Structure-Function Studies of Kinetoplastid Proteins

Jennifer Timm

Doctor of Philosophy

University of York

Chemistry

September 2014

Abstract

The class kinetoplastida include parasites responsible for devastating diseases like African sleeping sickness, Chagas's disease and Leishmaniases, mainly effecting people in the developing world. Current treatments are highly toxic and inefficient, leading to an urgent need of novel anti-parasitic compounds. This thesis focuses on the structural characterisation of potential drug targets against these parasites, namely adenosine kinase from *Trypanosoma brucei* (*Tb*AK), thymidine kinases from *T. brucei* (*Tb*TK) and *Leishmania major* (*Lm*TK) and the leucyl aminopeptidases from *T. brucei* (*Tb*LAP-A)*, T. cruzi* (*Tc*LAP-A) and *L. major* (*Lm*LAP-A).

Structures of *Tb*AK were solved in two conformations, open (apo) and closed (in complex with adenosine and ADP), both to 2.6 Å. Comprised of a big α/β-domain and a small lid domain, the structures confirm the large conformational change of the lid domain upon substrate binding. The structures of C-terminally truncated versions of *Lm*TK and *Tb*TK were determined as ligand-bound complexes with resolutions up to 2.4 Å and 2.2 Å, respectively. They show high similarity to structures of homologues in the PDB. The structures solved in this thesis give valuable information about ligand binding and aid rational drug design.

Leucyl aminopeptidase (LAP-A) was evaluated as a potential drug target in *T. brucei* parasites. It is not essential for *T. brucei* parasites grown *in vitro*, shown by generation and analysis of LAP-A-depleted parasites. Although this does not support LAP-A as a drug target in *T. brucei*, no conclusions can be drawn about the potential in *T. cruzi* and *L. major*. Several structures of the LAP-As were solved, the highest resolution ones to 2.3 Å, 2.3 Å and 2.5 Å for *Tb*LAP-A, *Tc*LAP-A and *Lm*LAP-A, respectively. These enzymes are hexameric and show the typical two-domain architecture of M17 LAPs. Although the physiological function remains elusive, the work in this thesis provides a firm basis for future studies.

Table of Contents

Abstract **2**

Table of Contents **3**

List of Figures **8**

List of Tables **12**

Acknowledgments **14**

Author's Declaration **15**

Chapter 1 Introduction **17**

1.1 Neglected tropical diseases **17**

1.2 Kinetoplastids and the TriTryps **17**

1.3 *Trypanosoma brucei* **19**

1.3.1 *T. brucei* cell cycle 22

1.3.2 *T. brucei* life cycle 23

1.3.3 Antigenic variation 25

1.3.4 Human African Trypanosomiasis (HAT) 26

1.3.5 Diagnosis and Current treatment of HAT 27

1.4 *Trypanosoma cruzi* **29**

1.4.1 *T. cruzi* life cycle 31

1.4.2 Chagas' Disease 32

1.4.3 Diagnosis and Current treatment of Chagas' disease 33

1.5 *Leishmania spp* **34**

1.5.1 *Leishmania spp*. life cycle 36

1.5.2 Leishmaniases 37

1.5.3 Diagnosis and Current treatment of leishmaniasis 39

1.5.4 HIV co-infection 41

1.6 Vaccine development against the TriTryps **41**

1.7 Drug discovery **44**

1.7.1 Classical Pharmacology – Forward Pharmacology 44

1.7.2 Target-based drug design – Reverse Pharmacology 44

1.8 Identification of potential drug targets in the TriTryps **45**

1.8.1 Nucleotide metabolism 45

1.8.2 Amino acid metabolism 53
1.8.3 Other potential drug targets in the TriTryps 60
1.9 Trypobase 60
1.10 Target selection in Trypobase 61
1.11 Aims 62
Chapter 2 Adenosine Kinase from *T. brucei brucei* 63
2.1 Introduction 63
2.1.1 Adenosine Kinase from *T. brucei brucei* 64
2.1.2 AK Mechanism 67
2.1.3 Transition state analogues for phosphoryl transfer 70
2.1.4 Aims 71
2.2 Experimental 72
2.2.1 Cloning of *Tb*AK constructs 72
2.2.2 Expression of *Tb*AK 74
2.2.3 Purification and characterisation of *Tb*AK 75
2.2.4 Crystallisation of *Tb*AK 77
2.2.5 Data collection, structure solution and refinement of *Tb*AK 79
2.3 Results 81
2.3.1 Purification and Characterisation of *Tb*AK 81
2.3.2 Crystallisation of *Tb*AK 82
2.3.3 Structures of *Tb*AK 85
2.4 Discussion 92
2.4.1 Comparison to other published structures 92
2.4.2 Modelling of *Tb*AK-2 and comparison to *Tb*AK-1 94
2.4.3 Cordycepin and its derivatives as anti-trypanosomal agents 95
2.5 Summary and future work 97
Chapter 3 Thymidine Kinases of the TriTryps 98
3.1 Introduction 98
3.1.1 Thymidine kinases from the TriTryps 101
3.1.2 TK Mechanism 104
3.1.3 TKs in drug development 107
3.1.4 Aims 107
3.2 Experimental 108
3.2.1 Cloning and expression testing in *E. coli* 108

3.2.2 Cloning of TK constructs for expression with the baculovirus system 110
3.2.3 Construction of baculoviruses and insect cells lines for expression 111
3.2.4 Expression tests of TKs with the Baculovirus system 113
3.2.5 Expression of *Lm*TK and *Tb*TK in Sf9 cells 113
3.2.6 Purification and Characterisation of *Lm*TK 114
3.2.7 Purification and Characterisation of *Tb*TK 116
3.2.8 Crystallisation of *Lm*TK and *Tb*TK 118
3.2.9 Data collection and Structure solution of *Lm*TK and *Tb*TK 120
3.3 Results 124
3.3.1 Expression and purification of *Lm*TK and *Tb*TK 124
3.3.2 Crystallisation of *Lm*TK and *Tb*TK 124
3.3.3 Structures of *Lm*TK and *Tb*TK 125
3.4 Discussion 141
3.5 Summary and future work 144

Chapter 4 Characterisation of the M17 family leucyl aminopeptidases from the TriTryps 145

4.1 Introduction 145
4.1.1 M17 family of leucyl aminopeptidases 145
4.1.2 M17 LAPs from the TriTryps 148
4.1.3 Mechanism of LAPs 153
4.1.4 Physiological Functions of LAPs 155
4.1.5 LAPs as drug targets and vaccine candidates 156
4.1.6 Aims 157
4.2 Experimental 158
4.2.1 Cloning of LAP-A constructs for structural studies 158
4.2.2 Expression of LAP-As in *E. coli* 159
4.2.3 Purification of LAP-As 159
4.2.4 Native PAGE zymography 160
4.2.5 Generation of anti-*Tb*LAP-A antisera in rabbit and affinity purification of polyclonal anti-*Tb*LAP-A antibody 161
4.2.6 Validation of polyclonal anti-*Tb*LAP-A antibody 162
4.2.7 SDS-PAGE and Western blots 162
4.2.8 Generation of the construct for ectopic expression of LAP-A 163
4.2.9 Generation of constructs for LAP-A RNAi 164

4.2.10 Generation of LAP-A cMyc-fusion constructs for overexpression in *T. brucei* .166
4.2.11 Generation of *T. brucei* LAP-A knockout cell lines .166
4.2.12 *Trypanosoma brucei* cell lines, culture and transfections .169
4.2.13 *T. brucei* cultures in minimal media .171
4.2.14 Determination of parasite growth upon RNAi induction and overexpression of LAP-A. .172
4.2.15 Statistical analysis of changes in parasite growth .172
4.2.16 Glycerol stocks of *T. brucei* clones .173
4.2.17 Cell cycle analysis by flow cytometry .173
4.2.18 Cell cycle analysis by fluorescence microscopy .173
4.2.19 Immunofluorescence .174
4.2.20 Crystallisation of the LAP-As .175
4.2.21 Data Collection and Structure Solution of the LAP-As .177
4.2.22 Identification of metal ions in the *Tb*LAP-A structure .183
4.2.23 SEC-MALLS .183
4.3 Results .184
4.3.1 Overexpression and purification of active protein .184
4.3.2 Validation of anti-*Tb*LAP-A antiserum .185
4.3.3 *Trypanosoma brucei brucei* cell lines, culture and transfections .187
4.3.4 Growth and adaptation to media .188
4.3.5 Cell cycle analysis: flow cytometry and fluorescence microscopy analysis .189
4.3.6 RNAi mediated depletion of LAP-A in bloodstream forms of *T. brucei* .190
4.3.7 RNAi mediated depletion of LAP-A and generation of LAP-A-dKO cells in procyclic forms of *T. brucei* .192
4.3.8 Overexpression of LAP-A in bloodstream and procyclic forms of *T. brucei* .193
4.3.9 Amino acid limitation in procyclic forms of *T. brucei* .195
4.3.10 Subcellular localisation of LAP-A in *T. brucei* .198
4.3.11 Crystallisation of the LAP-As .201
4.3.12 TriTryp LAP-A structures .201
4.3.13 Oligomeric State in solution .221
4.4 Discussion .227
4.4.1 LAP-A as potential drug target in *T. brucei* .227
4.4.2 Physiological function of the LAP-A .228
4.4.3 Potential use of LAP-A-dKO, RNAi and overexpression strains .229

4.4.4 Structure and oligomeric state of LAP-A and their effects on possible physiological substrates 229
4.4.5 The physiological metal ion in the active site of LAP-A 230
4.5 Summary and future work 231
Chapter 5 Summary and Conclusions 232
Appendices 234
A.1 Expression, purification, and crystallisation of thymidylate kinases 234
A.1.1 Introduction 234
A.1.2 Experimental 238
A.1.3 Results 241
A.1.4 Discussion 242
A.2 Sequences and plasmid maps 243
A.2.1 Protein sequences 243
A.2.2 Nucleotide sequences of *Tb*LAP-A RNAi fragment and UTRs for double KO 244
A.2.3 Primer sequences 245
A.2.3 List of constructs 247
A.2.4 Plasmid maps 249
A.3 Media composition for *T. brucei* parasite cultures 254
List of Abbreviations 257
References 261
Bibliography 278

List of Figures

Figure 1-1: Schematic representation of kinetoplastid cell morphology. 18
Figure 1-2: Positioning of nucleus, kinetoplast and flagellum. 19
Figure 1-3: Geographic endemicity of HAT caused by *T. brucei gambiense*, 2013.... 20
Figure 1-4: Geographic endemicity of HAT caused by *T. brucei rhodesiense,* 2013.. 21
Figure 1-5: Cell cycle of procyclic *T. brucei*. 22
Figure 1-6: Life cycle stages of *T. brucei*..... 23
Figure 1-7: Anti-parasitic compounds used for treatment of HAT. 28
Figure 1-8: Geographic endemicity of Chagas' disease, 2006-2009. 29
Figure 1-9: Life cycle stages of *T. cruzi*. 31
Figure 1-10: Anti-parasitic compounds used for treatment of Chagas' disease..... 34
Figure 1-11: Taxonomic tree of *Leishmania spp*. 35
Figure 1-12: Life cycle stages of *Leishmania spp*..... 36
Figure 1-13: Geographic endemicity of cutaneous leishmaniasis, 2012. 38
Figure 1-14: Geographic endemicity of viceral leishmaniasis, 2012. 39
Figure 1-15: Anti-parasitic compounds used for treatment of Leishmaniases..... 40
Figure 1-16: Purine nucleotide metabolism in kinetoplastids. 47
Figure 1-17: Pyrimidine metabolism in kinetoplastids. 48
Figure 1-18: Some medically relevant purine analogues. 50
Figure 1-19: Some medically relevant pyrimidine analogues. 53
Figure 1-20: Some well-characterised leucyl aminopeptidase inhibitors. 58
Figure 2-1: Overall view of adenosine kinase. 64
Figure 2-2: Structure-based alignment of *Tb*AK sequences 1 and 2 with other AKs. 66
Figure 2-3: Scheme of the ordered bi-bi mechanism of AK. 67
Figure 2-4: Schematic view of the substrate binding site of *Tg*AK with protein-substrate interactions. 68
Figure 2-5: Mechanism of γ-phosphoryl transfer in AK. 69
Figure 2-6: Transition state analogues in *Hs*PGK. 71
Figure 2-7: Agarose gel of *Tb*AK amplification. 73
Figure 2-8: Purification of *Tb*AK. 77
Figure 2-9: DLS and ESI-MS analysis of *Tb*AK purification batches. 82

Figure 2-10: *Tb*AK crystals. 83
Figure 2-11: Diffraction patterns of *Tb*AK crystals. 84
Figure 2-12: Overall structure of *Tb*AK. 86
Figure 2-13: Active site of *Tb*AK. 88
Figure 2-14: Electron density and difference maps of *Tb*AK ligands. 89
Figure 2-15: Anion hole motif. 91
Figure 2-16: Comparison of *Tb*AK structures. 93
Figure 2-17: Amino acid substitutions in *Tb*AK-2. 95
Figure 2-18: Cordycepin and derivatives. 96
Figure 3-1: Protomer structures of three deoxyribonucleoside kinases. 99
Figure 3-2: Overall structure of type II TKs. 100
Figure 3-3: Sequence alignment of type II TKs. 103
Figure 3-4: Schematic view of the substrate binding interactions in the *Tm*TK active site. 105
Figure 3-5: Mechanism of γ-phosphate transfers in TK. 106
Figure 3-6: SDS-PAGE of TK expression tests with the baculovirus system. 113
Figure 3-7: Purification of *Lm*TK. 116
Figure 3-8: SEC chromatograms of *Tb*TK purifications. 118
Figure 3-9: Structures of *Tb*TK and *Lm*TK. 125
Figure 3-10: The C-terminal domain of *Tb*TK. 126
Figure 3-11: Electron densities for the ligands bound in the *Lm*TK structures. 128
Figure 3-12: Electron densities for the ligands bound in the *Tb*TK structures. 129
Figure 3-13: TK substrate binding site. 130
Figure 3-14: Schematic representation of the coordination of AP_5dT in the active site of *Lm*TK. 132
Figure 3-15: Superposition of *Lm*TK and *Tb*TK tetramers. 133
Figure 3-16: Comparison of open and closed TK tetramers. 135
Figure 3-17: Conformations of the mainly disordered loop. 137
Figure 3-18: *Lm*TK-dTTP dimer connection to a symmetry-related dimer. 138
Figure 3-19: Model of *Tb*TK pseudo-dimer coloured by sequence identity. 140
Figure 3-20: Comparison of the active sites of *Lm*TK, *Tb*TK and *Hs*TK1. 142
Figure 4-1: Structure of *Pf*LAP. 147

Figure 4-2: Phylogenetic tree of the M17 LAPs. 149
Figure 4-3: Alignment of the TriTryp LAP-As. 151
Figure 4-4: Consensus sequence (> 70 %) of the TriTryps LAPs. 152
Figure 4-5: LAP mechanism. 154
Figure 4-6: LAP-A purification. 160
Figure 4-7: Native gel zymography of recombinant TriTryp LAP-As. 185
Figure 4-8: Validation of anti-*Tb*LAP-A by western blot analysis. 186
Figure 4-9: *T. brucei* growth adaptation 188
Figure 4-10: Western blot analysis of RNAi-mediated LAP-A depletion in bloodstream *T. brucei*. 190
Figure 4-11: Growth and cell cycle analysis of RNAi-mediated LAP-A depletion in bloodstream *T. brucei*. 191
Figure 4-12: Western blot analysis of RNAi-mediated LAP-A depletion and dKO in procyclic *T. brucei*. 192
Figure 4-13: Growth curve and cell cycle analysis of RNAi and dKO in procyclic *T. brucei*. 193
Figure 4-14: Western blot analysis of LAP-A overexpression in *T. brucei*. 194
Figure 4-15: Growth of LAP-A overexpressing *T. brucei* in standard media. 194
Figure 4-16: Growth and cell cycle analysis of LAP-A overexpression in procyclic *T. brucei* in MEM. 195
Figure 4-17: Growth and cell cycle analysis under amino acid limitation. 197
Figure 4-18: Subcellular localisation of LAP-A in *T. brucei*. 199
Figure 4-19: Colocalisation analysis with JACoP using Li's approach. 200
Figure 4-20: LAP-A crystals. 201
Figure 4-21: *Tb*LAP-A. 202
Figure 4-22: *Tb*LAP-A quaternary structure. 203
Figure 4-23: Comparison of *Tc*LAP-A with *Tb*LAP-A. 205
Figure 4-24: Comparison of *Lm*LAP-A with *Tb*LAP-A. 206
Figure 4-25: TriTryp LAP-A hexamers. 207
Figure 4-26: Superposition of the metal binding sites. 208
Figure 4-27: XRF scan of the *Tb*LAP-A-Mn crystal grown in presence of zinc. 209
Figure 4-28: Anomalous difference density in the active site of *Tb*LAP-A. 210

Figure 4-29: Superposition of apo *Tb*LAP-A and to *Tb*LAP-A-Mn-bestatin. 211
Figure 4-30: Electron density of bestatin bound in chain A of *Tb*LAP-A. 212
Figure 4-31: Bestatin binding in the active site. 213
Figure 4-32: Schematic view of bestatin binding in the *Tb*LAP-A active site. 214
Figure 4-33: Electron density of actinonin bound in *Lm*LAP-A. 216
Figure 4-34: Electron density of actinonin bound in *Tb*LAP-A. 217
Figure 4-35: Actinonin binding sites in *Lm*LAP-A and *Tb*LAP-A. 218
Figure 4-36: Schematic view of actinonin binding in the active sites of *Tb*LAP-A and *Lm*LAP-A. 219
Figure 4-37: Citrate binding in *Tc*LAP-A. 220
Figure 4-38: Annotated SEC-MALLS data for *Tb*LAP-A, *Tc*LAP-A and *Lm*LAP-A. 222
Figure 4-39: SEC-MALLS traces of *Tb*LAP-A. 223
Figure 4-40: SEC-MALLS traces of *Tc*LAP-A. 225
Figure 4-41: SEC-MALLS traces of *Lm*LAP-A. 226
Figure A.1-1: Structure of *P. falciparum* TMPK (*Pf*TMPK). 235
Figure A.1-2: Mechanism of human TMPK based on X-ray structures. 236
Figure A.1-3: Sequence alignment of TMPKs. 237
Figure A.1-4: SDS-PAGE of TMPK expression tests in the baculovirus system. 239
Figure A.1-5: *Lm*TMPK purification. 241
Figure A.1-6: Crystals of *Lm*TMPK. 242
Figure A.2-1: Plasmid map of the YSBLIC3C vector 250
Figure A.2-2: Plasmid map (linearised) of the pOPIN vector family 251
Figure A.2-3: Plasmid map of pGRV23b-*Tb*LAP-A 252
Figure A.2-4: Plasmid map of pGR19-*Tb*LAP-A-RNAi 253
Figure A.2-5: Plasmid map of pHD887B-*Tb*LAP-A-UTRs 254
Figure A.2-6: Plasmid map of pHD887H-*Tb*LAP-A-UTRs 254

List of Tables

Table 2-1: PCR reaction mix used for *Tb*AK. 72
Table 2-2: Temperature profile of the PCRs used for *Tb*AK. 72
Table 2-3: Reaction conditions of the T4 reaction. 73
Table 2-4: Reaction mix for colony PCRs. 74
Table 2-5: Temperature profile of the colony PCRs. 74
Table 2-6: Crystallographic data and statistics of the *Tb*AK crystals. 80
Table 3-1: Compositions of the resitriction digests for the pMAL-C2 constructs. 109
Table 3-2: Composition of T4 ligase dependent ligations for pMAL-C2 vectors. 109
Table 3-3: PCR condition of TK constructs for baculovirus expression. 110
Table 3-4: Temperature profile for PCRs of the baculovirus expression constructs. 110
Table 3-5: TK constructs cloned for expression with the baculovirus system. 112
Table 3-6: Crystallisation conditions of *Lm*TK and *Tb*TK. 119
Table 3-7: Crystallographic data and statistics for the *Lm*TK crystals. 121
Table 3-8: Crystallographic data and statistics for the *Lm*TK crystals 121
Table 3-9: Crystallographic data and statistics for the *Lm*TK crystals 122
Table 4-1: The TriTryp LAPs. 148
Table 4-2: Condition of control digests used to confirm cloning success. 158
Table 4-3: Compositions of the resitriction digests for the pGRV23b construct. 164
Table 4-4: Composition of the ligations for *Tb*LAP-A constructs. 164
Table 4-5: Conditions of Hind3 digests for first RNAi fragment insertion. 165
Table 4-6: Conditions of Apa1 digests for first RNAi fragment insertion. 165
Table 4-7: Conditions of restriction digests for second RNAi fragment insertion. 166
Table 4-8: Conditions of restriction digests for 5'-UTR fragment insertion. 167
Table 4-9: Composition of ligations for 5'-UTR construct. 167
Table 4-10: Conditions of restriction digests for 3'-UTR fragment insertion. 168
Table 4-11: Conditions of restriction digests for antibiotic resistance exchange. 169
Table 4-12: Conditions used to confirm HYG orientation. 169
Table 4-13: Selection drugs used for the different *T. brucei* parasite strains. 171
Table 4-14: Concentrations and screens used for crystallisation screening. 176

Table 4-15: Crystallisation conditions of the LAP-As and ligands bound..... 177
Table 4-16: Crystallographic data and statistics for *Tb*LAP-A crystals. 179
Table 4-17: Crystallographic data and statistics for *Tb*LAP-A-complex crystals. 180
Table 4-18: Crystallographic data and statistics for the *Tc*LAP-A crystals. 181
Table 4-19: Crystallographic data and statistics for the *Lm*LAP-A-Mn-actinonin. ... 182
Table 4-20: Expected molecular masses for different LAP-A oligomers..... 221
Table 4-21: M_w of the *Tb*LAP-A oligomers obtained from SEC-MALLS. 224
Table 4-22: M_w of the *Tc*LAP-A oligomers obtained from SEC-MALLS. 225
Table 4-23: M_w of the *Lm*LAP-A oligomers obtained from SEC-MALLS. 226
Table A.1-1: TMPK constructs for expression with the baculovirus system. 238
Table A.2-1: Primer sequences for the kinases for *E. coli* expression vectors..... 238
Table A.2-2: Primer sequences for the TriTyp LAP-As. 238
Table A.2-3: List of constructs for protein production. 238
Table A.2-4: List of constructs for protein production, continued. 238
Table A.2-5: List of constructs for parasitological work..... 238
Table A.3-1: Media composition. 254
Table A.3-2: Media composition, continued..... 255
Table A.3-3: Composition of "Non-essential amino acids". 238
Table A.3-4: Composition of "MEM Vitamins"..... 238

Acknowledgments

I would like to thank my supervisors Keith S. Wilson and Anthony J. Wilkinson for the opportunity to carry out my PhD, for their kind support and guidance and for always being available when I needed them. Additionally, I would like to thank Glyn, for guiding me through the lab in my first few months and for the continuous help, advice and patience.

I would like to thank Dolores Gonzales-Pacanowska for giving me the opportunity to carry out part of my work in her lab. Thanks to the members of Lab 103/104, for all their help and advice and making my stay in Granada so pleasant.

I thank Ray Owens and the members of OPPF, especially Louise and Heather, for their help and advice and for introducing me to the high-throughput way of protein production.

Very special thanks to the members of YSBL for all the support, advice and helpful attitude, for all the good laughs and for making working here so extremely enjoyable!

Finally, I would like to thank my friends and family for their continuous support.

Author's Declaration

I declare that I am the sole author of this thesis. The work presented here was neither published before nor used previously to obtain a degree at this or another university with the exception of the Master of Chemistry (MChem) obtained by Michael Norris using some of the SEC-MALLS data in Chapter 5.

I carried out all of the work with the exception of the tasks listed below and specifically mentioned in the text.

- Cloning and initial expression tests of thymidine kinase (TK) constructs in YSBLIC and YSBLIC3C vectors, mentioned in Chapter 4, were carried out by Dr. Glyn Hemsworth.
- Completion of baculovirus generation and expression of TK and TMPKs with the baculovirus system mentioned in Chapter 4 and the Appendix was performed by Heather Rada and Joanne Nettleship at the Oxford protein production facility (OPPF).
- Kinetic measurements of *Leishmania major* TK discussed in Chapter 4 were carried out by Cristina Bosch at Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain.
- Cloning of leucyl aminopeptidase from *Trypanosoma brucei* (*Tb*LAP-A) and *Leishmania major* (*Lm*LAP-A) into the YSBLIC vector, mentioned in Chapter 5, were carried out by Dr. Javier Garcia Nafria and Nam Duong (a project student of the Minority Health and Health Disparities International Research Training (MHIRT) program), respectively.
- For generation of anti-*Tb*LAP-A-specific antibody in Chapter 5, the injections into the rabbit and the extraction of the blood were done in collaboration with the staff of the Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain.
- The data of ESI-MS mentioned in Chapter 3 and SEC-MALLS in Chapter 5 were collected by the staff of the Technology Facility at the University of York.

- The sample preparation of *Tc*LAP-A and *Lm*LAP-A for SEC-MALLS and part of the data analysis in Chapter 5 were carried out by Michael Norris, a MChem undergraduate project student.

Figures reproduced or adapted from external sources are identified in the captions and copyright permission was obtained.

Part of this work has been published or submitted for publication. The published manuscript is attached at the end of this thesis.

Published:

Timm J., Gonzalez-Pacanowska D. & Wilson K. S. (2014), Structures of adenosine kinase from *Trypanosoma brucei brucei*. Acta Cryst. F70, 34-39.

Submitted for publication (in PlosOne):

Timm J., Bosch-Navarette C., Recio E., Nettleship J.E., Rada H., Gonzalez-Pacanowska D. & Wilson K. S., Structural and kinetic characterisation of thymidine kinase from *Leishmania major*.

Chapter 1 Introduction

1.1 Neglected tropical diseases

Neglected tropical diseases (NTDs) are a medically diverse group mainly affecting people in the low-income populations in the developing world. They have been largely overlooked partly due to general disregard of the developing world and focus on the big three infectious diseases: HIV/AIDS, tuberculosis and malaria (Feasey *et al.*, 2010). NTDs defined by the World Health Organisation (WHO) include devastating diseases caused by a variety of different pathogens, including viruses (e.g. dengue/severe dengue and rabies), bacteria (e.g. buruli ulcer, leprosy, trachoma and yaws), protozoa (Chagas' disease, Human African trypanosomiasis and Leishmaniases) and helminths (e.g. cysticercosis/taeniasis, dracunculiasis, echinococcosis, foodborne trematodiases, lymphatic filariasis, onchocerciasis, schistosomiasis and soil-transmitted helminthiases). Recently, the problem of NTDs has been addressed more extensively, e.g. by foundation of a specific department within the World Health Organisation (World Health Organisation, 2014b), the Global Network for Neglected Tropical Diseases Control at the Sabin Vaccine institute (http://www.globalnetwork.org) and the open access journal PLoS Neglected Tropical Diseases (http://www.plosntds.org).

1.2 Kinetoplastids and the TriTryps

Kinetoplastids are unicellular flagellate protozoa of the class kinetoplastida. They have either one or two flagella and therefore are divided into two groups, the uniflagellate trypanosomatida and the biflagellate bodonina. Some kinetoplastids are parasitic, targeting humans, animals or plants, while others are free-living, feeding on bacteria. Alongside the universal eukaryotic organelles, e.g. nucleus, mitochondrion, endoplasmic reticulum and Golgi apparatus and flagella, they also show many unique morphological features (Figure 1-1).

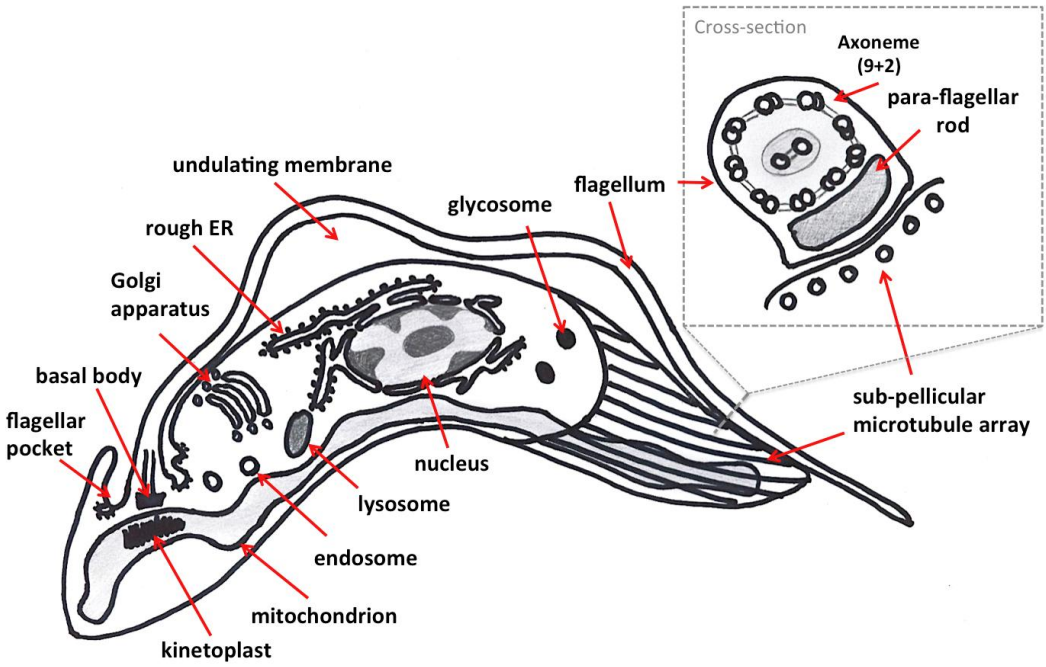


Figure 1-1: Schematic representation of kinetoplastid cell morphology.

One of the features is the kinetoplast, which contains the extensive genome of the single mitochondrion in the form of circular DNA molecules, structural proteins and polymerases. It is situated at the base of the flagellum associated with the flagella's basal body. The flagellum, containing a structure called the para-flagellar rod, starts from the flagellar pocket, the only site of endo- and exocytosis, and is sometimes attached to the cell body forming an undulating membrane. Kinetoplastids have only a single disc-shaped lysosome instead of the multiple lysosomes seen in other eukaryotes, and glycolysis takes place in specialised organelles called glycosomes. The cytoskeleton comprises mainly microtubules forming a regular structure under the cell's surface, the sub-pellicular microtubule array.

Parasitic kinetoplastids often have complex life cycles involving one or more hosts and show several different morphological forms, depending on their life cycle stage, e.g. trypomastigotes, epimastigotes, promastigotes and amastigotes (Figure 1-2). The different morphological forms can be distinguished by the length and attachment of the flagellum and by the position of the kinetoplast relative to the nucleus. In addition, these forms have differences in their metabolism adapting to their environment.

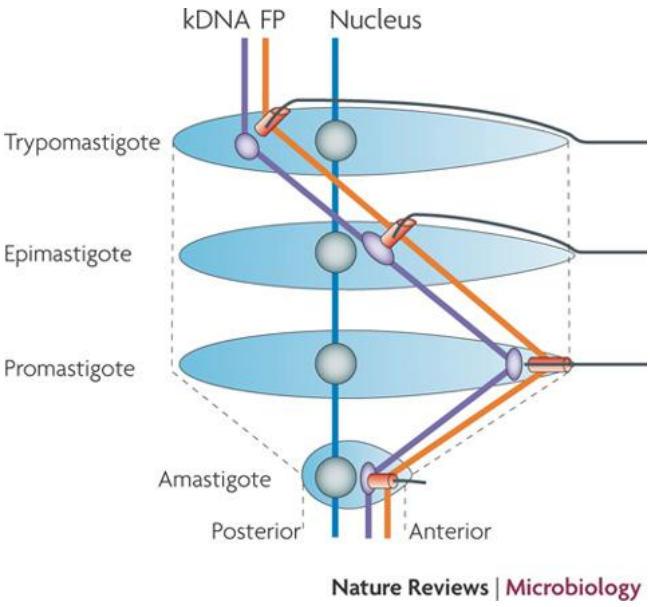


Figure 1-2: Positioning of nucleus, kinetoplast and flagellum.

Shown are the organelle positions in the different morphological forms during the kinetoplast life cycle. kDNA = kinetoplast in purple, FP = flagellar pocket in orange. Figure from (Field & Carrington, 2009).

The class kinetoplastida contains parasitic species causing devastating diseases in humans. The term "TriTryps" describes the human infectious kinetoplastida *Trypanosoma brucei*, causative agent of African sleeping sickness, *T. cruzi*, responsible for Chagas' disease, and *Leishmania spp.*, causing leishmaniases. Genome sequences, information on initial proteomic studies and some biological data can be found in a specialised database, TriTrypDB (Aslett *et al.*, 2010).

1.3 *Trypanosoma brucei*

T. brucei has the typical morphological features defining the kinetoplastida. There are three subspecies, *T. b. brucei*, *T. b. gambiense* and *T. b. rhodesiense,* which are morphologically almost indistinguishable, but vector preferences, drug sensitivity and the position of the nucleus in some *T. b. rhodesiense* cells (further towards the posterior end) made it possible to differentiate *T. b. gambiense* and *T. b. rhodesiense* (Gibson, 1986). Except for human infectivity, distinguishing *T. b. brucei* and *T. b. rhodesiense* is more difficult because of their similarity in endemicity and animal hosts and the high level of genetic heterogeneity in

T. b. rhodesiense from different geographical areas. The geographical distribution of the *T. brucei* subspecies is distinct and determined by their preferred insect vector species. *T. b. gambiense* and its main vector *Glossina palpalis* are endemic in western and central Africa (Figure 1-3), close to rivers and lakes. It is a mainly anthroponotic parasite with little involvement of animal vectors. *T. b. rhodesiense* and its main vector *G. morsitans* on the other hand are endemic in east Africa (Figure 1-4), preferentially in dry bush and woodland. Unlike *T. b. gambiense* it is a zoonotic disease, mainly infecting wild and domestic animals and being transmitted from animals to humans. *T. b. brucei* shares its main insect vector and therefore endemic area with *T. b. rhodesiense*.

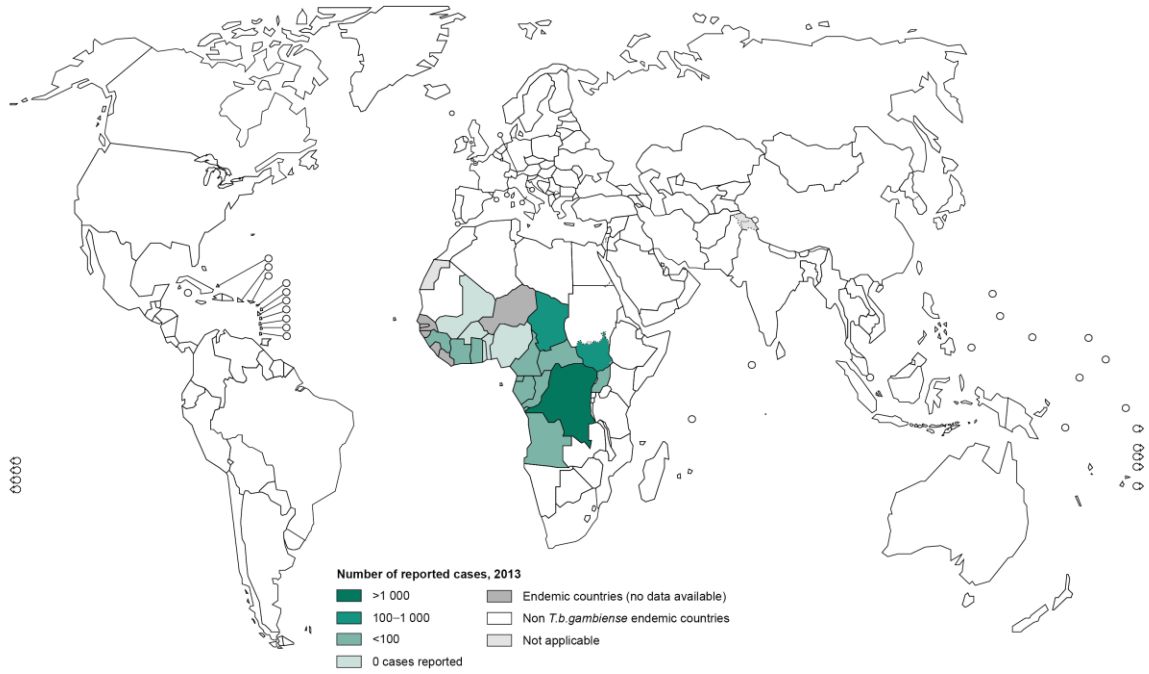


Figure 1-3: Geographic endemicity of HAT caused by *T. brucei gambiense*, 2013.
The colour (green) indicating the number of cases reported in 2013. Figure modified from WHO (World Health Organisation, 2013).

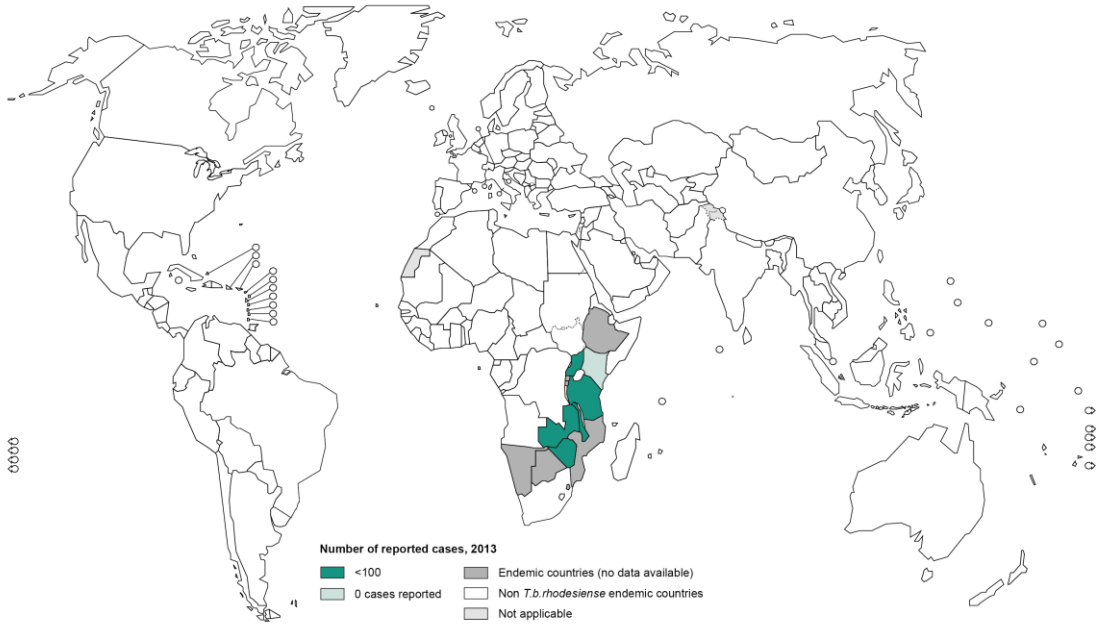


Figure 1-4: Geographic endemicity of HAT caused by *T. brucei rhodesiense,* 2013.
The colour (green) indicating the number of cases reported in 2013. Figure modified from WHO (World Health Organisation, 2013).

While *T. b. brucei* cannot infect humans because of the presence of the serum complexes TLF-1 and TLF-2 (named from **t**rypanosome **l**ytic **f**actors) containing the trypanolytic protein APOL1, the other two subspecies developed resistance to APOL1/TLF becoming human infective (Raper *et al.*, 2001). In *T. b. rhodesiense* a single gene, the human-serum-resistance-associated (SRA) gene, which encodes a truncated version of a variant surface glycoprotein, confers resistance (De Greef & Hamers, 1994, Xong *et al.*, 1998). Genetic exchange can occur between *T. b. rhodesiense* and *T. b. brucei* (Gibson, 1989) making it possible that the SRA gene may be transferred to other not-yet resistant *T. b. brucei* resulting in *T. b. rhodesiense* with other genetic backgrounds. This supports the definition of *T. b. rhodesiense* as a simple host-range variant of *T. b. brucei* (Tait *et al.*, 1985). In *T. b. gambiense* multiple factors play a role in resistance. First, a gene called *T. gambiense*-specific glycoprotein (TgsGP) mediates the APOL1/TLF resistance (Uzureau *et al.*, 2013), and second, differences in the TLF-1 receptor TbHpHbR in *T. b. gambiense* were found to be significant (Vanhollebeke *et al.*, 2008).

1.3.1 *T. brucei* cell cycle

As in other eukaryotes, the cell cycle of *T. brucei,* Figure 1-5 (reviewed in (Ooi & Bastin, 2013), is divided into G_1/G_0-, S-, M- and G_2-phases and DNA replication, segregation of the organelles and cytokinesis are taking place at specific time points.

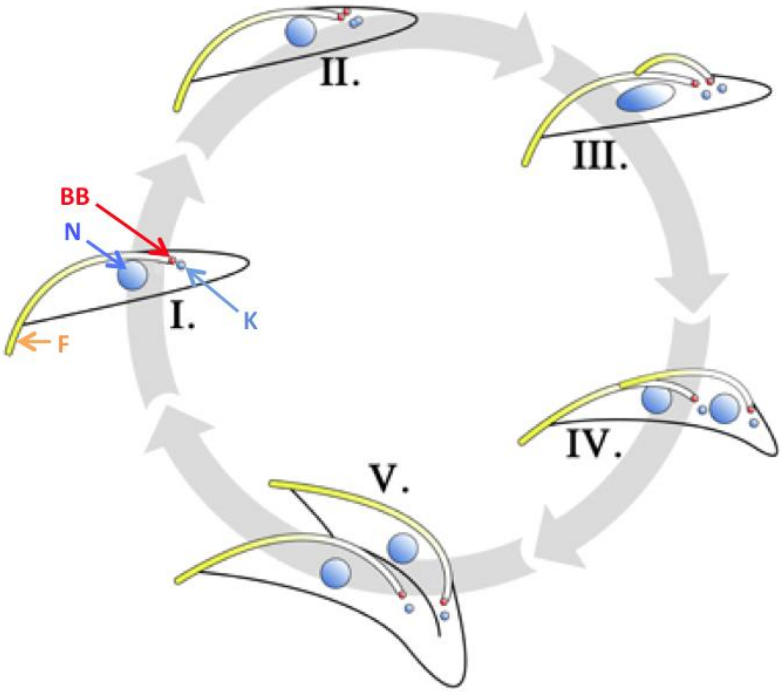


Figure 1-5: Cell cycle of procyclic *T. brucei*.
(I.) G_1/G_0-phase with 1N1K(1F). (II.) S-phase with 1N1K(1F). (III.) G_2-phase with 1N2K(2F). (IV.) M-phase with 2N2K(2F). (V.) Cytokinesis. N = nucleus, K = kinetoplast, F = flaggelum, BB = basal body. More details in the main text. Figure modified from (Ooi & Bastin, 2013).

Single cells are in G_1-phase (Figure 1-5 I.), called G_0 in cell cycle arrested cells, which is characterised by the presence of one nucleus (N), with one copy of the genome, one kinetoplast (K) and one flagellum (F) and therefore referred to as 1N1K(1F). Cells enter the S-phase (Figure 1-5 II.) with the initiation of DNA replication in the nucleus and the kinetoplast, both changing shape as a result of the increased DNA content. The basal body (BB) of the flagellum matures and forms a daughter basal body during this phase. The end of the S-phase and beginning of G_2-phase (Figure 1-5 III.) are marked by the presence of one nucleus, with more than one copy of the

genome, two kinetoplasts and two flagella attached to one another, referred to as 1N2K(2F). When DNA replication and nuclear segregation are complete, the cells enter the M-phase (Figure 1-5 IV.), with two nuclei, two kinetoplasts and two flagella, 2N2K(2F). This is followed by cytokinesis starting from the anterior end of the cell following the longitudinal axis (Figure 1-5 V.) and ultimately resulting in two daughter cells.

1.3.2 *T. brucei* life cycle

T. brucei has a complex life cycle involving differentiation processes within its human host and the tsetse fly (*Glossina spp*.) vector (Figure 1-6, reviewed in (Vickerman, 1985, Ooi & Bastin, 2013). All three subspecies show the same life cycle stages.

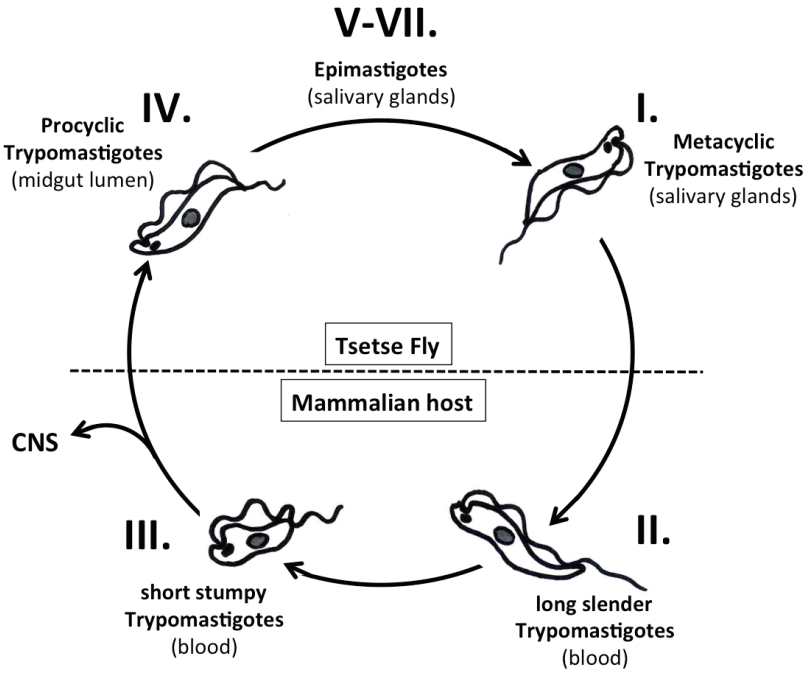


Figure 1-6: Life cycle stages of *T. brucei*.

I. Metacyclic trypomastigotes, II. Long slender trypomastigotes (= bloodstream form), III. Short stumpy trypomastigotes, IV. Procyclic trypomastigotes, V-VII. Epimastigote forms. More details in the main text.

At the moment of infection by the bite of the tsetse fly, *T. brucei* parasites are in their metacyclic trypomastigote form (Figure 1-6 I.), showing the classic

trypomastigote morphology with the nucleus in the centre of the cell, the kinetoplast at the posterior end and the flagellum starting at the posterior end being attached along the cell body forming an undulating membrane. As they are extracellular in the bloodstream (Figure 1-6 II.), exposed to the mammalian host's immune system, they undergo antigenic variation in the form of expression of one of the many variable surface glycoproteins (VSGs) encoded in their genome. The blood contains high levels of glucose, so the parasites' metabolism is fuelled mainly by glycolysis and no oxidative phosphorylation within the mitochondrion occurs. Therefore, the mitochondrion shows only little electron transport activity and lacks the typical cristae. While in the bloodstream the parasites are in a long slender form and replicate asexually by binary fission along the long axis of the cell. They differentiate into short, stumpy parasites (Figure 1-6 III.), which do not replicate and reactivate the mitochondria in preparation for the transfer into the tsetse fly. Some of the parasites cross the blood brain barrier, which marks the second stage of the disease. The remaining parasites in the blood stream are ready to be taken up by a tsetse fly.

When a tsetse fly ingests a blood meal, the bloodstream trypomastigotes within the fly's midgut differentiate into procyclic trypomastigotes (Figure 1-6 IV.). These parasites do not express VSGs and due to the low glucose levels within the fly's midgut, the parasites switch to an amino acid-based metabolism, showing high levels of oxidative phosphorylation within the mitochondrion. The main energy source at that stage is proline. Cristae develop in the mitochondrion and its electron transport activity is high. Procyclic parasites have a long, slender shape and replicate by binary fission. These parasites penetrate the ectoperitrophic space changing morphologically to become longer and the mitochondrion regresses.

Reaching the salivary glands, probably by following the digestive system backwards, the parasites differentiate into epimastigotes (Figure 1-6 V.), attaching to the epithelial cells *via* their flagella, which are starting from the middle of the cell, forming just a short undulating membrane. The nucleus is centrally located and the kinetoplast is closer, usually anterior, to the nucleus. Epimastigotes replicate further and eventually differentiate, *via* two intermediate stages (Figure 1-6 VI-VII.), into

mature metacyclic trypomastigotes. These metacyclics are not attached to the epithelium, show VSG expression and the mitochondrion is acristate. During the next blood meal, the tsetse fly transfers these parasites to the new host.

While in the tsetse fly *T. brucei* can undergo meiotic stages, exchanging genetic information between different parasites (Peacock *et al.*, 2008, Peacock *et al.*, 2011).

In the laboratory two *T. brucei* forms are cultured, the slender metacyclic trypomastigote (Figure 1-6 II.), often referred to as bloodstream form, and the procyclic trypomastigote form (Figure 1-6 IV.). Different laboratory strains are used, some of which can easily differentiate, while others do not.

1.3.3 Antigenic variation

T. brucei is a parasite residing in the blood stream of its mammalian host, which exposes it to the host's immune system. The immune system raises antibodies against the parasites, leading to their elimination. This results in strong fluctuation of parasitaemia during the course of an infection and the parasites in successive peaks display variant antigenic types (VATs). These VATs are the result of the expression of variant surface glycoproteins (VSGs), covering the parasite's surface as a thick coat and therefore providing the only antigen exposed to the host's immune system (Horn & McCulloch, 2010). The VSGs have a highly variable N-terminal domain, which covers the conserved C-terminal domain, and a GPI anchor attaches it to the parasite's plasma membrane.

The genome of *T. brucei* encodes for more than 1000 different VSGs, located mainly in subtelomeric regions, but also close to the telomeres of minichromosomes and very few at expression sites. Only one of these VSGs is expressed at a time by an individual parasite, so the host's immune system only raises antibodies against that one. The VSG expression is switched to another gene by *in situ* activation, telomere exchange or gene conversion. *In situ* activation occurs mainly in the beginning of an infection with the expression site of the original VSG being switched off, while a new one is switched on. In telomere exchange the telomere of the actively

expressed VSG is exchanged for an inactive one or one of the minichromosome and in case of the gene conversion an inactive VSG is copied into the site of the active VSG replacing it. Additional to the VSG genes, the *T. brucei* genome encodes for pseudogenes of VSGs containing frame shifts errors and in frame stop-codons. These sites undergo recombination resulting in novel and functional VSGs, which can be activated by gene conversion.

Switching expression of the VSG hides the parasites from the host's immune response developed against the previous VSG and the parasites with the new VSG multiply until the immune system raises a response against that one.

1.3.4 Human African Trypanosomiasis (HAT)

HAT, also called African sleeping sickness, is caused by *T. brucei* parasites and is ultimately fatal if left untreated. The disease was nearly eliminated by the late 1960s through systematic screening, treatment, patient follow-up and partial vector control, but negligence in subsequent years led to re-emergence of HAT in the 1980s (Simarro, Franco, Cecchi*, et al.*, 2012). After a big increase in the number of cases during the 1990s, the number of infections has been reduced over the past 20 years and the WHO estimates approximately 30 000 current infections with 7139 new cases reported in 2010 (World Health Organisation, 2014a). In addition to being a threat to human populations in endemic areas, *T. brucei* together with *T. vivax* and *T. congolense*, have big impacts on agriculture, causing a disease called nagana in cattle and other domestic and wild animals.

The clinical course of HAT is divided in two stages. During the first, haemolymphatic, stage, the parasites are found in the peripheral circulation, causing fever, strong headache, fatigue, aching muscles and joints, irritability and swollen lymph nodes among other symptoms (Centers for Disease Control and Prevention, 2014b). The symptoms are very similar to malaria and often misdiagnosed as such. In travellers the symptoms differ from people from endemic countries (Brun *et al.*, 2010). Travellers suffer high fever, severe haematological disorders, impaired kidney function, electrolyte disturbances, high concentrations of C-reactive protein and

liver enzymes, and a chancre at the site of the infective bite and trypanosomal rash are seen more frequently. In the second, or encephalitic, stage the parasites cross the blood brain barrier, entering the central nervous system, and patients suffer progressive confusion, changes in personality and other neurologic disorders, like alteration of the sleeping pattern with sleepiness in daytime and sleep disturbance at night. Depending on the subspecies the rate of disease onset and progression differs, but generally the symptoms are very similar for both subspecies. Coma and death will follow, if the disease remains untreated.

T. brucei rhodesiense is the causative agent of the acute form of HAT, which progresses rapidly. First symptoms occur within 1-3 weeks after the infective bite, the symptoms are severe and death follows months later, unless the disease is treated. These parasites, while mainly human infective, also infect domestic and wild animals, which serve as important reservoirs.

T. brucei gambiense causes the chronic form of HAT, with initially mild symptoms and which progresses slowly. First symptoms can occur several weeks or months after the infective bite and involvement of the central nervous system becomes apparent after 1-2 years. In most cases, the untreated disease kills within 3 years, but in some it can take up to 6-7 years before death. This parasite is anthroponotic and domestic and wild animals are rarely infected. In this case infective humans are the main reservoir.

1.3.5 Diagnosis and Current treatment of HAT

The clinical manifestations of HAT, especially during the first stage, are not sufficiently specific for diagnosis, so laboratory methods have to be applied. Most diagnoses are based on the detection of the parasites in the blood or lymph node aspirates *via* microscopic techniques (Centers for Disease Control and Prevention, 2014b). To address low parasitaemia in *T. brucei gambiense* infections and increase the chance of finding the parasite in the blood samples, blood concentration techniques are used. *T. brucei rhodesiense* infections show higher parasitaemia and detection of parasites is easier. Serological tests, e.g. the card agglutination test for

trypanosomiasis (CATT), immunofluorescence and enzyme-linked immunosorbent assays (ELISAs) are available for detection of *T. brucei gambiense*, but not for *T. brucei rhodesiense* (Brun *et al.*, 2010). None of the diagnostic techniques is 100 % sensitive, which requires physicians to pay close attention to clinical symptoms and the possibility of exposure.

After detection of parasites in the patient, the disease stage needs to be determined. For this the cerebro-spinal fluid is examined for the presence of parasites or more than five white blood cells per µL. Parasite invasion of the central nervous system increases the complications and different drugs have to be used for the treatment of second stage HAT compared to the first stage.

Drugs currently available for treatment of HAT (Figure 1-7) are highly toxic inducing a variety of severe side effects. Additionally, limited efficiency, high costs, long and complicated administration regimens and inadequate availability worsens the situation. The mechanism of action (MOA) of some of these drugs is poorly understood.

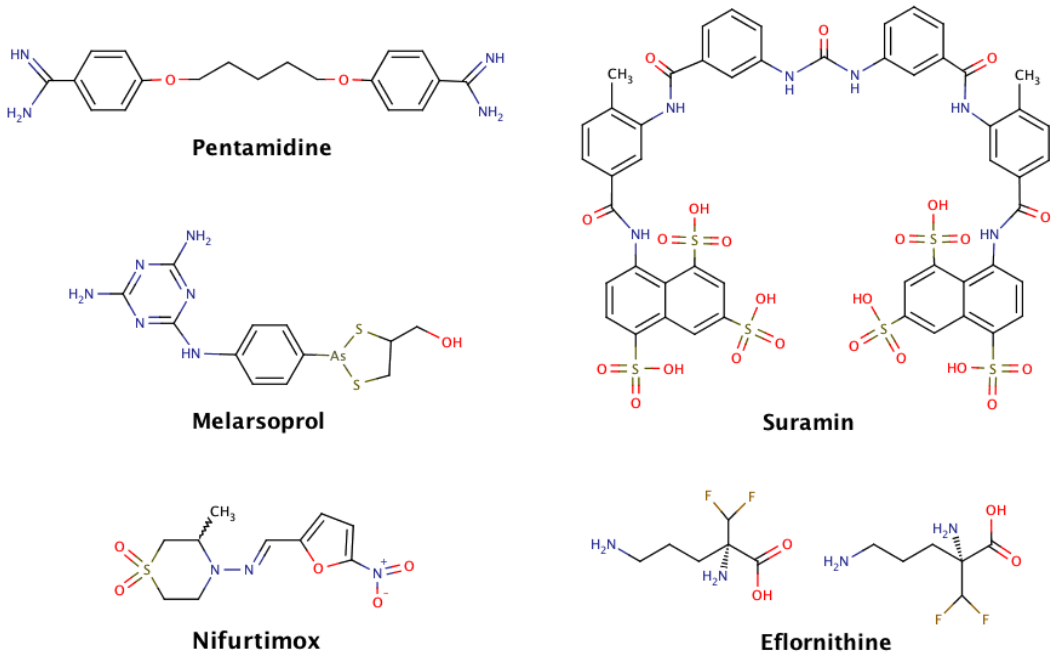


Figure 1-7: Anti-parasitic compounds used for treatment of HAT.

Details in the main text. Figure made with MarvinSketch (ChemAxon).

First stage HAT is treated with one of two drugs, either pentamidine or suramin. Treatment of second stage HAT is more complicated as the drugs are required to cross the blood brain barrier. The available drugs for second stage HAT are melarsoprol and eflornithine. Combinational therapy of eflornithine with nifurtimox, a drug registered for treatment of Chagas' disease, shows an increased efficiency and fewer side effects compared to eflornithine monotherapy (Simarro, Franco, Diarra*, et al.*, 2012).

1.4 *Trypanosoma cruzi*

T. cruzi, the causative agent of Chagas' disease in humans, is endemic in Central and South America (Figure 1-8).

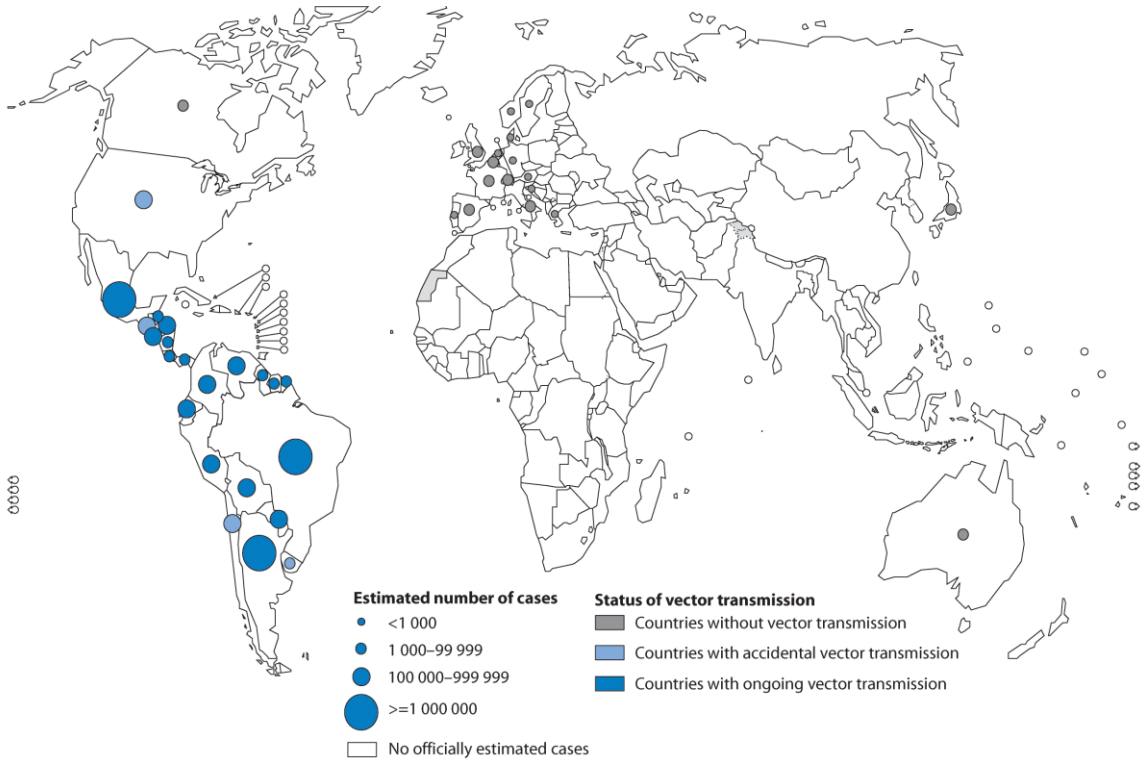


Figure 1-8: Geographic endemicity of Chagas' disease, 2006-2009.

The size of the circle indicating the number of cases reported, blue indicating endemicity. The data are based on official estimates and status of vector transmission by the WHO (World Health Organisation, 2013).

The *T. cruzi* genome shows a high degree of diversity and multiplicity. There are six discrete typing units (DTU), called TcI-TcVI, which show different geographical and environmental distributions (reviewed in (Zingales *et al.*, 2012). *T. cruzi* parasites are transmitted to animals and humans mostly through their insect vector, the triatomine or reduviid bug, but also congenital, through blood transfusions, organ transplants or through contaminated food.

Depending on the vector habitat two kinds of transmission cycles are distinguished, the sylvatic and the domestic. In the sylvatic transmission cycle *T. cruzi* rarely encounters human hosts and mainly infects animals, while the domestic transmission cycle occurs when the insect vector lives in close proximity to human habitation and therefore the rate of infection in humans is much higher. The transmission through the triatomine bug is very inefficient, but the bugs often live in poor housing with adobe walls, thatched roofs and animal stalls close to the house, so the ineffective transmission is compensated by the frequency of the encounter.

All *T. cruzi* DTUs are human infective. TcI is by far the most abundant and genetically diverse and uses both sylvatic and domestic transmission cycles. While TcIII and TcIV infections are rare due to mainly sylvatic transmission cycles, human infections with TcII, TcV and TcVI are more common because of their preferentially domestic transmission cycles. Co-infection with multiple DTUs is common, which provides the opportunity of hybridisation events to occur, as was the case for TcII and TcIII, which gave rise to the hybrids TcV and TcVI. The commonly used laboratory strains Sylvio X10/1 and CL Brener belong to the DTUs TcI and TcVI respectively. *In vivo* and *in vitro* the behaviour of different DTU's varies drastically and even within the DTUs differences occur, which makes investigations more complicated.

1.4.1 *T. cruzi* life cycle

T. cruzi's life cycle (Figure 1-9) differs significantly from that of *T. brucei*, as reviewed in (Vickerman, 1985), e.g. no mitochondrial cycle or switch in metabolism is seen in *T. cruzi* and the main replicative stage takes place inside a host cell.

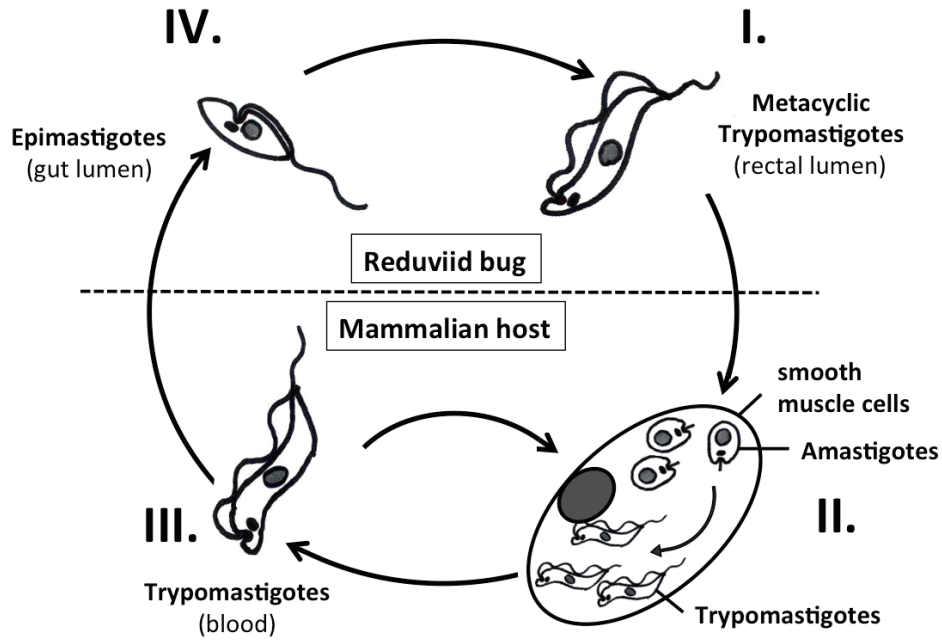


Figure 1-9: Life cycle stages of *T. cruzi*.

I. Metacyclic trypomastigotes, II. Intracellular stage of T. cruzi with amastigotes and trypomastigotes, III. Trypomastigotes residing the bloodstream, IV. Epimastigotes. More details in the main text.

When the infected triatomine bug takes a blood meal, it defecates. Infective metacyclic trypomastigote *T. cruzi* parasites (Figure 1-9 I.) in the faeces enter the human or animal host through the bite wound, mucous membranes or hair follicles. The trypomastigotes show the same general morphology as *T. brucei*. While in the bloodstream they do not divide, but invade macrophages. They enter the host cells *via* recruited lysosomes forming parasitophorous vacuoles, which they disrupt to gain access to the cytoplasm (Figure 1-9 II.). There they differentiate into the amastigote form. The amastigote parasites are intracellular in the cytoplasm and not motile with very short, buried flagella. Their kinetoplast is close to the nucleus. In this form the parasites replicate by binary fission and after several rounds of

division they fill up the host cell and differentiate back to trypomastigotes. These trypomastigotes are released into the bloodstream (Figure 1-9 III.) and infect other cells in target tissues, like heart, smooth muscle of the digestive system liver, spleen and lymph nodes, and repeat the differentiation cycle until the infection is cured. A triatomine bug taking a blood meal takes up trypomastigotes circulating in the bloodstream.

Within the midgut of the triatomine bug the parasites differentiate into epimastigotes (Figure 1-9 IV.). This form of the parasite is characterised by a centrally located kinetoplast anterior of the nucleus with the flagellum emerging from the middle of the cell forming a short undulating membrane. These parasites are less motile than trypomastigotes. They undergo binary fission for replication while they pass the digestive system and attach to the epithelium of the rectal gland of the insect vector. There they differentiate into mature metacyclic trypomastigotes, detaching from the epithelium, prepared for transmission to a mammalian host.

1.4.2 Chagas' Disease

Chagas' disease, named after the Brazilian physician Carlos Chagas, is caused by *T. cruzi* parasites. An estimated 8 million people in Central and South Americas are infected, most of whom are not aware of their infection. Globalisation and travel expose even more people to this disease. The clinical course of Chagas' disease has an acute and a chronic stage. Most of the clinical information is based on children infected through vector-borne transmission, but severity and course of the disease can vary with the *T. cruzi* strain involved, age and life stage of the patient and way of infection (Centers for Disease Control and Prevention, 2014c).

The acute phase, immediately after the infection, showing high parasitaemia, can last several weeks or months. This phase can be entirely a- or oligo-symptomatic and the symptoms vary greatly. Symptoms include initial skin chancre at the site of infection or unilateral purplish orbital oedema (Romaña's sign), which can be accompanied by fever, headache, cough, rash, abdominal pain, myocarditis or many

other symptoms (Centers for Disease Control and Prevention, 2014c, World Health Organisation, 2014c). The initial symptoms usually fade even if untreated after a few weeks or months, but the parasite infection remains. In young children, elderly or immunocompromised people the symptoms can be severe and in some cases fatal.

The chronic phase, after the initial symptoms fade away, can last decades or life-long, unless diagnosed and treated. During this phase, the parasites are hidden in target tissues, mostly heart or the smooth muscle of the digestive system. In most patients the chronic phase is asymptomatic. However, about 30 % of the patients develop problems, often decades after the infection, like cardiac complications, with cardiomyopathy, heart failure or cardiac arrest, or intestinal complications, with enlarged oesophagus or colon.

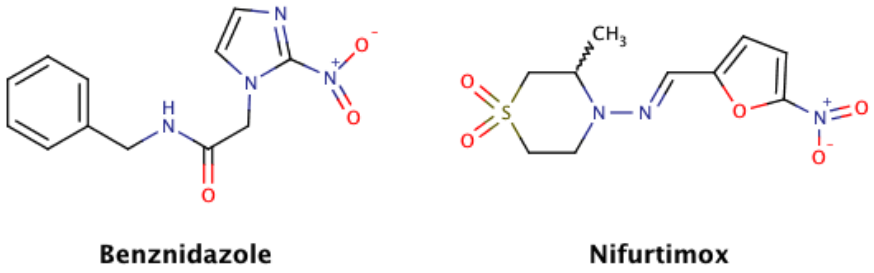
Reactivation of Chagas' infection can happen decades after the original infection, often due to immunosuppression, e.g. caused by HIV infection or use of immunosuppressants after transplantations.

1.4.3 Diagnosis and Current treatment of Chagas' disease

During the acute phase, parasitaemia is high and the parasites can be identified by microscopic examination of blood wet smears or a blood concentration technique, e.g. microhaematocrit (World Health Organisation, 2014c). During the chronic phase diagnosis is more difficult, especially if the disease is asymptomatic. Usually two independent methods are used to confirm infection. Serological tests are used, identifying *T. cruzi* antibodies, e.g. enzyme-linked immunosorbent assay (ELISA), indirect hemagglutination assay, indirect immunofluorescence assay, western blot or other tests. A common method with varying sensitivity is xenodiagnosis, where the infection of the insect vector after feeding on the patient's blood is detected.

There are two approaches for treatment, symptomatic and anti-parasitic (World Health Organisation, 2014c, Centers for Disease Control and Prevention, 2014c). The symptomatic treatment is especially important in the acute phase and in cases

of cardiac and intestinal complications. Anti-parasitic treatment to remove the parasites from the patient is effective in the early stages of infection, but effectivity is low during the chronic phase (Marin-Neto *et al.*, 2008, Issa & Bocchi, 2010). The two available drugs for treatment of Chagas' disease are benznidazole and nifurtimox (Figure 1-10).



Benznidazole **Nifurtimox**

Figure 1-10: Anti-parasitic compounds used for treatment of Chagas' disease. Details in main text. Figure made with MarvinSketch (ChemAxon).

Both drugs show strong toxicity and limited effectivity during the chronic phase of the disease. Strong side effects lead to very poor patient compliance (Viotti *et al.*, 2009). The mechanism of action is not well characterised and suspected to be related to production of radicals.

In general, effectiveness of treatment is difficult to assess and strongly dependent on the *T. cruzi* strain, life-cycle stage of the parasite and time after the primary infection. Also, evaluation of cure is problematic as parasitaemia is very low during the chronic phase and the parasites not only reside in the blood, but also in tissues, which make their detection and elimination challenging.

1.5 *Leishmania spp.*

Leishmania spp. is a very diverse family among the order Trypanosomatida, class Kinetoplastida (Figure 1-11) and the exact taxonomic positions of some are still under discussion. Currently 31 species are assigned to the genus, divided in groups (World Health Organisation, 2014d). The diversity of *Leishmania* species is further

complicated by the parasites' capacity for genetic exchange within the insect vector (Akopyants *et al.*, 2009).

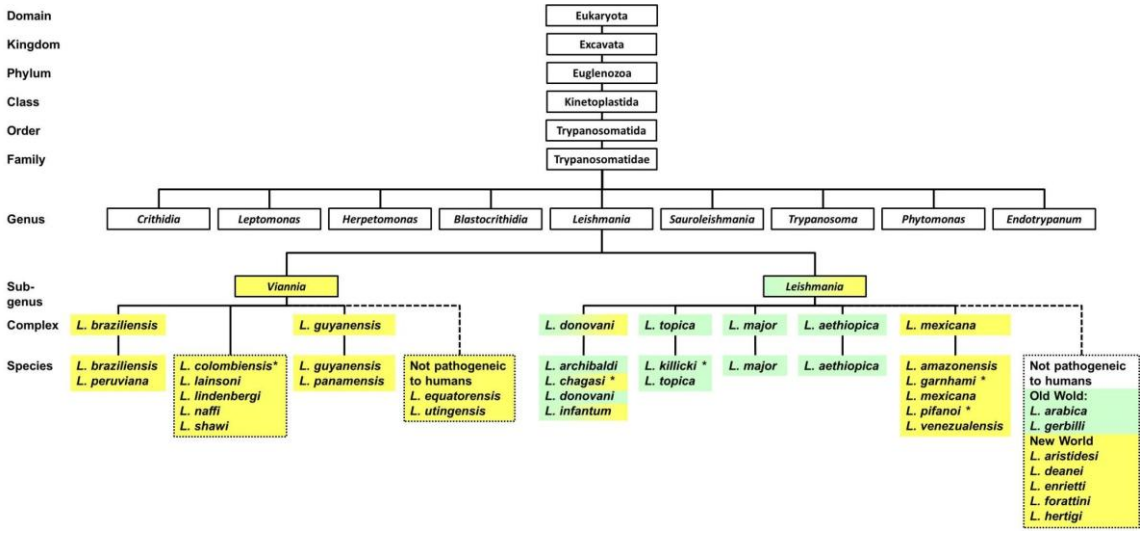


Figure 1-11: Taxonomic tree of *Leishmania spp*.

Species highlighted in green are endemic in the old world; the ones in yellow are endemic in the new world. The taxonomic positions of species marked with an asterisk are still under discussion. Figure taken from the PhD thesis of Johannes Doehl (Doehl, 2013).

Leishmania are endemic in tropical and subtropical regions of Africa, Middle East, Asia (= old world) and Southern and Middle America (= new world), but increased travel and global warming increases their range and vectors. There are 20 human infective *Leishmania* species and dependent on the species, different forms of Leishmaniasis are caused, namely cutaneous Leishmaniasis (CL), mucocutaneous Leishmaniasis (MCL) and visceral Leishmaniasis (VL). *L. major*, *L. tropica* and *L. aethiopica* in the old world and *L. mexicana* species complex (*L. mexicana*, *L. amazonensis*, and *L. venezuelensis*) and the *L. braziliensis* species complex (subgenus *Viannia, L. [V.] braziliensis*, *L. [V.] guyanensis*, *L. [V.] panamensis,* and *L. [V.] peruviana*) in the new world cause CL (Centers for Disease Control and Prevention, 2014a). Transfer of *Leishmania* parasites, mainly the subgenus *Viannia* and *L. amazonensis*, to the mucosal membranes causes MCL. The species *L. donovani* and *L. infantum* (= *L. chagasi*) cause VL.

1.5.1 *Leishmania spp*. life cycle

Leishmania's life cycle and morphological stages (Figure 1-12, (Gossage *et al.*, 2003) differ from those described for *T. brucei* and *T. cruzi*.

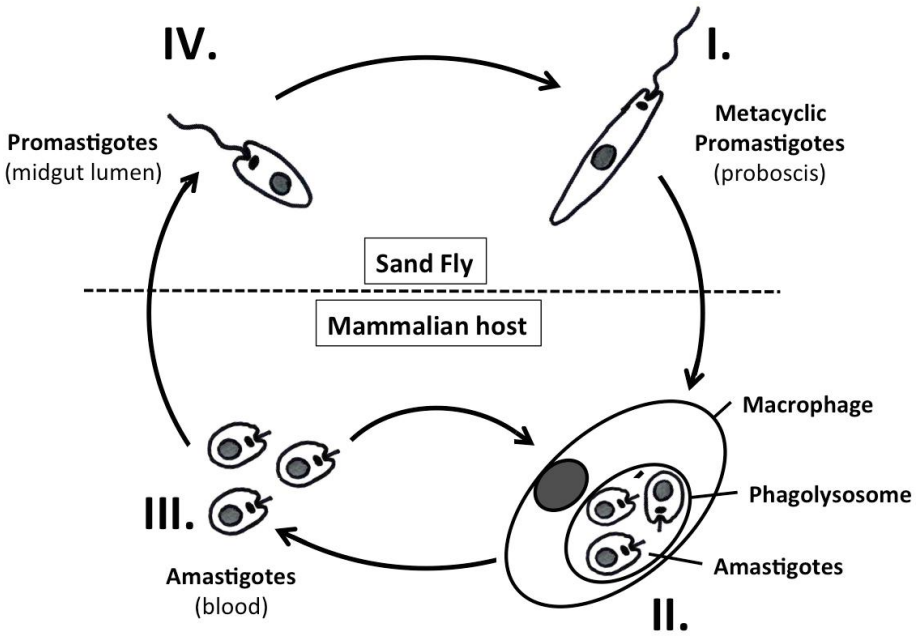


Figure 1-12: Life cycle stages of *Leishmania spp*.
I. metacyclic promastigotes, II. Amastigote parasites in phagolysosome, III. Amastigotes in the bloodstream, IV. Promastigotes. More details in the main text.

When infected sand flies of the genera *Phlebotomus* (old world) and *Lutzemyia* (new world) take a blood meal, they transfer the parasites. The *Leishmania* parasites in their metacyclic promastigote form (Figure 1-12 I.) reside in the anterior part of the gut and pharynx and the breakdown of the valve of the midgut entrance leads to a regurgitation of parasite containing material during the process of feeding. The metacyclic promastigotes are non-dividing and characterised by a free flagellum emerging from the anterior end of the cell. The nucleus is located in the centre of the cell and the kinetoplast anterior close to the basal body of the flagellum.

Once in the mammalian host the metacyclic parasites are phagocytised by macrophages and dendritic cells. Within the phagosome the parasites differentiate

into the amastigote form (Figure 1-12 II.), which allows them to survive the acidic pH and hydrolase-rich environment after lysosome fusion. The amastigotes replicate by binary fission and are released by the host cells into the bloodstream (Figure 1-12 III.). The released amastigotes are phagocytised by new host cells. Neither promastigotes nor amastigotes are able to actively enter cells, but require phagocytosis.

When a sand fly takes a blood meal, it takes up amastigotes, which differentiate into procyclic promastigotes (Figure 1-12 IV.), which proliferate and attach to the epithelium of the midgut. Procyclic promastigotes are not infective for mammalian hosts. Differentiation of the procyclic promastigotes to metacyclic promastigotes happens *via* two intermediate stages and is accompanied by a cessation of proliferation.

1.5.2 Leishmaniases

Leishmaniases, caused by parasites of the genus *Leishmania*, are transmitted by sand flies. Major risk factors include poor and crowded housing in or close to areas inhabited by the sand fly vector and malnutrition compromising the immune system. The WHO estimates 1.3 million new cases of Leishmaniases and 20 000 – 30 000 deaths annually (World Health Organisation, 2014d).

Cutaneous Leishmaniasis (CL) is the most common form of Leishmaniasis. It causes ulcers on the exposed areas of the body, which often heal spontaneously, but leave scars and, in severe cases, disability. About 95 % of CL cases occur in the Americas, the Mediterranean, Middle East and Central Asia (Figure 1-13). The WHO estimates 0.7 – 1.3 million new cases worldwide each year.

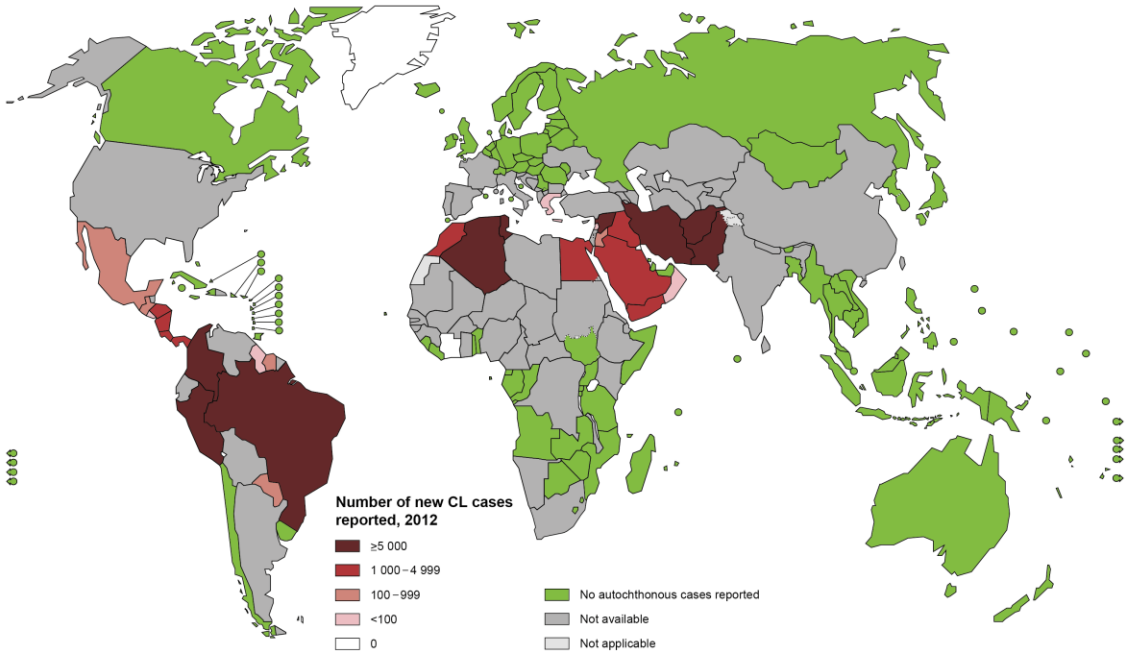


Figure 1-13: Geographic endemicity of cutaneous leishmaniasis, 2012.

Red colours indicate the number of new CL cases reported in 2012; countries highlighted in green did not report new cases. Figure modified from WHO (World Health Organisation, 2013).

Mucocutaneous Leishmaniasis (MCL) leads to destruction of mucosal membranes of mouth, nose and throat, severe scarring and disability, if left untreated. About 90 % of MCL cases occur in South America, mainly Bolivia, Brazil and Peru.

Visceral Leishmaniasis (VL), in its severe form called kala-azar, is the most dangerous form of Leishmaniasis and ultimately fatal if left untreated. It affects the inner organs, like liver and spleen, and is associated with high fever, weight loss and anaemia (Centers for Disease Control and Prevention, 2014a), but is often asymptomatic for long periods. There are an estimated 200 000 – 400 000 new cases each year (World Health Organisation, 2014d), 90 % in Bangladesh, Brazil, Ethiopia, India, South Sudan and Sudan (Figure 1-14). Following VL, patients often display post-kala-azar dermal leishmaniasis (PKDL) in the form of macular, popular or nodular rashes on face and other body parts.

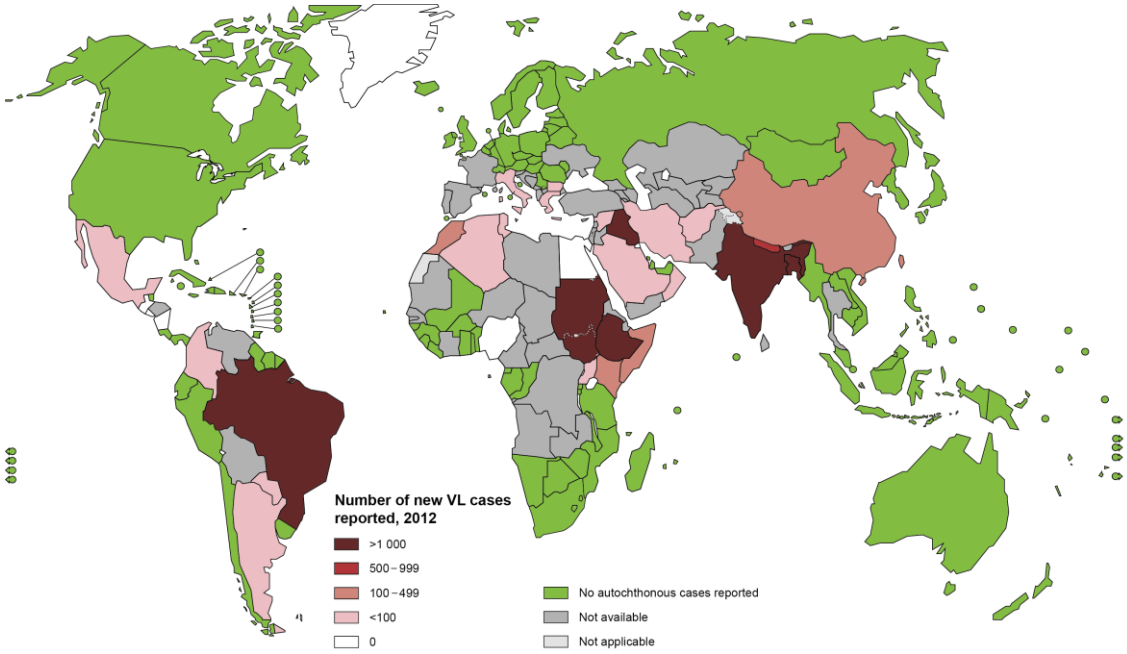


Figure 1-14: Geographic endemicity of viceral leishmaniasis, 2012.
Red colours indicate the number of new VL cases reported in 2012; countries highlighted in green did not report new cases. Figure from WHO (World Health Organisation, 2013).

