Exploring the ligand-binding potential of the Neuronal PAS proteins NPAS1 and NPAS3, and their partner ARNT2

Viktoriia Bazylianska

St. Hugh's College

Nuffield Department of Medicine
University of Oxford

Supervisors

Professor Fraydoon Rastinejad

Nuffield Department of Medicine
University of Oxford

Professor Paul Brennan

Nuffield Department of Medicine
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2025

Luck is not chance—
It's Toil—
Fortune's expensive smile
Is earned—

Emily Dickinson
(1830-1886)

Abstract

Mammalian basic helix-loop-helix PER-ARNT-SIM (bHLH-PAS) transcription factors play key roles in various human diseases and are emerging as promising drug targets due to their ability to bind small molecules within their PAS domains. Among them, the neuronal PAS proteins NPAS1 and NPAS3, together with their heterodimerisation partner ARNT2, have been linked to neurological conditions, notably schizophrenia, autism spectrum disorders, intellectual disabilities, and bipolar disorder, making them promising candidates for therapeutic targeting.

This DPhil project identifies novel small molecules for NPAS1, NPAS3, and their partner, ARNT2. A high-throughput biochemical screen comprising 67,847 compounds was conducted with NPAS1-ARNT2, NPAS3-ARNT2, and ARNT2-PAS-B proteins. Reconfirmed hits were evaluated for binding affinity using MST and SPR. Crystal structures of ligand-bound ARNT2-PAS-B revealed that compounds occupy a large cavity within the domain, and interact with His341, a residue critical for ligand stabilisation. Ligand binding induced conformational changes that enlarged the binding pocket, and disrupted interfaces critical for heterodimer stability.

Hit-to-lead optimisation using the known ARNT2 binder KG-548, and butein, yielded ligand series designed to reduce PAINS liabilities and improve affinity, metabolic stability, and selectivity for the ARNT2 PAS-B domain. Distinct yet complementary binding modes of these compounds highlighted the PAS-B pocket's flexibility in accommodating diverse scaffolds through hydrogen bonding, ionic interactions, and water-mediated networks.

Abstract

This study identifies ligands that directly target the ARNT2 PAS-B domain, and demonstrates their ability to modulate heterodimer stability through structural and biochemical characterisation. Compounds 3617-M16, FR-15, FR-20, and FR-26 emerge as promising probes, which may be used for exploring the functional role of ARNT2. These findings provide a framework for designing next-generation ARNT2 modulators and for investigating ligand-dependent regulation of NPAS-ARNT2 transcriptional activity in health and disease.

Acknowledgements

I should like to thank everyone who inspired, supported, and encouraged me throughout my entire scientific journey, both within and beyond the laboratory. This DPhil work would not have been possible without the guidance, collaboration, and assistance of so many remarkable people.

I would like to express my profound gratitude to my supervisor, Professor Fraydoon Rastinejad, for providing me with the incredible opportunity to work on this exciting project, and for his invaluable guidance and feedback throughout my doctoral research. His vast research experience in the field of drug discovery has made these four years truly inspiring. It has been a pleasure to learn from his wide erudition, and his insight and knowledge have steered me through this research. More than this, his gracious mentorship has left a lasting impact on my academic and personal development.

I am also deeply grateful to Professor Paul Brennan for his invaluable supervision, expert input on the chemical aspects of my project, and his steadfast professional and personal support over the years. I feel truly fortunate to have been a part of his group, and to have had the honour of learning more about medicinal chemistry under his guidance – knowledge that greatly enriched my work. I am immensely grateful for his unsparing support, which has gone far beyond the supervisor's usual call of duty, especially in offering assistance after the full-scale invasion of my beloved home country, Ukraine.

Special thanks go to Dr. Andrew Ball for giving so generously of his time and knowledge, and for always offering enthusiastic assistance in any way he could throughout the research

project. I am deeply grateful for his extensive support with crystallography, MST, TR-FRET assays, for detailed feedback on my thesis, and for his patient answers to countless questions.

My heartfelt thanks go to the wonderful members of our laboratory, who made these years both productive and joyful. After so many years together, it is a privilege to call you my friends. Thank you for your technical advice, your collaboration, and the many moments of laughter we shared. I am especially grateful to Dr. Sepideh Khorasanizadeh for her support throughout my DPhil, for our many memorable conversations, and for her detailed feedback on my thesis, which helped me become a better writer. I also thank Dr. Viacheslav Trush for his mentorship over the years, for teaching me protein purification, and for his thoughtful career advice and encouragement. I am particularly grateful to Dr. Srinivas Cheruku for teaching me how to interpret crystallographic data, for designing the butein analogues, and for his instructive feedback on my thesis; to Dr. Hua Pu for instruction on SPR experiments and crystallography; and to Dr. Minghao Zhang for assistance with molecular cloning and protein purification. I sincerely hope we will remain in touch for many years to come.

Many thanks to Dr. Lukas Scheibelberger, Dr. Ian de Toledo, and Tobias Bernard for designing and synthesising the KG-548 analogues.

I also wish to acknowledge the support of the Nuffield Department of Medicine for fostering an exceptional scientific environment in which to work, and for providing a full scholarship and generous stipend, which allowed me to dedicate myself fully to my research.

I wish to express my profound appreciation to all the teachers, mentors, and supervisors who have guided me throughout my education, from my early school years to universities in

Ukraine, Germany, and the United States. Their dedication, encouragement, and belief in my potential have greatly influenced my scientific path and personal growth.

I am especially thankful to my school biology teacher, Svitlana Lvivna Kovbas, who recognised my early interest in the biological sciences and nurtured my curiosity. Her late-night lessons on protein synthesis, while preparing me for the Biology Olympiad and national science competitions, sparked my lifelong passion for scientific enquiry. I am also deeply grateful to my BSc and MSc supervisors, Dr. Tetyana Mykolayivna Horid'ko and Dr. Halyna Vasylivna Kosiakova, who first opened the door to the wonders of experimental science and inspired me to pursue a career in research.

My sincere thanks go to Dr. María Moreno-Villanueva at the University of Konstanz, whose warm guidance eased my first experience living and studying abroad in Germany. She devoted countless hours to teaching me a range of biochemical assays that were instrumental to the success of my DPhil research.

I am also thankful to Dr. Siddhesh Aras, who made my time in the United States an exciting scientific adventure and taught me molecular cloning (a skill that proved invaluable during this work). My heartfelt appreciation extends to Dr. Anna B. Moszczynska, one of the most determined academics I have ever met, and whose dedication to science makes her one of my greatest female role models in STEM.

I am deeply grateful to my master's degree supervisors, Professor Maik Hüttemann at Wayne State University, and Professor Antonio Barrientos at the University of Miami, for their exemplary mentorship, and for planting the seed in my mind to pursue doctoral study.

I wish to express my heartfelt thanks to all my friends, both near and far, and above all to my beloved family, for their constant love, support, and encouragement throughout this journey. To Dr. Peter William Walsh: thank you for being at my side all these years. Your understanding, devotion, and constant willingness to help have been my greatest strength and mean more to me than words can say.

To my beloved grandmother, Viktoriia Derogan, who raised me with endless love, wisdom, and strength: your resilience and kindness have shaped the person I am today, and for this I am forever grateful.

To my mother, Nataliia: thank you for your unconditional love, moral support, and daily prayers. Your wicked sense of humour and light-hearted spirit have filled these years with joy and happiness.

To my darling brother, Dima: you are the light of my life. On so many occasions, your genuine laughter and warm smile were what I needed most.

To my Uncle Kosta: thank you for being the backbone of our family, and for looking after Granny during difficult times.

Finally, I would like to pay special tribute to my beloved grandfather, Dr. Dmytro Derogan, who taught me perseverance, no matter the challenge; and whose wisdom, kindness, and integrity have been guiding lights throughout my life. Though he is no longer with us, I hope he watches proudly from above. His example continues to inspire me every day.

To all of you who have had a hand in my journey – thank you, from the depths of my heart.

Від щирого серця дякую всім, хто вірив у мене.

Declaration

Except where acknowledgement is made in the text or figure legends, the work described in this thesis was performed by the author in the Nuffield Department of Medicine Research Building, the Target Discovery Institute and Centre for Medicine Discovery, at the University of Oxford. The work reported in this thesis has not been submitted for any other degree in this or any other university or institute of learning.

Table of Contents

Abstract i

Acknowledgements iii

Declaration vii

List of Figures xiii

List of Tables xx

List of Appendices xxi

List of Abbreviations xxii

PART I: Introduction and Literature Review 1

Chapter 1: Introduction to the Neuronal bHLH-PAS transcription factors, NPAS1 and NPAS3, and their partner ARNT2 2

Chapter 2: Literature review 5

2.1 NPAS1, NPAS3, and ARNT2 are members of the bHLH-PAS family of transcription factors 5

2.2 Dimerisation of NPAS1 and NPAS3 proteins 6

2.3 Expression of NPAS1 and NPAS3 in the nervous system 9

2.4 Aryl hydrocarbon receptor nuclear translocator 2 (ARNT2): a preferred NPAS partner in neuronal contexts 14

2.5 Domain structure of NPAS1, NPAS3, ARNT2, and other bHLH-PAS proteins 16

2.5.1 The bHLH (basic helix-loop-helix) DNA-binding domain 17

2.5.2 Tandem PAS-A and PAS-B domains 19

2.5.3 C-terminal regulatory domains 22

2.6 Structural analyses of NPAS1, NPAS3, and ARNT2 proteins 26

2.6.1 Crystal structures of NPAS1-ARNT and NPAS3-ARNT heterodimers 26

2.6.2 Crystal structure of ARNT2 protein in its complex with NPAS4 30

2.6.3 Ligand binding pockets within NPAS1, NPAS3, and ARNT2 33

2.7 Targeting ligand-binding pockets within bHLH-PAS proteins 37

2.8 Functional roles of NPAS1 and NPAS3 in neurodevelopment and disease 43

2.9 ARNT2 is a transcriptional integrator in neurodevelopment, metabolism, and human disease 50

2.10 KG-548 as a ligand for ARNT2: discovery, mechanism, and assays 56

Chapter 3: DPhil proposal and project aims 62

PART II: Results and Discussion 66

Chapter 4: Protein production and quality control 67

4.1 Construct selection and purification of the PAS-B domain of ARNT2 67

4.2 Comparison of human and mouse ARNT2 protein 73

4.3 Expression and purification of the single PAS-B domain of NPAS1 and NPAS3 76

4.4 Comparison of human and mouse NPAS1 and NPAS3 proteins 77

4.5 Generation and purification of NPAS1 and NPAS3 triple domains 83

4.6 Generation of NPAS1 and NPAS3 heterodimer complexes with ARNT2 88

4.7 Expression and purification of NPAS1 and NPAS3 heterodimers with ARNT2 90

4.8 Western blot analysis of the heterodimers NPAS1-ARNT2 and NPAS3-ARNT2 93

4.9 Mass spectrometry analysis of hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 heterodimers 95

Chapter 5: Hit identification 100

5.1 Thermal shift assay 100

5.2 TSA: setup and validation 102

5.3 High-throughput screening with mARNT2-PAS-B, hNPAS1-ARNT2, and hNPAS3-ARNT2 109

5.4 Dose-response follow-up for hit validation 114

5.5 Lead compound selection from the primary hits 116

5.6 Confirmation of primary hits and counter screen 125

Chapter 6: Structural analyses and crystallography 135

6.1. Ligand binding and induced conformational changes in the mARNT2-PAS-B domain upon interaction with KG-548 135

6.2 Crystallisation of the proteins of interest 143

6.3 Optimisation of mARNT2-PAS-B crystals 147

6.4 Crystal structures of mARNT2-PAS-B complexes with tested compounds 150

Chapter 7: Microscale thermophoresis binding assay 161

7.1 Principle of microscale thermophoresis 161

7.2 Development of the MST assay 164

7.3 MST binding assay in dose-response format 172

Chapter 8: Surface plasmon resonance 182

8.1 Introduction to the surface plasmon resonance assay 182

8.2 Development of the SPR assay 184

8.3 Immobilisation of proteins on the SA chip and optimisation of the SPR assay 188

8.4 The SPR competition assay in dose-response format 191

Chapter 9: Time-resolved FRET dimerisation assay 202

9.1 Principle of the TR-FRET assay 202

9.2 Development of the TR-FRET assay 204

9.3 Optimisation of the TR-FRET assay 210

9.4 TR-FRET assay with all the tested compounds 213

Chapter 10: Hit-to-lead optimisation for KG-548 **221**

10.1 Binding mode of KG-548 and initial design of analogues 221

10.2 Expanding a narrow sub-pocket of mARNT2-PAS-B 230

10.3 Attaching substituents to the tetrazole ring 233

10.4 Acid/tetrazole bioisosteres analogues 237

10.5 Aryl ring modifications 245

Chapter 11: Hit-to-lead optimisation of butein **262**

11.1 Butein: from herbal legacy to modern drug discovery 262

11.2. Butein binds the PAS-B domain of ARNT2 264

11.3 Hit-to-lead optimisation for butein 273

11.4 Design and evaluation of butein analogues to improve stability and target occupancy 280

11.5 Design of butein analogues to address selectivity concerns 289

Chapter 12: Closing remarks **298**

12.1 Overall conclusion 298

12.2 Future studies 302

PART III: Materials and Methods **307**

Chapter 13: Preparation of materials and buffers **308**

13.1 Lysogeny Broth (LB) media 308

13.2 LB Agar plates 308

13.3 Antibiotics 308

13.4 Agarose gel 309

13.5 Buffers for protein purification and HTS screening 309

Chapter 14: Methods and experimental approach **311**

14.1 Plasmid construction and general cloning procedure 311

14.2 Bacterial transformation and protein overexpression 313

14.3 Protein purification 314

14.3.1 Immobilised metal affinity chromatography for 6xHis-tagged proteins 314

14.3.2 Cation exchange-based isolation of target proteins 315

14.3.3 Gel filtration 315

14.4 SDS-PAGE and Western Blotting 316

14.4.1 SDS-PAGE analysis of protein samples 316

14.4.2 Western Blot analysis for protein validation 316

14.5 Protein analysis via mass spectrometry 317

14.6 Thermal shift analysis for protein stability evaluation 317

14.6.1 Compound handling and preparation for HTS chemical screen 317

14.6.2 Thermal shift assay 318

14.7 Protein crystallisation, structure determination, and refinement 319

14.8 Microscale thermophoresis analysis of protein-ligand interactions 320

14.9 Determination of binding affinity through Surface Plasmon Resonance 321

14.10 TR-FRET assay 323

Appendices **326**

References **343**

List of Figures

Figure 1. NPAS1 and NPAS3 are transcription factors in neuronal contexts. 3

Figure 2. NPAS1, NPAS3, and ARNT2 are members of the bHLH-PAS family. 6

Figure 3. Classification of the bHLH-PAS protein family. 8

Figure 4. Tissue and organ-wide expression profile of NPAS1 in humans. 11

Figure 5. Tissue and organ-wide expression profile of NPAS3 in humans. 12

Figure 6. Comparative tissue and organ-wide expression profiles of ARNT and ARNT2 in humans. 15

Figure 7. Structure of ordered domains in the bHLH-PAS proteins. 18

Figure 8. Domain variations in the bHLH-PAS family. 21

Figure 9. Representative structural analyses of the C-terminal flexible regions of bHLH-PAS proteins. 24

Figure 10. Structural analysis of NPAS1 and NPAS3 proteins. 28

Figure 11. Comparison of bHLH-PAS proteins. 29

Figure 12. Structural analysis of ARNT2 protein. 32

Figure 13. Representative structural analyses of ligand binding pockets for the NPAS1-ARNT heterodimer. 35

Figure 14. The functional role of bHLH-PAS proteins and their disease association. 38

Figure 15. Pharmacological targeting of the PAS-B domain within the bHLH-PAS family. 42

Figure 16. Known ARNT2 binders. 60

Figure 17. Preparation of the human and mouse ARNT2-PAS-B domains. 70

Figure 18. Mass analysis of purified ARNT2-PAS-B proteins by ESI-MS. 72

Figure 19. Sequence alignment of full-length human and mouse ARNT2 proteins. 75

Figure 20. Representative structural comparison of the human and mouse PAS-B binding pocket of ARNT2. 75

Figure 21. Purification results for the single PAS-B domains of hNPAS1/3. 77

Figure 22. Sequence comparison of human and mouse NPAS1 and NPAS3 proteins. 79

Figure 23. Structural comparison between human and mouse NPAS1 proteins. 81

Figure 24. Representative vector map of pETDuet-1 with hNPAS3 (residues 19-437). 84

Figure 25. DNA gels with one per cent agarose. 86

Figure 26. A representative example of NPAS3 triple domain purification. 87

Figure 27. Schematic representation of NPAS1 and NPAS3 constructs that were prepared. 88

Figure 28. DNA gels with one per cent agarose. 89

Figure 29. Purification of human NPAS1/3 heterodimers with ARNT2. 91

Figure 30. Purification of mouse NPAS3 heterodimer with ARNT2. 92

Figure 31. SDS-PAGE and Western blots with anti-His and anti-Flag antibodies. 95

Figure 32. Intact mass analysis of the purified hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 proteins by ESI-MS. 97

Figure 33. Spontaneous α-N-6-Phosphogluconoylation of 6xHis-tagged proteins. 99

Figure 34. TSA overview. 101

Figure 35. TSA analyses for protein concentration. 103

Figure 36. SYPRO Orange optimisation for TSA. 104

Figure 37. Buffer assessment for mARNT2-PAS-B. 105

Figure 38. The buffer pH optimisation for heterodimers. 106

Figure 39. The buffer-salt optimisation for heterodimers. 107

Figure 40. DMSO tolerance test. 108

Figure 41. Prestwick Drug-Fragment Library. 109

Figure 42. The primary HTS screen via TSA. 111

Figure 43. Structural comparison between primary hits and the known control, KG-548, and recently identified butein. 112

Figure 44. Crystal structure of the PAS-B domain of mARNT2 with the positions of three cysteine residues highlighted. 113

Figure 45. Representative examples of a follow-up DRC with one of the primary hits for mARNT2-PAS-B, hNPAS1-ARNT2, and hNPAS3-ARNT2 proteins. 115

Figure 46. Representative examples of DataWarrior structural similarity and clusters. 116

Figure 47. DRC for fifteen final primary hits for mARNT2-PAS-B. 119

Figure 48. DRC for fifteen final primary hits for hNPAS1-ARNT2 heterodimer. 119

Figure 49. DRC for fifteen final primary hits for hNPAS3-ARNT2 heterodimer. 120

Figure 50. Schematic representation of lead compound selection. 121

Figure 51. Fresh powder DRC-TSA profiles for fifteen primary hits, with KG-548 control, tested against mouse and human ARNT2-PAS-B in three independent runs. 128

Figure 52. Fresh powder DRC-TSA profiles for fifteen primary hits, with KG-548 control, tested against hNPAS1-ARNT2 in three independent runs. 129

Figure 53. Fresh powder DRC-TSA profiles for fifteen primary hits, with KG-548 control, tested against hNPAS3-ARNT2 in three independent runs. 130

Figure 54. Belzutifan control for the TSA studies with hHIF-2α PAS-B. 133

Figure 55. The crystal structures of mARNT2-PAS-B in the absence of a compound and as bound to KG-548. 137

Figure 56. Structural analysis of the mARNT2 PAS-B ligand-binding pocket and interactions with the control compound KG-548. 140

Figure 57. Structural analyses of mARNT2-NPAS4 (PDB: 7XI3) upon binding to KG-548. 142

Figure 58. Crystallisation screening. 146

Figure 59. Representative unoptimised mouse ARNT2-PAS-B crystals. 147

Figure 60. Optimisation of mARNT2-PAS-B protein crystallisation. 148

Figure 61. Diagram of crystal progression over time. 149

Figure 62. X-ray crystallography for 3617-M16 binding to ARNT2-PAS-B. 152

Figure 63. Conformation changes of apo-ARNT2-PAS-B upon binding 3617-M16. 154

Figure 64. X-ray crystal structure of the 50-E18 compound with ARNT2-PAS-B. 156

Figure 65. Conformation changes of apo-ARNT2-PAS-B upon 50-E18 ligand binding. 158

Figure 66. X-ray crystallography with 3613-D4 and ARNT2-PAS-B. 159

Figure 67. Representative MST trace for different concentrations of hNPAS1-ARNT2. 163

Figure 68. Protein curve analyses via MST. 165

Figure 69. Comparison of the MST analyses for KG-548 against mARNT2-PAS-B. 167

Figure 70. Structural analyses of lysine and cysteine residues inside the PAS-B binding pocket of mARNT2 proteins. 169

Figure 71. MST dose-response curve for KG-548 against mARNT2-PAS-B. 171

Figure 72. MST dose-response curve for 3617-M16 against mARNT2-PAS-B. 173

Figure 73. MST dose-response curve for 8-I21 against mARNT2-PAS-B. 174

Figure 74. MST dose-response curve for 3620-L22 against mARNT2-PAS-B. 175

Figure 75. MST dose-response curve for 3621-L6 against mARNT2-PAS-B. 176

Figure 76. MST dose-response curve for 41-B13 against hNPAS3-ARNT2. 178

Figure 77. MST dose-response curve for 3623-J16 against mARNT2-PAS-B. 180

Figure 78. Illustrative diagram of Surface Plasmon Resonance. 183

Figure 79. Purification and quality control of biotinylated NPAS1-ARNT2. 186

Figure 80. Purification of biotinylated NPAS3-ARNT2. 187

Figure 81. Representative example of a real-time dynamic process of biotinylated NPAS1-ARNT2 protein immobilisation on the surface of the SA chip. 189

Figure 82. Evaluation of DMSO via SPR. 191

Figure 83. Steady state affinity of the KG-548 positive control with biotinylated mARNT2-PAS-B. 192

Figure 84. Steady state affinity for 3617-M16 with biotinylated mARNT2-PAS-B. 194

Figure 85. Steady state affinity analyses for ARNT2 tested compounds with biotinylated mARNT2-PAS-B. 196

Figure 86. Dose-response kinetic model for 3624-N7 and 3626-N5 enantiomers. 198

Figure 87. Dose-response kinetic model for 3623-J16. 199

Figure 88. Illustration of the TR-FRET principle. 203

Figure 89. One per cent DNA agarose gels. 205

Figure 90. Protein purification for the TR-FRET assay. 206

Figure 91. Representative intact mass analysis of purified hNPAS1-ARNT2-eGFP. 207

Figure 92. Structural illustration of the distance between potential fluorophores. 209

Figure 93. Buffer optimisation and protein curve analyses for the TR-FRET assay. 211

Figure 94. Incubation time optimisation with mHIF-2α-ARNT2-eGFP. 212

Figure 95. TR-FRET with KG-548 control compound. 213

Figure 96. TR-FRET with 13 ARNT2-PAS-B tested compounds. 216

Figure 97. TR-FRET with NPAS1 compounds. 218

Figure 98. TR-FRET with NPAS3 compounds. 219

Figure 99. Protonation equilibrium of KG-548 at physiological pH 7.4. 222

Figure 100. Key interactions of KG-548 and its first analogues. 224

Figure 101. X-ray crystallography for Target-4 binding to ARNT2-PAS-B. 227

Figure 102. TR-FRET with Target-3, Target-4, and the KG-548 control. 229

Figure 103. SPR analyses of Target-4 and Target-3 analogues. 230

Figure 104. KG-548 analogues based on modifications of the benzofuran scaffold to address the back pocket strategy. 232

Figure 105. KG-548 analogues based on modifications of the heterocyclic hydrogen bond acceptors. 236

Figure 106. KG-548 analogues based on modifications with acid bioisosteres. 239

Figure 107. X-ray crystallography for LS02-121-01 binding to ARNT2-PAS-B. 241

Figure 108. Steady state affinity analysis and calculated K_D for several KG-548 analogues with the biotinylated PAS-B domain of ARNT2. 244

Figure 109. TR-FRET with several KG-548 analogues. 245

Figure 110. KG-548 analogues based on aryl ring modifications. 248

Figure 111. X-ray crystallography for LS002-024 binding to ARNT2-PAS-B. 251

Figure 112. TR-FRET with several KG-548 analogues progressed from the TSA analyses. 252

Figure 113. Steady state affinity analyses of several KG-548 analogues with the biotinylated PAS-B domain of ARNT2. 253

Figure 114. Similarities between analogues with aryl ring modifications. 254

Figure 115. X-ray crystallography for LS02-170-01 binding to ARNT2-PAS-B. 255

Figure 116. X-ray crystallography for LS02-174-01 binding to ARNT2-PAS-B. 257

Figure 117. Structural comparison of the analogues LS02-174-01 and LS02-170-01. 258

Figure 118. Conformational expansion of ARNT2 PAS-B upon KG-548 binding. 261

Figure 119. Butein, a multi-targeted compound with diverse biological effects. 263

Figure 120. TSA studies with butein. 265

Figure 121. X-ray crystal structure of the butein compound with ARNT2-PAS-B. 267

Figure 122. Structural analyses of mARNT2-NPAS4 upon binding to KG-548 and butein. 268

Figure 123. Structural analyses of mHIF2α upon binding to belzutifan. 270

Figure 124. TSA studies with belzutifan, butein, and KG-548 compounds. 271

Figure 125. SPR and TR-FRET studies with butein compound. 273

Figure 126. Butein analogues with reduced hydroxyl groups. 274

Figure 127. TSA and TR-FRET studies for butein analogues with reduced hydroxyl groups. 276

Figure 128. SPR studies for butein analogues with reduced hydroxyl groups. 278

Figure 129. Crystal structure of the FR-15 butein analogue with ARNT2-PAS-B. 279

Figure 130. Analogues with the unsaturated ketone replaced by an amide linkage. 281

Figure 131. Butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. 283

Figure 132. TSA studies for butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. 284

Figure 133. TR-FRET studies for butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. 285

Figure 134. SPR studies for butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. 286

Figure 135. Crystal structure of the FR-20 butein analogue with ARNT2-PAS-B. 288

Figure 136. Butein analogues with structural modifications on the right-hand side for potential selectivity toward ARNT2. 291

Figure 137. TSA and TR-FRET studies for the FR-26 analogue. 293

Figure 138. SPR steady state affinity analysis for the FR-26 analogue with the biotinylated single PAS-B domain of ARNT2. 294

Figure 139. Crystal structure of the FR-26 butein analogue with ARNT2-PAS-B. 295

Figure 140. Conformational expansion of ARNT2 PAS-B upon KG-548 and butein binding. 297

List of Tables

Table 1. Ligand-binding pocket volumes of bHLH-PAS family proteins. 35

Table 2. Amino acid differences between human and mouse NPAS1 in their functional domains. 82

Table 3. HTS summary from TSA. 115

Table 4. Lead compounds from the MedChemExpress and MedChemExpress Natural libraries. 123

Table 5. Calculated EC_{50} values for all tested compounds. 131

Table 6. Buffer composition for protein purification and TSA studies. 310

Table 7. hNPAS1, hNPAS3, and hARNT2 primers designed for the Gibson Assembly. 312

Table 8. General PCR reaction mixture. 312

Table 9. General programme for the PCR method. 313

Table 10. Primers' design for constructs used in the SPR assay. 323

Table 11. Primers for ARNT2-eGFP constructs for the TR-FRET assay. 324

List of Appendices

Appendix 1. Superdex 75-pg HiLoad 16/600 calibration curve (MW/Volume). 326

Appendix 2. Sequence analyses of human NPAS1 and NPAS3. 327

Appendix 3. Superdex 200-pg HiLoad 16/600 calibration curve (MW/Volume). 328

Appendix 4. Representative sequence analysis of hARNT2. 329

Appendix 5. Statistics of crystal diffraction and structure refinement for ARNT2-PAS-B ligands bound to mARNT2-PAS-B. 330

Appendix 6. MST dose-response curve for the rest of ARNT2-PAS-B tested compounds. 331

Appendix 7. MST dose-response curve for the hNPAS1-ARNT2 tested compounds. 332

Appendix 8. MST dose-response curve for the hNPAS3-ARNT2 tested compounds. 333

Appendix 9. SPR analyses for ARNT2 tested compounds. 335

Appendix 10. SPR analyses for NPAS1 tested compounds. 336

Appendix 11. SPR analyses for NPAS3 tested compounds. 337

Appendix 12. Representative ARNT2-eGFP construct sequence analysis. 338

Appendix 13. Purification of mHIF2α-ARNT2-eGFP. 339

Appendix 14. Statistics of crystal diffraction and structure refinement for KG-548 analogues bound to mARNT2-PAS-B. 340

Appendix 15. Statistics of crystal diffraction and structure refinement for KG-548 analogues and butein bound to mARNT2-PAS-B. 341

Appendix 16. Statistics of crystal diffraction and structure refinement for butein analogues bound to mARNT2-PAS-B. 342

List of Abbreviations

Abbreviation	Term
AHR	Aryl hydrocarbon receptor
AHRR	Aryl hydrocarbon receptor repressor
ARNT	Aryl hydrocarbon receptor nuclear translocator
ARNT2	Hydrocarbon receptor nuclear translocator 2
ARX	Aristaless-related homeobox gene
ASD	Autism spectrum disorders
BB	Binding buffer
BDNF	Brain-derived neurotrophic factor
bHLH-PAS	Basic helix-loop-helix PER-ARNT-SIM
BLI	Biolayer interferometry
BMAL	Brain and muscle ARNT-like protein
BMP4	Bone morphogenetic protein 4
cAMP	Cyclic adenosine monophosphate
CCM	Core Circadian Modulator
CE	Cation-exchange chromatography
CE-A	Cation exchange buffer A
CE-B	Cation exchange buffer B
CETSA	Cellular thermal shift assay
CGE	Caudal ganglionic eminences
CLOCK	Circadian locomotor output cycles kaput
CNS	Central nervous system
Co-IP	Co-immunoprecipitation

Abbreviation	Term
CoCoA	Coiled-coil coactivator
CBP	CREB-binding protein
CREB	cAMP response element-binding
DMSO	Dimethyl sulfoxide
dTmB	Change in melting temperature calculated from Boltzmann curve fitting
dTmD	Change in melting temperature determined by derivative analysis of the melt curve
E-box	Enhancer-box
eGFP	Enhanced green fluorescent protein
EPO	Erythropoietin
ESI-MS	Electrospray ionisation mass spectrometry
FDA	U.S. Food and Drug Administration
FGF	Fibroblast growth factor
FGFR1	FGF receptor subtype 1
FMER1	Fragile X mental retardation protein translational regulator 1
FP	Fluorescence polarisation
HAT	Histone acetyltransferase
HDACs	Histone deacetylases
HIF	Hypoxia-inducible factor proteins
HRE	Hypoxia-response element
HSP	Chaperone heat shock protein
HTS	High-throughput screening
IMAC	Immobilised metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside

Abbreviation	Term
IR	Infrared (laser)
ITC	Isothermal titration calorimetry
K_D	Equilibrium dissociation constant
MAP	Methionine aminopeptidase
MAPK	Mitogen-activated protein kinase
MGE	Medial ganglionic eminences
MST	Microscale thermophoresis
NAMPT	Nicotinamide phosphoribosyl transferase
NBEA	Neurobeachin
NCOA	Nuclear receptor coactivators
NCOR	Nuclear receptor corepressor
NES	Nuclear export signal
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
Notch	Neurogenic locus notch homolog protein 1
NPAS1	Neuronal PAS domain protein 1
NPAS3	Neuronal PAS domain protein 3
NR	Nuclear receptors
OEA	Oleoylethanolamide
PAINS	Pan-assay interference compounds
PCR	Polymerase chain reaction
PDB	Protein data bank
PHD	Prolyl hydroxylase domain
PLA	Proximity ligation assays

Abbreviation	Term
PVN	Paraventricular nuclei
RU	Response unit
RPD	Repressor domain
RXRs	Retinoid X receptors
SA	Streptavidin
SHH	Sonic hedgehog protein
SIM	Single-minded homologs
SMRT (NCOR2)	Silencing mediator for retinoid and thyroid receptors
SOC	Super optimal broth with catabolite repression
SON	Supraoptic nuclei
SPR	Surface plasmon resonance
SPRY2	Sprouty homolog 2
SST^+	Somatostatin-positive (interneurons)
ST	Soret coefficient
STAT6	Signal transducer and activator of transcription 6
TACC3	Transforming acidic coiled-coil containing protein 3
TAD	Transactivation domain
TAE	Tris-acetate-EDTA
Tb	Terbium
TH	Tyrosine hydroxylase
Trh	Trachealess
TR-FRET/HTRF	Time-resolved fluorescence
TRIC	Temperature-related intensity change
TSA/PTS/DSF	Protein thermal shift

Abbreviation	Term
Tm	Melting temperature of a protein
TXNIP	Thioredoxin interacting protein
UBE3A	Ubiquitin-protein ligase E3A
VEGF	Vascular endothelial growth factor
VIP^+	Vasoactive intestinal peptide-positive (interneurons)
VGF	Nerve growth factor inducible
VHL	Von Hippel-Lindau

PART I: Introduction and Literature Review

Chapter 1: Introduction to the Neuronal bHLH-PAS transcription factors, NPAS1 and NPAS3, and their partner ARNT2

This DPhil project seeks the discovery of first-in-class chemical ligands for neuronal PAS domain protein 1 (NPAS1), its closest paralogue, neuronal PAS domain protein 3 (NPAS3), and their dimerisation partner aryl hydrocarbon receptor nuclear translocator 2 (ARNT2). NPAS1 and NPAS3 are key neuronal transcription factors that, together with ARNT2, form part of the basic helix-loop-helix PER-ARNT-SIM (bHLH-PAS) protein family.[1] These regulatory proteins coordinate intricate gene expression programmes and play vital roles in neurodevelopment, neural regeneration, and synaptic plasticity.[2]

As transcription factors, NPAS1 and NPAS3 regulate the expression of target genes, primarily through direct interaction with their dimerisation partners, ARNT or ARNT2.[3] Upon forming heterodimeric complexes, NPAS1 and NPAS3 can recognise and directly bind specific DNA motifs in promoter or enhancer regions of target genes (Figure 1). Once bound to DNA, NPAS1 and NPAS3 govern the transcription of target genes by recruiting coactivator or corepressor complexes that promote histone acetylation or deacetylation, thereby altering chromatin accessibility to RNA polymerase II. Together with their partner ARNT2, NPAS1 and NPAS3 act as key regulatory factors that integrate diverse cellular and environmental signals to modulate gene expression programmes in neuronal contexts. Importantly, dysregulation of NPAS1, NPAS3, or ARNT2 has been strongly implicated in a

range of neuropsychiatric and neurodevelopmental disorders, including schizophrenia, bipolar disorder, intellectual disability, and autism spectrum disorders, highlighting their promising therapeutic potential with respect to these conditions.[4]

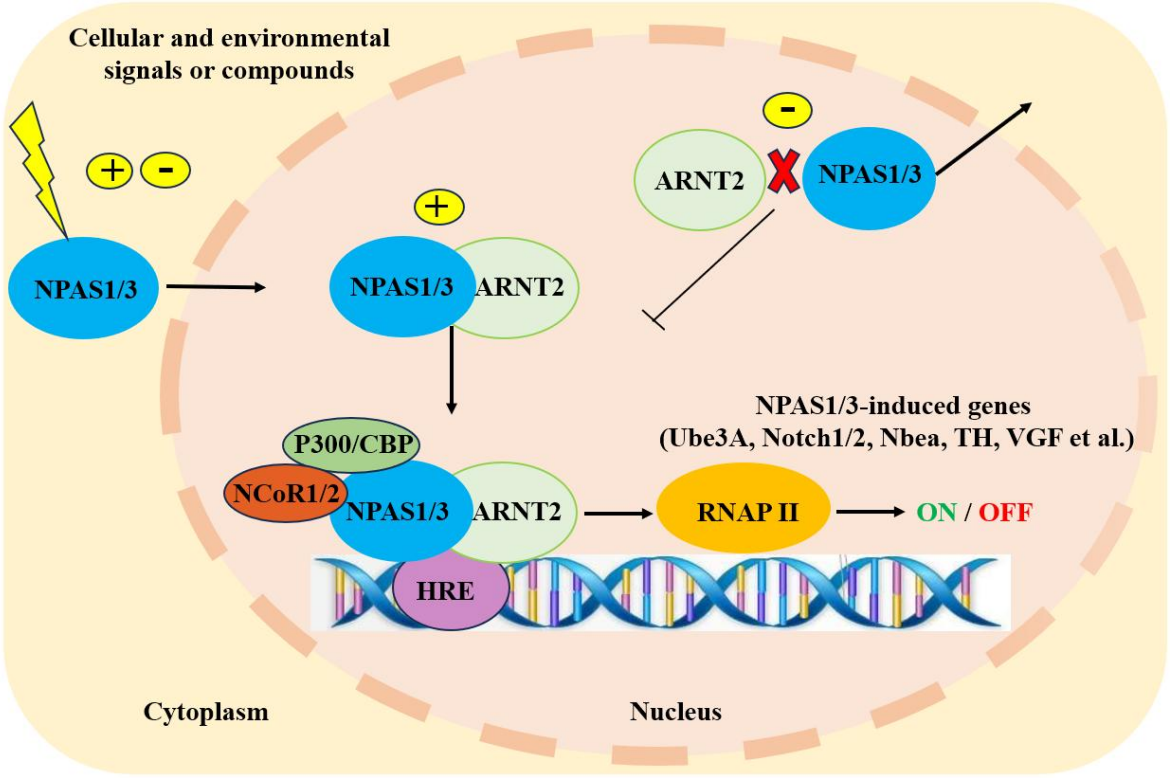


Figure 1. NPAS1 and NPAS3 are transcription factors in neuronal contexts.Schematic illustration detailing the current understanding of NPAS1 and NPAS3 functions. NPAS1 and NPAS3 translocate to the nucleus where, together with their heterodimeric partner ARNT2, they bind specific DNA sequences. Depending on the context, NPAS1/3 recruit coactivator (p300/CBP) or corepressor (NCOR1, NCOR2/SMRT) complexes, thereby modulating chromatin accessibility and influencing RNA polymerase II-mediated transcription.

The human genome encodes over 1,600 transcription factors, which regulate gene networks essential for cellular identity, development, environmental adaptation, and disease progression.[5] Despite their critical biological functions, transcription factors have

traditionally been considered among the most challenging targets for drug development. Nuclear receptors (NRs) have historically been regarded as the only human transcription factor family considered readily druggable, with numerous U.S. Food and Drug Administration (FDA)-approved drugs targeting this class.[6] This druggability is largely attributed to the presence of well-defined ligand-binding domains that accommodate endogenous molecules such as steroid and thyroid hormones, as well as structurally diverse ligands including heme.[7–9]

However, the recent discovery of ligand-binding pockets within bHLH-PAS proteins, along with the FDA approval of small-molecule drugs, such as tapinarof and belzutifan, have highlighted this family as a second major class of transcription factors that is amenable to drug discovery.[10–13] These latest findings underscore the bHLH-PAS family as compelling targets for intervention in complex disorders, with particular relevance to neurological conditions involving NPAS1, NPAS3, and ARNT2.

Chapter 2: Literature review

2.1 NPAS1, NPAS3, and ARNT2 are members of the bHLH-PAS family of transcription factors

The NPAS1 and NPAS3 proteins, together with their partner ARNT2, belong to the bHLH-PAS family of transcription factors. The bHLH-PAS family represents a highly conserved and functionally diverse group of regulatory proteins found exclusively in metazoans (Figure 2A).[14] The bHLH-PAS transcription factors play an essential role in regulating gene expression across various cell types, over the multiple stages of organismal development, and throughout adulthood.[15] From an evolutionary perspective, the bHLH-PAS family appears to have deep roots within the animal kingdom. The bHLH-PAS proteins are widely represented across species, from simple organisms, such as marine microalgae (diatoms and *Nannochloropsis*), to complex vertebrates, including mammals.[16,17] Insects like *Drosophila melanogaster* possess a subset of representative members of the bHLH-PAS family that are homologous to those found in vertebrates, providing some of the earliest insights into the developmental and physiological functions of this transcription factor family (Figure 2B).[18]

The broad phylogenetic distribution and high degree of conservation of bHLH-PAS proteins emphasise their fundamental roles in coordinating the gene expression programmes underlying fundamental biological processes that are conserved throughout evolution, such as development, environmental sensing, and homeostasis.

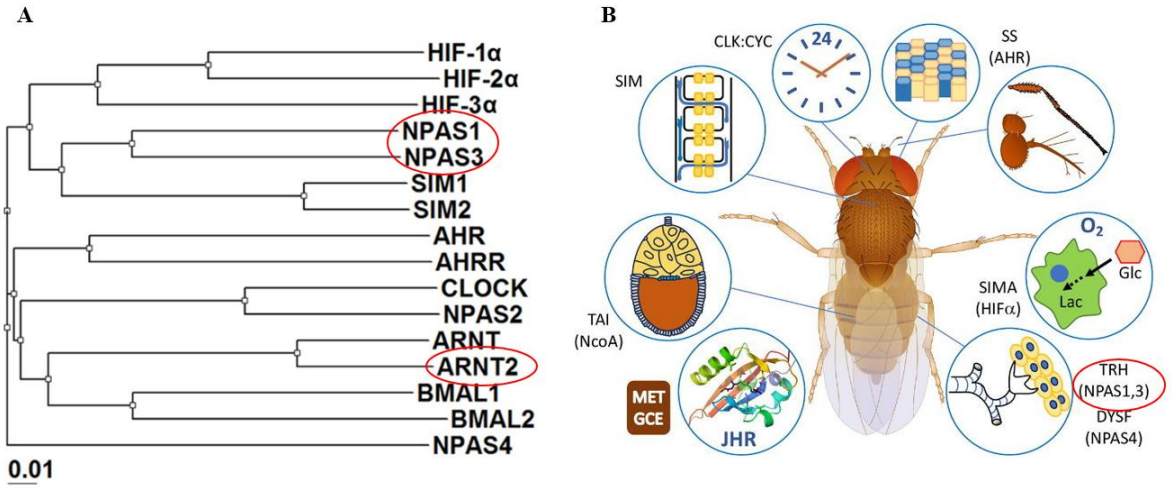


Figure 2. NPAS1, NPAS3, and ARNT2 are members of the bHLH-PAS family.(A) Phylogenetic tree of all mouse bHLH-PAS family members based on protein sequences at their bHLH-PAS-A-PAS-B regions. Figure adapted from Wu et al., 2016.[1] (B) The bHLH-PAS proteins and their functions in *Drosophila melanogaster*. Figure adapted from Tumova et al., 2024.[18]

2.2 Dimerisation of NPAS1 and NPAS3 proteins

Like other members of the bHLH-PAS family, NPAS1 and NPAS3 proteins must dimerise with their partner aryl hydrocarbon receptor nuclear translocator (ARNT), or ARNT2, to create a functional heterodimer capable of DNA binding and influencing transcription.[3] Interestingly, similar to the bHLH-PAS proteins, the NRs also typically form heterodimers to generate functional transcription factors, with retinoic acid receptors, particularly retinoid X receptors (RXRs), acting as the common heterodimerisation partner for many NRs.[19–22] However, in certain instances, NRs also operate as monomers, binding to an extended single half-site.[23]

The mammalian bHLH-PAS proteins can be divided into two classes (Figure 3). Both NPAS1 and NPAS3 belong to Class I, which includes specialised transcription factors that are typically activated in response to a signal, such as oxygen level, neuronal activity, or chemical compounds. The Class I group also consists of aryl hydrocarbon receptor (AHR), three hypoxia-inducible factor proteins (HIF-1α, HIF-2α, and HIF-3α), two other NPAS proteins (NPAS2 and NPAS4), two single-minded homologs (SIM1 and SIM2), and circadian locomotor output cycles kaput (CLOCK).

Class II members are necessary for dimer stabilisation with members of Class I.[3] The Class II group comprises four members: ARNT (also known as HIF-1β), ARNT2 (also known as HIF-2β), brain and muscle ARNT-like protein 1 (BMAL1, also known as ARNTL), and BMAL2 (also known as ARNTL2). The bHLH-PAS proteins from Class II can form a heterodimer with several proteins from Class I. Typically, BMAL1/2 proteins dimerise with CLOCK and NPAS2. However, recent reports suggest BMAL1 can also dimerise with HIF-1α.[24–26] To influence transcription, Class II members can also mediate function by homodimerisation. Although ARNT and ARNT2 are believed to mediate important functions through homodimerisation, experimentally their homodimeric conformations have not been characterised.[27]

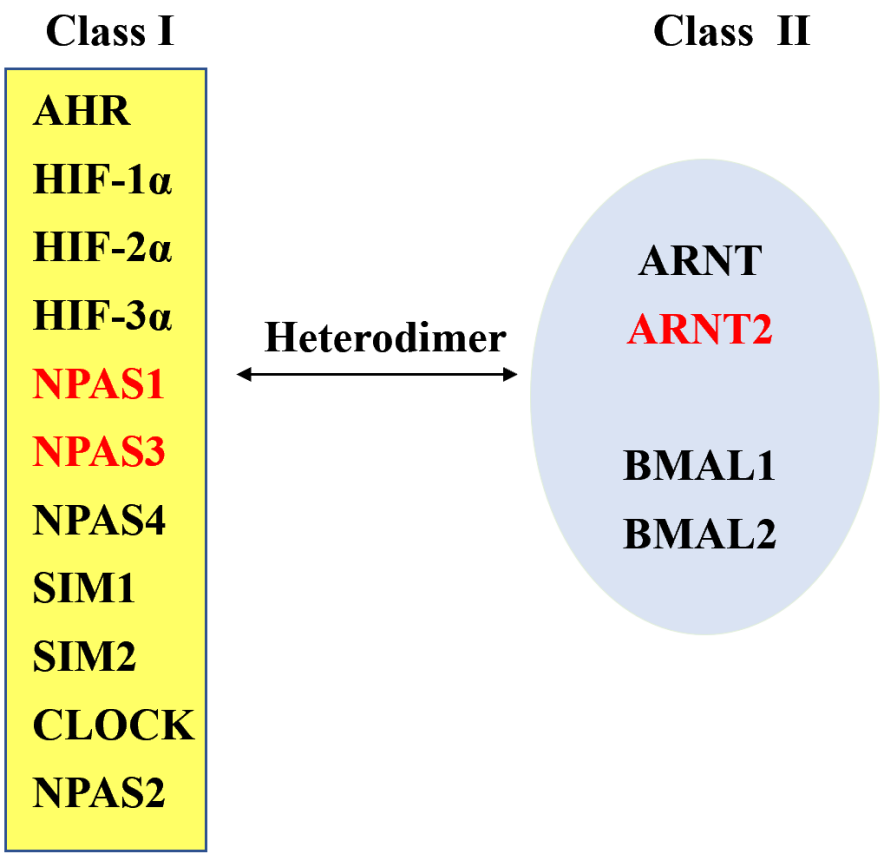


Figure 3. Classification of the bHLH-PAS protein family.The family is divided into Class I and Class II proteins.To form a functional TF, bHLH-PAS proteins must dimerise with a general heterodimeric partner, such as ARNT, ARNT2, or ARNT-like protein BMAL1 or BMAL2.

The aryl hydrocarbon receptor repressor (AHRR) and nuclear receptor coactivators (NCOA1, NCOA2, and NCOA3) differ from most bHLH-PAS proteins.[15] It is therefore difficult to classify these proteins as Class I or Class II. The AHRR protein is known to form a stable heterodimeric complex with ARNT for transcriptional repression.[28] However, a mutant AHRR protein that was unable to dimerise with ARNT to bind DNA, retains the

ability to repress AHR activation, suggesting AHRR might inhibit AHR activity using a novel mechanism.[29]

The NCOAs proteins were originally identified as coregulators of ligand-dependent nuclear hormone receptors.[30] Notably, NCOA1-3 bind to other proteins or coactivators to activate transcription.[31] As such, the NCOA1-PAS-B domain interaction with the transactivation domain of signal transducer and activator of transcription 6 (STAT6) has been characterised.[32] To date, however, a functional dimerisation of NCOA1-3 proteins remains elusive.

2.3 Expression of NPAS1 and NPAS3 in the nervous system

Expression analyses have revealed that NPAS1 and its closest paralogue, NPAS3, are predominantly expressed in the central nervous system (CNS) (Figure 4-5). The mRNA of NPAS1 was first detected at embryonic day 15 of mouse development, immediately following early organogenesis of the brain.[33] The expression of NPAS1 is observed in brain and spinal cord tissues, specifically in deep pyramidal cell layers of the neocortex, the polymorph layer of the hippocampus and dentate gyrus, the amygdala of basal ganglia, the ventroposterior area of the thalamus, as well as in neurons in the medial nuclei of the hypothalamus.[33]

Similar to NPAS1, the mRNA of NPAS3 was detected in early embryonic development, especially in the developing neural tube. Specifically, NPAS3 is expressed in the progenitor domains of the mouse basal ganglia, developing a pontine, neopallial layer of the cortex, cerebellar and medullary neuroepithelium, and to a lesser extent in the differentiating field

of the pons, cerebellum and medulla.[34,35] In the adult mouse brain, both NPAS1 and NPAS3 are primarily expressed in small inhibitory interneurons of the hippocampus, dentate gyrus, piriform cortex, medial septum, and cortical layers 1, 4, and 5.[4,34]

The NPAS1 and NPAS3 proteins have been identified as orthologs of *trachealess* (*Trh*) protein of *Drosophila melanogaster*, exhibiting 45% and 42% amino acid identity, respectively.[36] In flies, *Trh* is expressed in tracheal placodes, the embryonic precursor cells that form the tracheal system, and in a subset of the midline glia, specialised non-neuronal cells in the CNS that are crucial for the proper formation of the nervous system during embryonic development.[37,38] NPAS1 and NPAS3 are also expressed in murine embryonic lungs, promoting early lung development, including myofibroblast distribution, branching morphogenesis, and neuroendocrine cell differentiation.[39,40] The absence of such developmentally expressed transcription factors can alter downstream gene expression and multiple signalling pathways in organogenesis. NPAS3-null newborn mice exhibit respiratory distress and die shortly after birth, showing diminished alveolarisation, increased protein sprouty homolog 2 (*SPRY2*), and decreased sonic hedgehog protein (*SHH*), fibroblast growth factors 9 and 10 (*FGF9* and *FGF10*), and bone morphogenetic protein 4 (*BMP4*) mRNAs, which are crucial for embryonic development that regulates cell growth, specialisation, and body patterning.[41] During lung formation, the expression of NPAS1/3 genes in mice is tightly restricted to a specific developmental window. Thus, the expression of NPAS3 drops significantly to low levels after the first postnatal day, while NPAS1 mRNA was not detected in either adult murine lungs or whole embryos before embryonic day 13.[33]

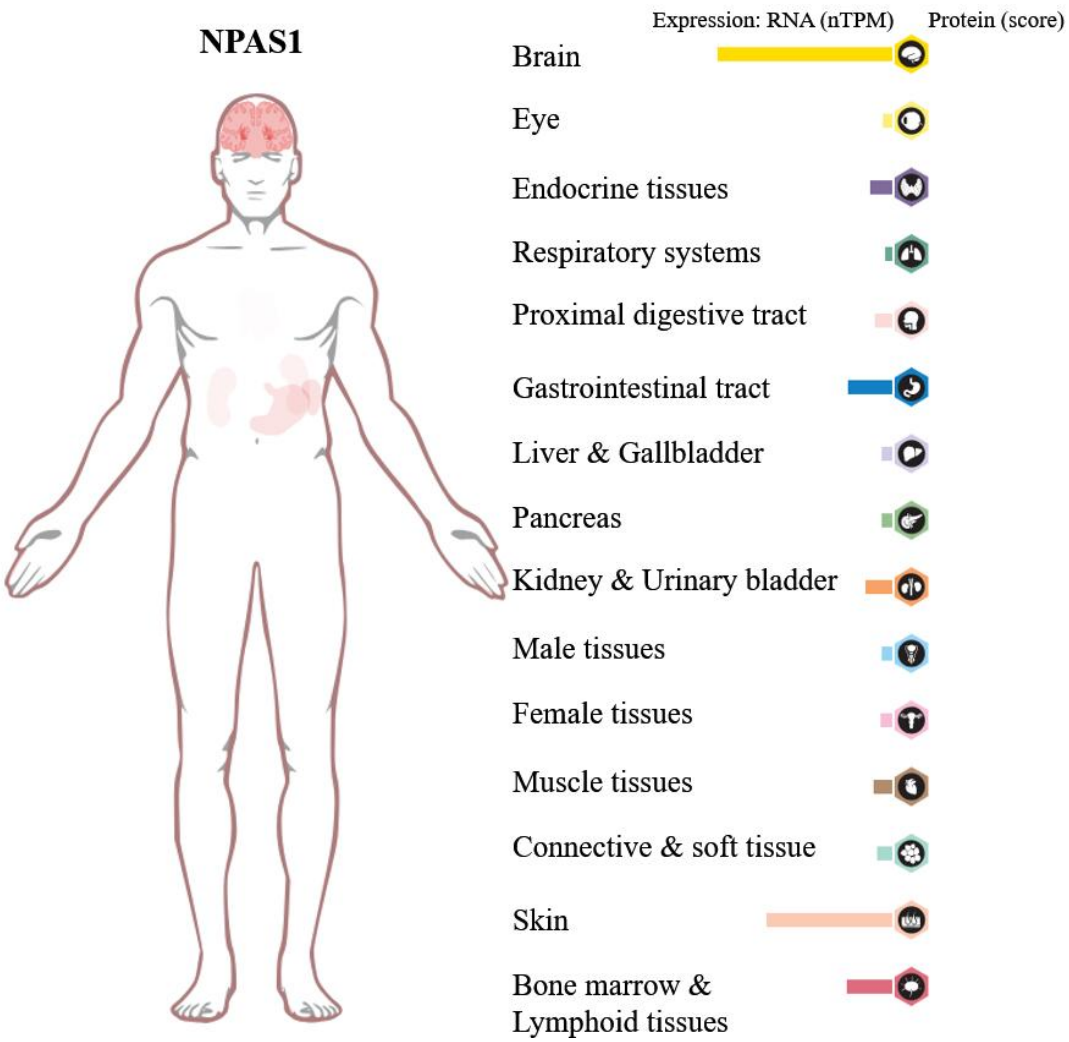


Figure 4. Tissue and organ-wide expression profile of NPAS1 in humans. The summary of the RNA and protein expression levels of NPAS1 across major human tissues and organs is based on data from the Human Protein Atlas. RNA expression levels are represented by transcript abundance (TPM), while protein expression is visualised based on immunohistochemistry staining intensity. The highest NPAS1 expression is observed in neural tissues, particularly in brain regions, with some expression detected in skin. Figure adapted from the Human Protein Atlas.

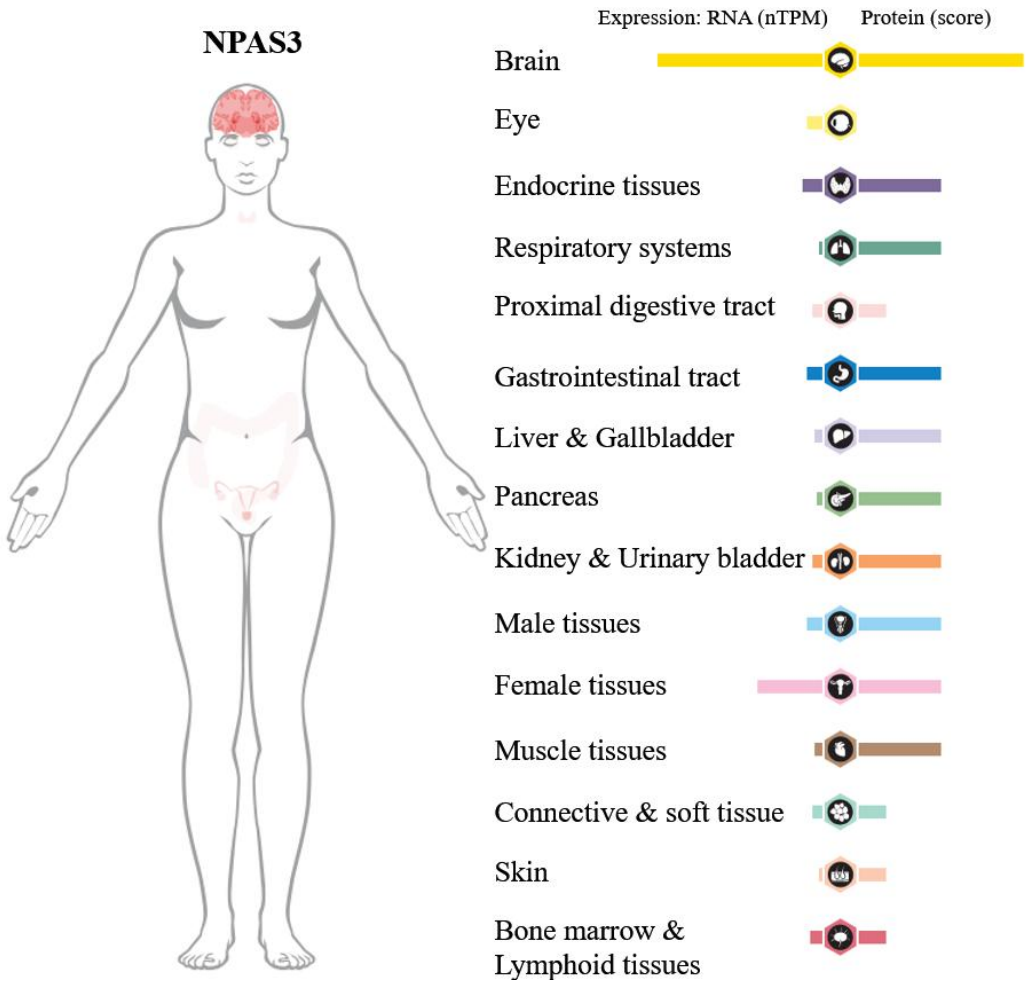


Figure 5. Tissue and organ-wide expression profile of NPAS3 in humans. The summary of the RNA and protein expression levels of NPAS3 across major human tissues and organs is based on data from the Human Protein Atlas. RNA expression levels are represented by transcript abundance (TPM), while protein expression is visualised based on immunohistochemistry staining intensity. NPAS3 shows predominant expression in the CNS, particularly in brain tissues, with minimal expression in most peripheral organs. Figure adapted from the Human Protein Atlas.

NPAS1 and NPAS3 are localised in the nucleus of interneurons.[4,42] However, this localisation is not constitutive. Instead, under basal conditions, the NPAS1 protein exhibits predominant cytosolic localisation. When NPAS1 was tagged to green fluorescent protein (GFP) and

expressed in MN9D murine cells, the fusion protein was excluded from the nucleus.[43] At the N-terminus of NPAS1, there is a short amino acid motif, LSLGLTIL, which corresponds to the nuclear export signal (NES) that mediates the transport of proteins out of the nucleus.[43] The nuclear translocation of NPAS1 requires direct interaction with dimerisation partners such as ARNT or ARNT2. Upon interaction with ARNT via their N-termini, ARNT is able to guide NPAS1 into the nucleus via its own nuclear localisation signal (NLS), masking the NES within the NPAS1.[43] Such a mechanism of nuclear translocation was also observed in *Drosophila*, suggesting a conserved mechanism throughout the animal kingdom.[44]

Compared to NPAS1, NPAS3 is more commonly found in the nucleus. Full-length NPAS3 showed both cytoplasmic and nuclear localisation.[45] The nuclear localisation was enhanced due to NPAS3's co-expression with ARNT.[45] The NPAS3 construct containing the C-terminus region, where an NLS motif has been predicted, was predominantly localised to the nucleus.[46] Recent analysis using NLStradamus showed a potential second NLS in NPAS3, but within the bHLH domain, specifically at 53-65 residues.[47] This prediction further appeared in one study demonstrating somewhat unexpected nuclear localisation of the N-terminal NPAS3 fragment.[48] Most studied bHLH constructs were found to be cytoplasmic, due to a predicted NES motif.[45,49] Such variable NPAS3 localisation may be actively regulated at several levels, which in turn can affect the output of gene regulation. Consistent with other bHLH-PAS constructs, to enter the nucleus and regulate the expression of downstream genes, each of the NPAS1 and NPAS3 constructs must form a stable heterodimer with their partner ARNT or ARNT2.[1,50]

2.4 Aryl hydrocarbon receptor nuclear translocator 2 (ARNT2): a preferred NPAS partner in neuronal contexts

The ARNT and ARNT2 proteins are very similar. The amino acid sequence identity between full-length proteins is 63.8%. Notably, ARNT and ARNT2 share twelve of the thirteen amino acids in the basic N-terminus region, including all amino acids known to bind DNA. Furthermore, the ARNT residues that were previously observed to participate at dimerisation interfaces are fully conserved in ARNT2.[1,51]

Compared to ARNT, ARNT2 is a brain-enriched TF. It has a more restricted tissue distribution and has the greatest expression in brain tissue among all ARNT proteins (Figure 6).[52] It was demonstrated that in the brain, HIF-1α dimerises with ARNT2 even more avidly than with ARNT.[53] While ARNT serves as the canonical partner for HIF-1α in many tissues, ARNT2 is the predominant dimerisation partner in neurons. It is also thought that neuronal Class I bHLH-PAS factors preferentially dimerise with the neuronally enriched Class II factor ARNT2.[54]

During human embryonic development, ARNT2 largely expresses within the CNS, with some expression also observed in the renal tract (Figure 6).[55] *In situ* hybridisation revealed ARNT2 expression in the developing brain and in both the posterior and anterior lobes of the pituitary gland at eight weeks of development. Strong expression was also detected in the hypothalamus, telencephalic cortex, thalamus, the developing neural retina, cochlear ducts of the inner ear, and in the trigeminal ganglion.[55]

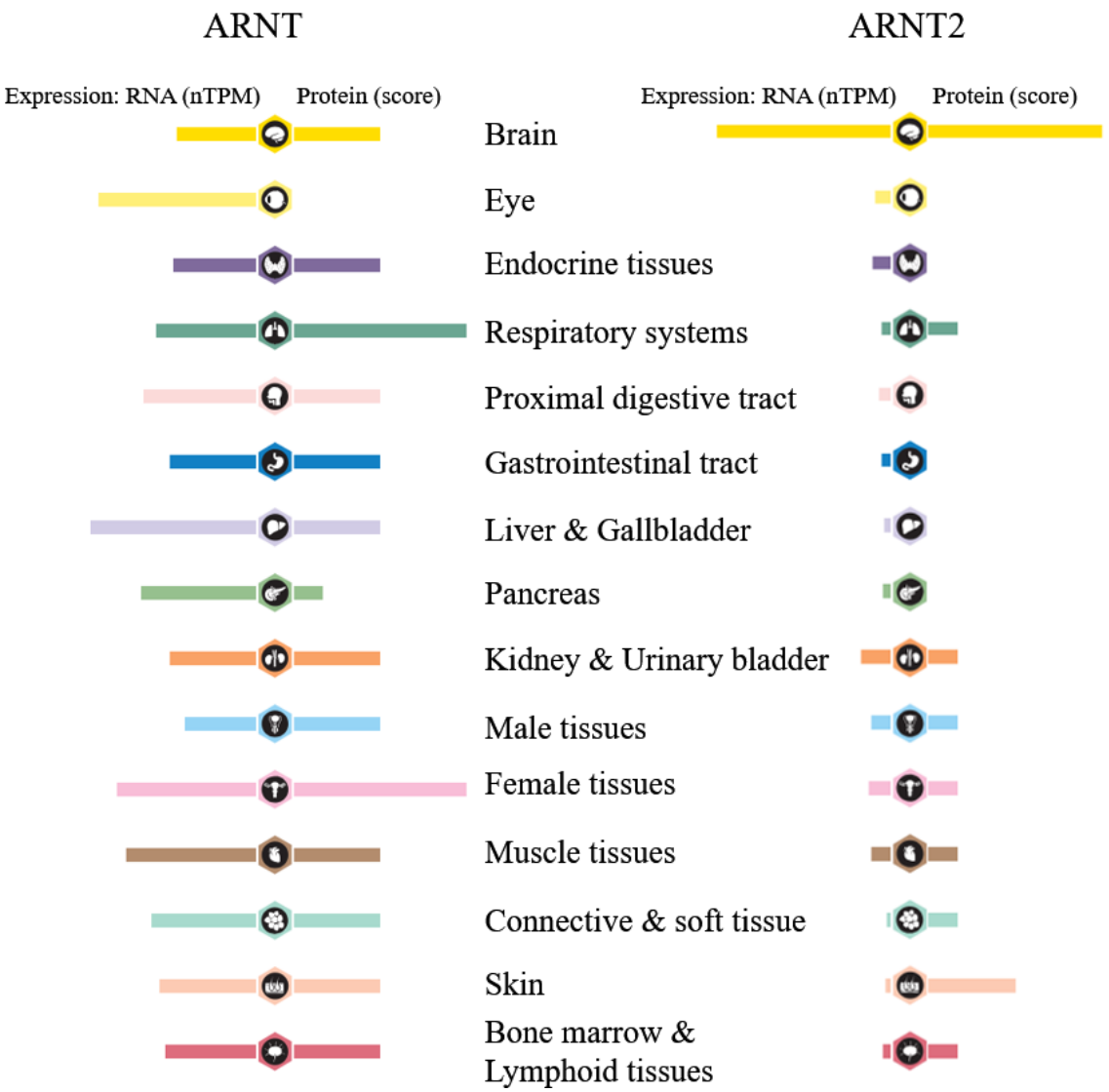


Figure 6. Comparative tissue and organ-wide expression profiles of ARNT and ARNT2 in humans. The summary of the RNA and protein expression levels of ARNT and ARNT2 across major human tissues and organs is based on data from the Human Protein Atlas. ARNT displays broad expression across multiple organ systems, reflecting its ubiquitous regulatory roles. In contrast, ARNT2 shows a more restricted expression pattern, with markedly higher levels in the CNS, particularly in brain regions, and minimal to undetectable expression in peripheral tissues. Expression data are visualised using transcript abundance (TPM) and immunohistochemistry staining. Figure adapted from the Human Protein Atlas.

The expression profile of ARNT2 has also been extensively studied in mouse embryos. Notably, ARNT2 is predominantly expressed in neuronal nuclei and select glial populations across multiple brain regions. As compared to NPAS proteins, ARNT2 is constitutively localised to the nucleus due to its bipartite NLS.[43] During murine neurodevelopment, ARNT2 is extensively expressed in the mitotic cells of the neuroepithelium of the neural tube, and differentiating fields of the neocortex, hypothalamus, thalamus, hippocampus, cerebellum, striatum, inner layer of the retina, and spinal cord, with expression gradually declining postnatally.[56–58]

Several studies have shown that ARNT2 is expressed continuously throughout the lifetime of rodents, and almost exclusively in the brain.[59,60] It has also been demonstrated that ARNT2 is co-expressed with NPAS4 in the individual cortical and hippocampal neurons of wildtype adult mice.[61] ARNT2 is also co-expressed with SIM1 in the developing paraventricular (PVN) and the supraoptic (SON) nuclei of the hypothalamus in newborn mice.[58] Cumulatively, these data strongly suggest that in the neuronal environment, ARNT2 but not ARNT is abundant and is the preferred dimerisation partner for NPAS1 and NPAS3 proteins.

2.5 Domain structure of NPAS1, NPAS3, ARNT2, and other bHLH-PAS proteins

The NPAS1, NPAS3, and ARNT2 proteins share a common domain architecture. A central aspect of the bHLH-PAS protein family is the conserved structural arrangement at their N-terminal regions, with all members containing the same three structural domains. These are

the bHLH (basic helix-loop-helix) DNA-binding domain, and two tandem PAS (Per-ARNT-Sim) domains, which are involved in the dimerisation process.[62]

2.5.1 The bHLH (basic helix-loop-helix) DNA-binding domain

Despite extensive evolutionary diversification, bHLH-PAS proteins are highly conserved among different organisms, including vertebrates and invertebrates.[63,64] This conservation is particularly notable in the N-terminal region of the α-helix within the bHLH DNA-binding domain, which is rich in the basic amino acids arginine and lysine, and is responsible for recognising and binding specific DNA motifs (Figure 7A, B).[65] The bHLH domain typically contains 60 residues. It can be divided into two functionally distinct parts: the basic region responsible for interactions with the phosphate backbone of DNA (approximately 15 residues), and the region consisting of two alpha helices joined by a loop (helix-loop-helix, HLH) that facilitates dimerisation.[66] Dimerisation is essential for effective DNA binding; therefore, the bHLH domain of one bHLH-PAS protein interacts with the bHLH domain of another bHLH-PAS protein to form a functional homodimeric or heterodimeric protein-DNA complex.

The DNA segment to which the bHLH domain binds is a short palindromic sequence with a degenerate CANNTG motif known as the Enhancer-box (E-box). The bHLH-PAS transcription factors exhibit a half-site binding preference to specific E-boxes. The basic region of each monomeric subunit of the dimeric structure recognises the DNA major groove and contacts the half-site in the asymmetric E-box CANNTG sequence. Both monomers engage in recognition of a “CAN” half-site. Therefore, E-boxes could be readily described by their strand-oriented half-sites, such as CAC-CAC, CAG-CAT, and CAG-CAG.[67]

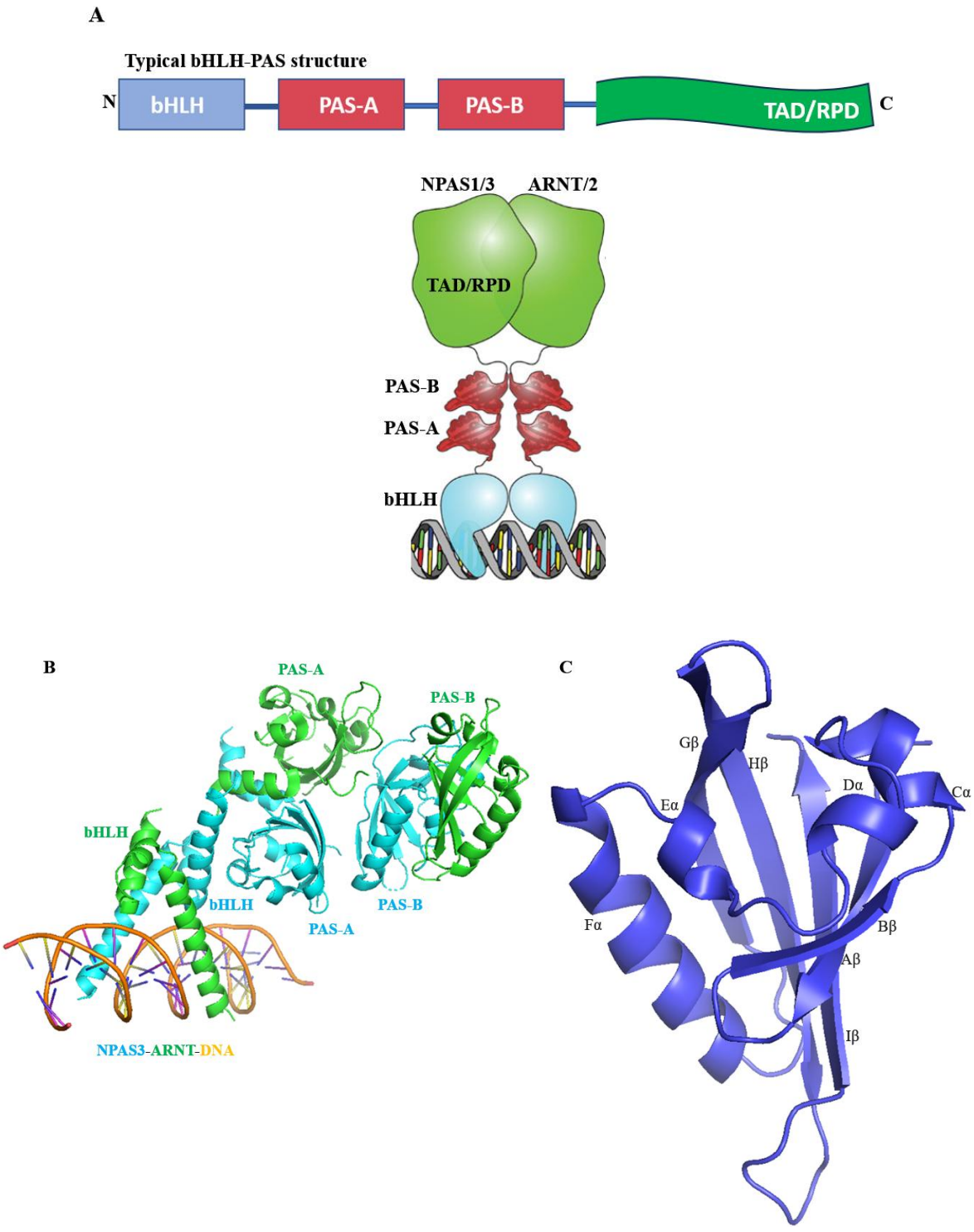


Figure 7. Structure of ordered domains in the bHLH-PAS proteins. (A) Schematic representation of a typical bHLH-PAS protein, illustrating key functional domains including the bHLH, PAS-A, PAS-B, and C-terminal transactivation (TAD) or repressor (RPD) domains. (B) Representative structure of the bHLH-PAS protein, demonstrating the bHLH-

DNA-binding domain and tandemly positioned PAS-A and PAS-B domains. The crystal structure of NPAS3-ARNT-DNA (PDB: 5SY7, resolution 4.2 Å) was used to illustrate bHLH-bound DNA complex. (C) Representative structure of the PAS domain. The crystal structure of the PAS-B domain of ARNT2, featuring characteristic secondary structural elements, like beta strands and alpha helices, labelled in order as Aβ, Bβ, Cα, Dα, Eα, Fα, Gβ, Hβ, and Iβ.

2.5.2 Tandem PAS-A and PAS-B domains

A unifying feature of this protein family is their reliance on PAS domains, which have evolved as versatile molecular sensors and interaction modules for detecting signals across bacteria, archaea, and eukaryotic organisms.[15] These domains enable bHLH-PAS proteins to sense a wide range of environmental cues, such as oxygen levels, light, and small molecules, and to regulate cellular responses through protein-protein interactions. As such, the bHLH-PAS family governs diverse biological processes, including, but not limited to, nitrogen fixation in bacteria, photomorphogenesis (light-responsive developmental growth) in plants, circadian rhythms (day-night cycles) in both prokaryotes and eukaryotes, and ion channel gating in vertebrates.[15,68–71]

In this family of transcription factors, the N-terminal bHLH DNA-binding domain is flanked by tandemly positioned PAS-A and PAS-B domains (Figure 7A, B). These PAS domains are therefore located in the central region of bHLH-PAS proteins, and typically consist of approximately 90-120 residues for PAS-A, and 100-130 residues for PAS-B.[72] [73] Despite a low degree of pairwise sequence identity of about 20%, the PAS-A and PAS-B domains are structurally conserved.[74]

These domains share a common globular core structure consisting of five antiparallel beta strands and four flanking alpha helices in the order Aβ, Bβ, Cα, Dα, Eα, Fα, Gβ, Hβ, and Iβ (Figure 7C). These antiparallel beta strands topologically arrange in the order Bβ, Aβ, Iβ, Hβ, and Gβ (2-1-5-4-3). A subset, such as the PAS-A, has an additional N-terminal helix referred to as A'α that maintains and stabilises PAS-A dimers. The connecting region between PAS-A and PAS-B domains is poorly conserved, and varies greatly in length, across even close species.[72]

Like with other bHLH-PAS family protein constructs, it was demonstrated that both the PAS and the bHLH domains of NPAS1 and NPAS3 proteins were found to be critical for heterodimerisation with a partner protein.[45] It is known that the PAS-A region is significant for the protein-protein interactions with bHLH-PAS partner protein. The PAS-A domain in all bHLH-PAS proteins may form common dimerisation surfaces that are roughly symmetric and mediated mainly by hydrophobic residues.[75] Thus, PAS-A domains are crucial for selecting a dimerisation partner and ensuring specific and high-affinity interactions between bHLH-PAS proteins.[50]

Compared to PAS-A, PAS-B has a more versatile role in bHLH-PAS proteins. While like PAS-A, the PAS-B domain also contributes to dimerisation, PAS-B also plays a crucial role in sensing a variety of exogenous and endogenous stimuli. These interactions can induce energetic and conformational changes that modulate the overall activity of the protein.[50] The PAS-B domain in bHLH-PAS proteins can also mediate protein-protein interactions with non-bHLH-PAS partners, including chaperone heat shock protein 90 (HSP90) that stabilises HIFs against oxygen-independent degradation.[76] Furthermore, HIF-1α interacts with both

chaperonin HSP60, a member of the heat shock protein family, and minichromosome maintenance proteins, as part of its role in regulating cellular responses to hypoxia.[77,78]

The PAS-B domain often contains an allosteric ligand site that impacts the domain's ability to promote binding. For example, the PAS-B domains of AHR and HIF-3α contain sites for endogenous ligand binding, which influence their functional activity.[79,80] By comparison, HIF-2α and BMALs do not have a known endogenous small molecule ligand; however, a ligand pocket exists within their PAS-B domain that can also bind small molecules, influencing their transcriptional activity.[81–84]

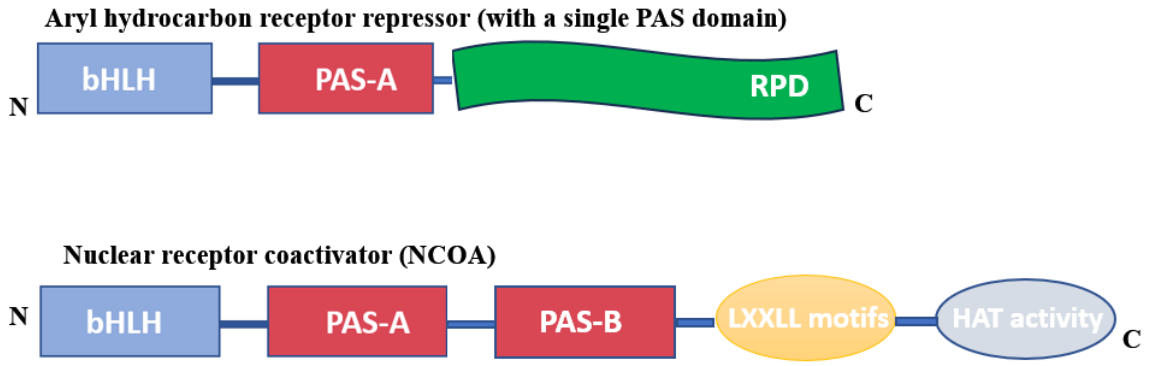


Figure 8. Domain variations in the bHLH-PAS family.Aryl hydrocarbon receptor repressor consists of a single PAS-A domain instead of tandem PAS-A and PAS-B domains. The nuclear receptor coactivators have several LxxLL motifs and a histone acetyltransferase (HAT) activity region.

It is also important to note that significant architectural variations exist within the bHLH-PAS family. For example, AHRR protein contains only a PAS-A domain and lacks the PAS-B (Figure 8). A subset of the bHLH-PAS proteins comprises the three distinct NCOA proteins, which in addition to having PAS-A and PAS-B domains contain multiple LxxLL

motifs, which mediate their engagement with members of the NR family, and also have histone acetyltransferase (HAT) activity regions to promote gene transcription through histone acetylation.[15] The LxxLL motifs (where L represents leucine and X any amino acid), form short α-helical structures, also known as NR boxes, that dock into a hydrophobic groove on the ligand-binding domain of NRs.[85] Upon binding of a ligand, the ligand-binding domain of the NR undergoes a conformational change, most notably involving the repositioning of helix 12.[86] This structural rearrangement creates a surface that favours the recruitment of coactivators, such as NCOAs in the presence of agonists, or corepressors in the presence of antagonists.[9,86] Corepressor complexes may include nuclear receptor corepressor 1 (NCOR1), NCOR2 (also known as SMRT, silencing mediator for retinoid and thyroid receptors), and various HDACs (histone deacetylases).[87,88]

Once recruited, the HAT domain of the coactivator catalyses the acetylation of histone tails, leading to a more relaxed chromatin structure and increased accessibility of DNA to the transcriptional machinery.[86] In addition to histone acetylation, coactivators can also recruit histone methyltransferases and various scaffolding proteins.[9] Together, these actions contribute to dynamic modifications of the histone code, further enhancing chromatin remodelling and transcriptional activation.

2.5.3 C-terminal regulatory domains

At the C-terminal region, the bHLH-PAS family members have so-called C-terminal regulatory regions, which are often found downstream of the PAS-B domain. This regulatory region contains variable transactivation domains (TAD) or repression domains (RPD), which influence the transcription of target genes, and often resembles LxxLL motifs and HAT activity regions found in NCOAs (Figure 9A). In contrast to conserved domains, the length

and composition of these regulatory domains are variable and predicted to be structurally disordered.[62]

The C-terminal regions of bHLH-PAS proteins play a critical role in mediating functional interactions by assembling with other protein regulators. The prolyl hydroxylase domain (PHD) proteins, which are the most well-known regulators for HIF-1α, facilitates hydroxylation of two specific proline residues (P402 and P564) in the oxygen-dependent degradation domain, promoting protein degradation (Figure 9B).[89] Due to the flexible conformation of the C-terminal regions of bHLH-PAS proteins, they are often studied as short peptide sequences bound to such coregulators (Figure 9B, C).

The C-terminal TAD can recruit essential protein coactivators, such as CREB-binding protein (CBP) and p300, enabling transcriptional activation of HIF-1α and other bHLH-PAS proteins.[90,91] The CBP and p300 proteins have a similar structure, and mediate a variety of important functions, including histone acetylation for RNA polymerase II-dependent transcriptional activity.[92] As such, the CBP and p300 proteins participate in most cellular processes, including cell growth, differentiation, development, and the cellular response to stress, including DNA damage.[93]

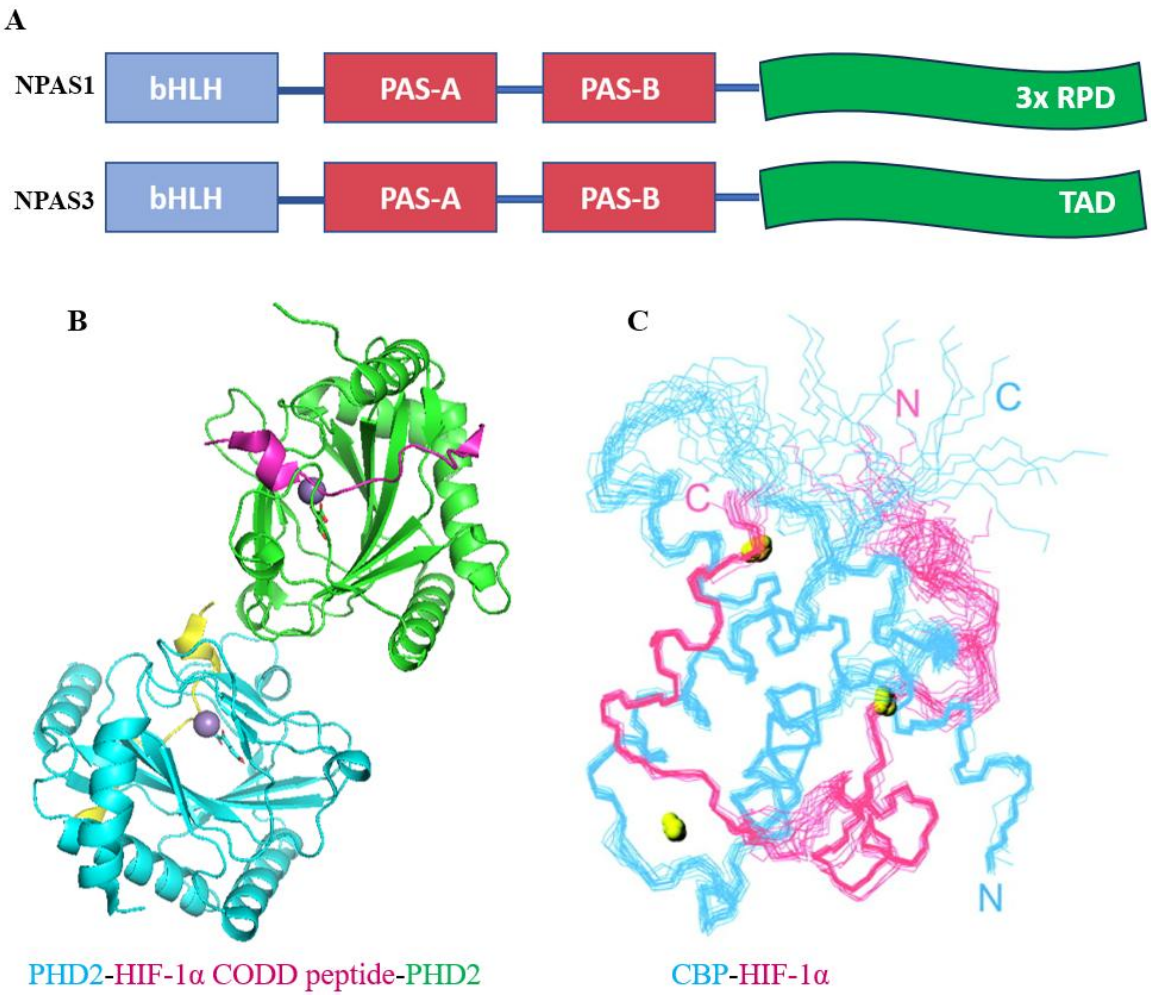


Figure 9. Representative structural analyses of the C-terminal flexible regions of bHLH-PAS proteins.The C-terminal regions are demonstrated by short peptide sequences that are bound to other protein regulators.(A) Schematic representation of NPAS1 and NPAS3 proteins, illustrating key functional domains including the bHLH, PAS-A, PAS-B, and TAD or RPD domains. (B) Structure of human HIF prolyl hydroxylase 2 (PHD2/EGLN1) enzyme in complex with its co-substrate 2-oxoglutarate and the C-terminal oxygen-dependent degradation domain (CODD) segment of HIF-1α, residues 556-574 (PDB: 5L9B, resolution 1.95 Å). (C) NMR structure (PDB: 1L8C) of the complex between the carboxy-terminal activation domain (CAD) of HIF-1α and the zinc-binding TAZ1 (CH1) motif of CREB-binding protein (CBP).

Like most of the bHLH-PAS members, NPAS3 encodes a TAD downstream of the PAS-B domain (Figure 9A).[45] Upon dimerisation with ARNT/2 and binding to the respective response element, TAD contributes to activation of gene expression. As no TADs have been found in NPAS1, it has been suggested that NPAS1 does not play a positive transcriptional role. Instead, three separate non-overlapping RPD domains were observed, suggesting a repressive function of the protein (Figure 9A).

The RPD of bHLH-PAS proteins functions as a modular regulatory element that opposes transcriptional activation by actively recruiting corepressor machinery, and remodelling chromatin into a less permissive state. The RPD can therefore serve as a docking platform for corepressor complexes, such as NCOR1, NCOR2, and different HDACs.[88] By recruiting these corepressor factors, the RPD helps shift the chromatin environment to a hypoacetylated, more compact state, thereby reducing nucleosomal accessibility to RNA polymerase II and limiting transcription initiation. The precise molecular interactions between the RPD and NCOR/SMRT complexes, however, have not been well characterised for bHLH-PAS proteins.

Supporting its role as a negative transcriptional regulator, NPAS1 was further shown to inhibit the expression of erythropoietin (*EPO*), suppress its target gene tyrosine hydroxylase (*TH*), and partially reduce the transactivation activity of both ARNT and ARNT2.[1,42,43] Strikingly, the same mutations in ARNT and NPAS1 proteins, which have been shown to destabilise the heterodimer, weakened NPAS1's ability to repress *TH* transcription. This effect was observed in both TH-driven and hypoxia-response element (HRE)-driven reporter assays when compared to the repression seen with the wildtype proteins.[1] These findings imply that the integrity of the NPAS1 heterodimer is crucial for its transcriptional functions.

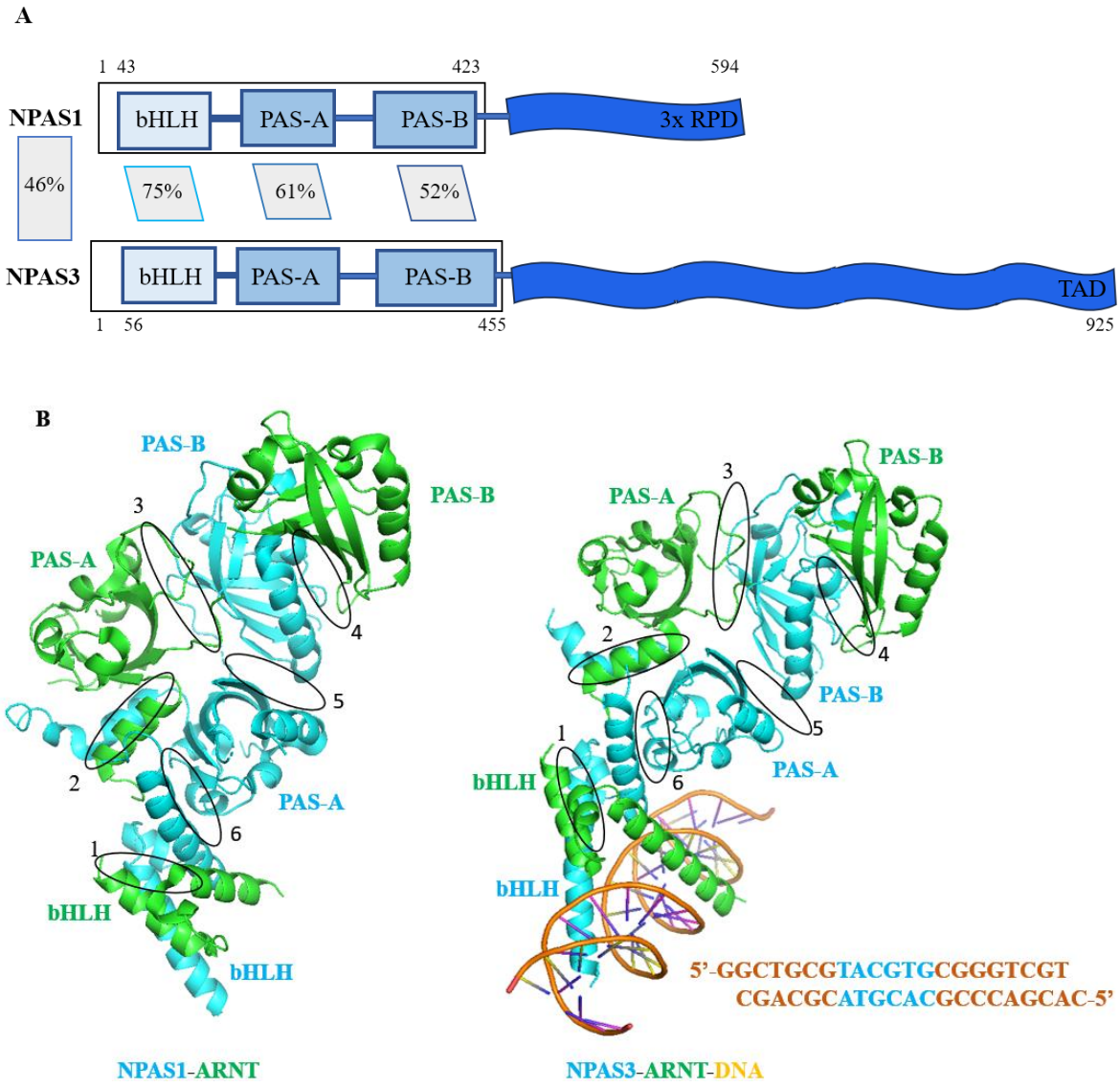
2.6 Structural analyses of NPAS1, NPAS3, and ARNT2 proteins

2.6.1 Crystal structures of NPAS1-ARNT and NPAS3-ARNT heterodimers

To date, no crystal structures are available for the individual full-length NPAS1 or NPAS3 proteins, their isolated domains, or their heterodimeric complexes with the partner protein ARNT2. However, structures of NPAS1 and NPAS3 heterodimers with the partner ARNT have been reported, providing significant insight into their protein structure and architecture.[1]

The resolution of the NPAS1-ARNT and NPAS3-ARNT crystal structures were 3.2 Å and 4.2 Å, respectively. These resolutions are in a similar range to the previously reported crystal structures of multidomain heterodimeric complexes, including HIF-1α-ARNT (3.9 Å), HIF-2α-ARNT (3.6 Å), AHR-ARNT (4.0 Å), and NPAS4-ARNT (4.0 Å). This resolution range enables the clear interpretation of domain structures in the electron density maps.

The NPAS1-ARNT and NPAS3-ARNT crystal structures reveal that the two proteins are structurally similar to each other (Figure 10).[1,83] Indeed, with a sequence identity of 46% between full-length proteins, NPAS1 is quite similar to NPAS3 (Figure 10A). Comparing structured domains, mouse NPAS1 and NPAS3 share 75% identity of amino acids within their bHLH domain, including residues that recognise HRE. The HRE is a DNA sequence of 5'-RCGTG-3', where R represents an adenine or guanine base.[1] NPAS1 and NPAS3 also showed 61% and 52% identities between their PAS-A and PAS-B, respectively.[1]



Interface	Domains
1	bHLH/bHLH
2	PAS-A/PAS-A
3	PAS-A/PAS-B
4	PAS-B/PAS-B
5	PAS-A/PAS-B
6	bHLH/PAS-A

Figure 10. Structural analysis of NPAS1 and NPAS3 proteins. (A) Protein homology between NPAS1 and NPAS3 proteins. The percentage of amino acid identities between the proteins and their domains is shown with light grey squares. (B) Crystal structures of the NPAS1-ARNT and NPAS3-ARNT-DNA complexes with detailed interactions at each of six interfaces. Both NPAS1 and NPAS3 are shown in cyan. ARNT is illustrated in green. Each junction is circled in black, and each interface is listed in the table according to the protein colour. The NPAS1-ARNT structure was solved at a resolution of 3.2 Å (PDB: 5SY5), and the NPAS3-ARNT-DNA complex at a resolution of 4.2 Å (PDB: 5SY7).

The contiguous bHLH-PAS-A-PAS-B constructs, which were used for the crystallisation studies, indicated that each domain of NPAS1 and NPAS3 forms mutual contacts or domain-domain interactions to form a contiguous surface and establish a well-integrated quaternary architecture (Figure 10B). The bHLH, PAS-A, and PAS-B domains of ARNT have no intra-molecular contacts. Instead, they rotate and twist around the outer surface of the NPAS1 and NPAS3 proteins. The flexibility and the lack of physical constraints within the ARNT domains could suggest the adaptability to accommodate its ten different dimerisation partners from Class I of the bHLH-PAS family.

These structural arrangements form six domain-domain junctions that are very similar between NPAS1 and NPAS3, as well as to the overall architecture of the previously discovered HIF-1α-ARNT and HIF-2α-ARNT.[83] The only difference is the PAS-B domain of ARNT, which is slightly displaced within the heterodimeric complex with NPAS1, further suggesting the flexibility of ARNT protein in accommodating its heteromeric partner (Figure 11A).

These six domain-domain interfaces in NPAS1/3-ARNT complexes are crucial for establishing and maintaining heterodimer stability. A single or double mutation in ARNT at interfaces 1-4 can significantly destabilise the heterodimer interactions with both NPAS1 and NPAS3.[1] Point mutations in NPAS1 at interfaces five and six have also been shown to compromise heterodimerisation stability with ARNT. Interestingly, the same ARNT mutations also demonstrated heterodimer destabilisation with SIM1 protein, further suggesting that these heterodimers share a similar overall architecture.[1]

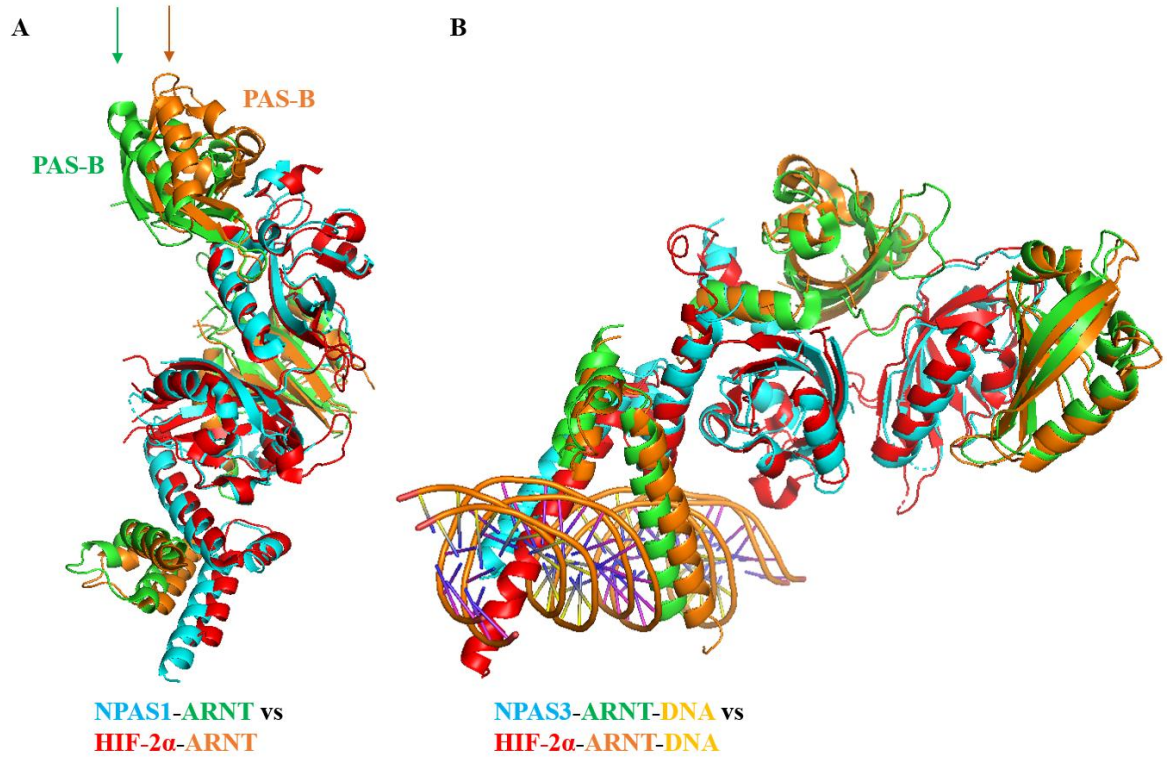


Figure 11. Comparison of bHLH-PAS proteins.(A) Superposition of NPAS1-ARNT (PDB: 5SY5, 3.20 Å) and HIF-2α-ARNT (PDB: 4ZP4, 2.35 Å) heterodimers. The top arrows indicate the shift in position for the PAS-B domain of ARNT. (B) Superposition of NPAS3-ARNT-DNA (PDB: 5SY7, 4.20 Å) and HIF-2α-ARNT-DNA (PDB: 4ZPK, 3.60 Å) heterodimers.

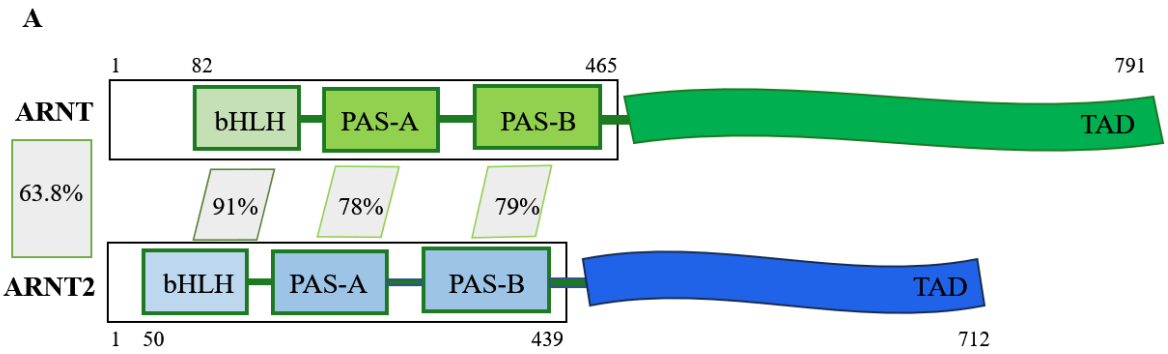
The DNA response elements for NPAS1-ARNT are known to match the consensus HRE sequence of HIFs.[1] The NPAS3-ARNT crystal was solved at 4.2 Å resolution in complex with the HRE element, indicating that NPAS1 and NPAS3 are highly related to HIF transcription factors (Figure 10B). The cooperation between the two subunits of NPAS3-ARNT creates a DNA-reading head that is very similar to those in HIF-2α-ARNT-DNA complexes, enabling the direct readout of the HRE sequence (Figure 11B).

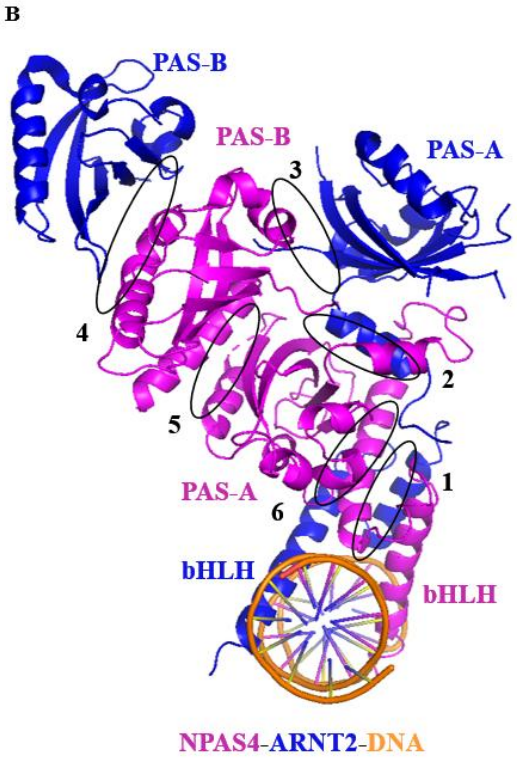
2.6.2 Crystal structure of ARNT2 protein in its complex with NPAS4

Like NPAS1 and NPAS3, ARNT and ARNT2 are very similar proteins, sharing 63.8% amino acid sequence identity (Figure 12A). Over the bHLH and PAS regions, the overall sequence is 80% identical, specifically 91% for bHLH, 78% for PAS-A, and 79% for PAS-B.[51] The major difference between the two proteins is at regions outside of the bHLH-PAS motif, which exhibit low sequence identity and differ in length.[59]

For ARNT, the PAS-A and PAS-B domain structures are characterised, and homodimerisation via PAS-A has been revealed. The heterodimeric crystal structures of the protein bHLH-PAS region with HIF-1α, HIF-2α, HIF-3α, AHR, AHRR, NPAS1, NPAS3, and NPAS4 dimerisation partners have also been determined. For ARNT2, however, the available protein crystal structure is limited to its DNA-bound bHLH-PAS heterodimeric complex with NPAS4, which was resolved at 4.3 Å resolution (Figure 12B).[51] The structure reveals that NPAS4 domains are in mutual contact, like the previously observed NPAS and HIF proteins. The three corresponding domains of ARNT2 were spatially separated and twisted around the NPAS4 domains, resembling previously observed ARNT architecture.

Since ARNT and ARNT2 proteins share high sequence identity at the bHLH-PAS region, it is not surprising that both proteins as a whole, as well as each of their individual domains, exhibit only minor structural differences when superimposing DNA-bound NPAS4-ARNT2 with DNA-bound NPAS3-ARNT structures (Figure 12C). One difference is that the PAS-A domain of ARNT2 exhibited a positional displacement, characterised by a ~30° twist, as compared to ARNT (Figure 12C). When the PAS-A domain undergoes this conformational change, it creates a unique junctional interface between NPAS4 and ARNT2. Since PAS-A plays a major role in the dimerisation process, such flexibility in orientation would suggest a better accommodation of a protein partner.





Interface	Domains
1	bHLH/bHLH
2	PAS-A/PAS-A
3	PAS-B/PAS-A
4	PAS-B/PAS-B
5	PAS-A/PAS-B
6	bHLH/PAS-A

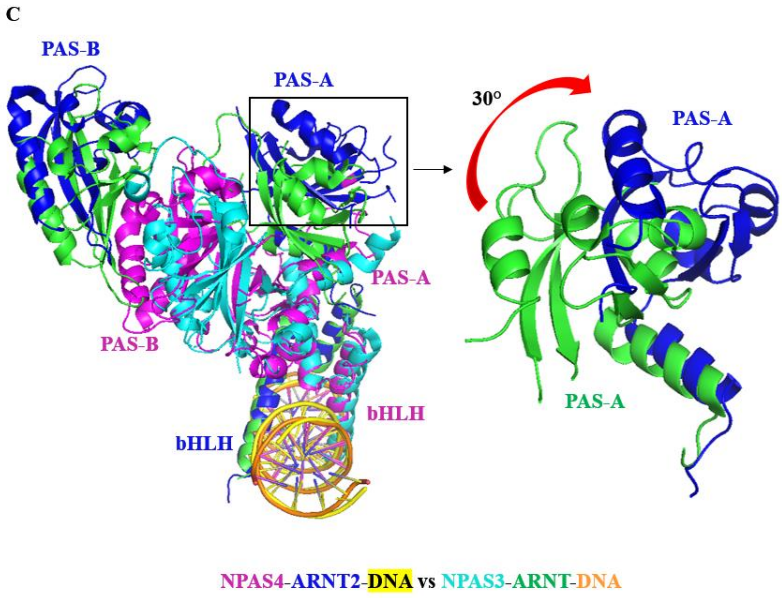
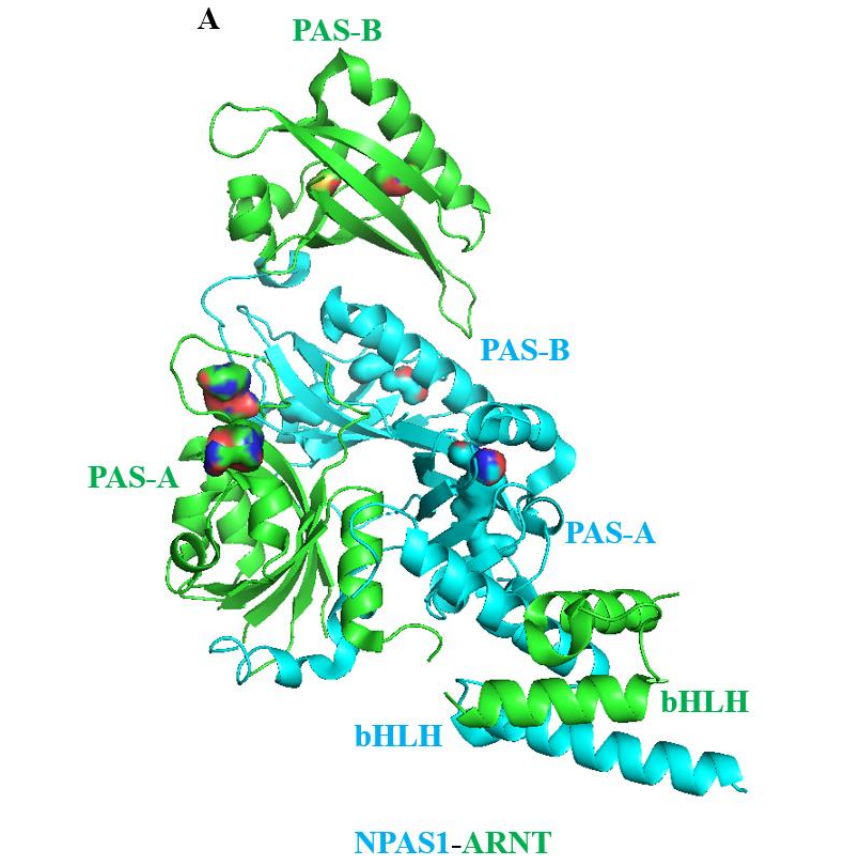


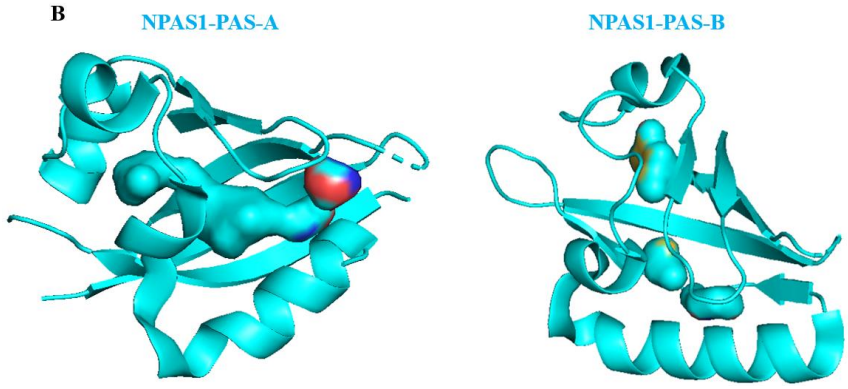
Figure 12. Structural analysis of ARNT2 protein. (A) Protein homology between ARNT and ARNT2 proteins. The percentage of amino acid identities between the proteins and their domains is shown with light grey squares. (B) DNA-bound ARNT2 crystal structure in a heterodimeric complex with NPAS4 diffracted to 4.3 Å (PDB: 7XI3). (C) Structural

comparison of ARNT2 and ARNT. Superimposition of NPAS4-ARNT2-DNA (PDB: 7XI3) and NPAS3-ARNT-DNA (PDB: 5SY7) structures. On the right is an enlarged view of the superimposed PAS-A domains of ARNT and ARNT2, demonstrating a ~30° twist.

2.6.3 Ligand binding pockets within NPAS1, NPAS3, and ARNT2

Previously, it was shown that both HIF-1α-ARNT and HIF-2α-ARNT have four hydrophobic ligand binding pockets encapsulated within the two PAS domains of HIFs, and the two PAS domains of ARNT.[83] Strikingly, the protein crystal structures revealed that the conserved multi-domain architecture of NPAS1 and NPAS3 contain similar distinct hydrophobic pockets within the PAS-A and PAS-B domains (Figure 13A).[1] Structurally, each of the PAS domains adopts a cupped conformation, with the β-sheet forming the base and the short α-helices creating an opposing surface to enclose a central cavity. The volumes of these internal cavities within each PAS-A and PAS-B domain were previously measured at 190 $Å^3$ and 180 $Å^3$ for NPAS1, and 100 $Å^3$ and 230 $Å^3$ for NPAS3 protein, respectively (Table 1).[1]


A
PAS-B
PAS-B
PAS-A
PAS-A
bHLH
bHLH
NPAS1-ARNT


B
NPAS1-PAS-A
NPAS1-PAS-B

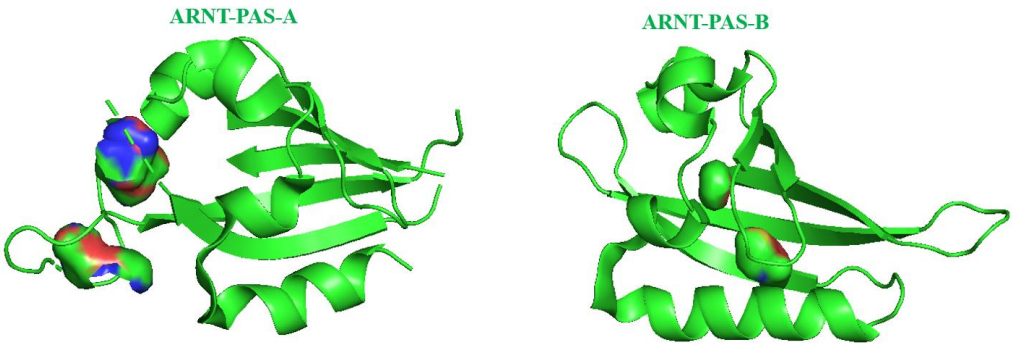


Figure 13. Representative structural analyses of ligand binding pockets for the NPAS1-ARNT heterodimer.(A) Visualisation of pocket cavities within the NPAS1-ARNT heterodimer using the PyMOL program (PBD: 5SY5). (B) Relative position of internal cavities within each PAS-A and PAS-B domain. The NPAS1 protein is in cyan, and the ARNT protein is in green.

Volumes of the PAS-A and PAS-B ligand-binding pockets in bHLH-PAS proteins (Å³)												
Location	**NPAS1**	**NPAS3**	**NPAS2**	**HIF-1α**	**HIF-2α**	**HIF-3α**	**CLOCK**	**BMAL1**	**SIM1**	**SIM2**	**ARNT**	**ARNT2**
PAS-A	190	100	230	100	150	170	120	200	210	210	110	
PAS-B	180	230	170	160	370	590	140	220	370	310	210	160*

Table 1. Ligand-binding pocket volumes of bHLH-PAS family proteins. Pocket volumes are reported in Å³. *The PAS-B pocket of ARNT2 was previously calculated using PyVOL, while those of other bHLH-PAS proteins were calculated using the CASTp program.[1,94]

Both ARNT and ARNT2 also contain sizable pocket cavities within the PAS-A and PAS-B domains. The PAS-B pockets of ARNT and ARNT2 were measured at 210 Å^3 and 160 Å^3, respectively (Table 1).[1,94] The PAS-A domain pocket was measured at 110 Å^3 for the ARNT protein, whereas it has yet to be calculated for the ARNT2 protein.[1]

The internal pocket volumes in the bHLH-PAS family range from 100-700 Å^3 (Table 1). The side-by-side comparison of the PAS-A and PAS-B pockets from several bHLH-PAS

members indicates that the PAS-B domain is generally larger in size. The PAS-B domain of HIF-3α possesses the largest internal cavity among the protein family members, enabling it to accommodate the 18-carbon chain of the endogenous lipid ligand oleoylethanolamide (OEA).[80] While the hydrophobic tail of OEA is deeply buried within the cavity, the hydrophilic ethanolamine head group remains largely exposed to the solvent; such mode of binding has been seen also in other fatty acid binding proteins.[95–97] Upon OEA binding, the volume of the PAS-B pocket increases significantly, from approximately 280 Å³ in the apo state to more than 700 Å³ in the ligand-bound complex, representing an estimated 150% expansion in cavity size.[80] A similar enlargement of the ligand-binding pocket is also observed in BMAL1, where binding of the small molecule ligand Core Circadian Modulator (CCM) expands the central pocket volume nearly three-fold, from 106 Å³ to 320 Å³.[84] There is slight variation in the reported volumes calculated for the PAS domain cavities, particularly noticeable for HIF-3α and BMAL1. This could be due to differences in the algorithms used by various programs for the calculations.

The pocket sizes for NPAS1/3 and ARNT/2 are at the lower end of the range among the family. In NPAS1, the PAS-A and PAS-B domains have similarly sized cavities, showing nearly identical volumes (Figure 13A). In contrast, NPAS3 displays a marked difference between the two domains, with the PAS-B cavity being substantially larger than PAS-A, a trend consistent with other proteins in this family. Notably, the ARNT PAS-A pocket is extremely small, making it difficult to detect any significant internal cavity (Figure 13B).

These findings strongly suggest that all three proteins of interest, NPAS1, NPAS3, and their partner ARNT2, can accommodate ligands within their PAS domain architecture. The ability

to bind small molecules would be extremely beneficial for potential physiological and pharmacological regulations of the NPAS1, NPAS3, and ARNT2 proteins.

Importantly, in all PAS-A and PAS-B pockets within the bHLH-PAS family, the residues lining interior cavities are mostly hydrophobic.[1] This hydrophobic character would allow for favourable van der Waals interactions with lipophilic ligands. The desolvation of hydrophobic ligands upon binding can further enhance the thermodynamic favourability of ligand association within these cavities. Additionally, each PAS domain relies on a unique constellation of amino acids to shape its internal cavity, enabling the pocket to bind selectively to distinct sets of both endogenous and chemical ligands.

2.7 Targeting ligand-binding pockets within bHLH-PAS proteins

The bHLH-PAS transcription factors act as molecular sensors, detecting both environmental and developmental signals within cells, and transmitting them to the nucleus to activate or repress target genes. Such regulation enables them to influence various cellular processes and a wide range of biological functions, from cell specification and differentiation to regulation of circadian rhythms, response to hypoxia, and the proper function of organs such as the brain and liver (Figure 14).[98–100]

Given that bHLH-PAS proteins play a pivotal role in diverse physiological processes, their dysfunction is regarded as a hallmark of numerous human disorders, including cancer, as well as cardiovascular, metabolic, and neurological disorders. The presence of potential ligand-binding pockets within the bHLH-PAS structural framework has established these proteins as promising therapeutic targets and biomarkers in a range of disease pathways,

thereby driving drug discovery efforts to uncover endogenous and synthetic modulators for therapeutic intervention in associated human disorders.

