Diboron-Enabled High-Precision Protein Engineering



Tim A. Mollner

University of Oxford, Department of Chemistry

Trinity Term 2021

A thesis submitted in partial fulfilment of the degree

Doctor of Philosophy

Diboron-Enabled High-Precision Protein Engineering

Tim A. Mollner

University of Oxford, Trinity Term 2021

Boronic acids are versatile synthetic building blocks, widely applied in molecular sensing applications and have more recently attracted attention as warheads in medicinal chemistry. Yet, boron containing organic compounds are rarely found in nature. Despite boron being considered a "probably essential" element which is found across all domains of life, organoboron compounds have only been reported to occur naturally in form of boric acid esters. These compounds are known to fulfil several biological functions, including cell-wall strength modulation and quorum sensing. In fact, boronic acids display unique reactivity towards Lewis bases, allowing for the dynamic and reversible formation of covalent interactions with suitable ligands. Precisely engineered boronic acids placed within biomacromolecules, such as proteins, could allow for altered or novel, previously not accessible, function and enable the selective manipulation of biological systems through specific programming of *de novo* binding.

Herein, a method for the site-selective introduction of a minimal boronic acid motif – boronoalanine (Bal) – into proteins via copper-catalysed borylation of dehydroalanine (Dha) residues with a diboron reagent is described. The compatibility of this borylation reaction with a range of different proteins was confirmed and detailed studies on the application of Bal as a binder of endogenous and exogenous ligands were conducted. Several proof-of-principle applications of Bal-dependent protein Lewis acid·base pair (PLABP) probing were developed, including modulation of protein stability, glycan sensing and utilization of Bal as a NMR probe to study nucleosome structure.

Lewis base binding was extended to fluoride, leading to the formation of trifluoroboronoalanine (3F-Bal). Trifluoroborates have been shown to directly enable ^{18}F-radiolabelling via ^{19}F$-$^{18}F isotope exchange. Introduction of 3F-Bal into proteins could thus enable radiolabelling in one step directly employing [^{18}F]fluoride.

Furthermore, aided by mechanistic observations, a diboron-promoted site-selective radical protein cleavage protocol was developed. This dehydroalanine-specific cleavage was found to proceed well on a range of proteins while providing a scarce example of a protein compatible, chemical proteolysis procedure.

Author's Declaration

The work presented in this thesis was conducted between October 2017 and February 2021 at the Department of Chemistry, University of Oxford under supervision of Prof. Ben G. Davis and Prof. Véronique Gouverneur and at Janssen Research & Development, Beerse under supervision of Dr. Daniel Oehlrich. All the work is my own, except where otherwise stated, and has not been submitted for any other degree at this or any other university.

The following experiments were performed in collaboration:

Protein CON NMR measurements were performed by Prof. D. Flemming Hansen, Institute of Structural and Molecular Biology, Division of Biosciences, University College London, UK.

Initial nucleosome phosphorylation and NMR experiments were conducted by Dr. Lukas Lercher, Department of Chemistry, Chemistry Research Laboratory, University of Oxford, UK.

Nucleosomal DNA preparation and nucleosome assembly was performed by Brian Josephson, Department of Chemistry, Chemistry Research Laboratory, University of Oxford, UK.

DFT calculations were performed by Prof. Andrew J. Baldwin; nucleosome NMR experiments were conducted and analysed by Charles Buchanan and Prof. Andrew J. Baldwin, Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, UK.

Titration of model compound Ac-Bal-NHBn was performed by Geert Pille, Janssen Research & Development, Beerse, Belgium.

AcrA, Histone H4, FLAG-Histone H3 and PanC mutants were kindly provided by Brian Josephson, Oluwatobi Arisa, Adeline Poh and Dr. Patrick Isenegger, Department of Chemistry, Chemistry Research Laboratory, University of Oxford, UK.

LC-MS/MS samples were processed at the Advanced Proteomics Facility, Department of Biochemistry, University of Oxford, UK by Svenja S. Hester and Dr. Marjorie Fournier.

Tim A. Mollner
April 2021

Chapter cover pages were designed using previously published protein structures (Chapter 1: PDB 5EIL, Chapter 3: PDB 1HVD, Chapter 4 and Chapter 6: 1KX5). Chapter covers are an artistic interpretation and not necessarily scientifically accurate.

Acknowledgements

This thesis would not have been possible in its current form without the generous support of many people, for which I am very grateful!

First of all, I would like to thank my supervisors Prof. Ben Davis and Prof. Véronique Gouverneur for giving me the opportunity to pursue my DPhil in their groups in Oxford. Their trust, support and advice has been crucial throughout my research. I am grateful for the scientific freedom I have received, without which conducting a project like this would not have been possible. Thank you!

I wish to thank the many people who have contributed to my project through various collaborations: Prof. Andrew Baldwin, Prof. Flemming Hansen, Dr. Patrick Isenegger, Brian Josephson and Charles Buchanan. Your hard work has helped to further this project in a direction that would not have been possible otherwise.

I am indebted to Dr. Patrick Isenegger for his support especially during my first months, to Brian Josephson for his patience and guidance with everything biology and to Aziz Khan for his help with flow cytometry. I would like to thank Elisabete Pires and Dr. John Walsby-Tickle for their support with mass spectrometry, Dr. Nader Amin and Prof. Tim Claridge for NMR advice, and Svenja Hester, Dr. Simon Nadal and Dr. Marjorie Fournier for their help with MS/MS experiments. I wish to thank Dr. Manuja Kaluarachchi-Soden for taking care of so many administrative tasks. I am grateful to members of both, the BGD and the VG group, for scientific discussions, advice and collegiality.

Oluwatobi Arisa, Dr. Alexander Dürr, Adeline Poh, Brian Josephson and Claudio Meyer very kindly proof-read this thesis. Thank you very much for your advice and suggestions!

During my DPhil, I had the privilege to spend a total of one year in Belgium working at Janssen. I am very grateful to my industrial supervisor Dr. Daniel Oehlrich for his warm welcome and can-do attitude. I would like to thank the whole team for discussions, advice and making my time in Belgium enjoyable. I wish to thank Michel Carpentier for HPLC purification of compounds, Geert Pille for pK_a measurements, Dr. Alex de Groot for help with NMR spectroscopy and Wouter Bruinzeel for help with protein purification and access to his facilities. Thank you to everyone in Building 020 for the nice and collegial atmosphere!

I would like to extend my thanks to all members of the FLUDD project for advice, good times and making this project successful.

During my time in Oxford, I have spent countless hours in various labs. I am grateful to everyone who made these hours truly enjoyable – you know who you are! I wish to thank my colleagues from ICL and CRL F11, where I spent most of my time, for the great atmosphere and all the banter.

Thanks to my colleagues and friends (in Oxford and around the world) for all the good times outside of the labs, chats, discussions, beers, jokes, shared meals, pints, formals, parties, tequila shots, fry-ups, trips and experiences that made my time in Oxford (and Belgium) most memorable. Let's stay in touch!

Finally, I indebted to my parents for their unconditional support throughout all the years of perusing chemistry. This would not have been possible without you!

Contents

Abstract i

Author's Declaration ii

Acknowledgements iv

List of Figures xiii

List of Schemes xix

List of Tables xxi

Abbreviations xxii

1 Introduction 2

1.1 Chemical Protein Modification 3

1.1.1 Contemporary Protein Modification Strategies: An Overview 3

1.1.2 'Tag-and-modify' Approach: Dehydroalanine (Dha) Chemistry 4

1.2 Proximity Enabled Covalent Trapping 7

1.2.1 Cys-reactive Artificial Covalent Linkers 8

1.2.2 Lys, His and Tyr Targeting Covalent Linkers 8

1.2.3 Discrimination between *De Novo* Binding and Covalent Trapping . . . 9

1.3 Engineered *De Novo* Binding Function 9

1.3.1 Introduction 9

1.3.2 Metal Binding 10

1.3.3 Reversible Binding 13

1.4 Conclusions and Outlook 15

2 Synthesis of Boronoproteins **24**

2.1 Introduction . . . 24

2.1.1 Introducing Boron Into Proteins . . . 24

2.1.2 Borylation of α,β-unsaturated Carbonyl Compounds . . . 26

2.2 Borylation of a Small-Molecule, Dha Containing Model Substrate . . . 28

2.3 Protein Borylation . . . 30

2.3.1 Influence of the Ligand and Diboron Reagent . . . 30

2.3.2 Influence of Buffer and pH . . . 32

2.3.3 Analysis of the Borylation Product . . . 33

2.4 Scope of Protein Borylation . . . 35

2.4.1 Preparation of Protein Substrates . . . 36

2.4.2 Borylation of Proteins . . . 38

2.4.3 Site-Selective Protein Deuteration . . . 40

2.5 Protein Compatibility of Borylation . . . 42

2.6 Conclusion . . . 45

3 Boronoalanine (Bal) Enabled *De Novo* Binding Function **51**

3.1 Boronic Acids as Reversible Covalent Binders . . . 51

3.1.1 Wulff-type Boronic Acids . . . 54

3.2 Protein Lewis Acid•Base Pair (PLABP) Probing . . . 56

3.2.1 Boron-Containing Natural Products . . . 56

3.2.2 Implications of Anchored Boron: B−C Bond Formation . . . 57

3.3 Intramolecular Binding . . . 59

3.3.1 Detection of Intramolecular Binding by Mass Spectrometry . . . 59

3.3.2 Intramolecular Engagement of Bal Modulates Thermostability . . . 61

3.4 Intermolecular Binding . . . 63

3.4.1 Binding Investigations Using Fluorinated Diols . . . 64

3.4.2 Interacting with Biological Systems . . . 73

3.5 Reactive PLABP . . . 76

3.6 Conclusion . . . 79

4 Application of Bal as a NMR Structural Probe **87**

4.1 Methods for the Detection of Long-Range Contacts in Intrinsically Disordered Proteins . . . 87

4.2 NMR Investigations on Small Molecule Model Compound . . . 88

4.2.1 NMR Studies on Isotopically Labelled Model Compound . . . 89

4.2.2 Probing Histone H3 Tail Conformations in a Nucleosome Context . . . 96

4.3 Conclusion . . . 101

5 Capturing Fluoride: Incorporation of Trifluoroboronoalanine **107**

5.1 From Boronic Acid to Trifluoroborate . . . 108

5.1.1 Understanding Trifluoroborate Stability . . . 109

5.1.2 Design of Stabilized Trifluoroborates . . . 111

5.2 Small-Molecule Studies on Trifluoroboronoalanine (3F-Bal) . . . 112

5.3 Formation of Trifluoroboronoalanine (3F-Bal) on Protein . . . 119

5.4 Conclusion & Outlook . . . 121

6 Mechanistic Insights & Protein Cleavage **128**

6.1 Introduction to Protein Cleavage . . . 129

6.1.1 Enzymatic Protein Cleavage . . . 129

6.1.2 Chemical Protein Cleavage . . . 130

6.2 Mechanistic Insights into Protein Borylation . . . 132

6.2.1 Small Molecule Studies . . . 132

6.2.2 Tetrahydroxydiboron-Promoted Reduction of Dha on Protein . . . 136

6.3 Diboron-Promoted, Site-Selective Radical Protein Cleavage . . . 137

6.3.1 Reaction Optimization . . . 138

6.3.2 Scope of Radical Protein Cleavage . . . 141

6.3.3 Analysis of the Cleavage Products: Identification of Terminal Modification 144

6.3.4 Mechanistic Investigations: Understanding Radical Protein Cleavage . 146

6.4 Conclusion & Outlook . . . 149

7 Experimental Methods & Materials **157**

7.1 General Methodology . . . 157

7.1.1 Synthetic Chemistry . . . 157

7.1.2 Chromatography . . . 157

7.1.3 Analytical Techniques . . . 158

7.2 Synthetic Chemistry . . . 159

7.2.1 Small Molecule Model . . . 159

7.2.2 Fluorinated Myrtenol Derivative . . . 168

7.2.3 Trifluoromethoxylated Myrtenol Derivative . . . 173

7.2.4 Myrtentriol . . . 174

7.2.5 2-Deoxy-2-fluoro-D-galactose (FDGal) . . . 175

7.2.6 Methyl 5-deoxy-5-fluoro-β,D-ribofuranoside (FDRibOMe) . . . 177

7.2.7 Ac-Gly-Dha-NHBn (**17**) . . . 179

7.2.8 Miscellaneous Compounds . . . 182

7.2.9 Mechanistic Experiments . . . 183

7.3 Model Substrate NMR Studies . . . 187

7.3.1 pH Dependency of Oxaborolane Formation . . . 188

7.3.2 Influence of Concentration . . . 188

7.3.3 ^{15}N-NMR Measurements . . . 189

7.4 Protein Expression and Purification . . . 190

7.4.1 Histone H3 Cys10 . . . 190

7.4.2 Histone H3 Cys9 . . . 193

7.4.3 [^{15}N]Histone H3 Cys9 and [^{15}N]Histone H3 Cys10 . . . 195

7.4.4 [^{13}C-^{15}N]Histone H3 Cys10 . . . 199

7.4.5 Annexin V Cys316 . . . 202

7.4.6 Npβ Cys61 . . . 205

7.4.7 pre-SUMO1 Cys51 . . . 208

7.4.8 pre-SUMO1 Cys51 Cys97 . . . 211

7.4.9 mCherry Cys131 . . . 214

7.4.10 PstS Cys197 . . . 217

7.5 Preparation of Protein Substrates 220
7.5.1 Histone H3 Dha10 220
7.5.2 [^{15}N]Histone H3 Dha10 221
7.5.3 [^{13}C-^{15}N]Histone H3 Dha10 221
7.5.4 Histone H3 Dha9 223
7.5.5 [^{15}N]Histone H3 Dha9 224
7.5.6 FLAG-Histone H3 Dha4 225
7.5.7 Histone H4 Dha16 226
7.5.8 Annexin V Dha316 227
7.5.9 Npβ Dha61 229
7.5.10 PanC Dha44 and Dha47 230
7.5.11 pre-SUMO1 Dha51 231
7.5.12 pre-SUMO1 Dha51 Dha97 232
7.5.13 mCherry Sulfonium131 233
7.5.14 PstS Dha197 234
7.5.15 AcrA Dha123 235
7.6 Protein Borylation & Reactions of Boronoproteins 236
7.6.1 General Borylation Protocol & Reaction Optimizations 236
7.6.2 Histone H3 Bal10 236
7.6.3 Histone H3 Ser10 238
7.6.4 Histone H3 Deuteration Study 239
7.6.5 [^{15}N]Histone H3 Bal10 241
7.6.6 [^{15}N]Histone H3 Ser10 242
7.6.7 [^{13}C-^{15}N]Histone H3 *d1*-Bal10 243
7.6.8 Histone H3 Bal9 245
7.6.9 [^{15}N]Histone H3 Bal9 247
7.6.10 [^{15}N]Histone H3 Ser9 248
7.6.11 Annexin V Bal316 249
7.6.12 Npβ Bal61 250
7.6.13 PanC Bal44 251

7.6.14 PanC Bal47 . . . 252
7.6.15 pre-SUMO1 Bal51 . . . 253
7.6.16 mCherry Bal131 . . . 254
7.6.17 PstS Bal197 . . . 255
7.6.18 AcrA Bal123 . . . 256
7.6.19 Histone H4 Bal16 . . . 257
7.7 Calculation of Residue Accessibility . . . 258
7.8 Reactive PLABP . . . 259
7.9 Trifluoroboronoalanine (3F-Bal) Formation & Hydrolysis . . . 260
7.9.1 3F-Bal Formation on Histone H3 Model Proteins . . . 260
7.9.2 Determination of Hydrolysis on Model Substrate . . . 260
7.10 Protein Cleavage . . . 261
7.10.1 Reaction Optimization . . . 261
7.10.2 Deuteration Experiments . . . 264
7.10.3 Protein Cleavage Scope . . . 266
7.10.4 Isotope Pattern Analysis . . . 279
7.11 Protein Sequencing . . . 282
7.12 Refolding and CD Spectroscopy of Selected Proteins . . . 283
7.12.1 General Procedure . . . 283
7.12.2 Histone H3 . . . 283
7.12.3 Annexin V . . . 284
7.12.4 PstS . . . 284
7.12.5 Npβ . . . 284
7.12.6 pre-SUMO1 . . . 285
7.13 Determination of Protein Melting Curves . . . 286
7.14 *In vitro* SUMOylation . . . 287
7.14.1 SUMOylation using SUMO1 Bal51 . . . 287
7.14.2 SUMOylation study using chemically cleaved SUMO1 . . . 289
7.15 Cell-based Binding Studies . . . 290
7.15.1 Annexin V Binding Study . . . 290

7.15.2 mCherry Binding Study 292
7.16 PstS Competition Assay 294
7.17 Observation of Diol Binding via NMR 295
7.17.1 K_D Determination on Protein 295
7.17.2 NMR-Titration of Histone H3 Bal9 with Fluorinated Diol **11** 297
7.17.3 Control Experiments 298
7.17.4 Determination of the Diastereomeric Ratio Using a Chiral Shift Reagent 300
7.18 Protein NMR 301
7.18.1 Protein CON NMR 301
7.18.2 Nucleosome NMR Experiments 301
7.19 Nucleosome NMR Study 301
7.19.1 Histone Expression and Purification 302
7.19.2 Nucleosome Assembly 302
7.19.3 Nucleosome Phosphorylation with Aurora B Kinase 302
7.19.4 NMR Aquisition and Processing for Borylated Nucleosomes 303

List of Figures

1.1 Reactive motifs for residue-specific, proximity enabled covalent cross linking . 7

1.2 Incorporation of metal binding unnatural amino acids into GFP for the engineering of metal-detecting biosensors 12

2.1 Influence of pH on protein borylation 33

2.2 ArgC digestion and LC-MS/MS peptide mapping of Histone H3 Bal10 and Histone H3 Ser10 34

2.3 ArgC digestion and LC-MS/MS peptide mapping using triol **4** 35

2.4 LC-MS analysis of Histone H3 *d1*-Ser 41

2.5 Protein compatibility of Dha borylation 44

3.1 Applications of boronic acid binding in a biological context 53

3.2 Enhancement of boronic acid·ligand binding through proximal Lewis bases . . 54

3.3 Examples of boron containing natural products 56

3.4 Intramolecular Lewis acid·base pairing 60

3.5 Determination of melting behaviour for selected proteins by nanoDSF 61

3.6 Polar interactions in selected proteins 62

3.7 Small molecule boronic acid based diol binders 63

3.8 Observation of the binding of fluorinated diols to small molecule model 66

3.9 Confirmation of PLABP with fluorinated diol **11** 68

3.10 Determination of dissociation constants for PLABP with fluorinated diols . . 70

3.11 Determination of diastereomeric ratio on protein using a ^{19}F-NMR diol probe 71

3.12 Borylation of mCherry yields a novel fluorescent protein sensor for the detection of cell surface diols 73

3.13 PLABP enables the engineering of selective and local proteolytic stability . . 75
3.14 Solvent-accessible surface area of the reaction site in different proteins 78

4.1 ^{15}N-NMR studies on ^{15}N-labelled model compound 91
4.2 HMBC NMR experiments suggest a constrained conformation for the C_β protons. 92
4.3 pH dependent oxaborolane and boronate formation 95
4.4 Application of Bal-mediated PLABP to probe structure using NMR 98
4.5 Observation of local structural features using a boronoalanine (Bal) NMR probe 99
4.6 Control experiments for the application of Bal as a NMR probe 100

5.1 Mechanistic considerations for trifluoroborate hydrolysis 110
5.2 Approaches for increasing the hydrolytic stability of trifluoroborates 111
5.3 Determination of hydrolysis of Ac-(3F-Bal)-NHBn (**15**) by ^{1}H-NMR 113
5.4 ^{19}F-NMR analysis of Ac-(3F-Bal)-NHBn (**15**) in different buffers 114
5.5 Determination of the influence of different buffer systems and additives on the position of the chemical equilibrium 115
5.6 ^{19}F-NMR analysis of Ac-(3F-Bal)-NHBn (**15**) in presence of different amino acids 117
5.7 Proposed modulation of the hydrolytic equilibrium driven 118
5.8 Observation of trifluoroborate formation on model protein Histone H3 120

6.1 Successful borylation of model compound Ac-Dha-NHBn (**1**) depends on the presence of oxygen. 133
6.2 Borylation under N_2, O_2 and H_2 atmosphere 134
6.3 Tetrahydroxydiboron-promoted reduction of Dha on protein 137
6.4 Additive screening for diboron-promoted site-selective protein cleavage 139
6.5 Influence of antioxidant concentration and temperature on protein cleavage . 140
6.6 Determination of cleavage conversions in a range of relevant buffers 141
6.7 Protein scope of the diboron-promoted, radical protein cleavage 142
6.8 Selective cleavage of preSUMO1 Dha51 Dha97 at Dha97 144
6.9 Analysis of the termini of the fragments observed after cleavage 145
6.10 Deuteration studies on diboron-mediated protein cleavage 147

7.1 Borylation under N_2, H_2 and O_2 atmosphere 184
7.2 Determination of Cu-particle size by filtration 185
7.3 TLC analysis of reduction with $B_2(OH)_4$ 186
7.4 NMR analysis of Ac-Bal-NHBn (**3**) 187
7.5 SDS-PAGE of Histone H3 Cys10 192
7.6 LC-MS analysis of Histone H3 Cys10 192
7.7 SDS-PAGE of Histone H3 Cys9 193
7.8 LC-MS analysis of Histone H3 Cys9 194
7.9 SDS-PAGE of [^{15}N]Histone H3 Cys9 197
7.10 LC-MS analysis of [^{15}N] Histone H3 Cys9 197
7.11 SDS-PAGE of [^{15}N]Histone H3 Cys10 198
7.12 LC-MS analysis of [^{15}N] Histone H3 Cys10 198
7.13 LC-MS analysis of [^{13}C-^{15}N]Histone H3 Cys10 201
7.14 SDS-PAGE of Annexin V Cys316 204
7.15 LC-MS analysis of Annexin V Cys316 204
7.16 SDS-PAGE of Npβ Cys61 207
7.17 LC-MS analysis of Npβ Cys61 207
7.18 SDS-PAGE of pre-SUMO1 Cys51 209
7.19 LC-MS analysis of pre-SUMO1 Cys51 210
7.20 SDS-PAGE of pre-SUMO1 Cys51 Cys97 213
7.21 LC-MS analysis of pre-SUMO1 Cys51 Cys97 213
7.22 SDS-PAGE of mCherry Cys131 215
7.23 LC-MS analysis of mCherry Cys131 216
7.24 SDS-PAGE of PstS Cys197 219
7.25 LC-MS analysis of PstS Cys197 219
7.26 LC-MS analysis of Histone H3 Dha10 220
7.27 LC-MS analysis of [^{15}N]Histone H3 Dha10 221
7.28 LC-MS analysis of [^{13}C-^{15}N]Histone H3 Dha10 222
7.29 LC-MS analysis of Histone H3 Dha9 223
7.30 LC-MS analysis of [^{15}N]Histone H3 Dha9 224

7.31 LC-MS analysis of Histone H4 Dha16 225
7.32 LC-MS analysis of Histone H4 Dha16 226
7.33 LC-MS analysis of Annexin V Dha316 227
7.34 LC-MS analysis of Npβ Dha61 229
7.35 LC-MS analysis of PanC Dha44 230
7.36 LC-MS analysis of pre-SUMO1 Dha51 231
7.37 LC-MS analysis of pre-SUMO1 Dha51 232
7.38 LC-MS analysis of mCherry Sulfonium131 233
7.39 LC-MS analysis of PstS Dha197 234
7.40 LC-MS analysis of AcrA Dha123 235
7.41 LC-MS analysis of Histone H3 Bal10 237
7.42 LC-MS analysis of Histone H3 Ser10 238
7.43 LC-MS analysis of Histone H3 *d1*-Bal10 239
7.44 LC-MS analysis of Histone H3 *d1*-Ser10 240
7.45 LC-MS analysis of [^{15}N]Histone H3 Bal10 241
7.46 LC-MS analysis of [^{15}N]Histone H3 Ser10 242
7.47 LC-MS analysis of [^{13}C-^{15}N]Histone H3 *d1*-Bal10 243
7.48 LC-MS analysis of Histone H3 Bal9 245
7.49 LC-MS analysis of [^{15}N]Histone H3 Bal9 247
7.50 LC-MS analysis of [^{15}N]Histone H3 Ser9 248
7.51 LC-MS analysis of Annexin V Bal316 249
7.52 LC-MS analysis of Npβ Bal61 250
7.53 LC-MS analysis of PanC Bal44 251
7.54 LC-MS analysis of PanC Bal47 252
7.55 LC-MS analysis of pre-SUMO1 Bal51 253
7.56 LC-MS analysis of mCherry Bal131 254
7.57 LC-MS analysis of PstS Bal197 255
7.58 LC-MS analysis of AcrA Bal123 256
7.59 LC-MS analysis of Histone H4 Bal16 257
7.60 Oxidation of borylated proteins 259

7.61 LC-MS analysis of cleaved FLAG-Histone H3 Dha4 266
7.62 LC-MS analysis of cleaved Histone H3 Dha9 267
7.63 LC-MS/MS analysis of cleaved Histone H3 Dha9 267
7.64 LC-MS analysis of cleaved Histone H3 Dha10 268
7.65 LC-MS analysis of cleaved Histone H4 Dha16 269
7.66 LC-MS analysis of the cleavage of Npβ Dha61 at room temperature 270
7.67 LC-MS analysis of the cleavage of Npβ Dha61 at 45°C 271
7.68 SDS-PAGE analysis of the cleavage of Npβ Dha61 271
7.69 SDS-PAGE analysis of the cleavage of preSUMO1 Dha51 273
7.70 LC-MS analysis of the cleavage of preSUMO1 Dha51 at room temperature . . 274
7.71 LC-MS analysis of the cleavage of preSUMO1 Dha51 Dha97 275
7.72 SDS-PAGE analysis of the cleavage of preSUMO1 Dha51 Dha97 276
7.73 LC-MS analysis of thia-Michael addition of BME to cleaved preSUMO1 Dha51 Dha97 277
7.74 LC-MS analysis of the cleavage of Annexin V Dha316 278
7.75 Isotope pattern analysis of cleaved of Npβ Dha61. 280
7.76 Isotope pattern analysis of cleaved of preSUMO1 Dha51. 281
7.77 CD spectra of selected borylated proteins 283
7.78 LC-MS analysis of matured SUMO1 Bal51 287
7.79 Detection of SUMOylation of a RanGAP1 fragment via Western Blot 288
7.80 Absorbance spectra of mCherry mutants 292
7.81 Determination of mCherry cell-surface binding by flow cytometry 293
7.82 ^{19}F-NMR spectrum of fluorinated diol **11** binding to Histone H3 Bal9 295
7.83 ^{19}F-NMR spectrum of fluorinated ribose derivative **12** binding to Histone H3 Bal9 296
7.84 ^{19}F-NMR spectrum of attempted binding of fluorinated galactose derivative **13** to Histone H3 Bal9 296
7.85 NMR-titration of Histone H3 Bal9 with fluorinated diol **11** 297
7.86 ^{19}F-NMR spectrum of fluorinated diol **11** binding to Histone H3 Bal9 with and without competitor 298

7.87 CMPG filtered ^{19}F-NMR spectrum of fluorinated diol **11** binding to Histone H3 Bal9 299
7.88 Determination of the enantiomeric ratio on Ac-Bal-NHBn 300

List of Schemes

1.1 Methods for residue-specific protein modification 3
1.2 Dehydroalanine (Dha) as a reactive 'tag' 5
1.3 Formation of C_β–C and C_β–heteroatom bonds from Dha 6
1.4 Genetically incorporated unnatural amino acids for *de novo* reversible covalent binding 14

2.1 Incorporation of boronic acids into proteins 25
2.2 Borylation of α,β-unsaturated carbonyl compounds 27
2.3 Synthesis of Ac-Dha-NHBn (**1**) 28
2.4 Borylation of model substrate Ac-Dha-NHBn (**1**) 28
2.5 Attempted Dha formation on mCherry Cys131 37
2.6 Protein Borylation: Reaction Sequence and Scope. 39
2.7 Borylation of mCherry Sulfonium131 40
2.8 Site-selective Protein Deuteration 40

3.1 Boronic acids are Lewis-acidic reversible covalent binders 52
3.2 Putative interactions of non-anchored and anchored boronic acids 57
3.3 Possible modes of interaction of boronoalanine (Bal) with intra- and intermolecular ligands. 58
3.4 Synthesis of ^{19}F-NMR diol probe **11** 64
3.5 Synthesis of fluorinated ribose derivative **12** 65
3.6 Synthesis of 2-Deoxy-2-fluoro-D-galactose (**13**) 65

4.1 Synthesis of ^{15}N-labelled isotopologues of small molecule model Ac-Bal-NHBn (**3**) 89

5.1 Boronic acid-based fluoride sensing 108
5.2 Synthesis of Ac-(3F-Bal)-NHBn (**15**) 112

6.1 Contemporary methods for the selective chemical cleavage of proteins 130
6.2 Proposed mechanism for Cu-catalyzed borylation and reduction of dehydroalanine 135
6.3 Cleavage of Histone H3 Dha10 in the presence of $B_2(OH)_4$ 138
6.4 Proposed mechanism for the site-selective diboron-mediated radical protein cleavage 148

7.1 Synthesis of Ac-Gly-Dha-NHBn 179
7.2 *In vitro* SUMOylation with pre-SUMO1 Bal51 287
7.3 Fluorescent labelling of Annexin V Bal316 with FITC 290

List of Tables

2.1 Optimization of Bal formation on peptidic model substrate. 29
2.2 Borylation of small molecule model substrate **1** with different diboron reagents. 30
2.3 Optimization of Bal formation on model protein Histone H3 Dha10. 31
2.5 Protein secondary structure content estimated from CD spectra 42
2.6 Change in protein secondary structure upon borylation 43

3.1 Distribution of Bal substitution 59
3.2 ^{19}F-NMR based K_D determination of different protein and diol combinations at pH 7.0 69
3.3 Reactive PLABP 77
3.4 Solvent-accessible for all proteins subjected to borylation 79

4.1 Formation of the novel species is not concentration dependent 89
4.2 Observed and calculated NMR shifts for possible species of Ac-Bal-NHBn . . 93

6.1 Borylation reaction under nitrogen, oxygen and hydrogen atmosphere 132
6.2 Boron reagents for the site-selective cleavage of proteins 139
6.3 Observed cleavage conversions and cleavage-site accessibility 143

7.1 oxaborolane formation at varying pH 188
7.2 oxaborolane formation at varying concentration 189
7.3 SDS-PAGE analysis of the cleavage of Npβ Dha61 272
7.4 SDS-PAGE analysis of the cleavage of preSUMO1 Dha51 274
7.5 Isotope pattern analysis 279

Abbreviations

3F-Bal	trifluoroboronoalanine
A_{280}	absorbance at 280 nm wavelength
Ac	acetyl
$AMBF_3$	alkylammoniomethyltrifluoroborate
ASA	accessible surface area
atm	standard atmospheres (1 atm = 101,325 Pa)
ATP	adenosine triphosphate
ATR	attenuated total reflection
BAC	boronate affinity chromatography
Bal	boronoalanine
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BDE	bond dissociation energy
BME	β-mercaptoethanol
Bn	benzyl
Bpa	L-boronophenylalanine
Bq	Becquerel
BSA	bovine serum albumin
Bz	benzoyl
cat	catechol, benzene-1,2-diol
CD	circular dichroism

CHES 2-(cyclohexylamino)ethanesulfonic acid

CHO Chinese hamster ovary (cell line)

CPMG Carr-Purcell-Meiboom-Gill pulse sequence

CSP chemical shift perturbation

CV column volume(s)

Da dalton

DAST *N*,*N*-diethylaminosuflur trifluoride

DBHDA 2,5-Dibromohexanediamide

DCM dichloromethane

DFT density-functional theory

Dha dehydroalanine

DIPEA *N*,*N*-diisopropylethylamine

DMAP 4-(dimethylamino)pyridine

DMF *N*,*N*-dimethyl formamide

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

DOPA 3,4-dihydroxyphenylalanine

DSC differential scanning calorimetry

DSF differential scanning fluorimetry

DTT DL-dithiothreitol

EDG electron donating group

EDTA ethylenediaminetetraacetic acid

equiv equivalent

ESI− electrospray ionization, negative mode

ESI+ electrospray ionization, positive mode

Et ethyl

EWG electron withdrawing group

FDGal	fluorodeoxygalactose
FDR	false discovery rate
FDRibOMe	methyl 5-deoxy-5-fluoro-β,D-ribofuranoside
FEP	fluorinated ethylene polypropylene copolymer
FITC	fluorescein isothiocyanate
FLAG	FLAG epitope tag; sequence: DYKDDDDK
FPLC	fast protein liquid chromatography
Gdn	guanidine
GFP	green fluorescent protein
h	hour(s)
HATU	1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium-3-oxide hexafluorophosphate
HIV	human immunodeficiency virus
HMBC	heteronuclear multiple-bond correlation
HPLC	high-performance liquid chromatography
HRMS	high-resolution mass spectrometry
HSQC	heteronuclear single-quantum correlation
IBCF	isobutyl chloroformate
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
iPr	isopropyl
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infrared
K	Kelvin
keto-ABNO	9-azabicyclo[3.3.1]nonan-3-one-*N*-oxyl

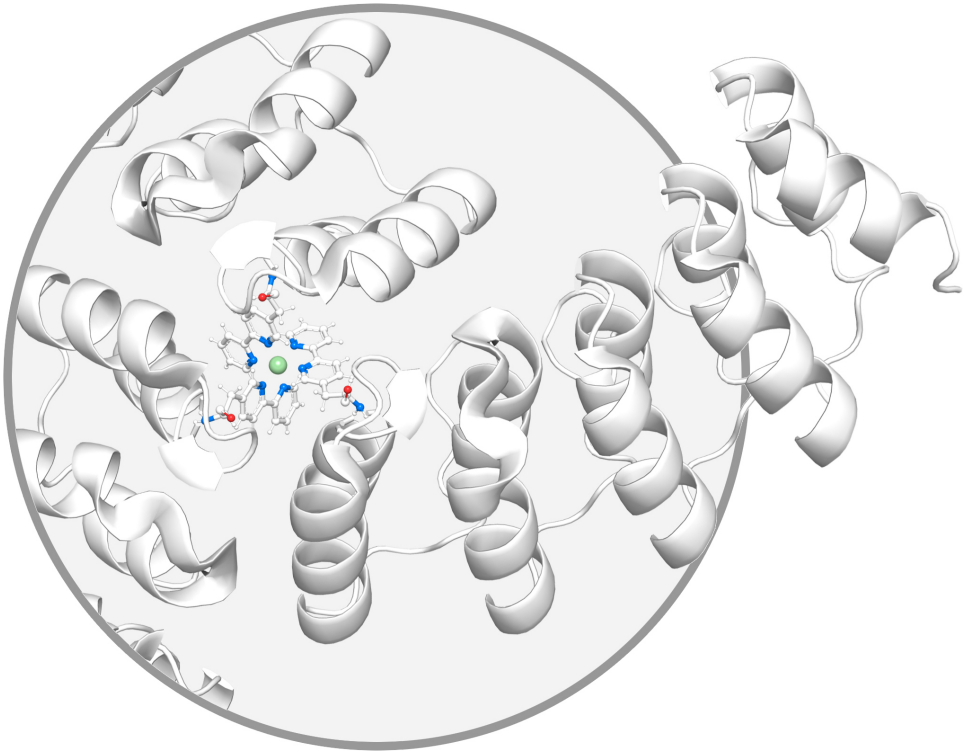
LB	lysogeny broth
LC-MS	liquid chromatography–mass spectrometry
LC-MS/MS	liquid chromatography coupled tandem mass spectrometry
m	*meta*
m.p.	melting point
$\mathbf{M_{calc}}$	calculated molecular weight
$\mathbf{M_{obs}}$	observed molecular weight
MDBPA	methyl 2,5-dibromopentanoate
Me	methyl
MES	2-(*N*-morpholino)ethanesulfonic acid
min	minute(s)
Ms	mesylate, methanesulfonate
MSH	*O*-(mesitylsulfonyl)hydroxylamine
MWCO	molecular weight cut-off
n.d.	not determined
$\mathbf{NaP_i}$	sodium phosphate
NBT	nitro blue tetrazolium
NHS	*N*-hydroxysuccinimide
NMM	*N*-methylmorpholine
NMO	*N*-methylmorpholine *N*-oxide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
o	*ortho*
OD	optical density
$\mathbf{OD_{600}}$	optical density measured at 600 nm wavelength
p	*para*

PBS	phosphate-buffered saline
PCS	pseudocontact shifts
PDB	Protein Data Bank
PEG	polyethylene glycol
PET	positron emission tomography
Ph	phenyl
pin	pinacol, 2,3-dimethylbutane-2,3-diol
PLABP	protein Lewis acid·base pair (probing)
PRE	paramagnetic relaxation enhancement
PTFE	polytetrafluoroethylene
PTM	post-translational modification
Py	pyridine
Q-TOF	quadrupole time-of-flight
rASA	relative solvent-accessible surface area
RDC	residual dipolar coupling
rfx	reflux
RG-II	rhamnogalacturonan II
ROS	reactive oxygen species
rpm	revolutions per minute
rt	room temperature
s	second(s)
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
Sep	phosphoserine
SET	single electron transfer
SNAC	sequence-specific nickel-assisted cleavage
SUMO	small Ubiquitin-like modifier

TB	terrific broth
TBAF	tetra-*n*-butylammonium fluoride
TBS-T	TRIS-buffered saline supplemented with Tween 20
TEMPOL	4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl
TEV	tobacco etch virus
Tf	trifluoromethanesulfonyl
THF	tetrahydrofuran
TLC	thin-layer chromatography
TMS	tetramethylsilane
TRIS	2-amino-2-(hydroxymethyl)propane-1,3-diol
tRNA	transfer ribonucleic acid
Ts	tosyl, toluenesulfonyl
UV	ultraviolet
vs.	versus
WT	wild-type

Chapter 1

Introduction



Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself.

— *Francis Crick*, What Mad Pursuit (1988)

Chapter 1

Introduction

Proteins have to fulfil a multitude of functions, including providing structure, catalysing reactions (enzymes), regulating biological processes, signalling, transporting materials across membranes and even converting chemical energy into mechanical work allowing for motion.[1] Understanding proteins on a molecular level will thus greatly expand our understanding of life. Detailed chemical knowledge of biological processes could allow for their modulation or even *de novo* engineering, advancing both medicine and technology.

Proteins also play an increasingly important role in the treatment of diseases. In 2019, eight out of the ten top selling drugs globally were proteins, and about one third of the drugs newly approved by the US Food and Drug Administration (FDA) were biologics.[2,3] While this highlights the great potential of proteins as therapeutics, several challenges have to be overcome, including limited oral bioavailability, rapid clearance and low stability, immunogenicity and limited cellular uptake.[4,5]

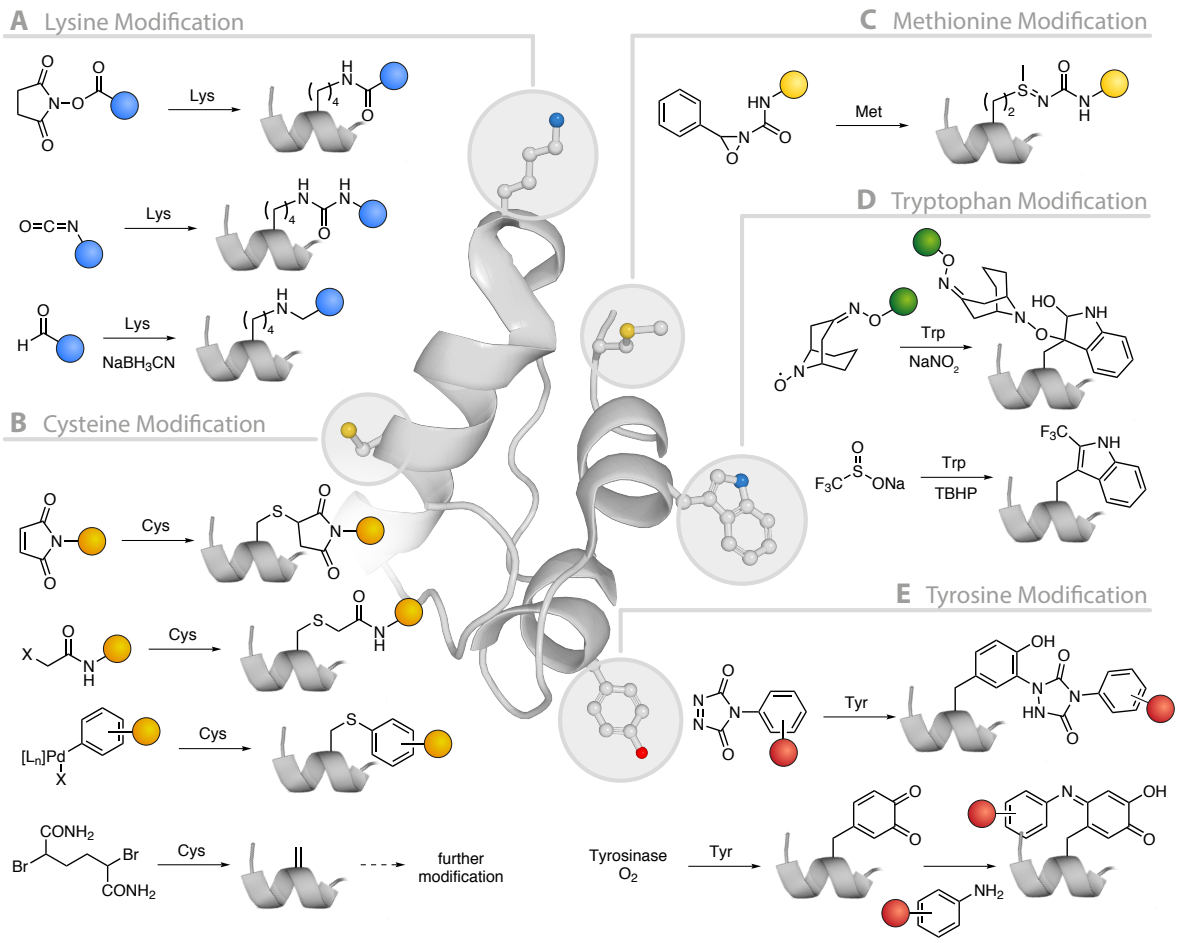
Chemical protein modification, such as PEGylation or protein conjugation, can help to address these shortcomings.[4] Furthermore, precise and site-selective alteration of proteins provide a powerful tool to investigate and understand biology on a molecular level.[6]

1.1 Chemical Protein Modification

Chemical protein modification has advanced significantly since the field was established about 100 years ago. Early protein modification protocols, such as the reaction with diazonium salts (1904) or halogenation and nitration (1930), were conducted under harsh, denaturing conditions.[7,8] Soon, the need for mild and non-denaturing protocols required for the retention of biological function was identified.[9] Contemporary approaches additionally aim to target a specific residue or site, leading to precisely defined reaction products.

1.1.1 Contemporary Protein Modification Strategies: An Overview

Modern protein modification strategies usually rely on residue-specific modification of endogenous amino acids with reactive side chains (**Scheme 1.1**).



Scheme 1.1: Methods for residue-specific protein modification. Current protocols typically lead to non-minimal constructs and do not allow for synthesis of minimal motifs.

Most commonly, lysine and cysteine side chains are targeted.[10] Typical strategies for lysine modification include reaction with *N*-hydroxysuccinimide esters (NHS esters), urea or thiourea formation with isocyanates or isothiocyanates and reductive amination with aldehydes and sodium cyanoborohydride (**Scheme 1.1 A**).[10] Routine protocols for cysteine conjugation include thia-Michael addition to maleimides as well as alkylation with α-halocarbonyls.[10] More recently, protocols for the *S*-arylation of cysteine residues using palladium(II) oxidative addition complexes were reported (**Scheme 1.1 B**).[11,12] In addition to targeting of nucleophilic lysine or cysteine residues, further strategies to target methionine, tryptophan or tyrosine residues have also been developed. Methionine was found to react with oxaziridines, yielding the corresponding sulfimide (**Scheme 1.1 C**). The oxaziridine hereby acts as a *N*-transfer oxidant.[13] Tryptophan was found to selectively react with derivatives of the persistent radical 9-azabicyclo[3.3.1]nonane *N*-oxyl, allowing for bioconjugation.[14] Furthermore, a procedure for the radical trifluoromethylation of tryptophan residues was reported (**Scheme 1.1 D**).[15] Examples of tyrosine modification (**Scheme 1.1 E**) include an ene-type reaction with azomaleimides[16] and the enzymatic oxidation to *o*-quinones followed by the coupling to aniline derivatives.[17]

Most of these methodologies share one shortcoming: they lead to constructs considerably larger than the unmodified natural amino acid they were targeting, which could influence protein structure and function. While these methods keep the natural scaffold of the targeted amino acid mostly intact, they do not allow for the installation of small and benign, unnatural amino acids within native proteins. Furthermore, site-selectivity can be hard to achieve if multiple accessible residues are present. Thus, while there is a selection of procedures available for the residue-specific bioconjugation of proteins, methods for the site-selective introduction of minimally sized, thus non-disturbing, motifs remain scarce.

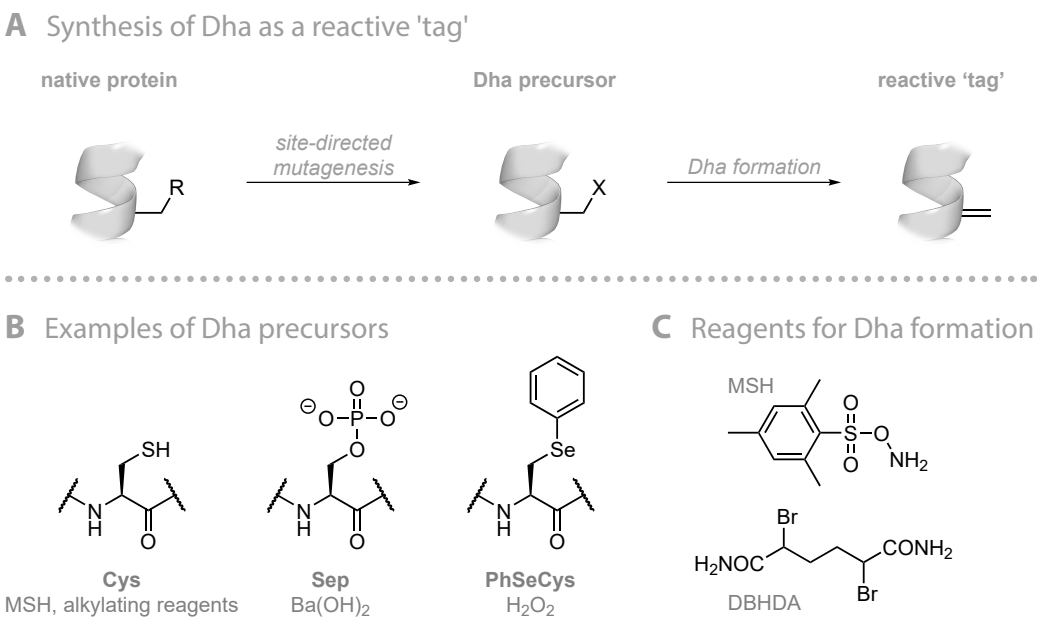
1.1.2 'Tag-and-modify' Approach: Dehydroalanine (Dha) Chemistry

A possible reactive 'tag' for the introduction of minimal functional motifs into proteins is dehydroalanine (Dha). Dha is found naturally in peptides and proteins, where it can be formed through phosphoserine or cysteine elimination.[18,19] Alternatively, a Dha precursor can be

introduced into a native protein, for example through site-directed mutagenesis. Subsequent Dha formation gives access to the reactive 'tag', which allows for further derivatization (**Scheme 1.2 A**).[20]

Dehydroalanine (Dha) formation

Several possible precursors for synthetic Dha formation have been described in literature, including cysteine (Cys), phosphoserine (Sep) and phenylseleoncysteine (PhSeCys) (**Scheme 1.2 B**).[21–23] While formation of Dha from Sep or PhSeCys requires relatively harsh conditions (the use of base or hydrogen peroxide respectively), a variety of different reagents is available for generating Dha from Cys under mild conditions.[21]

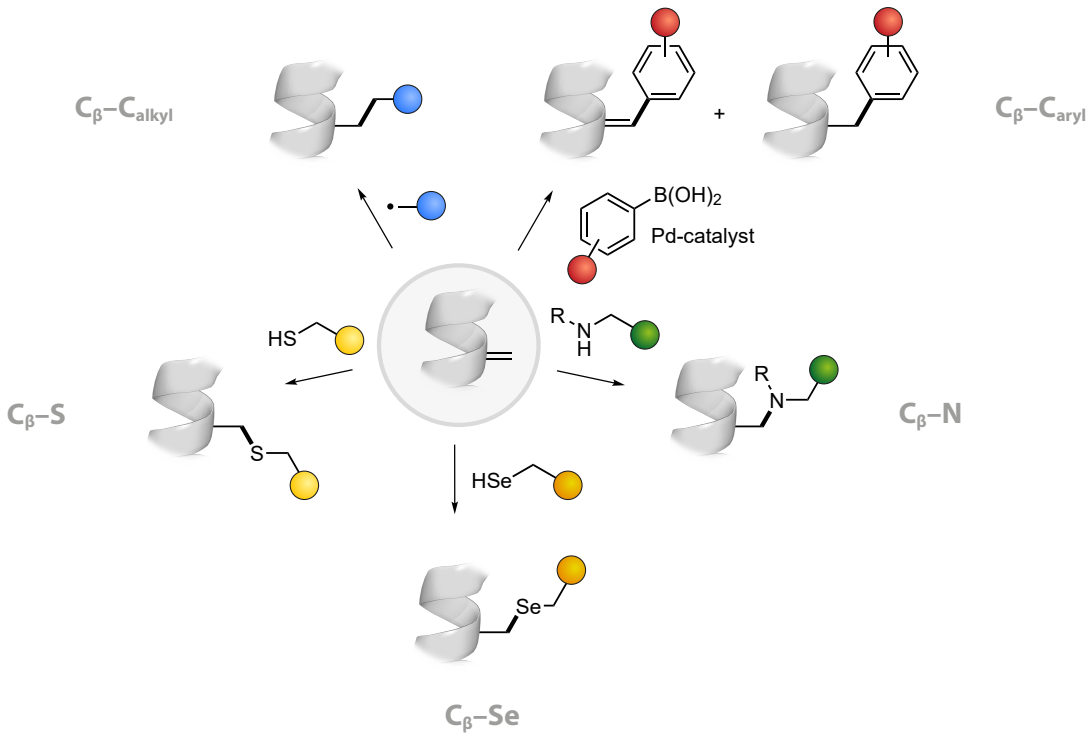


Scheme 1.2: Dehydroalanine (Dha) as a reactive 'tag'. (**A**) Dha can be introduced into native proteins using a genetically incorporated precursor. (**B**) Different precursors were employed for Dha formation. (**C**) Cysteine, possibly the most versatile Dha precursor, can be converted to Dha using various reagents including MSH and DBHDA.

Cysteine has a relatively low natural abundance (approximately 1%)[24] compared to other amino acids, and can readily be introduced into native proteins via site-directed mutagenesis. Furthermore, a wide selection of possible reagents for Dha formation from Cys (see **Scheme 1.2 C** for examples) are available making it an ideal Dha precursor.[21]

Direct C_β–C or C_β–Heteroatom Bond Formation

Dha has been shown to provide a versatile reactive 'tag' for the introduction of a multitude of minimally sized motifs via direct C_β–C or C_β–heteroatom bond formation (**Scheme 1.3**).[25] Michael addition to Dha was shown to conveniently allow for C_β–N, C_β–Se, C_β–S and C_β–P bond formation using *N*-nucleophiles (amines, hydroxylamines, hydrazines or *N*-heterocycles), thiols, allyl selenolate or phosphines respectively.[26–31] Alkylation forming a $C(sp_3)$—$C(sp_3)$ bond was achieved through radical addition to Dha. This methodology allowed for the site-selective introduction of a wide range of functional groups and the synthesis of proteins carrying post-translational modifications (such as lysine methylation and acetylation as well as glycosylation or phosphorylation mimics) and isotopically labelled amino acids.[22,32,33]



Scheme 1.3: Dehydroalanine (Dha) allows for formation of C_β–C_{alkyl}, C_β–C_{aryl}, C_β–N, C_β–Se and C_β–S bonds.

Reaction of Dha with arylboronic acids under palladium catalysis (using $Pd(EDTA)(OAc)_2$ as the catalyst) yielded a mixture of the Heck and reductive Heck product by C_β–C_{aryl} bond formation on peptides and a single protein example.[34] While Dha provides an extremely

versatile handle for protein modification, a few challenges have not yet been overcome: additions to Dha are typically not stereoselective, radical additions require an inert atmosphere and the formation of C_β–heteroatom bonds for elements other than found in the previously described nucleophiles are still lacking.[22,25,32,33]

1.2 Proximity Enabled Covalent Trapping

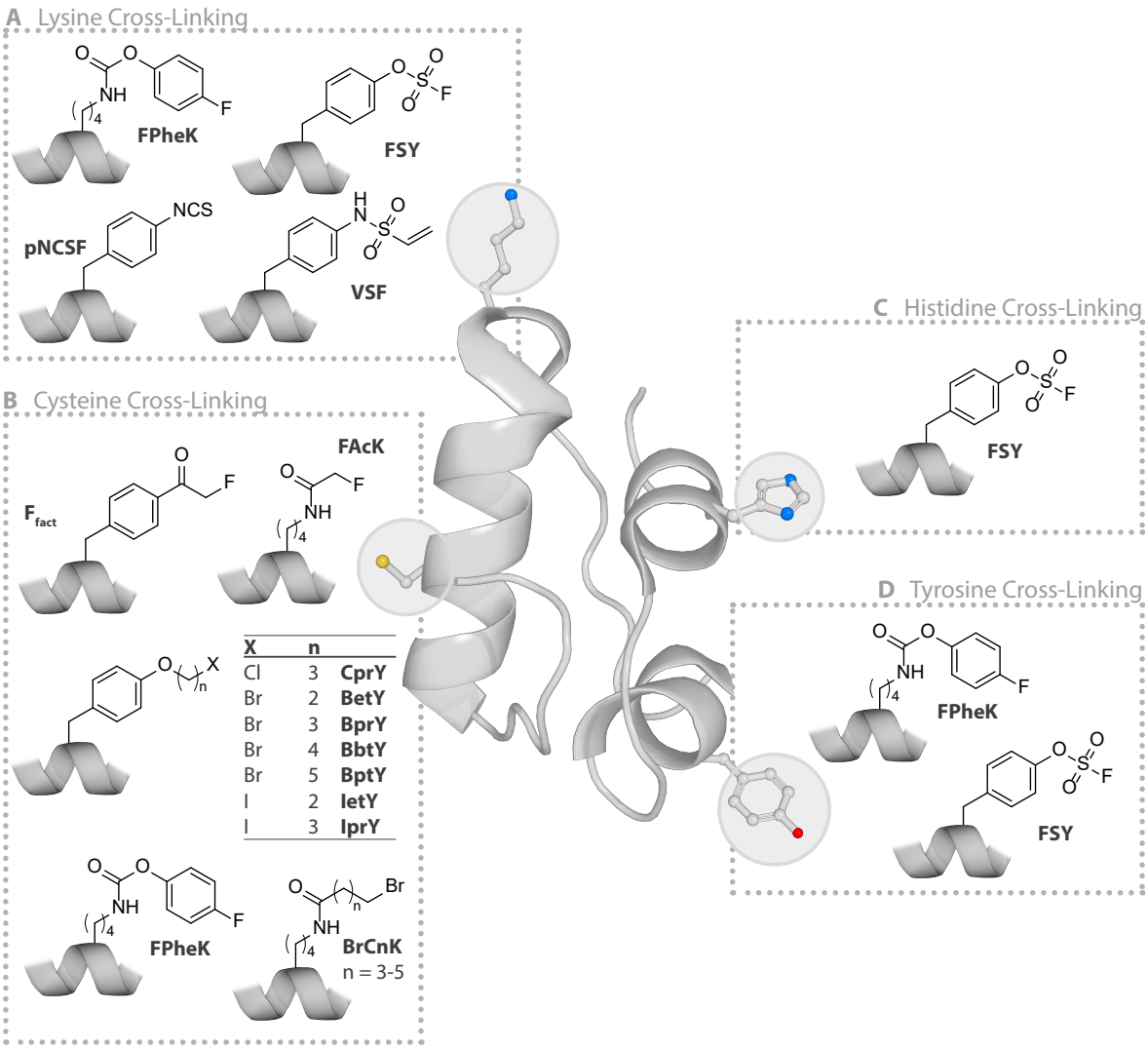


Figure 1.1: Reactive motifs genetically incorporated into proteins for residue-specific, proximity enabled covalent cross linking and their respective canonical amino acid target, including lysine (**A**), cysteine (**B**), histidine (**C**) and tyrosine (**D**).

Proximity enabled covalent trapping of proteins leading to stable cross-linking allows for detection and conservation of transient encounters between proteins with spatial precision

that would otherwise be difficult to achieve (**Figure 1.1**).[35,36] The first examples of proximity enabled cross-linking utilized the incorporation of non-canonical, photoreactive amino acids (either through a pre-charged tRNA or through amber codon suppression) into proteins followed by UV-irradiation to generate the reactive species.[37–39] Contemporary approaches usually rely on the biosynthetic introduction of unnatural amino acids bearing reactive side-chains. The choice of the reactive moiety then enables proximity dependent targeting of selected nucleophilic amino acid residues, while preventing off-target reactions.[40–47] Different reactive moieties were shown to provide suitable reactivity, achieving selectivity while remaining biocompatible.

1.2.1 Cys-reactive Artificial Covalent Linkers

Genetically introduced weak electrophile *p*-2'-fluoroacetylphenylalanine (F_{fact}) allowed for selective, proximity enabled cysteine binding (**Figure 1.1 B**). Due to fluoride being a relatively poor leaving group, minimal reactivity was observed in the presence of low cysteine concentrations. In contrast, higher effective concentrations, such as when F_{fact} and Cys are brought in close proximity, readily led to nucleophilic substitution. Consequently, inter- and intramolecular covalent cross-linking was achieved in mammalian cell culture and *in vitro*.[40] Genetically introduced tyrosine haloalkylethers (CprY, BetY, BprY, BbtY, BptY, IetY, IprY) were also shown to allow for Cys-specific cross-linking. Binding can be optimized by the choice of halide leaving group as well as alkyl chain length. Additionally, intramolecular cross-linking using an appropriate linker length was reported to significantly increase thermal stability (ΔT_M = 13.9 K) as seen in an affibody.[41] The same approaches were implemented using halogenated lysine derivatives. N^{ε}-fluoroacetyllysine (FAcK) was shown to enable inter- and intramolecular Cys-selective cross-linking *in vitro*.[45] Introduction of lysine bearing a bromoalkyl moiety allowed for covalent stabilization of weak protein complexes in cell culture and even enabled crystallization and structure determination.[42]

1.2.2 Lys, His and Tyr Targeting Covalent Linkers

Besides Cys-selective artificial covalent binders, a limited number of artificial covalent binders targeting other nucleophilic amino acid side chains have been reported. Genetically introduced

arylcarbamate derivatives of lysine (PheK and FPheK) showed proximity enabled reactivity towards lysine, cysteine and tyrosine side chains *in vitro* and in cell culture (**Figure 1.1 A** and **D**).[46] A fluorosulfate derivative of tyrosine (FSY) was found to selectively form cross-links with lysine, histidine and tyrosine side chains in an *E. coli* expression system, while also being compatible with mammalian cell culture (**Figure 1.1 A,C** and **D**).[44] For selective lysine targeting, other genetically encoded unnatural amino acids have been developed: an isothiocyanate derivative of phenylalanine (pNCSF) as well as aza-Michael acceptors N^{ε}-acryloyl-(*S*)-lysine (AcrK), *p*-acrylamido-(*S*)-phenylalanine (AcrF), and *p*-vinylsulfonamido-(*S*)-phenyl-alanine (VSF) were found to display the desired selectivity targeting proximal lysine side chains (**Figure 1.1 A**).[43,47]

1.2.3 Discrimination between *De Novo* Binding and Covalent Trapping

The examples discussed above share one common property: they lead to stable, covalently linked constructs, which do not allow for the reverse reaction to take place. Covalent cross-linking is driven by chemical reactivity and proximity-induced, local concentration effects. In contrast, *de novo* binding is defined here as an (at least theoretically) fully reversible equilibrium process. The presence of an equilibrium with association and dissociation events should allow equilibration towards an energetic minimum, conferring superior selectivity. Indeed, reversible inhibition has been linked with an increase in target selectivity and drugs operating with irreversible binding modes can have significant off-target effects.[48,49]

1.3 Engineered *De Novo* Binding Function

1.3.1 Introduction

Despite the limited number of canonical amino acids that proteins naturally comprise, nature has found ways to design proteins that are capable of binding to a multitude of motifs with high affinity. During evolution of these specific binders, nature managed to accommodate for an extensive structural diversity of the possible ligands and substrates: from a plethora of small molecules, including small inorganic compounds, such as metal ions or phosphate, up to large biomacromolecules – proteins and nucleic acids.[50–53] The ability to expand

binding towards novel targets and programme those interactions at will would open up endless possibilities in the fields of biomedicine and nanotechnology.[54] However, designing *de novo* binding functionality, while exclusively making use of the canonical amino acids, requires very precise prediction of those interactions, which is a major hurdle in *in silico* protein design.[55] Introduction of artificial binding sites, in contrast, for example through incorporation of non-canonical amino acids, could accommodate for unprecedented *de novo* binding by utilizing chemically pre-programmed, predictable binding interactions contained within that specific amino acid. Such biocompatible motifs would allow for highly specific and precisely programmed target engagement with minimal undesired background reactivity even in a complex biological matrix.

1.3.2 Metal Binding

In natural proteins, metal binding is typically achieved through imidazole groups (His), thiol or thioether (Cys and Met), hydroxyl moieties (Tyr, Ser or Thr) or carboxylates (Glu or Asp).[50] Artificial metal binding sites, typically polyhistidine tags for nickel(II) binding, have been used extensively over the last 40 years and are a standard methodology in protein purification nowadays.[56] While these binding sites exploit the canonical amino acid histidine, incorporation of unnatural amino acids tailored towards binding various metals has recently gained attention. The introduction of these unnatural amino acids allows for the synthesis of novel metalloenzymes, precise engineering of protein structure, and the design of metal-detecting biosensors.

Synthesis of *De Novo* Metalloenzymes

The incorporation of non-canonical amino acids, such as L-bipyridylalanine (BpyAla), was shown to give rise to predictable, novel metal binding sites.[57] These artificial metal binding sites were especially exploited for the design and synthesis of *de novo* metalloenzymes. A wide variety of methods were developed for the introduction of various binding motifs that facilitate the incorporation of different metals and therefore can enable novel function. A detailed review on this topic was published by Lewis and thus will not be discussed here.[58]

Design and Engineering of Protein Structure

Genetic incorporation of the metal binding amino acid BpyAla was also used to drive protein self-assembly. Baker and co-workers used computational design to create a homotrimeric protein carrying three BpyAla residues and an octahedral iron(III) centre at the interface between the monomers.[59] Yang *et al.* found that the introduction of six BpyAla residues into a homohexameric protein complex enabled controllable self assembly, creating distinct 1D and 2D structures in the presence of nickel(II). The obtained structures were mainly governed by the mutation site, metal concentration and reaction conditions, while no further protein engineering was necessary to shape the protein-protein interface for self-assembly.[60] Furthermore, artificial metal binding sites were found to enable structural stabilization of protein motifs.[61] Recently, Almhjell and Mills published a comprehensive review on the structural implications and use of metal-chelating non-canonical amino acids for the design of *de novo* metalloproteins.[62]

Artificial Metal-Detecting Biosensors

Besides their use in artificial metalloenzymes and in the engineering of protein structure, synthetic metal binding sites have gained some attention for the development of artificial metal-detecting biosensors. Strategic introduction of metal binding sites into proteins should allow for a change in their biophysical properties upon binding of the metal analyte, which – if it can be detected and quantified – would give rise to a protein-based metal sensor (**Figure 1.2**). The incorporation of the 8-hydroxyquinoline derivative of alanine (2-Amino-3-(8-hydroxyquinolin-3-yl)propanoic acid, HQ-Ala, **Figure 1.2 A**) into the Z-domain of staphylococcal protein A by genetic code expansion gave rise to a synthetic protein that exhibited concentration dependent fluorescence in the presence of zinc(II).[63]

When HQ-Ala was introduced into a super folder green fluorescent protein variant (cpsfGFP), an enhancement of fluorescence was detected. Confocal fluorescence microscopy of *E. coli* expressing cpsfGFP HQ-Ala66 allowed for *in vivo* Zn^{2+} detection. Another mutant of this protein, sfGFP HQ-Ala66 enabled the detection of copper(II) by the measurement of fluorescence quenching. This mutant was found to have sub-femptomolar Cu^{2+} binding affinity.[64] Likewise, genetic code expansion allowed for the incorporation of L-DOPA into GFP (**Figure**

1.2 B). This GFP mutant also exhibited Cu^{2+}-dependent fluorescence quenching. Hence, fabrication of a Cu^{2+}-biosensor was accomplished using soft lithography on an amine-coated glass surface.[65] In addition, incorporation of L-DOPA into GFP was found to render a GFP mutant that displayed a nearly 5-fold fluorescence increase in the presence of aluminium(III) allowing for the detection of Al^{3+} *in vitro* and *in vivo*.[66]

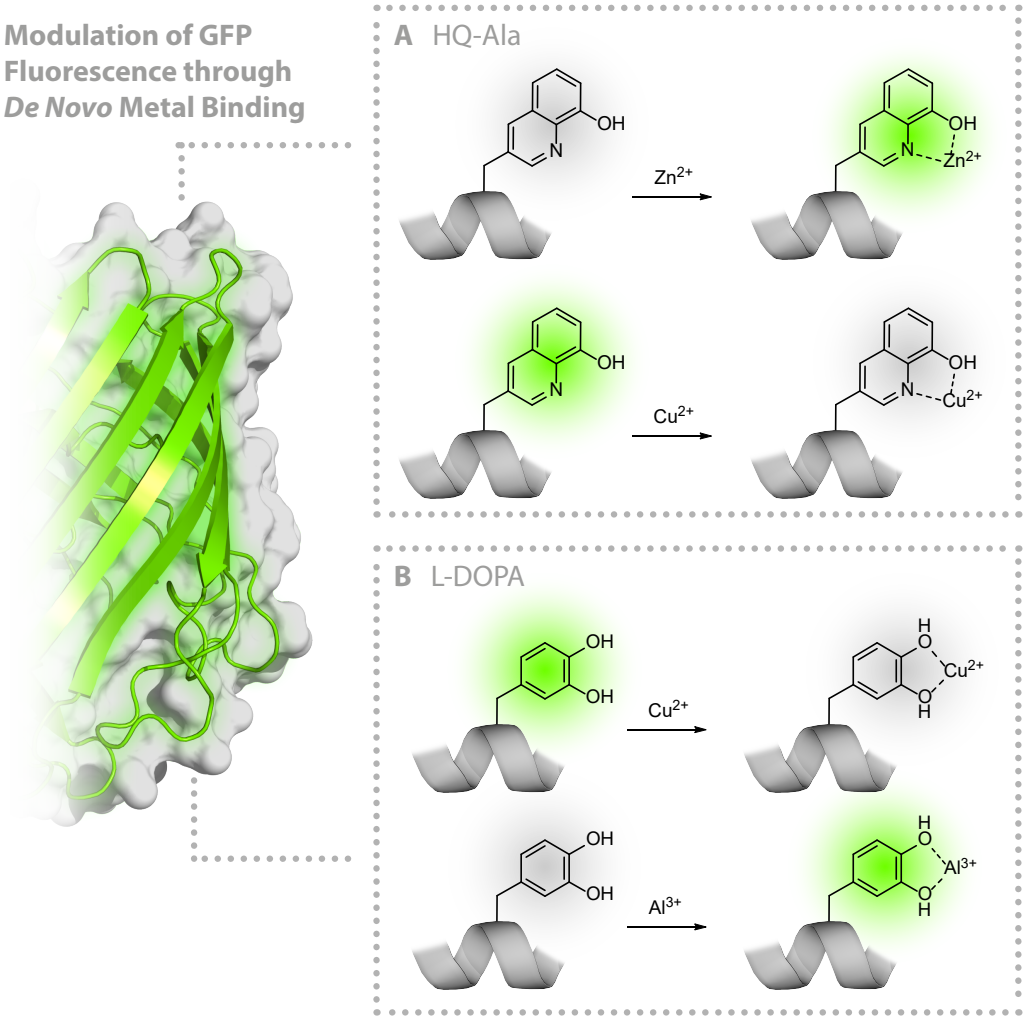


Figure 1.2: Incorporation of metal binding unnatural amino acids HQ-Ala (**A**) and L-DOPA (**B**) allowed for the engineering of metal-detecting biosensors through metal mediated fluorescence modulation. Fluorescence enhancement or quenching enabled metal detection *in vitro* and *in vivo*.

Aside from fluorescence-based sensors, BpyAla-facilitated cobalt(II) binding was used as a molecular NMR probe. Paramagnetic BpyAla-Co^{2+} complexes enabled NMR protein structure analysis via the resulting pseudocontact shifts (PCS).[67]

1.3.3 Reversible Binding

In nature, most of the non-covalent, reversible interactions between proteins and their ligands are based on hydrophobic, hydrogen bonding or π-stacking interactions.[68] Introduction of *de novo* reversible binding sites through non-canonical amino acids could vastly extend the portfolio of possible interaction modes. Hence, precise and programmed mimicry or even generation of completely unprecedented interactions could be engineered in a controllable fashion. The transient nature of these reversible interactions could allow for the interaction with biological systems in a more benign fashion than achievable with covalent binders, which can possibly interrupt native structure and function.

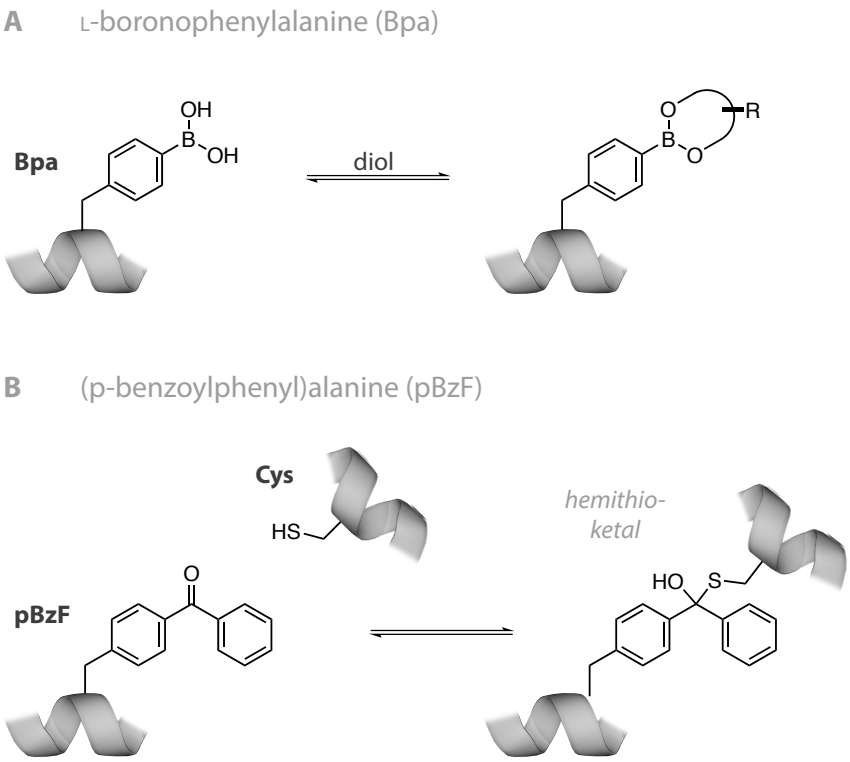
Reversible Covalent Binding

In 2008, Schultz and co-workers demonstrated the first introduction of a boronic acid into proteins using genetic code expansion.[69] Boronic acids are known for their ability to reversibly form covalent complexes with Lewis bases and have attracted great attention especially for saccharide sensing (**Scheme 1.4 A**).[70] The incorporation of L-boronophenylalanine (Bpa) into the T4 lysozyme allowed for binding to a glucamine conjugated fluorescent dye, a binding interaction that is absent in the wild-type protein.[69] The diol binding properties of Bpa were further exploited for the design of an artificial lectin. An anticalin, a small synthetic, cup-shaped β-barrel protein derived from the lipocalin scaffold, was equipped with a Bpa residue in its binding region. This led to a novel mutant capable of binding saccharides, such as sorbitol, galactose, mannose and glucose, with millimolar binding affinities and the catechol derivative 4-nitrocatechol with micromolar binding affinity.[71]

Furthermore, the ability of protein attached boronic acids to differentiate between different diols and polyols based on their binding behaviour was exploited to identify the individual components of a mixture containing several saccharides and even allowed for the recognition of the binding cyclic isomer. The construction of this saccharide sensor was achieved by chemically attaching a phenylboronic acid inside the protein pore of α-hemolysin (α-HL). Upon the application of a current, binding events can be monitored and characterized by the resulting reversible current blockades and their lifetimes. This allowed for identification of the binding cyclic isomers of D-glucose and D-fructose as well as for the differentiation between

D-glucose, D-fructose and D-maltose.[72]

The concept of saccharide binding using phenylboronic acid derivatives was further employed in the engineering of glucose responsive insulin. A phenylboronic acid tethered to insulin via an aliphatic, albumin binding linker gave rise to mutants that showed glucose-responsive behaviour in an *in vivo* model.[73]



Scheme 1.4: Genetically incorporated unnatural amino acids for *de novo* reversible covalent binding: Incorporation of L-boronophenylalanine (Bpa) allows for binding of natural and unnatural diols (**A**). Introduction of (*p*-benzoylphenyl)alanine enables reversible covalent cysteine engagement through hemithioketal formation (**B**).

Additionally, protein conjugation was achieved using a tailored boronic acid and diol pair. A phenylboronic acid was attached to a protein via a Cys-reactive maleimide. Tuning of the substitution pattern of the aromatic moiety and selection of a suitable diol then allowed for the formation of a stable boronate and enabled this methodology to be used for protein ligation in aqueous buffer.[74]

While all of the examples mentioned above rely on the reversible and intermolecular covalent formation of boronates between synthetically introduced boronic acids and external diols, the incorporation of the keto amino acid (*p*-benzoylphenyl)alanine (pBzF, **Scheme 1.4 B**) was

found to enable reversible binding of proximal cysteine side chains via hemithioketal formation. The introduction of pBzF into *E. coli* homoserine *O*-succinyltransferase (metA) at position 21 consequently led to a significant thermal stabilization of the enzyme (ΔT_M = 22 K), which is in agreement with the proposed covalent intramolecular binding.[75]

Host-Guest Interactions

Bayley and co-workers have pioneered the use of cyclodextrins (CD) as "molecular adapters" to engineer *de novo* non-covalent binding into the lumen of the multimeric protein pore α-hemolysin. This allowed for the stochastic sensing of small molecule binding events through single channel recording.[76] The assembly of two different cyclodextrin derivatives within the same pore enabled the construction of a nanocavity, in which single molecules could be trapped.[77] In agreement with the modular approach of using "molecular adapters", chemical modification of the adapter was shown to modulate small molecule binding behaviour.[78] Meanwhile, site-directed mutagenesis of the α-hemolysin host, could be used to modify the binding mode of the CD "molecular adapter" within the protein pore.[79] The harnessing of small molecule binding properties of cyclodextrins (and the possibility to modify these by chemical derivatization of the cyclodextrin) on one hand, and the possibility to modify cyclodextrin binding to the protein through protein engineering on the other hand, could allow for further precisely engineered, yet modular and highly tuneable, *de novo* binding.

1.4 Conclusions and Outlook

Diversification of protein function through the introduction of *de novo* binding mediated by artificial binding sites has gained increasing attention within the last decade. The wide toolkit of binding functionality available through (mostly) small molecule functional groups is increasingly being introduced into protein scaffolds enabling unprecedented function. Proximity enabled covalent binding has become an important tool to precisely capture transient protein-protein interactions. The introduction of artificial metal binding sites has allowed for the construction of *de novo* metalloenzymes, gave new opportunities in the design and engineering of protein structure and paved the way for metal-detecting protein-based biosensors for *in vitro* and *in vivo* applications. Engineered reversible binding was employed for additional

thermal protein stabilization and synthesis of artificial lectins. While the introduction of novel function into native proteins was previously fairly limited and mostly governed by non-selective bioconjugation (such as the case for bioconjugation of fluorescent reporter molecules),[80] novel approaches based on biosynthetic methods such as genetic code expansion or bespoke chemistry could enable the synthesis of truly bis- or multi-functional proteins. Recent advances in protein modification chemistry, such as the possibility to form minimally sized constructs through Dha chemistry, could allow for chemical modification with unprecedented precision. Such bespoke chemistry could lay the foundation for the tailored fusion of natural protein function with precise artificial binding sites. This would enable unprecedented function and is yet to be explored in depth.

This thesis aims to contribute towards the development of *de novo* protein function by the means of chemical protein modification. The introduction of a minimally sized boronic acid via the borylation of dehydroalanine residues is envisioned to create proteins with artificial function. The scope of this new motif will be explored within this work.

References

[1] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, 6th ed., (Eds.: J. Wilson, T. Hunt), Garland Science, **2017**, pp. 109–172.

[2] L. Urquhart, *Nat. Rev. Drug Discov.* **2020**, *19*, 228–228.

[3] C. Morrison, *Nat. Biotechnol.* **2020**, *38*, 126–131.

[4] A. Szlachcic, M. Zakrzewska, J. Otlewski, *Biotechnol. Adv.* **2011**, *29*, 436–441.

[5] P. J. Carter, *Exp. Cell Res.* **2011**, *317*, 1261–1269.

[6] N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103–113.

[7] H. Pauly, *Hoppe-Seyler´s Zeitschrift für Physiol. Chemie* **1904**, *42*, 508–518.

[8] A. Wormall, *J. Exp. Med.* **1930**, *51*, 295–317.

[9] R. M. Herriott, *Adv. Protein Chem.* **1947**, *3*, 169–225.

[10] O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174–2195.

[11] E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, *Nature* **2015**, *526*, 687–691.

[12] A. J. Rojas, B. L. Pentelute, S. L. Buchwald, *Org. Lett.* **2017**, *19*, 4263–4266.

[13] S. Lin, X. Yang, S. Jia, A. M. Weeks, M. Hornsby, P. S. Lee, R. V. Nichiporuk, A. T. Iavarone, J. A. Wells, F. D. Toste, C. J. Chang, *Science* **2017**, *355*, 597–602.

[14] Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki, M. Kanai, *J. Am. Chem. Soc.* **2016**, *138*, 10798–10801.

[15] M. Imiołek, G. Karunanithy, W.-L. Ng, A. J. Baldwin, V. Gouverneur, B. G. Davis, *J. Am. Chem. Soc.* **2018**, *140*, 1568–1571.

[16] H. Ban, J. Gavrilyuk, C. F. Barbas, *J. Am. Chem. Soc.* **2010**, *132*, 1523–1525.

[17] A. M. Marmelstein, M. J. Lobba, C. S. Mogilevsky, J. C. Maza, D. D. Brauer, M. B. Francis, *J. Am. Chem. Soc.* **2020**, *142*, 5078–5086.

[18] Z. Wang, B. Lyons, R. J. W. Truscott, K. L. Schey, *Aging Cell* **2014**, *13*, 226–234.

[19] R. Bar-Or, L. T. Rael, D. Bar-Or, *Rapid Commun. Mass Spectrom.* **2008**, *22*, 711–716.

[20] J. M. Chalker, G. J. L. Bernardes, B. G. Davis, *Acc. Chem. Res.* **2011**, *44*, 730–741.

[21] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem. Sci.* **2011**, *2*, 1666.

[22] A. Yang, S. Ha, J. Ahn, R. Kim, S. Kim, Y. Lee, J. Kim, D. Soll, H.-Y. Lee, H.-S. Park, *Science* **2016**, *354*, 623–626.

[23] J. Wang, S. M. Schiller, P. G. Schultz, *Angew. Chem. Int. Ed.* **2007**, *46*, 6849–6851.

[24] S. Itzkovitz, U. Alon, *Genome Res.* **2007**, *17*, 405–412.

[25] J. Dadová, S. R. Galan, B. G. Davis, *Curr. Opin. Chem. Biol.* **2018**, *46*, 71–81.

[26] A. M. Freedy, M. J. Matos, O. Boutureira, F. Corzana, A. Guerreiro, P. Akkapeddi, V. J. Somovilla, T. Rodrigues, K. Nicholls, B. Xie, G. Jiménez-Osés, K. M. Brindle, A. A. Neves, G. J. L. Bernardes, *J. Am. Chem. Soc.* **2017**, *139*, 18365–18375.

[27] J. Dadová, K.-J. Wu, P. G. Isenegger, J. C. Errey, G. J. L. Bernardes, J. M. Chalker, L. Raich, C. Rovira, B. G. Davis, *ACS Cent. Sci.* **2017**, *3*, 1168–1173.

[28] J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed.* **2012**, *51*, 1835–1839.

[29] C. L. Windle, K. J. Simmons, J. R. Ault, C. H. Trinh, A. Nelson, A. R. Pearson, A. Berry, *Proc. Natl. Acad. Sci.* **2017**, *114*, 2610–2615.

[30] Y. A. Lin, O. Boutureira, L. Lercher, B. Bhushan, R. S. Paton, B. G. Davis, *J. Am. Chem. Soc.* **2013**, *135*, 12156–12159.

[31] P.-y. He, H. Chen, H.-G. Hu, J.-j. Hu, Y.-j. Lim, Y.-m. Li, *Chem. Commun.* **2020**, *56*, 12632–12635.

[32] T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallee, A. Phanumartwiwath, O. D. Coleman, M.-L. Thezenas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed, B. G. Davis, *Science* **2016**, *354*, 597.

[33] B. Josephson, C. Fehl, P. G. Isenegger, S. Nadal, T. H. Wright, A. W. J. Poh, B. J. Bower, A. M. Giltrap, L. Chen, C. Batchelor-McAuley, G. Roper, O. Arisa, J. B. I. Sap, A. Kawamura, A. J. Baldwin, S. Mohammed, R. G. Compton, V. Gouverneur, B. G. Davis, *Nature* **2020**, *585*, 530–537.

[34] A. D. de Bruijn, G. Roelfes, *Chem. Eur. J.* **2018**, *24*, 12728–12733.

[35] R. Kluger, A. Alagic, *Bioorg. Chem.* **2004**, *32*, 451–472.

[36] A. Sinz, *Expert Rev. Proteomics* **2014**, *11*, 733–743.

[37] T. V. Kurzchalia, M. Wiedmann, A. S. Girshovich, E. S. Bochkareva, H. Bielka, T. A. Rapoport, *Nature* **1986**, *320*, 634–636.

[38] T. Kanamori, S.-I. Nishikawa, I. Shin, P. G. Schultz, T. Endo, *Proc. Natl. Acad. Sci.* **1997**, *94*, 485–490.

[39] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci.* **2002**, *99*, 11020–11024.

[40] Z. Xiang, H. Ren, Y. S. Hu, I. Coin, J. Wei, H. Cang, L. Wang, *Nat. Methods* **2013**, *10*, 885–888.

[41] Z. Xiang, V. K. Lacey, H. Ren, J. Xu, D. J. Burban, P. A. Jennings, L. Wang, *Angew. Chem. Int. Ed.* **2014**, *53*, 2190–2193.

[42] M. Cigler, T. G. Müller, D. Horn-Ghetko, M.-K. von Wrisberg, M. Fottner, R. S. Goody, A. Itzen, M. P. Müller, K. Lang, *Angew. Chem. Int. Ed.* **2017**, *56*, 15737–15741.

[43] J. L. Furman, M. Kang, S. Choi, Y. Cao, E. D. Wold, S. B. Sun, V. V. Smider, P. G. Schultz, C. H. Kim, *J. Am. Chem. Soc.* **2014**, *136*, 8411–8417.

[44] N. Wang, B. Yang, C. Fu, H. Zhu, F. Zheng, T. Kobayashi, J. Liu, S. Li, C. Ma, P. G. Wang, Q. Wang, L. Wang, *J. Am. Chem. Soc.* **2018**, *140*, 4995–4999.

[45] T. Kobayashi, C. Hoppmann, B. Yang, L. Wang, *J. Am. Chem. Soc.* **2016**, *138*, 14832–14835.

[46] W. Xuan, S. Shao, P. G. Schultz, *Angew. Chem. Int. Ed.* **2017**, *56*, 5096–5100.

[47] W. Xuan, J. Li, X. Luo, P. G. Schultz, *Angew. Chem. Int. Ed.* **2016**, *55*, 10065–10068.

[48] P. Beck, C. Dubiella, M. Groll, *Biol. Chem.* **2012**, *393*, 1101–1120.

[49] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.

[50] J. A. Tainer, V. A. Roberts, E. D. Getzoff, *Curr. Opin. Biotechnol.* **1992**, *3*, 378–387.

[51] M. Brune, J. L. Hunter, S. A. Howell, S. R. Martin, T. L. Hazlett, J. E. T. Corrie, M. R. Webb, *Biochemistry* **1998**, *37*, 10370–10380.

[52] I. M. Nooren, *EMBO J.* **2003**, *22*, 3486–3492.

[53] D. L. Ollis, S. W. White, *Chem. Rev.* **1987**, *87*, 981–995.

[54] P. S. Huang, S. E. Boyken, D. Baker, *Nature* **2016**, *537*, 320–327.

[55] R. J. Pantazes, M. J. Grisewood, C. D. Maranas, *Curr. Opin. Struct. Biol.* **2011**, *21*, 467–472.

[56] H. Block, B. Maertens, A. Spriestersbach, N. Brinker, J. Kubicek, R. Fabis, J. Labahn, F. Schäfer, *Methods Enzymol.* **2009**, *463*, 439–473.

[57] J. Xie, W. Liu, P. G. Schultz, *Angew. Chem. Int. Ed.* **2007**, *46*, 9239–9242.

[58] J. C. Lewis, *Curr. Opin. Chem. Biol.* **2015**, *25*, 27–35.

[59] J. H. Mills, W. Sheffler, M. E. Ener, P. J. Almhjell, G. Oberdorfer, J. H. Pereira, F. Parmeggiani, B. Sankaran, P. H. Zwart, D. Baker, *Proc. Natl. Acad. Sci.* **2016**, *113*, 15012–15017.

[60] M. Yang, W. J. Song, *Nat. Commun.* **2019**, *10*, 5545.

[61] X. Luo, T.-S. A. Wang, Y. Zhang, F. Wang, P. G. Schultz, *Cell Chem. Biol.* **2016**, *23*, 1098–1102.

[62] P. J. Almhjell, J. H. Mills, *Curr. Opin. Struct. Biol.* **2018**, *51*, 170–176.

[63] H. S. Lee, G. Spraggon, P. G. Schultz, F. Wang, *J. Am. Chem. Soc.* **2009**, *131*, 2481–2483.

[64] X. Liu, J. Li, C. Hu, Q. Zhou, W. Zhang, M. Hu, J. Zhou, J. Wang, *Angew. Chem. Int. Ed.* **2013**, *52*, 4805–4809.

[65] N. Ayyadurai, N. Saravanan Prabhu, K. Deepankumar, S.-G. Lee, H.-H. Jeong, C.-S. Lee, H. Yun, *Angew. Chem. Int. Ed.* **2011**, *50*, 6534–6537.

[66] D. Wu, K. Deepankumar, Y. Ping, G. Tang, N. Saravananprabhu, A. Miserez, W. Fang, *Sens. Actuators B Chem.* **2017**, *251*, 326–333.

[67] T. H. D. Nguyen, K. Ozawa, M. Stanton-Cook, R. Barrow, T. Huber, G. Otting, *Angew. Chem. Int. Ed.* **2011**, *50*, 692–694.

[68] R. Ferreira de Freitas, M. Schapira, *MedChemComm* **2017**, *8*, 1970–1981.

[69] E. Brustad, M. L. Bushey, J. W. Lee, D. Groff, W. Liu, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47*, 8220–8223.

[70] X. Wu, Z. Li, X.-X. Chen, J. S. Fossey, T. D. James, Y.-B. Jiang, *Chem. Soc. Rev.* **2013**, *42*, 8032.

[71] S. Edwardraja, A. Eichinger, I. Theobald, C. A. Sommer, A. J. Reichert, A. Skerra, *ACS Synth. Biol.* **2017**, *6*, 2241–2247.

[72] W. J. Ramsay, H. Bayley, *Angew. Chem. Int. Ed.* **2018**, *57*, 2841–2845.

[73] D. H.-C. Chou, M. J. Webber, B. C. Tang, A. B. Lin, L. S. Thapa, D. Deng, J. V. Truong, A. B. Cortinas, R. Langer, D. G. Anderson, *Proc. Natl. Acad. Sci.* **2015**, *112*, 2401–2406.

[74] B. Akgun, D. G. Hall, *Angew. Chem. Int. Ed.* **2016**, *55*, 3909–3913.

[75] J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, *J. Am. Chem. Soc.* **2018**, *140*, 15997–16000.

[76] L.-Q. Gu, O. Braha, S. Conlan, S. Cheley, H. Bayley, *Nature* **1999**, *398*, 686–690.

[77] L.-Q. Gu, S. Cheley, H. Bayley, *Science* **2001**, *291*, 636–640.

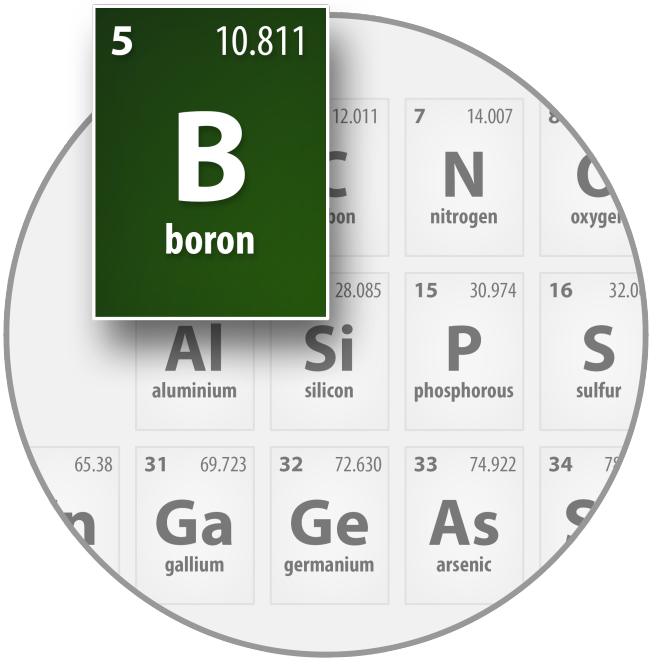
[78] W. W. Li, T. D. Claridge, Q. Li, M. R. Wormald, B. G. Davis, H. Bayley, *J. Am. Chem. Soc.* **2011**, *133*, 1987–2001.

[79] A. Banerjee, E. Mikhailova, S. Cheley, L.-Q. Gu, M. Montoya, Y. Nagaoka, E. Gouaux, H. Bayley, *Proc. Natl. Acad. Sci.* **2010**, *107*, 8165–8170.

[80] N. Stephanopoulos, M. B. Francis, *Nat. Chem. Biol.* **2011**, *7*, 876–884.

Chapter 2

Synthesis of Boronoproteins – Development of a Site-Selective Borylation Protocol



The structure known, but not yet accessible by synthesis, is to the chemist what the unclimbed mountain, the uncharted sea, the untilled field, the unreached planet, are to other men.

— *Robert B. Woodward*, Art and Science in the Synthesis of Organic Compounds: Retrospect and Prospect (1963)

Chapter 2

Synthesis of Boronoproteins – Development of a Site-Selective Borylation Protocol

For the past 60 years, organoboron compounds have played a crucial role in organic synthesis. Many protocols were developed to facilitate the formation of organoboron compounds as well as their use as building blocks in synthetic applications.[1] Likewise, biomacromolecules, in particular proteins, have gained increasing attention in the pharmaceutical industry over the past few decades.[2] Yet, despite their potential utility, the available methodologies for the introduction of organoboron moieties, namely boronic acids, into proteins remain very limited.

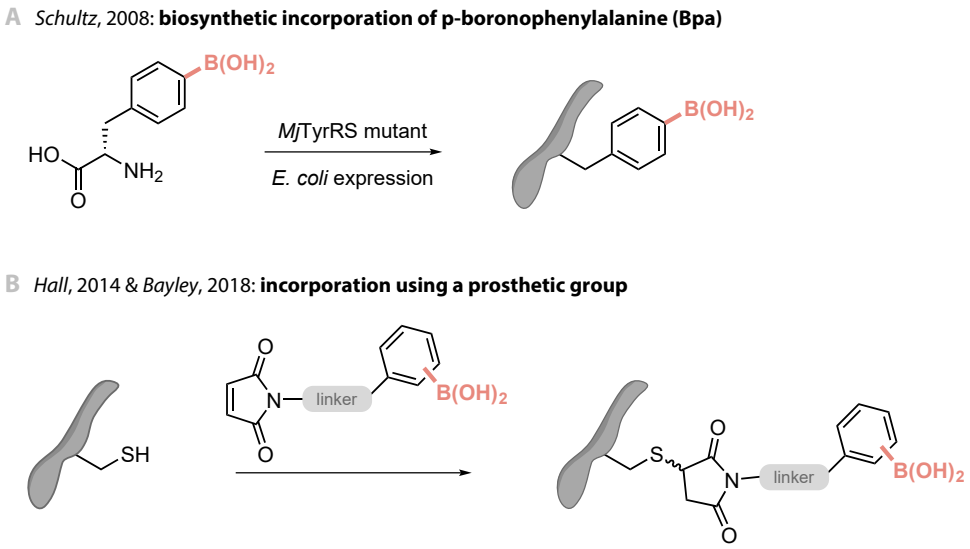
In this chapter, the development of a novel methodology for the chemical introduction of a minimal boronic acid motif into native proteins via direct C_β–B_γ bond formation is described.

2.1 Introduction

2.1.1 Introducing Boron Into Proteins

Multiple methodologies are available for the introduction of novel chemical moieties into proteins. Biosynthetic approaches traditionally rely on residue-specific or site-specific introduction of an unnatural amino acid. Residue-specific incorporation usually utilizes an

auxotrophic bacterial host system, which is starved of a natural amino acid and supplemented with its unnatural derivative, leading to global replacement of the natural amino acid with its unnatural counterpart.[3] This methodology can be combined with co-expression of an engineered aminoacyl-tRNA synthethase to allow for incorporation of a wider range of unnatural amino acids.[4] Site-specific incorporation of an unnatural amino acid, in contrast, enables the introduction of a novel, unnatural amino acid at a single specific site. This methodology requires an engineered aminoacyl-tRNA synthetase/tRNA pair which responds to a nonsense codon, such as the amber stop codon (TAG) and is able to recognize and bind the desired unnatural amino acid.[5] Biosynthetic incorporation can suffer from low efficiency of the engineered aminoacyl-tRNA synthetase, restricted cell permeability, as well as toxicity or limited chemical and metabolic stability of the unnatural amino acid. This typically leads to significantly lower expression yields compared to the wild-type protein.[5,6] The use of biosynthetic methodologies for the incorporation of boronic acids could be further complicated by their potential inhibitory activity. Indeed, several small-molecule boronic acids are known to act as powerful inhibitors of synthetases.[7] Despite those obstacles, biosynthetic incorporation of *p*-boronophenylalanine into proteins was previously reported using an evolved aminoacyl-tRNA synthetase (**Scheme 2.1 A**).[8]

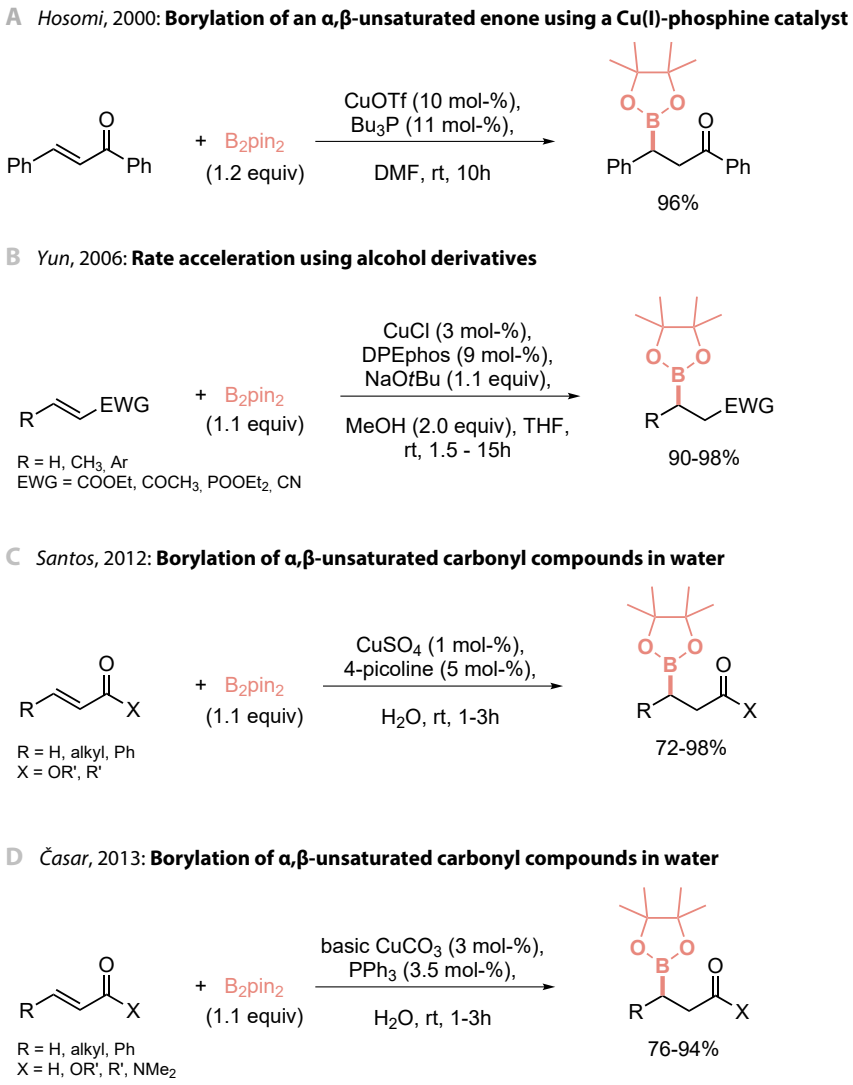


Scheme 2.1: Literature precedence for the incorporation of boronic acids into proteins. (**A**) Biosynthetic incorporation of *p*-boronophenylalanine via amber codon suppression. (**B**) Bioconjugation approach using a Cys reactive maleimide prosthetic group.

Chemical approaches often utilize the bioconjugation of the desired moiety using a reactive linker usually targeting Lys or Cys residues. Indeed, boronic acids have been conjugated to proteins using Cys-reactive maleimide prosthetic groups (**Scheme 2.1 B**).[9,10] While these methodologies allow for convenient introduction of boronic acids, they are not necessarily site-selective and do require the use of linkers leading to non-minimal boronic acid constructs. These prosthetic groups could potentially lead to perturbation of protein structure and hence prevent the precise probing of binding, which a minimal size boronic acid could allow for.

2.1.2 Borylation of α,β-unsaturated Carbonyl Compounds: From Small-Molecule to Protein?

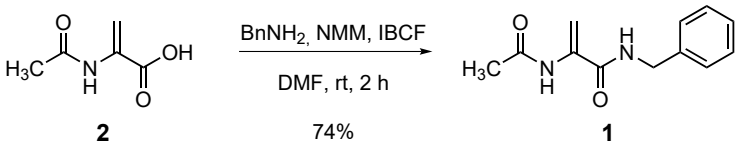
Multiple methodologies for the borylation of α,β-unsaturated carbonyl compounds, often ketones or esters, have been described. While transition-metal free and radical methodologies are known, most published procedures employ transition-metal catalysts.[11,12] Protocols using zinc, nickel, palladium, rhodium and platinum complexes were developed,[13–17] but most reports utilize copper-based systems (**Scheme 2.2**).[18–22] Cu-catalysed borylation of α,β-unsaturated carbonyl compounds usually takes place in organic solvent, often at elevated temperatures. Addition of an alcohol additive, such as methanol, was found to be beneficial and lead to significant rate acceleration (**Scheme 2.2 B**).[23] Site-selective chemical protein modification protocols require bespoke reactivity at a specific residue of interest. A "tag-and-modify" strategy is able to meet this requirement.[24] Dehydroalanine (Dha), which is readily chemically accessible (for example from cysteine residues), provides a chemical handle of unique reactivity.[25] Being, in essence, an α,β-unsaturated amide, a Dha residue could serve as a valuable handle for chemo-, regio- and site-selective protein borylation, if the reaction can be conducted under benign and aqueous conditions. Indeed, copper-catalysed β-borylation of α,β-unsaturated carbonyl compounds in aqueous media was previously described: several protocols were reported that allow for efficient borylation with anti-markovnikov selectivity in water. These procedures employ a copper(II) source, a diboron reagent and a phosphine or pyridine based ligand in solely aqueous media (**Scheme 2.2 C,D**).[17,26–29]



Scheme 2.2: Reported procedures for the borylation of α,β-unsaturated carbonyl compounds: (**A**) First procedure reported by Hosomi and co-workers in 2000 for the borylation of an α,β-enone using B_2pin_2 and a Cu(I)-phosphine complex. (**B**) Yun and co-workers disclosed a protocol in 2006 using MeOH as an additive to achieve rate acceleration, allowing for a broadened substrate scope. (**C**) Santos and co-workers published the first borylation of α,β-unsaturated carbonyl compounds in water in 2012, using copper sulphate and amine bases. (**D**) Časar and co-workers developed conditions employing a combination of basic copper carbonate and a phosphine ligand for the borylation of α,β-unsaturated carbonyl compounds in water in 2013. They also report the borylation of an α,β-unsaturated amide.

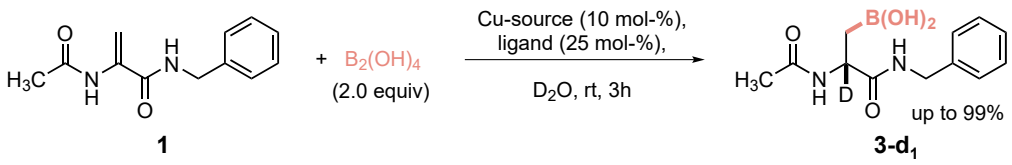
2.2 Borylation of a Small-Molecule, Dha Containing Model Substrate

Inspired by previous literature reports of borylation of α,β-unsaturated carbonyl compounds in aqueous media,[17,26–29] a small molecule model substrate was chosen as a basis to investigate the possibility of adapting this methodology for the borylation of dehydroalanine (Dha) residues. Such a model substrate should be readily accessible (short synthetic route and simple purification), water soluble, stable under aqueous conditions and exhibit sufficient similarity to a Dha residue within a protein.



Scheme 2.3: One-step synthesis of Ac-Dha-NHBn (**1**) from commercially available 2-acetamidoacrylic acid (**2**)

Model substrate Ac-Dha-NHBn (**1**) was found to fulfil those requirements. The compound can be accessed from commercially available 2-acetamidoacrylic acid (**2**) and benzylamine in one step, can be purified via column chromatography and is sufficiently water soluble.



Scheme 2.4: Borylation of model substrate Ac-Dha-NHBn (**1**) in D_2O yields *d*-Ac-Bal-NHBn (**3**).

First, the influence of different ligands and bases on borylation yield was investigated using model compound Ac-Dha-NHBn (**1**) as substrate and tetrahydroxydiboron ($B_2(OH)_4$) as the boron source (**Scheme 2.4**). The reaction was conducted in D_2O and the conversion to the boronoalanine (Bal) derivative Ac-Bal-NHBn (**3-d$_1$**) was analysed via ^{1}H-NMR spectroscopy (**Table 2.1**). Pyridine and 4-picoline gave excellent yields (96-99%), while bidentate ligands (2,2'-bipyridine and 'proton sponge' 1,8-bis(dimethylamino)naphthalene) only gave low yields (6-11%), possibly due to formation of inactive, coordinatively-saturated complexes. Surpris-

ingly, imidazole and even guanidine were able to catalyse the reaction in the presence of Cu(II) (88% and 18% respectively). This observation indicated that proteinogenic amino acid side-chains could not only be tolerated, but even promote the reaction. Indeed, addition of L-arginine as ligand enabled low conversion (18%) and addition of L-histidine led to good conversion (66%).

Table 2.1: Optimization of Bal formation on peptidic model substrate.

