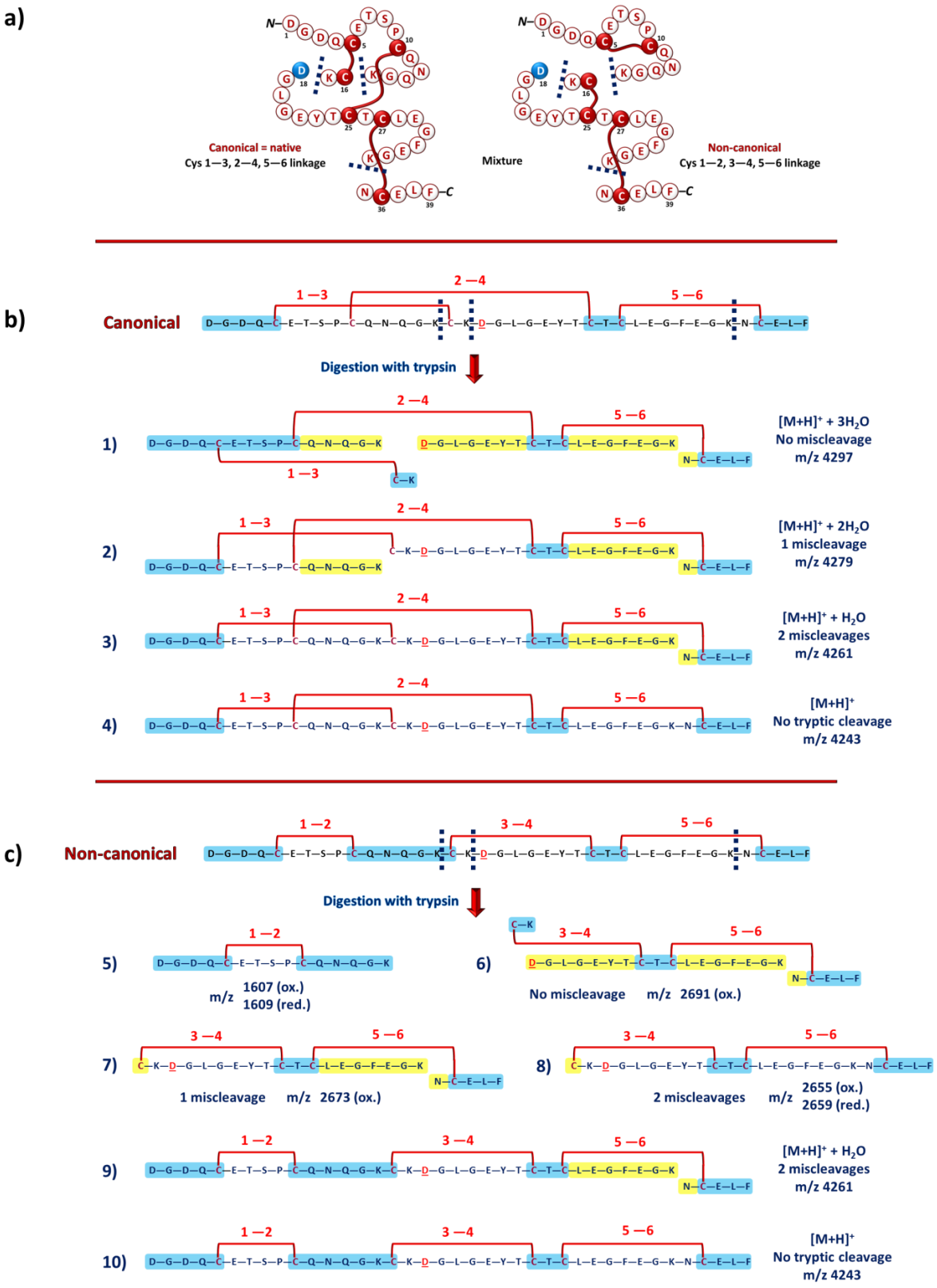


Fig. 3-11 | Schematic strategy for trypsin digestion of the hFX EGF1$_{39mer}$ material with intact disulfides. **a)** Graphical illustration of the theoretical cleavage sites. **b)** Trypsin digestion of the hFX EGF1$_{39mer}$ with canonical Cys1-3, 2-4, 5-6 disulfide linkages will still result in a connected 39mer peptide. **c)** Trypsin digestion of the hFX EGF1$_{39mer}$ with non-canonical Cys1-2, 3-4, 5-6 disulfide linkages, as observed in the 'top-down' MS/MS fragmentation data (**Fig. 3-7**), will result in two peptides, a cyclised 15mer with Cys 1-2 disulfide and a 24mer with peptide fragments connected by the disulfides.

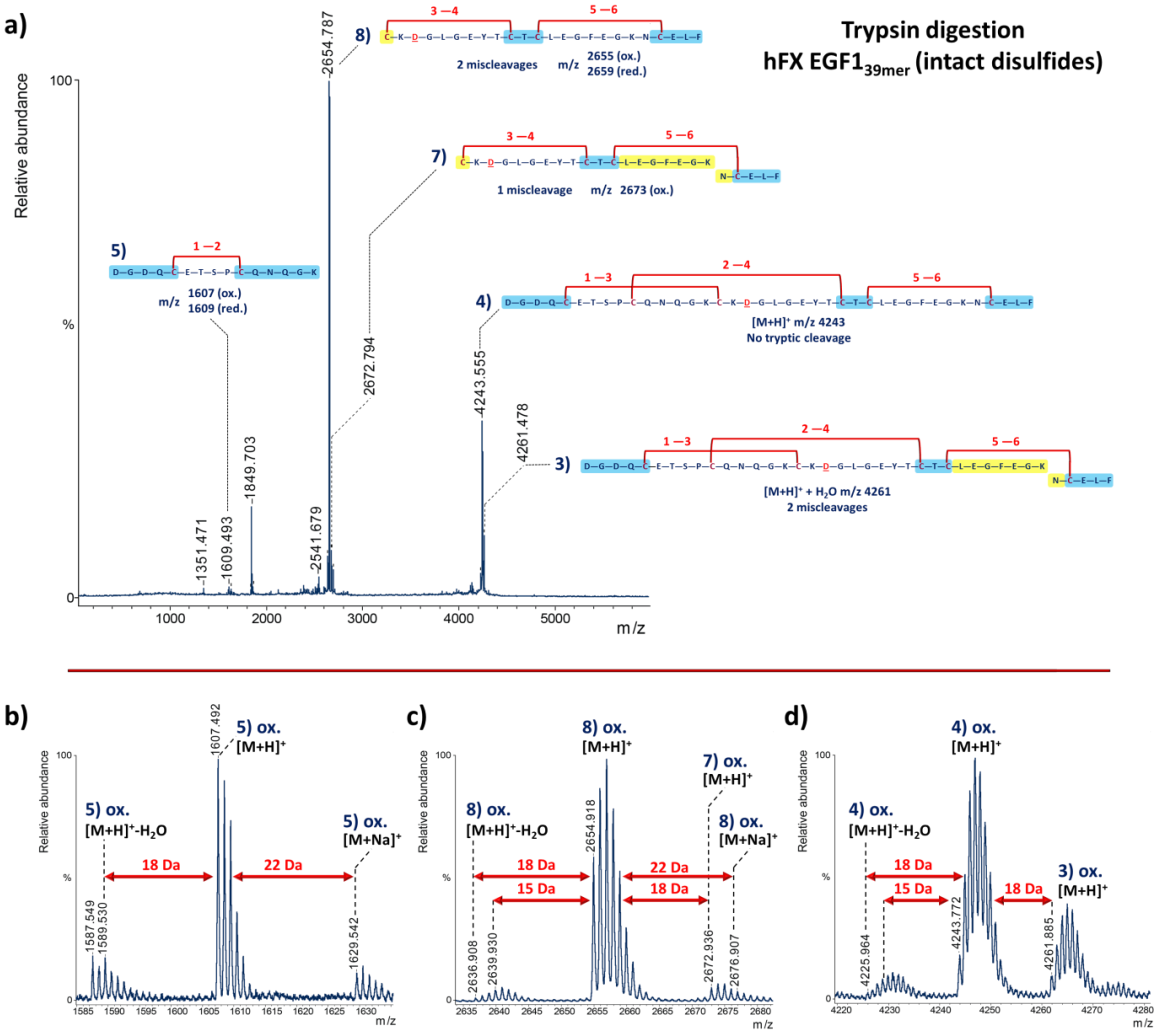


Fig. 3-12 | MALDI-ToF-MS data obtained from trypsin digestion of the hFX EGF1$_{39mer}$ material (with intact disulfides) under non-reducing conditions. **a)** Overview spectrum and assignment to peptides schematically shown in the preceding **Fig. 3-11**. Close-up showing peaks of **b)** a 15mer peptide species ([M+H]$^+$ *m/z* 1,607.5)and **c)** a 24mer peptide species ([M+H]$^+$ *m/z* 2,654.9) resulting from digestion of the hFX EGF1$_{39mer}$ isoform with non-canonical Cys1-2, 3-4, 5-6 disulfide linkages. **d)** Close-up showing peaks of 39mer peptide species ([M+H]$^+$ *m/z* 4,243.8 and 4,261.9), which may be derived from limited digestion of the hFX EGF1$_{39mer}$ with canonical Cys1-3, 2-4, 5-6 disulfide linkages.

The observed peptides differed strongly in their ionisation properties. Although the same quantities of the 15mer and the 24mer peptides derived from the non-canonical hFX EGF1$_{39mer}$ isoform with Cys1-2, 3-4, 5-6 disulfide linkages must have been generated, the 24mer showed peaks with very high intensity, whereas the signal-to-noise ratio of the 15mer species was comparatively low. Ionisation in positive ion mode is based on protonation of residues, and therefore mostly driven by the basicity of the peptide due to presence of Lys and Arg residues and the *N*-terminus.[455] The presence of certain amino acids such as Arg, Phe, Pro, and Leu has been reported to result in higher MALDI-ToF-MS peak intensities due to enhanced desorption/ionisation processes.[456] Guanidination of the Lys residues in the sample using *O*-methyl*iso*urea to give homoarginine residues[457-459] improved the ionisation of the 15mer peptide (species **5**), so that even comparable peak intensities of the 15mer and the 24mer peptides were observed, and the MS data quality was enhanced (**Fig. 3-13**).

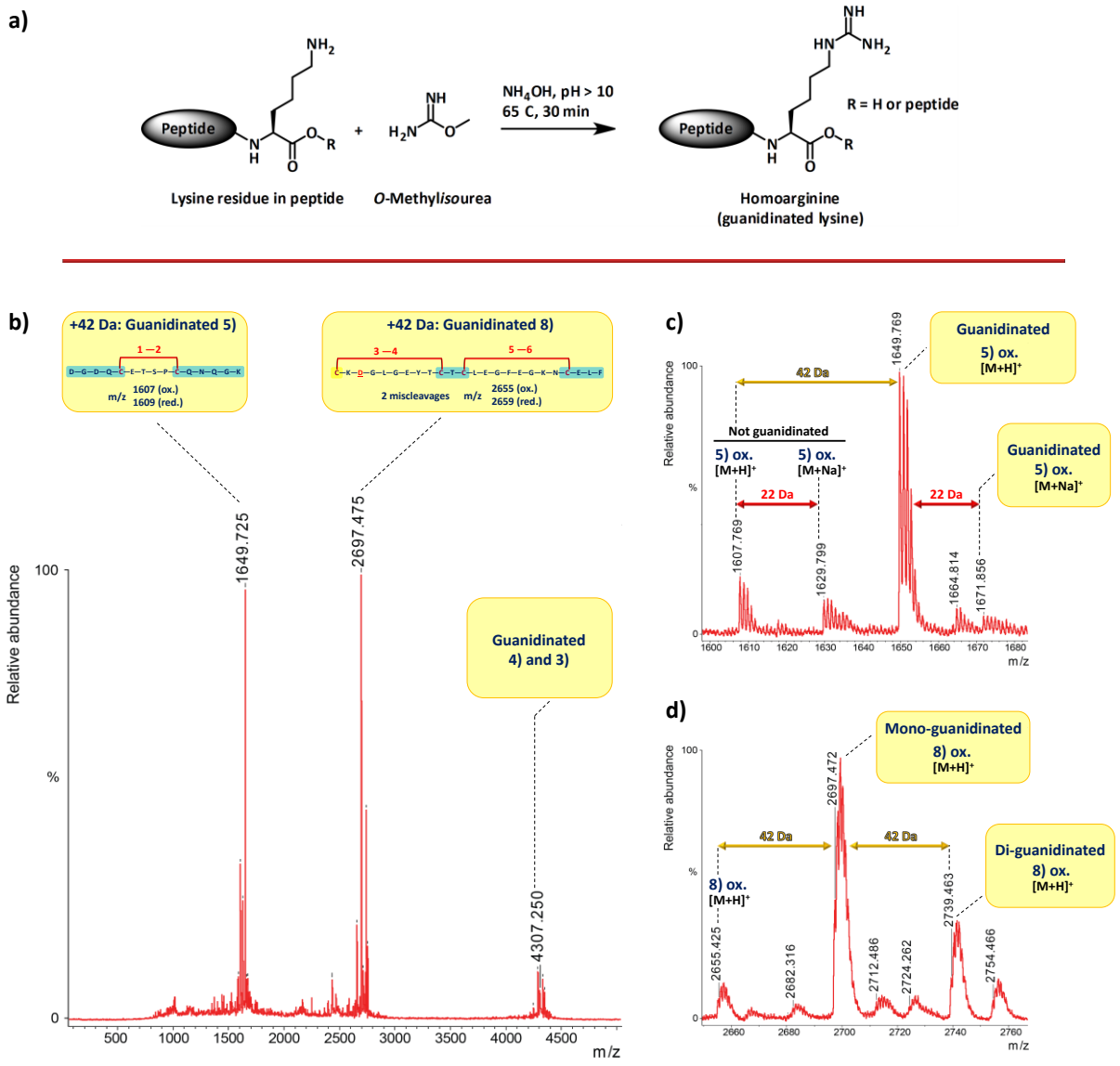


Fig. 3-13 | Guanidination of lysine residues using *O*-methyl*iso*urea enhanced the quality of the MS data by improving the ionisation properties of the peptides. **a)** Guanidination reaction scheme: Lysine residues (*C*-terminal: R = H, or within peptide resulting from tryptic miscleavage: R = peptide backbone) react with *O*-methyl*iso*urea to give homoarginine residues. MALDI-ToF-MS data of guanidinated peptides obtained from trypsin digestion of the hFX EGF1$_{39mer}$ material. **b)** Overview spectrum. **c)** Close-up showing peaks of guanidinated (Lys15) 15mer peptide ([M+H]$^+$ *m/z* 1,649.8). **d)** Close-up showing peaks of mono- and di-guanidinated (Lys17, Lys34) 24mer peptide species ([M+H]$^+$ *m/z* 2,697.5 and 2,739.5).

3.2 STUDIES ON ASPH SUBSTRATE REQUIREMENTS

3.2.1 Experiments with Canonical cbEGFD Proteins

To investigate if other EGFDs with well-characterised disulfide patterns were accepted as AspH substrates *in vitro*, recombinant EGFD proteins comprising several Ca^{2+}-binding (cb) EGFD repeats (~10-15 kDa) were tested in hydroxylation assays with AspH. The human Fibrillin-1 (hFib1) cbEGF32-33[460,349] pair (**Fig. 3-14a**) and the triple repeat cbEGF41-43,[461] the human Notch-1 (hNotch1) triple repeat cbEGF11-13[462,463] (**Fig. 3-14b**), and the hFib1 TB4cbEGF23[464,350] protein comprising a

cbEGFD preceded by a transforming growth factor β-binding protein-like (TB) domain (**Fig. 3-15**) have previously been shown to bind Ca^{2+} and to have canonical disulfides in the Cys1-3, 2-4, 5-6 pattern. These proteins were kindly provided by Professor Penny A. Handford, Department of Biochemistry, University of Oxford.

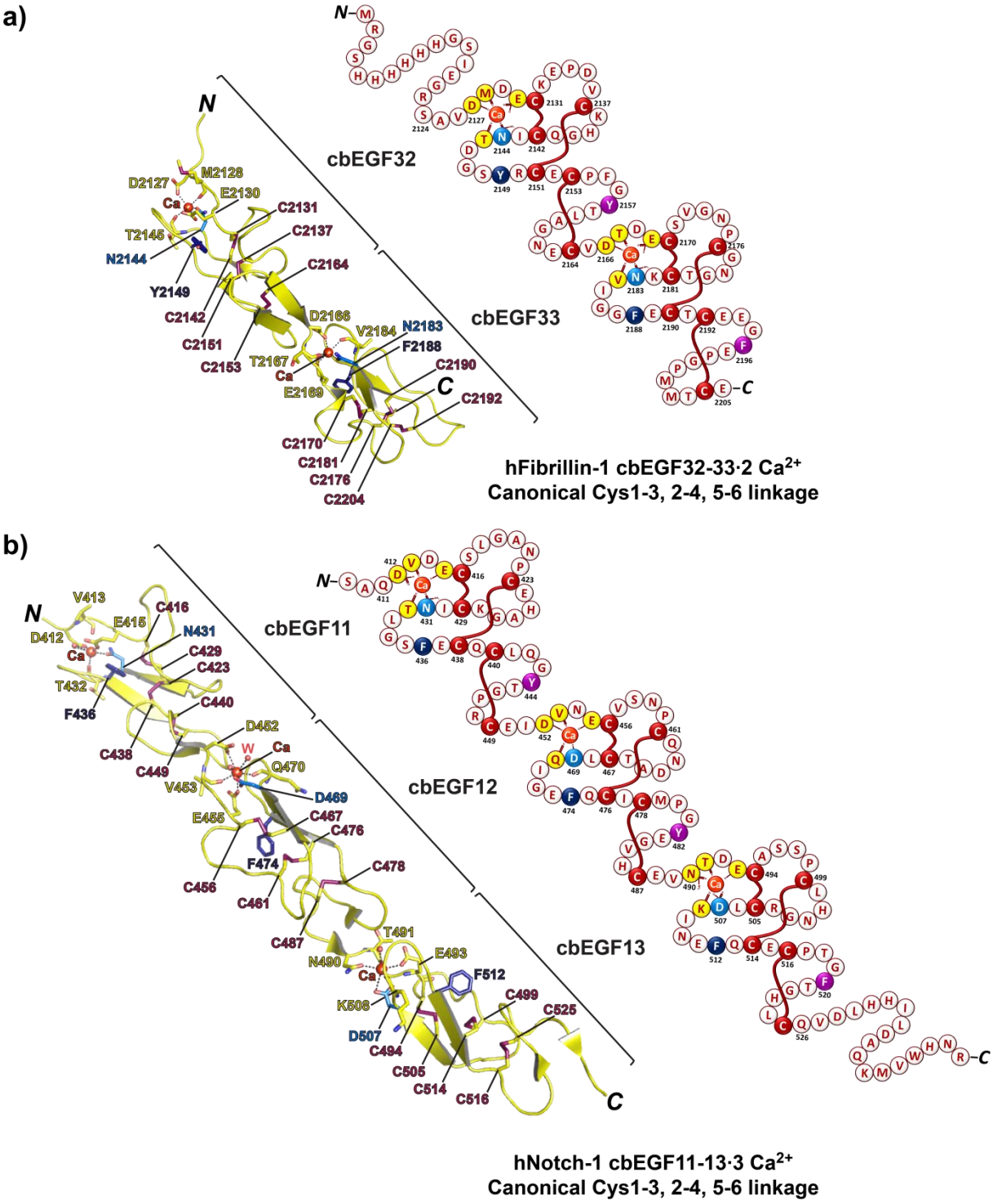


Fig. 3-14 | Structures and schematics of EGFD repeats. The structurally characterised EGFDs have canonical disulfides linked in the Cys1-3, 2-4, 5-6 pattern. **a)** A cbEGFD pair: human Fibrillin-1 cbEGF32-33 solution NMR structure (PDB: 1EMO[460,349]). **b)** A triple EGFD repeat: human Notch-1 cbEGF11-13 X-ray crystal structure (resolution: 2.60 Å; PDB: 2VJ3[462]).

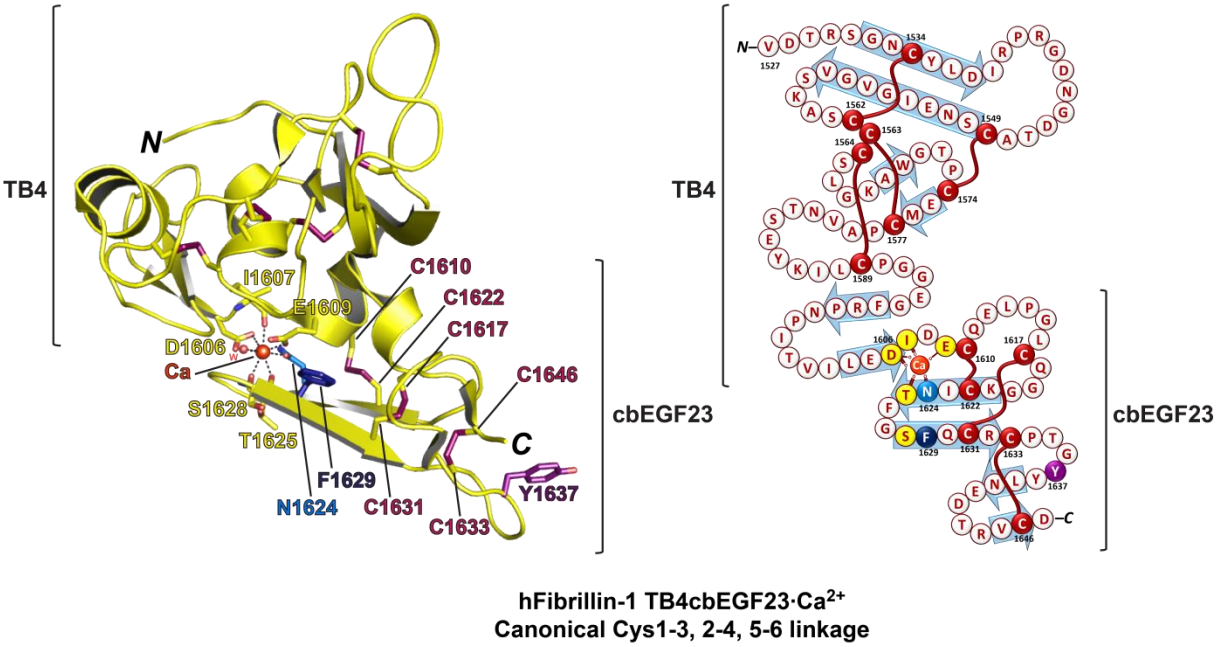


Fig. 3-15 | X-ray crystal structure (resolution: 2.25 Å; PDB: 1UJZ[464]) and schematic of human Fibrillin-1 TB4cbEGF23 comprising a cbEGFD preceded by a transforming growth factor *β*-binding protein-like (TB) domain. The cbEGF23 has canonical disulfides linked in the Cys1-3, 2-4, 5-6 pattern.

3.2.1.1 Hydroxylation Assays, Trypsin Digestion and MS/MS

$AspH_{315-758}$ hydroxylation assays with the cbEGFD proteins were performed under standard conditions with longer incubation times (180 min at 37 °C). For the hydroxylation product analysis using MS, the cbEGFD proteins required proteolytic digestion (using trypsin) and MS/MS measurements ('bottom-up' approach), because other methods such as MALDI-ToF-MS acquired in linear mode or denaturing protein LC-MS were not suitable for discrimination between hydroxylation and methionine oxidation (examples in **Fig. 3-16**).

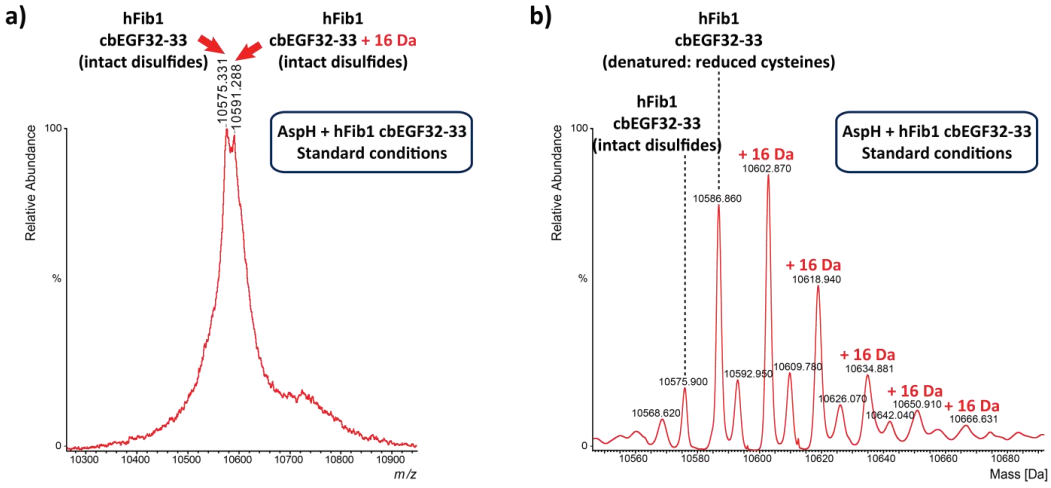


Fig. 3-16 | MS analysis of $AspH_{315-758}$-incubated cbEGFD protein samples, e.g. hFib1 cbEGF32-33. **a)** Intact protein MALDI-ToF-MS acquired in linear mode and **b)** denaturing protein LC-MS (represented as deconvoluted mass spectrum) were not suitable for hydroxylation detection because no discrimination between hydroxylation (mass addition of + 16 Da) and methionine oxidation (also corresponding to a mass addition of + 16 Da) was possible. Trypsin digestion and MS/MS measurements were required for appropriate analysis of the hydroxylated product.

Following the reaction quenching, the proteins in the hydroxylation assay mixtures were subjected to the standard in-solution digestion procedure under reducing conditions (see **Fig. 3-10**; denaturation using 6 M urea, disulfides reduction using DTT and alkylation using iodoacetamide to obtain *S*-carbamidomethylated cysteines, subsequent digestion using trypsin).

The MS analysis results revealed that all tested cbEGFD proteins were AspH substrates; however, only low hydroxylation levels (approximately 2-24 %; **Figs. 3-17**, **3-19**, **3-20** and **3-22**) were obtained under the standard conditions. The MS/MS analyses supported the proposal that the cbEGFD protein hydroxylation occurrs on a specific Asn residue in a Cys-Xaa-Asn motif between the third and fourth cysteine within one EGFD (**Figs. 3-18**, **3-21** and **3-23**).

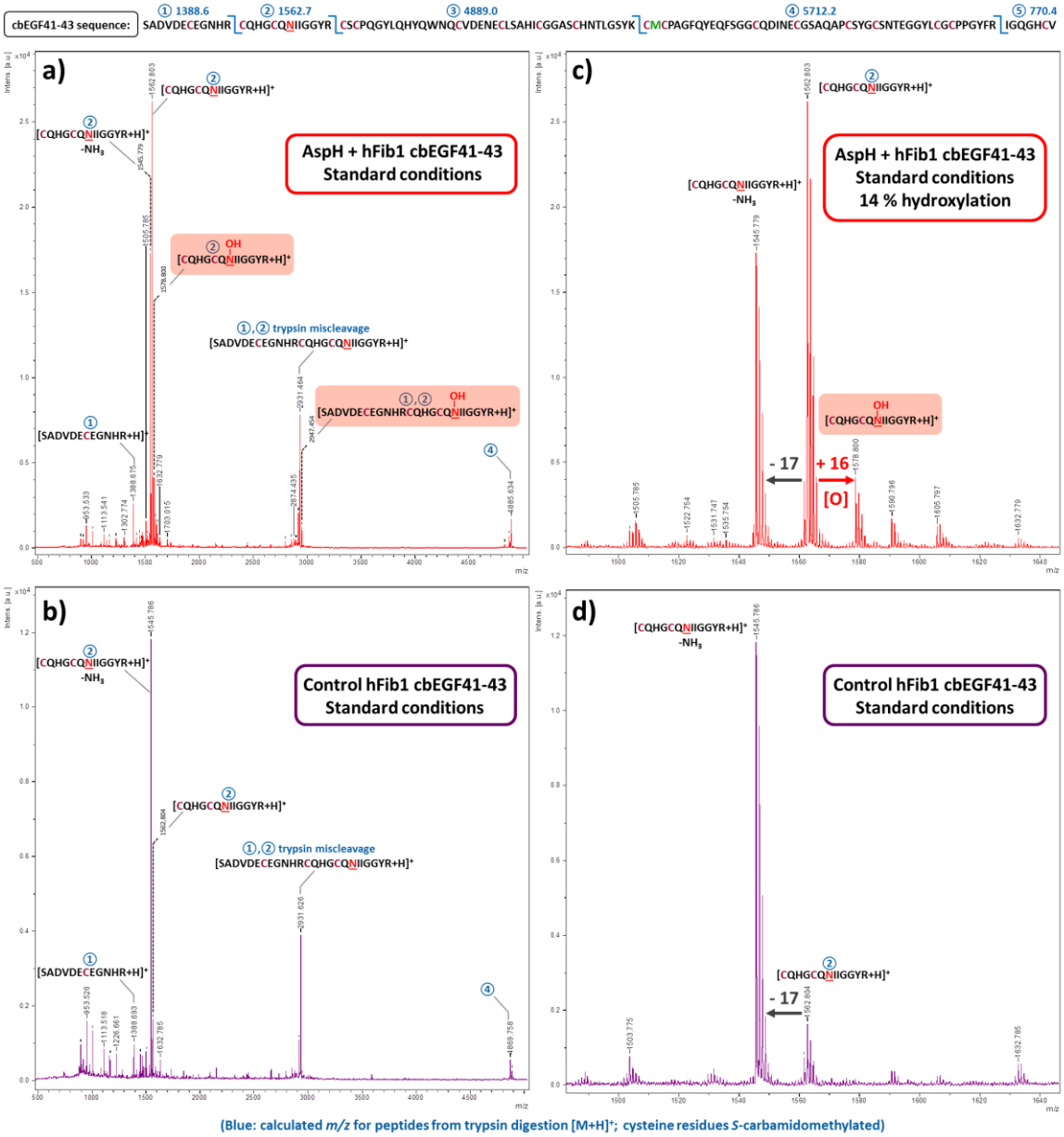


Fig. 3-17 | MALDI-ToF-MS results showing hFib1 cbEGF41-43 hydroxylation by $AspH_{315-758}$ in the cbEGF41 domain peptide ② CQHGCQNILGGYR (residues 2577-2589; $[M(OH) + H]^+$ *m/z* 1,5778.8; AspH to cbEGF41-43 ratio 1:25; incubation for 180 min at 37 °C). **a)** Overview spectrum of the digested $AspH_{315-758}$-incubated sample and **b)** the control sample (no $AspH_{315-758}$ incubation). Close-up of peptide ② in **c)** the $AspH_{315-758}$-incubated sample and **d)** the control sample. Prior to the analysis, hFib1 cbEGF41-43 was reduced, *S*-carbamidomethylated and digested using trypsin. Ammonia was lost from the *N*-terminal *S*-carbamidomethylated cysteine upon cyclisation (-17 Da peaks).

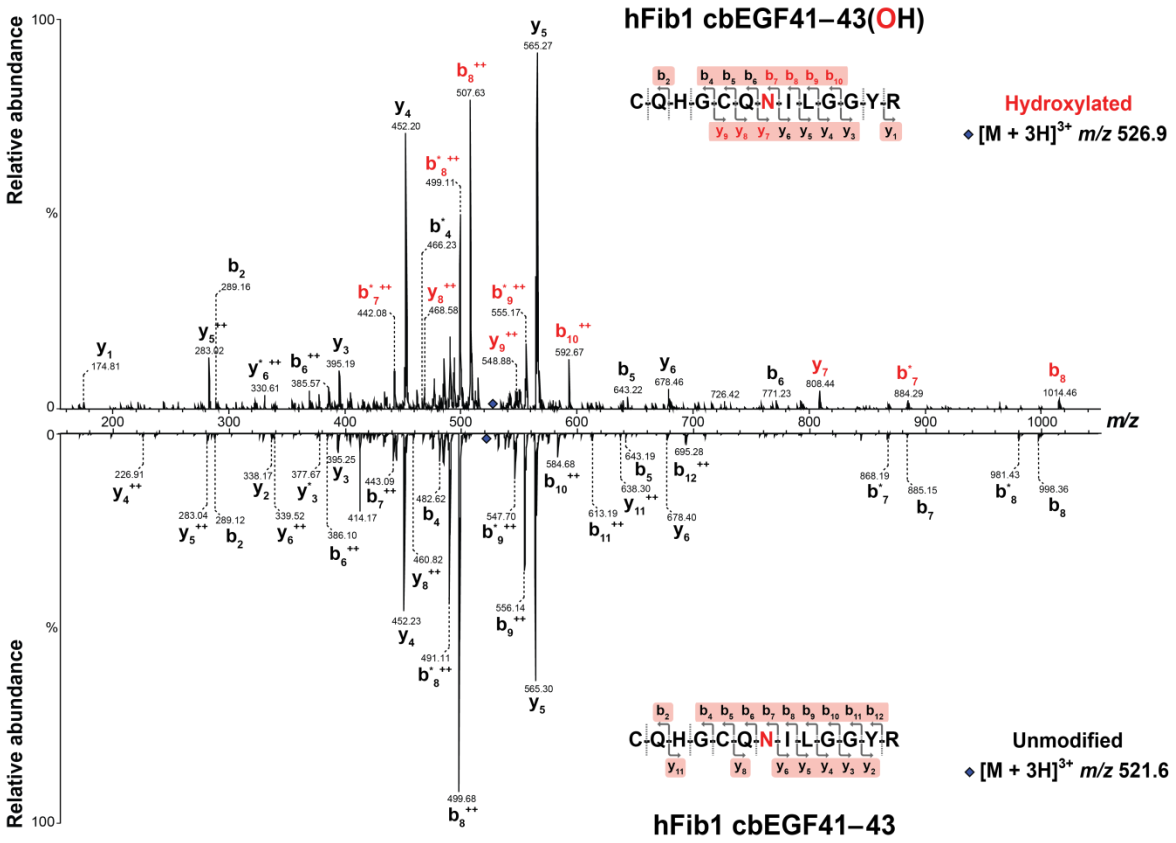


Fig. 3-18 | LC-MS/MS fragmentation of the hydroxylated (by the AspH hydroxylation assay) and the unmodified hFib1 cbEGF41-43 digested with trypsin ('bottom-up' approach; precursor ions from residues 2577-2589, *S*-carbamidomethylated [CQHGCQN(OH)ILGGYR + 3H]$^{3+}$ *m/z* 526.9 corresponding to the monoisotopic mass of 1,577.7 Da for the hydroxylated and [CQHGCQNILGGYR + 3H]$^{3+}$ *m/z* 521.6 corresponding to the monoisotopic mass of 1,561.7 Da for the unmodified peptide). Cysteine residues were reduced using dithiothreitol (DTT) and *S*-carbamidomethylated using iodoacetamide (IAM) prior to digestion. The data support modification at Asn2583. Ions coloured in red have the modification (+ 16 Da, relative to the unmodified amino acid in the control sample), black coloured ions are unmodified. The data were obtained using a *Bruker Daltonics* amaZon™ Ion Trap LC-MS/MS system.

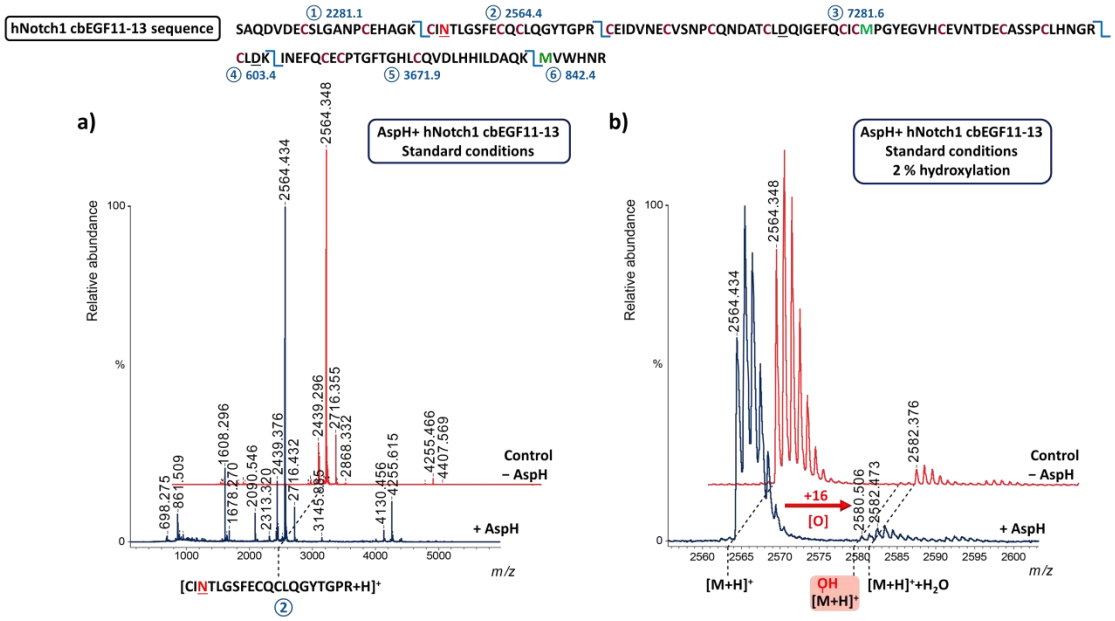


Fig. 3-19 | MALDI-ToF-MS results showing hNotch1 cbEGF11-13 hydroxylation by AspH$_{315\text{-}758}$ in the cbEGF11 domain peptide ② CINTLGSFECQCLQGYTGPR (residues 429-448; [M(OH) + H]$^{+}$ *m/z* 2,580.5; AspH to cbEGF11-13 ratio 1:10; incubation for 180 min at 37 °C). **a)** Overview spectrum of the digested sample. **b)** Close-up of peptide ②. Prior to the analysis, hNotch1 cbEGF11-13 was reduced, alkylated using *N*-ethylmaleimide (NEM) and digested using trypsin.

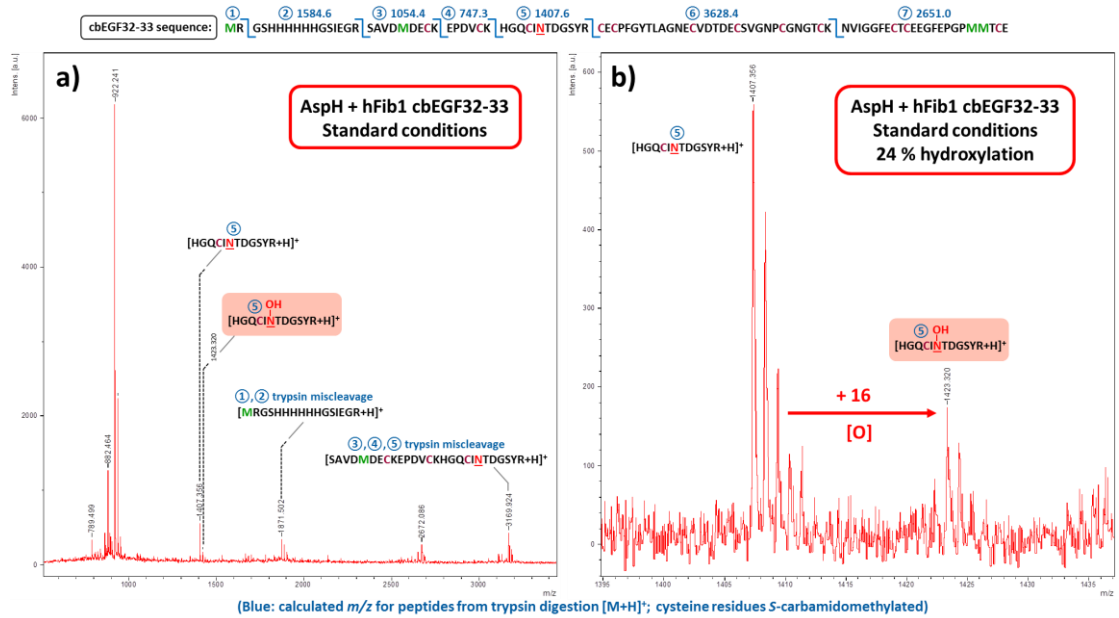


Fig. 3-20 | MALDI-ToF-MS results showing hFib1 cbEGF32-33 hydroxylation by AspH$_{315\text{-}758}$ in the cbEGF32 domain peptide ⑤ HGQCINTDGSYR (residues 2139-2150; [M(OH) + H]$^+$ *m/z* 1,423.3; AspH to cbEGF32-33 ratio 1:25; incubation for 180 min at 37 °C). **a)** Overview spectrum of the digested sample. **b)** Close-up of peptide ⑤. Prior to the analysis, hFib1 cbEGF32-33 was reduced, *S*-carbamidomethylated and digested using trypsin.

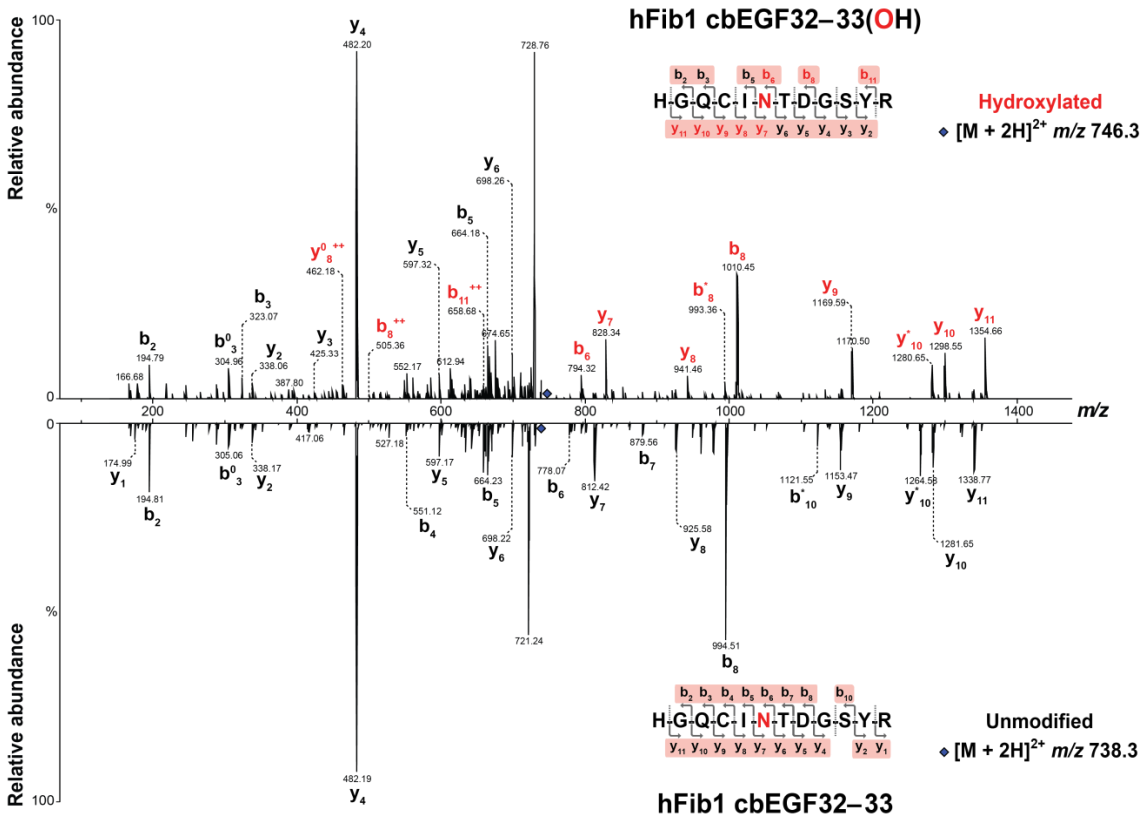


Fig. 3-21 | LC-MS/MS fragmentation of the hydroxylated (by the AspH hydroxylation assay) and the unmodified hFib1 cbEGF32-33 digested using trypsin ('bottom-up' approach; precursor ions from residues 2139-2150, [HGQCIN(OH)TDGSYR + 2H]$^{2+}$ *m/z* 746.3 corresponding to the monoisotopic mass of 1,490.6 Da for the hydroxylated and [HGQCINTDGSYR + 2H]$^{2+}$ *m/z* 738.3 corresponding to the monoisotopic mass of 1,474.6 Da for the unmodified peptide, *S*-(*N*-ethylsuccinimido)-modified Cys residues). The cysteine residues were reduced using dithiothreitol (DTT) and alkylated using *N*-ethylmaleimide (NEM) prior to digestion. The data support modification at Asn2144. Ions coloured in red have the modification (+ 16 Da, relative to the unmodified amino acid in the control sample), black coloured ions are unmodified. The data were obtained using a *Bruker Daltonics* amaZon™ Ion Trap LC-MS/MS system.

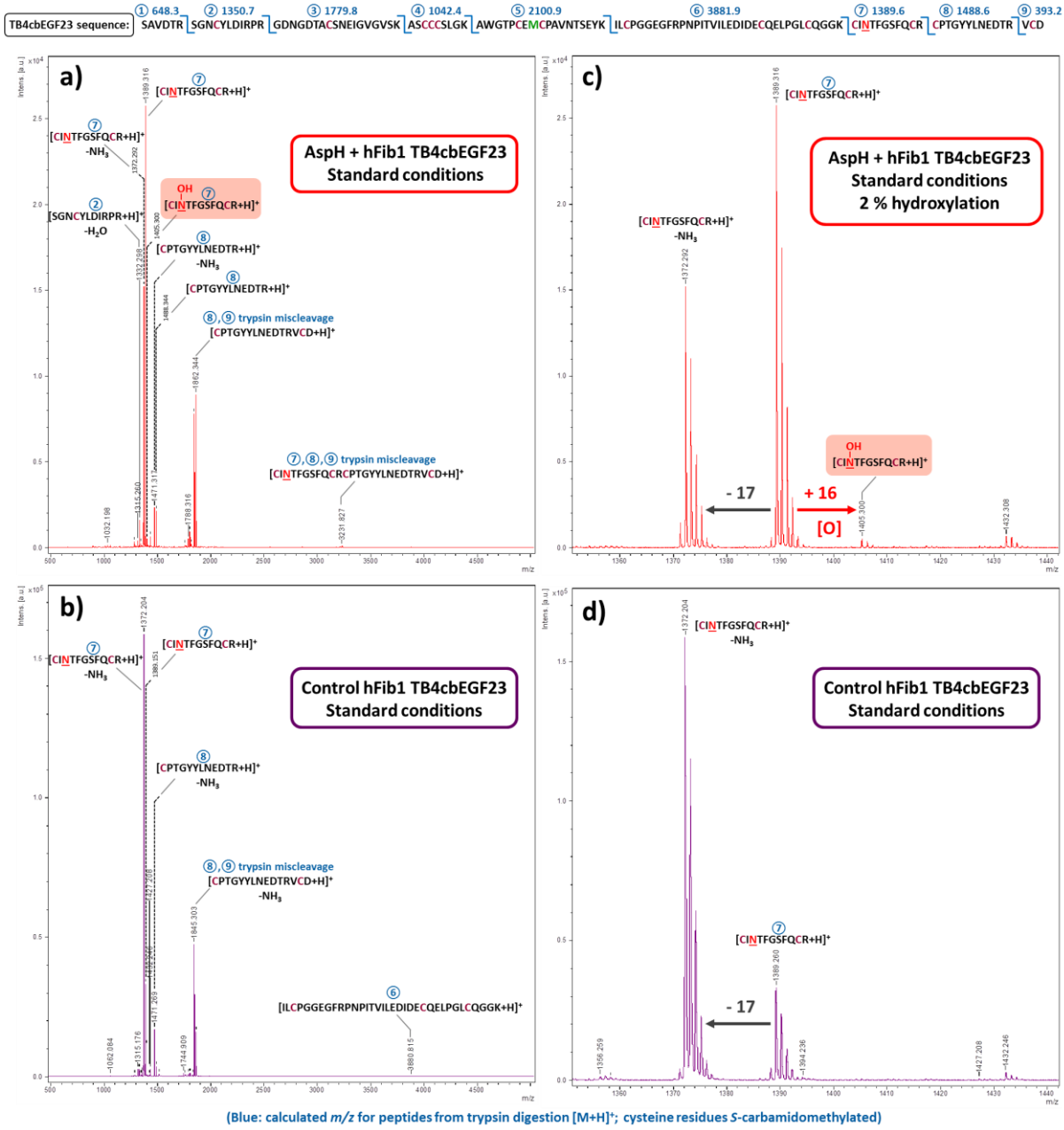


Fig. 3-22 | MALDI-ToF-MS results showing hFib1 TB4cbEGF23 hydroxylation by $AspH_{315-758}$ in the cbEGF23 domain peptide ⑦ CINTFGSFQCR (residues 1622-1632, $[M(OH) + H]^+$ *m/z* 1,405.3; AspH to TB4cbEGF23 ratio 1:25; incubation for 180 min at 37 °C). **a)** Overview spectrum of the digested $AspH_{315-758}$-incubated sample and **b)** the control sample (no $AspH_{315-758}$ incubation). Close-up of peptide ⑦ in **c)** the $AspH_{315-758}$-incubated sample and **d)** the control sample. Prior to the analysis, hFib1 TB4cbEGF23 was reduced, *S*-carbamidomethylated and digested using trypsin. An ammonia molecule was lost from the *N*-terminal *S*-carbamidomethylated cysteine upon cyclisation (-17 Da peaks).

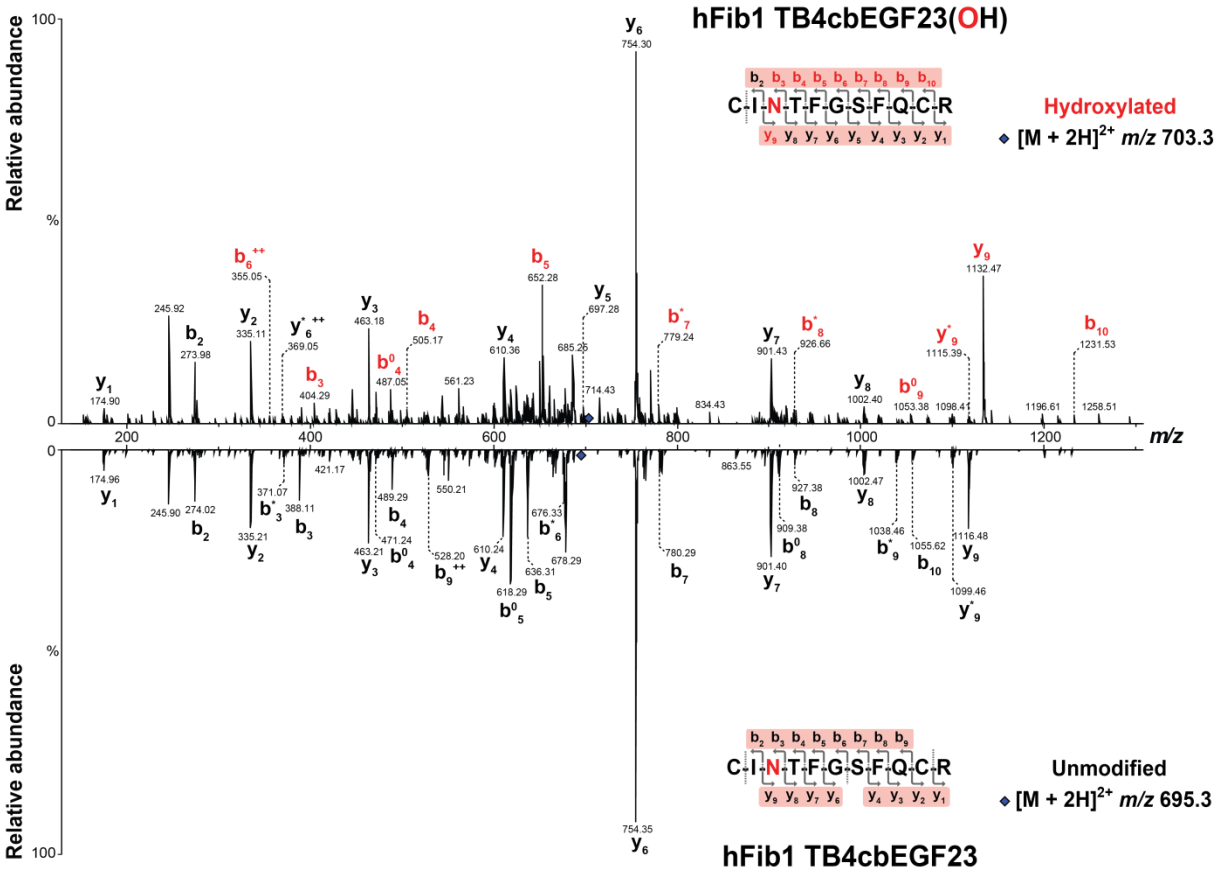


Fig. 3-23 | LC-MS/MS fragmentation of the hydroxylated (by the AspH hydroxylation assay) and the unmodified hFib1 TB4cbEGF23 digested with trypsin ('bottom-up' approach; precursor ions from residues 1622-1632, *S*-carbamidomethylated [CIN(OH)TFGSFQCR + 2H]$^{2+}$ *m/z* 703.3 corresponding to the monoisotopic mass of 1,404.6 Da for the hydroxylated and [CINTFGSFQCR + 2H]$^{2+}$ *m/z* 695.3 corresponding to the monoisotopic mass of 1,388.6 Da for the unmodified peptide). Cysteine residues were reduced using dithiothreitol (DTT) and *S*-carbamidomethylated using iodoacetamide (IAM) prior to digestion. The results support modification at Asn1624. Ions coloured in red have the modification (+ 16 Da, relative to the unmodified amino acid in the control sample), black coloured ions are unmodified. The data were obtained on a *Bruker Daltonics* amaZon™ Ion Trap LC-MS/MS system.

As visible in **Figs. 3-17** and **3-22**, ammonia was lost (-17 Da peaks) from the *N*-terminal *S*-carbamidomethylated cysteine (Cys^{CAM}), likely upon cyclisation to give a six-membered ring by intramolecular attack from the *N*-terminal amino group on the carbonyl of the carbamidomethyl group (**Fig. 3-24a**).[465,466,467] The reaction is similar to the commonly occurring cyclisation of an *N*-terminal glutamine residue to give a pyroglutamic acid residue (**Fig. 3-24b**).[465] To circumvent the undesirable *N*-terminal Cys^{CAM} cyclisation, *N*-ethylmaleimide (NEM, **Fig. 3-24c**) was used as cysteine alkylating reagent in further studies instead of iodoacetamide, since *S*-(*N*-ethylsuccinimido)-cysteine (Cys^{NES}) cannot undergo such a cyclisation.

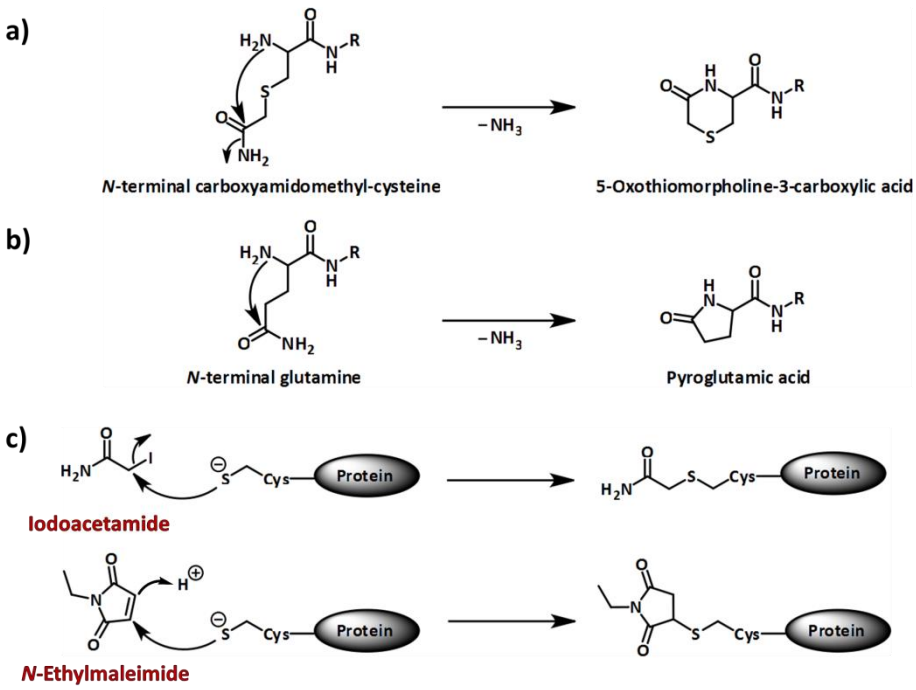


Fig. 3-24 | Common cyclisation reactions of *N*-terminal residues in peptides. **a)** Cyclisation of *N*-terminal cysteine residue alkylated with iodoacetamide; **b)** cyclisation of *N*-terminal glutamine to pyroglutamic acid. Figure adapted from Krokhin *et al.*[465]. **c)** Cysteine alkylation using the alkylating reagents iodoacetamide or *N*-ethylmaleimide.

3.2.2 AspH Cosubstrate and Cofactor Dependence

Studies on the $AspH_{315\text{-}758}$ substrate requirements and optimal assay conditions were then performed. Hydroxylation assays conducted to determine the cosubstrate/cofactor dependence *in vitro* showed that no reaction occurred when 2OG was omitted from the assay mixture (**Fig. 3-25**). $AspH_{315\text{-}758}$ strictly required the addition of 2OG for functioning. Externally added Fe(II) was not a requirement for hydroxylation activity, suggesting that $AspH_{315\text{-}758}$ may partially copurify with Fe(II). Ascorbate addition was not required but enhanced the activity of $AspH_{315\text{-}758}$ in the *in vitro* assays.

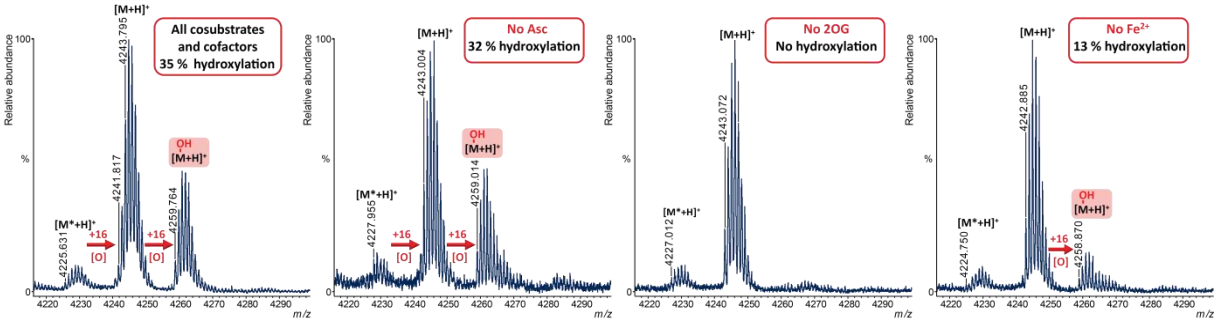


Fig. 3-25 | Representative MALDI-ToF-MS spectra obtained using $AspH_{315\text{-}758}$ for hFX $EGF1_{39mer}$ (1:25 ratio) hydroxylation under typical conditions in presence of all or absence of the stated cosubstrates/cofactors quenched after 30 min incubation at 37 °C.

3.2.3 Experiments with Different Buffer Systems

Only partial hFX EGF1$_{39mer}$ hydroxylation (< 40 %) was observed under the standard assay conditions; and no hydroxylation increase was attained upon subsequent addition of another aliquot of AspH$_{315-758}$ and cosubstrates/cofactors. Interestingly, the hydroxylation levels were strongly enhanced (> 90 %; **Fig. 3-26**) when the assays were performed in a buffer system containing reduced and oxidised glutathione at alkaline pH (GSH and GSSG; pH > 8; RedOx conditions).

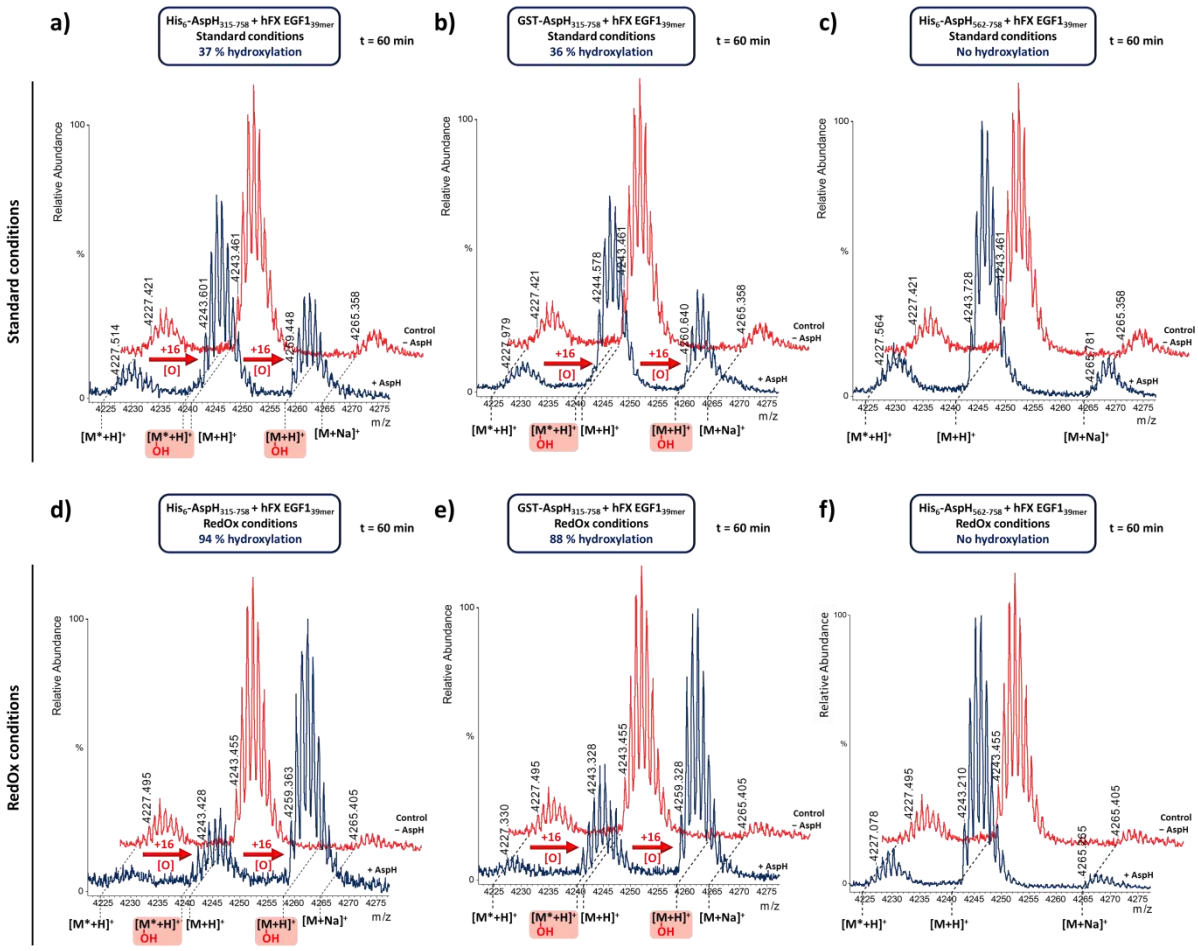


Fig. 3-26 | AspH activity in different buffer systems: **a)-c)** incubation under standard conditions (Standard Buffer: 50 mM HEPES, pH 7.5, 150 mM NaCl) or **d)-f)** incubation under RedOx conditions (here: 50 mM Tris, pH 8.0, 2.0 mM reduced *S*-(-)-glutathione (GSH), 1.0 mM oxidised *S*-(-)-glutathione (GSSG), 150 mM NaCl; the buffer system was later optimised). For comparison, the hydroxylation assays were performed using the hFX EGF1$_{39mer}$ and either His$_6$-AspH$_{315-758}$, GST-AspH$_{315-758}$, or His$_6$-AspH$_{562-758}$, to ensure that the AspH activity was not just enhanced under the RedOx conditions. [M*+H]$^+$ refers to a derivatised species, which was also hydroxylated (see **Chapter 6**).

RedOx buffer systems containing GSH and GSSG or *S*-cysteine and *S*-cystine (**Fig. 3-27a**) at alkaline pH are also used for *in vitro* refolding of recombinant EGFDs and have been proposed to enable oxido-shuffling of disulfides.[460,468] The RedOx conditions may hence enable reversible thiol-disulfide interchange[469-471] (**Fig. 3-27b**) of the canonical and non-canonical EGFD isoforms in the AspH hydroxylation assays.

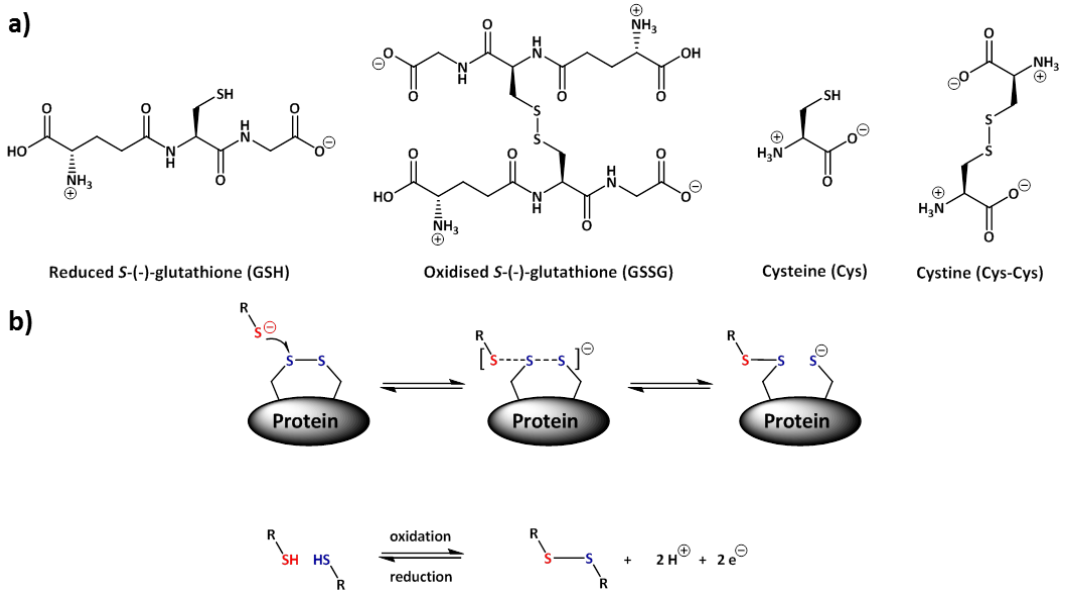


Fig. 3-27 | Scheme of **a)** common oxido-shuffling buffer (RedOx buffer) reagents, and **b)** the reversible thiol-disulfide interchange. At alkaline pH, a thiolate attacks a sulfur atom of a disulfide bond and forms a new disulfide bond.

3.2.3.1 Optimisation of Assay and Crystallisation Conditions

The buffer pH and ions can have strong influence on the activity of an enzyme. In order to optimise the buffer systems used for the AspH hydroxylation assays (standard or RedOx conditions) and crystallisation studies, $AspH_{315-758}$ activity and stability in different buffers were investigated using MALDI-ToF-MS activity assays and thermal shift measurements using differential scanning fluorimetry[472,473] (DSF, also thermofluor).

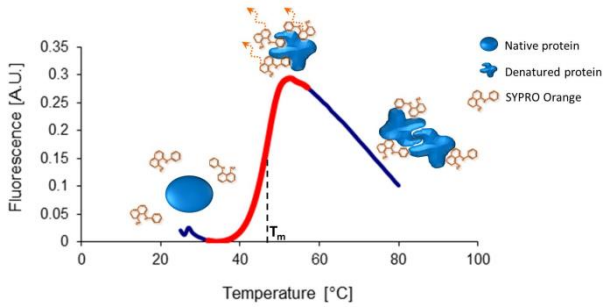


Fig. 3-28 | Typical protein melting curve from DSF.

Thermal unfolding of globular proteins leads to gradual exposure of hydrophobic patches. DSF is a method for monitoring the protein unfolding process using the dye SYPRO Orange that shows fluorescence enhancement upon binding to the hydrophobic protein regions, giving a melting curve (**Fig. 3-28**).[472] Fitting of a Boltzmann sigmoidal curve to the obtained melting curve allows determination of the midpoint representing the melting temperature (T_m) of the protein. Certain conditions such as different buffer systems or additives (e.g. metal ions, ligands) may alter the T_m of a protein, causing a positive or negative T_m shift indicating stabilisation or destabilisation respectively. DSF is a rapid and highly versatile technique for determination of optimal assay and crystallisation conditions and for high-throughput inhibitor screening.[472-474]

Hydroxylation activity assays conducted in a variety of buffers showed that $AspH_{315-758}$ was highly active in a wide pH range (from pH 6.0 to pH 9.0), and no significant difference in activity was observed in the different buffers in this range (**Table 3-1**). The RedOx Buffer system was optimised

with respect to the buffer pH, the concentrations and the ratio of GSH and GSSG. The higher the GSH concentration and GSH/GSSG ratio, the higher hydroxylation levels were obtained within 10 min incubation time (**Table 3-2**). High concentrations of GSH, however, led to high background noise. Concentrations of 3.0 mM GSH and 0.3 mM GSSG in Tris-HCl buffer, pH 8.5, appeared to provide optimal RedOx conditions for hydroxylation assays.

Table 3-1 | $AspH_{315-758}$ activity in selected buffers. The pH range and buffer ions for optimal enzymatic activity were investigated using hydroxylation assays of $AspH_{315-758}$ with the NC_{26mer} peptide in presence of all cosubstrates/cofactors, quenched after 10 min incubation at 37 °C and subsequently analysed using MALDI-ToF-MS. The average hydroxylation refers to duplicate assays.

	Buffer (50 mM)	pH	pK_a	Useful pH range	Average hydroxylation [%]
1	Citrate	3.0	3.1; 4.8; 6.4	5.8-7.2	0
2	MES	6.0	6.1	5.5-6.7	> 95
3	Bis-Tris	6.5	6.5	5.8-7.2	> 95
4	HEPES	7.0	7.5	6.8-8.2	> 95
5	HEPES	7.5	7.5	6.8-8.3	> 95
6	Tris-HCl	7.5	8.1	7.0-9.0	> 95
7	HEPES	8.0	7.5	6.8-8.5	> 95
8	Tris-HCl	8.0	8.1	7.0-9.0	> 95
9	EPPS	8.0	8.0	7.3-8.7	> 95
10	Tris-HCl	8.5	8.1	7.0-9.0	> 95
11	Bis-Tris-Propane	8.5	6.8; 9.0	6.3-9.5	90
12	TAPS	9.0	8.4	7.7-9.1	> 95
13	Bicine	9.0	8.3	7.6-9.0	13
14	CHES	9.5	9.3	8.6-10.0	11
15	CAPSO	10.0	9.6	8.9-10.3	0

Table 3-2 | RedOx Buffer optimisation with respect to buffer pH, concentrations and ratio of GSH and GSSG. The assays were performed using $AspH_{315-758}$ with the C_{26mer} peptide in presence of all cosubstrates/cofactors, quenched after 10 min incubation at 37 °C and subsequently analysed using MALDI-ToF-MS. The average hydroxylation refers to duplicate assays.

	Buffer, pH*	GSH Final conc. [mM]	GSSG Final conc. [mM]	Ratio GSH/GSSG	Average hydroxylation [%]
1	50 mM Tris-HCl, pH 8.0	3.0	0.3	10:1	63
2	50 mM Tris-HCl, pH 8.0	1.0	0.2	5:1	50
3	50 mM Tris-HCl, pH 8.0	2.0	0.5	4:1	63
4	50 mM Tris-HCl, pH 8.0	3.0	1.0	3:1	66
5	50 mM Tris-HCl, pH 8.0	2.0	1.0	2:1	63
6	50 mM Tris-HCl, pH 8.0	1.0	1.0	1:1	50
7	50 mM Tris-HCl, pH 8.0	1.0	-	1:0	51
8	50 mM Tris-HCl, pH 8.0	1.0	2.0	1:2	46
9	50 mM Tris-HCl, pH 8.5	3.0	0.1	30:1	71
10	50 mM Tris-HCl, pH 8.5	3.0	0.2	15:1	70
11	50 mM Tris-HCl, pH 8.5	3.0	0.3	10:1	73
12	50 mM Tris-HCl, pH 8.5	3.0	0.4	7.5:1	68
13	50 mM Tris-HCl, pH 8.5	3.0	0.6	5:1	67
14	50 mM Tris-HCl, pH 8.5	4.0	0.4	10:1	63
15	50 mM Tris-HCl, pH 8.5	5.0	0.5	10:1	71
16	50 mM Tris-HCl, pH 8.5	6.0	0.6	10:1	72

* The buffers included 150 mM NaCl.

$AspH_{315-758}$ stability in the buffers was assessed using differential scanning fluorimetry and correlated with the hydroxylation activity assay results (**Fig. 3-29**). Higher hydroxylation levels were obtained in buffers with higher pH, but the protein stability was decreased. GSH/GSSG in the buffer generally stabilised $AspH_{315-758}$. The optimised buffer compositions that were used for all subsequent studies are given in **Table 3-3**.

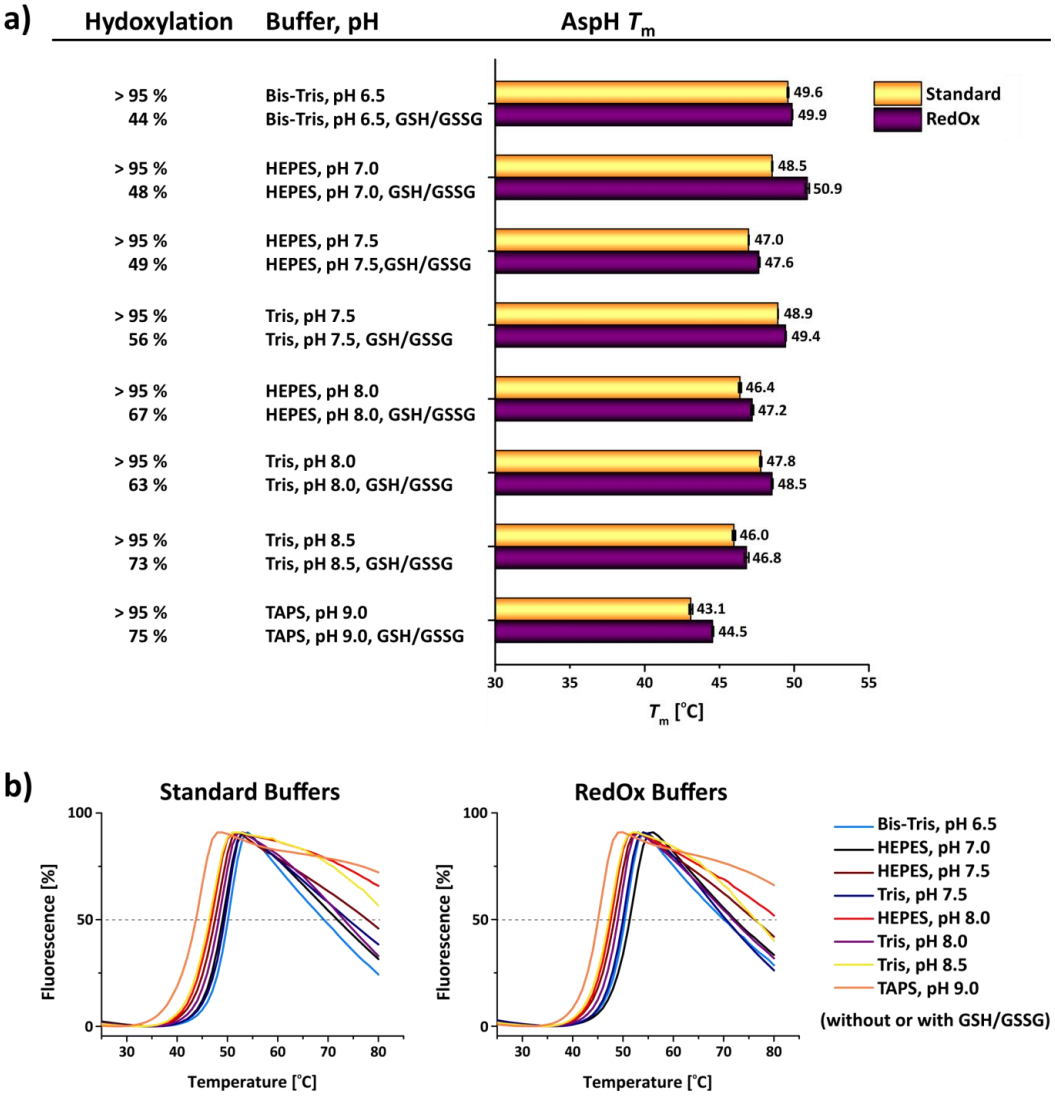


Fig. 3-29 | Optimisation of the hydroxylation assay buffer systems. **a)** Results from hydroxylation activity and DSF assays in standard or RedOx buffers (containing 3.0 mM GSH and 0.3 mM GSSG) as stated. **b)** Representative AspH melting curves in the given buffers (one curve of triplicates shown for each experiment). Hydroxylation values and error bars represent SD for triplicate assays.

Table 3-3 | Compositions of optimised buffer systems used for hydroxylation activity and DSF assays.

	Buffer system composition
Standard Buffer (Standard conditions)	50 mM HEPES, pH 7.5, 150 mM NaCl
RedOx Buffer (RedOx conditions)	50 mM Tris-HCl, pH 8.5, 3.0 mM reduced *S*-(-)-glutathione (GSH), 0.3 mM oxidised *S*-(-)-glutathione (GSSG), 150 mM NaCl
Refolding Buffer (Refolding conditions)	50 mM Tris-HCl, pH 8.3, 3.0 mM *S*-cysteine, 0.3 mM *S*-cystine, 10 mM $CaCl_2$

With respect to optimisation of crystallographic studies, the His_6-$AspH_{315-758}$ thermal stability in presence of various metal ions was screened using DSF (**Fig. 3-30**). Zn(II) ions destabilised His_6-$AspH_{315-758}$. Addition of Ni(II) and Co(II) appeared to lead to higher T_m values; however, these results were likely due to increased binding of these metal ions to the His_6-tag, which was not been cleaved from the protein for these studies. Only small differences in stability were observed with the other tested metal ions. As before, Mn(II) was therefore used to replace the active site Fe(II) in all subsequent crystallographic studies.

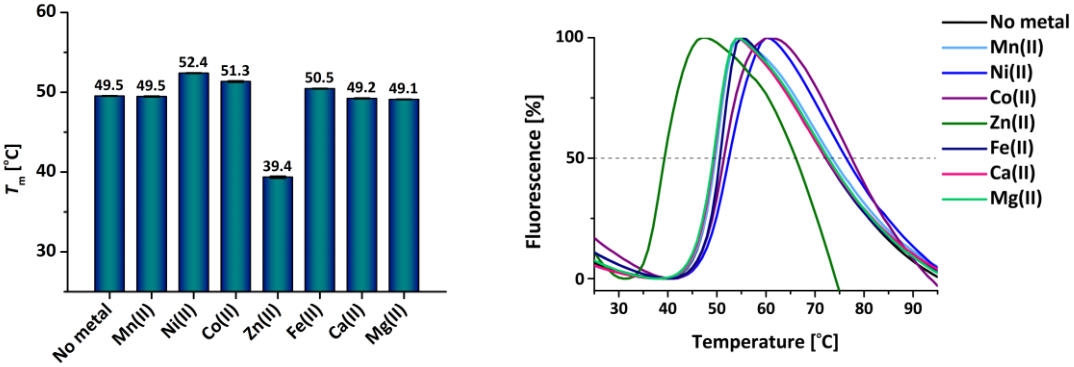


Fig. 3-30 | DSF metal ion screening using His_6-$AspH_{315-758}$ in presence of different metal ions in (50 µM metal ions, in 50 mM HEPES buffer) represented as T_m values and melting curves (one curve of triplicates shown for each experiment). Error bars represent SD for triplicate assays.

3.2.4 Disulfide Requirements

Since the characterisation of the synthetic hFX $EGF1_{39mer}$ oxidised by air oxidation had indicated that the material may contain several disulfide isoforms, further work was directed towards defining which disulfides were required in an AspH substrate.

The activity of $AspH_{315-758}$ with hFX EGF1-based 26mer peptides lacking the Cys5-6 disulfide-forming residues was investigated. A canonical peptide with Cys1-3, 2-4 linkages, the C_{26mer} (**Fig. 3-31c**), and a non-canonical peptide with Cys1-2, 3-4 linkages, the NC_{26mer} (**Fig. 3-31d**), were commercially obtained from *GL Biochem Shanghai Ltd*. As in reported procedures,[468,394] these peptides were synthesised using orthogonal protection of the cysteines enabling regioselective oxidation to ensure the canonical or the non-canonical disulfide pattern. The 26mer peptides simplified the analysis due to the shorter sequence and only three different possible disulfide patterns (Cys1-3, 2-4 linkages, as in the canonical C_{26mer}, Cys1-2, 3-4 linkages, as in the non-canonical NC_{26mer}, and Cys1-4, 2-3 linkages).

In addition, the activity of a modified 39mer peptide with a single disulfide in the non-canonical Cys3-4 linkage, the NC-$4Ser_{39mer}$ (**Fig. 3-31e**), was investigated. The other four cysteines of the original hFX $EGF1_{39mer}$ sequence were replaced by serine residues in the NC-$4Ser_{39mer}$ peptide, so that no other intramolecular disulfide bonding than the Cys3-4 linkage was possible.

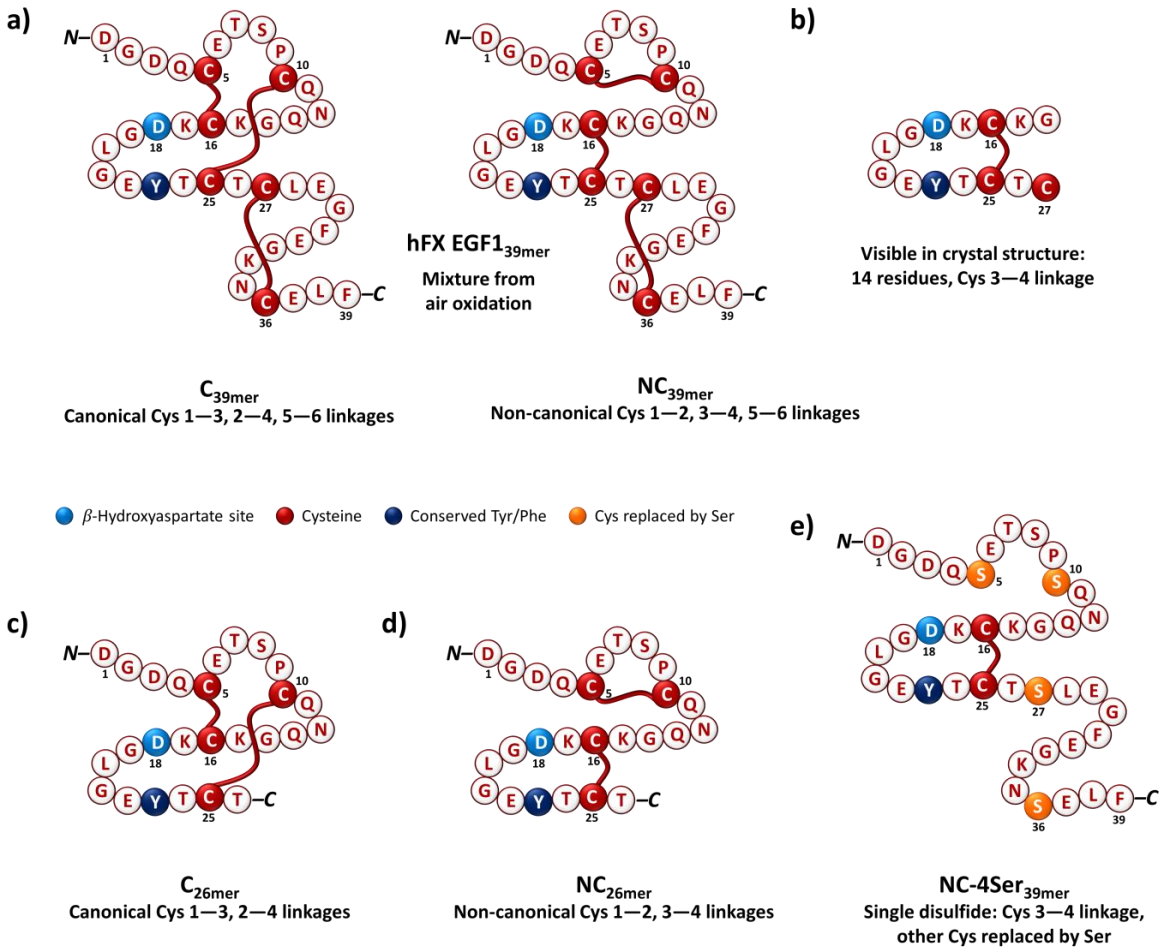


Fig. 3-31 | Schematics of the synthetic hFX EGF1-based peptides. **a)** Characterisation of the commercially obtained hFX EGF1$_{39mer}$, which was oxidised by air oxidation, indicated that this material was a mixture of several disulfide isoforms (with the canonical Cys1-3, 2-4, 5-6 and the non-canonical Cys1-2, 3-4, 5-6 disulfide linkages). **b)** In the obtained AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ crystal structure visible part (14 residues) of the peptide substrate. **c)** The C$_{26mer}$ peptide has canonical disulfides linked in Cys1-3, 2-4 pattern. **d)** The NC$_{26mer}$ peptide has non-canonical disulfides with Cys1-2, 3-4 linkages. **e)** The NC-4Ser$_{39mer}$ is a modified 39mer peptide with a single disulfide in the non-canonical Cys3-4 linkage; the other Cys were replaced by Ser residues.

3.2.4.1 Characterisation and Disulfide Pattern Analyses

Compared to the broad, unsymmetrical peak in the LC-MS elution profile of the hFX EGF1$_{39mer}$ material (**Fig. 3-2**, **Section 3.1**), the NC$_{26mer}$, the C$_{26mer}$, and the NC-4Ser$_{39mer}$ peptides eluted as sharp, symmetrical peaks under the same LC-MS conditions (**Fig. 3-31** and **Fig. 3-32**). The oxidised forms of the NC$_{26mer}$ and the C$_{26mer}$ peptides eluted at different retention times (t_R(NC$_{26mer}$) = 3.71 min vs. t_R(C$_{26mer}$) = 3.90 min), whereas a disulfide reduction led to the same linear 26mer peptide, LP$_{26mer}$, supported by the same elution profile of the reduced peptides (t_R(LP$_{26mer}$) = 3.92 min).

In the 'top-down' approach MS/MS fragmentation of the NC$_{26mer}$ and the C$_{26mer}$ peptides with intact disulfides very different, characteristic fragmentation patterns were observed (**Fig. 3-33**), which were consistent with the expected fragment ions (**Fig. 3-34**). Generally, the non-canonical NC$_{26mer}$ ionised readily, whereas a lower signal-to-noise ratio was usually observed in the measurements with the canonical C$_{26mer}$.

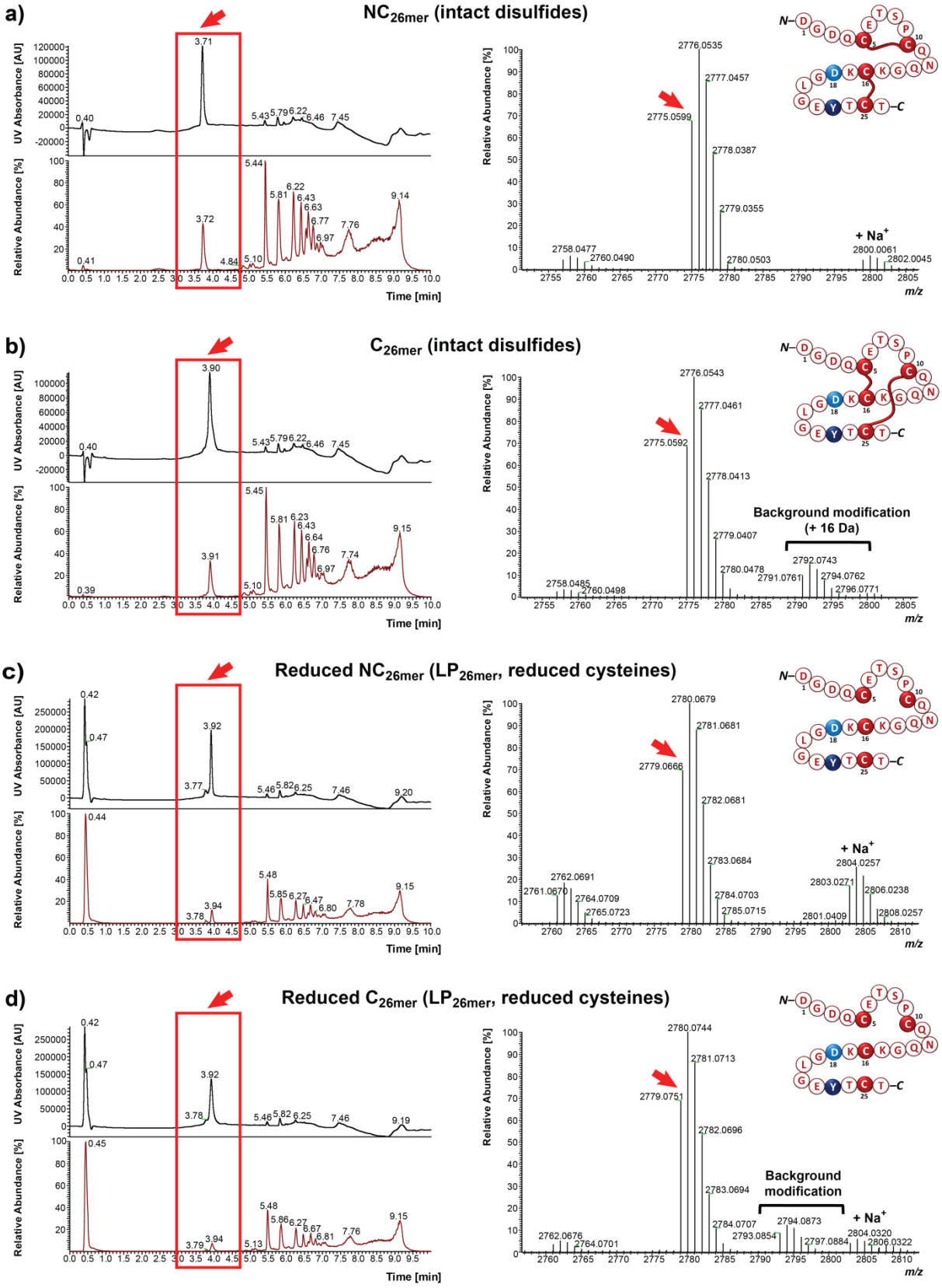


Fig. 3-32 | LC-MS analyses of **a)** the NC_{26mer} (intact disulfides linked in the non-canonical Cys1-2, 3-4 pattern) and **b)** the C_{26mer} (intact disulfides linked in the canonical Cys1-3, 2-4 pattern) in H_2O. Disulfide reduction of **c)** the NC_{26mer} and **d)** the C_{26mer} led to the same linear 26mer peptide, LP_{26mer}, supported by the same elution profile of the reduced peptides (reduction performed using TCEP in 50 mM citrate buffer, pH 3.0, which remained in the sample until analysis). A modification of the canonical C_{26mer} corresponding to a mass addition of +16 Da (**b**) or +14 Da in the reduced sample (**d**) was observed in all measurements. The data were obtained using a Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (*Thermo Scientific*), coupled to a Dionex™ UltiMate® 3000 HPLC. Represented as UV absorbance and MS traces (including deconvoluted mass spectra on the right hand side, monoisotopic mass indicated).

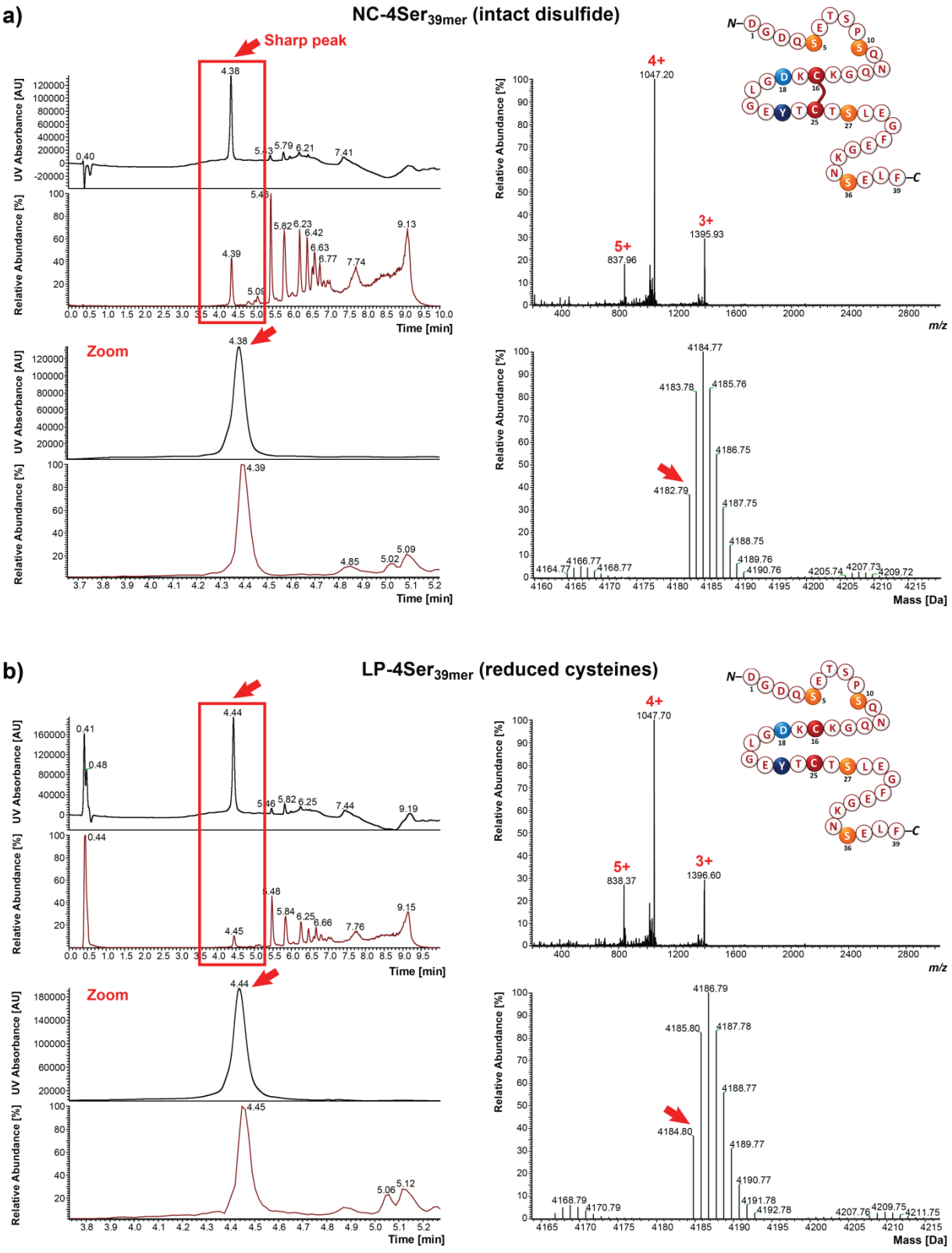


Fig. 3-33 | LC-MS analyses of **a)** the NC-4Ser$_{39mer}$ (intact disulfide) in H_2O (note the peak sharpness and symmetric shape compared to the observed peak shape in the hFX EGF1$_{39mer}$ analyses, **Fig. 3-2**) and **b)** the corresponding linear peptide, LP-4Ser$_{39mer}$ (cysteines reduced using TCEP in 50 mM citrate buffer, pH 3.0, which remained in the sample until analysis) performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (*Thermo Scientific*), coupled to a Dionex™ UltiMate® 3000 HPLC. Represented as UV absorbance and MS traces, and the mass spectra corresponding to the indicated peak (most abundant charge states and deconvoluted mass spectra, monoisotopic mass indicated).

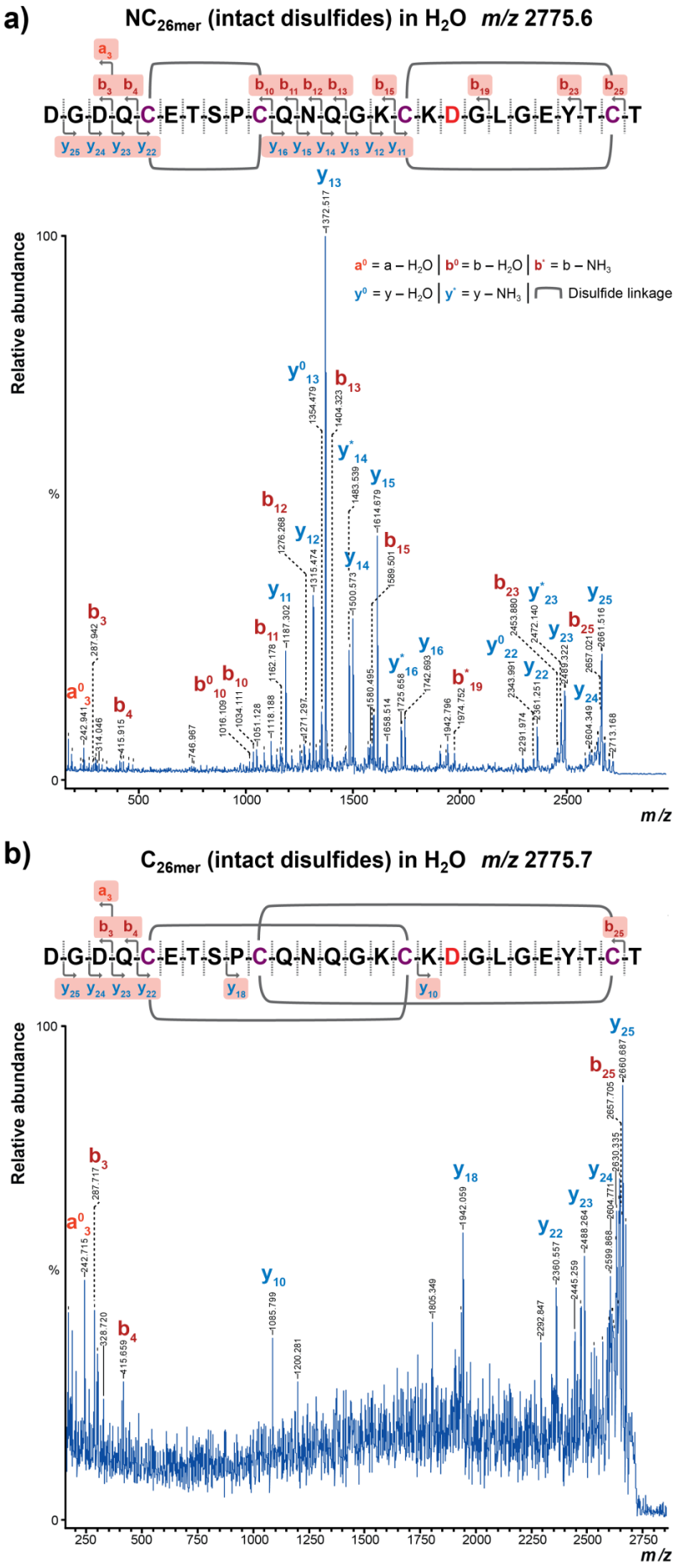


Fig. 3-34 | 'Top-down'approach MALDI-ToF-ToF MS/MS fragmentation data of **a)** the NC_{26mer} (Cys1-2, 3-4 linkage, $[M + H]^+$ *m/z* 2,775.6) and **b)** the C_{26mer} (Cys1-3, 2-4 linkage, $[M + H]^+$ *m/z* 2,775.7) peptides with intact disulfides in water.