1.5.3 Diagnosis and Current treatment of leishmaniasis

Leishmaniases are diagnosed using combinations of clinical symptomatic, parasitological and, for VL, serological tests, detecting antibodies (World Health Organisation, 2014d, Centers for Disease Control and Prevention, 2014a). VL diagnosis can be difficult, especially when the disease is asymptomatic. Treatment of leishmaniases is strongly dependent on species involved, stage of the disease and geographic location. CL often heals spontaneously, but as some *Leishmania* species can spread from the CL sores to mucosal membranes causing MCL, early treatment is necessary to avoid scarring and disability. VL requires prompt and complete treatment, as the disease can be lethal. Several drugs with anti-leishmanial activity are currently available (Figure 1-15), which often show strong toxicity and side effects and limited efficiency, reviewed by (Olliaro *et al.*, 2005, Singh *et al.*, 2012). The MOAs are varying from drug to drug with some being poorly characterised or not known. All drugs currently available have severe drawbacks, leading to the need of new compounds for anti-leishmanial therapy.

Combinational therapy, reviewed by (van Griensven *et al.*, 2010), could potentially reduce treatment duration, resulting in better patient compliance, lower costs and fewer toxic effects. Additionally, combinational therapy could increase the efficiency for complicated cases, like HIV co-infections. Several combinations have been studied, one of the most promising being a single application of liposomal amphotericin B followed by miltefosine (Sundar & Chakravarty, 2008), which showed good effectiveness and tolerance.

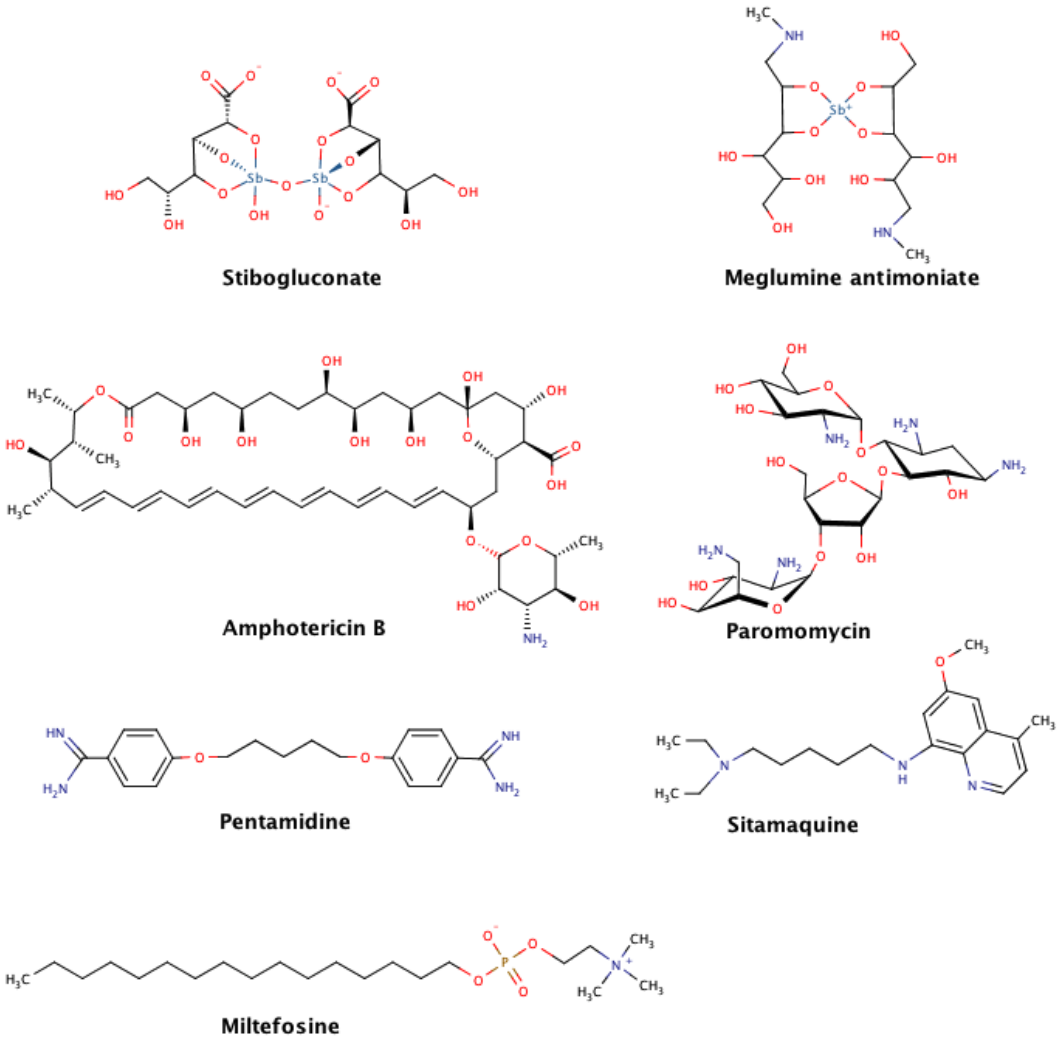


Figure 1-15: Anti-parasitic compounds used for treatment of Leishmaniases. Details in main text. Figure made with MarvinSketch (ChemAxon).

1.5.4 HIV co-infection

A major issue is HIV co-infection, which renders the symptoms of leishmaniases more severe and increases the rates of relapse and mortality (Alvar *et al.*, 2008). It also renders the diagnosis and treatment more difficult, as symptoms are more atypical and antibody-based diagnostics give false negatives. Anti-leishmanial therapeutics show reduced efficiency with HIV co-infection, but treatment with antiretroviral mediation reduces symptoms in and relapse and mortality of co-infected patients.

1.6 Vaccine development against the TriTryps

Vaccination is the most effective way of combating diseases as it provides long lasting protection against the pathogen without the need of constant monitoring or treatment. Currently, there are no vaccines effective against the TriTryps licensed for immunisation of humans.

Within the last 30 years extensive studies investigated the possible targets and strategies for development of anti-*T. brucei* vaccines without much success (reviewed in (Magez *et al.*, 2010, La Greca & Magez, 2011). Many studies using attenuated parasites (Esser *et al.*, 1982), VSGs (Cornelissen *et al.*, 1985), cytoskeletal proteins like tubulin and actin (Lubega, Byarugaba*, et al.*, 2002, Lubega, Ochola*, et al.*, 2002, Li *et al.*, 2007, Li *et al.*, 2009), microtubule-associated proteins (MAPs, (Balaban *et al.*, 1995), or flagellar pocket extracts (Mkunza *et al.*, 1995, Olenick *et al.*, 1988, Radwanska *et al.*, 2000) for vaccination failed to produce long-term B cell mediated immunity. *T. brucei* parasites evade the host's immune system by hiding most of the cell surface under the VSG coat, which is not only switched to generate a new antigenic variant, but is also recycled constantly, removing bound antibodies from the surface (Engstler *et al.*, 2007). Additionally, *T. brucei* parasites were shown to suppress the immune system upon infection (Murray *et al.*, 1974a, b, Askonas *et al.*, 1979, Clayton *et al.*, 1979) *via* destruction of the B cell compartments (Radwanska *et al.*, 2008) also removing acquired immune protection

against other pathogens. This renders long-term immunisation against this parasite extremely difficult.

Vaccine development against *T. cruzi* parasites has been more successful. Many studies in mice and dogs provide promising proof-of-concept data for preventative and disease-controlling/therapeutic vaccines. The studies include vaccines based on attenuated (Basombrio & Besuschio, 1982) or killed (Basombrio *et al.*, 1982) parasites, recombinant proteins like trans-sialidases or cruzipain (Frank *et al.*, 2003, Araujo *et al.*, 2005, de Alencar *et al.*, 2007, Hoft *et al.*, 2007, Giddings *et al.*, 2010, Cazorla *et al.*, 2010, Eickhoff *et al.*, 2010), DNA (Limon-Flores *et al.*, 2010, Garg & Tarleton, 2002, Chou *et al.*, 2010, Eickhoff *et al.*, 2011, Araujo *et al.*, 2005) and viral vectors (Machado *et al.*, 2006, Duan *et al.*, 2009, Miyahira *et al.*, 2005). Only immunisations causing strong cytotoxic T-cell responses seem effective, limiting the possible formulations for the vaccines. Despite the encouraging results, vaccine development has been hampered by worries of vaccination-stimulated autoimmunity (Kierszenbaum, 1999, Tarleton & Zhang, 1999, Engman & Leon, 2002), discussions about the need of a vaccine in the presence of (momentarily) effective vector control strategies (Camargo, 2009, Junqueira *et al.*, 2010) and problematic clinical trials due to the lowered infection rate and long course of the chronic phase of *T. cruzi* infection.

Hundreds of antigens as potential vaccine candidates are described for *Leishmania spp.* (reviewed in (Nagill & Kaur, 2011, Singh *et al.*, 2012), including killed and attenuated *Leishmania* parasites, recombinant proteins like kinetoplastid membrane protein 11 (KMP-11, (Jardim *et al.*, 1995), hydrophilic acetylated surface protein B (HASPB) and small hydrophilic endoplasmic reticulum associated protein (SHERP) (Sadlova *et al.*, 2010, Maroof *et al.*, 2012). So far, vaccine candidates performed poorly in clinical trials (Noazin *et al.*, 2008) and the limitation of animal models available for the different leishmaniases further complicates the development (Okwor & Uzonna, 2009). Two canine vaccines are licensed in Brazil, Leishmune and Leishtec (Evans & Kedzierski, 2012). Leishmune showing successful results in field trials, Leishtec producing more variable results.

An alternative to the immunisation against the parasites themselves might be the transmission blocking vaccines (TBVs). They aim to generate immunity against a different life cycle stage of the parasite or the parasite-insect vector interaction and therefore block infection of the insect vector or development of the parasite within the vector. It was shown that the presence of anti-procyclic parasite antibodies in the blood meal of the tsetse fly suppresses the development of the parasite in the vector (Maudlin *et al.*, 1984, Nantulya *et al.*, 1980). Possible targets for those immunisations are surface proteins of procyclic *T. brucei* parasites like GPEET, EP procyclins (Roditi & Lehane, 2008) and alanine-rich protein isoforms (BARP, (Urwyler *et al.*, 2007). Immunisation with tsetse fly midgut extracts (Kinyua *et al.*, 2005) was shown to reduce the transmission rate of the parasites and decreased the viability of the insect vector itself, indicating tsetse fly midgut antigens to be promising candidates for TBV development. The canine vaccine against visceral leishmaniasis Leishmune (FML), a fucose-mannose ligand, was shown to generate antibodies inhibiting the attachment of *Leishmania* parasites to the sand fly midgut (Saraiva *et al.*, 2006), thus interfering with the parasites development in the vector. Also, immunisation with sand fly salivary antigens was shown to reduce *Leishmania* transmission (Kamhawi *et al.*, 2000, Thiakaki *et al.*, 2005).

Another vaccination strategy regards the possibility for symptomatic vaccination, aiming to reduce the problems caused not by the parasites themselves but by the parasite-host interactions or the reaction of the host's immune system towards the parasite. This is of particular interest for the anti-*T. brucei* therapy in cattle where it was shown that the presence of antibodies against the cysteine protease congopain from *T. congolense* reduced the symptoms of *T. congolense* infection (Authie *et al.*, 1992, Lalmanach *et al.*, 2002). For *T. cruzi* infections a therapeutic vaccine is of high interest as the chronic phase of infection lasts many years without symptoms. The same antigens used for preventative immunisation could be shown to decrease the parasite burden and inflammation of the heart when administered as therapeutic vaccine (Dumonteil *et al.*, 2004, Zapata-Estrella *et al.*, 2006, Limon-Flores *et al.*, 2010, Sanchez-Burgos *et al.*, 2007).

Though vaccination is the most efficient way to combat diseases, the vaccine development proved difficult for the TriTryps.

1.7 Drug discovery

Drug discovery is the process of identification and partial validation of a drug candidate or lead compound for treatment of a specific disease. This includes the identification of an effective molecule and its modification in order to increase affinity, selectivity, potency, metabolic stability and bioavailability after administration.

1.7.1 Classical Pharmacology – Forward Pharmacology

In classical pharmacology (or forward pharmacology) chemical libraries containing small synthetic molecules, natural products and/or biological extracts are screened against whole cells or organisms in order to identify substances showing the desired effect, e.g. killing of a pathogen and survival of the human cells. Most drug candidates are small organic molecules, either of synthetic or biological origin, but sometimes bigger and more complex molecules are identified, e.g. antibiotics like vancomycin. The mechanism of action (MOA) of many of these active compounds is not initially known and the biological target and the MOA are sometimes identified in later stages of the drug discovery process.

1.7.2 Target-based drug design – Reverse Pharmacology

Here, a biological target is identified and molecules are screened or designed to specifically act on it. Biological targets are macromolecules, e.g. proteins, or pathways involved in the pathology of interest and often essential for cell proliferation and/or survival of an organism. The validation of a molecule or pathway as a potential drug target involves determination of its physiological role within the cell/organism and its involvement in the pathology. In the case of a

pathogenic organism, the potential drug target in the pathogen is usually compared to its human counterpart to identify and exploit differences between pathogen and host, and thus reduce possible side effects of the drug.

1.8 Identification of potential drug targets in the TriTryps

In the last decades many potential drug targets for the TriTryps were identified from a variety of different pathways throughout the parasite cells. Many metabolic pathways and/or enzymes within these pathways are essential for the survival and infectivity of the parasites and show significant differences between the TriTryps and humans. These provide potential targets for anti-parasite chemotherapy.

1.8.1 Nucleotide metabolism

The nucleic acids DNA and RNA are central macromolecules in all kingdoms of life. They are large polymers built from nucleotides, which in turn are composed of a nitrogenous base, a 5-membered sugar ring (deoxyribose for DNA and ribose for RNA) and a phosphate group. The precursors used for DNA and RNA biosynthesis are the nucleoside triphosphates dATP, dCTP, dGTP and dTTP (dNTPs) for DNA and ATP, CTP, GTP and TTP (NTPs) for RNA. Besides being essential for nucleic acid synthesis the NTPs carry out a variety of important intracellular functions, like energy transfer (ATP), cell signalling (cAMP, cGMP) and as components of cofactors for enzymatic processes (e.g. NAD and Coenzyme A). Because of their crucial role the nucleotide pool inside the cell has to be tightly controlled.

Cells synthesise nucleotides in two different ways, the *de novo* pathway and the salvage pathway. As the name suggests, the *de novo* pathway produces the nucleotides and all its components using basic components inside the cell. The salvage pathway takes advantage of nucleobases and nucleosides taken up from the cell's environment. Most cells are able to use both pathways for purine and pyrimidine nucleotide synthesis. However, kinetoplastids have lost the ability to

synthesise purines *de novo* and rely solely on the salvage pathway (de Koning *et al.*, 2005).

1.8.1.1 Purine nucleotide metabolism in *T. brucei*

The uptake of purine nucleobases and nucleosides is facilitated by the high affinity transporters P1 and P2, which are able to outcompete the host's purine uptake systems. Although they belong to the class of equilibrative nucleoside transporters (ENT) they use energy in the form of proton motive force to actively concentrate nucleosides inside the cell (de Koning & Jarvis, 1997, Molina-Arcas *et al.*, 2009, Young *et al.*, 2008). The P2 transporter has been shown to import trypanocidal nucleoside analogues, like cordycepin (Geiser *et al.*, 2005). Additionally, trypanosomes have purine base and pyrimidine base/nucleoside transporters, H1-3.

The salvaged nucleobases and nucleosides are interconverted by a complex and redundant network of enzymes, Figure 1-16. Because of their redundancy most of the enzymes in the salvage pathway are non-essential for proliferation (Hammond & Gutteridge, 1984). For example, AMP can be produced by adenosine (Ado) phosphorylation by adenosine kinase (AK), or by Ado hydrolysis through adenosine nucleosidase (purine nucleoside hydrolase, PNH) and subsequent phosphoribosyltransfer by adenine phosphoribosyltransferase (APRT) or from IMP *via* adenylosuccinate using the enzymes adenylosuccinate synthase (ADSS) and adenylosuccinate lyase (ASL). Generally all nucleosides can be hydrolysed to their nucleobases, which in turn are converted into nucleoside monophosphates. The monophosphates are further phosphorylated, converted into one another or used as precursors for signalling molecules.

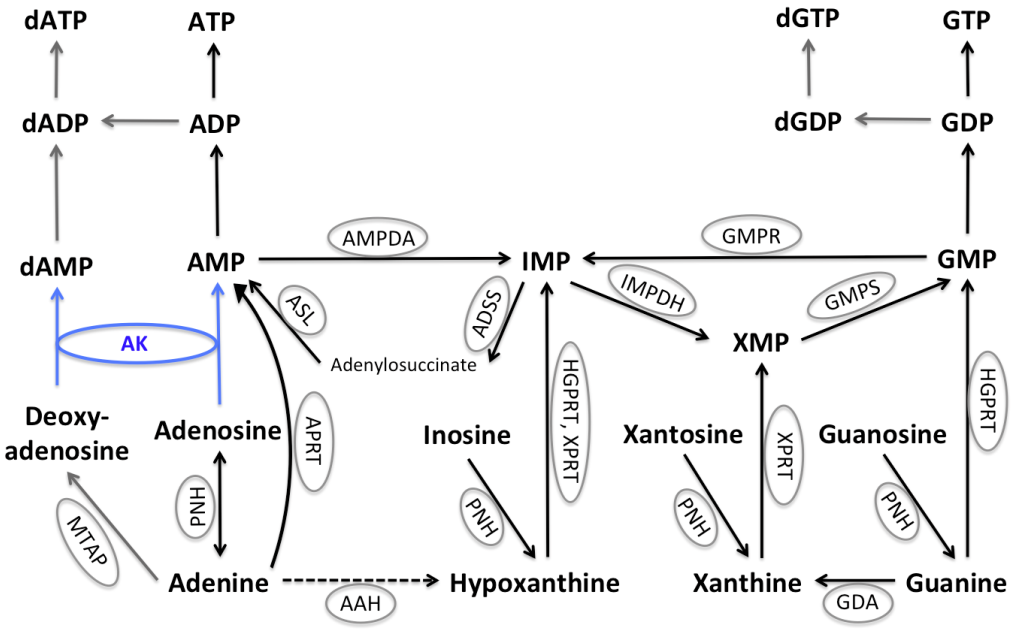


Figure 1-16: Purine nucleotide metabolism in kinetoplastids.

Shown are the purine nucleotides and the enzymes (grey circles) interconverting them. APRT = adenine phosphoribosyltransferase (PRT), EC 2.4.2.7; MTAP = methylthioadenosine phosphorylase; AK = adenosine kinase, EC 2.7.1.20; AMPDA = AMP deaminase, EC 3.5.4.6; ADSS = adenylosuccinate synthase, EC 6.3.4.4; ASL = adenylosuccinate lyase, EC 4.3.2.2; AAH = adenine deaminase (or adenine amino hydrolase), EC 3.5.4.2; PNH = purine nucleoside hydrolase, EC 3.2.2.1; HGPRT = hypoxanthine guanosine PRT, EC 2.4.2.8; XPRT = xanthine PRT, EC 2.4.2.22; IMPDH = IMP dehydrogenase, EC 1.2.1.14; GMPS = GMP synthase, EC 6.3.4.1; GMPR = GMP reductase, EC 1.6.6.8; GDA = guanine deaminase, EC 3.5.4.3. Dashed arrow indicates enzymes in *Leishmania sp*. and *T. vivax*.

1.8.1.2 Pyrimidine nucleotide metabolism in the TriTryps

While purine nucleotide metabolism in kinetoplastids is well studied, pyrimidine nucleotide metabolism is much less so. The TriTryps are able to use both the *de novo* and the salvage pathway to synthesise pyrimidine nucleotides with the exception of cytosine for which trypanosomes lack the enzymes of the salvage pathway (Hofer *et al.*, 2001). A schematic summary of pyrimidine nucleotide metabolism is shown in Figure 1-17.

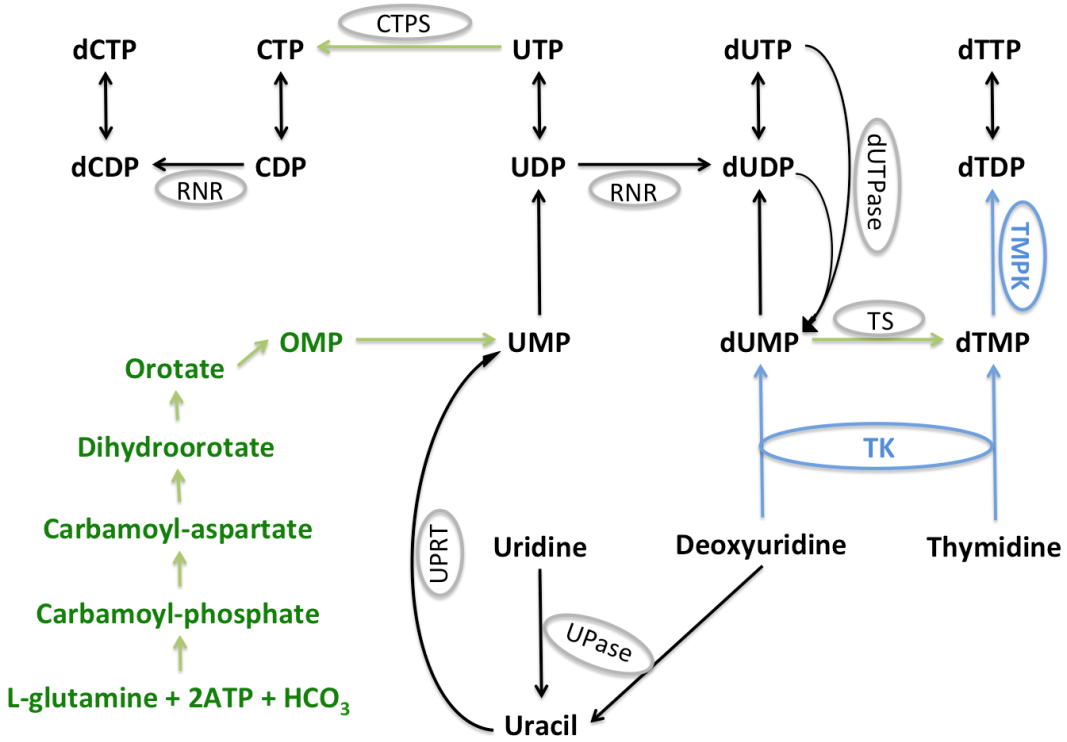


Figure 1-17: Pyrimidine metabolism in kinetoplastids.

Shown are the pyrimidine nucleotides and the enzymes (grey circles) interconverting them. The *de novo* synthesis pathway is shown in green, the salvage pathway in black and blue. CTPS = CTP synthase, EC 6.3.4.2; RNR = ribonucleotode reductase, EC 1.17.4.1/2; TK = thymidine kinase, EC 2.7.1.21; TMPK = thymidylate kinase, EC 2.7.4.9; TS = thymidylate synthase, EC 2.1.1.45; UPRT = uracil phosphoribosyltransferase, EC 2.4.2.9; UPase = uridine phosphorylase, EC 2.4.2.3; dUTPase = dUTP nucleotide hydrolase, EC 3.6.1.23.

The *de novo* pathway utilises L-glutamine, ATP and carbonate to synthesise uridine monophosphate (UMP) in six enzymatically-catalysed steps. UMP is readily phosphorylated and converted into other pyrimidine nucleotides. Using the salvage pathway, kinetoplastids take up uracil and its nucleosides from the cell's surroundings using the transporters U1 and U3 and thymidine (dThd) is taken up to a certain extent by the purine transporter P1 (Ali *et al.*, 2013). Imported uridine (Urd) is converted into uracil, which is used to generate UMP, a reaction catalysed by uracil phosphoribosyltransferase (UPRT). UMP can be further phosphorylated to UDP and UTP. UDP is reduced by ribonucleotide reductase (RNR) to yield dUDP, which is in turn either phosphorylated to dUTP or dephosphorylated to dUMP. dUMP can be methylated to dTMP by thymidylate synthase (TS). RNR is regulated by dTTP, which shifts its specificity from pyrimidine to purine nucleotide reduction (Thelander & Reichard, 1979, Xu *et al.*, 2006, Hofer *et al.*, 2012). Salvaged dThd and

dUrd are phosphorylated by thymidine kinase (TK) leading to dTMP and dUMP respectively. Further phosphorylation by thymidylate kinase (TMPK) results in dTDP and dUDP and subsequently dTTP and dUTP, direct precursors for DNA and RNA synthesis.

In eukaryotes, the intracellular dTTP pool is tightly regulated and coupled to the cell cycle and DNA synthesis (Hu & Chang, 2007, Mathews, 2006). While other dNDPs, including dUDP and purine nucleoside diphosphates, can also be synthesised by RNR, dTDP is generated only *via* phosphorylation of dTMP by thymidylate kinase (TMPK). Besides being essential for dTDP synthesis, TMPK has been shown to be involved in DNA damage repair, where blockage of TMPK results in dUTP incorporation into the DNA upon repair of double strand breaks (Hu *et al.*, 2012). The only nucleosides not made in the salvage pathway are cytosine nucleotides, whose only source is synthesis from UTP by CTP synthetase (CTPS).

1.8.1.3 Potential drug targets within the nucleotide metabolism

The dependence on tightly regulated nucleotide pools for correct DNA synthesis, cell signalling and energy balance, provides many potential drug targets, e.g. phosphoribosyltransferases (PRTs), which synthesise nucleotide monophosphate from the nucleobase and phosphoribosyl pyrophosphate (Carter *et al.*, 2008), deoxyuridine triphosphate nucleotidohydrolase (dUTPase), catalysing the hydrolysis of dUTP to dUMP and pyrophosphate (Hidalgo-Zarco *et al.*, 2001, Castillo-Acosta *et al.*, 2008), or cyclic AMP (cAMP) phosphodiesterases (PDEs), which degrades the secondary messenger molecule cAMP to AMP (Oberholzer *et al.*, 2007, de Koning *et al.*, 2012, Bland *et al.*, 2011).

1.8.1.4 Nucleoside analogues as chemotherapeutic agents

Nucleoside analogues, mimicking the natural nucleosides, often have cytotoxic properties. Most analogues are phosphorylated by nucleoside kinases and used as building blocks in RNA or DNA synthesis, which leads to premature strand

termination, resulting in cell death. This property was developed into a variety of laboratory applications, e.g. 3'-end labelling of RNA (Lingner & Keller, 1993). Additionally, many of them in their phosphorylated form function as inhibitors of DNA- or RNA polymerases, like retroviral reverse transcriptases.

Among the best-studied nucleoside analogues are the adenosine analogues (Figure 1-18), which are activated by their phosphorylation by AK and adenylate kinase (AMPK). Their ability to cross the blood brain barrier makes them attractive candidates in drug development and their pharmacokinetic properties are well characterised in humans.

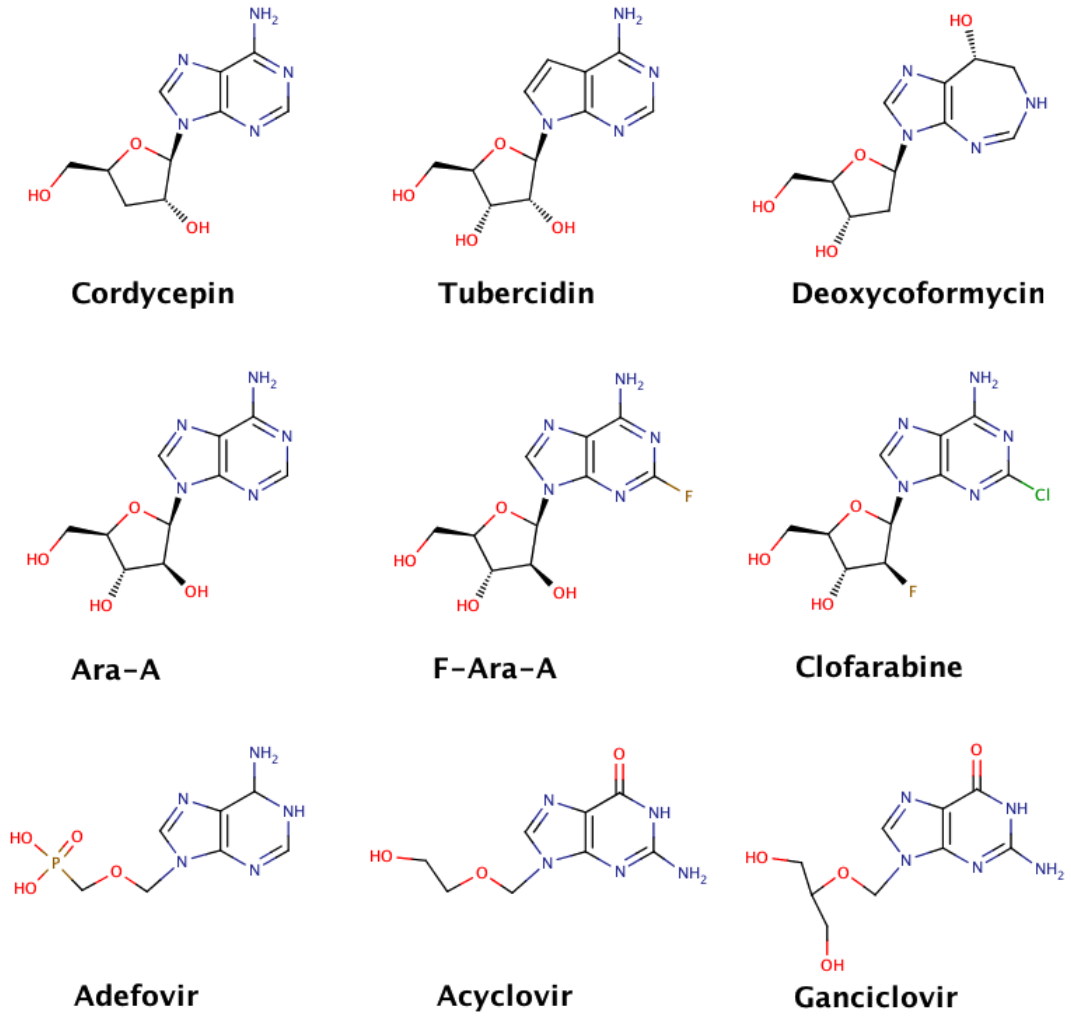


Figure 1-18: Some medically relevant purine analogues.

Details in main text. Ara-A = vidarabine, F-Ara-A = fludarabine. Figure made using MarvinSketch (ChemAxon).

While Ara-A was used against Herpes simplex infections (Shiota *et al.*, 1978), F-Ara-A, clofarabine and deoxycoformycin have been used as chemotherapeutics in cancer therapy (Gray *et al.*, 1982, Leahey *et al.*, 1997, Kline & Larson, 2005). Tubercidin has been shown to be highly toxic to most cells, including the TriTryps, being phosphorylated and built into nucleic acids, as well as affecting their glycolysis (Drew *et al.*, 2003). Anti-trypanosomal and anti-leishmanial activity was shown and its effectivity as well as a reduction of host toxicity could be achieved by co-administration with nitrobenzylthioinosine (NBMPR), a mammalian specific purine nucleoside transport inhibitor (el Kouni *et al.*, 1983, Ogbunude & Ikediobi, 1982, Aoki *et al.*, 2009). Some purine nucleoside analogues were shown to inhibit retroviral reverse transcriptase in their phosphorylated form (Richman, 1990, Wainberg *et al.*, 2005) and others stop dThd phosphorylation by TK (Susan-Resiga *et al.*, 2007, Lynx & McKee, 2006, Lynx *et al.*, 2006). This can be used in cases where the target cells express the type I TK or generally a TK with broader substrate specificity than the cytosolic human type II TK. Examples of pathogens overexpressing a lower specificity TK are human herpes simplex virus (HSV), varicella-zoster virus (VZV) and cytomegalovirus (CMV). The treatments for these viruses include nucleoside analogues, such as acyclovir for HSV and VZV infections and ganciclovir in CMV therapy. Adefovir in its phosphorylated form acting as a polymerase inhibitor is used in treatment of hepatitis B (Birkus *et al.*, 2003) and HSV.

Cordycepin is trypanocidal *in vitro* (Williamson, 1972, Williamson & Macadam, 1976). Besides its integration into RNA, terminating its elongation, cordycepin has been shown to inhibit mammalian poly-A polymerase and hinder nucleoside-activated protein kinase activity in *Trypanosoma spp.* (Walter & Ebert, 1977). Its trypanocidal properties were shown to depend on its activation by phosphorylation through AK indicated by RNAi experiments in which AK knock down in *T. brucei* did result in reduction of cordycepin and tubercidin sensitivity (Luscher *et al.*, 2007). Sub-lethal doses of AK inhibitors resulted in cordycepin resistance indicating a switch to AMP production *via* the PNH-APRT pathway. Tubercidin remained trypanocidal due to its cytotoxic modification being in the nucleobase rather than

the ribose ring. Unfortunately, *in vivo* studies in mice revealed a rapid deactivation of cordycepin *via* deamination by the host's adenosine deaminase (ADA) in the plasma (Aiyedun *et al.*, 1973, Agarwal *et al.*, 1975). Co-administration of cordycepin and the ADA inhibitor deoxycoformycin, blocking the metabolism of cordycepin, could cure mice suffering late stage African trypanosomiasis (Rottenberg *et al.*, 2005, Vodnala *et al.*, 2009). As deoxycoformycin has proven to be teratogen, a combinational therapy is not reasonable (Airhart *et al.*, 1996). To circumvent the problem of co-administration, it might be possible to modify cordycepin in a way which makes it resistant to deamination by human ADA, but does not affect its trypanocidal properties. In order to keep its cytotoxic properties the new compound would have to remain a substrate of the *T. brucei* nucleoside transport systems (P1 and P2) and the enzymes activating it inside the trypanosomes, e.g. AK and AMPK.

Not only purine nucleotide analogues have been successfully used as prodrugs, but also pyrimidine analogues (Figure 1-19). In this context, there has been a big interest in TKs and TMPKs in drug development because of differences in substrate specificity and expression profile of the human enzymes compared to the pathogens. While all TKs prefer dThd as substrate and are usually able to phosphorylate dUrd, type I TKs have generally lower substrate specificity and phosphorylate dC readily. This ability to accept a broader range of substrates enables the type I TKs to phosphorylate cytotoxic nucleoside analogues. The pyrimidine analogues (E)-1-*β-D*-(arabinofuranosyl)thymine (AraT) and (E)-5-(2-bromovinyl)-2'-deoxyuridine (BVDU) are well characterised for their antiherpetic properties, while 2'-fluoro-5-iodo-(1-*β-D*-arabinofuranosyl)uracil (FIAU) has been shown to be effective anti-hepatitis B agents (Tennant *et al.*, 1998). The substrate analogue 3'-azido-3'deoxythymidine (AZT) is used in HIV therapy as it is phosphorylated by the human TK1 and inhibits the HIV reverse transcriptase. Additionally, it was shown to inhibit growth of *L. amazonensis* promastigotes (Araujo *et al.*, 2011), indicating AZT to be a potential candidate for anti-leishmanial therapy. Uninfected resting human cells are much less affected by these nucleoside analogues, as the human cytosolic TK is only expressed during cell division, when

the cells are actually synthesising DNA. However, some toxicity has been observed for AZT and FIAU, probably mediated by the mitochondrial type I TK (Wang *et al.*, 2011).

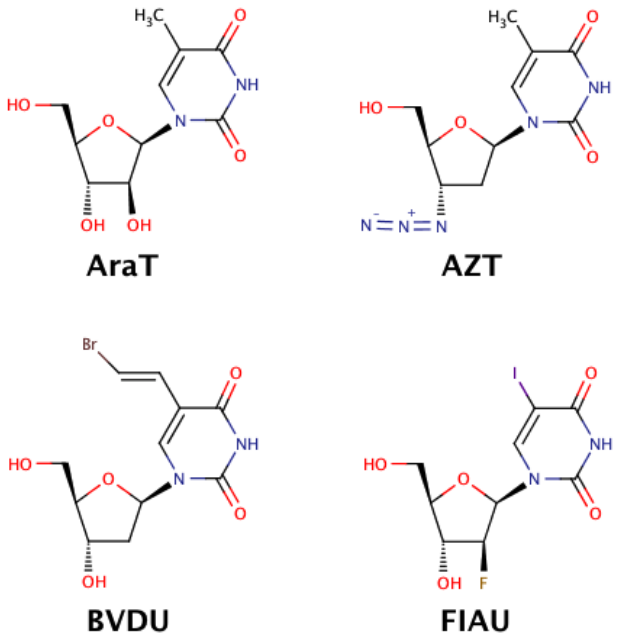


Figure 1-19: Some medically relevant pyrimidine analogues.
AZT = 3'-azido-3'deoxythymidine, AraT = 1-β-D-(arabinofuranosyl)thymine, BVDU = (E)-5-(2-bromovinyl)-2'-deoxyuridine, FIAU = 2'-fluoro-5-iodo-(1-β-D-arabinofuranosyl)uracil. Figure made using MarvinSketch (ChemAxon).

1.8.2 Amino acid metabolism

Amino acids are essential molecules within all organisms. Approximately 500 different amino acids are known, which are all comprised of carboxylic acid and amine groups with different side chains specific for each amino acid. Of these, 20 *L*-amino acids are components of proteins. Additionally, amino acids function as signalling molecules and precursors for a number of important metabolites, as well as fatty acids. Amino acid synthesis and catabolism show large differences between organisms.

Cells have the ability to synthesise certain amino acids *de novo* from inorganic compounds, amino acid remodelling and conversion of non-amino acid carbon skeletons into amino acids. Amino acids that cannot be synthesised, but have to be

taken up from the environment are considered essential for the given organism. The uptake from the environment can be either as free amino acids or in the form of proteins and peptides, which are broken down to amino acids by various proteases and peptidases.

Amino acids are catabolised into a variety of different metabolites. They are categorised as glucogenic or ketogenic. Glucogenic amino acids are transformed into pyruvate or citric acid cycle intermediates, e.g. oxaloacetate or α-ketoglutarate, which can be used in gluconeogenesis. Most amino acids are at least partially glucogenic, the only exception being leucine, which is solely ketogenic. Ketogenic amino acids are used to produce acetyl-CoA and acetoacetyl-CoA. Isoleucine, lysine, plenylalanine, tryptophan and tyrosine are both glucogenic and ketogenic and give rise to glucose and fatty acid precursors. Some amino acids, e.g proline, can serve as an energy source by oxidation of the carbon skeleton.

1.8.2.1 Amino acid metabolism in the TriTryps

Amino acid metabolism in the TriTryps is complex and varies depending on the life cycle stage and environment of the parasites. Generally the parasites maintain an amino acid pool of alanine, glutamate and proline as precursors of metabolites and for osmoregulation (Vieira *et al.*, 1996, Rohloff *et al.*, 2003, Darlyuk *et al.*, 2009), where amino acid efflux is used to reduce cell swelling upon hypotonic stress (Lefurgey *et al.*, 2005, Vieira *et al.*, 1996, Rohloff *et al.*, 2003). Additionally, the hydrolysis of peptides and proteins plays a crucial role in the parasites ability to invade the host's tissues and cells and in the evasion of the immune system.

Bloodstream form *T. brucei* parasites rely on *D*-glucose in the host's blood stream as energy source, which they metabolise to succinate and acetate. As soon as they are transferred to their insect vector and differentiate into the procyclic form parasites, their metabolism changes to use *L*-proline as main carbon and energy source, which is metabolised to succinate and further to alanine. The same applies for the insect stages of *T. cruzi* and *Leishmania spp*. (Bringaud *et al.*, 2012). However, procyclic *T. brucei* cultured in glucose-rich media prefer D-glucose as energy and carbon

source and convert it to succinate and acetate (Coustou *et al.*, 2008). The ability of amino acid biosynthesis of *T. brucei* parasites is less well studied and it is not clear which amino acids are essential for the parasites' survival and differentiation. However, threonine was shown to be essential for growth of procyclic *T. brucei* parasites. It is converted to glycine and acetate (Linstead *et al.*, 1977), and was shown to be a precursor in lipid biosynthesis (Gilbert *et al.*, 1983). In the databases homologues of only two out of the five enzymes necessary for leucine biosynthesis are present in *T. brucei*, none of which are characterised in the parasite. Therefore, leucine is expected to be essential in *T. brucei*. In *T. cruzi* and *Leishmania spp*. leucine is used for sterol and fatty acid synthesis (Ginger *et al.*, 2000). *Leishmania spp.* parasites were shown to produce up to 77 % of the sterol from leucine, while it was only 8 % in *T. cruzi*. However, experiments in *T. brucei* showed no integration of radioactively labelled leucine into sterol biosynthesis in this parasite (Guiomar Perez-Moreno *et al.*, personal communication). *T. cruzi* parasites were shown to metabolise asparagine, glutamine, aspartate, glutamate, leucine, isoleucine and proline (Zeledon, 1960, Sylvester & Krassner, 1976, Mancilla *et al.*, 1967). While glycolysis is active, alanine is produced from pyruvate (Zelada *et al.*, 1996). The growth of some *T. cruzi* life cycle stages, mainly epimastigotes, was shown to depend on the availability of arginine (Pereira *et al.*, 1999, Pereira, Alonso, Ivaldi*, et al.*, 2002, Canepa *et al.*, 2004) in spite of the necessary genes found in the *T. cruzi* genome, indicating the need for further investigations. The presence of glucose together with aspartate, glutamate and proline promote differentiation from epimastigote to metacylic trypomastigote *T. cruzi* (metacyclogenesis) (Contreras *et al.*, 1985). *Leishmania spp.* have three main amino acid transporters, taking up proline, alanine, methionine and valine from the environment (Bonay & Cohen, 1983, Mazareb *et al.*, 1999). As all kinetoplastids, *Leishmania spp.* can convert *L*-proline into glutamate (Coustou *et al.*, 2008, van Weelden *et al.*, 2003, van Weelden *et al.*, 2005, Krassner & Flory, 1972, Sylvester & Krassner, 1976), but unlike *T. brucei* and *T. cruzi* they also can catalyse the reverse reaction and synthesise proline from glutamate or ornithine (Bringaud *et al.*, 2012).

Overall, amino acid metabolism in the TriTryps is complex and tightly regulated depending on environmental conditions, species and life cycle stage of the parasite. Therefore, identification of potential drug targets in amino acid metabolism can be a difficult task.

1.8.2.2 Potential drug targets within the amino acid metabolism

Many potential drug targets have been identified, e.g. proline racemase, which converts L-proline into a racemic mixture (Chamond *et al.*, 2003), alanine aminotransferase, producing alanine from pyruvate (Spitznagel *et al.*, 2009), arginine kinase, which catalyses the reversible transphosphorylation between N-phospho-L-arginine and ADP (Pereira *et al.*, 1999, Pereira *et al.*, 2000, Pereira, Alonso, Ivaldi*, et al.*, 2002, Pereira, Alonso, Torres*, et al.*, 2002), the S9 family serine peptidases prolyl oligopeptidase (POP) and oligopeptidase B (OPB), involved in host tissue invasion of *T. cruzi* (Grellier *et al.*, 2001, Santana *et al.*, 1997, Burleigh *et al.*, 1997) and crossing the blood-brain-barrier of *T. brucei* (Bastos *et al.*, 2010, Lonsdale-Eccles & Grab, 2002).

1.8.2.3 Peptidases

Peptidases catalyse the hydrolysis of peptide bonds (EC 3.4.-.-, MEROPS, (Rawlings *et al.*, 2014) and are divided into “endopeptidases” which cleave peptide bonds within a polypeptide chain, and “exopeptidases” which cleave amino acid residues off the N- or C-terminus. There are two ways to classify peptidases: 1) by their substrates and products, and 2) homology based. Peptidases are divided into families depending on the substrates and products, but as substrate specificities are often difficult to define, the catalytic mechanism is often used to support this classification. Peptidase clans are homology-based classifications and represent one or more families of enzymes with a common evolutionary origin reflected in their structure, and/or conserved catalytic residues. Peptidase clans are assigned a two letter code, the first depending on their catalytic mechanism, e.g. clans of

metallopeptidases are assigned the first letter M. In case of an evolutionary divergence within the clan, it is divided into sub-clans.

1.8.2.4 Metallopeptidases

Many peptidases require co-catalytic metal ions for their activity and are classified into a number of metallopeptidase families, named M1-M98. For example M17 (Clan MF), M24 (Clan MG), M20 and M28 (Clan MH), all use penta-coordinated divalent metal ions. The M17 and M24 families contain aminopeptidases, M20 contains carboxypeptidases and M28 contains members of both.

1.8.2.5 Aminopeptidases

Aminopeptidases catalyse the cleavage of a single amino acid residue off the N-terminus of peptides and proteins (EC 3.4.11.-). They are divided into families dependent on their substrates, cofactors and mechanisms. Some of these are metallopeptidases of the families M17, M24 or M28, containing one or two co-catalytic metal ions, usually zinc, manganese or cobalt. The leucyl aminopeptidases (LAPs) were the first group to be characterised (Prescott & Wilkes, 1966, Dick *et al.*, 1970), but early studies did not distinguish between the different enzyme families, which makes it difficult to put studies of a specific enzyme family into context.

1.8.2.6 Aminopeptidase inhibitors and their use as drug candidates

There are many effective aminopeptidase inhibitors, including aminoketones and their derivatives (Birch *et al.*, 1972, Kettner *et al.*, 1974), microbial peptide analogues (Umezawa *et al.*, 1976, Aoyagi *et al.*, 1978, Umezawa, 1982) (Umezawa *et al.*, 1985, Rich *et al.*, 1984, Wilkes & Prescott, 1985), α-aminoboronic acids (Baker *et al.*, 1983, Shenvi, 1986), amino acid hydroxamates (Chan *et al.*, 1982, Baker *et al.*, 1983, Wilkes & Prescott, 1985), α-aminoaldehydes (Andersson *et al.*, 1982), amino acid thiols (Chan *et al.*, 1982) and (aminoalkyl)phosphonic acids (Giannousis &

Bartlett, 1987). Many of those inhibit leucyl aminopeptidases (LAPs) and some examples are shown in Figure 1-20.

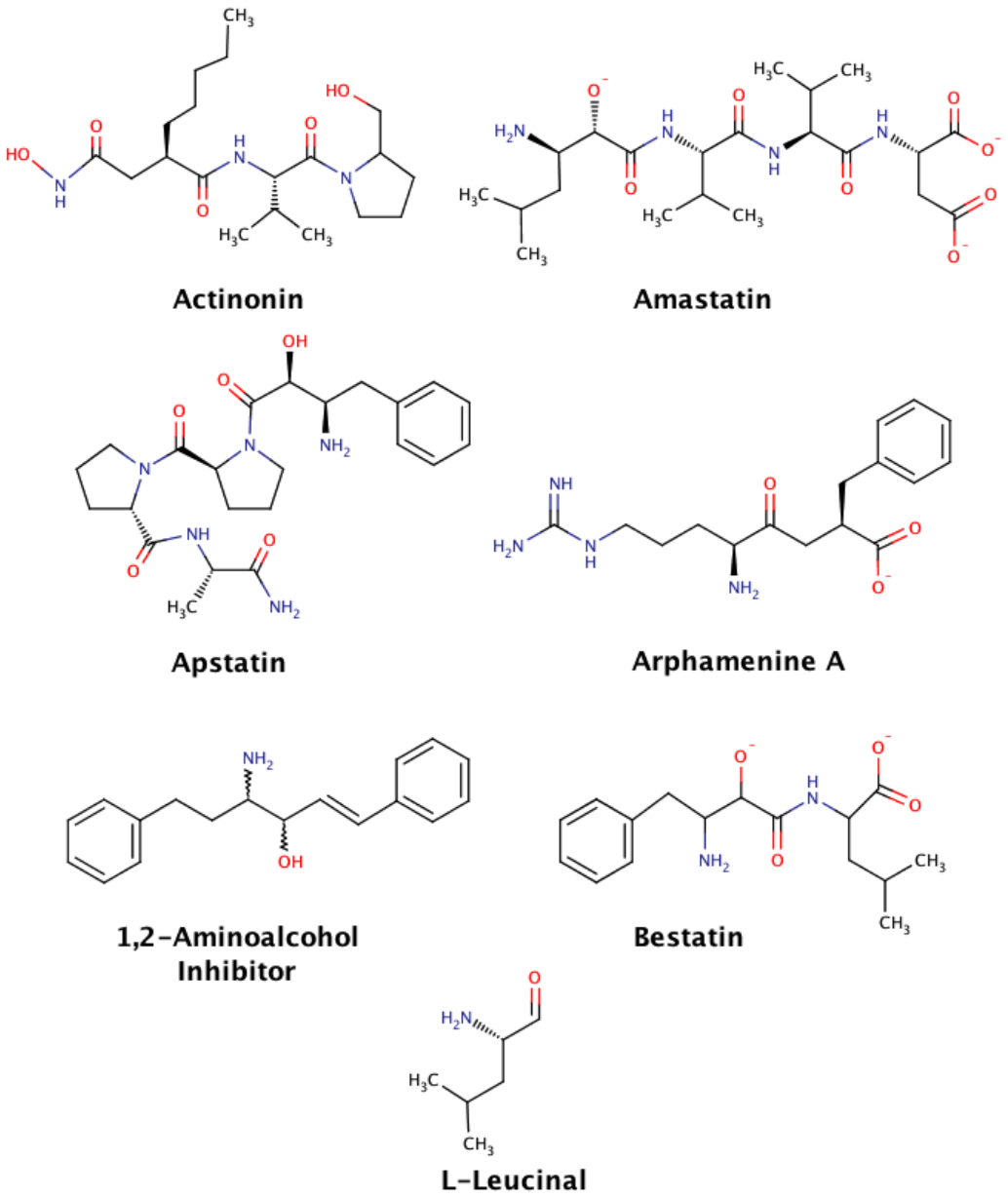


Figure 1-20: Some well-characterised leucyl aminopeptidase inhibitors. Details in main text. Figure made with Marvin Sketch (ChemAxon).

Among the microbial peptide analogue inhibitors actinonin (Gordon *et al.*, 1962, Umezawa *et al.*, 1985), amastatin (Aoyagi *et al.*, 1978, Rich *et al.*, 1984, Wilkes & Prescott, 1985), arphamenine A (Umezawa, 1982) and bestatin (Umezawa *et al.*,

1976, Rich *et al.*, 1984, Wilkes & Prescott, 1985) are the best-studied ones. A variety of derivatives, especially of actinonin and bestatin, have been synthesised and studied for their antibiotic and anti-parasitic properties (Anderson, Devlin*, et al.*, 1975, Anderson, Ollis*, et al.*, 1975, Rich *et al.*, 1984, Howarth & Lloyd, 2000).

Actinonin is a competitive inhibitor of aminopeptidase-N and LAP (Umezawa *et al.*, 1985) and peptide deformylase (Chen *et al.*, 2000) and has antibiotic properties against Gram-positive and Gram-negative bacteria and some phage strains *in vitro* (Gordon *et al.*, 1962). Amastatin is a slow-binding competitive inhibitor of aminopeptidase-A and -N and LAP, but not of aminopeptidase-B, while bestatin is a potent slow-binding inhibitor of aminopeptidase-B and LAP, but not aminopeptidase-A (Aoyagi *et al.*, 1978, Umezawa *et al.*, 1976, Tobe *et al.*, 1980) (Rich *et al.*, 1984, Wilkes & Prescott, 1985, Nagata *et al.*, 1991). Bestatin and its derivatives display anti-malarial activity, inhibiting the growth of *P. falciparum* (Nankya-Kitaka *et al.*, 1998). They also have been reported as immunomodifiers (Bruley-Rosset *et al.*, 1979), analgestic (Chaillet *et al.*, 1983) and inhibiting tumor growth (Umezawa, 1980). 1,2-aminoalcohols, based on the aminoalcohol core structure of bestatin, were shown to have anti-malarial properties (Howarth & Lloyd, 2000). Some of these derivatives act as LAP inhibitors, while others increase LAP activity. Apstatin, a not very efficient inhibitor of M24 prolyl aminopeptidases (Prechel *et al.*, 1995, Maggiora *et al.*, 1999, Cottrell *et al.*, 2000), was shown to be a potent slow-binding competitive inhibitor of *Leishmania* LAPs (Morty & Morehead, 2002). This compound has proven cardio-protective potential *via* a kinin-dependent pathway (Wolfrum *et al.*, 2001). Arphamenine A is a specific inhibitor of aminopeptidase B and is not effective against LAP or aminopeptidase A (Ohuchi *et al.*, 1984, Umezawa *et al.*, 1983). However, it was shown to have activity against *T.brucei* parasites *in vitro* (Knowles, 1993).

α-aminoaldehydes like L-leucinal, and aminophosphonic acids like L-leucine phosphonic acid, are potent, slow-binding inhibitors of porcine kidney LAP, cytosolic di- and tri-peptidase (Andersson *et al.*, 1982, Frick & Wolfenden, 1985), and aminopeptidase-N and LAP (Giannousis & Bartlett, 1987), respectively. Like

bestatin, they mimic the *gem*-diolate transition state of the reaction and therefore provide strong inhibition.

1.8.3 Other potential drug targets in the TriTryps

There are many potential drug targets identified in the TriTryps outside of the nucleotide or amino acid metabolisms, e.g. squalene synthase (SQS) and squalene epoxidase (SQE), which are involved in ergosterol biosynthesis (Perez-Moreno *et al.*, 2012), N-myristoyltransferase (NMT) involved in subcellular targeting of proteins (Price *et al.*, 2003), trypanothione reductase, which is crucial for maintenance of the redox homeostasis (Dumas *et al.*, 1997, Tovar *et al.*, 1998, Krieger *et al.*, 2000), mitogen-activated protein (MAP) kinase important for parasite differentiation (Wang *et al.*, 2005) and DNA topoisomerase which is involved in processes influencing the topology and organisation of the DNA (Das *et al.*, 2006).

1.9 Trypobase

Trypobase, an acronym for "Nucle**obase** derivatives as drugs against **tryp**anosomal diseases", was a collaborative research project funded for three years under the European Community's 7th Framework Programme, with collaborators in parasitology (the groups of Dolores Gonzalez-Pacanowska at "Consejo Superior de Investigaciones Cientificas", Granada, Spain, Carlos Robello at "Institute Pasteur de Montevideo-Uruguay, Montevideo, Uruguay and Reto Brun at "Swiss Tropical and Public Health Institute" Basel, Switzerland), medicinal chemistry (the groups of Ian Gilbert at the University of Dundee, Dundee, UK, Nils Gunnar Johansson at Medivir AB, Huddinge, Sweden and Ashis Baran Mandal at Syngene, Bangalore, India) and structural biology (the group of Keith Wilson, University of York, York, UK).

Trypobase addressed the neglected tropical diseases caused by the TriTryps (*T. brucei*, *T. cruzi* and *L. major*) and aimed to identify lead compounds for drug development against these pathogens, initially focussing on nucleobase derivatives and nucleotide metabolism. Two approaches were chosen to achieve this: 1)

classical pharmacology with phenotypic assays, screening compound libraries based on nucleobase derivatives against the parasites in culture; and 2) target-based drug design.

1.10 Target selection in Trypobase

Trypobase's target-based drug design approach initially focussed on the identification of inhibitors of the TriTryp dUTPases, essential enzymes for the parasites' survival. The dUTPases of the TriTryps belong to a distinct enzyme family from the human dUTPase and therefore provide the opportunity to selectively target the parasite enzymes, leaving the human counterpart unaffected. Unfortunately, the high specificity of these dUTPases for their substrates made drug candidate design and identification a difficult task. I joined the Trypobase project 18 months after its start, when other targets from nucleotide metabolism were included in the project to increase the chances of identification of drug candidates against the TriTryps. The nucleotide metabolism of the TriTryps has unique features compared to human, e.g. the parasites lack the *de novo* synthesis pathway for purines, making them depend entirely on the salvage pathway. Also, cytotoxic nucleoside analogues have been well characterised and established as treatments against several viral pathogens, but usually require activation *via* phosphorylation by enzymes from nucleotide metabolism. The differences in substrate specificity of human and pathogen enzymes are exploited leading to pro-drug activation only in infected cells. Because of the successful application of nucleoside-based pro-drugs against viral infections and differences of the pathways between human and TriTryps, the pro-drug activating enzymes adenosine kinase (AK), thymidine kinase (TK) and thymidylate kinase (TMPK) were chosen as potential targets.

In addition to the target-based approach, the Trypobase project applied phenotypic screens to identify compounds with trypanocidal activity. After extensive screening, a lead compound against *T. cruzi* parasites was identified and after investigation of the biological target using metabolomics, a possible mechanism of action (MOA) was suggested to be inhibition of a leucyl aminopeptidase, an enzyme shown

previously to be a potential drug target in *P. falciparum* (Harbut *et al.*, 2011). Following this hypothesis, the M17 family leucyl aminopeptidases of the TriTryps were included in the list of targets.

1.11 Aims

The initial work in this thesis was conducted as part of the Trypobase project, aiming to determine the X-ray structures of adenosine kinase (AK), thymidine kinase (TK), thymidylate kinase (TMPK) and leucyl aminopeptidase (LAP-A) from the TriTryps to aid target-based drug design. Additionally, this project was aimed to evaluate LAP-A as a potential drug target in the TriTryps, addressing the initial hypothesis of LAP-As as possible target for an anti-*T. cruzi* lead compound identified within the Trypobase project.

Chapter 2 Adenosine Kinase from *T. brucei brucei*

2.1 Introduction

Adenosine kinase (AK) catalyses the phosphorylation of adenosine (Ado) and deoxyadenosine (dAdo) to AMP or dAMP, respectively. This enzyme has been identified exclusively in eukaryotes with the exception of *Mycobacterium spp.* AKs of several species, including that of *T. brucei* (*Tb*AK), are well characterised and structures of the open (apo) form and the closed form in complex with a number of substrates and substrate analogues are available (Cook *et al.*, 2000, Kuettel *et al.*, 2011, Reddy *et al.*, 2007, Muchmore *et al.*, 2006, Mathews *et al.*, 1998, Schumacher *et al.*, 2000, Zhang *et al.*, 2007, 2006), e.g. *Toxoplasma gondii* AK, Figure 2-1. The protein is comprised of two α/β-domains, a large domain and small lid domain, highlighted in Figure 2-1 A and B. The active site is located between the two domains, which close upon substrate binding.

The AKs belong to the phosphofructokinase B family of carbohydrate kinases with two characteristic sequence motifs: the di-glycine switch close to the N-terminus and the signature sequence DXNGAGD close to the C-terminus (Park & Gupta, 2008), highlighted in Figure 2-1 C in dark blue and green, respectively. The two catalytically active residues R136 and S318 are conserved in all AKs and are highlighted in red. In solution, AKs are monomeric, with the exception of *M. tuberculosis* AK, which was shown to be dimeric (Reddy *et al.*, 2007).

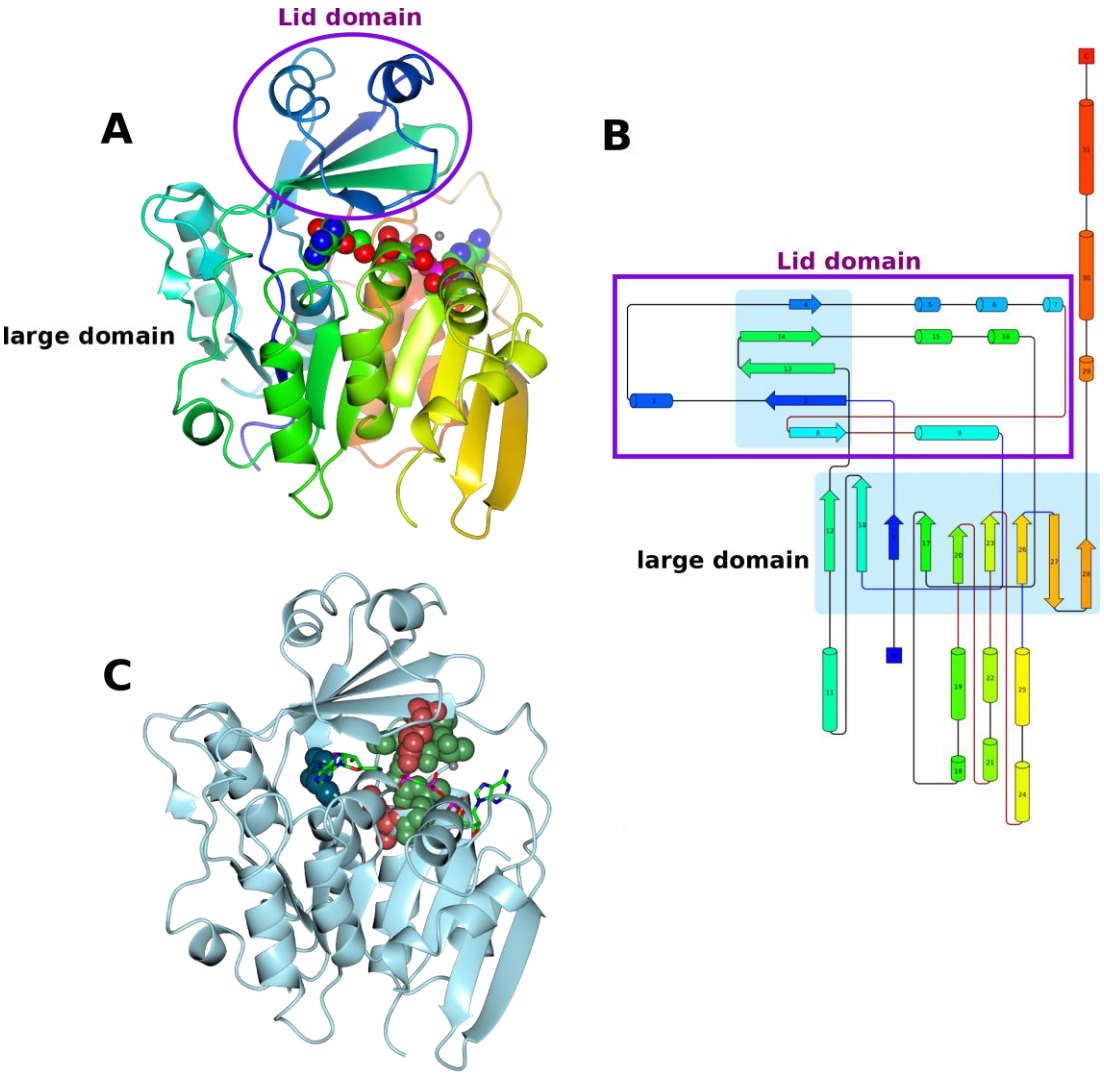


Figure 2-1: Overall view of adenosine kinase.

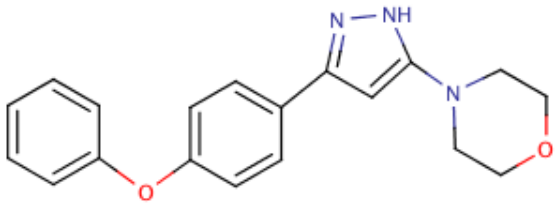
Shown is the adenosine kinase from *T. gondii* (*Tg*AK). (A) *Tg*AK-Ado-AppCp in ribbon representation coloured as rainbow from N-terminus (blue) to C-terminus (red), the ligands as spheres coloured by atom type. (B) Protein cartoon representation of *Tg*AK-Ado-AppCp coloured as rainbow with β-strands as arrows and α-helices as tubes. The β-sheets are underlaid in light blue. Cartoon generared with Pro-origami (Stivala *et al.*, 2011). (C) *Tg*AK-Ado-AppCp in ribbon representation with catalytically important residues (red), di-glycine switch (dark blue) and signature sequence (lawn green) as spheres. (A) and (C) were made with CCP4*mg* (McNicholas *et al.*, 2011).

2.1.1 Adenosine Kinase from *T. brucei brucei*

The genome of *T. brucei brucei* encodes two ORFs for AK, Tb927.6.2360 (*Tb*AK-1) and Tb927.6.2300 (*Tb*AK-2), each encoding a protein of 345 amino acid residues (37.9 kDa). The coding sequences differ in only eight nucleotides, resulting in just

four amino acid substitutions near the C-termini of the respective enzymes. The two genes are located on chromosome 6 in close proximity, probably the result of a tandem-duplication of a 7.8 kbp region including the *Tb*AK gene and four ORFs of unknown function. Both genes are transcribed in procyclic as well as bloodstream parasites and the produced proteins localise to the cytoplasm (Luscher *et al.*, 2007). This work focuses on *Tb*AK-1, the subject of most studies published to date. For simplicity, *Tb*AK-1 will be referred to as *Tb*AK in this thesis except when compared directly to *Tb*AK-2. Comparison of AKs from different species reveals a rather low sequence identity with 62 % to *T. cruzi* AK, 55 % to *L. major* AK and 41 % to *Hs*AK, Figure 2-2. Sequence similarities span the whole length of the protein with highest conservation in regions important for the function, e.g. the di-glycine switch, the P-loop and the adjacent helix, substrate coordinating and catalytically essential residues. Loop regions have low sequence similarity as expected.

AK is a non-essential enzyme in *T. brucei* as was shown by RNAi-mediated depletion of AK (Luscher *et al.*, 2007). This is not surprising as AMP can be produced not only by Ado phosphorylation, but also using the PNH-APRT pathway or from IMP using the ADSS-ASL pathway (as explained in more detail in section 1.8.1.1). However, hyper-activation of AK seems to have trypanocidal effects *in vitro* (Kuettel *et al.*, 2009). 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine (KRM) and some of its derivatives were shown to be trypanocidal (Kuettel *et al.*, 2007) and after identification of *Tb*AK as the protein binding to these compounds, it was shown that they increase *Tb*AK's activity and abolish substrate inhibition. The crystal structure of AK from *T. brucei rhodesiense* in complex with the activator (2xtb, Kuettel *et al.*, 2011) shows one molecule bound per AK chain, its binding site overlapping with the Ado binding site.



4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl] morpholine

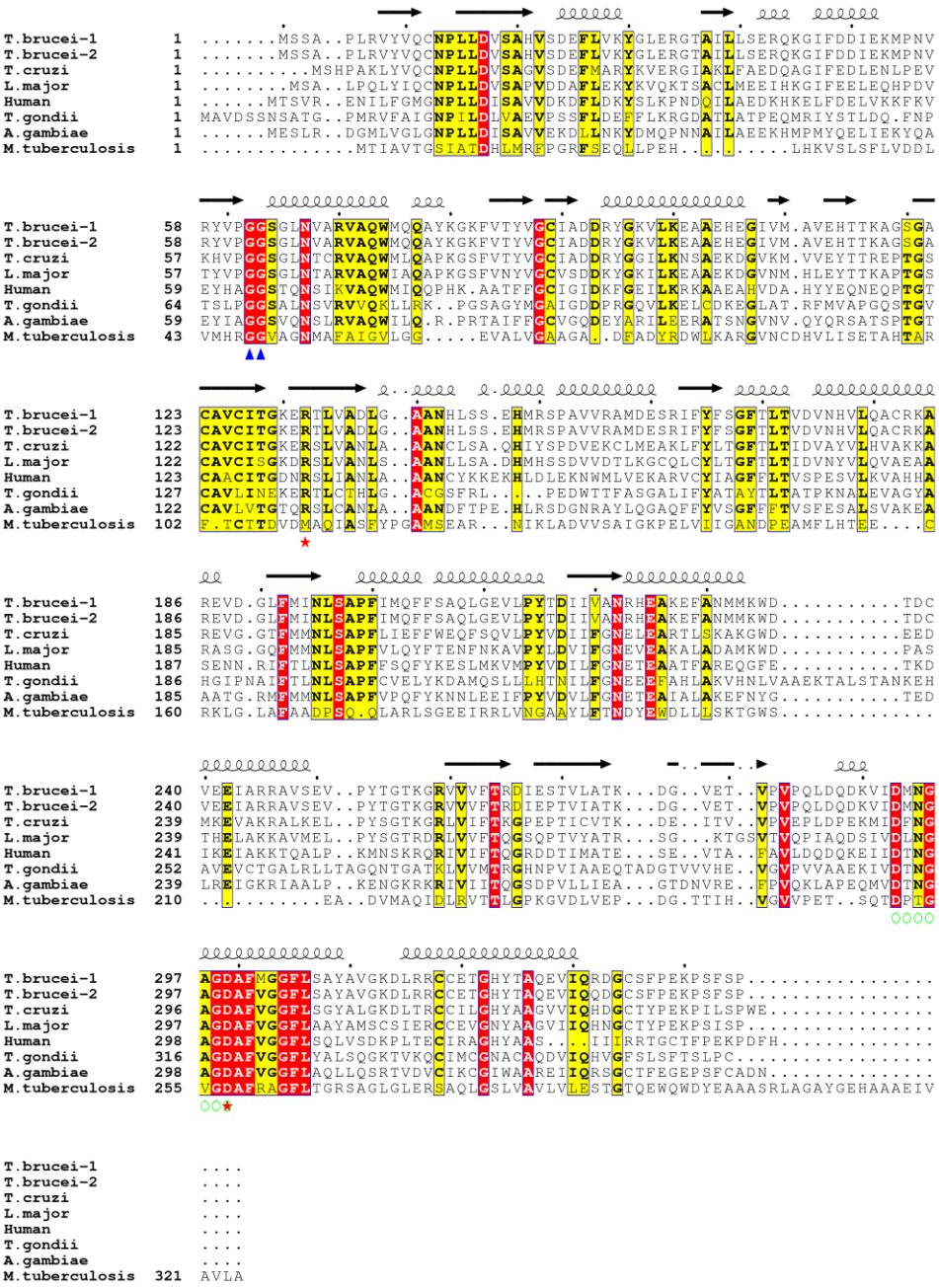


Figure 2-2: Structure-based alignment of *Tb*AK sequences 1 and 2 with other AKs.

Conserved residues highlighted in red, similar residues in yellow. The secondary structure elements (from *Tb*AK structure) are displayed above the sequences: spiral = α-helix, arrow = β-strand. The di-glycine switch is marked with blue triangles below the sequences, catalytically important residues with red stars and the P-loop residues with green circles. The alignment was generated using Expresso (from T-Coffee (Armougom *et al.*, 2006) and Espript 3.0 (Gouet *et al.*, 2003).

The preferred substrate of *Tb*AK is Ado, but it also phosphorylates dAdo, inosine and nucleoside analogues like cordycepin, Ara-A and F-Ara-A. It tolerates all NTPs as phosphate donors, but clearly prefers ATP and GTP as indicated by their lower K_M values (Vodnala *et al.*, 2008). *Tb*AK's activity is stimulated (increase in V_{max}) by inorganic phosphate without affecting the affinity for Ado. Like mammalian, yeast and plant AKs, *Tb*AK shows substrate inhibition, while AKs of most other protozoan parasites, like *L. donovani,* do not show inhibition by high Ado concentrations (Datta *et al.*, 1987).

2.1.2 AK Mechanism

The mechanism of γ-phosphoryl transfer from the ATP to the Ado was described as an ordered bi-bi mechanism for the human enzyme (*Hs*AK, Mathews *et al.*, 1998), Figure 2-3. As the Ado binding site is much more deeply buried than the ATP-binding site, it can be assumed that Ado is the first substrate to bind.

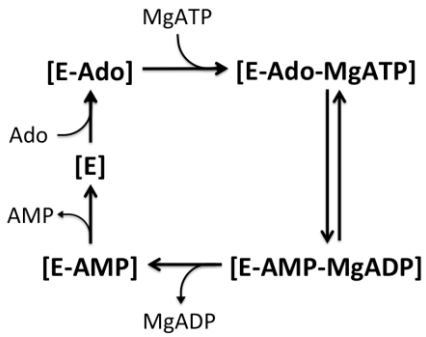


Figure 2-3: Scheme of the ordered bi-bi mechanism of AK.

The substrate binding site is conserved amoung different species, and the protein – substrate interactions are summarised in Figure 2-4. The proposed mechanism of AK based on the structural information of the active site of *Tg*TK is shown in Figure 2-5. The catalytic aspartate (D318 in *Tg*AK) acts as the catalytic residue and deprotonates the 5'-OH of the ribose ring of the Ado, which then performs a nucleophilic attack on the γ-phosphate of the phosphate donor, usually ATP. After formation of a transition state with a penta-coordinated γ-phosphate, the products

AMP and ADP leave the active site. An Mg^{2+} ion is essential for catalysis helping correct orientation of the substrate and acting as a counter ion.

AK shows strong, non-competitive substrate inhibition and it was proposed that the enzyme has an additional Ado-binding site differing from the substrate and triphosphate site, which was supported by chemical modification studies (Palella *et al.*, 1980, Hawkins & Bagnara, 1987). None of the structures published to date show a regulatory binding site for Ado, but that of *Hs*AK solved with Ados bound in both the Ado and the ATP-binding pockets, suggests a competitive inhibition mode of ATP binding by Ado (Mathews *et al.*, 1998).

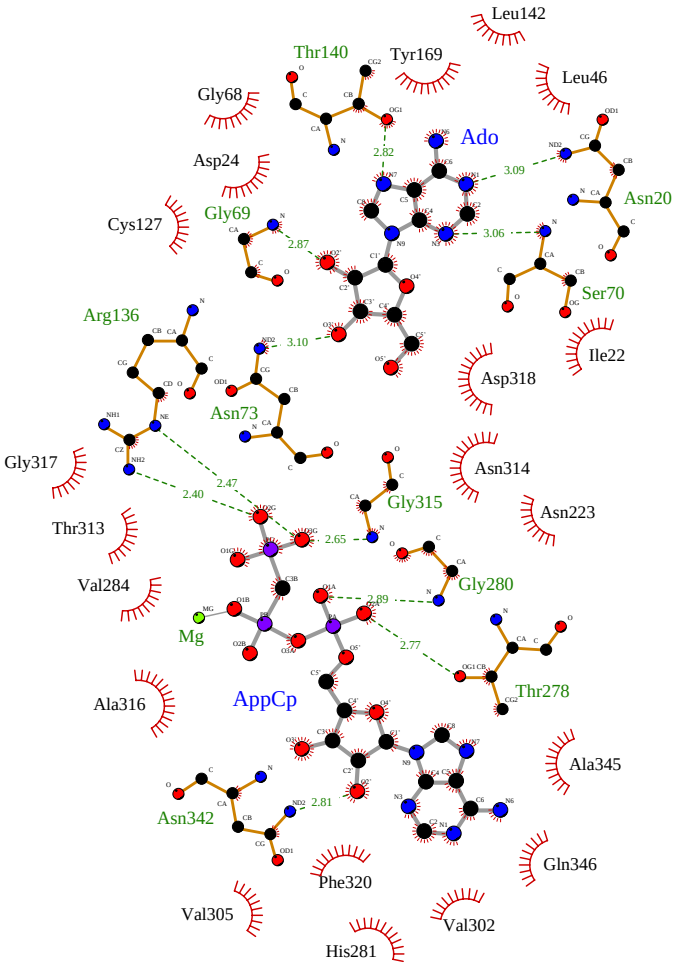


Figure 2-4: Schematic view of the substrate binding site of *Tg*AK with protein-substrate interactions.
Shown are the ligands and residues coordinating substrates *via* hydrogen bonds (green labels, hydrogen bonds as dashed lines), and hydrophobic interactions (black labels, red half circles). Ligand and hydrogen bonding residues are shown as ball and stick models coloured by atom type. Figure drawn with LigPlot+ (Laskowski & Swindells, 2011).

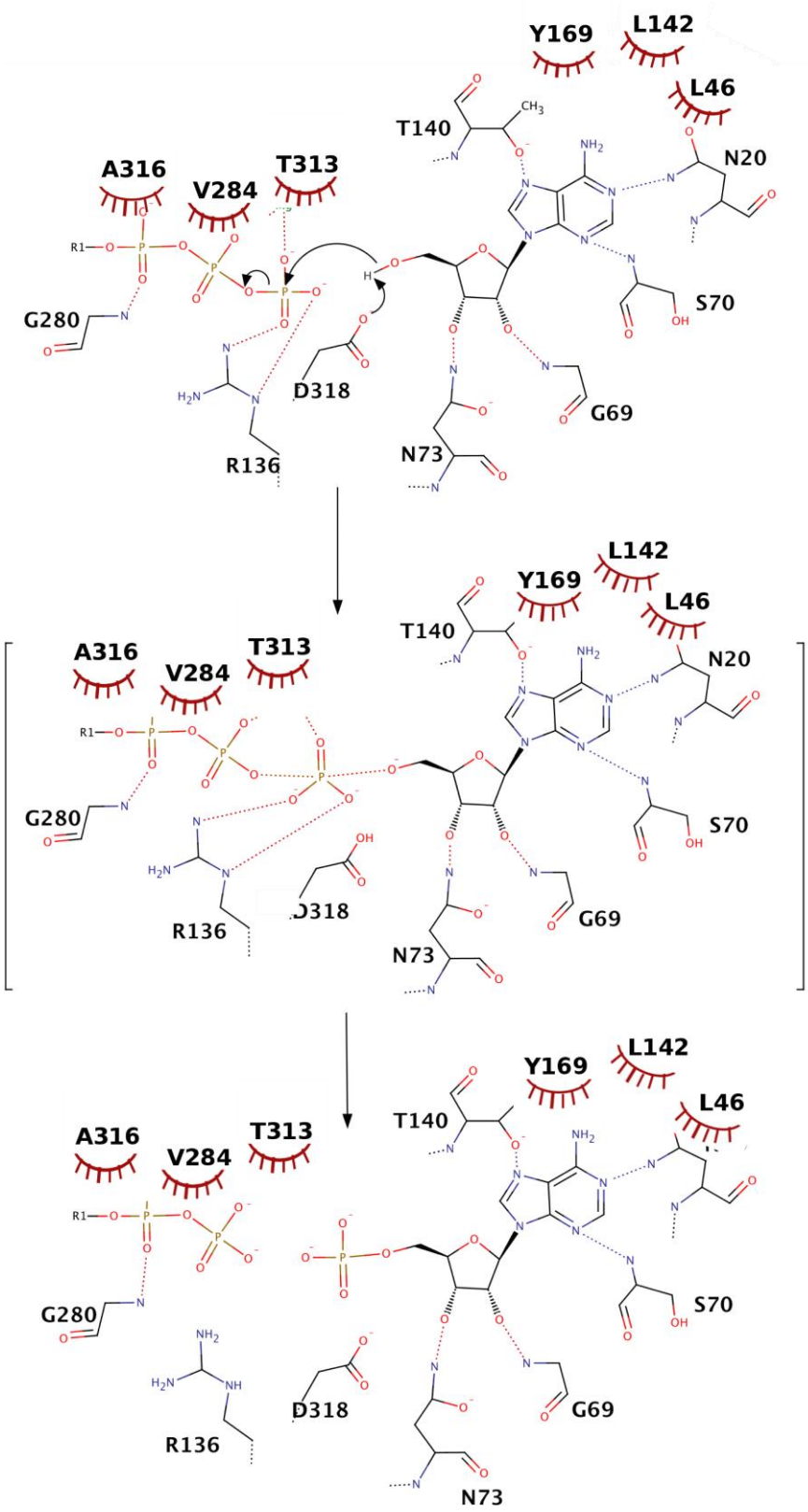


Figure 2-5: Mechanism of γ-phosphoryl transfer in AK.

R1 = adenosine moiety of the phosphate donor ATP, residue numbering after *Tg*AK. Hydrogen bonding indicated by dashed lines, hydrophobic interactions indicated in dark red. Figure made with MarvinSketch (ChemAxon).

2.1.3 Transition state analogues for phosphoryl transfer

Phosphoryl transfers in kinases happen *via* a planar, trigonal phosphoryl intermediate (middle panel in Figure 2-5, Cleland & Hengge, 2006, Rosta *et al.*, 2008). Compounds mimicking the reaction intermediate or transition states help to study the conformational changes an enzyme undergoes during the reaction. Transition state analogues with AlF_4^-, AlF_3^-, BeF_3^- and MgF_3^- mimicking the transferring phosphoryl group (Wittinghofer, 1997, Baxter *et al.*, 2006) have been used successfully for a variety of enzymes, e.g. the human phosphoglycerate kinase (*Hs*PGK), which has been co-crystallised with ADP, phosphogyceric acid (3PG) and AlF_4^-, AlF_3^- (+ H_2O) and MgF_3^-, Figure 2-6 (Cliff *et al.*, 2010). These analogues assemble spontaneously in the active site of the enzymes and lock them into place, therefore functioning as potent inhibitors. The MgF_3^- in the complex shows the typical trigonal bipyramidal geometry close to isosteric to a transferring phosphoryl group (Figure 2-6 A). AlF_4^- was shown to readily outcompete MgF_3^- (Baxter *et al.*, 2008) although it has non-isosteric octahedral geometry, like in Figure 2-6 B. Transition states with AlF_3^-, like in Figure 2-6 C, show the same geometry as AlF_4^- with a water molecule replacing one of the fluorines.

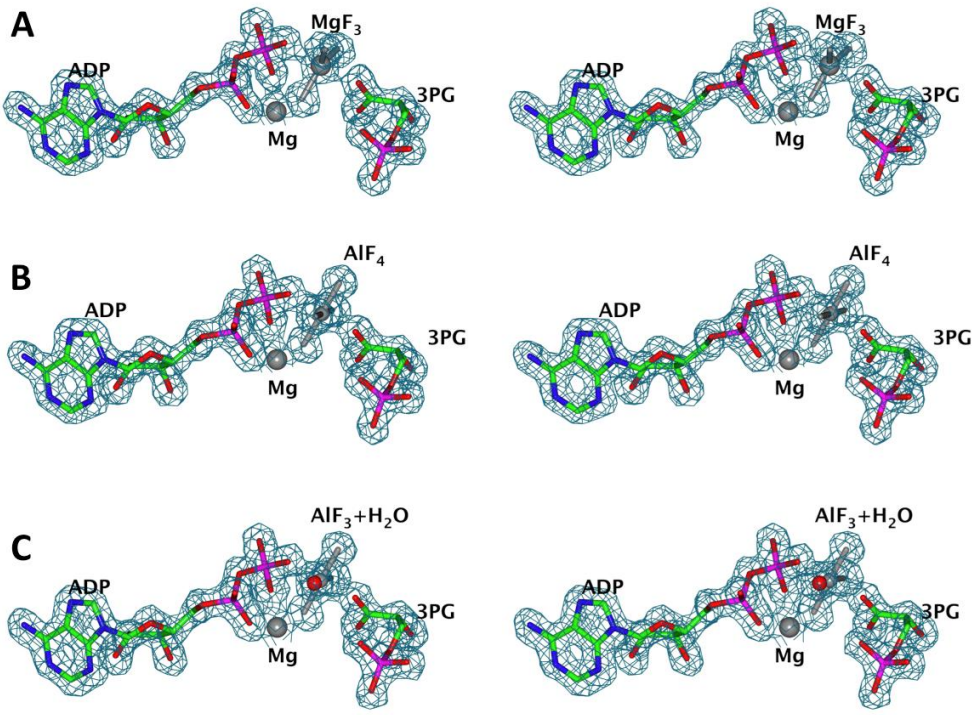


Figure 2-6: Transition state analogues in *Hs*PGK.

The ligands shown as cylinders, coloured by atom type, electron density map contoured at 1.5σ. (A) ADP-MgF_3-3PG +Mg, PDB ID: 2wzb. (B) ADP-AlF_4-3PG +Mg, PDB ID: 2wzc. (C) ADP-AlF_3-3PG +Mg, PDB ID: 2wzd. Figure made with CCP4*mg* (McNicholas *et al.*, 2011).

2.1.4 Aims

The aim was to determine the X-ray crystal structures of *T. brucei brucei* AK in its apo form and in complex with its substrate Ado plus the non-hydrolysable ATP analogue AppNHp. Insights gained from these structures and the structural information available for human ADA should lead to a better understanding of cordycepin modifications possible to improve its stability against deamination while retaining its trypanocidal properties.