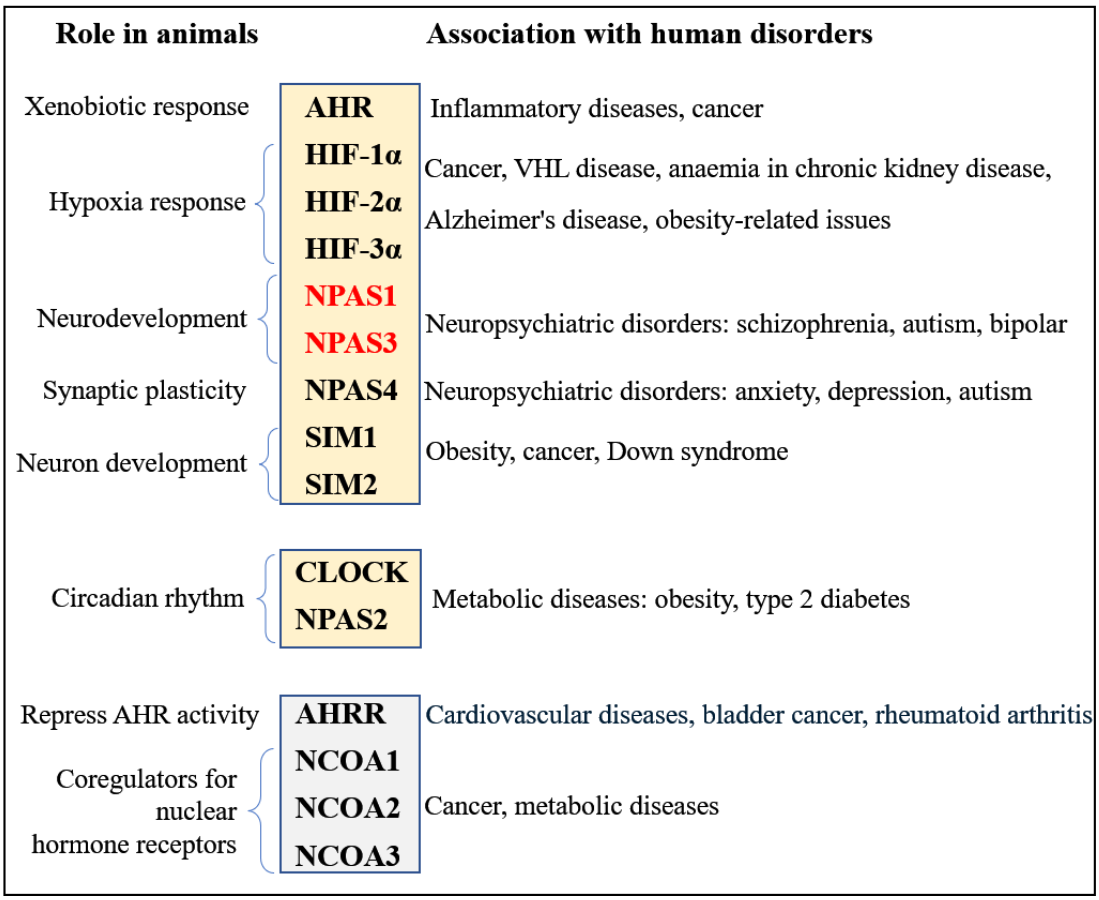


Figure 14. The functional role of bHLH-PAS proteins and their disease association.

As a ligand-dependent TF, AHR is activated by endogenous stimuli, such as tryptophan metabolites or environmental hydrocarbons. By binding xenobiotics, AHR initiates the transcription of a variety of detoxification enzymes within the liver, and modulates the differentiation of immune cells to cope with xenobiotic and toxic stresses.[101] Regulating a diverse range of genes involved in xenobiotic metabolism, skin homeostasis, and immune

response, AHR has been implicated in diseases that are driven by immune and inflammatory processes, including rheumatoid arthritis, asthma, allergic responses, and psoriasis. [102–104]

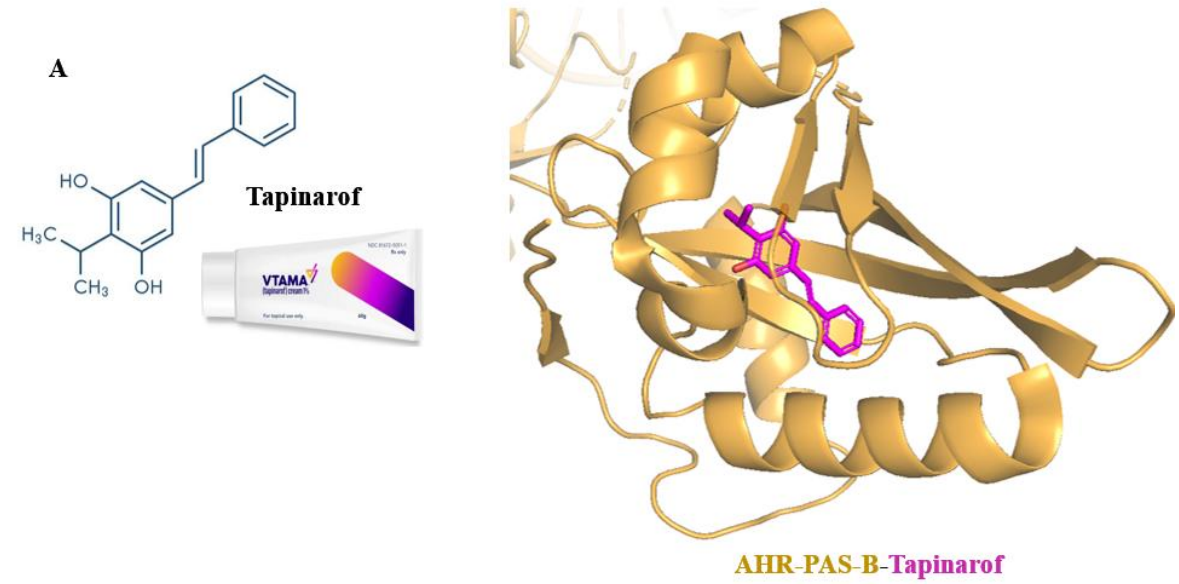
The FDA has approved tapinarof (benvitimod), a first-in-class, small-molecule topical therapeutic AHR agonist, for the treatment of psoriasis and atopic dermatitis (Figure 15A).[10,105] Tapinarof binds directly to the AHR-ARNT heterodimer, interacting specifically with the PAS-B domain, which is known as the ligand-binding region of AHR.[106] Since the success of tapinarof as an FDA-approved drug, a number of other AHR synthetic ligands have been described in the literature, including antagonists, allosteric modulators, and molecules with sub-nanomolar potency that bind to AHR. These have been suggested to be of potential therapeutic value in diseases ranging from ultraviolet (UV) induced skin damage, to Duchenne muscular dystrophy.[107–109]

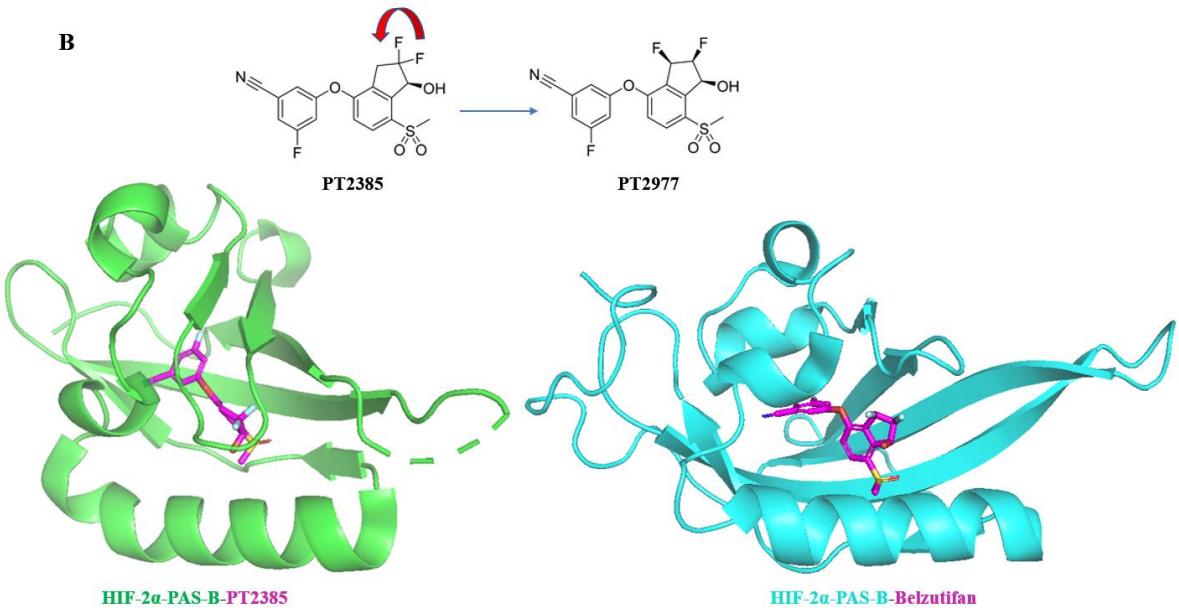
AHR might also play a role in tumorigenesis. The discovery of selective tryptophan-derived AHR ligands such as kynurenine, kynurenic acid, and 6-formylindolo[3,2-b] carbazole, which are all active against multiple types of tumour, is a subject of intensive research.[110,111]

HIFs have also been recognised as important cancer drug targets, as elevated levels of HIF-1α and HIF-2α correlate strongly with tumour growth, metastasis, and resistance to chemotherapy.[112,113] Activated by hypoxia, HIFs trigger cellular responses that maintain oxygen homeostasis and allow cells to survive in oxygen-deprived conditions. To counteract the low oxygen stress, HIFs up-regulate gene expression that enhances oxygen delivery to cells, including transcriptional programmes of erythropoiesis, angiogenesis, and metabolic adaptation.[114]

The PAS-B domain of HIF-2α possesses a water-filled hydrophobic cavity that can bind small molecules, which in turn could disrupt the HIF-2α dimer complex with ARNT.[82,115] The discovery of benzoxadiazole-1α and tetrazole-2 culminated in the development of PT2385 and FDA-approved belzutifan (Welireg, MK-6482, PT2977), the first HIF-2α small-molecule antagonists developed for treatment of clear cell renal cell carcinoma and Von Hippel-Lindau (VHL) disease (Figure 15B).[11–13,116] By contrast, small-molecule agonists that can enhance HIF-2α function as the key driver of *EPO* expression could be used to treat anaemia in patients with chronic kidney disease.[117] Additionally, several prolyl hydroxylase domain enzyme inhibitors, such as daprodustat, molidustat, roxadustat, and vadadustat are undergoing clinical testing for anaemia, because they upregulate HIF-2α activity.[118,119]

The CLOCK gene is a major regulator of circadian rhythm. Together with its paralog NPAS2, CLOCK regulates the molecular circadian clock through transcription-translation feedback loops.[120] Their complexes with BMAL1 drive the rhythmic transcription of key circadian target genes, whose protein products accumulate, form inhibitory complexes, and feed back to repress their own transcription, establishing a roughly 24-hour cycle.[121] The CLOCK-BMAL1 heterodimer regulates the circadian rhythms that play a fundamental role in broad physiological processes, including sleep-wake cycles, hormone secretion, immune function, and metabolism. Disruption of these rhythms leads to a spectrum of health conditions and human diseases, including cardiometabolic disorders, autoimmune diseases, cancers, neurodegenerative disorders, and psychiatric and depression conditions.[122–127] The recent discovery of CCM, a small molecule ligand targeting the PAS-B domain of BMAL1, underscores the potential benefits of BMAL1-CLOCK-directed modulators for addressing numerous disorders (Figure 15C).[84]


A
HO
H3C
CH3
OH
Tapinarof
VTAMA
AHR-PAS-B-Tapinarof


B
F
F
OH
N
O
F
S
O
O
PT2385
PT2977
HIF-2α-PAS-B-PT2385
HIF-2α-PAS-B-Belzutifan

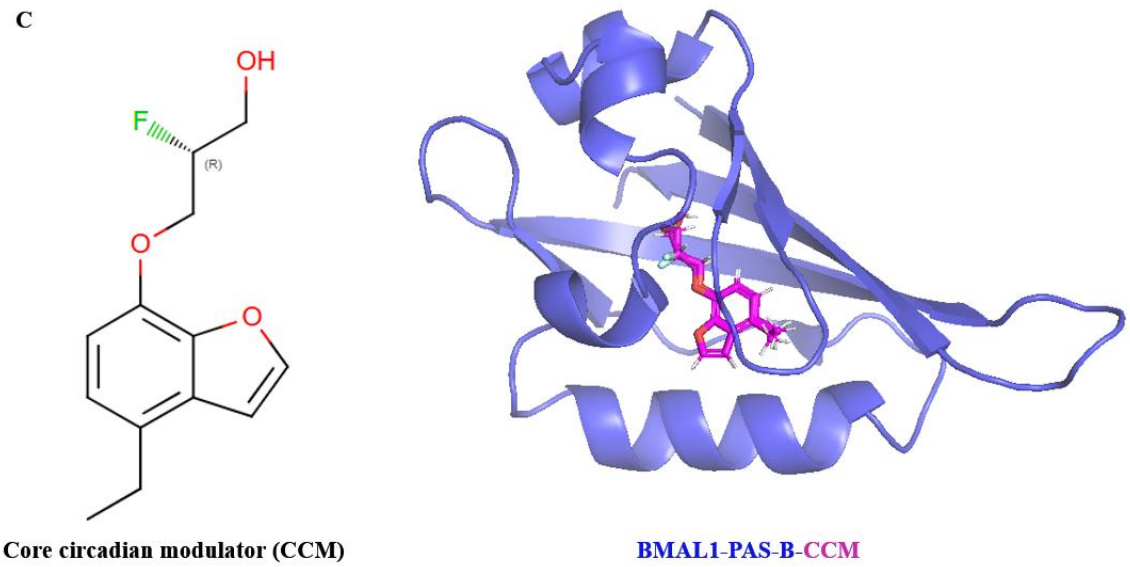


Figure 15. Pharmacological targeting of the PAS-B domain within the bHLH-PAS family. (A) FDA-approved tapinarof for the treatment of psoriasis binds to the PAS-B domain of AHR (PDB: 8XS6, 2.95 Å). (B) Chemical structures of compounds used in the development of belzutifan. Crystal structures of the PAS-B domain of HIF-2α bound to PT2385 (PDB: 5TBM, 1.85 Å) and belzutifan (PDB: 7W80, 2.75 Å). (C) Crystal structure of the BMAL1 PAS-B domain in complex with CCM (PDB: 8RW8, 2.16 Å).

Historically, the NR family was considered the primary group and the best example of human transcription factors that have been successfully targeted for drug discovery. However, these recent studies of the human bHLH-PAS protein family, supported by FDA approval of drugs targeting several of its members, have brought remarkable recognition to this family. These discoveries further underscore their druggability, and highlight their potential as compelling targets for intervention in complex disorders.

2.8 Functional roles of NPAS1 and NPAS3 in neurodevelopment and disease

NPAS1 and NPAS3 are brain-enriched transcription factors. The expression levels of NPAS1 and NPAS3 are related to neuronal development, differentiation, functions and growth, making these proteins crucial for the modulation of the CNS. Since the discovery of NPAS1 and NPAS3 as the causative loci of intellectual disability, their genetic deficiencies and abnormalities have been robustly associated with neurodevelopmental anomalies and linked to a range of neuropsychiatric diseases, including schizophrenia, autism spectrum disorders (ASD), bipolar disease, attention-deficit or hyperactivity disorder, and severe depression (Figure 14).[128–133]

Behavioural and neurochemical abnormalities have been further characterised in laboratory mice lacking NPAS1, NPAS3, or both genes. Several behavioural tests revealed that NPAS1/NPAS3-deficinet mice exhibited an impaired startle response, diminished social recognition, stereotypic darting behaviour, hyperactivity and increased locomotor activity.[4] General observation of mutant mice also revealed broader behavioural and neurological impairments. Mutant mice exhibited abnormal motor functions, characterised by gait changes, reduced muscle tone upon tail suspension, and general uncoordinated movement. After giving birth, mutant mothers exhibited poor nesting and nurturing instincts. Such behavioural and learning issues strongly suggest hippocampal dysfunction, including reduced performance on tasks dependent on hippocampal memory, and altered emotional tone. These behavioural features of the NPAS1/3 mutant mice demonstrate similarities to existing animal models of schizophrenia and are consistent with aberrations of the

glutamatergic and dopaminergic systems, which are features of human psychiatric conditions.[132]

NPAS1/3 mutant mice also display defects in brain structures and development, exhibiting direct parallels with neurological disorders. Specifically, the frontal cortex of adult mouse brain tissue from NPAS3-deficient and NPAS1:NPAS3 double-null animals demonstrated a profound reduction in reelin, an extracellular matrix serine protease secreted by pioneer neurons, which plays a role in neuronal layering in the cerebral cortex during brain development. The expression of reelin is also significantly reduced in patients with reduced cognitive ability and schizophrenia.[4] Additionally, clear defects were observed in the corpus callosum and cerebellum, indicating a potential correlation with altered coordination and motor phenotypes of NPAS3-deficient mice.[132] Probably the most striking structural change was observed within the dentate gyrus, a fundamental part of the hippocampal formation, as several studies demonstrated that NPAS3 mediates adult hippocampal neurogenesis.

In vertebrates, the NPAS function in embryonic neural progenitors likely reflects an evolutionarily conserved role shared with their *Trh* homolog in *Drosophila. Trh* is an essential protein that modulates fibroblast growth factor (*FGF*) signalling through transcriptional regulation of the FGF receptor, specifically during tracheal and neural development.[38] With analogy to *Trh*, NPAS3 regulates the FGF signalling pathway in the dentate gyrus of adult mice by controlling the expression of the FGF receptor subtype 1 (*FGFR1*), which in turn controls the proliferation of hippocampal granule neurons.[36] The deletion of NPAS3 in mice produced a profound deficit of *FGFR1* expression and an 84% reduction of neural precursor cell proliferation in the hippocampal dentate gyrus. Thus, the substantial deterioration of FGF-mediated hippocampal neurogenesis resulted in

malformation, electrophysiological dysfunction, and reduced size of the hippocampal dentate gyrus of NPAS3 mutant mice.[36]

Despite their high degree of similarity, NPAS1 and NPAS3 reciprocally regulate postnatal hippocampal neurogenesis. While the loss of NPAS3 inhibits postnatal hippocampal neurogenesis, it was demonstrated that the deletion of the NPAS1 gene promotes and increases neurogenesis in the dentate gyrus.[2,35,35] This opposite regulation could be due to the presence of several RPD domains within NPAS1, as compared to the confirmed TAD domain in the C-terminus of NPAS3 protein. Aberrations in hippocampal structure and functions, including reduced hippocampal volume, and altered hippocampal neurogenesis and neuronal activity, have been widely observed in the pathophysiology of schizophrenia, highlighting the importance of NPAS1 and NPAS3 in disease pathology and supporting their potential role in the aetiology of schizophrenia.[134,135]

Intriguingly, NPAS1 and NPAS3 also have opposite roles in regulating interneuron numbers. Several studies have demonstrated that NPAS1/3 proteins are novel regulators of cortex size and play an important role in excitatory-inhibitory balance in the brain. The neuronal expressions of NPAS1 and NPAS3 were confirmed to occur during intraneuronal development, specifically in the telencephalic progenitor domains of the cortex, and the caudal and medial ganglionic eminences (CGE and MGE), and later in both immature and mature cortical interneurons. It was previously shown that double knockout NPAS3 mice display reduced proliferation and mitogen-activated protein kinase (MAPK) signalling activity in progenitors of the ganglionic eminences.[35] Deletion of NPAS3 was also found to decrease the number of SST^+ (somatostatin-positive) and VIP^+ (vasoactive intestinal peptide-

positive) interneurons, which could lead to a disruption of the excitatory-inhibitory balance, which in turn is linked to schizophrenia and intellectual disability.[35]

In contrast, NPAS1 negatively regulated proliferation and MAPK signalling in CGE and MGE progenitors. NPAS1 represses the generation of cortical interneurons not by regulating *FGF*, but by suppressing the activity of the aristaless-related homeobox gene (*ARX*), a crucial TF for the development of cortical GABAergic interneurons.[35] Together with ARNT, NPAS1 directly interacts and inhibits *ARX* expression through its subpallial enhancer element, which mediates CGE progenitor cell proliferation. As a result, NPAS1-/- mutant mice exhibit increases in both SST^+ and VIP^+ cortical interneuron numbers, as well as enlarged cortical volume. As with NPAS1 mutant mice, enlargement of brain size is frequently observed in children with ASD. These observations suggest that both NPAS1 and NPAS3 transcription factors play a crucial role during neuronal differentiation, regulating the excitation-inhibition balance in the brain, and hence controlling regulatory pathways relevant to the aetiology of psychosis and schizophrenia.

Despite the importance of NPAS1/3 in neuronal function, very little is known about their target genes. It has been reported that NPAS1 binds to the enhancer region of *EPO* and negatively regulates EPO expression, a factor that stimulates the production of neuronal progenitors in the CNS.[42] One high-throughput microarray study in human HEK293 cells found that hundreds of genes are regulated differently by overexpression of NPAS3; however, which of them are direct targets has yet to be uncovered as correlative *in vivo* testing was not conducted.[136] It was demonstrated that co-expression of NPAS3 and ARNT up-regulated neuronal survival factor *VGF* (nerve growth factor inducible) and down-regulated *TXNIP* (thioredoxin interacting protein) through binding of their proximal promoters.

Therefore, both proteins might participate in dendritic arborisation, neuron formation, and oxidative stress response.[45] By inhibiting the pro-apoptotic function of *TXNIP*, NPAS3 may promote cellular survival, similar to its pro-neurogenic role in the hippocampus.[36]

The first *in vivo* chromatin immunoprecipitation next generation sequencing (ChiP-seq) transcriptome-scale study discovered multiple targets of NPAS1/3 proteins that also demonstrated significant functional enrichment for neuronal and neurogenesis-related processes. It was shown that in addition to *FGF* and *SHH* signalling, NPAS3 directly controls an ensemble of genes that regulate neuropsychiatric function, specifically: fragile X mental retardation protein translational regulator 1 (*FMER1*), ubiquitin-protein ligase E3A (*UBE3A*), and neurogenesis regulator neurogenic locus notch homolog protein 1 (*Notch*).[2] In the absence of NPAS3, both *FMER1* and *UBE3A* targets have decreased expression, which resulted in dysregulation of both downstream pathways. Dysfunction of *FMER1* is known to cause syndromic forms of intellectual disability and fragile X syndrome. The dysregulation of *UBE3A* is associated with Angelman syndrome.

Additionally, the loss of NPAS3 resulted in an approximately threefold greater transcriptional perturbation as compared to the loss of NPAS1. Out of 443 protein-coding transcripts, NPAS1 downregulated 69% of genes, showing mostly inhibitory effects. The study also revealed that NPAS1 and NPAS3 reciprocally regulate the expression of neurobeachin (*NBEA*), cannabinoid receptor-1, neuronal pentraxin receptor, MKL/myocardin-like-2, and microtubule-associated protein-2 in the mouse hippocampus, supporting the evidence of their different roles in postnatal hippocampal neurogenesis.[2] Strikingly, nine NPAS1 targets and ten NPAS3 targets have been implicated as

schizophrenia-related genes, suggesting a significant role for NPAS1/3 in the molecular mechanisms underlying schizophrenia and intellectual disability.

Aberrations in the regulation of genes, and excitatory-inhibitory balance within the brain, have both been associated with intellectual disability and neurological abnormalities. Interestingly, protein aggregation has also been reported to be significant in mental illness. Proteinopathy, or protein misfolding disease, is a well-established mechanism in classical neurodegeneration but has only recently been investigated in major mental illness.[137] In a small family where three individuals have schizophrenia, and one person has major depressive disorder, an ultrarare V304I point mutation in the NPAS3 protein reduces the solubility of endogenous protein and induces the aggregation propensity of NPAS3.[138] Residue 304 resides in the PAS-A domain of NPAS3. A recent postmortem study found insoluble full-length NPAS3 in 70% of 40 insular cortex samples across diagnostic groups, suggesting aggregation of NPAS3 could be a widespread phenomenon than the sole reported V304I family.[48] The investigation of the possible mechanism in SH-SY5Y human neuroblastoma cells revealed that oxidative stress could play a larger role in NPAS3 aggregation than the V304I mutation, indicating a plausible contributing combination of genetic and environmental factors.

Beyond V304I, multiple coding variants of NPAS3 have also been previously associated with psychiatric illnesses, including A552P, and one rare population variant, G697S.[46,139] The effect of these three human NPAS3 variants was assessed on *VGF* regulatory function, but no significant deficit was observed in their ability to activate expression driven by the *VGF* promoter. All mutants were also found to be normally expressed and localised to the nucleus, suggesting that point mutations did not affect interaction with the ARNT partner.[45]

From a clinical exome sequencing database, other low-frequency variants of NPAS3, which cluster around the PAS-A domain, have been identified and associated with developmental delays, which can be indicative of later intellectual disability. These include G201R, G229R, and non-functional 214 frameshift (fs) proteins.[54] The frameshift introduces 12 new amino acids and a premature termination codon, resulting in a significant part of NPAS3 being lost, including the C-terminal TAD.[54]

Another NPAS3 433fs variant leads to a frame shift near the end of PAS-B and disrupts the protein at an analogous position to the previously identified dimerisation hotspot framework with SIM1.[140] Both G201R and G229R had similar reporter activity to the wildtype NPAS3-ARNT/2 proteins.[54] Curiously, both 214fs and 433fs variants exhibited a near-complete loss in reporter gene activity with ARNT2. It was further shown that 214fs NPAS3 was expressed at a much lower level than the wildtype protein.

Furthermore, both 214fs and 433fs were unable to co-immunoprecipitate with ARNT2.[54] The inability to bind and heterodimerise with the ARNT2 partner protein could explain the loss of function mechanism and decreased protein expression. These observations might indicate that mutations or frameshifts in these regions make essential inter- and intra-molecular contacts, and that the absence of such contacts may destabilise the PAS-B domain and hinder the ability to form a dimer with a partner protein.

2.9 ARNT2 is a transcriptional integrator in neurodevelopment, metabolism, and human disease

As a paralog of ARNT, ARNT2 functions as an essential partner for sensor proteins within the bHLH-PAS family. ARNT2 forms various heterodimeric transcriptional complexes with HIF-1α, SIM1, and NPAS4, mediating cellular responses to hypoxia, metabolic regulation, neuroendocrine differentiation, and neuronal plasticity. Therefore, ARNT2 emerges as a multifunctional TF involved in the regulation of gene expression in response to various environmental and developmental signals within the CNS and beyond. It has been identified as a pivotal player in neurodevelopment, neuronal differentiation, and synaptic plasticity, and hence relevant to a variety of neurological and psychiatric disorders.

The expression of ARNT2 is largely restricted to the CNS and other neural tissues, suggesting a more specialised set of physiological functions.[60,141] Several studies have illuminated the significant role of ARNT2 in neurodevelopment and neuronal survival, particularly in regions such as the hypothalamus and hippocampus. During human embryonic development, ARNT2 is ubiquitously expressed in the hypothalamus, one of the most evolutionarily conserved and fundamental regions of the brain.[55]

In the hypothalamus, ARNT2 forms heterodimers with SIM1 to control the development and differentiation of neuroendocrine neurons in PVN and SON nuclei.[58,142] Mouse models deficient in ARNT2 demonstrated severe defective formation of the hypothalamus-pituitary axis. While progenitor markers like BRN2 appear in PVN and SON around embryonic day 13.5, no neurons expressing neurosecretory hormones, such as arginine vasopressin,

oxytocin, corticotropin-releasing hormone, and somatostatin, were found in these regions, demonstrating the essential role of ARNT2 in late-stage neuronal differentiation.

Furthermore, homozygous ARNT2 gene knockout mice died within 24 hours after birth, most likely due to disrupted hypothalamic development and loss of pituitary hormone.[143] These observations display a striking similarity to the phenotypes observed in SIM1-null mice, which also demonstrated a lack of terminal differentiation of these neuroendocrine neurons, and hence low secretion of neurohormones, suggesting the importance of both ARNT2 and SIM1.[58]

The conserved role of ARNT2 in hypothalamic development was also confirmed in studies with zebrafish, where it was shown that ARNT2 and SIM1 play a significant role in the differentiation of a defined population of dopaminergic and neuroendocrine neurons. The study also showed that ARNT2 and SIM1 control aspects of axon guidance, because their depletion resulted in the displacement of hypothalamo-spinal longitudinal axons towards the midline.[144] The involvement of both genes in long-range axonal guidance establishes the connections between the hypothalamus and the spinal cord in zebrafish.

The expression and function of ARNT2 in neuroendocrine and cortical circuits also suggest its involvement in psychiatric and neurodevelopmental disorders. Beyond embryonic development, ARNT2 is enriched in the cortex and hippocampus, regions critical for cognition, emotion regulation, and stress response, but is relatively underrepresented in midbrain dopaminergic neurons.[145] This distribution has led to hypotheses that ARNT2 contributes to neurological disorders such as schizophrenia, ASD, and depression via regulation of stress-response genes and neurotrophic factors.

Further support comes from several studies that have proposed that ARNT2 has a role in predisposition to autism spectrum conditions. After heat shock, ARNT2's expression level was altered in neural aggregates in human induced pluripotent stem cell models, comparable to that of a telencephalon in the early developmental stages.[146] Human genetic studies identified a significant association between the ARNT2 intronic single-nucleotide polymorphism, rs17225178, and Asperger Syndrome, a subgroup of ASD with no language or general cognitive delay.[147] Bioinformatic analysis suggests that this variant alters TF binding sites and regions that regulate the chromatin state in neural cells.

Within the first few months following birth, the absence of ARNT2, or mutations in ARNT2, result in a distinct congenital syndrome in humans characterised by subsequently developed hypopituitarism, seizures, severe visual and renal anomalies, and microcephaly with cerebral hypoplasia affecting the frontal and temporal lobes. Such a severe clinical phenotype strongly correlates with the loss of ARNT2 mRNA and the absence of detectable levels of ARNT2 in patient fibroblasts, most likely affected by nonsense-mediated decay caused by a frameshift mutation (c.1373_1374dupTC).[55] Children with secondary post-natal microcephaly displayed features of hypothalamic insufficiency, including obesity, diabetes insipidus, and adrenocorticotropic and thyroid-stimulating hormone deficiency. This suggests a progressive neurodegenerative condition, potentially resulting from decreased dendritic connections, or loss of the number or activity of neurons, which in turn could be explained by the post-natal requirement for ARNT2 expression. These observations underscore the essential function of ARNT2 in postnatal brain growth and the development of the hypothalamo-pituitary and visual systems.

The controlled expression of ARNT2 in the brain, throughout development and adult life, suggests its critical role in the neuronal regulation underlying cell division processes, cell life, and cell death. ARNT2 is involved in neuroprotective properties in ischemic insults and oxidative damage. ARNT2 gene transcript levels significantly declined after 2 hours of recirculation post-ischemic injury, preceding neuronal death at 24 hours of recirculation.[148] Furthermore, ARNT2 demonstrated an anti-apoptotic role in PC12 cells, a cell line derived from a pheochromocytoma of the rat adrenal medulla. The downregulation of ARNT2 by RNA interference inhibited PC12 cell proliferation and triggered apoptosis. Oxidative stress-induced PC12 cell death was also preceded by rapid and strong downregulation of ARNT2.[148] The effect of stressors involved in inflammatory and neurodegenerative processes on ARNT2 expression has also been illuminated in primary enriched cultures of cortical interneurons. The level of ARNT2 was rapidly upregulated upon exposure to oxidative stress or apoptotic stimuli.[149] Thus, ARNT2 is part of an early protective programme that is activated in response to neuronal stress, which regulates neuronal survival by modulating transcriptional programmes that protect neurons from apoptosis and other forms of cellular damage.

A neuroprotective role for ARNT2 is supported by other findings showing that its heterodimerisation with NPAS4 drives expression of the brain-derived neurotrophic factor (*BDNF*) gene, which is related to several human neurodevelopmental conditions. *BDNF* is a critical mediator of various physiological processes, including activity-dependent signalling, axonal growth, and neuronal survival, with crucial effects on memory, learning, and neurogenesis. The expression level of BDNF was reduced twofold in cultures where NPAS4 was knocked down by RNA interference, and the overexpression of dominant-negative ARNT2 and NPAS4 blocked *BDNF* induction.[150,151] Similarly, *in vivo* analyses demonstrate

that both ARNT2 and NPAS4 levels were upregulated prior to neuroinflammation disease onset, correlating with increased levels of BDNF. Interestingly, recent studies have shown that NPAS3 can also recognise and bind to the E-box sequence within the *BDNF* promoter.[152]

In the experimental autoimmune encephalomyelitis mouse models of multiple sclerosis, ARNT2 expression is significantly decreased in $NeuN^+$ neurons, $GFAP^+$ astrocytes, and $Olig2^+$ oligodendrocyte precursor cells at peak disease severity, correlating with inflammation, axonal loss, and impaired glial maturation. The downregulation of ARNT2 coincided with decreased expression of NPAS4 and is linked to a decrease in BDNF.[149] Together, these observations suggest that ARNT2 may form part of a neuroprotective transcriptional network that buffers neurons against inflammatory and oxidative insults.

ARNT2's protection of neurons from cell death also operates through interactions with the HIF signalling pathway. It is confirmed that ARNT2 can form HIF complexes *in vitro* and *in vivo*, specifically in primary neuronal cultures as well as in nerve growth factor (NGF)-induced PC12 cells.[148,153] Through heterodimerisation with HIF-1α, ARNT2 directly regulates hypoxia-induced gene expression and restores hypoxia-driven gene transcription in ARNT-null embryonic stem cells and hepatocytes.[153] ARNT2 also mediates the expression of angiogenic agent vascular endothelial growth factor (*VEGF*) in primitive neuroepithelial cells, playing an important role in the proper vascularisation and growth of ARNT-/- mutant tumours, indicating ARNT2's role in hypoxic gene induction.[153] These observations suggest an integral role of ARNT2 in hypoxic transcriptional responses by forming HIF complexes in the CNS. The neuron-specific hypoxia response of ARNT2 could have an important role in neural protection.

ARNT2 continues to regulate physiological functions in the adult brain, particularly within circuits controlling energy homeostasis. It was shown that a viable R74C mutation in the ARNT2 bHLH domain causes a significant decrease in transcriptional activity when co-expressed with SIM1, potentially due to the disrupted nuclear localisation of the ARNT2-SIM1 complex.[154] Homozygous mutant mice exhibit profound hyperphagia, increased linear growth, obesity, glucose intolerance, diabetes, and hepatic steatosis. These findings reveal ARNT2 as an important regulator of metabolic control, probably through its actions in the hypothalamus, and indicate that subtle alterations in its function may contribute to obesity and related disorders in humans.

Collectively, these findings define ARNT2 as a neuronal TF with a significant impact on normal development of the CNS, specifically affecting the hypothalamo-pituitary axis and visual pathways, as well as the maintenance of normal post-natal brain and CNS in humans. Orchestrating early neuronal differentiation and modulating adult brain plasticity, its actions extend to metabolic regulation, hypoxic adaptation, and disease defence. The phenotypes observed in both murine models and human patients underscore the importance of ARNT2 in maintaining homeostasis across multiple physiological systems. Furthermore, its dysregulation in disorders as diverse as neurodevelopmental syndromes, autoimmune diseases, obesity, and cancer, points to ARNT2 as a potential biomarker and therapeutic target in human health and disease.

2.10 KG-548 as a ligand for ARNT2: discovery, mechanism, and assays

Playing a pivotal role in neurological conditions, several studies aimed to identify targets within the NPAS1, NPAS3, and ARNT/2 regulatory pathways, with the ultimate goal of restoring normal neuronal development and function in disease contexts. Among these was the discovery of an aminopropyl carbazole designated P7C3, pro-neurogenic and neuroprotective compound, which demonstrated promising pharmacological properties. *In vivo* experiments indicated that P7C3 promotes neurogenesis by protecting newly formed neurons from undergoing apoptosis.[155] Mice lacking NPAS3 are devoid of hippocampal neurogenesis, along with structural abnormalities and functional impairments in the dentate gyrus.[36] However, extended treatment with P7C3 in these NPAS3-/- mice corrected these deficits by restoring normal levels of apoptosis among newborn hippocampal neurons.[155] Similarly, long-term administration of P7C3 to aged rats boosted neurogenesis in the dentate gyrus, reduced neuronal death, and helped maintain cognitive function during advanced ageing.[155]

Importantly, P7C3 neither directly binds to NPAS3 nor restores its transcriptional functions. Instead, P7C3 exerts its neuroprotective effects downstream of NPAS3 by binding and enhancing the activity of nicotinamide phosphoribosyl transferase (NAMPT), the rate-limiting enzyme involved in the conversion of nicotinamide into nicotinamide adenine dinucleotide within the NAD^+ salvage pathway.[156] This increase in NAMPT activity leads to elevated intracellular NAD^+ levels, which are crucial for cellular metabolism and survival. By boosting NAD^+ availability, P7C3 helps prevent apoptosis in vulnerable newborn neurons

within the dentate gyrus. Through this mechanism, P7C3 supports neurogenesis and maintains dentate gyrus function even when NPAS3 is absent.[156]

Through protein thermal shift (TSA/PTS/DSF) screening, a triazol-3-one bearing a 5-nitro-2-thiazolyl-thio moiety (BI-78D3) was identified as a hit that directly blocks the NPAS3-ARNT assembly (Figure 16A).[152] Time-resolved fluorescence (TR-FRET/HTRF) assays showed a potent inhibitory effect of BI-78D3 with an EC_{50} value of 4.48 ± 0.76 μM. Although BI-78D3 is a known inhibitor of Jun N-terminal kinase (IC_{50} ~280 nM) and extracellular regulated protein kinase, its inhibitory effect in this system did not require kinase activity and instead mapped to ARNT itself.[157,158] Biophysical characterisation via nuclear magnetic resonance (NMR) and mass-spectrometric (MS) analyses indicated adduct formation on ARNT and supported a covalent binding mode of BI-78D3. For NPAS3-ARNT, the equilibrium dissociation constant (K_D) for BI-78D3 was measured at 287 nM by microscale thermophoresis (MST). These findings demonstrate that BI-78D3 can effectively and covalently bind to the ARNT subunit and block NPAS3 interaction.[152]

Medicinal-chemistry focused optimisation of BI-78D3 nitrothiazole sulfydryl warhead yielded a more potent derivative, Compound 6, subsequently commercialised as BI-87G3/NPAS3-IN-1, which achieved a biochemical EC_{50} of 282 ± 61 nM for disrupting NPAS3-ARNT heterodimerisation via TR-FRET (Figure 16A).[152] Tandem mass spectrometry (MS-MS) identified that the Cys336 residue of ARNT is covalently modified by Compound 6. Further evaluation by TSA, NMR, and MST established covalent targeting of ARNT Cys336 as the operative mechanism, and quantified inhibitory potency. Additionally, demonstrating its cellular uptake and direct binding to the ARNT subunit, the immunoblotting results of a cellular thermal shift assay (CETSA) in HEK293T cells showed

that Compound 6 significantly increased the thermostability of the ARNT subunit at 48.25 and 50 °C. Compound 6 also disrupted the assembly of NPAS3 and ARNT subunits via co-immunoprecipitation (Co-IP). Transcriptional cellular assays also demonstrated a dose-dependent downregulation of NPAS3-dependent BDNF expression by disrupting the interface of NPAS3-ARNT complexes.[152] Together with prior crystallographic work that defined druggable pockets within NPAS3-ARNT, these results inaugurate a chemical strategy for modulating neuronal bHLH-PAS signalling via its partner ARNT.

Another small molecule that targets ARNT is KG-548 (5-[3,5-bis(trifluoromethyl) phenyl]-2H-tetrazole), which was discovered in a solution NMR screen against the PAS-B domain (Figure 16A).[159] A high-resolution ARNT PAS-B and KG-548 co-crystal structure (PDB: 8G4A; 1.97 Å) demonstrated direct binding of KG-548 to the external surface of ARNT PAS-B and illustrated how the tetrazole/aryl scaffold nests across the β-sheet surface that mediates coactivator engagement and dimerisation with HIF-2α PAS-B (Figure 16).[160,161] Both NMR readouts and mutational analysis using the Y456T variant showed that altering the surface modified the binding signature of KG-548, consistent with surface binding.[161]

When KG-548 binds the surface of ARNT PAS-B, it overlaps the recruitment site for the transforming acidic coiled-coil containing protein 3 transcriptional coactivator (TACC3), thereby disrupting ARNT-TACC3 association. Functionally, titrating KG-548 into a pulldown assay of His-ARNT PAS-B and GST-TACC3-CT fragments yielded a dose-dependent loss of complex formation, consistent with competitive engagement of the TACC3-recruitment surface on ARNT.[159] Co-IP from HEK293T lysates demonstrated that KG-548 diminished endogenous ARNT associated with immunoprecipitated TACC3,

supporting cellular target engagement at the ARNT-TACC3 node.[159] These experiments collectively demonstrate that KG-548 binds ARNT PAS-B, and functionally blocks TACC3 recruitment *in vitro* and in cell lysate. Subsequent studies by fluorine NMR spectroscopy using the fluorine atom in KG-548 showed the affinity for ARNT PAS-B was 350 μM.[161]

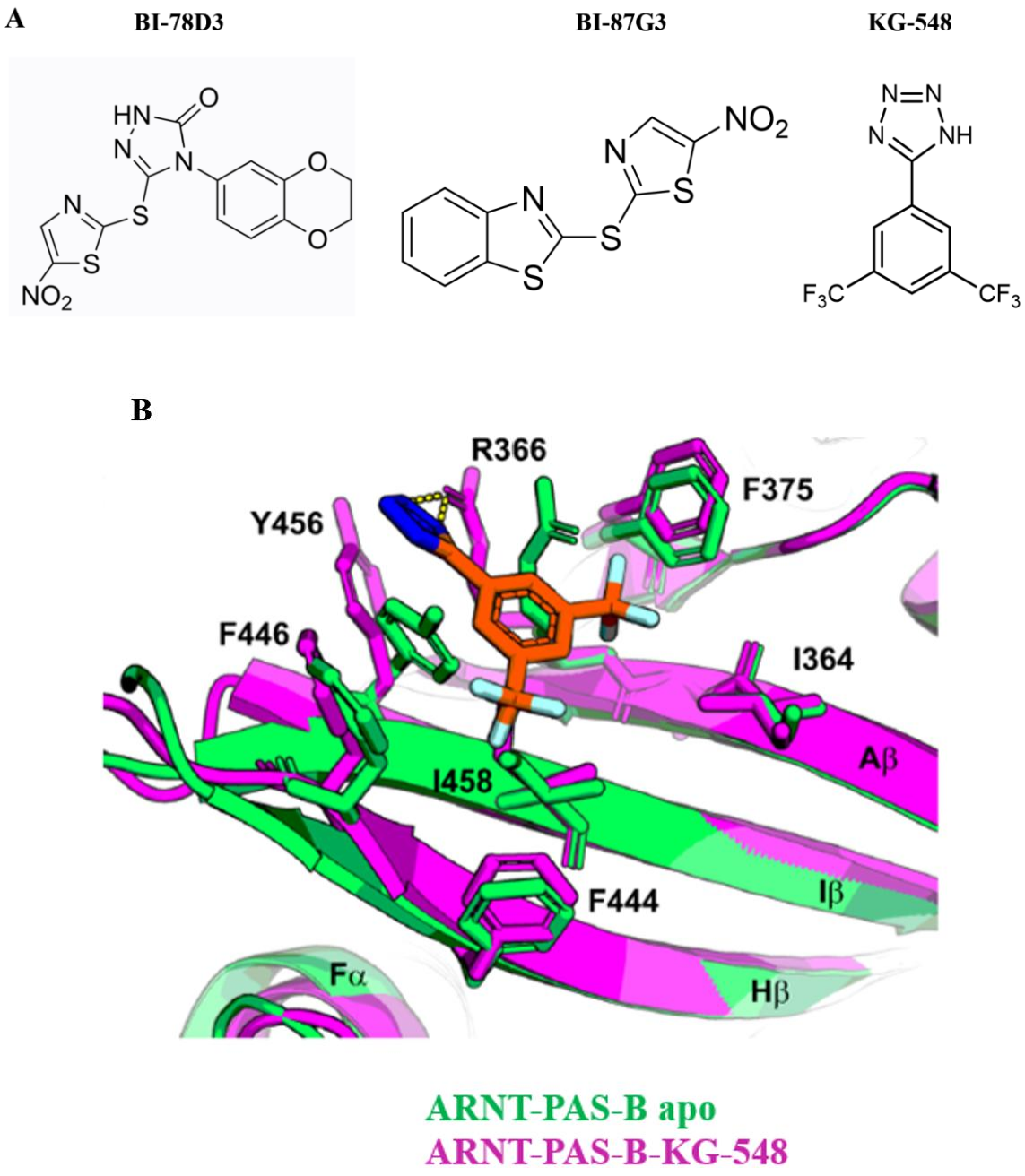


Figure 16. Known ARNT2 binders. (A) Chemical structures of ARNT2 binders. (B) Structural comparison of ARNT PAS-B apo and KG-548 bound states. Superimposition of the crystal structure of the ARNT PAS-B apo (PDB: 4EQ1, 1.60 Å, in green) and KG-548-bound ARNT-PAS-B (PDB: 8G4A, 1.97 Å, in magenta). Residues within 5Å of the ligand are shown in sticks. To prevent clashes, residue Tyr456 moves away from KG-548 in the ligand-bound state. Residue Arg366 also reorients and forms polar contacts with the tetrazole group of KG-548, shown with yellow dashed lines. Figure adapted from Xu et al., 2024.[161]

Strikingly, KG-548 also disrupts the homologous interaction between the paralog ARNT2 PAS-B and TACC3 in a pulldown experiment, indicating that KG-548 engages ARNT2 at a conserved TACC3-binding β-sheet surface, mirroring the ARNT results.[159] In contrast, no substantial KHS101-inhibiting effect on TAACC3 with ARNT or ARNT2 was observed in the same pulldown assay, underscoring that KG-548 uniquely engages the PAS-B surface on both paralogs. Notably, the required Glu398 and Lys417 residues for TACC3 binding on ARNT PAS-B are conserved with Glu372 and Lys391 in ARNT2, underlining cross-reactivity by sequence-structure homology.[159] The combined evidence of the dose-dependent disruption of ARNT2/TACC3 pulldown by KG-548 and conserved epitope mapping supports the conclusion that KG-548 binds ARNT2 PAS-B at the same β-sheet site used on ARNT.

Collectively, by occupying the β-sheet platform on ARNT/ARNT2 that recruits TACC3, KG-548 competes directly with coactivator binding and can therefore modulate the assembly of different HIF or AHR complexes. Although KG-548 binds to ARNT PAS-B very weakly, it is a rigorously validated chemical probe of the ARNT/ARNT2 PAS-B β-sheet surface. Its discovery by NMR screening, mechanistic deconvolution by NMR, mutagenesis,

crystallography, and extension to ARNT2 via conserved-site pulldown assays, establish a clear and direct mode of action. Thus, KG-548 is a useful starting point for designing more potent and selective coactivator-interface inhibitors within the bHLH-PAS family.

Chapter 3: DPhil proposal and project aims

This DPhil project will explore the drug-binding potential of NPAS1 and NPAS3, together with their partner ARNT2. Both NPAS1 and NPAS3 are closely related members of the bHLH-PAS family, sharing a high degree of sequence identity, neuronal expression patterns, structural homology, functional roles, and implications in neurodevelopmental and neuropsychiatric conditions. While ARNT serves as the canonical partner for several bHLH-PAS proteins in many tissues, ARNT2 is the predominant dimerisation partner in neurons and, like NPAS proteins, critically participates in various neurological processes.

To date, the involvement of NPAS1, NPAS3, and their partner ARNT2 in neurodevelopmental disorders is poorly understood. Despite evidence that their loss significantly disrupts neurodevelopment, NPAS1/3 and ARNT2 proteins have received limited attention as potential therapeutic targets. Given their genetic linkage to diseases such as schizophrenia, ASD, and bipolar disorder, their modulation might aid in early diagnosis or prognosis, as well as potential treatment.

The biological roles of NPAS1 and NPAS3 suggest a bidirectional therapeutic strategy in which the activation of NPAS3-ARNT2 restores pro-neurogenic and pro-plasticity programmes, whereas the inhibition of NPAS1-ARNT2 relieves repressive tone that restrains specific neuronal gene networks. Direct pharmacological control of NPAS1/NPAS3-ARNT2 heterodimers could therefore offer a mechanism-based route to normalise maladaptive neuronal transcription in disorders marked by impaired neurogenesis, cognitive dysfunction, or altered excitation-inhibition balance.

The bHLH-PAS family presents unique opportunities for small-molecule intervention. Both PAS domains are compact α/β folds that create deep, hydrophobic cavities and allosteric clefts at intra- and inter-domain interfaces. Although several modulators of certain bHLH-PAS family proteins have been successfully identified, the direct NPAS1 or NPAS3-targeted compound has yet to be discovered. This project therefore seeks the discovery of small molecules for functional modulation of NPAS1, NPAS3, and their partner ARNT2.

Similar to other bHLH proteins, recent crystal structures of NPAS1 and NPAS3 bound to ARNT family partners reveal four discrete pockets with internal ligandable cavities across the PAS domains of each complex subunit. These structural data provided a rationale for chemical intervention at the NPAS1/NPAS3-ARNT interface or within allosteric pockets. In contrast to other bHLH-PAS members, NPAS1/3 and ARNT2 have much smaller pockets within the key sensing and ligand-binding PAS-B domain. This suggests that the discovery of a compound capable of modulating NPAS1/3-ARNT2 function by directly binding to this pocket may be challenging. Carefully selected small-molecule chemical libraries might overcome this limitation and help to identify potential direct binders, which in turn may stabilise or destabilise specific protein-protein interfaces. Additionally, certain compounds may induce outward redisposition of pocket residues, effectively expanding and reshaping the pocket upon binding, as previously demonstrated in studies of HIFs and BMAL1.

The overarching objective of this proposal is, therefore, to deliver first-in-class small molecules that directly bind to NPAS1, NPAS3, and ARNT2 to modulate NPAS1/NPAS3-ARNT2 heterodimers with defined directionality and mechanistic clarity.

Several specific aims were proposed:

Aim 1: To identify small molecules that can directly bind NPAS1, NPAS3, and ARNT2 proteins.

To discover compounds that directly interact with target proteins, single PAS-B domains or heterodimers comprising the functional bHLH-PAS-A-PAS-B triple domains will be expressed in *Escherichia coli* and purified. The discovery phase will employ biochemical high-throughput screening (HTS) using a protein TSA to assess whether binding of the compound induces changes in the overall thermostability of the target protein. Diverse small molecules and fragment chemical libraries designated for screening might contain non-selective or Pan-assay interference compounds (PAINS). Therefore, the primary screening hits will be further filtered, and promiscuous and Pan-active binders will be eliminated from further consideration.

Aim 2: To obtain apo protein and protein-compound bound crystal structures to verify binding events.

A confirmatory X-ray crystallography will establish direct physical binding and delineate the binding sites of selected compounds. The compound-bound crystal structure depicts surrounding residues, providing detailed insights into the binding interactions and conformational changes induced upon ligand engagement, which will guide rational optimisation through medicinal chemistry and further functional studies.

Aim 3: To conduct a battery of comprehensive biochemical and biophysical assays to validate and prioritise the most promising hits.

A set of secondary studies will include the specificity of action for NPAS1/3 and ARNT2 versus other bHLH-PAS proteins. Specificity will be assessed using a TSA-based counter-

screen against other members of the bHLH-PAS family. To quantify compound binding affinities, orthogonal biophysical approaches, including MST and surface plasmon resonance (SPR), will be employed to determine binding kinetics and calculate K_D values. Where feasible, covalent binding capability will be assessed by MS to analyse the molecular weight change of the compound-bound and un-bound protein. Additionally, a dose-responsive ability of compounds to stabilise or destabilise NPAS1/3-ARNT2 heterodimers will be assessed using a TR-FRET assay.

Aim 4: To implement hit-to-lead optimisation.

In parallel with the identification of new chemotype, hit-to-lead optimisation will be conducted for the known ligand, such as KG-548, and recently identified butein, using a combination of SAR by catalogue and synthetic medicinal chemistry efforts. Follow-up biochemical, structural, and biophysical studies, outlined above, will be performed with the derivative compounds to accelerate the path towards preclinical development.

Together, these studies will provide a validated small molecule capable of modulating NPAS1 and NPAS3 proteins and their interactions with partner ARNT2. The proposed work represents a critical prerequisite for profiling the functional activities of these bHLH-PAS proteins. Identification of drug-like modulators may provide new mechanistic insights into the role of NPAS1 and NPAS3 in neurological disease pathogenesis, thereby advancing opportunities for the development of novel therapeutic strategies.

PART II: Results and Discussion

Chapter 4: Protein production and quality control

4.1 Construct selection and purification of the PAS-B domain of ARNT2

The bHLH-PAS transcription factor family is distinguished by having tandemly positioned PAS domains, referred to as PAS-A and PAS-B. While the PAS-A domains across the bHLH-PAS transcription factor family are structurally rigid, they have relatively small internal cavities compared to the PAS-B domains in this family.[1,162] These PAS-A domains are highly involved in mediating protein-protein interactions, including dimerisation and interaction with transcriptional coactivators like coiled-coil coactivator (CoCoA) proteins, rather than small-molecule binding.[91,163] In contrast, PAS-B domains have larger and more chemically diverse interior cavities, allowing them to frequently participate in small-molecule or cofactor binding, as well as environmental sensing.[1,164] It is more likely, therefore, that a selective ligand will be found specifically for the PAS-B domain than the PAS-A domain. Previously, it has been shown that synthetic compounds could bidirectionally modulate the transcriptional activities of AHR and HIF-2α through their PAS-B domains.[109,117,165] Given this success of targeting the PAS-B domain, the focus of this research was on the bacterially expressed recombinant PAS-B domains of NPAS1, NPAS3, and ARNT2.

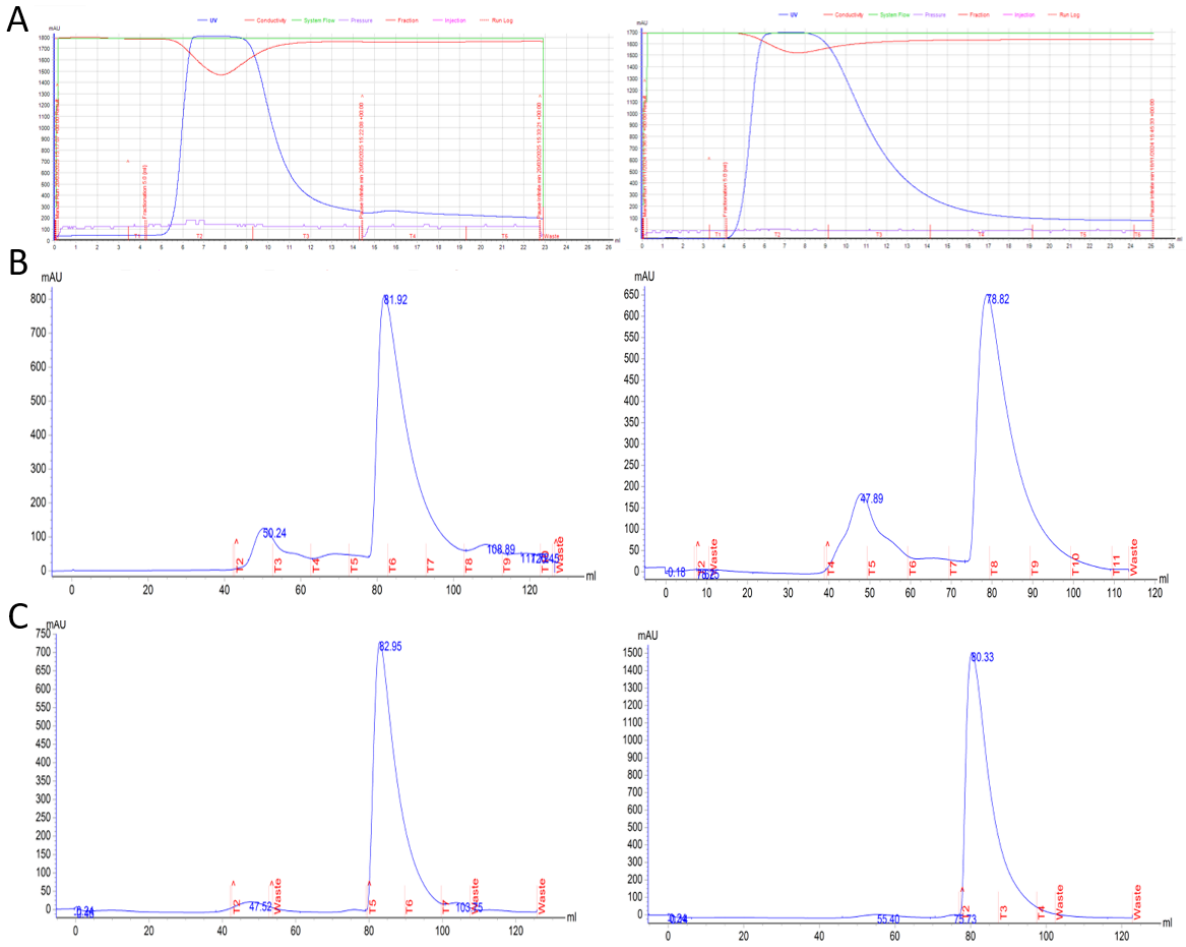
The recombinant human ARNT2-PAS-B domain (residues 330-444) has previously been purified, but no crystal structure was reported.[159] Mouse ARNT2-PAS-B (residues 332-438),

on the other hand, was successfully crystallised previously in our laboratory by Dr. Hua Pu (data not published). Therefore, it was decided to overexpress both human and mouse constructs for the ARNT2-PAS-B. The human protein was studied with the goal of identifying suitable crystallisation conditions and to characterise ligand specificity, while the mouse protein was purified as a backup in case crystallisation of the human protein proved unsuccessful.

The boundaries for human ARNT2-PAS-B were selected based on the mouse protein crystal structure. Human (residues 332-442) and mouse (residues 332-438) in-house ARNT2-PAS-B constructs were expressed in BL21-CodonPlus (DE3)-RIL cells to enhance translation efficiency and protein yield. This bacterial strain is commonly used for heterologous protein expressions due to its supply of tRNAs for rare codons (AGG, AGA, AUA, CUA), and its chromosomally encoded T7 RNA polymerase gene, which enables high-level transcription from T7 promoter-based plasmids.

Human and mouse expression vectors encode an N-terminal hexa-histidine 6xHis-tag followed by a TEV protease cleavage site (ENLYFQ/G). After sonication, soluble lysate fractions were loaded onto a HisTrap FF column prepacked with Ni Sepharose 6 Fast Flow resin, a 6% cross-linked agarose with an immobilised chelating group pre-charged with nickel ions, for preparative purification of histidine-tagged proteins via immobilised metal ion affinity chromatography (IMAC). After washing, 6xHis-tag proteins were eluted with imidazole-containing buffer (Figure 17A). Protein eluates were further purified by Superdex 75-pg size exclusion chromatography. Based on the S75 column calibration curve (MW/Volume), both proteins should be eluted after 75 mL column volume (Appendix 1). The size exclusion UV trace showed a single peak at 81 mL and 78 mL for the human and

mouse ARNT2-PAS-B, respectively, indicative of the presence of high-purity monomeric protein (Figure 17B).



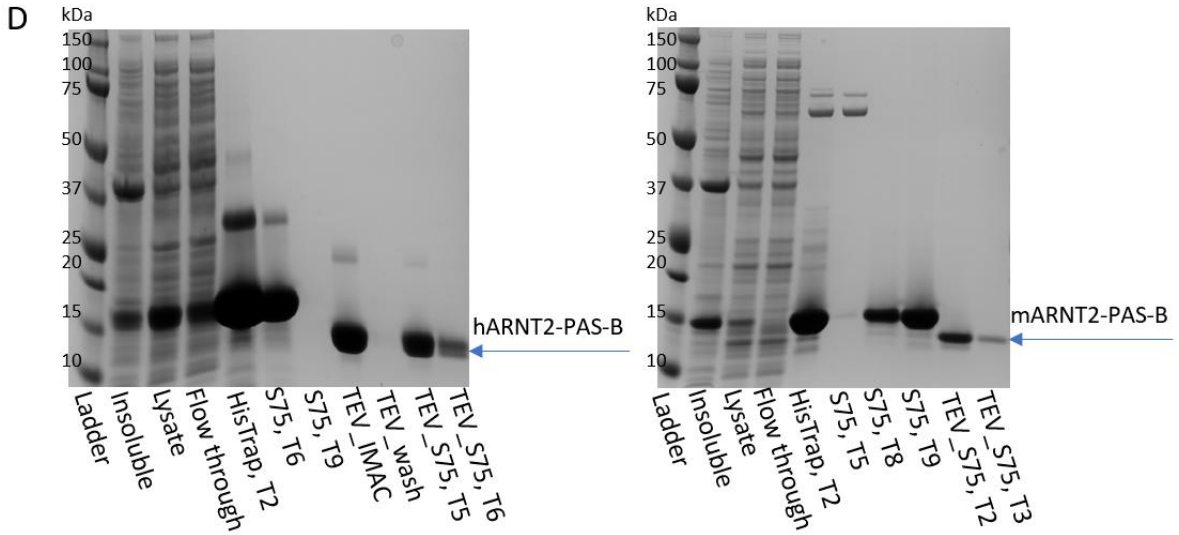


Figure 17. Preparation of the human and mouse ARNT2-PAS-B domains.(A) HisTrap FF affinity column UV traces for the lysate supernatant of human and mouse ARNT2-PAS-B constructs. (B) HiLoad 16/600 Superdex 75-pg size exclusion column UV traces. (C) HiLoad 16/600 Superdex 75-pg size exclusion column UV traces for TEV-cleaved human and mouse ARNT2-PAS-B domains. (D) Corresponding SDS-PAGE results for all steps in the purification of the constructs. The molecular weights of the histidine-tagged human and mouse ARNT2-PAS-B are 16.2 kDa and 15.1 kDa, respectively.

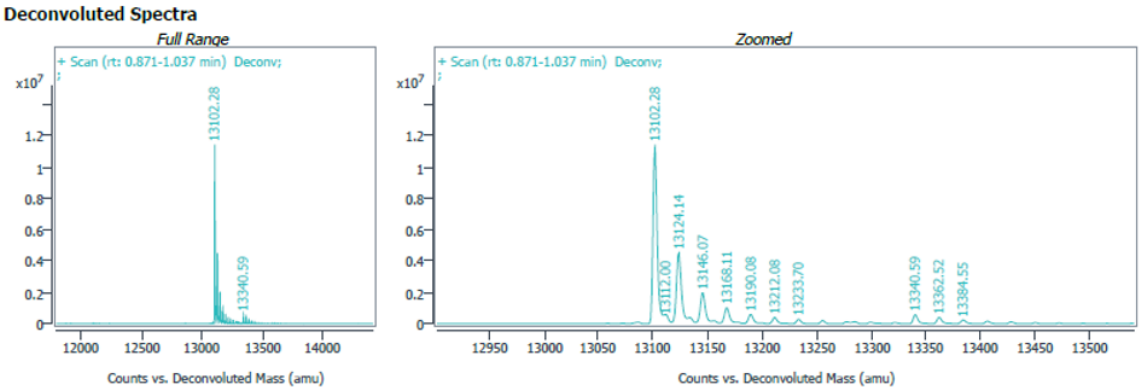
To remove the 6xHis-tag, both proteins were digested with the TEV protease. Proteins were re-purified with IMAC. The subsequent size exclusion chromatography showed a single peak at 82 mL and 80 mL for human and mouse PAS-B domains, respectively, showing the tag was efficiently removed for both proteins (Figure 17C). SDS-PAGE confirmed the purity of both proteins by showing clear bands at 13.1 kDa and 12.6 kDa, corresponding to human and mouse ARNT2-PAS-B, respectively (Figure 17D).

The intact mass and the purity of proteins were also assessed by electrospray ionisation mass spectrometry (ESI-MS). The ESI-MS results revealed that one major species was present in

each sample (Figure 18). The expected exact masses of 13,102 Da and 12,659 Da were observed for TEV-cleaved human and mouse ARNT2-PAS-B domains, respectively.

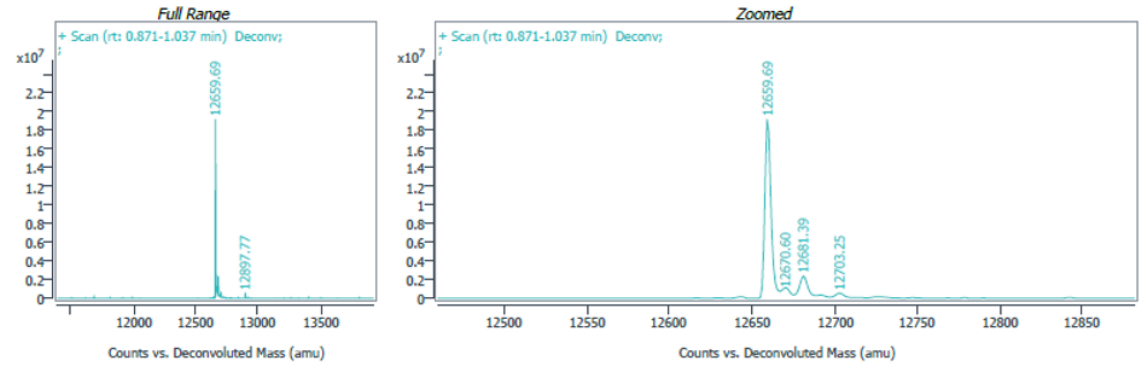
After obtaining pure proteins, both human and mouse ARNT2-PAS-B were subjected to crystallisation trials. While small crystals were obtained for the mouse protein, no crystals could be grown for the human ARNT2-PAS-B despite extensive screening using several attempts with a sparse matrix approach or condition-specific screens. Mouse but not human ARNT2-PAS-B was therefore selected for the HTS campaign. To understand how mouse protein could interact with ligands, sequence and structural comparison with the predicted conformation of the human protein was performed.

A



TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
pSJ2-hARNT2-PAS-B (332-442)	13102	13102	0

B



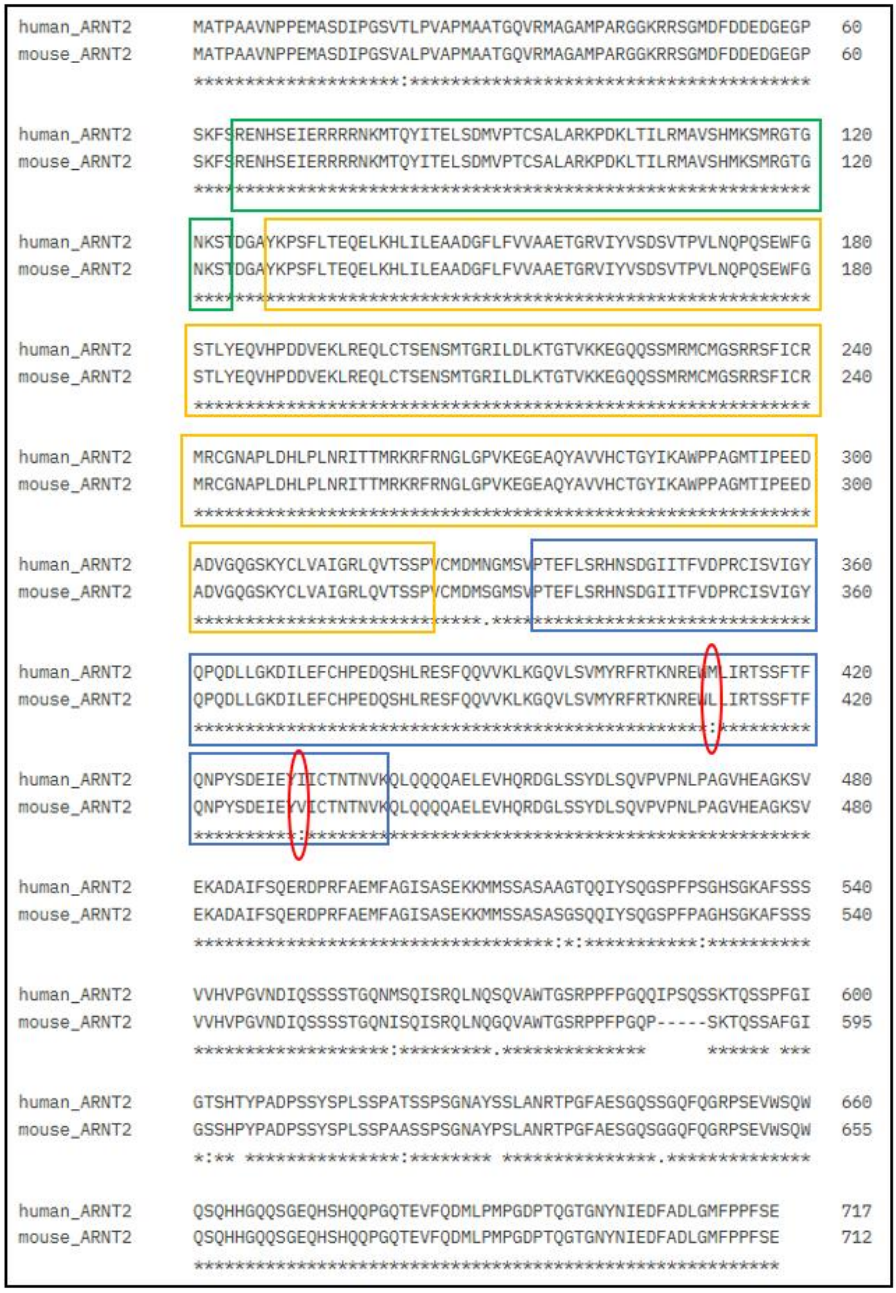
TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
pMKH-mARNT2-PAS-B (332-438)	12659	12659	0

Figure 18. Mass analysis of purified ARNT2-PAS-B proteins by ESI-MS. Deconvoluted spectra and masses for (A) human and (B) mouse TEV-cleaved ARNT2-PAS-B domains.

4.2 Comparison of human and mouse ARNT2 protein

The human and mouse ARNT2 proteins share 97.75% sequence identity (Figure 19). They contain three recognisable domains, with the sequence and boundaries of bHLH and PAS-A domains identical between human and mouse (Figure 19). The boundaries of their PAS-B domains are also identical; however, there are two positions where human and mouse differ (Figure 19). To investigate whether these species-specific differences could impact ligand binding, further structural comparison between two species was performed. By comparing the binding pockets where the control compound KG-548 binds, it can be assessed whether the mouse and human have a similar structural environment, which would suggest that KG-548 (or related compounds) might also bind effectively there.

There is a conserved amino acid difference at residue 412, a methionine in human and a leucine in mouse, that does not cause a charge change in the protein and occurs outside of the predicted ligand binding pocket; therefore, it should not interfere with ligand binding. Whereas the conserved amino acid difference at residue 431, an isoleucine in human and a valine in mouse, is inside the predicted ligand binding pocket, it is expected to change the size of the pocket cavity. With valine side-chain length being a bit shorter than isoleucine, the pocket is left slightly bigger and is expected to lead to the identification of grater hit numbers for the mouse protein. Valine and isoleucine both have two methyl groups pointing into the ligand pocket, but the extra CH_2 group in isoleucine's chain makes it extend more into the pocket (Figure 20). Therefore, it will be important to re-run final hits identified for the mouse protein with the human protein for the comparison between the two species and to distinguish the species selectivity.



Domain			
Protein	bHLH	PAS-A	PAS-B
hARNT2	65-123	128-323	334-439
mARNT2	65-123	128-323	334-439

Figure 19. Sequence alignment of full-length human and mouse ARNT2 proteins. The bHLH, PAS-A, and PAS-B domains are shown in green, yellow, and blue, respectively. The boundaries of functional bHLH, PAS-A, and PAS-B domains are based on the existing crystal structure of mNPAS4-mARNT2 (PDB: 7XI3). The human boundaries were selected after alignment with the mouse sequence. Two substitutions in the PAS-B domain are circled in red.

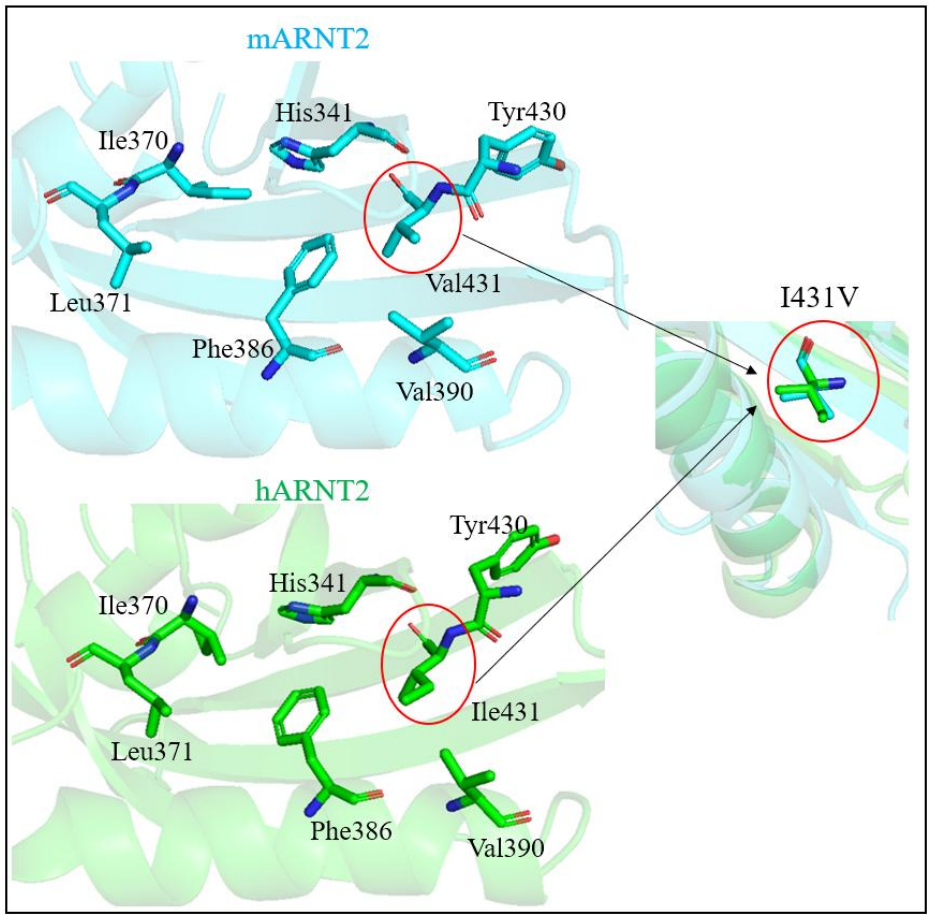


Figure 20. Representative structural comparison of the human and mouse PAS-B binding pocket of ARNT2. The analyses of the PAS-B domains were carried out using PyMOL.The AlphaFold predicted human protein (green) was aligned with the in-house crystal structure of the mouse protein (blue). The side chains of residues inside the binding pockets are shown in ball and stick representation. Position 431 is circled in red.

4.3 Expression and purification of the single PAS-B domain of NPAS1 and NPAS3

While the PAS-B domains of NPAS1 and NPAS3 were structurally characterised as part of the large heterodimeric NPAS-ARNT complexes using the mouse proteins, there is no evidence to date that the singular PAS-B domain of human NPAS1 or NPAS3 can be purified and structurally characterised.[1]

Similar to the ARNT2-PAS-B overexpression and purification described earlier, the recombinant plasmids pSJ2-hNPAS1-PAS-B (residues 302-421) and hNPAS3-PAS-B (residues 319-389) were transformed into BL21-CodonPlus (DE3)-RIL cells for overexpression. The boundaries for human NPAS1/3-PAS-B were based on the existing mouse protein crystal structures within the NPAS1/3-ARNT heterodimer complexes. Since both plasmids have an N-terminal 6xHis-tag, the supernatants containing proteins of interest were loaded on HisTrap FF column after sonication. The eluted proteins were applied on Superdex 75-pg size exclusion chromatography, and fractions were analysed by SDS-PAGE.

Despite numerous attempts, purification of the single PAS-B domain of NPAS1 and NPAS3 could not be achieved. Both singular domains should be eluted after 75 mL column volume. The size exclusion UV trace, however, did not show any protein peaks around that elution time (Figure 21). Importantly, as SDS-PAGE indicated, both proteins accumulated in insoluble fractions of the lysate, and the residual soluble protein was bound to chaperone complexes, supporting the S-75 earlier elution peaks. Various troubleshooting procedures, including different media, temperatures, induction times, induction strengths, additives, and additional ion exchange chromatography, did not help to release proteins from chaperones.

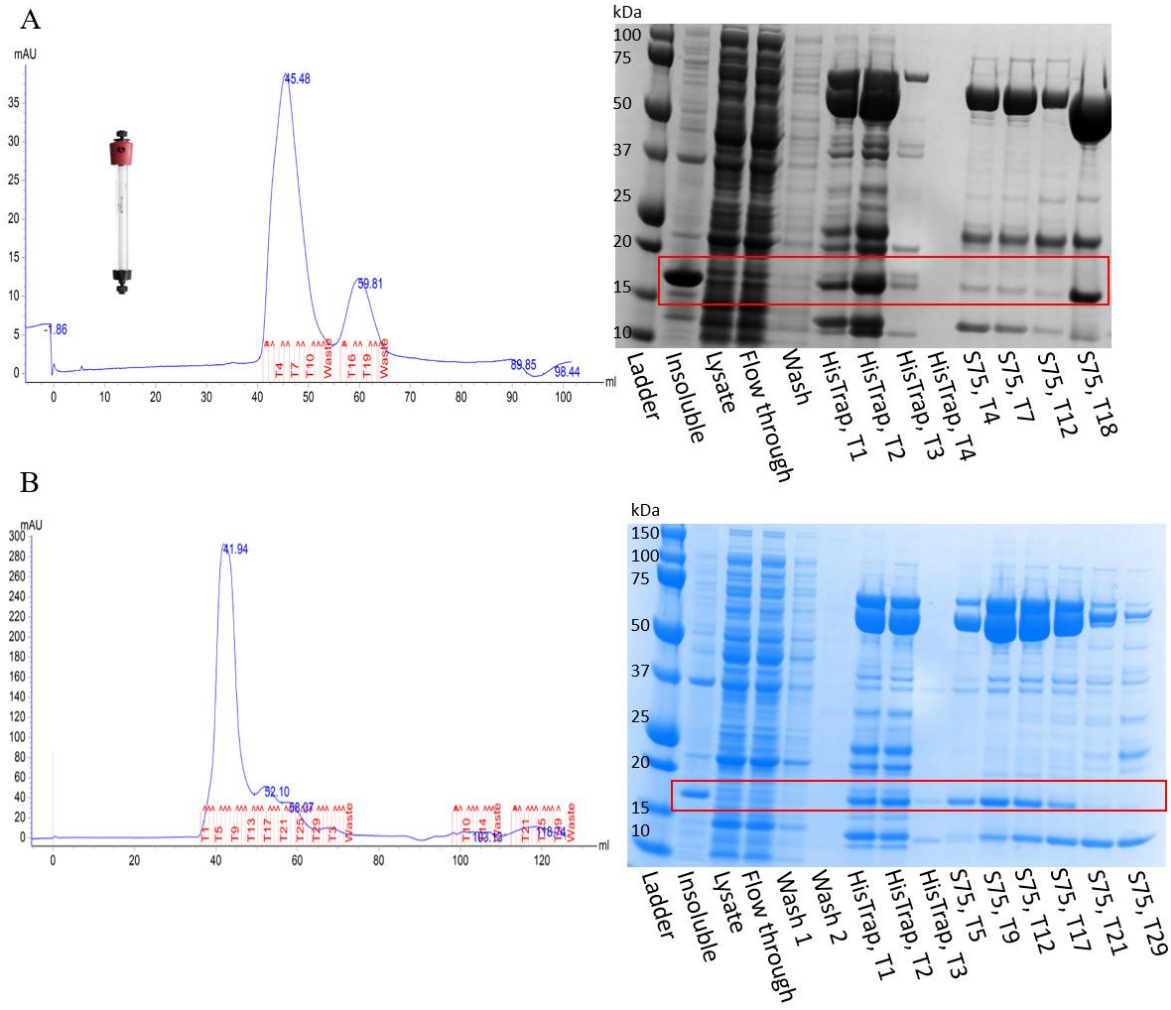


Figure 21. Purification results for the single PAS-B domains of hNPAS1/3. HiLoad 16/600 Superdex 75-pg size exclusion column UV traces and corresponding SDS-PAGE gels for (A) 16.4 kDa hNPAS1-PAS-B and (B) 17.3 kDa hNPAS3-PAS-B.

4.4 Comparison of human and mouse NPAS1 and NPAS3 proteins

As PAS-B domain purification for NPAS1 and NPAS3 was unsuccessful, it was decided to create their tandem bHLH-PAS-A-PAS-B domains. This tandem domain architecture

represents a conserved and functionally integrated contiguous triple-domain unit that mediates DNA binding, heterodimerisation, and potential ligand binding. To increase the chance of successful purification and subsequent crystallisation of proteins, it was decided to generate both human and mouse NPAS1 and NPAS3 constructs.

To identify the differences between human and mouse proteins and to understand whether both species should be used in ligand binding assays, the sequences of human and mouse were compared for NPAS1 and NPAS3. Full-length human and mouse NPAS1 and NPAS3 proteins show 87% and 95% sequence identity, respectively, with most of the amino acid changes predicted to be outside of the ligand binding pockets in the loops connecting secondary structural elements (Figure 22).

The amino acid sequence alignments shown in Figure 22 illustrate a comparison between the human and mouse for each of the NPAS1 and NPAS3 proteins. In both proteins, the bHLH domain is highly conserved. For NPAS1, there are numerous amino acid changes in the PAS-A and PAS-B domains (Figure 22A). Unlike NPAS1, NPAS3 PAS-B is absolutely conserved between human and mouse. The NPAS3 PAS-A exhibits subtle amino acid differences between human and mouse, and the three mismatches correspond to a flexible loop region, as indicated by previously published protein structures (Figure 22B).[1] Therefore, either mouse or human NPAS3 could be used for further chemical screening.

NPAS1 domains			
	bHLH	PAS-A	PAS-B
hNPAS1	52-109	128-291	298-420
mNPAS1	52-109	128-292	299-421

NPAS3 domains			
	bHLH	PAS-A	PAS-B
hNPAS3	52-111	143-319	331-448
mNPAS3	59-118	148-324	336-453

Figure 22. Sequence comparison of human and mouse NPAS1 and NPAS3 proteins. (A) NPAS1 and (B) NPAS3 full-length proteins. The bHLH, PAS-A, and PAS-B functional domains are shown in green, orange, and blue, respectively. The boundaries of functional bHLH, PAS-A, and PAS-B domains are based on the existing crystal structure of the heterodimeric mNPAS1-mARNT (PDB: 5SY5) and mNPAS3-mARNT (PDB: 5SY7). The

human boundaries were selected after alignment with the mouse sequence. Mismatches in each domain are circled in red.

Further structural analysis, using a 3D comparison of the predicted human NPAS1 with the mouse NPAS1 structure, was used to determine the exact location of each different residue, and whether it could result in potential differences in binding interactions. It was hypothesised that most of the substitutions would not cause any change in ligand binding, because they are either conservative substitutions, correspond to a connecting loop region, or point away from the predicted ligand binding pocket (Figure 23, Table 2). However, there are two amino acid differences which might influence ligand binding. In the PAS-A domain, residue 177 points its side chain inside the pocket. In the human NPAS1 PAS-A domain, the methylthio group of M177 might play an important role in ligand selectivity as compared to the L177 hydrophobic side chain in mouse NPAS1.[166] In the PAS-B domain, residue 364 that points away from the ligand binding pocket is a valine in mouse and a methionine in human, and may influence binding selectivity and interactions. These analyses suggest that it is better to focus on human NPAS1 chemical screening.

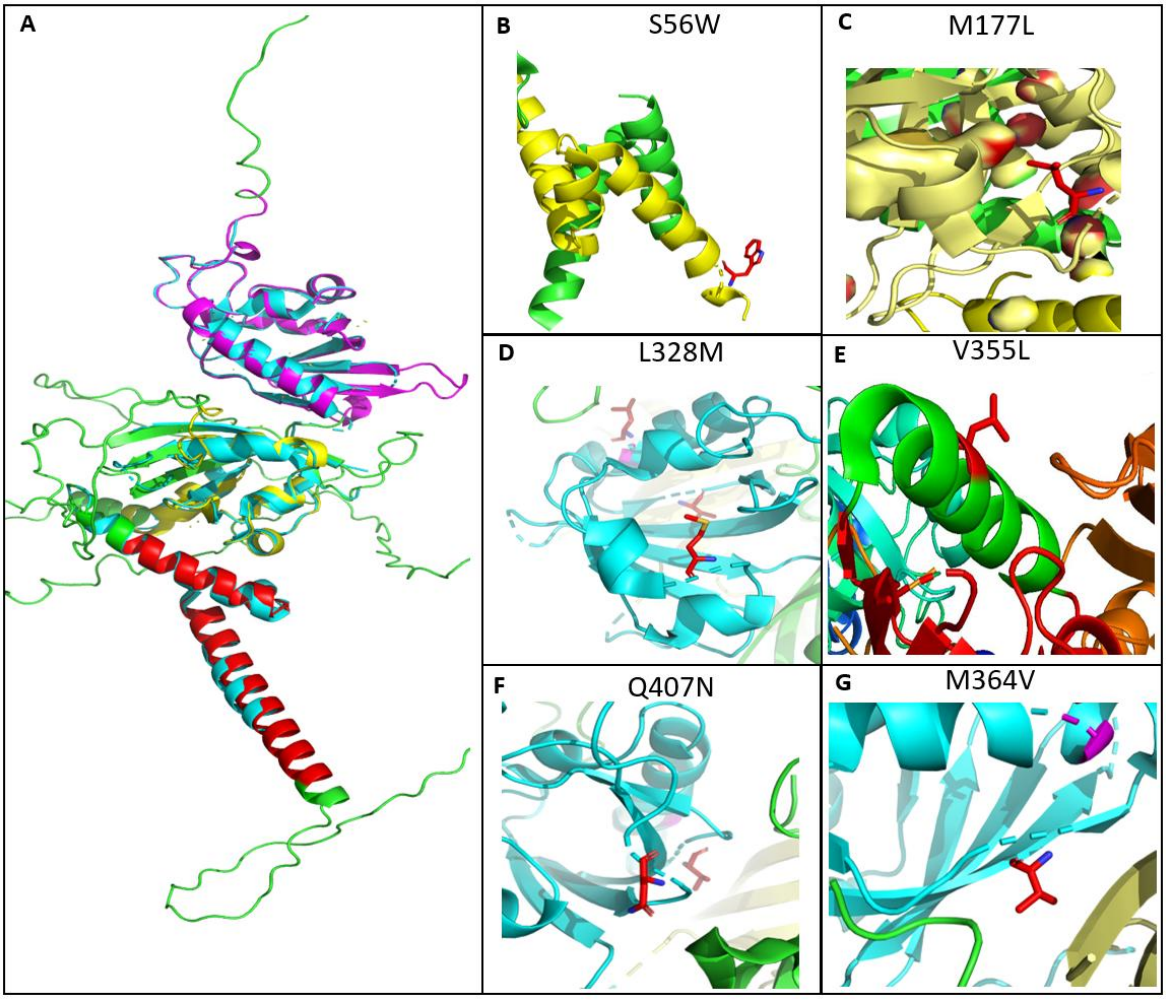


Figure 23. Structural comparison between human and mouse NPAS1 proteins. (A) 3D structural comparison between full-length human and mouse NPAS1 proteins using PyMOL. The AlphaFold predicted human protein (green) was aligned with the existing 5SY5 PDB structure (blue) of the mouse protein. The human bHLH, PAS-A, and PAS-B domains are shown in red, yellow, and purple, respectively. (B-G) Zoomed-in position of each different residue in NPAS1, including S56W, M177L, L328M, V355L, Q407N, and M364V.

Human vs. mouse NPAS1	Characteristics of mismatches in NPAS1 functional domains	Potential impact on ligand binding
S56W	In the bHLH domain pointing away from dimer interface and DNA.	Likely no impact.
M177L	In the PAS-A domain facing the hydrophobic side chain inside the pocket. Leucine is smaller, and methionine has the methylthio group.	Could potentially impact ligand binding.
TPTP→ AASI	Corresponds to a loop in the PAS-A domain.	Likely no impact.
SS→SSS	Corresponds to a loop in the PAS-A domain.	Likely no impact.
A224V	Corresponds to a loop in the PAS-A domain.	Likely no impact.
LTKVPPSSLV→ PTEASPAFRA	Corresponds to a loop in the PAS-A domain.	Likely no impact.
H261N	Corresponds to a loop in the PAS-A domain.	Likely no impact.
H279R	Corresponds to a loop in the PAS-A domain.	Likely no impact.
L328M	In the PAS-B domain facing the hydrophobic side chain towards the pocket. Leucine is smaller and methionine has the methylthio group.	Could have an impact.
V355L	In the PAS-B, the side chain of residue 356 points outside of the protein structure. A conservative substitution.	Likely no impact.
M364V	In the PAS-B, the hydrophobic side chain points outside of the PAS-B domain. Methionine has the methylthio group.	Could potentially impact ligand binding.
S389N and P393T	Both occur in loop regions in PAS-B.	Likely no impact.
Q407N and G411S	Both occur in loop regions in PAS-B.	Likely no impact.

Table 2. Amino acid differences between human and mouse NPAS1 in their functional domains.

4.5 Generation and purification of NPAS1 and NPAS3 triple domains

Based on structural analyses, it was concluded to generate both human and mouse contiguous bHLH-PAS-A-PAS-B triple domains only for NPAS3 protein. Due to the potential interference of mouse protein with ligand binding, only a human construct was decided to be designed for NPAS1. Therefore, DNA sequences corresponding to the contiguous bHLH-PAS-A-PAS-B triple domains of hNPAS1 (residues 43-423), hNPAS3 (residues 19-437), and mNPAS3 (residues 56-455) were amplified by polymerase chain reaction (PCR). Boundaries for the human constructs were determined based on their homologies with the existing mouse structures. For bacterial expression, an N-terminal 6xHis-tag, followed by a TEV protease cleavage site, were introduced in each construct.

It was hypothesised, however, that the isolated triple-domain protein architecture might not express or purify well. As was observed with both singular PAS-B domains, proper folding of these domains may require assistance from either their dimerisation partner or from cellular factors like molecular chaperones. It was anticipated that if overexpression or purification of the bHLH-PAS-A-PAS-B triple domains were unsuccessful, a heterodimeric protein construct with its partner ARNT2 would be generated instead.

Successful co-expression and purification of heterologous proteins is dependent on several factors, including plasmid compatibility, the *E. coli* host strain, and the nature of proteins being expressed. To avoid potential double transformation in case of unsuccessful single triple-domain purification, a pETDuet-1 vector was chosen as it is specifically designed for

the co-expression of two targeted genes. pETDuet-1 contains two multiple cloning sites, each of which is preceded by a T7 promoter, a lac operator, and a ribosome binding site (Figure 24).

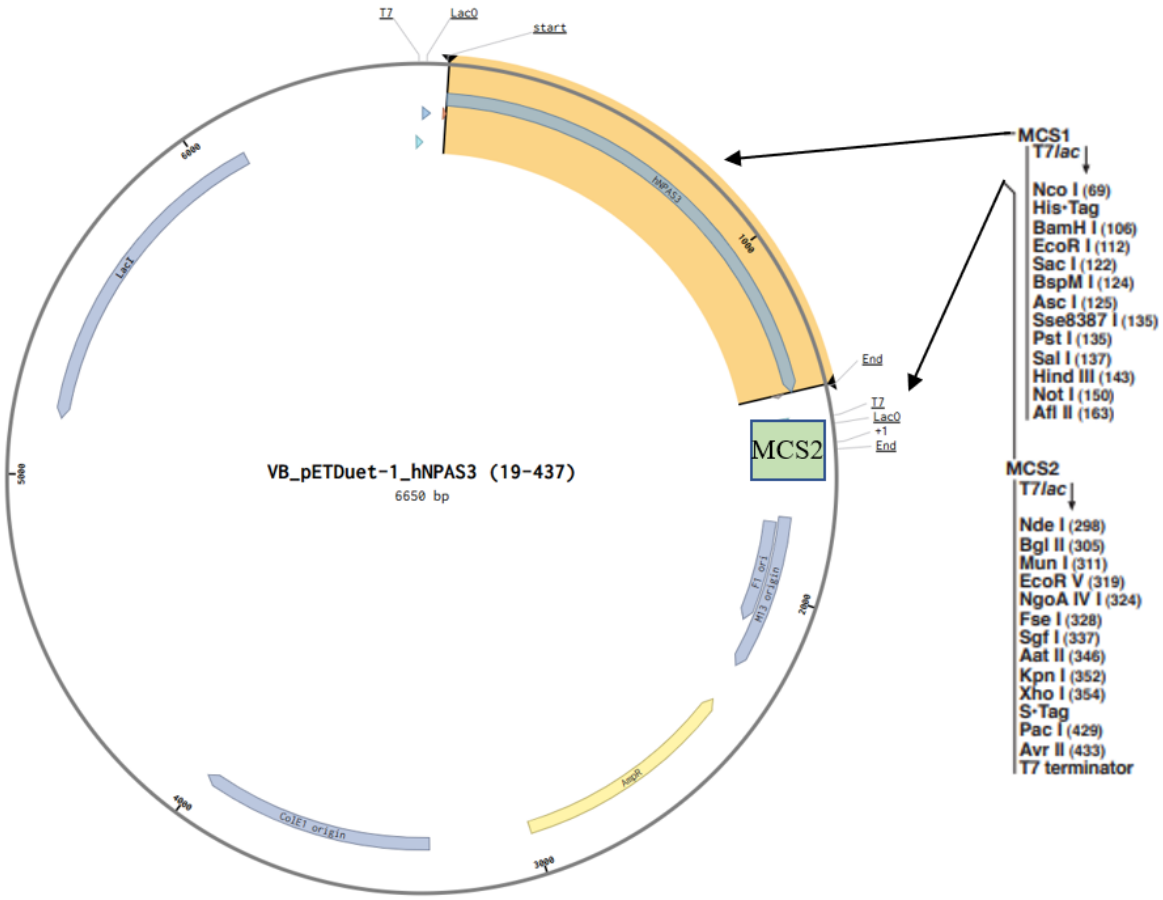


Figure 24. Representative vector map of pETDuet-1 with hNPAS3 (residues 19-437). The NPAS3 gene occupying the first cloning site is highlighted with orange. It is inserted between BamHI and HindIII restriction sites. The second cloning site would be used for cloning the ARNT2 gene if purification of the NPAS3 triple domains were unsuccessful and is represented with a green square. The ampicillin resistance gene is shown in yellow. Restriction enzyme options are listed to the right.

To confirm the size of the PCR products, purified samples and the pETDuet-1 vector, previously digested with BamHI and HindIII restriction enzymes, were run on an agarose gel. The DNA bands of approximately 1,300 bases correspond to the mass of hNPAS1, hNPAS3, and mNPAS3 triple domains (Figure 25A), and 5,000 bases bands match the mass of the linearised pETDuet-1 vector (Figure 25B).

Each of hNPAS1, hNPAS3, and mNPAS3 PCR fragments were combined with linearised pETDuet-1 in a Gibson assembly reaction. Gibson Assembly was used as it allows the joining of multiple DNA fragments in a vector without digestion of PCR products. Following DH5α transformation, plasmid DNAs were purified and run on an agarose gel (Figure 25C). The presence of two DNA bands of approximately 5,000 and 1,300 bases corresponding to the mass of pETDuet-1 and the triple domain of hNPAS1/3 and mNPAS3 indicates a successful Gibson Assembly reaction. Triple domain constructs pETDuet-1-hNPAS1, pETDuet-1-hNPAS3, and pETDuet-1-mNPAS3 were confirmed by sequencing (Appendix 2).

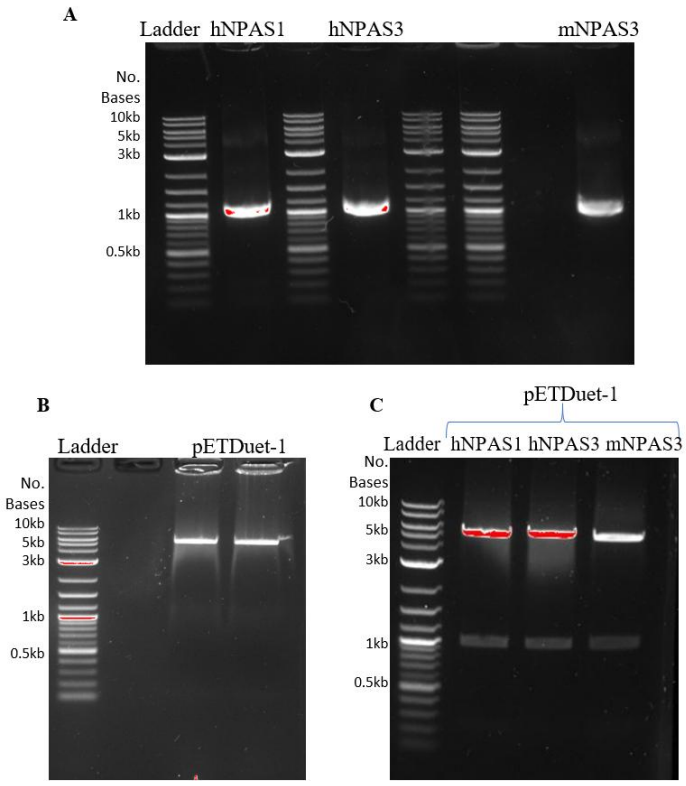


Figure 25. DNA gels with one per cent agarose. (A) Purified PCR products of hNPAS1, hNPAS3, and mNPAS3 with approximate mass of 1300 bases, correspond to the mass of triple domain of each protein. (B) Linearised pETDuet-1 vectors after BamHI and HindIII restriction enzymes digestion with a mass of 5,420 bases. (C) DNA products of the Gibson Assembly reaction digested with BamHI and HindIII restriction enzymes. Purified plasmids contain two DNA bands of approximately 5,000 and 1,300 bases, indicating the presence of pETDuet-1 and hNPAS1, or hNPAS3, or mNPAS3 triple domains.

The pETDuet-1-hNPAS1/3 and mNPAS3 were transformed into BL21-CodonPlus (DE3)-RIL cells and overexpressed. Purification of a three-domain protein complex, including a highly charged bHLH DNA binding domain, may require multiple chromatography steps. After IMAC, SDS-PAGE showed that whilst the proteins of interest were purified relative to the soluble lysate, other protein impurities were still retained (Figure 26D). The eluates were therefore further purified via cation-exchange chromatography (CE), which uses a negatively

charged resin (SP-Sepharose) to bind positively charged molecules like the bHLH domain. Despite this, CE resulted in multiple protein peaks and failed to dissociate proteins from chaperone complexes, which remained bound throughout the fractions (Figure 26B, D).

Proteins were further purified by Superdex 200-pg size exclusion chromatography. Based on the S200 column calibration curve (MW/Volume), both proteins should be eluted after 80 mL column volume (Appendix 3). The size exclusion UV trace, however, showed multiple protein peaks (Figure 26C). It seems that the previous hypothesis regarding a supporting partner for proper protein folding is correct, and both NPAS1 and NPAS3 constructs would therefore be co-expressed with their dimerisation partner ARNT2.

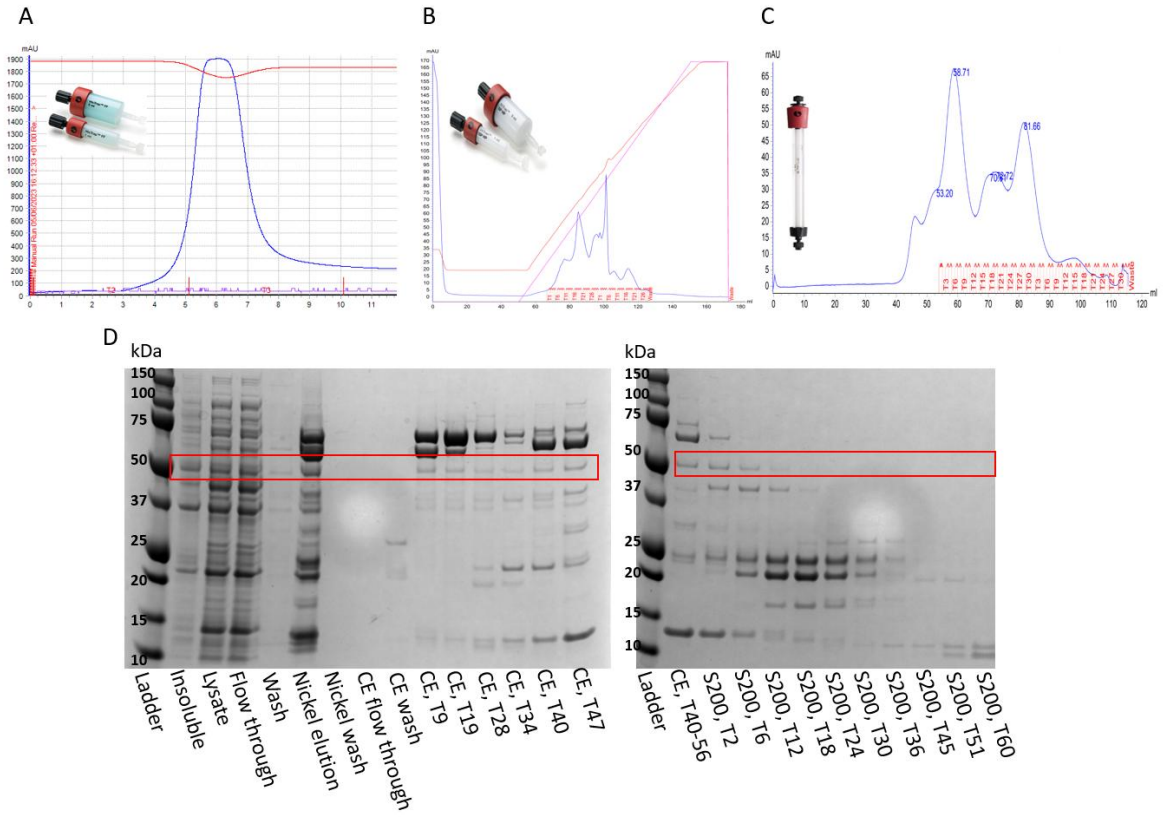


Figure 26. A representative example of NPAS3 triple domain purification. (A) HisTrap FF affinity column UV traces of the hNPAS3 triple bHLH-PAS-A-PAS-B domain. (B) HiTrap SP HP cation exchange column UV traces of the hNPAS3 triple domain. (C) HiLoad 16/600

Superdex 200-pg size exclusion column UV traces of the hNPAS3 triple domain. (D) Corresponding SDS-PAGE for each purification step for 50 kDa bHLH-PAS-A-PAS-B pETDuet-1-hNPAS3.

4.6 Generation of NPAS1 and NPAS3 heterodimer complexes with ARNT2

Since neither the singular domain nor the triple domain could be purified, triple domains of NPAS1 and NPAS3 will be overexpressed and purified as heterodimer complexes with functional domains of their partner ARNT2 (Figure 27). Previously, this heterodimer approach was successfully implemented for NPASs, HIFs, and AHR, to generate stable multidomain constructs. In this way, the crystal structures of HIF-2α-ARNT, NPAS1/3-ARNT, NPAS4-ARNT/2, and AHR-ARNT were successfully determined.[1,51,83,106]

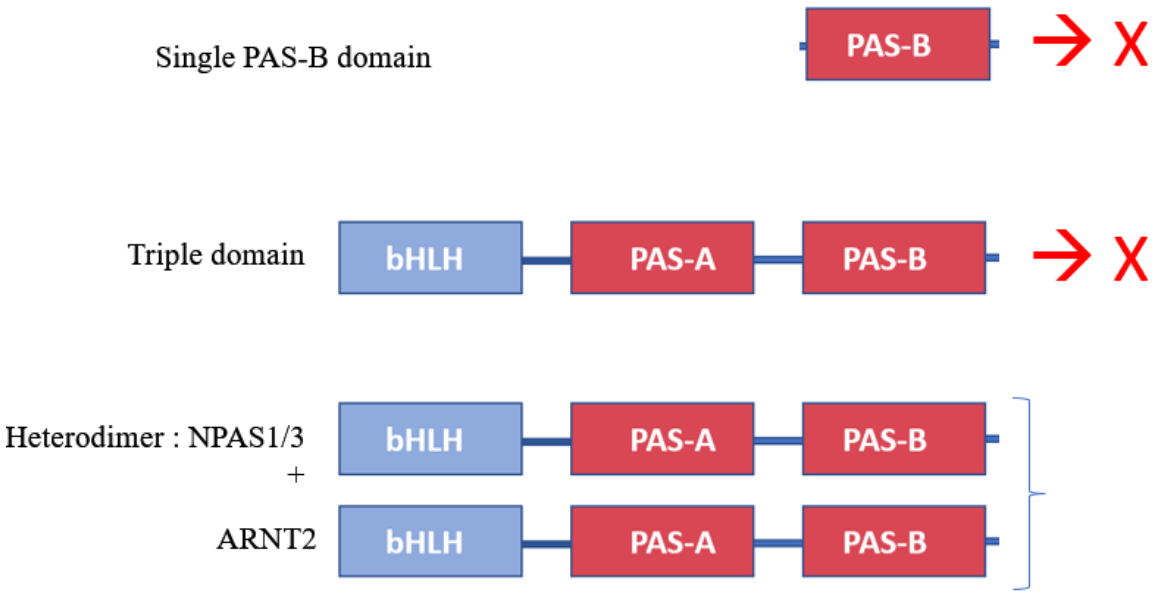


Figure 27. Schematic representation of NPAS1 and NPAS3 constructs that were prepared.

To generate a partner's inserts, DNA sequences of contiguous bHLH-PAS-A-PAS-B domains of hARNT2 (residues 57-438) and mARNT2 (residues 49-449) with N-terminal Flag-tag and a TEV protease cleavage site were amplified by PCR. Boundaries for the human constructs were determined based on their homologies with the existing mouse structures. Purified ARNT2 inserts were run on an agarose gel to confirm the correct size. The DNA bands of approximately 1,300 bases correspond to the mass of human and mouse ARNT2 triple domains (Figure 28A).

To assemble heterodimers, previously generated triple domain pETDuet-1-hNPAS1/3 and pETDuet-1-mNPAS3 constructs were digested with KpnI and BglII restriction enzymes. To insert the second gene of interest, linearised plasmids were combined with purified human or mouse ARNT2 fragments in a Gibson Assembly reaction. After DH5α transformation, purified plasmid DNA was digested with KpnI and BglII restriction enzymes and screened for the correct plasmid product (Figure 28B). The identities of hNPAS1-hARNT2, hNPAS3-hARNT2, and mNPAS3-mARNT2 heterodimer constructs in the pETDuet-1 vector were confirmed by sequencing (Appendix 4).

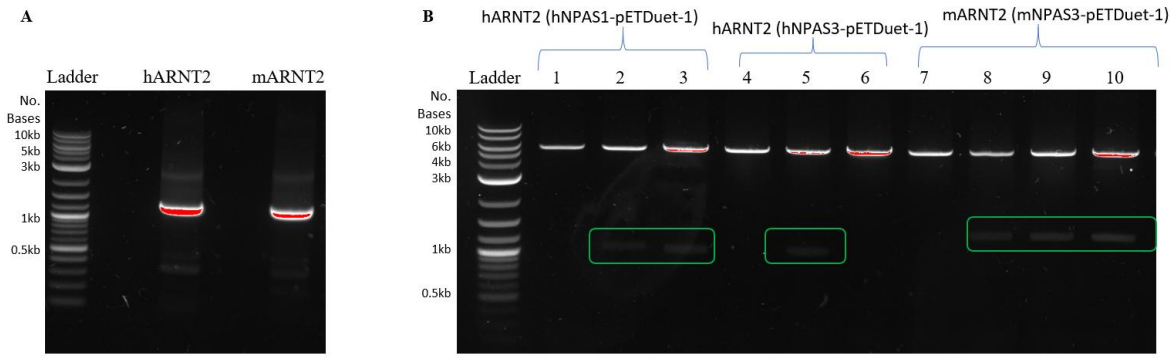


Figure 28. DNA gels with one per cent agarose. (A) Purified PCR products of the human and mouse ARNT2 triple domain with approximate mass of 1300 bases. (B) DNA products of the Gibson Assembly reaction digested with KpnI and BglII restriction enzymes. Plasmids

purified from colonies 2-3, 5, and 8-10 contain two DNA bands of approximately 6,000 and 1,300 bases, indicating the presence of pETDuet-1-hNPAS1/3 and mNPAS3 together with their partner's human or mouse ARNT2 triple domain.

4.7 Expression and purification of NPAS1 and NPAS3 heterodimers with ARNT2

The recombinant plasmids pETDuet-1-hNPAS1/3-ARNT2 and pETDuet-1-mNPAS3-ARNT2 were transformed into BL21-CodonPlus (DE3)-RIL cells, and proteins were expressed as previously described for similar heterodimer constructs[1,83,162]. Proteins were purified by IMAC, and all fractions were analysed by SDS-PAGE.

SDS-PAGE relies on boiling the protein solution with reducing and denaturing agents that break the heterodimer complex. Unlike NPAS1/3, ARNT2 does not have 6xHis-tag. Therefore, ARNT2 would be present only in the HisTrap Ni elution fraction as a stable complex with the NPAS partner. The gels clearly showed that expression did produce the desired proteins. Protein bands at 44 kDa, 50 kDa, and 48 kDa correspond to the mass of hNPAS1, hNPAS3, and mNPAS3 monomers (Figure 29E-F and 30C). The presence of 46 kDa and 48 kDa protein bands in human and mouse constructs suggests that ARNT2 is bound to NPAS1 and NPAS3 as a heterodimer.

Further purification of the proteins was achieved by using CE. The UV trace during CE showed one major protein peak for both hNPAS1/3-ARNT2 and mNPAS3-ARNT2 (Figure 29A-B and 30A). The SDS-PAGE results indicated that after CE, major protein impurities disappeared (Figure 29E-F and 30C).

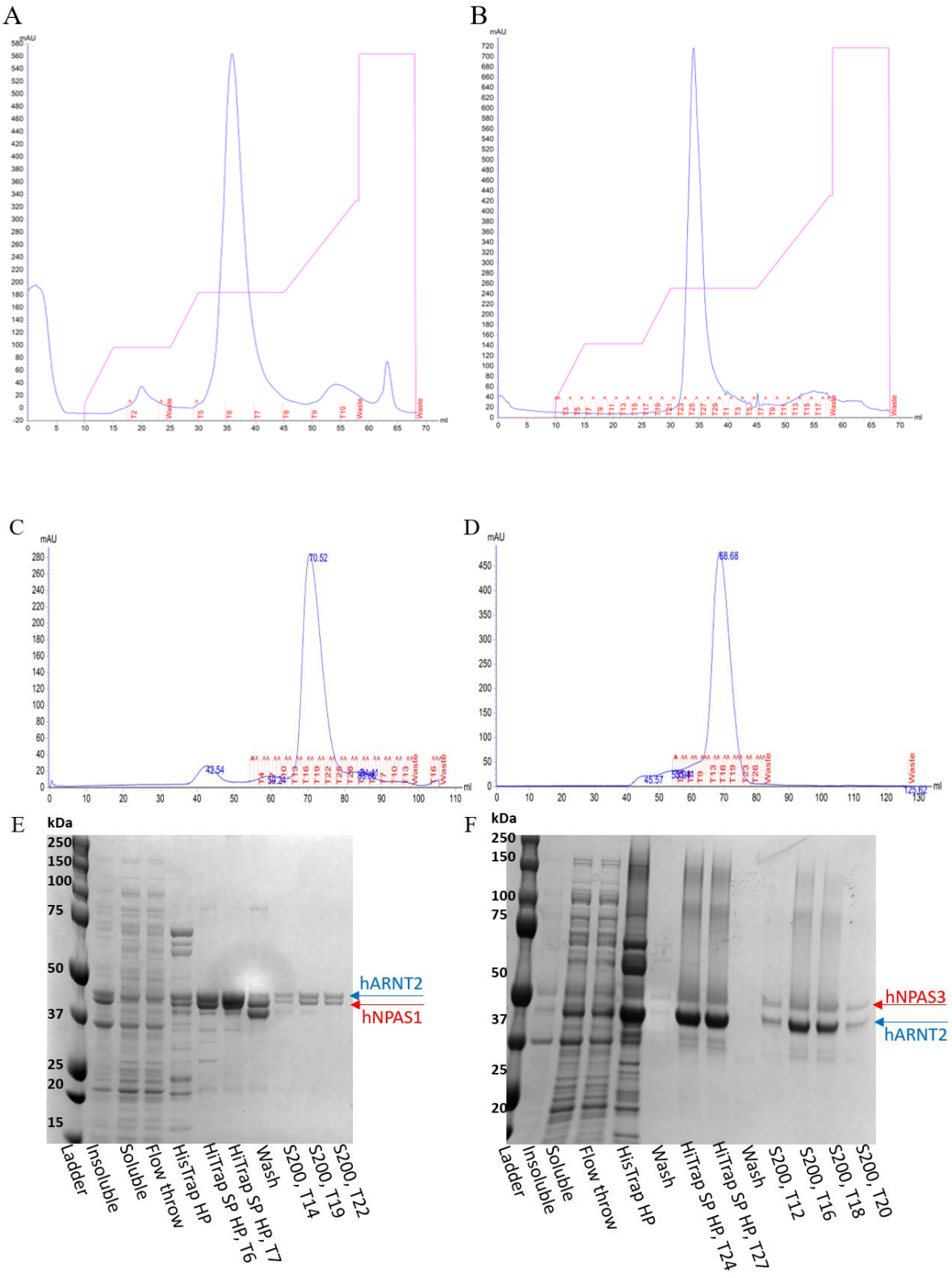


Figure 29. Purification of human NPAS1/3 heterodimers with ARNT2. Cation exchange and size exclusion chromatography, and corresponding SDS-PAGE. HiTrap SP HP cation

exchange column UV traces of (A) hNPAS1-ARNT2 and (B) hNPAS3-ARNT2. HiLoad 16/600 Superdex 200-pg size exclusion column UV traces of (C) hNPAS1-ARNT2 and (D) hNPAS3-ARNT2. SDS-PAGE for (E) hNPAS1-ARNT2 and (F) hNPAS3-ARNT2 protein purifications.