The y_{11}-y_{16} and b_{10}-b_{15} fragment ions expected for the non-canonical Cys1-2, 3-4 disulfide linkages in the NC_{26mer} peptide ionised readily (**Fig. 3-34**). These fragment ions were not observed in the C_{26mer} 'top-down' MS/MS fragmentation data, consistent with the expected pattern (**Fig. 3-35**).

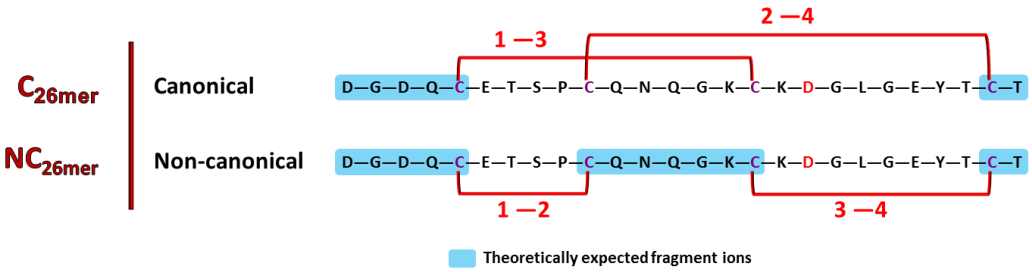


Fig. 3-35 | Schematic of the theoretically expected fragment ions in MS/MS studies of the canonical C_{26mer} with Cys1-3, 2-4 and the non-canonical NC_{26mer} with Cys1-2, 3-4 disulfide linkages.

3.2.4.2 Hydroxylation Assays

Hydroxylation assays with the His_6-$AspH_{315-758}$ and the synthetic EGFD peptides were then performed under standard or RedOx conditions and analysed using MALDI-ToF-MS.

The non-canonical NC_{26mer} was nearly fully hydroxylated under standard conditions, whereas no hydroxylation of the canonical C_{26mer} was observed under the same conditions (**Fig. 3-36a**). Under RedOx conditions, however, also the canonical C_{26mer} was nearly fully hydroxylated (**Fig. 3-36b**), supporting the proposed reversible thiol-disulfide interchange[469-471] of the canonical and non-canonical isoforms. The non-canonical NC_{26mer} and NC-$4Ser_{39mer}$ peptides were efficiently hydroxylated under both standard and RedOx conditions.

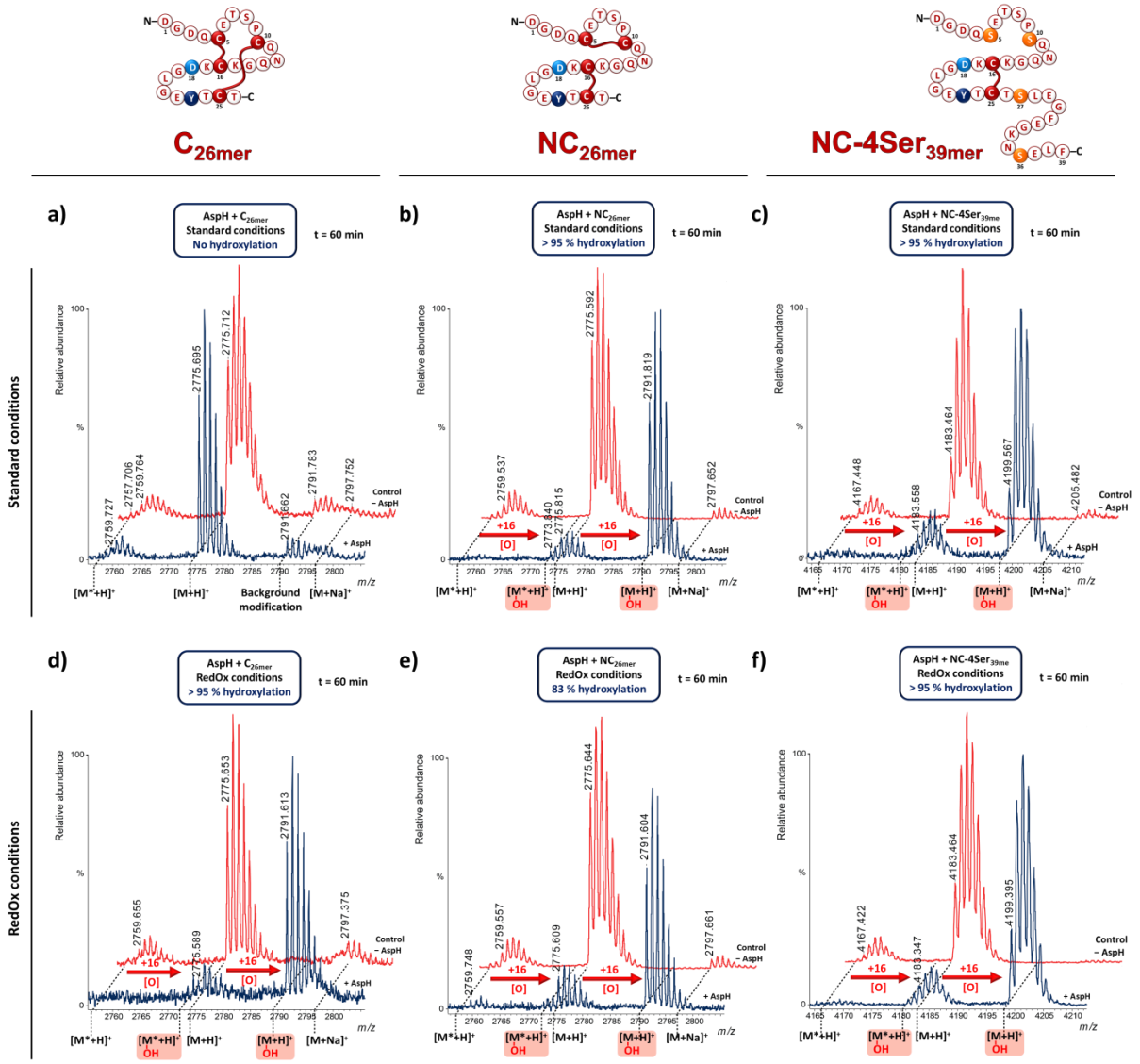


Fig. 3-36 | Hydroxylation of the synthetic hFX EGF1-based peptides, the canonical C_{26mer}, the non-canonical NC_{26mer} (both peptides were synthesised using orthogonal protection of the cysteines enabling regioselective oxidation; unhydroxylated $[M+H]^+$ *m/z* 2,775.6; hydroxylated $[M(OH)+H]^+$ *m/z* 2,791.6), and the NC-4Ser$_{39mer}$ peptide with a single disulfide in Cys3-4 linkage (unhydroxylated $[M+H]^+$ *m/z* 4,183.4; hydroxylated $[M(OH)+H]^+$ *m/z* 4,199.4): **a)-c)** incubation under standard conditions (Standard Buffer: 50 mM HEPES, pH 7.5, 150 mM NaCl) or **d)-f)** incubation under RedOx conditions (50 mM Tris-HCl, pH 8.5, 3.0 mM GSH, 0.3 mM GSSG, 150 mM NaCl). A modification of the canonical C_{26mer} corresponding to a mass addition of + 16 Da was observed (spectra in **a)**, approximately 7 % of the peptide) in all measurements; however, this modification was also found in the control sample (no incubation with AspH) and when samples of the peptide were freshly dissolved in H_2O without any additives and directly analysed (also see **Fig. 3-31**). $[M^*+H]^+$ refers to a derivatised species, which was also hydroxylated (see **Chapter 6**).

3.3 Discussion

Spontaneous folding (by air oxidation) has been proposed to be biased to a native conformation by amino acid sequence predetermination.[396] The results presented in this Chapter provided evidence that the hFX EGF1$_{39mer}$ material obtained using *in vitro* refolding by air-mediated oxidation, followed by HPLC purification of the major species, was a mixture of several disulfide isoforms, which co-eluted under the same peak in standard HPLC purification/LC-MS conditions. Air oxidation for the

disulfide formation was also used for reported synthetic AspH substrates.[341,340] The 'top-down' approach MS/MS fragmentation data and trypsin digestion performed under non-reducing conditions indicated that the hFX $EGF1_{39mer}$ material may be a mixture of the canonical Cys1-3, 2-4, 5-6 and the non-canonical Cys1-2, 3-4, 5-6 isoforms. It is possible that further disulfide isoforms were present in the hFX $EGF1_{39mer}$ material. However, the Cys3-4 disulfide in the $AspH_{315-758}$·Mn^{II}·NOG·hFX $EGF1_{39mer}$ crystal structure and the disordered hFX $EGF1_{39mer}$ *N*- and *C*-termini supports the proposal that the form bound to AspH was likely the non-canonical isoform with Cys1-2, 3-4, 5-6 linkages.

There are only a few reports of EGFDs with non-canonical disulfide linkages. The fifth EGFD of human thrombomodulin (hTRBM EGF5) has been reported to have the non-canonical Cys1-2, 3-4, 5-6 disulfide pattern (**Fig. 3-37**).[475-477] Interestingly, this domain is the thrombin-binding region,[477] and the more flexible non-canonical fold has been proposed to account for the high affinity to thrombin.[478] However, hTRBM EGF5 has no Ca^{2+}-binding or hydroxylation consensus residues. The hTRBM EGF4 and the Ca^{2+}-binding EGF6 had the typical canonical Cys1-3, 2-4, 5-6 disulfide linkages (**Fig. 3-37**).

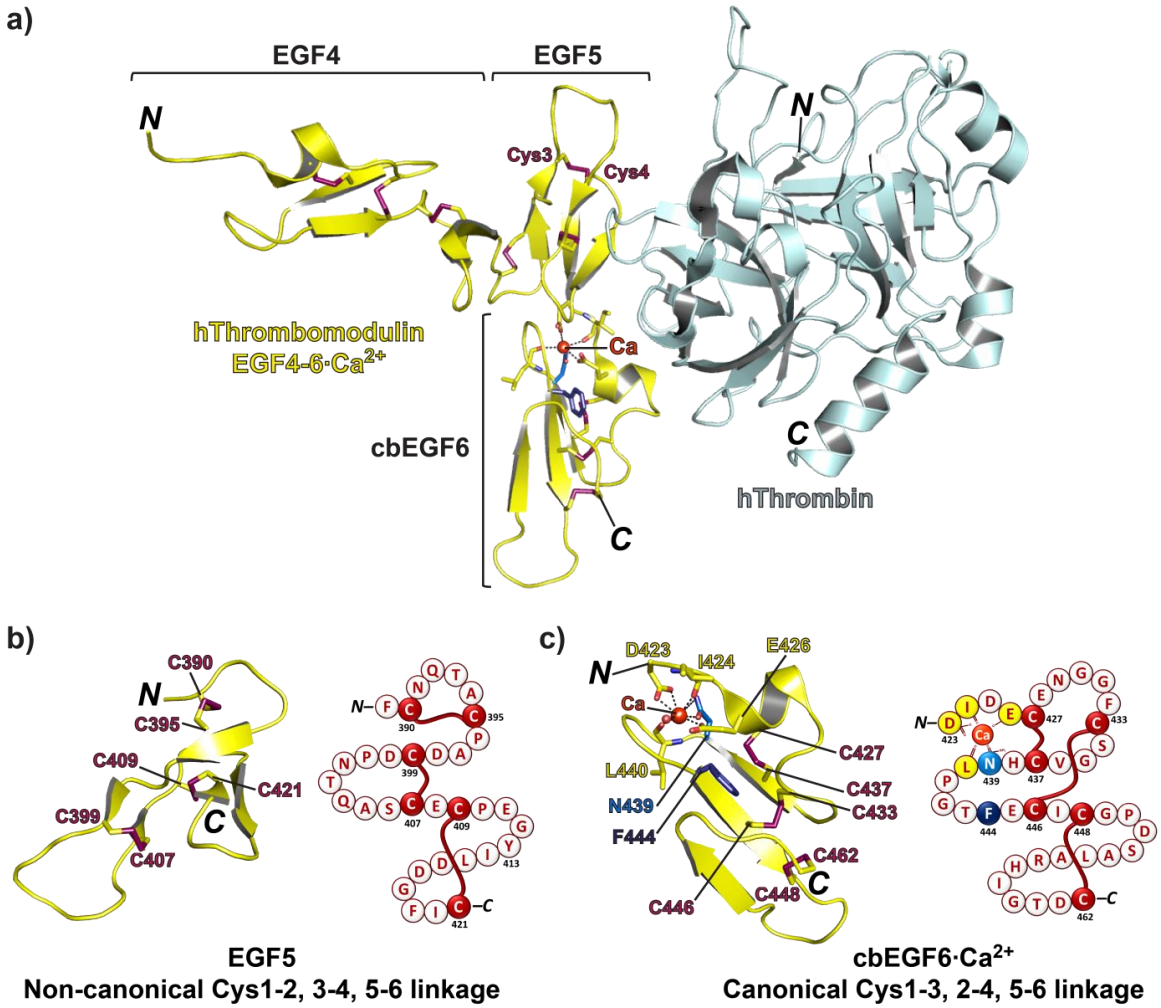


Fig. 3-37 | View from a crystal structure of a human thrombin-thrombomodulin EGF4-6 complex (resolution: 2.30 Å; PDB: 1DX5[477]). **a)** Overview of the complex structure. The human thrombomodulin (hTRBM) EGF5 domain is the only reported EGFD with a non-canonical disulfide pattern in Cys1-2, 3-4, 5-6 linkages. Close-ups of **b)** the non-canonical hTRBM EGF5 and **c)** the Ca^{2+}-binding hTRBM EGF6 with canonical Cys1-3, 2-4, 5-6 linkages.

Heat-denaturation of human epidermal growth factor (EGF) has been reported to lead to scrambled disulfide isomers with the non-canonical Cys1-2, 3-4, 5-6 disulfide form being the predominant isomer formed.[479] Structure calculation studies on murine EGF suggested a certain disulfide plasticity and determined that the non-canonical Cys1-2, 3-4, 5-6 and a Cys1-3, 2-5, 4-6 disulfide pattern satisfied the restraints indistinguishably from the canonical one with its Cys1-3, 2-4, 5-6 linkages.[480] A study on the folding pathway of human EGF showed proceeding via species with a single and two disulfides, and rendered the non-canonical Cys1-2, 3-4, 5-6 and a Cys1-6, 2-5, 3-4 isomer well-populated intermediates on the way to the native state.[481]

3.3.1 Canonical vs. Non-Canonical Disulfide Pattern

The studies on AspH substrate requirements indicated that the EGFDs with the crystallographically observed Cys3-4 disulfide reflected the AspH substrate. The low hydroxylation levels achieved with the well-characterised hFib1 and hNotch1 cbEGFD proteins under standard conditions may have resulted from "contaminations" with the non-canonical disulfide isoforms (**Fig. 3-38**), which could have eluted covered under the same peak in the HPLC purification (visible as a peak shoulder in the UV absorbance chromatogram, as observed in the HPLC purifications performed by Dr Sacha Jensen, personal communication). Hydroxylation assays with the hFib1 and hNotch1 cbEGFD proteins were later also performed under RedOx conditions and resulted in much higher hydroxylation levels, consistent with the proposed thiol-disulfide interchange between canonical and non-canonical disulfide isoforms (see **Chapter 7**).

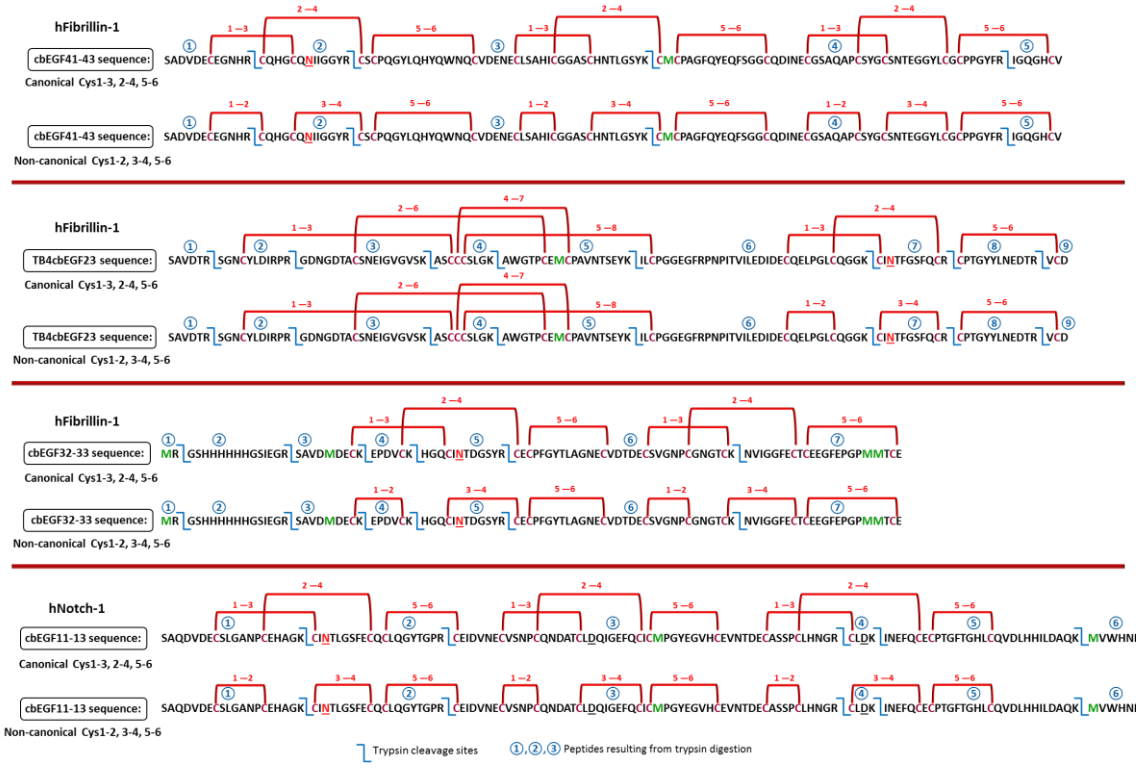


Fig. 3-38 | Schematic of potential canonical (Cys 1-3, 2-4, 5-6 disulfide linkages) and non-canonical (Cys 1-2, 3-4, 5-6 disulfide linkages) disulfide patterns in the hFib1 and hNotch1 cbEGFD proteins.

The hydroxylation activity assays showed that the C_{26mer} peptide with the canonical Cys 1-3, 2-4 disulfide linkages was not an AspH substrate using standard conditions. Using RedOx conditions, however, the peptide was turned into a substrate through the proposed thiol-disulfide interchange. Consistent with the Cys3-4 disulfide obverved in the crystal structure, the non-canonical NC_{26mer} (with Cys1-2, 3-4 linkages) was efficiently hydroxylated under all tested conditions. Furthermore, the assay results revealed that the presence of a single disulfide, i.e. the Cys3-4 linkage, in the NC-4Ser$_{39mer}$ peptide was sufficient for AspH hydroxylation.

Besides β-hydroxylation, further posttranslational modifications with unusual glycans unique to EGFDs have been identified on specific, partially conserved Ser or Thr residues (shown for hFIX EGF1 in **Fig. 3-39b**).[482] *O*-linked Xyl-Glc-(Ser/Thr) or Xyl-Xyl-Glc-(Ser/Thr)[483,484] are found within a proposed consensus sequence C_1XSXPC_2, and *O*-linked fucose[485-487] or tetrasaccharide Sia-Gal-GlcNAc-Fuc-(Ser/Thr)[488] within the consensus $C_2XXGG(S/T)C_3$. Compositional analyses of proteins containing βHya/βHyn or fucosylated residues raised the possibility that both modifications could be mutually exclusive.[489] EGFD glycosylation seems to be directly linked to signalling, as demonstrated for the case of the notch signalling pathway: Knockout of the fucosyltransferase POFUT1 in mice resulted in an embryonic-lethal phenotype, suggesting that notch required the *O*-linked fucose modification of its EGFDs allowing for subsequent elongation by the GlcNAc-transferase fringe and further glycosyltransferases for its function.[490] Interestingly, POFUT1 is localised to the ER[491] and has been reported to require proper folding of the EGFD substrate.[492]

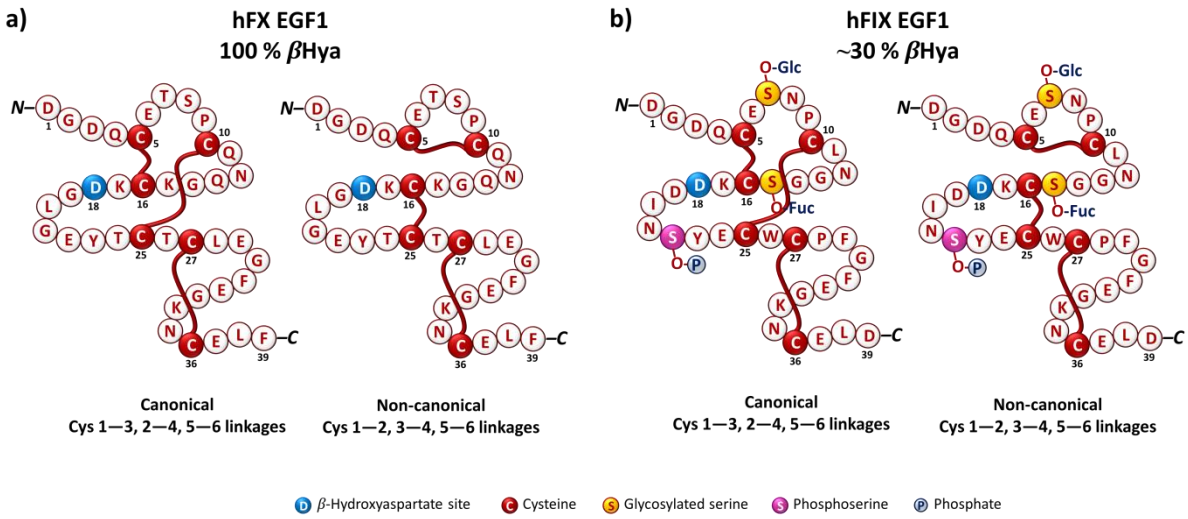


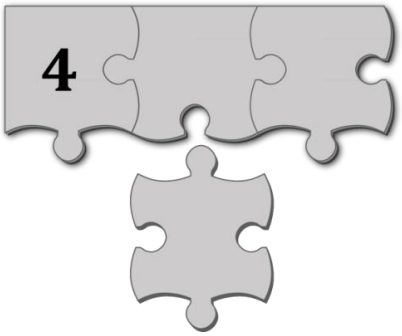
Fig. 3-39 | Schematic comparison of canonical and non-canonical disulfide linkages of **a)** hFX GF1, which was fully hydroxylated when extracted from mammalian tissues,[336] and **b)** hFIX EGF1, which was only ~30 % hydroxylated but additionally glycosylated and phosphorylated on partially conserved Ser and Thr residues.[482-484] Glc: glucosyl; Fuc: fucosyl.

The apparent AspH substrate selectivity for EGFDs with a non-canonical Cys3-4 disulfide suggested that may have a role in quality control of "correct" cbEGFD folding. However, the reversible nature of the thiol-disulfide interchange rendered a definitive identification of the disulfide

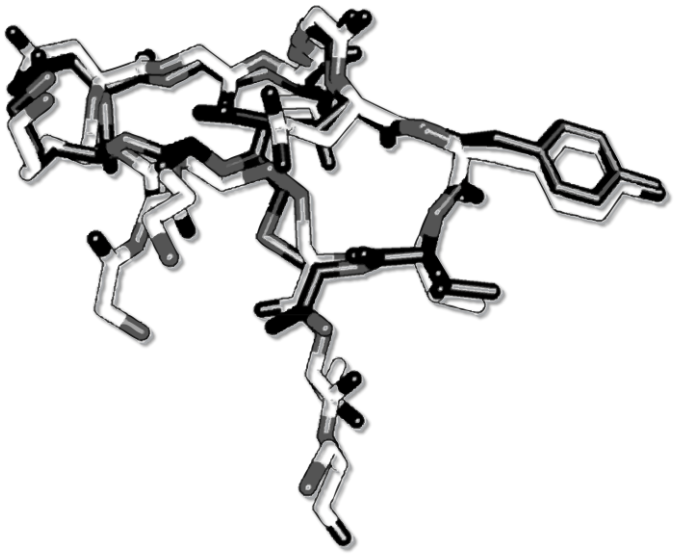
pattern of the substrate difficult. The activity of AspH with cyclic peptides representing irreversibly formed cyclic analogues of the Cys3-4 disulfide was thus investigated in the work presented in **Chapter 4**.

3.4 ACKNOWLEDGEMENTS FOR CHAPTER 3

Most of the MALDI-ToF-MS measurements were performed using the *Bruker Daltonics* Ultraflex I MALDI-ToF/ToF spectrometer based at the Oxion Proteomics Facility, Department of Physiology, Anatomy and Genetics, University of Oxford, and Dr Holger Kramer is kindly acknowledged for granting access to the instrument and for fruitful discussions. Dr James Wickens is acknowledged for kind technical assistance with operating the Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer coupled to a Dionex™ UltiMate® 3000 HPLC and for helpful discussions. Many thanks also go to Professor Penny A. Handford and her coworkers Dr Sacha Jensen and Dr Paul Taylor for providing the hFib1 and hNotch1 cbEGFD proteins.



Cyclic Peptides for AspH



Abstract

This Chapter describes work focusing on the design and synthesis of cyclic peptides as AspH substrates and probes to consolidate the results from prior crystallographic and substrate selectivity studies.

4.1 Design of a Cyclic Peptide Substrate

As described in **Chapter 3**, the hFX EGF1$_{39mer}$ AspH substrate can be fully hydroxylated in a RedOx buffer system, which is proposed to enable thiol-disulfide interchange of the canonical and non-canonical forms. Due to the reversible nature of thiol-disulfide interchange, a definitive identification of the disulfide pattern of the EGFD substrate under RedOx conditions is difficult. Therefore, an irreversibly linked (preventing possible disulfide opening or shuffling) cyclic peptide for AspH was required to support the results reported in **Chapter 3** and to prove that the ring structure formed by the Cys3-4 disulfide bond is the AspH substrate. In collaboration with Dr Martin Münzel and Dr Michael A. McDonough an irreversibly formed cyclic analogue of the Cys3-4 disulfide was thus devised based on the structural information from the crystal structures (**Fig. 4-1**).

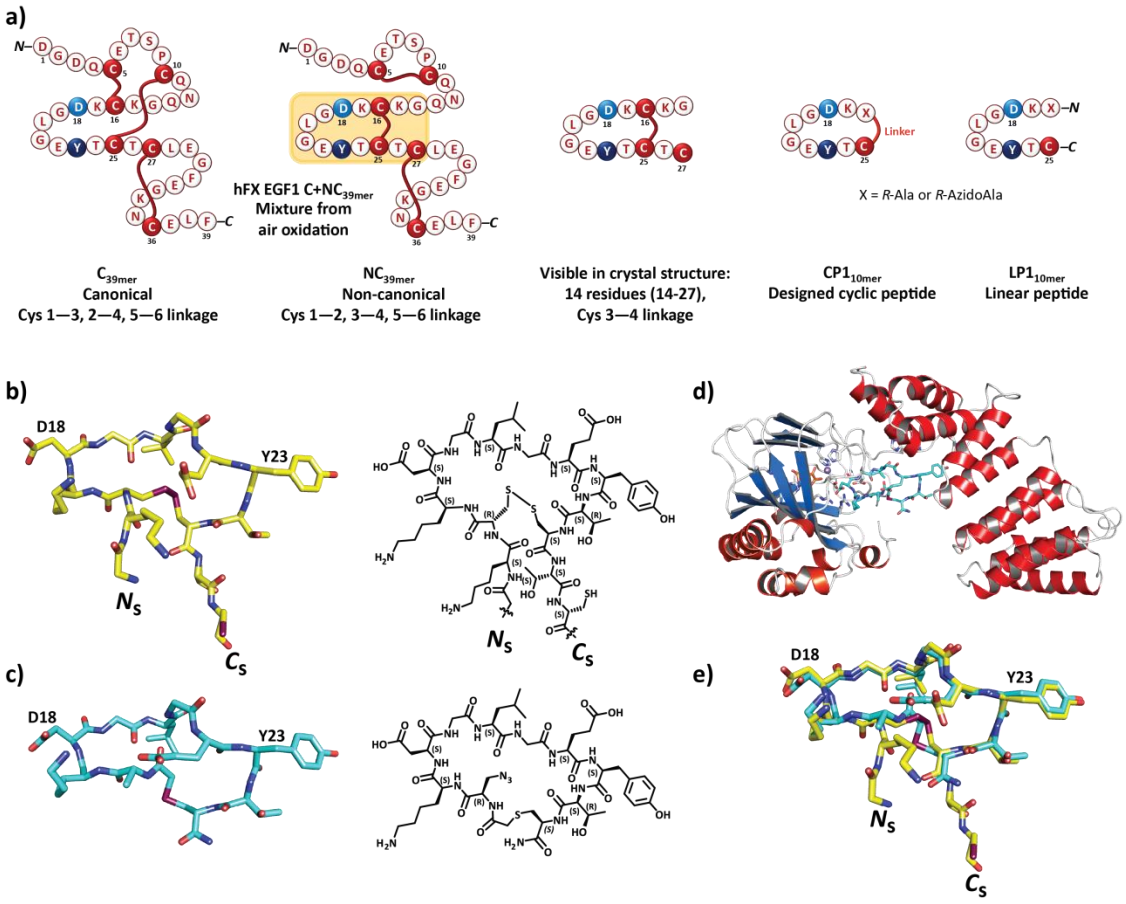


Fig. 4-1 | Design of a cyclic peptide for AspH based on the AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ crystal structure. **a)** Schematic representation of the hFX EGF1$_{39mer}$ peptide (mixture of the canonical and the non-canonical forms), visible part of the peptide in the crystal structure, and the designed cyclic and the corresponding linear peptide. **b)** Sticks representation of the visible hFX EGF1$_{39mer}$ substrate part in the crystal structure (residues 14-27) and skeletal formula. **c)** Model of the designed cyclic peptide (CP$_{10mer}$ comprising residues aa 16-25 of the substrate with C16 replaced by *R*-Ala or *R*-AzidoAla) made in PyMOL and skeletal formula. **d)** Cyclic peptide modelled in the crystal structure. **e)** Superimposition of the substrate and the cyclic peptide model.

Based on the AspH$_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ crystal structure, a 10mer cyclic peptide (CP$_{10mer}$), comprising the core ring residues aa 16-25 of the hFX EGF1$_{39mer}$ with non-canonical

disulfides (NC_{39mer}) substrate, was designed (**Fig. 4-1**). The Cys3-4 disulfide bond in the hFX $EGF1_{39mer}$ substrate was replaced by a stable thioether bridge in the cyclic peptide, using the methodology of Suga and coworkers.[493] The irreversibly linked thioether bond was incorporated to eliminate the possibility of any disulfide opening or shuffling and to make the scaffold more rigid. The *N*-terminus was functionalised with a chloroacetyl group, which was brought to reaction with a *C*-terminal cysteine in the peptide to give the thioether. The Asp and Tyr residues in the cyclic peptide correspond to the highly conserved residues $Asp18_{39mer}$ and $Tyr23_{39mer}$ respectively in the hFX EGF1 *in vitro* substrate, which are also part of the hydroxylation consensus sequence ($-C_3X(D/N)XXXX(Y/F)XC_4XC_5-$)[338]. These residues facilitate binding in the active site and the TPR domain pocket, supporting the key interactions with AspH. The *N*-terminal amino acid in the sequence, to which the chloroacetyl group was attached, was designed to have the *R*-stereochemistry, which orients the side chain of this linker acid in the model into the position of the peptide backbone of the original substrate. Both *R*-alanine (*R*-Ala) and *R*-azidoalanine (*R*-AzidoAla) were incorporated as the linker moiety. The latter allows postsynthetic functionalisation of the peptide in a conformationally sensible position. For example, inhibiting groups or fluorescence dyes could be attached to the azide moiety for use as molecular probes.

For the cyclic peptide design, a model was constructed in PyMOL (*Schrödinger, Inc.*; Version 1.5) by extraction of the visible substrate part from the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}$hFX $EGF1_{39mer}$ crystal structure and subsequent building of the desired connections using the "Builder" tool of the software. The resulting modified peptide was imported into ChemBio3D (*CambridgeSoft*, Version 12.0), and the energy was minimised using the "MM2" tool. The minimised structure was then re-imported into PyMOL and superimposed with the original substrate (**Fig. 4-1e**) in the crystal structure to check for the extent of deviation of the model from the structure and rule out obvious clashes with the protein before proceeding with solid-phase peptide synthesis.

4.1.1 Cyclic Peptide Synthesis

The designed 10mer peptide comprising residues 16-25 of the hFX $EGF1_{39mer}$ substrate with C16 replaced by *R*-AzidoAla was the first synthesised ($CP1_{10mer}$). Using protected amino acids, the linear precursor **1** of the cyclic peptide (**Fig. 4-2**) was prepared by standard fluorenyl-methyloxycarbonyl (Fmoc)-based solid-phase peptide synthesis (SPPS) on an automated peptide synthesiser, as described in **Chapter 10**. After deprotection of the *N*-terminal Fmoc group, a chloroacetyl moiety was attached to the *R*-amino acid by reaction of **1** with chloromethylcarbonyl-oxysuccinimide (**2**, ClAc-OSu). ClAc-OSu was synthesised from chloroacetic acid chloride (**3**) and *N*-hydroxysuccinimide (**4**) in the presence of sodium bicarbonate. ClAc-OSu coupling to the *N*-terminus of the peptide was performed on the resin either manually or on the peptide synthesiser. An excess of ClAc-OSu (5-10 equivalents) was added to a suspension of the resin-bound peptide in dimethylformamide

(DMF), and the mixture was shaken for 2 h at room temperature. The resin-bound peptide was subsequently collected by filtration and then dried. After cleavage from the resin, the linear peptide **6** was cyclised by an S_N2-reaction between the thiolate of the *C*-terminal cysteine and the *N*-terminal chloroacetyl moiety to give the cyclic peptide **7**. The reaction was performed at alkaline pH > 8 in a buffer system containing aqueous triethylammonium acetate ($NHEt_3OAc$) buffer, pH and acetonitrile (CH_3CN) in a microwave reactor at 100 °C with a reaction time of 10 min. A pH adjustment of the solution to a value in the alkaline range (pH > 8) using triethylamine (NEt_3) was usually required, as the peptide was cleaved from the resin using a mixture of trifluoroacetic acid (TFA), tri*iso*propylsilane (TIPS) and water (95 % TFA, 2.5 % TIPS, 2.5 % H_2O), resulting in the peptide TFA salt (Lys residue). The peptide was subsequently purified by high-performance liquid chromatography (HPLC) using the conditions stated in **Chapter 10.**

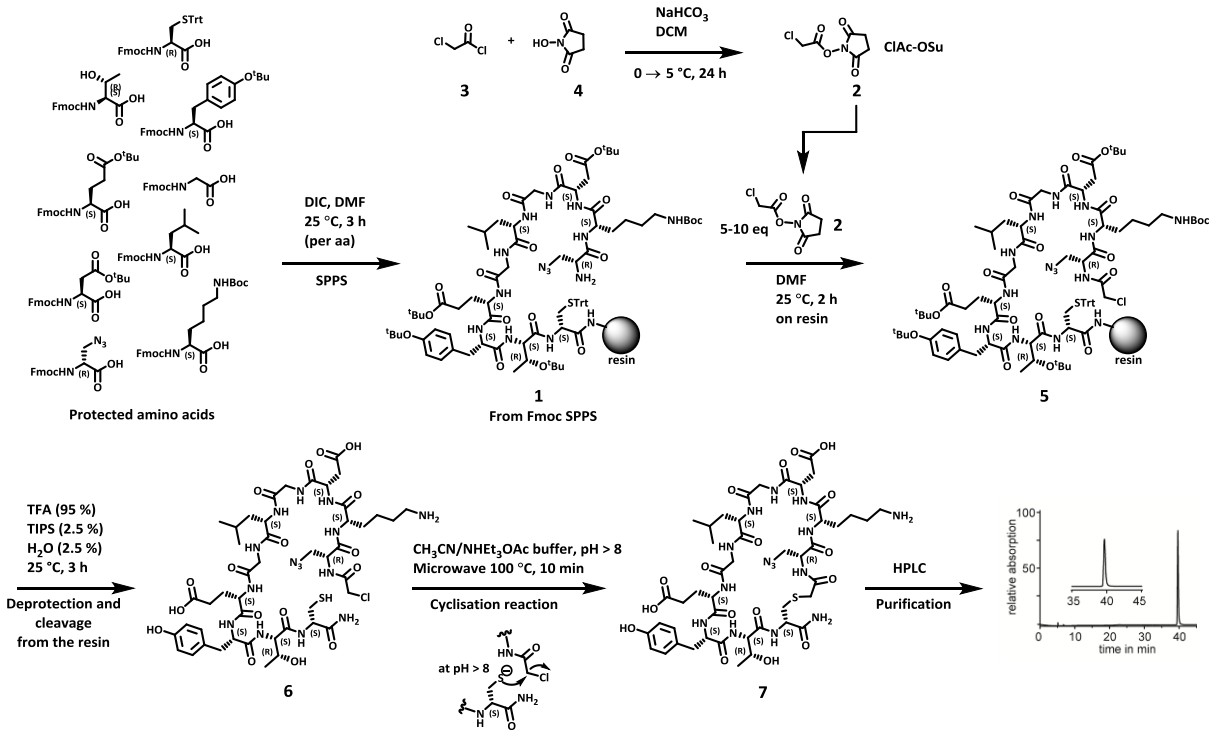


Fig. 4-2 | Synthesis of the cyclic peptide. The basic linear precursor peptide is synthesised by standard Fmoc solid-phase peptide synthesis (SPPS). After coupling of a chloroacetyl moiety to the *N*-terminus, the peptide is cyclised in a microwave reactor at pH > 8 and subsequently purified by HPLC. DIC = di*iso*propylcarbodiimide; DMF = *N,N*-dimethylformamide; ClAc-OSu = chloromethylcarbonyl-oxysuccinimide; TFA = trifluoroacetic acid; TIPS = tri*iso*propylsilane; DCM = dichloromethane.

4.1.2 The Cyclised Conformation Is Critical for Activity

The peptide was prepared both in cyclic ($CP1_{10mer}$, *m/z* 1,136 Da) and in linear ($LP1_{10mer}$, *m/z* 1,096 Da) forms. In tests for activity with AspH, only the cyclic but not the linear peptide was accepted as substrate (**Fig. 4-3**), supporting the crystallographically observed Cys3-4 disulfide as reflecting the preferred AspH substrate. This suggested that fixing of the peptide conformation was essential, consistent with the reported studies by Stenflo and coworkers.[338] The researchers tested

AspH activity with linear synthetic peptides from hFIX EGF1, synthesised based on the hydroxylation consensus sequence (-C_3X(D/N)XXXX(Y/F)XC_4XC_5-), but observed no activity of AspH with any of the linear peptides.[338] They, however, observed hydroxylation activity with a synthetic peptide comprising the whole hFX EGF1$_{39mer}$ (residues 45-87), which was synthesised and oxidised by air oxidation using a similar method as for the synthesis of the hFX EGF1$_{39mer}$ peptide used in this Thesis (as described in **Chapter 2**).[338] Thus far, this 39mer peptide containing all three disulfide bonds is the shortest reported AspH substrate.[338]

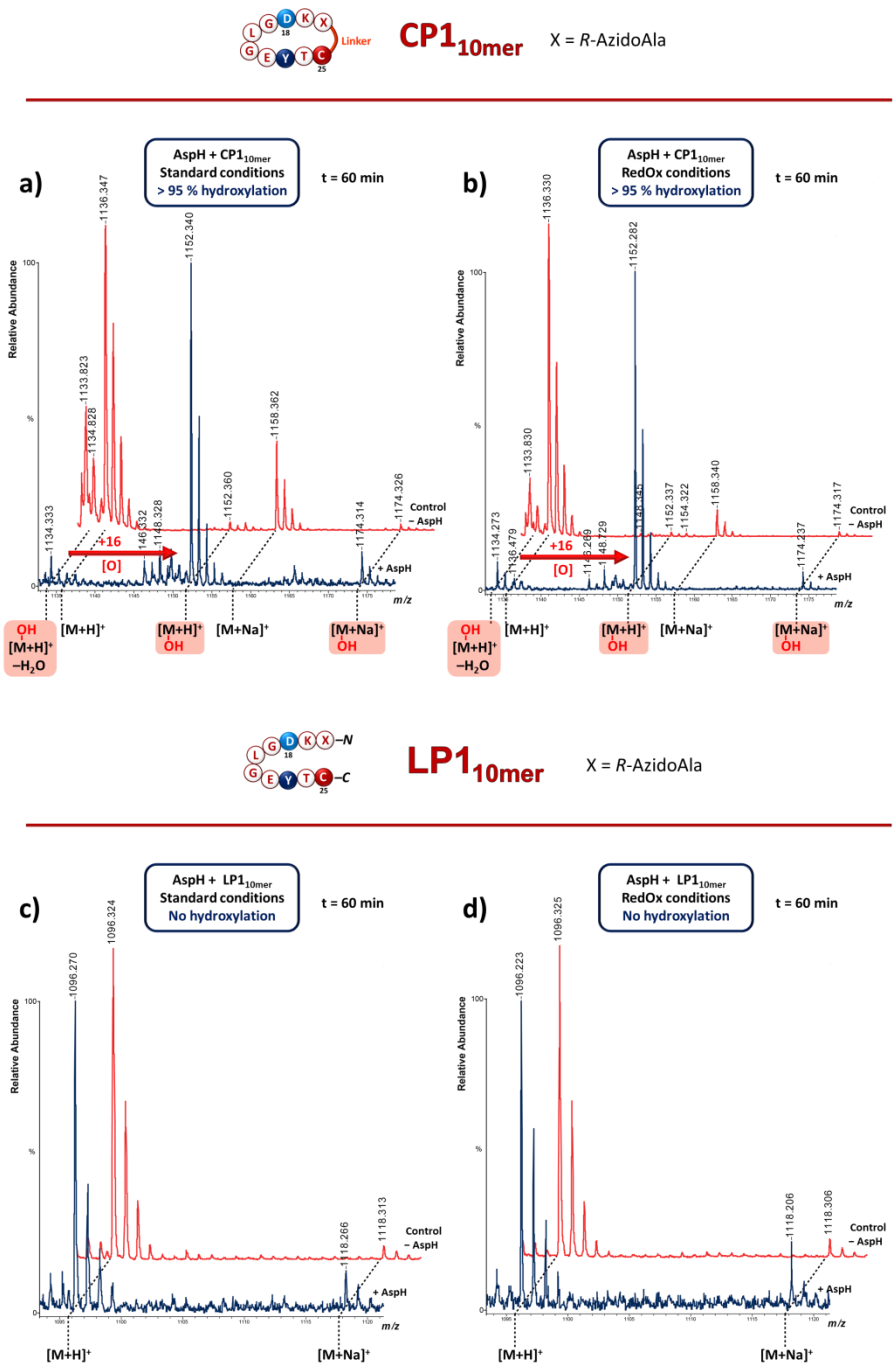


Fig. 4-3 | MALDI-ToF mass spectra showing results from AspH hydroxylation assays with the CP1$_{10mer}$ and the corresponding linear LP1$_{10mer}$. The CP1$_{10mer}$ ([M+H]$^+$ *m/z* 1,136 Da), is an efficient AspH substrate; it was fully converted to the hydroxylated peptide ([M(OH)+H]$^+$*m/z* 1,152 Da) using both **a)** Standard and **b)** RedOx buffer conditions. The peptides readily dehydrate (-H_2O) under MS analysis conditions (mass loss: -18 Da). The LP1$_{10mer}$ ([M+H]$^+$ *m/z* 1,096 Da), neither showed activity with AspH$_{315-758}$ using **c)** Standard nor using **d)** RedOx buffer conditions.

$CP1_{10mer}$ was found to be an efficient AspH substrate that reacted to give complete turnover to the hydroxylated peptide using Standard Buffer conditions within a short time (**Fig. 4-4**). The cyclic peptide poses an inexpensive alternative *in vitro* substrate that is stable in solution over long time, which is facile to synthesise and handle. Moreover, mass spectrometry analysis was considerably simplified compared to the hFX $EG1_{39mer}$ substrate with its disulfides due to the smaller size and good ionisation properties of the $CP1_{10mer}$. The peptide also showed comparable reactivity using RedOx conditions. In contrast, no activity under any tested conditions was observed with the linear $LP1_{10mer}$.

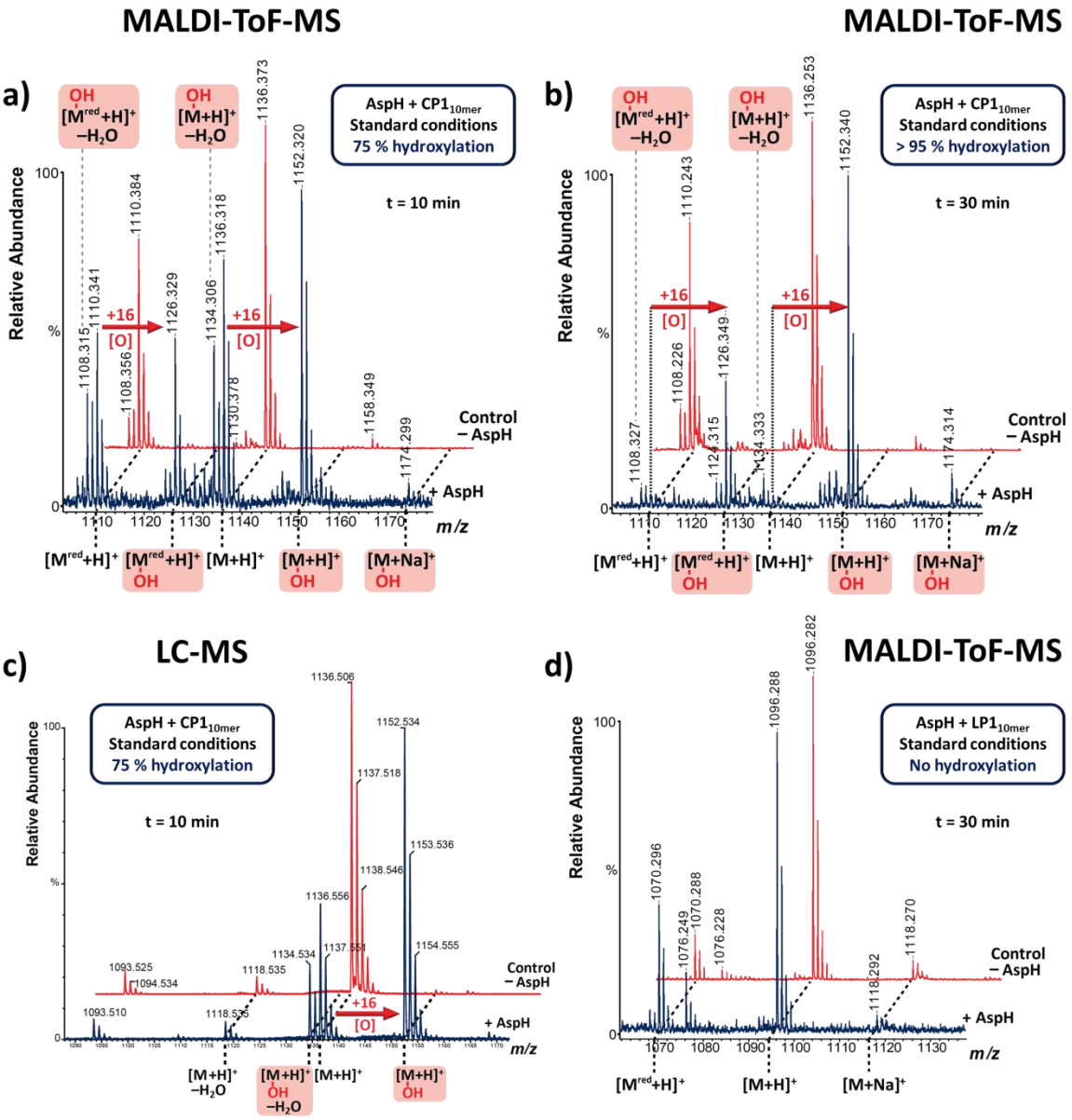


Fig. 4-4 | Mass spectra showing hydroxylation assay analyses. **a)** 75 % $CP1_{10mer}$ ($[M(OH)+H]^+$ *m/z* 1,152) hydroxylation by AspH within 10 min incubation at 37 °C, and **b)** full hydroxylation after 30 min. The wider range spectra show a partial reduction of the azide moiety to an amine ($[M^{red}+H]^+$ of $CP1_{10mer}$ *m/z* 1,110, corresponding hydroxylated peptide $[M^{red}(OH)+H]^+$ *m/z* 1,126), which was observed in samples analysed by MALDI-ToF-MS (mass loss of -26 Da). In addition, the peptides readily dehydrate ($-H_2O$) under MS analysis conditions (mass loss: -18 Da). **c)** The reduction of azides was not observed when the same samples were analysed by LC-MS. **d)** MALDI-ToF-MS spectra in wider mass range showing the reduced azide in $LP1_{10mer}$ ($[M^{red}+H]^+$ *m/z* 1,070).

When MALDI-ToF-MS was used for sample analysis of the $CP1_{10mer}$ or the $LP1_{10mer}$, apparent partial reduction of the azide moiety in $CP1_{10mer}$ ($[M+H]^+$ *m/z* 1,136 Da) and $LP1_{10mer}$ ($[M+H]^+$ *m/z* 1,096 Da) to an amine (corresponding to a mass loss of -26 Da; reduced azide in $CP1_{10mer}$: $[M^{red}+H]^+$ *m/z* 1,110; reduced azide in $LP1_{10mer}$: $[M^{red}+H]^+$ *m/z* 1,070) occurred under the analysis conditions (**Fig. 4-4**). The reduced azide peaks were usually not observed when the same samples were analysed by LC-MS, suggesting that the reduction resulted from interactions with the MALDI matrix likely caused by the high laser energy. In addition, peaks of dehydrated peptides were frequently observed in the mass spectra (corresponding to a mass loss of -18 Da).

Since the cyclic peptide lacks *N*- and *C*-termini, MS/MS analysis to confirm the aspartyl residue as the site of hydroxylation was difficult. Proteolytic digestion trials of $CP1_{10mer}$ samples in order to create *N*- and *C*-termini and enable facile MS/MS analysis proved unsuccessful, as the cyclic peptide appeared resistant to proteolytic cleavage using conditions that worked for simultaneously digested control protein samples. Uncleaved cyclic peptide was observed in all samples after treatment with four tested proteases (trypsin, GluC, LysC, chymotrypsin). The rigid small ring structure of the cyclic peptide may hinder binding to the proteases.

NMR spectroscopy was then used to confirm that the hydroxylation occurred on the Asp residue in the cyclic peptide that corresponds to $Asp18_{39mer}$ in the hFX EGF1 peptide. A $CP1_{10mer}$ sample was fully hydroxylated by AspH using the typical hydroxylation assay. Complete hydroxylation was confirmed by MALDI-ToF-MS analysis. After reaction quenching, the sample was lyophilised and resuspended in ^2H_2O for NMR analysis. The NMR data acquisition and analysis was performed by Dr Richard J. Hopkinson. Resonances arising from the β-hydroxylated Asp residue were observed in two-dimensional heteronuclear single-quantum correlation spectroscopy (2D ^1H,^{13}C-HSQC) experiments with the hydroxylated peptide sample (**Fig. 4-5**).

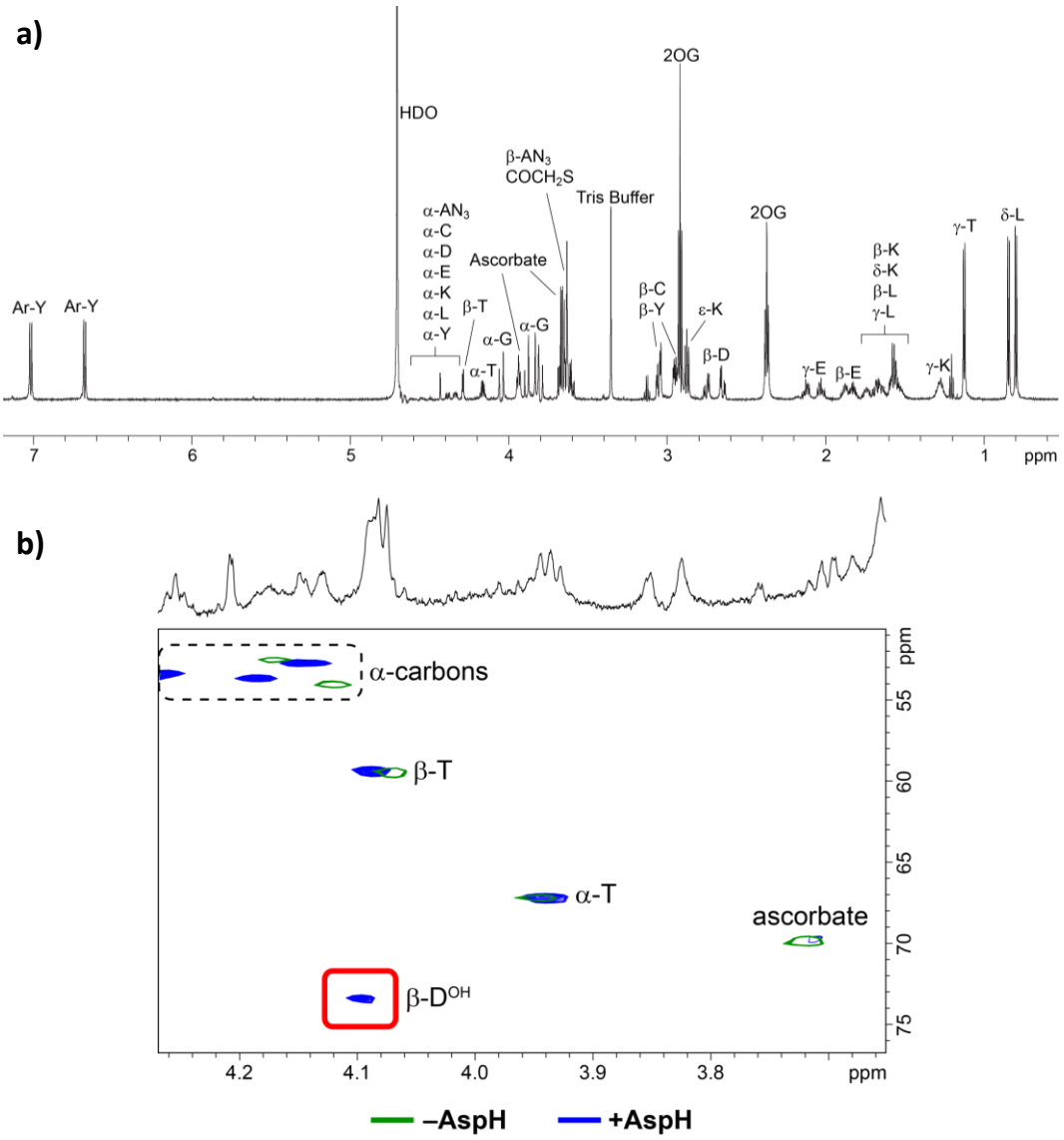


Fig. 4-5 | AspH-catalysed hydroxylation of the $CP1_{10mer}$ occurs at the Asp *β*-position. NMR spectroscopy analysis of the fully hydroxylated $CP1_{10mer}$ compared with the unhydroxylated peptide. Hydroxylated $CP1_{10mer}$ was produced by incubation with AspH using standard assay conditions. The peptide was subsequently lyophilised and resuspended in $^{2}H_2O$. **a)** ^{1}H NMR spectrum of the hydroxylated peptide; **b)** 2D ^{1}H,^{13}C-HSQC one-bond correlation spectrum of the hydroxylated (blue) compared with the unhydroxylated peptide (green). Shown is the *α*- and *β*-carbon region of interest. The resonance arising from the *β*-carbon of the hydroxylated Asp (*β*-D^{OH}, 73.2 ppm) is indicated, which is absent in the unhydroxylated peptide sample. Data acquisition and assignment were performed by Dr Richard J. Hopkinson who provided the figures.

4.2 Exploring the Cyclic Peptide Scaffold

4.2.1 Substrate Scope and Selectivity

The *in vitro* substrate scope and selectivity of AspH was then explored using the cyclic peptide scaffold as a model for an EGFD. Different cyclic peptides based on hFX EGF1 and other EGFDs have been synthesised and tested in assays with AspH. The key interaction points of AspH with its substrates are in the active site and the hydrophobic pocket within the TPR domain (see **Fig. 2-41**), as identified in the $AspH_{315-758}$-hFX $EGF1_{39mer}$ complex structure.

4.2.1.1 Cyclic Peptides Based on hFX EGF1$_{39mer}$

A selection of peptides based on the sequence of hFX EGF1 were synthesised (**Fig. 4-6**). Different residues were replaced by other proteinogenic or unnatural amino acids to test their importance and investigate the AspH substrate requirements, selectivity and scope of AspH catalysis.

Fig. 4-6 | Overview of synthesised peptides based on hFX EGF1. The changes made to the sequence are indicated (green). Either *R*-Ala or *R*-AzidoAla were incorporated in the linker moiety.

The cyclic peptides with both the linker amino acids *R*-AzidoAla in the CP1$_{10mer}$ and *R*-Ala in the CP2$_{10mer}$, respectively, were found to be AspH substrates. At first, the *R*-Ala linker position was replaced by a Gly residue (CP3$_{10mer}$) to investigate if the *R*-stereochemistry of this residue (that substitutes for the Cys in the original hFX EGF1$_{39mer}$) has any effect, or if the simple and cost-efficient natural amino acid (Gly) can be used for peptide synthesis instead. The peptides were tested side by side in hydroxylation assays with AspH. The CP3$_{10mer}$ with the Gly linker was an AspH substrate but showed significantly diminished hydroxylation compared to the peptides with *R*-AzidoAla or *R*-Ala linker (**Fig. 4-7**). Therefore, the *R*-amino acids were used for synthesis of further peptides.

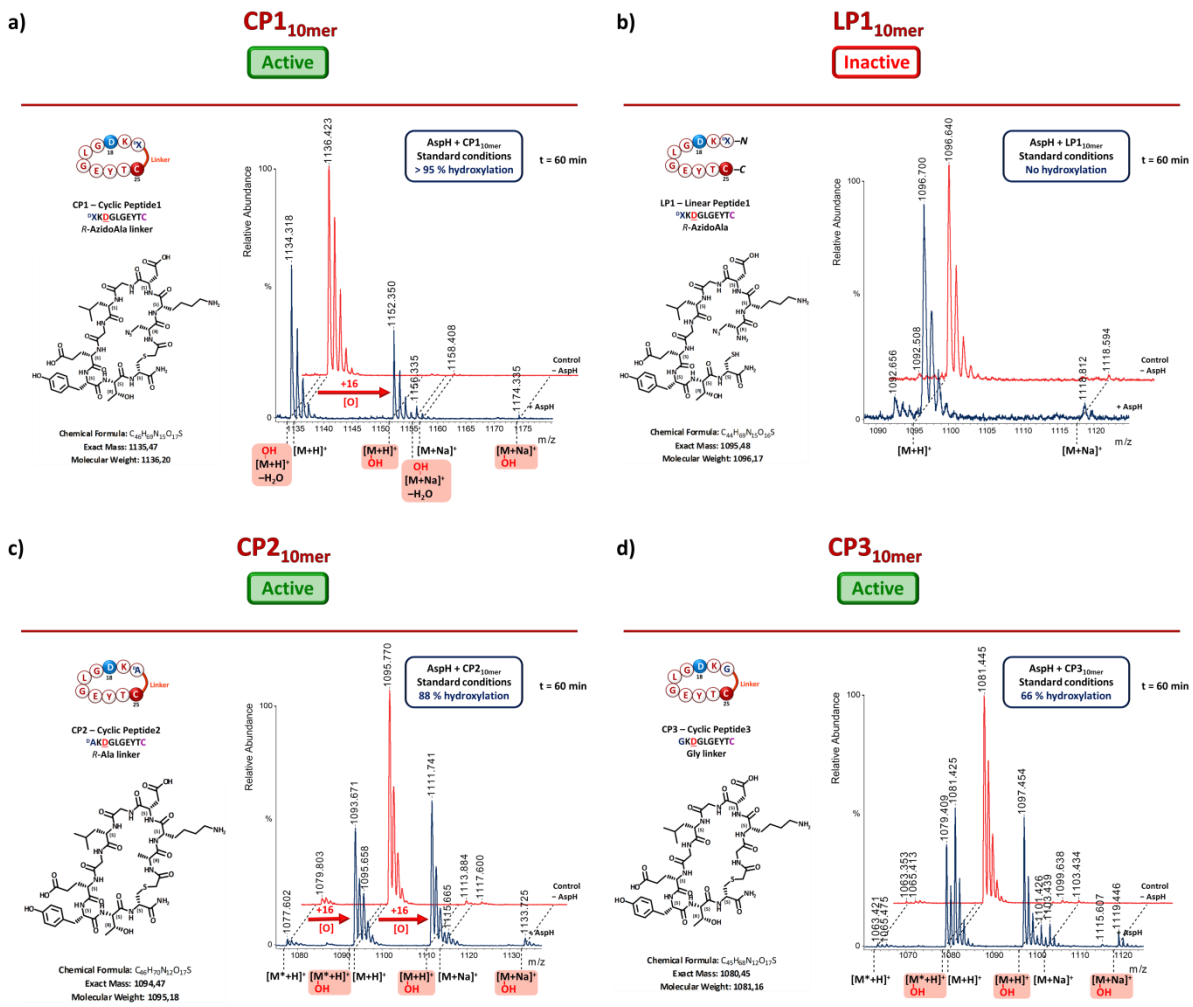


Fig. 4-7 | MALDI-ToF-MS results from activity assays of AspH with the cyclic and linear 10mer peptides synthesised based on hFX EGF1. **a)** AspH + $CP1_{10mer}$ with incorporated azide moiety (*R*-AzidoAla linker); **b)** AspH + $LP1_{10mer}$, the linear peptide corresponding to CP1; **c)** AspH + $CP2_{10mer}$ with *R*-Ala linker; **d)** AspH + $CP3_{10mer}$ with Gly linker. Incubations for 60 min were performed in using Standard Buffer conditions (10 mM HEPES, pH 7.5, 150 mM NaCl). $[M^*+H]^+$ refers to the derivatised, supposedly dehydrated, molecule species, which also gets hydroxylated (see **Chapter 6**).

Since AspH has been reported to catalyse hydroxylation of both Asp and Asn residues in EGFDs,[386] the Asp residue (**Fig. 4-8**) in the $CP1_{10mer}$ (*R*-AzidoAla linker) and $CP2_{10mer}$ (*R*-Ala linker) was replaced by an Asn residue. The peptide was synthesised in cyclic form with *R*-AzidoAla in the $CP4_{10mer}$ and *R*-Ala in the $CP5_{10mer}$ and in linear form ($LP4_{10mer}$). Both the $CP4_{10mer}$ and the $CP5_{10mer}$ were efficient AspH substrates and nearly fully hydroxylated (**Fig. 4-9**). No activity was observed with the $LP4_{10mer}$, consistent with the results obtained with the Asp-containing $LP1_{10mer}$.

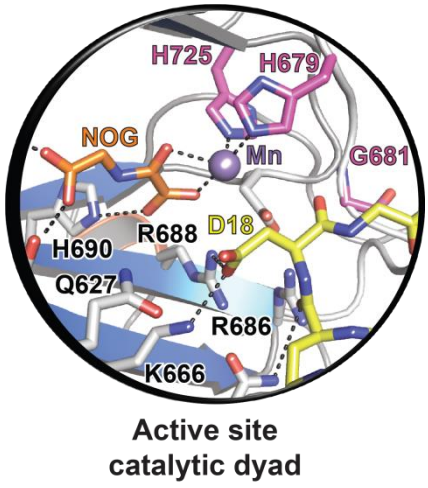


Fig. 4-8 | The AspH active site. $Asp18_{39mer}$ in the hFX EGF1 substrate was replaced by other amino acids in cyclic peptides to investigate substrate requirements, selectivity and scope.

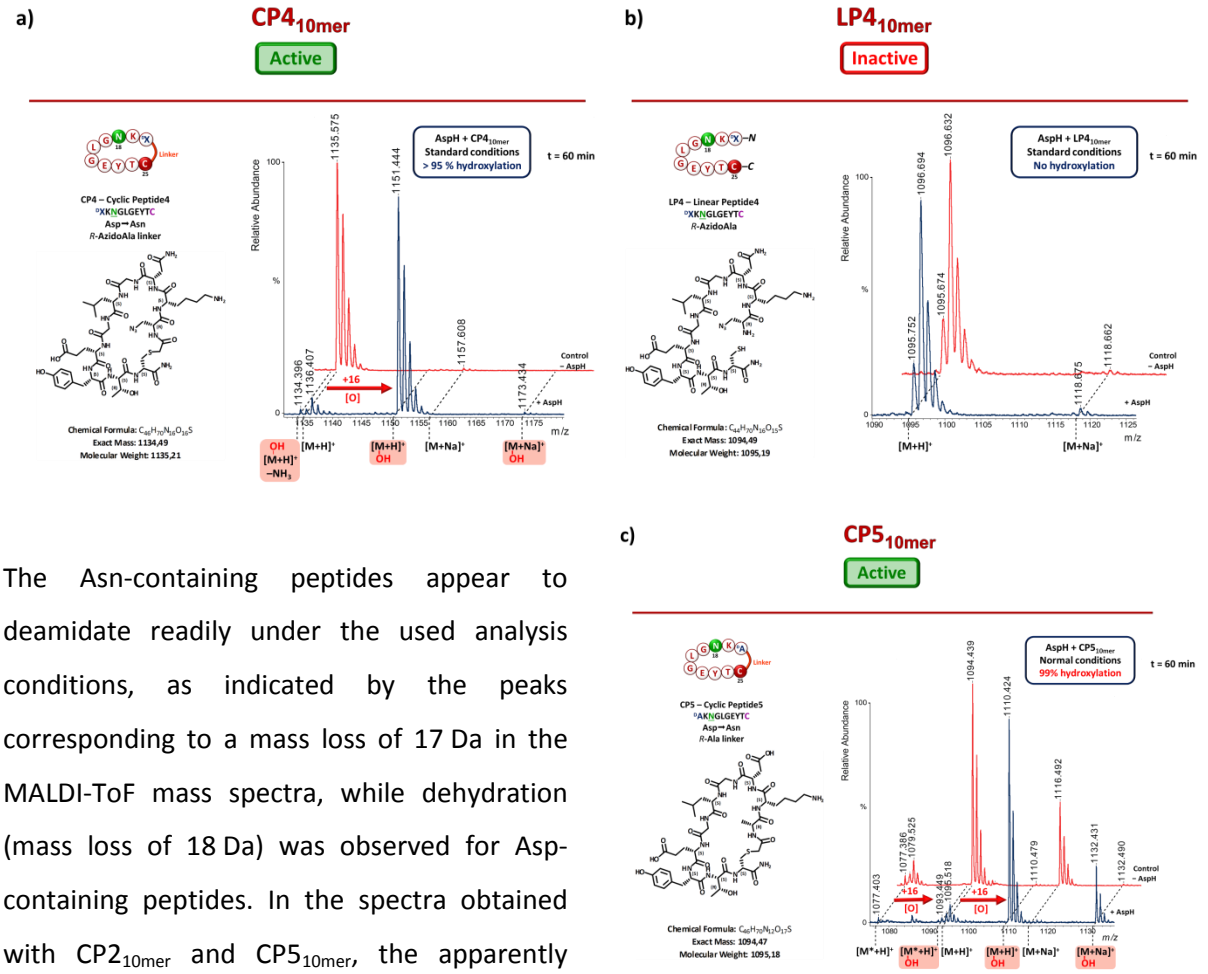


Fig. 4-9 | MALDI-ToF-MS results from activity assays of AspH with cyclic and linear peptides with Asn in place of the Asp residue. **a)** AspH + $CP4_{10mer}$ (*R*-AzidoAla linker); **b)** AspH + $LP4_{10mer}$ (linear peptide); **c)** AspH + $CP5_{10mer}$ (*R*–Ala linker).

The Asn-containing peptides appear to deamidate readily under the used analysis conditions, as indicated by the peaks corresponding to a mass loss of 17 Da in the MALDI-ToF mass spectra, while dehydration (mass loss of 18 Da) was observed for Asp-containing peptides. In the spectra obtained with $CP2_{10mer}$ and $CP5_{10mer}$, the apparently dehydrated or deamidated peptide species respectively ($[M^*+H]^+$) were also hydroxylated, as already observed with the hFX $EGF1_{39mer}$ based 39mer and 26mer peptides. These observations support the proposal that the dehydrated/deamidated species may have already existed in solution and have not been generated due to in-source fragmentation (see further investigations in **Chapter 6**).

The requirement for the conserved EGFD Tyr/Phe residue that binds in the hydrophobic pocket within the AspH TPR domain (**Fig. 4-10**) was also investigated. The residue corresponding to $Tyr23_{39mer}$ in the hFX EGF1 substrate appeared to have a key role for enzyme-substrate interaction in the crystal structure and is part of the consensus sequence ($-C_3X(D/N)XXXX(Y/F)XC_4XC_5-$).[338] The

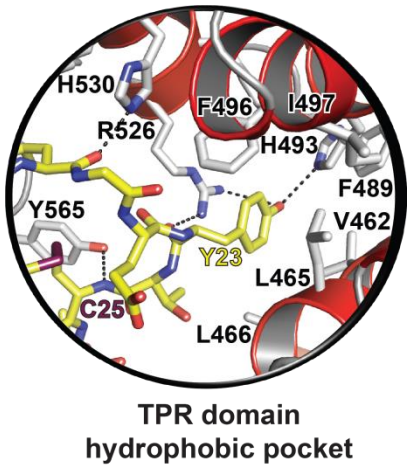


Fig. 4-10 | The hydrophobic pocket within the AspH TPR domain. $Tyr23_{39mer}$ in the hFX EGF1 substrate was replaced by Ala and Val.

residue was replaced by Ala ($CP6_{10mer}$) and Val ($CP7_{10mer}$) residues and the resultant peptide tested for activity with AspH. The $CP6_{10mer}$ showed no reaction, indicating that the hydrophobic interaction is critical for substrate recognition (**Fig. 4-11a**). The hydrophobic interaction with the Val residue in the $CP7_{10mer}$ was sufficient for AspH recognition, but only poor hydroxylation (5 %) was observed within the assay time (**Fig. 4-11b**), consistent with the results from reported cell-based EGFD mutagenesis studies.[494]

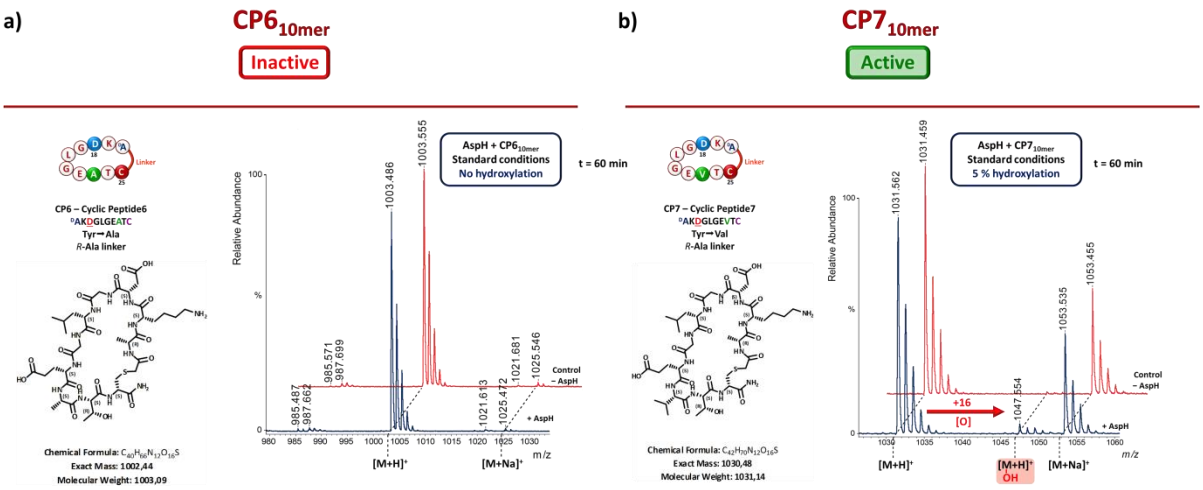


Fig. 4-11 | MALDI-ToF-MS results from activity assays of AspH with cyclic peptides synthesised based on hFactor X with change of the Tyr to an Ala or Val residue respectively. **a)** AspH + $CP6_{10mer}$ (Tyr ⇨ Ala). No hydroxylation was observed. **b)** AspH + $CP7_{10mer}$ (Tyr ⇨ Val). The peptide was a poor substrate.

The AspH substrate scope and selectivity were then investigated by substitution of the hydroxylation site Asp/Asn residue. The hydroxylation site Asp residue in the hFX $EGF1_{39mer}$-based cyclic peptide was replaced by Glu ($CP8_{10mer}$), Ser ($CP10_{10mer}$), and Pro ($CP13_{10mer}$) that could potentially still fit in the active site with their β-position. However, no activity was observed with any of these peptides (**Fig. 4-12**). Substitution of the hydroxylation site residue with unnatural amino acids of similar size, such as *R*-Asp ($CP9_{10mer}$), *S*-2-aminobutyric acid (Abu, $CP11_{10mer}$), and *S*-homoserine (Hse, $CP12_{10mer}$) also resulted in no activity (**Fig. 4-12**). These results suggested that AspH is selective for Asp and Asn residues. However, the cyclic peptide scaffold is an *in vitro* model, and different results may be obtained using whole EGFD proteins.

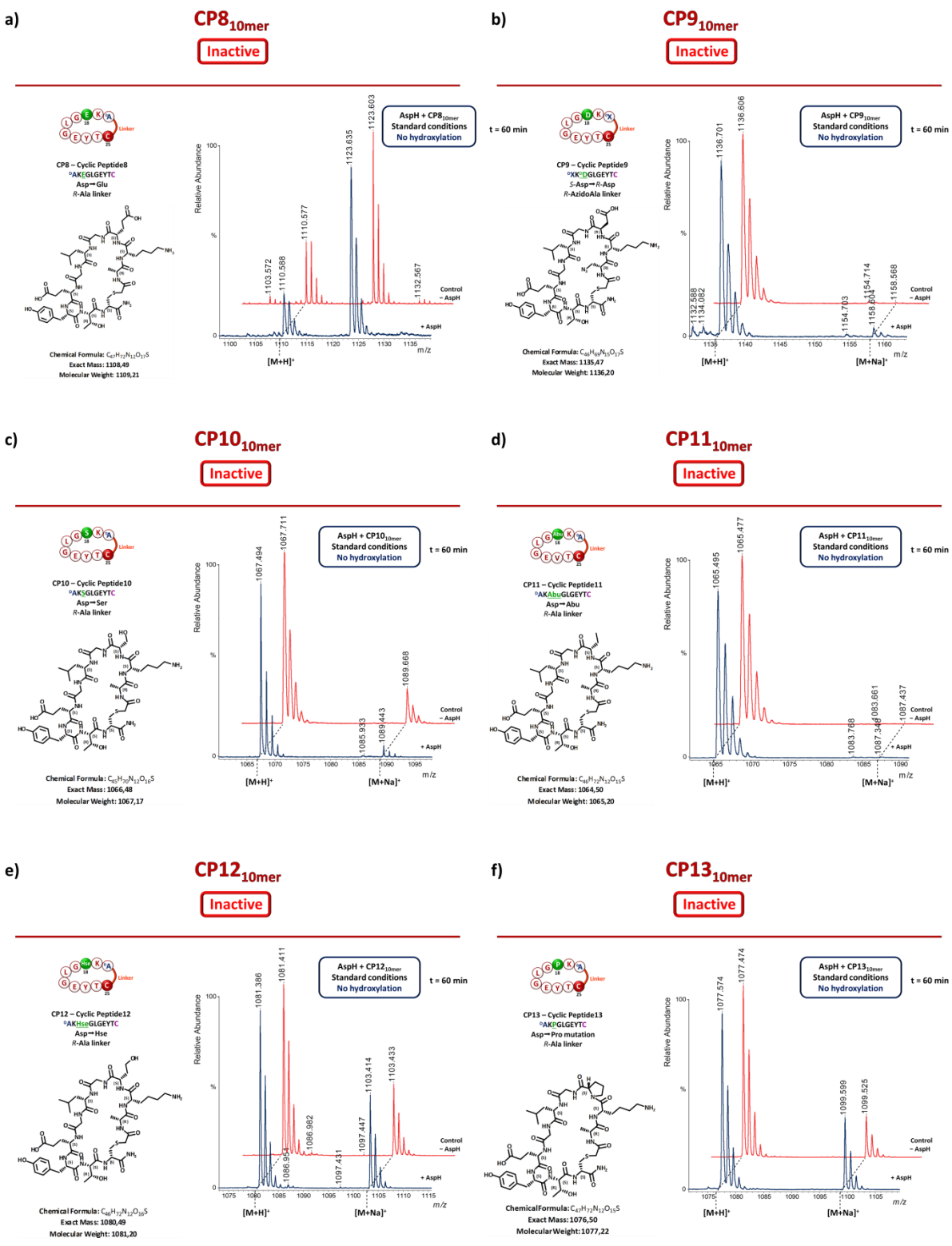


Fig. 4-12 | MALDI-ToF-MS results from activity assays of AspH with cyclic peptides containing hydroxylation site residue substitutes. **a)** AspH + $CP8_{10mer}$ (Asp ⇨ Glu); **b)** AspH + $CP9_{10mer}$ (L-Asp Asp ⇨ D-Asp); **c)** AspH + $CP10_{10mer}$ (Asp ⇨ Ser); **d)** AspH + $CP11_{10mer}$ (Asp ⇨ Abu)); **e)** AspH + $CP12_{10mer}$ (Asp ⇨ Hse)); **f)** AspH + $CP13_{10mer}$ (Asp ⇨ Pro). No hydroxylation was observed with either peptide.

Several of the inactive cyclic peptides were also tested as AspH inhibitors. However, none of the peptides inhibited $CP2_{10mer}$ hydroxylation when both peptides were simultaneously incubated with AspH (**Fig. 4-13**).

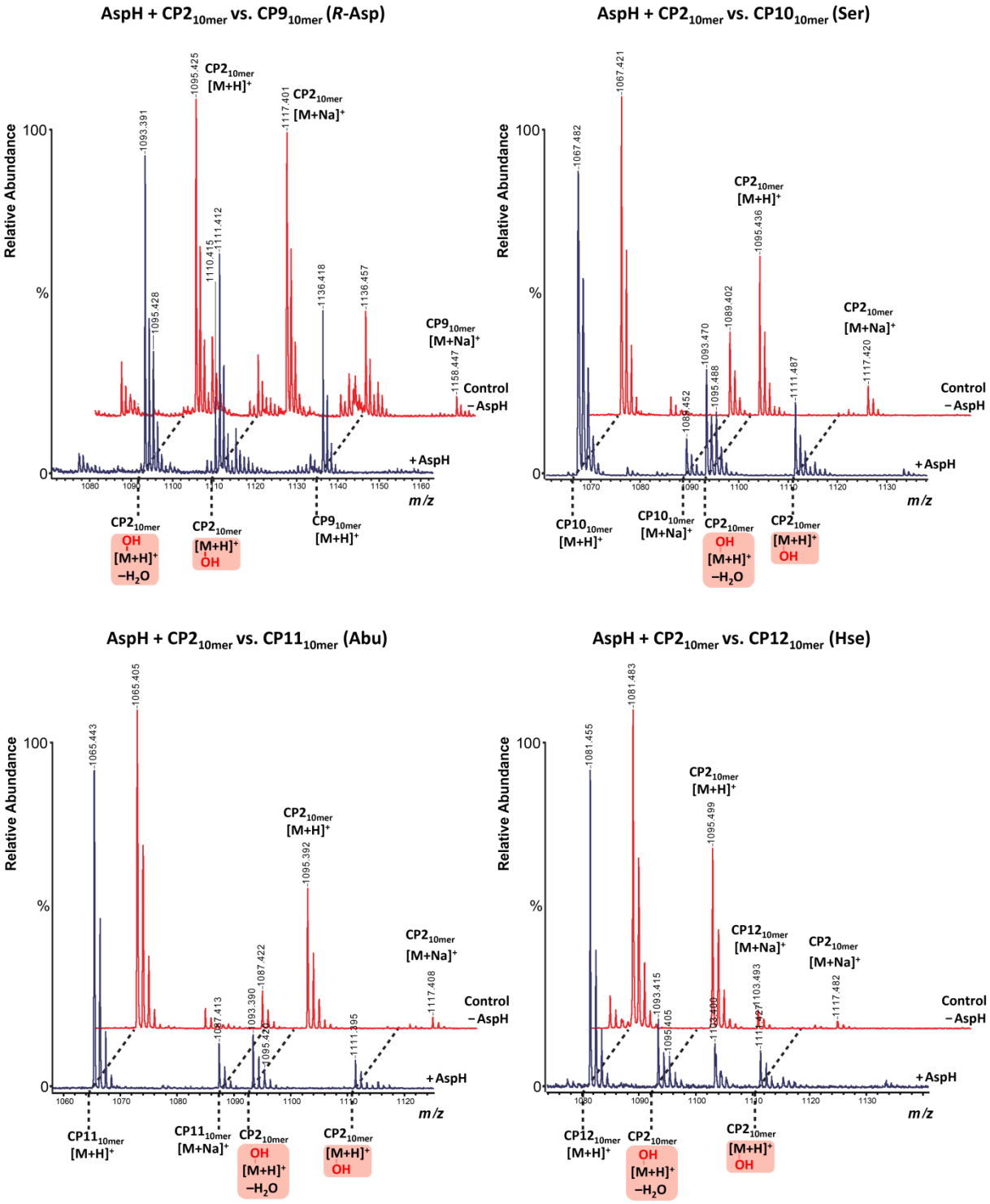


Fig. 4-13 | MALDI-ToF-MS results from activity assays of AspH simultaneously incubated with $CP2_{10mer}$ (20 µM) and inactive cyclic peptides (100 µM) in order to test these as AspH inhibitors. None of the tested peptides had any significant effect on hydroxylation of the $CP2_{10mer}$.

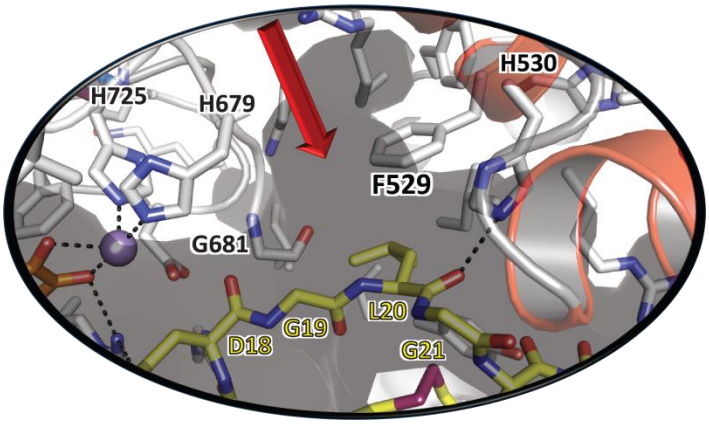


Fig. 4-14 | View from the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ crystal structure showing a cavity (dark grey cavity surface representation and indicated with the red arrow) close to the active site.

In the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ crystal structure, a cavity close to the active site was observed (**Fig. 4-14**). It was considered that substitution of the glycine residue corresponding to $Gly19_{39mer}$ in the hFX $EGF1_{39mer}$ substrate with a more bulky residue, could possibly increase the cyclic peptide binding affinity by interaction with the cavity surrounding residues. Further, this information could be useful for selective AspH inhibitor development or probes design. The residue in position of $Gly19_{39mer}$ was therefore replaced by Tyr ($CP14_{10mer}$) and Arg ($CP15_{10mer}$) to enable potential hydrophobic and possibly π-π-stacking or cation-π-stacking interactions between the residue side chain and the AspH Phe529 side chain in close proximity. Both peptides were AspH substrates, but while the $CP14_{10mer}$ was nearly fully hydroxylated, the $CP15_{10mer}$ was a poor substrate (**Fig. 4-15**), possibly due to clashes with the side chains of close protein residues.

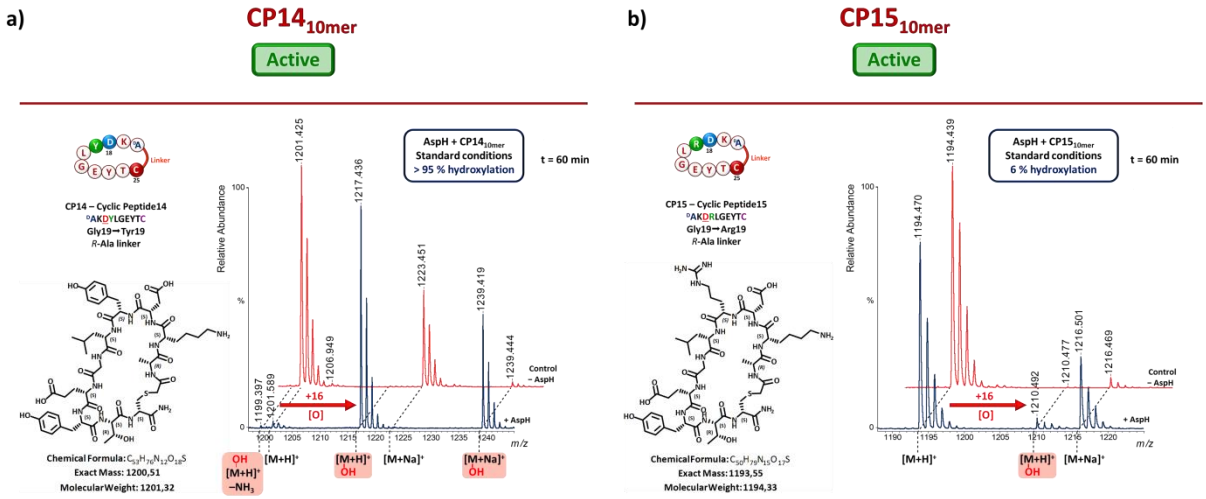


Fig. 4-15 | MALDI-ToF-MS results from activity assays of AspH with cyclic peptides with replaced residues to fill the cavity surrounding $Gly19_{39mer}$ in the bound hFX EGF1 substrate in the crystal structure. **a)** AspH + $CP14_{10mer}$ (Gly19 ⇨ Tyr19); **b)** AspH + $CP15_{10mer}$ (Gly19 ⇨ Arg19).

4.2.1.2 Cyclic Peptides Based on Other EGFDs

Based on the EGFD substrates from Professor Penny A. Handford's lab (Department of Biochemistry, University of Oxford), cyclic peptides corresponding to some of the EGFDs within the repeats were synthesised to see if AspH can tolerate other sequences in cyclic peptides than the hFX EGF1 based ones. The cyclic peptides corresponding to the Ca^{2+}-binding EGFDs in hFib1 cbEGF23 ($CP16_{10mer}$), hNotch1 cbEGF13 ($CP17_{10mer}$), and hFib1 cbEGF42 ($CP18_{10mer}$) were synthesised. The sequences all contain the hydroxylation consensus residues (Asp/Asn and Tyr/Phe in the correct positions) but differ strongly in their other residues from the hFX $EGF1_{39mer}$ sequence.

Due to its poor solubility, the very hydrophobic $CP16_{10mer}$ was not purified by HPLC, but the crude peptide was used for assays with AspH. The crude peptide was dissolved in 10% dimethyl sulfoxide (DMSO) instead of water, and this solution was then used for the preparation of the assay mixtures as usal.

$CP16_{10mer}$, $CP17_{10mer}$, and $CP18_{10mer}$ were AspH substrates and got hydroxylated to different extents (**Fig. 4-16**), suggesting that a 10mer cyclic peptide substrate can be made for any EGFD that contains the hydroxylation site Asp/Asn and the TPR domain-binding Tyr/Phe residue. These results were consistent with the hydroxylation consensus sequence proposed by Stenflo and coworkers.[338]

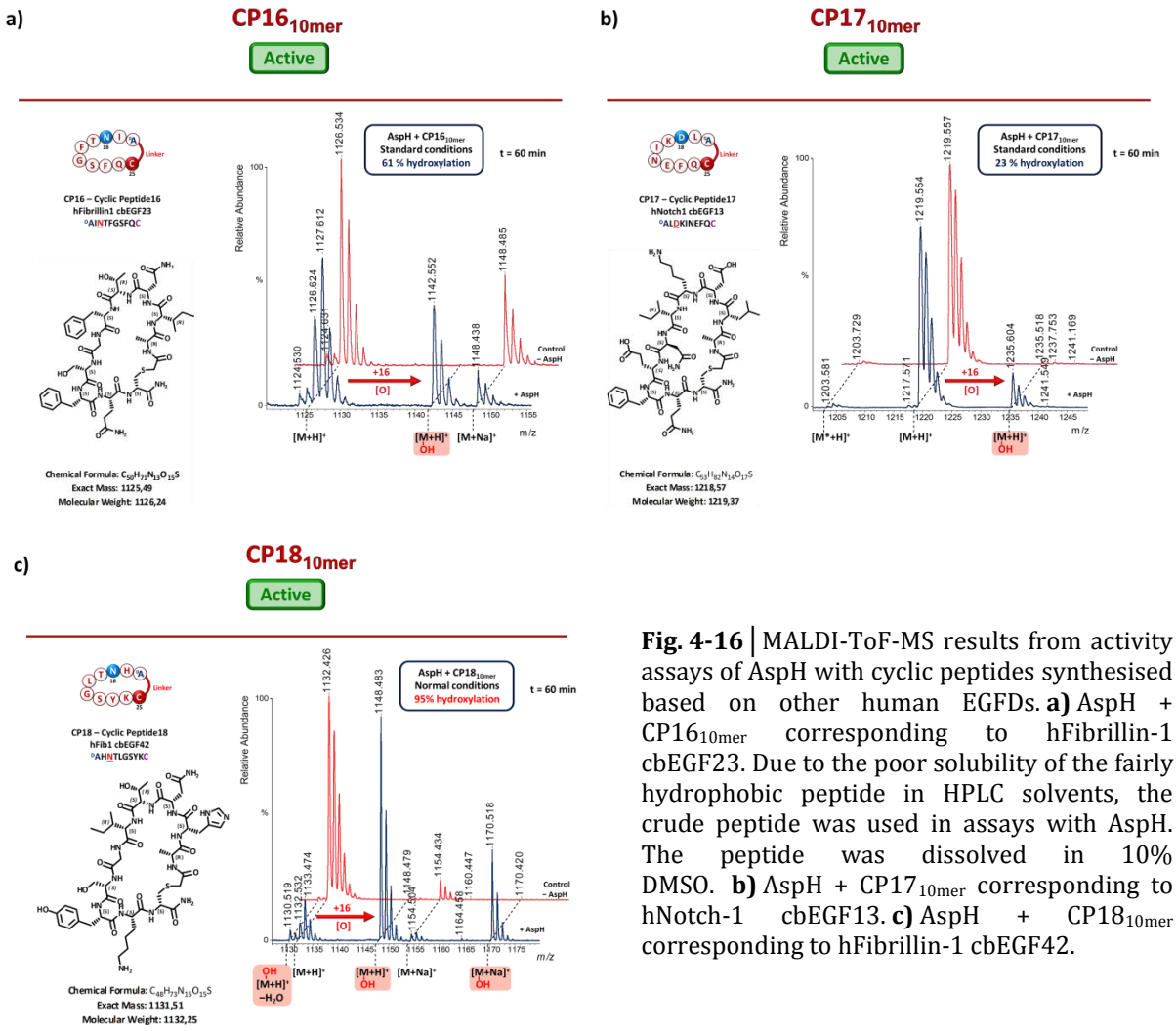
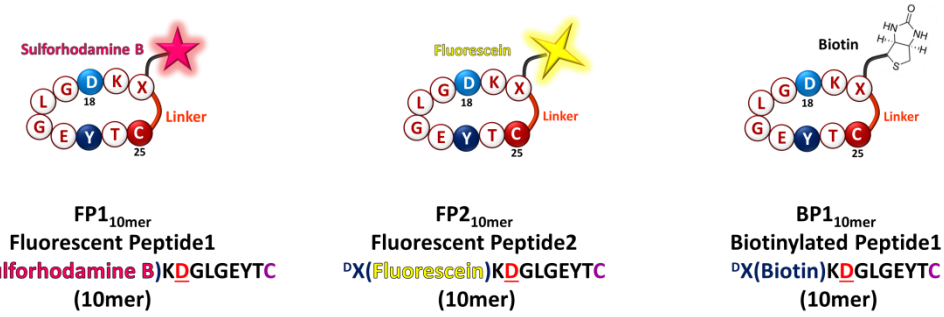


Fig. 4-16 | MALDI-ToF-MS results from activity assays of AspH with cyclic peptides synthesised based on other human EGFDs. **a)** AspH + $CP16_{10mer}$ corresponding to hFibrillin-1 cbEGF23. Due to the poor solubility of the fairly hydrophobic peptide in HPLC solvents, the crude peptide was used in assays with AspH. The peptide was dissolved in 10% DMSO. **b)** AspH + $CP17_{10mer}$ corresponding to hNotch-1 cbEGF13. **c)** AspH + $CP18_{10mer}$ corresponding to hFibrillin-1 cbEGF42.

4.2.2 Cyclic Peptide Conjugates

Fluorescent or biotinylated probes for AspH based on the cyclic peptide may be useful tools for *in vitro* (e.g. fluorescence polarisation) and cell-based assays. For the synthesis of such probes, the cyclic peptide was conjugated to fluorescence dyes such as sulforhodamine B (this peptide was synthesised and purified by Dr Martin Münzel) and fluorescein and to a biotin label (**Fig. 4-17**).



$FP1_{10mer}$
Fluorescent Peptide1
^{D}X(Sulforhodamine B)KDGLGEYTC
(10mer)

$FP2_{10mer}$
Fluorescent Peptide2
^{D}X(Fluorescein)KDGLGEYTC
(10mer)

$BP1_{10mer}$
Biotinylated Peptide1
^{D}X(Biotin)KDGLGEYTC
(10mer)

Fig. 4-17 | Schematics of small cyclic peptide probes as useful tools for *in vitro* and cell-based studies.

The cyclic peptide conjugates were synthesised using strain-promoted azide-alkyne [3+2] cycloaddition[495,496] facilitated by the incorporated azide moiety and a bicyclo[6.1.0]nonyne (BCN)-conjugate (**Fig. 4-18**).[497,498] In contrast to the azide-alkyne Huisgen cycloaddition,[499] no copper(I) catalyst is required for the strain-promoted 1,3-dipolar cycloaddition. The reaction proceeds under mild conditions at 37 °C or at room temperature within 24 h in a H_2O/CH_3CN solvent system. Due to the poor solubility of the fluorescein-BCN-conjugate, methanol had to be added as an additional solvent, as described in **Chapter 10.** The cyclic peptide conjugates could be useful probes in further AspH studies, e.g. by enabling the development of a fluorescence polarisation or cell-based assays.

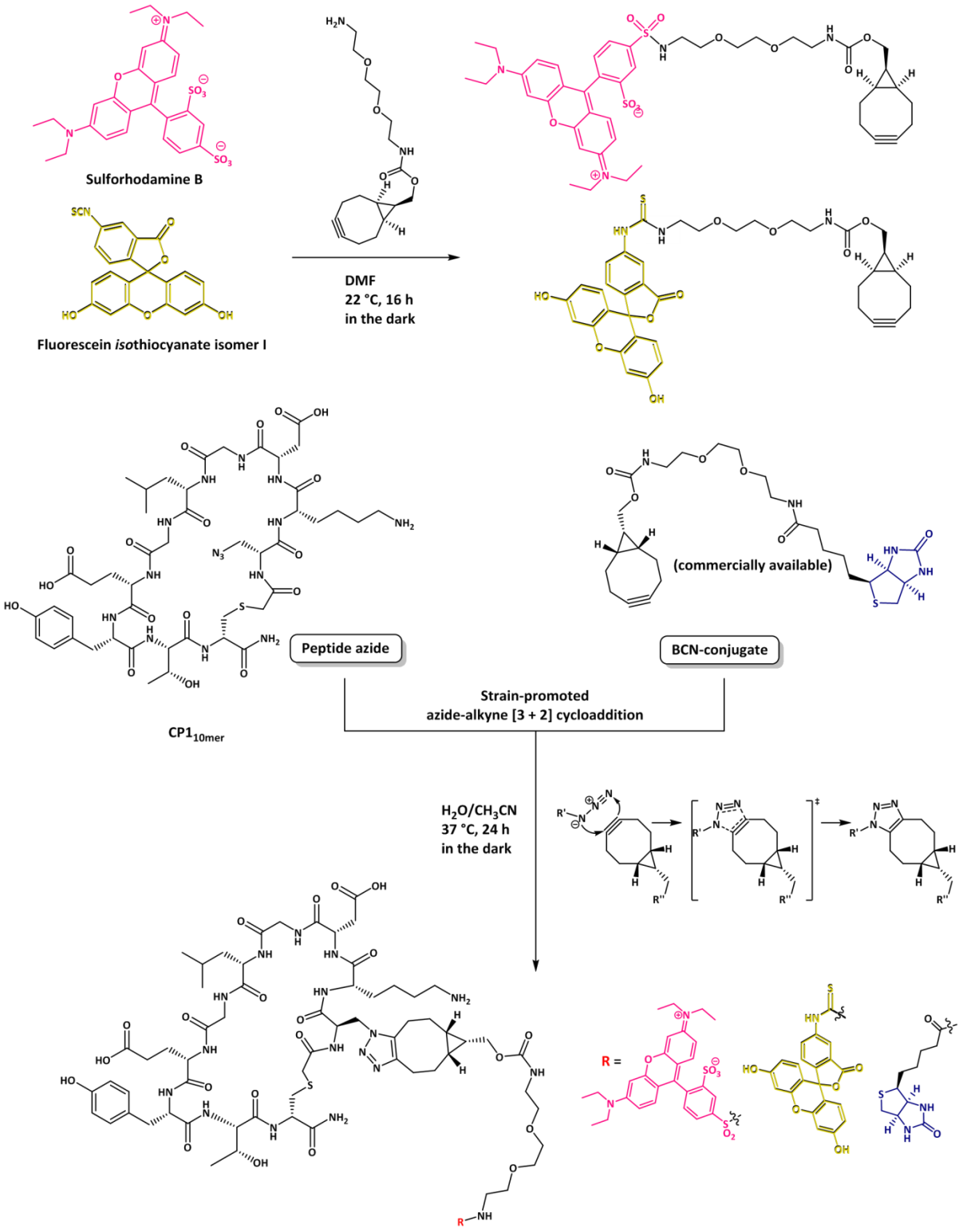


Fig. 4-18 | Conjugation of fluorescence dyes or a biotin label to the cyclic peptide by copper-free strain-promoted azide-alkyne cycloaddition. The incorporated azide moiety in the cyclic peptide and a fluorescence dye/biotin-bicyclo[6.1.0]nonyne (BCN)-conjugate react under [3+2] cycloaddition to the conjugated peptide under mild conditions. DMF = *N*,*N*-dimethylformamide.