2.2 Experimental

2.2.1 Cloning of *Tb*AK constructs

The gene encoding the AK from *T. brucei brucei* (Tb927.6.2360) was amplified by PCR from genomic DNA using the reaction conditions and temperature cycles detailed in Table 2-1 and Table 2-2. Primer sequences are listed in appendix A.2.3. The size of the amplified gene was checked by agarose gel electrophoresis (Figure 2-7) and then cloned into the YSBLIC and YSBLIC3C vectors using ligation-independent cloning (Aslanidis & de Jong, 1990, Fogg *et al.*, 2006).

Table 2-1: PCR reaction mix used for *Tb*AK.

Reagent	**Volume/reaction [µL]**
10 x Pfu buffer (incl. $MgSO_4$, Promega)	5
dNTP mix (10 mM each)	1
Forward primer (10 µM)	2.5
Reverse primer (10 µM)	2.5
ddH_2O	38.5
Pfu polymerase (2.5 U/µL, Promega)	0.5
Template (genomic DNA)	1 (= ~100 ng)
Total volume	**50**

Table 2-2: Temperature profile of the PCRs used for *Tb*AK.

Step	**Temperature [°C]**	**Time [s]**	**No. of cycles**
Initial denaturation	95	2	1
Denaturation	95	30	30
Annealing	60	30	
Extension	72	90	
Final extension	72	300	1

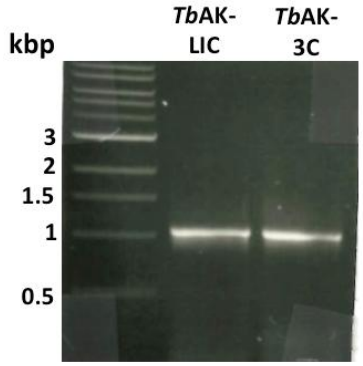


Figure 2-7: Agarose gel of *Tb*AK amplification.

Loaded onto a 1.5 % agarose gel: 1Kb ladder (lane 1), *Tb*AK gene for YSBLIC (lane 2) and *Tb*AK gene for YSBLIC3C (lane 3).

In brief, long overhangs on linearised vector and insert DNA fragments are created using the T4 DNA polymerase (reaction conditions specified in Table 2-3). Subsequently the T4 treated vector was purified using a Miniprep Kit (Qiagen) and stored in aliquots at -20°C, the insert was not purified. Insert and vector are then annealed by incubation of 2 µL insert (2 fmol) and 1 µL purified vector (0.3 fmol), both with long overhangs, for 10 min at room temperature before addition of 1 µL EDTA (6.25 mM) and another incubation for 10 min at room temperature. Chemically competent *E. coli* were then transformed with the mixture (4 µL total) where the construct is ligated.

Table 2-3: Reaction conditions of the T4 reaction.

Reagent	**Vector**	**Insert**
10x T4 pol. Buffer (Promega)	40 µL (1x)	2 µL (1x)
DTT (100 mM)	20 µL (0.5 mM)	1 µL (0.5 mM)
dTTP (25 mM)	2 µL (2.5 mM)	-
dATP (25 mM)	-	2 µL (2.5 mM)
DNA (linearised)	24 µL = 4 pmol	1.8 µL = 0.2 pmol
ddH_2O	268 µL	12.8 µL
T4 polymerase (10 U/µL, Promega)	8 µL = 80 units	0.4 µL = 4 units
Total volume	**400 µL**	**20 µL**
→ Incubation for 30 min at room temperature		
→ Incubation for 20 min at 75°C		

The generated construct, named *Tb*AK-LIC, encodes a Histag fusion to the N-terminus of the protein, whereas the construct *Tb*AK-3C contained a N-terminal Histag connected *via* a HRV 3C protease recognition site for subsequent removal of the tag. The cloning success was confirmed by colony PCR. For this colonies were picked and suspended in 50 µL ddH_2O. One µL was used to inoculate 1 mL LB cultures for plasmid propagation, the remaining 49 µL were heated at 95°C for 5 min and centrifuged for 2 min before the PCR. The PCR reaction was carried out using the conditions in Table 2-4 and Table 2-5. The positive clones were propagated and further verified by sequencing.

Table 2-4: Reaction mix for colony PCRs.

Reagent	**Volume/reaction [µL]**
10 x Pfu buffer (incl. $MgSO_4$, Promega)	5
dNTP mix (10 mM each)	1
T7 promotor primer (10 µM)	2.5
Reverse primer (10 µM)	2.5
ddH_2O	29.5
Pfu polymerase (2.5 U/µL, Promega)	0.5
Template (boiled colony in ddH_2O)	10
Total volume	**50**

Table 2-5: Temperature profile of the colony PCRs.

Step	**Temperature [°C]**	**Time [s]**	**No. of cycles**
Initial denaturation	95	2	1
Denaturation	95	30	30
Annealing	50	30	
Extension	72	90	
Final extension	72	300	1

2.2.2 Expression of *Tb*AK

For initial expression testing chemically competent *E. coli* BL21 gold (DE3), C41 and pLEMO cells were transformed with *Tb*AK-LIC and *Tb*AK-3C and plated on Luria-Bertani Miller (LB) plates containing 35 µg mL^{-1} kanamycin. The next day colonies were washed off the plates with 5 mL LB medium and these resuspended colonies were used for inoculation of new 5 mL cultures. The cultures were grown at 37°C to an optical density of 0.8 at 600 nm. One culture of each strain was transferred to

20°C and expression in all cultures was induced by addition of 1 mM isopropyl thio-β-*D*-galactoside (IPTG). The expression culture at 37°C was left for 4 h, the expression at 20°C was left overnight. After SDS-PAGE analysis of lysed 1 mL samples of expression culture, the expression of both constructs was found to be best in C41 cells.

For large scale expression the constructs were transformed into C41 cells as in the expression test and subsequently 8 x 500 mL LB was inoculated with cells washed off the plates of the fresh transformants. The cells were grown at 37°C with shaking at 180 rpm (Innova44, New Brunswick Scientific) until they reached an optical density of 0.2 at 600 nm and then transferred to 20°C with shaking at 251 rpm. Protein expression was induced by addition of 1 mM IPTG at an optical density of 0.6 at 600 nm for 6 h. Cells were harvested by centrifugation at 5000 x g for 30 min and washed with 100 mL PBS prior to freezing and storage at -20°C.

2.2.3 Purification and characterisation of *Tb*AK

The purification protocol for both *Tb*AK constructs used the same buffers and general procedure. Reducing agents were added to all buffers. Initial purification buffers contained 10 mM β-mercapto-ethanol and 100 µM PMSF, but later the reducing agent for the *Tb*AK-3C buffers was changed to 1 mM DTT and PMSF was left out (details discussed in 2.3.1).

Cell pellets were thawed on ice, resuspended in 80 mL lysis buffer A (20 mM Tris pH 7.5, 200 mM NaCl, 30 mM imidazole, 10 mM β-mercapto-ethanol and 100 µM PMSF) or B (20 mM Tris pH 7.5, 200 mM NaCl, 30 mM imidazole and 1 mM DTT) and lysed by sonication. Cell debris was removed by centrifugation at 5000 x g and 4°C for 40 min. The cleared lysate (supernatant) was loaded onto a pre-equilibrated 5 mL HisTrap FF crude column (GE Healthcare), which was then washed with three column volumes of lysis buffer (A or B respectively). The protein was eluted using an imidazole gradient from 30 mM to 500 mM imidazole (Figure 2-8 A). For purification of the *Tb*AK-3C the protein-containing fractions, identified by SDS-PAGE

(Figure 2-8 B) were pooled and concentrated to ~ 2 mg mL^{-1} before addition of HRV 3C protease in a ratio of 1:100 (protease:*Tb*AK). Proteolysis was monitored by 12 % SDS-PAGE and had reached completion overnight at 4°C. To remove the Histag and uncleaved *Tb*AK the protein buffer was exchanged into the lysis buffer using Centrifuge filters (Amicon® Ultra Centrifugal Filter from Millipore or Vivaspin6 from SartoriumStedim biotech, 30 kDa cutoff) and another nickel affinity purification was performed using the same buffers as before, resulting in untagged protein in the flowthrough (Figure 2-8 C). The resulting protein was concentrated and loaded onto a Superdex 75 column (GE Healthcare), in buffer containing 10 mM Tris pH 7.5, 200 mM NaCl and 10 mM β-mercapto-ethanol (BME) or 1 mM DTT as reducing agent (Figure 2-8 D). For purification of *Tb*AK-LIC the protein-containing fractions of the nickel affinity chromatography were pooled, concentrated and directly loaded onto a Superdex 75 column (without preceding proteolysis). Protein purity was monitored by SDS-PAGE (Figure 2-8 E).

Because of the higher degree of purity and better behaviour in the initial crystallisation screen, the main focus was set on the sample with the cleavable Histag.

After purification, *Tb*AK was concentrated to an approximate volume of 1 mL and dialyzed (12-14 kDa cutoff) against the size exclusion chromatography buffer overnight at 4°C. Subsequently, the protein was flash cooled in liquid nitrogen and stored at -80°C at a concentration of 46, 50 or 67 mg mL^{-1} dependent on the purification batch.

To confirm the identity of the purified protein, electrospray ionisation mass spectrometry (ESI-MS) was used. Before crystallisation experiments were carried out, the polydispersity of *Tb*AK was determined by dynamic light scattering (DLS). For this the protein was diluted to a concentration of 3 mg mL^{-1} at 20°C with a total sample volume of 20 µL. DLS measurements were carried out using a DynaPro™ (Protein Solutions™).

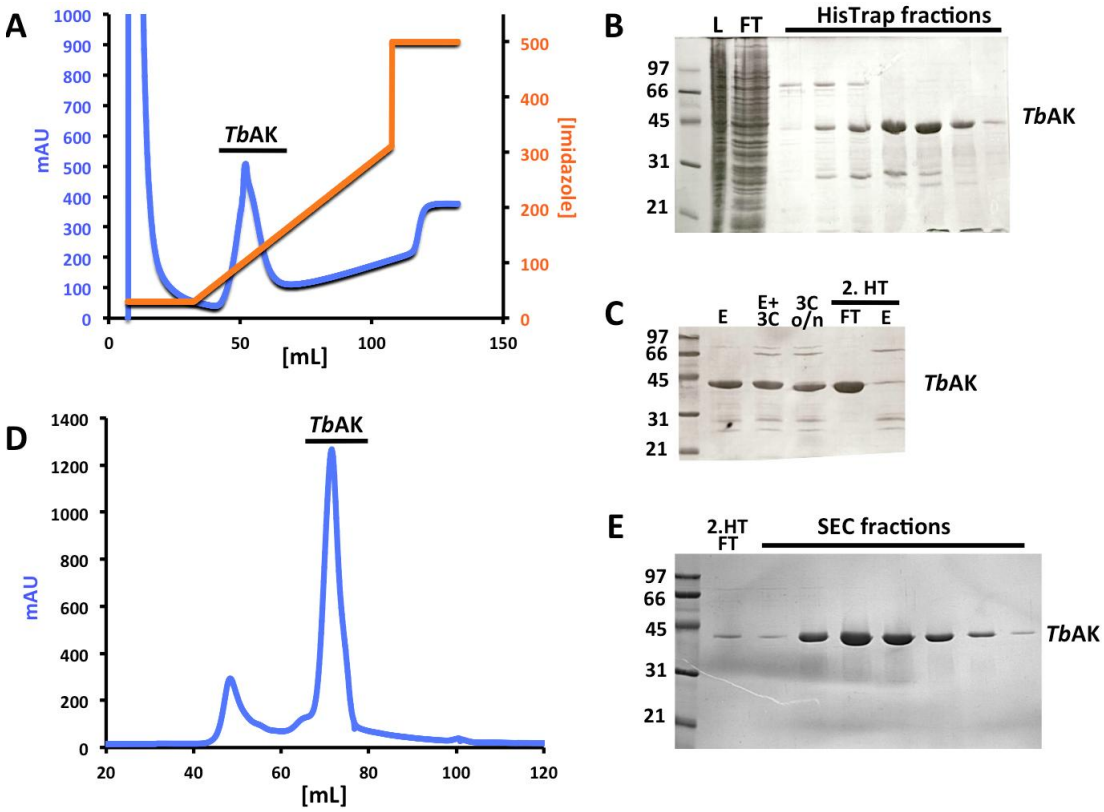


Figure 2-8: Purification of *Tb*AK.

(A) Nickel affinity chromatography trace from a 5 mL HisTrap crude column. UV trace shown in blue, imidazole concentration shown in orange, black bar indicates the fractions analysed by SDS-PAGE. (B) SDS gel after nickel affinity chromatography with L = cell lysate and FT = flowthrough. (C) SDS-gel after 3C digest and subsequent nickel affinity chromatography. E = TbAK-containing fractions of first HisTrap, 3C o/n = after o/n incubation with 3C protease, 2. HT = nickel affinity chromatography after digest, FT = flowthrogh. (D) Size exclusion trace from a 10/300 Superdex 75 column. The black bar indicates the elution volume of *Tb*AK, which was confirmed by SDS-PAGE. (E) SDS gel after size exclusion chromatography showing *Tb*AK as a single band without degradation or contamination. In all gels: lane 1 = low molecular weight marker with bands for 97, 66, 45, 31, 21 and 14 kDa sized proteins.

2.2.4 Crystallisation of *Tb*AK

The initial crystallisation screens, Index and PACT, were set up at *Tb*AK concentrations of 46 mg mL^{-1} with and without ligands, using sitting drop vapour diffusion at 20°C. Drops of 150 nL protein solution + 150 nL well solution were pipetted using a Mosquito robot (TTP Labtech Ltd. UK). Ligands used were 1 mM Ado and 1 mM β,γ-imidoadenosine 5′-triphosphate (AppNHp), and both *Tb*AK constructs, but the work focussed on the protein with its Histag removed as the histagged *Tb*AK-LIC did not show any crystallisation. After the initial screening, the

concentration of untagged *Tb*AK purified in the presence of BME was lowered to 30 mg mL^{-1}. Crystal conditions were optimised using hanging drop vapour diffusion with 1 µL *Tb*AK plus 1 µL well solution. The best crystals in complex with Ado and AppNHp grew in 0.1 M Bis Tris propane pH 8.5, 0.2 M sodium citrate and 24 % (w/v) PEG 3350 within 3 weeks to a maximum size of ~ 600 x 100 x 50 μm^3. The crystal used for structure solution was cryo-protected in mother liquor containing 5 % ethylene glycol and vitrified prior to data collection.

For the apo-*Tb*AK further screening with several commercially available screens (PACT, Index, JCSG, Hampton 1 and 2 and CSS) did not lead to crystals. Following the publication of the structure of the *Tb*AK complex with the activator KRM in the open conformation (Kuettel *et al.*, 2011) the screening focussed around the conditions reported therein, using a protein sample purified with buffers containing DTT as reducing agent. This led to a large number of crystals of which 200 were tested for diffraction. However, very few showed any diffraction, and finally only one provided a usable diffraction pattern with the rest showing very restricted resolution or substantial splitting and disorder. The single usable crystal grew at 4°C in 0.1 M Tris pH 8.8 and 60 % (v/v) MPD, and reached a maximum size of ~ 80 x 20 x 20 μm^3 after one week, and was not cryo-protected before vitrification.

Interestingly, the choice of reducing agent in purification appeared to have a strong influence on the crystallisation behaviour. While *Tb*AK purified in the presence of BME crystallised only in the closed conformation (with ligands added), *Tb*AK purified with DTT crystallised only in the open conformation (without ligands). Additionally, repetition of purification under the same or modified conditions did not lead to diffracting crystals, and even after extensive screening it was not possible to obtain crystals with improved diffraction. Extensive screening in order to crystallise *Tb*AK in complex with cordycepin and the transition state analogue ADP:MgF_4:Ado were unsuccessful. Crystallisation screening of *Tb*AK in complex with ADP:AlF_4:Ado, followed by optimisation with several rounds of seeding produced crystals diffracting to ~ 6 Å. Further optimisation did not improve the resolution.

2.2.5 Data collection, structure solution and refinement of *Tb*AK

Computations were made using programs from the CCP4 software package (Winn *et al.*, 2011) unless otherwise stated. The data were processed with XDS (Kabsch, 2010a, b), POINTLESS (Evans, 2011) and SCALA (Evans, 2006).

The X-ray diffraction data for both crystals were collected at the Diamond Light Source (DLS). The *Tb*AK-Ado-AppNHp crystal was in space group *C2* with four molecules of *Tb*AK in the asymmetric unit. The apo crystal was split, which limited data quality, and belonged to space group $P4_12_12$ with one molecule of *Tb*AK per asymmetric unit. The crystallographic data and statistics are summarised in Table 2-6.

The structures were solved by molecular-replacement using PHASER (McCoy, 2007). At the time, *Hs*AK (PDB ID: 1bx4) with a protein sequence identity of 40 % showed the highest similarity to *Tb*AK and was used as the search model for the *Tb*AK-Ado-AppNHp. For apo *Tb*AK the structure of the *Tb*AK-KRM (PDB ID: 2xtb) was used as the search model.

Structure refinement was carried out using maximum-likelihood restrained methods in REFMAC5 (Murshudov *et al.*, 2011) with automatically generated local NCS. The manual fitting of the electron-density maps, model building and fitting of the ligands were carried out in COOT (Emsley *et al.*, 2010). The models were validated using COOT and *MolProbity* (Chen *et al.*, 2010).

Table 2-6: Crystallographic data and statistics of the *Tb*AK crystals.

Data for the highest resolution shell shown in brackets, the Ramachandran Plot was calculated using COOT (Emsley *et al.* 2010), the poor rotamer analysis was carried out in *Molprobity* (Chen *et al.* 2010).

	apo *Tb*AK	***Tb*AK-Ado-AppNHp**
Data collection		
Diffraction Source	DLS beamline I03	DLS beamline I24
Wavelength (Å)	0.9763	0.9686
Temperature (K)	100	100
Detector	Pilatus 6M-F	Pilatus 6M Prosport+
Rotation range per image (°)	0.2	0.2
Total Rotation range (°)	180	220
Crystal data		
Space group	$P4_12_12$	C2
a, b, c (Å)	60.08, 60.08, 192.86	184.53, 90.22, 98.04
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 103.78, 90.00
Resolution (Å)	57.36 - 2.60 (2.74-2.60)	95.21 - 2.60 (2.74-2.60)
Total Reflections	132048 (20863)	160133 (23793)
Unique Reflections	11649 (1664)	47962 (6973)
Completeness (%)	99.90 (100)	99.5 (99.4)
Redundancy	11.3 (12.5)	3.3 (3.4)
R_{merge}	0.120 (0.395)	0.176 (0.991)
[I/σ (I)]	15.8 (6.3)	6.6 (2.3)
Matthew's coefficient	2.28	2.59
Refinement statistics		
Reflections, working set	10989	45507
Reflections, test set	660	2455
Resolution Range (Å)	57.36 - 2.60	95.21 - 2.60
R-factor	0.170	0.186
R_{free}	0.231	0.221
No. of non-H atoms		
Protein	2463	10276
Ligand	-	192
Water	157	148
Mean B factors (Å^2)		
Protein	38.0	45.2
Ligands	-	44.4
Water	40.5	38.3
RMS deviation from ideal		
Bond length (Å)	0.013	0.012
Bond angles (°)	1.466	1.531
Ramachandran Plot (%)		
Residues in favored region	93.6	96.05
Residues in allowed region	5.8	3.58
Outliers	0.6	0.37
Molprobity score	1.35	1.45
Poor Rotamers (%)	1.61	2.01
PDB ID	**4n08**	**4n09**

2.3 Results

2.3.1 Purification and Characterisation of *Tb*AK

*Tb*AK was expressed in *E. coli* C41 cells and purified using nickel affinity and size exclusion chromatography in sufficient quality and quantity for crystallisation experiments.

Initially, *Tb*AK was purified using buffers containing β-mercapto ethanol (BME) and PMSF for protease inhibition. However, characterisation by DLS revealed a polydispersity of approximately 75 % (Figure 2-9 A) and analysis by ESI-MS showed peaks of 80 and 156 Da higher than expected for the native protein, which could be the result of phosphorylation (80 Da) and a BME adduct (76 Da) or a PMSF adduct (156 Da) (Figure 2-9 B). This first batch was used for initial screening, which led to crystals of *Tb*AK grown in the presence of Ado and AppNHp.

After the characterisation and crystallisation of the first *Tb*AK preparation, the buffers used for purification were changed to avoid possible BME and PMSF adducts and reduce the polydispersity. DLS showed the protein to be monodisperse (10 % polydispersity) for this second batch (Figure 2-9 C) and ESI-MS indicated a single species with a M_W of 38187 Da (expected value 38184 Da for the protein with the extra three residues remaining at the N-terminus after cleavage of the Histag, Figure 2-9 D). This new *Tb*AK preparation was used for further crystallisation screening.

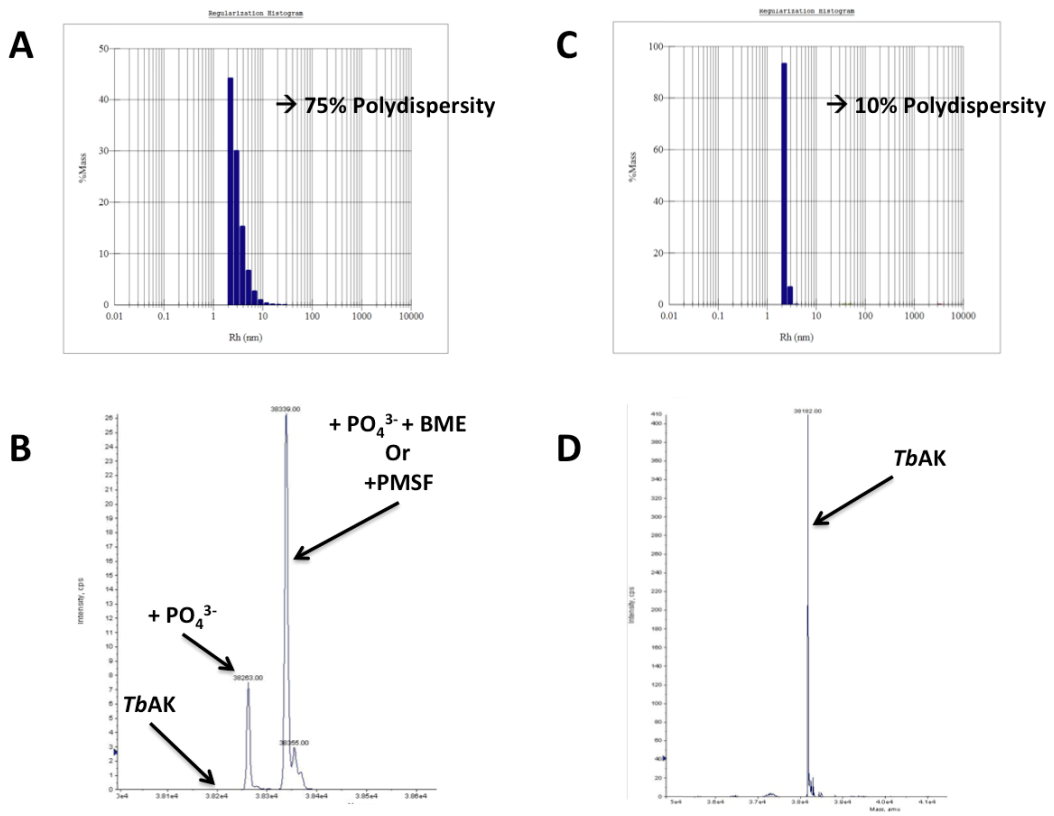


Figure 2-9: DLS and ESI-MS analysis of *Tb*AK purification batches.

(A) DLS Histogram and (B) ESI-MS trace of *Tb*AK purified with buffers containing BME as reducing agent. (C) DLS Histogram and (D) ESI-MS trace of *Tb*AK purified in presence of DTT as reducing agent.

2.3.2 Crystallisation of *Tb*AK

Crystals of *Tb*AK grown in the presence of Ado and AppNHp (Figure 2-10 A) using the protein purified with BME as reducing agent diffracted to 2.6 Å (Figure 2-11 A), and were in space group C2 with four *Tb*AK monomers in the asymmetric unit. All four chains had clear density for the Ados and the α- and β-phosphate of the AppNHp, but there was no density for the γ-phosphate and the ligand is therefore referred to as ADP in the figures. This could be the consequence of hydrolysis or of high flexibility of the γ-phosphate due to the lack of a coordinating Mg^{2+} ion. The crystallisation condition contained sodium citrate, which probably chelated any metal ions. There was no magnesium in the crystallisation buffer.

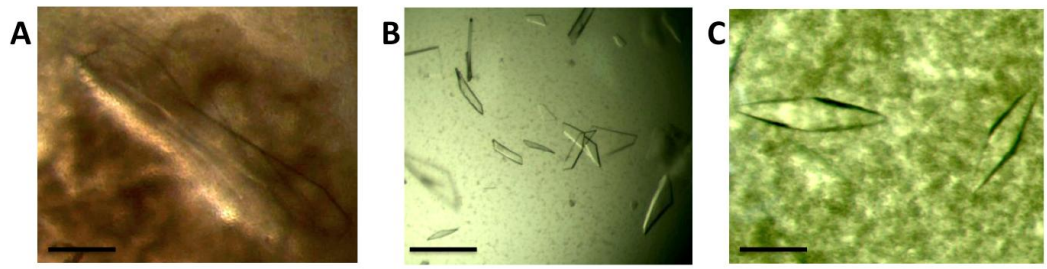


Figure 2-10: *Tb*AK crystals.

(A) *Tb*AK-Ado-ADP crystal used for structure determination. Black bar = ~ 200 µm. (B) Optimised *Tb*AK-ADP:AlF_4:Ado crystals after 2 rounds of seeding, diffracting to ~ 6 Å. Black bar = ~ 50 µm. (C) Apo *Tb*AK from the condition used to solve the structure. Black bar = ~ 40 µm.

Attempts to reproduce the *Tb*AK-Ado-AppNHp crystals failed with protein purified in the presence of DTT, as did attempts to co-crystallise *Tb*AK with cordycepin or with magnesium fluoride plus nucleotides. Crystallisation experiments of *Tb*AK in complex with ADP:AlF_4:Ado produced crystals (Figure 2-10 B), some of which diffracted to approximately 6 Å, but all showed multiple lattices. Further rounds of seeding and optimisation did not improve the resolution or diffraction quality. At this point, the structures of *Tb*AK in complex with the bi-substrate analogue diadenosine pentaphosphate (AP_5A) and the activator 4-[5-(4phenoxyphenyl)-2*H*-pyrazol-3-yl]morpholine (KRM) were published (Kuettel *et al.*, 2011). In order to obtain crystals of *Tb*AK in complex with the activator these authors crystallised the protein in its apo form and soaked in the activator. With the information about their crystallisation conditions it was possible to design an optimisation screen specifically for obtaining crystals of apo *Tb*AK using the protein from the DTT purification. Many single crystals grew (Figure 2-10 C) but showed no or very poor diffraction and inherent defects resulting in many overlapping lattices. However, one of the crystals diffracted up to 2.0 Å and allowed structure determination, although the diffraction was anisotropic and data quality remained rather poor with many overlapping lattices (Figure 2-11 B). The crystal was in space group $P4_12_12$ with one *Tb*AK molecule per asymmetric unit and the data were integrated to 2.6 Å.

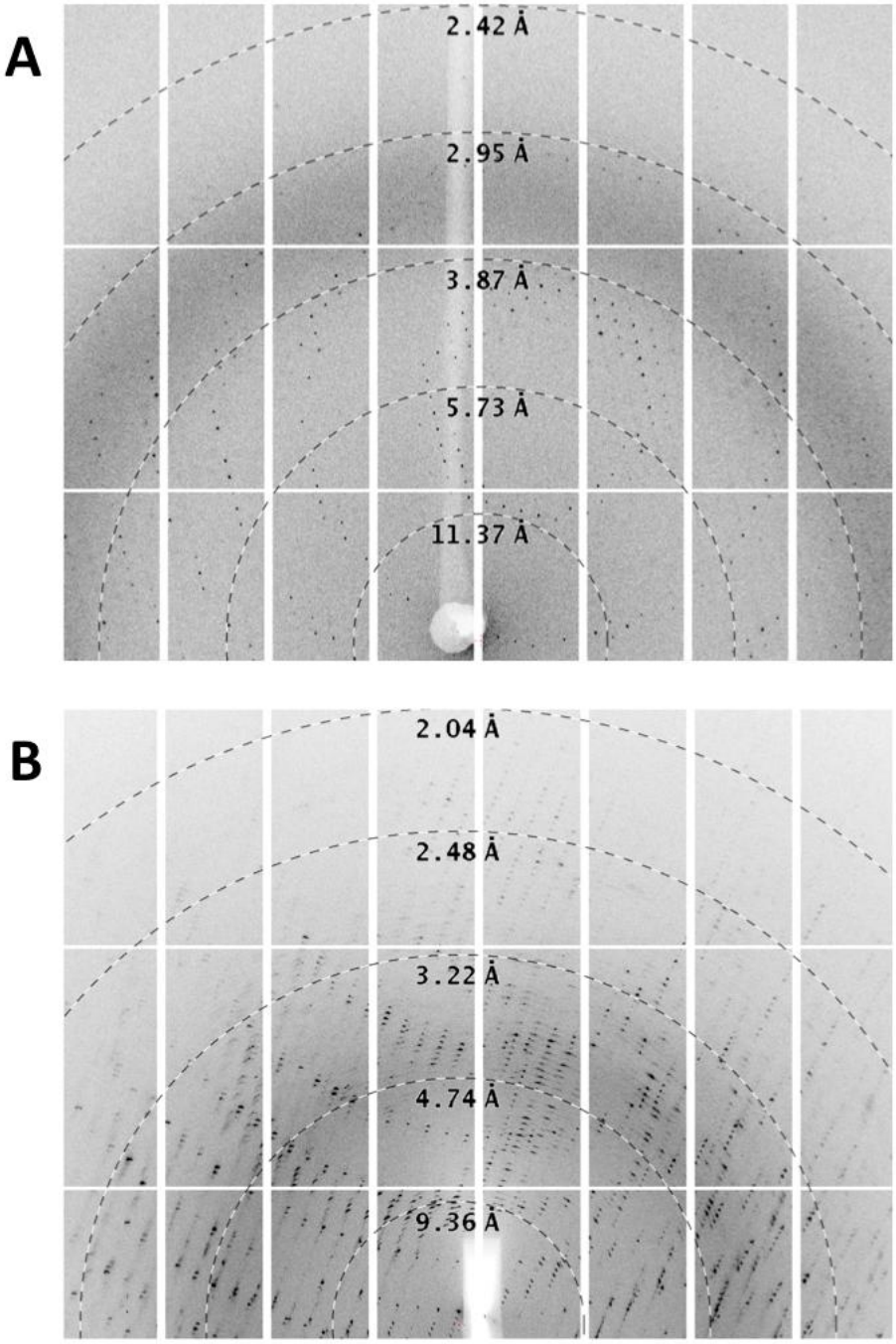


Figure 2-11: Diffraction patterns of *Tb*AK crystals.

This figure shows a close-up of images used for structure determination of the (A) *Tb*AK-Ado-ADP and (B) apo *Tb*AK structures with rings annotation the resolution. Although the diffraction of the apo *Tb*AK crystal seemed to extend to 2.0 Å, the diffraction was anisotropic and therefore only integrated to 2.6 Å overall. Figure was made using the *adxv* software (Arvai, 2012).

2.3.3 Structures of *Tb*AK

2.3.3.1 Overall fold

The structures of the *T. brucei brucei* enzyme reported here are highly similar to those from a closely related subspecies, *T. brucei rhodesiense* (Kuettel *et al.*, 2011), which has an identical amino acid sequence, the gene only differing by two silent mutations. Also they show only minor differences to the other published structures of AKs, like those from *Anopheles gambiae* (*Ag*AK, (Cassera *et al.*, 2011), human (*Hs*AK, (Mathews *et al.*, 1998, Muchmore *et al.*, 2006), *Toxoplasma gondii* (*Tgo*AK, (Cook *et al.*, 2000, Schumacher *et al.*, 2000, Zhang *et al.*, 2006, 2007) and *Mycobacterium tuberculosis* (*Mtu*AK, (Reddy *et al.*, 2007).

*Tb*AK shows the typical architecture consisting of two domains, a large ATP-binding α/β-domain and a small lid domain (Figure 2-12 A). The larger domain shows a α/β structure consisting of a central 9-stranded predominantly parallel β-sheet (β6, β5, β1, β9, β10, β11, β12, β13, β14) with β13 being the only antiparallel strand. This β-sheet is sandwiched between two layers of α–helices with 11 α–helices in total (α3-α13). The smaller lid domain consists of a 5-stranded β-sheet (β4, β2, β7, β8, β3) at the interface to the large domain and two α–helices (α1, α2) on top of it. Two alternative conformations are possible depending on the position of the lid domain, an open and a closed form (Figure 2-12 B). In the absence of substrate, AK is in the open conformation (apo), in presence of substrates the lid domain closes down, forming the closed conformation. The active site is located at the interface of the two domains.

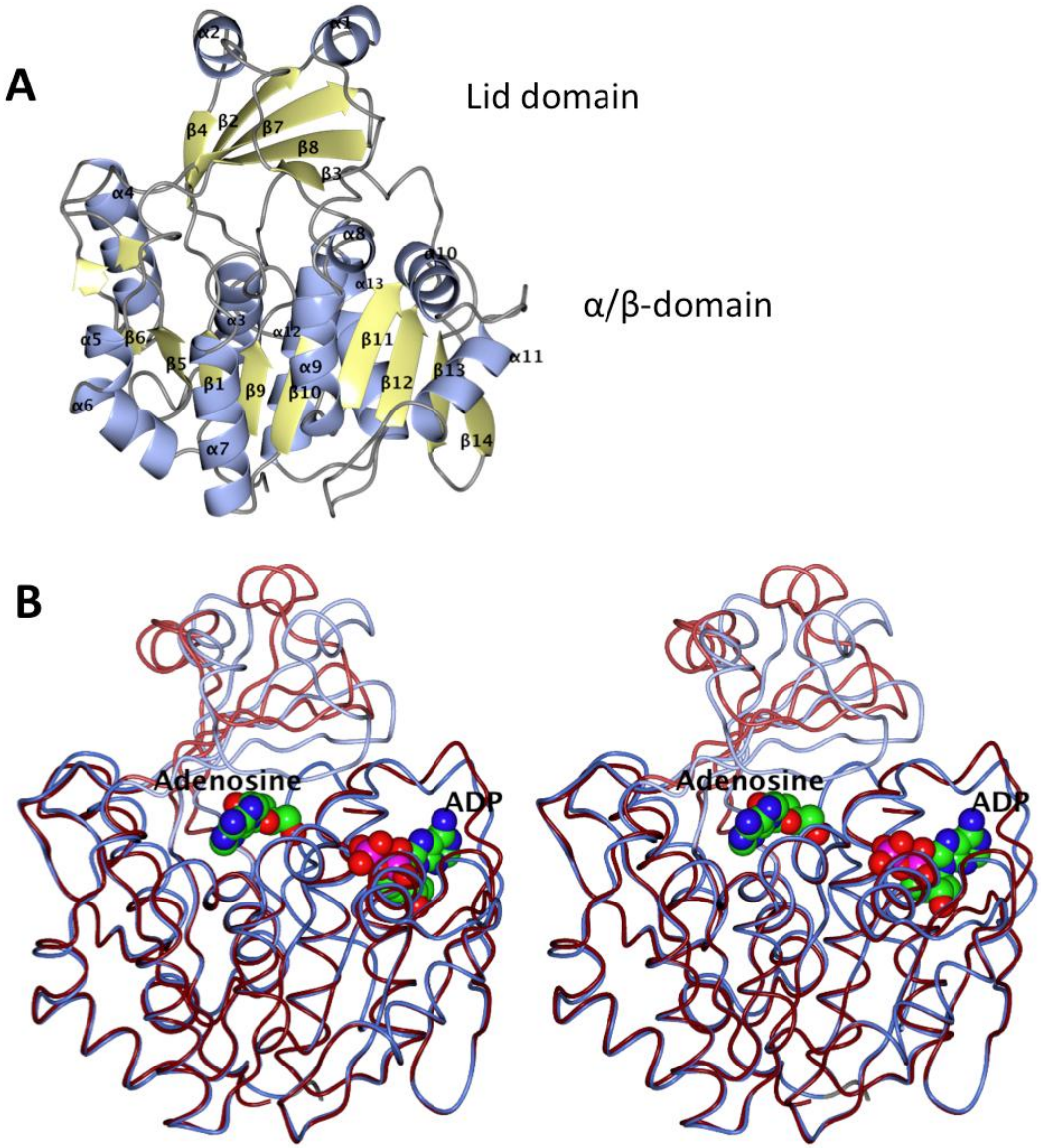


Figure 2-12: Overall structure of *Tb*AK.

(A) Structure in the closed conformation in ribbon representation coloured by secondary structure with α–helices in ice blue and β-strand in lemon. (B) Stereo view of a superposition structures in the open (tan for α/β-domain, dark crimson for lid domain) and closed (light blue and ice blue) conformations based on the core structure of the α/β-domain in worm representation, clearly showing the closure of the small lid domain. Ado and ADP bound in the active site are shown as spheres coloured by atom type. This and all subsequent structure figures were made using CCP4*mg* (McNicholas *et al.*, 2011) with all superpositions carried out using SSM (Krissinel & Henrick, 2004).

2.3.3.2 Substrate binding

In the structure co-crystallised with Ado and AppNHp, there are four independent protein molecules in the asymmetric unit, with well-defined density for both the Ado and the AppNHp missing the γ-phosphate. The four chains are very similar – the RMSD between all Cα atoms is 0.37, 0.39 and 0.29 Å for Chains B, C and D superposed on Chain A using SSM (Krissinel & Henrick, 2004). Chain A is therefore used in the rest of the discussion and other panels of the figures. *Tb*AK adopts a closed conformation similar to all other structures of AK complexes with substrates or substrate analogues in both of the binding pockets of the active site. Superposition of the structures of *Tb*AK in its closed and open forms confirmed that the smaller lid domain closes down over the active site and the large domain contracts slightly (Figure 2-12 B). This results from a rigid body rotation of 28° with hinge motions of the amino acids connecting the two domains. Among the critical residues for this motion are the two glycines of the di-glycine switch. In addition, *Tb*AK undergoes smaller structural rearrangements while binding to and orienting its substrates (Figure 2-13). The Ado forms hydrogen bonds to residues N13, D17, G63, N67, D299 and stacking interactions of its adenine ring with F169. The coordination of the 2'- and 3'-hydroxyl groups of its ribose ring to D17 assists in the closure of the lid domain. The ADP forms hydrogen bonds to N222, T264, T270, D266, G298 and Q288. The interactions with the ADP cause a slight closing of the loop between residues 266 and 268 and a rearrangement of the loop comprising residues 282-292. These smaller movements orient the substrates in the binding pockets and bury the two Ado moieties. Additionally, the 13 residues at the very N-terminus of the protein, disordered in the apo structure, form a loop behind the α13 helix.

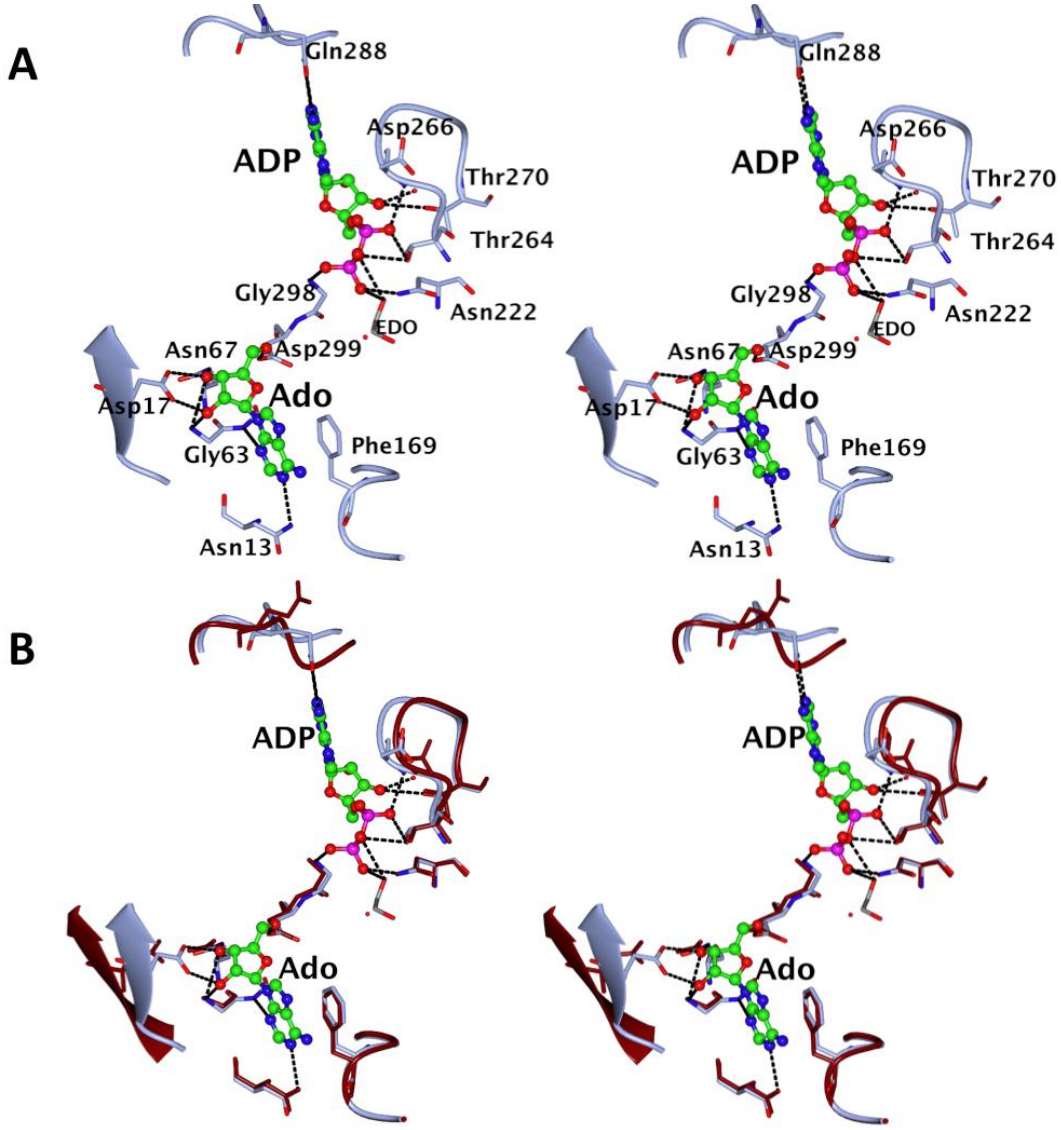


Figure 2-13: Active site of *Tb*AK.

The substrates Ado and ADP are shown in ball and stick representation coloured by atom type; the substrate-coordinating residues are shown as cylinders and the region around these residues are represented as ribbons. Dashed lines represent hydrogen bonds. (A) Active site with adenosine and ADP bound. (B) Overlay of the active sites of apo *Tb*AK (tan) and *Tb*AK-adenosine-ADP (ice blue).

*Tb*AK was co-crystallised with AppNHp, but only shows clear electron density for the α- and β-phosphate and none for the γ-phosphate. To address this, the structure was refined with REFMAC5 with ADP and AppNHp with γ-phosphate occupancies of 50 % and 100 %. Comparison of the electron density and difference maps of these refined structures with the unbiased map, generated before ligand modelling (Figure 2-14), strongly indicated the γ-phosphate to be absent. At this resolution it

is not possible to distinguish O and N atoms in the AppNHp from the maps, therefore, the ligand is referred to as ADP.

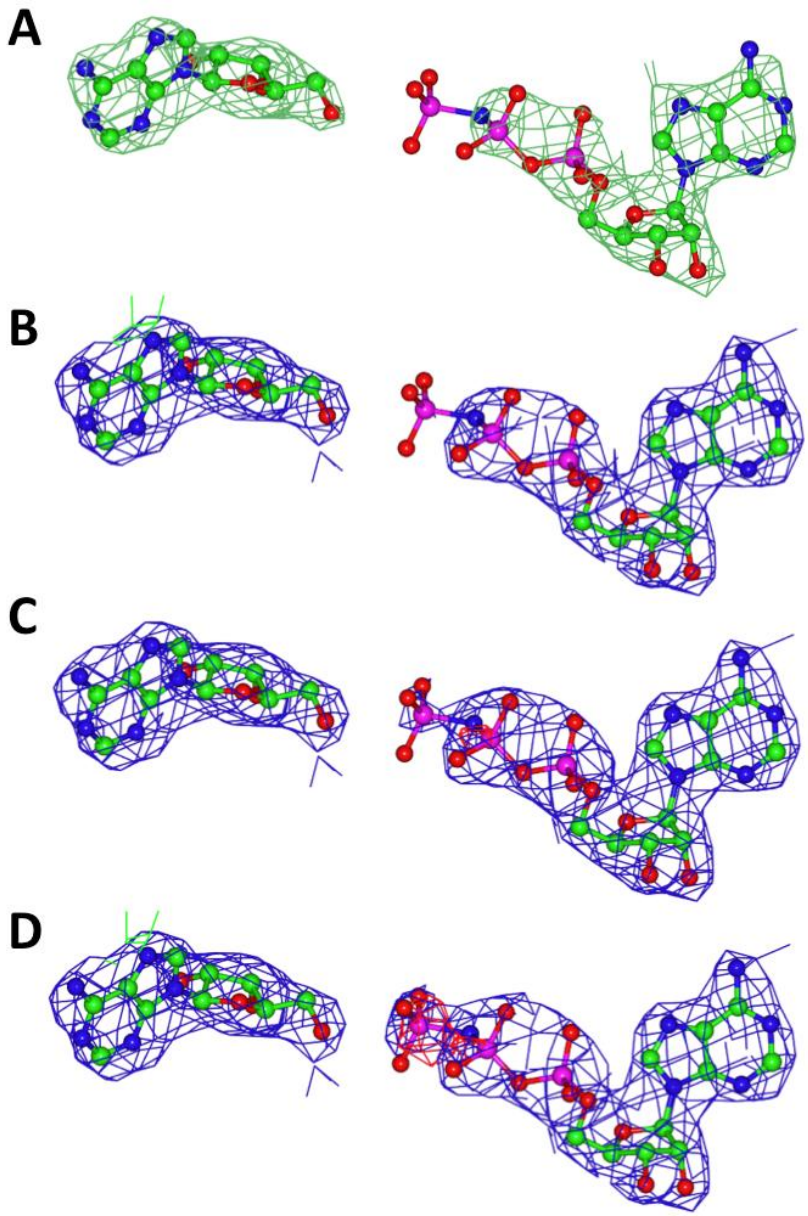


Figure 2-14: Electron density and difference maps of *Tb*AK ligands.

The weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c) were calculated using REFMAC5 (Murshudov *et al.*, 2011). Ado and AppNHp are shown as ball-and-stick models coloured by atom type, mF_o-DF_c maps contoured at 3σ in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue. (A) Unbiased maps generated before ligand modelling. (B) Maps after refinement with ADP (not AppNHp). (C) Maps after refinement with AppNHp, occupancies of the γ-phosphate atoms set to 50 %. (D) Maps after refinement with AppNHp, occupancies of the γ-phosphate atoms set to 100 %.

Important for the catalysis of nucleoside kinases is the anion hole motif (Schumacher *et al.*, 2000), a conserved sequence of residues (amino acids 290-300 in *Tb*AK) essential for binding and correct orientation of the phosphates (Figure 2-15). Superposition of the *Tb*AK-Ado-ADP structure (4n09) on those of *Tb*AK-AP_5A (3otx, (Kuettel *et al.*, 2011) and *Hs*AK in complex with two Ados (1bx4, (Mathews *et al.*, 1998) (Figure 2-15 A) shows it to be in an intermediate state between the latter two conformations, reflecting the importance of this motif. In our complex the γ-phosphate of the AppNHp (mimicking the phosphate donor ATP) is missing, so that its interactions cannot help to arrange the P-loop completely to form the anion hole, which is therefore not bent outwards as far as it would be with all three phosphates present. The structure of *Tb*AK-AP_5A (3otx) shows a fully developed but slightly distorted anion hole due to the presence of too many phosphates in the active site. The complex of *Hs*AK with two Ados (1bx4) lacks the anion hole motif and instead has an extended α11 helix, as there are no phosphates in the active site. In particular, residues 294-298 (centred around G296) take up a helical conformation in the absence of phosphate ligands, but change to a loop producing the anion hole when phosphates are present. Superposition of the apo *Tb*AK structure (4n08) with *Tb*AK-Ado-ADP (4n09) and *Tb*AK-KRM (2xbt, Figure 2-15 B) shows the anion hole being formed by the binding of the activator KRM (Kuettel *et al.*, 2011). In *Tb*AK-KRM and *Tb*AK-Ado-ADP the loop has nearly identical conformations, while in apo *Tb*AK the loop adopts a different conformation. How binding of KRM induces this change in the loop region of residues 294-298 is not clear, as its binding site is ~ 9 Å away from the loop.

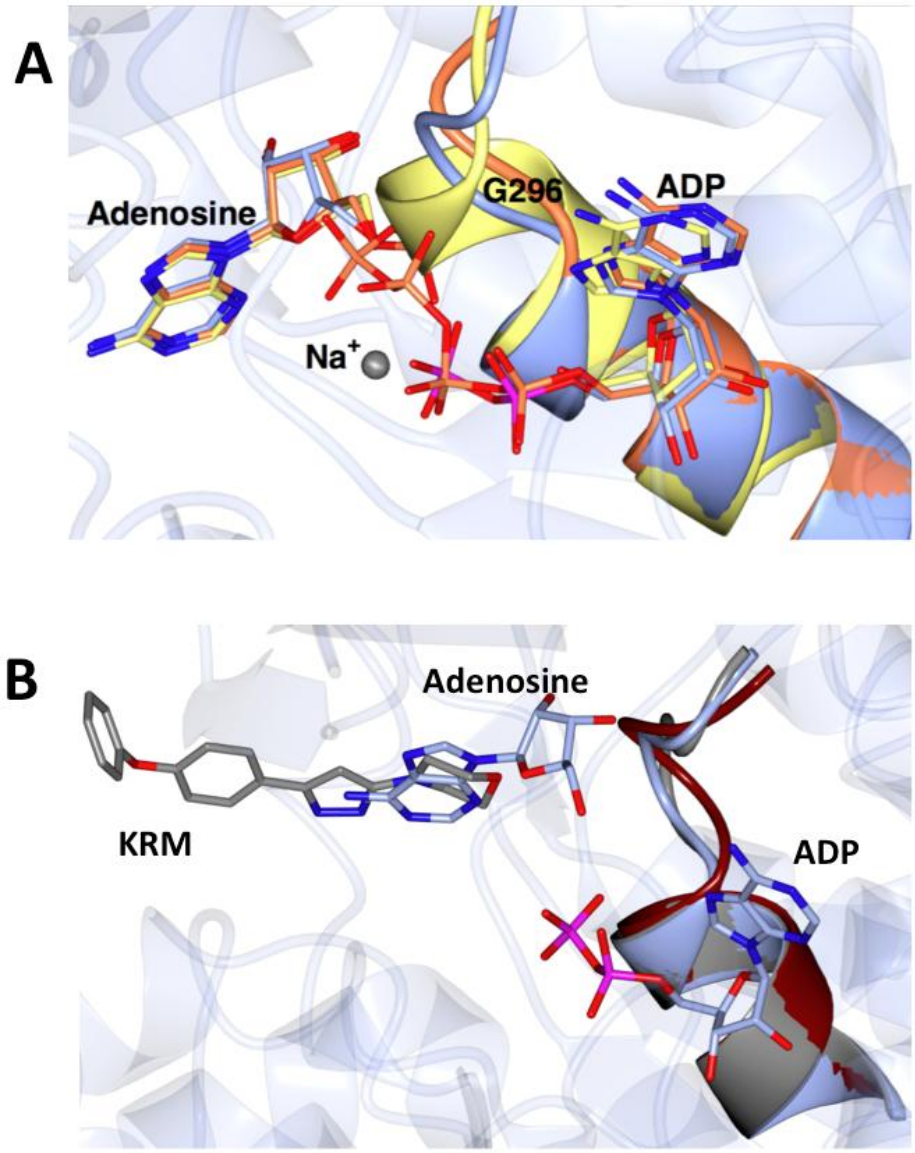


Figure 2-15: Anion hole motif.

(A) Superposition of the anion hole motifs of *Hs*AK (lemon, 1bx4, (Mathews *et al.*, 1998), *Tb*AK-AP_5A (coral, 3otx, (Kuettel *et al.*, 2011) and *Tb*AK-Ado-ADP (ice blue, *Tb*AK co-crystallised with AppNHp, but missing the γ-phosphate in the structure, 4n09). (B) Superposition of the anion hole motifs of apo *Tb*AK (tan, 4n08), *Tb*AK-Ado-ADP (ice blue, 4n09) and *Tb*AK-KRM (grey, 2xtb, (Kuettel *et al.*, 2011). The main chain is represented as ribbons; ligands are represented as cylinders coloured by atom type with the carbon and phosphorous atom coloured by chain.

2.4 Discussion

2.4.1 Comparison to other published structures

The apo structure is in the open conformation is essentially identical to that reported for *Tb*AK soaked with the activator KRM (PDB: 2xtb, (Kuettel *et al.*, 2011) and that of apo-*Tgo*AK (Schumacher *et al.*, 2000). Comparison with 2xtb showed that the overall structures are effectively the same (Figure 2-16 A), the only differences being in the flexible loop region of the anion hole motif, spanning residues 281-297 (Figure 2-15 B). Additionally, the activator complex seems slightly more open, but this probably does not directly reflect the binding of the activator but is rather an artefact of crystallisation. The side chain of L138, close to the activator, has a different position in the two structures but makes no direct or indirect interactions with the activator itself.

The structure of *Tb*AK-Ado-ADP in the closed conformation is very similar to that for the complex with the bi-substrate analogue *Tb*AK-AP_5A (3otx, (Kuettel *et al.*, 2011), with an RMSD on Cα positions of 0.57 Å (Figure 2-16 B). The adenines in both structures occupy essentially identical positions, as does the ribose on the ADP and AP_5A (Figure 2-16 C). The only regions showing differences are loop regions and in the anion hole motif (Figure 2-15 A).

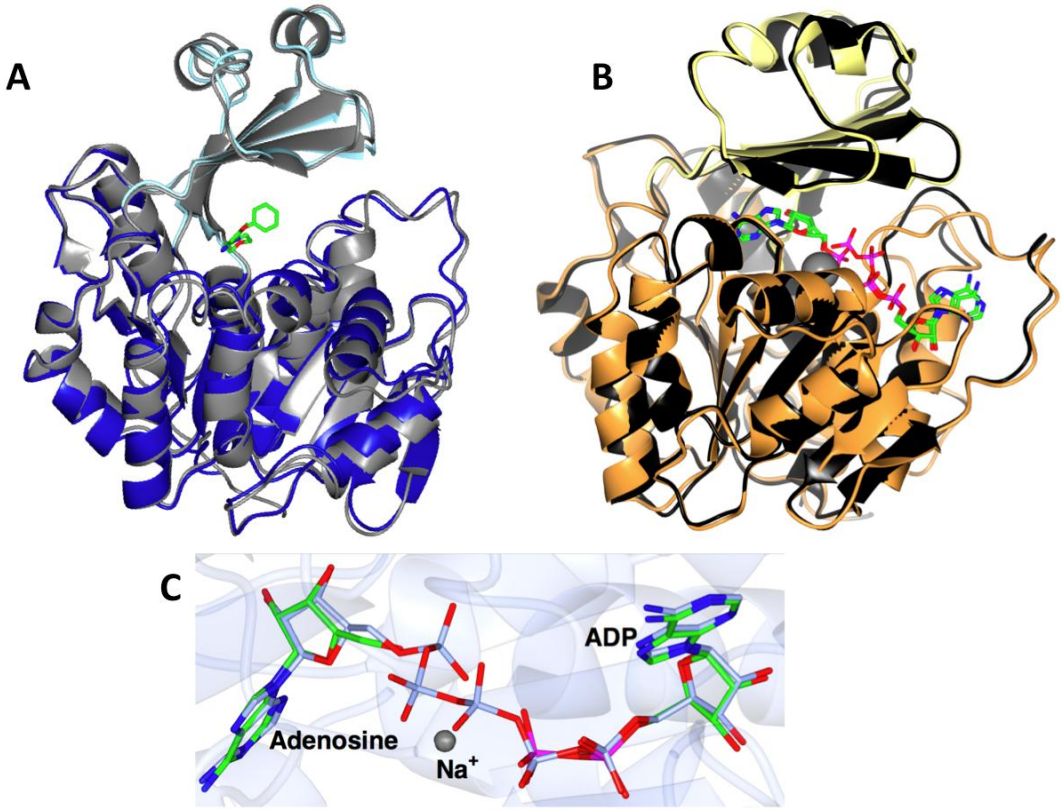


Figure 2-16: Comparison of *Tb*AK structures.

Proteins in ribbon representation and ligands are shown as cylinders coloured by atom type. (A)Superposition of *Tb*AK-KRM (grey, 2xtb, (Kuettel *et al.*, 2011) with apo *Tb*AK (blue and light blue, 4n08). (B) Superposition of *Tb*AK-AP_5A (black, 3otx, (Kuettel *et al.*, 2011) and *Tb*AK-Ado-ADP (orange and lemon, 4n09). (C) Close-up of the ligands in a superposition of *Tb*AK-AP_5A (3otx) and *Tb*AK-Ado-ADP (4n09). Ribbon in the background shows the *Tb*AK-Ado-ADP structure. Ligands are shown as cylinders coloured by atom type, with carbon and phosphor atoms of AP_5A in ice blue.

In the *Tb*AK-AP_5A complex, there are five phosphates between the two Ado moieties, two more than would be present in the transition state, so the phosphates would need to be in different conformations in the Michaelis complex and the transition state. The structures of AMPK from *Aquifex aeolicus* have been determined in complex with both AP_5A (2rgx, (Henzler-Wildman *et al.*, 2007) and with the phosphoryl transfer transition state mimic ADP:AlF_4:AMP in the active site (3sr0). The enzyme was in a very similar closed conformation in both, suggesting that the enzyme conformation and the adenine binding sites in the AP_5A analogues were indeed representative of the true ligand binding state. Unfortunately, extensive crystallisation screens with nucleotides and magnesium and aluminium fluorides with the aim of obtaining a better mimic of the transition state failed to

produce diffracting crystals of sufficient quality. One problem may lie in the fact that the aluminium fluoride in *Tb*AK would be coordinated by a P-O oxygen on one side and a ribose oxygen on the other, distinct from the AMPK system where the aluminium is coordinated by phosphate oxygens on both sides.

2.4.2 Modelling of *Tb*AK-2 and comparison to *Tb*AK-1

The amino acid sequences of *Tb*AK-1 and *Tb*AK-2 differ in only four positions. This high degree of identity allows the assumption that the structures of the two proteins are identical. Modelling of the four amino acid substitutions onto the structure of *Tb*AK-1 solved within this work (Figure 2-17), revealed no clashes or obvious potential effects on substrate binding or catalysis. The substitutions are located approximately 7 Å (S269P), 9.6 Å (L272I), 11 Å (M302V) and 12.5 Å (R332Q) away from the ATP-binding site and even further from the Ado-binding site. It is possible that S269P has an effect on the movement of the adjacent loop during ATP-binding, but the other amino acid substitutions are unlikely to have an effect on structure or function. Therefore, it can be assumed the two proteins have identical functions within the parasites, where one copy is free to mutate as long as the other retains its function.

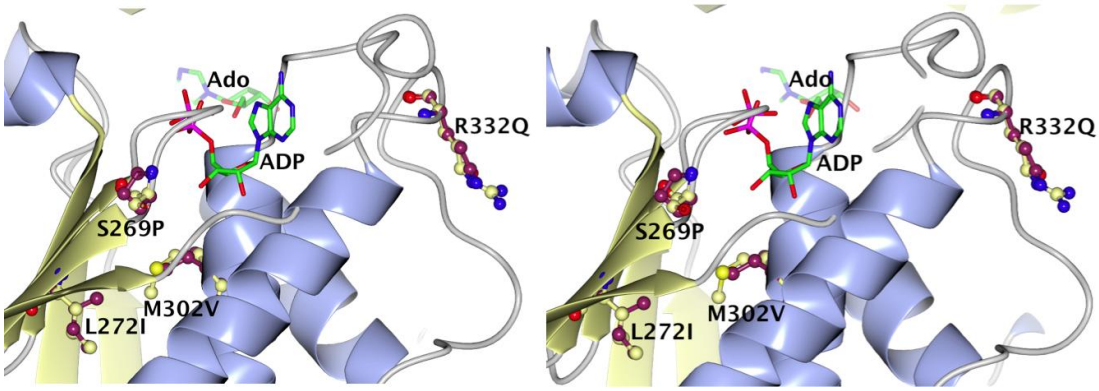


Figure 2-17: Amino acid substitutions in *Tb*AK-2.

Structure of *Tb*AK-Ado-ADP shown in ribbon representation coloured by secondary structure with α-helices in ice blue and β-sheets in lemon. Ado and ADP are shown as cylinders coloured by atom type. Residues that are substituted (S269P, L272I, M302V and R332Q) are shown as ball-and-stick coloured by atom type with the C^{α} atoms of *Tb*AK-1 in lemon and the C^{α} atoms of *Tb*AK-2 in dark purple.

2.4.3 Cordycepin and its derivatives as anti-trypanosomal agents

Cordycepin was shown to have trypanocidal properties *in vitro* (Williamson, 1972, Williamson & Macadam, 1976) and the toxicity of this pro-drug is strongly dependent on its activation by phosphorylation by AK (Luscher *et al.*, 2007). Somewhat disappointingly, attempts to crystallise *Tb*AK in complex with cordycepin failed to produce diffracting crystals. Nevertheless, the structure of *Tb*AK-Ado-ADP allows assumptions to be made about possible cordycepin modifications which might influence its binding to AK. Looking at the active site of *Tb*AK and Ado binding, key interactions can be identified, like the hydrogen bonding of the hydroxyl groups of the ribose ring and the adenine ring stacking to F169, which are crucial for binding of the substrate and therefore need to be maintained in modified substrates. *In vivo* cordycepin is inactivated by deamination by human ADA (Aiyedun *et al.*, 1973, Agarwal *et al.*, 1975), resulting in 3'-deoxyinosine (Figure 2-18). The structure of ADA and its substrate-binding mode are known (Wilson *et al.*, 1991), so that the possible cordycepin modifications can be discussed in respect to resistance against its deamination.

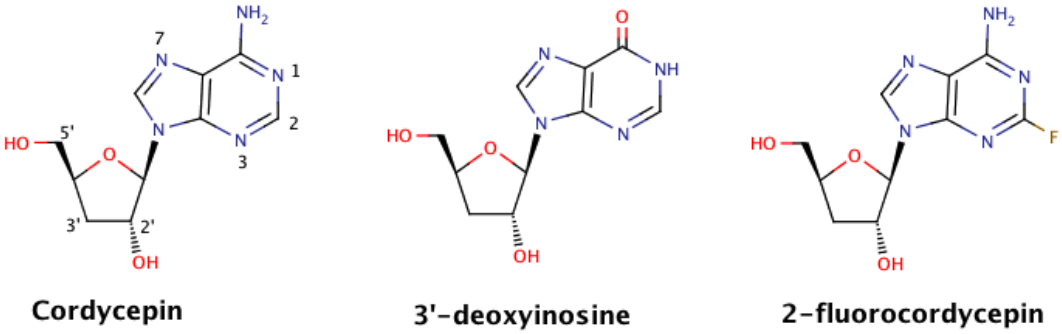


Figure 2-18: Cordycepin and derivatives.

In the course of this work, Vodnala *et al.* published a detailed study of cordycepin and its derivatives in respect of their trypanocidal activity *in vivo* and resistance to deamination (Vodnala *et al.*, 2013). They found the free amine at the C6 position of the adenine ring to be essential for trypanocidal activity. Moreover, modifications on the ribose moiety, such as changes in the stereochemistry of the 2'-OH group, inactivated the compound. Gillerman and Fischer found that substitutions at the C2 position of the adenine ring transform ADA substrates into non-substrates or inhibitors of ADA depending on the size of the substitution (Gillerman & Fischer, 2011). Based on this study combined with the structural information (Wilson *et al.*, 1991) Vodnala *et al.* tested different C2 substitutions using electron-withdrawing substituents to lower the hydrogen bonding capability and reactivity of the adenine nitrogens in the ADA active site (Vodnala *et al.*, 2013). This resistance to deamination due to C2 halogenation was shown previously for compounds like F-AraA (Brockman *et al.*, 1977). Indeed, fluorination of cordycepin at the C2 position abolished deamination by ADA, while retaining its trypanocidal activity *in vitro*. 2-fluorocordycepin was shown to be a substrate of the aminopurine transporter and *Tb*AK. *In vivo* experiments in mice showed administration of 2-fluorocordycepin was able to cure acute trypanosomiasis (Vodnala *et al.*, 2013). Further investigations to validate 2-fluorocordycepin's potential in anti-trypanosomal therapy in humans are needed.

2.5 Summary and future work

Obtaining *Tb*AK crystals with sufficient quality proved to be difficult. However, the structures in open (apo) and closed (in complex with Ado and AppNHp) conformation were solved, both to 2.6 Å. They are essentially identical to 2xtb and 3otx, respectively (Kuettel *et al.*, 2011), and show high homology to other published AK structures in the PDB. Unfortunately, attempts to solve the structures of *Tb*AK in complex with cordycepin and the transition state mimics Ado:AlF_4:ADP and Ado:MgF_3:ADP were not successful.

The *Tb*AK-Ado-ADP structure allows identification of the key interactions necessary for substrate binding and permits assumptions about the binding of cordycepin. An independent study by Vodnala *et al.* (Vodnala *et al.*, 2013) suggests the structures of *Tb*AK can be used for the rational design of further modifications optimising the pharmacokinetic properties of 2-fluorocordycepin.

Chapter 3 Thymidine Kinases of the TriTryps

3.1 Introduction

TKs catalyse the phosphorylation of dThd to dTMP (EC. 2.7.1.21, Arner & Eriksson, 1995, Jordan & Reichard, 1998) and are divided into two distinct families (Gentry, 1992, Birringer *et al.*, 2005), type I (e.g. herpes simplex virus 1 TK and human mitochondrial TK2) and type II (e.g. the human cytosolic TK1 and the trypanosomal TKs). The sequences and folds of these two types have no homology, a comparison shown in Figure 3-1, and they most likely have distinct evolutionary origins.

Type I TKs are monomeric or dimeric and show high structural similarity to other deoxyribonucleoside kinases, e.g. TMPK, human deoxycytidine kinase (dCK) and insect deoxyribonucleoside kinase (dNK, Figure 3-1 A & D). They can be subdivided into herpes-viral, pox-viral and cellular subfamilies, with the three subfamilies showing very low sequence identity (Wild *et al.*, 1997). One of the most prominent examples is the Herpes simplex virus enzyme (Figure 3-1 B), which is bigger than the other family members, but still shows a very similar fold. Each protomer contains a central five-stranded β-sheet, surrounded by α-helices. The substrate-binding site is formed by two helix pairs and a helical lid structure. Type I TKs generally have broader substrate specificity than type II, though they preferentially phosphorylate dThd (Vogt *et al.*, 2000, Johansson *et al.*, 2001). Although constitutively expressed throughout the cell cycle these enzymes show complex regulation of their activity, not only by product inhibition with the end product dTTP binding in the thymidine-binding site, but also in the form of negative cooperativity. Although showing this negative cooperativity the human TK2 appeared to be monomeric in solution and a second substrate-binding site could not be identified (Munch-Petersen, 1984, Munch-Petersen *et al.*, 1991, Knecht *et al.*, 2002).

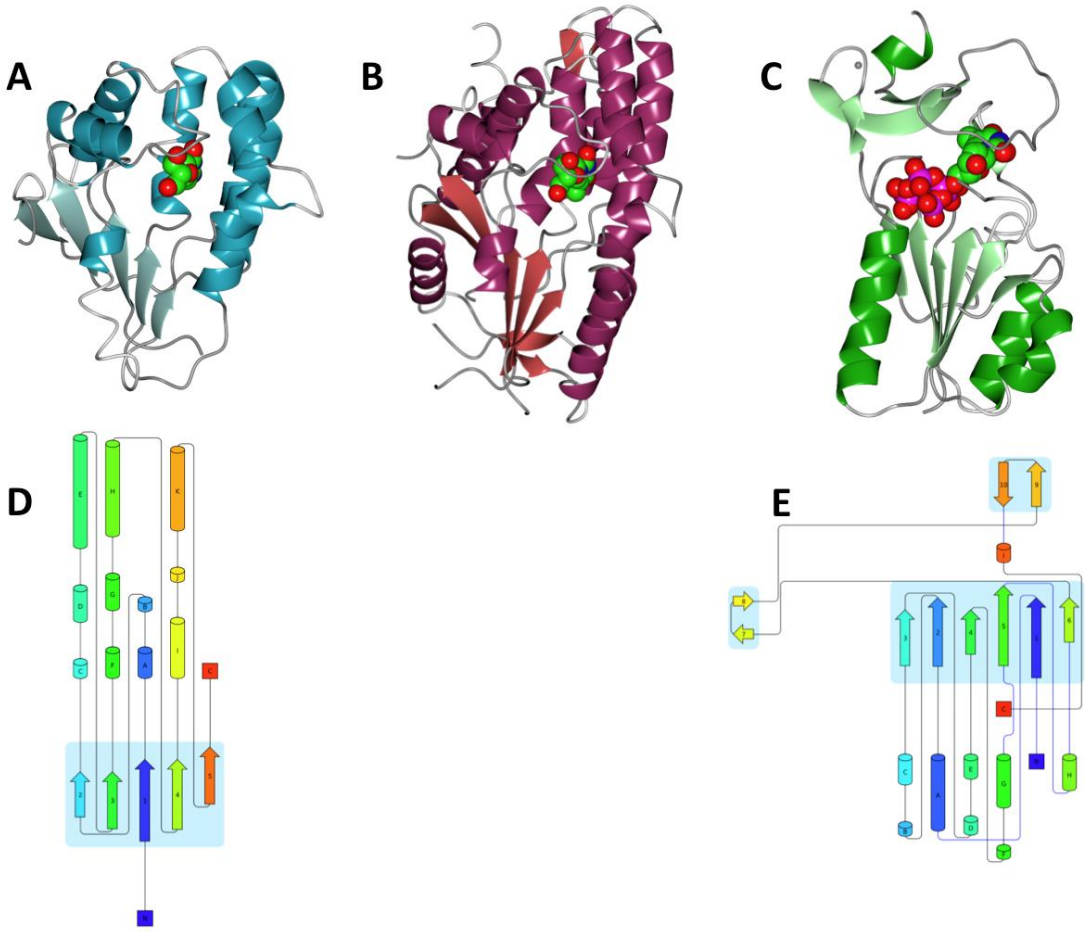


Figure 3-1: Protomer structures of three deoxyribonucleoside kinases.

(A) *Drosophila melanogaster* TK type 1-like enzyme, *Dm*dNK PDB: 1j90, in complex with deoxycytidine (B) HRV1 TK type 1 in complex with dThd, PDB: 1e2j, (C) human TK1 type II in complex with dTTP, PDB: 1w4r; protein in ribbon representation; ligands are shown as spheres coloured by atom type. (D) Topology cartoon of *Dm*dNK PDB: 1j90 coloured in rainbow from N-terminus (blue) to C-terminus (red). (E) Topology cartoon of human TK1, PDB: 1w4r coloured same as (D). Structure representations in this and all subsequent figures made with CCP4*mg* (McNicholas *et al.*, 2011), protein topology cartoons generated using Pro-Origami (Stivala *et al.*, 2011).

Type II TKs are very well studied and structures from several species have been solved including those from human (*Hs*TK1, Figure 3-2, Welin *et al.*, 2004, Birringer *et al.*, 2005), *Ureaplasma urealyticum* (*Uu*TK, Kosinska *et al.*, 2005), *Bacillus anthracis* and *B. cereus* (*Ba*TK and *Bc*TK, Kosinska *et al.*, 2007), *Thermotoga maritima* (*Tm*TK, Segura-Pena, Lichter*, et al.*, 2007, Segura-Pena, Lutz*, et al.*, 2007) and *Clostridium acetobutylicum* (*Ca*TK, Kuzin *et al.* not published). They occur as homodimers (Figure 3-2 B) and tetramers (Figure 3-2 C), where each protomer

consists of two lobes. The bigger of the two lobes contains a central six-stranded parallel β-sheet surrounded by α-helices and the P-loop. The second one comprises four conserved cysteine residues coordinating a Zn^{2+} ion, and a lasso structure, covering the substrate-binding site.

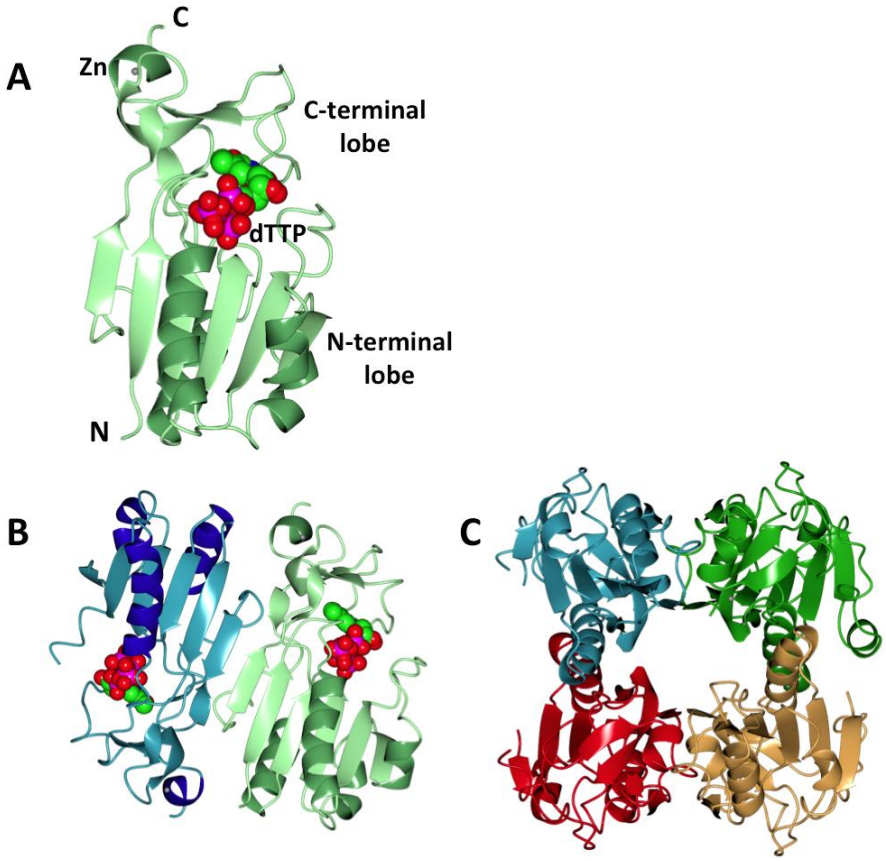


Figure 3-2: Overall structure of type II TKs.

Shown is the structure of a *Hs*TK1 protomer (A), homodimer (B), and homotetramer (C). The protein represented as ribbons coloured by chain and the ligands (dTTP and zinc) as spheres coloured by atom type. The figure was made using CCP4*mg*.

While the overall structure is highly conserved, the substrate specificity varies greatly between species. Type II TK's main function is proposed to be supplying cells with dTMP as a precursor of dTTP, which is used in DNA synthesis. This is supported by studies with the *Hs*TK1 showing the transcription level and activity being highest in cells immediately before and during DNA synthesis (late G1 until mid S phase of the cell cycle) and its degradation when DNA synthesis is not taking place (Dou &

Pardee, 1996, Ke & Chang, 2004, Munch-Petersen *et al.*, 1993, Munch-Petersen, 2009). The cell cycle dependent difference in TK concentration leads to an additional regulatory effect of TK activity. It was shown for TKs from higher vertebrates that the oligomeric state changes dependent on the concentration and by presence or in the absence of ATP; TK being dimeric at low concentration and absence of ATP and tetrameric at high concentrations and in the presence of ATP. The activity of the tetrameric TK was determined to be 30-fold higher than that of the dimeric TK (Munch-Petersen *et al.*, 1993, Munch-Petersen, 2009). TKs of bacteria and plants were shown to be predominantly dimeric even in the presence of ATP (Mutahir *et al.*, 2013). All TKs studied to date show product inhibition and cooperativity.

3.1.1 Thymidine kinases from the TriTryps

The TriTryp TKs belong to the type II TKs like the human cytosolic enzyme (*Hs*TK1). The overall structure and the assembly to dimers and tetramers is conserved among different species, and the amino acid sequences show an identity between ~ 35 % (*T. brucei* to bacterial) and ~ 50 % (*T. brucei* to human), Figure 2-2. Not surprisingly the degree of identity is highest in regions important for function, like the P-loop, the active site residues and the residues coordinating the structural zinc.

The TK from *L. major* (*Lm*TK) is a 40.6 kDa protein, with a C-terminal extension of unknown function most of which is predicted to be disordered using DISOPRED2 (Ward *et al.*, 2004). It is expressed in large amounts in *Leishmania* amastigotes (Bente *et al.*, 2003) and localises to the cytoplasm (Thiel *et al.*, 2008). It was shown to be involved in flagellum formation, as well as cell shape and infectivity of the parasites, making it a potential drug target.

T. brucei TK (*Tb*TK) is a tandem protein of 51 kDa total size, containing two domains with homology to the TK family, an N-terminal inactive one (*Tb*TK-N) and a C-terminal catalytically active one (*Tb*TK-C), forming a pseudo-dimer in solution (Ranjbarian *et al.*, 2012). These two TK-homology domains show an amino acid sequence identity of 67 % and their high sequence identity on the DNA level

indicates a relatively recent interchange of genetic material. Ranjbarian *et al*. investigated the kinetic parameters in detail, showing *Tb*TK had relaxed phosphate donor specificity and exhibited similar enzymatic activities with ATP, GTP, CTP and UTP. *Tb*TK's substrate specificity was shown to be broader than the *Hs*TK1, as it phosphorylated not only dThd and dUrd, but also dC and dG. *Tb*TK-N's inactivity can be traced back to the substitution of residues important for catalysis, namely D45, R47 and E87. When expressed individually the catalytically active *Tb*TK-C was predominantly monomeric and showed a lower binding affinity for the substrate than the full-length enzyme. The authors concluded that the inactive *Tb*TK-N domain helps substrate binding of the active *Tb*TK-C and is required for tetramerisation. *Tb*TK has a flexible extension on the C-terminus, which is shorter than that from *Lm*TK and whose function is not known.

The sequence of the *T. cruzi* TK (*Tc*TK) has two possible start codons, one resulting in a long N-terminal extension (omitted for the sequence alignment), leaving a protein of 34 kDa (*Tc*TK A) and the other leaving a protein of 30 kDa (*Tc*TK B). Like *Lm*TK and *Tb*TK, *Tc*TK has a flexible extension at the C-terminus. No detailed kinetic or functional studies on *Tc*TK have been published to date.

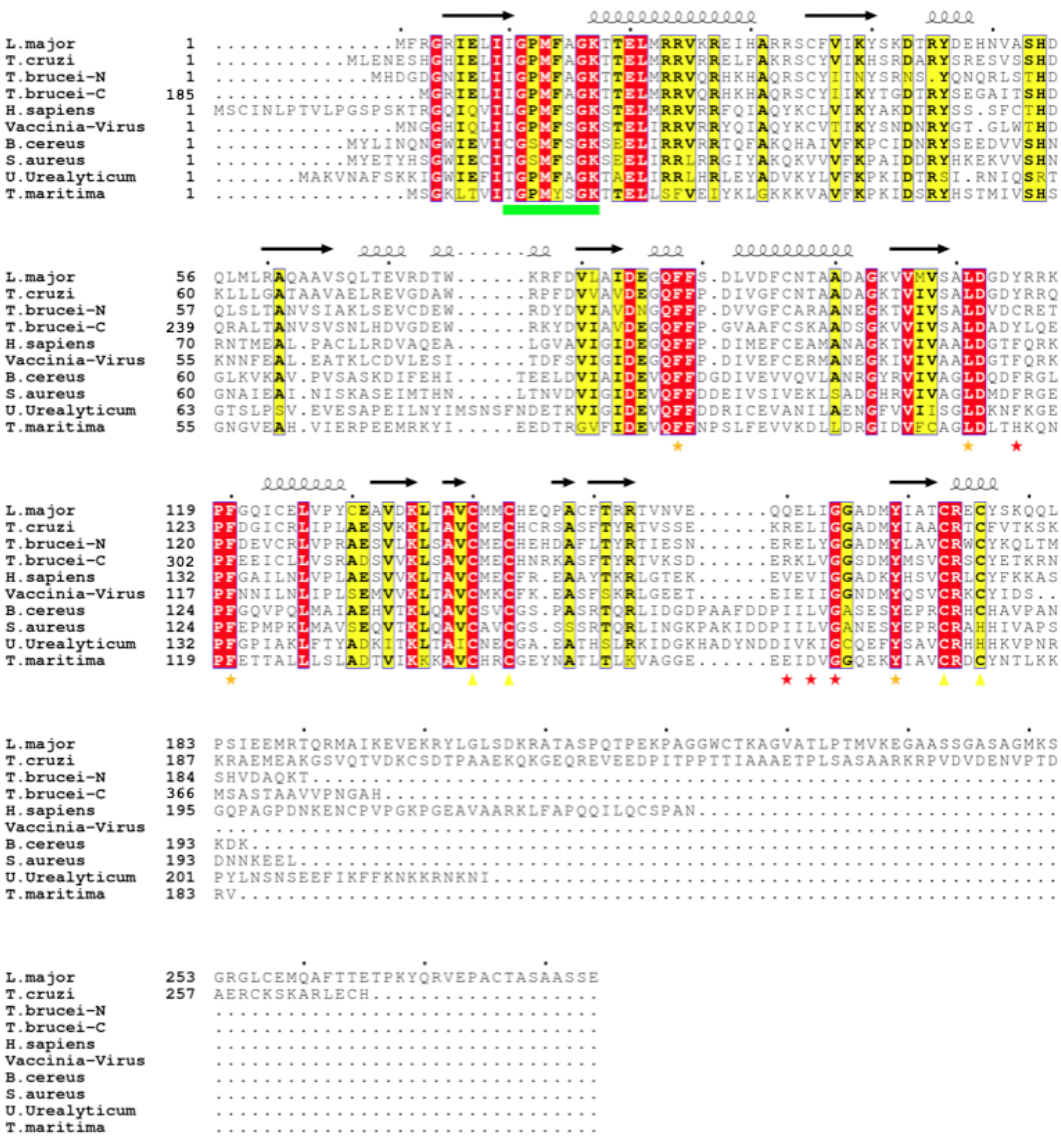


Figure 3-3: Sequence alignment of type II TKs.

Conserved residues with red (identical) and yellow (similar) background, residue numbering and secondary structure of *Lm*TK shown above the sequences (arrows = β-strands, spirals = α-helices), green bar = p-loop, red stars = residues coordinating dThd by main chain hydrogen bonds, orange stars = residues coordinating dThd *via* stacking and hydrophobic interactions, yellow triangles = Zn^{2+}-coordinating cysteines. Alignment carried out using T-coffee and Espript 3.0 (Notredame *et al.*, 2000, Gouet *et al.*, 2003).

3.1.2 TK Mechanism

The mechanism of thymidine phosphorylation in TKs of both types has been studied extensively (Wild *et al.*, 1997, Welin *et al.*, 2004). Many nucleoside kinases have a very similar mechanism (Eriksson *et al.*, 2002). The protein substrate-interactions based on structural analysis of *Tm*TK in complex with dThd and AppNHp are summarised in Figure 3-4. The proposed mechanism of TK based on this is shown in Figure 3-5. In brief, the carboxylate of a glutamate residue in the active site is proposed to act as base, deprotonating the 5'-OH group of the dThd. This activated 5'-O then makes a nucleophilic attack on the γ-phosphate of the ATP, leading to a transition state with a penta-coordinated γ-phosphate and subsequently the products ADP and dTMP. A magnesium ion is crucial for catalysis as coordinating counter ion for the phosphates. While TKs of both types exclusively utilise non-phosphorylated nucleosides as primary substrates, the HSV-1 TK is an exception being able to phosphorylate its product dTMP even further forming dTDP (Wild *et al.*, 1997). This ability has not been described for other TKs so far.