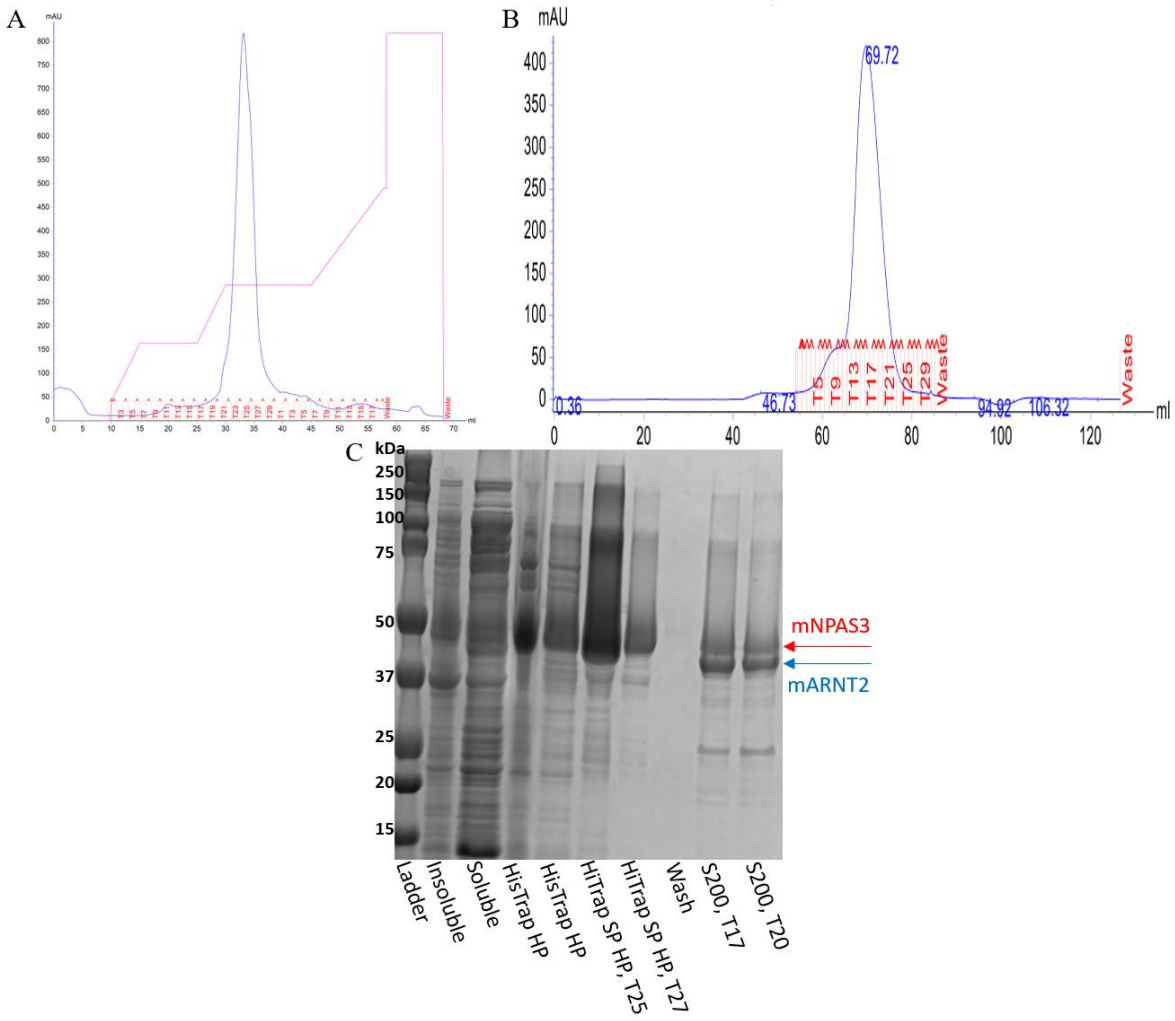


Figure 30. Purification of mouse NPAS3 heterodimer with ARNT2. Cation exchange and size exclusion chromatography, and corresponding SDS-PAGE for mNPAS3-ARNT2.(A) HiTrap SP HP cation exchange column UV traces of mNPAS3-ARNT2. (B) HiLoad 16/600 Superdex 200-pg size exclusion column UV traces of mNPAS3-ARNT2. (C) SDS-PAGE for each step of protein purification.

The subsequent size exclusion chromatography showed a single peak for both hNPAS1/3-ARNT2 and mNPAS3-ARNT2, indicating the presence of a high-purity protein (Figure 29C-D and 30B). The gels confirmed the purity of three proteins by showing two clear bands corresponding to hNPAS1, hNPAS3, mNPAS3, and their partner human or mouse ARNT2 (Figure 29E-F and 23C). Using size exclusion chromatography, the size of the native protein in solutions was investigated to determine whether proteins purified as monomers or heterodimers.

The molecular weights of hNPAS1, hNPAS3, and mNPAS3 monomers are 44 kDa, 50 kDa, and 48 kDa, respectively. Based on the S200 column calibration curve (MW/Volume), all monomeric proteins should be eluted at approximately 83-85 mL (Appendix 3). The expected elution volume of hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 heterodimers is approximately 70 mL, as their molecular weights are 91 kDa, 97 kDa, and 93 kDa, respectively. The size exclusion UV trace showed 70, 68, and 69 mL of elution volumes for hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2, respectively, indicating that all proteins were expressed and purified as heterodimers.

4.8 Western blot analysis of the heterodimers NPAS1-ARNT2 and NPAS3-ARNT2

To demonstrate that the two protein bands present in each SDS-PAGE sample (Figure 29E-F, and 30C) correspond to 6xHis-tagged hNPAS1, hNPAS3, mNPAS3, and Flag-tagged human and mouse ARNT2, Western blots were run to confirm the presence of the tags.

Different volumes from each of hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 protein samples previously concentrated to 1 mg/mL were loaded into three identical gels.

SDS-PAGE, which was run to visualise all protein content, detected a concentration-dependent pattern of purified hNPAS1/3-ARNT2 and mNPAS3-ARNT2 proteins (Figure 31A). The gel also indicated that at lower protein concentration, hNPAS1 and human and mouse NPAS3 have a 1:1 ratio with their partner ARNT2. Compared to hNPAS1-ARNT2, higher-concentrated human and mouse NPAS3-ARNT2 samples produce smears, and the NPAS3:ARNT2 ratio decreases. This generates the hypothesis that some NPAS3 molecules are getting trapped in the gel.

Western blots confirmed the presence of 6xHis-tagged hNPAS1/3 and mNPAS3, and Flag-tagged human and mouse ARNT2 (Figure 31B-C). Blots also demonstrated a concentration-dependent pattern of all proteins aligning with SDS-PAGE. High-concentration samples of human and mouse NPAS3 also showed smears with anti-His antibodies. This supports the hypothesis that it is NPAS3 molecules getting trapped in the gel, but not higher-MW-contaminants present in the protein sample.

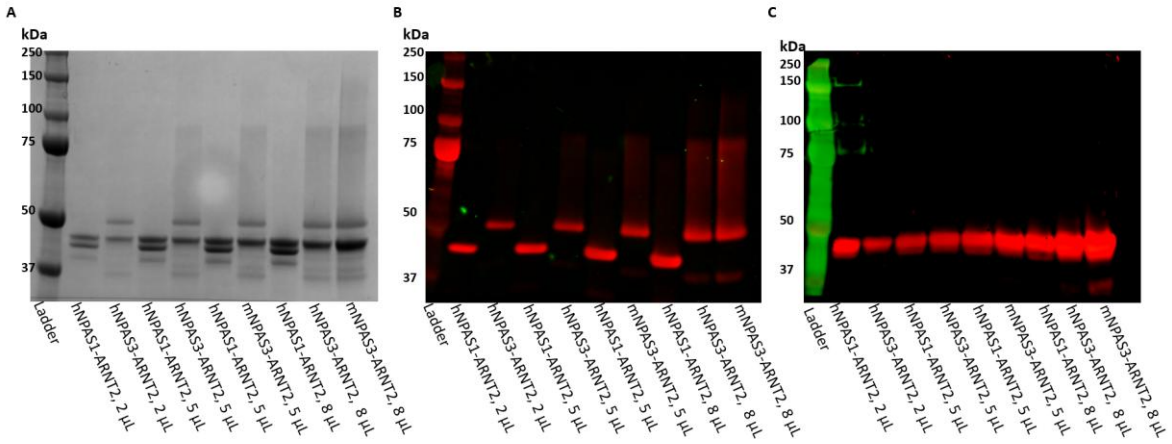


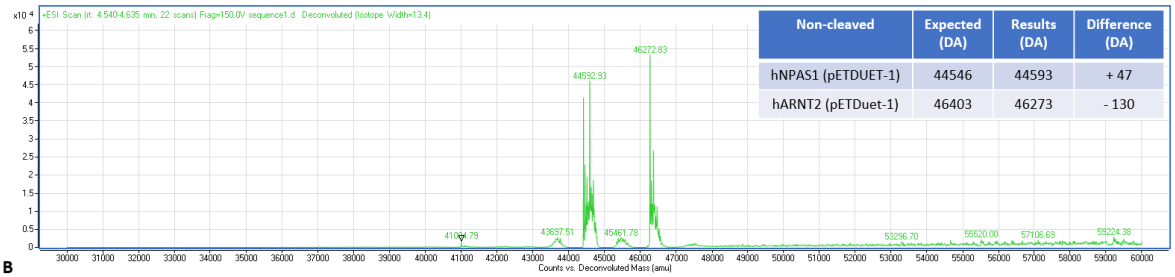
Figure 31. SDS-PAGE and Western blots with anti-His and anti-Flag antibodies. Purified hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 proteins were concentrated to 1 mg/mL. Different volumes from each hNPAS1-ARNT2, hNPAS3-ARNT2 and mNPAS3-ARNT2 samples were loaded into three identical gels: (A) Instant blue stain, (B) Anti-His blot, to visualise purified 6xHis-tagged hNPAS1/3 and mNPAS3 proteins, and (C) Anti-Flag blot, to confirm the presence of Flag-tagged human and mouse ARNT2.

4.9 Mass spectrometry analysis of hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 heterodimers

After successful purification of hNPAS1/3-ARNT2 and mNPAS3-ARNT2 heterodimers, the intact mass and the purity of the proteins were analysed by ESI-MS. According to ESI-MS results, each sample contained two major species, indicating that all three proteins were expressed as heterodimers with their partner ARNT2 (Figure 32).

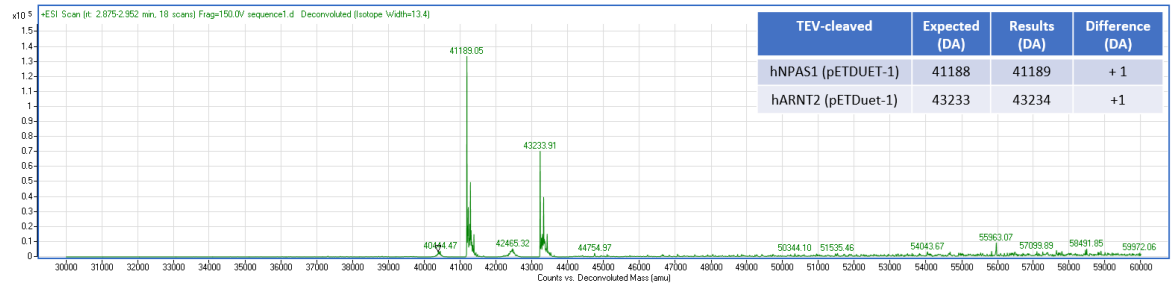
The purified proteins had a subtle mass difference from their predicted masses. The constructs used for purification, hNPAS1, hNPAS3, mNPAS3, hARNT2, and mARNT2 have the expected masses of 44,546 Da, 50,457 Da, 48,290 Da, 46,403 Da, and 48,593 Da, respectively (Figure 32A, C, E). The human and mouse ARNT2 obtained masses of 46,273 Da and 48,464 Da correspond to the protein having lost a methionine (-131 Da) and can be explained by the cleavage of the N-terminal initiator methionine.[167] Methionine aminopeptidase (MAP), present in *E. coli*, commonly cleaves methionine when adjacent to glycine and alanine. For both human and mouse ARNT2 constructs, N-terminal methionine is followed by alanine.

A



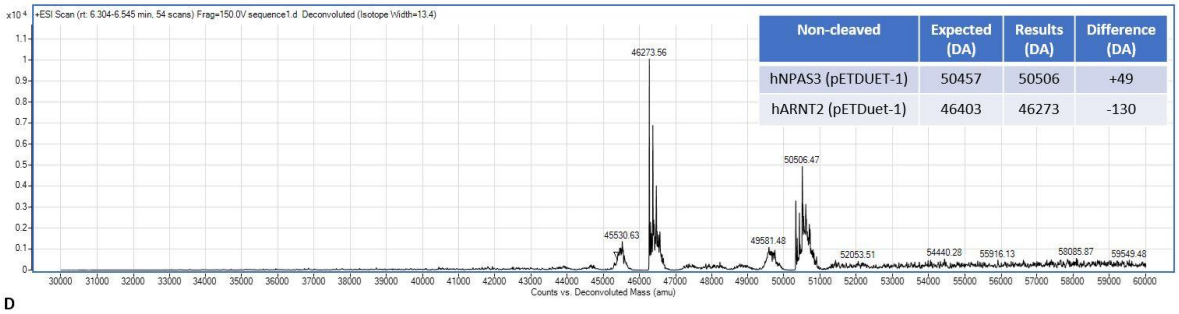
Non-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hNPAS1 (pETDUET-1)	44546	44593	+ 47
hARNT2 (pETDuet-1)	46403	46273	- 130

B



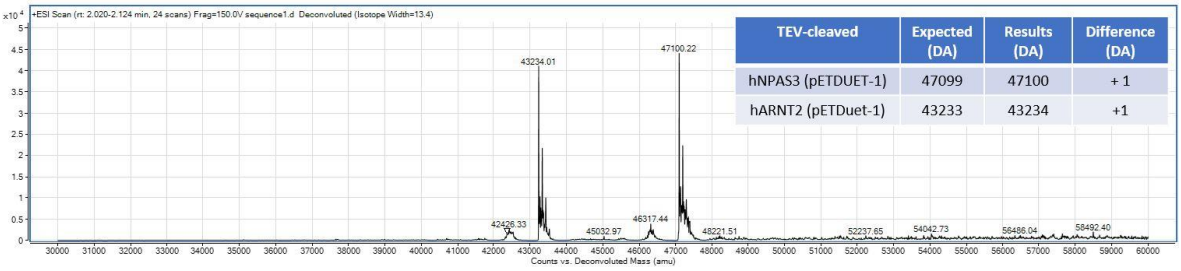
TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hNPAS1 (pETDUET-1)	41188	41189	+ 1
hARNT2 (pETDuet-1)	43233	43234	+1

C



Non-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hNPAS3 (pETDUET-1)	50457	50506	+49
hARNT2 (pETDuet-1)	46403	46273	-130

D



TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hNPAS3 (pETDUET-1)	47099	47100	+ 1
hARNT2 (pETDuet-1)	43233	43234	+1

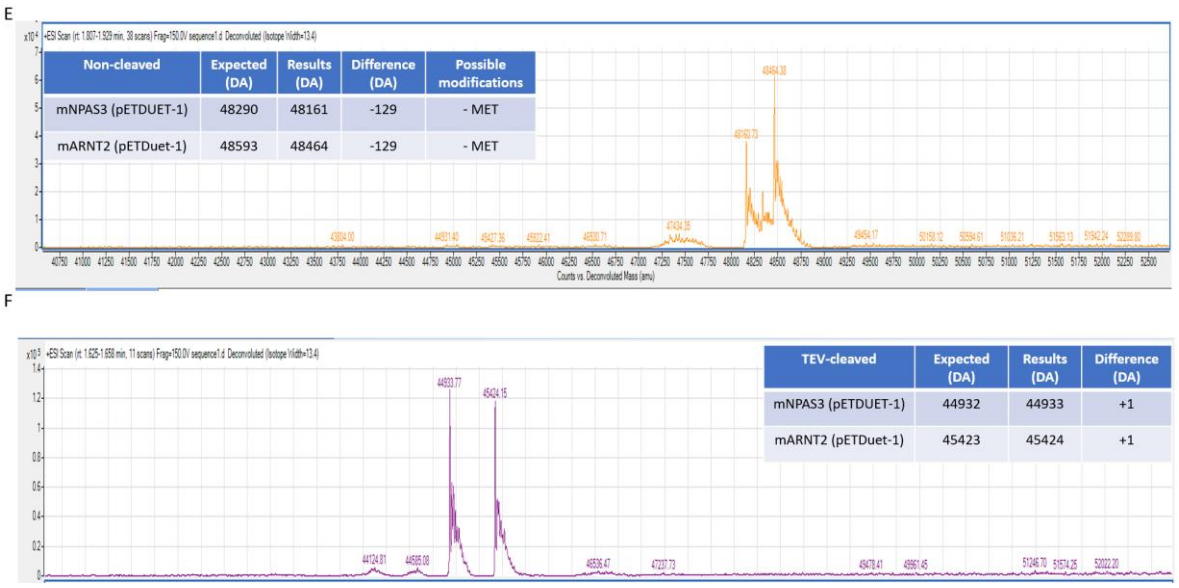


Figure 32. Intact mass analysis of the purified hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 proteins by ESI-MS. (A), (C), and (E): Deconvoluted masses for hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 proteins. (B), (D), and (F): Deconvoluted masses for TEV cleaved hNPAS1-ARNT2, hNPAS3-ARNT2, and mNPAS3-ARNT2 proteins.

The N-terminal methionine is also cleaved in hNPAS1, hNPAS3 and mNPAS3 proteins, as it is followed by glycine in all constructs. Besides the loss of initial methionine, the addition of a gluconate group (+178 Da) occurs on both hNPAS1 and hNPAS3 proteins. In *E. coli*, spontaneous α-N-6-Phosphogluconoylation occurs on proteins with the N-terminus Gly-Ser-Ser-6xHis - as in the case of hNPAS1 and hNPAS3 constructs (Figure 33).[168]

To confirm that the extra mass is located at the N-terminal region and hence does not affect the internal sequence of the proteins, three heterodimers were digested with the TEV protease to remove all tags. The expected exact masses of 41,188 Da, 47,099 Da, 43,233 Da, 44,933 Da, and 45,424 Da were observed for TEV-cleaved hNPAS1, hNPAS3, hARNT2, mNPAS3,

and mARNT2, respectively (Figure 32B, D, F). Since all extra masses are located at the N-terminal region, it is unlikely that these tags will interfere with the behaviour of the NPAS1/3 or ARNT2 proteins. For ease of use and to maximise the protein yields, uncut proteins were used for further HTS campaign.

As the purification of both human NPAS1-ARNT2 and NPAS3-ARNT2 heterodimers was possible, human but not mouse NPAS3-ARNT2 was advanced for future assays. The mouse heterodimer would still be used for crystallisation trials to increase the likelihood of obtaining protein crystals.

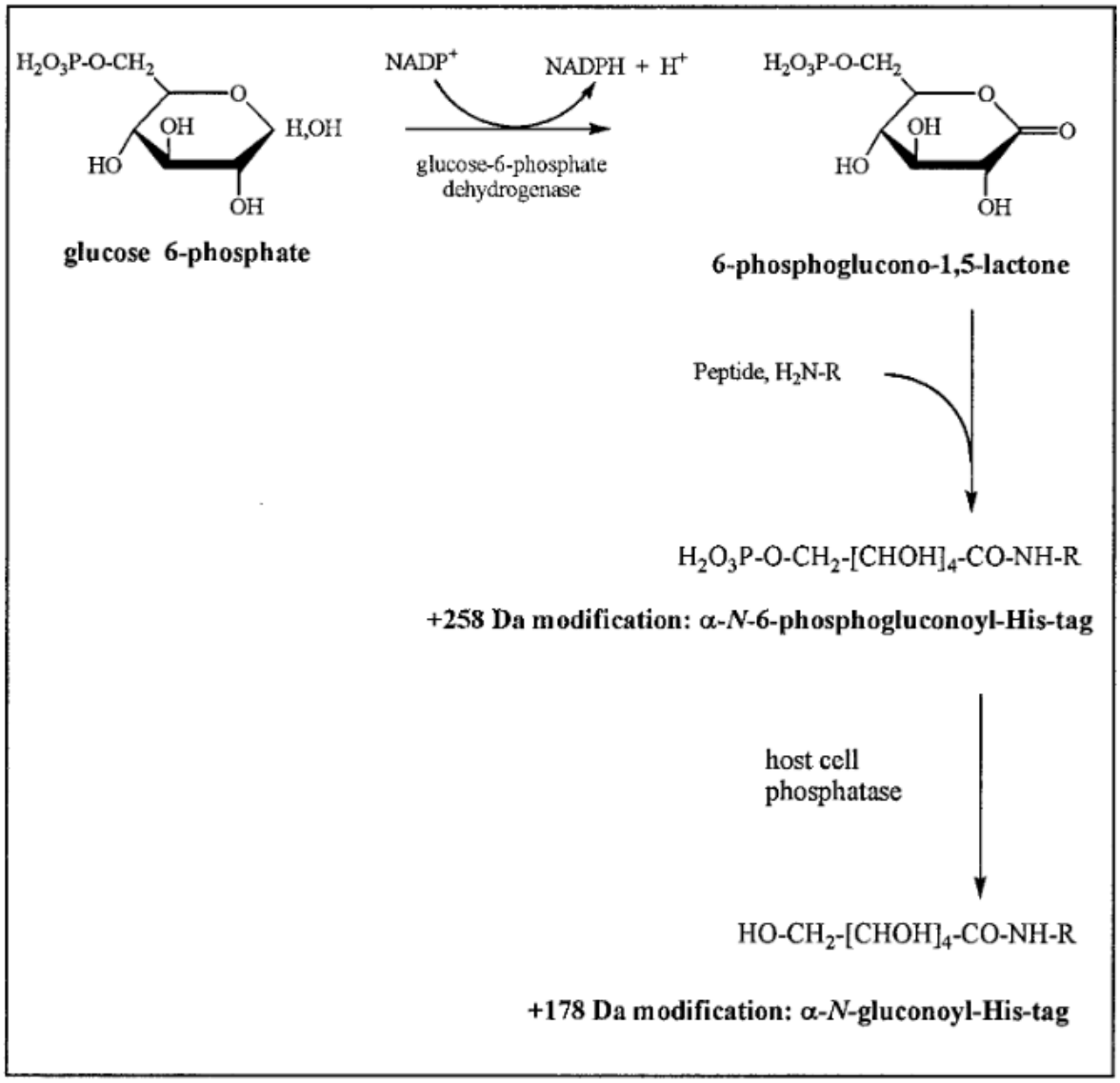


Figure 33. Spontaneous α-N-6-Phosphogluconoylation of 6xHis-tagged proteins. The addition of the gluconate group occurs by reaction of the exposed N-α amine with 6-phosphoglucono-1,5-lactone, which then has the phosphate group subsequently removed by a host-cell phosphatase. Figure taken from Geoghegan et al., 1999.[168]

Chapter 5: Hit identification

5.1 Thermal shift assay

To identify first-in-class ligands for NPAS1, NPAS3, and ARNT2 proteins, a TSA was applied as a high-throughput approach. TSA measures changes in the melting temperature (Tm) of a protein under the influence of a ligand (Figure 34).[169] Ligands that stabilise proteins upon binding cause an increase in the protein Tm, indicated by a positive shift of the curve. The denaturation of a protein is measured by the change in fluorescence of an environmentally sensitive dye. SYPRO Orange was chosen as it has a low quantum yield in aqueous solution and has an excitation and emission wavelength profile (the excitation maxima [Ex] are 300/472 nm and the emission maximum [Em] is 570 nm) compatible with the lab qPCR machine. When a protein starts to unfold, the dye binds non-specifically to the exposed hydrophobic residues of the denatured protein, resulting in a significant increase in fluorescence emission (Figure 34A). Individual melting curves are generated by plotting fluorescence intensity vs. temperature to obtain a sigmoidal curve with the midpoint corresponding to the protein's thermal unfolding transition (Figure 34B). To determine the Tm, the melting curve is further analysed by calculating its first derivative, expressed as dF/dT (change in fluorescence over change in temperature). This derivative curve highlights the point where the fluorescence changes most rapidly, producing a distinct peak that corresponds to the protein's melting temperature (Figure 34B). The derivative plot makes Tm estimation more precise than eyeballing the midpoint of the raw curve.

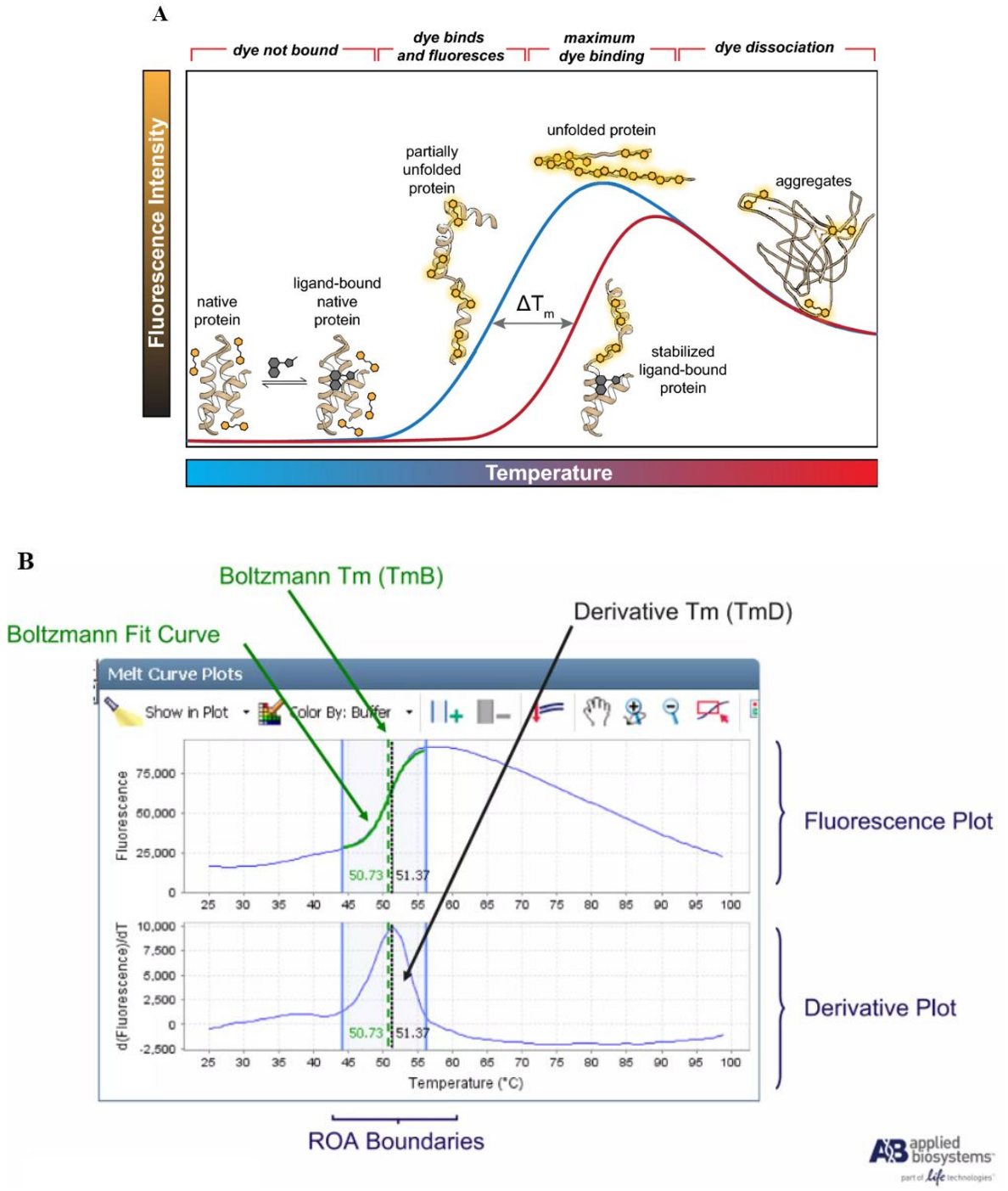


Figure 34. TSA overview. (A) The principle of TSA. Figure taken from Samuel et al., 2021.[170] (B) Applied Biosystems Protein Thermal Shift™ software calculates the Tm of the protein from the shape of the fluorescence plot (standard and derivative). Figure taken from Applied Biosystems, part of Life Technologies.

5.2 TSA: setup and validation

To enable testing large numbers of compounds, 384-well formatted TSA was employed. Several parameters were optimised for 384-well TSA with mARNT2-PAS-B and hNPAS1/3-ARNT2 heterodimers, including protein and SYPRO Orange concentrations, buffer pH, salt concentration, and DMSO tolerance.

To choose the optimal protein assay concentration, mARNT2-PAS-B and hNPAS1/3-ARNT2 heterodimers were titrated at several concentration points and were tested in TSA in the absence of any treatment. The raw data were visualised with GraphPad Prism (Figure 35). The Tm values were calculated by the Protein Thermal Shift™ software.

The Tm of hNPAS1-hARNT2 is 56 °C. Based on protein concentration, the Tm values for mARNT2-PAS-B and hNPAS3-ARNT2 are varied between 56-60 °C and 55-57 °C, respectively. It should be noted that the mARNT2-PAS-B background was relatively high (Figure 35A). Among all mARNT2-PAS-B concentrations, 0.4 mg/mL showed a slightly better assay window and was selected for TSA. Although 0.2 mg/mL protein concentration showed a higher fluorescent peak for both hNPAS1/3-ARNT2 heterodimers, 0.1 mg/mL was chosen for TSA due to the low yield of the protein expression (Figure 35B-C).

The SYPRO Orange test revealed that the 10x concentration has a sharper melting curve and a higher fluorescent peak than the 5x concentration for hNPAS1/3-ARNT2 heterodimers (Figure 36B-C). For mARNT2-PAS-B, higher concentrations of dye proportionally increase the protein curve background; thus, 5x SYPRO Orange was selected for the TSA (Figure 36A).

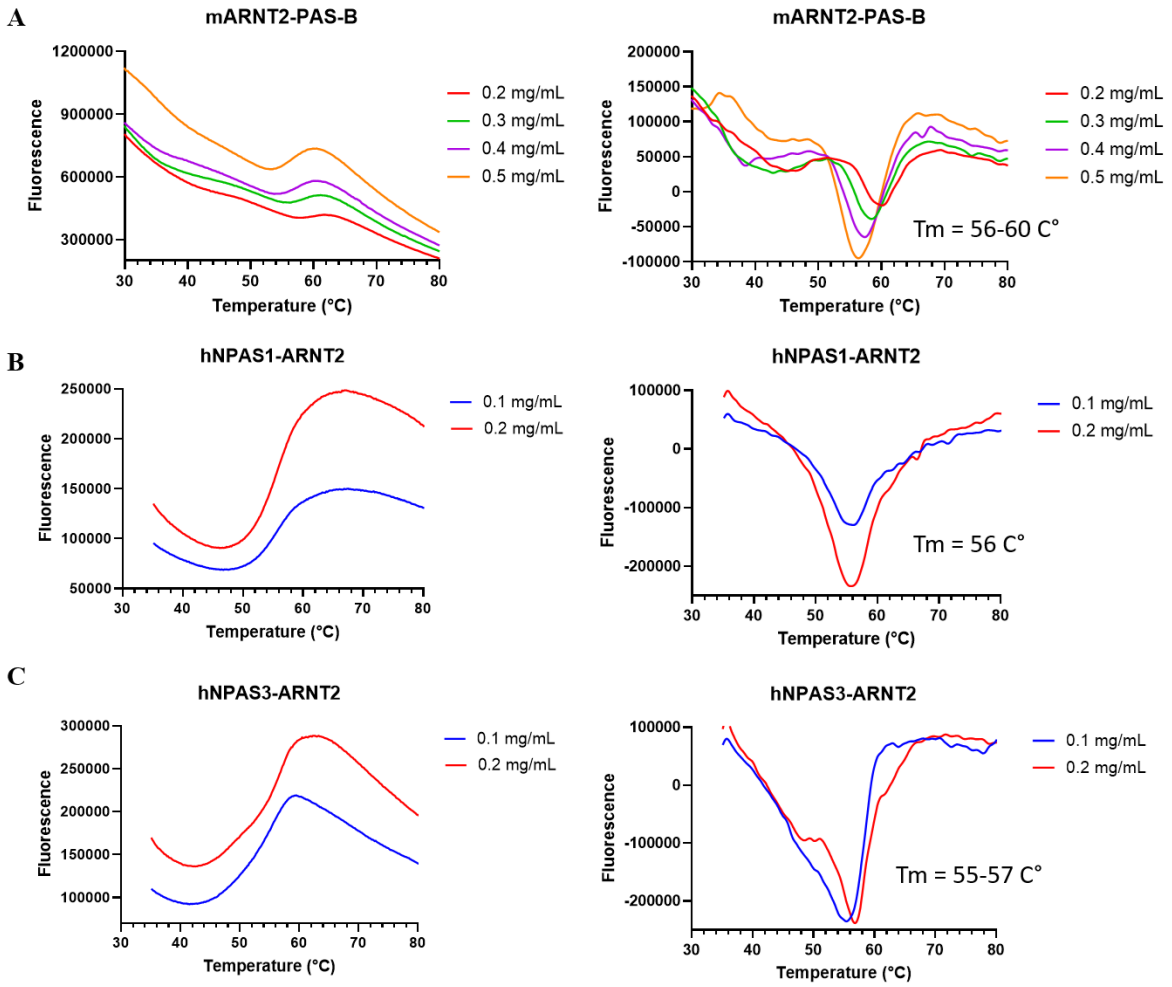


Figure 35. TSA analyses for protein concentration. (A) Melting curves and their derivatives for 0.2-0.5 mg/mL of purified mARNT2-PAS-B domain (B) Melting curves and their derivatives for 0.1-0.2 mg/mL of hNPAS1-ARNT2 heterodimer. (C) Melting curves and their derivatives for 0.1-0.2 mg/mL of hNPAS3-ARNT2 heterodimer.

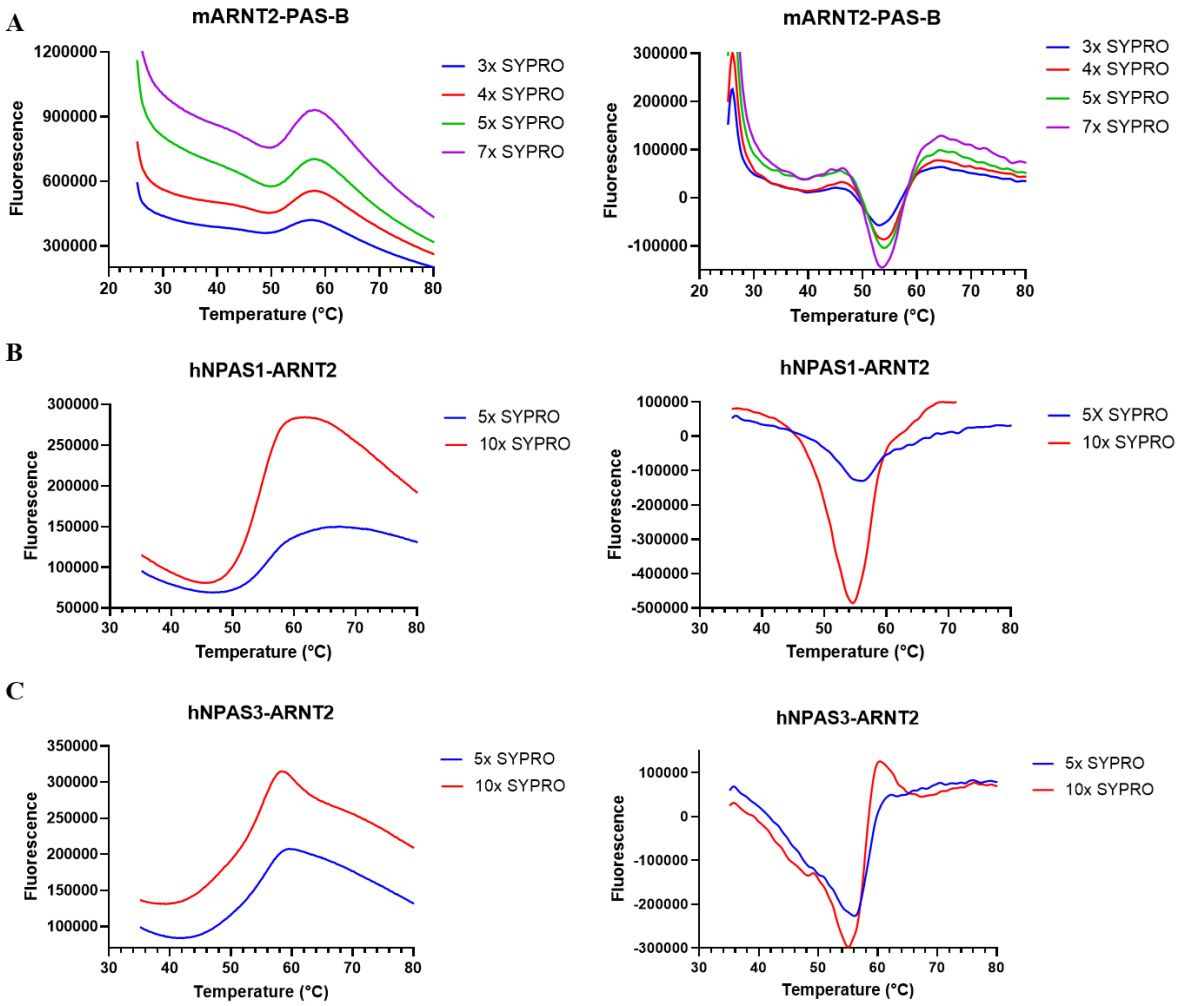


Figure 36. SYPRO Orange optimisation for TSA. (A) Melting curves and their derivatives for mARNT2-PAS-B with 3-10x SYPRO Orange concentrations. (B) Both 5x and 10x SYPRO Orange assessment for hNPAS1-ARNT2, and for (C) hNPAS3-ARNT2.

After identifying the favourable protein and SYPRO Orange concentrations, a buffer pH screen was performed to determine the optimal pH conditions that stabilise the protein and improve the assay window for ligand screening. After the buffer pH screen, intermediate stable conditions such as bis-tris or sodium phosphate pH 6.5 were selected for mARNT2-PAS-B (Figure 37A). Both buffers were then tested with different salt conditions. Sodium

phosphate pH 6.5 with 500 mM NaCl showed a higher fluorescent peak, and was hence chosen for HTS (Figure 37B).

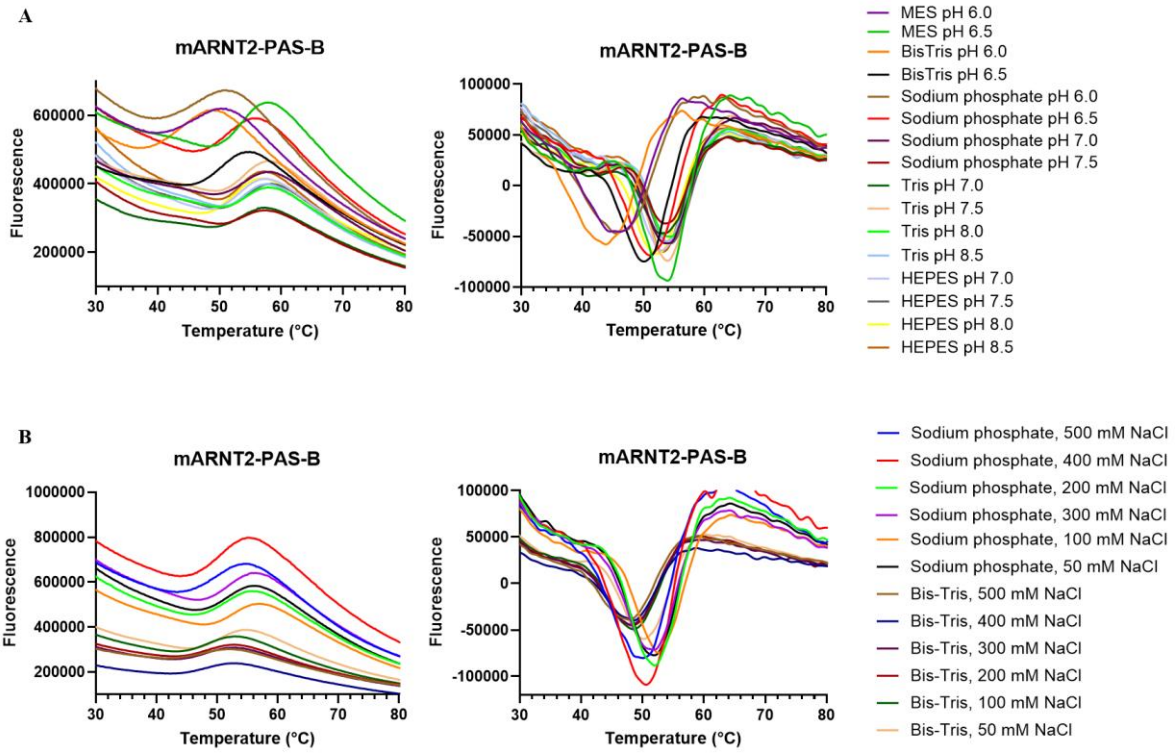


Figure 37. Buffer assessment for mARNT2-PAS-B. (A) The buffer pH screen consisting of different sets of MES, Bis-Tris, Sodium-Phosphate, HEPES, and Tris buffers, each at a concentration of 20 mM and with a pH range from 5.5 to 8.5. (B) Derivative and melting curves for salt screen, consisting of 20 mM Bis-Tris and Sodium-Phosphate pH 6.5, with varying NaCl from 50 mM to 500 mM.

Following buffer optimisation for both heterodimers, HEPES pH 7.5-8.0 showed significant stabilisation for NPAS1 and well-defined melting curves for NPAS3 heterodimers (Figure 38). The HEPES buffer has less temperature-dependent pH variability than the Tris buffer; hence it is better for TSA.[171] It should be noted that, as compared to a singular PAS-B domain of ARNT2, both heterodimers provide relatively small protein yields after purification. To

avoid the buffer exchange step and to potentially increase protein yield, the buffer composition for protein purification was changed from the previously reported Tris-HCl pH 8.0 to HEPES pH 7.5.[162] Medium salt concentration seems more favourable for both heterodimers; thus 150 mM NaCl was maintained from the original purification protocol (Figure 39).

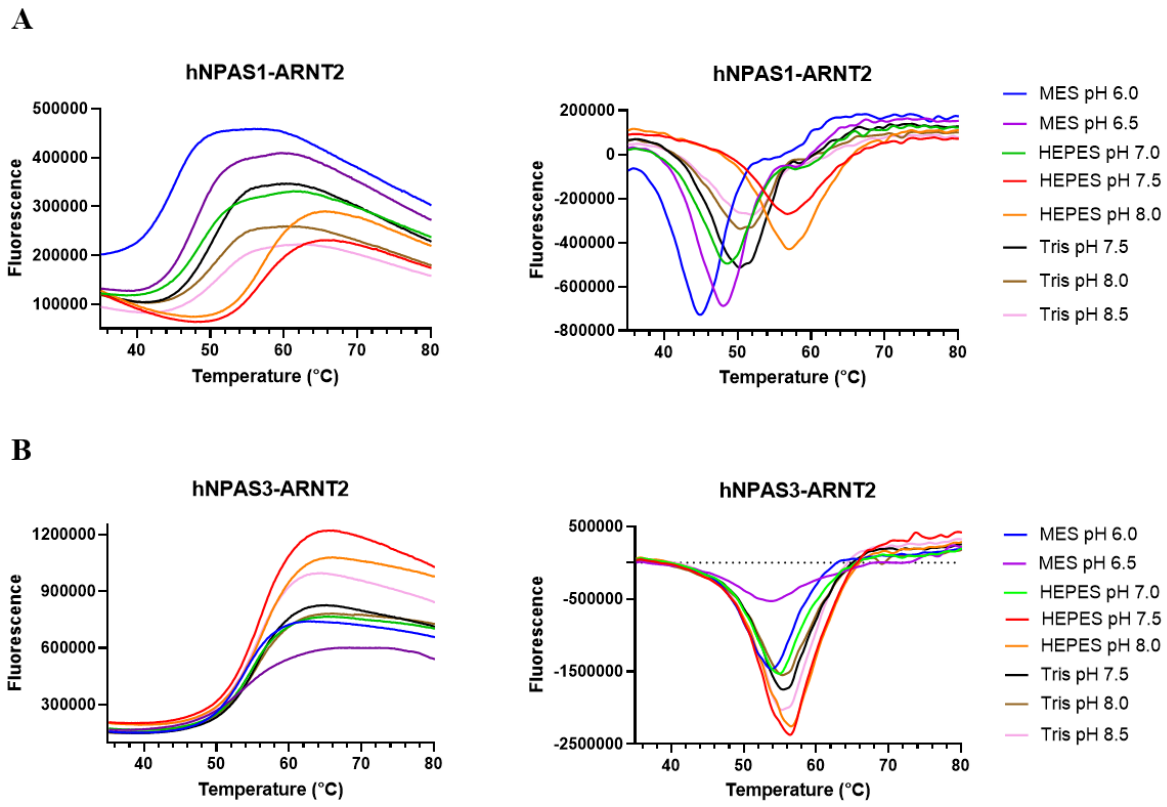


Figure 38. The buffer pH optimisation for heterodimers. The screen consists of different sets of 20 mM MES, HEPES, or Tris buffers with a pH range from 6.0 to 8.5. Melting curves and their derivatives are shown for different buffer pH for (A) hNPAS1-ARNT2, and for (B) hNPAS3-ARNT2.

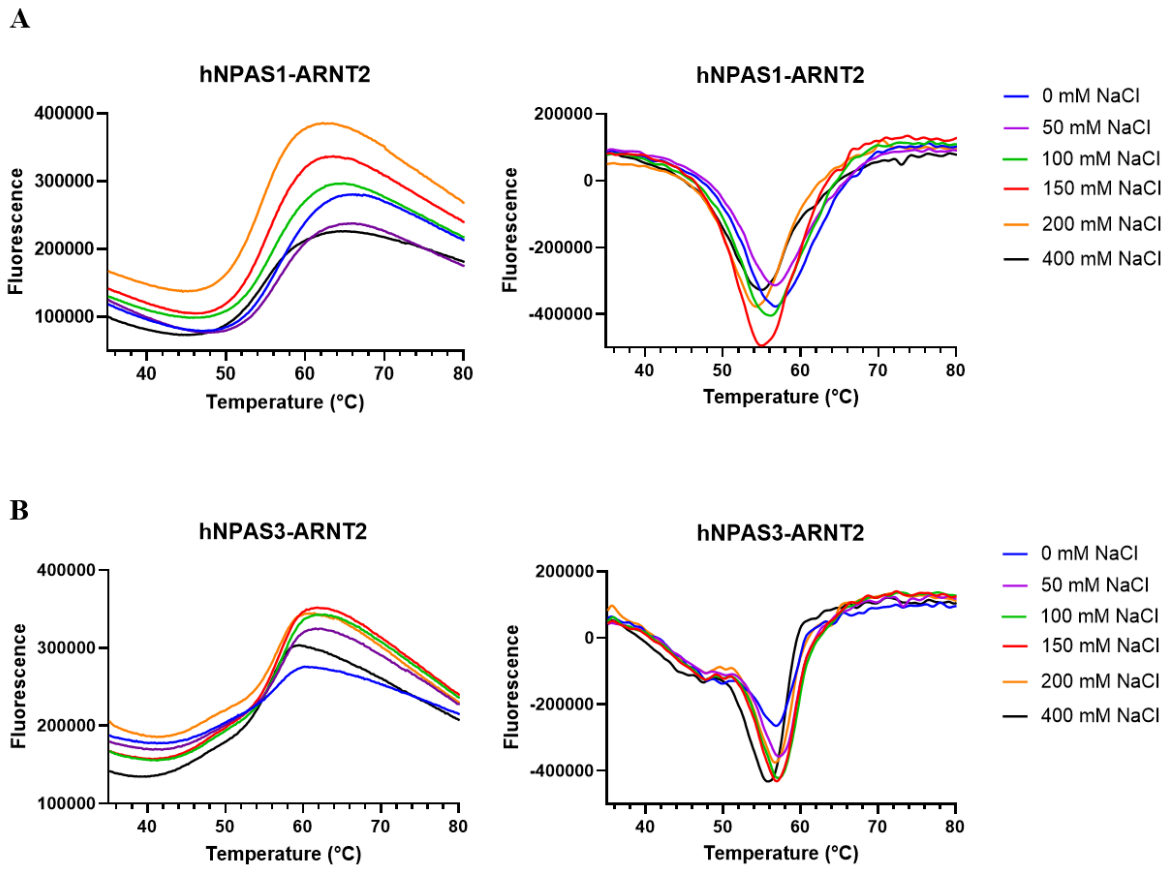


Figure 39. The buffer-salt optimisation for heterodimers. (A) Melting curves and their derivatives are shown for heterodimers in a salt screen in 20 mM HEPES pH 7.5, with NaCl ranging from 50 mM to 400 mM for hNPAS1-ARNT2, and for (B) hNPAS3-ARNT2.

Compounds for HTS are stored in dimethyl sulfoxide (DMSO) stocks. DMSO is a chaotropic agent, which can disrupt internal hydrogen bonding networks and hence reduce the stability of the protein.[172] To check which DMSO concentration is safe for the HTS, a DMSO screen was performed. The DMSO tolerance for mARNT2-PAS-B and hNPAS1/3-ARNT2 heterodimers is 2% (Figure 40). Low DMSO concentrations do not change the background of the melting curves, nor the Tm values of the proteins. For the HTS, 1% DMSO was therefore selected.

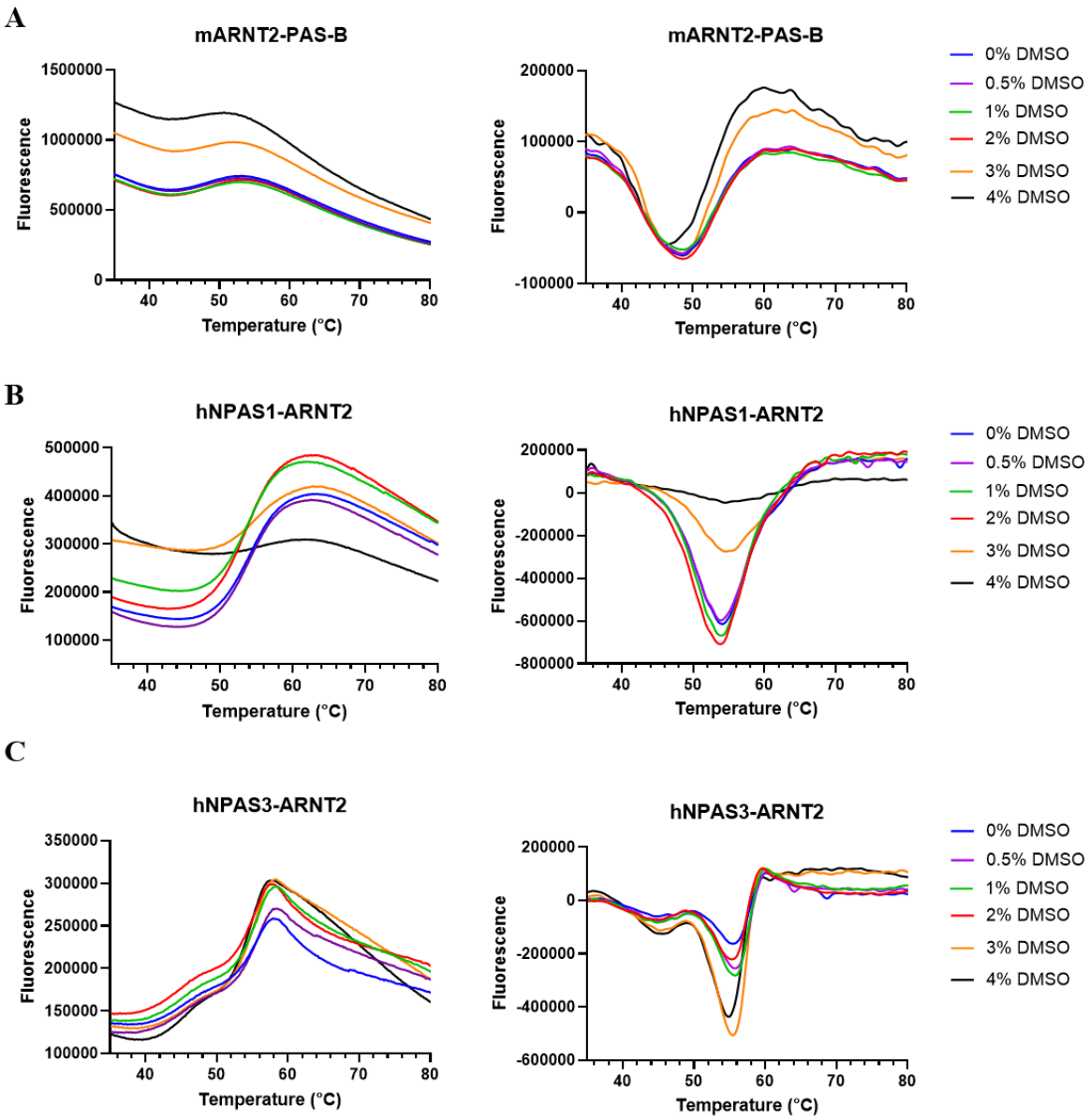


Figure 40. DMSO tolerance test. Final validated TSA conditions for each protein were tested in the presence of DMSO from 0% to 4%. Melting curves and their derivatives are shown for the DMSO screen for (A) mARNT2-PAS-B, (B) hNPAS1-ARNT2, and (C) hNPAS3-ARNT2.

5.3 High-throughput screening with mARNT2-PAS-B, hNPAS1-ARNT2, and hNPAS3-ARNT2

To find small molecules that bind NPAS1, NPAS3, and ARNT2, purified proteins were screened against commercially available libraries within the laboratory, totalling 67,847 compounds. A unique collection of libraries consists of core sets of chemical fragments arising from the smart fragmentation of approved drugs like the Prestwick library (Figure 41), and small molecules with MW ≤ 300 Da within the ChemDiv and Life Chem libraries. The CNS-targeted Neuro Enamine fragment library consists of compounds with molecular properties and structural features preferable for blood-brain barrier (BBB) permeability, which is extremely useful for this project given its relevance to neuronal regulation.

An exclusive set of 7,490 bioactive compounds within MedChemExpress and MedChemExpress Natural libraries consists of natural products, enzyme inhibitors, receptor ligands, and drugs for HTS and high content screening, with some having been approved by the FDA. High structural diversity ensures little overlap between the libraries. These libraries are widely used in research focusing on areas such as CNS, cancer, stem cells, neuronal signalling, immunity, and more.

Figure 41. Prestwick Drug-Fragment Library. An original collection of fragments derived from approved drugs. Each drug was manually "cut" into relevant fragments and retained

according to the ability to bind to a target, suitable for chemical expansion, and synthetic accessibility. Picture taken from Prestwick Chemical.

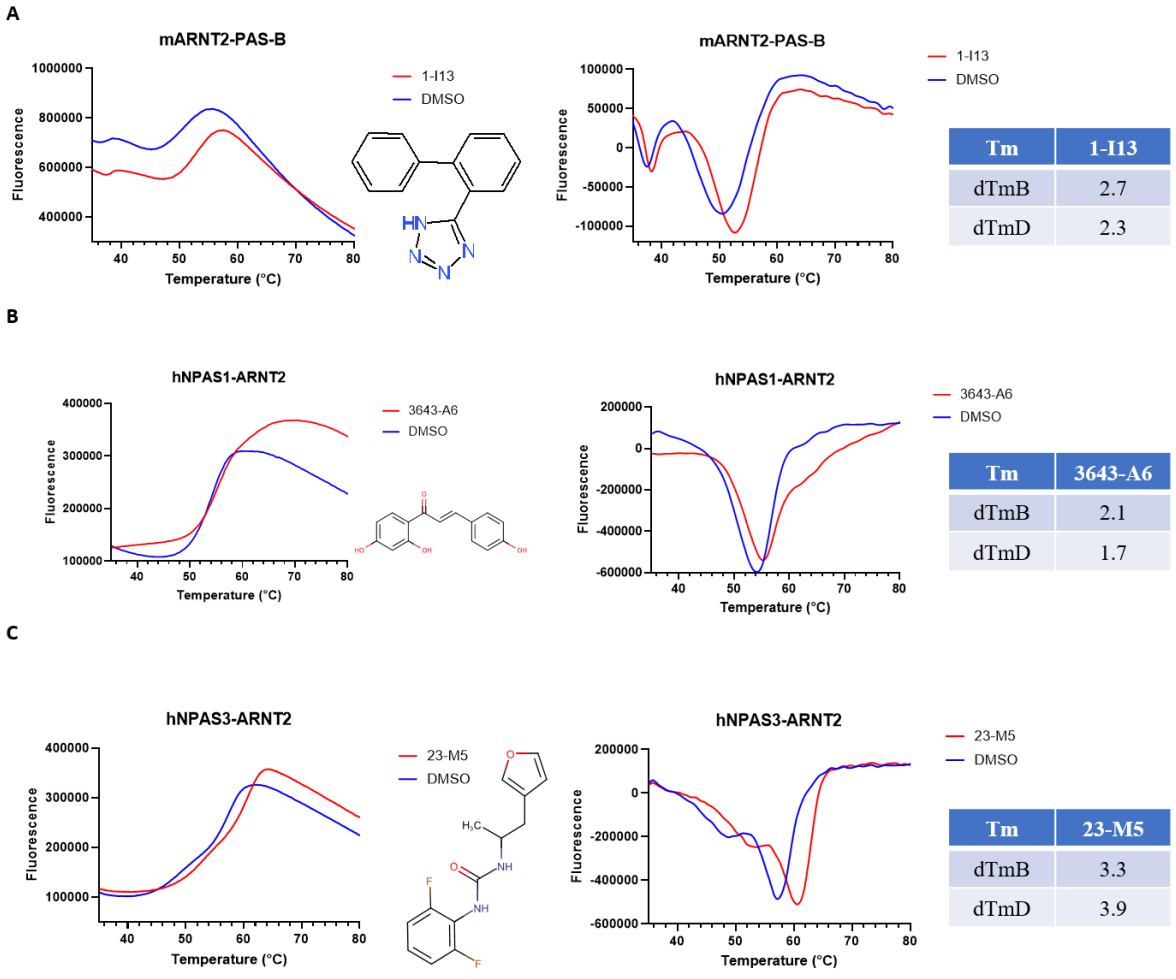
For the primary screen, large quantities of the mARNT2-PAS-B domain, hNPAS1-ARNT2, and hNPAS3-ARNT2 heterodimers were overexpressed and purified. Each purified protein batch that was used for HTS has a quality control assessment based on gel-filtration and SDS-PAGE profiles. To prepare assay-ready plates, all compounds were dispensed into 384-well plates for the final 1% DMSO assay concentration. Based on library stock concentration, compounds were therefore tested at 100, 200, or 500 μM.

A total of 67,847 compounds were screened against the mARNT2-PAS-B domain, hNPAS1-ARNT2, and hNPAS3-ARNT2 heterodimers at a single dose. Melting curves for each compound were analysed, and the Tm values were calculated by the Protein Thermal Shift™ software. The delta TmB (dTmB) values were calculated against reference curves with DMSO. Based on three times the standard deviation, a hit criterion was selected as dTmB ≥ 1.

After the primary screen, 751 compounds passed the selected dTmB criterion for hNPAS1 and 376 compounds for hNPAS3 heterodimers (Figure 42B, C). There are no known hit compounds reported for NPAS1 or NPAS3. However, the discovery of 3643-A6 as a primary hit for hNPAS1-ARNT2, which is similar to the recently identified ARNT2 target butein, only lacking the hydroxy group, confirms the efficiency of the screen (Figure 43B).

A total of 1,482 primary hits were identified for mARNT2-PAS-B (Figure 42A). A representative 1-I13 primary hit containing tetrazole and phenyl groups showed structural

similarities to the KG-548 control compound, indicating a successful screen (Figure 43A). It is important to note that, across the three primary screens, 47 compounds were repeatedly identified as hits in each screen. However, all 47 were subsequently eliminated during selectivity assessment and PAINS filtering, indicating that they were likely false positives arising from nonspecific or assay-interfering behaviour rather than genuine, selective target engagement. A summary of the primary hits from the HTS mARNT2-PAS-B and both of the hNPAS1 and hNPAS3 heterodimers is provided in Table 3.



Tm	1-I13
dTmB	2.7
dTmD	2.3

Tm	3643-A6
dTmB	2.1
dTmD	1.7

Tm	23-M5
dTmB	3.3
dTmD	3.9

Figure 42. The primary HTS screen via TSA. Representative examples of primary hits for (A) the mARNT2-PAS-B domain, (B) hNPAS1-ARNT2, and (C) hNPAS3-ARNT2 heterodimers. Normalised and derivative TSA traces of protein with a compound are shown

in red, and with DMSO control in blue. The compounds showed an increased dTmB and passed the selection hit criterion (dTmB ≥ 1).

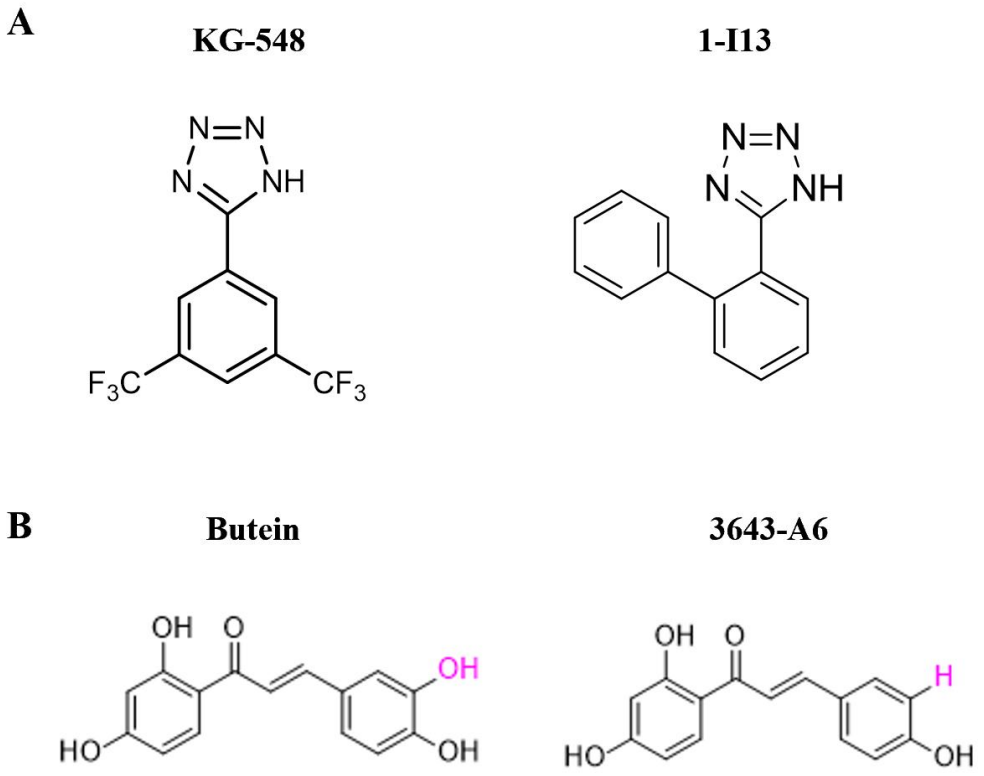


Figure 43. Structural comparison between primary hits and the known control, KG-548, and recently identified butein.(A) KG-548 control compound and 1-I13 primary hit for mARNT2-PAS-B. The hit molecule has tetrazole and phenyl groups. (B) Butein and 3643-A6 primary hit for hNPAS1-ARNT2. The hit molecule has one less hydroxyl group than butein.

Generally, the global hit rate in HTS (defined as the number of hits divided by the total measured) is usually 1%.[173] The 2.18% overall hit rate with ARNT2-PAS-B was mostly due to the overreactive covalent compound set within the ChemDiv2-3D diversity library (Table 3). Covalent inhibitors contain warheads, which are reactive functional groups that form a covalent bond with the target protein, usually at a cysteine residue. Inside the mARNT2-

PAS-B domain, there are three cysteines. Both cysteines 433 and 374 position their side chains internally, whereas cysteine 354 is surface exposed and solvent accessible (Figure 44). Thus, the ARNT2-PAS-B structure may accommodate covalent primary hits as legitimate or nonspecific binders.

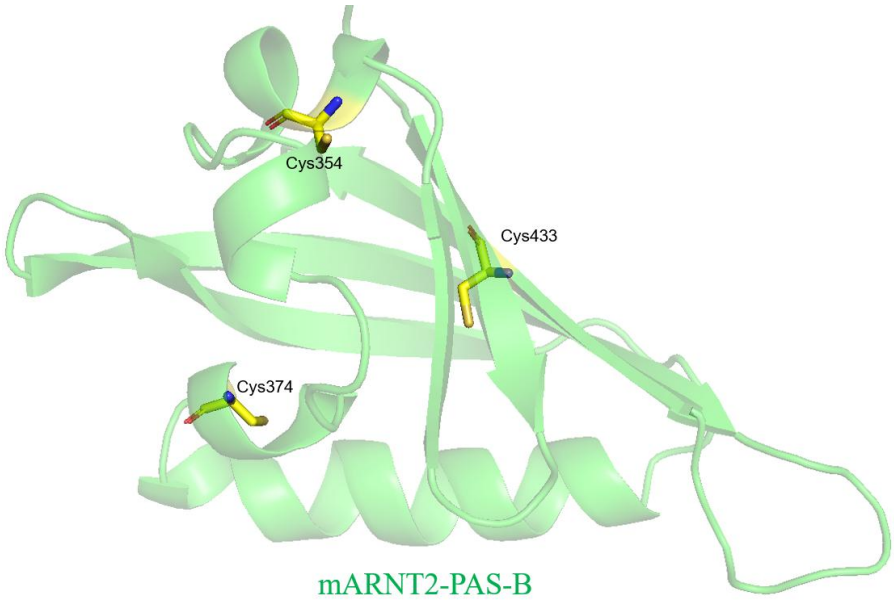


Figure 44. Crystal structure of the PAS-B domain of mARNT2 with the positions of three cysteine residues highlighted.The cysteine residues are labelled and shown as stick representations to indicate their spatial arrangement within the domain. Cysteine, with its highly reactive nucleophilic thiol (-SH) group, is a common target for covalent drug design due to its ability to form stable covalent bonds with electrophilic (electron-seeking molecules) covalent inhibitors.

The formation of a covalent bond between a warhead and cysteine can significantly alter the protein's structure and function, leading to its inactivation. Therefore, compounds that exhibit a covalent mechanism of action are often avoided in drug discovery programmes. On the other hand, covalent compounds offer significant potential as drug targets exhibiting

pharmacological advantages, such as prolonged duration of action and enhanced potency over non-covalent drugs. By carefully balancing the benefits and risks, covalent inhibitors can be a powerful tool in the fight against disease. Therefore, covalent ARNT2 hits were not eliminated after the primary screen, and all 1,482 compounds progressed to the next step of dose-response hit validation by TSA.

5.4 Dose-response follow-up for hit validation

Since the primary screen was conducted without a replication, a follow up TSA that determined the dose-response curves (DRC) was performed to reconfirm the primary hits. To validate primary hits, 1,482 promising compounds for mARNT2-PAS-B, 751 compounds for hNPAS1-ARNT2, and 377 compounds for hNPAS3-ARNT2 were tested in six-point concentrations in two independent runs. Results were analysed with GraphPad Prism. DRC studies identified 359 compounds for mARNT2-PAS-B, 135 compounds for hNPAS1-ARNT2, and 119 compounds for hNPAS3-ARNT2 that showed reproducible, dose-dependent increases in protein stabilisation (Figure 45, Table 3).

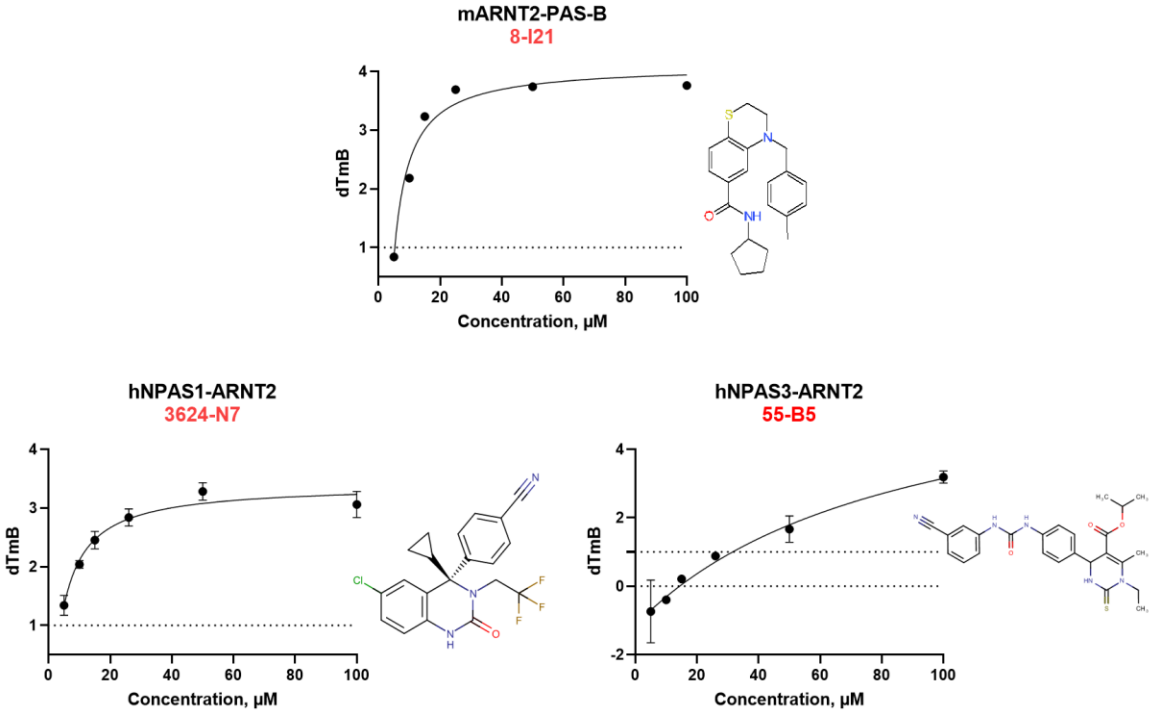


Figure 45. Representative examples of a follow-up DRC with one of the primary hits for mARNT2-PAS-B, hNPAS1-ARNT2, and hNPAS3-ARNT2 proteins. Data were analysed using TSA software, and dTmB values were plotted and fitted using GraphPad Prism.

Library	Total compounds	Final concentration (µM)	Primary hits			DRC hits		
			mARNT2-PAS-B	hNPAS1-ARNT2	hNPAS3-ARNT2	mARNT2-PAS-B	hNPAS1-ARNT2	hNPAS3-ARNT2
Prestwick	480	500	3	2	5	2	2	2
Electrophile	2,000	200	3	0	0	0	0	0
Enamine fragment	1,280	100	4	0	1	2	0	1
ChemDiv	10,000	500	94	23	30	14	14	9
Life Chem	16,150	500	203	108	47	63	18	23
MedChem Express (MCE)	6,486	100	149	40	6	69	19	5
MCE-Natural products	1,004	100	33	8	1	13	8	1
ChemDiv-2 (3D diversity)	30,447	100	1,004	570	287	196	74	78
Total	**67,847**		**1,482**	**751**	**377**	**359**	**135**	**119**

Table 3. HTS summary from TSA. Chemical libraries and the number of compounds screened against mARNT2-PAS-B, hNPAS1-ARNT2, and hNPAS3-ARNT2 in primary screens and in follow up DRC assays.

5.5 Lead compound selection from the primary hits

To find similar compounds between libraries, all primary hits that passed DRC test were examined for structural similarities and combined into clusters using DataWarrior software. Some clusters were found for each protein target (Figure 46A). It was expected, however, that many compounds discovered in the primary screen might be non-selective or false positives. To eliminate unselective compounds, DRC-confirmed hits were analysed against screening data previously examined in the laboratory from the same chemical libraries tested on other proteins from the bHLH-PAS family, including ARNT, HIF-1α-ARNT, HIF-2α-ARNT, HIF-3α-ARNT, SIM1-PAS-A, NCOA1, NCOA2, and NCOA3.

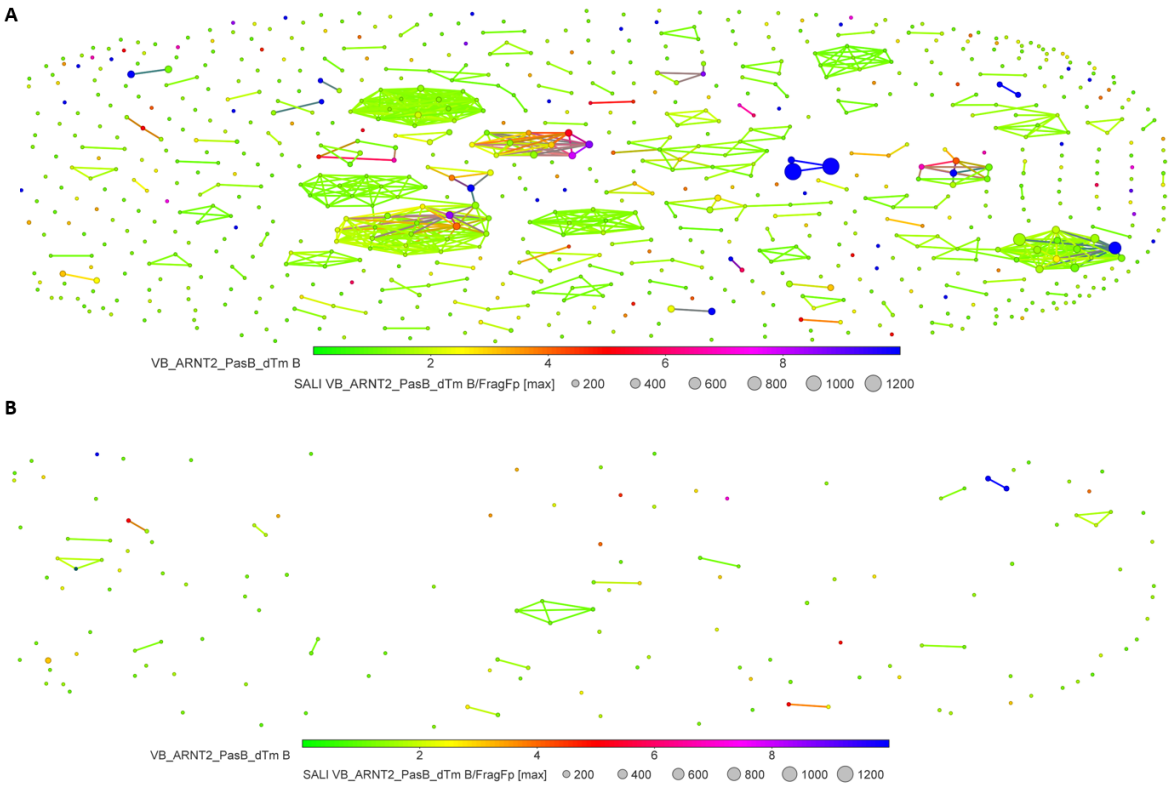


Figure 46. Representative examples of DataWarrior structural similarity and clusters. (A) Clustering output for DRC-confirmed hits targeting mARNT2-PAS-B. (B) Elimination of structural artefacts and unselective compounds following similarity and clustering analysis.

To remove artefacts, primary hits were further analysed for nasty functions, which are structural features or functionalities in molecules that are known to be problematic in drug discovery, and for PAINS, which are promiscuous molecules with multiple behaviours that masquerade as drug-like binding and which therefore yield false signals across a variety of assays.[174] A total of 89 compounds for ARNT2-PAS-B, 31 compounds for NPAS1-ARNT2, and 26 compounds for NPAS3-ARNT2 passed the selectivity and artefacts removal evaluation tests.

After removal of unselective and non-specific compounds, very few structurally similar clusters still remained for each target (Figure 46B). Such a significant reduction suggests that many of the initial hits were promiscuous or assay-interfering compounds. A substantial proportion of the excluded clusters may have included substructures associated with frequent hitters, such as reactive functional groups or scaffolds prone to non-specific binding. Although not all eliminated clusters are necessarily chemically reactive, their consistent lack of selectivity indicates limited target engagement specificity. Therefore, it is essential to implement structural filtering to prioritise high-quality compounds for follow-up studies.

To prioritise the best early candidates, these selected primary hits were subsequently evaluated based on several criteria, including potency of the dTmB curve, correlation between dTmB and dTmD, maximum dTmB value, cluster effects of several matching compounds, and chemical tractability for further synthesis of potential analogues. Lead compounds for mARNT2-PAS-B were also assessed based on the presence of key functional groups known to interact with the protein, as identified in previous crystal structures with

KG-548 and butein, specifically tetrazole or triazole moieties (KG-548-like) and α/β-keto phenyl groups (butein-like).

After evaluation of the most promising candidates, 15 best compounds were selected for each protein (Figures 47-49). To keep the diversity of final sets, one or a maximum of two representatives of each cluster were selected. Lead compounds were also chosen from several chemical libraries. The summary of lead compound selection from the primary screen for each protein is visualised in Figure 50.

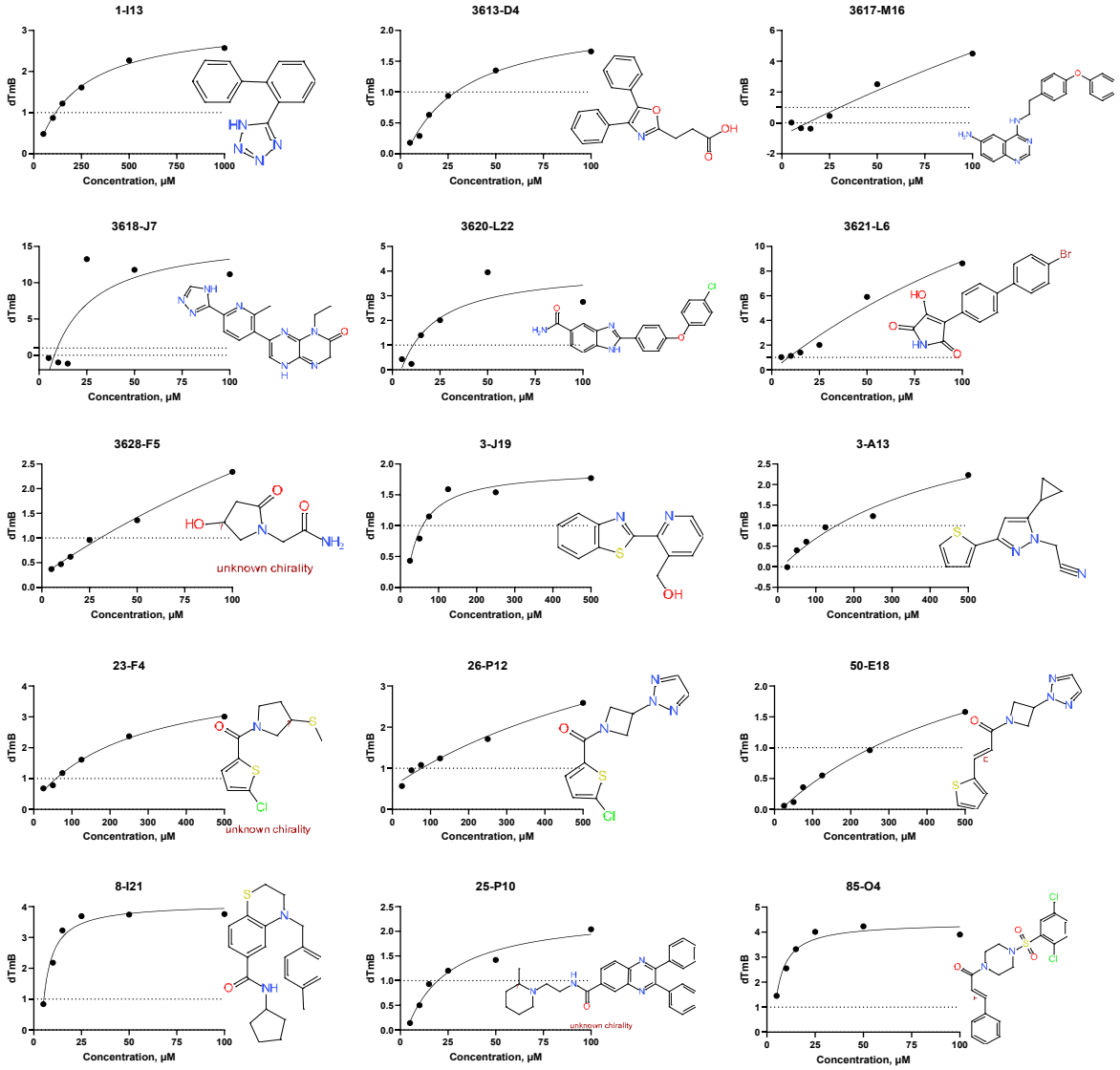


Figure 47. DRC for fifteen final primary hits for mARNT2-PAS-B. These hits were advanced for subsequent studies.

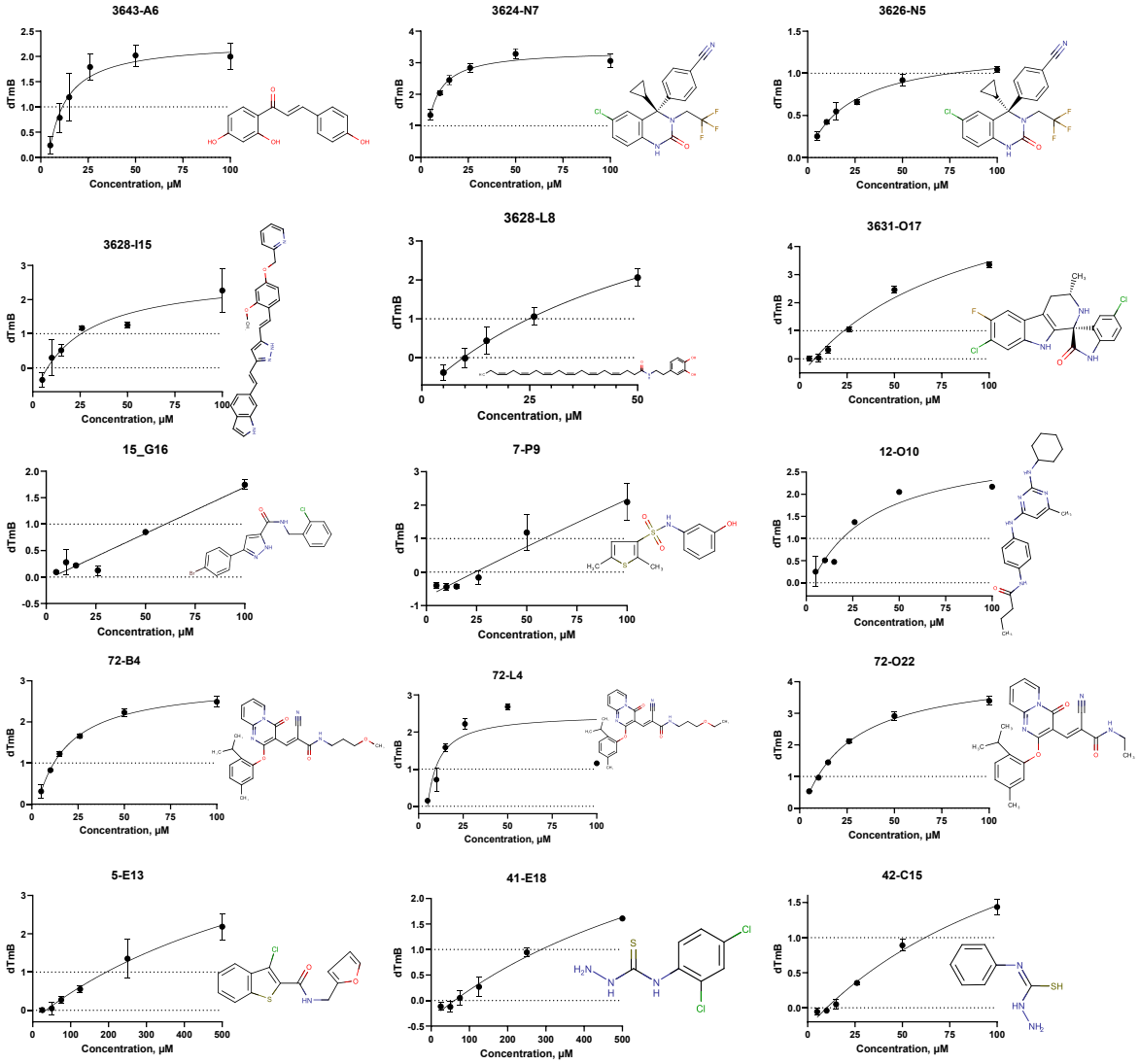


Figure 48. DRC for fifteen final primary hits for hNPAS1-ARNT2 heterodimer. These hits were advanced for secondary studies.

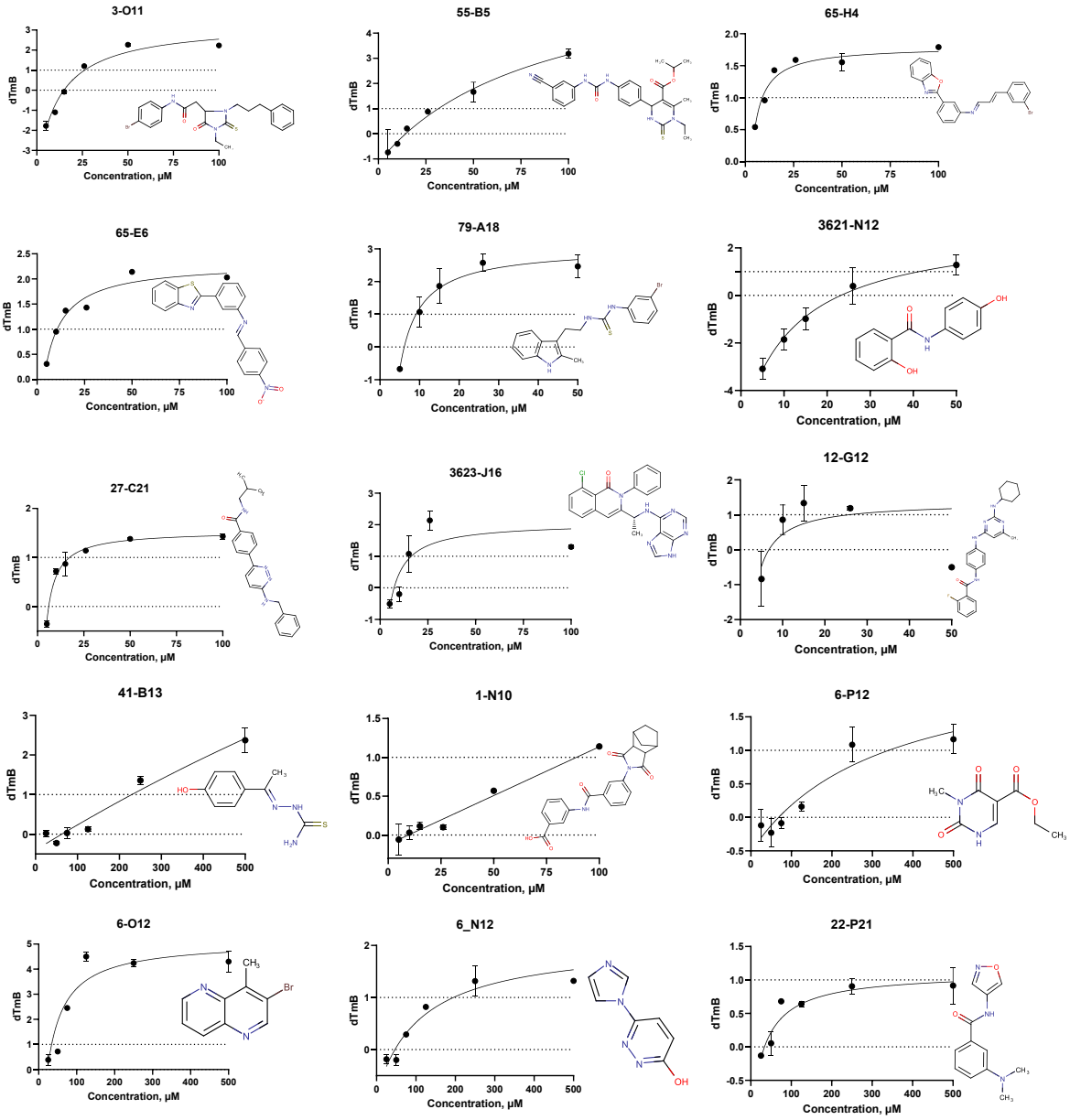


Figure 49. DRC for fifteen final primary hits for hNPAS3-ARNT2 heterodimer. These hits were advanced for follow-up studies.

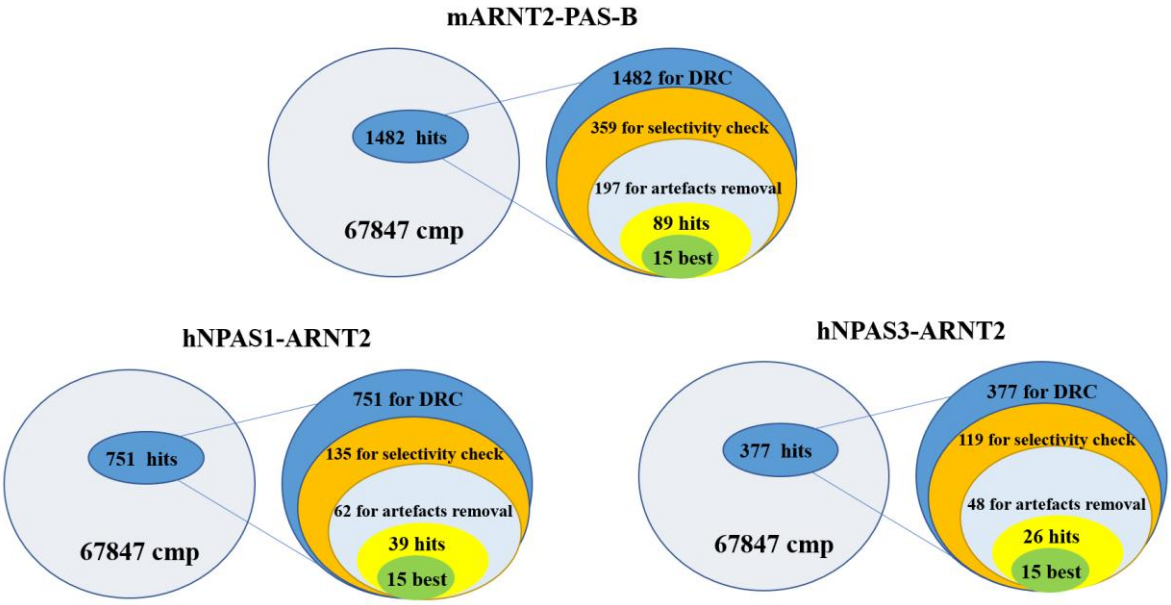
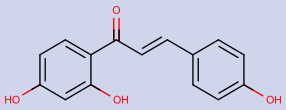
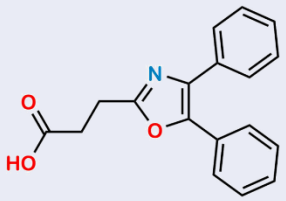
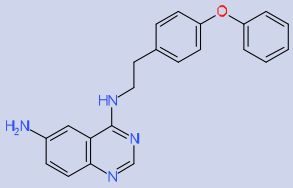
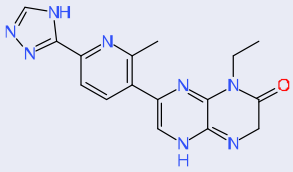
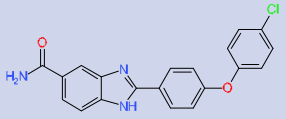
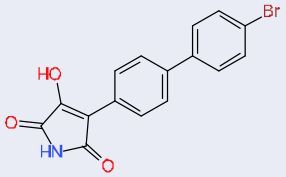


Figure 50. Schematic representation of lead compound selection. The diagrams represent the step-by-step selection of the final hits, starting from the primary screen for mARNT2-PAS-B, hNPAS1-ARNT2, and hNPAS3-ARNT2.

For the final selected ligands, the limited information currently available pertains primarily to compounds from the MedChemExpress libraries, which comprise known bioactive molecules and established drug targets (Table 4). The most interesting compounds are those with neurological activity or which relate to the targeting of CNS disorders. For example, 3617-M16 or QNZ, also known as EVP4593, is a quinazoline derivative that acts as a potent inhibitor of the NF-κB signalling pathway. It is known for its anti-inflammatory and neuroprotective properties. EVP4593 showed the ability to reduce store-operated calcium (SOC) entry, and protect neurons in human neuroblastoma SK-N-SH cells in a model for Huntington's disease.[175,176]

Compound	Structure	Information	Pathway
3643-A6 (Isoliquiriti-genin)		Isoliquiritigenin is a chalcone found in the roots of several Glycyrrhiza species which inhibits aldose reductase with an IC_{50} of 320 nM.[177]	Autophagy; protease
3613-D4 (Oxaprozin)		Oxaprozin is an inhibitor of COX-1 and COX-2 with IC_{50} of 2.2 μM and 36 μM for human platelet COX-1 and IL-1-stimulated human synovial cell COX-2, respectively.[178] Oxaprozin also inhibits the activation of NF-κB.[179]	Immunology; inflammation; NF-κB
3617-M16 (QNZ EVP4593)		QNZ (EVP4593) shows strong inhibitory effects on NF-κB transcriptional activation and TNF-α production with IC_{50} values of 11 and 7 nM, respectively.[180,181] QNZ is a neuroprotective inhibitor of SOC channel.[175]	Apoptosis; NF-κB
3618-J7 (CC-115)		CC-115 is a potent and dual DNA-PK and mTOR kinase inhibitor with IC_{50} values of 13 nM and 21 nM, respectively. CC-115 blocks both mTORC1 and mTORC2 signalling.[182–184]	Cell cycle; DNA damage; PI3K/mTOR
3620-L22 (BML-277)		BML-277 is a selective checkpoint kinase-2 inhibitor with an IC_{50} of 15 nM.[185]	Cell cycle; DNA damage
3621-L6 (Glycolic acid oxidase inhibitor-1)		Glycolic acid oxidase inhibitor-1 is a glycolate oxidase inhibitor.[186]	Other

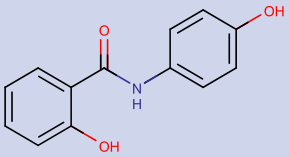
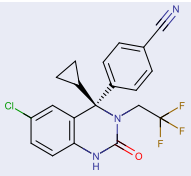
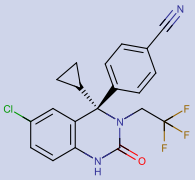
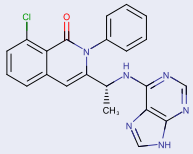
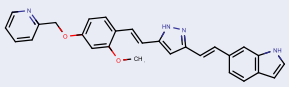
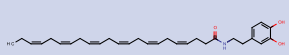
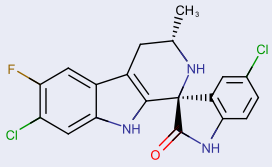
3621-N12 (Osalmid)		Osalmid is a ribonucleotide reductase small subunit M2 (RRM2) targeting compound; suppresses ribonucleotide reductase activity with an IC_{50} of 8.23 μM.[187,188]	Anti-infection
3624-N7 (TTA-Q6)		TTA-Q6 is a selective T-type Ca^{2+} channel antagonist, used in studies related to neurological disorders.[189,190]	Ion channel
3626-N5 (TTA-Q6 isomer)		TTA-Q6 (isomer) is an isomer of TTA-Q6.	Other
3626-J16 (PGD2-IN-1)		PGD2-IN-1 is a DP antagonist, which selectively targets the prostaglandin D2 receptor-2 (DP2).[191]	Other
3628-F5 (Oxiracetam)		Oxiracetam is a cyclic derivative of γ-aminobutyric acid (GABA), commonly used as a nootropic drug to treat cognitive impairments.[192–195]	Ion channel; neuronal signalling
3628-I15 (PE859)		PE859 is a potent inhibitor of both tau and Aβ aggregation with IC_{50} values of 0.66 and 1.2 μM, respectively.[196]	Cell cycle; DNA damage; cytoskeleton
3628-L8 (NMI 8739)		NMI 8739 is a dopamine D2 autoreceptor agonist. [197]	Neuronal signalling
3631-O17 (Cipargamin)		Cipargamin is a potent antimalarial compound, with IC_{50} of ~1 nM.[198]	Anti-infection

Table 4. Lead compounds from the MedChemExpress and MedChemExpress Natural libraries.

A second noteworthy compound is 3628-L8 or NMI 8739, also known as N-docosahexaenoyl dopamine. NMI 8739 binds to and activates the dopamine D2 auto-receptor, which is a receptor on dopamine-producing neurons. NMI 8739 is capable of penetrating the BBB and exhibits the potential of using fatty acid carriers to deliver small molecules, like dopamine, to the brain.[197] Therefore, NMI 8739 is often used in studies related to neurological diseases.

A butein similar 3643-A6 or isoliquiritigenin (ILG) is a naturally occurring chalcone found in liquorice root and other plants, known for its diverse pharmacological properties, including cancer treatment, inflammation reduction, and promotion of neurological health.[177] Several studies have shown the promising *in vivo* effects of ILG as a neuroprotective compound through inhibition of intracellular reactive oxygen species (ROS) generation, its antioxidative action, and its attenuation of synaptic dysfunction and neuronal damage, confirming its usefulness in neurodegenerative diseases.[199–202]

Another compelling ligand is 3628-I15 or PE859, which is a potent inhibitor of both tau and amyloid-beta (Aβ) aggregation, a key process involved in neurodegenerative diseases like Alzheimer's. It has been shown to significantly reduce aggregated tau and prevent the onset and progression of neural dysfunction *in vivo*, suggesting PE859's effectiveness for the treatment of tauopathies.[196,203]

Yet an intriguing small molecule, 3628-F5 or Oxiracetam, is a nootropic drug and a mild stimulant commonly used to treat cognitive impairments.[192] Oxiracetam has been shown to improve short and long-term memory, boost concentration and focus, improve cognition and increase cognitive energy.[193–195]

Other interesting chemicals are 3624-N7 (TTA-Q6) and its isomer 3626-N5. TTA-Q6 is a selective T-type Ca^{2+} channel antagonist that displayed *in vivo* CNS efficacy in epilepsy and tremor models.[189] TTA-Q6 is also a cognition-enhancing drug that belongs to the class of psychostimulants. It is an agonist at dopamine D2 receptors and has been shown to have a significant effect on locomotor activity and exhibits beneficial effects on working memory, as seen in rat studies.[190] The compounds 3624-N7 and 3626-N5 are enantiomers, which are molecules that are non-superimposable mirror images of each other. Usually, enantiomers act very differently, especially in biological and chiral environments. At this stage, both compounds seem to bind the NPAS1 protein. Since it remains unclear which isomer is the active form, both hits have been advanced to the next stage of validation.

In conclusion, while selectivity of the compounds was previously evaluated using screening data of other bHLH-PAS family proteins, known targets for some of the identified lead compounds have previously been discovered. Although this might introduce some off-target effects in potential future cellular or animal studies, these compounds remain largely uncharacterised. Their mechanisms of action are not fully elucidated, leaving open the possibility that they may still engage with NPAS1/3 or ARNT2 proteins. Accordingly, a total of 45 selected lead compounds were repurchased from fresh powder and advanced for secondary studies.

5.6 Confirmation of primary hits and counter screen

Both HTS and DRC studies used pre-dissolved compounds from the library plates. Despite careful handling of the chemical libraries, some compounds may have undergone partial

drying, spilled between wells causing inter-well cross-contamination, or precipitated over time. Therefore, it was necessary to reconfirm the hits using freshly purchased and prepared high-purity powders through DRC analyses. A total of 45 compounds were freshly dissolved in DMSO and titrated across 16 concentrations using a 2-fold serial dilution. DRC studies using TSA confirmed 13 compounds for mARNT2-PAS-B, and 11 compounds for each hNPAS1-ARNT2 and hNPAS3-ARNT2, as reproducible, dose-dependent leads (Figures 51-53). The EC_{50} values for each compound were calculated using GraphPad Prism (Table 5).

In TSA, the EC_{50} represents the ligand concentration that induces half-maximal thermal stabilisation of the protein, reflecting as half of the observed dTm shift. The EC_{50}, however, does not directly correspond to binding affinity as compared to the equilibrium-based binding assays. Protein concentration is one of the key factors limiting the accuracy of EC_{50} in TSA. When the protein concentration significantly exceeds the ligand concentration, the calculated EC_{50} may be markedly higher than the true K_D, due to ligand depletion and the non-equilibrium conditions under which TSA is performed. For both heterodimers, the tested protein concentration was 1 μM and the ligand concentration range was 500 μM - 0.015 μM. Thus, EC_{50} values should be reliable for comparative purposes. In contrast, for the ARNT2-PAS-B domain, a concentration of 25 μM is in excess of the same ligand concentration range, and ligand depletion effects may occur, especially in the lower concentration range of the titration series. As a result, the obtained EC_{50} values for the PAS-B domain might be significantly higher than the actual K_D.

Other confounding factors include slow-binding ligands that could artificially lower the EC_{50}. In addition, poor ligand solubility or chemical instability might lead to underestimation of effective concentrations, and as a result, increase EC_{50} values. Another consideration is

quenching or fluorescence interference: the interactions between some test compounds and the fluorescent dye, which can distort the thermal unfolding profile and influence the EC_{50} values. Therefore, EC_{50} calculated via TSA should be interpreted with caution, as it typically does not reflect true binding affinity, unless equilibrium conditions are met and the magnitude of the Tm shift correlates linearly with ligand binding.

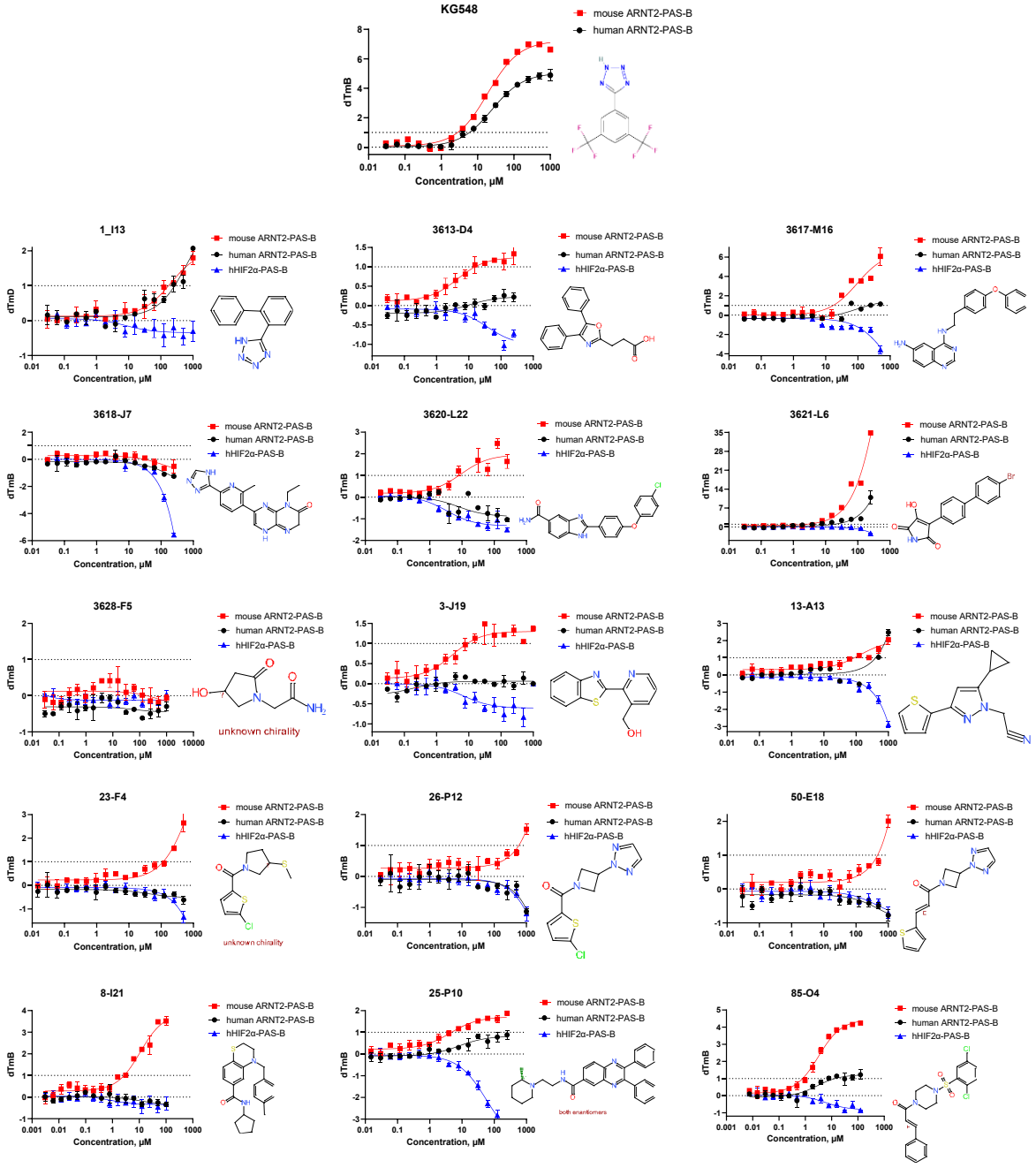


Figure 51. Fresh powder DRC-TSA profiles for fifteen primary hits, with KG-548 control, tested against mouse and human ARNT2-PAS-B in three independent runs. Thirteen of the 15 hits were reconfirmed for mouse ARNT2-PAS-B, and ten showed similar binding to the human protein. Counter-screening with hHIF-2α-PAS-B confirmed selectivity for ARNT2-PAS-B.

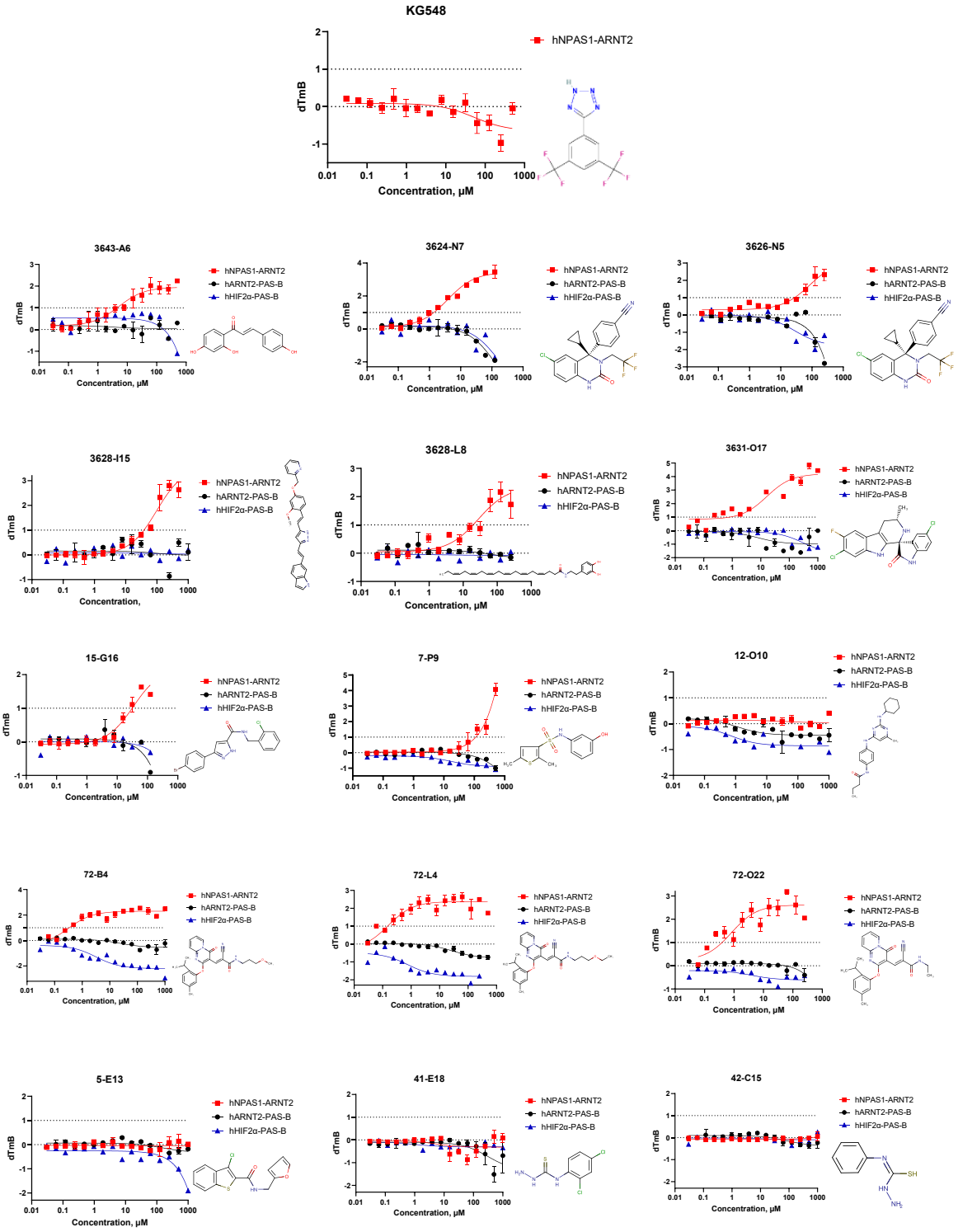


Figure 52. Fresh powder DRC-TSA profiles for fifteen primary hits, with KG-548 control, tested against hNPAS1-ARNT2 in three independent runs. Eleven of the 15 hits were reconfirmed from fresh powder. Counter-screening with hHIF-2α-PAS-B and hARNT2-PAS-B confirmed selectivity for hNPAS1.

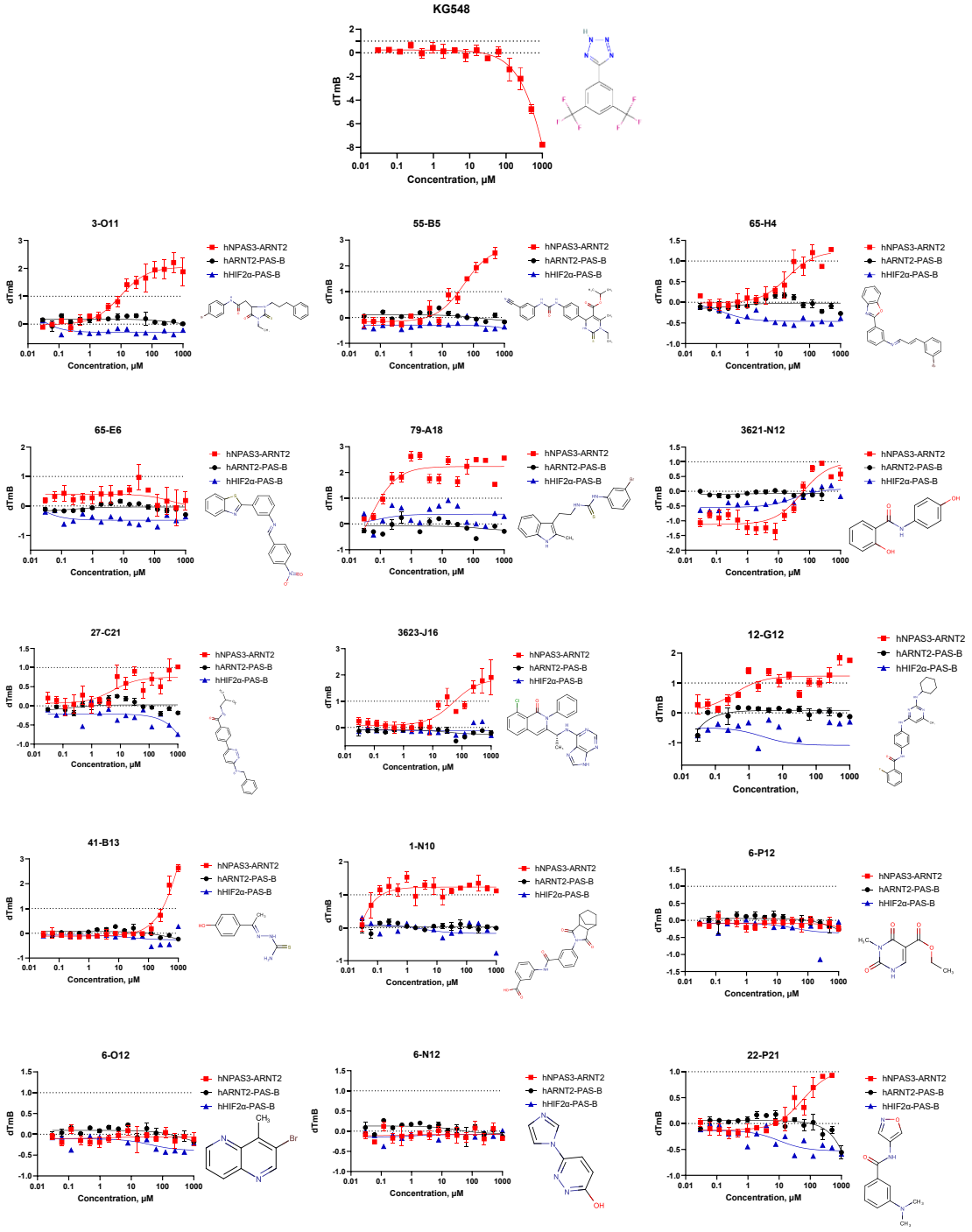


Figure 53. Fresh powder DRC-TSA profiles for fifteen primary hits, with KG-548 control, tested against hNPAS3-ARNT2 in three independent runs.Eleven of the 15 hits were reconfirmed from fresh powder. Counter-screening with hHIF-2α-PAS-B and hARNT2-PAS-B confirmed selectivity for hNPAS3.

EC_{50} (µM)			EC_{50} (µM)		EC_{50} (µM)	
Compound	mARNT2-PAS-B	hARNT2-PAS-B	Compound	hNPAS1-ARNT2	Compound	hNPAS3-ARNT2
KG-548	19	24	KG-548	n/a	KG-548	n/a
1-I13	203	622	3643-A6	5	3-O11	9
3613-D4	4	10	3624-N7	4	55-B5	50
3617-M16	58	90	3626-N5	85.1	65-H4	15
3618-J7	n/a	n/a	3628-I15	98	65-E6	n/a
3620-L22	9	n/a	3628-L8	27.7	79-A18	0.04
3624-L6	148	10	3631-O17	16	3621-N12	52
3628-F5	n/a	n/a	15-G16	34.3	27-C21	n/a
3-J19	3	n/a	7-P9	unstable	3623-J16	52
13-A13	133	unstable	12-O10	n/a	12-G12	0.3
23-F4	825	n/a	72-B4	0.3	41-B13	unstable
26-P12	4	n/a	72-L4	0.1	1-N10	0.1
50-E18	510	n/a	72-O22	0.8	6-P12	unstable
8-I21	11	n/a	5-E13	n/a	6-O12	n/a
25-P10	4	7.9	41-E18	n/a	6-N12	n/a
85-O4	3	3	42-C15	n/a	22-P21	n/a

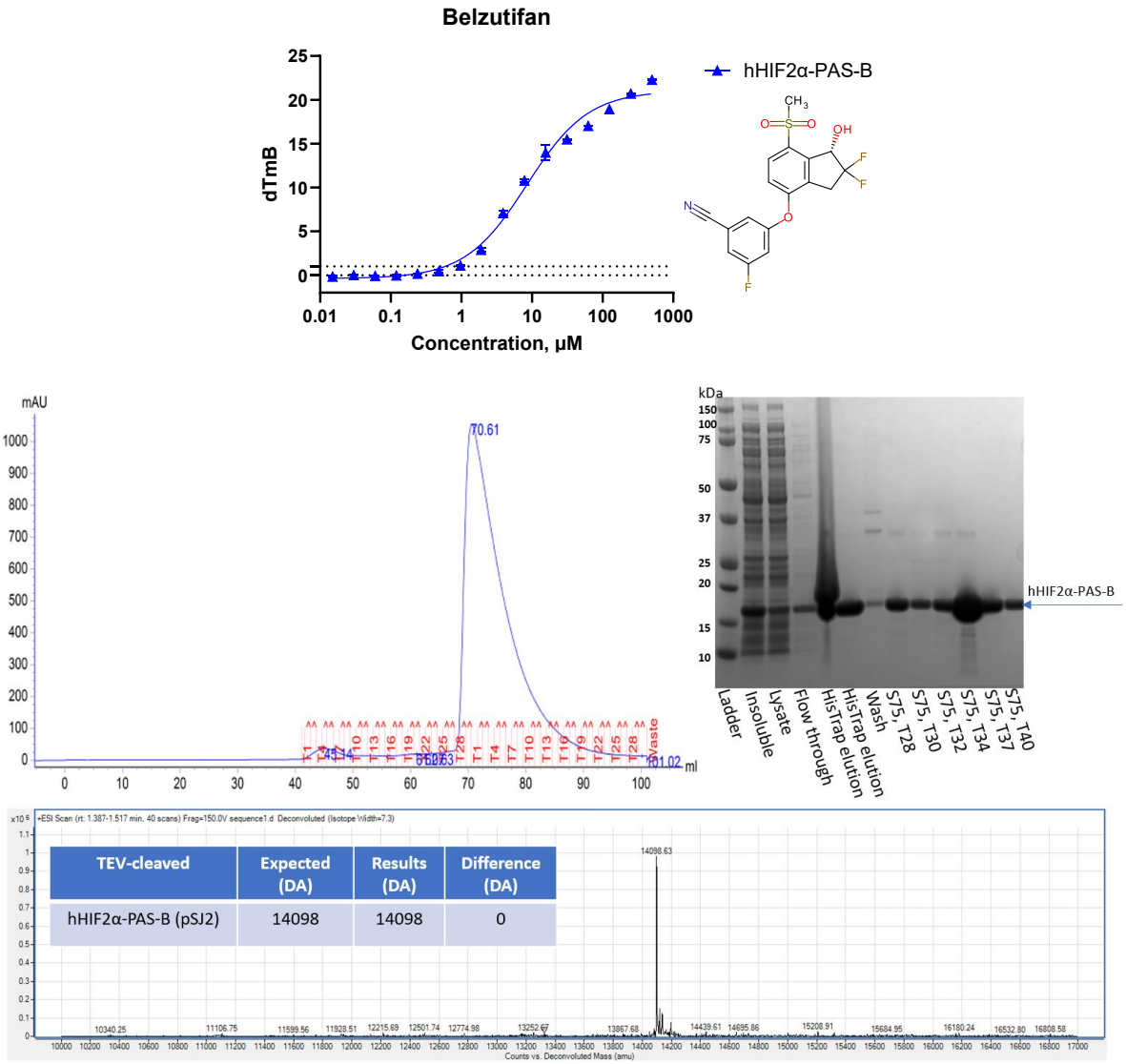
Table 5. Calculated EC_{50} values for all tested compounds. EC_{50} values were calculated using GraphPadPrism [Inhibitor] vs. response (three parameters). The not applicable (n/a) values indicate cases where the software could not fit a reliable curve or where the fit was too poor to extract an EC_{50} value. For the unstable values, the software did fit a curve, but the EC_{50} is statistically unreliable.