The $FP1_{10mer}$ was an efficient AspH substrate in hydroxylation activity assays (**Fig. 4-19**). As a proof of principle, the $FP1_{10mer}$ hydroxylation demonstrated the feasibility of the attachment of molecules to the cyclic peptide scaffold in a conformationally sensible position. Time course studies

showed that the reaction of AspH with the $FP1_{10mer}$ was slow compared to the reaction with the $CP1_{10mer}$. A possible reason could be the steric hindrance of substrate recognition by the bulky fluorescence dye tail.

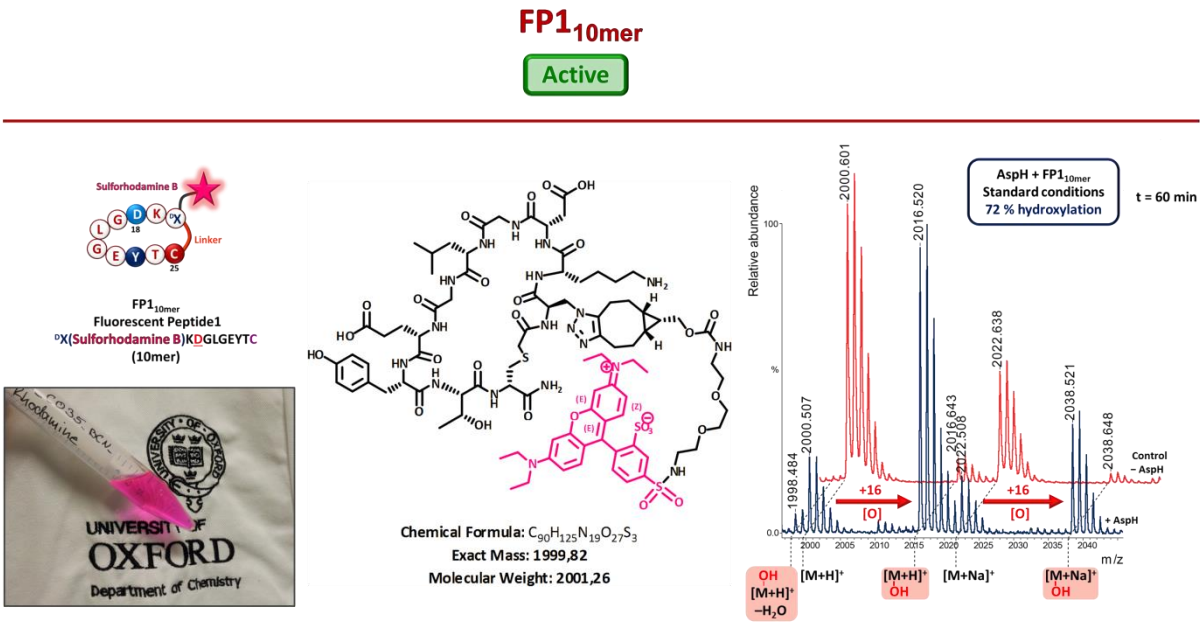


Fig. 4-19 | The $FP1_{10mer}$ (hFX $EGF1_{39mer}$-based $CP1_{10mer}$ conjugated to the fluorescence dye sulforhodamine B) is an AspH substrate. The photograph shows the peptide in solution after HPLC purification. The background oxidation found in the control samples, which could be due to sulfur oxidation, was subtracted.

No hydroxylation activity was observed with the $FP2_{10mer}$ with the attached fluorescein label (**Fig. 4-20a**). However, this may be due to the poor solubility of the peptide in aqueous solutions. The peptide was only soluble in solutions containing high amounts of methanol or DMSO, as observed during the synthesis. The peptide suspension in the buffer used for the activity assays may have therefore not been amenable to AspH hydroxylation. The reaction buffer system was not optimised at this stage of the studies, and the $FP2_{10mer}$ may show activity under addition of higher amounts of solvents.

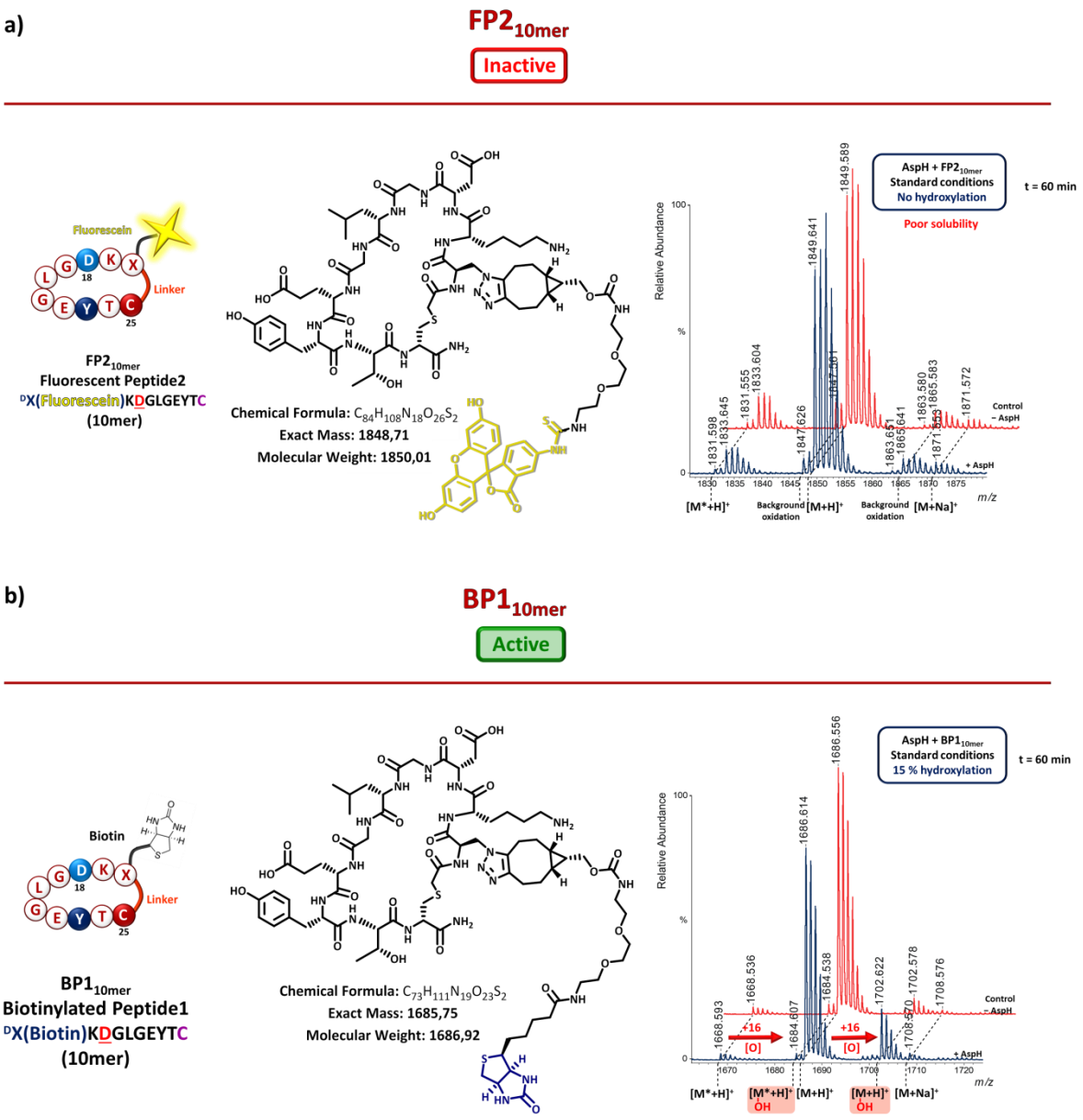


Fig. 4-20 | MALDI-ToF-MS results from activity assays of AspH with cyclic peptide conjugates. **a)** No activity was observed with the $FP2_{10mer}$ (hFX EGF1-based cyclic peptide conjugated to the fluorescence dye fluorescein) under the tested assay conditions, likely due to the poor solubility of the peptide in aqueous solutions. The reaction buffer system may require addition of solvents such as DMSO or methanol but was not optimised at this stage of the studies. **b)** The $BP1_{10mer}$ was an AspH substrate under the tested conditions. $[M^*+H]^+$ refers to the derivatised, supposedly dehydrated, molecule species, which also gets hydroxylated (see **Chapter 6**). The background oxidation found in the control samples, which could be due to sulfur oxidation, was subtracted.

The $BP1_{10mer}$ with attached biotin label was an AspH substrate (15% hydroxylation, **Fig. 4-20b**) using standard conditions, further supporting the utility of the conjugation technology. The hydroxylation levels were comparably low, possibly due to the higher flexibility of the relatively long biotin linker and the resulting steric hindrance of substrate recognition. The peptide was further optimised at a later stage of the studies (see **Chapter 5**).

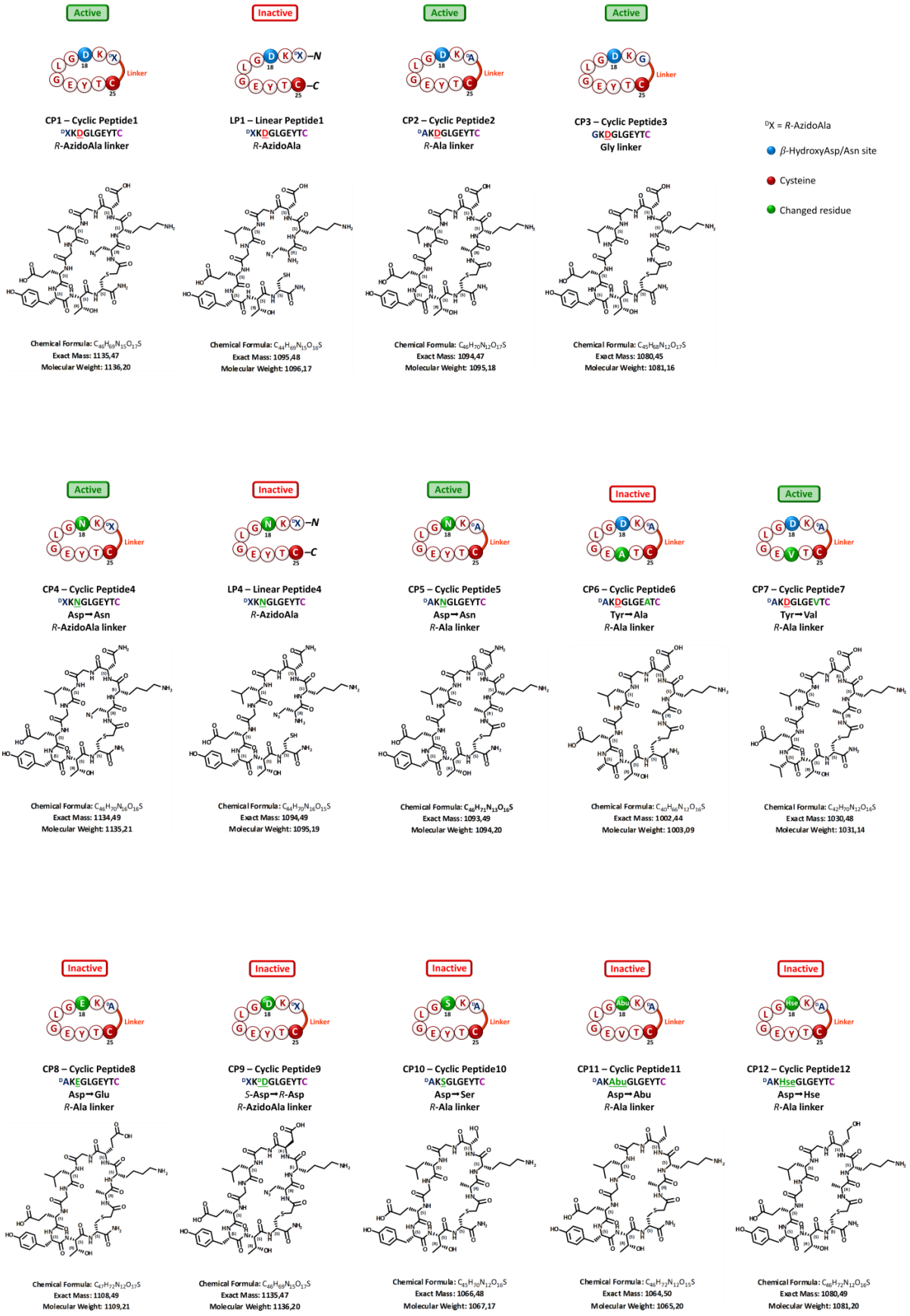


Fig. 4-21 | Activity overview of the synthesised cyclic and linear 10mer peptides reported on the preceding pages at a glance (Part A).

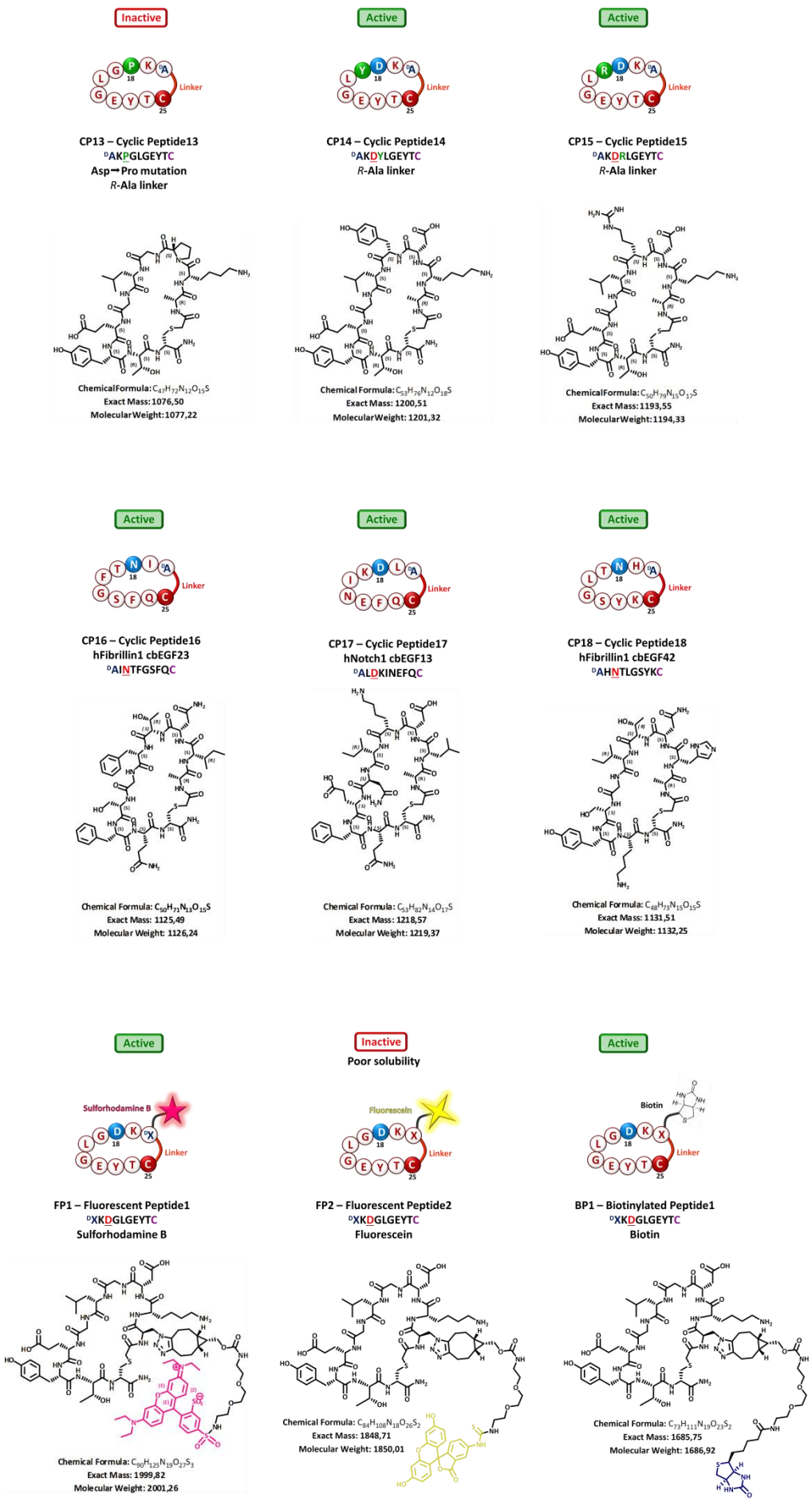


Fig. 4-22 | Activity overview of the synthesised cyclic 10mer peptides reported on the preceding pages at a glance (Part B).

Co-crystallisation trials of His_6-$AspH_{315-758}$ with the $CP1_{10mer}$ and the $CP2_{10mer}$ were then set up. It was considered that a crystal structure of AspH in complex with a synthesised cyclic peptide would verify the proposal that the enzyme-bound peptide adopts the binding mode predicted in the model. However, mostly haystack-shaped crystals were grown in similar well conditions as the "apo"-$AspH_{315-758}$ crystals. Some of the crystals were harvested and diffracted to 3.5 Å resolution. However, after molecular replacement no peptide electron density was apparent in the structure. Further crystallisation trials were set up after cyclic peptide optimisation (see **Chapter 5**).

4.3 Competition Experiments with AspH Substrates

For further characterisation of the AspH substrates in hand, substrate competition experiments were conducted in order to test whether AspH has a preferred substrate. AspH was incubated simultaneously with two peptide substrates. In each assay, the $CP1_{10mer}$ and the commercially obtained longer hFX $EGF1_{39mer}$-based peptides were present at the same concentration (final concentration of each peptide in the assay: 100 µM; final AspH concentration: 4 µM (enzyme-to-substrate ratio 1:25:25). The competition experiments were performed as time courses. Mass spectra of samples taken after 12 min incubation with AspH are representatively shown in **Figs. 4-23** to **4-29** to depict to which extent the substrates were hydroxylated after this incubation time. The levels of hydroxylation at this time point indicated the AspH substrate preferences under the tested conditions in the competition experiments.

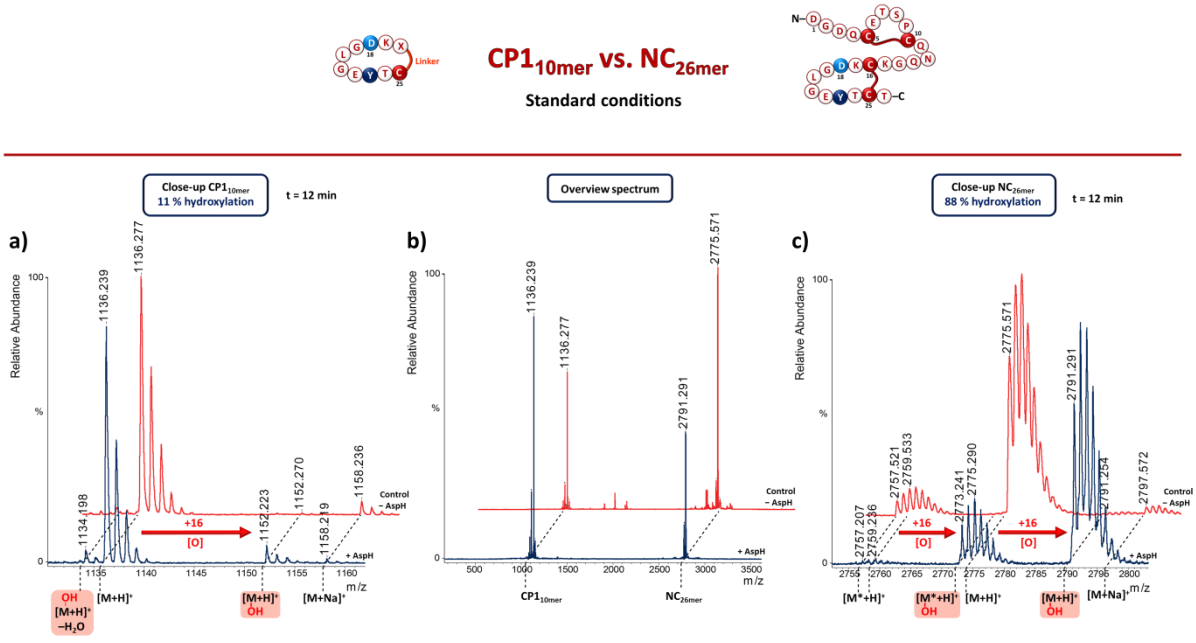


Fig. 4-23 | MALDI-ToF-MS results from competition assays in presence of both $CP1_{10mer}$ and NC_{26mer} (AspH + $CP1_{10mer}$ + NC_{26mer}, 1:25:25 ratio) in Standard Buffer conditions. Spectra obtained after 12 min showed that the 26mer peptide (~88 % hydroxylation) was clearly the preferred substrate since it is already almost fully hydroxylated after this time, whereas only low levels of hydroxylated $CP1_{10mer}$ were observed (~11 % hydroxylation).

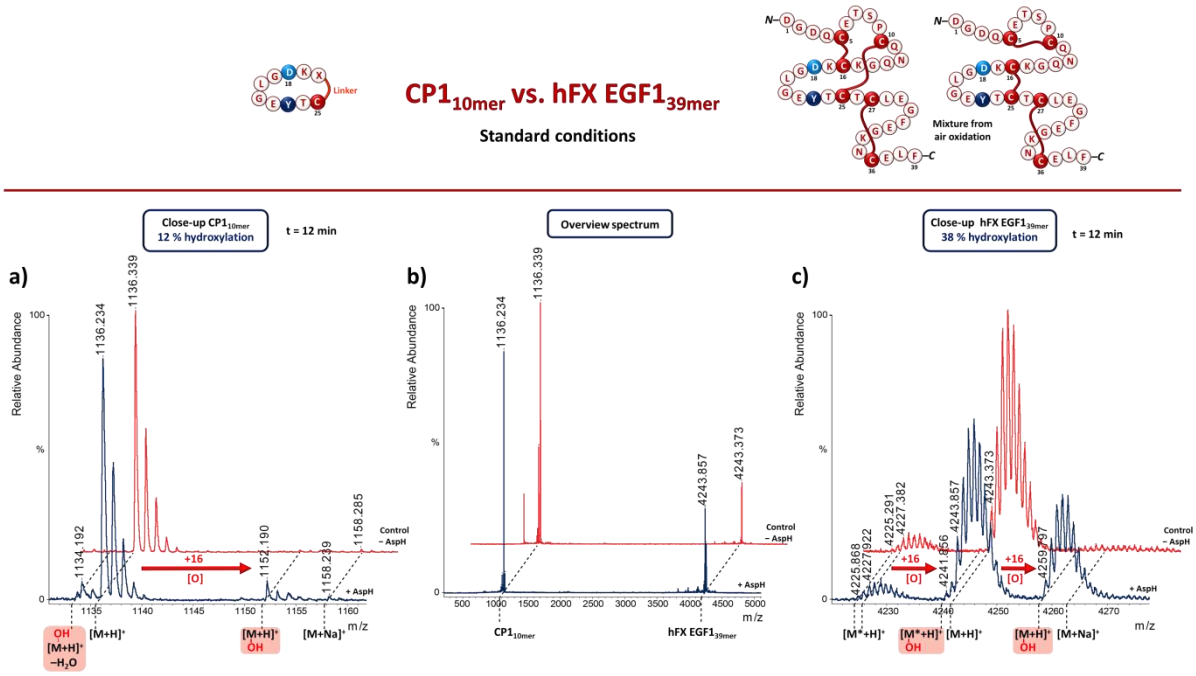


Fig. 4-24 | MALDI-ToF-MS results from competition assays in presence of both $CP1_{10mer}$ and the hFX $EGF1_{39mer}$ peptide mixture (AspH + $CP1_{10mer}$ + hFX $EGF1_{39mer}$, 1:25:25 ratio) in Standard Buffer conditions. Spectra obtained after 12 min showed that the hFX $EGF1_{39mer}$ peptide mixture of the canonical and non-canonical forms (C+NC_{39mer} mixture, ~38 % hydroxylation) is the preferred substrate since the non-canonical share of the mixture seemed to be nearly fully hydroxylated after this time, whereas the $CP1_{10mer}$ was hydroxylated later, as seen by the much smaller proportion of hydroxylated $CP1_{10mer}$ (~12 % hydroxylation).

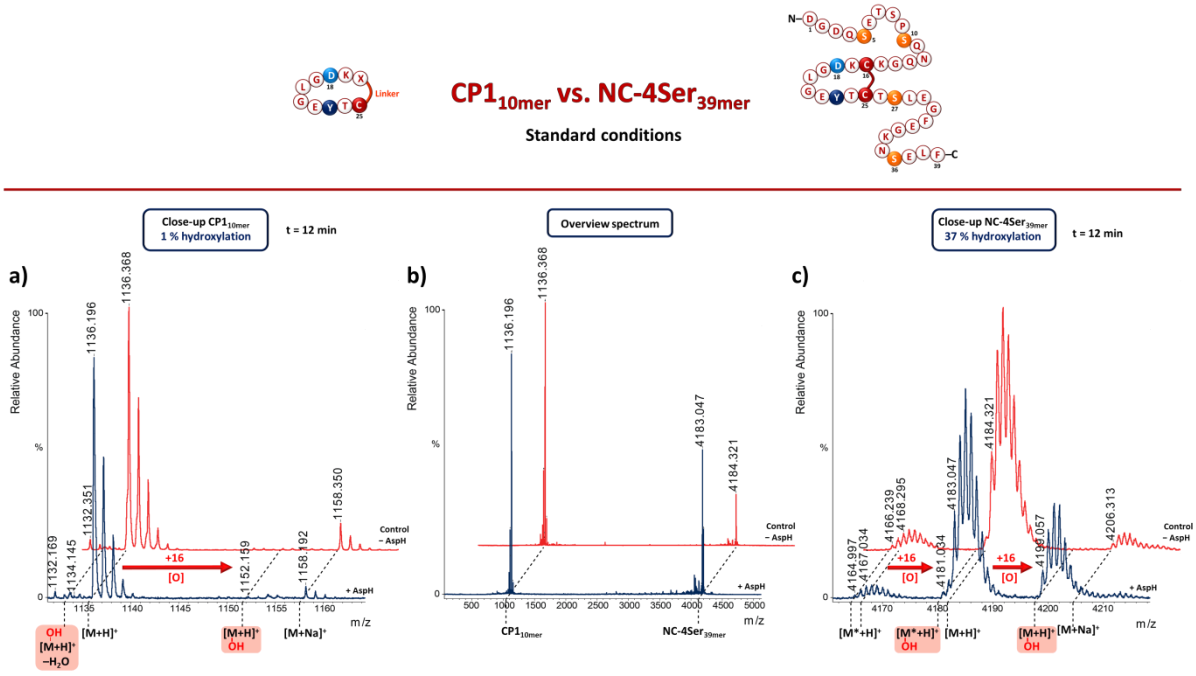


Fig. 4-25 | MALDI-ToF-MS results from competition assays in presence of both $CP1_{10mer}$ and $NC\text{-}4Ser_{39mer}$ (AspH + $CP1_{10mer}$ + $NC\text{-}4Ser_{39mer}$, 1:25:25 ratio) in Standard Buffer conditions. Spectra obtained after 12 min showed that $NC\text{-}4Ser_{39mer}$ (~37 % hydroxylation) is clearly the preferred substrate, since almost no $CP1_{10mer}$ hydroxylation was observed at this time point (~1 % hydroxylation).

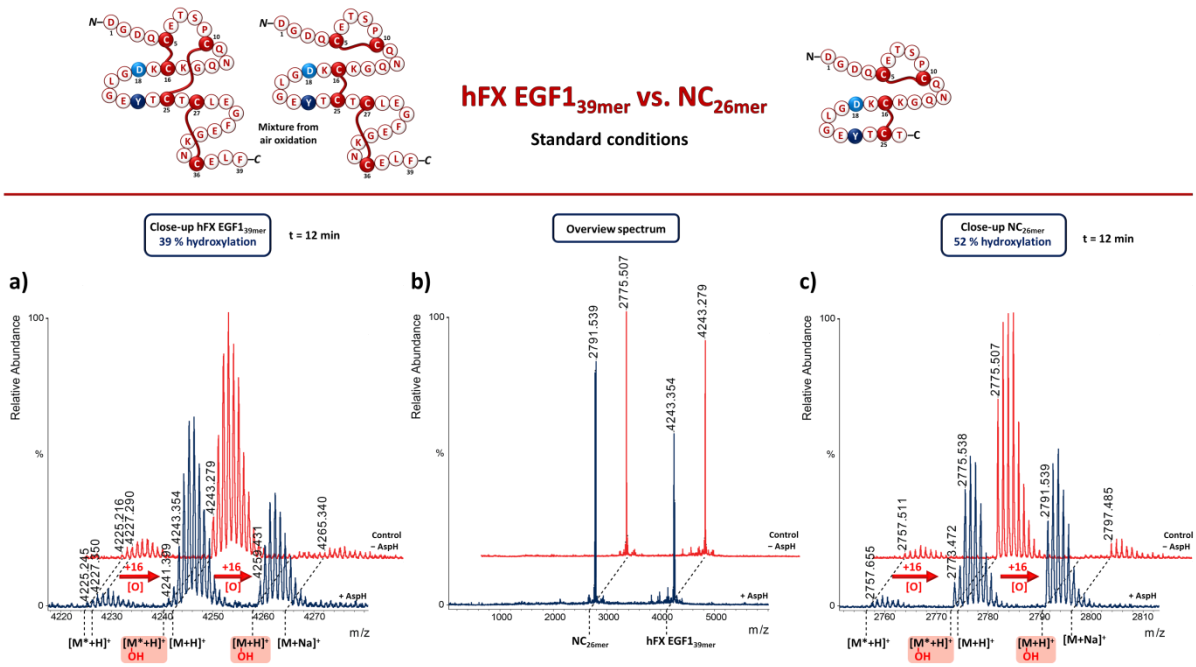


Fig. 4-26 | MALDI-ToF-MS results from competition assays in presence of hFX EGF1$_{39mer}$ (C+NC$_{39mer}$ mixture) and NC$_{26mer}$ (AspH + hFX EGF1$_{39mer}$ + NC$_{26mer}$, 1:25:25 ratio) in Standard Buffer conditions. Spectra obtained after 12 min showed that the non-canonical share of the hFX EGF1$_{39mer}$ peptide mixture seemed to be near to the maximal reached hydroxylation level und standard conditions (~40 %, see **Chapter 2**) after this time (~39 % hydroxylation). A great proportion of the NC$_{26mer}$ was hydroxylated at this time point but the reaction is incomplete (~52 % hydroxylation), indicating that the hFX EGF1$_{39mer}$ peptide became hydroxylated first.

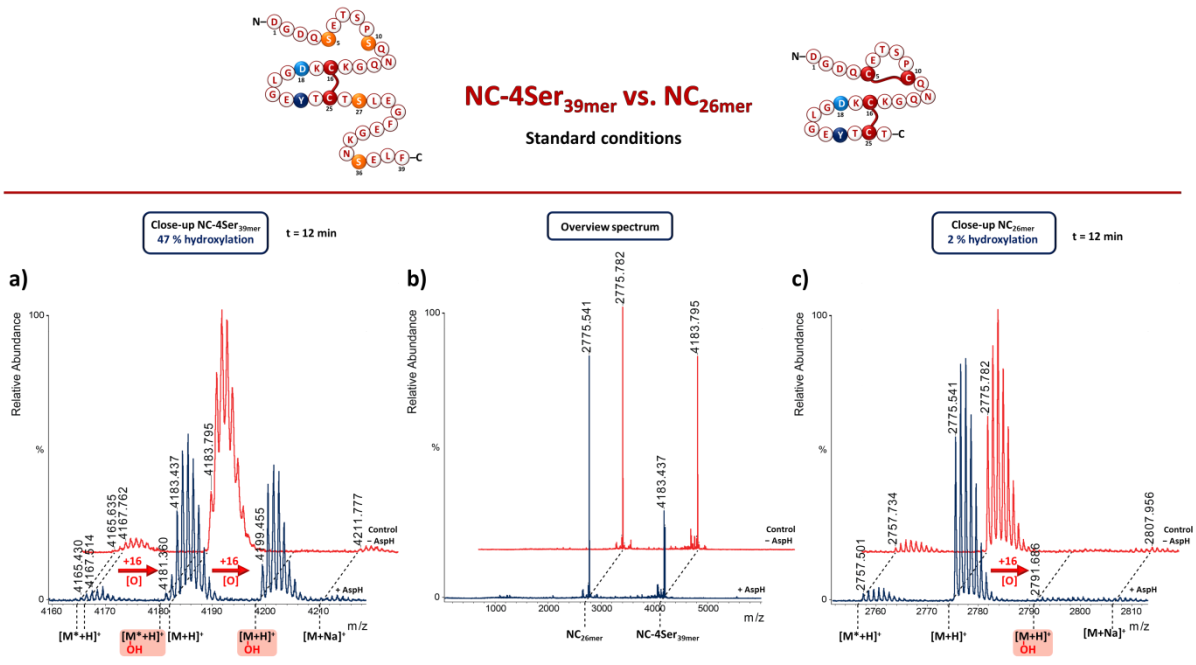


Fig. 4-27 | MALDI-ToF-MS results from competition assays in presence of both NC-4Ser$_{39mer}$ and NC$_{26mer}$ (AspH + NC-4Ser$_{39mer}$ + NC$_{26mer}$, 1:25:25 ratio) in Standard Buffer conditions. Spectra obtained after 12 min showed that the NC-4Ser$_{39mer}$ (~47 % hydroxylation) is clearly the preferred substrate over the 26mer that was hydroxylated at a very low level, if at all, at this time point (~2 % hydroxylation).

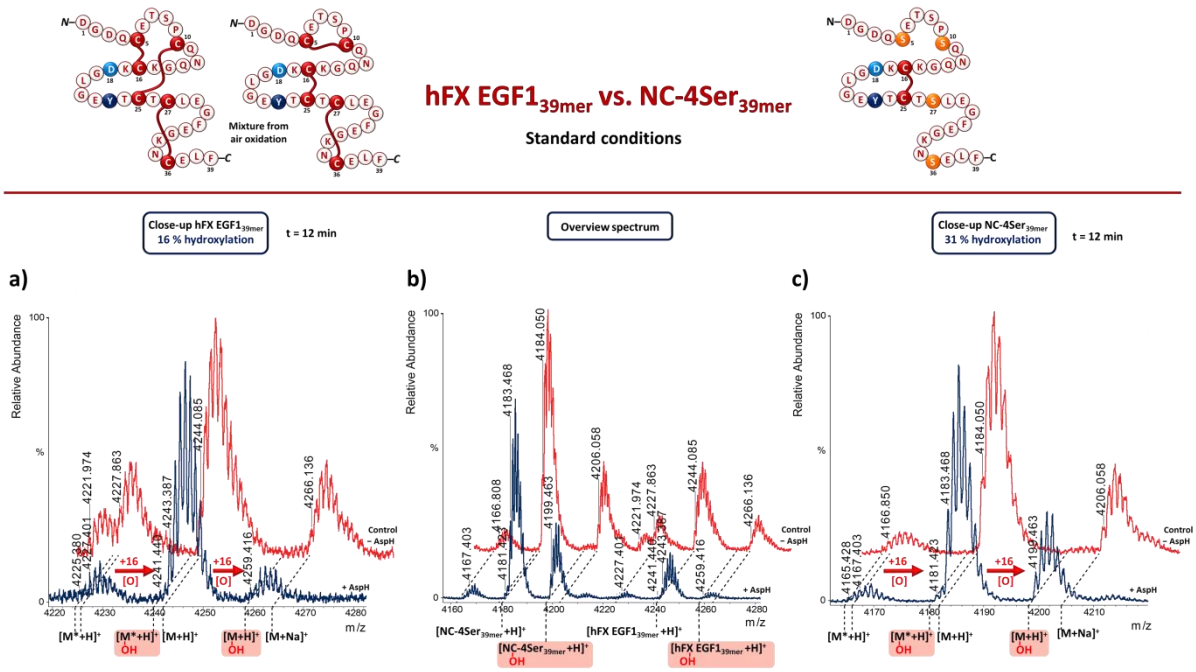


Fig. 4-28 | MALDI-ToF-MS results from competition assays in presence of both hFX EGF1$_{39mer}$ and NC-4Ser$_{39mer}$ (AspH + hFX EGF1$_{39mer}$ + NC-4Ser$_{39mer}$, 1:25:25 ratio) in Standard Buffer conditions. Spectra obtained after 12 min show that the NC-4Ser$_{39mer}$ with the single Cys3-4 disulfide (~31 % hydroxylation) appears to be preferred over the hFX EGF1$_{39mer}$ with the three disulfides (~16 % hydroxylation). However, since the hFX EGF1$_{39mer}$ was a mixture of both the non-canonical and canonical 39mer peptide forms, the quantitation may not be comparable.

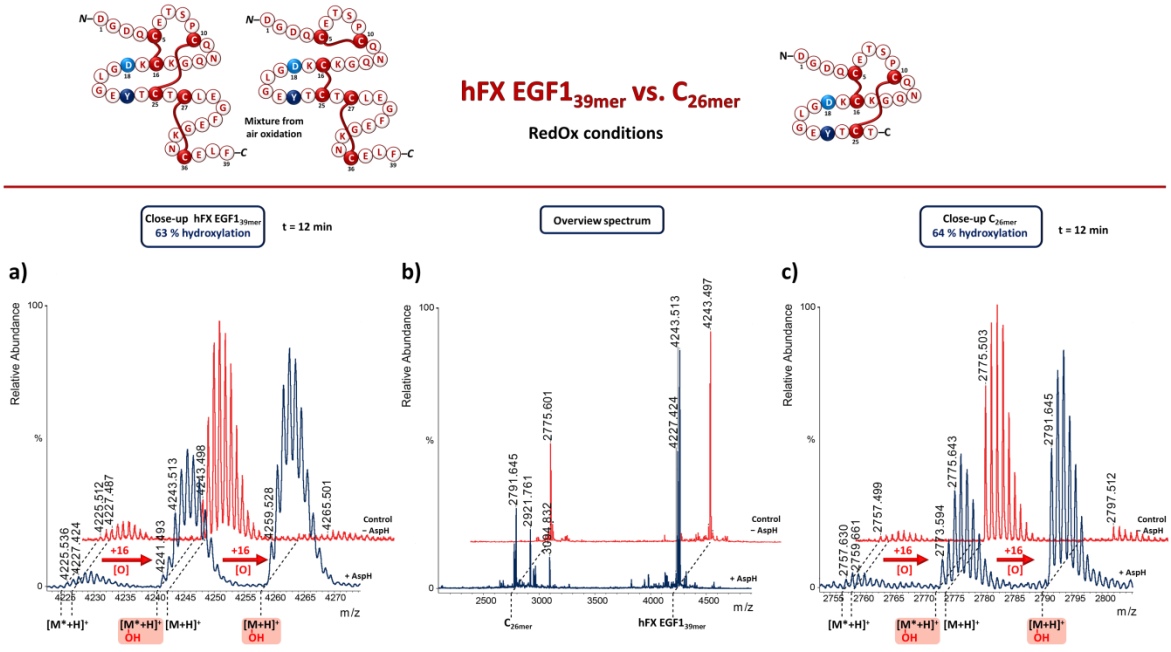


Fig. 4-29 | MALDI-ToF-MS results from competition assays performed in RedOx buffer conditions in presence of hFX EGF1$_{39mer}$ and C$_{26mer}$ (AspH + hFX EGF1$_{39mer}$ + C$_{26mer}$, 1:25:25 ratio). Spectra obtained after 12 min showed that both peptides were hydroxylated to comparable levels in RedOx buffer conditions within the incubation time.

The results of the competition experiments showed that AspH prefers the longer peptides over the shorter ones (**Fig. 4-30**). Thereby, the NC-4Ser$_{39mer}$ peptide with the single Cys3-4 disulfide was the preferred substrate over the hFX EGF1$_{39mer}$ with its three disulfides. However, since the hFX EGF1$_{39mer}$ was a mixture of both the canonical and the non-canonical disulfide forms (hFX EGF1$_{39mer}$), the quantitation may not be comparable. The NC$_{26mer}$ was the preferred substrate over the CP1$_{10mer}$. Note that over a longer incubation time also the cyclic peptide was fully hydroxylated.

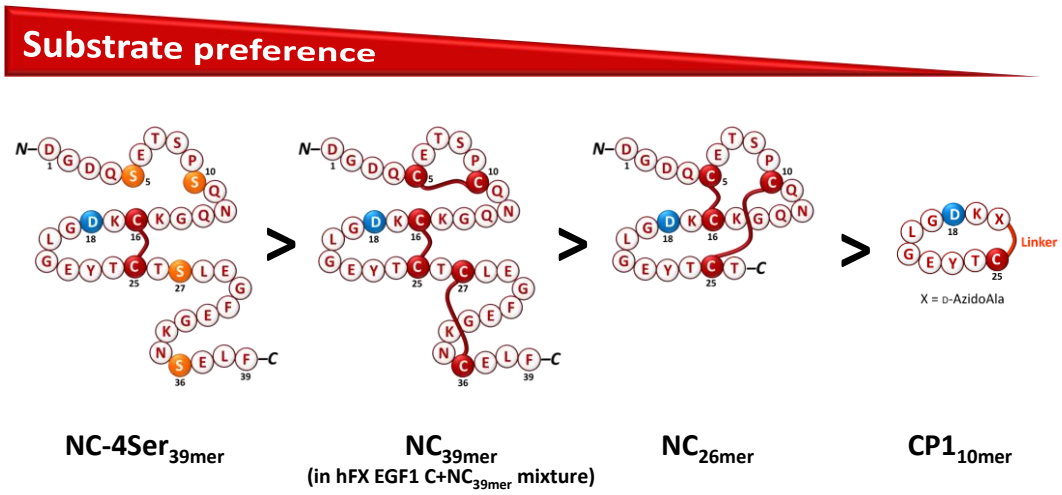


Fig. 4-30 | AspH substrate preference in the conducted time course-monitored competition experiments. AspH had a preference for the longer 39mer peptides over the 26mer, which was preferred over the CP1$_{10mer}$. The NC-4Ser$_{39mer}$ peptide was the most preferred substrate of all in these experiments.

4.4 Discussion

4.4.1 Substrate Requirements and Disulfide Plasticity

The results from the studies presented in this **Chapter** validate the consensus sequence for AspH-catalysed β-hydroxylation proposed by Stenflo and coworkers ($C_3X(D/N)XXXX(Y/F)XC_4XC_5$).[338] All cyclic peptides containing Asp/Asn in the hydroxylation site position as well as Tyr/Phe in the penultimate position preceding Cys4 were AspH substrates under the tested conditions. As shown here and consistent with the studies reported by Stenflo and coworkers,[338] linear peptides were not accepted as substrates. The cyclic structure and hence predetermined conformation of the peptide was a requirement for AspH substrate recognition (**Fig. 4-31**). The extent of EGFD hydroxylation could be influenced by the accessibility or steric hindrance of the hydroxylation site owing to the remaining residues in the peptide sequence. Efficient cyclic peptide model substrates are easily accessible through the established methods.

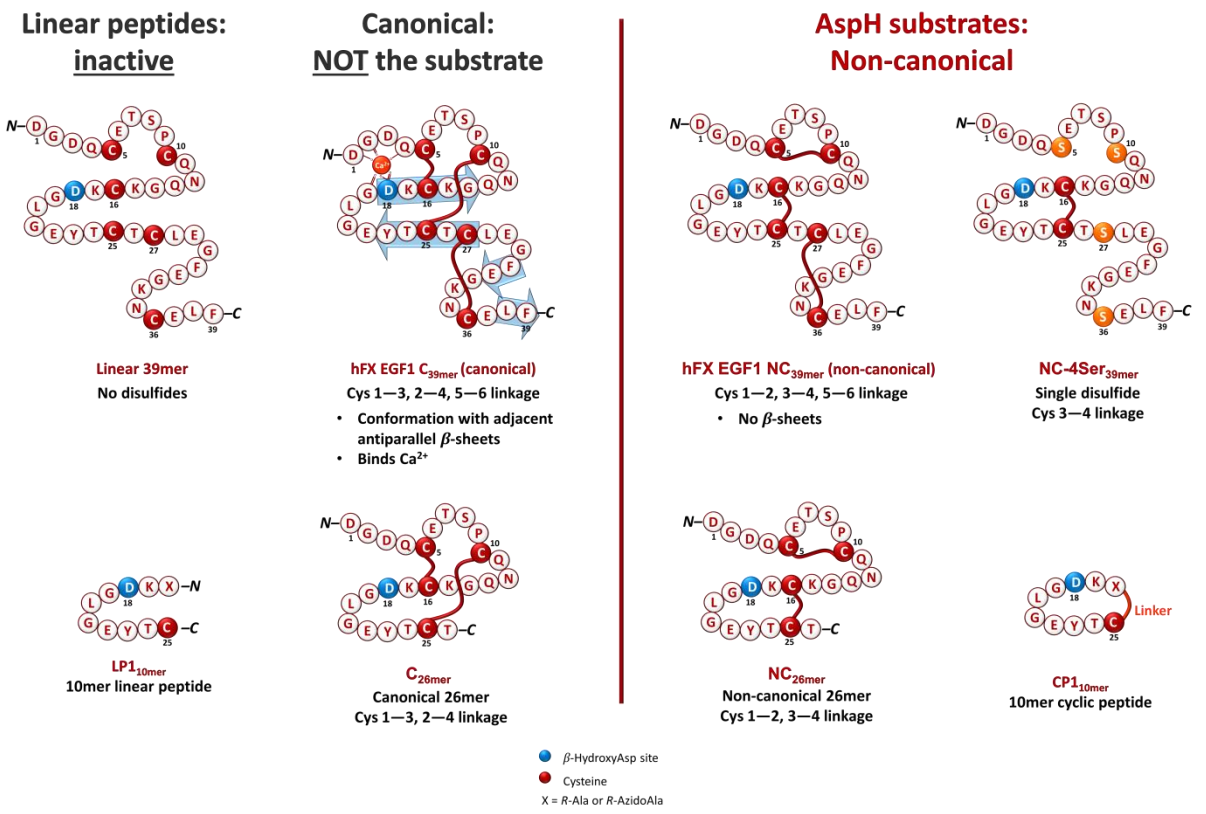


Fig. 4-31 | AspH requires cyclic substrates with a non-canonical Cys3-4 disulfide linkage or the corresponding thioether analogue in the synthesised 10mer cyclic peptides for activity.

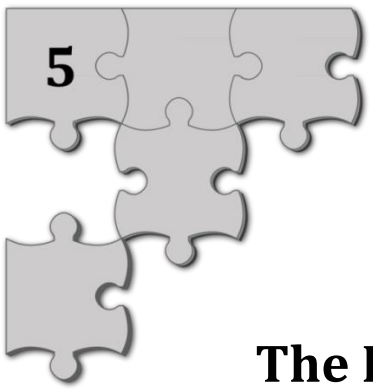
Any change of the hydroxylation site Asp/Asn residue to other amino acids resulted in complete loss of activity. Modelling of residues on the AspH-substrate crystal structure indicated that the side chain of Glu is sterically too demanding to fit in the AspH active site pocket. The cyclic peptides with either *R*-Asp, Ser, Abu, Hse, Pro replacing the Asp/Asn residue, however, were expected to fit in the active site pocket but they were not hydroxylated and were no inhibitors in the preliminary work. In summary, AspH appeared to be very selective with respect to the hydroxylation site in the synthesised cyclic peptide models. However, the enzyme can tolerate sequence differences in the residues between the conserved Asp/Asn and the Tyr/Phe. In contrast, the other known mammalian asparaginyl hydroxylase FIH has been reported to be promiscuous with respect to the residues it can oxidise, since the enzyme also accepts various other residues such as Ser, His, Leu as hydroxylation substrates.[500]

The poor performance of the 10mer cyclic peptide in the competition experiments could result from the decreased number of protein-protein interaction points. The results suggested that the 10mer cyclic peptide was the weakest binder of the tested *in vitro* substrates. The reaction of AspH with the 10mer cyclic peptides in absence of other substrates proceeded faster than the reaction with the 39mer peptides, but in the competition experiments the longer substrates were favoured. However, the studies presented in this **Chapter** show that the cyclic peptide is a good alternative substrate, and thus a model of a non-canonical EGFD. The preference for the 39mer

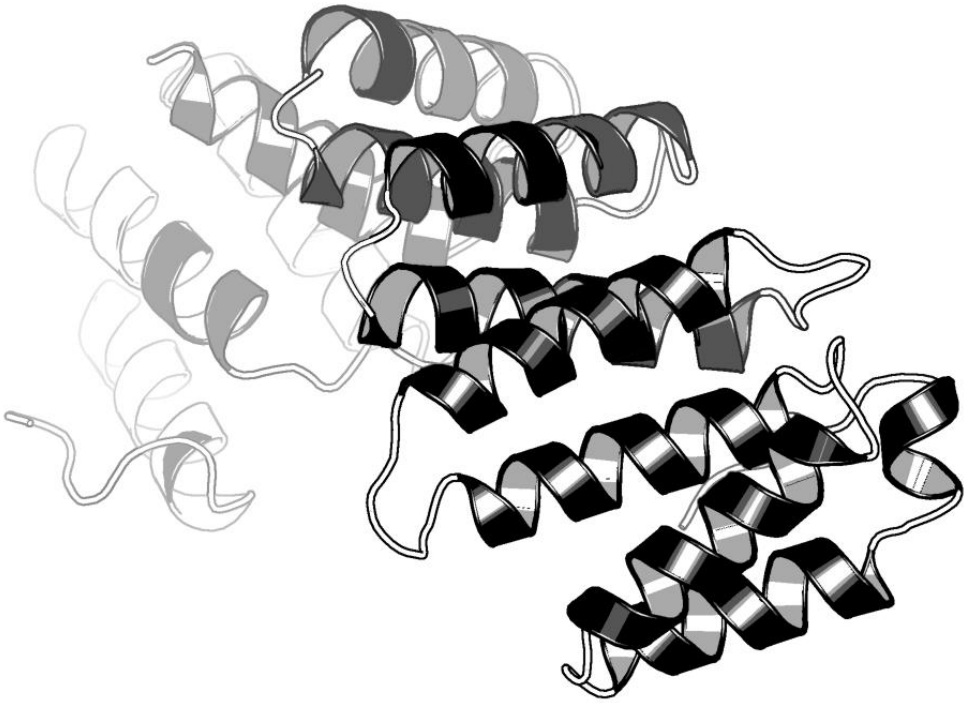
peptides over the 26mer peptides suggested further important interaction points of the *C*-terminal EGFD region with the AspH TPR domain. This region of the hFX $EGF1_{39mer}$ peptide was unfortunately not visible in the crystal structure. Therefore, further crystallographic studies on AspH-substrate complexes have been performed, which are presented in **Chapter 5**.

4.5 ACKNOWLEDGEMENTS FOR CHAPTER 4

Dr Martin Münzel and Dr Michael A. McDonough are kindly acknowledged for their contribution to the cyclic peptide development. Most of the MALDI-ToF-MS measurements were performed on the *Bruker Daltonics* Ultraflex I spectrometer based at the Oxion Proteomics Facility, Department of Physiology, Anatomy and Genetics, University of Oxford, and Dr Holger Kramer is kindly acknowledged for granting access to the instrument and fruitful discussions at all times. Dr Richard J. Hopkinson is kindly acknowledged for the NMR studies on the cyclic peptide hydroxylation site and for providing the NMR figures. Dr Sander van Berkel is kindly acknowledged for the fruitful discussions on strain-promoted cycloadditions and BCN-conjugates.



The Role of the TPR Domain in Substrate Recognition



Abstract

In this Chapter, further crystallographic, binding and activity studies on AspH-substrate complexes and AspH variants are presented; the results highlight the crucial role of the TPR domain in AspH substrate recognition.

5.1 Crystallographic Studies on AspH-Substrate Complexes

The $AspH_{315-758}$-substrate complex structures presented here were solved and refined by Dr Michael A. McDonough. The studies are still in progress with the final refinement data to be completed.

The results from the competition experiments presented in **Chapter 4** demonstrated that AspH has a preference for the NC-4Ser$_{39mer}$ and the hFX EGF1$_{39mer}$ over the NC$_{26mer}$ and the CP2$_{10mer}$ peptides, suggesting that there may be further important interaction points of the *C*-terminal EGFD region with the AspH TPR domain (**Fig. 5-1**). This region of the hFX EGF1$_{39mer}$ peptide was not visible in the $AspH_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ crystal structure. Further crystallographic studies on AspH-substrate complexes were therefore conducted.

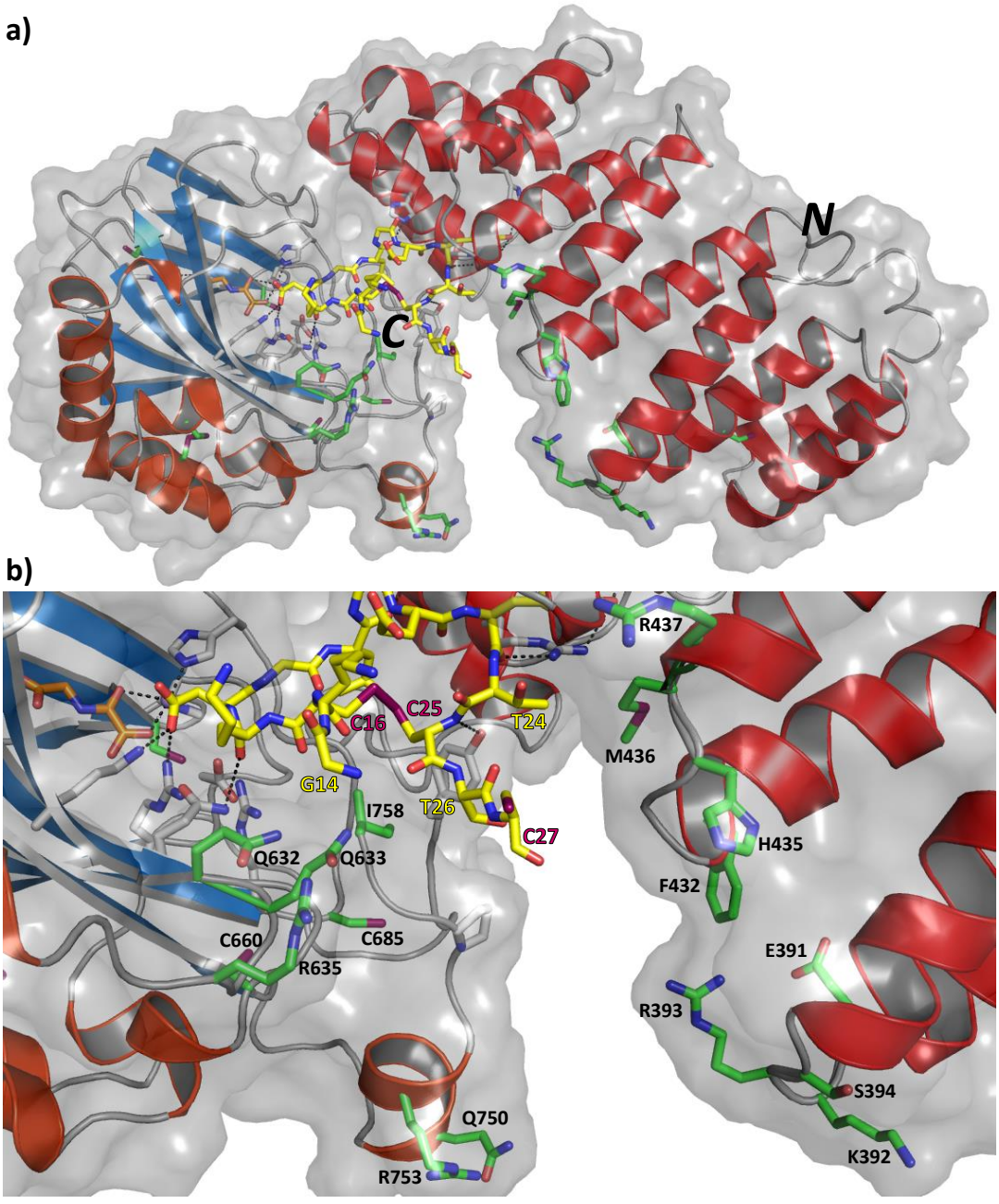


Fig. 5-1 | AspH residues along the surface (green) that are potentially involved in substrate recognition. **a)** Overview of the $AspH_{315-758}$·Mn^{II}·NOG·hFX EGF1$_{39mer}$ crystal structure; **b)** close-up view.

5.1.1 Crystal Structures of $AspH_{315-758}$ in Complex with NC-$4Ser_{39mer}$ and NC_{26mer}

Crystallisation trials of His_6-$AspH_{315-758}$ (18 mg/ml) with the NC-$4Ser_{39mer}$, NC_{26mer} and $CP2_{10mer}$ peptides were set up with 1 mM $MnCl_2{\cdot}4H_2O$ and 2 mM NOG (all at 4 °C). The trials were prepared with different enzyme-to-substrate ratios, since a higher substrate concentration may be required for the NC_{26mer} and $CP2_{10mer}$ peptides to achieve formation of a stable complex. An AspH-to-substrate ratio of 1:5 was used for the trials with the NC_{26mer} peptide and ratios of 1:5 and 1:10 with the $CP2_{10mer}$ peptide. Since the competition experiments had indicated that the NC-$4Ser_{39mer}$ peptide was the most preferred substrate, the co-crystallisation trials were set up with the same enzyme-to-substrate ratio (1:2.2) as for the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·hFX $EGF1_{39mer}$ trials. The plates were incubated at 4 °C for several weeks until crystals of orthorhombic morphology were obtained. The data were collected at Diamond Light Source Ltd, Didcot; crystals of the NC-$4Ser_{39mer}$ at 1.99 Å and 2.50 Å resolution for the NC_{26mer}, see **Table 5-1** in **Section 5.1.2.4**). The electron density maps revealed that the substrates were bound to AspH.

5.1.1.1 Crystallography Reveals Further Key Interactions for Binding to AspH

The substrate binding mode in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·NC-$4Ser_{39mer}$ and the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·$NC_{26mer}$ complex structures was very similar to the one observed in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·hFX $EGF1_{39mer}$ complex structure. The *N*-termini of the peptide substrates were also disordered in the new crystal structures, indicating a higher degree of flexibility in this region. However, 22 substrate residues were apparent in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·NC-$4Ser_{39mer}$ electron density map (**Fig. 5-2**; previously 14 substrate residues apparent in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·hFX $EGF1_{39mer}$ electron density map, see **Fig. 2-41**). Further important AspH-substrate interactions were revealed. The $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·NC-$4Ser_{39mer}$ crystal structure clearly showed that the *C*-terminal tail in the NC-$4Ser_{39mer}$ peptide interacts with TPR domain residues. In particular, $Phe31_{NC\text{-}4Ser39mer}$ appeared to be a key residue for interactions with AspH residues in a hydrophobic pocket surrounding Phe432 within the TPR domain.

Superimposition of the different peptides substrates in the crystal structures shows a very similar overall conformation in the AspH complexes and only slight variations in the orientation of the residues (**Fig. 5-3**). Interestingly, superimposition of the AspH-substrate complex structures revealed a minor but significant substrate-dependent change in the TPR domain conformation in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}$NOG·$NC_{26mer}$ complex structure to a more "open" position compared to the structures with the hFX $EGF1_{39mer}$ and the NC-$4Ser_{39mer}$ (**Fig. 5-3a**). This corroborated the importance of the *C*-terminal region of the hFX $EGF1_{39mer}$ and the NC-$4Ser_{39mer}$ peptides (missing in the NC_{26mer}) for tight binding during protein-protein interactions, even though this interaction did not seem to be

a requirement for hydroxylation. The results also supported the proposal of an important role of the TPR domain for substrate recognition and delivered a plausible explanation for the observed AspH substrate preferences. This substrate-dependent conformational change was subtle compared to the ~15 Å movement of the TPR domain upon substrate binding. The overall conformation of the oxygenase domain was retained in all AspH-substrate complex structures.

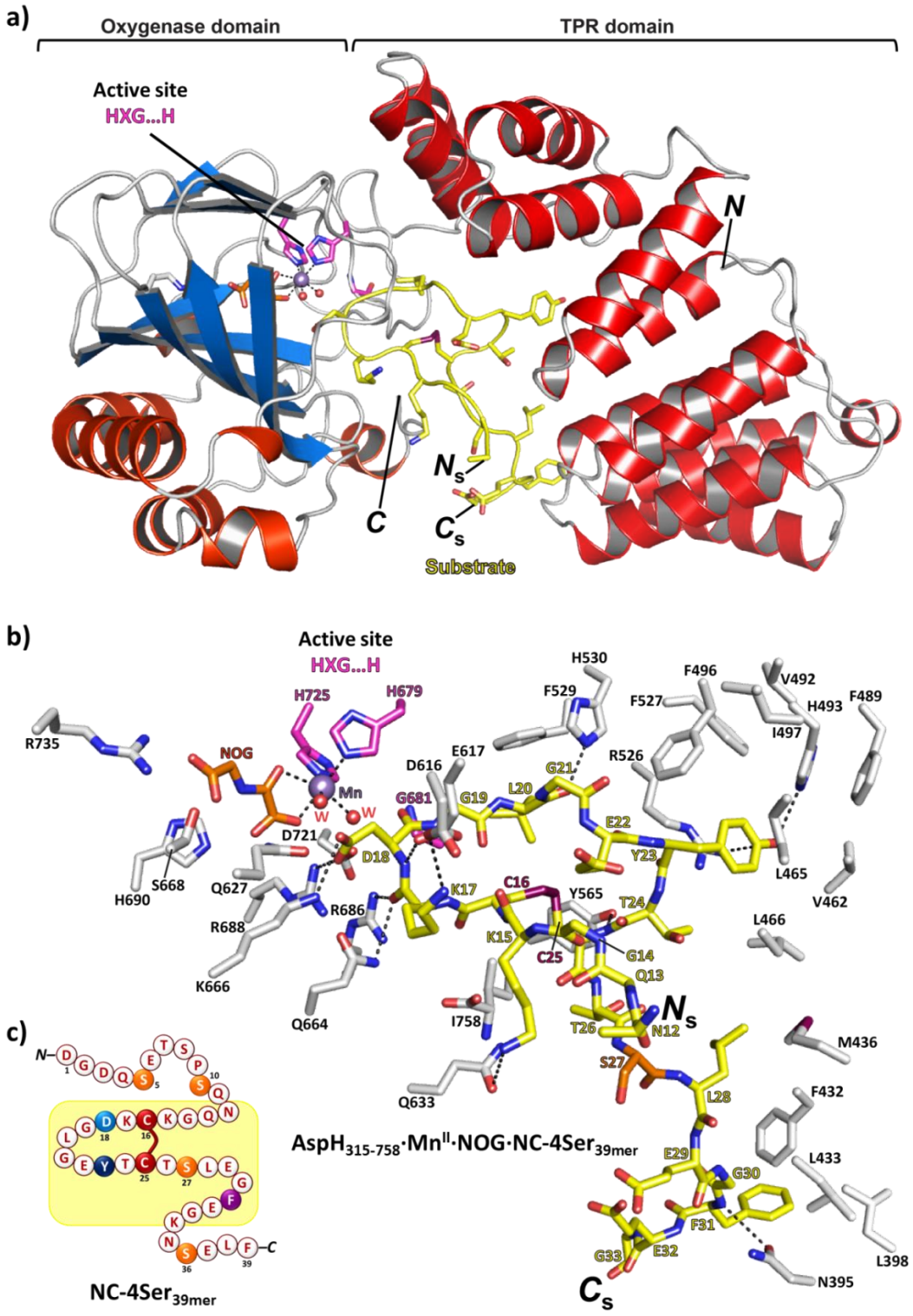


Fig. 5-2 | View from the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}NC\text{-}4Ser_{39mer}$ crystal structure (resolution: 1.99 Å). **a)** Overview of the structure. A number of 22 $NC\text{-}4Ser_{39mer}$ substrate residues were apparent in the electron density. **b)** Sticks representation of some residues involved in important interactions. **c)** Schematic of the $NC\text{-}4Ser_{39mer}$ substrate (visible part in the crystal structure shaded in yellow).

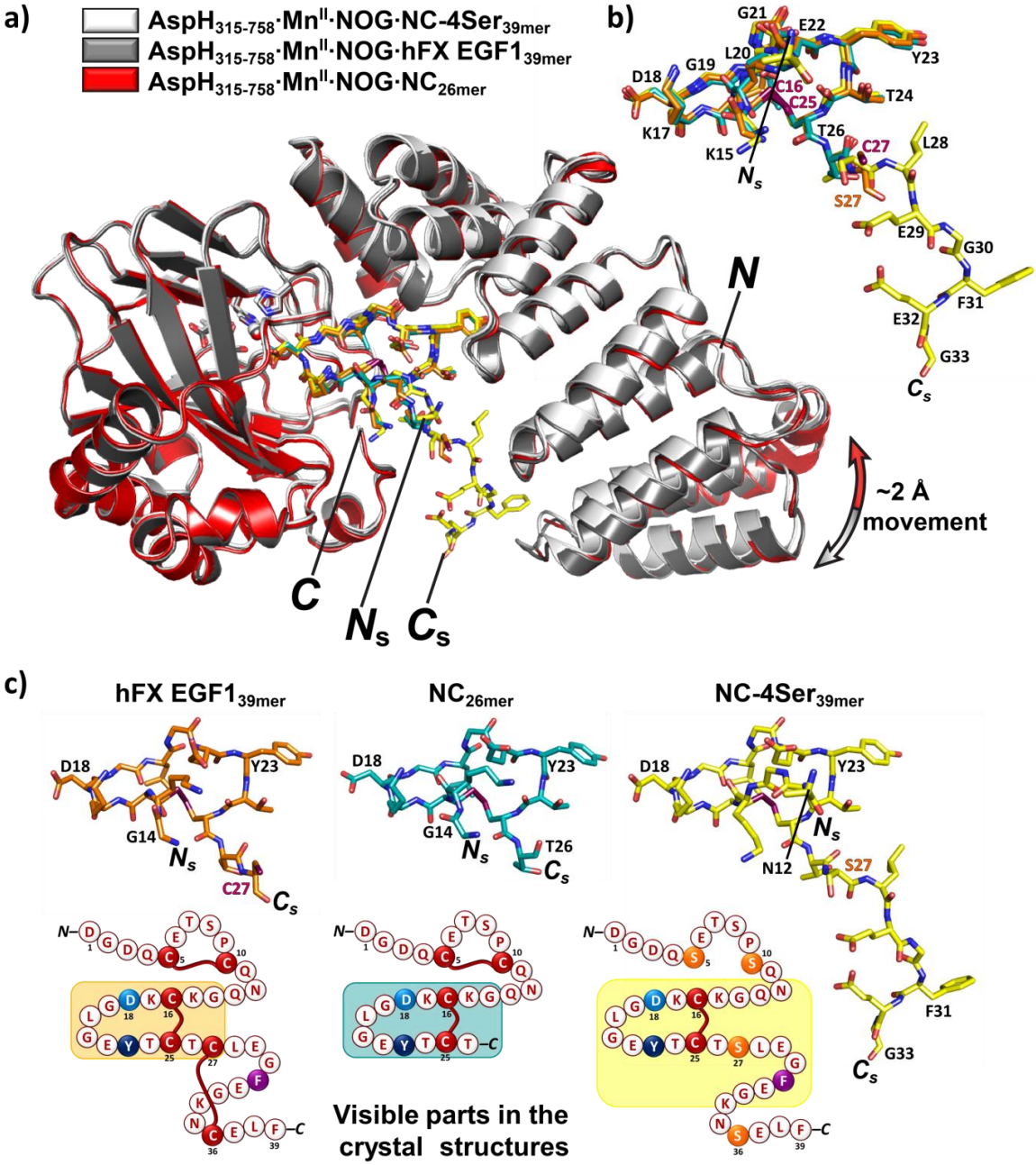
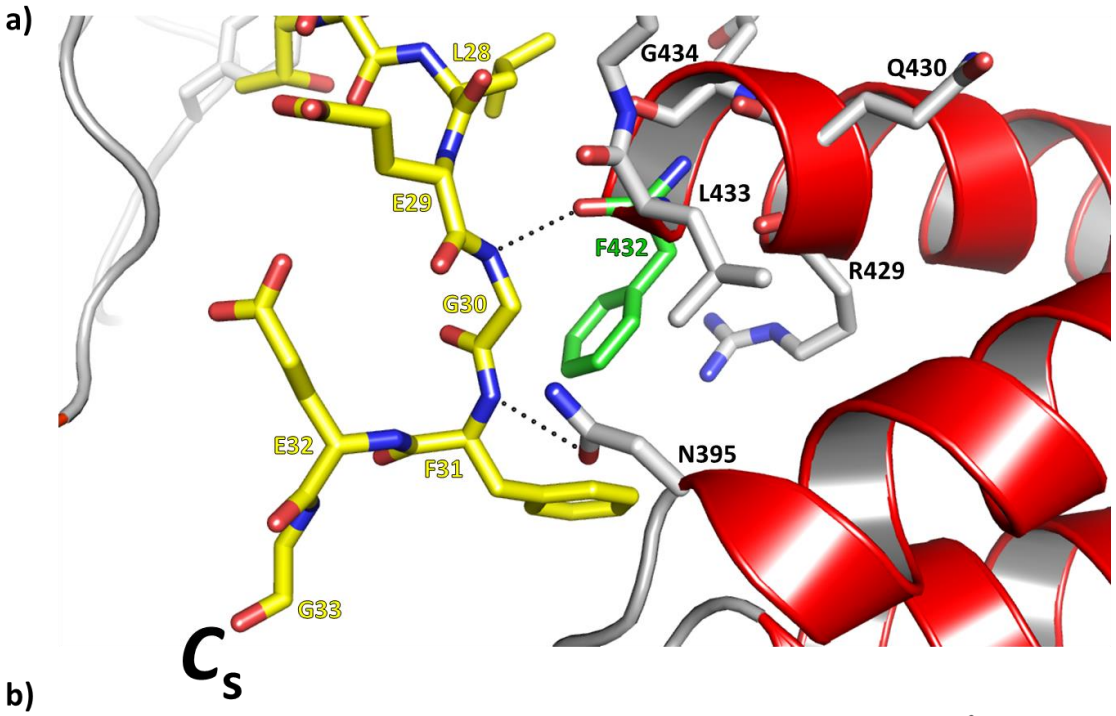


Fig. 5-3 | Crystallography reveals substrate-dependent conformational changes of the TPR domain. **a)** Superimposition of the obtained crystal structures of $AspH_{315\text{-}758}$ in complex with the different peptide substrates: NC-4Ser$_{39mer}$ (yellow), hFX EGF1$_{39mer}$ (orange), and NC$_{26mer}$ (teal). The overall conformation of the oxygenase domain was retained in all obtained structures. The TPR domain, however, underwent a minor conformational change to a more "open" position in the structure with the NC$_{26mer}$ peptide due to the lack of important interactions of the *C*-terminal region of the 39mer peptides with the TPR domain. **b)** Close-up of the superimposed peptides showing a very similar conformation of the different peptides in the AspH complexes. **c)** Visible parts of the peptide substrates in the crystal structures.

The side chain of Phe31$_{NC\text{-}4Ser39mer}$ can π-π-stack with the side chain of Phe432 from the TPR domain (**Fig. 5-4a**). In addition to the hydrophobic interactions within the binding pocket, the side chain of Asn395 forms a hydrogen bond with Phe31$_{NC\text{-}4Ser39mer}$ peptide backbone, and the AspH Phe432 backbone carbonyl interacts with the Gly30$_{NC\text{-}4Ser39mer}$ backbone. The sequence alignment of known *β*Hya/*β*Hyn-containing EGFDs shows, the aromatic residue in the position of Phe31 in the hFX EGF1-based NC-4Ser$_{39mer}$ is highly conserved (**Fig. 5-4b**).



b)

EGFD Protein			Amino Acid Sequence (with Residue Numbers)		Ca^{2+}-binding consensus	Ref
hFIX	EGF1	93:	DGDQCESNP----------CLNGG--SCKDDINSYECWCPFGFEG------KNCELD	:129	✓	337
bFIX	EGF1	47:	DGDQCESNP----------CLNGG--MCKDDINSYECWCQAGFEG------TNCELD	:85	✓	501
hFX	EGF1	86:	DGDQCETSP----------CQNQG--KCKDGLGEYTCTCLEGFEG------KNCELF	:124	✓	336
bFX	EGF1	86:	DGDQCEGHP----------CLNQG--HCKDGIGDYTCTCAEGFEG------KNCEFS	:124	✓	501
bProtein C	EGF1	94:	SGSPCDLP----------CCGRG--KCIDGLGGFRCDCAEGWEG------RFCLHE	:131	-	502,343
bProtein S	EGF1	117:	ISDQCNPLP----------CNEDGFMTCKDGQATFTCICKSGWQG------EKCESD	:157	-	503
bProtein S	EGF2	157	DINECKDPVNINGG-----CSQ----ICENTPGSYHCSCKNGFVM--LSNKKDCKDV	:202	✓	503
bProtein S	EGF3	201:	DVDECVLKPSI--------CGTA---VCKNIPGDFECECAEGYKY--NPVSKSCDDV	:244	✓	503
bProtein S	EGF4	243:	DVDECAENL----------CAQ----LCVNYPGGYSCYCDGKKGFKLAQDQKSCEAV	:285	✓	503
bProtein Z	EGF1	47:	GGSPCASQP----------CLNNG--SCQDSIRGYACTCAPGYEG------PNCAFA	:85	-	504
hC1r	EGF1	142:	DLDECASRSKSGEEDPQPQCQH----LCHNYVGGYFCSCRPGYEL--QEDTHSCQAE	:190	✓	505
hC1s	EGF1	131:	DINECTDFVDVP--------CSH----FCNNFIGGYFCSCPPEYFL--HDDMKNCGVN	:175	✓	505
bTRBM	EGF3	99:	DVDDCAQLPSP--------CPQ----RCVNTEGGFQCHCDTGYEL--VDGE--CVDP	:139	✓	506
bLDLR	EGF2	356:	DIDECQNPDT----------CSQ----LCVNLEGSYKCECEEGFRL--EPLTKACKAV	:397	✓	506
hEGF Prec.	EGF2	356:	DVNECAFWNHG--------CTL----GCKNTPGSYYCTCPVGFVL--LPDGKRCHQL	:398	✓	345
hEGF Prec.	EGF7	870:	DIDECEMGVPV--------CPPASS-KCINTEGGYVCRCSEGYQGD---GIH-CLDI	:913	✓	345
hEGF Prec.	EGF8	912:	DIDECQLGEHS--------CGENA--SCTNTEGGYTCMCAGRLSE---PGLI-CPDS	:954	✓	345
hUMOD	EGF1	65:	DLDECAIPGAHN-------CSANS--SCVNTPGSFSCVCPEGFRL--SPGLG-CDV-	:109	✓	505
hUMOD	EGF2	108:	DVDECAEPGLSH-------CHALA--TCVNVVGSYLCVCPAGYRGD---GWH-CECS	:158	✓	505

Fig. 5-4 | In addition to Tyr23 in hFX EGF1$_{39mer}$, another highly conserved residue in the *C*-terminal region of EGFDs plays a key role in binding to AspH. **a)** Close-up view of the AspH$_{315-758}$·Mn^{II}·NOG·NC-4Ser$_{39mer}$ crystal structure showing important interactions of the *C*-terminal region of the peptide within a hydrophobic pocket in the TPR domain. **b)** Sequence alignment (performed using Clustalω) of reported *β*Hya/*β*Hyn-containing EGFDs. Marked in red: hydroxylation site; light-blue: hydrophobic key residue for binding to the TPR domain corresponding to Tyr23 in hFX EGF1$_{39mer}$; dark-blue: highly conserved aromatic residue corresponding to Phe31 in hFX EGF1$_{39mer}$; purple: potential glycosylation sites. **Abbreviations:** h: human; b: bovine; FIX: coagulation factor IX; FX: coagulation factor X; C1r: complement component 1r; C1s: complement component 1s; TRBM: thrombomodulin; LDLR: low-density lipoprotein receptor; EGF Prec.: pro-epidermal growth factor (EGF precursor); UMOD: uromodulin.

5.1.2 Complex Formation and Crystallisation of AspH with a Cyclic Peptide

5.1.2.1 Optimisation of the Cyclic Peptide

A crystal structure of AspH in complex with a cyclic peptide could demonstrate that a peptide substrate adopts the binding mode predicted in the model and validate the results from the activity

assays. Co-crystallisation attempts of His_6-$AspH_{315-758}$ with the $CP1_{10mer}$ and the $CP2_{10mer}$ peptides (AspH-to-substrate ratios 1:5 and 1:10) were not successful, and only crystals of "apo"-AspH were obtained. As the results from the competition experiments (**Chapter 4**) had already indicated and the $AspH_{315-758}$·Mn^{II}·NOG·$NC\text{-}4Ser_{39mer}$ structure then revealed, the *C*-terminal region of the EGFD substrates was important for binding to AspH. This knowledge was then used for optimisation of the cyclic peptide scaffold.

Branched cyclic peptides with a tail covering the important *C*-terminal region of the $NC\text{-}4Ser_{39mer}$ substrate were produced in order to improve binding of the cyclic peptide to AspH. Based on the $NC\text{-}4Ser_{39mer}$ sequence, a 15mer ($CP1_{15mer}$ with *R*-AzidoAla linker and $CP2_{15mer}$ with *R*-Ala linker) and a 19mer peptide ($CP1_{19mer}$ with *R*-AzidoAla linker and $CP2_{19mer}$ with *R*-Ala linker) were designed and synthesised (**Fig. 5-5**). The same cyclisation method used for the 10mer peptides (see **Chapters 4** and **10**) was successfully employed for the synthesis of these longer cyclic peptides. The CP_{19mer} peptides were designed to contain the highly conserved aromatic residue corresponding to Phe31 in hFX $EGF1_{39mer}$, whereas the CP_{15mer} peptides lacked the residue.

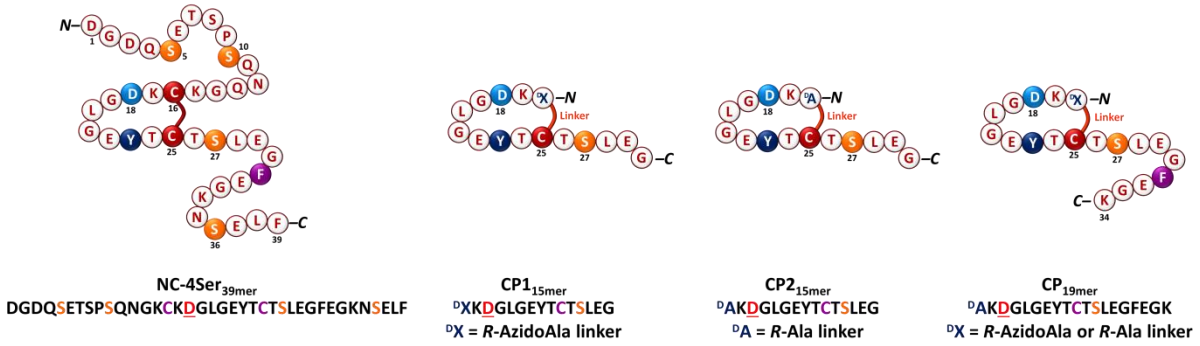


Fig. 5-5 | Branched cyclic peptides (CP_{15mer} and CP_{19mer} peptides) based on the $NC\text{-}4Ser_{39mer}$ sequence. The *N*-termini of the cyclic peptides are formally shown for clarity with respect to the sequences for the peptide synthesis.

The $CP1_{15mer}$, $CP2_{15mer}$ and $CP2_{19mer}$ peptides were tested in hydroxylation activity assays with $AspH_{315-758}$. All peptides were efficient $AspH_{315-758}$ substrates (**Fig. 5-6**), however, only the longer $CP2_{19mer}$ was fully hydroxylated within the assay time (60 min at 37 °C). The $CP1_{19mer}$ was also attached to a biotin-BCN conjugate ($BP1_{19mer}$) using the strain-promoted azide-alkyne cycloaddition technology described in **Chapter 4**. Although only low hydroxylation levels of the shorter $BP1_{10mer}$ were observed within 60 min incubation with AspH, almost full hydroxylation of the longer $BP1_{19mer}$ was achieved using the same conditions (**Fig. 5-7**). The results indicated that the *C*-terminal tail substantially improved binding of the cyclic peptide to AspH.

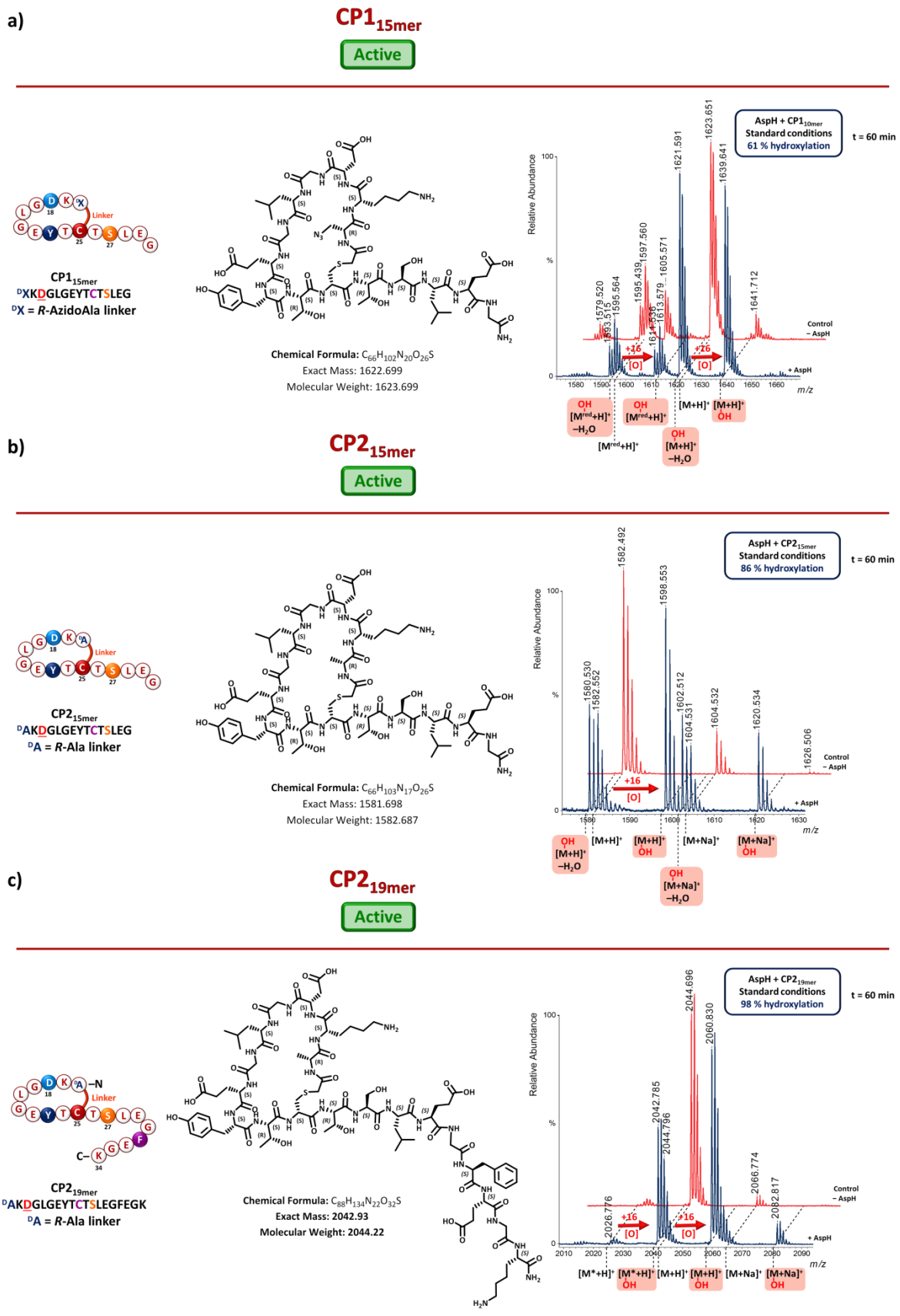


Fig. 5-6 | MALDI-ToF-MS results from activity assays of AspH with the branched cyclic peptides. **a)** AspH + $CP1_{15mer}$ with incorporated azide moiety (*R*-AzidoAla linker); **b)** AspH + $CP2_{15mer}$ (*R*-Ala linker); **c)** AspH + $CP2_{19mer}$ (*R*-Ala linker).

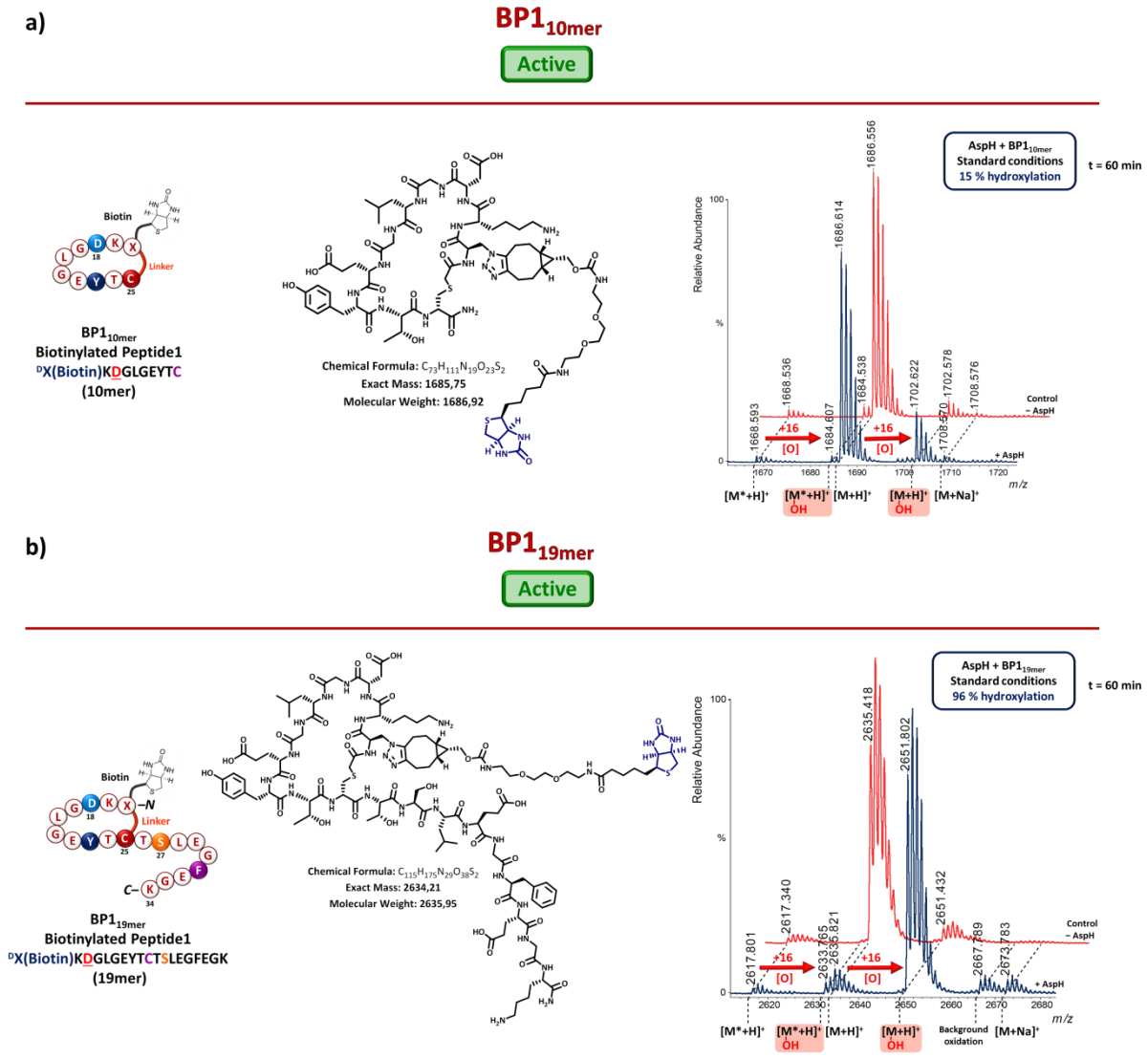


Fig. 5-7 | MALDI-ToF-MS results from activity assays of AspH with the biotinylated cyclic peptides. **a)** Low hydroxylation levels of the $BP1_{10mer}$ were observed within 60 min incubation, whereas **b)** almost full hydroxylation of the $BP1_{19mer}$ was achieved using the same conditions.

5.1.2.2 Binding Studies Using Differential Scanning Fluorimetry

Differential scanning fluorimetry assays (DSF) of $AspH_{315-758}$ with the different peptides at a 1:10 ratio (2 µM enzyme, 20 µM peptide in presence of 20 µM NOG and 50 µM $MnCl_2$) showed significant T_m shifts indicating binding in solution only for the NC-4Ser_{39mer} and the hFX $EGF1_{39mer}$ peptides but neither for the NC_{26mer}, the C_{26mer} nor for any of the cyclic peptides (**Fig. 5-8**). In assays with an AspH-to-peptide ratio of 1:100 (2 µM enzyme, 200 µM peptide in presence of 20 µM NOG and 50 µM $MnCl_2$) significant T_m shifts were also observed for the NC_{26mer} and the $CP2_{19mer}$ but not for the C_{26mer} and the shorter cyclic peptides. The higher peptide concentration required to produce significant shifts in the assay indicated that a higher substrate concentration may also be required for crystallisation to achieve formation of a sufficiently stable complex.

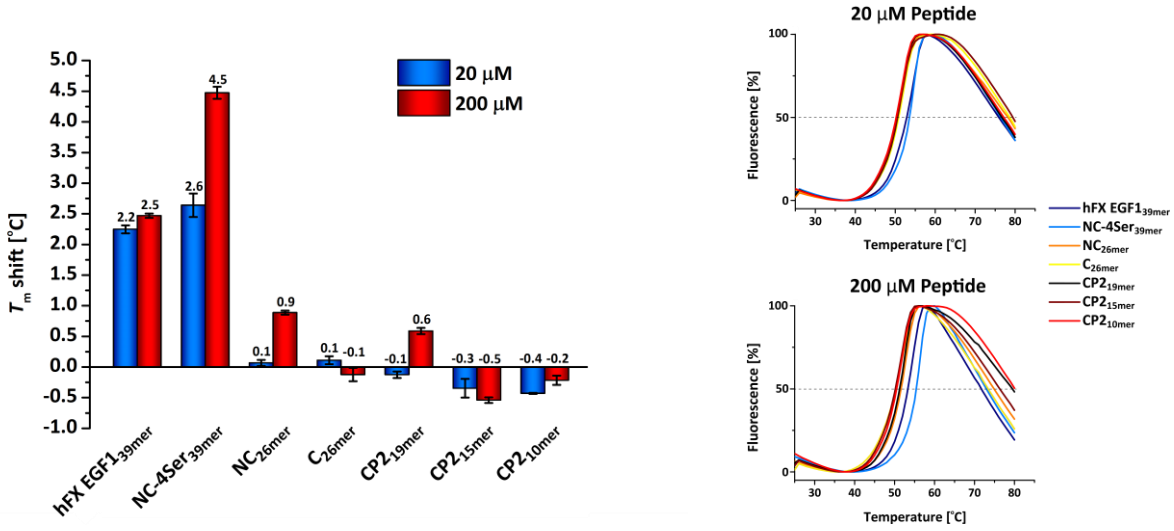


Fig. 5-8 | Results from differential scanning fluorimetry assays of AspH with different peptides at 1:10 or 1:100 ratio respectively (2 µM AspH, 20 µM or 200 µM peptide). Representation as T_m shifts and melting curves (one curve of triplicates shown for each experiment). Error bars represent SD for triplicate assays.

5.1.2.3 Binding Studies Using Non-Denaturing Mass Spectrometry

Non-denaturing electrospray ionisation mass spectrometry (native ESI-MS)[507,508] was used to study AspH complex formation in the gas phase. The experiments were performed in the positive ion mode using a *Waters* Q-ToF Micro mass spectrometer equipped with *Advion* NanoMate®. Initial experiments were conducted together with Hanna Tarhonskaya, who provided training on the instrument. The analysis conditions were adapted from those successfully used for other 2OG oxygenases such as PHD2,[509,510] and were later further optimised as described below.

The volatile ammonium acetate buffer (NH_4OAc) was used for the native ESI-MS studies to avoid non-specific binding of Na^+ or K^+ ions to the protein. His$_6$-AspH$_{315-758}$ was buffer exchanged into 15 mM NH_4OAc, pH 7.5, prior to analysis (double buffer exchange to minimise background noise, as described in **Chapter 10**). With its mass of 54.5 kDa His$_6$-AspH$_{315-758}$ is a relatively big protein compared to PHD2 (27.6 kDa). However, His$_6$-AspH$_{315-758}$ was ionisable under the applied conditions, and the apo-protein gave strong and sharp signals (**Fig. 5-9**). An initial non-deconvoluted ESI-MS spectrum showed the molecular peaks of the different charge states of the native protein (mass-to-charge ratios, *m/z*, **Fig. 5-9a**). The most abundant peaks were observed for the $[M + 16\ H]^{16+}$ and $[M + 15\ H]^{15+}$ charge states of His$_6$-AspH$_{315-758}$. The deconvolution (performed using MassLynx™ version 4.1) combines all charge states, giving the molecular mass of the protein (**Fig. 5-9b**).

The spectra from protein concentrations of 15 µM, 20 µM, 25 µM His$_6$-AspH$_{315-758}$ differed only slightly. The lowest background signal was obtained with 15 µM His$_6$-AspH$_{315-758}$, and this concentration was therefore used for further studies.

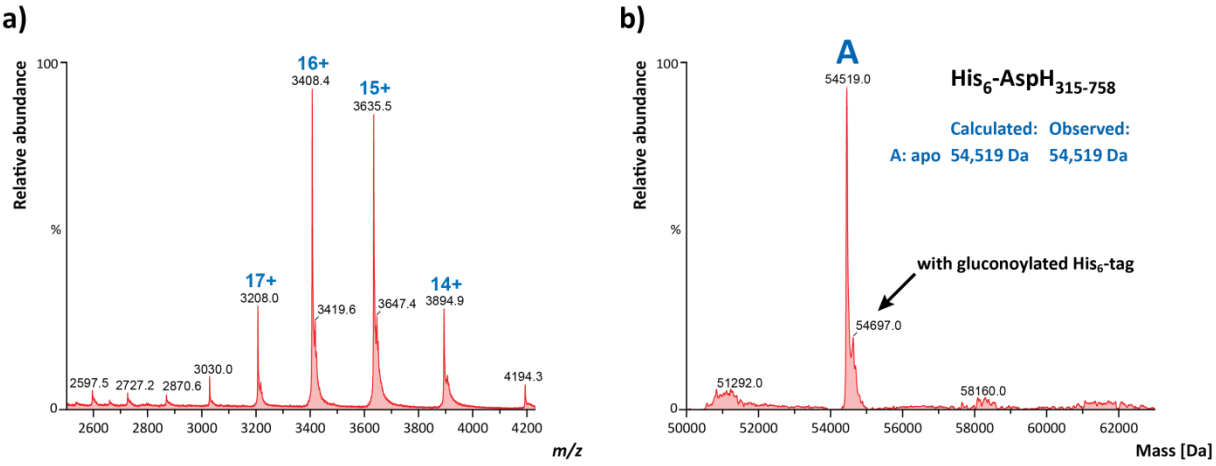


Fig. 5-9 | Non-denaturing ESI-MS of purified His_6-$AspH_{315-758}$ (15 µM) in NH_4OAc (15 mM, pH 7.5) at 80 V sample cone voltage. Data represented as **a)** non-deconvoluted mass spectrum (charge states (blue) and molecular peaks), and **b)** deconvoluted mass spectrum. The molecular mass of 54,519 Da observed for His_6-$AspH_{315-758}$ corresponds to the calculated mass. The peak at 54,697 Da likely corresponds to the protein with a gluconoylated His_6-tag[390,391] (a mass addition of +178 Da). All samples and measurements were repeated at least three times, and representative spectra are shown throughout.

Non-denaturing ESI-MS was then employed to study binding of metals such as Fe(II) and Mn(II), the cosubstrate 2OG (in form of the free acid) or its analogue NOG, and the different peptide substrates to His_6-$AspH_{315-758}$. The background noise in the spectra obtained for the non-covalent complex samples was generally higher than for the apo-protein sample. However, the data showed that His_6-$AspH_{315-758}$ was suitable for native ESI-MS analysis. Tested different buffer concentrations of 15 mM, 50 mM, 200 mM NH_4OAc, pH 7.5, showed no significant differences but the samples in 15 mM NH_4OAc, pH 7.5, appeared to give the most stable signals with the lowest background. Variation of the source temperature (40 °C, 60 °C or 80 °C) had no significant influence on the quality of the observed signals. All samples thereafter were analysed in 15 mM NH_4OAc, pH 7.5, at a source temperature of 40 °C.