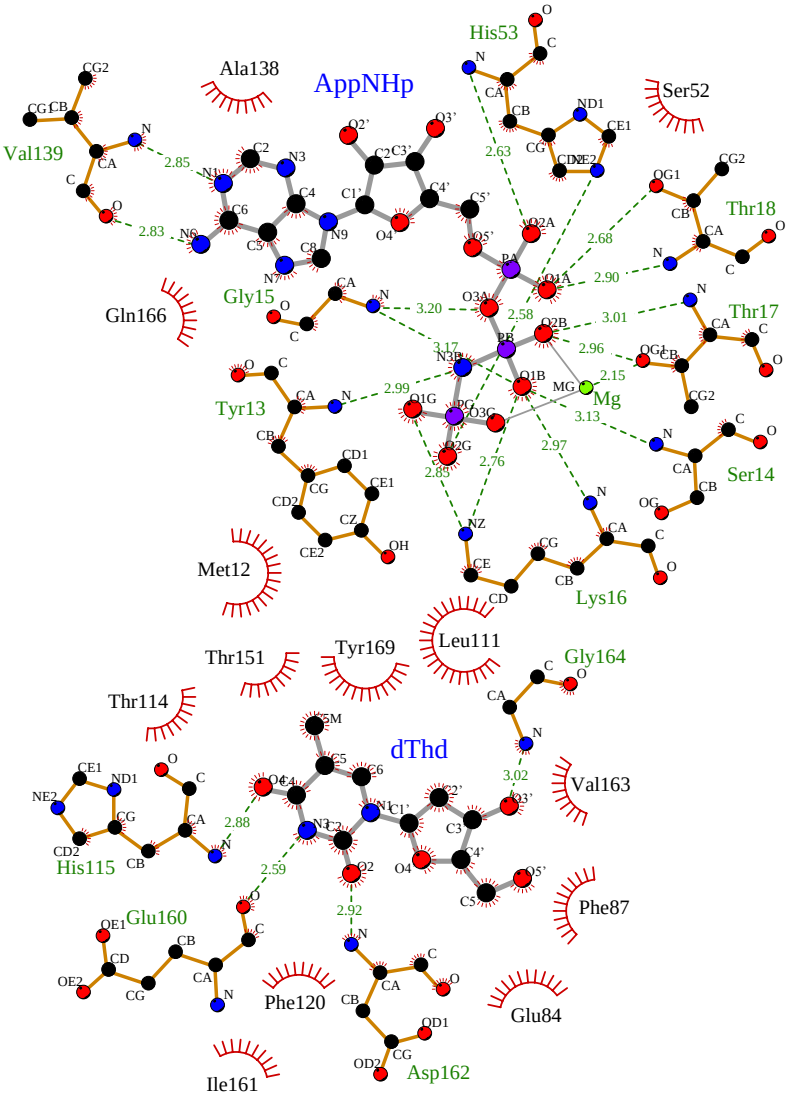


Figure 3-4: Schematic view of the substrate binding interactions in the *Tm*TK active site.

Shown are the ligands and residues coordinating substrates *via* hydrogen bonds (green labels, hydrogen bonds as dashed lines), and hydrophobic interactions (black labels, red half circles). Ligand and hydrogen bonding residues are shown as ball and stick models coloured by atom type. Figure drawn with LigPlot+ (REF).

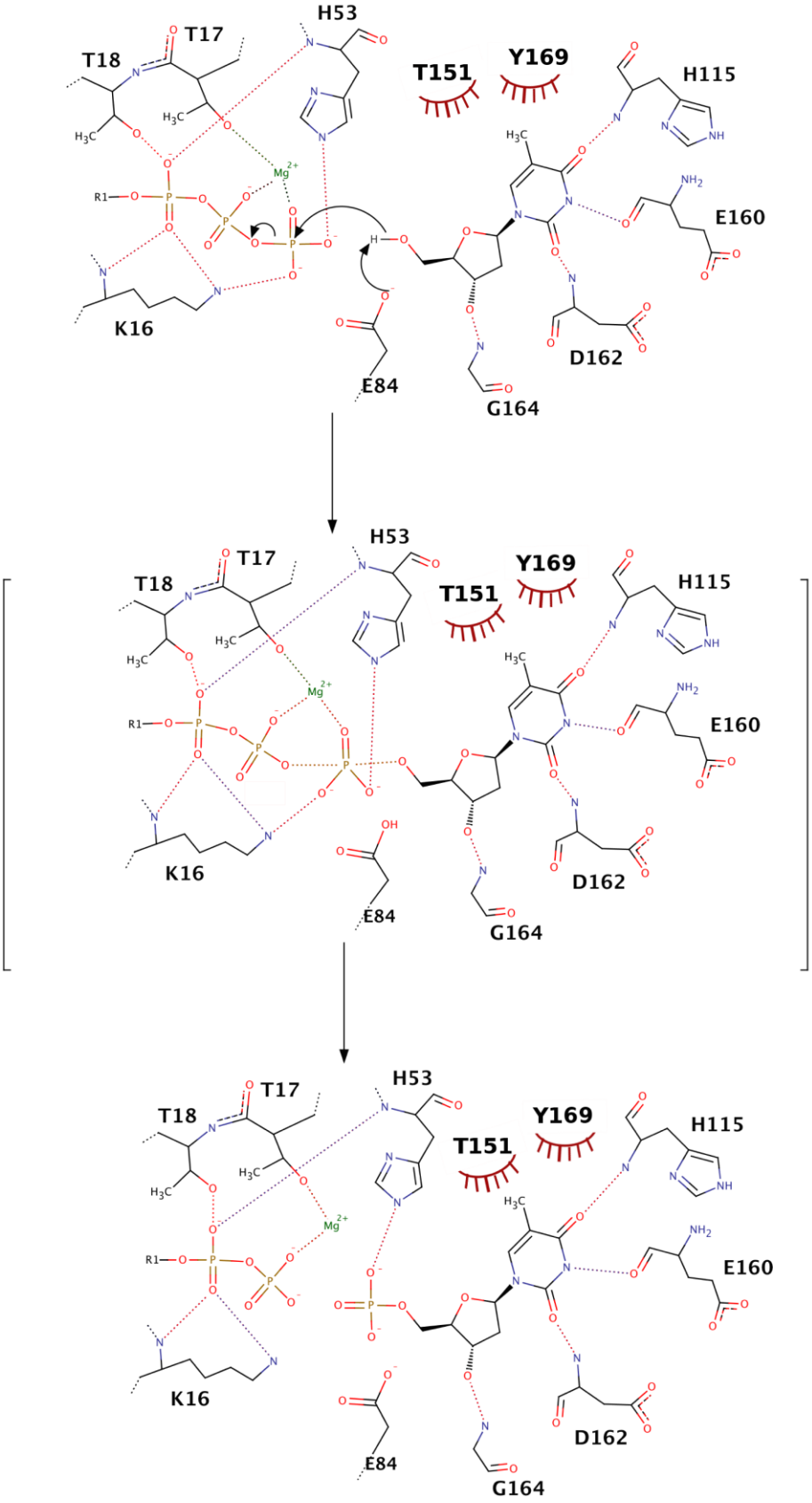


Figure 3-5: Mechanism of γ-phosphate transfers in TK.

R1 is the adenosine moiety of the phosphate donor ATP. Residue numbering after *Tm*TK. Figure made with MarvinSketch (ChemAxon).

3.1.3 TKs in drug development

The viral type I TKs have been of great interest as they activate cytotoxic nucleoside analogues with substrate specificities differing from the type II TKs, resulting in the possibility to specifically target pathogens with these prodrugs. Additionally, type II TKs are of great interest in cancer monitoring and therapy. The cell cycle specific expression of human cytosolic TK and its influence on nucleoside uptake can be exploited to monitor active tumor proliferation in PET scans (Barthel *et al.*, 2003, Chao, 2006, Salskov *et al.*, 2007, de Langen *et al.*, 2009, Shields *et al.*, 2008) and has been shown to be a useful tool for distinguishing slow-growing from aggressive fast-growing cancers (Aufderklamm *et al.*, 2012). The high substrate specificity of the *Hs*TK1 (type II) did not permit the development of nucleoside-based cytotoxic prodrugs for cancer therapy at this point.

3.1.4 Aims

The main focus of the work discussed in this chapter was to overexpress, purify and determine the crystal structures of the TKs of the TriTryps and compare them to *Hs*TK1. This should highlight unique features of the TriTryp TKs, which could then be exploited for target-based drug design.

3.2 Experimental

3.2.1 Cloning and expression testing in *E. coli*

The full-length genes as well as various truncations of TKs from *T. brucei* Lister427, *T. cruzi* Sylvio X10/1 and *L. major* Friedlin were originally cloned into YSBLIC and YSBLIC3C vectors using ligation-independent cloning (Aslanidis & de Jong, 1990, Fogg *et al.*, 2006) and tested for expression by Dr. Glyn Hemsworth. I continued with design, cloning and expression testing of additional constructs (21 in total, including MBP fusion tags) using different *E. coli* strains (BL21 (DE3), BL21 (DE3) gold, Rosetta2 (DE3), pLysS (DE3), pLEMO (DE3), C41 and ArcticExpress), temperatures (10, 16, 20, 25, 37°C) and media (LB, autoinduction, 2XYT, M9).

For these additional constructs the genes and truncated versions were amplified from genomic DNA by PCR using the same reaction conditions as described in Table 2-1 and Table 2-2, with the primer sequences given in appendix A.2.3. The templates for the MBP constructs were the constructs *Lm*TK-F-3C, *Tb*TK-F-3C and *Tc*TK-F-3C. The MBP-fusion constructs were cloned using conventional cloning methods with the restriction enzymes EcoR1 and Sal1 (Promega) and subsequent ligation with the T4 ligase (New England Biolabs, NEB). In brief, approximately 2 µg of the amplified TK genes and pMAL-C2 vector, purified using the PCR purification Kit (Qiagen) or Miniprep Kit (Qiagen) respectively, were digested with EcoR1 and Sal1 for 3 h at 37°C. The reaction mix compositions are summarised in Table 3-1. Subsequently the digested pMAL-C2 vector was dephosphorylated by incubation with alkaline phosphatase for 1 h at 37°C. After purification with the PCR purification kit, inserts and vector were ligated by incubation of the reaction mixture specified in Table 3-2 at room temperature for 24h. The ligated constructs were transformed into *E. coli* (NEB5α, NEB) cells and the cloning success was confirmed by colony PCR and sequencing.

Table 3-1: Compositions of the resitriction digests for the pMAL-C2 constructs.

	***Lm*TK-F**	***Lm*TKdC**	***Tc*TK-A**	***Tc*TK-B**	***Tb*TK-F**	**pMAL-C2**
10x buffer D (Promega)	5 µL					
DNA (~ 2 µg)	24 µL	21 µL	24 µL	32 µL	25 µL	35 µL
BSA (10 µg/µL, Promega)	0.5 µL					
ddH_2O	15.5 µL	18.5 µL	15.5 µL	8.5 µL	14.5 µL	5.5 µL
EcoR1 (12 U/µL, Promega)	2.5 µL					
Sal1 (10 U/µL, Promega)	2.5 µL					
Total volume	50 µL					

Table 3-2: Composition of T4 ligase dependent ligations for pMAL-C2 vectors.

Reagent	**Volume [µL]**
10x T4 ligase buffer (NEB)	2
ddH_2O	9.8
Vector (digested)	5 (= ~100 ng)
Insert (digested)	3 (= ~100 ng)
T4 ligase (400 U/µL, NEB)	0.2
Total volume	20

All expression tests were carried out in 5 mL culture volumes using different *E. coli* strains, temperatures and expression media (as mentioned above). For each strain and expression condition 1 mL culture samples were lysed chemically using BugBuster® (Millipore) and soluble and insoluble fractions were analysed *via* SDS-PAGE.

While *Tb*TK and *Tc*TK did not give soluble protein, some of the *Lm*TK constructs showed soluble expression, but these proved to be unstable during purification and only very small amounts at low concentrations could be obtained, of insufficient purity and quantity for crystallisation experiments. After Ranjbarian *et al.* (Ranjbarian *et al.*, 2012) successfully expressed full-length *Tb*TK in BL21 (DE3) codon plus cells, I tested their expression protocol, which resulted in soluble but heavily degraded protein showing half of the expected molecular weight. No soluble full-length *Tb*TK was obtained.

3.2.2 Cloning of TK constructs for expression with the baculovirus system

A more high throughput approach was applied for which 30 constructs were cloned for expression in insect cells using the baculovirus system with a ligation-independent cloning method developed at the Oxford protein production facility (OPPF, (Berrow *et al.*, 2007) with the commercially available In-Fusion™ enzyme (Clontech-Takara Bio Europe). In parallel, the same approach was taken with constructs for TMPK expression, discussed in the appendix A.1. I spent two weeks at the OPPF in Harwell for cloning and expression testing.

Full-length and re-designed truncated versions of TK from the TriTryps were amplified by PCR from genomic DNA (*T. brucei* Lister427, *T. cruzi* Sylvio X10/1 and *L. major* Friedlin) in 96-well format using the reaction conditions and temperature profile summarised in Table 3-3 and Table 3-4.

Table 3-3: PCR condition of TK constructs for baculovirus expression.

Reagent	**Volume/reaction [µL]**
2x Phusion™ Flash Mastermix	25
Forward primer (10 µM)	3
Reverse primer (10 µM)	3
ddH_2O	18
Template (genomic DNA)	1
Total volume	**50**

Table 3-4: Temperature profile for PCRs of the baculovirus expression constructs.

Step	**Temperature [°C]**	**Time [s]**	**No. of cycles**
Initial denaturation	98	10	1
Denaturation	98	5	30
Annealing	70	5	
Extension	72	60	
Final extension	72	60	1

The PCR products were purified using Agencourt AMPure XP magnetic beads according to manufacturers instructions and analysed on a 1.6 % agarose gel.

Each TK version was cloned into three different vectors:

- pOPINI (N-terminal non-cleavable Histag fusion),
- pOPINF (N-terminal HRV3C cleavable Histag) and
- pOPINE (C-terminal non-cleavable Histag).

The cloning was carried out using the dried In-Fusion™ enzyme mix to which 7 µL ddH_2O, 2 µL purified PCR product and 1 µL prepared vector were added before incubation at 42°C for 30 min. Each reaction was stopped by addition of 40 µL Tris-EDTA buffer. Three µL of each of the reactions were transformed into competent OmniMAX2-TI^R cells (Invitrogen) and plated in 24 well plates containing LB, carbenicillin, IPTG and XGal for blue-white screening (positive clones in white). Positive clones were confirmed by colony PCR.

For tracking of the constructs within the text, an OPPF number is given. Constructs tested for expression and their identification numbers are given in Table 3-5.

3.2.3 Construction of baculoviruses and insect cells lines for expression

For generation of the baculoviruses the protocol developed at OPPF was followed (Berrow *et al.*, 2007). *Sf9* cells were grown in Sf900 II-SFM media and plated in 24-well plates to a density of ~ 5 x 10^5 cells mL^{-1} at 500 µL per well. Attachment to the well surface took place within 1 h incubation at room temperature prior to co-transfection. For transfection 4 µL (= 100-500 ng) of each pOPIN vector, 2.5 µL (= 500 ng) linearised *Autographa califonica* bacmid (Durocher *et al.*, 2002), 0.75 µL FuGene® HD (Promega) and 50 µL Sf900 II-SFM were mixed and pre-incubated for 30 min at room temperature before addition to the adherent *Sf*9 cells. The primary viral stocks were harvested manually from the media 168 h post-transfection. These virus stocks were used for infection of fresh *Sf*9 cells. For this, *Sf*9 cells were grown to a density of ~ 1 x 10^6 cells mL^{-1}, plated in 24-well plates with 500 µL per well and allowed to attach to the well surface 1 h prior to infection. Adherent cells were infected with 5 µL primary virus stock, and harvested manually from the

supernatant 168 h post-infection and stored frozen at -80°C. The final virus stocks were harvested and stored at OPPF.

Table 3-5: TK constructs cloned for expression with the baculovirus system.
Cloning performed in 96 well plate format (using half of the plate); "well" gives the location in the plate, OPPF number for tracking in the OPPF database.

Well	Gene name	Amino acid range	M_w of expected protein (+ tag) [Dalton]	Vector pOPIN	OPPF No.
A01	*Lm*TK	2-284	31471.0 (+969)	E	11318
B01	*Lm*TK	2-185	21045.2 (+969)	E	11319
C01	*Lm*TK	2-185	21045.2 (+2158)	F	11320
D01	*Tc*TK	2-313	34604.4 (+2158)	F	11435
E01	*Tc*TK	2-228	25402.2 (+2158)	F	11438
F01	*Tc*TK	45-313	29573.6 (+2158)	F	11441
G01	*Tc*TK	45-228	20371.4 (+2158)	F	11444
H01	*Tc*TK	2-313	34604.4 (+1274)	I	11433
A02	*Tc*TK	2-228	25402.2 (+1274)	I	11436
B02	*Tc*TK	45-313	29573.6 (+1274)	I	11439
C02	*Tc*TK	45-228	20371.4 (+1274)	I	11442
D02	*Tc*TK	2-313	34604.4 (+969)	E	11434
E02	*Tc*TK	2-228	25402.2 (+969)	E	11437
F02	*Tc*TK	45-313	29573.6 (+969)	E	11440
G02	*Tc*TK	45-228	20371.4 (+969)	E	11443
H02	*Tb*TK	2-455	50713.2 (+2158)	F	11335
A03	*Tb*TK	2-382	42753.4 (+2158)	F	11338
B03	*Tb*TK	2-180	20435.0 (+2158)	F	11341
C03	*Tb*TK	198-455	28496.2 (+2158)	F	11344
D03	*Tb*TK	198-382	20536.3 (+2158)	F	11347
E03	*Tb*TK	2-455	50713.2 (+1274)	I	11333
F03	*Tb*TK	2-382	42753.4 (+1274)	I	11336
G03	*Tb*TK	2-180	20435.0 (+1274)	I	11339
H03	*Tb*TK	198-455	28496.2 (+1274)	I	11342
A04	*Tb*TK	198-382	20536.3 (+1274)	I	11345
B04	*Tb*TK	2-455	50713.2 (+969)	E	11334
C04	*Tb*TK	2-382	42753.4 (+969)	E	11337
D04	*Tb*TK	2-180	20435.0 (+969)	E	11340
E04	*Tb*TK	198-455	28496.2 (+969)	E	11343
F04	*Tb*TK	198-382	20536.3 (+969)	E	11346

3.2.4 Expression tests of TKs with the Baculovirus system

Expression tests from the TK baculoviruses were carried out by Heather Rada at OPPF following standard procedures. *Sf*9 cells were infected with 50 µL of each virus stock described in 3.2.3. Cells were harvested 72 h after infection, lysed by incubation for 30 min in lysis buffer (NPI-10 Tween) complemented with lysozyme and either benzonase or DNase type 1. Subsequently cell debris was removed by centrifugation and expressed Histagged protein was isolated using magnetic Ni-NTA beads. After washing twice with buffer NPI-20, bound Histagged protein was eluted with NPI-250 (containing 250 mM Imidazole). The eluates were analysed by SDS-PAGE, Figure 3-6.

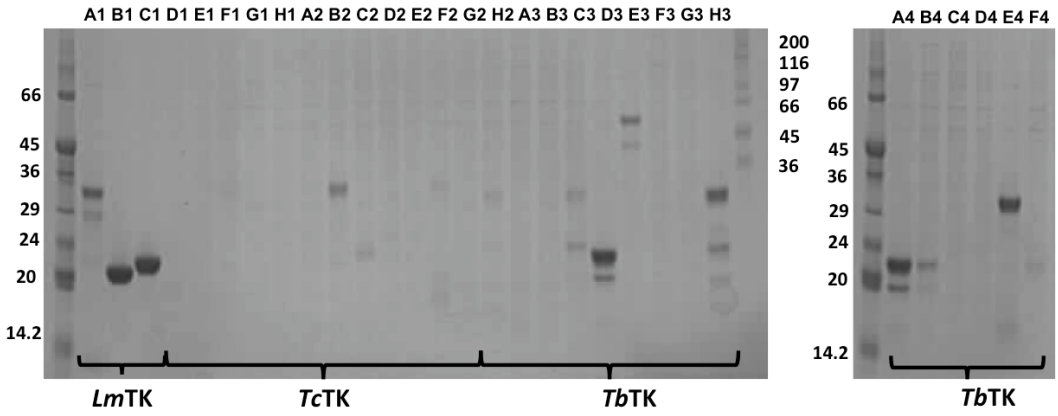


Figure 3-6: SDS-PAGE of TK expression tests with the baculovirus system.

First and last lanes are protein standards. Each other lane contains a NiNTA purified expression test sample from the baculovirus expression test. Lane labelling corresponds to the well numbers in Table 3-5.

3.2.5 Expression of *Lm*TK and *Tb*TK in Sf9 cells

After identification of baculovirus stocks leading to soluble expression of *Lm*TK and *Tb*TK in *Sf*9 cells, these were amplified and used for larger scale expression. Up-scaled expression of these was carried out at OPPF by Joanne Nettleship. Cells of 1 L cultures were harvested and frozen at -80°C before shipment to York.

3.2.6 Purification and Characterisation of *Lm*TK

The cell pellets from 1 L Sf9 cell cultures expressing the protein (11319 or 11320) were resuspended in 35 mL buffer A (20 mM Tris pH 8.0, 300 mM NaCl, 30 mM imidazole, 1 mM DTT and 100 µM PMSF) and lysed by sonication. This was followed by addition of 25-29 U benzonase, incubating for 10 min at room temperature. The cell debris was removed by centrifugation for 30 min at 5000 rcf and 4°C. The supernatant was loaded onto a 1 mL HisTrap crude FF column for 1 h in a loop on ice. After washing with 10 mL buffer A the column was connected to an AKTA purifier washed with another 10 mL of buffer A and eluted with an imidazole gradient from 30 to 500 mM in buffer A. The protein containing fractions were pooled and concentrated up to 4 mL total volume. In order to stabilise the protein, it was diluted with 1 mL SEC buffer (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM DTT) containing 1 µM $ZnSO_4$.

*Lm*TK 11319 (containing a non-cleavable C-terminal Histag) was concentrated further to 250 µL total volume and directly loaded onto a Superdex 200 column (Figure 3-7 A). Size exclusion chromatography was performed using SEC buffer (composition described above). Protein purity was assessed by SDS-PAGE and the protein-containing fractions were pooled (Figure 3-7 B). As the protein started to precipitate, it was split into two fractions and stabilised by addition of either 1 mM dThd + 1 mM AppNHp + 3 mM $MgCl_2$ or 1 mM dTTP + 3 mM $MgCl_2$. The stabilised protein was concentrated to 2.4 mg mL^{-1} (dThd + AppNHp + $MgCl_2$) and 7.3 mg mL^{-1} (dTTP + $MgCl_2$), aliquoted, flash frozen in liquid nitrogen and then stored at -80°C.

*Lm*TK 11320 (containing a cleavable N-terminal Histag) was diluted with SEC buffer in the ratio 1:2. To cleave off the Histag, HRV 3C protease was added in a ratio of 1:100 (protease:*Lm*TK) and incubated at 4°C overnight. The proteolysis was monitored by SDS-PAGE and reached ~ 90 % after 16 h. To remove the Histag and non-cleaved *Lm*TK the protein buffer was exchanged into lysis buffer using Centrifuge filters (Amicon® Ultra Centrifugal Filter from Millipore, 30 kDa cutoff) and nickel affinity purification was performed with the same buffers as before. The

cleaved (untagged) *Lm*TK in the flowthrough was concentrated to 250 µL total volume and injected onto an equilibrated Superdex 200 column (Figure 3-7A). The same SEC buffer was used as for *Lm*TK 11319. The separation of cleaved and uncleaved *Lm*TK and the final purity was analysed by SDS-PAGE (Figure 3-7C). A high molecular weight band on the SDS-PAGE indicated the presence of a HRV 3C contamination, which was not removed, as it did not bind to the HisTrap column and co-eluted with the *Lm*TK in size exclusion chromatography. Alternatively, this band might have been caused by *Lm*TK dimers. Due to the limited amount of protein available, this was not further investigated. Protein-containing fractions were pooled and concentrated to 15.1 mg mL^{-1} in the presence of 1 mM dThd, 1 mM AppNHp and 3 mM $MgCl_2$. Concentrated protein was aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

The polydispersity of *Lm*TK was determined by dynamic light scattering (DLS) at 20°C, with the protein diluted in SEC buffer to a total concentration of 3 mg mL^{-1} with a total sample volume of 20 µL. DLS measurements were carried out using a DynaPro™ (Protein Solutions™).

Although the protocols were kept identical *Lm*TK from subsequent purifications could not be stabilised and precipitated completely during lysis (11320) or after nickel affinity chromatography (11319).

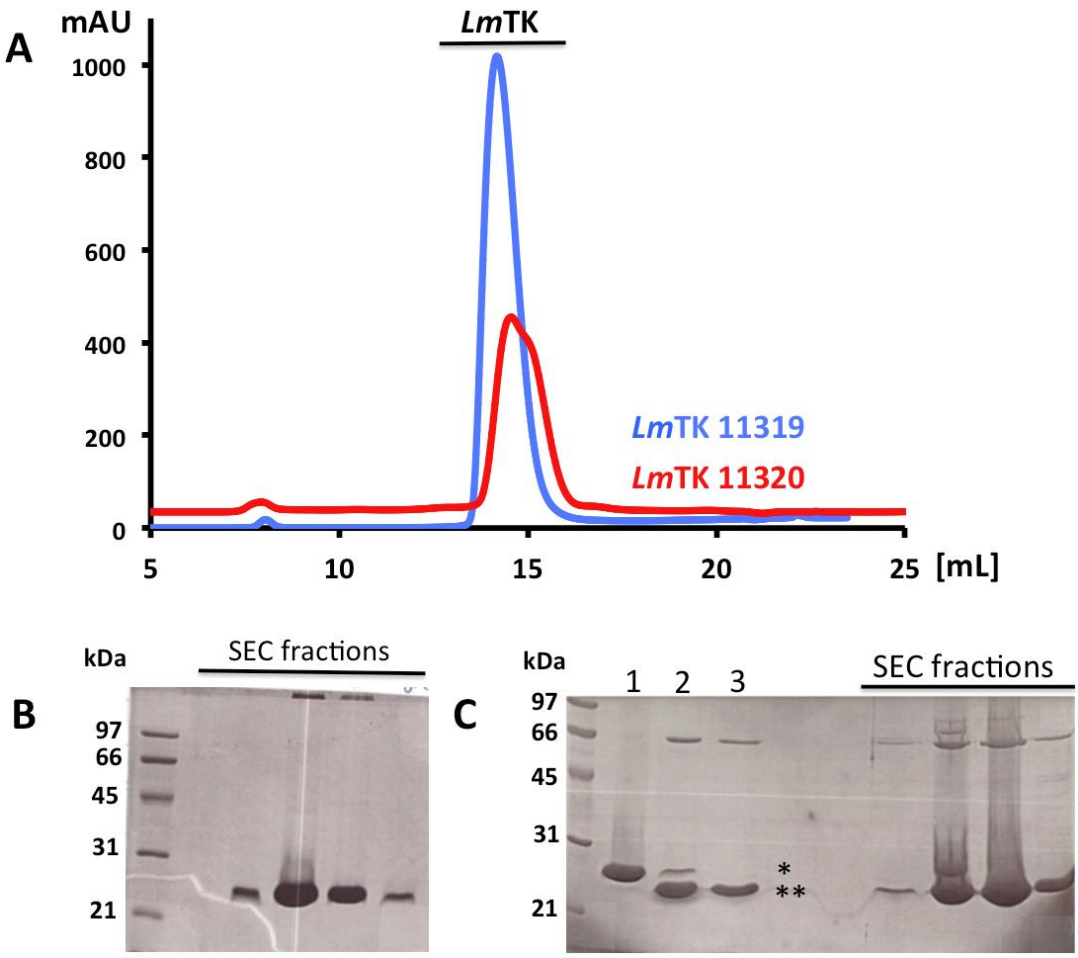


Figure 3-7: Purification of *Lm*TK.

(A) Superposition of the chromatogram (absorption at 280 nm) of construct 11319 (blue) with that from 11320 (red). (B) SDS-PAGE analysis of protein containing fractions of *Lm*TK11319 after SEC. (C) SDS-PAGE analysis of *Lm*TK11320 purification showing pooled protein fractions from the first HisTrap column (lane 1), the protein after HRV3C proteolysis overnight (lane 2) and flowthrough of the second HisTrap column (lane 3), followed by the protein containing fractions of the SEC. * *Lm*TK 11320 with Histag, ***Lm*TK 11320 without the Histag.

3.2.7 Purification and Characterisation of *Tb*TK

For purification of *Tb*TK constructs 11343, 11345 and 11347 the pellets of 1 L *Sf*9 culture were thawed and lysed as described for *Lm*TK. Nickel affinity purification followed the same protocol and used the same buffer as was used for *Lm*TK. After nickel affinity chromatography of 11343 (non-cleavable C-terminal Histag) the

protein yield was very poor and many contaminants were still present, so this construct was not further purified.

*Tb*TK 11345 (non-cleavable N-terminal Histag) was stabilised after nickel affinity chromatography by addition of 1 µM $ZnSO_4$, 5 mM $MgCl_2$, 0.5 mM dThd and 0.5 mM AppNHp. It was concentrated to 500 µL total volume and injected onto an equilibrated Superdex 200 column in SEC buffer (10 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT) (Figure 3-8 A). The protein containing fractions were pooled and after addition of 10 mM $MgCl_2$, 1 mM dThd and 1 mM AppNHp, concentrated to 35 mg mL^{-1} and aliquoted. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

*Tb*TK 11347 (cleavable N-terminal Histag) appeared stable after nickel affinity chromatography, but similar to 11343, the yield was very low and contaminations were present. The Histag was cleaved by HRV 3C proteolysis over 36 h at 4°C in dialysis against buffer A. During proteolysis the protein showed strong precipitation, which was removed by centrifugation before nickel affinity purification. Flowthrough and elution fractions contained *Tb*TK and HRV 3C protease and were concentrated separately before loading onto Superdex 200 column (Figure 3-8 B). SDS-PAGE analysis revealed the presence of 3C protease contamination in all fractions of both size exclusion chromatography runs and only a very small amount of *Tb*TK (< 200 µg), insufficient for further studies.

A second batch of *Tb*TK 11347, which was mislabelled at the OPPF as *Lm*TMPK 11357, was stabilised after nickel affinity chromatography by addition of 1 mM TMP, 1 mM AppNHp and 3 mM $MgCl_2$. The protein was concentrated to ~500 µL and was loaded onto a Superdex 200 column (Figure 3-8 C). Protein purity was monitored by SDS-PAGE. The protein-containing fractions were pooled and concentrated to 35 mg mL^{-1} in the presence of 1 mM TMP, 1 mM AppNHp and 3 mM $MgCl_2$. Aliquots were flash frozen in liquid nitrogen and stored at -80°C. Electrospray ionisation mass spectrometry (ESI-MS) and tandem mass spectrometry (MS/MS, using a Bruker autoflex III MALDI-TOF/TOF) after trypsin digest, confirmed the protein to be *Tb*TK, most likely construct 11347. The polydispersity was

determined by DLS at 20°C using protein diluted in SEC buffer to a concentration of 3 mg mL^{-1} with a total sample volume of 20 µL.

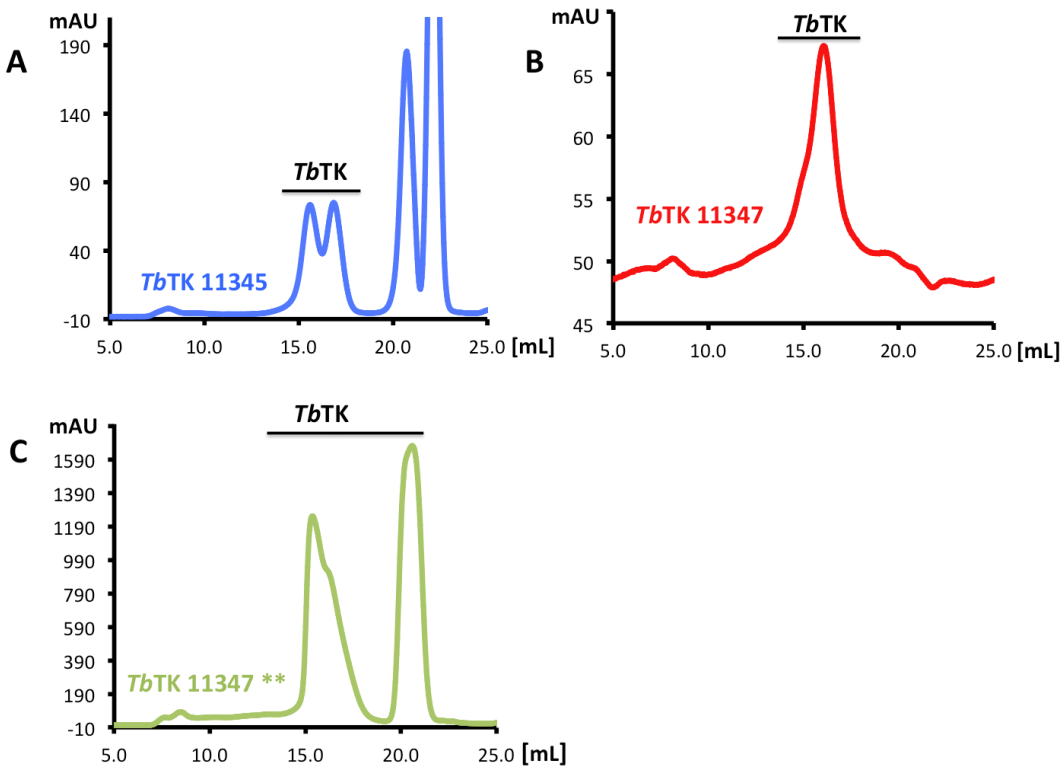


Figure 3-8: SEC chromatograms of *Tb*TK purifications.

The black bars indicate the peaks caused by *Tb*TK (identified by SDS-PAGE analysis), peaks at elution volumes above 20 mL are most likely caused by nucleotide that was added for protein stabilisation. (A) Size exclusion chromatogram of *Tb*TK 11345 (stabilised with dThd). (B) SEC chromatogram of *Tb*TK 11347 (not stabilised). (C) Size exclusion chromatogram of *Tb*TK 11347 (stabilised with TMP because of mislabelling by the OPPF).

3.2.8 Crystallisation of *Lm*TK and *Tb*TK

Initial screens (Index, PACT and JCSG) were set up in sitting-drop vapour diffusion format with 150 nL protein solution including additives plus an equal volume of precipitant, using a Mosquito robot (TTP Labtech, UK). Crystals were optimised using hanging-drop vapour diffusion with 1 µL protein solution plus an equal volume of precipitant. To produce crystals in complex with the bi-substrate analogue AP_5dT 1 mM ligand was added to *Lm*TK 11319 containing dThd and AppNHp and incubated on ice, allowing AP_5dT to replace the substrates in the

active site. In order to crystallise *Tb*TK in a different crystal form, cross seeding experiments were carried out using an Oryx 4 robot (Douglas Instruments, UK), using the Index screen in sitting-drop vapour diffusion with 250 nL protein solution plus 170 nL precipitant and 80 nL seed stock. The seed stock was made up using seed beads from *Lm*TK 11320 crystals in buffer as for the *Lm*TK-dThd crystal. The *Tb*TK crystals obtained without seeding were cryo-protected before vitrification in precipitant solution containing 10 % glycerol. All other crystals were not cryo-protected. The conditions leading to crystals that allowed structure solution as well as the ligands bound in the active sites are summarised in Table 3-6.

Table 3-6: Crystallisation conditions of *Lm*TK and *Tb*TK.

Crystal name	Additives in protein solution	Crystallisation condition	Ligand bound in the active site
*Lm*TK-dThd	1 mM dThd, 1 mM AppNHp, 3 mM $MgCl_2$	1.4 M Na/K PO_4 pH 5.4	dThd, PO_4^{2-}
*Lm*TK-dTTP	1 mM dTTP 3 mM $MgCl_2$	50 mM HEPES pH 8.0, 36% (v/v) Pentaerythritol propoxylate	dTTP, Mg^{2+}
*Lm*TK- AP_5dT	1 mM AP_5dT, 1 mM dThd, 1 mM AppNHp, 3 mM $MgCl_2$	0.1 M tri-Na citrate pH 5.5, 20 % (w/v) PEG 3K	AP_5dT
*Tb*TK-dThd/dTMP	1 mM TMP, 1 mM AppNHp, 3 mM $MgCl_2$	0.1 M HEPES pH 7.0, 0.9 M Succinic acid	dThd/dTMP
*Tb*TK- dTMP-PO_4^{2-}-a	1 mM TMP, 1 mM AppNHp, 3 mM $MgCl_2$	0.1 M HEPES pH 7.0, 0.9 M Succinic acid, 1% (w/v) PEG 2K MME	dTMP, PO_4^{2-}
*Tb*TK- dThd-Glycerol	1 mM TMP, 1 mM AppNHp, 3 mM $MgCl_2$	0.1 M HEPES pH 7.0, 0.8 M Succinic acid	dThd, glycerol
*Tb*TK- dTMP-PO_4^{2-}-b	1 mM TMP, 1 mM AppNHp, 3 mM $MgCl_2$ Seed stock of LmTK-dThd sister crystals	0.15 M DL-Malic acid pH 7.0, 20% (w/v) PEG 3350	dTMP, PO_4^{2-}

3.2.9 Data collection and Structure solution of *Lm*TK and *Tb*TK

Diffraction data of *Lm*TK and *Tb*TK crystals were collected on beamline I04 of the Diamond Light Source (DLS). Binding of zinc ions was confirmed by X-ray fluorescence experiments for two of the *Lm*TK and *Tb*TK crystals. All computations were carried out using the CCP4 software package (Winn *et al.*, 2011). The diffraction images were processed with MOSFLM (Battye *et al.*, 2011) and AIMLESS (Evans & Murshudov, 2013). The image quality of the *Lm*TK-dTTP was poor at some rotation angles resulting a reduction of the completeness to 94.7 %. Crystallographic data and statistics are summarised in Table 3-7 - 9.

The *Lm*TK and *Tb*TK structures were determined by molecular replacement using PHASER (McCoy, 2007). At the time the *Hs*TK1 (PDB ID: 1w4r) had the highest homology with a sequence identity of 49% and was selected as search model for *Tb*TK. Subsequently, the structure of *Tb*TK-dThd/dTMP with its 58 % sequence identity to *Lm*TK was chosen as search model for *Lm*TK. Refinement was performed using the maximum-likelihood restrained method in REFMAC5 (Murshudov *et al.*, 2011) with one TLS group per chain and automatically generated local NCS. For the first few rounds external restraints generated from *Tb*TK-dThd/dTMP using ProSMART (Nicholls *et al.*, 2012) in REFMAC5 were used. The last rounds of refinement were carried out with the jelly body refinement option (with sigma 0.02). Manual fitting of the electron-density maps, model building and fitting of the ligands were carried out in COOT (Emsley *et al.*, 2010). The models were validated in COOT and *MolProbity* (Chen *et al.*, 2010).

Table 3-7: Crystallographic data and statistics for the *Lm*TK crystals.

In this and all subsequent data tables: Data for the highest resolution shell shown in brackets, the Ramachandran Plot was calculated using COOT (Emsley *et al.* 2010), the poor rotamer analysis was carried out in *Molprobity* (Chen *et al.* 2010).

	LmTK-dThd- PO_4^{2-}	***Lm*TK-dTTP**	***Lm*TK-AP_5dT**
Data collection			
Diffraction Source	DLS beamline I04	DLS beamline I04	DLS beamline I24
Wavelength (Å)	0.9795	0.9795	0.9686
Temperature (K)	100	100	100
Detector	Pilatus 6M-F	Pilatus 6M-F	Pilatus3 6M
Rotation range per image (°)	0.2	0.2	0.1
Total Rotation range (°)	72	80	90
Crystal data			
Space group	$P6_522$	P1	$P6_522$
a, b, c (Å)	121.1, 121.1, 116.2	61.60, 61.63, 110.1	64.15, 64.15, 324.6
α, β, γ (°)	90, 90, 120	81.07, 85.8, 74.94	90, 90, 120
Resolution (Å)	52.46-2.74 (2.88-2.74)	59.45-3.00 (3.16-3.00)	45.84-2.40 (2.49-2.40)
Total Reflections	271650 (35991)	58503 (8708)	137341 (12676)
Unique Reflections	13223 (1772)	29290 (4271)	16728 (1684)
Completeness (%)	96.3 (100)	94.7 (94.7)	99.9 (99.6)
Redundancy	20.5 (20.3)	2.0 (2.0)	8.2 (7.5)
R_{merge}	0.110 (1.174)	0.135 (0.600)	0.118 (0.690)
Mean [I/σ (I)]	17.3 (2.9)	6.2 (2.1)	11.0 (2.7)
Matthew's coefficient	2.93	2.26	2.19
Refinement statistics			
Reflections, working set	12506	27884	15762
Reflections, test set	648	1404	829
Resolution Range (Å)	52.46-2.74	59.45-3.00	45.84-2.40
R-factor	0.2822	0.2603	0.1955
R	0.3104	0.2784	0.2569
No. of non-H atoms			
Protein	2370	9673	2598
Ligand	49	232	112
Water	13	53	57
Mean B factors ($Å^2$)			
Protein	88.8	49.5	29.8
Ligands	71.7	38.1	51.6
Water	57.1	22.8	40.4
RMS deviation from ideal			
Bond length (Å)	0.0065	0.0094	0.0124
Bond angles (°)	1.2916	1.4449	1.6943
Ramachandran Plot (%)			
Residues in favored region	94.94	96.00	95.43
Residues in allowed region	5.06	4.00	4.57
Outliers	0	0	0
Molprobity score	1.67	1.69	1.07
Poor Rotamers (%)	0	0.11	0.75
PDB ID	**4uxi**	**4uxj**	**4uxh**

Table 3-8: Crystallographic data and statistics for *Tb*TK crystals.

	***Tb*TK-dThd-glycerol**	***Tb*TK-dThd/dTMP**
Data collection		
Diffraction Source	DLS beamline I04	DLS beamline I02
Wavelength (Å)	0.9795	0.9795
Temperature (K)	100	100
Detector	Pilatus 6M	Pilatus 6M-F
Rotation range per image (°)	0.2	0.2
Total Rotation range (°)	220	180
Crystal data		
Space group	$P4_322$	$P4_322$
a, b, c (Å)	114.7, 114.7, 105.6	115, 115, 104.8
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	44.25-2.30 (2.39-2.30)	64.24-2.20 (2.27-2.20)
Total Reflections	456390 (47217)	422851 (33374)
Unique Reflections	31864 (3298)	36299 (3079)
Completeness (%)	99.9 (99.9)	100 (100)
Redundancy	14.3 (14.3)	11.6 (10.8)
R_{merge}	0.085 (1.024)	0.105 (0.885)
Mean [I/σ (I)]	17.0 (2.6)	11.8 (2.6)
Matthew's coefficient	4.01	3.81
Refinement statistics		
Reflections, working set	30146	34470
Reflections, test set	1708	1757
Resolution Range (Å)	44.25-2.30	64.24-2.20
R-factor	0.1680	0.1800
R_{free}	0.1888	0.1889
No. of non-H atoms		
Protein	2598	2588
Ligand	59	36
Water	94	123
Mean B factors (Å^2)		
Protein	42.2	39.3
Ligands	56.7	52.8
Water	55.5	59.0
RMS deviation from ideal		
Bond length (Å)	0.0202	0.0201
Bond angles (°)	2.0820	1.9079
Ramachandran Plot (%)		
Residues in favoured region	97.26	96.93
Residues in allowed region	2.44	3.07
Outliers	0.30	0
Molprobity score	1.44	1.36
Poor Rotamers (%)	0.36	0
PDB ID		

Table 3-9: Crystallographic data and statistics for *Tb*TK crystals.

	***Tb*TK-dTMP-PO_4^{2-}-a**	***Tb*TK-dTMP-PO_4^{2-}-b**
Data collection		
Diffraction Source	DLS beamline I02	DLS beamline I04
Wavelength (Å)	0.9795	
Temperature (K)	100	100
Detector	Pilatus 6M-F	Pilatus 6M
Rotation range per image (°)	0.2	0.2
Total Rotation range (°)	180	220
Crystal data		
Space group	$P4_322$	C222
a, b, c (Å)	114.8, 114.8, 104.5	87.61, 97.0, 304.4
α, β, γ (°)	90, 90, 90	90, 90, 89.99
Resolution (Å)	43.92-2.20 (2.27-2.20)	65.02-2.80 (2.95-2.80)
Total Reflections	441889 (37765)	173354 (20274)
Unique Reflections	36045 (3273)	31168 (4599)
Completeness (%)	100 (100)	95.9 (98.7)
Redundancy	12.3 (11.5)	5.6 (4.4)
R_{merge}	0.071 (1.215)	0.098 (0.373)
Mean [I/σ (I)]	18.9 (2.3)	8.5 (2.7)
Matthew's coefficient	3.97	2.34
Refinement statistics		
Reflections, working set	34261	29477
Reflections, test set	1749	1570
Resolution Range (Å)	43.92-2.20	65.02-2.80
R-factor	0.1810	0.2062
R_{free}	0.1887	0.2344
No. of non-H atoms		
Protein	2582	7420
Ligand	54	162
Water	76	32
Mean B factors (Å^2)		
Protein	43.1	48.6
Ligands	60.3	60.5
Water	60.2	47.5
RMS deviation from ideal		
Bond length (Å)	0.0203	0.0132
Bond angles (°)	2.0027	1.7817
Ramachandran Plot (%)		
Residues in favoured region	96.60	96.06
Residues in allowed region	3.40	3.94
Outliers	0	0
Molprobity score	1.43	1.39
Poor Rotamers (%)	0.36	0.26
PDB ID		

3.3 Results

3.3.1 Expression and purification of *Lm*TK and *Tb*TK

In *E. coli*, constructs coding for N-terminal Histagged TKs from all three species did not give soluble and stable protein in sufficient quantities. Fusion to an MBP tag helped *Lm*TK and *Tb*TK constructs to express soluble protein, but these were not stable after removal of the MBP tag and precipitated when separated from non-cleaved protein. The C-terminal Histagged *Lm*TK showed soluble expression, but the protein was not stable during purification showing rapid precipitation. Collaborators at the Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain, expressed full-length *Lm*TK in *E. coli* BL21 (DE3) cells, obtaining sufficient amounts for kinetic measurements. The amounts obtained with their protocol were not high enough for crystallisation experiments and problems with protein stability hindered further use of protein produced with this method.

When expressed in *Sf*9 cells using the baculovirus system at OPPF two *Lm*TK constructs, spanning amino acid residues 2-185, and three *Tb*TK constructs, spanning amino acid residues 198-455 and 198-383, showed soluble expression and potentially good yield. These constructs were chosen for expression in bigger volumes (1 L cell cultures) and four of these could be purified in sufficient quality and quantity for crystallisation experiments. While purification of each of the four constructs was successful once, subsequent purifications of the same constructs gave protein precipitating during purification. The reason for the instability of TKs in the second batch is unclear and precipitated protein could not be refolded.

3.3.2 Crystallisation of *Lm*TK and *Tb*TK

*Lm*TK constructs 11319 and 11320 crystallised in several conditions, but showed limited diffraction quality and seeding experiments did not improve the crystal quality. The *Tb*TK construct 11347 crystallised, 11343 and 11345 did not.

3.3.3 Structures of *Lm*TK and *Tb*TK

3.3.3.1 Overall structure

The TK constructs forming diffracting crystals were of heavily truncated versions, removing the C-terminal extensions of unknown function. For *Tb*TK only the catalytically active of the two TK domains expressed in soluble form, so that the solved structure represents less than half of the full-length protein. However, the core structures of *Lm*TK and *Tb*TK were solved, showing a high sequence homology to type II TK from other species. They show nearly identical folds with an overall RMSD of 0.30 Å (superposition shown in Figure 3-9 B) and to *Hs*TK1 (PDB ID: 1w4r, (Welin *et al.*, 2004) with an overall RMSD of 0.31 Å. Each TK protomer shows the typical bilobal structure with the active site sitting between the domains (Figure 3-9).

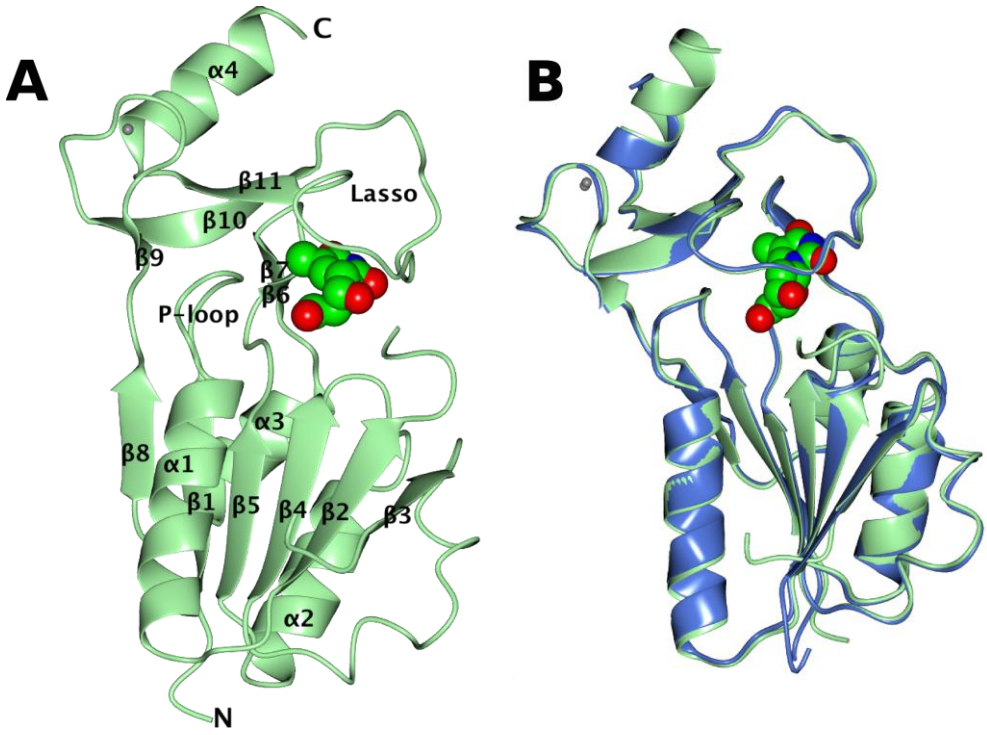


Figure 3-9: Structures of *Tb*TK and *Lm*TK.

(A) *Tb*TK-dThd-glycerol with dThd shown as spheres coloured by atom type. (B) Overlay of *Tb*TK-dThd-glycerol (green) and *Lm*TK (blue, ligand not shown). Superposition in this figure and all subsequent ones was carried out using SSM (Krissinel & Henrick, 2004).

The large N-terminal α/β-domain consists of a central six-stranded parallel β-sheet (β1, β2, β3, β4, β5, β8), flanked by a long α-helix (α1) and a loop region of 15-20 amino acids on one side and 2 shorter α-helices (α2 and α3) and a loop on the other. The conserved P-loop, responsible for phosphate binding and essential for the kinase activity, is located between β1 and α1. This domain shows similarity to ATP-binding domains of the RecA-F_1-ATPase family, which is also found in helicases and DNA repair proteins (Sawaya *et al.*, 1999, Welin *et al.*, 2004). There is a flexible loop region between residues 245 and 260 in *Tb*TK (42-59 in *Lm*TK), which is disordered in most TK structures published to date.

The smaller C-terminal domain (Figure 3-10) contains three β-strands (β9-11) and a short α-helix (α4) at the very C-terminus of the structure. It contains a structural Zn^{2+} ion tetrahedrally coordinated by four cysteine residues, C341 and C344 (*Tb*TK) in the loop between β9 and β10 and the C374 and C377 (*Tb*TK) at the start of the α4-helix. The long loop between β10 and β11 forms a lasso structure covering the thymidine-binding site, held in place by conserved R354 and Y370 residues, stabilising the main chain by hydrogen bonds. The conformations of the corresponding side chains in the *Lm*TK structures (C140, C143, C173, C176, R153 and Y169) are in identical positions.

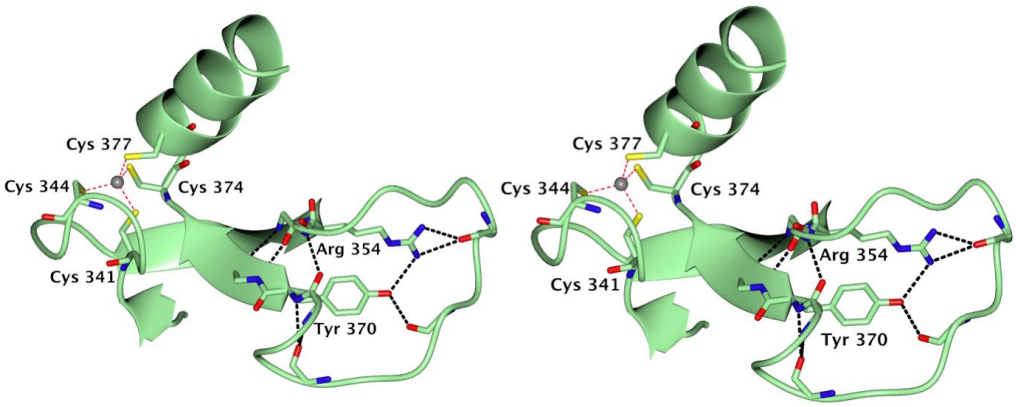


Figure 3-10: The C-terminal domain of *Tb*TK.

The Zn^{2+} coordinating cysteine residues (C341, C344, C374 and C377) and lasso-stabilising arginine (R354) and tyrosine (Y370) residues shown as cylinders coloured by atom type.

3.3.3.2 Substrate and Inhibitor binding

*Lm*TK crystallised in the presence of the substrate dThd + AppNHp, and the inhibitors dTTP and AP_5dT. The crystal grown with dThd + AppNHp shows clear electron density for dThd (Figure 3-11 A). While there was no density for AppNHp, there seems to be a partially occupied phosphate where the β-phosphate of the AppNHp would be expected. The *Lm*TK-dTTP crystal clearly shows electron density for dTTP in the thymidine-binding site (Figure 3-11 B), the phosphates being coordinated by a magnesium ion while the ATP-binding site is empty. The *Lm*TK-AP_5dT crystal has two chains in the AU of which both show clear density for the thymidine moiety and the 5 phosphates, but only chain A exhibits good density for the adenosine moiety (Figure 3-11 C).

*Tb*TK was crystallised in complex with dThd (+glycerol), dThd/dTMP and dTMP + PO_4^{2-}. The *Tb*TK-dThd-glycerol crystal was soaked in cryo-protectant containing 10 % glycerol, dThd and AppNHp in order to replace the co-crystallised TMP. This crystal shows clear electron density for the dThd, but none for AppNHp and seems to have glycerol molecules bound from the cryo-protection (Figure 3-12 A). The *Tb*TK-dThd/dTMP crystal shows clear electron density for the dThd from the dTMP it was co-crystallised with (Figure 3-12 B), as well as additional density most probably resulting from a partially occupied PO_4^{2-} (from the dTMP), which was not modelled because of the bad electron density fit upon refinement with dTMP. There is no density for AppNHp and the expected position of its phosphate groups seems occupied by water molecules. The *Tb*TK-dTMP-PO_4^{2-}-a crystal shows very similar density to the *Tb*TK-dThd/dTMP crystal, with a well-resolved dThd, but a higher occupied PO_4^{2-} (from dTMP) in the active site (Figure 3-12 C). Additionally, it shows clear electron density for a PO_4^{2-} where the β-PO_4^{2-} of the AppNHp was expected to be bound. To obtain *Tb*TK crystals in a different space group, which allows AppNHp to bind in the site expected from comparison to other TK structures, seeding experiments were carried out. This led to the crystals *Tb*TK-dTMP-PO_4^{2-}-b with the phosphate in the β-position (Figure 3-12 D).

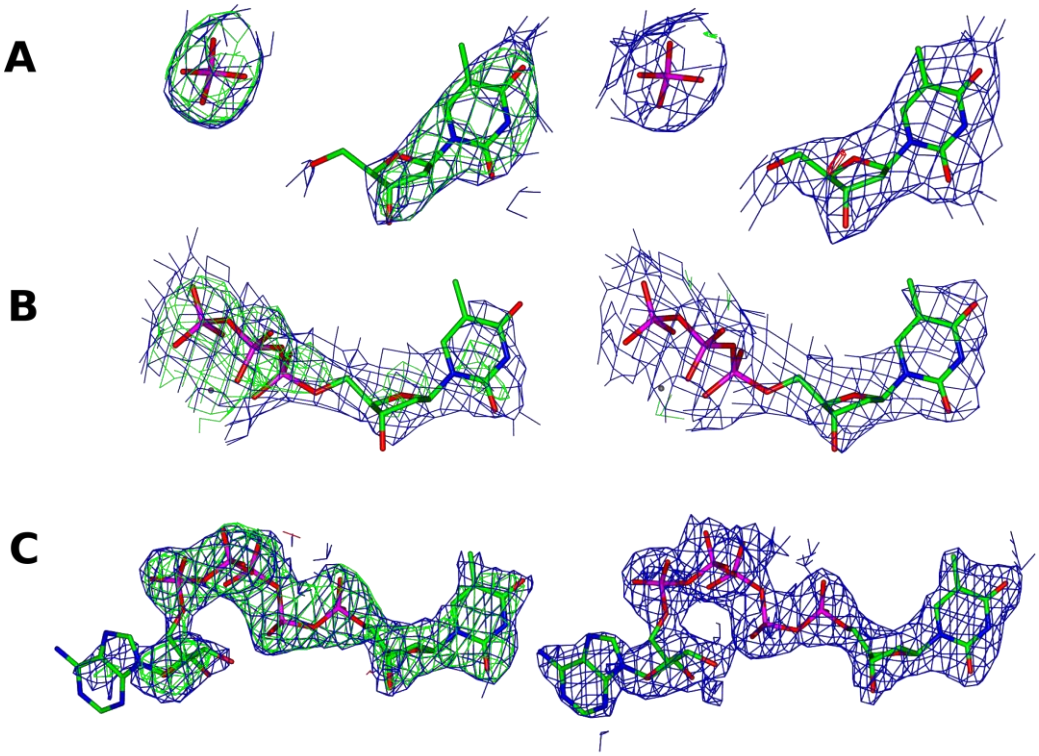


Figure 3-11: Electron densities for the ligands bound in the *Lm*TK structures.
Shown are the ligands in the active sites of the chain A from (A) *Lm*TK-dThd-PO_4^{2-} (B) *Lm*TK-dTTP (C) *Lm*TK-AP_5dT. The ligands are shown as cylinders coloured by atom type, mF_o-DF_c maps contoured at 3σ in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue. Shown are the maps generated before (left) and after (right) ligand modelling. The weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c) were calculated using REFMAC5 (Murshudov *et al.*, 2011).

The active site residues forming the thymidine-binding site have identical conformations in all structures (Figure 3-13 A) and very similar to the conformation in other TKs with the exception of F321 in *Tb*TK, which is pushed outwards. A detailed description of the active site was given for *Hs*TK1 (Welin *et al.*, 2004), *Ba*TK and *Bc*TK (Kosinska *et al.*, 2007) among others, so that it can be described concisely for *Lm*TK and *Tb*TK. In brief, the dThd molecule, deeply buried between the two domains, is coordinated by hydrogen bonds to the main chain of residues, R361, L363 and G365 in *Tb*TK and Q160, L162 and G164 in *Lm*TK, located in the lasso domain, and Y316/Y115 (*Tb*TK/*Lm*TK) from the α/β-domain (Figure 3-13 A). Additionally, stacking interactions with Y370/Y169 (*Tb*TK/*Lm*TK) of the lasso domain and hydrophobic interactions of F289, L312 and F321 (*Tb*TK)/F88, L111 and F120

(*Lm*TK) on the α/β-domain are holding the thymidine ring in place. In the structures with dThd in the active site, the E286/E85 (*Tb*TK/*Lm*TK), essential for catalysis, forms a hydrogen bond to the 5'-oxygen of the ribose ring. It is therefore in a position to serve as a catalytic base, abstracting a proton from the oxygen, which is then enabled to perform a nuceophilic attack on the γ-phosphate of the phosphate donor.

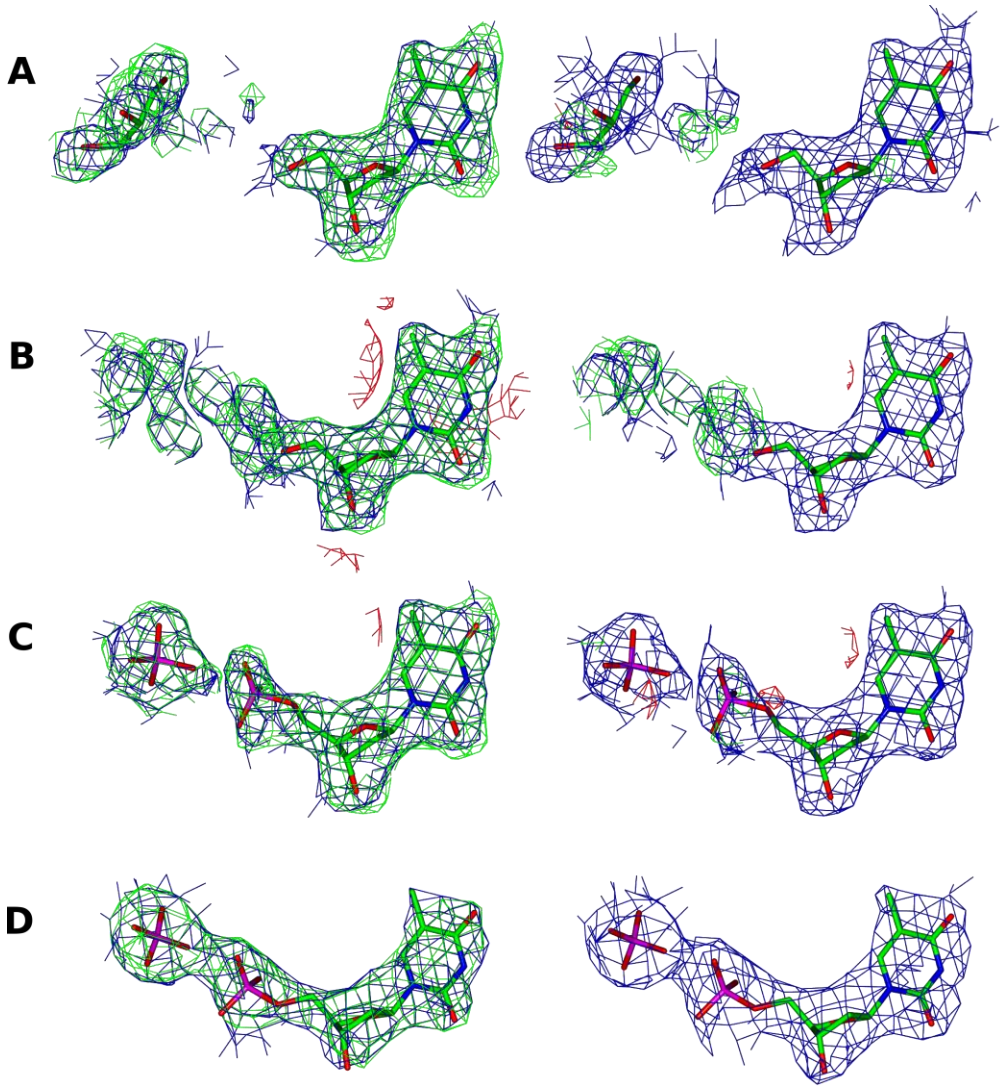


Figure 3-12: Electron densities for the ligands bound in the *Tb*TK structures.

Shown are the ligands in the active sites of the chain A from (A) *Tb*TK-dThd-glycerol, (B) *Tb*TK-dThd/dTMP, (C) *Tb*TK-dTMP-PO_4^{2-}-a and (D) *Tb*TK-dTMP-PO_4^{2-}-b. The ligands are shown as cylinders coloured by atom type, mF_o-DF_c maps contoured at 3σ in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue. Shown are the maps generated without (left) and with (right) ligands present in refinement. The weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c) were calculated using REFMAC5 (Murshudov *et al.*, 2011).

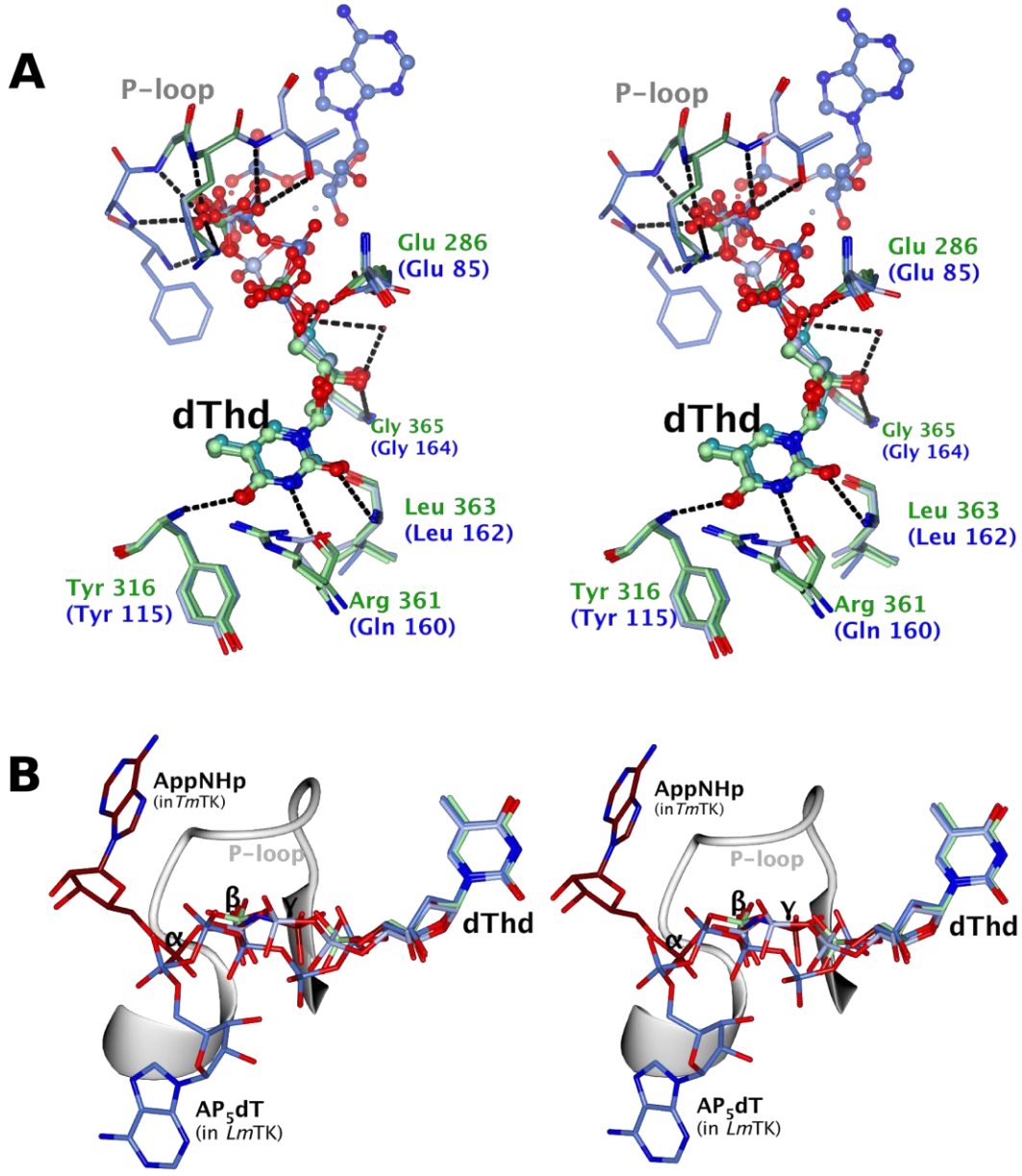


Figure 3-13: TK substrate binding site.

(A) Thymidine-binding site and P-loop in cylinder representation, coloured by atom type with carbons and phosphates coloured as follows: *Tb*TK-dThd/dTMP in light green, *Tb*TK-TMP-PO_4^{2-}-b in lawn green, *Lm*TK-dThd-PO_4^{2-} in ice blue, *Lm*TK-dTTP in light blue. Numbering according to *Tb*TK and *Lm*TK in brackets. Hydrogen bonds are shown as dashed lines. (B) Overlay of the substrates of *Tb*TK-TMP-PO_4^{2-}-b (light green), *Lm*TK-dTTP (ice blue), *Lm*TK-AP_5dT (light blue) and *Tm*TK-dThd-AppNHp, PDB ID: 2qq0 (tan). The position of the P-loop is identical for all structures, shown here is the one from *Lm*TK-AP_5dT.

None of the structures contains the phosphate donor mimic AppNHp. However, in some structures a phosphate is coordinated by the main chain of the P-loop in the expected position of the β-phosphate of the phosphate donor, as seen in the structure of *Tm*TK-dThd-AppNHp (PDB ID: 2qq0, Figure 3-13 B). In the structure of *Lm*TK-AP_5dT the dThd moiety is bound in the same site as in the other structures, but the phosphates show different positions. Four of the five phosphates of the AP_5dT show identical positions to the phosphates in the *Hs*TK1 structure in complex with AP_4dT (PDB ID: 2orv, Segura-Pena, Lutz*, et al.*, 2007), which has no electron density for the adenine moiety.

The expected ATP-binding site, according to the structures of *Ca*TK and *Tm*TK with adenosine bound in the phosphate donor site, is blocked in the structures of *Lm*TK and *Tb*TK by residues of the α1-helix of the adjacent, symmetry-related dimer, forming a closed tetramer, discussed in 3.3.3.3. The structure of *Lm*TK-AP_5dT shows clear electron density not only for the dThd moiety and the five phosphates, but also for the adenosine. The adenosine is not bound in the expected ATP-binding site identified in structures of *Ca*TK (PDB ID: 1xx6), *Tm*TK (PDB ID: 2qq0) and *Bc*TK (PDB ID: 2ja1), but instead is flipped about 180°, lying on the α/β-domain (Figure 3-13 B). The adenosine moiety is coordinated *via* a hydrogen bond of residue T19 and hydrophobic interactions of residues M22, K39 and Y46 (Figure 3-14).

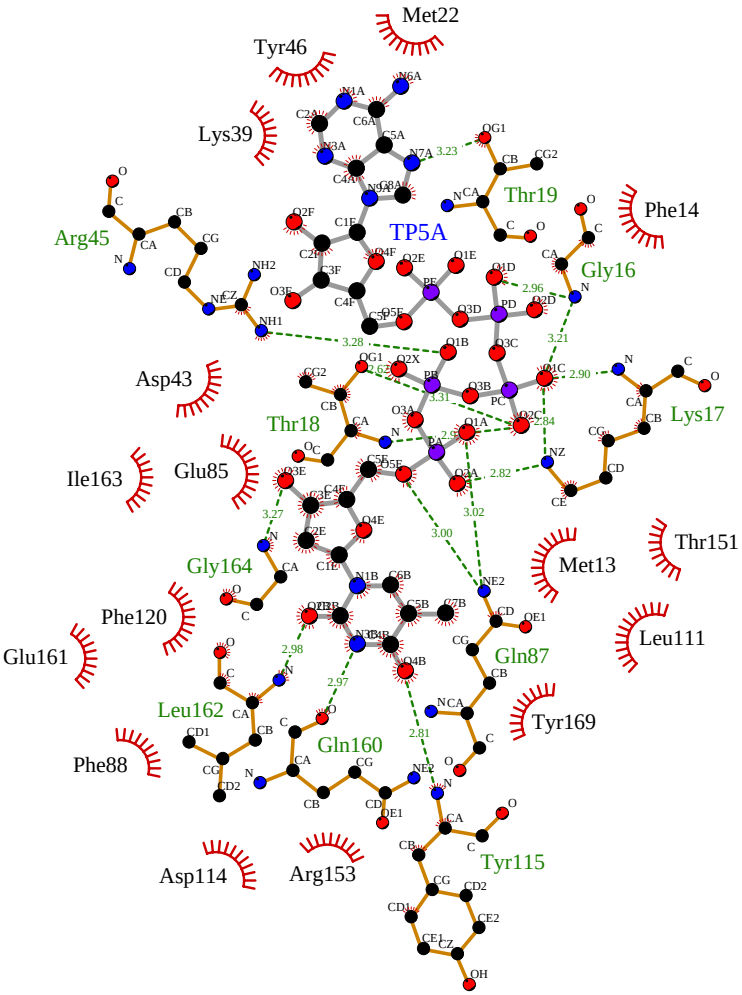


Figure 3-14: Schematic representation of the coordination of AP_5dT in the active site of *Lm*TK.
Shown are the ligand AP_5dT (TP5A) and residues coordinating it *via* hydrogen bonds (green labels, hydrogen bonds as dashed lines), and hydrophobic interactions (black labels, red half circles). Ligand and hydrogen bonding residues are shown as ball and stick models coloured by atom type. Figure drawn with LigPlot+ (Laskowski & Swindells, 2011).

This position may be an artefact caused by tetramer formation in solution prior to binding of the AP_5dT. It could indicate a low binding affinity and specificity for the nucleobase of the phosphate donor with the phosphate donor at the concentrations used not being able to change the conformation of the tetramer. Also, we cannot rule out a different binding site of the phosphate donor in *Lm*TK compared to other solved structures. While it is not possible to pinpoint the amino acid residues coordinating the adenosine in *Lm*TK and *Tb*TK, the structures of LmTK-AP_5dT and *Lm*TK-dTTP give insight into their mode of inhibition. Its dThd moiety

binds in the dThd binding site and the phosphates are bound by the P-loop. In this, dTTP's γ-phosphate takes the place of ATP's β-phosphate, blocking the site. The phosphates of AP_5dT have different positions, but still binding tightly to the P-loop.

3.3.3.3 Quaternary structure and subunit-subunit interactions

Most of the TK crystals contain one dimer in the asymmetric unit, with the tetramers generated by applying the symmetry operators (Figure 3-15). The crystals displaying lower symmetry have either one tetramer and a dimer (*Tb*TK-dTMP-PO_4^{2-}-b) or two tetramers (*Lm*TK-dTTP) in the asymmetric unit. SSM superposition reveals tetramer formation of *Lm*TK and *Tb*TK to be identical.

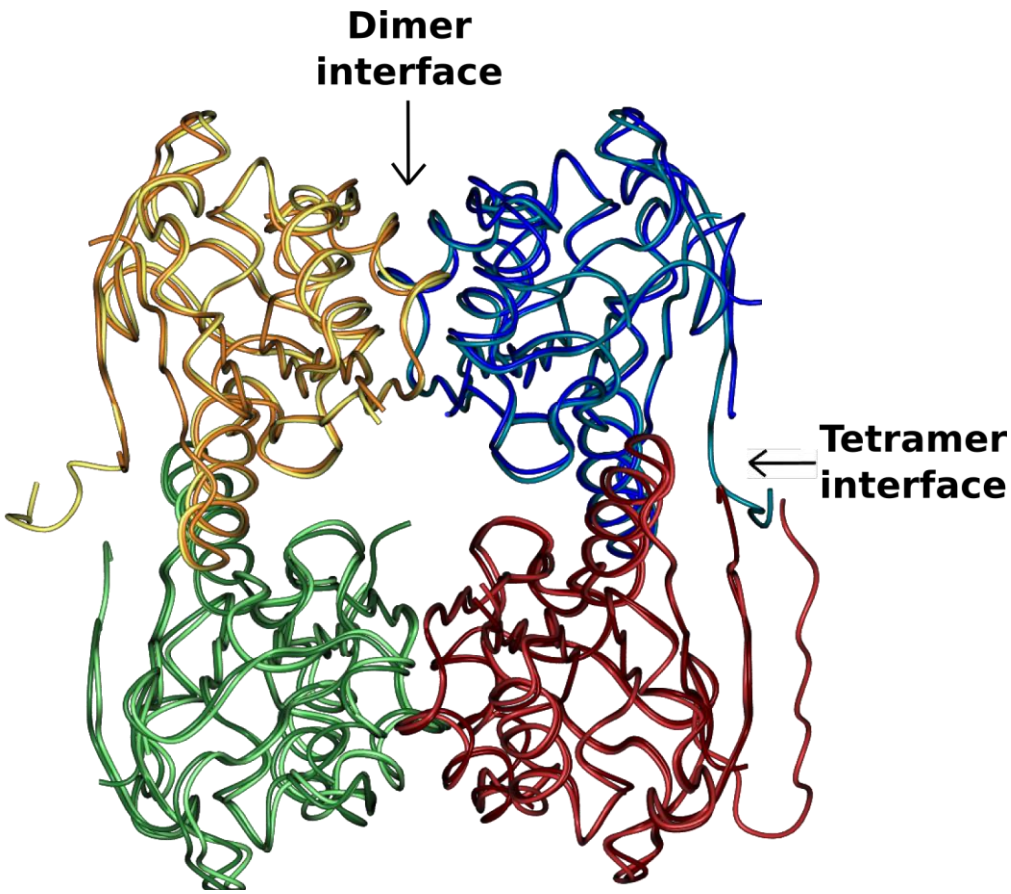


Figure 3-15: Superposition of *Lm*TK and *Tb*TK tetramers.

Protein chains are represented as worms, *Lm*TK-dTTP in lemon, dark cyan, light green and pink, *Tb*TK-dThd/dTMP in orange, dark blue, lawn green and tan.

Analysis using PISA (Krissinel & Henrick, 2007) suggested *Lm*TK to be tetrameric independent of the ligand bound or its presence during the PISA analysis. *Tb*TK was predicted to be a dimer or the tetramer in solution, depending on the presence or absence of phosphates in the P-loop during PISA analysis. If the phosphate was used in analysis, *Tb*TK was predicted to be tetrameric, if all ligands were ignored, *Tb*TK was suggested to be dimeric. SSM superposition shows *Lm*TK and *Tb*TK to form the same subunit-subunit interactions.

There are two protomer-protomer interfaces, one defined by the central β-sheets of two protomers, which come together in an anti-parallel orientation to form one continuous 12-stranded β-sheet. This results in the formation of a stable dimer. The other interface, resulting in tetramer formation, is defined by the anti-parallel interaction of the long α1-helices of the α/β-domains of two adjacent subunits.

In solution, TKs were reported to form tetramers at high concentrations and depending on the species, tetramer formation is induced by the presence of the phosphate donor (Mutahir *et al.*, 2013). In crystal structures two different types of TK tetramers are observed depending on the presence or absence of phosphate donor, the open and the closed tetramer (Figure 3-16). In the absence of ligand in the phosphate donor site, TKs form a closed tetramer with the long α1-helices of the α/β-domains interacting directly. In the presence of the phosphate donor the tetramers formed are more open and the α1-helices are ~ 3 Å apart, so there is space for the nucleobase to bind. The dimer interfaces remain identical in both tetramers. The formation of the open and the closed tetramer can be observed in TKs shown to be dimeric in solution, e.g. *Bc*TK (Kosinska *et al.*, 2007, Mutahir *et al.*, 2013), while TKs, that have been shown to form ATP-induced tetramers in solution, e.g. *Hs*TK1 (Welin *et al.*, 2004, Mutahir *et al.*, 2013), have only been crystallised as closed tetramers. The conformations of the dimer interfaces do not change upon phosphate donor binding.

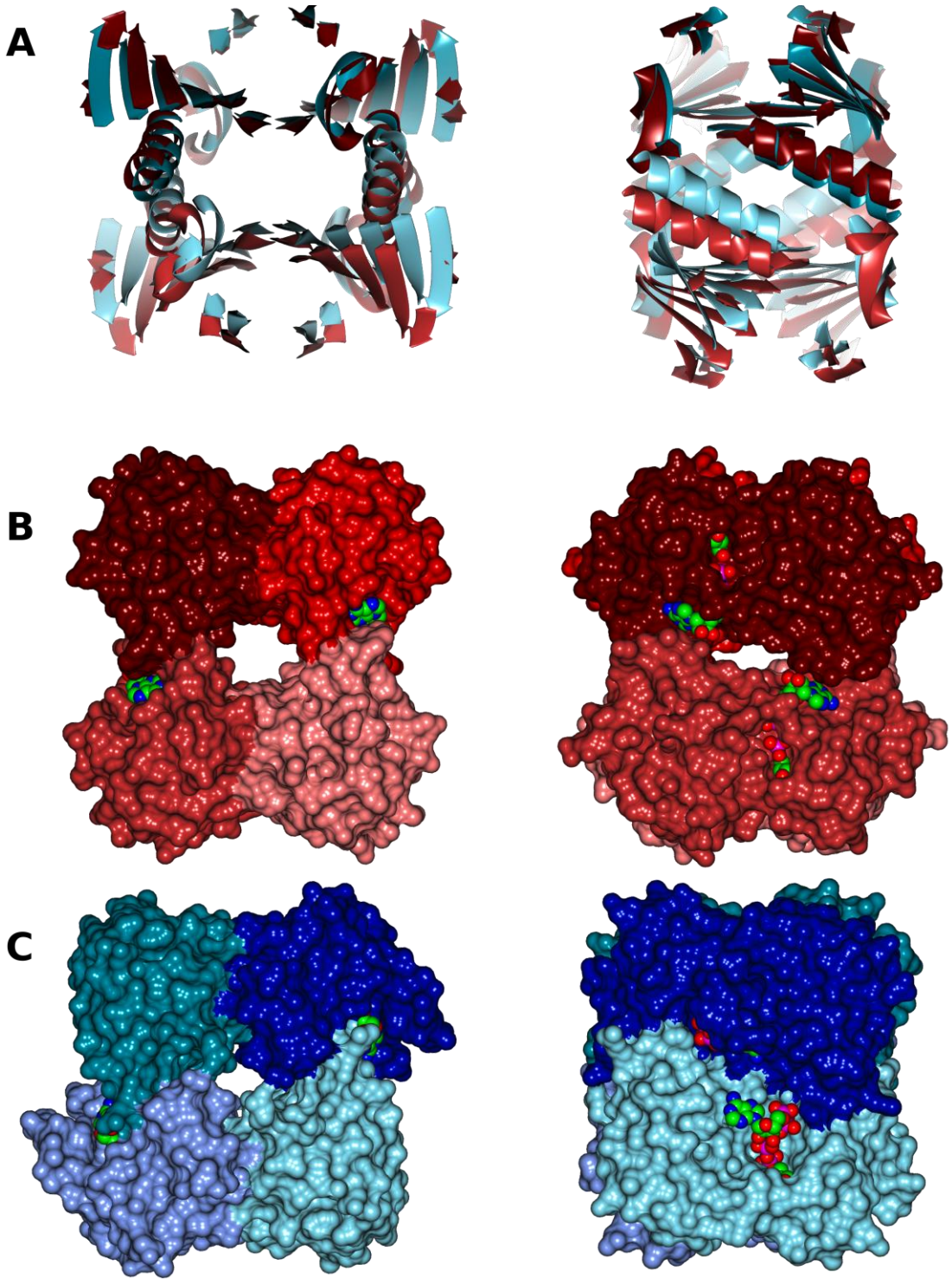


Figure 3-16: Comparison of open and closed TK tetramers.

(A) Closed *Lm*TK-AP$_5$dT (blue tones) and open *Tm*TK-dThd-AppNHp (2qq0, red tones) tetramers superposed on one dimer each (upper dimer). Shown are the β-sheets and α1 helices of the tetramers viewed from two angles 90° rotated. (B) Surface representation of the open *Tm*TK-dThd-AppNHp tetramer. (C) Surface representation of the closed *Lm*TK-AP$_5$dT tetramer.