Previously, primary hits were already cross checked against screening data from other bHLH-PAS proteins in the lab. As only one compound concentration was used in the HTS, to reconfirm the selectivity of primary hits, compounds were investigated through DRC-TSA counter-screen studies with hHIF-2α-PAS-B, where belzutifan served as a control (Figure 54). The PAS-B domain of hHIF-2α was expressed and purified as previously described.[82] None of the 45 ordered compounds for three protein targets showed a positive TSA response

with hHIF-2α-PAS-B, establishing selectivity of the identified primary hits among the bHLH-PAS family (Figure 51-53).

To check whether hits for mouse ARNT2-PAS-B can also bind human protein, compounds were re-tested with hARNT2-PAS-B via DRC-TSA. With human protein, 10 out of 15 compounds showed a similar TSA pattern to the mouse protein (Figure 51). The remaining five compounds, including 3620-L22, 23-F4, 26-P12, 50-E18, and 8-I21, were selective to mouse ARNT2-PAS-B. These data support the hypothesis that a subtle alteration in the ligand pocket (human and mouse I431V substitution) influences hit discovery. This further suggests that, despite high sequence identity between species, minor variations in protein sequences influence ligand binding.

Given that heterodimer constructs consist of hNPAS1/3 and their dimerisation partner hARNT2, some identified compounds could be hits for hARNT2 protein. To establish further selectivity, hits for both heterodimers were also screened against hARNT2-PAS-B. None of the 30 hits for NPAS1/3 heterodimers showed a positive TSA response with hARNT2-PAS-B, indicating the likelihood of selectivity for NPAS1 or NPAS3 proteins (Figure 52-53). The putative ligand binding pocket in the PAS-B domain is the most probable binding site among other domains; however, these compounds may still interact with other regions (bHLH or PAS-A) of NPAS1, NPAS3, or ARNT2 proteins.



TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hHIF2α-PAS-B (pSJ2)	14098	14098	0

Figure 54. Belzutifan control for the TSA studies with hHIF-2α PAS-B. Purification and quality control assessment of hHIF-2α PAS-B based on size-exclusion chromatography and mass spectrometry, confirming sample homogeneity and expected molecular weight.

KG-548 was used as reference compound in each TSA studies, including with mouse and human ARNT2-PAS-B, as well as with the hNPAS1/3-ARNT2 heterodimers. As expected, KG-548 showed potent TSA curves for both mouse and human ARNT2-PAS-B domains with EC_{50} 19 and 24 μM, respectively (Figure 51, Table 5). It was also anticipated that KG-

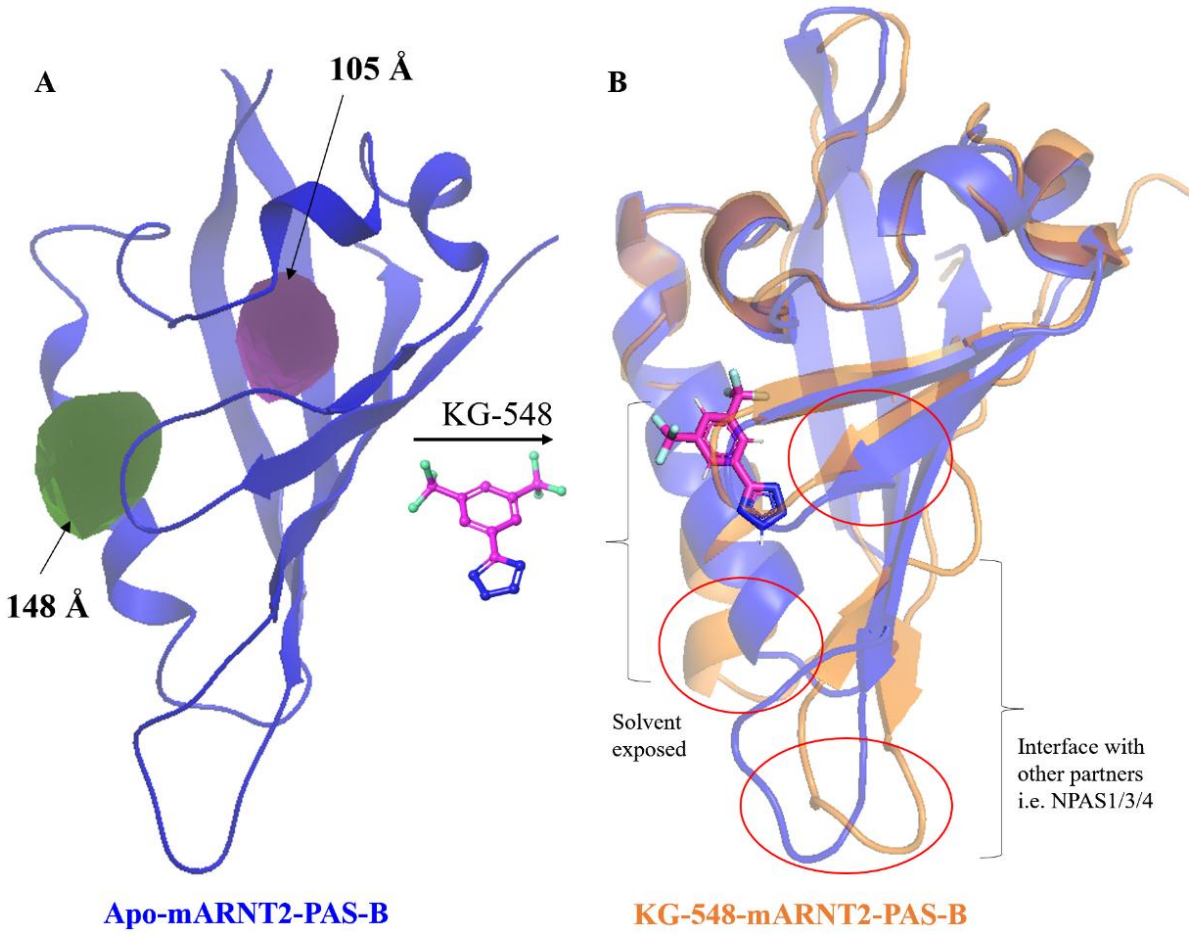
548 would bind to both NPAS1/3-ARNT2 heterodimers in a similar manner. However, KG-548 did not show binding to either of the heterodimers, suggesting the single PAS-B domain has a distinct conformation from that in the heterodimeric complex, which enables its selectivity for these compounds (Figure 52-53).

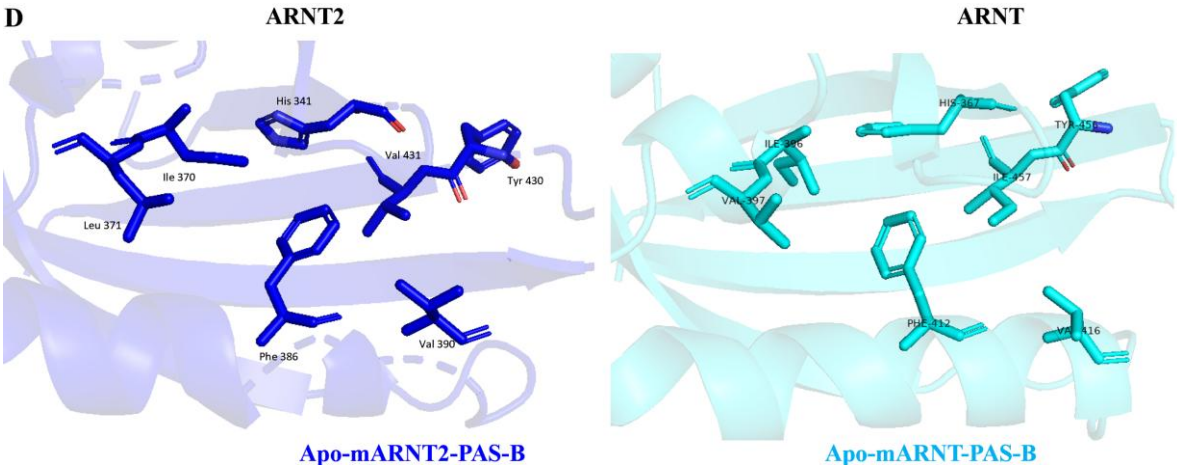
Chapter 6: Structural analyses and crystallography

6.1. Ligand binding and induced conformational changes in the mARNT2-PAS-B domain upon interaction with KG-548

To visualise how KG-548 binds to the PAS-B domain of ARNT2, the unpublished crystal structures of mARNT2-PAS-B and its complex with KG-548, solved previously by Dr. Hua Pu in the lab, were analysed. Binding pocket analysis in the ICM-pro program identified two adjacent cavities with measurements 148 $Å^3$ and 105 $Å^3$ in the PAS-B domain (Figure 55A). Since cavity size measurements depend on the algorithms used by different programs, the small difference between the larger cavity measured at 148 $Å^3$ with ICM-Pro and the previously reported 160 $Å^3$ obtained using PyVOL is considered consistent.[94]

The calculated cavity volume of 148 $Å^3$ indicates that the ligand-binding pocket of ARNT2 is among the smallest within the bHLH-PAS family. The control compound, KG-548, is relatively small, with a molecular weight of 282.15 Da. The crystal structure of the KG-548-bound ARNT2 PAS-B domain revealed that this low-molecular-weight ligand occupies the 148 $Å^3$ cavity within the PAS-B domain and induces some conformational changes (Figure 55B). Interestingly, the crystal structure also indicates a binding mode distinct from previously reported ARNT structures, in which KG-548 associates with the β-sheet surface of the PAS-B domain rather than the pocket cavity (Figure 55C).[159–161] This altered binding mode may result from differences in the amino acid composition of the ligand-binding pocket between ARNT2 and ARNT.


A
105 Å
148 Å
KG-548
B
Solvent exposed
Interface with other partners i.e. NPAS1/3/4
Apo-mARNT2-PAS-B
KG-548-mARNT2-PAS-B
C
KG-548-mARNT2-PAS-B
KG-548-mARNT-PAS-B



ARNT2	ARNT
His 341	His 367
Ile 370	Ile 396
Leu 371	Val 397
Phe 386	Phe 412
Val 390	Val 416
Tyr 430	Tyr 456
Val 431	Ile 457

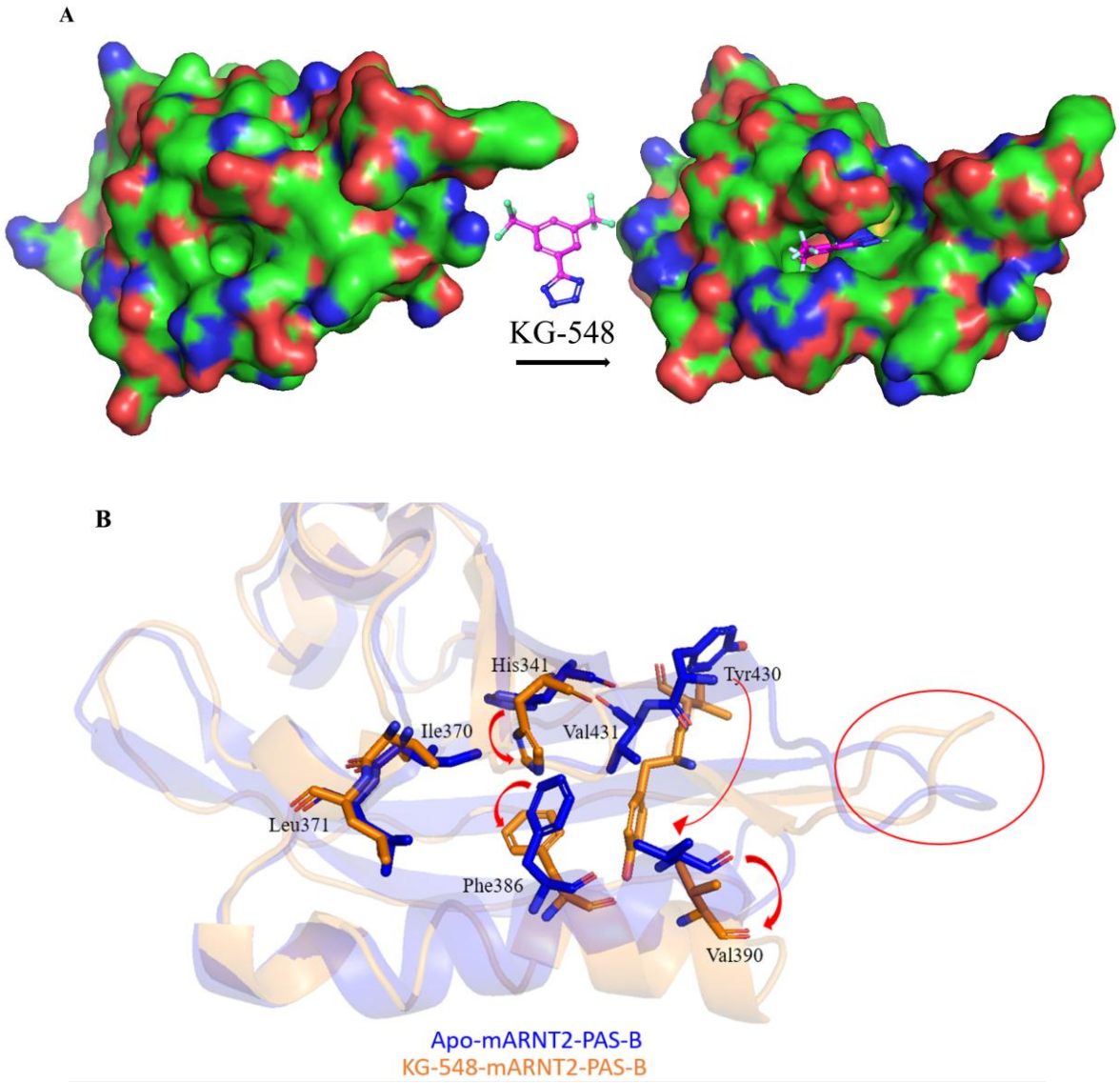
Figure 55. The crystal structures of mARNT2-PAS-B in the absence of a compound and as bound to KG-548. (A) The cavity analysis by ICM-Pro for the apo structure of mARNT2-PAS-B (blue) found two adjacent pockets of 148 $Å^3$ (green) and 105 $Å^3$ (purple). (B) A structural comparison of KG-548-bound ARNT2-PAS-B complex (orange) and apo (blue) using ribbon diagrams. Locations of conformational change are circled in red. (C) A structural comparison of KG-548-bound ARNT2-PAS-B complex (orange) and ARNT-PAS-B complex (pink) using ribbon diagrams. (D) Differences in the amino acid composition of the ligand-binding pocket between ARNT2 (blue) and ARNT (cyan).

The protein sequences of ARNT2 and ARNT differ at two amino acids near the binding site. Specifically, ARNT2 has Leu371, which is replaced by Val397 in ARNT (Figure 55D). Additionally, ARNT2 has Val431, which in ARNT is replaced by Ile457. The Leu371 to Val397 substitution is solvent-exposed and is therefore unlikely to significantly impact ligand binding. In contrast, the Val431 to Ile457 substitution faces the binding pocket and may affect ligand binding. It was hypothesised that Ile457 in ARNT introduces an extra methyl group into the binding pocket, which could block ligand access and prevent the conformational changes required for binding.

Upon binding to ARNT2, KG-548 induce some conformational changes and enlarge the binding pocket of PAS-B domain (Figure 55B, 56). Specifically, the sidechains of residues Ile370, Phe386, Val431, and Val390 are organised to have strong hydrophobic interactions within the apo-ARNT2 structure (Figure 56C). Upon binding, KG-548 causes the redisposition of Ile370, Leu371, Phe386, and Val390 residues, disrupting the hydrophobic interactions, as the observed distance between these amino acids has been increased, particularly to Val431.

As secondary modifications, the side chain of Phe386 flipped inside the pocket, and the side chain of His341 twisted to interact with KG-548 and Phe386 (Figure 56D). It was also hypothesised that if the imidazole ring of His341 is protonated, it can form cation-π interactions with the benzyl group of Phe386. As Phe386 and Val390 repositioned in binding to KG-548, Val431's sidechain loses hydrophobic interactions, opening space within the pocket. To fill the resulting space, Val431 flipped-out as this space is bigger than Val, but Tyr430 flipped-in and forms a hydrogen bond with the Phe386 backbone, adding to the stability of the complex. These rearrangements are expected to change the interaction

surfaces in the β-sheet, α-helices, and HI loop regions that are critical for protein-protein interactions and for forming heterodimers.



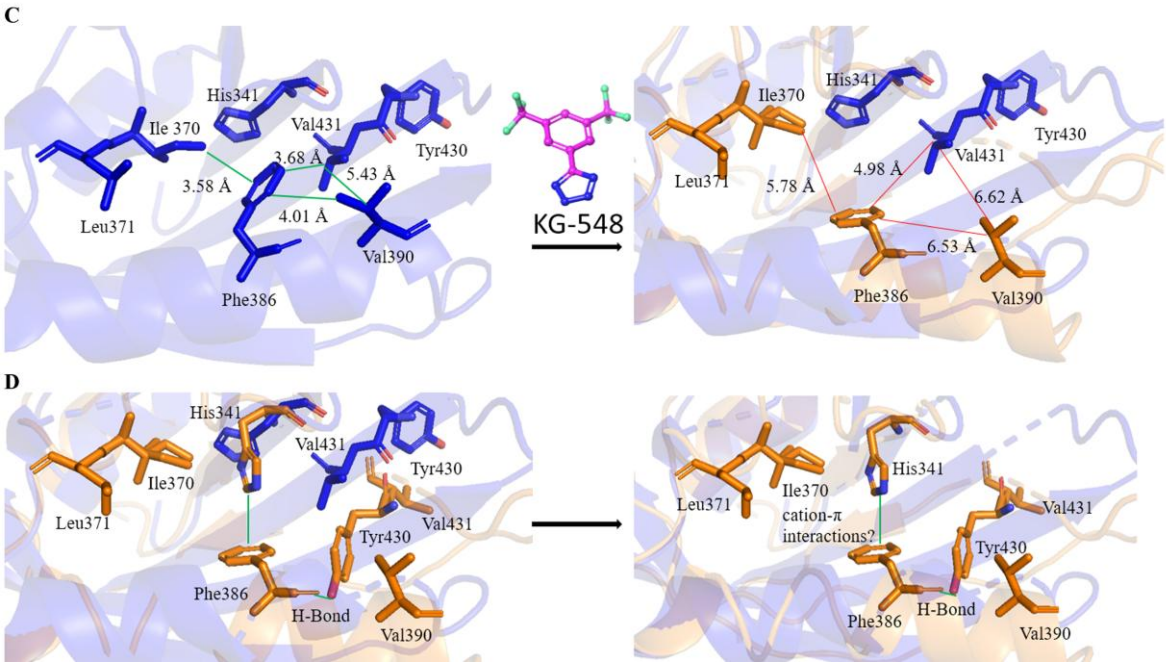


Figure 56. Structural analysis of the mARNT2 PAS-B ligand-binding pocket and interactions with the control compound KG-548. (A) A surface representation generated in PyMOL showing ARNT2 PAS-B surface alters upon binding KG-548. (B) Overview of the mARNT2 PAS-B ligand-binding pocket highlighting residues that interact with KG-548. (C) Primary modifications upon binding KG-548. Hydrophobic interactions within the apo-ARNT2 structure are measured and shown in green. The disrupted interactions upon KG-548 binding are measured and shown in red. (D) Secondary modifications upon binding KG-548, with potential cation-π interactions between His341 and Phe386 sidechains, illustrated with a green line. Apo-mARNT2-PAS-B is shown in blue, and mARNT2-PAS-B bound to KG-548 is shown in orange.

The next step was to determine the implications of KG-548 binding on the ability of ARNT2 to form heterodimers, and to understand why no binding of KG-548 to NPAS1/3 heterodimers was detected via TSA. However, no crystal structure for the NPAS1/3-ARNT2 heterodimers exist to allow accurate structural analysis. A low-resolution structure of a

related complex, mNPAS4-ARNT2 at 4.27 Å resolution (PDB: 7XI3), was examined to understand the impact of KG-548 on heterodimer formation.[51]

In the mNPAS4-mARNT2 complex, residues on the β-sheet of ARNT2 contribute to the dimer interface (Figure 57). In this context, the Val431 sidechain is buried at the dimer interface, while the Tyr430 sidechain is exposed to form a hydrogen bond with the backbone of Tyr253 from NPAS4. As shown in Figure 56, KG-548 binding causes the Tyr430 sidechain to flip, potentially disrupting this critical hydrogen bond between ARNT2 and NPAS4, explaining the observed disruption leading to weaker dimerisation. Additionally, Val341, which is typically positioned internally, is displaced outward, resulting in a steric clash with Leu261 of NPAS4 (measured at a close distance of 2.3 Å). Furthermore, the HI loop of ARNT2 shifts, losing essential contacts with the F helix in PAS-B of NPAS4. These structural alterations are induced at domain-domain interfaces responsible for maintaining heterodimer stability. Therefore, such conformational changes likely destabilise the heterodimer interface, impairing the formation or stability of the ARNT2-containing heterodimer.

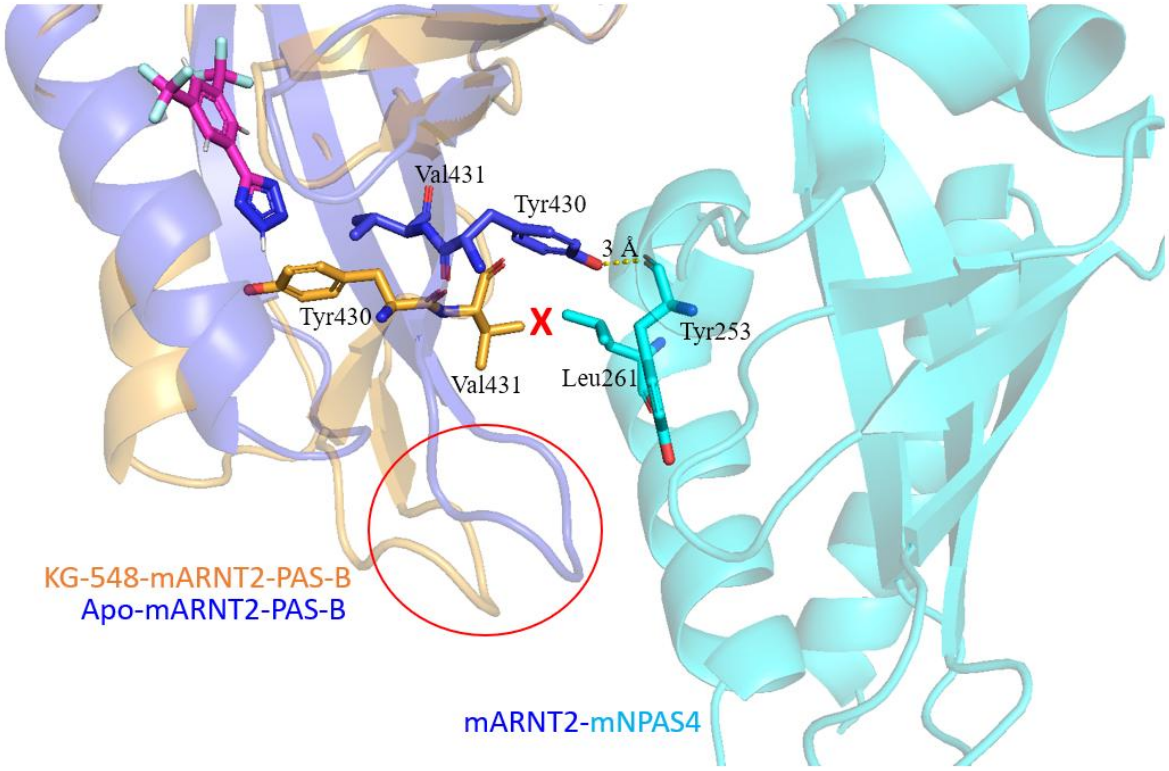


Figure 57. Structural analyses of mARNT2-NPAS4 (PDB: 7XI3) upon binding to KG-548. Apo mARNT2-PAS-B is shown in blue, and mARNT2-PAS-B bound to KG-548 in orange. The PAS-B domain of NPAS4 is illustrated in cyan. The hydrogen bond between Tyr430 of apo-ARNT2 and the backbone of Tyr253 from NPAS4 is measured and shown with a dashed yellow line.

These structural observations demonstrate that ligand binding can induce conformational plasticity and expansion of PAS-B domains, confirming that within the bHLH family, these domains are not rigid but possess considerable structural flexibility that allows pocket remodelling and enlargement upon ligand engagement. This detailed structural analysis underscores how KG-548 binding can induce structural changes in the ARNT2-PAS-B architecture. This structural rationale further explains the unfavourable binding of KG-548

to NPAS1/3-ARNT2 heterodimers as previously observed in TSA. Therefore, KG-548 is not a suitable control compound for examining heterodimeric complexes with ARNT2.

6.2 Crystallisation of the proteins of interest

The aim of this project is to identify compounds that bind the transcription factors NPAS1, NPAS3, and ARNT2, and to elucidate how their modulation influences the 3D structure. X-ray crystallography is a powerful method for determining the three-dimensional structure of proteins and their complexes with other macromolecules or ligands. Crystallography provides precise information about protein conformation in crystals, revealing bond lengths, angles, and interaction surfaces with profound implications for understanding molecular mechanisms of ligand specificity.

The laboratory has extensive experience in crystallising the bHLH-PAS proteins, and the mARNT2-PAS-B structure is available from Dr. Hua Pu. The initial aim was to obtain apo crystals of the target proteins using crystallisation conditions previously reported for several related bHLH-PAS protein constructs, such as mNPAS1/3-ARNT, and mNPAS4-ARNT2.

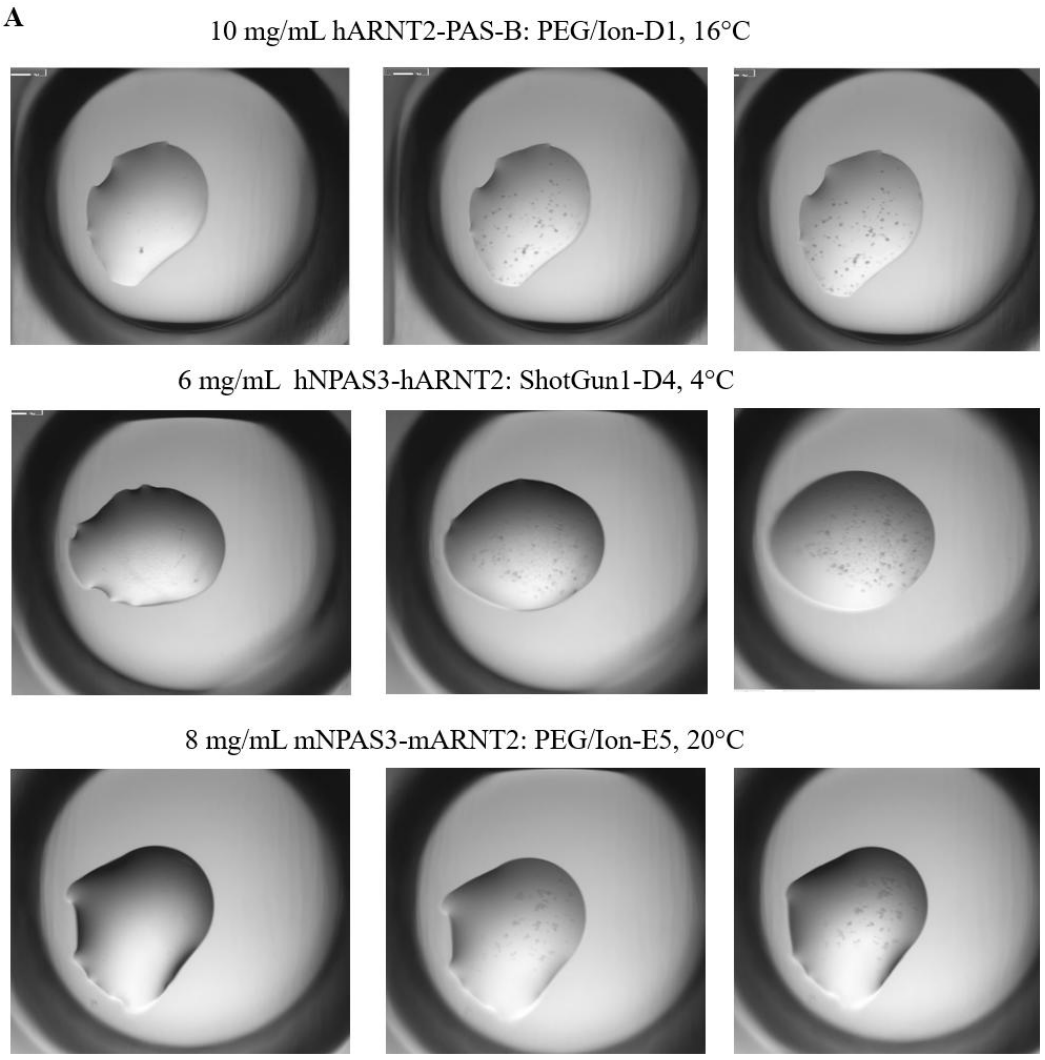
As such, the mNPAS1-mARNT heterodimer was crystallised using the sitting drop vapour diffusion method at 16 °C in solution containing 2% Tacsimate pH 7.0, 3% PEG-3350.[1] The NPAS3-ARNT structure was solved in the complex with DNA, utilising the sitting drop method at 16 °C and crystallisation buffer consisting of 100 mM NH_4F and 9% PEG-3350.[1] The same method was used to crystallise mNPAS4-ARNT2-DNA complex at 16 °C in a solution consisting of 200 mM sodium citrate tribasic and 12% PEG-3350.[51] The PAS-B

domain of ARNT2 has been crystallised in condition 0.1 M HEPES pH 8.2, 25% PEG-3350, 0.2 M $(NH_4)_2SO_4$, and with 0.1 M HEPES pH 8.0, 3 M NaCl (data not published).

The aforementioned crystallisation conditions and related variations were tested with several concentrations of uncut and TEV-cleaved hARNT2-PAS-B, hNPAS1-hARNT2, and both human and mouse NPAS3-ARNT2 heterodimers, yet no crystals developed. To identify conditions which are favourable for crystallisation of the target constructs, new crystallisation screens using the sitting drop vapour diffusion method at both 4 °C and 16 °C were carried out with TEV-cleaved and purified by size-exclusion chromatography target proteins. Several primary screen libraries, including Index-HT, PEG/Ion, Morpheus, ShotGun1, Large Molecular Fragment (LMF), and JCSG+, have been used with the goal of optimising crystal formation. These libraries are diverse and combine both grid and sparse matrix approaches, utilising a portfolio of 96 reagents. Each screen incorporates various crystallisation reagents, such as salts (ammonium sulphate, sodium chloride), different buffers at various pH levels, neutralised organic acids, and polymers like PEG and MPD.

Efforts to obtain crystals for the human ARNT2-PAS-B single domain, hNPAS1-hARNT2, and human or mouse NPAS3-ARNT2 heterodimers, were not successful. Protein samples either precipitated or formed salt crystals (Figure 58). To distinguish potential protein crystals from salt or other non-protein crystals, all crystal formations were examined under UV light based on the intrinsic fluorescence of aromatic amino acids (tryptophan and tyrosine). When exposed to UV light, tryptophan absorbs light at 280 nm and emits fluorescence at 330-350 nm. An imaging system equipped with appropriate UV detectors able to capture and visualise the fluorescence signal was used. Under these conditions, protein crystals often fluoresce, whereas salt crystals and other small molecules do not, remaining dark or dull. The

corresponding crystallisation drops did not fluoresce, indicating crystals of salts, buffer components, or other non-protein precipitates. A range of troubleshooting approaches, including different protein buffers, varying protein concentrations, the size and the ratio of protein-screen drop, incubation temperature and time, did not yield protein crystals.



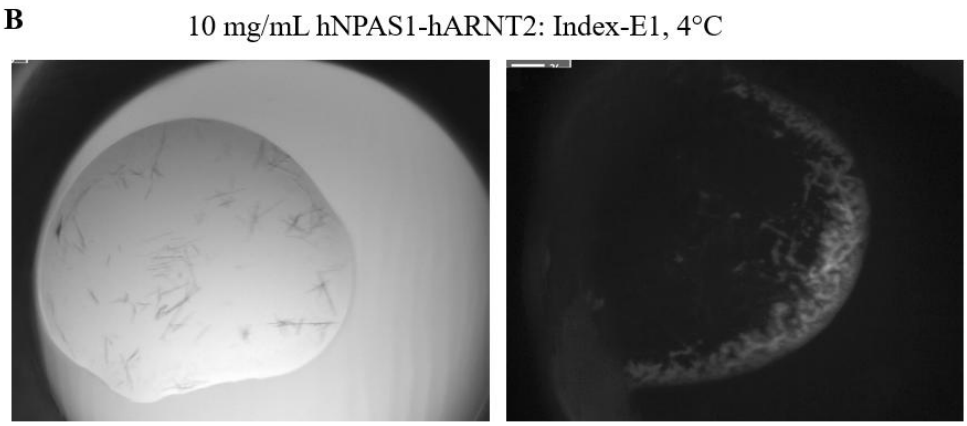


Figure 58. Crystallisation screening. (A) Representative crystal screening snapshots with different proteins and conditions. (B) The examination of crystallisation drops with a 280 nm UV light source. Salt crystals do not fluoresce under UV light.

To obtain crystals for mARNT2-PAS-B, known crystallisation conditions developed in the laboratory by Dr. Hua Pu were used. Protein was concentrated to 10 mg/mL in buffer 20 mM BTP pH 8.0, and 150 mM NaCl. Crystals were grown with 0.1 M HEPES pH 8.0, 3 M NaCl by sitting drop vapour diffusion. After three days, small crystals began to appear (Figure 59). If left longer, crystals then formed clusters which were not ideal for crystal picking and X-ray diffraction. To obtain larger single crystals, the optimisation of crystallisation conditions was performed.

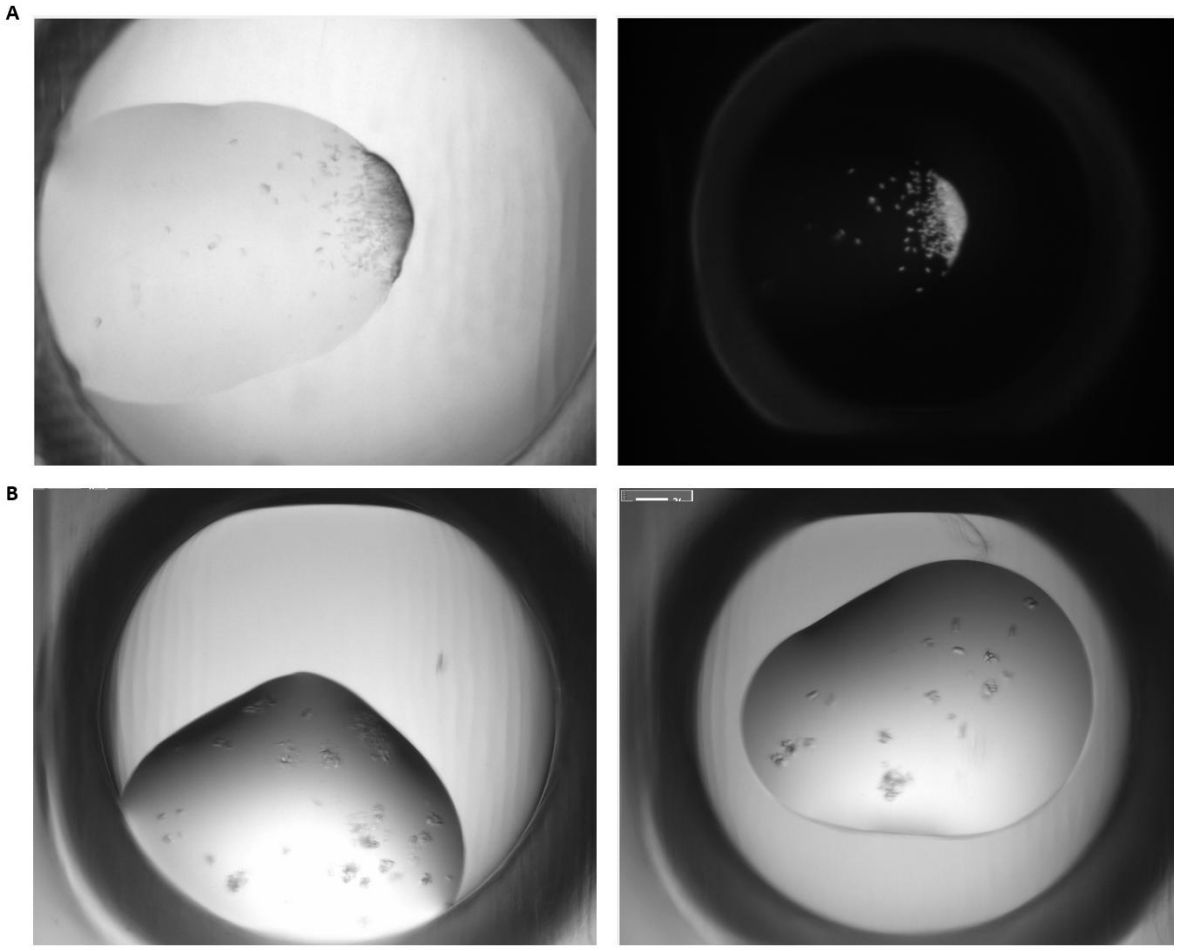


Figure 59. Representative unoptimised mouse ARNT2-PAS-B crystals. Crystals were grown in 0.1 M HEPES pH 8.0, and 3 M NaCl by sitting drop vapour diffusion at 16 °C. Protein was previously concentrated to 10 mg/mL in buffer 20 mM BTP pH 8.0, with 150 mM NaCl. (A) Large number of small crystals of mARNT2-PAS-B were developed and confirmed using a 280 nm UV light source. (B) The formation of crystal clusters was noted on the third day after the crystals developed.

6.3 Optimisation of mARNT2-PAS-B crystals

A crystallisation experiment usually begins with finding stabilising conditions for a protein. It was observed that mARNT2-PAS-B eluted from S75 column in gel filtration buffer formed

large, single, well-ordered crystals (Figure 60A). Therefore, it was decided to omit buffer exchange with 20 mM BTP pH 8.0, and 150 mM NaCl, and instead concentrate protein to 10 mg/mL in the original gel filtration buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, and 5% glycerol.

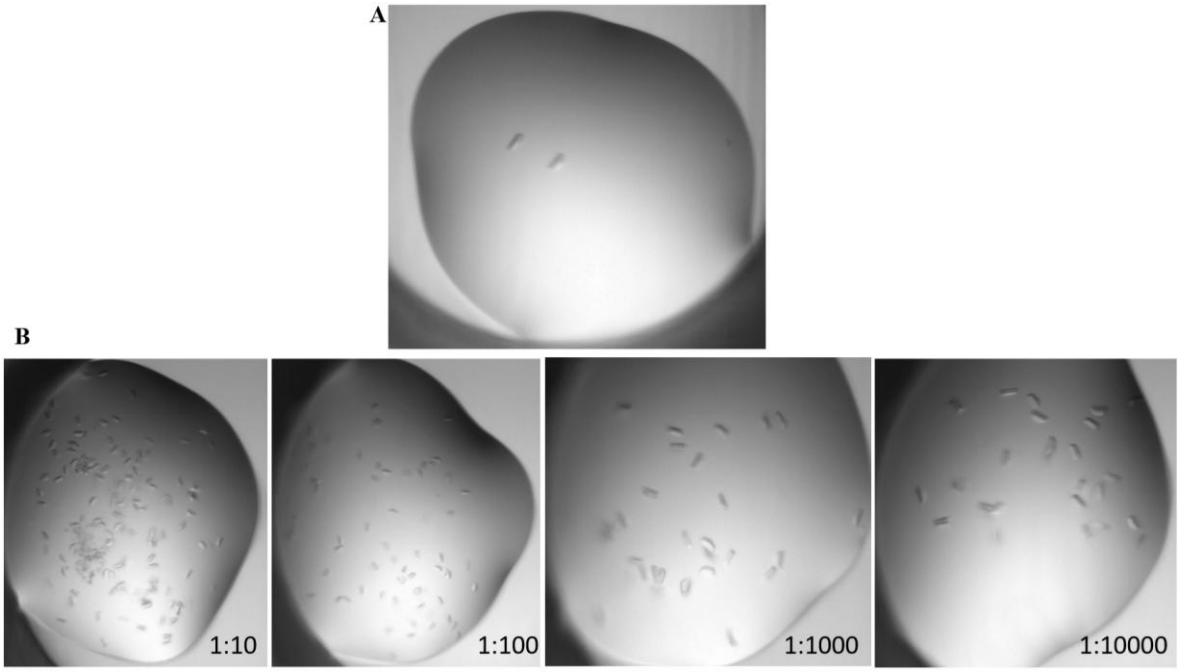


Figure 60. Optimisation of mARNT2-PAS-B protein crystallisation. Crystallisation was carried out using sitting drop vapour diffusion at 16 °C, by mixing 200 nL of protein (10 mg/mL) and 600 nL of reservoir solution containing 0.1 M HEPES pH 8.0 and 3 M NaCl. (A) A formation of a few single mARNT2-PAS-B crystals after the protein concentration step was carried out in HEPES pH 7.5 gel filtration buffer instead of BTP pH 8.0 buffer. (B) Optimisation of mARNT2-PAS-B crystals with seeding protocol. Using several dilutions, different concentrations of seed stock were added to the protein reservoir drops, which led to the formation of large individual crystals in three days.

However, the formation of such crystals was not consistent, probably due to the uncontrolled nucleation. To ensure reproducibility of crystal formation and to increase the number of

individual crystals, a seeding protocol was implemented. To promote crystal growth when nucleation is difficult, seeding was relied on to introduce a small crystal seed as a starter to trigger the growth of crystals with consistent results (Figure 61). An advantage of seeding is the significantly shorter time required for crystal formation compared to spontaneous nucleation.

Previous cluster crystals were therefore crushed with a crystal crusher probe, and seed stock was created. Concentrated mARNT2-PAS-B in gel filtration buffer was mixed with reservoir solution, and different dilutions of seed stock, such as 1:10, 1:100, 1:1000, and 1:10000, were added to the drop. In three days, crystals started to appear in accordance with the ratio of the seed stock (Figure 60B). Thus, the seeding protocol provided reproducible, larger, and individual crystals per seeding drop.

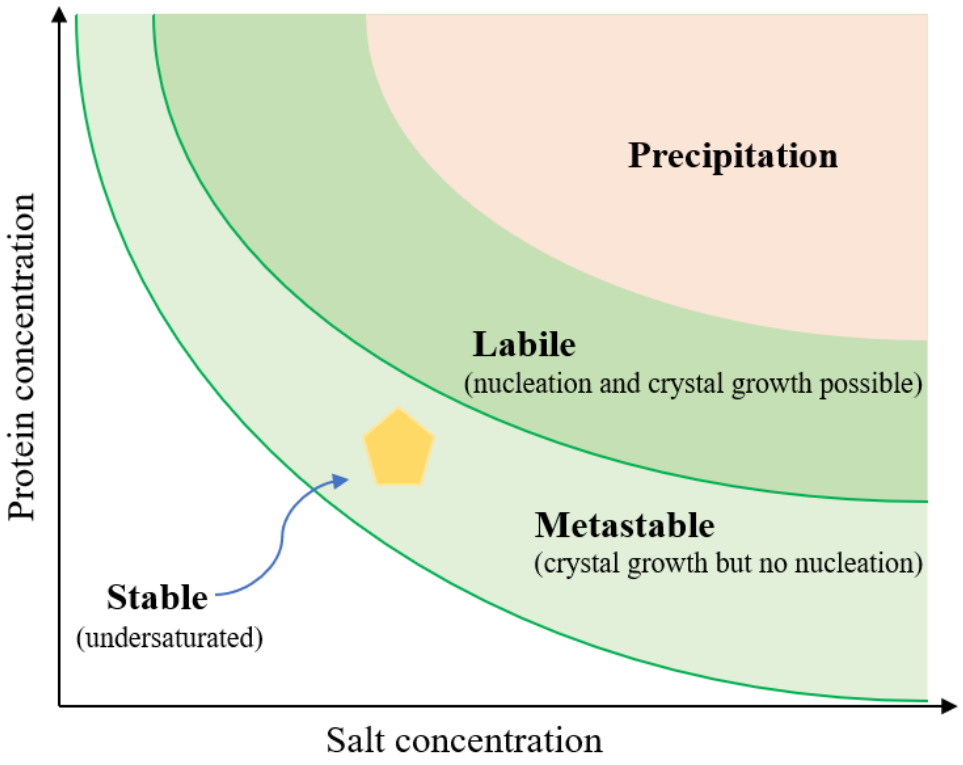


Figure 61. Diagram of crystal progression over time. At the beginning of the sitting drop experiment, a clear drop represents a stable or undersaturated zone, as crystal nucleation and

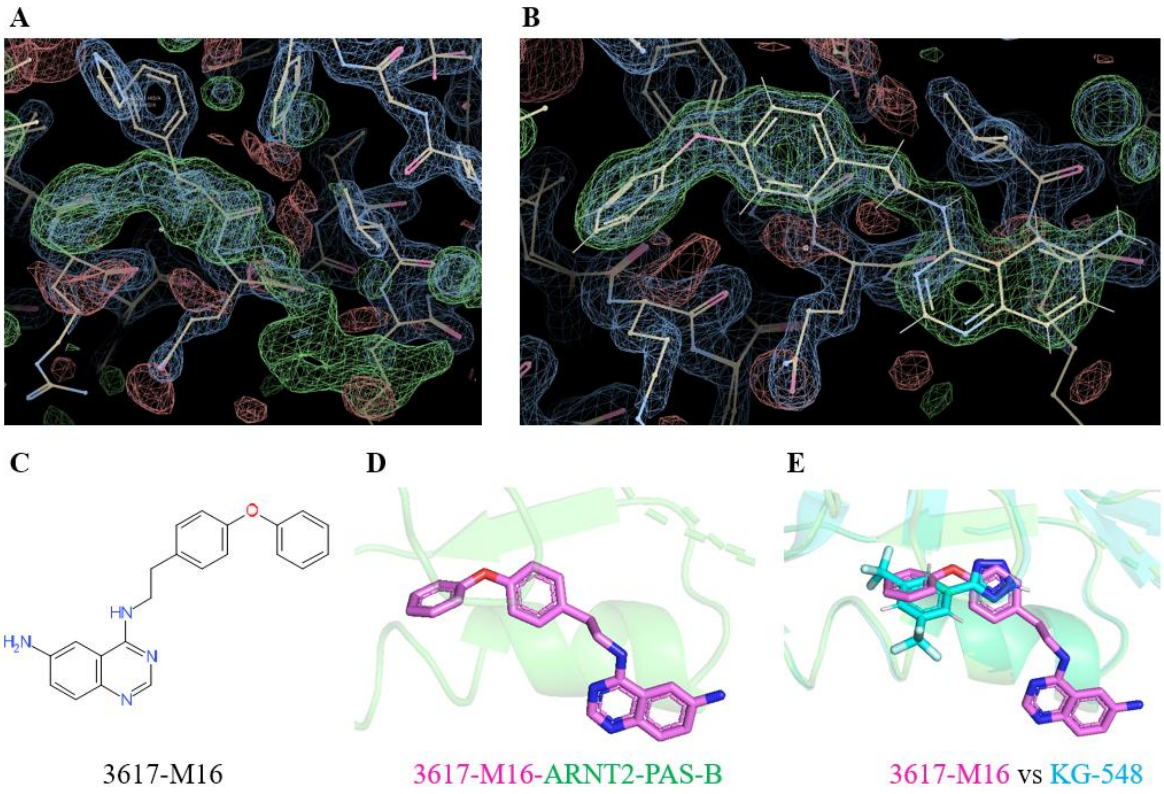
growth are not possible. As water evaporates, the concentrations of protein and precipitants increase, and the system is driven from undersaturated to supersaturated conditions. In the metastable zone, spontaneous homogeneous nuclei cannot occur, but the crystals can grow. The formation of nuclei can occur only in the labile zone with subsequent crystal growth. In a precipitation zone, nucleation of crystals and growth are not possible, and the protein sample precipitates. By placing starting seeds into a metastable zone, the often inefficient and unpredictable nucleation step can be bypassed, promoting more consistent crystal growth. The administered crystal seed is shown in yellow.

6.4 Crystal structures of mARNT2-PAS-B complexes with tested compounds

To decipher the basis for compound selectivity by the ARNT2-PAS-B domain, crystal structures of the ligand-bound complexes were sought. Using the seeding protocol, protein crystals were grown in several 96-well crystallisation plates through sitting drop vapour diffusion for three weeks. The confirmed ARNT2-PAS-B hits were soaked in 24 hours before crystal harvesting. Picked crystals were flash frozen in liquid nitrogen with 30% glycerol as the cryoprotectant.

Out of 13 soaked compounds, three resulted in ligand-bound protein complex structures at high resolution (Appendix 5). ARNT2-PAS-B complexes were solved in the P62_2_2 space group. Based on “omit” electron density maps, the 3617-M16 binding location was unambiguously visualised within the large pocket of ARNT2 PAS-B at 1.67 Å resolution (Figure 62A-B). As expected, the 3617-M16 ligand binds the same binding pocket as KG-548 (Figure 62D-E). Inside ARNT2-PAS-B, 3617-M16 contacts multiple residues by forming hydrophobic interactions. Like KG-548, 3617-M16 is also stabilised by a hydrogen

bond with His341 at a distance of 3 Å (Figure 62F). Residue His341 might be a key factor in keeping certain ligands inside the PAS-B pocket. Apart from a direct hydrogen bond with His341, 3617-M16 also has several water-mediated hydrogen bonds with Gln387, Arg383, Val390, and Lys391 (Figure 62G).



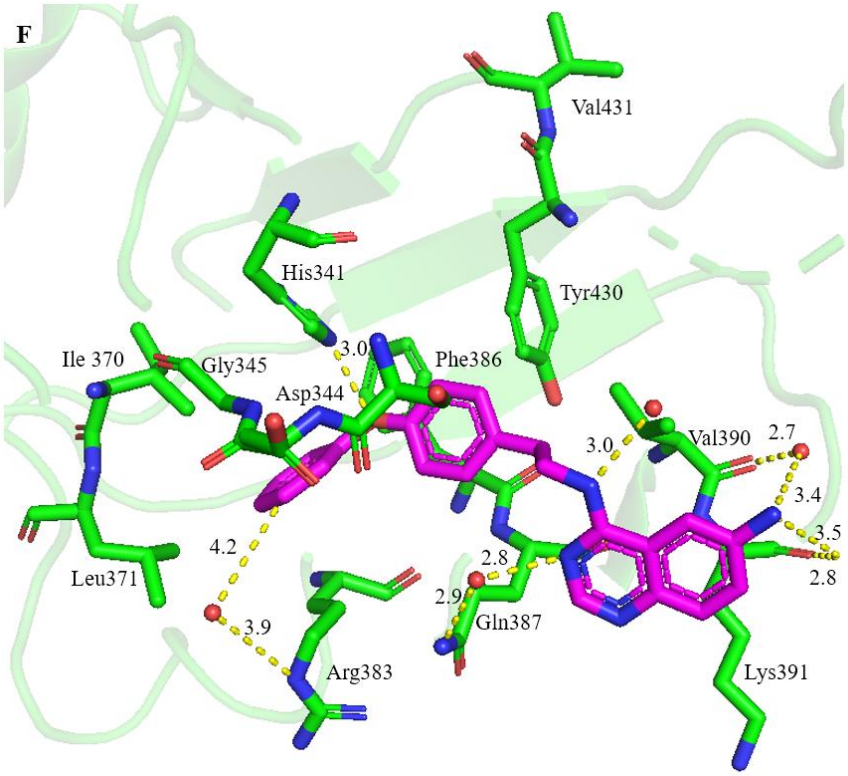
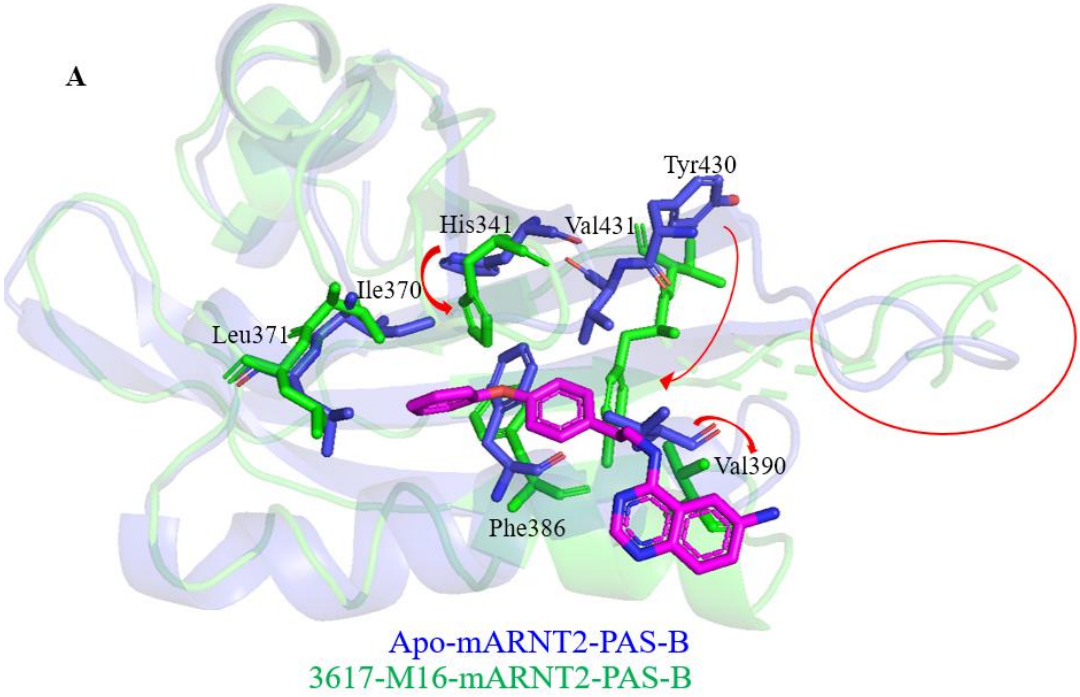


Figure 62. X-ray crystallography for 3617-M16 binding to ARNT2-PAS-B. (A) Density map without ligand. (B) Density map with the 3617-M16 ligand within the ARNT2-PAS-B domain. (C) Chemical structure of 3617-M16. (D) Close-up look at the location of 3617-M16 inside the ARNT2-PAS-B domain; 3617-M16 is shown in pink. (E) The KG-548-bound structure (blue) was superposed on the 3617-M16 (pink) to show ligand binding overlap. (F) Interaction of 3617-M16 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts or interactions with protein residues and water molecules in the crystal structure.

Like KG-548, binding of 3617-M16 induced positional shifts in several amino acids, including Phe386, His341, Val390, Val431, and Tyr430 (Figure 63A). Specifically, 3617-M16 binding caused Phe386 to twist inward, which in turn pushed Val390 of the α-helix downward to accommodate the ligand. Ligand binding further caused secondary modifications. His341 rotated inward and formed a hydrogen bond that stabilised the ligand

within the binding pocket. This conformational change disrupted local interactions. Notably, Val431 lost its hydrophobic contacts and was displaced outward. To compensate, Tyr430 flipped inward and formed a hydrogen bond with the backbone of Phe386. Therefore, binding of 3617-M16 caused substantial conformation changes across the β-sheet, α-helices, and loops, which may influence heterodimerisation of ARNT2 with its partners. As compared to apo-ARNT2-PAS-B, these structural changes significantly expanded the binding pocket to accommodate a ligand (Figure 63B).



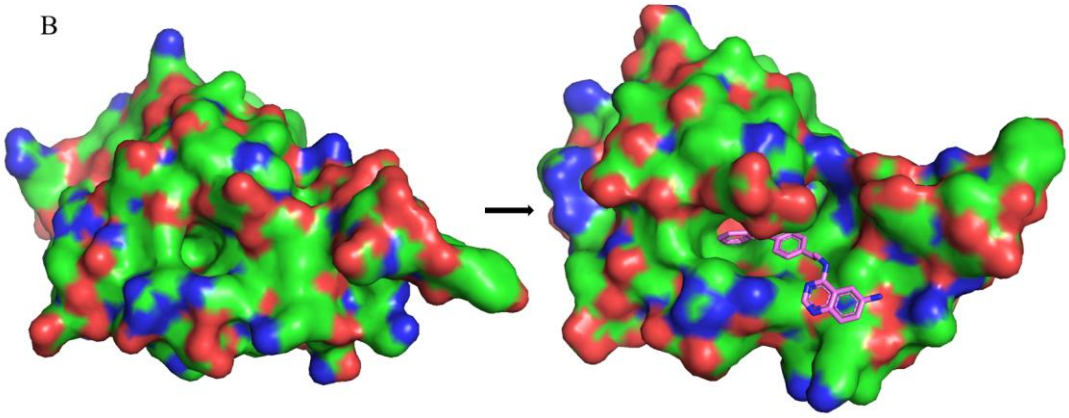
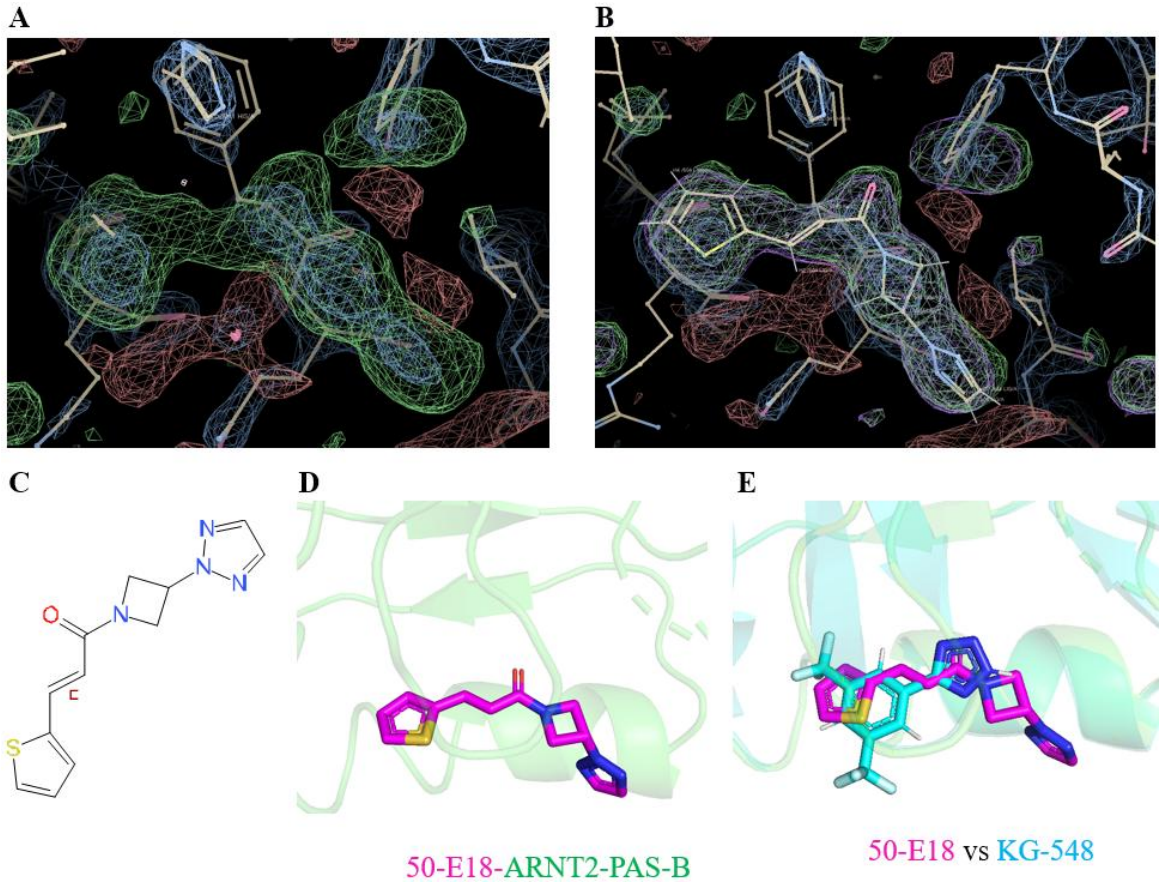


Figure 63. Conformation changes of apo-ARNT2-PAS-B upon binding 3617-M16. (A) Structural analyses of mARNT2-PAS-B upon binding 3617-M16 ligand. Unbound mARNT2-PAS-B is shown in blue, and mARNT2-PAS-B bound to 3617-M16 is illustrated in green. Ligand 3617-M16 itself is shown in pink. Direction and positional shifts of Phe386, His341, Val390, Val431, and Tyr430 upon binding to 3617-M16 are indicated with red arrows. Change in the HI loop is circled in red. (B) Expansion of the ARNT2-PAS-B binding pocket upon binding 3617-M16.

The binding of 50-E18 to ARNT2-PAS-B was determined at 1.82 Å resolution (Appendix 5). Like 3617-M16, the 50-E18 crystal structure was obtained by forming crystals after 24-hour soaking. The 50-E18 ligand was clearly visible in the electron density map, allowing for the determination of its binding mode, orientation, and key interactions within the ARNT2-PAS-B binding pocket (Figure 64A-B). The omit map shows that 50-E18 also binds the pocket where KG-548 and 3617-M16 bound (Figure 64D-E). Not surprisingly, the 50-E18 ligand formed a hydrogen bond between His341 and the carbonyl oxygen of the compound within a distance of 3 Å, supporting the hypothesis that His341 plays a key role in the stability of a ligand inside the binding pocket (Figure 64F). Inside the ARNT2-PAS-B

pocket, 50-E18 contacts several residues, forming hydrophobic interactions with Ile370, Leu371, and Gly345 via its thiophene ring (Figure 64G).



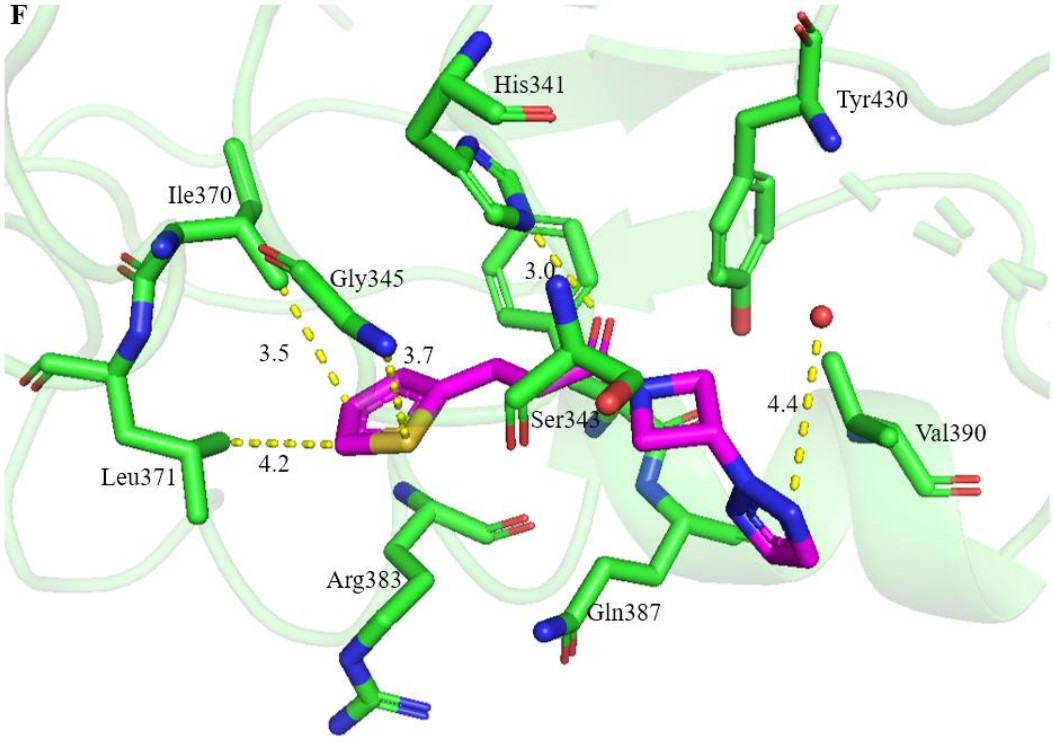
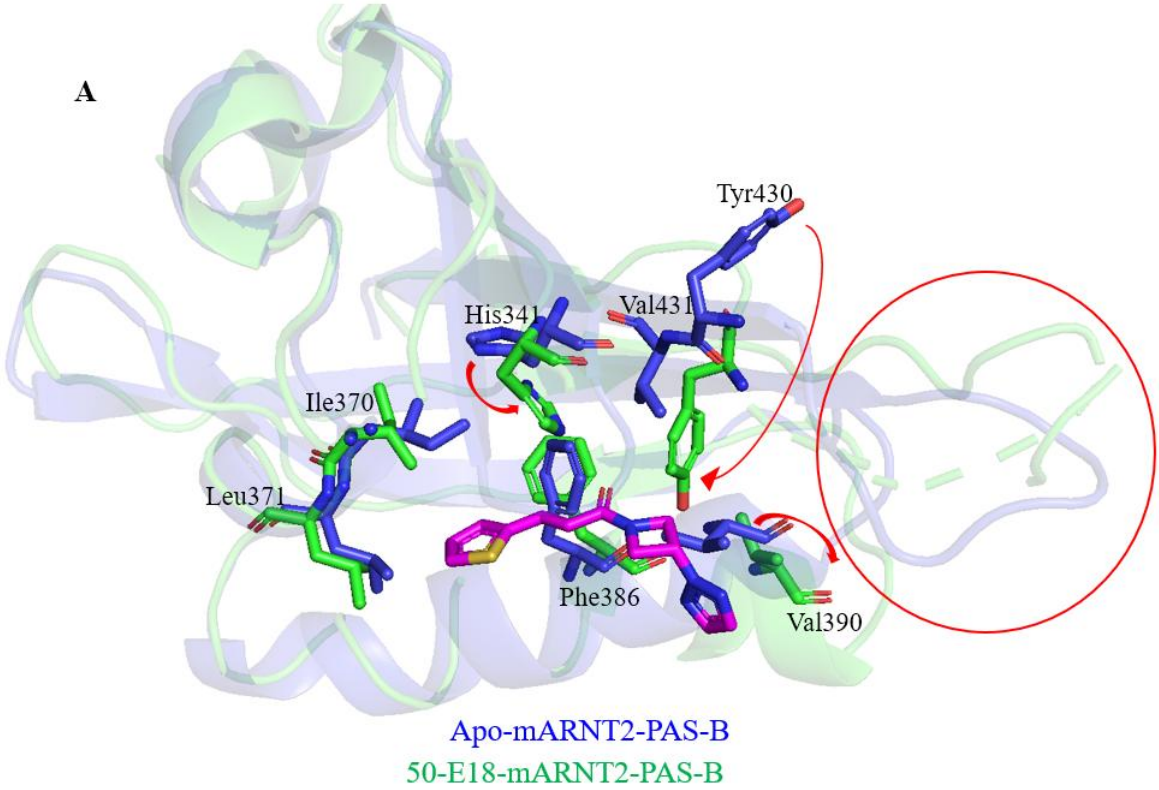


Figure 64. X-ray crystal structure of the 50-E18 compound with ARNT2-PAS-B. (A) Density map without ligand. (B) Density map with the 50-E18 ligand within the ARNT2-PAS-B domain. (C) Chemical structure of 50-E18. (D) Close-up look at the location of 50-E18 inside the ARNT2-PAS-B domain; 50-E18 is shown in pink. (E) The KG-548-bound structure (cyan) was superposed on the 50-E18 (pink) to show the overlapping binding location. (F) Interactions of 50-E18 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts or interactions with protein residues and water molecules in the crystal structure.

Comparable adjustments were noted in the ARNT2-PAS-B binding pocket for accommodating the 50-E18 ligand. The same amino acids within the binding pocket underwent distinct conformational rearrangements (Figure 65A). Both Ile370 and Leu371 side chains were repositioned and created additional space for the accommodation of the 50-E18 ligand. The Phe386 ring exhibited a notable inward rotation, making room to

accommodate the ligand. Additionally, Val390 was displaced downward, altering the lower region of the binding pocket. In contrast, His341was flipped outward, possibly to disengage from prior intramolecular contacts, and to stabilise the ligand within the pocket by forming a hydrogen bond. This mARNT2-PAS-B-50-E18 crystal structure did not reveal the Val431 position. However, analogous tweaks of Tyr430, which was reoriented inward, suggest similar conformational changes for Val431 upon binding KG-548 and 3617-M16. The displaced Tyr430 is likely to fill the space created by the ligand and to further stabilise the enlarged binding pocket (Figure 65B). In summary, these rearrangements suggest a coordinated structural adaptation of the binding site in response to ligand engagement.



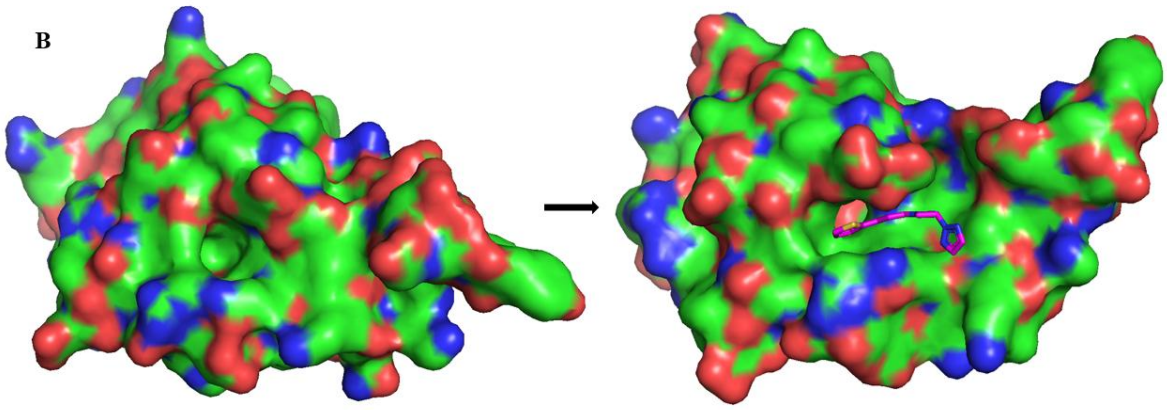


Figure 65. Conformation changes of apo-ARNT2-PAS-B upon 50-E18 ligand binding. (A) Structural analyses of mARNT2-PAS-B upon binding 50-E18. Unbound mARNT2-PAS-B is shown in blue, and mARNT2-PAS-B bound to 50-E18 as illustrated in green colour. Ligand 50-E18 itself is shown in pink. Direction and positional shifts of Phe386, His341, Val390, and Tyr430 upon binding 50-E18 are shown with red arrows. The red circle points to changes in the HI loop. (B) Expansion of the ARNT2-PAS-B binding pocket upon binding 50-E18.

An additional mARNT2-PAS-B-3613-D4 co-crystal structure was also characterised. The initial 24-hour soaking protocol resulted in limited electron density, with partial visibility of the ligand at 2.13 Å resolution (Figure 66A-B). From the 3613-D4 electron density map, it is clear that one phenyl ring is absent. To increase resolution and enhance ligand occupancy, an extended soaking period of 72 hours was implemented before crystal harvesting. While the prolonged soaking resulted in an improved overall crystal structure resolution of 1.71 Å, it did not significantly increase ligand-associated electron density (Figure 66D). One of the two phenyl rings remains unresolved. The overlay of the KG-548-bound structure suggests that 3613-D4 occupies the same PAS-B binding site as previously characterised compounds.

The binding of 3613-D4 could therefore potentially lead to similar ligand-induced structure changes (Figure 66E-F).

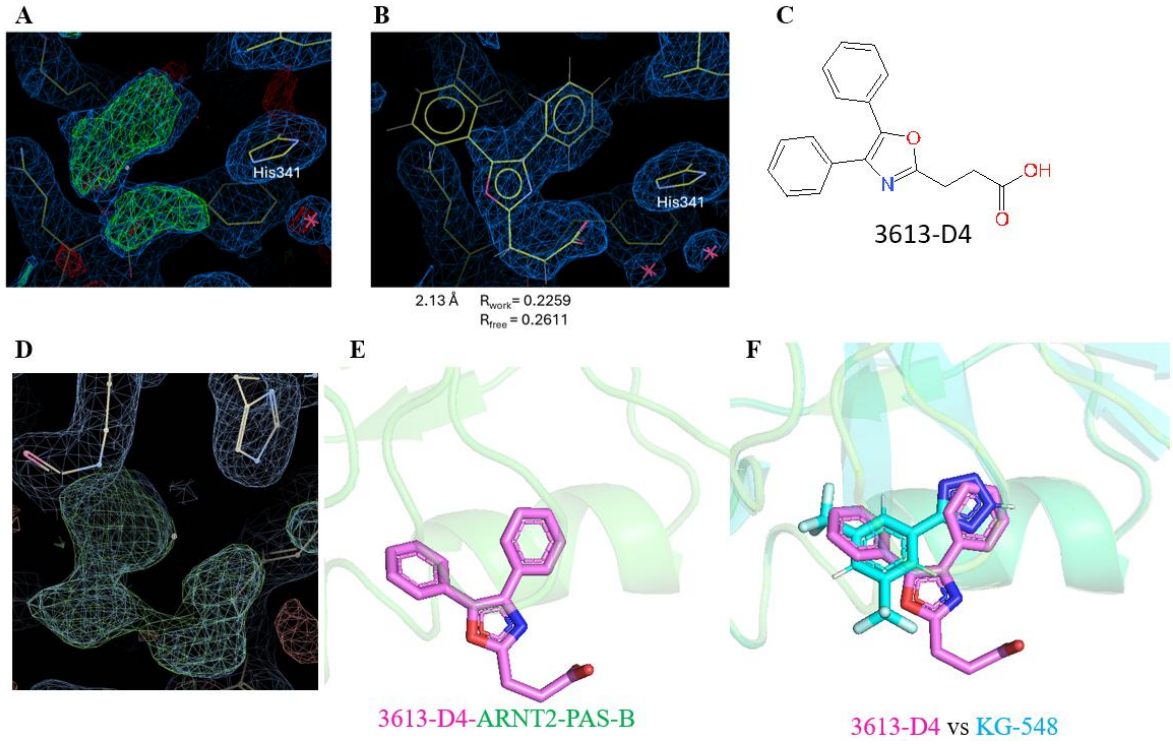


Figure 66. X-ray crystallography with 3613-D4 and ARNT2-PAS-B. (A) Density map without ligand with 24-h soaking protocol. (B) Density map with 3613-D4 ligand within the ARNT2-PAS-B domain with 24-h soaking protocol. (C) Chemical structure of 3613-D4. (D) Density map without ligand with 72-h soaking protocol. (E) Close-up look at the location of 3613-D4 inside the ARNT2-PAS-B domain; 3613-D4 is shown in pink. (F) The KG-548-bound structure (blue) was superposed on the 3613-D4 (pink) to show the same binding location.

To conclude, X-ray crystallography confirmed that several ligands bind directly to the ARNT2 PAS-B domain. Detailed structural analyses revealed that all hit compounds occupy the same binding cavity and interact with residues similar to those involved in KG-548 binding. Analogous to KG-548, these ligands induce conformational plasticity and expansion

of the PAS-B domain. The observed structural rearrangements occur at the interfaces between the β-sheet, α-helices, and HI loop regions, areas that are critical for protein-protein interactions, particularly for the formation and stabilisation of heterodimers. Consequently, it is expected that, like KG-548, all identified compounds would be unfavourable for heterodimer stability and, therefore, would not exhibit binding to the heterodimers in the TSA assay.

Chapter 7: Microscale thermophoresis binding assay

7.1 Principle of microscale thermophoresis

After obtaining several crystals, the next step is to determine the binding affinity between target proteins and tested compounds. To study molecular interactions between biomolecules, MST was employed. Due to its high sensitivity and low sample requirement, MST seems the most suitable technique for the characterisation and quantification of the binding effect. Another advantage is that MST is performed entirely in free solutions and does not require surface immobilisation. Preserving the native structure and activity of the biomolecules enables the characterisation of binding events under near-physiological conditions.

MST is a biophysical technique that relies on both thermophoresis and the temperature-related intensity change (TRIC) of the fluorescent probe. Thermophoresis is the directed movement of molecules in a microscopic temperature gradient. When an infrared (IR) laser induces a temperature gradient, biomolecules diffuse away from the hotter area according to their thermophoretic properties. The biophysical basis of this movement is described by the Soret coefficient (S_T) and follows the equation:

$$\mathbf{c_{hot}/c_{cold} = \exp(-S_T \Delta T)}$$

Where c_{hot}/c_{cold} are the concentrations of the molecule in the hot and cold regions, S_T is the Soret coefficient, and ΔT is a temperature difference (usually 2-5 K) between the hot and

cold states. The Soret coefficient reflects the diffusion characteristics of the molecules, such as their biomolecular size, charge, conformation, and solvation shell. Upon binding of a ligand, changes in these physicochemical properties affect the value of S_T and alter the thermophoretic behaviour of a fluorescently labelled protein.[204]

At the same time, TRIC describes the immediate change in fluorescence intensity caused by the local heating of the sample. This fluorescence response may also be affected by binding-induced modifications in the molecular environment, including conformational changes, fluorophore exposure, or flexibility. Both thermophoresis and TRIC therefore contribute to the observed changes in fluorescence signal when a non-fluorescent binding partner is titrated against a constant concentration of a fluorescently labelled protein. This change depends on the ligand concentration, allowing the characterisation of binding affinities with a low detection limit, enabling MST to be highly responsive and sensitive to even minimal changes in protein structure. Such characteristics make MST very suitable for detecting the binding of a small molecule to a protein.

The raw fluorescence data can be divided into five events: the initial fluorescence, T-jump, thermophoresis, the steady-state, and back-diffusion (Figure 67). The initial fluorescence (F_{cold}) represents the state of the sample before the IR laser is turned on. The T-jump follows after IR heating and refers to the immediate fluorescence intensity change of the dye molecule upon rapid heating of the sample. The differences in the signal during this phase can also indicate the binding event, if the dye is sensitive to its local environment and if the binding alters this local environment. Afterwards, the thermophoresis follows. A pronounced difference in bound and unbound thermophoretic movement of fluorescent molecules in the sample could be observed. At the steady-state, thermophoresis reaches equilibrium, and F_{hot}

values could be measured. The IR laser is switched off at the back-diffusion state, or fluorescence recovery. As the temperature returns to equilibrium, molecules start to diffuse back into the measuring region of the capillary, increasing the fluorescence.

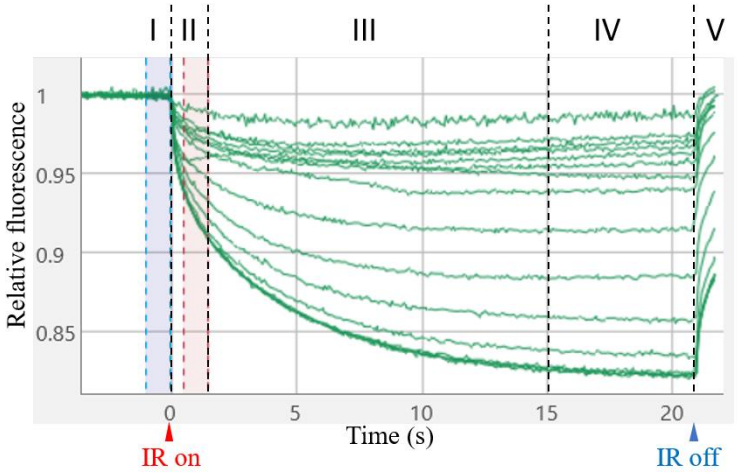


Figure 67. Representative MST trace for different concentrations of hNPAS1-ARNT2.The blue and red regions indicate the windows where the F_{cold} and F_{hot} measurements were taken, respectively. The red and blue triangles show when the IR was turned on and off. The graph is divided into five regions: (I) the initial fluorescence, (II) T-jump, (III) thermophoresis, (IV) the steady-state, and (V) back-diffusion.

The MST signal is recorded as a normalised fluorescence difference between the initial state and the steady state:

$$\mathbf{\Delta F_{norm}} = \frac{\mathbf{F_{hot} - F_{cold}}}{\mathbf{F_{cold}}} \times \mathbf{100\%}$$

By plotting this fluorescence change ΔF_{norm} against the ligand concentration, a binding curve can be generated, from which, using nonlinear regression, key parameters such as the K_D value can be extracted. The data are generally fitted to a 1:1 binding equilibrium:

$$\mathbf{\Delta F_{norm}} = \mathbf{\Delta F_{max}} \frac{[\mathrm{L}]}{\mathbf{K_D} + [\mathrm{L}]}$$

[L] is the concentration of the ligand, ΔF_{max} is the maximum change in normalised fluorescence, and K_D is the dissociation constant, the concentration of a ligand at which half of the fluorescently labelled molecules are bound. Depending on the system, K_D provides a reliable quantitative measure of binding affinity in the nanomolar to millimolar range.

7.2 Development of the MST assay

To test the affinities of all confirmed hits, target proteins were first labelled with a fluorophore. The RED-Tris-NTA dye was chosen as it was designed to specifically bind the 6xHis expression tag of recombinant proteins with high affinity. All three proteins of interest, mARNT2-PAS-B, hNPAS1/3-ARNT2, have a 6xHis tag.

To achieve optimal sensitivity and accurate determination of the binding affinity (K_D), it is critical to choose a protein concentration that gives a strong and stable fluorescence signal. To detect binding, MST measures small changes in fluorescence, and if the signal is too weak or unstable, those changes can be lost in background noise. Following incubation with the fluorophores, different concentrations starting from 4 μM to 0.12 nM of labelled proteins were tested via MST to obtain a protein titration curve (fluorescence intensity vs. protein concentration). Ideally, the fluorescence signal from the labelled protein should not be lower than 150 fluorescence units (FUs).

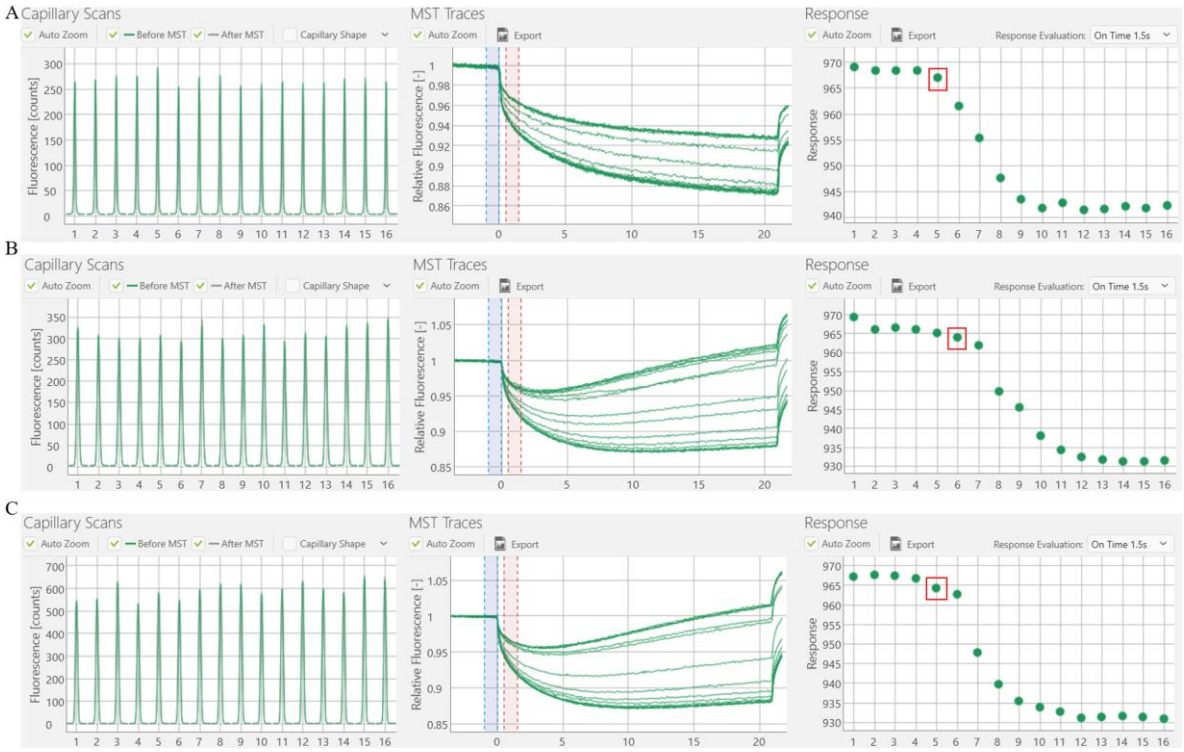


Figure 68. Protein curve analyses via MST.MST traces of RED-Tris-NTA-labelled target proteins from 4 μM to 0.12 nM with two-fold dilution. Protein concentration curves with capillary scans, MST traces, and plotted fluorescence response for (A) mARNT2-PAS-B, (B) hNPAS1-ARNT2, and (C) hNPAS3-ARNT2 labelled proteins. The chosen optimal concentration where the signal is steady and reproducible is indicated with red squares.

No sudden jumps or drops in fluorescence were observed for any of the three proteins, which might indicate aggregation or poor labelling, suggesting the stability of biomolecules (Figure 68). A good signal-to-noise ratio (SNR) means the fluorescence emitted by the labelled protein is clearly above the background noise and does not fluctuate too much. When SNR is high, the measurements are considered more reliable and sensitive to binding interactions. Therefore, the optimal protein concentration should be chosen from the plateau or linear stable region. With the measured K_D between 70-60 nM for all labelled proteins, the optimal

assay concentration was selected at 250 nM for both mARNT2-PAS-B and hNPAS3-ARNT2 (Figure 68). For hNPAS1-ARNT2 labelled protein, the concentration was picked at 125 nM.

As previously, KG-548 did not prove itself to be reliable control with either of the heterodimers; it was only assessed with ARNT2-PAS-B protein. Thus, 250 nM of mARNT2-PAS-B, previously labelled with first-generation RED-Tris-NTA dye, was incubated with a 16-point serial dilution series of KG-548 control compound. Based on the previously predicted EC_{50} of 19 μM from the TSA experiment, KG-548 concentration therefore varied from 500 μM to 15 nM. Unfortunately, the KG-548 did not appear to show any activity with mARNT2-PAS-B protein (Figure 69A).

Generally, MST is a sensitive method; however, it can fail to detect binding due to several reasons. The most frequent one is when the labelled protein concentration is too low, because the fraction of the bound complex may be too small to produce a measurable thermophoretic shift. Or, if the ligand titration range does not exceed the K_D, full binding saturation will not occur. Additionally, the binding effect might be sensitive to assay conditions, including pH, salt concentration, or the presence of certain additives.

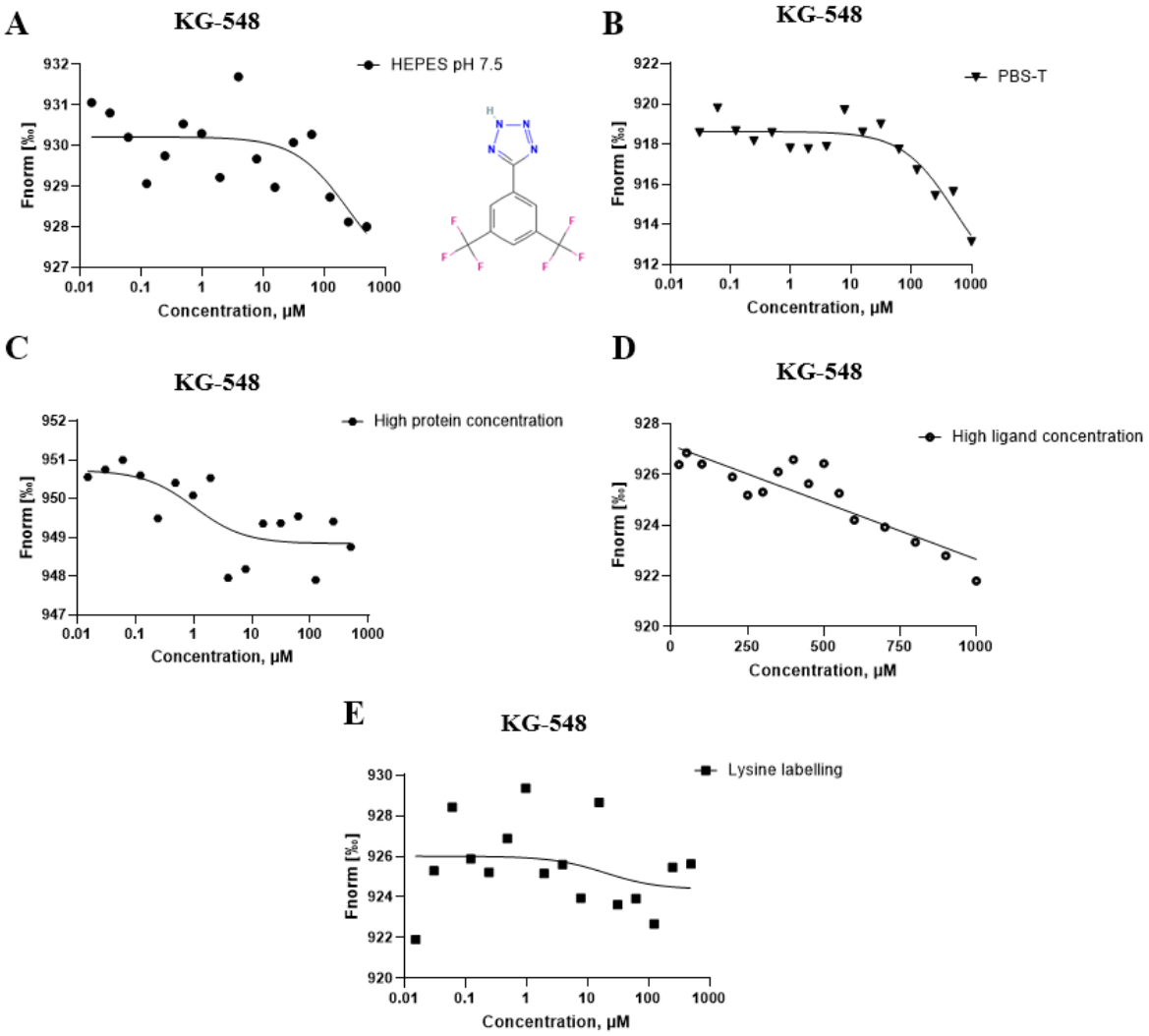


Figure 69. Comparison of the MST analyses for KG-548 against mARNT2-PAS-B.(A) 250 nM of ARNT2-PAS-B was labelled with the first generation of RED-Tris-NTA dye using 20 mM HEPES pH 7.5, 150 mM NaCl, and 0.1% Tween-20. (B) Buffer switch to the kit provided PBS-T. (C) Protein concentration of labelled ARNT2-PAS-B was increased to 1200 nM. (D) The dose-response of the KG-548 control compound was increased up to 1000 μM and decreased stepwise to 25 μM. (E) KG-548 affinity with covalently labelled mARNT2-PAS-B with lysine residue.

To address these possible reasons, the assay buffer was changed from 20 mM HEPES pH 7.5, 150 mM NaCl, and 0.1% Tween-20 to the recommended Phosphate-Buffered Saline

with Tween-20 (PBS-T). Tween-20 is a mild non-ionic detergent which is usually added in small concentrations (0.05%–0.1%) to reduce nonspecific binding. Increasing the labelled protein concentration proportionally increases the total number of binding sites, which might help drive the binding equilibrium toward a more bound complex. Therefore, the concentration of the labelled protein was increased to 1200 nM, which is 20 times the protein/dye K_D measured during the protein concentration curve assessment. Unfortunately, none of these adjustments provided a well-defined binding interaction curve (Figure 69B-C).

To adjust the ligand concentration, it would have been useful to know the binding affinity of KG-548. However, to date, no precise K_D for the binding of KG-548 to the wild-type ARNT2-PAS-B domain has been published. Existing NMR studies do not provide exact values but rather qualitative insights by showing, similar to ARNT, slow-exchange chemical shift changes, strongly indicating binding of KG-548 upon ARNT2.[159] The only reported K_D ~ 684 µM measurement refers to KG-548 binding to a Y456T mutant of ARNT-PAS-B, but not the wild-type ARNT2 or ARNT-PAS-B.[205]

To cover the Y456T mutant literature concentration range for K_D measurement, a standard 16-point KG-548 dilution series was therefore increased and started from 1000 µM with decreasing stepwise to 25 µM. The concentrations used were: 1000, 900, 800, 700, 600, 550, 500, 450, 400, 350, 300, 250, 200, 100, 50, and 25 µM. This combination of linear and serial dilution was intended to provide good coverage across both high and low KG-548 concentrations. However, the expected binding curve was still not observed (Figure 69D).

Since a His-tag-specific labelling strategy did not provide a measurable binding response with KG-548 control compound, it was decided to transition to a covalent labelling strategy.

Covalent labelling can increase sensitivity and the robustness of the measurements as it provides a stable, non-reversible binding of the fluorophore, eliminating the possibility of dye dissociation during the measurements. Although the His-tag labelling dye should not interfere with the ligand binding as the 6xHis-tag locates further away from the ARNT2 binding site, covalent dye labels several surface-exposed cysteine or lysine residues, minimising the risk of conformational constraints near the ligand binding site.

The cysteine labelling was not appropriate, as two out of three cysteines, such as Cys374 and Cys433, face inside the binding pocket and may influence ligand binding (Figure 70). A covalent labelling approach targeting lysine residues was tested; however, it did not improve the KG-548 binding curve (Figure 69E).

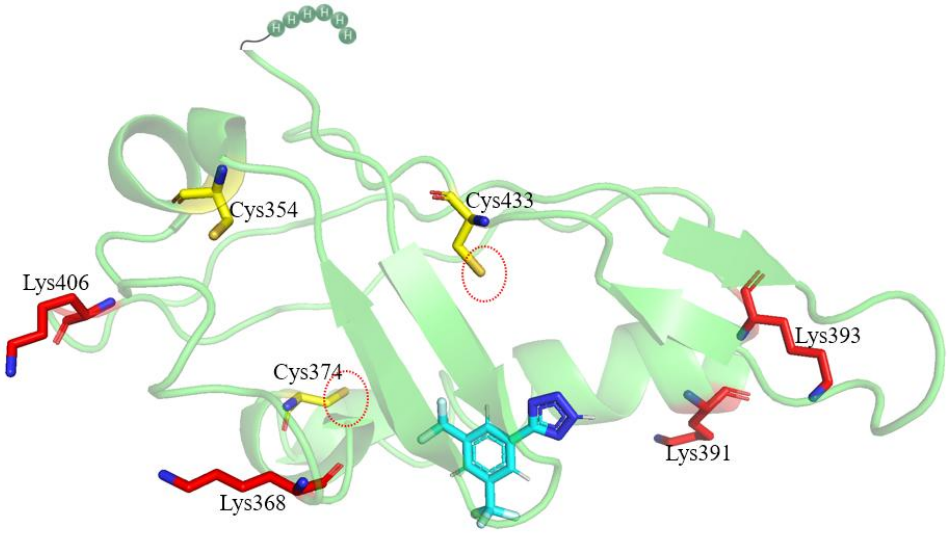


Figure 70. Structural analyses of lysine and cysteine residues inside the PAS-B binding pocket of mARNT2 proteins.In-house KG-548 bound ARNT2 structure was used. Lysine and cysteine residues are shown in red and yellow, respectively. Both 374 and 433 cysteines that faced inside the binding pocket are circled in dashed red. The hexa-histidine (6xHis) expression tag of the uncut-TEV mARNT2-PAS-B recombinant protein, where the fluorescence dye binds, is added separately and shown with six green H-circles.

The switch from the first to the second generation of His-tag-specific RED-Tris-NTA dye, however, showed a slightly better shape of the KG-548 binding curve (Figure 71). The capillary scan showed good fluorescence intensity without variation of absorption. The MST traces confirmed the absence of aggregates and showed no ligand-induced changes in the photobleaching rate, suggesting that the fluorescent label remained stable throughout the MST measurements and was not affected by ligand binding. The SNR was large enough to conclude binding. The estimated measured K_D ~70.9 μM has a quite wide ±57.9 μM confidence interval, suggesting the true K_D could lie between ~13 μM and 129 μM, indicating moderate binding affinity between the PAS-B domain of ARNT2. However, it is important to note that KG-548 data was not always consistent, suggesting that the MST assay might not be ideal for measuring its affinity, or other weak binders.

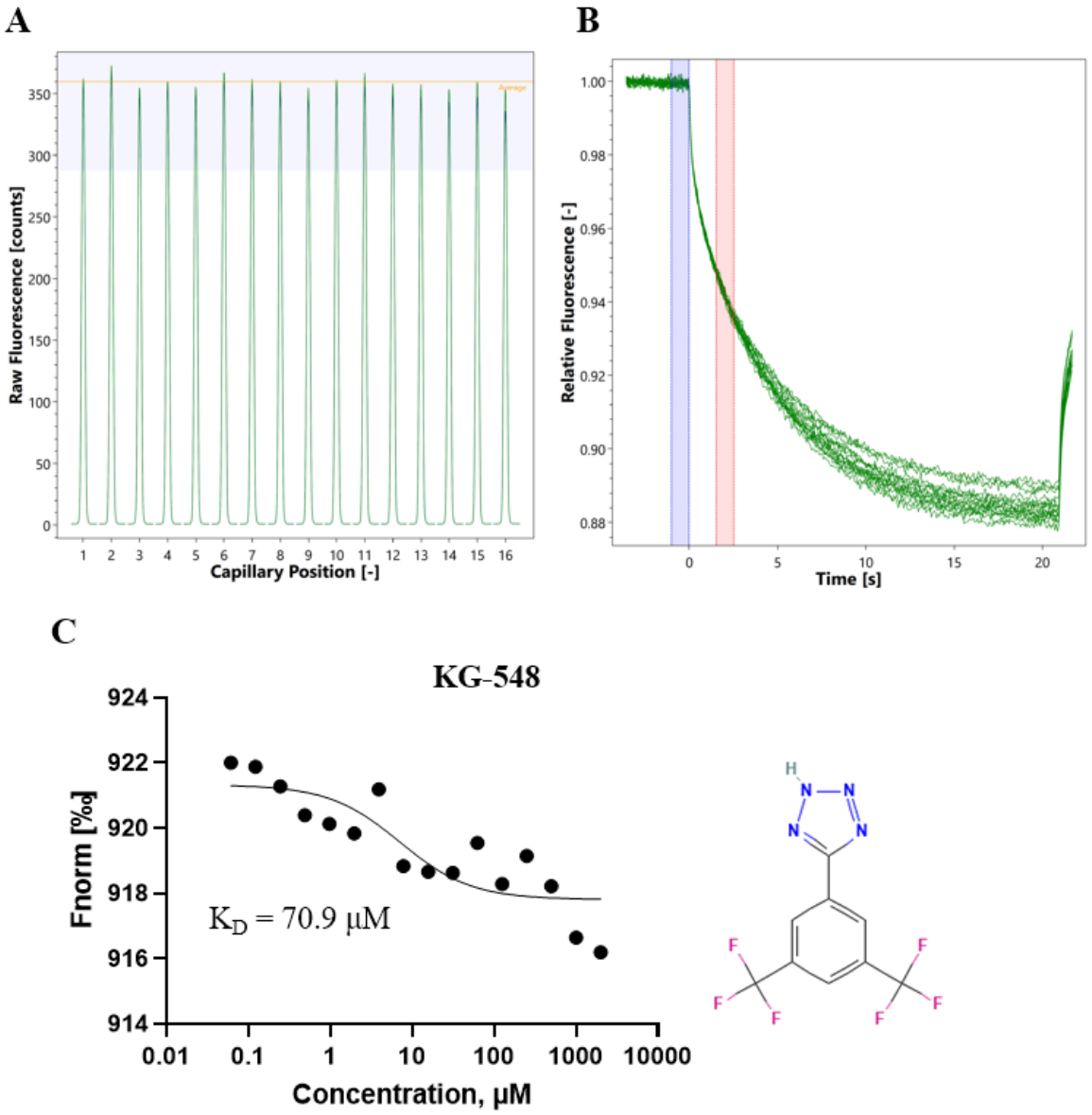
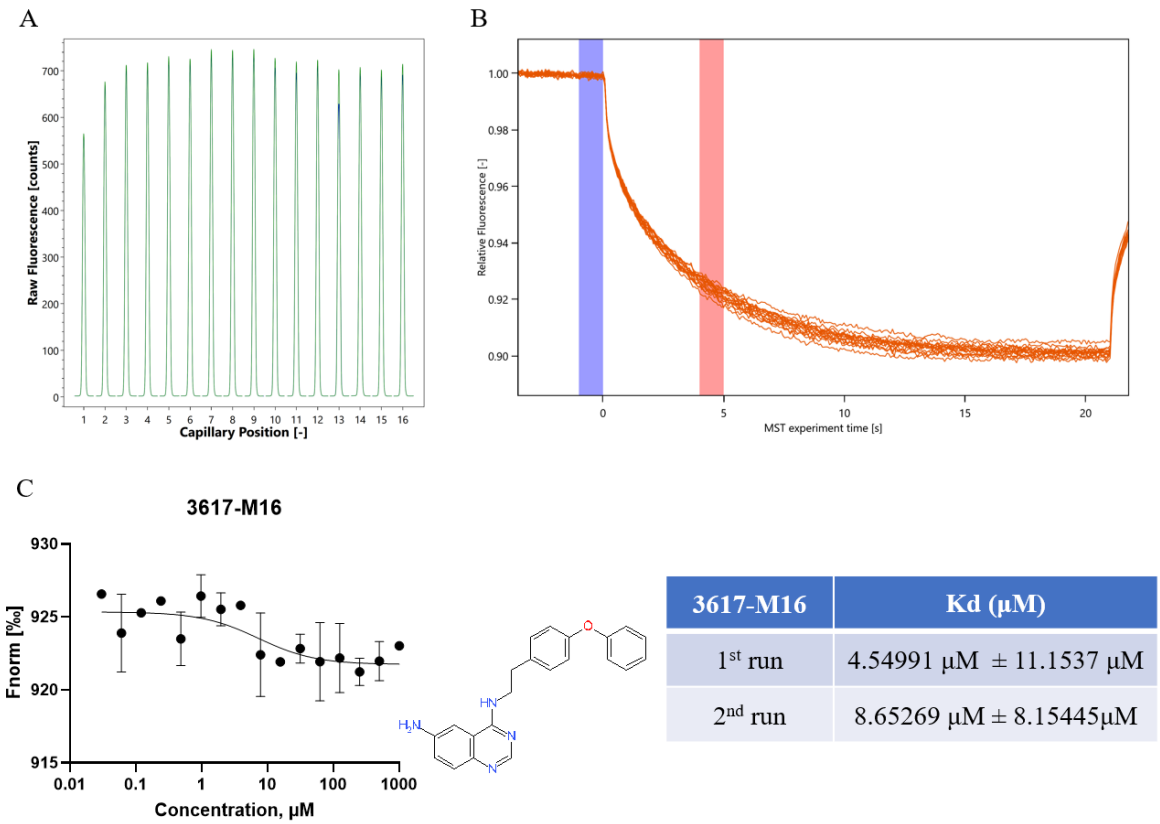


Figure 71. MST dose-response curve for KG-548 against mARNT2-PAS-B.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The concentration of KG-548 decreases from left to right. (B) MST traces. (C) Plot and binding curve fit for the intensity changes.

7.3 MST binding assay in dose-response format

Given that MST is a relatively swift and easy assay, and despite the inconsistency of KG-548 control data, it was decided to still test all compounds for MST binding affinity. As the second generation of RED-Tris-NTA dye showed slightly better results with KG-548 control compound, this labelling strategy was employed to evaluate other tested ligands. Protein concentration binding curves, which were re-evaluated with the second-generation RED-Tris-NTA dye, showed similar K_D values. Therefore, 250 nM of mARNT2-PAS-B and hNPAS3-ARNT2, and 125 nM of hNPAS1-ARNT2 were used to fluorescently label proteins. Depending on solubility properties, almost all compounds were re-titrated from 1,000 μM to 30 nM for 16-point dilution and incubated with fluorescently labelled proteins.

Among all ARNT2 tested compounds, only recently crystallised 3617-M16, and several other ligands like 8-I21, 3621-L6, and 3620-L22 showed prospective ligand binding curves, although without a clear sigmoidal response. Similar to the KG-548, the 3617-M16, 8-I21, and 3620-L22 showed strong initial fluorescence, and capillary performance was generally consistent (Figure 72-74A). The MST traces across the series indicated that no aggregation, photobleaching, or surface adsorption effects were present (Figure 72-74B). The overall change in the normalised fluorescence (F_{norm}) values across ligand concentration series showed modest variability, ranging from ~921 to ~935 ‰ for 3620-L22, or approximately ~7‰ difference for both 3617-M16 and 8-I21 (Figure 72-74C, D). The relatively low amplitude of signal change (<15 ‰) usually falls near the threshold of typical MST detection sensitivity. A subtle increase in fluorescence signal was still observed with these ligands, suggesting a weak or low-occupancy binding interaction with the compounds. The lack of a clear sigmoidal curve limits confidence in defining a precise K_D based solely on these results.

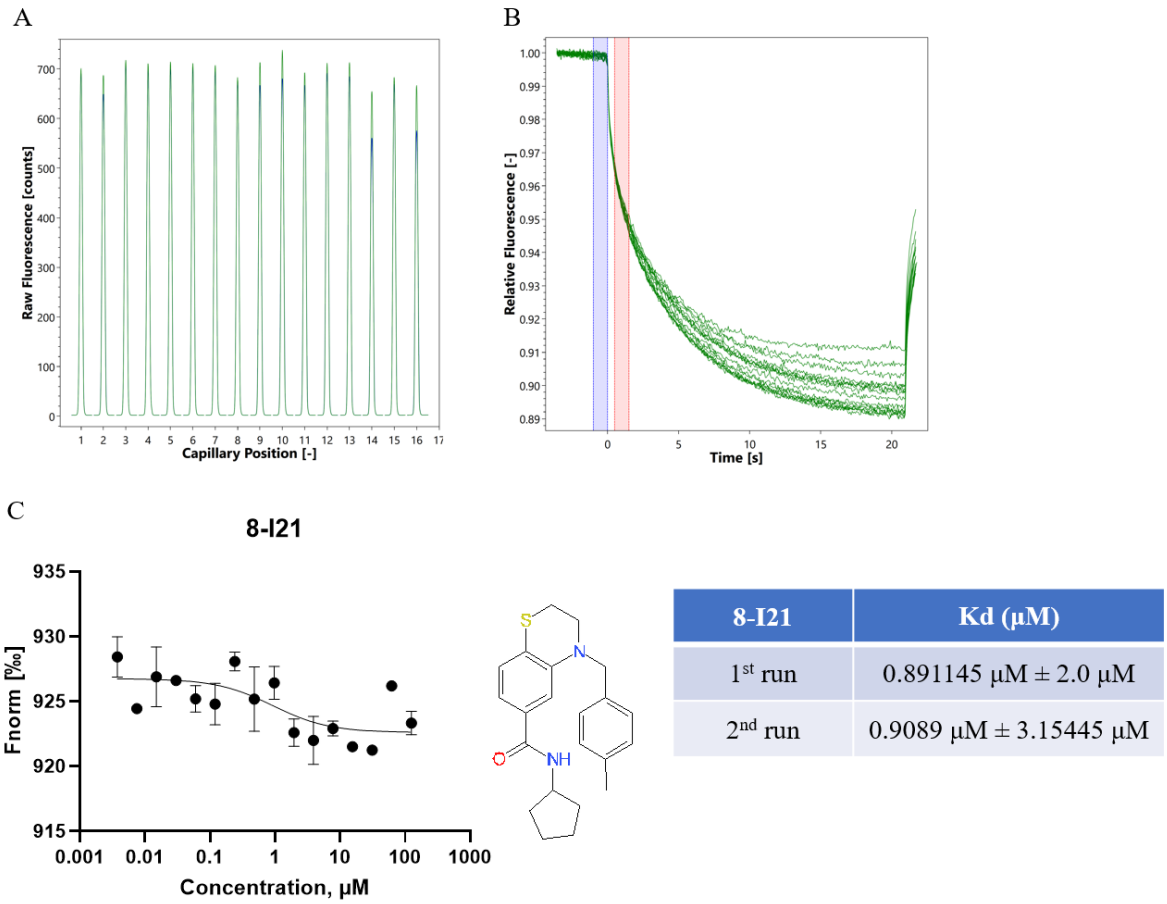


3617-M16	Kd (µM)
1st run	4.54991 µM ± 11.1537 µM
2nd run	8.65269 µM ± 8.15445µM

Figure 72. MST dose-response curve for 3617-M16 against mARNT2-PAS-B.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The concentration of 3617-M16 decreases from left to right. (B) MST traces. (C) Plot and binding curve fit for the intensity changes, n=2. Approximate K_D measurements are shown in the blue table.

Additionally, the hot measurement window is different across compounds. Ideally, all F_{hot} measurements should have been taken from the steady state region at the same time for all compounds to allow direct quantitative comparison. Variation in the hot window across compounds could occur for several reasons, including differences in signal quality, binding kinetics, photobleaching, or aggregation behaviour. Sometimes, such adjustments are made to select a time range that best captures the stable portion of the thermophoretic signal while

avoiding artefacts such as early signal noise, delayed drift, or compound-induced fluorescence changes.



8-I21	Kd (μM)
1st run	0.891145 μM ± 2.0 μM
2nd run	0.9089 μM ± 3.15445 μM

Figure 73. MST dose-response curve for 8-I21 against mARNT2-PAS-B.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The concentration of 8-I21 decreases from left to right. (B) MST traces. (C) Plot and binding curve fit for the intensity changes, n=2. Approximate K_D measurements are shown in the blue table.

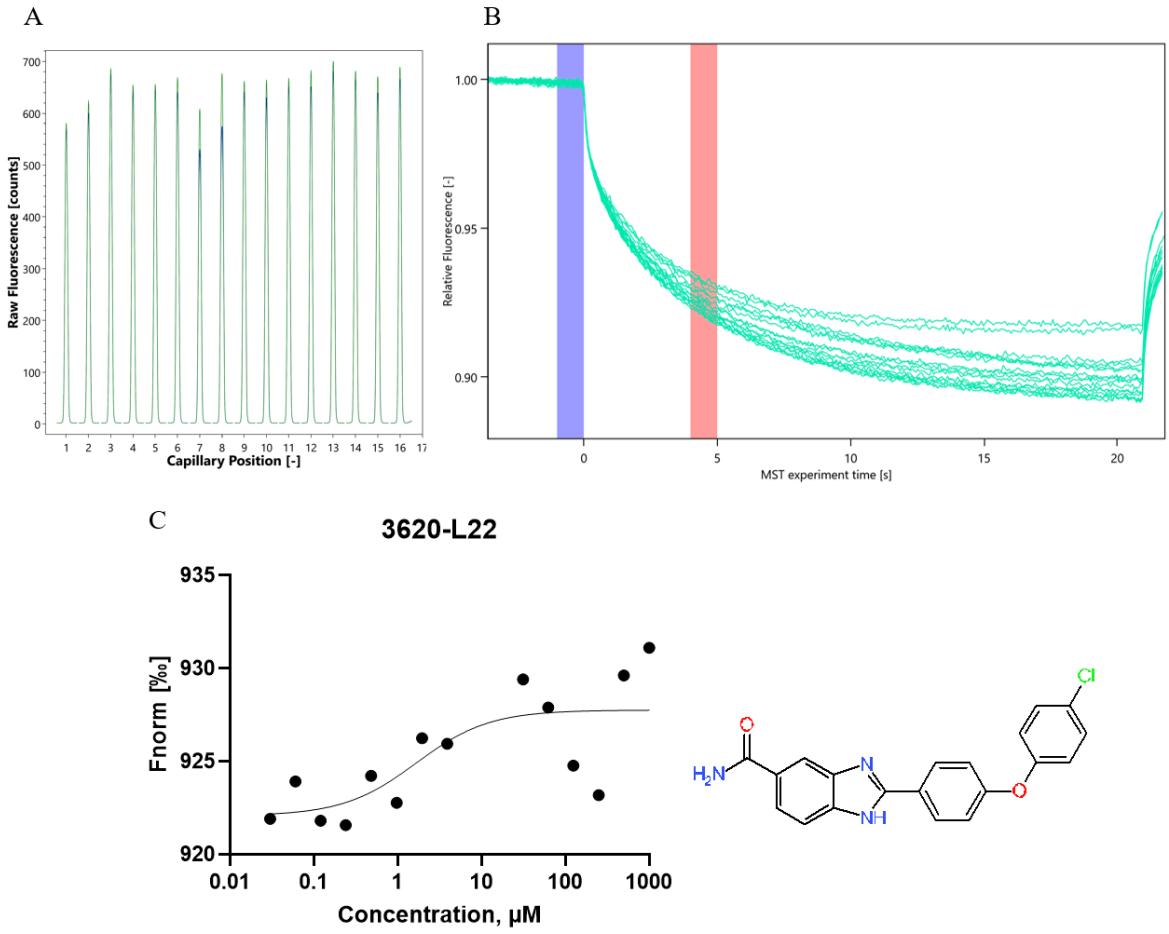


Figure 74. MST dose-response curve for 3620-L22 against mARNT2-PAS-B.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The concentration of 3620-L22 decreases from left to right. (B) MST traces. (C) Plot and binding curve fit for the intensity changes.