Non-covalent interactions can usually be studied at sample cone voltages around 50 V-80 V.[511] However, lowering of the sample cone voltages is usually accompanied by an increased background signal. Complex formations of AspH with metals, cosubstrates or analogues, and peptides were monitored using a sample cone voltage of 60 V unless otherwise indicated. Prior to analysis, a preincubation time of 15 min was allowed for complex formation.

Following incubation of His_6-$AspH_{315-758}$ with Fe(II) or Mn(II), no clear peaks indicating binding and AspH·Fe^{II} or AspH·Mn^{II} complex formation respectively were observed. Instead, the major peaks in both cases corresponded to apo-AspH (**Fig. 5-10**). Upon addition of NOG, stable peaks corresponding to the AspH·Fe^{II}·NOG or the AspH·Mn^{II}·NOG complex respectively were observed.

A peak at 54,697 Da (a mass addition of +178 Da to the apo-protein), likely corresponding to His_6-$AspH_{315-758}$ with a gluconoylated His_6-tag,[390,391] complicated the analysis and may have masked other peaks. The His_6-tag was therefore cleaved off using thrombin, as described in **Chapter 10.**

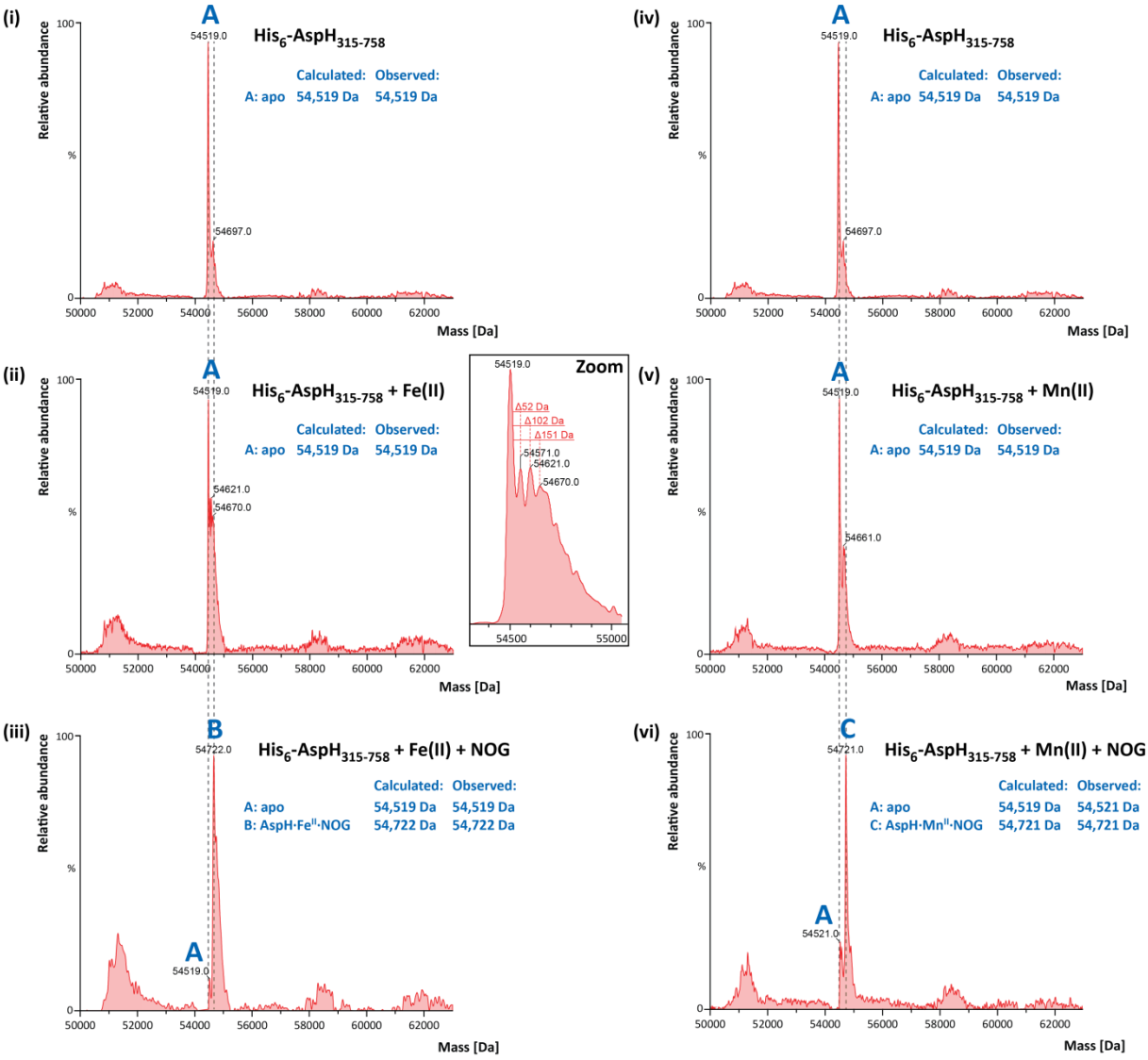


Fig. 5-10 | Deconvoluted non-denaturing ESI-MS spectra of His_6-$AspH_{315-758}$ (15 µM) in the presence of equimolar Fe(II) or Mn(II) and NOG, as indicated, at 60 V sample cone voltage (15 min preincubation time for complex formation prior to analysis). A clear peak corresponding to His_6-$AspH_{315-758}$ in complex with Fe(II), AspH·Fe^{II} (mass addition of 56 Da to the mass of apo-AspH, expected: 54,575 Da), or Mn(II) respectively, AspH·Mn^{II} (mass addition of 55 Da, expected: 54,574 Da), was not observed but mainly a peak corresponding to the apo-protein (A: 54,519 Da) instead (**(ii)** and **(v)**). After incubation with both metal and NOG (mass: 147 Da), peaks corresponding to the complexes **(iii)** AspH·Fe^{II}·NOG (B: 54,722 Da) and **(vi)** AspH·Mn^{II}·NOG (C: 54,721 Da) were the observed major species.

The His_6-tag cleavage from $AspH_{315-758}$ was monitored using denaturing ESI-MS (**Fig. 5-11**). After 20 h incubation with thrombin (Restriction Grade, *Novagen* Thrombin Cleavage Capture Kit) at 4 °C, two cleavage products were observed (expected product mass: 52,768.2 Da, observed masses: 52,767.5 Da and 51,250 Da). In addition to the typical thrombin cleavage site sequence (cleavage site 1: LVPR^GS), a second cleavage site (GR^GS) was identified in the His_6-$AspH_{315-758}$ sequence (in the pET-28a(+) vector). After continued incubation for 40 h, the cleavage reaction was completed and a single product was detected (mass: 51,250 Da).

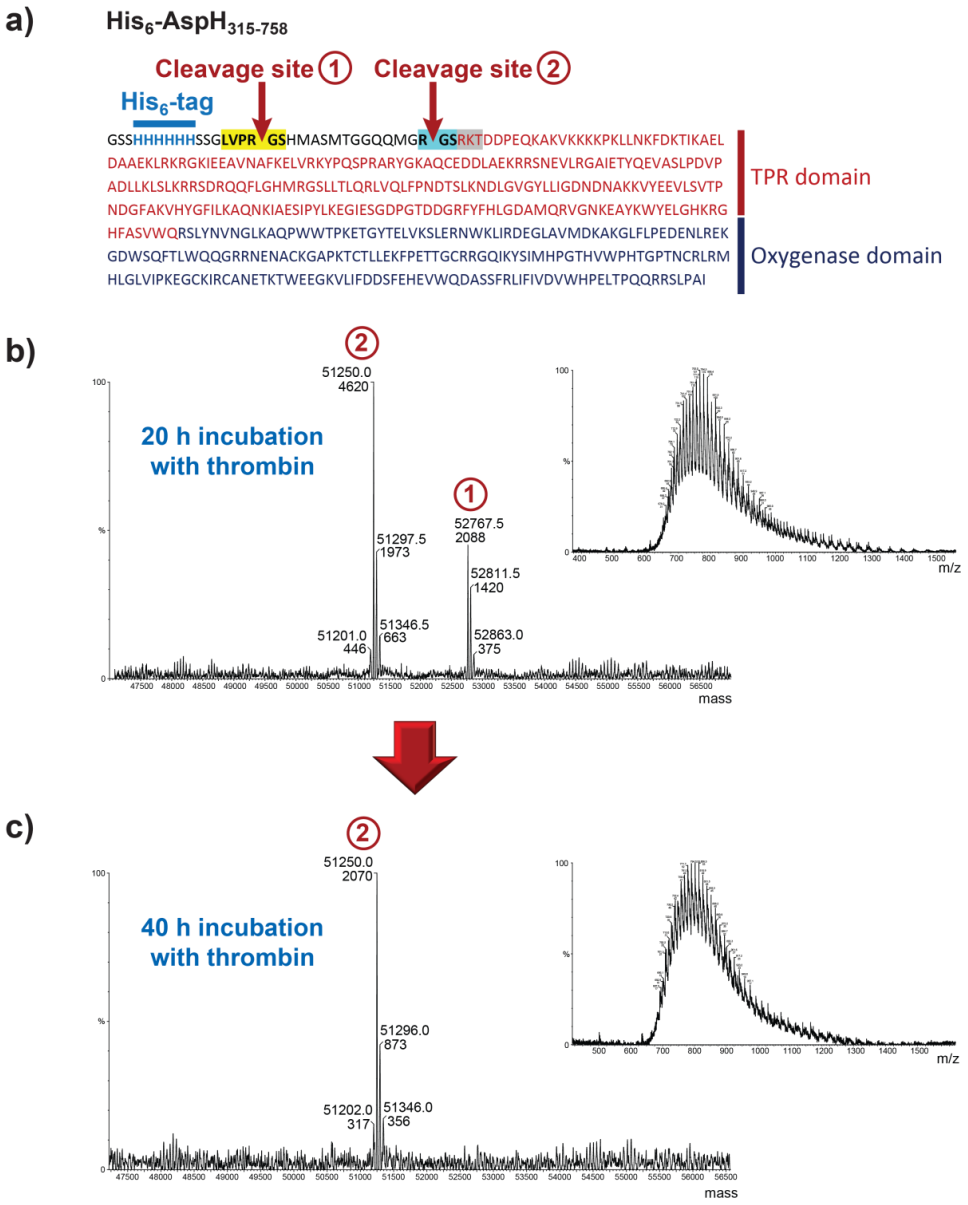


Fig. 5-11 | His_6-tag cleavage monitoring using denaturing ESI-MS (represented as deconvoluted mass spectra and charge state envelopes). **a)** His_6-$AspH_{315-758}$ sequence (in the pET-28a(+) vector) with indicated observed thrombin cleavage sites. The typical thrombin cleavage site sequence is cleavage site 1 (LVPR^GS). A cleavage site 2 (GR^GS) was identified. **b)** After 20 h incubation of His_6-$AspH_{315-758}$ with thrombin two cleavage products were observed. **c)** After 40 h incubation a single cleavage product was found.

In the non-denaturing ESI-MS analysis of $AspH_{315-758}$ (15 µM, His_6-tag cleaved) with equimolar Mn(II) and NOG at 60 V sample cone voltage several peaks were observed in the deconvoluted mass spectra (**Fig. 5-12**). The major peak was assigned to the AspH·Mn^{II}·NOG complex (C: 51,452 Da). In addition, a peak corresponding to the apo-protein (A: 51,250 Da) was observed, suggesting that the AspH·Mn^{II}·NOG complex was not stable under the conditions employed. The sample cone voltage was therefore varied to optimise transmission of the protein complex and improve the quality of the spectra. Increasing the sample cone voltage can provide the ions with higher kinetic energy and improve the resulting spectrum by "collisional cleaning".[512,513] However, at too high voltages the protein may begin to unfold and dissociation of non-covalent complexes occurs.[512,514] A lowering of

the sample cone voltage to 50 V, resulted in an unsatisfactory signal-to-noise ratio. At 80 V sample cone voltage, the AspH·Mn^{II}·NOG complex peak intensity was strongly reduced, indicating that specific binding and non-covalent complex formation were suppressed. Simultaneously, an increased intensity of a peak that may correspond to an AspH·Mn^{II} complex (B: 51,304 Da) was observed. This peak, however, was also stable at 200 V sample cone voltage, which likely indicates unspecific Mn(II)-binding. Another peak observed in the spectra at 60 V sample cone voltage (D: 51,496 Da, corresponding to a mass addition of +246 Da to the apo-protein mass) could not be assigned but likely corresponded to a non-covalent complex because the peak was not observed at 200 V sample cone voltage (**Fig. 5-12**).

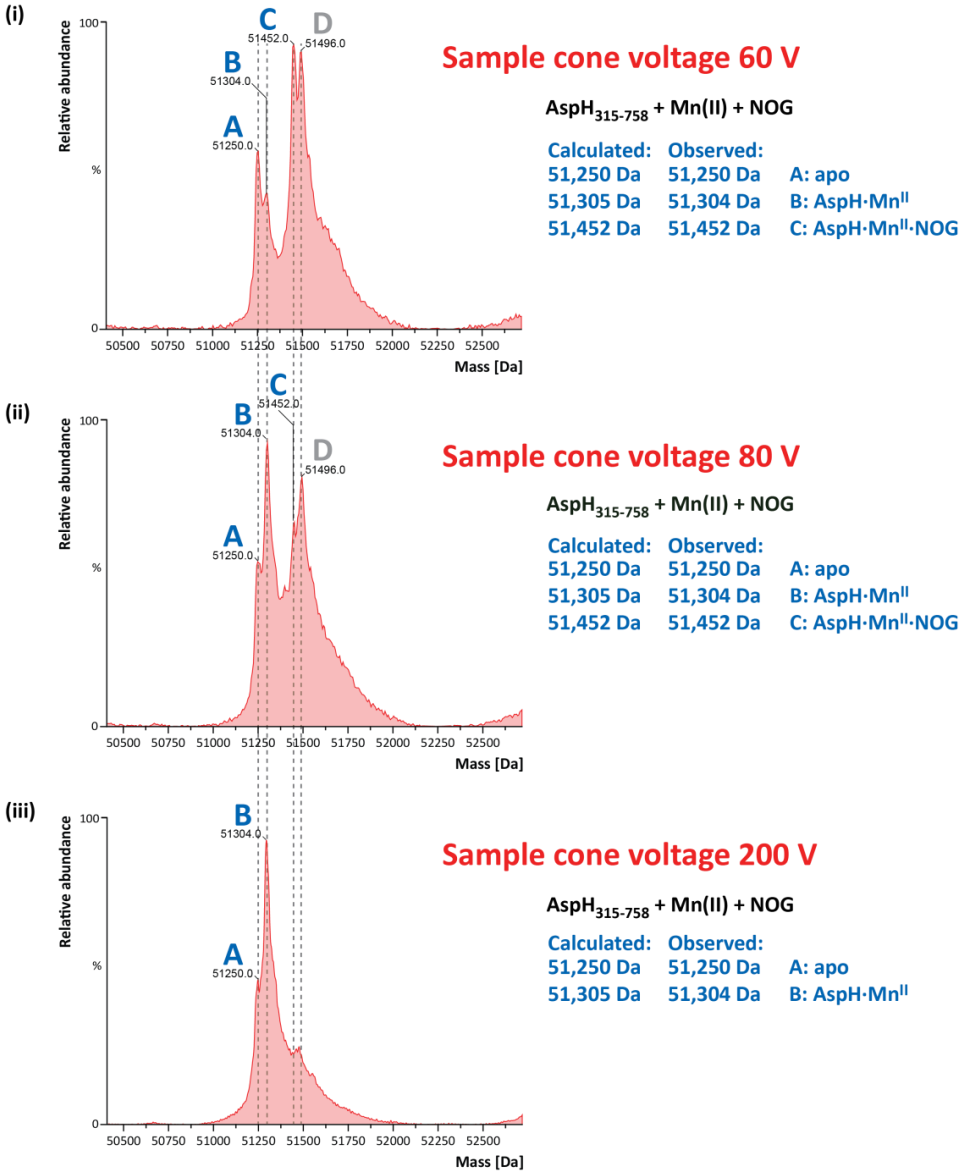


Fig. 5-12 | Sample cone voltages variation and complex formation. Deconvoluted non-denaturing ESI-MS spectra of $AspH_{315-758}$ (15 µM, His_6-tag cleaved) in the presence of equimolar Mn(II) and NOG (15 min preincubation time for complex formation prior to analysis). **(i)** At 60 V sample cone voltage several peaks were observed, corresponding to: apo-protein (A: 51,250 Da), AspH·Mn^{II} (B: 51,304 Da), AspH·Mn^{II}·NOG (C: 51,452 Da). **(ii)** At 80 V sample cone voltage specific binding and non-covalent complex formation are suppressed (diminished signal for C: AspH·Mn^{II}·NOG). **(iii)** Dissociation of non-covalent complexes at 200 V sample cone voltage. The stable peak with a mass of 51,304 Da (B) may indicate non-specific Mn(II)-binding. The apparently non-covalent complex D (51,496 Da, corresponding to a mass addition of +246 Da relative to the apo-protein mass) could not be assigned.

In non-denaturing ESI-MS studies of $AspH_{562-758}$ (oxygenase domain, 15 µM, His_6-tag cleaved) no peak indicating Mn(II)-binding was observed after incubation with Mn(II) at sample cone voltages lowered to 20 V (**Fig. 5-13**). Following incubation with Mn(II) and NOG, peaks for both the apo-protein and a binary complex, $AspH_{562-758}$·NOG, were observed. However, a sample cone voltage as low as 30 V was required, and the complex was labile at higher sample cone voltages. Following incubation with the NC-$4Ser_{39mer}$ peptide in presence of Mn(II) and NOG, peaks corresponding to the apo-protein, the binary complex, $AspH_{562-758}$·NOG, and a ternary complex, $AspH_{562-758}$·Mn^{II}·NOG were observed but no binding of the NC-$4Ser_{39mer}$. The non-deconvoluted mass spectra showed uncomplexed NC-$4Ser_{39mer}$ peptide in the same sample (**Fig. 5-13b**).

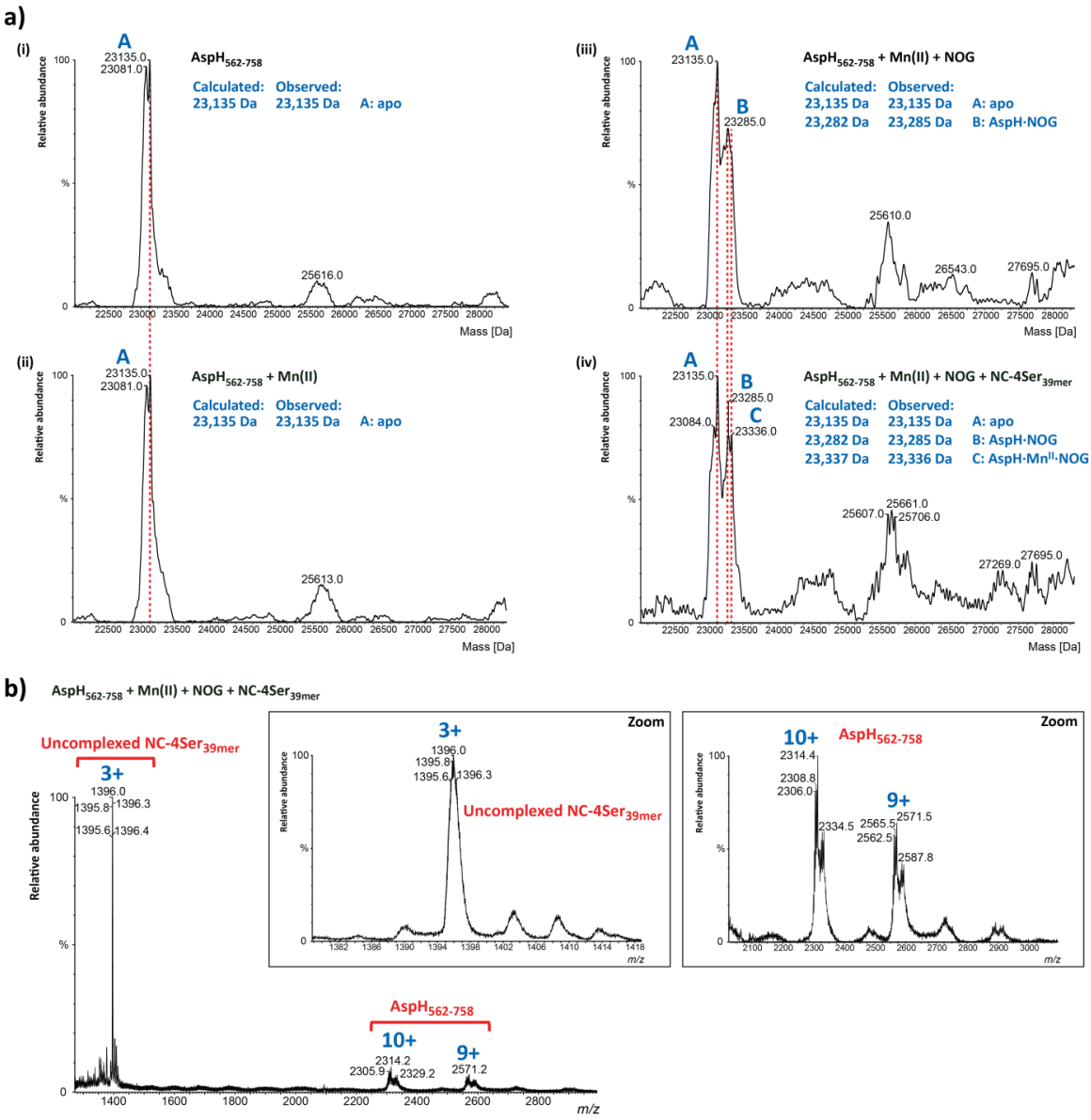
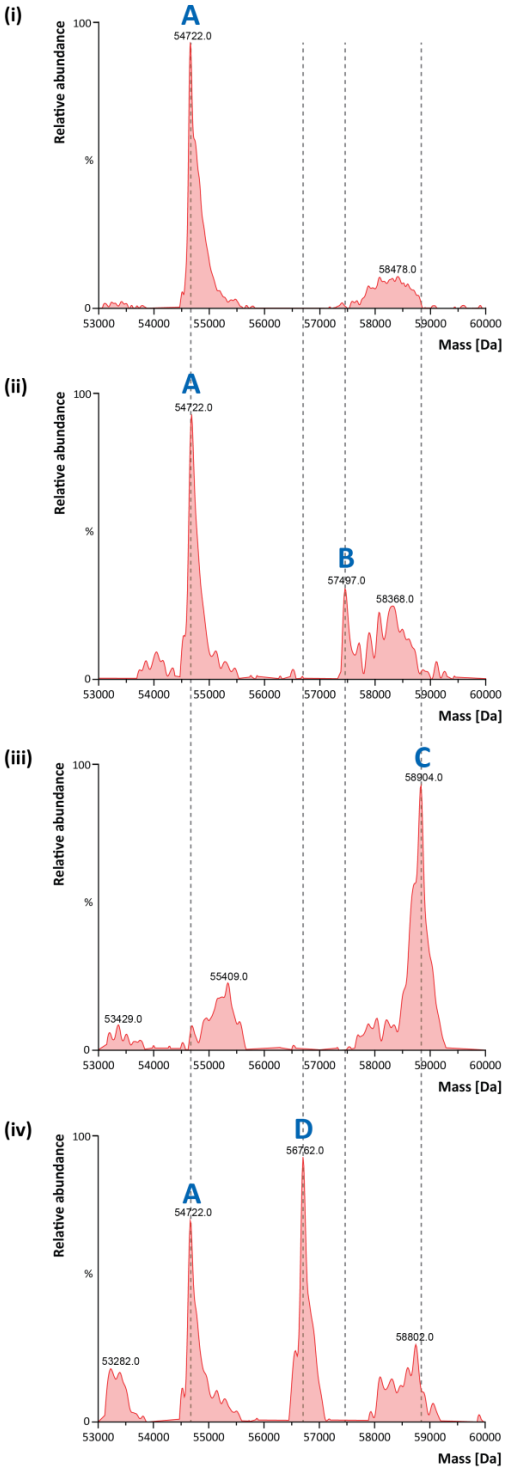


Fig. 5-13 | Non-denaturing ESI-MS of $AspH_{562-758}$ (oxygenase domain, 15 µM, His_6-tag cleaved, 23,135 Da) in the presence of equimolar Mn(II), NOG and the NC-$4Ser_{39mer}$ peptide at 30 V sample cone voltage (no complex formation was observed at higher sample cone voltages; 15 min preincubation time for complex formation prior to analysis). **a)** Deconvoluted mass spectra. No peak indicating Mn(II)-binding to $AspH_{562-758}$ was observed **(ii)**. Incubation with Mn(II) and NOG appeared to result in formation of a binary complex, $AspH_{562-758}$·NOG (B: 23,285 Da, **(iii)**). Incubation with the NC-$4Ser_{39mer}$ peptide in presence of Mn(II) and NOG led to peaks corresponding to the apo-protein, the binary complex, $AspH_{562-758}$·NOG (B: 23,285 Da), and a ternary complex, $AspH_{562-758}$·Mn^{II}·NOG (C: 23,336 Da, **(iv)**). **b)** Non-deconvoluted mass spectra corresponding to **(iv)** showing uncomplexed NC-$4Ser_{39mer}$ peptide in the sample.

Substrate binding studies using His_6-$AspH_{315-758}$ with equimolar amounts of Fe(II), NOG and NC-$4Ser_{39mer}$, NC_{26mer}, or $CP2_{19mer}$ respectively supported the findings from the preceding peptide binding and activity experiments, and showed that the NC-$4Ser_{39mer}$ peptide formed a stable complex with His_6-$AspH_{315-758}$ (**Fig. 5-14**). Complexes with the NC_{26mer} and the $CP2_{19mer}$ were also observed but appeared less stable under the conditions employed, since a peak representing AspH·Fe^{II}·NOG was also found. The $CP2_{19mer}$ (approximately 60% complexed by AspH) appeared to be a stronger binder than the NC_{26mer} peptide in these experiments, supporting the importance of the described interactions between the *C*-terminal peptide region and the AspH TPR domain.

Substrate binding studies using $AspH_{315-758}$ with equimolar amounts of Mn(II), NOG and the C_{26mer} or the NC_{26mer} respectively showed no binding of the canonical C_{26mer} but quaternary complex formation with the non-canonical NC_{26mer} (AspH·Mn^{II}·NOG·NC_{26mer}, **Fig. 5-15**). These results were consistent with the activity assay (**Chapter 3**) and the differential scanning fluorimetry binding assay data (**Section 5.1.2.2**). Partial quaternary complex formation was also observed with the hFX $EGF1_{39mer}$ (AspH·Mn^{II}·NOG·hFX $EGF1_{39mer}$). Interestingly, while no $CP2_{10mer}$-binding was found, both a quaternary complex, AspH·Mn^{II}·NOG·$CP2_{19mer}$, and a binary complex, AspH·$CP2_{19mer}$, were apparent (**Fig. 5-15**). In the non-deconvoluted spectra of the samples, where no or only partial binding of the peptides were found, peaks corresponding to the uncomplexed peptides were simultaneously observed. In the case of the hFX $EGF1_{39mer}$ this may have been the canonical disulfide isoform in the peptide mixture.

The results from the binding studies suggested that, in contrast to the shorter $CP2_{10mer}$, the $CP2_{19mer}$ could enable stable complex formation with AspH sufficient for crystallisation.



His_6-$AspH_{315-758}$ + Fe(II) + NOG

Calculated:	Observed:	
54,722 Da	54,722 Da	A: AspH·Fe^{II}·NOG

His_6-$AspH_{315-758}$ + Fe(II) + NOG + NC_{26mer}

Calculated:	Observed:	
54,722 Da	54,722 Da	A: AspH·Fe^{II}·NOG
57,498 Da	57,497 Da	B: AspH·Fe^{II}·NOG·NC_{26mer}

His_6-$AspH_{315-758}$ + Fe(II) + NOG + NC-$4Ser_{39mer}$

Calculated:	Observed:	
58,904 Da	58,904 Da	C: AspH·Fe^{II}·NOG·NC-$4Ser_{39mer}$

His_6-$AspH_{315-758}$ + Fe(II) + NOG + $CP2_{19mer}$

Calculated:	Observed:	
54,722 Da	54,722 Da	A: AspH·Fe^{II}·NOG
56,765 Da	56,762 Da	D: AspH·Fe^{II}·NOG·$CP2_{19mer}$

Fig. 5-14 | Deconvoluted non-denaturing ESI-MS spectra of His_6-$AspH_{315-758}$ (15 µM) in the presence of equimolar Fe(II), NOG and peptides at 60 V sample cone voltage (15 min preincubation time for complex formation prior to analysis). After incubation with the NC_{26mer} peptide two peaks corresponding to the AspH·Mn^{II}·NOG (A: 54,722 Da) and the AspH·Mn^{II}·NOG·NC_{26mer} (~30% B: 57,497 Da) complexes were observed **(ii)**. Incubation with the NC-$4Ser_{39mer}$ peptide resulted in quaternary complex formation of AspH·Mn^{II}·NOG·NC-$4Ser_{39mer}$ only (C: 58,904 Da, **(iii)**). After incubation with the $CP2_{19mer}$, the observed major peak corresponded to the quaternary complex AspH·Mn^{II}·NOG·$CP2_{19mer}$ (~60% D: 56,762 Da, **(iv)**).

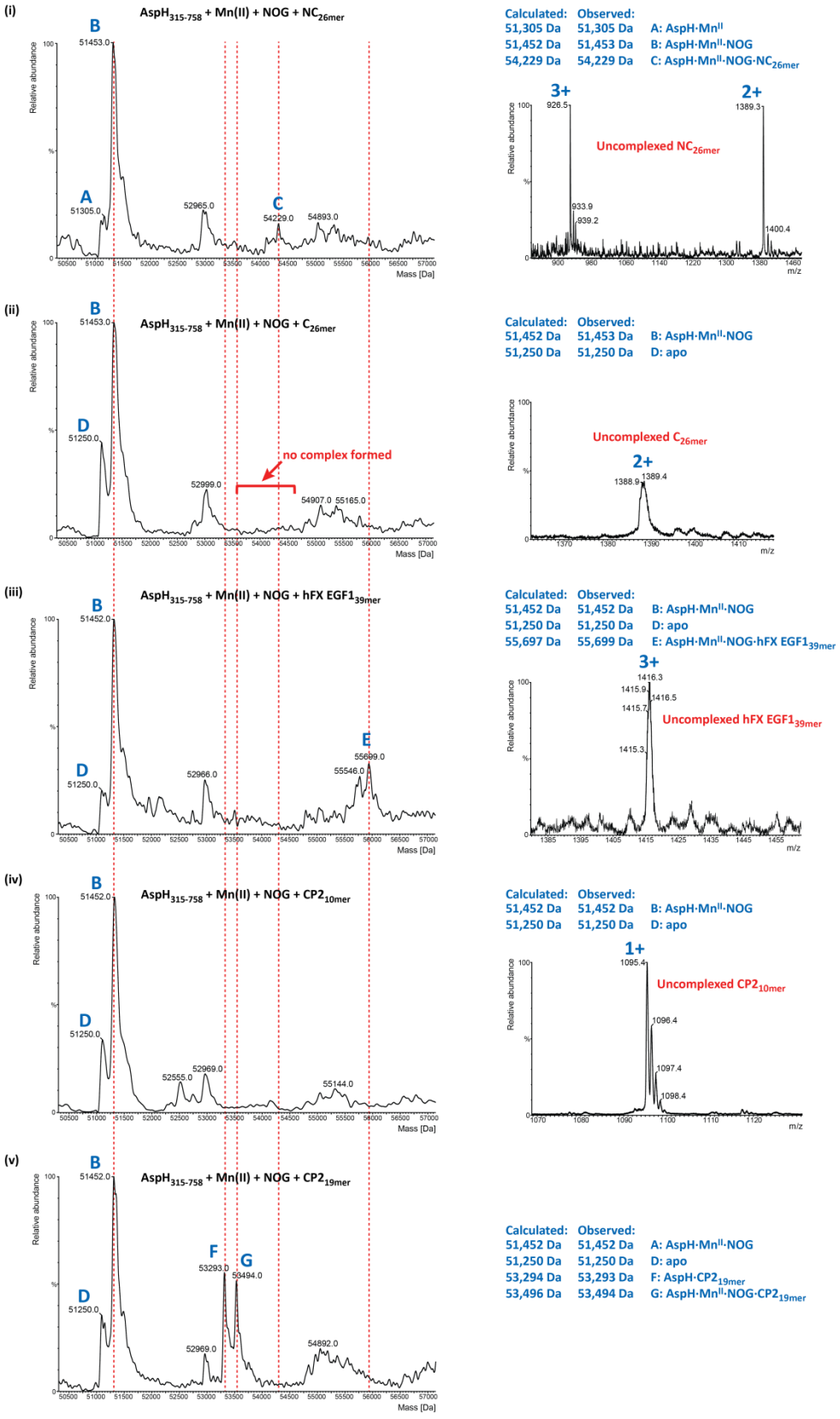


Fig. 5-15 | Non-denaturing ESI-MS spectra of $AspH_{315\text{-}758}$ (15 µM, His_6-tag cleaved) in the presence of equimolar Mn(II), NOG and peptides at 60 V sample cone voltage (15 min preincubation time for complex formation prior to analysis). In the cases, where no or only partial binding of the peptides were observed, peaks corresponding to the uncomplexed peptides were also observed in the same samples. Quaternary complex formation with the non-canonical NC_{26mer} **(i)**, but no sign of binding was observed after incubation of $AspH_{315\text{-}758}$ with the canonical C_{26mer} peptide **(ii)**. Incubation with the hFX $EGF1_{39mer}$ peptide led to partial quaternary complex formation (E: 55,699 Da, **(iii)**). No peak indicating binding of the $CP2_{10mer}$ to $AspH_{315\text{-}758}$ was observed **(iv)**. The $CP2_{19mer}$ appeared to form both a quaternary complex, $AspH{\cdot}Mn^{II}{\cdot}NOG{\cdot}CP2_{19mer}$ (G: 53,494 Da), and a binary complex, $AspH{\cdot}CP2_{19mer}$ (F: 53,293 Da, **(v)**).

5.1.2.4 Crystal Structure of AspH$_{315\text{-}758}$ in Complex with CP2$_{19mer}$

Co-crystallisation trials of His$_6$-AspH$_{315\text{-}758}$ with 1 mM $MnCl_2$, 2 mM NOG and the CP2$_{19mer}$ were then set up with an AspH-to-substrate ratio of 1:10 and 1:5. The crystals grown within a few weeks in the co-crystallisation trials of AspH with the CP2$_{19mer}$ peptide appeared to have a different morphology compared to the orthorhombic AspH-substrate complex crystals obtained before (**Fig. 5-16**).

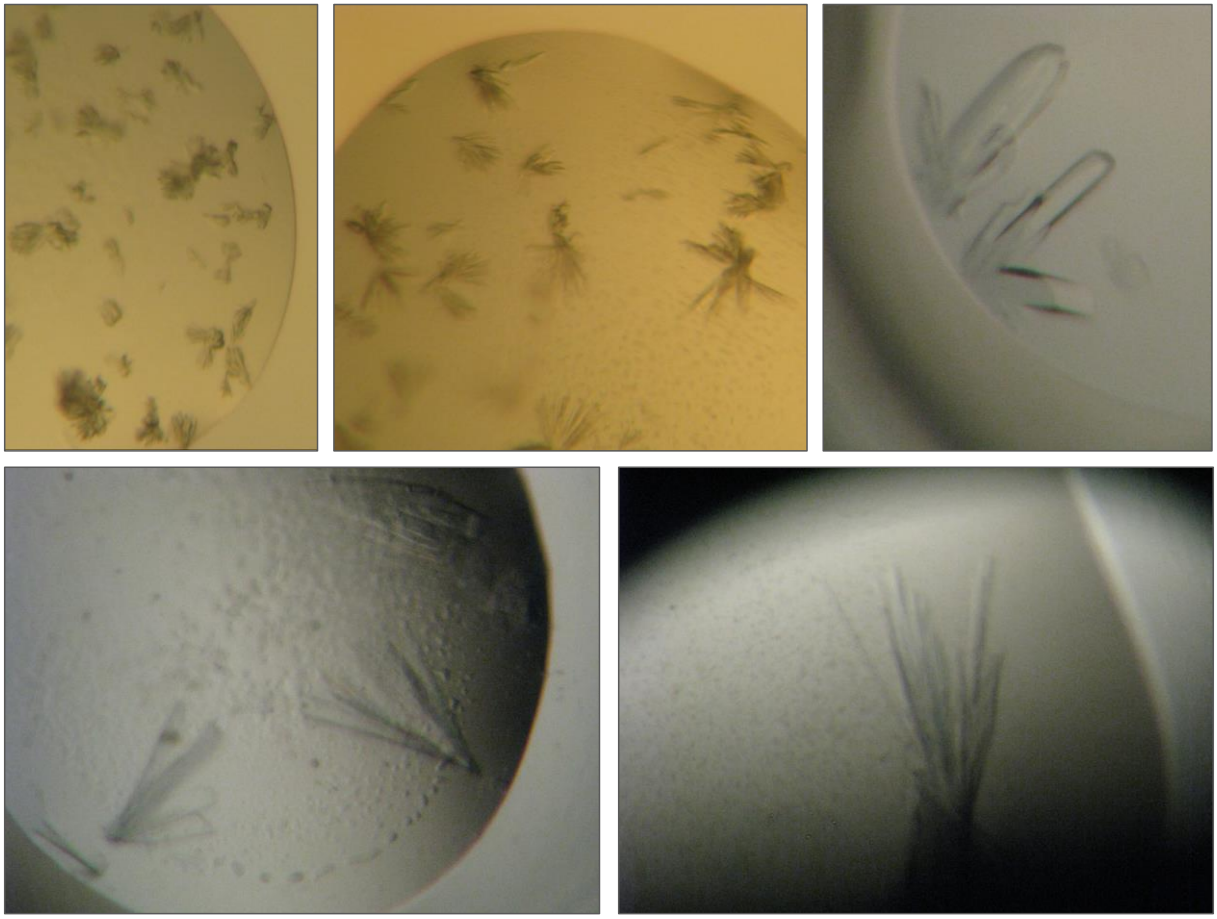


Fig. 5-16 | Morphology of the AspH$_{315\text{-}758}$·Mn^{II}·NOG·CP2$_{19mer}$ crystals, space group *P*1. The crystals grew in sitting drops at 4 °C.

Diffraction (2.60 Å, **Table 5-1**) of the harvested crystals was carried out using in-house X-ray radiation and were assigned to triclinic symmetry, space group *P*1. The crystal structure was solved by Dr Michael A. McDonough by molecular replacement. Interestingly, two molecules were found in the asymmetric unit but only one molecule contained the bound CP2$_{19mer}$, corresponding to the complex AspH$_{315\text{-}758}$·Mn^{II}·NOG·CP2$_{19mer}$, the other molecule corresponded to "apo"-AspH (**Fig. 5-17**).

The peptide ring and the *C*-terminal tail were apparent in the electron density map. In good agreement with the predicted model, the peptide adopted a binding mode very similar to the disulfide-containing substrates in the other AspH-substrate complex structures. The key interactions of the CP2$_{19mer}$ with AspH in the active site and the hydrophobic pockets within the TPR domain were consistent with the prior crystallographic studies on AspH-substrate complexes.

Table 5-1 | Crystallisation conditions and data collection for the $AspH_{315-758}$ structures determined by molecular replacement (work still in progress; refinements performed by Dr Michael A. McDonough).

	$AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}NC\text{-}4Ser_{39mer}$	$AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}NC_{26mer}$	$AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}CP2_{19mer}$
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P1$
Cell dimensions			
a, *b*, *c* [Å]	50.17, 89.54, 123.55	50.23, 90.81, 124.09	49.47, 59.27, 95.66
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00	104.00, 91.50, 92.70
Resolution [Å]	1.99	2.50	2.60
Completeness [%]	100.0	99.9	99.0
Symmetry	Orthorhombic	Orthorhombic	Triclinic
Crystal size	250 x 100 µm	150 x 100 µm	150 x 20 µm
Method	Vapour diffusion, sitting drop (300 nl), protein-to-well ratio 2:1	Vapour diffusion, sitting drop (200 nl), protein-to-well ratio 1:1	Vapour diffusion, sitting drop (200 nl), protein-to-well ratio 1:1
Temp. [K]	277	277	277
Precip. cond.	PACT Premier, A5-well1 18 mg/ml His_6-$AspH_{315-758}$, 1 mM $MnCl_2$, 2 mM NOG, 100 mM SPG buffer pH 8.0, 25 % w/v PEG1500, 726 µM $NC\text{-}4Ser_{39mer}$	PACT Premier, F10-well2 18 mg/ml His_6-$AspH_{315-758}$, 1 mM $MnCl_2$, 2 mM NOG, 100 mM Bis Tris Propane, pH 6.5, 20 mM sodium/ potassium phosphate, 20 % w/v PEG3350, 1.65 mM NC_{26mer}	PACT Premier, F4-well2 18 mg/ml His_6-$AspH_{315-758}$, 1 mM $MnCl_2$, 2 mM NOG, 100 mM Bis Tris Propane, pH 6.5, 200 mM potassium thiocyanate, 20 % w/v PEG3350, 3.3 mM $CP2_{19mer}$
AspH:Substrate ratio	1:2.2 (330 µM:726 µM)	1:5 (330 µM:1.65 mM)	1:10 (330 µM:3.3 mM)
Cryoprotectant	25 % glycerol	25 % glycerol	25 % glycerol
Diffraction Source	Synchrotron	Synchrotron	Copper rotating anode
Site	Diamond Light Source	Diamond Light Source	In-house diffractometer
Type	Diamond Beamline I04	Diamond Beamline I04	CuKα
Wavelength List [Å]	0.9795	0.9795	1.542
Detector	Pilatus 6M-F	Pilatus 6M-F	Saturn 944+ CCD detector

The *C*-terminal peptide region including the highly conserved aromatic residue corresponding to Phe31 in hFX $EGF1_{39mer}$ was observed to form interactions in the hydrophobic pocket within the TPR domain, consistent with the AspH-substrate interactions observed in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}NC\text{-}4Ser_{39mer}$ crystal structure (**Fig. 5-17b**). Only the three *C*-terminal $CP2_{19mer}$ peptide residues (corresponding to Glu32, Gly33, Lys34 in the hFX $EGF1_{39mer}$) were not apparent in the electron density map and likely disordered to higher flexibility.

Superimposition of the two molecules found in the asymmetric unit showed the differences in the orientation of important residues, including Ile758 and Gln633, due to the "induced fit" binding of the $CP2_{19mer}$ (**Fig. 5-17c**). Interestingly, the orientation of the $Lys17_{CP2\text{-}19mer}$ side chain was different from the orientation of the $Lys17_{NC\text{-}4Ser39mer}$ side chain in the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}NC\text{-}4Ser_{39mer}$ structure. The $Lys17_{CP2\text{-}19mer}$ side chain interacted with the carboxylate of Ile758 (**Fig. 5-17c** and **Fig. 5-18**), whereas the $Lys17_{NC\text{-}4Ser39mer}$ side chain formed a hydrogen bond with Asp616 (**Fig. 5-2**).

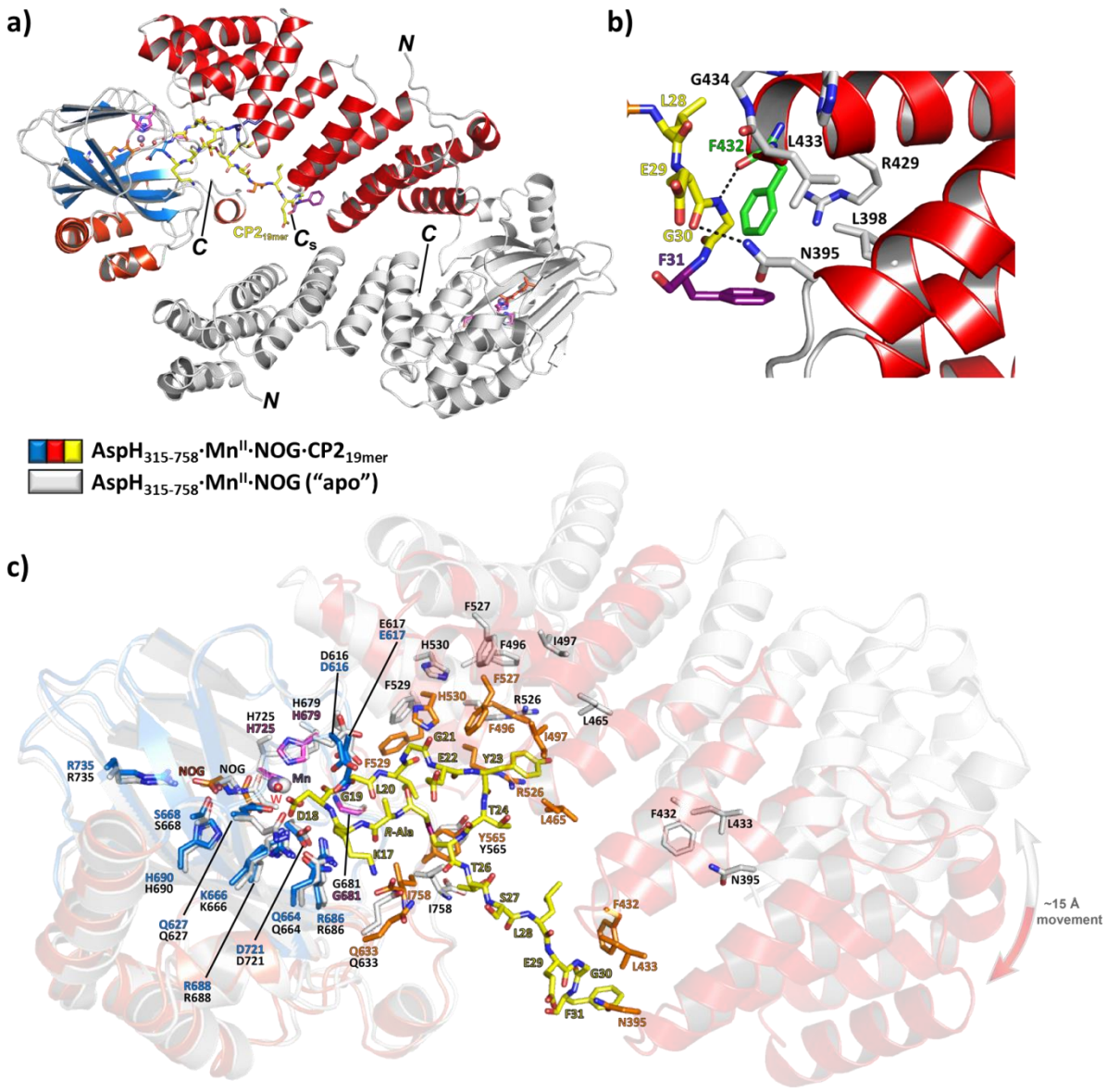


Fig. 5-17 | View from a AspH$_{315-758}$·Mn^{II}·NOG·CP2$_{19mer}$ complex crystal structure (resolution: 2.60 Å). **a)** Two molecules were found in the asymmetric unit, but only one contained the bound peptide; the other one corresponded to "apo"-AspH. **b)** Close-up showing the important interactions of the *C*-terminal region of the peptide, including the highly conserved aromatic key residue corresponding to Phe31 in the hFX EGF1$_{39mer}$, with a hydrophobic pocket around Phe432 in the TPR domain. **c)** Superimposition of the two molecules showing differences in the orientation of important residues.

The thioether link in the peptide was clearly observed in the electron density map (**Fig. 5-18**). Superimposition of the CP2$_{19mer}$ peptide in the crystal structure and the model constructed for the CP2$_{10mer}$ (see **Chapter 4**) showed small deviations in the orientation of the residues. Interestingly, as already predicted in the model, the peptide ring structure was stabilised by an intramolecular hydrogen bond between the carbonyl of the linker moiety and the backbone N-H of Gly21$_{CP2-19mer}$ (**Fig. 5-18d**).

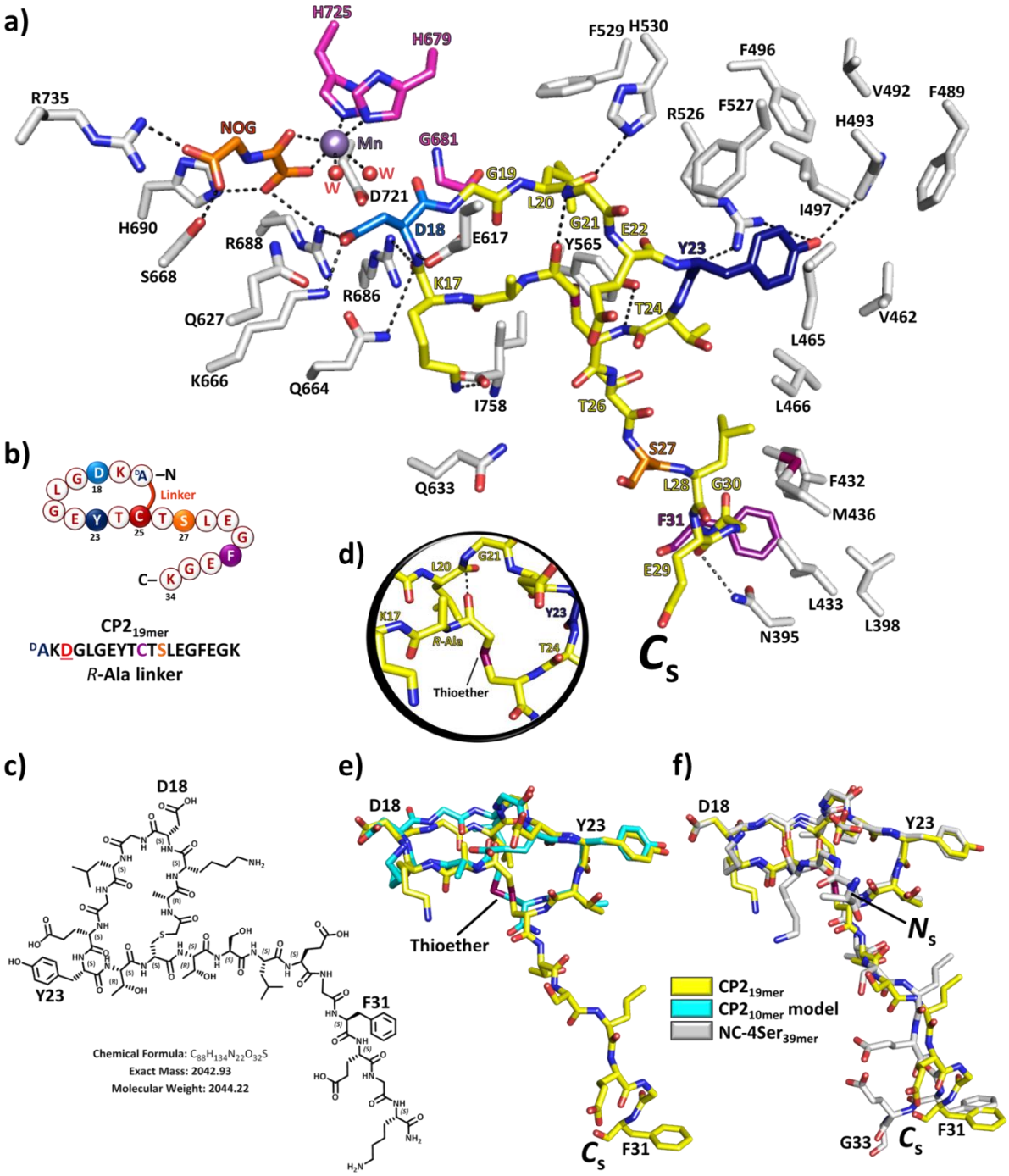


Fig. 5-18 | Important interactions between AspH and the CP2$_{19mer}$ peptide. **a)** Excerpt from the AspH$_{315-758}$·Mn^{II}·NOG·CP2$_{19mer}$ crystal structure. Hydrophobic residues lining the 2OG-binding pocket in the active site are omitted for clarity. **b)** Schematic of the CP2$_{19mer}$, **c)** skeletal formula. **d)** Close-up of the thioether linkage and an intramolecular hydrogen bond. **e)** Superimposition of the CP2$_{19mer}$ (yellow) in the crystal structure with the CP2$_{10mer}$ model (cyan). **f)** Superimposition of the CP2$_{19mer}$ (yellow) with the NC-4Ser$_{39mer}$ (grey) in the AspH$_{315-758}$·Mn^{II}·NOG·NC-4Ser$_{39mer}$ crystal structure.

5.2 MUTAGENESIS STUDIES

Crystal structures can only capture a steady state of a protein, but proteins in solution under physiological conditions will always have a degree of flexibility and be dynamic. The roles and importance of residues found to be involved in key interactions in the crystal structures were

therefore further investigated and validated by mutagenesis studies. Substituting alanine for interacting residues can disable hydrogen bonds, electrostatic and (substantially) hydrophobic interactions between amino acids. Substitution of potentially crucial residues by alanine can therefore reveal their contribution to recognition and binding of the substrate. Based on the AspH-substrate crystal structures, several residues involved in interactions were selected for an alanine scan[515,516] and studied using hydroxylation activity assays.

The ConSurf server[517,518] is a bioinformatics tool to estimate the evolutionary conservation of amino acid positions in a protein based on the phylogenetic relations between homologous sequences. Using the ConSurf server tool, conserved residues in AspH were identified mainly in the active site and the substrate-binding pockets (**Fig. 5-19**). These results supported the choice of important residues for mutagenesis studies.

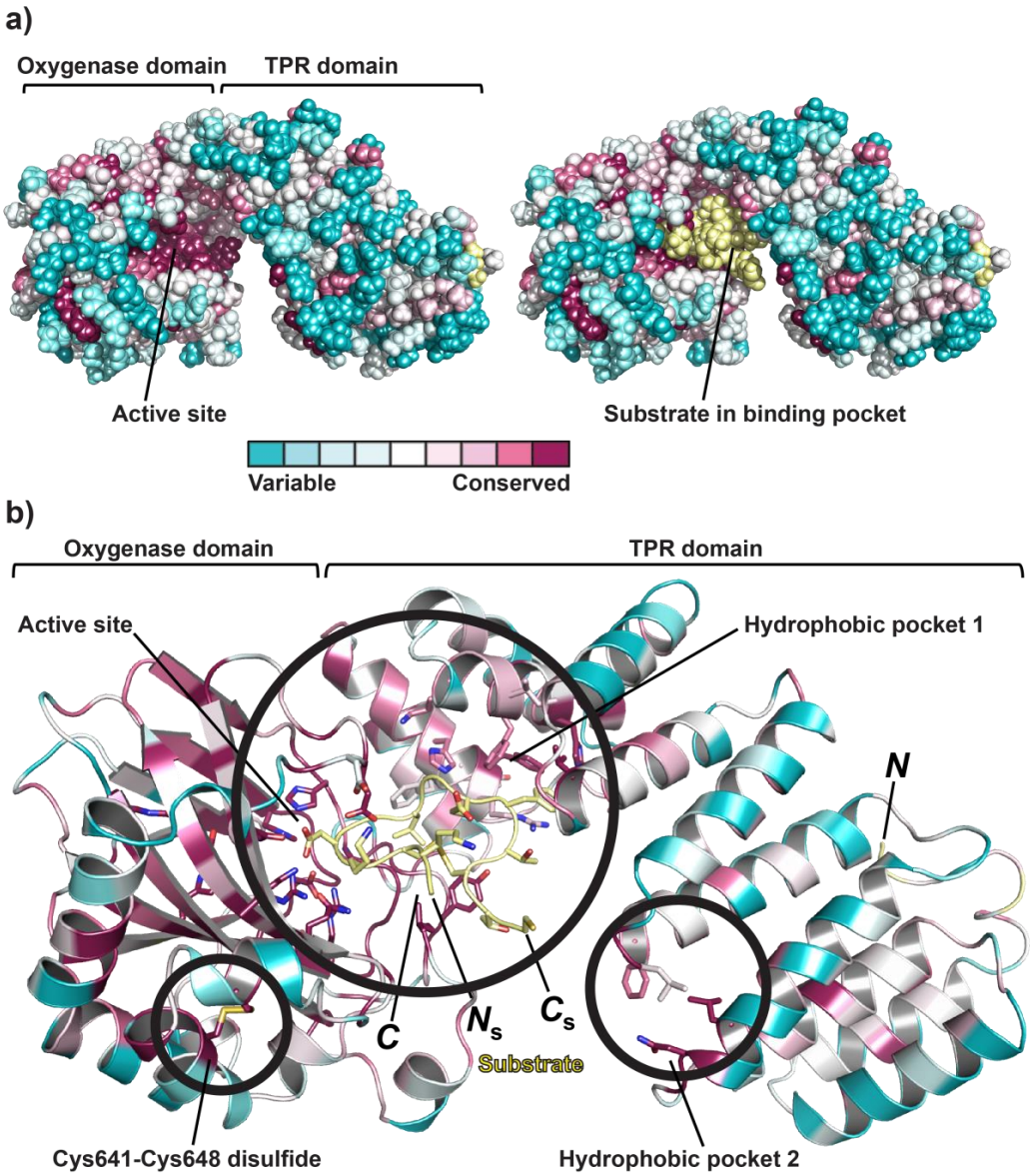


Fig. 5-19 | ConSurf server[517,518] analysis of evolutionary conservation of amino acid positions in the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ complex crystal structure data. **a)** Surface representation. **b)** Overview showing that conserved residues in AspH were identified mainly in the active site and the other substrate-binding pockets.

The substitutions were incorporated in the sequence encoding for His_6-$AspH_{315-758}$ by site-directed mutagenesis, as described in **Chapter 10.** Successful mutagenesis was verified by DNA sequencing, and the correct masses of the produced variant proteins were confirmed by denaturing protein ESI-MS.

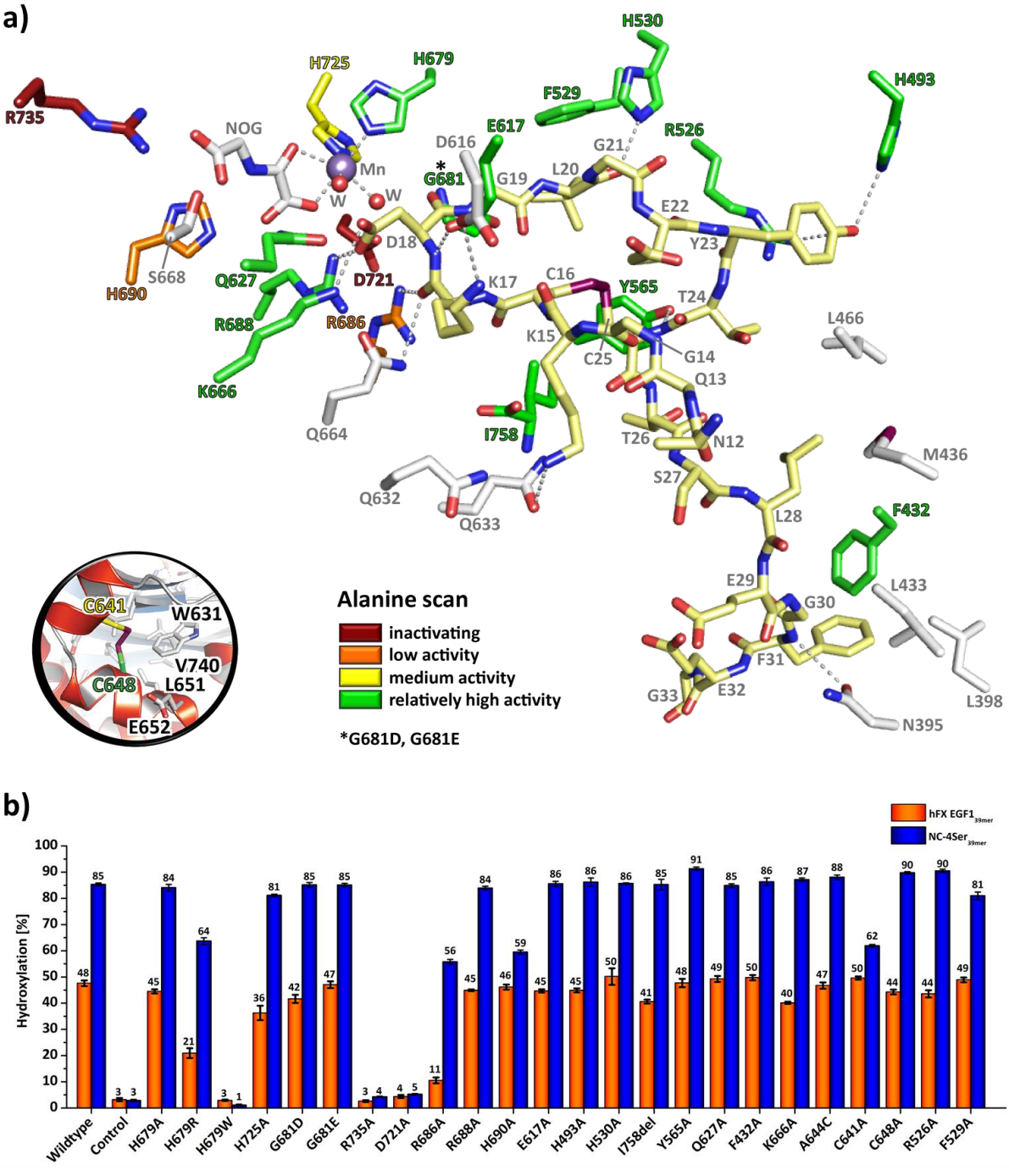


Fig. 5-20 | Mapping of the produced and tested variants on the $AspH_{315-758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}NC\text{-}4Ser_{39mer}$ structure. The activities were rated using the shown colour coding scheme. Important active site and TPR domain residues, and the AspH Cys641-Cys648 residues forming a disulfide in all crystal structures were replaced by an alanine residue to disable hydrogen bonds, electrostatic and hydrophobic interactions between residues (alanine scan, except the Gly681 mutation to Asp or Glu). For the hydroxylation activity assays, a new batch of the hFX $EGF1_{39mer}$ was used that showed about ~50% maximum hydroxylation activity using standard conditions (i.e. non-RedOx). The error bars represent the SD of triplicate measurements.

Table 5-2 | Overview of the produced $AspH_{315-758}$ variants and their activities (including data for wildtype and the background (no enzyme) control). The data are graphically represented in **Fig. 5-20**. The errors represent the SD of triplicate experiments. Ox. domain: oxygenase domain.

Variant	Role/Interactions	Region	Hydroxylation activity [%]	
			hFX $EGF1_{39mer}$	NC-$4Ser_{39mer}$
Wildtype	-	-	47.6 ± 1.1	85.3 ± 0.5
Control	-	-	3.2 ± 0.6	2.8 ± 0.5
H679A	HXG…H, Fe-binding	Active site	44.5 ± 0.8	84.1 ± 1.2
H679R	HXG…H, Fe-binding	Active site	20.9 ± 1.9	63.7 ± 1.3
H679W	HXG…H, Fe-binding	Active site	2.9 ± 0.2	1.1 ± 0.3
H725A	HXG…H, Fe-binding	Active site	36.3 ± 2.7	81.1 ± 0.4
G681D	HXG…H	Active site	41.6 ± 1.6	85.1 ± 0.9
G681E	HXG…H	Active site	47.1 ± 1.3	85.1 ± 0.6
R735A	2OG-binding	Active site	2.6 ± 0.4	4.3 ± 0.3
D721A	binds R686	Active site	4.4 ± 0.6	5.2 ± 0.3
R686A	binds K17-backbone	Active site	10.5 ± 1.1	55.8 ± 0.9
R688A	binds D18, 2OG-binding	Active site	44.9 ± 0.3	83.9 ± 0.7
H690A	2OG-binding	Active site	46.2 ± 1.0	59.5 ± 0.8
E617A	binds D18-backbone	Flexible loop	44.7 ± 0.6	85.5 ± 0.9
H493A	binds Y23	TPR domain	44.9 ± 0.8	86.2 ± 1.6
H530A	binds L20-backbone	TPR domain	50.2 ± 3.2	85.7 ± 0.3
I758del	potential interaction	*C*-terminus	40.6 ± 0.8	85.2 ± 2.0
Y565A	binds C25-backbone	Ox. domain	47.7 ± 1.6	91.3 ± 0.7
Q627A	potential interaction D18	Active site	49.3 ± 1.1	84.9 ± 0.7
F432A	hydrophobic interaction F31	TPR domain	49.8 ± 1.0	86.3 ± 1.5
K666A	binds D18	Active site	40.1 ± 0.4	87.1 ± 0.7
A644C	AspH disulfide proximity	Ox. domain	46.8 ± 1.2	88.1 ± 0.9
C641A	AspH disulfide	Ox. domain	49.6 ± 0.6	61.9 ± 0.5
C648A	AspH disulfide	Ox. domain	44.3 ± 0.9	89.7 ± 0.4
R526A	binds Y23-backbone	TPR domain	43.6 ± 1.4	90.5 ± 0.6
F529A	hydrophobic interaction L20	TPR domain	48.9 ± 0.9	81.0 ± 1.4

All variants were produced from gene expression in *E. coli* BL21 (DE3) for 16 h at 18 °C after induction with 0.1 mM IPTG at OD_{600} 1.3-1.5, as described in **Chapter 10.** Consistent with the wildtype protein purification, the variant proteins were purified using Ni^{2+}-affinity and subsequent size-exclusion chromatography, as required, until sufficient purity for activity tests and crystallisation was achieved. The variants were then tested for hydroxylation activity with both the hFX $EGF1_{39mer}$ and the NC-$4Ser_{39mer}$ substrates using standard conditions, followed by analysis using MALDI-ToF-MS (**Fig. 5-20** and **Table 5-2**). Both the hFX $EGF1_{39mer}$ and the NC-$4Ser_{39mer}$ substrates were tested with the variants because only partial hydroxylation of the hFX $EGF1_{39mer}$ peptide was observed, since the material was a mixture of the canonical and the non-canonical peptides, as described in **Chapter 3**. A new batch of the hFX $EGF1_{39mer}$ peptide material was used for the assays with the variants. Levels of up to approximately 50 % hydroxylation were achieved with this newly synthesised batch (produced

by *Peptide Synthetics*, Hampshire; oxidised using air oxidation as before). The NC-4Ser$_{39mer}$ was used because full hydroxylation could be achieved with this peptide. However, the NC-4Ser$_{39mer}$ sequence was not the same as in the native hFX EGF1$_{39mer}$ substrate and contained only a single disulfide instead of the three disulfide links in the native hFX EGF1$_{39mer}$. Both peptide substrates were therefore tested side by side with the variant proteins and incubated at 37 °C for 60 min in presence of all cosubstrates/cofactors.

Interestingly, only a few of the substitutions were inactivating (**Fig. 5-20**). The R735A substitution led to disruption of the critical salt bridge interaction with the 2OG-binding Arg735, and eliminated the hydroxylase activity. The D721A substitution was also inactivating, and the R686A mutation strongly diminished the activity.

The most surprising finding was that the activity was retained when either of the HXG…H catalytic dyad residues, His679 or His725, was mutated to Ala. The hydroxylation levels achieved with the H679A variant after 60 min incubation were indistinguishable from the ones resulting from incubation with the wildtype. However, the reaction catalysed by the variants proceeded much more slowly. Samples taken after 15 min incubation time showed almost no hydroxylation with the variants, whereas more than 30 % of the peptides were hydroxylated after incubation with the wildtype at this time point. The His679 residue was then also mutated to Arg because this residue is found in the corresponding position in the AspH homologue AspHD2 (see **Fig. 8-13**). Even the H679R variant showed medium activity (21 % hFX EGF1$_{39mer}$ and 64 % NC-4Ser$_{39mer}$ hydroxylation after 60 min), although the positively charged Arg residue may not be able to bind the active site metal. The observed activities suggested that AspH may not require the catalytic dyad to bind the metal, cosubstrate and prime substrate. Alternatively, a conformational change within the active site to allow other residues in close proximity act as Fe(II)-binding surrogates was possible. The H679W substitution was completely inactivating, which demonstrated that the bulky side chain likely induced clashes in the active site with other protein residues interfering with substrate binding.

Crystallisation studies with the H679A variant performed by Dr Rasheduzzaman A. Chowdhury using the same conditions as for the wildtype AspH$_{315\text{-}758}$·Mn^{II}·NOG·NC-4Ser$_{39mer}$ complex (PACT Premier screen at 4 °C) resulted in an AspH$_{315\text{-}758}$-H679A·Mn^{II}·NOG·NC-4Ser$_{39mer}$ complex structure (**Fig. 5-21**; resolution: 1.75 Å, **Table 5-2**). Consistent with the activity assays results, the structure revealed that a single protein ligand was sufficient for metal and substrate binding by AspH. The metal was octahedrally coordinated in the AspH$_{315\text{-}758}$-H679A active site. Besides the His725 ligand, NOG bound the Mn(II) in the typical bidentate manner, and three water molecules took up the remaining ligand positions. The NC-4Ser$_{39mer}$ substrate was bound to the variant in the same way as to the wildtype AspH.

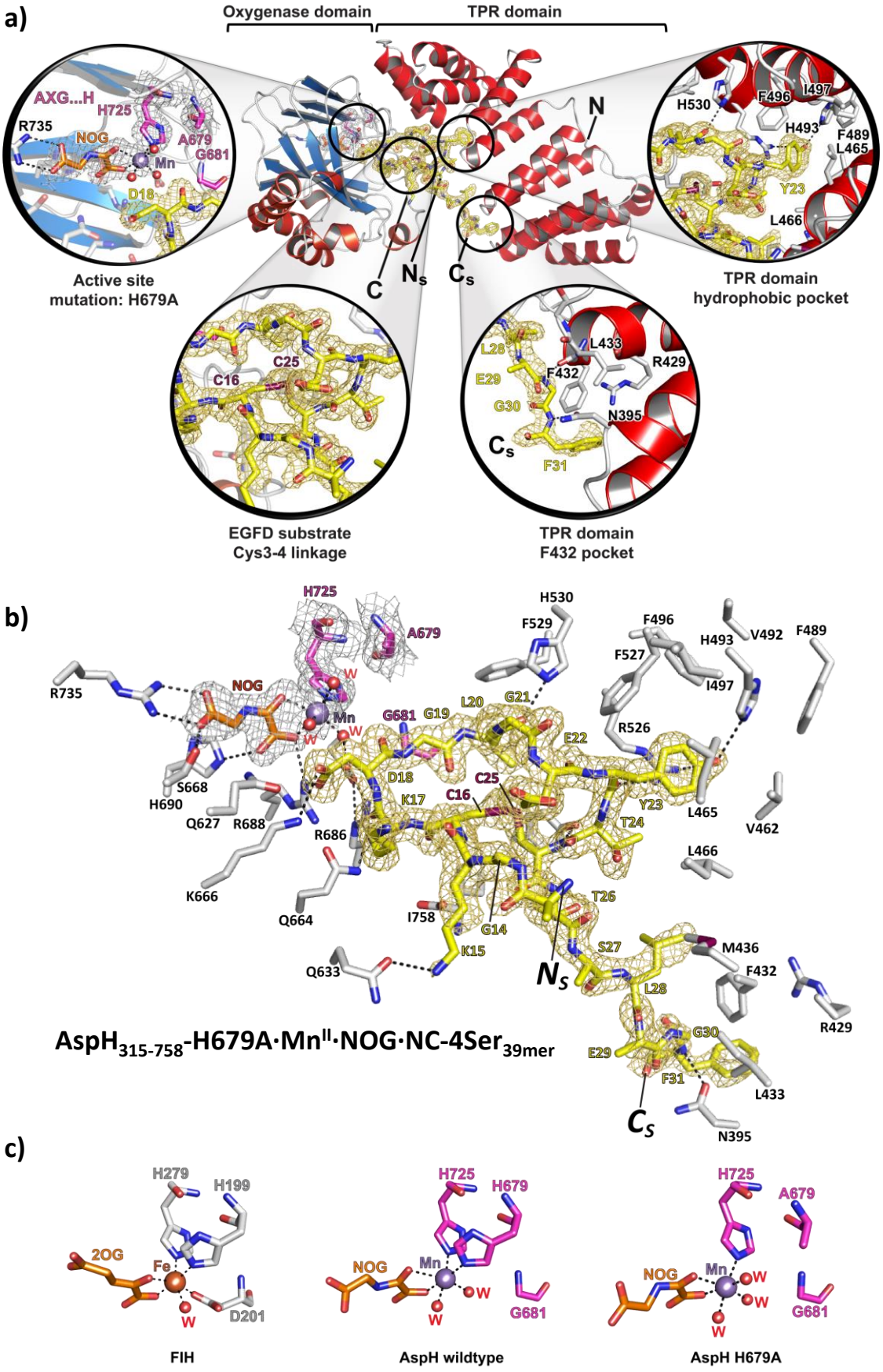


Fig. 5-21 | View from the AspH$_{315\text{-}758}$-H679A·Mn^{II}·NOG·NC-4Ser$_{39mer}$ crystal structure (resolution: 1.75 Å). The experimental F_o-F_c omit electron density maps contoured at 3σ are shown around the metal-coordinating residues and (grey mesh) and the NC-4Ser$_{39mer}$ substrate (yellow mesh). **a)** Overview of the key interactions. **b)** Sticks representation of some residues involved in important interactions. **c)** Active site close-ups of the catalytic triad in FIH (PDB 3P3N[66]), the catalytic dyad in wildtype AspH$_{315\text{-}758}$ and the variant AspH$_{315\text{-}758}$-H679A.

Table 5-2 | Crystallisation conditions and data collection for the $AspH_{315-758}$ H679A structure determined by molecular replacement by Dr Rasheduzzaman Chowdhury.

	$AspH_{315-758}$ H679A·Mn^{II}·NOG·NC-4Ser_{39mer}
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* [Å]	49.81, 91.56,122.92
α, β, γ [°]	90.00, 90.00, 90.00
Resolution [Å]	1.75
Completeness [%]	100
Symmetry	Orthorhombic
Crystal size	300 x 50 µm
Method	Vapour diffusion, sitting drop (200 nl), protein-to-well ratio 1:1
Temp. [K]	277
Precip. cond.	PACT Premier, D7-well2 18 mg/ml His_6-$AspH_{315-758}$, 1 mM $MnCl_2$, 2 mM NOG, 100 mM Tris, pH 8.0, 200 mM sodium chloride, 20 % w/v PEG 6000, 726 µM NC-4Ser_{39mer}
AspH:Substrate ratio	1:2.2 (330 µM:726 µM)
Cryoprotectant	25 % glycerol
Diffraction Source	Synchrotron
Site	Diamond Light Source
Type	Diamond Beamline I04
Wavelength List [Å]	0.9795
Detector	Pilatus 6M-F

5.3 DISCUSSION

The structural, binding and activity studies in this Chapter corroborate the crucial role of the TPR domain for substrate recognition and the importance of the *C*-terminal region of EGFDs for protein-protein interactions.

TPR domains are thought to mediate protein-protein interactions in many biological systems.[519-521] This structural motif is found in many proteins,[522] including some other 2OG oxygenases such as the C-P4Hs and C-P3Hs, UTX and UTY of the KDM6 subfamily. A similar role in substrate recognition and protein-protein interactions to that in AspH has also been reported for the TPR domains in the C-P4Hs[523] and in several protein-chaperone complexes.[524-526]

The key interactions revealed in the crystallographic studies of AspH-substrate complexes deliver a plausible explanation for the observed AspH substrate requirements and preferences. The results from the binding studies in solution using DSF or in the gas phase using non-denaturing ESI-MS are consistent with the hydroxylation activity assays data (**Chapter 3** and **Chapter 4**). The results demonstrate that the oxygenase domain alone ($AspH_{562-758}$) is unable to bind the peptide substrates, whereas the longer construct including the TPR domain ($AspH_{315-758}$) can form enzyme-substrate complexes. The canonical EGFD peptides showed no sign of binding to $AspH_{315-758}$ in the DSF or the non-denaturing ESI-MS studies. In contrast, the non-canonical peptides bound to $AspH_{315-758}$.

5.3.1 Disulfide Plasticity and Cooperativity

The $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}CP2_{19mer}$ complex crystal structure validates the results obtained from the *in vitro* hydroxylation assays and supports the proposal that an EGFD with a non-canonical Cys3-4 disulfide is the natural AspH substrate.

The *N*-termini of the peptide substrates were substantially disordered in all the obtained crystal structures, indicating a higher degree of flexibility and that there are probably no further stable contacts with AspH in this region. The results from binding and complex formation studies consistently showed that the $NC\text{-}4Ser_{39mer}$ peptide is the strongest binder of the tested peptides. The *C*-terminal peptide tail, including the highly conserved hydrophobic residue corresponding to Phe31 in the hFX $EGF1_{39mer}$, is important for tight binding. The results explain why the crystallisation trials with the shorter $CP2_{10mer}$ peptide were unsuccessful and led to the design and characterisation of the longer $CP2_{19mer}$ peptide, enabling crystallisation of the complex. The contribution of the residues in the *C*-terminal part may play an important role in the first recognition contact in protein-protein interactions of AspH with its natural substrates and could help reach the required assembly for complex formation.

The aromatic residue in the position of Phe31 in hFX $EGF1_{39mer}$ is highly conserved in EGFDs. The residue in this position is also known to be important for EGFD inter-domain packing interactions.[527] As EGFDs usually occur in repetitive sequences, there is likely a degree of cooperativity between the domains, determining fold and overall geometry of the repeats (**Fig. 5-22**). The expected higher flexibility of a non-canonical EGFD with Cys1-2, 3-4, 5-6 disulfide pattern may therefore also have effects on the folds of adjacent EGFDs.

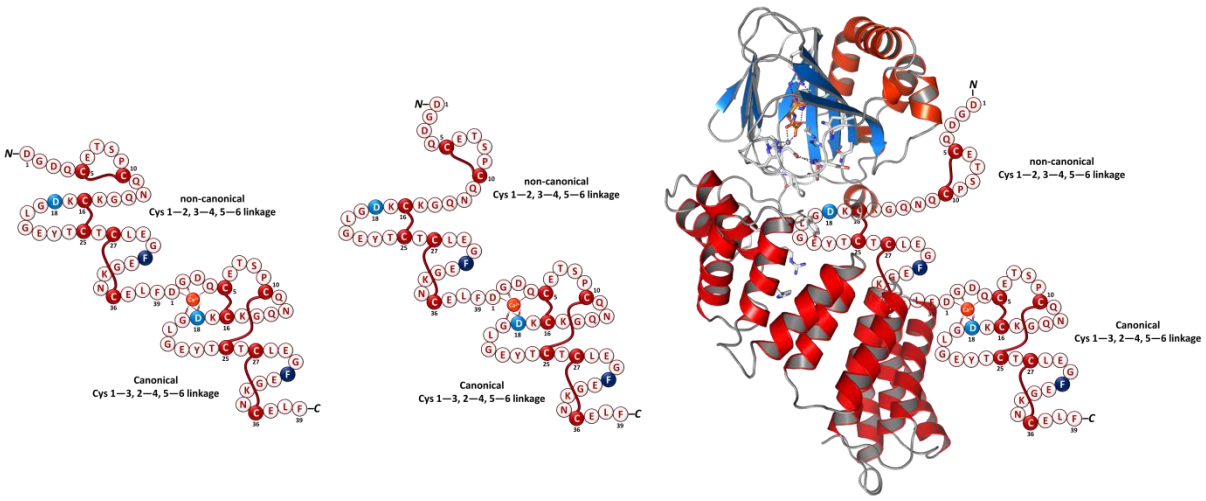


Fig. 5-22 | A non-canonical disulfide pattern with Cys1-2, 3-4, 5-6 disulfide linkage may result in higher flexibility and have effects on the folds of adjacent domains. The peptide core ring (residues 16-25) in the non-canonical EGFD is essential for the interactions with AspH; the *C*-terminal region is important for tight binding to AspH. The *N*-terminal regions of the EGFD peptides were disordered in all obtained AspH-substrate complex structures, indicating a higher degree of flexibility of these residues.

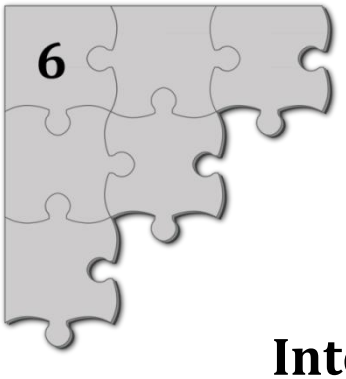
5.3.2 AspH Is a Highly Specific but Tolerant Enzyme

The crystallographic studies suggest that substrate recognition by AspH involving both the oxygenase and TPR domains is very specific. However, AspH can thereby tolerate variations in the amino acids between the key residues by forming many interactions with the substrate peptide backbone, which allows for sequence flexibility within the EGFD substrates. The number of interactions with the EGFD backbone partially explains the mutagenesis results. Due to the number of protein-protein contacts within both the active site and the TPR domain pockets, the loss of a single interaction can be tolerated if one residue is mutated to Ala at a time. However, some of the key interactions such as the salt bridge interaction between Arg735 and the cosubstrate 2OG appear to be essential for activity.

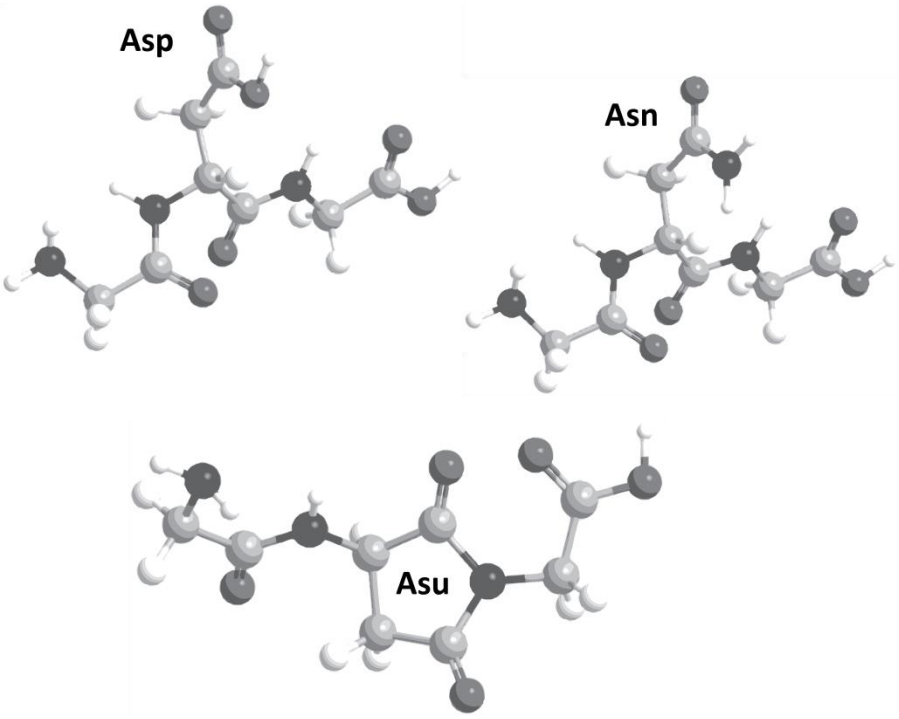
In the case of FIH, the two active site His ligands have been reported to be sufficient for Fe(II)-binding and catalysis; i.e. the catalytic triad carboxylate is dispensable for FIH activity.[428] The mutagenesis studies reported in this Chapter provide evidence that a single AspH active site ligand (His) is sufficient for Fe(II)- and substrate binding and catalysis. This phenomenon has not been observed for any other 2OG oxygenase thus far, and may be unique to AspH. The specific binding mode and strong interaction with the EGFD substrate may thereby help metal- and 2OG-binding. The non-denaturing ESI-MS results indicate that the metal and the cosubstrate bind simultaneously to AspH and that the complex is relatively labile. However, further extensive studies will be required to elucidate the mechanism.

5.5 ACKNOWLEDGEMENTS FOR CHAPTER 5

Dr Michael A. McDonough harvested crystals, collected data, solved and refined the $AspH_{315-758}$–substrate complex structures and is kindly acknowledged for these contributions. Hanna Tarhonskaya is kindly acknowledged for training in non-denaturing ESI-MS and for fruitful discussions. Dr Rasheduzzaman A. Chowdhury crystallised the $AspH_{315-758}$-H679A variant with the NC-$4Ser_{39mer}$ peptide, collected the data, solved and refined the structure presented here. Most of the MALDI-ToF-MS measurements were performed using the *Bruker Daltonics* Ultraflex I spectrometer based at the Oxion Proteomics Facility, Department of Physiology, Anatomy and Genetics, University of Oxford, and Dr Holger Kramer is kindly acknowledged for granting access to the instrument and for fruitful discussions. Many thanks also go to the researchers and staff at Diamond Light Source, Didcot, for their kind technical assistance.



Interference from Aspartyl Succinimides in MS-based Peptide Assays



Abstract

This Chapter provides a characterisation of the aspartyl succinimide peptide species interfering in MS-based hydroxylation activity assays.

6.1 Interference in MS Analyses of Hydroxylation Assays

In the previous Chapters, the presented MS analyses of hydroxylation activity assays frequently showed peaks of two peptide species that became hydroxylated, the expected peptide species and a lower-mass species (shown for NC-4Ser$_{39mer}$ hydroxylation in **Fig. 6-1**). The lower-mass peptide species interfering in the MS analysis of the hydroxylation assays was further characterised as described in this Chapter.

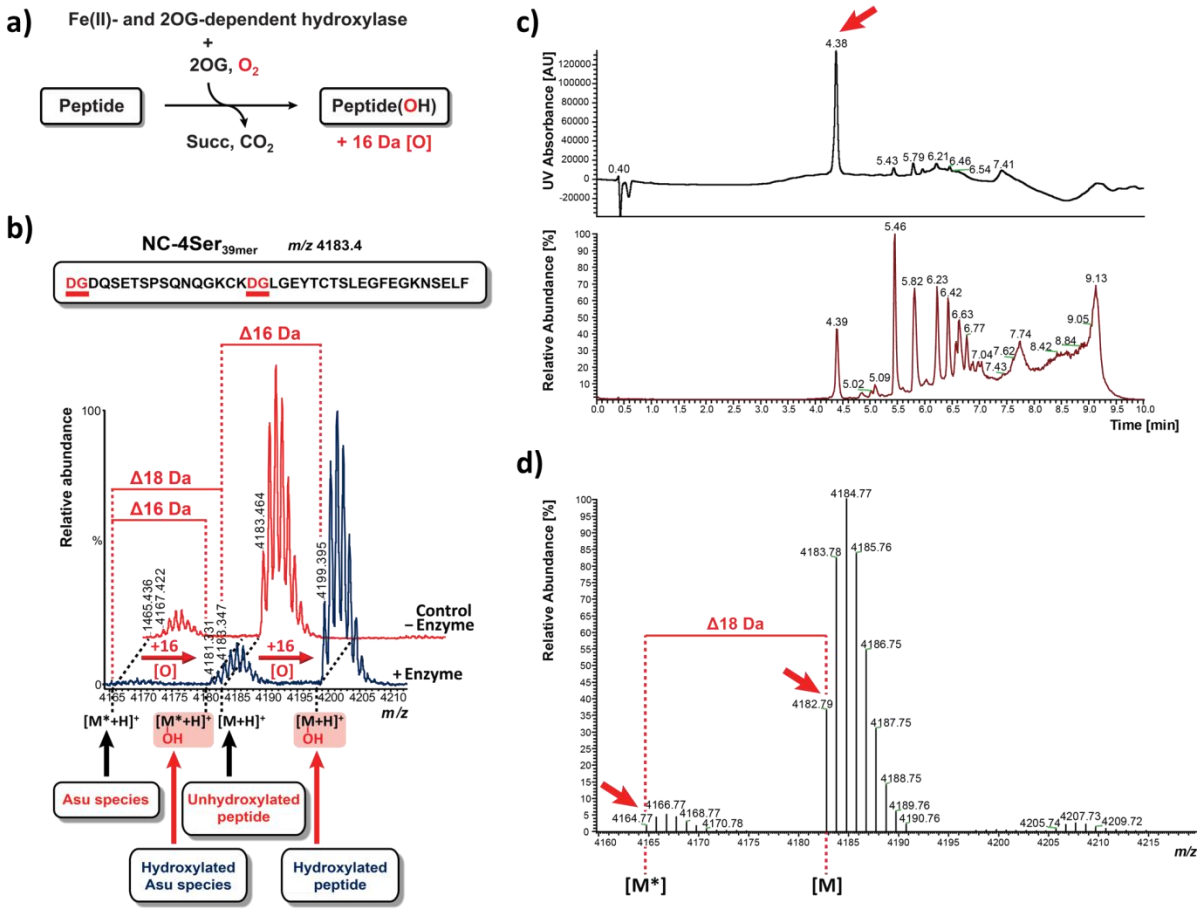


Fig. 6-1 | MS-based peptide hydroxylation assay. **a)** Scheme of a 2OG oxygenase-catalysed peptide hydroxylation assay. Hydroxylation occurs with a mass addition of 16 Da corresponding to the insertion of an oxygen atom during the reaction. **b)** Typical MALDI-ToF mass spectra (AspH-incubated and control samples) for determination of AspH-catalysed Asp hydroxylation in the synthetic NC-4Ser$_{39mer}$ peptide (unhydroxylated species $[M+H]^+$ m/z 4,183.3; hydroxylated species $[M(OH)+H]^+$ m/z 4,199.4). A further species, characterised as aspartyl succinimide (Asu) derivative, ($[M^*+H]^+$ m/z 4,165.4; corresponding to a mass loss of 18 Da from the calculated NC-4Ser$_{39mer}$ mass) also became hydroxylated ($[M^*(OH)+H]^+$ m/z 4,181.3). The isotope pattern of this hydroxylated peak overlaps with the initial unhydroxylated NC-4Ser$_{39mer}$ species peak ($[M+H]^+$ m/z 4,183.4), complicating the analysis. **c)** LC-MS analysis of the unhydroxylated NC-4Ser$_{39mer}$ in H_2O performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific), coupled to a Dionex™ UltiMate® 3000 HPLC. Represented as UV absorbance and MS traces. **d)** Deconvoluted mass spectrum corresponding to the peak eluted after 4.39 min. Asu: aspartyl succinimide.

In all cases, one peak in the mass spectra corresponded to the expected peptide species (calculated m/z for the unhydroxylated peptide $[M+H]^+$ or the hydroxylated peptide $[M(OH)+H]^+$

respectively), the other peak corresponded to a dehydrated peptide species, i.e. a mass loss of -18 Da from the calculated *m/z* (unhydroxylated $[M^*+H]^+$ or hydroxylated $[M^*(OH)+H]^+$). This dehydrated species has also been observed in hydroxylation assays performed using 2OG oxygenases other than AspH with their appropriate substrate peptides (e.g. PHD2 with the $CODD_{19mer}$ substrate; sequence DLDLEMLAPYIPMDDDFQL; D.Phil. Thesis of Dr Rasheduzzaman Chowdhury). The dehydrated peptide species was only observed in peptides containing at least one Asp residue. In addition, deamidated peptide species, i.e. a mass loss of 17 Da from the calculated *m/z* of the peptide, were found in peptides containing an Asn residue. The isotope pattern of the peak of the hydroxylated dehydrated species ($[M^*(OH)+H]^+$) overlapped with the initial unhydroxylated peptide species peak ($[M+H]^+$), complicating the MS analysis for accurate determination of hydroxylation levels (**Fig. 6-1**).

Frequently, the peaks of the dehydrated or deamidated species in MALDI-ToF mass spectra misleadingly appeared as a peak corresponding to a mass loss of 16 Da. However, LC-MS analysis of the peptides performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (*Thermo Scientific*) coupled to a Dionex™ UltiMate® 3000 HPLC (as described in **Chapter 10**), supported that the mass loss was -18 Da in Asp-containing or -17 Da in Asn-containing peptides respectively (**Fig. 6-1c** and **d**).

6.1.1 Aspartyl Succinimide Formation

Asp and Asn residues in peptides are known to undergo spontaneous degradation under loss of water and formation of an aspartyl succinimide species (aspartimide, Asu).[528] Asu species have previously been identified in the analysis of therapeutic proteins,[529-531] and have long been recognised as intermediates in both the deamidation reaction of Asn and isomerisation and racemisation of Asp residues (**Fig. 6-2**).[532] Asu formation has been reported to be favoured in Asp-Gly and Asn-Gly motifs,[528,533] as found in the hFX $EGF1_{39mer}$-based peptides used as AspH substrates in this Thesis (e.g. NC-4Ser_{39mer} sequence, **Fig. 6-1b**).

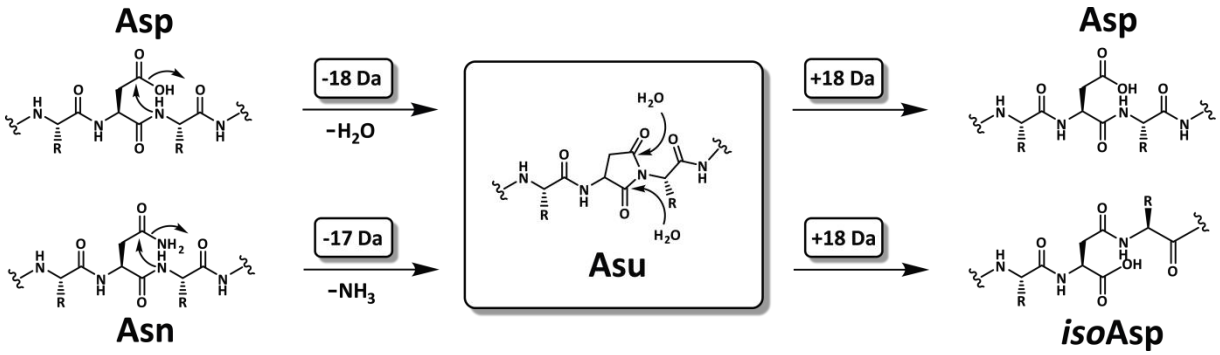


Fig. 6-2 | Mechanism of aspartyl succinimide (Asu) formation and hydrolysis from Asp and Asn residues. Hydrolysis of Asu residues leads to formation of mixtures of Asp and *iso*Asp products due to opening of the succinimide ring by attack of water at two alternative carbonyl groups.

6.2 Asu Is Formed in Solution and Not Due to In-Source Fragmentation

Peptides used in 2OG oxygenase-catalysed hydroxylation assays are frequently synthetically produced. Asu formation can occur during Fmoc-based solid-phase peptide synthesis.[534-537] However, following the development of sufficiently mild digestion conditions employing a high-fidelity digestion technique (incubation with trypsin for 1 h at 37 °C, **Fig. 6-3**),[538] Asu formation was also observed in a number of proteins of diverse origins during the work in this Thesis. Trypsin digestion of several protein standards, horse heart myoglobin, chicken egg lysozyme, and bovine cytochrome C, was performed by Dr Holger Kramer, and the corresponding modified peptides were characterised by MS/MS fragmentation using a high capacity ion trap tandem mass spectrometer. The extracted ion chromatograms indicated that the Asu modifications were present in solution, rather than results of mere in-source fragmentation. In contrast, standard digestion conditions in proteomics workflows involving thiol reducing agents and elevated pH conditions lead to facile succinimide hydrolysis. Therefore, most proteomic studies are not suitable for the observation of Asu-containing peptides.

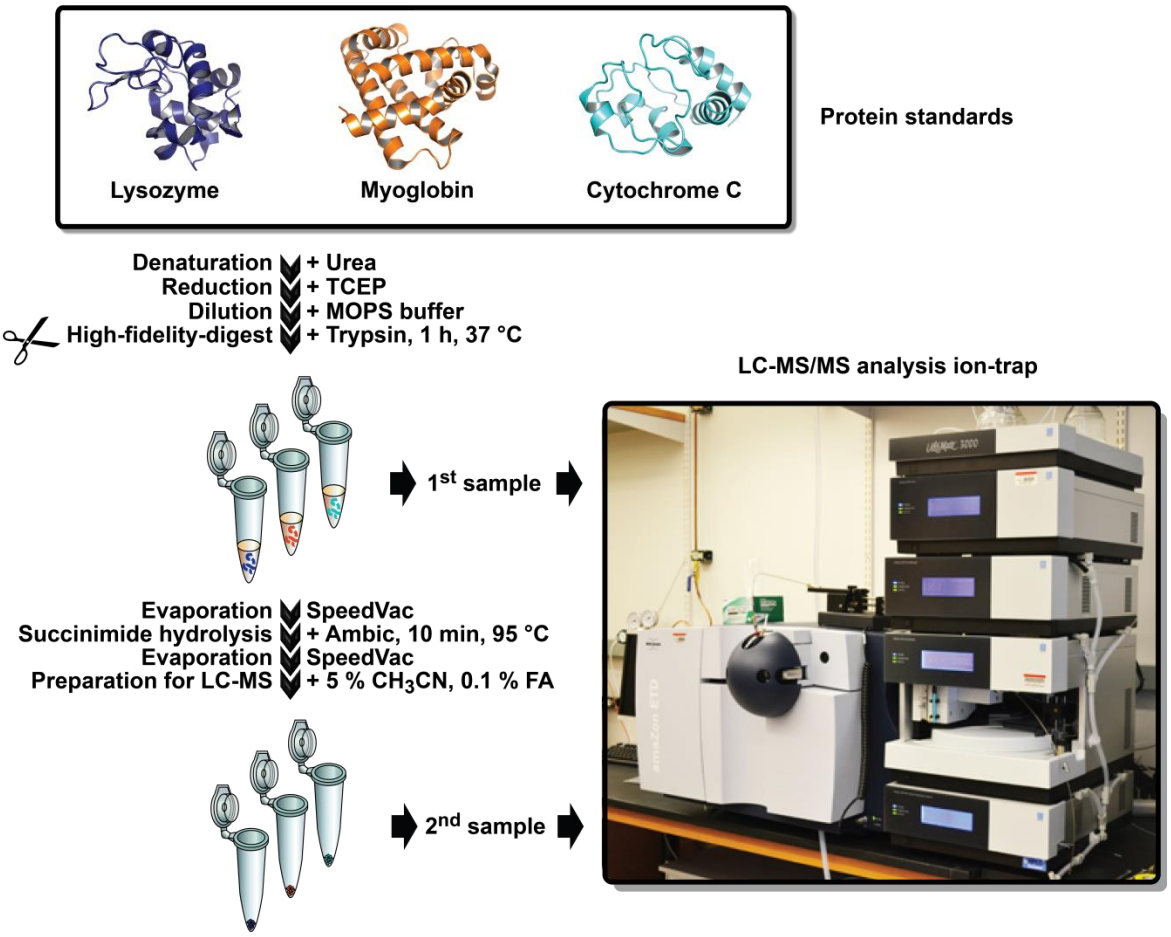


Fig. 6-3 | Workflow for the rapid and mild digestion of proteins following denaturation and reduction with urea and TCEP based on a published protocol.[538] Hydrolysis was triggered by heating to 95 °C in ammonium bicarbonate (Ambic).