Entry	Variable[#]	NMR-Yield
Influence of the Ligand/Base (Using $CuSO_4$ as Copper Source)		
1	pyridine	96%
2	4-picoline	99%
3	2,2'-bipyridine	6%
4	DMAP	25%
5	proton sponge	11%
6	urea	3%
7	imidazole	88%
8	Gdn·HCl	0%
9	Gdn·HCl, Na_2HPO_4 (1.0 equiv)	18%
10	Na_2HPO_4 (1.0 equiv)	5%
11	no ligand/base	2%
Using Amino Acids as Ligands (Using $CuSO_4$ as Copper Source)		
12	L-histidine	66%
13	L-tryptophan	4%
14	L-proline	0%
15	glycine	3%
16	L-lysine·HCl	2%
17	L-arginine	19%
Influence of the Copper Source (Using 4-picoline as Ligand)		
18	$CuSO_4$	99%
19	$Cu(NO_3)_2$	95%
20	$Cu(OAc)_2$	99%
21	$Cu_2(OH)_2CO_3$	59%
22	$Cu(OTf)_2$	99%
23	no copper	0%

[#]If not specified otherwise, standard reaction conditions are used. Standard reaction conditions: 2-acetylamino-*N*-benzyl-acrylamide (10.0 mg, 45.8 µmol, 1.00 equiv) was dissolved in D_2O (1.0 mL). Tetrahydroxydiboron (8.2 mg, 91.6 µmol, 2.00 equiv), copper source (4.58 µmol Cu(II), 0.10 equiv) and base (11.5 µmol, 0.25 equiv) were added and the mixture was stirred at room temperature for 3 hours. NMR-yields were determined against an internal standard (maleic acid).

Next, the influence of the copper source was investigated. Most copper sources ($CuSO_4$, $Cu(NO_3)_2$, $Cu(OAc)_2$, $Cu(OTf)_2$) led to excellent conversions in the presence of 4-picoline as ligand. Employing basic copper carbonate ($Cu_2(OH)_2CO_3$) only afforded moderate yield

(59%), possibly due to limited water solubility. Different diboron reagents were investigated (**Table 2.2**). Tetrahydroxydiboron ($B_2(OH)_4$) proved to give the highest yield, while bis(pinacolato)diboron (B_2pin_2) and bis(neopentylglycolato)diboron gave lower yields, possibly due to their very limited solubility in water.

Table 2.2: Borylation of small molecule model substrate **1** with different diboron reagents.

Entry	Diboron Reagent	NMR-Yield
1	$B_2(OH)_4$	99%
2	B_2cat_2	90%
3	Bis(neopentylglycolato)diboron	<20%*
4	B_2pin_2	70%*

Reaction conditions: 2-acetylamino-*N*-benzyl-acrylamide (10.0 mg, 45.8 µmol, 1.00 equiv) was dissolved in D_2O (1.0 mL). Diboron reagent (91.6 µmol, 2.00 equiv), $CuSO_4$ (1.1 mg, 4.58 µmol, 0.10 equiv) and 4-picoline (1.11 µL, 11.5 µmol, 0.25 equiv) were added and the mixture was stirred at room temperature for 3 hours. NMR-yields were determined against an internal standard (maleic acid).* NMR-yield based on acetyl protons.

Overall, tetrahydroxydiboron as boron source, 4-picoline as ligand and copper sulphate as metal source proved optimal for the borylation of the model substrate allowing for the regioselective conversion of dehydroalanine (Dha) to boronoalanine (Bal). Thus, these conditions were selected as a starting point for protein borylation.

2.3 Translation to Protein Chemistry: Protein Borylation

Histone H3 was chosen as a model protein for reaction optimization. Histone H3 Cys10 was readily converted to its Dha derivative, Histone H3 Dha10, using the *bis*-alkylating reagent DBHDA (2,5-dibromohexanediamide).[25] Proceeding from the optimal conditions found for the borylation of the small-molecule model compound (**1**), borylation of Histone H3 Dha10, affording the boronoalanine (Bal) mutant Histone H3 Bal10 was investigated.

2.3.1 Influence of the Ligand and Diboron Reagent

Based on the data gathered for borylation of the small molecule model substrate (see **Chapter 2.2**), copper(II) sulphate was chosen as the copper source due to its good water solubility and availability.

Table 2.3: Optimization of Bal formation on model protein Histone H3 Dha10.

Entry	Variable[#]	Yield
Influence of the Ligand/Base (equiv)		
1	2,2'-bipyridine (12.5)	72%
2	DMAP (12.5)	62%
3	proton sponge (12.5)	0%
4	4-picoline (12.5)	92%
5	4-picoline (50.0)	87%
6	none	82%
Influence of the Temperature		
7	room temperature	92%
8	0°C	79%
9	37°C	85%
Influence of the Boron Source		
10	10 equiv $B_2(OH)_4$	16%
11	50 equiv $B_2(OH)_4$	92%
12	250 equiv $B_2(OH)_4$	87%
13	4×15 equiv $B_2(OH)_4$	85%
Influence of Stoichiometry		
14	1.0 equiv $CuSO_4$	86%
15	5.0 equiv $CuSO_4$, 12.5 equiv 4-picoline	92%
16	5.0 equiv $CuSO_4$, 50.0 equiv 4-picoline	87%
17	50.0 equiv $CuSO_4$, 50.0 equiv 4-picoline	62%
Influence of Addition Sequence)		
18	$B_2(OH)_4 \rightarrow CuSO_4 \rightarrow$ 4-picoline	85%
19	4-picoline $\rightarrow B_2(OH)_4 \rightarrow CuSO_4$	87%
20	4-picoline $\rightarrow CuSO_4 \rightarrow B_2(OH)_4$	92%
Influence of Urea as Denaturant		
21	1 M urea	75%
22	3 M urea	82%
23	5 M urea	84%
24	7 M urea	89%
Use of Common Buffer Systems		
25	50 mM NH_4OAc, 3 M Gdn·HCl, pH 7.0	91%
26	50 mM TRIS base, 3 M Gdn·HCl, pH 7.0	92%
27	50 mM HEPES (sodium salt), 3 M Gdn·HCl, pH 7.0	87%

[#]If not specified otherwise, standard conditions are used. Standard conditions: Histone H3 Dha10 in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) at 1 mg/mL, 5 equiv $CuSO_4$, 12.5 equiv 4-picoline, 50 equiv $B_2(OH)_4$, 40 µL reaction volume. Reactions were conducted at room temperature and conversions were determined by LC-MS after 30 min.

Tetrahydroxydiboron ($B_2(OH)_4$) was chosen as the diboron reagent because of its water solubility, allowing for preparation of an aqueous stock solution. The reaction was conducted in NaP_i buffer (100 mM, pH 7.0) using guanidine hydrochloride as denaturant (3 M). The influence of different variables on the borylation yield was determined (**Table 2.3**). First, the influence of different ligands was investigated. While 4-picoline gave the highest yield (92%), the reaction still proceeded with good yield (82%) even in absence of a ligand. This observation is in agreement with observations made on the model substrate, which indicated that specific side-chain residues or even guanidine could be catalytically active ligands. Addition of 2,2'-bipyridine or DMAP slightly inhibited the reaction and led to decreased yields (72% and 62% respectively), while addition of proton sponge (1,8-bis(*N*,*N*-dimethylamino)naphthalin) led to complete inhibition of borylation. Next, the influence of different equivalents of diboron reagent were examined. The reaction proceeded best with 50 equivalents, while reduction to 10 equivalents led to a significant drop in yield (16%). Higher concentrations (250 equiv) or multiple additions (4 × 15 equivalents) did not lead to higher yields. Furthermore, the influence of the addition sequence was investigated. Addition of 4-picoline and $CuSO_4$ followed by $B_2(OH)_4$ gave the highest yield (92%), but deviation from this sequence did not lead to significant drops in yield (85–87%).

2.3.2 Influence of Buffer and pH

The borylation of Histone H3 Dha10 was successful at a wide range of relevant pH values (pH 6.5 – pH 8.5) in sodium phosphate buffer using 3 M Gdn·HCl as denaturant (88–92% yield), but the best results were obtained at pH 7.0 (92% yield) (**Figure 2.1**). Both, urea and guanidinium hydrochloride were tolerated as denaturants and allowed for excellent yields (89% with 7 M urea and 92% with 3 M Gdn·HCl respectively). Increased guanidinium hydrochloride concentrations (5 M Gdn·HCl) did not lead to higher yields, while the yield dropped significantly in the absence of denaturant (65–69%). This indicates that denaturant is beneficial, probably by increasing accessibility to the reaction site, albeit not essential for reactivity. Next, a range of common buffers was tested. Gratifyingly, the reaction proceeded well in all buffer systems (NaP_i, NH_4OAc, TRIS and HEPES) with satisfying yield (87–92%) (**Table 2.3**). In summary, optimal conversions for successful C_β–B_γ protein borylation were reached using 4-picoline

(12.5 equiv) and copper sulphate (5.0 equiv) as ligand and copper source, respectively, and tetrahydroxydiboron (50 equiv) as diboron reagent. Sodium phosphate buffer (100 mM) at pH 7.0 using 3 M Gdn·HCl as denaturant was found to be the optimal buffer choice.

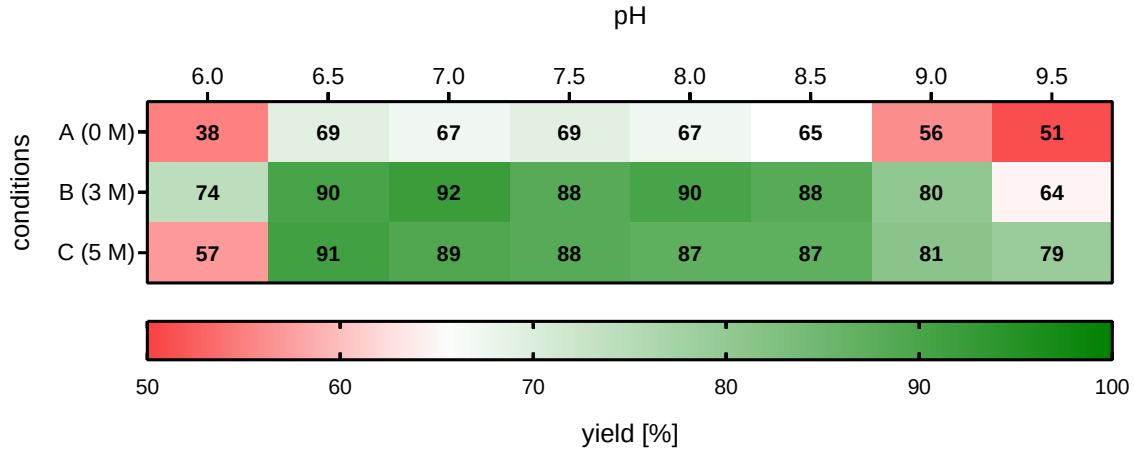


Figure 2.1: Influence of pH on protein borylation at different denaturant concentrations. Numbers in % yield. Reaction conditions: Histone H3 Dha10 in NaP_i buffer (100 mM) at specified pH and 1 mg/mL, 5.0 equiv $CuSO_4$, 12.5 equiv 4-picoline, 50 equiv $B_2(OH)_4$, 40 µL reaction volume. Reactions were conducted at room temperature and conversions were determined after 30 min. A: no addition of Gdn·HCl; B: 3 M Gdn·HCl; C: 5 M Gdn·HCl.

2.3.3 Analysis of the Borylation Product

The reaction product Histone H3 Bal10 was assessed via LC-MS and LC-MS/MS peptide mapping. Interestingly, the deconvoluted mass spectrum of Histone H3 Bal10 showed three distinct peaks for three different borylated species. These peaks corresponded directly to free Bal at site 10 as well as two species accounting for the loss of water, as expected for mono- or di-substituted boronates (also see **Chapter 3, Figure 3.4**). Notably, upon digestion with ArgC and subsequent LC-MS/MS peptide mapping, no peptides corresponding to Bal at position 10 were detected (**Figure 2.2 A**). The lack of coverage around position 10 and the fact that excellent coverage (overall coverage: 90%) was achieved on the remaining sequence indicates possible cross-reactivity of Bal, leading to its disappearance. This hypothesis was tested by oxidation of Histone H3 Bal10 to Histone H3 Ser10 using H_2O_2. Indeed, excellent sequence coverage was observed after oxidation (95%), confirming site selective protein borylation-oxidation (**Figure 2.2 B,C**).

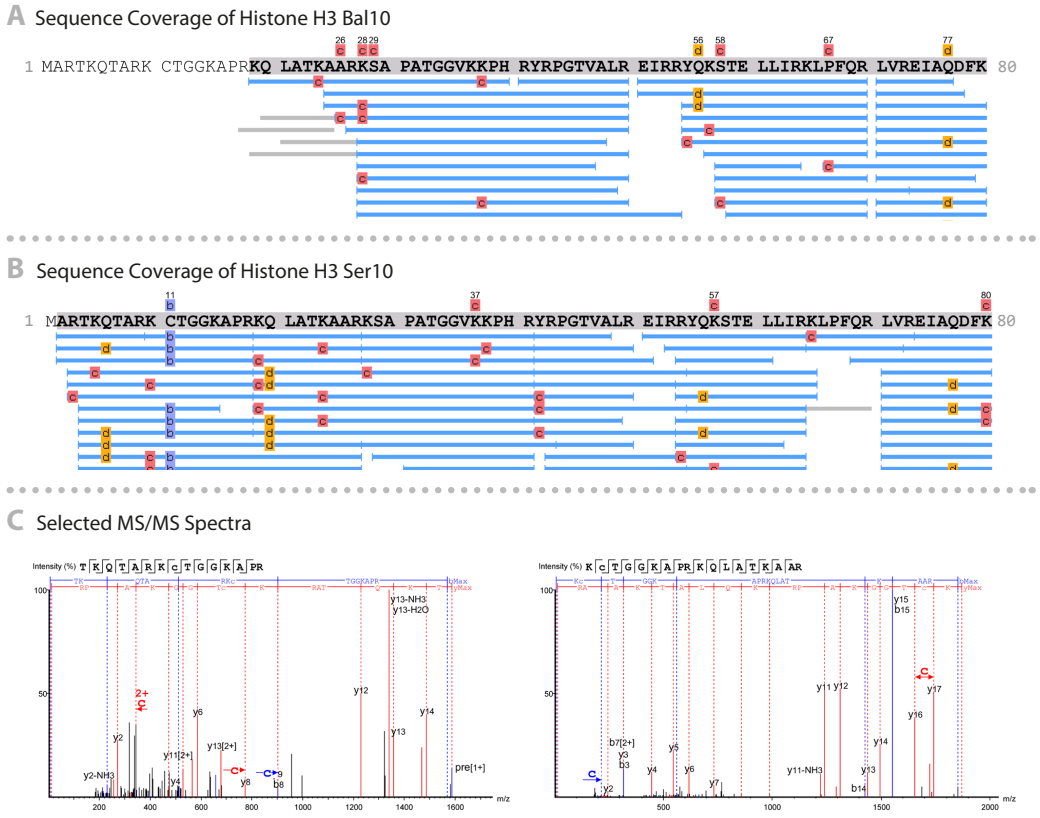


Figure 2.2: ArgC digestion and LC-MS/MS peptide mapping of Histone H3 Bal10 and Histone H3 Ser10. (**A**) Excellent sequence coverage (90%) was detected for Histone H3 Bal10, while no peptide carrying Bal10 was observed. (**B**) Excellent sequence coverage (95%) was detected for Histone H3 Ser10 obtained via oxidation of Histone H3 Bal10. Multiple peptides carrying the desired modification were detected. (**C**) MS/MS spectra for peptides TKQARKSTGGKAPR and KSTGGKAPRKQLATXAAR confirm site-selective modification.

Next, the fate of Bal during digestion and LC-MS/MS analysis and the possibility to inhibit the previously postulated cross-reactivity was investigated (**Figure 2.3**). Digested Histone H3 Bal10 was analysed in the presence of excess boronic acid binding triol **4**, which should allow for blocking of undesired boronate formation by acting as a transient masking group (**Figure 2.3 A,B**). Consequently, this should allow for good coverage of the relevant sequence and direct detection of Bal at position 10 after LC-MS/MS peptide mapping. Gratifyingly, good sequence coverage (88%) was obtained. Peptides confirming the presence of Bal at position 10 were detected and enabled the direct confirmation of site-selective borylation of Histone H3 Dha10 (**Figure 2.3 C,D**). With these promising results on the model protein in hand, a protein scope was established.

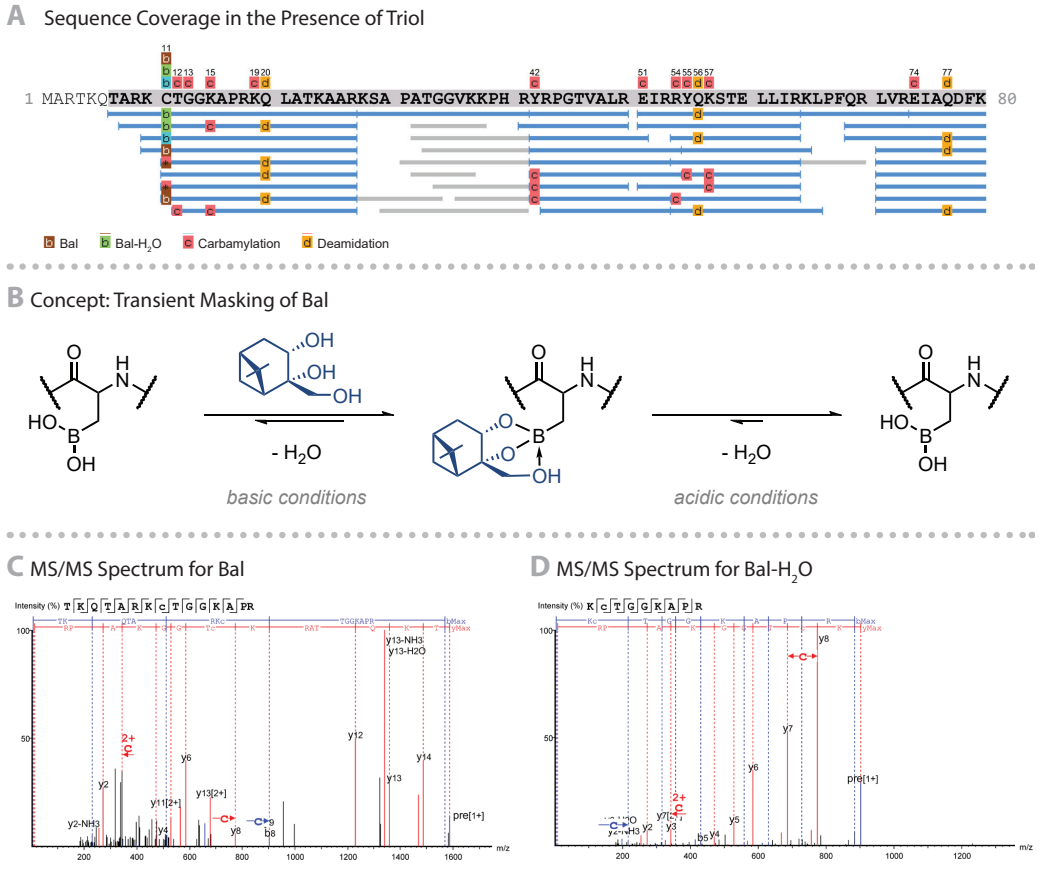


Figure 2.3: ArgC digestion and LC-MS/MS peptide mapping using triol **4** as a transient masking group to inhibit unspecific boronate formation. (**A**) Good sequence coverage (88%) was observed in the presence of triol **4**. (**B**) Triol **4** was used as a transient masking group for Bal allowing for its detection via LC-MS/MS. MS/MS spectra for peptides TKQTARK-**Bal**-TGGKAPR (**C**) and K-(**Bal**–$\mathbf{H_2O}$)-TGGKAPR (**D**) directly confirming site-selective borylation of Histone H3 Dha10.

2.4 Scope of Protein Borylation

A range of proteins differing in size, fold, stability, biological function and reaction site were selected to assess the general applicability of the borylation methodology: Histones H3 and H4 (small α-helical nuclear proteins), Npβ (a pentapeptide repeat protein consisting mostly of β-sheets),[30] acrA (a membrane protein and part of a drug efflux pump complex),[31] mCherry (a red fluorescent protein comprising a β-barrel and being known to be sensitive to oxidation),[32] preSUMO1 (a small globular protein consisting of α-helices and β-sheets and containing a cleavage site relevant in the SUMOylation pathway),[33] Annexin V (a

globular α-helical protein, widely known for its ability to bind phosphatidylserine residues and commonly applied in flow cytometry for apoptosis detection),[34,35] PanC (pantothenate synthetase, an enzyme)[36] and PstS (a protein involved in bacterial phosphate transport and applied for phosphate sensing).[37] This selection of proteins should enable the determination of the generality and limitations of the methodology as well as its protein compatibility.

2.4.1 Preparation of Protein Substrates

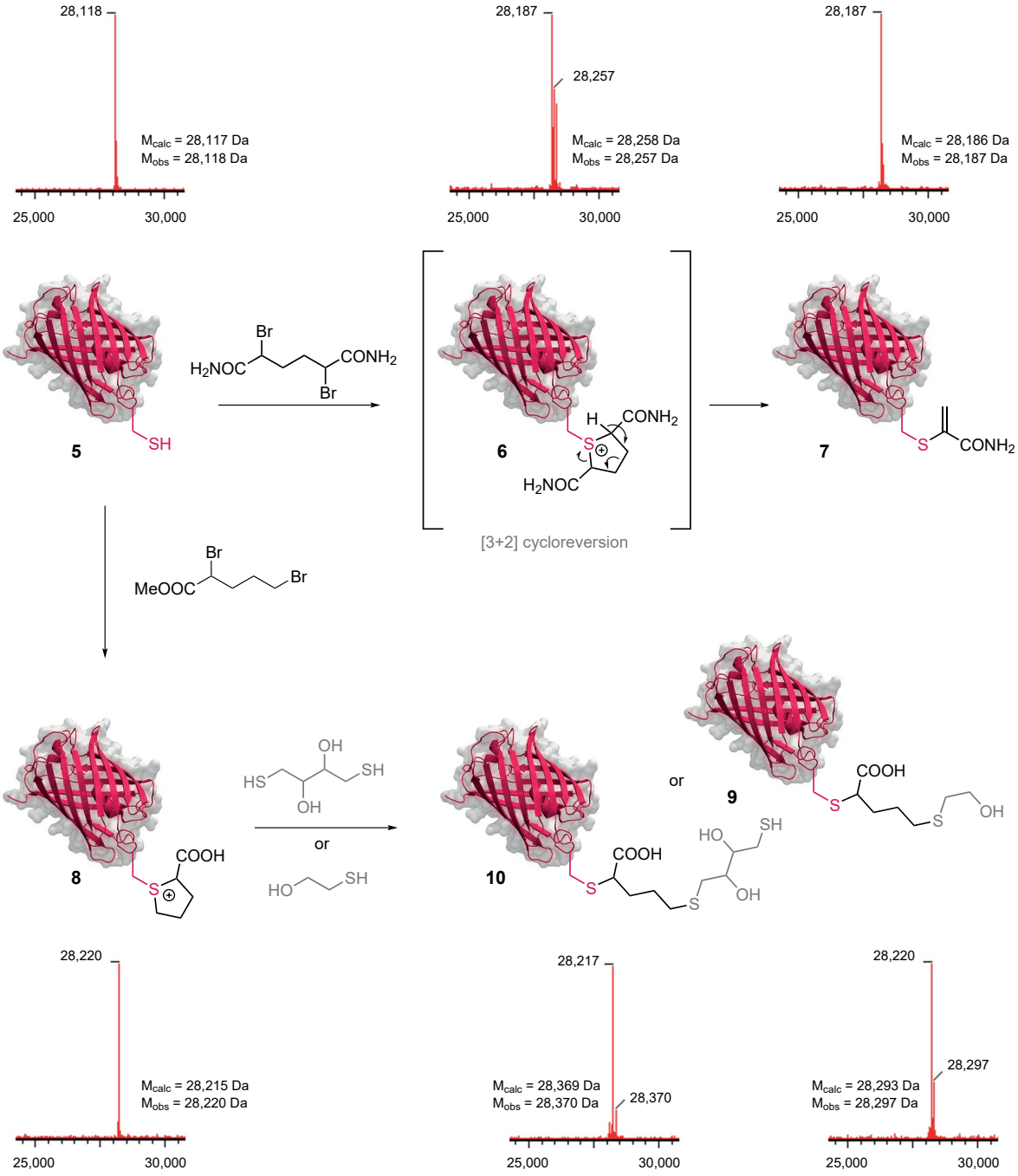
Dha formation on Histones H3 and H4, Npβ, PanC and acrA was conducted following published literature procedures.[38,39] Dha formation on Annexin V was achieved with a modified literature procedure,[38] while dedicated procedures were developed for mCherry, preSUMO1 and PstS.

A previously published procedure for Dha formation on Annexin V 316C using high concentrations (500 equiv) of 2,5-dibromohexanediamide (DBHDA) did not allow for reliable and clean product formation due to high degrees of additional unspecific alkylation observed. Therefore, a modified procedure was developed using methyl 2,5-dibromopentanoate (MDBPA, 100 equiv, then 50 equiv). This procedure was found to reliably afford product with only minor double alkylation observed.

Dehydroalanine formation on preSUMO1 C51 was first attempted using DBHDA at varying concentrations (50–500 equiv). While the desired product could be observed via LC-MS, significant double alkylation was present even before all starting material was consumed. When the alkylating reagent was changed to MDBPA (500 equiv), no double alkylation was observed even though prolonged reaction times (12–17 h at 37°C) were necessary to drive Dha formation to completion.

Attempted Dha formation on PstS C197 under conditions employed for Annexin V or preSUMO1 did not yield any product and no conversion was observed even at high concentrations of alkylating reagent (500 equiv) and long incubation times (20 h at 37°C). Next, alkylation using Mukaiyama's reagent followed by base-promoted elimination was attempted.[25] While the protocol allowed for successful dehydroalanine formation, significant adduct formation was observed possibly due to unspecific alkylation. Low accessibility of the targeted Cys residue likely contributes to the low reactivity and specificity observed. Hence, the reaction was conducted under denaturing conditions (3 M Gdn·HCl). Excellent alkylation selectivity

was detected under these conditions leading to clean formation of PstS Dha197 using DBHDA (250 equiv) as alkylating reagent.



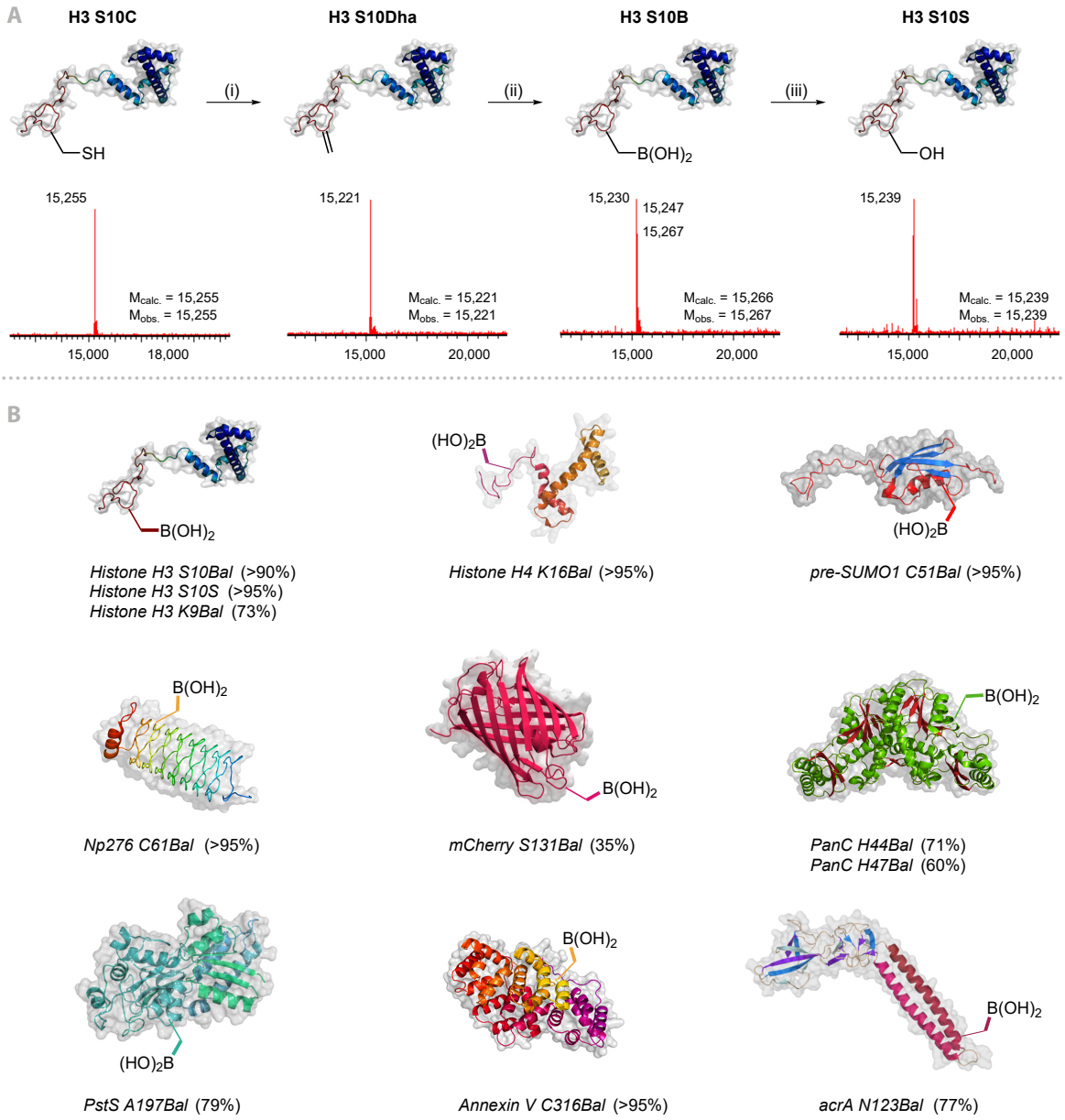
Scheme 2.5: Attempted Dha formation on mCherry Cys131 (**5**) led to different products dependent on the *bis*-alkylation reagent used. Treatment with DBHDA led to the formation of vinyl sulfid **7**, while treatment with MDBPA yielded sulfonium **8**. Sulfonium **8** was found to be reactive towards DTT and BME. All reaction products were identified by LC-MS.

Dehydroalanine formation on mCherry Cys131 (**5**) was attempted using both DBHDA and MDBPA (100 equiv) (**Scheme 2.5**). In both cases, adducts were obtained that did not match the expected mass for dehydroalanine formation. Treatment with DBHDA likely led to formation of the expected sulfonium (**6**) via *bis*-alkylation of the targeted Cys residue. The sulfonium then underwent [3+2]-cycloreversion resulting in vinyl sulfide (**7**) as previously reported on green fluorescent protein (GFP).[40] In contrast, treatment with MDBPA likely led to formation of the corresponding sulfonium intermediate (**8**) in excellent conversion and purity with no signs of elimination or [3+2]-cycloreversion detectable. While the reason for the lack of reactivity towards elimination remains unknown, [3+2]-cycloreversion might be suppressed by the presence of the carboxylic acid inhibiting α-deprotonation. When sulfonium **8** was treated with thiols such as β-mercaptoethanol (BME) or dithiothreitol (DTT), adduct formation was observed via LC-MS. This is likely caused by nucleophilic attack of the sulfanyl moiety at the α-carbon of the thiophenium moiety and subsequent ring opening, leading to sulfides **9** (reaction with DTT) and **10** (reaction with BME). While no LC-MS/MS peptide mapping was conducted for these samples, the postulated reaction products are in agreement with all LC-MS observations (**Scheme 2.5**). Furthermore, sulfonium formation on the related GFP was previously reported and employed for the conjugation of S-nucleophiles.[40,41]

Comparable reactivity was observed when Dha formation was attempted on α-hemolysin. Hence, the β-barrel structural motif seems to greatly affect Dha formation, possibly inhibiting α-deprotonation of the *bis*-alkylated cysteine residue. This would lead to either a stable sulfonium derivative or the vinyl sulfide, if [3+2]-cycloreversion is favoured and could explain the reactivity observed in β-barrel proteins.

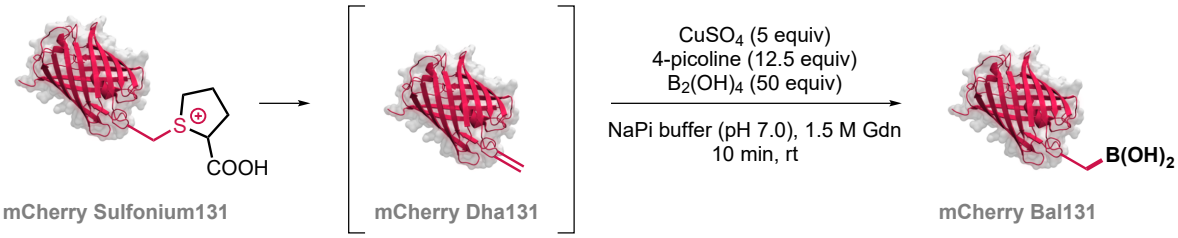
2.4.2 Borylation of Proteins

With a wide range of Dha mutants in hand, protein borylation was attempted on these substrates using the conditions optimized on model protein Histone H3 Dha10 (50.0 equiv $B_2(OH)_4$, 12.5 equiv 4-picoline, 5.0 equiv $CuSO_4$, NaP_i buffer (100 mM, pH 7.0), 3 M Gdn·HCl). Good to excellent conversions were achieved for most proteins after optimization (**Scheme 2.6**). In some cases, more equivalents of reagents were necessary to allow for optimal conversion. In the case of Annexin V, elevated temperatures and more equivalents of reagents (45°C,



Scheme 2.6: Protein Borylation: Reaction Sequence and Scope. (**A**) Borylation of Histone H3 Cys10 allows for post-translational mutagenesis from Cys to Ser via Dha and Bal. The transformations were confirmed by LC-MS (deconvoluted mass spectra for Histone H3 Cys10, Dha10, Bal10 and Ser10 are depicted). Reaction conditions: (*i*) DBHDA, NaP_i buffer (100 mM, pH 8.0), rt—37°C, 4 h; (*ii*) tetrahydroxydiboron ($B_2(OH)_4$; 50 equiv), $CuSO_4$ (5.0 equiv), 4-picoline (12.5 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0), rt, 5 min.; (*iii*) H_2O_2 (5 mM) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0), rt, 10 min. (**B**) Protein scope of the borylation methodology.

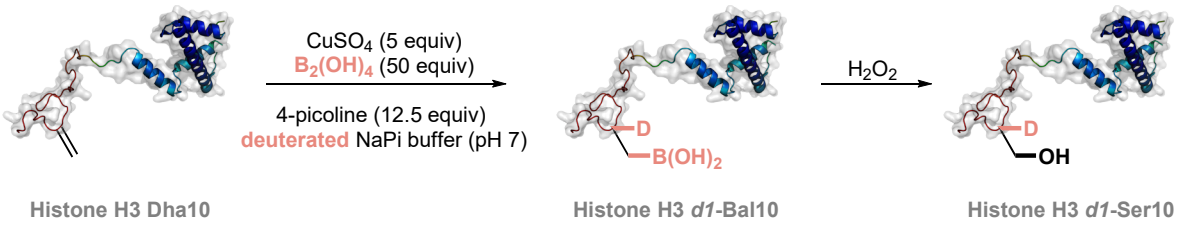
75.0 equiv $CuSO_4$, 187.5 equiv 4-picoline, 750 equiv $B_2(OH)_4$) led to full conversion. Likewise, the borylation of acrA Dha123 (50 equiv $CuSO_4$, 125 equiv 4-picoline, 500 equiv $B_2(OH)_4$; 77%), PanC Dha44 and PanC Dha47 (8.0 equiv $CuSO_4$, 20 equiv 4-picoline, 80 equiv $B_2(OH)_4$; 71% and 60% respectively), preSUMO1 Dha51 (60.0 equiv $CuSO_4$, 150 equiv 4-picoline, 600 equiv $B_2(OH)_4$; >95%) and PstS Dha197 (7.5 equiv $CuSO_4$, 18.8 equiv 4-picoline, 75 equiv $B_2(OH)_4$; 79%) proceeded with good to excellent yields using adjusted equivalents of reagents.



Scheme 2.7: Borylation of mCherry Sulfonium131 via *in situ* elimination under borylation conditions.

Borylation of mCherry was attempted using mCherry Sulfonium131, which hopefully would eliminate to form mCherry Dha131 *in situ* under borylation conditions. Indeed, borylation was observed, albeit in low yield (35%), using standard borylation conditions in NaP_i buffer (100 mM, pH 7.0, 1.5 M Gdn·HCl) (**Scheme 2.7**).

2.4.3 Site-Selective Protein Deuteration



Scheme 2.8: Site-selective Protein Deuteration via deutero-borylation allows for access to *d1*-Bal and *d1*-Ser.

Carbon–deuterium bonds are known to absorb in a region (~2100–2200 cm^{-1}) in the IR spectrum that is typically free of other protein absorptions.[42] These motifs therefore enable the generation of vibrational "transparent window" probes, which have been used for appli-

cations including the characterization of protein folding and dynamics, the study of protein environment, and live-cell Raman spectroscopy.[42–44] Besides a recently published method for the chemical introduction of deutero-Cys, the methods available for deuterium incorporation into proteins are currently limited.[40] They are mainly based on peptide synthesis and native chemical ligation or on biosynthetic approaches.[44] In the described borylation reaction, water likely provides the H-source (see also **Chapter 6.2**).

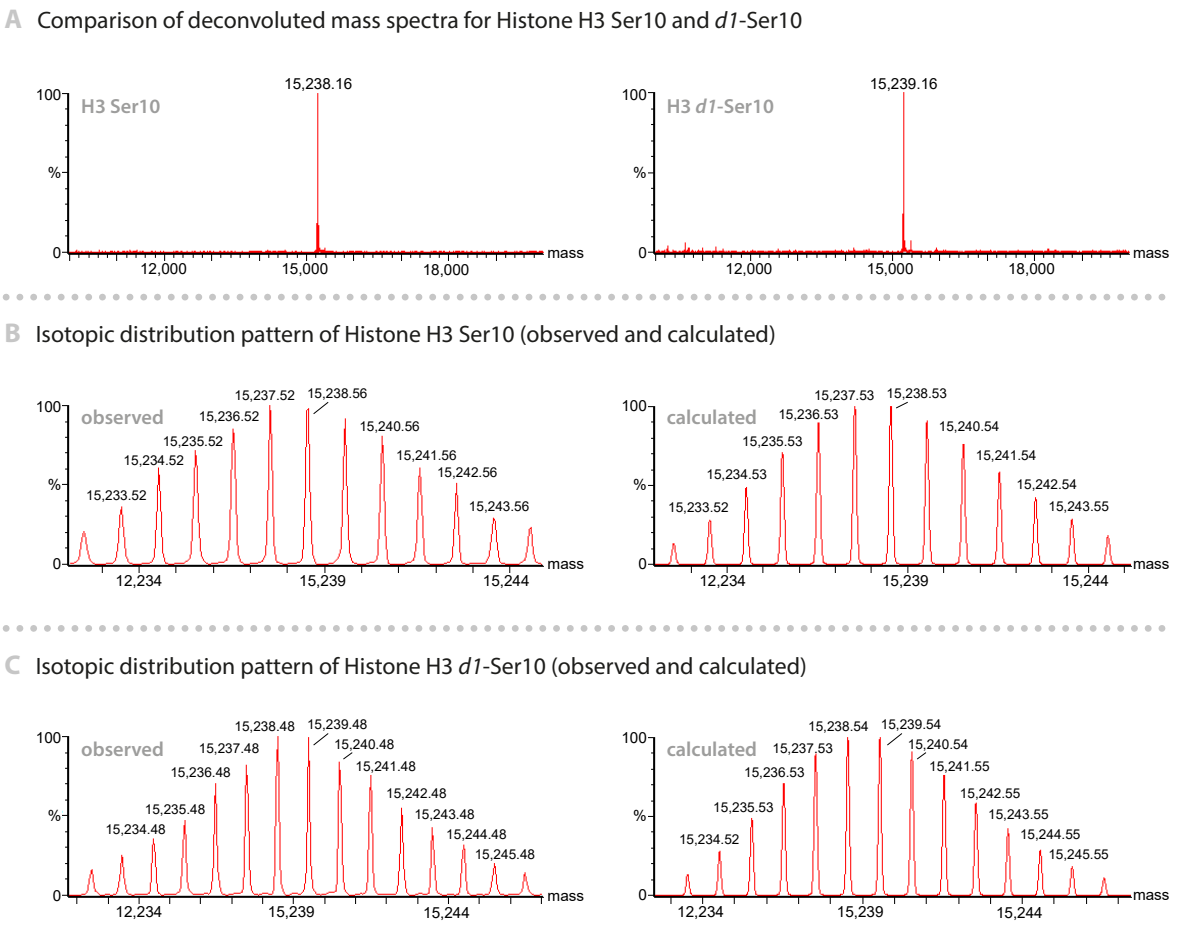


Figure 2.4: LC-MS analysis of Histone H3 *d1*-Ser: (**A**) Comparison of deconvoluted mass spectra of Histone H3 Ser10 and *d1*-Ser10. (**B**) Comparison of observed and calculated isotopic distribution patterns for Histone H3 Ser10 and Histone H3 *d1*-Ser10 (**C**). Measured patterns were obtained by high resolution Q-TOF mass spectrometry and deconvolution using the MaxEnt1 algorithm (mass resolution: 0.04 Da/channel, uniform gaussian, half height: 0.04 Da, minimum intensity ratios: 33 % left and right) preinstalled on MassLynx 4.1 (Waters). Calculated patterns were obtained using the isotope modelling feature (Create Continuum spectrum, peak width at half-height: 0.10, separation: 1.00, min abundance (%): 1.00, full range) preinstalled on MassLynx 4.1 (Waters) using the molecular formula $C_{670}H_{1131}N_{215}O_{186}S_2$ for Histone H3 Ser10 and $C_{670}H_{1130}N_{215}O_{186}S_2D$ for Histone H3 *d1*-Ser10. The atomic mass of D was defined as 2.014102 Da.

Therefore, if the reaction is conducted in deuterium oxide (D_2O), α-deuteration should be possible. Indeed, upon borylation of model substrate **1** in D_2O, a high amount of deuterium incorporation was observed (92%). Consequently, borylation of Histone H3 Dha10 in deuterated NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) led to the deuterated product Histone H3 *d1*-Bal10. Subsequent oxidation with hydrogen peroxide (H_2O_2) gave access to Histone H3 *d1*-Ser10 (**Scheme 2.8**). LC-MS analysis revealed a high degree of deuteration with a 1 Da difference in mass between the deuterated mutant and the control as expected (**Figure 2.4**).

2.5 Protein Compatibility of Borylation

The influence of the borylation reaction conditions on protein structure and function was investigated for several different proteins by structural and functional analyses. First, CD spectra were recorded for selected proteins (Annexin V, Histone H3, Npβ, preSUMO1 and PstS). An analysis of the secondary structure contents derived directly from CD spectra was conducted by multiple algorithms (K2D3,[45] CDSSTR[46–48] and BeStSel[49]). Secondary structure contents derived from the corresponding crystal structures using different algorithms (DSSP_CONT, P-SEA, PALSSE, STICKS, STRIDE and XTLSSTR as implemented in 2Struc[50]) were generally in agreement with the data derived directly from measured CD spectra (**Table 2.5**).

Table 2.5: Protein secondary structure content estimated from CD spectra for Cys and Bal mutants and comparison to theoretical secondary structure content derived from crystal structures.

Entry	Protein	α-helix [%]			β-sheet [%]		
		Cys	*Bal*	*calc*	*Cys*	*Bal*	*calc*
1	Annexin V C316X	65.3±2.9	48.5±3.3	72.1±4.9	2.6±1.9	9.0±4.3	1.5±1.9
2	Histone H3 S10X	66.9±24.9	65.7±26.7	52.1±5.0	7.8±7.3	8.3±7.8	2.4±4.1
3	Npβ C61X	13.5±5.6	9.9±5.0	5.4±0.5	26.0±2.9	23.4±4.9	20.8±31.1
4	preSUMO1 C51X	17.2±12.4	9.5±3.9	16.0±6.3	21.3±6.8	19.1±5.0	31.7±5.4
5	PstS C197X	77.4±8.9	76.2±5.3	41.5±4.3	6.7±4.1	5.3±3.0	24.0±6.3

Protein secondary structures were estimated from CD spectra using K2D3,[45] CDSSTR[46–48] and BeStSel.[49] Reference values were determined from the corresponding crystal structures deposited on PDB using the algorithms DSSP_CONT, P-SEA, PALSSE, STICKS, STRIDE and XTLSSTR implemented in 2Struc.[50] The single values obtained from the different algorithms were averaged and the standard deviation was calculated.

An analysis of the difference between the Cys-mutant and the Bal-mutant in secondary

structure revealed minimal perturbation in most cases (**Table 2.6**). This observation is consistent with only minor local changes in global structure as expected for a single point mutation at a non critical site.[51,52]

Table 2.6: Change in protein secondary structure content upon borylation as determined via CD spectroscopy for selected borylated proteins.

Entry	Protein	Δ α-helix	Δ β-sheet
1	Annexin V Bal316	−15.6±0.9	6.4±2.4
2	Histone H3 Bal10	−1.2±1.8	0.5±0.5
3	Npβ Bal61	−3.7±3.8	−2.6±2.7
4	preSUMO1 Bal51	−7.7±12.4	−2.2±7.4
5	PstS Bal197	−1.2±3.8	−1.4±1.1

Protein secondary structures were estimated from CD spectra using K2D3,[45] CDSSTR[46–48] and BeStSel.[49] Changes in secondary structure were calculated as Δ α-helix = $\alpha\text{-helix}_{\text{Bal}} - \alpha\text{-helix}_{\text{Cys}}$ and Δ β-sheet = $\beta\text{-sheet}_{\text{Bal}} - \beta\text{-sheet}_{\text{Cys}}$. The single values obtained from the different algorithms were averaged and the standard deviation was calculated.

Surprisingly, a major difference in secondary structure composition was observed for Annexin V Cys316 and Annexin V Bal316. While Annexin V Cys316 showed 65.3±2.9% α-helix content, a decrease of 15.6±0.9% was observed upon borylation, indicating that in this specific case Bal316 causes global structural perturbation (**Figure 2.5 A**). The observation of an altered structure immediately raised the question of its impact on function. Annexin V is typically employed in flow cytometry for apoptosis imaging due to its ability to bind phosphatidyl serine displayed on the cell surface of apoptotic cells.[35] Hence, a Annexin V Bal316-fluorescein isothiocyanate (FITC) conjugate was synthesized and analysed using flow cytometry. Annexin V Bal316-FITC showed binding to apoptotic Jurkat cells and allowed for distinction between apoptotic and non-apoptotic cells (**Figure 2.5 B**). This confirms that function is retained upon borylation. Nevertheless, the fluorescence signal was weaker than observed when using commercial Annexin V-FITC, which could be due to decreased binding affinity or less efficient fluorescein labelling. Next, the absorption properties of mCherry Bal131 were characterized and compared to mCherry Cys131 and its sulfonium derivative **8**. In all cases, an absorption maximum of 586 nm was observed, which is in agreement with published values (**Figure 2.5 C**).[53] Emission and absorption wavelengths strongly depend on the chromophore environment and thus the structural integrity of the protein.[54]

Finally, the phosphate binding activity of PstS was examined using a competition assay. A

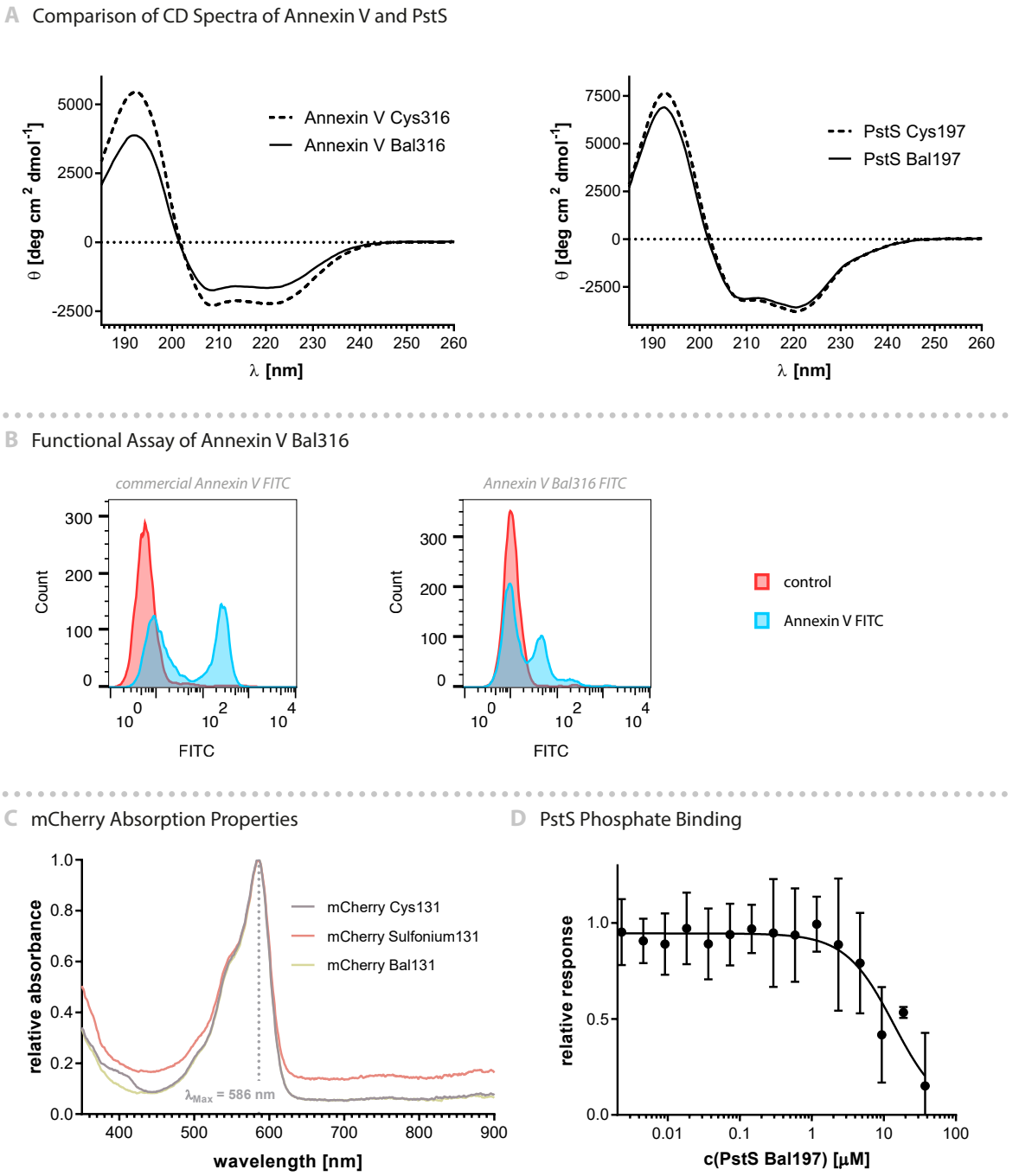


Figure 2.5: Protein compatibility of Dha borylation: (**A**) CD spectra of Annexin V Cys316 and Bal316 and PstS Cys197 and Bal197. While no major structural change was observed after borylation of PstS, significant loss of α-helix content was observed upon borylation of Annexin V Dha316. (**B**) Functional studies on Annexin V Bal316: Despite the observed change in secondary structure, binding of Annexin V Bal316-FITC to apoptotic Jurkat cells was observed. (**C**) Spectrophotometric investigations confirm an absorption maximum for the mCherry mutants mCherry Cys131, mCherry Sulfonium131 and mCherry Bal131. (**D**) Titration of commercial Phosphate Sensor against PstS Bal197 confirmed that the protein retained its ability to bind phosphate.

mixture of a commercially available fluorescent phosphate sensor and inorganic phosphate was titrated against PstS Bal197. PstS Bal197 was found to have retained its phosphate binding ability (**Figure 2.5 D**), which is in agreement with the lack of structural variation observed between PstS Cys197 and Bal197.

2.6 Conclusion

A borylation reaction for the synthesis of boronalanine (Bal) from dehydroalanine (Dha) was optimized on a small molecule model substrate and successfully translated to protein chemistry. The methodology was shown to allow for precise C_β–B_γ bond formation on protein, enabling the first introduction of a minimal boronic acid motif into native proteins. The reaction showed good to excellent yields on a range of different proteins and borylation did not change overall protein fold or function in most cases. This protein compatibility in combination with the ability to precisely introduce a minimal boronic acid motif could allow for the engineering of novel reactivity and unprecedented function.

References

[1] H. DeFrancesco, J. Dudley, A. Coca in *ACS Symp. Ser. Vol. 1236*, **2016**, pp. 1–25.

[2] M. S. Kinch, *Drug Discov. Today* **2015**, *20*, 393–398.

[3] A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.

[4] M. Ibba, H. Hennecke, *FEBS Lett.* **1995**, *364*, 272–275.

[5] T. S. Young, P. G. Schultz, *J. Biol. Chem.* **2010**, *285*, 11039–11044.

[6] P. O'Donoghue, J. Ling, Y.-S. Wang, D. Söll, *Nat. Chem. Biol.* **2013**, *9*, 594–598.

[7] S. J. Baker, J. W. Tomsho, S. J. Benkovic, *Chem. Soc. Rev.* **2011**, *40*, 4279.

[8] E. Brustad, M. L. Bushey, J. W. Lee, D. Groff, W. Liu, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47*, 8220–8223.

[9] B. Akgun, D. G. Hall, *Angew. Chem. Int. Ed.* **2016**, *55*, 3909–3913.

[10] W. J. Ramsay, H. Bayley, *Angew. Chem. Int. Ed.* **2018**, *57*, 2841–2845.

[11] A. Bonet, H. Gulyás, E. Fernández, *Angew. Chem. Int. Ed.* **2010**, *49*, 5130–5134.

[12] S.-C. Ren, F.-L. Zhang, A.-Q. Xu, Y. Yang, M. Zheng, X. Zhou, Y. Fu, Y.-F. Wang, *Nat. Commun.* **2019**, *10*, 1934.

[13] K. Hirano, H. Yorimitsu, K. Oshima, *Org. Lett.* **2007**, *9*, 5031–5033.

[14] W. J. Fleming, H. Müller-Bunz, V. Lillo, E. Fernández, P. J. Guiry, *Org. Biomol. Chem.* **2009**, *7*, 2520–2524.

[15] T. Shiomi, T. Adachi, K. Toribatake, L. Zhou, H. Nishiyama, *Chem. Commun.* **2009**, 5987.

[16] N. J. Bell, A. J. Cox, N. R. Cameron, J. S. O. Evans, T. B. Marder, M. A. Duin, C. J. Elsevier, X. Baucherel, A. A. D. Tulloch, R. P. Tooze, *Chem. Commun.* **2004**, *6*, 1854–1855.

[17] S. Kobayashi, P. Xu, T. Endo, M. Ueno, T. Kitanosono, *Angew. Chem. Int. Ed.* **2012**, *51*, 12763–12766.

[18] I. H. Chen, L. Yin, W. Itano, M. Kanai, M. Shibasaki, *J. Am. Chem. Soc.* **2009**, *131*, 11664–11665.

[19] J. E. Lee, J. Yun, *Angew. Chem. Int. Ed.* **2008**, *47*, 145–147.

[20] J. A. Schiffner, K. Müther, M. Oestreich, *Angew. Chem. Int. Ed.* **2010**, *49*, 1194–1196.

[21] H. Chea, H. S. Sim, J. Yun, *Adv. Synth. Catal.* **2009**, *351*, 855–858.

[22] H. Ito, H. Yamanaka, J.-i. Tateiwa, A. Hosomi, *Tetrahedron Lett.* **2000**, *41*, 6821–6825.

[23] S. Mun, J.-E. Lee, J. Yun, *Org. Lett.* **2006**, *8*, 4887–4889.

[24] J. M. Chalker, G. J. L. Bernardes, B. G. Davis, *Acc. Chem. Res.* **2011**, *44*, 730–741.

[25] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem. Sci.* **2011**, *2*, 1666.

[26] S. B. Thorpe, J. A. Calderone, W. L. Santos, *Org. Lett.* **2012**, *14*, 1918–1921.

[27] T. Kitanosono, P. Xu, S. Kobayashi, *Chem. Asian J.* **2014**, *9*, 179–188.

[28] G. Stavber, Z. Časar, *Appl. Organomet. Chem.* **2013**, *27*, 159–165.

[29] G. Stavber, Z. Časar, *ChemCatChem* **2014**, *6*, 2162–2174.

[30] M. W. Vetting, S. S. Hegde, K. Z. Hazleton, J. S. Blanchard, *Protein Sci.* **2007**, *16*, 755–760.

[31] D. Ma, D. N. Cook, M. Alberti, N. G. Pon, H. Nikaido, J. E. Hearst, *J. Bacteriol.* **1993**, *175*, 6299–6313.

[32] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer, R. Y. Tsien, *Nat. Biotechnol.* **2004**, *22*, 1567–1572.

[33] S. Müller, C. Hoege, G. Pyrowolakis, S. Jentsch, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 202–210.

[34] R. Huber, J. Römisch, E. P. Paques, *Eur. Mol. Biol. Organ. J.* **1990**, *9*, 3867–3874.

[35] M. van Engeland, L. J. Nieland, F. C. Ramaekers, B. Schutte, C. P. Reutelingsperger, *Cytometry* **1998**, *31*, 1–9.

[36] R. Zheng, J. S. Blanchard, *Biochemistry* **2001**, *40*, 12904–12912.

[37] Y. Qi, Y. Kobayashi, F. M. Hulett, *J. Bacteriol.* **1997**, *179*, 2534–2539.

[38] T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallee, A. Phanumartwiwath, O. D. Coleman, M.-L. Thezenas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed, B. G. Davis, *Science* **2016**, *354*, 597.

[39] J. Dadová, K.-J. Wu, P. G. Isenegger, J. C. Errey, G. J. L. Bernardes, J. M. Chalker, L. Raich, C. Rovira, B. G. Davis, *ACS Cent. Sci.* **2017**, *3*, 1168–1173.

[40] S. R. G. Galan, J. R. Wickens, J. Dadova, W.-L. Ng, X. Zhang, R. A. Simion, R. Quinlan, E. Pires, R. S. Paton, S. Caddick, V. Chudasama, B. G. Davis, *Nat. Chem. Biol.* **2018**, *14*, 955–963.

[41] R. Nathani, P. Moody, M. E. B. Smith, R. J. Fitzmaurice, S. Caddick, *ChemBioChem* **2012**, *13*, 1283–1285.

[42] J. Zimmermann, M. C. Thielges, W. Yu, P. E. Dawson, F. E. Romesberg, *J. Phys. Chem. Lett.* **2011**, *2*, 412–416.

[43] J. K. Chin, R. Jimenez, F. E. Romesberg, *J. Am. Chem. Soc.* **2001**, *123*, 2426–2427.

[44] R. Adhikary, J. Zimmermann, F. E. Romesberg, *Chem. Rev.* **2017**, *117*, 1927–1969.

[45] C. Louis-Jeune, M. A. Andrade-Navarro, C. Perez-Iratxeta, *Proteins Struct. Funct. Bioinforma.* **2012**, *80*, 374–381.

[46] L. A. Compton, W. Johnson, *Anal. Biochem.* **1986**, *155*, 155–167.

[47] P. Manavalan, W. Johnson, *Anal. Biochem.* **1987**, *167*, 76–85.

[48] N. Sreerama, R. W. Woody, *Anal. Biochem.* **2000**, *287*, 252–260.

[49] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, *Proc. Natl. Acad. Sci.* **2015**, *112*, E3095–E3103.

[50] D. P. Klose, B. A. Wallace, R. W. Janes, *Bioinformatics* **2010**, *26*, 2624–2625.

[51] S. Shanthirabalan, J. Chomilier, M. Carpentier, *Proteins Struct. Funct. Bioinforma.* **2018**, *86*, 853–867.

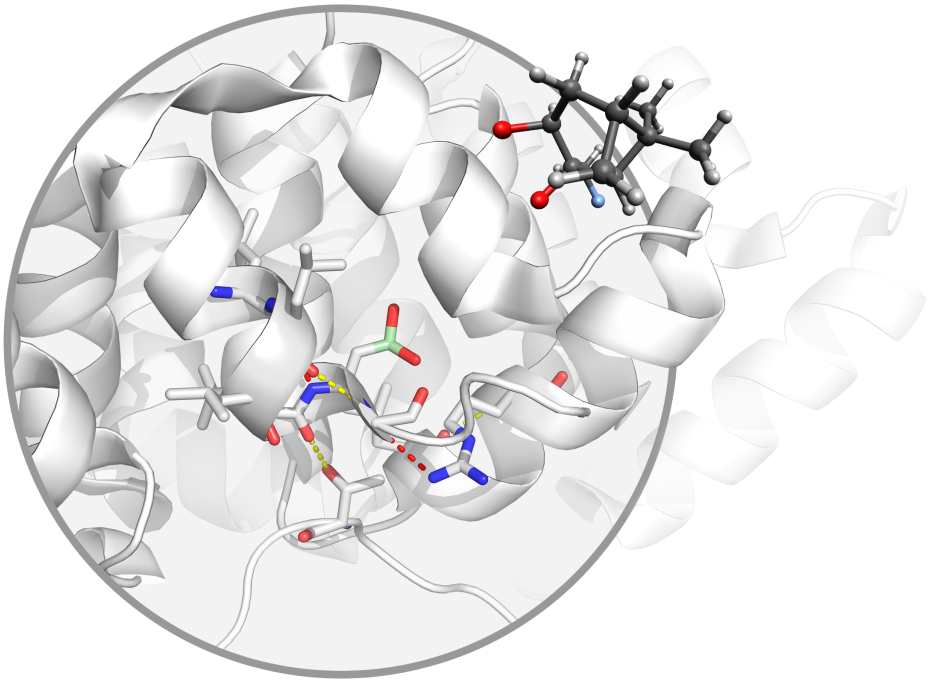
[52] A. Kumar, P. Biswas, *Proteins Struct. Funct. Bioinforma.* **2019**, *87*, 760–773.

[53] C. S. Thomas, M. J. Glassman, B. D. Olsen, *ACS Nano* **2011**, *5*, 5697–5707.

[54] H.-L. Ng, M. Z. Lin, *Curr. Opin. Struct. Biol.* **2016**, *39*, 124–133.

Chapter 3

Boronoalanine (Bal) Enabled *De Novo* Binding Function



Each new insight into how the atoms in their interactions express themselves in structure and transformations, not only of inanimate matter, but particularly also of living matter, provides a thrill.

— *Henry Taube*, Nobel Banquet speech (1983)

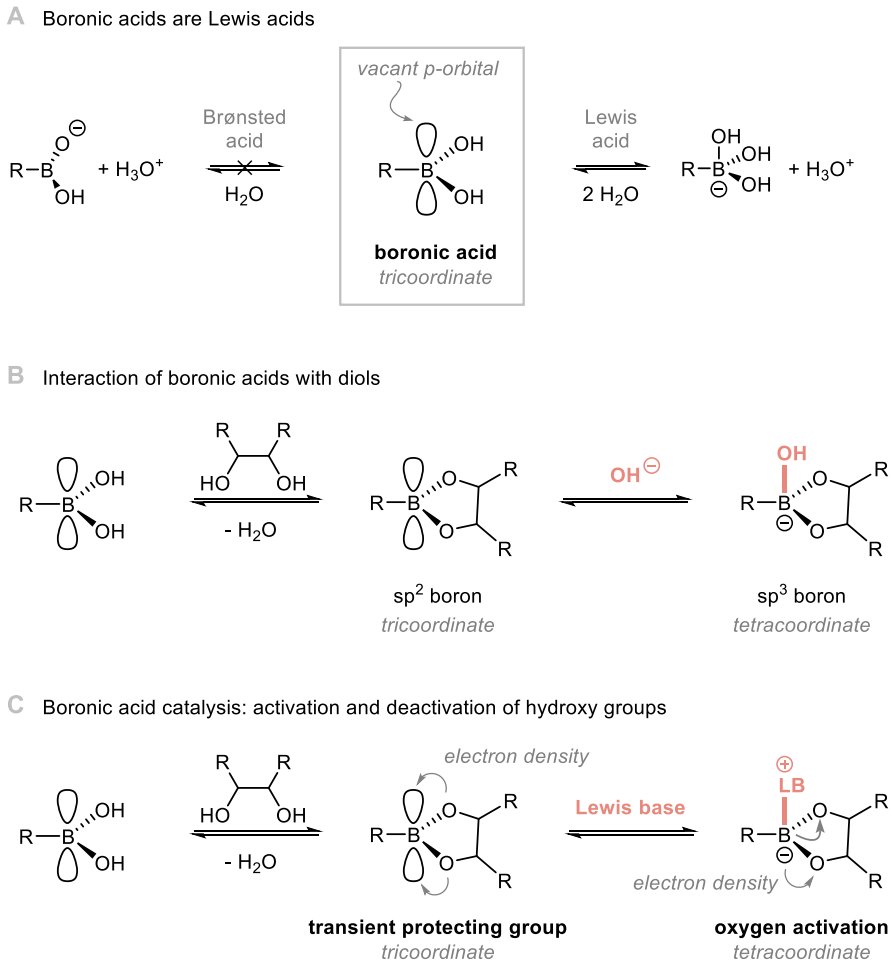
Chapter 3

Boronoalanine (Bal) Enabled *De Novo* Binding Function

Besides their great versatility as building blocks in synthetic organic chemistry, the use of boronic acids recently emerged in multiple fields including catalysis, molecular sensing, bioconjugation and medicinal chemistry. The application of boronic acids in these fields is driven by their unique ability to form reversible covalent bonds with Lewis bases, most notably with *cis*-diols yielding boronic esters (**Scheme 3.1 B**). Several applications were developed that make use of boronic acids and their bespoke binding behaviour. These include examples of enzyme inhibition, boronic acid catalysts that allow for either electrophilic or nucleophilic activation of carbonyl compounds, chemical sensors as well as materials for affinity chromatography.

3.1 Boronic Acids as Reversible Covalent Binders

Boronic acids feature a vacant p-orbital at the sp^2-hybridized boron atom, explaining their Lewis acidity. Thus, boronic acids typically do not behave as Brønsted acids in aqueous solutions (through direct loss of a proton), but rather act as Lewis acids by formally accepting an electron pair from hydroxide and forming a tetracoordinate boron species (**Scheme 3.1 A**).[1]



Scheme 3.1: Boronic acids are Lewis-acidic reversible covalent binders. (**A**) Boronic acids behave as Lewis acids through their vacant p-orbital. (**B**) Boronic acids form reversible covalent bonds with Lewis bases and particularly with diols. The central boron atom can adapt sp^2-hybridization in a tricoordinate, trigonal planar species or sp^3-hybridization in a tetracoordinate, tetrahedral species. (**C**) Formation of tricoordinate esters leads to deactivation through flow of electron density into the vacant p-orbital and allows for boronic acids to serve as transient protecting groups. In presence of a Lewis base, a tetracoordinate species can be formed allowing for enhanced nucleophilicity at oxygen through flow of electron density from boron to oxygen.

Boronic acid based catalysts were employed in reactions such as esterifications, amide bond formations, silylation, aldol reactions, hydrolysis of alkyl halides and the Biginelli reaction.[2–4] Strikingly, boronic acids can act both as transient protecting groups for sugar alcohols and as activating catalysts driving acylation.[5,6] Switching from a tricoordinate boronic acid derivative to a tetracoordinate species is accompanied by a change from deactivating to activating properties. Tricoordinate boronic esters lead to decreased nucleophilicity of the corresponding alcohol due to the flow of electron density of oxygen into the vacant p-orbital

of boron. In contrast, tetraccordinate species lead to activation of the alcohol by increasing electron density and thus nucleophilicity (**Scheme 3.1 C**).[6–8]

A multitude of molecular sensors were developed exploiting the unique ability of boronic acids to form reversible covalent bonds with Lewis bases and in particular with *cis*-diols, such as found in saccharides. Molecular sensors for the detection of saccharides, catecholamines, reactive oxygen species, fluoride, cyanide and metal ions such as Cu^{2+} were described.[9–12] The unique ability of boronic acids to form reversible, covalent bonds with alcohols was also exploited in the development of a tetraserine-selective switch-on fluorescent sensor. The rhodamine based, cell-permeable *bis*-boronic acid RhoBo was shown to display nanomolar affinity to –SSPGSS– tags, enabling fluorescent labelling in cell culture (**Figure 3.1 A**).[13] Furthermore, a boronic acid modified insulin derivative was engineered affording an insulin conjugate that was glucose-responsive *in vivo*.[14] These examples highlight the possible utility of boronic acid based sensors in a biological context. Boronic acid based molecular sensors are covered in depth in several reviews.[9–12]

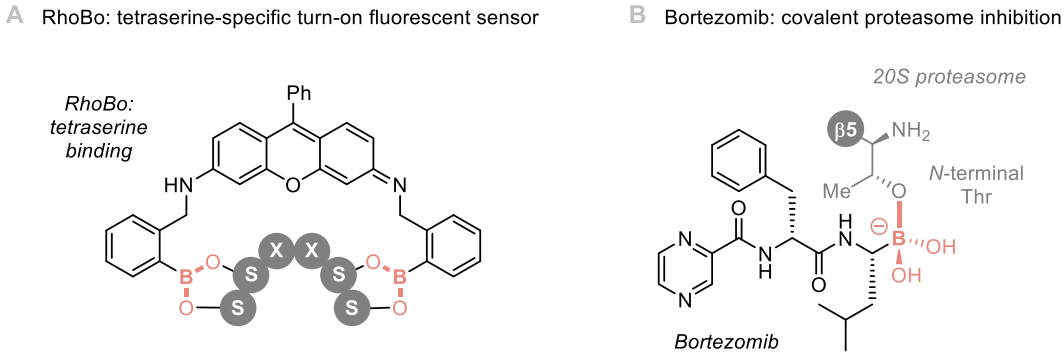


Figure 3.1: Applications of boronic acid binding in a biological context: (**A**) Detection of the tetraserine motif using a switch-on boronic acid based fluorescent sensor. (**B**) Proteasome inhibition through formation of a tetragonal covalent complex between the catalytic *N*-terminal Thr residue of the β5-subunit of the 20S proteasome and the boronic acid warhead of bortezomib. This complex mimics the tetrahedral intermediate present during peptide bond cleavage.

More recently, boronic acids have begun to gain attention as reactive warheads in medicinal chemistry.[15–18] The ability to act as reactive electrophiles towards nucleophilic amino acid residues (such as hydroxyl, imidazole or carboxyl) leading to tetracoordinate boronates, renders boronic acids versatile motifs for enzyme inhibition. Indeed, boronic acid containing structures were proposed for anticancer, antibacterial, antifungal and antiviral applications.[16,19] So far,

five boronic acid containing drugs were approved by the FDA.[19] Bortezomib, which is used in the treatment of multiple myeloma, was the first FDA approved drug containing a boronic acid and also the first approved proteasome inhibitor.[20] The boronic acid hereby binds the catalytic, *N*-terminal threonine residue of the β5-subunit of the 20S proteasome, leading to inhibition through formation of a tetragonal covalent species (**Figure 3.1 B**).[18,21,22]

3.1.1 Wulff-type Boronic Acids

The ability of a boronic acid to engage with diols depends on multiple factors, including its pK_a as well as the external environment and in particular its pH.[9,23] While the external environment might be predetermined and hard to control, direct modulation of the boronic acid is possible. pK_a modulation can be achieved by introduction of electron-withdrawing (increasing acidity) or electron-donating (decreasing acidity) groups.[24] A different strategy

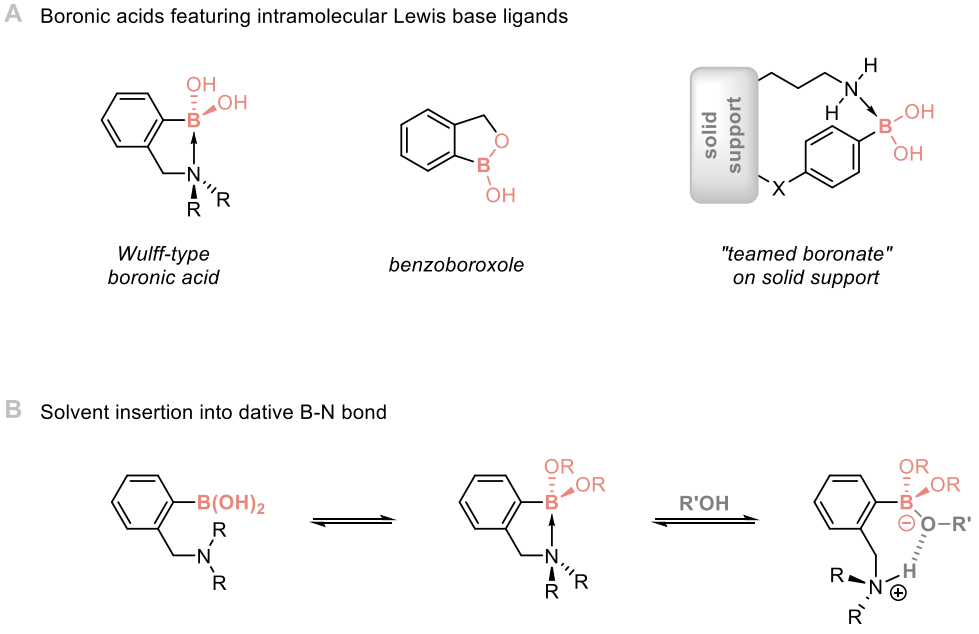


Figure 3.2: Enhancement of boronic acid·ligand binding through proximal Lewis bases. (**A**) Several boronic acids featuring intramolecular Lewis bases were described for enhanced binding. (**B**) Solvent insertion into dative B—N bonds as described for Wulff-type boronic acids in protic solvents leads to hydrogen bonding between the intramolecular nitrogen and the solvent molecule.

was pioneered by Wulff and co-workers: introduction of a Lewis-basic functional group, such as an amino group, in proximity to the boronic acid would allow for the formation of a B—Lewis base bond increasing the sp^3-character of the boron centre (**Figure 3.2 A**).[25]

Indeed, *ortho*-dialkylaminomethyl phenylboronic acid derivatives, so called Wulff-type boronic acids, were shown to not only feature increased acidity but also show enhanced binding kinetics, affinity and specificity.[23–28] In this case an intramolecular dative B—N bond can be formed. In protic solvents, solvent insertion into this bond is possible whereby the nitrogen atom acts as a hydrogen bond acceptor for the boron bound solvent molecule (**Figure 3.2 B**).[29] The general concept of increased boronic acid·ligand engagement through intramolecular binding was further extended to benzoboroxoles (*ortho*-hydroxymethyl boronic acids) featuring intramolecular B—O coordination (**Figure 3.2 A**). Benzoboroxoles show an increased affinity for saccharides and even found application in medicinal chemistry.[30–32] Furthermore, the synthesis of monolithic co-polymers from both phenylboronic acid and amino functionalized monomers allowed this concept to be adapted for boronate affinity chromatography (BAC). The resulting monoliths allowed for the efficient capture of *cis*-diols under neutral conditions (**Figure 3.2 A**).[33] This concept of "teamed boronate affinity", where a proximal Lewis base enhances binding of exogenous ligands, was further employed for the synthesis of magnetic nanoparticles for the selective capture of suitable ligands, including glycoproteins, at neutral pH.[27]

All applications of boronic acids highlighted in this section depend on the unique ability of boronic acids to interact with Lewis bases in a reversible covalent fashion. The concept of enhanced binding through Wulff-type boronic acids featuring a proximal Lewis base as an intramolecular coordination site is appealing in the context of a protein-bound boronic acid, as nearby Lewis-basic residues could act as intramolecular ligands to modulate or even engineer binding. All of this renders boronic acids exceptionally versatile motifs for the engineering of *de novo* binding function into a myriad of different scaffolds, potentially including proteins. The immense wealth of applications based on boronic acids as dynamic covalent binders prompted the further investigations towards utilization of these unique properties in the context of boronoalanine (Bal) containing proteins as described in this chapter.

3.2 Protein Lewis Acid•Base Pair (PLABP) Probing

3.2.1 Boron-Containing Natural Products

Despite its wide abundance in the earth's crust,[34] only a few boron-containing natural products have been described. Examples include the macrolides borophycin[35] and boromycin,[36] rhamnogalacturonan II (RG-II)[37] and the bacterial signalling molecule AI-2[38] (**Figure 3.3**). The microbial polyketides borophycin and boromycin were both found to exhibit antibiotic activities.[34,39,40] Borophycin additionally exhibited potent cytotoxicity against different human cancer cell lines, while boromycin showed *in vitro* activity against HIV as well as *in vivo* activity against *Coccidia* (kingdom *Protozoa*) infestation in chicken.[40,41] RG-II is a pectic polysaccharide found in plant cell walls. Boric acid here provides a cross-link of two homogalacturonan chains via apiose residues contributing to primary cell wall stability.[37] The autoinducer AI-2 was originally found in the bioluminescent marine bacterium *Vibrio harveyi*, where it contributes to the regulation of cell density dependent bioluminescence.[38] In all of these cases leading to diverse function, boron is incorporated as an ester of boric acid exclusively forming B—O bonds (**Figure 3.3**). While compounds featuring B—C bonds, such as boronic acids, do not occur naturally, their main mode of engagement with biological systems is also through Lewis acid·base interactions.[18]

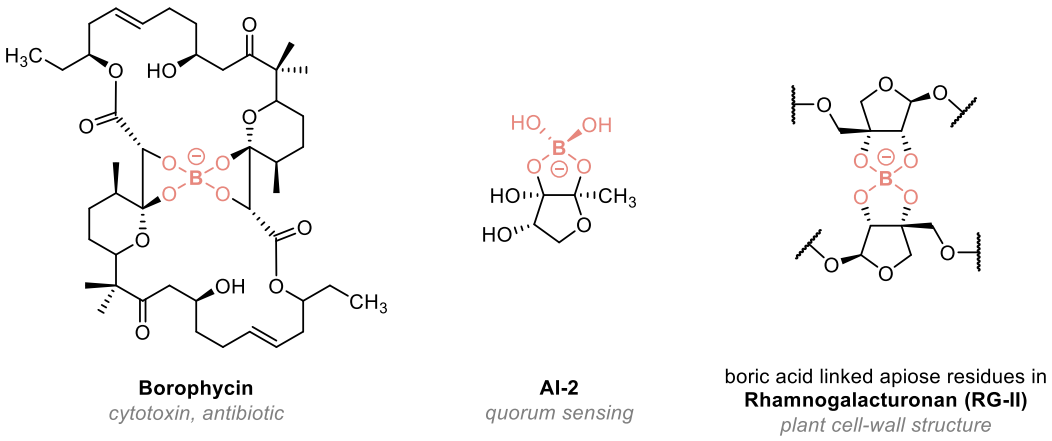


Figure 3.3: Examples of boron containing natural products: boron is exclusively incorporated in form of boric acid esters.

3.2.2 Implications of Anchored Boron: B–C Bond Formation

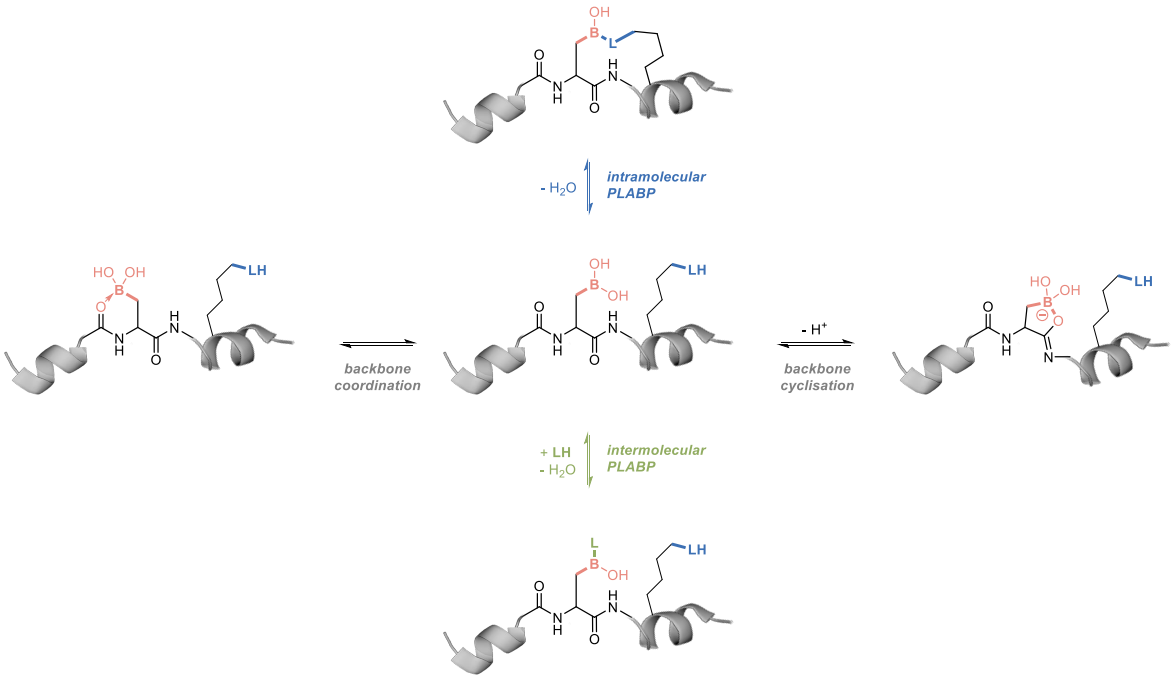
The transient nature and the resulting reversibility of the interactions of boron with Lewis basic moieties make boron containing compounds highly interesting for the modulation of biological systems. Indeed, the application of small molecule bound boronic acids as reversible warheads to engage with proteins has gained attention over the last years.[18,19]



Scheme 3.2: Comparison of putative interactions of a non-anchored boronic acid or boric acid (**A**) and an anchored boronic acid (Bal) (**B**) with a protein of interest.

Nevertheless, the transient character of B–O bonds[42] does not allow for precise and permanent localization of boron in a protein context. In contrast, a boronic acid anchored via a permanent B–C bond (such as found in Bal) would allow for precise spatial control and could provide a useful tool for the investigation of biological systems (**Scheme 3.2**).

Several different modes of interaction of boronoalanine (Bal) with Lewis basic moieties are easily conceivable (**Scheme 3.3**). One can differentiate between intramolecular interactions and intermolecular binding. Possible intramolecular interactions include binding of amino acid side chain residues in proximity to the respective Bal residue (Wulff-type interaction) as well as interactions with the peptidic backbone. Backbone interactions could entail backbone coordination or the formation of cyclic species, in the simplest case a five-membered heterocycle (oxoborolane),[43] through backbone cyclisation (see **Chapter 4**). Intermolecular binding would allow for the targeting of unnatural and natural Lewis basic ligands, including biomolecules such as saccharides.



Scheme 3.3: Possible modes of interaction of boronoalanine (Bal) with intra- and intermolecular ligands: reversible covalent intra- and intermolecular PLABP as well as backbone coordination or backbone cyclisation can be readily conceived.

The introduction of Bal should allow for precise probing of all of these interactions on native proteins (protein Lewis acid·base pair probing, PLABP), enabling the probing of biological systems and re-programming of function.

3.3 Intramolecular Binding

3.3.1 Detection of Intramolecular Binding by Mass Spectrometry

The wide abundance of putative intramolecular ligands available for protein Lewis acid·base pairing due to their proximity to chemically introduced boronoalanine (Bal) prompted investigations towards their possible interactions. Strikingly, upon mass spectrometric analysis of a range of borylated proteins (**Chapter 2, Scheme 2.6**), up to three peaks were observed for the borylated derivative.

Table 3.1: Distribution of Bal substitution across a range of borylated proteins reveals protein- and site-dependent substitution patterns.

Protein	**Distribution of Bal substitution[#] [%]**		
	un-substituted	*mono-substituted*	*di-substituted*
acrA Bal123*	0	100	0
Annexin V Bal316	0	56±3	44±3
Histone H3 Bal9	80±4	12±3	8±4
Histone H3 Bal10	40±5	19±1	41±6
Histone H4 Bal16	48±4	52±4	0
mCherry Bal131	100±0	0	0
Npβ Bal61	21±1	50±2	29±2
panC Bal44*	0	100	0
preSUMO1 Bal51	35±4	54±6	11±3
PstS Bal197	6±5	79±7	15±11

[#] Bal substitution distribution was determined by LC-MS. Unless stated otherwise, samples were measured in triplicate. * n = 1.

These peaks correspond to the mass expected for the borylated protein, for the borylated protein after loss of one water molecule and for the borylated protein after loss of two water molecules. This pattern can be attributed to the ability of the boronic acid to engage in binding with side-chain residues – formally a condensation reaction that proceeds under the loss of water.

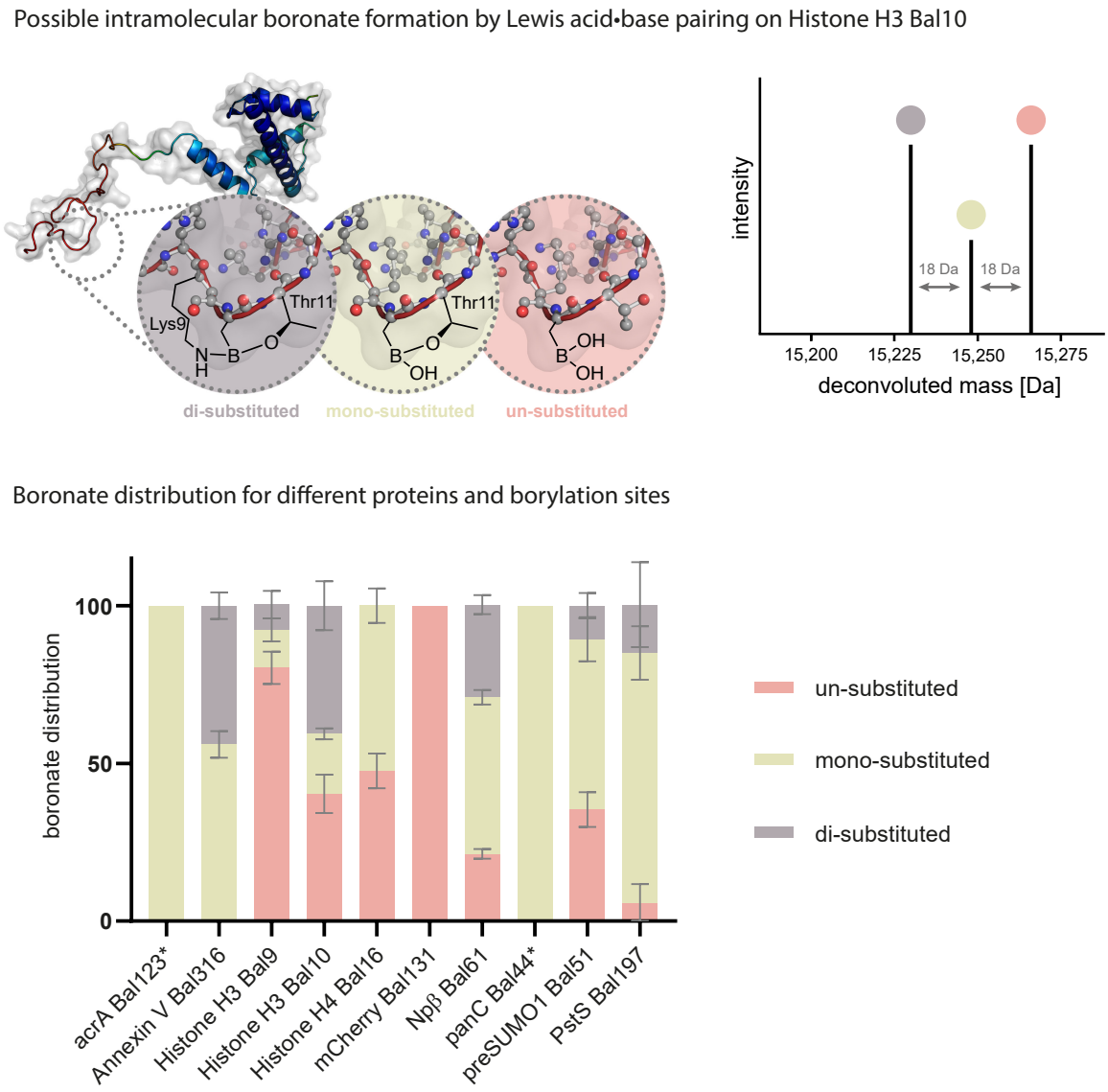


Figure 3.4: Intramolecular Lewis acid·base pairing between Bal and side chain residues leads to diverse boronate distribution patterns. (**A**) Boronate formation on model protein Histone H3 Bal10 possibly involves Lys9 and Thr11. Three different species differing by 18 Da can be observed in the deconvoluted mass spectrum. (**B**) Boronate distribution patterns vary widely for a range of different proteins and borylation sites. Samples were measured in triplicate, samples labelled with an asterisks (*) were only measured once.

Depending on Lewis base 'availability', the presence of an unsubstituted, mono-substituted or di-substituted Bal residue is possible (**Figure 3.4 A**). Interestingly, the ratios of this Lewis acid·base engagement change in a protein-to-protein and site-to-site dependent manner (**Figure 3.4 B**). Hereby, a wide range of substitution patterns is observed: exclusively unsubstituted (mCherry Bal131), exclusively mono-substituted (PanC Bal44), mixtures of either unsubstituted and mono-substituted (Histone H4 Bal16) or mono- and di-substituted

(Annexin V Bal316) as well as varying mixtures of all three substitution states (Histone H3 Bal10 and Npβ Bal61) (**Figure 3.4** and **Table 3.1**). This provides a protein- and site-specific fingerprint of Bal's coordination state which can be linked to the respective Lewis base environment.

3.3.2 Intramolecular Engagement of Bal Modulates Thermostability

Having established that Bal engages with proximal Lewis bases, the influence of these interactions on thermal protein stability was investigated. Indeed, intramolecular covalent protein side-chain cross-linking can lead to dramatically increased thermal stability.[44] A range of proteins (Annexin V Bal316, Npβ Bal61 and PstS Bal197) was selected for thermal stability analysis based on the presence of mono- and di-substitution as observed by LC-MS. Melting temperatures (T_M) and onset (T_{ON}) were determined by nanoDSF (**Figure 3.5**).

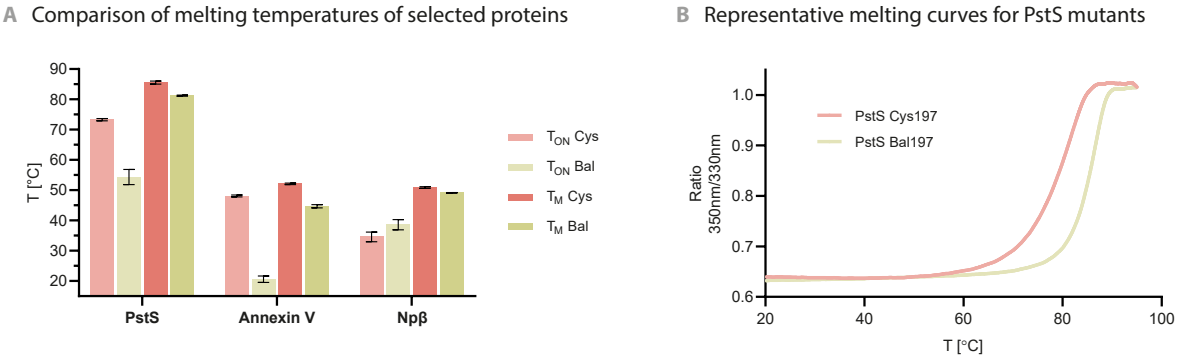


Figure 3.5: Determination of melting behaviour for selected proteins by nanoDSF. (**A**) Introduction of Bal allows for engineered melting behaviour: Engagement of Bal can lead to expansion ($\Delta\Delta T > 0$) of the melting range ($\Delta\Delta T = \Delta T_M - \Delta T_{ON}$; $\Delta\Delta T = +20.1°C$ for Annexin V Bal316 and $\Delta\Delta T = +14.6°C$ for PstS Bal197) consistent with the disturbance of local polar interactions. Tightening of the melting range ($\Delta\Delta T < 0$) can be observed in cases where Bal can cooperatively interact with existing polar interactions such as in Npβ ($\Delta\Delta T = -5.8°C$). (**B**) Representative melting curve for PstS mutants show expansion of melting range consistent with the destabilization of local stabilizing interactions.

Interestingly, decreased melting temperatures and onsets were observed for Annexin V Bal316 ($\Delta T_{ON} = -27.5°C$; $\Delta T_M = -7.4°C$) and PstS Bal197 ($\Delta T_{ON} = -18.9°C$; $\Delta T_M = -4.3°C$) with an increase in 'melting range' ($\Delta\Delta T = \Delta T_M - \Delta T_{ON}$; $\Delta\Delta T = +20.1°C$ for Annexin V Bal316 and $\Delta\Delta T = +14.6°C$ for PstS Bal197). This thermal destabilization is in agreement with Bal 'disturbing' local polar interactions (**Figure 3.6**). Annexin V for example exhibits a hydrogen bond with the carbonyl at the mutation site C_{316}=O acting as an acceptor with

Arg285 acting as the donor. Intra-residue oxaborolane formation could interfere with this hydrogen bond leading to the observed broadened melting range. Bal-enabled locally restricted interference is also in agreement with previous observations of retention of function and global structure while minor structural changes are detectable (see **Chapter 2.5**). In the case of Annexin V Bal316, the local disruption of structure was observed as a loss in α-helical character (−15.6±0.9%) and a gain in β-sheet (+6.4±2.4%) character. For Npβ only a minor decrease in melting temperature ($\Delta T_M = -1.8°C$) and an increase in melting onset ($\Delta T_{ON} = +4.0°C$) was observed. This results in tightening of the 'melting range' ($\Delta\Delta T = -5.8°C$).

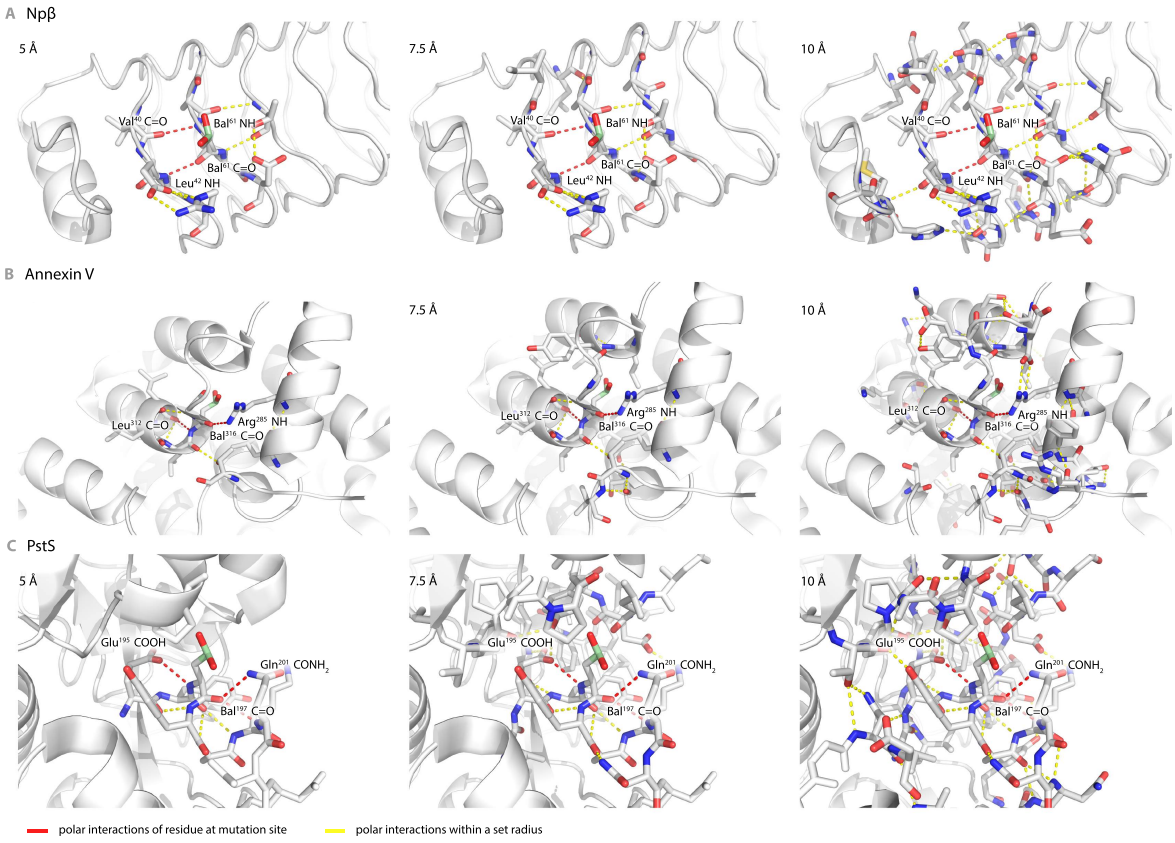


Figure 3.6: Polar interactions in Npβ (**A**), Annexin V (**B**) and PstS (**C**) within a specific distance. PDB deposited crystal structures of Npβ (PDB-ID 2J8K), Annexin V (PDB-ID 1HVD) and PstS (PDB-ID 1A40) were mutated to Bal at the relevant site using PyMol. Polar interactions of the residue at the mutation site (red) and within a set radius (red) of 5 Å (left), 7.5 Å (middle) or 10 Å (right) of a putative Bal residue are displayed as dashed lines.

Engagement of Bal with its local environment (site 61 in Npβ sits within a β-strand which is stabilized by hydrogen bonds to adjacent β-strands[45]) apparently leads to thermal stabilization. Hence, the introduction of Bal, not at an arbitrarily chosen, but at a carefully selected site

should allow for at-will modulation of a protein's thermal stability depending on the *de novo* introduction of stabilizing interactions or the disturbance of existing motifs leading to destabilization.

3.4 Intermolecular Binding

Boronic acids are well known for their ability to bind Lewis basic moieties, especially *cis*-diols. This behaviour has been widely exploited for the synthesis of boronic acid based diol sensors, which can enable diol sensing through different observation modes including ^{19}F-NMR and fluorescence based approaches (**Figure 3.7 A**).[11,28,46] Furthermore, this unique property was utilized as the mode of action of the topical antifungal agent tavaborole, which obtained FDA approval in 2014.[47] Tavaborole acts as an inhibitor of fungal leucyl tRNA synthetase, acting within the editing site of the enzyme. Its boron atom hereby binds to the 2'- and 3'-oxygen atoms of the ribose unit of the 3'-terminal adenosine residue within the corresponding tRNA, locking enzyme and tRNA in the editing conformation and leading to inhibition with a low µM IC_{50}.[32,48] The small molecular size and physicochemical properties of tavaborole were designed to allow for good nail penetration, a property crucial for the topical treatment of onychomycosis and a major contributor to the compound's success.[49]

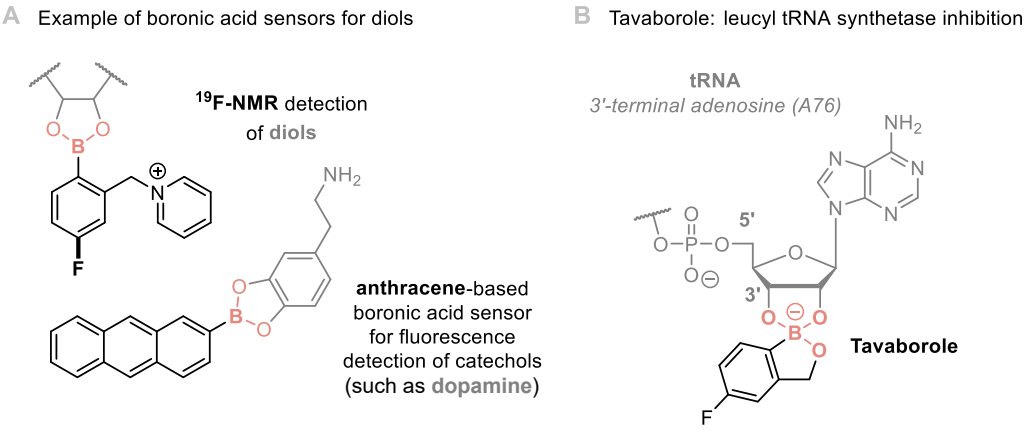


Figure 3.7: Small molecule boronic acid based diol binders: (**A**) Examples of detection of diols using boronic acid based sensors.[28,46] (**B**) Leucyl tRNA synthetase inhibitor tavaborole acts by boronic acid based ribose binding.[32]

Consequently, the ability of boronoalanine to bind different diols and its behaviour in biological

systems was investigated.

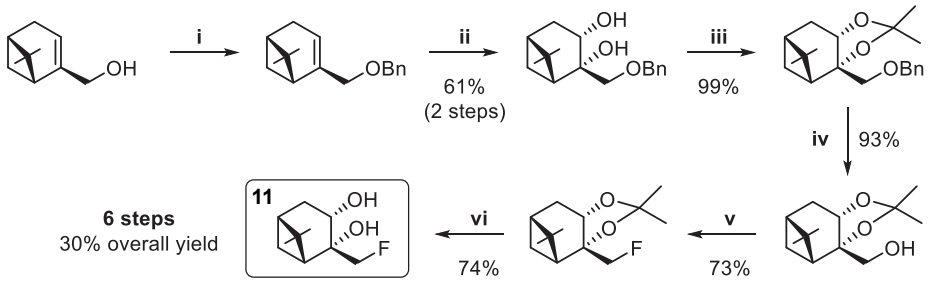
3.4.1 Binding Investigations Using Fluorinated Diols

Saccharides are ubiquitous in biological systems and fulfil many crucial functions, including cellular signalling and recognition as well as providing structure.[50] Saccharides typically exhibit a high abundance of vicinal diols which readily interact with boronic acids.

Synthesis of Fluorinated Diols

To probe these interactions on borylated proteins, several fluorinated diols were synthesized. The introduction of fluorine should hereby aid analysis and allow for detection and characterization of binding interactions via ^{19}F-NMR. Three different diols and polyols were selected for further studies: a fluorinated diol derived from the terpenoid (1*R*)-(−)-myrtenol, a fluorinated ribose derivative and a fluorinated galactose derivative.

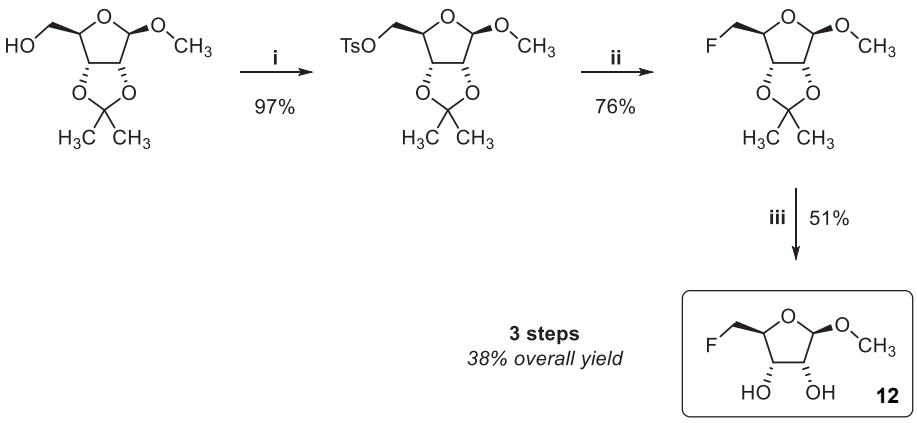
Myrtenol derivative **11** was synthesized in six steps from commercially available (1*R*)-(−)-myrtenol with an overall yield of 30% (**Scheme 3.4**). After benzyl protection of the primary alcohol and osmium tetroxide catalysed Upjohn dihydroxylation, the corresponding diol was obtained. The *cis*-diol was protected by acetonide formation and the benzyl protecting group was cleaved to enable deoxyfluorination with DAST. After acetonide deprotection, the fluorinated diol **11** was obtained.



Scheme 3.4: Synthesis of ^{19}F-NMR diol probe **11**: (i) NaH, BnBr, DMF; (ii) K_2OsO_4, NMO, 4-picoline, iPrOH; (iii) 2,2-dimethoxypropane, *p*TsOH·Py, acetone; (iv) H_2, Pd/C, MeOH; (v) DAST, Et_3N, toluene; (vi) Amberlyst 15, EtOH, H_2O.