*Lm*TK and *Tb*TK form closed tetramers in the crystals and do not show electron density for the AppNHp they were co-crystallised with. *Lm*TK co-crystallised with the bi-substrate analogue AP_5dT, is also seen as a closed tetramer in the crystal. This could be caused by formation of the closed tetramer in solution prior to addition of AP_5dT to the protein, blocking the phosphate donor site. The adenosine part of the AP_5dT apparently was not able to induce the change into the open tetramer, either a result of a low binding affinity of the phosphate donor nucleobase or indicating a different binding site compared to the one in *Ca*TK (PDB ID: 1xx6), *Tm*TK (PDB ID: 2qq0) and *Bc*TK (PDB ID: 2ja1).

3.3.3.4 Conformations of the mainly disordered loop region

The 15-20 amino acid residue loop region between amino acid residues 245 and 260 in *Tb*TK (42-59 in *Lm*TK) is disordered in most structures reported in the PDB. However, in some chains of the structures of *Uu*TK (2b8t, Kosinska *et al.*, 2005), *Bc*TK (2ja1, Kosinska *et al.*, 2007), *Hs*TK1 (1w4r, Birringer *et al.*, 2005) and *Ca*TK (1xx6, Kuzin *et al.* unpublished) this region is ordered with their conformations varying greatly (Figure 3-17).

This loop was proposed to be involved in the binding of the phosphate donor (Kosinska *et al.*, 2005, Kosinska *et al.*, 2007), although none of the structures solved so far directly supports this. In most chains of the *Lm*TK and *Tb*TK structures this loop is disordered. However, chain A of the *Lm*TK-AP_5dT structure shows a partial folding of this loop over the adenosine moiety, which is bound in a different position than the expected phosphate donor site, defined for other TK structures (*Ca*TK 1xx6, *Tm*TK 2qq0, *Bc*TK 2ja1). The β-hairpin formed by the loop region of *Bc*TK-dTTP (2ja1) is clashing with the site binding the adenosine moiety in the *Lm*TK-AP_5dT structure (Figure 3-17 B).

Additionally, in the *Lm*TK-dTTP structure, this region can be traced in two of the eight chains. However, this can be assumed to be a crystallographic artefact since the loop forms an additional β-strand to extend the central β-sheet, connecting it to one of the central β-sheets of a neighbouring, symmetry-related tetramer. Dimers

of these are shown in Figure 3-18. The formation of a continuing β-sheet between tetramers is not likely to be of physiological importance.

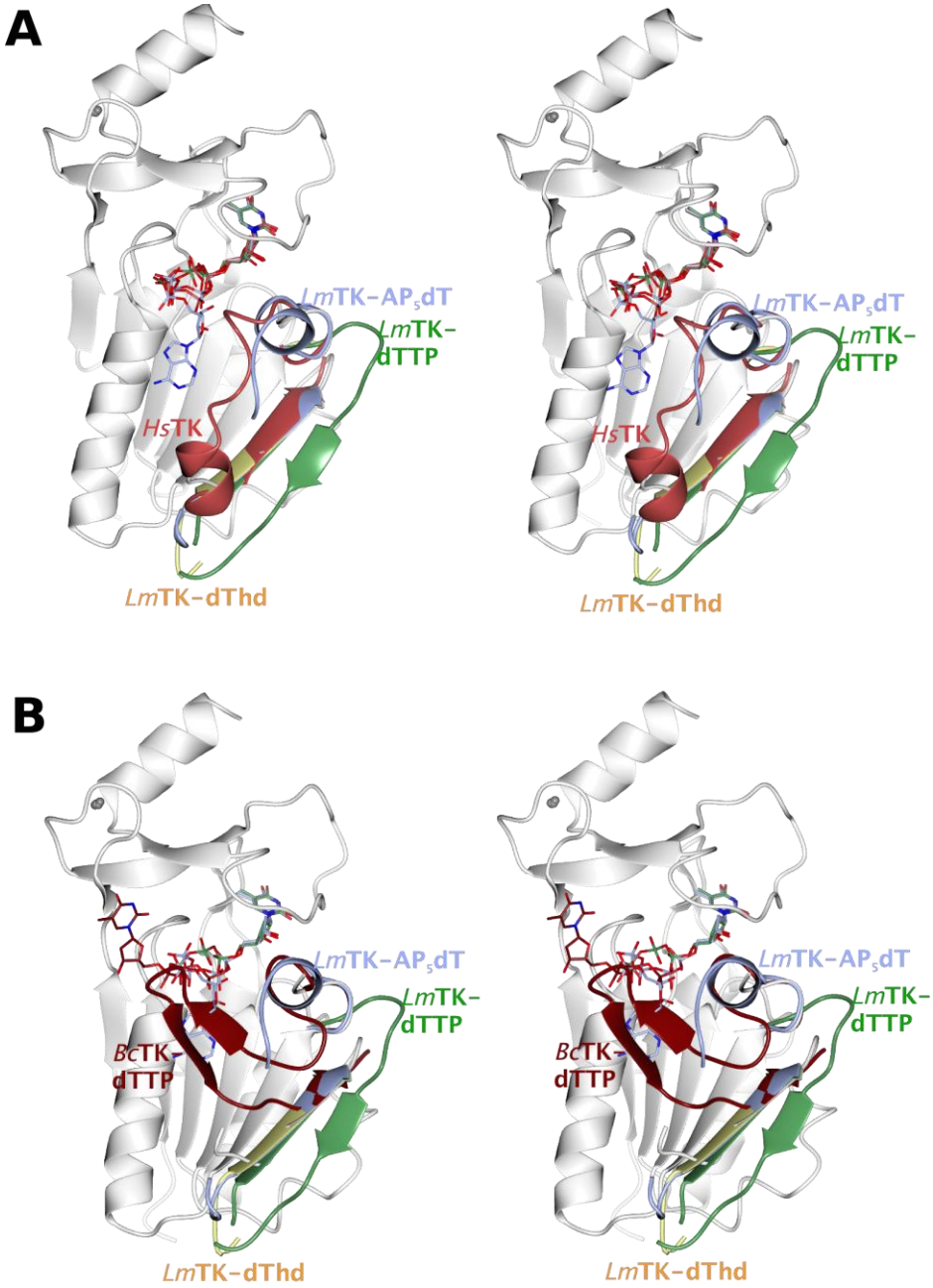


Figure 3-17: Conformations of the mainly disordered loop.

The peptide chains are shown in ribbon representation, the ligands as cylinders coloured by atom type with C and P atoms according to model. The overall structure shows the chain A of *Tb*TK in white with the relevant loop being completely disordered. Displayed are the loops of *Lm*TK-AP$_5$dT (ice blue, chain A), *Lm*TK-dTTP (lawn green, chain C), *Lm*TK-dThd (lemon, chain A) together with (A) *Hs*TK1 (pale crimson, PDB: 1w4r, chain A), and (B) *Bc*TK-dTTP (tan, PDB: 2ja1, chain A) after SSM superposition of the whole structures.

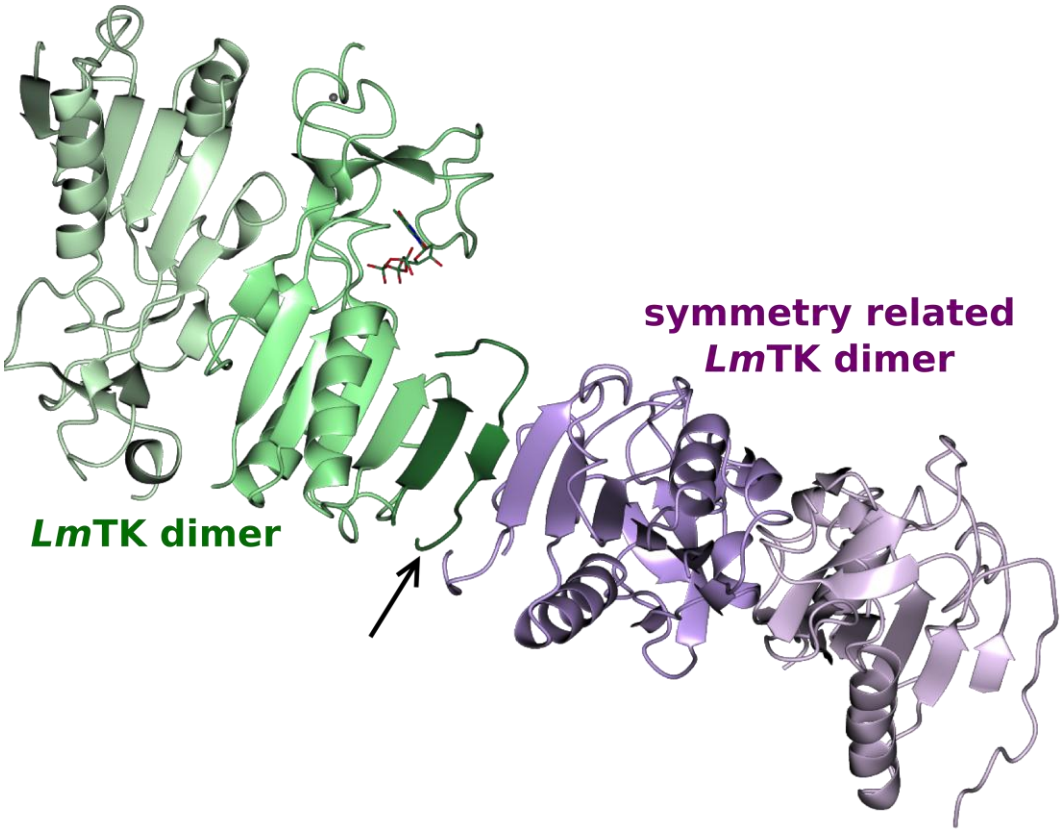


Figure 3-18: *Lm*TK-dTTP dimer connection to a symmetry-related dimer.
The black arrow indicates the β-strand formed in order to connect the two β-sheets. The dimers (half of the tetramers) shown are in adjacent symmetry-related tetramers (not in the same asymmetric unit).

3.3.3.5 The *Tb*TK pseudo-dimer

*Tb*TK is a tandem enzyme, containing two TK-homology domains fused to one another and it has been shown previously that the catalytically active domain does not dimerise efficiently in solution, when expressed as thioredoxin fusion in *E. coli* (Ranjbarian *et al.*, 2012). Full-length *Tb*TK forms a pseudo-dimer in solution and assembles into a dimer (pseudo-tetramer) at high concentrations and in the presence of ATP. For this assembly, the C-terminus of the N-terminal TK-homology domain needs to be close enough to the N-terminus of the other, so that the linker region of 18 amino acids (aa 181-197) is able to bridge this gap. To address this, the sequence of the N-terminal TK-homology domain was modelled onto the structure of the catalytically active, C-terminal TK-homology domain and coloured by sequence identity (Figure 3-19). As there are two different protomer-protomer

interfaces in a TK tetramer, one responsible for dimer formation and one for tetramer formation, there are two possible pseudo-dimers formed by full-length *Tb*TK. In both modelled pseudo-dimer structures (Figure 3-19 A and B), the termini of the two protomers seem indeed close enough to be connected by an 18 amino acid linker (shown as dashed grey line in the figure). After visualisation of the sequence identity in respect of the structure, high sequence identity of the central β-sheet as well as the α1-helix becomes apparent, while the outside facing parts of helices α2 and α3 and the region between the β-strands β5 and β6 show amino acid replacements. Looking at the dimer interface (Figure 3-19 C), only approximately 50 % of the residues involved are identical; the tetramer interface (Figure 3-19 D) on the other hand seems not to contain any amino acid replacements. Analysis with PISA indicated the pseudo-dimer shown in Figure 3-19 A is stable in solution, not so the pseudo-dimer in Figure 3-19 B. Therefore, the full-length *Tb*TK is most likely forming a pseudo-dimer similar to the one shown in Figure 3-19 A.

Comparison of the active site residues of structures of catalytically active TKs with our modelled inactive TK-homology domain highlights the amino acid substitution D286N. Asparagine is not able to act as a catalytic base, which renders the protein inactive. Additionally the transition state stabilising residue R246 is deleted and the substitution D244N might influence the activity. All remaining active site residues are identical in the thwo TK-homology domains.

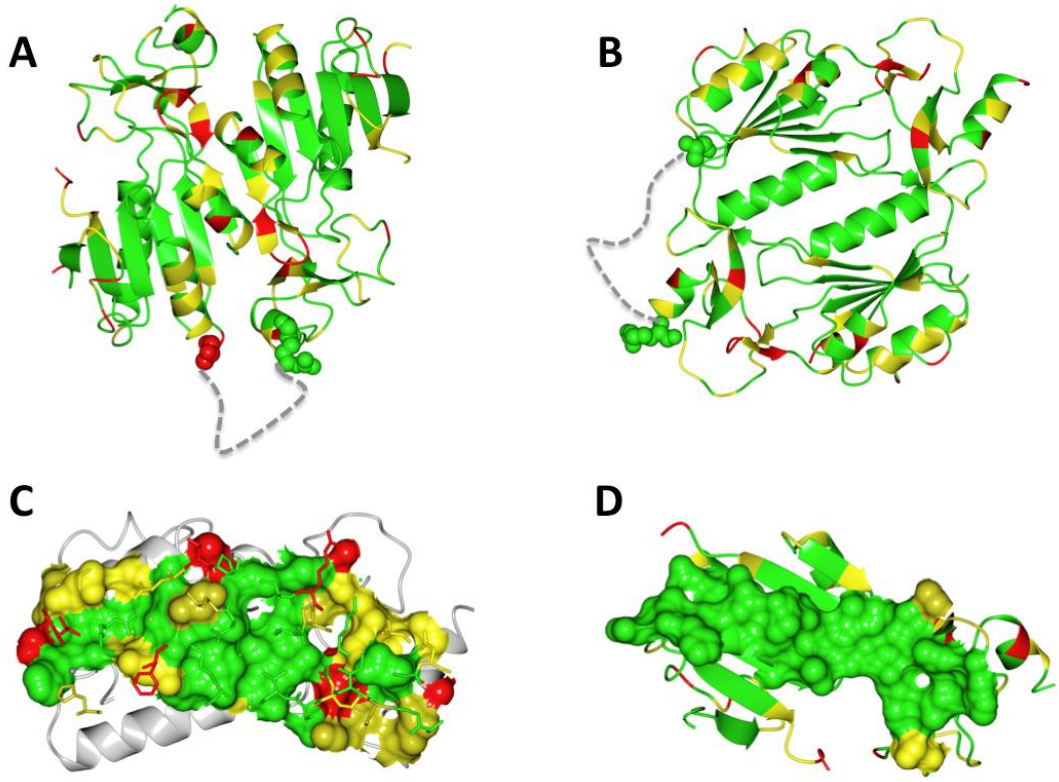


Figure 3-19: Model of *Tb*TK pseudo-dimer coloured by sequence identity.

Green = identical residues, yellow = similar residues, red = amino acid replacements with very different residues. (A) Ribbon representation of the *Tb*TK pseudo-dimer; first residue of chain A modelled TK-homology domain) and last residue of chain B (catalytically active domain, crystal structure from *Tb*TK-dThd-glycerol) are shown as spheres; the dashed line represents the linker between the two TK-homology domains of full-length *Tb*TK. (B) Alternative pseudo-dimer formed by the full-length *Tb*TK. (C) Surface view of the *Tb*TK dimer-interface coloured by sequence identity. (D) Surface view of the *Tb*TK tetramer-interface coloured by sequence identity.

3.4 Discussion

Expression and stabilisation of TKs of the TriTryps turned out to be difficult, which was surprising given that many structures of type II TKs have already been solved. It was possible to obtain soluble and stable protein of TKs from *L. major* and *T. brucei*, but all attempts to express *T. cruzi* TK failed. Somewhat disappointingly, only heavily truncated versions of the enzymes expressed in soluble form and structural work was limited to the core proteins closely homologous to other known structures in the PDB. However, analysis and comparison of the *Lm*TK and *Tb*TK structures with the *Hs*TK1 structure alongside kinetic data provided by collaborators (Table 3-), allow conclusions about the potential use of TKs as potential targets for anti-trypanosomal chemotherapeutics.

The structures of *Lm*TK and *Tb*TK are highly similar to those of other species (*Hs*TK1, (Welin *et al.*, 2004, Birringer *et al.*, 2005); *Uu*TK, (Kosinska *et al.*, 2005); *Ba*TK and *Bc*TK, (Kosinska *et al.*, 2007); *Tm*TK, (Segura-Pena, Lichter*, et al.*, 2007, Segura-Pena, Lutz*, et al.*, 2007); *Ca*TK, Kuzin *et al.* unpublished). The protomers overlay nicely and dimers and tetramers retain the same interfaces and overall architecture. The *Lm*TK and *Tb*TK structures form closed tetramers with the blocked phosphate donor site and therefore overlay better with other closed tetramers than with open tetramers.

While the active site residues whose side chains coordinate the dThd are highly conserved, comparison of the actual thymidine-binding pockets of *Tb*TK with the *Hs*TK1 shows differences. *Tb*TK has more space around the bound dThd, especially around the C5 position of the dThd ring. The thymidine-binding site has a widened cavity beside the dThd ring into the α/β-domain, where the main chain is bent outwards and the stacking F321 is ~ 1.1 Å further away from the dThd than in *Hs*TK1 or *Lm*TK (Figure 3-20 A). The region around the 3'-oxygen of the ribose ring seems to be more open in most *Lm*TK and *Tb*TK structures (Figure 3-20 B) compared to *Hs*TK1. This might be an artefact of the crystallisation as the loop, harbouring two amino acid residues, which coordinate this region of the dThd, is disordered in all *Lm*TK and *Tb*TK structures except for the *Lm*TK-AP_5dT, where this loop is partially ordered in chain A. In this structure the side chain conformations of these residues are identical to those reported for *Hs*TK1. The alignment in Figure 3-3 indicated

these residues (D43 and R45 in *Lm*TK) to be conserved in all TKs except the inactive TK-homology domain of *Tb*TK.

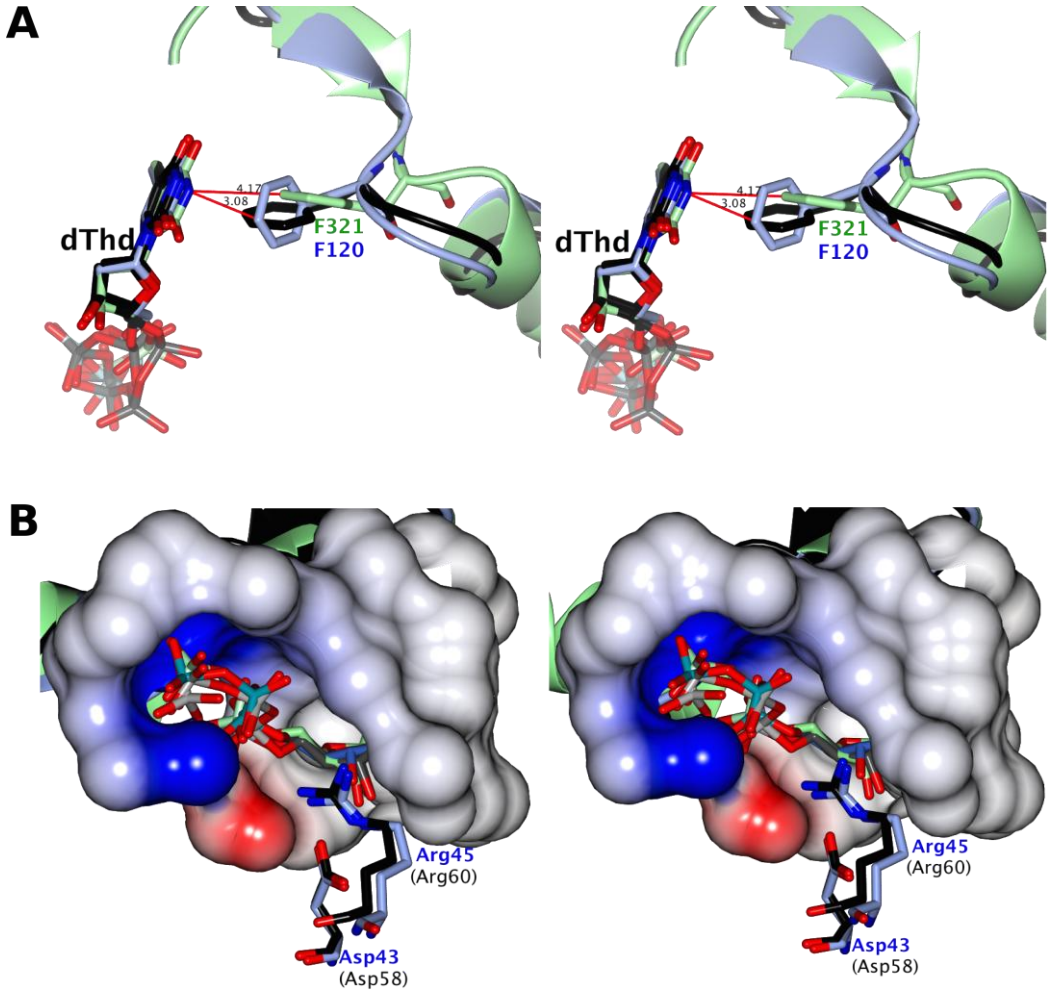


Figure 3-20: Comparison of the active sites of *Lm*TK, *Tb*TK and *Hs*TK1.

(A) Stereo view of the substrate environment focusing on F321 in *Tb*TK and F120 in *Lm*TK. Phenylalanines and dThd in cylinder representations, the helix of *Tb*TK (light green) is extended and pushed outwards compared to *Lm*TK (ice blue) and *Hs*TK1 (1W4R, black). (B) Stereo view of the surface of *Lm*TK active site focusing on the region around the 3'-oxygen of the ribose ring, D43 and R45 of *Lm*TK in ice blue (D58 and R60 in *Hs*TK1, black) cover the ribose region of the active site, while these residues are disordered in other *Lm*TK and *Tb*TK structures.

The structural differences in the active site of *Tb*TK indicate lower substrate specificity compared to the human enzyme. Lower substrate specificity was demonstrated previously (Ranjbarian *et al.*, 2012), showing *Tb*TK to readily phosphorylate deoxyinosine (dIno) and deoxyguanosine (dG), while the activity of

*Hs*TK1 against these substrates was shown to be poorer. This leads to the possibility of developing nucleoside analogues and other ligands that selectively bind to *Tb*TK either as alternative substrates or inhibitors. During the course of this work, collaborators at the Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain, showed TK being essential for the survival and proliferation of *T. brucei* parasite, which confirms this enzyme as a potential drug target in these parasites.

The structure of *Lm*TK on the other hand does not show major differences in the active site, so that design of *Lm*TK-specific ligands might be difficult. However, during the course of this work, collaborators at the Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain, studied the substrate specificity and inhibition kinetics of *Lm*TK (Table 3-11). They found *Lm*TK to have as high substrate specificity as *Hs*TK1, only phosphorylating dThd, dUrd and to some extent uridine (Urd). Additionally, they were able to show the phosphorylation of 5'-modified dUrd, as well as AZT, which was shown to be toxic to *L. amazonenses* (Araujo *et al.*, 2011) and therefore is a potential candidate for anti-leishmanial therapy.

Table 3-11: Kinetic parameters of full-length *Lm*TK.

Parameters are averages of two or more independent experiments. N.D. = not determined. [a] activity measurements at 10 mM substrate concentration. Data provided by collaborators at the Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain.

Substrate	**K_M (µM)**	**V_{max} (µmol·min^{-1}·mg^{-1})**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($s^{-1}.M^{-1}$)**
dThd	1.10 ± 0.33	4.66 ± 0.63	2.62	2.38×10^6
dUrd	11.1 ± 2.7	5.45 ± 0.77	3.06	2.76×10^5
dC	N.D.	0.07[a]	-	N.D.
dA	N.D.	0.01[a]	-	N.D.
dG	N.D.	0.04[a]	-	N.D.
Urd	1462 ± 158	0.79 ± 0.12	0.44	3.03×10^2
Cyd	N.D.	0.00[a]	-	N.D.
Ado	N.D.	0.00[a]	-	N.D.
Guo	N.D.	0.00[a]	-	N.D.
dIno	N.D.	0.04[a]	-	N.D.
AZT	3.04 ± 0.76	3.52 ± 0.56	1.97	6.51×10^5
5F-2'dU	5.72 ± 0.56	8.29 ± 1.12	4.66	8.14×10^5
5Br-2'dU	2.05 ± 0.33	6.17 ± 0.19	3.47	1.69×10^6

3.5 Summary and future work

The C-terminally truncated versions of *Lm*TK, spanning amino acid residues 2-185, and *Tb*TK, spanning amino acid residues 198-382, were soluble and produced in sufficient amounts for structural studies when expressed using the baculovirus expression system. Therefore, the structural work was restricted to the core proteins with high homology to other TK structures in the PDB. Comparison of *Lm*TK and *Tb*TK with *Hs*TK1 allows conclusions to be drawn about the use of TKs as potential targets against *Leishmania spp.* and *T. brucei*. All *Lm*TK and *Tb*TK structures show the formation of closed tetramers as seen in many other TK structures in the PDB. The dThd binding site is well resolved in all of the structures and allows analysis of the substrate site. *Lm*TK's dThd site is very similar to the *Hs*TK1, but *Tb*TK shows a slightly widened cavity. This, together with its essentiality in the parasite (unpublished data) and the lower substrate specificity demonstrated previously (Ranjbarian *et al.*, 2012), supports *Tb*TK's potential as a drug target. In future studies, the *Tb*TK structures will aid identification and design of lead compounds to treat HAT.

*Lm*TK phosphorylates AZT, a compound shown to inhibit the growth of *Leishmania* amastigote growth (Araujo *et al.*, 2011). In this context, the structures of *Lm*TK will allow future studies to focus on the structural analysis of ligand binding and phosphorylation in respect to possible modifications of AZT to maximise its potential as an anti-leishmanial chemotherapeutic.

Chapter 4 Characterisation of the M17 family leucyl aminopeptidases from the TriTryps

4.1 Introduction

4.1.1 M17 family of leucyl aminopeptidases

This family is subdivided into eukaryotic leucyl aminopeptidases (LAP, EC 3.4.11.1), bacterial leucyl aminopeptidases (EC 3.4.11.10), prolyl aminopeptidases (PAP, EC3.4.11.5) and PepB aminopeptidases (PepB, EC 3.4.11.23).

LAPs are well studied and in the literature are referred to under several alternative names, e.g. cytosol aminopeptidase, leucine aminopeptidase and peptidase S. They are found in all kingdoms of life and species can have more than one M17 LAP. The eukaryotic LAPs have been characterised biochemically, mainly the porcine and bovine lens enzymes (Allen *et al.*, 1983, Carpenter & Vahl, 1973, Cappiello *et al.*, 2006, Matsushima *et al.*, 1991, Strater & Lipscomb, 1995b, a, Strater, Sun*, et al.*, 1999, Van Wart & Lin, 1981) and those of a number of plant species, (Gu & Walling, 2002, Scranton *et al.*, 2012). Protozoan LAPs have been studied in *Babesia gibsoni* (*Bg*LAP, Jia *et al.*, 2009), *Leishmania spp.* (Morty & Morehead, 2002), *Plasmodium falciparum* (*Pf*LAP, Harbut *et al.*, 2011), *Toxoplasma gondii* (*Tg*LAP, Jia *et al.*, 2010) and *Trypanosoma cruzi* (Cadavid-Restrepo *et al.*, 2011). Bacterial enzymes were studied from *E. coli* (Stirling *et al.*, 1989, Bhosale *et al.*, 2010), *Helicobacter pylori* (Dong *et al.*, 2005), *Pseudomonas putida* (Kale *et al.*, 2010), *P. aeruginosa* (Woolwine *et al.*, 2001), *Vibrio cholerae* (Behari *et al.*, 2001) and *Xanthomonas oryzae* (Kim *et al.*, 2013).

The LAP structures of many species have been solved, e.g. bovine lens (blLAP, Burley *et al.*, 1990, Kim & Lipscomb, 1993, Kim *et al.*, 1993, Strater & Lipscomb, 1995a, b), tomato (tLAP, DuPrez *et al.* unpublished), *Caenorhabditis elegans* (*Ce*LAP, Patskovsky *et al.* unpublished) *Coxiella burnetii* (*Cb*LAP, Minasov *et al*. unpublished), *E. coli* (*Ec*LAP, PepA, Strater, Sherratt*, et al.*, 1999), *Francisella tularensis* (*Ft*LAP, Maltseva *et al.* unpublished), *P. falciparum* (*Pf*LAP, Figure 4-1, McGowan *et al.*,

2010), *P. putida* (*Pp*LAP, Kale *et al.*, 2010), *Staphylococcus aureus* (*Sa*LAP, Hattne *et al.* unpublished), *T. brucei* (*Tb*LAP, Wernimont *et al.* unpublished) and *X. oryzae* (*Xo*LAP, Nataranjan *et al*. unpublished). All of these structures are homologous with overall RMSDs of 0.27 – 0.35 Å, while sequence identity towards the TriTryp LAPs is 32-45 %. LAPs are bi-lobal proteins (Figure 4-1 A) with a highly conserved C-terminal domain, containing the active site, and a smaller N-terminal domain with lower sequence similarity between species. The C-terminal domain shows some structural homology to family M14 carboxypeptidases and the M20 and M28 families of metallopeptidases (Artymiuk *et al.*, 1992), but the position of the metal binding sites between these families is not conserved. While the fold of the C-terminal domain is absolutely identical for all LAPs, the N-terminal domain shows differences, the secondary structure elements being conserved but their positioning towards each other differing among species. The overall LAP fold is represented as cartoon in Figure 4-1 B. All LAPs described so far form trimers (Figure 4-1 C) and hexamers (as dimers of trimers, Figure 4-1 D) in solution with the hexameric state being predominant.

LAP activity is dependent on the presence of two co-catalytic zinc ions, but activity has been shown with a number of divalent metal ions, e.g. manganese and magnesium (Carpenter & Vahl, 1973). LAPs hydrolyse peptides and amino acyl substrates with N-terminal leucine efficiently, but the substrate specificity is broad and varies between species and between different LAPs within one species. These enzymes generally accept substrates with most amino acids as terminal residue and the penultimate residue plays a role in the catalytic efficiency of the enzyme (Gu & Walling, 2002).

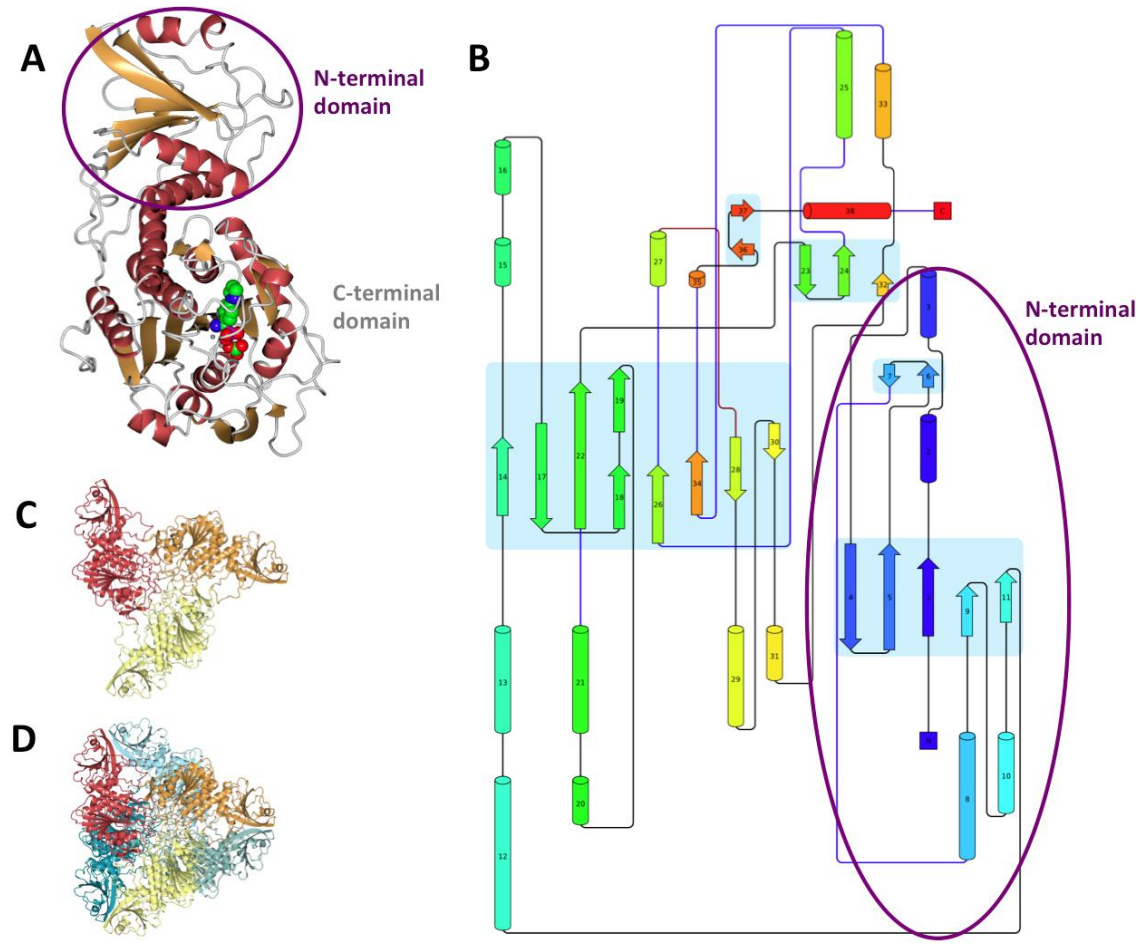


Figure 4-1: Structure of *Pf*LAP.

(A) Protomer; (B) Cartoon representation of *Pf*LAP secondary structure elements coloured with a colour gradient from N-terminus (blue) to C-terminus (red) with β-sheets as arrows and α-heleces as tubes. (C) Half a hexamer (trimer); (D) Hexamer (dimer of trimers). For (A), (C) and (D): Protein in ribbon representation, ligand indicating the position of the active site as spheres. This and all subsequent structure figures were made with CCP4*mg* (McNicholas *et al.*, 2011). The protein cartoon was generated using Pro-Origami (Stivala *et al.*, 2011).

PAPs were shown to be manganese-dependent aminopeptidases (Sarid *et al.*, 1962, Nordwig & Mayer, 1973). Later studies showed the mammalian enzyme not to be restricted to prolyl bonds and functional characterisation suggested the PAPs are identical to LAPs (Turzynski & Mentlein, 1990). PepB, characterised in *E. coli*, is a zinc-dependent aminopeptidase cleaving preferentially acidic amino acid residues (aspartic and glutamic acid) off the peptide substrate (Mathew *et al.*, 2000). It also accepts other amino acids, e.g. leucine, methionine, histidine cysteine and

glutamine. Its pH optimum is alkaline and the activity is stimulated by potassium ions. Like LAPs it forms homohexamers.

4.1.2 M17 LAPs from the TriTryps

Four distinct M17 family LAPs have been identified on the TriTryps genomes (Table 4-1), classified according to their isoelectric points as acidic (LAP-A), neutral (LAP-N) or basic (LAP-B).

Table 4-1: The TriTryp LAPs.

The subcellular localisations were predicted with PSORT II, accessed on 22.6.2014 (Nakai & Horton, 1999), except those marked with an asterisk (*), which were determined experimentally; ER = endoplasmic reticulum.

Name	Gene ID (in TriTrypDB)	Number of amino acids	MW [Da]	Isoelectric point	Subcellular localisation
*Lm*LAP-A	LmjF.11.0630	538	57 148	5.98	Cytoplasmic
*Lm*LAP-A2	LmjF.11.0620	292	31 229	5.91	Cytoplasmic
*Lm*LAP-N	LmjF.33.2570	528	55 757	6.67	ER
*Lm*LAP-B	LmjF.23.0950	565	60 267	9.21	Cytoplasmic* (Mitochondrial)
*Tb*LAP-A	Tb427tmp.02.4440	521	55 400	6.04	Cytoplasmic
*Tb*LAP-A2	Tb11.v5.0360	216	23 376	5.67	Extracellular
*Tb*LAP-N	Tb427tmp.02.0070	565	60 625	7.51	ER
*Tb*LAP-B	Tb427.08.3060	671	71 455	9.33	Mitochondrial
*Tc*LAP-A	TcCLB.508799.240	520	55 892	6.06	Cytoplasmic
*Tc*LAP-A2	TCSILVIO_001670	217	23 512	5.44	Extracellular
*Tc*LAP-N	TCSYLVIO_005560	557	59 915	6.98	ER
*Tc*LAP-B	TCSILVIO_000171	584	62 309	9.19	ER

Structure-based and sequence based multiple alignments clearly indicate a large divergence between the three distinct types, Figure 4-2, suggesting all three were present in a common ancestor of the TriTryps. The LAPs of most other species, including bacteria, plants and protozoa of other genera, are closest related to the B-type LAPs with sequence identities between 34 and 45 %. The LAP with the highest identity to the TriTryp LAP-As and LAP-Ns is the human NPEPL1 with an overall

sequence identity of 37 % to *Lm*LAP-A, 41 % to *Tb*LAP-A and 51 % to *Tc*LAP-A, while all other enzymes analysed showed lower identities between 28 % and 38 %.

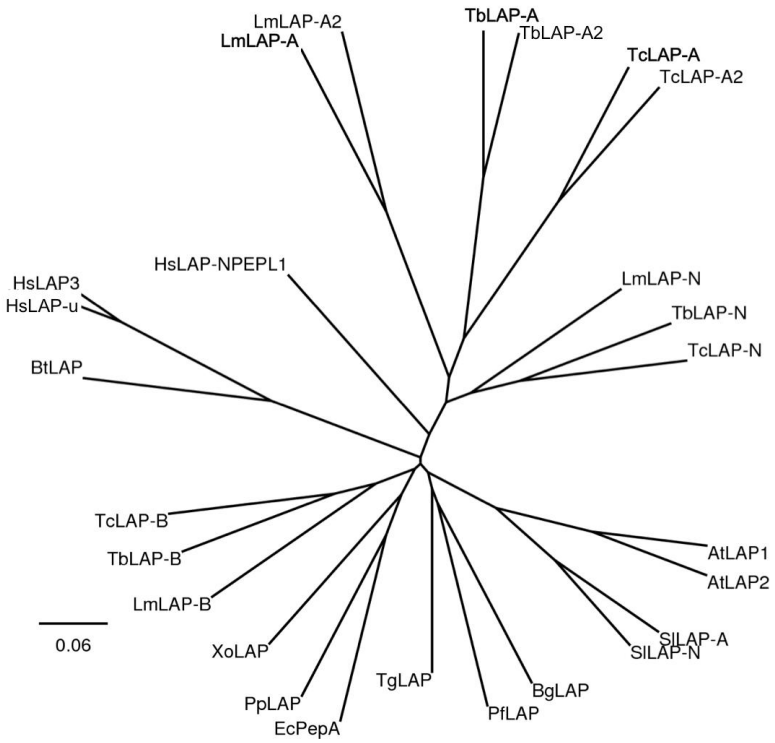


Figure 4-2: Phylogenetic tree of the M17 LAPs.

***Hs*LAP-NPEPL1 = human LAP called NPEPL1, *Hs*LAP3 = human LAP3, *Hs*LAP-u = human LAP, still classified as unknown product, *Bt*LAP = bovine LAP, *Xo*LAP = *X. oryzae*, *Pp*LAP = *P. putida* LAP, *Ec*PepA = *E. coli* LAP, *Tg*LAP = *T. gondii* LAP, *Pf*LAP = *P. falciparum*, *Bg*LAP = *B. gibsonii* LAP, *Sl*LAP = tomato LAP, AtLAP = *A. thaliana* LAP. Radial tree was drawn using TreeDyn 198.3 (Chevenet *et al.*, 2006) after structure-based multiple sequence alignment using Expresso in T-coffee (Armougom *et al.*, 2006).**

Alignment of the distinct TriTryp LAP types, e.g. all three LAP-As (Figure 4-3), shows a very high sequence similarity.

Sequence alignment of the four LAPs (A, A2, N and B) from *T. brucei*, *T. cruzi* and *L. major* (Figure 4-4) shows low sequence identity, but common core secondary structure elements, interspaced with longer loop regions mainly for the LAP-Bs. It can be assumed that the core structures of the three long LAP types, LAP-A, LAP-N and LAP-B, are highly similar to one another and only differ in the length, conformation and charge of surface loops. The LAP signature sequence and active

site residues are conserved among all LAPs with the exception of the LAP-A2s, which are missing three of them. This work focuses on the LAP-As. This, together with analysis of the genomic surrounding of the *lap*-A2 genes, indicates the divergence of the LAP-As and LAP-A2s to be either an error in the annotation or alternatively non-functional gene duplication.

Only two of the TriTryp LAPs have been characterised so far: LAP-B from *L. major*, *L. donovani* and *L. amazonensis* (Morty & Morehead, 2002) and LAP-A from *T. cruzi* (Cadavid-Restrepo *et al.*, 2011).

The *Lm*LAP-Bs were shown to have very restricted substrate specificity, hydrolysing only substrates with leucine, cysteine and methionine residues as N-terminal amino acid efficiently. The peptidase activity was dependent on the metal ion with the catalytic efficiencies in the order $Mn^{2+} >> Mg^{2+} > Co^{2+} > Ni^{2+} > Zn^{2+}$ and inhibition by Ca^{2+}, Fe^{2+} and Cu^{2+}. However, the physiological metal was determined to be zinc (by inductively coupled plasma atomic emission spectroscopy (ICP-AES)) (Morty & Morehead, 2002). Inhibition of the aminopeptidase function could be obtained by metal chelators, bestatin, apstatin, actinonin, L-leucinal and anti-porcine LAP IgG. Although analysis with PSORT II indicated mitochondrial localisation, Morty *et al.* suggested its localisation in the cytoplasm after immunocytochemical localisation experiments using anti porcine LAP IgG.

*Tc*LAP-A, like *Lm*LAP-B, was shown to have limited substrate specificity with amidolytic activity only detected for amino coumarin substrates with N-terminal leucine (Leu-AMC) and none with N-terminal proline, aspartic acid, glycine-phenylalanine, glycine-arginine, or glycine-proline and also did not accept bovine serum albumin, immunoglobulin G or gelatine as substrates. The enzyme showed a clear metal preference for $Mn^{2+} > Ca^{2+} > Zn^{2+}$ with considerably less activity in the presence of Fe^{2+} and Mg^{2+} and almost none with Co^{2+} or Al^{3+}. Inhibition could be obtained with metal chelators, bestatin and 1,10-phenanthroline. Interestingly, Cadavid-Restrepo *et al*. found distinct activity profiles for natively expressed and recombinant *Tc*LAP-A with differences in temperature stability and activity optima. It was shown to be mainly hexameric in solution like other M17 LAPs. *Tc*LAP-A is

localised in cytosolic foci or vesicles within the cytosol in *T. cruzi* amastigote, epimastigote and trypomastigote forms.

Neither *Lm*LAP-B, nor *Tc*LAP-A has been validated as a potential drug target.

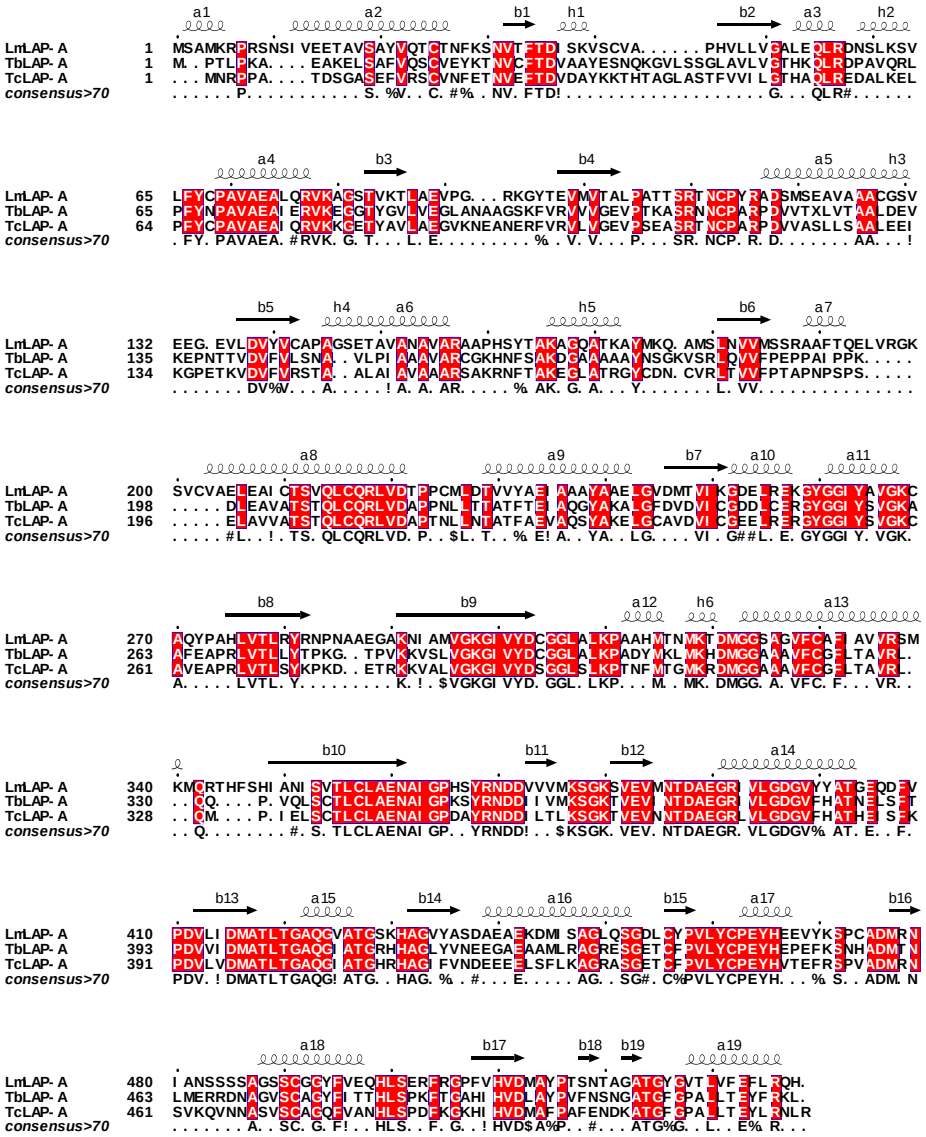


Figure 4-3: Alignment of the TriTryp LAP-As.

Consensus sequence (>70 %) is shown below the sequences, secondary structure of *Lm*LAP-A above the sequences. Figure made with Espript 3.0 (Gouet *et al.*, 2003) after sequence alignment in M-coffee (Notredame *et al.*, 2000).

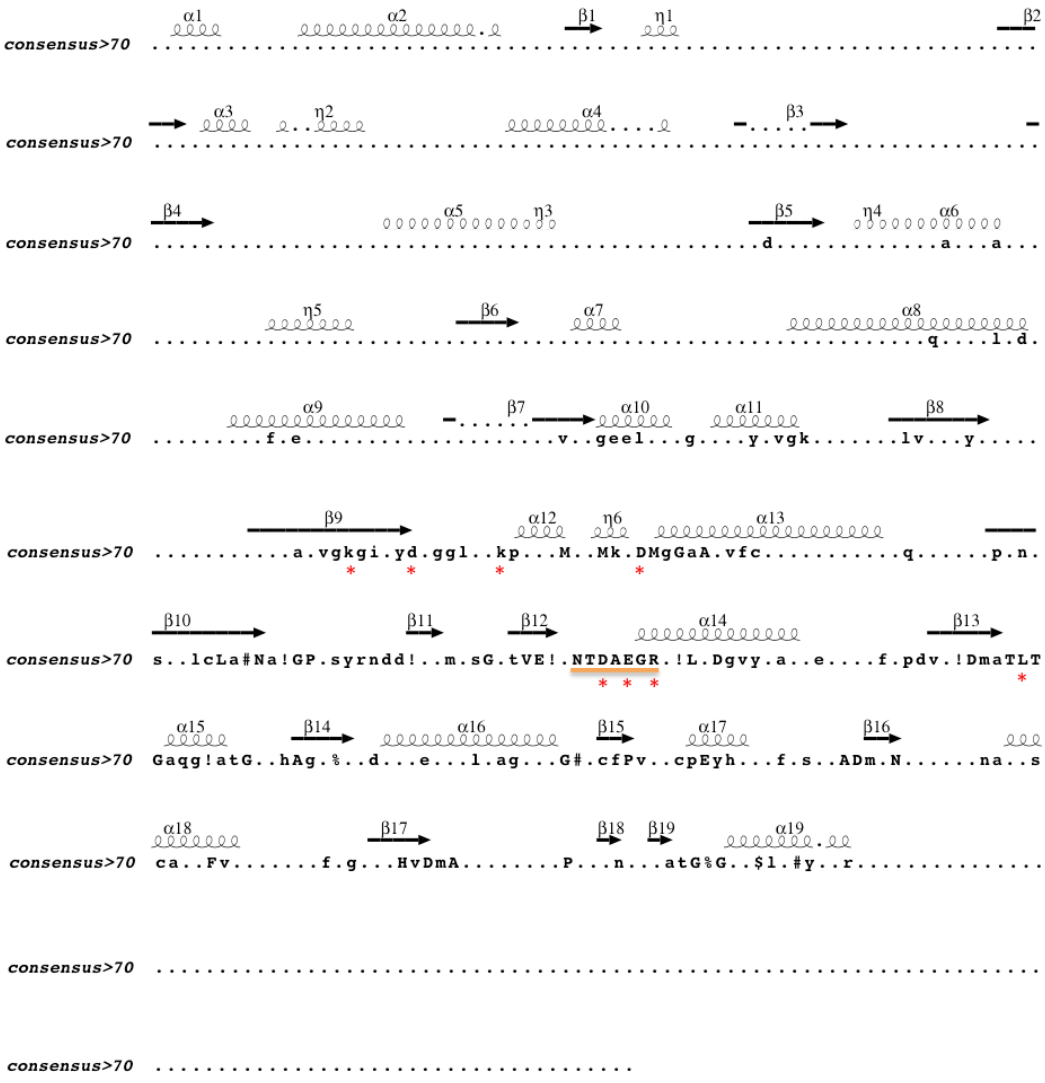


Figure 4-4: Consensus sequence (> 70 %) of the TriTryps LAPs.

LAP signature sequence is underlined in orange, red asterisks indicate active site residues. Conserved residues are written as capital letters, partially conserved residues in small letters, and secondary structure of *Lm*LAP-A above the sequences. Figure made with Espript 3.0 (Gouet *et al.*, 2003) after sequence alignment in M-coffee (Notredame *et al.*, 2000).

4.1.3 Mechanism of LAPs

The amidolytic mechanism of M17 LAPs was characterised using structural information gained from the bovine lens LAP (Strater, Sun*, et al.*, 1999) and led to the now commonly accepted model for catalysis, Figure 4-5.

In brief, the active site contains two divalent metal ions in two non-equivalent binding sites. One site is reported to have a higher affinity for the metal as it is buried deeper within the protein than the other, which is closer to the surface. The physiological metal is believed to be zinc, but it can be exchanged *in vitro* for some other divalent ions preserving the activity (Kim & Lipscomb, 1993, Jia *et al.*, 2010, Morty & Morehead, 2002, Cadavid-Restrepo *et al.*, 2011). Depending on the species, LAPs seem to have different metal preferences. Besides the two zinc ions the active site contains a HCO_3^- in most but not all structures. The two metal ions coordinate the substrate and are involved in stabilisation of the *gem*-diolate intermediate (middle panel in Figure 4-5) by polarisation of the substrate carbonyl group. The HCO_3^- coordinated by R375 (in *Tb*LAP-A) functions as a general base in the catalysis by accepting a proton from the metal-bridging water (the catalytic nucleophile). Then the HCO_3^- transfers a proton to the leaving amino group of the peptide, which results in breakage of the *gem*-diolate intermediate. The residues involved in metal coordination and HCO_3^- binding are conserved in all LAPs.

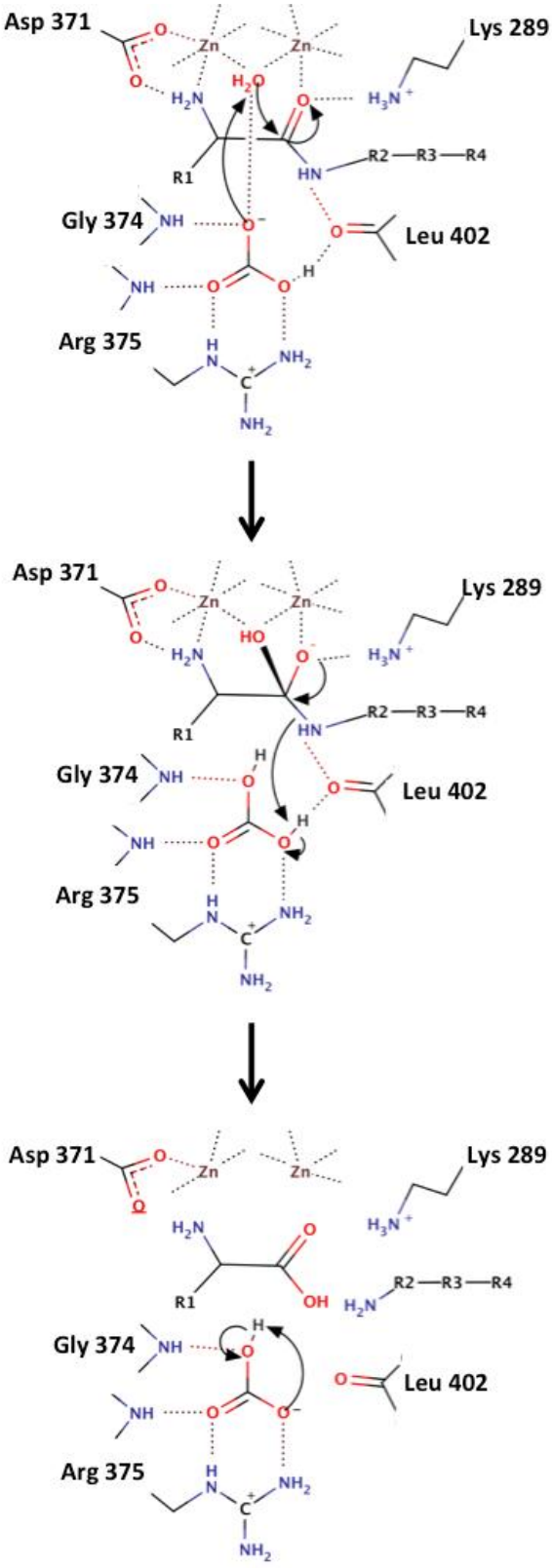


Figure 4-5: LAP mechanism.

Amino acid residues numbered after *Tb*LAP-A. Figure made with MarvinSketch (ChemAxon).

4.1.4 Physiological Functions of LAPs

The physiological functions proposed for LAPs are diverse and differ between species. In animals, LAPs were implicated to play a role in proteolytic degradation and protein turnover based on the accumulation of peptides in rat liver cells after treatment with the aminopeptidase inhibitor bestatin (Botbol & Scornik, 1991). Whether this peptide accumulation was caused by inhibition of LAP or another bestatin-sensitive aminopeptidase is not clear. Another study (Cappiello *et al.*, 2006) proposed a role in maintenance of the cell redox status based on the observation that rat liver and bovine lens LAPs are able to cleave Cys-Gly bonds, which are components of glutathione, the cellular response against oxidative stress. Although *E. coli* and porcine LAP also utilise this substrate, plant LAPs show limited activity against it (Gu & Walling, 2002), arguing against this role in plants. Chaerkady and Sharma (Chaerkady & Sharma, 2004) suggested LAP plays a role in the formation of cataracts, due to the correlation of accumulation of partially degraded proteins and subsequent formation of opaque protein aggregates with the decline of LAP activity during lens maturation and aging. Mammalian LAPs are involved in MHC-1 antigen presentation by the trimming of the N-terminus of the peptides to be presented (Beninga *et al.*, 1998, Stoltze *et al.*, 2000, Kloetzel & Ossendorp, 2004), but LAP is not essential for antigen presentation *in vivo* (Towne *et al.*, 2005).

Plants encode a varying number of LAPs, which are classified by their isoelectric points with LAP-As having acidic, LAP-Ns neutral and LAP-Bs basic pIs. Some plant LAPs are expressed as 65 kDa pre-proteins and subsequently processed to their mature 55 kDa form. LAP-As are transcriptionally regulated (Chao *et al.*, 1999) responding to biotic and abiotic stresses (Gu & Walling, 2002). LAPs in tomato (*Sl*LAP) and *A. thaliana* (*At*LAP) are bifunctional, showing aminopeptidase activity and in addition acting as molecular chaperones (Scranton *et al.*, 2012). These two functions are independent from each other with active site mutations eliminating peptidase activity having no adverse effect on the ATP-independent chaperone activity. Also, disruption of LAP's hexamer formation (by introduction of point mutations) increased the chaperone activity.

Bacterial LAPs have rather broad substrate specificities and were proposed to be housekeeping enzymes facilitating amino acid recycling within the cell. The N-terminal domain of *E. coli* LAP, called aminopeptidase A, *Ec*PepA or *Ec*XerB, was shown to have an additional, completely unrelated function. *Ec*PepA (XerB) aids site-specific recombination events between directly repeated *cer* sites on ColE1 plasmids and *psi* sites in pSC101 (Gourlay & Colloms, 2004) by binding to specific sites and bending the DNA to allow other proteins to interact. In addition, *Ec*PepA was shown to be involved in c*re*-Lox recombination, which maintains the lysogenic state of the bacteriophage P1 (Paul & Summers, 2004). Another proposed function of *Ec*PepA is transcriptional repression of the *carAB* operon, encoding an essential protein in the pyrimidine biosynthesis pathway (reviewed in Gourlay & Colloms, 2004), and is working as a repressor of its own transcription. *V. cholera* PepA was shown to be involved in expression regulation of the toxT-dependent virulence genes (Behari *et al.*, 2001, Woolwine *et al.*, 2001), repressing them under non-inducing pH conditions. The *P. aeruginosa* LAP, called PhpA, causes transcriptional repression of *algD*, which encodes for the first enzyme in the alginate biosynthesis pathway (Woolwine *et al.*, 2001). Unlike EcPepA, PhpA's aminopeptidase function and transcriptional repression are coupled, so that mutations abolishing the aminopeptidase function hamper the transcriptional repression.

In summary, the physiological functions of M17 family LAPs are extremely diverse and vary with the species. This makes it very difficult, if not impossible, to predict the function in other species not characterised so far, e.g. kinetoplastid parasites.

4.1.5 LAPs as drug targets and vaccine candidates

As peptidases are central enzymes in protein and amino acid metabolism, there has been lots of interest in them as drug targets. For example, broad-spectrum aminopeptidase and 1,2-aminoalcohol-based LAP inhibitors were shown to have anti-malarial activity (Nankya-Kitaka *et al.*, 1998, Howarth & Lloyd, 2000) and arphamenine-A proved to have anti-trypanosomal properties (Knowles, 1993). More recently, *P. falciparum* LAP (*Pf*LAP), involved in haemoglobin digestion, was

identified as an essential enzyme and hence a potential drug target (Stack *et al.*, 2007, Harbut *et al.*, 2011).

In addition, LAPs have potential in vaccine development. Thus *Fasciola hepatica* LAP (FhLAP) was shown to be a vaccine candidate against ruminant fasciolosis (Acosta *et al.*, 2008, Marcilla *et al.*, 2008) and when combined with different adjuvants conferred a high level of immunity in sheep (Maggioli *et al.*, 2011).

4.1.6 Aims

The primary aim was to evaluate the cytosolic M17 LAP-A as a potential drug target in *T. brucei* and additional experiments were to shed light on the physiological function of the LAP-As. The crystal structures of the LAP-As of the TriTryps in apo form and in complex with ligands were to be determined, and the oligomeric state in solution to be confirmed.

4.2 Experimental

4.2.1 Cloning of LAP-A constructs for structural studies

The genes encoding for the LAP-As were amplified from genomic DNA (*T. brucei* Lister427, *T. cruzi* Sylvio X10/1 and *L. major* Friedlin) and cloned into the YSBLIC3C vector using ligase independent cloning (LIC, Aslanidis & de Jong, 1990, Fogg *et al.*, 2006). PCR amplification and cloning of the LAP-As followed the protocol used for *Tb*AK (Table 2-1 and Table 2-2) including identical reaction composition and temperature profile with primer sequences listed in appendix A.2.3. The resulting constructs (*Tb*LAP-3C, *Tc*LAP-3C and *Lm*LAP-3C) encoded the LAP-As with an N-terminal fusion to a hexahistidine tag, separated from the N-terminus of the protein by the recognition site of the HRV 3C protease. Cloning of the *Lm*LAP-A construct was carried out by Nam Duong (a student of the Minority Health and Health Disparities International Research Training (MHIRT) program). The *Tb*LAP-A gene was cloned into the YSBLIC vector by Dr. Javier Garcia Nafria using LIC, resulting in a construct encoding *Tb*LAP-A with a non-cleavable N-terminal fusion to a hexahistidine tag (*Tb*LAP-A-LIC). The success of the cloning was confirmed by restriction endonuclease digestion after plasmid propagation. The restriction endonuclease digestion was carried out using Nco1 and Nde1 (NEB). Reaction conditions are specified in Table 4-2 and were incubated for 1.5 h at 37°C. Subsequently, samples were analysed *via* agarose gel electrophoresis. Positive clones were further verified by sequencing.

Table 4-2: Condition of control digests used to confirm cloning success.

Reagent	Volume per reaction [µL]
10x NEB buffer 4	2
ddH_2O	6
Nco1 (10 U/µL, NEB)	1
Nde1 (20 U/µL, NEB)	1
DNA (= ~50-100 ng/µL)	10
Total volume	20

4.2.2 Expression of LAP-As in *E. coli*

The *Tb*LAP-LIC and *Tb*LAP-3C constructs were transformed into *E. coli* BL21 (DE3) gold cells and plated on LB plates containing 35 µg mL^{-1} kanamycin. The *Tc*LAP-3C and *Lm*LAP-3C constructs were transformed into pLEMO (DE3) cells and plated on LB plates supplemented with 35 µg mL^{-1} kanamycin and 34 µg mL^{-1} chloramphenicol. For all constructs, the next day the colonies were washed off with 3 mL LB each and used for inoculation of fresh LB containing the appropriate antibiotics. The cells were grown at 37°C with shaking at 180 rpm (Innova44, New Brunswick Scientific) till they reached an optical density of 0.4 at 600 nm, where *Tb*LAP-A-LIC and -3C expression was induced by addition of 1 mM IPTG for 4 h and *Tc*LAP-A-3C and *Lm*LAP-A-3C cultures transferred to 20°C with shaking at 251 rpm. *Tc*LAP-A-3C and *Lm*LAP-A-3C expression was induced at 20°C at an optical density of ~ 0.6 at 600 nm over night (~ 16 h) also by addition of 1 mM IPTG. Cells were harvested by centrifugation at 5000 rcf for 30 min and washed with PBS prior to freezing.

4.2.3 Purification of LAP-As

Cells of 1 L culture were thawed on ice and lysed in 35 mL buffer A (20 mM Tris pH8.0, 300 mM NaCl, 30 mM imidazole, 1 mM DTT and 110 µM PMSF) by sonication. Cell debris was removed by centrifugation at 5000 rcf for 30 min and supernatant was loaded onto an equilibrated 5 mM HisTrap FF crude column (GE Healthcare). The column was washed with 3 column volumes of buffer A before the protein was eluted using an imidazole gradient from 30 mM to 500 mM in buffer A. For LAP-As containing a HRV 3C cleavable histidine-tag the protein fractions were pooled and concentrated to a volume of ~ 30 mL before addition of HRV 3C protease in a ratio 1:100 (protease:LAP-A). The proteolysis was monitored by SDS-PAGE and reached completion after incubation overnight at 4°C. In order to remove the histidine-tag and non-cleaved LAP, the protein was exchanged into buffer A using Centrifuge filters (Amicon® Ultra Centrifugal Filter from Millipore or Vivaspin6

from SartoriumStedim biotech, 50 kDa cutoff) and another Nickel affinity purification was performed using the same conditions as before. The flowthrough contained untagged LAP-A. The resulting protein was concentrated and loaded onto a Superdex200 16/60 column (GE Healthcare), and run in buffer containing 10 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT (Figure 4-6 A) at a flowrate of 0.5 mL min^{-1}. The *Tb*LAP-A fused to a non-cleavable histidine-tag was concentrated and loaded onto a Superdex 200 16/60 column directly after the first nickel affinity column. The purifications of the LAP-As were monitored by SDS-PAGE (Figure 4-6 B).

Purified LAP-As were buffer exchanged into 20 mM Tris pH 8.0, 300 mM NaCl and 1 mM DTT concentrated, flash frozen in liquid nitrogen and stored at -80°C.

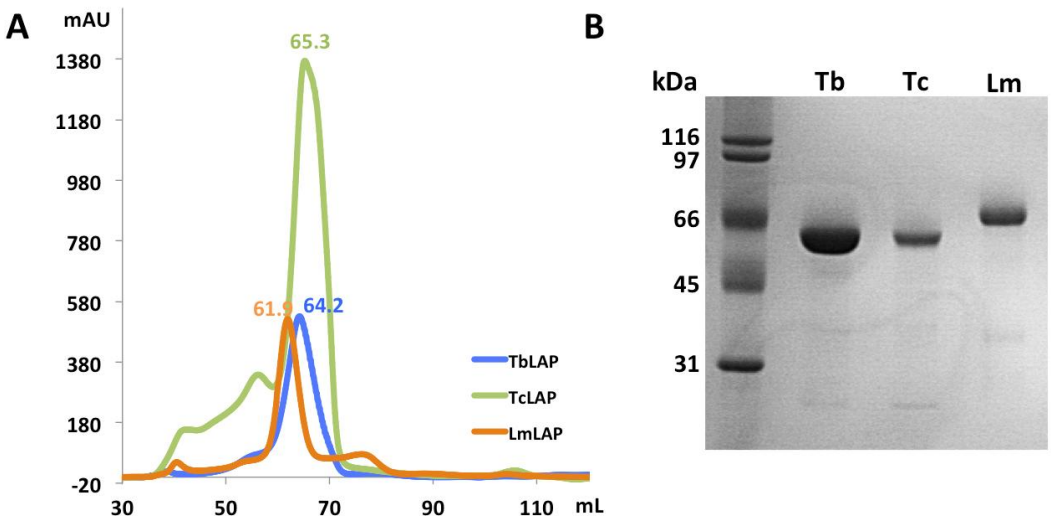


Figure 4-6: LAP-A purification.

(A) Overlay of the size exclusion chromatography traces of *Tb*LAP-A, *Tc*LAP-A and *Lm*LAP-A. (B) SDS-PAGE analysis of purified LAP-As (Tb = *T. brucei*, Tc = *T. cruzi*, Lm = *L. major*).

4.2.4 Native PAGE zymography

The pure LAP-As were mixed with native gel loading buffer (100 mM Tris pH 6.8, 20 % (v/v) glycerol, 0.1 % (w/v) bromophenol blue) and loaded on a 5 % polyacrylamide gel. Each LAP-A was loaded three times on the same gel, which was run at 100 V for 2 h at 4°C to prevent heat denaturation during the electrophoresis.

Subsequently, the gel was washed 3 x 20 min with 50 mM Tris pH 8.0 and cut into three pieces (each containing all three LAP-As). Each piece was incubated with activity assay buffer containing 50 mM Tris pH 8.0, 100 mM KCl and 1 mM $MnCl_2$, supplemented with 7.5 µM fluorescent substrate for 10 min at 37°C. The fluorescent substrates tested were L-Leucine-(4-methyl-7-coumarinylamide) (Leu-AMC), L-Valine-7-amido-4-methylcoumarin (Val-AMC) and L-Proline-7-amido-4-methylcoumarin (Pro-AMC). The fluorescent products were detected using an UV light imager. Subsequently the same gel pieces were stained with Coomassie.

4.2.5 Generation of anti-*Tb*LAP-A antisera in rabbit and affinity purification of polyclonal anti-*Tb*LAP-A antibody

Ten mg of purified *Tb*LAP-A were lyophilised and resuspended in PBS and mixed with Freud's adjuvant prior to injecting into a rabbit. Four inoculations of ~ 400 µg of protein were carried out and antisera were extracted prior to the inoculations and tested for recognition of LAP-A in *T. brucei* lysates and purified protein. For antiserum preparation, rabbit blood was taken and incubated at room temperature for 2 h. Subsequently the blood was incubated at 37°C for 20 min and on ice for 4 h. After the incubations the antisera were separated from cells and insoluble material by centrifugation (4°C, 3000 rpm, 15 min). Sodium azide was added for preservation. The antisera were stored in 500 µL aliquots at 4°C, -20°C and -80°C. To increase specificity, the polyclonal antibody was affinity-purified using pure *Tb*LAP-A coupled to an Affi-Gel® 15 Gel resin (Biorad) following the manufacturer's instructions. Anti-*Tb*LAP-A antiserum was bound to the *Tb*LAP-A coupled to the Affi-Gel® 15 Gel resin, washed and eluted in 1 mL fractions. Elution fractions were tested for sensitivity and specificity *via* western blot analysis as mentioned below.

The injections into the rabbit and the extraction of the blood were done in collaboration with the staff of the Instituto de Parasitologia y Biomedicina "Lopez-Neyra" in Granada, Spain.

4.2.6 Validation of polyclonal anti-*Tb*LAP-A antibody

To test sensitivity 2.5 x 10^6 lysed *T. brucei* bloodstream form and procyclic parasites were loaded in alternating lanes of one SDS-gel and after separation the proteins were transferred to a nitrocellulose membrane. After the blocking step, the membrane was cut in pieces (two lanes each) and incubated in different dilutions (1:50 000, 1:75 000, 1:100 000 and 1:200 000) of primary antibody. This ensures comparability between the different antibody dilutions tested. For secondary antibody incubation the same dilution was used for all membrane pieces. For imaging, the pieces were recombined and the cut sites together with lane marks and marker were marked on the developed film. As control, prepared pre-immune serum was treated as the anti-*Tb*LAP-A antiserum from the other extraction.

To test specificity anti-*Tb*LAP-A antiserum was saturated with a 30-fold excess of purified *Tb*LAP-A protein in blocking solution overnight at 4°C. Western blot analysis of 50, 100 and 500 ng pure *Tb*LAP-A as well as 2.5 x 10^6 bloodstream and procyclic parasites was carried out using untreated and the saturated anti-*Tb*LAP-A antiserum at a dilution of 1:100 000 for incubation. As for the sensitivity test, all samples were loaded onto the same SDS-gel, which was cut into pieces after blocking for incubation with the two different primary antibody solutions.

The cross-reactivity against *T. cruzi* and *L. major* LAP-As was determined by loading 5 x 10^6 and 10 x 10^6 lysed *T. cruzi* trypomastigotes, 10 x 10^6 *T. cruzi* amastigotes (plus host cell background) and 5 x 10^6 lysed *L. major* promastigotes and analysing the western blot. For this anti-*Tb*LAP-A antiserum from the second extraction was used in a dilution of 1:15 000.

4.2.7 SDS-PAGE and Western blots

For the preparation of *T. brucei* lysates (bloodstream form or procyclic), parasites were harvested by centrifugation, washed in PBS and lysed. In each lane of a 12 % SDS gel 2.5 x 10^6 lysed parasites were loaded. After separation on the SDS gel, proteins were transferred to a nitrocellulose membrane. The membranes were

blocked in 5 % (w/v) defatted milk in wash buffer (PBS + 0.1 % (v/v) Tween20®) for 30 min at room temperature, before being incubated for 1 h at room temperature with anti-*Tb*LAP-A primary antibody from the second extraction after immunisation in a dilution of 1:15 000 or anti-*Tb*LAP-A primary antibody of the final extraction in a dilution of 1:200 000. Subsequently, the membranes were washed three times for 10 min in wash buffer and incubated with HRP-conjugated anti-rabbit secondary antibody (Promega) in a dilution of 1:5000 for 1 h at room temperature. The membranes were then developed using chemiluminescence, detected by exposure to photo film.

To compare the amount of LAP-A in different samples, *Tb*SMT was chosen as loading control because its expression level is independent from the one from LAP-A and it is clearly distinguishable on SDS-PAGE and Western blot by its molecular weight. For this, rabbit anti-*Tb*SMT antiserum in 1:10 000 dilution was added with the primary anti-*Tb*LAP-A antibody.

4.2.8 Generation of the construct for ectopic expression of LAP-A

The *Tb*LAP-A gene (Tb427tmp.02.4440) was amplified from genomic DNA (*T. brucei* Lister427) by PCR using the same reaction conditions and temperature profle as described in Table 2-1 and Table 2-2 with the annealing temperature increased to 65°C. The primer sequences can be found in appendix A.2.3 and contained overhangs with the recognition sites for the restriction endonucleases Nde1 (forward primer) and BamH1 (reverse primer). PCR product and the pGRV23b vector (Castillo-Acosta *et al.*, 2012) were digested for 5 h at 37°C using the restriction endonucleases for Nde1 and BamH1 following the composition summarised in Table 4-3. The vector backbone of pGRV23b was purified from a 1.6 % agarose gel after separation from its previous insert by electrophoresis using the Gel extraction Kit (Qiagen) and manufacturers instructions. The insert was purified using the PCR purification Kit (Qiagen). Subsequently vector backbone and insert were ligated using the reaction conditions summarised in Table 4-4, which

were incubated at room temperature for 24 h before transformation into competent *E. coli* (NEB5α) cells.

Table 4-3: Compositions of the resitriction digests for the pGRV23b construct.

Reagent	*Tb*LAP-A	pGRV23b
10x NEB buffer 4	5 µL	5 µL
DNA	39.5 µL (= ~0.5 µg)	10 µL (= ~2 µg)
BSA (10 µg/µL, NEB)	0.5 µL	0.5 µL
ddH_2O	-	30 µL
Nde1 (20 U/µL, NEB)	3 µL	3 µL
BamH1 (20 U/µL, NEB)	2 µL	2 µL
Total volume	50.5 µL	50.5 µL

Table 4-4: Composition of the ligations for *Tb*LAP-A constructs.

Reagent	Volume [µL]
10x T4 ligase buffer (NEB)	2
Vector (digested)	8.5 (= ~100 ng)
Insert (digested)	8.5 (= ~100 ng)
T4 ligase (400 U/µL, NEB)	1
Total volume	20

The cloning success was confirmed by restriction endonuclease digestion after plasmid propagation using the same enzymes as for the cloning and subsequent verification *via* sequencing.

4.2.9 Generation of constructs for LAP-A RNAi

To create a hairpin structure of the transcribed RNA from this vector, the same fragment was inserted into the vector in both 5' → 3' and 3' → 5' orientation at locations separated from eachother by a non-coding linker region. A 465 bp fragment of the *Tb*LAP-A gene (bp 10-475) was selected using the server http://trypanofan.path.cam.ac.uk/software/RNAit.html (Redmond *et al.*, 2003), accessed on 20. 8. 2012) and amplified from genomic DNA (*T. brucei* Lister427) using primers containing the recognition sites for BamH1 and Hind3 (forward primer) and for Hpa1 and Apa1 (reverse primer). PCR conditions and temperature profile were identical to the ones used for the previous LAP-A constructs. Firstly, the

fragment was cloned into the pGR19 vector (Clayton *et al.*, 2005) using the restriction sites Hind3 and Apa1. As these two enzymes require buffer conditions not compatible to oneanother the resitriction digests were carried out separately and the DNA was purified between each step using the PCR purification Kit (Qiagen). Both restriction digests were incubated for 3 h, Hind3 at 37°C and Apa1 at room temperature. The reaction conditions are summarised in Table 4-5 and Table 4-6. Additionally the vector (pGR19) was dephosphorylated by incubation at 37°C for 1 h after addition of 10 µL 10x alkaline phosphatase buffer (Promega), 10 µL alkaline phosphatase (= 10 U, Promega) and 30 µL ddH_2O to the last digest. After DNA purification, digested vector and insert were ligated using the same conditions as described for the construct for ectopic expression of LAP-A in *T. brucei* (section 4.2.8) and the successful insertion was confirmed by restriction digest (with Hind3 and Apa1) after plasmid propagation.

Table 4-5: Conditions of Hind3 digests for first RNAi fragment insertion.

	RNAi fragment	**pGR19**
DNA (= ~ 2 µg)	40 µL	10 µL
10 x NEB buffer 2	5 µL	5 µL
ddH_2O	2.5 µL	32.5 µL
Hind3 (NEB)	2.5 µL	2.5 µL
Total volume	**50 µL**	**50 µL**

Table 4-6: Conditions of Apa1 digests for first RNAi fragment insertion.

	RNAi fragment (cut with Hind3)	**pGR19 (cut with Hind3)**
DNA	44 µL	44 µL
10 x NEB buffer 4	5 µL	5 µL
BSA (NEB)	0.5 µL	0.5 µL
Apa1 (NEB)	0.5 µL	0.5 µL
Total volume	**50 µL**	**50 µL**

To insert the fragment in opposite orientation from the first insert, 2 µg of each, insert and previous vector, were incubated for 3 h at 37°C with BamH1 and Hpa1 under the conditions specified in Table 4-7. The vector was dephosphorylated as described above and both (digested vector and insert) were ligated after DNA purification using the same protocol as above (section 4.2.8).

Table 4-7: Conditions of restriction digests for second RNAi fragment insertion.

	RNAi fragment	pGR19 + first insert
DNA (= ~ 2 µg)	39.5 µL	10 µL
10 x NEB buffer 4	5 µL	5 µL
BSA (NEB)	0.5 µL	0.5 µL
ddH_2O	-	29.5 µL
Hpa1 (NEB)	2.5 µL	2.5 µL
BamH1 (NEB)	2.5 µL	2.5 µL
Total volume	**50 µL**	**50 µL**

Cloning success was confirmed by restriction digestion (with Hpa1 and BamH1) after plasmid propagation. A plasmid map of the final vector is shown in appendix A.2.5.4.

4.2.10 Generation of LAP-A cMyc-fusion constructs for overexpression in *T. brucei*

The LAP-A gene from *T. brucei* was amplified by PCR and cloned into the backbone of the pGRV33 vector (Castillo-Acosta *et al.*, 2012) using the restriction endonucleases Nde1 and Hpa1 and ligation with T4 ligase, following the same protocol as described for the construct for ectopic expression of LAP-A in *T. brucei* (section 4.2.8) with the exception of the restriction digest which was incubated over night at 37°C. The primer sequences are listed in appendix A.2.3. Cloning success was confirmed by restriction endonuclease digestion (with Nde1 and Hpa1) and sequencing.

4.2.11 Generation of *T. brucei* LAP-A knockout cell lines

To generate the two replacement cassettes for the double knockouts, untranslated regions (UTRs) upstream (5') and downstream (3') of the *Tb*LAP-A gene on the genome were chosen and cloned into a vector each side of a blasticidin resistance marker. To create the second vector (for removal of the second allel on the genome) the blasticidin resistance cassette was exchanged for a hygromycin

resistance cassette. Transfection of the final vectors into *T. brucei* cells results in the exchange of the *Tb*LAP-A gene from the genome for the antibiotic resistances by homologous recombination.

For this a 538 bp fragment of the 5'UTR upstream of the *Tb*LAP-A gene (-61 - -599) and a 467 bp fragment of the 3'UTR downstream of the *Tb*LAP-A gene (+7 - +475) were amplified from genomic DNA (*T. brucei* Lister427) by PCR using the same conditions as described in Table 2-1 and Table 2-2. The primers used for the 5'UTR fragment contained a Not1 site (forward) and a Xho1 site (reverse) and the primers used for the 3' UTR fragment contained a Stu1 site (forward) and Nhe1 and Not1 sites (reverse). The primer sequences can be found in appendix A.2.3. Firstly, the 5'UTR fragment was inserted into the pHD887 vector (kindly provided by V. Castillo-Acosta, Instituto de Parasitologia y Biomedicina "Lopez-Neyra", Granada, Spain) using incubation with the endonucleases Xho1 and Not1 for 2 h at 37°C under the conditions specified in Table 4-8. This was followed by dephosphorylation of the vector and DNA purification of vector and insert using the PCR purification Kit (Qiagen). Insert and vector were ligated overnight at room temperature following the conditions in Table 4-9.

Table 4-8: Conditions of restriction digests for 5'-UTR fragment insertion.

	5'-UTR	**pHD887**
DNA (= ~ 2 µg)	42.5 µL	13 µL
10 x NEB buffer 3	5 µL	5 µL
BSA (NEB)	0.5 µL	0.5 µL
ddH_2O	-	29.5 µL
Not1 (NEB)	1 µL	1 µL
Xho1 (NEB)	1 µL	1 µL
Total volume	**50 µL**	**50 µL**

Table 4-9: Composition of ligations for 5'-UTR construct.

Reagent	**Volume [µL]**
10x T4 ligase buffer (NEB)	2
Vector (digested)	8 (= ~100 ng)
Insert (digested)	4.25 (= ~100 ng)
ddH_2O	4.75
T4 ligase (400 U/µL, NEB)	1
Total volume	**20**

Ligated vector was transformed into *E. coli* cells and the successful insertion of the 5'-UTR was confirmed by control digests with Xho1 and Not1.

Subsequently, the 3'UTR fragment was cloned into the resulting vector from the previous step using the restriction endonucleases Stu1 and Nhe1. The incubation of digests was at 37°C and the reaction conditions are summarised in Table 4-10. As the Stu1 stock was very old, it was added to the reaction 3 h earlier than the Nhe1, which was incubated for 2 h with the DNA (so the total reaction time was 5 h for Stu1 and 2 h for Nhe1). The vector was dephosphorylated as described before. Ligations were carried out using incubation with T4 ligase over night at room temperature following the conditions summarised in Table 4-9. Cloning was confirmed by restriction endonuclease digestion.

Table 4-10: Conditions of restriction digests for 3'-UTR fragment insertion.

	3'-UTR	**pHD887-5'-UTR**
DNA (= ~ 2 µg)	30 µL	30 µL
10 x NEB buffer 4	5 µL	5 µL
BSA (NEB)	0.5 µL	0.5 µL
ddH_2O	9.5 µL	9.5 µL
Stu1 (NEB)	3 µL	3 µL
Nhe1 (NEB)	2 µL	2 µL
Total volume	**50 µL**	**50 µL**

In the resulting vector (pHD887**B**/UTRs) the two UTR fragments flank the blasticidin S transferase (BSD) resistance gene and it was used to transform parental bloodstream form and procyclic cells as described below. To remove the second *Tb*LAP-A allele from the genome, the hygromycin phosphotransferase (HYG) selectable marker from pGR19 (Clayton *et al.*, 2005) was sub-cloned into pHD887**B**/UTRs. For this pHD887**B**/UTRs and pGR19 were digested with the restriction endonuclease SnaB1 for 3 h at 37°C under the conditions specified in Table 4-11. The pHD887**B**/UTRs backbone and HYG from pGR19 were purified from a 2 % agarose gel using the Gel extraction Kit (Qiagen). The vector backbone was dephosphorylated as described above and purified using a PCR purification kit (Qiagen). HYG and vector were ligated using the same protocol as in Table 4-4 in section 4.2.8.

Table 4-11: Conditions of restriction digests for antibiotic resistance exchange.

	pGR19	pHD887B/UTRs
DNA (= ~ 2 µg)	10 µL	5 µL
10 x NEB buffer 4	5 µL	5 µL
BSA (NEB)	0.5 µL	0.5 µL
ddH_2O	34 µL	39 µL
SnaB1 (NEB)	1 µL	1 µL
Total volume	**50 µL**	**50 µL**

Positive clones were identified by control digests with SnaB1 and propagated. To confirm the right orientation of HYG, the clones were digested with Xho1 and Nco1 for 2 h at 37°C. The reaction condition is summarised in Table 4-12. This restriction digest led to a 740 bp or a 1055 bp fragment for plasmids containing HYG in the correct or the wrong orientation respectively, which were distinguished by agarose gel electrophoresis. Additionally, successful construction of the vectors was confirmed by sequencing and plasmid maps can be found in appendix A.2.5.5.

Table 4-12: Conditions used to confirm HYG orientation.

	Volume/reaction [µL]
DNA	5
10 x NEB buffer 4	1.5
ddH_2O	8
Xho1 (NEB)	0.25
Nco1 (NEB)	0.25
Total volume	**15**

The resulting vector pHD887**H**/UTRs was used to transfect the positive single knockouts as described below.