6.2.1 Analysis of Asu in Protein Standards

LC-MS/MS fragmentation analyses of the protein standard peptides demonstrated loss of water (-18 Da) from Asp or Asn (Asx) residues. As described below, an abrogation of fragmentation at the Asu modification due to the formation of a five-membered ring was observed in the analysed peptides of the digested protein standards (**Fig. 6-4**).

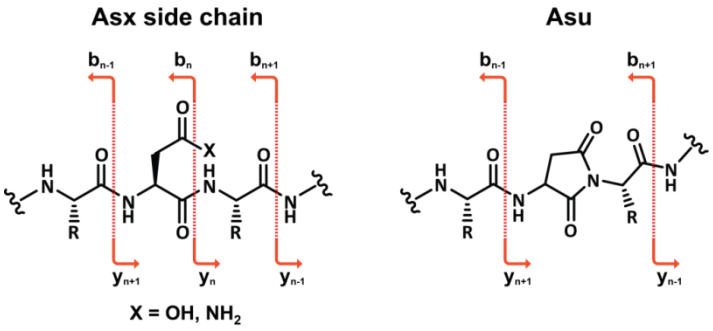


Fig. 6-4 | MS/MS spectra demonstrate a loss of water (-18 Da) or ammonia (-17 Da) from Asp or Asn (Asx) residues and abrogation of fragmentation at the Asu modification due to the formation of a five-membered ring.

A first sample was analysed using LC-MS/MS directly following the high-fidelity trypsin digestion of the protein standards. Asu was identified in several dehydrated peptides resulting from the digestion. The horse heart myoglobin$_{2\text{-}17}$, chicken egg lysozyme$_{64\text{-}79}$, and chicken egg lysozyme$_{116\text{-}130}$ peptides all contained an Asp-Gly motif. In a second sample analysis using LC-MS/MS, following heating of the same samples for 10 min at 95 °C in ammonium bicarbonate, no dehydrated species of these peptides were observed. These results indicated that the proposed Asu species was easily hydrolysed under these conditions.

Comparison of the MS/MS spectra of the dehydrated and corresponding unmodified peptide species showed differences in the fragmentation patterns. No fragments at the expected Asu modification were found in the dehydrated peptides, whereas the fragments at this position were observed in the unmodified peptides (**Fig. 6-5a** to **Fig. 6-7a**). These results were consistent with the proposed abrogation of fragmentation at the Asu five-membered ring.

Comparison of the extracted ion chromatograms showed that the peptides with Asu modifications eluted earlier from the C_{18} PepMap nano-HPLC column (*Dionex*™) than the corresponding unmodified peptides, indicating that succinimide-containing peptides are more polar or less hydrophobic than the corresponding unmodified peptides. Importantly, the earlier retention times also suggested that the dehydration occurred in solution and before the samples reached the electrospray source (**Fig. 6-5b** to **Fig. 6-7b**). This strongly suggested that this observation was not a mere in-source artefact.

One analysed bovine cytochrome $C_{93\text{-}100}$ peptide (**Fig. 6-8**) contained an *N*-terminal Glu residue. The dehydrated peptide species was therefore likely the result of pyroglutamate formation from the *N*-terminal Glu residue rather than from Asu formation. Comparison of the retention times showed that both the dehydrated and the unmodified peptide species eluted at the same time in this case, suggesting that Asu formation may be discriminated from *N*-terminal pyroglutamate.

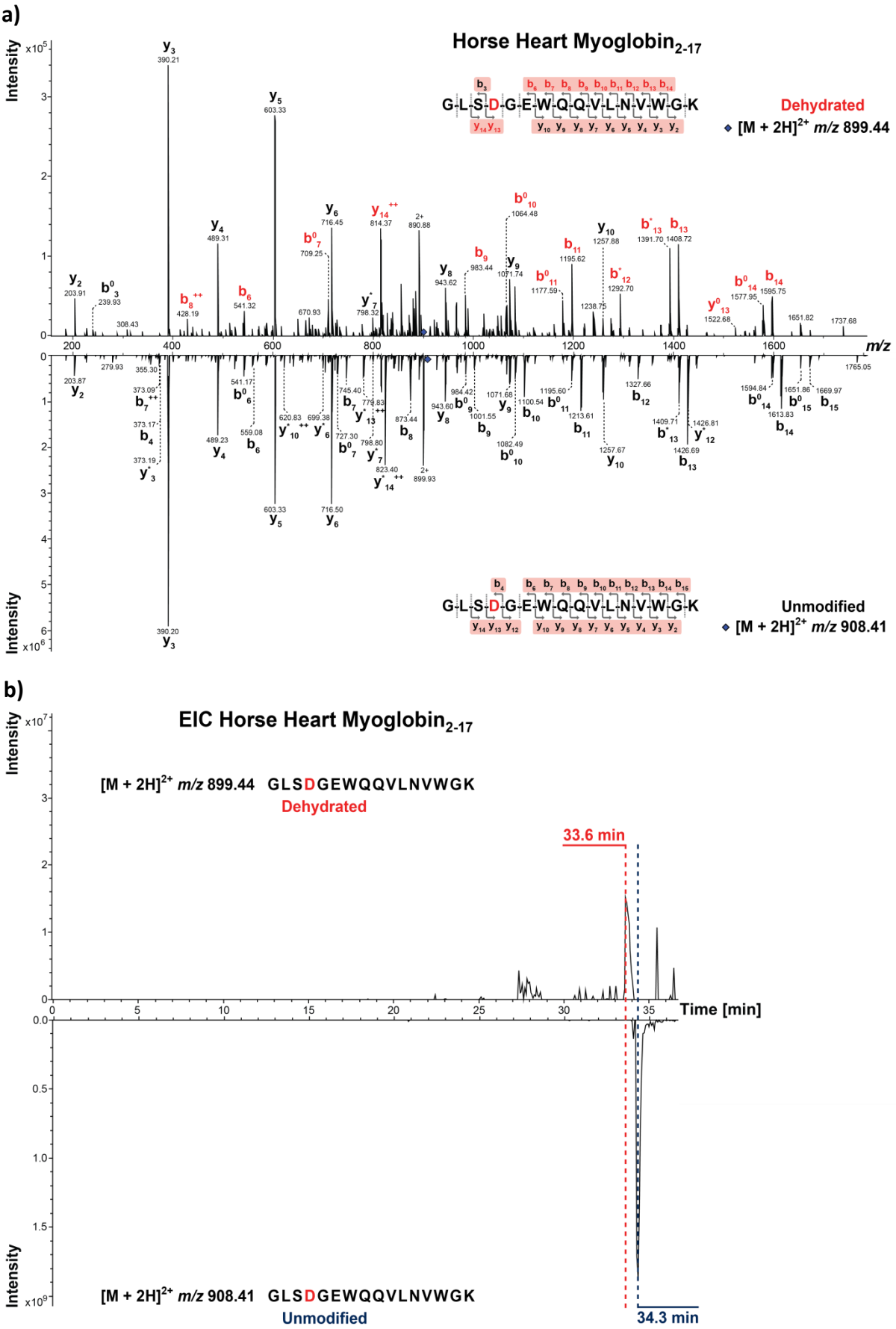


Fig. 6-5 | Identification of Asu formation in the horse heart myoglobin$_{2-17}$ peptide from trypsin digestion. **a)** MS/MS spectra showing loss of water (-18 Da) from residue Asp5; fragment ions exhibiting loss of water due to Asu formation highlighted in red. **b)** Extracted ion chromatograms (EICs) showing a comparison of the retention times for dehydrated and unmodified peptides.

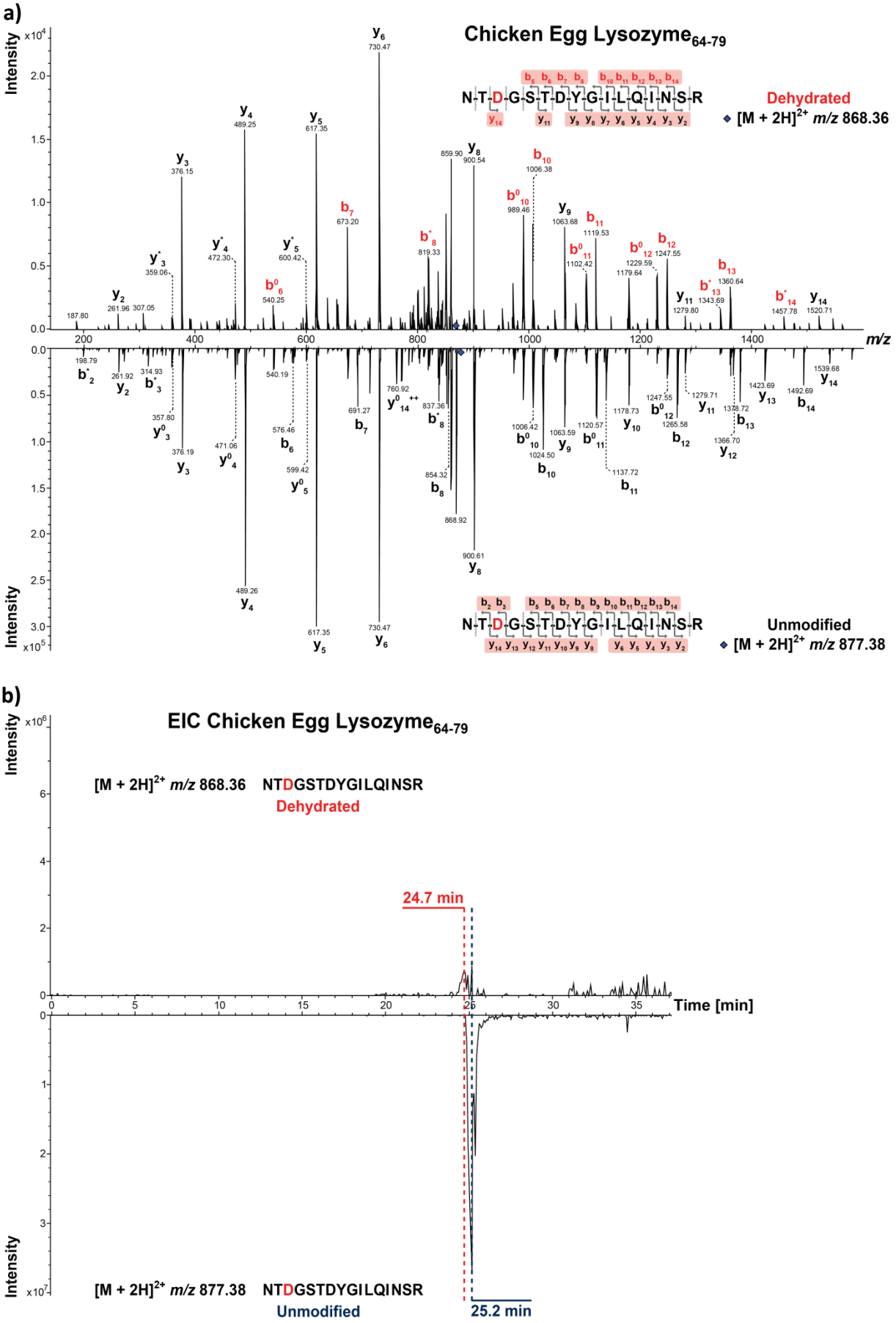


Fig. 6-6 | Identification of Asu formation in the chicken egg lysozyme$_{64\text{-}79}$ peptide from trypsin digestion. **a)** MS/MS spectra showing loss of water (-18 Da) from residue Asp66; fragment ions exhibiting loss of water due to Asu formation highlighted in red. **b)** Extracted ion chromatograms (EICs) showing a comparison of the retention times for dehydrated and unmodified peptides.

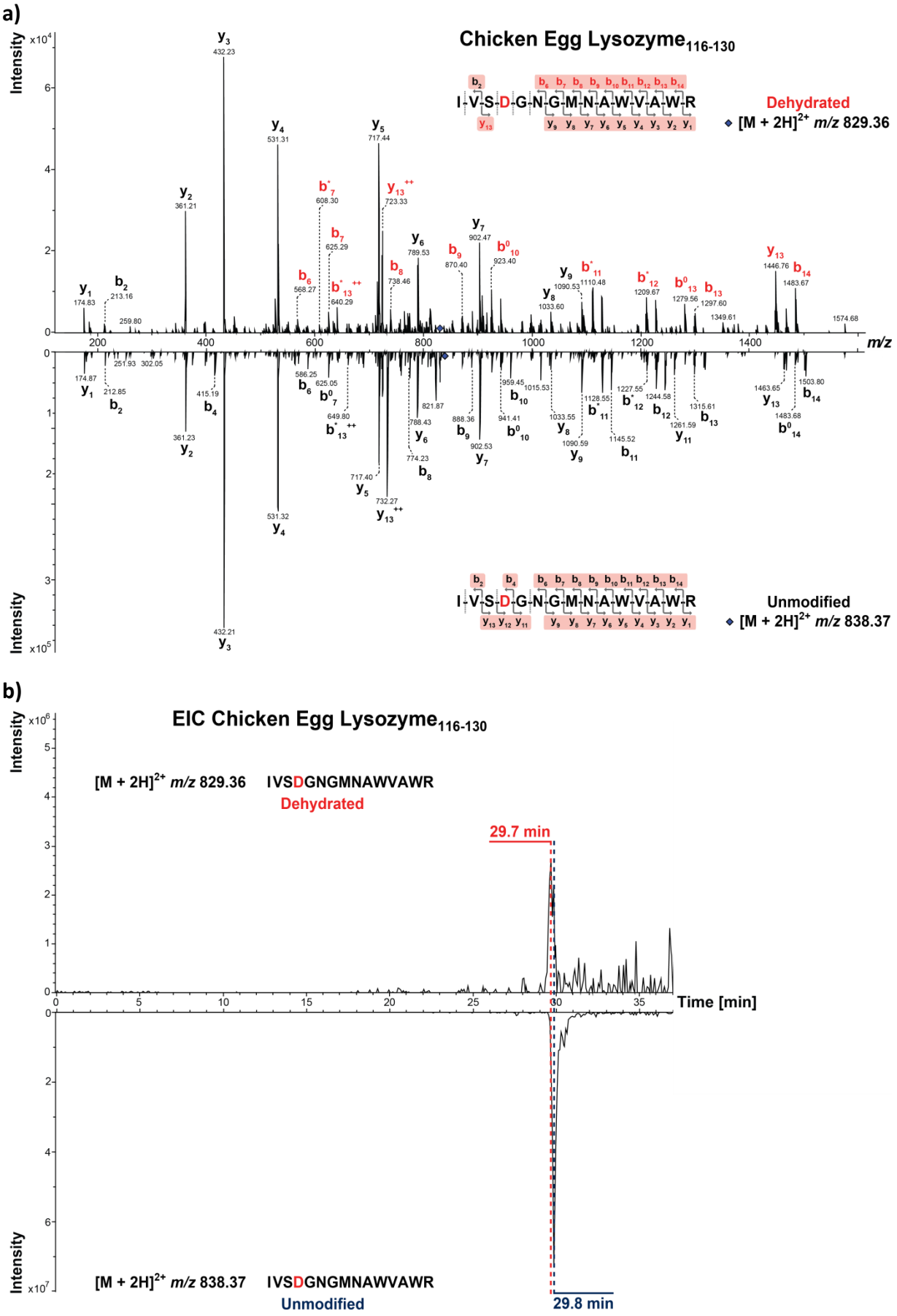


Fig. 6-7 | Identification of Asu formation in the chicken egg lysozyme$_{116-130}$ peptide from trypsin digestion. **a)** MS/MS spectra showing loss of water (-18 Da) from residue Asp119; fragment ions exhibiting loss of water due to Asu formation highlighted in red. **b)** Extracted ion chromatograms (EICs) showing a comparison of the retention times for the dehydrated and unmodified peptides.

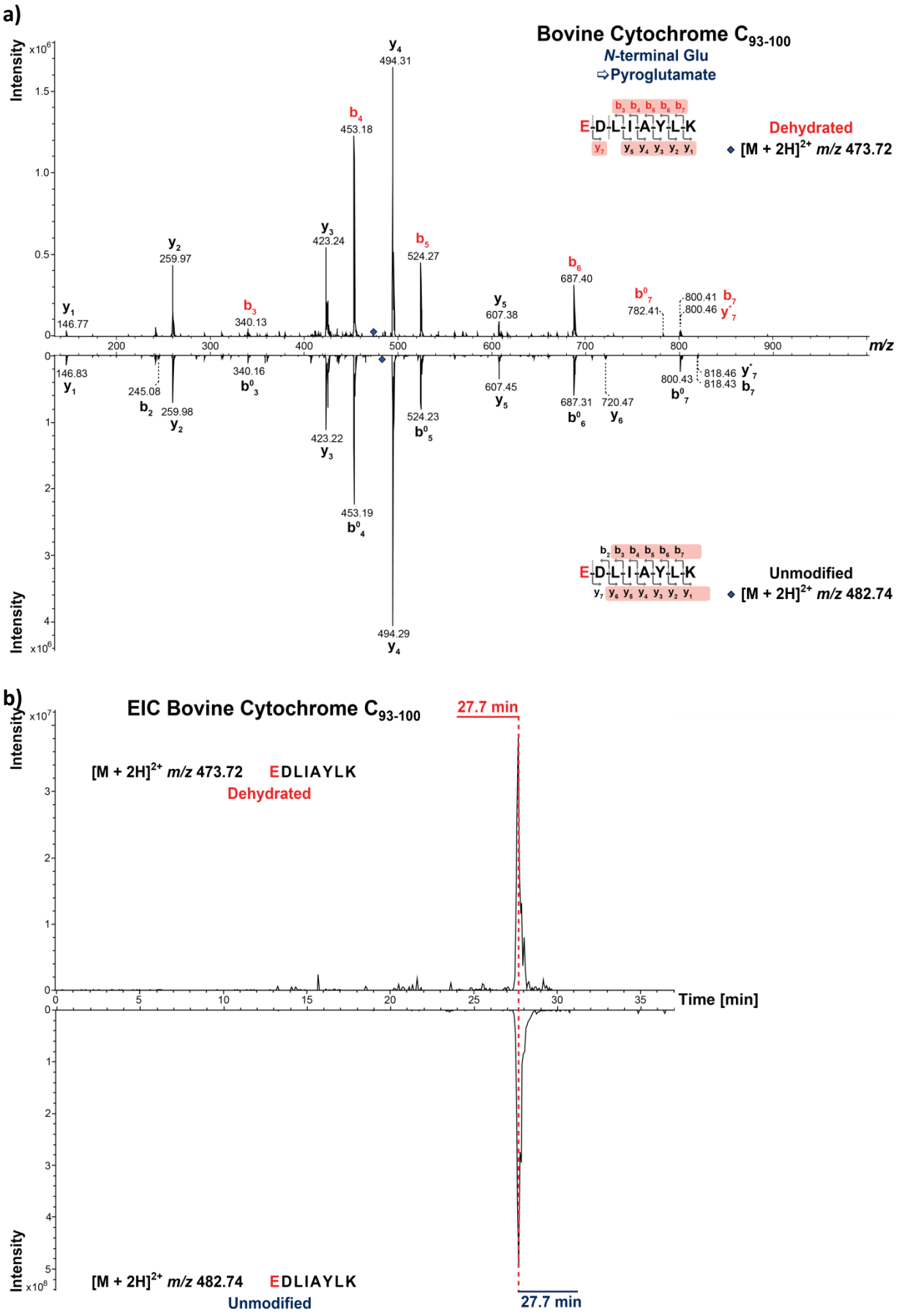


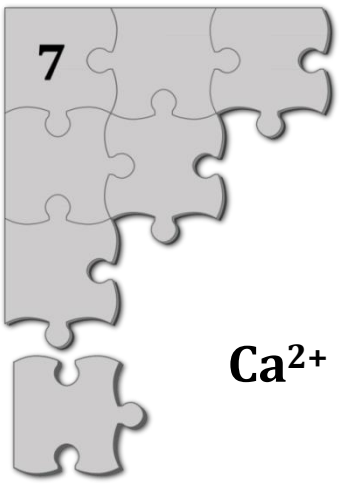
Fig. 6-8 | The dehydrated peptide species in the bovine cytochrome $C_{93\text{-}100}$ peptide from trypsin digestion was likely the result of pyroglutamate formation from the *N*-terminal Glu residue. **a)** MS/MS spectra showing loss of water (-18 Da) from residue Glu93; fragment ions exhibiting loss of water due to pyroglutamate formation highlighted in red. **b)** Extracted ion chromatograms (EICs) showing a comparison of the retention times for the dehydrated and unmodified peptides.

6.3 DISCUSSION

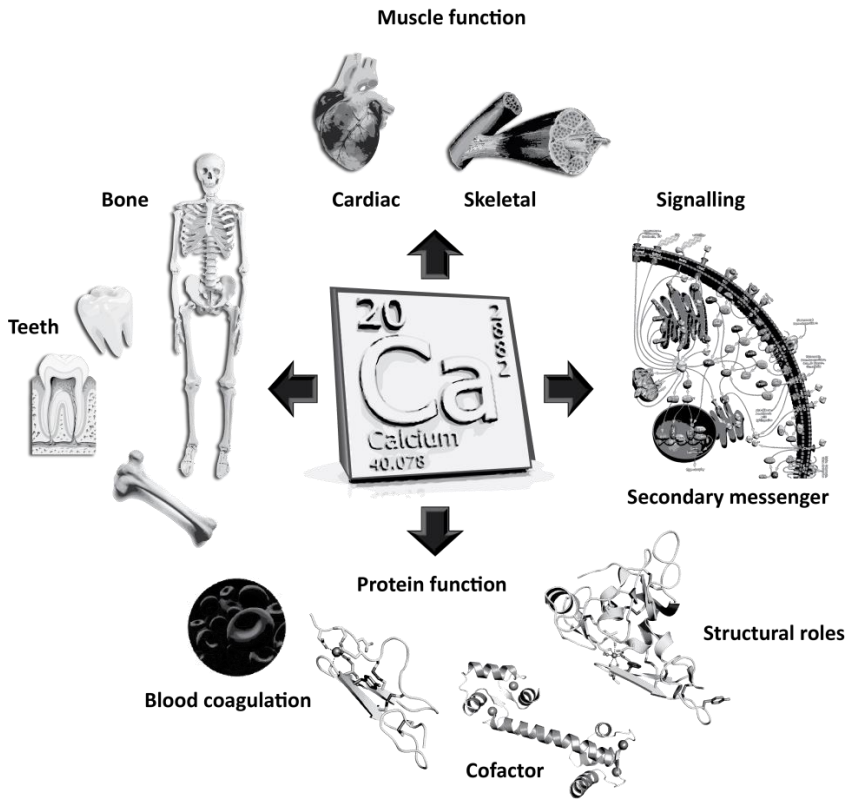
The results in this Chapter suggest that the interfering peaks frequently observed in 2OG oxygenase-catalysed hydroxylation assays likely result from Asu formation. The dehydration or deamidation of Asp or Asn residues respectively to form Asu is favoured under acidic conditions, as they are often employed to quench the assayed reaction (lowering of the pH in the sample by addition of acids such as formic or trifluoroacetic acid to stop the enzymatic activity). Asu in synthetic peptides used for 2OG oxygenase assays may be a result from formation during the peptide synthesis or sample preparation. The results from the protein standards analyses suggest that an additional high pH step for succinimide hydrolysis would be advantageous for the determination of hydroxylation levels and could simplify the analysis of modification sites by MS/MS experiments.

6.4 ACKNOWLEDGEMENTS FOR CHAPTER 6

Dr Holger Kramer is kindly acknowledged for preparation of the protein standards samples and data acquisition using the *Bruker Daltonics* amaZon™ Ion Trap LC-MS/MS system located at the Oxion Proteomics Facility, Department of Physiology, Anatomy and Genetics, University of Oxford, and for fruitful discussions.



Ca^{2+} Promotes Canonical EGFD Disulfide Formation



Abstract

The work described in this Chapter is focused on the effects of different Ca^{2+} levels on EGFD hydroxylation, which were investigated *in vitro*.

7.1 THE PHYSIOLOGICAL ROLES OF CA^{2+}

Ca^{2+} ions play a vital role in physiology (**Fig. 7-1**) and are required for many biochemical processes, and Ca^{2+} is a secondary messenger in signal transduction pathways.[539-542] Many proteins depend on Ca^{2+} as a cofactor. Ca^{2+} allows proteins to interact with (phospholipid) membrane surfaces.[543] The physiological concentration of free Ca^{2+} is reported to be in the range of ca. 1.0 mM–1.2 mM in the extracellular matrix.[350]

Ca^{2+} is not only a vital mineral in bone and teeth but also facilitates the contraction of muscle cells.[539] Ca^{2+} ions are important signalling molecules and can induce activation of ion channels or act as secondary messengers.[539,540] Ca^{2+} is also essential for protein function and participates in many important biological processes such as blood coagulation and protein folding.[539,542]

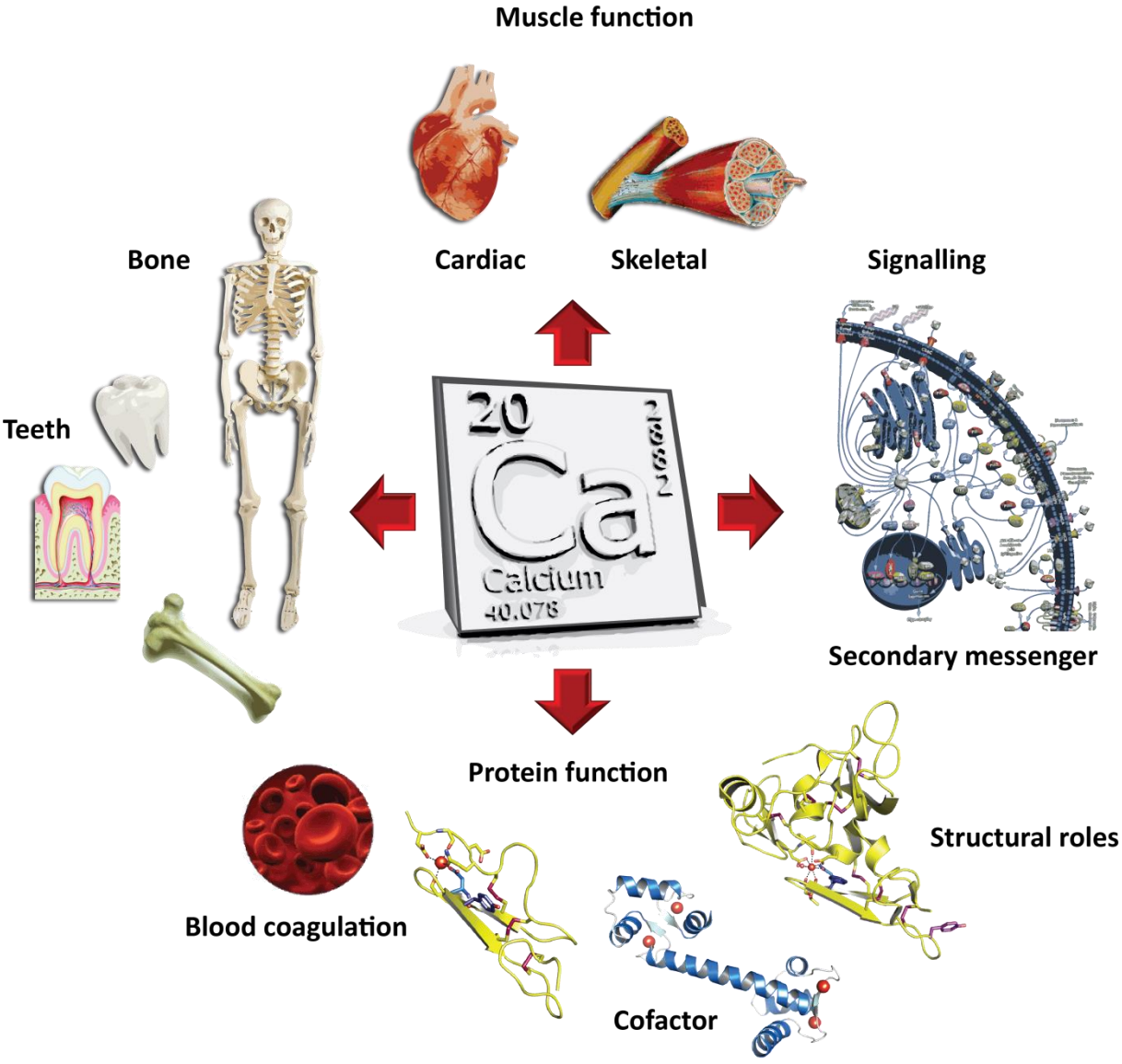


Fig. 7-1 | The physiological roles of calcium (Ca^{2+}). Ca^{2+} is not only a vital mineral in bone and teeth but also facilitates the contraction of muscle cells. Ca^{2+} ions are important signalling molecules and can induce activation of ion channels or act as secondary messengers. Ca^{2+} is also essential for protein function and participates in many important biological processes such as blood coagulation and protein folding.

7.1.1 The Hydroxylation Site Residue in Ca^{2+}-Binding EGFDs

The AspH crystal structures (**Chapters 2** and **5**) showed that the hydroxylation site residue, $Asp18_{39mer}$, is buried in the AspH active site, suggesting that Ca^{2+} may impair hydroxylation (**Fig: 7-2**). Concurrent Ca^{2+} coordination by the $Asp18_{39mer}$ side chain appeared highly unlikely.

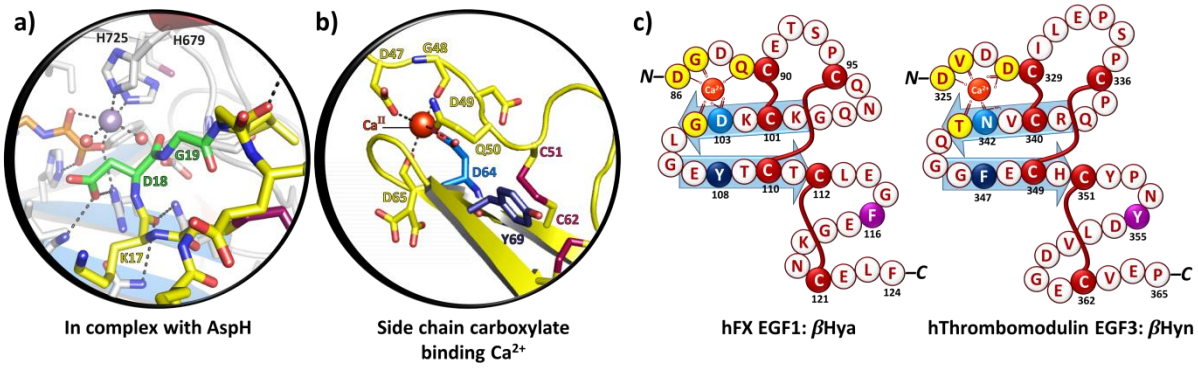


Fig. 7-2 | The EGFD hydroxylation site is buried in the AspH active site. Simultaneous Ca^{2+}-binding is highly unlikely. **a)** The $Asp18_{39mer}$ side chain is orientated towards the AspH active site in the $AspH_{315\text{-}758}{\cdot}Mn^{II}{\cdot}NOG{\cdot}hFX\ EGF1_{39mer}$ crystal structure. **b)** The $Asp64_{FIX\ EGF1}$ (residue corresponding to $Asp18_{39mer}$ in hFX EGF1) side chain coordinates the Ca^{2+} ion in the X-ray crystal structure of hFIX EGF1 in complex with Ca^{2+} (PDB: 1EDM[351]; side chain of Asp65 found in two conformations).

7.1.2 Ca^{2+}-Binding of EGFDs

EGFDs in the intact proteins typically have Ca^{2+} affinities in the micromolar to nanomolar range.[544,350] Isolated cbEGFDs bind Ca^{2+} relatively weakly with reported dissociation constants in the millimolar range,[545-546a,546] whereas the *C*-terminal cbEGF of a cbEGF-cbEGF pair typically has a higher Ca^{2+} affinity than the *N*-terminal cbEGFD.[547] Supposedly, the calcium-binding site is thereby stabilised by the presence of the preceding EGFD.[548] One of the cbEGFD proteins investigated in this Thesis, hFib1 TB4cbEGF23, binds Ca^{2+} with very high affinity (K_d of cbEGF23 = 16 nM).[350]

For molecules such as Fibrillin, variations in the Ca^{2+} affinity across the molecule are believed to be biologically significant, as Ca^{2+} rigidifies the canonical fold.[350] At physiological concentrations of free Ca^{2+} moderate- to high-affinity cbEGFs with K_d values < 100 µM would be saturated, and thus rigid regions, whereas low-affinity cbEGFs are proposed to be more flexible junction regions that allow for a degree of elasticity within the molecule.[350]

7.2 The Effects of Ca^{2+} on Hydroxylation

7.2.1 Effects of Ca^{2+} on Hydroxylation of cbEGFD Proteins

hFib1 TB4cbEGF23 was a poor AspH substrate when standard assay conditions were employed (~2 % hydroxylation), but it was an excellent substrate under RedOx conditions (~86 % hydroxylation). In

the presence of 3 mM Ca^{2+} (without preincubation time) the hydroxylation levels were strongly reduced (~9 %; **Fig. 7-3**) , whereas no effect was seen in the presence of 3 mM Mg^{2+} and a reduced effect (~24 %) in the presence of 1.0 mM Ca^{2+} (**Fig. 7-4**).

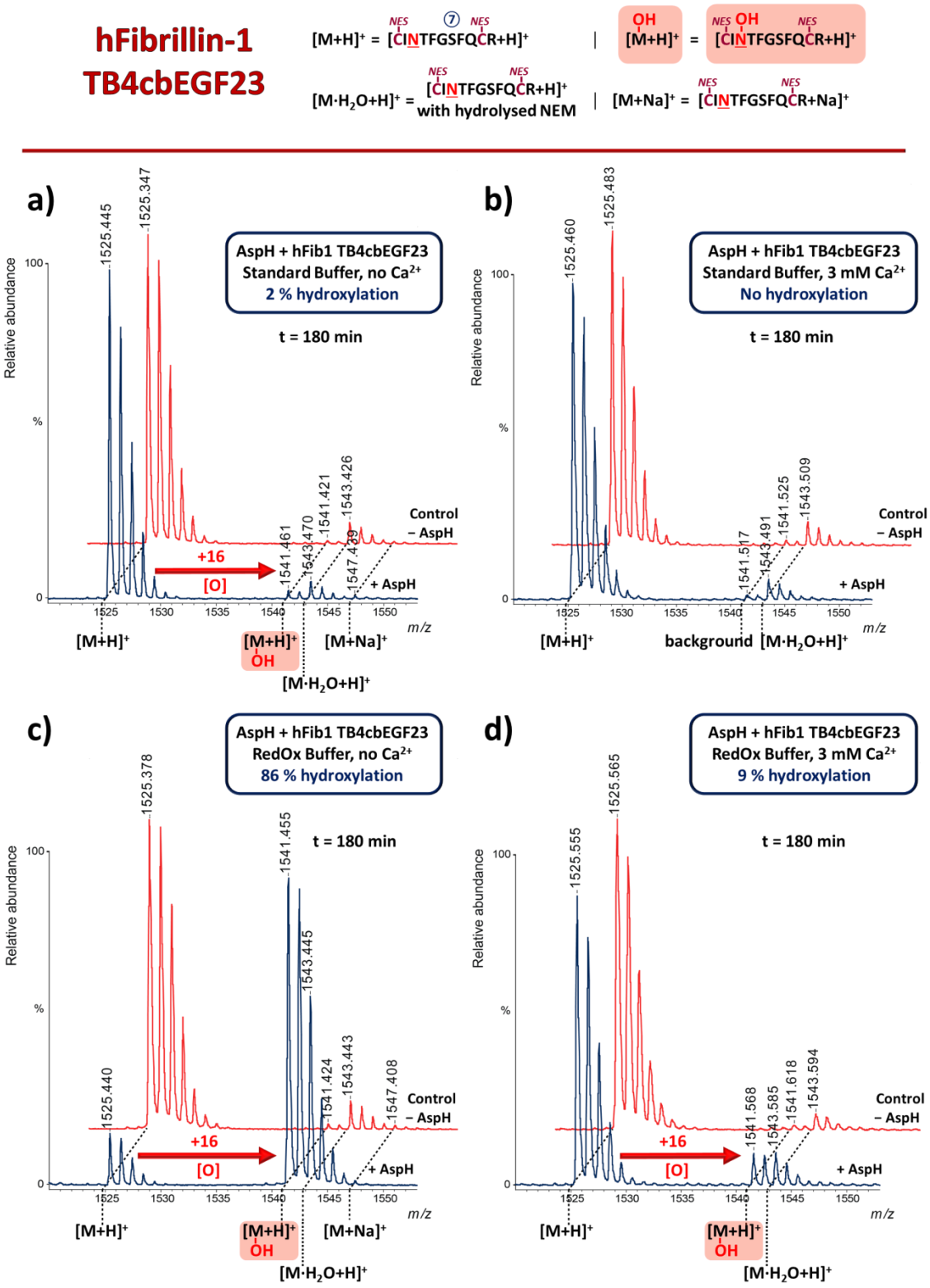


Fig. 7-3 | MALDI-ToF-MS results of hFib1 TB4cbEGF23 incubated with AspH (AspH to substrate ratio 1:10) and control samples in different buffer systems and presence or absence of Ca^{2+}. Close-ups on the shown peptide from trypsin digestion, which contains the hydroxylation site of cbEGF23 (unhydroxylated $[M+H]^+$ *m/z* 1,525.4; hydroxylated $[M(OH)+H]^+$ *m/z* 1,541.4); cysteine residues alkylated with NEM. **a)** Incubation using standard conditions without Ca^{2+} and **b)** in the presence of 3 mM Ca^{2+}. **c)** Incubation using RedOx conditions without Ca^{2+} and **d)** in the presence of 3 mM Ca^{2+}.

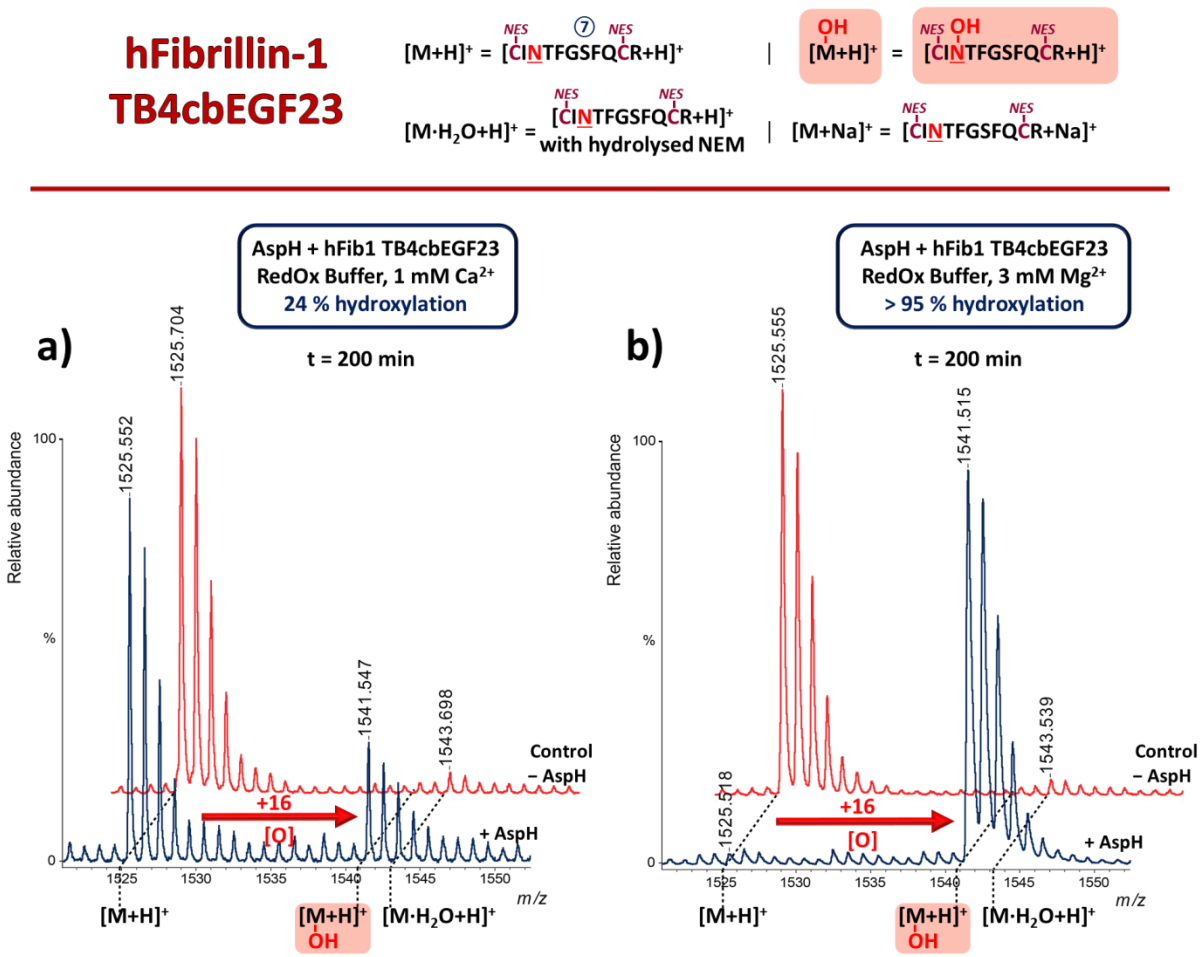


Fig. 7-4 | MALDI-ToF-MS results of hFib1 TB4cbEGF23 incubated with AspH (AspH to substrate ratio 1:10) and control samples in different buffer systems and presence or absence of Ca^{2+}. Close-up views of the shown peptide from trypsin digestion, which contains the hydroxylation site of cbEGF23 (unhydroxylated $[M+H]^+$ *m/z* 1,525.4; hydroxylated $[M(OH)+H]^+$ *m/z* 1,541.4); cysteine residues alkylated with NEM. **a)** Incubation using RedOx conditions in the presence of 1.0 mM Ca^{2+} and **b)** in the presence of 3 mM Mg^{2+}.

For the hFib1 cbEGF41-43 (**Fig. 7-5**), hFib1 cbEGF32-33 (**Fig. 7-6**), and hNotch1 cbEGF11-13 (**Fig. 7-7**) the observed effects of the presence of Ca^{2+} under RedOx conditions were similar, albeit not as distinctive as for hFib1 TB4cbEGF23. However, only the data for the hydroxlation sites of the *N*-terminal EGFDs in the peptides resulting from trypsin digestion of the cbEGFD proteins were investigated. Other proteolytic digests may allow access to hydroxylation level determination of the other cbEGFDs in the repeats in future studies.

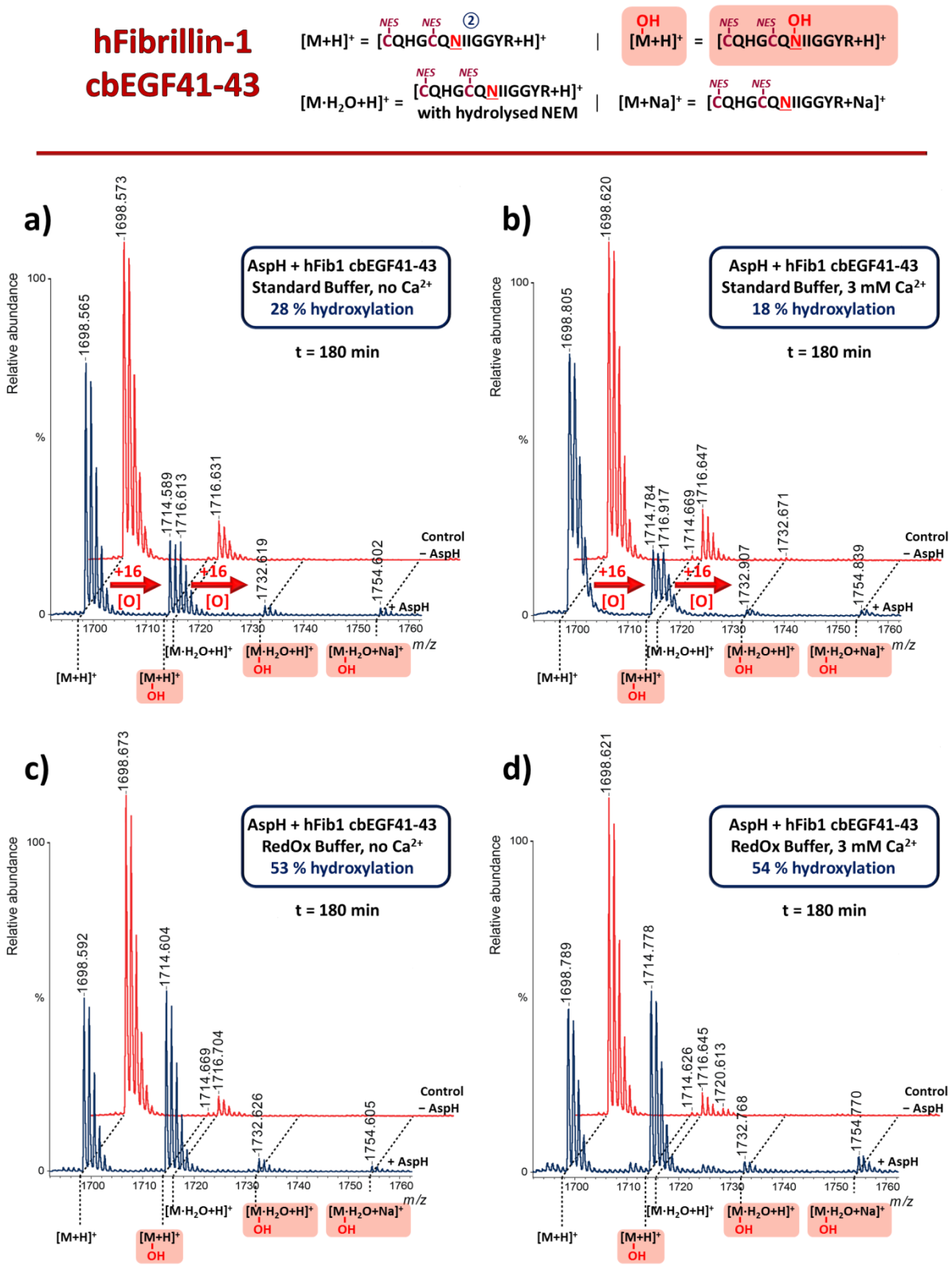


Fig. 7-5 | MALDI-ToF-MS results of hFib1 cbEGF41-43 incubated with AspH (AspH to substrate ratio 1:10) and control samples in different buffer systems and presence or absence of Ca^{2+}. Close-up views of the shown peptide from trypsin digestion, which contains the hydroxylation site of cbEGF41 (unhydroxylated $[M+H]^+$ *m/z* 1,698.7; hydroxylated $[M(OH)+H]^+$ *m/z* 1,714.7); cysteine residues alkylated with NEM. **a)** Incubation using standard conditions without Ca^{2+} and **b)** in the presence of 3 mM Ca^{2+}. **c)** Incubation using RedOx conditions without Ca^{2+} and **d)** in the presence of 3 mM Ca^{2+}.

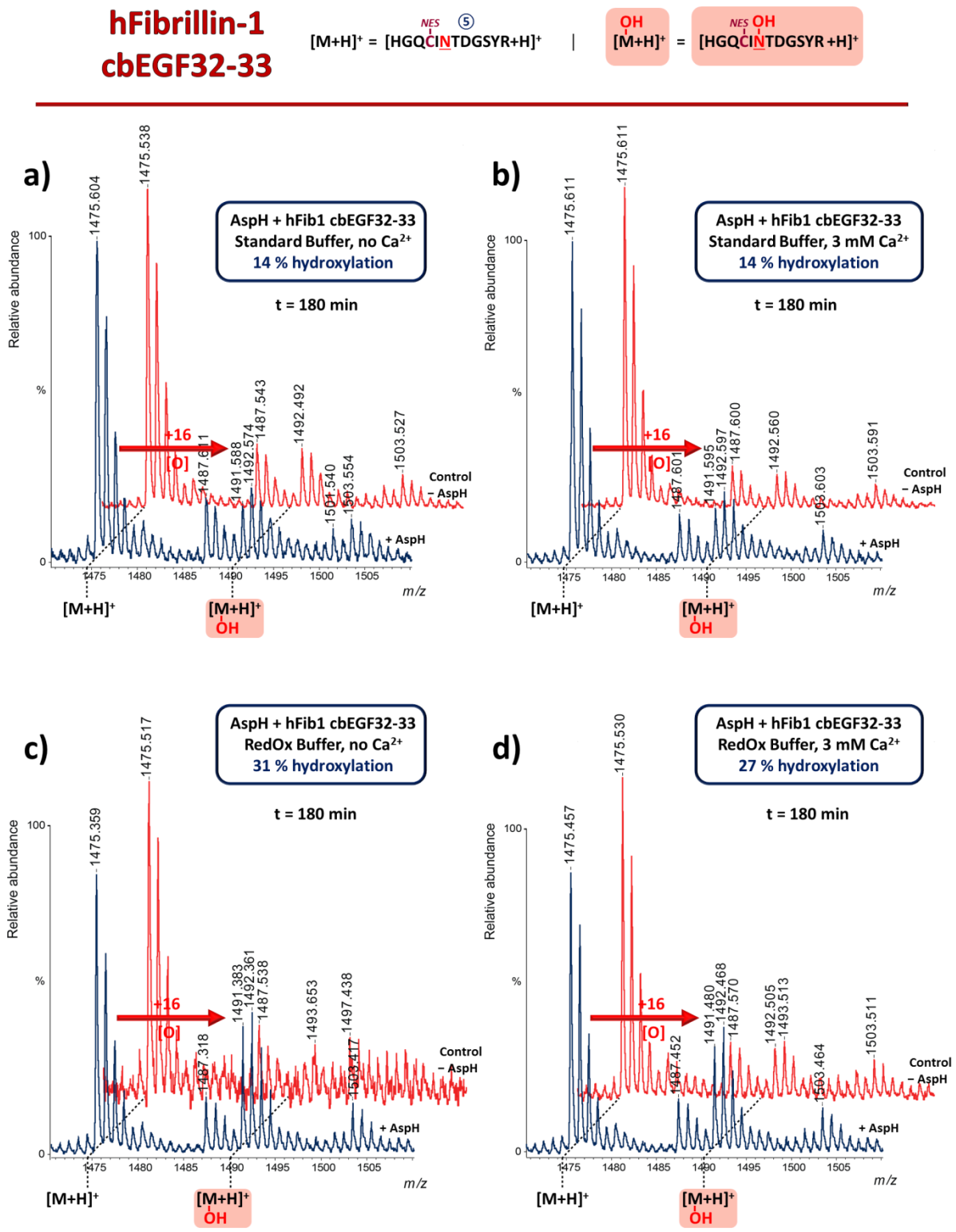


Fig. 7-6 | MALDI-ToF-MS results of hFib1 cbEGF32-33 incubated with AspH (AspH to substrate ratio 1:10) and control samples in different buffer systems and presence or absence of Ca^{2+}. Close-up views of the shown peptide from trypsin digestion, which contains the hydroxylation site of cbEGF32 (unhydroxylated $[M+H]^+$ *m/z* 1,475.5; hydroxylated $[M(OH)+H]^+$ *m/z* 1,491.5); cysteine residues alkylated with NEM. **a)** Incubation using standard conditions without Ca^{2+} and **b)** in the presence of 3 mM Ca^{2+}. **c)** Incubation using RedOx conditions without Ca^{2+} and **d)** in the presence of 3 mM Ca^{2+}.

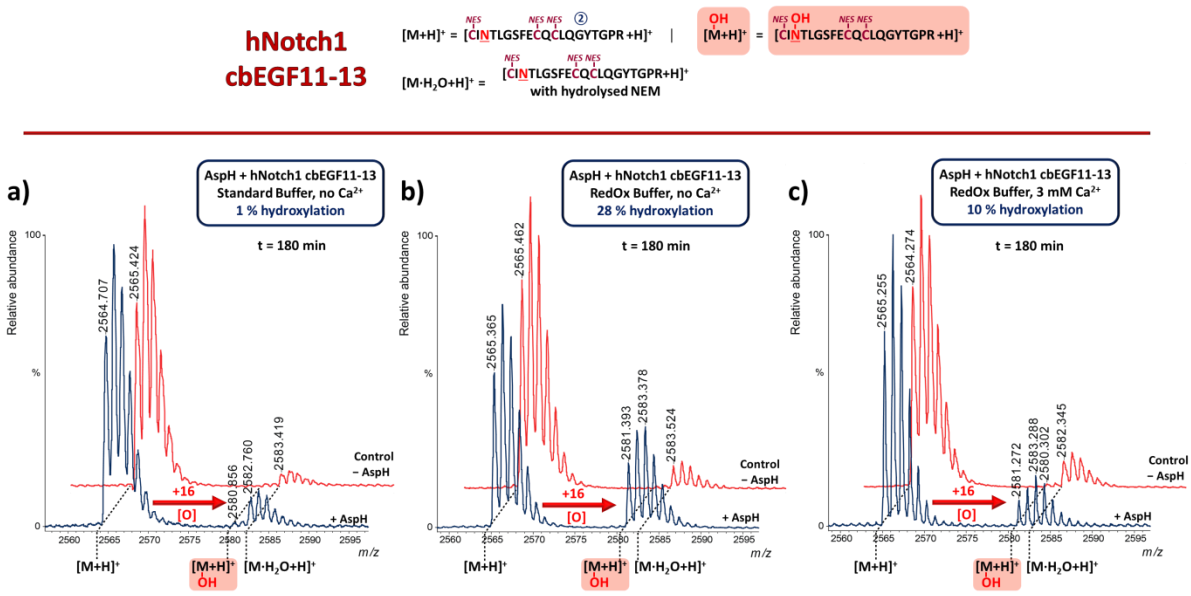


Fig. 7-7 | MALDI-ToF-MS results of hNotch1 cbEGF11-13 incubated with AspH (AspH to substrate ratio 1:10) and control samples in different buffer systems and presence or absence of Ca^{2+}. Close-up views of the shown peptide from trypsin digestion, which contains the hydroxylation site of cbEGF11 (unhydroxylated $[M+H]^+$ *m/z* 2,655.3; hydroxylated $[M(OH)+H]^+$ *m/z* 2681.3); cysteine residues alkylated with NEM. **a)** Incubation using standard conditions without Ca^{2+} and **b)** incubation using RedOx conditions without Ca^{2+} and **d)** in the presence of 3 mM Ca^{2+}.

7.2.2 Effects of Ca^{2+} on Hydroxylation of the Synthetic EGFD Peptides

In preliminary assays, the effects of Ca^{2+} ions on hydroxylation were investigated using the hFX $EGF1_{39mer}$ peptide under standard conditions. $CaCl_2$ was added to the assay to a final concentration of 500 µM, 5 mM and 50 mM. The results showed no effect of Ca^{2+} on hydroxylation under standard conditions (**Fig. 7-8**).

Hydroxylation assays were then performed in Standard Buffer (50 mM HEPES, pH 7.5, 150 mM NaCl), RedOx Buffer (50 mM Tris-HCl, pH 8.5, 3.0 mM reduced S-(-)-glutathione (GSH), 0.3 mM oxidised S-(-)-glutathione (GSSG), 150 mM NaCl), and Refolding Buffer (3 mM L-cysteine, 0.3 mM L-cystine, 10 mM $CaCl_2$, 50 mM Tris, pH 8.3). Refolding Buffer is an oxido-shuffling buffer system that enables disulfide exchange but is supplemented with Ca^{2+}, as it is used for in vitro refolding of recombinant EGFDs.[350]

Diminished hydroxylation of the hFX $EGF1_{39mer}$ was observed when the assay was performed in Refolding Buffer. Likewise, diminished hydroxylation levels of the C_{26mer} and the NC_{26mer} were observed in Refolding Buffer (**Fig. 7-9a** and **b**). No effect of added Ca^{2+} was observed in the hydroxylation assays using peptides that cannot undergo disulfide shuffling to the canonical isoforms, i.e. the NC-$4Ser_{39mer}$ (**Fig. 7-9c**), and the assayed cyclic peptides, $CP1_{10mer}$, $LP1_{10mer}$, and $FP1_{10mer}$ (**Fig. 7-10**).

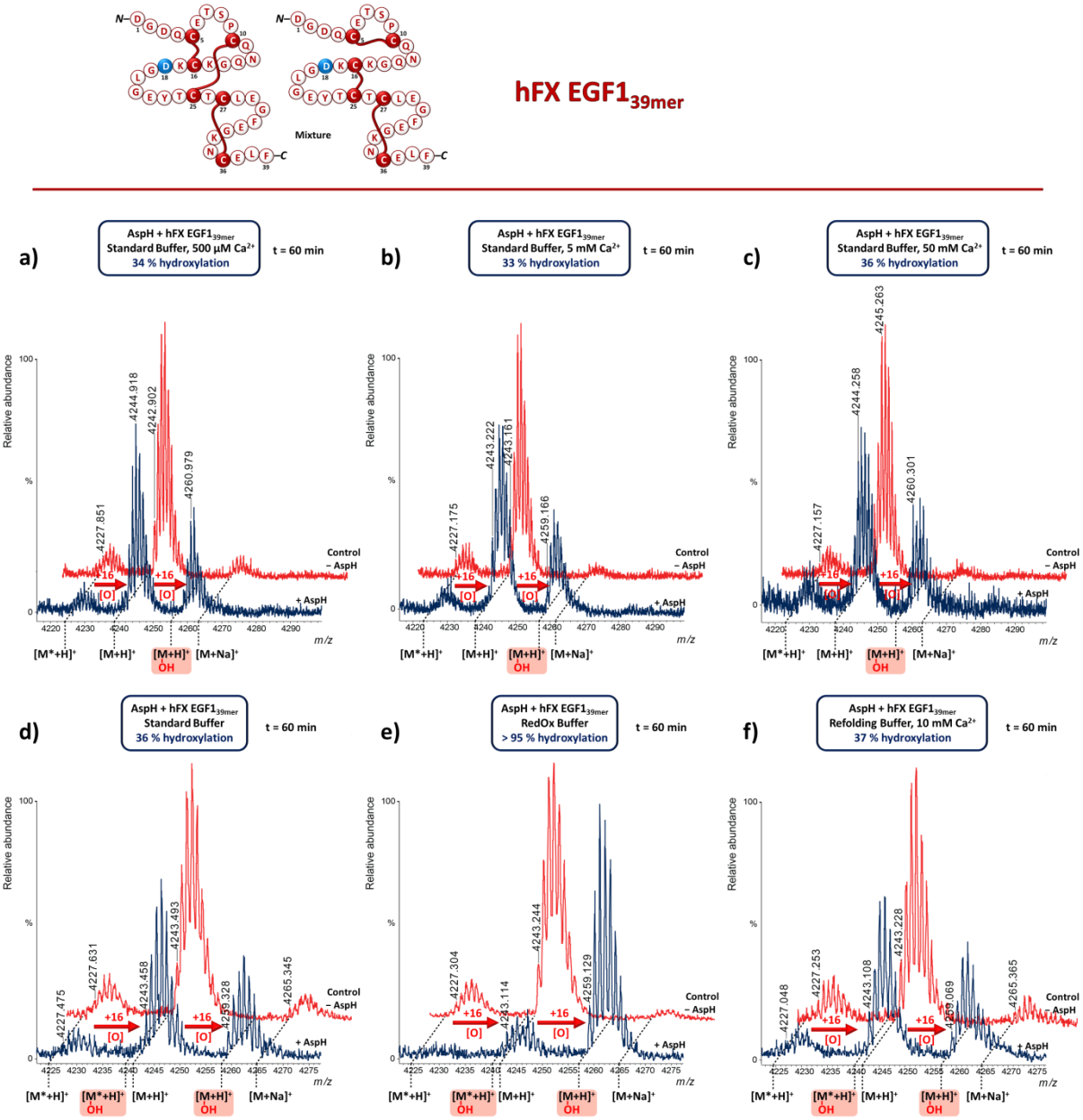


Fig. 7-8 | MALDI-ToF-MS results from preliminary studies of the effect of Ca^{2+} on hFX EGF1$_{39mer}$ hydroxylation. Assays performed using standard conditions and addition of **a)** 500 µM, **b)** 5 mM, or **c)** 50 mM $CaCl_2$ and 15 min preincubation of the EGFD peptide with Ca^{2+}. Solid-phase extraction of the samples was required to achieve a better signal-to-noise ratio. Assays performed using **d)** Standard, **e)** RedOx or **f)** Refolding Buffer (in the presence of 10 mM Ca^{2+}).

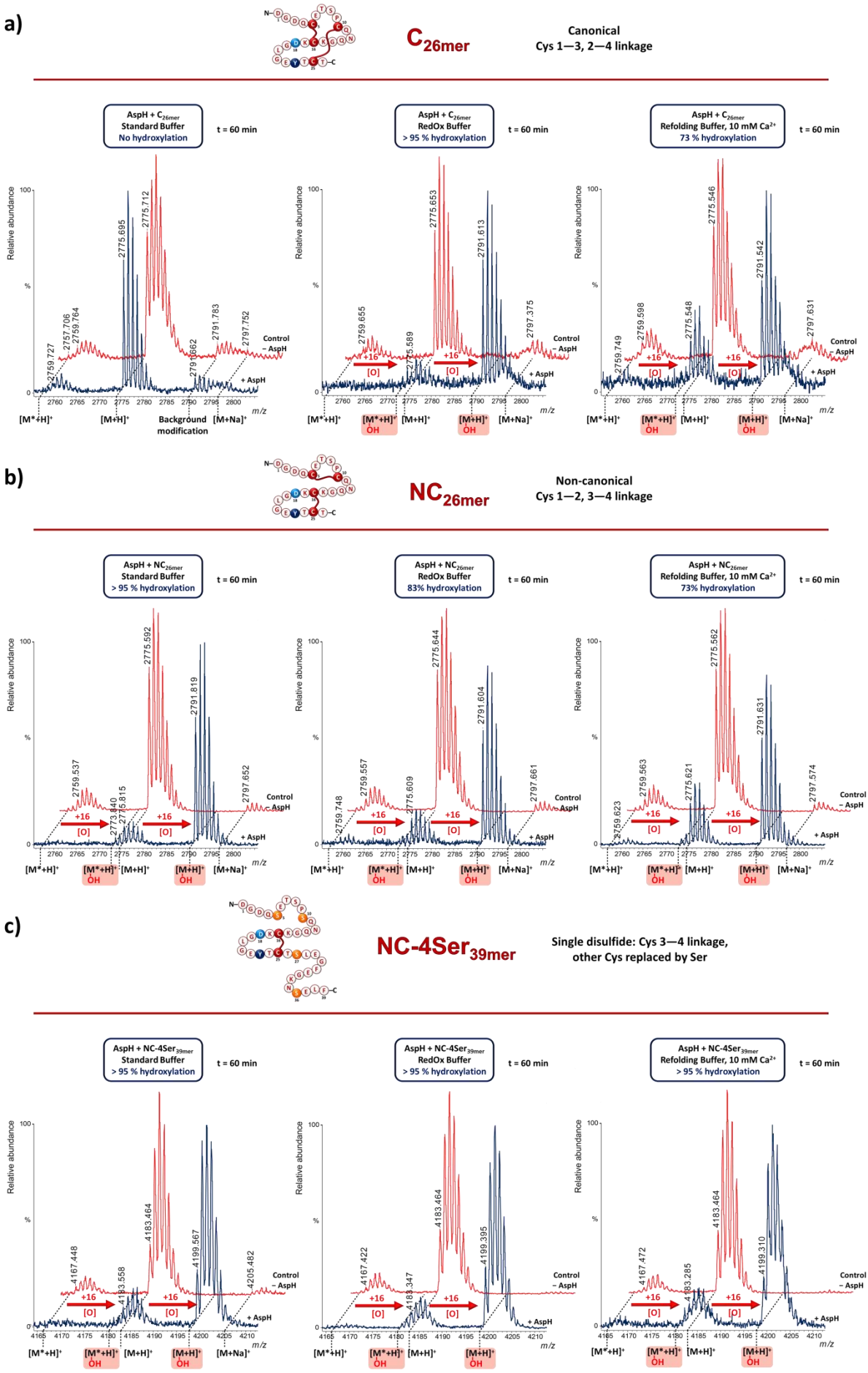


Fig. 7-9 | MALDI-ToF-MS results of **a)** C_{26mer}, **b)** NC_{26mer}, **c)** $NC\text{-}4Ser_{39mer}$ incubated with AspH and cosubstrates/cofactors in Standard, RedOx or Refolding Buffer (in the presence of 10 mM Ca^{2+}).

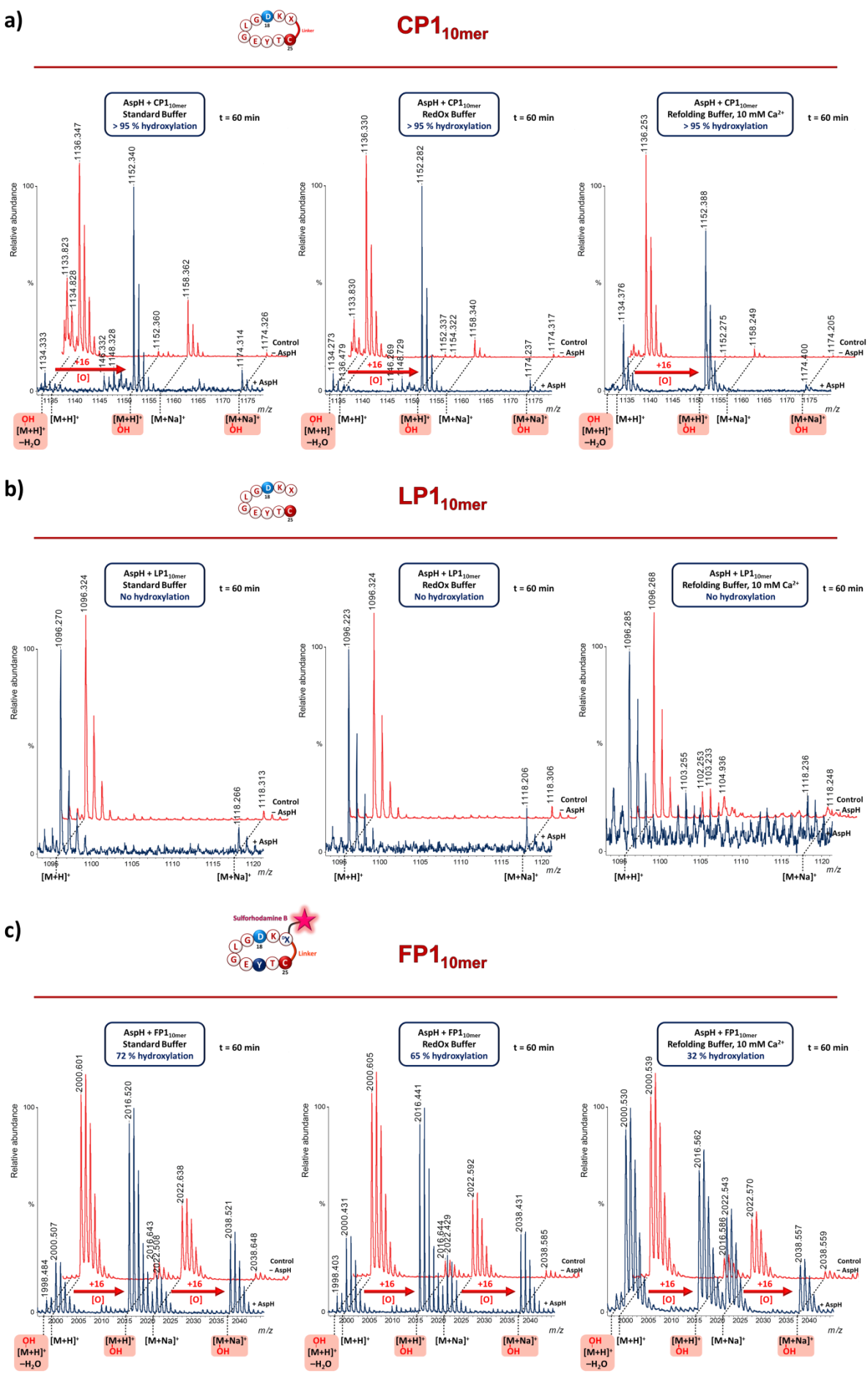


Fig. 7-10 | MALDI-ToF-MS results of **a)** $CP1_{10mer}$, **b)** $LP1_{10mer}$, **c)** $FP1_{10mer}$ incubated with AspH and cosubstrates/cofactors in Standard, RedOx or Refolding Buffer.

7.3 DISCUSSION

Ca^{2+} ions have an impairing effect on hydroxylation of EGFDs under RedOx, but not under standard conditions in the cases where several disulfide arrangements are possible. The results suggest that Ca^{2+} impedes hydroxylation by sequestration of the EGFD substrate and "locking" it in the canonical fold, which makes it unavailable for recognition by AspH. Non-canonical peptides do not appear to bind Ca^{2+}, as suggested by the results obtained using the NC-4Ser$_{39mer}$.

In a RedOx environment sufficiently high Ca^{2+} levels appear to promote formation of the canonical EGFD disulfide pattern, thereby preventing hydroxylation by AspH (**Fig. 7-11**).

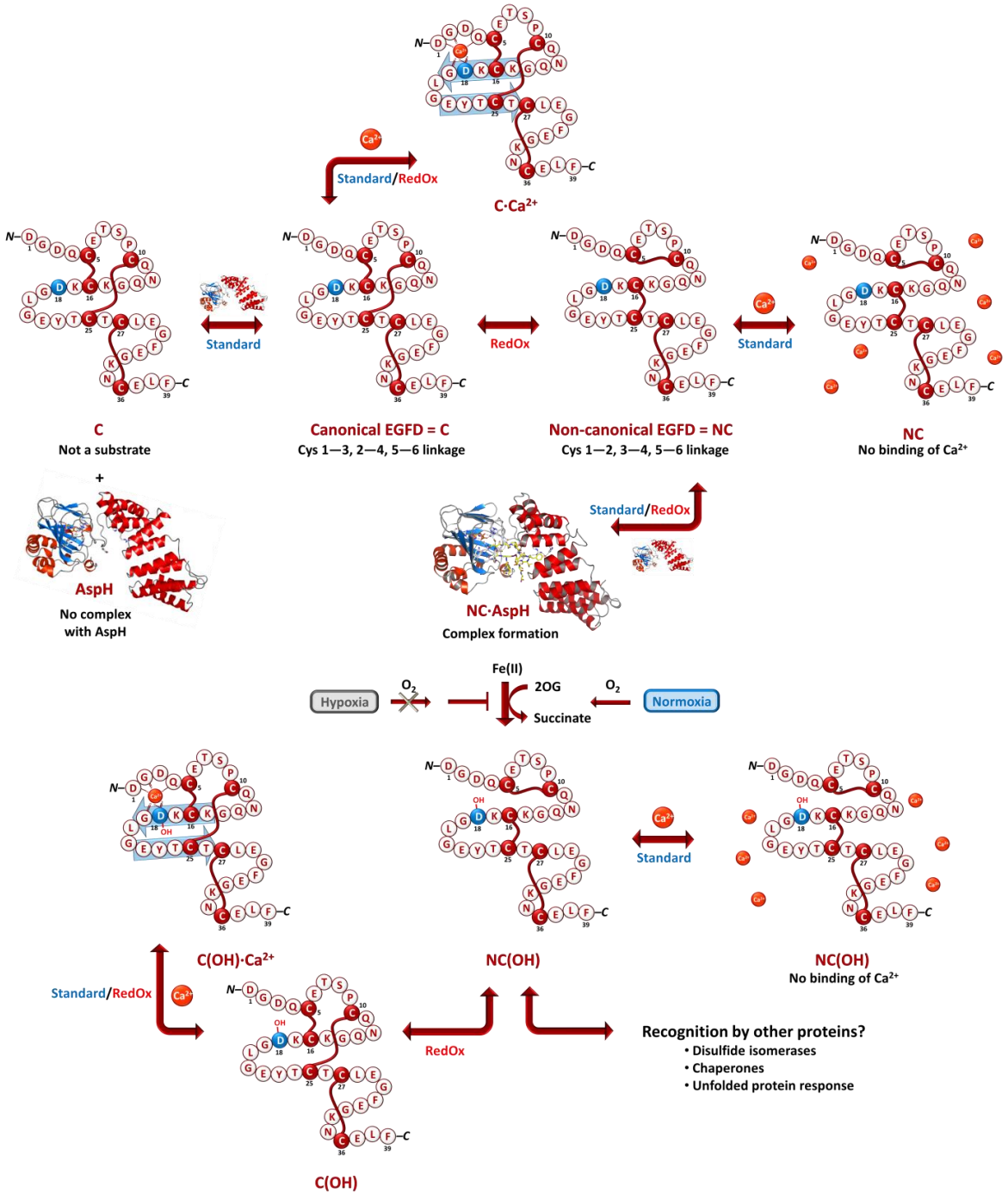
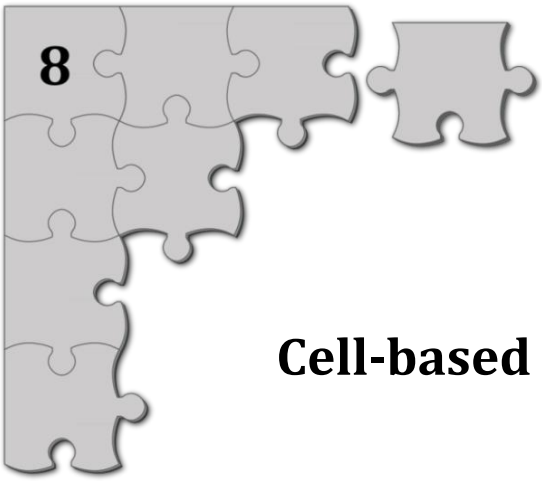


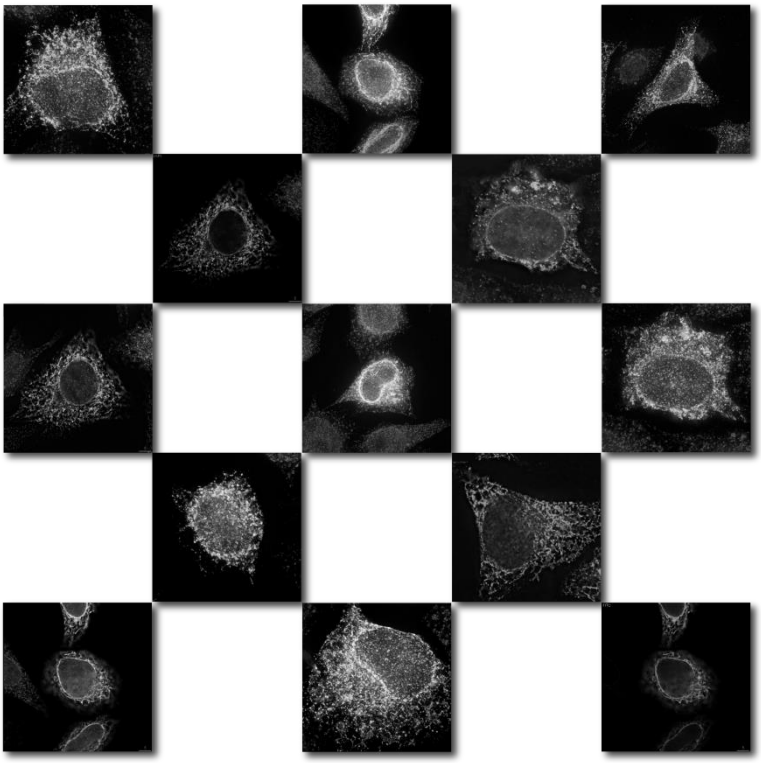
Fig. 7-11 | Ca^{2+} appears to promote the canonical EGFD disulfide pattern under RedOx conditions.

7.4 ACKNOWLEDGEMENTS FOR CHAPTER 7

Most of the MALDI-ToF-MS measurements were performed using the *Bruker Daltonics* Ultraflex I MALDI-ToF/ToF spectrometer based at the Oxion Proteomics Facility, Department of Physiology, Anatomy and Genetics, University of Oxford, and Dr Holger Kramer is kindly acknowledged for granting access to the instrument and for fruitful discussions. Professor Penny A. Handford and her coworkers Dr Sacha Jensen and Dr Paul Taylor are kindly acknowledged for providing the hFib1 and hNotch1 cbEGFD proteins.



Cell-based Studies of AspH



Abstract

This Chapter describes the results of subcellular localisation studies of full-length AspH constructs and relates the results presented in the preceding Chapters to potential physiological implications.

8.1 PROPERTIES OF THE DISTINCTIVE ASPH SPLICING ISOFORMS

The human AspH splicing isoforms partially differ in their tissue expression patterns. flAspH (AspHA), junctate (AspHB) and humbug (AspHC) are universally expressed in a variety of tissues,[342,365,366] while junctin (AspHD) is mainly found in skeletal and cardiac muscle (cardiac junctin, AspHE).[367] Junctate and junctin have been localised to the ER or sarcoplasmic reticulum (SR) membranes respectively,[365,367] junctional sites of close proximity of the ER and plasma membranes (ER-PM junctions). The *N*-termini are on the cytoplasmic side, the *C*-termini localised to the ER lumen. The *N*-terminal domains of flAspH (758 aa) and junctate (299 aa) only differ in their *N*-termini, as AspHA has a longer, serine-rich cytoplasmic tail (which humbug (313 aa) shares as well), and a small internal fragment comprising 15 residues (-AKAKDFRYNLSEVLQ-), which is found in the sequence for junctate but not for flAspH (**Fig. 8-1**). AspHF (729 aa) and junctate have the same *N*-terminus. Junctin (210 aa, cardiac junction 225 aa) differs strongly from the other AspH isoforms and does not share the humbug/junctate domain (see **Fig. 1-19** and **Appendix B**). The humbug/junctate domain is characterised by predominantly acidic residues, resulting in a net negative charge. In contrast, junctin has predominantly basic residues, and thus a net positive charge (also see **Section 8-4** for graphical illustrations).

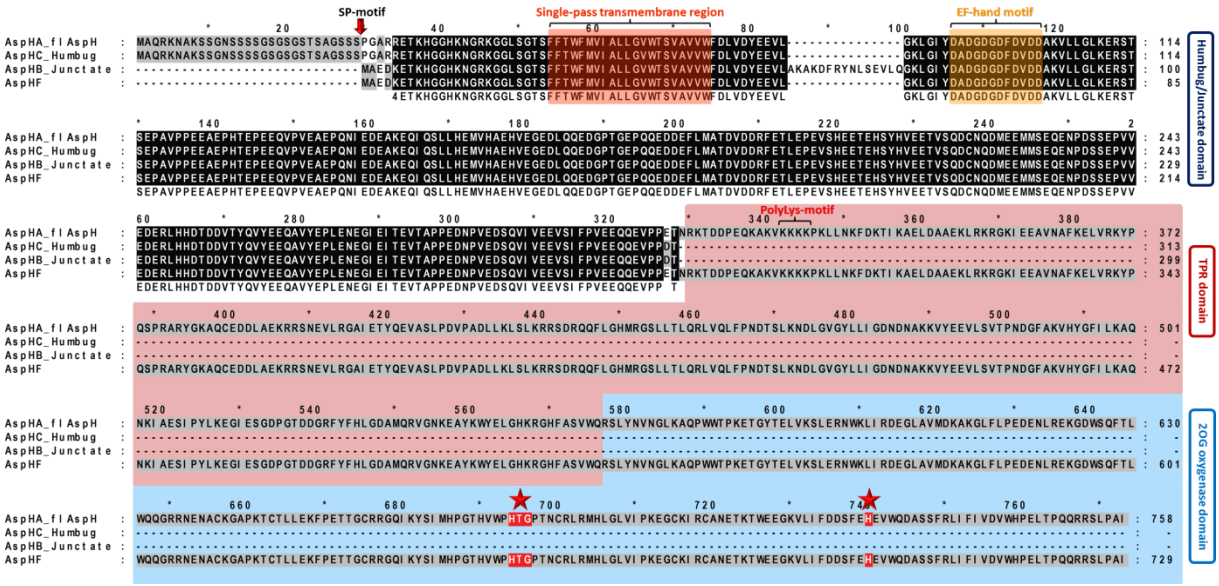


Fig. 8-1 | Sequence alignment (performed using Clustalω) of selected AspH isoforms (AspH NCBI gene ID: 444), demonstrating that AspHA = flAspH (758 aa) and AspHC = humbug (313 aa), or AspHB = junctate (299 aa) and the alternative longer isoform AspHF (729 aa) respectively, have an identical *N*-terminal domain. Humbug and junctate are truncated versions lacking the TPR and oxygenase domains of AspHA and AspHF respectively.

8.1.1 Proteolytic Processing of AspH and Subcellular Localisation

The isolation of fully active 52 kDa and 56 kDa AspH species from bovine liver has been reported.[386,387] These species were proposed products from proteolytic cleavage of the *C*-terminal

domains of the bovine full-length AspH (analogy shown for potential proteolytic processing of human flAspH in **Fig. 8-2**). If the human flAspH is processed *in vivo* in analogous manner to the bovine species remains unclear. However, as AspH immunoreactivity was detected on tumour cell surface membranes and in sera from patients, it was also proposed that AspH overexpression may enable its transportation to the cell surface.[252] Subcellular localisation data have so far only been reported for the AspH isoforms junctate[365] and junctin[367] that lack the TPR and catalytic domains. In another study, the AspH distribution in a variety of tumour cell lines was investigated by immunofluorescence staining[255] but the antibody used was raised against an isoform of the *N*-terminal domain of murine AspH, i.e. humbug/junctate, which was additionally truncated in its cytoplasmic region.[259]

8.2 Subcellular Localisation Studies on FLAG-Tagged AspH

Since no information on the subcellular localisation of catalytically active AspH isoforms has been reported thus far, subcellular localisation studies on human flAspH using immunofluorescence detection were undertaken during the work for this Thesis. In addition, the localisation and characteristics of the isoform humbug and $AspH_{315-758}$ were investigated, which would be the products from potential proteolytic cleavage of human flAspH (**Fig. 8-2**).

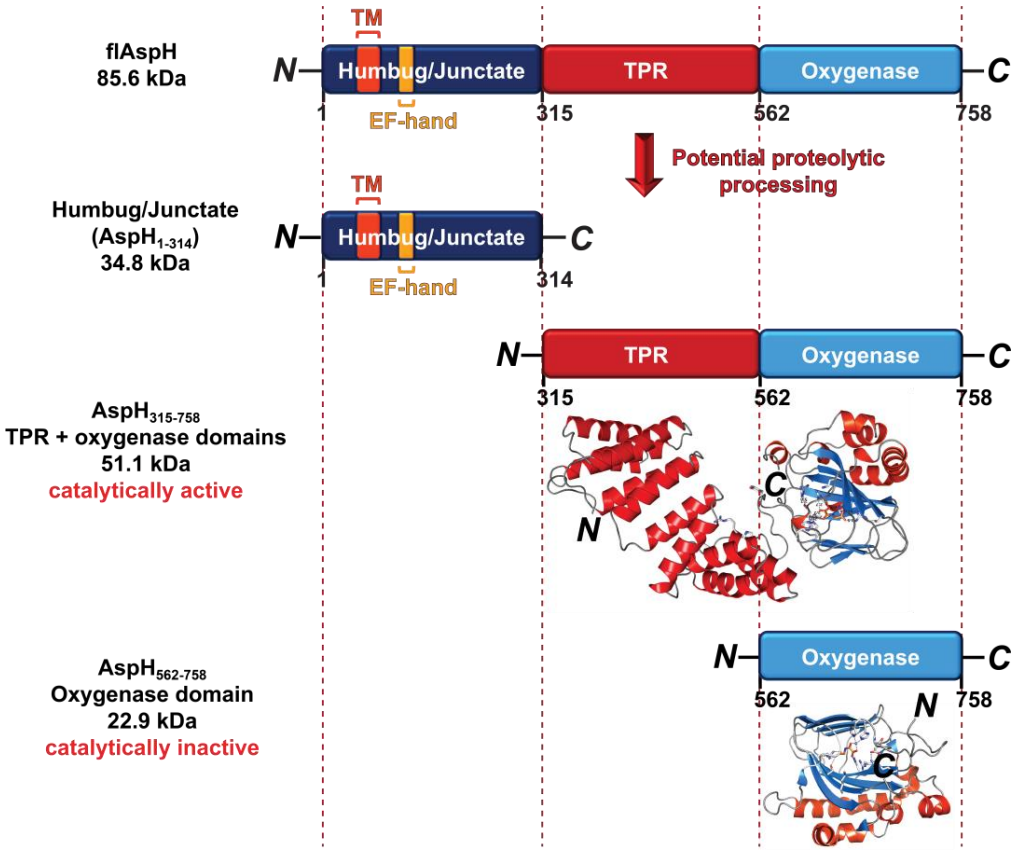


Fig. 8-2 | Potential proteolytic processing of human flAspH *in vivo* may lead to cleaved humbug and $AspH_{315-758}$, and possibly further cleavage of the oxygenase domain, $AspH_{562-758}$. How the highly conserved humbug/junctate domain containing the Ca^{2+}-binding EF-hand motif is potentially involved in EGFD hydroxylation remains unclear.

8.2.1 Preparation of FLAG-Tagged *AspH* Constructs

Ligation-independent cloning (LIC)[549,388] was used to prepare the FLAG-tagged human *AspH* constructs (**Fig. 8-3**). LIC requires no use of restriction endonucleases or DNA ligase, and enables facile and efficient cloning of the same PCR product into multiple different vectors (the LIC vectors used for this work were constructed by the SGC Oxford, see **Chapter 10**).[549,388] Using specific LIC primers (see **Chapter 10**), the DNA insert of interest was amplified by PCR, while the LIC vector was digested with the non-palindromic restriction enzyme *Bsa*I to cut out a stuffer fragment (containing the *Bacillus subtilis sacB* gene, which confers sensitivity to sucrose[388]). The purified amplified insert and the purified digested vector were then separately treated with T4 DNA polymerase, which has 3' → 5' exonuclease activity.[550] During the LIC treatment of the insert with T4 DNA polymerase and deoxycytidine triphosphate (dCTP) sticky ends were generated. Complementary single-stranded regions were produced in the LIC vector by treatment with T4 DNA polymerase in presence of deoxyguanosine triphosphate (dGTP). The insert and the vector were then briefly annealed (LIC reaction) and subsequently transformed into *E. coli*, so enabling plasmid preparation.

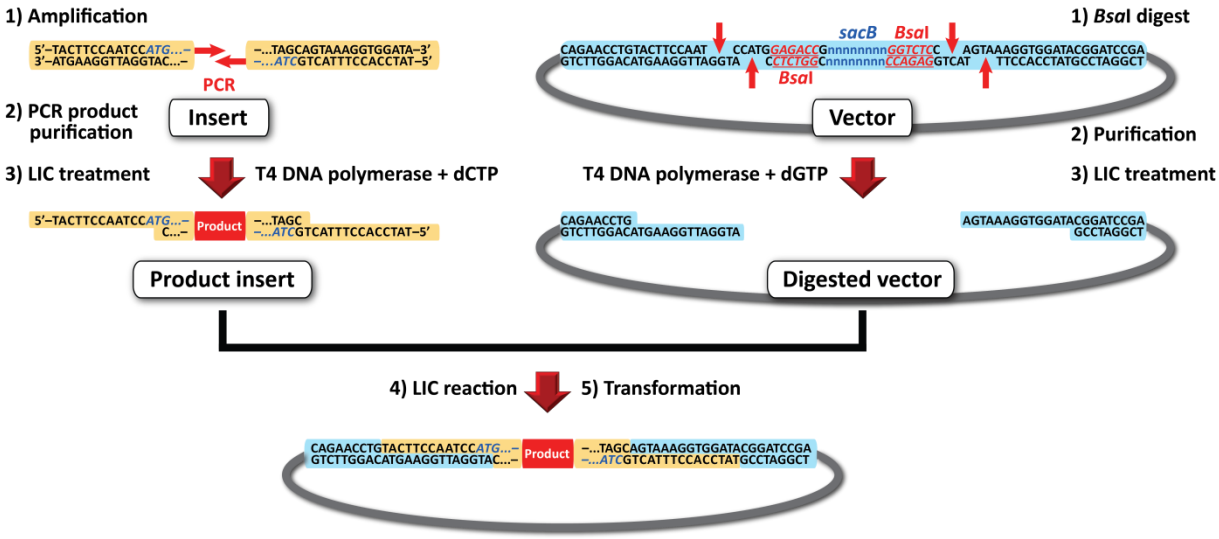


Fig. 8-3 | Scheme of a typical ligation-independent cloning (LIC) workflow. The DNA insert of interest is amplified by PCR using specific LIC primers, while the LIC vector is digested with the non-palindromic restriction enzyme *Bsa*I to cut out the *sacB* gene-containing stuffer fragment. Following PCR product purification, a LIC treatment of the insert with T4 DNA polymerase and dCTP generates sticky ends. Complementary single-stranded regions are exposed in the LIC vector by treatment with T4 DNA polymerase in presence of dGTP. In the LIC reaction, the insert and the vector are briefly annealed and subsequently transformed into *E. coli*.

AspH constructs with *N*- or *C*-terminal tags were designed (**Fig. 8-4a**), enabling detection using an appropriate antibody in immunofluorescence studies. The localisation results for the *N*- and *C*-terminally tagged proteins were later compared in order to take a potential proteolytic processing into account, which would result in cleavage of one fragment with and another without a tag.

To date, no integrated molecular analysis of genomes and their expression complementary DNA clone (I.M.A.G.E.[389] cDNA clone) for AspH is available. The sequence encoding for flAspH was therefore cloned from U-2 OS (human osteosarcoma) and HepG2 cells (human liver carcinoma) cDNA (**Fig. 8-4b**; isolated as described in **Chapter 10**), which have been reported to express high levels of *AspH*.[342,252,551] The sequence was amplified using the appropriate LIC primers for generation of an *N*-terminal FLAG-, His_6-, or GFP-tag fusion construct. Successful cloning of the constructs was verified by DNA sequencing (use of the 'middle primers for sequencing' given in **Chapter 10** allowed complete sequencing of the long constructs). The sequences encoding for flAspH with a *C*-terminal FLAG-tag (flAspH-FLAG), *N*- and *C*-terminally FLAG-tagged humbug and $AspH_{315-758}$ constructs were cloned using the FLAG-flAspH plasmid as template (the cloned FLAG-tagged humbug constructs in this work therefore differ from the natural humbug (AspHC, see **Fig. 8-1**) with the presence of D312E mutation and the presence of an additional residue, N314).

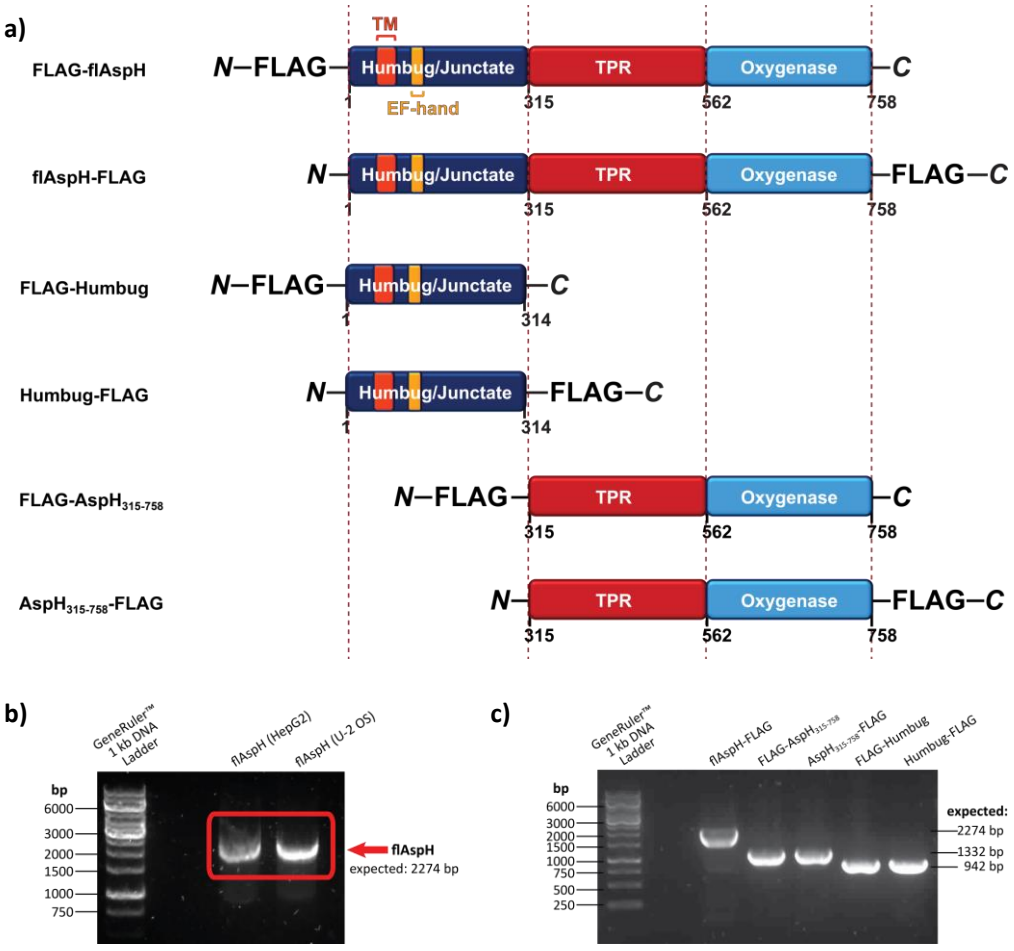


Fig. 8-4 | Production of AspH constructs for immunofluorescence studies. **a)** Constructs encoding for flAspH, humbug ($AspH_{1-314}$), and $AspH_{315-758}$ with either an *N*- or *C*-terminal FLAG-tag were generated, enabling detection using an antibody raised against the FLAG-tag (αFLAG). Agarose gels (1%, stained with SYBR® Safe DNA Gel Stain, image of the gel under UV light) representing **b)** cloning of the sequence encoding for flAspH from HepG2 or U-2 OS cDNA, amplified using LIC primers for an *N*-terminal FLAG-, His_6-, or GFP-tag fusion construct (see **Chapter 10**); **c)** cloning of the sequences encoding for the stated constructs from the flAspH template (in **b**), amplified using the appropriate LIC primers for an *N*- or *C*-terminal FLAG-tag fusion construct. Due to the use of flAspH as template, the cloned FLAG-tagged humbug constructs in this work differ from the natural humbug (AspHC, see **Fig. 8-1**) with the presence of D312E mutation and the presence of an additional residue, N314.

8.2.2 Immunofluorescence Studies

The immunofluorescence studies presented in this Chapter were conducted with assistance from Dr Joan Gannon, who obtained some of the presented data and provided training in the technique, in particular sample preparation, microscopy, and data analysis.

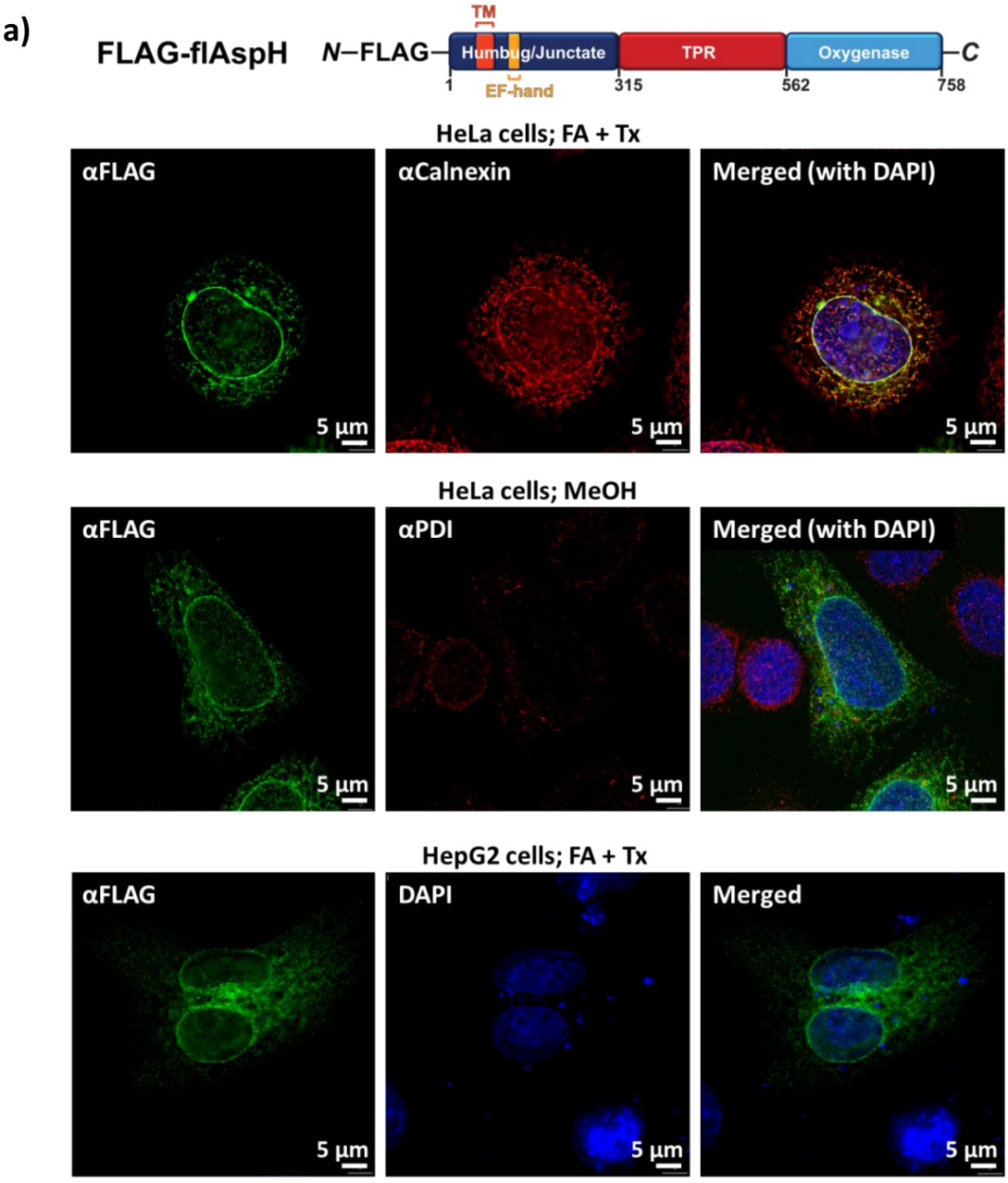
Immunofluorescence studies on the FLAG-tagged *AspH* constructs were then conducted using fixed cells. HeLa and HepG2 cells were grown on glass coverslips to an appropriate confluence for the microscopy studies, as described in **Chapter 10**. Transfection of the cells with the constructed plasmids for transient overexpression of either *N*- or *C*-terminally FLAG-tagged *AspH* constructs was performed using FuGENE® HD Transfection Reagent (*Promega*). The cells were fixed and permeabilised either using 3.7 % paraformaldehyde and subsequent treatment with 0.1 % Triton-X (FA + Tx) or using methanol (MeOH). Following fixation, the cells were stained with an antibody raised against the FLAG-tag sequence (αFLAG), 4',6'-diamidino-2-phenylindole (DAPI) to visualise DNA, and antibodies for either the ER membrane marker calnexin (αCalnexin), the soluble ER marker protein PDI (αPDI), or the exosomes marker protein HSC70 (HSPA8) respectively. AlexaFluor® dye conjugates were used as secondary antibodies for staining of the cells, enabling detection of the proteins using wide-field fluorescence microscopy.