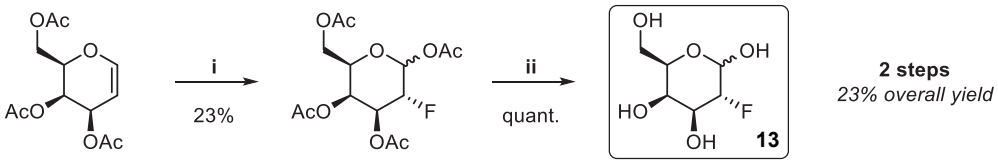
Ribose derivative **12** was synthesized in three steps in 38% yield starting from a commercial precursor (**Scheme 3.5**). The starting material was tosylated and subjected to nucleophilic

fluorination with TBAF. Selective cleavage of the acetonide protecting group with acetic acid finally afforded fluorinated ribose derivative **12**.



Scheme 3.5: Synthesis of fluorinated ribose derivative **12**: (i) TsCl, pyridine; (ii) TBAF, THF; (iii) AcOH, H_2O.

2-Deoxy-2-fluoro-D-galactose (**13**) was synthesized starting from commercially available 3,4,6-Tri-*O*-acetyl-D-galactal following a literature procedure (**Scheme 3.6**).[51] The desired fluorinated saccharide was obtained after fluorination of the commercial precursor with SelectFluor and subsequent deprotection in 23% yield.



Scheme 3.6: Synthesis of 2-Deoxy-2-fluoro-D-galactose (**13**). (i) SelectFluor, AcOH, MeCN; (ii) Na in MeOH, then Dowex 50WX8.

Assessment of Diol Binding on Borylated Small Molecule Model 3

First, the characterization of the engagement of these fluorinated diols and polyols with boronoalanine (Bal) was attempted using small molecule model compound Ac-Bal-NHBn (**3**). In all cases reactivity could be observed: treatment of Ac-Bal-NHBn with fluorinated myrtenol derivative **11** led to immediate precipitation preventing NMR analysis, fluorinated ribose derivative **12** gave a complex mixture of eight different species when treated with

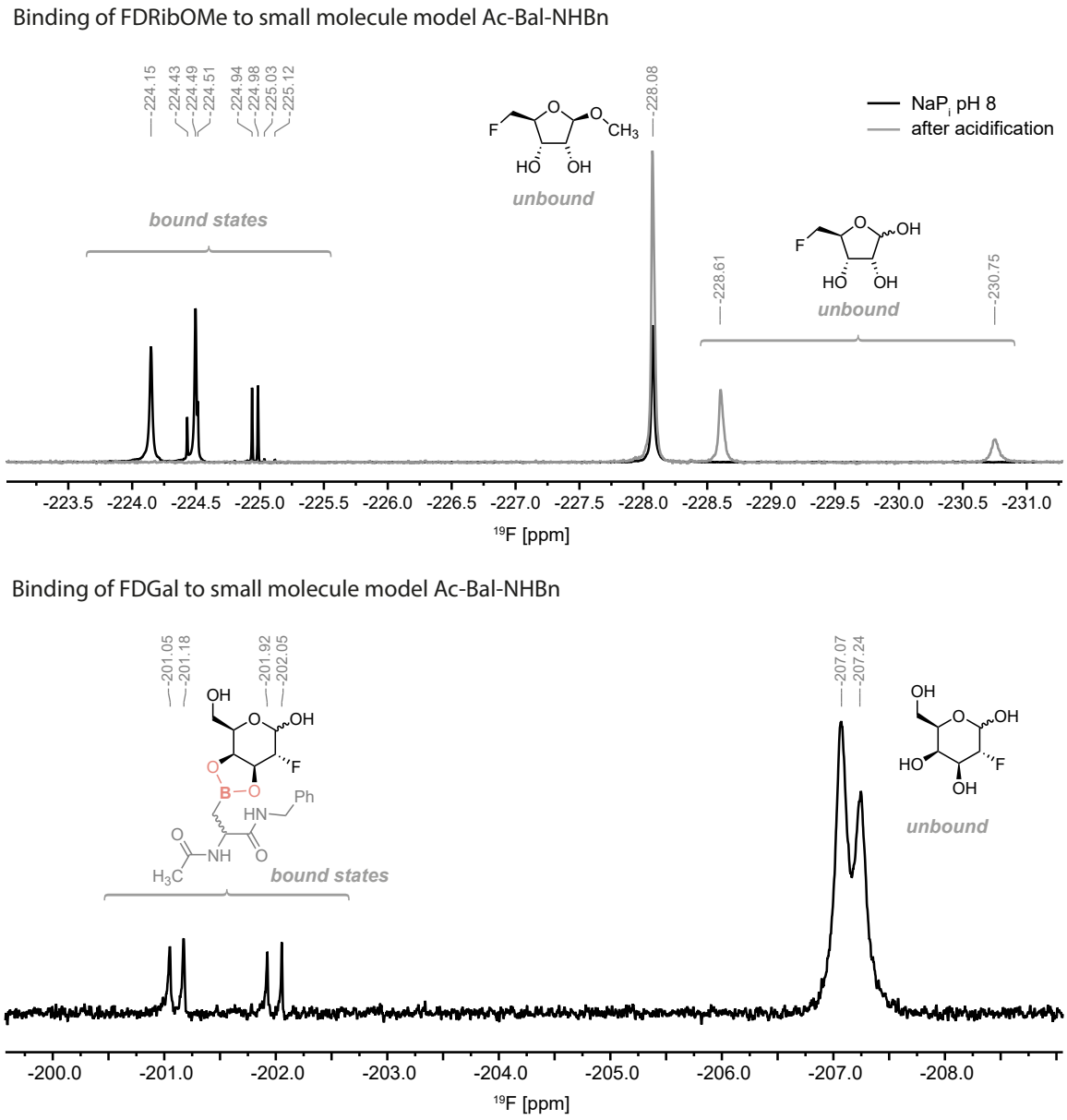


Figure 3.8: Observation of the binding of fluorinated diols to small molecule model compound Ac-Bal-NHBn (**3**) by ^{19}F-NMR. (**A**) Mixing of fluorinated ribose derivative FDRibOMe (**12**) with an equimolar amount of small molecule model compound Ac-Bal-NHBn (**3**) under basic conditions results in eight peaks corresponding to bound species as determined by decoupled fluorine NMR. Acidification with concentrated HCl leads to disappearance of the bound species and the appearance of two new peaks corresponding to methoxy deprotection (−228.61 ppm: β-anomer, −230.75 ppm: α-anomer[52]). (**B**) Binding of 2-Deoxy-2-fluoro-D-galactose (**13**) to Ac-Bal-NHBn (**3**) can be observed in the presence of a large excess of the boronic acid (10 equiv) under basic conditions (NaP$_i$ buffer, pH 8). The ^{19}F-{^{1}H}-NMR spectrum reveals the presence of four bound species corresponding to all possible permutations for the α- and β-anomer of the fluorosugar and the *R*- and *S*- isomers of Ac-Bal-NHBn.

Ac-Bal-NHBn under basic conditions and 2-Deoxy-2-fluoro-D-galactose (**13**) showed weak binding as indicated by four peaks (binding of the α- and β-anomer of the fluorosugar to the respective *R*- and *S*- isomers of Ac-Bal-NHBn) when treated with a large excess of model compound **3** (**Figure 3.8**). These first successful diol binding attempts prompted more thorough analysis of diol binding on model protein Histone H3.

Diol Binding on Borylated Model Protein Histone H3

Next, binding of fluorinated myrtenol derivative **11** to Histone H3 Bal9 was assessed by ^{19}F-NMR. Histone H3 Bal9 was incubated with 10 equivalents of the fluorinated diol leading to the emergence of a new peak. To confirm the identity of this new peak, several control experiments were conducted (**Figure 3.9**).

In the absence of boronoalanine, only one peak belonging to unbound ligand was observed (**Figure 3.9 A**). When a sample containing Histone H3 Bal9 and 10 equivalents of the fluorinated diol was incubated with a large excess of competitor (triol **4**, 200 equivalents), the peak indicating bound fluorinated diol disappeared (**Figure 3.9 A**). Likewise, no peak corresponding to bound ligand was observed in the absence of Bal (**Figure 3.9 B**). Finally, to confirm that the fluorinated ligand is indeed protein bound, a ^{19}F-NMR spectrum of Histone H3 Bal9 and 10 equivalents of fluorinated diol **11** was acquired using a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence with a 200 ms filter. CPMG experiments are commonly used in ligand-observed NMR experiments for the observation of ligand binding to protein targets.[53–55] Unbound ligands exhibit fast random rotational motions, which are described by short rotational correlation times (τ_c). The rotational correlation time is proportional to the molecular weight for a spherical molecule. Consequently, protein-ligand complexes exhibit much longer rotational correlation times.[56] The transversal relaxation time (T_2) is dependent on the rotational correlation time. Therefore, large molecules show fast transversal relaxation (short T_2).[54,57] CPMG experiments hereby exploit differences in transversal relaxation times (T_2) between bound and unbound ligand. While small and unbound ligands typically exhibit longer transversal relaxation times (longer T_2) leading to sharp peaks, transversal relaxation for protein-bound ligands is much faster (shorter T_2), usually giving rise to broadened peaks. Hence, a weakened signal is expected for a ligand bound to a protein target compared to an

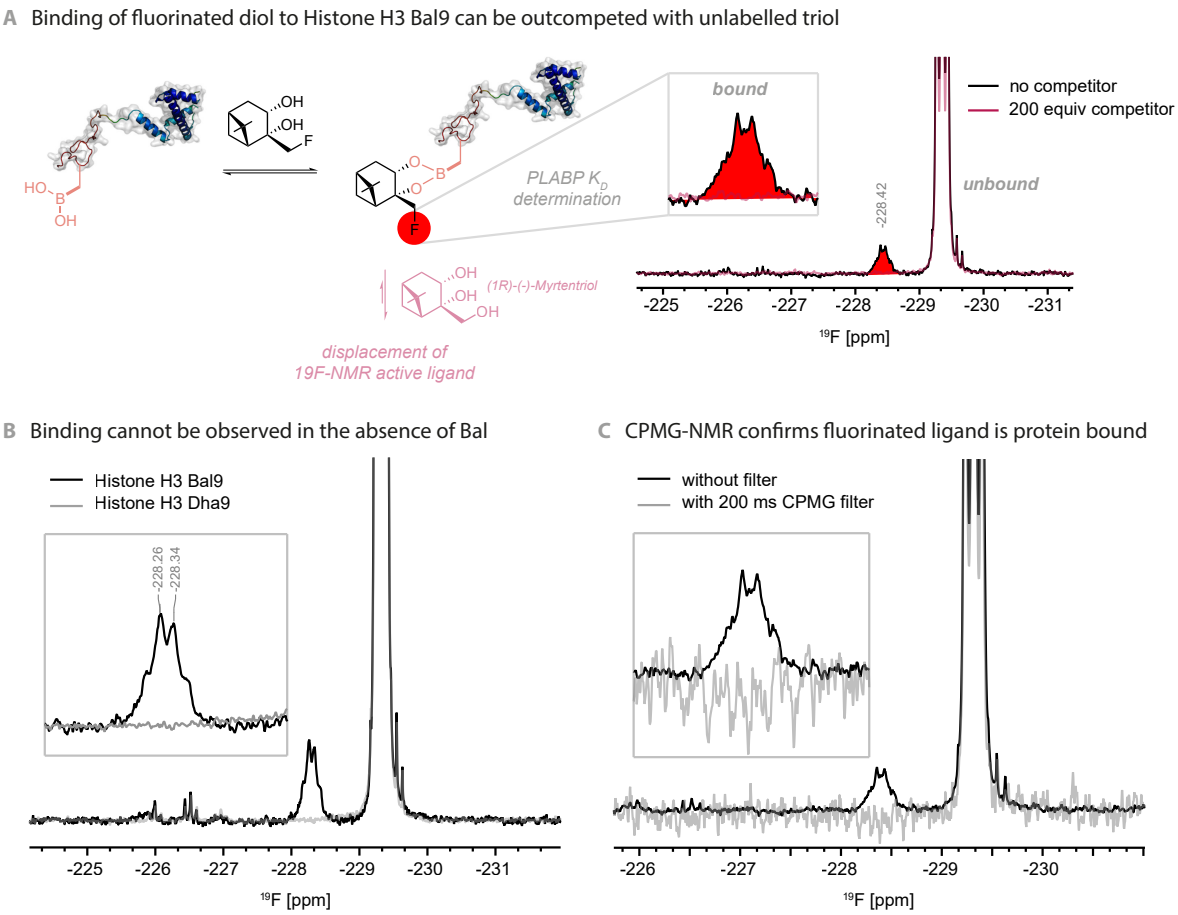


Figure 3.9: Confirmation of PLABP with fluorinated diol **11** (10 equiv): (**A**) ^{19}F-NMR reveals the presence of two peaks corresponding to protein bound and free ligand respectively. In the presence of a large excess of a competitor (triol **4**) complete displacement of the fluorinated ligand can be observed. (**B**) Binding of fluorinated ligand **11** to Histone H3 Bal9 is facilitated exclusively through Bal and no binding can be observed on non-borylated Histone H3 Dha9. (**C**) A ^{19}F-NMR spectrum acquired of a mixture of Histone H3 Bal9 and fluorinated diol **11** (10 equiv) using a CPMG pulse sequence with a 200 ms filter shows complete loss of signal for bound ligand indicating a protein bound species.

unbound ligand when a CPMG filter is employed.[53–55,58] Indeed, significant loss of signal was observed for the peak corresponding to bound diol compared to unbound diol when employing a 200 ms CPMG filter. This further confirms that fluorinated diol **11** is protein bound via engagement with boronoalanine (**Figure 3.9 C**). Having confirmed that the observed peak stems from fluorinated diol **11** binding to a Bal residue, the dependence of this interaction on Bal local environment and employed diol was investigated. Histones H3 Bal9 and H3 Bal10 were chosen to probe for the effect of the boronic acid local environment. In both cases, Bal sits within a portion of the flexible and intrinsically disordered *N*-terminal tail (having the sequence $\cdots \mathrm{T_6A_7R_8}\mathbf{K_9}\mathbf{S_{10}}\mathrm{T_{11}G_{12}G_{13}}\cdots$ in wild-type Histone H3) which boasts richly

functionalized side chains. These Lewis-basic side chain residues seemingly engage Bal in a site-dependent manner (**Section 3.3**), possibly giving rise to a handle for tuning of Wulff-type binding enhancement.[25] The two Histone mutants employed thus only differ by having Bal placed at two different, but neighbouring sites which might experience different degrees of Wulff-type binding enhancements through different proximal side chains. The previously synthesized fluorinated diols (fluorinated myrtenol derivative **11**, fluorinated ribose **12** and 2-Deoxy-2-fluoro-D-galactose (**13**)) were used as putative ligands to test this hypothesis.

In contrast to previous observations when boronic acid·diol interactions were studied on small molecule model compound (see above), ^{19}F-NMR binding studies on Histone H3 mutants H3 Bal9 and H3 Bal10 using both fluorinated myrtenol derivative **11** and fluorinated ribose **12** revealed a single peak for the bound ligand. This is in agreement with increased selectivity through Wulff-type enhancement mediated by proximal side chain residues. Indeed, interactions of small molecule boronic acids with *cis*-diols can lead to the formation of complex mixtures and oligomerization through the presence of multiple possible binding modes.[24,59]

Next, binding affinities between the two borylated Histone H3 derivatives and the selected ligands were measured by ^{19}F-NMR (**Table 3.2** and **Figure 3.10 A**). Strikingly, a clear ligand selectivity was observed with myrtenol derivative **11** showing a more than 10-fold lower K_D than ribose derivative **12** and galactose derivative **13** which did not exhibit any measurable binding at all. Moreover, binding affinities showed a clear site dependency. Shifting Bal from residue 10 to residue 9 as in Histone H3 Bal9 led to a 2-fold increase in binding affinity allowing for a low mM K_D for the interaction with terpene derivative **11** at neutral pH.

Table 3.2: ^{19}F-NMR based K_D determination of different protein and diol combinations at pH 7.0.

Entry	Protein + Diol	K_D [mM]
1	Histone H3 Bal9 + diol **11**	3.31±0.49
2	Histone H3 Bal9 + FDRibOMe (**12**)	43.2±7.0
3	Histone H3 Bal10 + diol **11**	6.22±0.55
4	Histone H3 Bal10 + FDRibOMe (**12**)	71.4±30.6
5	Histone H3 Bal9 + FDGal (**13**)	not detected

Samples were prepared and measured in triplicate for K_D determination. The amount of bound ligand was determined via the respective relative integrals. Protein concentrations were determined by A_{280} spectrophotometry and 10 equivalents of ligand were used in each case.

To confirm the dissociation constants derived from single concentration ^{19}F-NMR measurements with known ligand and protein concentrations, a NMR titration of Histone H3 Bal9 with fluorinated myrtenol derivative **11** was conducted (**Figure 3.10 B**). A K_D of 1.09 mM was determined, which is slightly lower than previously determined values but within the expected order of magnitude. A Hill coefficient of 1.14 was determined, which could be well within error for non-cooperative binding. This indicates binding of a single molecule of ligand to a single binding site as the predominant binding mode[60] and further confirms the formation of a boronic acid·diol adduct as postulated previously.

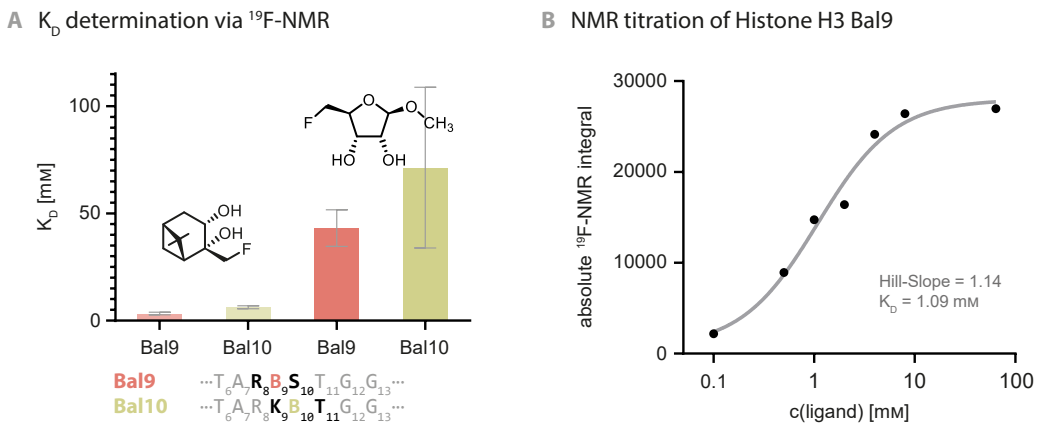


Figure 3.10: Determination of dissociation constants for PLABP with fluorinated diols. (**A**) Dissociation constants are strongly dependent not only on the specific diol but also on the borylation site. Comparison of binding affinities for Histone H3 Bal9 and Histone H3 Bal10 reveals pronounced site-specific modulation of binding. (**B**) ^{19}F-NMR titration of Histone H3 Bal9 with myrtenol derived diol **11**.

Utilizing Diol Binding to Analyse Diastereoselectivity

Having established binding affinities and an 'ideal' protein·diol combination, exploiting this reactivity for subsequent applications became possible. For the binding of both, ribose derivative **12** and myrtenol derived diol **11**, to model protein Histone H3, a single peak was detected in the ^{19}F-NMR spectrum. Given the likely presence of *R* and *S* isomers of Bal for the non-diastereoselective borylation of Dha residues in proteins, two peaks would be expected (if the difference in chemical shift of the two complexes formed is sufficient). As only one, albeit broadened, peak was observed, the synthesis of a fluorinated diol that enables differentiation of both diastereomers by ^{19}F-NMR was envisioned. Such a diol would

require the ability for tight binding and an organofluorine moiety that experiences sufficient change in chemical environment upon binding to each diastereomer. Additionally, multiple chemically equivalent fluoride atoms would be beneficial to increase signal. A trifluoromethoxy derivative (**14**) of myrtenol derived diol **11** was anticipated to fulfil these requirements, possessing a trifluoromethyl group (allowing for a strong signal) in direct proximity of an additional oxygen atom possibly enabling the formation of a dative B—O bond (**Figure 3.11**). This trifluoromethoxy derivative was synthesized following the general route developed for the synthesis of fluorinated myrtenol derived diol **11** (**Scheme 3.4**), but introducing a trifluoromethyl ether instead of deoxyfluorination using a method developed by Qing and co-workers.[61]

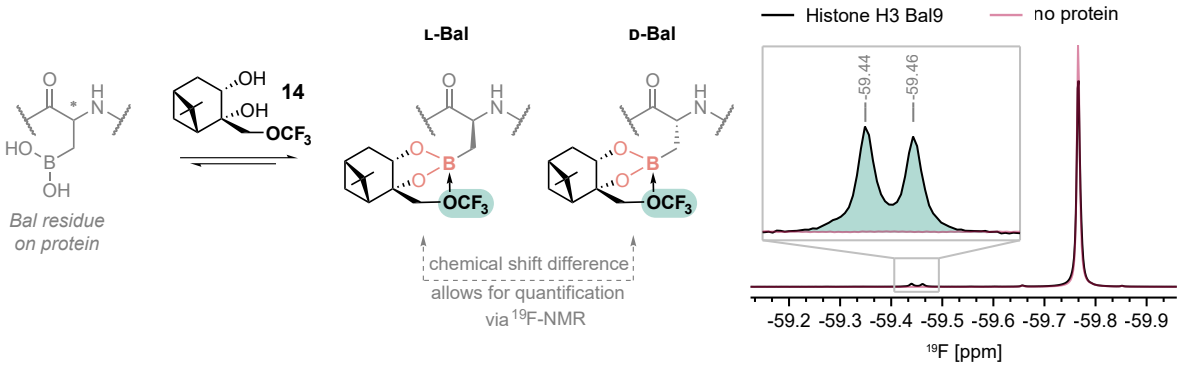


Figure 3.11: Proof-of-concept for the determination of the diastereomeric ratio of D- and L-Bal after protein borylation using a ^{19}F-NMR diol probe. Two distinct peaks are obtained after treatment of model protein Histone H3 Bal9 (black) with excess ligand (10 equiv; myrtenol derived diol **14**). In the absence of borylated Histone, only unbound ligand is observed (pink).

Indeed, two singlets were observed upon ^{19}F-NMR analysis in the presence of Histone H3 Bal9. Despite these peaks only showing a difference in chemical shift of 0.02 ppm, the determination of the diastereomeric ratio was possible indicating an approximately one-to-one mixture (**Figure 3.11**). While different binding affinities for L-Bal and D-Bal are a theoretical possibility, the observed ratio is within expectation. Nevertheless, further validation would be required to practically apply this methodology. The ability to readily determine the diastereoselectivity – or lack thereof – of a chemical reaction on protein could be extremely useful. Not only would it enable one to measure site-dependent induction of chirality, but it would also aid in the development of stereoselective reactions on proteins.[62]

In summary, the sequence-dependent modulation of binding affinity makes future applications in the design of efficient and specific *de novo* binding sites and even the construction of proteins featuring dual or multiple binding motifs readily imaginable. In fact, both Annexin V Bal316 and PstS Bal197 retained their native binding affinities towards their natural ligands phosphatidyl serine[63] and inorganic phosphate[64] respectively. While the properties of these dual binders was not further explored here, the utility of at-will engineering of bispecific binding can be readily conceived.[65]

3.4.2 Interacting with Biological Systems

Synthesis of an Artificial Glycan Sensor

Gaining an understanding of boronoalanine's ability to engage with *cis*-diols, such as found in saccharides, immediately raised the question whether this unique ability can be exploited in biological systems.

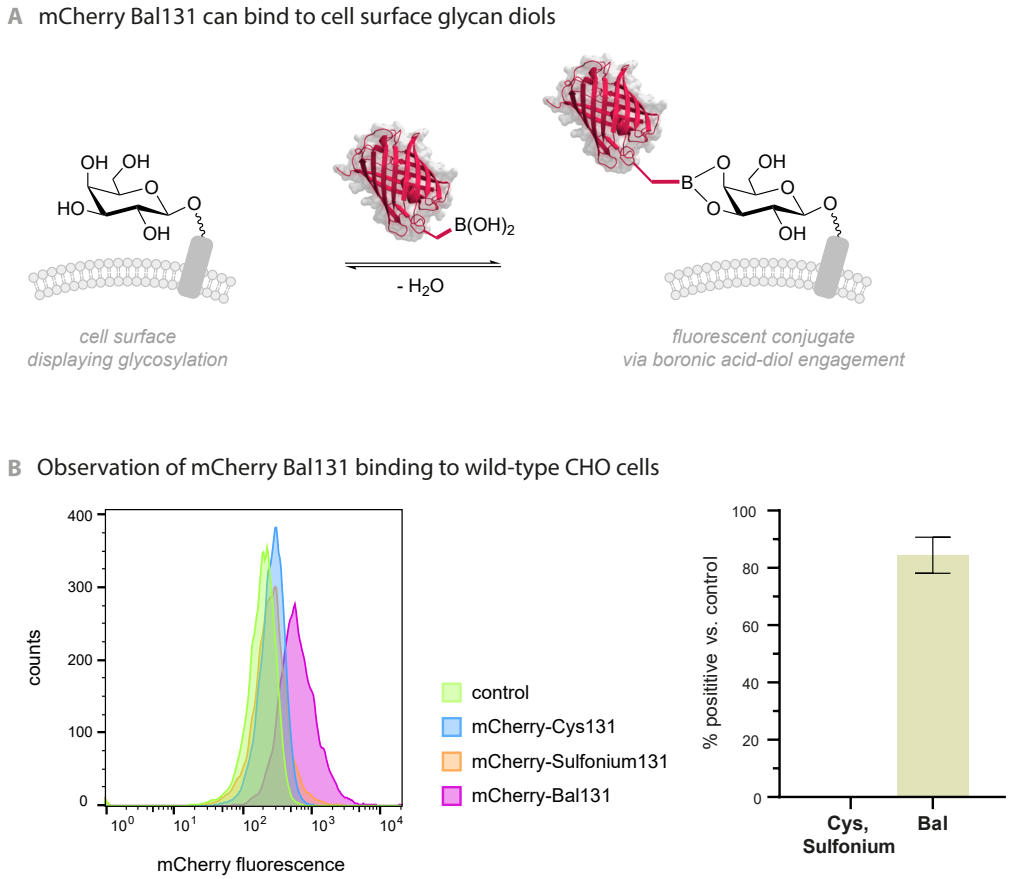


Figure 3.12: Borylation of mCherry yields a novel fluorescent protein sensor for the detection of cell surface diols. (**A**) mCherry Bal131 can bind to cell surface glycans to form fluorescent adducts via boronic acid·diol engagement. (**B**) This binding behaviour can be utilized to detect cell surface glycosylation via flow cytometry. While no significant difference in fluorescence was observed for treatment with mCherry Cys131 (blue) or mCherry Sulfonium131 (orange) compared to the untreated control, mCherry Bal131 treated cells exhibited significant fluorescence (84.4±6.3% positive vs. control) as detected by averaging the values obtained from data analysis using both the Super-Enhanced D_{max} Subtraction (SED) and the Overton cumulative histogram subtraction algorithms as implemented in FlowJo.

In fact, glycans are ubiquitously present on cell surfaces which could provide a handle for the interaction with borylated proteins.[66] Consequently, the binding of the borylated

red fluorescent protein mCherry Bal131 to a mammalian cell surface via boronic acid·diol interactions was investigated. Treatment of Chinese hamster ovary (CHO) cells with mCherry Bal131 and subsequent analysis by flow cytometry revealed binding of the borylated mutant (84.4±6.3% positive vs. untreated control), whereas no binding could be observed for either mCherry Cys131 or mCherry Sulfonium131 (<5% positive vs. untreated control) (**Figure 3.12**). This confirms that the presence of a single boronic acid allows for de novo binding function. Borylation therefore enabled the synthesis of a novel and artificial, fluorescent protein-based sensor for the detection of cell surface glycosylation.

Modulation of Proteolytic Stability

While the binding of Bal to diols could be readily characterized by ^{19}F-NMR using a range of fluorinated diols, its interactions with exogenous ligands featuring just an alcohol were yet to be determined. However, observations of intramolecular boronate formation (**Chapter 3.4**) suggested binding of simple alcohols, such as serine residues, if they are located in suitable proximity. The formation of esters with protein side chain alcohols that assume a suitable confirmation is indeed the typical mechanism of action of boronic acid based enzyme inhibitors.[15,17,18] While this has been investigated in depth for protein-small molecule systems, there are currently no literature records on the effect of a protein-anchored boronic acid on its interactions with other proteins. To investigate this question, preSUMO1 Bal51 and its interactions with different proteases was chosen as a model system.

The covalent attachment of small ubiquitin-like modifier (SUMO) to cellular proteins via an isopeptide bond – SUMOylation – is an important post-translational modification involved in the regulation of many biological processes in all eukaryotic kingdoms. SUMOylation is governed by a range of different enzymes responsible for SUMO maturation, activation and ligation.[68] First, pre-SUMO maturation is achieved by a carboxy-terminal hydrolase such as SENP1.[69] The resulting *C*-terminal diglycine motif of matured SUMO is then adenylated by SUMO activating enzyme (SAE1) and subsequently transferred to its active site cysteine forming a thioester bond with the *C*-terminal glycine residue. This thioester is next transferred to the active site cysteine of SUMO conjugating enzyme such as Ubc9. Finally, an isopeptide bond between the *C*-terminus of SUMO and an ε-amino group of a suitable acceptor substrate

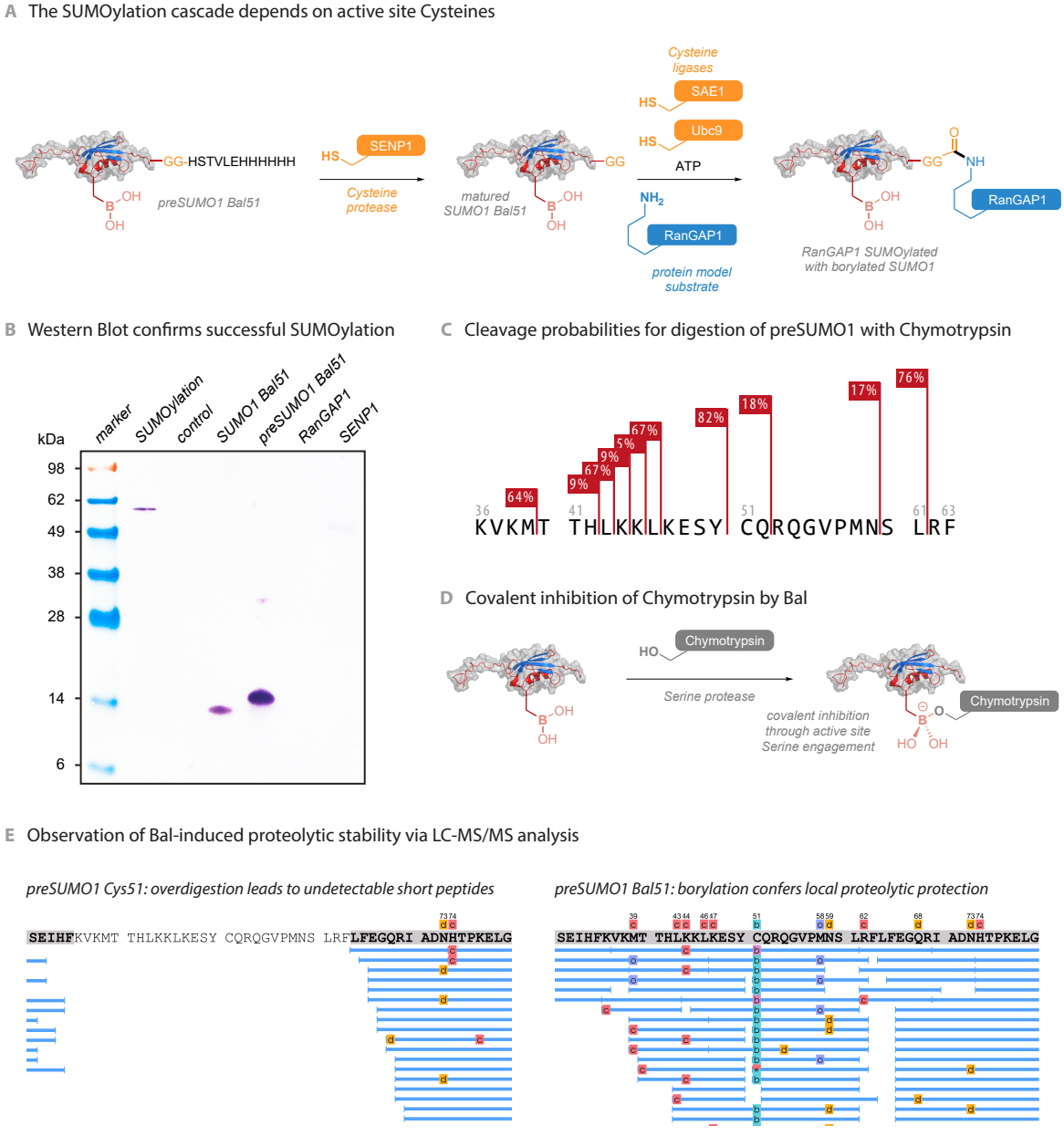


Figure 3.13: PLABP enables the engineering of selective and local proteolytic stability. (**A**) The SUMOylation cascade heavily depends on enzymes (proteases and ligases) featuring an active site cysteine residue. Borylation of SUMO does not affect these reactions. (**B**) Western Blot analysis confirms successful SUMOylation of a RanGAP1 fragment utilizing preSUMO1 Bal51. (**C**) The 28-residue motif (residues 36-63) within which Bal was placed contains eleven putative cleavage sites for chymotrypsin rendering this section highly prone to proteolytic cleavage. Cleavage probabilities were calculated using the ExPASy PeptideCutter tool.[67] (**E**) Digestion of preSUMO1 Cys51 and preSUMO1 Bal51 with chymotrypsin and subsequent LC-MS/MS analysis revealed proteolytic protection in the presence of Bal. In the absence of Bal the sequence of 28-residue motif (residues 36-63) was not covered due to overdigestion leading to undetectable short peptides.

is formed completing the SUMOylation cascade (**Figure 3.13 A**).[70,71]

SUMOylation was attempted using preSUMO1 Bal51, which was maturated with SENP1. Mature SUMO1 Bal51 was then conjugated to a RanGAP1 fragment acting as model acceptor using SAE1 and Ubc9 (**Figure 3.13 A**). Successful SUMO maturation by SENP1 (full conversion) was observed by mass spectrometry and SUMOylation of the RanGAP1 fragment was confirmed by Western Blot using an anti-SUMO primary antibody (**Figure 3.13 B**). Despite the presence of a boronic acid on SUMO, the reaction cascade proceeded smoothly and all SUMO1 Bal51 present in the reaction mixture was consumed. This is in agreement with the literature indicating that cysteine proteases are typically not susceptible to inhibition by boronic acids.[18]

Next, the influence of borylation on the proteolytic stability towards the serine protease chymotrypsin was investigated. In contrast to cysteine proteases, serine proteases are known targets for small molecule boronic acid inhibitors.[15,18] In preSUMO Bal51, boronoalanine is placed strategically within a 28-residue motif (residues 36-63) containing eleven putative cleavage sites for chymotrypsin (**Figure 3.13 C**). Both wild-type preSUMO1 and preSUMO1 Bal51 were subjected to chymotrypsin digestion and subsequent LC-MS/MS analysis (**Figure 3.13 D** and **Figure 3.13 E**). In the case of wild-type preSUMO1, the central region most susceptible to proteolysis was unobservable by LC-MS/MS. This indicates extensive overdigestion leading to undetectable short peptide fragments. Typically, short peptides of six or less residues are not detected in LC-MS/MS experiments (commonly detected peptides are 7-35 residues in length).[72] Conversely, MS/MS analysis of preSUMO1 Bal51 provided full sequence coverage in the presence of masking triol **4** (**Figure 3.13 E**). This is in agreement with a drastically increased resistance to serine protease hydrolysis. Borlyation thus confers local proteolytic protection in proximity of the introduced boronic acid residue.

3.5 Reactive PLABP

Previous examples of protein-protein Lewis acid·base pairing (PLABP) utilized the transient nature of boronic acid·Lewis base interactions to probe or convey biological structure and function. In contrast, reactive PLABP could serve as a permanent reporter of these transient contacts. The irreversible reaction with a suitable partner, such as reactive oxygen species

(ROS) would leave a permanent mark that can be interpreted at a later stage. While ROS are known to play a crucial role in many biological processes, including signalling, disease and aging, their exact roles in many processes remain elusive.[73,74] Owing to the importance of ROS in biology, many different methods have been developed for their detection.[75] Boronoalanine could serve as a direct reporter of a protein's interaction with reactive oxygen species. In fact, hydrogen peroxide was previously shown to efficiently oxidize Bal to the corresponding Ser residue (**Chapter 2**).

Treatment of different borylated proteins with hydrogen peroxide (H_2O_2) revealed the protein and site dependent requirement of varying H_2O_2 concentrations to achieve full oxidation of Bal to the corresponding alcohol (**Table 3.3**). While for Histone H3 Bal10 full conversion to Histone H3 Ser10 was observed at a hydrogen peroxide concentration of 5 mM, moving to Histone H3 Bal9 led to a much higher tolerance towards oxidation with 20 mM H_2O_2 required for full conversion using the same buffer. Neither Npβ Bal61 nor PstS Bal197 showed any sign of oxidation at this hydrogen peroxide concentration.

Table 3.3: Reactive PLABP: Concentration of H_2O_2 required for oxidation of Bal to Ser correlates with rASA at the site of interest.

Entry	Protein	rASA[#] [%]	c(H_2O_2) for full conversion to Ser*
1	Histone H3 Bal10	108	5 mM
2	Histone H4 Bal16	60	20 mM
3	Histone H3 Bal9	34	20 mM
4	Npβ Bal61	25	decomposition at 20 mM
5	PstS Bal197	0	no oxidation at 20 mM

[#]solvent-accessible surface area of the Cys mutant relative to Gly-Cys-Gly calculated with a probe size of 2.8 Å. * Reaction in NaP_i (100 mM, pH 7.0, 3 M Gdn·HCl) for 10 min.; H_2O_2 concentrations tested: 1 mM, 5 mM, 20 mM. Full conversion to Ser is considered achieved once no masses belonging to Bal can be detected.

Predictably, the concentration of ROS required for oxidation of Bal to Ser correlated well with the accessibility of the borylated site as judged by the relative solvent accessible surface area (rASA). Histone H3 Bal10 exhibited the highest relative solvent accessible surface area (solvent accessible surface area of the Cys mutant relative to Gly-Cys-Gly as determined using the Lee & Richards algorithm and varying probe size[76,77]) of 71 – 108% for probe sizes of 1.0 – 2.8 Å. The relative solvent accessibility for Histone H3 Bal9 was 34 – 45%, while 29 – 25%

were calculated for Npβ (**Figure 3.14** and **Table 3.3**). The protein mutants selected for this

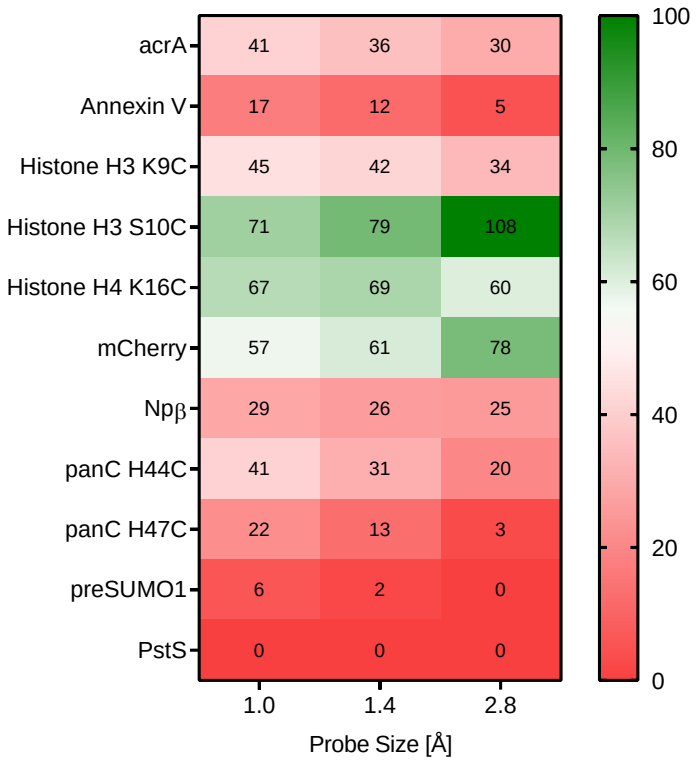


Figure 3.14: Heat map for the solvent-accessible surface area relative to Gly-Cys-Gly for Cys residues at the borylation site in different proteins (in %) as calculated using the Lee & Richards algorithm[76] on PDB protein structures with NACCESS atom radii. Calculations were conducted using FreeSASA.[77]

study covered a wide range of possible accessibilities including the most accessible (Histone H3 Bal10), mutants of intermediate accessibility (Histone H4 Bal16, Histone H3 Bal9 and Npβ Bal61) as well as the least accessible mutant (PstS Bal196) (**Figure 3.14** and **Table 3.4**).

This allows for the conclusion that reactive PLABP can not only be applied to create a permanent record of the contact with reactive oxygen species but also enables the estimation of the accessibility of a Bal residue via the hydrogen peroxide concentration required for complete oxidation. Furthermore, this could also be used to engineer ROS sensing proteins of different sensitivity, solely based on the careful selection of the borylation site (and hence accessibility of the boronic acid). It should be noted that the hydrogen peroxide concentrations found in cells are typically in nanomolar range,[78] whereas much higher concentrations (and short reaction times) were employed in this proof-of-principle study. As (very slow) oxidation of accessible Bal to Ser was observed under ambient atmosphere at room temperature, the

oxidation of Bal by cellular concentrations of hydrogen peroxide seems nevertheless feasible using adjusted reaction times. The possibility of site-selective deuteration upon installation of Bal (**Chapter 2.4.3**) additionally could allow for discrimination in a biological context between native protein and artificially introduced protein carrying a deuterated Ser residue (stemming from *d*-Bal oxidation).

Table 3.4: Solvent-accessible surface area (ASA) and relative solvent-accessible surface area (rASA) for all proteins subjected to borylation (compare **Chapter 2.4**)

Protein	Site	PDB ID	solvent-accessible surface area [Å²]			rel. solvent-accessible surface area [%]		
			1.0 Å	1.4 Å	2.8 Å	1.0 Å	1.4 Å	2.8 Å
acrA	C127	2FM1	68.45	60.18	50.03	41	36	30
Annexin V	C316	1HVD	28.09	19.56	8.05	17	12	5
Histone H3	C9	1KX5	74.88	70.47	56.28	45	42	34
Histone H3	C10	1KX5	118.13	131.56	179.69	71	79	108
Histone H4	C16	1KX5	112.36	115.57	100.63	67	69	60
mCherry	C131	4ZIN	95.41	102.49	129.64	57	61	78
Npβ	C61	2J8K	47.71	44.15	42.00	29	26	25
panC	C44	1N2E	68.73	52.23	33.42	41	31	20
panC	C47	1N2E	36.45	21.19	4.61	22	13	3
preSUMO1	C51	1A5R	9.26	3.79	0.00	6	2	0
PstS	C197	1A40	0.81	0.00	0.00	0	0	0

The solvent accessible surface area was calculated for Cys residues in different proteins using the Lee & Richards algorithm[76] on PDB protein structures with NACCESS atom radii. Calculations were conducted using FreeSASA.[77] Relative solvent accessible surface areas were calculated relative to Gly-Cys-Gly and are given in %.

3.6 Conclusion

In the vast majority of cases, reversible protein−ligand binding is mediated by hydrophobic, hydrogen bonding or π-stacking interactions.[79] The introduction of anchored boronic acid into proteins via direct C−B bond formation creates *de novo* binding sites allowing for reversible, yet covalent Lewis acid·base pairing. This facilitates the binding of endogenous

as well as exogenous ligands. The presence of endogenous ligands provided by Lewis-basic side-chain residues (effectively leading to a Wulff-type boronic acid) allows for sequence-specific affinity modulation. Furthermore, the possibility to engineer *de novo* binding enables the deliberate adjustment of protein stability. While thermal stability can be increased or decreased depending on boronoalanine's intramolecular interactions, increased proteolytic stability is provided by the local inhibition of (serine)proteases. At the same time, engagement of Bal is sufficiently selective that it does not interfere with natural protein function. Furthermore, the installation of artificial binding sites into functional proteins, such as mCherry, enabled the engineering of dual function. In the case of mCherry, the natural fluorescent properties were retained, while a *de novo* glycan binding site was created. When preSUMO1 was borylated, it preserved its natural reactivity towards the SUMOylation cascade. Here, the presence of a boronic acid that does not interfere with natural function does not only confer additional proteolytic resistance, but could also be applied for ROS sensing via reactive PLABP. The utility of such proteins engineered towards artificial bi-functionality is readily conceivable, yet has only been sparsely explored. In this chapter, a few examples illustrating this utility were developed as a proof-of-concept. Furthermore, it was shown that these artificial binding modes can be exploited to selectively interfere with complex biological systems – possibly providing a valuable tool to answer biological or biochemical questions.

References

[1] D. G. Hall in *Boronic Acids*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2011**, pp. 1–133.

[2] J. N. Payette, H. Yamamoto in *Boronic Acids*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2011**, pp. 551–590.

[3] D. G. Hall, *Chem. Soc. Rev.* **2019**, *48*, 3475–3496.

[4] M. S. Taylor, *Acc. Chem. Res.* **2015**, *48*, 295–305.

[5] S. Manhas, Y. C. Lin, G. Wang, L. T. Kyne, M. S. Taylor, *Green Chem.* **2019**, *21*, 5363–5367.

[6] N. Shimada, Y. Nakamura, T. Ochiai, K. Makino, *Org. Lett.* **2019**, *21*, 3789–3794.

[7] K. Oshima, E.-i. Kitazono, Y. Aoyama, *Tetrahedron Lett.* **1997**, *38*, 5001–5004.

[8] K. Oshima, Y. Aoyama, *J. Am. Chem. Soc.* **1999**, *121*, 2315–2316.

[9] X. Wu, Z. Li, X.-X. Chen, J. S. Fossey, T. D. James, Y.-B. Jiang, *Chem. Soc. Rev.* **2013**, *42*, 8032.

[10] X. Zhang, G. Liu, Z. Ning, G. Xing, *Carbohydr. Res.* **2017**, *452*, 129–148.

[11] G. Fang, H. Wang, Z. Bian, J. Sun, A. Liu, H. Fang, B. Liu, Q. Yao, Z. Wu, *RSC Adv.* **2018**, *8*, 29400–29427.

[12] R. Nishiyabu, Y. Kubo, T. D. James, J. S. Fossey, *Chem. Commun.* **2011**, *47*, 1124–1150.

[13] T. L. Halo, J. Appelbaum, E. M. Hobert, D. M. Balkin, A. Schepartz, *J. Am. Chem. Soc.* **2009**, *131*, 438–439.

[14] D. H.-C. Chou, M. J. Webber, B. C. Tang, A. B. Lin, L. S. Thapa, D. Deng, J. V. Truong, A. B. Cortinas, R. Langer, D. G. Anderson, *Proc. Natl. Acad. Sci.* **2015**, *112*, 2401–2406.

[15] R. Smoum, A. Rubinstein, V. M. Dembitsky, M. Srebnik, *Chem. Rev.* **2012**, *112*, 4156–4220.

[16] P. C. Trippier, C. McGuigan, *MedChemComm* **2010**, *1*, 183.

[17] S. J. Baker, J. W. Tomsho, S. J. Benkovic, *Chem. Soc. Rev.* **2011**, *40*, 4279.

[18] D. B. Diaz, A. K. Yudin, *Nat. Chem.* **2017**, *9*, 731–742.

[19] J. Plescia, N. Moitessier, *Eur. J. Med. Chem.* **2020**, *195*, 112270.

[20] P. Bonvini, E. Zorzi, G. Basso, A. Rosolen, *Leukemia* **2007**, *21*, 838–842.

[21] A. L. Goldberg, *J. Cell Biol.* **2012**, *199*, 583–588.

[22] J. Schrader, F. Henneberg, R. A. Mata, K. Tittmann, T. R. Schneider, H. Stark, G. Bourenkov, A. Chari, *Science* **2016**, *353*, 594–598.

[23] F. Duval, T. A. van Beek, H. Zuilhof, *Analyst* **2015**, *140*, 6467–6472.

[24] J. A. Peters, *Coord. Chem. Rev.* **2014**, *268*, 1–22.

[25] G. Wulff, M. Lauer, H. Böhnke, *Angew. Chem. Int. Ed.* **1984**, *23*, 741–742.

[26] G. Wulff, *Pure Appl. Chem.* **1982**, *54*, 2093–2102.

[27] L. Liang, Z. Liu, *Chem. Commun.* **2011**, *47*, 2255.

[28] G. F. Whyte, R. Vilar, R. Woscholski, *J. Chem. Biol.* **2013**, *6*, 161–174.

[29] L. Zhu, S. H. Shabbir, M. Gray, V. M. Lynch, S. Sorey, E. V. Anslyn, *J. Am. Chem. Soc.* **2006**, *128*, 1222–1232.

[30] M. Dowlut, D. G. Hall, *J. Am. Chem. Soc.* **2006**, *128*, 4226–4227.

[31] M. Bérubé, M. Dowlut, D. G. Hall, *J. Org. Chem.* **2008**, *73*, 6471–6479.

[32] F. L. Rock, W. Mao, A. Yaremchuk, M. Tukalo, T. Crépin, H. Zhou, Y. K. Zhang, V. Hernandez, T. Akama, S. J. Baker, J. J. Plattner, L. Shapiro, S. A. Martinis, S. J. Benkovic, S. Cusack, M. R. Alley, *Science* **2007**, *316*, 1759–1761.

[33] L. Ren, Z. Liu, Y. Liu, P. Dou, H.-Y. Chen, *Angew. Chem. Int. Ed.* **2009**, *48*, 6704–6707.

[34] T. Řezanka, K. Sigler in *Stud. Nat. Prod. Chem. Vol. 35*, C, **2008**, pp. 835–921.

[35] T. Hemscheidt, M. P. Puglisi, L. K. Larsen, G. M. Patterson, R. E. Moore, J. L. Rios, J. Clardy, *J. Org. Chem.* **1994**, *59*, 3467–3471.

[36] R. Hütter, W. Keller-Schien, F. Knüsel, V. Prelog, G. C. Rodgers, P. Suter, G. Vogel, W. Voser, H. Zähner, *Helv. Chim. Acta* **1967**, *50*, 1533–1539.

[37] M. A. O'Neill, T. Ishii, P. Albersheim, A. G. Darvill, *Annu. Rev. Plant Biol.* **2004**, *55*, 109–139.

[38] X. Chen, S. Schauder, N. Potier, A. Van Dorsselaer, I. Pelczer, B. L. Bassler, F. M. Hughson, *Nature* **2002**, *415*, 545–549.

[39] W. Moreira, D. B. Aziz, T. Dick, *Front. Microbiol.* **2016**, *7*, 1–7.

[40] A. M. Burja, B. Banaigs, E. Abou-Mansour, J. Grant Burgess, P. C. Wright, *Tetrahedron* **2001**, *57*, 9347–9377.

[41] V. M. Dembitsky, A. A. A. Al Quntar, M. Srebnik, *Chem. Rev.* **2011**, *111*, 209–237.

[42] H. Steinberg, D. L. Hunter, *Ind. Eng. Chem.* **1957**, *49*, 174–181.

[43] S. J. Kaldas, T. Rogova, V. G. Nenajdenko, A. K. Yudin, *J. Org. Chem.* **2018**, *83*, 7296–7302.

[44] J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, *J. Am. Chem. Soc.* **2018**, *140*, 15997–16000.

[45] M. W. Vetting, S. S. Hegde, K. Z. Hazleton, J. S. Blanchard, *Protein Sci.* **2007**, *16*, 755–760.

[46] J. Axthelm, S. H. Askes, M. Elstner, G. Upendar Reddy, H. Görls, P. Bellstedt, A. Schiller, *J. Am. Chem. Soc.* **2017**, *139*, 11413–11420.

[47] S. Jinna, J. Finch, *Drug Des. Devel. Ther.* **2015**, *9*, 6185–6190.

[48] E. Seiradake, W. Mao, V. Hernandez, S. J. Baker, J. J. Plattner, M. R. Alley, S. Cusack, *J. Mol. Biol.* **2009**, *390*, 196–207.

[49] M. R. Alley, S. J. Baker, K. R. Beutner, J. Plattner, *Expert Opin. Investig. Drugs* **2007**, *16*, 157–167.

[50] A. Lakshminarayanan, M. Richard, B. G. Davis, *Nat. Rev. Chem.* **2018**, *2*, 148–159.

[51] B. P. Rempel, S. G. Withers, *Org. Biomol. Chem.* **2014**, *12*, 2592–2595.

[52] X. G. Li, S. Dall'Angelo, L. F. Schweiger, M. Zanda, D. O'Hagan, *Chem. Commun.* **2012**, *48*, 5247–5249.

[53] E. H. Mashalidis, P. Śledź, S. Lang, C. Abell, *Nat. Protoc.* **2013**, *8*, 2309–2324.

[54] D. R. Calabrese, C. M. Connelly, J. S. Schneekloth, *Methods Enzymol.* **2019**, *623*, 131–149.

[55] A. K. Urick, L. P. Calle, J. F. Espinosa, H. Hu, W. C. Pomerantz, *ACS Chem. Biol.* **2016**, *11*, 3154–3164.

[56] A. D. Gossert, W. Jahnke, *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *97*, 82–125.

[57] M. Čuperlović-Culf in *NMR Metabolomics Cancer Res.* Elsevier, **2013**, pp. 139–213.

[58] S. Meiboom, D. Gill, *Rev. Sci. Instrum.* **1958**, *29*, 688–691.

[59] P. Metola, B. M. Chapin, E. V. Anslyn in *Boron Sensing, Synthesis and Supramolecular Self-Assembly*, **2015**, pp. 44–60.

[60] J. N. Weiss, *FASEB J.* **1997**, *11*, 835–841.

[61] J. B. Liu, X. H. Xu, F. L. Qing, *Org. Lett.* **2015**, *17*, 5048–5051.

[62] R. Meledin, S. M. Mali, S. K. Singh, A. Brik, *Org. Biomol. Chem.* **2016**, *14*, 4817–4823.

[63] M. van Engeland, L. J. Nieland, F. C. Ramaekers, B. Schutte, C. P. Reutelingsperger, *Cytometry* **1998**, *31*, 1–9.

[64] Y. Qi, Y. Kobayashi, F. M. Hulett, *J. Bacteriol.* **1997**, *179*, 2534–2539.

[65] B. Husain, D. Ellerman, *BioDrugs* **2018**, *32*, 441–464.

[66] A. Varki, *Glycobiology* **2017**, *27*, 3–49.

[67] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, A. Bairoch in *Proteomics Protoc. Handb.* **2005**.

[68] S. Müller, C. Hoege, G. Pyrowolakis, S. Jentsch, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 202–210.

[69] J. Mikolajczyk, M. Drag, M. Békés, J. T. Cao, Z. Ronai, G. S. Salvesen, *J. Biol. Chem.* **2007**, *282*, 26217–26224.

[70] A. Flotho, F. Melchior, *Annu. Rev. Biochem.* **2013**, *82*, 357–385.

[71] J. R. Gareau, C. D. Lima, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 861–871.

[72] D. L. Swaney, C. D. Wenger, J. J. Coon, *J. Proteome Res.* **2010**, *9*, 1323–1329.

[73] B. D'Autréaux, M. B. Toledano, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 813–824.

[74] H. Sies, D. P. Jones, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 363–383.

[75] Y. Zhang, M. Dai, Z. Yuan, *Anal. Methods* **2018**, *10*, 4625–4638.

[76] B. Lee, F. Richards, *J. Mol. Biol.* **1971**, *55*, 379.

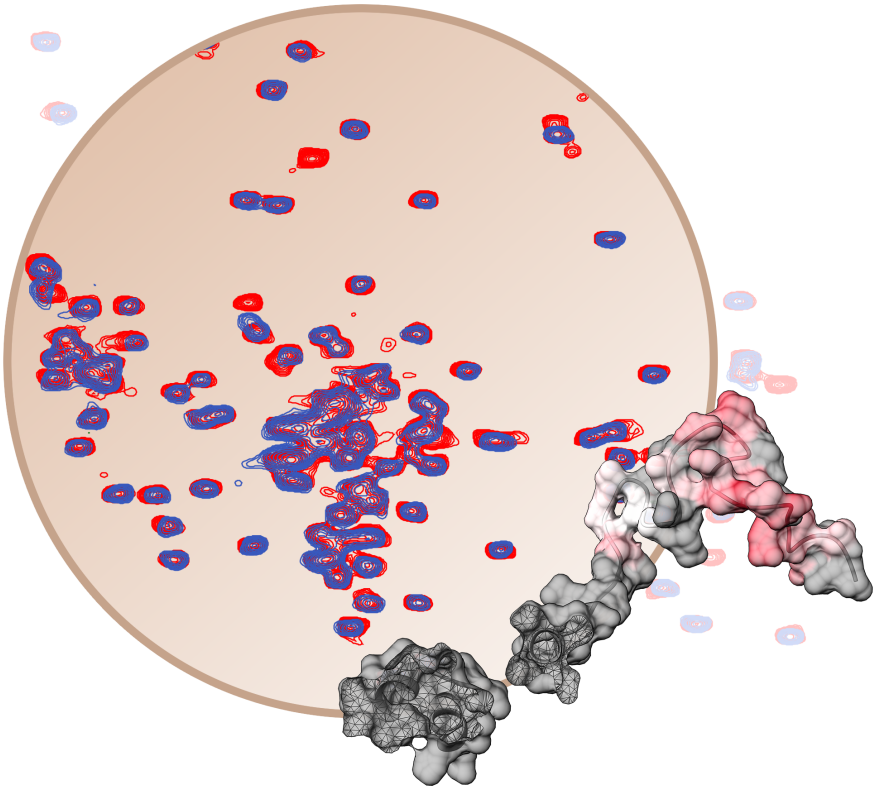
[77] S. Mitternacht, *F1000Research* **2016**, *5*, 1–11.

[78] N. Di Marzo, E. Chisci, R. Giovannoni, *Cells* **2018**, *7*, 156.

[79] R. Ferreira de Freitas, M. Schapira, *MedChemComm* **2017**, *8*, 1970–1981.

Chapter 4

Application of Bal as a NMR Structural Probe



It is structure that we look for

whenever we try to understand anything.

— *Linus Pauling*, The Place of Chemistry in the Integration of the Sciences (1949)

Chapter 4

Application of Bal as a NMR Structural Probe

For a long time, protein function was believed to depend on static, ordered and three-dimensional structure programmed by the amino acid sequence.[1] Following this structure-function paradigm, the possible presence of disordered and thus apparently non-functional proteins *in vivo* was debated and the observation of unstructured domains was suggested to be an *in vitro* artefact.[2] Only recent advances in NMR spectroscopy facilitated a shift of this paradigm and allowed for the detailed characterization of intrinsically disordered proteins.[1,3]

4.1 Methods for the Detection of Long-Range Contacts in Intrinsically Disordered Proteins

Several NMR approaches have been described for the structural characterization of intrinsically disordered proteins (IDPs). NMR is specifically suitable for these investigations as IDPs can adopt an ensemble of different states in solution and NMR spectroscopy allows for the measurement of ensemble-averaged structural parameters.[3,4] This enables the determination of structures that can not be determined by x-ray crystallography. Multiple NMR techniques are used routinely for the detection of long-range contacts including long-range nuclear Overhauser effects (NOEs) as well as techniques that depend on a paramagnetic system, such as residual dipolar couplings (RDC), pseudo contact shifts (PCS) and paramagnetic relaxation

enhancement (PRE).[3–6] Long-range NOE experiments are typically used in well-structured proteins and their application to IDPs is challenging as transient contacts can be too short lived to be observed reliably.[3,7] Paramagnetic relaxation enhancement (PRE) is widely used for the detection of long-range contacts making use of the interaction between the spin of an unpaired electron and nuclear spins.[3] Typically, an unpaired electron is introduced by chemical conjugation (often via cysteine) of a nitroxide spin label or a chelated paramagnetic metal ion.[6] As spins in spacial proximity to the spin label relax more efficiently, residues experiencing long-range contacts can be detected by resonance broadening.[3] Depending on the spin label, contacts within a distance of 10—35 Å from the label can usually be detected.[8] Drawbacks of this methodology include the required introduction of a (relatively) bulky spin label that can potentially interfere with protein structure[6] and the complete loss of any structural information on residues adjacent to the spin label as these are permanently in its proximity always leading to resonance broadening.[3]

Consequently, a minimally-sized and readily introduced structural NMR probe that allows for the detection of transient long-range contacts as well as for the detection of structural features within its spacial proximity could be of great utility.

4.2 NMR Investigations on Small Molecule Model Compound

NMR investigations were conducted on a small molecule model compound to gain a deeper understanding of intramolecular binding. Such a simplified system was envisioned to aid the understanding of intramolecular interactions on a more complex protein system.

When NMR spectra of boronoalanine model compound Ac-Bal-NHBn (**3**) were recorded under basic conditions in aqueous buffer, the formation of a novel species could be observed. Interactions of Bal with the peptidic backbone are readily conceivable (**Chapter 3.2.2**, **Scheme 3.3**) and indeed, the formation of cyclic boronates involving adjacent carbonyl groups was previously reported.[9,10] Boronic acids are also known for their ability to form dimers in crystalline solids as well as in solution.[11,12] First, dimerization was precluded as a reason for the appearance of the novel species in the NMR spectra. As the formation of this species was found not to be concentration dependent, dimerization could be ruled out for its origin (**Table 4.1**). To further investigate the proposed formation of cyclic boronates with

either of the two carbonyl groups, more extensive NMR studies were necessary.

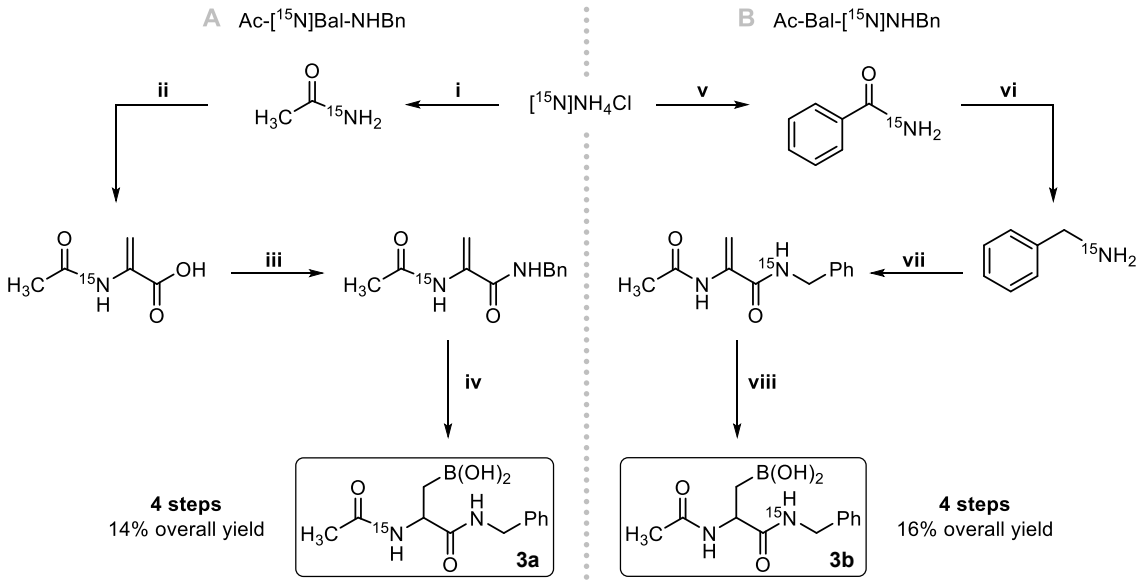
Table 4.1: Formation of the novel species is not concentration dependent

c [mg/mL]	c[mM]	^{1}H-NMR		^{11}B-NMR	^{13}C-NMR	
		C_β–H	CH_3	B_γ	C_β	C_α
30	113.6	32.9	36.9	34.5	33.6	33.5
15	56.8	33.0	34.7	36.5	33.3	33.0
7.5	28.4	33.0	35.4	37.5	31.9	33.6
3.75	14.2	33.2	35.3	n.d.	n.d.	n.d.
1.875	7.1	33.1	37.4	n.d.	n.d.	n.d.

The detected amount of novel species is given in percent. Percentages are relative to the total amount of Ac-Bal-NHBn (**3**) present is solution and are derived from the respective relative integrals. Samples were measured in deuterated, aqueous buffer (20 mM NaP_i, pH 8.5, D_2O).

4.2.1 NMR Studies on Isotopically Labelled Model Compound

Synthesis of Isotopically Labelled Model Compound



Scheme 4.1: Synthesis of ^{15}N-labelled isotopologues of small molecule model Ac-Bal-NHBn (**3**). (**A**) Synthesis of Ac-[^{15}N]Bal-NHBn (**3a**): (i) AcCl, NaOH, PhMe; (ii) pyruvic acid, PhMe, Dean-Stark conditions; (iii) $BnNH_2$, *N*-methylmorpholine, isobutyl chloroformate, THF; (iv)$B_2(OH)_4$, $CuSO_4$, 4-picoline, H_2O. (**B**) Synthesis of Ac-Bal-[^{15}N]NHBn (**3b**): (v) BzCl, NaOH, PhMe; (vi) $LiAlH_4$, THF, rfx; (vii) 2-acetamidoacrylic acid, *N*-methylmorpholine, isobutyl chloroformate, THF; (viii) $B_2(OH)_4$, $CuSO_4$, 4-picoline, H_2O.

The introduction of ^{15}N-labels for either of the two amide nitrogen atoms in Ac-Bal-NHBn (**3**) would enable the ^{15}N-NMR detection of engagement of either of the amide carbonyl groups with the central boronic acid moiety. ^{15}N-labelled isotopologues of Ac-Bal-NHBn (**3**) were synthesized starting from commercially available [^{15}N]ammonium chloride (**Scheme 4.1**). The synthesis of Ac-[^{15}N]Bal-NHBn (**3a**) was achieved reacting [^{15}N]ammonium chloride with acetyl chloride followed by condensation of the resulting [^{15}N]acetamide with pyruvic acid to yield [^{15}N]acetamidoacrylic acid. Subsequent amidation with benzylamine followed by borylation led to formation of Ac-[^{15}N]Bal-NHBn (**3a**) in 14% overall yield (**Scheme 4.1 A**). The synthesis of Ac-Bal-[^{15}N]NHBn (**3b**) commenced with the synthesis of [^{15}N]benzylamine as described in literature.[13] Coupling to commercially available acetamidoacrylic acid and successive borylation allowed for the synthesis of Ac-Bal-[^{15}N]NHBn (**3b**) in 16% overall yield (**Scheme 4.1 B**).

NMR Analysis of Small Molecule Model Compound

^{15}N-NMR analysis in aqueous buffer under basic conditions (20 mM NaP_i in D_2O, pH 8.50) revealed involvement of the *C*-terminal amide in the formation of the novel species (**Figure 4.1 A**). A 40.9 ppm difference in chemical shift was observed for the two species in the case of Ac-Bal-[^{15}N]NHBn (**3b**), while only 5.7 ppm chemical shift difference were observed for Ac-[^{15}N]Bal-NHBn (**3a**) (**Figure 4.1 A**). Following this clear indication of engagement of the boronic acid moiety with the *C*-terminal amide, nitrogen substitution was further investigated. When the sample was measured in H_2O (containing 5% v/v D_2O) instead of D_2O, a singlet was observed for the unknown species (**Figure 4.1 B**). In the presence of an amidic proton, a doublet would be expected due to ^{15}N—^{1}H coupling ($^1J_{N-H}$ approx. 90 Hz).[14] Indeed, a broad doublet is observed in the case of the free boronic acid featuring a coupling constant within the expected range ($^1J_{N-H}$ = 72.4 Hz). This observation and the observed change in chemical shift are in agreement with the formation of an oxaborolane, as described by Yudin and co-workers on a similar substrate.[9] Further NMR experiments were conducted to confirm this hypothesis. For the novel species, two sets of peaks corresponding to the C_β protons were observed. Only one of these showed a cross peak with the *C*-terminal carbonyl group in the ^{1}H-^{13}C-HMBC. In the respective ^{1}H-^{15}N-HMBC spectra, for each of the two

amides a cross peak with only one C_β—H was observed. This indicates a species with a highly constrained conformation leading to dihedral angles (ϕ) (and thus coupling constants) that do not allow observation in HMBC spectra. This degree of constraint would be probable for a cyclic species, providing further evidence for the formation of an oxaborolane (**Figure 4.2**).

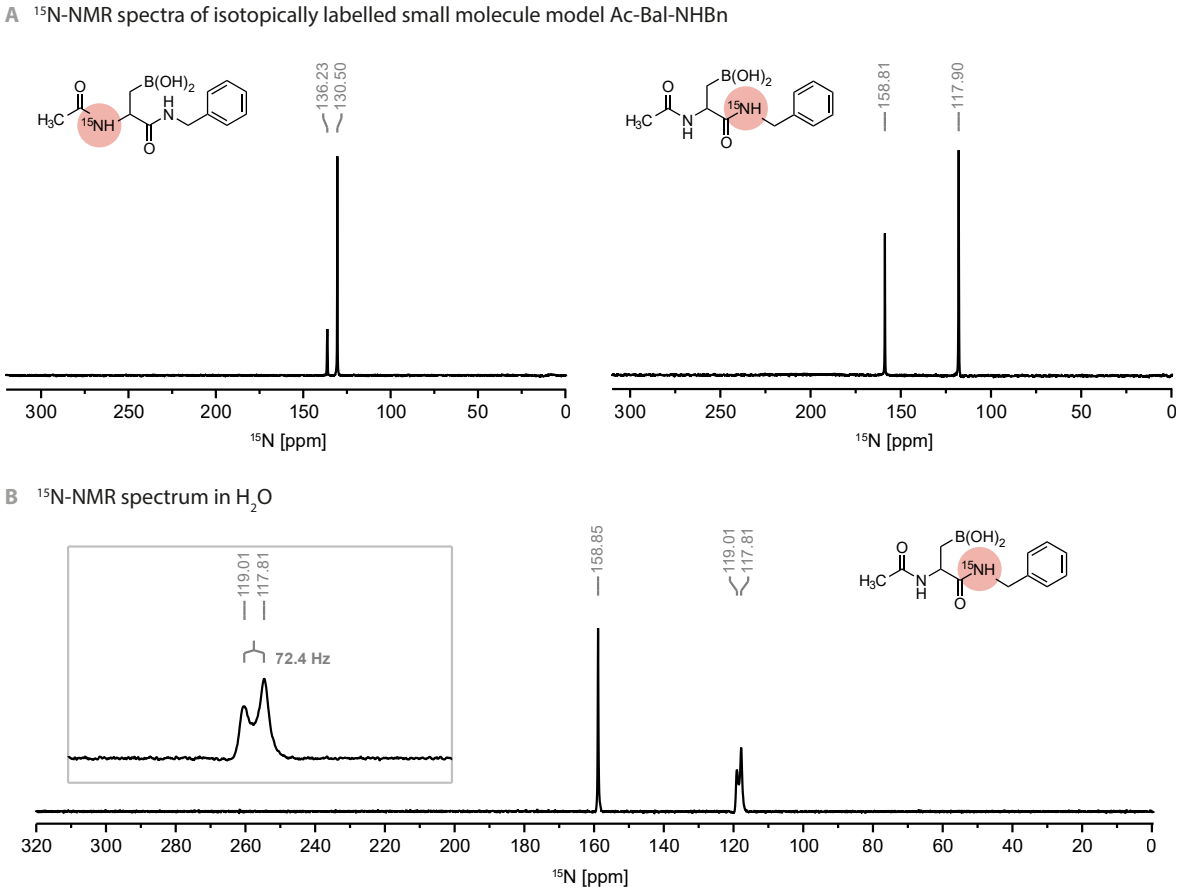


Figure 4.1: ^{15}N-NMR studies on ^{15}N-labelled model compound Ac-Bal-NHBn (**3a** and **3b**). (**A**) Chemical shifts suggest involvement of the *C*-terminal amide in formation of the novel species. (**B**) NMR analysis in H_2O reveals the absence of an amidic proton in the novel species.

These observations were additionally confirmed by comparison of observed NMR shifts with calculated NMR shifts. Taking into consideration ^{1}H-, ^{11}B-, ^{13}C- and ^{15}N-NMR chemical shifts, chemical shift perturbations were calculated allowing for assignment of the observed peaks to a respective species (**Table 4.2**). The unbound boronic acid hereby likely exists as the free acid, which could exhibit some degree of coordination to the *N*-terminal acetylamide carbonyl oxygen. The oxaborolane seems to exist exclusively as its tetracoordinate, negatively charged boronate. This is further supported by the observed ^{11}B-NMR chemical shift, which

is consistent with a sp^3-hybridized boron centre.[15] It should be noted that the nature of these species is highly pH dependent and changes are likely at different pH.

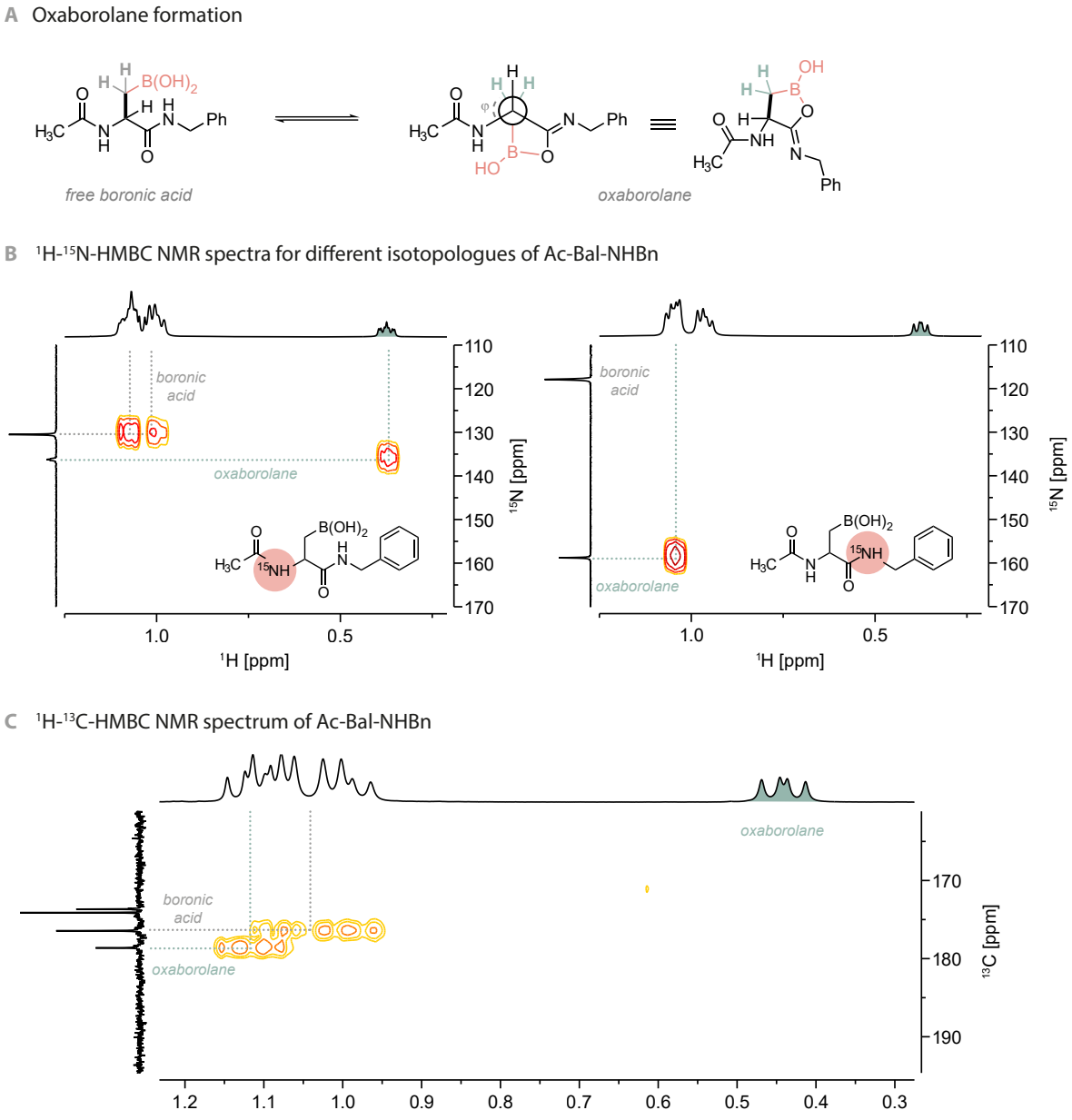
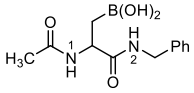
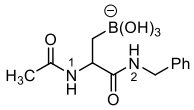
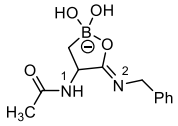
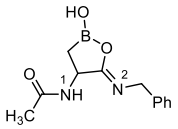
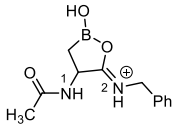
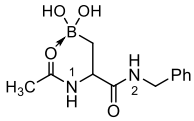
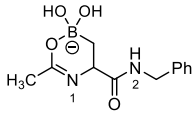
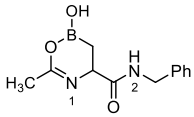
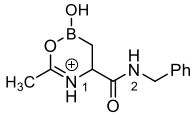


Figure 4.2: HMBC NMR experiments suggest a constrained conformation for the C_β protons. (**A**) Formation of a cyclic species – an oxaborolane – would lead to a constrained conformation for the C_β protons. This constrained conformation is confirmed by ^{1}H-^{15}N-HMBC (**B**) and ^{1}H-^{13}C-HMBC (**C**) spectra which do not allow for simultaneous observation of both C_β protons for the cyclic species.

To investigate the interconversion between these species as a function of varying pH, NMR spectra were recorded at a range of different pH values. It was found that formation of both, the oxaborolane and boronate is promoted by increasingly basic conditions. The formation

Table 4.2: Observed and calculated NMR shifts for possible species of Ac-Bal-NHBn.

No	Species	1H-NMR [ppm]		11B-NMR [ppm]	13C-NMR [ppm]		15N-NMR [ppm]		CSP (boronic acid)	CSP (cyclic form)
		C_α–H	C_β–H	B_γ	C_α	C_β	N_1	N_2		
1	observed (boronic acid)	4.28	0.94-1.18	21.4	52.6	20.4	130.5	117.9		
2	observed (cyclic form)	4.49	0.44 0.94-1.18	4.9	53.2	24.5	136.2	158.8		
3		3.3	0.54 2.51	26.3	55.3	19.1	123	109	*1.31*	5.61
4		3.79	0.27 0.46	2.2	61.7	23.3	128	104	2.57	5.61
5		3.92	0.34 1.00	8.3	57.1	26.2	136	167	5.16	*0.99*
6		3.39	1.26 1.89	32.2	55.7	15.7	122	233	11.61	8.09
7		3.53	0.92 1.97	31.3	59.1	12.3	116	228	11.20	7.80
8		3.74	0.63 1.21	9.9	61.4	15.9	132	114	*1.57*	4.68
9		3.68	-0.34 0.95	2.1	62.6	22.7	202	113	7.49	8.08
10		4.14	0.61 1.74	28.7	57.7	10.1	228	120	9.85	10.36
11		3.56	0.91 0.98	11.4	60.4	17	83	138	5.32	5.84

Isotropic chemical shifts were calculated using the Gaussian09.[16] Three dimensional models were created and their geometry was first optimised, before calculation of the shielding tensor from which the isotropic chemical shift can be calculated. All DFT calculations were carried out using the B3LYP density functional with the 6-31G(d) basis set, a combination known to give values in excellent agreement with experimental data.[17,18] Data for observed species was collected on ^{15}N-labelled Ac-Bal-NHBn (**3**) in D_2O (20 mM NaP_i, pH 8.5). CSP was calculated as $CSP = \sqrt{\sum_{i,n}(\omega_A - \omega_B)^2 \times \frac{\gamma_H^2}{10^2 \times \gamma_n^2}}$ using calculated chemical shift ω_A, observed chemical shift ω_B, $\gamma_H = 267 \times 10^6\ rad\,T^{-1}s^{-1}$, $\gamma_C = 67.28 \times 10^6\ rad\,T^{-1}s^{-1}$, $\gamma_N = -27.12 \times 10^6\ rad\,T^{-1}s^{-1}$ and $\gamma_B = 85.84 \times 10^6\ rad\,T^{-1}s^{-1}$. *DFT calculations were performed by Prof. Andrew J. Baldwin.*

of both species could be resolved by observation of the C_β protons in the ^{1}H-NMR spectra (1.30−0.25 ppm) (**Figure 4.3 A**). The formation of the oxaborolane (rose) is manifested in the appearance of additional peaks, which is indicative for slow exchange. In contrast, formation of the boronate (turquoise) is revealed by a change in chemical shift (over a range of approximately 0.6 ppm) indicating fast exchange. The appearance of new peaks (oxaborolane) or change in chemical shift (boronate) allowed for the determination of the respective pK_a values (**Figure 4.3 B**). The pK_a values for both, oxaborolane (pK_a = 8.50±0.026) and boronate formation (pK_a = 8.94±0.084) were found to be almost two orders of magnitude lower than expected for simple alkyl boronic acids and more in the range of simple arylboronic acids (phenylboronic acid: pK_a = 8.8, methylboronic acid: pK_a = 10.4).[19–22] It should be noted that these values were determined in D_2O and a small correction needs to be applied to allow for comparison with pK_a values determined in H_2O.[23] This unusually low pK_a was further confirmed by acid-base titration indicating an experimental value of 8.31±0.02. Indeed, this unusually high acidity could explain Bal's outstanding ability to engage in PLABP (**Chapter 3**). Having established and explicitly characterized "backbone" interactions of boronoalanine on a small molecule model compound, the effect of this behaviour on a protein was yet to be explored. Isotopically (^{13}C, ^{15}N) labelled model protein Histone H3 Cys10 was expressed and borylated following the previously described procedure (**Chapter 2**). Intriguingly, upon analysis of the $^{13}CO^{15}N$-NMR spectrum recorded in denaturing buffer (50 mM NaP_i, 3 M Gdn·HCl, 5% D_2O, pH 7.0), the loss of six signals became visible in the case of Histone H3 Bal10 (blue) when compared to Histone H3 Cys10 (red) (**Figure 4.3 C**). No further major changes to the NMR spectrum could be detected upon borylation. This loss of multiple signals is in agreement with PLABP in an intermediate exchange regime. The greatly increased local concentration of Lewis basic backbone carbonyl groups in the protein sample compared to the small molecule model system can explain the shift from slow to intermediate exchange.

The availability of a probe that can modulate NMR signal intensity by PLABP in a selective way is intriguing. If the observed selectivity is indeed a function of distance, this would render Bal a readily introducable and useful structural NMR probe.

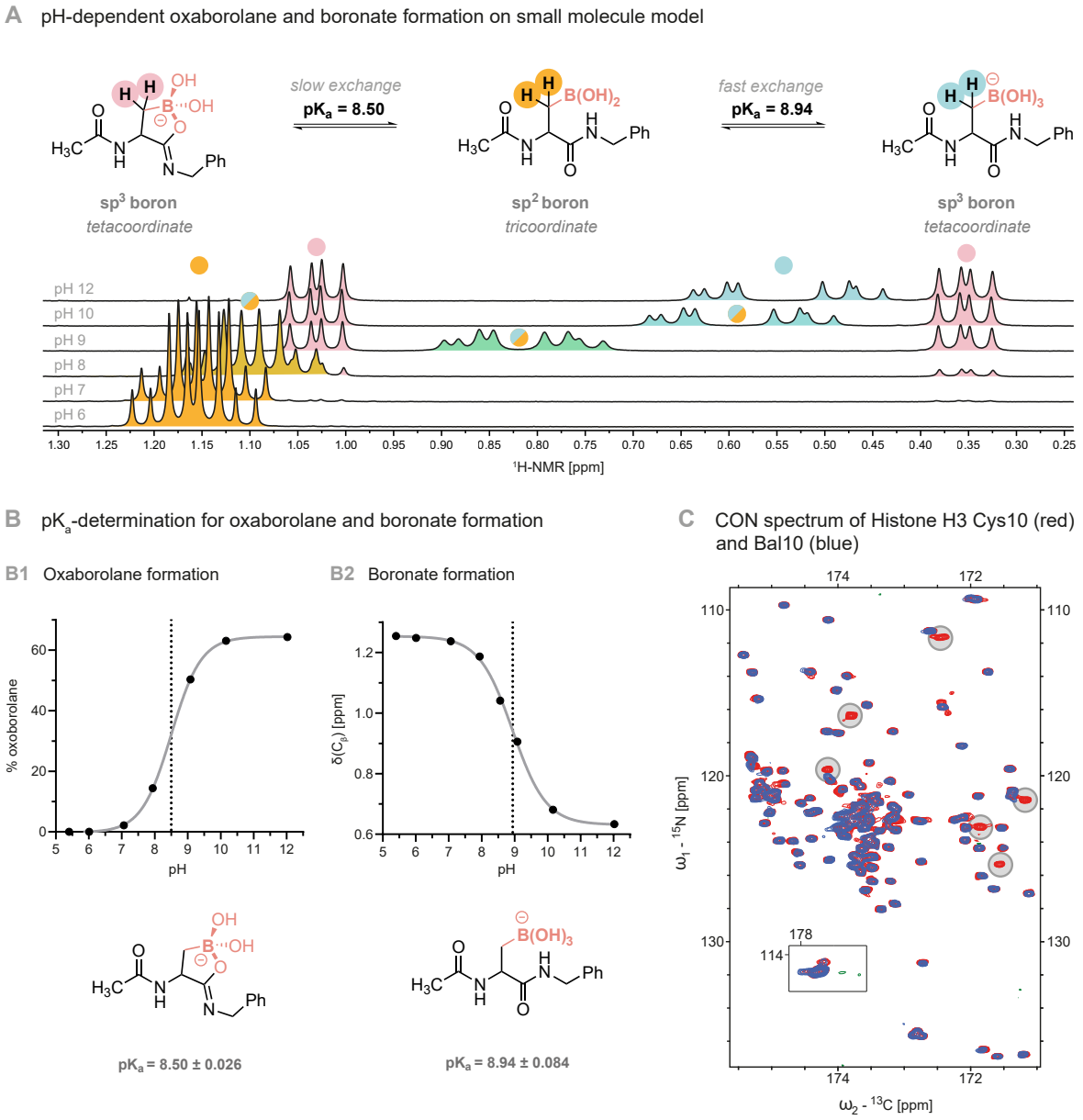


Figure 4.3: Oxaborolane and boronate formation are pH dependent. (**A**) oxaborolane (rose) formation is a special case of PLABP, featuring engagement of the boronic acid with the *C*-terminal C=O on small molecule model. In contrast to boronate (turquoise) formation, which takes place in a fast exchange regime, slow exchange was observed for oxaborolane formation. (**B**) The pK_a values for oxaborolane formation (**B1**, $pK_a = 8.50 \pm 0.026$) and boronate formation (**B2**, $pK_a = 8.94 \pm 0.084$) were determined by ^{1}H-NMR. (**C**) The $^{13}CO^{15}N$ protein NMR spectrum of model protein Histone H3 Bal10 (blue) shows the loss of six signals (grey circles) when compared to Histone H3 Cys10 (red). This can be attributed to backbone engagement via PLABP in an intermediate exchange regime. The shift from slow to intermediate exchange can be rationalized with the greatly increased local concentration of available backbone carbonyls.

4.2.2 Probing Histone H3 Tail Conformations in a Nucleosome Context

Eukaryotic cells store their DNA in chromosomes which are formed from chromatin fibers. These fibers consist of an array of nucleosomes often figuratively referred to as 'beads on a string'.[24,25] Each nucleosome consists of a Histone octamer (containing two copies each of Histone H2A, H2B, H3 and H4) which is wrapped by approximately 150 base pairs of DNA.[26] The Histones possess disordered *N*-terminal tails which fulfil multiple functions.[27] These tails are known to often be populated by a myriad of post-translational modifications (PTMs) which contribute to the regulation of multiple biological processes.[28–30] One of these PTMs is the phosphorylation of serine mediated by Aurora B kinase. Intriguingly, phosphorylation of Histone H3 at serine 10 was shown to cause two seemingly contrary effects: chromatin condensation and gene silencing at mitosis and chromatin relaxation and gene expression in interphase.[29,31–33] A model for these seemingly opposing observations proposed Histone H3 Ser10 phosphorylation as a structural regulator of chromatin organization.[34] While the structure of the central, well-ordered part of the nucleosome was elucidated at high resolution,[26,29,35] detailed structural studies on the intrinsically disordered *N*-terminal tail regions (and consequently the elucidation of the structural implications of the introduction of PTMs such as phosphorylation) are difficult to conduct and remain lacking.[7,27,29,31,36]

Having established that boronoalanine is able to engage with the peptidic backbone via the transient formation of cyclic species through C=O···B_γ bond formation (**Figure 4.3**) and consequently modulates NMR signal via chemical exchange (**Figure 4.3 C**), the applicability of Bal as a structural NMR probe to investigate transient contacts was put to test using Histone H3 phosphorylation in a reconstituted nucleosome context as a model system. Previous studies indicated that the tail of Histone H3 contains residual α-helical structure which can be described as two separate, α-helical domains (residues 3-10 and 14-28) possibly connected by a β-turn. Upon Ser10 phosphorylation transient structure in the tail is released leading to increased flexibility.[37] This structural change upon phosphorylation by Aurora B kinase should be observable as a modulation of NMR intensity, if a suitably placed Bal residue acts as a structural probe as predicted.

In order to construct a borylated nucleosome suitable for NMR studies, ^{15}N-labelled Histone

H3 Cys9 was expressed. Following the previously described protocol (**Chapter 2**), the isotopically labelled protein was converted into its respective Bal mutant [^{15}N]Histone H3 Bal9 in which Bal replaced the native residue Lys9. Site 9 was chosen strategically for borylation as it does not only sit directly next to phosphorylation site Ser10, but the native residue Lys9 also plays an important biological role which is modulated by Ser10 phosphorylation. Heterochromatin protein 1 (HP1) is recruited by Lys9 trimethylation and ejected upon Ser10 phosphorylation.[38,39] It was shown that Ser10 phosphorylation disrupts HP1 ejection independent of Lys9 trimethylation levels acting like a binary switch.[34,38,40] This suggests that site 9 might experience a significant structural change upon phosphorylation leading to HP1 ejection, which – if existent – should be detectable by PLABP-enabled modulation of NMR intensity.

Nucleosome formation proceeded smoothly in spite of the presence of Bal following a literature procedure.[41] Nucleosomes containing either [^{15}N]Histone H3 Bal9 or wild-type [^{15}N]Histone H3 derived from oxidation of [^{15}N]Histone H3 Ser10 were assembled (**Figure 4.4 A**) and analysed by NMR spectrometry. Excitingly, ^{1}H-^{15}N-HSQC NMR measurements conducted before and after phosphorylation (H3 Ser10 phosphorylation was confirmed by Western blot) revealed a clear change in Bal-mediated intensity modulation upon the Aurora B kinase mediated introduction of phosphoserine residues (**Figure 4.4 B**). This change is consistent with the postulated release of tail structure and hence increased flexibility (**Figure 4.4 C**).[37] Bal is held in closer proximity to the backbone carbonyl groups of the second α-helical domain (residues 14-28) in the non-phosphorylated, more structured case, leading to increased backbone engagement via PLABP and consequently loss of intensity. The underlying structural features (two α-helices) are directly visible as a 'double dip' pattern in the intensity ratio plots comparing the borylated and non-borylated nucleosomes (**Figure 4.4 B**). Once phosphorylated, the increased flexibility leads to less backbone engagement in the second α-helical domain (residues 14-28), resulting in the recovery of the observed intensities. Again, this change can be directly read from the corresponding intensity ratio plot.

Intriguingly – and in contrast to typically employed PRE measurements – Bal-mediated residue-specific intensity modulation did not only expose long range contacts but also provided further information on structural features in the boronic acid probe's direct spacial environment

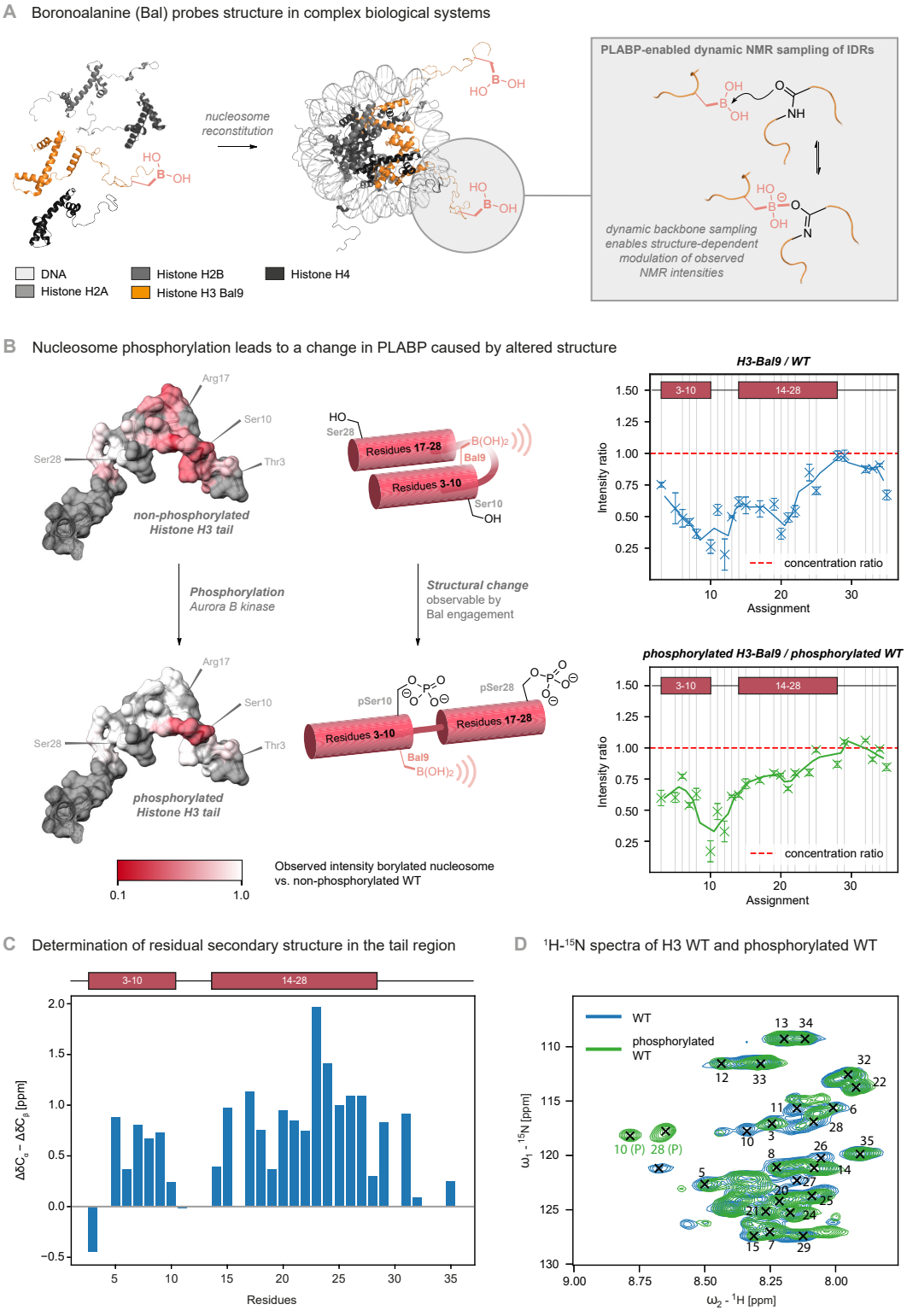


Figure 4.4: Application of Bal-mediated PLABP to probe structure in a complex biological system using NMR spectroscopy. (**A**) Introduction of ^{15}N-labelled, borylated Histone H3 Bal9 into nucleosomes allows for PLABP-enabled dynamic NMR sampling of intrinsically disordered regions (IDRs). (**B**) Boronoalanine (Bal) acts as a structural NMR probe for the detection of changed structure in the *N*-terminal tail region upon phosphorylation. (**C**) Previously reported residual secondary structure determination of non-phosphorylated Histone H3 in a nucleosome context via $\Delta\delta C_\alpha - \Delta\delta C_\beta$ chemical shift analysis allowed for the detection of two α-helical domains (residues 3-10 and residues 14-28).[37] (**D**) Overlay of ^{1}H-^{15}N-HSQC NMR spectra of wild-type (blue) and phosphorylated wild-type (green) nucleosomal Histone H3.

(**Figure 4.5**). Intensity modulation observed for residues adjacent to Bal showed a pattern that is agreement with a β-turn connecting the two α-helical domains as postulated earlier (**Figure 4.5 C**). The placement of Bal herein leads to preferential interactions with specific carbonyls (i+1 and i+3, with i being the borylation site) leading to a pronounced loss of intensity. The simultaneous observation of long range contacts and structural features within the local environment of the NMR probe (Bal) has not yet been described and demonstrates the possible utility of boronoalanine for structural NMR studies.

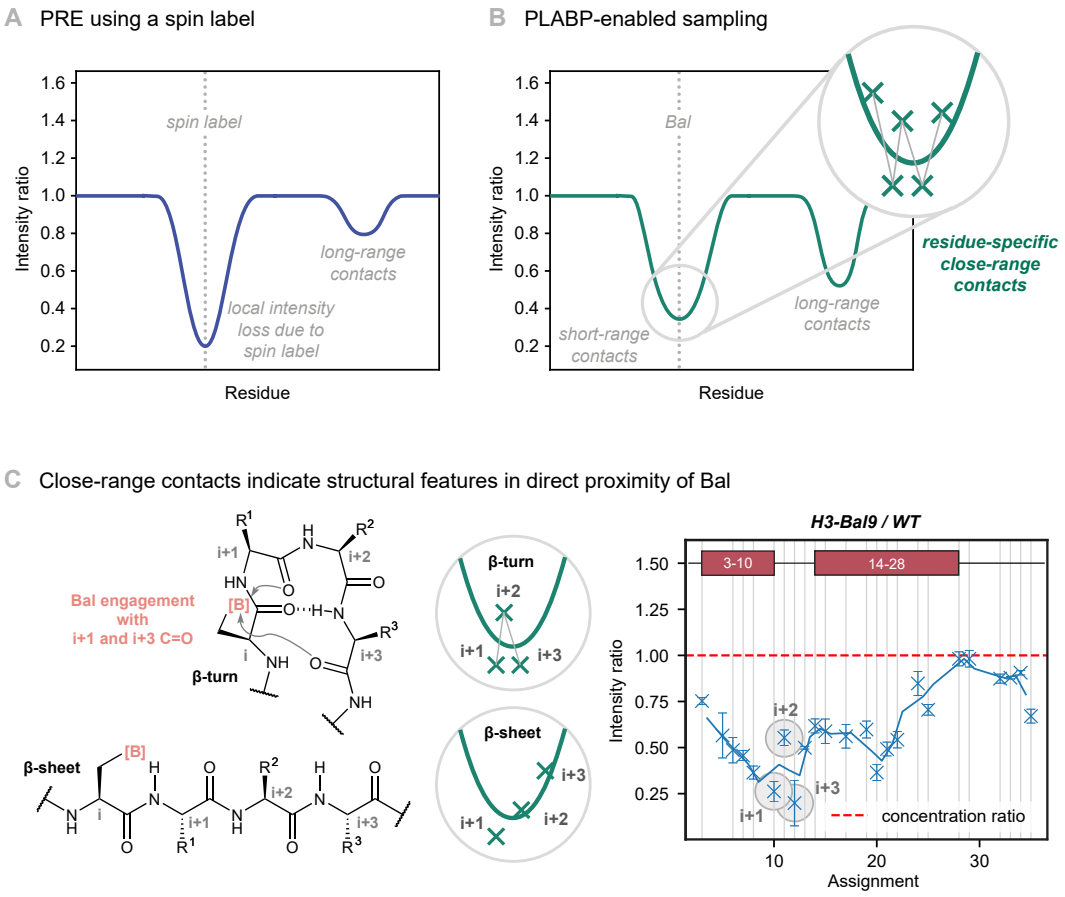


Figure 4.5: Observation of local structural features using a boronoalanine (Bal) NMR probe. (**A**) The application of spin labels for PRE measurements leads to a local intensity loss around the spin label masking any local structural features. Long range contacts are observed via intensity modulation. (**B**) PLABP-enabled sampling using Bal enables the observation of long range contacts as well as residue-specific close-range contacts. (**C**) The observed intensity patterns in direct proximity of Bal allow for the identification of local structural motifs. In the case of the Histone H3 tail in a nucleosomal context, the observed intensity pattern indicates the presence of a β-turn between the two α-helical domains.

In order to confirm that the observed intensity modulation is indeed a direct result of residual structure and that the observed changes are a consequence of Aurora B kinase dependent phosphorylation, several control experiments were conducted. The intensity ratios observed for glycine, serine and threonine residues (**Figure 4.6 B**) were shown to be temperature dependent, which is in agreement with chemical exchange (**Figure 4.6 A**). Additionally, a drastically reduced amount of intensity loss was observed for denatured Histone H3 (green) when compared to more structured, nucleosomal Histone H3 (blue: non-phosphorylated, yellow: phosphorylated) (**Figure 4.6 D**).

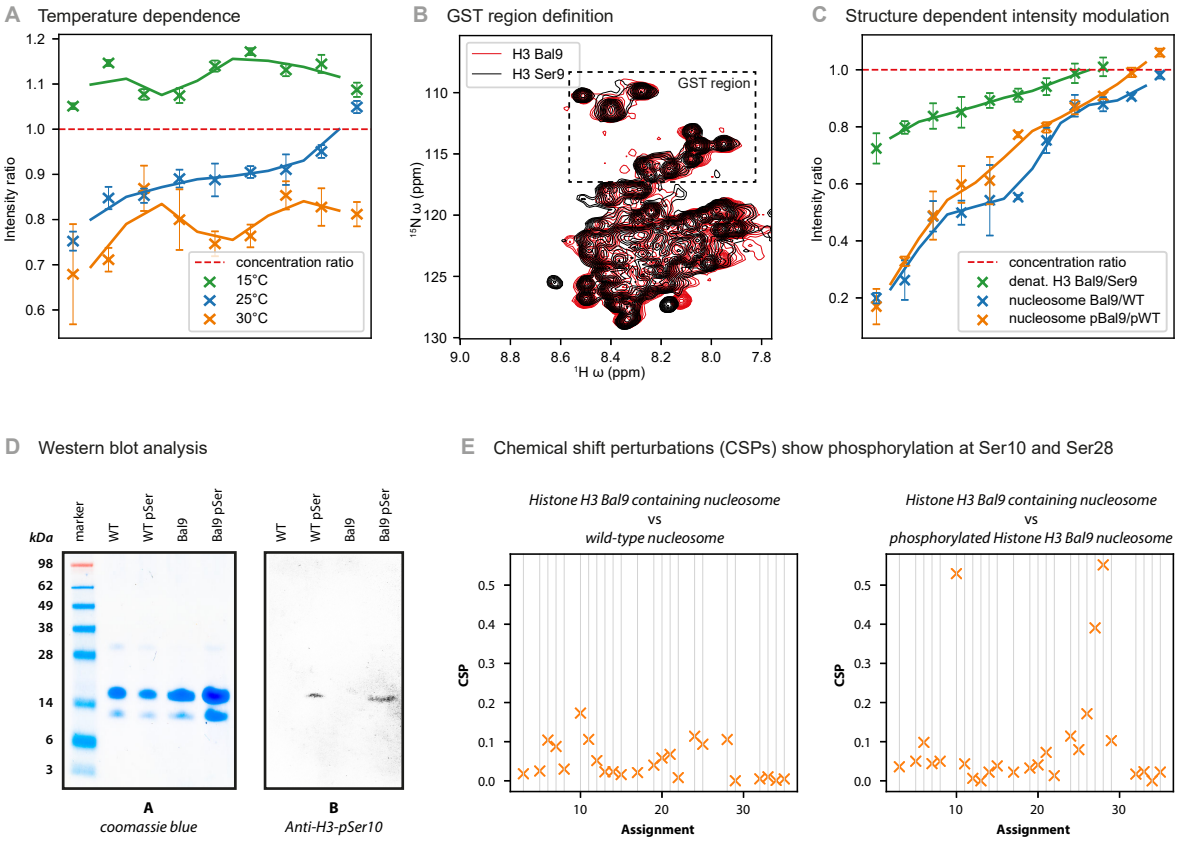


Figure 4.6: Control experiments for the application of Bal as a NMR probe. (**A**) Bal-enabled intensity modulation in the glycine, serine and threonine (GST) region of denatured Histone H3 Bal9 is temperature dependent. (**B**) Contour plot showing the definition of the GST region. (**C**) Comparison of Bal-enabled intensity modulation in denatured Histone H3 Bal9 and in a nucleosomal context reveals structure dependency. (**D**) Western blot analysis of phosphorylated nucleosomes confirms phosphorylation of Histone H3 Ser10 in presence and absence of Bal9. (**E**) Chemical shift perturbations (CSPs) confirm phosphorylation of Histone H3 at the native phosphorylation sites Ser10 and Ser28 in the presence of Bal9.