4.2.12 *Trypanosoma brucei* cell lines, culture and transfections

The cell lines used in this work were *Trypanosoma brucei brucei* bloodstream form S16 (Wirtz *et al.*, 1999) and the procyclic cell line 449 (Estevez *et al.*, 2001). The blood stream parasites were cultured at 37°C and 5 % CO_2 in HMI-9 media, containing 10 % (v/v) fetal bovine serum (FBS) (Hirumi & Hirumi, 1994). The

procyclic trypanosomes were cultured at 28°C in SDM-79 media supplemented with 10 % (v/v) FBS and 7.5 μg mL^{-1} Haemin (Brun & Schonenberger, 1979).

To generate double knockout and knockdown (RNAi) parasite cells lines as well as a parasite line overexpressing LAP-A and LAP-A cMyc-fusion protein, parental cell lines of bloodstream form and procyclics were transfected with the respective targeting vectors. Firstly, the circular plasmids were linearised by digestion with endonuclease NotI and subsequently purified by ethanol precipitation. For each transfection 24 x 10^6 parasites were transfected with 3-10 μg of linearised targeting vectors using electroporation as previously described (Wirtz *et al.*, 1998, Wirtz *et al.*, 1999). Selection of positive clones was carried out by limited dilution in 24-well cell culture plates in the presence of selection drugs, Table 4-13. Each transfection led to 1-3 clones (bloodstream form) or 6-12 clones (procyclics), respectively. Selection of clones was completed after 7-14 days.

Growing clones were tested by determination of the LAP-A amount in the cells in comparison with the parental cell lines using western blot analysis after induction of RNAi or overexpression of the LAP-A. Positive RNAi clones did show a significant reduction of LAP-A upon induction, overexpressing clones showed a strong increase of LAP-A (with and without cMyc-fusion). Uninduced RNAi and overexpression clones showed the same amount of LAP-A than the parental cells.

Table 4-13: Selection drugs used for the different *T. brucei* parasite strains.

TbBF: bloodstream form, Tb449: procyclic form

Targeting vector – purpose	Selective antibiotics	Concentration for TbBF [μg mL^{-1}]	Concentration for Tb449 [μg mL^{-1}]
Replacement of first LAP-A allele from the genome (sKO)	Blasticidin	10	20
Replacement of second LAP-A allele from the genome (dKO)	Hygromycin (Additionally to the blasticidine from the sKO)	5	50
LAP-A RNAi	Hygromycin	5	50
Ectopic copy control vector for LAP-A overexpression	Puromycin	0.1	1
Ectopic copy control for expression of LAP-A cMyc fusion	Puromycin	0.1	1

4.2.13 *T. brucei* cultures in minimal media

Bloodstream form parasites were cultured in HMI-9 media, containing 10 % (v/v) dialysed FBS to reduce the amount of undefined small molecule components.

For further investigations into the physiological role of LAP-A, procyclic parasites were cultured in minimal essential media (MEM) containing 7.5 μg mL^{-1} haemin and 10 % (v/v) FBS or dialysed FBS to remove glucose as energy source and other undefined small molecules from the media. To test the influence of amino acid limitation the same MEM with dialysed FBS was prepared either without leucine or without the “Non-essential amino acids”-mix (100x, Sigma M7145), thus omitting L-alanine, L-asparagine, L-aspartic acid, L-glutamic acid, glycine and L-serine, and reducing the amount of L-proline. The exact composition of the media can be found in appendix A.3.

Generally, when culture media or volume was changed, the parasites were cultured for 6-9 days (2-3 dilutions) to give them time to adapt to the new conditions and reach a reproducible growth rate. Growth rates in media with dialysed FBS were

lower, so cultures were incubated at culture temperature for 3 days (instead of 2) before the next dilution. For the procyclic parasite cultures in amino acid depleted MEM the cultures were diluted to 1 000 000 parasites mL^{-1} and incubated for 3 days for adaptation.

4.2.14 Determination of parasite growth upon RNAi induction and overexpression of LAP-A

The targeting vectors of RNAi and ectopic copy control contained the tetracycline-inducible promotor. RNAi and overexpression were induced by the addition of 1 $\mu g\ mL^{-1}$ doxycycline to the cultures directly after dilution in fresh media.

For recording the growth curves, all procyclic cultures (induced, un-induced and parental cell lines) were diluted to and induced at 200 000 parasites mL^{-1} and cultured for two days before counting and dilution in fresh media. The cell growth was followed for 6-10 days. For the growth curves of bloodstream form cultures the cells were diluted to and induced at 5 000 parasites mL^{-1} and cultured for two days before the next dilution. Additionally, the growth curves for bloodstream form parasites were recorded when diluting to and inducing RNAi/over-expression at 100 000 parasites mL^{-1} and letting the cultures grow for 24 h before the next dilution. The two different protocols addressed the possible influence of the high dilution itself on parasite growth.

4.2.15 Statistical analysis of changes in parasite growth

The significance of changes in generation times of *T. brucei* dKO clones and upon induction of RNAi and overexpression was evaluated using the Welch's t-test (two-sample, unpaired) in GraphPad Prism (version 6). Results giving $p \leq 0.05$ were considered significant.

4.2.16 Glycerol stocks of *T. brucei* clones

All clones generated were frozen at -80°C. For each frozen aliquot (bloodstream form and procyclics) 20 - 30 x 10^6 parasites were harvested by centrifugation at 1400 rcf for 10 min and resuspended in 250 µL HMI-9 media, containing 10 % (v/v) FBS and 10 % Glycerol (bloodstream form) or 250 µL SDM-79 media with 10 % (v/v) FBS, 7.5 µg mL^{-1} Heamin and 10% Glycerol (procyclic form).

4.2.17 Cell cycle analysis by flow cytometry

Bloodstream form and procyclic parasites (5-20 x 10^6) were harvested by centrifugation at 1400 rcf for 10 min and washed twice with PBS before being fixed overnight at -20°C in 70 % ethanol (in PBS). Fixed parasites were washed twice in PBS, resuspended in 500 µL PBS containing 40 µg mL^{-1} propidium iodide and 10 µg mL^{-1} RNase A and incubated for 30 min at 37°C in the dark. The propidium iodide intercalates into the DNA of the cells and the DNA content of the cells was analysed using a FACSCalibur (Becton Dickinson, USA) analytical flow cytometer and the CellQuest software. The fluorescence signals were recorded in FL3-H for the cell population gated in SSC-H vs. FSC-H. Data were analysed using FlowJo software v7.6.1 calculating the percentages of each cell cycle stage (Sub-G_1, G_1, S, G_2/M and Super-G_2). After normalisation, data from three independent cultures per condition and strain were averaged and the standard deviation calculated. The analysed data were plotted using Microsoft Excel 2010.

4.2.18 Cell cycle analysis by fluorescence microscopy

Bloodstream form and procyclic parasites were harvested by centrifugation, washed in PBS and fixed in 4 % p-formaldehyde in wash solution (PBS + 0.2 % (v/v) Tween20®) on a poly-L-lysine coated glass slide for 20 min at room temperature. After washing twice the slides were dehydrated in 100 % methanol for 1 min. Once the methanol had evaporated the slides were stained and mounted with

Vectashield-DAPI (Vector Laboratories, Inc.), covered with cover slips and sealed with nail polish. Images were taken using a Zeiss Axiophot microscope (Carl Zeiss, INC) and the percentages of 1N1K, 1N2K, 2N2K and pathological (0N1K 1N0K, 2N1K and xNxK) cells were analysed manually.

4.2.19 Immunofluorescence

For preparation of immunofluorescence in bloodstream and procyclic forms, 10×10^6 parasites from mid-log cultures were harvested by centrifugation at 1400rcf for 10 min. Cells were resuspended in 1 mL PBS containing MitoTracker® (Invitrogen) and incubated in the dark at culture temperature for 15 min. After incubation the cells were pelleted by centrifugation and washed with 1 mL PBS before resuspension in a solution of 4 % *p*-formaldehyde and wash solution (PBS + 0.2 % (v/v) Tween20®). The cells were fixed on a poly-L-lysine coated glass slide for 20 min at room temperature, washed twice and incubated for 75 min in a mixture of wash solution with 1 % Blocking Reagent (Roche) and 1 % IGEPAL (Sigma) in order to block and permeabilise the cells. The slides were incubated for 1 h at room temperature with primary rabbit anti-*Tb*LAP-A antibody (affinity-purified, diluted 1:500 for over-expressing cells, 1:100 for all others in blocking solution) or monoclonal anti-cMyc antibody (Sigma, 1:100 in blocking solution) to find the optimal conditions. After 6 washes of 5 min with wash solution, the slides were incubated with secondary antibody for 1 h at room temperature: FITC-conjugated anti-rabbit antibody (Sigma, 1:40 diluted in blocking solution) for anti-*Tb*LAP-A primary antibody or AlexaFluor488-conjugated anti-mouse antibody (Sigma, 1:40 diluted in blocking solution). Subsequently, the slides were washed 6 times for 5 min and dehydrated by incubation in 100 % methanol for 1 min. After evaporation of the methanol the slides were stained and mounted with Vectashield-DAPI (Vector Laboratories, Inc.), covered with cover slips and sealed with nail polish.

Vertical stacks of 20 slices (0.2 µm steps) were taken with a Leica TCS SP5 confocal microscope with 63x magnification. Image preparation and analysis were carried out with the LAS-AF software (Leica) and ImageJ (version 1.37; National Institutes of

Health). The confocal microscopy images of bloodstream forms and procyclics (parental and LAP-A-cMyc overexpressing, 12 images each) were analysed for colocalisation of LAP-A with DAPI and MitoTracker using Li's approach (Li *et al.*, 2004) with the JACoP Plugin (Bolte & Cordelieres, 2006) in ImageJ.

4.2.20 Crystallisation of the LAP-As

Prior to setting up crystallisation screens, the LAP-As were thawed at room temperature and pre-incubated with the desired metal salts for 2 h and subsequently with ligand for another 2 h on ice. Precipitated protein was removed by centrifugation and protein concentration was determined from the absorbance at 280 nm. Concentrations of proteins and ligands used in crystallisation as well as the initial commercially available screens used are summarised in Table 4-14. The initial screens were pipetted in sitting drop vapour diffusion using a Mosquito robot (TTP Labtech, UK) with a drop size of 150 nL protein plus an equal amount of well solution.

Table 4-14: Concentrations and screens used for crystallisation screening.

Protein	Protein concentration [mg mL^{-1}]	Additives [mM]	Inhibitors/Ligands and concentrations [mM]	Initial screens
*Tb*LAP-A (tagged)	20, 30, 50	5, 20 mM $CaCl_2$, 5, 20 mM $CoSO_4$, 5, 20 mM $MgCl_2$, 5, 20 mM $MnSO_4$, 1, 5 mM $MnCl_2$, 1 µM $ZnSO_4$, 1, 5 mM $ZnSO_4$, 1 µM $ZnCl_2$	5, 10 mM Actinonin, 5, 10 mM Amastatin, 5, 10 mM Bestatin, 1, 10 mM L-leucine	PACT, JCSG, Index, MPD, Ammonium sulfate
*Tb*LAP-A (not tagged)	30	1 mM $MnCl_2$, 1 µM $ZnCl_2$	5 mM Bestatin, 1 mM L-Leucine	JCSG, Ammonium sulfate
*Tc*LAP-A	16.6 48.8	5 mM $MnSO_4$, 1 mM $MnCl_2$, 1 µM $ZnSO_4$/ $ZnCl_2$	5, 10 mM Actinonin, 5, 10 mM Amastatin, 5, 10 mM Bestatin	JCSG, Index, Ammonium sulphate, Morpheus, MPD
*Lm*LAP-A	26	5 mM $MnSO_4$, 1 mM $MnCl_2$, 1 µM $ZnSO_4$	5, 10 mM Actinonin, 5, 10 mM Amastatin, 5, 10 mM Bestatin	JCSG, Ammonium sulfate

Hits were optimised using sitting and hanging drop vapour diffusion with a drop size of 1 µL protein plus 1 µL well solution set up manually. For the *Lm*LAP-A the additive screen (Hampton Research) was used to identify a compound that enhanced the diffraction of the crystals. All crystals were grown at 20°C. The crystallisation conditions leading to LAP-A crystals that allowed structure solution, as well as the ligands bound in the active sites of these structures are listed in Table 4-15.

Table 4-15: Crystallisation conditions of the LAP-As and ligands bound.

Protein	Crystal name	Crystallisation condition	Ligand bound in the active site in the structure
*Tb*LAP-A	apo *Tb*LAP-A	0.8 M Na citrate pH 7.0, 5 mM $CoSO_4$	-
*Tb*LAP-A	*Tb*LAP-A-Mn-bestatin	0.1 M HEPES pH 7.0, 40 % (v/v) MPD, 5 mM $MnCl_2$, 5 mM bestatin	Mn^{2+}, HCO_3^-, bestatin
*Tb*LAP-A	*Tb*LAP-A-Mn-actinonin	0.1 M Tris pH 8.0, 2.4 M $(NH_4)_2SO_4$, 5 mM $MnSO_4$, 5 mM actinonin	Mn^{2+}, actinonin
*Tb*LAP-A	*Tb*LAP-A-Mn	0.1 M Tris pH 6.8, 0.2 M $MgCl_2$, 11 % (w/v) PEG 8000, 10 µM $ZnCl_2$, 10 mM L-leucine	Mn^{2+}, HCO_3^-
*Tc*LAP-A	*Tc*LAP-A-Mn-citrate	0.1 M Phosphate-citrate pH 4.2, 40 % (v/v) PEG 300, 5 mM $MnSO_4$, 5 mM amastatin	Mn^{2+}, SO_4^{2-}, citrate
*Tc*LAP-A	apo *Tc*LAP-A	0.1 M Citric acid pH 4.0, 1.6 M $(NH_4)_2SO_4$, 1 mM $MnCl_2$, 5 mM actinonin	SO_4^{2-}
*Lm*LAP-A	*Lm*LAP-A-Mn-actinonin	0.8 M Succinic acid pH 7.0, 3 % (w/v) 1,5-Diamino-pentane di-hydrochloride, 5 mM $MnSO_4$, 5 mM actinonin	Mn^{2+}, HCO_3^-, actinonin

4.2.21 Data Collection and Structure Solution of the LAP-As

Unless otherwise indicated, all computations were performed using programs of the CCP4 suite (Winn *et al.*, 2011). The data were processed with MOSFLM (Battye *et al.*, 2011) and AIMLESS (Evans & Murshudov, 2013). The crystal of apo *Tb*LAP-A was cryo-protected in well solution containing 20% (v/v) glycerol prior to

vitrification. The other LAP-A crystals were not cryo-protected. Crystallographic data and statistics are shown in Table 4-16 - 8.

The structures were solved by molecular replacement (MR) using PHASER (McCoy, 2007) and MOLREP (Vagin & Teplyakov, 1997). The highest structural homology to the *Tb*LAP-A was predicted to be the *C. elegans* LAP using BLAST searching against entries in the PDB, and the protomer was used in MR to generate initial phases from which the *Tb*LAP-A model was built. For all subsequent LAP-A structures the first *Tb*LAP-A model (apo) was used for MR. Model building was carried out using BUCCANEER (Cowtan, 2006). Manual fitting of the electron-density maps, as well as model building and fitting of the ligands were carried out with COOT (Emsley *et al.*, 2010). The structures were refined using REFMAC5 (Murshudov *et al.*, 2011) and validated using COOT and *Molprobity* (Chen *et al.*, 2010).

Table 4-16: Crystallographic data and statistics for *Tb*LAP-A crystals.

In this and all subsequent data tables: Data for the highest resolution shell shown in brackets, the Ramachandran Plot was calculated using COOT (Emsley *et al.* 2010), the poor rotamer analysis was carried out in *Molprobity* (Chen *et al.* 2010).

	apo *Tb*LAP-A	***Tb*LAP-A-Mn**
Data collection		
Diffraction Source	DLS beamline I04	DLS beamline I04
Wavelength (Å)	0.9795	0.9795
Temperature (K)	100	100
Detector	Pilatus 6M-F	Pilatus 6M-F
Rotation range per image (°)	0.5	0.2
Total Rotation range (°)	180	220
Crystal data		
Space group	$P2_12_12_1$	$P2_1$
a, b, c (Å)	162.33, 162.45, 176.99	87.37, 143.8, 270.1
α, β, γ (°)	90, 90, 90	90, 95.47, 90
Resolution (Å)	72.61-2.60 (2.64-2.60)	32.73-2.40 (2.44-2.40)
Total Reflections	1065695 (53405)	803022 (32773)
Unique Reflections	143992 (7106)	244412 (11362)
Completeness (%)	100 (100)	94.5 (88.7)
Redundancy	7.4 (7.5)	3.3 (2.9)
R_{merge}	0.165 (1.019)	0.117 (0.351)
[I/σ (I)]	9.5 (2.0)	5.7 (2.0)
Matthew's coefficient	3.51	2.40
Refinement statistics		
Reflections, working set	136637	232199
Reflections, test set	7257	11981
Resolution Range (Å)	72.61-2.60	32.73-2.40
R-factor	0.1742	0.1877
R_{free}	0.1934	0.2062
No. of non-H atoms		
Protein	22152	45624
Ligands	-	72
Water	500	661
Mean B factors (Å^2)		
Protein	29.0	25.7
Ligands	-	30.5
Water	31.4	28.9
RMS deviation from ideal		
Bond length (Å)	0.0119	0.0096
Bond angles (°)	1.4252	1.2938
Ramachandran Plot (%)		
Residues in favored region	97.88	97.68
Residues in allowed region	2.02	2.06
Outliers	0.10	0.26
Molprobity score	0.92	1.14
Poor Rotamers (%)	1.19	2.04
PDB ID		

Table 4-17: Crystallographic data and statistics for *Tb*LAP-A-complex crystals.

	***Tb*LAP-A-Mn-bestatin**	***Tb*LAP-A-Mn-actinonin**
Data collection		
Diffraction Source	DLS beamline I03	DLS beamline I04-1
Wavelength (Å)	0.97625	0.9200
Temperature (K)	100	100
Detector	Pilatus 6M	Pilatus 2M
Rotation range per image (°)	0.2	0.2
Total Rotation range (°)	220	180
Crystal data		
Space group	$P2_1$	$P2_1$
a, b, c (Å)	87.26, 143.45, 268.03	95.91, 165.73, 121.98
α, β, γ (°)	90, 95.54, 90	90, 112.45, 90
Resolution (Å)	80.33-2.30 (2.34-2.30)	41.50-3.00 (3.07-3.00)
Total Reflections	1132102 (52607)	217196 (14160)
Unique Reflections	290030 (14130)	68219 (4424)
Completeness (%)	99.7 (99.0)	97.0 (98.0)
Redundancy	3.9 (3.7)	3.2 (3.2)
R_{merge}	0.091 (0.523)	0.166 (0.440)
[I/σ (I)]	9.1 (2.1)	4.5 (2.0)
Matthew's coefficient	2.51	2.70
Refinement statistics		
Reflections, working set	275670	63948
Reflections, test set	14322	3327
Resolution Range (Å)	80.33-2.30	41.50-3.00
R-factor	0.1649	0.2531
R_{free}	0.1910	0.2710
No. of non-H atoms		
Protein	45631	22529
Ligands	436	174
Water	2798	7
Mean B factors (Å^2)		
Protein	22.3	2.0
Ligands	47.3	5.5
Water	37.2	2.0
RMS deviation from ideal		
Bond length (Å)	0.0120	0.0120
Bond angles (°)	1.4022	1.3810
Ramachandran Plot (%)		
Residues in favored region	97.27	97.75
Residues in allowed region	2.52	2.06
Outliers	0.21	0.19
Molprobity score	1.21	1.15
Poor Rotamers (%)	2.37	1.77
PDB ID		

Table 4-18: Crystallographic data and statistics for the *Tc*LAP-A crystals.

	***Tc*LAP-A-Mn-citrate**	**apo *Tc*LAP-A**
Data collection		
Diffraction Source	DLS beamline I04	DLS beamline I04
Wavelength (Å)	0.9795	0.9795
Temperature (K)	100	100
Detector	Pilatus 6M-F	Pilatus 6M-F
Rotation range per image (°)	0.5	0.2
Total Rotation range (°)	180	180
Crystal data		
Space group	H32 / R32	H32 / R32
a, b, c (Å)	160.12, 160.12, 203.67	161.42, 161.42, 205.11
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	57.31-2.32 (2.40-2.32)	48.14-2.30 (2.38-2.30)
Total Reflections	477971 (47316)	380916 (39032)
Unique Reflections	43546 (4249)	43976 (4437)
Completeness (%)	100 (100)	96.2 (100)
Redundancy	11.0 (11.1)	8.7 (8.8)
R_{merge}	0.188 (1.202)	0.181 (0.987)
[I/σ (I)]	11.0 (2.1)	9.1 (2.4)
Matthew's coefficient	2.28	2.34
Refinement statistics		
Reflections, working set	41361	41496
Reflections, test set	2183	2061
Resolution Range (Å)	57.31-2.32	48.14-2.30
R-factor	0.1652	0.2052
R_{free}	0.2076	0.2558
No. of non-H atoms		
Protein	7648	7537
Ligands	51	20
Water	382	139
Mean B factors (Å^2)		
Protein	26.89	26.07
Ligands	47.69	41.65
Water	33.17	25.54
RMS deviation from ideal		
Bond length (Å)	0.0101	0.0111
Bond angles (°)	1.4295	1.3625
Ramachandran Plot (%)		
Residues in favored region	95.95	96.27
Residues in allowed region	3.85	3.54
Outliers	0.20	0.20
Molprobity score	1.31	1.21
Poor Rotamers (%)	1.86	0.90
PDB ID		

Table 4-19: Crystallographic data and statistics for the *Lm*LAP-A-Mn-actinonin.

	***Lm*LAP-A-Mn-actinonin**
Data collection	
Diffraction Source	DLS beamline I24
Wavelength (Å)	0.97781
Temperature (K)	100
Detector	Pilatus 6M
Rotation range per image (°)	0.2
Total Rotation range (°)	100
Crystal data	
Space group	P321
a, b, c (Å)	116.44, 116.44, 91.49
α, β, γ (°)	90, 90, 120
Resolution (Å)	38.11-2.50 (2.60-2.50)
Total Reflections	125585 (14243)
Unique Reflections	21850 (2528)
Completeness (%)	87.8 (90.2)
Redundancy	5.7 (5.6)
R_{merge}	0.150 (0.747)
[I/σ (I)]	8.2 (2.0)
Matthew's coefficient	3.09
Refinement statistics	
Reflections, working set	20818
Reflections, test set	1031
Resolution Range (Å)	38.11-2.50
R-factor	0.1877
R_{free}	0.2408
No. of non-H atoms	
Protein	3893
Ligands	33
Water	50
Mean B factors (Å^2)	
Protein	23.8
Ligands	43.1
Water	27.5
RMS deviation from ideal	
Bond length (Å)	0.0134
Bond angles (°)	1.6575
Ramachandran Plot (%)	
Residues in favored region	95.14
Residues in allowed region	4.67
Outliers	0.19
Molprobity score	1.76
Poor Rotamers (%)	2.26
PDB ID	

4.2.22 Identification of metal ions in the *Tb*LAP-A structure

X-ray fluorescence (XRF) scans were collected for the crystal grown in the presence of $ZnCl_2$ and L-leucine. These crystals were set up using protein treated with EDTA during purification to remove any metals and no other metal except $ZnCl_2$ was added prior to crystallisation. The XRF scans were analysed using the PyMca software (Sole *et al.*, 2007). A specific edge scan for zinc did not give a signal.

X-ray data were collected at each side of the theoretical zinc edge at wavelengths of 0.9795 Å and 1.3051 Å. After data processing with Xia2 (Winter, 2010) and structure solution and refinement as described above, reflection data from REFMAC5 (including the phases) were combined with the Friedel difference (DANO) and the standard deviation of F2 (SIGDANO) from the unmerged data file (from Xia2) using CAD from CCP4. Anomalous difference maps were generated using the FFT program in CCP4.

4.2.23 SEC-MALLS

Purified and frozen *Tb*LAP-A, *Tc*LAP-A and *Lm*LAP-A were thawed and diluted to concentrations of 0.1, 0.5, 1.0 and 3.0 mg mL^{-1}. Additionally *Tb*LAP-A was diluted to 5 mg mL^{-1}. The LAP-As were incubated at the respective concentrations for a minimum of 3 h at room temperature before automatic injection into a SEC-MALLS system (Wyatt Dawn HELEOS-II 18-angle light scattering detector and Wyatt Optilab rEX refractive index monitor linked to a Shimadzu HPLC system comprising LC-20AD pump, SIL-20A Autosampler and SPD20A UV/Vis detector) using a Superdex200 10/300 size exclusion column and a running buffer containing 10 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT. In addition, *Tb*LAP-A was run in a buffer containing 50 mM Tris pH 8.0, 100 mM KCl, 1 mM $MnCl_2$ and 1 mM DTT.

The sample preparation of *Tc*LAP-A and *Lm*LAP-A and part of the data analysis were carried out by Michael Norris (MChem undergraduate project student). The data were analysed using the ASTRA software (version 5.3.4.14) and the molecular masses were calculated for each sample.

4.3 Results

The first aim was to explore the essential character of LAP-A and its potential as a drug target in *T. brucei* parasites. For this purpose two different life cycle stages of the parasites, bloodstream form and procyclic, were cultured and used for generation of LAP-A knockout, knock down and overexpression strains, which were subsequently characterised. *Tb*LAP-A-specific polyclonal antibody was generated and purified for determination of the intracellular LAP-A amount and its localisation and characterisation. Secondly, structural studies of the TriTryp LAP-As and the determination of their oligomeric states in solution were to aid rational drug design.

4.3.1 Overexpression and purification of active protein

The TriTryp LAP-As were expressed and purified successfully as described above. The activities against the substrate analogues Leu-AMC, Val-AMC and Pro-AMC were confirmed by native gel zymography (Figure 4-7). All three LAP-As were active against Leu-AMC. *Tb*LAP-A showed activity against Val-AMC and Pro-AMC, though much less than for Leu-AMC. For *Tc*LAP-A and *Lm*LAP-A no activity against Val-AMC and Pro-AMC could be detected.

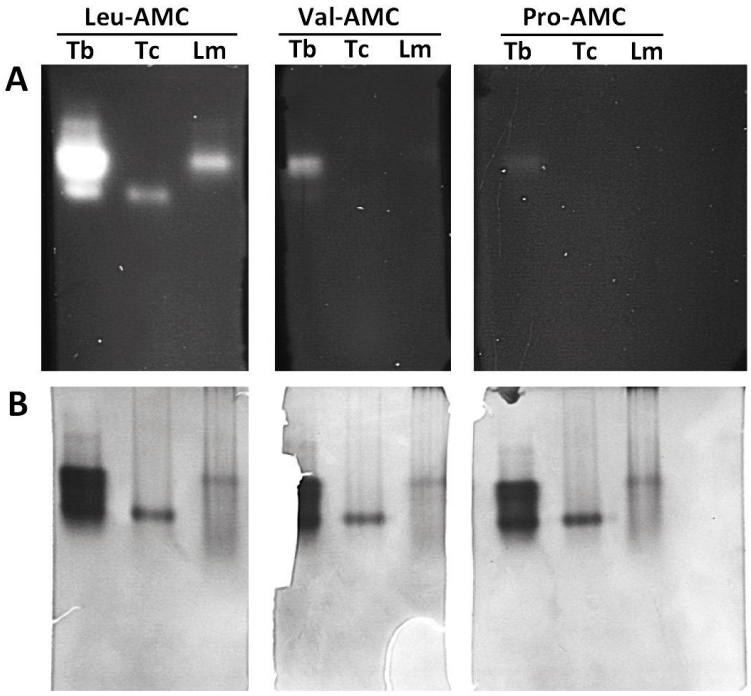


Figure 4-7: Native gel zymography of recombinant TriTryp LAP-As.

The figure shows the native gel pieces containing the LAP-As, which were incubated with Leu-AMC, Val-AMC or Pro-AMC for 10 min at 37°C. (A) Visualisation of the gel pieces using UV light, showing the AMC products hydrolysed by the LAP-As. (B) Coomassie staining of the same native gel pieces.

4.3.2 Validation of anti-*Tb*LAP-A antiserum

The generated anti-*Tb*LAP-A antiserum recognises *Tb*LAP-A in cell extracts of bloodstream forms and procyclics indicated by a strong band at the expected molecular weight, Figure 4-8 A. In western blots the anti-serum gives strong signals in dilutions up to 1:200 000 and recognises an additional product of ~ 33 kDa in the cell lysates, especially when used at lower dilutions. This additional product is recognised in the lysate of bloodstream form parasites, and less in procyclics. As it is recognised also in the procyclic double knockout and bloodstream and procyclic knock down parasites, it can be assumed to be a different protein and not a degradation product of LAP. Saturation experiments with recombinant, purified *Tb*LAP-A, Figure 4-8 B, identify the band at ~ 55 kDa as *Tb*LAP-A. The antibody recognises the purified *Tb*LAP-A loaded in spite of the saturation due to the high concentration on the blot.

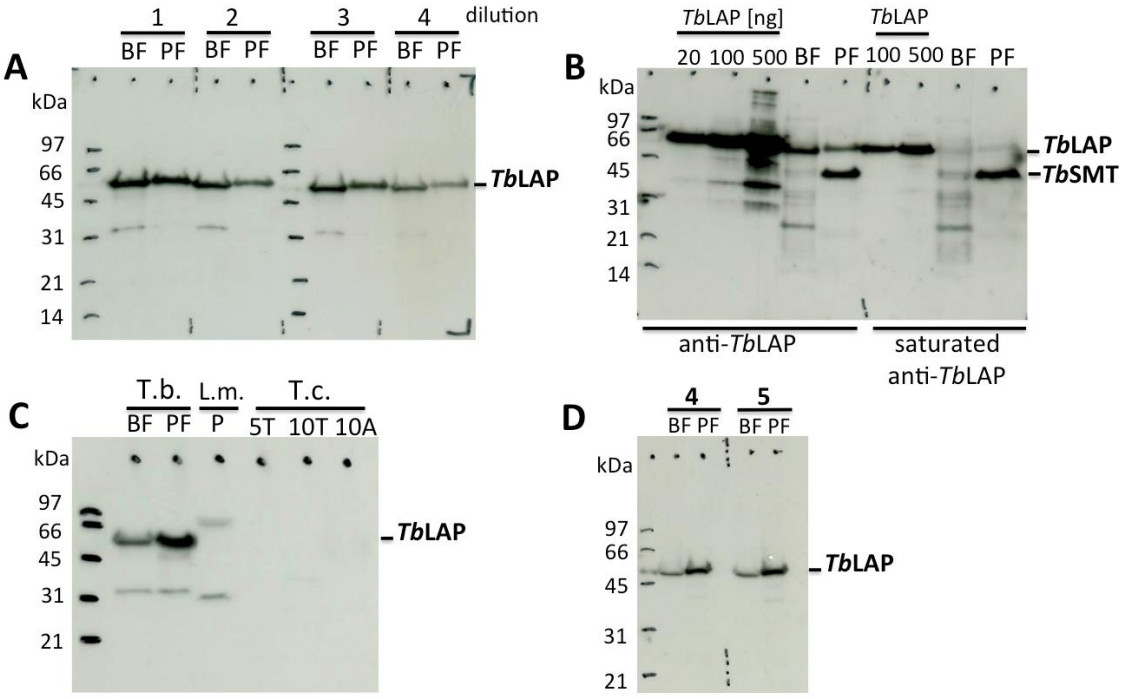


Figure 4-8: Validation of anti-*Tb*LAP-A by western blot analysis.

(A) Sensitivity of anti-*Tb*LAP-A: western blot analysis with anti-*Tb*LAP-A as primary antibody in dilutions 1:50 000 (1), 1:75 000 (2), 1:100 000 (3) or 1:200 000 (4); each lane contains 2.5 x 10^6 parasites (BF = bloodstream forms, PF = procyclics); the dotted line indicates where the membrane was cut after the blocking step. (B) Specificity of anti-*Tb*LAP-A: western blot analysis with saturated or untreated anti-*Tb*LAP-A in dilution 1:100 000 and anti-*Tb*SMT 1:15 000. The lanes contain either 50, 100 or 500 ng pure *Tb*LAP-A or 2.5 x 10^6 parasites; the dotted line indicates where the membrane was cut after the blocking step. (C) Cross-reactivity of anti-*Tb*LAP-A: western blot analysis with anti-*Tb*LAP-A (1:15 000) as primary antibody. The lanes contain 2.5 x 10^6 *T. brucei* (bloodstreams and procyclics), 5 x 10^6 *L. major* promastigotes (L.m. P), 5 x 10^6 or 10 x 10^6 *T. cruzi* trypomastigotes (5T and 10T respectively) or 10 x 10^6 *T. cruzi* amastigotes (with host cell background). (D) Analysis of affinity purified polyclonal anti-*Tb*LAP-A antibody: western blot analysis with antibody elution fractions 4 and 5 (from affinity purification) used in 1:100 000 dilution. Each lane contains 2.5 x 10^6 parasites.

Cross-reactivity with *L. major* LAP-A was shown, Figure 4-8 C, although the signal is much weaker, and could indicate poor recognition or a generally lower expression level of LAP-A. *L. major* also shows a lower molecular weight product being recognised, like that detected in *T. brucei* cell lysates. No cross-reactivity could be detected towards LAP-A in *T. cruzi* cell lysates although unfortunately the amount of protein loaded was very low as indicated by Ponceau red staining of the

membranes (data not shown), and therefore a certain level of cross-reactivity cannot be ruled out.

After affinity purification with purified recombinant *Tb*LAP-A, the elution fractions 4 and 5 contained the anti-*Tb*LAP-A antibody as shown by western blot against cell lysates of blood stream and procyclic *T. brucei* (Figure 4-8 D). At an antibody dilution of 1:100 000 the additional low molecular weight band was recognised only very weakly.

4.3.3 *Trypanosoma brucei brucei* cell lines, culture and transfections

For confirmation of the successful generation of the strains, the LAP-A levels in induced and un-induced cultures were determined by western blots using *Tb*LAP-A-specific antiserum. In bloodstream form parasites, strains with inducible LAP-A RNAi and overexpression, as well as strains overexpressing a LAP-A-cMyc fusion protein, were generated successfully. However, the generation of a LAP-A double knockout was not successful. Parasites exhibiting resistance to the drugs used for selection were attained, yet it is not possible to say at this point whether spontaneous resistance was generated, if the marker integrated into the wrong locus, and therefore did not replace the LAP-A gene, or if gene arrangements or duplications have occurred thus generating an additional copy of the LAP-A gene. All bloodstream form strains grew in HMI-9 with generation times of ~ 6 h.

In procyclic parasites, strains with LAP-A RNAi, double knockouts and cells overexpressing LAP-A and a LAP-A-cMyc fusion protein were generated. Interestingly, the double knockout was created using the same constructs as for the bloodstream form parasites, where they had not been successful. All the procyclic strains grew in SDM-79 with generation times of ~ 8 h.

4.3.4 Growth and adaptation to media

Growth variability was observed according to culture conditions and batches of media. Also, for culture of bloodstream forms in dialysed serum and changes in culture volumes a period of adaptation was required for obtaining constant growth rates (Figure 4-9).

Cells were adapted to different media and culture volumes (until they showed constant generation times), before RNAi or overexpression were induced and growth curves were recorded.

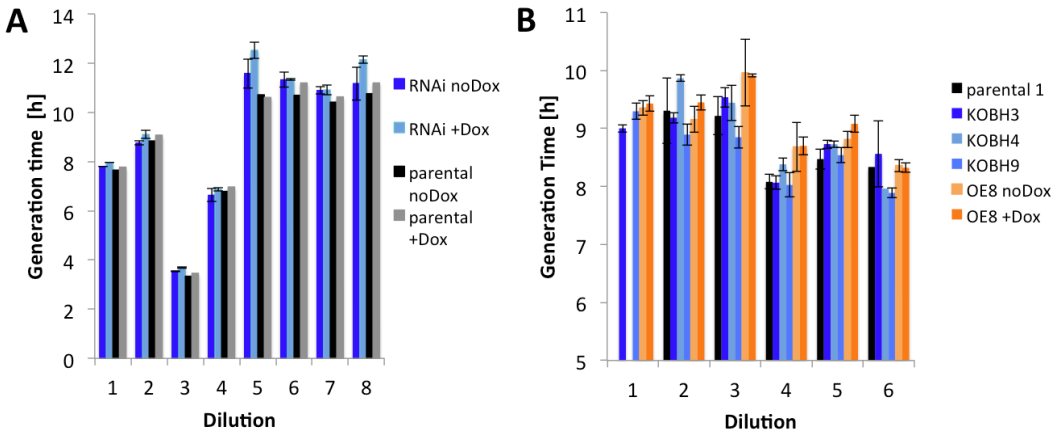


Figure 4-9: *T. brucei* growth adaptation.

Given are the generation times in hours with error bars indicating the standard deviation of three independent cultures. (A) Bloodstream parasite cultures (parental and RNAi) in HMI-9 media with dialysed FBS after transfer from HMI-9 with non-dialysed FBS. (B) Procyclic parasite cultures (parental, double knockout clones 3,4 and 9 (KOBH3, KOBH4, KOBH9) and overexpression (induced and un-induced)) in SDM-79 after transfer from 5 mL cultures in flasks into 3 mL cultures in 6-well cell culture plates.

While there is no glucose available in the tsetse fly and procyclic *T. brucei* use proline as the main energy source (Bringaud *et al.*, 2006), standard SDM-79 media contains high concentrations of glucose, which provides the parasites with an additional energy source, changing their energy metabolism (Lamour *et al.*, 2005, Coustou *et al.*, 2008). Glucose-free MEM media attempts to mimic the *in vivo* conditions more closely. Adaptation to MEM with dialysed FBS had to be performed step-wise with lower dilutions. Generally, the generation time observed in the

glucose-free minimal media (MEM with dialysed FBS) was considerably longer than in SDM-79 media for all procyclic strains.

The generation times observed for bloodstream form parasites were significantly longer in HMI-9 with dialysed serum (~ 11 h) than with non-dialysed FBS (6-7 h).

4.3.5 Cell cycle analysis: flow cytometry and fluorescence microscopy analysis

To determine the effect of LAP-A depletion and overexpression on the cell cycle, cells were analysed by flow cytometry and fluorescence microscopy.

Flow cytometry allows determination of the DNA content per parasite cell, and to distinguish parasites in G_1-, S- and G_2/M-phase by measuring the fluorescence signal of the stained DNA compared to the particle size and internal complexity. Cells in M-phase have the same DNA content as G_2-phase cells and therefore cannot be distinguished from these. Cells having lower DNA content, usually fragmented, are identified as the Sub-G_1 population. Cell fragments are not detected, because data were recorded of a gated cell population, which shows overall size and inner complexity of trypanosomal cells. Cells containing the same amount of DNA as one of the other cell cycle stages will be misinterpreted as such.

Using fluorescence microscopy of DAPI-stained cells, it is possible to gain more details about the cell cycle stage M and cytokinesis since cells with modifications in the kinetoplast or nuclear DNA content can be identified.

Both, flow cytometry cell cycle analysis and DAPI staining were employed for analysis of the effects by growth conditions and LAP-A depletion and overexpression.

4.3.6 RNAi mediated depletion of LAP-A in bloodstream forms of *T. brucei*

In bloodstream form *T. brucei* LAP-A knockdown over the whole time of the experiment was confirmed by western blots (Figure 4-10). LAP-A depletion approached 95 % and was maintained throughout the experiment.

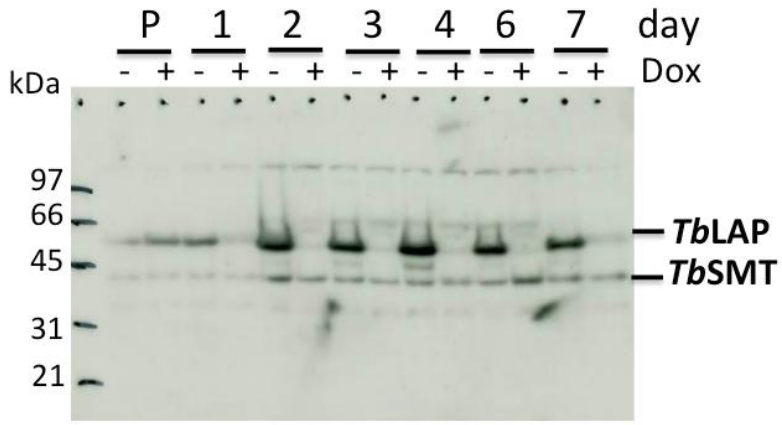


Figure 4-10: Western blot analysis of RNAi-mediated LAP-A depletion in bloodstream *T. brucei*.
Western blot analysis of LAP-A knockdown with RNAi over 7 days. P = parental cell line, all other lanes *Tb*BF RNAi clone, +/- doxycycline (Dox), cultures diluted and samples taken daily (days 1-7); Loading control = rabbit anti-*Tb*SMT 1:10 000 dilution. Each lane contains 2.5 x 10^6 parasites.

Bloodstream form parasites were cultured in HMI-9 media containing FBS or dialysed FBS and after adaptation and observation of constant growth the generation time was calculated and averaged over time (Figure 4-11 A). The average generation time of parental bloodstreams in the presence and absence of doxycycline was recorded, as well as the generation times of RNAi induced and un-induced parasites. The generation times of cultures containing doxycycline were marginally longer compared to cultures without it, independent of LAP-A depletion. In both media, the increase in generation time upon LAP-A depletion was identical to that of parental cells upon doxycycline addition and is considered not significant (evaluated using Welch's t-test). Additionally, the cell cycle was analysed *via* flow cytometry and fluorescence microscopy. Neither in HMI-9 with FBS (Figure 4-11 B)

nor with dialysed FBS (Figure 4-11 C), did cells display cell cycle defects upon LAP-A depletion and no increase in cell numbers with pathological nuclei or kinetoplast numbers was detected with fluorescent microscopy. Therefore, it can be concluded that LAP-A is not essential in bloodstream *T. brucei* under the culture conditions used.

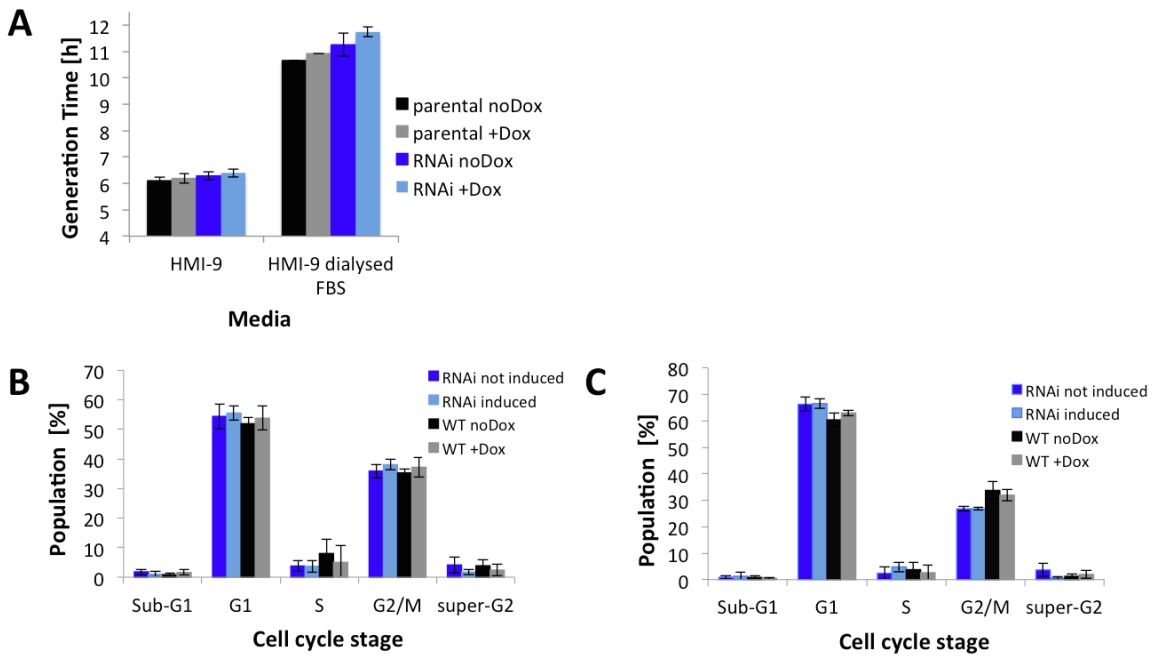


Figure 4-11: Growth and cell cycle analysis of RNAi-mediated LAP-A depletion in bloodstream *T. brucei*.

(A) Generation times calculated for bloodstream parasites with RNAi induced and un-induced, in HMI-9 media containing FBS or dialysed FBS. Given are the generation times in hours with error bars indicating the standard deviation of three independent cultures. Cell cycle analysis *via* flow cytometry of parental (= WT) and RNAi grown in HMI-9 containing FBS (B) and HMI-9 containing dialysed FBS (C).

4.3.7 RNAi mediated depletion of LAP-A and generation of LAP-A-dKO cells in procyclic forms of *T. brucei*

In procyclic LAP-A knockdown *via* RNAi as well as the generation of double knockout (LAP-A-dKO) strains was successful and confirmed by western blots (Figure 4-12). The RNAi-mediated depletion of LAP-A approached 100 % and could be maintained over more than 10 days.

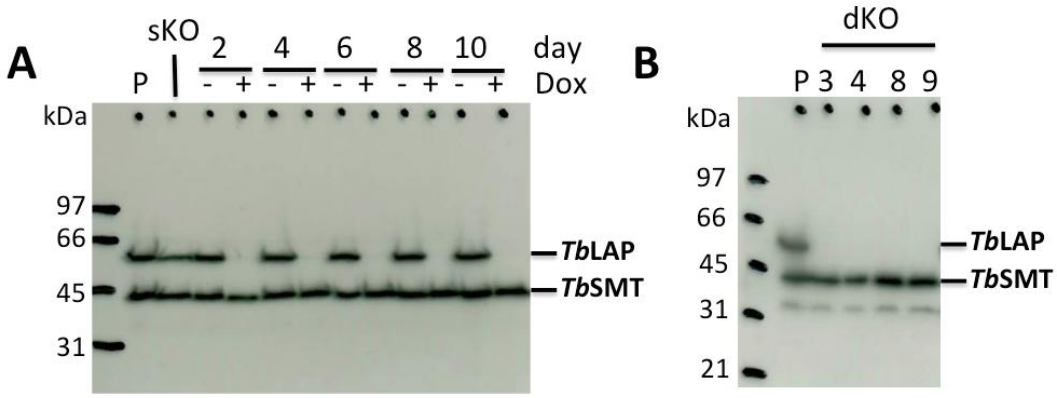


Figure 4-12: Western blot analysis of RNAi-mediated LAP-A depletion and dKO in procyclic *T. brucei*.
(A) Western blot analysis of single knockout (sKO) and LAP-A knock down *via* RNAi in comparison to parental (P) procyclic *T. brucei*. The lanes contain 2.5 x 10^6 parasites each, either induced or un-induced RNAi (+/- doxycycline) over 10 days with dilution of cultures every 2 days. Anti-*Tb*LAP-A antiserum from the second extraction was used in a dilution 1:15 000, anti-*Tb*SMT (as loading control) in 1:10 000 dilution. (B) Western blot analysis of double knockout (dKO) clones 3, 4, 8 and 9, compared to the parental (P) cells. Anti-*Tb*LAP-A was used in 1:15 000 dilution, anti-*Tb*SMT in 1:10 000.

Neither knockdown (data not shown) nor LAP-A-dKO showed a strong effect on the growth profile (> 10 % increase in generation time, Figure 4-13 A) or the cell cycle (Figure 4-13 B), either in standard SDM-79 or in MEM media with dialysed FBS. However, the increase of generation times of LAP-A-dKO clones 3 and 9 in MEM media were considered significant based on the Welch's t-test. No increase in cell numbers with pathological numbers of nuclei or kinetoplast was detected with fluorescent microscopy. Therefore it can be assumed that LAP-A is not essential in procyclic forms of *T. brucei* under the conditions tested.

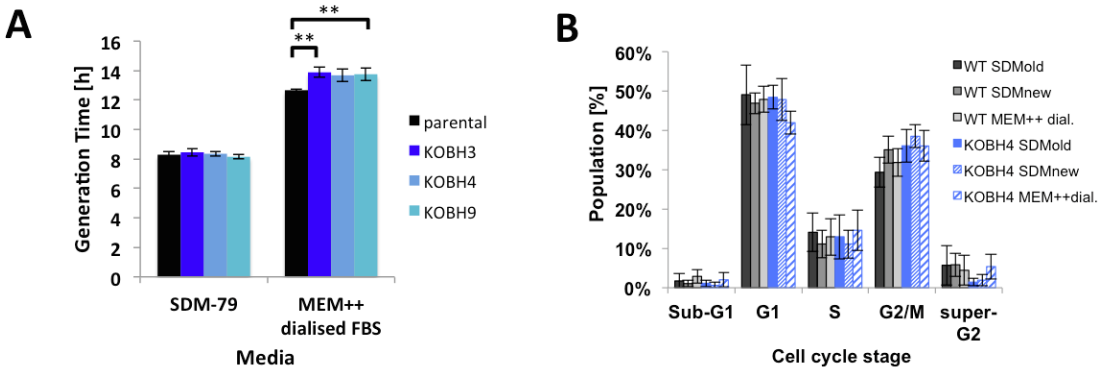


Figure 4-13: Growth curve and cell cycle analysis of RNAi and dKO in procyclic *T. brucei*.

(A) Generation times in hours of parental procyclics compared with dKO clones 3, 4 and 9 in SDM-79 (+ 10% FBS, + Haemin) and MEM media with dialysed FBS (+ Haemin). Given are the generation times in hours with error bars indicating the standard deviation of three independent cultures. ** indicates significant differences between samples ($p \leq 0.05$). (B) Cell cycle analysis *via* flow cytometry of procyclic parasites (parental and dKO clone 4) in different media (SDM-79 and MEM with dialysed FBS).

4.3.8 Overexpression of LAP-A in bloodstream and procyclic forms of *T. brucei*

Overexpression of LAP-A in bloodstream form and procyclic parasites was induced by addition of doxycycline and confirmed by western blot analysis (Figure 4-14).

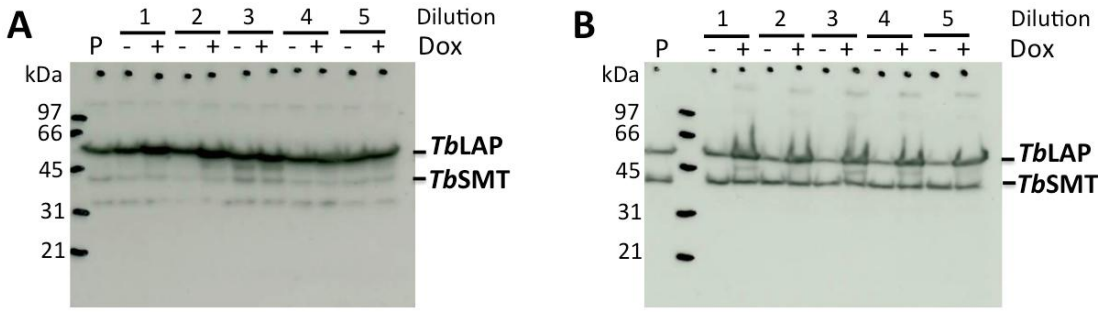


Figure 4-14: Western blot analysis of LAP-A overexpression in *T. brucei*.

Western blot analysis of (A) bloodstream and procyclic (B) *T. brucei* with LAP-A overexpression induced by addition of doxycycline over 10 days with dilutions of cultures and samples taken every 2 days. Uninduced cultures were treated identical to induced ones. The anti-*Tb*LAP-A from the second extraction was used in 1:15 000 dilution, anti-*Tb*SMT 1:10 000.

LAP-A overexpression in bloodstream parasites grown in HMI-9 media did not effect the growth rate of the parasites (Figure 4-15 A). LAP-A overexpressing procyclic parasites in SDM-79 media show a significant reduction of growth (longer generation time) even if not induced ($p \leq 0.05$, Figure 4-15 B). This might be caused by the selective marker (puromycin) or by leakiness of the construct, although the latter was not observed in comparison to the parental cell line on the western blot. Cell cycle analysis using flow cytometry and fluorescence microscopy did not reveal adverse effects of LAP-A overexpression on the cell cycle in these media.

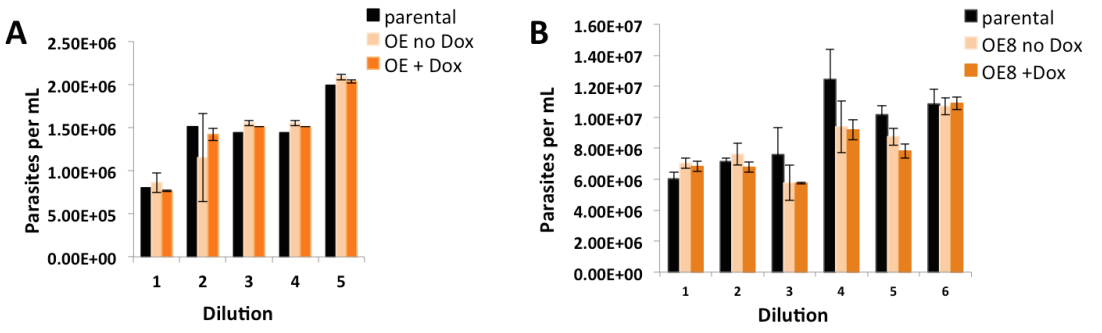


Figure 4-15: Growth of LAP-A overexpressing *T. brucei* in standard media.

Shown are the parasite counts per mL of LAP-A overexpressing bloodstream (A) and procyclic (B) parasites over the course of 10/12 days with dilution of the cultures every 2 days. Error bars indicate the standard deviation.

When procyclic forms were grown in minimal essential medium (MEM), LAP-A overexpression had no immediate adverse effects, but growth of induced and un-induced cultures was reduced significantly ($p \leq 0.05$) after subsequent dilutions, possibly due to the puromycin (Figure 4-16 A). Compared to parental cells no increase of LAP-A expression for un-induced cells could be detected by western blots. After six days of culture (two dilutions) the induced cells show a reduction in growth rate and after 12 days (three dilutions) a proportion of the LAP-A overexpressing cells died, shown most clearly by the increase of sub-G_1 cells in cell cycle analysis *via* flow cytometry (Figure 4-16 B).

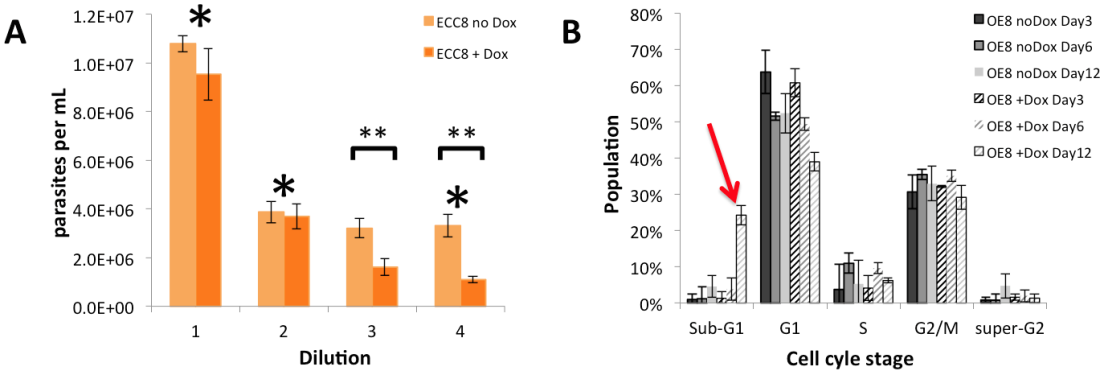


Figure 4-16: Growth and cell cycle analysis of LAP-A overexpression in procyclic *T. brucei* in MEM.

(A) Culture growth: cells counted every three days (Dilution 1 = day 3, dilution 2 = day 6, etc.) as parasites per mL before cultures were diluted again to 200 000 parasites mL^{-1}. The error bars indicate the standard deviation, * indicates the cultures that were harvested for cell cycle analysis by flow cytometry, ** indicates statistical significance between samples ($p < 0.05$). ECC8 = OE8 = clone name. (B) Cell cycle analysis by flow cytometry. The red arrow indicates the strong increase in Sub-G_1 = dead cells of the LAP-A overexpressing culture after 12 days (4 dilutions).

4.3.9 Amino acid limitation in procyclic forms of *T. brucei*

LAPs cleave amino acids, preferentially leucine, from the N-terminus of proteins or peptides and therefore play a role in protein recycling and turnover and the depletion or overexpression of an aminopeptidase is expected to have an increased effect on the cells wellbeing under suboptimal extracellular amino acid conditions. In the case of LAP-A depletion, the extracellular leucine levels might be particularly

important. To test the effect of leucine limitation on the growth and cell cycle of procyclic forms of *T. brucei*, cultures were grown in glucose-free MEM prepared without leucine and with dialysed FBS to make sure no peptides and undefined amino acids are present. Parental and LAP-A-dKO cells from cultures grown in MEM containing leucine (with dialysed FBS) were transferred to MEM without leucine and incubated at culture temperature. The parental cells continued to grow for approximately 24 h before they entered stationary phase (Figure 4-17 A) where they remained for ~ 2 days. In contrast, the LAP-A-dKO parasites continued to grow for a further 12 h (36 h in total) before entering stationary phase.

Both parental and LAP-A-dKO parasites showed the morphology typical for starvation with very elongated and thin cell body and reduced motility, analysed visually under the microscope.

The cell cycle stages of the parasites grown under leucine starvation were analysed by flow cytometry after 2, 3 and 4 days (Figure 4-17 B). The number of sub-G_1 cells in the LAP-A-dKO parasite population after 3 and 4 days was slightly increased compared to the parental cells despite the observation that the LAP-A-dKO cells continued to grow and enter stationary phase at a later time point. Similarly, LAP-A-overexpressing parasites showed an increased number of sub-G_1 cells than controls. In fluorescence microscopy, the number of cells with pathological morphology was increased in LAP-A-dKO and overexpressing cells. While 18.5 % of the parental cell population showed morphological defects, it was 23.2 % in LAP-A-dKO and 24.5 % in overexpressing cells.

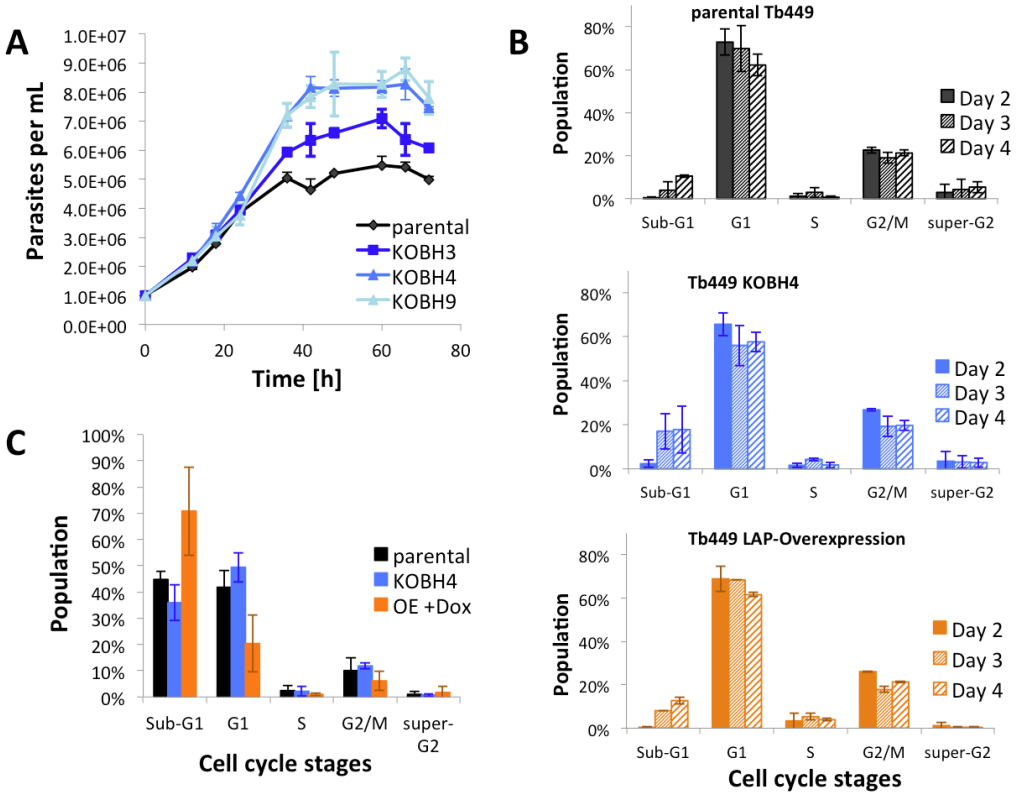


Figure 4-17: Growth and cell cycle analysis under amino acid limitation.

(A) Growth curve of procyclic parasites in MEM without leucine. KOBH3, 4 and 9 are different LAP-A double knockout strains. (B) Cell cycle analysis of procyclic parasites (parental, double knockout and overexpression) in MEM without leucine after 2, 3 and 4 days *via* flow cytometry. (C) Cell cycle analysis of procyclic parasites (parental, double knockout and overexpressing cells) grown in MEM without "non-essential amino acids" after 3 days *via* flow cytometry.

When procyclic parasites were grown in MEM in the absence of non-essential amino acids, cell division arrested (no increase in the number of cells could be observed) and the parasites died rapidly as indicated by cell cycle analysis with flow cytometry (Figure 4-17 C), where about half of the parasites were seen as sub-G_1 population after 3 days. LAP-A-overexpressing procyclics had a higher number of sub-G_1 cells than parental cells, which in turn showed sub-G_1 cells than the LAP-A-dKO. These findings were confirmed by fluorescence microscopy where 50.8 % of the parental, 42.3 % of the LAP-A-dKO and 59.0 % of the overexpressing parasites showed morphological defects.

4.3.10 Subcellular localisation of LAP-A in *T. brucei*

Bloodstream and procyclic forms of *T. brucei* were grown to mid-log phase and fixed on poly-lysine coated glass slides. Mitochondria were stained using red fluorescent MitoTracker and DNA in the nuclei and kinetoplasts were stained with DAPI. *Tb*LAP-A was visualised using the affinity purified rabbit anti-*Tb*LAP-A anti-serum as described in 4.2.5. Cells were imaged by confocal fluorescence microscopy.

Parental bloodstream form and procyclic *T. brucei* (Figure 4-18) show a clear *Tb*LAP-A staining that disappears in RNAi-mediated LAP-A knockdowns and LAP-A-dKO cells, thus confirming the result obtained by western blotting. cMyc-tagged LAP-A staining (with cMyc-specific antibody) exhibited a similar pattern to that obtained using the anti-*Tb*LAP-A antibody in the parental cell line, although in bloodstream form parasites overexpressing the cMyc-tagged LAP-A a distribution in foci was observed while the untagged LAP-A seems more homogeneous.

Co-localisation analysis using JACoP in ImageJ (Bolte & Cordelieres, 2006, Schneider *et al.*, 2012) gave some indication of partial co-localisation of *Tb*LAP-A with the nucleus and the mitochondrion in bloodstream forms and procyclics, when analysed with Li's approach (Li *et al.*, 2004), however, LAP-A's localisation in the nucleus is not supported when images are inspected manually. An example analysis of parental bloodstream form cells is shown in Figure 4-19. The distribution of Classical intensity correlation coefficient-based (ICCB) methods, like Pearson's coefficient, Costes' approach (Costes *et al.*, 2004) and Van Steensel's approach (van Steensel *et al.*, 1996), did not give conclusive results. This may be caused by the inherent heterogeneity of the LAP-A and MitoTracker signals, background noise and/or bleeding of the DAPI channel into FITC, as well as limitations of resolution in the z-axis of the samples and resulting signal bleeds from lower and higher planes.

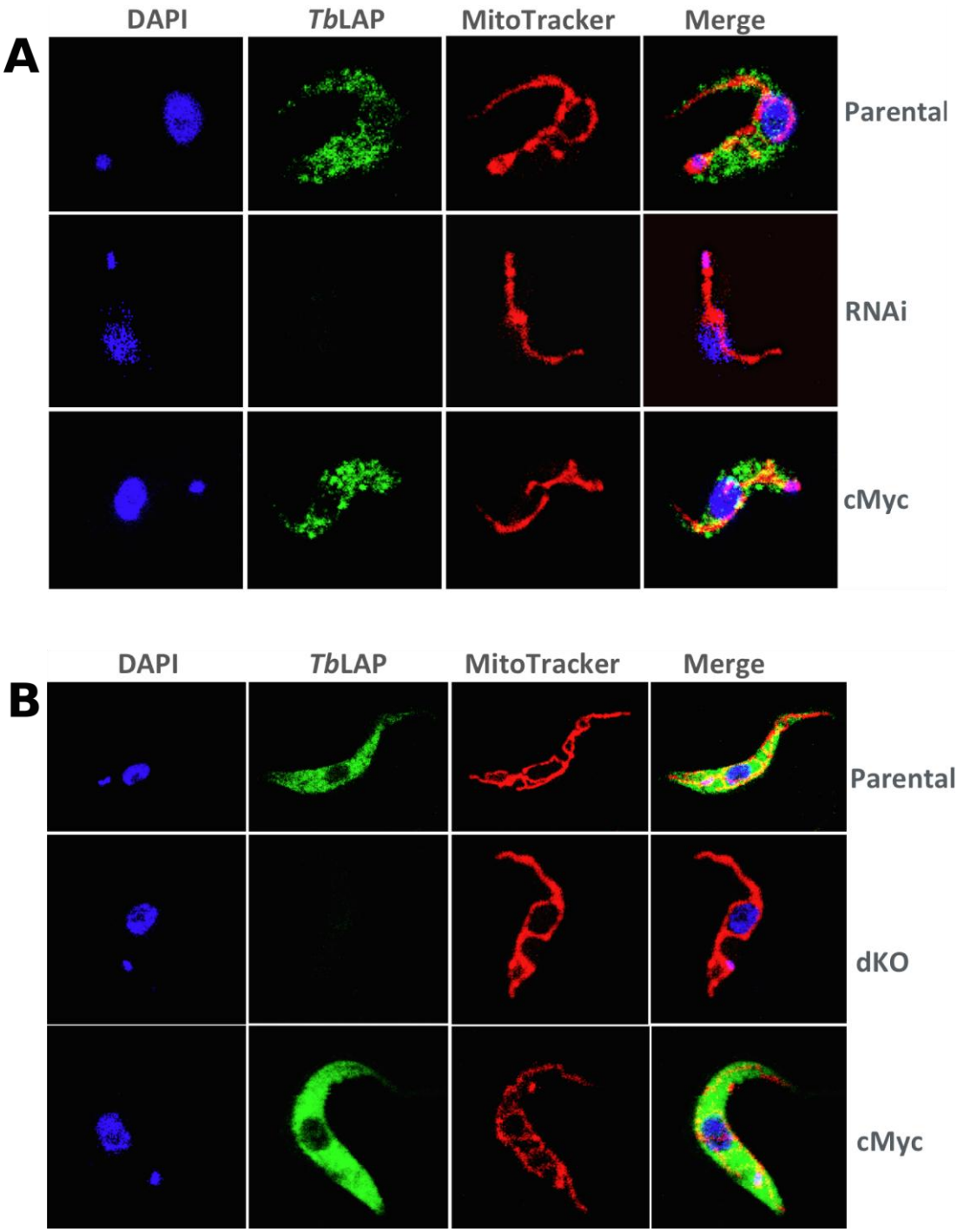


Figure 4-18: Subcellular localisation of LAP-A in *T. brucei*.

Shown are immunofluorescence images of bloodstream form (A) and procyclic (B) *T. brucei* taken at different fluorescence channels. Top rows show parental parasites, middle rows are RNAi (clone 1) of bloodstream form parasites with RNAi induced for 4 days before fixation of the cells, or LAP-A-dKO of procyclics, and the bottom rows show LAP-A-cMyc overexpressing cells, induced for 2 days before localisation. The Images were taken with identical microscope settings (and antibody dilution for parental and RNAi) for comparison of the different cell lines. The figure shows a 3D reconstruction of images taken with a Leica TCS SP5 confocal microscope at 63x magnification.

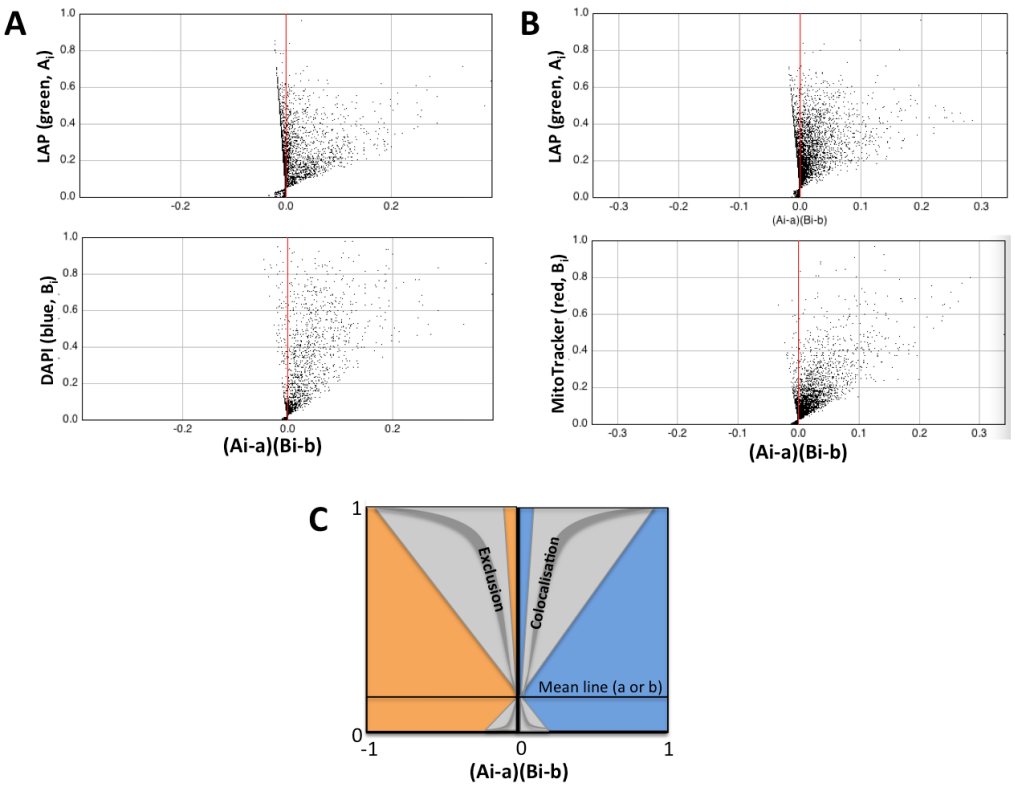


Figure 4-19: Colocalisation analysis with JACoP using Li's approach.

Shown are example analyses of colocalisation of LAP-A with DAPI (A) and MitoTracker (B) in parental bloodstream form cells. The value on the x-axis depends on the covariance of the channels that are compared, while the y-value reflects the intensities of the channel in question. The upper graphs show the graphs for the LAP-A channel, the lower ones for DAPI and MitoTracker respectively. (C) Schematic representation of the interpretation of the graphs, with dark grey indicating the distribution in case of complete colocalisation and exclusion and the light grey areas showing the expected distribution in case of colocalisation in presence of noise or partial colocalisation.

In conclusion, the immunofluorescence analysis indicates LAP-A localisation in the cytoplasm in both life cycle stages of the parasites, although the pattern of distribution shows differences. So LAP-A in the blood stream form seems more heterogeneous and for the cMyc-tagged LAP-A localised to foci, which might be an artefact of overexpression, while in procyclic parasites LAP-A distribution is more homogeneous. A localisation to the endoplasmic reticulum (ER) would cause a very similar pattern and would have to be confirmed by co-localisation studies using ER-specific and cytosolic marker proteins.

4.3.11 Crystallisation of the LAP-As

*Tb*LAP-A, *Tc*LAP-A and *Lm*LAP-A crystallised in various conditions. *Tb*LAP-A crystals in $P2_12_12_1$ with six chains in the asymmetric unit grew in citrate-containing conditions and comprise apo *Tb*LAP-A, missing the metal ions in the active site (Figure 4-20 A). *Tb*LAP-A complexes crystallised under many different conditions, preferentially in space group $P2_1$ with 6 or 12 chains per asymmetric unit (Figure 4-20 B). *Tc*LAP-A crystallised only in citrate-containing conditions, resulting in crystals in space group *R32* with one or both metal ions missing from the active site (Figure 4-20 C). No ligand complexes of this protein could be crystallised. *Lm*LAP-A crystallised in many different conditions forming hexagonal plate-like crystals, of which only two diffracted (Figure 4-20 D). One of these, grown in the presence of actinonin, was in space group *P321* and diffracted to 2.5 Å.

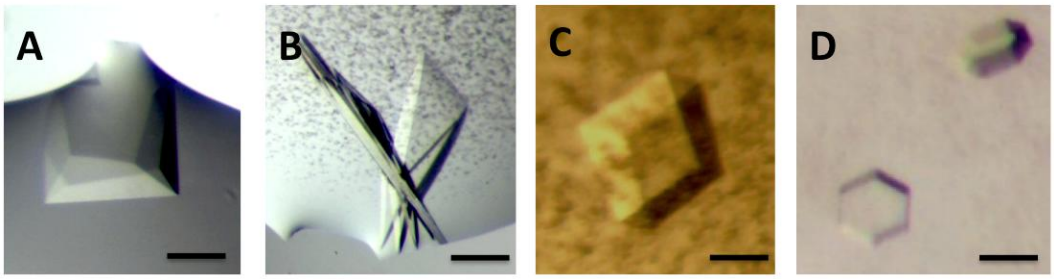


Figure 4-20: LAP-A crystals.

(A) apo *Tb*LAP-A crystal, black bar = ~200 µm; (B) *Tb*LAP-A-Mn-bestatin crystal, black bar = ~ 100 µm; (C) *Tc*LAP-A-Mn-citrate crystal, black bar = ~50 µm; (D) *Lm*LAP-A-Mn-actinonin crystal, black bar = ~ 50 µm.

4.3.12 TriTryp LAP-A structures

The structures of the TriTryp LAP-As were solved by molecular replacement and show a very high structure similarity to each other and LAPs of other species. The structure description will focus on the *Tb*LAP-A with differences in *Tc*LAP-A and *Lm*LAP-A indicated where relevant.

4.3.12.1 Overall and quaternary structure in the crystals

*Tb*LAP-A has two domains (N- and C-terminal) each with a mixed α/β structure (Figure 4-21). The smaller N-terminal domain (aa 2-197) shows a six-stranded β-sheet (β1, β6, β5, β2, β4, β3) flanked by five α-helices (α2 – α6). At the N-terminus there is an α-helix (α1), which interacts with the neighbouring protomer in the hexamer. The structures of other LAPs in the PDB do not have this N-terminal helix. The N-terminal domain is connected to the C-terminal domain by a long α-helix (α7). The highly conserved C-terminal domain (aa 215-521) is nearly identical to the structures of other LAPs in the PDB and is composed of a central eight-stranded β-sheet (β7, β8, β10, β9, β13, β17, β14, β15), which is surrounded by α-helices (α8 – α17) and a small β-sheet (β11, β12, β16).

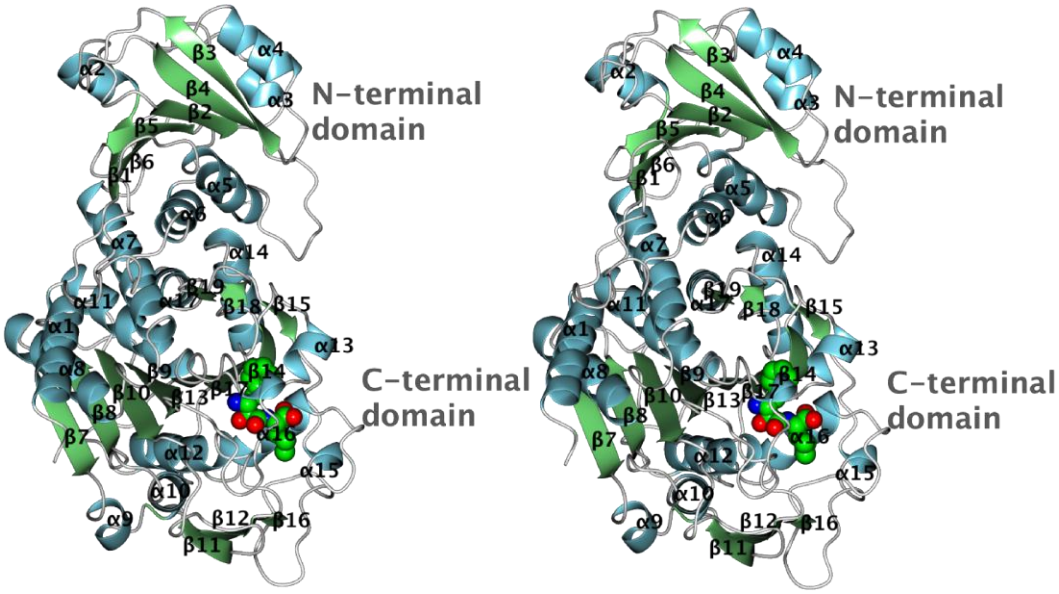


Figure 4-21: *Tb*LAP-A.

(A) *Tb*LAP-A in complex with Mn^{2+}, HCO_3^- and bestatin. Ligands are shown as spheres coloured by atom type, protein is shown as ribbon. The secondary structure elements are numbered from N- to C-terminus.

*Tb*LAP-A shows the typical hexameric architecture of the M17 metallopeptidases, where six LAP-A protomers form a dimer of trimers (Figure 4-22 A & B). The C-terminal domain, containing the active site, forms the core of the hexamer, packed around the central 3-fold axis with the active sites facing into the central cavity, which is secluded from the surrounding bulk solvent. This results in low accessibility for substrates as long as the assembly as hexamer persists.

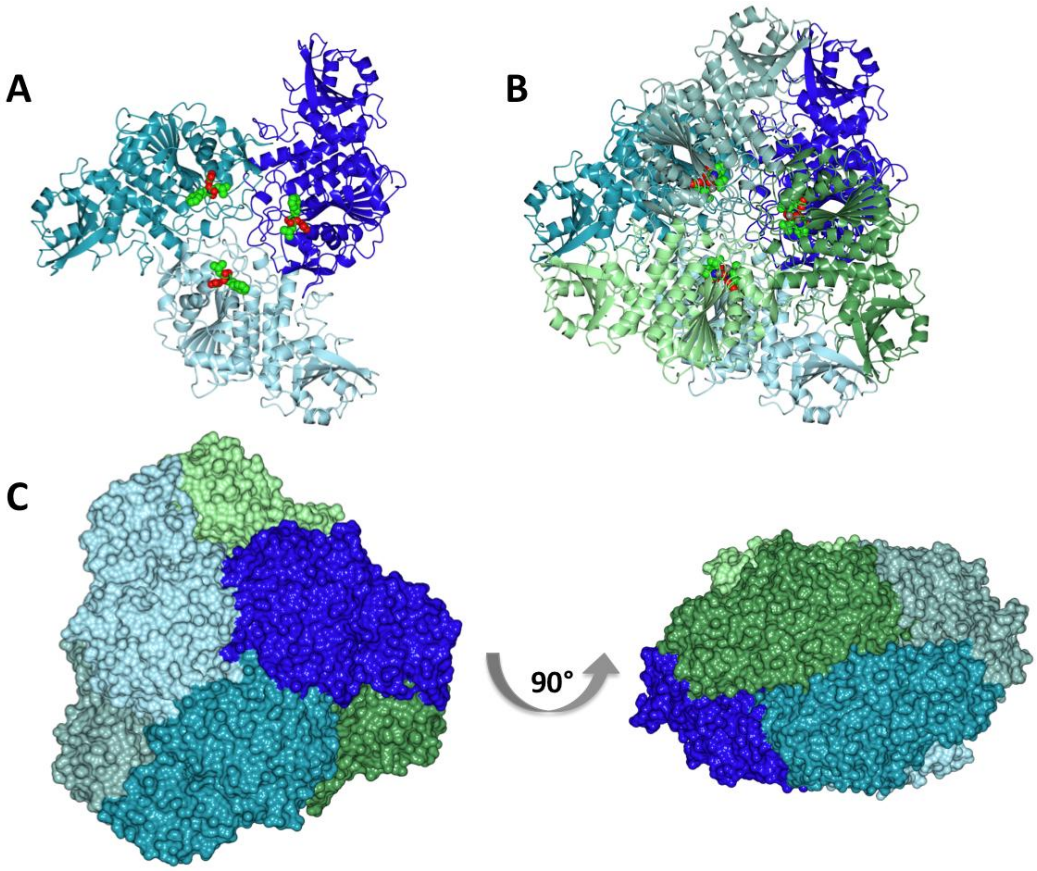


Figure 4-22: *Tb*LAP-A quaternary structure.

Two trimers (A) form a dimer of trimers (hexamer) (B). Protein chains are shown as ribbons coloured by chain. The ligands in the active site (bestatin, manganese and HCO_3^-) are shown as spheres coloured by atom type. (C) Surface view of the hexamer along the three-fold axis (left) and rotated about 90° (right), coloured by chain.

4.3.12.2 Comparison of the TriTryp LAP-A structures

The TriTryp LAP-As have identical core structures with overall RMSDs for *Tc*LAP-A and *Lm*LAP-A towards *Tb*LAP-A of 0.31 Å and 0.33 Å respectively with differences limited to loop regions and slight changes in the length of secondary structure elements (Figure 4-23). For example, the N-terminal helix and β-strands β3 and β4 in the N-terminal domain are slightly shorter than those in *Tb*LAP-A, while β5 and β6 are slightly longer. Some loop regions in the N-terminal domain (between α2 and β2, β3 and β4, α5 and β5, and α6 and β6) and in the C-terminal domain (between β8 and β9 and β10 and α11) show different conformations.

Structural differences of *Lm*LAP-A compared to *Tb*LAP-A (Figure 4-24) and *Tc*LAP-A are mainly in the loop regions. The secondary structure elements and their positions in *Lm*LAP-A are nearly identical to those in *Tb*LAP-A with the exception of the β1, β3 and β4, which are shortened and α2, which is replaced by a loop region. The N-terminus of *Lm*LAP-A is extended and forms a loop before the start of the α1-helix. The loop regions noted as different for *Tc*LAP-A are extended in *Lm*LAP-A and adopt different conformations. Additionally, the loop between β6 and α7 is extended, harbouring a small α-helix and helix α11 in the C-terminal domain is extended by one turn.