8.2.2.1 Localisation of *N*- or *C*-FLAG-Tagged flAspH, Humbug and $AspH_{315\text{-}758}$

All investigated transmembrane domain-containing AspH proteins (i.e. FLAG-flAspH, flAspH-FLAG, FLAG-humbug, humbug-FLAG) exhibited highly defined ER network staining, reaching out to the cell periphery with distinct perinuclear accentuation (**Fig. 8-5**, **Fig. 8-6**, **Fig. 8-7**). Both the *N*- and *C*-terminally FLAG-tagged flAspH and humbug forms colocalised well with the ER membrane marker calnexin but less well with the soluble ER marker PDI. No significant differences in the subcellular localisation of the tagged flAspH and humbug proteins were observed in HeLa or HepG2 cells.

For the truncated *N*- or *C*-terminally FLAG-tagged $AspH_{315\text{-}758}$ no fixed, clearly transfected cells were found. In some cases, sparse and diffuse cytoplasmic staining and partial colocalisation with the exosomes marker HSC70 were observed (**Fig. 8-8**), which could indicate that $AspH_{315\text{-}758}$ may be secreted out of the cell.

For all investigated *AspH* constructs, at least three independent experiments were conducted, and several cells were examined and imaged on each coverslip. Consistent results were obtained in the replicate experiments, and representative images are shown below for each construct.



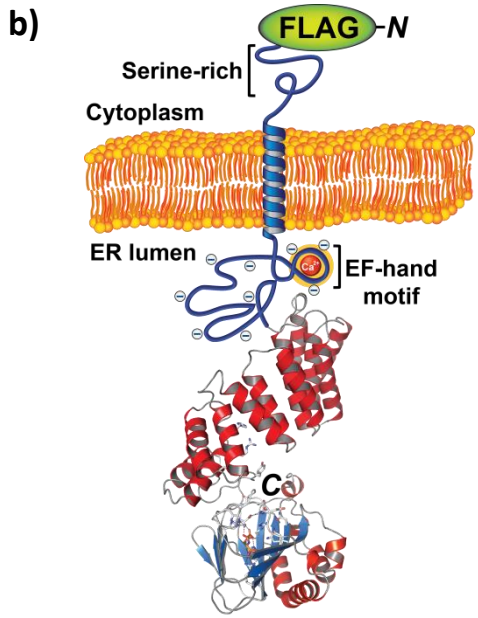
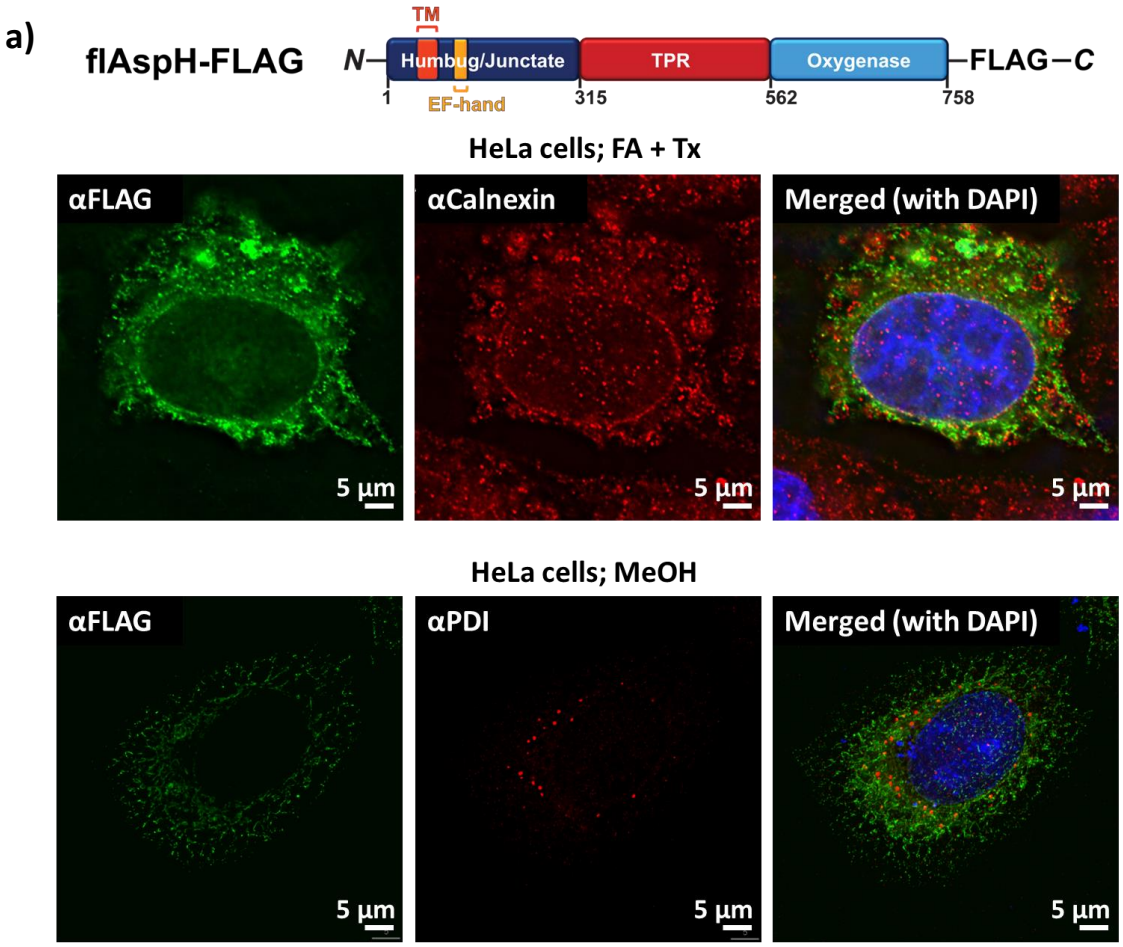


Fig. 8-5 | Immunofluorescence microscopy. **a)** Images of HeLa or HepG2 cells transfected with *N*-terminally FLAG-tagged flAspH; fixation and permeabilisation performed using 3.7 % paraformaldehyde and 0.1 % Triton X 100 (FA + Tx) or methanol (MeOH), as indicated. FLAG-flAspH (green) localises to the ER membrane, as seen by the network structure around the nuclear envelope and co-localisation with the ER membrane marker protein calnexin (red, top panel). Comparably sparse co-localisation of FLAG-flAspH with the soluble ER marker protein disulfide isomerase (PDI, red, middle panel) was observed. Scale bars represent 5 µm. The imaged cells varied in size. **b)** Cartoon illustrating the proposed FLAG-flAspH topology based on the localisation studies results.



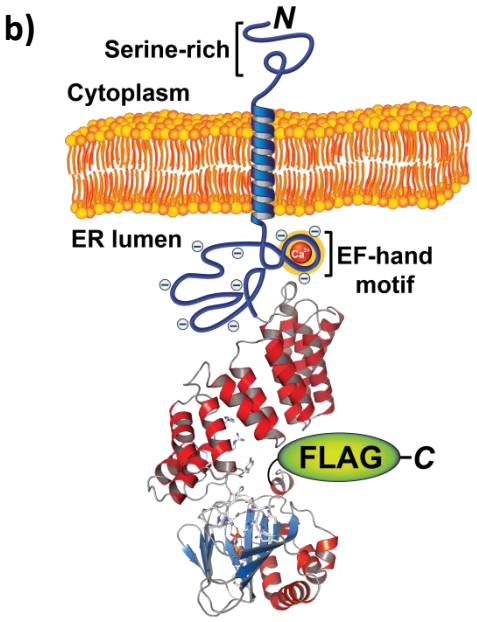


Fig. 8-6 | Immunofluorescence microscopy. **a)** Images of HeLa or HepG2 cells transfected with *N*-terminally FLAG-tagged flAspH; fixation and permeabilisation performed using 3.7 % paraformaldehyde and 0.1 % Triton X 100 (FA + Tx) or methanol (MeOH), as indicated. FLAG-flAspH (green) localises to the ER membrane, as seen by the network structure around the nuclear envelope and co-localisation with the ER membrane marker protein calnexin (red, top panel). Comparably sparse co-localisation of FLAG-flAspH with the soluble ER marker protein disulfide isomerase (PDI, red, middle panel) was observed. Scale bars represent 5 µm. **b)** Cartoon illustrating the proposed FLAG-flAspH topology based on the localisation studies results.

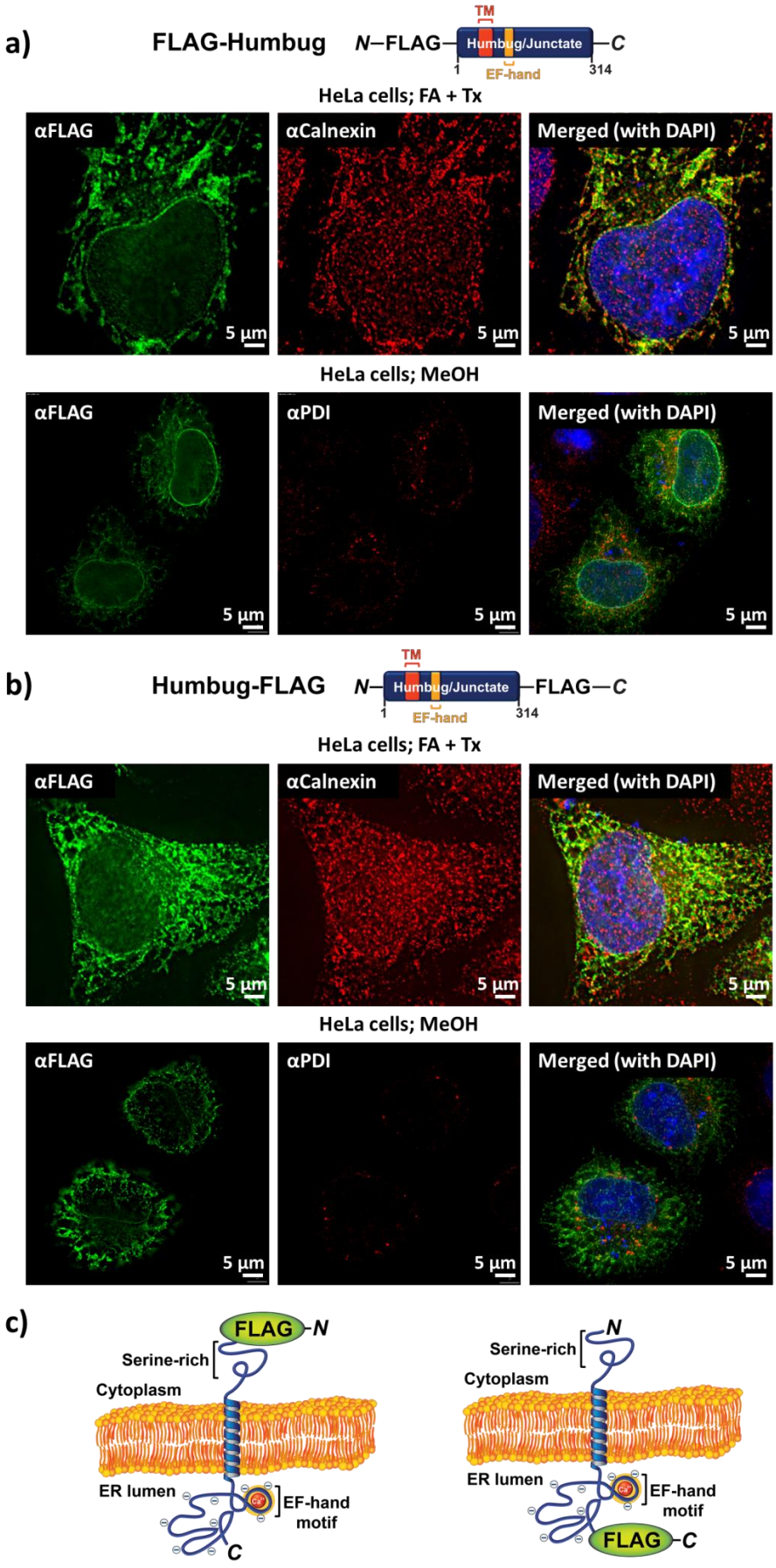


Fig. 8-7 | Immunofluorescence microscopy images of HeLa cells transfected with **a)** *N*-terminally or **b)** *C*-terminally FLAG-tagged humbug ($AspH_{1-314}$). Both FLAG-humbug and humbug-FLAG (green) localise to the ER membrane and co-localise with the ER membrane marker protein calnexin (red). Scale bars represent 5 µm. The imaged cells varied in size. **c)** Cartoons illustrating the proposed FLAG-humbug and humbug-FLAG topologies based on the localisation studies results.

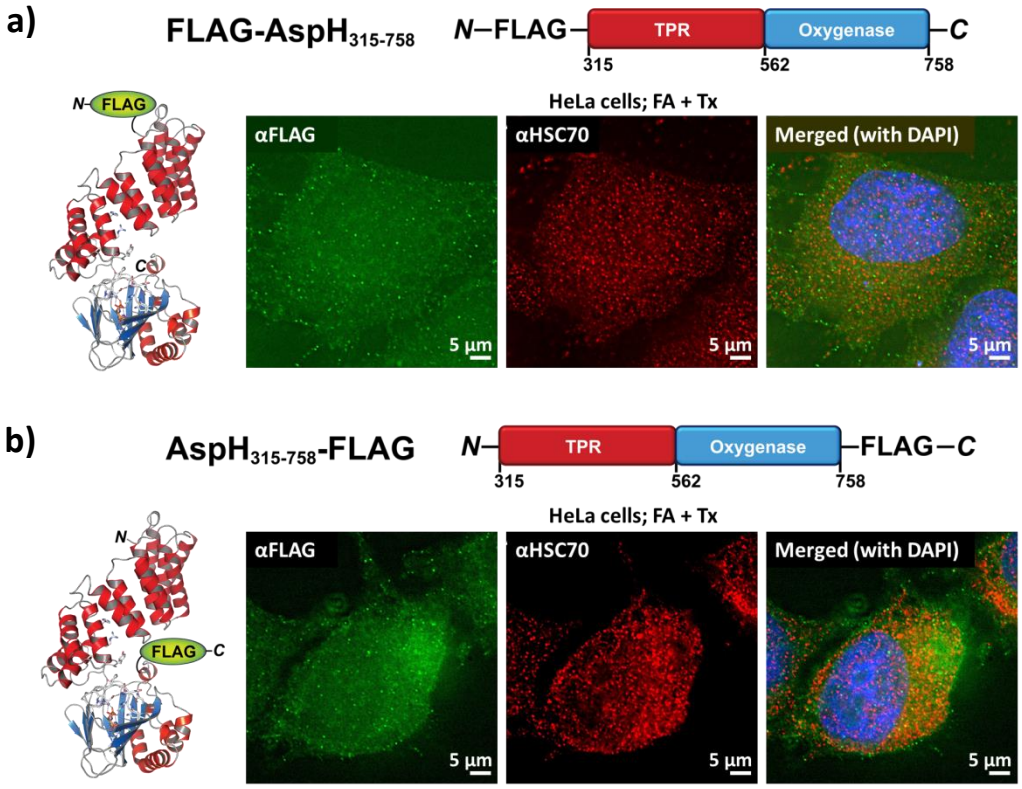


Fig. 8-8 | Immunofluorescence microscopy images of HeLa cells transfected with **a)** *N*-terminally or **b)** *C*-terminally FLAG-tagged $AspH_{315-758}$; fixation and permeabilisation performed using 3.7 % paraformaldehyde and 0.1 % Triton X 100 (FA + Tx). No fixed, clearly transfected cells were observed in the studies with FLAG-$AspH_{315-758}$ and $AspH_{315-758}$-FLAG; the images with high background are just shown here as examples of the sparse, diffuse cytoplasmic staining (green) and partial colocalisation with the exosomes marker HSC70 observed in some cells. Scale bars represent 5 µm.

8.3 Work Towards Longer AspH *in vitro* Proteins

Since the studies presented in the preceding Chapters were conducted using the truncated but catalytically active $AspH_{315-758}$, a potential involvement and role of the *N*-terminal humbug/junctate domain in EGFD hydroxylation remains unclear. Since the humbug/junctate domain contains a Ca^{2+}-sensing EF-hand motif, constructs containing this motif could enable investigation of the potential interplay of the different AspH domains.

The sequence encoding for flAspH was cloned into the bacterial pNIC-28 Bsa4 LIC vector for production of an *N*-terminal His_6-tag fusion protein. However, no clear overexpression as soluble or insoluble protein was observed, and no protein could be produced from small-scale expression trials in *E. coli* host strains BL21 (DE3), Rosetta 2 (DE3), or Origami 2 (DE3). Purification trials using Ni^{2+}-affinity purification to enrich the protein remained unsuccessful, and staining with InVision™ His-Tag In-Gel Stain (*Invitrogen*) for visualisation of the His_6-tag was inconclusive.

As the construct encoding for the full-length protein including the transmembrane domain may be difficult to express in soluble form in *E. coli*, *N*-terminally truncated *AspH* constructs containing no transmembrane domain but the EF-hand motif were then cloned (**Fig. 8-9**). However,

trial expression studies (1.2 l media) using the same conditions as optimised for His_6-$AspH_{315-758}$ (expression for 16 h at 18 °C after induction with 0.1 mM IPTG at OD_{600} 1.1-1.3) and trial Ni^{2+}-affinity purification (1 ml HisTrap™) produced none of the desired proteins (His_6-$TMtAspH_{76-758}$ expression in *E. coli* BL21 (DE3), Rosetta 2 (DE3), or Origami 2 (DE3), both His_6-$TMtNAspH_{76-314}$ and His_6-$TMtAspH_{76-561}$ expression in *E. coli* BL21 (DE3), and subsequent trial purification). Due to the required excessive optimisation and time constraints of this work, the production of these proteins was not further investigated.

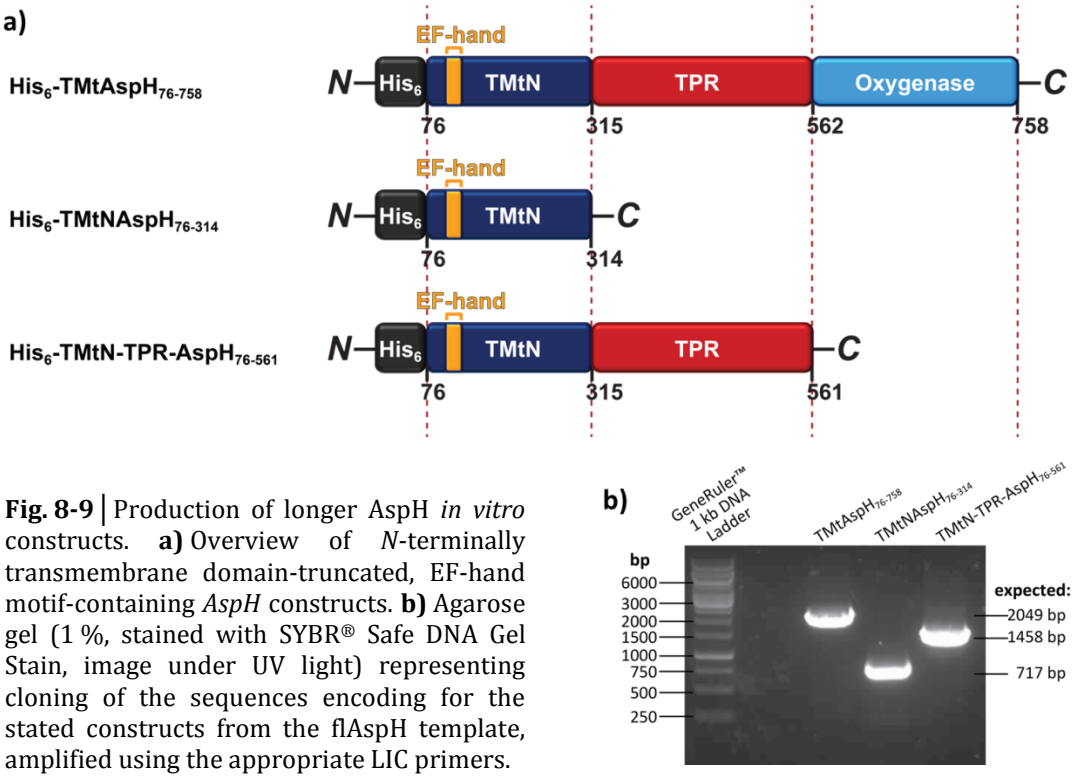


Fig. 8-9 | Production of longer AspH *in vitro* constructs. **a)** Overview of *N*-terminally transmembrane domain-truncated, EF-hand motif-containing *AspH* constructs. **b)** Agarose gel (1 %, stained with SYBR® Safe DNA Gel Stain, image under UV light) representing cloning of the sequences encoding for the stated constructs from the flAspH template, amplified using the appropriate LIC primers.

8.4 DISCUSSION

The immunofluorescence studies in this Chapter showed that all investigated AspH proteins containing the transmembrane domain localise to the ER (**Fig. 8-10**), visible as the highly defined network structure around the nuclear envelope. The obtained flAspH and humbug localisation data exhibited the same characteristics as the data reported for the isoform junctate.[365]

The *N*- or *C*-terminally FLAG-tagged $AspH_{315-758}$ proteins, which had no transmembrane domain anchoring the protein to the ER membrane, may have been released from the cells via exosomes.[552] In that case, the proteins could not be visualised by immunofluorescence in the fixed cells, which would explain the lack of clearly transfected cells. However, the expression of these constructs may have been impaired, or rapid protein degradation in the HeLa or HepG2 cells may have occurred, thus preventing immunofluorescence detection. More extensive studies involving

'live' cell imaging using the GFP-tagged proteins would be required to determine whether the proteins were produced and to investigate the localisation.

Despite the observed FLAG-flAspH and flAspH-FLAG localisation to the ER, a proportion of the overexpressed flAspH may be proteolytically processed, which could be masked by satisfying detection of the full-length protein. A much lower number of transfected cells were observed in the specimens prepared using flAspH-FLAG, compared to the other transmembrane domain-containing constructs (i.e. FLAG-flAspH, FLAG-humbug, humbug-FLAG). This could be due to proteolytic processing of flAspH-FLAG, leading to transportation of the *C*-terminal domains including the FLAG-tag out of the cell. Live cell imaging using flAspH-GFP would allow investigating the results further.

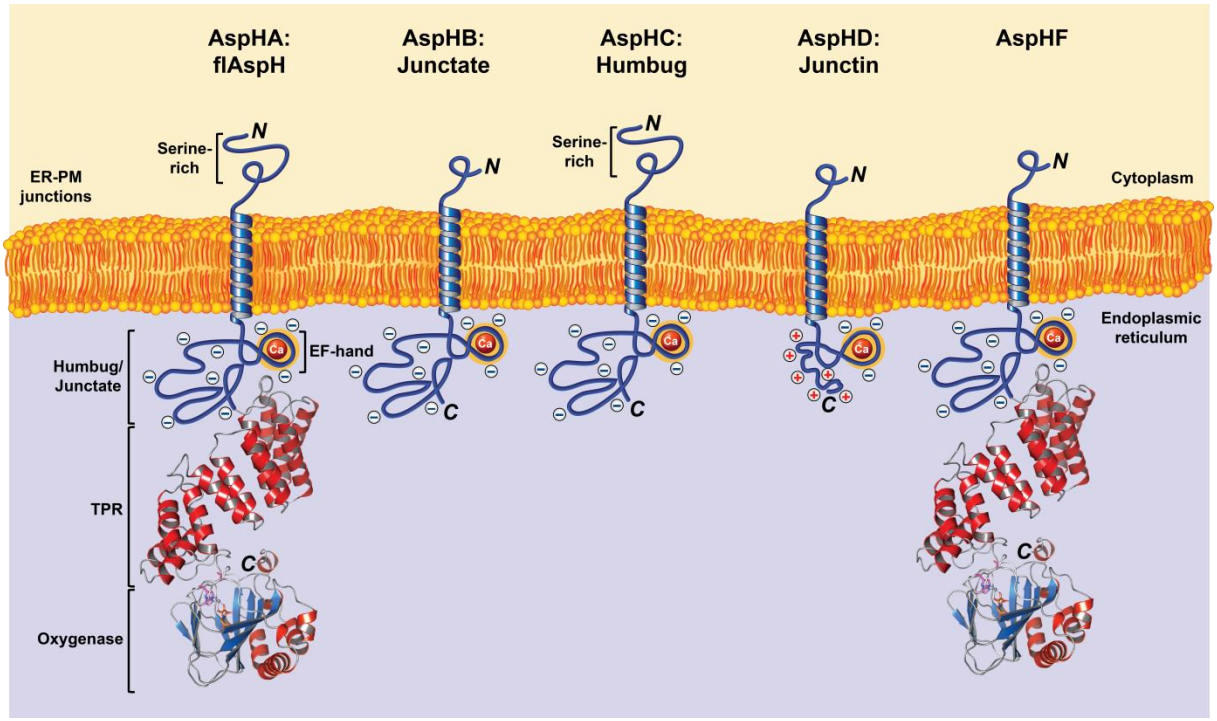


Fig. 8-10 | Cartoon showing the topologies of distinct AspH isoforms. All contain a transmembrane domain and a Ca^{2+}-binding EF-hand motif on the luminal side. Humbug is a truncated version of AspHA = flAspH, and shares the same Ser-rich cytoplasmic *N*-terminus. Junctate is the corresponding truncated version of AspHF; both lack the Ser-rich *N*-terminal region. The luminal domain of junctate and humbug is identical to that in AspHA and AspHF and characterised by the presence of multiple acidic residues, which result in an overall net negative charge. In contrast, junctin differs significantly in the luminal domain and is characterised by predominantly basic residues, resulting in an overall net positive charge

Since no structure of the *N*-terminal humbug/junctate domain is available and work towards longer AspH *in vitro* proteins including the *N*-terminal domain came to nothing, the true flAspH topology remains unclear (**Fig. 8-11**). The oxygenase domain may directly interact with the *N*-terminal humbug/junctate domain, or the TPR domain may be in contact with the ER membrane (possibly through interaction between a poly-Lys motif found in AspH (see **Fig. 8-1**) with phospholipids in the membrane). The conformations of the protein domains may be flexible

depending on the local environment. Crystal structures of the full-length protein or truncated forms reaching up to the transmembrane region will be required for further characterisation.

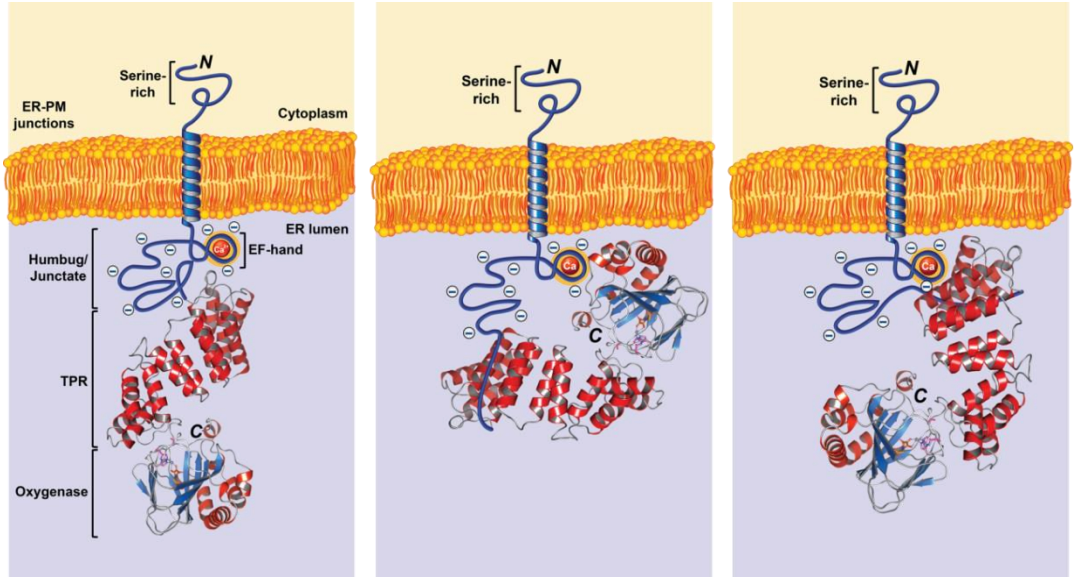


Fig. 8-11 | Cartoon illustrating some other possible flAspH topologies. The true flAspH topology and domain interactions remain unclear.

8.4.1 High Similarity Between flAspH and Junctate

The AspH isoform junctate has been reported to modulate Ca^{2+} entry induced by activation of the inositol 1,4,5-trisphosphate receptor ($InsP_3R$).[369] In addition, Ca^{2+} entry involving junctate is also directly mediated by Ca^{2+} depletion in the ER independently of receptor activation.[369] More recently, junctate has been reported to be a structural component of orai1 and stromal interaction molecule 1[553] (STIM1).[371,554] These proteins form Ca^{2+} release-activated Ca^{2+} (CRAC) channels that mediate store-operated Ca^{2+} entry (SOCE) from the extracellular space into the cytoplasm upon detection of ER Ca^{2+} depletion.[555] SOCE is suppressed during cell division,[556] and regulation by serine phosphorylation of STIM1 in serine-proline motifs (SP-motifs) has been reported.[556,557] Notably, the cytoplasmic tail of some AspH isoforms such as flAspH and humbug is Ser-rich and also contains an SP-motif, which could potentially be phosphorylated (see **Fig. 8-1**).

The AspH isoform junctin is also known as calsequestrin-binding protein and has been reported to associate with the ryanodine receptor (RyR), calsequestrin (CASQ), and triadin (TDN) to form a Ca^{2+} release complex at the junctional SR of striated muscle cells.[370] Since the Ca^{2+}-sensing EF-hand motif is strictly conserved, all AspH isoforms may have roles in Ca^{2+} homeostasis.

In excitable cells, i.e. cardiac and skeletal muscle, junctin in the SR appears to fulfil an analogous role to junctate/flAspH in the ER in non-excitable cells. This is reflected by parallels in the assemblies of the interacting partners (**Fig. 8-12**): a Ca^{2+} channel localised to the plasma membrane (① the voltage-dependent L-type Ca^{2+} channel Ca_V vs. orai1), a ligand-gated receptor (② RyR vs. $InsP_3R$), and partner proteins in the SR/ER (③ e.g. triadin, calsequestrin vs. e.g. STIM1, calreticulin).[369,371,370]

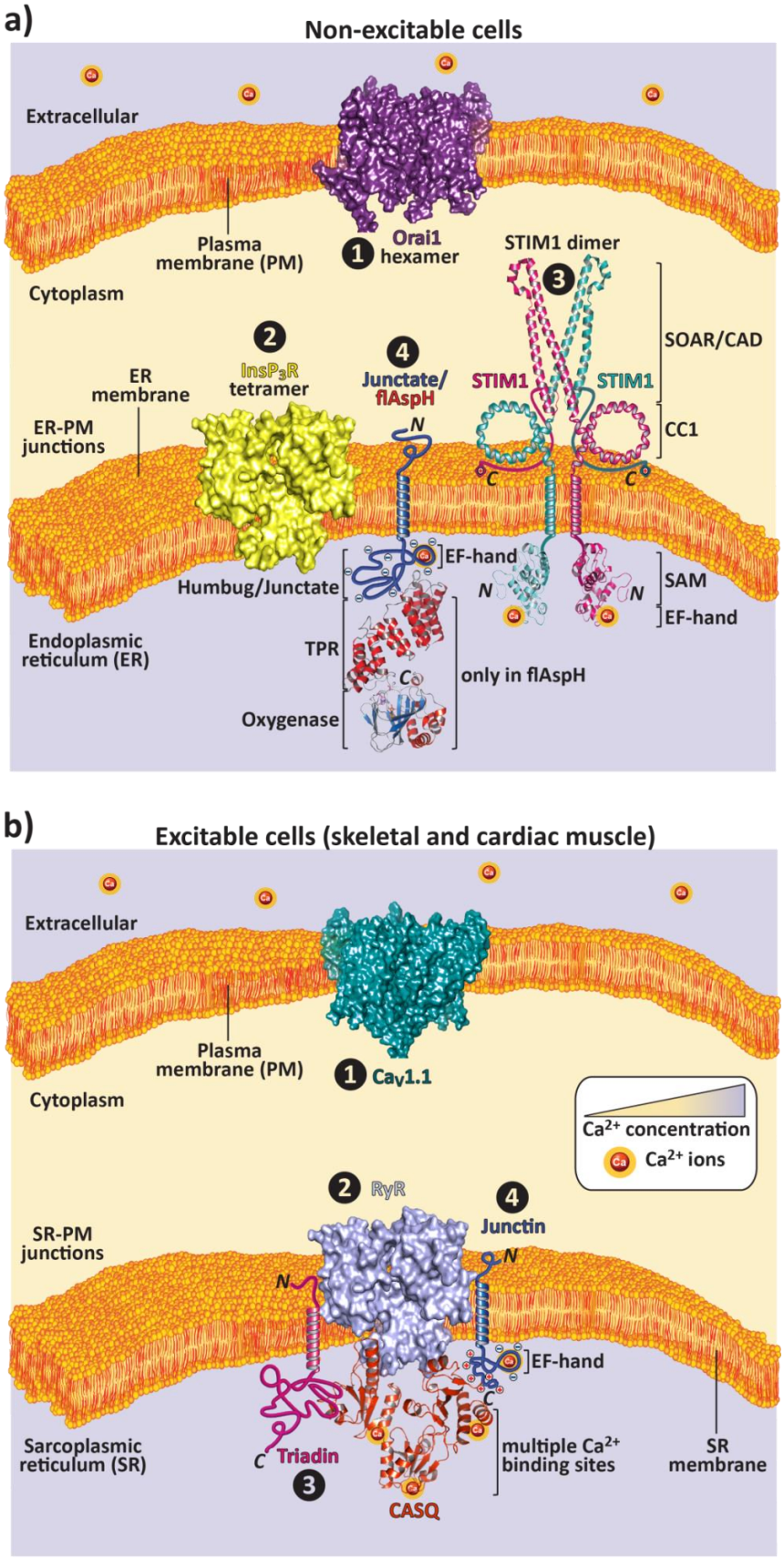


Fig. 8-12 | All AspH isoforms may have roles in Ca^{2+} homeostasis. Analogy between multi-protein assemblies enabling Ca^{2+} entry in **a)** non-excitable and **b)** excitable cells. Due to the high similarity to its isoform junctate, flAspH (④) may also interact with the store-operated Ca^{2+} entry machinery proteins including the Ca^{2+} channel orai1 (①; PDB: 4HKR[558]), the ER membrane-bound protein STIM1 (③; parts from PDBs: 3TEQ[559] and 2K60[560]), and the ligand-gated $InsP_3R$ (②). In a similar assembly linked to the voltage-dependent L-type Ca^{2+} channel Ca_V (①), the AspH isoform junctin (④) interacts as a Ca^{2+}-sensor with the ligand-gated RyR (②) and partner proteins in the SR such as triadin and calsequestrin (③; CASQ, from PDB: 1A8Y[561]).

The *N*-terminal domain of flAspH differs from junctate only in its additional Ser-rich *N*-terminal extension (the small internal segment -AKAKDFRYNLSEVLQ- is also missing in the study reporting junctate as a structural component of orai1 and STIM1[371]). The high degree of sequence identity, the highly similar subcellular localisation and tissue distribution may endow flAspH with the same Ca^{2+}-binding properties as they have been reported for junctate.[369,371] Furthermore, flAspH may thus share the abilities to interact with proteins such as orai1, STIM1, and the $InsP_3R$, but flAspH has additional capabilities to act as EGFD hydroxylase.

flAspH may have a role in quality control of 'correct' cbEGFD folding in the ER that is potentially linked to the SOCE machinery. As demonstrated by the data presented in the previous **Chapters**, AspH selectively hydroxylates non-canonical EGFDs (regarded as "misfolded") *in vitro*. The hydroxylation creates a new site for potential attachment of groups and for interaction or recognition. The downstream pathway of the hydroxylated non-canonical EGFD substrate, however, remains unclear. Ubiquitinylation of the hydroxyl group followed by proteasomal degradation is no plausible pathway, since significant amounts of βHya/βHyn from partial to full hydroxylation have commonly been reported in functional secreted and cell surface-bound proteins. The hydroxyl group, however, could mark for recognition by proteins possessing disulfide isomerase and chaperone activity that may refold the domains by forming the canonical disulfide linkages before the "misfolded" EGFD can get transported[562,563] out of the ER. An activation of the unfolded protein response[564-568] (UPR) mediators or the endoplasmic reticulum-associated protein degradation[569-572] (ERAD) pathway may also be triggered. Known disulfide isomerases and chaperones rely on Ca^{2+} for their action, including PDI,[573] BiP,[574] calnexin[575] and calreticulin.[576] However, Ca^{2+} apparently acts locally only.[539] EGFD misfolding will be more relevant under ER stress conditions with prolonged Ca^{2+} depletion as in cancer, hypoxia, or inflammation, and a folding control will hence be of particular importance.[577]

Detection of a non-canonical EGFD may effectuate a conformational change within flAspH, setting off the trigger for orai1-STIM1-mediated Ca^{2+} entry to replenish the ER Ca^{2+} levels. SOCE would be inactivated again when the EF-hand motif is replete with Ca^{2+}. As EGFDs typically have higher Ca^{2+} affinities (in the intact proteins in the micromolar to nanomolar range[544,350]) than EF-hand motifs (moderate affinities in the millimolar to micromolar range[578-580]), sufficient Ca^{2+} levels required for refolding of the EGFDs would be ensured in this way.

Notch and its ligands have roles in cell fate determination and cellular differentiation and are known to be oncogenic.[581] Perturbed coagulation with promoted fibrin clot formation and platelet activation is strongly linked to tumour progression and angiogenesis (tumours can be seen as wounds that do not heal).[582] Extracellular matrix molecules containing EGFDs have been associated with cell motility and cancer invasiveness.[582] A 'correct' EGFD folding may play a role for the functioning of these Ca^{2+}-dependent proteins, and thus for cancer growth. SOCE will lead to rapid entry of Ca^{2+} into the cytosol in the first instance.[555] Increased intracellular concentrations of Ca^{2+} can provide a growth

advantage for tumours by promoting aberrant Ca^{2+} signalling.[583] This may lead to the activation of transcription factors such as nuclear factor of activated T-cells (NFAT) that has been reported to translocate to the nucleus and lead to stimulation of growth and differentiation.[584] The Ca^{2+} channel $Ca_V1.2$ can even directly modulate transcription when it is proteolytically cleaved to form the Ca^{2+} channel-associated transcriptional regulator (CCAT) in response to changes in intracellular Ca^{2+} levels.[585]

Whether enzymatically active flAspH is involved in Ca^{2+} homeostasis in a similar manner as its isoform junctate remains unclear. However, flAspH may be a contributor to unfolded protein response mechanisms in the ER.

8.4.2 AspH Homologues AspHD1 and AspHD2

Only little information about the AspH homologues aspartyl/asparaginyl-β-hydroxylase domain-containing proteins 1 and 2 (AspHD1 and AspHD2, encoded by the *AspHD1* or *AspHD2* genes respectively) is available. As for AspH, these proteins are proposed to contain a single-pass transmembrane domain spanning the ER membrane. No EF-hand motif or significant sequence similarity with flAspH in the TPR domain region have been identified, but they exhibit good homology within the oxygenase domain (**Fig. 8-13**). The oxygenase domains in these proteins may therefore have a similar geometry. Interestingly, the AspH active site catalytic dyad HXG…H is not conserved in AspHD1 but an Arg residue is found in the position of the first His residue, resulting in a RCG…H motif (**Fig. 8-13**). Furthermore, both AspH homologues contain two vicinal cysteines in a CXXC motif[433] (CRRC in AspHD1, CRKC in AspHD2), which is not found in flAspH. However, the first Cys residue is also conserved in flAspH ($Cys641_{flAspH}$) and forms a disulfide with $Cys648_{flAspH}$ in all obtained crystal structures. Most of the other residues important for AspH-substrate interactions within the AspH oxygenase domain are conserved in AspHD1 and AspHD2.

The mutagenesis studies in **Chapter 5** showed that the active site H679R mutant had strongly diminished hydroxylation activity, compared to the AspH wildtype. Whether the H679R mutant can also bind βHya/βHya residues in the hydroxylated (non-canonical) EGFD substrate, requires further investigation.

CXXC motifs are also found in thioredoxins[433] and can enable intermolecular disulfide formation and exchange, and thus the ability to potentially act as a disulfide isomerase and folding aid.[433,586] Extensive studies will be required to investigate the potential roles of AspHD1 and AspHD2 in EGFD hydroxylation or as disulfide isomerases in downstream processing and potential refolding of hydroxylated non-canonical EGFDs.

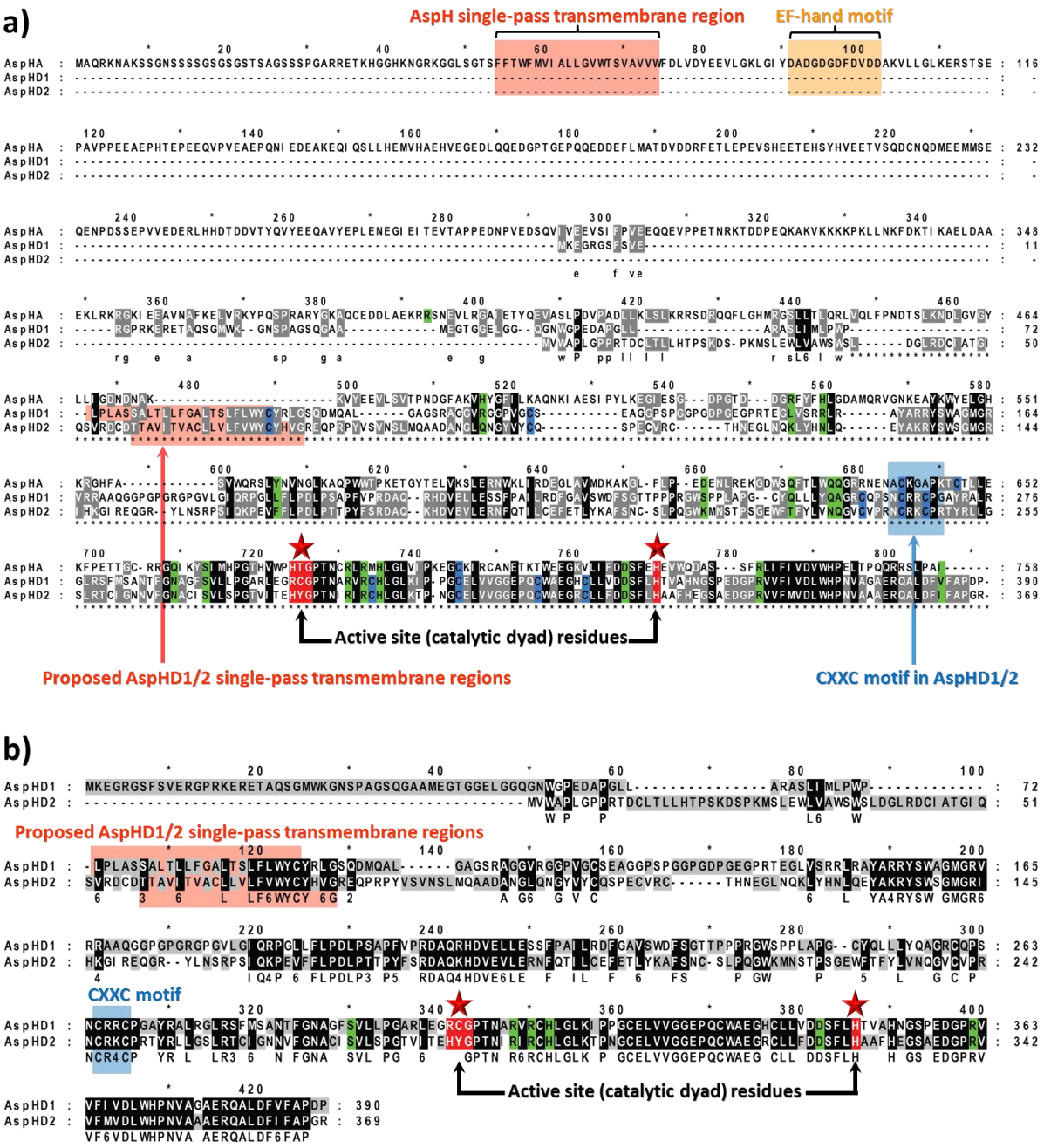
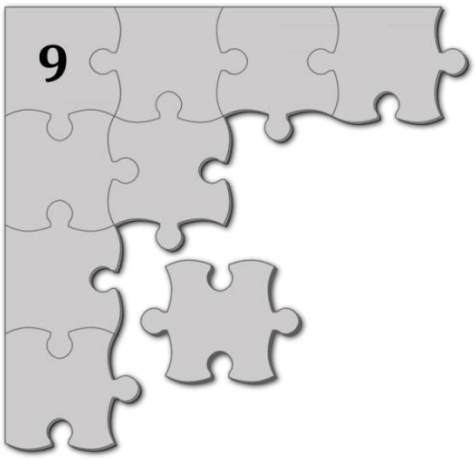


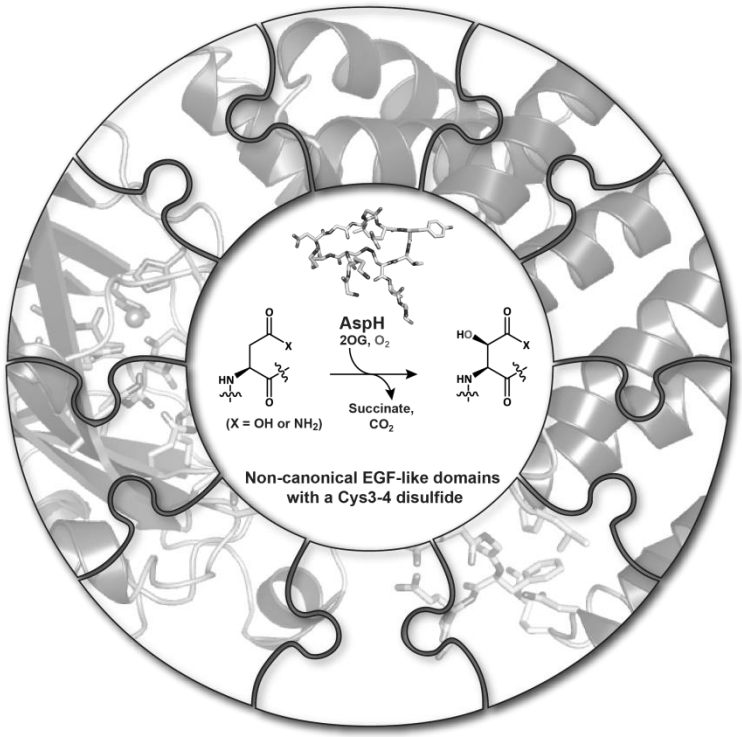
Fig. 8-13 | Sequence alignment (performed using Clustalω[364]) of **a)** AspHA = flAspH, AspHD1 and AspHD2 demonstrating that these are homologues with only ~18 % sequence identity, and **b)** AspHD1 and AspHD2 (sequence identity ~40 %). Conserved residues important for AspHA-substrate interactions, as known from the preceding studies, are marked on the alignment.

8.5 Acknowledgements for Chapter 8

Dr Joan Gannon is kindly acknowledged for training in the immunofluorescence technique, for providing some of the presented immunofluorescence data and for fruitful discussions. Many thanks also go to the researchers and staff at the Micron Oxford Advanced Bioimaging Unit, Department of Biochemistry, University of Oxford, who granted access to the instruments and provided training.



Summary,
Conclusions and Future Prospects



Abstract

This Chapter summarises the results presented in this Thesis and provides an outlook to consequent future studies.

9.1 SUMMARY AND CONCLUSIONS

Many of the apparent links between members of the 2OG oxygenase superfamily and cancer are not fully understood. However, 2OG oxygenases are tractable, 'druggable' targets. Both the inhibition and activation of 2OG oxygenases is of therapeutic interest. Selectivity for the target enzyme(s) over other 2OG oxygenases will likely be very important for the development of effective drugs. The roles of the enzymes at biochemical, cellular and physiological levels need to be assigned to understand the impact of potential drugs. Structure and function are closely linked. Protein structures can therefore provide valuable insights in functional aspects and assist in the rational design of inhibitors and probes that can be useful for cancer treatment and diagnostics.

The work described in this Thesis provides a structural and biochemical characterisation of the unusual 2OG oxygenase AspH and its substrate interactions. The results contribute to a better understanding of the biological role of AspH and provide valuable methods and tools for future work.

The structural studies presented in **Chapter 2** revealed that AspH has an unusual active site geometry with a HXG…H catalytic dyad Fe(II)-binding motif. The established assay methods demonstrated that both the oxygenase and the TPR domain are required for substrate recognition. A crystal structure of AspH in complex with its substrate showing an unexpected non-canonical Cys3-4 disulfide formed the basis for further characterisation of the AspH substrates.

The work described in **Chapter 3** investigated the substrate disulfide pattern and revealed the unanticipated AspH substrate selectivity for EGFDs with a non-canonical Cys3-4 disulfide.

Chapter 4 showed how the cyclic peptides synthesised for AspH helped support the results presented in the preceding Chapters. Some of the cyclic peptides will be useful probes for further AspH *in vitro* and cell-based studies.

The results in **Chapter 5** illustrated the highly specific AspH-substrate interactions involving the TPR domain. The mutagenesis studies revealed that a single His ligand is sufficient for Fe(II)-binding and catalysis in the case of AspH.

The work described in **Chapter 6** addressed a common problem in MS-based assays and characterised the interfering peptide species.

The results in **Chapter 7** suggested that the present Ca^{2+} levels in a RedOx environment have an influence on the disulfide pattern of EGFDs. Sufficiently high Ca^{2+} levels appeared to prevent hydroxylation by promoting the formation of the canonical EGFD disulfide pattern.

Chapter 8 provided evidence that transmembrane domain-containing AspH forms localise to the ER membrane. The reported data suggested that flAspH has similar characteristics to its isoform junctate that has been reported to be a Ca^{2+}-sensing component linked with the orai1-STIM1 machinery.[371] Further studies will be required to elucidate if enzymatically active flAspH is also involved in Ca^{2+} homeostasis in the ER with a gain of additional function as EGFD hydroxylase compared to junctate.

These results taken together suggest that enzymatically active AspH may have a role in quality control of EGFD folding. As described in **Chapter 8** (see **Section 8.4.1**), AspH may also contribute to signalling in the unfolded protein response. Proper EGFD folding is critical for functioning, as the case of POFUT1 that only attaches fucose to properly folded Notch EGFDs demonstrates (see **Section 3.3.1**).[492] Increased EGFD misfolding may occur under ER stress with prolonged Ca^{2+} depletion as in cancer, hypoxia, or inflammation;[577] EGFD folding quality control will therefore be particularly important under these conditions.

Notch and its ligands play critical roles in cell fate determination and cellular differentiation, and are strongly linked to cancer.[581,587] A 'correct' EGFD folding enabling Ca^{2+}-binding may have direct influence on the functioning of these Ca^{2+}-dependent proteins, and thus on cancer growth. Notch signalling pathway activation by a cell surface form of AspH in pancreatic cancer has recently been reported.[588] A major contribution to malignant transformation has been attributed to AspH.[588] Moreover, the authors provided evidence that small molecule inhibitors of AspH hydroxylase activity reduced pancreatic tumour growth and progression by downregulating the Notch signalling pathway.[588] Similar effects of the inhibitors have also been reported in hepatocellular carcinoma.[589] Inhibition by probe molecules is an established strategy for understanding the physiological roles of 2OG oxygenases.[318] Although the subcellular localisation studies described in this Thesis support that flAspH is localised to the ER membrane, certain forms of the protein may be transported out of the cell in cancers, as proposed in some studies.[252,588,589] It is uncertain if such an AspH form can be enzymatically active outside the cell, since the cosubstrate 2OG and free Fe(II) as cofactor are required for activity. However, the reported data[588,589] and the results in this Thesis suggest that specific AspH inhibitors can potentially be useful in cancer therapies. The structural and biochemical characterisation of $AspH_{315-758}$ in this Thesis will aid the development of specific probes or inhibitors for AspH including those based on the cyclic peptide scaffold.

9.2 FUTURE PROSPECTS

The work presented in this Thesis has mainly focused on a characterisation of the AspH compartment that contains the catalytic function. However, more work on the full-length protein will be required to elucidate the AspH function and its role in cancer (**Fig. 9-1**). Assays with the isolated protein provide useful information, which may or may not reflect *in vivo* activities. More physiologically relevant assays will be required to define the downstream pathway of the hydroxylated substrate and the potential links to Ca^{2+} homeostasis in the ER. AspH hydroxylase activity may be compromised in hypoxia; whether AspH can act as an O_2 sensor, needs further mechanistic studies. Interaction partners of flAspH need to be identified to determine the physiological role of the enzymatically active isoforms.

Whether inhibition of AspH activity can be of therapeutic use to help control aberrant Ca^{2+} signalling requires further investigation. AspH itself may have no causal role in malignancy. Its potential action as EGFD folding control, Ca^{2+} sensor and SOCE activator (see **Section 8.4.1**), however, could cultivate oncogenic pathways.

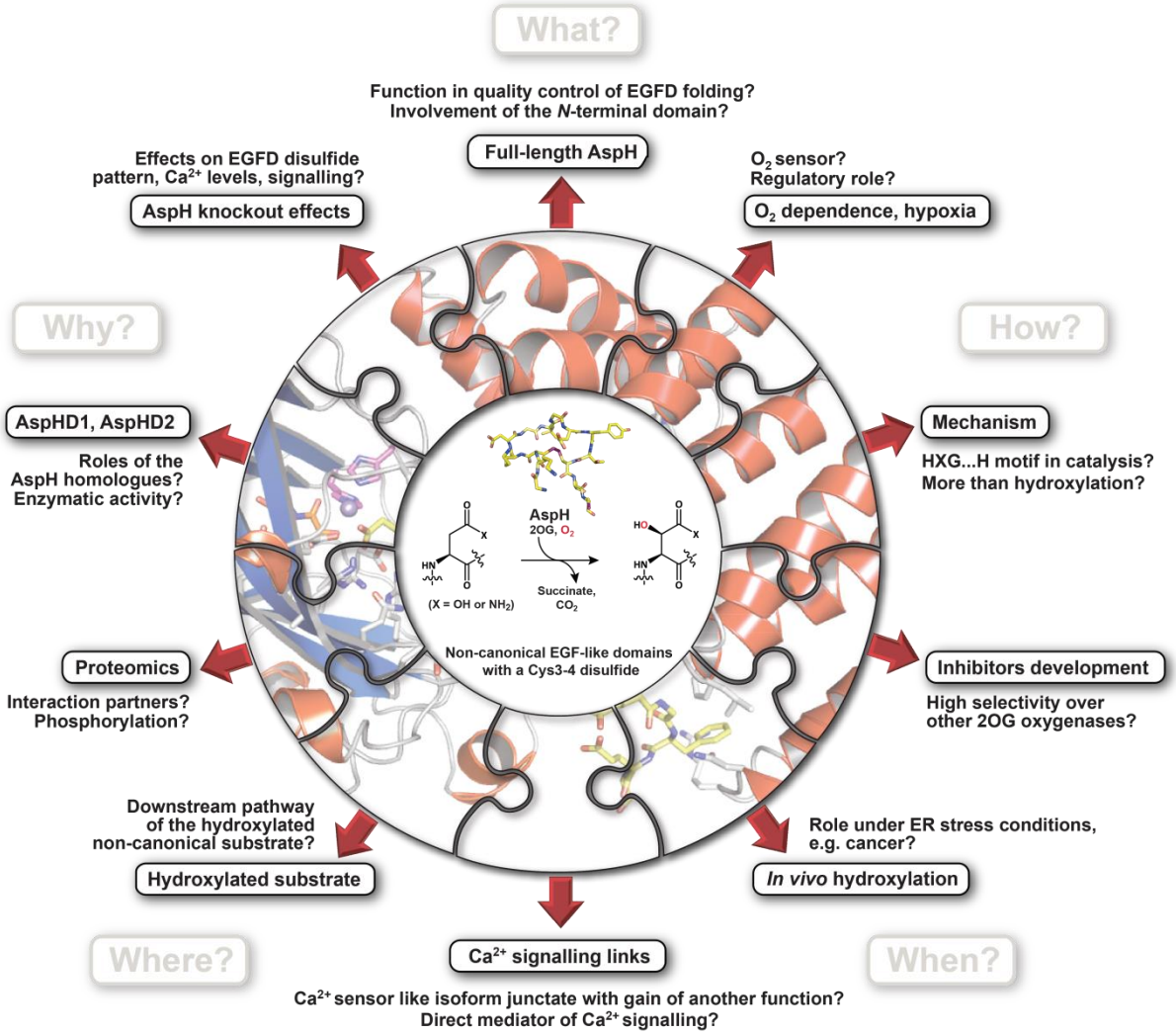
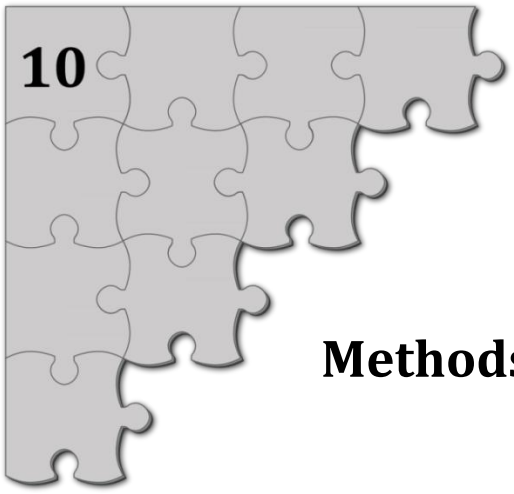


Fig. 9-1 | AspH studies future prospects: outline of some research questions arising from the work described in this Thesis.



Methods and Materials

Abstract

This Chapter lists the tools that were used to perform the studies presented in this Thesis.

10.1 General Experimental Procedures

10.1.1 Chemicals, Antibiotics, Molecular Markers

All Chemicals were obtained in the highest quality available from *Sigma-Aldrich, Fluka, Acros* or *AlfaAesar* unless stated otherwise. Tryptone, yeast extract and agar (bacteriological) for use in culture media were from *Oxoid*. *Iso*propyl-*β*-D-thiogalacto-pyranoside (IPTG), ampicillin and kanamycin were from *AppliChem*. Plasmids and restriction enzymes were from *Sigma-Aldrich* and *New England Biolabs* respectively. *PfuTurbo*® DNA polymerase and competent cells were from *Stratagene*. Perfect Protein™ Markers (10-225 kDa, each protein marker carries a His-tag sequence, which allows for use as positive control for anti-His western blots) were from *Novagen*®, and GeneRuler™ 1 kb DNA Ladder (250-10,000 bp) was from *Fermentas*. Water was purified using a *Millipore* Milli-Q system with a 0.22 µm filter on the outlet. All solvents used for LC-MS were HPLC/LC-MS grade quality (e.g. CHROMASOLV®, *Sigma-Aldrich*).

10.1.2 AspH *in vitro* Substrates and EGFD Peptides

Schematics and sequences including the disulfide links of commercially obtained peptides are given in **Fig. 10-1** and **Table 10-1**. The cyclised 39mer peptide (hFX EGF1$_{39mer}$) representing the first EGFD of human coagulation factor X EGF1 (hFX EGF1, aa 86-124) was synthesised and purified by *Peptide Synthetics* (Hampshire, UK). The three disulfide bonds were formed by spontaneous air oxidation, as in a reported procedure.[395] Details of the synthesis were not disclosed by the company. The peptide material was later found to be a mixture of likely the canonical C$_{39mer}$ with Cys1-3, 2-4, 5-6 linkages (corresponding to disulfides between C5-C16, C10-C25, and C27-C36) and the non-canonical NC$_{39mer}$ with Cys1-2, 3-4, 5-6 linkages (corresponding to disulfides between C5-C10, C16-C25, and C27-C36; **Fig. 10-1**).

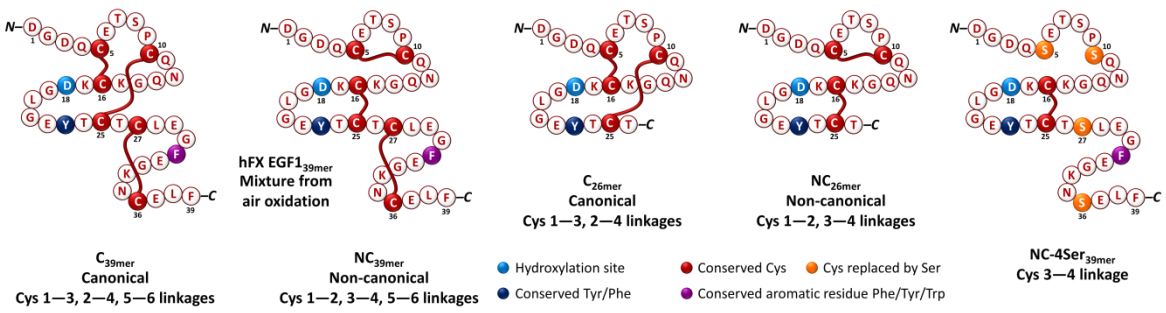


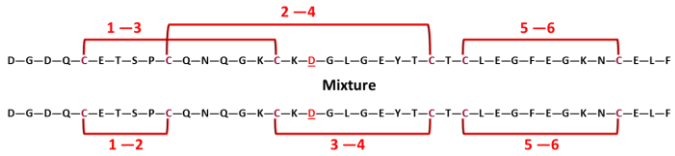
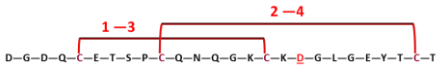
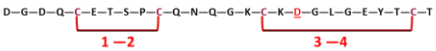
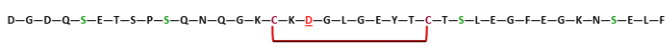
Fig. 10-1 | Schematics of commercially obtained hFX EGF1-based peptides.

The following peptides were synthesised and purified by *GL Biochem Shanghai Ltd.*: the canonical C$_{26mer}$ with Cys1-3, 2-4 linkages (corresponding to two disulfides between C5-C16 and

C10-C25), the non-canonical NC_{26mer} with Cys1-2, 3-4 linkages (corresponding to two disulfides between C5-C10 and C16-C25), and the NC-$4Ser_{39mer}$ with a single disulfide in Cys3-4 linkage between C16-C25. In the cases of the C_{26mer} and NC_{26mer} peptides, orthogonal protective groups for the cysteines were used to ensure regioselective oxidation to form the specific disulfides after solid-phase peptide synthesis, as in reported procedures.[394,468] Details of the syntheses were not disclosed by the company.

Cyclic peptides (10mer, 15mer, and 19mer) with a thioether bond were synthesised and characterised as stated below in **Section 10.4**.

Table 10-1 | Commercially obtained hFX EGF1-based peptides.

Peptide	Nr. of Cys	Supplier	Calculated mono-isotopic mass [Da]	Observed mono-isotopic mass* [Da]	Purity** (HPLC)
$hFX\ EGF1_{39mer}$ (air oxidation)	6	*Peptide Synthetics*	4,242.71	4,242.69	> 95 %
					
C_{26mer} (regioselective oxidation)	4	*GL Biochem Shanghai Ltd.*	2,775.08	2,775.05	> 95 %
					
NC_{26mer} (regioselective oxidation)	4	*GL Biochem Shanghai Ltd.*	2,775.08	2,775.06	> 95 %
					
NC-$4Ser_{39mer}$ (single disulfide)	2	*GL Biochem Shanghai Ltd.*	4,182.81	4,182.79	> 95 %
					

* The values for masses refer to the monoisotopic mass [M] determined by analysis using Q Exactive Orbitrap as described in Section **10.5.4**.

** Purity according to the supplier as determined by HPLC.

EGFD proteins were a gift from Professor Penny A. Handford, Department of Biochemistry, University of Oxford. The human Fibrillin-1 (hFib1) EGFD proteins, cbEGF32-33 (aa 2124-2205), cbEGF41-43 (aa 2567-2687), TB4cbEGF23 (aa 1527-1647) were prepared recombinantly by Dr Sacha Jensen, and human Notch-1 (hNotch1) cbEGF11-13 (aa 411-526) was produced by Dr Paul Taylor, following reported methods.[460] In summary, these constructs were cloned into the pQE-30 vector

(*Qiagen*) for production of proteins with an *N*-terminal His_6-tag coding sequence in *E. coli* strain NM554. The resultant cells were lysed under denaturing conditions in 6 M guanidine-HCl, 50 mM sodium phosphate, 5 mM 2-mercaptoethanol, pH 6.0, and the proteins were purified by Ni^{2+} affinity chromatography, followed by reduction with 100 mM DTT, 100 mM Tris, pH 8.3, and subsequent refolding *in vitro* by dialysis against 3 mM L-cysteine, 0.3 mM L-cystine, 100 mM Tris, pH 8.3, in presence of 10 mM $CaCl_2$ in order to favour formation of disulfides with the canonical stereochemistry. The pH was adjusted to pH 3.0 prior to HPLC purification.[590] Folding of the EGFD proteins with presence of canonical disulfides was then confirmed by NMR employing the reported Ca^{2+}-binding titration assay.[590]

Table 10-2 | EGFD proteins obtained from Professor Penny A. Handford's laboratory, Department of Biochemistry, University of Oxford.

Peptide	Nr. of aa	Nr. of Cys	Sequence	Calculated average mass [Da]
hFib1 cbEGF32-33	98	12	MRGSHHHHHHGSIEGRSAVDMDECKEPDVCKH GQCINTDGSYRCECPFGYTLAGNECVDTDECSVG NPCGNGTCKNVIGGFECTCEEGFEPGPMMTCE	10,575.65
hFib1 cbEGF41-43	123	18	SADVDECEGNHRCQHGCQNIIGGYRCSCPQGYL QHYQWNQCVDENECLSAHICGGASCHNTLGSYK CMCPAGFQYEQFSGGCQDINECGSAQAPCSYGC SNTEGGYLCGCPPGYFRIGQGHCV	13,210.56
hFib1 TB4cbEGF23	123	14	SAVDTRSGNCYLDIRPRGDNGDTACSNEIGVGVS KASCCCSLGKAWGTPCEMCPAVNTSEYKILCPGG EGFRPNPITVILEDIDECQELPGLCQGGKCINTFGS FQCRCPTGYYLNEDTRVCD	13,118.77
hNotch1 cbEGF11-13	135	18	SAQDVDECSLGANPCEHAGKCINTLGSFECQCLQ GYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCI CMPGYEGVHCEVNTDECASSPCLHNGRCLDKINE FQCECPTGFTGHLCQVDLHHILDAQKMVWHNR	14,826.58

10.1.3 Microbiological Techniques

Standard sterile practices were followed throughout. All prepared biological solutions and equipment such as media, plastic and glassware used in handling of microbial cultures, water, buffers and reagents for DNA applications were sterilised by autoclaving at 121 °C for 20 min using a *LTE Scientific* Falcon tabletop autoclave prior to use. All biological work was performed in a *Kendro* HERAsafe KS-12 Class II laminar flow cabinet. Solutions of heat-sensitive substances were sterilised using 0.22 µm filters (Minisart®, *Sartorius AG*).

10.1.4 Water Baths

A thermostatic Grant W14 waterbath with controller (*Grant Instruments*) was used for transformations with heat-pulsing at 42 °C and Grant Y14 for incubations at 37 °C.

10.1.5 pH Measurement

A *Hanna Instruments* HI 9321 microprocessor pH meter fitted with a 5 mm diameter electrode was used for determination of the pH of solutions. Prior to each use, the probe was calibrated in the range pH 4-7 or pH 7-10, appropriate to the required pH, using standard calibration solutions (*Fisher Scientific*). Between uses, the electrode was stored in 4 M KCl solution. The pH of solutions was adjusted using stock solutions of NaOH or HCl as required.

10.1.6 Bacterial Growth Media

For cultivation of *E. coli* strains, 2TY medium or LB medium was used. Before autoclaving, the pH of the media was adjusted to 7.0. For growth on agar plates, Agar No 1 was added to LB medium to a concentration of 1.5 % (w/v) prior to autoclaving (121 °C, 1.5 bar, 20 min). Antibiotics for selection were added to the autoclaved medium (temperature below 55 °C) at the following concentrations: ampicillin 100 µg/ml, kanamycin 30 µg/ml, chloramphenicol 33 µg/ml. Sterile SOC medium was used for transformations (2TY medium supplemented with 10 mM $MgCl_2$, 20 mM glucose). 2TY medium for large-scale growths was prepared by Mr Herminio Manso-Jubier using a TouchClave® II autoclave (*LTE Scientific*).

Table 10-3 | Bacterial growth media; **a)** 2TY medium, **b)** LB medium.

a)

2TY medium (1 l)	
Tryptone	16 g
Yeast extract	10 g
Sodium chloride	5 g

b)

LB medium (1 l)	
Tryptone	10 g
Yeast extract	5 g
Sodium chloride	10 g

10.1.7 Centrifugation and Incubation

Centrifugation was usually carried out at 4 °C using a *Beckman* Avanti J25 centrifuge using a JLA-9.1, a JLA-16.250 or a JA-25.50 rotor at 7,000 rpm, 14,000 rpm or 24,000 rpm respectively, depending on the size of the sample. Samples with a volume less than 1.5 ml were centrifuged at room temperature using a benchtop CENTRISART® A-14 microcentrifuge (*Sartorius Stedim Biotech*) at 14,800 rpm. For 50 ml falcon tubes, a *Beckman* Allegra 21R bench top centrifuge was used with a S-4180 rotor at 4,000 rpm and 4 °C.

For incubation, *New Brunswick Scientific* G25 Environmental Incubator shakers were used at the appropriate temperatures and shaking at 180-200 rpm.

10.1.8 Optical Density Measurements (OD_{600})

Optical density at 600 nm (OD_{600}) was measured using a 1 ml sample of bacterial culture, compared to 1 ml reference solution of initial growth medium using a Novaspec II spectrophotometer *(Pharmacia)*.

10.1.9 Lyophilisation

Samples were flash-frozen in liquid N_2 and then lyophilised under vacuum using a *Thermo Scientific* Micro Modulyo® Freeze Dryer.

10.2 MOLECULAR BIOLOGY TECHNIQUES

10.2.1 Vectors

Table 10-4 | Vectors used for cloning of the indicated constructs for recombinant protein production.

Vector	Tag	Tag cleavage site	Antibiotic resistance	Promotor	Recombinant proteins
pET-28a(+)	*N*-terminal His_6-tag	Thrombin	Kanamycin	T7	$AspH_{315-758}$ and variants
pGEX-6P-1	*N*-terminal GST-tag	3C Protease	Ampicillin	tac	GST-$AspH_{315-758}$
pNIC28-Bsa4 (LIC)	*N*-terminal His_6-tag	TEV	Kanamycin	T7	$AspH_{562-758}$, $flAspH_{1-758}$, $TMtAspH_{76-758}$, $TMtNAspH_{76-314}$, TMtN-TPR-$AspH_{76-561}$
pcDNA3-*N*-FLAG-LIC	*N*-terminal FLAG-tag	-	Ampicillin	CMV	$flAspH_{1-758}$, Humbug ($AspH_{1-314}$), $AspH_{315-758}$
pcDNA3-*C*-FLAG-LIC	*C*-terminal FLAG-tag	-	Ampicillin	CMV	$flAspH_{1-758}$, Humbug ($AspH_{1-314}$), $AspH_{315-758}$
pcDNA3-*N*-GFP-LIC	*N*-terminal GFP-tag	-	Ampicillin	CMV	$flAspH_{1-758}$
pcDNA3-*C*-GFP-LIC	*C*-terminal GFP-tag	-	Ampicillin	CMV	$flAspH_{1-758}$

10.2.2 Preparation of Oligonucleotides

Oligonucleotides were synthesised to order by *Sigma-Aldrich* and were obtained as lyophilised, desalted pellets on a 0.025 µmol scale. Before use, the oligonucleotides were resuspended in sterile Milli-Q water to give a 100 µM stock solution, which was then diluted to 10 µM with sterile Milli-Q water for use in PCR experiments.

10.2.3 DNA Amplification by Polymerase Chain Reactions

Polymerase chain reactions (PCRs) were performed in thin-walled 0.5 ml Eppendorf tubes (STARLAB) using a Genius Thermal Cycler (*Techne*) with a preheated lid. The thermal cycling conditions are given below (**Table 10-7**).

The sequence encoding $AspH_{315\text{-}758}$ was subcloned from the pGEX-6p-1 vector into the pET-28a(+) vector (*Novagen*) between *BamH*I and *Not*I restriction sites for production of the protein with an *N*-terminal His_6-tag coding sequence, and this construct was used as template for mutagenesis experiments. FLAG-tagged AspH constructs were prepared by LIC cloning as described below (**Section 10.2.10**).

10.2.4 Site-directed Mutagenesis

Single site-directed mutageneses (for making point mutations or enabling deletion of single amino acids) were carried out using *Stratagene* QuikChange® site-directed mutagenesis kit according to the manufacturer's instructions. Primers (forward and reverse) incorporating the desired mutation were used to amplify the sequence in opposite directions by PCR. In all PCR experiments, high fidelity *PfuTurbo*® DNA polymerase (*Stratagene*) was used, which allows for proofreading of misincorporated nucleotides. A typical PCR reaction set-up and the thermal cycling conditions are given below; the final volume was maintained at 50 µl. Primers used for site-directed mutageneses of $AspH_{315\text{-}758}$ template are given in **Table 10-6**.

Table 10-5 | Set-up of a typical PCR reaction for site-directed mutagenesis.

Component		Amount for 50 µl reaction
dsDNA template	(10-20 ng/µl)	1 µl
Forward primer	(10 µM)	2 µl
Reverse primer	(10 µM)	2 µl
10 x Reaction buffer		5 µl
dNTP Mix	(2 mM)	5 µl
Sterile Milli-Q H_2O		34 µl
PfuTurbo® DNA polymerase	(2.5 U/µl)	1 µl

Table 10-6 | Oligonucleotide primers used for site-directed mutageneses of wildtype $AspH_{315-758}$.

Variant primer name		Sequence
R735A	*Fwd*	5’-GGATGCCTCATCTTTC**GC**GCTGATATTCATCGTG-3’
	Rev	5’-CACGATGAATATCAGC**GC**GAAAGATGAGGCATCC-3’
R735W	*Fwd*	5'-GGATGCCTCATCTTTC**T**GGCTGATATTCATCGTG-3'
	Rev	5'-CACGATGAATATCAGCC**A**GAAAGATGAGGCATCC-3'
H679A	*Fwd*	5’-CTCACGTGTGGCCG**GC**CACAGGGCCC-3’
	Rev	5’-GGGCCCTGTG**GC**CGGCCACACGTGAG-3’
H679R	*Fwd*	5'-CTCACGTGTGGCCG**AGG**ACAGGGCCC-3'
	Rev	5'-GGGCCCTGT**CCT**CGGCCACACGTGAG-3'
H725A	*Fwd*	5’-GATGACTCCTTTGAG**GC**CGAGGTATGGCAGG-3’
	Rev	5’-CCTGCCATACCTCG**GC**CTCAAAGGAGTCATC-3’
H690A	*Fwd*	5’-CAGGCTCCGAATG**GC**CCTGGGCTTGGTG-3’
	Rev	5’-CACCAAGCCCAGG**GC**CATTCGGAGCCTG-3’
R686A	*Fwd*	5’-CCCACAAACTGC**GC**GCTCCGAATGCAC-3’
	Rev	5’-GTGCATTCGGAGC**GC**GCAGTTTGTGGG-3’
R688A	*Fwd*	5’-CACAAACTGCAGGCTC**GCT**ATGCACCTGGGCTTG-3’
	Rev	5’-CAAGCCCAGGTGCAT**AGC**GAGCCTGCAGTTTGTG-3’
E617A	*Fwd*	5’-CTGCCTGAGGATG**C**AAACCTGAGGGAAAAAG-3’
	Rev	5’-CTTTTTCCCTCAG**G**TTTGCATCCTCAGGCAG-3’
D721A	*Fwd*	5’-GCTCATCTTTGATG**C**CTCCTTTGAGCACG-3’
	Rev	5’-CGTGCTCAAAGGAG**G**CATCAAAGATGAGC-3’
K666A	*Fwd*	5'-CAGAAGAGGACAGATC**GC**ATATTCCATCATGCACC-3'
	Rev	5'-GGTGCATGATGGAATAT**GC**GATCTGTCCTCTTCTG-3'
Q627A	*Fwd*	5'-GAAAAAGGGGACTGGAGC**GCA**TTCACGCTGTGGCAGCAAG-3'
	Rev	5'-CTTGCTGCCACAGCGTGAA**TGC**GCTCCAGTCCCCTTTTTC-3'
H493A	*Fwd*	5'-GATGGCTTTGCTAAAGTC**GCA**TATGGCTTCATCCTGAAG-3'
	Rev	5'-CTTCAGGATGAAGCCATA**TGC**GACTTTAGCAAAGCCATC-3'
H530A	*Fwd*	5'-GGGAGATTTTATTTC**GC**CCTGGGGGATGCCATG-3'
	Rev	5'-CATGGCATCCCCCAGG**GC**GAAATAAAATCTCCC-3'
R526A	*Fwd*	5'-CTGGCACTGATGATGGG**GC**ATTTTATTTCCACCTG-3'
	Rev	5'-CAGGTGGAAATAAAAT**GC**CCCATCATCAGTGCCAG-3'
Y565A	*Fwd*	5'-GGCAACGCTCACTC**GC**CAATGTGAATGGAC-3'
	Rev	5'-GTCCATTCACATTG**GC**GAGTGAGCGTTGCC-3'
F432A	*Fwd*	5'-CTCAGACAGGCAACAA**GC**TCTAGGTCATATGAGAG-3'
	Rev	5'-CTCTCATATGACCTAGA**GC**TTGTTGCCTGTCTGAG-3'
F529A	*Fwd*	5'-GATGATGGGAGATTTTAT**GCG**CACCTGGGGGATGCCATG-3'
	Rev	5'-CATGGCATCCCCCAGGTG**CGC**ATAAAATCTCCCATCATC-3'
G681D	*Fwd*	5'-GTGGCCGCACACAG**AT**CCCACAAACTGCAG-3'
	Rev	5'-CTGCAGTTTGTGGG**AT**CTGTGTGCGGCCAC-3'
G681E	*Fwd*	5'-GTGGCCGCACACAG**AA**CCCACAAACTGCAG-3'
	Rev	5'-CTGCAGTTTGTGGG**TT**CTGTGTGCGGCCAC-3'
I758del	*Fwd*	5'-CGCAGCCTTCCAGCA**TGA**TAGTAAGCGGC-3'
	Rev	5'-GCCGCTTACTA**TCA**TGCTGGAAGGCTGCG-3'
A644C	*Fwd*	5'-GAAAATGCCTGCAAAGGA**TGC**CCTAAAACCTGTACCTTAC-3'
	Rev	5'-GTAAGGTACAGGTTTTAGG**GCA**TCCTTTGCAGGCATTTTC-3'
C641A	*Fwd*	5'-GAAGAAGAAATGAAAATGCC**GCG**AAAGGAGCTCCTAAAACCTG-3'
	Rev	5'-CAGGTTTTAGGAGCTCCTTT**CGC**GGCATTTTCATTTCTTCTTC-3'
C648A	*Fwd*	5'-CAAAGGAGCTCCTAAAACC**GCG**ACCTTACTAGAAAAGTTC-3'
	Rev	5'-GAACTTTTCTAGTAAGGT**CGC**GGTTTTAGGAGCTCCTTTG-3'
N395A	*Fwd*	5'-CTGAGAAGAGGAGAAGT**GCG**GAGGTGCTACGTGGAGC-3'
	Rev	5'-GCTCCACGTAGCACCTC**CGC**ACTTCTCCTCTTCTCAG-3'
T578P*	*Fwd*	5'-CACAGCCTTGGTGGACC**C**C**G**AAAGAAACGGGCTACAC-3'
	Rev	5'-GTGTAGCCCGTTTCTTT**C**G**G**GGTCCACCAAGGCTGTG-3'

*R526A mutation repeatedly induced a second mutation (P578T), which was mutated back using R526A as template.

Table 10-7 | PCR thermal cycling conditions used for site-directed mutageneses.

Function	Temperature	Duration	Number of cycles
Denaturation of template DNA	95 °C	1 min	1
Denaturation	95 °C	30 s	16 (point mutation)
Annealing	55 °C	1 min	
Extension	68 °C	1 min/kb	
Final extension	68 °C	10 min	1

10.2.5 *Dpn*I Digestion of the PCR Products

Following PCR, the PCR mixtures containing mutated plasmid and parental template DNA were treated with 1 µl of the restriction endonuclease *Dpn*I[591] from the *Stratagene* QuikChange® site-directed mutagenesis kit for 3 h at 37 °C according to manfacturer's instructions. Since the template DNA was extracted from a dam+ *E. coli* strain and therefore methylated, in contrast to the mutated PCR product, *Dpn*I that is specific for methylated and hemimethylated DNA was used to digest the parental DNA. The mutated DNA was then transformed into XL10-Gold ultracompetent cells for growth and subsequent plasmid DNA extraction (as described below).

10.2.6 DNA Gel Electrophoresis

DNA samples were analysed by gel electrophoresis on agarose gels (typically 1 %), which were prepared by boiling 0.4 g agarose (*Bioline*) in 40 ml Tris-acetate-EDTA (TAE) buffer using a SuperWave 750 microwave (*Tecnolec*) until a clear solution was obtained. The mixture was allowed to cool to approximately 55 °C prior to addition of 6 µl SYBR Safe™ DNA gel stain (*Invitrogen*). Gels were cast in an H2-SET gel tray (*Anachem*) to a depth of approximately 0.7 cm, and run on a *Bio-RAD* system at a constant potential of 100 V for typically 20-35 min until sufficient fragment separation was achieved. Bands were then visualised by illumination of the gel with UV light using a Gel Logic 200 imaging system (*Kodak*) with a 535 nm emission filter. Samples were loaded after mixing with 6x sample loading buffer and run in TAE buffer. The size of fragments was referenced to *Fermentas* GeneRuler™ 1 kb DNA ladder (250-10,000 bp), allowing for molecular weight determination.

Table 10-8 | DNA gel electrophoresis buffers; **a)** TAE running buffer, **b)** DNA sample loading buffer.

a)

50 x TAE (running buffer solution)
2 M Tris
1 M acetic acid (glacial)
0.5 mM EDTA
Adjust to pH 8.0

b)

6 x Sample loading buffer
70 % glycerol
30 % 1 x TAE
(25 % Bromophenol Blue, if appropriate)

10.2.7 Plasmid DNA Preparation and Gel Extraction

For plasmid DNA preparation a QIAprep Spin Miniprep Kit and QIAGEN Plasmid Maxi Kit (*QIAGEN*) or GeneJET Plasmid Miniprep Kit and GeneJET Plasmid Maxiprep Kit (*Fermentas*) were used. Typically 5 ml (for Miniprep) or 300 ml (for Maxiprep) 2TY medium were inoculated with a single colony or a glycerol stock and grown overnight. Portions of 3 ml or 250 ml respectively of the culture were transferred into an appropriate tube and centrifuged. The supernatant was discarded; the resulting pellet was used for plasmid purification following the kit manufacturer's protocols. DNA elution for Miniprep kits was typically performed using 30 µl sterile H_2O; for Maxiprep kits the elution buffer provided in the kit was used.

For gel extraction, bands were excised from agarose gels using a clean scalpel, weighed, and DNA extracted using *QIAGEN* or *Fermentas* Gel Extraction kit, following the manufacturer's protocol. DNA was eluted in 30 µl sterile H_2O.

DNA concentrations were determined by measurement of absorbance at 260 nm using a NanoDrop ND-1000 UV/vis spectrophotometer (*Thermo Fisher Scientific*).

10.2.8 DNA Restriction Digest and Subcloning

DNA digests were carried out using the appropriate restriction enzymes (*New England Biolabs*) in the recommended buffers and incubation at 37 °C (water bath) for 1-3 h. A typical reaction volume was 20 µl, containing 2-10 U enzyme and BSA added as necessary. Subcloning of constructs between vectors with compatible restriction sites was achieved by restriction digest of purified plasmid DNA templates for both vector and insert.

Table 10-9 | Typical set-up for DNA restriction digest.

For insert	
DNA template	2 µl
10 x *NEB* buffer	2 µl
Restriction enzyme A	1 µl
Restriction enzyme B	1 µl
Sterile Milli-Q H_2O	14 µl

For vector	
Vector template	2 µl
10 x *NEB* buffer	2 µl
Restriction enzyme A	1 µl
Restriction enzyme B	1 µl
Sterile Milli-Q H_2O	14 µl

10.2.9 DNA Ligation

Restriction digestion products were run on an agarose gel and purified by gel extraction. Insert and vector DNA with complementary overhanging ends were ligated at room temperature for 1 h, followed by incubation at 4 °C overnight and subsequent transformation into XL10-Gold ultracompetent cells.

Table 10-10 | A typical set-up for DNA ligation mixture with a molar ratio of 1:3 vector to insert.

Component	Amount for 20 µl ligation
Doubly digested insert DNA (3 kb)	~50 ng (0.025 pmol)
Doubly digested vector DNA (1 kb)	~50 ng (0.076 pmol)
10 x T4 ligase buffer	2 µl
T4 DNA ligase	1 µl
Sterile Milli-Q H_2O	to final volume of 20 µl

10.2.10 Ligation-Independent Cloning (LIC)

Ligation-independent cloning[388,549] (LIC) vectors (as listed in **Table 10-4**) were a gift from Dr Grazyna Kochan, Structural Genomics Consortium Oxford. Primers (forward and reverse) with appropriate LIC sequences at either end were used to amplify the desired DNA sequence in opposite directions by PCR. Phusion® High-Fidelity DNA Polymerase (*New England Biolabs*) was used for PCR with either Phusion® HF or Phusion® GC Buffer. A typical PCR reaction set-up and thermal cycling conditions are given below in **Tables 10-11** and **10-12**; the final volume was maintained at 50 µl. LIC vectors were digested with *Bsa*I for 3 h at 50 °C for linearisation. The primers used for LIC cloning are given in **Table 10-13**.

Table 10-11 | Reaction set-up for **a)** PCR for LIC cloning, **b)** *Bsa*I digest of LIC vectors.

a)

Component	Amount for 50 µl reaction
DNA template (25 ng/µl)	1 µl
5 x Phusion® HF or GC Buffer	10 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
dNTP Mix (10 mM)	1 µl
Sterile Milli-Q H_2O	35.5 µl
Phusion® DNA Polymerase	0.5 µl

b)

For LIC vector digest (50 µl)	
Vector template	2.5 µg
10 x *NEB* buffer 3	5 µl
BSA	0.5 µl
*Bsa*I	1.5 µl
Sterile Milli-Q H_2O	To 50 µl

Table 10-12 | Typical PCR thermal cycling conditions used for LIC cloning.

Function	Temperature	Duration	Number of cycles
Denaturation of template DNA	98 °C	1 min	1
Denaturation	98 °C	10 s	30
Annealing	5 °C below primer T_m	30 s	
Extension	72 °C	30 s-1 min/kb	
Final extension	72 °C	10 min	1

Table 10-13 | Oligonucleotide primers used LIC cloning.

Construct	Primer name		Sequence
***N*-FLAG-flAsph$_{1-758}$, His$_6$-flAsph$_{1-758}$, *N*-GFP-flAsph$_{1-758}$**	**LIC_AspH**	*Fwd*	5'-**TACTTCCAATCCATG**GCCCAGCGTAAG-3'
		Rev	5'-**TATCCACCTTTACTGCTA**ATTGCTGGAAGGCTG-3'
Middle primers for sequencing of flAspH	**Mid_AspH**	*Fwd*	5'-GTGTGAGGATGATTTGGCTG-3'
		Rev	5'-CAAATCATCCTCACACTGCG-3'
flAsph$_{1-758}$-*C*-FLAG, flAsph$_{1-758}$-*C*-GFP	**LIC_AspH**	*Fwd*	5'-**TACTTCCAATCCATG**GCCCAGCGTAAG-3'
	CLIC_AspH	*Rev*	5'-**TATCCACCTTTACTGGAAAT**TGCTGGAAGGCTG-3'
***N*-FLAG-AspH$_{315-758}$, *N*-GFP-AspH$_{315-758}$**	**LIC_tAspH**	*Fwd*	5'-**TACTTCCAATCCATG**CGTAAAACAGATGATCC-3'
		Rev	5'-**TATCCACCTTTACTGTCA**AATTGCTGGAAGGC-3'
AspH$_{315-758}$-*C*-FLAG, AspH$_{315-758}$-*C*-GFP	**LIC_tAspH**	*Fwd*	5'-**TACTTCCAATCCATG**CGTAAAACAGATGATCC-3'
	CLIC_AspH	*Rev*	5'-**TATCCACCTTTACTGGAAAT**TGCTGGAAGGCTG-3'
***N*-FLAG-Humbug, *N*-GFP-Humbug (AspH$_{1-314}$)**	**LIC_AspH**	*Fwd*	5'-**TACTTCCAATCCATG**GCCCAGCGTAAG-3'
	LIC_tNAspH	*Rev*	5'-**TATCCACCTTTACTGTCA**ATTTGTTTCTGGTGGTAC-3'
Humbug-*C*-FLAG, Humbug-*C*-GFP (AspH$_{1-314}$)	**LIC_AspH**	*Fwd*	5'-**TACTTCCAATCCATG**GCCCAGCGTAAG-3'
	CLIC_tNAspH	*Rev*	5'-**TATCCACCTTTACTGGAATT**TGTTTCTGGTGGTACTTC-3'
***N*-FLAG-Junctate, *N*-GFP-Junctate**	**LIC_Junc**	*Fwd*	5'-**TACTTCCAATCCATG**GCTGAAGATAAAGAGAC-3'
	LIC_Junc	*Rev*	5'-**TATCCACCTTTACTGTTA**AGTATCTGGTGGTACTTCC-3'
Junctate-*C*-FLAG, Junctate-*C*-GFP	**LIC_Junc**	*Fwd*	5'-**TACTTCCAATCCATG**GCTGAAGATAAAGAGAC-3'
	CLIC_Junc	*Rev*	5'-**TATCCACCTTTACTGGAAGT**ATCTGGTGGTACTTCCT-3'
***N*-FLAG-Junctin, *N*-GFP-Junctin**	**LIC_JNT**	*Fwd*	5'-**TACTTCCAATCCATG**GCTGAAGATAAAGAGAC-3'
	LIC_JNT	*Rev*	5'-**TATCCACCTTTACTGTTA**GCCGTTTCTTTTCTG-3'
Junctin-*C*-FLAG, Junctin-*C*-GFP	**LIC_JNT**	*Fwd*	5'-**TACTTCCAATCCATG**GCTGAAGATAAAGAGAC-3'
	CLIC_JNT	*Rev*	5'-**TATCCACCTTTACTGGAGCC**GTTTCTTTTCTGG-3'
His$_6$-TMtAspH$_{76-758}$	**LIC_TMtAspH**	*Fwd*	5'-**TACTTCCAATCCATG**GATCTTGTTGACTATGAGG-3'
	LIC_AspH	*Rev*	5'-**TATCCACCTTTACTGCTA**AATTGCTGGAAGGCTG-3'
His$_6$-TMtNAspH$_{76-314}$	**LIC_TMtAspH**	*Fwd*	5'-**TACTTCCAATCCATG**GATCTTGTTGACTATGAGG-3'
	LIC_tNAspH	*Rev*	5'-**TATCCACCTTTACTGTCA**ATTTGTTTCTGGTGGTAC-3'
His$_6$-TMtN-TPR-AspH$_{76-561}$	**LIC_TMtAspH**	*Fwd*	5'-**TACTTCCAATCCATG**GATCTTGTTGACTATGAGG-3'
	LIC_TMtNTPR	*Rev*	5'-**TATCCACCTTTACTGCTA**TTGCCAGACAGATGCA-3'

Table 10-14 | Treatment with T4 DNA Polymerase as preparation for the LIC reaction.

For insert	
PCR product	5 µl
Sterile Milli-Q H_2O	1.25 µl
10 x *NEB* buffer 2	1 µl
10 x BSA	1 µl
DTT (100 mM)	0.5 µl
dCTP (10 mM)	1 µl
T4 DNA Polymerase	0.25 µl

For vector	
*Bsa*I-digested vector	5 µl
Sterile Milli-Q H_2O	1.25 µl
10 x *NEB* buffer 2	1 µl
10 x BSA	1 µl
DTT (100 mM)	0.5 µl
dGTP (10 mM)	1 µl
T4 DNA Polymerase	0.25 µl

The PCR products and vectors were then analysed by agarose gel (1 %) electrophoresis and purified by gel extraction, followed by treatment with T4 DNA Polymerase (*New England Biolabs*;

Table 10-4) for 30 min at 22 °C to generate sticky ends, then 20 min at 75 °C to heat-inactivate the enzyme. For the LIC reaction, 2 µl insert and 1 µl vector were incubated for 10 min at 22 °C, and then transformed into XL10-Gold ultracompetent cells.

10.2.11 DNA Sequencing

Sequencing of plasmids was carried out by the Source Bioscience DNA Sequencing Service Facility with collection point in the Department of Biochemistry, University of Oxford.

10.2.12 Competent Cell Strains

E. coli BL21 (DE3), Rosetta 2 (DE3), and Origami 2 (DE3) were used for protein production, and *E. coli* XL10-Gold (both *Stratagene*) for transformation of ligation and mutagenesis products. Genotypes of these strains are given in **Table 10-15**.

Table 10-15 | Genotypes of *E. coli* strains used in this work.

E. coli Strain	Genotype	Supplier
BL21 (DE3)	*E. coli* B F^- *dcm ompT hsd*S(r^-m^-) *gal* λ(DE3)	*Stratagene*
XL10-Gold	Tet^R Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte* [F′ *proAB* $lacI^q$Z Δ*M15* Tn*10* (Tet^R) Amy Cam^R]	*Stratagene*
Rosetta 2 (DE3)	F^- *ompT* $hsdS_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)	*Novagen*
Origami 2 (DE3)	*Δ(ara-leu)7697 ΔlacX74 ΔphoA Pvu*II *phoR araD139 ahpC galE galK rpsL* F′[lac^+ $lacI^q$ *pro*] (DE3) *gor522*::Tn*10 trxB* (Str^R, Tet^R)	*Novagen*

10.2.13 Competent Cell Preparation

Aliquots of competent *E. coli* BL21 (DE3) were prepared in antibiotic-free media from commercially obtained stocks (*Stratagene*), which were streaked onto LB agar plates. Colonies were then grown overnight. A single colony was used to inoculate 100 ml sterile 2TY medium and incubated at 37 °C with shaking (220 rpm) overnight. A portion of 800 µl of this starter culture was used to inoculate 400 ml sterile 2TY medium and then grown at 37°C with shaking to an OD_{600} of 0.6-0.8. The cells were harvested by centrifugation (750 x g, 4 °C, 5 min) in pre-chilled centrifuge tubes. Cell pellets were gently resuspended in 200 ml ice-cold 100 mM $CaCl_2$/20 % glycerol, and the suspension was incubated on ice for 20 min. Cells were pelleted by centrifugation (750 x g, 4 °C, 5 min) and gently resuspended in 8 ml 100 mM $CaCl_2$/20 % glycerol. Cells were stored as 100 µl aliquots in freshly sterilised, pre-chilled 0.5 ml Eppendorf tubes, then immediately stored at -80 °C.

10.2.14 Transformations

Competent cells were transformed according to the supplier's protocol (*Stratagene* or *Novagen* as stated in **Table10-15**). After thawing of the competent cells on ice, typically 20-100 ng (~0.5 µl-2 µl) plasmid DNA was added to 10-30 µl *E. coli* BL21 (DE3) and incubated on ice for 30 min before heat-pulsing at 42 °C for 45 s and returning to ice for 2 min. 300-450 µl sterile 2TY or SOC medium was added to the cells, which were then incubated at 37 °C with shaking (300 rpm) for 1 h.

To 15-50 µl (35 µl for site-directed mutagenesis products) *E. coli* XL10-Gold ultracompetent cells first 2 µl 2–mercaptoethanol was added and incubated on ice for 10 min, followed by addition of typically 10-100 ng plasmid DNA, or 1 µl of a ligation mixture, or 2 µl of a PCR product from site-directed mutagenesis, and incubation on ice for 30 min. The cells were then subjected to a heat-pulse for 30 s at 42 °C and subsequently chilled on ice for 2 min, followed by addition of 200-450 µl (200 µl for PCR products from site-directed mutagenesis) sterile 2TY or SOC medium and incubation at 37 °C with shaking (300 rpm) for 1 h.

Typically 100 µl of the transformation mixture was plated onto an agar plate containing the appropriate antibiotic for selection and incubated at overnight at 37 °C.

10.2.15 Glycerol Stocks

A cell pellet obtained from 1.0 ml of an overnight starter culture was resuspended in 400 µl sterile 2TY medium, followed by addition of 400 µl sterile glycerol stock solution (65 % glycerol, 0.1 M $MgSO_4$, 25 mM Tris, pH 7.5) and storage at -80 °C.

10.3 PROTEIN PRODUCTION

10.3.1 Starter Culture Preparation and Large-Scale Growths

As starter culture, 100 ml 2TY medium (in 500 ml Duran® flasks, *Schott UK*) containing the appropriate antibiotic for selection was inoculated with a single colony from an agar plate or from a glycerol stock and grown overnight at 37 °C with shaking (220 rpm). For large-scale growths 12 flasks each containing 600 ml 2TY media (in 2,000 ml *PYREX*® narrow-mouth graduated Erlenmeyer flasks) and the appropriate antibiotic for selection were inoculated with 6 ml starter culture and grown at 37 °C with shaking (180 rpm) until an OD_{600} of 0.6-0.8 (or OD_{600} of 1.1-1.3 for subsequent incubations at 18 °C overnight) was reached. The appropriate concentration of *iso*propyl-β-D-thiogalacto-pyranoside (IPTG) was then added to induce expression. Cultures were further incubated under the optimal conditions (determined by expression trials) and then harvested by centrifugation (typically using a JLA-9.1 or JA-10 rotor at 7,000 rpm, 4 °C, 8 min). The resulting pellet was stored at -80 °C.