This proves that intensity modulation does depend on residual structure. Furthermore, the presence of phosphoserine residues was verified via chemical shift perturbations (CSPs) and western blot (**Figure 4.6 D** and **Figure 4.6 E**). CSPs did indicate phosphorylation at Ser10 and Ser28, which is in agreement with the reported selectivity of Aurora B kinase.[32] Western blot using a primary anti-Histone H3 pSer10 antibody further confirmed phosphorylation at this position regardless of the presence of an adjacent boronic acid.

4.3 Conclusion

Boronoalanine was shown to act as a powerful NMR probe for detection of transient contacts in an intrinsically disordered region. While the detection of long range contacts is well established for example using proteins tagged with spin labels (typically attached via standard bioconjugation protocols) or chelated paramagnetic metals for paramagnetic relaxation enhancement (PRE) measurements, simultaneously gathering information on local structural features via short range contacts is not feasible using this methodology.[3,42] Furthermore, the introduction of bulky spin labels can affect protein structure and function.[6] In contrast, the introduction of a minimally sized Bal residue allowed for the detection of long range contacts (comparable to PRE) as well as short range contacts providing information on structural features in its proximity – a unique feature not yet described for any other structural NMR probe. Additionally, and in agreement with previous studies indicating the retention of fold and function upon borylation (**Chapter 2**), introduction of Bal did neither affect nucleosome reconstitution nor phosphorylation by Aurora B kinase. The latter is especially remarkable and a testament to Bal's selectivity and biocompatibility, considering that Bal was placed at site 9 directly next to phosphorylation site Ser10 prior to phosphorylation. The unique compatibility of Bal with biological systems (yet giving rise to selective *de novo* binding function) was further highlighted by successful western blot analysis of the phosphorylated, borylated nucleosome. Phosphorylation at Ser10 was detected in spite of the presence of Bal9 using a highly specific anti-Histone H3 pSer10 antibody.

The wealth of structural information that can be gathered using boronoalanine as a minimally sized, non-disruptive and biocompatible NMR probe could indeed make PLABP-enabled

dynamic sampling of transient contacts a very powerful technique for structural NMR studies – that could possibly be extended far beyond the proof-of-principle example presented in this chapter.

References

[1] A. Dunker, J. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, Z. Obradovic, *J. Mol. Graph. Model.* **2001**, *19*, 26–59.

[2] P. Tompa, *Trends Biochem. Sci.* **2012**, *37*, 509–516.

[3] D. Eliezer, *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.

[4] M. R. Jensen, R. W. Ruigrok, M. Blackledge, *Curr. Opin. Struct. Biol.* **2013**, *23*, 426–435.

[5] D. Eliezer in *Protein Folding Protocols*, *Vol. 350*, Humana Press, New Jersey, **2007**, pp. 49–68.

[6] C. A. Softley, M. J. Bostock, G. M. Popowicz, M. Sattler, *J. Biomol. NMR* **2020**, *74*, 287–309.

[7] J. Habchi, P. Tompa, S. Longhi, V. N. Uversky, *Chem. Rev.* **2014**, *114*, 6561–6588.

[8] G. M. Clore, J. Iwahara, *Chem. Rev.* **2009**, *109*, 4108–4139.

[9] S. J. Kaldas, T. Rogova, V. G. Nenajdenko, A. K. Yudin, *J. Org. Chem.* **2018**, *83*, 7296–7302.

[10] A. Krajnc, J. Brem, P. Hinchliffe, K. Calvopiña, T. D. Panduwawala, P. A. Lang, J. J. Kamps, J. M. Tyrrell, E. Widlake, B. G. Saward, T. R. Walsh, J. Spencer, C. J. Schofield, *J. Med. Chem.* **2019**, *62*, 8544–8556.

[11] D. G. Hall in *Boronic Acids*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2011**, pp. 1–133.

[12] V. V. Zhdankin, P. Persichini, L. Zhang, S. Fix, P. Kiprof, *Tetrahedron Lett.* **1999**, *40*, 6705–6708.

[13] H. J. Reich, W. S. Goldenberg, B. Ö. Gudmundsson, A. W. Sanders, K. J. Kulicke, K. Simon, I. A. Guzei, *J. Am. Chem. Soc.* **2001**, *123*, 8067–8079.

[14] J. P. Marchal, D. Canet, *Org. Magn. Reson.* **1981**, *15*, 344–346.

[15] C. R. Cooper, *Chem. Commun.* **1998**, 1365–1366.

[16] M. J. F. and Fox, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, Gaussian 09, Gaussian, Inc., Wallingford CT, **2009**.

[17] G. Karunanithy, A. Cnossen, H. Müller, M. D. Peeks, N. H. Rees, T. D. W. Claridge, H. L. Anderson, A. J. Baldwin, *Chem. Commun.* **2016**, *52*, 7450–7453.

[18] M. W. Lodewyk, C. Soldi, P. B. Jones, M. M. Olmstead, J. Rita, J. T. Shaw, D. J. Tantillo, *J. Am. Chem. Soc.* **2012**, *134*, 18550–18553.

[19] L. Babcock, R. Pizer, *Inorg. Chem.* **1980**, *19*, 56–61.

[20] D. L. Yabroff, G. E. Branch, B. Bettman, *J. Am. Chem. Soc.* **1934**, *56*, 1850–1857.

[21] S. Soundararajan, M. Badawi, C. M. Kohlrust, J. H. Hageman, *Anal. Biochem.* **1989**, *178*, 125–134.

[22] J. Yan, G. Springsteen, S. Deeter, B. Wang, *Tetrahedron* **2004**, *60*, 11205–11209.

[23] A. Krężel, W. Bal, *J. Inorg. Biochem.* **2004**, *98*, 161–166.

[24] D. J. Tremethick, *Cell* **2007**, *128*, 651–654.

[25] S. Baldi, P. Korber, P. B. Becker, *Nat. Struct. Mol. Biol.* **2020**, *27*, 109–118.

[26] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251–260.

[27] C. Zheng, J. J. Hayes, *Biopolymers* **2003**, *68*, 539–546.

[28] P. Cheung, C. D. Allis, P. Sassone-Corsi, *Cell* **2000**, *103*, 263–271.

[29] A. J. Bannister, T. Kouzarides, *Cell Res.* **2011**, *21*, 381–395.

[30] K. Zhou, G. Gaullier, K. Luger, *Nat. Struct. Mol. Biol.* **2019**, *26*, 3–13.

[31] Y. Wei, C. A. Mizzen, R. G. Cook, M. A. Gorovsky, C. D. Allis, *Proc. Natl. Acad. Sci.* **1998**, *95*, 7480–7484.

[32] C. Prigent, *J. Cell Sci.* **2003**, *116*, 3677–3685.

[33] A. Sawicka, C. Seiser, *Biochimie* **2012**, *94*, 2193–2201.

[34] K. M. Johansen, J. Johansen, *Chromosom. Res.* **2006**, *14*, 393–404.

[35] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, *J. Mol. Biol.* **2002**, *319*, 1097–1113.

[36] D. A. Potoyan, G. A. Papoian, *Proc. Natl. Acad. Sci.* **2012**, *109*, 17857–17862.

[37] L. Lercher, DPhil Thesis, University of Oxford, **2014**.

[38] W. Fischle, S. T. Boo, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki, C. D. Allis, *Nature* **2005**, *438*, 1116–1122.

[39] H. L. Dormann, S. T. Boo, C. D. Allis, H. Funabiki, W. Fischle, *Cell Cycle* **2006**, *5*, 2842–2851.

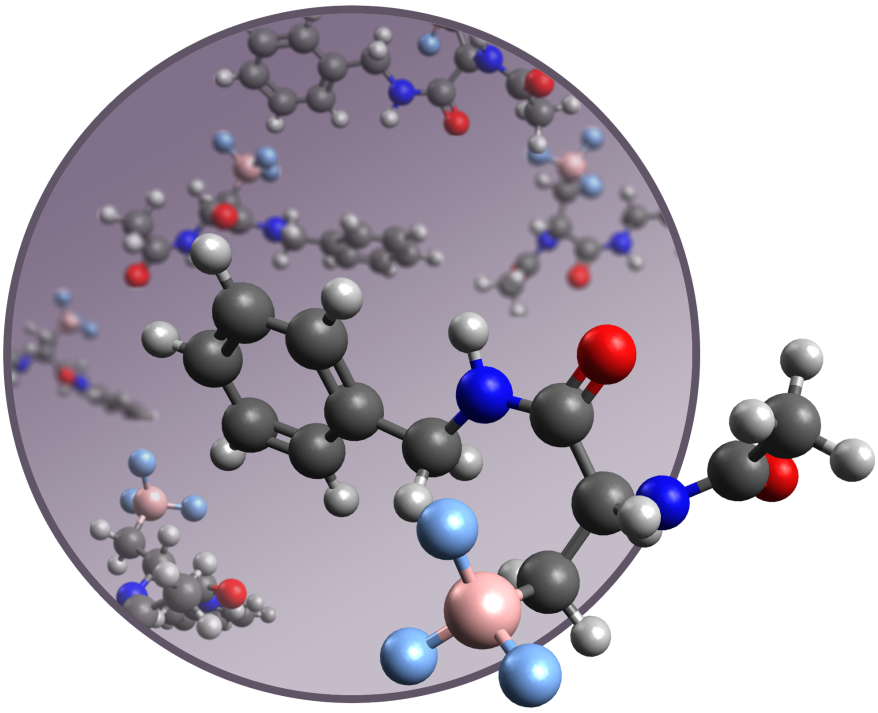
[40] T. Banerjee, D. Chakravarti, *Mol. Cell. Biol.* **2011**, *31*, 4858–4873.

[41] P. N. Dyer, R. S. Edayathumangalam, C. L. White, Y. Bao, S. Chakravarthy, U. M. Muthurajan, K. Luger, *Methods Enzymol.* **2003**, *375*, 23–44.

[42] G. M. Clore, C. Tang, J. Iwahara, *Curr. Opin. Struct. Biol.* **2007**, *17*, 603–616.

Chapter 5

Capturing Fluoride: Incorporation of Trifluoroboronoalanine (3F-Bal) into Proteins



We speak erroneously of "artificial" materials, "synthetics", and so forth. [...] Nature wrote all the rules of structuring; [...] All the chemist can do is find out what Nature permits, and any substances that are thus developed or discovered are inherently natural.

— *R. Buckminster Fuller*, Ideas and Integrities (1969)

Chapter 5

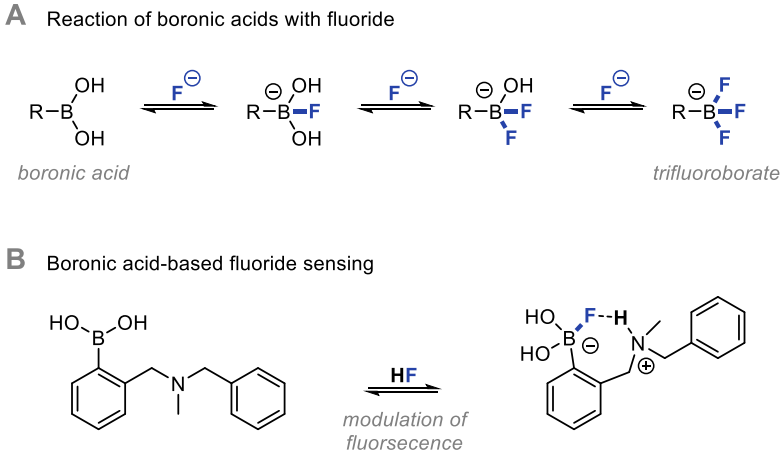
Capturing Fluoride: Incorporation of Trifluoroboronoalanine (3F-Bal) into Proteins

Proteins capable of binding inorganic fluoride are rare in nature, yet several examples exist.[1–5] Known examples share a common binding mode, where the fluoride anion sits in a hydrophobic pocket and is bound via hydrogen bonding, which leads to at least partial dehydration.[1–5] Fluoride has been shown to act as an inhibitor of different enzymes when binding in their active site.[1,3] The ability of the fluoride anion to bind to the oxyanion hole within the active site of enzymes was proposed as a general methodology to obtain crystal structures trapping enzymes "in action".[2,6] Yet, enzymes rarely possess native fluoride binding sites. In fact, only two enzymes – the fluorinase and the fluoroacetate dehalogenase – are currently known that make use of a tailored fluoride binding pocket.[1,3] The low abundance of fluoride binding proteins as well as their reliance on the placement of fluoride in hydrophobic pockets might be a direct effect of fluoride's intrinsically low reactivity towards nucleophilic fluorination in aqueous solution.[6,7] Fluoride anions are highly hydrated in aqueous media, hampering basicity and reactivity.[7,8] Consequently, engineering of a fluoride capturing protein would either entail the precise construction of a tight fluoride binding site, in which the fluoride

anion can be stabilized by hydrogen bonding, or the introduction of an unnatural amino acid capable of interaction with aqueous fluoride for example by formation of a X—F bond.

5.1 From Boronic Acid to Trifluoroborate

Boron occurs naturally mostly in form of borate or borosilicate minerals. Strikingly, the only known exception to this are tetrafluoroborate containing minerals, which could be explained by the exceptional strength of the boron—fluoride σ-bond (BDE B—F bond: 766 kJ/mol).[9–12] Yet, man-made, synthetic boron containing compounds have found a wide range of applications. Boronic acids in particular were utilized for various sensing applications, usually enabled by Lewis acid·base interactions (**Chapter 3.4**). Not surprisingly, boronic acids were found to interact with Lewis basic fluoride anions leading to an equilibrium in which OH^- is exchanged for F^- (**Scheme 5.1 A**).[13] This inherent tendency of boron to interact with Lewis basic fluoride anions was thus utilized to enable the development of fluorescent sensors for the detection of fluoride in aqueous media (**Scheme 5.1 B**).[6,13,14]



Scheme 5.1: Boronic acid-based fluoride sensing: Boronic acids interact with inorganic fluoride (**A**) allowing for their application as fluorescent fluoride sensors (**B**).

The selective capture of aqueous fluoride using an engineered protein would not only be of great interest for sensing applications, but also for ^{18}F-radiolabelling. As a positron emitter, fluorine-18 has found wide application in molecular imaging using positron emission tomography (PET).[15] It exhibits favourable properties for clinical imaging, such as a suitable

half-life (110 min), relatively low positron energy (and thus good resolution) and a high degree of positron emission (>97%).[15,16] Typically, fluorine-18 is produced by bombardment of [^{18}O]water with protons in a cyclotron yielding aqueous [^{18}F]fluoride (^{18}O(p,n)^{18}F).[15,16] Due to the low nucleophilicity of aqueous, hydrated fluoride, trapping, elution and azeotropic drying are generally required prior to application in nucleophilic radiolabelling protocols.[15,16] Trifluoroborates provide an exception to this general rule, as they were found to undergo rapid chemical exchange in aqueous media allowing for radiofluorination from aqueous [^{18}F]fluoride via ^{19}F−^{18}F isotope exchange.[17]

5.1.1 Understanding Trifluoroborate Stability

One of the challenges in the application of trifluoroborates for ^{18}F-radiolabelling is the low hydrolytic stability of many simple trifluoroborates. Indeed, the half-life of simple alkyl trifluoroborates is often in the range of minutes under aqueous conditions.[18] *In vivo*, hydrolysis would lead to a low signal-to-noise ratio and a high background signal through bone uptake.[19]

Gaining a mechanistic understanding of the factors that govern trifluoroborate hydrolysis is thus crucial for the rational design of trifluoroborates exhibiting sufficient stability to be applied as a handle for radiolabelling. The mechanistic details of trifluoroborate hydrolysis are complex, involving multiple species and being influenced by various factors that are partially hard to control. Thus, these mechanistic intricacies remain under-explored and the few studies conducted focus on tightly defined systems.[18,20,21] Solvolysis rates of various trifluoroborates span five orders of magnitude leading to half-lives in aqueous media ranging from minutes to years.[22] Hydrolysis of trifluoroborates was found to depend on a plethora of factors, including pH, reaction vessel shape, material and size, and stirring rate.[18] Perrin and co-workers showed that the hydrolysis of aryltrifluoroborates follows pseudo-first-order kinetics limited by the release of the first fluoride ion (**Figure 5.1 A**). Electron-withdrawing groups were found to enhance hydrolytic stability, while electron-donating substituents increase hydrolysis rates. Intermediate species (such as R−BF_2, R−$BF_2(OH)^-$, R−BF(OH) or R−$BF(OH)_2{}^-$) were not observed by NMR.[20] In a later report, the charged species R−$BF_2(OH)^-$ and R−$BF(OH)_2{}^-$ could be observed using a bespoke ESI-MS setup.[21] Further investigations by Lloyd-Jones

and co-workers revealed complex hydrolysis profiles in THF/water mixtures, depending on the substituent attached to the trifluoroborate.[18] While enhancement of trifluoroborate hydrolysis by fluorophiles such as silica gel or glass is known,[23] the same effect was found for glass. A rapid formation of an equilibrium was observed in the absence of glass, while addition of glass powder led to full hydrolysis (**Figure 5.1 A**).[18]

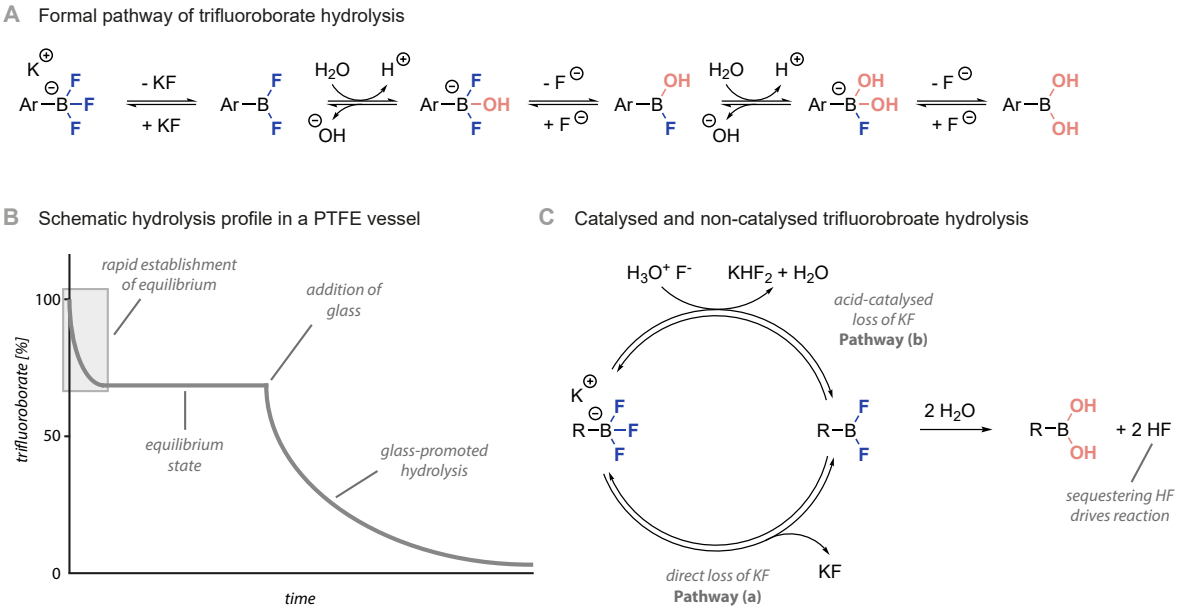


Figure 5.1: Mechanistic considerations for trifluoroborate hydrolysis: (**A**) Formal hydrolysis pathway and possible species. The loss of the first fluoride ion was found to be rate-determining.[20] (**B**) Schematic hydrolysis profile of an aryltrifluoroborate in THF/water and a PTFE reaction vessel. An equilibrium is rapidly established and addition of "HF-sinks", such as glass, is required to drive the reaction towards the boronic acid.[18] (**C**) Two pathways for hydrolysis, the direct loss of KF (a) or the acid catalysed loss of KF (b), were proposed by Lloyd-Jones and co-workers.[18] Figure adapted from Perrin and co-workers and Lloyd-Jones and co-workers.[18,20]

Lloyd-Jones and co-workers proposed two general pathways of hydrolysis: (a) the direct dissociation of KF or (b) the acid-catalysed loss of KF (**Figure 5.1 C**). The hydrolytic stability of a trifluoroborate ($R-BF_3^-$) was linked to the ability of the substituent R to donate into the vacant p-orbital on the boron atom. Substituents capable of π-donation or hyperconjugation would thus decrease hydrolytic stability by stabilizing $R-BF_2$ or $R-B(OH)_2$, whereas increased s-character on substituent R would lead to a stabilized trifluoroborate. For sufficiently stabilized trifluoroborates, pathway (b) is dominant. Consequently, Lloyd-Jones and co-workers found a good correlation between the B—F bond length (serving as a proxy

for donation into the p-orbital, leading to longer B—F bonds) and the observed hydrolytic equilibrium.[18] In search of a simpler prediction of trifluoroborate hydrolytic rates, PERRIN and co-workers reported an excellent correlation between trifluoroborate ($R—BF_3^-$) hydrolysis and pK_a of the analogous carboxylic acid (R—COOH) for non-aromatic trifluoroborates. More acidic carboxylic acids (R—COOH) were found to correlate to analogous trifluoroborates of higher hydrolytic stability.[22]

5.1.2 Design of Stabilized Trifluoroborates

Based on the understanding of the factors that mediate trifluoroborate stability, stabilized derivatives were designed for radiolabelling applications. More acidic boronic acids, such as 3,5-bis(trifluoromethyl)phenyltrifluoroborate, were found to yield trifluoroborates with greatly increased hydrolytic stability (**Figure 5.2 A**).[18]

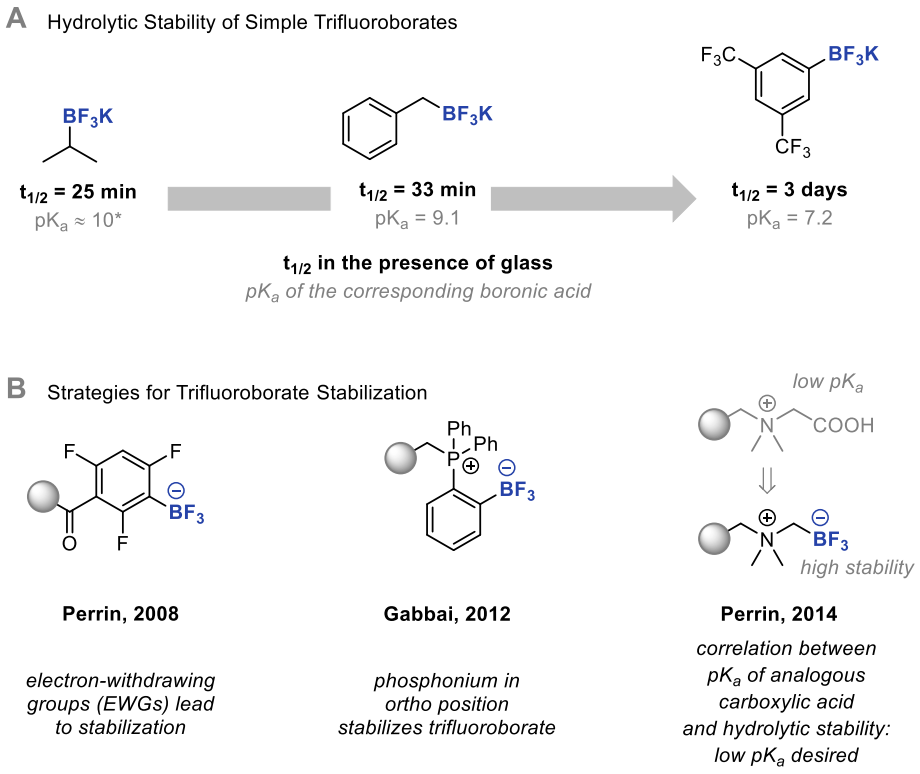


Figure 5.2: Approaches for increasing the hydrolytic stability of trifluoroborates to allow for their application in ^{18}F-radiolabelling: (**A**) Simple trifluoroborates often show low hydrolytic stability ($t_{1/2}$ for hydrolysis in the range of minutes).[18] (**B**) Different strategies were developed to increase the hydrolytic stability of trifluoroborates and enable their use in ^{18}F-radiolabelling.[24–26]

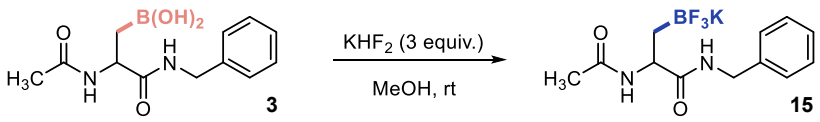
Consequently, the possibility to harness modulation of the chemical environment to design stabilized trifluoroborates was explored and led to the development of several trifluoroborate-based radioprosthetic groups. Attachment of these to a molecule of interest enabled radiofluorination from aqueous [^{18}F]fluoride directly via ^{19}F−^{18}F isotope exchange.[17,19,27]

Examples for these engineered trifluoroborates include aryl trifluoroborates carrying electron-withdrawing groups (preventing weakening of the B−F bond through dislocation of electron density from the π-system into the σ*-orbital),[19,28] the installation of a phosphonium-bearing moiety *ortho* to an aromatic trifluoroborate,[25] as well as zwitterionic, alkylammoniomethyltrifluoroborate ($AMBF_3$) based systems, which were derived from the betaine scaffold and engineered based on the observation that alkyl trifluoroborate analogues of more acidic alkyl carboxylic acids show decreased rates of solvolysis (**Figure 5.2 B**).[26] Trifluoroborate-based radiolabelling approaches were shown to allow for good radiochemical yields with high molar activities (up to 555 GBq·µmol^{-1}) and engineered hydrolytic stability enabled the application of these radiotracers for pre-clinical *in vivo* imaging.[27,29–32]

Therefore, the introduction of a fluoride-capturing moiety into proteins could not only provide new avenues in protein engineering, but could also furnish a versatile handle to directly access site-selectively ^{18}F-radiolabelled proteins under biocompatible conditions employing aqueous [^{18}F]fluoride. Conveniently, such a methodology would avoid azeotropic drying of fluoride as well as time-consuming post-labelling steps.

5.2 Small-Molecule Studies on Trifluoroboronoalanine (3F-Bal)

To evaluate the utility of boronoalanine (Bal) derived trifluoroboronoalanine (3F-Bal), detailed investigations were carried out on the small-molecule model compound Ac-(3F-Bal)-NHBn (**15**).



Scheme 5.2: Synthesis of Ac-(3F-Bal)-NHBn (**15**) from Ac-Bal-NHBn (**3**)

Model compound Ac-(3F-Bal)-NHBn (**15**) was synthesized by treatment of its boronic acid derivative Ac-Bal-NHBn (**3**) with potassium bifluoride (KHF_2) in methanol and purified by Soxhlet extraction (**Scheme 5.2**).

First, the hydrolytic stability of the trifluoroborate was investigated by ^{1}H-NMR (**Figure 5.3 A**). An experimental half-life of $t_{1/2}$ = 48 h was determined for a solution in unbuffered D_2O in a glass NMR tube. The data showed an excellent fit to a two-phase exponential decay equation (R^2 = 0.998) with a half-life of $t_{1/2}$ = 13.4 h for the fast component and a half-life of $t_{1/2}$ = 8.8 days for the slow component (**Figure 5.3 B**).

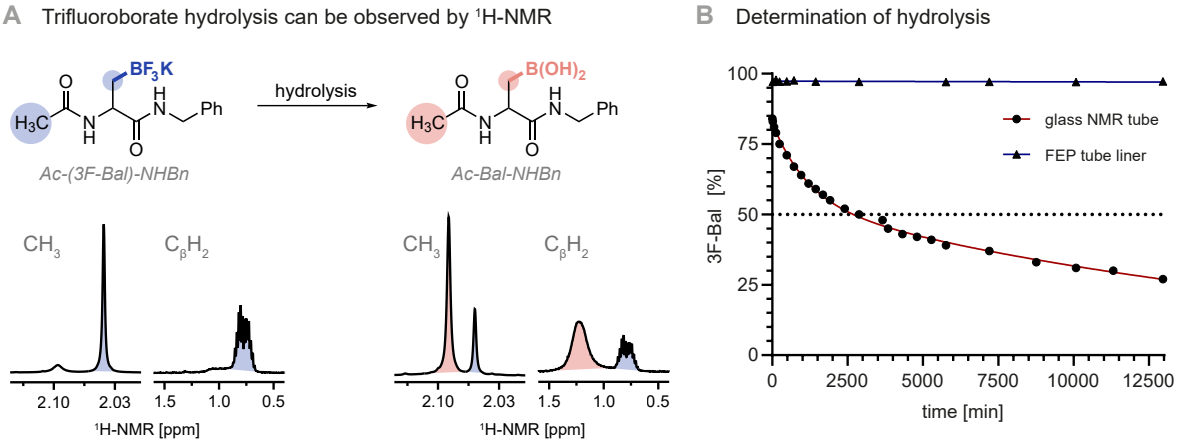


Figure 5.3: Determination of hydrolysis of Ac-(3F-Bal)-NHBn (**15**) by ^{1}H-NMR. (**A**) Trifluoroborate hydrolysis in aqueous medium (D_2O) can be observed by ^{1}H-NMR. Both the acetyl CH_3 group as well as the C_β protons show distinct peaks for the trifluoroborate or boronic acid derivative respectively. (**B**) Trifluoroborate hydrolysis was observed as a function of time. While no hydrolysis can be observed using a FEP tube liner even after 9 days (blue), hydrolysis in a glass NMR tube shows a good fit for a two-phase exponential decay equation (R^2 = 0.998) with a half-life of $t_{1/2}$ = 13.4 h for the fast component and a half-life of $t_{1/2}$ = 8.8 days for the slow component.

While previous reports of simple trifluoroborates showed good fits for one-phase exponential decay functions, the underlying mechanisms for the hydrolysis of trifluoroborates are still not fully understood.[18,20] They seem to be governed by a plethora of parameters, which could give rise to the observation of two different hydrolysis pathways in this specific case. Under similar conditions, when using a FEP (fluorinated ethylene polypropylene copolymer) tube liner preventing the direct contact with the glass surface, no significant hydrolysis was observed within 9 days (**Figure 5.3 B**), while addition of a glass capillary to the sample restored hydrolytic decay. Indeed, silicon oxides such as silica gel or glass are known to significantly

enhance trifluoroborate hydrolysis, likely by sequestering released fluoride and thus shifting the equilibrium towards hydrolysis.[18,23] Next, having shown that trifluoroboronoalanine is stable under aqueous conditions in the absence of fluoride scavengers, the influence of different buffers, additives and amino acids on the equilibrium was investigated. This might have important implications on the choice of (reaction) buffer. Additionally, if different amino acid residues were to affect the equilibrium between trifluoroborate and boronic acid differently, this could allow for at-will programming of trifluoroborate stability simply by choice of a suitable amino acid sequence within which 3F-Bal is embedded. In fact, previous studies on the engagement of Bal with exogenous ligands (diols) showed a clear modulation of binding affinities through adjacent amino acid residues (**Chapter 3.4.1**). A similar modulation of trifluoroborate stability by competition between fluoride binding and engagement with proximal Lewis-basic side chains therefore seems likely.

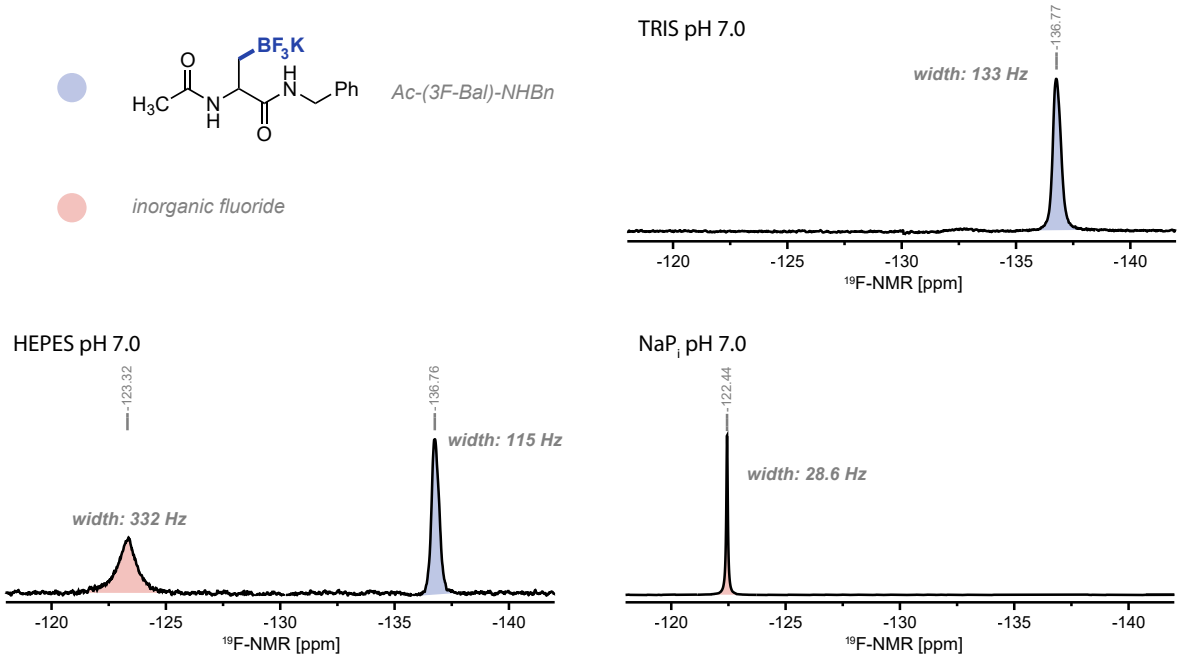


Figure 5.4: ^{19}F-NMR analysis of Ac-(3F-Bal)-NHBn (**15**) in different buffers. Significant differences in peak width were observed in different buffer systems indicating buffer-dependent chemical exchange.

The amount of trifluoroborate present under different buffer conditions was determined by ^{19}F-NMR. Absolute integrals were used for quantification, as released fluoride showed excessive

broadening making the quantification relative to fluoride inaccurate in some cases (**Figure 5.4**). All measurements were conducted in FEP NMR tube liners to prevent glass from promoting hydrolysis.[18] Measurements conducted over a period of time revealed that the equilibrium between fluoride and trifluoroborate was established rapidly enough not to affect any subsequent NMR measurements. The observed line broadening is in agreement with chemical exchange via an equilibrium between inorganic fluoride and free boronic acid on one side and trifluoroborate on the other side.[33] When compared to unbuffered D_2O, significantly smaller ^{19}F-NMR integrals were detected for most buffer systems (**Figure 5.5**). This indicates the likely involvement of buffer constituents in the previously mentioned equilibrium.

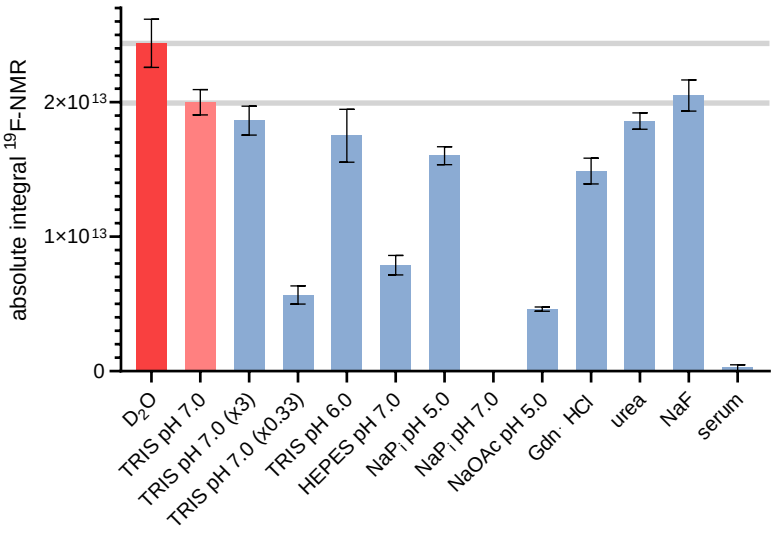


Figure 5.5: Determination of the influence of different buffer systems and additives on the position of the chemical equilibrium between trifluoroborate Ac-(3F-Bal)-NHBn (**15**) and boronic acid Ac-Bal-NHBn (**3**). ^{19}F-NMR measurements were conducted in FEP NMR tube liners using a 16.7 mM solution of Ac-(3F-Bal)-NHBn (**15**) in deuterated solvent. A standard zg30 pulse sequence, 1024 scans and 5 s relaxation delay were used. Additives Gdn·HCl, urea and NaF were assessed in TRIS buffer (100 mM, D_2O) at pH 7.0. Adult bovine serum was supplemented with 10% v/v D_2O to provide a lock signal. The integrals of samples using higher (3x) or lower (x0.33) trifluoroborate concentration were adjusted accordingly.

Interestingly, in different neutral buffers (pH 7.0) a wide range of remaining trifluoroborate was observed. While an absolute integral of 82% of the value detected for unbuffered D_2O was observed for TRIS, only 32% were detected when HEPES was used. Sodium phosphate (NaP_i) buffer led to disappearance of the trifluoroborate peak at pH 7.0, while an integral of 66%

relative to unbuffered D_2O was detected at pH 5.0. Sodium acetate buffer at the same pH value only led to an average integral of 19% relative to D_2O. This indicates that both the buffer substance and the pH value of the solution play a crucial role in determining the position of the chemical equilibrium. Furthermore, the influence of common denaturants and additives was investigated. While a significant reduction of the absolute integral by 26% was observed when 100 mM guanidine hydrochloride was present in the buffer (100 mM TRIS, pH 7.0), only a minor reduction by 7% was observed for urea. Indeed, both guanidine and urea derivatives are known to engage in hydrogen bonding (likely even under aqueous conditions) with fluoride and thus might be able to shift the equilibrium towards hydrolysis by sequestering released fluoride or even promoting fluoride release.[34–36] When the concentration of the trifluoroborate in TRIS buffer was lowered to one third, the value of the concentration corrected integral was only 28% relative to the more concentrated sample. Yet, no significant shift in the position of the equilibrium (towards the trifluoroborate) was observed when the buffer was supplemented with additional sodium fluoride (100 mM). In adult bovine serum, a reduction of the absolute integral by 99% when compared to an unbuffered aqueous solution was observed.

As described previously, trifluoroborate hydrolysis in general is driven by the sequestration of HF formed during this reaction (**Figure 5.1 C**) and can reach completion if the capacity of the "HF-sink" is sufficient.[18,23] The observations made for the position of the equilibrium in different buffers suggests that buffer components might act as "HF-sinks", sequestering fluoride and shifting the equilibrium towards the boronic acid. The direct interaction of buffer components with fluoride is further suggested by the observed, buffer-dependent broadening of the fluoride ^{19}F-NMR peak, which is indicative of chemical exchange (**Figure 5.3**). Besides removing fluoride from the equilibrium, a second mechanism for the modulation of the equilibrium position might be (competitive) binding of buffer components to the boronic acid displacing bound fluoride. Indeed, polyols such as TRIS have previously been observed to interact with boronic acids.[37]

This stability modulation by external factors, and especially the high amount of hydrolysis observed in plasma, indicate that the *in vivo* stability of trifluoroboronoalanine (3F-Bal) could present a major problem for PET imaging applications. In such a scenario, the stabilization of the trifluoroborate by an engineered protein environment will be essential. Such an envi-

ronment could, for example, be presented by a designed hydrophobic pocket, limiting solvent exposure of the trifluoroborate. To confer stabilizing properties, the pocket itself would need to be inert towards the trifluoroborate. The presence of amino acid side chains which could either act as "HF-sinks" via hydrogen bonding to fluoride or promote hydrolysis by competitive binding to the boronic acid (see **Chapter 3**) could hereby be employed to modulate, at will, $^{19}F—^{18}F$ isotope exchange.

To address the possibility of sequence dependent modulation of trifluoroborate hydrolytic stability by proximal amino acid residues, spiking experiments were conducted. A buffered solution of model compound **15** (16.7 mM solution in 100 mM TRIS, pH 7.0) was spiked with all canonical amino acids individually, and the amount of trifluoroboronoalanine present in the mixture was determined by absolute ^{19}F-NMR integrals (**Figure 5.6**).

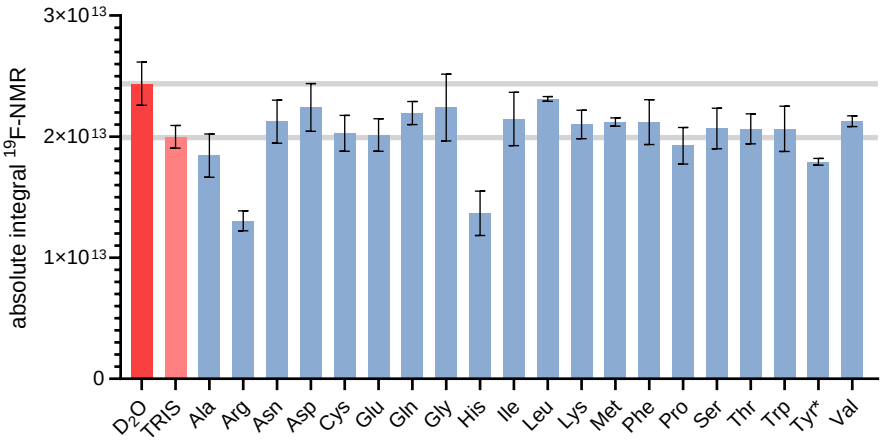


Figure 5.6: Determination of the influence of amino acid residues on the position of the chemical equilibrium between trifluoroborate and boronic acid. A 16.7 mM solution of Ac-(3F-Bal)-NHBn (**15**) in deuterated buffer (100 mM TRIS, pH 7.0, D_2O) was spiked with 10 mM of free amino acid. ^{19}F-NMR measurements were conducted in FEP NMR tube liners using a standard zg30 pulse sequence, 1024 scans and 5 s relaxation delay. *Tyrosine was used at 5 mM due to its limited solubility.

While most amino acids did not affect the equilibrium, samples spiked with arginine (Arg), histidine (His) and tyrosine (Tyr) showed a clear deviation from the control. Spiking with Arg showed a 35% loss in absolute integral, while spiking with His led to a loss of 32% compared to the buffered control. This is in agreement with fluoride sequestration by these moieties (**Figure 5.7 A**). Arg features a guanidine moiety, which can likely sequester released fluoride

via hydrogen bonding.[35,36] Histidine carries an imidazole moiety, which likely plays a similar role. Indeed, both guanidine and imidazole derivatives were reported as fluoride binding motifs in colorimetric fluoride sensors even under aqueous conditions. Fluoride is believed to engage in hydrogen bonding with the guanidine and imidazole moieties, forming a N—H···F hydrogen bond.[38–41] The formation of an arginine—fluoride or a histidine—fluoride complex, consequently, would withdraw fluoride from the equilibrium explaining the loss in integral observed for the trifluoroborate.

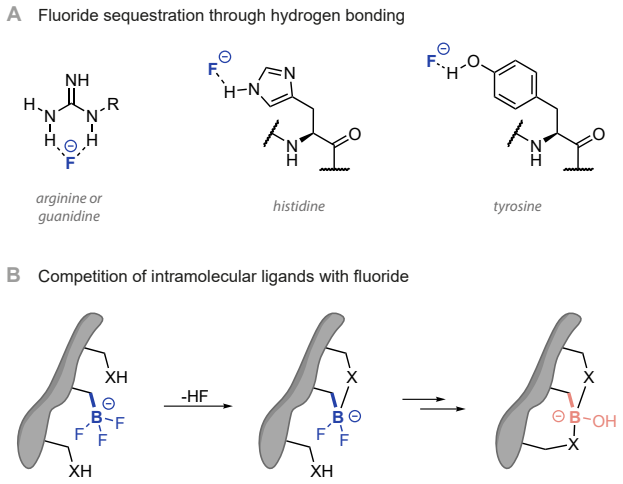


Figure 5.7: Proposed modulation of the hydrolytic equilibrium driven by fluoride sequestration through amino acid side chains or buffer components (**A**) and competitive, intramolecular binding of suitable residues to the central boron atom (**B**).

Spiking with tyrosine led to a 10% loss in integral (**Figure 5.6**). Alcohols are known for their ability to form O—H···F hydrogen bonds with inorganic fluoride and are employed routinely as nucleophilicity-modulating solvents in nucleophilic fluorination reactions.[42,43] While the stability of such hydrogen bonds with phenols like tyrosine was not yet described, this interaction was shown to be strong enough to enable its utilization in colorimetric 'naked-eye' fluoride sensing in aqueous media.[44] Consequently, modulation of the equilibrium between trifluoroborate and boronic acid by tyrosine is plausible (**Figure 5.7 A**). The high degree of hydrolysis observed in serum is in agreement with this observation and likely stems from the presence of several fluoride sequestering species in this complex biological matrix.

Additionally, the possibility for sequence specific intramolecular competition between fluoride and endogenous ligands, such as presented by amino acid residues, should be considered

(**Figure 5.7 B**). The ability of proximal side chains to engage in PLABP could enhance the rate-determining dissociation of the first fluoride ion greatly affecting stability. This should also be taken into account for the rational design of boronoalanine-based fluoride binding pockets.

5.3 Formation of Trifluoroboronoalanine (3F-Bal) on Protein

Trifluoroborate formation was attempted on model protein Histone H3. Attempted LC-MS analysis revealed the presence of the corresponding boronic acid but no trifluoroborate was detected. This is likely due to rapid hydrolysis during data acquisition, making trifluoroborate formation on protein unobservable via mass spectrometry. LC-MS analysis of trifluoroborate formation on model compound Ac-Bal-NHBn (**3**) always resulted in a mixture of trifluoroborate and boronic acid in spite of NMR indicating full conversion to the corresponding trifluoroborate. Full hydrolysis to the corresponding boronic acid seems therefore probable in a more complex protein system. Consequently, trifluoroboronoalanine formation was monitored by ^{19}F-NMR (**Figure 5.8**). Trifluoroborates exhibit a narrow range of ^{19}F-NMR chemical shifts typically between −130 ppm and −140 ppm.[45] Inorganic tetrafluoroborates, which might be formed in the presence of boric acid, can be distinguished easily by their unique appearance in the ^{19}F-NMR spectrum. The tetrafluoroborate anion is typically observed at chemical shifts around −150 ppm and exhibits a unique 1:1:1:1 quartet with ^{19}F—^{11}B coupling constants in the range of $^{1}J_{^{11}\mathrm{B}-^{19}\mathrm{F}} = 1.5 \pm 0.05$ Hz.[46] A second smaller peak is observed 0.05 ppm downfield, which stems from the boron-10 isotopologue $^{10}BF_4{}^-$.[47] Both peaks show a characteristic ratio of approximately 80:20 ($^{11}BF_4{}^-$: $^{10}BF_4{}^-$) reflecting the natural abundance of both boron isotopes.[48] Under aqueous conditions, the formation of the hydroxytrifluoroborate anion (BF_3OH^-) through hydrolysis of the corresponding tetrafluoroborate is possible.[47,49] This anion exhibits a characteristic 1:1:1:1 quartet around −140 ppm showing a coupling constant of approximately 14.5 Hz.[49] Both, the tetrafluoroborate anion and the hydroxytrifluoroborate anion are thus readily distinguishable from organic trifluoroborates. Fluoride and bifluoride can be observed at a wide range of chemical shifts depending on the pH, typically showing an average chemical shift due to fast exchange.[50,51] On Histone H3 Bal10, trifluoroborate formation could be observed by ^{19}F-NMR after treatment with potassium bifluoride (KHF_2)

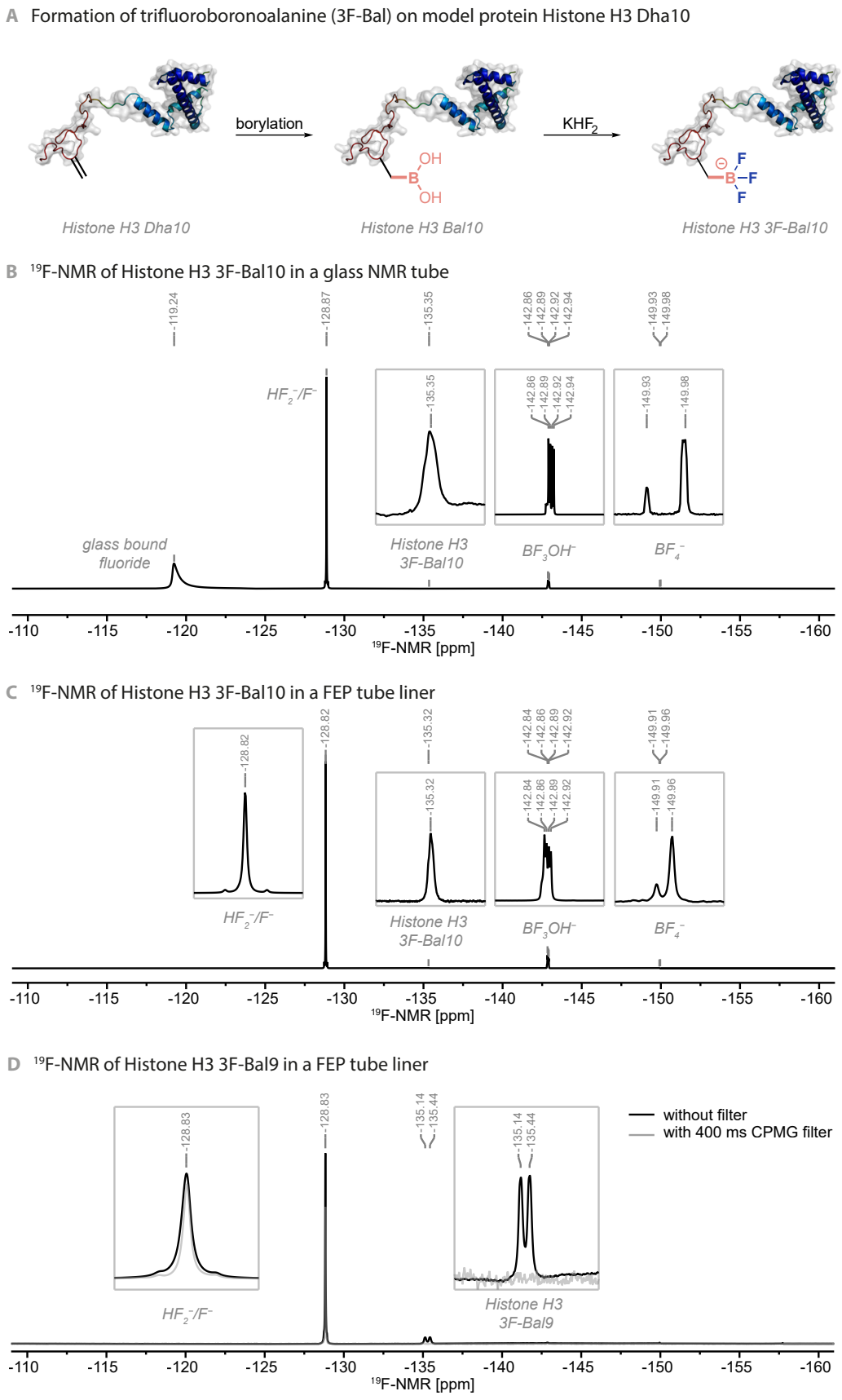


Figure 5.8: Observation of trifluoroborate formation on model protein Histone H3 via ^{19}F-NMR. (**A**) Protein borylation and subsequent trifluoroboronoalanine (3F-Bal) formation. (**B**) Observation of Histone H3 3F-Bal10 in a glass NMR tube and a FEP tube liner (**C**). (**D**) ^{19}F-NMR of Histone H3 3F-Bal9 with and without a CPMG filter.

in acidic buffer (final concentrations: 47 mM KHF_2, 10.5 mM NaP_i, 316 mM Gdn·HCl, pH 4.4, approx. 2 mg/mL protein concentration) (**Figure 5.8 A**). When the NMR spectrum was recorded in a glass NMR tube, peaks corresponding to fluoride/bifluoride as well as to silicon-fluoride compounds obtained by reaction with the glass surface were observed (**Figure 5.8 B**).[51] Cleaner spectra were obtained in FEP NMR tube liners when no fluorophilic glass was present. Under optimized conditions, a single peak corresponding to trifluoroboronoalanine (3F-Bal) was observed at −135.3 ppm (**Figure 5.8 C**). Additionally, peaks corresponding to aqueous fluoride species (fluoride/bifluoride), tetrafluoroborate and hydroxytrifluoroborate (*vide supra*) could be observed. Next, trifluoroborate formation was attempted on Histone H3 Bal9. Two peaks were observed at −135.1 ppm and −135.4 ppm which is in agreement with the formation of two diastereomers (L-Bal and D-Bal) during borylation (**Figure 5.8 D**). Most likely both diastereomers do not exhibit sufficient difference in their chemical shifts for Histone H3 Bal10, leading to a single observable peak in this case. Application of a 400 ms CMPG filter to the same sample of Histone H3 3F-Bal9 confirmed short transversal relaxation rates, which is indicative for large molecular size.[52] This and the observed chemical shifts confirm trifluoronoalanine (3F-Bal) formation on Histone H3 Bal9.

5.4 Conclusion & Outlook

The formation of trifluoroboronoalanine (3F-Bal) was successfully demonstrated on protein, showing that boronoalanine (Bal) does not only confer *de novo* binding function for typical boronic acid ligands such as diols (**Chapter 3**), but also allows for 'capture' of Lewis basic fluoride anions. The introduction of a fluorophilic boron atom via direct C_β−B_γ bond formation enables proteins to bind fluoride through unprecedented and engineered direct covalent fluoride 'capture'. Investigations on a small molecule model provided detailed insights into the intrinsic stability of the newly formed trifluoroborate and how exogenous factors can influence the equilibrium between boronic acid and trifluoroborate. External factors such as pH, buffer substance and denaturants were found to have substantial influences on the position of the equilibrium. Furthermore, screening of the canonical amino acids as additives revealed that arginine and histidine exert a significant influence on this equilibrium. Addition of tyrosine induced a moderate change, while no significant effect was detected for the remaining

canonical amino acids. This is in agreement with the ability of these residues to form hydrogen bonds with fluoride even in an aqueous environment (**Figure 5.7 A**). Formation of such complexes depletes fluoride from the solution and shifts the equilibrium towards the boronic acid. In alignment with previously reported 'Wulff'-type modulation of binding affinities for diol ligands (**Chapter 3**), fluoride 'capture' and consequently fluoride release might be modulated not only by external factors, but also through careful engineering of the local protein environment (**Figure 5.7 B**). While this multifaceted modulation of stability unarguably poses challenges, precise engineering of the trifluoroboronoalanine (3F-Bal) system will likely be able to overcome these. Detailed understanding of 3F-Bal stability will likely allow for utilization of the high number of adjustment points to tailor the system for its intended application. For example, the stability of a trifluoroborate species plays an essential role in its application for ^{18}F-radiolabelling: the compound needs to be labile enough to be susceptible to ^{19}F−^{18}F isotope exchange (under labelling conditions), while the stability must be high enough to efficiently prevent *in vivo* hydrolysis leading to the release of [^{18}F]fluoride and thus high background and low signal-to-noise in PET imaging.[19] The *in vivo* stability might be further increased by exploiting stabilization through limitation of solvent accessibility. Indeed, previous experiments on oxidation of the boronic acid with reactive oxygen species revealed a high degree of stabilization through selective placement of the boronoalanine residue in less accessible sites (**Chapter 3.5**).

The modularity and tunability trifluoroboronoalanine could make this non-canonical amino acid residue a versatile toolkit for the efficient radiolabelling of a wide range of relevant proteins. Borylation and trifluoroborate formation proceed under mild and fully aqueous conditions ensuring the compatibility with a range of proteins. Additionally, there is no need for dry [^{18}F]fluoride and time consuming drying steps can be omitted completely. Finally, a radiolabelling protocol based on the results presented herein is expected to be operationally simple, allowing for the wide application of this *de novo* fluoride 'capture' in a molecular imaging context.

References

[1] C. Dong, F. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O'Hagan, J. H. Naismith, *Nature* **2004**, *427*, 561–565.

[2] M. Mileni, J. Garfunkle, C. Ezzili, B. F. Cravatt, R. C. Stevens, D. L. Boger, *J. Am. Chem. Soc.* **2011**, *133*, 4092–4100.

[3] P. W. Chan, A. F. Yakunin, E. A. Edwards, E. F. Pai, *J. Am. Chem. Soc.* **2011**, *133*, 7461–7468.

[4] X. Y. Pei, K. M. Erixon, B. F. Luisi, F. J. Leeper, *Biochemistry* **2010**, *49*, 1727–1736.

[5] F. Nachon, E. Carletti, M. Wandhammer, Y. Nicolet, L. M. Schopfer, P. Masson, O. Lockridge, *Biochem. J.* **2011**, *434*, 73–82.

[6] M. Cametti, K. Rissanen, *Chem. Soc. Rev.* **2013**, *42*, 2016–2038.

[7] L. Yang, T. Dong, H. M. Revankar, C.-P. Zhang, *Green Chem.* **2017**, *19*, 3951–3992.

[8] C. G. Zhan, D. A. Dixon, *J. Phys. Chem. A* **2004**, *108*, 2020–2029.

[9] N. N. Greenwood, A. Ernshaw in *Chemistry of the Elemements*, Elsevier, **1997**, pp. 139–211.

[10] A. Garavelli, F. Vurro, *Am. Mineral.* **1994**, *79*, 381–384.

[11] R. Thomas, *Am. Mineral.* **2002**, *87*, 56–68.

[12] H. Kameo, S. Sakaki, *Chem. Eur. J.* **2015**, *21*, 13588–13597.

[13] C. R. Cooper, *Chem. Commun.* **1998**, 1365–1366.

[14] Z. Guo, I. Shin, J. Yoon, *Chem. Commun.* **2012**, *48*, 5956.

[15] L. Cai, S. Lu, V. W. Pike, *Eur. J. Org. Chem.* **2008**, *2008*, 2853–2873.

[16] O. Jacobson, D. O. Kiesewetter, X. Chen, *Bioconjug. Chem.* **2015**, *26*, 1–18.

[17] D. M. Perrin, *Acc. Chem. Res.* **2016**, *49*, 1333–1343.

[18] A. J. Lennox, G. C. Lloyd-Jones, *J. Am. Chem. Soc.* **2012**, *134*, 7431–7441.

[19] V. Bernard-Gauthier, J. J. Bailey, Z. Liu, B. Wängler, C. Wängler, K. Jurkschat, D. M. Perrin, R. Schirrmacher, *Bioconjug. Chem.* **2016**, *27*, 267–279.

[20] R. Ting, C. W. Harwig, J. Lo, Y. Li, M. J. Adam, T. J. Ruth, D. M. Perrin, *J. Org. Chem.* **2008**, *73*, 4662–4670.

[21] I. Omari, L. P. E. Yunker, J. Penafiel, D. Gitaari, A. San Roman, J. S. McIndoe, *Chem. – A Eur. J.* **2021**, *27*, 3812–3816.

[22] Z. Liu, D. Chao, Y. Li, R. Ting, J. Oh, D. M. Perrin, *Chem. Eur. J.* **2015**, *21*, 3924–3928.

[23] G. A. Molander, L. N. Cavalcanti, B. Canturk, P. S. Pan, L. E. Kennedy, *J. Org. Chem.* **2009**, *74*, 7364–7369.

[24] R. Ting, C. Harwig, U. Auf Dem Keller, S. McCormick, P. Austin, C. M. Overall, M. J. Adam, T. J. Ruth, D. M. Perrin, *J. Am. Chem. Soc.* **2008**, *130*, 12045–12055.

[25] Z. Li, K. Chansaenpak, S. Liu, C. R. Wade, P. S. Conti, F. P. Gabbaï, *MedChemComm* **2012**, *3*, 1305.

[26] Z. Liu, M. Pourghiasian, M. A. Radtke, J. Lau, J. Pan, G. M. Dias, D. Yapp, K.-S. Lin, F. Bénard, D. M. Perrin, *Angew. Chem. Int. Ed.* **2014**, *53*, 11876–11880.

[27] K. Chansaenpak, B. Vabre, F. P. Gabbaï, *Chem. Soc. Rev.* **2016**, *45*, 954–971.

[28] R. Ting, M. J. Adam, T. J. Ruth, D. M. Perrin, *J. Am. Chem. Soc.* **2005**, *127*, 13094–13095.

[29] A. A. W. L. Wong, J. Lozada, M. L. Lepage, C. Zhang, H. Merkens, J. Zeisler, K.-S. Lin, F. Bénard, D. M. Perrin, *RSC Med. Chem.* **2020**, *11*, 569–576.

[30] H. T. Kuo, M. L. Lepage, K. S. Lin, J. Pan, Z. Zhang, Z. Liu, A. Pryyma, C. Zhang, H. Merkens, A. Roxin, D. M. Perrin, F. xois Bénard, *J. Nucl. Med.* **2019**, *60*, 1160–1166.

[31] X. Yang, Z. Liu, H. Zhang, Z. Li, J. P. Munasinghe, G. Niu, G. Teng, X. Chen, *Eur. J. Nucl. Med. Mol. Imaging* **2018**, *45*, 585–592.

[32] Á. Roxin, C. Zhang, S. Huh, M. L. Lepage, Z. Zhang, K. S. Lin, F. Bénard, D. M. Perrin, *Nucl. Med. Biol.* **2018**, *61*, 11–20.

[33] C. A. Waudby, M. Ouvry, B. Davis, J. Christodoulou, *J. Biomol. NMR* **2020**, *74*, 95–109.

[34] L. Pfeifer, K. M. Engle, G. W. Pidgeon, H. A. Sparkes, A. L. Thompson, J. M. Brown, V. Gouverneur, *J. Am. Chem. Soc.* **2016**, *138*, 13314–13325.

[35] M. R. Silva, J. A. Paixão, A. M. Beja, L. A. Da Veiga, *J. Fluor. Chem.* **2001**, *107*, 117–120.

[36] O. D. Bonner, *J. Phys. Chem.* **1977**, *81*, 2247–2249.

[37] B. Akgun, D. G. Hall, *Angew. Chem. Int. Ed.* **2016**, *55*, 3909–3913.

[38] P. Jayasudha, R. Manivannan, K. P. Elango, *Sens. Actuators B Chem.* **2016**, *237*, 230–238.

[39] R. Manivannan, A. Satheshkumar, K. P. Elango, *New J. Chem.* **2013**, *37*, 3152.

[40] A. Jain, R. Gupta, M. Agarwal, *Synth. Commun.* **2017**, *47*, 1307–1318.

[41] W. Kim, S. K. Sahoo, G.-d. Kim, H.-j. Choi, *RSC Adv.* **2016**, *6*, 7872–7878.

[42] S. Liang, G. B. Hammond, B. Xu, *Chem. Eur. J.* **2017**, *23*, 17850–17861.

[43] D. W. Kim, Jeong, S. T. Lim, M.-H. Sohn, J. A. Katzenellenbogen, D. Y. Chi, *J. Org. Chem.* **2008**, *73*, 957–962.

[44] L. Wang, G. Fang, D. Cao, *J. Fluoresc.* **2014**, *24*, 1757–1766.

[45] R. A. Oliveira, R. O. Silva, G. A. Molander, P. H. Menezes, *Magn. Reson. Chem.* **2009**, *47*, 873–878.

[46] B. N. Chernyshov, V. A. Shcherbakov, R. L. Davidovich, *Spectrosc. Lett.* **1972**, *5*, 421–428.

[47] R. J. Gillespie, J. S. Hartman, M. Parekh, *Can. J. Chem.* **1968**, *46*, 1601–1610.

[48] J. Xiao, Y. K. Xiao, Z. D. Jin, M. Y. He, C. Q. Liu, *Aust. J. Earth Sci.* **2013**, *60*, 431–447.

[49] G. Ochoa, C. D. Pilgrim, J. Kerr, M. P. Augustine, W. H. Casey, *Geochim. Cosmochim. Acta* **2019**, *244*, 173–181.

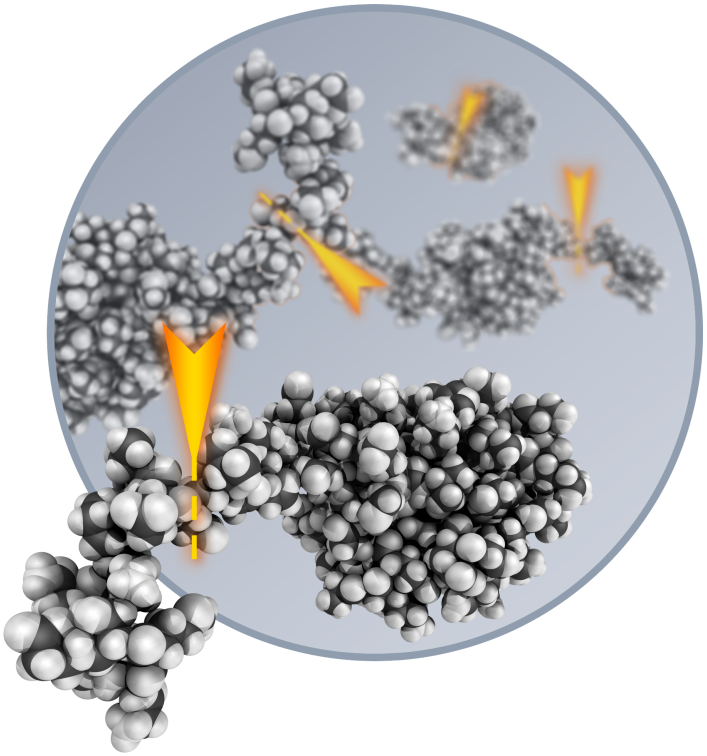
[50] J. B. Gerken, *J. Fluor. Chem.* **2011**, *132*, 68–70.

[51] K. Schaumburg, C. Deverell, *J. Am. Chem. Soc.* **1968**, *90*, 2495–2499.

[52] D. R. Calabrese, C. M. Connelly, J. S. Schneekloth, *Methods Enzymol.* **2019**, *623*, 131–149.

Chapter 6

Mechanistic Insights & Protein Cleavage



Die Welt der chemischen Vorgänge gleicht einer Bühne,
auf welcher sich in unablässiger Aufeinanderfolge Scene um Scene abspielt.
The world of chemical processes resembles a stage,
on which scene after scene is played in relentless succession.

— *Clemens A. Winkler*, Ueber die Entdeckung neuer Elemente im Verlaufe der letzten fünfundzwanzig Jahre und damit zusammenhängende Fragen (1897)

Chapter 6

Mechanistic Insights & Protein Cleavage

Diboron reagents have found widespread applications in synthetic methodology.[1] The superior stability and easy handling of tetraalkoxy diboron reagents such as *bis*(pinacolato)diboron (B_2pin_2) or *bis*(catecholato)diboron (B_2cat_2) have significantly contributed to their vast prevalence in a plethora of borylation protocols.[1–4] Besides their application in borylation chemistry, diboron reagents such as B_2pin_2 or tetrahydroxydiboron ($B_2(OH)_4$) recently gained attention as mild reductants for a wide array of substrates. Diboron-promoted reductions of alkenes, alkynes, carbonyls, α,β-unsaturated ketones, quinolines, nitroarenes, imines, *N*-oxides, sulfoxides, hydroxybenzotriazoles and hydroxyazabenzotriazoles were described.[5–15] Diboron reagents were also employed in the reduction of water for the *in situ* generation of hydrogen.[16–18] Some of these reactions were further shown to proceed under benign, biocompatible conditions allowing for reductions to proceed on DNA or even inside mammalian cells.[19,20] Bertozzi and co-workers demonstrated that non-toxic diboron reagents can penetrate cell membranes and thus activate inactive precursors by *N*-oxide reduction within living cells.[20] During the course of this research, it was further shown that tetrahydroxydiboron can be utilized as a radical initiator for the generation of fluoroalkyl radicals from the corresponding bromides.[21]

6.1 Introduction to Protein Cleavage

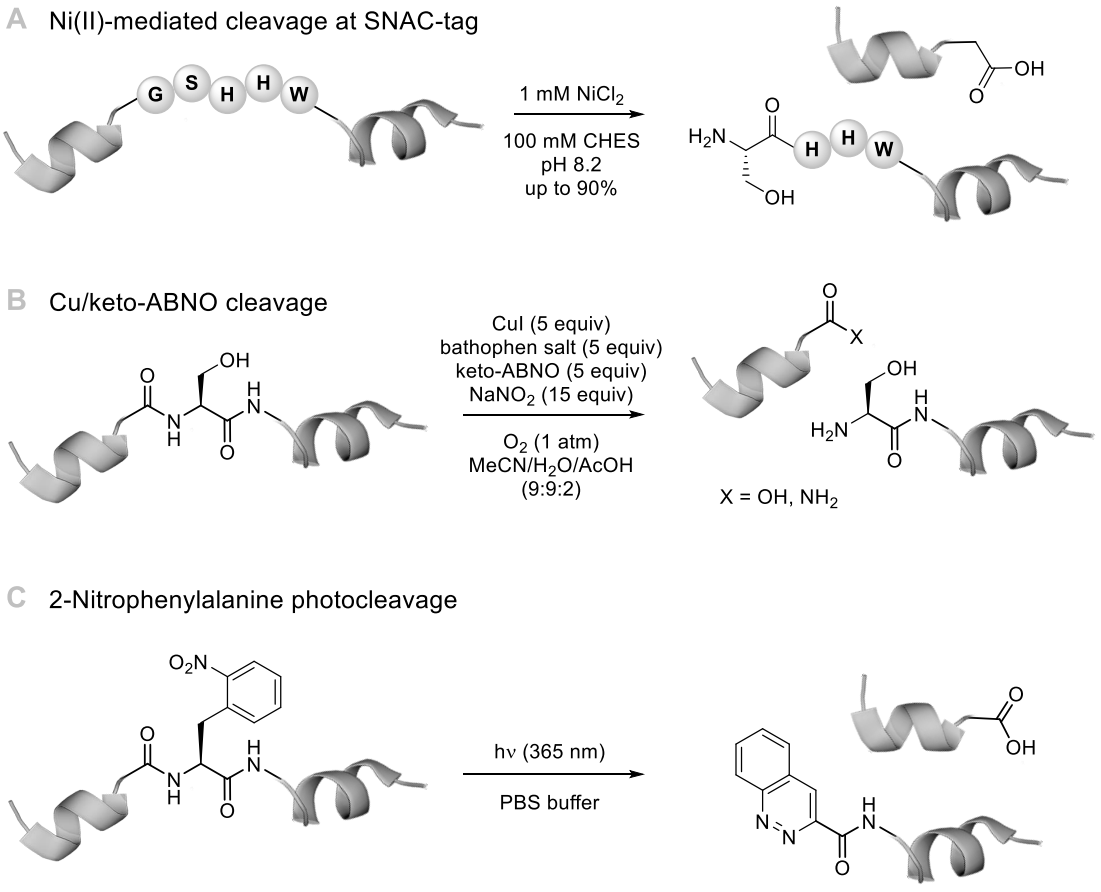
The cleavage of proteins, proteolysis, is an important biological process that fulfils multiple functions. Site-selective proteolysis is one of the most common post-translational modifications and often crucial for protein structure and function.[22,23] Indispensable biological processes such as zymogenesis, prohormone processing and protein degradation are driven by enzymatic protein cleavage and are involved in the regulation of a plethora of physiological and pathological pathways.[22–24] Indeed, proteolysis is crucial in almost all biological pathways, including cell cycle regulation.[25,26]

6.1.1 Enzymatic Protein Cleavage

A wide range of different proteases catalyse protein cleavage. Depending on their biological function, some of them exhibit an extraordinarily high specificity while others are more promiscuous.[24,27] The availability of a broad selection of proteases differing in cleavage sites, efficiency and scope has made them irreplaceable in several areas of biochemical research, such as proteomics and cleavage of recombinant affinity tags.[28–30] More specifically, for mass spectrometry-based proteomics, multiple cleavage events distributed equally amongst the protein sequence and yielding detectable peptides of 7-35 residues in length are required.[31] This is often achieved employing proteases such as trypsin, chymotrypsin or pepsin.[30] In contrast, efficient affinity tag removal depends on the highly selective cleavage at a single, pre-determined site. While affinity tags can increase expression yields, improve solubility and are often crucial for efficient protein purification, these polypeptide motifs can interfere with protein structure and function.[28,32] Therefore, their post-translational removal is often necessary requiring benign and selective methods for efficient affinity tag cleavage. In most cases, highly specific endopeptidases recognizing a unique sequence placed between the affinity tag and the protein of interest are employed.[28,33] The highly specific Tobacco Etch Virus (TEV) protease is one of the proteases commonly employed for affinity tag cleavage. The TEV protease exhibits a very high specificity for a seven residue sequence (ENLYFQ↓G/S), preventing the undesired cleavage at secondary sites and therefore rendering it extremely useful for this task.[34–36]

6.1.2 Chemical Protein Cleavage

In spite of the great importance of site-selective protein cleavage, non-enzymatic protocols remain scarce and are practically not employed. The development of purely chemical methodologies is challenging, as such cleavage reactions would need to be highly selective but also proceed in purely aqueous buffer, ideally at or below room temperature, at benign pH and without any reagents that could affect protein fold or function. Consequently, only very few procedures have been published that aim at filling this gap. Typical chemical cleavage protocols such as cyanogen bromide (BrCN) cleavage (*C*-terminal of methionine) lack selectivity and lead to protein denaturation and decomposition.[37,38]



Scheme 6.1: Contemporary methods for the selective chemical cleavage of proteins: (**A**) SNAC-tag for Ni(II)-mediated cleavage at serine residues. (**B**) Cu/keto ABNO mediated oxidative cleavage at serine residues. (**C**) Backbone photocleavage using the genetically incorporated non-canonical amino acid 2-nitrophenylalanine.

More recent approaches utilise a genetically engineered tag (SNAC-tag), the incorporation of an unnatural amino acid (2-nitrophenylalanine) or a Copper promoted system (**Scheme 6.1**).[39–41] The SNAC-tag (sequence-specific nickel-assisted cleavage) (**Scheme 6.1 A**) was shown to allow for efficient Ni(II)-promoted cleavage *N*-terminal of serine within a −GSHHW− sequence motif. The method is driven by coordinative binding of Ni(II) within the cleavage motif and a subsequent *N*-to-*O* acyl shift of the glycine residue located *N*-terminal of serine. Hydrolysis of the resulting ester yields the cleaved protein. This method was shown to proceed well under benign ambient conditions (up to 90% conversion after 18 h) but does – comparable to enzymatic methods – require the introduction of an engineered tag.[39] As a tag-free method, a protocol using a copper−organoradical conjugate for the serine-selective cleavage of peptides and proteins was reported (**Scheme 6.1 B**). CuI, keto-ABNO (9-azabicyclo[3.3.1]nonan-3-one *N*-oxyl), bathophenanthrolinedisulfonic acid disodium salt and sodium nitrite ($NaNO_2$) were found to promote cleavage *N*-terminal of serine residues under oxygen atmosphere. While the authors were able to demonstrate cleavage on a small protein (Ubiquitin), the large amounts of organic co-solvents required (MeCN/H_2O/AcOH 9:9:2) will likely forbid the use of this methodology on larger, functional proteins.[41] Further methods for the serine-selective cleavage include the application of *N*,*N*'-disuccinimidyl carbonate (DSC) or scandium(III) triflate, but are limited to peptides.[42,43] The introduction of the non-canonical amino acid 2-nitrophenylalanine was shown to allow for the photocleavage of proteins (**Scheme 6.1 C**). Cleavage leading to a *C*-terminal carboxylate and a *N*-terminal cinnoline moiety was observed upon irradiation at 365 nm. While this methodology allows for spatiotemporal control over protein cleavage, only low cleavage efficiencies (approx. 30%) were reported.[40] The small number of chemical protocols available for the selective cleavage of proteins as well as the disadvantages of the published methods are a testament to the challenges encountered developing such methodologies. Site-selective cleavage at a single canonical amino acid residue is difficult to achieve, while just the introduction of a non-canonical, possibly cleavable residue itself can be equally demanding. Furthermore, chemical protocols often rely on organic co-solvents and do not proceed under protein-compatible, fully aqueous conditions. Consequently, a cleavage protocol operating under these conditions could be of high synthetic utility.

6.2 Mechanistic Insights into Protein Borylation

Previously, an efficient methodology for the borylation of dehydroalanine (Dha) residues in intact proteins was developed (**Chapter 2**) and the utility of the resulting boronoalanine (Bal) residue for the introduction of *de novo* function was explored (**Chapter 3**, **Chapter 4** and **Chapter 5**). Yet, mechanistic details of this transformation remained underexplored. To further mechanistic understanding of protein borylation, additional experiments were conducted on small molecule model Ac-Dha-NHBn (**1**) and model protein Histone H3 Dha10.

6.2.1 Small Molecule Studies

While the borylation of Ac-Dha-NHBn (**1**) to yield small molecule model compound Ac-Bal-NHBn (**3**) proceeded well at milligram scale, initial attempts to conduct the reaction at gram scale failed (**Figure 6.1**). The reaction was found to proceed much slower and the main product was identified as Ac-Ala-NHBn (**16**) stemming from reduction of the dehydroalanine precursor (**Figure 6.1 A**). Indeed, diboron reagents were shown to act as powerful reductants by various mechanisms.[5–14,16–18] Interestingly, when reduction was observed, the reaction mixture retained its black colour for multiple days (**Figure 6.1 B**). In contrast, when the reaction was conducted at milligram scale, the reaction mixture turned colourless after a few hours. Subsequently, it was found that gram scale reactions proceeded smoothly if they were conducted under vigorous stirring open to air and not in a capped flask (**Figure 6.1 C**). This indicates that the presence of oxygen might be crucial for borylation or detrimental for reduction. This hypothesis was confirmed by conducting milligram scale (10 mg, 45.8 µmol) reactions under nitrogen, oxygen and hydrogen atmosphere in thoroughly degassed D_2O (**Table 6.1**).

Table 6.1: Borylation reaction under nitrogen, oxygen and hydrogen atmosphere.

Entry	Atmosphere	Ac-Bal-NHBn [%]*	Ac-Ala-NHBn[%]*
1	nitrogen	11	88
2	oxygen	73	0
3	hydrogen	73	27

* Yields were determined by ^{1}H-NMR against maleic acid as internal standard.

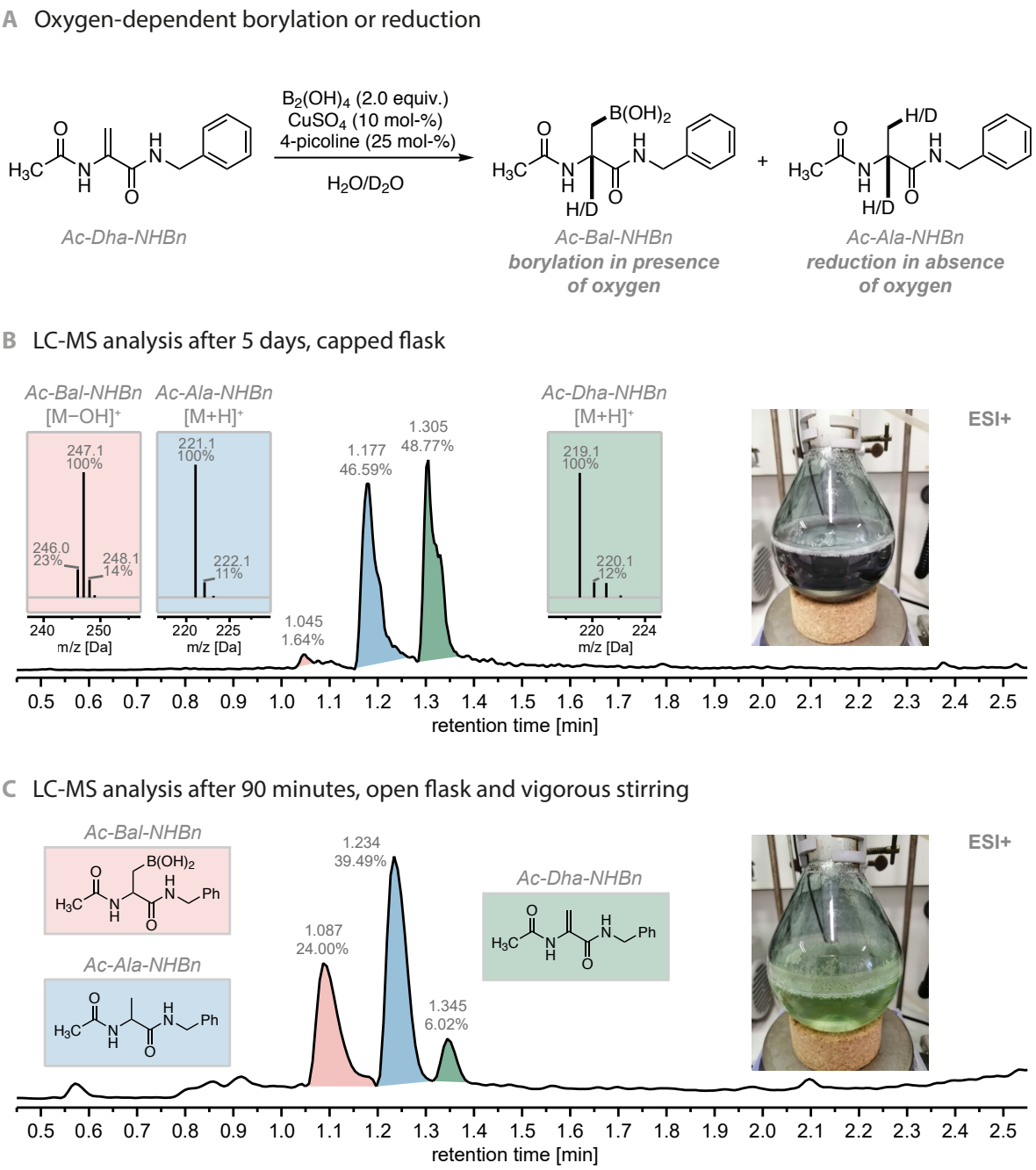


Figure 6.1: Successful borylation of model compound Ac-Dha-NHBn (**1**) depends on the presence of oxygen. (**A**) Presence or absence of oxygen allows for switching between borylation and reduction. (**B**) On larger scale (1 g substrate), only traces of borylation were observed after 5 days and the major product was Ac-Ala-NHBn (**16**) stemming from reduction of Ac-Dha-NHBn (**1**). The reaction mixture stayed black. (**C**) After the same mixture was treated with another two equivalents of tetrahydroxydiboron and stirred vigorously open to air for 90 minutes, a clear solution was observed. LC-MS analysis indicated that most of the remaining substrate was converted to the corresponding boronic acid.

It should be noted that thorough degassing by multiple "freeze-pump-thaw" cycles was found to be absolutely crucial on this small scale, as borylation was observed to proceed readily in the presence of residual oxygen. As expected, mainly reduction product Ac-Ala-NHBn-d_2 (**16-d_2**) was observed under nitrogen atmosphere (88%, **Table 6.1**, entry 1). Consistent with the required presence of oxygen for efficient borylation or its inhibitory effect on reduction, no reduction was observed under oxygen atmosphere while good conversion to Ac-Bal-NHBn-d_1 (**3-d_1**) could be detected (73%, **Table 6.1**, entry 2). The decreased yield when compared to atmospheric conditions (99%) might be due to decomposition of tetrahydroxydiboron, Ac-Dha-NHBn (**1**) or Ac-Bal-NHBn (**3**) under oxygen atmosphere. Under hydrogen atmosphere, a mixture of borylation and reduction was observed (**Table 6.1**, entry 3). This surprising finding can be due to minor oxygen contamination enabling borylation on small scale. For reductions under nitrogen as well as under hydrogen atmosphere, the reaction mixture retained the black colouration initially observed (**Figure 6.2 A**).

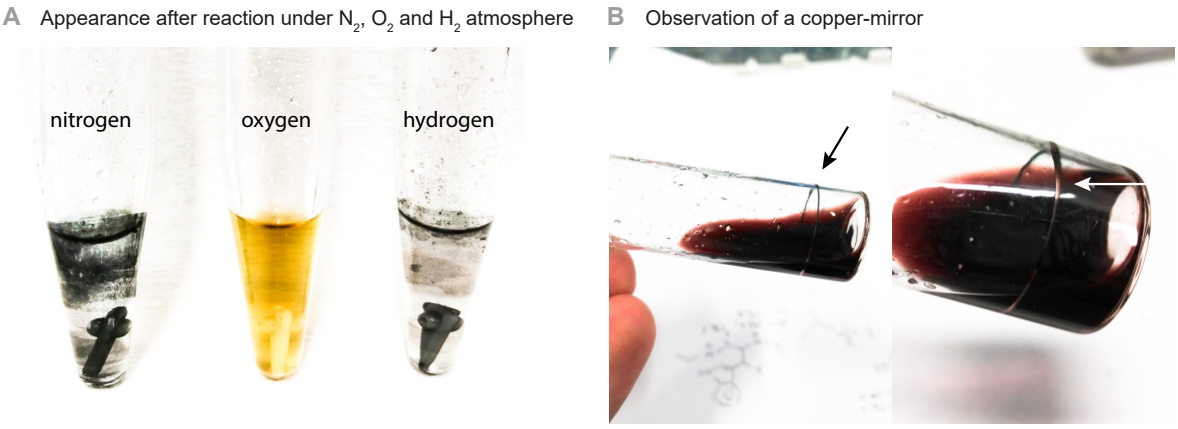
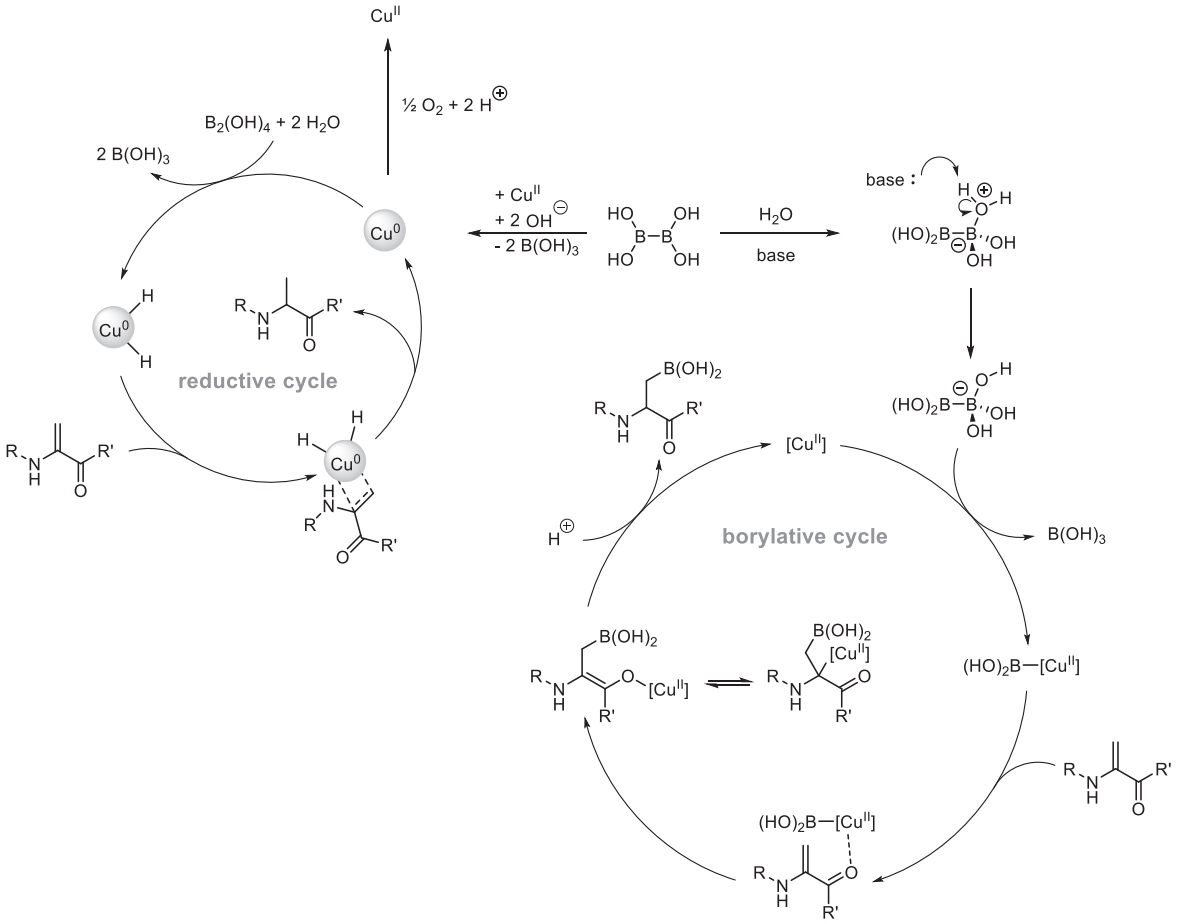


Figure 6.2: Appearance of borylation reactions under nitrogen, oxygen and hydrogen atmosphere (**A**) and formation of a copper-mirror by a mixture of $CuSO_4$, $B_2(OH)_4$ and 4-picoline in water (**B**).

In both cases, the formation of both deuterated species Ac-Bal-NHBn-d_1 (**3-d_1**) and Ac-Ala-NHBn-d_2 (**16-d_2**) was observed. To aid further characterization, Ac-Ala-NHBn-d_2 (**16-d_2**) was synthesized on preparative scale (50 mg substrate) under nitrogen atmosphere and isolated in a yield of 73%. Analysis by ^{1}H-NMR spectroscopy confirmed high degrees of deuterium incorporation (97% at C_α and 91% at C_β) which is in agreement with formal transfer deuteration from D_2O. While tetrahydroxydiboron-based protocols for the hydrogenation of alkenes were

described under palladium catalysis, no such protocol has been reported using copper.[11,16,18] In spite of copper-catalyzed (transfer)hydrogenations being known in the literature,[44–48] completely excluding the responsibility of other species for the reduction would be challenging and beyond the scope of this work. Indeed, trace impurities can often be sufficient for catalytic transformations.[49] Furthermore, slow reduction could be observed under anaerobic conditions even in the absence of copper, making a metal free reduction mechanism possible.

The mechanism of the Cu(II)-catalyzed borylation of α,β-unsaturated carbonyl compounds was previously studied in detail utilizing both experimental and computational methods (**Scheme 6.2**).[50,51]



Scheme 6.2: Proposed mechanism for Cu-catalyzed borylation[50,51] and reduction of dehydroalanine residues. A reductive 'off'-cycle is proposed involving the formation of finely dispersed Cu(0) which leads to the observation of persistent black colouration under anaerobic conditions.

The reaction is believed to entail the formation of a copper-boron species, likely aided by base activation and *in situ* generation of an sp^2—sp^3 diboron reagent readily undergoing transmetalation with Cu(II). After coordination of this copper-species to the substrate carbonyl, a C—B bond and a copper enolate are formed. Subsequent protonation yields the product and releases the Cu(II) catalyst.[50,51] While this mechanism seems plausible for the dehydroalanine system, it cannot explain the observation of reduction under anaerobic conditions (**Scheme 6.2**). Therefore, in accordance with this observation, the parallel existence of an 'off'-cycle for a Cu(0)-mediated reduction is proposed. Tetrahydroxydiboron is hypothesized to reduce Cu(II) to Cu(0) under the reaction conditions. In fact, in the absence of substrate and at high concentration the formation of a copper-mirror can be observed on glass vials (**Figure 6.2 B**) and the black coloration observed might stem from colloidal elemental copper (formed through reduction of $CuSO_4$ by $B_2(OH)_4$) supporting this claim. The presence of particles was confirmed by membrane filtration and the particle size was estimated between 800 nm and 450 nm. Under aerobic conditions, Cu(0) is expected to be readily oxidized to Cu(II) again. In contrast, under anaerobic conditions this re-oxidation will not take place. This could then lead to the formation of a copper—H species from tetrahydroxydiboron and water (possibly under base-activation), which might be the active reductant.

6.2.2 Tetrahydroxydiboron-Promoted Reduction of Dha on Protein

Having explored the reduction of dehydroalanine on a small molecule model substrate, the reduction of Dha in a protein context was investigated. Model protein Histone H3 Dha10 was incubated with tetrahydroxydiboron (250 equiv), copper sulphate (5 equiv) and 4-picoline (12.5 equiv) under inert atmosphere (nitrogen-filled glove box, <6 ppm O_2) and analysed via LC-MS (**Figure 6.3**).

Significantly lower conversion to Histone H3 Bal10 was observed when compared to reactions conducted under ambient conditions. The remaining reactivity towards borylation might be a result of residual trace oxygen. This reaction was conducted in both H_2O and D_2O and the molecular masses observed by LC-MS were compared. Comparison of the masses of both Histone H3 Bal10 mutants revealed a difference of 1 Da consistent with deuteroboration and

hydroboration, respectively. When both reduction products were compared, a mass difference of 2 Da was observed. This observation is consistent with the incorporation of two deuterium atoms as predicted by previous small molecule investigations (*vide supra*).

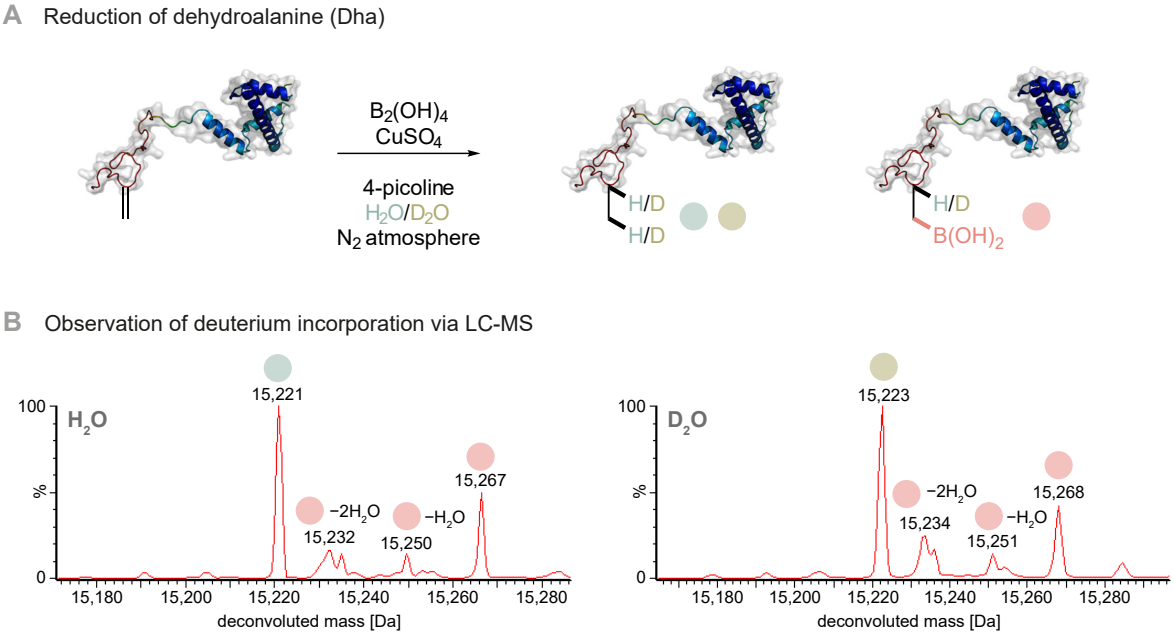
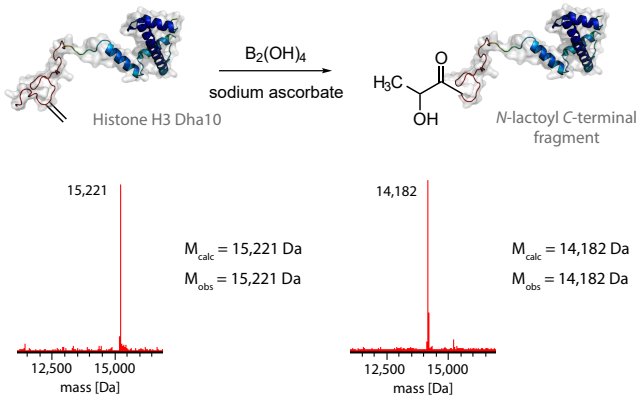


Figure 6.3: Tetrahydroxydiboron-promoted reduction of Dha on protein. (**A**) Transfer hydrogenation of dehydroalanine on model protein Histone H3 Dha10 allows for selective deuterium incorporation. (**B**) LC-MS analysis reveals incorporation of a single deuterium atom in the borylated product and two deuterium atoms in the reduced product.

6.3 Diboron-Promoted, Site-Selective Radical Protein Cleavage

Surprisingly, site-selective protein cleavage at Dha was observed when Histone H3 Dha10 was treated with tetrahydroxydiboron in the absence of copper and under ambient atmosphere (**Scheme 6.3**). The formation of a product with a molecular mass corresponding to a *C*-terminal fragment carrying a *N*-terminal lactoyl moiety was detected. This observation and the lack of published protocols for the site-selective chemical cleavage of native proteins prompted further studies investigating the utility of this novel transformation.



Scheme 6.3: Cleavage of Histone H3 Dha10 in the presence of $B_2(OH)_4$: The formation of a putative a *C*-terminal fragment carrying a *N*-terminal lactoyl moiety was detected by LC-MS.

6.3.1 Reaction Optimization

As significant protein oxidation was observed during initial experiments, different additives were screened for their ability to inhibit unwanted oxidative damage while allowing for selective cleavage. Transition metal reductants iron(II) sulphate ($FeSO_4$) and manganese(II) sulphate ($MnSO_4$), radical scavenger TEMPOL (4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl) and antioxidant sodium ascorbate were employed as additives and their time-dependent influence on protein cleavage and oxidation was observed by mass spectrometry (**Figure 6.4**). Next, the influence of the concentration of the antioxidant sodium ascorbate was investigated. An increase of the amount of sodium ascorbate employed was found to significantly reduce oxidative damage while also inhibiting cleavage (**Figure 6.5 A**). Appropriate choice of antioxidant concentration should therefore be considered on a case-by-case basis. Furthermore, the role of the boron source was investigated (**Table 6.2**). Different diboron reagents (tetrahydroxydiboron, B_2cat_2 and B_2pin_2), borohydrides ($NaBH_4$ and $NaCNBH_3$) and boranes (HBpin and HBcat) were screened for their ability to induce site-selective protein cleavage. The reaction was found to proceed well with both tetrahydroxydiboron ($B_2(OH)_4$) and *bis*(catecholato)diboron (B_2cat_2), while *bis*(pinacolato)diboron (B_2pin_2) gave drastically reduced conversions and required the use of organic co-solvent for solubilisation. The use of neither the borohydrides nor the boranes employed in this screening resulted in any of the desired cleavage products. These results are in agreement with the proposed mechanism (*vide infra*).

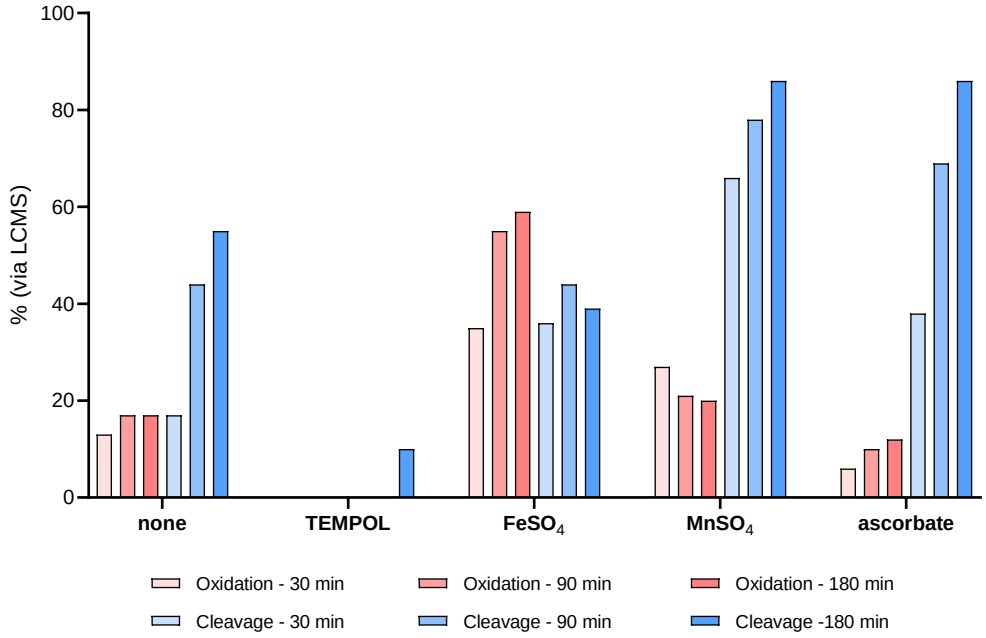


Figure 6.4: Screening of additives for the inhibition of protein oxidation under cleavage conditions. Conditions: Histone H3 Dha10 was dissolved in water at a concentration of 1 mg/mL and treated with additive (50 equiv) and tetrahydroxydiboron (200 equiv) using freshly prepared aqueous stock solutions.

Table 6.2: Boron reagents for the site-selective cleavage of proteins.

Entry	Reagent	Conversion [%]*
1	$B_2(OH)_4$	79
2	B_2cat_2	80
3	B_2pin_2	31
4	$NaBH_4$	0
5	$NaCNBH_3$	0
6	HBpin	0
7	HBcat	0

* Conditions: Histone H3 Dha10 was dissolved in water at a concentration of 1 mg/mL and treated with sodium ascorbate (50 equiv) and reagent (200 equiv) using freshly prepared aqueous stock solutions. The reactions were incubated at room temperature for 3 h before being analysed by LC-MS. 10% v/v DMF was employed as organic co-solvent to solubilize B_2pin_2.

Having established tetrahydroxydiboron as the most suitable diboron reagent (due to its solubility and cleavage efficiency), the temperature dependence of the reaction was investigated (**Figure 6.5 B**). The reaction was found to proceed best at elevated temperature (37°C),

where full conversion was observed after 6 hours on model protein Histone H3 Dha10. At room temperature, 86% conversion were observed after 6 hours and >95% after a day. In both cases the reaction showed high initial rates and was slowed down towards completion. Cleavage was found to even proceed at temperatures as low as 4°C, albeit at a much slower rate when compared to room temperature. After one day, 73% conversion were observed and the reaction was found to proceed to completion (>95% conversion) within 4 days.

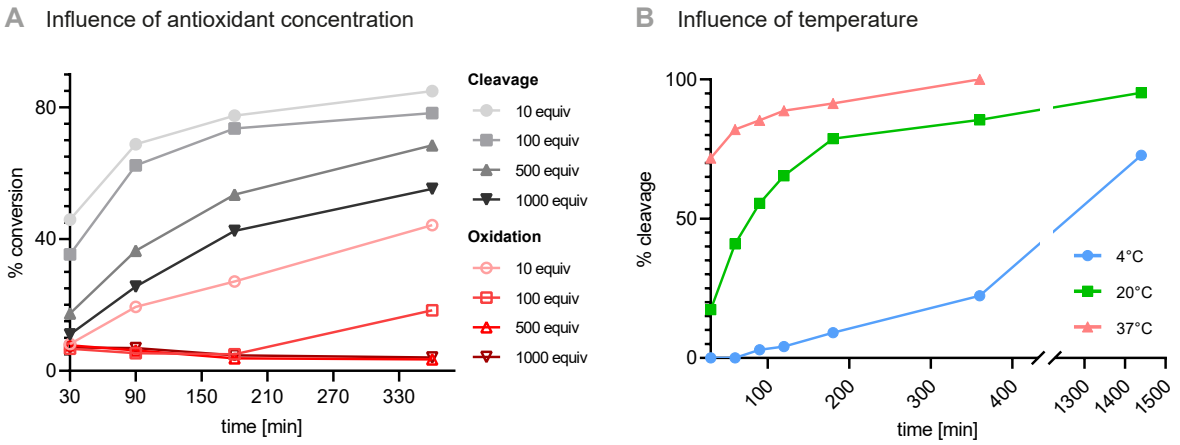


Figure 6.5: Influence of antioxidant concentration and temperature on the cleavage of Histone H3 Dha10. (**A**) Influence of the concentration of antioxidant sodium ascorbate on cleavage and oxidative damage as observed by LC-MS. (**B**) Temperature- and time-dependent progression of cleavage as observed by LC-MS. Conditions: Histone H3 Dha10 was dissolved in water at a concentration of 1 mg/mL and treated with sodium ascorbate (variable for A, 50 equiv for B) and tetrahydroxydiboron (200 equiv) using freshly prepared aqueous stock solutions.