Interestingly, the binding curve of 3621-L6 demonstrates a sigmoidal profile with a defined transition zone (Figure 75D). The F_{norm} values for 3621-L6 showed a pronounced and consistent decrease from approximately 980 ‰ at the highest ligand concentration to nearly 920 ‰ at the lowest concentrations. This amplitude, spanning nearly 60 ‰, indicates a clear and concentration-dependent thermophoretic response, suggesting a specific molecular interaction. It should be noted, however, that the initial fluorescence capillaries scan for

3621-L6 appeared to decrease with increasing ligand concentration, a factor that potentially contributed to the analyses of the ligand (Figure 75A).

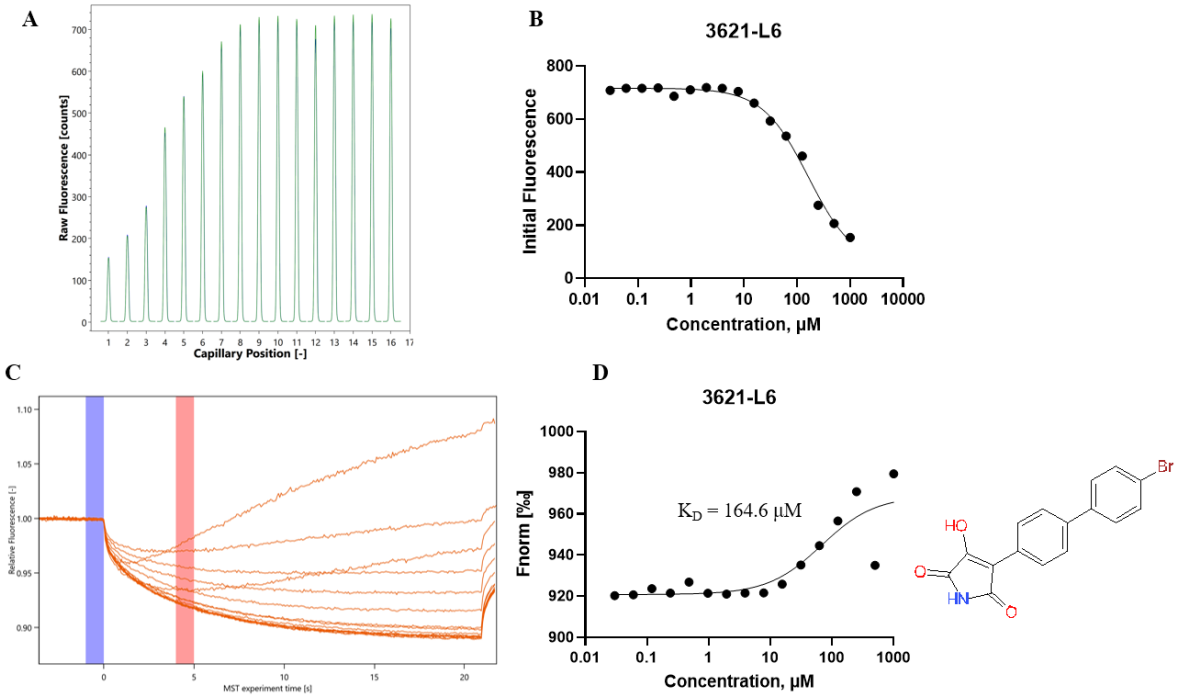


Figure 75. MST dose-response curve for 3621-L6 against mARNT2-PAS-B.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The concentration of 3621-L6 decreases from left to right. (B) Initial fluorescence was plotted against ligand concentration. (C) MST traces. (D) Plot and binding curve fit for the intensity changes.

This effect is frequently observed when proteins adhere to the surface of the plastic vessels employed during assay preparation, which leads to a reduced protein concentration in the solution that is subsequently transferred to the capillary. To prevent this phenomenon, Monolith's premium capillaries were always used to evaluate all the compounds. Apart from having a higher surface area and more uniform pore structure, leading to better flow and analyte separation, premium capillaries are usually coated with materials that make the inner

surface hydrophilic to reduce hydrophobic interactions that often cause proteins to stick. However, this effect was seen only with 3621-L6, suggesting that the compound itself can also affect the extent of these surface-bound aggregations.

MST experimental traces are considered compatible only when the initial fluorescence values across the titration series differ by less than 10%. When traditional MST traces are unreliable, it is recommended to plot the initial fluorescence intensity against the ligand concentration to ensure that the observed changes in the MST signal are truly due to molecular interactions and not caused by other factors. This approach allowed for a dose-response analysis, as it can still unveil the binding pattern and estimates of binding affinity. The plotted dose-response against 3621-L6 seems relevant and is considered a result of binding (Figure 75B). This data further supports the conclusion that the apparent K_D of 164.6 μM is not a result of non-specific binding, but rather that the 3621-L6 binds the mARNT2-PAS-B with moderate affinity. Strikingly, the measured K_D of 164.6 μM perfectly matches the predicted EC_{50} of 148 μM calculated via TSA (Table 5).

Interestingly, the recently crystallised 50-E18 ligand did not show any binding affinity via MST (Appendix 6). This could be explained by crystallography providing a protein-ligand complex under conditions that usually stabilise binding. MST, on the other hand, measures binding in solution, where interacting partners move freely and might not adopt the same conformation or interaction as seen in the crystal.

Among all NPAS1/3 tested compounds, only 41-B13 demonstrated a sigmoidal profile with a defined transition zone. Only one capillary showed some aggregation, which was later removed from the analyses (Figure 76A). The smooth and reproducible nature of the signal

across the rest of the titration points indicates that no other capillary adsorption, fluorescence quenching, or significant aggregation occurred under the experimental conditions.

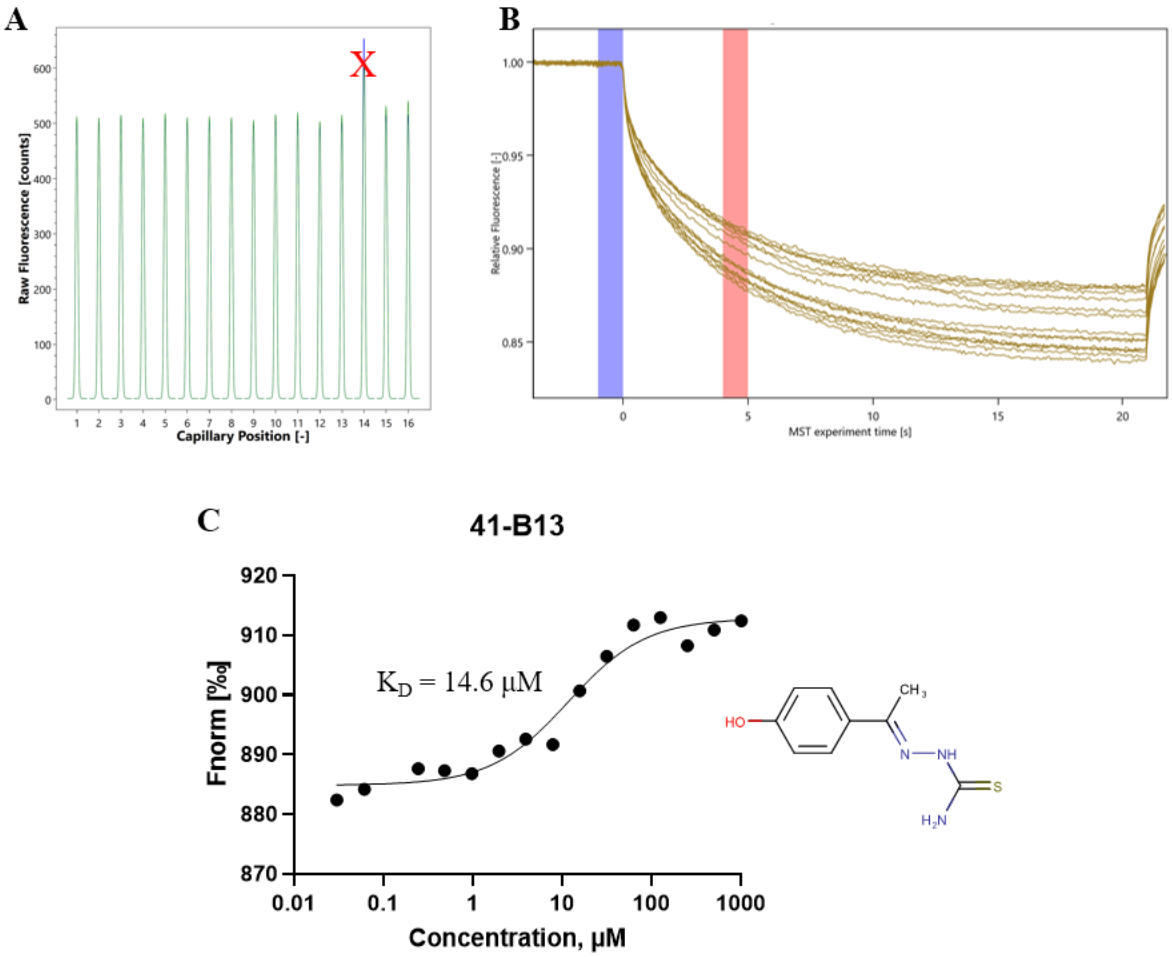


Figure 76. MST dose-response curve for 41-B13 against hNPAS3-ARNT2.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The 14th capillary that displayed aggregation and was later removed from further analyses is shown with a red cross. (B) MST traces. (C) Plot and binding curve fit for the intensity changes.

The F_{norm} values showed a pronounced and consistent decrease from approximately 979 ‰ at the highest ligand concentration to around 920 ‰ at the lowest concentrations. The

magnitude of 60 ‰ F_{norm} change, and the sharpness of the transition indicate good binding specificity and signal-to-noise performance in the assay. The measured K_D of approximately 14.6 μM, however, has a wide K_D confidence interval of ±145 μM. When the error is much larger than the K_D itself, this usually suggests high uncertainty in the measurement, a very weak binding interactions, or perhaps even poor data quality.

Another compound, 3623-J16, also exhibited a somewhat sigmoidal binding profile (Figure 77). However, the three lowest-concentration capillaries showed anomalously low signals, suggesting that these data points may be unreliable and are likely distorting the overall fit. The MST data nonetheless display a concentration-dependent change in F_{norm}, indicative of some degree of interaction between the compound and the target protein. Fitting the data to a one-to-one binding model yielded an apparent K_D of approximately 0.45 μM, which nominally corresponds to moderate-affinity binding in the submicromolar range.

Despite this, the reliability of the result is questionable. The upper plateau (low-concentration region) is poorly defined due to outlier capillaries, and the high-concentration baseline shows fluctuations that increase the uncertainty of the fit (± 0.23 μM, nearly half the calculated K_D). These inconsistencies likely reflect experimental artifacts such as incomplete equilibration, or compound aggregation at higher concentrations. Therefore, while a sigmoidal response is observed, the overall dataset does not provide robust evidence for specific binding, and the compound should be considered a weak or unreliable binder under the tested conditions.

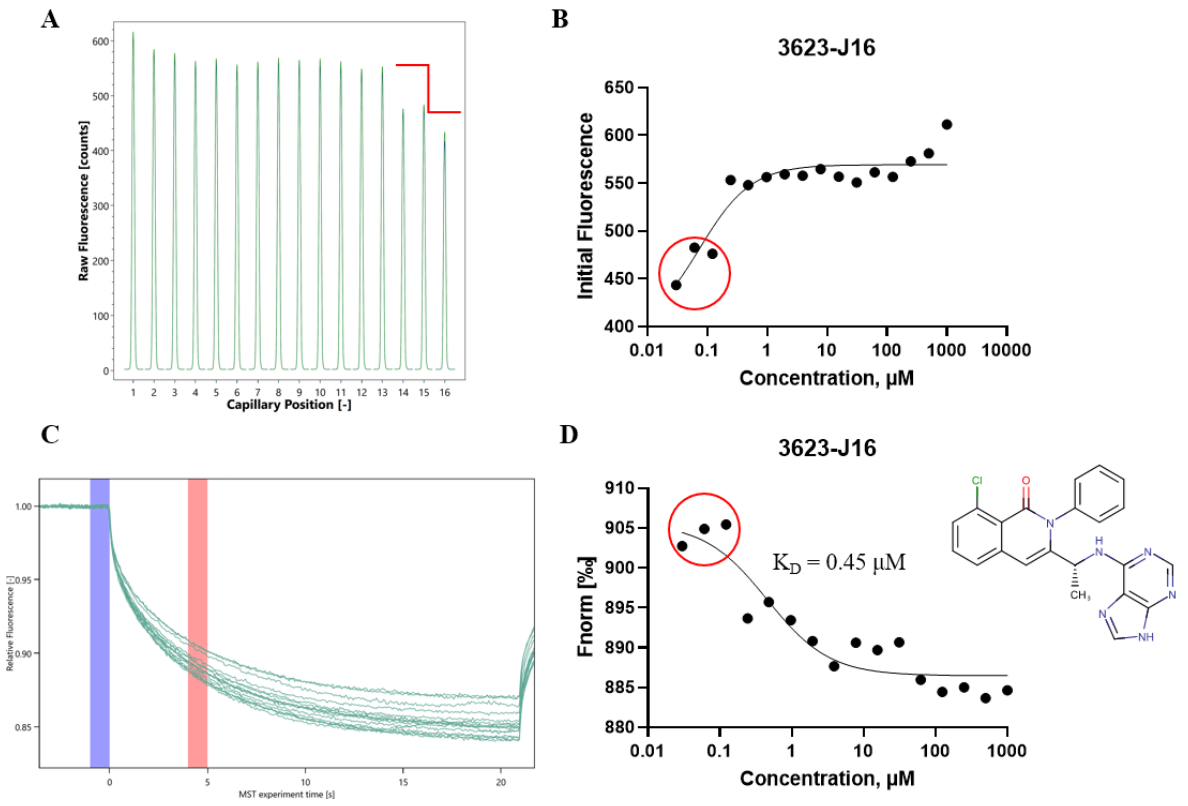


Figure 77. MST dose-response curve for 3623-J16 against mARNT2-PAS-B.Protein was labelled with the second generation of RED-Tris-NTA dye. (A) Fluorescence scan of the capillaries used. The concentration of 3623-J16 decreases from left to right. (B) Initial fluorescence was plotted against ligand concentration. (C) MST traces. (D) Plot and binding curve fit for the intensity changes.

The rest of the NPAS1/3 tested compounds did not show a monotonic sigmoidal response (Appendix 7-8). The absence of a clear, concentration-dependent shift in F_{norm} may indicate either a lack of binding affinity between the interacting biomolecules or that no specific binding interaction could be reliably detected under the tested conditions.

To conclude, MST is a powerful tool, but only when it comes to analysing biomolecule interactions with affinities in the nano- to picomolar concentration range.[206] Weaker

interactions may not achieve the saturation point within the tested ligand concentrations and eventually can be missed. MST relies on fluorescence-based detection, which is susceptible to interference from photobleaching, fluorescence quenching, and compound-related fluorescence artefacts. These MST effects might distort the fluorescent signal, leading to inconclusive or false-negative results.[207]

Furthermore, MST may also hinder the detection of a clear binding event as the thermophoretic signal change is too small.[208] Another limitation of the method is that MST results may be highly affected by the buffer or assay conditions, such as ionic strength, pH, and additives, which may influence protein conformation and ligand solubility.[207] Considering these limitations of the MST assay, another biophysical method with greater sensitivity was used to further quantity and investigate the strength of binding affinity of the tested compounds.

Chapter 8: Surface plasmon resonance

8.1 Introduction to the surface plasmon resonance assay

The binding affinity of the tested compounds was further studied with another biophysical technique: SPR. This assay is a label-free, powerful approach to monitor molecular interactions in real time and to measure ligand kinetics, and is used extensively in early-phase drug discovery.[209] The fundamental principle underlying SPR is based on the propagation of electromagnetic waves (plasmons) at the interface between a typically gold-coated sensor surface and an aqueous sample medium.[210]

In a standard SPR, a beam of light is passed through a prism at a gold-coated surface of a glass sensor chip (Figure 78). The light is reflected off the gold chip surface into a detector at a specific incident angle, known as the resonant angle. Some of the light energy is absorbed by the free electrons in the sensor chip surface, exciting charge density oscillations, known as surface plasmons, that propagate along the metal-dielectric interface. As plasmon resonance occurs, some of the light energy is absorbed, reducing the intensity of the reflected light. This can be detected as a dip in the SPR reflection intensity curve. The binding of biomolecules or dissociation occurring near the surface results in changes in the refractive index at the surface of the sensor chip. When a target protein is immobilised on the sensor chip and a tested compound binds to it, the interaction results in a mass increase at the sensor surface. This mass increase leads to a proportional rise in the refractive index, which is observed as a change in the SPR signal. This change is monitored and plotted over time, generating a response curve known as a sensorgram.

The SPR phenomenon is thus highly sensitive to even minimal perturbations, enabling the quantitative analysis of interactions between molecules. Such interactions could occur between proteins, proteins and small molecules, and proteins and nucleic acids. One of the key advantages of SPR is its ability to generate binding results using only submicromolar quantities of protein, making it particularly valuable for studying proteins that are produced in small yields, like the NPAS1-ARNT2 heterodimer. The technique, however, relies on the immobilisation of one interaction partner, typically the protein, onto the sensor chip surface, which can necessitate optimisation for each specific system to preserve biological activity. Despite this limitation, SPR has become the gold standard in molecular interaction analysis due to its sensitivity, versatility, and broad applicability across various types of biomolecular systems.

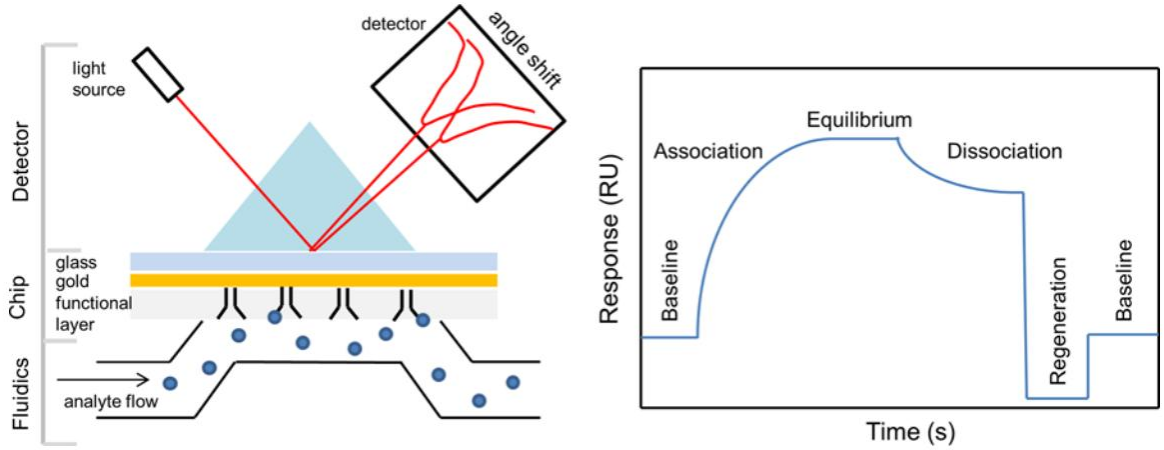


Figure 78. Illustrative diagram of Surface Plasmon Resonance.Polarised light from a monochromatic light source is directed through a prism before reaching the surface of a gold-coated sensor chip. As the light undergoes total internal reflection at a specific angle, known as the resonance angle, it excites surface plasmons at the metal-dielectric interface. This results in a characteristic dip in the intensity of the reflected light, which is measured by a detector. When a binding event occurs on the sensor surface, such as the interaction between an immobilised ligand and an analyte, the associated increase in mass leads to a change in

the local refractive index. This change causes a shift in the resonance angle, reflecting the binding event. The corresponding variation in signal is recorded over time and displayed as a sensorgram. Figure adapted from Damborsky et al., 2016.[211]

8.2 Development of the SPR assay

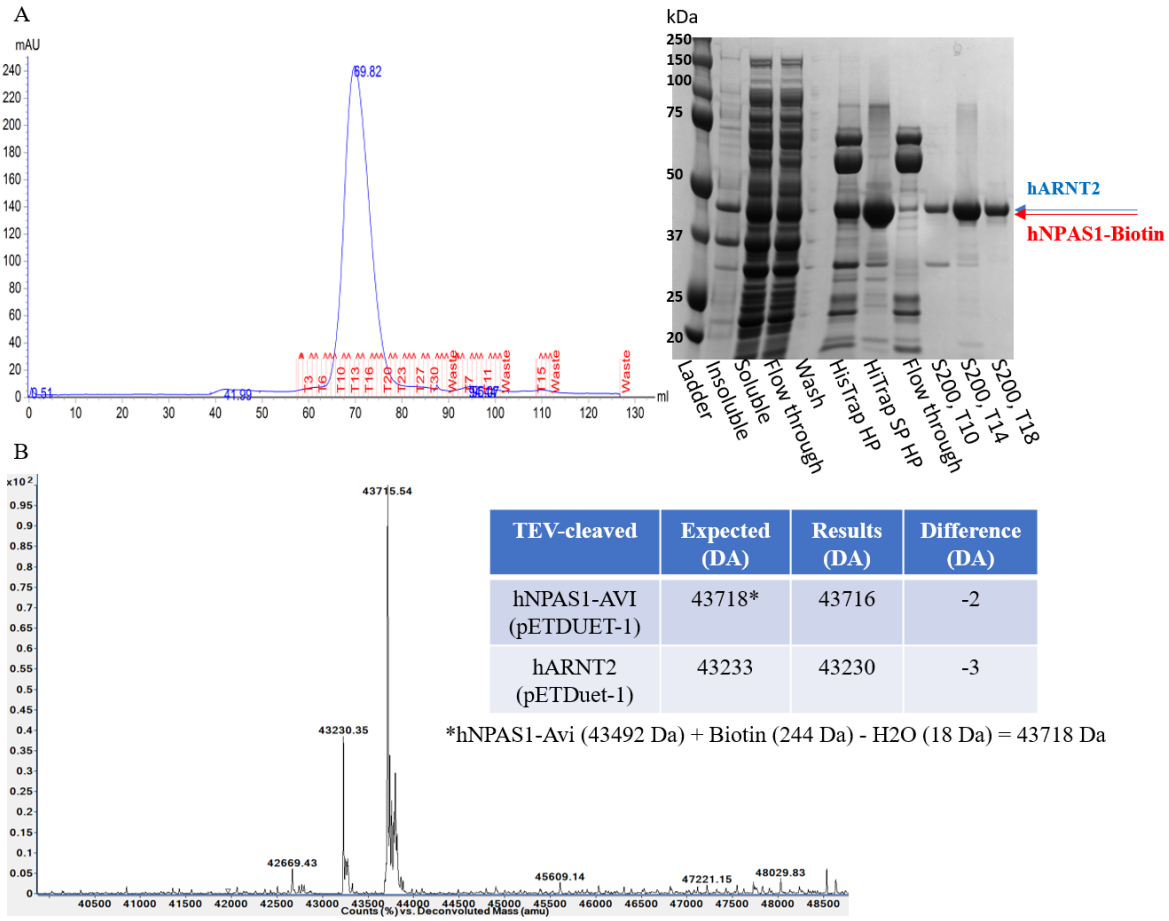
To enable efficient and stable immobilisation of a target protein on the sensor surface for SPR, biotinylation of a target protein is usually employed. Biotin is a small molecule (MW 244.3) found to be present at very low levels in all living cells and is crucial for many biological processes. In SPR, biotinylation is carried out to facilitate specific and high-affinity binding to streptavidin, which has been pre-immobilised on the surface of commercially available streptavidin-coated sensor chips. Streptavidin is a biotin-binding protein isolated from *Streptomyces avidinii*; it is similar in size and affinity to biotin. Streptavidin-biotin binding is the strongest known non-covalent interaction ($K_D \sim 10^{-14}$ M).[212] Under standard flow conditions, this binding occurs rapidly and irreversibly, resulting in a uniform and robust protein layer suitable for subsequent accurate and reliable measurements of protein interactions.

Biotinylation of a protein of interest is usually achieved through the incorporation of a specific biotinylation Avi-tag sequence, Gly-Leu-Asn-Asp-Ile-Phe-Glu-Ala-Gln-Lys-Ile-Glu-Trp-His-Glu. This 15-amino acid peptide sequence serves as a specific recognition site for biotinylation by the *E. coli* biotin ligase, BirA. When genetically fused to either the N- or C-terminus of the protein of interest, the Avi tag enables enzymatic covalent attachment of a single biotin molecule to a single lysine residue within the tag. This site-specific biotinylation strategy offers significant advantages over non-specific chemical methods, ensuring a

homogeneous modified protein population and preservation of the functional integrity of the protein. The resulting biotinylated target protein exhibits firm and oriented attachment to streptavidin-coated surfaces.

To generate biotinylated NPAS1 and NPAS3 proteins, Avi-tag was added to the C-terminus of both proteins within pETDuet-1-hNPAS1/3-ARNT2 constructs. To achieve *in vivo* biotinylation, a pCDF-BirA plasmid was co-expressed with both proteins in the presence of biotin. Proteins were purified and analysed by mass spectrometry to detect the addition of biotin (Figure 79, 80). When analysing biotinylated proteins with mass spectrometry, the expected mass should increase by +226 Da due to the addition of a biotin +244 Da as well as elimination of a water molecule -18 Da.[213] The NPAS1 and NPAS3 proteins provided the exact expected mass for biotinylated constructs together with their ARNT2 partners (Figure 79, 80).

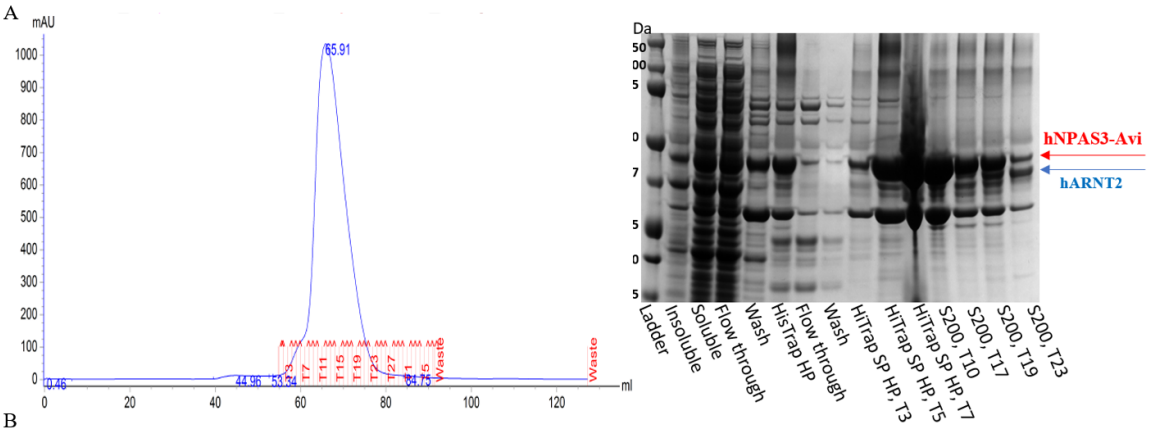
Biotinylation of NPAS1 increased its apparent molecular weight, causing its band to migrate very close to that of ARNT2 on SDS-PAGE. As a result, the two bands are difficult to distinguish due to the overlap. In the biotinylated NPAS3 sample, apart from the ARNT2 protein band, there is another band corresponding to the mass of approximately 35 kDa. Mass spectrometry confirmed the presence of another protein peak at 35.4 kDa. Further purification with additional ion-exchange and size exclusion chromatography failed to remove this protein band. It is known that endogenous biotinylated proteins in *E. coli* often appear around 35 kDa on SDS-PAGE gels.[214] This is primarily due to the BirA biotin ligase, which is a 35-kDa protein catalyses biotinylation. BirA, which attaches biotin to the biotin carboxyl carrier protein (BCCP), is likely the source of the observed additional 35 kDa band.

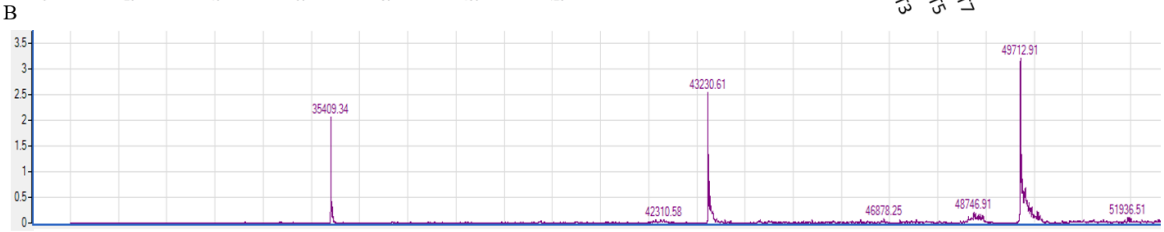


TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hNPAS1-AVI (pETDUET-1)	43718*	43716	-2
hARNT2 (pETDuet-1)	43233	43230	-3

*hNPAS1-Avi (43492 Da) + Biotin (244 Da) - H2O (18 Da) = 43718 Da

Figure 79. Purification and quality control of biotinylated NPAS1-ARNT2.(A) HiLoad 16/600 Superdex 200-pg size exclusion column UV traces showed one protein peak corresponding to the heterodimer elution. SDS-PAGE confirmed the purity of the protein complex, illustrating one protein band, corresponding to the overlap of biotinylated NPAS1 protein and its partner ARNT2. The molecular weight of both proteins is approximately 43 kDa. (B) Deconvoluted spectra and intact mass analysis of TEV-cleaved biotinylated NPAS1-ARNT2. Mass spectrometry confirmed the presence of two proteins with the expected masses of biotinylated NPAS1 and its ARNT2 partner.





TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)
hNPAS3-AVI (pETDUET-1)	49715*	49713	-2
hARNT2 (pETDuet-1)	43233	43231	-2

*hNPAS3-Avi (49489.61Da) + Biotin (244 Da) - H2O (18 Da) = 49715.61 Da

Figure 80. Purification of biotinylated NPAS3-ARNT2.(A) HiLoad 16/600 Superdex 200-pg size exclusion column UV traces showed one protein peak corresponding to the heterodimer elution. SDS-PAGE, however, showed three protein bands, corresponding to biotinylated NPAS3 protein, its ARNT2 partner, and an additional band at approximately 35 kDa. (B) Deconvoluted spectra and intact mass analysis of TEV-cleaved biotinylated NPAS3-ARNT2. Mass spectrometry confirmed the biotinylation of NPAS3 protein. Apart from the ARNT2 partner, it also showed the presence of additional protein at 35.4 kDa, which is more likely to be a BirA biotin ligase.

8.3 Immobilisation of proteins on the SA chip and optimisation of the SPR assay

Purified NPAS1/3-ARNT2 and in-house ARNT2-PAS-B biotinylated proteins were immobilised onto the streptavidin (SA) sensor chip. SA is a general-purpose sensor chip with pre-immobilised streptavidin for the capture of biotinylated ligands. The SA chip provides excellent performance in a broad range of applications due to its high affinity and binding capacity, chemical resistance, and results reproducibility.

The immobilisation of proteins was monitored in real time. Upon injection, a gradual increase of approximately 2,000 response units (RU) was observed, reflecting the high-affinity and near-instantaneous binding of the biotinylated proteins to the immobilised streptavidin (Figure 81). After washing, the binding curve remained stable and flat, indicating the stability of the heterodimer. Considering the SPR binding equation, immobilisation of 2,000 RU for a six-domain 100 kDa heterodimer will provide a theoretical maximum 6 RU value (RUmax) upon binding on average 300 Da small molecules. With such a low signal, it would be challenging to distinguish a real binding event from the background noise. Therefore, additional protein was added until it achieved stabilisation at the immobilisation level of approximately 8,000 RU, which falls within the optimal range for downstream kinetic measurements while avoiding mass transport limitations.

The stable baseline did not show any drift or significant fluctuations, demonstrating efficient capture and retention of the protein. In addition, the background noise remained negligible, which will not interfere with data interpretation. Depending on the exact size of a molecule,

the RUmax should be approximately 24 RU for both heterodimers and 140 RU for a single PAS-B domain.

$$\mathbf{RU_{max} = RL \times \frac{MWA}{MWL} \times n}$$

RU_{max} = Maximum theoretical response in response units (RU)

RL = Amount of immobilised ligand or protein on the surface (in RU)

MWA = Molecular weight of the analyte (Da)

MWL = Molecular weight of the ligand or protein (Da)

n = Stoichiometry of binding (number of analyte molecules bound per ligand molecule)

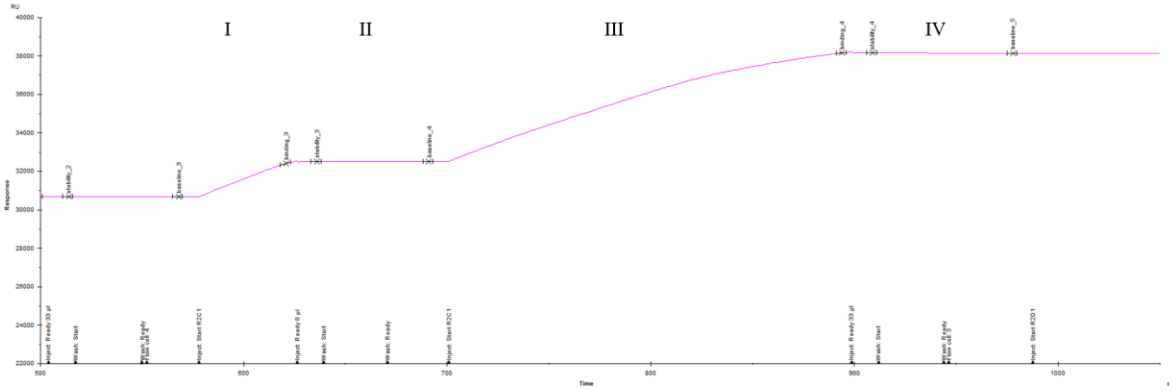


Figure 81. Representative example of a real-time dynamic process of biotinylated NPAS1-ARNT2 protein immobilisation on the surface of the SA chip.The process demonstrated four characteristic phases: (I) Initial immobilisation of NPAS1 was confirmed by an approximately 2,000 RU increase. (II) The binding curve remained stable and flat after washing with loading buffer. (III) Additional protein was added to the chip until the final immobilisation level reached 8,000 RU. (IV) After washing with the assay buffer, the immobilisation was highly stable, with minimal baseline drift, confirming the irreversible nature of the biotin-streptavidin interaction.

As all tested compounds were dissolved in DMSO, it was essential to account for DMSO-related artefacts in the SPR measurements. DMSO has a significantly higher refractive index than aqueous buffers, and even small variability in concentration (as little as 0.1%) can cause artificial shifts in RU that mimic true binding events. This effect is particularly crucial for studies involving small molecule-protein interactions, where signal changes are often subtle and easily obscured by solvent-induced refractive index changes.

To address this, several DMSO volumes were tested with each of the immobilised proteins. The DMSO tolerance is 2.5 µL as it does not show a significant signal after subtracting the background (Figure 82A). For the SPR, only 0.48 µL of compound DMSO stocks were used, which is a substantially lower volume than the DMSO tolerance. As a negative control, a blank injection containing only DMSO without analyte was used in each SPR analysis. As expected, DMSO did not show any binding effect with any of the three proteins, NPAS1/3 or ARNT2-PAS-B (Figure 82B). To adjust for RU shifts resulting from minor DMSO concentration variations across samples, a solvent correction curve was generated using DMSO concentrations ranging from 0% to 0.8% in 0.1% increments.

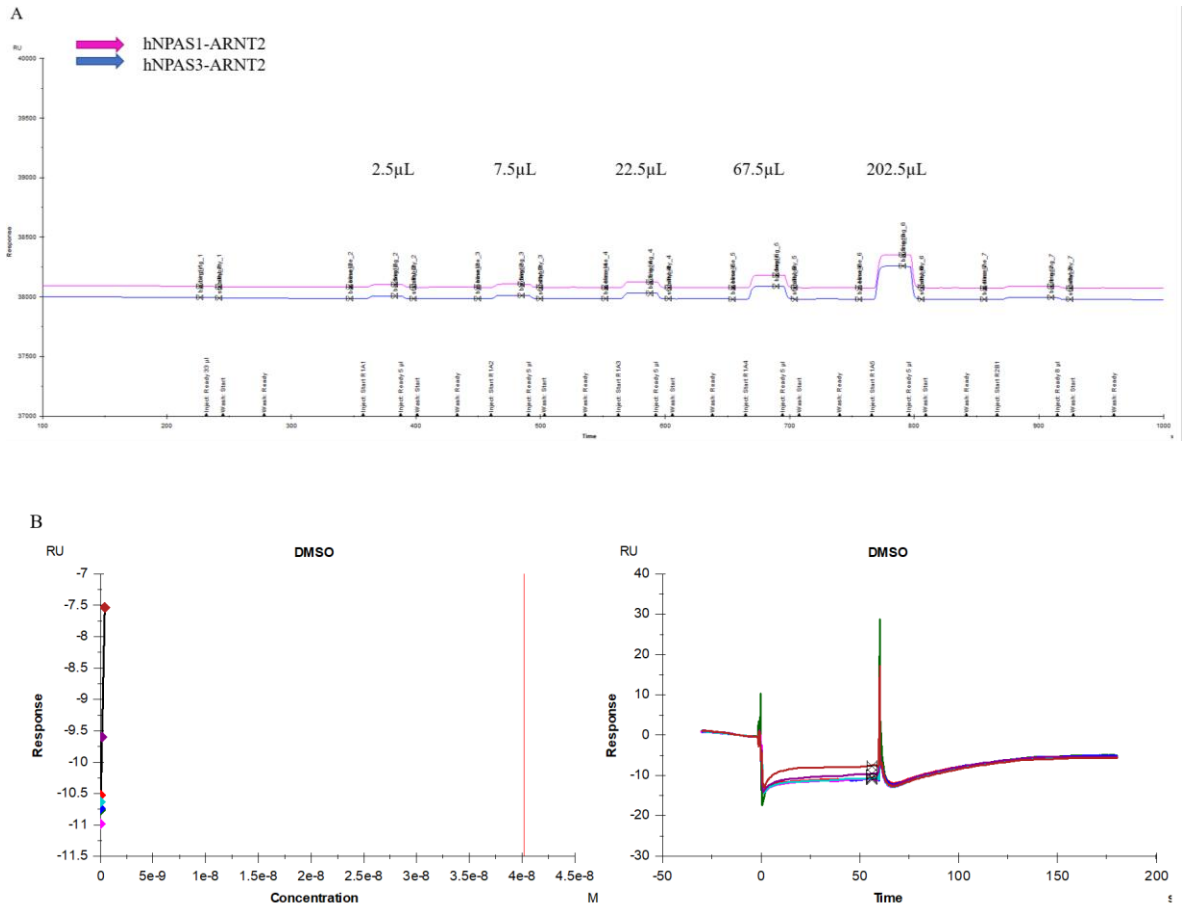


Figure 82. Evaluation of DMSO via SPR.(A) Different volumes of DMSO were measured with each protein. DMSO tolerance is 2.5 µL. (B) To ensure the specificity and reliability of SPR data, a DMSO-only injection (without analyte) was tested in each SPR test as a negative control.

8.4 The SPR competition assay in dose-response format

After successfully loading ARNT2-PAS-B and both NPAS1/3 heterodimers onto streptavidin chips, immobilised biosensors were then treated with dose-response format titrations of the tested compounds. To obtain the binding curve, all compound titrations were referenced against an untreated biosensor. The reference data was later subtracted from the obtained results. To gain real-time signal curves of the binding of all tested compounds, to

fit kinetics or affinity curves, and to calculate K_D values, the data were processed using Biacore S200 result analysis software (Biacore S200 Evaluation Software, V3.2).

As KG-548 could serve only as a control for ARNT2-PAS-B, it was only used with ARNT2-PAS-B tested compounds. KG-548 showed weak binding to the single PAS-B domain of ARNT2, as the binding effect had not reached the saturation point (Figure 83). Due to the weak binding interaction, poor fit of the 1:1 kinetic model, and relatively rapid association and dissociation rates, a steady-state affinity analysis was employed to evaluate the KG-548 data. The affinity found via this method was quite poor, with an approximate K_D ~938 µM. As there are no published data reporting K_D values for the binding of KG-548 to the ARNT2 protein, it is challenging to properly validate the assay. This measured K_D is somewhat in the range of the previously reported affinity of KG-548 to the Y456T mutant of ARNT-PAS-B, with a calculated K_D ~684 µM.[205]

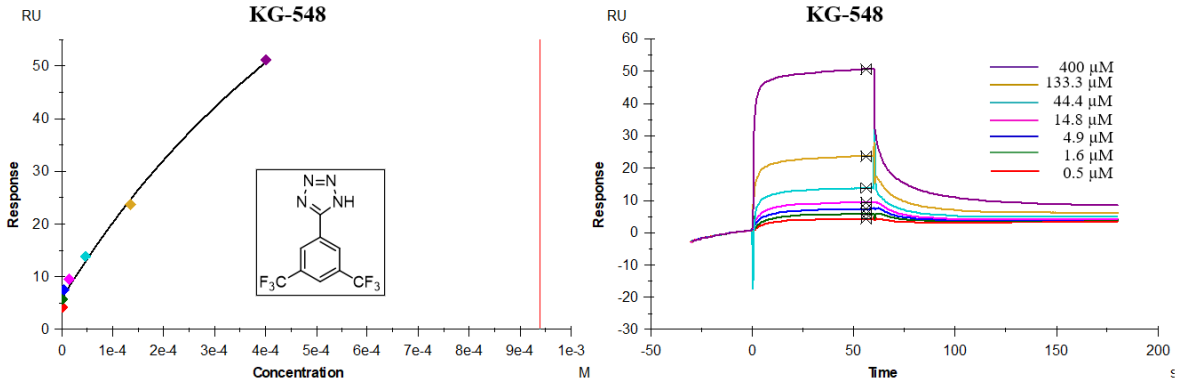
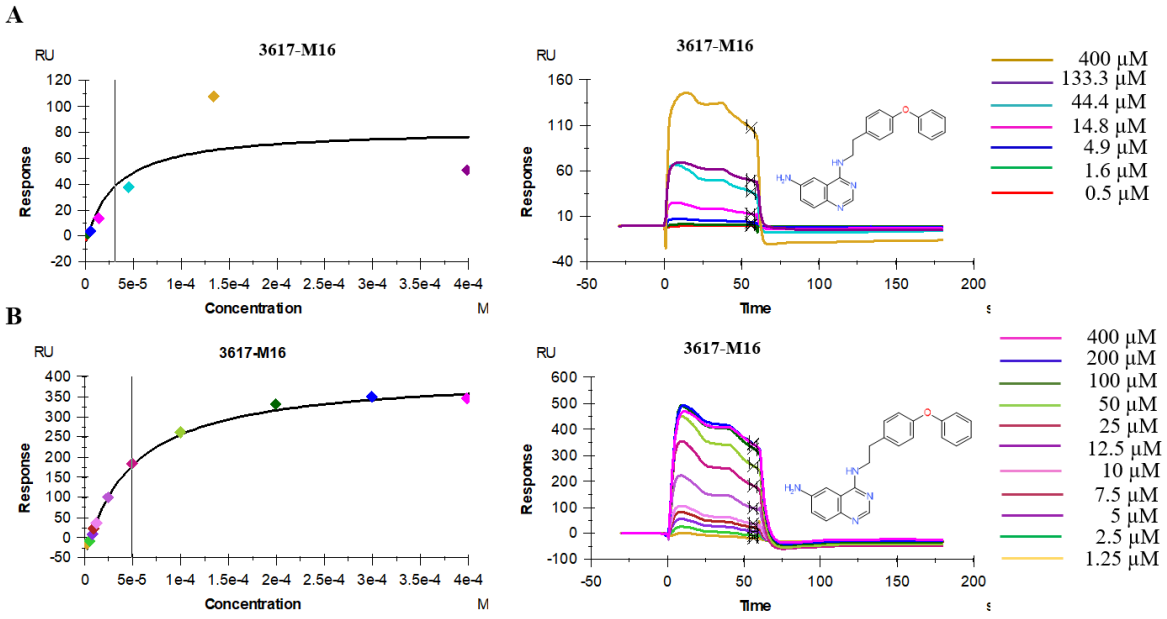


Figure 83. Steady state affinity of the KG-548 positive control with biotinylated mARNT2-PAS-B.

Interestingly, the 1:1 binding site model provided a poor fit for the association and dissociation traces observed for all ARNT2-tested compounds. As kinetic constants were

within instrument specification, but were difficult to determine, the steady state affinity analyses were also applied to all thirteen ARNT2-PAS-B compounds. For several hits, the affinities found via this method were poorer than expected based on preliminary EC_{50} obtained from the previous TSA experiment. Half of the compounds did not reach the saturation point, despite high concentration titration from 400 μM (Appendix 9). The poor binding affinity might be explained by the fact that some conformations are potentially sterically hindered upon immobilisation of the proteins. The binding affinity may also be affected by the quality of the proteins, which in turn could be compromised as a result of immobilisation. In general, the process of immobilising protein to a surface can potentially cause protein conformational changes, poor exposure of binding sites, as well as interactions with the biolayer surface that impact protein stability.[215,216]

Several compounds, however, still reached the saturation point and confirmed their affinity to the previously observed EC_{50} via TSA. Interestingly, these compounds are mostly those which showed some preliminary affinity via MST. For example, the crystallised binder 3617-M16 exhibited K_D ~31 μM for the first SPR run, validating its preliminary EC_{50} of 58 μM observed via TSA (Figure 84). The binding curves, however, showed some wavy patterns. There are many potential reasons for this effect. One of the possible explanations might be mass transport limitations, with the compound diffusing too slowly from the bulk solution to the sensor surface, making the binding rate transport-limited rather than reaction-limited.



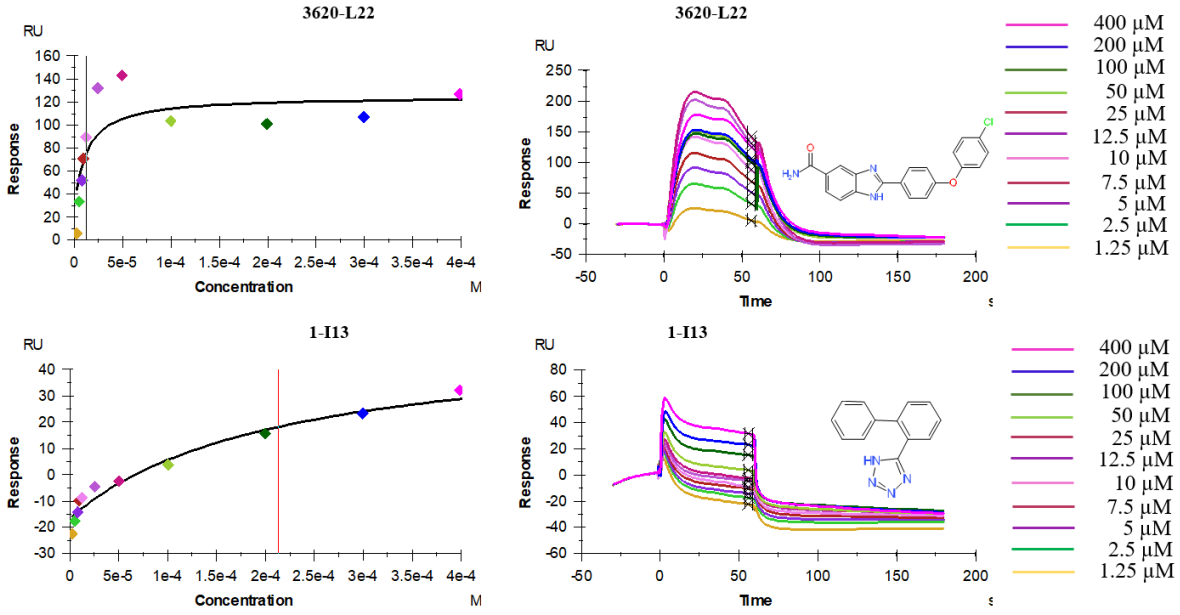
3617-M16	First run	Second run
K_D (µM)	31.7	49.4

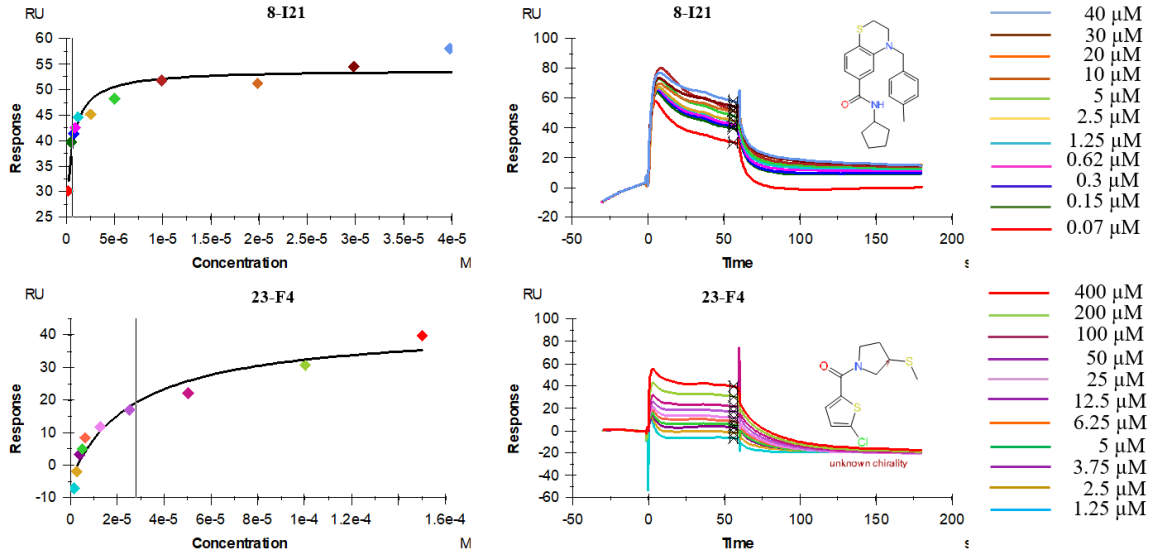
Figure 84. Steady state affinity for 3617-M16 with biotinylated mARNT2-PAS-B.(A) First SPR run with a flow rate of 20 µL/min. (B) The second SPR experiment with more concentration points and a higher flow rate of 30 µL/min. The calculated K_D values are shown in the blue table at the bottom of the figure.

To reduce the mass transport effect, the flow rate was increased from 20 µL/min to 30 µL/min. However, sensorgrams were still wavy. A second run was also performed with more concentration points. The calculated K_D ~49 µM again verified the expected affinity of 3617-M16 to the ARNT2-PAS-B domain. To further troubleshoot the wavy pattern and to eliminate the impact from the buffer pH, an additional trial was performed with a buffer swap from 50 mM Tris pH 7.0 to 50 mM HEPES pH 7.5, while keeping 150 mM NaCl and 0.01% Tween-20. It was expected that a higher pH 7.5 would improve protein solubility or reduce potential aggregation. HEPES buffer is also more stable with temperature changes, which

might be helpful for longer measurements. However, these adjustments did not help to correct the wavy curves (data is almost identical and therefore not shown).

In addition to 3617-M16, 3620-L22 also showed a saturation curve, and the calculated K_D ~12.7 µM matched the expected EC_{50} of 9 µM via TSA (Figure 85). The affinity and calculated values K_D ~212 µM of 1-I13 also closely align with the expected EC_{50} data of 203 µM. Strikingly, 8-I21 and 23-F4 showed much higher affinity via SPR with calculated K_D ~0.61 µM and K_D ~27.9 µM, than those predicted via TSA with EC_{50} of 11 µM and 825 µM, respectively. It should be emphasised again that TSA only indicates a binding event but not the affinity; the real affinity and the K_D values are those calculated with biophysical binding assays like SPR.





Compound	K_D (µM)
3620-L22	12.7
1-I13	212
8-I21	0.61
23-F4	27.9

Figure 85. Steady state affinity analyses for ARNT2 tested compounds with biotinylated mARNT2-PAS-B.The calculated K_D values are shown in the blue table at the bottom of the figure.

Given the small mass of the compounds and the relatively large mass of both heterodimers, it was not necessarily expected that a large enough change could be detected upon binding. As expected, most of the NPAS1/3 tested compounds showed a quite weak signal, on average 5 RU (Appendix 10-11). Even though the kinetic model was a good fit for some of the compounds, a 5 RU signal is right on the edge of the SPR detection limit; therefore, the calculated K_D values should be interpreted with caution.

To enhance the SPR signal, several troubleshooting procedures were implemented. Despite changing buffer compositions, adjusting contact and dissociation times, flow rate, as well as changing Avi-tag from the C-terminus to the N-terminus to potentially promote the accessibility of the protein binding pocket, the binding signal and affinity of tested compounds to NPAS1/3 proteins could not be improved.

Out of all NPAS1/3 compounds, only for 3624-N7 an RUmax signal of 26.5 was detected with biotinylated NPAS1 protein (Figure 86A). The 1:1 kinetic model provided a good fit, and the affinity to the NPAS1 protein was calculated with an approximate K_D ~16.9 μM. Based on the biological activity of enantiomers and on recent TSA data, it was expected that 3626-N5 isomer would show a lower binding affinity as compared to 3624-N7. As predicted, 3626-N5 enantiomer showed 12.8 RU, which is just half that of 3624-N7. However, the kinetic model was not a proper fit for 3626-N5. Therefore, the K_D ~139 μM is not considered reliable. For the NPAS3 tested compounds, only 3623-J16 showed a medium signal of 9.4 RU and a good kinetic fit with K_D ~453 μM (Figure 87A).

As there is no positive control for NPAS1/3, an additional to DMSO negative control, biotinylated hHIF-2α-PAS-B, was used. The protein was immobilised onto another channel within the same SA biosensor. Due to the negative counter screen data with hHIF-2α-PAS-B and NPAS1/3 tested compounds via TSA, no signal was expected from the SPR analyses. Surprisingly, all tested compounds, including 3624-N7, 3626-N5, and 3623-J16, demonstrated measurable binding to a single hHIF-2α-PAS-B which is proportionally similar to those RU signals with hNPAS1/3-ARNT2 heterodimers (Figure 86-87B). This non-specific interaction pattern suggests that the compounds may exhibit a tendency to adhere non-selectively to the sensor chip surface. Such behaviour indicates that surface-related

artefacts may be influencing the SPR results, and consequently, SPR may not be the most suitable assay for accurately characterising the binding properties of these compounds.

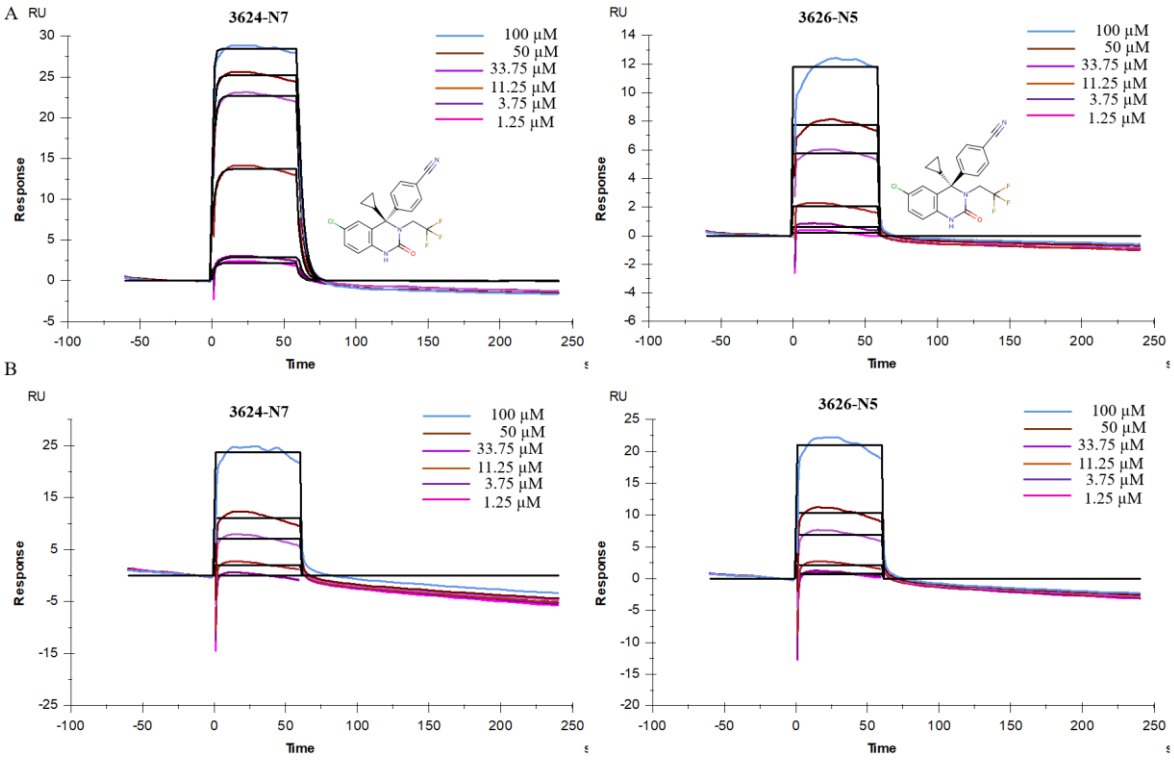


Figure 86. Dose-response kinetic model for 3624-N7 and 3626-N5 enantiomers.(A) Dose-response SPR analyses with biotinylated hNPAS1-ARNT2. (B) Dose-response SPR analyses with biotinylated hHIF-2α-PAS-B as a negative control.

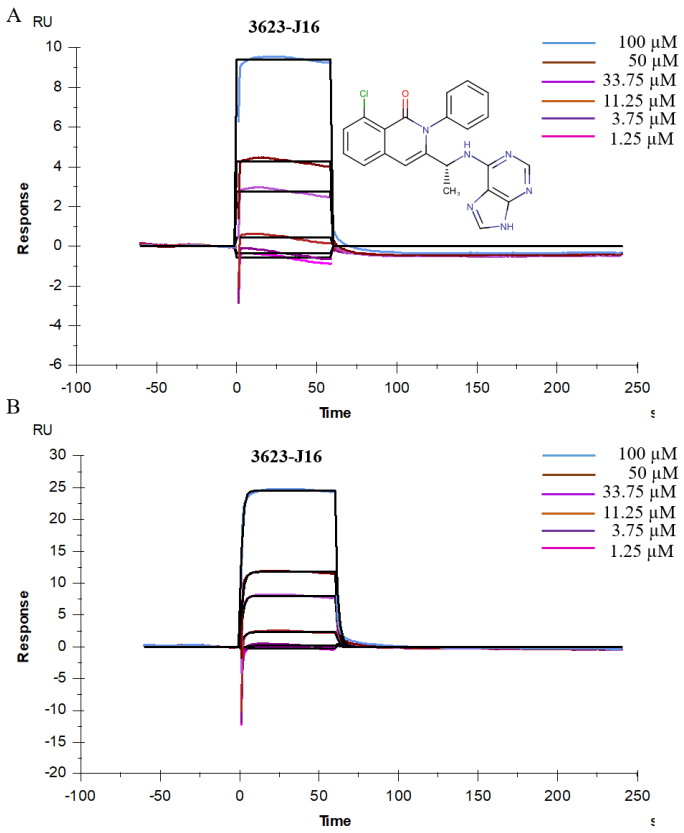


Figure 87. Dose-response kinetic model for 3623-J16.(A) Dose-response SPR analyses with biotinylated hNPAS3-ARNT2. (B) Dose-response SPR analyses with biotinylated hHIF-2α-PAS-B as a negative control.

To further investigate the affinity between tested compounds and NPAS1/3 heterodimers, orthogonal techniques might be used, such as biolayer interferometry (BLI), fluorescence polarisation (FP), or isothermal titration calorimetry (ITC). BLI is similar to SPR, as both are label-free, real-time, and optical biosensing techniques used to study biomolecular interactions. Instead of an immobile chip, BLI utilises a mobile optical probe with an immobilised biomolecule, which is dipped and rapidly oscillated in static solutions within a microplate.[217] BLI measures changes in the interference pattern of white light reflected from two surfaces: a reference layer and a layer containing immobilised biomolecules on a

biosensor tip. When a binding event occurs at the sensor surface, the optical thickness changes, causing a measurable shift in the interference pattern. This shift is directly proportional to the number of biomolecules that are bound. While BLI offers several advantages, such as ease of use, high throughput, and cost-effectiveness, it has limited sensitivity for low-molecular-weight analytes (<500 Da), making it less suitable for detecting weak or subtle binding events involving small molecules. Therefore, BLI may not be the optimal choice to study large heterodimer interactions with small molecules, which are on average less than 300 Da and require high-sensitivity detection.

FP is another technique that can utilise a low protein concentration (nanomolar range) to study the binding effect between biomolecules. The measurement, however, mostly depends on the size change of a biomolecule upon binding. For 100 kDa heterodimers, the mass increase of ~300 Da upon binding a small molecule would not be sensitive enough. Even for a single domain like ARNT2-PAS-B, the increase in mass from the ~12.6 kDa protein to 12.9 kDa protein-bound to a small molecule was not expected to generate a change in polarisation great enough to be detected.

ITC, on the other hand, is highly effective for studying interactions between proteins and small molecules because it doesn't rely on molecular size or optical properties, making it sensitive and quantitative. ITC measures the heat that is released (exothermic reaction) or absorbed (endothermic reaction) upon the binding of two molecules, a ligand and a protein.[218] One of the major advantages of ITC is its ability to provide comprehensive thermodynamic information, not only the K_D, but its ability to help determine the enthalpy change (ΔH), entropy change (ΔS), Gibbs free energy (ΔG), binding stoichiometry (n), and the strength of the interaction (the equilibrium association constant $K\alpha$), from which the K_D could be

derived. However, ITC also has limitations. In addition to long-time measurements, the major drawback is that it typically requires relatively high concentrations and quantities of purified protein. Given the low yield of NPAS1/3 heterodimers, ITC would not be viable. Although ARNT2 PAS-B is well expressed and amenable to biophysical characterisation, ITC was not attempted. SPR measurements showed unsaturated binding curves for KG-548 and most tested compounds, even at the highest concentrations, indicating very weak interactions, with estimated K_D values ranging from several hundred micromolar to over 400 μM, with this higher range of values being typical for KG-548. Such affinities fall outside the practical sensitivity range of ITC, and weak binders will likely produce very small heat changes, which are often buried in noise. ITC is generally suitable for moderate affinities (~1 nM to 100 μM), where the heat signal is detectable. Outside this range, the signal is too small relative to noise, and achieving saturation would require impractically high ligand concentrations or exceed solubility and DMSO limits. Therefore, ITC was unlikely to yield reliable, interpretable thermodynamic parameters and was not pursued.

Chapter 9: Time-resolved FRET dimerisation assay

9.1 Principle of the TR-FRET assay

After addressing the binding affinity of identified ligands that directly bind to ARNT2-PAS-B and NPAS1/3-ARNT2 heterodimers, the next step is to determine whether these compounds act as activators or inhibitors. To facilitate this, developing an assay based on protein-protein interaction would enable efficient screening for compounds that either stabilise or disrupt the heterodimer.

The TR-FRET is particularly well-suited for this purpose, as it combines the low background aspect of time-resolved fluorescence (TRF) with the homogeneous assay format of fluorescence resonance energy transfer (FRET). TR-FRET is based on the non-radiative transfer of energy from an excited donor fluorophore to a nearby acceptor. GFP is commonly used as the acceptor fluorophore, particularly when paired with a terbium (Tb) cryptate donor. This energy transfer only occurs when the two fluorophores are within a short distance of each other, typically 10-100 Å, and have significant spectral overlap, specifically between the donor's emission and the acceptor's excitation spectra. The energy transferred is emitted as fluorescence at the acceptor's characteristic wavelength (Figure 88).

The efficiency of this process is known as FRET efficiency. The process is highly sensitive to the distance between the two molecules, with small changes causing large differences in energy transfer. FRET efficiency is mathematically defined as:

$$E = 1 / [1 + (r/R_0)^6]$$

where r is the distance between donor and acceptor, and R_0 is the Förster distance at which energy transfer efficiency is 50%. The small distance range of this technique has proven particularly suited to evaluate protein-protein interactions and conformational changes in real time, as many complexes typically span under 100 Å.[80,117]

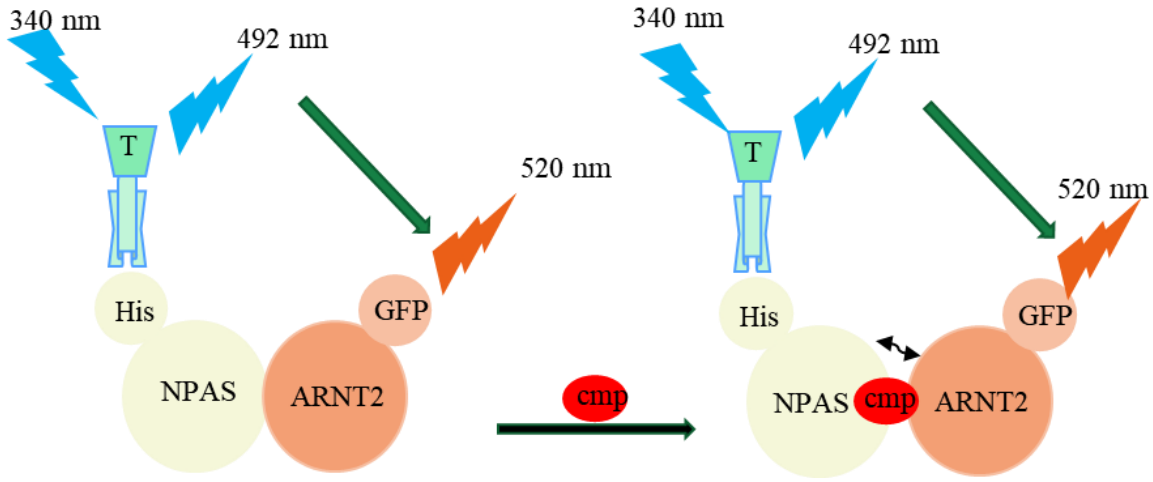


Figure 88. Illustration of the TR-FRET principle.When the terbium donor is excited (340 nm) by an external energy source, such as a flash lamp or laser, and if the GFP acceptor is in close proximity, energy transfer occurs and is detected by fluorescence emission at 520 nm. The assay readout is calculated as the ratio of emissions at 520 nm to 492 nm. This provides a quantitative measure of heterodimer stabilisation or destabilisation upon binding a ligand.

TR-FRET is widely used in drug discovery as it provides increased flexibility, reliability, and sensitivity. A key advantage of TR-FRET is the use of lanthanide donors like terbium chelates, often conjugated to antibodies that bind to His-tags proteins. These complexes have an exceptionally long fluorescence lifetime, enabling the detection of FRET signals after

background autofluorescence and scattered light have decayed. This leads to high assay sensitivity and significantly reduces both false positives and false negatives.

9.2 Development of the TR-FRET assay

Since all tested compounds are presumed to bind the PAS-B domain of the target proteins, the interaction between the PAS-B domain of ARNT2 and the PAS-B domain of its partner represents a key interface to study. Previously, TR-FRET was successfully used with GFP-tagged ARNT together with His-tagged HIF-2α and NPAS3 proteins within the laboratory.[80,117] Similarly, enhanced GFP (eGFP) was fused to the C-terminus of both mouse and human ARNT2 proteins via a two-step PCR strategy.

The DNA sequences of eGFP, and both mouse and human ARNT2, with overlapping ends, were amplified by initial PCR and run on an agarose gel. The DNA bands of approximately 700 bases correspond to the mass of eGFP, and the bands of approximately 1,300 bases correspond to mouse or human ARNT2 (Figure 89). The purified PCR products of eGFP and mouse or human ARNT2 were subsequently mixed at a 1:1 molar ratio and used in the secondary PRC, where the overlapping regions anneal. The final PCR product consisted of a seamless ARNT2-eGFP fusion construct (Appendix 12).

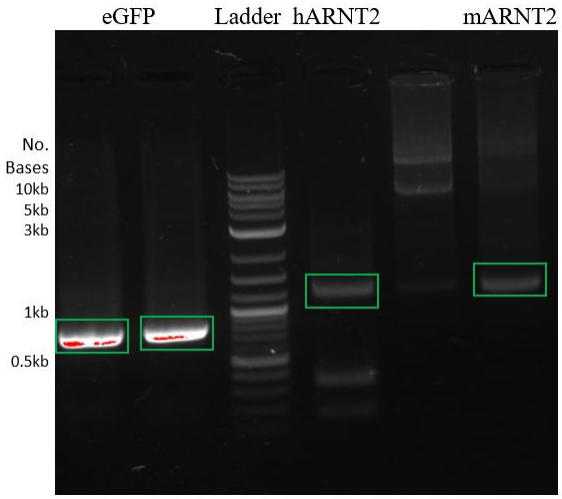


Figure 89. One per cent DNA agarose gels.Purified PCR products of eGFP (723 bases) and both mouse and human ARNT2 constructs with approximate mass of 1,300 bases, correspond to the mass of triple domain of the protein.

To evaluate ARNT2-PAS-B hits and given the unknown dimerisation partner for these compounds, a generated full-length mouse ARNT2-eGFP fusion product was cloned into the pMKH expression vector and later co-expressed and purified alongside several in-house C-terminal His-tagged triple-domain proteins from the bHLH-PAS family, including mHIF-2α, mNPAS1, mNPAS3, and mNPAS4 (Appendix 13).

For NPAS1/3 tested compounds, hARNT2-eGFP was cloned into existing N-terminus-His-tagged-pETDuet-1-hNPAS1/3 constructs previously used for protein purification. Proteins were successfully overexpressed and purified (Figure 90). The MS analyses, however, showed only the presence of hNPAS1 or hNPAS3 proteins, without fused hARNT2-eGFP (Figure 91). The elution volume of approximately 70 mL from S200 size exclusion chromatography suggested that both proteins eluted as heterodimer complexes.

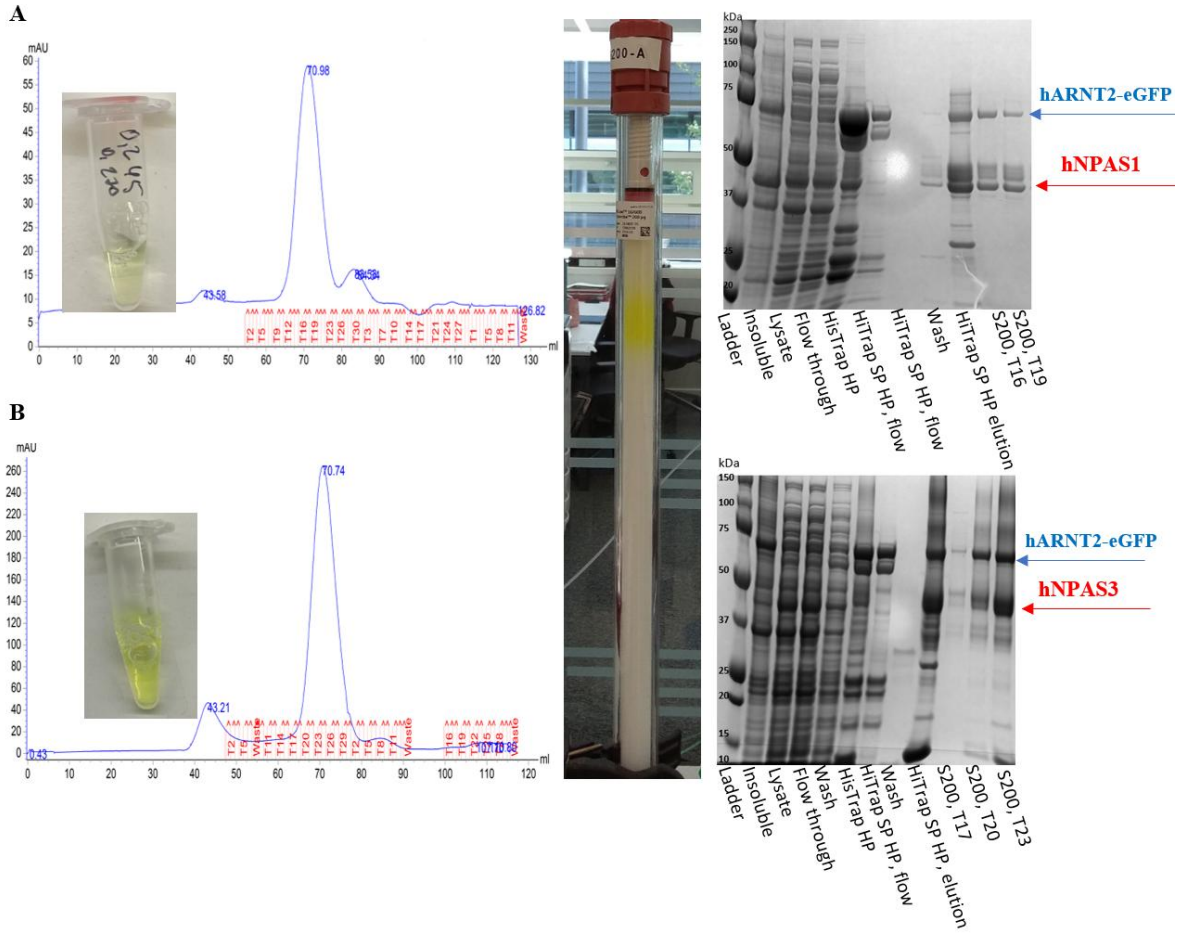
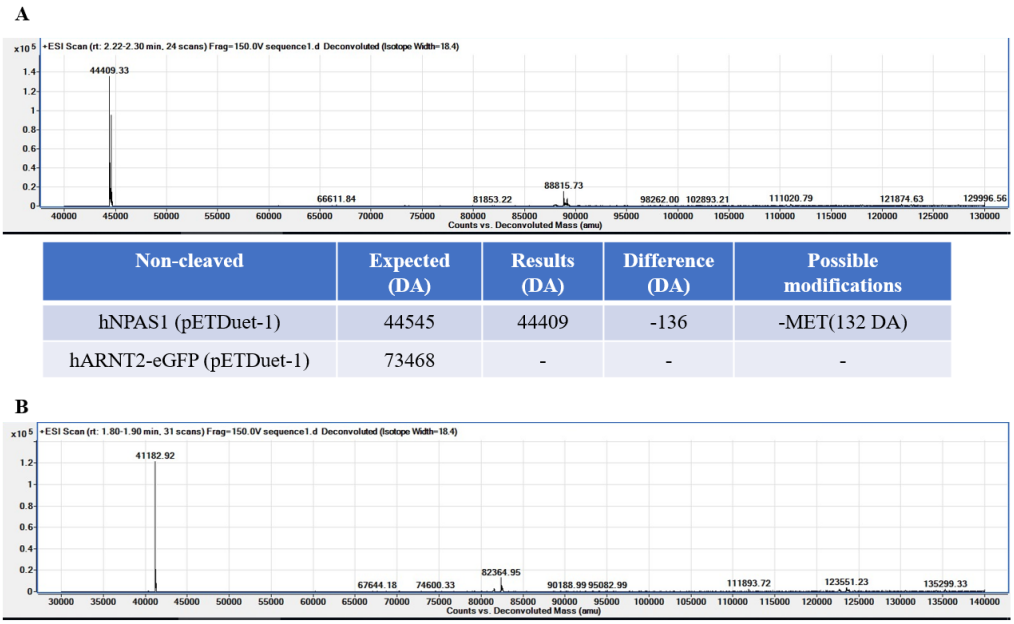


Figure 90. Protein purification for the TR-FRET assay.Both (A) hNPAS1-ARNT2-eGFP and (B) hNPAS1-ARNT2-eGFP constructs were expressed and purified via HisTrap HP metal ion affinity chromatography, followed by HiTrap SP HP cation-exchange, and S200 size exclusion chromatography. Fused eGFP (26.9 kDa) to the hARNT2 (46 kDa) provides a bright green-yellow colour. SDS-PAGE confirmed the presence of a 73 kDa protein band, which represents fused ARNT2-eGFP (26.9 + 46 = 72.9 kDa). The protein band of 44 kDa or 50 kDa showed the presence of NPAS1 or NPAS3 proteins, suggesting heterodimeric protein organisation.

The SDS-PAGE gels clearly showed the presence of approximately 73 kDa bands corresponding to the mass of hARNT2-eGFP, suggesting that eGFP is bound to ARNT2

(Figure 90). Protein bands at 44 kDa and 50 kDa in each sample correspond to the mass of hNPAS1 and hNPAS3 proteins, implying a heterodimeric protein organisation. Additionally, a bright green-yellow colour of both NPAS1/3-ARNT2-eGFP samples indicates the presence of eGFP fused to ARNT2.



Non-cleaved	Expected (DA)	Results (DA)	Difference (DA)	Possible modifications
hNPAS1 (pETDuet-1)	44545	44409	-136	-MET(132 DA)
hARNT2-eGFP (pETDuet-1)	73468	-	-	-

TEV-cleaved	Expected (DA)	Results (DA)	Difference (DA)	Possible modifications
hNPAS1 (pETDuet-1)	41188	41183	-5	
hARNT2 -eGFP(pETDuet-1)	70298	-	-	-

Figure 91. Representative intact mass analysis of purified hNPAS1-ARNT2-eGFP.Deconvoluted spectra and masses for (A) non-cleaved and (B) TEV-cleaved hNPAS1-ARNT2-eGFP protein.

The ARNT2-eGFP protein is a relatively large complex. It is possible that during ESI-MS, larger proteins produce multiple low-intensity charge states, leading to dispersed signals that can be difficult to detect with high sensitivity. Additionally, a compact and hydrophilic β-barrel structure within eGFP could hinder efficient protonation during ESI, resulting in poor

ionisation and reduced detectability in the mass spectrometer. Thus, considering protein quality control based on size exclusion elution volume, SDS-PAGE, and the colour of the sample, both proteins were used for TR-FRET analyses.

Both proteins, however, did not produce a strong signal via the TR-FRET assay (data is not shown). This is likely due to the relatively long distance between the fluorophores. Due to the limitation of the ARNT2 heterodimeric structure with NPAS1/3, the existing crystal structure of a closely related partner, mNPAS4-ARNT2, was employed. The structure of mNPAS4-ARNT2 shows that the distance between the N-termini of NPAS4 and the C-termini of ARNT2 is 98.2 Å, which falls just within the upper range for effective FRET detection (Figure 92A). In contrast, the distance between the C-termini of NPAS4 and ARNT2 is just 45.9 Å. Considering that NPAS1/3-ARNT2 and other bHLH-PAS family heterodimers are similar in size to NPAS4-ARNT2, the fluorophores fused to the C-termini of these proteins are likely to fall within the FRET-compatible range of under 100 Å (Figure 92B). However, the purification of human C-terminus-His-tagged-NPAS1/3 proteins expressed from pETDuet-1 constructs was not successful. Therefore, in-house mouse constructs encoding C-terminal His-tagged-NPAS1/3-ARNT2-eGFP heterodimers, used for testing ARNT2-PAS-B compounds, were instead employed for testing NPAS1/3-targeting compounds.

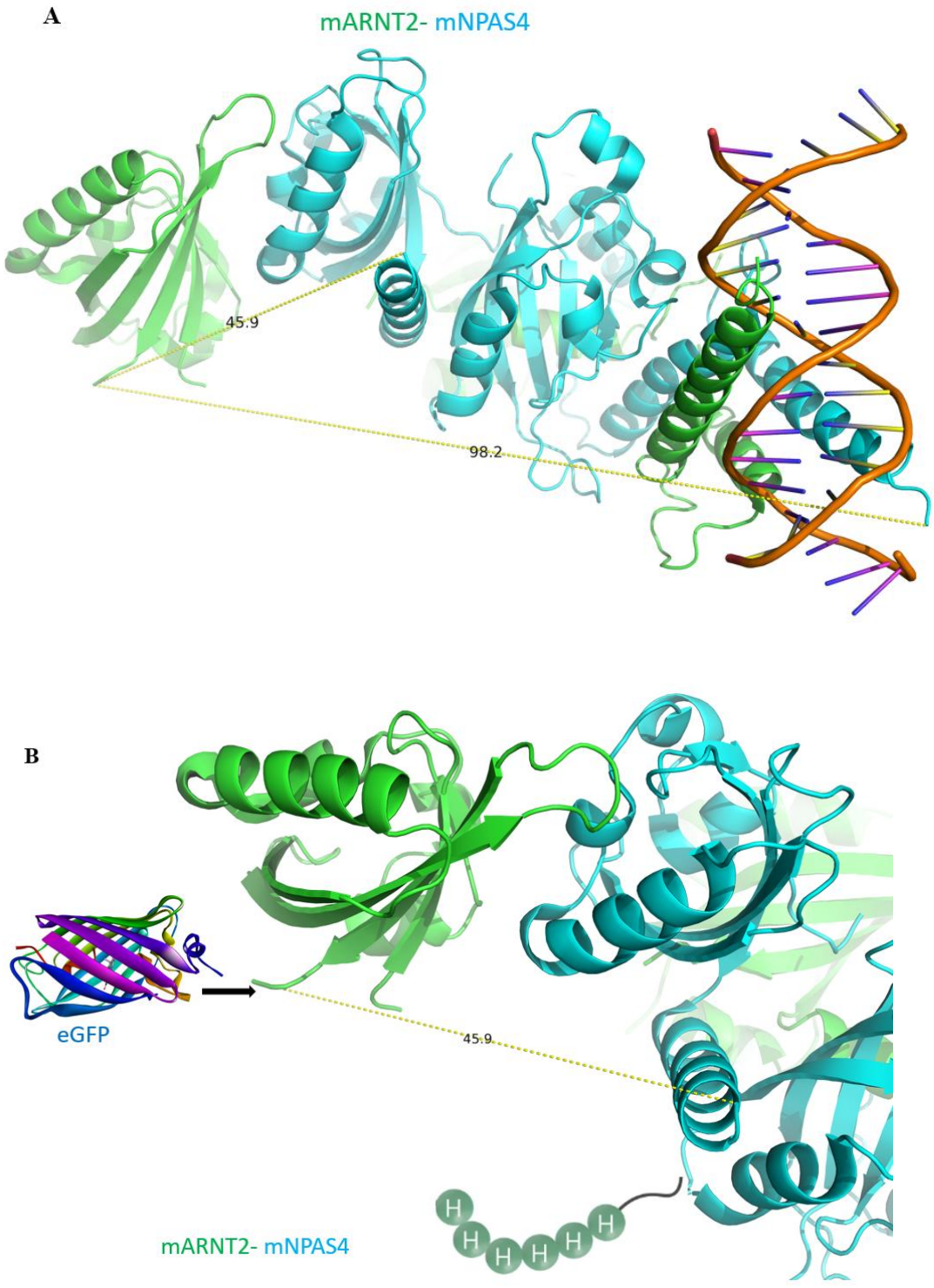


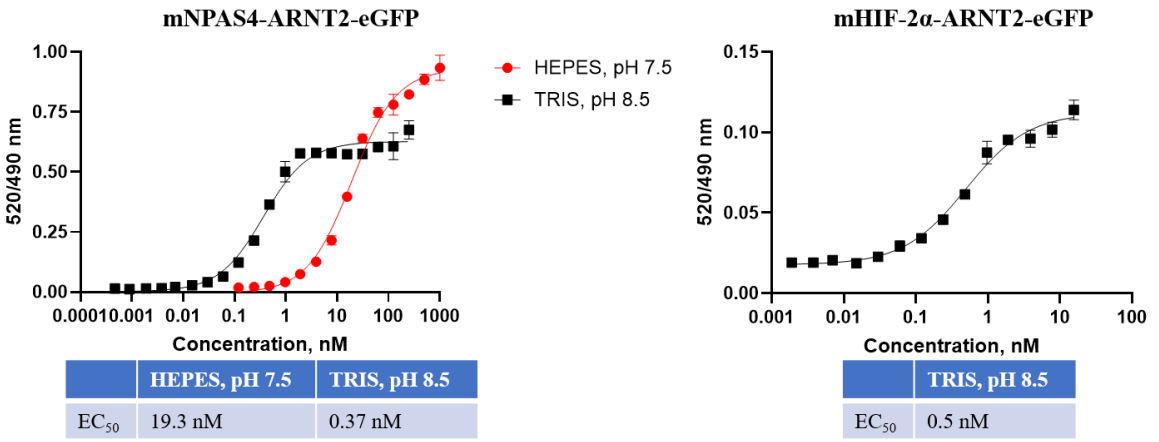
Figure 92. Structural illustration of the distance between potential fluorophores.(A) A representative existing crystal structure of mNPAS4-ARNT2 (PDB: 7XI3) was used to measure the distance between the C-termini of NPAS4 and the C-termini of ARNT2 (45.9

Å), and between the N-termini of NPAS4 and the C-termini of ARNT2 (98.2 Å). (B) Schematic structural representation of the TR-FRET construct. eGFP is fused to the C-terminus of mARNT2 and His-tag is added to the C-terminus of mNPAS4.

9.3 Optimisation of the TR-FRET assay

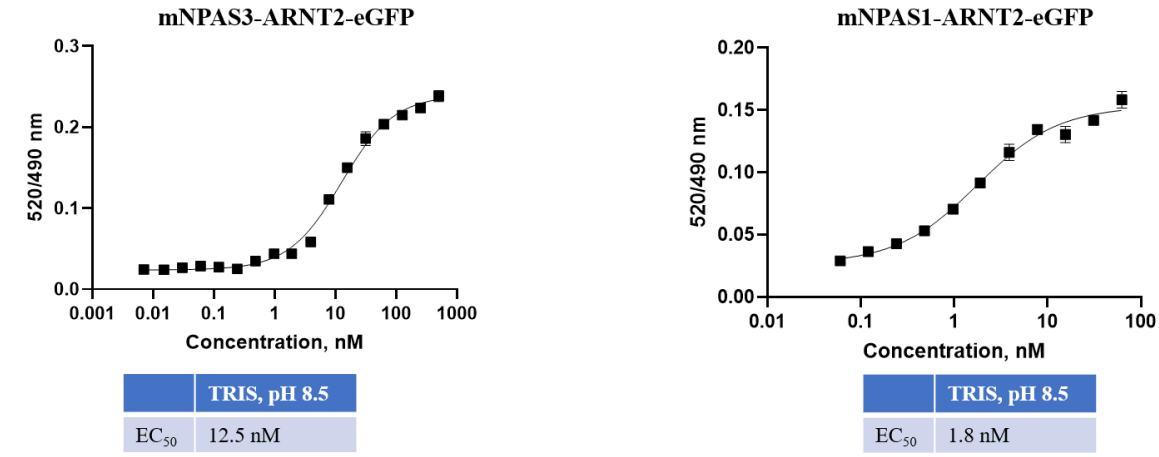
Optimising a TR-FRET assay involves carefully adjusting experimental conditions to maximise sensitivity and signal stability and validating the system using known ligands. To evaluate buffer conditions for optimal assay performance, mNPAS4-ARNT2-eGFP was tested in its original HEPES buffer pH 7.5 used during purification, and with previously reported Tris buffer pH 8.5.[117] The Tris buffer yielded higher assay sensitivity, resulting in a more potent EC_{50} value of 0.37 nM as compared to 19 nM for HEPES (Figure 93).

To measure optimum protein concentration for the assay, ARNT2-eGFP protein constructs with mHIF-2α, mNPAS1, mNPAS3, and mNPAS4 were tested at multiple concentrations using DRC in Tris buffer. For TR-FRET, each protein was used at a double concentration of its measured EC_{50} value (Figure 93).



	HEPES, pH 7.5	TRIS, pH 8.5
EC_{50}	19.3 nM	0.37 nM

	TRIS, pH 8.5
EC_{50}	0.5 nM

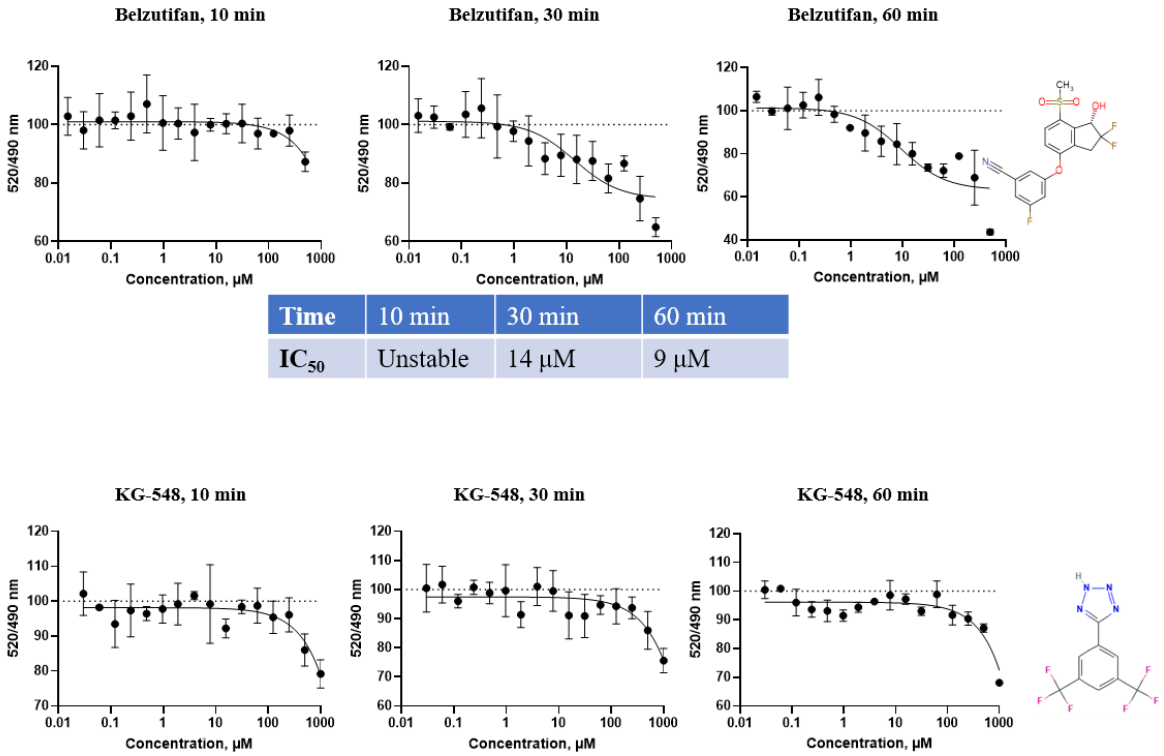


	TRIS, pH 8.5
EC_{50}	12.5 nM

	TRIS, pH 8.5
EC_{50}	1.8 nM

Figure 93. Buffer optimisation and protein curve analyses for the TR-FRET assay.

To optimise the incubation time with the compounds, known ligands like belzutifan for HIF-2α, and KG-548 for ARNT2-PAS-B were used with the mHIF-2α-ARNT2-eGFP construct. Belzutifan is a potent inhibitor and showed 40% inhibition of the mHIF-2α-ARNT2 heterodimer after 30 minutes' incubation (Figure 94). These results are very similar to the previously reported mHIF-2α-ARNT construct.[117] After 1hour incubation, belzutifan showed even further destabilisation of the mHIF-2α-ARNT2 heterodimer at approximately 60% with IC_{50} of 9 µM. Importantly, inhibition with belzutifan does not reach zero, suggesting that it does not fully separate the heterodimer complex in the TR-FRET assay.



Time	10 min	30 min	60 min
IC_{50}	Unstable	14 μM	9 μM

Figure 94. Incubation time optimisation with mHIF-2α-ARNT2-eGFP.Belzutifan and KG-548 control compounds were incubated with mHIF-2α-ARNT2-eGFP protein for 10 min, 30 min, and 1 hour.

There was not much difference between 10, 30, and 60 minutes of incubation for KG-548; therefore, a medium incubation time of 30 minutes was chosen for the TR-FRET assay. As expected, KG-548 showed a minimal and barely detectable inhibition effect with HIF-2α-ARNT2 heterodimer, limited to the two highest concentration points throughout the incubation time course (Figure 94). Similar to results with other heterodimers, KG-548 did not show any activity via TR-FRET (Figure 95). These results support the TSA data and the structural analyses of KG-548 binding heterodimers, indicating conformational changes which are unfavourable for the heterodimeric constructs.

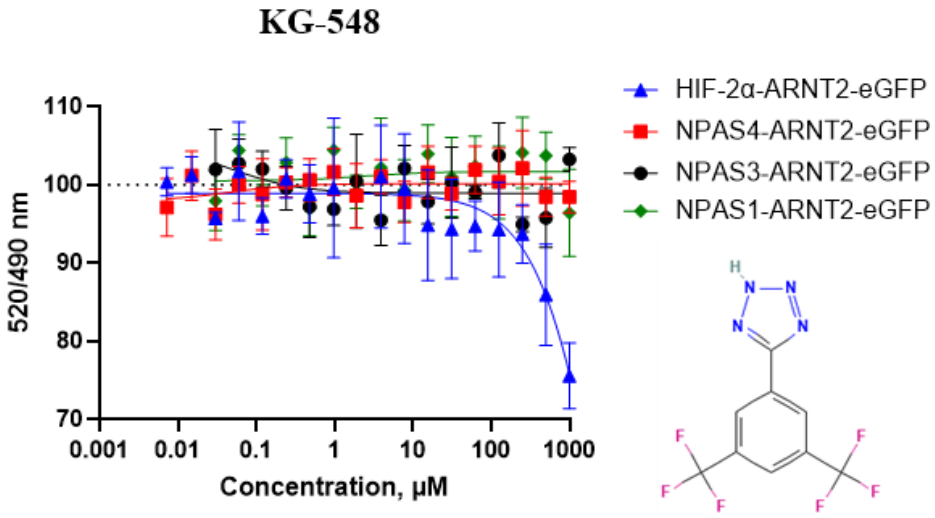


Figure 95. TR-FRET with KG-548 control compound.Since it is a reference for ARNT2-PAS-B, KG-548 was tested with several protein partners within the bHLH-PAS family via a 30 min TR-FRET assay.

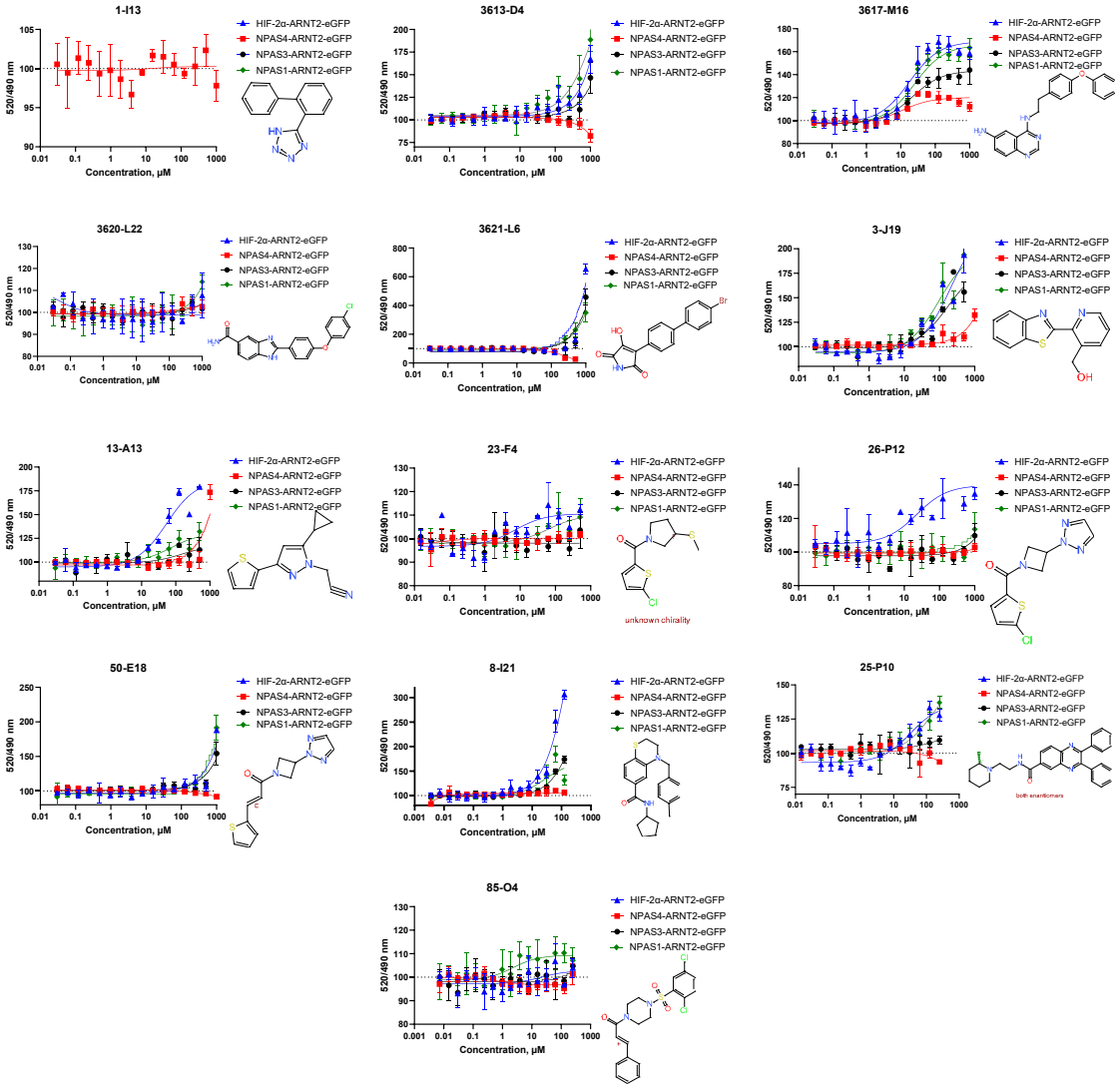
9.4 TR-FRET assay with all the tested compounds

To assess whether the tested compounds act as enhancers or suppressors, 13 compounds for mARNT2-PAS-B and 11 compounds for each of the hNPAS1/3-ARNT2 heterodimers were evaluated via TR-FRET. As the precise dimerisation partner for ARNT2-PAS-B binding compounds is unknown, TR-FRET was performed with several ARNT2 heterodimers from the bHLH-PAS family, including mHIF-2α, mNPAS1, mNPAS3, and mNPAS4.

Based on the analogy with the KG-548 crystal structure, it was expected that all crystallised compounds would have a similar inhibitory effect on heterodimeric complexes. As 3617-M16, 50-E18, and 3613-D4 bind the same place inside the PAS-B cavity, interact with identical residues, and cause similar conformational changes as KG-548, it was anticipated that ligands would also promote unfavourable interactions for the heterodimer complexes.

Surprisingly, TR-FRET did not identify any inhibitors; all tested compounds showed increased signal, suggesting enhanced heterodimer formation. Three compounds, 3617-M16, 3-J19, and 13-A13, exhibited a signal increase from 20% to 80% across all heterodimer complexes (Figure 96). In particular, 3617-M16 showed a potent saturated signal with EC_{50} from 10-24 μM throughout all heterodimer pairs. Although the calculated EC_{50} is somewhat in agreement with other assays, including MST, SPR, and TSA, the crystal data strongly suggested an inhibitory effect for hit compound, while weak binders are expected to show no significant impact, similar to those observed with KG-548, but definitely not a strong favourable effect on formation of heterodimers.

Two other crystals, 50-E18 and 3613-D4, together with 3621-L6 and 8-I21 also showed increased activity with all heterodimer partners except mNPAS4. Two compounds, 26-P12 and 25-P10, showed activity only with mHIF-2α or mHIF-2α and mNPAS1, respectively. Four compounds did not show any activity via TR-FRET, including 1-I13, 3620-L22, 23-F4, and 85-O4.


1-I13
3613-D4
3617-M16
3620-L22
3621-L6
3-J19
13-A13
23-F4
26-P12
50-E18
8-I21
25-P10
85-O4
HIF-2α-ARNT2-eGFP
NPAS4-ARNT2-eGFP
NPAS3-ARNT2-eGFP
NPAS1-ARNT2-eGFP
520/490 nm
Concentration, μM
unknown chirality
both enantiomers

EC_{50} (μM)				
Compound	HIF-2α-ARNT2	NPAS4-ARNT2	NPAS3-ARNT2	NPAS1-ARNT2
1-I13		n/a		
3613-D4	unstable	n/a	unstable	unstable
3617-M16	17.8	10.5	21.3	24.7
3620-L22	n/a	n/a	n/a	n/a
3624-L6	unstable	n/a	unstable	unstable
3-J19	214	unstable	169	94.8
13-A13	54	unstable	150	72.8
23-F4	7.1	n/a	n/a	63.7
26-P12	22.4	n/a	n/a	n/a
50-E18	unstable	n/a	unstable	unstable
8-I21	219.4	n/a	382.6	22.1
25-P10	32.4	n/a	n/a	143.6
85-O4	n/a	n/a	n/a	2.3

Figure 96. TR-FRET with 13 ARNT2-PAS-B tested compounds.To access the activity of the compounds, mouse ARNT2-eGFP was co-expressed and purified with several potential partners, including mHIF-2a, mNPAS1, mNPAS3, and mNPAS4.

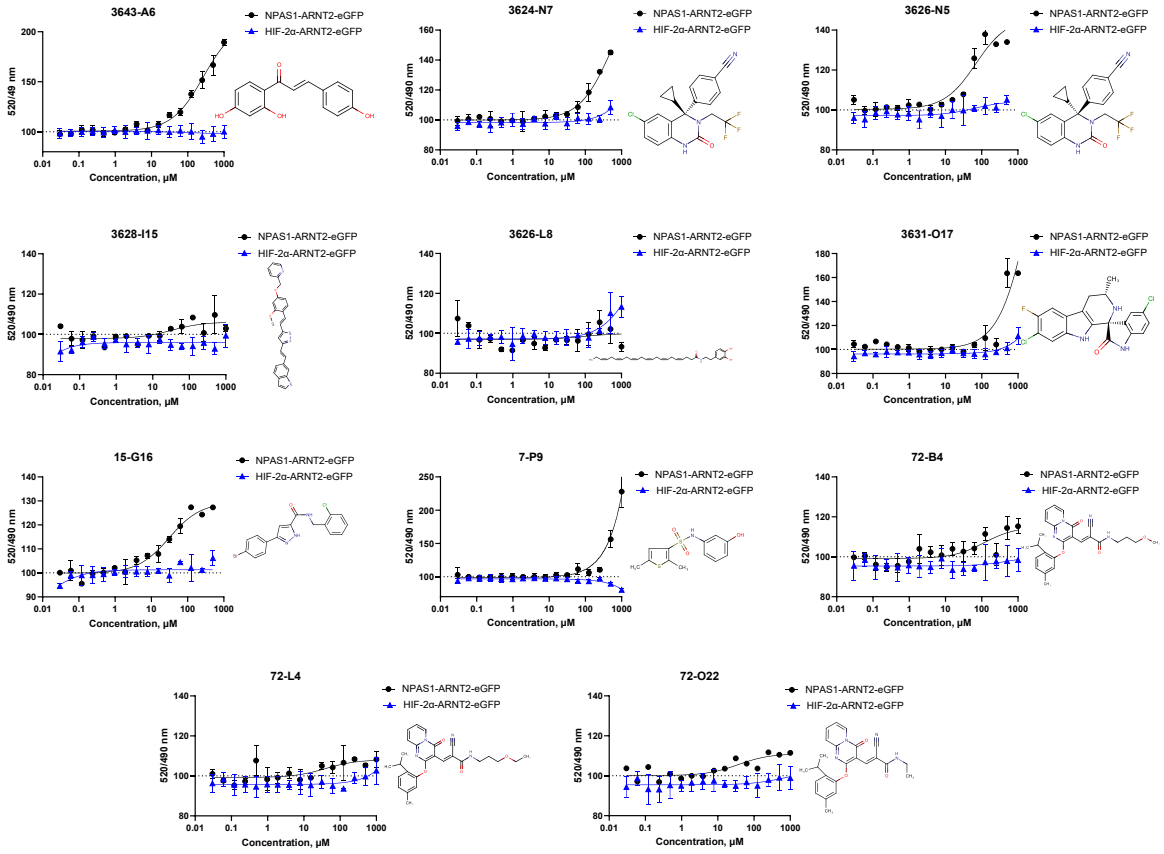
As FRET efficiency is highly dependent on the distance and relative orientation of donor and acceptor fluorophores, one plausible interpretation is that fluorophores might come closer to each other due to the potential flexibility of PAS-B domains in the solution. Another reason could be tag interference, as fusion tags such as GFP may affect protein folding or binding. Additionally, compound autofluorescence may also cause signal distortion by quenching signals at similar wavelengths as donor or acceptor fluorophores.

Without NPAS1/3-ARNT2 crystal data, it is challenging to predict potential activities of the tested compounds. Similar to ARNT2 ligands, TR-FRET identified only activators across 22

NPAS1 and NPAS3 tested compounds (Figures 97, 98). Six compounds, including 3643-A6, 3624-N7, 3626-N5, 3631-O17, 15-G15, and 7-P9 showed increased NPAS1-ARNT2 heterodimer stabilisation. Interestingly, it was 3624-N7 that produced more prominent data as compared to its isomer throughout all assays; however, the TR-FRET calculated EC_{50} of ~ 70 μM for 3626-N5 is not only seven times higher than that of 3624-N7, but also matches the TSA predicted EC_{50} of ~ 85 μM. Additionally, 15-G15 showed a potent activation curve with EC_{50} of ~ 31 μM, perfectly matching the calculated EC_{50} of ~ 34 μM via TSA.

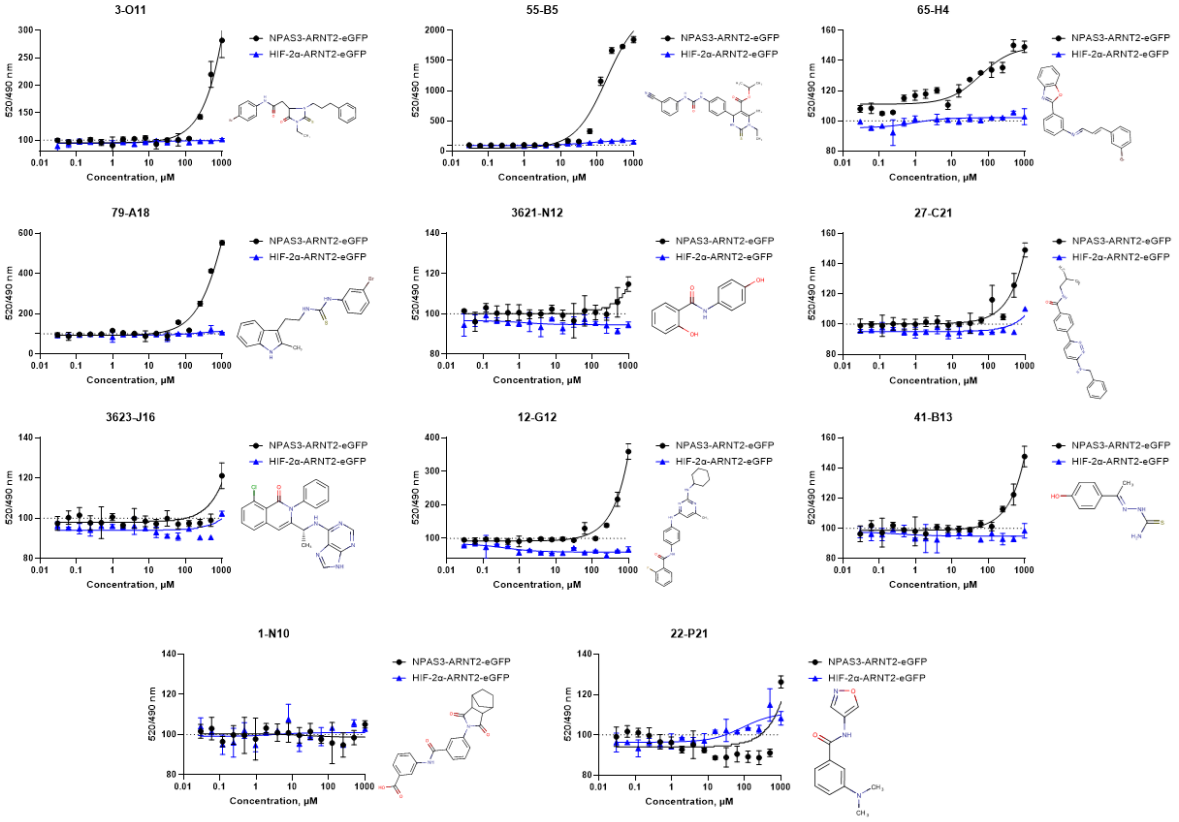
For NPAS3 tested compounds, 3-O11, 55-B5, 65-H4, 79-A18, 27-C21, 12-G12, and 41-B13 showed enhanced heterodimer formation with ARNT2 protein. However, three compounds, 55-B5, 79-A18, and 12-G12, demonstrated a relatively strong TR-FRET signal, specifically a large increase in the normalised GFP/Terbium emission ratio, possibly indicating a strong artificial effect, like fluorescence interference, or compound aggregation. Interestingly, mHIF-2α-ARNT2-eGFP did not show a high signal with these ligands, indicating potential assay artefacts with NPAS3 protein rather than compound-related effects.

To cross-check results, mHIF-2α-ARNT2-eGFP was tested along with NPAS1/3 compounds as a negative control. As expected, none of the compounds showed any activity with mHIF-2α-ARNT2 proteins confirming negative counter screen with mHIF-2α-PAS-B domain via TSA.



EC_{50} (μM)	
Compound	NPAS1-ARNT2
3643-A6	271.5
3624-N7	520
3626-N5	69.6
3628-I15	32.9
3628-L8	n/a
3631-O17	unstable
15-G16	31.1
7-P9	unstable
72-B4	118.4
72-L4	39.4
72-O22	36.4

Figure 97. TR-FRET with NPAS1 compounds.To access the activity of the compounds, mouse ARNT2-eGFP was co-expressed and purified with mNPAS1.



EC_{50} (μM)	
Compound	NPAS3-ARNT2
3-O11	unstable
55-B5	179.5
65-H4	61.4
19-A18	unstable
3621-N12	n/a
27-C21	unstable
3623-J16	n/a
12-G12	unstable
41-B13	unstable
1-N10	n/a
22-P21	n/a

Figure 98. TR-FRET with NPAS3 compounds.To access the activity of the compounds, mouse ARNT2-eGFP was co-expressed and purified with mNPAS3.

TR-FRET is a relatively easy and well-established assay within the laboratory. However, considering the crystallisation data, TR-FRET might not be the most suitable biochemical assay for assessing the functional activity of the heterodimers. Perhaps, future studies, with cell-based assays, like luciferase, where both protein partners are separated and yet to be heterodimerised, would be a better condition to establish ligand activity.

Chapter 10: Hit-to-lead optimisation for KG-548

In parallel to the discovery of new compounds and potential phenotypes for the transcription factors NPAS1, NPAS3, and ARNT2, another aim of the project was to conduct hit-to-lead optimisation for known binders, like reported KG-548, and recently identified butein. Hit-to-lead optimisation was carried out by using a combination of SAR by catalogue and synthetic medicinal chemistry efforts. KG-548 analogues were designed and synthesised by the research group of Professor Paul Brennan, including Dr. Lukas Scheibelberger, Dr. Ian de Toledo, and Tobias Bernard. The following butein analogues were designed by Dr. Srinivas Cheruku and ordered from convenient vendors.

10.1 Binding mode of KG-548 and initial design of analogues

KG-548 has a tetrazole ring attached to the bis(trifluoromethyl)phenyl ring. Tetrazoles are five-membered heterocyclic compounds containing four nitrogen atoms and one carbon atom. Due to their approximate pKa of 4.9, tetrazoles are known for their acidic properties. Thus, at physiological pH 7.4, which is significantly above their pKa, tetrazoles exist predominantly in their deprotonated form, carrying a negative charge (Figure 99).

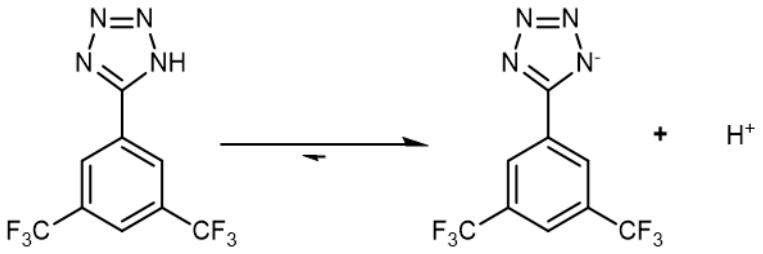
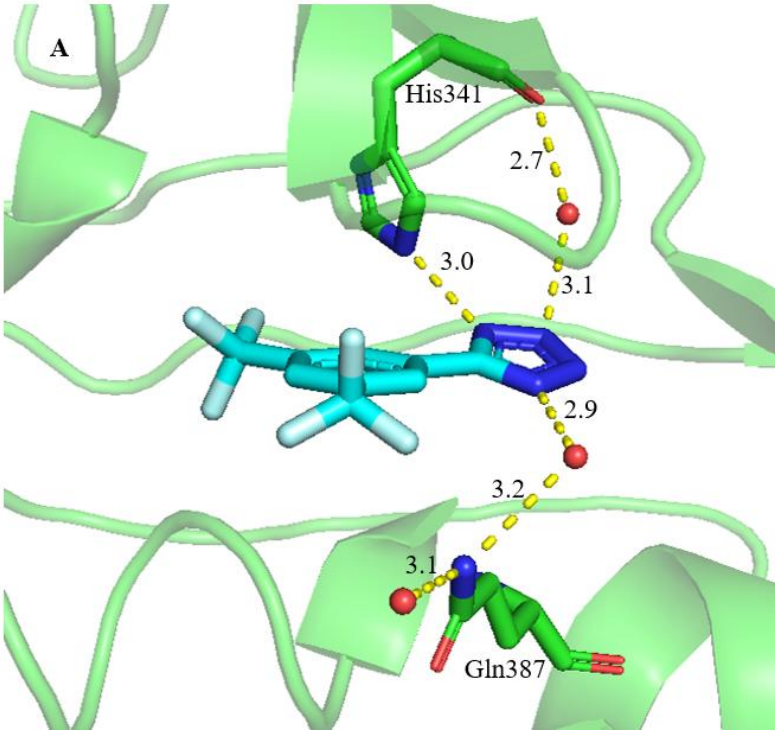


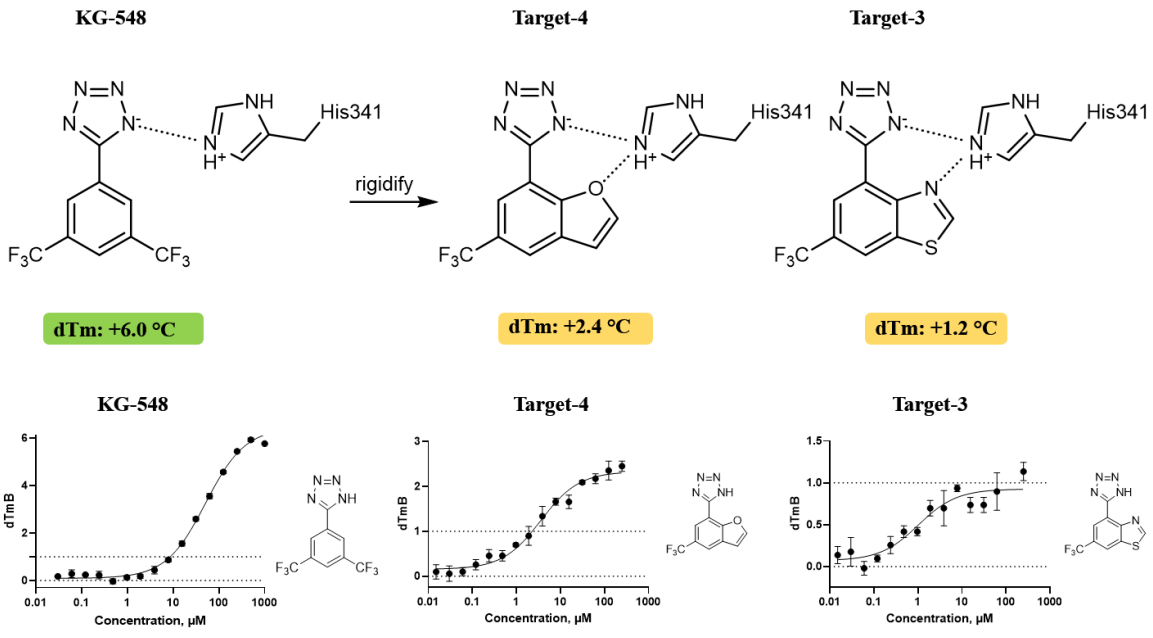
Figure 99. Protonation equilibrium of KG-548 at physiological pH 7.4.