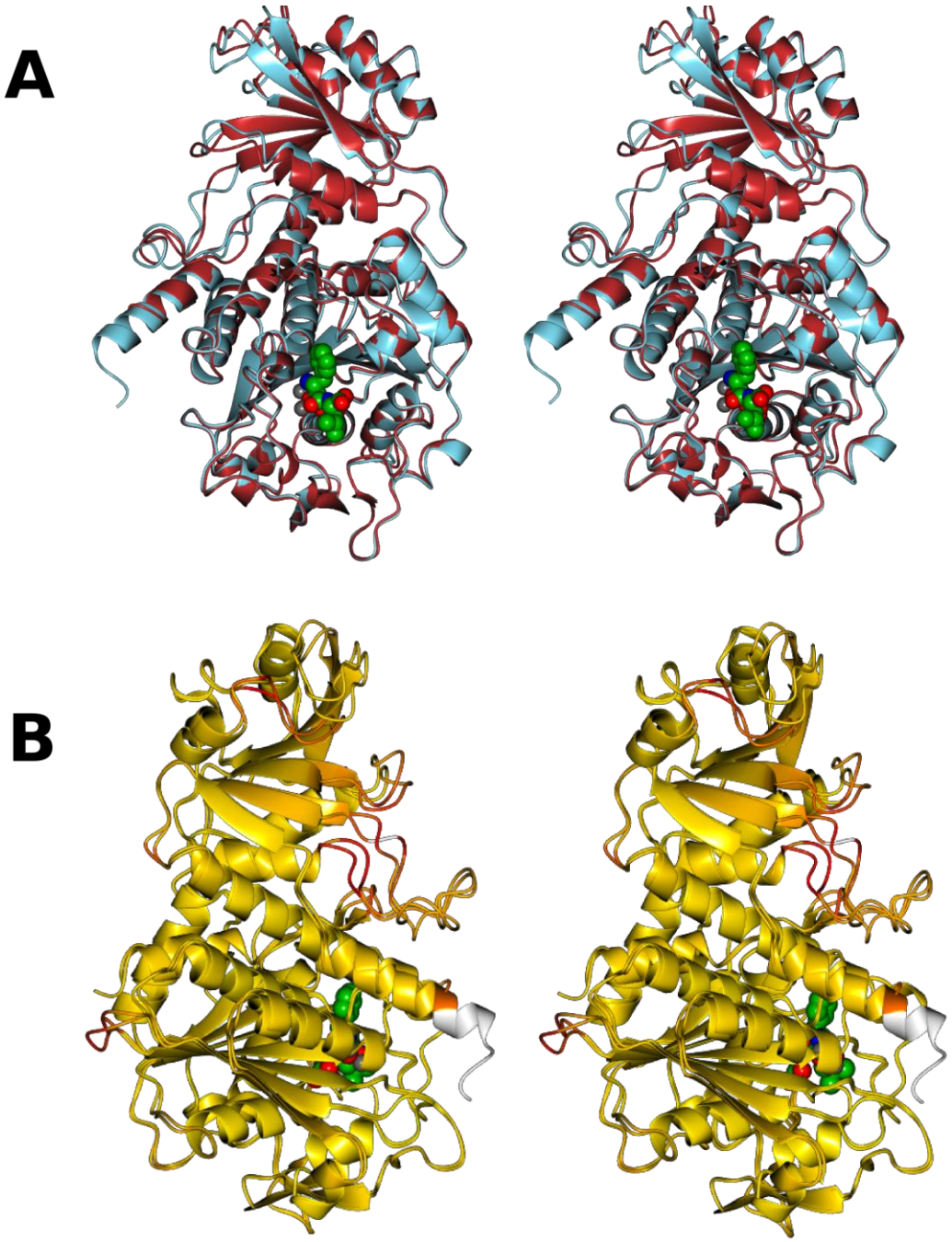


Figure 4-23: Comparison of *Tc*LAP-A with *Tb*LAP-A.

(A) Superposition of *Tc*LAP-A (crimson) and *Tb*LAP-A in complex with manganese, HCO_3^- and bestatin (ice blue). (B) Superposition of the same two proteins coloured by ProSMART (Nicholls *et al.*, 2012) from yellow (identical structures) to orange and red (differing parts of the structures) and white indicating the missing part of the shortened N-terminal helix in the *Tc*LAP-A structure. Superpositions in this and all subsequent figures were made using SSM (Krissinel & Henrick, 2004).

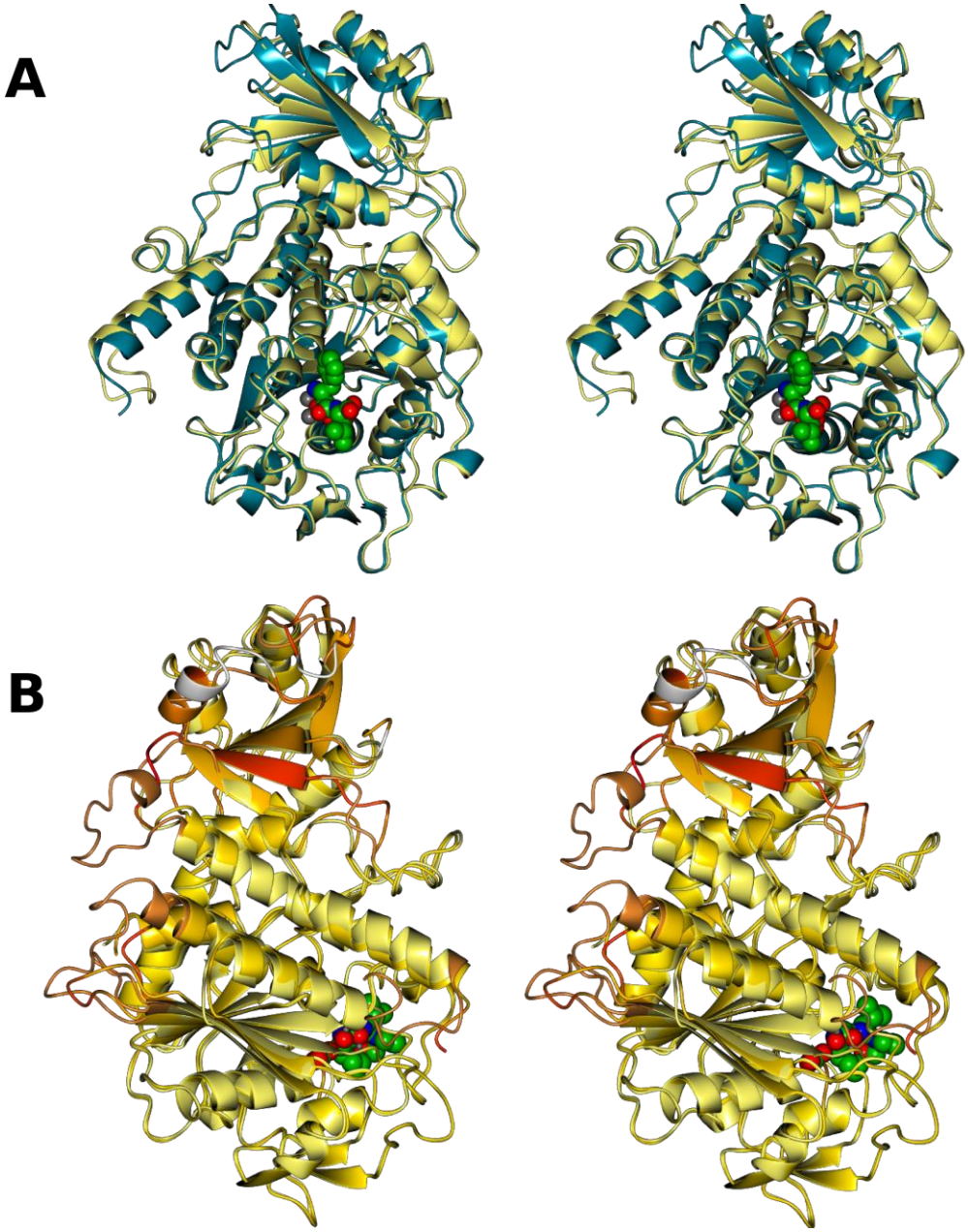


Figure 4-24: Comparison of *Lm*LAP-A with *Tb*LAP-A.

(A) Superposition of *Lm*LAP-A (lemon) and *Tb*LAP-A in complex with manganese, HCO_3^- and bestatin (dark cyan). (B) Superposition of the same two chains coloured by ProSMART (Nicholls *et al.*, 2012) from yellow (identical structures) to orange and red (differing parts of the structures) and white indicating the helix missing and loop region shortened in the *Lm*LAP-A structure. The proteins are shown as ribbons and the ligands as spheres coloured by atom type.

All of the TriTryp LAP-As form hexamers in the crystals (Figure 4-25) and analysis using PISA (Krissinel & Henrick, 2007) confirmed the hexamer as the likely oligomer in solution. All conformational differences seen between the TriTryp LAP-As are at the surface of the hexamer and away from the active site and are unlikely to influence oligomerisation or enzymatic activity.

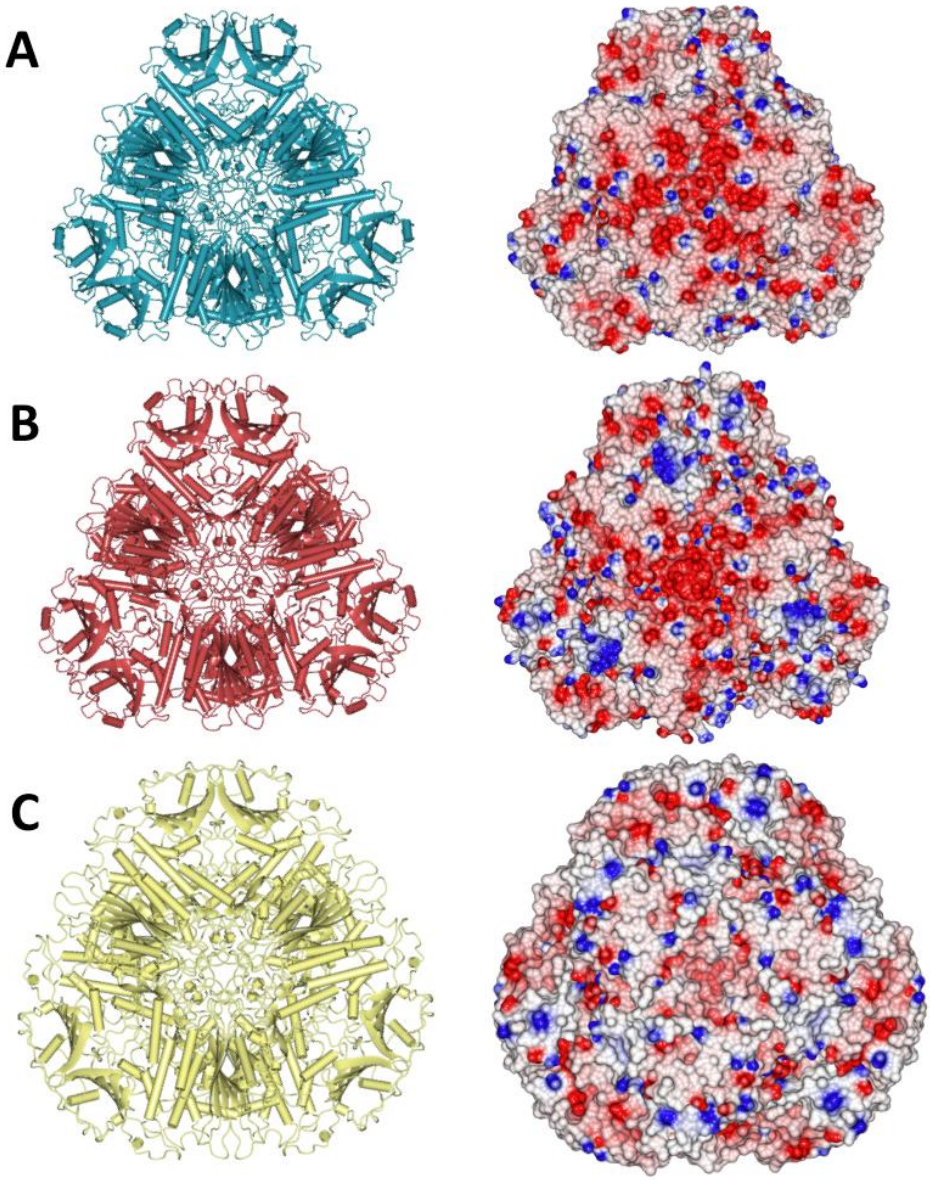


Figure 4-25: TriTryp LAP-A hexamers.

(A) *Tb*LAP-A; (B) *Tc*LAP-A; (C) *Lm*LAP-A. The LAP-As are shown in worm-and-tube representation (left) and the surfaces are shown coloured by electrostatic potential (right).

4.3.12.3 Metal coordination in the active site and apo LAP-A

The metal-coordinating amino acid residues in the TriTryp LAP-As are conserved and identical to those in previously solved LAP structures, Figure 4-26. The two metal binding sites are reported to have different binding affinities for the metal ions, with site 1 having lower affinity than site 2 (Carpenter & Vahl, 1973, Burley *et al.*, 1990). In *Tb*LAP-A, the carboxylate oxygen atoms of D294, E373 and the main chain and carboxylate oxygens of D371 coordinate the metal ion in site 1, and the carboxylate oxygens of D294, E373 and D312 and the ε-amino group of K289 coordinate that in site 2. The corresponding residue numbers in *Tc*LAP-A are K288, D293, D311, D370 and E372, and in *Lm*LAP-A are K299, D304, D322, D389 and E391.

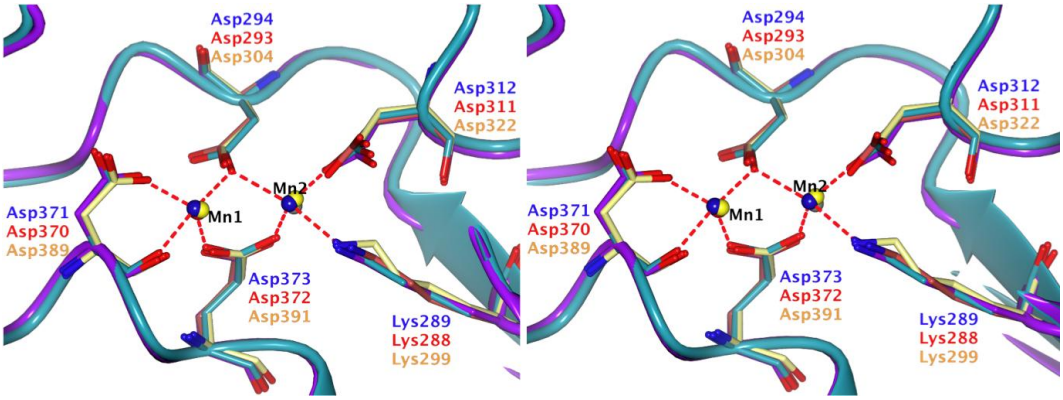


Figure 4-26: Superposition of the metal binding sites.

The main chains of *Tb*LAP-A (dark cyan) and *Pf*LAP (purple) are shown as ribbons, metal coordinating residues are shown as cylinders coloured by atom type with the C atoms coloured by model, *Tb*LAP-A = dark cyan, *Tc*LAP-A = crimson, *Lm*LAP-A = lemon, *Pf*LAP = purple. The manganese ions in the TriTryp LAP-A structures and the amino acid labelling are coloured by model with blue for *Tb*LAP-A, red for *Tc*LAP-A and yellow/orange for *Lm*LAP-A. The coordination contacts between amino acid residues and metals are shown as red dashed lines; an additional coordination partner is either a water molecule or ligand and not shown in this figure.

While the unliganded *Tb*LAP-A (crystallised in presence of leucine), *Tb*LAP-A complexes and *Lm*LAP-A structures have two metals bound in the active site, the *Tc*LAP-A structures show clear density for only one metal ion, in site 1, but no metal in site 2. However, the active site residues adopt the same conformations as if two metal ions were bound.

Most LAP-A crystals grew in the presence of manganese ions. However, LAPs are reported to be zinc-dependent enzymes and thus crystals of EDTA-treated *Tb*LAP-A were grown in the presence of $ZnCl_2$. However, XRF scans (Figure 4-27) on the resulting crystals indicated the presence of manganese ions and subsequent data collection at different wavelengths and calculation of anomalous difference maps confirmed the absence of Zn^{2+} ions from both metal binding sites, by the presence of anomalous difference peaks at a wavelength above the X-ray absorption edge of zinc (1.3052 Å), Figure 4-28.

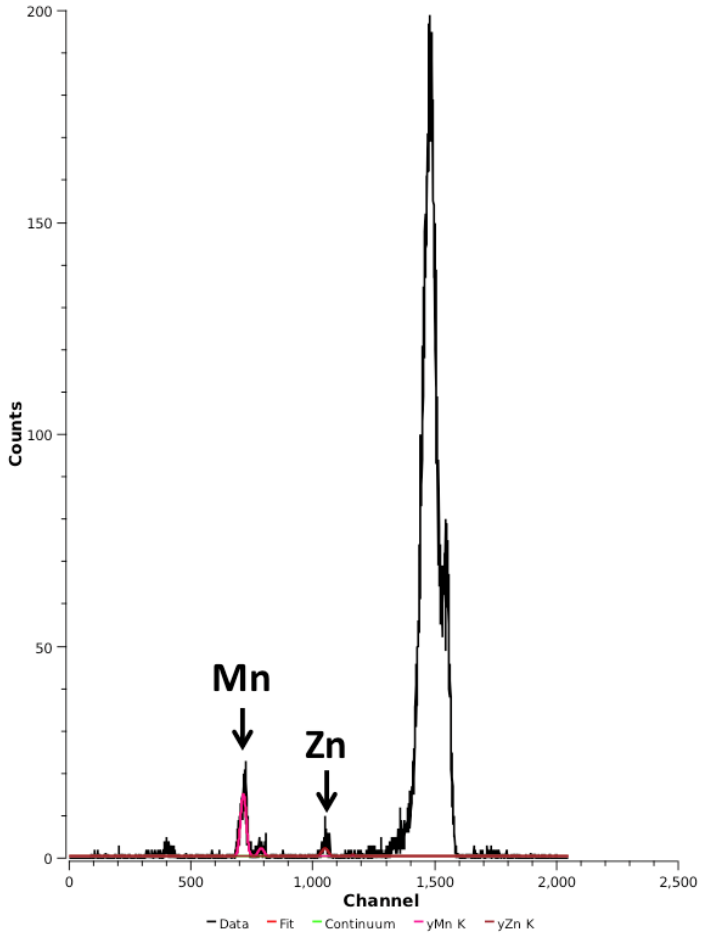


Figure 4-27: XRF scan of the *Tb*LAP-A-Mn crystal grown in presence of zinc.

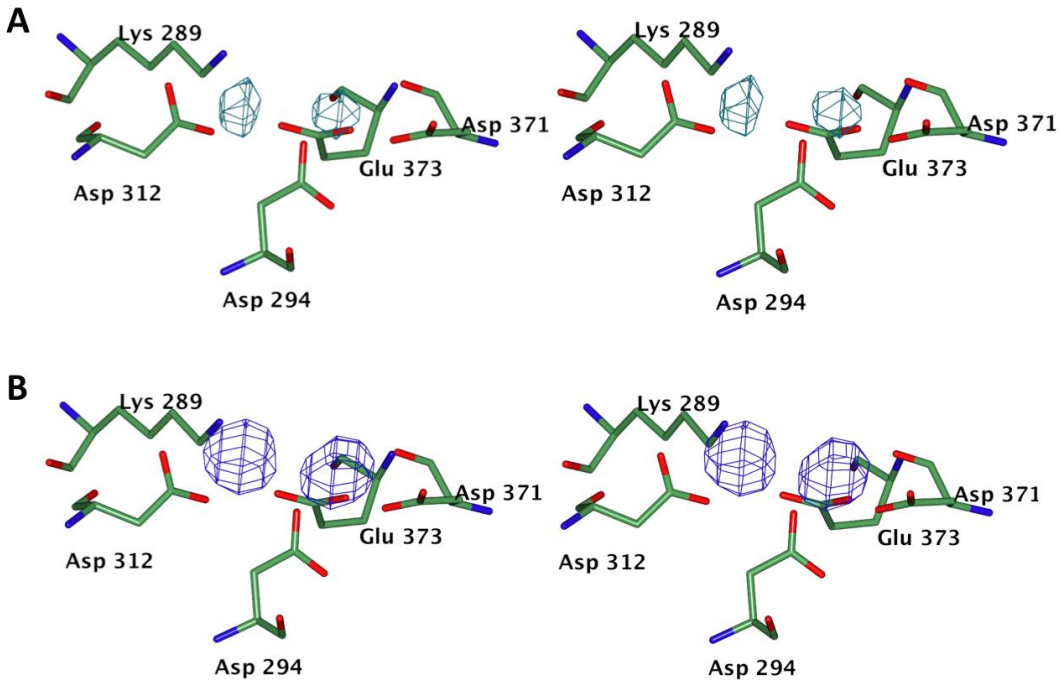


Figure 4-28: Anomalous difference density in the active site of *Tb*LAP-A.

(A) Anomalous difference map calculated from data collected at a wavelength of 0.9795 Å, contoured at 5σ.

(B) Anomalous difference map calculated from data collected at a wavelength of 1.3051 Å, contoured at 5σ.

The structures of LAP-As in their apo form, unliganded and in complex with inhibitors are practically identical. However, the active site loop (aa 293-305) of the apo *Tb*LAP-A is disordered and no electron density for bound metal ions could be detected, most likely due to the high concentration of citrate in the crystallisation buffer chelating the metal ions. Apo *Tc*LAP-A as well as the other LAP-A structures shows fully formed active sites independent of metal or ligand binding.

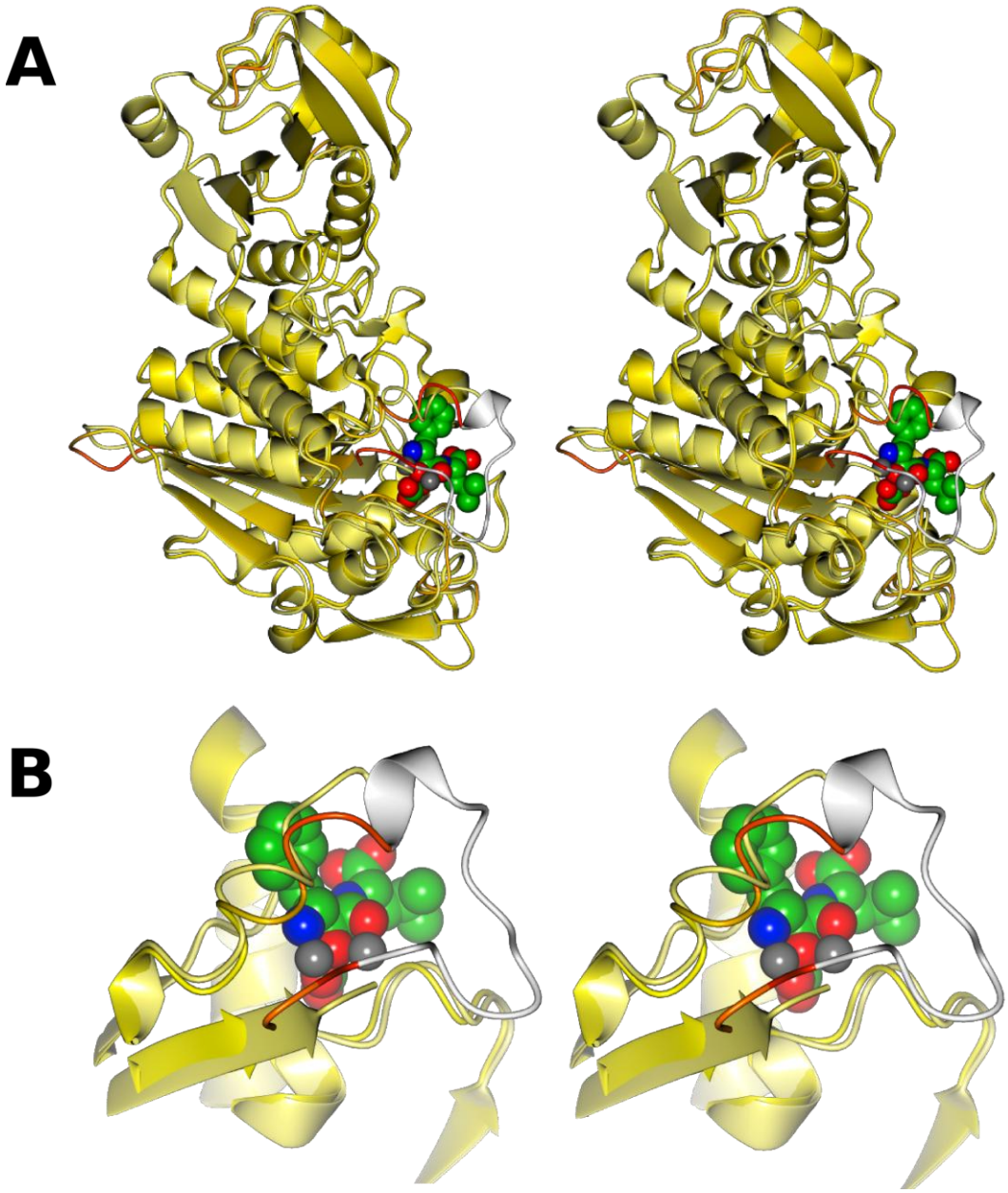


Figure 4-29: Superposition of apo *Tb*LAP-A and to *Tb*LAP-A-Mn-bestatin.

(A) Superposition of the A chains and (B) a close-up of the active site region of apo *Tb*LAP-A and *Tb*LAP-A in complex with manganese, HCO_3^- and bestatin with the protein in ribbon representation and the ligands as spheres coloured by atom type. Apo *Tb*LAP-A is coloured in lemon, while *Tb*LAP-A complex is coloured by ProSMART (Nicholls *et al.*, 2012) from yellow (identical to apo structure) to red (differing to apo structure) and white indicating the loop region missing in the apo structure.

4.3.12.4 The *Tb*LAP-A-bestatin complex

*Tb*LAP-A crystallised in the presence of the inhibitor bestatin. The electron density of the ligand was well defined in all 12 chains in the asymmetric unit and was modelled in the same conformation (Figure 4-30). The bestatin conformation mimics the *gem*-diolate intermediate of the peptide substrate in the process of its cleavage.

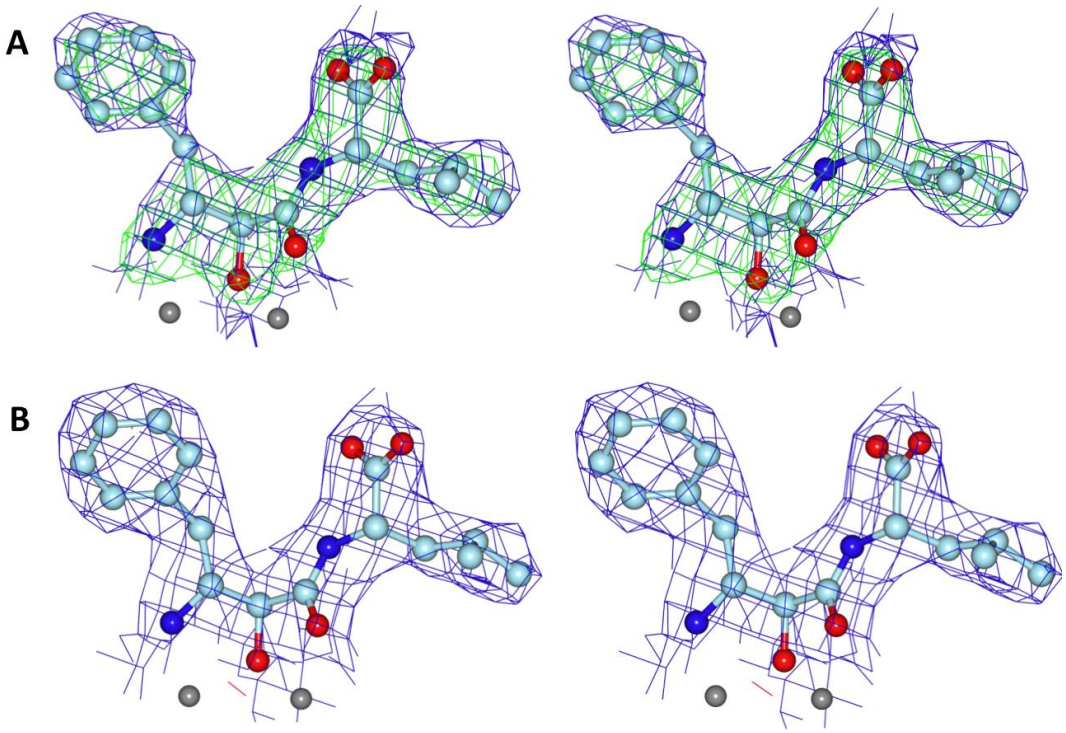


Figure 4-30: Electron density of bestatin bound in chain A of *Tb*LAP-A.
The weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c) were calculated using REFMAC5 (Murshudov *et al.*, 2011). Bestatin and Mn^{2+} are shown as ball-and-stick models coloured by atom type, mF_o-DF_c maps contoured at 3σ in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue. (A) Unbiased maps generated before ligand modelling. (B) Maps after refinement with bestatin.

The coordination of bestatin (Figure 4-31 A) is facilitated mainly through the Mn^{2+} ions and hydrogen bonding. The Mn^{2+} ions coordinate the nitrogen atom N2 and oxygen atoms O2 and O3. Additionally, N2 is coordinated through hydrogen bonds to the side chain oxygens of D294 and D312 as well as the main chain oxygen of

T401. The main chain oxygen of D371, and the side chains of K289 and D294 form hydrogen bonds to O2, while the side chain of K301 forms a hydrogen bond to O3. Furthermore the bound HCO_3^- forms a hydrogen bond to the O2 atom. While hydrophobic residues surround the phenyl ring of bestatin, its carboxyl group forms hydrogen bonds to G404 and R467. The bestatin binding in *Tb*LAP-A is identical to that in *Pf*LAP (Figure 4-31 B).

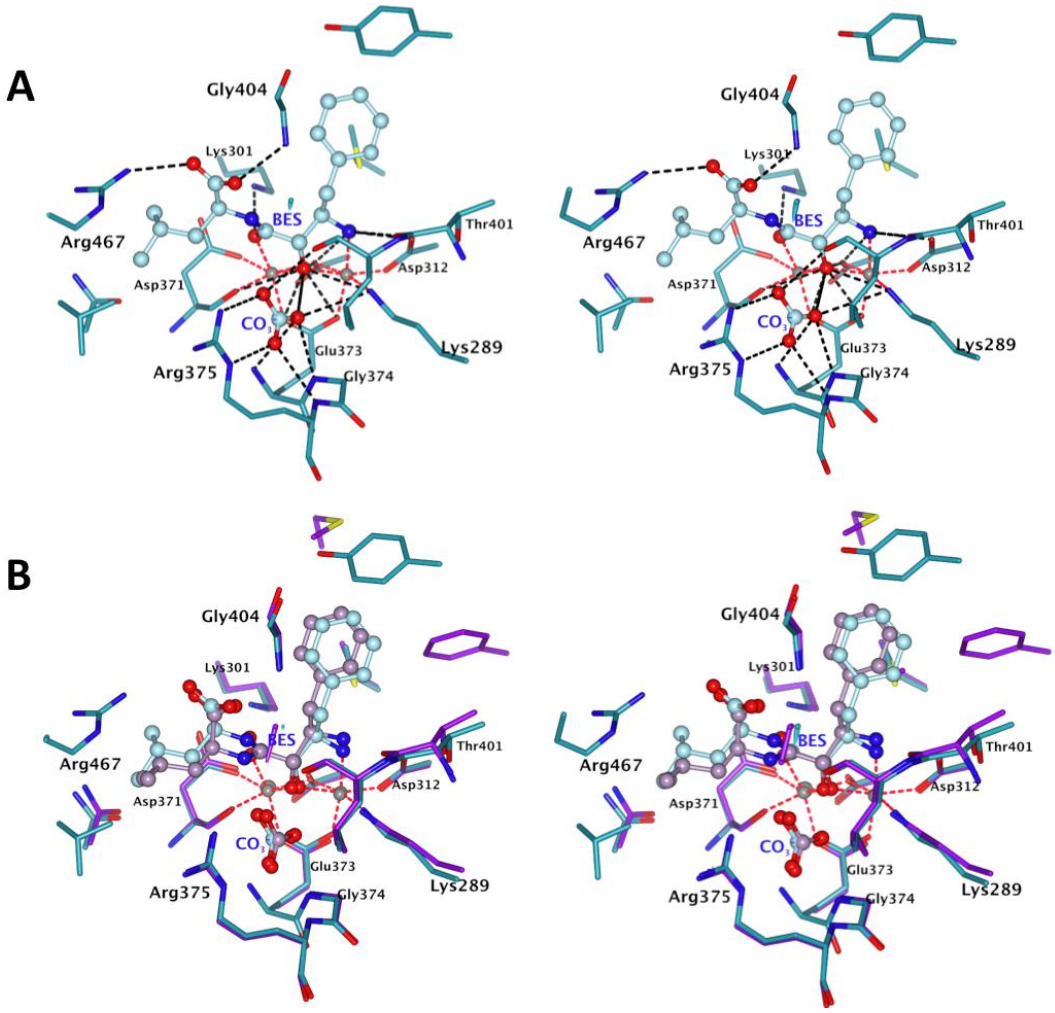


Figure 4-31: Bestatin binding in the active site.

Bestatin, Mn^{2+} and HCO_3^- are shown as ball-and-stick models and active site residues are shown as cylinders, coloured by atom type with C according to model, amino acid numbering according to *Tb*LAP-A. (A) Bestatin coordination in the active site of *Tb*LAP-A with hydrogen bonds as black dashed lines and metal coordination shown as red dashed lines. (B) Superposition of the active sites of the bestatin complex structures of *Tb*LAP-A (dark cyan with ice blue ligand) and *Pf*LAP (3KR4, purple with lilac ligand).

A schematic view of the bestatin binding in *Tb*LAP-A is shown in Figure 4-32.

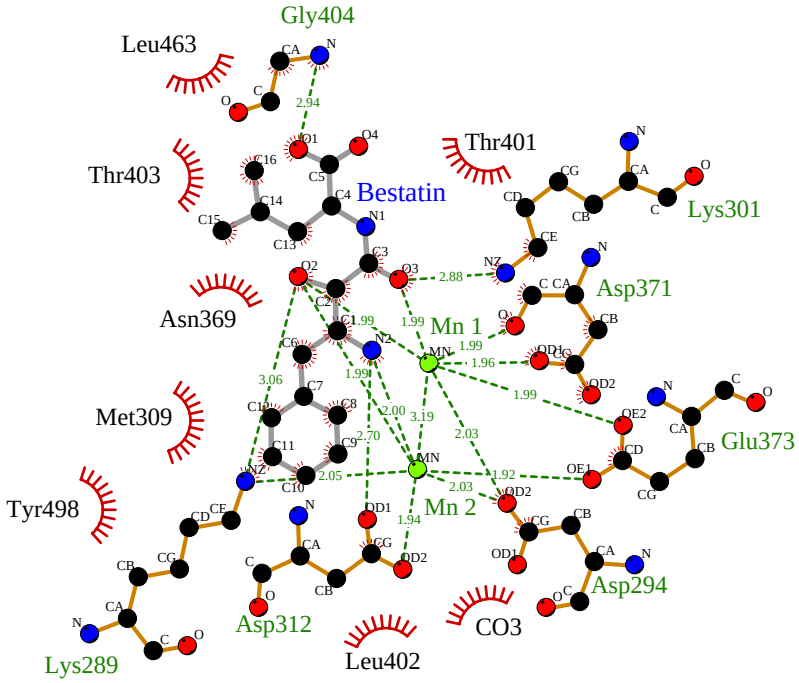


Figure 4-32: Schematic view of bestatin binding in the *Tb*LAP-A active site.

Shown are bestatin and the residues coordinating it *via* hydrogen bonds (green labels, hydrogen bonds as dashed lines), and hydrophobic interactions (black labels, red half circles). Ligand and hydrogen bonding residues are shown as ball and stick models coloured by atom type. Figure made with LigPlot+ (Laskowski & Swindells, 2011).

4.3.12.5 The actinonin complexes

For both *Tb*LAP-A and *Lm*LAP-A the electron density was sufficient to model the ligand in its approximate conformations (Figure 4-33 and Figure 4-34). However, its occupancy was below 100 % and the limited resolution and quality of the data sets does not allow accurate determination of the ligand's conformation.

The *Tb*LAP-A complex with actinonin had six molecules in the asymmetric unit, so that local NCS could be applied for the protein. The electron density for the actinonin on the other hand was only sufficient in four chains and varied slightly, so that the ligand was finally modelled in each chain individually. However, the electron density of two of the chains (D and F) was very poor, so that these were not included in analysis of the binding mode. The actinonin from two of the well-defined chains with the electron densities of chain A are shown in Figure 4-34. The

binding mode of actinonin in *Tb*LAP-A was identical in the four chains (Figure 4-35 B) and differed only slightly to the one in *Lm*LAP-A (Figure 4-35 C).

Actinonin in *Lm*LAP-A and *Tb*LAP-A is coordinated by the Mn^{2+} ions and hydrogen bonds. In both structures the two Mn^{2+} atoms coordinate the oxygen atoms O2 and O4 of the ligand. In *Lm*LAP-A the side chains of K299, D304, K311 and D389 form hydrogen bonds to O2 and O4, and the main chain oxygen of L420 and the main chain nitrogen of G422 form hydrogen bonds to N1 and O13 respectively. The N14 atom of actinonin coordinates a water molecule. The corresponding residues in *Tb*LAP-A are K289, D294, K301, D371, L402 and G404 and additionally, A405 forms a hydrogen bond to O27 (in four out of six chains). Because of the poor data quality no water molecules were modelled in the *Tb*LAP-A structure with the exception of those replacing the HCO_3^-. Superposition of the active sites including the actinonin molecules (Figure 4-35 C) reveals the high similarity of the conformations actinonin adopts, its coordination involving the same amino acid residues. The only difference can be observed in the hydroxymethyl-pyrrolidine moiety, which forms a hydrogen bond in *Tb*LAP-A but not in *Lm*LAP-A where it is bent away from the corresponding residue (A424).

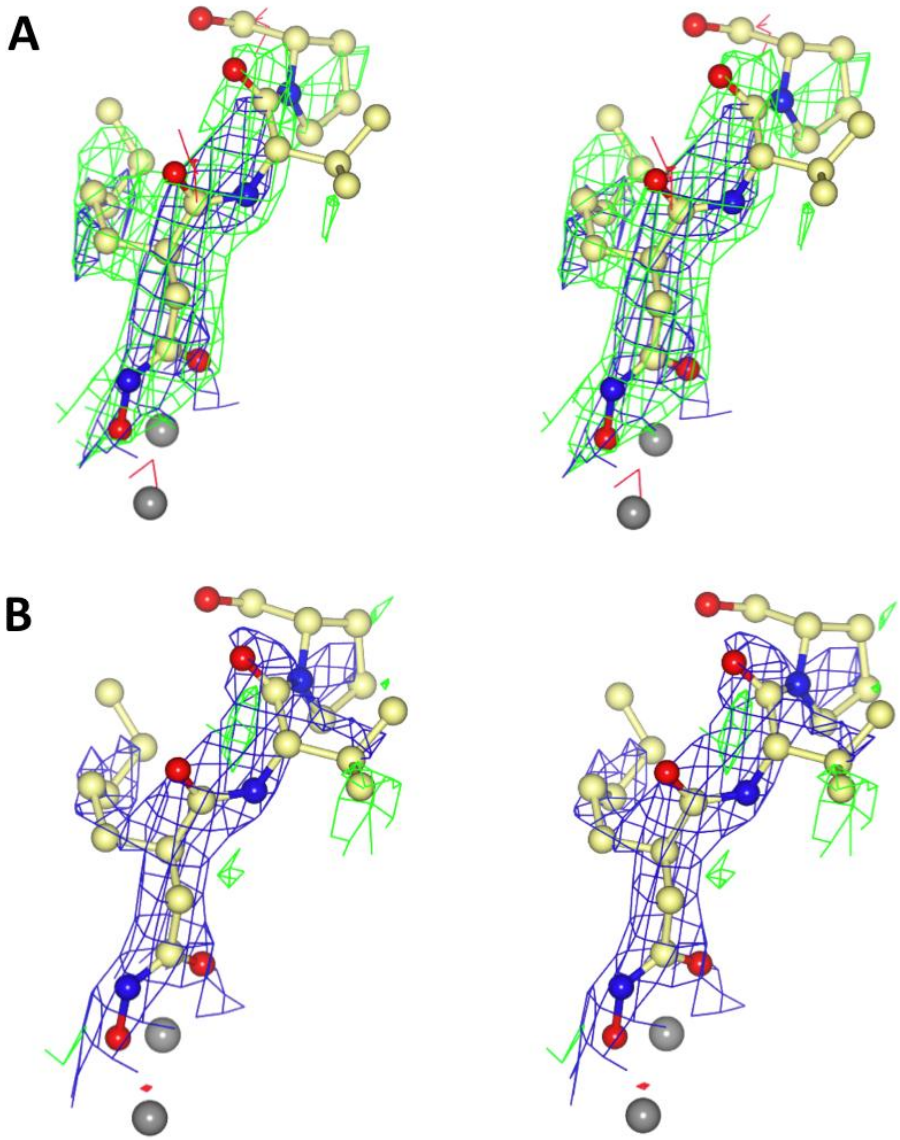


Figure 4-33: Electron density of actinonin bound in *Lm*LAP-A.

The weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c) were calculated using REFMAC5 (Murshudov *et al.*, 2011). (A) Unbiased maps generated before ligand modelling. (B) Maps after refinement with actinonin. Actinonin and Mn^{2+} are shown as ball-and-stick models coloured by atom type, mF_o-DF_c maps contoured at 2σ for (A) and 3σ in (B), in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue.

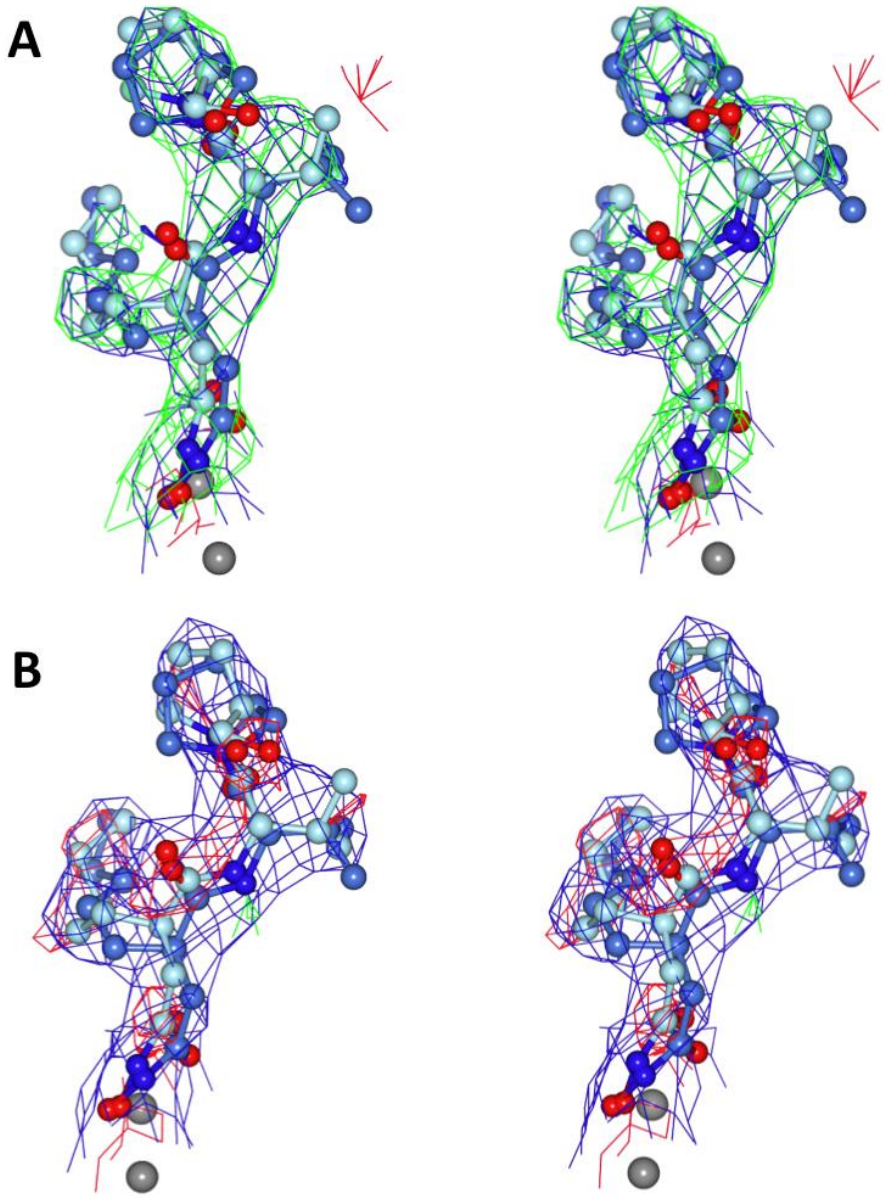


Figure 4-34: Electron density of actinonin bound in *Tb*LAP-A.

The weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c) were calculated using REFMAC5 (Murshudov *et al.*, 2011). Shown are the actinonin molecules of chain A and E representing the 6 molecules in the asymmetric unit superposed. (A) Unbiased maps generated before ligand modelling. (B) Maps after refinement with actinonin. Actinonin and Mn^{2+} are shown as ball-and-stick models coloured by atom type, mF_o-DF_c maps contoured at 2σ for (A) and 3σ in (B), in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue.

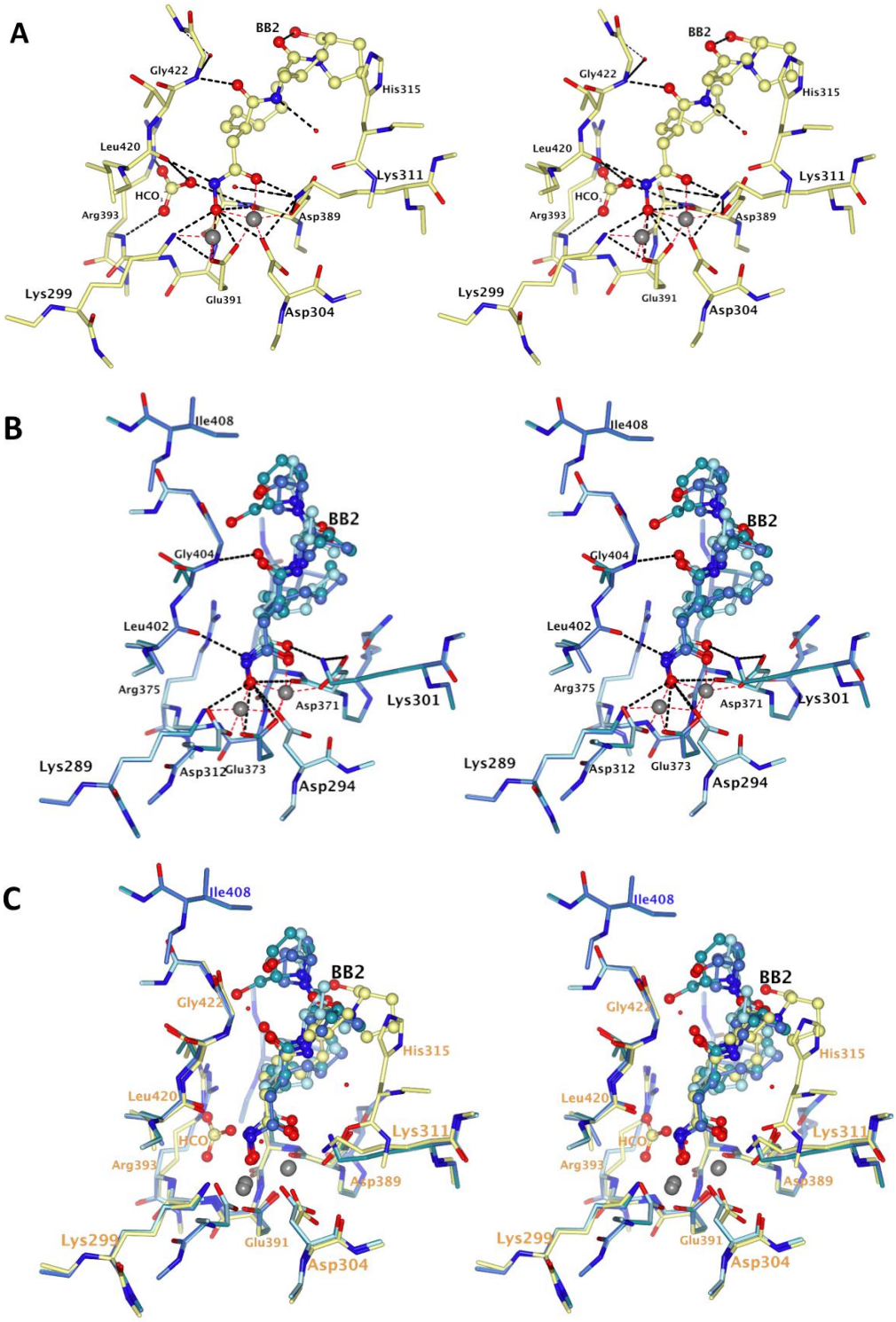


Figure 4-35: Actinonin binding sites in *Lm*LAP-A and *Tb*LAP-A.

(A) *Lm*LAP-A active site with bound actinonin. (B) *Tb*LAP-A active site with bound actinonin. (C) Superposition of the active sites of *Lm*LAP-A and *Tb*LAP-A with active site residues numbered after *Lm*LAP-A (orange) and *Tb*LAP-A (blue). Actinonin (BB2), Mn^{2+} and HCO_3^- are shown as ball-and-stick models and active site residues are shown as cylinders, coloured by atom type with C according to model, hydrogen bonds as black dashed lines and metal coordination shown as red dashed lines.

The actinonin interactions in *Tb*LAP-A and *Lm*LAP-A are plotted in Figure 4-36.

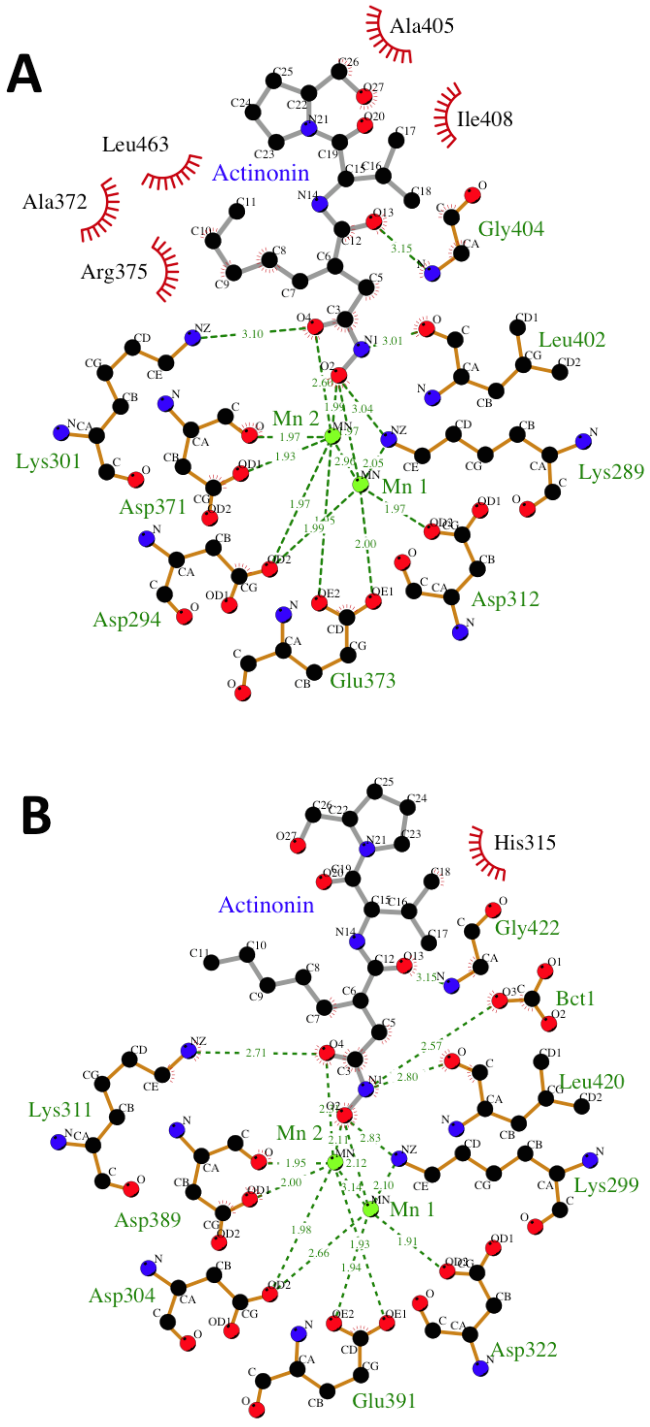


Figure 4-36: Schematic view of actinonin binding in the active sites of *Tb*LAP-A and *Lm*LAP-A.

(A) *Tb*LAP-A, (B) *Lm*LAP-A. Shown are actinonin and the residues coordinating it *via* hydrogen bonds (green labels, hydrogen bonds as dashed lines), and hydrophobic interactions (black labels, red half circles). Ligand and hydrogen bonding residues are shown as ball and stick models coloured by atom type. Figure made with LigPlot+ (Laskowski & Swindells, 2011).

4.3.12.6 *Tc*LAP-A in complex with citrate

*Tc*LAP-A crystallised only in citrate containing conditions, which led to the absence of one or both of the Mn^{2+} ions from the active site and no electron density for the inhibitors. However, the structure obtained from co-crystallisation with the inhibitor amastatin showed clear density for one partially occupied Mn^{2+} ion and a citrate molecule bound in the active site, but no amastatin (Figure 4-37 A & B). The HCO_3^- is replaced by a SO_4^{2-}.

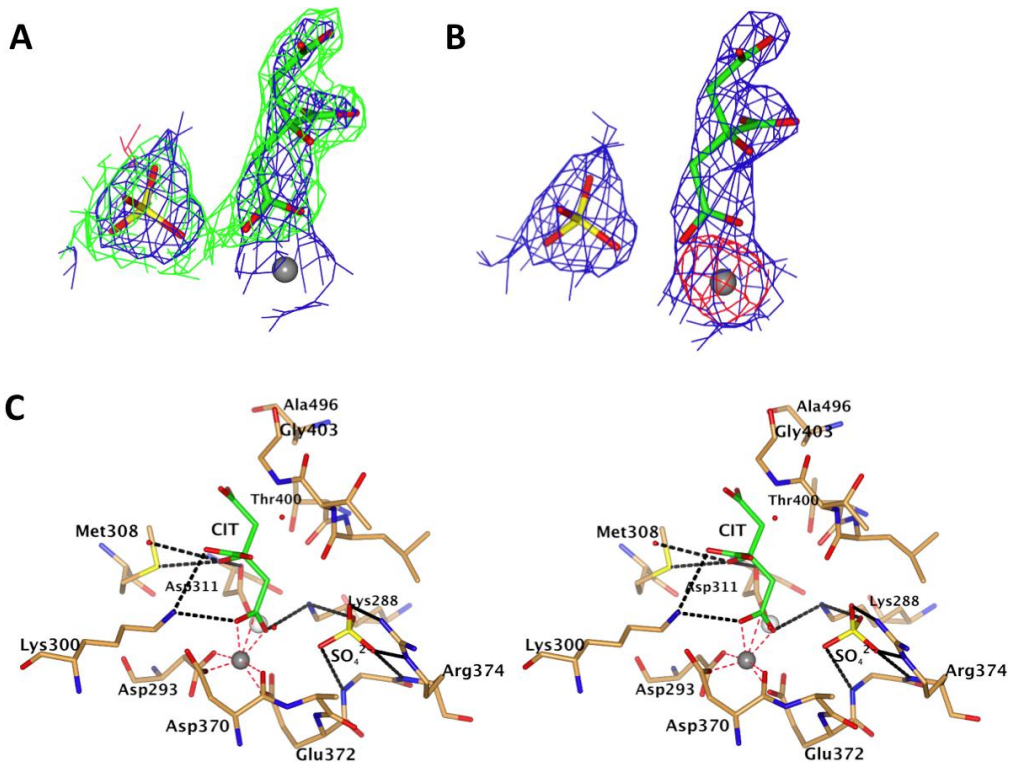


Figure 4-37: Citrate binding in *Tc*LAP-A.

Shown are the weighted electron density maps ($2mF_o$-DF_c) and difference maps (mF_o-DF_c), calculated using REFMAC5 (Murshudov *et al.*, 2011): (A) Unbiased maps generated before ligand modelling (but with the Mn^{2+}). (B) Maps after refinement with citrate (CIT) and SO_4^{2-}. CIT and SO_4^{2-} shown as cylinder coloured by atom type, Mn^{2+} as grey sphere, mF_o-DF_c maps contoured at 2σ for (A) and 3σ in (B), in green (positive) and red (negative), the $2mF_o$-DF_c maps are contoured at 1σ and drawn in blue. (C) Mn^{2+}, SO_4^{2-} and CIT coordination in the active site. CIT, SO_4^{2-} and coordinating side chains shown as cylinders coloured by atom type with the side chain C atoms in orange Mn^{2+} as grey sphere, the white sphere indicates the location of the empty metal binding site. The metal coordination and hydrogen bonds are drawn as dashed lines in red and black respectively.

The citrate molecule is coordinated by the Mn^{2+} and hydrogen bonding (Figure 4-37 C), with Mn^{2+} coordinating the oxygen atoms O3 and O4. While K288 forms a hydrogen bond to O4, K300 forms hydrogen bonds with O3 and O5. Oxygen atom O7 from the citrate forms hydrogen bonds to D311, M308 and a water molecule. Main and side chains or R374, G373 and K288 coordinate the SO_4^{2-} molecule.

4.3.13 Oligomeric State in solution

The TriTryp LAP-As crystallised as hexamers, like in other LAPs structures in the PDB. To confirm the hexamer as the oligomeric state in solution, SEC-MALLS analyses of all three LAP-As were carried out at a range of protein concentrations. The expected molecular masses expected for the different oligomeric states are summarised in Table 4-20 and were used to annotate the elution peaks obtained from the SEC-MALLS data allowing for an error of ± 10 %. An overlay of the differential refractive index traces at a protein concentration of 1.0 mg mL^{-1} with annotated peaks according to the oligomeric states corresponding to the molecular masses is shown in Figure 4-38.

Table 4-20: Expected molecular masses for different LAP-A oligomers.

	Monomer [kDa]	Dimer [kDa]	Trimer [kDa]	Hexamer [kDa]
*Tb*LAP-A	55.4	110.8	166.2	332.4
*Tc*LAP-A	55.9	111.8	167.7	335.4
*Lm*LAP-A	57.1	114.2	171.3	342.6

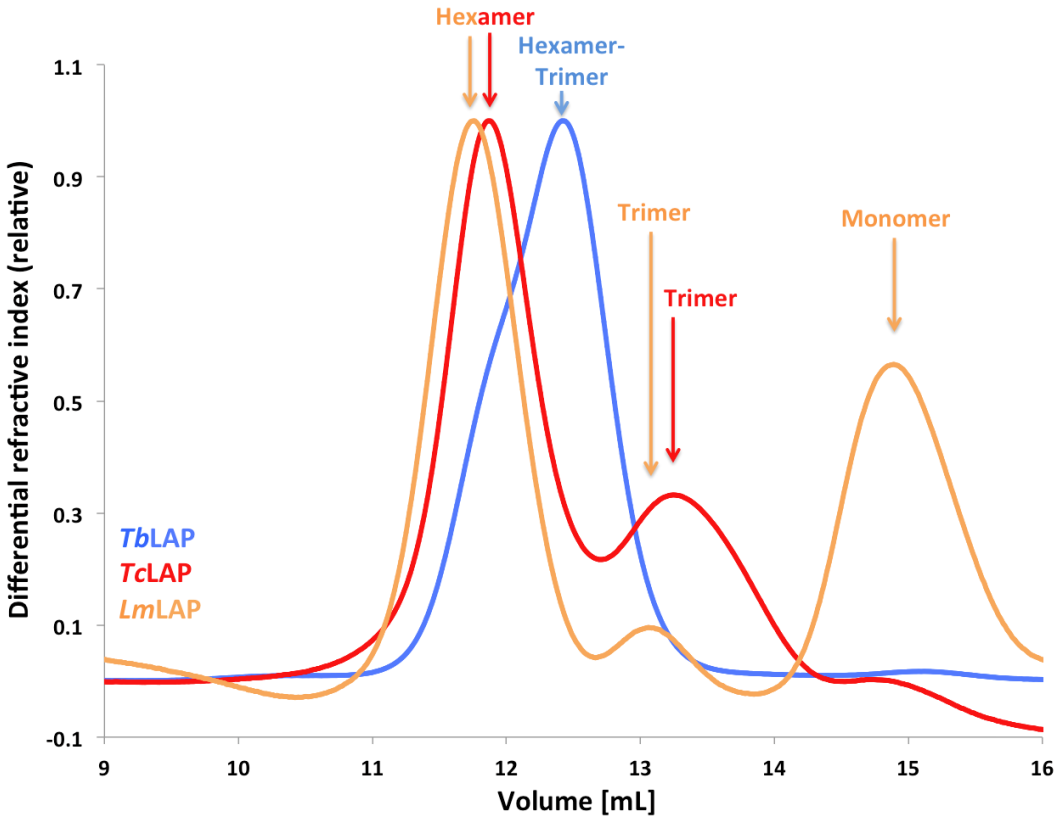


Figure 4-38: Annotated SEC-MALLS data for *Tb*LAP-A, *Tc*LAP-A and *Lm*LAP-A.

Shown are the overlaid differential refractive index traces of the three LAP-As each at 1 mg mL^{-1}.

*Tb*LAP-A (Figure 4-39) analysis in the NaCl-containing buffer showed an asymmetric double peak of ~ 270 kDa at one side of the peak and ~ 240 kDa at the other (Table 4-21). These calculated masses would correspond approximately to a tetrameric state of the protein, which is very unlikely when the X-ray structures of this and other LAPs are considered. More likely this peak is caused by a dynamic equilibrium of hexamers and trimers with the speed of equilibration being faster than the time of running the SEC. The asymmetric peak shape indicated the lower molecular mass species to be more abundant than the higher. The SEC-MALLS traces were identical for all concentrations tested. To investigate whether the buffer composition is destabilising the expected hexamer, a different buffer was tested exchanging the 200 mM NaCl with 100 mM KCl and increasing the Tris concentration from 10 mM to 50 mM. Enzymatic activity of *Tb*LAP-A could be shown in the KCl-containing buffer (see 4.3.1), but the protein was inactive in the NaCl-containing one (not shown). SEC-MALLS analysis of *Tb*LAP-A in the KCl-containing buffer showed a more

symmetric double peak with ~ 280 and ~ 270 kDa as molecular masses calculated from each side of the peak (Table 4-21). These data also suggest a concentration-independent mixture of hexamer and trimer, but in contrast to the previous buffer the two oligomeric states seem to be present in equal amounts. As the KCl-containing buffer shows a higher amount of higher molecular mass species and can also be used for measurements of the enzymatic activity, it can be assumed that the protein is active in the higher oligomeric state. Overall these data demonstrate a significant influence of the buffer composition on the stability of the oligomeric state of *Tb*LAP-A.

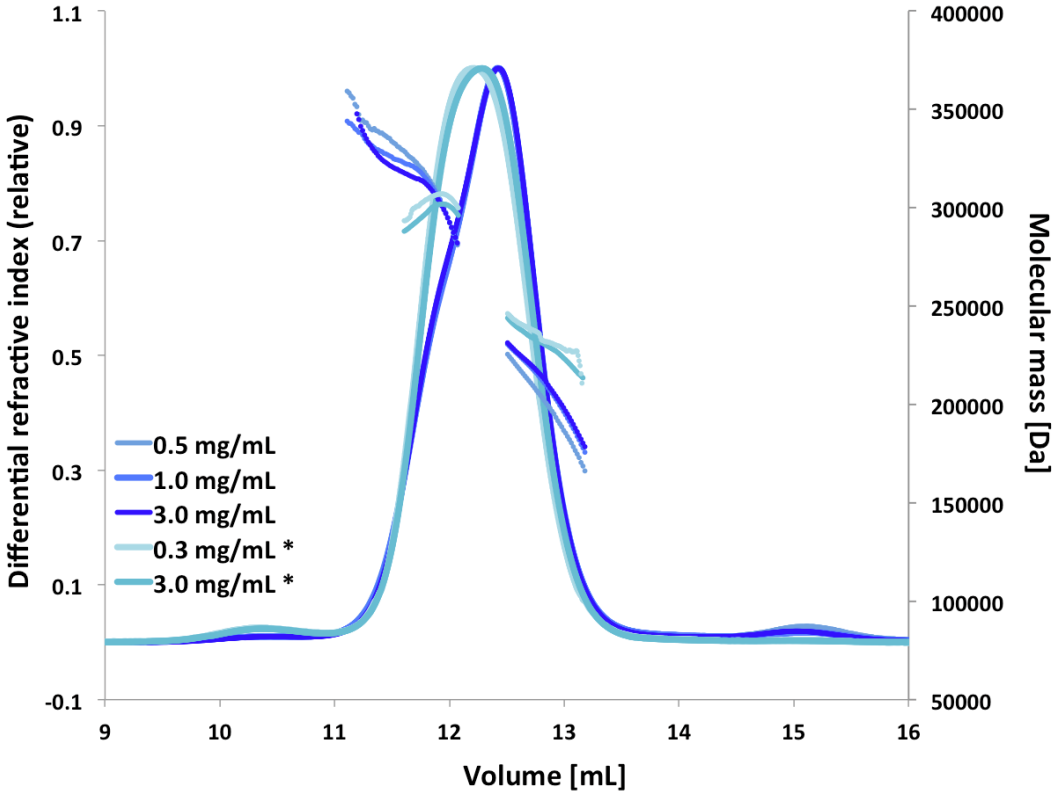


Figure 4-39: SEC-MALLS traces of *Tb*LAP-A.

Shown are the normalised traces of the differential refractive index and at the different protein concentrations loaded. Asterisks indicate the data obtained in KCl-containing buffer while the other traces are from NaCl-containing buffer. The molecular weights calculated from the peaks are shown as dashed lines in the same colours as the corresponding differential refractive index trace.

Table 4-21: M_w of the *Tb*LAP-A oligomers obtained from SEC-MALLS.

The asterisks indicate the samples obtained from KCl-containing buffer, all others obtained with buffer containing NaCl.

Protein	Concentration [mg mL^{-1}]	Aggregation in sample	Peak 1 [kDa]	Peak 2 [kDa]
*Tb*LAP-A	0.3	No*	281.2*	274.3*
	0.5	No	284.5	246.4
	1.0	No	273.8	238.8
	3.0	No	266.6	240.6
	3.0	No*	274.8*	269.0*
	5.0	No	264.6	236.5

The analysis of *Tc*LAP-A (Figure 4-40) showed the presence of two peaks, one with a calculated molecular mass of ~ 320 kDa and the other with ~ 170 kDa (Table 4-22), which correspond to the hexamer and trimer, respectively. While the elution volume of the hexamer peak was the same at all protein concentrations analysed, the peak maxima of the trimer peaks varies due to the influence of the hexamer species on the trimer. The higher the proportion of the hexamer compared to the trimer, the bigger its influence on the behavior of the trimer. At all concentrations analysed the hexameric species was the most abundant and as the concentration was lowered the proportion of trimers increased. Unlike the *Tb*LAP-A data, the hexameric and trimeric species separate well on the column, so that their masses can be determined more accurately. This suggests a slower conversion of the hexamer and trimeric *Tc*LAP-A species, than in *Tb*LAP-A.

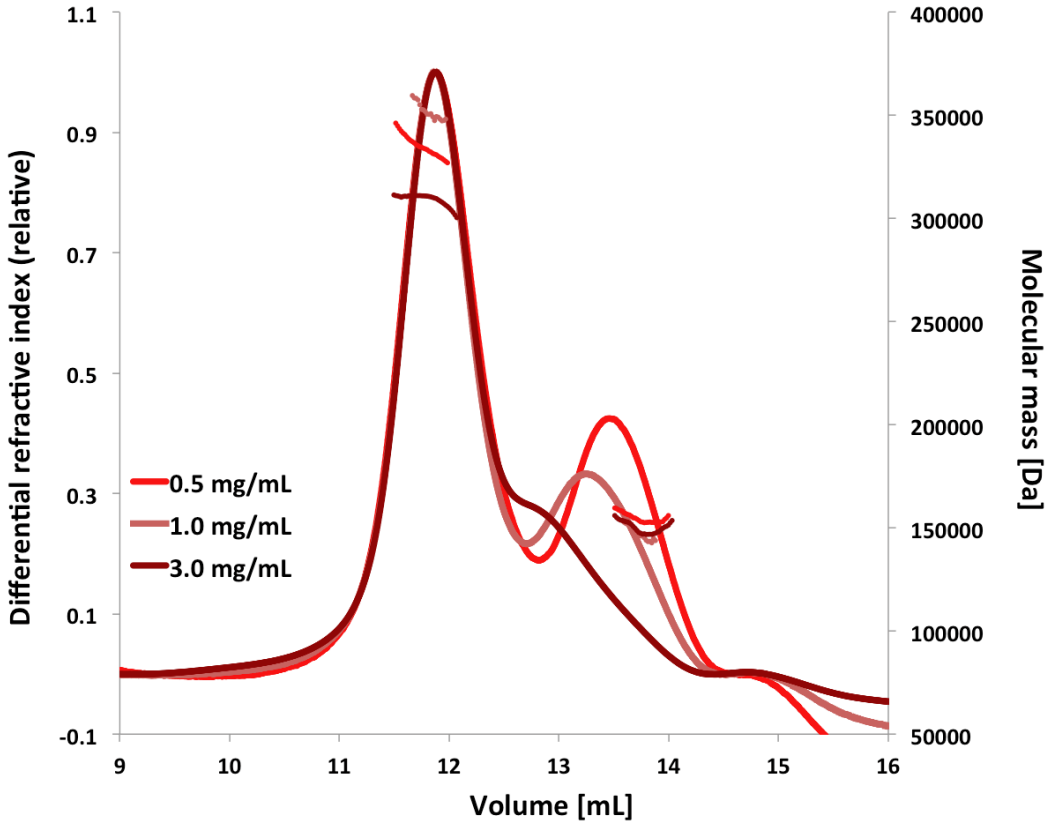


Figure 4-40: SEC-MALLS traces of *Tc*LAP-A.

Shown are the normalised traces of the differential refractive index and at the different protein concentrations loaded. The molecular weights calculated from the peaks are shown as dashed lines in the same colours as the corresponding differential refractive index trace.

Table 4-22: M_w of the *Tc*LAP-A oligomers obtained from SEC-MALLS.

Protein	Concentration [mg mL^{-1}]	Aggregation in sample	Peak 1 [kDa]	Peak 2 [kDa]
*Tc*LAP-A	0.5	No	331.1	162.2
	1.0	No	318.5	167.4
	3.0	No	309.9	183.2

Analysis of the *Lm*LAP-A (Figure 4-41) revealed three distinct molecular mass species plus a considerable amount of aggregation (not shown in the figure). This influences the baseline considerably and reduces the accuracy of the molecular mass calculations. The three elution peaks have molecular masses of ~ 300 kDa, ~ 125 kDa and ~ 60 kDa (Table 4-23), which correspond to the hexamer, trimer and monomer respectively. All three species present in the sample are well separated from each other and the aggregates in the void volume (not shown in the figure).

The lower the protein concentration the higher is the proportion of monomers in the mixture. Whether the monomers are a result of instability and subsequently result in their aggregation or if the monomeric state is physiologically relevant is not clear.

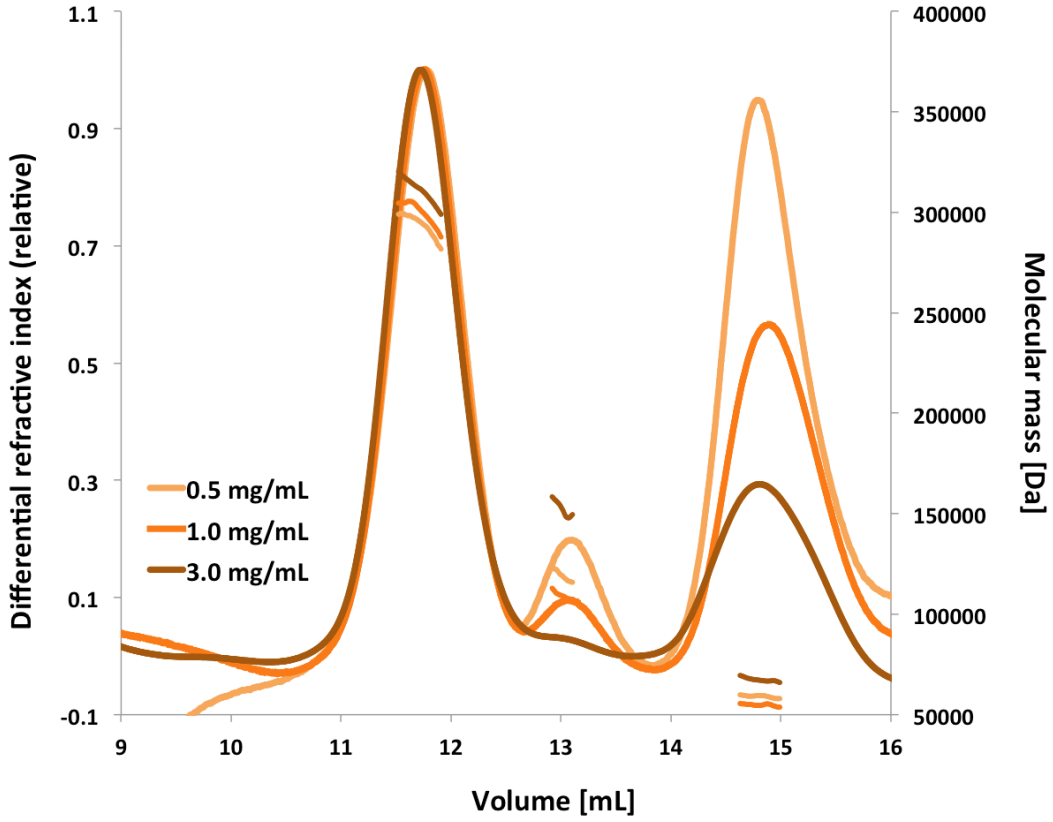


Figure 4-41: SEC-MALLS traces of *Lm*LAP-A.

Shown are the normalised traces of the differential refractive index and at the different protein concentrations loaded. The molecular weights calculated from the peaks are shown as dashed lines in the same colours as the corresponding differential refractive index trace.

Table 4-23: M_w of the *Lm*LAP-A oligomers obtained from SEC-MALLS.

Protein	Concentration [mg mL^{-1}]	Aggregation in sample	Peak 1 [kDa]	Peak 2 [kDa]	Peak 3 [kDa]
*Lm*LAP-A	0.5	Yes	293.1	118.9	59.1
	1.0	Yes	299.2	109.0	54.8
	3.0	Yes	310.3	152.2	67.4

4.4 Discussion

4.4.1 LAP-A as potential drug target in *T. brucei*

The data obtained indicate that LAP-A is not essential for survival and proliferation of bloodstream and procyclic *T. brucei brucei* parasites under the conditions tested. RNAi-mediated depletion in bloodstream parasites and both RNAi-mediated depletion and the generation of dKO mutants in procyclics support this conclusion. Whether this is the result of redundancy caused by the presence of additional M17 LAPs, paralogues of the one investigated, requires further research. The LAP-A paralogues are predicted to localise to the endoplasmic reticulum and the mitochondrion and therefore should have distinct functions from LAP-A, which is most likely cytosolic. It is not possible to predict if the paralogues can compensate for the loss of LAP-A. The overexpression of LAP-A in procyclics was not toxic. However, a decrease in proliferation and increase in sub-G_1 cells were observed after prolonged induction of LAP-A overexpression in parasites growing in minimal media suggesting a reduction in parasite fitness when media resources are limiting.

Overall the results indicate LAP-A cannot be considered as a drug target in *T. brucei*, unlike in *P. falciparum*, where its potential has been shown (Stack *et al.*, 2007, Harbut *et al.*, 2011). However, its role *in vivo* remains to be established since LAP-A might be required for infectivity, survival of the parasite inside the host or the crossing of the blood brain barrier. Due to the profound differences in metabolism and life cycle stages between the TriTryps, the results obtained with *T. brucei brucei* parasites do not allow to draw conclusions about LAP-A's potential as drug target in *T. cruzi* and *L. major.*

4.4.2 Physiological function of the LAP-A

The establishment of LAP-A's subcellular localisation could aid in the identification of the physiological function. Indeed a cytosolic localisation would suggest functions distinct from an enzyme that resides in the endoplasmic reticulum. Thus, while in the cytosol LAP-A would perform metabolic or housekeeping functions, in the endoplasmic reticulum the enzyme could be involved in the post-translational modification of N-terminal signal peptides of specific proteins and therefore perform a more regulatory role. In the present study the enzyme appears to be mainly cytosolic although colocalisation studies with a marker for endoplasmic reticulum have not been performed and the latter cannot be excluded. Fluorescence microscopy does not give convincing evidence for the presence of the enzyme in the nucleus and hence an involvement in site-specific homologous recombination, as in *E. coli* (Gourlay & Colloms, 2004, Paul & Summers, 2004), is unlikely.

T. brucei parasites do not exhibit a clear and specific phenotype upon depletion or overexpression of LAP-A, so that the protein's physiological function remains elusive.

When procyclic *T. brucei* are exposed to leucine starvation, parental cells stop proliferation and enter a short stationary phase before they show DNA fragmentation. LAP-A-dKO parasites on the other hand appear to exhibit a growth advantage under these conditions and proliferate further entering the stationary phase later. This distinct and somewhat surprising growth behaviour might indicate a role of LAP-A in the sensing of amino acid levels and its influence in regulation of the growth rate. LAP-A's involvement in these processes is speculative and its connection to the growth behaviour remains to be elucidated.

One possible mode of regulation would be the participation of LAP-A in the trimming of N-termini of specific proteins. These modifications may be essential for protein function and sensitive to the levels of leucine. This trimming could be triggered by the lack of leucine as direct regulator of LAP-A activity or alternatively the proteins whose N-termini are modified sense the lack of leucine. Another

possibility is a more indirect role of LAP-A in amino acid sensing, related to the overall amino acid level inside the cells. A lack of LAP-A as metabolic enzyme could lead to an increased level of di- and tri-peptides and overall lower levels of free amino acids in the cells. This could induce an increased expression and activation of amino acid transporters, which would in turn increase the amino acid uptake from the media. This would lead to better growth initially, but ultimately lead to a faster depletion of resources in the media and therefore result in cell starvation after a shorter period of time.

4.4.3 Potential use of LAP-A-dKO, RNAi and overexpression strains

The *T. brucei* LAP-A-dKO, -RNAi and overexpressing strains can be used to study the physiological function of the LAP-A and other aminopeptidases. They can serve as starting strains for the knockout, knock down or overexpression of other enzymes involved in amino acid metabolism to gain insight in these complex processes of *T. brucei*, which might give clues about other potential drug targets.

4.4.4 Structure and oligomeric state of LAP-A and their effects on possible physiological substrates

The TriTryp LAP-As form hexamers in solution and also crystallise as hexamers. Although it cannot be ruled out that these are artefacts of the buffer conditions used to conduct the experiments, homology to other LAPs indicates this oligomeric state to be a key features of this family of enzymes. Analysis of the structure shows the active sites of each protomer facing towards the inner cavity, sequestered from the bulk solvent. Although access channels allow diffusion into the inner cavity and the active sites, the accessibility seems restricted to small, unfolded substrates. This favours LAP-A's function as a mainly metabolic enzyme with housekeeping function or trimming of the N-termini of specific unfolded peptides. However, the bacterial

M17 family LAPs PepA and PepB were shown to cleave the loosely folded protein casein *in vitro* (Bhosale *et al.*, 2010), indicating the LAP's potential to process larger substrates.

4.4.5 The physiological metal ion in the active site of LAP-A

While early studies on the bovine lens LAP showed Zn^{2+} to be the physiological metal ion in the active site of the M17 family LAPs (Carpenter & Vahl, 1973), the enzyme was also active with other divalent metal ions (Kim & Lipscomb, 1993, Kim *et al.*, 1993, Jia *et al.*, 2010, Morty & Morehead, 2002, Cadavid-Restrepo *et al.*, 2011). It was noted before (Kale *et al.*, 2010) that *Pp*LAP produced in *E. coli* has Mn^{2+} in its active site. It was shown that substrate specificity is influenced by the metal identity in the active site (Cappiello *et al.*, 2006), with sulfhydryl-containing substrates only being accepted by Mn^{2+}-containing LAPs but not by Zn^{2+}-containing ones.

Within this work, the best crystals of the TriTryp LAP-As grew in the presence of Mn^{2+}. Indeed, crystals grown in Zn^{2+}-containing conditions also had Mn^{2+} in the active site. This might implicate Mn^{2+} to be the physiological metal co-purified with the enzyme expressed in *E. coli* or, but more likely, indicates Mn^{2+} contamination of one or multiple solutions used throughout the process. The high degree of identity of the active site residues of the TriTryp LAP-As with those from other characterised species, suggests similar characteristics and Zn^{2+} to be the physiological metal.

4.5 Summary and future work

The structures of the TriTryp LAP-As were solved in different ligand complexes, which gives insight into ligand binding and confirm the homology to other family members. The parasitological data presented in this thesis show LAP-A not to be essential in *T. brucei brucei* when cultured *in vitro*. However, LAP-A might still be indispensible for infectivity, immune evasion or crossing of the blood brain barrier, which remains to be investigated. The experiments in *T. brucei* do not allow conclusions to be drawn about LAP-A's essentiality in *T. cruzi* and *Leishmania spp*.

LAP-A's physiological function within *T. brucei* parasites remains elusive and further studies are necessary to confirm its localisation to the cytoplasm and to determine whether its function is metabolic or regulative and can be compensated for by the presence of its paralogues. In future investigations, kinetic characterisation of the peptidase activity would be useful for determining the physiological metal ion and substrates, as well as the possible regulation of its function by leucine or other amino acids. Examination of the paralogues LAP-B and LAP-N and their potential to compensate for the loss of LAP-A will expand the understanding of amino acid metabolism in these parasites and potentially aid identification of novel targets for rational drug design.

Chapter 5 Summary and Conclusions

The TriTryps cause a set of devastating diseases, mainly in the developing world. In addition, the poverty of the effected population led to general neglect by the pharmaceutical industry. However, in recent years neglected tropical diseases have been addressed more extensively.

The Trypobase project was one of the initiatives aimed to identify lead compounds for drug development against the TriTryps using both classical pharmacology with phenotypic assays and a target-based drug design approach. The targets included *Tb*AK, and the TKs and TMPKs from all three species. In this thesis the structures of *Tb*AK, *Lm*TK and *Tb*TK were solved to aid the process of rational drug design. With the exception of cordycepin binding to *Tb*AK, no lead compound was identified for these targets during the limited time of the Trypobase project. However, the structures obtained can aid future drug discovery projects.

Extensive phenotypic screening of compounds for trypanocidal activity identified a compound active against *T. cruzi*. Initial metabolomic studies investigating the possible biological target indicated leucyl aminopeptidase to be effected. Based on this hypothesis the LAP-A was added to the target list and in scope of this work, the structures of LAP-A from all three species were solved. Later, assays carried out by collaborators revealed the lead compound not to act on LAP-A. Hence, within this work, LAP-A's potential as a drug target was evaluated in *T. brucei* parasites. Neither RNAi-mediated LAP-A depletion nor dKO nor LAP-A overexpression were lethal to the bloodstream and procyclic form parasites, which does not support LAP-A as potential drug target. Nevertheless, the parasite strains generated can be used to shed light on the complex interactions within amino acid metabolism in *T. brucei*, which might lead to the identification of novel drug targets. The data do not allow conclusions about LAP-A's potential as drug target in *T. cruzi* and *L. major*, which requires further investigation.

Taken together, my results highlight the difficulties of classical pharmacology and target-based drug design. The classical pharmacology approach identifies compounds, whose mechanism of action (MOA) is not known, and identification of

the biological targets is challenging with a high risk of false positives. An unknown MOA results in difficulties with optimisation of the pharmacokinetic properties and reduction of off-target effects, which prolongs the process of drug development considerably. The target-based approach on the other hand struggles with identification of ligands for known targets. Screening carried out *in vitro* often does not reflect the physiological binding situation and especially weak-binding compounds might not be identified as potential candidates. However, crystal structures aid the process of compound optimisation by identifying features that can be used to increase specificity or to suggest modifications to facilitate resistance to degradation by other enzymes, as seen clearly in case of 2-fluorocordycepin (Vodnala *et al.*, 2013).

Many drug discovery processes suffer from short funding periods, which do not leave much room for adaptations of the strategy after obtaining initial results.