To further confirm the synthetic utility of this transformation, cleavage was attempted in a wide range of relevant buffers (sodium phosphate, TRIS, HEPES, sodium acetate, sodium borate) at varying pH (pH 6 – pH 9) (**Figure 6.6**). Cleavage was found to proceed satisfactory in most buffers. In general, good conversions were achieved under mildly basic conditions (pH 8 and pH 9). Borate buffer at pH 8 (88% conversion after 3 hours at room temperature) was found to yield the highest conversions and thus be most suitable for the cleavage of Histone H3 Dha10. As the cleavage reaction did not proceed under acidic conditions, acidification can be used to halt the reaction if required.

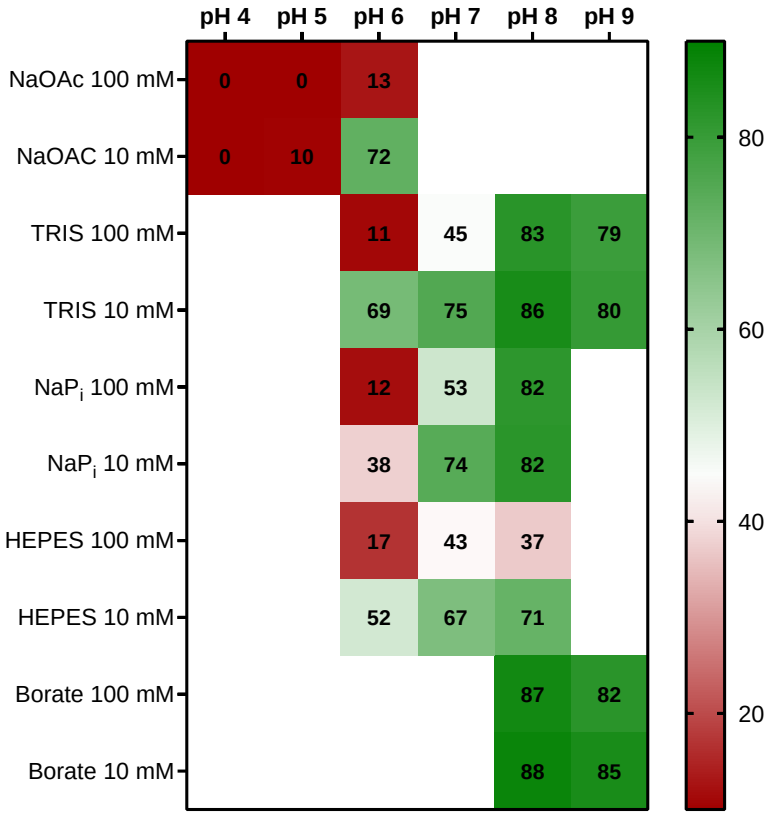


Figure 6.6: Determination of cleavage conversions in a range of relevant buffers: Conversions are given in % and were determined after 3 hours at room temperature. Conditions: Histone H3 Dha10 was dissolved in buffer at a concentration of 1 mg/mL, treated with sodium ascorbate (50 equiv) and tetrahydroxydiboron (200 equiv) using freshly prepared aqueous stock solutions, incubated at room temperature and analysed by LC-MS.

6.3.2 Scope of Radical Protein Cleavage

Having established optimal cleavage conditions, the synthetic utility of the methodology was put to test by expanding the substrate scope to a range of different proteins.

A range of proteins differing in size, fold and accessibility of the cleavage site were subjected to the previously optimized cleavage conditions. Various mutants of Histones H3 and H4 (small α-helical proteins with accessible tails),[52] Npβ (pentapeptide repeat, β-helical),[53] Annexin V (globular α-helical protein)[54] and preSUMO1 (SUMO1 precursor protein consisting of α-helices and β-sheets)[55] were employed as substrates (**Figure 6.7**). Excellent cleavage efficiencies (>95%) were observed for cleavage sites placed within the highly accessible tails of Histone H3 and Histone H4 (*C*-terminal FLAG-HA Histone H3 Dha4 >95%; Histone H3 Dha9

>95%; Histone H3 Dha10 >95%; Histone H4 Dha16 >95%). Less accessible sites such as Npβ Dha61 (85%) or Annexin V Dha316 (39%) required elevated temperatures (45°C). This is in agreement with increased accessibility through partial melting (Npβ Cys61: T_M = 51°C; Annexin V: T_M = 52°C). For preSUMO1 Dha51, some conversion was observed after prolonged incubation at room temperature (48 h, 14%). This behaviour is mimicking the behaviour of proteases which also highly depend on the accessibility of a potential cleavage site (**Table 6.3**).[56]

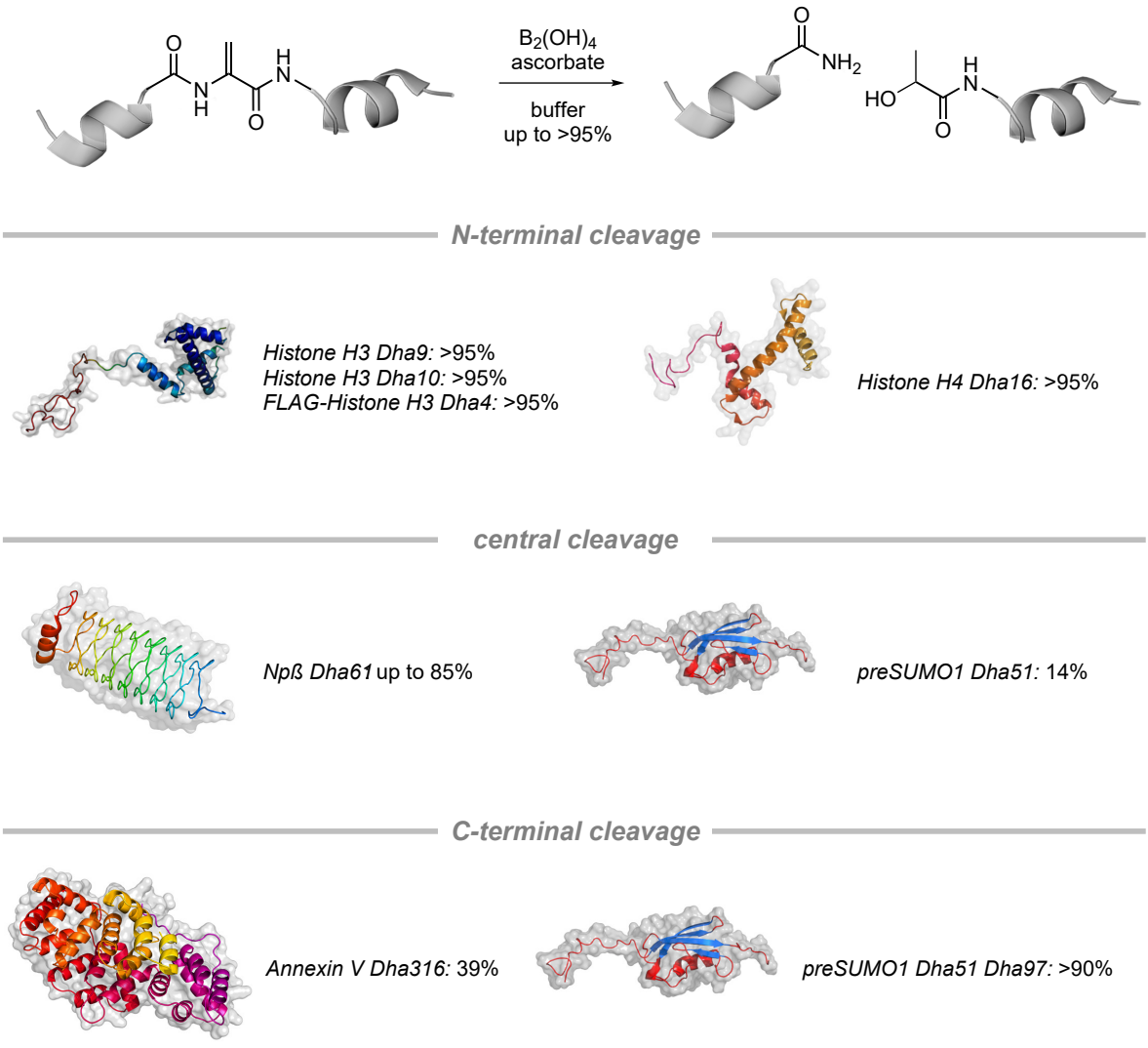


Figure 6.7: Protein scope of the Dha-selective, diboron-promoted radical protein cleavage. Conversions were determined by LC-MS, for preSUMO1 Dha51 conversions were determined by densitometric SDS-PAGE using Coomassie staining.

Table 6.3: Observed cleavage conversions depend on cleavage-site accessibility under the employed conditions.

Entry	Protein	rASA [%]	Conditions	Conversion [%]
1	Histone H3 Dha9	42	23 h, rt	>95%
2	Histone H3 Dha10	79	24 h, rt	>95%
3	FLAG-Histone H3 Dha4	61	3 h, rt	>95%
4	Histone H4 Dha16	69	17 h, rt	>95%
5	Npβ Dha61	26	24 h, 45°C	85%
6	preSUMO1 Dha51	2	48 h, rt	14%*
7	preSUMO1 Dha51 Dha97	103	18 h, rt	>90%*
8	Annexin V Dha316	12	3 h, 45°C	39%

Relative solvent-accessible surface areas (rASA) were calculated for the respective Cys mutants relative to the accessibility of Cys within the tripeptide Gly-Cys-Gly. Calculations were conducted using the Lee & Richards algorithm, NACCESS atom radii and a probe size of 1.4 Å size as implemented in FreeSASA.[57,58] Conversions were determined by LC-MS unless stated otherwise. * Conversion determined by SDS-PAGE, Coomassie staining and densitometry. Image analysis was performed with ImageJ 1.53f.[59]

The dependency on site accessibility and the resulting possibility to selectively cleave at one Dha residue over another was further investigated by the preparation of a preSUMO1 mutant featuring two Dha residues. An additional Cys residue was introduced into preSUMO1 Cys51 by site-directed mutagenesis and placed at site 97. Enzymatic SUMO1 maturation by SENP1 naturally involves cleavage *N*-terminal of site 97.[60] A strategically placed Dha residue at site 97 therefore should allow for chemical SUMO1 maturation. After conversion to preSUMO1 Dha51 Dha97 employing Mukaiyama's reagent,[61] cleavage was conducted in TRIS buffer (pH 8.0). Analysis by LC-MS revealed >90% cleavage at Dha97 leading to chemically matured SUMO1 (**Figure 6.8 A**). The superior accessibility of site 97 over site 51 enabled retention of Dha51 and allowed for its use as a handle for subsequent modifications via thia-Michael addition of β-mercaptoethanol. The selectivity of cleavage at Dha97 over cleavage at Dha51 was further confirmed by SDS-PAGE and Coomassie staining (**Figure 6.8 B**). Densitometric analysis confirmed only minor cleavage at Dha51 (22%).

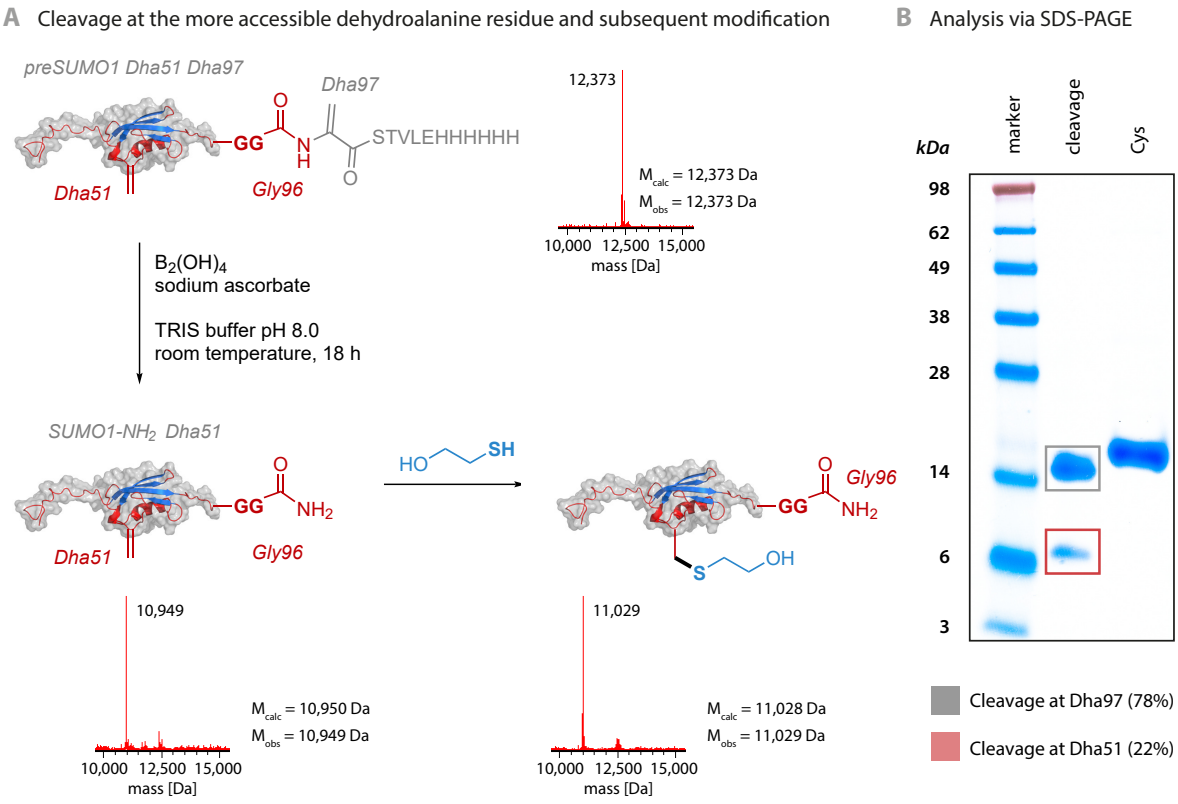


Figure 6.8: Cleavage of preSUMO1 Dha51 Dha97 at the most accessible dehydroalanine residue. (**A**) Selective cleavage allows for subsequent modification via 'tag-and-modify' approach. (**B**) Cleavage at site 97 was determined to be 78% selective using SDS-PAGE and densitometry. Image analysis was conducted with ImageJ 1.53f.[59]

6.3.3 Analysis of the Cleavage Products: Identification of Terminal Modification

Based on LC-MS analysis and the putative mechanism of the diboron-promoted protein cleavage, *C*-terminal amidation and the presence of an *N*-terminal lactamide were postulated. Multiple experiments were conducted to confirm these terminal modifications (**Figure 6.9**). Proteolytic peptide mapping on cleaved Histone H3 Dha9 and protein sequencing via LC-MS/MS confirmed the *N*-terminal lactoyl residue on the *C*-terminal fragment (**Figure 6.9 B**).

The mass spectrometric discrimination between a free *C*-terminal carboxyl group (COOH) and the corresponding primary amide ($CONH_2$) is challenging due to the mass difference of only 1 Da and typically requires derivatization.[62] Hence, high-resolution mass spectrometry and isotope pattern analysis were employed to analyse the fragments obtained after cleavage of preSUMO1 Dha51 and Npβ Dha61. Comparison with the corresponding calculated isotope

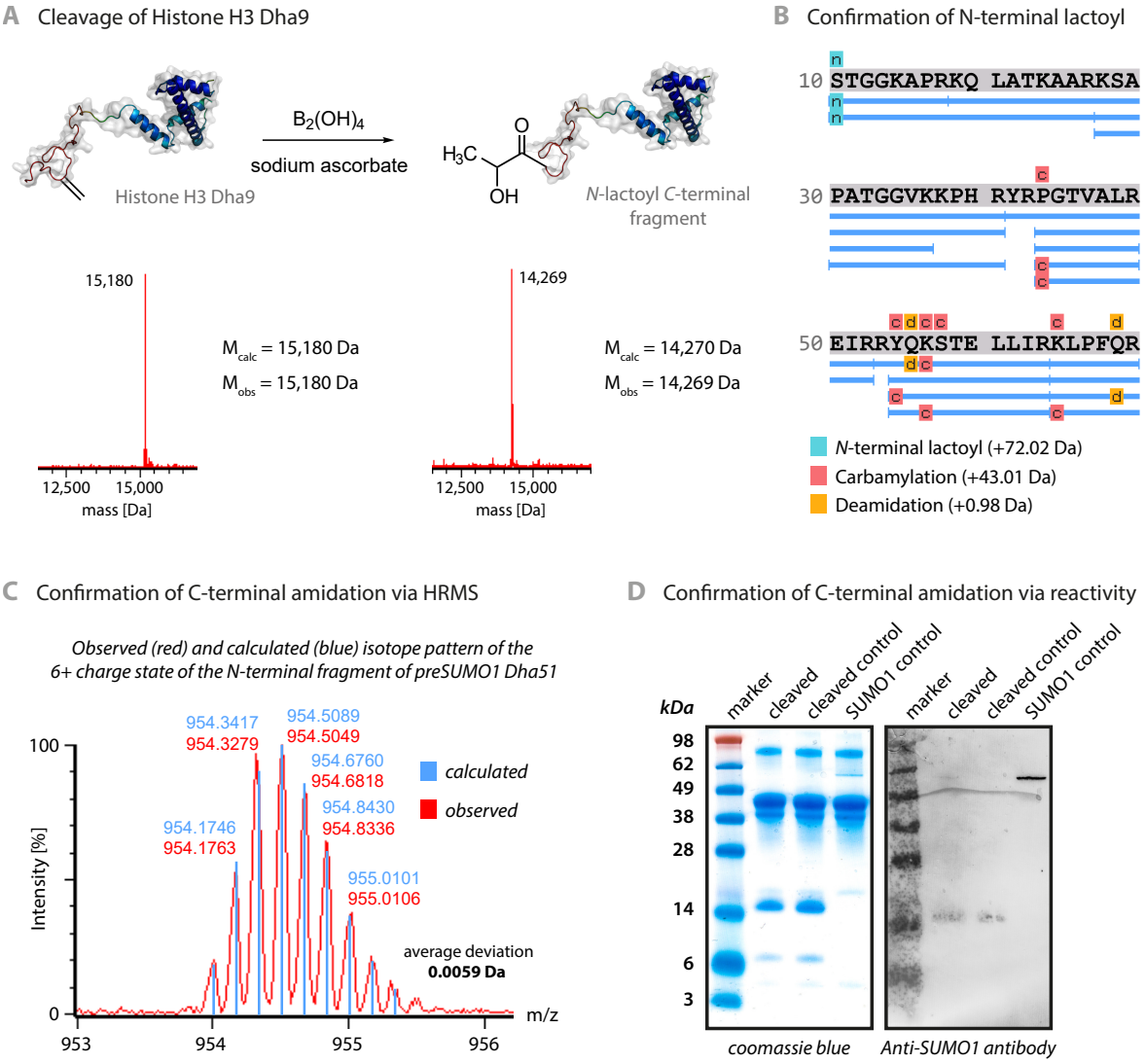


Figure 6.9: Analysis of the termini of the fragments observed after cleavage. (**A**) *N*-terminal cleavage of Histone H3 Dha9 leads to a *C*-terminal fragment carrying a *N*-lactoyl modification. (**B**) The presence of the lactoylated *N*-terminus was confirmed by proteolytic digestion and LC-MS/MS analysis. (**C**) The presence of an amidated *C*-terminus was confirmed by high resolution mass spectrometry and analysis of the isotope patterns of the *N*-terminal fragments obtained after cleavage of preSUMO1 Dha51 (6+ charge state depicted here) and Npβ Dha61. (**D**) preSUMO1 is naturally cleaved by SENP1 *N*-terminally of site 97 during maturation leaving a *C*-terminus (COOH) reactive towards SUMOylation. Introduction of Dha97 allowed for diboron-promoted SUMO maturation giving access to an α-amidated 'protected' ($CONH_2$) SUMO that cannot participate in SUMOylation and can be observed via Western blot (SUMOylation of a RanGAP1 fragment using recombinant SEA1 and Ubc9; cleaved: diboron-promoted SUMO1 maturation, cleaved control: diboron-promoted SUMO1 maturation, no ATP, SUMO1 control: commercial SUMO1 carrying a *C*-terminal carboxylate, positive control)

patterns confirmed *C*-terminal α-amidation (**Figure 6.9 C**).

The possibility to synthesize matured SUMO1-$CONH_2$ (as opposed to 'natural' SUMO1-COOH) by chemical protein cleavage should allow for the creation of a 'protected' SUMO1 mutant that can not participate in SUMOylation reactions. SUMOylation typically proceeds via activation of the *C*-terminal carboxylate through ATP-dependent thioester formation with SUMO-activating enzyme subunit 1 (SAE1) and subsequent transfer to Ubc9 catalysing the formation of an isopeptide bond between the *C*-terminus and lysine residues.[63] The absence of a *C*-terminal carboxyl group in chemically matured SUMO1-$CONH_2$ should therefore prevent its attachment to Lys residues via isopeptide bond formation. Indeed, when SUMO1-$CONH_2$ was subjected to SUMOylation conditions using a RanGAP1 fragment as substrate, no SUMOylated RanGAP1 could be detected by Western blot analysis employing an anti-SUMO1 polyclonal antibody. In contrast, commercial SUMO1-COOH showed clear SUMOylation of RanGAP1 (**Figure 6.9 D**). This confirms *C*-terminal α-amidation rendering SUMO1-$CONH_2$ inactive.

6.3.4 Mechanistic Investigations: Understanding Radical Protein Cleavage

Having shown the general applicability of the diboron-promoted protein cleavage protocol to a range of proteins, mechanistic investigations were conducted. As the cleavage reaction was inhibited by the common radical trap TEMPOL, a radical mechanism was proposed. Additionally, deuteration experiments revealed the incorporation of almost three deuterium atoms into the *C*-terminal cleavage product (**Figure 6.10 A**). When Histone H3 Dha10 was incubated under cleavage conditions under nitrogen atmosphere (nitrogen-filled glove box, 2.1 ppm O_2), no cleavage could be observed. Instead, the incorporation of an average of 1.6 deuterium atoms was observed after 3 hours (**Figure 6.10 B**). This is in agreement with reduction under anaerobic conditions. Treatment of Histone H3 Dha10 with tetrahydroxydiboron (200 equiv) and TEMPOL (50 equiv) for 24 hours under ambient conditions only led to incorporation of an average of 0.9 deuterium atoms with some cleavage being observed (**Figure 6.10 C**). This indicates that not only cleavage, but also reduction are likely radical pathways under these conditions. Depletion of radical scavenger TEMPOL over time thus

enables reduction and cleavage under these conditions.

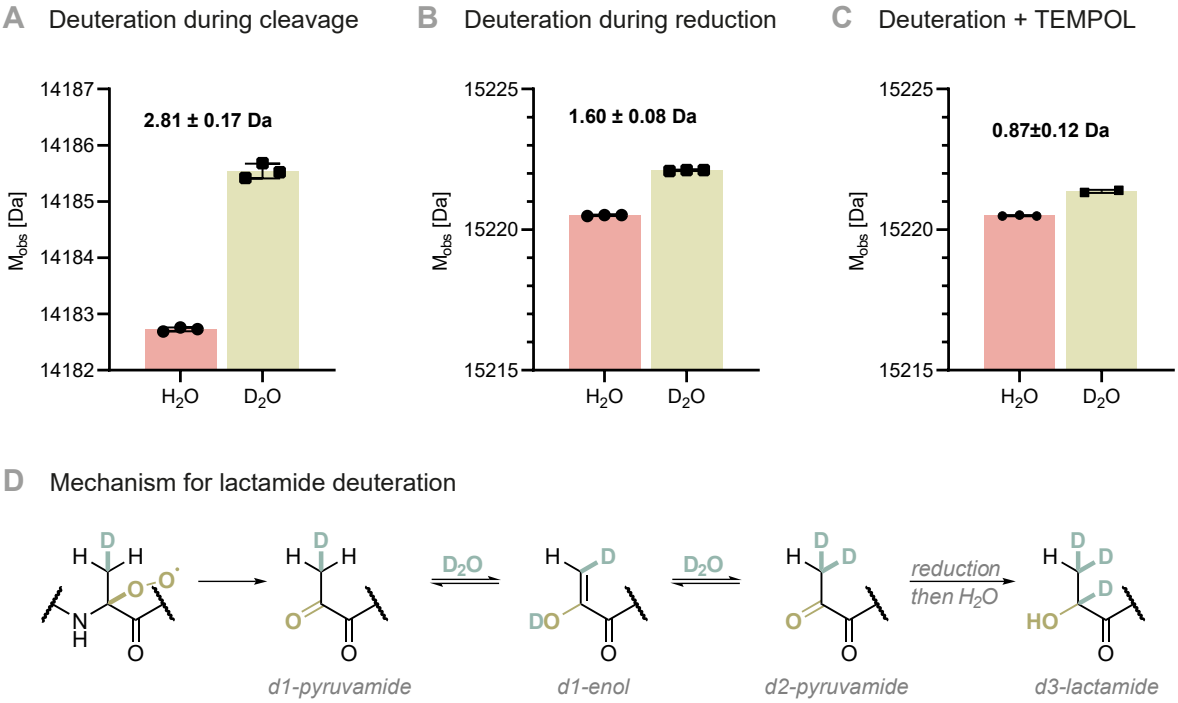
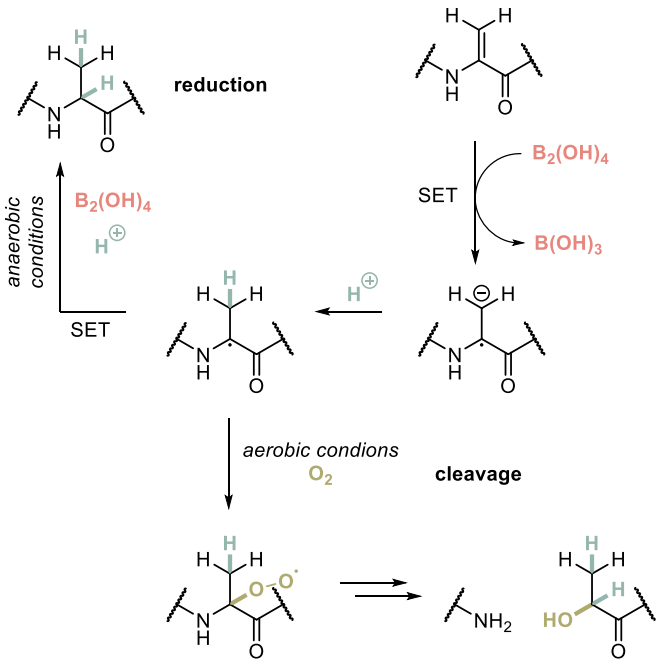


Figure 6.10: Deuteration studies on diboron-mediated protein cleavage: (**A**) Incorporation of an average of 2.81±0.17 deuterium atoms into the *C*-terminal fragment upon cleavage in D_2O. (**B**) Incorporation of an average of 1.60±0.08 deuterium atoms upon reduction under anaerobic conditions. (**C**) Incorporation of an average of 0.87±0.12 deuterium atoms upon reduction under ambient conditions in the presence of radical scavenger TEMPOL. (**D**) Proposed mechanism for the incorporation of multiple deuterium atoms during lactamide formation.

Furthering mechanistic investigations, a small molecule model substrate Ac-Gly-Dha-NHBn (**17**) was synthesized. After treatment of Dha derivative **17** with tetrahydroxydiboron, the formation of the *N*-terminal fragments acetyl glycine and acetyl glycinamide could be observed by ^{1}H-NMR and comparison to the authentic references. The formation of the free carboxylic acid likely proceeds from the primary amide. Possibly, the high concentration of boric acid released under small-molecule conditions is a driving factor of this reaction. Boric acid was reported in the literature as a transamidation catalyst and amide hydrolysis could proceed through a comparable mechanism.[64] While previously reports of oxidative radical protein cleavage reported the formation of a ketoamide derivative,[65] no such motif could be detected in this case. Treatment of the expected ketoamide derivative *N*-benzylpyruvamide (**18**) with $B_2(OH)_4$ under aqueous conditions revealed reduction to the corresponding *N*-

benzyllactamide (**19**). A mechanism was proposed based on these observations (**Scheme 6.4**): the diboron reagent decomposes under the employed conditions by homolytic B—B bond cleavage, yielding an active radical species capable of reaction initiation. Indeed, a procedure utilizing tetrahydroxydiboron as a radical initiator was recently reported.[21]



Scheme 6.4: Proposed mechanism for the site-selective diboron-mediated radical cleavage of the protein backbone under ambient conditions and explanation of observed reduction under anaerobic conditions. Although an intermolecular SET mechanism between the protein and a boron species is depicted here, the reaction could in fact proceed via an intramolecular mechanism from a previously protein-bound boron compound.

Homolytic B—B bond cleavage would also explain the observed reactivity differences between tetrahydroxydiboron, B_2cat_2 and B_2pin_2. B_2pin_2 features a more stable B—B bond than $B_2(OH)_4$ as indicated by bond dissociation energies (BDEs B—B bond: B_2pin_2: 460 kJ/mol, $B_2(OH)_4$: 434 kJ/mol).[66] While B_2cat_2 features an even more stable B—B bond (BDE B—B bond: B_2cat_2: 488 kJ/mol), catechol boronic esters are known for a far lower hydrolytic stability in aqueous media when compared to their pinacol equivalents.[66,67] Consequently, the formation of $B_2(OH)_4$ from B_2cat_2 in aqueous media is likely and would explain the observation of comparable reactivities. Next, the radical anion derived from single electron reduction of dehydroalanine by a boron-centered radical is protonated yielding a C_α radical. In the

absence of molecular oxygen, this radical can undergo a second single electron transfer (SET) followed by protonation giving rise to the alanine derivative. Under aerobic conditions, the C_α radical can react with molecular oxygen. Subsequent cleavage results in formation of a primary amide (*N*-terminal fragment) and a ketoamide (*C*-terminal fragment) as described previously for the radical cleavage of the protein backbone.[68] The proposed mechanism (**Scheme 6.4**) might hereby proceed via an intramolecular rather than an intermolecular SET mechanism. Previous studies (see **Chapter 3** and **Chapter 4**) indicated diverse binding modes for the boron-containing amino acid boronoalanine (Bal). A similarly diverse range of binding interactions could be anticipated for diboron reagents such as tetrahydroxydiboron. If the diboron reagent is recruited to the reaction site by Lewis acid·base pairing prior to formation of the reactive species, an intramolecular mechanism could be proposed. Following backbone cleavage, the ketoamide is reduced to the corresponding alcohol under the reaction conditions yielding a *N*-lactoyl derivatized *N*-terminus on the *C*-terminal fragment. Formation of the lactamide via the keto derivative also explains the observed incorporation of almost three deuterium atoms when the reaction is conducted in D_2O. A keto derivative will undergo keto-enol tautomerism leading to the further incorporation of deuterium (**Figure 6.10 D**).

6.4 Conclusion & Outlook

In summary, mechanistic studies on the borylation of dehydroalanine residues under Copper(II) catalysis revealed the presence of a previously neglected, 'off'-cycle process leading to reduction rather than borylation under limited oxygen supply. Further investigations allowed for the development of a site-specific, diboron-mediated radical protein cleavage protocol. This protocol follows a general 'tag-and-modify' approach[69] enabling selective cleavage at readily introduced dehydroalanine residues. The method was shown to be efficient and work on a range of proteins under benign and aqueous conditions. The method is metal-free, operationally simple and only requires inexpensive and commercially available reagents producing only non-toxic side products. A range of therapeutically relevant proteins are either expressed as their zymogen precursor or carrying affinity tags to aid purification – and therefore require proteolytic processing.[28,70] The cleavage procedure presented herein could provide a cost-efficient and protease-free alternative for the processing of commercially and pharmaceutically

relevant proteins. Furthermore, this diboron-promoted protocol provides several advantages over typical protocols employing highly specific proteases: the introduction of multi-residue recognition sites can be omitted and post-cleavage purification is simplified as no protease has to be separated from the desired protein. Additionally, the *C*-terminal fragment obtained by this method only contains a small *N*-lactoyl residue, while the *N*-terminal fragment contains an amidated *C*-terminus. In fact, *C*-terminal amidation is found naturally in peptides and proteins and was shown to be crucial for the biological activity of numerous therapeutic peptides.[71] Additionally, *C*-terminal amidation was predicted to be a common post-translational modification in the human proteome, possibly involved in mediating protein stability and function.[72] Hence, attempted *in vitro* SUMOylation using SUMO1-$CONH_2$ produced with this methodology confirmed protection towards *C*-terminal activation. The introduction of a *C*-terminal primary amide into recombinant peptides and proteins can be challenging using traditional methodologies and usually requires *in vitro* treatment with or co-expression of a suitable enzyme.[71,73,74] In contrast, the method presented herein allows for the simultaneous cleavage of *C*-terminal tags under introduction of a primary amide.

References

[1] E. C. Neeve, S. J. Geier, I. A. Mkhalid, S. A. Westcott, T. B. Marder, *Chem. Rev.* **2016**, *116*, 9091–9161.

[2] R. J. Brotherton, A. L. McCloskey, J. L. Boone, H. M. Manasevit, *J. Am. Chem. Soc.* **1960**, *82*, 6245–6248.

[3] X. Liu, *Synlett* **2003**, 2442–2443.

[4] J. W. Fyfe, A. J. Watson, *Chem* **2017**, *3*, 31–55.

[5] W. Ding, Q. Song, *Org. Chem. Front.* **2016**, *3*, 14–18.

[6] D. Pi, H. Zhou, Y. Zhou, Q. Liu, R. He, G. Shen, Y. Uozumi, *Tetrahedron* **2018**, *74*, 2121–2129.

[7] X. Han, J. Hu, C. Chen, Y. Yuan, Z. Shi, *Chem. Commun.* **2019**, *55*, 6922–6925.

[8] H. Bao, B. Zhou, H. Jin, Y. Liu, *J. Org. Chem.* **2019**, *84*, 3579–3589.

[9] X. Lin, Y. Wang, Y. Hu, W. Zhu, X. Dou, *Eur. J. Org. Chem.* **2020**, *2020*, 1046–1049.

[10] M. Zhou, K. Li, D. Chen, R. Xu, G. Xu, W. Tang, *J. Am. Chem. Soc.* **2020**, *142*, 10337–10342.

[11] S. P. Cummings, T. Le, G. E. Fernandez, L. G. Quiambao, B. J. Stokes, *J. Am. Chem. Soc.* **2016**, *138*, 6107–6110.

[12] A. Londregan, D. Piotrowski, J. Xiao, *Synlett* **2013**, *24*, 2695–2700.

[13] V. Gurram, H. K. Akula, R. Garlapati, N. Pottabathini, M. K. Lakshman, *Adv. Synth. Catal.* **2015**, *357*, 451–462.

[14] Y.-T. Xia, X.-T. Sun, L. Zhang, K. Luo, L. Wu, *Chem. Eur. J.* **2016**, *22*, 17151–17155.

[15] F. Takahashi, K. Nogi, H. Yorimitsu, *Eur. J. Org. Chem.* **2020**, *2020*, 3009–3012.

[16] D. P. Ojha, K. Gadde, K. R. Prabhu, *Org. Lett.* **2016**, *18*, 5062–5065.

[17] M. Flinker, H. Yin, R. W. Juhl, E. Z. Eikeland, J. Overgaard, D. U. Nielsen, T. Skrydstrup, *Angew. Chem. Int. Ed.* **2017**, *56*, 15910–15915.

[18] K. A. Korvinson, H. K. Akula, C. T. Malinchak, D. Sebastian, W. Wei, T. A. Khandaker, M. R. Andrzejewska, B. Zajc, M. K. Lakshman, *Adv. Synth. Catal.* **2020**, *362*, 166–176.

[19] H.-C. Du, N. Simmons, J. C. Faver, Z. Yu, M. Palaniappan, K. Riehle, M. M. Matzuk, *Org. Lett.* **2019**, *21*, 2194–2199.

[20] J. Kim, C. R. Bertozzi, *Angew. Chem.* **2015**, *127*, 16003–16007.

[21] Z.-Y. Sun, S. Zhou, K. Yang, M. Guo, W. Zhao, X. Tang, G. Wang, *Org. Lett.* **2020**, *22*, 6214–6219.

[22] H. Neurath, K. A. Walsh, *Proc. Natl. Acad. Sci.* **1976**, *73*, 3825–3832.

[23] T. Klein, U. Eckhard, A. Dufour, N. Solis, C. M. Overall, *Chem. Rev.* **2018**, *118*, 1137–1168.

[24] M. Vizovišek, R. Vidmar, M. Drag, M. Fonović, G. S. Salvesen, B. Turk, *Trends Biochem. Sci.* **2018**, *43*, 829–844.

[25] R. W. King, R. J. Deshaies, J.-M. Peters, M. W. Kirschner, *Science* **1996**, *274*, 1652–1659.

[26] M. D. Kazanov, Y. Igarashi, A. M. Eroshkin, P. Cieplak, B. Ratnikov, Y. Zhang, Z. Li, A. Godzik, A. L. Osterman, J. W. Smith, *J. Proteome Res.* **2011**, *10*, 3642–3651.

[27] M. Schauperl, J. E. Fuchs, B. J. Waldner, R. G. Huber, C. Kramer, K. R. Liedl, *PLoS One* **2015**, *10*, e0142658.

[28] D. S. Waugh, *Protein Expr. Purif.* **2011**, *80*, 283–293.

[29] P. Giansanti, L. Tsiatsiani, T. Y. Low, A. J. R. Heck, *Nat. Protoc.* **2016**, *11*, 993–1006.

[30] L. Tsiatsiani, A. J. R. Heck, *FEBS J.* **2015**, *282*, 2612–2626.

[31] D. L. Swaney, C. D. Wenger, J. J. Coon, *J. Proteome Res.* **2010**, *9*, 1323–1329.

[32] M. E. Kimple, A. L. Brill, R. L. Pasker, *Curr. Protoc. Protein Sci.* **2013**, *73*, 1–23.

[33] J. Arnau, C. Lauritzen, G. E. Petersen, J. Pedersen, *Protein Expr. Purif.* **2006**, *48*, 1–13.

[34] T. Parks, E. D. Howard, T. J. Wolpert, D. J. Arp, W. G. Dougherty, *Virology* **1995**, *210*, 194–201.

[35] R. B. Kapust, J. Tözsér, T. D. Copeland, D. S. Waugh, *Biochem. Biophys. Res. Commun.* **2002**, *294*, 949–955.

[36] L. D. Cabrita, D. Gilis, A. L. Robertson, Y. Dehouck, M. Rooman, S. P. Bottomley, *Protein Sci.* **2007**, *16*, 2360–2367.

[37] A. Li, R. C. Sowder, L. E. Henderson, S. P. Moore, D. J. Garfinkel, R. J. Fisher, *Anal. Chem.* **2001**, *73*, 5395–5402.

[38] D. L. Crimmins, S. M. Mische, N. D. Denslow, *Curr. Protoc. Protein Sci.* **2005**, *41*, 1–11.

[39] B. Dang, M. Mravic, H. Hu, N. Schmidt, B. Mensa, W. F. DeGrado, *Nat. Methods* **2019**, *16*, 319–322.

[40] F. B. Peters, A. Brock, J. Wang, P. G. Schultz, *Chem. Biol.* **2009**, *16*, 148–152.

[41] Y. Seki, K. Tanabe, D. Sasaki, Y. Sohma, K. Oisaki, M. Kanai, *Angew. Chem. Int. Ed.* **2014**, *53*, 6501–6505.

[42] H. E. Elashal, M. Raj, *Chem. Commun.* **2016**, *52*, 6304–6307.

[43] J. Ni, Y. Sohma, M. Kanai, *Chem. Commun.* **2017**, *53*, 3311–3314.

[44] W. Li, X. Cui, K. Junge, A.-E. Surkus, C. Kreyenschulte, S. Bartling, M. Beller, *ACS Catal.* **2019**, *9*, 4302–4307.

[45] J. E. De Vrieze, C. A. Urbina Blanco, J. W. Thybaut, M. Saeys, *ACS Catal.* **2019**, *9*, 8073–8082.

[46] M. J. Moran, K. Martina, F. Baricco, S. Tagliapietra, M. Manzoli, G. Cravotto, *Adv. Synth. Catal.* **2020**, *362*, 2689–2700.

[47] S. E. Sloane, A. Reyes, Z. P. Vang, L. Li, K. T. Behlow, J. R. Clark, *Org. Lett.* **2020**, *22*, 9139–9144.

[48] B. M. Zimmermann, S. C. K. Kobosil, J. F. Teichert, *Chem. Commun.* **2019**, *55*, 2293–2296.

[49] I. Thomé, A. Nijs, C. Bolm, *Chem. Soc. Rev.* **2012**, *41*, 979.

[50] S. B. Thorpe, J. A. Calderone, W. L. Santos, *Org. Lett.* **2012**, *14*, 1918–1921.

[51] M. Isegawa, W. M. Sameera, A. K. Sharma, T. Kitanosono, M. Kato, S. Kobayashi, K. Morokuma, *ACS Catal.* **2017**, *7*, 5370–5380.

[52] C. Zheng, J. J. Hayes, *Biopolymers* **2003**, *68*, 539–546.

[53] M. W. Vetting, S. S. Hegde, K. Z. Hazleton, J. S. Blanchard, *Protein Sci.* **2007**, *16*, 755–760.

[54] R. Huber, J. Römisch, E. P. Paques, *Eur. Mol. Biol. Organ. J.* **1990**, *9*, 3867–3874.

[55] S. Müller, C. Hoege, G. Pyrowolakis, S. Jentsch, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 202–210.

[56] A. A. Belushkin, D. V. Vinogradov, M. S. Gelfand, A. L. Osterman, P. Cieplak, M. D. Kazanov, *Proteomics* **2014**, *14*, 42–50.

[57] B. Lee, F. Richards, *J. Mol. Biol.* **1971**, *55*, 379.

[58] S. Mitternacht, *F1000Research* **2016**, *5*, 1–11.

[59] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nat. Methods* **2012**, *9*, 671–675.

[60] L. Shen, M. H. Tatham, C. Dong, A. Zagórska, J. H. Naismith, R. T. Hay, *Nat. Struct. Mol. Biol.* **2006**, *13*, 1069–1077.

[61] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem. Sci.* **2011**, *2*, 1666.

[62] A. Kaya, H. C. Karakaya, D. E. Fomenko, V. N. Gladyshev, A. Koc, *Mol. Cell. Biol.* **2009**, *29*, 3665–3674.

[63] A. Flotho, F. Melchior, *Annu. Rev. Biochem.* **2013**, *82*, 357–385.

[64] T. B. Nguyen, J. Sorres, M. Q. Tran, L. Ermolenko, A. Al-Mourabit, *Org. Lett.* **2012**, *14*, 3202–3205.

[65] T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallee, A. Phanumartwiwath, O. D. Coleman, M.-L. Thezenas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed, B. G. Davis, *Science* **2016**, *354*, 597.

[66] J. Wang, W. Zheng, Y. Zheng, *RSC Adv.* **2017**, *7*, 49251–49272.

[67] A. J. J. Lennox, G. C. Lloyd-Jones, *Chem. Soc. Rev.* **2014**, *43*, 412–443.

[68] C. L. Hawkins, M. J. Davies, *Biochim. Biophys. Acta Bioenerg.* **2001**, *1504*, 196–219.

[69] J. M. Chalker, G. J. L. Bernardes, B. G. Davis, *Acc. Chem. Res.* **2011**, *44*, 730–741.

[70] G. Walsh, *Drug Discov. Today* **2010**, *15*, 773–780.

[71] D. J. Merkler, *Enzyme Microb. Technol.* **1994**, *16*, 450–456.

[72] A. A. Tokmakov, A. Kurotani, T. Takagi, M. Toyama, M. Shirouzu, Y. Fukami, S. Yokoyama, *J. Biol. Chem.* **2012**, *287*, 27106–27116.

[73] G. Walsh, R. Jefferis, *Nat. Biotechnol.* **2006**, *24*, 1241–1252.

[74] T. Nuijens, E. Piva, J. A. Kruijtzer, D. T. Rijkers, R. M. Liskamp, P. J. Quaedflieg, *Tetrahedron Lett.* **2012**, *53*, 3777–3779.

Chapter 7

Experimental Methods & Materials

Chapter 7

Experimental Methods & Materials

7.1 General Methodology

7.1.1 Synthetic Chemistry

For air and moisture sensitive reactions, all used glassware was dried by heating in high vacuum and purging with nitrogen. This procedure was repeated three times prior to use. Syringes were purged with nitrogen and airtight septa were used for the addition of liquids to reactions performed under nitrogen atmosphere. Solid substances were added using nitrogen countercurrent where needed. Solvents were removed on a rotary evaporator where applicable.

7.1.2 Chromatography

Chromatographic separations were performed by flash chromatography on CombiFlash instruments (Teledyne ISCO, USA). Analytical thin-layer chromatography (TLC) was performed on Merck aluminium backed sheets coated with silica gel 60 F_{254}. UV-active substances were visualized by short wavelength (254 nm) ultraviolet light. Other substances were visualized using a $KMnO_4$ staining solution and heating the TLC plate after application.

7.1.3 Analytical Techniques

Nuclear Magnetic Resonance Spectroscopy (NMR Spectroscopy)

Synthesized compounds were characterized by NMR spectroscopy. ^{1}H- and ^{13}C-NMR spectra were recorded using Bruker Advance-III HD 400 MHz, Bruker Advance-III HD 500 MHz and Bruker Advance-III HD 600 MHz spectrometers. ^{1}H- and ^{13}C-NMR spectra were referenced to the NMR solvent shift or to tetramethylsilane (TMS). Chemical shifts are given in parts per million (ppm) and coupling constants are declared in Hertz (Hz). The multiplicities of the signals are reported as singlet (s), doublet (d), triplet (t), quartet (q), pentet (pent), multiplet (m), broad (br) or any combination of these. Signals showing unexpected multiplicities are indicated by the prefix "app".

Protein Mass Spectrometry

Protein mass spectra were recorded on a Waters Xevo G2-S QTof spectrometer, on a Waters Xevo G2-XS QTof spectrometer or an AB Sciex TripleTOF 6600 spectrometer. Spectra were deconvoluted either using MassLynx (Waters) and the "MaxEnt1" deconvolution algorithm or Analyst (AB Sciex) using the "Reconstruct Protein" algorithm. Conversions were calculated from peak intensities (Waters) or peak areas (AB Sciex) dividing the value for the product by the sum of the values for products and (residual) starting material. Impurities present prior to the reaction were not considered.

IR Spectrometry

Infrared (IR) spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer equipped with a diamond ATR probe or on a Thermo Nexus 670 equipped with a silicon probe. Samples were loaded directly onto the probe.

Melting Points

Melting points were measured via differential scanning calorimetry (DSC) using a Mettler Toledo DSC 1 STAR^{e} system. Heating was conducted from 30°C to 300°C at a heating rate of 10°C/minute and peak values are given. Uncorrected melting points were determined using

a Leica Galen III microscope equipped with a heating stage and a Testo 720 thermocouple probe. Heating was conducted up to 200°C at a heating rate of 2°C/minute.

Optical Rotations

Optical rotations were recorded on a Schmidt+Haensch Unipol polarimeter using a cell with a length of 10 cm. Background measurements were performed before loading of the sample.

CD Spectroscopy

Circular dichroism (CD) spectra were recorded on a Chirascan device at 25°C. Experiments were performed using 300 µL of protein solution in a 1 mm cuvette in buffer. Spectra were collected from 200 to 260 nm (0.5 nm step, 1 s per point) as a smoothened average of three scans. Buffer solution was used as a blank sample and subtracted from the measured data.

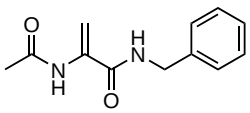
pK_a Determination

pK_a values were determined using a Sirius T3 automated titrator (Pion Inc., United Kingdom). A stock solution of compound in DMSO was prepared and titrated from pH 2.0 to pH 12.0 in the presence of three varying concentrations of MeOH. HCl was used for acidification and KOH was used for titration. The pK_a value in a solely aqueous system was extrapolated from the measured pK_a values for the three different concentrations of organic co-solvent using Yasuda-Shedlovsky extrapolation.[1]

7.2 Synthetic Chemistry

7.2.1 Small Molecule Model

2-Acetylamino-*N*-benzyl-acrylamide (1)



The title compound was synthesized following a literature procedure.[2] To a stirred solution of 2-acetamidoacrylic acid (1.29 g, 10.0 mmol, 1.00 equiv) and 4-methylmorpholine (1.21 mL, 11.0 mmol, 1.10 equiv) in THF (100 mL) were added subsequently isobutyl chloroformate (1.43 mL, 11 mmol, 1.10 equiv) and benzylamine (1.20 mL, 11.0 mmol, 1.10 equiv). The mixture was stirred

at room temperature for 2 h, before it was filtered and the solvent was evaporated. The residue was purified by flash chromatography (*n*-heptane / EtOAc; 10-100% EtOAc) yielding compound **1** as a white solid (1.62 g, 7.43 mmol, 74%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 9.08 (s, 1H, CON*H*), 8.86 (t, J = 6.1 Hz, 1H, CON*H*), 7.35 – 7.20 (m, 5H, 5 × Ar-*H*), 6.03 (s, 1H, C=CH_a), 5.44 (s, 1H, C=CH_b), 4.36 (d, J = 6.1 Hz, 2H, CH_2), 2.00 (s, 3H, CH_3).

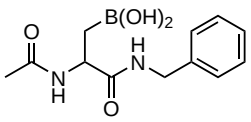
^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 169.2 (*C*=O), 164.1 (*C*=O), 139.3 (C_{Ar}), 136.2 (*C*=CH_2), 128.2 (C_{Ar}), 127.2 (C_{Ar}), 126.8 (C_{Ar}), 103.0 (C=*C*H_2), 42.5 (*C*H_2), 23.8 (*C*H_3).

IR (silicone, neat): $\tilde{\nu}$ (cm^{-1}) = 3404, 3295, 3033, 2941, 1723, 1666, 1648, 1622, 1533, 1497, 1423, 1401, 1371, 1346, 1322, 1283, 1267, 1232, 1215, 1190, 1125, 1100, 1070, 1025, 1017, 1001, 990, 968, 954, 939, 923, 907, 867, 850, 839, 827, 806, 752, 726, 707, 694, 660.

m.p.: 126.8°C

HRMS (ESI+): m/z for $C_{12}H_{15}O_2N_2$ $[M+H]^+$: calculated: 219.1128; found: 219.1129.

(2-acetamido-3-(benzylamino)-3-oxopropyl)boronic acid (3)



To a stirred suspension of 2-Acetylamino-*N*-benzyl-acrylamide (**1**) (1.00 g, 4.58 mmol, 1.00 equiv) and tetrahydroxydiboron (822 mg, 9.16 mmol, 2.00 equiv) in 200 mL of water were added $CuSO_4 \cdot 5H_2O$ (114 mg, 0.46 mmol, 0.10 equiv) and 4-methylpyridine (111 µL, 1.15 mmol, 0.25 equiv). The reaction mixture was stirred vigorously (1700 rpm) in a flask open to air for 90 min. The resulting solution was lyophilized and dissolved in 18 mL of H_2O for purification. Purification was performed via preparative HPLC (stationary phase: RP XBridge Prep C18 OBD-10 µm, 50 × 250 mm, mobile phase: 0.25% NH_4HCO_3 solution in water, MeCN). The title compound **3** was afforded after lyophilisation as a white solid (730 mg, 2.76 mmol, 60%).

^{1}H-NMR (400 MHz, NaP_i (20 mM, pH 1.5, D_2O), 300 K): δ (ppm) = 7.47 – 7.24 (m, 5H, Ar-*H*), 4.45 – 4.33 (m, 3H, C*H* + CH_2NHBn), 2.02 (s, 3H, CH_3), 1.23 (app qd, J = 15.4, 8.0 Hz, 2H, CH_2).

^{13}C-NMR (101 MHz, NaP_i (20 mM, pH 1.5, D_2O), 300 K): δ (ppm) = 175.3 (*C*=O), 174.1 (*C*=O), 137.8 (C_{Ar}), 128.7 (C_{Ar}), 127.4 (C_{Ar}), 127.0 (C_{Ar}), 51.6 (C*H*), 42.8 (*C*H_2), 21.7 (*C*H_3), 18.2 (br s, *C*H_2).

^{11}B-NMR (128 MHz, NaP_i (20 mM, pH 1.5, D_2O), 300 K): δ (ppm) = 29.7 (br s, $B(OH)_2$).

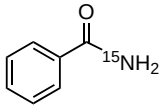
IR (silicone, neat): $\tilde{\nu}$ (cm^{-1}) = 3282, 3078, 2933, 1631, 1531, 1498, 1454, 1420, 1374, 1259, 1080, 1029, 897, 734, 699.

m.p.: not determined due to decomposition

HRMS (ESI+): *m/z* for $C_{12}H_{17}O_4N_2{}^{11}BNa$ $[M+Na]^+$: calculated: 287.1174; found: 287.1175.

pK_a (aqueous solution, 25°C): 8.31 ± 0.02

[^{15}N]Benzamide (20)



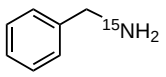
[^{15}N]Benzamide was synthesized following a reported procedure.[3] [^{15}N]NH_4Cl (98% ^{15}N, 1.00 g, 18.4 mmol, 1.00 equiv) was dissolved in water (15 mL) and toluene (2 mL) was added. The mixture was cooled to 0°C. NaOH (4 M, 10 mL, 40.4 mmol, 2.20 equiv) was added followed by a solution of benzoyl chloride (2.2 mL, 18.7 mmol, 1.02 equiv) in toluene (80 mL) was added slowly. After addition, the mixture was allowed to warm to room temperature and stirred overnight. The white precipitate was collected by filtration, washed with ice cold water (3 × 10 mL) and dried *in vacuo* to yield the title compound (**20**) as a white solid (1.70 g, 13.9 mmol, 76%).

^{1}H-NMR (400 MHz, $CDCl_3$, 295 K): δ (ppm) = 7.84 – 7.79 (m, 2H, Ar-*H*), 7.55 – 7.49 (m, 1H, Ar-*H*), 7.47 – 7.41 (m, 2H, Ar-*H*), 6.24 (d, *J* = 88.7 Hz, 2H, CONH_2).

^{13}C-NMR (101 MHz, $CDCl_3$, 295 K): δ (ppm) = 169.8 (d, *J* = 15.4 Hz, *C*ONH_2), 133.5 (d, *J* = 8.1 Hz, C_{Ar}), 132.1 (C_{Ar}), 128.7 (C_{Ar}), 127.5 (C_{Ar}).

HRMS (ESI+): *m/z* for $C_7H_8{}^{15}NO$ $[M+H]^+$: calculated: 123.0571; found: 123.0571.
The analytical data are in agreement with the literature.[3]

[^{15}N]Benzylamine (21)



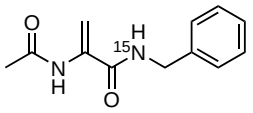
[^{15}N]Benzylamine was synthesized following a reported procedure.[3] [^{15}N]Benzylamide (**20**) (1.00 g, 8.25 mmol, 1.00 equiv) was dissolved in dry THF and added to a suspension of $LiAlH_4$ (934 mg, 24.8 mmol, 3.00 equiv) in THF (40 mL) at 0°C. The solution was warmed to room temperature and refluxed for 2.5 h. After cooling back to room temperature, the resulting mixture was slowly added to ice water (200 mL) and filtered through a plug of celite. NaOH (4 M, 10 mL) was added and the aqueous solution was extracted with EtOAc (3 × 100 mL). The combined organic extracts were dried over Na_2SO_4 and the solvent was evaporated to afford [^{15}N]benzylamine (**21**) as a yellow oil (784 mg, 8.19 mmol, 89%).

^{1}H-NMR (400 MHz, $CDCl_3$, 295 K): δ (ppm) = 7.39 – 7.22 (m, 5H, Ar-*H*), 3.89 (s, 2H, C*H*$_2$), 1.55 (br s, 2H, N*H*$_2$).

^{13}C-NMR (101 MHz, $CDCl_3$, 295 K): δ (ppm) = 143.5 (C_{Ar}), 128.7 (C_{Ar}), 127.2 (C_{Ar}), 126.9 (C_{Ar}), 46.7 (d, J = 3.9 Hz, CH$_2$).

HRMS (ESI+): *m/z* for $C_7H_9{}^{15}N$ $[M+H]^+$: calculated: 109.0778; found: 109.0780.
The analytical data are in agreement with the literature.[3]

2-Acetylamino-*N*-benzyl-acryl-[^{15}N]-amide (1b)



2-Acetamidoacrylic acid (837 mg, 6.49 mmol, 1.10 equiv) was suspended in THF (40 mL). *N*-methylmorpholine (713 µL, 6.49 mmol, 1.10 equiv), isobutyl chloroformate (841 µL, 6.49 mmol, 1.10 equiv) and [^{15}N]benzylamine (**21**) (650 µL, 5.90 mmol, 1.00 equiv) were added subsequently. The mixture was stirred overnight at room temperature, filtered through a plug of celite and purified by flash chromatography (0-100% EtOAc in petroleum ether) and recrystallized from EtOH/Et_2O yielding the title compound (**1b**) as a white solid (715 mg, 3.26 mmol, 55%).

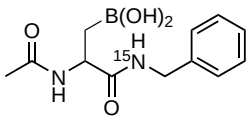
^{1}H-NMR (400 MHz, DMSO-d_6, 295 K): δ (ppm) = 9.08 (br s, 1H, CON*H*), 8.86 (dt, J = 92.8, 6.1 Hz, 1H, CO^{15}N*H*), 7.36 – 7.19 (m, 5H, Ar-*H*), 6.03 (s, 1H, C=C*H*), 5.43 (s, 1H, C=C*H*), 4.36 (d, J = 5.6 Hz, 2H, C*H*$_2$), 2.00 (s, 3H, C*H*$_3$).

^{13}C-NMR (101 MHz, DMSO-d_6, 295 K): δ (ppm) = 169.2 (*C*ONH), 164.1 (d, J = 17.1 Hz,

$CO^{15}NH$), 139.3 (C_{Ar}), 136.1 (d, J = 9.2 Hz, C=CH_2), 128.2 (CAr), 127.2 (C_{Ar}), 126.8 (C_{Ar}), 102.9 (C=CH), 42.5 (d, J = 10.7 Hz, CH_2), 23.8 (CH_3).

HRMS (ESI+): m/z for $C_{12}H_{14}N^{15}NO_2Na$ $[M+Na]^+$: calculated: 242.0918; found: 242.0918.

(2-Acetamido-3-(benzyl-[^{15}N]-amino)-3-oxopropyl)boronic acid (3b)



2-Acetylamino-N-benzyl-acryl-[^{15}N]-amide (300 mg, 1.37 mmol, 1.00 equiv), $B_2(OH)_4$ (245 mg, 2.74 mmol, 2.00 equiv) and $CuSO_4 \cdot 5H_2O$ (34 mg, 0.14 mmol, 10 mol-%) were suspended in distilled water (30 mL). 4-picoline was added (33 µL, 0.34 mmol, 25 mol-%) and the mixture was stirred vigorously open to air at room temperature for 3 h. Another portion of $B_2(OH)_4$ (245 mg, 2.74 mmol, 2.00 equiv) was added and stirring was continued overnight. The solution was lyophilized and dissolved in 18 mL H_2O for purification. Purification was performed via preparative HPLC (stationary phase: RP XBridge Prep C18 OBD-10 µm, 50 × 250 mm, mobile phase: 0.25% NH_4HCO_3 solution in water, MeCN). The title compound (**22**) was afforded after lyophilisation as an off-white solid (155 mg, 0.58 mmol, 43%).

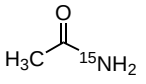
^{1}H-NMR (400 MHz, D_2O (5% conc. DCl), 295 K): δ (ppm) = 7.21 – 7.06 (m, 5H, Ar-H), 4.25 – 4.10 (m, 3H, CH + CH_2NHBn), 1.89 (s, 1H, CH_3), 1.05 – 0.81 (m, 2H, CH_2).

^{13}C-NMR (101 MHz, D_2O (5% conc. DCl), 295 K): δ (ppm) = 174.5 (d, J = 16.7 Hz, C=O), 174.4 (C=O), 137.5 (C_{Ar}), 128.6 (C_{Ar}), 127.3 (C_{Ar}), 127.0 (C_{Ar}), 52.9 – 51.7 (m, C_α), 42.9 (dd, J = 9.9, 4.9 Hz), 21.3 (CH_3), 17.2 ($C_\beta H_2$).

^{11}B-NMR (128 MHz, D_2O (5% conc. DCl), 295 K): δ (ppm) = 21.66 (br s).

HRMS (ESI+): m/z for $C_{12}H_{17}O_4N^{15}N^{11}BNa$ $[M+Na]^+$: calculated: 288.1144; found: 288.1147.

[^{15}N]-Acetamide (23)



To a solution of [^{15}N]NH_4Cl (98% ^{15}N, 1.00 g, 18.4 mmol, 1.00 equiv) in NaOH (10 M, 4.1 mL, 41.1 mmol, 2.20 equiv) and toluene (40 mL) was added acetyl chloride (1.5 mL, 20.6 mmol, 1.10 equiv) at 0°C. The solution was allowed to warm to room temperature and vigorously stirred overnight. The solvent was evaporated and the solid residue was extracted with boiling acetone (3 × 50 mL). The combined extracts were evaporated to

dryness to yield [^{15}N]-acetamide (**23**) as white needles (724 mg, 18.7 mmol, 66%).

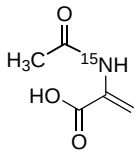
^{1}H-NMR (400 MHz, $CDCl_3$, 295 K): δ (ppm) = 5.75 (dd, J = 88.3, 3.1 Hz, 1H, CON*H*), 5.59 (dd, J = 88.9, 3.2 Hz, 1H, CON*H*), 2.00 (d, J = 1.2 Hz, 3H, CH_3).

^{13}C-NMR (101 MHz, $CDCl_3$, 295 K): δ (ppm) = 173.0 (d, J = 14.1 Hz, *C*=O), 22.8 (d, J = 9.0 Hz, *C*H_3).

HRMS (ESI+): m/z for $C_2H_6{}^{15}NO$ $[M+H]^+$: calculated: 61.0414; found: 61.0420.

The spectroscopic data is in agreement with the literature.[4]

[^{15}N]-Acetamidoacrylic acid (24)



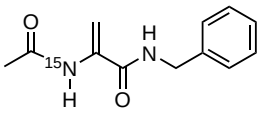
Following a literature procedure,[5] a solution of [^{15}N]-acetamide (**23**) (600 mg, 10.0 mmol, 1.00 equiv) and pyruvic acid (1.5 mL, 22.0 mmol, 2.20 equiv) in toluene (80 mL) were refluxed under Dean-Stark conditions overnight. The solvent was evaporated and the solid residue was titurated with ether to afford the title compound as light brown solid containing impurities (667 mg). The compound was used without further purification.

^{1}H-NMR (400 MHz, DMSO-d_6, 295 K): δ (ppm) = 13.27 (br s, 1H, COO*H*), 9.09 (dd, J = 91.9, 1.1 Hz, 1H, CO^{15}N*H*), 6.23 (t, J = 1.2 Hz, 1H, C=C*H*), 5.65 (dt, J = 4.3, 1.1 Hz, 1H, C=C*H*), 2.02 (d, J = 1.4 Hz, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 295 K): δ (ppm) = 169.7 (d, J = 12.8 Hz, *C*O^{15}NH), 165.5 (d, J = 2.7 Hz, COO*H*), 133.8 (d, J = 15.7 Hz, *C*=CH_2), 109.0 (C=*C*H_2), 24.3 (d, J = 9.4 Hz, *C*H_3).

HRMS (ESI+): m/z for $C_5H_8{}^{15}NO_3$ $[M+H]^+$: calculated: 131.0469; found: 131.0469.

2-Acetyl-[^{15}N]-amino-*N*-benzyl-acrylamide (1a)



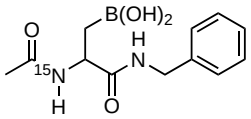
[^{15}N]-2-Acetamidoacrylic acid (**24**) (400 mg, 3.07 mmol, 1.00 equiv) was suspended in THF (30 mL). *N*-methylmorpholine (372 µL, 3.38 mmol, 1.10 equiv), isobutyl chloroformate (440 µL, 3.38 mmol, 1.10 equiv) and benzylamine (369 µL, 3.38 mmol, 1.10 equiv) were added subsequently. The mixture was stirred at room temperature for 1 h, filtered through a plug of celite and purified by flash chromatography (0-100% EtOAc in petroleum ether) affording the title compound (**1a**) as a pale yellow solid (252 mg, 1.15 mmol, 37%).

^{1}H-NMR (400 MHz, DMSO-d_6, 295 K): δ (ppm) = 9.08 (d, J = 91.5 Hz, 1H, CO^{15}NH), 8.86 (t, J = 6.1 Hz, 1H, CONH), 7.36 – 7.19 (m, 5H, Ar-H), 6.03 (d, J = 1.4 Hz, 1H, C=CH), 5.43 (d, J = 4.4 Hz, 1H, C=CH), 4.36 (d, J = 6.1 Hz, 2H, CH_2), 2.00 (d, J = 1.3 Hz, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 295 K): δ (ppm) = 169.2 (d, J = 12.9 Hz, CO^{15}NH), 164.1 (d, J = 2.5 Hz, CONH), 139.3 (C_{Ar}), 136.2 (d, J = 15.6 Hz, C=CH2), 102.9 (C=CH$_2$), 42.5 (CH$_2$), 23.8 (d, J = 9.4 Hz, CH$_3$).

HRMS (ESI+): *m/z* for $C_{12}H_{14}N^{15}NO_2Na$ $[M+Na]^+$: calculated: 242.0918; found: 242.0918.

(2-Acet-[^{15}N]-amido-3-(benzylamino)-3-oxopropyl)boronic acid (3a)



2-Acetyl-[^{15}N]-amino-*N*-benzyl-acrylamide (150 mg, 0.68 mmol, 1.00 equiv), $B_2(OH)_4$ (123 mg, 1.37 mmol, 2.00 equiv) and $CuSO_4 \cdot 5H_2O$ (17 mg, 0.07 mmol, 10 mol-%) were suspended in distilled water (15 mL). 4-picoline was added (16.6 µL, 0.17 mmol, 25 mol-%) and the mixture was stirred vigorously open to air at room temperature for 3 h. Another portion of $B_2(OH)_4$ (123 mg, 1.37 mmol, 2.00 equiv) was added and stirring was continued overnight. The solution was lyophilized and dissolved in 18 mL H_2O for purification. Purification was performed via preparative HPLC (stationary phase: RP XBridge Prep C18 OBD-10 µm, 50 × 250 mm, mobile phase: 0.25% NH_4HCO_3 solution in water, MeCN). The title compound was afforded after lyophilisation as an off-white solid (101 mg, 0.38 mmol, 56%).

^{1}H-NMR (400 MHz, D_2O (5% conc. DCl), 295 K): δ (ppm) = 7.27 – 7.02 (m, 5H, Ar-H),

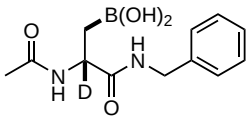
4.24 – 4.09 (m, 3H, C*H* + C*H*$_2$NHBn), 1.89 (d, J = 1.5 Hz, 3H, C*H*$_3$), 1.04 – 0.80 (m, 2H, C*H*$_2$).

^{13}C-NMR (101 MHz, D_2O (5% conc. DCl), 295 K): δ (ppm) = 174.5 (*C*=O), 174.4 (d, J = 15.6 Hz, *C*=O), 137.5 (C_{Ar}), 128.6 (C_{Ar}), 127.3 (C_{Ar}), 127.0 (C_{Ar}), 52.3 (dd, J = 9.9, 5.1 Hz, C_α), 42.9 (app t, J = 4.7 Hz, *C*H$_2$), 21.3 (d, J = 6.2 Hz, *C*H$_3$), 17.2 (*C*H$_2$).

^{11}B-NMR (128 MHz, D_2O (5% conc. DCl), 295 K): δ (ppm) = 20.65 (br s).

HRMS (ESI+): m/z for $C_{12}H_{17}O_4N^{15}N^{11}BNa$ $[M+Na]^+$: calculated: 288.1144; found: 288.1145.

(2-Acetamido-3-(benzylamino)-3-oxopropyl-2-*d*)boronic acid (3-d$_1$)



2-Acetylamino-*N*-benzyl-acrylamide (**1**) (100 mg, 458 µmol, 1.00 equiv), $CuSO_4 \cdot {}_5H_2O$ (11 mg, 45.8 µmol, 10 mol-%), 4-picoline (11 µL, 115 µmol, 25 mol-%) and $B_2(OD)_4$ (123 mg, 1.38 mmol, 3.00 equiv) were dissolved in D_2O (5 mL) and stirred vigorously (1700 rpm) in a vial open to air overnight. The solvent was removed by lyophilisation and the crude product was purified by preparative HPLC (stationary phase: RP XBridge Prep C_{18} OBD-10 µm, 50 × 150 mm, mobile phase: 0.25% NH_4HCO_3 solution in water, MeCN). Ac-Bal-NHBn-d$_1$ (**3-d$_1$**) was obtained as a white solid (69 mg, 260 µmol, 57%). A deuterium incorporation of 92% for D−C$_\alpha$ was observed by ^{1}H-NMR.

^{1}H-NMR (400 MHz, D_2O, 295 K): δ (ppm) = 7.45 – 7.23 (m, 5H, Ar-*H*), 4.40 (dd, J = 22.1, 15.4 Hz, 2H, C*H*$_2$), 2.04 (s, 3H, C*H*$_3$), 1.22 (dd, J = 33.1, 15.1 Hz, 2H, $C_\beta H_2$).

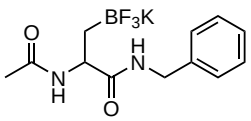
^{13}C-NMR (101 MHz, D_2O, 295 K): δ (ppm) = 175.1 (*C*=O), 174.1 (*C*=O), 137.8 (C_{Ar}), 128.7 (C_{Ar}), 127.3 (C_{Ar}), 127.0 (C_{Ar}), 51.4 (t, J = 21.4 Hz, C_αD), 42.8 (*C*H$_2$), 21.6 (*C*H$_3$), 17.8 (br s, $C_\beta H_2$).

^{11}B-NMR (128 MHz, D_2O, 295 K): δ (ppm) = 28.2 (br s, *B*(OH)$_2$).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3273, 2929, 1631, 1534, 1370, 1297, 1030, 732, 697.

HRMS (ESI+): m/z for $C_{12}H_{16}DO_4N_2{}^{11}BNa$ $[M+Na]^+$: calculated: 288.1236; found: 288.1238.

Potassium (2-acetamido-3-(benzylamino)-3-oxopropyl)trifluoroborate (15)



Ac-Bal-NHBn (450 mg, 1.70 mmol, 1.00 equiv) was dissolved in MeOH (50 mL). A solution of potassium bifluoride in water (4.5 M, 1.14 mL, 5.11 mmol, 3.00 equiv) was added and the mixture was stirred overnight. The solvent was evaporated and the residual solid was transferred to a Soxhlet cellulose thimble. Soxhlet extraction was performed for 20 h using 60 mL of acetone. The solvent was evaporated and the product was dried *in vacuo* to yield the title compound as an off-white solid (357 mg, 1.09 mmol, 64%).

^{1}H-NMR (400 MHz, DMSO-d_6, 295 K): δ (ppm) = 7.77 (t, J = 6.1 Hz, 1H, CONH), 7.38 – 7.13 (m, 5H, Ar-H), 6.97 (d, J = 5.6 Hz, 1H, CONH), 4.22 (d, J = 6.0 Hz, 2H, CH_2Ph), 3.92 (dt, J = 10.5, 5.4 Hz, 1H, C$_\beta H_2$), 1.79 (s, 3H, CH_3), 0.57 – 0.26 (m, 2H, C$_\alpha H_2$).

^{13}C-NMR (101 MHz, DMSO-d_6, 295 K): δ (ppm) = 175.2 (C=O), 168.5 (C=O), 140.1 (C_{Ar}), 128.0 (C_{Ar}), 126.8 (C_{Ar}), 126.3 (C_{Ar}), 52.6 (C_αH$_2$), 41.7 (CH$_2$Ph), 22.9 (CH$_3$). C$_\beta$ was not observed.

^{11}B-NMR (128 MHz, DMSO-d_6, 295 K): δ (ppm) = 4.1 (br s, BF$_3^-$).

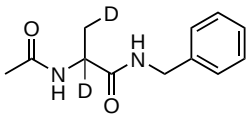
^{19}F-NMR (376 MHz, DMSO-d_6, 295 K): δ (ppm) = −134.90 (s, BF_3^-).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3276, 1655, 1620, 1544, 1524, 1496, 1454, 1421, 1370, 1274, 1236, 1072, 1031, 1012, 968, 928, 751, 729, 696.

m.p.: 215.3°C.

HRMS (ESI+): m/z for $C_{12}H_{15}O_2N_2{}^{11}BF_3$ [M−K]$^-$: calculated: 287.1173; found: 287.1181.

2-Acetamido-*N*-benzylpropanamide-d_2 (16-d$_2$)



D$_2$O (5.0 mL) was thoroughly degassed using five consecutive freeze-pump-thaw cycles. 2-Acetylamino-*N*-benzyl-acrylamide (50.0 mg, 229 µmol, 1.00 equiv), $CuSO_4{\cdot}_5H_2O$ (5.72 mg, 22.9 µmol, 10 mol-%) and $B_2(OD)_4$ (41.1 mg, 458 µmol, 2.00 equiv) were added to a microwave vial. The vial was capped and degassed by applying vacuum and refilling with N_2 for five times. Degassed D$_2$O and 4-picoline (5.6 µL, 57.3 µmol) were added using a syringe. The mixture was stirred

at room temperature for 20 h before the vial was opened. The reaction mixture was then filtered through a syringe filter (0.45 µm) and diluted with MeCN (5 mL) and water (5 mL). The solution was purified by preparative HPLC (stationary phase: RP XBridge Prep C_{18} OBD-10 µm, 50×150 mm, mobile phase: 0.25% NH_4HCO_3 solution in water, MeCN) and lyophilized to afford the title compound as a white solid (37 mg, 166 µmol, 73%). A deuterium incorporation of 97% for D−C_α and 91% for D−C_β was observed.

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 8.35 (t, J = 6.1 Hz, 1H, CONH), 8.02 (s, 1H, CONH), 7.34 – 7.19 (m, 5H, Ar-H), 4.27 (d, J = 6.0 Hz, 2H, CH_2Ph), 1.84 (s, 3H, CH_3), 1.22 – 1.17 (m, 2H, $C_\beta H_2$D).

^{13}C-NMR (151 MHz, DMSO-d_6, 300 K): δ (ppm) = 176.4 (C=O), 174.1 (C=O), 173.4 (C=O), 173.1 (C=O), 172.6 (C=O), 169.5 (C=O), 166.2 (C=O), 161.5 (C=O), 140.0 (C_{Ar}), 138.2 (C_{Ar}), 138.1 (C_{Ar}), 136.3 (C_{Ar}), 133.6 (C_{Ar}), 129.3 (C_{Ar}), 128.9 (C_{Ar}), 128.8 (C_{Ar}), 128.7 (C_{Ar}), 127.9 (C_{Ar}), 127.7 (C_{Ar}), 127.6 (C_{Ar}), 127.5 (C_{Ar}), 127.4 (C_{Ar}), 127.3 (C_{Ar}), 127.3 (C_{Ar}), 127.2 (C_{Ar}), 127.1 (C_{Ar}), 110.9 (C_{Ar}), 42.9 (CH$_2$Ph), 42.7 (CH$_2$Ph), 42.2 (CH$_2$Ph), 24.2 (C_β), 24.0 (t, J = 20.0 Hz, CH$_2$D), 22.5 (CH$_3$), 21.9 (CH$_3$), 21.5 (CH$_3$). A high number of peaks is observed due to the presence of four different isotopologues.

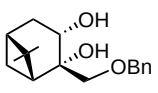
IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3290, 3068, 1663, 1648, 1549, 1496, 1452, 1427, 1377, 1295, 1183, 1030, 736, 695, 656.

m.p.: 142.4°C.

HRMS (ESI+): m/z for $C_{12}H_{14}D_2O_2N_2Na$ [M+Na]$^+$: calculated: 245.1230; found: 245.1229.

7.2.2 Fluorinated Myrtenol Derivative

(1*R*)-(−)-Myrtenoldiol benzylether (25)



To an ice cold solution of (1R)-(−)-Myrtenol (4.0 mL, 25.2 mmol, 1.00 equiv) in DMF (100 mL) was added NaH (dispersion, 60% in mineral oil) (1.5 g, 37.7 mmol, 1.50 equiv) in small portions. The mixture was stirred for 30 min before benzyl bromide (3.6 mL, 30.2 mmol, 1.20 equiv) was added. The mixture was allowed to warm to room temperature and stirred overnight. The reaction mixture was cooled to 0°C, 300 mL

of H_2O and 100 mL of brine were added and the mixture was extracted with EtOAc (4 × 100 mL). The combined organic extracts were washed with brine (3 × 100 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The residual oil was dissolved in 125 mL of iPrOH and 4-picoline (3.2 mL, 32.7 mmol, 1.30 equiv), a solution of 4-methylmorpholine *N*-oxide monohydrate (4.42 g, 32.7 mmol, 1.30 equiv) in 5 mL of H_2O and $K_2OsO_4 \cdot 2H_2O$ (185 mg, 0.50 mmol, 2 mol-%) were added to the solution. The mixture was refluxed for 20 h. Most of the solvent was evaporated and the crude product was purified by flash chromatography (silica gel, 0—40% EtOAc in petroleum ether). Fractions containing product were pooled and the solvent was evaporated. The oily residue was treated with petroleum ether yielding compound **25** as white needles (4.23 g, 15.3 mmol, 61%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 7.40 – 7.22 (m, 5H, Ar-*H*), 4.97 (d, J = 5.7 Hz, 1H, O*H*), 4.50 (dd, J = 26.3, 12.6 Hz, 2H, CH_2), 4.36 (s, 1H, O*H*), 3.81 (dt, J = 10.1, 5.3 Hz, 1H, C*H*), 3.29 (dd, J = 43.8, 10.3 Hz, 2H, CH_2), 2.30 (dddd, J = 13.6, 9.3, 3.5, 2.0 Hz, 1H, C*H*), 2.13 – 2.01 (m, 2H, 2 × C*H*), 1.78 (tdd, J = 5.9, 3.6, 2.6 Hz, 1H, C*H*), 1.54 (ddd, J = 13.8, 5.1, 2.5 Hz, 1H, C*H*), 1.38 (d, J = 9.0 Hz, 1H, C*H*), 1.17 (s, 3H, CH_3), 0.74 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 138.9 (C_{Ar}), 128.1 (C_{Ar}), 127.4 (C_{Ar}), 127.3 (C_{Ar}), 76.8 ($C\mathrm{H}_2$), 74.6 (C_{quart}), 72.4 ($C\mathrm{H}_2$), 63.8 (*C*H), 48.4 (*C*H), 40.1 (*C*H), 38.1 (C_{quart}), 37.8 ($C\mathrm{H}_2$), 27.7 ($C\mathrm{H}_3$), 27.2 ($C\mathrm{H}_2$), 23.9 ($C\mathrm{H}_2$).

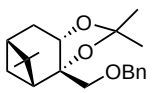
IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3364, 2991, 2955, 2910, 2859, 2161, 2032, 1496, 1455, 1439, 1389, 1366, 1331, 1296, 1227, 1200, 1116, 1087, 1038, 1016, 986, 960, 940, 903, 850, 810, 742, 698, 647.

m.p.: 83.6°C.

$[\alpha]_D^{20}$ = −2.03 (c = 0.30, DMF)

HRMS (ESI+): m/z for $C_{17}H_{24}O_3Na$ $[M+Na]^+$: calculated: 299.1618; found: 299.1616.

(1*R*)-(−)-Myrtenoldiol benzylether acetonide (26)



Compound **25** (2.00 g, 7.24 mmol, 1.00 equiv) was dissolved in 75 mL of acetone. 2,2-dimethoxypropane (2.7 mL, 21.7 mmol, 3.00 equiv) and pyridinium *p*-toluenesulfonate (273 mg, 1.09 mmol, 0.15 equiv) were added and the mixture was stirred for 16 h at room temperature. The solvent was evaporated and the residue was suspended in 10 mL of a mixture of petroleum ether and EtOAc (8:2, v/v). The solid was removed by filtration and washed with 5 mL of petroleum ether/EtOAc (8:2, v/v) twice. The combined filtrates were evaporated to afford compound **26** as pale yellow oil (2.27 g, 7.19 mmol, 99%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 7.54 – 7.10 (m, 5H, Ar-*H*), 4.50 (dd, J = 33.1, 11.8 Hz, 2H, CH_2OBn), 4.19 (d, J = 6.9 Hz, 1H, C*H*), 3.64 – 3.55 (m, 2H, CH_2), 2.18 – 2.11 (m, 1H, C*H*), 2.11 – 2.02 (m, 2H, 2 × CH_2), 1.90 – 1.77 (m, 2H, 2 × CH_2), 1.59 (d, J = 10.0 Hz, 1H, CH_2), 1.37 (s, 3H, CH_3), 1.30 (s, 3H, CH_3), 1.21 (s, 3H, CH_3), 0.71 (s, 3H, CH_3).

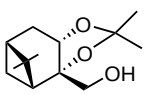
^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 138.3 (C_{Ar}), 128.1 (C_{Ar}), 127.4 (C_{Ar}), 127.3 (C_{Ar}), 107.1 (C_{quart}), 86.2 (C_{quart}), 75.3 ($C\mathrm{H}_2$), 72.2 ($C\mathrm{H}_2$), 71.2 (CH), 46.1 (CH), 37.6 (C_{quart}), 31.5 ($C\mathrm{H}_2$), 27.4 ($C\mathrm{H}_3$), 27.0 ($C\mathrm{H}_3$), 26.0 ($C\mathrm{H}_3$), 24.1 ($C\mathrm{H}_2$) 22.6 ($C\mathrm{H}_3$).

IR (silicone, neat): $\tilde{\nu}$ (cm^{-1}) = 2984, 2936, 2869, 2361, 2160, 2028, 1722, 1454, 1368, 1315, 1258, 1206, 1141, 1096, 1057, 1036, 1000, 980, 926, 893, 865, 851, 822, 736, 712, 697, 645.

$[\alpha]_D^{20} = -1.54$ (c = 0.33, DMF)

HRMS (ESI+): m/z for $C_{20}H_{28}O_3Na$ $[M+Na]^+$: calculated: 339.1931; found: 339.1931.

(1*R*)-(−)-Myrtenoldiol acetonide (27)



To a stirred solution of compound **26** (2.00 g, 6.32 mmol, 1.00 equiv) in methanol (60 mL) was added palladium on carbon (loading 10% wt, 673 mg, 0.63 mmol, 10 mol-%). The suspension was purged with nitrogen for 10 min, purged with hydrogen for another 10 min and stirred under hydrogen atmosphere for 1 h. The mixture was filtered through a plug of celite and the solvent was removed *in vacuo* yielding compound **27** as colorless oil (1.38 g, 6.23 mmol, 96%).

^{1}H-NMR (400 MHz, $CDCl_3$, 300 K): δ (ppm) = 4.08 (d, J = 7.2 Hz, 1H, C*H*), 3.75 – 3.61

(m, 1H, CH_2), 2.46 (br s, 1H, OH), 2.26 (dd, J = 6.1, 5.0 Hz, 1H, CH), 2.18 – 2.05 (m, 2H, CH_2), 1.97 – 1.84 (m, 2H, CH_2 + CH), 1.59 (d, J = 10.4 Hz, 1H, CH_2), 1.46 (s, 3H, CH_3), 1.36 (s, 3H, CH_3), 1.24 (s, 3H, CH_3), 0.78 (s, 3H, CH_3).

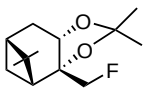
^{13}C-NMR (101 MHz, $CDCl_3$, 300 K): δ (ppm) = 108.0 (C_{quart}), 87.5 (C_{quart}), 73.4 (CH), 68.1 (CH$_2$), 45.7 (CH), 40.2 (CH), 38.1 (C_{quart}), 32.4 (CH$_2$), 27.6 (CH$_3$), 27.1 (CH$_3$), 27.0 (CH$_3$), 24.8 (CH$_2$), 23.7 (CH$_3$).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3461, 2985, 2936, 2160, 1456, 1369, 1242, 1206, 1187, 1141, 1095, 1045, 999, 976, 918, 892, 862, 822, 766, 688, 645.

$[\alpha]_D^{20} = -6.67$ (c = 0.38, DMF)

HRMS (ESI+): m/z for $C_{13}H_{22}O_3Na$ [M+Na]$^+$: calculated: 249.1461; found: 249.1464.

Fluorodeoxy-(1*R*)-(−)-myrtenoldiol acetonide (28)



Alcohol **27** (300 mg, 1.33 mmol, 1.00 equiv) was dissolved in dry toluene (15 mL). Triethylamine (0.55 mL, 3.98 mmol, 3.00 equiv) and DAST (0.53 mL, 3.98 mmol, 3.00 equiv) were added and the solution was heated to 75°C for 20 h. The mixture was cooled to room temperature and 1 mL of methanol was added. The solvent was evaporated and the crude product was purified by flash chromatography (silica gel, 0−20% EtOAc in petroleum ether) affording compound **28** as yellow oil (220 mg, 0.96 mmol, 73%).

^{1}H-NMR (400 MHz, $CDCl_3$, 300 K): δ (ppm) = 4.58 (dd, J = 24.4, 10.1 Hz, 1H, CH_2), 4.46 (dd, J = 23.1, 10.1 Hz, 1H, CH_2), 4.28 (d, J = 7.0 Hz, 1H, CH), 2.23 – 2.14 (m, 2H, CH_2), 2.11 (t, J = 5.6 Hz, 1H, CH), 2.02 (dd, J = 14.7, 4.1 Hz, 1H, CH_2), 1.97 –1.90 (m, 1H, CH), 1.72 (d, J = 10.3 Hz, 1H, CH_2), 1.51 (s, 3H, CH_3), 1.40 (s, 3H, CH_3), 1.29 (s, 3H, CH_3), 0.85 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, $CDCl_3$, 300 K): δ (ppm) = 108.6 (C_{quart}), 87.4 (d, J = 175.9 Hz, CH$_2$F), 86.4 (d, J = 16.7 Hz, C_{quart}), 70.8 (d, J = 6.3 Hz, CH), 45.9 (d, J = 5.9 Hz, CH), 40.1 (CH), 38.1 (C_{quart}), 32.2 (CH$_2$), 27.7 (CH$_3$), 27.3 (CH$_3$), 26.32 (d, J = 5.4 Hz, CH$_3$), 24.6 (CH$_2$), 23.6 (CH$_3$).

^{19}F-NMR (376 MHz, $CDCl_3$, 300 K): δ (ppm) = = −225.76 (t, J = 47.6 Hz).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 2994, 2957, 2932, 2885, 2161, 2031, 1461, 1377, 1366, 1281, 1253, 1203, 1178, 1144, 1123, 1099, 1057, 1038, 1012, 972, 936, 919, 894, 863, 853, 820, 768, 692, 648.

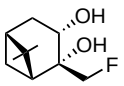
m.p.: 41.3°C.

$[\alpha]_D^{20}$ = −11.64 (c = 0.28, DMF)

HRMS (ESI+): m/z for $C_{13}H_{21}O_2FNa$ $[M+Na]^+$: calculated: 251.1418; found: 251.1419.

Fluorodeoxy-(1*R*)-(−)-myrtenoldiol (11)

To a solution of **28** (230 mg, 1.01 mmol, 1.00 equiv) in ethanol (25 mL) and water (25 mL) was added Amberlyst 15 strong cation exchange resin (300 mg). The mixture was refluxed for two days, filtered through a plug of celite and purified by flash chromatography (silica gel, 0-60% EtOAc in petroleum ether) to yield fluorinated diol **11** as pale yellow oil (141 mg, 0.75 mmol, 74%).



^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 5.26 (d, J = 6.3 Hz, 1H, OH), 4.52 (s, 1H, OH), 4.24 (dd, J = 23.8, 9.4 Hz, 1H, CH_2F), 4.12 (dd, J = 24.3, 9.4 Hz, 1H, CH_2F), 3.87 (dddd, J = 9.4, 6.3, 5.1, 1.2 Hz, 1H, CH), 2.35 (dddd, J = 13.9, 9.4, 3.6, 2.3 Hz, 1H, CH_2), 2.11 (dtdd, J = 9.6, 6.0, 2.3, 0.9 Hz, 1H, CH_2), 2.00 (t, J = 6.2 Hz, 1H, CH), 1.82 (tdd, J = 5.9, 3.6, 2.6 Hz, 1H, CH), 1.57 (ddd, J = 13.8, 5.1, 2.5 Hz, 1H, CH_2), 1.40 (d, J = 9.9 Hz, 1H, CH_2), 1.21 (s, 3H, CH_3), 0.88 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 88.4 (d, J = 173.2 Hz, CH$_2$F), 73.8 (d, J = 16.0 Hz, C_{quart}), 62.2 (d, J = 5.6 Hz, CH), 47.6 (d, J = 4.3 Hz, CH), 40.1 (CH), 38.1 (C_{quart}), 37.5 (CH$_2$), 27.6 (CH$_3$), 27.1 (CH$_2$), 24.0 (CH$_3$).

^{19}F-NMR (376 MHz, DMSO-d_6, 300 K): δ (ppm) = −228.40 (t, J = 48.0 Hz).

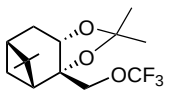
IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3362, 2925, 2361, 2161, 2031, 1455, 1386, 1368, 1335, 1273, 1216, 1154, 1127, 1066, 1048, 1023, 956, 933, 904, 847, 830, 743, 671, 625.

$[\alpha]_D^{20}$ = −2.57 (c = 0.35, DMF)

HRMS (ESI+): m/z for $C_{10}H_{17}O_2FNa$ $[M+Na]^+$: calculated: 211.1105; found: 211.1105.

7.2.3 Trifluoromethoxylated Myrtenol Derivative

Trifluoromethoxy-(1*R*)-(−)-myrtenoldiol acetonide (29)



Following a literature procedure,[6] KF (581 mg, 10.0 mmol, 4.00 equiv) was dried by heating under high vacuum. Alcohol **27** (566 mg, 2.50 mmol, 1.00 equiv), AgOTf (1.93 g, 7.50 mmol, 3.00 equiv) and SelectFluor (1.33 g, 3.75 mmol, 1.50 equiv) were added and everything was dissolved in dry EtOAc (12.5 mL, dried over activated molecular sieves 4 Å) under nitrogen atmosphere. $TMSCF_3$ (1.1 mL, 7.50 mmol, 3.00 equiv) and 2-fluoropyridine (0.65 mL, 7.50 mmol, 3.00 equiv) were added and the mixture was stirred at room temperature for one day. The mixture was filtered through a plug of celite. The solvent was evaporated and the crude product was purified by flash chromatography (silica gel, 0−20% EtOAc in petroleum ether) affording the title compound as a colorless oil (260 mg, 0.88 mmol, 35%).

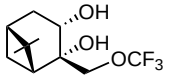
^{1}H-NMR (400 MHz, $CDCl_3$, 295 K): δ (ppm) = 4.22 (d, J = 7.1 Hz, 1H, CH), 4.11 (s, 2H, CH_2), 2.24 – 2.14 (m, 3H, 2 × CH_2 + CH), 2.07 – 1.98 (m, 1H, CH_2), 1.98 – 1.90 (m, 1H, CH), 1.75 – 1.65 (m, 1H, CH_2), 1.51 (d, J = 0.8 Hz, 3H, CH_3), 1.40 (d, J = 0.8 Hz, 3H, CH_3), 1.30 (s, 3H, CH_3), 0.86 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, $CDCl_3$, 295 K): δ (ppm) = 121.8 (q, J = 255.0 Hz, CF$_3$), 108.9 (C_{quart}), 85.1 (C_{quart}), 71.9 (CH), 71.6 (q, J = 2.7 Hz, CH$_2$), 46.1 (CH), 40.0 (CH), 38.1 (C_{quart}), 32.2 (CH$_2$), 27.7 (CH$_3$), 27.2 (CH$_3$), 26.5 (CH$_3$), 24.6 (CH$_2$), 23.3 (CH$_3$).

^{19}F-NMR (376 MHz, $CDCl_3$, 295 K): δ (ppm) = −60.60 (OCF_3).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 2988, 2941, 1460, 1405, 1382, 1370, 1269, 1211, 1132, 1055, 1028, 1000, 979, 926, 895, 866, 848, 822, 765, 633, 616.

Trifluoromethoxy-(1*R*)-(−)-myrtenoldiol (30)



To a solution of acetonide **29** (200 mg, 0.68 mmol, 1.00 equiv) in ethanol (3 mL) and water (3 mL) was added Amberlyst 15 strong cation exchange resin (300 mg). The mixture was refluxed for four days, filtered through a syringe filter (pore size 0.45 µm) and purified by flash chromatography (silica gel, 0-20% EtOAc in petroleum

ether). The resulting oil was dried *in vacuo* for a week to yield the title compound (**30**) as white needles (145 mg, 0.57 mmol, 84%).

^{1}H-NMR (400 MHz, DMSO-d_6, 295 K): δ (ppm) = 5.38 (d, J = 6.2 Hz, 1H, OH), 4.59 (s, 1H, OH), 3.95 – 3.78 (m, 3H, CH + CH_2), 2.43 – 2.31 (m, 1H, CH_2), 2.13 (dtd, J = 9.9, 6.1, 2.3 Hz, 1H, CH_2), 2.06 (t, J = 5.9 Hz, 1H, CH), 1.84 (tdd, J = 5.9, 3.6, 2.6 Hz, 1H, CH), 1.59 (ddd, J = 13.8, 5.0, 2.5 Hz, 1H, CH_2), 1.38 (d, J = 9.9 Hz, 1H, CH_2), 1.22 (s, 3H, CH_3), 0.89 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 295 K): δ (ppm) = 121.6 (q, J = 252.5 Hz, CF$_3$), 73.7 (q, J = 2.2 Hz, CH$_2$), 73.0 (C_{quart}), 62.8 (CH), 47.5 (CH), 39.9 (CH), 37.9 (C_{quart}), 37.4 (CH$_2$), 27.5 (CH$_2$), 27.0 (CH$_3$), 23.6 (CH$_3$).

^{19}F-NMR (376 MHz, DMSO-d_6, 295 K): δ (ppm) = −58.47 (OCF_3).

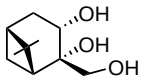
IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3266, 2993, 2917, 2876, 1457, 1411, 1256, 1218, 1132, 1064, 1037, 1015, 934, 906, 875, 827, 698, 611.

m.p.: 34—37°C.

HRMS (ESI+): m/z for $C_{11}H_{17}O_3F_3Na$ [M+Na]$^+$: calculated: 277.1022; found: 277.1024.

7.2.4 Myrtentriol

(1*R*)-(−)-Myrtentriol (4)



(1R)-(−)-Myrtenol (3.2 mL, 20.0 mmol, 1.00 equiv), $K_2OsO_4 \cdot 2H_2O$ (147 mg, 0.40 mmol, 2 mol-%) and 4-picoline (2.5 mL, 26.0 mmol, 1.30 equiv) were dissolved in 40 mL of iPrOH. A solution of 4-methylmorpholine N-oxide monohydrate (3.51 g, 26.0 mmol, 1.30 equiv) in 3 mL of H_2O was added and the mixture was refluxed for 24 h. Most of the solvent was evaporated and the residue was diluted with 5 mL of EtOAc. The crude compound was purified by flash chromatography (silica gel, 20—100% EtOAc in petroleum ether) and dried *in vacuo* to yield (1R)-(−)-myrtentriol (**4**) as a white solid (3.06 g, 16.4 mmol, 82%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 4.86 (br s, 1H, OH), 4.30 (br s, 1H, OH), 4.20 (br s, 1H, OH), 3.82 (dd, J = 9.5, 5.1 Hz, 1H, CH), 3.24 (dd, J = 55.9, 11.0 Hz, 2H,

CH_2OH), 2.31 (dddd, J = 13.7, 9.4, 3.6, 2.3 Hz, 1H, CH_2), 2.06 (dtd, J = 9.7, 6.0, 2.3 Hz, 1H, CH_2), 1.99 (t, J = 5.9 Hz, 1H, CH), 1.79 (tdd, J = 5.9, 3.6, 2.6 Hz, 1H, CH), 1.54 (ddd, J = 13.7, 5.1, 2.6 Hz, 1H, CH_2), 1.37 (d, J = 9.7 Hz, 1H, CH_2), 1.19 (s, 3H, CH_3), 0.85 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 75.1 (C_{quart}), 68.6 (CH$_2$), 63.8 (CH), 48.1 (CH), 40.2 (CH), 38.1 (C_{quart}), 37.8 (CH$_2$), 27.7 (CH$_3$), 27.3 (CH$_2$), 24.1 (CH$_3$).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3385, 2983, 2947, 2920, 2863, 2361, 2161, 2029, 1413, 1383, 1367, 1320, 1269, 1224, 1148, 1125, 1067, 1028, 958, 929, 895, 860, 829, 738, 645.

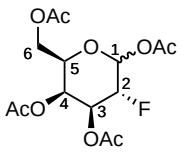
m.p.: 61.7°C.

$[\alpha]_D^{20} = -7.31$ (c = 0.26, DMF)

HRMS (ESI+): m/z for $C_{10}H_{18}O_3Na$ [M+Na]$^+$: calculated: 209.1148; found: 209.1148.

7.2.5 2-Deoxy-2-fluoro-D-galactose (FDGal)

1,3,4,6-tetra-*O*-acetyl-2-deoxy-2-fluoro-D-galactopyranose (31)



Compound **31** was synthesized following a known literature procedure.[7] 3,4,6-Tri-*O*-acetyl-D-galactal (1.04 g, 3.82 mmol, 1.00 equiv) and SelectFluor (2.03 g, 5.73 mmol, 1.50 equiv) were dissolved in a mixture of MeCN (40 mL) and AcOH (20 mL). The solution was heated to 70°C for 3 hours. The solvent was evaporated, and the residue was dissolved in 100 mL of EtOAc, washed with water (100 mL), saturated $NaHCO_3$ solution (100 mL) and brine (100 mL) and the solvent was evaporated. The crude residue was purified by flash chromatography (10-100% EtOAc in n-heptane) to yield the title compound (**31**) as a white solid (302 mg, 0.86 mmol, 23%).

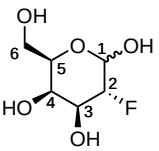
^{1}H-NMR (400 MHz, $CDCl_3$, 300 K): δ (ppm) = 6.47 (d, J = 4.0 Hz, 0.6H, H-C1α), 5.79 (dd, J = 8.0, 4.1 Hz, 1H, H-C1β), 5.52 (td, J = 3.4, 1.4 Hz, 0.6H, H-C4α), 5.49 – 5.43 (m, 1H, H-C4β), 5.44 (ddd, J = 11.0, 10.2, 3.5 Hz, 0.6H, H-C3α), 5.18 (ddd, J = 13.2, 9.8, 3.6 Hz, 1H, H-C3β), 4.89 (ddd, J = 49.2, 10.2, 4.0 Hz, 0.6H, H-C2α), 4.65 (ddd, J = 51.5, 9.9, 8.1 Hz, 1H, H-C2β), 4.31 (td, J = 6.9, 1.3 Hz, 0.6H, H-C5α), 4.20 – 4.04 (m, 4.2H, 2 × H-C6α + 2 × H-C6β + H-C5β), 2.19 (2 × s, 4.8H, CH_3α + CH_3β), 2.15 (s, 4.8H, CH_3α + CH_3β), 2.07 (s,

3H, CH_3β), 2.06 (s, 1.8H, CH_3α), (2 × s, 4.8H, CH_3α + CH_3β).

^{19}F-NMR (376 MHz, $CDCl_3$, 300 K): δ (ppm) = −208.10 (dddd, J = 51.4, 13.1, 4.1, 2.6 Hz, β-anomer), −209.12 (ddd, J = 49.2, 11.1, 3.6 Hz, α-anomer).

The spectroscopic data are in accordance with known literature data for compound **31**.[7]

2-Deoxy-2-fluoro-D-galactopyranose (13)



1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-fluoro-D-galactopyranose (**31**) (100 mg, 0.29 mmol, 1.00 equiv) was dissolved in 2 mL of MeOH. 66 µL (28.5 µmol, 10 mol-%) of a solution of sodium metal (10 mg) in MeOH (1 mL) were added. The mixture was stirred for 1 h at room temperature before being neutralized by addition of Dowex 50WX8 strongly acidic cation exchange resin (15 mg). The mixture was stirred for an additional 10 min, the resin was filtered off and the solvent was evaporated to yield a sticky oil which was dissolved in 1 mL of water. The solution was lyophilized to afford the title compound (**13**) as a colorless powder (52 mg, 0.29 mmol, quantitative yield).

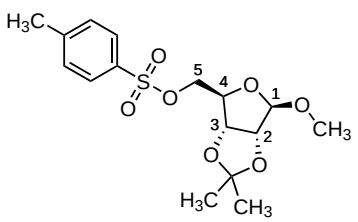
^{1}H-NMR (400 MHz, D_2O, 300 K): δ (ppm) = 5.51 (d, J = 4.0 Hz, 0.6H, *H*-C1α), 4.88 (dd, J = 7.8, 3.4 Hz, 1H, *H*-C1β), 4.77 – 4.62 (m, 0.6H, *H*-C2α), 4.37 (ddd, J = 52.0, 9.5, 7.8 Hz, 1H, *H*-C2β), 4.19 – 4.11 (m, 1.2H, *H*-C5α + *H*-C3α), 4.10 (td, J = 3.7, 1.2 Hz, 0.6H, *H*-C4α), 4.04 (app t, J = 3.4 Hz, 1H, *H*-C4β), 3.98 (ddd, J = 14.3, 9.5, 3.6 Hz, 1H, *H*-C3β), 3.82 – 3.76 (m, 4.2H, 2 × *H*-C2α + 2 × *H*-C2β + *H*-C5β).

^{19}F-NMR (376 MHz, D_2O, 300 K): δ (ppm) = −207.36 (ddt, J = 52.1, 14.5, 3.0 Hz, β-anomer), −207.53 (ddd, J = 49.7, 12.5, 3.8 Hz, α-anomer).

The spectroscopic data are in accordance with known literature data for compound **13**.[8,9]

7.2.6 Methyl 5-deoxy-5-fluoro-β,D-ribofuranoside (FDRibOMe)

Methyl 2,3-*O*-isopropylidene-5-*O*-(toluenesulfonyl)-β,D-ribofuranoside (32)

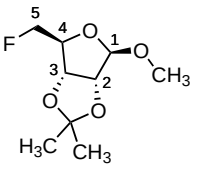


Following a literature procedure,[10] methyl 2,3-*O*-isopropylidene-5-*O*-β,D-ribofuranoside (1.50 g, 7.35 mmol, 1.00 equiv) was dissolved in dry pyridine (4.5 mL) and cooled to 0°C. *p*-Toluenesulfonyl chloride (2.10 g, 11.0 mmol, 1.50 equiv) was added in portions. The mixture was stirred for 15 min at 0°C and an additional 2 h at room temperature, before being poured into 20 mL of ice water. The white precipitate was collected by filtration, washed with ice water (3 × 10 mL) and dried *in vacuo* to afford tosylate **32** as a white solid (2.56 g, 7.14 mmol, 97%).

^{1}H-NMR (400 MHz, $CDCl_3$, 300 K): δ (ppm) = 7.82 – 7.79 (m, 2H, Ar-*H*), 7.36 (d, *J* = 7.9 Hz, 2H, Ar-*H*), 4.93 (app s, 1H, *H*-C1), 4.60 (dd, *J* = 5.9, 1.0 Hz, 1H, *H*-C3), 4.53 (d, *J* = 5.9 Hz, 1H, *H*-C2), 4.31 (td, *J* = 7.1, 1.0 Hz, 1H, *H*-C4), 4.02 (app dd, *J* = 7.2, 1.9 Hz, 2H, *H*-C5), 3.24 (s, 3H, OCH_3), 2.46 (s, 1H, Ar-CH_3), 1.45 (s, 3H, CH_3), 1.29 (s, 3H, CH_3).

The spectroscopic data are consistent with the literature.[10]

Methyl 2,3-*O*-isopropylidene-5-deoxy-5-fluoro-β,D-ribofuranoside (33)



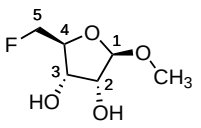
Tosylate **32** (2.00 g, 5.58 mmol, 1.00 equiv) was dissolved in TBAF solution (20 mL, 1 M in THF) under nitrogen atmosphere. The solution was heated to 100°C in a sealed vessel for 15 h. The solvent was evaporated, and the crude oil was purified by flash chromatography (0-10% EtOAc in *n*-heptane) yielding the title compound (**33**) as a colorless oil (872 mg, 4.23 mmol, 76%).