It was hypothesised that the negatively charged tetrazole ring of KG-548 forms a crucial salt bridge with the positively charged side chain of His341 at a distance of 3 Å within the ARNT2 binding pocket (Figure 100A). The nitrogen atoms from the tetrazole ring of KG-548 also form hydrogen bonds with water molecules in the pocket, mediating contacts with the backbone of His341 and the NH_2 group of Gln387. The bis(trifluoromethyl)phenyl ring occupies the hydrophobic region of the binding site.

From the KG-548-bound crystal structure with ARNT2-PAS-B, it was apparent that the ligand does not fully occupy the binding cavity. Binding of KG-548 induces conformational changes that result in an expansion of the pocket. Based on this observation, the first design strategy was to fill the hydrophobic pocket with bicyclic heteroaromatic rings. These rigid, planar systems not only provide a good steric and hydrophobic fit within the pocket, but also have the potential to engage in favourable electronic interactions, to form hydrogen bonds to the His341 side chain. As such, Dr. Ian de Toledo generated and synthesised KG-548 analogues Target-3 and Target-4 (Figure 100B). Both compounds were dissolved in DMSO and serially diluted using a two-fold dilution scheme across 16 concentration points to generate a dose-response curve.


A
His341
2.7
3.0
3.1
2.9
3.2
3.1
Gln387

B

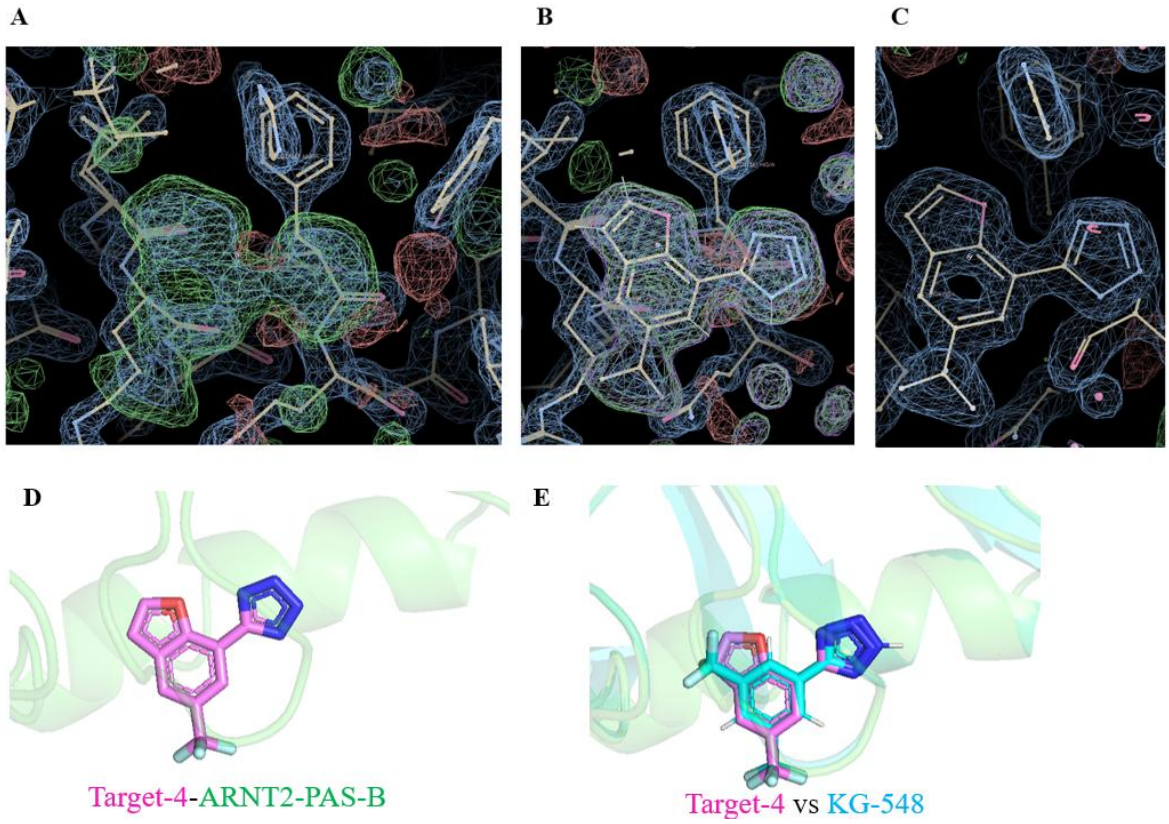

KG-548
Target-4
Target-3
His341
rigidify
dTm: +6.0 °C
dTm: +2.4 °C
dTm: +1.2 °C
KG-548
Target-4
Target-3
dTmB
Concentration, μM

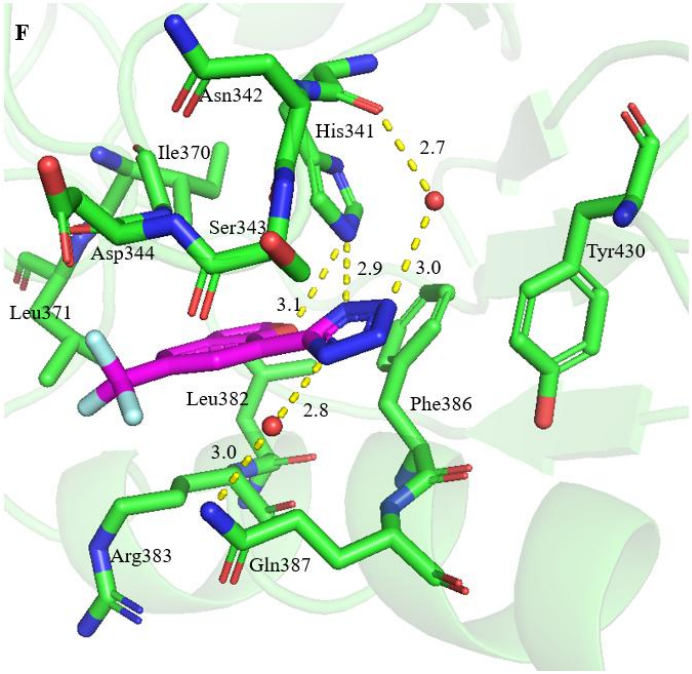
EC_{50} (μM)	
Compound	mARNT2-PAS-B
KG-548	39.5
Target-4	3.59
Target-3	1

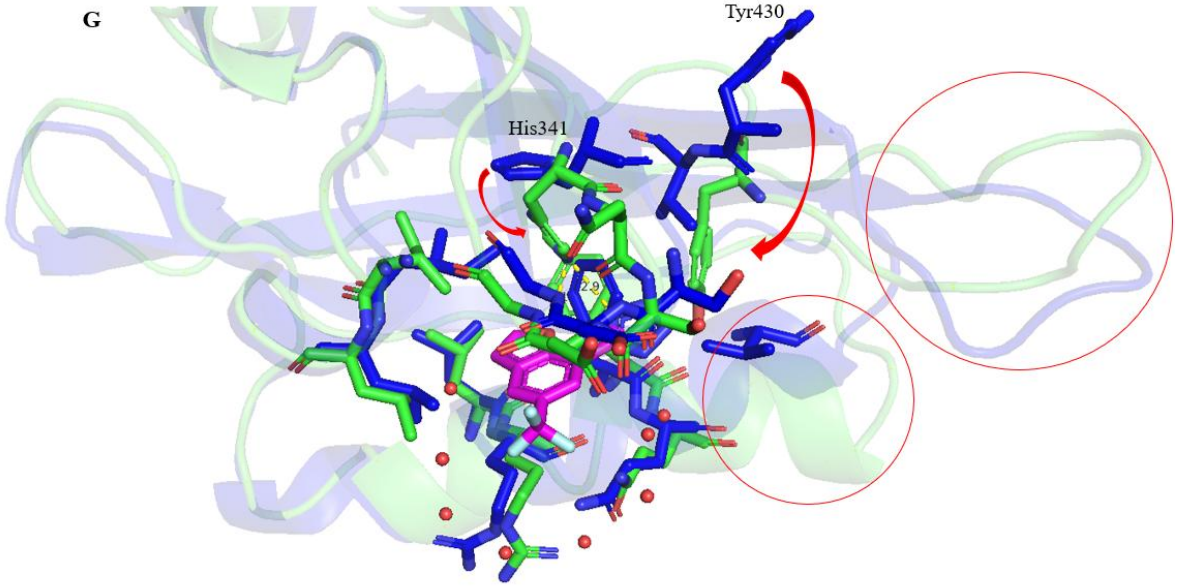
Figure 100. Key interactions of KG-548 and its first analogues.(A) Crystal structure of KG-548 bound to the ARNT2-PAS-B domain, highlighting all key ligand-protein interactions. (B) Initial rigidification strategy leading to the design of Target-4 and Target-3 analogues. The thermal shift (dTm) values for each compound are shown beneath their respective chemical structures. TSA fitting curves are included to illustrate the binding effects of each ligand on the ARNT2-PAS-B domain.

The TSA analysis with mARNT2-PAS-B revealed that both Target-4 and Target-3 compounds bind to the protein with EC_{50} of 3.59 and 1 μM, respectively (Figure 100B). Target-4, however, showed a higher dTmB of 2.4 °C as compared to 1.2 °C for Target-3. Although the EC_{50} values obtained for Target-3 and Target-4 appear more favourable than that of KG-548, it should be noted that EC_{50} values derived from TSA do not directly reflect binding affinity in the same manner as equilibrium-based binding assays. High protein concentrations relative to ligand concentration can cause ligand depletion, and non-equilibrium conditions may inflate EC_{50} values, as observed with ARNT2-PAS-B at 25 μM protein concentration. Other factors, such as slow-binding ligands, limited solubility, chemical instability, or dye interactions, can further distort thermal shift data. Therefore, EC_{50} values from TSA may considerably overestimate the true binding affinity and therefore should be interpreted with caution.

The binding of Target-4 to mARNT2-PAS-B was further confirmed via X-ray crystallography (Figure 95). The Target-4-bound-ARNT2-PAS-B crystal structure was solved at 1.54 Å resolution (Appendix 14). The overlay of the KG-548-bound structure suggests that Target-4 binds the PAS-B domain at the same binding site as KG-548 and other previously characterised compounds (Figure 101 D-E). Similar to KG-548, Target-4 interacts with His341 and Gln387 (Figure 101F). Interacting directly with the key His341, Target-4 produced similar ligand-induced structure changes and expanded the putative ligand-binding pocket (Figure 100G-H). Interestingly, a crystal structure comparison of pocket size between Target-4 and KG-548 bound to the ARNT2-PAS-B domain, revealed that the tunnel with Target-4 shows a wider opening than with KG-548 (Figure 101I). Due to the high chemical similarity between Target-4 and Target-3, it was hypothesised that Target-3 will bind the PAS-B domain in the same manner.







Target-4-ARNT2-PAS-B
Apo-ARNT2-PAS-B

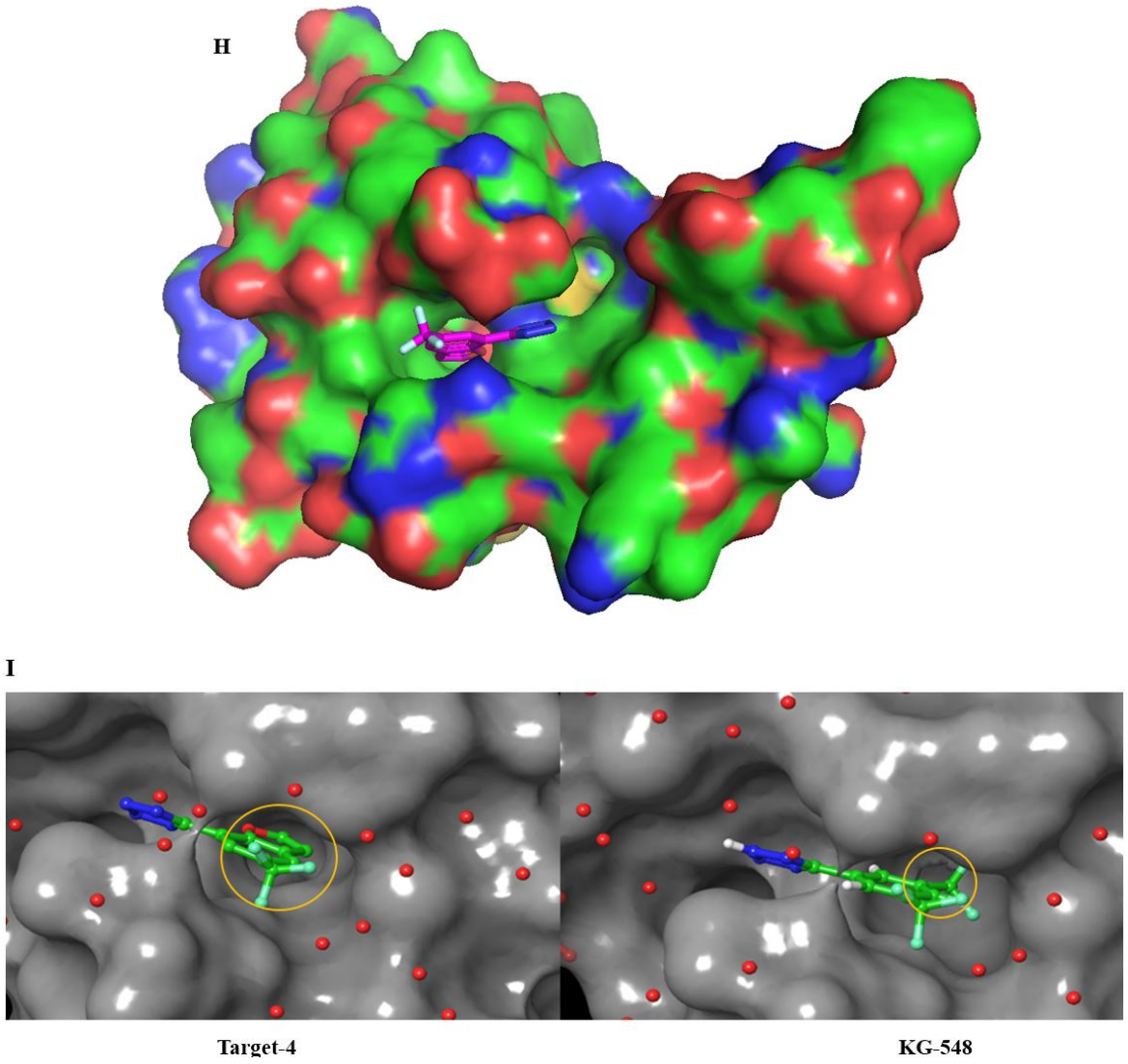


Figure 101. X-ray crystallography for Target-4 binding to ARNT2-PAS-B.(A) Density map without ligand. (B-C) Density map with Target-4 ligand within the ARNT2-PAS-B domain. (D) Close-up look at the location of Target-4 inside the ARNT2-PAS-B domain. Target-4 is shown in pink. (E) The KG-548-bound structure (cyan) was superposed on the Target-4 (pink) to show ligand binding overlap. (F) Interaction of Target-4 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts/interactions with protein residues and water molecules in the crystal structure. (G-H) The Target-4-bound structure (pink) was superposed on the apo ARNT2-PAS-B (blue) to show conformational changes and expansion of the ligand-binding pocket. (I) Crystal structure comparison of pocket size between Target-4 (left) and KG-548 (right) bound to the ARNT2-PAS-B domain.

The tunnel on the left with Target-4 shows a wider opening than that with KG-548, illustrated with an orange circle.

Target-4 and Target-3 were further utilised with several follow-up assays, to investigate whether the analogues have a better affinity profile than the previously examined KG-548 control compound. To access the TR-FRET activity of analogues, mouse ARNT2-eGFP was co-expressed and purified with mNPAS4 and mHIF2α proteins. Since Target-4 binds at the same place and interacts with the same amino acids as the KG-548 control, it was anticipated that Target-4 and, potentially Target-3, would demonstrate a similar activity profile to the KG-548 control. As expected, both Target-4 and Target-3 did not show any activity with either mNPAS4-ARNT2-eGFP or HIF2α-ARNT2-eGFP proteins via TR-FRET (Figure 102).

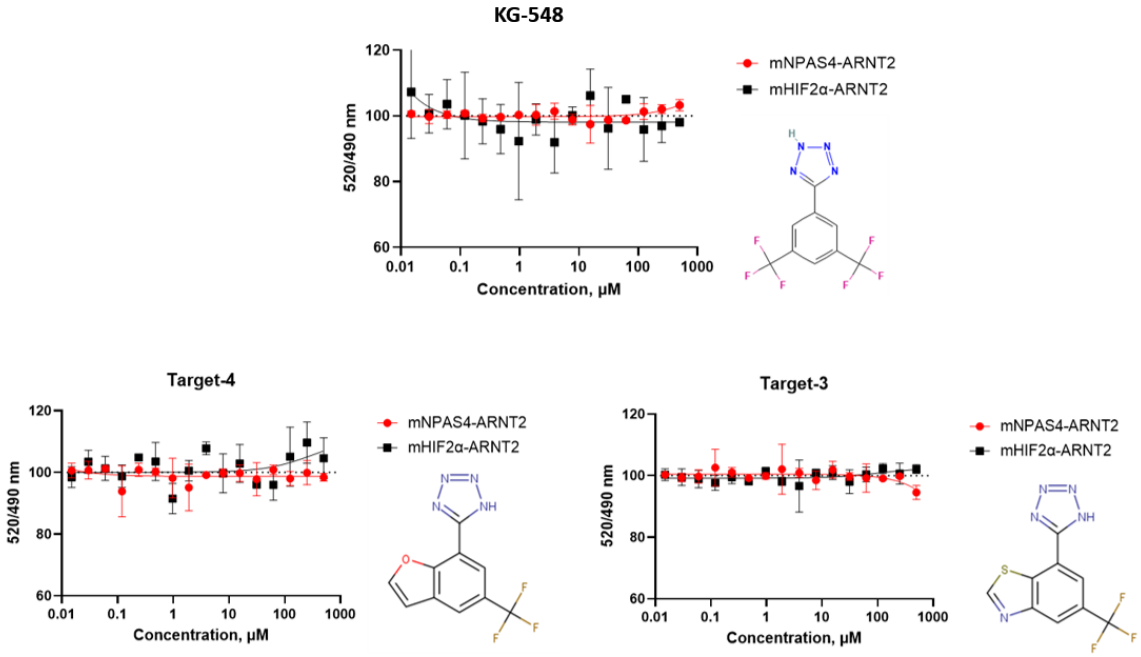


Figure 102. TR-FRET with Target-3, Target-4, and the KG-548 control.Compounds were tested with mNPAS4-ARNT2-eGFP and mHIF2α-ARNT2-eGFP via a 30 min TR-FRET assay.

The binding affinity of both compounds was then further tested by SPR assay. Although KG-548 did not reach saturation, suggesting a K_D > 400 μM, the affinity of both analogues measured via SPR was notably worse than that of the KG-548 control (Figure 103). The SPR data are consistent with the thermal shift results, with both analogues showing considerably lower dTm values: 2.4 °C and 1.4 °C for Target-4 and Target-3, respectively, compared to 6 °C for KG-548 (Figure 102). However, the EC_{50} values obtained from TSA contradict the SPR data, indicating a ligand depletion limitation, particularly at the lower end of the titration range, which is due to the high protein concentration used in TSA studies. Based on these observations, it was decided that subsequently designed analogues would be tested in secondary studies only if their dTm shows more than 4 °C.

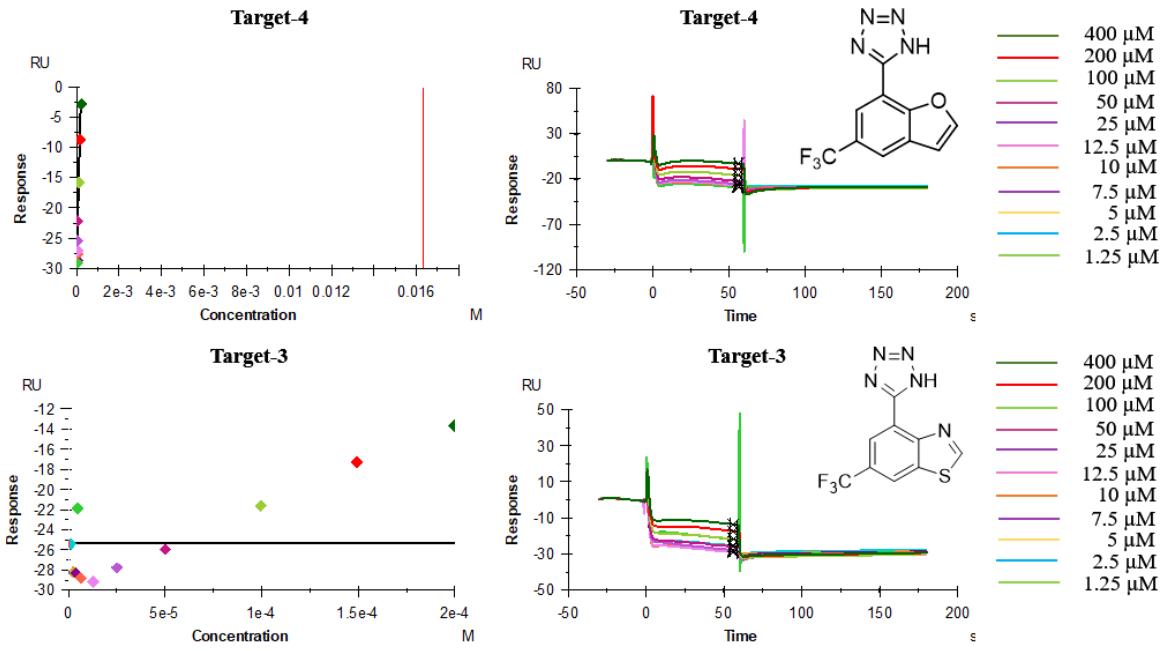


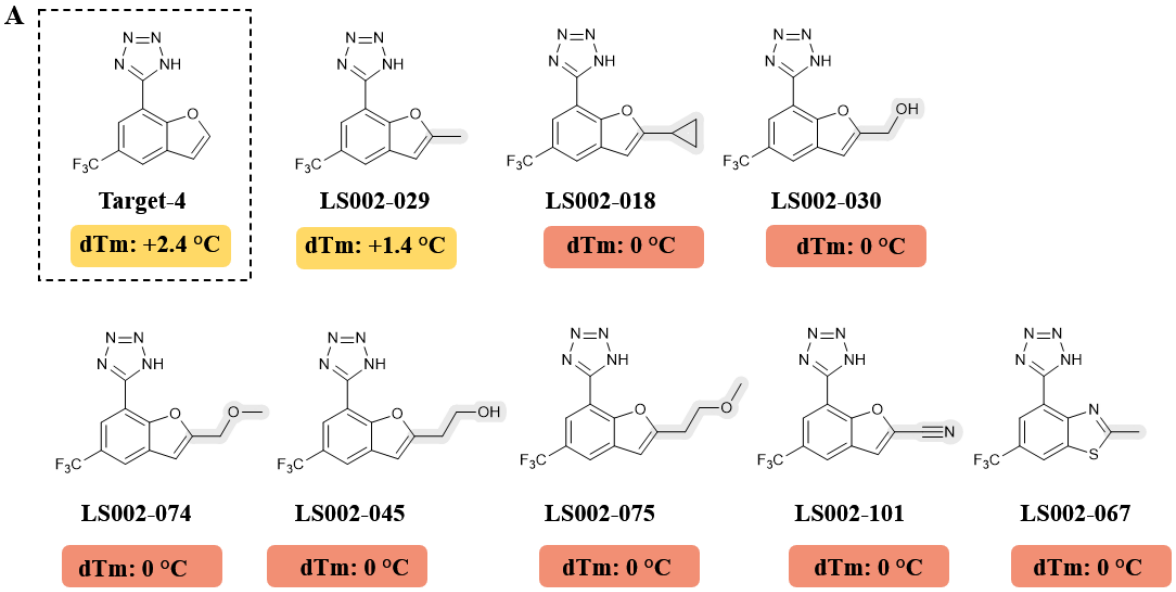
Figure 103. SPR analyses of Target-4 and Target-3 analogues.Steady state affinity of both analogues with the biotinylated single PAS-B domain of ARNT2.

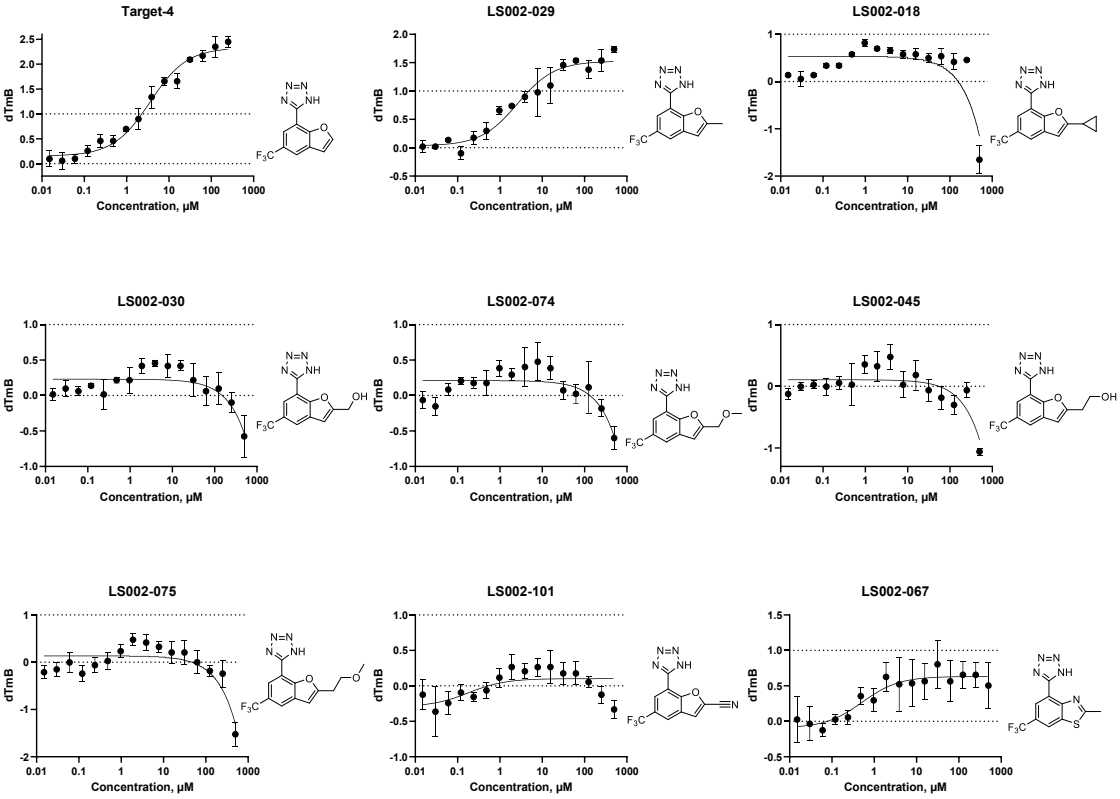
10.2 Expanding a narrow sub-pocket of mARNT2-PAS-B

From the crystal structure of Target-4 (benzofuran scaffold), a narrow sub-pocket located deep within the PAS-B domain appeared to be accessible (Figure 101I). To exploit this feature, the next step in the design strategy was focused on introducing small substituents at the 2-position of the benzofuran scaffold. This modification was intended to allow the molecule to extend into the narrow tunnel leading toward the back pocket of the binding site, potentially enhancing interactions within this otherwise hidden region.

Based on this design strategy, Dr. Lukas Scheibelberger designed and synthesised a series of eight compounds. However, the TSA analysis with the ARNT2-PAS-B domain revealed that most of these compounds did not exhibit any measurable binding (Figure 104A). Although

the sub-pocket appeared accessible in the crystal structure, it is possible that even small substituents at the 2-position of the benzofuran caused steric clashes, either directly within the narrow tunnel, or by inducing unfavourable conformational changes. As such, even a small modification like the addition of the methyl group in LS002-029 appears to be accommodated within the pocket, but it resulted in weaker binding compared to Target-4. The pocket may be more constrained than it appears in a static crystal structure.





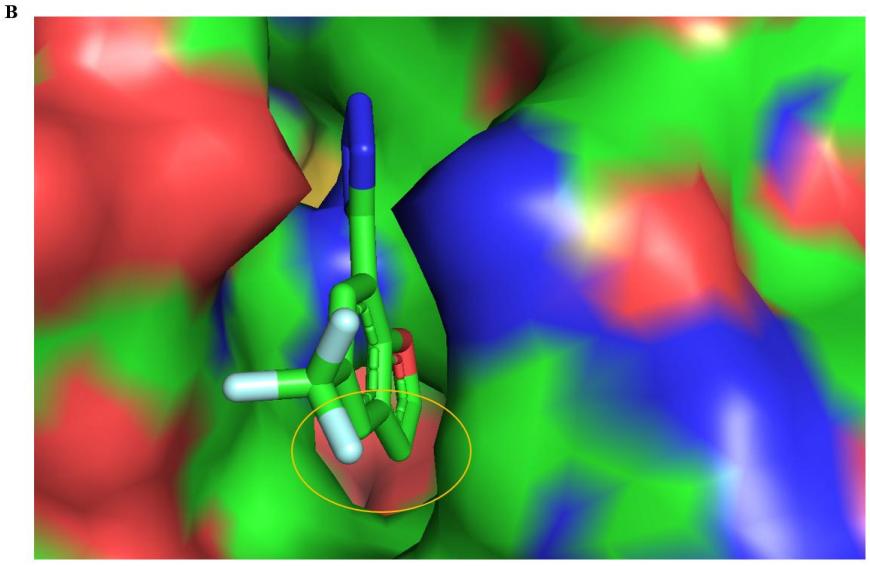


Figure 104. KG-548 analogues based on modifications of the benzofuran scaffold to address the back pocket strategy.(A) The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with

the ARNT2-PAS-B domain. (B) A close-up view of the Target-4 crystal structure, suggesting a hydrophilic environment of the sub-pocket.

Furthermore, a close-up view of the Target-4 crystal structure suggests that the narrow tunnel may be hydrophilic in nature (Figure 104B). The presence of red colouring in PyMOL indicates a polar or negatively charged environment, potentially lined with acidic residues or water molecules. Introducing a substituent into this region could require displacing ordered water molecules or overcoming an energetic barrier. If the gain in binding energy does not compensate for the cost of desolvation, binding may not be favourable.

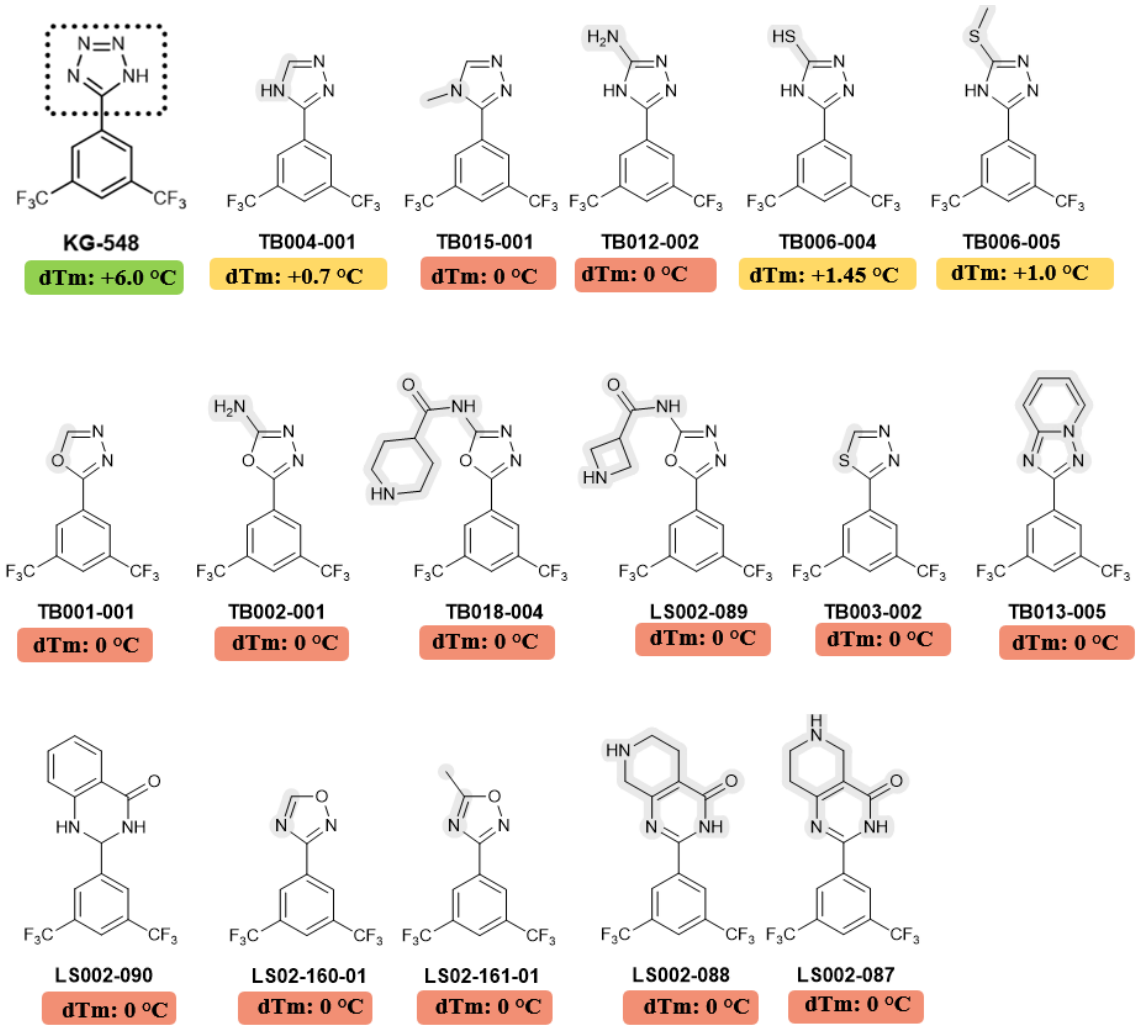
Among the series, only compound LS002-029 showed a positive thermal shift, with a dTmB of 1.4 °C. This modest stabilisation was, however, significantly lower than that observed for Target-4 and KG-548, suggesting that LS002-029 is unlikely to display an improved binding affinity compared to these reference compounds. Therefore, the LS002-029 analogue was not advanced for secondary assay trials, leading to the eventual discontinuation of this chemical-design approach.

10.3 Attaching substituents to the tetrazole ring

The crystal structure of KG-548 bound to the ARNT2-PAS-B domain revealed that a significant portion of the binding pocket remains unoccupied. This observation initially suggested an opportunity to improve binding affinity by designing ligands that extend further into the vacant region of the pocket. Attaching substituents to the tetrazole ring of KG-548 to grow into the pocket already showed negative results in the initial publication from the

Gardner lab.[159] As an alternative strategy, it was decided to find other small heterocycles that would allow them to grow into the rest of the pocket while maintaining the interaction with the sidechain of histidine.

A series of small heterocycles was designed by Dr. Lukas Scheibelberger and Tobias Bernard with the idea of retaining potential interactions with His341 while avoiding the acidic character of the tetrazole ring. The TSA analysis, however, did not show competitive binding of these analogues (Figure 105). These results suggest that the interaction between the tetrazole and His341 is not a generic hydrogen bond, but instead a key electrostatic interaction, most likely a salt bridge. Given that tetrazoles are predominantly deprotonated and carry a negative charge at physiological pH (~7.4), these findings support the preliminary hypothesis that the tetrazole ring engages in a crucial ionic interaction with a protonated histidine side chain. Thus, the presence of a negative charge appears to be essential for binding, and its removal in the new analogues likely accounts for the observed loss of affinity.


KG-548
dTm: +6.0 °C
TB004-001
dTm: +0.7 °C
TB015-001
dTm: 0 °C
TB012-002
dTm: 0 °C
TB006-004
dTm: +1.45 °C
TB006-005
dTm: +1.0 °C
TB001-001
dTm: 0 °C
TB002-001
dTm: 0 °C
TB018-004
dTm: 0 °C
LS002-089
dTm: 0 °C
TB003-002
dTm: 0 °C
TB013-005
dTm: 0 °C
LS002-090
dTm: 0 °C
LS02-160-01
dTm: 0 °C
LS02-161-01
dTm: 0 °C
LS002-088
dTm: 0 °C
LS002-087
dTm: 0 °C

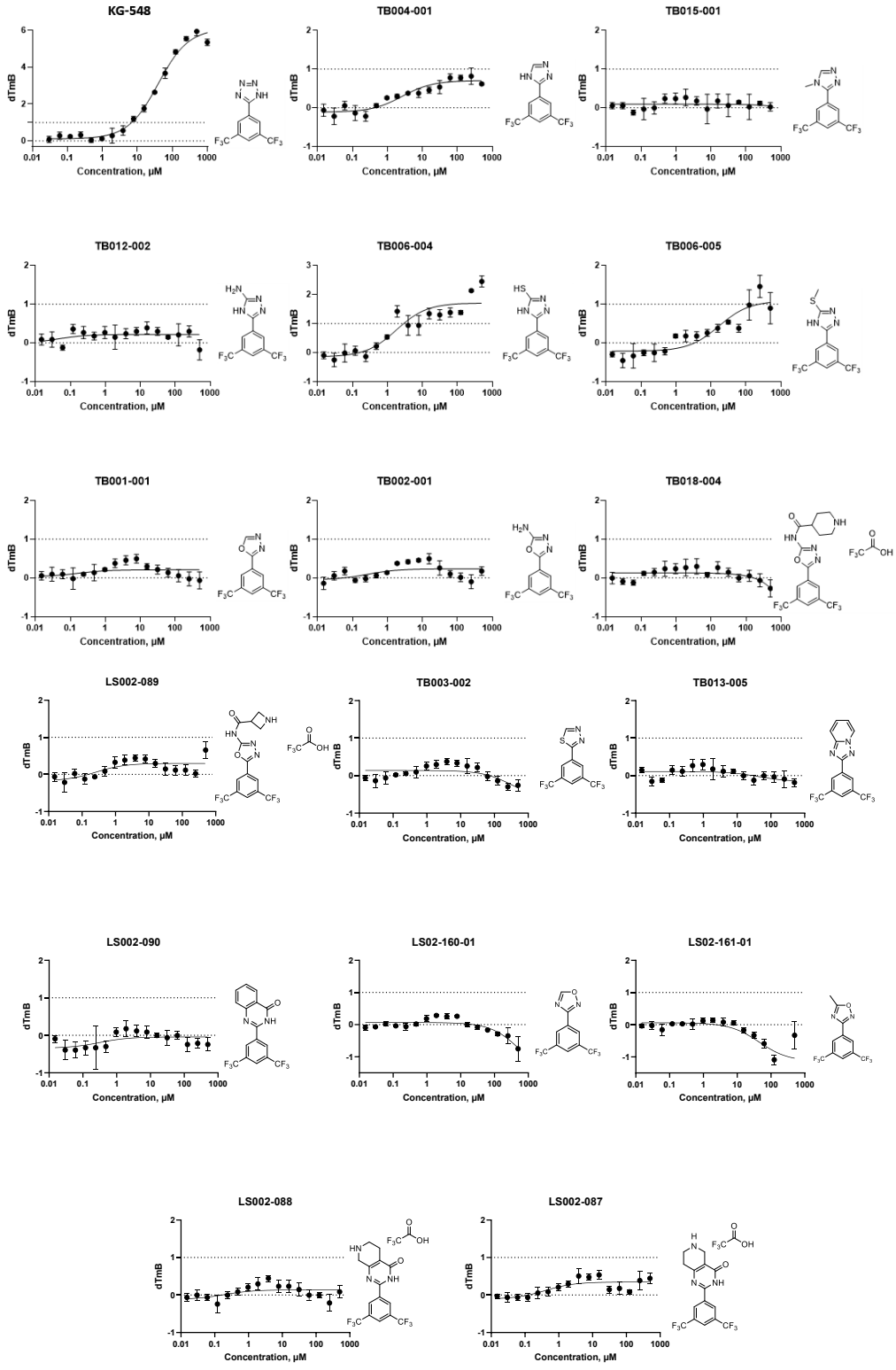


Figure 105. KG-548 analogues based on modifications of the heterocyclic hydrogen bond acceptors.The thermal shift (dTm) for each compound is shown below its corresponding

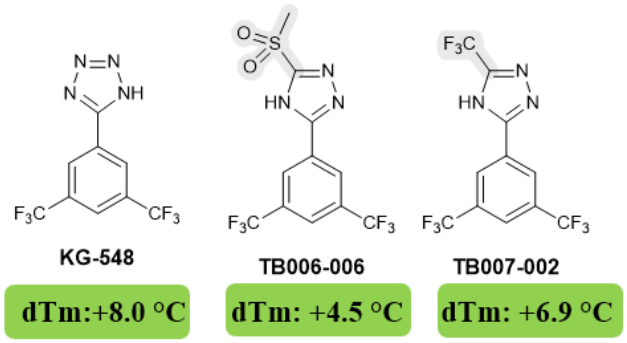
chemical structure, alongside TSA curves depicting the binding effect with the ARNT2-PAS-B domain.

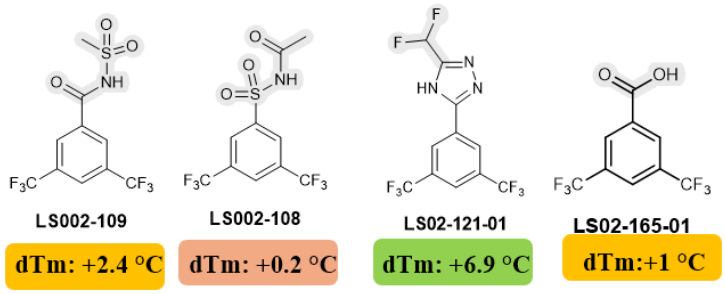
10.4 Acid/tetrazole bioisosteres analogues

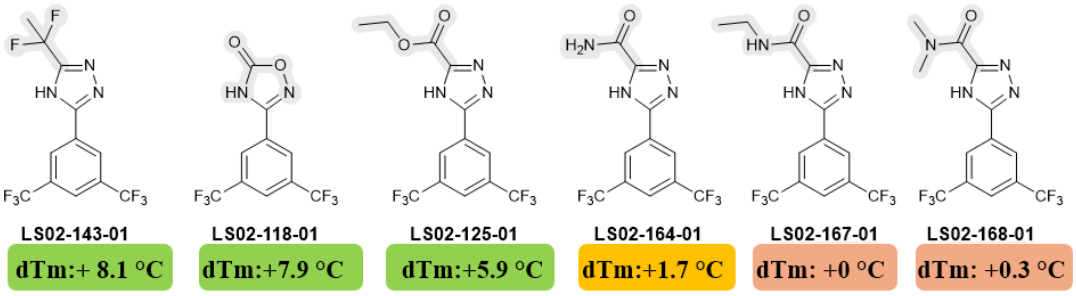
After exploring several analogue design strategies, it became clear that an acidic moiety or ring had to be incorporated into the molecule to enable the formation of a salt bridge. Although the tetrazole group plays a crucial role in binding, there are multiple reasons to investigate potential alternatives. Tetrazoles can present challenges such as poor metabolic stability, low membrane permeability due to high polarity, or synthetic complexity.[219,220] In some cases, they may also have limited selectivity or undesirable pharmacokinetic properties.[221] Therefore, finding suitable bioisosteres that preserve the required acidity and binding interactions, while improving drug-like properties, is a common strategy in lead optimisation.

Several reports have described acidic triazoles bearing strong electron-withdrawing groups such as $-CF_3$ or $-SO_2Me$.[222,223] This led to the design of compounds TB007-002 and TB006-006 by Tobias Bernard (Figure 106). Both compounds exhibited a promising thermal shift, with dTmB of 6.9 °C and 4.5 °C for TB007-002 and TB006-006, respectively. As such, an additional set of acidic compounds was designed by Dr. Lukas Scheibelberger. Follow-up compounds such as LS02-121-01, LS002-143-01, and LS02-125-01 also demonstrated that $-CF_2H$, -CF2CH3 substituents, or an ester could withdraw electron density enough to make the triazole proton sufficiently acidic (Figure 106). Unfortunately, the amide derivatives,

such as LS02-164-01, LS02-167-01, and LS02-168-01, were not sufficiently electron-withdrawing and therefore were inactive (Figure 106).







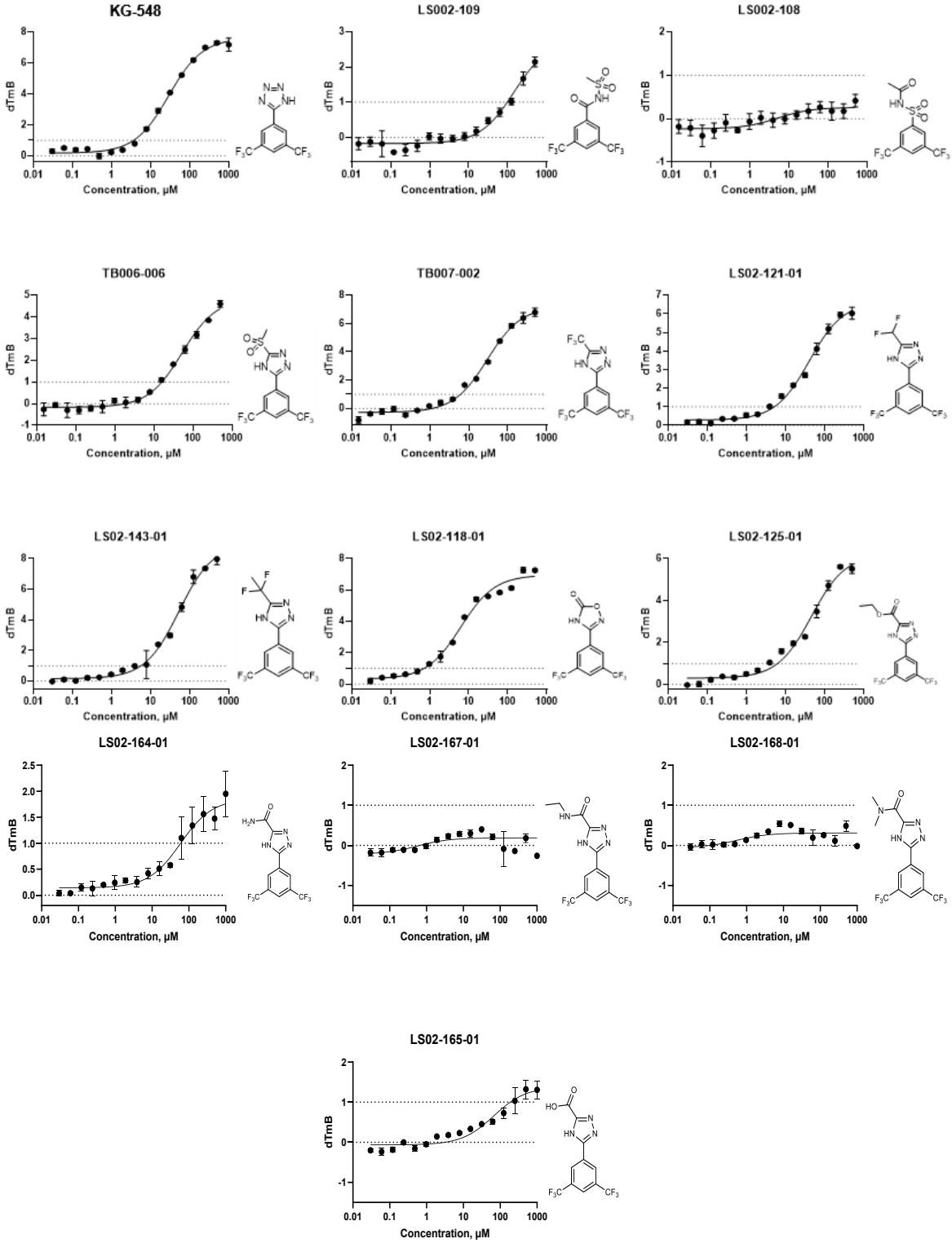


Figure 106. KG-548 analogues based on modifications with acid bioisosteres.The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with the ARNT2-PAS-B domain.

The analogues TB007-002, LS02-121-01, and LS02-143-01 are structurally similar. The sole difference lies in the substituent attached to the triazole nitrogen: a trifluoromethyl group (-CF_3) in TB007-002, a difluoromethyl group (-CHF_2) in LS02-121-01, and a 1,1-difluoroethyl group (-CF_2CH_3) in LS02-143-01. These variations slightly decrease the electron-withdrawing strength and increase steric bulk and lipophilicity, which should influence compounds' physicochemical and biological properties. Both TB007-002 and LS02-121-01 showed the same dTm of 6.9 °C in TSA studies. This suggests that the difference in electron-withdrawing strength between -CF_3 and -CHF_2 may not be large enough to significantly change the binding interaction, especially if other forces like hydrophobicity or fit are similar.

The binding of LS02-121-01 was confirmed via X-ray crystallography (Figure 107 and Appendix 14), which revealed that it occupies the same binding site on mARNT2-PAS-B and engages the same key residues as KG-548 and related compounds (Figure 107E-F). Furthermore, the difluoromethyl group forms direct contact with Ser345 at a distance of 2.9 Å, along with several additional hydrogen bonds with water molecules, further stabilising the ligand within the pocket. Although no crystal structure was obtained for TB007-002, its similar chemical structure suggests that it would bind in a comparable manner and have a similar fit to LS02-121-01, supporting the TSA results.

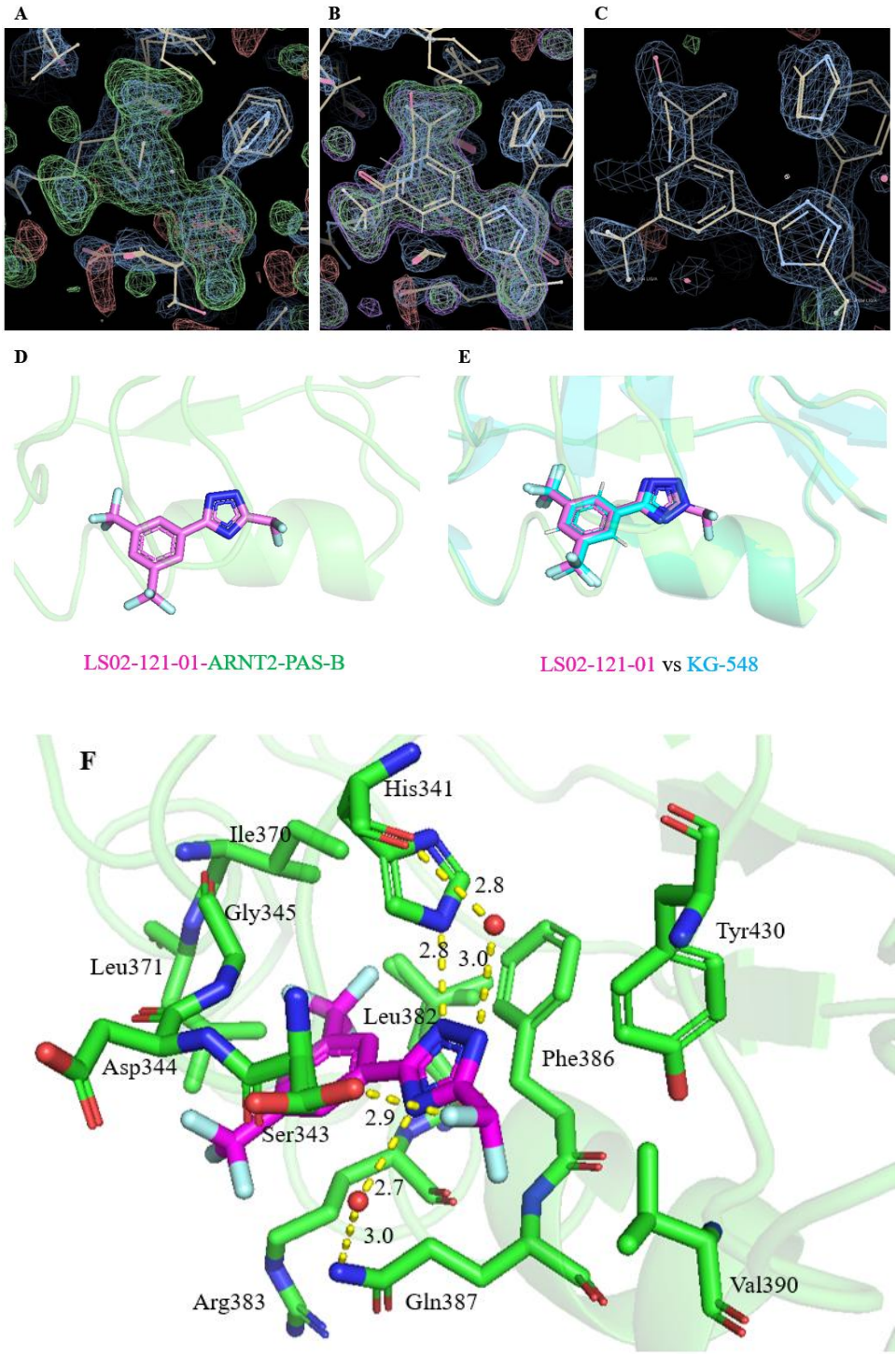


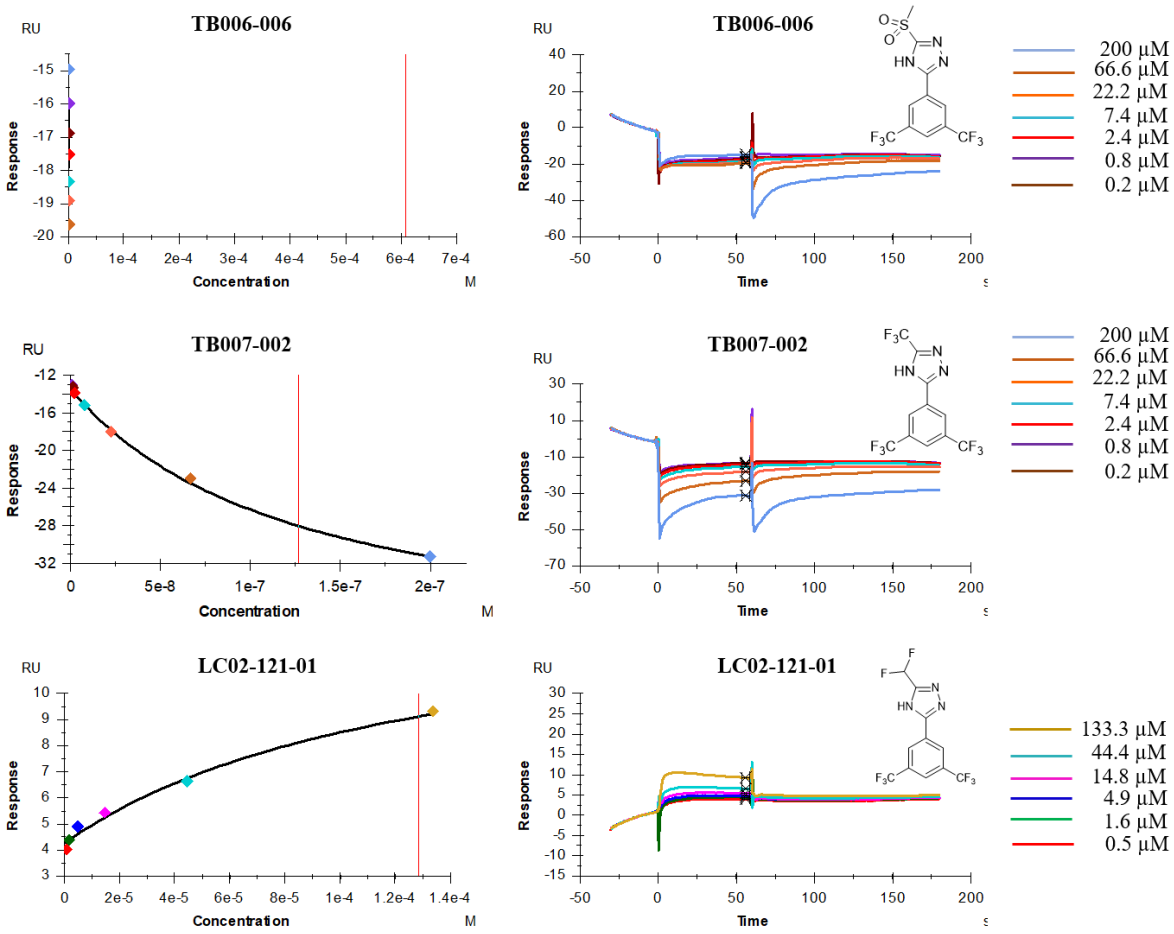
Figure 107. X-ray crystallography for LS02-121-01 binding to ARNT2-PAS-B. (A) Density map without ligand. (B-C) Density map with the LS02-121-01 ligand within the ARNT2-PAS-B domain. (D) Close-up look at the location of LS02-121-01 inside the ARNT2-PAS-B domain. LS02-121-01 is shown in pink. (E) The KG-548-bound structure (cyan) was

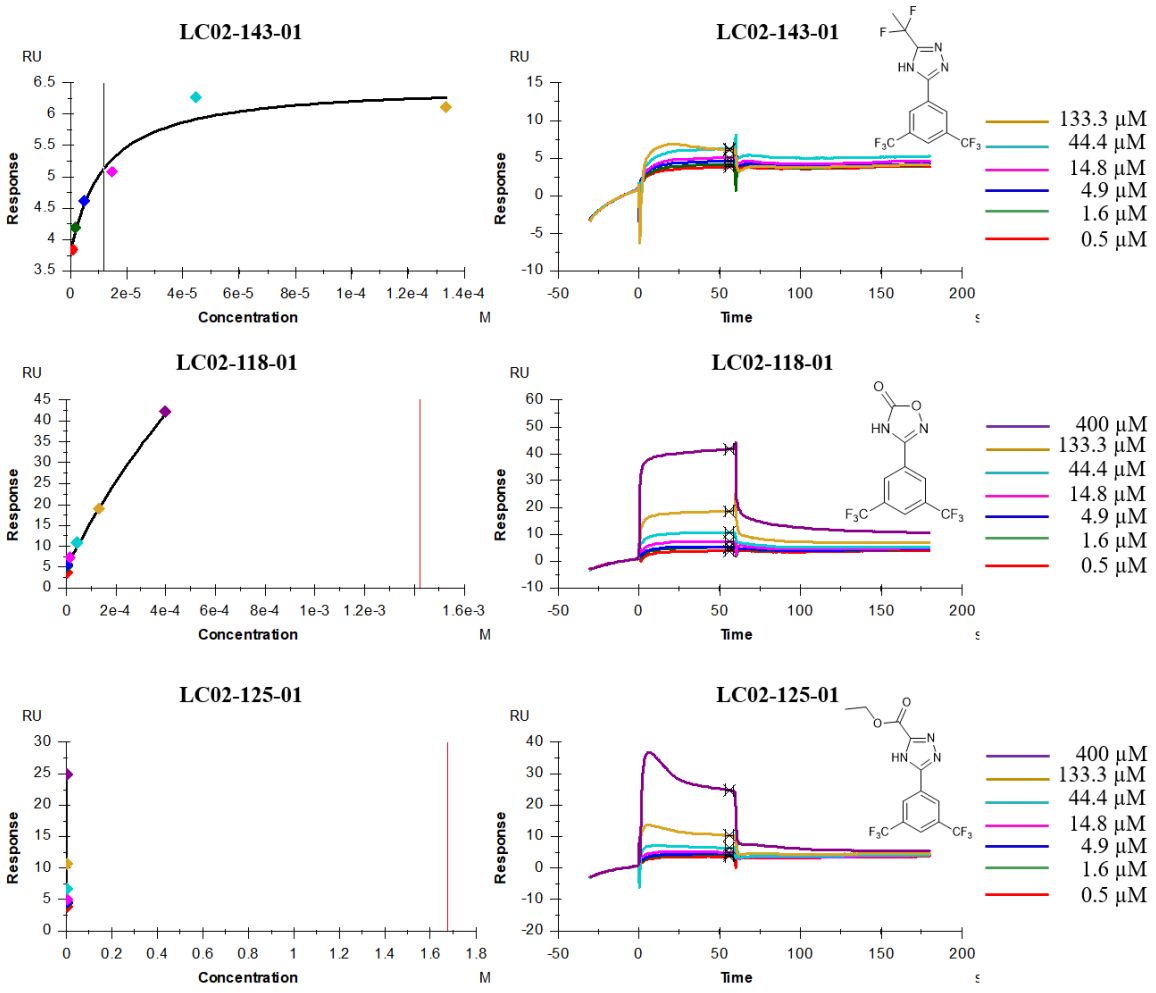
superposed on LS02-121-01 (pink) to show ligand binding overlap. (F) Interaction of LS02-121-01 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts/interactions with protein residues and water molecules in the crystal structure.

Interestingly, the -CF_2CH_3 group of LS02-143-01 has a weaker electron-withdrawing nature than the trifluoromethyl group (-CF_3) of TB007-002. The TSA studies, however, exhibited a higher thermal shift with LS02-143-01 (dmT: +8.1 °C vs. 6.9 °C). This suggests that factors beyond triazole acidity contribute significantly to binding affinity. While stronger electron-withdrawing groups can enhance triazole proton acidity and facilitate hydrogen bonding, an increased hydrophobicity, steric complementarity, and favourable conformational effects introduced by the bulkier -CF_2CH_3 substituent could strengthen the overall ligand-protein interactions. Based on the LS02-121-01 crystal structure, the presence of an additional methyl group in LS02-143-01 may enable extra interactions within the pocket, which in turn may enhance thermal stabilisation. Thus, the increased thermal shift in LS02-143-01 may be attributed to a better overall binding profile due to hydrophobic interactions, better fit, or conformational advantages, despite the slightly weaker acidity.

Although the steady-state affinity analysis showed that both LS02-121-01 and LS02-143-01 reached saturation as compared to other derivatives in this set, the average RU signals were relatively low. The calculated K_D values should therefore be interpreted with caution (Figure 108). Compounds that exhibited promising TSA shifts with more than 4.0 °C for dTmB, including TB006-006, TB007-002, LS02-121-01, LS02-143-01, LS02-118-01, and LS02-

125-01, were also tested with the TR-FRET assay. As predicted, TR-FRET did not show any activity with either NPAS4-ARNT2, nor with HIF2α-ARNT2 heterodimers (Figure 109).





Compound	KD (M)
LS02-121-01	128.5
LS02-121-01	11.9

Figure 108. Steady state affinity analysis and calculated K_D for several KG-548 analogues with the biotinylated PAS-B domain of ARNT2.

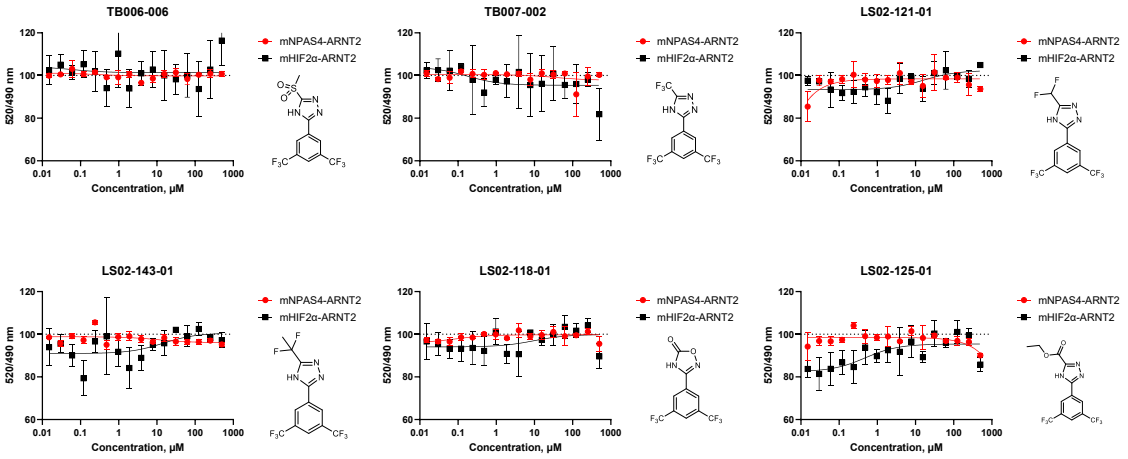
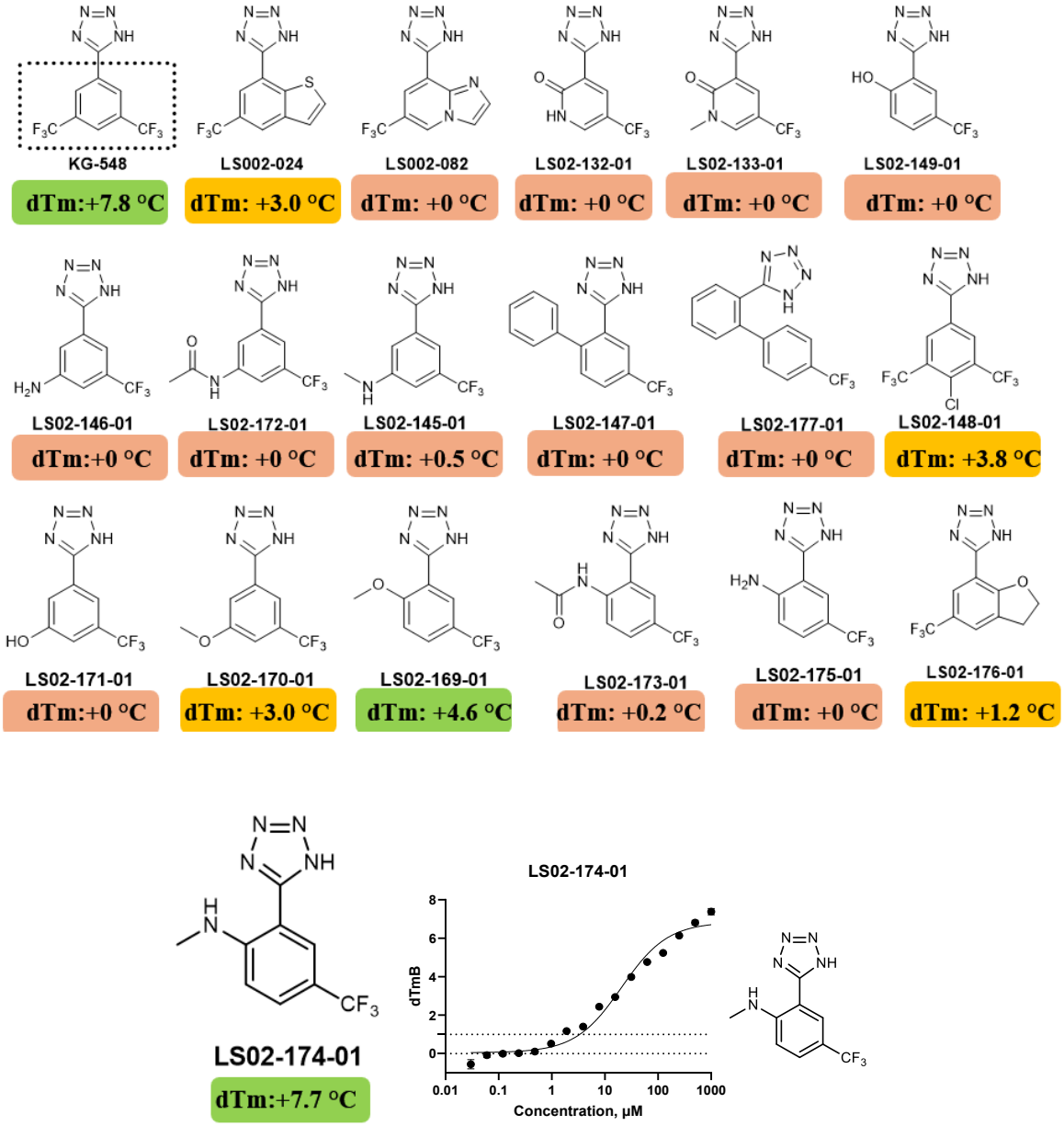


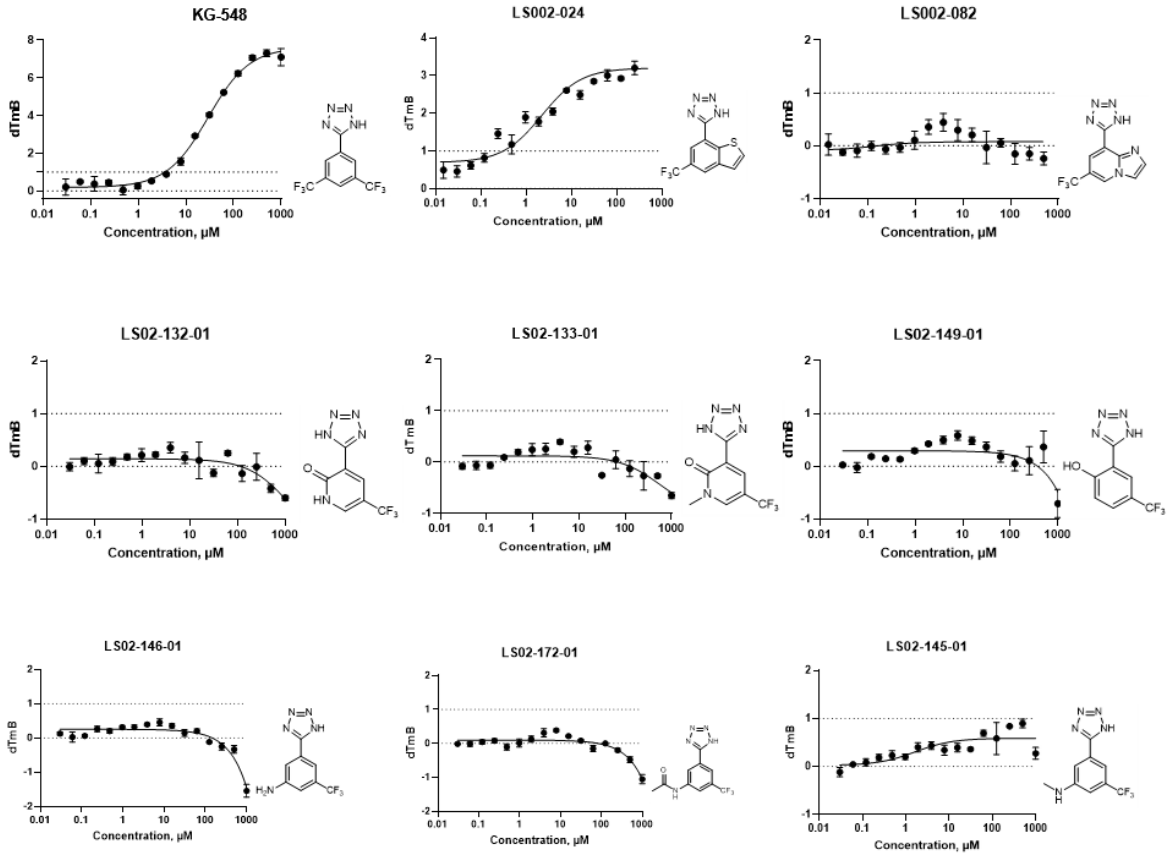
Figure 109. TR-FRET with several KG-548 analogues.

10.5 Aryl ring modifications

The aryl (bis-trifluoromethylphenyl) ring in KG-548 likely plays a key role in hydrophobic interactions, shaping complementarity within the binding pocket. There may be unexploited opportunities for increased target engagement or improved selectivity through structural diversification of this region. The approach to designing new analogues focused on systematically evaluating alternative aryl substitutions to determine if binding affinity could be maintained. Additionally, it sought to investigate other regions of the binding pocket, targeting new residues via hydrogen bonding, electrostatic, or halogen interactions, while also aiming to enhance physicochemical properties such as solubility and metabolic stability. A set of analogues was designed by Dr. Lukas Scheibelberger with the goal of enhancing the binding affinity of the analogues (Figure 110).


KG-548
dTm:+7.8 °C
LS002-024
dTm: +3.0 °C
LS002-082
dTm: +0 °C
LS02-132-01
dTm: +0 °C
LS02-133-01
dTm: +0 °C
LS02-149-01
dTm: +0 °C
LS02-146-01
dTm:+0 °C
LS02-172-01
dTm: +0 °C
LS02-145-01
dTm: +0.5 °C
LS02-147-01
dTm: +0 °C
LS02-177-01
dTm: +0 °C
LS02-148-01
dTm: +3.8 °C
LS02-171-01
dTm:+0 °C
LS02-170-01
dTm: +3.0 °C
LS02-169-01
dTm: +4.6 °C
LS02-173-01
dTm: +0.2 °C
LS02-175-01
dTm: +0 °C
LS02-176-01
dTm: +1.2 °C

LS02-174-01
dTm:+7.7 °C
LS02-174-01
dTmB
Concentration, μM


KG-548
LS002-024
LS002-082
LS02-132-01
LS02-133-01
LS02-149-01
LS02-146-01
LS02-172-01
LS02-145-01
dTmB
Concentration, µM

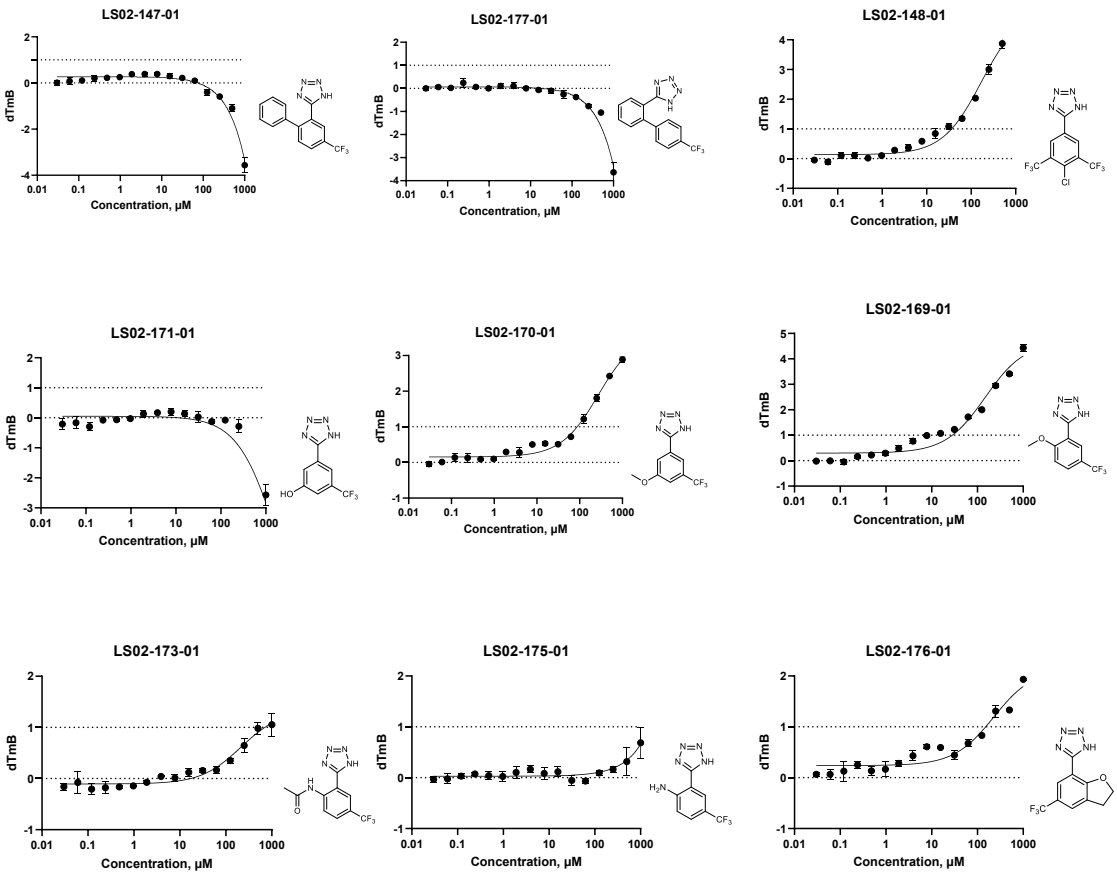


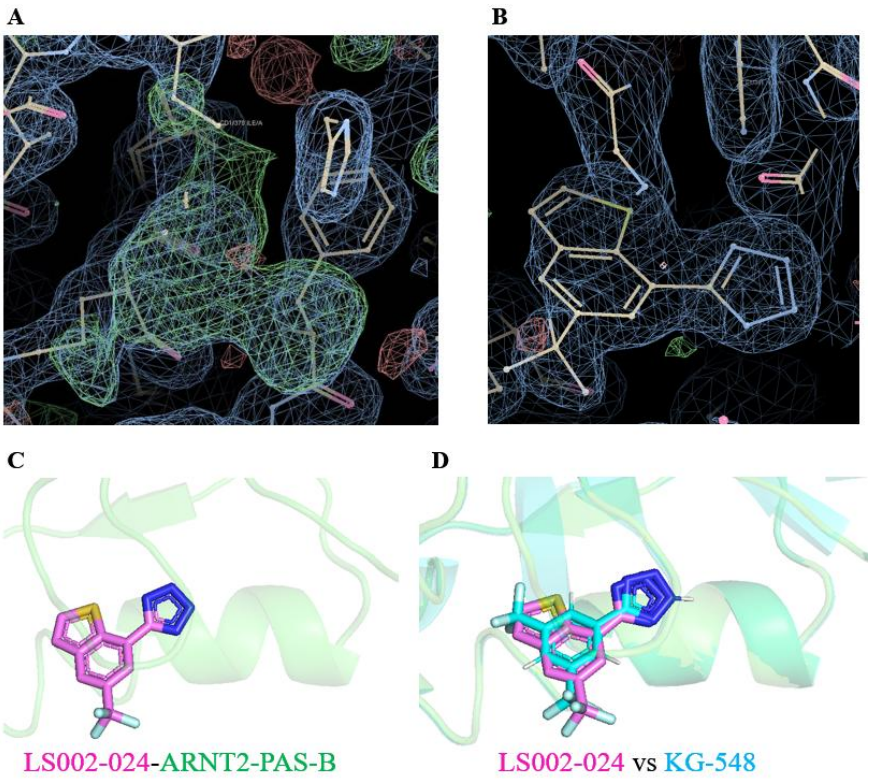
Figure 110. KG-548 analogues based on aryl ring modifications. The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with the mARNT2-PAS-B domain.

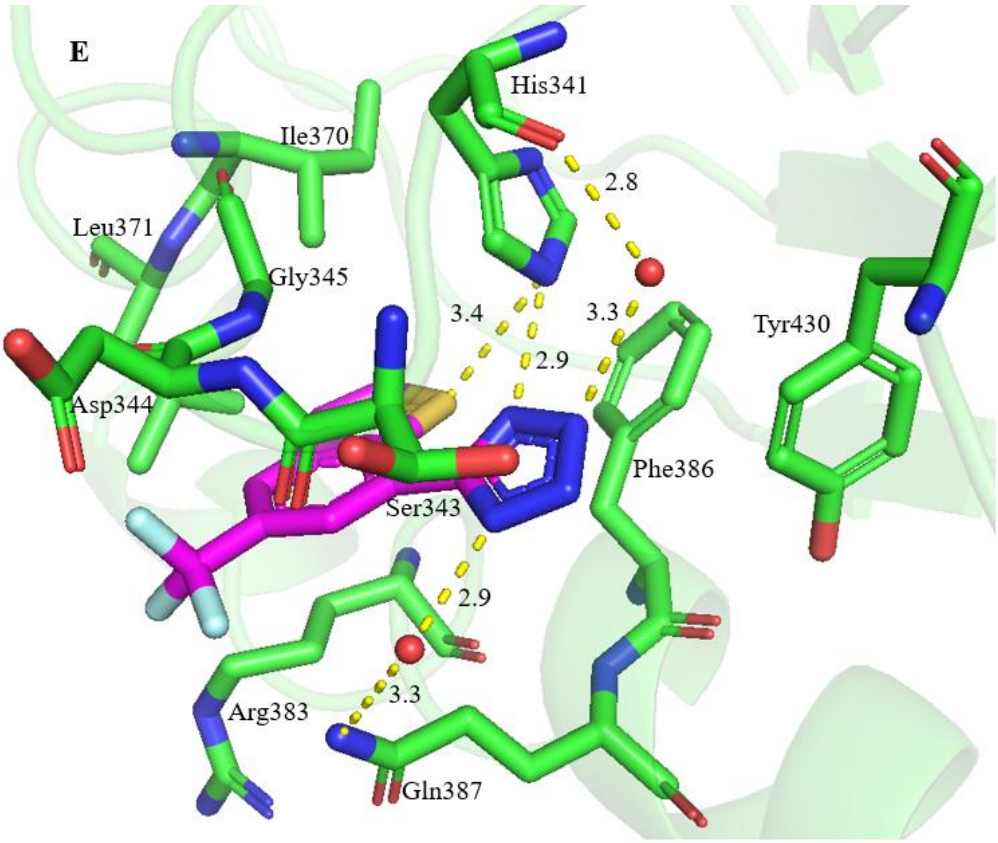
Mixed results were achieved with analogues based on aryl ring modifications. While most of the compounds did not show binding in TSA studies, several analogues demonstrated promising thermal shifts, including LS002-024, LS02-148-01, LS02-169-01, and LS02-170-01 (Figure 110). The best results showed the LS02-174-01 analogue with dTm of 7.7 °C, which is similar to the dTm of the KG-548 control compound.

Interestingly, LS002-024 is structurally similar to Target-4. The main difference lies in the fused heterocycle: LS002-024 contains a benzothiophene core (sulphur), while Target-4 has a benzofuran core (oxygen). This single-atom substitution alters electronic and physicochemical properties, potentially impacting binding, activity, or metabolic stability.

Based on the crystal structure of LS002-024 bound to ARNT2-PAS-B, the ligand appears to occupy a similar binding site and interact with the same key amino acid residues as Target-4 (Figure 111A-E, and Appendix 14). Superimposition of both ligand-protein complexes reveals that LS002-024 and Target-4 induce comparable conformational changes in the ARNT2-PAS-B domain, suggesting a conserved binding mode despite slight structural differences between the two compounds (Figure 111F).

Both LS002-024 and Target-4 also display comparable biological profiles. Similar to Target-4, LS002-024 did not show any TR-FRET activity with either mNPAS4 or mHIF2α in complexes with ARNT2 (Figure 112). A benzothiophene moiety appears to slightly enhance LS002-024's binding affinity to the ARNT2-PAS-B domain, as evidenced by a modest improvement in binding strength compared to Target-4 (Figure 113). However, in both cases, binding remained weak (K_D > 400 μM) and non-saturating, indicating that this region of the molecule plays a limited role in driving high-affinity interaction. These findings suggest that modification of the fused heterocyclic core may have fine-tune binding affinity, but more substantial structural changes are likely needed to achieve functional activity.


A
B
C
D
LS002-024-ARNT2-PAS-B
LS002-024 vs KG-548


E
His341
Ile370
Leu371
Gly345
Asp344
Ser343
Arg383
Gln387
Phe386
Tyr430
2.8
3.4
3.3
2.9
2.9
3.3

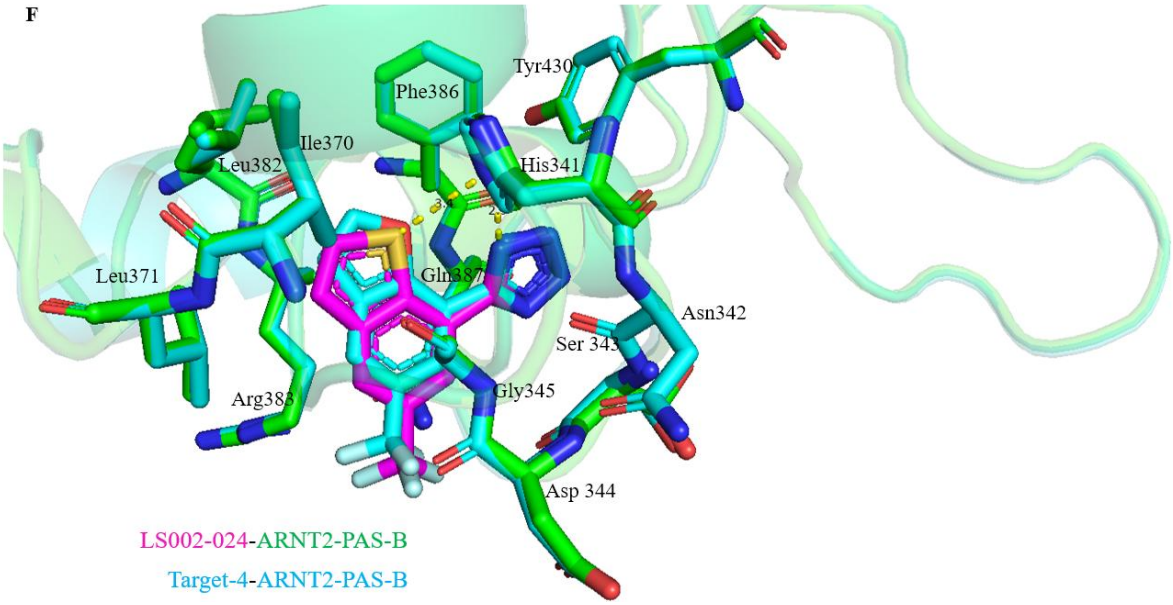


Figure 111. X-ray crystallography for LS002-024 binding to ARNT2-PAS-B. (A) Density map without ligand. (B) Density map with the LS002-024 ligand within the ARNT2-PAS-B domain. (C) Close-up look at the location of LS002-024 inside the ARNT2-PAS-B domain. LS002-024 is shown in pink. (D) The KG548-bound structure (cyan) was superposed on LS002-024 (pink) to show ligand binding overlap. (E) Interaction of LS002-024 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts/interactions with protein residues and water molecules in the crystal structure. (F) The Target-4-bound structure (cyan) was superposed on LS002-024 (pink) to show ligand binding overlap.

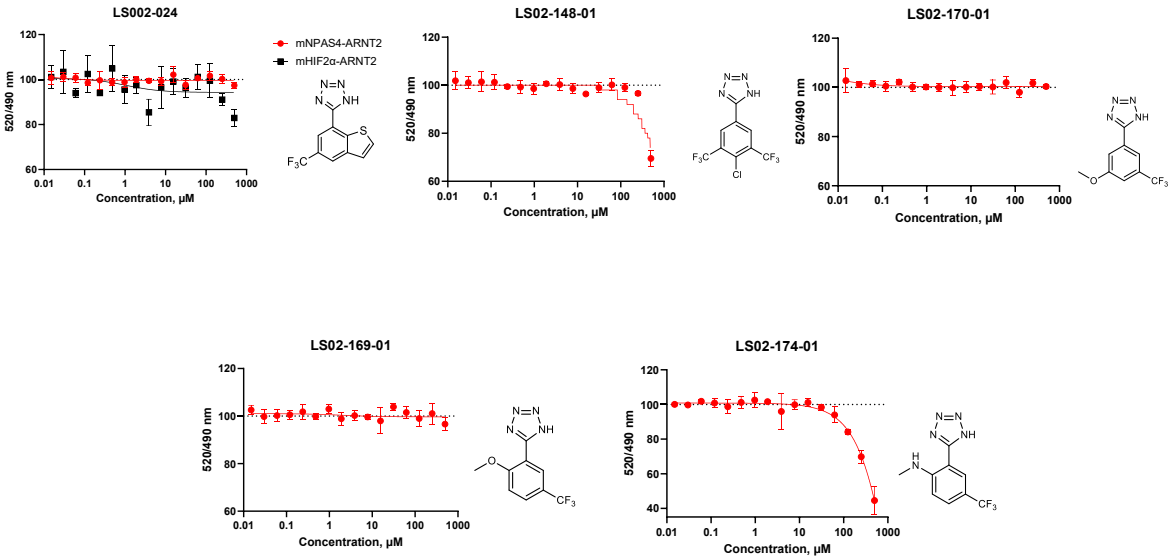
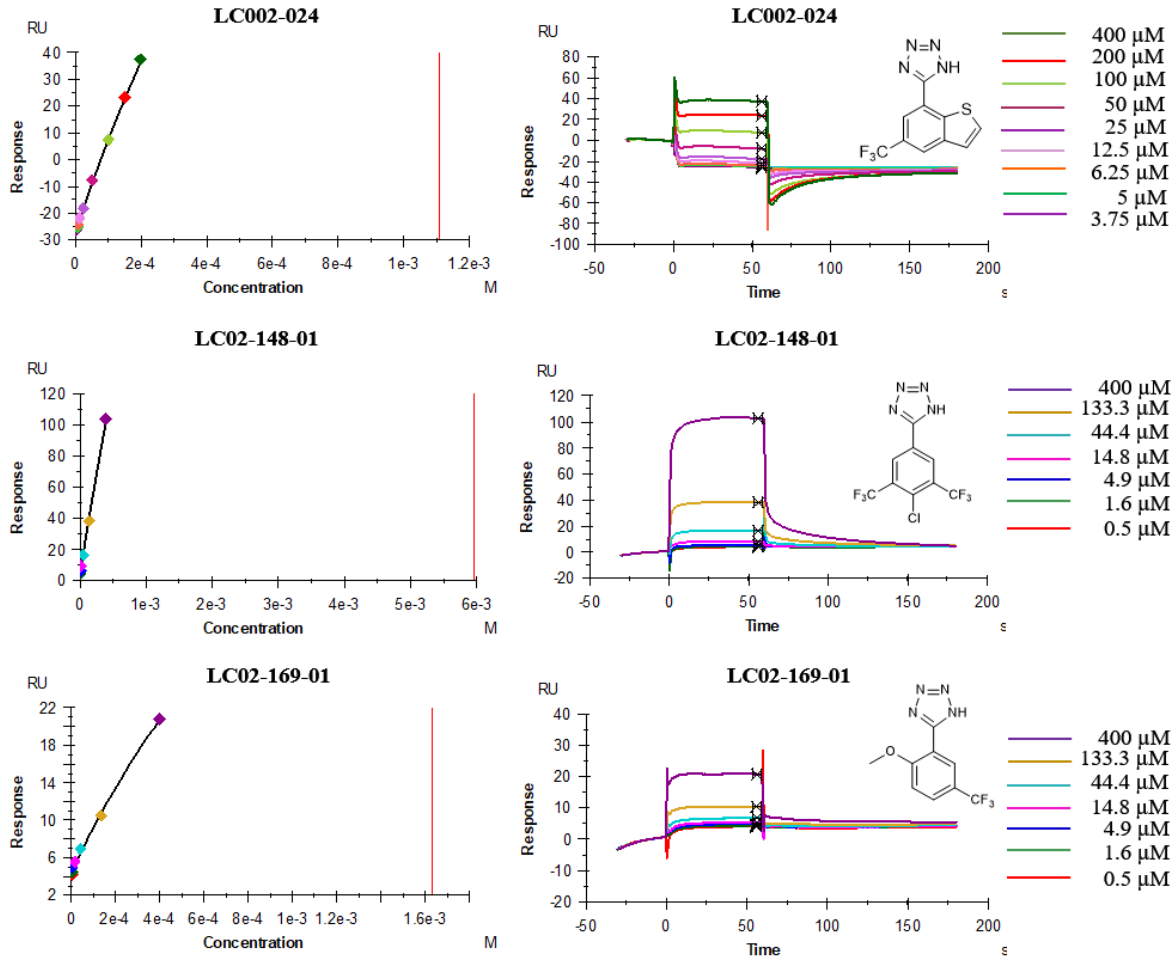


Figure 112. TR-FRET with several KG-548 analogues progressed from the TSA analyses.



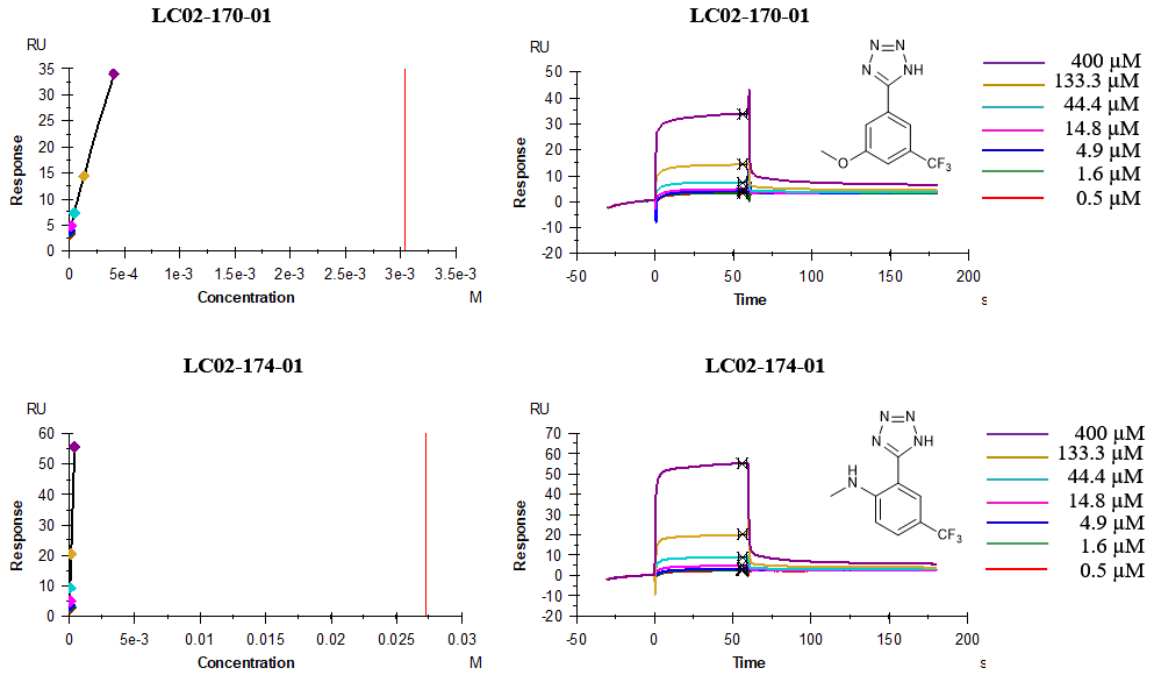


Figure 113. Steady state affinity analyses of several KG-548 analogues with the biotinylated PAS-B domain of ARNT2.

Two analogues, LS02-169-01 and LS02-170-01, are almost identical and differ only by the position of the methoxy (-OCH_3) substituent on the aromatic ring (Figure 114). In LS02-169-01, it is ortho to the tetrazole group, while in LS02-170-01 it is meta to it. Binding of LS02-170-01 was further confirmed by X-ray crystallography (Figure 115 and Appendix 15). As expected, the ligand occupied the same binding pocket within the PAS-B domain and formed interactions comparable to those of KG-548 and other analogues. Consistent with previous data, LS02-170-01 showed no TR-FRET activity, and SPR analysis revealed no significant improvement in binding affinity (Figures 112-113).

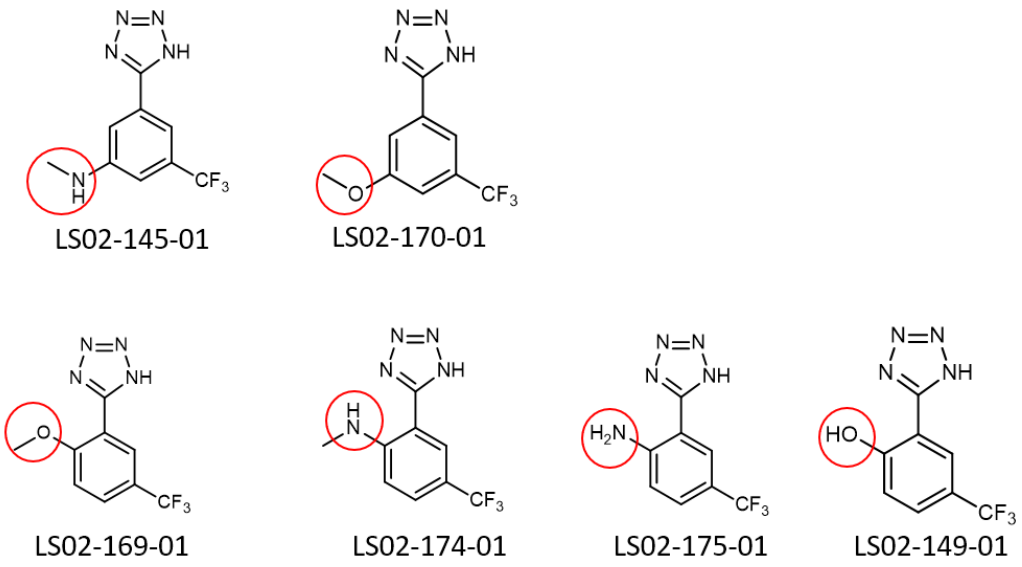
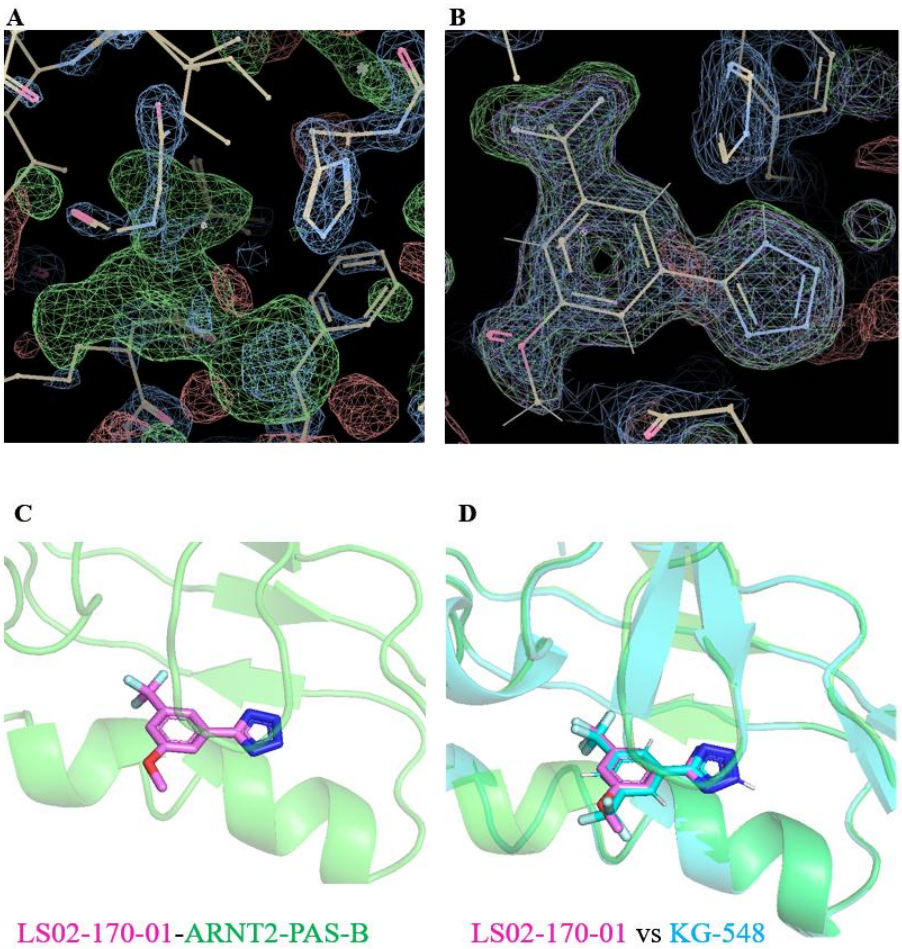


Figure 114. Similarities between analogues with aryl ring modifications. Differences in substitutions are circled in red.



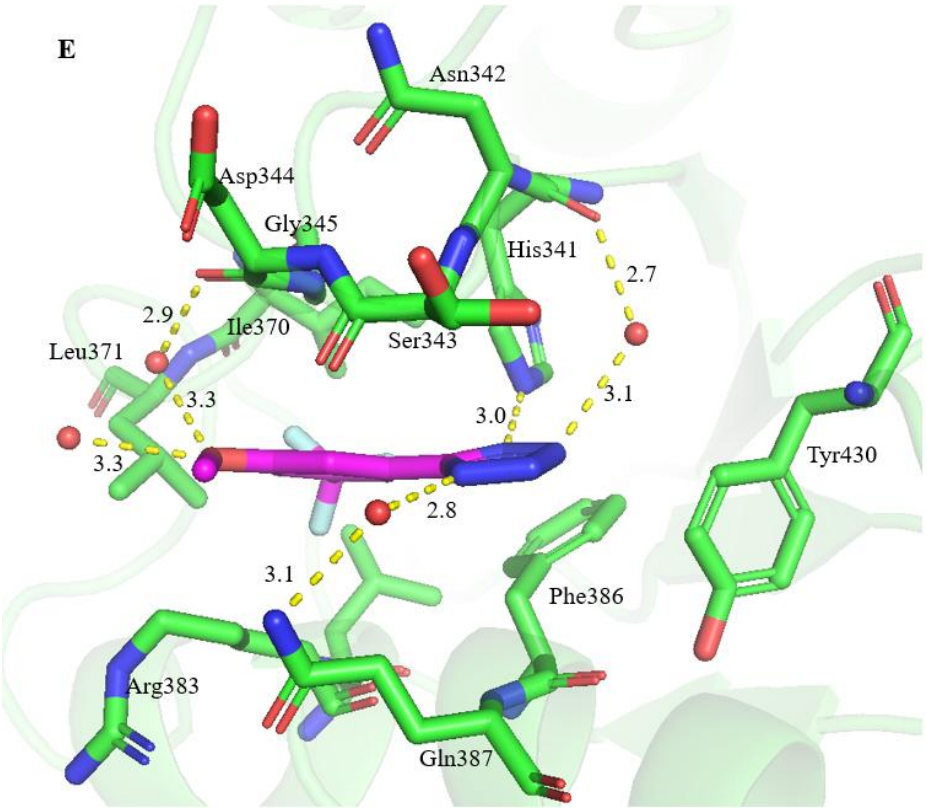


Figure 115. X-ray crystallography for LS02-170-01 binding to ARNT2-PAS-B. (A) Density map without ligand. (B) Density map with the LS02-170-01 ligand within the ARNT2-PAS-B domain. (C) Close-up look at the location of LS02-170-01 inside the ARNT2-PAS-B domain. LS02-170-01 is shown in pink. (D) The KG548-bound structure (cyan) was superposed on LS02-170-01 (pink) to show ligand binding overlap. (E) Interaction of LS02-170-01 (pink) with surrounding residues in the pocket (green). Yellow dashed lines represent measured interaction distances between the compound and nearby residues or water molecules.

In addition to LS02-169-01, LS02-170-01 is also structurally very similar to LS02-145-01. However, LS02-145-01 did not show any binding in TSA experiments (Figure 110). The only difference between the two compounds lies in the substituent meta to the tetrazole linkage: LS02-145-01 carries an N-methylamino group (-$NHCH_3$), while LS02-170-01 has a

methoxy group (-OCH_3) (Figure 114). It was therefore hypothesised that the methoxy substitution might be both electronically and sterically more favourable than that of the N-methylamino group, as it showed better TSA results.

Two further similar compounds, LS02-169-01 and LS02-174-01, showed prominent binding in TSA studies (Figure 110). Compared to LS02-169-01, LS02-174-01 displayed a higher thermal shift (dTm: +7.7 vs 4.6 °C). Both ligands differ by a single heteroatom (oxygen vs. nitrogen) at the ortho substituent, a methoxy group in LS02-169-01, and an N-methylamino group in LS02-174-01 (Figure 114). Similar to the comparison between LS02-170-01 and LS02-145-01, it is believed that the N-methyl substitution may be electronically and sterically more favourable than that of the O-methyl group.

The large difference in activity between these substitution patterns was made clear by solving the crystal structure of LS02-174-01 bound to ARNT2-PAS-B. By contrast with the meta-substituted analogue, the LS02-174-01 bound to ARNT2-PAS-B showed that the ortho-substituted group is instead oriented inward (Figure 116 and Appendix 15). The methyl group of the N-methyl substituent occupies a hydrophobic pocket and locates the nearby carbonyl groups of Ile370 and Leu371.

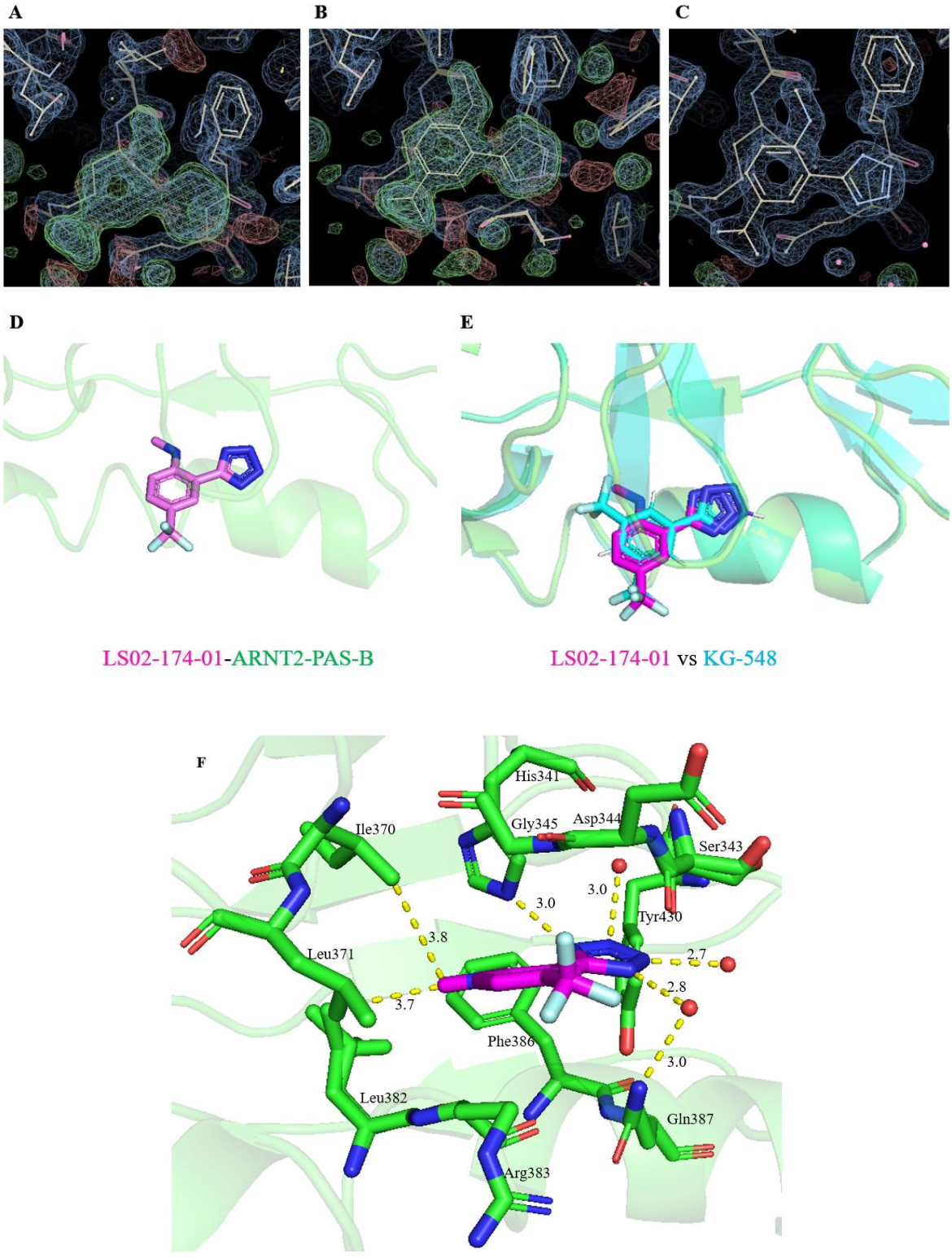


Figure 116. X-ray crystallography for LS02-174-01 binding to ARNT2-PAS-B. (A) Density map without ligand. (B-C) Density map with the LS02-174-01 ligand within the ARNT2-PAS-B domain. (D) Close-up look at the location of LS02-174-01 inside the ARNT2-PAS-

B domain. LS02-174-01 is shown in pink. (E) The KG548-bound structure (cyan) was superposed on LS02-174-01 (pink) to show ligand binding overlap. (F) Interaction of LS02-174-01 (pink) with surrounding residues in the pocket (green). Yellow dashed lines represent measured interaction distances between the compound and nearby residues or water molecules.

When polar groups such as methoxy or N-methyl were incorporated at the meta position, they tended to come out of the binding pocket due to the hydrophobic environment surrounding the C3 position. In contrast, at the ortho position, the N-methyl group was positioned inside the pocket, as it lies in closer proximity to His341. It is believed that the inward orientation of the N-methyl group allows for additional interactions with the protein, resulting in improved binding compared to the C-substituted analogues, in which the substituents protrude from the pocket and are exposed to solvent (Figure 117).

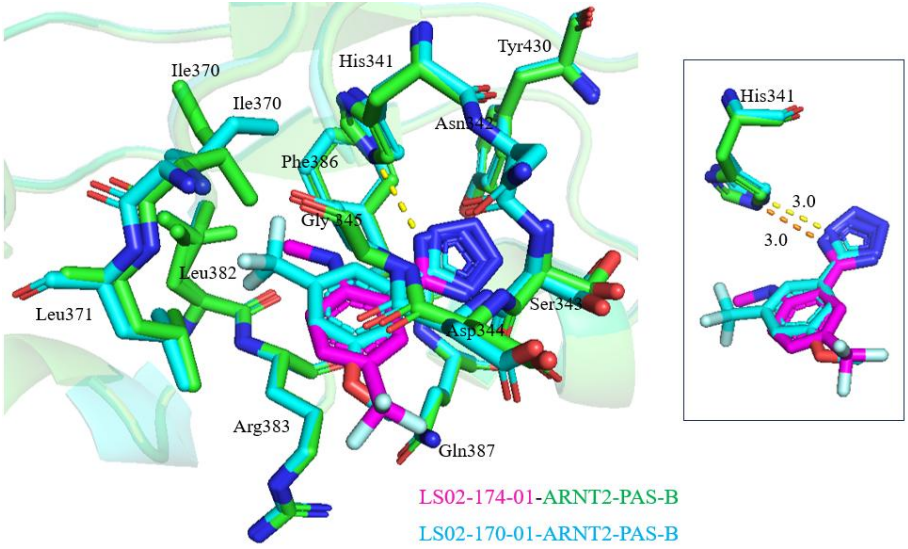


Figure 117. Structural comparison of the analogues LS02-174-01 and LS02-170-01.

To further understand why LS02-174-01 is a prominent analogue, it was decided to compare it with similar derivatives. Although LS02-174-01 and LS02-175-01 differ by only a single methyl group, this subtle change has a significant impact on binding. In LS02-174-01, the N-methylamino group points into the hydrophobic PAS-B pocket, where the methyl group contributes favourable hydrophobic interactions and helps orient the NH group for potential hydrogen bonding. In contrast, the primary amine (-NH_2) in LS02-175-01 is more polar and introduces an additional hydrogen-bond donor, which is likely unfavourable in a nonpolar environment. These factors likely explain why LS02-174-01 shows strong thermal stabilisation in TSA, whereas LS02-175-01 does not bind. A similar trend is also observed in the comparison between LS02-169-01 and LS02-149-01.

Based on these structural analyses, it was expected that LS02-174-01 would demonstrate competitive affinity to KG-548. Despite the high dTmB values observed in the TSA, the SPR data did not indicate strong binding affinity, as the compounds did not reach a saturation plateau in the steady-state affinity analysis (Figure 113). Unexpectedly, LS02-174-01 showed profound inhibition, reducing TR-FRET activity with mNPAS4-ARNT2 by 60% (Figure 112). Based on crystal data, no TR-FRET activity should be observed with these derivatives with the analogy to the KG-548 control. Such TR-FRET activity could be due to the off-target effect, or some interference between LS02-174-01 and the assay components.

To conclude, several KG-548 analogues were designed and tested in several assays, including TSA, X-ray crystallography, TR-FRET, and SPR, to identify a derivative compound with a higher affinity than that observed in the control compound. Although several analogues were able to crystallise, confirming binding to the PAS-B domain of ARNT2, none of them demonstrated an improved binding affinity to that observed with the original KG-548. As the

quest to identify a potent binder for ARNT2-PAS-B remains, the design and subsequent testing of KG-548 analogues can be continued. However, further investigation is required to identify a potent compound.

The sets of analogues used in this study could provide a valuable understanding of compound preferences inside the PAS-B pocket of ARNT2. Overall, the results from all the compound comparisons suggest that the binding pocket strongly favours small, non-polar substituents that can occupy a confined hydrophobic cavity adjacent to the tetrazole-phenyl core. Methyl and methoxy groups improve binding by potentially providing favourable van der Waals contacts and reducing desolvation costs, while more polar or hydrogen-bond-donating groups such as -NH_2 or -OH weaken affinity by increasing polarity and disrupting the optimal geometry. Larger or extended aromatic systems, like the extra-ring analogues LS02-177-01 and LS02-147-01, are too bulky to fit within the restricted space, resulting in the loss of productive interactions.

Additionally, it was also indicated that binding of KG-548 induces profound expansion of the ARNT2 PAS-B binding pocket, creating additional space for potential ligand interactions. This observation suggests an opportunity to potentially design even larger analogues of KG-548 – specifically, by further extending the tetrazole moiety and even projecting it beyond the confines of the binding pocket to engage with surrounding residues to improve binding affinity (Figure 118). Although further expansion beyond the tetrazole core is not chemically feasible without modifying the ring system, this study demonstrates that a triazole ring can be accommodated within the binding pocket when appropriately substituted with electron-withdrawing groups. This insight may inform future design strategies aimed at optimising ligand interactions within the PAS-B domain.

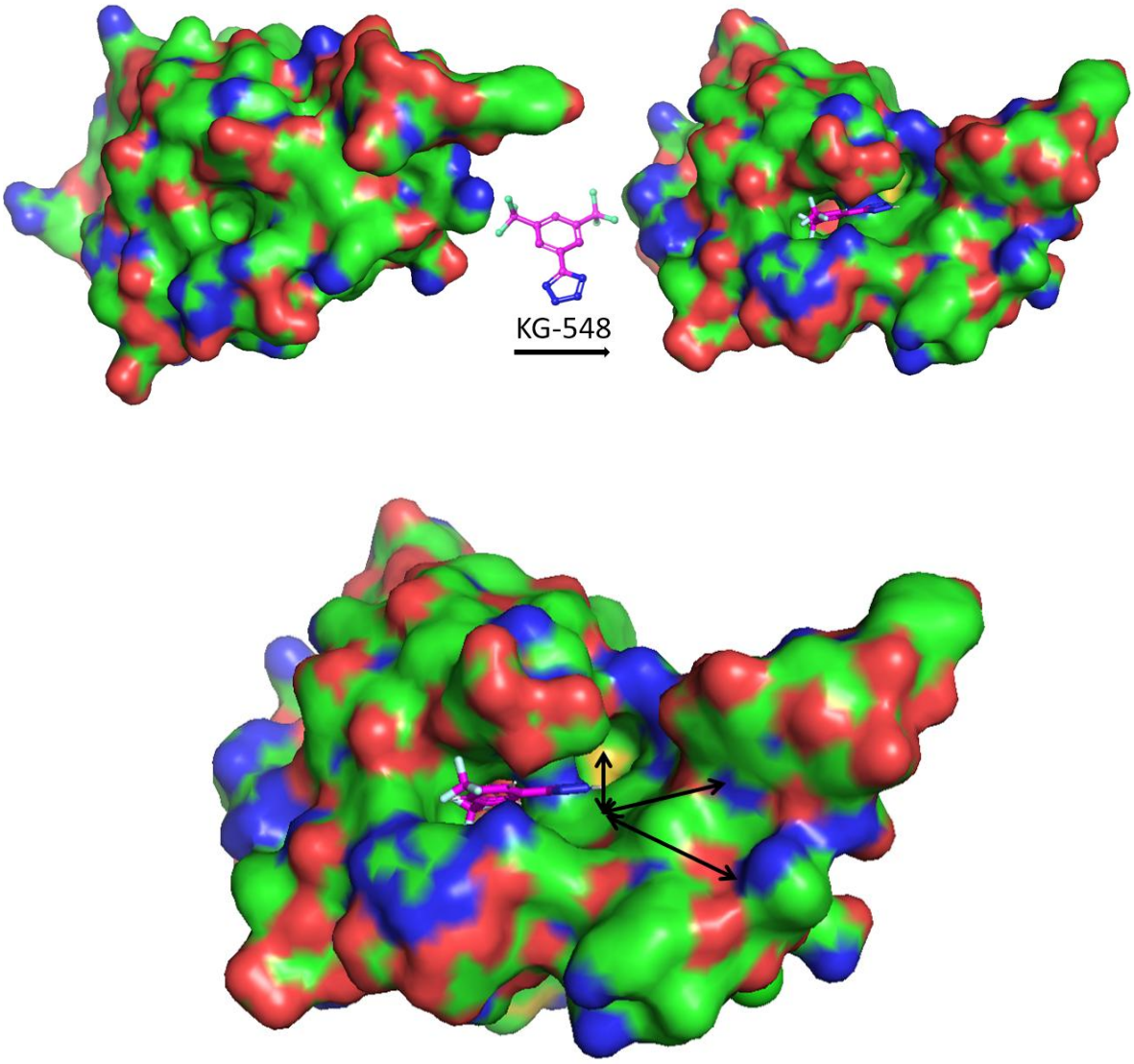


Figure 118. Conformational expansion of ARNT2 PAS-B upon KG-548 binding. The rational ligand design strategies for enhanced binding affinity are shown with black arrows.