Table 10-16 | Expression conditions for production of recombinant proteins.

Protein Construct	Temperature after induction [°C]	Duration after induction [h]	IPTG concentration [µM]	OD_{600} at point of induction
His_6-$AspH_{315-758}$	18/28/37	16/16/5	0.1/1.0/1.0	1.1-1.3/0.6-0.8
His_6-$AspH_{315-758}$ variants	18	16	0.1	1.1-1.3
GST-$AspH_{315-758}$	37	4	1.0	0.6-0.8
His_6-$AspH_{562-758}$	18	16	0.1	0.9-1.2
His_6-fl$AspH_{1-758}$	No expression observed (in *E. coli* BL21 (DE3), Rosetta 2 (DE3), Origami 2(DE3))			
His_6-TMt$AspH_{76-758}$	No expression observed (in *E. coli* BL21 (DE3), Rosetta 2 (DE3), Origami 2(DE3))			
His_6-TMtN$AspH_{76-314}$	No expression observed (in *E. coli* BL21 (DE3))			
His_6-TMtN-TPR-$AspH_{76-561}$	No expression observed (in *E. coli* BL21 (DE3))			

10.3.2 Expression Trials for Small-Scale Protein Production

For expression trials, flasks containing 100 ml sterile 2TY media and the appropriate antibiotic for selection were inoculated with 1 ml of a starter culture (prepared as described above) and grown until induction with 0.1 mM, 0.5 mM or 1.0 mM *iso*propyl-*β*-D-thiogalacto-pyranoside (IPTG), then incubated at typically 18 °C, 28 °C and 37 °C with shaking (220 rpm) for 3-4 h, and 50 ml of each culture harvested by centrifugation in 50 ml falcon tubes (4,000 rpm, 10 min, 4 °C). Resulting pellets were stored at -80 °C. Incubation of the remaining cultures was retained at the appropriate temperatures for further 12 h to achieve an incubation time of in total ~16 h. Cells were harvested as described above. Cell pellets were resuspended in 5 ml lysis buffer for the appropriate purification method (see below), then lysed by sonication on ice with three 3 x 30 s pulses, each separated by a 30 s pause, using a *MSE* Soniprep 150 sonicator at an amplitude of 10 µm or a *Sonics* Vibra-Cell™ VCX-500 sonicator with 13 mm probe at 40 % intensity for 6 x 15 s bursts, each separated by 15 s breaks. A volume of 1 ml of the cell lysate was then transferred to a 1.5 ml Eppendorf tube and centrifuged (9,000 rpm, 2 min, 25 °C). Soluble protein was removed, and the remaining insoluble pellet resuspended in 1 ml lysis buffer. Typically 10 µl of the protein samples were analysed by SDS-PAGE, revealing conditions resulting in greatest production of soluble protein. Additionally, an uninduced (-IPTG) sample used as negative control was analysed for comparison.

10.3.3 Large-Scale Cell Lysate Preparation

Cell pellets obtained from large-scale growths were resuspended in approximately 4-5 times their mass (w/v) of the appropriate lysis buffer with addition of approximately 0.2 mg *DNase*I (bovine

pancreas, grade II, *Roche*), $MgCl_2$ (of a 1 M stock to a final concentration of 10 mM Mg^{2+} in the suspension), and *Roche* complete EDTA-free protease inhibitor cocktail tablets (1 tablet/50 ml suspension). The mixture was stirred at 4 °C until homogenous, and the cells were then lysed on ice by sonication using a *Sonics* Vibra-Cell™ VCX-500 sonicator with 13 mm probe at 60 % intensity for 6 x 45 s bursts, each separated by 45 s breaks. Insoluble cell debris was removed by centrifugation (24,000 rpm, 20 min, 4 °C). An additional centrifugation step (14,500 rpm, 20 min, 4 °C) was carried out where necessary. Supernatants were filtered before purification using 0.4 µm filters.

10.3.4 Small-Scale Protein Purification Trials

A cell pellet (typically ~6 g) obtained from 1 l culture was resuspended in approximately four-five times its mass (w/v) of the appropriate lysis buffer with addition of approximately 0.02 mg *DNase*I, and *Roche* complete EDTA-free protease inhibitor cocktail tablets (half a tablet/25 ml suspension), then lysed on ice by sonication with *Sonics* Vibra-Cell™ VCX-500 sonicator with 13 mm probe at 60 % intensity for 4 x 30 s bursts, each separated by 30 s breaks. Insoluble cell debris was removed by centrifugation (24,000 rpm, 20 min, 4 °C). The supernatant was filtered through a 0.4 µm filter. The lysate was then manually loaded onto Ni^{2+} Sepharose High Performance HisTrap™ HP 1 ml (*GE Healthcare*); the flow-through was collected. The column was then washed with 6 CV binding buffer, followed by 10 CV was buffer. Protein was eluted with 5 CV elution buffer; fractions of 0.5 ml were collected. For analysis, SDS-PAGE was performed as described below.

10.3.5 Visualisation of His-tagged Proteins Using InVision™ His-Tag In-Gel Stain

An SDS-PAGE gel was incubated in 50 ml fixing solution (40 % EtOH, 10 % glacial acetic acid) with gentle shaking for 1 h directly after electrophoresis. The gel was then washed twice for each 10 minutes with 50 ml H_2O, followed by staining for 1 h in 25 ml InVision™ His-Tag In-Gel Stain (*Invitrogen*) with gentle shaking. The gel was then washed twice for each 5 minutes with 50 ml of 20 mM NaH_2PO_4, pH 7.8, then transferred to a gel imager with transluminator, and His-tagged proteins were visualised under UV light (254 nm or 365 nm). The gel was then stained with Coomassie using Instant*Blue*™ (*Expedeon*).

10.3.6 Large-Scale Protein Purification

Fast Protein Liquid Chromatography (FPLC) protein purification was carried out at 4 °C using ÄKTA™ FPLC systems (*GE Healthcare*, P-920 pump system, UPC-900 UV detector, Frac-900 fraction collector, controlled by *Amersham Pharmacia* Unicorn Software V5.1). All buffers were freshly

prepared with Milli-Q water and filtered through a 0.2 µm filter prior to use. Purification of proteins was monitored using UV absorbance at 280 nm with the detector on the FPLC.

Nickel (Ni^{2+})-affinity purification of His_6-tagged proteins was carried out as immobilised metal affinity chromatography (IMAC) using Ni^{2+} Sepharose High Performance HisTrap™ HP 5 ml (*GE Healthcare*), which was manually charged and equilibrated using 5 column volumes (CV) H_2O, 3 CV charge buffer, 5 CV binding buffer. Buffer compositions are given in **Table 10-17**. The filtered cell lysate was loaded onto the column at 1 ml/min. The column was then washed with 5 CV binding buffer and 10 CV wash buffer until the signal in the UV absorbance chromatogram (monitored at 280 nm) had returned to background level. Bound protein was eluted using a step-gradient increasing the elution buffer concentration in steps of 5 %, 10 %, 15 %, 20 %, 25 %, 30 %-100 % over 0.5 CV with hold for 3-15 CV at each step, and 8 CV at 100 % elution buffer. Fractions of 2.5 ml were collected and analysed by SDS-PAGE to determine those containing the greatest amount of protein, which were pooled and concentrated to approximately 2 ml using 50, 30 or 10 kDa molecular weight cut-off (MWCO) Amicon® ultrafiltration devices for loading onto the size-exclusion column. After completed elution, the column was washed with 5 CV H_2O, and stripped of Ni^{2+} using 5 CV strip buffer, if required, then cleaned, recharged and stored according to the manufacturer's recommendations.

Table 10-17 | Composition of buffers used for Ni^{2+}-affinity chromatography.

Buffer	HEPES [mM]	NaCl [mM]	Imidazole [mM]	EDTA [mM]	$NiSO_4$ [mM]	pH
Binding	50	500	5	-	-	7.5
Wash	50	500	40	-	-	7.5
Elution	50	500	500	-	-	7.5
Strip	50	300	-	100	-	7.5
Charge	-	-	-	-	100	-

Affinity purification of glutathione *S*-transferase (GST) fusion proteins was performed with Glutathione Sepharose 4 Fast Flow GSTrap™ FF 5 ml (all *GE Healthcare*), which was manually equilibrated using 5 CV GST binding buffer (**Table 10-18**). The filtered cell lysate was loaded onto the column at 1 ml/min. The column was then washed with 12 CV GST binding buffer until a stable UV trace at background level was achieved. Bound protein was eluted using a gradient increasing the elution buffer concentration 0 %-100 % over 8 CV. Fractions of 2.5 ml were collected and analysed by SDS-PAGE. Fractions of interest were pooled and concentrated to approximately 2 ml using 50, 30 or 10 kDa MWCO Amicon® ultrafiltration devices. After completed elution, the column was stripped, cleaned, and stored according to the manufacturer's protocol.

Table 10-18 | Composition of buffers used for purification of GST-fusion proteins.

Buffer	HEPES [mM]	NaCl [mM]	L-Glutathione, reduced [mM]	Guanidine HCl [M]	pH
Binding	40	500	-	-	7.5
Elution	40	500	10	-	8.0
Strip	-	-	-	6	-

Size-exclusion purification was performed using HiLoad™ 16/60 Superdex™ 75 120 ml (*GE Healthcare*) or HiLoad™ 26/60 Superdex™ 75 300 ml (320 ml column (*Pharmacia*) manually packed with Superdex™ 75 resin (*Amersham Biosciences*)) with gel filtration buffer (50 mM HEPES, pH 7.5, 500 mM NaCl). The column was equilibrated by washing with 1 CV H_2O and 1.1 CV gel filtration buffer. Previously purified and concentrated protein was loaded onto the appropriate column via a 2 ml injection loop at a flow rate of 0.5 ml/min, and then subsequently run at 1.2 ml/min or 2.0 ml/min respectively, eluting with 1 CV gel filtration under collection of 5 ml fractions when the UV trace exceeded a set threshold absorption limit. SDS-PAGE was used to determine the fractions to be combined and concentrated.

10.3.7 Protein Concentrating, Buffer Exchange and Protein Concentration Determination

Fractions containing purified protein were pooled and concentrated by centrifugation using 50, 30 or 10 kDa MWCO Amicon® Ultra-4 Ultracel® PL ultrafiltration spin concentrators (*Millipore*). The resulting protein was exchanged into the appropriate buffer for storage by repeated dilution in the storage buffer and further concentration. Protein concentration was determined by measurement of the absorbance at 280 nm (A_{280}) using a NanoDrop ND-1000 UV/Vis spectrophotometer (*Thermo Fisher Scientific*). The concentration c can be determined, if the extinction coefficient ε is known, based on as $A_{280} = \varepsilon \cdot c \cdot l$. Protein extinction coefficients and molecular weights were calculated using the ExPASy ProtParam tool.[592] For the measurements, 2 µl of the appropriate buffer was used for the blank, and 2 µl of protein for concentration determination. Concentrated, purified protein in the storage buffer was flash-frozen in liquid nitrogen and stored at -80 °C in appropriate aliquots.

10.3.8 SDS-PAGE Analysis of Protein Fractions

Protein fractions from chromatographic purification were analysed by SDS-PAGE using either precast NuPAGE® Novex® 4-12 % Bis-Tris Mini Gels (8 x 8 cm, 1.0 mm, 12 well or 17 well) with MOPS buffer

(*Invitrogen*), run in the XCell *SureLock*™ Mini-Cell MultiGel Unit, or Tris-glycine gels prepared according to the recipes given in **Table 10-19**, run with running buffer (**Table 10-20a**) in a *Bio-RAD* Mini PROTEAN II system at constant potential of 200 V. Ammonium persulfate and TEMED were added immediately prior to pouring of the gels. Typically, 10 µl protein samples were mixed with 10 µl SDS-PAGE sample loading buffer (**Table 10-20b**), heated at 95 °C for 5 min to denature the protein, and 15 µl of the sample loaded onto the gel. Perfect Protein™ Markers (10-225 kDa, *Novagen*, each protein marker carries a His-tag sequence) were used as molecular weight markers. After electrophoresis, gels were stained with gentle shaking at room temperature with Coomassie using Instant*Blue*™ (*Expedeon*).

Table 10-19 | Recipes for Tris-glycine SDS-PAGE. **a)** Separating gel (15 % or 12 % acrylamide) and **b)** stacking gel (4 % acrylamide).

a)

Separating gel (15 %/12 % acrylamide)	
Milli-Q H_2O	1.15/1.65 ml
1 M Tris·HCl, pH 8.8	1.25 ml
30% Acrylamide	2.5/2.0 ml
10% (w/v) SDS	50 µl
10% (w/v) APS	50 µl
TEMED	4 µl

b)

Stacking gel (4 % acrylamide)	
Milli-Q H_2O	1.57 ml
0.5 M Tris·HCl, pH 6.8	0.63 ml
30% Acrylamide	0.25 ml
10% (w/v) SDS	25 µl
10% (w/v) APS	25 µl
TEMED	2.5 µl

Table 10-20 | Tris-glycine SDS-PAGE buffer recipes. **a)** 10x Running buffer and **b)** 2 x sample loading buffer.

a)

10 x SDS-PAGE running buffer
250 mM Tris·HCl
1.92 M glycine
10 % (w/v) SDS
pH 8.8

b)

For 100 ml of 2 x SDS-PAGE sample loading buffer	
0.5 M Tris·HCl, pH 6.8	10 ml
Bromophenol Blue	0.2 g
10 % (w/v) SDS	20 ml
Glycerol	12 ml
2-Mercaptoethanol	0.5 ml
Milli-Q H_2O	57.5 ml

10.3.9 Native PAGE

A precast NativePAGE™ Novex® 4-16 % Bis-Tris Mini Gel (8 x 8 cm, 1.0 mm, 10 well) was used to perform native PAGE analysis with NativePAGE™ Running Buffer (20x) and NativePAGE™ Dark Blue Cathode Buffer Additive (20x), run in the XCell *SureLock*™ Mini-Cell MultiGel Unit (all *Invitrogen*). Native PAGE electrophoresis was conducted following the manufacturer's recommendations. All buffers were chilled at 4 °C before starting the experiment. Samples were prepared containing 1 µg protein (18 µl sample containing no or low salt levels < 50 mM), mixed with 6 µl NativePAGE™

Sample Buffer (4 x), and 24 µl of the sample was loaded onto the gel. NativeMark™ Unstained Protein Standard (20-1,200 kDa, *Invitrogen*) was used as molecular weight markers. Electrophoresis was performed at a constant potential of 150 V for 110 min. Afterwards, the gel was placed in 100 ml fixing solution (40% MeOH, 10% glacial acetic acid) and heated using a microwave (950-1100 W) for 45 s. The gel was then gently shaken at room temperature for 15 min. The fixing solution was decanted, 100 ml destain solution (8 % glacial acetic acid) was added, and the gel was heated again in the microwave (950-1100 W) for 45 s. The gel was then gently shaken at room temperature until the desired background was obtained.

10.3.10 Tris-Tricine Gel Analysis

Tricine SDS-PAGE[424] was used for analysis of protein-peptide complex fractions following the method of Schägger and coworkers.[424] The samples were mixed with Tricine Sample Buffer (*Bio-RAD*) including 2 % 2-mercaptoethanol and heated at 95 °C for 5 min. They were then run on a 10-20 % Mini-PROTEAN® Tris/Tricine Gel using Tris/Tricine/SDS eletrophoresis buffer (both *Bio-RAD*) in a *Bio-RAD* Mini PROTEAN II system at constant potential of 100 V for approximately 100 min. Ultra-low Range Molecular Weight Marker (1,060-26,600 Da, *Sigma-Aldrich*) was run as reference on the gel, and Instant*Blue*™ (*Expedeon*) was used as Coomassie stain for the gel after electrophoresis.

10.3.11 His_6-Tag Cleavage

Cleavage of the His_6-tag from a recombinant protein was performed following Ni^{2+} affinity purification and buffer exchange into 50 mM HEPES, pH 7.5. His_6-Tobacco Etch Virus (TEV) protease (~3 mM) was kindly provided by Dr Jürgen Brem; Thrombin, Restriction Grade was from *Novagen* Thrombin Cleavage Capture Kit. For cleavage using thrombin, the His_6-tagged protein was incubated with 0.25 units Thrombin per 1.0 mg of protein for 40 h at 4 °C in Thrombin Cleavage Buffer (20 mM Tris-HCl, pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$). For cleavage with TEV protease, the His_6-tagged protein was incubated with TEV protease at a protein to TEV ratio of 10:1 by molarity for 24 h at 4 °C in 50 mM HEPES, pH 7.5. To remove the His_6-TEV protease, the solution was re-applied to the Ni^{2+}-affinity column. The cleaved protein was collected in the flow-through and concentrated for further purification.

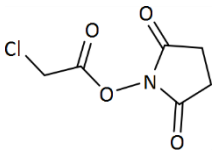
Alternatively, frozen aliquots of purified, concentrated protein in storage buffer (50 mM HEPES, pH 7.5) were diluted to 10 mg/ml in a volume of 50-60 µl of thrombin cleavage buffer (20 mM Tris-HCl, pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$), and incubated with 0.25 units thrombin per 1.0 mg of protein for 40h at 4 °C. For cleavage using TEV protease, frozen aliquots of concentrated protein in storage buffer (50 mM HEPES, pH 7.5) were diluted to 10 mg/ml in a volume of 50-60 µl in 50 mM HEPES, pH 7.5, and incubated with TEV protease (ratio protein to TEV 10:1 by molarity) for

24 h at 4 °C. His_6-TEV was then removed by incubation with 20 µl Ni^{2+}-NTA beads (washed and pre-equilibrated in 50 mM HEPES, pH 7.5), for 60 min at 22 °C with mixing in Thermomixer® comfort (300 rpm, *Eppendorf*). The cleavage reaction was monitored by LC-MS analysis. After successful tag cleavage, the protein samples were then directly loaded onto Micro Bio-Spin columns (*Bio-RAD*) for buffer exchange into 15 mM ammonium acetate, pH 7.5, as described in **Section 10.5.8** for non-denaturing mass spectrometry.

10.4 SYNTHETIC PROCEDURES

10.4.1 Organic Synthesis

10.4.1.1 Chloromethylcarbonyloxysuccinimide (ClAc-OSu)



A previously reported procedure[593] was adapted for the synthesis of ClAc-OSu. Chloroacetyl chloride (1.59 ml/20.0 mmol) was added dropwise to a suspension of *N*-hydroxysuccinimide (2.30 g/20.0 mmol) and $NaHCO_3$ (1.68 g/20.0 mmol) in 40 ml DCM at 0 °C and then stirred at room temperature for 2 h. The reaction mixture was filtered to remove all remaining solids, and the solvent was evaporated *in vacuo* to give a white crystalline material. The residue was taken up in 15 ml $CHCl_3$ and warmed up to 40 °C. An equal volume of cHex was added, and the product left to crystallise overnight. The supernatant was discarded and the product dried *in vacuo* to yield 3.39 g (88 %) of a white crystalline solid.

^{1}H NMR ($CDCl_3$, 400MHz): δ [ppm] = 4.37 (s, 2H, ClC***H***$_2$C(O)), 2.85 (s, 4H, 2x C***H***$_2$C(O)). **^{13}C NMR** ($CDCl_3$, 101MHz): δ [ppm] = 168.4 ($ClCH_2$***C***(O)), 163.2 (CH_2***C***(O)), 37.8 (***C***H_2C(O)), 25.5 (Cl***C***H_2C(O)). **MP** 119-122 °C. R_f ≈ 0.2 (cHex/EE (1:1)). **IR:** ν_{max} = 1734 cm^{-1}. **MS (FI)**: calculated for $C_6H_6NO_4Cl$: *m*/*z* 190.9985, found: *m*/*z* 190.9983.

10.4.1.2 General Procedure for the Synthesis of Fluorescence Dye-BCN-Conjugates

Fluorescence dye-BCN conjugates were synthesised as in previously reported procedures.[497,498] Sulforhodamine B (230162 ALDRICH, dye content 75 %), fluorescein isothiocyanate isomer I (F4274 SIGMA, ~98 % (HPLC)), and the bicyclo[6.1.0]nonyne (BCN)-biotin-conjugate *N*-[(1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl-oxycarbonyl]-*N*'-biotinyl-1,8-diamino-3,6-dioxaoctane (742783 ALDRICH) were purchased from *Sigma-Aldrich* and the BCN-conjugate (*N*-(1R,8S,9s)-bicyclo[6.1.0]non-4-yn-9-ylmethyloxycarbonyl 1,8-diamino-3,6-dioxaoctane (SX-A1004, > 95 %, BCN-POE_3-NH_2) from *SynAffix* (Nijmegen, Netherlands).

One equivalent (equiv.), typically about 20 mg, of the BCN-conjugate BCN-POE_3-NH_2 (SX-A1004) was stirred with one equiv. of fluorescein *iso*thiocyanate isomer I or sulforhodamine B respectively in 1.0 ml of anhydrous DMF at 22 °C for ~16 h in the dark until no starting material was left (monitored by LC-MS and TLC in DCM/MeOH (9:1), 0.5% acetic acid). The solvent was then removed *in vacuo* (three times co-evaporated with each 3 ml toluene to remove residual amounts of DMF). The desired product was typically obtained in apparently quantitative yield, as determined by LC-MS analysis, and used without further purification.

10.4.1.3 General Procedure for Strain-Promoted Azide-Alkyne Cycloaddition

Strain-promoted azide-alkyne cycloaddition reactions were performed as in previously reported procedures.[495-497] One equiv. of the peptide containing the azide, typically about 2-3 mg, was mixed with 1.0-1.5 equiv. of the BCN-conjugate in 100 µl H_2O/CH_3CN (2:1; the poorly soluble FITC-BCN-conjugate was first dissolved in 50 µl MeOH and then added to the peptide solution) and shaken in the dark at 37 °C, 600 rpm using a Thermomixer® comfort (*Eppendorf*). The reaction progress was monitored by MALDI-ToF MS or LC-MS. The reaction times depend on the steric demand of the azide; azidoalanine reacted in ~24 h. The desired product was typically obtained in apparently quantitative yield and the remaining BCN starting material was separated by HPLC purification (*Agilent* 1200 Series, *Waters* Sunfire column), or the peptide was used without further purification.

10.4.2 Cyclic Peptide Design

The model of the $CP2_{10mer}$ was constructed in PyMOL (*Schrödinger, Inc.*; Version 1.5) as described in **Section 4-1**, and the linear precursor was synthesised by solid-phase peptide synthesis (**Section 10.4.3**). The Cys3-4 disulfide bond in the original hFX $EGF1_{39mer}$ peptide substrate was replaced by a stable thioether bond using the methodology of Suga and coworkers.[493] The *N*-terminal amino acid was designed to have *R*-stereochemistry to orient the side chain towards the position of the peptide backbone of the original substrate, and the *N*-terminus was functionalised with a chloroacetyl moiety, which reacted with the *C*-terminal cysteine to give the thioether. Both *R*-alanine and *R*-azidoalanine were incorporated.

10.4.3 Solid-Phase Peptide Synthesis

Protected amino acids for solid-phase peptide synthesis[594] were from *CS Bio*, *Novabiochem*, *Sigma-Aldrich*, *TCI Chemicals*, *Alfa Aesar*, and *AGTC Bioproducts*. Azide-containing amino acids were from *Iris Biotech*. Linear precursors of cyclic peptides were synthesised using solid-phase peptide synthesis (SPPS) on a *CS-Bio* CS336X automated peptide synthesiser (100 µmol scale) with PL-AMS resin

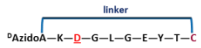
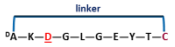
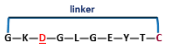
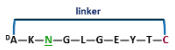
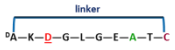
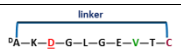
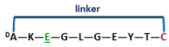
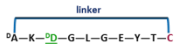
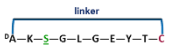
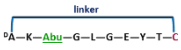
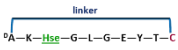
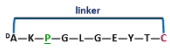
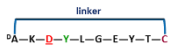
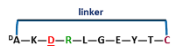
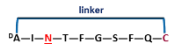
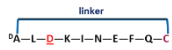
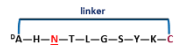
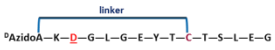
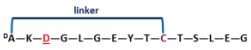
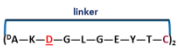
(*Polymer Laboratories*) and a Rink amide linker. Fluorenylmethyloxycarbonyl(Fmoc)-protected amino acids were activated using di*iso*propyl-carbodiimide (DIC)/1-hydroxybenzotriazole. Linear peptides were synthesised sequentially from their *C*- to *N*-termini. After cleavage of the *N*-terminal Fmoc-protecting group a solution of 150 mg of chloromethylcarbonyloxysuccinimide (ClAc-OSu, synthesis in **Section 10.4.1.1**), or alternatively chloroacetic anhydride (150 mg, *Sigma-Aldrich*), in 4 ml DMF was added to the resin and the mixture shaken for 3 h. The resin was filtered off, and cleavage of the completed linear peptide from the resin was performed by subsequent treatment with 4 ml of deprotection solution (95 % trifluoroacetic acid, 2.5 % tri*iso*propylsilane, 2.5 % water). After 3 h, the mixture was filtered, and the peptide was precipitated with ~45 ml cold diethylether. The mixture was centrifuged (4,000 rpm, 10 min, 4 °C) and the supernatant discarded. The peptide mass in the crude mixture was analysed by MALDI-ToF-MS.

10.4.4 Peptide Cyclisation Reaction and HPLC Purification

Cyclisation of peptides[594] was performed just prior to HPLC purification to avoid precipitation of peptides. The peptides were cyclised by S_N2-reaction between the sulfur atom of the *C*-terminal cysteine and the *N*-terminal chloroacetyl moiety. For this, the solid linear peptide was taken up in 1.5 ml of 1 M triethylammoniumacetate buffer, pH 8.5, the pH readjusted to > 8 with triethylamine if necessary, and the peptide solution transferred to a microwave vial (0.5 – 2 ml V-bottom glass vials). The mixture was heated to 100 °C for 10 min using a microwave (*Biotage* Initiator) and subsequently purified by HPLC (typically 0-30 % or up to 0-60 % CH_3CN in 45 min, 0.1 % TFA, depending on solubility of the peptide, conditions given in **Table 10-21**). Preparative HPLC purifications were carried out on a Dionex™ UltiMate® 3000 (*Thermo Scientific*) system with a *Grace* Vydac® 218TP101522 HPLC column, and the peptide mass was confirmed by MALDI-ToF-MS. Purity of the peptide was analysed using an *Agilent* 1200 series LC-MS system (6120 quadrupole MS, *Agilent Technologie*s) with a *Waters* C_{18} Sunfire column (0-45 % CH_3CN in 45 min) or Dionex™ UltiMate® 3000 coupled to Q Exactive Orbitrap (*Thermo Scientific*) as described below in **Section 10.5.4**.

Synthesised peptides and applied HPLC purification conditions are listed in **Table 10-21**, and an overview of peptides conjugated to fluorescence dyes or biotin is given in **Table 10-22**.

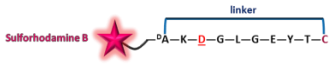
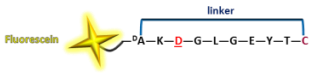
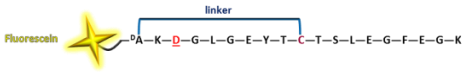
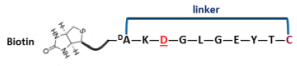
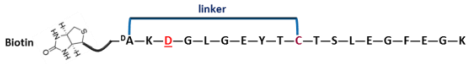
Table 10-21 | Peptides synthesised by SPPS and purified by HPLC.

Peptide	Sequence	HPLC gradient	Calculated mass [Da]	Observed* mass [Da]	Purity**
CP1		0-30 % B	1,136.20	1,136.48	> 95 %
LP1		0-30 % B	1,096.18	1095.98	> 95 %
CP2		0-30 % B	1,095.18	1,095.48	> 95 %
CP3		0-38 % B	1,081.16	1,081.46	> 95 %
CP4		0-38 % B	1,135.21	1,135.49	> 95 %
LP4		0-30 % B	1,095.19	1,094.99	> 95 %
CP5		0-30 % B	1,094.18	1,094.49	> 95 %
CP6		0-50 % B	1, 003.09	1,003.45	> 95 %
CP7		0-35 % B	1,031.14	1,031.48	> 95 %
CP8		0-50 % B	1,109.21	1,110.52	> 60 %
CP9		0-38 % B	1,136.20	1,136.48	> 95 %
CP10		0-35 % B	1,067.17	1,067.48	> 95 %
CP11		0-38 % B	1,065.20	1,065.50	> 95 %
CP12		0-30 % B	1,081.20	1,081.50	> 95 %
CP13		0-30 % B	1,077.22	1,077.50	> 95 %
CP14		0-30 % B	1,201.32	1,201.52	> 95 %
CP15		0-30 % B	1,194.33	1,194.56	> 95 %
CP16		Not purified	1,126.24	1,126.50	> 65 %
CP17		0-30 % B	1,219.37	1,219.58	> 95 %
CP18		0-60 % B	1,132.25	1,132.52	> 95 %
CP1$_{15mer}$		0-35 % B	1,623.70	1,623.70	> 95 %
CP2$_{15mer}$		0-35 % B	1,582.69	1,582.70	> 95 %
CP1$_{19mer}$		0-35 % B	2,085.23	2,085.71	> 65 %
CP2$_{19mer}$		0-35 % B	2,044.22	2,044.93	> 95 %
CP2$_{double}$		0-30 % B	2,190.37	2,189.94	> 95 %

* The values for masses refer to the monoisotopic mass [M] determined by analysis using Q Exactive Orbitrap as described in Section **10.5.4**.

** Purity as determined by LC-MS.

Table 10-22 | Peptides conjugated to fluorescence dyes or biotin.

Peptide	Sequence	Calculated mass [Da]	Observed* mass [Da]	Purity**
$FP1_{10mer}$		2,000.82	2,000,82	> 95 %
$FP2_{10mer}$		1,849.99	1,850.71	> 90 %
$FP2_{19mer}$		2,799.04	2,799.15	> 65 %
$BP1_{10mer}$		1,686.92	1,686.74	> 90 %
$BP1_{19mer}$		2,635.92	2,636.18	> 65 %

* The values for masses refer to the monoisotopic mass [M] determined by analysis using Q Exactive Orbitrap as described in Section **10.5.4**.
** Purity according to the supplier as determined by HPLC.

10.5 Analytical and Biophysical Assay Methods

10.5.1 Mass Spectrometry-Based AspH Hydroxylation Activity Assays

Peptides to be tested as AspH substrates were usually dissolved in Milli-Q H_2O to give a 1 mM stock solution. The cbEGFD proteins from Professor Penny A. Handford's laboratory received in solution (water) were lyophilised and reconstituted in Milli-Q H_2O at 1 mM stock concentration. Control samples were prepared by replacing AspH with buffer. The total assay volume was variable, with 10 µl as the smallest assay volume, and scaled up to the desired volume as appropriate, while maintaining the final concentrations and ratios. The assay was typically performed in triplicates using either individual 0.5 ml Eppendorf tubes incubated in a Thermomixer® comfort (*Eppendorf*) or clear non-skirted 96-well PCR plates (*STARLAB*) in a 96-well GeneAmp® PCR System 9700 (*Applied Biosystems*) at 37 °C for variable time, typically 60-180 min, as indicated.

Reagents 2-oxoglutaric acid disodium salt dihydrate (2OG) and sodium ascorbate (Asc; both from *Sigma-Aldrich*) were prepared in Standard Buffer, as given in **Table 10-23**. $(NH_4)_2Fe(SO_4)_2 \cdot 7\ H_2O$ (Fe(II)) was dissolved in 20 mM HCl and then diluted to the appropriate stock concentration with Milli-Q H_2O. An 'Enzyme Mix' containing reaction buffer, typically 4-10 µM AspH (for an AspH to substrate ratio 1:25 or 1:10 respectively, as stated in the experiments), 400 µM Asc, and 50 µM Fe(II), and a 'Substrate Mix' containing reaction buffer, 100 µM substrate, 300 µM 2OG, and additional reagents such as $CaCl_2$, $MgCl_2$, and ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetraacetic acid

(EGTA), where appropriate, were prepared separately. A 'Control Mix' was prepared to replace the 'Enzyme Mix' in the control samples (enzyme replaced by buffer). The assayed reaction was started by mixing equal volumes of 'Substrate Mix' and 'Enzyme or Control Mix' respectively. The tubes or plates were then immediately incubated at 37 °C. At the end of the assay time, the reaction was stopped by addition of an equal volume of 1 % formic acid (HCOOH). For time course measurements, at each time point (as indicated in the given experiment), typically a sample of 5-10 µl of the reaction mixture was removed and mixed immediately with an equal volume of 1 % HCOOH to quench the reaction. Whenever a reagent was omitted from the reaction mixture (e.g. for cofactor dependence analysis), it was replaced by the appropriate buffer. Samples were stored at -20 °C until solid-phase extraction, where required, and analysis by MALDI-ToF-MS or ESI-MS.

For K_m value determination assays containing 1 µM AspH, serial dilutions of the peptide substrates were made in Milli-Q H_2O from a stock concentration; serial dilutions of 2OG were made in Standard Buffer. For inhibition assays, the compounds were dissolved at appropriate stock concentration in buffer or DMSO and serially diluted in Standard Buffer. The enzyme was pre-incubated with the inhibitor compound for 30 min prior to initiation of the reaction (cofactors for 'Enzyme Mix' were added prior to the 'Substrate Mix').

Table 10-23 | Nomenclature and composition of specific buffer systems frequently used in this Thesis.

	Buffer system composition
Standard Buffer (Standard conditions)	50 mM HEPES, pH 7.5, 150 mM NaCl
RedOx Buffer (RedOx conditions)	50 mM Tris-HCl, pH 8.5, 3.0 mM reduced *S*-(-)-glutathione (GSH), 0.3 mM oxidised *S*-(-)-glutathione (GSSG), 150 mM NaCl
Refolding Buffer (Refolding conditions)	50 mM Tris-HCl, pH 8.3, 3.0 mM *S*-cysteine, 0.3 mM *S*-cystine, 10 mM $CaCl_2$

10.5.2 MALDI-ToF-MS and MS/MS

Matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-ToF-MS) or MS/MS was performed using a *Bruker Daltonics* Ultraflex I™ MALDI-ToF/ToF located at the Oxion Proteomics Facility at the Department of Physiology, Anatomy and Genetics, University of Oxford, operated with flexControl™ 3.0 software. Spectra were recorded in positive ion reflectron mode with typically 32-60 % laser energy. Calibration was performed on each day prior to the measurements using Peptide Calibration Standard II (*Bruker Daltonics*). Data were processed using *Bruker Daltonics* flexAnalysis™ 3.0 software. Alternatively, MALDI-ToF-MS data were acquired on a *Waters* Micromass® MALDI micro MX™ mass spectrometer in positive ion reflectron mode, with flight tube voltage at 12 kV and reflectron voltage at 5.2 kV. The laser fire rate was set at 10 Hz with 10 shots per spectrum, and the following parameters were typically used: The laser energy was set to

160-200, pulse voltage 1950 V, detector voltage 2750 V, mass suppression 200 Da. Data were then processed using MassLynx™ version 4.1.

For MALDI measurements, when trypsin digestion was not required, typically 1 µl of the sample was mixed with 4 µl of matrix, either 2,5-dihydroxybenzoic acid (DHB, *Bruker Daltonics*, 20 mg/ml in 50 % CH_3CN/0.1 % TFA), *α*-cyano-4-hydroxy-cinnamic acid (CHCA, *LaserBio Labs* or *Bruker Daltonics*, 10 mg/ml in 50 % CH_3CN/0.1 % TFA), sinapinic acid (SA, *Bruker Daltonics*, 10 mg/ml in 50 % CH_3CN/0.1 % TFA) or alternatively, especially for peptides containing disulfides, with 4 µl of a 10:1 mixture of CHCA/2-(4'-hydroxybenzeneazo)benzoic acid (HABA, *Bruker Daltonics*) as matrix (10 mg/ml in 50 % CH_3CN/0.1 % TFA), and 1-2 µl of this sample-matrix mixture spotted onto 24 x 16 MTP AnchorChip™ 384 T F MALDI target or a 12 x 8 MALDI target plate (*Waters*) respectively, and allowed to dry before analysis. For measurements on *Waters* Micromass™ MALDI micro MX™ mass spectrometer, alternatively a sample-matrix (in this case CHCA) ratio of 1:1 was used when sample and matrix were directly mixed on the target.

Data processing included background subtraction and automated plus manual peak picking, as required. Data were quantified (% hydroxylation) by referencing to the intensity of the unhydroxylated peak (and to the aspartyl succinimide (Asu) containing peaks).

10.5.3 MS/MS Data Acquisition and Processing

MS/MS data were acquired either using a *Bruker Daltonics* Ultraflex I™ MALDI-ToF/ToF using the LIFT technique based on post-source and post-decay acceleration of fragment ions in positive ion reflectron mode with typically 50-70 % laser energy and 17 kV acceleration voltage, or using a *Bruker Daltonics* amaZon™ Ion Trap LC MS/MS system equipped with a Dionex™ UltiMate® 3000 nano-HPLC system comprising a WPS-3000 micro auto sampler, a FLM-3000 flow manager and column compartment, a UVD-3000 UV detector, an LPG-3600 dual-gradient micro-pump, and an SRD-3600 solvent rack controlled by HyStar™ software version 3.2 (*Bruker Daltonics*) and DCMS link 2.0 software (*Dionex*).

Samples were concentrated on a trapping column (*Dionex*), 300 µm internal diameter (i.d.), 0.1 cm) at a flow rate of 20 µl/min. For separation with a C_{18} PepMap column (75 µm i.d., 15 cm, *Dionex*), a flow rate of 250 nl/min was used as generated by a cap-flow splitter cartridge (1/1000), and the peptides were eluted with a multi-step gradient over ~30 min (solvent A 98 % H_2O/2 % CH_3CN/0.1 % formic acid, solvent B 80 % CH_3CN/20 % H_2O/0.1 % formic acid; 0-3 min 2 % B, 3-18 min 10 % B, 18-30 min 25 % B, 30-30.5 min 90 % B). The interfaced amaZon 3D high capacity ion trap mass spectrometer utilised 10 µm i.d. distal coated SilicaTips (*New Objective*, Woburn, USA) and nano-ESI mode. Parameter settings on the ion trap were tuned for a target mass of *m/z* 850, compound stability 100 %, and a smart ICC target of 250,000. MS/MS analysis was initiated on a contact closure signal triggered by HyStar™, and up to four precursor ions were

selected per cycle with active exclusion (0.5 min) in collision-induced dissociation (CID) mode. CID fragmentation was achieved using helium gas and 30 %-100 % collision energy sweep with amplitude 1.0 (ions are ejected from the trap as soon as they fragment).

MS/MS spectra (peak lists) were inspected and assigned manually, and, where possible in cases of reduced, alkylated and proteolytically digested proteins listed in databases, confirmed with searches against the UniProtKB/Swiss-Prot database using BioTools™ version 3.0 (*Bruker Daltonics*) or Mascot version 2.4 (*Matrix Science*, 2011) on an in-house server (Computational Biology Research Group Cluster (CBRG), University of Oxford) with the following parameters: database searches restricted to human species; peptide tolerance 0.3 Da; #^{13}C 1; MS/MS tolerance 0.2 Da; 2 missed cleavages, peptide charge 1+, 2+ and 3+ ions; instrument type ESI-Trap; fixed modification either 'Carbamidomethyl (C)' or 'Nethylmaleimide (C)'; and variable modifications oxidation (M, N, D, K, Y, P) and/or dehydration/deamidation (D, N, Q). Raw LC-MS/MS data from amaZon™ Ion Trap were processed to generate Mascot-compatible files (mgf) using DataAnalysis 4.0 software (*Bruker Daltonics*) with the following parameters: compounds (autoMS) threshold 10,000; number of compounds 500; and retention time windows 0.5 min for C_{18} (30 min gradient).

10.5.4 LC-MS Analysis of EGFD Peptides and Proteins

Analysis of EGFD peptides and proteins by LC-MS was performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (*Thermo Scientific*), operated with Xcalibur™ 2.2 software, coupled to a Dionex™ UltiMate® 3000 with a DAD-3000RS Rapid Separation Diode Array Detector (*Thermo Scientific*) controlled through Chromeleon® 7.2 Chromatography Data System software. A *Phenomenex* Aeris™ WIDEPORE 3.6 µm XB-C8 column (C_8, 50 x 2.1 mm, pore size 200 Å) was used for chromatographic separation at a temperature of 40.0°C. A multi-step gradient over 10 min was run (solvent A 99.9 % H_2O/0.1 % formic acid, solvent B 99.9 % CH_3CN/0.1 % formic acid; 0-1 min 5 % B for equilibration, followed by a linear gradient to 99 % B over 5 min, then 99 % B for an additional 1 min, followed by a linear gradient over 3 min back to 5 % B to re-equilibrate the column) to separate and elute the peptide/protein samples at a flow rate of 0.3 ml/min.

The peptides are detected on a Q Exactive™ mass spectrometer equipped with an Orbitrap mass analyser. A heated electrospray source ionisation (HESI) probe was used as an ion source. The source temperatures and voltages were adjusted to maximise the signal; the spray voltage was set to 3.50 kV. The instrument was operated in positive ionisation mode, run in Fullscan MS with data-dependent MS^2 acquisition. The scan range for full MS was *m*/*z* 200-3000, collecting data with 140,000 full width at half maximum (FWHM) resolution. Data-dependent scans were acquired with 17,500 FWHM) resolution and a normalised collision energy (NCE) of 30.0, and the scan range was selected by the software. Data were analysed using Xcalibur™ 2.2 Qual Browser software. UV absorbance chromatograms are presented in a UV wavelength range from 190.0-215.0 nm.

10.5.5 In-Solution Digestion of Proteins with Proteases

Porcine Sequencing Grade Modified Trypsin, Sequencing Grade Chymotrypsin and Endoproteinase LysC were from *Promega*, endoproteinases GluC (*Staphylococcus aureus* Protease V8) and AspN (Flavastacin) was from *New England Biolabs*.

All reagent solutions were freshly prepared just prior to use. Proteases were stored in aliquots at -20 °C. Digested samples were subsequently purified and desalted by solid-phase extraction. Protein or peptide solutions from quenched assays were concentrated to dryness using a Vacufuge® vacuum concentrator (*Eppendorf*), equipped with either fixed-angle rotors F-45-72-8 (72 x 0.5 ml Eppendorf tubes) or F-45-48-11 (48 x 1.5 ml/2.0 ml) respectively, or swing-bucket rotor A-2-VC (2 x MTP or skirted PCR plates or 2 x DWP), connected to an external diaphragm pump.

For analysis of hydroxylation using denaturing conditions, the dried protein sample (< 300 µg) was resuspended in 30 µl of 6 M urea in 100 mM phosphate buffer, pH 6.6. Reduction of disulfides was accomplished with the addition of 1.5 µl of reducing agent 194 mM dithiothreitol (DTT; in 100 mM phosphate buffer, pH 6.6) and incubation for 30 min at room temperature, followed by alkylation for 120 min at room temperature with 6 µl of 195 mM *N*-ethylmaleimide (NEM; in 100 mM phosphate buffer, pH 6.6). Alternatively, the 6 M urea and 194 mM DTT solutions were made up in 100 mM Tris buffer, pH 7.8, when 195 mM iodoacetamide (in 100 mM Tris buffer, pH 7.8), was used for alkylation with an incubation time of 30-60 min. Unreacted alkylating reagent was consumed by addition of 6 µl DTT solution to the sample, which was then allowed to stand for 30 min at room temperature. The sample was diluted with 235 µl phosphate or Tris buffer respectively, to reduce the urea concentration, and mixed with 1 µl of 200 ng/µl porcine Sequencing Grade Modified Trypsin (*C*-terminal cleavage at Lys and Arg; *Promega*) dissolved in 100 mM Tris buffer, pH 7.8, for digestion at 37 °C for 8-16 h. Alternatively, 1-2 µl of 100 ng/µl endoproteinase GluC (*Staphylococcus aureus* Protease V8; *C*-terminal cleavage at Glu), endoproteinase AspN (*N*-terminal cleavage at Asp; both *New England Biolabs*), Sequencing Grade Chymotrypsin (*C*-terminal cleavage at Tyr, Phe, Trp and Leu), or endoproteinase LysC (*C*-terminal cleavage at Lys; both *Promega*), each dissolved in water, was used. Digestion was stopped by adjusting the sample to pH 3-4 by addition of concentrated acetic acid (~3-8 µl). The samples were typically stored at -20 °C until solid-phase extraction and analysis.

For analysis of peptides with intact disulfides, the dried peptide or protein sample (< 300 µg) was resuspended in 30 µl of 6 M urea in 100 mM phosphate buffer, pH 6.6, diluted with 250 µl of phosphate buffer, and digested with 1-2 µl of protease as described above.

For sequential double digest, GluC digest was typically performed first as described above for 16 h at 37 °C without lowering of the pH (after this time protease activity has faded), directly followed by digest with Trypsin as described.

10.5.6 Solid-Phase Extraction (SPE)

Solid-phase extraction (SPE) for purification and desalting of samples (e.g. after proteolytic digestion) was performed using either single ISOLUTE® PH 25 mg/1 ml SPE cartridges or ISOLUTE®-96 PH Fixed Well SPE Plates 25 mg/1 ml (*Biotage*; phenyl (non end-capped) functionalised silica, manufactured using trifunctional silane, average particle 50 µm, nominal porosity 60 Å) using a SPE vacuum manifold connected to a vacuum pump. Cartridges were first flushed with 500 µl solution B (**Table 10-24**), then equilibrated with 1 ml solution A. Digested samples were each mixed with 200 µl solution A and loaded onto the cartridge. The cartridges were washed with 1 ml solution A, and samples were eluted into fresh collection wells with 2 x 200 µl solution B. The samples were then concentrated to dryness in either single 1.5 ml Eppendorf tubes or V-bottom 96-well plates (*Axygen*; capacity 500 µl, clear, nonsterile) in Vacufuge® vacuum concentrator (*Eppendorf*) using fixed-angle rotor F-45-48-11 or swing-bucket rotor A-2-VC respectively. Dry samples were re-dissolved for analysis by MALDI-ToF-MS in typically 15-30 µl of 50 % CH_3CN/0.1 % TFA, or for LC-MS in 15-30 µl of 5 % CH_3CN/0.1 % formic acid and diluted as appropriate.

Alternatively, Sep-Pak C_{18} Plus Light Cartridges (*Waters*, 130 mg sorbent per cartridge, 55-105 µM particle size) were used for SPE. These cartridges were manually operated using syringes. The cartridges were first flushed with 5 ml of solution C (**Table 10-24**), then equilibrated with 10 ml solution A. Digested samples were each mixed with 200 µl solution A and loaded onto the cartridge. The cartridges were then washed with 10 ml solution A, samples were eluted into fresh collection tubes with 2 x 1 ml of solution C, and subsequently dried and re-dissolved as described above.

Table 10-24 | Composition of solutions used for solid-phase extraction.

Solution A	Solution B	Solution C
97.9 % Milli-Q H_2O	19.9 % Milli-Q H_2O	34.9 % Milli-Q H_2O
2 % CH_3CN	80 % CH_3CN	65 % CH_3CN
0.1 % HCOOH	0.1 % HCOOH	0.1 % HCOOH

10.5.7 Denaturing Protein Mass Spectrometry

Liquid chromatography-mass spectrometry (LC-MS) was performed using a *Waters* Micromass LCT Premier™ time-of-flight mass spectrometer with electrospray ionisation (ESI-ToF-MS) coupled to a *Waters* 1525 Micro Binary HPLC Pump using a Chromolith® FastGradient RP-18 Endcapped 50-2 HPLC column, made of monolithic silica (C_{18}, 2 x 50 mm, macropores with 1.6 µm diameter, *Merck*) or a Chromolith® RP-18 Endcapped 5-2 Guard Cartridge (C_{18}, 2 x 5 mm, macropores with 1.5 µm diameter, *Merck*). The instrument was connected to a CTC-autosampler inlet system. A multi-step gradient over 10 min was run (solvent A 99.9 % H_2O/0.1 % formic acid, solvent B 100.0 % CH_3CN; 0-1 min 5 % B for equilibration, followed by a linear gradient to 100 % B over 4 min, then 100 % B for

an additional 3 min, followed by a linear gradient over 2 min back to 5 % B to re-equilibrate the column) to separate and elute protein samples at flow rates of 0.4 ml/min for the first 5 min and then 1.0 ml/min for the remaining time. The electrospray source was operated with a capillary voltage of 3.0 kV and a cone voltage of 100 V. Nitrogen was used as the nebuliser and desolvation gas at a flow of 600 l/h. Protein material typically elutes as a peak between 3 min and 5 min under these conditions. For analysis, the mass spectra for all protein material in the peak were combined using *Waters* MassLynx™ version 4.1. The resulting combined ion series was deconvoluted to convergence using a maximum entropy algorithm that is preinstalled on the software.

10.5.8 Non-Denaturing (Native) Mass Spectrometry

An adaptation of a reported method[510] was used for non-denaturing ESI-MS studies. The protein samples were desalted and buffer exchanged into 15 mM ammonium acetate, pH 7.5, using Micro Bio-Spin columns (*Bio-RAD*). The concentration was determined using NanoDrop ND-1000 UV/Vis spectrophotometer (*Thermo Fisher Scientific*), and the sample was diluted with the same buffer to a stock concentration of 100 µM. $(NH_4)_2Fe(SO_4)_2 \cdot 7\ H_2O$ was dissolved in 20 mM HCl at a concentration of 500 mM, and then diluted with Milli-Q H_2O to get a stock concentration of 100 µM Fe(II). 2-Oxoglutaric acid and peptides were directly dissolved (or peptide stock concentrations in Milli-Q H_2O used) in 15 mM ammonium acetate, pH7.5, and diluted to 100 µM stock solutions with the same buffer. For measurements, the protein (15 µM final assay concentration) was mixed with one or more equivalents of the other components and incubated at room temperature for 15 min prior to ESI-MS analysis.

Data were acquired on a *Waters* Micromass Q-ToF micro™ interfaced with a NanoMate™ chip-based nano-ESI source (*Advion Biosciences*). Samples were measured in 384-well plates (Microplate for PCR polypropylene natural 384-well, skirt; *FisherScientific*) and injected into the Q-ToF through the ESI chip at a flow-rate of approximately 100 nl/min. A delivery pressure of 0.5 psi and a spraying voltage of 1.8 kV were applied through a standard Z-spray source block. The cone voltage was 60 V with a source temperature of 40 °C and an acquisition/scan time of 10 s/1 s. Scanning was performed in positive ion mode in the range of *m/z* 500-5,500. The interface pressure between the atmospheric source and the high vacuum was set at 6.6 mbar. Data processing was performed using MassLynx™ 4.1 software (*Waters*).

10.5.9 Differential Scanning Fluorimetry

An adaptation of reported methods[472] was used for differential scanning fluorimetry (DSF) studies performed to determine the T_m values of proteins. DSF was performed using a MiniOpticon™ Real-Time PCR Detection System (*Bio-RAD*). The assay was carried out in white 8 x 6 PCR wells. The total

assay volume per well was 50 µl with final concentrations of 2 µM protein and, where appropriate, 50 µM metal (typically $MnCl_2$ or chloride salts of other metals). SYPRO Orange dye (*Invitrogen*) was used for unspecific binding to hydrophobic residues, and its increase in fluorescence was monitored as a function of temperature. The dye stock solution in DMSO (5,000 x concentrate) was diluted 1:2,500 in the appropriate buffer (typically 50 mM HEPES buffer, pH 7.5; 200 mM HEPES buffer, pH 7.5, was used when high concentrations of TCA cycle compounds were included). For measurements of the effect of TCA cycle intermediates on the stability of proteins, the compounds were dissolved in 200 mM HEPES, pH 7.5, to appropriate stock concentrations. Compounds of the 2OG mimetics library and substances insoluble in buffer solutions were dissolved in DMSO at appropriate stock concentrations and diluted in buffer so that a final concentration of 2 % DMSO in the assay was not exceeded. Measurements were typically carried out in triplicates. Fluorescence readings (monitored at 492 nm excitation and 610 nm emission, FAM and ROX filters) were taken between 25 °C and 90 °C, increasing the temperature linearly in steps of 1 °C every minute. For determination of T_m values, melting curves for each triplicate data set were exported into GraphPad Prism 5.0. A Boltzmann curve was fitted between the values of minimum and maximum intensities. The midpoint (i.e. the T_m value) was calculated by the software.

10.5.10 Crystallisation Methods

Purified AspH (> 95 % purity as judged by SDS-PAGE) was crystallised at a concentration of 18 mg/ml in 50 mM HEPES, pH 7.5 with 1 mM $MnCl_2$ (dissolved and diluted in 50 mM HEPES, pH 7.5) and 2 mM NOG or 2,4-PDCA (both *Sigma-Aldrich*, dissolved and diluted in 50 mM HEPES, pH 7.5, to 100 mM stock solution, pH adjusted to ~7.5 with 100 mM NaOH controlled using pH paper). For screens with peptide substrates, the solid peptides were dissolved in 50 mM HEPES, pH 7.5, diluted to the appropriate stock concentrations (pH adjusted to ~7.5 with 100 mM NaOH), and added last to the AspH mixture to the desired ratios. Small amounts of precipitate, formed upon addition of $MnCl_2$, NOG or other 2OG analogues, and substrate to the protein solution, were removed by centrifugation (14,000 rpm, 4 °C, 2 x 5 min with transfer to a fresh tube in between).

A Phoenix RE crystallisation robot (*Art Robbins Instruments*) was used to set up 96-well crystallisation plates (Intelliplate Low Profile Protein Crystallisation Tray 96-3 well, *Alpha Biotech Ltd.*) with sitting drops of 100-200 nl protein and various *Hampton Research* Molecular Dimensions crystallisation screens (JCSG, PACT, INDEX, PEG/Ion, Structure, Salt, MIDAS) or prepared optimisation screens, with protein to precipitants mix ratios of 2:1, 1:1, and 1:2 (300 nl, 200 nl, or 300 nl total volume respectively). Plates were incubated at 4 °C or 20 °C (277 K or 293 K), and crystals were grown at 4 °C by the vapour diffusion method (no crystals were obtained at 20 °C).

Crystals were harvested using nylon loops under cryoprotection using well solution with 25 % (v/v) glycerol and flash-frozen in liquid nitrogen. Data (typically 0.2 ° oscillation/image, 180 °

total rotation) were collected on single crystals using in-house X-ray radiation (copper rotating anode CuKα 1.542 Å) from a *Rigaku* FR-E+ SuperBright™ diffractometer, equipped with Osmic HF optics and Saturn 944+ CCD detector, at 100 K (with flash-cooling in an *Oxford Cryosystems* N_2 gas stream) or at Diamond Light Source synchrotron science facility MX beamlines (Macromolecular Crystallography, I02, I04, I04-1), Didcot, at the respective wavelengths as indicated in the data collection table for each dataset.

10.5.11 Structure Determination

Data were indexed, integrated and scaled using HKL-3000®.[595] The "apo" structure of $AspH_{315-758}$ including both oxygenase and TPR domains was determined by the molecular replacement (MR) method using the MR_Auto (Phaser)[596] subroutine in PHENIX[423] (against the "apo" structural data) performed by Dr Michael A. McDonough. The coordinates of the AspH oxygenase domain ($AspH_{562-758}$, current PDB entry 3RCQ) were first used to successfully identify a rotation and translation solution for the oxygenase domain (RFZ X, TFZ X, LLG X). Initial attempts to find a rotation and translation solution for the TPR using the *C*-terminal TPR domain of peroxisomal targeting signal 1 receptor (PEX5) were complicated by the repeating nature of the TPR fold. In order to overcome the problems associated with the repeating fold, the search model was trimmed to include three sequential repeats (PEX5 Gln3000-Tyr395). The three *C*-terminal repeats of $AspH_{315-758}$ were identified (RFZ, TFZ, LLG) in a search after the oxygenase domain was fixed. This allowed manual fitting of three further TPR repeats to the *N*-terminus of $AspH_{315-758}$. Rigid body refinement of the three separate fragments followed by simulated annealing refinement (both cartesian and torsion) led to a model with R_{free} 42.15. Iterative rounds of model building in COOT[422] and refinement in PHENIX were performed until the decreasing *R*-factor and R_{free} no longer converged and the final model was obtained. Due to domain movements between different crystal forms and complexes, all other structures were determined by MR using PHASER with the oxygenase and TPR domains searched for independently. The oxygenase domain was always searched first followed by the TPR domain.

10.5.12 [1-^{14}C]-2OG Turnover Assays

An adaptation of reported methods[428,393] was used for [1-^{14}C]-2OG turnover assays. A mixture of 50 mM HEPES, pH 7.5, 400 µM sodium ascorbate, and 300 µM 2OG containing 8 % [1-^{14}C]-2OG (*Perkin Elmer*) was prepared and dispensed into a 5 ml polycarbonate tube. Typically 100 µM substrate and an 'Enzyme or Control Mix' with 4 µM enzyme and 50 µM Fe(II), made from $(NH_4)_2Fe(SO_4)_2{\cdot}7\ H_2O$, were dispensed as separate spots in the tube. The total assay volume was 100 µl. A 0.2 ml Eppendorf tube with 200 µl of hyamine hydroxide was placed in a 0.5 ml Eppendorf

tube and then placed into the 5 ml tube to capture the evolved $^{14}CO_2$. A Suba-Seal® septum was used to close the tube during the incubation. Reactions were initiated by tube shaking time and incubated for 40 min at 37 °C. As reaction stop reagent 100 µl of 20% formic acid was added, and the tubes were further incubated for 20 min at 22 °C after the addition. The 0.2 ml Eppendorf tube containing hyamine hydroxide was removed and placed into a scintillation vial with 5 ml of scintillation fluid (ScintiSafe, *Fisher Scientific*), mixed well by shaking, and scintillation was then counted using a *Beckman Coulter* LS6500 Multipurpose Scintillation Counter.

10.5.13 2OG Turnover Assays by NMR

An adaptation of reported methods[409] was used for 2OG turnover assays monitored using NMR spectroscopy. A mixture of 1 mM sodium ascorbate, 2 mM 2OG, 100 µM Fe(II), 50 µM substrate (made up in H_2O, other reagents made up in deuterated Tris except Fe(II), which was made up in 20 mM HCl, then diluted in ^2H_2O), buffer 50 mM [$^2H_{11}$]-Tris-^{2}HCl, p^{2}H 7.5 (pH 7.1) was prepared, and 10 µM enzyme was added last (total assay volume 75 µl), just prior to transfer into a 2 mm Bruker Match™ NMR tube (*Hilgenberg GmbH*) and analysis on a *Bruker* Avance III 700 MHz spectrometer, operated with TopSpin™ 3.2 software. All spectra were acquired at 298 K. The lag time between mixing and the first NMR analysis was typically 180 s-480 s.

10.5.14 NMR Analysis of Hydroxylated CP2$_{10mer}$

For the analysis of the hydroxylated CP2$_{10mer}$ peptide, the peptide was fully hydroxylated by incubation with AspH in presence of Fe(II), 2OG, Asc in a typical hydroxylation activity assay (conducted for 120 min) as described above. The reaction was quenched by the addition of an equal volume of 1 % formic acid, and the peptide then purified by SPE using a Sep-Pak C_{18} Plus Light Cartridge (*Waters*) as described above. The eluted sample was dried in Vacufuge® vacuum concentrator (*Eppendorf*), the residue then taken up in 20 µl ^2H_2O, flash-frozen in liquid N_2 and lyophilised. The peptide was then reconstituted in 75 µl ^2H_2O or 75 µl [$^2H_6$]-DMSO and analysed in 2 mm Bruker Match™ NMR tubes (*Hilgenberg GmbH*) on a *Bruker* Avance III 700 MHz spectrometer, operated with TopSpin™ 3.2 software.

10.6 MAMMALIAN CELL-BASED METHODS

10.6.1 Mammalian Cell Culture

Mammalian cell culture procedures were carried out in a *NuAire* LabGard ES (Energy Saver) NU 437 Class II, Type A2 Biosafety Cabinet. *H. sapiens* HeLa (human cervical carcinoma), HEK 293T (human

embryonic kidney), U-2 OS (human osteosarcoma), and HepG2 (human hepatocellular carcinoma) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, *Lonza*), supplemented with 10% fetal bovine serum (FBS, *Sigma-Aldrich*) and 1 % GlutaMAX™-I (*Life Technologies*), in a *BINDER* BD-53 incubator with a humidified 5 % CO_2 atmosphere at 37 °C until confluence was attained. Phosphate buffered saline (PBS, *Lonza*) was used as buffer for wash steps.

10.6.1.1 Cell Storage

Cells were frozen in aliquots in DMEM containing 20 % FBS and 10 % DMSO using Mr. Frosty™ Freezing Container (*Thermo Scientific*) stored at -80 °C for short term storage and then in a liquid N_2 dewar for long term storage. For cell culture use, an aliquot of cells was then thawed and added into 10 ml of pre-warmed media in a 15 ml falcon tube, mixed well, and then centrifuged (1.5 rpm, 3 min, 22 °C). The supernatant containing media and DMSO was removed, the cells were resuspended in 5 ml of fresh media and added to an appropriate vessel for cell culture at 37 °C.

10.6.1.2 Cell Passaging

At sufficient confluence, strongly adherent HeLa, U-2 OS, and HepG2 cells were washed with PBS and then enzymatically removed from the vessel using 0.2 mg/ml Trypsin-EDTA (*Lonza*) at 37 °C. The cells were then taken up in fresh media to stop the reaction (as the medium contains the trypsin inhibitors Mg^{2+} and Ca^{2+}) and resuspended by pipetting to break up cell clumps prior to dispensing the appropriate volume of cells into new vessels containing fresh pre-warmed media.
Semi-adherent HEK 293T cells were gently washed with PBS, then removed from the vessel by resuspension in fresh media and dispensed into new vessels containing fresh pre-warmed media.

Table 10-25 | Cell culture vessels and typical cell numbers.

Culture Vessel	Surface area [cm^2]	Approximate number of cells at 100 % confluence	Media volume [ml]
T25 flask	25	2.5×10^6	7
12-well plate	3.8	3.8×10^5	1
6-well plate	9.5	9.5×10^5	2
10 cm diameter petri dish	55	5.5×10^6	10
15 cm diameter petri dish	148	1.48×10^7	20

10.6.2 Transfection

Typically 2h prior to the desired transfection, the medium was removed from the cells and replaced with fresh media to initiate cell division. HeLa and HEK 293T cells were transiently transfected at ~30-50 % confluence (Forward transfection) using FuGENE® HD Transfection Reagent (*Promega*) and typically 1 µg plasmid DNA in 100 µl Opti-MEM® I Reduced Serum Media (*Life Technologies*) per 1 ml of media at a FuGENE® HD to DNA ratio of 3:1 (following the reagent manufacturer's protocol) when grown on sterile glass coverslips (15-18 mm) for immunofluorescence in 12-well or 6-well plates. DNA was diluted in Opti-MEM® I, then mixed well with transfection reagent and incubated for 15 min at 22 °C prior to dropwise addition to the cells, which were then incubated at 37 °C for typically further 24-48 h. For transfection of cells grown in 15 cm diameter petri dishes, 1.0 mg/ml linear polyethyleneimine (linear PEI, *PolySciences Inc.*, made up in hot water and filter-sterilised using a 0.2 µm syringe filter) and 20 µg DNA in 100 µl Opti-MEM® I at a linear PEI to DNA ratio of 12:1 was used. Due to the potential cell toxicity of the reagent, media containing linear PEI was exchanged to fresh media 3-4 h after transfection, and cells were incubated at 37 °C for 24-48 h in total.

Reverse transfection was employed for HepG2 cells, which were washed with PBS, then trypsinised and resuspended in fresh media prior to plating in the appropriate cell culture dishes containing transfection reagent and DNA in Opti-MEM® I prepared as described above (except that a FuGENE® HD to DNA ratio of 4:1 was used). Cells were then incubated for 36-48 h. When linear PEI was used for transfection, the media was exchanged to fresh media after 16 h. Cells on coverslips for immunofluorescence were processed at 60-80 % confluence.

10.6.3 Immunofluorescence

Typically 24h after transfection using FuGENE® HD as described above, the medium was aspirated and the live cells grown on coverslips were washed twice with PBS prior to fixation using 3.7 % paraformaldehyde made up in PBS (made fresh from a 15 % stock solution in PBS stored in the dark at -20 °C) for 15 min at 22 °C, followed by three washes with PBS, and subsequent permeabilisation using 0.1 % Triton-X for 4 min at 22 °C. Alternatively, fixation and permeabilisation were performed using ice-cold MeOH for 6 min under chilling on ice. After three washes with PBS, the coverslips were incubated in 0.2 % fish skin gelatin (FSG) blocking solution (0.2 % Gelatin from cold water fish skin, *Sigma-Aldrich*, in PBS) for 30 min, followed by incubation with primary antibodies in FSG solution (typically F3165 SIGMA Monoclonal ANTI-FLAG® M2 antibody produced in mouse, *Sigma-Aldrich*, dilution 1 in 500; ADI-SPA-860 Calnexin, pAb antibody produced in rabbit, *Enzo Life Sciences*, dilution 1 in 200; anti-HSPA8, i.e. HSC 70 (K-19): sc-1059, *Santa Cruz Biotechnology, Inc.*, produced in rabbit, dilution 1 in 50) for 60 min at 22 °C in the dark in a

humidified chamber. After three washes with FSG solution, the coverslips were incubated with secondary antibodies (typically Alexa Fluor® 488 Goat Anti-Mouse IgG (H+L) Antibody, and Alexa Fluor® 594 Goat Anti-Rabbit IgG (H+L) Antibody, *Molecular Probes® Life Technologies*) for 45 min at 22 °C in the dark in the chamber, followed by three washed with FSG solution and staining with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI, *Sigma-Aldrich*, 0.2 µg/ml in FSG solution) for 5 min at 22 °C. After three washes with FSG solution and a final wash in Milli-Q H_2O, the coverslips were mounted on glass slides using ProLong® Gold Antifade Reagent (*Life Technologies*).

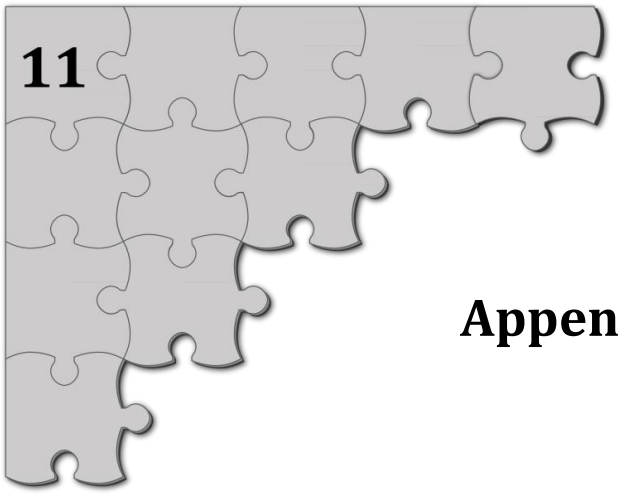
Cells on the prepared slides were then imaged by oil immersion (Immersion Oil 1.514, *GE Healthcare*) on a DeltaVision Elite Imaging System wide-field fluorescence microscope (*Applied Precision*, *GE Healthcare*) with a 100 x objective and CoolSNAP HQ2 CCD camera (*Photometrics*) operated with softWoRx Suite 2.0 (*Applied Precision, GE Healthcare*; typical settings: aperture 100 % open, illumination always Köhler, Alexa filter sets, polychroic Quad-mch, 100 % transmission, image size 1024 x 1024, varying exposure times as appropriate, image deconvolution: Ratio conservative) using the parameters given in **Table 10-25**.

Table 10-26 | Fluorescence microscopy excitation and emission parameters.

Fluorophore	Excitation wavelength [nm]	λ_{max} absorption [nm]	λ_{max} emission [nm]
DAPI	445	345	455
Alexa Fluor® 488	525	495	519
Alexa Fluor® 594	543	590	617

10.6.4 cDNA Preparation

For cDNA preparation, RNA was extracted using a RNeasy Plus Mini Kit (*QIAGEN*, including gDNA Eliminator column) following the manufacturer's protocol. HeLa, U-2 OS, or HepG2 cDNA was prepared from a 10 cm diameter petri dish (approximately 5.5 x 10^6 cells) and used for cloning of human flAspH$_{1-758}$ by LIC cloning as described above. RNase-free tubes and water (Nuclease-Free Water, not DEPC-treated, *Ambion®*, *Life Technologies*) were used throughout, and the working area, pipettes and gloves were wiped with RNase*Zap*® RNase Decontamination Solution (*Ambion®*, *Life Technologies*) prior to start. A High Capacity cDNA Reverse Transcription Kit (*Applied Biosystems*, including RNase inhibitor) was then used to synthesise cDNA from 2 µg of the extracted RNA in 20 µl reactions according to the manufacturer's protocol with thermal cycling using a Genius Thermal Cycler (*Techne*) for 10 min at 25 °C, then 120 min at 37 °C, followed by 5 min at 85 °C, and final cooling to 4 °C.



Appendices

A PROTEIN AND DNA SEQUENCES

GST-AspH$_{315-758}$

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKH
NMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLY
DALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSR
KTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRR
SNEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDN
DNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNKEAYKW
YELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLR
EKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLV
IPKEGCKIRCANETKTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI

Blue = GST tag Green = TPR domain Red = AspH oxygenase domain

M_W = 77,930.6 Da; 675 aa

His$_6$-AspH$_{562-758}$

MHHHHHHSSGVDLGTENLYFQSMRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDEGLAVMDKAKGL
FLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNC
RLRMHLGLVIPKEGCKIRCANETKTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI

Blue = His$_6$ tag Red = AspH oxygenase domain

M_W = 25,601.1 Da; 220 aa

His$_6$-AspH$_{315-758}$

GSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRG
KIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFL
GHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYL
KEGIESGDPGTDDGRFYFHLGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGY
TELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETT
GCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETKTWEEGKVLIFDDSFEHEVWQDASS
FRLIFIVDVWHPELTPQQRRSLPAI

Blue = His$_6$ tag Green = TPR domain Red = AspH oxygenase domain

M_W = 54,519.1 Da; 477 aa

B SUPPLEMENTARY ASPH ISOFORMS SEQUENCE ALIGNMENTS

The distinctive AspH splicing isoforms have been introduced in **Section 1.3.3**. The *N*-termini are on the cytoplasmic side, the *C*-termini localised to the ER lumen. All isoforms contain a single-pass transmembrane region followed by the luminal EF-hand motif (Ca^{2+}-binding). The following section provides descriptions of the domain architecture of all isoforms.

AspHA: flAspH

Key features: Serine-rich cytoplasmic tail (this may be phosphorylated or *O*-linked glycosylated); humbug/junctate domain with Glu-rich luminal segment (i.e. predominantly negative charges in luminal part of this domain); TPR domain, which contains a poly-Lys motif of four adjacent Lys residues (and further Lys nearby) at the *N*-terminal start; an oxygenase domain with the catalytic site (with potentially glycosylated residues[342] *N*-linked GlcNAc on N452 within the TPR domain and N706 within the oxygenase domain).

AspHB: Junctate

Key features: Very short cytoplasmic tail lacking the serine-rich segment; has a Glu-rich luminal segment identical to humbug and corresponding segment of flAspH apart from small internal difference (AKAKDFRYNLSEVLQ on luminal side preceding EF-hand motif); the TPR and oxygenase domains are absent.

AspHC: Humbug

Key features: Ser-rich cytoplasmic tail as AspHA; humbug/junctate domain with Glu-rich luminal segment; the TPR and oxygenase domains are absent.

AspHF: alternative full-length isoform

Key features: Very short cytoplasmic tail lacking the serine-rich segment; humbug/junctate domain with a Glu-rich luminal segment (i.e. predominantly negative charges in luminal part of this domain); TPR domain, which contains a poly-Lys motif at the *N*-terminal start, and an oxygenase domain (with potentially glycosylated residues[342] *N*-linked GlcNAc on N452 within the TPR domain and N706 within the oxygenase domain).

AspHD: Junctin, AspHE: Junctin-1 (cardiac junctin)

Key features: Very short cytoplasmic tail identical to junctate (lacking the serine-rich segment); the luminal segment of this domain is atypical, as it is Lys-rich (i.e. predominantly positive charges in the luminal part of this domain); the TPR and oxygenase domains are absent; AspHE: junctin-1 (cardiac junctin) differs from AspHD only by the same small internal difference (-AKAKDFRYNLSEVLQ- on the luminal side preceding the EF-hand motif) that is also found in junctate and some other isoforms.

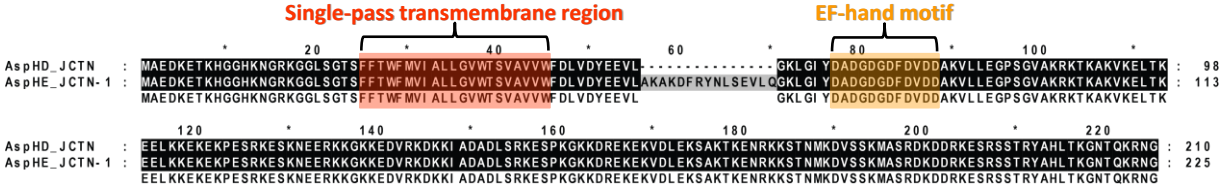


Fig. S-1 | Sequence alignment of selected AspH isoforms, demonstrating that AspHD = junctin and AspHE = junctin-1 (cardiac junctin) differ in a small internal segment only.

AspHG, AspHH, AspHI: Junctate-like isoforms

Key features: *N*-terminus identical to junctate with a very short cytoplasmic tail lacking the serine-rich segment. As junctate, the isoforms AspHG, AspHH, and AspHI all contain the internal segment AKAKDFRYNLSEVLQ on the luminal side preceding the EF-hand motif. AspHF lacks the internal segment although it shares the same *N*-terminus with the junctate-like isoforms. Compared to junctate, further internal segments are missing in AspHH and AspHI, and a single amino acid deletion is found in AspHG. TPR and oxygenase domains are absent in AspHG, AspHH, and AspHI.

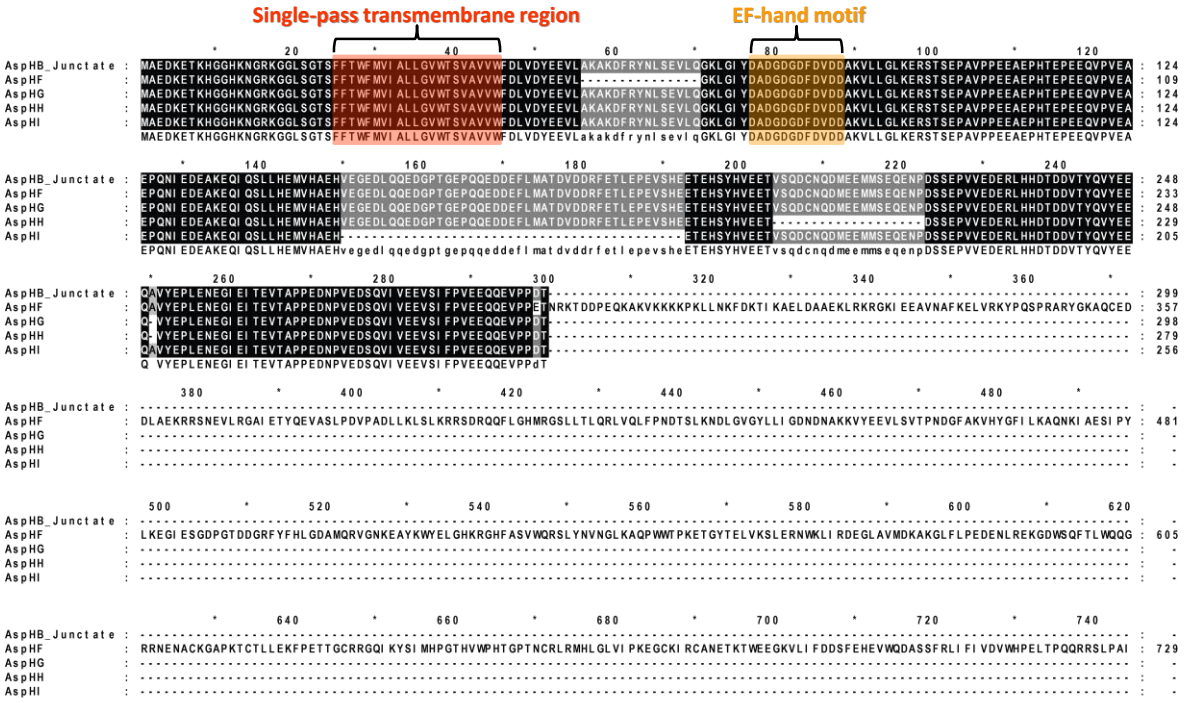


Fig. S-2 | Sequence alignment of selected AspH isoforms, demonstrating that junctate-like isoforms are all very similar and only differ in length missing different internal segments.

AspHJ, AspHK: Humbug-like isoforms

Key features: Same Ser-rich cytoplasmic tail as AspHA and humbug; humbug/junctate domain with Glu-rich luminal segment; the TPR and oxygenase domains are absent; compared to humbug, further internal segments are missing in AspHH and AspHI.

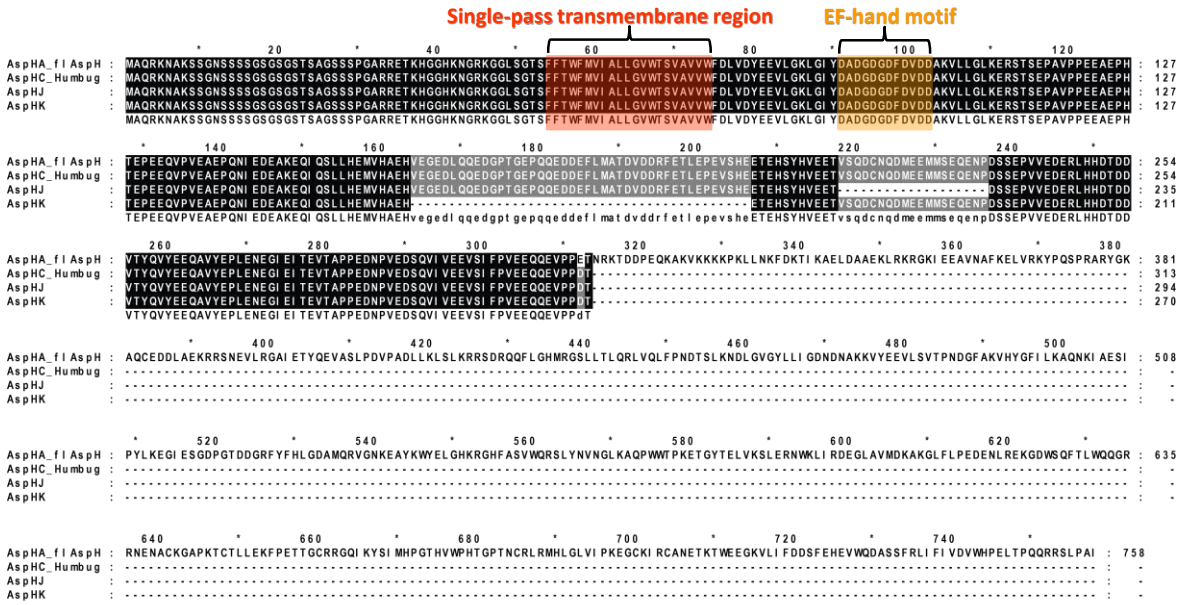


Fig. S-3 | Sequence alignment of selected AspH isoforms, demonstrating that humbug-like isoforms are all very similar and only differ in the same missing internal segments as found in junctate-like isoforms.

AspHL:

Key features: Same Ser-rich cytoplasmic tail as AspHA and humbug but the luminal segment following the EF-hand motif is very short and has no similarity to humbug/junctate although this region is also Glu-rich (contains a poly-Glu motif of six adjacent Glu residues), the TPR and oxygenase domains are absent.

Permission for the reproduction of a figure: Chapter 2, Fig. 2-45a

Title: Absence of Post-translational Aspartyl β-Hydroxylation of Epidermal Growth Factor Domains in Mice Leads to Developmental Defects and an Increased Incidence of Intestinal Neoplasia
Authors: Joseph E. Dinchuk, Richard J. Focht, Jennifer A. Kelley, *et al.*
Publication: Journal of Biological Chemistry
Publisher: The American Society for Biochemistry and Molecular Biology
Date: Apr 12, 2002

11200 Rockville Pike
Suite 302
Rockville, Maryland 20852

August 19, 2011

American Society for Biochemistry and Molecular Biology

To whom it may concern,

It is the policy of the American Society for Biochemistry and Molecular Biology to allow reuse of any material published in its journals (the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research) in a thesis or dissertation at no cost and with no explicit permission needed. Please see our copyright permissions page on the journal site for more information.

Best wishes,

Sarah Crespi

American Society for Biochemistry and Molecular Biology
11200 Rockville Pike, Rockville, MD
Suite 302
240-283-6616
JBC | MCP | JLR

Tel: 240-283-6600 • Fax: 240-881-2080 • E-mail: asbmb@asbmb.org

Permission for the reproduction of a figure: Chapter 2, Fig. 2-45b

ELSEVIER LICENSE
TERMS AND CONDITIONS

This is a License Agreement between Inga Pfeffer ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Inga Pfeffer
Customer address	University of Oxford Oxford, OX1 3TA
License number	3473770558959
License date	Sep 21, 2014
Licensed content publisher	Elsevier
Licensed content publication	The American Journal of Human Genetics
Licensed content title	Mutations in ASPH Cause Facial Dysmorphism, Lens Dislocation, Anterior-Segment Abnormalities, and Spontaneous Filtering Blebs, or Traboulsi Syndrome
Licensed content author	Nisha Patel,Arif O. Khan,Ahmad Mansour,Jawahir Y. Mohamed,Abdullah Al-Assiri,Randa Haddad,Xiaofei Jia,Yong Xiong,André Mégarbané,Elias I. Traboulsi,Fowzan S. Alkuraya
Licensed content date	1 May 2014
Licensed content volume number	94
Licensed content issue number	5
Number of pages	5
Start Page	755
End Page	759
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of your thesis/dissertation	Interplay between 2-Oxoglutarate Oxygenases and Cancer
Expected completion date	Oct 2014
Estimated size (number of pages)	300
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 GBP
VAT/Local Sales Tax	0.00 GBP / 0.00 GBP
Total	0.00 GBP

REFERENCES

1. Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000,** *100* (1), 57-70.
2. Kroemer, G.; Pouyssegur, J., Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell* **2008,** *13* (6), 472-82.
3. Croce, C. M., Oncogenes and cancer. *N. Engl. J. Med.* **2008,** *358* (5), 502-11.
4. Vander Heiden, M. G.; Lunt, S. Y.; Dayton, T. L.*, et al.*, Metabolic Pathway Alterations that Support Cell Proliferation. *Cold Spring Harb Symp Quant Biol* **2012**.
5. Jordan, C. T., Cancer stem cell biology: from leukemia to solid tumors. *Curr. Opin. Cell Biol.* **2004,** *16* (6), 708-12.
6. Hanahan, D.; Weinberg, R. A., Hallmarks of cancer: the next generation. *Cell* **2011,** *144* (5), 646-74.
7. Kinzler, K. W.; Vogelstein, B., Landscaping the cancer terrain. *Science* **1998,** *280* (5366), 1036-7.
8. Giancotti, F. G.; Ruoslahti, E., Integrin signaling. *Science* **1999,** *285* (285), 5430.
9. Aplin, A. E.; Howe, A.; Alahari, S. K.*, et al.*, Signal transduction and signal modulation by cell adhesion receptors: the role of integrins, cadherins, immunoglobulin-cell adhesion mole-cules, and selectins. *Pharmacol. Rev.* **1998,** *50* (2), 197-263.
10. Lukashev, M. E.; Werb, Z., ECM signaling: orchestrating cell behaviour and misbehaviour. *Trends Cell Biol.* **1998,** *8* (11), 437-41.
11. Olumi, A. F.; Grossfeld, G. D.; Hayward, S. W.*, et al.*, Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* **1999,** *59* (19), 5002-11.
12. Fynan, T. M.; Reiss, M., Resistance to inhibition of cell growth by transforming growth factor-β and its role in oncogenesis. *Crit. Rev. Oncog.* **1993,** *4* (5), 493-540.
13. Weinberg, R. A., The retinoblastoma protein and cell cycle control. *Cell* **1995,** *81* (3), 323-30.
14. Moses, H. L.; Yang, E. Y.; Pietenpol, J. A., TGF-β stimulation and inhibition of cell proliferation: new mechanistic insights. *Cell* **1990,** *63* (2), 245-7.
15. Kahn, P.; Frykberg, L.; Brady, C.*, et al.*, v-erbA cooperates with sarcoma oncogenes in leukemic cell transformation. *Cell* **1986,** *45* (3), 349-56.
16. Harris, C. C., p53 tumor suppressor gene: from the basic research laboratory to the clinic—an abridged historical perspective. *Carcinogenesis* **1996,** *17* (6), 1187-98.
17. Levine, A. J., p53, the cellular gatekeeper for growth and division. *Cell* **1997,** *88* (3), 323-31.
18. Ishizaki, Y.; Cheng, L.; Mudge, A. W.*, et al.*, Programmed cell death by default in embryonic cells, fibroblasts, and cancer cells. *Mol. Biol. Cell* **1995,** *6* (11), 1443-58.
19. Vaux, D. L.; Cory, S.; Adams, T. M., Bcl-2 promotes the survival of hematopoietic cells and cooperates with c-myc to immortalize pre-B cells. *Nature* **1988,** *335* (6189), 440-2.
20. Wyllie, A. H.; Kerr, J. F.; Currie, A. R., Cell death: the significance of apoptosis. *Int. Rev. Cytol.* **1980,** *68*, 251-306.
21. Lotem, J.; Sachs, L., Control of apoptosis in hematopoiesis and leukemia by cytokines, tumor suppressor and oncogenes. *Leukemia* **1996,** *10* (6), 925-31.
22. Sporn, M. B., The war on cancer. *Lancet* **1996,** *347* (9012), 1377-81.
23. Christofori, G.; Naik, P.; Hanahan, D., A second signal supplied by insulin-like growth factor II in oncogene-induced tumorigenesis. *Nature* **1994,** *369* (6479), 414-8.
24. Johnson, J. P., Cell adhesion molecules of the immunoglobulin supergene family and their role in malignant transformation and progression to metastatic disease. *Cancer Metastasis Rev.* **1991,** *10* (1), 11-22.
25. Coussens, L. M.; Werb, Z., Matrix metalloproteinases and the development of cancer. *Chem. Biol.* **1996,** *3* (11), 895-904.
26. Fischer, K.; Hoffmann, P.; Voelkl, S.*, et al.*, Inhibitory effect of tumor cell-derived lactic acid on human T cells. *Blood* **2007,** *109* (9), 3812-9.
27. Cramer, T.; Yamanishi, Y.; Clausen, B. E.*, et al.*, HIF-1alpha is essential for myeloid cell-mediated inflammation. *Cell* **2003,** *112* (5), 645-57.
28. Puccetti, P.; Grohmann, U., IDO and regulatory T cells: a role for reverse signalling and non-canonical NF-kappaB activation. *Nat. Rev. Immunol.* **2007,** *7* (10), 817-23.

29. Wright, W. E.; Pereira-Smith, O. M.; Shay, J. W., Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Mol. Cell. Biol.* **1989,** *9* (7), 3088-92.
30. Hayflick, L., Mortality and immortality at the cellular level. A review. *Biochemistry* **1997,** *62* (11), 1180-90.
31. Stetler-Stevenson, W. G., Matrix metalloproteinases in angiogenesis: a moving target for therapeutic intervention. *J. Clin. Invest.* **1999,** *103* (9), 1237-41.
32. Vander Heiden, M. G.; Locasale, J. W.; Swanson, K. D.*, et al.*, Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science* **2010,** *329* (5998), 1492-9.
33. Airley, R. E.; Mobasheri, A., Hypoxic regulation of glucose transport, anaerobic metabolism and angiogenesis in cancer: novel pathways and targets for anticancer therapeutics. *Chemotherapy* **2007,** *53* (4), 233-56.
34. Kaelin, W. G.; Thompson, C. B., Q&A: Cancer: Clues from cell metabolism. *Nature* **2010,** *465* (7298), 562-564.
35. Lehninger, A. L.; Nelson, D. L.; Cox, M. M., *Principles of Biochemistry*. 5th ed.; W. H. Freeman: New York, 2008.
36. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Biochemistry*. 6th ed.; W. H. Freeman and Company: New York, 2007.
37. Warburg, O., Posener, K., Negelein, E., *Biochem Z* **1924,** *152*, 1319.
38. Warburg, O., On the origin of cancer cells. *Science* **1956,** *123* (3191), 309-14.
39. Vander Heiden, M. G.; Cantley, L. C.; Thompson, C. B., Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **2009,** *324* (5930), 1029-1033.
40. Gao, P.; Zhang, H.; Dinavahi, R.*, et al.*, HIF-dependent antitumorigenic effect of antioxidants in vivo. *Cancer Cell* **2007,** *12* (3), 230-8.
41. Zhou, S.; Kachhap, S.; Sun, W.*, et al.*, Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. *Proc. Natl. Acad. Sci. USA* **2007,** *104* (18), 7540-5.
42. Chowdhury, R.; Hardy, A.; Schofield, C. J., The human oxygen sensing machinery and its manipulation. *Chem. Soc. Rev.* **2008,** *37* (7), 1308-19.
43. Lando, D.; Gorman, J. J.; Whitelaw, M. L.*, et al.*, Oxygen-dependent regulation of hypoxia-inducible factors by prolyl and asparaginyl hydroxylation. *Eur. J. Biochem.* **2003,** *270* (5), 781-90.
44. Lu, H.; Forbes, R. A.; Verma, A., Hypoxia-inducible factor 1 activation by aerobic glycolysis implicates the Warburg effect in carcinogenesis. *J. Biol. Chem.* **2002,** *277* (26), 23111-5.
45. Schofield, C. J.; Ratcliffe, P. J., Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (5), 343-54.
46. Hewitson, K. S.; Schofield, C. J., The HIF pathway as a therapeutic target. *Drug Discov. Today* **2004,** *9* (16), 704-11.
47. Wenger, R. H.; Stiehl, D. P.; Camenisch, G., Integration of oxygen signaling at the consensus HRE. *Sci STKE* **2005,** *2005* (306), re12.
48. Loenarz, C.; Schofield, C. J., Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **2008,** *4* (3), 152-6.
49. Hausinger, R. P., Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **2004,** *39* (1), 21-68.
50. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.*, et al.*, Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* **2006,** *100* (4), 644-69.
51. Costas, M.; Mehn, M. P.; Jensen, M. P.*, et al.*, Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem Rev* **2004,** *104* (2), 939-86.
52. Zhang, Z.; Ren, J.; Stammers, D. K.*, et al.*, Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **2000,** *7* (2), 127-33.
53. Zhou, J.; Gunsior, M.; Bachmann, B. O.*, et al.*, Substrate Binding to the α-Ketoglutarate-Dependent Non-Heme Iron Enzyme Clavaminate Synthase 2: Coupling Mechanism of Oxidative Decarboxylation and Hydroxylation. *J. Am. Chem. Soc.* **1998,** *120* (51), 13539–40.
54. Lloyd, M. D.; Lee, H. J.; Harlos, K.*, et al.*, Studies on the active site of deacetoxycephalosporin C synthase. *J. Mol. Biol.* **1999,** *287* (5), 943-60.

55. Roach, P. L.; Clifton, I. J.; Hensgens, C. M.*, et al.*, Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **1997,** *387* (6635), 827-30.
56. Watanabe, M.; Sumida, N.; Murakami, S.*, et al.*, A phosphonate-induced gene which promotes Penicillium-mediated bioconversion of cis-propenylphosphonic acid to fosfomycin. *Appl. Environ. Microbiol.* **1999,** *65* (3), 1036-44.
57. Vaillancourt, F. H.; Yin, J.; Walsh, C. T., SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (29), 10111-6.
58. Baldwin, J. E.; Adlington, R. M.; Coates, J. B.*, et al.*, Purification and initial characterization of an enzyme with deacetoxycephalosporin C synthetase and hydroxylase activities. *Biochem. J.* **1987,** *245* (3), 831-41.
59. Flashman, E.; Schofield, C. J., The most versatile of all reactive intermediates? *Nat. Chem. Biol.* **2007,** *3* (2), 86-7.
60. Higley, M. J.; Sabatini, B. L., Calcium signaling in dendritic spines. *Cold Spring Harb Perspect Biol* **2012,** *4* (4), a005686.
61. Ozer, A.; Bruick, R. K., Non-heme dioxygenases: cellular sensors and regulators jelly rolled into one? *Nat. Chem. Biol.* **2007,** *3* (3), 144-53.
62. McDonough, M. A.; Loenarz, C.; Chowdhury, R.*, et al.*, Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr. Opin. Struct. Biol.* **2010,** *20* (6), 659-72.
63. Hegg, E. L.; Que, L., Jr., The 2-His-1-Carboxylate Facial Triad - An Emerging Structural Motif in Mononuclear Non-Heme Iron(II) Enzymes. *Eur. J. Biochem.* **1997,** *250* (3), 625-9.
64. Aik, W.; McDonough, M. A.; Thalhammer, A.*, et al.*, Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **2012,** *22* (6), 691-700.
65. Chowdhury, R.; McDonough, M. A.; Mecinovic, J.*, et al.*, Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* **2009,** *17* (7), 981-9.
66. Coleman, M. L.; McDonough, M. A.; Hewitson, K. S.*, et al.*, Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J. Biol. Chem.* **2007,** *282* (33), 24027-38.
67. Price, J. C.; Barr, E. W.; Glass, T. E.*, et al.*, Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). *J. Am. Chem. Soc.* **2003,** *125* (43), 13008-9.
68. Price, J. C.; Barr, E. W.; Tirupati, B.*, et al.*, The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2003,** *42* (24), 7497-508.
69. Zhou, J.; Kelly, W. L.; Bachmann, B. O.*, et al.*, Spectroscopic studies of substrate interactions with clavaminate synthase 2, a multifunctional alpha-KG-dependent non-heme iron enzyme: correlation with mechanisms and reactivities. *J. Am. Chem. Soc.* **2001,** *123* (30), 7388-98.
70. Borowski, T.; Bassan, A.; Siegbahn, P. E., Mechanism of dioxygen activation in 2-oxoglutarate-dependent enzymes: a hybrid DFT study. *Chemistry* **2004,** *10* (4), 1031-41.
71. Bugg, T. D.; Ramaswamy, S., Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. *Curr. Opin. Chem. Biol.* **2008,** *12* (2), 134-40.
72. Valegard, K.; van Scheltinga, A. C.; Lloyd, M. D.*, et al.*, Structure of a cephalosporin synthase. *Nature* **1998,** *394* (6695), 805-9.
73. Welford, R. W.; Kirkpatrick, J. M.; McNeill, L. A.*, et al.*, Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett.* **2005,** *579* (23), 5170-4.
74. Hoffart, L. M.; Barr, E. W.; Guyer, R. B.*, et al.*, Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (40), 14738-43.
75. Flashman, E.; Hoffart, L. M.; Hamed, R. B.*, et al.*, Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *FEBS J.* **2010,** *277* (19), 4089-99.
76. Semenza, G. L., Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. *Trends Mol. Med.* **2001,** *7* (8), 345-50.

77. Sanchez-Fernandez, E. M.; Tarhonskaya, H.; Al-Qahtani, K.*, et al.*, Investigations on the oxygen dependence of a 2-oxoglutarate histone demethylase. *Biochem. J.* **2013,** *449* (2), 491-6.
78. Tausendschon, M.; Dehne, N.; Brune, B., Hypoxia causes epigenetic gene regulation in macrophages by attenuating Jumonji histone demethylase activity. *Cytokine* **2011,** *53* (2), 256-62.
79. Pollard, P. J.; Loenarz, C.; Mole, D. R.*, et al.*, Regulation of Jumonji-domain-containing histone demethylases by hypoxia-inducible factor (HIF)-1alpha. *Biochem. J.* **2008,** *416* (3), 387-94.
80. Loenarz, C.; Schofield, C. J., Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **2011,** *36* (1), 7-18.
81. Semenza, G. L., HIF-1 and human disease: one highly involved factor. *Genes Dev.* **2000,** *14* (16), 1983-91.
82. Johansson, C.; Tumber, A.; Che, K.*, et al.*, The roles of Jumonji-type oxygenases in human disease. *Epigenomics* **2014,** *6* (1), 89-120.
83. Osmekhina, E.; Neubauer, A.; Klinzing, K.*, et al.*, Sandwich ELISA for quantitative detection of human collagen prolyl 4-hydroxylase. *Microb. Cell Fact.* **2010,** *9*, 48.
84. Gilkes, D. M.; Chaturvedi, P.; Bajpai, S.*, et al.*, Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Res.* **2013,** *73* (11), 3285-96.
85. Eisinger-Mathason, T. S.; Zhang, M.; Qiu, Q.*, et al.*, Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. *Cancer Discov* **2013,** *3* (10), 1190-205.
86. Rajkumar, T.; Sabitha, K.; Vijayalakshmi, N.*, et al.*, Identification and validation of genes involved in cervical tumourigenesis. *BMC Cancer* **2011,** *11*, 80.
87. Noda, T.; Yamamoto, H.; Takemasa, I.*, et al.*, PLOD2 induced under hypoxia is a novel prognostic factor for hepatocellular carcinoma after curative resection. *Liver Int.* **2012,** *32* (1), 110-8.
88. Dong, S.; Nutt, C. L.; Betensky, R. A.*, et al.*, Histology-based expression profiling yields novel prognostic markers in human glioblastoma. *J. Neuropathol. Exp. Neurol.* **2005,** *64* (11), 948-55.
89. Reis, P. P.; Waldron, L.; Perez-Ordonez, B.*, et al.*, A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. *BMC Cancer* **2011,** *11*, 437.
90. Mares, J.; Szakacsova, M.; Soukup, V.*, et al.*, Prediction of recurrence in low and intermediate risk non-muscle invasive bladder cancer by real-time quantitative PCR analysis: cDNA microarray results. *Neoplasma* **2013,** *60* (3), 295-301.
91. Cheng, L.; Wang, P.; Yang, S.*, et al.*, Identification of genes with a correlation between copy number and expression in gastric cancer. *BMC Med Genomics* **2012,** *5*, 14.
92. Schiarea, S.; Solinas, G.; Allavena, P.*, et al.*, Secretome analysis of multiple pancreatic cancer cell lines reveals perturbations of key functional networks. *J. Proteome Res.* **2010,** *9* (9), 4376-92.
93. Ho, J. R.; Chapeaublanc, E.; Kirkwood, L.*, et al.*, Deregulation of Rab and Rab effector genes in bladder cancer. *PloS one* **2012,** *7* (6), e39469.
94. Di Maro, G.; Maria Orlandella, F.; Claudio Bencivenga, T.*, et al.*, Identification of targets of Twist1 transcription factor in thyroid cancer cells. *J. Clin. Endocrinol. Metab.* **2014**, jc20133799.
95. Gilkes, D. M.; Bajpai, S.; Chaturvedi, P.*, et al.*, Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. *J. Biol. Chem.* **2013,** *288* (15), 10819-29.
96. Pan, P. W.; Zhang, Q.; Bai, F.*, et al.*, Profiling and comparative analysis of glycoproteins in Hs578BST and Hs578T and investigation of prolyl 4-hydroxylase alpha polypeptide II expression and influence in breast cancer cells. *Biochemistry (Mosc.)* **2012,** *77* (5), 539-45.
97. Chakravarthi, B. V.; Pathi, S. S.; Goswami, M. T.*, et al.*, The miR-124-Prolyl Hydroxylase P4HA1-MMP1 axis plays a critical role in prostate cancer progression. *Oncotarget* **2014**.
98. Xiong, G.; Deng, L.; Zhu, J.*, et al.*, Prolyl-4-hydroxylase alpha subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition. *BMC Cancer* **2014,** *14*, 1.
99. Mackay, A.; Jones, C.; Dexter, T.*, et al.*, cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene* **2003,** *22* (17), 2680-8.
100. Reis, P. P.; Waldron, L.; Goswami, R. S.*, et al.*, mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol.* **2011,** *11*, 46.