^{1}H-NMR (400 MHz, $CDCl_3$, 300 K): δ (ppm) = 4.99 (d, *J* = 2.4 Hz, 1H, *H*-C1), 4.70 (d, *J* = 6.0 Hz, 1H, *H*-C3), 4.59 (d, *J* = 5.9 Hz, 1H, *H*-C2), 4.47 – 4.27 (m, 3H, *H*-C4 + *H*-C5), 3.33 (s, 3H, OCH_3), 1.49 (s, 3H, CH_3), 1.33 (s, 3H, CH_3).

^{19}F{^{1}*H*}-NMR (376 MHz, $CDCl_3$, 300 K): δ (ppm) = −225.01.

The spectroscopic data are consistent with the literature.[10]

Methyl 5-deoxy-5-fluoro-β,D-ribofuranoside (FDRib-OMe) (12)



Acetonide **33** (800 mg, 3.88 mmol, 1.00 equiv) was dissolved in AcOH (7 mL) and water (3 mL). The solution was stirred for 3 h at 75°C in a sealed vessel. The solvent was evaporated and the product was purified by flash chromatography (0-10% MeOH in DCM) yielding FDRib-OMe (**12**) (331 mg, 1.99 mmol, 51%) as a pale yellow oil.

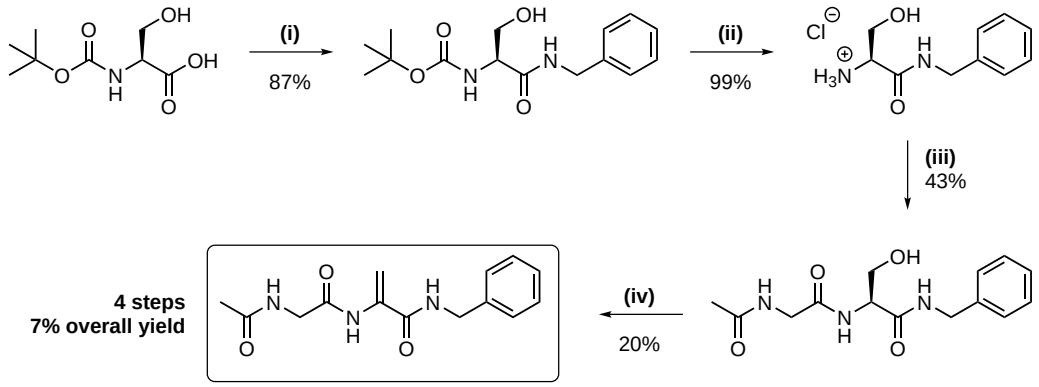
^{1}H-NMR (400 MHz, D_2O, 300 K): δ (ppm) = 4.98 (s, 1H, *H*-C1), 4.77 – 4.49 (m, 2H, *H*-C5), 4.31 (dd, J = 7.2, 4.6 Hz, 1H, *H*-C4), 4.21 (dddd, J = 24.8, 7.4, 5.0, 2.5 Hz, 1H), 4.09 (dt, J = 4.7, 1.4 Hz, 1H) 3.41 (s, 3H, OCH_3).

^{13}C-NMR (101 MHz, D_2O, 300 K): δ (ppm) = 108.1 (C1), 83.5 (d, J = 168.5 Hz, C5), 80.8 (d, J = 18.0 Hz, C4), 73.9 (d, J = 1.8 Hz, C2), 69.6 (d, J = 7.2 Hz, C3), 55.2 (CH3).

^{19}F-NMR (376 MHz, D_2O, 300 K): δ (ppm) = −228.21 (td, J = 47.2, 25.0 Hz).

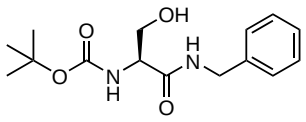
The spectroscopic data are consistent with the literature.[11]

7.2.7 Ac-Gly-Dha-NHBn (17)



Scheme 7.1: Synthesis of Ac-Gly-Dha-NHBn: (**i**) NMM, IBCF, $BnNH_2$, EtOAc, −10°C—rt; (**ii**) HCl, dioxane, DCM; (**iii**) *N*-acetyl glycine, NMM, IBCF, DMF; (**iv**) MsCl, Et_3N, DMF.

Boc-Ser-NHBn (34)



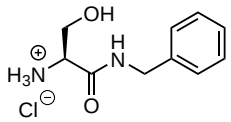
Boc-Ser-OH (1.00 g, 4.87 mmol, 1.00 equiv) and *N*-methylmorpholine (589 µL, 5.36 mmol, 1.10 equiv) were dissolved in EtOAc (30 mL). The solution was cooled to −10°C before isobutyl chloroformate (732 µL, 5.36 mmol, 1.10 equiv) and a solution of benzylamine (586 µL, 5.36 mmol, 1.10 equiv) in EtOAc (10 mL) were added subsequently. The mixture was kept at 10 to 15°C for 2 hours, water was added (20 mL) and stirring continued for 10 min. The organic layer was separated and washed with 1 M HCl (20 mL), saturated aqueous Na_2CO_3 solution (20 mL) and brine (20 mL) and the solvent was evaporated. The residue was titurated with petroleum ether and dried *in vacuo* to yield the title compound as a white solid (1.25 g, 4.23 mmol, 87%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 8.30 (t, J = 5.8 Hz, 1H, CON*H*), 7.33 – 7.17 (m, 5H, Ar-*H*), 6.65 (d, J = 7.9 Hz, 1H, CON*H*), 4.84 (t, J = 5.7 Hz, 1H, O*H*), 4.29 (app qd, J = 15.3, 5.9 Hz, 2H, CH_2), 4.05 – 3.96 (m, 1H, C*H*), 3.64 – 3.51 (m, 2H, CH_2), 1.39 (s, 9H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 170.5 (*C*=O), 155.2 (*C*=O), 139.4 (C_{Ar}), 128.1 (C_{Ar}), 126.9 (C_{Ar}),126.6 (C_{Ar}), 78.2 (C_{quart}), 61.8 (CH_2), 57.0 (*C*H), 42.0 (CH_2), 28.2 (CH_3).

The NMR data are in agreement with literature values.[12]

H-Ser-NHBn Hydrochloride (35)



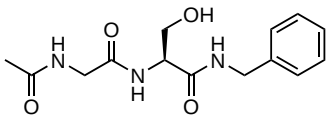
Boc-Ser-NHBn (**34**) (3.75 g, 12.7 mmol, 1.00 equiv) was dissolved in DCM (120 mL). HCl (4 M in dioxane, 31.9 mL, 127 mmol, 10.0 equiv) was added and the mixture was stirred at room temperature for an hour. The solvent was evaporated and the title compound (**35**) was afforded as a white solid (2.92 g, 12.6 mmol, 99%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 9.09 (t, J = 5.9 Hz, 1H, CONH), 8.29 (s, 3H, NH_3Cl), 7.35 – 7.22 (m, 5H, Ar-H), 5.55 (t, J = 4.9 Hz, 1H, OH), 4.34 (d, J = 5.9 Hz, 2H, CH_2), 3.88 (t, J = 5.0 Hz, 1H, CH), 3.85 – 3.73 (m, 2H, CH_2).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 166.8 (C=O), 138.7 (C_{Ar}), 128.3 (C_{Ar}), 127.2 (C_{Ar}), 126.9 (C_{Ar}), 60.3 (CH_2), 54.4 (CH), 42.2 (CH_2).

The NMR data are in agreement with literature values.[12]

Ac-Gly-Ser-NHBn (36)



N-acetyl glycine (1.02 g, 8.67 mmol, 1.00 equiv), *N*-methylmorpholine (2.38 mL, 21.7 mmol, 2.50 equiv) and isobutyl chloroformate (1.12 mL, 8.67 mmol, 1.00 equiv) were dissolved in DMF (100 mL). The mixture was stirred at room temperature for 15 min, before H-Ser-NHBn hydrochloride (2.00 g, 8.67 mmol, 1.00 equiv) was added as a solid in one portion. The mixture was stirred overnight and the solvent was evaporated. The crude product was purified by reversed-phase flash column chromatography (C_{18} silica, 0-100% MeCN in H_2O) and lyophilized to yield Ac-Gly-Ser-NHBn (**36**) as a white solid (1.10 g, 3.75 mmol, 43%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 8.36 (t, J = 6.0 Hz, 1H, CONH), 8.13 (t, J = 5.6 Hz, 1H, CONH), 7.96 (d, J = 7.9 Hz, 1H, CONH), 7.35 – 7.17 (m, 5H, Ar-H), 4.94 (t, J = 5.4 Hz, 1H, OH), 4.35 – 4.26 (m, 3H, CH + CH_2), 3.81 – 3.56 (m, 4H, 2 × CH_2), 1.84 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 169.9 (C=O), 169.8 (C=O), 169.2

(*C*=O), 139.3 (C_{Ar}), 128.2 (C_{Ar}), 127.0 (C_{Ar}),126.6 (C_{Ar}), 61.6 (CH_2), 55.3 (*C*H), 42.2 (CH_2), 42.0 (CH_2), 22.4 (CH_3).

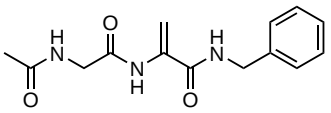
IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3273, 1672, 1650, 1562, 1428, 1284, 1092, 1062, 1032, 725, 691.

m.p.: 184−187°C.

$[\alpha]_D^{20}$ = −7.31 (c = 0.26, DMF)

HRMS (ESI+): *m/z* for $C_{14}H_{20}O_4N_3$ $[M+H]^+$: calculated: 294.1448; found: 294.1449.

Ac-Gly-Dha-NHBn (17)



Ac-Gly-Ser-NHBn (**36**) (200 mg, 0.68 mmol, 1.00 equiv) was dissolved in DMF (5 mL) and cooled to 0°C. MsCl (74 µL, 0.95 mmol, 1.40 equiv) was added followed by Et_3N (237 µL, 1.70 mmol, 2.50 equiv). The mixture was allowed to slowly warm to room temperature and stirred overnight. Another portion of MsCl (74 µL, 0.95 mmol, 1.40 equiv) and Et_3N (237 µL, 1.70 mmol, 2.50 equiv) were added and the mixture was stirred for an additional day. The title compound (**17**) was obtained after flash chromatography (twice, silica, 0-10% MeOH in DCM) as a white solid (37.5 mg, 0.14 mmol, 20%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 9.11 (s, 1H, CON*H*), 8.95 (t, J = 6.0 Hz, 1H, CON*H*), 8.29 (t, J = 5.9 Hz, 1H, CON*H*), 7.37 – 7.19 (m, 5H, Ar-*H*), 6.13 (s, 1H, C=C*H*), 5.55 (s, 1H, C=C*H*), 4.37 (d, J = 6.0 Hz, 2H, CH_2), 3.80 (d, J = 5.9 Hz, 2H, CH_2), 1.88 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 169.9 (*C*=O), 168.6 (*C*=O), 163.6 (*C*=O), 139.1 (C_{Ar}), 135.0 (*C*=CH_2), 128.3 (C_{Ar}), 127.2 (C_{Ar}), 126.8 (C_{Ar}), 102.8 (C=*C*H_2), 43.1 (CH_2), 42.6 (CH_2), 22.3 (CH_3).

IR (diamond ATR, neat): $\tilde{\nu}$ (cm^{-1}) = 3275, 1678, 1658, 1625, 1554, 1532, 1305, 1246, 972, 730, 712, 696, 642.

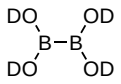
m.p.: 169−172°C.

HRMS (ESI+): *m/z* for $C_{14}H_{17}O_3N_3Na$ $[M+Na]^+$: calculated: 298.1162; found: 298.1161.

7.2.8 Miscellaneous Compounds

Tetradeuteroxydiboron ($B_2(OD)_4$, 37)

Tetradeuteroxydiboron ($B_2(OD)_4$, **37**) was synthesized following a literature procedure.[13]



A 50 mL Schlenk flask was equipped with a dropping funnel and evacuated. The flask was thoroughly heated with a heat gun and allowed to cool under vacuum before being filled with N_2. Tetrakis(dimethylamino)diboron (3 mL, 13.0 mmol, 1.00 equiv) and D_2O (99.8% D, 5 mL) were added and the flask was cooled to 0°C. Deuterium chloride (35% wt in D_2O, 99% D) (15 mL) was added dropwise via the dropping funnel over 20 min leading to formation of a white precipitate. The mixture was allowed to warm to room temperature and stirred for an additional hour. The precipitate was collected by filtration, washed with 3 mL of 10% DCl in D_2O and 3 mL of $CDCl_3$ and dried *in vacuo* to yield the title compound (**37**) as a white solid (854 mg, 9.09 mmol, 70%).

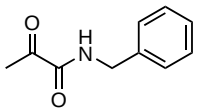
98% deuterium incorporation was determined by ^{1}H-NMR in DMSO-d_6 against mesitylene as internal standard.

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 8.66 (s, 0.008H), 7.62 (s, 0.013H).

^{11}B-NMR (128 MHz, DMSO-d_6, 300 K): δ (ppm) = 31.1 (s).

The NMR data are in agreement with previous literature reports.[13]

N-Benzyl-2-oxopropanamide (38)



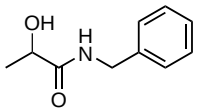
Pyruvic acid (355 µL, 5.04 mmol, 1.10 equiv) was dissolved in DMF (15 mL) and cooled to 0°C. HATU (2.26 g, 5.95 mmol, 1.30 equiv) and DIPEA (2.4 mL, 13.7 mmol, 3.00 equiv) were added subsequently and the mixture was stirred for 10 min. Benzylamine (500 µL, 4.58 mmol, 1.00 equiv) was added dropwise, the solution was allowed to warm to room temperature and stirred for 2 hours. The mixture was diluted with 1 mM HCl (20 mL) and extracted with EtOAc (3 × 20 mL). The crude product was purified by flash chromatography (silica gel, 0-100% EtOAc in petroleum ether) to yield the title compound (**38**) as a yellow oil (142 mg, 0.80 mmol, 18%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 9.10 (t, J = 5.8 Hz, 1H, CONH), 7.35 – 7.19 (m, 5H, Ar-H), 4.31 (d, J = 6.5 Hz, 2H, CH_2), 2.36 (s, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 197.2 (C=O), 161.2 (C=O), 138.7 (C_{Ar}), 128.3 (C_{Ar}), 127.4 (C_{Ar}), 126.9 (C_{Ar}), 42.1 (CH$_2$), 25.0 (CH$_3$).

The analytical data are in accordance with the literature.[14]

N-Benzyl-2-hydroxypropanamide (39)



N-Benzyl-2-oxopropanamide (**38**) (50.0 mg, 0.28 mmol, 1.00 equiv) was dissolved in MeOH (5 mL). To the stirred solution was added $NaBH_4$ (21.4 mg, 0.56 mmol, 2.00 equiv) in one portion and the mixture was stirred for one hour. The solvent was evaporated and the crude product was purified by flash chromatography (silica gel, 2-10% MeOH in DCM) to yield the title compound (**39**) as a colorless oil (22.8 mg, 0.13 mmol, 45%).

^{1}H-NMR (400 MHz, DMSO-d_6, 300 K): δ (ppm) = 8.22 (t, J = 6.1 Hz, 1H, CONH), 7.36 – 7.17 (m, 5H, Ar-H), 5.51 (br s, 1H, OH), 4.28 (d, J = 6.3 Hz, 2H, CH_2), 4.01 (q, J = 6.8 Hz, 1H, CH), 1.23 (d, J = 6.8 Hz, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6, 300 K): δ (ppm) = 174.5 (C=O), 139.8 (C_{Ar}), 128.2 (C_{Ar}), 127.1 (C_{Ar}), 126.6 (C_{Ar}), 67.3 (CH), 41.6 (CH$_2$), 21.1 (CH$_3$).

The analytical data are in accordance with the literature.[15]

7.2.9 Mechanistic Experiments

Reaction Optimization on Ac-Dha-NHBn

$CuSO_4{\cdot}5H_2O$ (1.1 mg, 4.58 µmol, 10 mol-%), 4-picoline (1.11 µL, 11.5 µmol, 25 mol-%), 2-Acetylamino-*N*-benzyl-acrylamide (**1**) (10 mg, 45.8 µmol, 1.00 equiv) and $B_2(OH)_4$ (8.2 mg, 91.6 µmol, 2.00 equiv) were dissolved in D_2O (1 mL). The reaction mixture was stirred vigorously (1500 rpm) at room temperature for 3 h. The reaction mixture was filtered through a syringe filter (0.45 µm pore size) and 500 µL of the mixture were taken for NMR analysis. Internal standard (maleic acid, 0.5 M solution in D_2O, 45.8 µL, 22.9 µmol, 1.00 equiv) was

added to the previously taken aliquot and the mixture was briefly vortexed. Yields were determined by ^{1}H-NMR by integration of the C_β proton signals against the maleic acid internal standard.

Screenings were conducted by deviating the parameter of interest from these standard conditions.

Borylation under N_2, H_2 and O_2 atmosphere

$CuSO_4 \cdot 5H_2O$ (1.1 mg, 4.58 µmol, 10 mol-%) was dissolved in D_2O (1 mL). The solution was degassed by four freeze-pump-thaw cycles. 2-Acetylamino-*N*-benzyl-acrylamide (**1**) (10 mg, 45.8 µmol, 1.00 equiv) and $B_2(OD)_4$ (12.3 mg, 137 µmol, 3.00 equiv) were put under nitrogen atmosphere in a capped microwave vial. The previously prepared $CuSO_4$ solution was added. For reactions under hydrogen or oxygen atmosphere, the solution was purged with H_2 or O_2 for 15 seconds. 4-picoline (1.11 µL, 11.5 µmol, 25 mol-%) was added with a Hamilton glass syringe and the mixture was stirred overnight. Yields were determined by ^{1}H-NMR against maleic acid as internal standard.



Figure 7.1: Borylation under N_2, H_2 and O_2 atmosphere. From left to right: nitrogen, oxygen and hydrogen atmosphere.

Determination of Cu-particle size

$CuSO_4 \cdot 5H_2O$ (5 mg) and $B_2(OH)_4$ (10 mg) were dissolved in water. 4-picoline (20 µL) was added and the mixture was vortexed briefly. The resulting black mixture was passed through syringe filters of different pore size to estimate the size of particular matter. No retention was observed for 0.8 µm pore size (Sartorius, Catalogue No. 16592-K) while a colourless, clear solution was obtained for 0.45 µm pore size (Sartorius, Catalogue No. 16555-K) and 0.2 µm pore size (VWR, Catalogue No. 28145-483).



Figure 7.2: Determination of Cu-particle size by filtration using syringe filters with various pore sizes. Resulting solution in 2 mL glass vials from left to right: 0.8 µm, 0.45 µm, 0.2 µm.

Ketoamide reduction with $B_2(OH)_4$

N-Benzyl-2-oxopropanamide (**38**) (5.0 mg, 28.2 µmol, 1.00 equiv) was dissolved in water (2 mL). Tetrahydroxydiboron (25.3 mg, 282 µmol, 10.0 equiv) was added and the mixture was stirred at room temperature. After 24 h, the mixture was analysed by TLC (EtOAc/petroleum ether 1:1, $KMnO_4$ staining). Comparison with starting material and authentic reference confirmed formation of *N*-Benzyl-2-hydroxypropanamide (**39**).

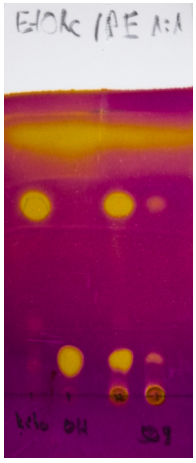


Figure 7.3: TLC analysis of the reduction of *N*-Benzyl-2-oxopropanamide with tetrahydroxydiboron. From left to right: *N*-Benzyl-2-oxopropanamide (**38**), *N*-Benzyl-2-hydroxypropanamide (**39**), co-spot, reaction mixture.

Cleavage on Model Substrate Ac-Gly-Dha-NHBn (17)

Ac-Gly-Dha-NHBn (**17**) (100 mg, 0.36 mmol, 1.00 equiv) was suspended in water (30 mL). Tetrahydroxydiboron (325 mg, 3.63 mmol, 10.0 equiv) was added and the mixture was vortexed for 5 min. The mixture was stirred at room temperature for 24 h, lyophilized and purified by column chromatography (0-10% MeOH in DCM). *N*-acetylglycine (5.2 mg, 40.4 µmol, 12%) and *N*-acetylglycinamide (5.9 mg, 50.8 µmol, 14%) were isolated and identified by ^{1}H-NMR analysis and comparison with authentic references.

7.3 Model Substrate NMR Studies

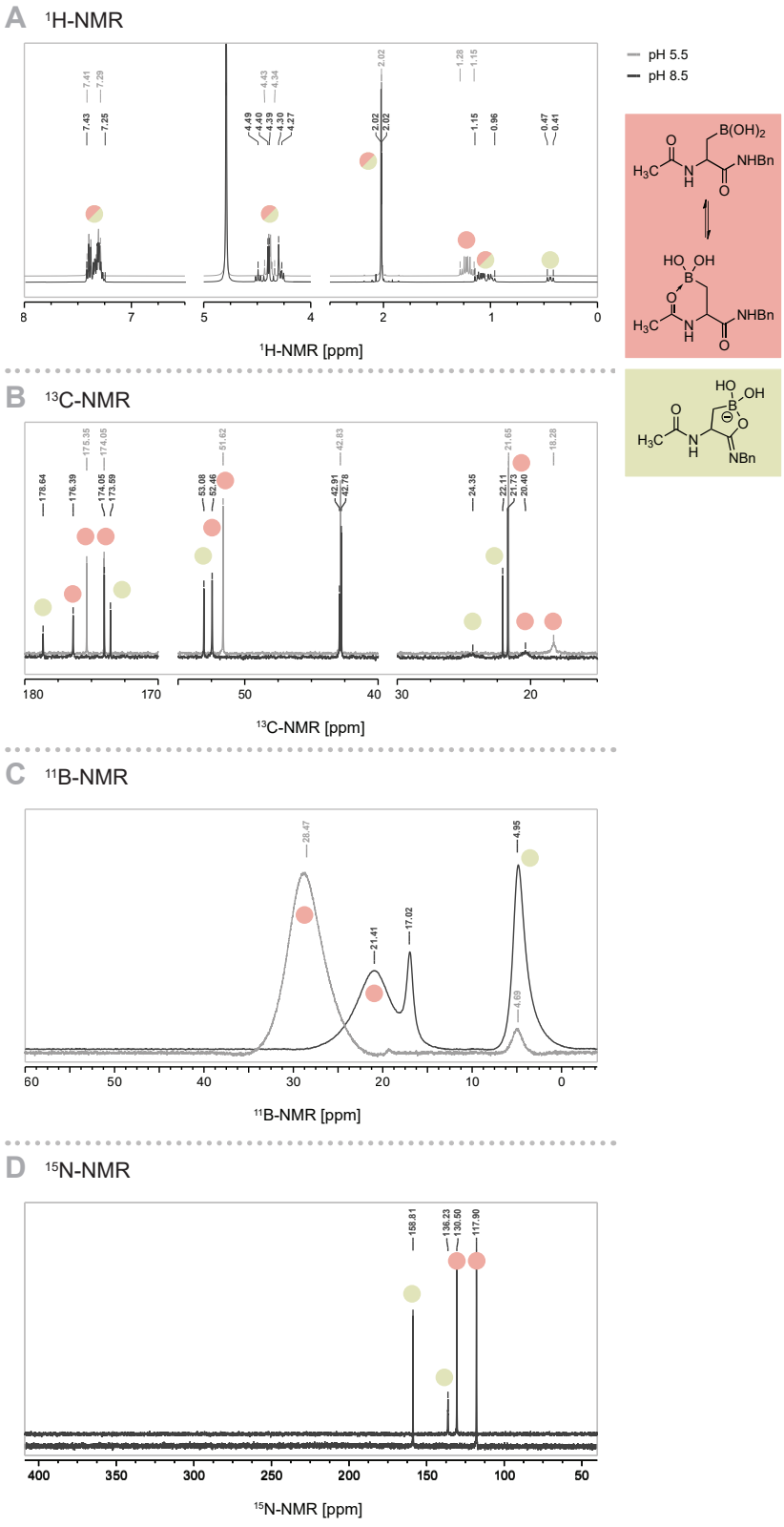


Figure 7.4: Comparison of ^{1}H-, ^{13}C-, ^{11}B- and ^{15}N-NMR spectra of Ac-Bal-NHBn (**3**)

7.3.1 pH Dependency of Oxaborolane Formation

The pH dependency of oxaborolane formation on model compound Ac-Bal-NHBn (**3**) was observed by ^{1}H-NMR using solutions of Ac-Bal-NHBn (20 mg/mL) in NaP_i buffer (20 mM in D_2O). The pH was adjusted with conc. DCl and 1 M NaOH in D_2O. Quantification was conducted by integration of the C_β-H signals, which are distinctly different for the boronic acid and the oxaborolane. The data were processed with GraphPad Prism 8.0.0 (GraphPad Software Inc., San Diego, USA) and fitting was performed using a sigmoidal standard curve:

```
Y = Bottom + (Top-Bottom)/(1+10^((logIC50-X)*HillSlope))
```

Table 7.1: oxaborolane formation at varying pH observed via ^{1}H-NMR

Entry	measured pH	oxoborolane [%]
1	5.40	0
2	6.01	0
3	7.06	2
4	7.94	14
5	9.08	50
6	10.16	63
7	12.02	64

oxaborolane formation relative to boronic acid as determined by ^{1}H-NMR.

7.3.2 Influence of Concentration

A stock solution of Ac-Bal-NHBn (**3**) (30 mg/mL) in NaP_i buffer (20 mM, pH 8.50, D_2O) was prepared and the pH was adjusted to pH 8.50 using 1 M NaOH in D_2O. A two-fold dilution series was prepared by dilution of the stock solution into the same volume of NaP_i buffer (20 mM, pH 8.50, D_2O). The pH of the resulting solutions was checked before analysis via NMR spectroscopy.

Table 7.2: oxaborolane formation at varying concentration observed via ^{1}H-, ^{11}B- and ^{13}C-NMR

c [mg/mL]	**c [mM]**	**^{1}H-NMR**		**^{11}B-NMR**	**^{13}C-NMR**	
		C_β−H	CH_3		C_β	C_α
30.00	113.6	32.9	36.9	34.5	33.6	33.5
15.00	56.8	33.0	34.7	36.5	33.3	33.0
7.50	28.4	33.0	35.4	37.5	31.9	33.6
3.75	14.2	33.2	35.3	nd	nd	nd
1.88	7.1	33.1	37.4	nd	nd	nd

oxaborolane formation is given in % relative to all species.

7.3.3 ^{15}N-NMR Measurements

^{15}N-NMR studies were conducted on a Bruker AVIII HD 600 equipped with a Prodigy N2 broadband cryoprobe. The [^{15}N]-labelled model substrate was dissolved in NaP_i buffer (20 mM, pH 8.50, D_2O) and the pH was adjusted to pH 8.50 using 1 M NaOH in D_2O.

7.4 Protein Expression and Purification

7.4.1 Histone H3 Cys10

Protein Sequence

```
1          11         21         31         41         51
ARTKQTARKC TGGKAPRKQL ATKAARKSAP ATGGVKKPHR YRPGTVALRE IRRYQKSTEL
61         71         81         91         101        111
LIRKLPFQRL VREIAQDFKT DLRFQSSAVM ALQEASEAYL VALFEDTNLA AIHAKRVTIM
121        131
PKDIQLARRI RGERA
```

Formula: $C_{670}H_{1131}N_{215}O_{185}S_3$

Calculated molecular weight: 15254.9 g/mol

Extinction coefficient $\varepsilon = 4470$ M^{-1} cm^{-1}

Plasmid Amplification

The plasmid was transformed into Agilent XL10-Gold ultra-competent *E. coli* (Agilent) according to the manufacturer's transformation protocol and grown on agar plates (supplemented with ampicillin) over night. A single colony was selected and transferred to 5 mL of LB media containing 100 µg/mL ampicillin. The culture was incubated at 37°C for 14 h and the cell pellet was harvested. Plasmid purification was conducted using a QIAprep Spin Miniprep kit and following the manufacturer's instructions.

Protein Expression and Purification

Protein Expression

BL21(DE3)pLysS competent *E. coli* (Agilent) were transformed with the plasmid following the manufacturer's instructions and grown on agar plates (supplemented with 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol) over night. Single colonies were selected and transferred to 5 mL of LB media (containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol) each. The starter cultures were incubated at 37°C for 16 h. One starter culture was prepared for each 450 mL of expression media. Eight 2.5 L Erlenmeyer flasks charged with

450 mL of LB media supplemented with carbenicillin (at final concentration of 100 µg/mL) and chloramphenicol (at a final concentration of 34 µg/mL) were inoculated with the starter cultures. The bacterial cultures were incubated at 37°C with vigorous shaking (250 rpm) an optical density value (OD_{600}) of 0.6 – 0.8 was observed. Protein expression was induced by addition of 1 mM IPTG and the culture was further incubated at 37°C for 2 h. The cells were harvested by centrifugation (7,500 rpm, 20 min, 4 °C). The cell pellet was suspended in 25 mL of lysis buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 7.5) containing a tablet of cOmplete™ mini EDTA-free protease inhibitor cocktail. The sample was frozen in liquid nitrogen and stored at −80 °C until purification.

Protein Purification

The samples were thawed on ice and DNase I (1 mg) was added. The cell suspension was then sonicated on ice (5 cycles, 40% amplitude, 30 sec on, 60 sec off). The samples were centrifuged (20,000 rpm, 20 min, 4°C) and the supernatant was discarded. The cell pellet was re-suspended in 20 mL of TW buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, 1% Triton X-100, pH 7.5) using a mixer and sonicated on ice (2 cycles, 40% amplitude, 30 sec on, 60 sec off). The mixture was centrifuged (20,000 rpm, 10 min, 4°C) and the supernatant was discarded. The pellet was again suspended in 20 mL of TW buffer, centrifuged (20,000 rpm, 10 min, 4°C) and the supernatant was discarded. Washing was repeated one more time with TW buffer and one more time with wash buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 7.5). The pellet was incubated with 1 mL of DMSO at room temperature for 10 min before 10 mL of unfolding buffer (7 M Gdn·HCl, 10 mM TRIS base, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine) were added. The mixture was shaken at room temperature for 1 h and centrifuged (20,000 rpm, 10 min, 4°C). The supernatant was purified by FPLC using a size-exclusion column running with SAU-100 buffer (7 M urea, 20 mM NaOAc, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 5.2). The Histone H3 containing fractions were further purified by anion exchange chromatography using a linear gradient from SAU-100 buffer to SAU-1000 buffer (7 M urea, 20 mM NaOAc, 1000 mM NaCl, 1 mM EDTA, 5 mM BME, pH 5.2). The clean fractions were pooled together and dialyzed four times against 4 L of water containing 2 mM BME. The protein solution was lyophilized

yielding 230 mg of Histone H3 Cys10 (64 mg/L expression volume). The lyophilized protein stored at −20°C until needed.

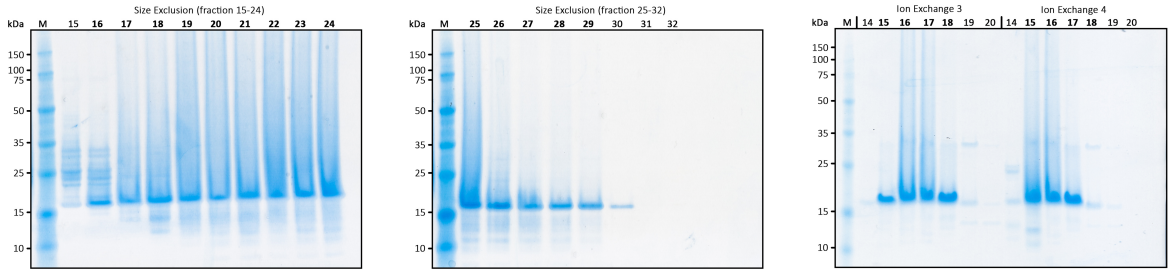


Figure 7.5: Purification and analysis of of purified fractions of Histone H3 Cys10 via SDS-PAGE and coomassie staining (M = marker).

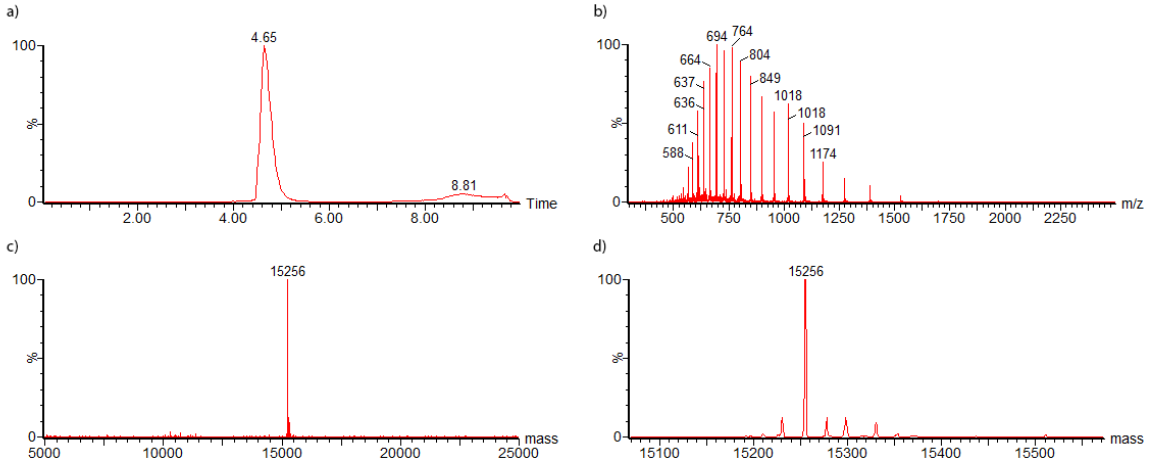


Figure 7.6: LC-MS analysis of Histone H3 Cys10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15255 g/mol; observed masses: 15256 g/mol.

7.4.2 Histone H3 Cys9

Protein Sequence

```
1           11          21          31          41          51
ARTKQTARCS TGGKAPRKQL ATKAARKSAP ATGGVKKPHR YRPGTVALRE IRRYQKSTEL
61          71          81          91          101         111
LIRKLPFQRL VREIAQDFKT DLRFQSSAVM ALQEASEAYL VALFEDTNLA AIHAKRVTIM
121         131
PKDIQLARRI RGERA
```

Formula: $C_{667}H_{1124}N_{214}O_{186}S_3$

Calculated molecular weight: 15213.8 g/mol

Extinction coefficient $\varepsilon = 4470\ \text{M}^{-1}\ \text{cm}^{-1}$

Protein Expression and Purification

Histone H3 Cys9 was expressed in BL21(DE3)pLysS competent *E. coli* (Agilent) using a total of 3.6 L culture volume and purified as described earlier. The protein solution was lyophilized yielding 187 mg of Histone Cys9 (52 mg/L expression volume). The lyophilized protein stored at −20°C until needed.

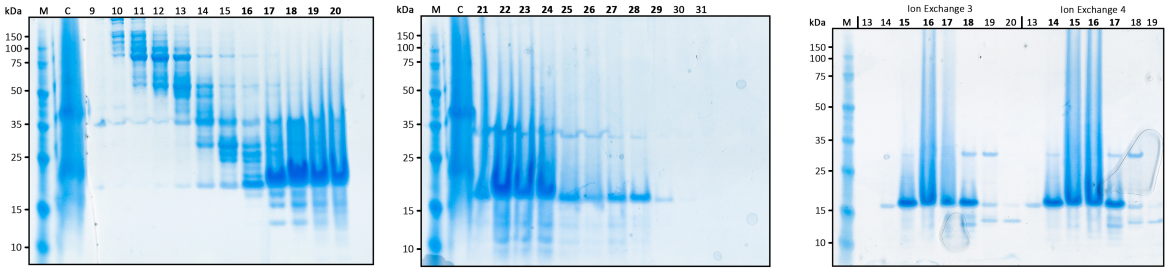


Figure 7.7: Purification and analysis of of purified fractions of Histone H3 Cys9 via SDS-PAGE and coomassie staining (M = marker).

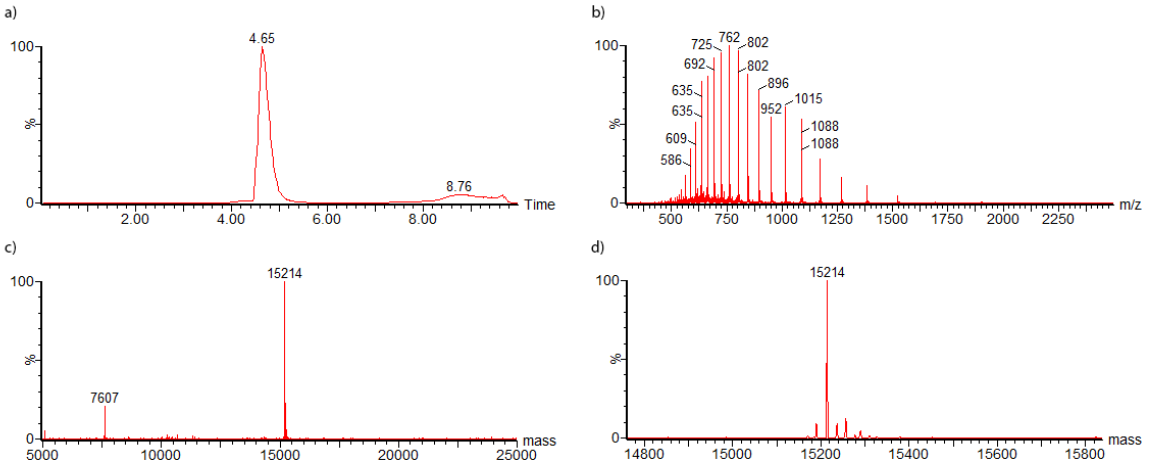


Figure 7.8: LC-MS analysis of Histone H3 K9C: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15214 g/mol; observed masses: 15214 g/mol.

7.4.3 [^{15}N]Histone H3 Cys9 and [^{15}N]Histone H3 Cys10

Transformation and Expression

M9 Media Production

M9 media was produced following a modified literature procedure.[16]

In 1 L of milli-Q H_2O were dissolved

6.0 g	Na_2HPO_4
3.0 g	KH_2PO_4
0.5 g	NaCl
1.0 g	[^{15}N]NH_4Cl

The pH was adjusted to 7.4 with HCl/NaOH as needed and the solution was autoclaved and cooled to room temperature. Before use,

11 mg	$CaCl_2$
120 mg	$MgSO_4$
3.0 g	glucose
10 mg	biotin
10 mg	thiamine

and antibiotics (100 μg/mL carbenicillin and 34 μg/mL chloramphenicol) were added.

Protein Expression and Purification

Protein Expression

BL21(DE3)pLysS competent *E. coli* (Agilent) were transformed with the plasmid encoding for Histone H3 Cys9 or Histone H3 Cys10 following the manufacturer's instructions and grown on agar plates (supplemented with 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol) over night. Single colonies were selected and transferred to 2 × 15 mL LB medium (supplemented with 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol). The starter cultures were

grown at 37°C for 3 h, centrifuged (3,500 rpm, 10 min, 4°C) and the LB medium was replaced with M9 medium (35 mL each). The starter cultures were grown overnight at 30°C and used to inoculate 2 × 500 mL M9 medium. Expression cultures were grown at 37°C until $OD_{600} = 0.7$, induced with 0.5 mM IPTG and grown at 30°C for 14 h. The cell pellets was harvested and suspended in 25 mL of lysis buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 7.5) containing a tablet of cOmplete™ mini EDTA-free protease inhibitor cocktail.

Protein Purification

DNase I (1 mg) was added and the cell suspension was then sonicated on ice (5 cycles, 40% amplitude, 30 sec. on, 60 sec. off). The sample was centrifuged (20,000 rpm, 20 min, 4°C) and the supernatant was discarded. The cell pellet was re-suspended in 20 mL of TW buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, 1% Triton X-100, pH 7.5) and sonicated on ice (2 cycles, 40% amplitude, 30 sec. on, 60 sec. off). The mixture was centrifuged (20,000 rpm, 10 min, 4 °C) and the supernatant was discarded. Washing was repeated wash buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 7.5). The pellet was incubated with 1 mL of DMSO at room temperature for 10 min before 10 mL of unfolding buffer (7 M Gdn·HCl, 10 mM TRIS base, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine) were added. The mixture was shaken at room temperature for 1 h and centrifuged (20,000 rpm, 10 min, 4°C). The supernatant was concentrated to approximately 1 mL using a VivaSpin concentrator (5,000 Da MWCO) and purified by FPLC using a size-exclusion column running with SAU-100 buffer (7 M urea, 20 mM NaOAc, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 5.2). The clean Histone H3 containing fractions were pooled together and dialyzed four times against 4 L of water containing 2 mM BME. The protein solution was lyophilized yielding 28 mg of Histone H3 Cys9 (28 mg/L expression volume) or 38 mg Histone H3 Cys10, respectively. The lyophilized protein stored at −80°C until needed. LC-MS analysis showed 99% 15N incorporation.

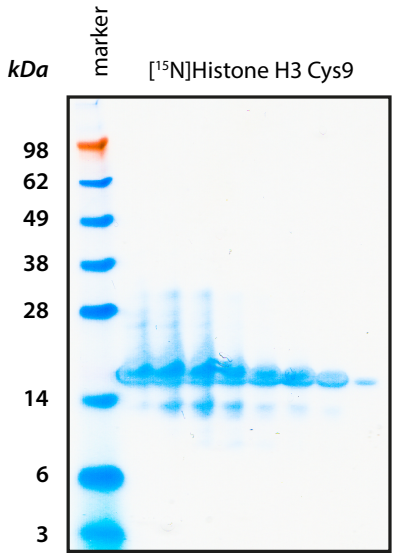


Figure 7.9: Analysis of purified fractions of [^{15}N]Histone H3 Cys9 via SDS-PAGE (MES buffer, 200 V, 40 min, 10% *bis*-TRIS gel) and coomassie staining (marker: SeeBlue Plus2 Protein Standard).

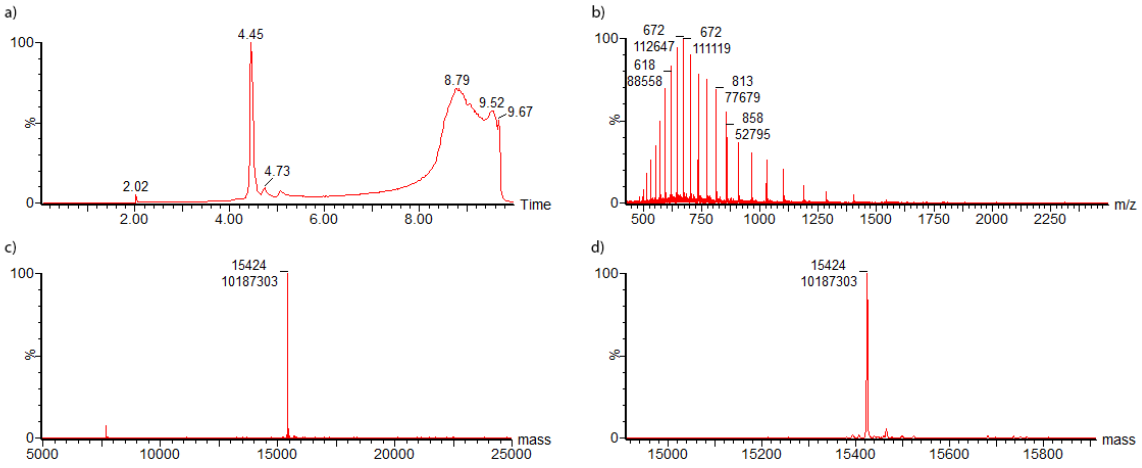


Figure 7.10: LC-MS analysis of [^{15}N]Histone H3 Cys9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15426 g/mol; observed mass: 15424 g/mol.

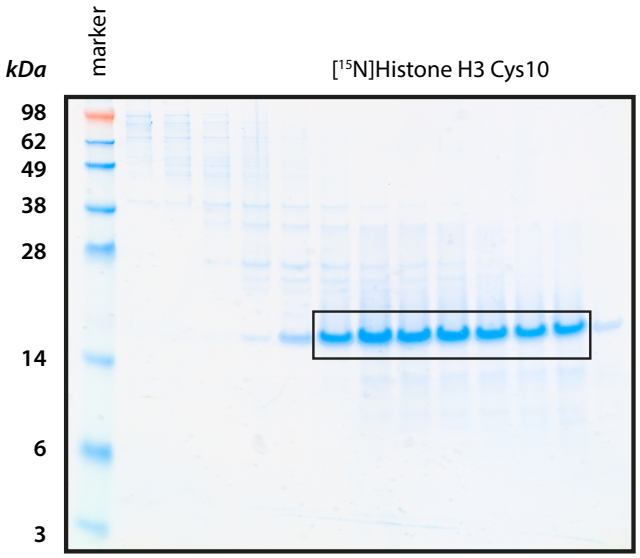


Figure 7.11: Analysis of purified fractions of [^{15}N]Histone H3 Cys10 via SDS-PAGE (MES buffer, 200 V, 40 min, 10% *bis*-TRIS gel) and coomassie staining (marker: SeeBlue Plus2 Protein Standard).

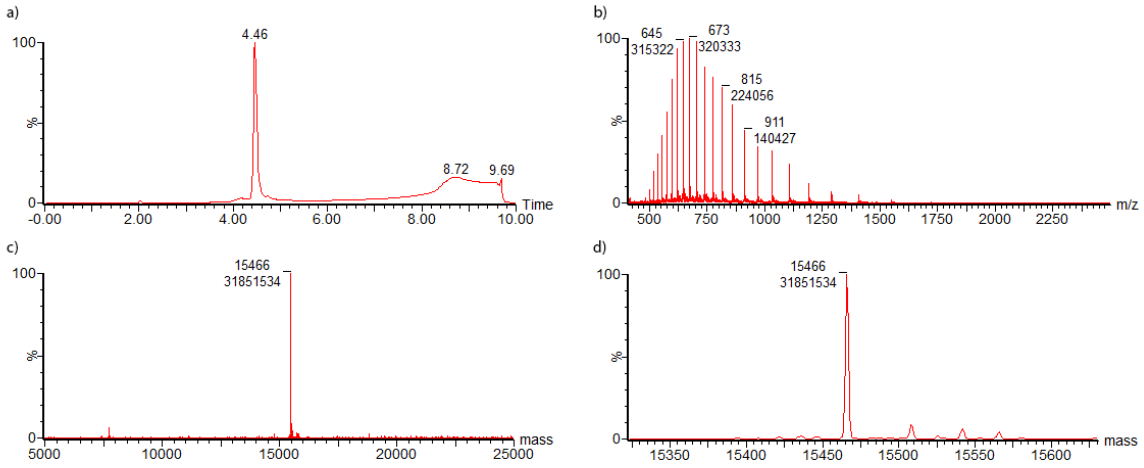


Figure 7.12: LC-MS analysis of [^{15}N]Histone H3 Cys10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15468 g/mol; observed mass: 15466 g/mol.

7.4.4 [^{13}C-^{15}N]Histone H3 Cys10

Media Preparation

M9 medium was prepared by dissolving the following in 1 L of distilled water:

6.0 g	Na_2HPO_4
3.0 g	KH_2PO_4
0.5 g	NaCl
1.0 g	[^{15}N]NH_4Cl
1.0 mL	1.0 M $MgSO_4$
1.0 mL	100 mM $CaCl_2$
1.0 mL	Microsolution 1
0.1 mL	Microsolution 2

Microsolution 1

Microsolution 1 was prepared by dissolving the following in 10 mL of distilled water:

150 mg	$CaCl_2$
300 mg	Na_2EDTA
250 mg	$FeCl_3$

Microsolution 2

Microsolution 2 was prepared by dissolving the following in 10 mL of distilled water:

240 mg	$CuSO_4 \cdot 5H_2O$
180 mg	$MnSO_4 \cdot H_2O$
27 mg	$ZnSO_4 \cdot 7H_2O$
27 mg	$CoCl_2$

Microsolution 3

Microsolution 3 was prepared by dissolving the following in 10 mL of distilled water followed by sterile filtration:

100 mg	Thiamine
100 mg	Biotin

The pH was adjusted to 7.4 and the medium was sterilized and cooled to room temperature before [^{13}C]glucose (2.0 g), Microsolution 3 (1 mL) and antibiotics (final concentration: chloramphenicol: 34 µg/mL, carbenicillin: 100 µg/mL) were added.

Protein Expression and Purification

Protein Expression

BL21(DE3)pLysS competent *E. coli* (Agilent) were transformed with the plasmid following the manufacturer's instructions and grown on agar plates (supplemented with 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol) over night. Single colonies were selected and transferred to 5 mL of LB media (containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol) each. The starter cultures were incubated at 37°C until an optical density value (OD_{600}) of 0.5 was observed. The cultures were pelleted by centrifugation (4,000 g, 10 min, 4°C) and resuspended in 10 mL M9 medium. One starter culture was prepared for each 500 mL of expression media. Two 2.5 L Erlenmeyer flasks charged with 500 mL of M9 media each were inoculated with the starter cultures. The bacterial cultures were incubated at 37°C with vigorous shaking (250 rpm) an optical density value (OD_{600}) of 0.7 was observed. Protein expression was induced by addition of 1 mM IPTG and the culture was further incubated at 37°C for 3 h. The cells were harvested by centrifugation (8,000 rpm, 20 min, 4°C). The cell pellet was suspended in 25 mL of lysis buffer (50 mM TRIS base, 100 mM NaCl, 1 mM EDTA, 5 mM BME, pH 7.5) containing a tablet of cOmplete™ mini EDTA-free protease inhibitor cocktail. The sample was frozen in liquid nitrogen and stored at −80°C until purification.

Protein Purification

Protein purification was conducted as previously described for Histone H3 Cys10 to yield a total of 2.5 mg of [^{13}C-^{15}N]Histone H3 Cys10 as a white solid. The molecular mass was determined by LC-MS (M_w = 16117 g/mol) and corresponds to 97% heavy atom incorporation.

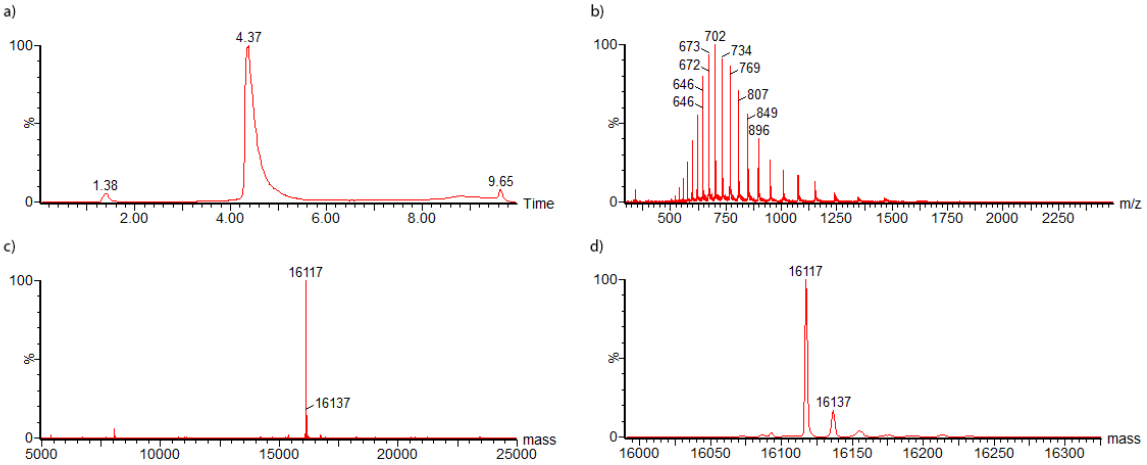


Figure 7.13: LC-MS analysis of [^{13}C-^{15}N]Histone H3 Cys10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 16140 g/mol (full incorporation of ^{13}C and ^{15}N); observed mass: 16117 g/mol (97% heavy atom incorporation).

7.4.5 Annexin V Cys316

Protein Sequence

```
1          11         21         31         41         51
AQVLRGTVTD FPGFDERADA ETLRKAMKGL GTDEESILTL LTSRSNAQRQ EISAAFKTLF
61         71         81         91         101        111
GRDLLDDLKS ELTGKFEKLI VALMKPSRLY DAYELKHALK GAGTNEKVLT EIIASRTPEE
121        131        141        151        161        171
LRAIKQVYEE EYGSSLEDDV VGDTSGYYQR MLVVLLQANR DPDAGIDEAQ VEQDAQALFQ
181        191        201        211        221        231
AGELKWGTDE EKFITIFGTR SVSHLRKVFD KYMTISGFQI EETIDRETSG NLEQLLLAVV
241        251        261        271        281        291
KSIRSIPAYL AETLYYAMKG AGTDDHTLIR VMVSRSEIDL FNIRKEFRKN FATSLYSMIK
301        311
GDTSGDYKKA LLLLCGEDD
```

Formula: $C_{1585}H_{2535}N_{423}O_{502}S_8$

Calculated molecular weight: 35805.6 g/mol

Extinction coefficient $\varepsilon = 23380$ M^{-1} cm^{-1}

Plasmid Amplification

The plasmid was transformed into Invitrogen One Shot TOP10 chemically competent *E. coli* according to the manufacturer's transformation protocol and grown on agar plates (supplemented with ampicillin) over night. A single colony was selected and transferred to 5 mL of LB media containing 100 µg/mL ampicillin. The culture was incubated at 37°C for 14 h and the cell pellet was harvested. Plasmid purification was conducted using a QIAprep Spin Miniprep kit and following the manufacturer's instructions.

Protein Expression and Purification

Protein Expression

T7 Express Competent *E. coli* (New England BioLabs) were transformed with the plasmid

following the manufacturer's instructions and grown on agar plates (supplemented with 100 µg/mL ampicillin) over night. Single colonies were selected and transferred to 10 mL of LB media (containing 100 µg/mL ampicillin) each. The starter culture was incubated at 37°C for 15 h. One starter culture was prepared for every 500 mL of expression media. Four 2.5 L Erlenmeyer flasks charged with 500 mL of LB media supplemented with carbenicillin (at final concentration of 100 µg/mL) were inoculated with the starter cultures. The bacterial cultures were incubated at 37°C with vigorous shaking (250 rpm) until an optical density value (OD_{600}) of 0.45 – 0.55 was observed. Protein expression was induced by addition of 1 mM IPTG and the culture was further incubated at 37°C for 3 h. The cells were harvested by centrifugation (8,000 rpm, 10 min, 4°C). The cell pellet was suspended in 60 mL of TBS buffer (50 mM TRIS base, 150 mM NaCl, pH 8.0) and re-pelleted by centrifugation (10,000 rpm, 10 min, 4°C), then re-suspended in 60 mL of $CaCl_2$ buffer (50 mM TRIS base, 10 mM $CaCl_2$, pH 7.2) containing a tablet of cOmplete™ mini EDTA-free protease inhibitor cocktail. The samples were frozen in liquid nitrogen and stored at −80 °C until purification.

Protein Purification

The samples were thawed on ice and DNase I was added. The cell suspension was then sonicated on ice (5 cycles, 50% amplitude, 30 sec on, 60 sec off). The samples were centrifuged (18,000 rpm, 20 min, 4°C) and the supernatant was discarded. The cell pellet was re-suspended in 40 mL of EDTA buffer (50 mM TRIS base, 20 mM EDTA, pH 7.2) and the cell debris was removed by centrifugation (18,000 rpm, 20 min, 4°C). The annexin V containing supernatant was filtered through a 0.45 µm syringe filter and dialysed against 3 × 4 L of TRIS buffer (20 mM TRIS base, pH 7.8). The dialysed protein solution was purified by FPLC using an anion exchange column (10 mL) and running a gradient from TRIS buffer (20 mM TRIS base, pH 7.8) to NaCl buffer (20 mM TRIS base, 500 mM NaCl, pH 7.8). The purified protein was dialyzed against 3 × 4 L of NaP_i buffer (50 mM NaP_i, pH 8.0). The protein solution was concentrated using a VivaSpin 20 centrifugal concentrator (MWCO 10,000 Da, 4,000 rpm, 10 min, 4°C) to yield 10 mL of a protein solution. The protein concentration was determined to be 5.40 mg/mL using an A_{280} spectrophotometer. The expression yield was determined to be 27 mg/L. The protein solution was divided into aliquots, frozen in liquid nitrogen and

stored at −80°C until needed.

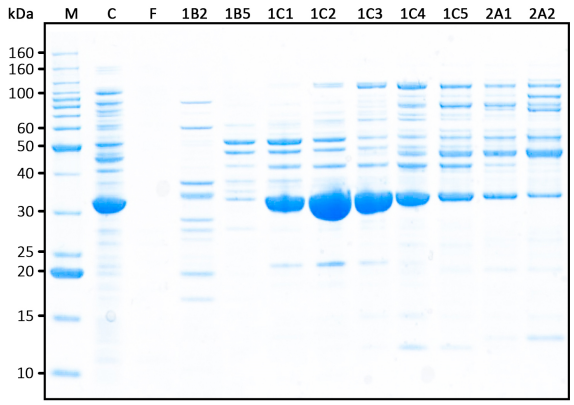


Figure 7.14: Analysis of purified fractions of Annexin V via SDS-PAGE and coomassie staining (M = BenchMark™ Unstained Protein Ladder, C = crude, F = flow through).

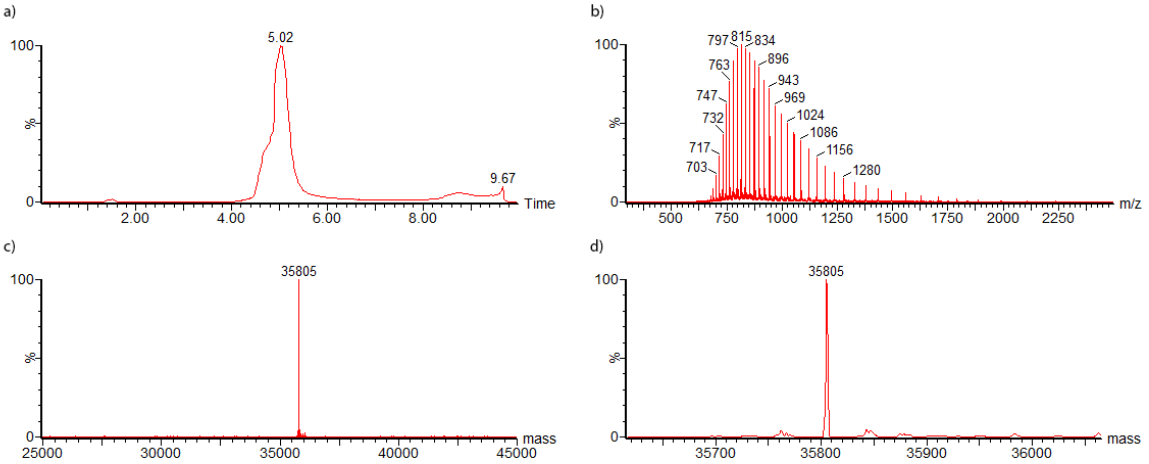


Figure 7.15: LC-MS analysis of Annexin V Cys316: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 35806 g/mol; observed mass: 35805 g/mol.

7.4.6 Npβ Cys61

Protein Sequence

```
1           11          21          31          41          51
MFSSHHHHHH SSGLVPRGSH IDVGKLRQLY AAGERDFSIV DLRGAVLENI NLSGAILHGA
61          71          81          91          101         111
CLDEANLQQA NLSRADLSGA TLNGADLRGA NLSKADLSDA ILDNAILEGA ILDEAVLNQA
121         131         141         151         161         171
NLKAANLEQA ILSHANIREA DLSEANLEAA DLSGADLAIA DLHQANLHQA ALERANLTGA
181         191         201
NLEDANLEGT ILEGGNNNLA T
```

Formula: $C_{898}H_{1456}N_{276}O_{303}S_2$

Calculated molecular weight: 21031.2 g/mol

Extinction coefficient $\varepsilon = 1490\ \text{M}^{-1}\ \text{cm}^{-1}$

Plasmid Amplification

The plasmid was transformed into Invitrogen One Shot TOP10 chemically competent *E. coli* according to the manufacturer's transformation protocol and grown on agar plates (supplemented with kanamycin) over night. A single colony was selected and transferred to 10 mL of LB media containing 50 µg/mL kanamycin. The culture was incubated at 37°C for 14 h and the cell pellet was harvested. Plasmid purification was conducted using a QIAprep Spin Miniprep kit and following the manufacturer's instructions.

Protein Expression and Purification

Protein Expression

T7 Express Competent *E. coli* (New England BioLabs) were transformed with the plasmid following the manufacturer's instructions and grown on agar plates (supplemented with 50 µg/mL kanamycin) over night. Single colonies were selected and transferred to 10 mL of LB media (containing 50 µg/mL kanamycin) each. The starter culture was incubated at

37°C for 15 h. One starter culture was prepared for every 900 mL of expression media. Two 2.5 L Erlenmeyer flasks charged with 900 mL of LB media supplemented with kanamycin (at final concentration of 30 µg/mL) were inoculated with the starter cultures. The bacterial cultures were incubated at 37°C with vigorous shaking (250 rpm) until an optical density value (OD_{600}) of 0.7 – 0.8 was observed. Protein expression was induced by addition of 1 mM IPTG and the culture was further incubated at 37°C for 4 h. The cells were harvested by centrifugation (8,000 rpm, 15 min, 4°C). The cell pellet was suspended in 25 mL of lysis buffer (20 mM imidazole, 20 mM TRIS base, 500 mM NaCl, 2 mM DTT, 5% glycerol, pH 7.8) containing half a tablet of cOmplete™ mini EDTA-free protease inhibitor cocktail. The sample was frozen in liquid nitrogen and stored at −80°C until purification.

Protein Purification

The samples were thawed on ice and DNase I (2 mg) was added. The cell suspension was then sonicated on ice (10 cycles, 60% amplitude, 15 sec on, 60 sec off). The samples were centrifuged (18,000 rpm, 10 min, 4°C) and the supernatant was filtered through a 0.45 µm syringe filter and purified via FPLC. The sample was loaded onto a 5 mL GE HisTrap™ HP column at a flow rate of 1 mL/min. The column was washed with 20 CV of buffer A (40 mM imidazole, 20 mM TRIS base, 500 mM NaCl, 2 mM DTT, 5% glycerol, pH 7.8) and the protein was eluted running a linear gradient of 20 CV to 100% buffer B (500 mM imidazole, 20 mM TRIS base, 500 mM NaCl, 2 mM DTT, 5% glycerol, pH 7.8). The fractions were analyzed by SDS-PAGE and clean fractions containing Npβ were pooled together yielding 60 mL of protein solution. The purified protein solution was dialysed against 3 × 4 L of NaP_i buffer (50 mM NaP_i, pH 7.8) and concentrated using four VivaSpin 20 centrifugal concentrators (MWCO 10,000 Da, 4,000 rpm, 15 min, 4°C) to yield 10.3 mL of a protein solution. The protein concentration was determined to be 5.48 mg/mL using an A_{280} spectrophotometer. The expression yield was determined to be 31 mg/L. The protein solution was divided into aliquots, frozen in liquid nitrogen and stored at −80°C until needed.

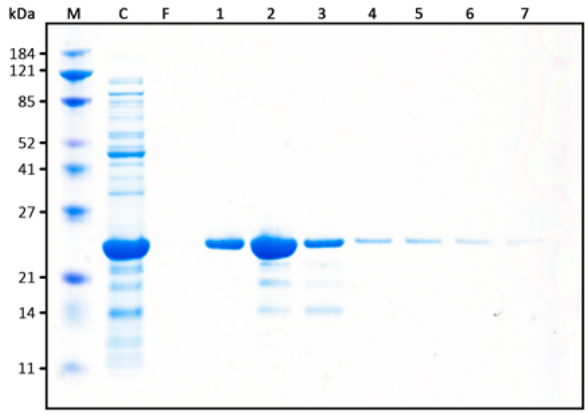


Figure 7.16: Analysis of purified fractions of Npβ Cys61 via SDS-PAGE and coomassie staining (M = BenchMark™ Unstained Protein Ladder, C = crude, F = flow through).

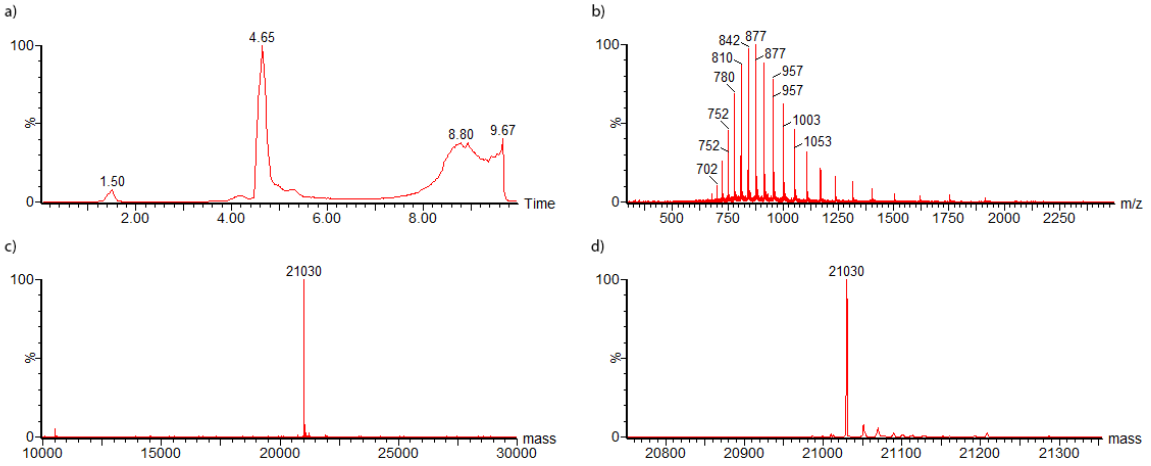


Figure 7.17: LC-MS analysis of Npβ Cys61: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 21031 g/mol; observed mass: 21030 g/mol.

7.4.7 pre-SUMO1 Cys51

Protein Sequence

```
1          11         21         31         41         51
ADQEAKPSTE DLGDKKEGEY IKLKVIGQDS SEIHFKVKMT THLKKLKESY CQRQGVPMNS
61         71         81         91         101
LRFLFEGQRI ADNHTPKELG MEEEDVIEVY QEQTGGHSTV LEHHHHHH
```

Formula: $C_{543}H_{851}N_{157}O_{173}S_4$

Calculated molecular weight: 12474.9 g/mol

Extinction coefficient $\varepsilon = 4470$ M^{-1} cm^{-1}

Plasmid Amplification

A pET28a plasmid encoding the gene for pre-SUMO1 was a gift from Guy Salvesen distributed by Addgene (Addgene plasmid #25101) as an agar stab. Bacteria were streaked and grown on agar plates (supplemented with kanamycin) over night. A single colony was selected and transferred to 5 mL of LB media containing 50 µg/mL kanamycin. The culture was incubated at 37°C overnight and the cell pellet was harvested. Plasmid purification was conducted using a QIAprep Spin Miniprep kit (Qiagen) and following the manufacturer's instructions.

Protein Expression and Purification

Protein Expression

Chemically competent *E. coli* BL21(DE3) pLysS (Agilent) were transformed with pET28a-pre-SUMO1 and plated on agar plates (supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol) according to the manufacturer's instructions. After over-night incubation, single colonies were transferred to 2 × 10 mL of freshly prepared LB media (supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol). The starter cultures were incubated overnight and used to inoculate 2 × 900 mL of LB media (supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol). The cultures were shaken at 37°C (250 rpm) until an optical density value (OD_{600}) of 0.8 – 1.0 was reached. Protein expression was induced by addition of 1 mM IPTG and the cultures were shaken for an additional 3 hours. The cells

were harvested by centrifugation (8,500 rpm, 30 min, 4°C). The cell pellets were resuspended in 15 mL of lysis buffer (50 mM NaH_2PO_4, 10 mM imidazole, 300 mM NaCl, 10 mM BME, pH 8.0, one tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche)), flash frozen in liquid nitrogen and stored at −80°C until purification.

Protein Purification

The samples were thawed and 1 mg of DNase I was added. The suspension was sonicated (10 cycles, 40% amplitude, 15 sec on, 60 sec off) and the cell debris was removed by centrifugation (20,000 rpm, 30 min, 4°C). The solution was loaded onto two HisTrap HP 5 mL columns (GE Healthcare) operated in series, the columns were washed with 20 CV of binding buffer (50 mM NaH_2PO_4, 20 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0) and eluted with a 20 CV gradient 0-100% elution buffer (50 mM NaH_2PO_4, 500 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0). The fractions were analysed by SDS-PAGE, clean fractions were pooled together and dialysed twice against phosphate buffer supplemented with BME (50 mM NaH_2PO_4, 10 mM BME, pH 8.0) and once phosphate buffer (50 mM NaH_2PO_4, pH 8.0). The protein concentration was determined using an A_{280} spectrophotometer. The expression yield was determined to be 81.3 mg (45 mg/L expression volume). The protein solution was divided into aliquots, frozen in liquid nitrogen and stored at −80°C until needed.

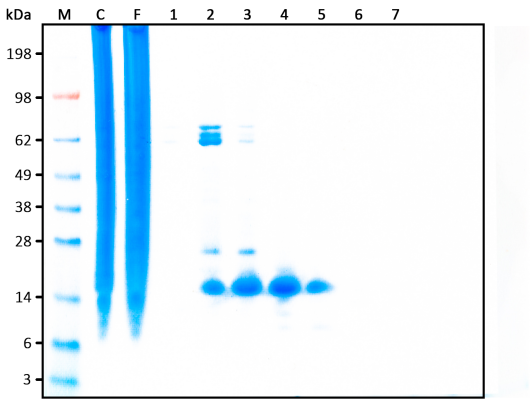


Figure 7.18: Analysis of purified fractions of pre-SUMO1 Cys51 via SDS-PAGE (MES buffer, 200 V, 40 min, 10% *bis*-TRIS gel) and coomassie staining (M = SeeBlue Plus2 Protein Standard, C = crude lysate, F = flow through).

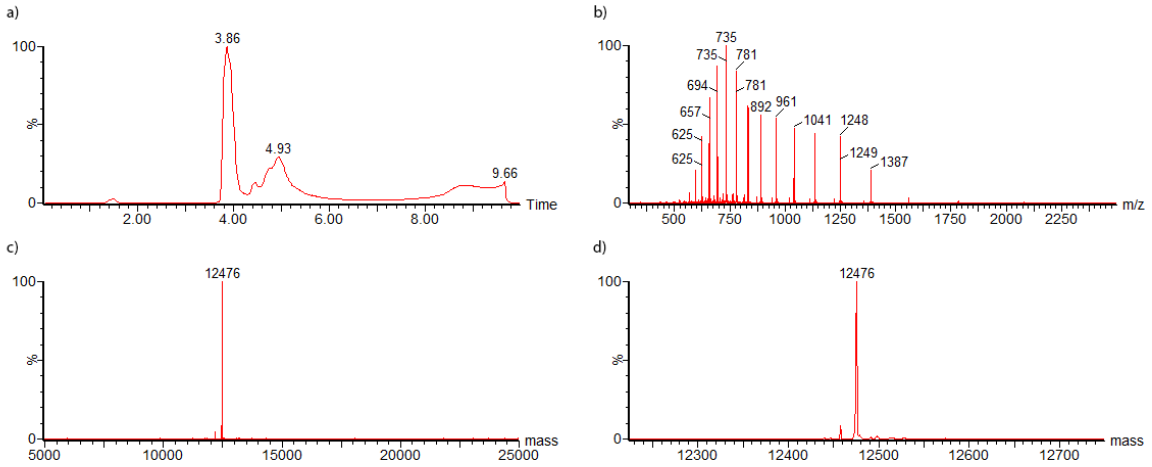


Figure 7.19: LC-MS analysis of preSUMO1-Cys1: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 12475 g/mol; observed mass: 12476 g/mol.

7.4.8 pre-SUMO1 Cys51 Cys97

Protein Sequence

```
1          11         21         31         41         51
ADQEAKPSTE DLGDKKEGEY IKLKVIGQDS SEIHFKVKMT THLKKLKESY CQRQGVPMNS
61         71         81         91         101
LRFLFEGQRI ADNHTPKELG MEEEDVIEVY QEQTGGCSTV LEHHHHHH
```

Formula: $C_{540}H_{849}N_{155}O_{173}S_5$

Calculated molecular weight: 12440.9 g/mol

Extinction coefficient $\varepsilon = 4470$ M^{-1} cm^{-1}

Site-Directed Mutagenesis

Site-directed mutagenesis was conducted on a template pET28a plasmid encoding the gene for pre-SUMO1 (received as a gift from Guy Salvesen and distributed by Addgene; Addgene plasmid #25101). Site-directed mutagenesis was performed using the QuikChange II XL site directed mutagenesis kit (Agilent, Catalogue No. 200521-5) following the manufacturer's instructions. The following primers (purchased from Merck Life Science UK) were used:

Forward Primer	`5'-TCAGGAACAAACGGGAGGTTGTTCAACAGTTCTCGAGCAC-3'`
Reverse Primer	`5'-GTGCTCGAGAACTGTTGAACAACCTCCCGTTTGTTCCTGA-3'`

The reaction mixture was transformed into One Shot™ MAX Efficiency™ DH5α-T1R chemically competent cells (Invitrogen, Catalogue No. 12297016) and grown on agar plates (supplemented with kanamycin at 50 µg/mL) over night. Single colonies were picked and grown in 5 mL of LB media shaking (225 rpm) at 37°C for 18 hours. The cells were pelleted by centrifugation (8,000 rpm, 15 min) and the plasmid was extracted and purified using a QIAprep Spin Miniprep kit (Qiagen, Catalogue No. 27104) according to the manufacturer's instructions.

The presence of the desired mutation was confirmed by Sanger sequencing (Source BioScience, Cambridge, UK) using T7 forward and reverse primers.

5'-ATGGCTGACCAGGAGGCAAAACCATCAACTGAGGACTTGGGGGATAAGAAGGAAGGTGAATATATTAAA
CTCAAAGTCATTGGACAGGATAGCAGTGAGATTCACTTCAAAGTGAAAATGACAACACATCTCAAGAAACTC
AAAGAATCATACTGTCAAAGACAGGGTGTTCCAATGAATTCACTCAGGTTTCTCTTTGAGGGTCAGAGAATT
GCTGATAATCATACTCCAAAAGAACTGGGAATGGAGGAAGAAGATGTGATTGAAGTTTATCAGGAACAAACG
GGAGGTTGTTCAACAGTTCTCGAGCACCACCACCACCACCAC-3'

Protein Expression and Purification

Protein Expression

Chemically competent *E. coli* BL21(DE3) pLysS (Agilent) were transformed with the mutated pET28a-pre-SUMO1 C51 C97 plasmid and plated on agar plates (supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol) according to the manufacturer's instructions. After over-night incubation, single colonies were transferred to 2 × 10 mL of freshly prepared LB media (supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol). The starter cultures were incubated overnight and used to inoculate 2 × 500 mL of LB media (supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol). The cultures were shaken at 37°C (250 rpm) until an optical density value (OD_{600}) of 0.8 – 1.0 was reached. Protein expression was induced by addition of 1 mM IPTG and the cultures were shaken for an additional 3 hours. The cells were harvested by centrifugation (8,500 rpm, 30 min, 4°C). The cell pellets were resuspended in 15 mL of lysis buffer (50 mM NaH_2PO_4, 20 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0, one tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche)), flash frozen in liquid nitrogen and stored at −80°C until purification.

Protein Purification

1 mg of DNase I was added to the suspended cell pellet. The suspension was sonicated (10 cycles, 40% amplitude, 15 sec on, 60 sec off) and the cell debris was removed by centrifugation (20,000 rpm, 30 min, 4°C). The solution was loaded onto a HisTrap HP 5 mL column (GE Healthcare) and the column was washed with 10 CV of binding buffer (50 mM NaH_2PO_4, 20 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0) and eluted with elution buffer (50 mM NaH_2PO_4, 500 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0). The fractions were analysed by SDS-PAGE, clean fractions were pooled together and dialysed twice against

milli-Q water supplemented with BME (500 µL/L). The protein concentration was determined using an A_{280} spectrophotometer. The expression yield was determined to be 25.8 mg. The protein solution was lyophilized and stored at −80°C until needed.

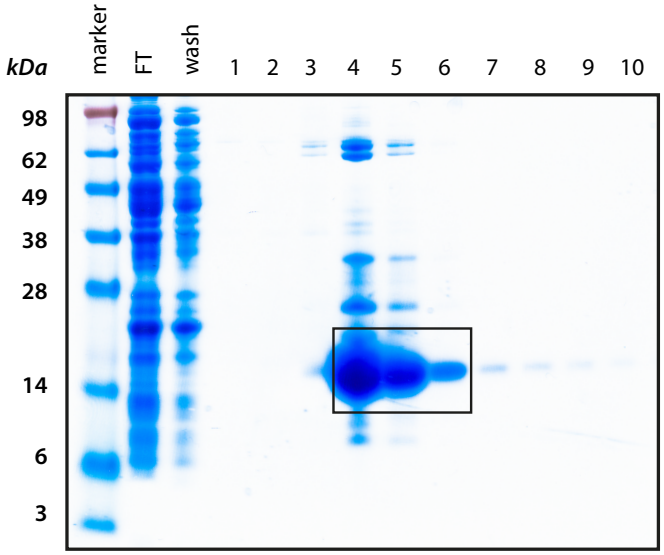


Figure 7.20: Analysis of purified fractions of pre-SUMO1 Cys51 via SDS-PAGE (MES buffer, 200 V, 50 min, 4-12% *bis*-TRIS gel) and coomassie staining (marker: SeeBlue Plus2 Protein Standard, FT = flow through).

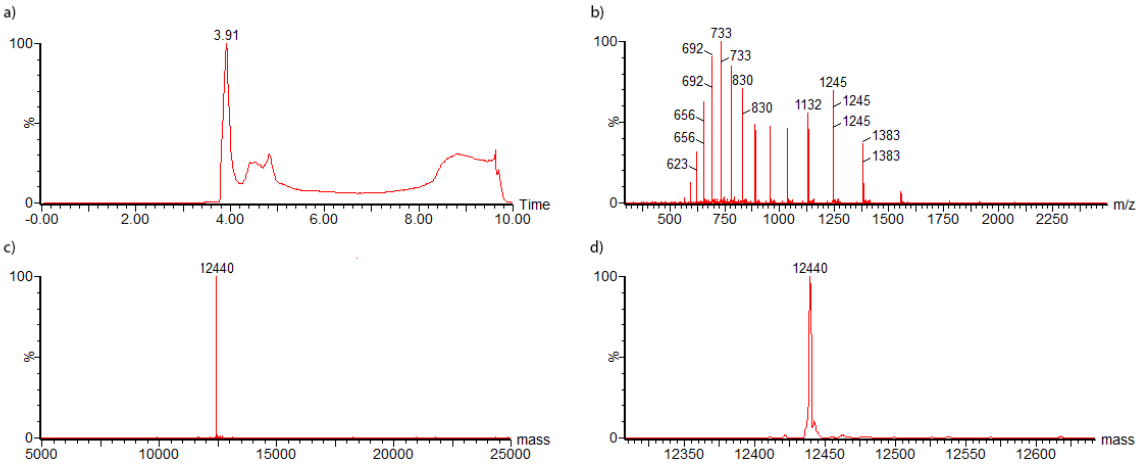


Figure 7.21: LC-MS analysis of preSUMO1-Cys1: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 12441 g/mol; observed mass: 12440 g/mol.

7.4.9 mCherry Cys131

Protein Sequence

```
1           11          21          31          41          51
MRGSHHHHHH GSMVSKGEED NMAIIKEFMR FKVHMEGSVN GHEFEIEGEG EGRPYEGTQT
61          71          81          91          101         111
AKLKVTKGGP LPFAWDILSP QFMYGSKAYV KHPADIPDYL KLSFPEGFKW ERVMNFEDGG
121         131         141         151         161         171
VVTVTQDSSL QDGEFIYKVK LRGTNFPCDG PVMQKKTMGW EASSERMYPE DGALKGEIKQ
181         191         201         211         221         231
RLKLKDGGHY DAEVKTTYKA KKPVQLPGAY NVNIKLDITS HNEDYTIVEQ YERAEGRHST
241
GGMDELYK
```

Formula: $C_{1250}H_{1915}N_{341}O_{377}S_{12}$

Calculated molecular weight: 28136.7 g/mol

Extinction coefficient $\varepsilon = 34380$ M^{-1} cm^{-1}

Protein Expression and Purification

Protein Expression

E. coli SG13009 (Qiagen) carrying the plasmid pQE-9-mCherryS131C was kindly donated by Bradley Olsen and distributed by Addgene (Addgene plasmid #48732). Bacteria were streaked on agar plates (supplemented with 50 µg/mL kanamycin and 100 µg/mL ampicillin) and grown overnight. Single colonies were transferred to 2 × 10 mL freshly prepared terrific broth (TB media) supplemented with 50 µg/mL kanamycin and 100 µg/mL ampicillin and incubated at 37°C (250 rpm) for 16 hours. 2 × 900 mL TB media (containing 50 µg/mL kanamycin, 100 µg/mL ampicillin) were inoculated with the previously prepared starter cultures. The cultures were shaken at 37°C (200 rpm) until an optical density value (OD_{600}) of 1.0 – 1.2 was reached. Protein expression was induced by addition of 1 mM ITPG and the cultures were shaken for another 5 hours. The cells were harvested by centrifugation (9,000 rpm, 20 min, 4°C), suspended in 12.5 mL of lysis buffer each (50 mM NaH_2PO_4, 10 mM imidazole,

300 mM NaCl, 10 mM BME, pH 8.0, one tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche)) and frozen in liquid nitrogen.

Protein Purification

The samples were thawed and 1 mg of DNase I was added. The suspension was sonicated (5 cycles, 40% amplitude, 30 sec on, 60 sec off) and the cell debris was removed by centrifugation (20,000 rpm, 20 min, 4°C). The solution was loaded onto two HisTrap HP 5 mL columns (GE Healthcare) operated in series, the columns were washed with 20 CV of binding buffer (50 mM NaH_2PO_4, 20 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0) and eluted with a 20 CV gradient 0-100% elution buffer (50 mM NaH_2PO_4, 500 mM imidazole, 500 mM NaCl, 10 mM BME, pH 8.0). The fractions were analysed by SDS-PAGE, clean fractions were pooled together and dialysed three times against phosphate buffer supplemented with BME (50 mM NaH_2PO_4, 10 mM BME, pH 8.0). The protein concentration was determined using an A_{280} spectrophotometer. The expression yield was determined to be 294 mg (163 mg/L expression volume). The protein solution was divided into aliquots, frozen in liquid nitrogen and stored at −80°C until needed.

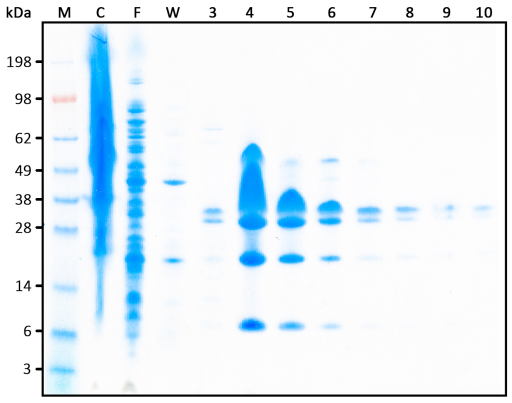


Figure 7.22: Analysis of purified fractions of mCherry Cys131 via SDS-PAGE (MES buffer, 200 V, 40 min, 10% *bis*-TRIS gel) and coomassie staining (M = SeeBlue Plus2 Protein Standard, C = crude lysate, F = flow through, W = wash).

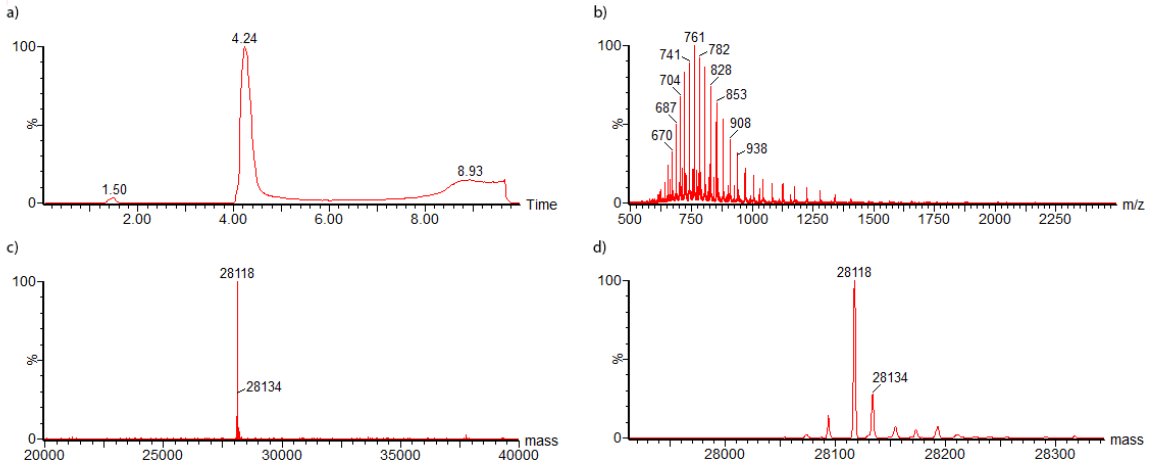


Figure 7.23: LC-MS analysis of mCherry Cys131: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 28137 g/mol (pre maturation), 28117 g/mol (matured mCherry Cys131); observed masses: 28134 g/mol (pre maturation), 28118 g/mol (matured mCherry Cys131).

7.4.10 PstS Cys197

Protein Sequence

```
1          11         21         31         41         51
MEASLTGAGA TFPAPVYAKW ADTYQKETGN KVNYQGIGSS GGVKQIIANT VDFGASDAPL
61         71         81         91         101        111
SDEKLAQEGL FQFPTVIGGV VLAVNIPGLK SGELVLDGKT LGDIYLGKIK KWDDEAIAKL
121        131        141        151        161        171
NPGLKLPSQN IAVVRRADGS GTSFVFTSYL AKVNEEWKNN VGTGSTVKWP IGLGGKGNDG
181        191        201        211        221        231
IAAFVQRLPG AIGYVEYCYA KQNNLAYTKL ISADGKPVSP TEENFANAAK GADWSKTFAQ
241        251        261        271        281        291
DLTNQKGEDA WPITSTTFIL IHKDQKKPEQ GTEVLKFFDW AYKTGAKQAN DLDYASLPDS
301        311        321
VVEQVRAAWK TNIKDSSGKP LY
```

Formula: $C_{1563}H_{2426}N_{406}O_{476}S_2$

Calculated molecular weight: 35585.0 g/mol

Extinction coefficient $\varepsilon = 61880$ M^{-1} cm^{-1}

Plasmid Amplification

E. coli DH5α carrying the plasmid pET22b-PstSA197C was kindly donated by Martin Webb and distributed by Addgene (Addgene plasmid #78198) as an agar stab. Bacteria were streaked on agar plates (supplemented with 100 µg/mL ampicillin) and grown overnight. A single colony was selected and transferred to 5 mL of LB media containing 100 µg/mL ampicillin. The culture was incubated at 37°C overnight and the cell pellet was harvested. Plasmid purification was conducted using a QIAprep Spin Miniprep kit (Qiagen) and following the manufacturer's instructions.

Protein Expression and Purification

Protein Expression

Chemically competent *E. coli* BL21(DE3) (Agilent) were transformed with pET22b-PstS-A197C and plated on agar plates (supplemented with 100 µg/mL ampicillin) according to the manufacturer's instructions. Single colonies were transferred to 2 × 10 mL freshly prepared LB media supplemented with 100 µg/mL ampicillin and incubated at 37°C (250 rpm) for 16 hours. 2 × 900 mL LB media (containing 100 µg/mL ampicillin) were inoculated with the previously prepared starter cultures. The cultures were shaken at 37°C (200 rpm) until an optical density value (OD_{600}) of 0.6 – 0.8 was reached. Protein expression was induced by addition of 1 mM ITPG and the cultures were shaken for another 4 hours. The cells were harvested by centrifugation (9,000 rpm, 20 min, 4°C), suspended in 12.5 mL of lysis buffer each (10 mM TRIS base, 1 mM DTT, pH 8.0, one tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche)), frozen in liquid nitrogen and stored at −80°C until purification.

Protein Purification

The samples were thawed and 1 mg of DNase I was added. The suspension was sonicated (10 cycles, 40% amplitude, 15 sec on, 60 sec off) and the cell debris was removed by centrifugation (30,000 rpm, 30 min, 4°C). The solution was loaded onto a HiTrap Q HP column (5 mL) (GE Healthcare) and washed with 10 CV of binding buffer (10 mM TRIS base, 1 mM DTT, pH 8.0) and eluted with a 20 CV gradient 0−100% elution buffer (10 mM TRIS base, 200 mM NaCl, 1 mM DTT, pH 8.0). The fractions were analysed by SDS-PAGE and clean fractions were pooled together. The protein concentration was determined using an A_{280} spectrophotometer. The expression yield was determined to be 13.7 mg (15.2 mg/L expression volume). The protein solution was divided into aliquots, frozen in liquid nitrogen and stored at −80°C until needed.

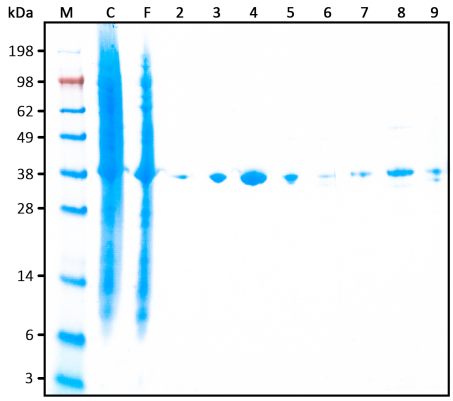


Figure 7.24: Analysis of purified fractions of PstS Cys197 via SDS-PAGE (MES buffer, 200 V, 40 min, 10% *bis*-TRIS gel) and coomassie staining (M = SeeBlue Plus2 Protein Standard, C = crude lysate, F = flow through, W = wash).

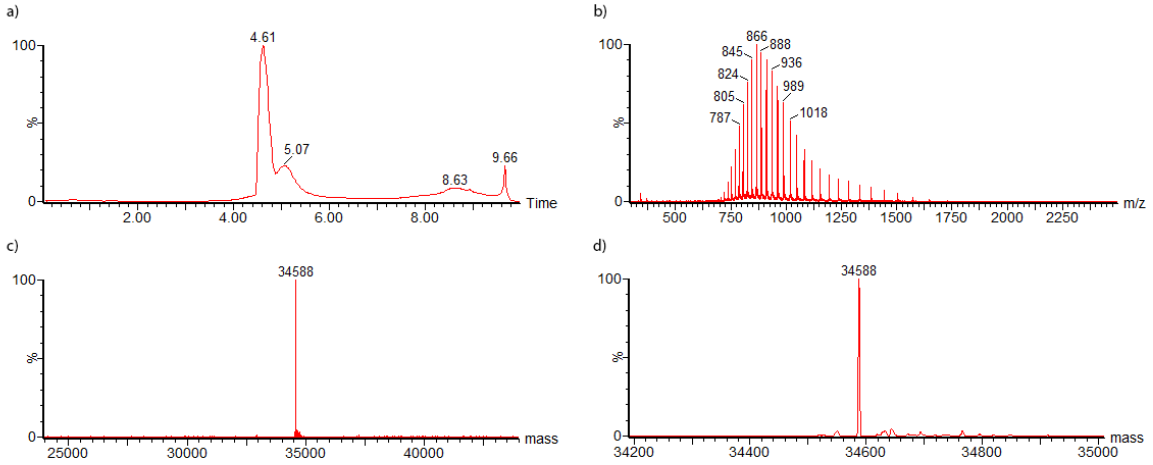
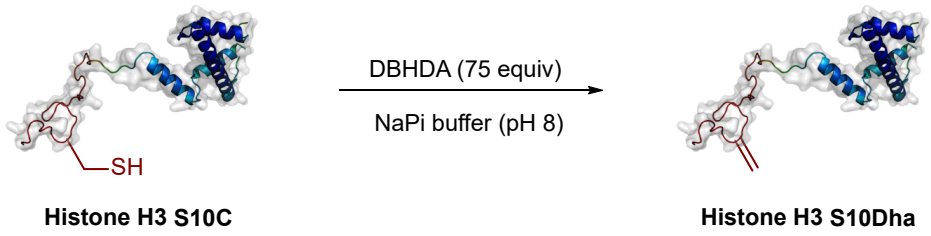


Figure 7.25: LC-MS analysis of PstS Cys197: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 34585 g/mol; observed mass: 34588 g/mol.

7.5 Preparation of Protein Substrates

7.5.1 Histone H3 Dha10



Histone H3 Cys10 (11.4 mg, 0.75 µmol, 1.00 equiv) and DTT (30 mg, 194 µmol, 260 equiv) were dissolved in 500 µL of NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) by sonication. The solution was incubated at room temperature for 1 h, before it was desalted by passing through a GE MiniTrap G-25 desalting column pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) according to the manufacturer's instructions. To the solution was added a solution of DBHDA (18.0 mg, 59.8 µmol, 75 equiv) in DMF (100 µL).

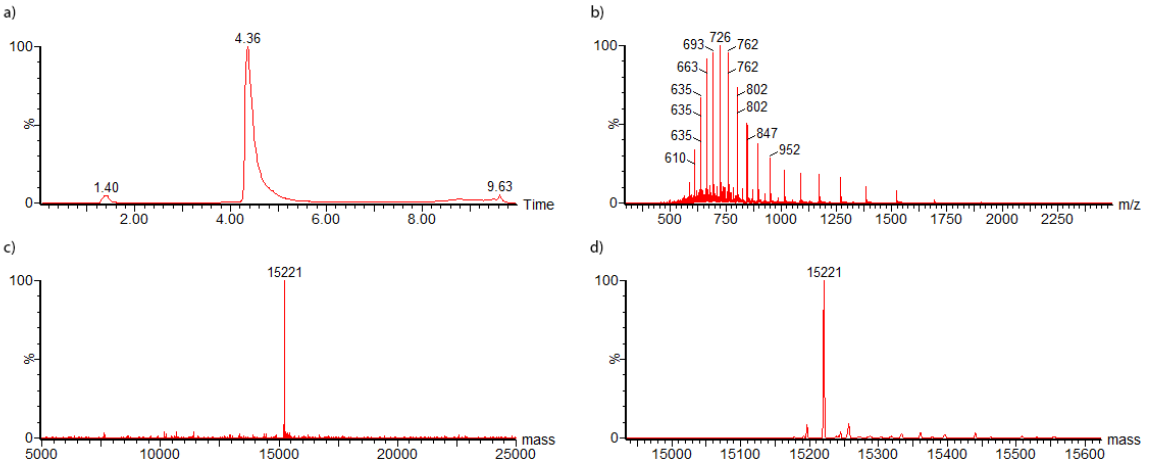


Figure 7.26: LC-MS analysis of Histone H3 Dha10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15221 g/mol, observed mass: 15221 g/mol.

The mixture was vortexed and gently shaken at room temperature for 75 min until no starting material could be detected anymore by LC-MS. The temperature was raised to 37°C and the reaction mixture was shaken for an additional 2.5 h. The protein was desalted by passing through a GE MiniTrap G-25 column pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) to yield 2.00 mL of a protein solution with a concentration of 4.58 mg/mL as determined

by A_{280} spectrophotometry (9.16 mg, 0.60 µmol, 80%). The sample was aliquoted, frozen in liquid nitrogen and stored at −80°C.

7.5.2 [^{15}N]Histone H3 Dha10

[^{15}N]Histone H3 Dha10 was synthesized following the procedure for Histone H3 Dha10. Full conversion was observed via LC-MS analysis.

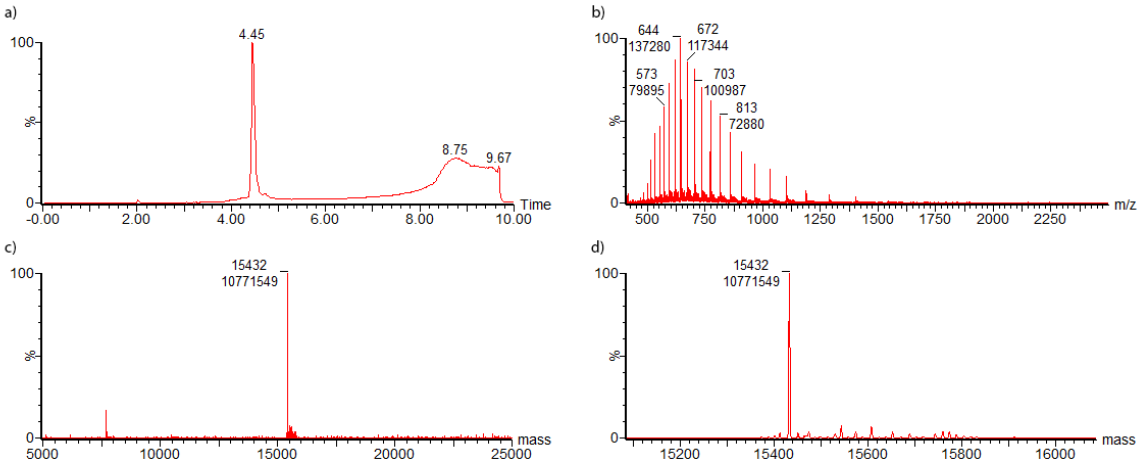
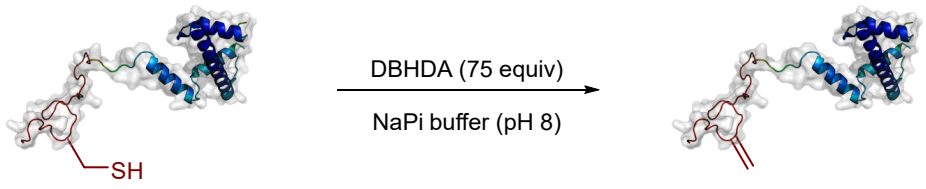


Figure 7.27: LC-MS analysis of [^{15}N]Histone H3 Dha10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15434 g/mol; observed mass: 15432 g/mol.

7.5.3 [^{13}C-^{15}N]Histone H3 Dha10



[^{13}C-^{15}N]Histone H3 S10C **[^{13}C-^{15}N]Histone H3 S10Dha**

[^{13}C-^{15}N]Histone H3 Cys10 (2.5 mg, 164 nmol, 1.00 equiv) was dissolved in 330 µL of NMR buffer (50 mM NaP_i, 3 M Gdn·HCl, 5% D_2O). After completed NMR measurements, the sample was incubated with DTT (2.5 mg, 16.4 µmol, 100 equiv) for 1 h.

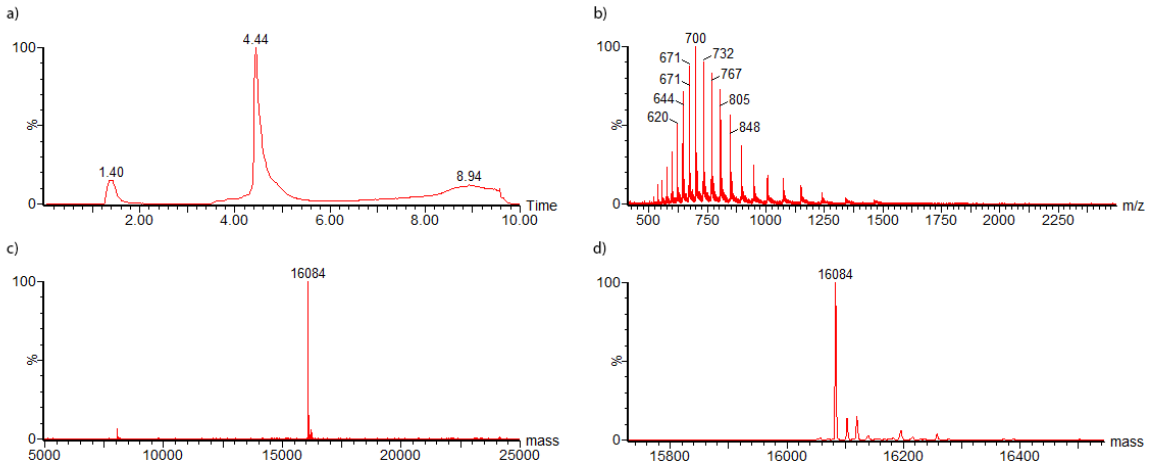
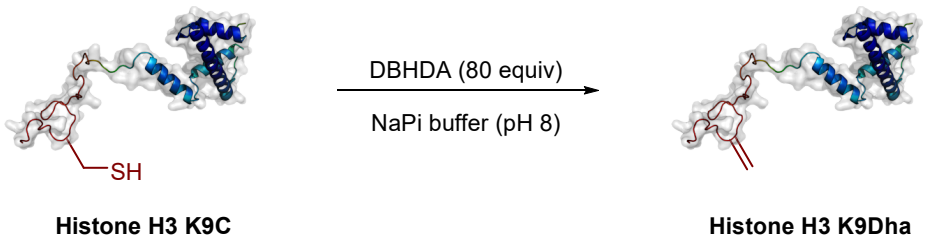


Figure 7.28: LC-MS analysis of [^{13}C-^{15}N]Histone H3 Dha10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 16083 g/mol; observed mass: 16084 g/mol.

The sample was desalted by passing through a GE MidiTrap G-25 desalting column pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) according to the manufacturer's instructions. To the solution was added a solution of DBHDA (3.7 mg, 12.3 µmol, 75 equiv) in DMF (80 µL), the mixture was vortexed and shaken (500 rpm) at room temperature for 45 min. The temperature was increased to 37°C and shaking was continued for 2 h. The reaction mixture was purified by passing through two GE MidiTrap G-25 desalting columns pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0). The sample was then dialysed against milli-Q water (3 × 4 L) using a dialysis cassette (MWCO 3,500 Da) and lyophilized. The lyophilized powder was redissolved in deuterated NaP_i buffer (1600 µL) and the concentration was determined at 1.20 mg/mL by A_{280} spectrophotometry (1.92 mg, 126 nmol, 77% recovery).

7.5.4 Histone H3 Dha9



Histone H3 Cys9 (15.0 mg, 0.99 µmol, 1.00 equiv) and DTT (37.5 mg, 243 µmol, 250 equiv) were dissolved in 1000 µL of NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) by repeated vortexing and sonication.

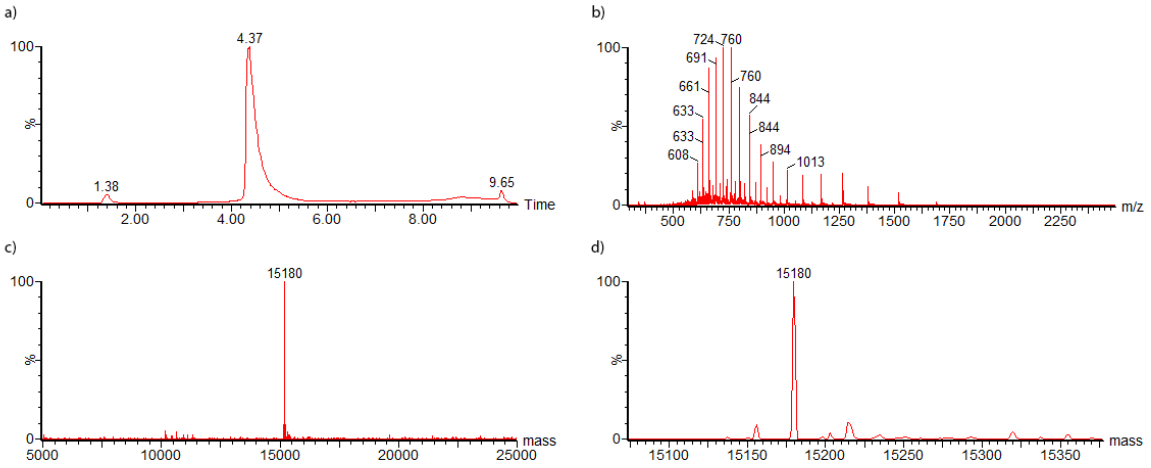


Figure 7.29: LC-MS analysis of Histone H3 Dha9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15180 g/mol; observed mass: 15180 g/mol.

The solution was incubated at room temperature for 1 h before being desalted using GE MiniTrap G-25 desalting columns preconditioned with NaP_i buffer (100 µM, 3 M Gdn·HCl, pH 8.0) according to the manufacturer's specifications. A solution of DBHDA (23.8 mg, 78.9 µmol, 80 equiv) in 90 µL of DMF was added and the mixture was briefly vortexed and shaken at room temperature for 2 h. The temperature was raised to 37°C and shaking was continued for an additional 2.5 h until completion of the reaction was observed by LC-MS. The solution was purified using GE MiniTrap G-25 columns preconditioned with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) to yield 4000 µL of a protein solution with a concentration of

2.54 mg/mL as determined by A_{280} spectrophotometry (10.16 mg, 0.67 µmol, 68% recovery). The sample was aliquoted, frozen in liquid nitrogen and stored at −80°C.