Chapter 11: Hit-to-lead optimisation of butein

11.1 Butein: from herbal legacy to modern drug discovery

Recently in the laboratory, Dr. Hua Pu identified butein as a binder of the PAS-B domain of ARNT2 (unpublished data). Butein (3,4,2′,4′-tetrahydroxychalcone) is a natural polyphenolic compound that belongs to the chalcone family of flavonoids and predominantly found in many plant species, including *Rhus verniciflua Stokes*, *Dalbergia odorifera, Cyclopiasubternata,* and the stems of *Semecarpus anacardium.*[224,225] The medicinal properties of the source plants of butein have been extensively explored and implemented in the traditional medicinal systems of Asian countries for the treatment of a variety of conditions, including diabetes, dysentery, tuberculosis, hypertension, infectious diseases, inflammatory diseases, ischaemia, malaria, paralysis, and rheumatoid arthritis.[226] Butein isolated from these plants is also a well-characterised compound that exhibits a wide range of biological activities, including antioxidant, anti-inflammatory, anticancer, antinephritic, antidiabetic, and antibacterial effects.[227–229]

Owing to its multifaceted bioactivity, butein has attracted considerable attention as a lead compound for drug discovery and as a dietary phytochemical with potential clinical relevance. In various preclinical settings, butein has demonstrated therapeutic benefits across a range of chronic conditions, including inflammation, neurological disorders, neoplasms, atherosclerosis, cancer, diabetes. Most of these actions were accomplished by the inhibition of the transcription factor NF-κB and its downstream targets.[227,228,230] Butein, however, like many natural polyphenolic compounds, exhibits a multitargeted mechanism of action, suggesting pleiotropic activity. Rather than acting on a single, specific target, as is common

with many synthetic drugs, butein can simultaneously modulate multiple molecular targets and signalling pathways (Figure 119).[231,232] Studies have demonstrated that butein influences the expression and activity of various genes, transcription factors, enzymes, and proteins involved in critical cellular processes such as tumour initiation, progression, and chemoresistance.[233] This broad bioactivity may be partly attributed to its catechol moiety, which can participate in redox reactions and metal chelation. These structural features contribute to its high PAINS score of 0.97 (on a scale from 0 to 1), indicating a high potential for nonspecific interactions in biochemical assays.

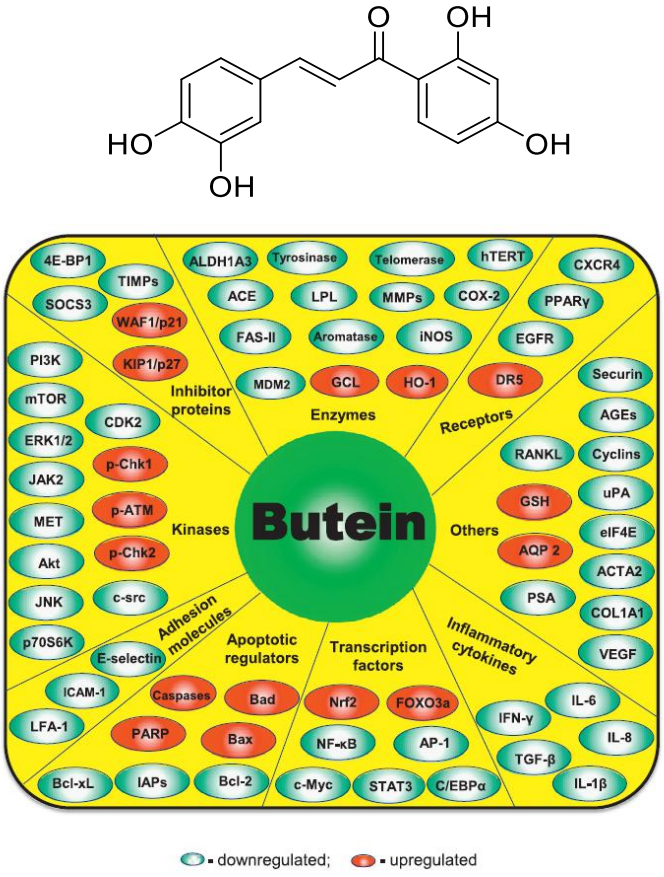
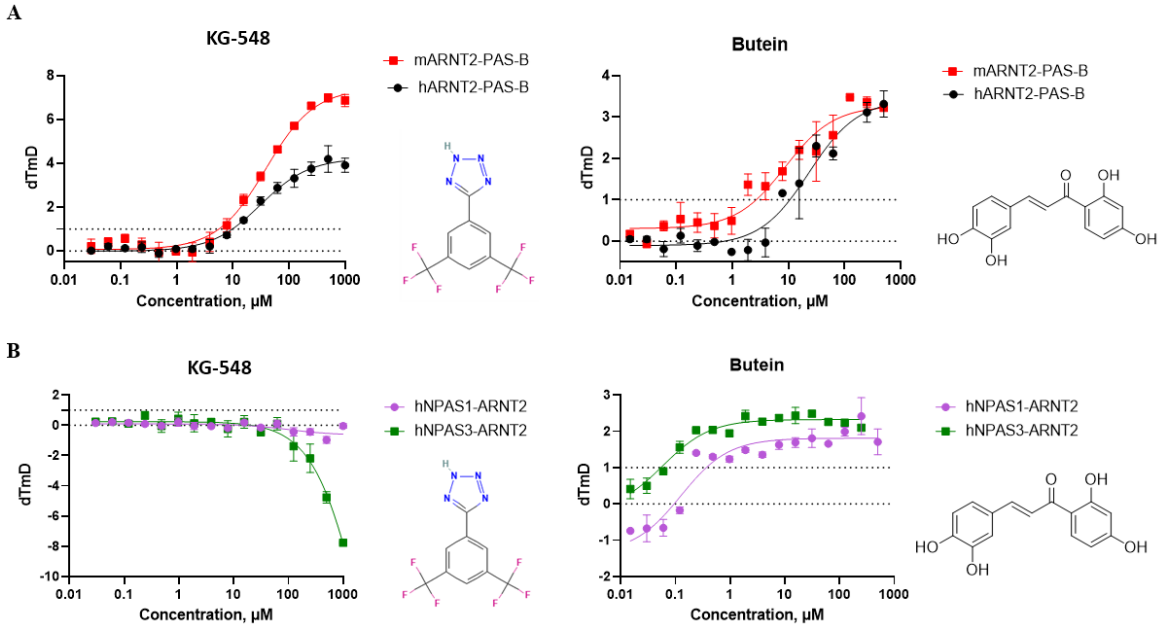


Figure 119. Butein, a multi-targeted compound with diverse biological effects. Figure adapted from Padmavathi et al., 2017.[233]

11.2. Butein binds the PAS-B domain of ARNT2

Like KG-548, butein binds to both mouse and human ARNT2-PAS-B domains via TSA, with calculated EC_{50} values of 8 µM for the mouse protein, and 23 µM for the human protein (Figure 120A). It was expected that butein would not produce positive TSA results with ARNT2-containing heterodimers, as its binding was presumed to induce conformational changes similar to those of KG-548, which compromise heterodimer stability. Butein, however, demonstrated clear thermal shift signals with both hNPAS1-ARNT2 and hNPAS3-ARNT2 heterodimers (Figure 120B). One possible explanation for this unexpected result is that butein, exhibiting PAINS activity, may cause assay artefacts that mimic true binding or stabilisation events.

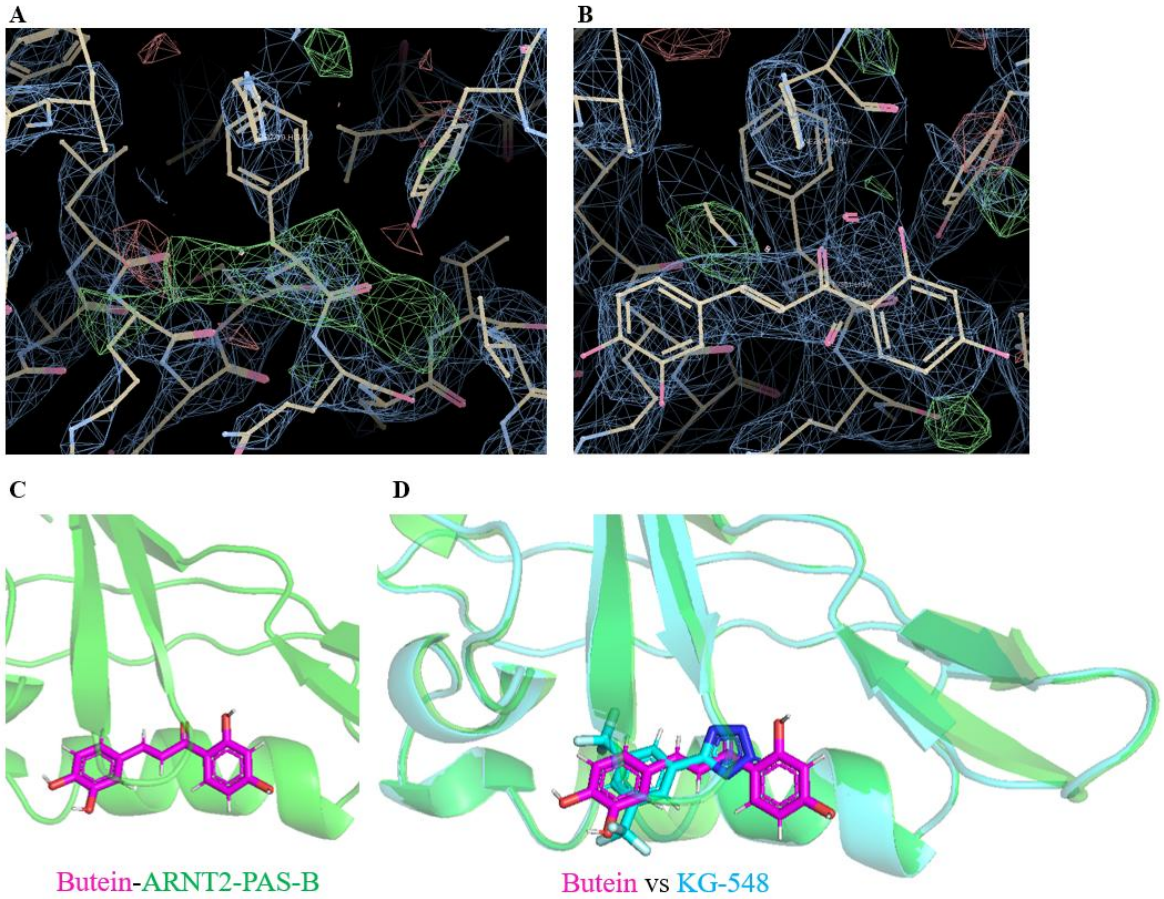


EC_{50} (μM)				
Compound	mARNT2-PAS-B	hARNT2-PAS-B	hNPAS1-ARNT2	hNPAS3-ARNT2
KG-548	38.6	31.3	n/a	n/a
Butein	8.6	23.2	0.1	0.06

Figure 120. TSA studies with butein. The compound was serially diluted to prepare 16-point concentration gradients. Butein was tested in DRC profile with (A) the mouse and human PAS-B domains of ARNT2, as well as with (B) hNPAS1/3-ARNT2 heterodimers in three independent runs. The calculated EC_{50} values are presented in the table below.

To understand the structural effect upon binding butein, the ligand was soaked with mouse ARNT2-PAS-B crystals (Figure 121, Appendix 15). The protein-ligand-bound crystal structure obtained from the seeding protocol was solved at 3.0 Å resolution (Figure 121, Appendix 15). To enable clearer structural analyses, the unpublished crystal structure of butein bound to mARNT2 PAS-B, which had been previously solved by Dr. Hua Pu in the laboratory, was used. The omit map shows that butein unambiguously binds to the PAS-B domain of ARNT2, occupying the same binding pocket as KG-548 (Figure 121A-D). Through hydrophobic contacts, butein interacts with several residues inside the PAS-B domain, notably with Ser343, Asp344, and Asp369 at distances of 2.7 Å, 3.4 Å and 3.3 Å, respectively (Figure 121E). Furthermore, butein may engage in a water-mediated interaction with Ile428, forming a tight bridge characterised by distances of 2.9 Å between the ligand and the water molecule, and 2.2 Å between the water and Ile428. Although the 2.2 Å distance is unusually short, it may reflect a highly ordered water molecule that significantly contributes to binding stabilisation.

Similar to KG-548, butein is further stabilised by a hydrogen bond between its carbonyl group and His341 at a distance of 3.3 Å. In this context, His341 acts as a hydrogen bond donor, donating a proton from its imidazole ring to the carbonyl oxygen of butein, which serves as the acceptor. This observation is consistent with the binding behaviour of KG-548 and its analogues, which show improved affinity when they include well-positioned hydrogen bond acceptors, such as ether oxygens, carbonyls, or tetrazole rings, capable of interacting with the His341donor. Together, these observations support a model where favourable hydrogen bonding and electrostatic interactions with His341 enhance ligand binding.



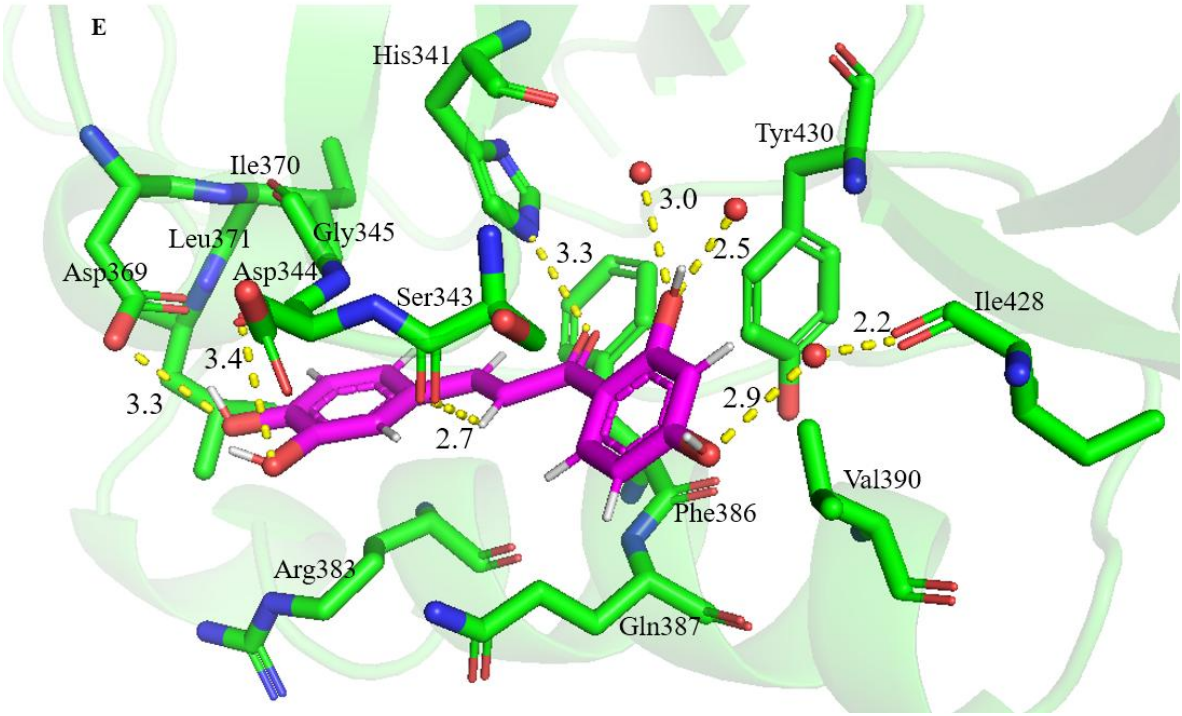


Figure 121. X-ray crystal structure of the butein compound with ARNT2-PAS-B. (A) Density map without ligand. (B) Density map with the butein ligand within the ARNT2-PAS-B domain. (C) Close-up look at the location of butein inside the ARNT2-PAS-B domain; butein is shown in pink. (D) The KG-548-bound structure (cyan) was superposed on the butein (pink) to show the overlapping binding location. (E) Interactions of butein (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts/interactions with protein residues and water molecules in the crystal structure.

Structurally, Butein is slightly larger than KG-548. To understand why butein showed positive TSA results with both NPAS1/3-ARNT2 heterodimers, both ARNT2-PAS-B structures bound to KG-548 and butein were superimposed onto the mNPAS4-ARNT2 complex structure (PDB: 7XI3). As expected, butein binding induces several conformational changes similar to those observed with KG-548 (Figure 122). Notably, the side chain of Tyr430 undergoes a flip, potentially disrupting a key hydrogen bond between ARNT2 and

NPAS4. Additionally, Val341, which is normally buried, is displaced outward, creating a steric clash with Leu261 of NPAS4, with a close interatomic distance of just 2.3 Å. The HI loop of ARNT2 also shifts position, resulting in the loss of critical interactions with the F helix in the PAS-B domain of NPAS4. These structural rearrangements occur at crucial domain-domain interfaces that are essential for maintaining heterodimer integrity, suggesting that butein binding, similar to KG-548, compromises complex stability by altering these key interaction sites.

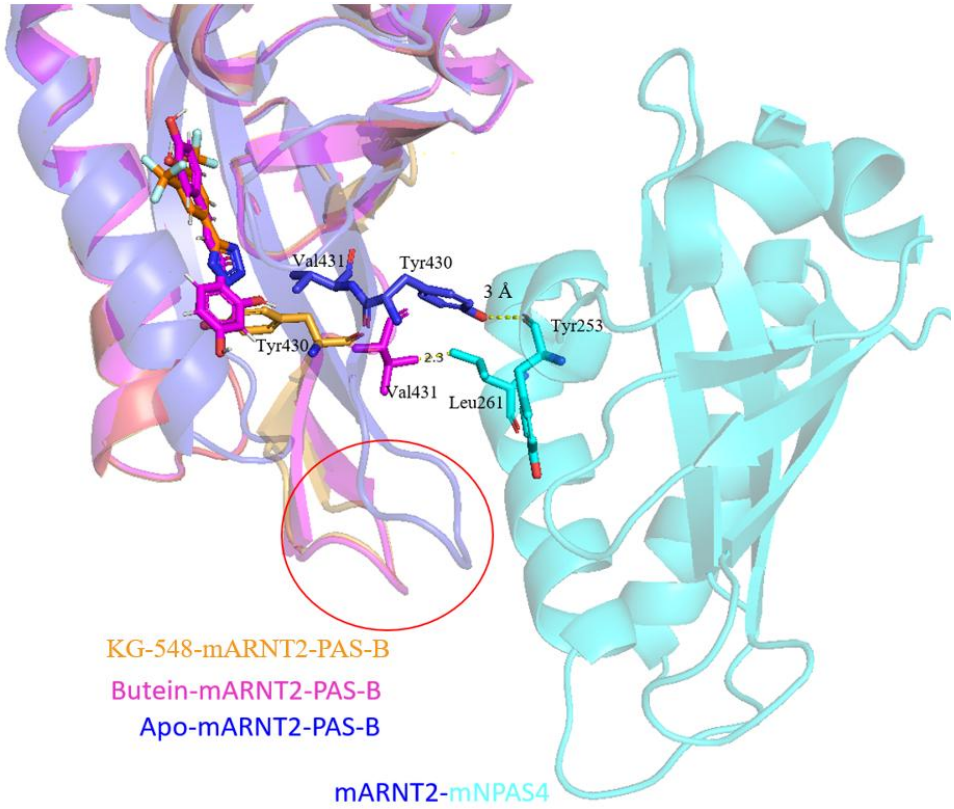
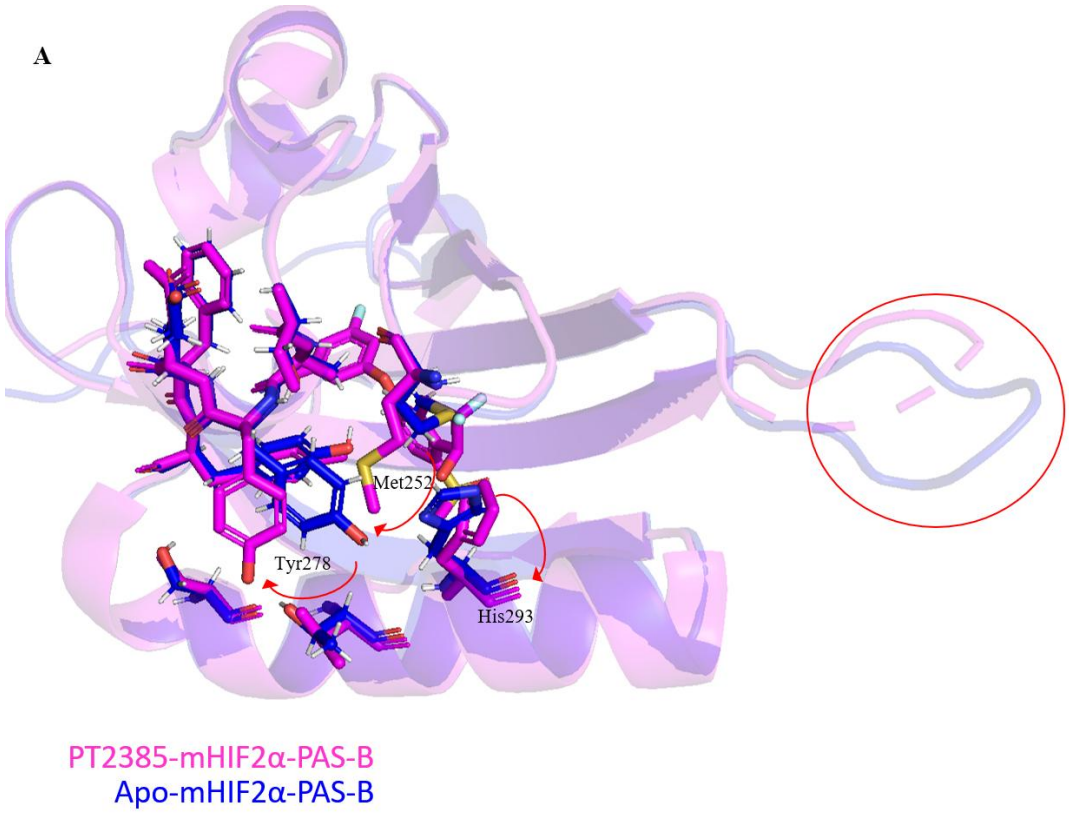


Figure 122. Structural analyses of mARNT2-NPAS4 upon binding to KG-548 and butein. The mARNT2-NPAS4 crystal structure is taken from PDB: 7XI3. Apo mARNT2-PAS-B is shown in blue, mARNT2-PAS-B bound to KG-548 in orange, and mARNT2-PAS-B bound to butein in pink. The PAS-B domain of NPAS4 is illustrated in cyan. The hydrogen bond between Tyr430 of apo-ARNT2 and the backbone of Tyr253 from NPAS4 is measured and shown with a dashed yellow line. The bond between Val431 of ligand-bound mARNT2-

PAS-B and the backbone of Leu261 from NPAS4 is measured and shown with a dashed yellow line.

Based on this structural analysis, it was hypothesised that butein might possess a high affinity for the PAS-B domain of ARNT2, and that the heterodimeric complex might also therefore undergo conformational changes to accommodate the ligand. A similar phenomenon was observed with HIF2α and its potent inhibitor belzutifan. Upon binding to HIF2α-PAS-B, belzutifan causes structural changes that prevent the formation of heterodimeric complexes with ARNT (Figure 123A). However, the belzutifan-bound HIF2α-ARNT crystal structure illustrates that the pre-formed heterodimer complex might undergo additional changes to accommodate the potent ligand (Figure 123B).



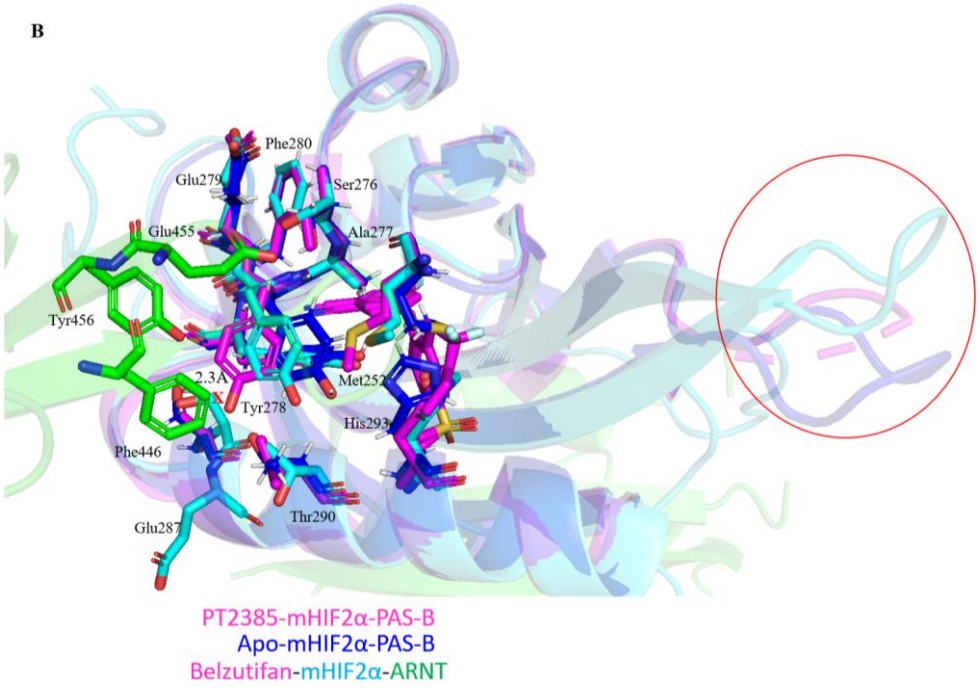


Figure 123. Structural analyses of mHIF2α upon binding to belzutifan. (A) Apo mHIF2α-PAS-B (PDB: 3F1P, resolution 1.17 Å) is shown in blue, and mHIF2α-PAS-B bound to belzutifan precursor PT2385 in pink (PDB: 5TBM, resolution 1.85 Å). The Met252 sidechain, which is typically positioned internally within the PAS-B domain of HIF2α, is repositioned and pushes the Tyr278 sidechain sideways to accommodate the ligand. Displacement of Tyr278 residue creates an unfavourable van der Waals clash with ARNT: Phe446, resulting in dimer disruption (measured at a close distance of 2.3 Å). (B) Both single domain structures were overlaid with belzutifan-bound mHIF2α-ARNT heterodimer complex (PDB: 7W80, resolution 2.75 Å). The PAS-B domain of mHIF2α within the belzutifan-bound heterodimer complex is illustrated in cyan, whereas the PAS-B domain of ARNT is shown in green. The HI loop of HIF2α-PAS-B, which takes part in the formation of heterodimer complexes, shifts position in all three structures, apo-HIF2α-PAS-B, precursor-bound PT2385-mHIF2α-PAS-B complex, and belzutifan-HIF2α-ARNT heterodimer, suggesting additional structural modification to accommodate potent belzutifan.

In support of this hypothesis, TSA studies showed that belzutifan appeared to bind not only to a single HIF2α-PAS-B, but also to both HIF2α-ARNT and HIF2α-ARNT2 heterodimers (Figure 124). It is important to note that while belzutifan induces a substantial dTmB shift of nearly 25 °C with a single PAS-B domain, the observed dTmB values for both heterodimers are significantly lower, approximately 7 °C, representing a threefold reduction in thermal stabilisation. Butein, however, did not show a profound difference between binding a single ARNT2-PAS-B domain and heterodimeric complexes with either HIF2α-ARNT2 or HIF2α-ARNT. Interestingly, binding to HIF2α-ARNT suggests that butein might also bind the PAS-B domain of ARNT protein. In contrast, the KG-548 has very weak affinity to the PAS-B domain of ARNT2; therefore, no binding was observed with either of ARNT2's heterodimeric complexes (Figure 124, 52-53).

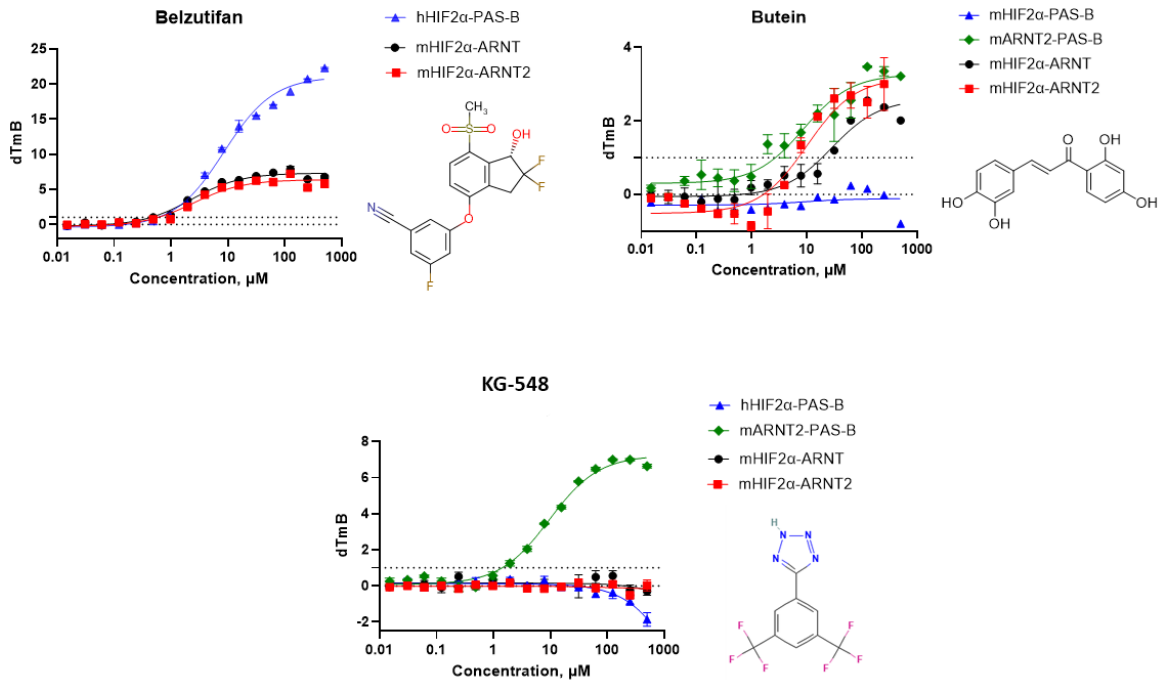
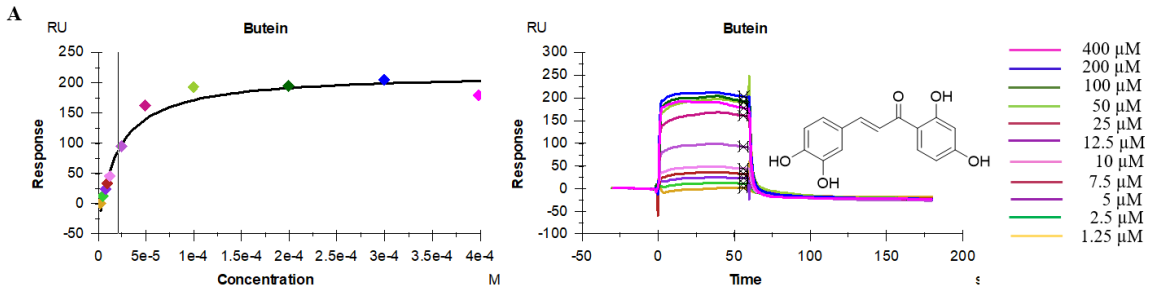


Figure 124. TSA studies with belzutifan, butein, and KG-548 compounds. Compounds were titrated against protein targets, in three independent runs. Compounds were tested with

relevant single domains, and both heterodimeric complexes with HIF2α-ARNT2 and HIF2α-ARNT.

To check the affinity hypothesis, SPR analyses were performed with biotinylated ARNT2-PAS-B and butein ligand. Compared to the previously measured KG-548, butein reached saturation; its calculated K_D of approximately 21 µM, determined via the steady-state affinity method, is closely aligned with its observed EC_{50} of around 8 µM in the TSA assay (Figure 125A). This SPR-derived affinity of 21 µM is significantly stronger than that of KG-548, which did not reach saturation even at the highest tested concentration of 400 µM. This stronger butein binding may explain the positive stabilisation observed via TSA across all heterodimeric complexes.



Compound	K_D (µM)
Butein	21

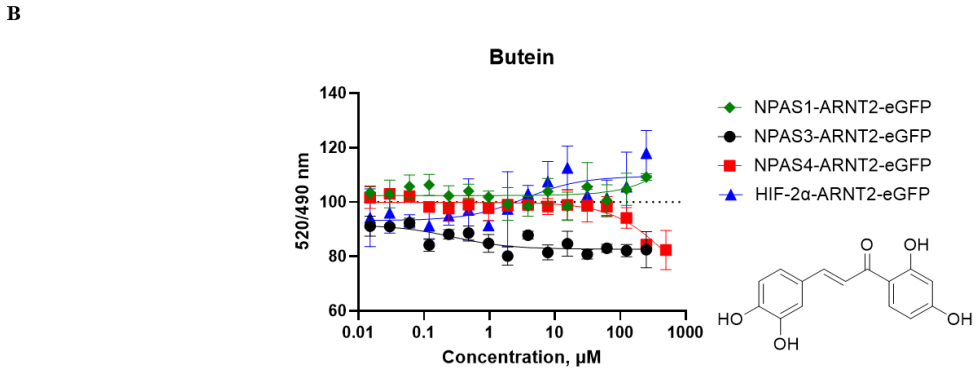


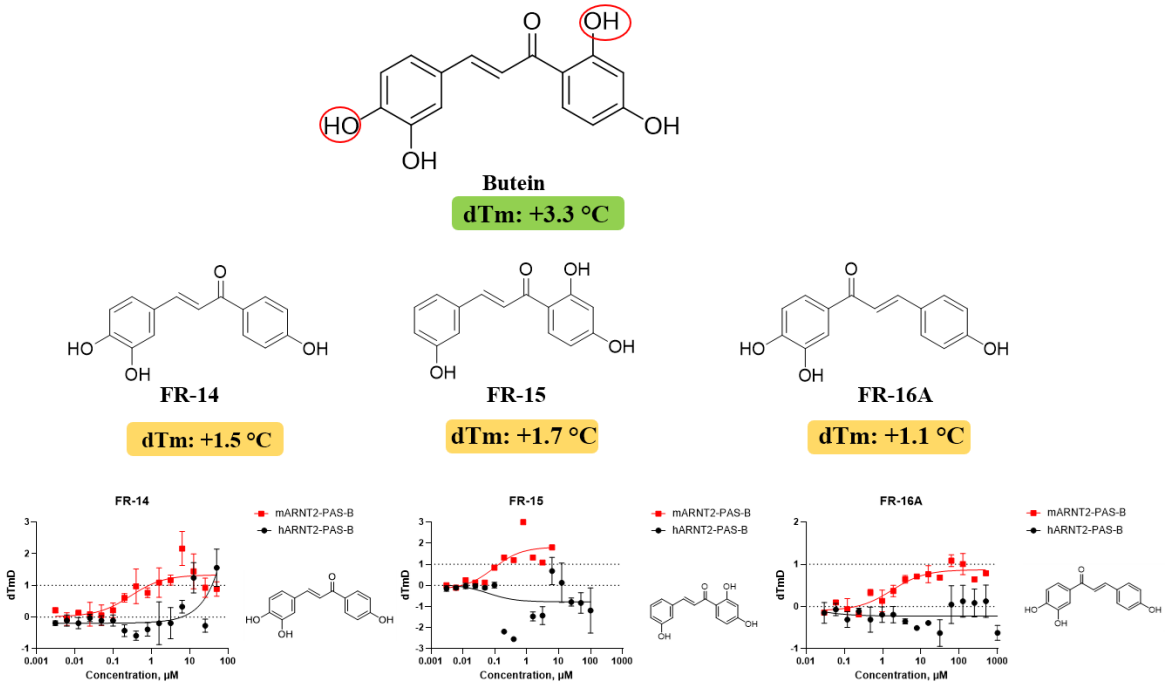
Figure 125. SPR and TR-FRET studies with butein compound. (A) Steady state affinity and calculated K_D for butein with the biotinylated single PAS-B domain of ARNT2. (B) The TR-FRET activity of the butein against ARNT2 with several potential partners, including mNPAS1, mNPAS3, mNPAS4, and mHIF2α.

Although butein was expected to show some activity in the TR-FRET assay, the observed 20% activation with HIF2α-ARNT2 and 20% inhibition with NPAS3-ARNT2 are relatively weak and cannot be considered biologically meaningful (Figure 125B). Therefore, it was decided to investigate whether hit-to-lead optimisation of butein ligand could yield compounds with improved selectivity, affinity, or activity via TR-FRET.

11.3 Hit-to-lead optimisation for butein

The chemical structure of butein consists of two aromatic rings connected by a three-carbon α,β-unsaturated carbonyl system (chalcone), with four hydroxyl groups located at positions 2', 3, 4, and 4'. In the hit-to-lead optimisation of butein, metabolic stability emerged as a key concern due to the presence of four hydroxyl groups on the scaffold. These hydroxyl functionalities, while potentially contributing to target affinity, are prone to extensive phase II metabolism, particularly glucuronidation and sulfation, leading to high clearance and reduced systemic exposure.[234] Additionally, the high degree of hydroxylation may negatively impact membrane permeability, further limiting the compound's bioavailability. To address these liabilities, it was proposed to remove hydroxyl groups, or to replace them with less metabolically labile substituents. Thus, compounds FR-14, FR-15, and FR-16A, designed by

Dr. Srinivas Cheruku, feature a reduced number of hydroxyl groups, which contributes to a lower PAINS score, thereby enhancing the overall drug-likeness of the scaffold (Figure 126).

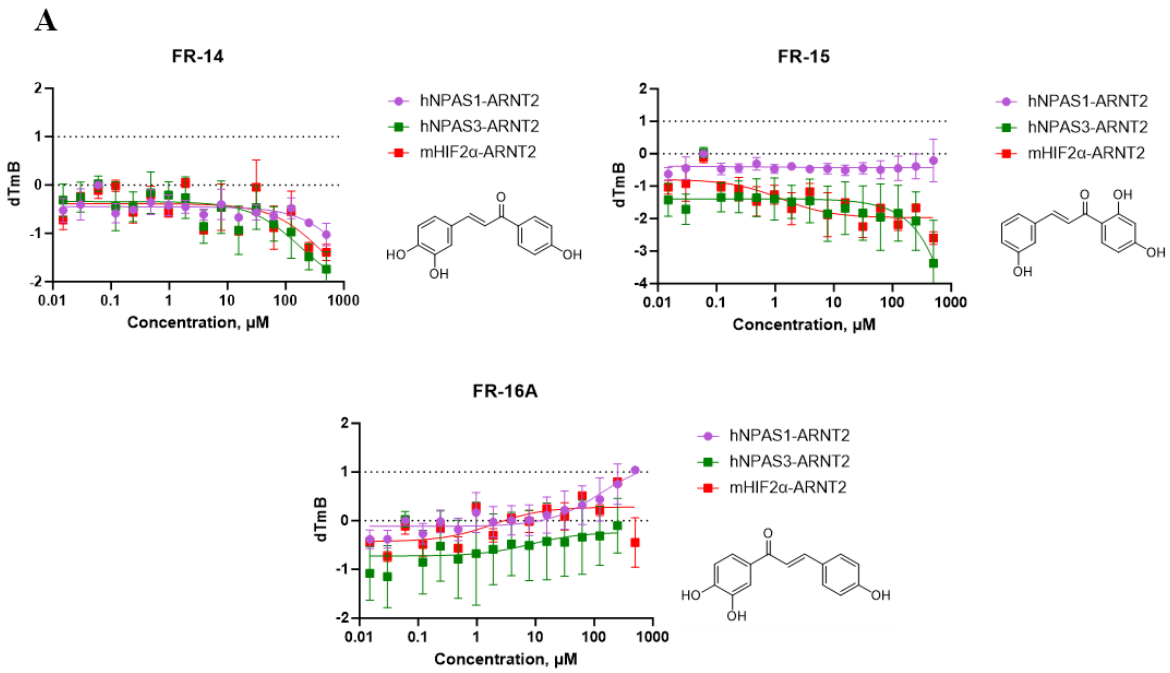


EC_{50} (µM)		
Compound	mARNT2-PAS-B	hARNT2-PAS-B
FR-14	0.3	unstable
FR-15	0.1	n/a
FR-16A	1.7	n/a

Figure 126. Butein analogues with reduced hydroxyl groups. The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with mouse and human ARNT2-PAS-B domain.

All three analogues, FR-14, FR-15, and FR-16A exhibited a positive thermal shift in TSA experiments when tested against the isolated mouse PAS-B domain, indicating direct binding and thermal stabilisation of the domain (Figure 126). The observed changes in melting

temperature were modest but consistent, with increases of 1.5 °C, 1.7 °C, and 1.1 °C for FR-14, FR-15, and FR-16A, respectively. However, when the compounds were tested against heterodimeric complexes formed by NPAS1, NPAS3, or HIF2α in association with partner ARNT2, no thermal shift was detected (Figure 127A). This lack of shift implies that the compounds do not bind to the PAS-B domain in the context of the heterodimer, as compared to butein, potentially due to the lower affinity. In agreement with these binding data, no activity was observed in TR-FRET with either of the heterodimeric complexes (Figure 127B).



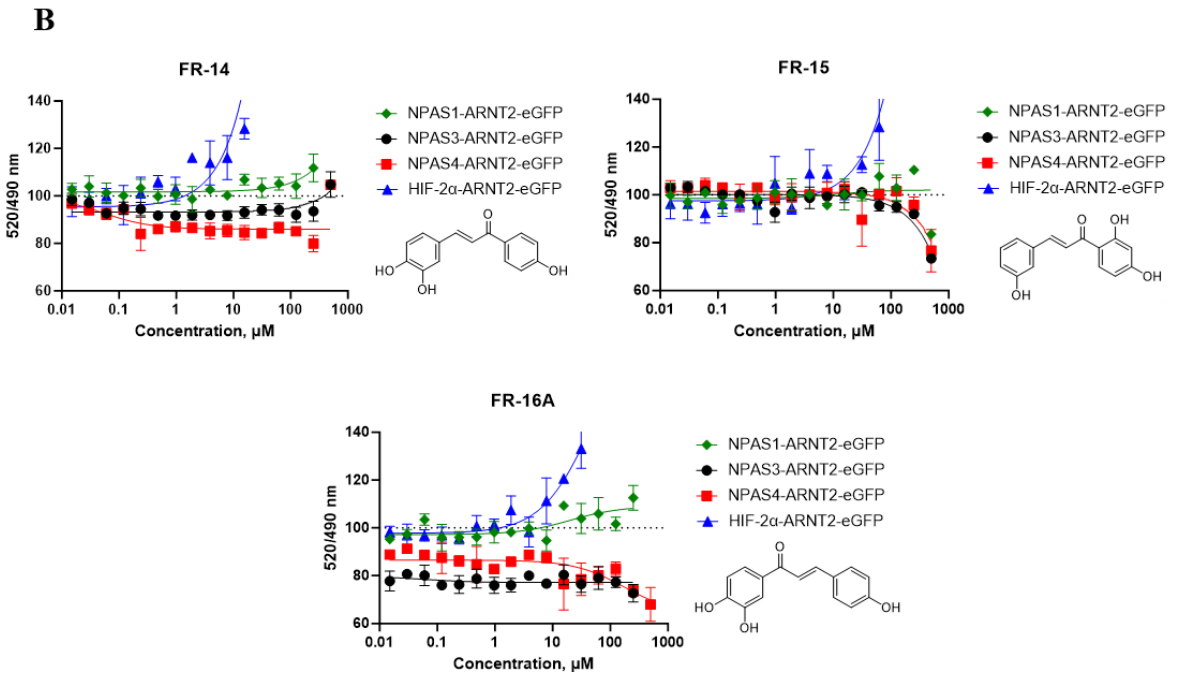
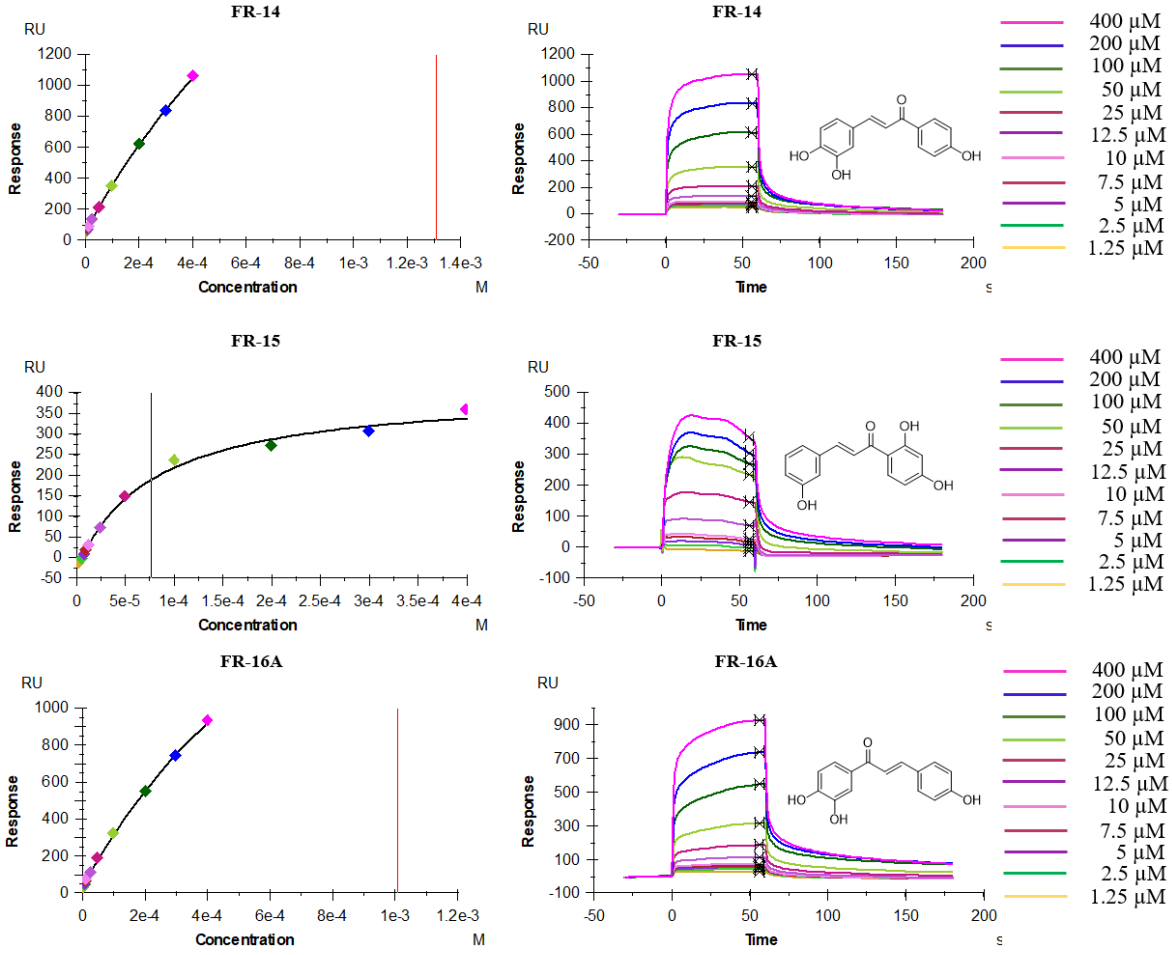


Figure 127. TSA and TR-FRET studies for butein analogues with reduced hydroxyl groups. (A) Butein analogues were tested in DRC profile with freshly purified hNPAS1/3-ARNT2 and mHIF2α-ARNT2 heterodimers in three independent runs using the TSA assay. (B) The TR-FRET activity of the butein analogues against ARNT2 with several potential partners, including mNPAS1, mNPAS3, mNPAS4, and mHIF2α.

The SPR experiment confirmed that FR-15 has a lower affinity than butein, reaching equilibrium binding with an estimated steady-state K_D of approximately 76 μM (Figure 128). Although this affinity is still within the low micromolar range and is slightly weaker than that of butein, FR-15 exhibited a significantly improved PAINS score of 0.33 compared to 0.97 for butein, suggesting reduced assay interference and enhanced chemical tractability.

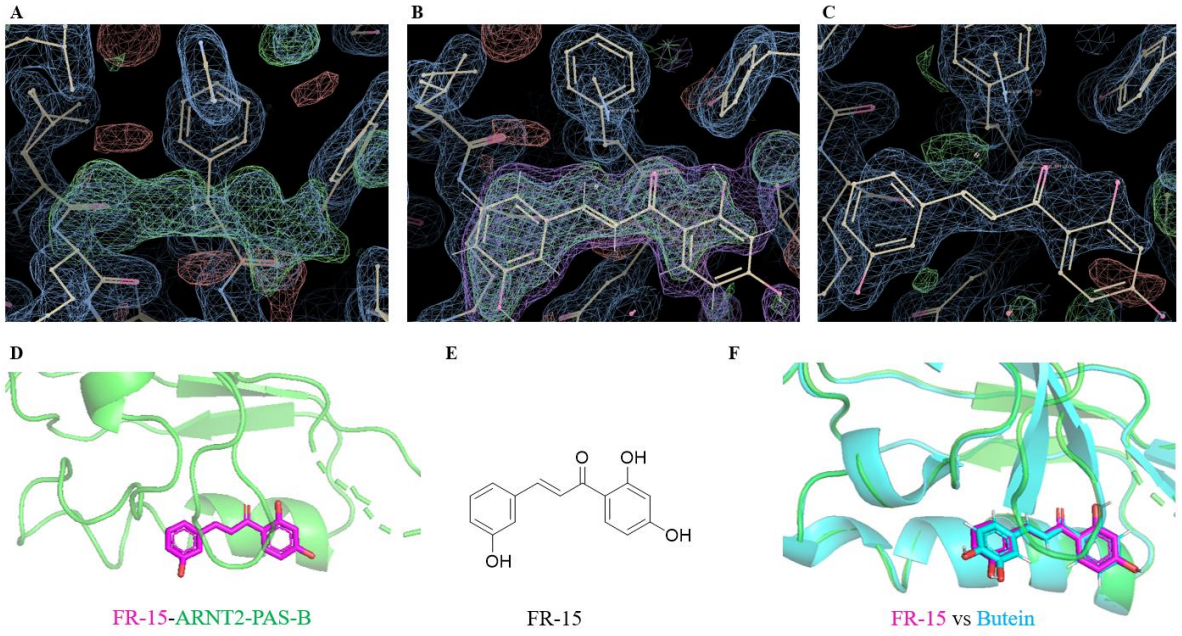
In contrast to FR-15, neither FR-14 nor FR-16A reached saturation within the tested concentration range. As a result, their affinities could not be directly determined, and the data imply weak binding, with estimated K_D values greater than 400 μM. Both compounds also

produced very large SPR responses (~1,000 RU), which are inconsistent with 1:1 binding given their molecular weights. Based on the SPR binding equation and an immobilisation level of ~8,000 RU, the theoretical RUmax for a single PAS-B domain should be approximately 140 RU. The observed responses therefore greatly exceed the expected RUmax, indicating that the signals are not compatible with simple 1:1 binding and likely reflect nonspecific binding, aggregation, or bulk refractive index effects.



Compound	K_D (µM)
FR-15	76

Figure 128. SPR studies for butein analogues with reduced hydroxyl groups. Steady state affinity for FR-14, FR-15, and FR-16A with the biotinylated single PAS-B domain of ARNT2. The calculated K_D for FR-15 is shown in the blue table.



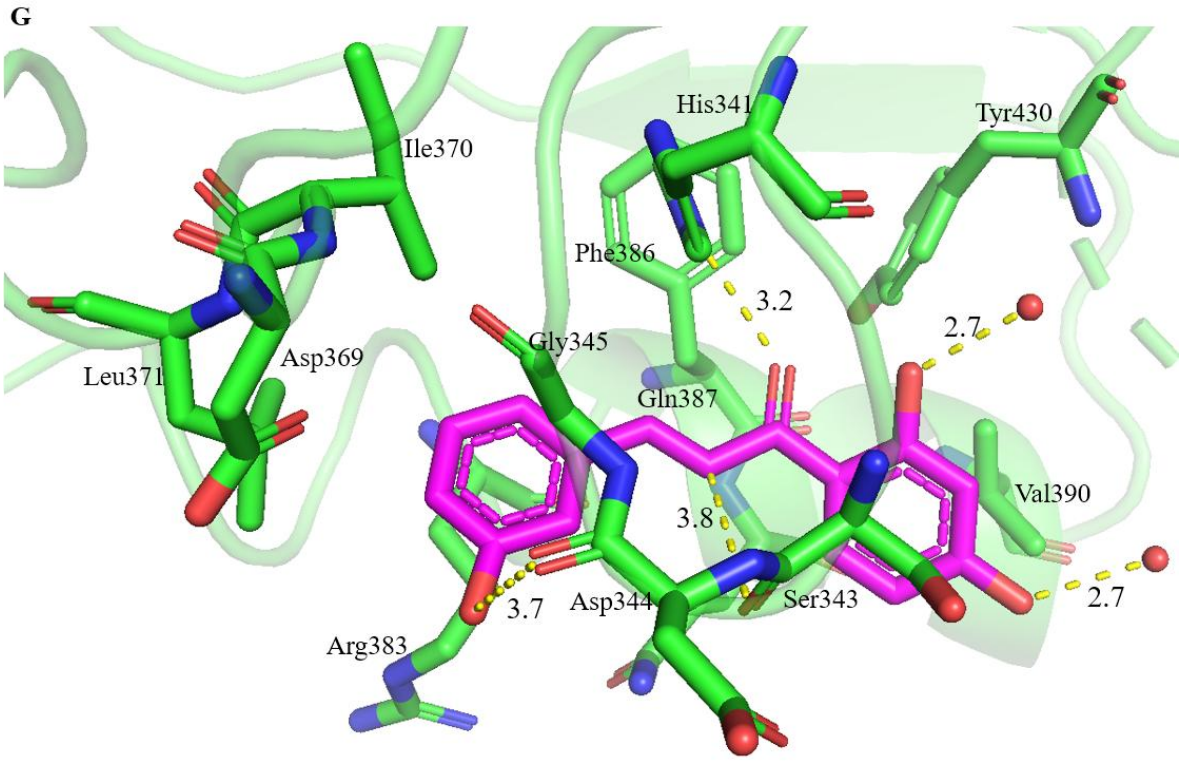


Figure 129. Crystal structure of the FR-15 butein analogue with ARNT2-PAS-B. (A) Density map without ligand. (B-C) Density map with the FR-15 ligand within the ARNT2-PAS-B domain. (D) Close-up look at the location of FR-15 inside the ARNT2-PAS-B domain; FR-15 is shown in pink. (E) Chemical structure of FR-15 analogue. (F) The butein-bound structure (cyan) was superposed on FR-15 (pink) to show the overlapping binding location. (G) Interactions of FR-15 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts or interactions with protein residues and water molecules in the crystal structure.

Crystallographic analysis was also successful for FR-15, whose complex with the PAS-B domain was resolved at 1.61 Å resolution (Figure 129, Appendix 16). Structural comparison revealed that FR-15 binds to the same hydrophobic pocket previously identified for butein. Notably, FR-15 forms a hydrogen bond with His341 at a distance of 3.2 Å, a key interaction also observed in the butein-bound structure, further supporting the conservation of the binding mode across different ligands. The absence of the hydroxyl group, however, resulted in the loss of the hydrogen bond previously observed between butein and Asn369. Furthermore, interactions with Ser343 and Asp344 were weakened, as evidenced by increased distances between the ligand and these residues, now measuring approximately 3.8 Å and 3.7 Å, respectively. These structural observations are consistent with the TSA and SPR data, which also indicate a reduction in binding affinity for FR-15 compared to butein.

11.4 Design and evaluation of butein analogues to improve stability and target occupancy

In addition to the metabolic liability associated with hydroxylation, the α,β-unsaturated carbonyl system in butein also raises stability concerns. The conjugated double bond is a potential site for oxidative metabolism, particularly by cytochrome P450 (CYP) enzymes, which can lead to metabolic instability and the formation of reactive intermediates.[235] While saturation of the double bond could mitigate this metabolic vulnerability, such a modification would disrupt the planarity of the molecule, which is critical for maintaining optimal interactions within the flat binding pocket of the target.

To address this, an alternative strategy under consideration involves replacing the unsaturated ketone with an amide linkage, which led to the design of FR-17a and FR-21 analogues. This modification was expected to preserve the necessary planarity for binding while improving metabolic stability by eliminating the electrophilic double bond and reducing the likelihood of CYP-mediated oxidation. Both analogues, however, were not active in TSA studies (Figure 130). Structural analysis of butein and FR-15 suggests that the pyrazole ring introduced in FR-21 is too bulky to fit within the binding pocket. Similarly, the amide linkage in FR-17 is predicted to cause steric clashes with Ser343. In contrast, the distance between butein and Ser343 is approximately 2.7 Å, indicating a close and favourable interaction that is likely disrupted by the bulkier amide group in FR-17.

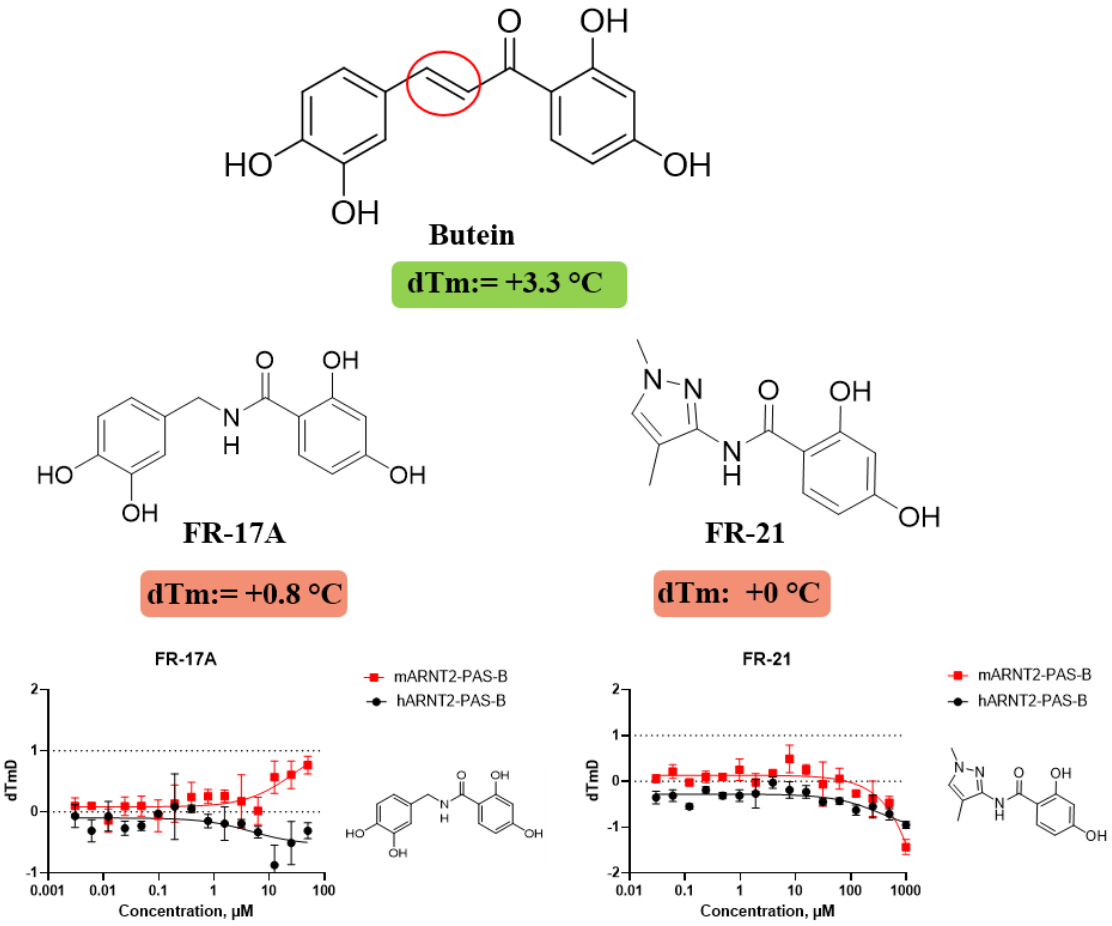
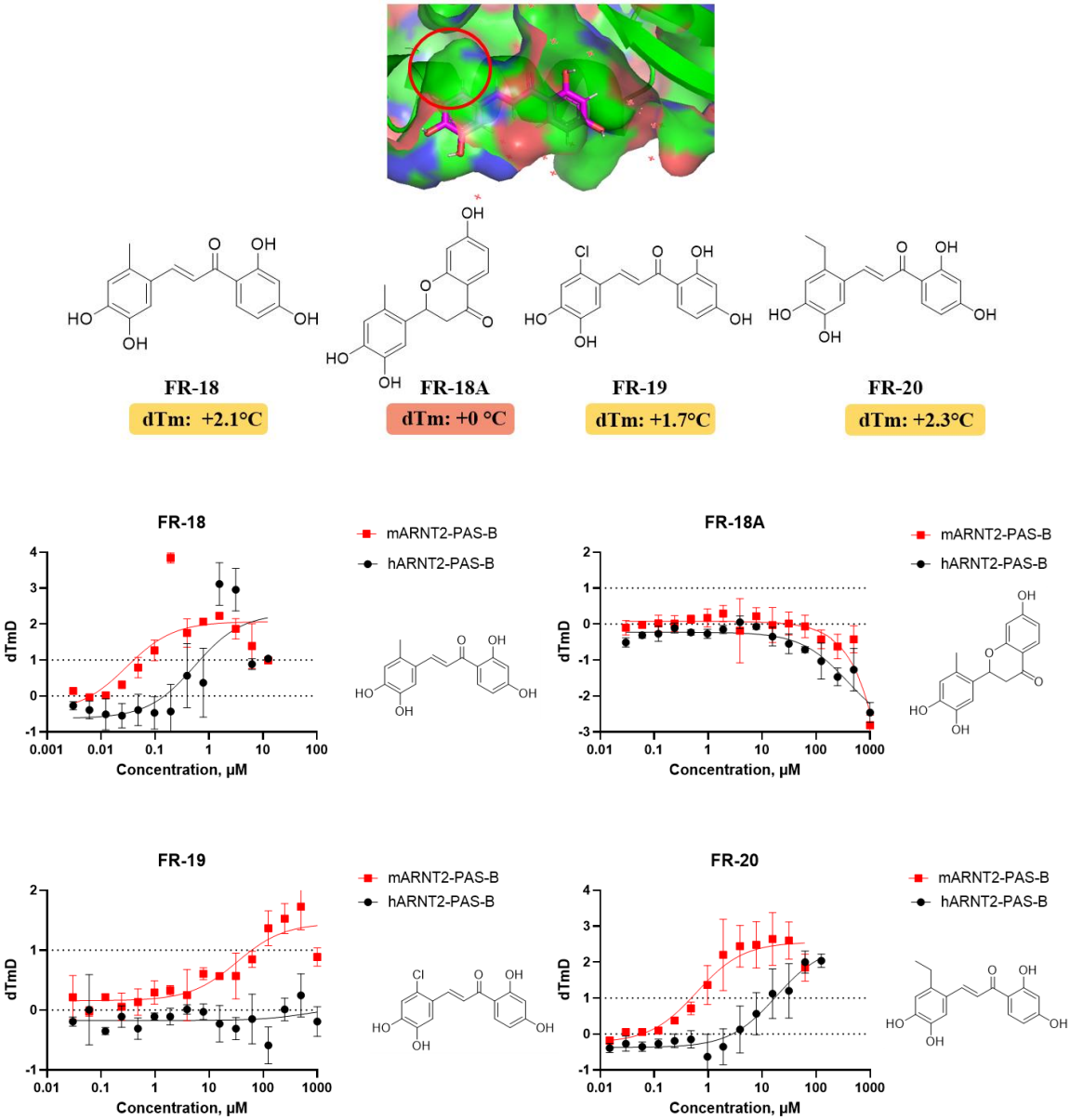


Figure 130. Analogues with the unsaturated ketone replaced by an amide linkage. The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with mouse and human ARNT2-PAS-B domain.

From the crystal structure analysis, it was observed that butein compound, like the KG-548, does not fully occupy the hydrophobic pocket of the ARNT2 PAS-B domain. Therefore, alternative design strategies were pursued to improve the occupancy of the PAS-B pocket and to enhance interactions within the hydrophobic groove of the ARNT2 binding site. This led Dr. Srinivas Cheruku to design a series of analogues, including FR-18, FR-18a, FR-19,

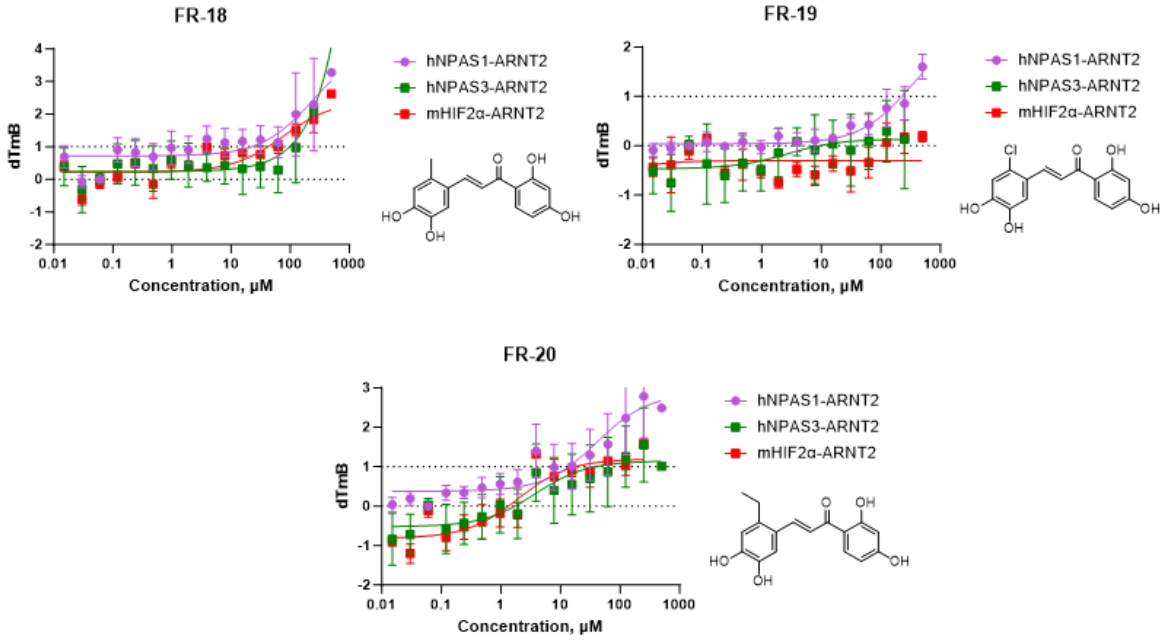
and FR-20. Among these, FR-18, FR-19, and FR-20 ligands exhibited binding to mARNT2-PAS-B in TSA studies (Figure 131). Additionally, FR-18 and FR-20 showed binding to the human ARNT2 PAS-B domain.



EC_{50} (μM)		
Compound	mARNT2-PAS-B	hARNT2-PAS-B
FR-18	0.03	0.5
FR-19	32.8	n/a
FR-20	0.6	18

Figure 131. Butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with mouse and human ARNT2-PAS-B domain.

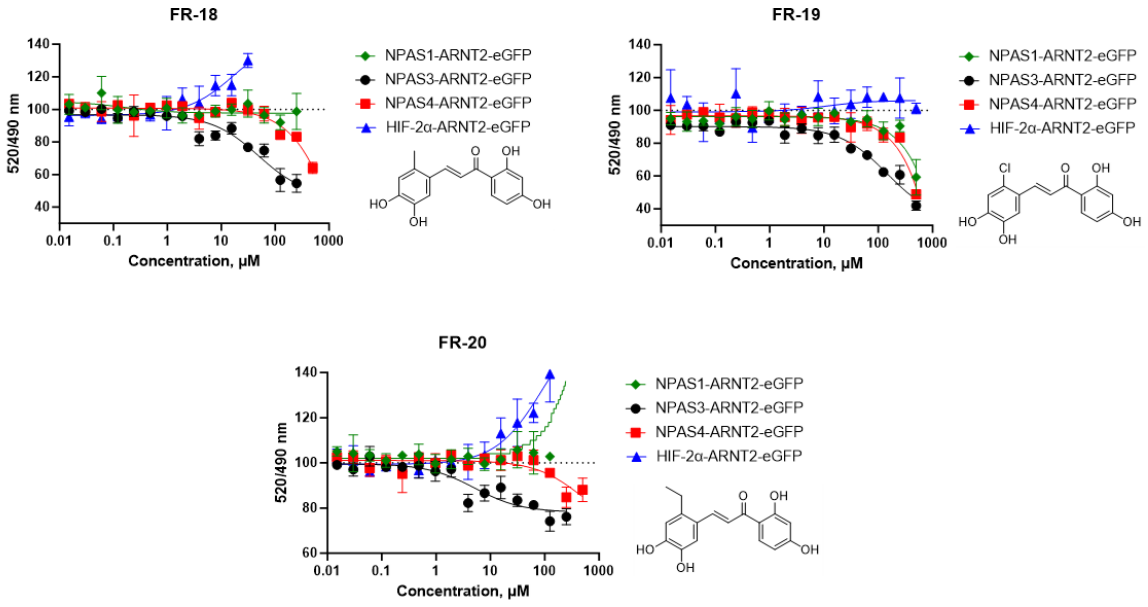
While FR-18 displayed an unsaturated binding curve with the heterodimeric complexes formed by NPAS1, NPAS3, or HIF2α with ARNT2, FR-20 demonstrated fully saturated binding with all three heterodimers (Figure 132). Calculated EC_{50} values for FR-20 binding to NPAS1-ARNT2, NPAS3-ARNT2, and HIF2α-ARNT2 heterodimers were approximately 40 μM, 3.7 μM, and 1.4 μM, respectively, which are somewhat lower than those determined for butein.



EC_{50} (μM)			
Compound	hNPAS1-ARNT2	hNPAS3-ARNT2	mHIF2α-ARNT2
FR-18	244	unstable	76
FR-19	328	n/a	n/a
FR-20	40	3.7	1.9

Figure 132. TSA studies for butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. Butein analogues were tested in a DRC profile with freshly purified hNPAS1/3-ARNT2 and mHIF2α-ARNT2 heterodimers in three independent runs using the TSA assay.

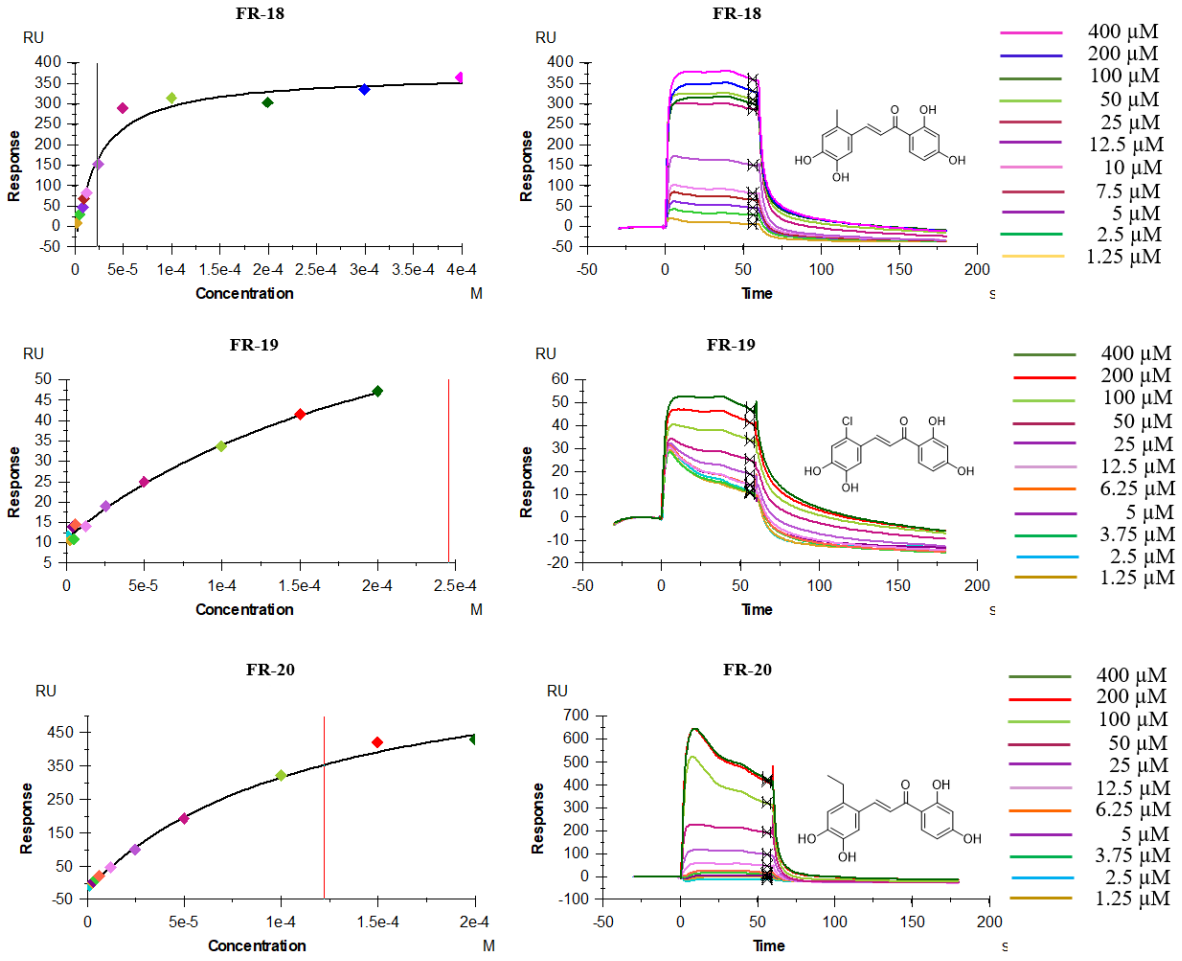
TR-FRET assays revealed significant inhibition of heterodimer activity, particularly in the NPAS3-ARNT2 complex, with FR-18, FR-19, and FR-20 showing 45%, 60%, and 25% inhibition, respectively (Figure 133). Given the TR-FRET inhibition activity and the fact that these compounds were designed to better occupy the binding pocket to form potential additional interactions, higher affinity binding to ARNT2 was anticipated.



IC_{50} (μM)				
Compound	NPAS1-ARNT2-eGFP	NPAS3-ARNT2-eGFP	NPAS4-ARNT2-eGFP	mHIF2α-ARNT2-eGFP
FR-18	n/a	49.6	unstable	unstable
FR-19	unstable	145	unstable	n/a
FR-20	n/a	5	492	unstable

Figure 133. TR-FRET studies for butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. The TR-FRET activity of the butein analogues against ARNT2 with several potential partners, including mNPAS1, mNPAS3, mNPAS4, and mHIF2α.

However, SPR analysis showed that FR-18 had a binding affinity comparable to that of butein, with a K_D of 23 μM as determined by steady-state affinity methods (Figure 134). FR-20 exhibited slightly weaker binding, with a K_D of 122 μM. FR-19 did not reach saturation in its binding curve, and its calculated K_D of 245 μM should be interpreted with caution.



Compound	K_D (µM)
FR-18	23
FR-19	245
FR-20	122

Figure 134. SPR studies for butein analogues designed to fill the hydrophobic pocket of the PAS-B domain of ARNT2. Steady state affinity analyses for FR-18, FR-19, and FR-20 with the biotinylated single PAS-B domain of ARNT2. The calculated K_D values for all three compounds are shown in the blue table.

Although the crystal structure of FR-18 is not available, the structure of FR-20 provides insight into potential binding interactions (Figure 135, Appendix 16). The ethyl group substitution in FR-20 extends deeper into the binding pocket and makes additional hydrophobic contact with Leu382 at a measured distance of approximately 3.8 Å (Figure 135). By comparison, structural modelling suggests that the methyl group in FR-18 would be positioned in close proximity to Ile370, with an estimated distance of ~3.6 Å (Figure 135G). This interaction may contribute to the comparable binding affinity of FR-18 relative to butein.

Interestingly, despite being less hydrophobic, the methyl-substituted FR-18 displays slightly higher affinity than the ethyl analogue, as determined by SPR. This can be explained by differences in desolvation and fit. The ethyl group presents a larger hydrophobic surface area, which necessitates the displacement of more ordered water molecules upon binding, resulting in a higher desolvation penalty. Unless this energetic cost is fully compensated by stronger van der Waals or hydrophobic interactions, the overall binding free energy becomes less favourable. In contrast, the smaller methyl group incurs a lower desolvation cost and fits more snugly within the pocket, without introducing steric strain or unnecessary entropic penalties. These factors are likely to contribute to the improved binding affinity observed for FR-18.

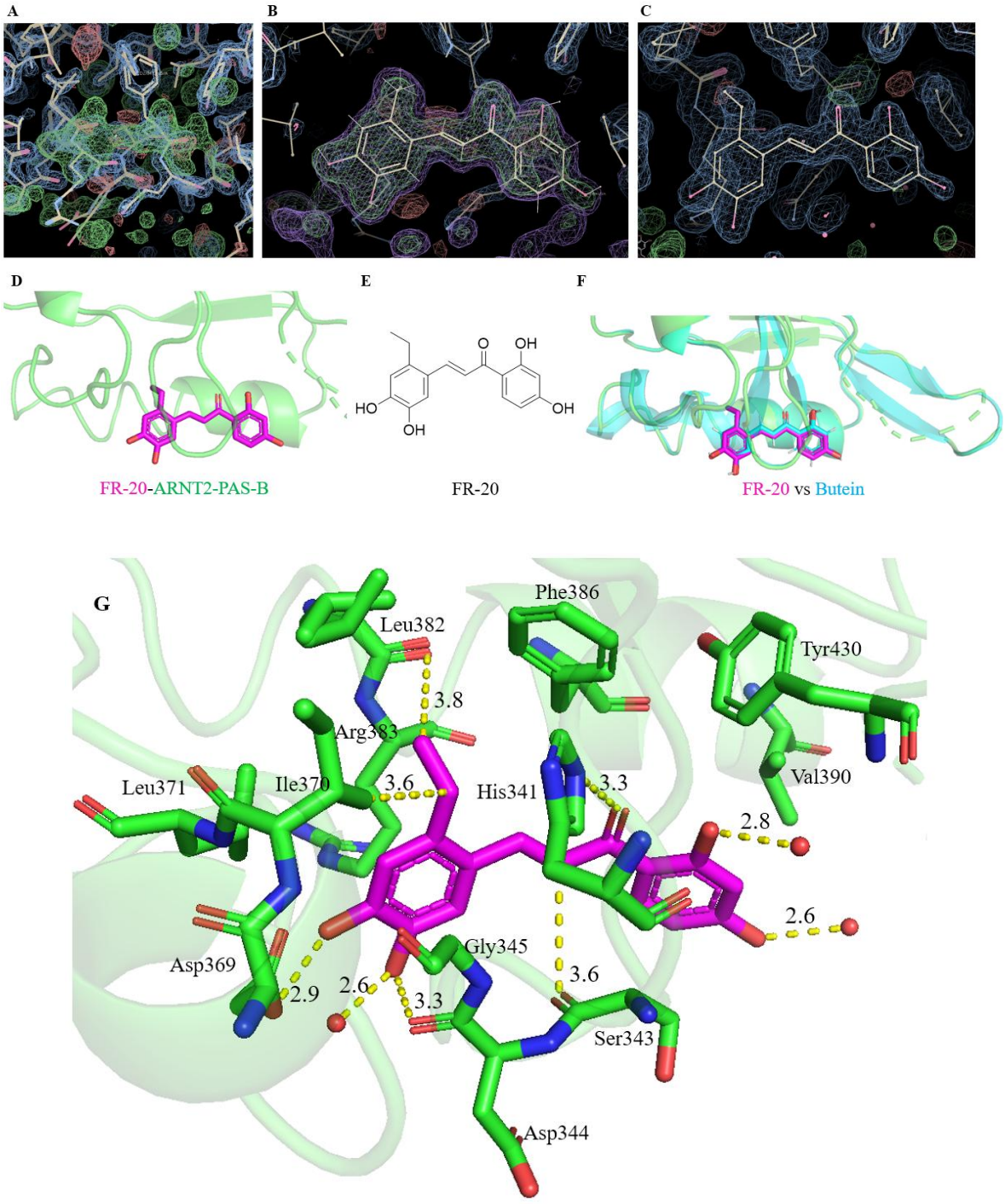


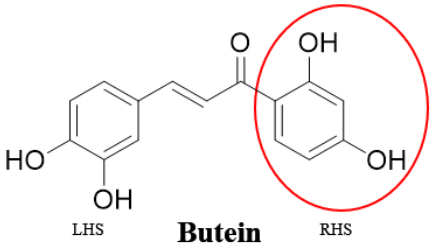
Figure 135. Crystal structure of the FR-20 butein analogue with ARNT2-PAS-B. (A) Density map without ligand. (B-C) Density map with the FR-20 ligand within the ARNT2-PAS-B domain. (D) Close-up look at the location of FR-20 inside the ARNT2-PAS-B domain; FR-20 is shown in pink. (E) Chemical structure of the FR-20 analogue. (F) The butein-bound structure (cyan) was superposed on FR-20 (pink) to show the overlapping

binding location. (G) Interactions of FR-20 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts or interactions with protein residues and water molecules in the crystal structure.

11.5 Design of butein analogues to address selectivity concerns

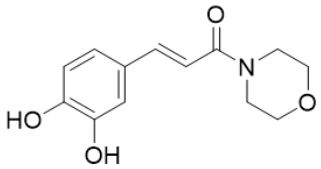
Another critical concern in the hit-to-lead optimisation of butein was its limited selectivity. Butein is known to interact with a broad range of protein targets, raising the potential for off-target effects and associated safety liabilities. This promiscuity is likely driven by the combination of its polyphenolic structure and planar conformation, both of which are features commonly associated with non-specific binding. To improve target specificity and reduce off-target interactions, the design strategy was focused on introducing structural modifications on the right-hand side of butein that would potentially enhance selectivity towards ARNT2 while minimising interactions with unrelated proteins. In this context, a series of novel analogues, including FR-22, FR-23, FR-24, FR-25, and FR-26, were rationally designed by Dr. Srinivas Cheruku.

Efforts to shift the binding profile from broadly active compounds to more target-focused ligands, which represents a key step toward the development of selective ARNT2 modulators, revealed that FR-26 analogue demonstrated measurable binding not only to the mouse ARNT2-PAS-B domain but also to the human protein, as confirmed by TSA studies (Figure 136). Preliminary EC_{50} values, determined by the dose-response binding TSA assay, were 1.7 μM for mARNT2-PAS-B and 17.8 μM for hARNT2-PAS-B, suggesting significant improvement of potential binding affinity compared to butein.



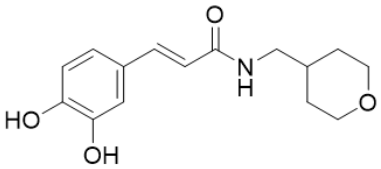
Butein

dTm:= +3.3 °C



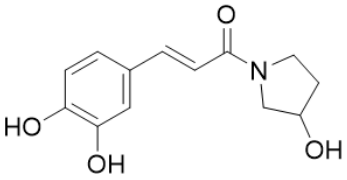
FR-22

dTm: +0 °C



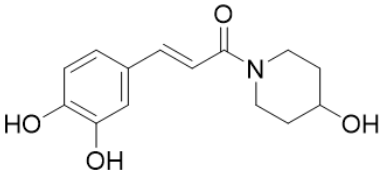
FR-23

dTm: +0 °C



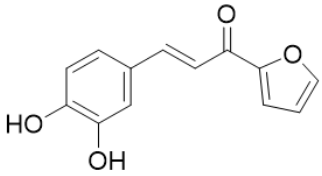
FR-24

dTm: +0 °C



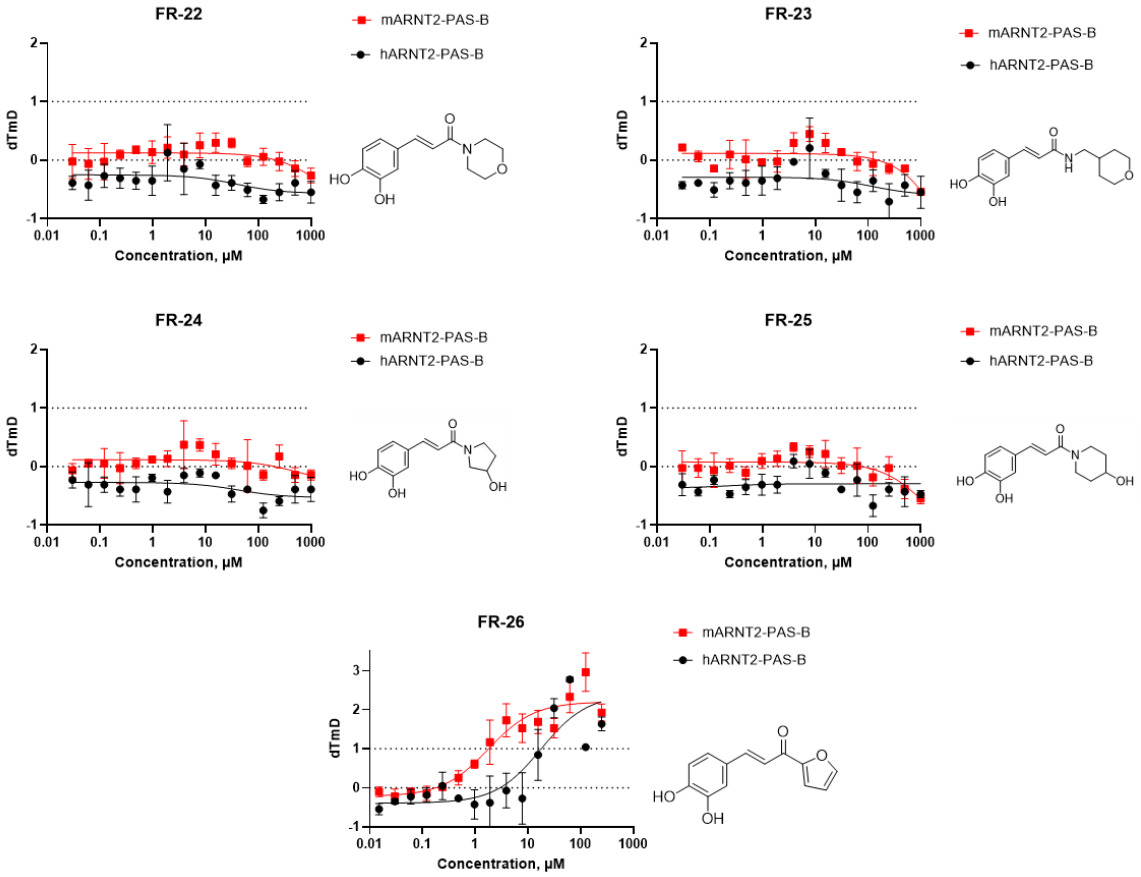
FR-25

dTm: +0 °C



FR-26

dTm: +2.8°C

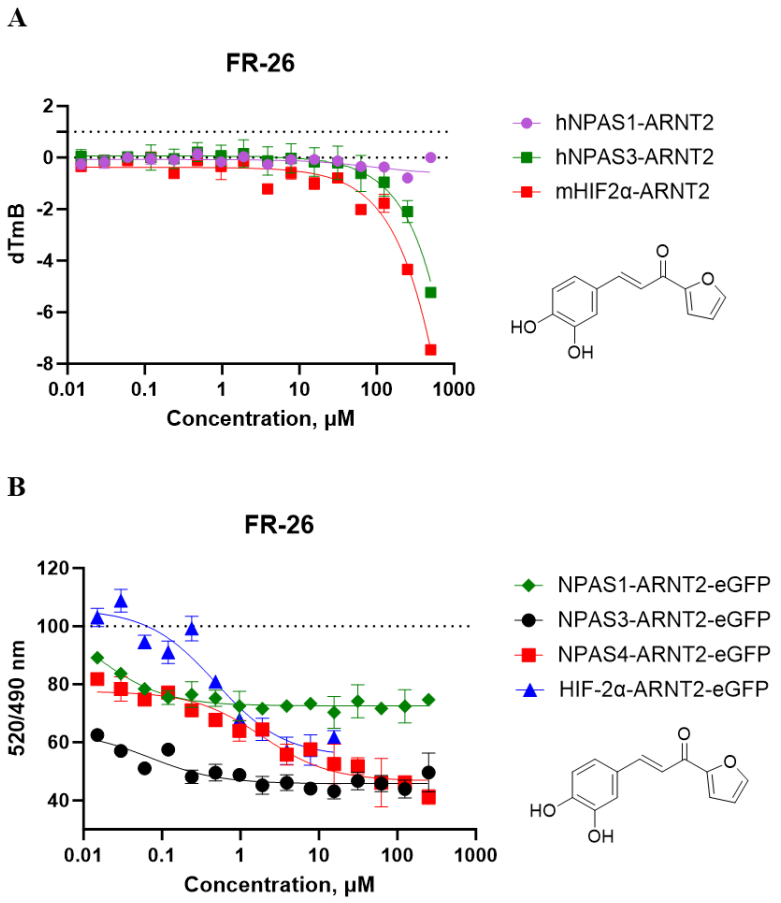


EC_{50} (µM)		
Compound	mARNT2-PAS-B	hARNT2-PAS-B
FR-26	1.7	17.8

Figure 136. Butein analogues with structural modifications on the right-hand side for potential selectivity toward ARNT2. The thermal shift (dTm) for each compound is shown below its corresponding chemical structure, alongside TSA curves depicting the binding effect with mouse and human ARNT2-PAS-B domain.

While TSA measurements did not show any stabilisation of heterodimeric complexes, TR-FRET assays revealed robust functional inhibition of ARNT2 heterodimer activity in all tested complexes (Figure 137). Although the calculated IC_{50} values derived from the TR-

FRET assays consistently fell within the ~1 μM range and were broadly consistent with preliminary EC_{50} values obtained for the isolated PAS-B domain using TSA, the observed TR-FRET activity is likely influenced by assay interference associated with the furan moiety. Furan-containing compounds are known to interfere with fluorescence- and luminescence-based assays due to intrinsic photophysical properties, including compound autofluorescence and absorption within assay-relevant spectral windows.[236] In the context of TR-FRET, such effects can lead to distortion of the donor-acceptor emission ratio through direct emission, acceptor channel bleed-through, or apparent donor quenching, thereby producing false-positive or misleading activity signals. This behaviour is attributable to the aromatic nature of the furan ring, which possesses a delocalised π-electron system capable of absorbing excitation energy and emitting light that may overlap with the donor or acceptor emission wavelengths used in TR-FRET measurements.

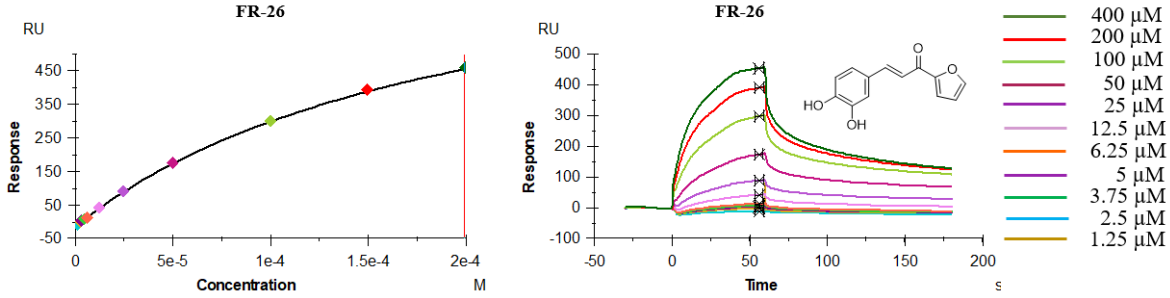


IC_{50} (μM)				
Compound	NPAS1-ARNT2-eGFP	NPAS3-ARNT2-eGFP	NPAS4-ARNT2-eGFP	mHIF2α-ARNT2-eGFP
FR-26	unstable	0.9	1.9	0.5

Figure 137. TSA and TR-FRET studies for the FR-26 analogue. (A) FR-26 was tested in a DRC profile with freshly purified hNPAS1/3-ARNT2 and mHIF2α-ARNT2 heterodimers in three independent runs using the TSA assay. (B) The TR-FRET activity of the FR-26 butein analogue against ARNT2 with several potential partners, including mNPAS1, mNPAS3, mNPAS4, and mHIF2α.

To assess the binding affinity of FR-26, SPR was employed with the biotinylated PAS-B domain of ARNT2. Despite promising TSA and TR-FRET results, affinity analyses revealed

weak binding of FR-26. While SPR measurements produced a saturation binding curve, steady-state analysis yielded a relatively weak K_D of 199 µM, indicative of low-affinity binding under the experimental conditions (Figure 138).



Compound	K_D (µM)
FR-26	199

Figure 138. SPR steady state affinity analysis for the FR-26 analogue with the biotinylated single PAS-B domain of ARNT2. The calculated K_D value is shown in the blue table.

To confirm direct binding, the X-ray crystal structure of the FR-26-mARNT2-PAS-B complex was determined at 1.51 Å resolution (Figure 139, Appendix 16). Structural analysis revealed that FR-26 occupies the same internal hydrophobic cavity within the PAS-B domain as butein and other ligands identified in this study. Consistent with previous ligand-bound structures, including those of KG-548 and butein, FR-26 engages key residues lining the pocket, notably forming interactions with His341. Overall, FR-26 binding induces conformational rearrangements similar to those observed with other ligands in this series, particularly subtle shifts in the β-sheet region, the adjacent α-helix, and the HI loop, suggesting a conserved mode of allosteric modulation across this ligand class.

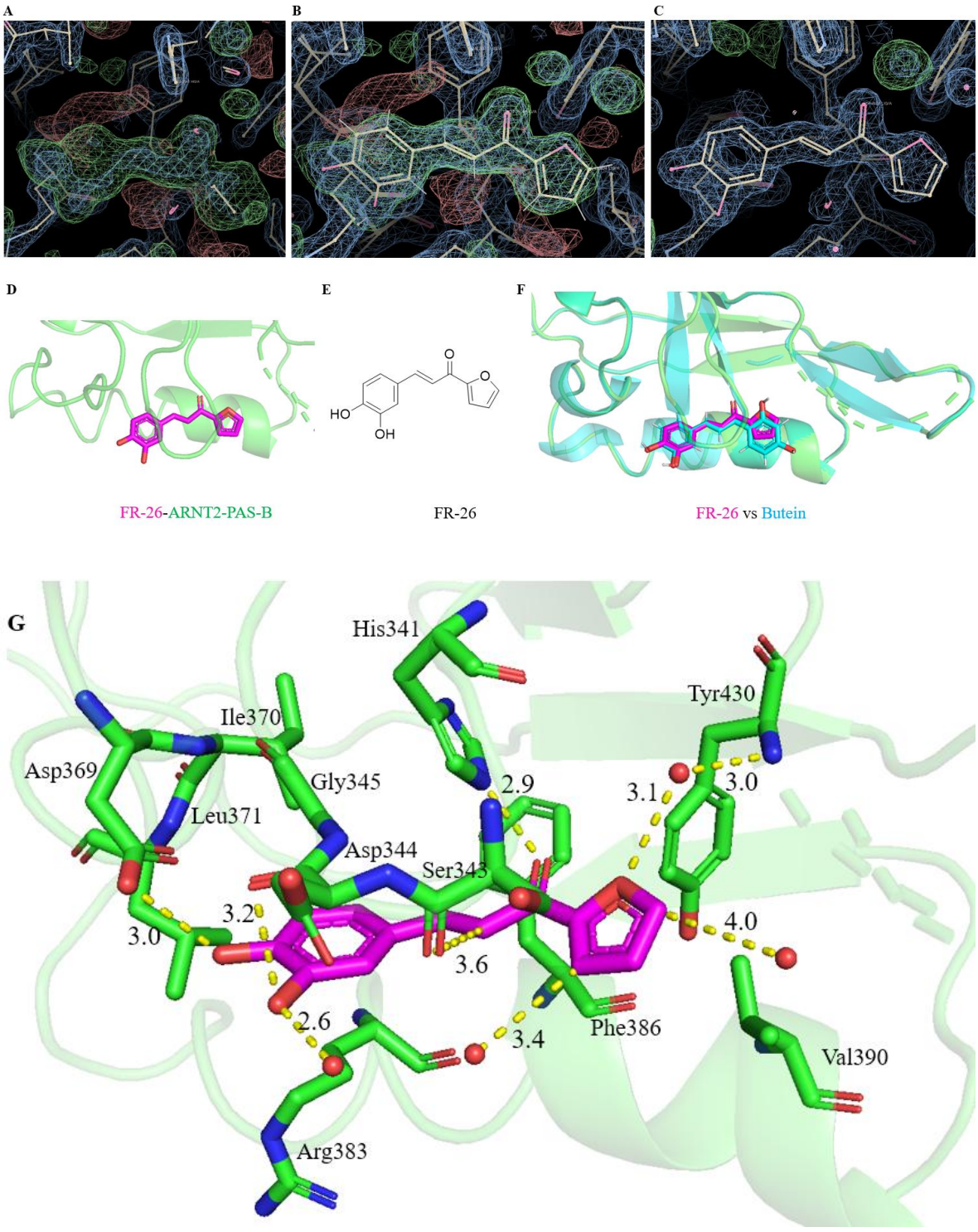


Figure 139. Crystal structure of the FR-26 butein analogue with ARNT2-PAS-B. (A) Density map without ligand. (B-C) Density map with the FR-26 ligand within the ARNT2-PAS-B domain. (D) Close-up view of the location of FR-26 inside the ARNT2-PAS-B domain; FR-26 is shown in pink. (E) Chemical structure of the FR-26 analogue. (F) The

butein-bound structure (cyan) was superposed on FR-26 (pink) to show the overlapping binding location. (G) Interactions of FR-26 (pink) with surrounding residues in the pocket (green). The yellow dashed lines represent close contacts or interactions with protein residues and water molecules in the crystal structure.

A distinctive feature of FR-26 is the slightly inward orientation of its furan ring, which enables a water-mediated hydrogen bond with Tyr430. Positioned 3.1 Å from the ligand, the water molecule forms a hydrogen bond with the backbone amide nitrogen of Tyr430 at a distance of 3.0 Å. This interaction is analogous to the water-bridged contact observed between butein and Ile428. Additionally, butein is further stabilised by two water molecules oriented into the pocket, forming favourable interactions at 2.5 Å and 3.0 Å. In contrast, FR-26 also engages two additional water molecules; however, both are oriented toward the solvent and are therefore unlikely to contribute significantly to ligand stabilisation. These subtle differences in water-mediated interactions and inward orientation may explain the slightly reduced SPR affinity observed for FR-26 compared to butein, despite their overall similar binding profiles.

To conclude, several butein analogues were designed and evaluated using TSA, X-ray crystallography, TR-FRET, and SPR, with the goal of improving affinity, stability, and selectivity for the ARNT2 PAS-B domain. Multiple analogues were successfully co-crystallised with ARNT2 PAS-B, confirming direct binding. Although the affinity of several analogues, including FR-15, FR-20, and FR-26, is comparable to or slightly lower than that of butein, they offer potential advantages such as reduced PAINS liability and improved

chemical stability. These features make them promising tools for future studies targeting ARNT2 PAS-B, and they may serve as valuable chemical probes alongside butein.

Furthermore, butein and its derivatives, similar to KG-548 and related compounds, induce significant conformational changes in the PAS-B domain, including expansion of the binding cavity (Figure 139). This observed plasticity supports previous proposals that expanding ligand size, specifically by extending KG-548 derivatives beyond the tetrazole moiety, could improve binding affinity. Butein, which shows higher affinity than KG-548, partially occupies this newly accessible space within the expanded pocket, providing structural validation for this design strategy. These findings further support the potential for developing more potent ARNT2 PAS-B modulators through rational ligand expansion.

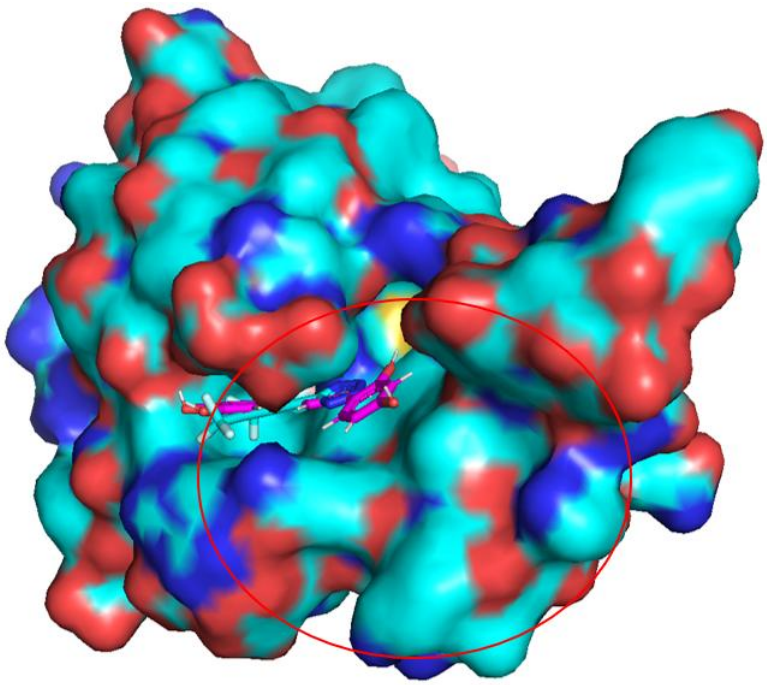


Figure 140. Conformational expansion of ARNT2 PAS-B upon KG-548 and butein binding. The potential ligand design strategies to use the created pocket space are circled in red.

Chapter 12: Closing remarks

12.1 Overall conclusion

The mammalian bHLH-PAS transcription factors are critically involved in a variety of human disease processes. They are increasingly viewed as viable drug targets, as their three-dimensional architectures reveal an ability to bind small molecules within their PAS domains. The neuronal PAS proteins NPAS1 and NPAS3, together with their partner ARNT2, are members of this family and have been genetically linked to human neurological disorders, including schizophrenia, cognitive impairment, bipolar disorders, attention-deficit/hyperactivity disorder, intellectual disability, and major and seasonal depression. These associations make them promising targets for therapeutic intervention in these conditions.

To become functional transcription factors, NPAS1 and NPAS3 must form stable heterodimeric complexes with ARNT or their preferred neuronal partner, ARNT2. In this study, human NPAS1-ARNT2 and NPAS3-ARNT2 variants were successfully generated, overexpressed in *Escherichia coli*, and purified. This represents the first successful production of recombinant human NPAS1 and NPAS3 proteins. In parallel, a substantial quantity of the PAS-B single domain of mouse ARNT2 was also purified.

A high-throughput biochemical screen was conducted using diverse chemical libraries comprising 67,847 compounds. The screening was performed with NPAS1-ARNT2, NPAS3-ARNT2, and ARNT2-PAS-B proteins, using TSA to identify potential small molecule modulators. Primary hits from the initial screen were subsequently validated

through follow-up dose-response analyses using TSA. Artefacts were eliminated by analysing and evaluating the compounds for PAINS characteristics. To remove non-specific compounds, counter-screens were performed using other members of the bHLH-PAS protein family using TSA studies. The most promising compounds identified for NPAS1-ARNT2, NPAS3-ARNT2, and ARNT2-PAS-B were re-ordered as fresh powder for further investigation.

Binding affinities were further evaluated using biophysical techniques, including MST and SPR. While MST proved to be a less reliable technique for this system, SPR measurements successfully quantified binding affinities. Notably, compound 3617-M16 exhibited a substantially stronger binding affinity for the ARNT2-PAS-B domain, with a dissociation constant of 49 μM, as compared to over 400 μM for the reference compound KG-548. These results suggest that 3617-M16 could serve as a potential chemical probe for future studies assessing ARNT2 function.

To assess the functional activity of the tested compounds and determine whether they act as enhancers or suppressors, a TR-FRET assay was carried out using several heterodimeric complexes within the bHLH-PAS protein family. The results revealed that most compounds promoted heterodimerisation, indicating an overall upregulation of activity. Although the calculated EC_{50} values were generally consistent with results from other assays, including MST, SPR, and TSA, subsequent crystal structure analyses strongly suggested that the lead compounds may in fact exert an inhibitory effect.

For structural studies, the crystal structure of the apo mouse ARNT2-PAS-B domain was successfully determined, whereas NPAS1-ARNT2 and NPAS3-ARNT2 proteins proved

difficult to crystallise, limiting structural validation of ligand binding to these complexes. Several crystal structures of the ARNT2 PAS-B domain in complex with small molecules were resolved with defined orientation and mechanistic clarity. Structural analyses revealed that all ligands bind within a sizable cavity in the PAS-B domain, measuring approximately 148 $Å^3$ in volume. In each complex, the ligands consistently interacted with the His341 residue, suggesting that this amino acid plays a critical role in stabilising ligand binding within the PAS-B pocket.

Ligand binding consistently induced similar conformational changes within the ARNT2 PAS-B domain to accommodate the compounds. In particular, the repositioning of surrounding residues led to structural alterations in β-sheets, α-helices, and the HI loop, which are regions critical for maintaining heterodimer stability. These conformational changes are likely to destabilise the heterodimer interface, thereby impairing the formation or stability of ARNT2-containing heterodimers. Compared to the apo structure of mARNT2 PAS-B, these rearrangements also caused a significant expansion of the binding pocket, suggesting promising opportunities for further hit-to-lead optimisation.

In parallel with the identification of new chemotype, hit-to-lead optimisation was carried out for the control compounds KG-548 and butein. TSA studies identified LS02-174 as a promising analogue of KG-548. The SPR analysis, however, revealed no significant improvement in binding affinity compared to the parent compound. Among the butein analogues, FR-15, FR-20, and FR-26 exhibited similar or slightly lower binding affinities relative to butein. Nevertheless, these analogues may offer advantages as chemical tools due to their lower PAINS scores, reduced off-target activity, and potentially enhanced stability and selectivity for the ARNT2 PAS-B domain.

The structural and biochemical analyses of KG-548, butein, and their analogues provide important insights into the nature of ligand recognition at the ARNT2-PAS-B domain, which may inform the future development of more potent and selective modulators. Although both ligand classes occupy the same hydrophobic cavity, their binding modes and functional group preferences highlight distinct interaction strategies for achieving affinity and specificity.

Butein engages the binding site through its α,β-unsaturated carbonyl group, which functions as a key hydrogen bond acceptor within the site. This carbonyl forms a direct interaction with the conserved His341 residue, contributing to a stable and favourable binding conformation. The planar and rigid scaffold of butein allows optimal positioning of the carbonyl and neighbouring hydroxyl groups to form water-mediated hydrogen bonding networks, maximising contacts within the hydrophobic environment, providing its relatively high affinity to the protein.

In contrast, KG-548 and its analogues lack the α,β-unsaturated carbonyl group and instead rely on acidic moieties, such as a tetrazole ring, carboxylate, or sulfonate, to mediate binding. These acidic functional groups act as strong hydrogen bond acceptors and are essential for forming direct hydrogen bonds, including potential salt bridges, with the positively charged His341 residue. These electrostatic interactions compensate for the absence of a carbonyl group and are critical for anchoring the ligand within the binding pocket.

The differing binding requirements of these ligand classes underscore the versatility of the ARNT2 PAS-B domain. The pocket is able to accommodate chemically distinct scaffolds by supporting either hydrogen bonding through carbonyl groups or ionic interactions via acidic functionalities. This plasticity suggests that the pocket possesses multiple interaction

hotspots and that ligand affinity and specificity can be tuned by targeting different residues or interaction networks within the pocket.

From the perspective of drug design, this study has proven that the ARNT2-PAS-B domain is a flexible and ligand-responsive pocket capable of accommodating chemically diverse scaffolds through distinct interaction strategies, including hydrogen bonding, ionic contacts, and water-mediated networks. The comparative analysis of butein and KG-548 analogues not only reveals how different functional groups drive binding affinity but also illustrates how subtle changes in ligand structure can dramatically alter binding capability, either enabling or completely disrupting interaction with the protein. These insights provide a strong foundation for the rational design of next-generation ARNT2 modulators. The data also support the strategy of expanding into newly accessible regions within the binding pocket to improve both potency and selectivity. Alternatively, another viable strategy involves targeting a distinct, smaller cavity within the PAS-B domain, with an approximate volume of 105 $Å^3$, which may also be accessible to ligands. Exploring this secondary pocket with novel compounds could lead to the discovery of alternative binders with improved affinity and potentially unique binding modes.

12.2 Future studies

Despite the importance of NPAS1, NPAS3, and ARNT2 in neuronal processes, their molecular functions and target genes are largely uncharacterised. Although there is growing genetic and transcriptomic evidence linking these transcription factors to human neurological disorders, the molecular and cellular mechanisms by which NPAS1, NPAS3, and ARNT2

influence brain development and neuronal function remain poorly understood. Understanding their roles could open the door to novel diagnostic markers and therapeutic targets for a wide range of currently intractable neuropsychiatric disorders.

This DPhil study employed a range of *in vitro* biophysical and biochemical techniques to characterise the interaction between NPAS1/NPAS3 and ARNT2, and to identify small molecule compounds that can bind within their protein architecture. Initial compound screening via TSA allowed the identification of ligands with potential binding affinity. These hits were further validated and quantified using MST and SPR, providing insight into binding kinetics and affinities. Structural analysis through X-ray crystallography confirmed compound binding at putative interaction sites, and TR-FRET assays were subsequently used to assess compound effects on heterodimerisation.

Overall, this DPhil study demonstrated that the transcription factors NPAS1, NPAS3, and their dimerisation partner ARNT2, are capable of binding and accommodating diverse small molecules within their ligand-binding pockets. This discovery opens new avenues for understanding the functional roles of these proteins, particularly the possibility that their transcriptional activity may be modulated in a ligand-dependent manner. Future research, therefore, should aim to establish a coherent framework linking ligand binding to functional outcomes, with a specific focus on how small molecules influence NPAS1-ARNT2 and NPAS3-ARNT2 heterodimerisation and downstream transcriptional regulation in a cellular context.

While the TR-FRET assay offered valuable preliminary functional insights, signal variability and limited sensitivity in complex mixtures highlighted the need for more physiologically

relevant methods. To build upon these current *in vitro* findings and to validate NPAS1/3-ARNT2 heterodimerisation by modulation with small molecules in live-cell environments, future studies could employ a NanoBRET-based protein-protein interaction assay. NanoBRET is a luminescence-based resonance energy transfer technique that enables real-time monitoring of protein interactions in live cells. Unlike TR-FRET, NanoBRET utilises a NanoLuc luciferase as an energy donor, eliminating the need for external excitation light and thereby significantly reducing background autofluorescence and improving the signal-to-noise ratio. This feature makes NanoBRET particularly well-suited for studying dynamic protein interactions in living cells with higher sensitivity and reliability.

Complementary methods such as Co-IP and proximity ligation assays (PLA) could also be employed to orthogonally validate protein-protein interactions and ligand effects in cells. Together, these approaches will help establish the cellular relevance of the NPAS1/NPAS3-ARNT2 interaction, and determine whether heterodimerisation could be modulated with selective small molecule ligands *in vivo*.

If a molecule potently stabilises or destabilises the formation of NPAS1/3-ARNT2 heterodimers in the primary screen, and during TR-FRET or NanoBRET assays, this potentiation is expected to manifest itself in enhanced gene expression within cells. To identify a compound that biochemically modulates heterodimerisation could also pass through a cell membrane and activate the transcriptional function of the protein of interest in terms of target gene expression, a cell-based transcriptional reporter assay could be employed.

A luciferase reporter assay is a powerful and widely used molecular biology technique that allows the quantitative measurement of gene expression and intracellular signalling events. Fundamentally, it enables determination of whether a protein of interest can activate or repress the expression of a target gene using luciferase as a reporter protein.

The fundamental principle of the assay is the coupling of a luciferase gene to a promoter or response element of interest. When this engineered construct is added into cells, usually via transfection, the activity of the promoter drives the expression of luciferase. Upon addition of the substrate, the amount of produced bioluminescence is directly proportional to the transcriptional activity of the promoter, thus acting as an indirect measure of gene expression or signalling pathway activity. This system therefore enables the investigation of promoter activity, binding of TF, and intracellular responses to stimuli such as compounds or drugs.

Validated and optimised hits could further undergo more rigorous analysis of cell activity, relying on RNA-sequencing and RT-qPCR analyses to detect whether ligands can affect the target gene expressions from several neuronal cell lines expressing NPAS1 or NPAS3, and ARNT2 proteins, as well as knockout control cells.

To identify which genes could be activated and repressed in the presence of ligands, gene expression could be quantified by RNA-sequencing. Gene expression of novel and known target genes, including but not limited to *FGF*, *EPO, BDNF, FMR1, UBE3A*, *NOTCH,* could also be evaluated by a follow-up RT-qPCR analysis.

The future studies proposed here could not only serve as a prerequisite for identifying a potent ligand but also help to elucidate the critical and complex mechanistic links between ligand binding and the transcriptional regulation of these proteins. This knowledge will help define

the genetic programs regulated by NPAS1 and NPAS3 in partnership with ARNT2, offering new insights into the molecular mechanisms underlying neurological disease pathogenesis. Ultimately, these findings will provide a foundation for the therapeutic targeting of these transcriptional regulators, paving the way for the development of novel treatment strategies for neurological disorders.

PART III: Materials and Methods

Chapter 13: Preparation of materials and buffers

13.1 Lysogeny Broth (LB) media

To prepare the LB media, 20 g of LB Broth (Lennox) powder (Sigma-Aldrich, L3022-1KG) was dissolved in 1 L of deionised water and autoclaved. The autoclaved media were allowed to cool to room temperature before usage.

13.2 LB Agar plates

To formulate agar plates, 8 g of LB Broth (Lennox) powder was mixed with 6 g of Agar powder (Sigma-Aldrich, A6686-1KG). The powder mixture was dissolved in 400 mL of deionised water and autoclaved. Autoclaved media was allowed to cool to 55 °C, and then appropriate antibiotics were added. The mixture was gently swirled and then poured into sterile petri dishes (20 mL per dish). Once solidified, plates were dried in an incubator at 37 °C before usage.