101. Chang, K. P.; Yu, J. S.; Chien, K. Y.*, et al.*, Identification of PRDX4 and P4HA2 as metastasis-associated proteins in oral cavity squamous cell carcinoma by comparative tissue proteomics of microdissected specimens using iTRAQ technology. *J. Proteome Res.* **2011,** *10* (11), 4935-47.
102. Jarzab, B.; Wiench, M.; Fujarewicz, K.*, et al.*, Gene expression profile of papillary thyroid cancer: sources of variability and diagnostic implications. *Cancer Res.* **2005,** *65* (4), 1587-97.
103. Berger, B.; Capper, D.; Lemke, D.*, et al.*, Defective p53 antiangiogenic signaling in glioblastoma. *Neuro Oncol* **2010,** *12* (9), 894-907.
104. Erez, N.; Milyavsky, M.; Eilam, R.*, et al.*, Expression of prolyl-hydroxylase-1 (PHD1/EGLN2) suppresses hypoxia inducible factor-1alpha activation and inhibits tumor growth. *Cancer Res.* **2003,** *63* (24), 8777-83.
105. Zhang, Q.; Gu, J.; Li, L.*, et al.*, Control of cyclin D1 and breast tumorigenesis by the EglN2 prolyl hydroxylase. *Cancer Cell* **2009,** *16* (5), 413-24.
106. Peurala, E.; Koivunen, P.; Bloigu, R.*, et al.*, Expressions of individual PHDs associate with good prognostic factors and increased proliferation in breast cancer patients. *Breast Cancer Res. Treat.* **2012,** *133* (1), 179-88.
107. Parker, L. P.; Taylor, D. D.; Kesterson, S.*, et al.*, Gene expression profiling in response to estradiol and genistein in ovarian cancer cells. *Cancer Genomics Proteomics* **2009,** *6* (3), 189-94.
108. Yi, Y.; Mikhaylova, O.; Mamedova, A.*, et al.*, von Hippel-Lindau-dependent patterns of RNA polymerase II hydroxylation in human renal clear cell carcinomas. *Clin Cancer Res* **2010,** *16* (21), 5142-52.
109. Hsu, Y. C.; Yuan, S.; Chen, H. Y.*, et al.*, A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung cancer. *Clin Cancer Res* **2009,** *15* (23), 7309-15.
110. Su, Y.; Loos, M.; Giese, N.*, et al.*, Prolyl hydroxylase-2 (PHD2) exerts tumor-suppressive activity in pancreatic cancer. *Cancer* **2012,** *118* (4), 960-72.
111. Heindryckx, F.; Kuchnio, A.; Casteleyn, C.*, et al.*, Effect of prolyl hydroxylase domain-2 haplodeficiency on the hepatocarcinogenesis in mice. *J. Hepatol.* **2012,** *57* (1), 61-8.
112. Kamphues, C.; Wittschieber, D.; Klauschen, F.*, et al.*, Prolyl hydroxylase domain 2 protein is a strong prognostic marker in human gastric cancer. *Pathobiology* **2012,** *79* (1), 11-7.
113. Jokilehto, T.; Rantanen, K.; Luukkaa, M.*, et al.*, Overexpression and nuclear translocation of hypoxia-inducible factor prolyl hydroxylase PHD2 in head and neck squamous cell carcinoma is associated with tumor aggressiveness. *Clin Cancer Res* **2006,** *12* (4), 1080-7.
114. Tennant, D. A.; Gottlieb, E., HIF prolyl hydroxylase-3 mediates alpha-ketoglutarate-induced apoptosis and tumor suppression. *J. Mol. Med. (Berl.)* **2010,** *88* (8), 839-49.
115. Sciorra, V. A.; Sanchez, M. A.; Kunibe, A.*, et al.*, Suppression of glioma progression by Egln3. *PloS one* **2012,** *7* (8), e40053.
116. Su, Y.; Loos, M.; Giese, N.*, et al.*, PHD3 regulates differentiation, tumour growth and angiogenesis in pancreatic cancer. *Br. J. Cancer* **2010,** *103* (10), 1571-9.
117. Tanaka, T.; Kitamura, H.; Torigoe, T.*, et al.*, Autoantibody against hypoxia-inducible factor prolyl hydroxylase-3 is a potential serological marker for renal cell carcinoma. *J. Cancer Res. Clin. Oncol.* **2011,** *137* (5), 789-94.
118. Su, C.; Huang, K.; Sun, L.*, et al.*, Overexpression of the HIF hydroxylase PHD3 is a favorable prognosticator for gastric cancer. *Med Oncol* **2012,** *29* (4), 2710-5.
119. Morris, M. R.; Maina, E.; Morgan, N. V.*, et al.*, Molecular genetic analysis of FIH-1, FH, and SDHB candidate tumour suppressor genes in renal cell carcinoma. *J Clin Pathol* **2004,** *57* (7), 706-11.
120. Chen, T.; Yao, L. Q.; Shi, Q.*, et al.*, MicroRNA-31 contributes to colorectal cancer development by targeting factor inhibiting HIF-1alpha (FIH-1). *Cancer Biol Ther* **2014,** *15* (5), 516-23.
121. Liu, C. J.; Tsai, M. M.; Hung, P. S.*, et al.*, miR-31 ablates expression of the HIF regulatory factor FIH to activate the HIF pathway in head and neck carcinoma. *Cancer Res.* **2010,** *70* (4), 1635-44.
122. Wang, E.; Zhang, C.; Polavaram, N.*, et al.*, The role of factor inhibiting HIF (FIH-1) in inhibiting HIF-1 transcriptional activity in glioblastoma multiforme. *PloS one* **2014,** *9* (1), e86102.
123. Hyseni, A.; van der Groep, P.; van der Wall, E.*, et al.*, Subcellular FIH-1 expression patterns in invasive breast cancer in relation to HIF-1alpha expression. *Cell Oncol (Dordr)* **2011,** *34* (6), 565-70.

124. Tan, E. Y.; Campo, L.; Han, C.*, et al.*, Cytoplasmic location of factor-inhibiting hypoxia-inducible factor is associated with an enhanced hypoxic response and a shorter survival in invasive breast cancer. *Breast Cancer Res* **2007,** *9* (6), R89.
125. Rizwani, W.; Schaal, C.; Kunigal, S.*, et al.*, Mammalian Lysine Histone Demethylase KDM2A Regulates E2F1-Mediated Gene Transcription in Breast Cancer Cells. *PloS one* **2014,** *9* (7), e100888.
126. He, J.; Nguyen, A. T.; Zhang, Y., KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* **2011,** *117* (14), 3869-80.
127. Tzatsos, A.; Paskaleva, P.; Ferrari, F.*, et al.*, KDM2B promotes pancreatic cancer via Polycomb-dependent and -independent transcriptional programs. *J Clin Invest* **2013,** *123* (2), 727-39.
128. Osawa, T.; Tsuchida, R.; Muramatsu, M.*, et al.*, Inhibition of histone demethylase JMJD1A improves anti-angiogenic therapy and reduces tumor-associated macrophages. *Cancer Res.* **2013,** *73* (10), 3019-28.
129. Qi, J.; Nakayama, K.; Cardiff, R. D.*, et al.*, Siah2-dependent concerted activity of HIF and FoxA2 regulates formation of neuroendocrine phenotype and neuroendocrine prostate tumors. *Cancer Cell* **2010,** *18* (1), 23-38.
130. Uemura, M.; Yamamoto, H.; Takemasa, I.*, et al.*, Jumonji domain containing 1A is a novel prognostic marker for colorectal cancer: in vivo identification from hypoxic tumor cells. *Clin Cancer Res* **2010,** *16* (18), 4636-46.
131. Yamada, D.; Kobayashi, S.; Yamamoto, H.*, et al.*, Role of the hypoxia-related gene, JMJD1A, in hepatocellular carcinoma: clinical impact on recurrence after hepatic resection. *Ann Surg Oncol* **2012,** *19 Suppl 3*, S355-64.
132. Guo, X.; Shi, M.; Sun, L.*, et al.*, The expression of histone demethylase JMJD1A in renal cell carcinoma. *Neoplasma* **2011,** *58* (2), 153-7.
133. Hu, Z.; Gomes, I.; Horrigan, S. K.*, et al.*, A novel nuclear protein, 5qNCA (LOC51780) is a candidate for the myeloid leukemia tumor suppressor gene on chromosome 5 band q31. *Oncogene* **2001,** *20* (47), 6946-54.
134. Kim, J. Y.; Kim, K. B.; Eom, G. H.*, et al.*, KDM3B is the H3K9 demethylase involved in transcriptional activation of lmo2 in leukemia. *Mol. Cell. Biol.* **2012,** *32* (14), 2917-33.
135. MacKinnon, R. N.; Kannourakis, G.; Wall, M.*, et al.*, A cryptic deletion in 5q31.2 provides further evidence for a minimally deleted region in myelodysplastic syndromes. *Cancer Genet* **2011,** *204* (4), 187-94.
136. Tang, M. H.; Varadan, V.; Kamalakaran, S.*, et al.*, Major chromosomal breakpoint intervals in breast cancer co-localize with differentially methylated regions. *Front Oncol* **2012,** *2*, 197.
137. Wolf, S. S.; Patchev, V. K.; Obendorf, M., A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. *Arch. Biochem. Biophys.* **2007,** *460* (1), 56-66.
138. Cloos, P. A.; Christensen, J.; Agger, K.*, et al.*, The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* **2006,** *442* (7100), 307-11.
139. Mallette, F. A.; Richard, S., JMJD2A promotes cellular transformation by blocking cellular senescence through transcriptional repression of the tumor suppressor CHD5. *Cell Rep* **2012,** *2* (5), 1233-43.
140. Kogure, M.; Takawa, M.; Cho, H. S.*, et al.*, Deregulation of the histone demethylase JMJD2A is involved in human carcinogenesis through regulation of the G(1)/S transition. *Cancer Lett.* **2013,** *336* (1), 76-84.
141. Black, J. C.; Manning, A. L.; Van Rechem, C.*, et al.*, KDM4A lysine demethylase induces site-specific copy gain and rereplication of regions amplified in tumors. *Cell* **2013,** *154* (3), 541-55.
142. Berry, W. L.; Shin, S.; Lightfoot, S. A.*, et al.*, Oncogenic features of the JMJD2A histone demethylase in breast cancer. *Int. J. Oncol.* **2012,** *41* (5), 1701-6.
143. Suikki, H. E.; Kujala, P. M.; Tammela, T. L.*, et al.*, Genetic alterations and changes in expression of histone demethylases in prostate cancer. *Prostate* **2010,** *70* (8), 889-98.
144. Shin, S.; Janknecht, R., Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D. *Biochem. Biophys. Res. Commun.* **2007,** *359* (3), 742-6.
145. Kim, T. D.; Shin, S.; Berry, W. L.*, et al.*, The JMJD2A demethylase regulates apoptosis and proliferation in colon cancer cells. *J. Cell. Biochem.* **2012,** *113* (4), 1368-76.

146. Hu, C. E.; Liu, Y. C.; Zhang, H. D.*, et al.*, JMJD2A predicts prognosis and regulates cell growth in human gastric cancer. *Biochem. Biophys. Res. Commun.* **2014,** *449* (1), 1-7.
147. Kauffman, E. C.; Robinson, B. D.; Downes, M. J.*, et al.*, Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol. Carcinog.* **2011,** *50* (12), 931-44.
148. Chu, C. H.; Wang, L. Y.; Hsu, K. C.*, et al.*, KDM4B as a Target for Prostate Cancer: Structural Analysis and Selective Inhibition by a Novel Inhibitor. *J. Med. Chem.* **2014,** *57* (14), 5975-85.
149. Toyokawa, G.; Cho, H. S.; Iwai, Y.*, et al.*, The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6. *Cancer Prev Res (Phila)* **2011,** *4* (12), 2051-61.
150. Shi, L.; Sun, L.; Li, Q.*, et al.*, Histone demethylase JMJD2B coordinates H3K4/H3K9 methylation and promotes hormonally responsive breast carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (18), 7541-6.
151. Kawazu, M.; Saso, K.; Tong, K. I.*, et al.*, Histone demethylase JMJD2B functions as a co-factor of estrogen receptor in breast cancer proliferation and mammary gland development. *PloS one* **2011,** *6* (3), e17830.
152. Berry, W. L.; Janknecht, R., KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer Res.* **2013,** *73* (10), 2936-42.
153. Yang, J.; Jubb, A. M.; Pike, L.*, et al.*, The histone demethylase JMJD2B is regulated by estrogen receptor alpha and hypoxia, and is a key mediator of estrogen induced growth. *Cancer Res.* **2010,** *70* (16), 6456-66.
154. Fu, L.; Chen, L.; Yang, J.*, et al.*, HIF-1alpha-induced histone demethylase JMJD2B contributes to the malignant phenotype of colorectal cancer cells via an epigenetic mechanism. *Carcinogenesis* **2012,** *33* (9), 1664-73.
155. Kim, J. G.; Yi, J. M.; Park, S. J.*, et al.*, Histone demethylase JMJD2B-mediated cell proliferation regulated by hypoxia and radiation in gastric cancer cell. *Biochim. Biophys. Acta* **2012,** *1819* (11-12), 1200-7.
156. Li, W.; Zhao, L.; Zang, W.*, et al.*, Histone demethylase JMJD2B is required for tumor cell proliferation and survival and is overexpressed in gastric cancer. *Biochem. Biophys. Res. Commun.* **2011,** *416* (3-4), 372-8.
157. Wissmann, M.; Yin, N.; Muller, J. M.*, et al.*, Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell Biol.* **2007,** *9* (3), 347-53.
158. Ishimura, A.; Terashima, M.; Kimura, H.*, et al.*, Jmjd2c histone demethylase enhances the expression of Mdm2 oncogene. *Biochem. Biophys. Res. Commun.* **2009,** *389* (2), 366-71.
159. Luo, W.; Chang, R.; Zhong, J.*, et al.*, Histone demethylase JMJD2C is a coactivator for hypoxia-inducible factor 1 that is required for breast cancer progression. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (49), E3367-76.
160. Yang, Z. Q.; Imoto, I.; Fukuda, Y.*, et al.*, Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. *Cancer Res.* **2000,** *60* (17), 4735-9.
161. Liu, G.; Bollig-Fischer, A.; Kreike, B.*, et al.*, Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. *Oncogene* **2009,** *28* (50), 4491-500.
162. Berdel, B.; Nieminen, K.; Soini, Y.*, et al.*, Histone demethylase GASC1--a potential prognostic and predictive marker in invasive breast cancer. *BMC Cancer* **2012,** *12*, 516.
163. Sun, L. L.; Holowatyj, A.; Xu, X. E.*, et al.*, Histone demethylase GASC1, a potential prognostic and predictive marker in esophageal squamous cell carcinoma. *Am J Cancer Res* **2013,** *3* (5), 509-17.
164. Kim, T. D.; Oh, S.; Shin, S.*, et al.*, Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2D/KDM4D. *PloS one* **2012,** *7* (4), e34618.
165. Chen, W. D.; Fu, X.; Dong, B.*, et al.*, Neonatal activation of the nuclear receptor CAR results in epigenetic memory and permanent change of drug metabolism in mouse liver. *Hepatology* **2012,** *56* (4), 1499-509.
166. Wang, G. G.; Song, J.; Wang, Z.*, et al.*, Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **2009,** *459* (7248), 847-51.

167. de Rooij, J. D.; Hollink, I. H.; Arentsen-Peters, S. T.*, et al.*, NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. *Leukemia* **2013,** *27* (12), 2280-8.
168. Sharma, S. V.; Lee, D. Y.; Li, B.*, et al.*, A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **2010,** *141* (1), 69-80.
169. Chicas, A.; Kapoor, A.; Wang, X.*, et al.*, H3K4 demethylation by Jarid1a and Jarid1b contributes to retinoblastoma-mediated gene silencing during cellular senescence. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (23), 8971-6.
170. Paolicchi, E.; Crea, F.; Farrar, W. L.*, et al.*, Histone lysine demethylases in breast cancer. *Crit Rev Oncol Hematol* **2013,** *86* (2), 97-103.
171. Ohta, K.; Haraguchi, N.; Kano, Y.*, et al.*, Depletion of JARID1B induces cellular senescence in human colorectal cancer. *Int. J. Oncol.* **2013,** *42* (4), 1212-8.
172. Li, Q.; Shi, L.; Gui, B.*, et al.*, Binding of the JmjC demethylase JARID1B to LSD1/NuRD suppresses angiogenesis and metastasis in breast cancer cells by repressing chemokine CCL14. *Cancer Res.* **2011,** *71* (21), 6899-908.
173. Barrett, A.; Santangelo, S.; Tan, K.*, et al.*, Breast cancer associated transcriptional repressor PLU-1/JARID1B interacts directly with histone deacetylases. *Int. J. Cancer* **2007,** *121* (2), 265-75.
174. Lu, P. J.; Sundquist, K.; Baeckstrom, D.*, et al.*, A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J. Biol. Chem.* **1999,** *274* (22), 15633-45.
175. Kuzbicki, L.; Lange, D.; Straczynska-Niemiec, A.*, et al.*, JARID1B expression in human melanoma and benign melanocytic skin lesions. *Melanoma Res.* **2013,** *23* (1), 8-12.
176. Niu, X.; Zhang, T.; Liao, L.*, et al.*, The von Hippel-Lindau tumor suppressor protein regulates gene expression and tumor growth through histone demethylase JARID1C. *Oncogene* **2012,** *31* (6), 776-86.
177. Dalgliesh, G. L.; Furge, K.; Greenman, C.*, et al.*, Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **2010,** *463* (7279), 360-3.
178. Perinchery, G.; Sasaki, M.; Angan, A.*, et al.*, Deletion of Y-chromosome specific genes in human prostate cancer. *J. Urol.* **2000,** *163* (4), 1339-42.
179. Puda, A.; Milosevic, J. D.; Berg, T.*, et al.*, Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am. J. Hematol.* **2012,** *87* (3), 245-50.
180. Simon, C.; Chagraoui, J.; Krosl, J.*, et al.*, A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev.* **2012,** *26* (7), 651-6.
181. Kim, J. H.; Sharma, A.; Dhar, S. S.*, et al.*, UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. *Cancer Res.* **2014,** *74* (6), 1705-17.
182. van Haaften, G.; Dalgliesh, G. L.; Davies, H.*, et al.*, Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat. Genet.* **2009,** *41* (5), 521-3.
183. Patani, N.; Jiang, W. G.; Newbold, R. F.*, et al.*, Histone-modifier gene expression profiles are associated with pathological and clinical outcomes in human breast cancer. *Anticancer Res.* **2011,** *31* (12), 4115-25.
184. Dasari, V. K.; Goharderakhshan, R. Z.; Perinchery, G.*, et al.*, Expression analysis of Y chromosome genes in human prostate cancer. *J. Urol.* **2001,** *165* (4), 1335-41.
185. Ene, C. I.; Edwards, L.; Riddick, G.*, et al.*, Histone demethylase Jumonji D3 (JMJD3) as a tumor suppressor by regulating p53 protein nuclear stabilization. *PloS one* **2012,** *7* (12), e51407.
186. Yamamoto, K.; Tateishi, K.; Kudo, Y.*, et al.*, Loss of histone demethylase KDM6B enhances aggressiveness of pancreatic cancer through downregulation of C/EBPalpha. *Carcinogenesis* **2014**.
187. Wei, Y.; Chen, R.; Dimicoli, S.*, et al.*, Global H3K4me3 genome mapping reveals alterations of innate immunity signaling and overexpression of JMJD3 in human myelodysplastic syndrome CD34+ cells. *Leukemia* **2013,** *27* (11), 2177-86.
188. Anderton, J. A.; Bose, S.; Vockerodt, M.*, et al.*, The H3K27me3 demethylase, KDM6B, is induced by Epstein-Barr virus and over-expressed in Hodgkin's Lymphoma. *Oncogene* **2011,** *30* (17), 2037-43.

189. McLaughlin-Drubin, M. E.; Crum, C. P.; Munger, K., Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (5), 2130-5.
190. Svotelis, A.; Bianco, S.; Madore, J.*, et al.*, H3K27 demethylation by JMJD3 at a poised enhancer of anti-apoptotic gene BCL2 determines ERalpha ligand dependency. *The EMBO journal* **2011,** *30* (19), 3947-61.
191. Dubuc, A. M.; Remke, M.; Korshunov, A.*, et al.*, Aberrant patterns of H3K4 and H3K27 histone lysine methylation occur across subgroups in medulloblastoma. *Acta Neuropathol.* **2013,** *125* (3), 373-84.
192. Pereira, F.; Barbachano, A.; Silva, J.*, et al.*, KDM6B/JMJD3 histone demethylase is induced by vitamin D and modulates its effects in colon cancer cells. *Hum. Mol. Genet.* **2011,** *20* (23), 4655-65.
193. Ramadoss, S.; Chen, X.; Wang, C. Y., Histone demethylase KDM6B promotes epithelial-mesenchymal transition. *J. Biol. Chem.* **2012,** *287* (53), 44508-17.
194. Shen, Y.; Guo, X.; Wang, Y.*, et al.*, Expression and significance of histone H3K27 demethylases in renal cell carcinoma. *BMC Cancer* **2012,** *12*, 470.
195. Sun, X.; Qiu, J. J.; Zhu, S.*, et al.*, Oncogenic features of PHF8 histone demethylase in esophageal squamous cell carcinoma. *PloS one* **2013,** *8* (10), e77353.
196. Sinha, S.; Singh, R. K.; Alam, N.*, et al.*, Alterations in candidate genes PHF2, FANCC, PTCH1 and XPA at chromosomal 9q22.3 region: pathological significance in early- and late-onset breast carcinoma. *Mol Cancer* **2008,** *7*, 84.
197. Arteaga, M. F.; Mikesch, J. H.; Qiu, J.*, et al.*, The histone demethylase PHF8 governs retinoic acid response in acute promyelocytic leukemia. *Cancer Cell* **2013,** *23* (3), 376-89.
198. Bjorkman, M.; Ostling, P.; Harma, V.*, et al.*, Systematic knockdown of epigenetic enzymes identifies a novel histone demethylase PHF8 overexpressed in prostate cancer with an impact on cell proliferation, migration and invasion. *Oncogene* **2012,** *31* (29), 3444-56.
199. Osawa, T.; Muramatsu, M.; Wang, F.*, et al.*, Increased expression of histone demethylase JHDM1D under nutrient starvation suppresses tumor growth via down-regulating angiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (51), 20725-9.
200. Hsia, D. A.; Tepper, C. G.; Pochampalli, M. R.*, et al.*, KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (21), 9671-6.
201. Oh, S.; Janknecht, R., Histone demethylase JMJD5 is essential for embryonic development. *Biochem. Biophys. Res. Commun.* **2012,** *420* (1), 61-5.
202. Lee, Y. F.; Miller, L. D.; Chan, X. B.*, et al.*, JMJD6 is a driver of cellular proliferation and motility and a marker of poor prognosis in breast cancer. *Breast Cancer Res* **2012,** *14* (3), R85.
203. Zhang, J.; Ni, S. S.; Zhao, W. L.*, et al.*, High expression of JMJD6 predicts unfavorable survival in lung adenocarcinoma. *Tumour Biol* **2013,** *34* (4), 2397-401.
204. Boeckel, J. N.; Guarani, V.; Koyanagi, M.*, et al.*, Jumonji domain-containing protein 6 (Jmjd6) is required for angiogenic sprouting and regulates splicing of VEGF-receptor 1. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (8), 3276-81.
205. Wang, F.; He, L.; Huangyang, P.*, et al.*, JMJD6 promotes colon carcinogenesis through negative regulation of p53 by hydroxylation. *PLoS Biol.* **2014,** *12* (3), e1001819.
206. Bodmer, D.; Schepens, M.; Eleveld, M. J.*, et al.*, Disruption of a novel gene, DIRC3, and expression of DIRC3-HSPBAP1 fusion transcripts in a case of familial renal cell cancer and t(2;3)(q35;q21). *Genes Chromosomes Cancer* **2003,** *38* (2), 107-16.
207. Suzuki, C.; Takahashi, K.; Hayama, S.*, et al.*, Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer. *Mol. Cancer Ther.* **2007,** *6* (2), 542-51.
208. Zhang, Y.; Lu, Y.; Yuan, B. Z.*, et al.*, The Human mineral dust-induced gene, mdig, is a cell growth regulating gene associated with lung cancer. *Oncogene* **2005,** *24* (31), 4873-82.
209. Komiya, K.; Sueoka-Aragane, N.; Sato, A.*, et al.*, Mina53, a novel c-Myc target gene, is frequently expressed in lung cancers and exerts oncogenic property in NIH/3T3 cells. *J. Cancer Res. Clin. Oncol.* **2010,** *136* (3), 465-73.

210. Tsuneoka, M.; Koda, Y.; Soejima, M.*, et al.*, A novel myc target gene, mina53, that is involved in cell proliferation. *J. Biol. Chem.* **2002,** *277* (38), 35450-9.
211. Teye, K.; Tsuneoka, M.; Arima, N.*, et al.*, Increased expression of a Myc target gene Mina53 in human colon cancer. *Am. J. Pathol.* **2004,** *164* (1), 205-16.
212. Tan, X. P.; Zhang, Q.; Dong, W. G.*, et al.*, Upregulated expression of Mina53 in cholangiocarcinoma and its clinical significance. *Oncol Lett* **2012,** *3* (5), 1037-1041.
213. Ogasawara, S.; Komuta, M.; Nakashima, O.*, et al.*, Accelerated expression of a Myc target gene Mina53 in aggressive hepatocellular carcinoma. *Hepatol. Res.* **2010,** *40* (4), 330-6.
214. Zhang, Q.; Hu, C. M.; Yuan, Y. S.*, et al.*, Expression of Mina53 and its significance in gastric carcinoma. *Int. J. Biol. Markers* **2008,** *23* (2), 83-8.
215. Tan, X. P.; Dong, W. G.; Zhang, Q.*, et al.*, Potential effects of Mina53 on tumor growth in human pancreatic cancer. *Cell Biochem. Biophys.* **2014,** *69* (3), 619-25.
216. Ishizaki, H.; Yano, H.; Tsuneoka, M.*, et al.*, Overexpression of the myc target gene Mina53 in advanced renal cell carcinoma. *Pathol. Int.* **2007,** *57* (10), 672-80.
217. Tsuneoka, M.; Fujita, H.; Arima, N.*, et al.*, Mina53 as a potential prognostic factor for esophageal squamous cell carcinoma. *Clin Cancer Res* **2004,** *10* (21), 7347-56.
218. Fukahori, S.; Yano, H.; Tsuneoka, M.*, et al.*, Immunohistochemical expressions of Cap43 and Mina53 proteins in neuroblastoma. *J Pediatr Surg* **2007,** *42* (11), 1831-40.
219. Teye, K.; Arima, N.; Nakamura, Y.*, et al.*, Expression of Myc target gene mina53 in subtypes of human lymphoma. *Oncol. Rep.* **2007,** *18* (4), 841-8.
220. Komiya, K.; Sueoka-Aragane, N.; Sato, A.*, et al.*, Expression of Mina53, a novel c-Myc target gene, is a favorable prognostic marker in early stage lung cancer. *Lung Cancer* **2010,** *69* (2), 232-8.
221. Marina, O.; Hainz, U.; Biernacki, M. A.*, et al.*, Serologic markers of effective tumor immunity against chronic lymphocytic leukemia include nonmutated B-cell antigens. *Cancer Res.* **2010,** *70* (4), 1344-55.
222. Hu, L.; Wu, C.; Zhao, X.*, et al.*, Genome-wide association study of prognosis in advanced non-small cell lung cancer patients receiving platinum-based chemotherapy. *Clin Cancer Res* **2012,** *18* (19), 5507-14.
223. Fujii, T.; Shimada, K.; Anai, S.*, et al.*, ALKBH2, a novel AlkB homologue, contributes to human bladder cancer progression by regulating MUC1 expression. *Cancer Sci.* **2013,** *104* (3), 321-7.
224. Bethke, L.; Murray, A.; Webb, E.*, et al.*, Comprehensive analysis of DNA repair gene variants and risk of meningioma. *J Natl Cancer Inst* **2008,** *100* (4), 270-6.
225. Johannessen, T. C.; Prestegarden, L.; Grudic, A.*, et al.*, The DNA repair protein ALKBH2 mediates temozolomide resistance in human glioblastoma cells. *Neuro Oncol* **2013,** *15* (3), 269-78.
226. Gao, W.; Li, L.; Xu, P.*, et al.*, Frequent down-regulation of hABH2 in gastric cancer and its involvement in growth of cancer cells. *J. Gastroenterol. Hepatol.* **2011,** *26* (3), 577-84.
227. Tasaki, M.; Shimada, K.; Kimura, H.*, et al.*, ALKBH3, a human AlkB homologue, contributes to cell survival in human non-small-cell lung cancer. *Br. J. Cancer* **2011,** *104* (4), 700-6.
228. Shimada, K.; Fujii, T.; Tsujikawa, K.*, et al.*, ALKBH3 contributes to survival and angiogenesis of human urothelial carcinoma cells through NADPH oxidase and tweak/Fn14/VEGF signals. *Clin Cancer Res* **2012,** *18* (19), 5247-55.
229. Yamato, I.; Sho, M.; Shimada, K.*, et al.*, PCA-1/ALKBH3 contributes to pancreatic cancer by supporting apoptotic resistance and angiogenesis. *Cancer Res.* **2012,** *72* (18), 4829-39.
230. Koike, K.; Ueda, Y.; Hase, H.*, et al.*, anti-tumor effect of AlkB homolog 3 knockdown in hormone- independent prostate cancer cells. *Curr. Cancer Drug Targets* **2012,** *12* (7), 847-56.
231. Nakao, S.; Mabuchi, M.; Shimizu, T.*, et al.*, Design and synthesis of prostate cancer antigen-1 (PCA-1/ALKBH3) inhibitors as anti-prostate cancer drugs. *Bioorg. Med. Chem. Lett.* **2014,** *24* (4), 1071-4.
232. Thalhammer, A.; Bencokova, Z.; Poole, R.*, et al.*, Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1alpha (HIF-1alpha). *PloS one* **2011,** *6* (1), e16210.
233. Fu, D.; Jordan, J. J.; Samson, L. D., Human ALKBH7 is required for alkylation and oxidation-induced programmed necrosis. *Genes Dev.* **2013,** *27* (10), 1089-100.

234. Shimada, K.; Nakamura, M.; Anai, S.*, et al.*, A novel human AlkB homologue, ALKBH8, contributes to human bladder cancer progression. *Cancer Res.* **2009,** *69* (7), 3157-64.
235. da Cunha, P. A.; de Carlos Back, L. K.; Sereia, A. F.*, et al.*, Interaction between obesity-related genes, FTO and MC4R, associated to an increase of breast cancer risk. *Mol. Biol. Rep.* **2013,** *40* (12), 6657-64.
236. Kaklamani, V.; Yi, N.; Sadim, M.*, et al.*, The role of the fat mass and obesity associated gene (FTO) in breast cancer risk. *BMC Med. Genet.* **2011,** *12*, 52.
237. Tarabra, E.; Actis, G. C.; Fadda, M.*, et al.*, The obesity gene and colorectal cancer risk: a population study in Northern Italy. *Eur J Intern Med* **2012,** *23* (1), 65-9.
238. Frayling, T. M.; Timpson, N. J.; Weedon, M. N.*, et al.*, A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **2007,** *316* (5826), 889-94.
239. Keller, L.; Xu, W.; Wang, H. X.*, et al.*, The obesity related gene, FTO, interacts with APOE, and is associated with Alzheimer's disease risk: a prospective cohort study. *J Alzheimers Dis* **2011,** *23* (3), 461-9.
240. Lappalainen, T.; Kolehmainen, M.; Schwab, U. S.*, et al.*, Association of the FTO gene variant (rs9939609) with cardiovascular disease in men with abnormal glucose metabolism--the Finnish Diabetes Prevention Study. *Nutr Metab Cardiovasc Dis* **2011,** *21* (9), 691-8.
241. Huang, H.; Jiang, X.; Li, Z.*, et al.*, TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (29), 11994-9.
242. Lorsbach, R. B.; Moore, J.; Mathew, S.*, et al.*, TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **2003,** *17* (3), 637-41.
243. Abdel-Wahab, O.; Mullally, A.; Hedvat, C.*, et al.*, Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **2009,** *114* (1), 144-7.
244. Liu, C.; Liu, L.; Chen, X.*, et al.*, Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. *PloS one* **2013,** *8* (5), e62828.
245. Yang, Q.; Wu, K.; Ji, M.*, et al.*, Decreased 5-hydroxymethylcytosine (5-hmC) is an independent poor prognostic factor in gastric cancer patients. *J Biomed Nanotechnol* **2013,** *9* (9), 1607-16.
246. Wielscher, M.; Liou, W.; Pulverer, W.*, et al.*, Cytosine 5-Hydroxymethylation of the LZTS1 Gene Is Reduced in Breast Cancer. *Transl Oncol* **2013,** *6* (6), 715-21.
247. Hsu, C. H.; Peng, K. L.; Kang, M. L.*, et al.*, TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. *Cell Rep* **2012,** *2* (3), 568-79.
248. Ko, M.; Huang, Y.; Jankowska, A. M.*, et al.*, Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **2010,** *468* (7325), 839-43.
249. Langemeijer, S. M.; Jansen, J. H.; Hooijer, J.*, et al.*, TET2 mutations in childhood leukemia. *Leukemia* **2011,** *25* (1), 189-92.
250. Langemeijer, S. M.; Kuiper, R. P.; Berends, M.*, et al.*, Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* **2009,** *41* (7), 838-42.
251. Jawert, F.; Hasseus, B.; Kjeller, G.*, et al.*, Loss of 5-hydroxymethylcytosine and TET2 in oral squamous cell carcinoma. *Anticancer Res.* **2013,** *33* (10), 4325-8.
252. Lavaissiere, L.; Jia, S.; Nishiyama, M.*, et al.*, Overexpression of human aspartyl(asparaginyl)beta-hydroxylase in hepatocellular carcinoma and cholangiocarcinoma. *J Clin Invest* **1996,** *98* (6), 1313-23.
253. Lee, J. H., Overexpression of humbug promotes malignant progression in human gastric cancer cells. *Oncol. Rep.* **2008,** *19* (3), 795-800.
254. Palumbo, K. S.; Wands, J. R.; Safran, H.*, et al.*, Human aspartyl (asparaginyl) beta-hydroxylase monoclonal antibodies: potential biomarkers for pancreatic carcinoma. *Pancreas* **2002,** *25* (1), 39-44.
255. Yang, H.; Song, K.; Xue, T.*, et al.*, The distribution and expression profiles of human Aspartyl/Asparaginyl beta-hydroxylase in tumor cell lines and human tissues. *Oncol. Rep.* **2010,** *24* (5), 1257-64.
256. Luu, M.; Sabo, E.; de la Monte, S. M.*, et al.*, Prognostic value of aspartyl (asparaginyl)-beta-hydroxylase/humbug expression in non-small cell lung carcinoma. *Hum. Pathol.* **2009,** *40* (5), 639-44.
257. Wang, J.; de la Monte, S. M.; Sabo, E.*, et al.*, Prognostic value of humbug gene overexpression in stage II colon cancer. *Hum. Pathol.* **2007,** *38* (1), 17-25.

258. Feldmann, G.; Nattermann, J.; Nischalke, H. D*., et al.*, Detection of human aspartyl (asparaginyl) beta-hydroxylase and homeobox B7 mRNA in brush cytology specimens from patients with bile duct cancer. *Endoscopy* **2006,** *38* (6), 604-9.
259. Xue, T.; Xue, X. P.; Huang, Q. S*., et al.*, Monoclonal antibodies against human aspartyl (asparaginyl) beta-hydroxylase developed by DNA immunization. *Hybridoma (Larchmt)* **2009,** *28* (4), 251-7.
260. Elvidge, G. P.; Glenny, L.; Appelhoff, R. J*., et al.*, Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. *J. Biol. Chem.* **2006,** *281* (22), 15215-26.
261. Ince, N.; de la Monte, S. M.; Wands, J. R., Overexpression of human aspartyl (asparaginyl) beta-hydroxylase is associated with malignant transformation. *Cancer Res.* **2000,** *60* (5), 1261-6.
262. Wang, K.; Liu, J.; Yan, Z. L*., et al.*, Overexpression of aspartyl-(asparaginyl)-beta-hydroxylase in hepatocellular carcinoma is associated with worse surgical outcome. *Hepatology* **2010,** *52* (1), 164-73.
263. Dinchuk, J. E.; Focht, R. J.; Kelley, J. A*., et al.*, Absence of post-translational aspartyl beta-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia. *J. Biol. Chem.* **2002,** *277* (15), 12970-7.
264. Maeda, T.; Taguchi, K.; Aishima, S*., et al.*, Clinicopathological correlates of aspartyl (asparaginyl) beta-hydroxylase over-expression in cholangiocarcinoma. *Cancer Detect. Prev.* **2004,** *28* (5), 313-8.
265. Vaz, F. M.; Wanders, R. J., Carnitine biosynthesis in mammals. *Biochem. J.* **2002,** *361* (Pt 3), 417-29.
266. Loenarz, C.; Chowdhury, R.; Schofield, C. J*., et al.*, Oxygenases for oxygen sensing. *Pure Appl. Chem.* **2008,** *80* (8), 1837–47.
267. Qiu, J., Epigenetics: unfinished symphony. *Nature* **2006,** *441* (7090), 143-5.
268. Di Lullo, G. A.; Sweeney, S. M.; Korkko, J*., et al.*, Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* **2002,** *277* (6), 4223-31.
269. Hulmes, D. J., Building collagen molecules, fibrils, and suprafibrillar structures. *J. Struct. Biol.* **2002,** *137* (1-2), 2-10.
270. Hutton, J. J., Jr.; Kaplan, A.; Udenfriend, S., Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase. *Arch. Biochem. Biophys.* **1967,** *121* (2), 384-91.
271. Berg, R. A.; Prockop, D. J., Affinity column purification of protocollagen proline hydroxylase from chick embryos and further characterization of the enzyme. *J. Biol. Chem.* **1973,** *248* (4), 1175-82.
272. Helaakoski, T.; Vuori, K.; Myllyla, R*., et al.*, Molecular cloning of the alpha-subunit of human prolyl 4-hydroxylase: the complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (12), 4392-6.
273. Pihlajaniemi, T.; Helaakoski, T.; Tasanen, K*., et al.*, Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *The EMBO journal* **1987,** *6* (3), 643-9.
274. Myllyharju, J., Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **2003,** *22* (1), 15-24.
275. Vuori, K.; Pihlajaniemi, T.; Myllyla, R*., et al.*, Site-directed mutagenesis of human protein disulphide isomerase: effect on the assembly, activity and endoplasmic reticulum retention of human prolyl 4-hydroxylase in Spodoptera frugiperda insect cells. *The EMBO journal* **1992,** *11* (11), 4213-7.
276. Murad, S.; Grove, D.; Lindberg, K. A*., et al.*, Regulation of collagen synthesis by ascorbic acid. *Proc. Natl. Acad. Sci. U. S. A.* **1981,** *78* (5), 2879-82.
277. Murad, S.; Sivarajah, A.; Pinnell, S. R., Regulation of prolyl and lysyl hydroxylase activities in cultured human skin fibroblasts by ascorbic acid. *Biochem. Biophys. Res. Commun.* **1981,** *101* (3), 868-75.
278. Jenkins, C. L.; Raines, R. T., Insights on the conformational stability of collagen. *Nat. Prod. Rep.* **2002,** *19* (1), 49-59.

279. Ruotsalainen, H.; Sipila, L.; Vapola, M.*, et al.*, Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. *J. Cell Sci.* **2006,** *119* (Pt 4), 625-35.
280. Turpeenniemi-Hujanen, T. M.; Puistola, U.; Kivirikko, K. I., Isolation of lysyl hydroxylase, an enzyme of collagen synthesis, from chick embryos as a homogeneous protein. *Biochem. J.* **1980,** *189* (2), 247-53.
281. Kivirikko, K. I.; Ryhanen, L.; Anttinen, H.*, et al.*, Further hydroxylation of lysyl residues in collagen by protocollagen lysyl hydroxylase in vitro. *Biochemistry* **1973,** *12* (24), 4966-71.
282. Koukoulis, G. K.; Patriarca, C.; Gould, V. E., Adhesion molecules and tumor metastasis. *Hum. Pathol.* **1998,** *29* (9), 889-92.
283. Juliano, R. L.; Haskill, S., Signal transduction from the extracellular matrix. *J. Cell Biol.* **1993,** *120* (3), 577-85.
284. Lester, B. R.; McCarthy, J. B., Tumor cell adhesion to the extracellular matrix and signal transduction mechanisms implicated in tumor cell motility, invasion and metastasis. *Cancer Metastasis Rev.* **1992,** *11* (1), 31-44.
285. Canty, E. G.; Kadler, K. E., Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* **2005,** *118* (Pt 7), 1341-53.
286. Lochter, A.; Bissell, M. J., Involvement of extracellular matrix constituents in breast cancer. *Semin. Cancer Biol.* **1995,** *6* (3), 165-73.
287. Provenzano, P. P.; Eliceiri, K. W.; Campbell, J. M.*, et al.*, Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med* **2006,** *4* (1), 38.
288. Semenza, G. L., Hypoxia-inducible factor 1 (HIF-1) pathway. *Sci STKE* **2007,** *2007* (407), cm8.
289. Semenza, G. L., Hypoxia-inducible factor 1: master regulator of O2 homeostasis. *Curr. Opin. Genet. Dev.* **1998,** *8* (5), 588-94.
290. Semenza, G. L., Hydroxylation of HIF-1: oxygen sensing at the molecular level. *Physiol. (Bethesda)* **2004,** *19*, 176-82.
291. Hon, W. C.; Wilson, M. I.; Harlos, K.*, et al.*, Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature* **2002,** *417* (6892), 975-8.
292. Masson, N.; Willam, C.; Maxwell, P. H.*, et al.*, Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. *The EMBO journal* **2001,** *20* (18), 5197-206.
293. Jaakkola, P.; Mole, D. R.; Tian, Y. M.*, et al.*, Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **2001,** *292* (5516), 468-72.
294. Lando, D.; Peet, D. J.; Whelan, D. A.*, et al.*, Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. *Science* **2002,** *295* (5556), 858-61.
295. Hewitson, K. S.; McNeill, L. A.; Riordan, M. V.*, et al.*, Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J. Biol. Chem.* **2002,** *277* (29), 26351-5.
296. McNeill, L. A.; Hewitson, K. S.; Claridge, T. D.*, et al.*, Hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) catalyses hydroxylation at the beta-carbon of asparagine-803. *Biochem. J.* **2002,** *367* (Pt 3), 571-5.
297. Stolze, I. P.; Tian, Y. M.; Appelhoff, R. J.*, et al.*, Genetic analysis of the role of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible factor (FIH) in regulating hypoxia-inducible factor (HIF) transcriptional target genes [corrected]. *J. Biol. Chem.* **2004,** *279* (41), 42719-25.
298. Sang, N.; Fang, J.; Srinivas, V.*, et al.*, Carboxyl-terminal transactivation activity of hypoxia-inducible factor 1 alpha is governed by a von Hippel-Lindau protein-independent, hydroxylation-regulated association with p300/CBP. *Mol. Cell. Biol.* **2002,** *22* (9), 2984-92.
299. Freedman, S. J.; Sun, Z. Y.; Poy, F.*, et al.*, Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (8), 5367-72.
300. Wenger, R. H., Cellular adaptation to hypoxia: O2-sensing protein hydroxylases, hypoxia-inducible transcription factors, and O2-regulated gene expression. *FASEB J.* **2002,** *16* (10), 1151-62.
301. Greijer, A. E.; van der Groep, P.; Kemming, D.*, et al.*, Up-regulation of gene expression by hypoxia is mediated predominantly by hypoxia-inducible factor 1 (HIF-1). *J. Pathol.* **2005,** *206* (3), 291-304.

302. Henze, A. T.; Acker, T., Feedback regulators of hypoxia-inducible factors and their role in cancer biology. *Cell Cycle* **2010,** *9* (14), 2749-63.
303. Maynard, M. A.; Ohh, M., The role of hypoxia-inducible factors in cancer. *Cellular and molecular life sciences : CMLS* **2007,** *64* (16), 2170-80.
304. Zhong, H.; De Marzo, A. M.; Laughner, E.*, et al.*, Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. *Cancer Res.* **1999,** *59* (22), 5830-5.
305. Semenza, G. L., Evaluation of HIF-1 inhibitors as anticancer agents. *Drug Discov. Today* **2007,** *12* (19-20), 853-9.
306. Krishnamachary, B.; Semenza, G. L., Analysis of hypoxia-inducible factor 1alpha expression and its effects on invasion and metastasis. *Methods Enzymol.* **2007,** *435*, 347-54.
307. Bertout, J. A.; Patel, S. A.; Simon, M. C., The impact of O2 availability on human cancer. *Nat. Rev. Cancer* **2008,** *8* (12), 967-75.
308. Kaelin, W. G., Jr., New cancer targets emerging from studies of the Von Hippel-Lindau tumor suppressor protein. *Ann. N. Y. Acad. Sci.* **2010,** *1210*, 1-7.
309. Hewitson, K. S.; McNeill, L. A.; Schofield, C. J., Modulating the hypoxia-inducible factor signaling pathway: applications from cardiovascular disease to cancer. *Curr. Pharm. Des.* **2004,** *10* (8), 821-33.
310. Nagel, S.; Talbot, N. P.; Mecinovic, J.*, et al.*, Therapeutic manipulation of the HIF hydroxylases. *Antioxid. Redox Signal.* **2010,** *12* (4), 481-501.
311. Yan, L.; Colandrea, V. J.; Hale, J. J., Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: small molecule-based therapeutics for anemia. *Expert Opin Ther Pat* **2010,** *20* (9), 1219-45.
312. Lachner, M.; O'Sullivan, R. J.; Jenuwein, T., An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **2003,** *116* (Pt 11), 2117-24.
313. Tsukada, Y.; Fang, J.; Erdjument-Bromage, H.*, et al.*, Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **2006,** *439* (7078), 811-6.
314. Cloos, P. A.; Christensen, J.; Agger, K.*, et al.*, Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **2008,** *22* (9), 1115-40.
315. Klose, R. J.; Zhang, Y., Regulation of histone methylation by demethylimination and demethylation. *Nat. Rev. Mol. Cell Biol.* **2007,** *8* (4), 307-18.
316. Ng, S. S.; Kavanagh, K. L.; McDonough, M. A.*, et al.*, Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **2007,** *448* (7149), 87-91.
317. Loenarz, C.; Ge, W.; Coleman, M. L.*, et al.*, PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. *Hum. Mol. Genet.* **2010,** *19* (2), 217-22.
318. Rose, N. R.; McDonough, M. A.; King, O. N.*, et al.*, Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **2011,** *40* (8), 4364-97.
319. Arrowsmith, C. H.; Bountra, C.; Fish, P. V.*, et al.*, Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* **2012,** *11* (5), 384-400.
320. Sedgwick, B.; Bates, P. A.; Paik, J.*, et al.*, Repair of alkylated DNA: recent advances. *DNA Repair (Amst)* **2007,** *6* (4), 429-42.
321. Delaney, J. C.; Essigmann, J. M., Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (39), 14051-6.
322. Frick, L. E.; Delaney, J. C.; Wong, C.*, et al.*, Alleviation of 1,N6-ethanoadenine genotoxicity by the Escherichia coli adaptive response protein AlkB. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (3), 755-60.
323. Duncan, T.; Trewick, S. C.; Koivisto, P.*, et al.*, Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (26), 16660-5.
324. Li, M. M.; Nilsen, A.; Shi, Y.*, et al.*, ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nat Commun* **2013,** *4*, 1832.
325. Zheng, G.; Dahl, J. A.; Niu, Y.*, et al.*, ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **2013,** *49* (1), 18-29.
326. Cetica, V.; Genitori, L.; Giunti, L.*, et al.*, Pediatric brain tumors: mutations of two dioxygenases (hABH2 and hABH3) that directly repair alkylation damage. *J Neurooncol* **2009,** *94* (2), 195-201.

327. Gerken, T.; Girard, C. A.; Tung, Y. C.*, et al.*, The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **2007,** *318* (5855), 1469-72.
328. Tahiliani, M.; Koh, K. P.; Shen, Y.*, et al.*, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009,** *324* (5929), 930-5.
329. Ito, S.; Shen, L.; Dai, Q.*, et al.*, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011,** *333* (6047), 1300-3.
330. Kamiya, H.; Tsuchiya, H.; Karino, N.*, et al.*, Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells. *Journal of biochemistry* **2002,** *132* (4), 551-5.
331. Ge, W.; Wolf, A.; Feng, T.*, et al.*, Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat. Chem. Biol.* **2012,** *8* (12), 960-2.
332. Lu, Y.; Chang, Q.; Zhang, Y.*, et al.*, Lung cancer-associated JmjC domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3. *Cell Cycle* **2009,** *8* (13), 2101-9.
333. Sinha, K. M.; Yasuda, H.; Coombes, M. M.*, et al.*, Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *The EMBO journal* **2010,** *29* (1), 68-79.
334. Singleton, R. S.; Liu-Yi, P.; Formenti, F.*, et al.*, OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (11), 4031-6.
335. Loenarz, C.; Sekirnik, R.; Thalhammer, A.*, et al.*, Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (11), 4019-24.
336. McMullen, B. A.; Fujikawa, K.; Kisiel, W., The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood coagulation zymogens. *Biochem. Biophys. Res. Commun.* **1983,** *115* (1), 8-14.
337. McMullen, B. A.; Fujikawa, K.; Kisiel, W.*, et al.*, Complete amino acid sequence of the light chain of human blood coagulation factor X: evidence for identification of residue 63 as beta-hydroxyaspartic acid. *Biochemistry* **1983,** *22* (12), 2875-84.
338. Stenflo, J.; Holme, E.; Lindstedt, S.*, et al.*, Hydroxylation of aspartic acid in domains homologous to the epidermal growth factor precursor is catalyzed by a 2-oxoglutarate-dependent dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (2), 444-7.
339. Derian, C. K.; VanDusen, W.; Przysiecki, C. T.*, et al.*, Inhibitors of 2-ketoglutarate-dependent dioxygenases block aspartyl beta-hydroxylation of recombinant human factor IX in several mammalian expression systems. *J. Biol. Chem.* **1989,** *264* (12), 6615-8.
340. Gronke, R. S.; VanDusen, W. J.; Garsky, V. M.*, et al.*, Aspartyl beta-hydroxylase: in vitro hydroxylation of a synthetic peptide based on the structure of the first growth factor-like domain of human factor IX. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (10), 3609-13.
341. Gronke, R. S.; Welsch, D. J.; VanDusen, W. J.*, et al.*, Partial purification and characterization of bovine liver aspartyl beta-hydroxylase. *J. Biol. Chem.* **1990,** *265* (15), 8558-65.
342. Korioth, F.; Gieffers, C.; Frey, J., Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase. *Gene* **1994,** *150* (2), 395-9.
343. Drakenberg, T.; Fernlund, P.; Roepstorff, P.*, et al.*, beta-Hydroxyaspartic acid in vitamin K-dependent protein C. *Proc. Natl. Acad. Sci. U. S. A.* **1983,** *80* (7), 1802-6.
344. Carpenter, G.; Cohen, S., Epidermal growth factor. *J. Biol. Chem.* **1990,** *265* (14), 7709-12.
345. Valcarce, C.; Bjork, I.; Stenflo, J., The epidermal growth factor precursor. A calcium-binding, beta-hydroxyasparagine containing modular protein present on the surface of platelets. *Eur. J. Biochem.* **1999,** *260* (1), 200-7.
346. Boonstra, J.; Rijken, P.; Humbel, B.*, et al.*, The epidermal growth factor. *Cell Biol. Int.* **1995,** *19* (5), 413-30.
347. Harris, R. C.; Chung, E.; Coffey, R. J., EGF receptor ligands. *Exp. Cell Res.* **2003,** *284* (1), 2-13.
348. Heath, W. F.; Merrifield, R. B., A synthetic approach to structure-function relationships in the murine epidermal growth factor molecule. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83* (17), 6367-71.
349. Downing, A. K.; Knott, V.; Werner, J. M.*, et al.*, Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* **1996,** *85* (4), 597-605.

350. Jensen, S. A.; Corbett, A. R.; Knott, V.*, et al.*, Ca2+-dependent interface formation in fibrillin-1. *J. Biol. Chem.* **2005,** *280* (14), 14076-84.
351. Rao, Z.; Handford, P.; Mayhew, M.*, et al.*, The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell* **1995,** *82* (1), 131-41.
352. Selander-Sunnerhagen, M.; Ullner, M.; Persson, E.*, et al.*, How an epidermal growth factor (EGF)-like domain binds calcium. High resolution NMR structure of the calcium form of the NH2-terminal EGF-like domain in coagulation factor X. *J. Biol. Chem.* **1992,** *267* (27), 19642-9.
353. Sunnerhagen, M. S.; Persson, E.; Dahlqvist, I.*, et al.*, The effect of aspartate hydroxylation on calcium binding to epidermal growth factor-like modules in coagulation factors IX and X. *J. Biol. Chem.* **1993,** *268* (31), 23339-44.
354. Nelson, R. M.; VanDusen, W. J.; Friedman, P. A.*, et al.*, beta-Hydroxyaspartic acid and beta-hydroxyasparagine residues in recombinant human protein S are not required for anticoagulant cofactor activity or for binding to C4b-binding protein. *J. Biol. Chem.* **1991,** *266* (31), 20586-9.
355. Stenflo, J., Structure-function relationships of epidermal growth factor modules in vitamin K-dependent clotting factors. *Blood* **1991,** *78* (7), 1637-51.
356. Fernlund, P.; Stenflo, J., Beta-hydroxyaspartic acid in vitamin K-dependent proteins. *J. Biol. Chem.* **1983,** *258* (20), 12509-12.
357. Astermark, J.; Hogg, P. J.; Stenflo, J., The gamma-carboxyglutamic acid and epidermal growth factor-like modules of factor IXa beta. Effects on the serine protease module and factor X activation. *J. Biol. Chem.* **1994,** *269* (5), 3682-9.
358. PanaceaPharmaceuticals www.panaceapharma.com.
359. PanaceaGlobal www.panaceaglobalinc.com.
360. Patel, N.; Khan, A. O.; Mansour, A.*, et al.*, Mutations in ASPH Cause Facial Dysmorphism, Lens Dislocation, Anterior-Segment Abnormalities, and Spontaneous Filtering Blebs, or Traboulsi Syndrome. *Am. J. Hum. Genet.* **2014**.
361. Geer, L. Y.; Marchler-Bauer, A.; Geer, R. C.*, et al.*, The NCBI BioSystems database. *Nucleic Acids Res.* **2010,** *38* (Database issue), D492-6.
362. Feriotto, G.; Finotti, A.; Volpe, P.*, et al.*, Myocyte enhancer factor 2 activates promoter sequences of the human AbetaH-J-J locus, encoding aspartyl-beta-hydroxylase, junctin, and junctate. *Mol. Cell. Biol.* **2005,** *25* (8), 3261-75.
363. Feriotto, G.; Finotti, A.; Breveglieri, G.*, et al.*, Multiple levels of control of the expression of the human A beta H-J-J locus encoding aspartyl-beta-hydroxylase, junctin, and junctate. *Ann. N. Y. Acad. Sci.* **2006,** *1091*, 184-90.
364. Sievers, F.; Wilm, A.; Dineen, D.*, et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **2011,** *7*, 539.
365. Treves, S.; Feriotto, G.; Moccagatta, L.*, et al.*, Molecular cloning, expression, functional characterization, chromosomal localization, and gene structure of junctate, a novel integral calcium binding protein of sarco(endo)plasmic reticulum membrane. *J. Biol. Chem.* **2000,** *275* (50), 39555-68.
366. Dinchuk, J. E.; Henderson, N. L.; Burn, T. C.*, et al.*, Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved isoform of Asph missing the catalytic domain share exons with junctin. *J. Biol. Chem.* **2000,** *275* (50), 39543-54.
367. Jones, L. R.; Zhang, L.; Sanborn, K.*, et al.*, Purification, primary structure, and immunological characterization of the 26-kDa calsequestrin binding protein (junctin) from cardiac junctional sarcoplasmic reticulum. *J. Biol. Chem.* **1995,** *270* (51), 30787-96.
368. Lim, K. Y.; Hong, C. S.; Kim, D. H., cDNA cloning and characterization of human cardiac junctin. *Gene* **2000,** *255* (1), 35-42.
369. Treves, S.; Franzini-Armstrong, C.; Moccagatta, L.*, et al.*, Junctate is a key element in calcium entry induced by activation of InsP3 receptors and/or calcium store depletion. *J. Cell Biol.* **2004,** *166* (4), 537-48.
370. Zhang, L.; Kelley, J.; Schmeisser, G.*, et al.*, Complex formation between junctin, triadin, calsequestrin, and the ryanodine receptor. Proteins of the cardiac junctional sarcoplasmic reticulum membrane. *J. Biol. Chem.* **1997,** *272* (37), 23389-97.
371. Srikanth, S.; Jew, M.; Kim, K. D.*, et al.*, Junctate is a Ca2+-sensing structural component of Orai1 and stromal interaction molecule 1 (STIM1). *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (22), 8682-7.

372. Lancaster, D. E.; McDonough, M. A.; Schofield, C. J., Factor inhibiting hypoxia-inducible factor (FIH) and other asparaginyl hydroxylases. *Biochem. Soc. Trans.* **2004,** *32* (Pt 6), 943-5.
373. Graslund, S.; Nordlund, P.; Weigelt, J.*, et al.*, Protein production and purification. *Nat Methods* **2008,** *5* (2), 135-46.
374. Mooij, W. T.; Mitsiki, E.; Perrakis, A., ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. *Nucleic Acids Res.* **2009,** *37* (Web Server issue), W402-5.
375. Gheyi, T.; Rodgers, L.; Romero, R.*, et al.*, Mass spectrometry guided in situ proteolysis to obtain crystals for X-ray structure determination. *Journal of the American Society for Mass Spectrometry* **2010,** *21* (10), 1795-801.
376. Banci, L.; Bertini, I.; Cusack, S.*, et al.*, First steps towards effective methods in exploiting high-throughput technologies for the determination of human protein structures of high biomedical value. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 10), 1208-17.
377. Fontana, A.; de Laureto, P. P.; Spolaore, B.*, et al.*, Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* **2004,** *51* (2), 299-321.
378. Notredame, C.; Higgins, D. G.; Heringa, J., T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000,** *302* (1), 205-17.
379. Altschul, S. F.; Gish, W.; Miller, W.*, et al.*, Basic local alignment search tool. *J. Mol. Biol.* **1990,** *215* (3), 403-10.
380. Holm, L.; Rosenstrom, P., Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **2010,** *38* (Web Server issue), W545-9.
381. Ward, J. J.; McGuffin, L. J.; Bryson, K.*, et al.*, The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **2004,** *20* (13), 2138-9.
382. Ginalski, K.; Elofsson, A.; Fischer, D.*, et al.*, 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **2003,** *19* (8), 1015-8.
383. Rose, N. R.; Ng, S. S.; Mecinovic, J.*, et al.*, Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. *J. Med. Chem.* **2008,** *51* (22), 7053-6.
384. Yue, W. W.; Hozjan, V.; Ge, W.*, et al.*, Crystal structure of the PHF8 Jumonji domain, an Nepsilon-methyl lysine demethylase. *FEBS Lett.* **2010,** *584* (4), 825-30.
385. Sengoku, T.; Yokoyama, S., Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev.* **2011,** *25* (21), 2266-77.
386. Wang, Q. P.; VanDusen, W. J.; Petroski, C. J.*, et al.*, Bovine liver aspartyl beta-hydroxylase. Purification and characterization. *J. Biol. Chem.* **1991,** *266* (21), 14004-10.
387. Jia, S.; McGinnis, K.; VanDusen, W. J.*, et al.*, A fully active catalytic domain of bovine aspartyl (asparaginyl) beta-hydroxylase expressed in Escherichia coli: characterization and evidence for the identification of an active-site region in vertebrate alpha-ketoglutarate-dependent dioxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91* (15), 7227-31.
388. Gileadi, O.; Burgess-Brown, N. A.; Colebrook, S. M.*, et al.*, High Throughput Production of Recombinant Human Proteins for Crystallography. In *Methods in Molecular Biology; Structural Proteomics: High-throughput Methods*, Kobe, B. G., M.; Huber, T., Ed. Humana Press: 2008; Vol. 426, pp 221-246.
389. I.M.A.G.E.Consortium http://www.imageconsortium.org/.
390. Geoghegan, K. F.; Dixon, H. B.; Rosner, P. J.*, et al.*, Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **1999,** *267* (1), 169-84.
391. Yan, Z.; Caldwell, G. W.; McDonell, P. A., Identification of a gluconic acid derivative attached to the N-terminus of histidine-tagged proteins expressed in bacteria. *Biochem. Biophys. Res. Commun.* **1999,** *262* (3), 793-800.
392. El-Aneed, A.; Cohen, A.; Banoub, J., Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectroscopy Rev.* **2009,** *44* (3), 210-330.
393. Zhang, J. H.; Qi, R. C.; Chen, T.*, et al.*, Development of a carbon dioxide-capture assay in microtiter plate for aspartyl-beta-hydroxylase. *Anal. Biochem.* **1999,** *271* (2), 137-42.
394. Gali, H.; Sieckman, G. L.; Hoffman, T. J.*, et al.*, Chemical synthesis of Escherichia coli ST(h) analogues by regioselective disulfide bond formation: biological evaluation of an (111)In-DOTA-

Phe(19)-ST(h) analogue for specific targeting of human colon cancers. *Bioconjug. Chem.* **2002,** *13* (2), 224-31.
395. Garsky, V. M.; Lumma, P. K.; Freidinger, R. M.*, et al.*, Chemical synthesis of echistatin, a potent inhibitor of platelet aggregation from Echis carinatus: synthesis and biological activity of selected analogs. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (11), 4022-6.
396. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
397. Frecklington, D., General Method for MALDI-MS Analysis of Proteins and Peptides. *CSH Protoc* **2007,** *2007*, pdb prot4679.
398. Cotter, R. J., Time-of-flight mass spectrometry: an increasing role in the life sciences. *Biomed Environ Mass Spectrom* **1989,** *18* (8), 513-32.
399. Cotter, R. J., Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal. Chem.* **1992,** *64* (21), 1027A-1039A.
400. Jaskolla, T. W.; Lehmann, W. D.; Karas, M., 4-Chloro-alpha-cyanocinnamic acid is an advanced, rationally designed MALDI matrix. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (34), 12200-5.
401. Fournier, I.; Marinach, C.; Tabet, J. C.*, et al.*, Irradiation effects in MALDI, ablation, ion production, and surface modifications. Part II. 2,5-dihydroxybenzoic acid monocrystals. *Journal of the American Society for Mass Spectrometry* **2003,** *14* (8), 893-9.
402. Teearu, A.; Vahur, S.; Haljasorg, U.*, et al.*, 2,5-Dihydroxybenzoic acid solution in MALDI-MS: ageing and use for mass calibration. *J. Mass Spectrom.* **2014,** *49* (10), 970-9.
403. Ghafouri, B.; Karlsson, H.; Mortstedt, H.*, et al.*, 2,5-Dihydroxybenzoic acid instead of alpha-cyano-4-hydroxycinnamic acid as matrix in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for analyses of in-gel digests of silver-stained proteins. *Anal. Biochem.* **2007,** *371* (1), 121-3.
404. Gobom, J.; Schuerenberg, M.; Mueller, M.*, et al.*, Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. *Anal. Chem.* **2001,** *73* (3), 434-8.
405. Hewitson, K. S.; Schofield, C. J.; Ratcliffe, P. J., Hypoxia-inducible factor prolyl-hydroxylase: purification and assays of PHD2. *Methods Enzymol.* **2007,** *435*, 25-42.
406. Myllyla, R.; Majamaa, K.; Gunzler, V.*, et al.*, Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *J. Biol. Chem.* **1984,** *259* (9), 5403-5.
407. Flashman, E.; Davies, S. L.; Yeoh, K. K.*, et al.*, Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem. J.* **2010,** *427* (1), 135-42.
408. Horton, J. R.; Upadhyay, A. K.; Qi, H. H.*, et al.*, Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **2010,** *17* (1), 38-43.
409. Hopkinson, R. J.; Hamed, R. B.; Rose, N. R.*, et al.*, Monitoring the activity of 2-oxoglutarate dependent histone demethylases by NMR spectroscopy: direct observation of formaldehyde. *Chembiochem* **2010,** *11* (4), 506-10.
410. Stueber, D.; Ibrahimi, I.; Cutler, D.*, et al.*, A novel in vitro transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. *The EMBO journal* **1984,** *3* (13), 3143-8.
411. Miller, O. J.; Bernath, K.; Agresti, J. J.*, et al.*, Directed evolution by in vitro compartmentalization. *Nat Methods* **2006,** *3* (7), 561-70.
412. Davidson, E. A.; Dlugosz, P. J.; Levy, M.*, et al.*, Directed evolution of proteins in vitro using compartmentalization in emulsions. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2009,** *Chapter 24*, Unit 24 6.
413. Murthy, T. V.; Wu, W.; Qiu, Q. Q.*, et al.*, Bacterial cell-free system for high-throughput protein expression and a comparative analysis of Escherichia coli cell-free and whole cell expression systems. *Protein Expr. Purif.* **2004,** *36* (2), 217-25.
414. Schwarz, D.; Dotsch, V.; Bernhard, F., Production of membrane proteins using cell-free expression systems. *Proteomics* **2008,** *8* (19), 3933-46.

415. Galloway, C. A.; Sowden, M. P.; Smith, H. C., Increasing the yield of soluble recombinant protein expressed in E. coli by induction during late log phase. *BioTechniques* **2003,** *34* (3), 524-6, 528, 530.
416. Doublie, S., Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **1997,** *276*, 523-30.
417. Kleywegt, G. J., Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr* **2000,** *56* (Pt 3), 249-65.
418. Hendrickson, W. A., Analysis of Protein Structure from Diffraction Measurement at Multiple Wavelengths. *Trans. ACA* **1985,** *21*.
419. Hendrickson, W. A.; Ogata, C. M., Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **1997,** *276*, 494-523.
420. Ealick, S. E., Advances in multiple wavelength anomalous diffraction crystallography. *Curr. Opin. Chem. Biol.* **2000,** *4* (5), 495-9.
421. Drenth, J., Principles of Protein X-Ray Crystallography. 3rd ed.; Springer: 2007; p. 332.
422. Emsley, P.; Lohkamp, B.; Scott, W. G*., et al.*, Features and development of Coot. *Acta Cryst.* **2010,** *D66*, 486-501.
423. Adams, P. D.; Afonine, P. V.; Bunkóczi, G*., et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **2010,** *D66*, 213-21.
424. Schagger, H., Tricine-SDS-PAGE. *Nat Protoc* **2006,** *1* (1), 16-22.
425. Schagger, H.; von Jagow, G., Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **1987,** *166* (2), 368-79.
426. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970,** *227* (5259), 680-5.
427. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T*., et al.*, Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006,** *440* (7082), 368-71.
428. Hewitson, K. S.; Holmes, S. L.; Ehrismann, D*., et al.*, Evidence that two enzyme-derived histidine ligands are sufficient for iron binding and catalysis by factor inhibiting HIF (FIH). *J. Biol. Chem.* **2008,** *283* (38), 25971-8.
429. Matthews, M. L.; Neumann, C. S.; Miles, L. A*., et al.*, Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (42), 17723-8.
430. Gruber, C. W.; Cemazar, M.; Heras, B*., et al.*, Protein disulfide isomerase: the structure of oxidative folding. *Trends Biochem. Sci.* **2006,** *31* (8), 455-64.
431. Wilkinson, B.; Gilbert, H. F., Protein disulfide isomerase. *Biochim. Biophys. Acta* **2004,** *1699* (1-2), 35-44.
432. Lundstrom, J.; Holmgren, A., Protein disulfide-isomerase is a substrate for thioredoxin reductase and has thioredoxin-like activity. *J. Biol. Chem.* **1990,** *265* (16), 9114-20.
433. Ellgaard, L.; Ruddock, L. W., The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.* **2005,** *6* (1), 28-32.
434. Thornton, J. M., Disulphide bridges in globular proteins. *J. Mol. Biol.* **1981,** *151* (2), 261-87.
435. Creighton, T. E., Disulphide bonds and protein stability. *BioEssays* **1988,** *8* (2), 57-63.
436. Matsumura, M.; Signor, G.; Matthews, B. W., Substantial increase of protein stability by multiple disulphide bonds. *Nature* **1989,** *342* (6247), 291-3.
437. Maleknia, S. D.; Johnson, R., Mass Spectrometry of Amino Acids and Proteins. In *Amino Acids, Peptides and Proteins in Organic Chemistry: Analysis and Function of Amino Acids and Peptides*, Hughes, A. B., Ed. Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2011; Vol. 5.
438. Hardouin, J., Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry. *Mass spectrometry reviews* **2007,** *26* (5), 672-82.
439. Angel, T. E.; Aryal, U. K.; Hengel, S. M*., et al.*, Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.* **2012,** *41* (10), 3912-28.
440. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.
441. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009,** *11*, 49-79.

442. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat Methods* **2007,** *4* (10), 817-21.
443. Tran, J. C.; Zamdborg, L.; Ahlf, D. R.*, et al.*, Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011,** *480* (7376), 254-8.
444. Nemeth-Cawley, J. F.; Tangarone, B. S.; Rouse, J. C., "Top Down" characterization is a complementary technique to peptide sequencing for identifying protein species in complex mixtures. *J. Proteome Res.* **2003,** *2* (5), 495-505.
445. Kelleher, N. L., Top-down proteomics. *Anal. Chem.* **2004,** *76* (11), 197A-203A.
446. Reid, G. E.; McLuckey, S. A., 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **2002,** *37* (7), 663-75.
447. Johnson, R. S.; Martin, S. A.; Biemann, K.*, et al.*, Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987,** *59* (21), 2621-5.
448. Badock, V.; Raida, M.; Adermann, K.*, et al.*, Distinction between the three disulfide isomers of guanylin 99-115 by low-energy collision-induced dissociation. *Rapid Commun. Mass Spectrom.* **1998,** *12* (23), 1952-6.
449. Bean, M. F.; Carr, S. A., Characterization of disulfide bond position in proteins and sequence analysis of cystine-bridged peptides by tandem mass spectrometry. *Anal. Biochem.* **1992,** *201* (2), 216-26.
450. Gorman, J. J.; Ferguson, B. L.; Speelman, D.*, et al.*, Determination of the disulfide bond arrangement of human respiratory syncytial virus attachment (G) protein by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein Sci.* **1997,** *6* (6), 1308-15.
451. Kalkhof, S.; Witte, K.; Ihling, C. H.*, et al.*, A novel disulfide pattern in laminin-type epidermal growth factor-like (LE) modules of laminin beta1 and gamma1 chains. *Biochemistry* **2010,** *49* (38), 8359-66.
452. Gorman, J. J.; Wallis, T. P.; Pitt, J. J., Protein disulfide bond determination by mass spectrometry. *Mass spectrometry reviews* **2002,** *21* (3), 183-216.
453. Olsen, J. V.; Ong, S. E.; Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004,** *3* (6), 608-14.
454. Rodriguez, J.; Gupta, N.; Smith, R. D.*, et al.*, Does trypsin cut before proline? *J. Proteome Res.* **2008,** *7* (1), 300-5.
455. Banerjee, S.; Mazumdar, S., Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry* **2012,** *2012*, 282574.
456. Baumgart, S.; Lindner, Y.; Kuhne, R.*, et al.*, The contributions of specific amino acid side chains to signal intensities of peptides in matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004,** *18* (8), 863-8.
457. Warwood, S.; Mohammed, S.; Cristea, I. M.*, et al.*, Guanidination chemistry for qualitative and quantitative proteomics. *Rapid Commun. Mass Spectrom.* **2006,** *20* (21), 3245-56.
458. Cockrill, S. L.; Foster, K. L.; Wildsmith, J.*, et al.*, Efficient micro-recovery and guanidination of peptides directly from MALDI target spots. *BioTechniques* **2005,** *38* (2), 301-4.
459. Beardsley, R. L.; Reilly, J. P., Optimization of guanidination procedures for MALDI mass mapping. *Anal. Chem.* **2002,** *74* (8), 1884-90.
460. Knott, V.; Downing, A. K.; Cardy, C. M.*, et al.*, Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1. *J. Mol. Biol.* **1996,** *255* (1), 22-7.
461. Jensen, S. A.; Aspinall, G.; Handford, P. A., C-terminal propeptide is required for fibrillin-1 secretion and blocks premature assembly through linkage to domains cbEGF41-43. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (28), 10155-60.
462. Cordle, J.; Johnson, S.; Tay, J. Z.*, et al.*, A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat. Struct. Mol. Biol.* **2008,** *15* (8), 849-57.
463. Hambleton, S.; Valeyev, N. V.; Muranyi, A.*, et al.*, Structural and functional properties of the human notch-1 ligand binding region. *Structure* **2004,** *12* (12), 2173-83.
464. Lee, S. S.; Knott, V.; Jovanovic, J.*, et al.*, Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization. *Structure* **2004,** *12* (4), 717-29.

465. Krokhin, O. V.; Ens, W.; Standing, K. G., Characterizing degradation products of peptides containing N-terminal Cys residues by (off-line high-performance liquid chromatography)/matrix-assisted laser desorption/ionization quadrupole time-of-flight measurements. *Rapid Commun. Mass Spectrom.* **2003,** *17* (22), 2528-34.
466. Geoghegan, K. F.; Hoth, L. R.; Tan, D. H.*, et al.*, Cyclization of N-terminal S-carbamoylmethylcysteine causing loss of 17 Da from peptides and extra peaks in peptide maps. *J. Proteome Res.* **2002,** *1* (2), 181-7.
467. Zhang, X.; Hojrup, P., Cyclization of the N-terminal X-Asn-Gly motif during sample preparation for bottom-up proteomics. *Anal. Chem.* **2010,** *82* (20), 8680-5.
468. Ng, A. S.; Kini, R. M., Structural determinants in protein folding: a single conserved hydrophobic residue determines folding of EGF domains. *ACS Chem Biol* **2013,** *8* (1), 161-9.
469. Singh, R.; Whitesides, G. M., Thiol—disulfide interchange. In *Sulphur-Containing Functional Groups*, Pata, S.; Rappoport, Z., Eds. John Wiley & Sons, Inc.: Chichester, UK, 1993.
470. Sevier, C. S.; Kaiser, C. A., Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002,** *3* (11), 836-47.
471. Sela, M.; Lifson, S., The reformation of disulfide bridges in proteins. *Biochim. Biophys. Acta* **1959,** *36*, 471-8.
472. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2007,** *2* (9), 2212-21.
473. Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.*, et al.*, Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **2006,** *357* (2), 289-98.
474. Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.*, et al.*, Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (43), 15835-40.
475. White, C. E.; Hunter, M. J.; Meininger, D. P.*, et al.*, The fifth epidermal growth factor-like domain of thrombomodulin does not have an epidermal growth factor-like disulfide bonding pattern. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93* (19), 10177-82.
476. Wood, M. J.; Sampoli Benitez, B. A.; Komives, E. A., Solution structure of the smallest cofactor-active fragment of thrombomodulin. *Nat. Struct. Biol.* **2000,** *7* (3), 200-4.
477. Fuentes-Prior, P.; Iwanaga, Y.; Huber, R.*, et al.*, Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* **2000,** *404* (6777), 518-25.
478. Hunter, M. J.; Komives, E. A., Thrombin-binding affinities of different disulfide-bonded isomers of the fifth EGF-like domain of thrombomodulin. *Protein Sci.* **1995,** *4* (10), 2129-37.
479. Chang, J. Y.; Li, L., The disulfide structure of denatured epidermal growth factor: preparation of scrambled disulfide isomers. *J. Protein Chem.* **2002,** *21* (3), 203-13.
480. Sampoli Benitez, B. A.; Komives, E. A., Disulfide bond plasticity in epidermal growth factor. *Proteins* **2000,** *40* (1), 168-74.
481. Chang, J. Y.; Schindler, P.; Ramseier, U.*, et al.*, The disulfide folding pathway of human epidermal growth factor. *J. Biol. Chem.* **1995,** *270* (16), 9207-16.
482. Harris, R. J.; Spellman, M. W., O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **1993,** *3* (3), 219-24.
483. Hase, S.; Nishimura, H.; Kawabata, S.*, et al.*, The structure of (xylose)2glucose-O-serine 53 found in the first epidermal growth factor-like domain of bovine blood clotting factor IX. *J. Biol. Chem.* **1990,** *265* (4), 1858-61.
484. Nishimura, H.; Kawabata, S.; Kisiel, W.*, et al.*, Identification of a disaccharide (Xyl-Glc) and a trisaccharide (Xyl2-Glc) O-glycosidically linked to a serine residue in the first epidermal growth factor-like domain of human factors VII and IX and protein Z and bovine protein Z. *J. Biol. Chem.* **1989,** *264* (34), 20320-5.
485. Kentzer, E. J.; Buko, A.; Menon, G.*, et al.*, Carbohydrate composition and presence of a fucose-protein linkage in recombinant human pro-urokinase. *Biochem. Biophys. Res. Commun.* **1990,** *171* (1), 401-6.
486. Harris, R. J.; Leonard, C. K.; Guzzetta, A. W.*, et al.*, Tissue plasminogen activator has an O-linked fucose attached to threonine-61 in the epidermal growth factor domain. *Biochemistry* **1991,** *30* (9), 2311-4.

487. Bjoern, S.; Foster, D. C.; Thim, L.*, et al.*, Human plasma and recombinant factor VII. Characterization of O-glycosylations at serine residues 52 and 60 and effects of site-directed mutagenesis of serine 52 to alanine. *J. Biol. Chem.* **1991,** *266* (17), 11051-7.
488. Harris, R. J.; van Halbeek, H.; Glushka, J.*, et al.*, Identification and structural analysis of the tetrasaccharide NeuAc alpha(2-->6)Gal beta(1-->4)GlcNAc beta(1-->3)Fuc alpha 1-->O-linked to serine 61 of human factor IX. *Biochemistry* **1993,** *32* (26), 6539-47.
489. Harris, R. J.; Ling, V. T.; Spellman, M. W., O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C. *J. Biol. Chem.* **1992,** *267* (8), 5102-7.
490. Haines, N.; Irvine, K. D., Glycosylation regulates Notch signalling. *Nat. Rev. Mol. Cell Biol.* **2003,** *4* (10), 786-97.
491. Luo, Y.; Haltiwanger, R. S., O-fucosylation of notch occurs in the endoplasmic reticulum. *J. Biol. Chem.* **2005,** *280* (12), 11289-94.
492. Wang, Y.; Spellman, M. W., Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. *J. Biol. Chem.* **1998,** *273* (14), 8112-8.
493. Goto, Y.; Ohta, A.; Sako, Y.*, et al.*, Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem. Biol.* **2008,** *3* (2), 120-29.
494. Hughes, P. E.; Morgan, G.; Rooney, E. K.*, et al.*, Tyrosine 69 of the first epidermal growth factor-like domain of human factor IX is essential for clotting activity. *J. Biol. Chem.* **1993,** *268* (24), 17727-33.
495. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.*, et al.*, Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (43), 16793-7.
496. Lutz, J.-F., Copper-Free Azide–Alkyne Cycloadditions: New Insights and Perspectives. *Angew. Chem. Int. Ed.* **2008,** *47* (12), 2182-84.
497. Jawalekar, A. M.; Malik, S.; Verkade, J. M.*, et al.*, Oligonucleotide tagging for copper-free click conjugation. *Molecules* **2013,** *18* (7), 7346-63.
498. Dommerholt, J.; Schmidt, S.; Temming, R.*, et al.*, Readily accessible bicyclononynes for bioorthogonal labeling and three-dimensional imaging of living cells. *Angew Chem Int Ed Engl* **2010,** *49* (49), 9422-5.
499. Huisgen, R., The Concerted Nature of 1,3-Dipolar Cycloadditions and the Question of Diradical Intermediates. *J. Org. Chem.* **1976,** *41* (3), 403-19.
500. Yang, M.; Hardy, A. P.; Chowdhury, R.*, et al.*, Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl* **2013,** *52* (6), 1700-4.
501. Sugo, T.; Fernlund, P.; Stenflo, J., erythro-beta-Hydroxyaspartic acid in bovine factor IX and factor X. *FEBS Lett.* **1984,** *165* (1), 102-6.
502. Fernlund, P.; Stenflo, J., Amino acid sequence of the light chain of bovine protein C. *J. Biol. Chem.* **1982,** *257* (20), 12170-9.
503. Stenflo, J.; Lundwall, A.; Dahlback, B., beta-Hydroxyasparagine in domains homologous to the epidermal growth factor precursor in vitamin K-dependent protein S. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84* (2), 368-72.
504. Hojrup, P.; Jensen, M. S.; Petersen, T. E., Amino acid sequence of bovine protein Z: a vitamin K-dependent serine protease homolog. *FEBS Lett.* **1985,** *184* (2), 333-8.
505. Przysiecki, C. T.; Staggers, J. E.; Ramjit, H. G.*, et al.*, Occurrence of beta-hydroxylated asparagine residues in non-vitamin K-dependent proteins containing epidermal growth factor-like domains. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84* (22), 7856-60.
506. Stenflo, J.; Ohlin, A. K.; Owen, W. G.*, et al.*, beta-Hydroxyaspartic acid or beta-hydroxyasparagine in bovine low density lipoprotein receptor and in bovine thrombomodulin. *J. Biol. Chem.* **1988,** *263* (1), 21-4.
507. Sanglier, S.; Atmanene, C.; Chevreux, G.*, et al.*, Nondenaturing mass spectrometry to study noncovalent protein/protein and protein/ligand complexes: technical aspects and application to the determination of binding stoichiometries. *Methods Mol. Biol.* **2008,** *484*, 217-43.
508. Lorenzen, K.; van Duijn, E., Native mass spectrometry as a tool in structural biology. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **2010,** *Chapter 17*, Unit17 12.
509. McNeill, L. A.; Flashman, E.; Buck, M. R.*, et al.*, Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol Biosyst* **2005,** *1* (4), 321-4.

510. Mecinovic, J.; Chowdhury, R.; Flashman, E.*, et al.*, Use of mass spectrometry to probe the nucleophilicity of cysteinyl residues of prolyl hydroxylase domain 2. *Anal. Biochem.* **2009,** *393* (2), 215-21.
511. Selevsek, N.; Tholey, A.; Heinzle, E.*, et al.*, Studies on ternary metallo-beta lactamase-inhibitor complexes using electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2006,** *17* (7), 1000-4.
512. Kondrat, F. D.; Struwe, W. B.; Benesch, J. L., Native mass spectrometry: towards high-throughput structural proteomics. *Methods Mol. Biol.* **2015,** *1261*, 349-71.
513. Shukla, A. K.; Futrell, J. H., Tandem mass spectrometry: dissociation of ions by collisional activation. *J. Mass Spectrom.* **2000,** *35* (9), 1069-90.
514. Ashcroft, A. E., *Ionization Methods in Organic Mass Spectrometry*. Royal Society of Chemistry: 1997.
515. Morrison, K. L.; Weiss, G. A., Combinatorial alanine-scanning. *Curr. Opin. Chem. Biol.* **2001,** *5* (3), 302-7.
516. Weiss, G. A.; Watanabe, C. K.; Zhong, A.*, et al.*, Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (16), 8950-4.
517. Glaser, F.; Pupko, T.; Paz, I.*, et al.*, ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **2003,** *19* (1), 163-4.
518. Armon, A.; Graur, D.; Ben-Tal, N., ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* **2001,** *307* (1), 447-63.
519. Blatch, G. L.; Lassle, M., The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* **1999,** *21* (11), 932-9.
520. Cortajarena, A. L.; Regan, L., Ligand binding by TPR domains. *Protein Sci.* **2006,** *15* (5), 1193-8.
521. Lamb, J. R.; Tugendreich, S.; Hieter, P., Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **1995,** *20* (7), 257-9.
522. D'Andrea, L. D.; Regan, L., TPR proteins: the versatile helix. *Trends Biochem. Sci.* **2003,** *28* (12), 655-62.
523. Pekkala, M.; Hieta, R.; Bergmann, U.*, et al.*, The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. *J. Biol. Chem.* **2004,** *279* (50), 52255-61.
524. Scheufler, C.; Brinker, A.; Bourenkov, G.*, et al.*, Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **2000,** *101* (2), 199-210.
525. Cliff, M. J.; Harris, R.; Barford, D.*, et al.*, Conformational diversity in the TPR domain-mediated interaction of protein phosphatase 5 with Hsp90. *Structure* **2006,** *14* (3), 415-26.
526. Das, A. K.; Cohen, P. W.; Barford, D., The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *The EMBO journal* **1998,** *17* (5), 1192-9.
527. Suk, J. Y.; Jensen, S.; McGettrick, A.*, et al.*, Structural consequences of cysteine substitutions C1977Y and C1977R in calcium-binding epidermal growth factor-like domain 30 of human fibrillin-1. *J. Biol. Chem.* **2004,** *279* (49), 51258-65.
528. Stephenson, R. C.; Clarke, S., Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J. Biol. Chem.* **1989,** *264* (11), 6164-70.
529. Huang, H. Z.; Nichols, A.; Liu, D., Direct identification and quantification of aspartyl succinimide in an IgG2 mAb by RapiGest assisted digestion. *Anal. Chem.* **2009,** *81* (4), 1686-92.
530. Violand, B. N.; Schlittler, M. R.; Kolodziej, E. W.*, et al.*, Isolation and characterization of porcine somatotropin containing a succinimide residue in place of aspartate129. *Protein Sci.* **1992,** *1* (12), 1634-41.
531. Teshima, G.; Stults, J. T.; Ling, V.*, et al.*, Isolation and characterization of a succinimide variant of methionyl human growth hormone. *J. Biol. Chem.* **1991,** *266* (21), 13544-7.
532. Yang, H.; Zubarev, R. A., Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. *Electrophoresis* **2010,** *31* (11), 1764-72.

533. Nicolás, E.; Pedroso, E.; Girald, E., Formation of aspartimide peptides in Asp-Gly sequences. *Tetrahedron Lett.* **1989,** *30* (4), 497-500.
534. Mergler, M.; Dick, F.; Sax, B.*, et al.*, The aspartimide problem in Fmoc-based SPPS. Part I. *J. Pept. Sci.* **2003,** *9* (1), 36-46.
535. Mergler, M.; Dick, F.; Sax, B.*, et al.*, The aspartimide problem in Fmoc-based SPPS. Part II. *J. Pept. Sci.* **2003,** *9* (8), 518-26.
536. Mergler, M.; Dick, F., The aspartimide problem in Fmoc-based SPPS. Part III. *J. Pept. Sci.* **2005,** *11* (10), 650-7.
537. Palasek, S. A.; Cox, Z. J.; Collins, J. M., Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. *J. Pept. Sci.* **2007,** *13* (3), 143-8.
538. Yu, X. C.; Joe, K.; Zhang, Y.*, et al.*, Accurate determination of succinimide degradation products using high fidelity trypsin digestion peptide map analysis. *Anal. Chem.* **2011,** *83* (15), 5912-9.
539. Clapham, D. E., Calcium signaling. *Cell* **2007,** *131* (6), 1047-58.
540. Bootman, M. D.; Berridge, M. J.; Roderick, H. L., Calcium signalling: more messengers, more channels, more complexity. *Curr. Biol.* **2002,** *12* (16), R563-5.
541. Berridge, M. J.; Bootman, M. D.; Roderick, H. L., Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.* **2003,** *4* (7), 517-29.
542. Berridge, M. J.; Lipp, P.; Bootman, M. D., The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* **2000,** *1* (1), 11-21.
543. Stenflo, J.; Suttie, J. W., Vitamin K-dependent formation of gamma-carboxyglutamic acid. *Annu. Rev. Biochem.* **1977,** *46*, 157-72.
544. Stenflo, J.; Stenberg, Y.; Muranyi, A., Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions. *Biochim. Biophys. Acta* **2000,** *1477* (1-2), 51-63.
545. Handford, P.; Downing, A. K.; Rao, Z.*, et al.*, The calcium binding properties and molecular organization of epidermal growth factor-like domains in human fibrillin-1. *J. Biol. Chem.* **1995,** *270* (12), 6751-6.
546. Whiteman, P.; Downing, A. K.; Smallridge, R.*, et al.*, A Gly --> Ser change causes defective folding in vitro of calcium-binding epidermal growth factor-like domains from factor IX and fibrillin-1. *J. Biol. Chem.* **1998,** *273* (14), 7807-13.
547. Smallridge, R. S.; Whiteman, P.; Doering, K.*, et al.*, EGF-like domain calcium affinity modulated by N-terminal domain linkage in human fibrillin-1. *J. Mol. Biol.* **1999,** *286* (3), 661-8.
548. Werner, J. M.; Knott, V.; Handford, P. A.*, et al.*, Backbone dynamics of a cbEGF domain pair in the presence of calcium. *J. Mol. Biol.* **2000,** *296* (4), 1065-78.
549. Haun, R. S.; Serventi, I. M.; Moss, J., Rapid, Reliable Ligation-Independent Cloning of PCR Products Using Modified Plasmid Vectors. *BioTechniques* **1992,** *13* (4), 515-18.
550. Kucera, R. B.; Nichols, N. M., DNA-dependent DNA polymerases. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2008,** *Chapter 3*, Unit3 5.
551. Xian, Z. H.; Zhang, S. H.; Cong, W. M.*, et al.*, Expression of aspartyl beta-hydroxylase and its clinicopathological significance in hepatocellular carcinoma. *Mod. Pathol.* **2006,** *19* (2), 280-6.
552. Keller, S.; Sanderson, M. P.; Stoeck, A.*, et al.*, Exosomes: from biogenesis and secretion to biological function. *Immunol. Lett.* **2006,** *107* (2), 102-8.
553. Hogan, P. G.; Lewis, R. S.; Rao, A., Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annu. Rev. Immunol.* **2010,** *28*, 491-533.
554. Soboloff, J.; Rothberg, B. S.; Madesh, M.*, et al.*, STIM proteins: dynamic calcium signal transducers. *Nat. Rev. Mol. Cell Biol.* **2012,** *13* (9), 549-65.
555. Luik, R. M.; Wang, B.; Prakriya, M.*, et al.*, Oligomerization of STIM1 couples ER calcium depletion to CRAC channel activation. *Nature* **2008,** *454* (7203), 538-42.
556. Smyth, J. T.; Petranka, J. G.; Boyles, R. R.*, et al.*, Phosphorylation of STIM1 underlies suppression of store-operated calcium entry during mitosis. *Nat. Cell Biol.* **2009,** *11* (12), 1465-72.
557. Pozo-Guisado, E.; Martin-Romero, F. J., The regulation of STIM1 by phosphorylation. *Commun Integr Biol* **2013,** *6* (6), e26283.
558. Hou, X.; Pedi, L.; Diver, M. M.*, et al.*, Crystal structure of the calcium release-activated calcium channel Orai. *Science* **2012,** *338* (6112), 1308-13.

559. Yang, X.; Jin, H.; Cai, X.*, et al.*, Structural and mechanistic insights into the activation of Stromal interaction molecule 1 (STIM1). *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (15), 5657-62.
560. Stathopulos, P. B.; Zheng, L.; Li, G. Y.*, et al.*, Structural and mechanistic insights into STIM1-mediated initiation of store-operated calcium entry. *Cell* **2008,** *135* (1), 110-22.
561. Wang, S.; Trumble, W. R.; Liao, H.*, et al.*, Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. *Nat. Struct. Biol.* **1998,** *5* (6), 476-83.
562. Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884-90.
563. Dobson, C. M., Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **2004,** *15* (1), 3-16.
564. Ron, D.; Walter, P., Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **2007,** *8* (7), 519-29.
565. Schroder, M., The unfolded protein response. *Mol. Biotechnol.* **2006,** *34* (2), 279-90.
566. Travers, K. J.; Patil, C. K.; Wodicka, L.*, et al.*, Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **2000,** *101* (3), 249-58.
567. Kaneko, M.; Nomura, Y., ER signaling in unfolded protein response. *Life Sci.* **2003,** *74* (2-3), 199-205.
568. Chakrabarti, A.; Chen, A. W.; Varner, J. D., A review of the mammalian unfolded protein response. *Biotechnol. Bioeng.* **2011,** *108* (12), 2777-93.
569. Meusser, B.; Hirsch, C.; Jarosch, E.*, et al.*, ERAD: the long road to destruction. *Nat. Cell Biol.* **2005,** *7* (8), 766-72.
570. Eisele, F.; Schafer, A.; Wolf, D. H., Ubiquitylation in the ERAD Pathway. *Subcell. Biochem.* **2010,** *54*, 136-48.
571. Vashist, S.; Ng, D. T., Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control. *J. Cell Biol.* **2004,** *165* (1), 41-52.
572. Vembar, S. S.; Brodsky, J. L., One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* **2008,** *9* (12), 944-57.
573. Lucero, H. A.; Kaminer, B., The role of calcium on the activity of ERcalcistorin/protein-disulfide isomerase and the significance of the C-terminal and its calcium binding. A comparison with mammalian protein-disulfide isomerase. *J. Biol. Chem.* **1999,** *274* (5), 3243-51.
574. Lievremont, J. P.; Rizzuto, R.; Hendershot, L.*, et al.*, BiP, a major chaperone protein of the endoplasmic reticulum lumen, plays a direct and important role in the storage of the rapidly exchanging pool of Ca2+. *J. Biol. Chem.* **1997,** *272* (49), 30873-9.
575. Schrag, J. D.; Bergeron, J. J.; Li, Y.*, et al.*, The Structure of calnexin, an ER chaperone involved in quality control of protein folding. *Mol. Cell* **2001,** *8* (3), 633-44.
576. Corbett, E. F.; Michalak, K. M.; Oikawa, K.*, et al.*, The conformation of calreticulin is influenced by the endoplasmic reticulum luminal environment. *J. Biol. Chem.* **2000,** *275* (35), 27177-85.
577. Yoshida, H., ER stress and diseases. *FEBS J.* **2007,** *274* (3), 630-58.
578. Yanez, M.; Gil-Longo, J.; Campos-Toimil, M., Calcium binding proteins. *Adv. Exp. Med. Biol.* **2012,** *740*, 461-82.
579. Persechini, A.; Moncrief, N. D.; Kretsinger, R. H., The EF-hand family of calcium-modulated proteins. *Trends Neurosci.* **1989,** *12* (11), 462-7.
580. Kawasaki, H.; Nakayama, S.; Kretsinger, R. H., Classification and evolution of EF-hand proteins. *Biometals* **1998,** *11* (4), 277-95.
581. Radtke, F.; Raj, K., The role of Notch in tumorigenesis: oncogene or tumour suppressor? *Nat. Rev. Cancer* **2003,** *3* (10), 756-67.
582. Dvorak, H. F., Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N. Engl. J. Med.* **1986,** *315* (26), 1650-9.
583. Monteith, G. R.; McAndrew, D.; Faddy, H. M.*, et al.*, Calcium and cancer: targeting Ca2+ transport. *Nat. Rev. Cancer* **2007,** *7* (7), 519-30.
584. Macian, F., NFAT proteins: key regulators of T-cell development and function. *Nat. Rev. Immunol.* **2005,** *5* (6), 472-84.
585. Gomez-Ospina, N.; Tsuruta, F.; Barreto-Chang, O.*, et al.*, The C terminus of the L-type voltage-gated calcium channel Ca(V)1.2 encodes a transcription factor. *Cell* **2006,** *127* (3), 591-606.

586. Berndt, C.; Lillig, C. H.; Holmgren, A., Thioredoxins and glutaredoxins as facilitators of protein folding. *Biochim. Biophys. Acta* **2008,** *1783* (4), 641-50.
587. Bray, S. J., Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* **2006,** *7* (9), 678-89.
588. Dong, X.; Lin, Q.; Aihara, A.*, et al.*, Aspartate beta-Hydroxylase expression promotes a malignant pancreatic cellular phenotype. *Oncotarget* **2015,** *6* (2), 1231-48.
589. Aihara, A.; Huang, C. K.; Olsen, M. J.*, et al.*, A Cell surface beta-Hydroxylase is a biomarker and therapeutic target for hepatocellular carcinoma. *Hepatology* **2014**.
590. Handford, P. A.; Baron, M.; Mayhew, M.*, et al.*, The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *The EMBO journal* **1990,** *9* (2), 475-80.
591. Papworth, C.; Bauer, J. C.; Braman, J.*, et al.*, Site-directed mutagenesis in one day with >80% efficiency. *Strategies* **1996,** *9* (3), 3-4.
592. http://web.expasy.org/protparam/.
593. Wojciechowski, F.; Suchy, M.; Li, A. X.*, et al.*, A Robust and Convergent Synthesis of Dipeptide−DOTAM Conjugates as Chelators for Lanthanide Ions: New PARACEST MRI Agents. *Bioconjug. Chem.* **2007,** *18* (5), 1625-36.
594. Feng, T.; Yamamoto, A.; Wilkins, S. E.*, et al.*, Optimal Translational Termination Requires C4 Lysyl Hydroxylation of eRF1. *Mol. Cell* **2014**.
595. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307-26.
596. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.*, et al.*, Phaser crystallographic software. *J. Appl. Cryst.* **2007,** *40*, 658-74.