7.5.5 [^{15}N]Histone H3 Dha9

[^{15}N]Histone H3 Dha9 was synthesized following the procedure for Histone H3 Dha9. Full conversion was observed via LC-MS analysis.

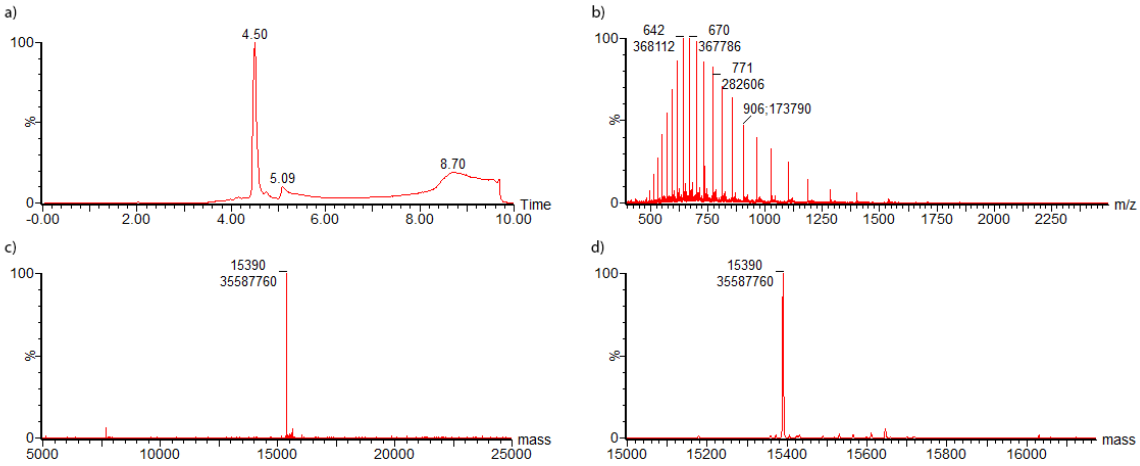
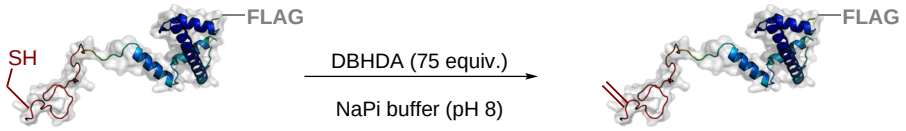


Figure 7.30: LC-MS analysis of [^{15}N]Histone H3 Dha9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15392 g/mol; observed mass: 15390 g/mol.

7.5.6 FLAG-Histone H3 Dha4



FLAG-Histone H3 Cys4 **FLAG-Histone H3 Dha4**

FLAG-Histone H3 Dha4 mutants was kindly provided by Adeline Poh. FLAG-Histone H3 Dha4 was prepared following a previously published procedure.[17]

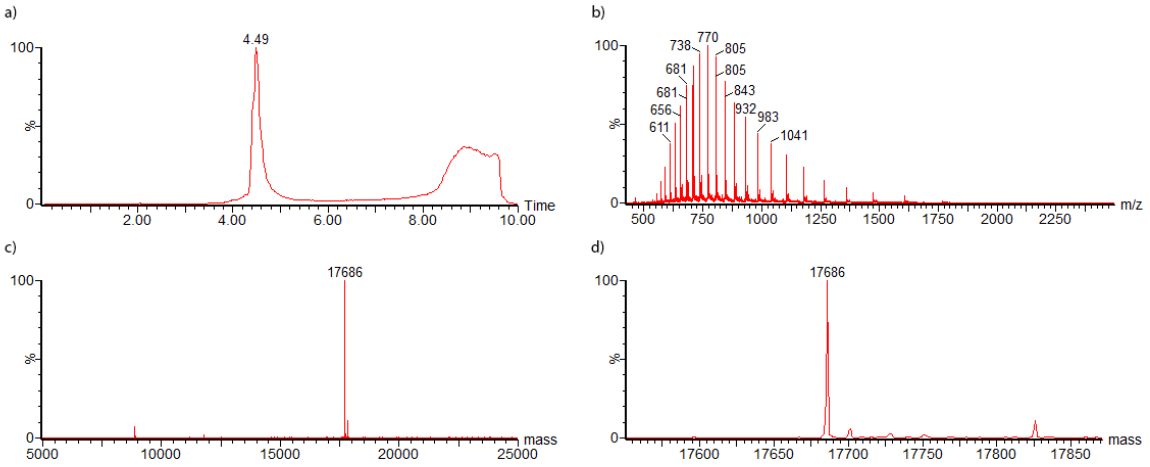


Figure 7.31: LC-MS analysis of Histone H4 Dha16: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 11177 g/mol; observed mass: 11176 g/mol.

7.5.7 Histone H4 Dha16

Histone H4 Dha16 mutants were kindly provided by Brian Josephson and Oluwatobi Arisa. Histone H4 Dha16 was prepared following a previously published procedure.[18]

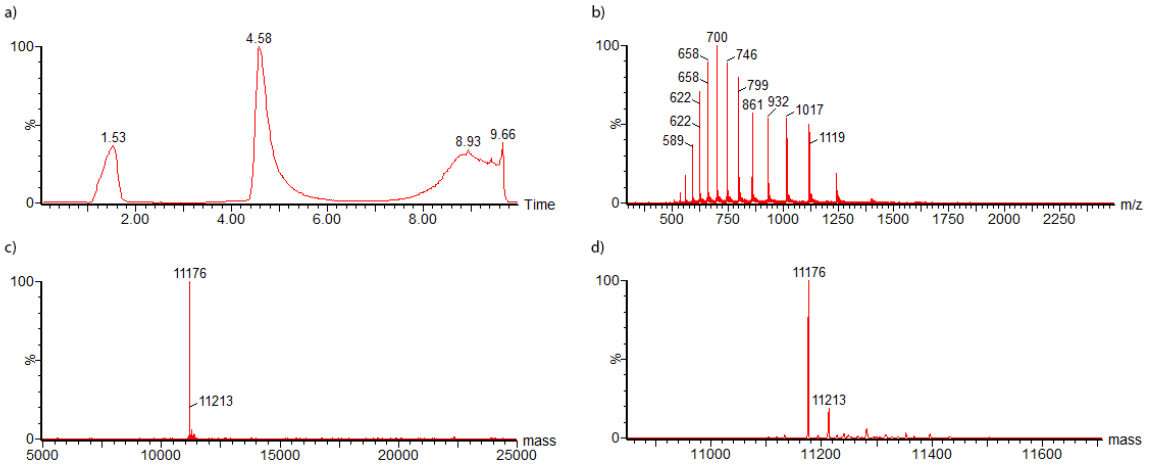
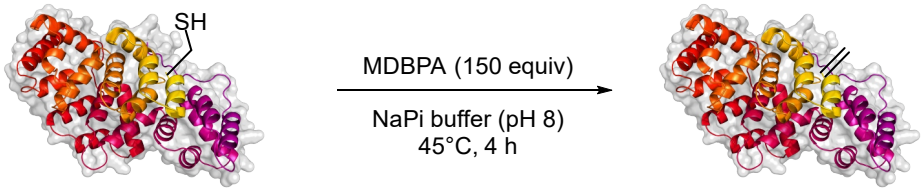


Figure 7.32: LC-MS analysis of Histone H4 Dha16: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 11177 g/mol; observed mass: 11176 g/mol.

7.5.8 Annexin V Dha316



To 450 µL of a solution of Annexin V (2.00 mg/mL, 25.1 nmol, 1.00 equiv) in NaP_i buffer (50 mM, pH 7.8) was added DTT (0.19 mg, 1.26 µmol, 50 equiv). The mixture was incubated at room temperature for 20 min and desalted by passing through a GE MiniTrap G-25 column pre-equilibrated with NaP_i buffer (100 mM, pH 8.0) according to the manufacturer's instructions.

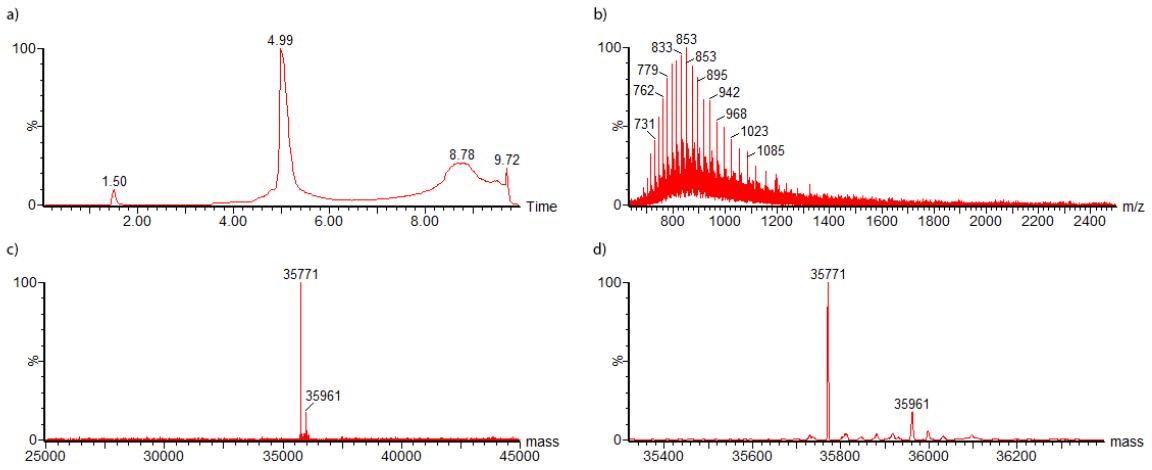
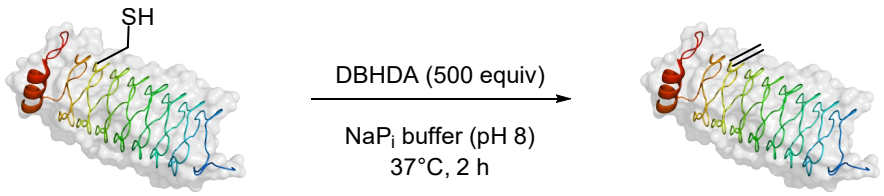


Figure 7.33: LC-MS analysis of Annexin V Dha316: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 34772 g/mol; observed mass: 34771 g/mol.

A solution of MDBPA (0.69 mg, 2.51 µmol, 100 equiv) in 45 µL of DMSO was added and the mixture was shaken at 45°C for 3.5 h. Another 50 equiv of MDBPA in DMSO were added and the mixture was shaken for another 25 min until complete conversion was indicated by LC-MS. The mixture was passed through a GE MiniTrap G-25 column pre-equilibrated with NaP_i buffer (100 mM, pH 7.0) and yielding 1.5 mL of a solution of Annexin V Dha316 with a concentration of 0.60 mg/mL as determined by A_{280} spectrophotometry (900 µg, 25.1 nmol,

quantitative recovery). The product contained 15% double alkylated product. The sample was aliquoted, frozen in liquid nitrogen and stored at −80°C.

7.5.9 Npβ Dha61



To a solution of Npβ Cys61 (2.74 mg, 130 nmol, 1.00 equiv) in 500 µL of NaP_i buffer (50 mM, pH 7.8) was added DTT (1.00 mg, 6.51 µmol, 50 equiv). The mixture was vortexed and incubated at room temperature for 30 min before it was purified using a GE PD MiniTrap G-25 which was previously equilibrated with NaP_i buffer (100 mM, pH 8.0) according to the manufacturer's instructions. A solution of DBHDA (19.7 mg, 65.1 µmol, 500 equiv) in 160 µL of DMF was added to the purified protein and the reaction mixture was shaken at 37°C and 500 rpm for 2 h. The protein was purified using a GE PD MiniTrap G-25 which was previously equilibrated with NaP_i buffer (100 mM, pH 7.0) to yield 2000 µL of protein solution with a concentration of 0.92 mg/mL as determined by A_{280} spectrophotometry (1.84 mg, 87.6 nmol, 67% recovery).

The sample was aliquoted, frozen in liquid nitrogen and stored at −80°C.

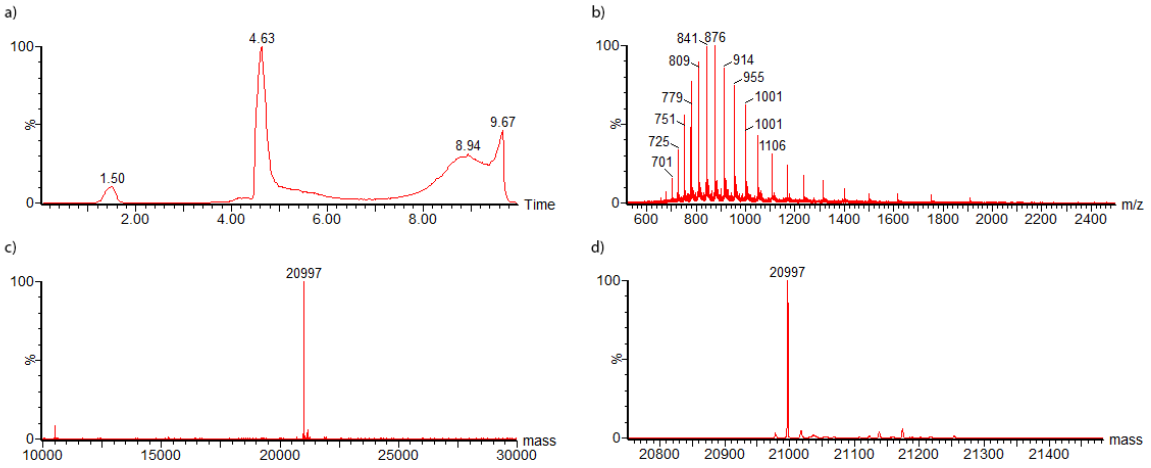


Figure 7.34: LC-MS analysis of Npβ Dha61: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 20997 g/mol; observed mass: 20997 g/mol.

7.5.10 PanC Dha44 and Dha47

PanC Dha mutants were kindly provided by Dr. Patrick Isenegger. PanC Dha44 and PanC Dha47 were prepared following a previously published procedure.[19]

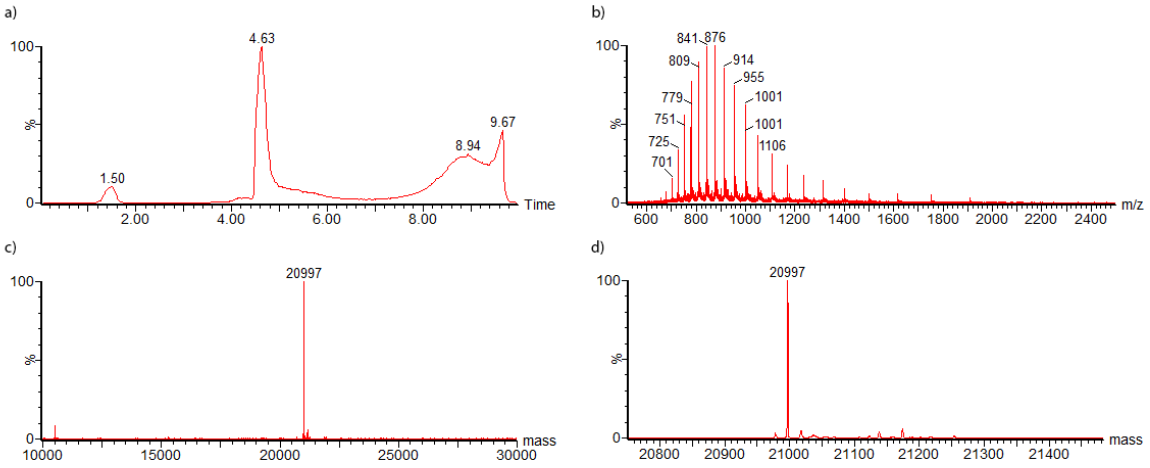
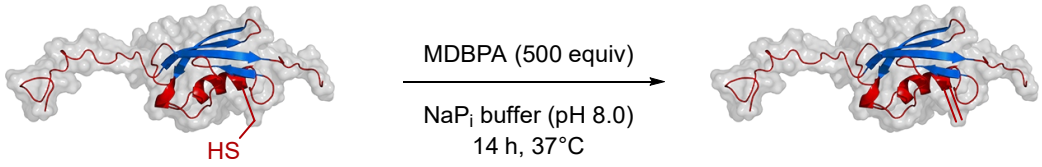


Figure 7.35: LC-MS analysis of PanC Dha44: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 33998 g/mol; observed mass: 33998 g/mol.

7.5.11 pre-SUMO1 Dha51



A solution of pre-SUMO1 Cys51 (4000 µL, 2.30 mg/mL, 737 nmol, 1.00 equiv) was incubated with DTT (5.7 mg, 36.9 µmol, 50.0 equiv) for 30 min at room temperature. The protein was purified using GE PD MidiTraps G-25 which were previously equilibrated with NaP_i buffer (100 mM, pH 8.0) according to the manufacturer's instructions. To the purified protein solution was added a solution of MDBPA (57.4 µL, 369 µmol, 500 equiv) in DMSO (142.6 µL), the mixture was vortexed and shaken at 37°C and 600 rpm for 14 h. The solution was concentrated to 2500 µL using a VivaSpin 20 concentrator (MWCO 3,000 Da). 2000 µL of this solution were purified using GE PD MidiTrap G-25 which were previously equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) yielding 3000 µL of a solution of pre-SUMO1 Dha51 (2.01 mg, 199 nmol, 27% recovery).

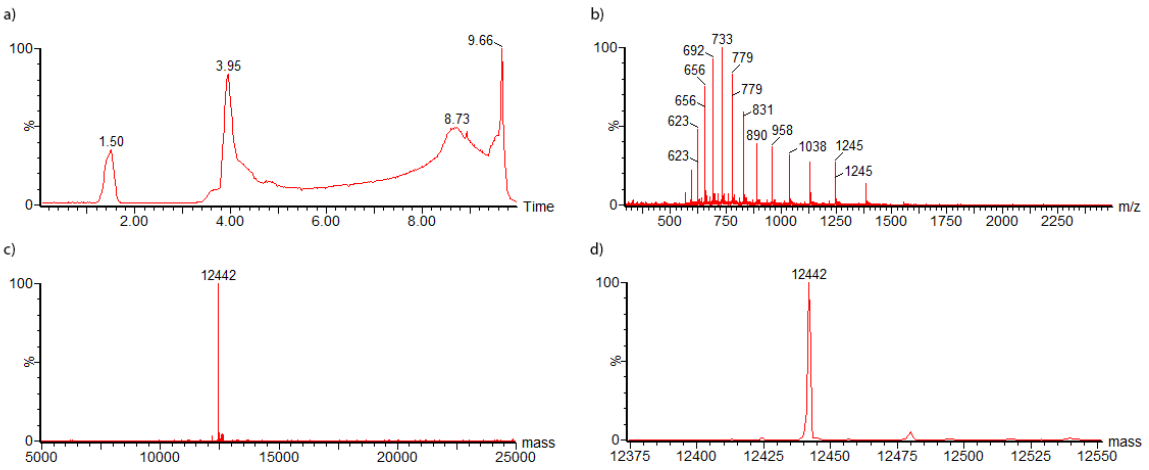
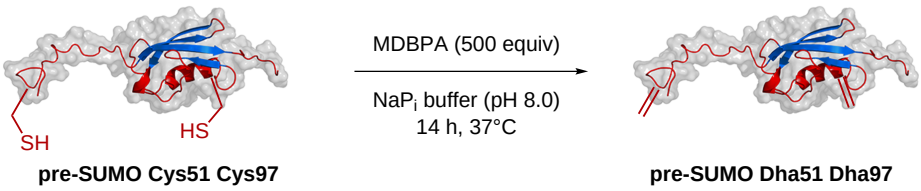


Figure 7.36: LC-MS analysis of pre-SUMO1 Dha51: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 12441 g/mol; observed mass: 12442 g/mol.

7.5.12 pre-SUMO1 Dha51 Dha97



Solid preSUMO1 Cys51 Cys97 (1.0 mg, 80.4 nmol, 1.00 equiv) was dissolved in sodium phosphate buffer (500 µL, 50 mmM, pH 8.0). The solution was cooled to 4°C before Mukaiyama's reagent (20 µL of a stock solution of 6.4 mg in 500 µL of milli-Q H_2O, 1.00 µmol, 12.5 equiv) was added. The solution was mixed and incubated at 4°C. After 10 min, 60 µL of aqueous NaOH (1.0 M) was added and the mixture was incubated for another 60 min. The reaction was quenched with sodium acetate buffer (500 µL, 200 mM, pH 4.0) before being allowed to warm to room temperature. The reaction mixture was purified via gel filtration using a PD MidiTrap G-25 previously equilibrated with TRIS buffer (100 mM, pH 8.0) to yield 1500 µL of a solution of preSUMO1 Dha51 Dha97 (0.39 mg/mL, 585 µg, 47.0 nmol, 59% recovery).

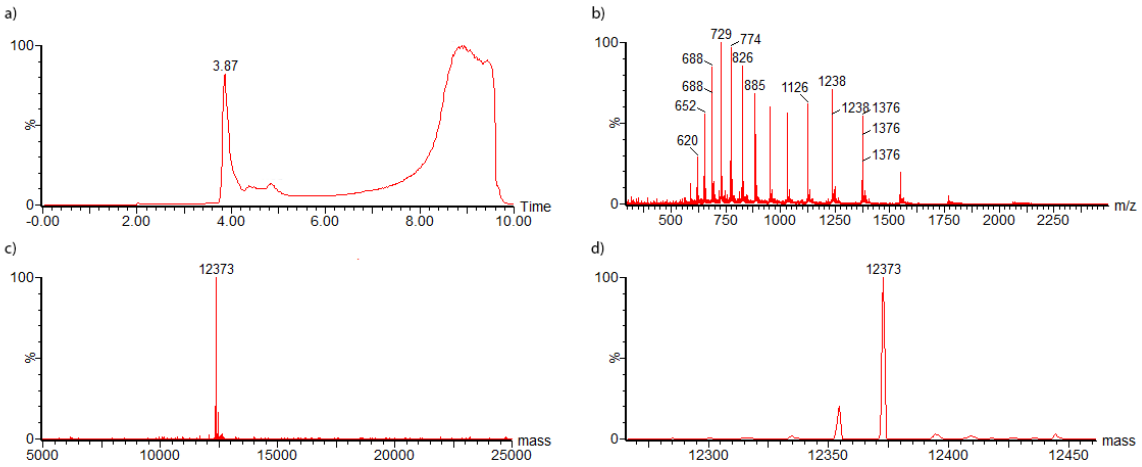
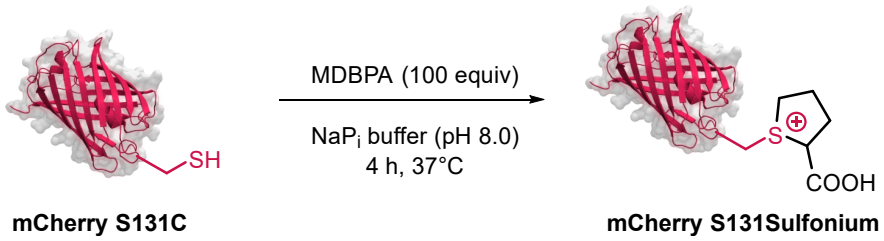


Figure 7.37: LC-MS analysis of pre-SUMO1 Dha51 Dha97: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 12373 g/mol; observed mass: 12373 g/mol.

7.5.13 mCherry Sulfonium131



A solution of mCherry Cys131 (1000 µL, 4.56 mg/mL, 366 nmol) in storage buffer was transferred to a glove box and degassed for 14 h at 4°C. The protein was purified using a GE PD MidiTrap G-25 which was previously equilibrated with NaP$_i$ buffer (100 mM, pH 8.0) according to the manufacturer's instructions. A solution of MDBPA (6.20 µL, 36.6 µmol, 100 equiv) in 50 µL of MeCN was added and the mixture was shaken at 37°C and 500 rpm for 4 h. The mixture was transferred out of the glove box, concentrated to 1000 µL using VivaSpin 500 concentrators (MWCO 5,000 Da) and purified using a GE PD MidiTrap G-25 which was previously equilibrated with NaP$_i$ buffer (100 mM, pH 7.0) according to the manufacturer's instructions to yield 1500 µL of a solution of mCherry Sulfonium131 (2.41 mg/mL, 3.62 mg, 289 nmol, 79% recovery).

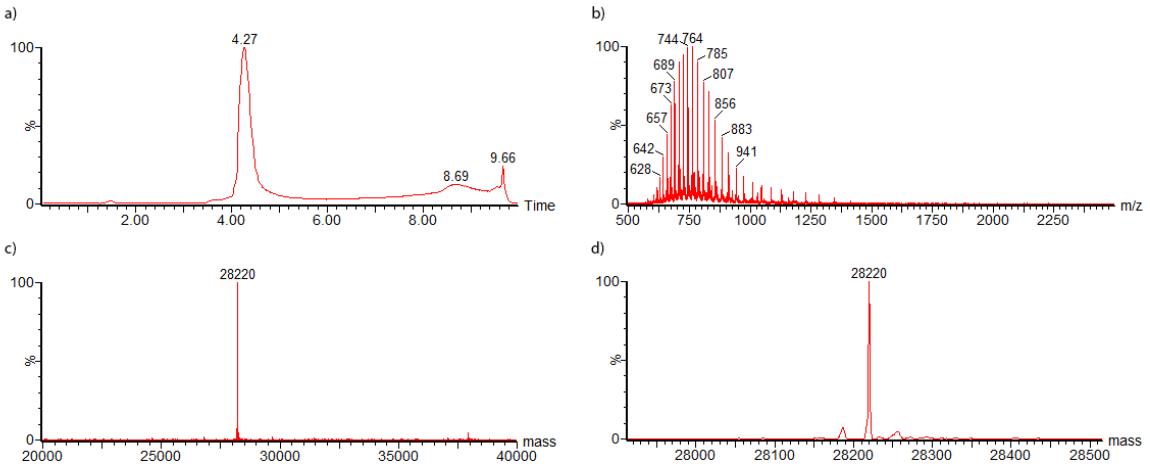
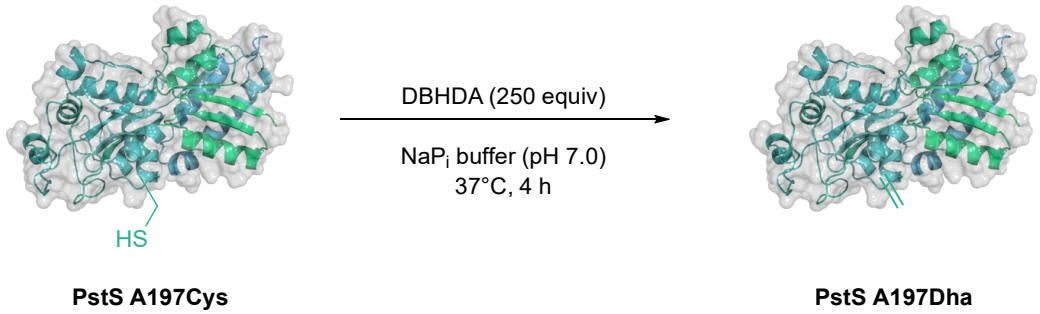


Figure 7.38: LC-MS analysis of mCherry Sulfonium131: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 28215 g/mol; observed mass: 28220 g/mol.

7.5.14 PstS Dha197



A solution of PstS Cys197 (500 µL, 2.07 mg/mL, 1.0 mg, 29.9 nmol) was incubated with DTT (0.23 mg, 1.50 µmol, 50 equiv) for 30 min at room temperature. The protein solution was purified using a GE PD MidiTrap G-25 which was previously equilibrated with NaP_i buffer (100 mM, pH 8.0, 3 M Gdn·HCl) according to the manufacturer's instructions. To the purified protein was added a solution of DBHDA (2.3 mg, 7.48 µmol, 250 equiv) in 50 µL of DMF. The mixture was shaken at 37°C for 4 h (500 rpm) before being purified using a GE PD MidiTrap G-25 which was previously equilibrated with NaP_i buffer (100 mM, pH 7.0, 3 M Gdn·HCl) to yield 1500 µL of a solution of PstS Dha197 with a concentration of 0.44 mg/mL (32% recovery).

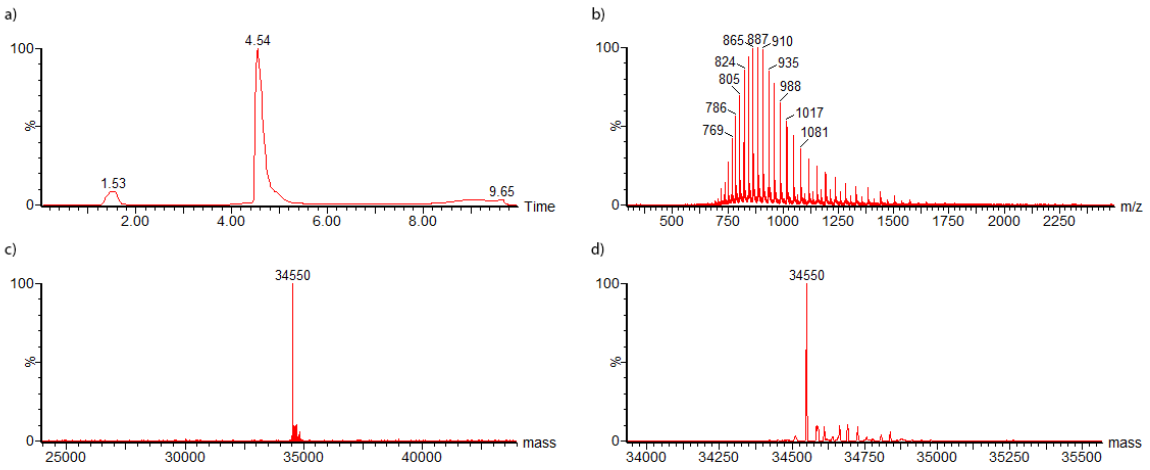


Figure 7.39: LC-MS analysis of PstS Dha197: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 34551 g/mol; observed mass: 34550 g/mol.

7.5.15 AcrA Dha123

AcrA Dha123 mutants was kindly provided by Brian Josephson. AcrA Dha123 was prepared following a previously published procedure.[18]

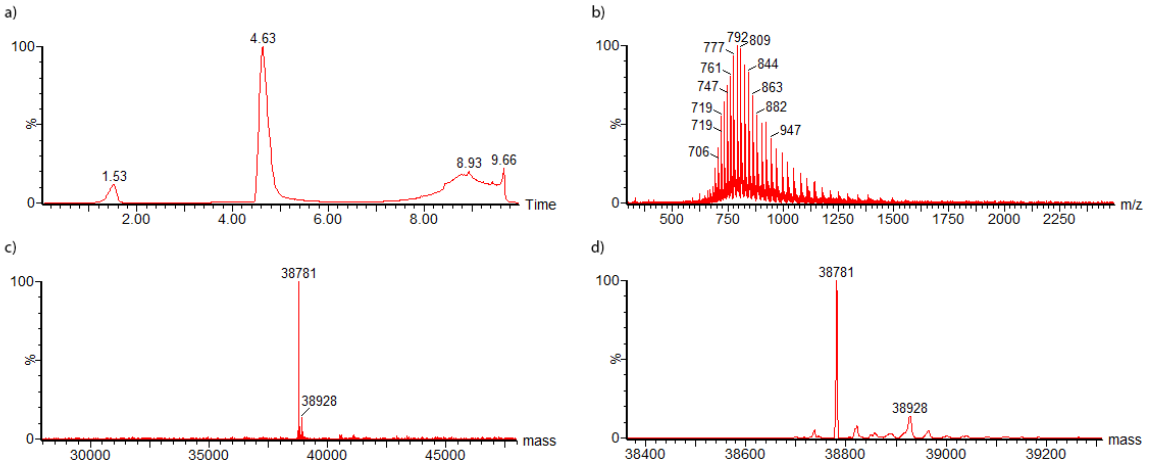


Figure 7.40: LC-MS analysis of AcrA Dha123: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 38783 g/mol; observed mass: 38781 g/mol.

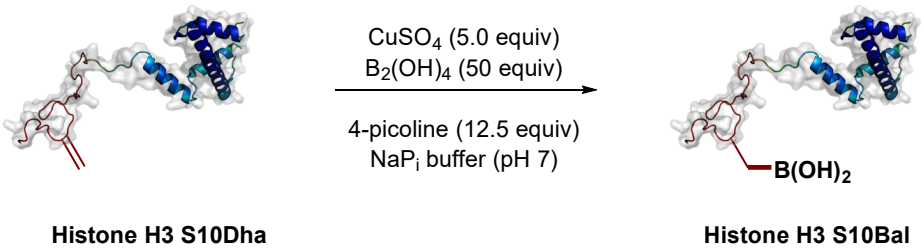
7.6 Protein Borylation & Reactions of Boronoproteins

7.6.1 General Borylation Protocol & Reaction Optimizations

Protein borylation was performed at a protein concentration of approximately 1 mg/mL in sodium phosphate buffer (pH 8.0, 100 mM, 3 M Gdn·HCl) on 20-60 µL scale. Stock solutions of $CuSO_4 \cdot 5H_2O$, 4-picoline and diboron reagent were freshly prepared so that a volume of 1 µL to 5 µL of the stock solution were needed for the addition of the required equivalents of reagent ($CuSO_4$: 5.0 equiv, $B_2(OH)_4$: 50 equiv, 4-picoline: 12.5 equiv). The calculated amount of each stock solution was added and the reaction mixtures were mixed using a vortex mixer. The mixtures were then incubated at room temperature and analysed by LC-MS. Typical sample preparation consisted of dilution of 3 µL of reaction mixture with 100 µL of 0.1% formic acid in milli-Q water.

Screening reactions for reaction optimization were conducted following this general protocol while modifying the variable that the screening was conducted for (ligand/base, buffer, pH, temperature, stoichiometry).

7.6.2 Histone H3 Bal10



Stock solutions of $CuSO_4 \cdot 5H_2O$ (11.8 mg in 1.00 mL of H_2O), 4-picoline (11.5 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (14.1 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of Histone H3 Dha10 (500 µL, 2.87 mg/mL, 94.3 nmol, 1.00 equiv) NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) were added previously prepared stock solutions of 4-picoline (10.0 µL, 1.18 µmol, 12.5 equiv), $CuSO_4$ (10.0 µL, 471 nmol, 5.00 equiv) and $B_2(OH)_4$ (30.0 µL, 4.71 µmol, 50.0 equiv) subsequently.

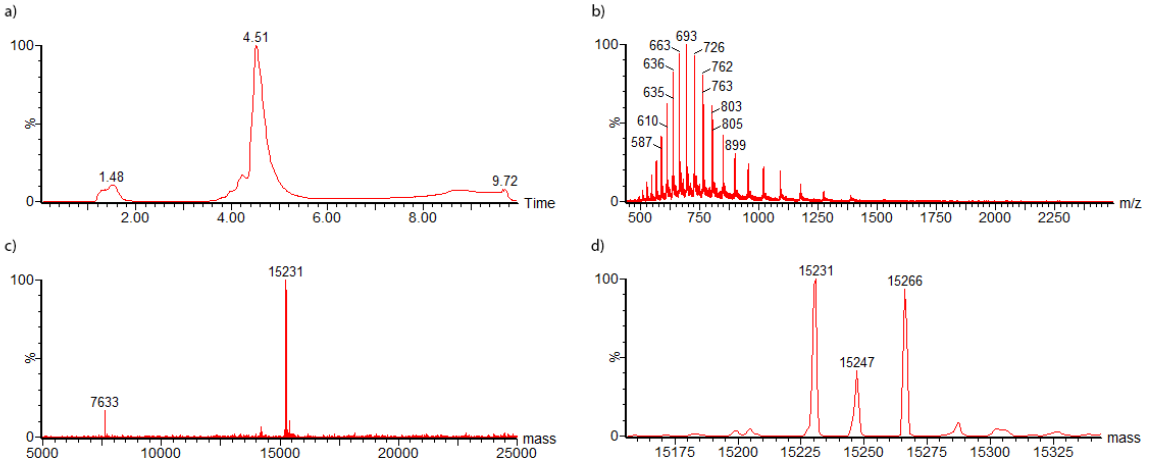
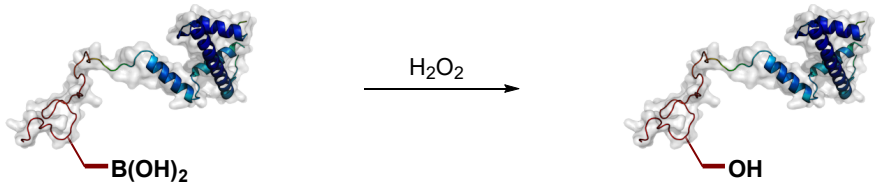


Figure 7.41: LC-MS analysis of Histone H3 Bal10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 15267 g/mol (Bal), 15249 g/mol (Bal$-H_2O$), 15231 g/mol (Bal$-2H_2O$); observed masses: 15266 g/mol (Bal), 15247 g/mol (Bal$-H_2O$), 15231 g/mol (Bal$-2H_2O$).

The mixture was vortexed and shaken for 30 min at room temperature. The protein was desalted by passing through a GE MiniTrap G-25 column pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) according to the manufacturer's instructions. The solution dialysed against NaP_i buffer (100 mM, pH 8.0) to remove any residual boric acid yielding 1000 µL of a solution of Histone Bal10 with a concentration of 1.31 mg/mL as determined by A_{280} spectrophotometry (1.31 mg, 85.8 nmol, 91% recovery, up to 98% conversion).

7.6.3 Histone H3 Ser10



Histone H3 S10Bal **Histone H3 S10S**

To a solution of Histone H3 Bal10 (200 µL, 1.31 mg/mL, 261 µg, 17 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.2) were added 10 µL of a solution of H_2O_2 in water (510 nmol, 30 equiv).

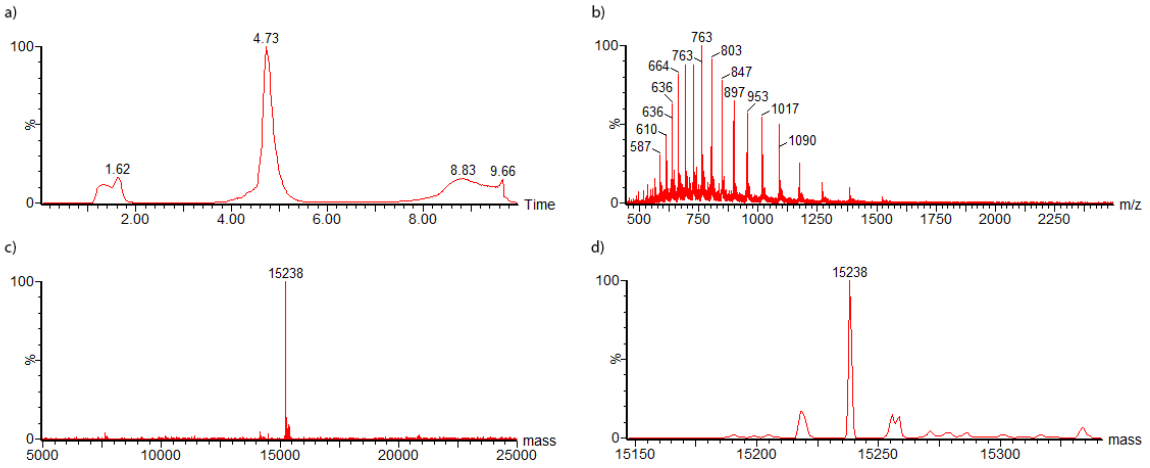
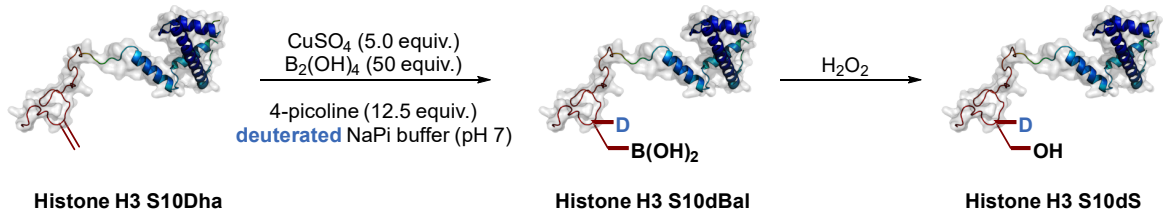


Figure 7.42: LC-MS analysis of Histone H3 Ser10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15239 g/mol; observed mass: 15238 g/mol.

The mixture was vortexed and shaken for 10 min at room temperature. The solution was dialysed against NaP_i buffer (100 mM, pH 8.0) to remove any residual boric acid or H_2O_2 yielding 350 µL of a solution of Histone H3 Ser10 with a concentration of 0.70 mg/mL as determined by A_{280} spectrophotometry (245 µg, 16 nmol, 94% recovery, full conversion).

7.6.4 Histone H3 Deuteration Study



A solution of Histone H3 Dha10 (250 µL, 0.78 mg/mL) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) was diluted with 750 µL D_2O and flash frozen in liquid N_2. The mixture was lyophilized, redissolved in 1000 µL of D_2O, shaken at room temperature for 10 min and flash frozen. This step was repeated twice before the protein solution was reconstituted with 250 µL of D_2O. Borylation was executed as described for Histone H3 Bal10, while all reactant solutions were prepared in D_2O. The reaction mixture was purified using a GE MiniTrap G-25 column and the purified protein was concentrated using VivaSpin 500 concentrators (5000 Da MWCO). The resulting solution was split in half and one aliquot was oxidized. Oxidation was performed with H_2O_2 (30 equiv) as reported previously. Both aliquots were then dialysed against NaP_i buffer (100 mM, pH 8.0).

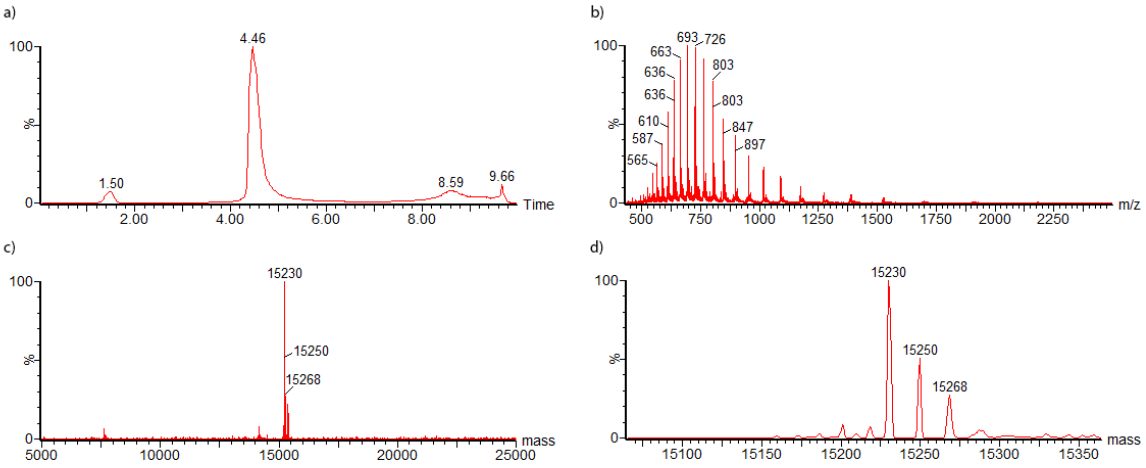


Figure 7.43: LC-MS analysis of Histone H3 *d1*-Bal10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 15268 g/mol (Bal), 15250 g/mol (Bal−H_2O), 15232 g/mol (Bal−$2H_2O$); observed masses: 15268 g/mol (Bal), 15250 g/mol (Bal−H_2O), 15230 g/mol (Bal−$2H_2O$).

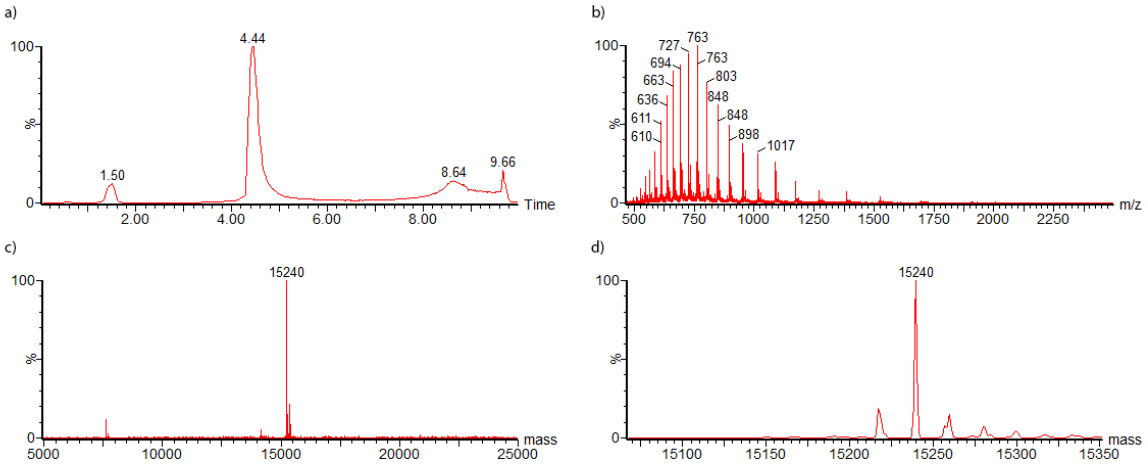


Figure 7.44: LC-MS analysis of Histone H3 *d1*-Ser10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 12240 g/mol; observed mass: 12240 g/mol.

7.6.5 [^{15}N]Histone H3 Bal10

[^{15}N]Histone H3 Bal10 was synthesized following the procedure for Histone H3 Bal10. To 3 × 1 mL of [^{15}N]Histone H3 Dha10 (2.04 mg/mL, 402 nmol, 1.00 equiv) in borylation buffer (100 mM NaP_i, 3 M Gdn·HCl, pH 7.0) were each added stock solutions of 4-picoline (12.5 equiv), $CuSO_4$ (5.0 equiv) and $B_2(OH)_4$ (50.0 equiv). The reaction mixture was dialyzed against milli-Q water and lyophilized.

95% conversion was observed via LC-MS analysis.

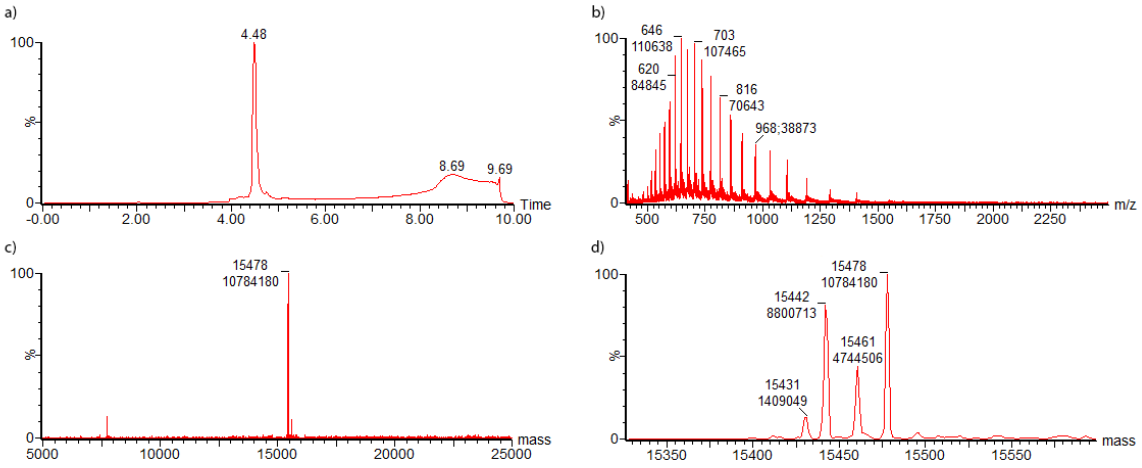


Figure 7.45: LC-MS analysis of [^{15}N]Histone H3 Bal10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 15480 g/mol (Bal), 15462 g/mol (Bal−H_2O), 15444 g/mol (Bal−$2H_2O$); observed masses: 15478 g/mol (Bal), 15461 g/mol (Bal−H_2O), 15442 g/mol (Bal−$2H_2O$).

7.6.6 [^{15}N]Histone H3 Ser10

[^{15}N]Histone H3 Ser10 was synthesized by incubation of Histone H3 Bal10 (approx. 4 mg/mL) with 10 mM H_2O_2 for 30 min in unfolding buffer (7 M Gdn·HCl, 10 mM TRIS base, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine).

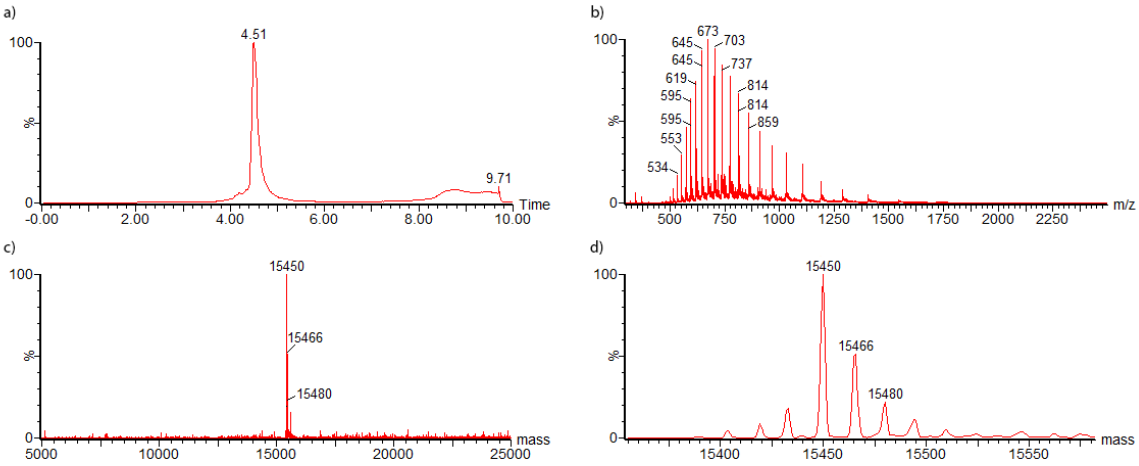
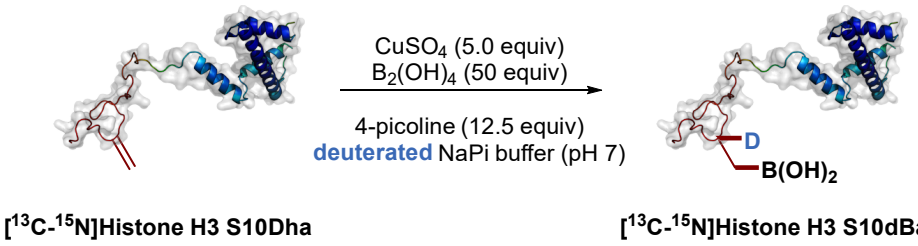


Figure 7.46: LC-MS analysis of [^{15}N]Histone H3 Bal10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15452 g/mol; observed masses: 15450 g/mol, 15466 g/mol (+ 1 × oxidation), 15480 g/mol (+ 2 × oxidation).

7.6.7 [^{13}C-^{15}N]Histone H3 *d1*-Bal10



Deuterated NaP_i buffer was produced using the following procedure: NaP_i buffer (100 mM NaP_i, 3 M Gdn·HCl, pH 7.0) was frozen in liquid nitrogen and lyophilized. The lyophilized power was reconstituted with D_2O and frozen and lyophilized again. The lyophilized power was then reconstituted with D_2O and used to dissolve previously lyophilized [^{13}C-^{15}N]Histone H3 Dha10.

Stock solutions of $CuSO_4 \cdot 5H_2O$ (15.7 mg in 1.00 mL of D_2O), 4-picoline (15.4 µL in 1.00 mL of D_2O) and $B_2(OH)_4$ (3.4 mg in 1.00 mL of D_2O) were freshly prepared before the reaction was conducted.

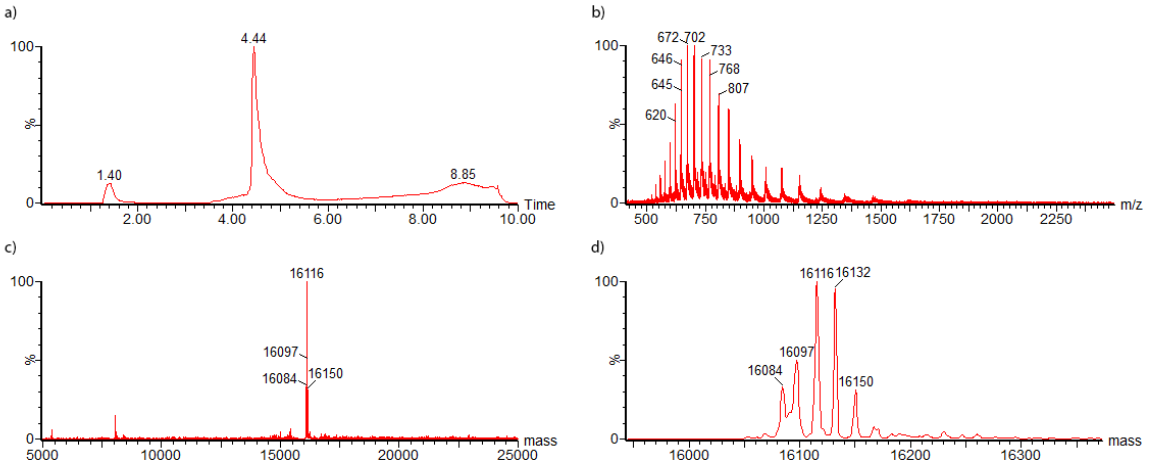
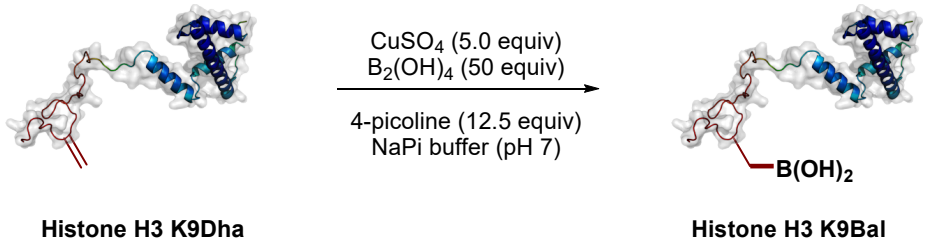


Figure 7.47: LC-MS analysis of [^{13}C-^{15}N]Histone H3 *d1*-Bal10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 16083 g/mol ([^{13}C-^{15}N]Histone H3 Dha10), 16094 g/mol ([^{13}C-^{15}N]Histone H3 *d1*-Bal10−2H_2O), 16112 g/mol ([^{13}C-^{15}N]Histone H3 *d1*-Bal10−H_2O), 16130 g/mol ([^{13}C-^{15}N]Histone H3 *d1*-Bal10); observed masses: 16084 g/mol ([^{13}C-^{15}N]Histone H3 Dha10), 16097 g/mol ([^{13}C-^{15}N]Histone H3 *d1*-Bal10−2H_2O), 16116 g/mol ([^{13}C-^{15}N]Histone H3 *d1*-Bal10−H_2O), 16132 g/mol ([^{13}C-^{15}N]Histone H3 *d1*-Bal10).

To the solution of [^{13}C-^{15}N]Histone H3 Dha10 in deuterated NaP_i buffer (1600 µL, 1.20 mg/mL,

1.92 mg, 126 nmol) were added previously prepared stock solutions of 4-picoline (10 µL, 1.57 µmol, 12.5 equiv), $CuSO_4$ (10 µL, 0.63 µmol, 5.00 equiv) and $B_2(OH)_4$ (130 µL, 6.30 µmol, 50 equiv). The mixture was incubated at room temperature for 30 min before being purified using two GE MidiTrap G-25 columns pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) according to the manufacturer's instructions. The sample was then dialysed against milli-Q water (3 × 4 L) using a dialysis cassette (MWCO 3,500 Da) and lyophilized. The lyophilized power was dissolved in 330 µL of NMR buffer (50 mM NaP_i, 3 M Gdn·HCl, 5% D_2O). The concentration of the sample was determined at 5.04 mg/mL by A_{280} photospectrometry (1.66 mg, 109 nmol, 86% recovery).

7.6.8 Histone H3 Bal9



Stock solutions of $CuSO_4 \cdot 5H_2O$ (2.35 mg in 1.00 mL of H_2O), 4-picoline (2.29 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (2.81 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of Histone H3 Dha9 (500 µL, 2.54 mg/mL, 1.27 mg, 84 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) were added previously prepared stock solutions of 4-picoline (10.0 µL, 1.05 µmol, 12.5 equiv), $CuSO_4$ (10.0 µL, 418 nmol, 5.00 equiv) and $B_2(OH)_4$ (30.0 µL, 4.18 µmol, 50.0 equiv) subsequently.

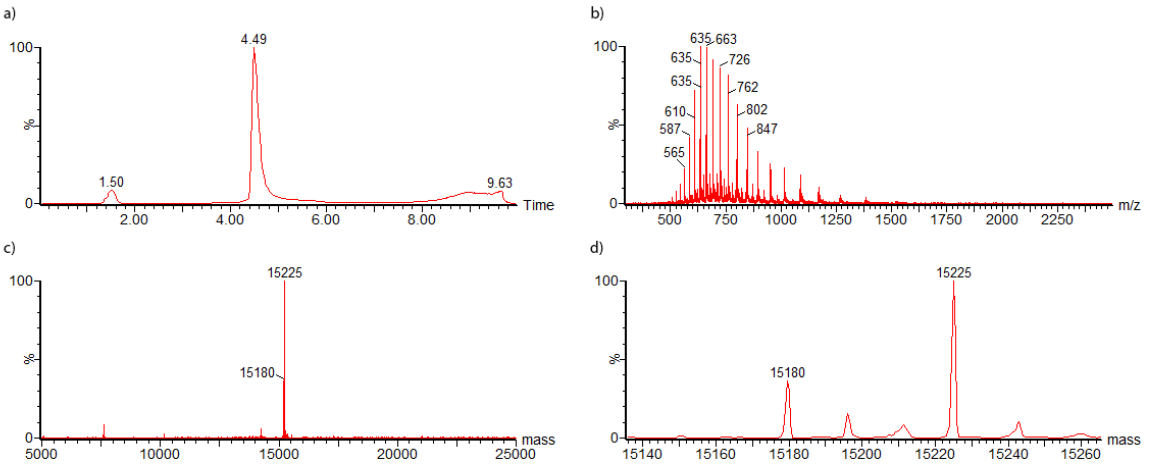


Figure 7.48: LC-MS analysis of Histone H3 Bal9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15226 g/mol (Bal); observed mass: 15225 g/mol (Bal).

The mixture was vortexed and shaken for 30 min at room temperature. The protein was desalted by passing through a GE MiniTrap G-25 column pre-equilibrated with NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) according to the manufacturer's instructions. The solution dialysed against NaP_i buffer (100 mM, pH 8.0) to remove any residual boric acid yielding

1000 µL of a solution of Histone H3 Bal9 with a concentration of 1.11 mg/mL as determined by A_{280} spectrophotometry (0.89 mg, 72 nmol, 86% recovery, up to 73% conversion).

7.6.9 [^{15}N]Histone H3 Bal9

[^{15}N]Histone H3 Bal10 was synthesized following the procedure for Histone H3 Bal9. To 3 × 1 mL of [^{15}N]Histone H3 Dha9 (1.75 mg/mL, 346 nmol, 1.00 equiv) in borylation buffer (100 mM NaP_i, 3 M Gdn·HCl, pH 7.0) were each added stock solutions of 4-picoline (12.5 equiv), $CuSO_4$ (5.0 equiv) and $B_2(OH)_4$ (50.0 equiv). The reaction mixture was dialyzed against milli-Q water and lyophilized.

Full conversion was observed via LC-MS analysis.

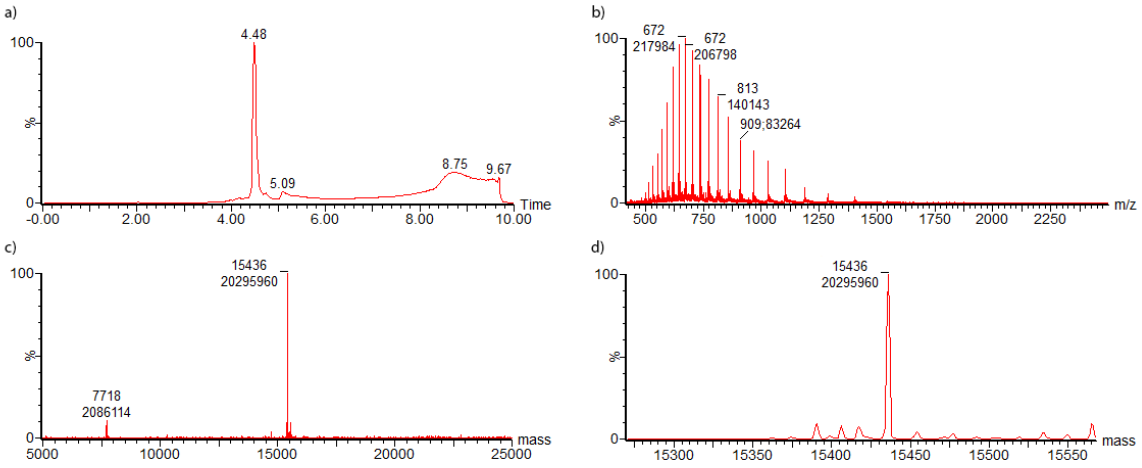


Figure 7.49: LC-MS analysis of [^{15}N]Histone H3 Bal9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15438 g/mol (Bal); observed masses: 15436 g/mol (Bal).

7.6.10 [^{15}N]Histone H3 Ser9

[^{15}N]Histone H3 Ser9 was synthesized by incubation of Histone H3 Bal9 (approx. 4 mg/mL) with 10 mM H_2O_2 for 30 min in unfolding buffer (7 M Gdn·HCl, 10 mM TRIS base, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine).

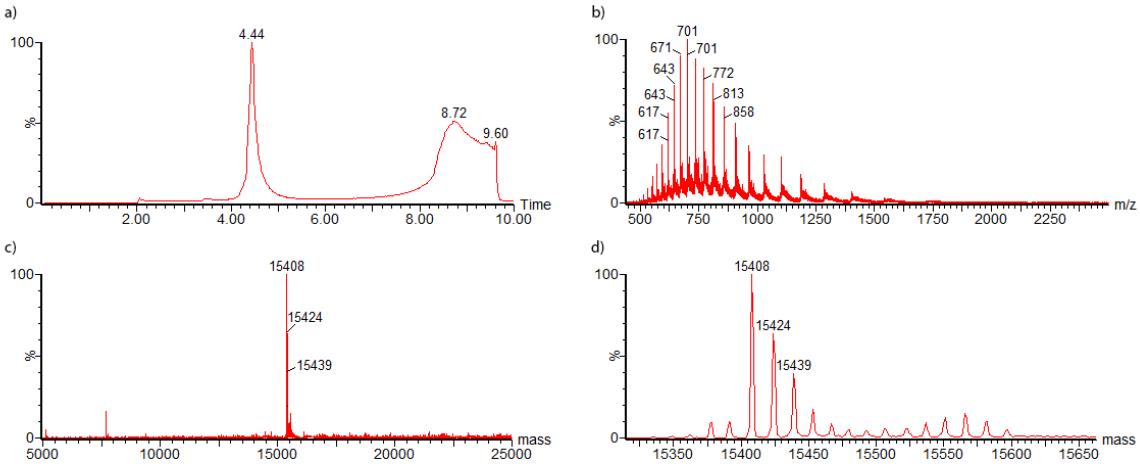
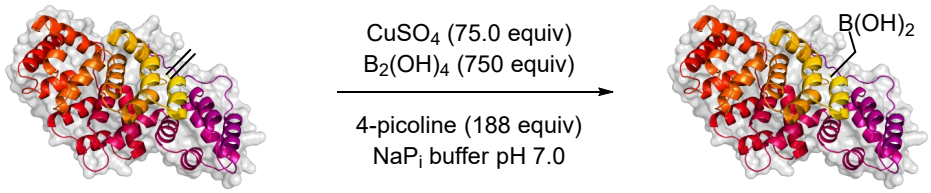


Figure 7.50: LC-MS analysis of [^{15}N]Histone H3 Ser9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 15410 g/mol; observed masses: 15408 g/mol, 15424 g/mol (+ 1 × oxidation), 15439 g/mol (+ 2 × oxidation).

7.6.11 Annexin V Bal316



Annexin V C316Dha **Annexin V C316Bal**

Stock solutions of $CuSO_4 \cdot 5H_2O$ (5.2 mg in 1.00 mL of H_2O), 4-picoline (5.10 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (7.5 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of Annexin V Dha316 (40 µL, 0.20 mg/mL, 0.56 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) pre-warmed to 45°C were added previously prepared and pre-warmed stock solutions of 4 picoline (2.00 µL, 104.8 nmol, 188 equiv), $CuSO_4$ (2.00 µL, 41.9 nmol, 75.0 equiv) and $B_2(OH)_4$ (5.00 µL, 419 nmol, 750 equiv) subsequently. The mixture was vortexed and shaken for 10 min at 45°C. LC-MS indicated full conversion.

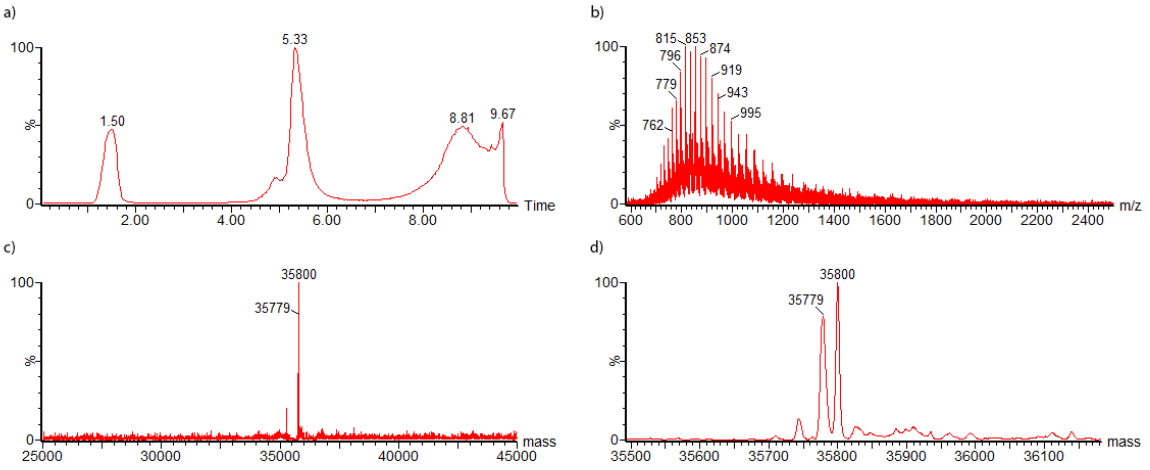
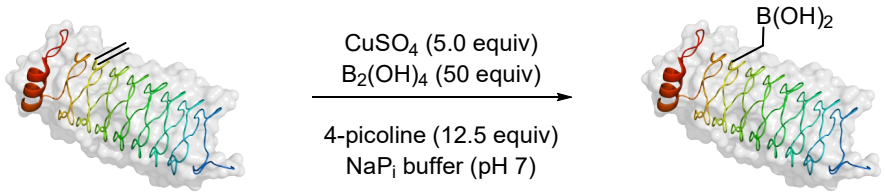


Figure 7.51: LC-MS analysis of Annexin V Bal316: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 35799 g/mol (Bal−H_2O), 35781 g/mol (Bal−$2H_2O$); observed masses: 35800 g/mol (Bal−H_2O), 35779 g/mol (Bal−$2H_2O$).

7.6.12 Npβ Bal61



Stock solutions of $CuSO_4 \cdot 5H_2O$ (3.5 mg in 1.00 mL of H_2O), 4-picoline (1.71 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (6.3 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of Npβ Dha61 (50 µL, 1.12 mg/mL, 2.67 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) were subsequently added stock solutions of 4-picoline (1.90 µL, 13.4 nmol, 12.5 equiv), $CuSO_4$ (0.95 µL, 13.4 nmol, 5.00 equiv) and $B_2(OH)_4$ (1.90 µL, 134 nmol, 50 equiv). The mixture was briefly vortexed and incubated at room temperature for 10 min (>95% conversion).

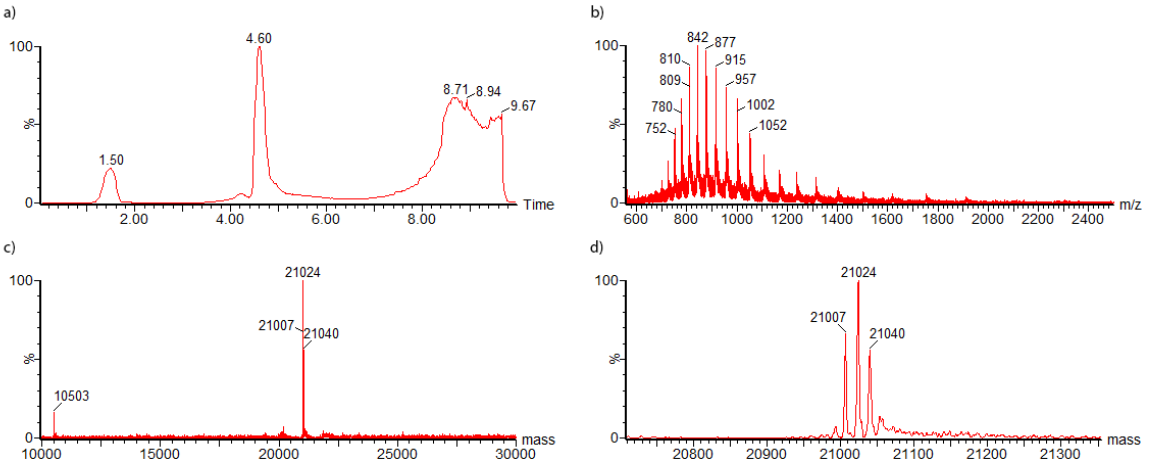
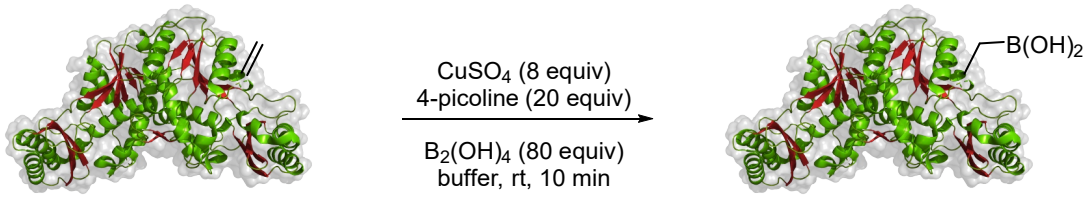


Figure 7.52: LC-MS analysis of Npβ Bal61: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 21043 g/mol (Bal), 21025 g/mol (Bal−H_2O), 21007 g/mol (Bal−$2H_2O$); observed masses: 21040 g/mol (Bal), 21024 g/mol (Bal−H_2O), 21007 g/mol (Bal−$2H_2O$).

7.6.13 PanC Bal44



Stock solutions of $CuSO_4 \cdot 5H_2O$ (3.5 mg in 1.00 mL of H_2O), 4-picoline (1.71µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (6.3 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of PanC Dha44 (kindly provided by Dr. P. Isenegger) (50 µL, 0.99 mg/mL, 1.46 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) were subsequently added stock solutions of 4-picoline (1.70 µL, 29.2 nmol, 20 equiv), $CuSO_4$ (0.85 µL, 11.7 nmol, 8.00 equiv) and $B_2(OH)_4$ (1.70 µL, 117 nmol, 80 equiv). The mixture was briefly vortexed and incubated at room temperature for 10 min (71% conversion).

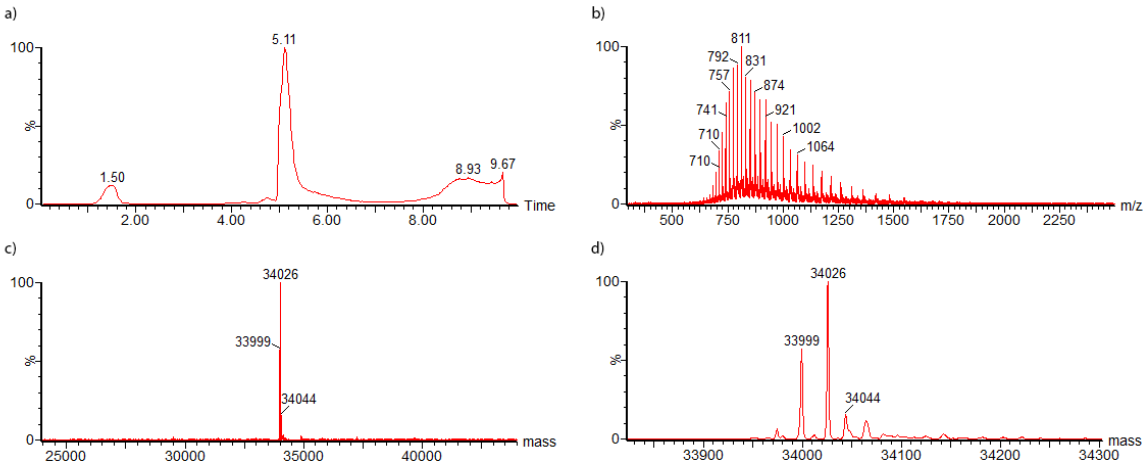
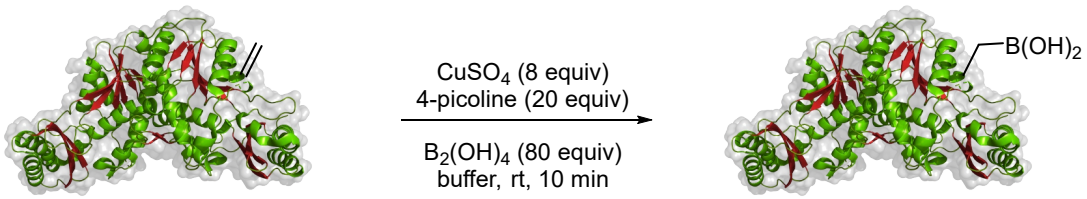


Figure 7.53: LC-MS analysis of PanC Bal44: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 34044 g/mol (Bal), 34026 g/mol (Bal−H_2O); observed masses: 34044 g/mol (Bal), 34026 g/mol (Bal−H_2O).

7.6.14 PanC Bal47



Stock solutions of $CuSO_4 \cdot 5H_2O$ (3.5 mg in 1.00 mL of H_2O), 4-picoline (1.71µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (6.3 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of PanC Dha47 (kindly provided by Dr. P. Isenegger) (50 µL, 1.16 mg/mL, 1.71 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) were subsequently added stock solutions of 4-picoline (2.00 µL, 34.1 nmol, 20 equiv), $CuSO_4$ (1.00 µL, 13.7 nmol, 8.00 equiv) and $B_2(OH)_4$ (2.00 µL, 137 nmol, 80 equiv). The mixture was briefly vortexed and incubated at room temperature for 10 min (60% conversion).

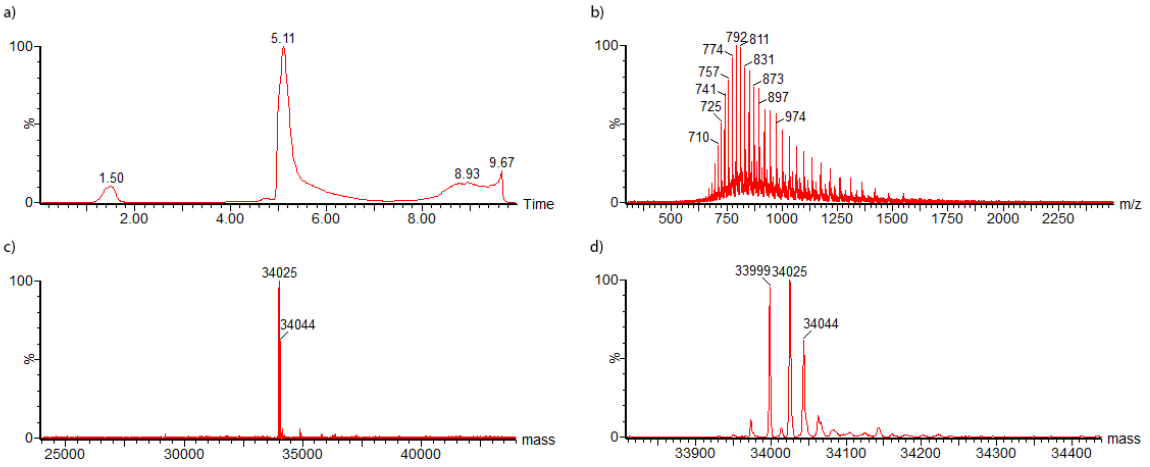
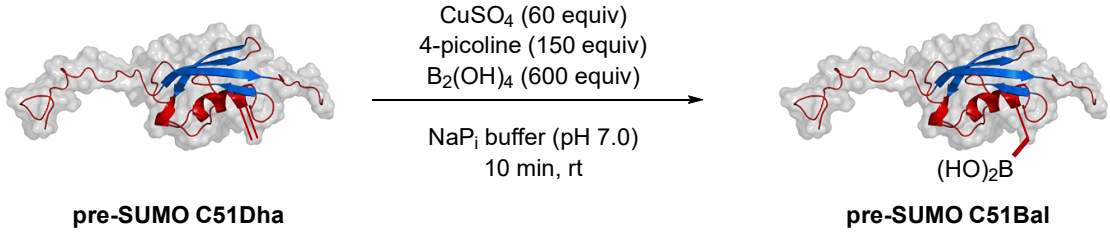


Figure 7.54: LC-MS analysis of PanC Bal47: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 34044 g/mol (Bal), 34026 g/mol (Bal−H_2O); observed masses: 34044 g/mol (Bal), 34025 g/mol (Bal−H_2O).

7.6.15 pre-SUMO1 Bal51



Stock solutions of $CuSO_4 \cdot 5H_2O$ (40.3 mg in 1.00 mL of H_2O), 4-picoline (39.3 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (8.1 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of pre-SUMO1 Dha51 (500 µL, 0.67 mg/mL, 26.9 nmol, 1.00 equiv) in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) were subsequently added stock solutions of 4-picoline (10.0 µL, 4.04 µmol, 150 equiv), $CuSO_4$ (10.0 µL, 1.62 µmol, 60 equiv) and $B_2(OH)_4$ (180 µL, 16.2 µmol, 600 equiv). The mixture was briefly vortexed and incubated at room temperature for 10 min (>95% conversion). The mixture was dialysed against SUMO maturation buffer (50 mM TRIS base, 20 mM NaCl, 2 mM DTT).

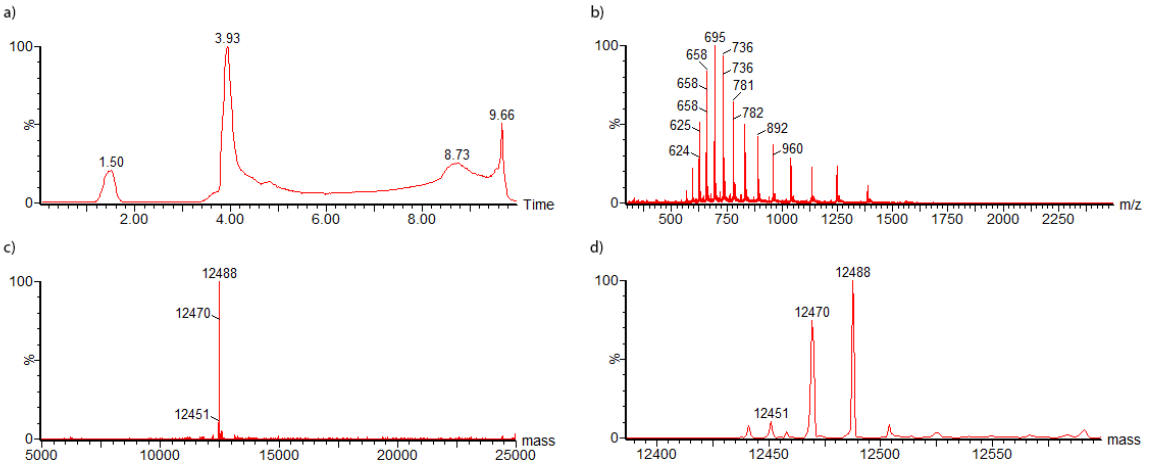
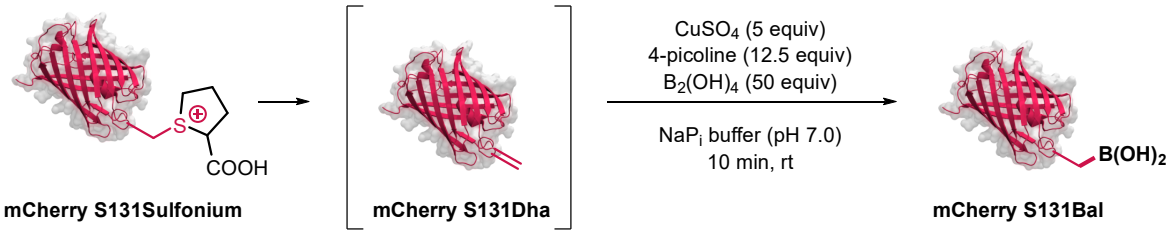


Figure 7.55: LC-MS analysis of pre-SUMO1 Bal51: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 12487 g/mol (Bal), 12469 g/mol (Bal−H_2O), 12451 g/mol (Bal−2H_2O); observed masses: 12488 g/mol (Bal), 12470 g/mol (Bal−H_2O), 12451 g/mol (Bal−2H_2O).

7.6.16 mCherry Bal131



Stock solutions of $CuSO_4 \cdot 5H_2O$ (5.3 mg in 1.00 mL of H_2O), 4-picoline (5.19 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (4.8 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

A solution of mCherry Sulfonium131 (100 µL, 2.41 mg/mL, 8.54 nmol, 1.00 equiv) in NaP_i buffer (100 mM, pH 7.0) was diluted with 100 µL of NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0). Stock solutions of 4-picoline (2.0 µL, 107 nmol, 12.5 equiv), $CuSO_4$ (2.0 µL, 42.7 nmol, 5.0 equiv) and $B_2(OH)_4$ (8.0 µL, 427 nmol, 50 equiv) were added subsequently. The mixture was briefly vortexed and incubated at room temperature for 10 min (35% conversion).

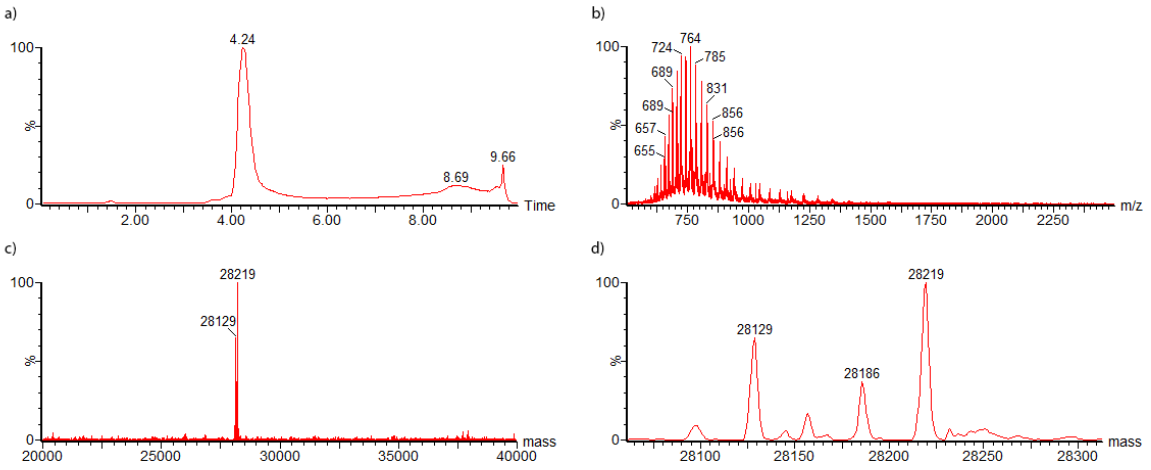
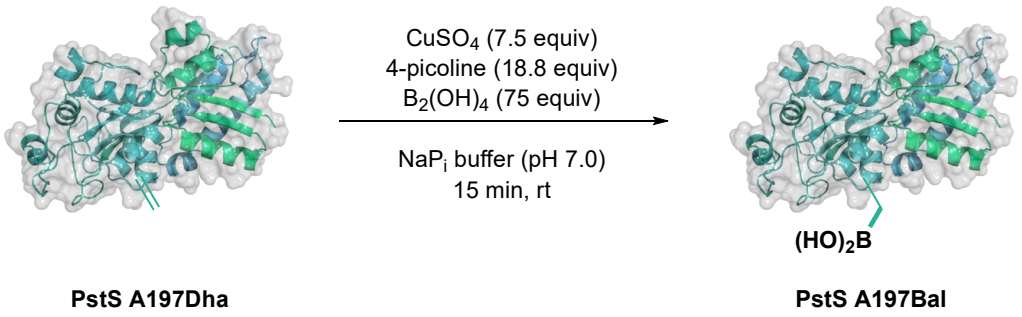


Figure 7.56: LC-MS analysis of mCherry Bal131: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 28128 g/mol (Bal); observed masses: 28129 g/mol (Bal).

7.6.17 PstS Bal197



Stock solutions of $CuSO_4 \cdot 5H_2O$ (3.0 mg in 1.00 mL of H_2O), 4-picoline (2.90 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (5.4 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To a solution of PstS Bal197 (500 µL, 0.44 mg/mL, 220 µg, 6.37 nmol) in NaP_i buffer (100mM, 3 M Gdn·HCl, pH 7.0) were added the previously prepared stock solutions of 4-picoline (4.00 µL, 119 nmol, 18.8 equiv), $CuSO_4$ (4.00 µL, 47.8 nmol, 7.50 equiv) and $B_2(OH)_4$ (8.00 µL, 478 nmol, 75.0 equiv). The sample was vortexed and incubated at room temperature for 15 min (79% conversion).

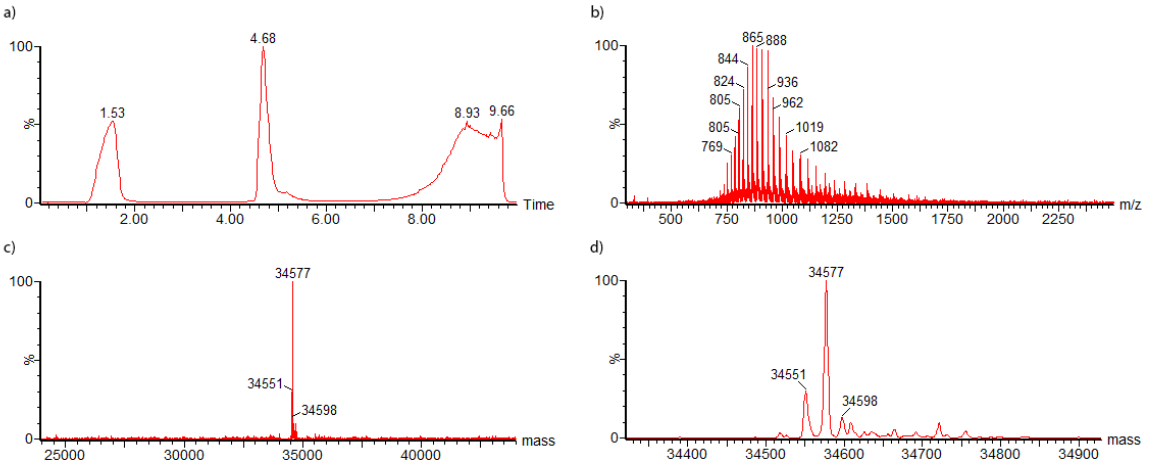
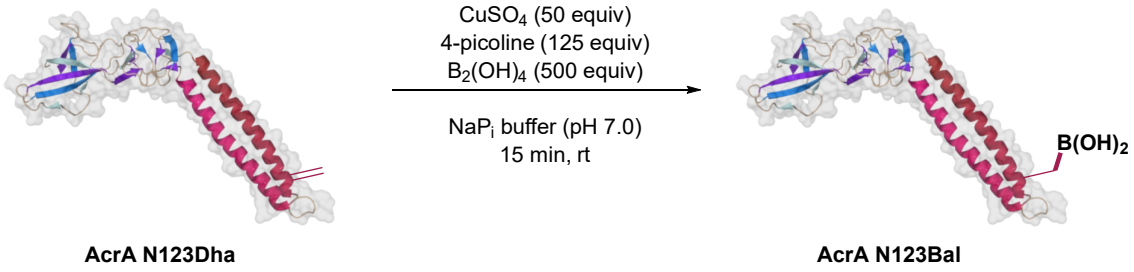


Figure 7.57: LC-MS analysis of PstS Bal197: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 34597 g/mol (Bal), 34579 g/mol (Bal–H_2O); observed masses: 34598 g/mol (Bal), 34577 g/mol (Bal–H_2O).

7.6.18 AcrA Bal123



Stock solutions of $CuSO_4 \cdot 5H_2O$ (2.2 mg in 1.00 mL of H_2O), 4-picoline (2.09 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (3.9 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To 50 µL of a solution of AcrA Dha123 (0.13 mg/mL, 0.17 nmol) in PBS were added 50 µL of NaP_i buffer (100 mM, pH 7.0, 6 M Gdn·HCl). The previously prepared stock solutions of 4-picoline (1.00 µL, 21.7 nmol, 125 equiv), $CuSO_4$ (1.00 µL, 8.70 nmol, 50 equiv) and $B_2(OH)_4$ (2.00 µL, 87.0 nmol, 500 equiv) were added, the sample was briefly vortexed and incubated for 15 min (77% conversion).

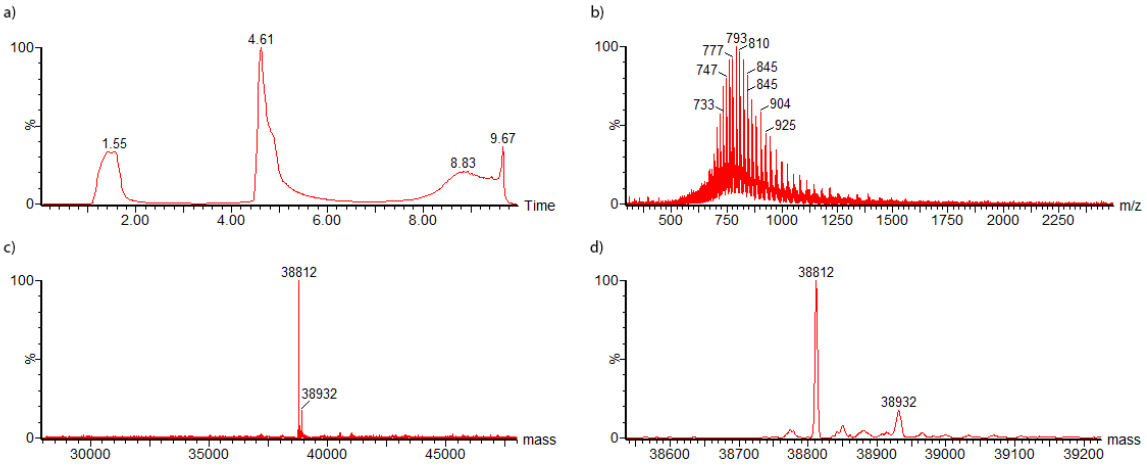
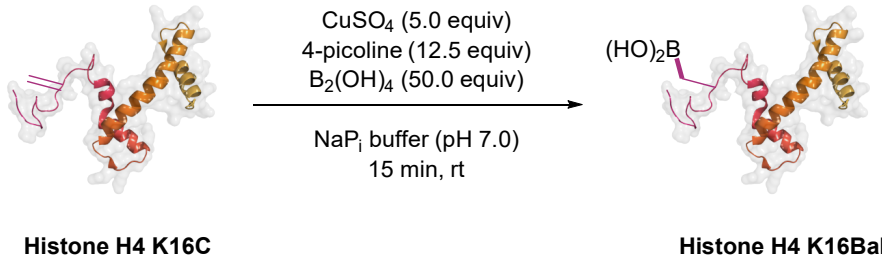


Figure 7.58: LC-MS analysis of AcrA Bal123: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 38811 g/mol (Bal−H_2O); observed mass: 38812 g/mol (Bal−H_2O).

7.6.19 Histone H4 Bal16



Stock solutions of $CuSO_4 \cdot 5H_2O$ (2.1 mg in 1.00 mL of H_2O), 4-picoline (2.02 µL in 1.00 mL of H_2O) and $B_2(OH)_4$ (3.7 mg in 1.00 mL of H_2O) were freshly prepared before the reaction was conducted.

To 50 µL of a solution of Histone H4 Dha16 (0.37 mg/mL, 1.66 nmol) in NaP_i buffer (100 mM, pH 7.0, 3 M Gdn·HCl) were added the previously prepared stock solutions of 4-picoline (1.00 µL, 20.7 nmol, 12.5 equiv), $CuSO_4$ (1.00 µL, 8.28 nmol, 5.00 equiv) and $B_2(OH)_4$ (2.00 µL, 82.8 nmol, 50 equiv) were added, the sample was briefly vortexed and incubated for 15 min. Full conversion was determined by LC-MS.

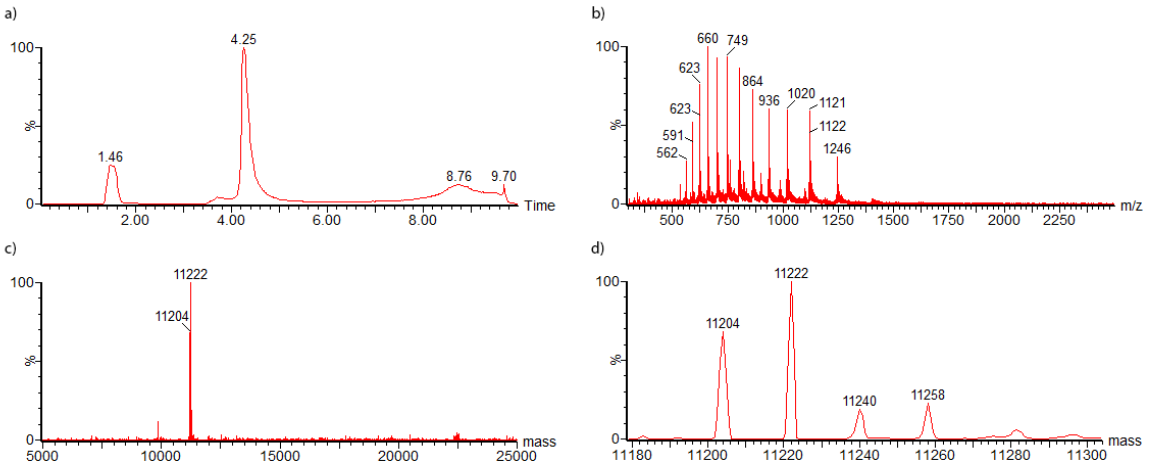


Figure 7.59: LC-MS analysis of Histone H4 Bal16: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 11223 g/mol (Bal), 11205 g/mol (Bal–H_2O); observed masses: 11222 g/mol (Bal), 11204 g/mol (Bal–H_2O).

7.7 Calculation of Residue Accessibility

The solvent-accessible surface area was calculated using FreeSASA.[20] The Lee & Richards algorithm[21] was used on PDB structures having the site of interest mutated to Cys (mutation conducted in Open-Source PyMol Molecular Graphics System 2.4). The number of slices was set to 100, atom radii from NACCESS were used and different probe sizes (1.00 Å, 1.40 Å and 2.80 Å) were utilized. The absolute values obtained for the solvent-accessible surface area of the respective Cys residues were set in relation to a tripeptide Gly-Cys-Gly using the value published by Tien et al.[22]

7.8 Reactive PLABP

To purified solution of the borylated protein of interest in borylation buffer (NaP_i 100 mM, pH 7.0, 3 M Gdn·HCl) (20-50 µL) was added H_2O_2 to a final concentration of 1 mM to 20 mM. The sample was incubated for 10 min before being diluted with MS buffer (1:50, 0.1% formic acid in milli-Q water) and analysed via LC-MS.

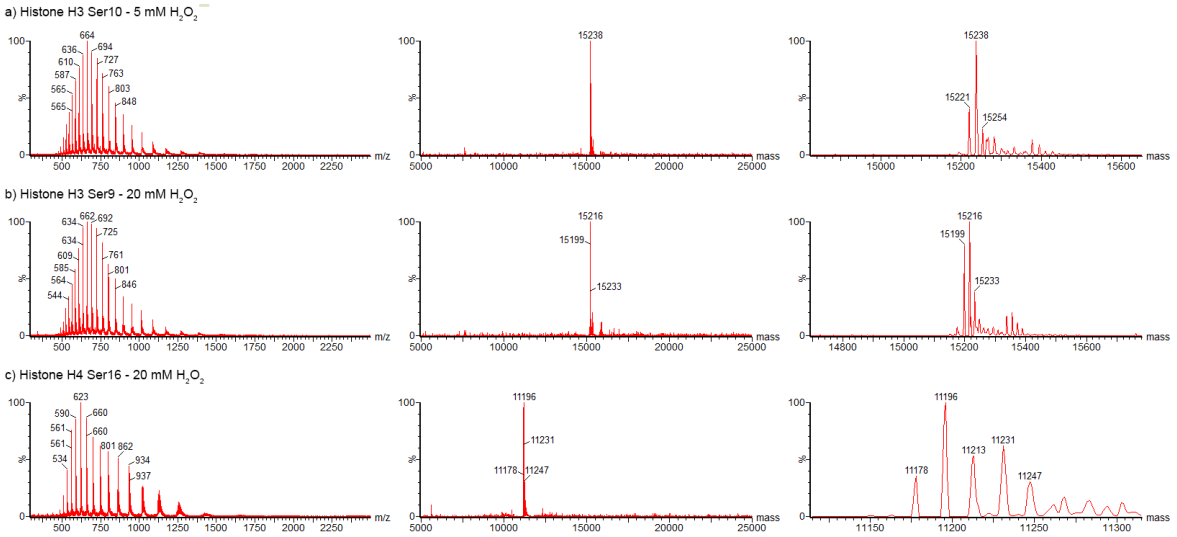


Figure 7.60: Ion series and deconvoluted spectra for oxidation of borylated mutants to Histone H3 Ser10 (M_{calc} = 15,239 Da; M_{obs} = 15,238 Da; 15,254 Da (mono-oxidation)) (**a**), Histone H3 Ser9 (M_{calc} = 15,198 Da; M_{obs} = 15,199 Da; 15,216 Da (mono-oxidation); 15,233 Da (di-oxidation)) (**b**) and Histone H4 Ser16 (M_{calc} = 11,195 Da; M_{obs} = 11,196 Da; 11,213 Da (mono-oxidation); 11,213 Da (di-oxidation); 11,231 Da (tri-oxidation)) (**c**).

7.9 Trifluoroboronoalanine (3F-Bal) Formation & Hydrolysis

7.9.1 3F-Bal Formation on Histone H3 Model Proteins

Lyophilized Histone H3 Bal9 or Bal10 was dissolved in 3F-Bal formation buffer (47 mM KHF_2, 10.5 mM NaP_i, 316 mM Gdn·HCl, pH 4.4, 10% v/v D_2O) in a microcentrifuge tube. The solution was transferred to a NMR tube and analysed by ^{19}F-NMR on a Bruker Avance III HD 600 MHz NMR spectrometer equipped with a Prodigy N2 broadband cryoprobe.

7.9.2 Determination of Hydrolysis on Model Substrate

General Procedure

Ac-(3F-Bal)-NHBn (**15**) (3.00 mg, 9.20 µmol) was dissolved in D_2O (control) or buffer in D_2O (550 µL). The mixture was transferred to a FEP NMR tube liner and sonicated briefly to remove air bubbles. Samples were analysed by ^{19}F-NMR using a Bruker NEO 400 MHz nanobay instrument equipped with a 5 mm Z-gradient broadband multinuclear SMART probe. A standard pulse sequence (*zg30*), 5 s relaxation delay and 1024 scans were used. NMR spectra were processed using Bruker TopSpin 4.0.7.

Screening of Amino Acid Additives

The influence of the canonical amino acids was investigated following the general procedure and dissolving Ac-(3F-Bal)-NHBn (**15**) in TRIS buffer (100 mM, pH 7.0, D_2O) supplemented with 10 mM of the free amino acid of interest. In the case of tyrosine, an amino acid concentration of 5 mM was used due to limited solubility.

7.10 Protein Cleavage

7.10.1 Reaction Optimization

Cleavage Additive Screening

Lyophilized Histone H3 Dha10 was dissolved in milli-Q water at 1.0 mg/mL. Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL), TEMPOL (4-Hydroxy-2,2,6,6-tetramethylpiperidinyl-oxyl, 14.2 mg in 1000 µL), iron(II) sulfate heptahydrate (22.8 mg in 1000 µL) and sodium ascorbate (16.3 mg in 1000 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. For each reaction, 50 µL of Histone H3 Dha10 (50 µg, 3.29 nmol, 1.00 equiv) were incubated at room temperature with 2 µL of additive (164 nmol, 50 equiv) and 10 µL of tetrahydroxydiboron (657 nmol, 200 equiv). Samples were taken after 30 min, 60 min and 180 min and directly analysed by LC-MS. Cleavage and oxidation were determined from relative intensities after deconvolution of the corresponding mass spectra.

Cleavage Boron Reagent Screening

Lyophilized Histone H3 Dha10 was dissolved in milli-Q water at 1.0 mg/mL. Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL), *bis*(catecholato)diboron (15.6 mg in 1000 µL), $NaBH_4$ (2.5 mg in 1000 µL), $NaBH_3CN$ (4.1 mg in 1000 µL) and sodium ascorbate (13.0 mg in 1000 µL) in milli-Q water were prepared. A stock solution of *bis*(pinacolato)diboron (16.7 mg in 1000 µL) was prepared in DMF. Stock solutions of catecholborane (14.1 µL, 15.8 mg in 1000 µL) and pinacolborane (19.1 µL, 16.8 mg in 1000 µL) were prepared in dry MeCN. All stock solutions were prepared freshly before use.

40 µL of Histone H3 Dha10 (40 µg, 2.63 nmol) were treated with sodium ascorbate (2 µL, 131 nmol, 50 equiv) followed by boron reagent (8 µL, 526 nmol, 200 equiv) and incubated for 3 hours at room temperature. The samples were analysed by LC-MS and cleavage was determined from relative intensities after deconvolution of the corresponding mass spectra.

Antioxidant Concentration Screening

Lyophilized Histone H3 Dha10 was dissolved in milli-Q water at 1.0 mg/mL. Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL) and sodium ascorbate (32.5 mg in 500 µL) in milli-Q water were prepared freshly directly before the reactions were conducted.

50 µL of Histone H3 Dha10 (50 µg, 3.29 nmol, 1.00 equiv) were incubated at room temperature with varying concentrations of sodium ascorbate (10 µL, 3.29 mmol, 1000 equiv; 5 µL, 1.64 mmol, 500 equiv; 1.0 µL, 329 µmol, 100 equiv; 0.1 µL, 16.4 µmol, 10 equiv) and 10 µL of tetrahydroxydiboron (657 nmol, 200 equiv). Samples were taken after 30 min, 90 min, 180 min and 360 min and directly analysed by LC-MS. Cleavage and oxidation were determined from relative intensities after deconvolution of the corresponding mass spectra.

Temperature Screening

Lyophilized Histone H3 Dha10 was dissolved in milli-Q water at 1.0 mg/mL. Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL) and sodium ascorbate (13.0 mg in 1000 µL) in milli-Q water were prepared.

40 µL of Histone H3 Dha10 (40 µg, 2.63 nmol) were treated with sodium ascorbate (2 µL, 131 nmol, 50 equiv) followed by tetrahydroxydiboron (8 µL, 526 nmol, 200 equiv) and incubated at 4°C, room temperature or 37°C. Samples were taken after 30 min, 60 min, 90 min, 120 min, 180 min, 360 min, 24 h and 4 days. No further samples were taken once the reaction was found to have gone to completion. The samples were analysed by LC-MS and cleavage was determined from relative intensities after deconvolution of the corresponding mass spectra.

Buffer Screening

Lyophilized Histone H3 Dha10 was dissolved in milli-Q water at 4.0 mg/mL. A 2× stock of each buffer was prepared. Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL) and sodium ascorbate (13.0 mg in 1000 µL) in milli-Q water were prepared.