13.3 Antibiotics

Antibiotic stock solutions were prepared at the following working concentrations: 100 mg/mL for ampicillin (Sigma-Aldrich, A9518-100G), and 50 mg/mL for both kanamycin (Fluorochem, M02038) and streptomycin (Sigma-Aldrich, S9137-100G), using autoclaved deionised water as the solvent. Working stock of 34 mg/mL chloramphenicol (Serva, 16785.02) was dissolved in molecular biology grade ethanol (Total Lab Supplies,

MO330). Antibiotics were used at a ratio of 1 mL of antibiotic solution per 1 L of culture medium.

13.4 Agarose gel

To make 1% w/v agarose gel, 1 g of agarose powder (Fisher Scientific, BP161-100) was dissolved in 100 mL of 1x Tris-acetate-EDTA (TAE) buffer (Alfa Aesar, J69677) by heating in a microwave until dissolution was complete. After the solution cooled to 45 °C, 10 µL of SYBR Safe DNA Gel Stain (Invitrogen, S33102) was added. The solution was poured into a gel cast and allowed to solidify for 30 minutes. The agarose gel was run at 100 V for 30 minutes in 1x TAE buffer.

13.5 Buffers for protein purification and HTS screening

All buffers were prepared according to the formulations described in Table 6. Protein buffers were filtered through a 0.22 µm sterile filter (Starlab, CC8221-5226) prior to use.

Buffer	Buffer composition
Binding buffer (BB)	50 mM HEPES pH 7.5, 20 mM Imidazole, 500 mM NaCl, 0.5 mM TCEP, 10% Glycerol
Washing buffer	Binding buffer containing 1 M NaCl
Elusion buffer	Binding buffer containing 250 mM Imidazole
Cation exchange A (CE-A) buffer	20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 5% Glycerol
Cation exchange B (CE-B) buffer	CE-A buffer containing 1 M NaCl
Gel-filtration buffer	20 mM HEPS pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 5% Glycerol
Screening buffer for NPAS1/3-ARNT2 heterodimers	20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP
Screening buffer for single PAS-B domain of ARNT2	50 mM Sodium phosphate pH 6.5, 500 mM NaCl, 0.5 mM TCEP

Table 6. Buffer composition for protein purification and TSA studies.

Chapter 14: Methods and experimental approach

14.1 Plasmid construction and general cloning procedure

To generate hNPAS1 and hNPAS3 constructs with their partner hARNT2 in the pETDuet-1 vector, Gibson assembly was used as a molecular cloning method. A set of oligos was designed to amplify contiguous bHLH-PAS-A-PAS-B segments of hNPAS1, hNPAS3, and hARNT2 by PCR with high-fidelity DNA polymerase (New England Biolabs (NEB), M0493S) (Table 7). cDNAs of human NPAS1 (HsCD00515966), NPAS3 (HsCD00080332), and ARNT2 (HsCD00039862) were purchased from DNASU and used as templates. The reaction mixture was prepared as directed in Table 8. PCR reactions were conducted in a thermocycler, using the general programme (Table 9)

PCR products were purified using the DNA Clean & Concentrator Kit (Zymo Research, D4080) following the manufacturer's protocol. Purified hNPAS1 (residues 43-423), hNPAS3 (residues 19-437), and hARNT2 (residues 57-438) PCR fragments were cloned into the pETDuet-1 vector using a Gibson assembly reaction. To clone hNPAS1 and hNPAS3 constructs, pETDuet-1 was digested for one hour at 37 °C with BamHI-HF (NEB, R3136S) and HindIII-HF (NEB, R3104S) restriction enzymes. The linearised vector and purified hNPAS1 and hNPAS3 fragments were added to a Gibson Assembly Master Mix (NEB, E2611S) and incubated at 50 °C for 1 hour. Both constructs were then transformed into DH5α competent *E. coli* cells (Zymo Research, NC9911229) and plated into LB agar plates with ampicillin. Selected colonies were grown in LB medium supplemented with ampicillin, and

plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen, 12125) following the manufacturer's protocol.

To clone hARNT2, previously generated pETDuet-1 vectors with hNPAS1 and hNPAS3 inserts were digested with BglII (NEB, R0144S) and KpnI-HF (NEB, R3142S) restriction enzymes. Purified hARNT2 PCR product was cloned into both linearised vectors followed by DH5α bacterial transformation and plasmid's purification protocols. BioDrop µLITE+ (Montreal Biotech) was used to quantify and assess the purity of DNA.

Name	Primer's sequence
hNPAS1 (43–423)_FP	5′CCGCGAAATTAATACGACTCACTATAGGGG 3′
hNPAS1 (43–423)_RP	5′TTAAGCATTATGCGCGCGCAAGCTTTTAGCTGGCTGGAAGCTGGAAG 3′
hNPAS3 (56–455)_FP	5′ACCATCATCACCACAGCCAGGATCCAGGTTCTGATTACGATATCCCAA CGACCGAAAACCTGTATTTTCAGGGACAAGCATTGAGAAAGGAGAA 3′
hNPAS3 (56–455)_RP	5′TTAAGCATTATGCGCGCGCAAGCTTTTAATGGGGAGCTGTGCGATGT 3′
hARNT2 (57–438)_FP1	5′AAGGAGATATACATATGGCAGATCTCGATTACAAGGATGACGACGAT AAGGTTCTTG 3′
hARNT2 (57–438)_FP2	5′TGACGACGATAAGGGTTCTGATTACGATATCCCAACGACCGAAAACCT GTATTTTCAGGGAGAAGGCCCCAGTAAATTTTC 3′
hARNT2 (57–438)_RP	5′TTTACCAGACTCGAGGGTACCTTAGACGTTGGTGTTGGTGCAGA 3′

Table 7. hNPAS1, hNPAS3, and hARNT2 primers designed for the Gibson Assembly.

Chemicals	Quantity (µL)	Final concentration
Polymerase buffer (5x)	10	1x
dNTP solution (10 mM)	1	200 µM
Forward primer (FP) (10 µM)	1.5	0.3 µM
Reverse primer (RP) (10 µM)	1.5	0.3 µM
DNA template	0.5	0.2 ng µL
DNA polymerase	0.5	
Deionised water	Up to 50 µL	

Table 8. General PCR reaction mixture.

Cycle	Denaturation	Annealing	Extension
1	95 °C for 3:00	-	-
2-34	95 °C for 0:15	63 °C for 1:00 *	72 °C for 0:15 #
35	-	-	72 °C for 5:00

Table 9. General programme for the PCR method. * 63 °C was calculated as the lowest Tm of primers minus - 5 °C; # 0:15 was set based on the speed of the polymerase and amplification length as per supplier's instruction.

14.2 Bacterial transformation and protein overexpression

For protein overexpression, hNPAS1-ARNT2 and hNPAS3-ARNT2 constructs in pETDuet-1 vector, as well as in-house recombinant plasmids pSJ2-hNPAS1-PAS-B (residues 302-421), pSJ2-hNPAS3-PAS-B (residues 319-389), and human (residues 332-442) and mouse (residues 332-438) pMKH-ARNT2-PAS-B plasmids, were used to transform competent BL21-Codon Plus (DE3)-RIL cells (Agilent Technologies, 230245) using the heat shock method.

Plasmids containing genes of interest were added to each tube of BL21-Codon Plus cells at a concentration of 50 ng. The tubes with the transformation mixture were incubated on ice for 30 minutes and then placed in a 42 °C water bath for 45 seconds. After the heat shock procedure, tubes were placed on ice for 2 minutes, followed by 1 hour of shaking incubation at 37 °C with super optimal broth with catabolite repression (SOC) outgrowth medium (Sigma-Aldrich, S1797). To grow selective colonies, the transformation mixture was plated on LB-agar plates containing 100 µg/mL ampicillin for pSJ2 plasmids, or 50 µg/mL

kanamycin for pMKH plasmids, together with 34 μg/mL chloramphenicol. Plates were incubated overnight at 37 °C.

Selected colonies were transferred into 10 mL of LB under either ampicillin or kanamycin together with chloramphenicol resistance control, and were grown overnight at 37 °C under constant shaking. The overnight culture was added into 990 mL of LB medium containing the same antibiotics, and the culture grew at 37 °C until an OD_{600} of 0.6-0.8 was reached. The protein expression was induced by the addition of 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Fluorochem, MO2726), and the culture was expressed for 18 hours at 18 °C. The bacterial cells were pelleted by centrifugation at 6000 rpm for 25 minutes at 4 °C. The cell pellets were instantly frozen at -80 °C.

14.3 Protein purification

14.3.1 Immobilised metal affinity chromatography for 6xHis-tagged proteins

Bacterial cell pellets were thawed and resuspended in binding buffer (Table 6) supplemented with a protease inhibitor cocktail (Roche, 05056489001), DNase I (Sigma, DN25-100MG), and Lysozyme Type VI (MP Biomedicals, 195303). For each 10 g of pellet, 100 mL of binding buffer was used. Whilst kept on ice, bacterial cells were lysed by sonication for 40 minutes with 10 second pulses followed by 20 seconds recovery time per cycle. To remove cell debris, the cell lysates were centrifugated for 45 minutes at 21,000 rpm at 4 °C. The supernatant was filtrated and applied to a HisTrap FF affinity column (Cytiva, 17531901) that had been previously equilibrated with binding buffer on ÄKTA start™ machine (Cytiva, 29022094). The column was washed with binding and wash buffers (Table 6) until no more

unbound protein was eluted. The 6xHis-tagged protein of interest was then eluted by elution buffer (Table 6). The fractions containing protein of interest were analysed by SDS-PAGE.

14.3.2 Cation exchange-based isolation of target proteins

The bound heterodimers were further purified by a HiTrap SP HP cation exchange column (Cytiva, 17115201), which had previously been equilibrated with CE-A buffer (Table 6). The eluate from the HisTrap column was resuspended in CE-A buffer in ratio 1:5 and applied to the column. The column was then washed with CE-A buffer, and protein of interest was eluted in a 20% and 35% steps isocratic elution by CE-B buffer. The fractions containing protein of interest were analysed by SDS-PAGE.

14.3.3 Gel filtration

After cation exchange chromatography, the eluted heterodimeric fractions were concentrated, and 1 mL of protein solution was injected via sample loop to a HiLoad 16/600 Superdex 200-pg size exclusion column (Cytiva, 28989335) previously equilibrated with gel-filtration buffer (Table 6). For the purification of a single PAS-B domain, protein fractions eluted from metal affinity chromatography were concentrated and loaded on a HiLoad 16/600 Superdex 75-pg size exclusion chromatography column (Cytiva, 28989333). The proteins of interest were eluted by gel-filtration buffer, and the protein fractions were analysed by SDS-PAGE. The concentration of protein was measured on the BioDrop µLITE+ using extinction coefficients calculated by the ExPASy ProtParam tool.

14.4 SDS-PAGE and Western Blotting

14.4.1 SDS-PAGE analysis of protein samples

Protein samples were resolved using SDS-PAGE. For SDS-PAGE, 5 µL of protein sample was mixed with 20 µL of 4x NuPAGE™ LDS sample buffer (Invitrogen, NP0007). The mixture was heated at 95 °C for 10 minutes. The samples and the Precision Plus protein ladder (BioRad, 1610393) were loaded into NuPAGE 4-12% Bis-Tris gel (Invitrogen, NP0322BOX). The gel was run in NuPAGE MES SDS running buffer (Invitrogen, NP0002) at 170 V for 45 minutes until the bromophenol blue dye reached the bottom of the gel.

Gels were stained by immersion in QuickBlue Protein Stain solution (Lubio science, LU001000) and incubated overnight at room temperature under constant shaking. Gels were then washed with water for several hours to remove any background staining, and protein bands were visualised using the Coomassie blue filter settings on a BioRad Universal Hood III.

14.4.2 Western Blot analysis for protein validation

Protein samples were prepared and run on an SDS-PAGE gel as described in section 14.4.1. For Western Blotting, protein samples were then transferred at 60 V for 1 hour to a PVDF membrane (0,2 µm, BioRad, 1620177) using a semi-dry apparatus (BioRad, 1703940). Following transfer, the membrane was blocked with 5% non-fat dry milk (Millipore, 70166) for 1 hour at room temperature. The blots were then probed with anti-His tag (1:5000 dilution, Biolegend, 362602) and anti-FLAG tag (1:2500 dilution, Sigma, F1804-1MG) primary antibodies overnight at 4 °C in blocking buffer (LI-COR, 927-70001). The membrane was then incubated with fluorescent anti-Mouse-HRP (1:10000 dilution, LI-COR, 926-68072)

secondary antibodies for 1 hour at room temperature. Each incubation with antibodies was followed by washing the membrane three times for 10 minutes with 1x TBS-Tween (Sigma-Aldrich, T5030). The signal was detected using Odyssey Gel Imager (LI-COR, Model 9140).

14.5 Protein analysis via mass spectrometry

To remove the 6xHis-tag, TEV Protease, previously synthesised by Dr. Minghao Zhang in the lab, was added into protein solution in a molar ratio of 1:100. The solution was left overnight at room temperature for cleavage. To remove TEV protease, the overnight sample was applied to a small column containing 3 mL of Ni-NTA resin (Millipore, 70666-5) pre-equilibrated with binding buffer to perform reverse IMAC. All fractions (flowthrough, wash, and eluate) along with uncut protein were analysed by SDS-PAGE. Samples were then uploaded to the RapidFire High-throughput Mass Spectrometry System (Agilent Technologies), and the intact mass was analysed by 6530 Accurate-Mass Q-TOF LC/MS (Agilent Technologies). The results were analysed by MassHunter Qualitative Analysis B.07.00 software.

14.6 Thermal shift analysis for protein stability evaluation

14.6.1 Compound handling and preparation for HTS chemical screen

All custom chemical libraries, consisting of fragments and small molecules sourced from Prestwick, ChemDiv, LifeChem, Electrophile, MedChemExpress, and Enamine, are stored as DMSO stock solutions. For the primary screen, 50 nL of each compound was dispensed

into 384-well PCR plates (Azenta Life Sciences, 4ti-0384/C) using the Echo 555 liquid handler (Labcyte, 001-10080), so that a final assay concentration of 1% DMSO was achieved. For the DRC analysis, six compound concentrations, including 50 nL, 25 nL, 12.5 nL, 7.5 nL, 5 nL, and 2.5 nL, were dispensed into 384-well plates, followed by DMSO backfill to reach a total volume of 50 nL.

Fresh-powder analyses required resuspending newly purchased powder compounds in DMSO stocks. To generate stock Echo plates, compounds were titrated through a 16-point serial dilution, starting from 100 mM or 50 mM stock, depending on the compound's solubility. For DRC analysis via TSA, 50 nL of each concentration was dispensed from stock plates into PCR plates.

14.6.2 Thermal shift assay

Each hNPAS1-ARNT2 and hNPAS3-ARNT2 heterodimer was diluted to a concentration of 0.1 mg/mL or 1.1 μM with screening buffer (Table 6). To monitor protein unfolding, the fluorescent dye SYPRO Orange (Invitrogen, S6651) was added to the protein mixture at a 10x final concentration. For a single ARNT2-PAS-B domain, the protein was diluted to a concentration of 0.4 mg/mL or 25 μM with 5x SYPRO Orange concentration in screening buffer. These mixtures were dispensed into 384-well plates containing compounds using Multidrop Combi Reagent Dispenser (Thermo Fisher Scientific, 5840340) to a final assay volume of 5 μL. Plates were then sealed with clear heat-sealing films (Azenta Life Sciences, 4ti-0500) and centrifuged briefly at 1,000 rpm to ensure even liquid distribution in the bottom of the wells.

Plates were incubated at room temperature for more than 15 minutes before measuring. The thermal shift measurements were carried out in 0.13 °C/s increments from 35-85 °C in the QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, 4485701). The results were analysed using Protein Thermal Shift™ Software v1.4, QuantStudio Real-Time PCR Software, and visualised by GraphPad Prism 9. Screening data were analysed in DataWarrior version 6.1.0. The terms dTmB and dTmD refer to the shift of protein melting temperature, calculated by the Boltzmann fitting method (dTmB) and the derivative fitting method (dTmD), respectively.

14.7 Protein crystallisation, structure determination, and refinement

Purified Tev-cleaved ARNT2-PAS-B protein was concentrated in the original HEPES-containing gel filtration buffer using concentrators with a molecular weight cut-off of 5 kDa (Thermo Scientific, 15502023), centrifuging at 4,000 rpm at 4 °C until a concentration of at least 10 mg/mL was reached. Crystals were grown in a 96-well crystallisation plate (SwissCi, MRC96T-UVP) through sitting-drop vapour diffusion at 16 °C. The reservoir solution consisted of 0.1 M HEPES pH 8.0 and 3 M NaCl crystal buffer. Crystal clusters aged one week were mechanically crushed using a crystal crusher probe to generate seed stock, which was subsequently diluted at a 1:1,000 ratio with crystal buffer. Crystallisation of the ARNT2-PAS-B was carried out by mixing 30 nL of crystal seed solution with an equal volume of protein (10 mg/mL) and reservoir solution containing crystal buffer using the Mosquito liquid handler (SPT Labtech, HTS LV model). Crystal plates were sealed (SwissCi, UVXPO-

SEAL) and incubated at 16 °C for three weeks before harvesting. Compound soaking was performed 24 hours before harvesting at a concentration of 1.25 mM. Before flash freezing in liquid nitrogen, crystals were soaked in reservoir plus 30% glycerol as the cryoprotectant.

Diffraction datasets were collected at the Diamond Light Source on beamline I03, using a wavelength of 0.9763 Å. Data were recorded on an Eiger2 XE 16M detector at a crystal temperature of 100 K. The initial phases were determined by molecular replacement with apo-ARNT2-PAS-B refined model with Phaser. The ligand-ARNT2-PAS-B structures were solved in the P62_2_2 space group. The model of ligand-ARNT2-PAS-B complex structures was built and refined using Coot.[237] Further structure refinements were done in Refmac5.[238] All the structural figures were prepared using PyMoL.

14.8 Microscale thermophoresis analysis of protein-ligand interactions

The RED-tris-NTA second generation dye (NanoTemper, MO-L018) was suspended in a small volume of PBS-T to prepare a 50 nM working solution. A His-tagged protein was prepared at a concentration of 4 μM in PBS-T. To create a concentration gradient, a 1:1 serial dilution was performed by mixing 10 μL of protein solution with 10 μL of PBS-T across a 16-point concentration series. Following the dilution, 10 μL of 50 nM RED-tris-NTA dye solution was added to each well and mixed gently by pipetting. The samples were then incubated for 30 minutes at room temperature to allow for labelling. The samples were then loaded into Monolith premium capillaries (NanoTemper, MO-K025) and measured at 40% LED/excitation power and 40 % MST power with the temperature set to 25 °C using the

Monolith NT.115 system (NanoTemper). The protein/dye interaction K_D was determined using the MO.Affinity analysis program with the K_D fit model.

To generate assay-ready plates for DRC analysis with tested compounds, 100 nL of each compound concentration was dispensed from stock Echo plates into a black, low-volume 384-well microplate (Corning, 4514). For protein-ligand interactions, 250 nM of mARNT2-PAS-B and hNPAS3-ARNT2, and 125 nM of hNPAS1-ARNT2 protein were labelled with 25 nM of RED-tris-NTA dye for 30 minutes in the dark at room temperature. Following incubation, the labelled protein solution was centrifuged at 12,000 rpm for 15 minutes at 4 °C. Then, 10 μL of the supernatant was dispensed into each well of the assay plates containing the test compounds. After gentle mixing, the plates were sealed and centrifuged at 1,000 rpm for 1 minute. The plates were then incubated on a gentle rotary shaker at room temperature for 1 hour to allow binding interactions to take place. Solution from each well was then loaded into Monolith premium capillaries and placed in the sample tray. MST experiments were conducted using the Monolith NT.115 system at 40% LED/excitation power and 40% MST power, with the temperature set to 25 °C. The binding affinity and K_D were determined using MO.Affinity with the K_D fit model.

14.9 Determination of binding affinity through Surface Plasmon Resonance

To generate biotinylated NPAS1 and NPAS3 proteins, a set of oligos was designed to add Avi-tag to the C-terminus of both proteins within pETDuet-1-hNPAS1/3-ARNT2 constructs via general PCR procedure (Table 10). Both hNPAS1-Avi-ARNT2 and hNPAS3-Avi-ARNT2 constructs in the pETDuet-1 vector were then co-transformed with a pCDF-BirA

plasmid into BL21-Codon Plus (DE3)-RIL competent cells using the heat shock method and plated into LB agar plates with ampicillin, chloramphenicol, and streptomycin. Both proteins were expressed in the presence of 200 nM biotin (Fluorochem, MO2926) and 500 μM of IPTG and were purified as previously described in section 14.3.

The ARNT2-PAS-B domain (residues 332-438) was previously subcloned into the pET-28a vector with an N-terminal His-tag and a C-terminal Avi-tag by Dr. Hua Pu. The plasmid was then transformed into BL21 (DE3) competent cells that overexpressed BirA. Following bacterial transformation, the protein was expressed in the presence of 200 nM biotin along with IPTG, and was purified as previously described in section 14.3.

SPR was performed on a Biacore S200 (Cytiva, 29136649) with sensor chip SA (Cytiva, BR100531). SA chip channels were first washed with 1 M NaCl and 50 mM NaOH, and proteins were then immobilised directly on the SA chip through capture with pre-immobilised streptavidin. The immobilisation level was set to 8,000 RU. Assay cycles were set at 20 °C using a running buffer of 50 mM BTP pH 8.0, 150 mM NaCl, and 0.4% DMSO. A DMSO range from 0-0.8% was used for solvent correction. The analyte was injected at 30 μl min^{-1} for 60 seconds with a dissociation time of 120 seconds. The data were fitted with a steady-state affinity binding model.

Name	Primer's sequence
hNPAS1 (43-423) Avi_FP	tttgaagcgcagaagattgaatggcatgaatgaAAGCTTGCGGCCGCATAA
hNPAS1 (43-423) Avi_RP	gatatcattcaggccgtatccacctttactgctAGCTGGAAGCTGGAAGGC
hNPAS3 (56-455) Avi_FP	tttgaagcgcagaagattgaatggcatgaatgaAAGCTTGCGGCCGCATAAT
hNPAS3 (56-455) Avi_RP	gatatcattcaggccgtatccacctttactgctATGGGGGAGCTGTGCGAT

Table 10. Primers' design for constructs used in the SPR assay. Primers were designed to add Avi-tag to the C-terminus of both NPAS1 and NPAS3 proteins within pETDuet-1-hNPAS1/3-ARNT2 constructs.

14.10 TR-FRET assay

To generate constructs with fluorophore, eGFP was fused to C-terminus of mouse and human ARNT2 via a two-step PCR strategy (Tables 8, 9). Initially, DNA sequences encoding ARNT2 and eGFP were amplified separately by PCR using primers designed to introduce overlapping regions at the junction of the two genes (Table 11). The resulting PCR products were analysed on an agarose gel, and the correct bands were excised and purified using Gel DNA Recovery Kit (Zymo Research, D4008) following the manufacturer's protocol.

For the fusion reaction, purified PCR products of eGFP and ARNT2 were subsequently mixed at a 1:1 molar ratio and used as templates in a secondary PCR. In this reaction, the overlapping regions between ARNT2 and eGFP annealed, allowing DNA polymerase to extend and generate a full-length ARNT2-eGFP fusion construct. This fusion product was amplified using outer primers flanking the entire construct, resulting in a seamless, in-frame fusion.

Human ARNT2-eGFP was cloned into existing pETDuet-1-hNPAS1/3 constructs by Gibson assembly. The pETDuet-1-hNPAS1/3 prepared previously were digested with BglII and KpnI-HF restriction enzymes and mixed with the ARNT2-eGFP PCR product, and a Gibson assembly reaction was performed as previously described in section 14.1. Constructs were transformed into competent BL21-Codon Plus (DE3)-RIL cells, followed by the overexpression and purification of proteins as described in sections 14.2 and 14.3.

The mouse ARNT2-eGFP construct was cloned into the pMKH expression vector. The construct was co-transformed into competent BL21-Codon Plus (DE3)-RIL cells alongside several in-house C-terminal His-tagged triple-domain proteins from the bHLH-PAS family, including mHIF-2α, mNPAS1, mNPAS3, and mNPAS4. Proteins were overexpressed and purified as described in sections 14.2 and 14.3.

Name	Primer's sequence
hARNT2_eGFP_VB24_FP	AAGGAGATATACATATGGCAGATCTCGATTACAAGGATGACG
hARNT2_eGFP_VB25_RP	TGCTCACCATGACGTTGGTGTTGGTGCAG
eGFP_hARNT2_VB26_FP	CACCAACGTCATGGTGAGCAAGGGCGAG
eGFP_hARNT2_VB27_RP	TTTACCAGACTCGAGGGTACTTAAGCTGCCTTGTACAGCTC

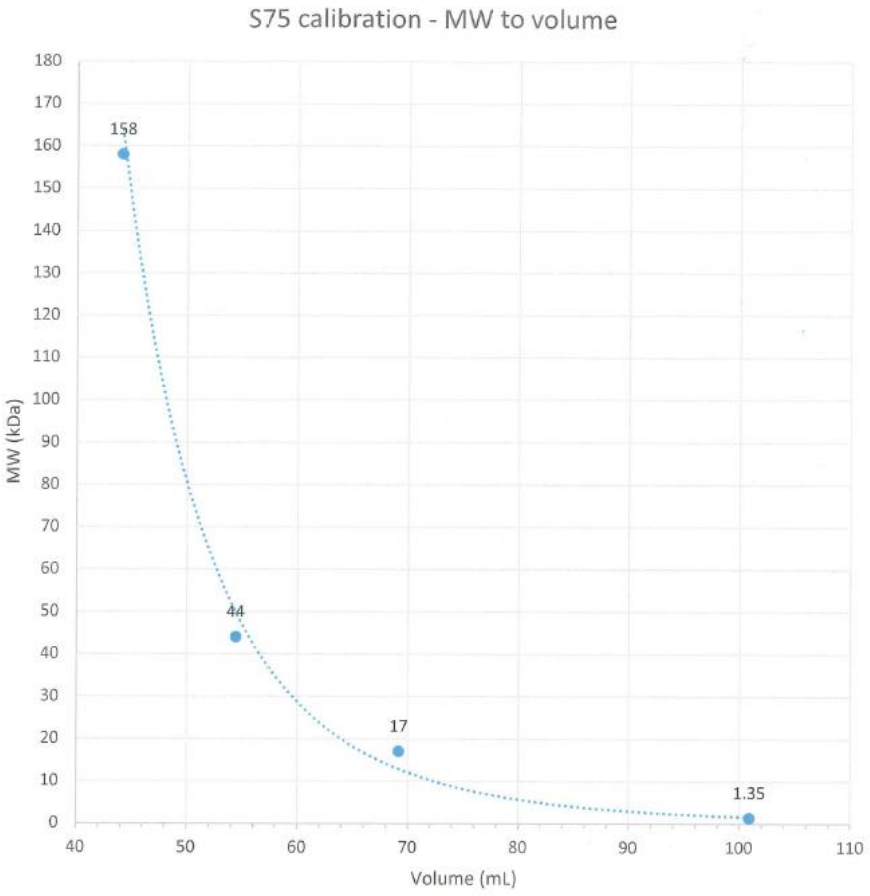
Name	Primer's sequence
mARNT2-eGFP_FP	gtatattagttaagtataagaaggagatatacatatggatggtgaaggccccagtaaattttcaagagagaatc
mARNT2-eGFP_RP	gttattgctcagcggtggcagcagcctaggttaattaagctgcCTTGTACAGCTCGTCCATG CCGAGAGTGATC
eGFP-mARNT2_FP	cagaactggaggtacatcagcgagatgggctgtcgatggtgagcaa
eGFP-mARNT2_RP	caccccggtgaacagctcctcgcccttgctcaccatcgaacagccca

Table 11. Primers for ARNT2-eGFP constructs for the TR-FRET assay.

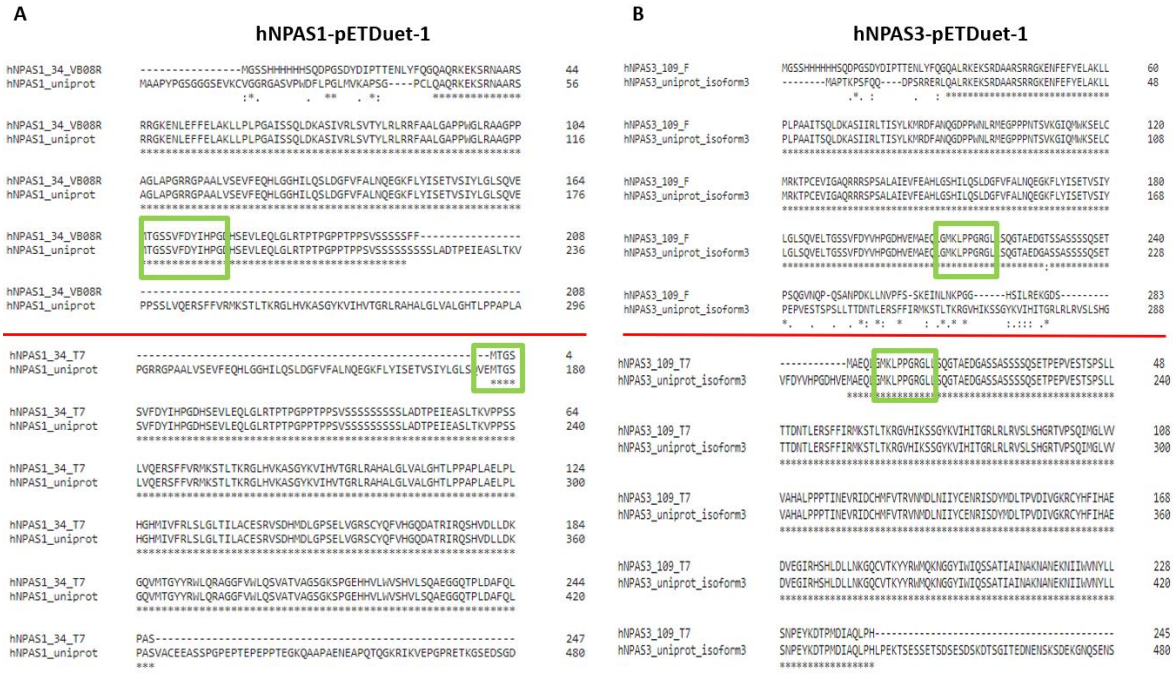
Purified protein complexes of ARNT2-eGFP with His-tagged HIF-2α, NPAS1, NPAS3, and NPAS4 were serially diluted in assay buffer containing 20 mM Tris pH 8.5, 100 mM NaCl, 1mM DTT, and 0.05% Tween-20. Following the addition of LanthaScreen Tb-anti-His-Tag Antibody (Invitrogen, PV5863) at a final concentration of 5 nM, protein interactions were monitored via the energy transfer signal with a PHERAstar FSX plate reader (BMG LABTECH). The TR-FRET values were determined as a ratio of the signal measured at 520 nm (eGFP) to the signal measured at 492 nm (terbium). The apparent EC_{50} values of each protein complex were calculated by plotting the ratios (520 nm/492 nm) against the protein concentrations in GraphPad Prism. For the subsequent TR-FRET assay with tested compounds, each protein was used at a concentration double its measured EC_{50} value, such as 1 nM, 4 nM, 25 nM, and 0.8 nM for ARNT2-eGFP complexes with HIF-2α, NPAS1, NPAS3, and NPAS4, respectively.

To generate assay-ready plates for DRC analysis with tested compounds, 100 nL of each compound concentration was dispensed from stock Echo plates into a black, low-volume 384-well microplate (Corning, 4514). Protein complexes diluted in assay buffer were mixed with LanthaScreen Tb-anti-His-Tag Antibody at a final concentration of 5 nM per well. The resulting protein-terbium mixture was dispensed into assay-ready plates at a volume of 10 µL per well. Plates were sealed and centrifuged at 1,000 rpm for 1 minute, then incubated on a gentle rotary shaker at room temperature in the dark for 1 hour to allow binding interactions to occur. The protein interactions were measured, and TR-FRET values were determined as a ratio of 520 nm to 492 nm. Apparent IC_{50} or EC_{50} values for each protein complex in the presence of a ligand were calculated by plotting the 520 nm/492 nm ratios against ligand concentrations using GraphPad Prism

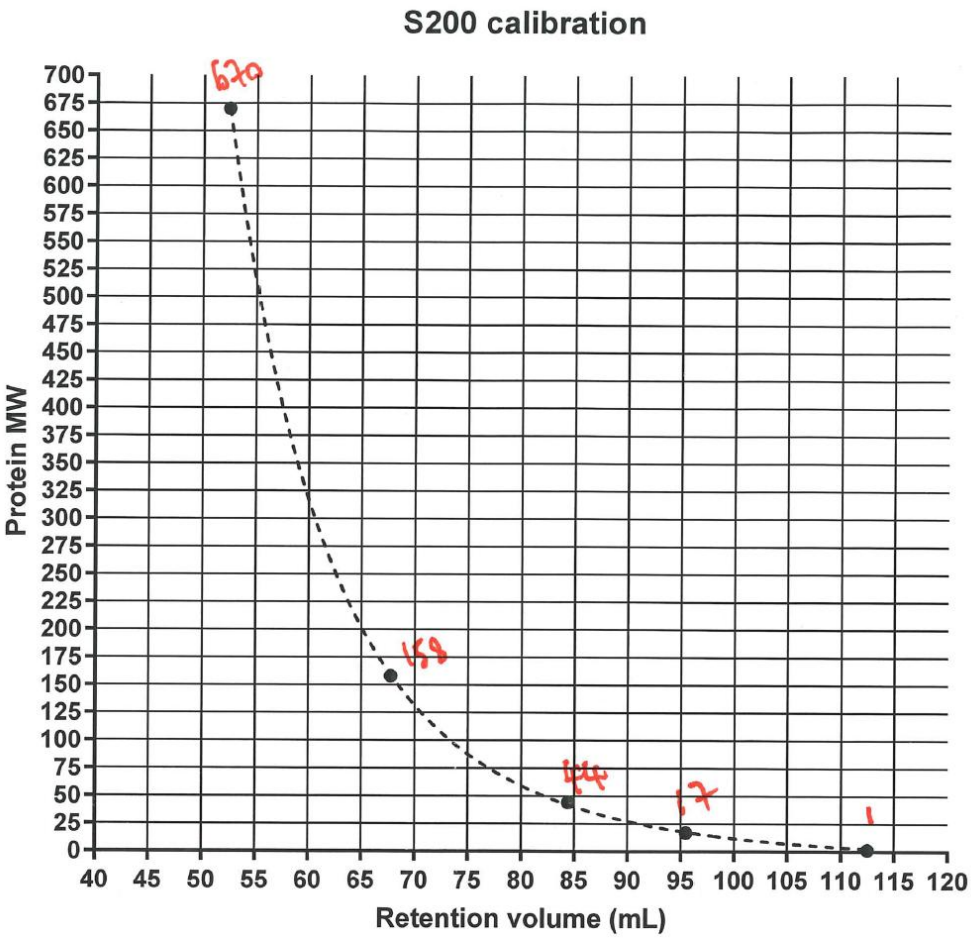
Appendices



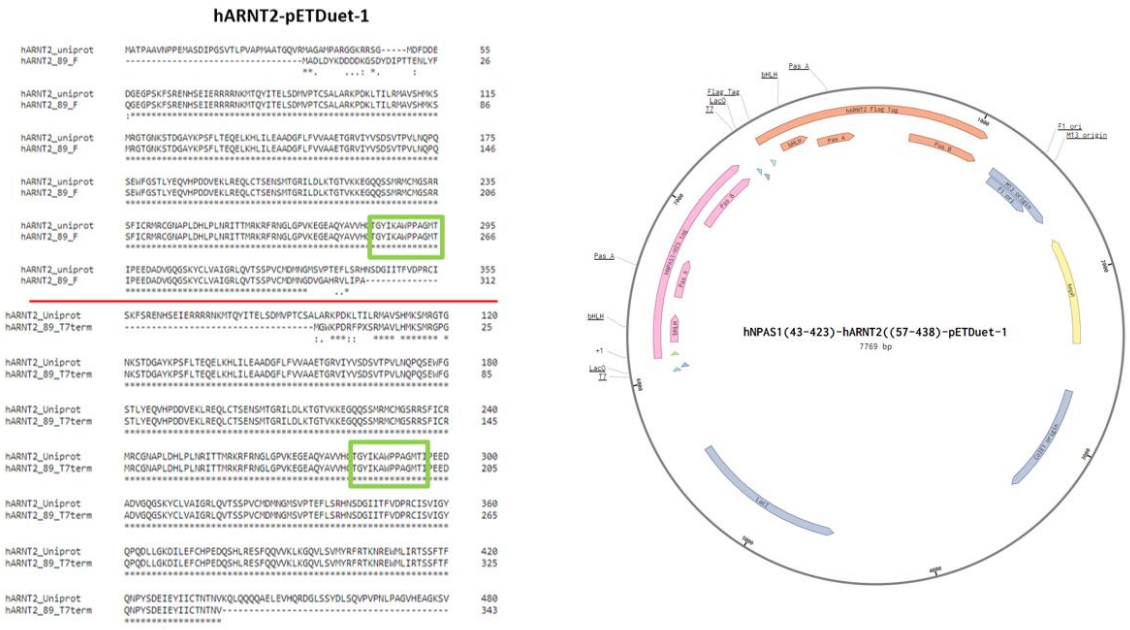
Appendix 1. Superdex 75-pg HiLoad 16/600 calibration curve (MW/Volume). The Gel Filtration Standard Kit (BioRad, 1511901) was previously used to calibrate the Superdex 75-pg HiLoad 16/600 size exclusion chromatography column. The kit contains lyophilised molecular weight standards, including thyroglobulin (670,000 Da), γ-globulin (158,000 Da), ovalbumin (44,000 Da), myoglobin (17,000 Da), and vitamin B12 (1,350 Da). These standards cover a broad molecular weight range, allowing for accurate calibration of the column and enabling reliable molecular weight estimation of unknown protein samples. Superdex 75 columns are designed, however, for smaller proteins, typically separating molecules between 70,000 Da and 3,000 Da. The higher molecular weight standard, like thyroglobulin, eluted in the void volume. The lower molecular weight proteins, including γ-globulin, ovalbumin, myoglobin, and vitamin B12, eluted well within the Superdex 75 range, and their corresponding molecular weights are marked on the graph.



Appendix 2. Sequence analyses of human NPAS1 and NPAS3. Both constructs were cloned into pETDuet-1. **(A)** Sequence confirmation of hNPAS1 (residues 43-423), and **(B)** hNPAS3 (residues 19-437) constructs in pETDuet-1. Green boxes correspond to regions where sequencing with forward and reverse primers was carried out.



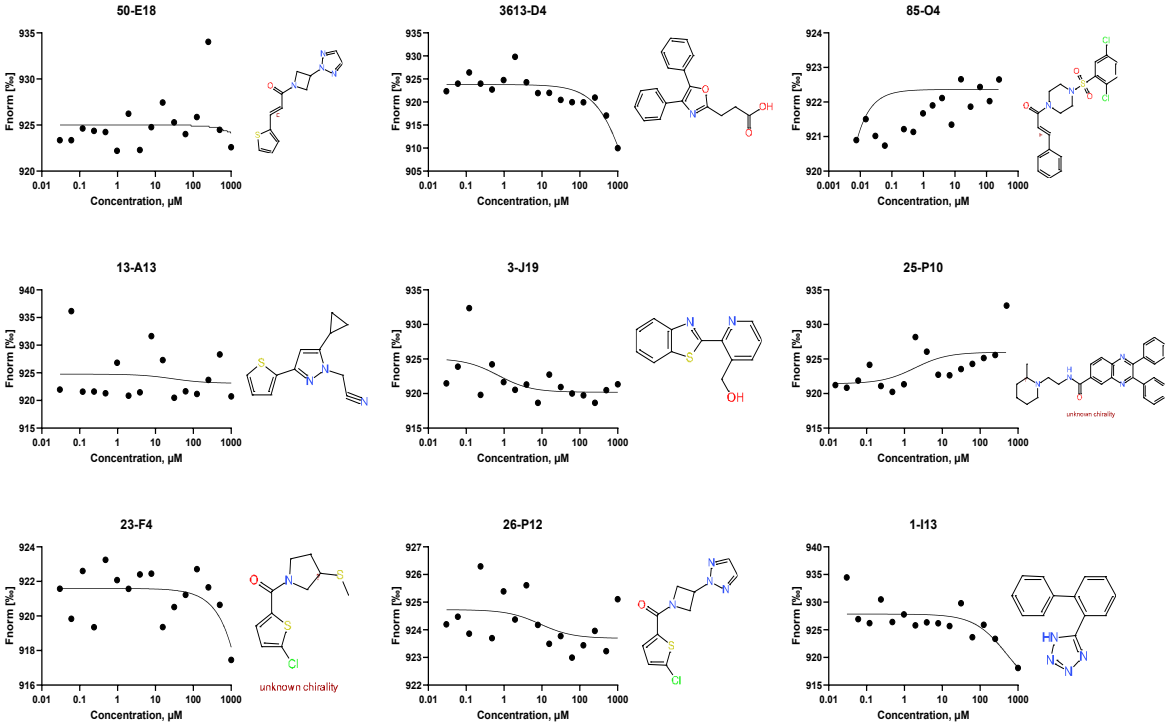
Appendix 3. Superdex 200-pg HiLoad 16/600 calibration curve (MW/Volume).The Gel Filtration Standard Kit (BioRad, 1511901) was used to calibrate the Superdex 200-pg HiLoad 16/600 size exclusion chromatography column. The kit contains lyophilised molecular weight standards, including thyroglobulin (670,000 Da), γ-globulin (158,000 Da), ovalbumin (44,000 Da), myoglobin (17,000 Da), and vitamin B12 (1,350 Da). These standards eluted well from the Superdex 200-pg HiLoad 16/600 column, and their corresponding masses were marked on the curve.



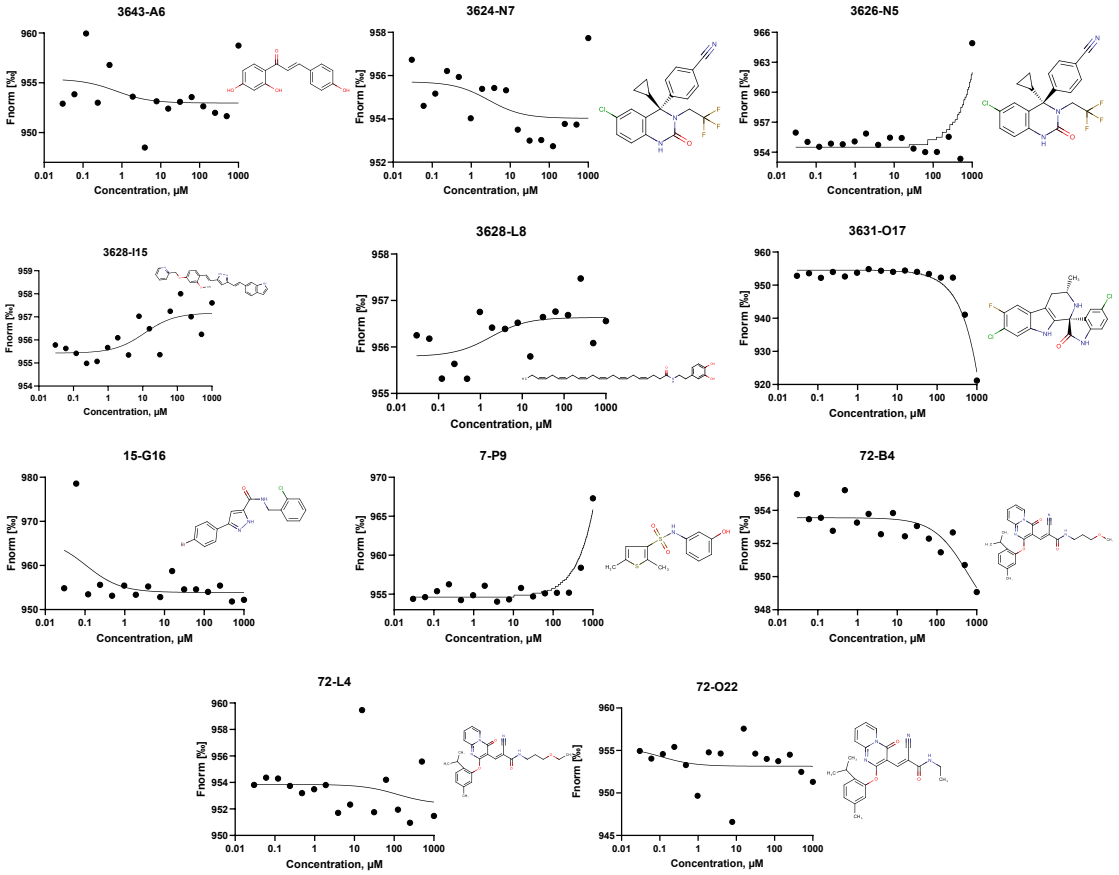
Appendix 4. Representative sequence analysis of hARNT2. Human ARNT2 (residues 57-438) was cloned into pETDuet-1. Green boxes indicate the overlapping region sequenced with forward and reverse sequencing primers. The hARNT2 insert was subsequently cloned into previously generated pETDuet-1-hNPAS1/3 to create a heterodimeric construct.

	3617-M16	50-E18	3613-D4
Data collection			
Space group	P62_2_2	P62_2_2	P62_2_2
Cell dimensions			
a, b, c (Å)	81.222, 81.222, 125.343	80.961, 80.961, 125.711	81.336, 81.336, 125.631
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range (Å)	70.340-1.666 (1.695-1.666)	70.114-1.821 (1.853-1.821)	70.439-1.705 (1.734-1.705)
R_{merge}	0.206 (6.356)	0.510 (34.048)	0.221 (11.387)
Multiplicity	38.4 (25.2)	39.6 (41.0)	39.7 (33.4)
Completeness (%)	99.8 (98.1)	100.0 (100.0)	98.8 (94.3)
Mean(I)/sd(I)	15.8 (0.7)	9.2 (0.5)	14.1 (0.4)
Half-set correlation CC1/2	0.999 (0.488)	0.997 (0.424)	0.999 (0.350)
Refinement			
Resolution (Å)	70.44-1.67	70.21-1.82	
No. reflections all/free	29142/1438	21922/1159	
R_{work}/R_{free}	0.207/0.220	0.225/0.257	
No. atoms	1113	1088	
macromolecules	982	989	
ligand/ion	27/4	18/4	
solvent	100	77	
R. m. s. deviations			
RMS (bonds) (Å)	0.0118	0.0086	
RMS (angels) (Å)	1.809	1.695	
Average B-factor (Å^2)			
macromolecules	25.32	35.7	
ligand	32.51	47.7	
solvent	38.84	45.8	

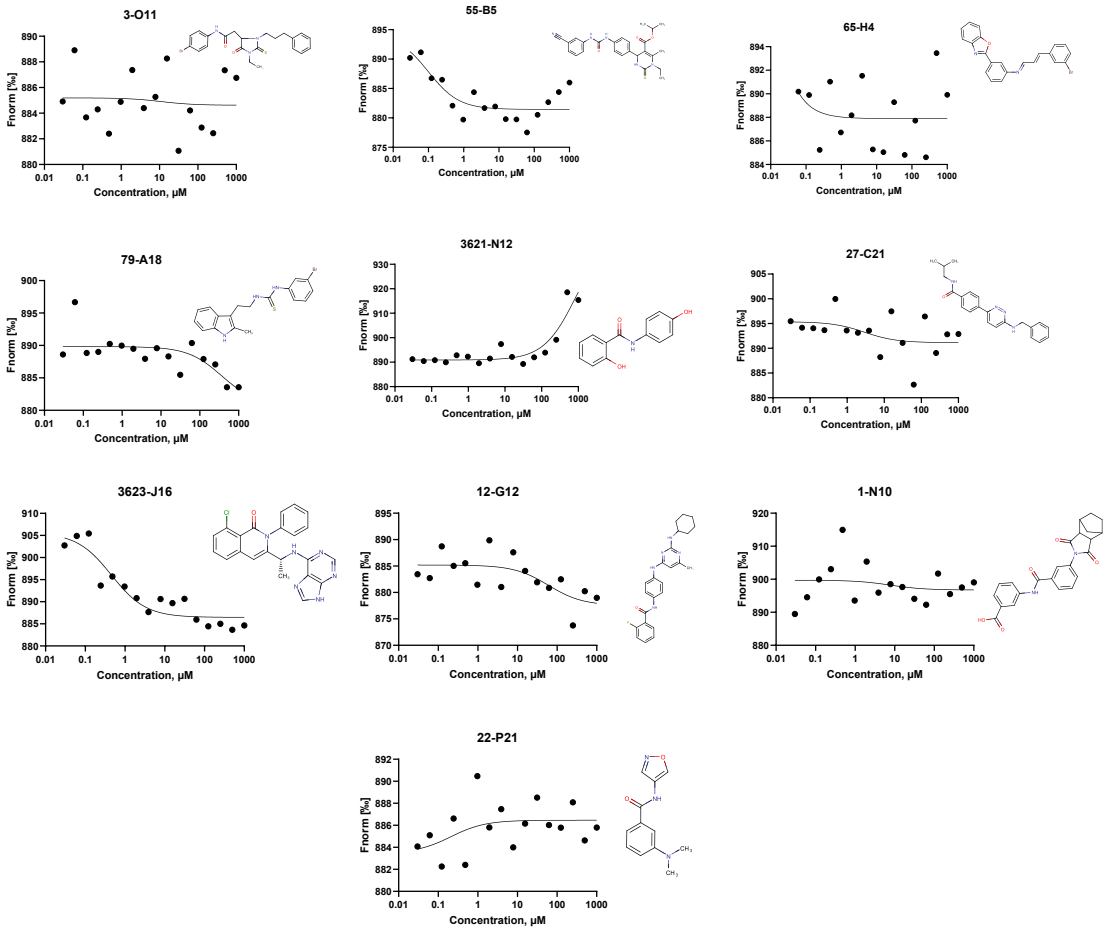
Appendix 5. Statistics of crystal diffraction and structure refinement for ARNT2-PAS-B ligands bound to mARNT2-PAS-B.



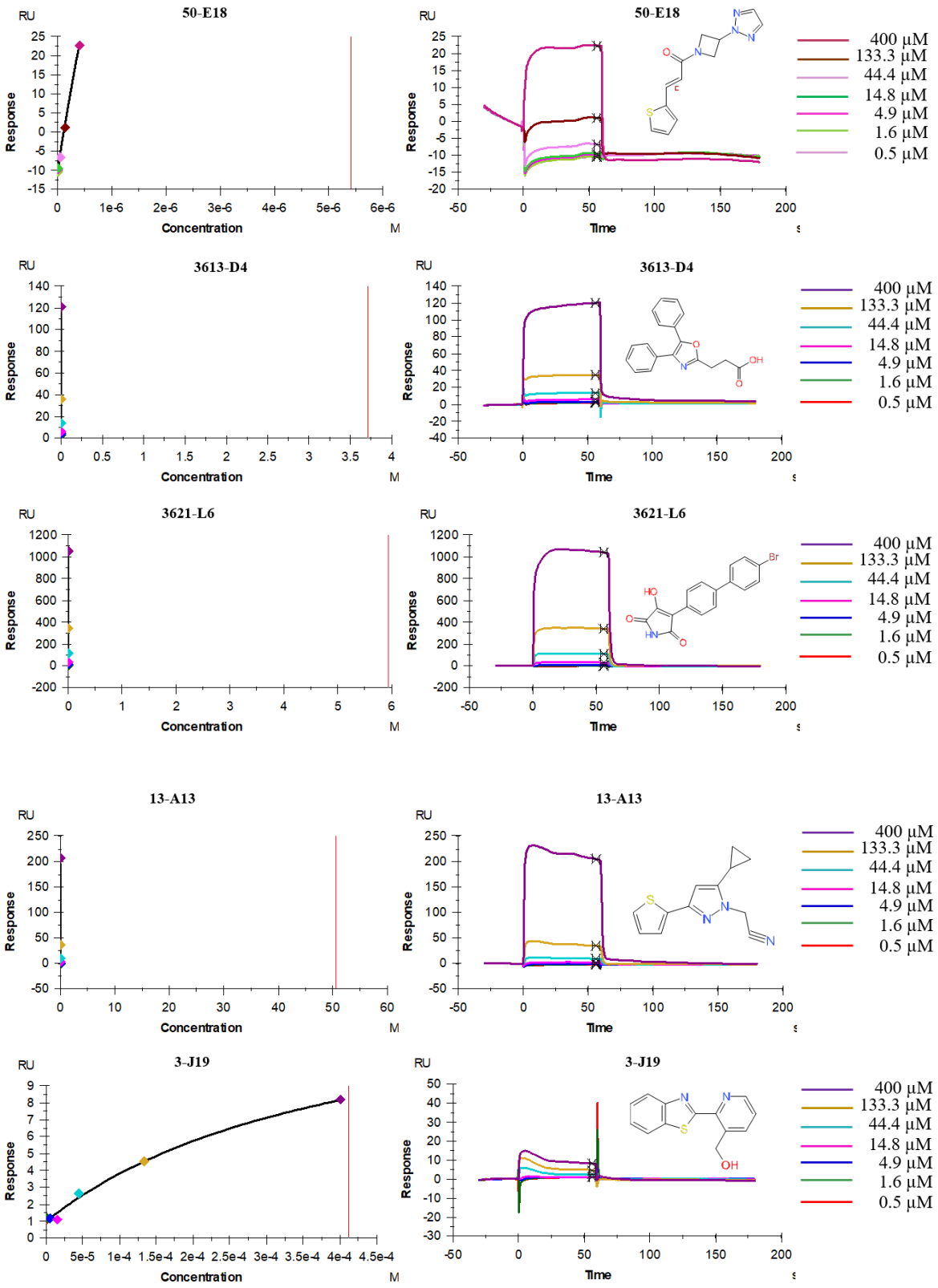
Appendix 6. MST dose-response curve for the rest of ARNT2-PAS-B tested compounds. 250 nM of the target protein was labelled with the second generation of RED-Tris-NTA dye.

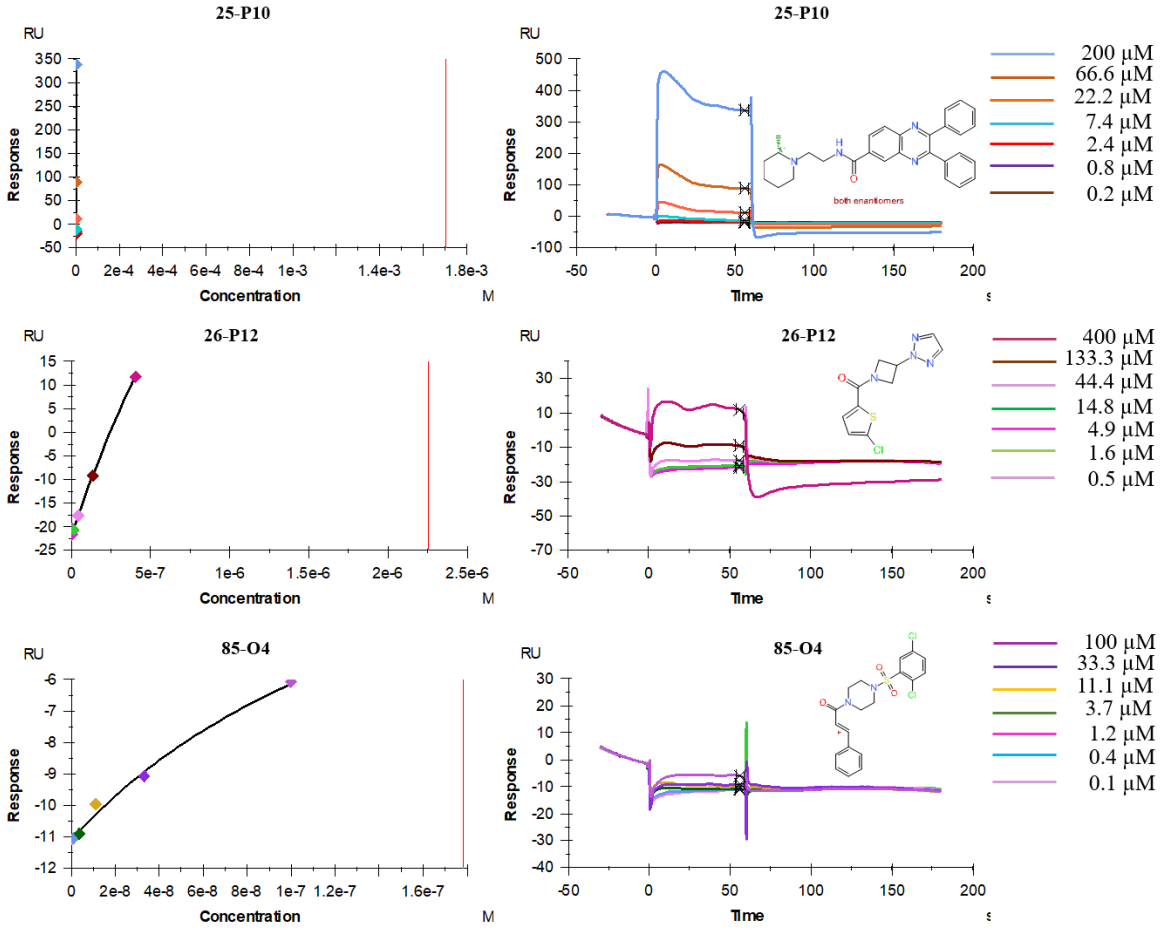


Appendix 7. MST dose-response curve for the hNPAS1-ARNT2 tested compounds. 125 nM of the target protein was labelled with the second generation of RED-Tris-NTA dye.

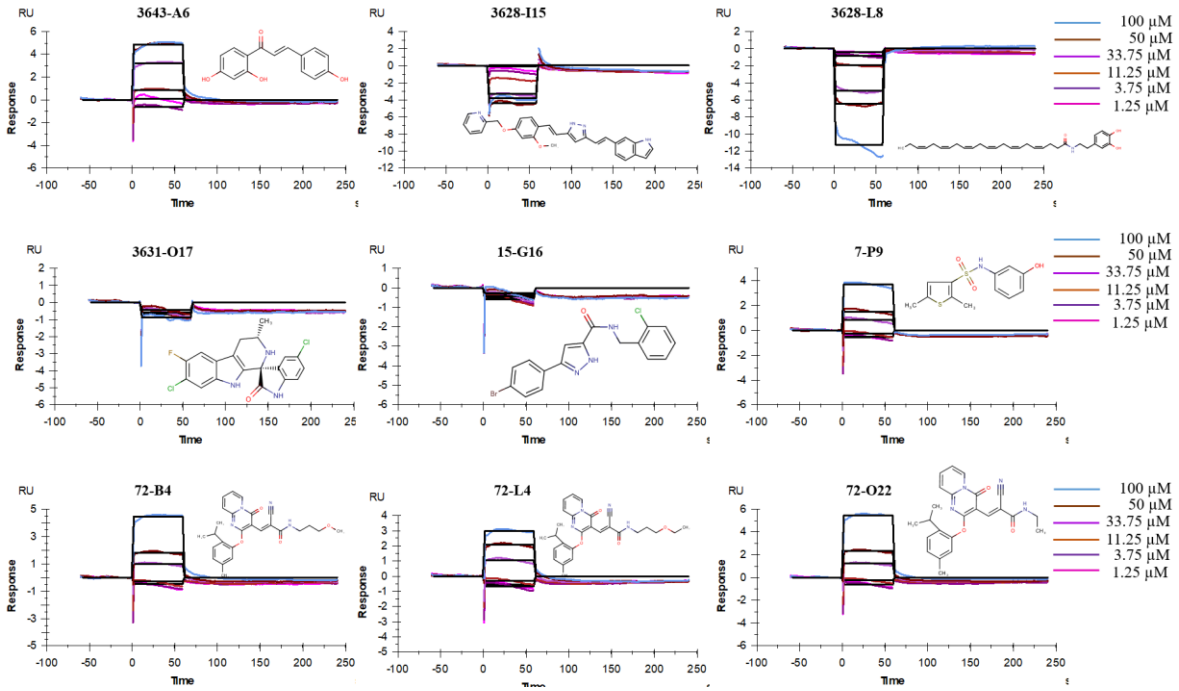


Appendix 8. MST dose-response curve for the hNPAS3-ARNT2 tested compounds. 250 nM of the target protein was labelled with the second generation of RED-Tris-NTA dye.

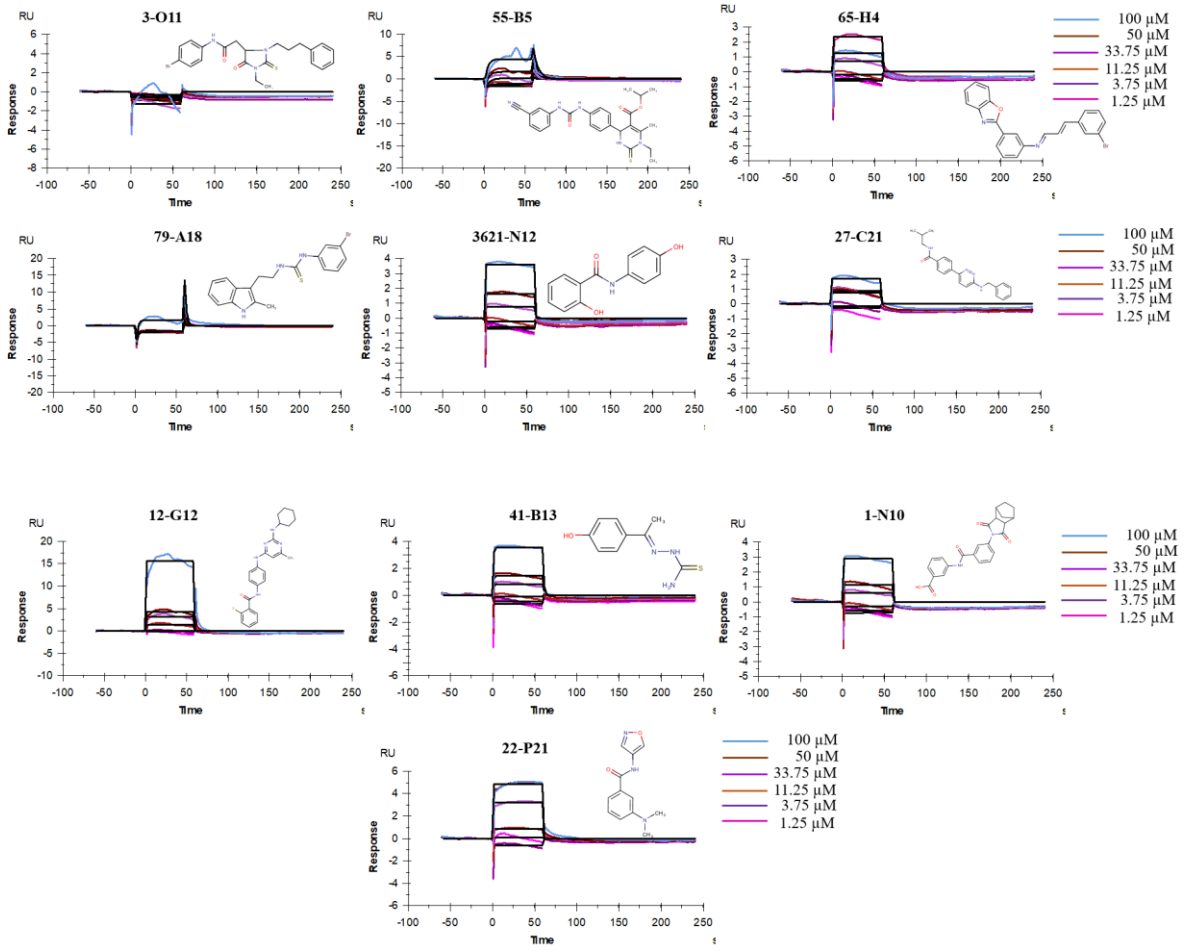

50-E18
RU
Response
Concentration
M
Time
s
400 µM
133.3 µM
44.4 µM
14.8 µM
4.9 µM
1.6 µM
0.5 µM
3613-D4
3621-L6
13-A13
3-J19



Appendix 9. SPR analyses for ARNT2 tested compounds. Dose-response steady state affinity for the biotinylated single PAS-B domain of ARNT2 and the remaining ARNT2 tested compounds that did not reach the saturation point.

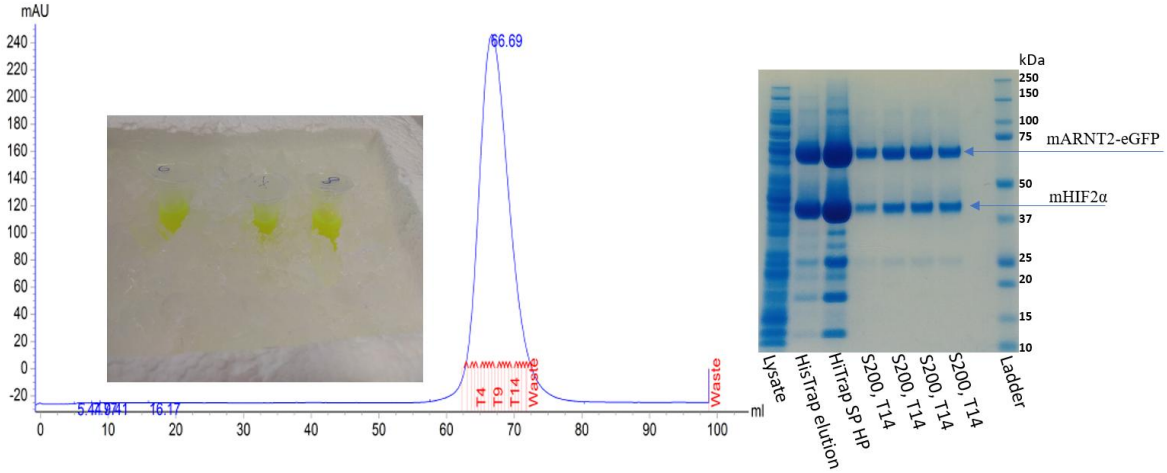


Appendix 10. SPR analyses for NPAS1 tested compounds. Dose-response SPR kinetic model for NPAS1 tested compounds and biotinylated hNPAS1-ARNT2 protein.



Appendix 11. SPR analyses for NPAS3 tested compounds. Dose-response SPR kinetic model for NPAS3 tested compounds and biotinylated hNPAS3-ARNT2 protein.

Appendix 12. Representative ARNT2-eGFP construct sequence analysis. Final PCR product consisting of a seamless hARNT2-eGFP construct. The green square represents the overlapping ARNT2 region between two sequence primers. The blue square represents the beginning of the eGFP sequence.



Appendix 13. Purification of mHIF2α-ARNT2-eGFP. Representative example of mHIF2α-ARNT2-eGFP purification for the TR-FRET assay. Quality control assessment of mHIF2α-ARNT2-eGFP is based on gel-filtration, bright green-yellow colour samples, and SDS-PAGE profiles.

	Target-4	LS02-121-01	LS002-024
Data collection			
Space group	P62_2_2	P62_2_2	P62_2_2
Cell dimensions			
a, b, c (Å)	81.149, 81.149, 125.086	80.827, 80.827, 124.937	80.83, 80.83, 124.82
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range (Å)	70.277-1.536 (1.562-1.536)	69.998-1.671 (1.700-1.671)	70.001-1.836 (1.868 -1.836)
R_{merge}	0.131 (6.048)	0.261 (14.696)	0.580 (18.063)
Multiplicity	33.8 (12.3)	38.6 (26.5)	39.5 (41.1)
Completeness (%)	99.6 (96.7)	99.7 (97.4)	100.0 (100.0)
Mean (I)/sd(I)	15.9 (0.4)	13.5 (0.4)	9.2 (0.7)
Half-set correlation CC1/2	0.999(0.333)	0.999 (0.341)	0.978 (0.369)
Refinement			
Resolution (Å)	70.38-1.54	70.10-1.67	70.10-1.84
No. reflections all/free	36228/1740	28206/1390	21636/1122
R_{work}/R_{free}	0.210/0.230	0.218/0.238	0.227/0.256
No. atoms	1135	1125	1103
macromolecules	990	1013	1013
ligand/ion	18/4	22/4	18/4
solvent	123	86	68
R. m. s. deviations			
RMS (bonds) (Å)	0.0122	0.0107	0.0089
RMS (angels) (Å)	1.796	1.648	1.483
Average B-factor (Å²)			
macromolecules	24.9	28.93	31.97
ligand	29.7	36.59	51.31
solvent	40.3	43.51	43.42

Appendix 14. Statistics of crystal diffraction and structure refinement for KG-548 analogues bound to mARNT2-PAS-B.

	LS02-170-01	LS02-174-01	Butein
Data collection			
Space group	P62_2_2	P62_2_2	P62_2_2
Cell dimensions			
a, b, c (Å)	81.182, 81.182, 125.044	81.017, 81.017, 124.970	81.333, 81.333, 124.795
A, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range (Å)	70.306 -1.510 (1.536-1.510)	70.163-1.589 (1.617-1.589)	70.436-3.016 (3.068-3.016)
R_{merge}	0.124 (4.688)	0.138 (5.053)	1.669 (8.720)
Multiplicity	32.7 (10.3)	36.2 (18.3)	38.0 (34.2)
Completeness (%)	99.7 (97.4)	99.1 (92.9)	100.0 (100.0)
Mean (I)/sd(I)	18.2 (0.5)	18.3 (0.8)	5.2 (1.4)
Half-set correlation CC1/2	0.999 (0.450)	0.999 (0.408)	0.916 (0.429)
Refinement			
Resolution (Å)	70.41-1.51	70.26-1.59	70.54-3.02
No. reflections all/free	38549/1929	32995/1642	5230/272
R_{work}/R_{free}	0.208/0.233	0.217/0.247	0.221/0.290
No. atoms	1167	1150	932
macromolecules	1005	1013	910
ligand/ion	17/4	17/4	20/2
solvent	135	116	-
R. m. s. deviations			
RMS (bonds) (Å)	0.0119	0.0121	0.0065
RMS (angels) (Å)	1.757	1.804	1.799
Average B-factor (Å²)			
macromolecules	21.75	23.86	27.91
ligand	27.94	27.17	54.29
solvent	38.04	39.95	-

Appendix 15. Statistics of crystal diffraction and structure refinement for KG-548 analogues and butein bound to mARNT2-PAS-B.

	FR-15	FR-20	FR-26
Data collection			
Space group	P62_2_2	P62_2_2	P62_2_2
Cell dimensions			
a, b, c (Å)	80.89, 80.89, 125.07	80.915, 80.915, 125.422	81.290, 81.290, 124.816
A, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range (Å)	70.053 -1.611 (1.639 -1.611)	70.074 -1.520 (1.547-1.520)	70.399 -1.512 (1.538-1.512)
R_{merge}	0.132 (6.561)	0.124 (5.907)	0.112 (3.583)
Multiplicity	36.8 (20.0)	33.9 (13.4)	32.6 (10.1)
Completeness (%)	99.4 (94.5)	97.3 (79.9)	100.0 (99.4)
Mean (I)/sd(I)	18.8 (0.6)	18.4 (0.4)	18.2 (0.6)
Half-set correlation CC1/2	0.999 (0.437)	0.999 (0.382)	0.999 (0.428)
Refinement			
Resolution (Å)	70.15-1.61	70.17-1.52	70.50-1.51
No. reflections all/free	31724/1619	35814/1770	38319/1910
R_{work}/R_{free}	0.217/0.252	0.211/0.231	0.204/0.223
No. atoms	1096	1123	1121
macromolecules	974	982	982
ligand/ion	19/4	22/4	17/4
solvent	99	115	118
R. m. s. deviations			
RMS (bonds) (Å)	0.0117	0.0133	0.0132
RMS (angels) (Å)	1.815	1.887	1.872
Average B-factor (Å^2)			
macromolecules	27.25	25.12	25.44
ligand	43.16	35.2	32.89
solvent	41.84	39.58	39.79

Appendix 16. Statistics of crystal diffraction and structure refinement for butein analogues bound to mARNT2-PAS-B.

References

1. Wu, D., Su, X., Potluri, N., Kim, Y. & Rastinejad, F. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. *Elife* 5, (2016).

2. Michaelson, J. J. *et al.* Neuronal PAS Domain Proteins 1 and 3 Are Master Regulators of Neuropsychiatric Risk Genes. *Biol Psychiatry* 82, 213–223 (2017).

3. Fribourgh, J. L. & Partch, C. L. Assembly and function of bHLH-PAS complexes. *Proc Natl Acad Sci U S A* 114, 5330–5332 (2017).

4. Erbel-Sieler, C. *et al.* Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc Natl Acad Sci U S A* 101, 13648–13653 (2004).

5. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* 172, 650–665 (2018).

6. Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors, RXR, and the big bang. *Cell* 157, 255–266 (2014).

7. Huang, P., Chandra, V. & Rastinejad, F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu Rev Physiol* 72, 247–272 (2010).

8. Raghuram, S. *et al.* Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nat Struct Mol Biol* 14, 1207–1213 (2007).

9. Rastinejad, F., Huang, P., Chandra, V. & Khorasanizadeh, S. Understanding nuclear receptor form and function using structural biology. *J Mol Endocrinol* 51, T1–T21 (2013).

10. Smith, S. H. *et al.* Tapinarof Is a Natural AhR Agonist that Resolves Skin Inflammation in Mice and Humans. *J Invest Dermatol* 137, 2110–2119 (2017).

11. Courtney, K. D. *et al.* HIF-2 Complex Dissociation, Target Inhibition, and Acquired Resistance with PT2385, a First-in-Class HIF-2 Inhibitor, in Patients with Clear Cell Renal Cell Carcinoma. *Clin Cancer Res* 26, 793–803 (2020).

12. Deeks, E. D. Belzutifan: First Approval. *Drugs* 81, 1921–1927 (2021).

13. Zhou, J. & Gong, K. Belzutifan: a novel therapy for von Hippel-Lindau disease. *Nat Rev Nephrol* 18, 205–206 (2022).

14. Kolonko, M. & Greb-Markiewicz, B. bHLH-PAS Proteins: Their Structure and Intrinsic Disorder. *Int J Mol Sci* 20, (2019).

15. Edwards, H. E. & Gorelick, D. A. The evolution and structure/function of bHLH–PAS transcription factor family. *Biochem Soc Trans* 50, 1227–1243 (2022).

16. Thiriet-Rupert, S. *et al.* Transcription factors in microalgae: Genome-wide prediction and comparative analysis. *BMC Genomics* 17, (2016).

17. Annunziata, R. *et al.* BHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom Phaeodactylum tricornutum. *Proc Natl Acad Sci U S A* 116, 13137–13142 (2019).

18. Tumova, S., Dolezel, D. & Jindra, M. Conserved and Unique Roles of bHLH-PAS Transcription Factors in Insects – From Clock to Hormone Reception. *J Mol Biol* 436, 168332 (2024).

19. Rastinejad, F. Retinoic acid receptor structures: the journey from single domains to full-length complex. *J Mol Endocrinol* 69, T25–T36 (2022).

20. Rastinejad, F. Retinoid X receptor and its partners in the nuclear receptor family. *Curr Opin Struct Biol* 11, 33–38 (2001).

21. Rastinejad, F., Perlmann, T., Evans, R. M. & Sigler, P. B. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375, 203–211 (1995).

22. Huang, P., Chandra, V. & Rastinejad, F. Retinoic Acid Actions Through Mammalian Nuclear Receptors. *Chem Rev* 114, 233 (2013).

23. Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M. A. & Rastinejad, F. Structural elements of an orphan nuclear receptor-DNA complex. *Mol Cell* 1, 849–861 (1998).

24. Ghorbel, M. T., Coulson, J. M. & Murphy, D. Cross-talk between hypoxic and circadian pathways: Cooperative roles for hypoxia-inducible factor 1α and CLOCK in transcriptional activation of the vasopressin gene. *Mol Cell Neurosci* 22, 396–404 (2003).

25. Hogenesch, J. B., Gu, Y. Z., Jain, S. & Bradfield, C. A. The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc Natl Acad Sci U S A* 95, 5474–5479 (1998).

26. Takahata, S. *et al.* Transcriptionally active heterodimer formation of an Arnt-like PAS protein, Arnt3, with HIF-1a, HLF, and clock. *Biochem Biophys Res Commun* 248, 789–794 (1998).

27. Sogawa, K. *et al.* Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation. *Proc Natl Acad Sci U S A* 92, 1936 (1995).

28. Sakurai, S., Shimizu, T. & Ohto, U. The crystal structure of the AhRR–ARNT heterodimer reveals the structural basis of the repression of AhR-mediated transcription. *J Biol Chem* 292, 17609 (2017).

29. Evans, B. R. *et al.* Repression of Aryl Hydrocarbon Receptor (AHR) signaling by AHR repressor: Role of DNA binding and competition for AHR nuclear translocator. *Mol Pharmacol* 73, 387–398 (2008).

30. Oñate, S. A., Tsai, S. Y., Tsai, M.-J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* 270, 1354–1357 (1995).

31. Bersten, D. C., Sullivan, A. E., Peet, D. J. & Whitelaw, M. L. BHLH-PAS proteins in cancer. *Nat Rev Cancer* 13, 827–841 (2013).

32. Litterst, C. M. & Pfitzner, E. Transcriptional Activation by STAT6 Requires the Direct Interaction with NCoA-1. *J Biol Chem* 276, 45713–45721 (2001).

33. Zhou, Y. D. *et al.* Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. *Proc Natl Acad Sci U S A* 94, 713–718 (1997).

34. Brunskill, E. W., Witte, D. P., Shreiner, A. B. & Potter, S. S. Characterization of Npas3, a novel basic helix-loop-helix PAS gene expressed in the developing mouse nervous system. *Mech Dev* 88, 237–241 (1999).

35. Stanco, A. *et al.* NPAS1 Represses the Generation of Specific Subtypes of Cortical Interneurons. *Neuron* 84, 940–953 (2014).

36. Pieper, A. A. *et al.* The neuronal PAS domain protein 3 transcription factor controls FGF-mediated adult hippocampal neurogenesis in mice. *Proc Natl Acad Sci U S A* 102, 14052–14057 (2005).

37. Sonnenfeld, M. *et al.* The Drosophila tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Development* 124, 4571–4582 (1997).

38. Ohshiro, T. & Saigo, K. Transcriptional regulation of breathless FGF receptor gene by binding of TRACHEALESS/dARNT heterodimers to three central midline elements in Drosophila developing trachea. *Development* 124, 3975–3986 (1997).

39. Levesque, B. M. *et al.* NPAS1 regulates branching morphogenesis in embryonic lung. *Am J Respir Cell Mol Biol* 36, 427–434 (2007).

40. Sunday, M. E., Zhou, S., Degan, S., Potts, E. & Foster, W. M. NPAS3 promotes lung development by up-regulating FGF signaling. *The FASEB Journal* 23, 572.1-572.1 (2009).

41. Zhou, S., Degan, S., Potts, E. N., Foster, W. M. & Sunday, M. E. NPAS3 is a trachealess homolog critical for lung development and homeostasis. *Proc Natl Acad Sci U S A* 106, 11691–11696 (2009).

42. Ohsawa, S., Hamada, S., Kakinuma, Y., Yagi, T. & Miura, M. Novel function of neuronal PAS domain protein 1 in erythropoietin expression in neuronal cells. *J Neurosci Res* 79, 451–8 (2005).

43. Teh, C. H. L. *et al.* Neuronal PAS domain protein 1 is a transcriptional repressor and requires arylhydrocarbon nuclear translocator for its nuclear localization. *J Biol Chem.* 281, 34617–34629 (2006).

44. Ward, M. P., Mosher, J. T. & Crews, S. T. Regulation of bHLH-PAS protein subcellular localization during Drosophila embryogenesis. *Development* 125, 1599–1608 (1998).

45. Luoma, L. M. & Berry, F. B. Molecular analysis of NPAS3 functional domains and variants. *BMC Mol Biol* 19, 14 (2018).

46. Macintyre, G. *et al.* Association of NPAS3 exonic variation with schizophrenia. *Schizophr Res* 120, 143–149 (2010).

47. Nguyen Ba, A. N., Pogoutse, A., Provart, N. & Moses, A. M. NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* 10, (2009).

48. Samardžija, B. *et al.* Protein Aggregation of NPAS3, Implicated in Mental Illness, Is Not Limited to the V304I Mutation. *J Pers Med* 11, 1070 (2021).

49. Kreft, L. *et al.* ConTra v3: a tool to identify transcription factor binding sites across species, update 2017. *Nucleic Acids Res* 45, W490 (2017).

50. Kewley, R. J., Whitelaw, M. L. & Chapman-Smith, A. The mammalian basic helix–loop–helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol* 36, 189–204 (2004).

51. Sun, X. *et al.* Structures of NPAS4-ARNT and NPAS4-ARNT2 heterodimers reveal new dimerization modalities in the bHLH-PAS transcription factor family. *Proc Natl Acad Sci U S A* 119, e2208804119 (2022).

52. Doering, J. A. *et al.* Aryl hydrocarbon receptor nuclear translocators (ARNT1, ARNT2, and ARNT3) of white sturgeon (Acipenser transmontanus): Sequences, tissue-specific expressions, and response to β-naphthoflavone. *Comp Biochem Physiol C Toxicol Pharmacol* 231, 108726 (2020).

53. Drutel, G. *et al.* Two splice variants of the hypoxia-inducible factor HIF-1α as potential dimerization partners of ARNT2 in neurons. *Eur J Neurosci.* 12, 3701–3708 (2000).

54. Rossi, J. J. *et al.* Molecular characterisation of rare loss-of-function NPAS3 and NPAS4 variants identified in individuals with neurodevelopmental disorders. *Sci Rep* 11, 6602 (2021).

55. Webb, E. A. *et al.* ARNT2 mutation causes hypopituitarism, post-natal microcephaly, visual and renal anomalies. *Brain* 136, 3096–3105 (2013).

56. Aitola, M. H. & Pelto-Huikko, M. T. Expression of Arnt and Arnt2 mRNA in developing murine tissues. *J Histochem Cytochem.* 51, 41–54 (2003).

57. Jain, S., Maltepe, E., Lu, M. M., Simon, C. & Bradfield, C. A. Expression of ARNT, ARNT2, HIF1α, HIF2α and Ah receptor mRNAs in the developing mouse. *Mech Dev* 73, 117–123 (1998).

58. Michaud, J. L., Rosenquist, T., May, N. R. & Fan, C.-M. Development of neuroendocrine lineages requires the bHLH–PAS transcription factor SIM1. *Genes Dev* 12, 3264–3275 (1998).

59. Hirose, K. *et al.* cDNA Cloning and Tissue-Specific Expression of a Novel Basic Helix-Loop-Helix/PAS Factor (Arnt2) with Close Sequence Similarity to the Aryl Hydrocarbon Receptor Nuclear Translocator (Arnt). *Mol Cell Biol* 16, 1706–1713 (1996).

60. Drutel, G., Kathmann, M., Heron, A., Schwartz, J. C. & Arrang, J. M. Cloning and selective expression in brain and kidney of ARNT2 homologous to the Ah receptor nuclear translocator (ARNT). *Biochem Biophys Res Commun* 225, 333–339 (1996).

61. Sharma, N. *et al.* ARNT2 tunes activity-dependent gene expression through NCoR2-mediated repression and NPAS4-mediated activation. *Neuron* 102, 390–406 (2019).

62. Wu, D. & Rastinejad, F. Structural characterization of mammalian bHLH-PAS transcription factors. *Curr Opin Struct Biol* 43, 1–9 (2016).

63. Crews, S. T. & Fan, C. M. Remembrance of things PAS: regulation of development by bHLH–PAS proteins. *Curr Opin Genet Dev* 9, 580–587 (1999).

64. Xing, J., Gumerov, V. M. & Zhulin, I. B. Origin and functional diversification of PAS domain, a ubiquitous intracellular sensor. *Sci Adv* 9, (2023).

65. Mitchell, P. J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science (1979)* 245, 371–378 (1989).

66. Li, X. *et al.* Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol* 141, 1167–1184 (2006).

67. de Martin, X., Sodaei, R. & Santpere, G. Mechanisms of Binding Specificity among bHLH Transcription Factors. *Int J Mol Sci* 22, 9150 (2021).

68. Crosthwaite, S. K., Dunlap, J. C. & Loros, J. J. Neurospora wc-1 and wc-2: Transcription, photoresponses, and the origins of circadian rhythmicity. *Science (1979)* 276, 763–769 (1997).

69. King, D. P. *et al.* Positional cloning of the mouse circadian Clock gene. *Cell* 89, 641–653 (1997).

70. Christie, J. M., Swartz, T. E., Bogomolni, R. A. & Briggs, W. R. Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function. *Plant Journal* 32, 205–219 (2002).

71. Tuckerman, J. R., Gonzalez, G. & Gilles-Gonzalez, M. A. Complexation precedes phosphorylation for two-component regulatory system FixL/FixJ of Sinorhizobium meliloti. *J Mol Biol* 308, 449–455 (2001).

72. Crews, S. T. Control of cell lineage-specific development and transcription by bHLH-PAS proteins. www.genesdev.org (1998).

73. Ponting, C. P. & Aravind, L. PAS: a multifunctional domain family comes to light. *Curr Biol* 7, (1997).

74. Möglich, A., Ayers, R. A. & Moffat, K. Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* 17, 1282–1294 (2009).

75. Pongratz, I., Antonsson, C., Whitelaw, M. L. & Poellinger, L. Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. *Mol Cell Biol* 18, 4079–4088 (1998).

76. Katschinski, D. M. *et al.* Interaction of the PAS B domain with HSP90 accelerates hypoxia-inducible factor-1α stabilization. *Cell Physiol Biochem.* 14, 351–360 (2004).

77. Hubbi, M. E., Luo, W., Baek, J. H. & Semenza, G. L. MCM Proteins Are Negative Regulators of Hypoxia-Inducible Factor 1. *Mol Cell* 42, 700–712 (2011).

78. Ban, H. S., Shimizu, K., Minegishi, H. & Nakamura, H. Identification of heat shock protein 60 as the regulator of the hypoxia-inducible factor subunit hif-1α. *Pure Appl. Chem.* 84, 2325–2337 (2012).

79. Xing, Y. *et al.* Identification of the Ah-receptor structural determinants for ligand preferences. *Toxicol Sci* 129, 86–97 (2012).

80. Diao, X. *et al.* Identification of oleoylethanolamide as an endogenous ligand for HIF-3α. *Nat Commun* 13, (2022).

81. Key, J., Scheuermann, T. H., Anderson, P. C., Daggett, V. & Gardner, K. H. Principles of ligand binding within a completely buried cavity in HIF2α PAS-B. *J Am Chem Soc* 131, 17647–17654 (2009).

82. Scheuermann, T. H. *et al.* Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nat Chem Biol* 9, 271–276 (2013).

83. Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* 524, 303–308 (2015).

84. Pu, H. *et al.* Pharmacological targeting of BMAL1 modulates circadian and immune pathways. *Nat Chem Biol* 21, 736–745 (2025).

85. Darimont, B. D. *et al.* Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev* 12, 3343–3356 (1998).

86. Khorasanizadeh, S. & Rastinejad, F. Visualizing the Architectures and Interactions of Nuclear Receptors. *Endocrinology* 157, 4212–4221 (2016).

87. Sun, Z. & Xu, Y. Nuclear Receptor Coactivators (NCOAs) and Corepressors (NCORs) in the Brain. *Endocrinology* 161, bqaa083 (2020).

88. You, S. H. *et al.* Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nat Struct Mol Biol* 20, 182–187 (2013).

89. Strowitzki, M. J., Cummins, E. P. & Taylor, C. T. Protein Hydroxylation by Hypoxia-Inducible Factor (HIF) Hydroxylases: Unique or Ubiquitous? *Cells* 8, 384 (2019).

90. Carrero, P. *et al.* Redox-Regulated Recruitment of the Transcriptional Coactivators CREB-Binding Protein and SRC-1 to Hypoxia-Inducible Factor 1α. *Mol Cell Biol* 20, 402–415 (2000).

91. Partch, C. L. & Gardner, K. H. Coactivator recruitment: A new role for PAS domains in transcriptional regulation by the bHLH-PAS family. *J Cell Physiol* 223, 553–557 (2010).

92. Ghosh, A. K. & Varga, J. The transcriptional coactivator and acetyltransferase p300 in fibroblast biology and fibrosis. *J Cell Physiol* 213, 663–671 (2007).

93. Fraguas, S. *et al.* CREB-binding protein (CBP) gene family regulates planarian survival and stem cell differentiation. *Dev Biol* 476, 53–67 (2021).

94. Smith, R. H. B., Dar, A. C. & Schlessinger, A. PyVOL: a PyMOL plugin for visualization, comparison, and volume calculation of drug-binding sites. *bioRxiv* 816702 (2019).

95. Wright, C. S., Mi, L. Z., Lee, S. & Rastinejad, F. Crystal structure analysis of phosphatidylcholine-GM2-activator product complexes: evidence for hydrolase activity. *Biochemistry* 44, 13510–13521 (2005).

96. Wright, C. S., Mi, L. Z. & Rastinejad, F. Evidence for lipid packaging in the crystal structure of the GM2-activator complex with platelet activating factor. *J Mol Biol* 342, 585–592 (2004).

97. Wright, C. S., Zhao, Q. & Rastinejad, F. Structural analysis of lipid complexes of GM2-activator protein. *J Mol Biol* 331, 951–964 (2003).

98. Schmidt, J. V, Su, G. H., Reddy, J. K., Simon, M. C. & Bradfield, C. A. Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proc Natl Acad Sci U S A* 93, 6731–6736 (1996).

99. Iyer, N. V. *et al.* Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1α. *Genes Dev* 12, 149–162 (1998).

100. Gekakis, N. *et al.* Role of the CLOCK Protein in the Mammalian Circadian Mechanism. *Science (1979)* 280, 1564–1569 (1998).

101. Moura-Alves, P. *et al.* AhR sensing of bacterial pigments regulates antibacterial defence. *Nature* 512, 387–392 (2014).

102. Neavin, D. R., Liu, D., Ray, B. & Weinshilboum, R. M. The Role of the Aryl Hydrocarbon Receptor (AHR) in Immune and Inflammatory Diseases. *Int J Mol Sci* 19, 3851 (2018).

103. Gutiérrez-Vázquez, C. & Quintana, F. J. Regulation of the Immune Response by the Aryl Hydrocarbon Receptor. *Immunity* 48, 19–33 (2018).

104. Rico-Leo, E. M. *et al.* Aryl hydrocarbon receptor controls skin homeostasis, regeneration, and hair follicle cycling by adjusting epidermal stem cell function. *Stem Cells* 39, 1733–1750 (2021).

105. Bissonnette, R., Stein Gold, L., Rubenstein, D. S., Tallman, A. M. & Armstrong, A. Tapinarof in the treatment of psoriasis: A review of the unique mechanism of action of a novel therapeutic aryl hydrocarbon receptor–modulating agent. *J Am Acad Dermatol* 84, 1059–1067 (2021).

106. Diao, X. *et al.* Structural basis for the ligand-dependent activation of heterodimeric AHR-ARNT complex. *Nat Commun* 16, 1–14 (2025).

107. Ondrová, K. *et al.* Monoterpenoid aryl hydrocarbon receptor allosteric antagonists protect against ultraviolet skin damage in female mice. *Nat Commun* 14, (2023).

108. Flegel, J. *et al.* The Highly Potent AhR Agonist Picoberin Modulates Hh-Dependent Osteoblast Differentiation. *J Med Chem* 65, 16268–16289 (2022).

109. Wilkinson, I. V. L. *et al.* Chemical Proteomics and Phenotypic Profiling Identifies the Aryl Hydrocarbon Receptor as a Molecular Target of the Utrophin Modulator Ezutromid. *Angew Chem Int Ed Engl* 59, 2420–2428 (2020).

110. Walczak, K. *et al.* Effect of Tryptophan-Derived AhR Ligands, Kynurenine, Kynurenic Acid and FICZ, on Proliferation, Cell Cycle Regulation and Cell Death of Melanoma Cells—In Vitro Studies. *Int J Mol Sci* 21, 7946 (2020).

111. DiNatale, B. C. *et al.* Kynurenic Acid Is a Potent Endogenous Aryl Hydrocarbon Receptor Ligand that Synergistically Induces Interleukin-6 in the Presence of Inflammatory Signaling. *Toxicol Sci* 115, 89–97 (2010).

112. Masoud, G. N. & Li, W. HIF-1α pathway: role, regulation and intervention for cancer therapy. *Acta Pharm Sin B* 5, 378–389 (2015).

113. Zhao, J., Du, F., Shen, G., Zheng, F. & Xu, B. The role of hypoxia-inducible factor-2 in digestive system cancers. *Cell Death Dis* 6, e1600–e1600 (2015).

114. Keith, B., Johnson, R. S. & Simon, M. C. HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer* 12, 9–22 (2011).

115. Scheuermann, T. H. *et al.* Artificial ligand binding within the HIF2α PAS-B domain of the HIF2 transcription factor. *Proc Natl Acad Sci U S A* 106, 450–455 (2009).

116. Cho, H. & Kaelin, W. G. Targeting HIF2 in Clear Cell Renal Cell Carcinoma. *Cold Spring Harb Symp Quant Biol* 81, 113–121 (2016).

117. Wu, D. *et al.* Bidirectional modulation of HIF-2 activity through chemical ligands. *Nat Chem Biol* 15, 367–376 (2019).

118. Gupta, N. & Wish, J. B. Hypoxia-Inducible Factor Prolyl Hydroxylase Inhibitors: A Potential New Treatment for Anemia in Patients With CKD. *Am J Kidney Dis* 69, 815–826 (2017).

119. Maxwell, P. H. & Eckardt, K.-U. HIF prolyl hydroxylase inhibitors for the treatment of renal anaemia and beyond. *Nat Rev Nephrol* 12, 157–168 (2016).

120. DeBruyne, J. P., Weaver, D. R. & Reppert, S. M. CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nat Neurosci* 10, 543–545 (2007).

121. Rutter, J., Reick, M., Wu, L. C. & McKnight, S. L. Regulation of crock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science (1979)* 293, 510–514 (2001).

122. Dong, Z. *et al.* Targeting glioblastoma stem cells through disruption of the circadian clock. *Cancer Discov* 9, 1556–1573 (2019).

123. Bass, J. & Takahashi, J. S. Circadian integration of metabolism and energetics. *Science (1979)* 330, 1349–1354 (2010).

124. Awuah, W. A. *et al.* Circadian rhythm in systemic autoimmune conditions: Potential of chrono-immunology in clinical practice: A narrative review. *Medicine (United States)* 102, E34614 (2023).

125. Difrancesco, S. *et al.* Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. *Depress Anxiety* 36, 975–986 (2019).

126. Nguyen, C., Murray, G., Anderson, S., Filipowicz, A. & Ingram, K. K. In vivo molecular chronotyping, circadian misalignment, and high rates of depression in young adults. *J Affect Disord* 250, 425–431 (2019).

127. Crnko, S., Du Pré, B. C., Sluijter, J. P. G. & Van Laake, L. W. Circadian rhythms and the molecular clock in cardiovascular biology and disease. *Nat Rev Cardiol* 16, 437–447 (2019).

128. Pickard, B. S., Malloy, M. P., Porteous, D. J., Blackwood, D. H. R. & Muir, W. J. Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. *Am J Med Genet B Neuropsychiatr Genet* 136B, 26–32 (2005).

129. Weber, H. *et al.* Cross-disorder analysis of bipolar risk genes: further evidence of DGKH as a risk gene for bipolar disorder, but also unipolar depression and adult ADHD. *Neuropsychopharmacology* 36, 2076–2085 (2011).

130. Huang, J. *et al.* Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry* 167, 1254–1263 (2010).

131. Pickard, B. S. *et al.* Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. *Mol Psychiatry* 14, 874–884 (2009).

132. Pickard, B. S., Pieper, A. A., Porteous, D. J., Blackwood, D. H. & Muir, W. J. The NPAS3 gene--emerging evidence for a role in psychiatric illness. *Ann Med* 38, 439–48 (2006).

133. Kamnasaran, D. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. *J Med Genet* 40, 325–332 (2003).

134. Ledoux, A. A. *et al.* Structural Hippocampal Anomalies in a Schizophrenia Population Correlate with Navigation Performance on a Wayfinding Task. *Front Behav Neurosci* 8, 88 (2014).

135. Nakahara, S., Matsumoto, M. & van Erp, T. G. M. Hippocampal subregion abnormalities in schizophrenia: A systematic review of structural and physiological imaging studies. *Neuropsychopharmacol Rep* 38, 156 (2018).

136. Sha, L. *et al.* Transcriptional regulation of neurodevelopmental and metabolic pathways by NPAS3. *Mol Psychiatry* 17, 267–279 (2012).

137. Bradshaw, N. J. & Korth, C. Protein misassembly and aggregation as potential convergence points for non-genetic causes of chronic mental illness. *Mol Psychiatry* 24, 936–951 (2019).

138. Nucifora, L. G. *et al.* A Mutation in NPAS3 That Segregates with Schizophrenia in a Small Family Leads to Protein Aggregation. *Mol Neuropsychiatry* 2, 133–144 (2016).

139. Yu, L. *et al.* A mutation in NPAS3 segregates with mental illness in a small family. *Mol Psychiatry* 19, 7–8 (2014).

140. Sullivan, A. E. *et al.* Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2. *Biochem J* 461, 403–412 (2014).

141. Abbott, B. D. & Probst, M. R. Developmental expression of two members of a new class of transcription factors: II. Expression of aryl hydrocarbon receptor nuclear translocator in the C57BL/6N mouse embryo. *Developmental Dynamics* 204, 144–155 (1995).

142. Michaud, J. L., Derossi, C., May, N. R., Holdener, B. C. & Fan, C. M. ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus. *Mech Dev* 90, 253–261 (2000).

143. Hosoya, T. *et al.* Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice. *Genes to Cells* 6, 361–374 (2001).

144. Schweitzer, J., Löhr, H., Bonkowsky, J. L., Hübscher, K. & Driever, W. Sim1a and Arnt2 contribute to hypothalamo-spinal axon guidance by regulating Robo2 activity via a Robo3- dependent mechanism. *Development (Cambridge)* 140, 93–106 (2013).

145. Dela Cruz, J. A. D., Schmidt-Kastner, R., Stevens, J. A. A., Steinbusch, H. W. M. & Rutten, B. P. F. Differential distribution of hypoxia-inducible factor 1-beta (ARNT or ARNT2) in mouse substantia nigra and ventral tegmental area. *J Chem Neuroanat* 61–62, 64–71 (2014).

146. Lin, M. *et al.* Heat shock alters the expression of schizophrenia and autism candidate genes in an induced pluripotent stem cell model of the human telencephalon. *PLoS One* 9, (2014).

147. Di Napoli, A., Warrier, V., Baron-Cohen, S. & Chakrabarti, B. Genetic variant rs17225178 in the ARNT2 gene is associated with Asperger Syndrome. *Mol Autism* 6, (2015).

148. Drutel, G. ARNT2, a transcription factor for brain neuron survival? *Eur J Neurosci* 11, 1545–1553 (1999).

149. Rahim, T., Becquart, P., Baeva, M. E. & Quandt, J. Expression of the neuroprotective protein aryl hydrocarbon receptor nuclear translocator 2 correlates with neuronal stress and disability in models of multiple sclerosis. *J Neuroinflammation* 15, 1–17 (2018).

150. Lin, Y. *et al.* Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature* 455, 1198–1204 (2008).

151. Pruunsild, P., Sepp, M., Orav, E., Koppel, I. & Timmusk, T. Identification of cis-elements and transcription factors regulating neuronal activity-dependent transcription of human BDNF gene. *J Neurosci* 31, 3295–3308 (2011).

152. Li, P. *et al.* Discovery of a highly potent NPAS3 heterodimer inhibitor by covalently modifying ARNT. *Bioorg Chem* 139, 106676 (2023).

153. Maltepe, E., Keith, B., Arsham, A. M., Brorson, J. R. & Simon, M. C. The Role of ARNT2 in Tumor Angiogenesis and the Neural Response to Hypoxia. *Biochem Biophys Res Commun* 273, 231–238 (2000).

154. Turer, E. E. *et al.* A viable hypomorphic Arnt2 mutation causes hyperphagic obesity, diabetes and hepatic steatosis. *Dis Model Mech* 11, (2018).

155. Pieper, A. A. *et al.* Discovery of a Pro-neurogenic, Neuroprotective Chemical. *Cell* 142, 39 (2010).

156. Wang, G. *et al.* P7C3 neuroprotective chemicals function by activating the rate-limiting enzyme in NAD salvage. *Cell* 158, 1324–1334 (2014).

157. Kaoud, T. S. *et al.* Modulating multi-functional ERK complexes by covalent targeting of a recruitment site in vivo. *Nat Commun* 10, 1–15 (2019).

158. Stebbins, J. L. *et al.* Identification of a new JNK inhibitor targeting the JNK-JIP interaction site. *Proc Natl Acad Sci U S A* 105, 16809–16813 (2008).

159. Guo, Y. *et al.* Regulating the ARNT/TACC3 axis: Multiple approaches to manipulating protein/protein interactions with small molecules. *ACS Chem Biol* 8, 626–635 (2013).

160. Gagné, D. *et al.* Use of High Pressure NMR Spectroscopy to Rapidly Identify Proteins with Internal Ligand-Binding Voids. *bioRxiv* (2023).

161. Xu, X. *et al.* Identification of small-molecule ligand-binding sites on and in the ARNT PAS-B domain. *bioRxiv* (2024).

162. Wu, D., Potluri, N., Kim, Y. & Rastinejad, F. Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain. *Mol Cell Biol* 33, 4346–4356 (2013).

163. Erbel, P. J. A., Card, P. B., Karakuzu, O., Bruick, R. K. & Gardner, K. H. Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor. *Proc Natl Acad Sci U S A* 100, 15504–15509 (2003).

164. Henry, J. T. & Crosson, S. Ligand binding PAS domains in a genomic, cellular, and structural context. *Annu Rev Microbiol* 65, 261 (2011).

165. Zhuang, J., Shang, Q., Rastinejad, F. & Wu, D. Decoding Allosteric Control in Hypoxia-Inducible Factors. *J Mol Biol* 436, 168352 (2024).

166. Lim, J. M., Kim, G. & Levine, R. L. Methionine in Proteins: It's not just for protein initiation anymore. *Neurochem Res* 44, 247 (2018).

167. Wingfield, P. T. N-terminal methionine processing. *Curr Protoc Protein Sci* 2017, 6.14.1-6.14.3 (2017).

168. Geoghegan, K. F. *et al.* Spontaneous α-N-6-Phosphogluconoylation of a "His Tag" inEscherichia coli:The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Anal Biochem* 267, 169–184 (1999).

169. Lo, M.-C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* 332, 153–159 (2004).

170. Samuel, E. L. G., Holmes, S. L. & Young, D. W. Processing binding data using an open-source workflow. *J Cheminform* 13, 1–11 (2021).

171. Samuelsen, L., Holm, R., Lathuile, A. & Schönbeck, C. Buffer solutions in drug formulation and processing: How pKa values depend on temperature, pressure and ionic strength. *Int J Pharm* 560, 357–364 (2019).

172. Oh, K. I., Rajesh, K., Stanton, J. F. & Baiz, C. R. Quantifying Hydrogen-Bond Populations in Dimethyl Sulfoxide/Water Mixtures. *Angewandte Chemie - International Edition* 56, 11375–11379 (2017).

173. Nissink, J. W. M., Schmitt, S., Blackburn, S. & Peters, S. Stratified high-throughput screening sets enable flexible screening strategies from a single plated collection. *J Biomol Screen* 19, 369–378 (2014).

174. Baell, J. & Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* 513, 481–483 (2014).

175. Wu, J. *et al.* Neuronal store-operated calcium entry pathway as a novel therapeutic target for Huntington's disease treatment. *Chem Biol* 18, 777–793 (2011).

176. Grekhnev, D. A., Kruchinina, A. A., Vigont, V. A. & Kaznacheyeva, E. V. The Mystery of EVP4593: Perspectives of the Quinazoline-Derived Compound in the Treatment of Huntington's Disease and Other Human Pathologies. *Int J Mol Sci* 23, 15724 (2022).

177. Prajapati, R. *et al.* Isoliquiritigenin, a potent human monoamine oxidase inhibitor, modulates dopamine D1, D3, and vasopressin V1A receptors. *Sci Rep* 11, 23528 (2021).

178. Zhao, J. H. *et al.* Clinically approved small-molecule drugs for the treatment of rheumatoid arthritis. *Eur J Med Chem* 256, (2023).

179. Bevilacqua, M. Effects of oxaprozin and of other 2-arylpropionic acid derivatives on nuclear factor κB(NF-κB) activation. *Inflammopharmacology* 10, 173–183 (2002).

180. Tobe, M. *et al.* Discovery of quinazolines as a novel structural class of potent inhibitors of NF-κB activation. *Bioorg Med Chem* 11, 383–391 (2003).

181. Tobe, M. *et al.* A novel structural class of potent inhibitors of NF-κB activation: structure–activity relationships and biological effects of 6-aminoquinazoline derivatives. *Bioorg Med Chem* 11, 3869–3878 (2003).

182. Munster, P. N. *et al.* Phase I trial of a dual TOR kinase and DNA-PK inhibitor (CC-115) in advanced solid and hematologic cancers. *J Clin Oncol* 34, 2505–2505 (2016).

183. Mortensen, D. S. *et al.* Optimization of a Series of Triazole Containing Mammalian Target of Rapamycin (mTOR) Kinase Inhibitors and the Discovery of CC-115. *J Med Chem* 58, 5599–5608 (2015).

184. Thijssen, R. *et al.* Dual TORK/DNA-PK inhibition blocks critical signaling pathways in chronic lymphocytic leukemia. *Blood* 128, 574–583 (2016).

185. Arienti, K. L. *et al.* Checkpoint kinase inhibitors: SAR and radioprotective properties of a series of 2-arylbenzimidazoles. *J Med Chem* 48, 1873–1885 (2005).

186. Cabrera, N. *et al.* Searching glycolate oxidase inhibitors based on QSAR, molecular docking, and molecular dynamic simulation approaches. *Sci Rep* 12, 1–19 (2022).

187. Liu, X. *et al.* Inhibition of hepatitis B virus replication by targeting ribonucleotide reductase M2 protein. *Biochem Pharmacol* 103, 118–128 (2016).

188. Tang, Q. *et al.* Osalmid, a Novel Identified RRM2 Inhibitor, Enhances Radiosensitivity of Esophageal Cancer. *Int J Radiat Oncol Biol Phys* 108, 1368–1379 (2020).

189. Barrow, J. C. *et al.* Discovery of 4,4-disubstituted quinazolin-2-ones as T-type calcium channel antagonists. *ACS Med Chem Lett* 1, 75–79 (2010).

190. Uslaner, J. M. *et al.* T-type calcium channel antagonism produces antipsychotic-like effects and reduces stimulant-induced glutamate release in the nucleus accumbens of rats. *Neuropharmacology* 62, 1413–1421 (2012).

191. Domingo, C., Palomares, O., Sandham, D. A., Erpenbeck, V. J. & Altman, P. The prostaglandin D2 receptor 2 pathway in asthma: A key player in airway inflammation. *Respir Res* 19, 1–8 (2018).

192. Malykh, A. G. & Sadaie, M. R. Piracetam and Piracetam-Like Drugs. *Drugs 2010 70:3* 70, 287–312 (2012).

193. Liu, T. *et al.* Effect of L-oxiracetam and oxiracetam on memory and cognitive impairment in mild-to-moderate traumatic brain injury patients: Study protocol for a randomized controlled trial. *Aging Medicine* 7, 341–349 (2024).

194. Gouliaev, A. H. & Senning, A. Piracetam and other structurally related nootropics. *Brain Res Rev* 19, 180–222 (1994).

195. Li, W. *et al.* (S)-Oxiracetam is the Active Ingredient in Oxiracetam that Alleviates the Cognitive Impairment Induced by Chronic Cerebral Hypoperfusion in Rats. *Scientific Reports 2017 7:1* 7, 1–14 (2017).

196. Okuda, M. *et al.* Design and synthesis of curcumin derivatives as tau and amyloid β dual aggregation inhibitors. *Bioorg Med Chem Lett* 26, 5024–5028 (2016).

197. Shashoua, V. E. & Hesse, G. W. N-docosahexaenoyl, 3 hydroxytyramine : A dopaminergic compound that penetrates the blood-brain barrier and suppresses appetite. *Life Sci* 58, 1347–1357 (1996).

198. Van Pelt-Koops, J. C. *et al.* The spiroindolone drug candidate NITD609 potently inhibits gametocytogenesis and blocks Plasmodium falciparum transmission to anopheles mosquito vector. *Antimicrob Agents Chemother* 56, 3544 (2012).

199. Zhu, X. *et al.* Isoliquiritigenin attenuates lipopolysaccharide-induced cognitive impairment through antioxidant and anti-inflammatory activity. *BMC Neurosci* 20, (2019).

200. Zhu, X. *et al.* Neuroprotective and anti-inflammatory effects of isoliquiritigenin in kainic acid-induced epileptic rats via the TLR4/MYD88 signaling pathway. *Inflammopharmacology* 27, 1143–1153 (2019).

201. Shi, D. *et al.* The antioxidant activity and neuroprotective mechanism of isoliquiritigenin. *Free Radic Biol Med* 152, 207–215 (2020).

202. Zhu, X. *et al.* Neuroprotective effects of isoliquiritigenin against cognitive impairment via suppression of synaptic dysfunction, neuronal injury, and neuroinflammation in rats with kainic acid-induced seizures. *Int Immunopharmacol* 72, 358–366 (2019).

203. Okuda, M. *et al.* PE859, a novel tau aggregation inhibitor, reduces aggregated tau and prevents onset and progression of neural dysfunction in vivo. *PLoS One* 10, e0117511 (2015).

204. Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc Natl Acad Sci U S A* 103, 19678–19682 (2006).

205. Xu, X., Dikiy, I., Evans, M. R., Marcelino, L. P. & Gardner, K. H. Fragile protein folds: sequence and environmental factors affecting the equilibrium of two interconverting, stably folded protein conformations. *Magnetic Resonance* 2, 63–76 (2021).

206. Asmari, M. *et al.* Studying molecular interactions via capillary electrophoresis and microscale thermophoresis: A review. *Electrophoresis* 44, 1114–1142 (2023).

207. Seidel, S. A. I. *et al.* Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* 59, 301–315 (2013).

208. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct* 1077, 101–113 (2014).

209. Nguyen, H. H., Park, J., Kang, S. & Kim, M. Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. *Sensors* 15, 10481–10510 (2015).

210. Lameirinhas, R. A. M., Torres, J. P. N., Baptista, A. & Martins, M. J. M. A New Method to Analyse the Role of Surface Plasmon Polaritons on Dielectric-Metal Interfaces. *IEEE Photonics J* 14, (2022).

211. Damborský, P., Švitel, J. & Katrlík, J. Optical biosensors. *Essays Biochem* 60, 91–100 (2016).

212. Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. & Salemme, F. R. Structural origins of high-affinity biotin binding to streptavidin. *Science (1979)* 243, 85–88 (1989).

213. Renuse, S. *et al.* Signature Fragment Ions of Biotinylated Peptides. *J Am Soc Mass Spectrom* 31, 394–404 (2020).

214. Wang, W. W. S., Das, D. & Suresh, M. R. Biotin carboxyl carrier protein co-purifies as a contaminant in core-streptavidin preparations. *Mol Biotechnol* 31, 29–40 (2005).

215. Edwards, P. R. *et al.* Kinetics of Protein-Protein Interactions at the Surface of an Optical Biosensor. *Anal Biochem* 231, 210–217 (1995).

216. Edwards, P. R. & Leatherbarrow, R. J. Determination of association rate constants by an optical biosensor using initial rate analysis. *Anal Biochem* 246, 1–6 (1997).

217. Jug, A., Bratkovič, T. & Ilaš, J. Biolayer interferometry and its applications in drug discovery and development. *TrAC Trends in Analytical Chemistry* 176, 117741 (2024).

218. Saponaro, A. Isothermal Titration Calorimetry: A Biophysical Method to Characterize the Interaction between Label-free Biomolecules in Solution. *Bio Protoc* 8, (2018).

219. Yuan, Y. *et al.* Tetrazoles: A multi-potent motif in drug design. *Eur J Med Chem* 279, (2024).

220. Young, A. M., Audus, K. L., Proudfoot, J. & Yazdanian, M. Tetrazole compounds: The effect of structure and pH on Caco-2 cell permeability. *J Pharm Sci* 95, 717–725 (2006).

221. Zou, Y., Liu, L., Liu, J. & Liu, G. Bioisosteres in drug discovery: focus on tetrazole. *Future Med Chem* 12, 91–93 (2020).

222. Roberts, R. L. *et al.* Acidic triazoles as soluble guanylate cyclase stimulators. *Bioorg Med Chem Lett* 21, 6515–6518 (2011).

223. Boezio, A. A. *et al.* 1,2,4-Triazolsulfone: A novel isosteric replacement of acylsulfonamides in the context of NaV1.7 inhibition. *Bioorg Med Chem Lett* 28, 2103–2108 (2018).

224. Song, N. J. *et al.* Butein is a novel anti-adipogenic compound. *J Lipid Res* 54, 1385–1396 (2013).

225. Lee, S. D. *et al.* Butein effects in colitis and interleukin-6/signal transducer and activator of transcription 3 expression. *World J Gastroenterol* 21, 465–474 (2015).

226. Semwal, R. B., Semwal, D. K., Combrinck, S. & Viljoen, A. Butein: From ancient traditional remedy to modern nutraceutical. *Phytochem Lett* 11, 188–201 (2015).

227. Kojima, R., Kawachi, M. & Ito, M. Butein suppresses ICAM-1 expression through the inhibition of IκBα and c-Jun phosphorylation in TNF-α- and PMA-treated HUVECs. *Int Immunopharmacol* 24, 267–275 (2015).

228. Rasheed, Z., Akhtar, N., Khan, A., Khan, K. A. & Haqqi, T. M. Butrin, isobutrin, and butein from medicinal plant Butea monosperma selectively inhibit nuclear factor-kappaB in activated human mast cells: suppression of tumor necrosis factor-alpha, interleukin (IL)-6, and IL-8. *J Pharmacol Exp Ther* 333, 354–363 (2010).

229. Lee, S. H., Seo, G. S. & Sohn, D. H. Inhibition of lipopolysaccharide-induced expression of inducible nitric oxide synthase by butein in RAW 264.7 cells. *Biochem Biophys Res Commun* 323, 125–132 (2004).

230. Szuster-Ciesielska, A., Mizerska-Dudka, M., Daniluk, J. & Kandefer-Szerszeń, M. Butein inhibits ethanol-induced activation of liver stellate cells through TGF-β, NFκB, p38, and JNK signaling pathways and inhibition of oxidative stress. *J Gastroenterol* 48, 222–237 (2013).

231. Padmavathi, G. *et al.* Potential of butein, a tetrahydroxychalcone to obliterate cancer. *Phytomedicine* 22, 1163–1171 (2015).

232. Lagunin, A., Filimonov, D. & Poroikov, V. Multi-targeted natural products evaluation based on biological activity prediction with PASS. *Curr Pharm Des* 16, 1703–1717 (2010).

233. Padmavathi, G. *et al.* Butein in health and disease: A comprehensive review. *Phytomedicine* 25, 118–127 (2017).

234. Cisneros, K. V., Agarwal, V. & James, M. O. Sulfonation and Glucuronidation of Hydroxylated Bromodiphenyl Ethers in Human Liver. *Chemosphere* 226, 132–139 (2019).

235. Guengerich, F. P. & Yoshimoto, F. K. Formation and Cleavage of C-C Bonds by Enzymatic Oxidation-Reduction Reactions. *Chem Rev* 118, 6573–6655 (2018).

236. Adusumalli, V. N. K. B., Woźny, P., Lis, S. & Park, Y. Il. Furoic Acid-Capped LaF3:Tb Nanoparticles for Al3+ Detection Based on Aggregation-Induced Luminescence Enhancement. *ACS Appl Nano Mater* 7, 17678–17687 (2024).

237. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486–501 (2010).

238. Agirre, J. *et al.* The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr D Struct Biol* 79, 449–461 (2023).