A mixture of 2× buffer (20 µL), Histone H3 Dha10 (10 µL, 40 µg, 2.63 nmol), sodium ascorbate

(2 µL, 131 nmol, 50 equiv) and tetrahydroxydiboron (8 µL, 526 nmol, 200 equiv) was incubated at room temperature for 3 hours. A sample was taken and directly analysed by LC-MS. Cleavage was determined from relative intensities after deconvolution of the corresponding mass spectrum.

7.10.2 Deuteration Experiments

LC-MS Analysis of Protein Deuteration

LC-MS analysis of protein deuteration was conducted on high resolution QTof mass spectrometers (Waters Xevo G2-S or Waters Xevo G2-XS). All samples were measured and analysed in triplicate. Samples were analysed in MassLynx 4.1 (Waters) using the integrated MaxEnt1 algorithm for deconvolution. For deuteration experiments, the mass range was set to 5000 Da to 25000 Da, the mass resolution was set to 0.04 Da/channel, the damage model was set to uniform Gaussian with 0.4 Da width at half-height. Minimum intensity ratios were set to 33% left and 33% right.

Deuterative Cleavage

Lyophilized Histone H3 Dha10 was dissolved in D_2O at 1.0 mg/mL. Stock solutions of $B_2(OD)_4$ (5.9 mg in 1000 µL) and sodium ascorbate (13.0 mg in 1000 µL) in D_2O were prepared.

40 µL of Histone H3 Dha10 (40 µg, 2.63 nmol) were treated with sodium ascorbate (2 µL, 131 nmol, 50 equiv) followed by $B_2(OD)_4$ (8 µL, 526 nmol, 200 equiv) and incubated room temperature for 3 hours. A control experiment was conducted under identical conditions using H_2O and $B_2(OH)_4$ instead of D_2O and $B_2(OD)_4$. Samples were taken and directly analysed by LC-MS as described above.

Transfer Deuteration under Anaerobic Conditions

Lyophilized Histone H3 Dha10 was dissolved in D_2O at 1.0 mg/mL and transferred to a nitrogen filled glove box (2.1 ppm O_2). The sample was degassed by standing open overnight. Fresh stock solutions of $B_2(OD)_4$ (5.9 mg in 1000 µL) and sodium ascorbate (13.0 mg in 1000 µL) in D_2O were prepared in the glove box.

100 µL of Histone H3 Dha10 (100 µg, 6.57 nmol) were treated with sodium ascorbate (5 µL, 329 nmol, 50 equiv) followed by $B_2(OD)_4$ (20 µL, 1.31 µmol, 200 equiv) and incubated room temperature for 3 hours. A control experiment was conducted under identical conditions using H_2O instead of D_2O. Samples were taken and directly analysed by LC-MS as described above.

Deuteration under Aerobic Conditions in the Presence of TEMPOL

Lyophilized Histone H3 Dha10 was dissolved in D_2O at 1.0 mg/mL. Stock solutions of $B_2(OD)_4$ (5.9 mg in 1000 µL) and TEMPOL (11.3 mg in 1000 µL) in D_2O were prepared.

100 µL of Histone H3 Dha10 (100 µg, 6.57 nmol) were treated with TEMPOL (5 µL, 329 nmol, 50 equiv) followed by $B_2(OD)_4$ (20 µL, 1.31 µmol, 200 equiv) and incubated room temperature for 24 hours. A control experiment was conducted under identical conditions using H_2O and $B_2(OH)_4$ instead of D_2O and $B_2(OD)_4$. Samples were taken and directly analysed by LC-MS as described above.

7.10.3 Protein Cleavage Scope

FLAG-Histone H3 Dha4

FLAG-Histone H3 Dha4 was desalted into milli-Q water using a PD SpinTrap G-25 microspin column (Cytiva). Stock solutions of tetrahydroxydiboron (5.6 mg in 1000 µL) and sodium ascorbate (14.0 mg in 500 µL) in milli-Q water were prepared freshly directly before the reactions were conducted.

To 100 µL of a solution of FLAG-Histone H3 Dha4 in milli-Q water (100 µg, 5.64 nmol, 1.00 equiv) were added aliquots of the previously prepared stock solutions of sodium ascorbate (2 µL, 282 nmol, 50 equiv) and tetrahydroxydiboron (18 µL, 1.13 µmol, 200 equiv). The solution was briefly vortexed and incubated at room temperature for 21 hours before being analysed by LC-MS (>95% conversion).

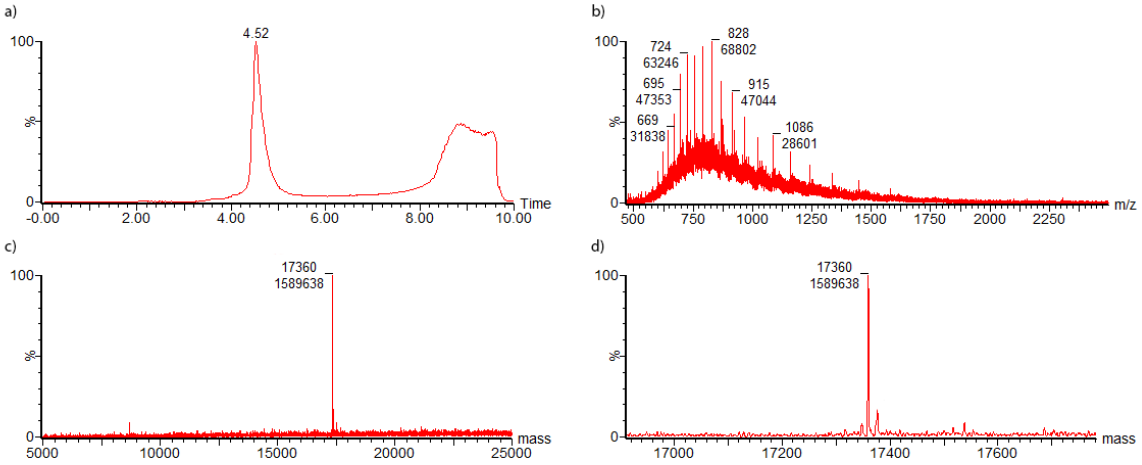


Figure 7.61: LC-MS analysis of cleaved FLAG-Histone H3 Dha4: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 17361 g/mol; observed mass: 17360 g/mol.

Histone H3 Dha9

Lyophilized Histone H3 Dha9 (500 µg) was dissolved in milli-Q water (500 µL). Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL) and sodium ascorbate (16.3 mg in 500 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. To the solution of Histone H3 Dha9 (500 µL, 500 µg, 32.9 nmol, 1.00 equiv) were added aliquots of the previously prepared stock solutions of sodium ascorbate (10 µL, 1.64 µmol, 50 equiv) and tetrahydroxydiboron (100 µL, 6.59 µmol, 200 equiv). The solution was briefly vortexed and incubated at room temperature for 23 hours before being analysed by LC-MS (>95% conversion).

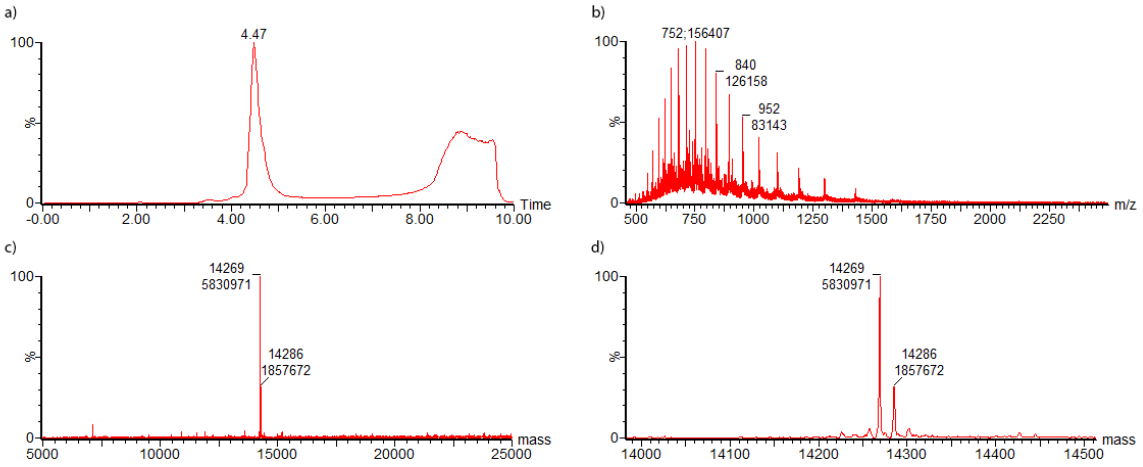


Figure 7.62: LC-MS analysis of cleaved Histone H3 Dha9: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 14270 g/mol; observed masses: 14269 g/mol, 14286 g/mol (*C*-terminal fragment + single methionine oxidation).

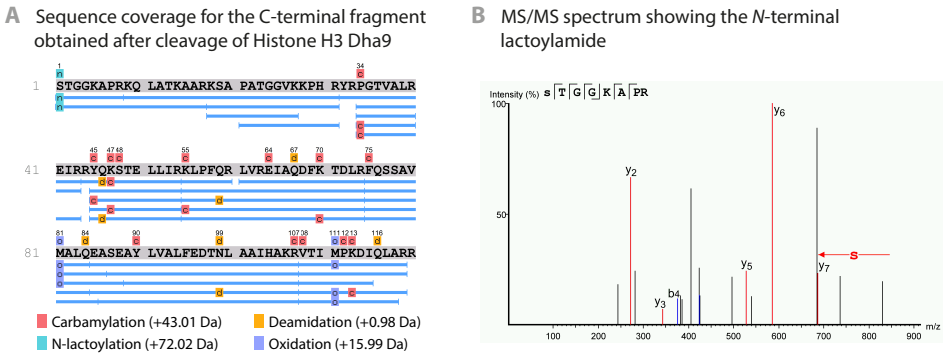


Figure 7.63: LC-MS/MS analysis of cleaved Histone H3 Dha9. (**A**) 95% sequence coverage were observed (1% FDR). (**B**) MS/MS spectrum of peptide fragment sTGGKAPR. s depicts the N-terminally lactoyl modified serine residue.

Histone H3 Dha10

Lyophilized Histone H3 Dha10 was dissolved in milli-Q water at 1.0 mg/mL. Stock solutions of tetrahydroxydiboron (5.9 mg in 1000 µL) and sodium ascorbate (13.0 mg in 1000 µL) in milli-Q water were prepared.

40 µL of Histone H3 Dha10 (40 µg, 2.63 nmol) were treated with sodium ascorbate (2 µL, 131 nmol, 50 equiv) followed by tetrahydroxydiboron (8 µL, 526 nmol, 200 equiv) and incubated at room temperature for 24 hours before being analysed by LC-MS (>95% conversion).

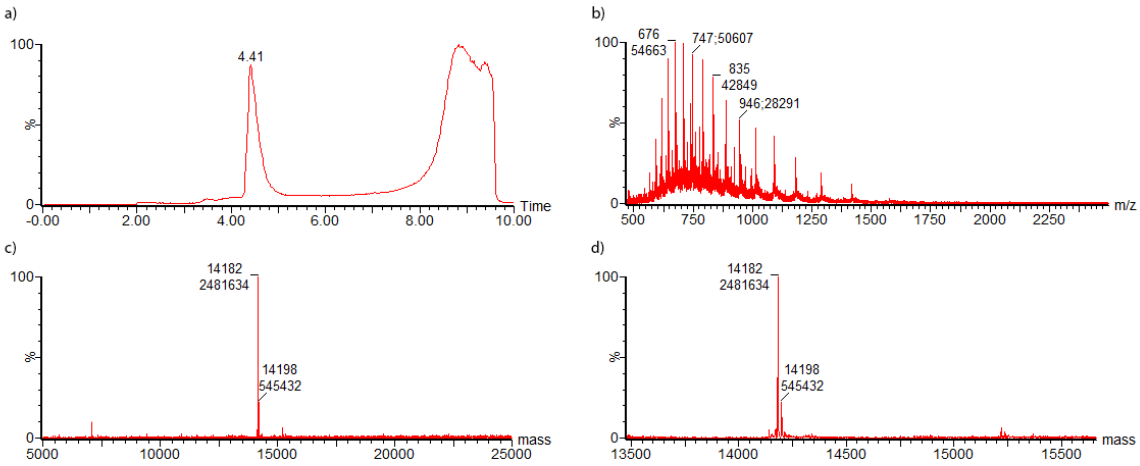


Figure 7.64: LC-MS analysis of cleaved Histone H3 Dha10: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 14182 g/mol; observed masses: 14182 g/mol, 14199 g/mol (*C*-terminal fragment + single methionine oxidation).

Histone H4 Dha16

Histone H4 Dha16 was buffer exchanged into sodium borate buffer (100 mM, pH 8.0) using a PD SpinTrap G-25 microspin column (Cytiva) and following the manufacturer's instructions. Stock solutions of tetrahydroxydiboron (5.8 mg in 1000 µL) and sodium ascorbate (13.3 mg in 500 µL) in milli Q water were prepared freshly directly before the reactions were conducted.

To 90 µL of a solution of Histone H4 Dha16 in borate buffer (1.0 mg/mL, 90 µg, 8.05 nmol, 1.00 equiv) were added aliquots of the previously prepared stock solutions of sodium ascorbate (3 µL, 403 nmol, 50 equiv) and tetrahydroxydiboron (25 µL, 1.61 µmol, 200 equiv). The solution was briefly vortexed and incubated at room temperature for 17 hours before being analysed by LC-MS (>95% conversion).

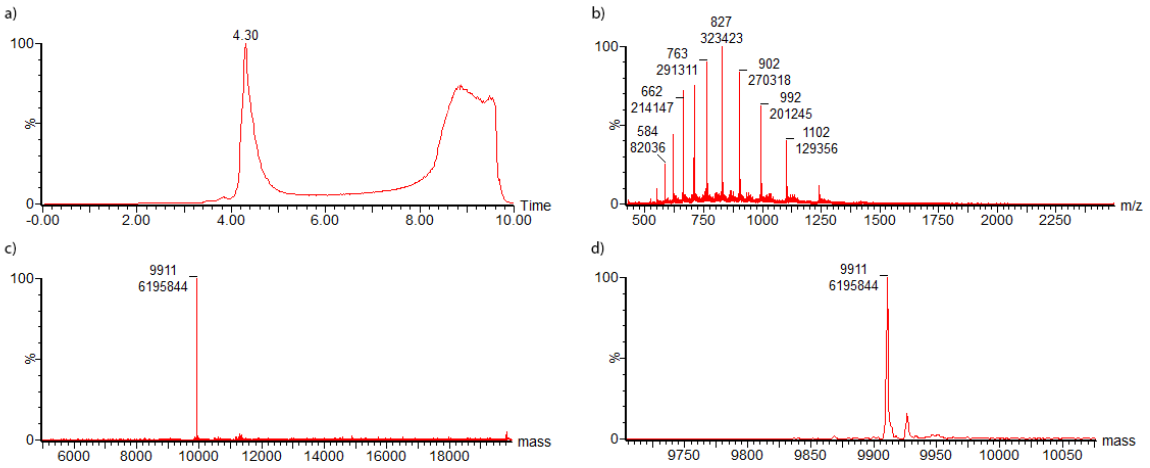


Figure 7.65: LC-MS analysis of cleaved Histone H4 Dha16: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 9912 g/mol; observed mass: 9911 g/mol.

Npβ Dha61

Npβ Dha61 was buffer exchanged into sodium borate buffer (100 mM, pH 8.0) using a PD SpinTrap G-25 microspin column (Cytiva) and following the manufacturer's instructions. The protein concentration was determined by A_{280} spectrophotometry (0.90 mg/mL).

Reaction at room temperature

Stock solutions of tetrahydroxydiboron (5.5 mg in 1000 µL) and sodium ascorbate (15.3 mg in 1000 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. To 60 µL of a solution of Npβ Dha61 (0.90 mg/mL, 54.0 µg, 2.57 nmol) were added aliquots of the previously prepared stock solutions of sodium ascorbate (2.5 µL, 193 nmol, 75 equiv) and tetrahydroxydiboron (12.5 µL, 771 nmol, 300 equiv). The solution was briefly vortexed and incubated at room temperature for 24 h before being analysed by LC-MS (21% conversion).

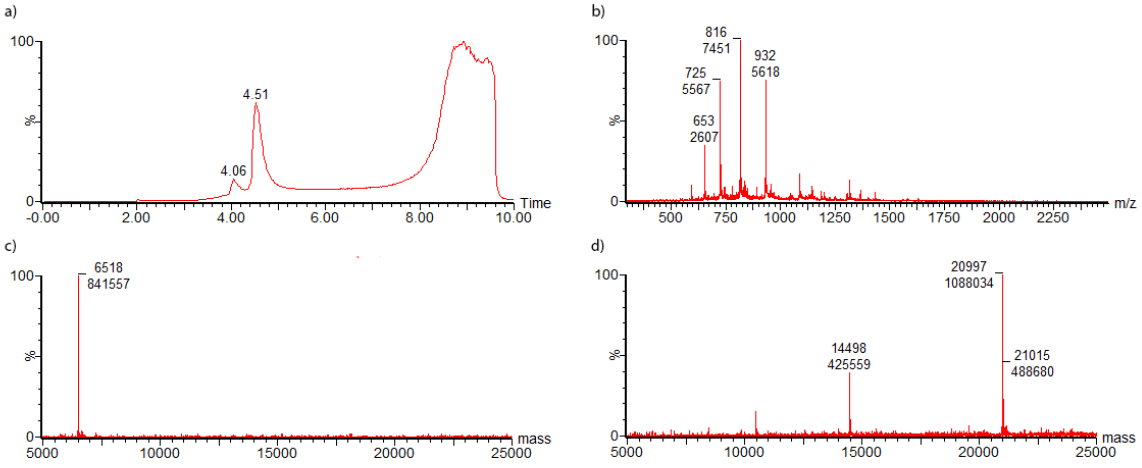


Figure 7.66: LC-MS analysis of the cleavage of Npβ Dha61 at room temperature: a) total ion chromatogram b) ion series of the *N*-terminal fragment c) deconvoluted spectrum of the *N*-terminal fragment d) deconvoluted spectrum of the *C*-terminal fragment; calculated mass for the *N*-terminal fragment: 6519 g/mol; observed mass: 6518 g/mol; calculated mass for the *C*-terminal fragment: 14499 g/mol; observed mass: 14498 g/mol.

Reaction at 45°C

Stock solutions of tetrahydroxydiboron (6.0 mg in 1000 µL) and sodium ascorbate (9.9 mg in 500 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. To 65 µL of a solution of Npβ Dha61 (0.90 mg/mL, 58.5 µg, 2.79 nmol) were added aliquots of the previously prepared stock solutions of sodium ascorbate (5 µL, 5 mM final concentration) and

tetrahydroxydiboron (30 µL, 20 mM final concentration). The mixture was briefly vortexed and shaken at 45°C and 500 rpm for 24 hours before being analysed by LC-MS (85% conversion).

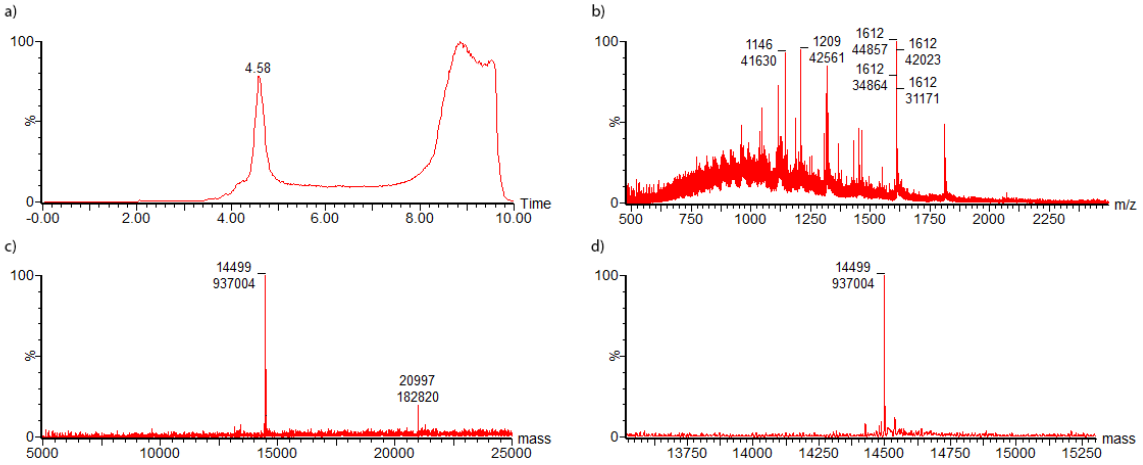


Figure 7.67: LC-MS analysis of the cleavage of Npβ Dha61 at 45°C: a) total ion chromatogram b) ion series of the *C*-terminal fragment c) deconvoluted spectrum of the *C*-terminal fragment d) magnification of the product peak; calculated mass for the *C*-terminal fragment: 14499 g/mol; observed mass: 14499 g/mol.

Reaction analysis by SDS-PAGE

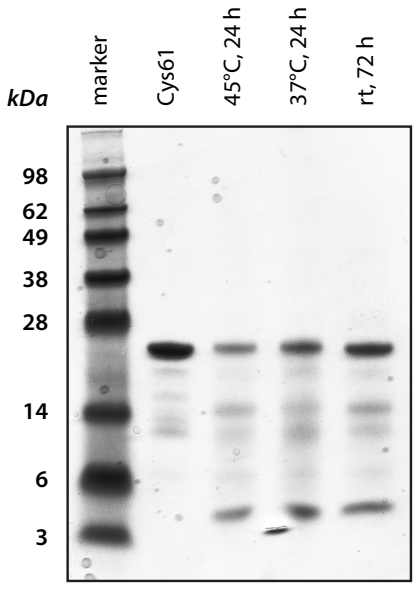


Figure 7.68: SDS-PAGE analysis of the cleavage of Npβ Dha61 (Coomassie staining, Invitrogen NuPAGE 10% *bis*-TRIS gel, MES buffer, 200 V, 75 min, 4°C, marker: SeeBlue Plus2 Pre-Stained Protein Standard, Cys61: uncleaved Npβ Cys61 standard)

20 µL of sample were mixed with 5 µL of Laemmli buffer (5×) and heated at 95°C for 5 minutes. 10 µL of the sample was analysed by gel electrophoresis (Invitrogen NuPAGE 10% *bis*-TRIS gel, MES buffer, 200 V, 75 min, 4°C). Protein bands were visualized by Coomassie staining. Stained gels were analysed by densitometric analysis in ImageJ 1.53f.[23] Conversions were calculated based on the area of cleavage products compared to the area of intact protein.

Table 7.3: SDS-PAGE analysis of the cleavage of Npβ Dha61 under various conditions.

Entry	Conditions	Cleavage [%]
1	45°C, 24 h, borate buffer (100 mM, pH 8.0)	52
2	37°C, 24 h, borate buffer (100 mM, pH 8.0)	45
3	room temperature, 3 days, borate buffer (100 mM, pH 8.0)	46

Cleavage was quantified by Coomassie staining and densitometric analysis in ImageJ 1.53f.[23]

preSUMO1 Dha51

preSUMO1 Dha51 was buffer exchanged into sodium borate buffer (100 mM, pH 8.0) or TRIS buffer (100 mM, pH 8.0) using PD SpinTrap G-25 microspin columns (Cytiva) and following the manufacturer's instructions. The protein concentration was determined by A_{280} spectrophotometry (borate buffer: 0.41 mg/mL, TRIS buffer: 0.30 mg/mL).

Stock solutions of tetrahydroxydiboron (5.4 mg in 1000 µL) and sodium ascorbate (14.9 mg in 1000 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. To 30 µL of a solution of preSUMO1 Dha51 in borate or TRIS buffer (9.0 µg – 12.3 µg, 0.72 nmol – 0.99 nmol) were added aliquots of the previously prepared stock solutions of sodium ascorbate (1.2 µL, final concentration 2.5 mM) and tetrahydroxydiboron (6.25 µL, final concentration 10 mM). The solution was briefly vortexed and incubated at room temperature for 48 hours or shaken at 37° and 500 rpm for 24 hours before being analysed by SDS-PAGE (up to 14% conversion).

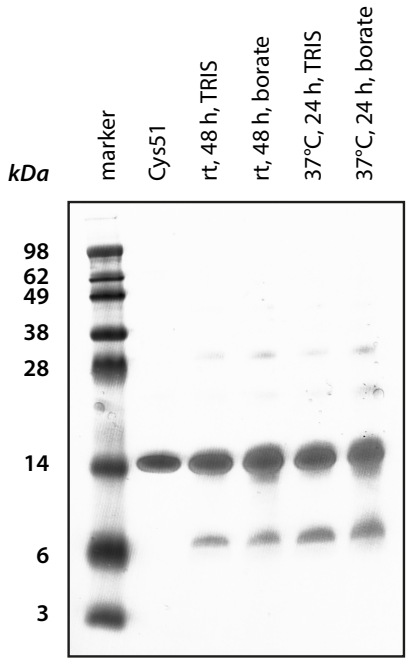


Figure 7.69: SDS-PAGE analysis of the cleavage of preSUMO1 Dha51 (Coomassie staining, Invitrogen NuPAGE 10% *bis*-TRIS gel, MES buffer, 200 V, 120 min, 4°C, marker: SeeBlue Plus2 Pre-Stained Protein Standard, Cys51: uncleaved preSUMO1 Cys51 standard)

20 µL of sample were mixed with 5 µL of Laemmli buffer (5×) and heated at 95°C for 5 minutes.

10 µL of the sample were analysed by gel electrophoresis (Invitrogen NuPAGE 10% *bis*-TRIS gel, MES buffer, 200 V, 120 min, 4°C). Protein bands were visualized by Coomassie staining. Stained gels were analysed by densitometric analysis in ImageJ 1.53f.[23] Conversions were calculated based on the area of cleavage products compared to the area of intact protein.

Table 7.4: SDS-PAGE analysis of the cleavage of preSUMO1 Dha51 under various conditions.

Entry	Conditions	Cleavage [%]
1	room temperature, 2 days, TRIS buffer (100 mM, pH 8.0)	14
2	room temperature, 2 days, borate buffer (100 mM, pH 8.0)	12
3	37°C, 24 h, TRIS buffer (100 mM, pH 8.0)	11
4	37°C, 24 h, borate buffer (100 mM, pH 8.0)	11

Cleavage was quantified by Coomassie staining and densitometric analysis in ImageJ 1.53f.[23]

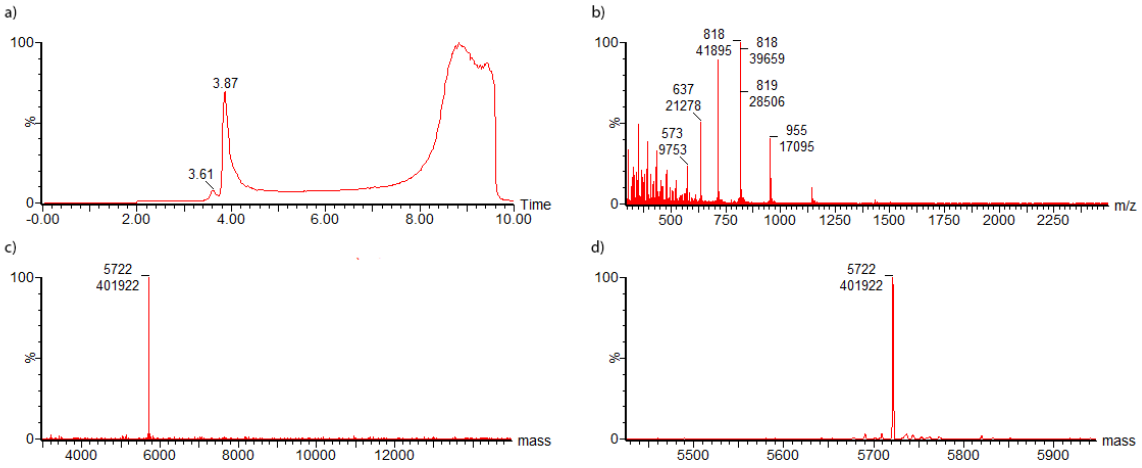


Figure 7.70: LC-MS analysis of the cleavage of preSUMO1 Dha51 at room temperature (24 h): a) total ion chromatogram b) ion series of the *N*-terminal fragment c) deconvoluted spectrum of the *N*-terminal fragment d) magnification of the product peak (*N*-terminal fragment); calculated mass for the *N*-terminal fragment: 5722 g/mol; observed mass: 5722 g/mol.

preSUMO1 Dha51 Dha97

Stock solutions of tetrahydroxydiboron (5.6 mg in 1000 µL) and sodium ascorbate (6.2 mg in 500 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. To 60 µL of a solution of preSUMO1 Dha51 Dha97 in TRIS buffer (100 mM, pH 8.0) (0.39 mg/mL, 23.4 µg, 1.88 nmol, 1.00 equiv) were added aliquots of the previously prepared stock solutions of sodium ascorbate (3.0 µL, 188 nmol, 100 equiv) and tetrahydroxydiboron (12 µL, 752 nmol, 400 equiv). The mixture was vortexed and another portion of reagents was added after 3 hours. The reaction mixture was incubated at room temperature for a total of 18 h and analysed by LC-MS (>90% conversion).

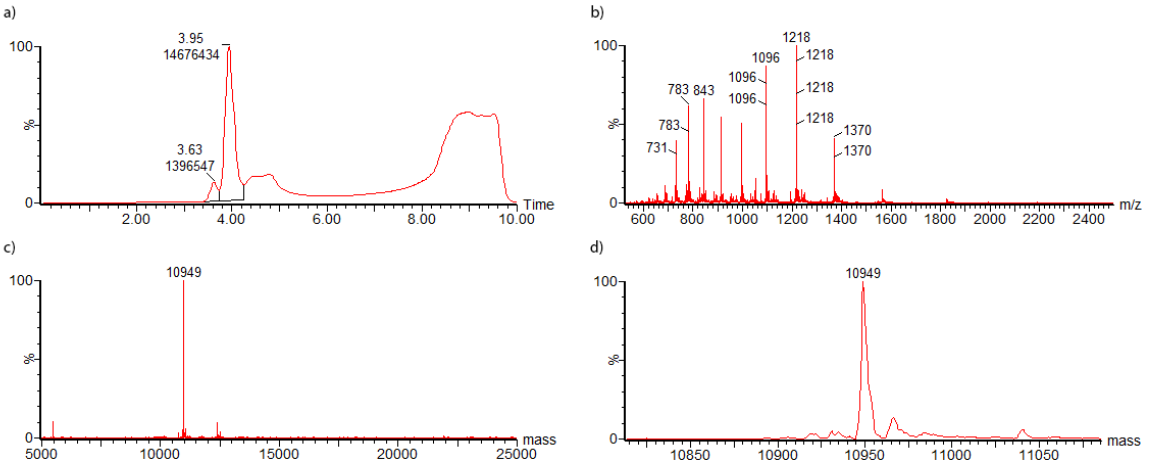


Figure 7.71: LC-MS analysis of the cleavage of preSUMO1 Dha51 Dha97: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 10950 g/mol; observed mass: 10949 g/mol.

Reaction analysis by SDS-PAGE

10 µL of sample were mixed with 2.5 µL of Laemmli buffer (5×) and heated at 95°C for 5 minutes. 10 µL of the sample were analysed by gel electrophoresis (Invitrogen NuPAGE 10% bis-TRIS gel, MES buffer, 200 V, 50 min, room temperature). Protein bands were visualized by Coomassie staining. Stained gels were analysed by densitometric analysis in ImageJ 1.53f.[23] Cleavage at site 51 was calculated based on the area of cleavage products compared to the area of intact protein. 22% of protein was cleaved at site 51.

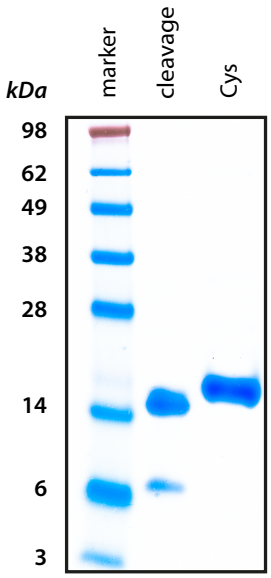


Figure 7.72: SDS-PAGE analysis of the cleavage of preSUMO1 Dha51 Dha97 (Coomassie staining, Invitrogen NuPAGE 10% *bis*-TRIS gel, MES buffer, 200 V, 50 min, room temperature, marker: SeeBlue Plus2 Pre-Stained Protein Standard, Cys: uncleaved preSUMO1 Cys51 Cys97 standard)

20 µL of sample were mixed with 5 µL of Laemmli buffer (5×) and heated at 95°C for 5 minutes. 10 µL of the sample was analysed by gel electrophoresis (Invitrogen NuPAGE 10% *bis*-TRIS gel, MES buffer, 200 V, 75 min, 4°C). Protein bands were visualized by Coomassie staining. Stained gels were analysed by densitometric analysis in ImageJ 1.53f.[23] Conversions were calculated based on the area of cleavage products compared to the area of intact protein.

Subsequent functionalization: Thia-Michael addition

50 µL of the reaction mixture were treated with 0.5 µL of BME (1%, v/v). The mixture was shaken (500 rpm) at 37°C for 3 h before being analysed by LC-MS. 81% conversion to the thia-Michael adduct was observed.

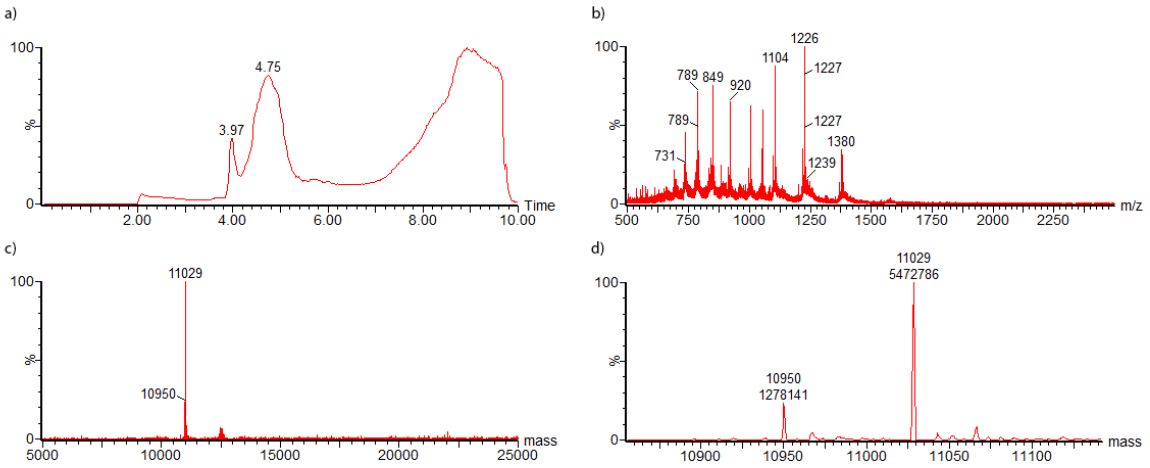


Figure 7.73: LC-MS analysis of thia-Michael addition of BME to cleaved preSUMO1 Dha51 Dha97: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated mass: 11028 g/mol; observed mass: 11029 g/mol.

Annexin V Dha316

Annexin V Dha316 was buffer exchanged into sodium borate buffer (100 mM, pH 8.0) using a PD SpinTrap G-25 microspin column (Cytiva) and following the manufacturer's instructions. The protein concentration was determined by A_{280} spectrophotometry (0.11 mg/mL).

Stock solutions of tetrahydroxydiboron (6.0 mg in 1000 µL) and sodium ascorbate (9.9 mg in 500 µL) in milli-Q water were prepared freshly directly before the reactions were conducted. To 65 µL of a solution of Annexin V Dha316 (0.11 mg/mL, 7.15 µg, 0.20 nmol) were added aliquots of the previously prepared stock solutions of sodium ascorbate (5 µL, 5 mM final concentration) and tetrahydroxydiboron (30 µL, 20 mM final concentration). The mixture was briefly vortexed and shaken at 45°C and 500 rpm for up to 24 hours before being analysed by LC-MS (39% after 3 hours, 52% after 24 hours).

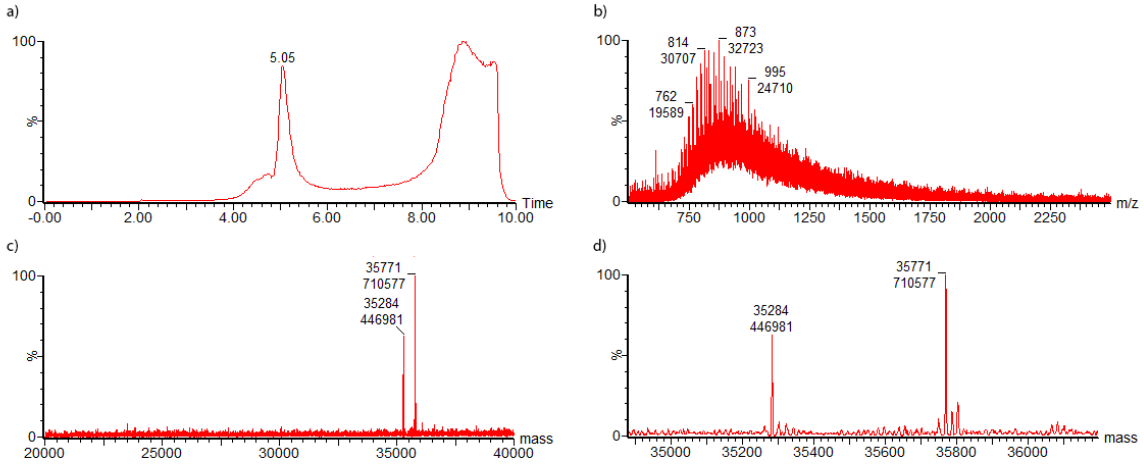


Figure 7.74: LC-MS analysis of the cleavage of Annexin V Dha316 after 3 h at 45°C: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the product peak; calculated mass: 35285 g/mol; observed mass: 35284 g/mol.

7.10.4 Isotope Pattern Analysis

Isotope pattern analysis was conducted for the *N*-terminal fragments observed for the cleavage of Npβ Dha61 and preSUMO1 Dha51 in order to distinguish between an unmodified and amidated *C*-terminus. In both cases, three different charge states (Npβ: 7+, 8+, 9+; preSUMO1: 6+, 7+, 8+) were analysed. Isotope patterns were calculated for the *N*-terminal cleavage fragments containing either unmodified and amidated *C*-termini using the isotope modelling algorithm included in MassLynx 4.1 (Waters). Calculated isotope patterns were aligned with the observed isotope patterns by their relative intensities and the deviation of the calculated masses for the unmodified (OH) and amidated (NH_2) *C*-termini from the observed spectrum was calculated. An average deviation was calculated for each of the analysed charge states. Isotopic pattern analysis confirmed the presence of α-amidated *C*-termini for both cases.

Table 7.5: Isotope pattern analysis for the *N*-terminal fragments of Npβ Dha61 and preSUMO1 Dha51 confirmed the presence of α-amidated *C*-termini.

Entry	Protein	Charge State	Deviation OH [Da]	Deviation NH_2 [Da]
1	Npβ Dha61	7+	0.1447	0.0054
2		8+	0.1285	0.0059
3		9+	0.1145	0.0052
4	preSUMO1 Dha51	6+	0.1672	0.0059
5		7+	0.1465	0.0067
6		8+	0.1317	0.0086

Deviation is reported as the average deviation of the isotopologue *m/z* observed for the respective charge state.

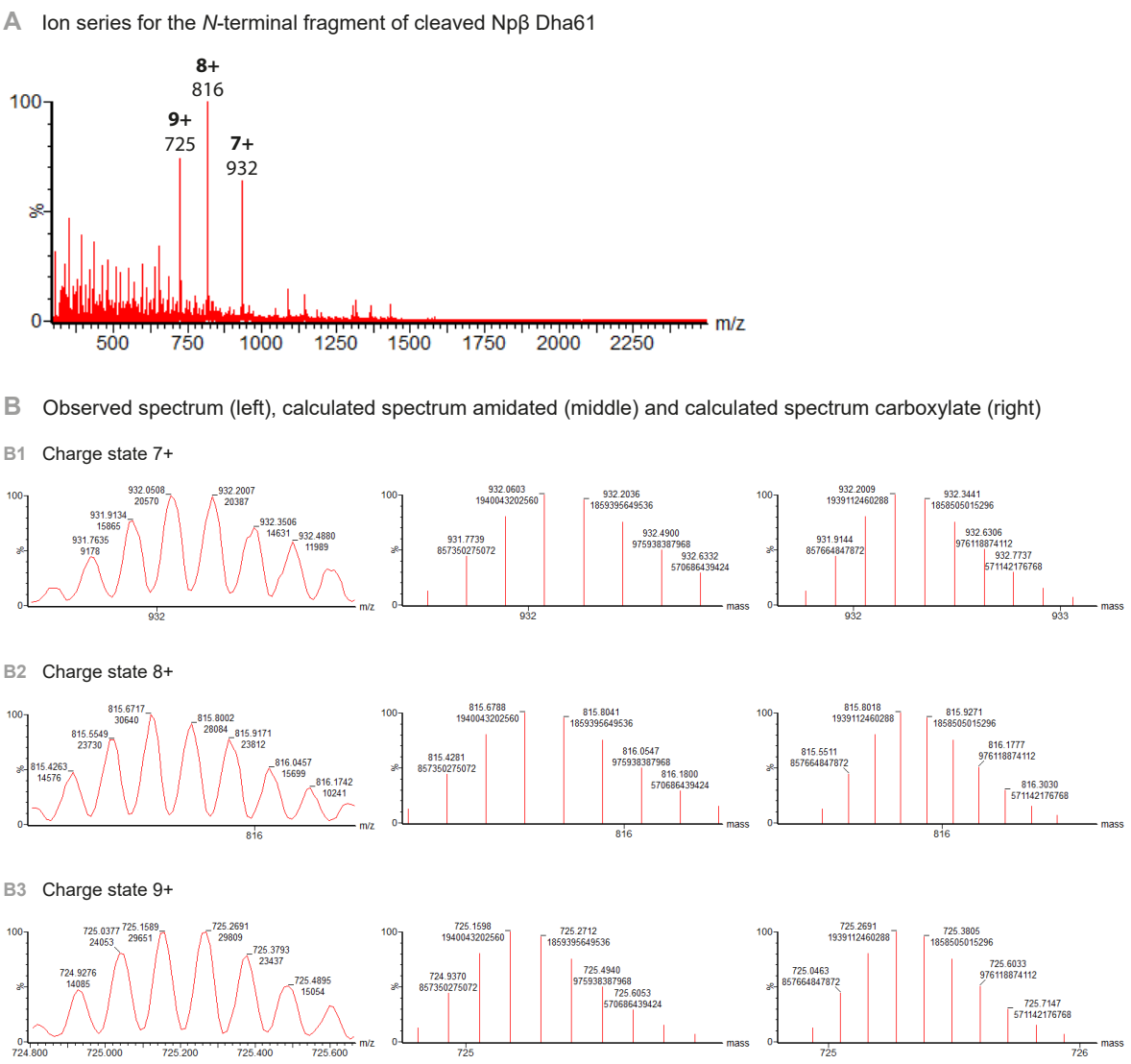


Figure 7.75: Isotope pattern analysis for the *N*-terminal fragment obtained after cleavage of Npβ Dha61. (**A**) Ion series and corresponding charge states (**B**) Observed spectra (*left*) and calculated spectra for the amidated ($CONH_2$) (*middle*) or unmodified (COOH) species (*right*) of the analysed charge states.

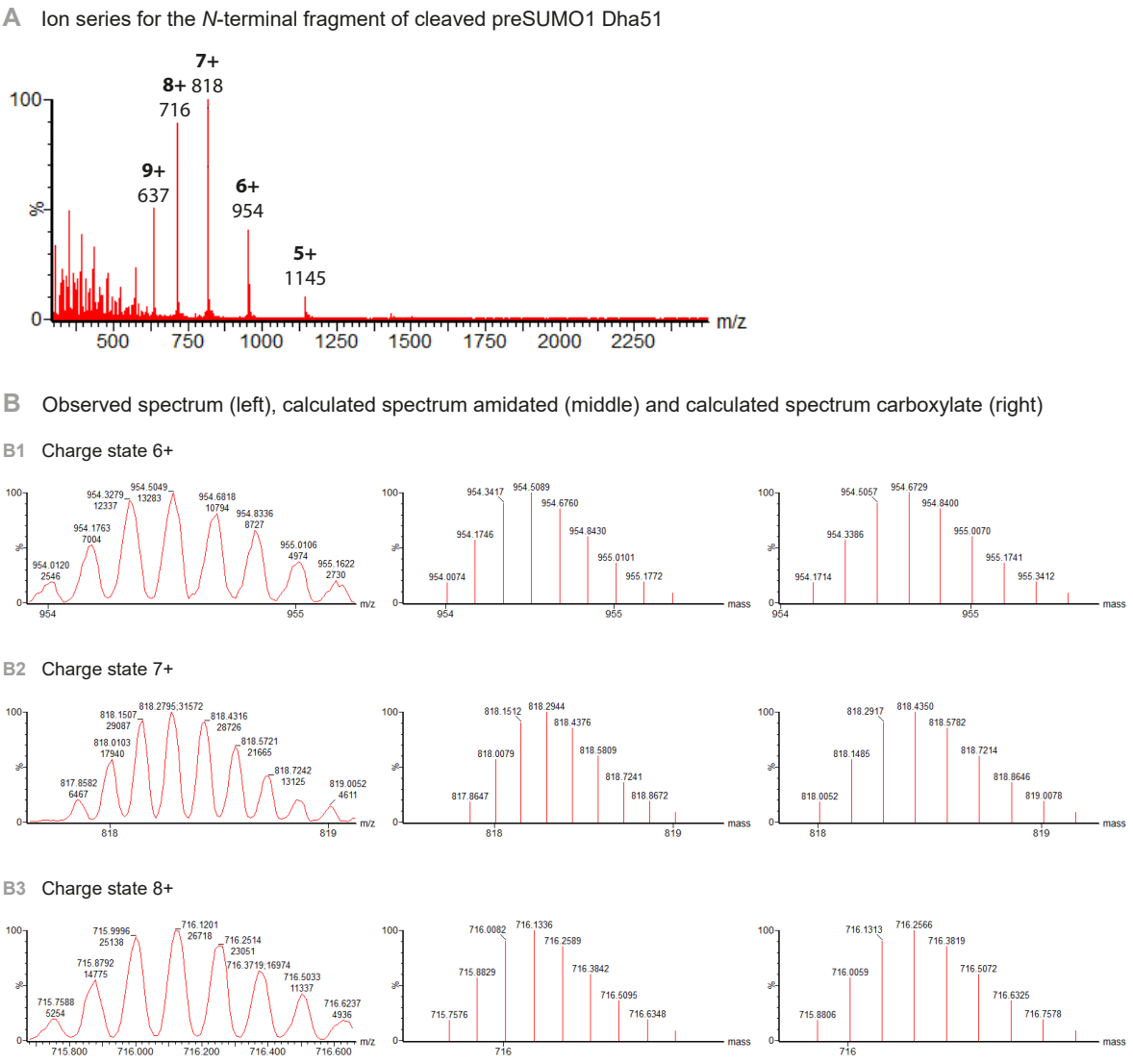


Figure 7.76: Isotope pattern analysis for the *N*-terminal fragment obtained after cleavage of preSUMO1 Dha51. (**A**) Ion series and corresponding charge states (**B**) Observed spectra (*left*) and calculated spectra for the amidated ($CONH_2$) (*middle*) or unmodified (COOH) species (*right*) of the analysed charge states.

7.11 Protein Sequencing

ArgC digestion

50 µg of protein sample were dissolved in freshly prepared denaturing buffer (50 µL, 4 M urea, 100 mM ammonium bicarbonate) and incubated at room temperature for 10 min. The sample was diluted to a final concentration of 2 M urea using a 100 mM ammonium bicarbonate solution. $CaCl_2$ was added to a final concentration of 2 mM. Activation solution (50 mM DTT, 5 mM EDTA in water) was added (1:10, v/v) before ArgC (1 µg, sequencing grade, Promega, Catalogue No. V1881) was added. The sample was shaken at 37°C (800 rpm) for 20 hours. The reaction was quenched by addition of formic acid (5%, v/v) and centrifuged (30 min, 14,000 rpm, 4°C). The sample was desalted using a C_{18} resin containing tip and lyophilized in a protein LoBind microcentrifuge tube (Eppendorf, Catalogue No. 0030108116).

Data Acquisition

Resulting peptides were separated using a nano-flow reversed-phase liquid chromatography Ultimate 3000 UHPLC system (Thermo Fisher Scientific) and analysed using a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fischer Scientific). Data acquisition was performed by the Advanced Proteomics Facility, University of Oxford.

Data Analysis

Data analysis was performed with PEAKS Studio 8.5 (Bioinformatics Solutions Inc.). The data was searched against the given protein sequence. The precursor mass tolerance was set to 15 ppm and the fragment mass tolerance was set to 0.5 Da. A maximum number of 3 missed cleavages and non-specific cleavage at one end of the peptide were specified. Methionine oxidation (+15.99 Da), deamidation (+0.98 Da; asparagine, glutamine) and carbamylation (+43.01 Da; lysine) were set as variable post-translational modifications. Additionally, the modification of interest was included in the search. A maximum of 4 variable modifications was defined and a FDR of 1% on peptide level was applied.

7.12 Refolding and CD Spectroscopy of Selected Proteins

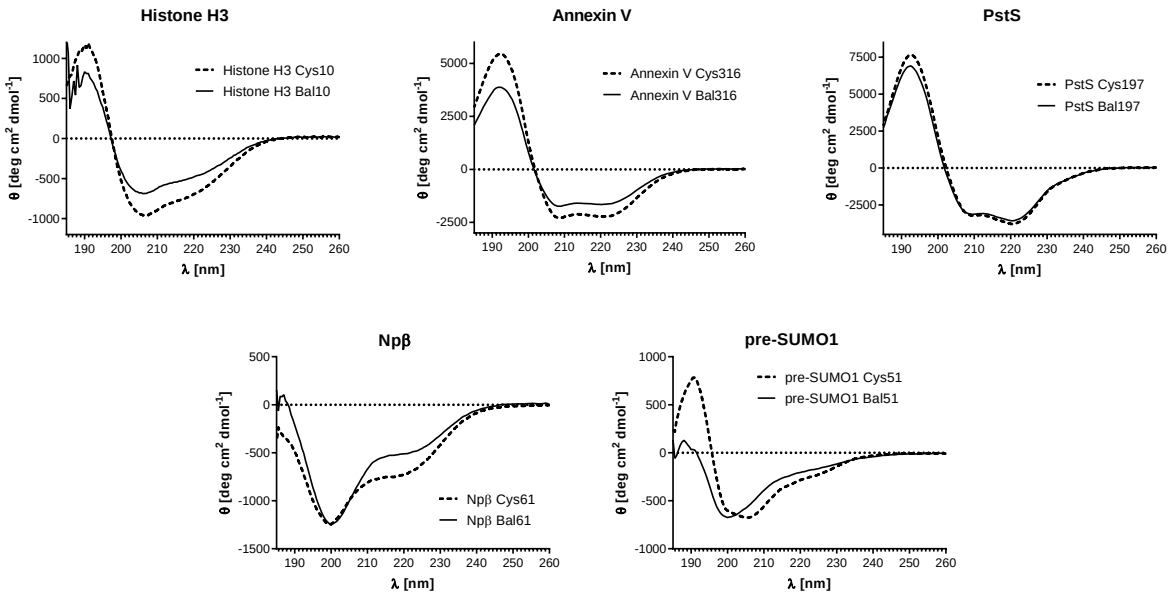


Figure 7.77: CD spectra of borylated proteins PstS Bal197, Annexin V Bal316, Npβ Bal61, pre-SUMO1 Bal51 and Histone H3 Bal10 in comparison with their respective cysteine mutants.

7.12.1 General Procedure

A solution of the desired protein (0.1 – 0.5 mg/mL) was dialysed against refolding buffer and/or CD buffer (50 – 250 mM KF or NaF, 20 – 25 mM NaP_i). Experiments were performed using 300 µL of buffered protein solution in a 1 mm quartz cuvette. Spectra were collected from 185 to 260 nm (0.5 nm step, 1 s per point) as a smoothed average of three scans. CD buffer solution was used as a blank sample.

7.12.2 Histone H3

A sample of Histone H3 Bal was dialysed against refolding buffer (250 mM KF, 25 mM NaP_i, pH 8.4). The concentration of the sample was determined by A_{280} spectrophotometry (0.32 mg/mL). A sample of Histone Cys10 in refolding buffer (250 mM KF, 25 mM NaP_i, pH 8.4) was prepared at the same concentration.

7.12.3 Annexin V

Borylated Annexin V in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) was dialysed against refolding buffer (20 mM TRIS·HCl, 100 mM NaCl, pH 7.4) at 4°C overnight. The solution was then dialysed against CD buffer (20 mM NaP_i, 50 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry (0.13 mg/mL).

Annexin V wt was dialysed against CD buffer (20 mM NaP_i, 50 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry and adjusted to 0.13 mg/mL by dilution with CD buffer.

7.12.4 PstS

Borylated PstS in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 8.0) was dialysed against refolding buffer (20 mM TRIS·HCl, 100 mM NaCl, pH 7.4) at 4°C overnight. The solution was then dialysed against CD buffer (20 mM NaP_i, 50 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry (0.19 mg/mL).

PstS Cys197 was dialysed against CD buffer (20 mM NaP_i, 50 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry and adjusted to 0.19 mg/mL by dilution with CD buffer.

7.12.5 Npβ

Borylated Npβ in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) was dialysed against refolding buffer (20 mM TRIS·HCl, 100 mM NaCl, pH 7.4) at 4°C overnight. The solution was then dialysed against CD buffer (20 mM NaP_i, 100 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry (0.32 mg/mL).

Npβ Cys61 was reduced in NaP_i buffer (50 mM, 100 mM DTT, pH 7.8) at 37°C for 30 min. The protein solution was purified PD Minitrap G-25 and eluted into CD buffer. The concentration was determined by A_{280} spectrophotometry and adjusted to 0.32 mg/mL by dilution with CD buffer.

7.12.6 pre-SUMO1

Borylated pre-SUMO1 in NaP_i buffer (100 mM, 3 M Gdn·HCl, pH 7.0) was dialysed against refolding buffer (20 mM TRIS·HCl, 100 mM NaCl, pH 7.4) at 4°C overnight. The solution was then dialysed against CD buffer (20 mM NaP_i, 100 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry (0.29 mg/mL).

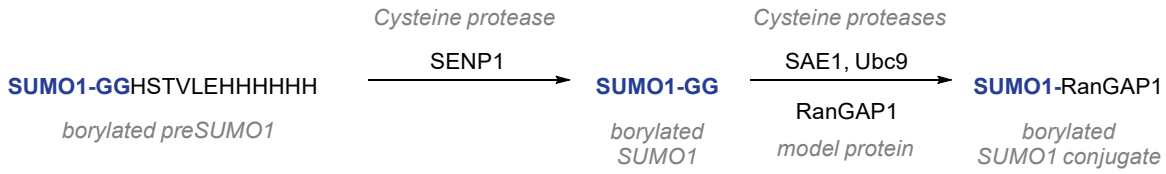
Pre-SUMO1 Cys51 was dialysed against CD buffer (20 mM NaP_i, 50 mM NaF, pH 7.4) five times. The concentration was determined by A_{280} spectrophotometry and adjusted to 0.29 mg/mL by dilution with CD buffer.

7.13 Determination of Protein Melting Curves

Melting curves were recorded on a Prometheus NT.48 differential scanning fluorimeter (NanoTemper Technologies, Munich, Germany). Samples were loaded into high sensitivity quartz capillaries. Unfolding was detected during heating using a linear temperature gradient (20°C to 95°C, 1°C/min) and an excitation power of 50% (100% for Npβ). Melting curves were recorded at three different concentrations in phosphate buffer (20 mM NaP_i, 50 mM NaF, pH 7.4). The data were analysed using PR.Stability Analysis v1.0.2 software (NanoTemper Technologies, Munich, Germany) and GraphPad Prism 8.0.0 (GraphPad Software Inc., San Diego, USA). Melting points (T_M) were calculated as the average of the melting points determined at different concentrations.

7.14 *In vitro* SUMOylation

7.14.1 SUMOylation using SUMO1 Bal51



Scheme 7.2: pre-SUMO1 Bal51 was used in an *in vitro* SUMOylation reaction for the SUMOylation of model substrate RanGAP1 fragment.

pre-SUMO1 Maturation

200 µL (0.67 mg/mL) of a solution of pre-SUMO1 Bal51 in TRIS buffer (100 mM TRIS base, pH 7.0) were incubated with 10 µL of SENP1 (0.14 mg/mL) at 37°C for 4 h. 20 µL of fresh Ni-NTA resin was washed with water (2 × 500 µL) and added to the reaction mixture. The mixture was incubated for 15 min and the resin was removed by filtration. The filtrate was concentrated to 50 µL using a VivaSpin concentrator (MWCO 5,000 Da) and used directly for *in vitro* SUMOylation.

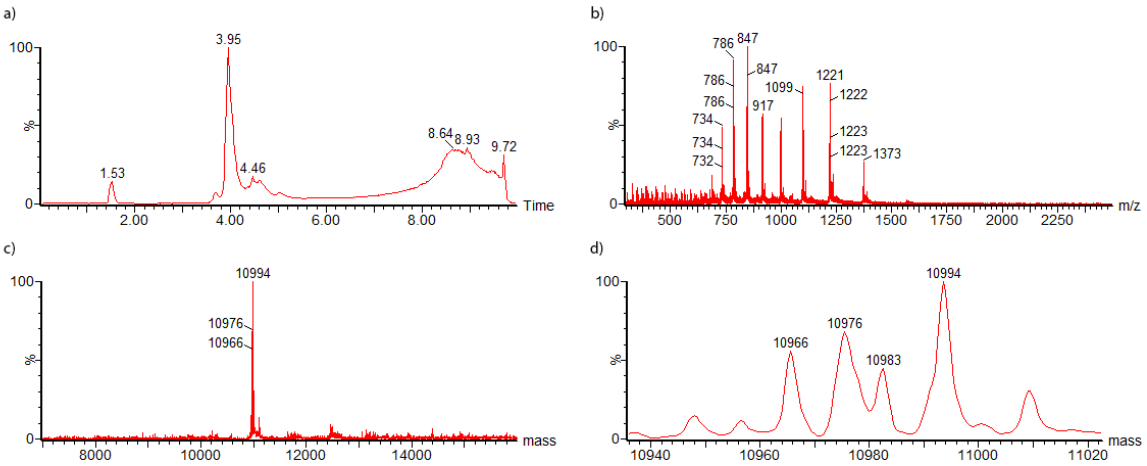


Figure 7.78: LC-MS analysis of pre-SUMO1 Bal51 after treatment with SENP1 yielding matured SUMO1 Bal51: a) total ion chromatogram b) ion series c) deconvoluted spectrum d) magnification of the major peak; calculated masses: 10997 g/mol (Bal), 10979 g/mol (Bal−H_2O), 10961 g/mol (Bal−$2H_2O$); observed masses: 10997 g/mol (Bal), 10976 g/mol (Bal−H_2O), 10966 g/mol (Bal−$2H_2O$).

SUMOylation of a Model Protein

In vitro SUMOylation of RanGAP1 fragment was conducted using an *in vitro* SUMOylation assay kit (Abcam, Catalogue No. ab139470) on a 20 µL scale using matured SUMO1 Bal51 and following the manufacturer's general protocol. A control experiment was conducted in the absence of matured SUMO1 Bal51.

RanGAP1 SUMOylated with SUMO1 Bal51 was detected via Western Blot using the supplied primary antibody and a polyclonal goat anti-rabbit IgG alkaline phosphatase secondary antibody. Controls and reaction were analysed via SDS-PAGE (10% *bis*-TRIS gel, 150 V, 45 min, MES buffer). The proteins were transferred to a nitrocellulose membrane. The membrane was blocked for 1 h in TBS-T (pH 7.5) supplemented with 5% BSA. The membrane was incubated with the primary rabbit anti-SUMO1 polyclonal antibody for 1 h at room temperature and washed three times for 5 min before the secondary goat anti-rabbit polyclonal antibody was added. The membrane was incubated for another hour, washed three times and SUMO1 was visualized using NBT/BCIP substrate solution (ThermoFisher).

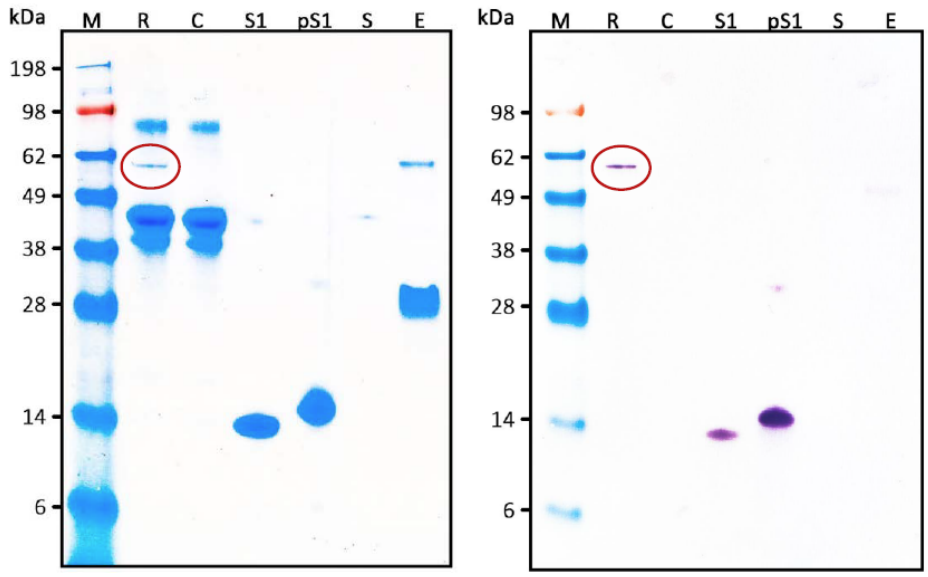


Figure 7.79: Coomassie-stained gel (left) and Western Blot (right) of *in vitro* SUMOylation of RanGAP1 using SUMO1 Bal51. M = marker, R = reaction, C = control (in absence of SUMO1 Bal51), S1 = SUMO1 Bal51, pS1 = pre-SUMO1 Bal51, S = substrate (RanGAP1 fragment), E = SENP1.

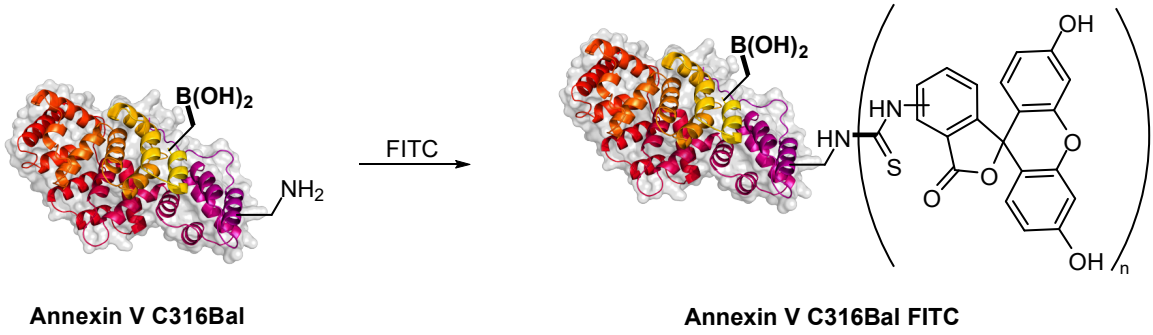
7.14.2 SUMOylation study using chemically cleaved SUMO1

SUMOylation was conducted using a commercially available SUMOylation assay kit (Abcam, Catalogue No. ab139470) following the manufacturer's instructions. 5 µL (approx. 0.3 mg/mL) of a solution of chemically (diboron-promoted) cleaved preSUMO1 in TRIS buffer (100 mM, pH 8.0) were treated with milli-Q water (10 µL), 10× SUMOylation buffer (2 µL), 20× Mg-ATP (1 µL), 20× SUMO E1 (1 µL) and 20× RanGAP1 (1 µL). Control reactions containing no Mg-ATP or 20×SUMO1 (1 µL) were conducted. The amount of milli-Q water used was adjusted to achieve a total reaction volume of 20 µL. The reaction mixtures were incubated at 37°C for 60 minutes before being quenched by addition of 5× Laemmli buffer (5 µL) and heating to 95°C (10 min).

The reactions were analysed by SDS-PAGE (MES running buffer, 10% *bis*-TRIS precast gel, 200 V, 40 min, room temperature) and Western blot. Western blot analysis was conducted using the primary rabbit anti-SUMO1 polyclonal antibody supplied with the SUMOylation kit (1:1000 dilution) and a secondary goat anti-rabbit IgG–alkaline phosphatase fusion (1:1000 dilution; Sigma Aldrich, Catalogue No. A3687). NBT/BCIP substrate solution (Thermo Scientific, Catalogue No. 34042) was used to visualize proteins. Blocking was conducted using 5% (w/v) skimmed milk powder in TBS-T buffer. In agreement with a *C*-terminally amidated SUMO1 mutant, no SUMOylation was detected for chemically cleaved SUMO1, while SUMOylation was detected for the positive control.

7.15 Cell-based Binding Studies

7.15.1 Annexin V Binding Study



Scheme 7.3: Fluorescent labelling of Annexin V Bal316 with fluorescein isothiocyanate (FITC)

FITC-Labelling of Annexin V Bal316

A solution of Annexin V Bal316 in phosphate buffer (500 µL, 0.15 mg/mL, 2.09 nmol, 1.00 equiv) was exchanged into sodium bicarbonate buffer (100 mM, pH 9.4) and concentrated to 200 µL using VivaSpin 500 concentrators (5,000 Da MWCO). 20 µL (81.6 µg, 209 nmol, 100 equiv) of a freshly prepared solution of fluorescein 5(6)-isothiocyanate (FITC) (4.1 mg) in DMSO (1000 µL) was added to the protein solution and the mixture was vortexed briefly. The mixture was shaken (400 rpm) in the dark at 37°C for 2 h and purified using a PD Minitrap™ G-25 size exclusion column (GE Healthcare) previously equilibrated with HEPES buffer (10 mM HEPES, 140 mM NaCl, pH 7.4) according to manufacturer instructions. The purified solution of FITC-labelled Annexin V Bal316 was concentrated using VivaSpin 500 concentrators (5,000 Da MWCO) and stored on ice until used.

The dye-to-protein ratio was determined spectrophotometrically on a NanoDrop 8000 spectrophotometer (Thermo Scientific) measuring at $\lambda = 280$ nm and $\lambda = 494$ nm. A dye-to-protein ratio of 0.95 was detected.

Evaluation of Annexin V Bal316 FITC in Flow Cytometry

Jurkat cells (clone E6.1) (5×10^5 cells/mL, 10 mL) were incubated with etoposide at a final concentration of 25 µM (addition of 10 µL of 25 mM stock in DMSO) for 390 min at 37°C. Approximately 100,000 cells were pelleted and resuspended in 100 µL of Annexin V binding buffer (10 mM HEPES, 150 mM NaCl, 2.5 mM $CaCl_2$, pH 7.4). 45 µL Annexin V Bal316 FITC or 10 µL of commercial Annexin V-FITC (Miltenyi Biotech) were added and the cells were incubated at room temperature for 15 min and stored on ice before being analysed by flow cytometry.

Flow cytometry was performed on a BD FACSCalibur™ cell analyser using BD FACSDiva software. A minimum of 10,000 cells per sample was analysed using the 488 nm laser and a 530/30 nm bandpass filter. The data were analysed using FlowJo Version 10 software (FlowJo LLC, USA).

7.15.2 mCherry Binding Study

Determination of Absorption Maxima

Absorption spectra of different mCherry mutants were measured on a BMG Labtech SPECTROstar Nano (full spectrum, 350 – 900 nm, 1 nm interval) using 200 µL protein solution in a clear 96-well plate at a concentration of approximately 10 µM. Measured OD values were used to calculate the relative absorbance of the respective sample.

The absorbance maximum was determined to be at 586 nm for all three mutants – mCherry S131C, mCherry S131Sulfonium and mCherry S131Bal.

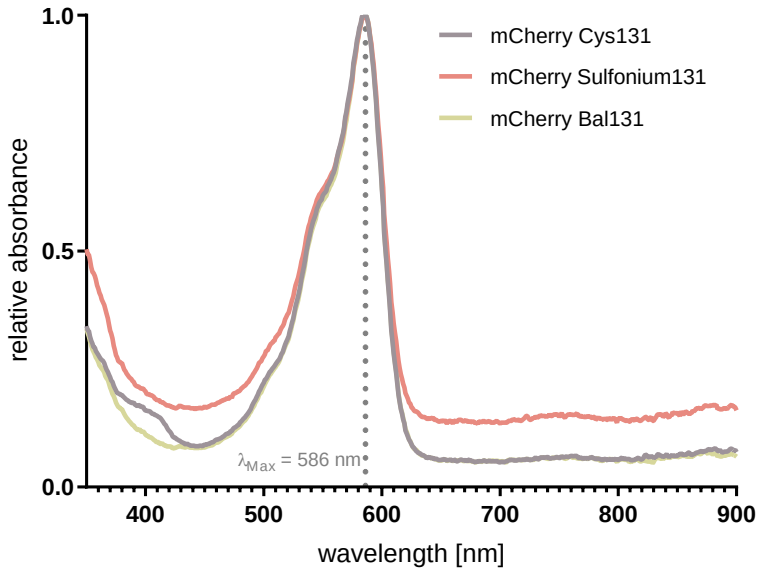


Figure 7.80: Absorbance spectra of mCherry-C131, mCherry-Sulfonium131 and mCherry-Bal131 show unaltered absorbance properties with a maximum at 586 nm

Determination of Cell-Surface Binding by Flow Cytometry

To 100 µL of CHO wt cells (approx. 10^6 cells) in FACS buffer (Dulbecco's phosphate-buffered saline, pH 8.0, 2% FBS) were added 300 µL of mCherry mutant (0.28 mg/mL, 10 µM) or 300 µL of FACS buffer (control). The cells where shaken on ice for 20 min at 300 rpm. The samples were centrifuged at 400 g for 3 min at 4°C. The liquid was removed and the cell pellet was resuspended in 1000 µL of FACS buffer, centrifuged at 400 g for 3 min at 4°C and the

buffer was removed. The washing step was repeated once before the cell pellet was suspended in 400 µL of FACS buffer.

Flow cytometry was performed on a BD LSRFortessa™ X-20 cell analyser using BD FACSDiva software. A minimum of 10,000 cells per sample was analysed using the 561 nm laser and a 610/20 nm bandpass filter. The 640 nm laser in combination with a 780/60 nm bandpass filter was used as a reference channel. The data was analysed using FlowJo Version 10 software. Histograms depict mCherry fluorescence versus cell count, dot plots depict mCherry fluorescence versus the reference channel.

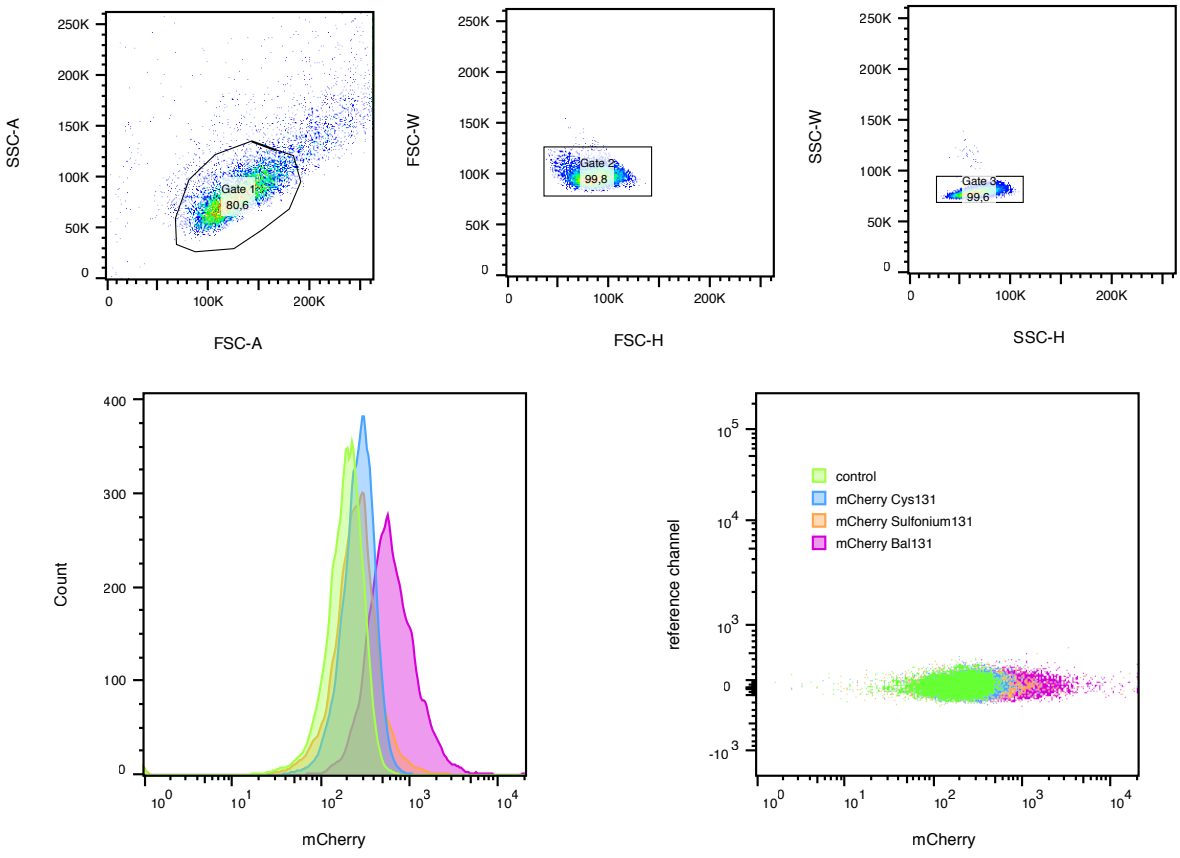


Figure 7.81: Determination of mCherry cell-surface binding by flow cytometry.

7.16 PstS Competition Assay

A stock solution of Phosphate Sensor (ThermoFisher Scientific, USA) (1.0 µM) and sodium phosphate (1.0 µM) in TRIS buffer (20 mM, pH 7.6) was prepared. 10 µL of this stock solution were mixed with 10 µL of a solution of PstS Bal197 in TRIS buffer in a black clear bottom 384-well plate in triplicates. Final PstS Bal197 concentrations of 2.28 nM to 37.4 µM were investigated. Measurements were captured on a BMG Labtech ClarioSTAR plate reader (BMG Labtech, Germany) with excitation at 430 nm (8 nm bandwidth) and emission at 450 nm (8 nm bandwidth).

The data were fitted to a four-parameter dose response curve using GraphPad Prism 8.0 (GraphPad Software, USA).

$$Y = Bottom + (Top - Bottom)/(1 + ((X^{HillSlope})/(IC_{50}^{HillSlope})))$$

7.17 Observation of Diol Binding via NMR

7.17.1 K_D Determination on Protein

Histone H3 Dha9 or H3 Dha10 was borylated following the standard procedure. The mixture was purified via dialysis into H_2O and lyophilized before being resuspended in NMR binding buffer buffer (NaP_i 40 mM, pH 7.0, 5 M urea, 10% D_2O). The conversion to the Bal mutant was determined by LC-MS and the amount of diol to be added was determined based on the concentration of the Bal mutant. 10 equiv of the fluorinated diol were added from a stock solution (200 mM in DMSO) before transfer to an NMR tube. The mixture was analysed via ^{19}F-NMR on a Bruker Avance III HD 600 MHz NMR spectrometer equipped with a Prodigy N2 broadband cryoprobe. The K_D was determined by integration of the peaks for bound and free diol and experiments were conducted in triplicates.

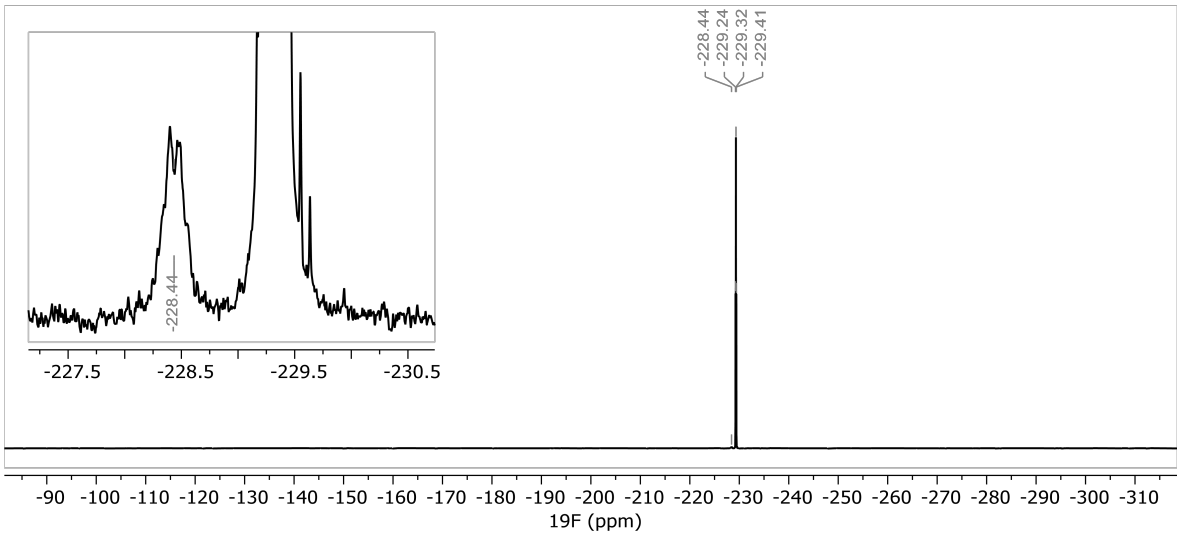


Figure 7.82: Representative ^{19}F-NMR spectrum (d_1 = 2 s, 4096 scans) of fluorinated diol **11** binding to Histone H3 Bal9. Bound diol leads to a broad peak at −228.44 ppm.

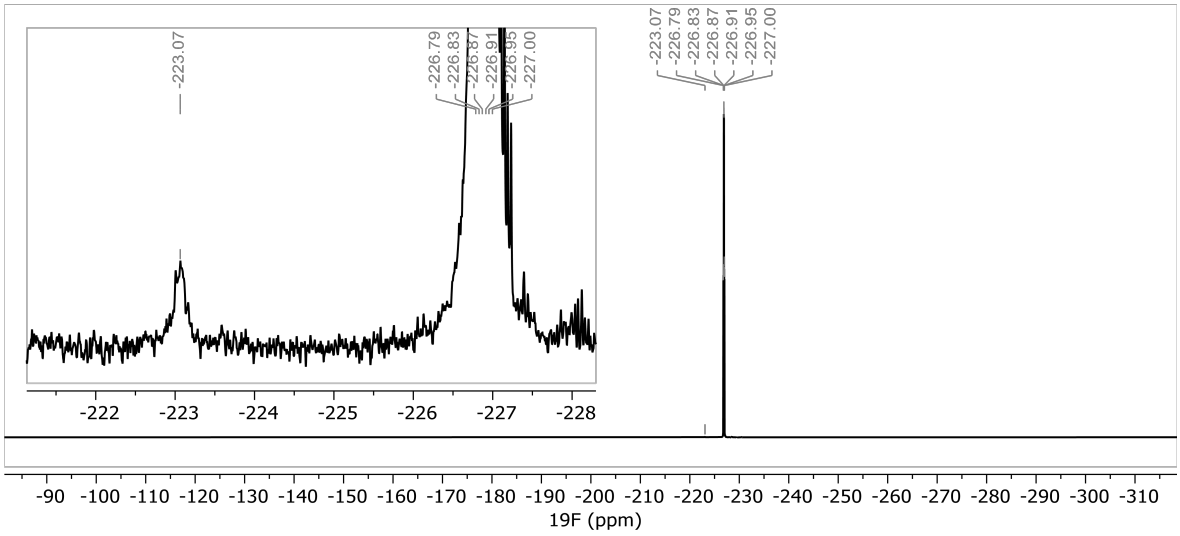


Figure 7.83: Representative ^{19}F-NMR spectrum ($d_1 = 2$ s, 20480 scans) of fluorinated ribose derivative **12** binding to Histone H3 Bal9. Bound ribose derivative leads to a broad peak at −223.07 ppm.

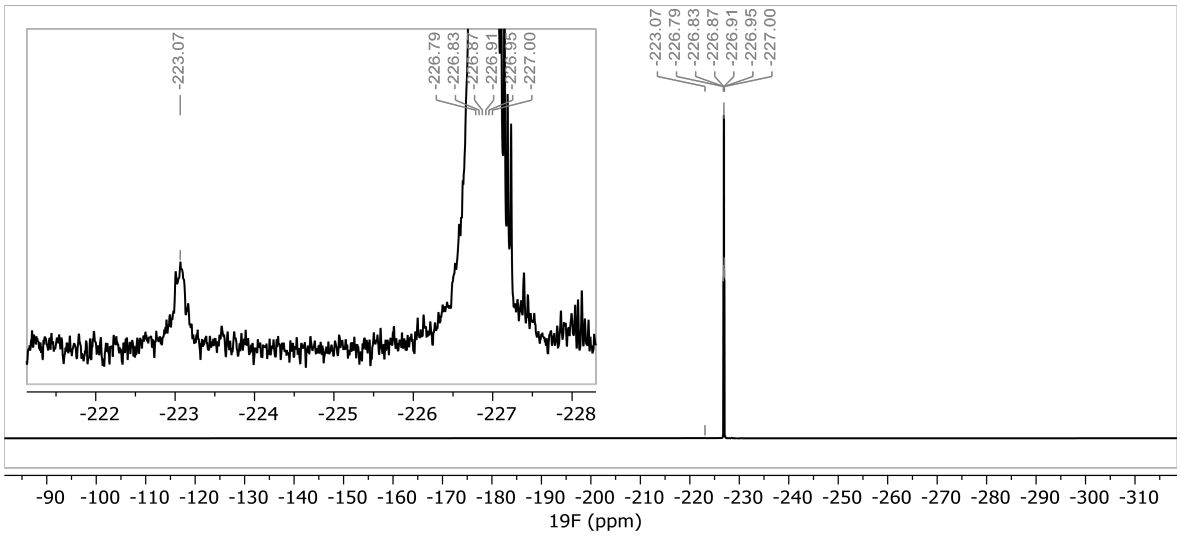


Figure 7.84: Representative ^{19}F-NMR spectrum ($d_1 = 2$ s, 25395 scans) of attempted binding of fluorinated galactose derivative **13** to Histone H3 Bal9

7.17.2 NMR-Titration of Histone H3 Bal9 with Fluorinated Diol 11

Histone H3 Dha9 was borylated following the standard procedure. The mixture was purified via dialysis into H_2O and lyophilized before being resuspended in NMR binding buffer buffer (NaP_i 40 mM, pH 7.0, 5 M urea, 10% D_2O). The protein was titrated with fluorinated diol **11** (stock solution in DMSO) covering a concentration range from 100 µM to 64 mM and ^{19}F-NMR spectra (d_1 = 2 s, 3072 scans) were recorded after each addition. The data were analysed using GraphPad Prism 8.0.0 (GraphPad Software Inc., San Diego, USA) and fitted to a four-parameter dose-response curve ($Y = Bottom + (X^{Hillslope}) * (Top - Bottom)/(X^{HillSlope} + K_D^{HillSlope})$).

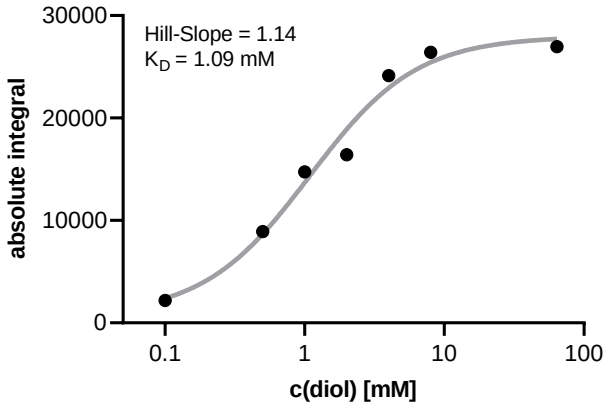


Figure 7.85: Dose-response curve obtained for ^{19}F-NMR-titration of Histone H3 Bal9 with fluorinated diol **11**

7.17.3 Control Experiments

Competition with Non-Fluorinated Triol 4

Samples were prepared following the previously described procedure for K_D determination. (1*R*)-(−)-Myrtentriol (**4**) (200 equiv) was added as a competitor.

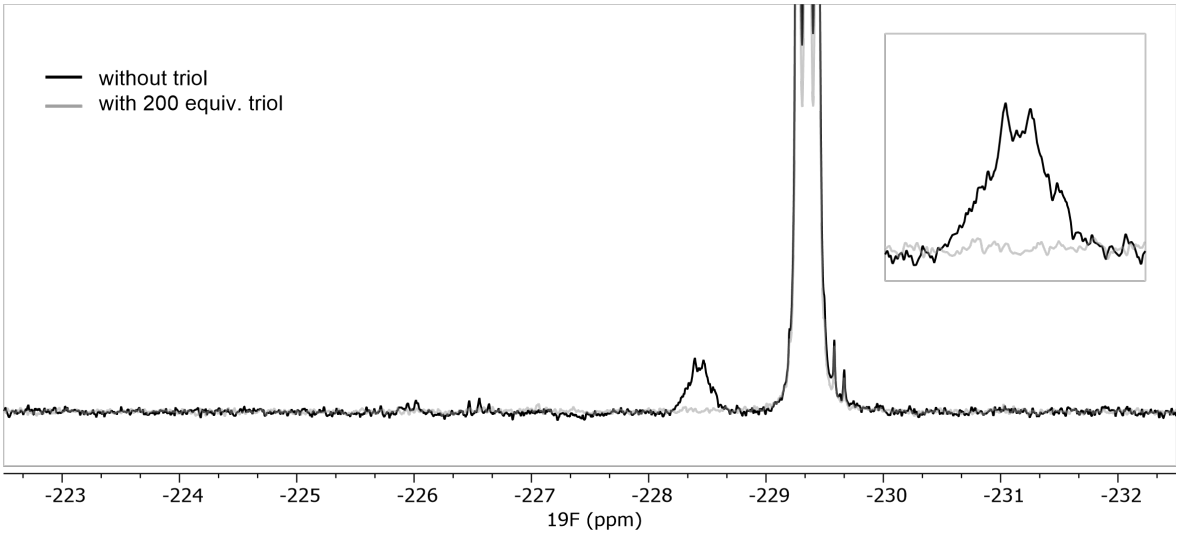


Figure 7.86: ^{19}F-NMR spectrum of fluorinated diol **11** binding to Histone H3 Bal9 with (grey spectrum) and without (black spectrum) competitor (triol **4**).

CMPG Filtered NMR

Samples were prepared following the previously described procedure for K_D determination. The control was measured using a 200 ms CMPG filter.

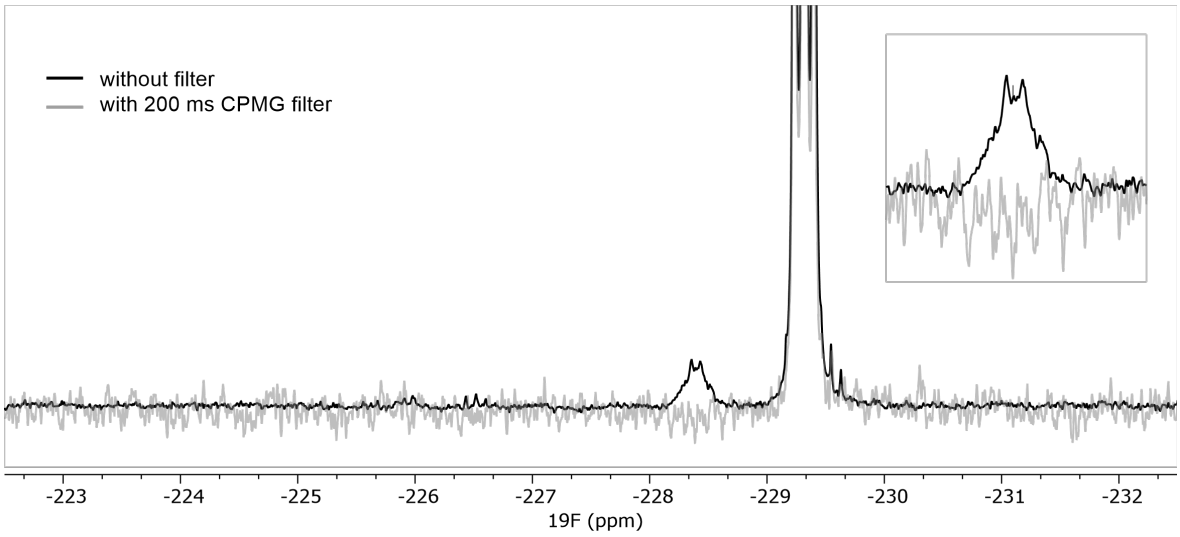


Figure 7.87: ^{19}F-NMR spectra of Histone H3 Bal9 in the presence of fluorinated diol **11** using a standard pulse sequence (black spectrum) or a 200 ms CPMG filter (grey spectrum). Peak intensities were normalized based on unbound diol.

7.17.4 Determination of the Diastereomeric Ratio Using a Chiral Shift Reagent

Determination of the Enantiomeric Ratio on Ac-Bal-NHBn

Ac-Bal-NHBn (**3**) was dissolved in binding buffer (40 mM NaP_i, 5 M urea, pH 7.0, 10% D_2O) at a final concentration of 500 µM. 10 equiv of chiral shift reagent **30** were added and the sample was vortexed. The sample was transferred to a NMR tube and analysed on a Bruker AVIII HD 500 MHz spectrometer equipped with BBFO SMART probe (2060 scans, d1 = 2 s). A ratio of approximately one-to-one was detected after integration of the peaks.

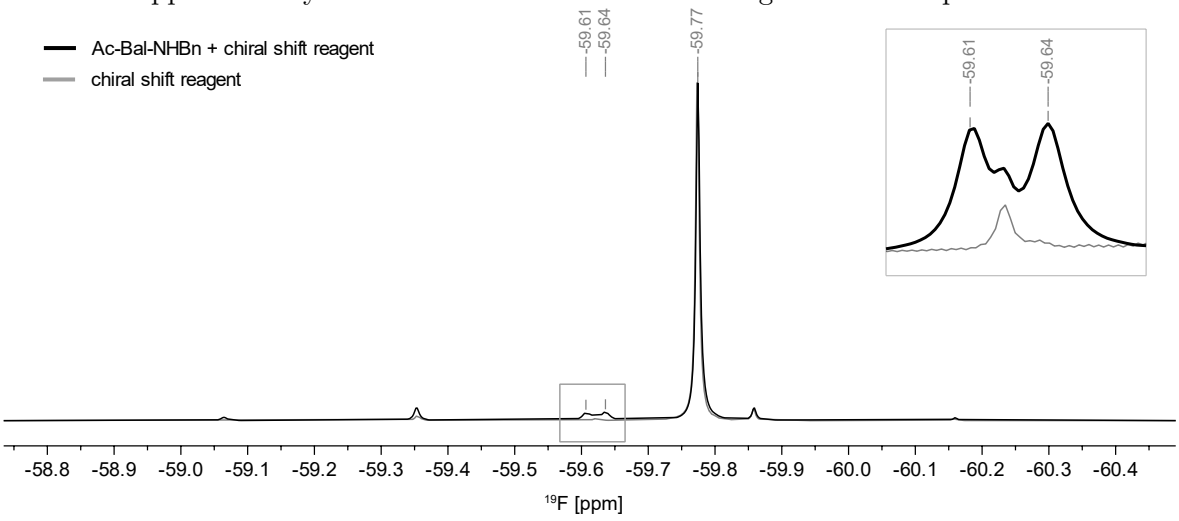


Figure 7.88: ^{19}F-NMR enables the determination of the enantiomeric ratio of Ac-Bal-NHBn (**3**) using chiral shift reagent **30**]

Determination of the Diastereomeric Ratio on Histone H3 Bal9

Histone H3 Bal9 was dissolved in binding buffer (40 mM NaP_i, 5 M urea, pH 7.0, 10% D_2O) at a final concentration of 144 µM. 10 equiv of chiral shift reagent **30** were added and the sample was vortexed. The sample was transferred to a NMR tube and analysed on a Bruker AVIII 600 MHz spectrometer equipped with a Prodigy N2 broadband cryoprobe (3500 scans, d1 = 2 s). A ratio of approximately one-to-one was detected after integration of the peaks.

7.18 Protein NMR

7.18.1 Protein CON NMR

Sample Preparation

The lyophilized samples were dissolved in 330 µL of NMR buffer (50 mM NaP_i, 3 M Gdn·HCl, 5% D_2O, pH 7.0).

NMR Spectroscopy

NMR measurements were conducted by Prof. D. Flemming Hansen, Institute of Structural and Molecular Biology, Division of Biosciences, University College London, UK.

The ^{13}C-detected backbone $^{13}CO-^{15}N$ correlation spectra were performed on a Bruker Avance II 600 MHz (14.1 T) NMR spectrometer using a TXO ^{13}C-optimised cryogenic probe with cooled ^{1}H, ^{13}C and ^{15}N preamplifiers, (^{1}H outer coil and ^{13}C inner coil). The standard $^{13}CO-^{15}N$ pulse sequence in the Bruker library (*c_caco_ia*)[24] was used with 160 real points recorded in the indirect ^{15}N dimension (SW = 35 ppm) and 64 transients recorded per FID. The spectra were processed using NMRPipe[25] and analysed using NMRFAM-Sparky.[26] All spectra were recorded at a temperature of 298 K.

7.18.2 Nucleosome NMR Experiments

Nucleosome NMR experiments are described in section 7.19.

7.19 Nucleosome NMR Study

Nucleosome NMR experiments were conducted by Charles Buchanan and Prof. Andrew Baldwin, Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, UK. Preliminary NMR studies were conducted by Dr. Lukas Lercher.[27] DNA preparation and nucleosome assembly was conducted by Brian Josephson, Department of Chemistry, Chemistry Research Laboratory, University of Oxford, UK.

7.19.1 Histone Expression and Purification

WT *Xenopus laevis* core Histones H2A, H2B, and H4 were expressed and purified as described previously (see section 7.4). ^{15}N-labelled Histone H3 mutants were expressed in M9 minimal media (see section 7.4.3). Borylation and oxidation were conducted as described previously (see section 7.6).

7.19.2 Nucleosome Assembly

Octamer Reconstitution, Nucleosome Assembly and 145 bp DNA product were all performed essentially as previously described.[28]

NMR Sample Preparation

Nucleosomes were dialyzed three times against NMR buffer (10 mM NaP_i, 1 mM EDTA, pH 6.5) (4°C). The last dialysis step was performed against slightly higher buffer concentration to account for addition of 5% D_2O (v/v). After dialysis, nucleosomes were concentrated to 190 µL (VivaSpin 500, 5 kDa MWCO) and D_2O (10 µL) was added. The samples were transferred to 3 mm Bruker Match NMR tubes.

7.19.3 Nucleosome Phosphorylation with Aurora B Kinase

After NMR measurement, samples were retrieved from the NMR tubes and diluted with 1800 µL of kinase buffer (10 mM TRIS, 30 mM NaCl, 1 mM DTT, 3 mM $MgCl_2$, 2 mM ATP, 10 mM β-glycerolphosphate, pH 7.5). In the case of Bal-containing nucleosomes, dilution was conducted in a nitrogen filled glove box using kinase buffer which was degassed overnight (4°C). This precaution was taken to prevent oxidation of Bal to Ser, which was observed on Histone H3 Bal9 upon prolonged exposure to kinase buffer at elevated temperature (30°C). Recombinant human Aurora B kinase (3 µg, Abcam, Catalogue No. ab51435) was added and the mixture was shaken (400 rpm) at 30°C for 48 h, after which the samples were again prepared for NMR measurement (see above).

Western Blot Analysis

Wild-type, phosphorylated wild-type, Histone H3 Bal9-containing and phosphorylated Histone H3 Bal9-containing nucleosomes were analysed by Western blot. After SDS-PAGE analysis (10% *bis*-TRIS precast gel, 200 V, 50 min, MES running buffer, 4°C), the proteins were transferred to a nitrocellulose membrane (20 V, 10 min). The membrane was blocked shaking for one hour in TBS-T (pH 7.5) supplemented with 5% (w/v) skimmed milk powder. The membrane was incubated with the primary mouse anti-Histone H3 pSer10 monoclonal antibody (GeneTex, Catalogue No. GTX630185) for one hour using a 1:2000 dilution before being washed three times for 5 min each. The membrane was incubated with a secondary goat anti-mouse IgG–alkaline phosphatase antibody (Sigma Aldrich, Catalogue No. A3562) using a 1:2000 dilution for another hour and again washed three times for 5 min each. Histone H3 pSer10 was visualized using NBT/BCIP substrate solution (Thermo Scientific, Catalogue No. 34042).

7.19.4 NMR Aquisition and Processing for Borylated Nucleosomes

Nucleosome NMR experiments were conducted by Charles Buchanan and Prof. Andrew Baldwin, Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, UK.

NMR experiments on borylated nucleosome samples were performed on a Bruker AVANCE III HD 950 MHz spectrometer equipped with a 5 mm TCI Cryoprobe. 2D $^{1}H—^{15}N$ BEST-TROSY experiments from the standard Bruker library were recorded with ^{1}H (^{15}N) 1024 (128) complex points, spectral widths of 10416 Hz (3465 Hz), maximum acquisition times of 98.3 ms (36.0 ms), 256 scans per FID with an interscan delay of 0.2 s for a total acquisition time of 2 h 52 min. To calculate errors, these spectra were recorded three times for the following samples: wild type nucleosome, phosphorylated wild type nucleosome, Histone H3 Bal9-containing nucleosome, denatured Histone H3 Ser10 (WT) and denatured Histone H3 Bal9. One measurement was made of the phosphorylated Histone H3 Bal9-containing sample, and errors for this sample were calculated by using the average peak noise found in the other spectra. These spectra were processed using a sine-bell window function, zero filling and phase

correction in both dimensions. The peaks were then manually identified in SPARKY[26] by mapping the assignments from Zhou *et al.*[29] With the exception of phosphorylated Histone H3 Bal9-containing nucleosome, the intensities were averaged across the three spectra and the standard deviations calculated. All nucleosome samples were measured in NMR buffer (1 mM EDTA, 10 mM NaP_i pH 6.5, 5% v/v D_2O). Denatured Histones were measured in denaturing NMR buffer (50 mM NaP_i, 3 M Gdn·HCl, 5% v/v D_2O, pH 7.0).

Concentration Determination

To quantify the concentration of nucleosomes, ^1H spectra were recorded with 8192 complex points and a sweep width of 10417 Hz for an acquisition time of 786 ms. 32 scans per FID were recorded with an interscan delay of 1 s. The spectra were processed in NMRPipe[25] using a sine-bell window function, zero filling, phase correction and linear baselining and then the methyl regions (0.62 ppm to 0.98 ppm) were extracted for comparison.

To quantify the concentration of denatured H3 Histones, ^1H spectra were recorded with 31248 complex points and a sweep width of 10417 Hz for an acquisition time of 3 s. 32 scans per FID were recorded with an interscan delay of 1 s. The spectra were processed in NMRPipe[25] using a sine-bell window function, zero filling, phase correction and linear baselining and then extracted the methyl regions (0.7 ppm to 1.14 ppm) for comparison.

Temperature Variation Experiments

The same experimental protocol was applied to denatured Histone H3 Ser9 and denatured Histone H3 Bal9 samples. 2D ^1H—^{15}N BEST-TROSY spectra were recorded three times at three temperatures; 288 K, 298 K and 303 K. Processing again involved a sine-bell window function, zero filling and phase correction in both dimensions. In the absence of assignments, isolated peaks in the GST region were peak picked in SPARKY.[26] Peak intensities were averaged and standard deviations calculated for errors. Denatured Histones were measured in denaturing NMR buffer (50 mM NaP_i, 3 M Gdn·HCl, 5% v/v D_2O, pH 7.0).

References

[1] A. Avdeef, K. Box, J. Comer, M. Gilges, M. Hadley, C. Hibbert, W. Patterson, K. Tam, *J. Pharm. Biomed. Anal.* **1999**, *20*, 631–641.

[2] B. N. Naidu, M. E. Sorenson, T. P. Connolly, Y. Ueda, *J. Org. Chem.* **2003**, *68*, 10098–10102.

[3] H. J. Reich, W. S. Goldenberg, B. Ö. Gudmundsson, A. W. Sanders, K. J. Kulicke, K. Simon, I. A. Guzei, *J. Am. Chem. Soc.* **2001**, *123*, 8067–8079.

[4] D. R. Morgan, H. C. Dorn, *J. Label. Compd. Radiopharm.* **1991**, *29*, 777–779.

[5] B. Dedeoğlu, I. Uğur, I. Değirmenci, V. Aviyente, B. Barçin, G. Çayli, H. Y. Acar, *Polymer* **2013**, *54*, 5122–5132.

[6] J. B. Liu, X. H. Xu, F. L. Qing, *Org. Lett.* **2015**, *17*, 5048–5051.

[7] B. P. Rempel, S. G. Withers, *Org. Biomol. Chem.* **2014**, *12*, 2592–2595.

[8] C. C. Geilen, N. Loch, W. Reutter, K. Seppelt, F. Oberdorfer, *Tetrahedron Lett.* **1992**, *33*, 2435–2438.

[9] J. N. Barlow, J. S. Blanchard, *Carbohydr. Res.* **2000**, *328*, 473–480.

[10] X. G. Li, S. Dall'Angelo, L. F. Schweiger, M. Zanda, D. O'Hagan, *Chem. Commun.* **2012**, *48*, 5247–5249.

[11] M. Sharma, Y. X. Li, M. Ledvina, M. Bobek, *Nucleosides Nucleotides Nucleic Acids* **1995**, *14*, 1831–1852.

[12] A. Yang, F. Hu, Z. Li, M. Chen, J. Cai, L. Wang, T. Zhang, C. Zhao, F. Zhang, *Org. Process Res. Dev.* **2019**, *23*, 818–824.

[13] S. P. Cummings, T. Le, G. E. Fernandez, L. G. Quiambao, B. J. Stokes, *J. Am. Chem. Soc.* **2016**, *138*, 6107–6110.

[14] T. Ghosh, A. Mukherji, P. K. Kancharla, *Org. Lett.* **2019**, *21*, 3490–3495.

[15] M. Trincado, K. Kühlein, H. Grützmacher, *Chem. Eur. J.* **2011**, *17*, 11905–11913.

[16] P. Lundström, P. Vallurupalli, D. F. Hansen, L. E. Kay, *Nat. Protoc.* **2009**, *4*, 1641–1648.

[17] B. Josephson, C. Fehl, P. G. Isenegger, S. Nadal, T. H. Wright, A. W. J. Poh, B. J. Bower, A. M. Giltrap, L. Chen, C. Batchelor-McAuley, G. Roper, O. Arisa, J. B. I. Sap, A. Kawamura, A. J. Baldwin, S. Mohammed, R. G. Compton, V. Gouverneur, B. G. Davis, *Nature* **2020**, *585*, 530–537.

[18] T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallee, A. Phanumartwiwath, O. D. Coleman, M.-L. Thezenas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed, B. G. Davis, *Science* **2016**, *354*, 597.

[19] J. Dadová, K.-J. Wu, P. G. Isenegger, J. C. Errey, G. J. L. Bernardes, J. M. Chalker, L. Raich, C. Rovira, B. G. Davis, *ACS Cent. Sci.* **2017**, *3*, 1168–1173.

[20] S. Mitternacht, *F1000Research* **2016**, *5*, 1–11.

[21] B. Lee, F. Richards, *J. Mol. Biol.* **1971**, *55*, 379.

[22] M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman, C. O. Wilke, *PLoS One* **2013**, *8*, (Ed.: A. Porollo), e80635.

[23] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nat. Methods* **2012**, *9*, 671–675.

[24] W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, *J. Magn. Reson.* **2006**, *178*, 56–64.

[25] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.

[26] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.

[27] L. Lercher, DPhil Thesis, University of Oxford, **2014**.

[28] P. N. Dyer, R. S. Edayathumangalam, C. L. White, Y. Bao, S. Chakravarthy, U. M. Muthurajan, K. Luger, *Methods Enzymol.* **2003**, *375*, 23–44.

[29] B.-R. Zhou, H. Feng, R. Ghirlando, H. Kato, J. Gruschus, Y. Bai, *J. Mol. Biol.* **2012**, *421*, 30–37.