Appendices

A.1 Expression, purification, and crystallisation of thymidylate kinases

A.1.1 Introduction

Thymidylate kinases (TMPKs) catalyse the phosphorylation of dTMP to dTDP (EC 2.7.4.9) with the preferred phosphate donor being ATP. TMPKs have been extensively studied for their ability to phosphorylate cytotoxic nucleoside analogues and as potential drug targets, e.g. in *P. falciparum*, causative agent of malaria (Kandeel *et al.*, 2009). Several structures have been solved (Figure A.1-1) and discussed with respect to mechanism, substrate specificity and suitability as drug target (Whittingham *et al.*, 2010, Choi *et al.*, 2012, Leibly *et al.*, 2011). TMPKs are approximately 25 kDa in size and form homodimers in solution. Each subunit contains a five-stranded parallel β-sheet surrounded by 7-11 α-helices. Each subunit forms its own active site, composed of a so-called critical loop with the DR(Y/H) motif, containing the catalytically important aspartic acid and arginine residues (Reynes *et al.*, 1996), the P-loop, responsible for binding of the α- and β-phosphates of the phosphate donor and their orientation (Saraste *et al.*, 1990) and a LID region, a flexible loop which closes upon substrate binding.

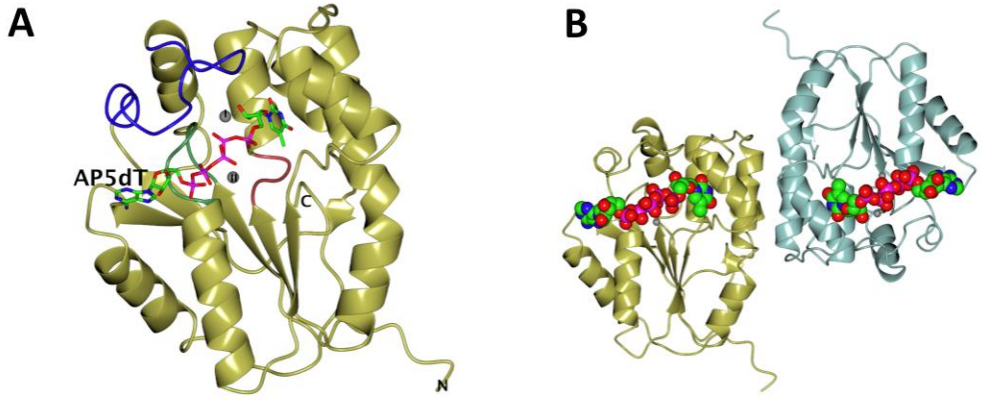


Figure A.1-1: Structure of *P. falciparum* TMPK (*Pf*TMPK).

(A) *Pf*TMPK (PDB ID: 2wwh, Whittingham *et al.*, 2010) protomer shown in ribbon representation with P-loop in lawn green, critical loop in pale crimson and LID region in blue, AP_5dT in cylinder representation coloured by atom type. (B) *Pf*TMPK dimer shown as ribbon with AP_5dT atoms represented as spheres. Figure made with CCP4*mg* (McNicholas *et al.*, 2011).

There are two types of TMPK, distinguished by the presence (type I) or absence (type II) of an additional basic residue in the P-loop. Depending on species the details of the catalytic mechanism differ slightly, e.g. in movements of P-loop and lid during catalysis or orientation of specific side chains to stabilise the transition state. The phosphoryltransfer reaction of mouse TMPK and the homologous guanylate (GMPK) and adenylate kinases (AMPK) have been determined to be random bi-bi mechanisms with both substrates (NMP and phosphate donor) binding independently of each other (Cheng & Prusoff, 1973, Li *et al.*, 1996). It has been analysed in great detail for the human enzyme using structural data for complexes with substrates, products and transition state mimics (Figure A.1-2, Ostermann *et al.*, 2000). In brief, upon substrate binding the P-loop, adenosine-binding loop and LID region close into their active conformation orienting the substrates (Figure A.1-2 A). The γ-phosphate is essential for adoption of the active conformations of substrates and the active site residues of the TMPK. The correct conformations allow the dTMP phosphate, held in place by the catalytic arginine (R97 in human TMPK), to make a nucleophilic attack on the γ-phosphate of the phosphate donor forming a transition state, looking similar to the dTMP-AlF_3-ADP complex in Figure A.1-2 B (green). After the reaction is completed the product moves into a

conformation unfavourable for the back reaction (Figure A.1-2 C and D). The products are released thereafter.

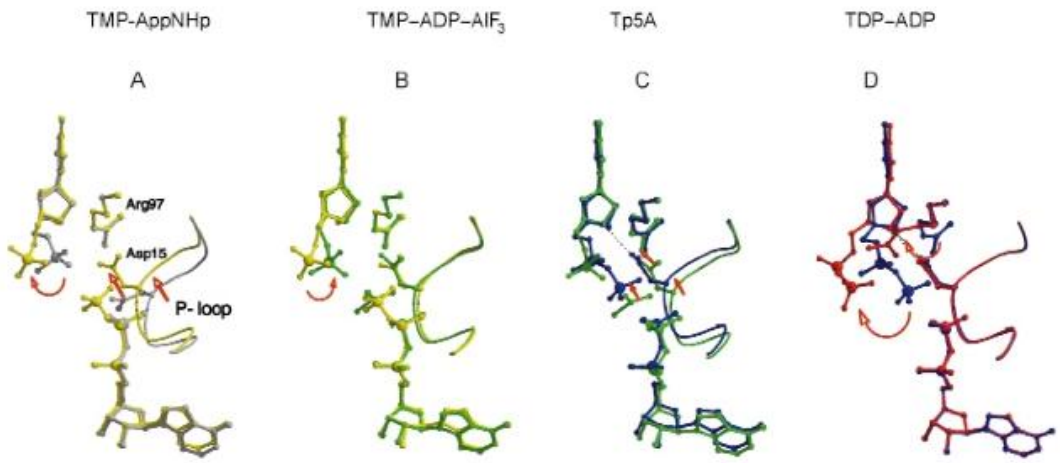


Figure A.1-2: Mechanism of human TMPK based on X-ray structures.
(A) Overlay of TMPK complexes with dTMP-ADP (grey) and dTMP-AppNHp (yellow); (B) dTMP-AppNHp (yellow) and dTMP-AlF_3-ADP (green); (C) dTMP-AlF_3-ADP (green) and Ap_5dT (blue); (D) Ap_5dT (blue) and dTDP-ADP (red); more details are given in the text. Figure modified from (Ostermann *et al.*, 2000).

A.1.1.1 TMPKs in drug development

TMPKs, while playing a crucial role in activation of cytotoxic nucleoside analogues, are known to be essential enzymes and therefore potential targets in gram-positive bacteria like *Staphylococcus aureus* (Choi *et al.*, 2012). Additionally, it was shown that inhibition of human TMPK improves the efficiency of anticancer drugs like doxorubicin (Hu & Chang, 2008).

A.1.1.2 Thymidylate Kinases from the TriTryps

The TriTryp TMPKs are much bigger than those of other species with masses of 40.1, 39.5 and 39.4 kDa for *L. major*, *T. brucei* and *T. cruzi* respectively, due to the presence of an extra N-terminal domain of unknown function. This domain of approximately 130 amino acid residues has no homologues outside the kinetoplastida and is predicted to contain 4-5 β-strands and 4 α-helices by the

PSIPRED and PHD servers (Buchan *et al.*, 2013, Rost *et al.*, 2004). An alignment of TMPK sequences, whose X-ray structures are known, is shown in Figure A.1-3. As expected, the catalytically important regions, P-loop and the critical loop, are highly conserved among different species. Besides the N-terminal extension, TriTryp TMPKs have an additional helix inserted between the critical loop and lid domain. To date no kinetic or biophysical data have been published for any kinetoplastid TMPK.

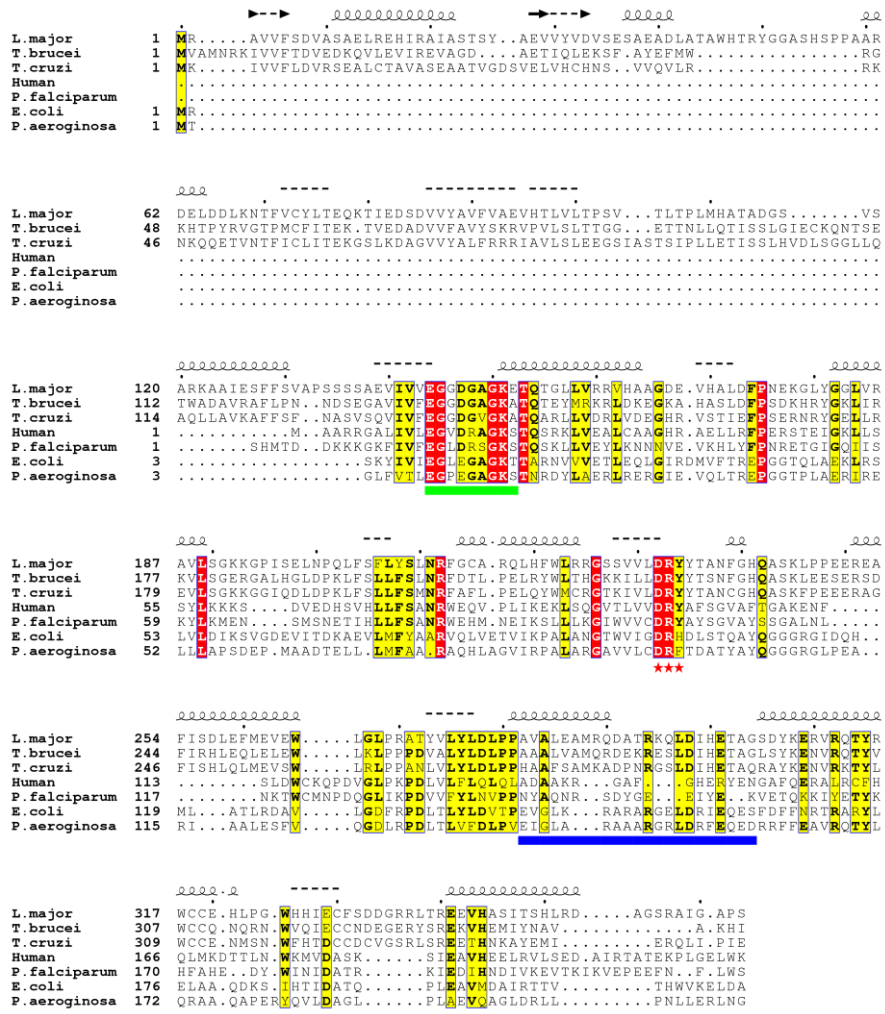


Figure A.1-3: Sequence alignment of TMPKs.

Similar residues highlighted in yellow, identical residues shown in red. Secondary structure was predicted for *Lm*TMPK using the PHD server (Rost *et al.*, 2004) and is shown above the sequences with α-helices as spirals and β-sheets as dashed lines (confidence of the prediction < 90%) or arrows (confidence >90%). Important structural features determined in the *Pf*TMPK structure are highlighted below the sequences: P-loop region (green bar), critical loop (red stars) and lid region (blue bar). Figure was made in ESPript 3.0 (Gouet *et al.*, 2003).

A.1.2 Experimental

A.1.2.1 Cloning and Expression of TMPKs in the baculovirus system

Previous expression tests of a number of TMPK constructs in *E. coli* failed to produce soluble and stable protein. Therefore, 18 constructs were cloned and tested for expression with the baculovirus system at OPPF, Table A.1-1. The SDS-PAGE analysis of the expression test is shown in Figure A.1-4.

Table A.1-1: TMPK constructs for expression with the baculovirus system.
Cloning performed in 96 well plate format; "well" gives location in the 96 plate, OPPF number was given for identification of the process in the OPPF database.

Well	Gene name	Amino acid range	M_r of expected protein (+ tag) [Dalton]	Vector pOPIN	OPPF No.
G04	*Tb*TMPK	2-345	39383.7 (+2158)	F	11350
H04	*Tb*TMPK	120-345	26180.6 (+2158)	F	11353
A05	*Tb*TMPK	2-345	39383.7 (+1274)	I	11348
B05	*Tb*TMPK	120-345	26180.6 (+1274)	I	11351
C05	*Tb*TMPK	2-345	39383.7 (+969)	E	11349
D05	*Tb*TMPK	120-345	26180.6 (+969)	E	11352
E05	*Tc*TMPK	2-351	39302.1 (+2158)	F	11362
F05	*Tc*TMPK	122-351	26447.1 (+2158)	F	11365
G05	*Tc*TMPK	2-351	39302.1 (+1274)	I	11360
H05	*Tc*TMPK	122-351	26447.1 (+1274)	I	11363
A06	*Tc*TMPK	2-351	39302.1 (+969)	E	11361
B06	*Tc*TMPK	122-351	26447.1 (+969)	E	11364
C06	*Lm*TMPK	2-363	39952.8 (+2158)	F	11356
D06	*Lm*TMPK	131-363	26034.3 (+2158)	F	11359
E06	*Lm*TMPK	2-363	39952.8 (+1274)	I	11354
F06	*Lm*TMPK	131-363	26034.3 (+1274)	I	11357
G06	*Lm*TMPK	2-363	39952.8 (+969)	E	11355
H06	*Lm*TMPK	131-363	26034.3 (+969)	E	11358

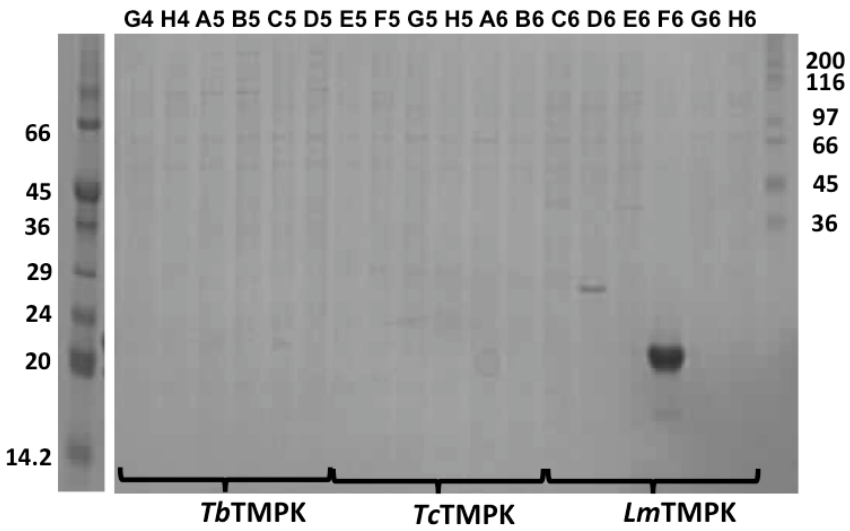


Figure A.1-4: SDS-PAGE of TMPK expression tests in the baculovirus system.

First and last lanes are protein standards. Each other lane contains a NiNTA purified expression test sample from the baculovirus expression test. Lane labelling corresponds to the well numbers in Table A.1-1. Performed by Louise Bird and Joanne Nettleship at OPPF.

A.1.2.2 Purification of *Lm*TMPK

Cells from 1 L Sf9 cultures expressing *Lm*TMPK were resuspended in 40 mL lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 25 mM imidazole, 1 mM DTT, 1 mM EDTA, 100 µM PMSF and 25-29 U Benzonase) and lysed by sonication. Cell debris was removed by centrifugation (5000 rcf, 30 min, 4°C). The supernatant was loaded onto a pre-equilibrated 5 mL HisTrap crude FF column for 1 h in a loop on ice. After washing with 10 mL lysis buffer the column was connected to an AKTA FPLC and washed with 15 mL lysis buffer. The protein was eluted with an imidazole gradient from 25 mM to 500 mM. The protein-containing fractions, identified by SDS-PAGE analysis, were pooled and concentrated to ~250 µL in the presence of 1 mM $MgCl_2$, 1 mM dTMP and 1 mM AppNHp. The stable protein was loaded onto an equilibrated Superdex 200 column in SEC buffer (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM DTT). After SDS-PAGE analysis the protein-containing fractions were pooled and concentrated to 5.2 mg mL^{-1} in the presence of 1 mM $MgCl_2$, 1 mM dTMP and 1 mM AppNHp. The protein was aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

A.1.2.3 Crystallisation of *Lm*TMPK

The initial screen (JCSG + Index) was carried out in sitting-drop vapour diffusion format with 150 nL protein solution mixed with an equal amount of precipitant, using a Mosquito robot (TTP Labtech, UK). Additionally, seeding experiments were set up as described for *Tb*TK in 3.2.8 using seed stocks prepared from initial hits as well as *Lm*TK 11320 crystals for cross-seeding. Crystals grew within two months and were optimised in hanging-drop and sitting drop vapour diffusion format with 1 µL or 0.5 µL protein solution respectively plus an equal volume of precipitant.

After mounting and vitrification of optimised *Lm*TMPK crystals, no diffraction was obtained. Therefore, the small microcrystals found in the initial screens were tested for diffraction *in situ* at beamline I04-1 of the DLS.

A.1.3 Results

A.1.3.1 Expression, purification and crystallisation

Expression in *Sf*9 cells did yield soluble protein for the construct 11357, which encodes *Lm*TMPK in which the first 130 amino acids have been deleted (spanning amino acids 131-363), fused to an N-terminal non-cleavable Histag. None of the other constructs tested produced soluble protein. Purification of *Lm*TMPK 11357 was successful using nickel affinity and size exclusion chromatography (Figure A.1-5) and the protein could be stabilised by addition of dTMP and AppNHp.

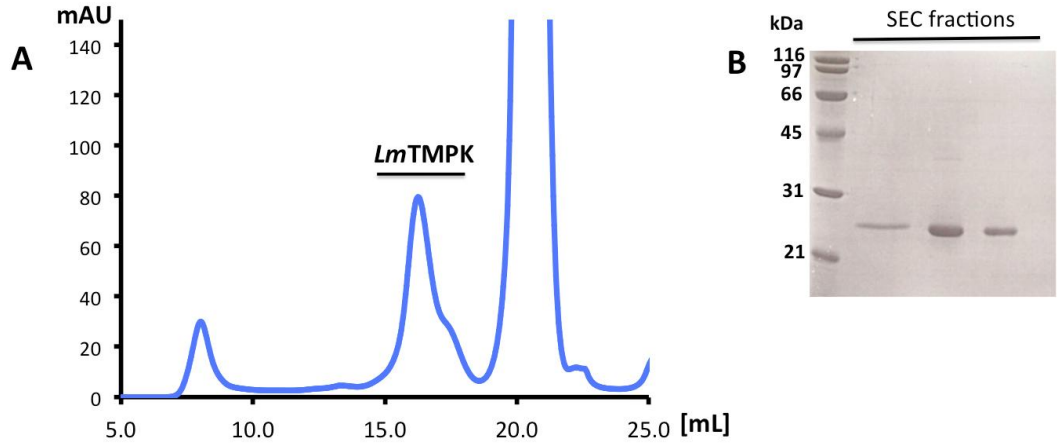


Figure A.1-5: *Lm*TMPK purification.

(A) Size exclusion trace (absorption at 280 nm) of LmTK 11357. The black bar indicated the fractions containing *Lm*TMPK, identified by SDS-PAGE analysis (B).

The best crystals grew in following conditions:

- 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris pH 5.5, 25 % (w/v) PEG 3350, in sitting drop format, with seeding (Figure A.1-6 A).
- 0.2 M Li_2SO_4, 0.1 M Bis-Tris pH 5.5, 25 % (w/v) PEG 3350 in sitting drop format, with seeding (Figure A.1-6 B).
- 0.2 M Li_2SO_4, 0.1 M Bis-Tris pH 5.5, 20 % (w/v) PEG 3350 in hanging drop format, (no seeding) (Figure A.1-6 C).
- 0.2 M Li_2SO_4, 0.1 M Bis-Tris pH 6.5, 19 % (w/v) PEG 3350 in hanging drop format, with seeding (Figure A.1-6 D).

*Lm*TMPK formed small bi-pyramidal crystals (Figure A.1-6 A & B) after 2 months. After optimisation and seeding experiments the resulting crystals (Figure A.1-6 C & D) mounted and vitrified did not diffract. *In situ* exposure at the I04-1 beamline of the DLS revealed the initial crystals (not optimised) also to be non-diffracting.

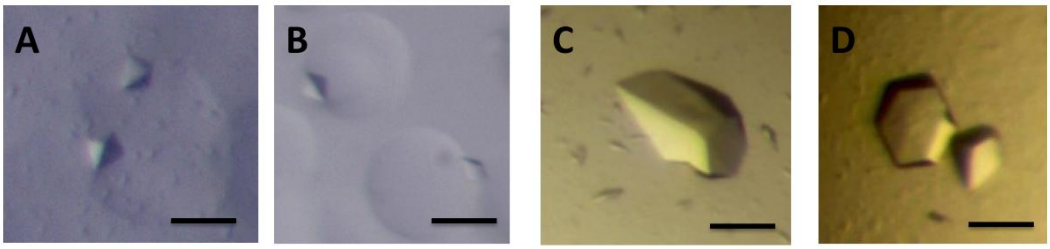


Figure A.1-6: Crystals of *Lm*TMPK.

(A) Crystals grown after 2 months with seeding in 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris pH 5.5, 25 % (w/v) PEG3350, black bar ~ 100 µm. (B) Crystals grown after 2 months with seeding in 0.2 M Li_2SO_4, 0.1 M Bis-Tris pH5.5, 25 % (w/v) PEG 3350; black bar ~ 100 µm. (C) Crystals grown in 0.2 M Li_2SO_4, 0.1 M Bis-Tris pH 5.5, 20% (w/v) PEG 3350, black bar ~ 50 µm. (D) Crystals grown with seeding in 0.2 M Li_2SO_4, 0.1 M Bis-Tris pH 6.5, 19% (w/v) PEG 3350, black bar ~ 50 µm.

A.1.4 Discussion

Expression and stabilisation of the TriTryp TMPKs turned out to be very difficult. With the baculovirus expression system it was possible to obtain soluble and stable N-terminally truncated *Lm*TMPK, but all attempts to express full-length *Lm*TMPK or any construct of *Tb*TMPK and *Tc*TMPK failed. This truncated *Lm*TMPK was purified and stabilised by addition of TMP and AppNHp. Unfortunately, *Lm*TMPK did not form diffracting crystals, so that no structural information could be obtained to help rational drug design.

A.2 Sequences and plasmid maps

A.2.1 Protein sequences

The full-length protein sequences from the databases are shown below with the amino acid range seen in the crystal structures in bold.

>*Tb*AK (Tb427.06.2360)

**MSSAPLRVYVQCNPLLDVSAHVSDEFLVKYGLERGTAILLSERQKGIFDDIEKMPNVRYVPGGSGLNV
ARVAQWMQQAYKGKFVTYVGCIADDRYGKVLKEAAEHEGIVMAVEHTTKAGSGACAVCITGKERTLVA
DLGAANHLSSEHMRSPAVVRAMDESRIFYFSGFTLTVDVNHVLQACRKAREVDGLFMINLSAPFIMQF
FSAQLGEVLPYTDIIVANRHEAKEFANMMKWDTDCVEEIARRAVSEVPYTGTKGRVVVFTRDIESTVL
ATKDGVETVPVPQLDQDKVIDMNGAGDAFMGGFLSAYAVGKDLRRCCETGHYTAQEVIQRDGCSFPEK
PSFS**P

>*Lm*TK (LmjF.21.1210)

MF**RGRIELIIGPMFAGKTTELMRRVKREIHARRSCFVIKYSKDTRYDEHNVASHDQLMLRAQAAVSQL
TEVRDTWKRFDVLAIDEGQFFSDLVDFCNTAADAGKVVMVSALDGDYRRKPGQICELVPYCEAVDKLT
AVCMMCHEQPACFTRRTVNVEQQELIGGADMYIATCRECYSKQ**QLPSIEEMRTQRMAIKEVEKRYLGL
SDKRATASPQTPEKPAGGWCTKAGVATLPTMVKEGAASSGASAGMKSGRGLCEMQAFTTETPKYQRVE
PACTASAASSE

>*Tb*TK (Tb427.10.880)

MHDGDGNIELIIGPMFAGKTTELMRRVQRHKHAQRSCYIINYSRNSYQNQRLSTHDQLSLTANVSIAK
LSEVCDEWRDYDVIAVDNGQFFPDVVGFCARAANEGKTVIVSALDVDCRETPFDEVCRLVPRAESVLK
LSAVCMECHEHDAFLTYRTIESNERELYGGADMYLAVCRWCYKQLTMSHVDAQKTSASTAAVVPNGA**H
GRIELIIGPMFAGKTTELMRRVQRHKHAQRSCYIIKYTGDTRYSEGAITSHDQRALTANVSVSNLHDV
GDEWRKYDVIAVDEGQFFPGVAAFCSKAADSGKVVIVSALDADYLQEPFEEICLLVSRADSVVKLSAV
CMECHNRKASFTYRTVKSDERKLVGGSDMYMSVCRSCYETKRN**MVQTEKYIYSCVGINEGSYSECSPG
PSERSSAGTSGVQTSVKVDEQNCTEPNTEAKKMPLKRKRNQMAVDTT

>*Tc*TK (TCSILVIO_009141)

MFASHTSTCTLKLRYAEDQRNFRLAFGFCASSIQQLFADISLQMMLENESHGHIELIIGPMFAGKTTE
LMRRVRRELFAKRSCYVIKHSRDARYSRESVSSHDKLLLGATAAVAELREVGDAWRPFDVVAVDEGQF
FPDIVGFCNTAADAGKTVIVSALDGDYRRQPFDGICRLIPLAESVKKLTAVCMECHCRSASFTYRTVS
SEKRELIGGADMYIAACRTCFVTKSKKRAEMEAKGSVQTVDKCSDTPAAEKQKGEQREVEEDPITPPT
TIAAAETPLSASAARKRPVDVDENVPTDAERCKSKARLECH

>*Lm*LAP-A (LmjF.11.0630)

M**SAMKRPRSNSIVEETAVSAYVQTCTNFKSNVTFTDISKVSCVAPHVLLVGALEQLRDNSLKSVLFYC
PAVAEALQRVKAGSTVKTLAEVPGRKGYTEVMVTALPATTSRTNCPYRADSMSEAVAAACGSVEEGEV
LDVYVCAPAGSETAVANAVARAAPHSYTAKAGQATKAYMKQAMSLNVVMSSRAAFTQELVRGKSVCVA
ELEAICTSVQLCQRLVDTPPCMLDTVVYAEIAAAYAAELGVDMTVIKGDELREKGYGGIYAVGKCAQY
PAHLVTLRYRNPNAAEGAKNIAMVGKGVYDCGGLALKPAAHMTNMKTDMGGSAGVFCAFIAVVRSMKM
QRTHFSHIANISVTLCLAENAIGPHSYRNDDVVVMKSGKSVEVMNTDAEGRIVLGDGVYYATGEQDFV
PDVLIDMATLTGAQGVATGSKHAGVYASDAEAEKDMISAGLQSGDLCYPVLYCPEYHEEVYKSPCADM
RNIANSSSSAGSSCGGYFVEQHLSERFRGPFVHVDMAYPTSNTAGATGYGVTLVFEFLRQH**

>*Tb*LAP-A (Tb427tmp.02.4440)

MP**TLPKAEAKELSAFVQSCVEYKTNVCFTDVAAYESNQKGVLSSGLAVLVGTHKQLRDPAVQRLPFYN**
PAVAEAIERVKEGGTYGVLVEGLANAAGSKFVRVVVGEVPTKASRNNCPARPDVVTXLVTAALDEVKE
PNTTVDVFVLSNAVLPIAAAVARCGKHNFSAKDGAAAAAYNSGKVSRLQVVFPEPPAIPPKDLEAVAT
STQLCQRLVDAPPNLLTTATFTEIAQGYAKALGFDVDVICGDDLCERGYGGIYSVGKAAFEAPRLVTL
LYTPKGTPVKKVSLVGKGIVYDCGGLALKPADYMKLMKHDMGGAAAVFCGFLTAVRLQQPVQLSCTLC
LAENAIGPKSYRNDDIIVMKSGKTVEVINTDAEGRIVLGDGVFHATNELSFTPDVVIDMATLTGAQGI
ATGRHHAGLYVNEEGAEAAMLRAGRESGETCFPVLYCPEYHEPEFKSNHADMTNLMERRDNAGVSCAG
YFITTHLSPKFTGAHIHVDLAYPVFNSNGATGFGPALLTEYFRKL

>*Tc*LAP-A (Tc00.1047053508799.240)

MNRPPATD**SGASEFVRSCVNFKTNVEFTDVDAYKKTHTAGLASTFVVILGTHAQLREDALKELPFCPA**
VAEAIQRVKKGETYAVLAEGVKNEANERFVRVLVGEVPSEASRTNCSARPDVVVSLLSAALEEIKGPE
TKVDVFVCSTAALAIAVAAARSAKRNFTAKEGLAARGYCDNCVRLTVVFPTAPNPSPSELVVVATSTQ
LCQRLVDAPTNLLNTATFAEVAQSYAKELGCAVDVICGEELRERGYGGIYSVGKCAAEAPRLVTLSYK
PKDETRKKVALVGKGIVYDSGGLSLKPTNFMTGMKRDMGGAAAVFCGFLTAVRLQMPIELSCTLCLAE
NAIGPDAYRNDDILTLKSGKTVEVNNTDAEGRLVLGDGVFHATHEISFKPDVLVDMATLTGAQGIATG
HRHAGIFVNDEEEELSFLKAGRVSGETCFPVLYCPEYHVTEFRSPVADMRNSVKQVNNASVSCAGHFV
ANHLSPDFKGKHIHVDMAFPAFENDKATGFGPALLMEYLRNLR

A.2.2 Nucleotide sequences of *Tb*LAP-A RNAi fragment and UTRs for double KO

*Tb*LAP-A RNAi fragment (reverse complement)

TACCCAAGGCCGAAGCAAAAGAGCTTTCTGCGTTCGTCCAGTCGTGCGTCGAATACAAAACCAATGTT
TGCTTCACCGATGTCGCCGCCTACGAAAGCAACCAGAAGGGTGTACTGTCCTCAGGACTCGCTGTTCT
TGTTGGAACTCATAAGCAACTCCGTGACCCCGCCGTCCAGCGTTTACCATTTTACAATCCCGCTGTTG
CTGAAGCTATTGAGCGGGTGAAGGAAGGTGGGACGTACGGCGTTTTGGTGGAGGGACTTGCGAATGCG
GCAGGCTCAAAGTTCGTGCGCGTTGTCGTTGGTGAGGTGCCCACGAAGGCATCTCGCAACAACTGCCC
TGCCCGACCGGATGTTGTGACTGCNCTTGTAACTGCTGCATTGGACGAAGTTAAAGAGCCTAATACAA
CGGTGGATGTCTTTGTGCTTTCCAATGCTGTCCTTCCCATCGCTGCTGCTGTAGCACG

5'-UTR (reverse complement)

GGGAGGCGGTAATGACAAGTCGCCCGCAGGATGGATAAAATTTACGGGCCAAACAAGTGCTAAATCAG
GAGACAGGTCGTACTTTTATTATTATTATTATTATCATTATCACCCTTACTGCTGTTTTTAATGTTTC
TTTGTCCCTTTCTCTGTTTCTACGCTTCCACTGTTGCGTTGACGGAAGAAATGGGGTAACTTTGTCTT
TGTCTTTTTTTTTTTTTGAAAGGCAAAAAAATTCTTAAACTTGACAGATACCAAGGCTCTCACACCCT
TGATACCAATATGCGGTCACTCCTGCGCCTCTATTTTCTCCTTTTTTTTCTTCTTGCCATGGAAAAGC
AAATCTCGACTTTTTTTTTTGAAAGGAGTACGTTCGAACTTCACCGGTATATCGCCATTGAGTTGCTA
ACAACCGAAGTACATATCAGTTGACAGAAGTGCAATATTTCTTGCTTGGGACTTACGTTAGCAGGAAT
TTGTTAAAAAGGGAGTGTAGCAACAACAATAAGAGATAAGGTGCATAAGGGGACCTGAAGC

3'-UTR (reverse complement)

GGCGTGCAAGTATCGGTTCCCGAATTTGGAAAATCGTTGTAGGGGAAACGAATGTGAATGAAGGGTCT
TGACTGCCCATGTTGGAGATTTCTTCTACATGATTTCCTTAACCTGCTCTCTTGGGGTTTGTGTAGGC
GCGGATTGGGATGAGGTCGTGGAGGAAAAAGGGAGAACACATACACACACACTCATCCACCCACTCAC
AAGGGGAACTGCGGGAAAGGGGAAGGGAAGGAAAATAAATACATGGTGAATTCTTCTTTGTTTTTGAA
AAAAAAATCCGAAAATCATTCGGCTTCCTTTTATTTATTTATGTACTTCCTGCAGTCTTCGTCATCGT
TACTGCTGTCACTTGCTACTGCAGCCGTTGTGGTTGCGGGGTTCATTTATGCCTCTCTAGGTAACCCT
AACGGTCATCGCTATGCCCAGTGTGCGGTGTTTTGTGAGTTTACTCGTTTACTCGCTAC

A.2.3 Primer sequences

Constructs of adenosine and thymidine kinases were cloned using the primers listed in Table A.2-1. The primer sequences used to clone the pOPIN constructs at OPPF are not listed here. Other constructs tested for expression were cloned by Dr. Glyn Hemsworth and Cristina Bosch.

Table A.2-1: Primer sequences for the kinases for *E. coli* expression vectors.

Name	**Vector backbone**	**Sequence 5' --> 3'** **(Restriction enzyme site/overhang for cloning)**	**Restriction enzyme/ Overhang**
TbAK-LIC-F	YSBLIC	CACCACCACCACATGTCATCCGCTCCTCTGAGG	Overhang
TbAK-3C-F	YSBLIC3C	CCAGGGACCAGCAATGTCATCCGCTCCTCTGAGG	Overhang
TbAK-Rev	YSBLIC,	GAGGAGAAGGCGCGTTATCAAGGAGAAAAACTGGGCTTC	Overhang
LmTK-F_3C_F	YSBLIC3C	CCAGGGACCAGCAATGTTCCGCGGTCGTATAGAGC	Overhang
LmTK-F_3C_R	YSBLIC3C	GAGGAGAAGGCGCGTTATCACTCTGAGGATGCAGC	Overhang
TcTK-F_3C_F-1	YSBLIC3C	CCAGGGACCAGCAATGTTTGCCTCCCACACG	Overhang
TcTK-F_3C_F-2	YSBLIC3C	CCAGGGACCAGCAATGCTGGAGAACGAATCG	Overhang
TcTK-F_3C_R	YSBLIC3C	GAGGAGAAGGCGCGTTATTAGTGACACTCCAGCCG	Overhang
TKs-3C_M_F	pMAL-C2	ACTGGAATTCATGGGCAGCAGCCATCAT	EcoR1
LmTK-F_M_R	pMAL-C2	CAGTGTCGACTCACTCTGAGGATGCAGC	Sal1
LmTK-dC_M_R	pMAL-C2	CAGTGTCGACTCACTTCGAGTAGCACTCGC	Sal1
TbTK-F_M_R	pMAL-C2	CAGTGTCGACCTAAGTAGTATCAACGGCC	Sal1
TcTK_M_R	pMAL-C2	CAGTGTCGACTTAGTGACACTCCAGCCG	Sal1

Primer sequences used to clone the TriTryp LAP-A constructs and the vectors used for *T. brucei* parasitology are listed in Table A.2-2.

Table A.2-2: Primer sequences for the TriTryp LAP-As.

Name	**Vector backbone**	**Sequence 5' --> 3'** **(Restriction enzyme site/overhang for cloning)**	**Restriction enzyme/ Overhang**
Overexpression in *E. coli*:			
*Tb*LAP_3Cfor	YSBLIC3C	CCAGGGACCAGCAATGCCTACTTTACCCAAGGCC	Overhang
*Tb*LAP_3Crev	YSBLIC3C	GAGGAGAAGGCGCGTTACTACAATTTTCGGAAGTACTCGG	Overhang
*Tc*LAP_3Cfor	YSBLIC3C	CCAGGGACCAGCAATGAACAGACCTCCTGCTACAGATTCA	Overhang
*Tc*LAP_3Crev	YSBLIC3C	GAGGAGAAGGCGCGTTATTATCGTAAATTACGAAGATATT	Overhang
*Lm*LAP_3Cfor	YSBLIC3C	CCAGGGACCAGCAATGTCCGCGATGAAGCGCCCG	Overhang
*Lm*LAP_3Crev	YSBLIC3C	GAGGAGAAGGCGCGTTACTAGTGCTGGCGCAGAAACTCG	Overhang
Parasitology:			
ECCFor	pGRV23	GCCATATGATGCCTACTTTACCCAAGGCC	Nde1
ECCRev	pGRV23	GCGGATCCCTACAATTTTCGGAAGTACTCGG	BamH1
cmycFor	pGRV33	GCCATATGATGCCTACTTTACCCAAGGCC	Nde1
cmycRev	pGRV33	GCGTTAACCTACAATTTTCGGAAGTACTCGG	Hpa1
RNAiFor	pGR19	GCGGATCCAAGCTTTACCCAAGGCCGAAGCAAAA	BamH1,
RNAiRev	pGR19	GCGTTAACGGGCCCCGTGCTACAGCAGCAGCGAT	Hpa1, Apa1
5UTRFor	pHD887	CGCGGCCGCGGGAGGCGGTAATGACAAGTC	Not1
5UTRRev	pHD887	GCCTCGAGGCTTCAGGTCCCCTTATGCAC	Xho1
3UTRFor	pHD887	GCAGGCCTGGCGTGCAAGTATCGGTTCCC	Stu1
3UTRRev	pHD887	GGCTAGCGCGGCCGCGTAGCGAGTAAACGAGTAAACTCAC	Nhe1, Not1

A.2.3 List of constructs

Table A.2-3: List of constructs for protein production.

OE-Ec = overexpression in *E. coli*, OE-Bv = overexpression with the baculovirus system. Constructs marked with an asterisk (*) were cloned by Dr. Glyn Hemsworth, constructs marked with two asterisks (**) were cloned by Cristina Bosch.

Name	Protein	Amino acid range	Vector backbone	Application
Adenosine kinase (Chapter 2)				
*Tb*AK-LIC	*Tb*AK	2-345	YSBLIC	OE-Ec
*Tb*AK-3C	*Tb*AK	2-345	YSBLIC3C	OE-Ec
Thymidine kinases (Chapter 3)				
*Lm*TK-F_3C	*Lm*TK	2-284	YSBLIC3C	OE-Ec
*Tc*TK-A_3C	*Tc*TK	2-313	YSBLIC3C	OE-Ec
*Tc*TK-B_3C	*Tc*TK	45-313	YSBLIC3C	OE-Ec
*Lm*TK-F_M	*Lm*TK	2-284	pMAL-C2	OE-Ec
*Lm*TK-dC_M	*Lm*TK	2-178	pMAL-C2	OE-Ec
*Tb*TK-F_M	*Tb*TK	2-455	pMAL-C2	OE-Ec
*Tc*TK_M	*Tc*TK	45-313	pMAL-C2	OE-Ec
LmTK-F_28**	*Lm*TK	2-284	pET28	OE-Ec
LmTKdC_3C*	*Lm*TK	2-178	YSBLIC3C	OE-Ec
LmTKdC_LIC*	*Lm*TK	2-178	YSBLIC	OE-Ec
TbTK-F_3C*	*Tb*TK	2-455	YSBLIC3C	OE-Ec
TbTK-F_LIC*	*Tb*TK	2-455	YSBLIC	OE-Ec
TbTKdC_3C*	*Tb*TK	2-382	YSBLIC3C	OE-Ec
TbTKdC_LIC*	*Tb*TK	2-382	YSBLIC	OE-Ec
TbTK-D1_3C*	*Tb*TK	2-179	YSBLIC3C	OE-Ec
TbTK-D1_LIC*	*Tb*TK	2-179	YSBLIC	OE-Ec
TbTK-D2_3C*	*Tb*TK	198-382	YSBLIC3C	OE-Ec
TbTK-D2_LIC*	*Tb*TK	198-382	YSBLIC	OE-Ec
TcTK-A_28**	*Tc*TK	2-313	pET28	OE-Ec
TcTK-TdC_3C*	*Tc*TK	87-225	YSBLIC3C	OE-Ec
TcTK-TdC_LIC*	*Tc*TK	87-225	YSBLIC	OE-Ec
Leucyl aminopeptidases (Chapter 4)				
*Tb*LAP_LIC	*Tb*LAP-A	2-521	YSBLIC	OE-Ec
*Tb*LAP_3C	*Tb*LAP-A	2-521	YSBLIC3C	OE-Ec
*Tc*LAP_3C	*Tc*LAP-A	2-520	YSBLIC3C	OE-Ec
*Lm*LAP_3C	*Lm*LAP-A	2-538	YSBLIC3C	OE-Ec

Table A.2-4: List of constructs for protein production, continued.

OE-Bv = overexpression with the baculovirus system.

Name	Protein	Amino acid range	Vector backbone	Application
Thymidine kinases (Chapter 3)				
11318	*Lm*TK	2-284	pOPIN-E	OE-Bv
11319	*Lm*TK	2-185	pOPIN-E	OE-Bv
11320	*Lm*TK	2-185	pOPIN-F	OE-Bv
11435	*Tc*TK	2-313	pOPIN-F	OE-Bv
11438	*Tc*TK	2-228	pOPIN-F	OE-Bv
11441	*Tc*TK	45-313	pOPIN-F	OE-Bv
11444	*Tc*TK	45-228	pOPIN-F	OE-Bv
11433	*Tc*TK	2-313	pOPIN-I	OE-Bv
11436	*Tc*TK	2-228	pOPIN-I	OE-Bv
11439	*Tc*TK	45-313	pOPIN-I	OE-Bv
11442	*Tc*TK	45-228	pOPIN-I	OE-Bv
11434	*Tc*TK	2-313	pOPIN-E	OE-Bv
11437	*Tc*TK	2-228	pOPIN-E	OE-Bv
11440	*Tc*TK	45-313	pOPIN-E	OE-Bv
11443	*Tc*TK	45-228	pOPIN-E	OE-Bv
11335	*Tb*TK	2-455	pOPIN-F	OE-Bv
11338	*Tb*TK	2-382	pOPIN-F	OE-Bv
11341	*Tb*TK	2-180	pOPIN-F	OE-Bv
11344	*Tb*TK	198-455	pOPIN-F	OE-Bv
11347	*Tb*TK	198-382	pOPIN-F	OE-Bv
11333	*Tb*TK	2-455	pOPIN-I	OE-Bv
11336	*Tb*TK	2-382	pOPIN-I	OE-Bv
11339	*Tb*TK	2-180	pOPIN-I	OE-Bv
11342	*Tb*TK	198-455	pOPIN-I	OE-Bv
11345	*Tb*TK	198-382	pOPIN-I	OE-Bv
11334	*Tb*TK	2-455	pOPIN-E	OE-Bv
11337	*Tb*TK	2-382	pOPIN-E	OE-Bv
11340	*Tb*TK	2-180	pOPIN-E	OE-Bv
11343	*Tb*TK	198-455	pOPIN-E	OE-Bv
11346	*Tb*TK	198-382	pOPIN-E	OE-Bv
Thymidylate kinases (Appendix A1)				
11350	*Tb*TMPK	2-345	pOPIN-F	OE-Bv
11353	*Tb*TMPK	120-345	pOPIN-F	OE-Bv
11348	*Tb*TMPK	2-345	pOPIN-I	OE-Bv
11351	*Tb*TMPK	120-345	pOPIN-I	OE-Bv
11349	*Tb*TMPK	2-345	pOPIN-E	OE-Bv
11352	*Tb*TMPK	120-345	pOPIN-E	OE-Bv
11362	*Tc*TMPK	2-351	pOPIN-F	OE-Bv
11365	*Tc*TMPK	122-351	pOPIN-F	OE-Bv
11360	*Tc*TMPK	2-351	pOPIN-I	OE-Bv
11363	*Tc*TMPK	122-351	pOPIN-I	OE-Bv
11361	*Tc*TMPK	2-351	pOPIN-E	OE-Bv
11364	*Tc*TMPK	122-351	pOPIN-E	OE-Bv
11356	*Lm*TMPK	2-363	pOPIN-F	OE-Bv
11359	*Lm*TMPK	131-363	pOPIN-F	OE-Bv
11354	*Lm*TMPK	2-363	pOPIN-I	OE-Bv
11357	*Lm*TMPK	131-363	pOPIN-I	OE-Bv
11355	*Lm*TMPK	2-363	pOPIN-E	OE-Bv
11358	*Lm*TMPK	131-363	pOPIN-E	OE-Bv

Table A.2-5: List of constructs for parasitological work.

KO = knockout in *T. brucei*, RNAi = RNA interfecerence in *T. brucei*, OE-Tb = overexpression in *T. brucei*, OE-myc-Tb = overexpression of cMyc-tagged protein in *T. brucei*.

Name	Protein	Amino acid/ base pair range	Vector backbone	Application
Leucyl aminopeptidase (Chapter 4)				
ECC	*Tb*LAP-A	2-521	pGRV23	OE-Tb
cmyc	*Tb*LAP-A	2-521	pGRV33	OE-myc-Tb
RNAi	(*Tb*LAP-A)	bp 10-475	pGR19	RNAi
5UTR	-	bp -598 - -62	pHD887B	KO
3UTR	-	bp +7 - +473	pHD887H	KO2

A.2.4 Plasmid maps

A.2.4.1 YSBLIC and YSBLIC3C

For overexpression in *E. coli* the sequences of the proteins of interest (POI) were cloned into YSBLIC and YSBLIC3C vectors between the LIC-for and LIC-rev sites.

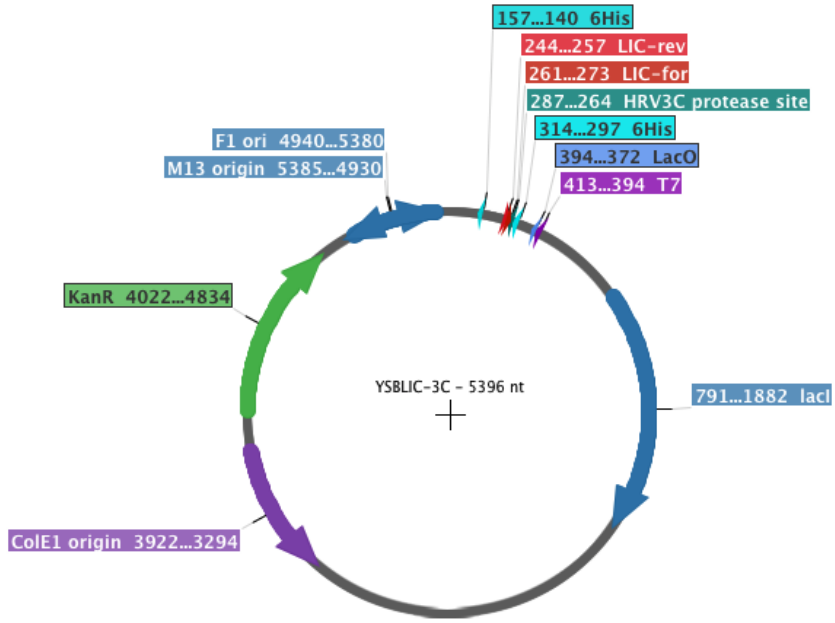


Figure A.2-1: Plasmid map of YSBLIC3C vector.

This and all subsequent plasmid maps were drawn using Serial cloner (version 2.6.1, from RealBasic). KanR = Kanamycin resistance.

A.2.4.2 pOPIN-I, -E and –F

The pOPIN vectors (plasmids for OPPF In-fusion, (Berrow *et al.*, 2007) are ampicillin resistant plasmids with identical backbones, their only differences being the different tags fused t the protein of interest (POI). The three pOPIN vectors used for thymidine kinase baculovirus generation are shown below.

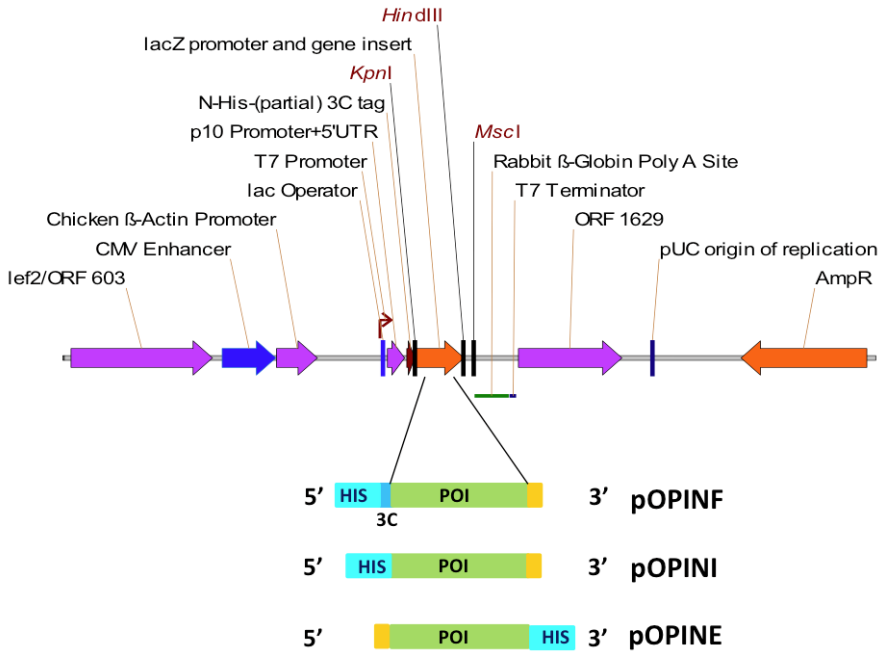


Figure A.2-2: Plasmid map (linearised) of the pOPIN vector family.

The three pOPIN vectors used for the thymidine and thymidylate kinases were pOIPNF, pOPINI and pOPINE. AmpR= Ampicillin resistance.

A.2.4.3 pGRV23b-*Tb*LAP-A and pGRV33b-*Tb*LAP-A

The vectors pGRV23b-*Tb*LAP-A and pGRV33-*Tb*LAP-A for overexpression of LAP-A and cMyc-tagged LAP-A in *T. brucei* parasites are very similar to each other. The major difference is the presence of the cMyc-tag fused to the C-terminal end of *Tb*LAP-A in pGRV33-*Tb*LAP-A (missing in pGRV23b-*Tb*LAP-A).

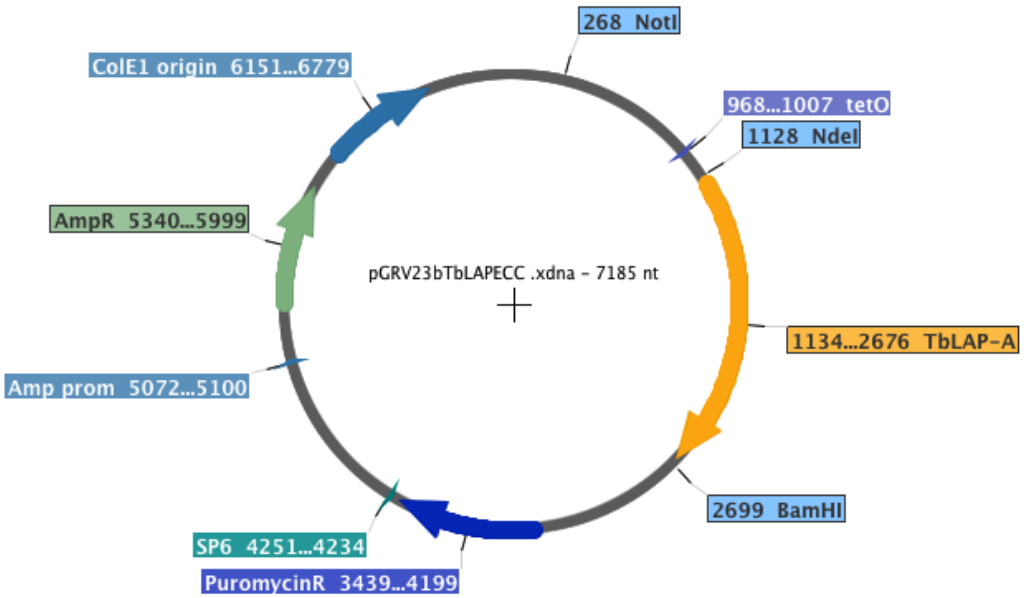


Figure A.2-3: Plasmid map of pGRV23b-*Tb*LAP-A.

AmpR= Ampicillin resistance. PuromycinR = puromycin resistance.

A.2.4.4 pGR19-*Tb*LAP

The vector for LAP-A RNAi in *T. brucei* parasites encodes the same *Tb*LAP-A fragment in opposite orientation, separated by a DNA linker.

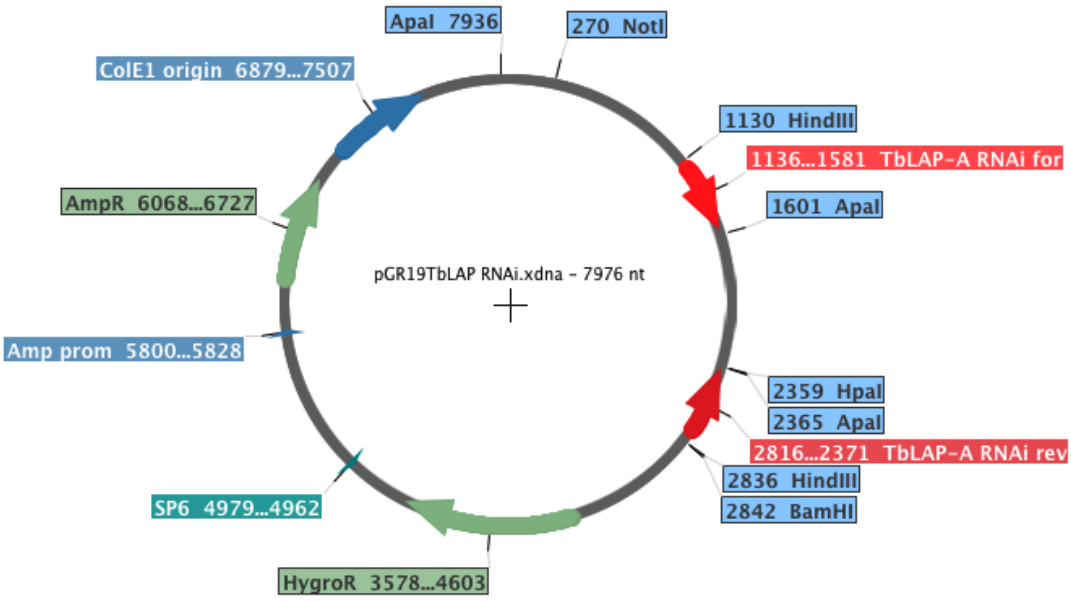


Figure A.2-4: Plasmid map of pGR19-*Tb*LAP-A-RNAi.

The RNAi fragment was cloned into the vector in two different orientations, indicated by the red arrows. HygroR = hygromycin resistance marker, AmpR= Ampicillin resistance.

A.2.4.5 pH887B/UTRs and pH887H/UTRs

For generation of LAP-A double knockouts in *T. brucei* parasites.

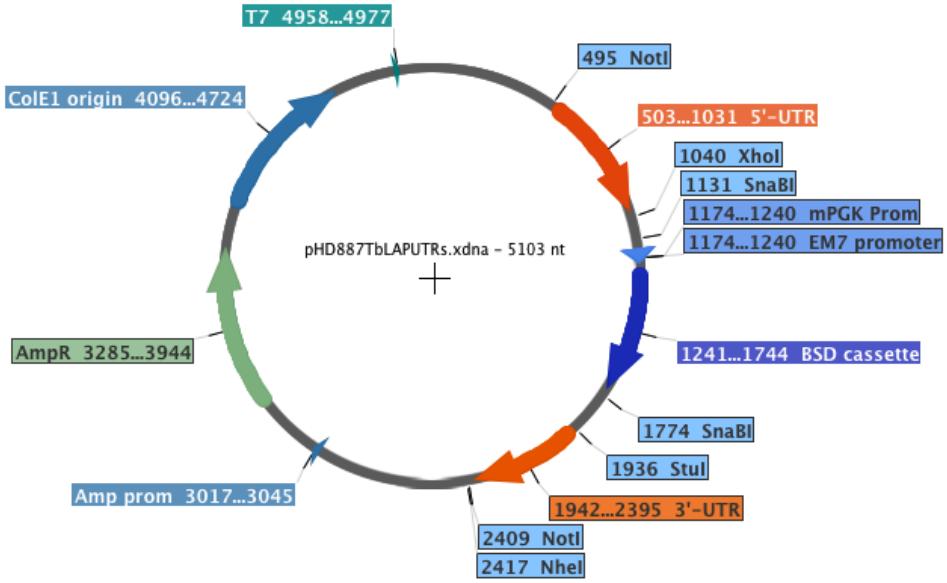


Figure A.2-5: Plasmid map of pHD887B-*Tb*LAP-A-UTRs.

AmpR= Ampicillin resistance, the BSD cassette facilitated blasticidin resistance.

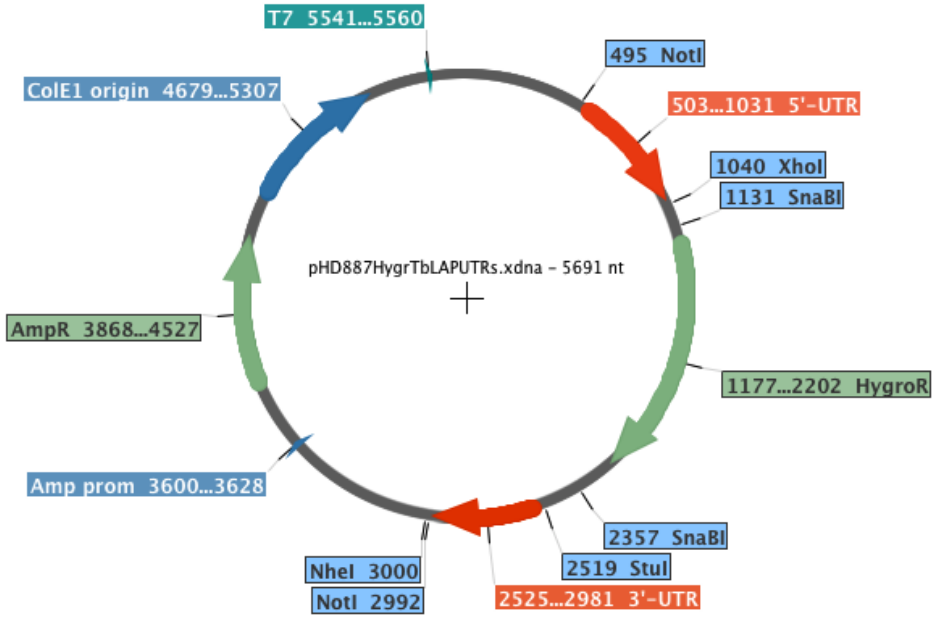


Figure A.2-6: Plasmid map of pHD887H-*Tb*LAP-A-UTRs.

AmpR= Ampicillin resistance, HygroR = hygromycin resistance marker.

A.3 Media composition for *T. brucei* parasite cultures

Table A.3-1: Media composition.

MEM = minimal essential media, MEM –Leu = MEM without leucine, MEM –neaa = MEM without the "Non-essential amino acids". In this and the following table: *Compounds added in form of "Non-essential amino acids" (100x, Sigma M7145). ** Compounds added in form of "MEM Vitamins" (100x, Sigma M6895). Concentrations are given in mg L^{-1} unless stated otherwise.

	HMI-9	SDM-79	MEM	MEM -Leu	MEM -neaa
Salts, etc.					
HEPES	5960	8000	714	714	714
MOPS	-	5000	-	-	-
$CaCl_2 \bullet 2H_2O$	165	180	301	301	301
Fe(NO3)3 • H2O	-	0.14	-	-	-
KCl	330	360	402	402	402
KH_2PO_4	-	-	2**	2**	2**
KNO_3	0.076	-	-	-	-
$MgSO_4$	98	-	-	-	-
$MgSO_4 \bullet 7H_2O$	-	180	200	200	200
NaCl	4500	6120	760	760	760
$NaHCO_3$	3020	2000	-	-	-
$Na_2HPO_4 \bullet H_2O$	-	-	12**	12**	12**
$NaH_2PO_4 \bullet H_2O$	125	126	211	211	211
$Na_2SeO_3 \bullet 5H_2O$	0.017	-	-	-	-
Mercaptoethnol	15	-	-	-	-
Bathocuproline disulfonate $\bullet Na_2$	28	-	-	-	-
Amino acids					
DL-Alanine	25	200	9*	9*	-
L-Arginine • HCl	84	215	126	126	126
L-Asparagine	25	8	15*	15*	-
L-Aspartic acid	30	14	13*	13*	-
L-Cysteine	182	0.02	-	-	-
L-Cystine	91	29	24	24	24
L-Glutamine	584	320	292	292	292
L-Glutamic acid	75	24	15*	15*	-
Glycine	30	15	8*	8*	-
L-Histidine • HCl $\bullet H_2O$	42	38	42	42	42
Hydroxyproline	-	2	-	-	-
L-Isoleucine	105	46	52	52	52
L-Leucine	105	54	26	-	26
L-Lysine • HCl	146	72	73	73	73
DL-Methionine	30	85	16	16	16
L-Phenylalanine	66	111	100	100	100
L-Proline	40	615	612*	612*	600
DL-Serine	42	71	11*	11*	-
Taurine	-	160	-	-	-
DL-Threonine	95	394	48	48	48
L-Tryptophan	16	10	10	10	10
L-Tyrosine	104	150	100	100	100
L-Valine	94	42	47	47	47

Table A.3-2: Media composition continued.

***** Dialysed FBS was replaced by non-dialysed FBS for the standard culture media HMI-9 and SDM-79 used initially and the MEM for the stepwise adaptation to glucose-free MEM.**

	HMI-9	SDM-79	MEM	MEM -Leu	MEM -neaa
Vitamins etc.					
p-aminobenzoic acid	-	2.01	-	-	-
Ascorbic acid	-	0.01	-	-	-
DL-alpha-tocopherol-Na_2PO_4	-	0.002	-	-	-
Vitamin B-12	0.013	-	-	-	-
Biotin	0.013	0.2	-	-	-
Choline chloride	4	0.8	1.0**	1.0**	1.0**
Calciferol		0.02	-	-	-
D-Pantothenic Acid •½Ca	4	0.7	1.0**	1.0**	1.0**
Folic acid	4	4.7	1.0**	1.0**	1.0**
L-glutathione (reduced)	-	0.01	-	-	-
Menadione	-	0.002	-	-	-
myo-Inositol	7.2	1.41	2.0**	2.0**	2.0**
Niacine	-	0.005	-	-	-
Niacinamide	4	0.705	1.0**	1.0**	1.0**
Pyridoxal • HCl	4	0.705	1.0**	1.0**	1.0**
Pyridoxine • HCl	-	0.005	-	-	-
Riboflavin	0.4	0.072	0.1**	0.1**	0.1**
Thiamine • HCl	4	0.702	1.0**	1.0**	1.0**
Vitamin A acetate	-	0.028	-	-	-
Other compounds					
Sodium acetate	-	10	-	-	-
Pyruvate • Na	114	100	-	-	-
Adenine		2	-	-	-
Adenosine		10	12.22	12.22	12.22
AMP • H_2O		0.04	-	-	-
ATP • Na_2 • $2H_2O$		0.2	-	-	-
Guanine • HCl • $2H_2O$		0.06	-	-	-
Guanosine		10	-	-	-
Hypoxanthine	136	0.06	-	-	-
Ornithine • HCl			10.00	10.00	10.00
Thymine		0.06	-	-	-
Uracil	10	0.06	-	-	-
Cytosine	10		-	-	-
Xanthine • Na		0.06	-	-	-
Cholesterol		0.04	-	-	-
Tween 80		4	-	-	-
Phenol red • Na	15	11	10.00	10.00	10.00
Glucose	4500	1000	-	-	-
Glucosamine • HCl	-	50	-	-	-
D-ribose	-	0.1	-	-	-
Deoxy-D-ribose	-	0.1	-	-	-
Haemin [µg mL^{-1}]	-	7.5	7.5	7.5	7.5
FBS (non-dialysed)	-	-	1 %	1 %	1 %
dialysed FBS***	10 %	10 %	9 %	9 %	9 %

Table A.3-3: Composition of "Non-essential amino acids". (100x, Sigma M7145), in g L^{-1}.

L-Alanine (free base)	0.89
L-Asparagine (free base)	1.5
L-Aspartic acid (anhyd.)	1.33
L-Glutamic acid	1.47
Glycine	0.75
L-Proline	1.15
L-Serine	1.05

Table A.3-4: Composition of "MEM Vitamins" (100x, Sigma M6895), in g L^{-1}.

Choline chloride	0.1
Folic acid	0.1
Myo-Inositol	0.2
Niacinamide	0.1
D-Pantothenic Acid •½Ca	0.1
Pyridoxal • HCl	0.1
Riboflavin	0.01
Thiamine • HCl	0.1
KCl	0.2
KH_2PO_4	0.2
NaCl	8.0
Na_2HPO_4	1.15

HMI-9 and SDM-79 media were obtained from the central facilities of the Instituto de Parasitologia y Biomedicina "Lopez-Neyra", Granada, Spain. The MEM media were prepared manually.

List of Abbreviations

aa	Amino acid residue
*Ac*MNPV	*Autographa californica* multiple nuclear polyhydrosis virus
ADA	Adenosine deaminase
Ado	Adenosine
ADP	Adenosine diphophate
AK	Adenosine kinase
AMC	4-methyl-7-coumarinylamide/7-amido-4-methylcoumarin
AMPK	Adenylate kinase
AppNHp	β , γ -imidoadenosine 5’-triphosphate
AP_5A	Adenosine (5’) pentaphospho (5’) adenosine
AP_4dT	P1-(5'-adenosyl)-P5-(5'-thymidyl) tetraphosphate
AP_5dT	P1-(5'-adenosyl)-P5-(5'-thymidyl) pentaphosphate
ATP	Adenosine tryphosphate
AU	Asymmetric unit
AZT	3’-azido-3’deoxythymidine
BB2	Actinonin
BES	Bestatin
BF	Bloodstream form (*T. brucei*)
BME	β -mercapto-ethanol
bp	Base pairs
BSD	Gene for blasticidin deaminase
CIT	Citrate
CL	Cutaneous Leishmaniasis

CNS	Central nervous system
Cyt	Cytosine
Da	Dalton
dAdo	Deoxyadenosine
DAPI	4’,6-diamidino-2-phenylindole
dC	Deoxycytosine
dG	Deoxyguanosine
dIno	Deoxyinosine
DLS	Dynamic light scattering or Diamond light source
dKO	Double knockout
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
Dox	Doxycycline
dThd	Thymidine
dTMP	Thymidine monophosphate
DTT	Dithiothreitol
dTTP	Thymidine triphosphate
DTU	Discrete typing units
dUrd	Deoxyuridine
dUTPase	Deoxyuridine triphosphate nucleotidohydrolase
ECC	Ectopic copy control
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ESI-MS	Electrospray ionisation mass spectrometry
FBS	Foetal bovine serum

Guo	Guanosine
HAT	Human African trypanosomiasis
HRP	Horseradish peroxidase
HygroR	Hygromycin resistance protein
ICCB	Intensity correlation coefficient-based
IPTG	Isopropyl thio- β -D-galactoside
KRM	4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine
LAP	Leucyl aminopeptidase
LB	Luria-Bertani media
LIC	Ligation-independent cloning
Lm	*Leishmania major*
MCL	Mucocutaneous Leishmaniasis
MEM	Minimal essential media
MOA	Mechanism of action
MPD	2-Methyl-2,4-pentanediol
M_W	Molecular weight
NCS	Non-crystallographic symmetry
NTDs	Neglected tropical diseases
OE	Overexpression
OPPF	Oxford protein production facility
ORF	Open reading frame
PBS	Phosphate buffered saline
PDB	Protein data bank
PEG	Polyethylene glycol
PF	Procyclic form (*T. brucei*)

pI	Isoelectric point
PKDL	Post-kala-azar dermal leishmaniasis
PMSF	Phenylmethanesulfonylfluoride
POI	Protein of interest
rcf	Relative centrifugal force
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC-MALLS	Size exclusion chromatography multi angle laser light scattering
sKO	Single knockout
SMT	Sterol C-24-methyltransferase
SSM	Secondary structure matching
Tb	*Trypanosoma brucei*
Tc	*Trypanosoma cruzi*
TMPK	Thymidylate kinase
TK	Thymidine kinase
TriTryps	The three trypanosomatids *T. brucei*, *T. cruzi* and *Leishmania spp*.
Urd	Uridine
UTR	Un-translated region (upstream (5') and downstream (3') of a gene)
VAT	Variant antigenic type
VL	Visceral Leishmaniasis
VSG	Variable surface glycoprotein
WHO	World Health Organisation

References

Acosta, D., Cancela, M., Piacenza, L., Roche, L., Carmona, C. & Tort, J. F. (2008). *Molecular and Biochemical Parasitology* **158**, 52-64.

Agarwal, R. P., Sagar, S. M. & Parks, R. E., Jr. (1975). *Biochemical pharmacology* **24**, 693-701.

Airhart, M. J., Robbins, C. M., Knudsen, T. B., Church, J. K. & Skalko, R. G. (1996). *Teratology* **53**, 361-373.

Aiyedun, B. A., Williamson, J. & Amodu, A. A. (1973). *Acta tropica* **30**, 276-278.

Akopyants, N. S., Kimblin, N., Secundino, N., Patrick, R., Peters, N., Lawyer, P., Dobson, D. E., Beverley, S. M. & Sacks, D. L. (2009). *Science* **324**, 265-268.

Ali, J. A., Creek, D. J., Burgess, K., Allison, H. C., Field, M. C., Maser, P. & De Koning, H. P. (2013). *Molecular pharmacology* **83**, 439-453.

Allen, M. P., Yamada, A. H. & Carpenter, F. H. (1983). *Biochemistry* **22**, 3778-3783.

Alvar, J., Aparicio, P., Aseffa, A., Den Boer, M., Canavate, C., Dedet, J. P., Gradoni, L., Ter Horst, R., Lopez-Velez, R. & Moreno, J. (2008). *Clinical microbiology reviews* **21**, 334-359, table of contents.

Anderson, N. H., Devlin, J. P., Jones, S., Ollis, W. D. & Thorpe, J. E. (1975). *Journal of the Chemical Society. Perkin transactions 1*, 852-857.

Anderson, N. H., Ollis, W. D., Thorpe, J. E. & Ward, A. D. (1975). *Journal of the Chemical Society. Perkin transactions 1*, 825-830.

Andersson, L., Isley, T. C. & Wolfenden, R. (1982). *Biochemistry* **21**, 4177-4180.

Aoki, J. I., Yamashiro-Kanashiro, E. H., Ramos, D. C. & Cotrim, P. C. (2009). *Parasitology research* **104**, 223-228.

Aoyagi, T., Tobe, H., Kojima, F., Hamada, M., Takeuchi, T. & Umezawa, H. (1978). *The Journal of antibiotics* **31**, 636-638.

Araujo, A. F., de Alencar, B. C., Vasconcelos, J. R., Hiyane, M. I., Marinho, C. R., Penido, M. L., Boscardin, S. B., Hoft, D. F., Gazzinelli, R. T. & Rodrigues, M. M. (2005). *Infection and immunity* **73**, 6017-6025.

Araujo, C. A., Araujo, A. A., Batista, C. L., Oliveira, M. A., Oliveira, V., Lino Junior, R. S., Vinaud, M. C. & Bezerra, J. C. (2011). *Parasitology research* **108**, 547-551.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. & Notredame, C. (2006). *Nucleic Acids Research* **34**, W604-608.

Arner, E. S. & Eriksson, S. (1995). *Pharmacology & Therapeutics* **67**, 155-186.

Artymiuk, P. J., Grindley, H. M., Park, J. E., Rice, D. W. & Willett, P. (1992). *FEBS letters* **303**, 48-52.

Arvai, A. (2012). *ADXV - a program to display X-ray diffraction images,* http://www.scripps.edu/~arvai/adxv.html.

Askonas, B. A., Corsini, A. C., Clayton, C. E. & Ogilvie, B. M. (1979). *Immunology* **36**, 313-321.

Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Research* **18**, 6069-6074.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Logan, F. J., Miller, J. A., Mitra, S., Myler, P. J., Nayak, V., Pennington, C., Phan, I., Pinney, D. F.,

Ramasamy, G., Rogers, M. B., Roos, D. S., Ross, C., Sivam, D., Smith, D. F., Srinivasamoorthy, G., Stoeckert, C. J., Jr., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G. & Wang, H. (2010). *Nucleic Acids Research* **38**, D457-462.

Aufderklamm, S., Todenhofer, T., Gakis, G., Kruck, S., Hennenlotter, J., Stenzl, A. & Schwentner, C. (2012). *Cancer letters* **316**, 6-10.

Authie, E., Muteti, D. K., Mbawa, Z. R., Lonsdale-Eccles, J. D., Webster, P. & Wells, C. W. (1992). *Molecular and Biochemical Parasitology* **56**, 103-116.

Baker, J. O., Wilkes, S. H., Bayliss, M. E. & Prescott, J. M. (1983). *Biochemistry* **22**, 2098-2103.

Balaban, N., Waithaka, H. K., Njogu, A. R. & Goldman, R. (1995). *The Journal of infectious diseases* **172**, 845-850.

Barthel, H., Cleij, M. C., Collingridge, D. R., Hutchinson, O. C., Osman, S., He, Q., Luthra, S. K., Brady, F., Price, P. M. & Aboagye, E. O. (2003). *Cancer research* **63**, 3791-3798.

Basombrio, M. A. & Besuschio, S. (1982). *Infection and immunity* **36**, 351-356.

Basombrio, M. A., Besuschio, S. & Cossio, P. M. (1982). *Infection and immunity* **36**, 342-350.

Bastos, I. M., Motta, F. N., Charneau, S., Santana, J. M., Dubost, L., Augustyns, K. & Grellier, P. (2010). *Microbes and infection / Institut Pasteur* **12**, 457-466.

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. (2011). *Acta crystallographica. Section D, Biological crystallography* **67**, 271-281.

Baxter, N. J., Blackburn, G. M., Marston, J. P., Hounslow, A. M., Cliff, M. J., Bermel, W., Williams, N. H., Hollfelder, F., Wemmer, D. E. & Waltho, J. P. (2008). *Journal of the American Chemical Society* **130**, 3952-3958.

Baxter, N. J., Olguin, L. F., Golicnik, M., Feng, G., Hounslow, A. M., Bermel, W., Blackburn, G. M., Hollfelder, F., Waltho, J. P. & Williams, N. H. (2006). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 14732-14737.

Behari, J., Stagon, L. & Calderwood, S. B. (2001). *Journal of bacteriology* **183**, 178-188.

Beninga, J., Rock, K. L. & Goldberg, A. L. (1998). *Journal of Biolological Chemistry* **273**, 18734-18742.

Bente, M., Harder, S., Wiesgigl, M., Heukeshoven, J., Gelhaus, C., Krause, E., Clos, J. & Bruchhaus, I. (2003). *Proteomics* **3**, 1811-1829.

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. (2007). *Nucleic Acids Research* **35**, e45.

Bhosale, M., Pande, S., Kumar, A., Kairamkonda, S. & Nandi, D. (2010). *Biochemical and biophysical research communications* **395**, 76-81.

Birch, P. L., el-Obeid, H. A. & Akhtar, M. (1972). *Archives of biochemistry and biophysics* **148**, 447-451.

Birkus, G., Gibbs, C. S. & Cihlar, T. (2003). *Journal of viral hepatitis* **10**, 50-54.

Birringer, M. S., Claus, M. T., Folkers, G., Kloer, D. P., Schulz, G. E. & Scapozza, L. (2005). *FEBS letters* **579**, 1376-1382.

Bland, N. D., Wang, C., Tallman, C., Gustafson, A. E., Wang, Z., Ashton, T. D., Ochiana, S. O., McAllister, G., Cotter, K., Fang, A. P., Gechijian, L., Garceau,

N., Gangurde, R., Ortenberg, R., Ondrechen, M. J., Campbell, R. K. & Pollastri, M. P. (2011). *Journal of Medicinal Chemistry* **54**, 8188-8194.
Bolte, S. & Cordelieres, F. P. (2006). *Journal of microscopy* **224**, 213-232.
Bonay, P. & Cohen, B. E. (1983). *Biochimica et biophysica acta* **731**, 222-228.
Botbol, V. & Scornik, O. A. (1991). *Journal of Biological Chemistry* **266**, 2151-2157.
Bringaud, F., Barrett, M. P. & Zilberstein, D. (2012). *Frontiers in bioscience* **17**, 349-374.
Bringaud, F., Riviere, L. & Coustou, V. (2006). *Molecular and Biochemical Parasitology* **149**, 1-9.
Brockman, R. W., Schabel, F. M., Jr. & Montgomery, J. A. (1977). *Biochemical pharmacology* **26**, 2193-2196.
Bruley-Rosset, M., Florentin, I., Kiger, N., Schulz, J. & Mathe, G. (1979). *Immunology* **38**, 75-83.
Brun, R., Blum, J., Chappuis, F. & Burri, C. (2010). *Lancet* **375**, 148-159.
Brun, R. & Schonenberger (1979). *Acta tropica* **36**, 289-292.
Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. (2013). *Nucleic Acids Research* **41**, W349-357.
Burleigh, B. A., Caler, E. V., Webster, P. & Andrews, N. W. (1997). *The Journal of cell biology* **136**, 609-620.
Burley, S. K., David, P. R., Taylor, A. & Lipscomb, W. N. (1990). *Proceedings of the National Academy of Sciences of the United States of America* **87**, 6878-6882.
Cadavid-Restrepo, G., Gastardelo, T. S., Faudry, E., de Almeida, H., Bastos, I. M., Negreiros, R. S., Lima, M. M., Assumpcao, T. C., Almeida, K. C., Ragno, M., Ebel, C., Ribeiro, B. M., Felix, C. R. & Santana, J. M. (2011). *BMC biochemistry* **12**, 46.
Camargo, E. P. (2009). *Memorias do Instituto Oswaldo Cruz* **104 Suppl 1**, 275-280.
Canepa, G. E., Silber, A. M., Bouvier, L. A. & Pereira, C. A. (2004). *FEMS microbiology letters* **236**, 79-84.
Cappiello, M., Alterio, V., Amodeo, P., Del Corso, A., Scaloni, A., Pedone, C., Moschini, R., De Donatis, G. M., De Simone, G. & Mura, U. (2006). *Biochemistry* **45**, 3226-3234.
Carpenter, F. H. & Vahl, J. M. (1973). *Journal of Biological Chemistry* **248**, 294-304.
Carter, N. S., Yates, P., Arendt, C. S., Boitz, J. M. & Ullman, B. (2008). *Advances in experimental medicine and biology* **625**, 141-154.
Cassera, M. B., Ho, M. C., Merino, E. F., Burgos, E. S., Rinaldo-Matthis, A., Almo, S. C. & Schramm, V. L. (2011). *Biochemistry* **50**, 1885-1893.
Castillo-Acosta, V. M., Aguilar-Pereyra, F., Vidal, A. E., Navarro, M., Ruiz-Perez, L. M. & Gonzalez-Pacanowska, D. (2012). *The international journal of biochemistry & cell biology* **44**, 1555-1568.
Castillo-Acosta, V. M., Estevez, A. M., Vidal, A. E., Ruiz-Perez, L. M. & Gonzalez-Pacanowska, D. (2008). *The international journal of biochemistry & cell biology* **40**, 2901-2913.
Cazorla, S. I., Frank, F. M., Becker, P. D., Arnaiz, M., Mirkin, G. A., Corral, R. S., Guzman, C. A. & Malchiodi, E. L. (2010). *The Journal of infectious diseases* **202**, 136-144.

Centers for Disease Control and Prevention (2014a). *Parasites - Leishmaniasis,* http://www.cdc.gov/parasites/leishmaniasis/.
Centers for Disease Control and Prevention (2014b). *Parasites - African Trypanosomiasis,* http://www.cdc.gov/parasites/sleepingsickness/.
Centers for Disease Control and Prevention (2014c). *Parasites - American Trypanosomiasis,* http://www.cdc.gov/parasites/chagas/.
Chaerkady, R. & Sharma, K. K. (2004). *Investigative ophthalmology & visual science* **45**, 1214-1223.
Chaillet, P., Marcais-Collado, H., Costentin, J., Yi, C. C., De La Baume, S. & Schwartz, J. C. (1983). *European journal of pharmacology* **86**, 329-336.
Chamond, N., Gregoire, C., Coatnoan, N., Rougeot, C., Freitas-Junior, L. H., da Silveira, J. F., Degrave, W. M. & Minoprio, P. (2003). *Journal of Biological Chemistry* **278**, 15484-15494.
Chan, W. W., Dennis, P., Demmer, W. & Brand, K. (1982). *Journal of Biological Chemistry* **257**, 7955-7957.
Chao, K. S. (2006). *Seminars in oncology* **33**, S59-63.
Chao, W. S., Gu, Y. Q., Pautot, V. V., Bray, E. A. & Walling, L. L. (1999). *Plant physiology* **120**, 979-992.
Chen, D. Z., Patel, D. V., Hackbarth, C. J., Wang, W., Dreyer, G., Young, D. C., Margolis, P. S., Wu, C., Ni, Z. J., Trias, J., White, R. J. & Yuan, Z. (2000). *Biochemistry* **39**, 1256-1262.
Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta crystallographica. Section D, Biological crystallography* **66**, 12-21.
Cheng, Y. C. & Prusoff, W. H. (1973). *Biochemistry* **12**, 2612-2619.
Chevenet, F., Brun, C., Banuls, A. L., Jacq, B. & Christen, R. (2006). *BMC bioinformatics* **7**, 439.
Choi, J. Y., Plummer, M. S., Starr, J., Desbonnet, C. R., Soutter, H., Chang, J., Miller, J. R., Dillman, K., Miller, A. A. & Roush, W. R. (2012). *J Med Chem* **55**, 852-870.
Chou, B., Hiromatsu, K., Hisaeda, H., Duan, X., Imai, T., Murata, S., Tanaka, K. & Himeno, K. (2010). *Biochemical and biophysical research communications* **392**, 277-282.
Clayton, C. E., Estevez, A. M., Hartmann, C., Alibu, V. P., Field, M. & Horn, D. (2005). *Methods in molecular biology* **309**, 39-60.
Clayton, C. E., Sacks, D. L., Ogilvie, B. M. & Askonas, B. A. (1979). *Parasite immunology* **1**, 241-249.
Cleland, W. W. & Hengge, A. C. (2006). *Chemical reviews* **106**, 3252-3278.
Cliff, M. J., Bowler, M. W., Varga, A., Marston, J. P., Szabo, J., Hounslow, A. M., Baxter, N. J., Blackburn, G. M., Vas, M. & Waltho, J. P. (2010). *Journal of the American Chemical Society* **132**, 6507-6516.
Contreras, V. T., Salles, J. M., Thomas, N., Morel, C. M. & Goldenberg, S. (1985). *Molecular and Biochemical Parasitology* **16**, 315-327.
Cook, W. J., DeLucas, L. J. & Chattopadhyay, D. (2000). *Protein Sci* **9**, 704-712.
Cornelissen, A. W., Johnson, P. J., Kooter, J. M., Van der Ploeg, L. H. & Borst, P. (1985). *Cell* **41**, 825-832.
Costes, S. V., Daelemans, D., Cho, E. H., Dobbin, Z., Pavlakis, G. & Lockett, S. (2004). *Biophysical journal* **86**, 3993-4003.

Cottrell, G. S., Hooper, N. M. & Turner, A. J. (2000). *Biochemistry* **39**, 15121-15128.
Coustou, V., Biran, M., Breton, M., Guegan, F., Riviere, L., Plazolles, N., Nolan, D., Barrett, M. P., Franconi, J. M. & Bringaud, F. (2008). *Journal of Biological Chemistry* **283**, 16342-16354.
Cowtan, K. (2006). *Acta crystallographica. Section D, Biological crystallography* **62**, 1002-1011.
Darlyuk, I., Goldman, A., Roberts, S. C., Ullman, B., Rentsch, D. & Zilberstein, D. (2009). *Journal of Biological Chemistry* **284**, 19800-19807.
Das, B. B., Sen, N., Dasgupta, S. B., Ganguly, A., Das, R. & Majumder, H. K. (2006). *The Indian journal of medical research* **123**, 221-232.
Datta, A. K., Bhaumik, D. & Chatterjee, R. (1987). *Journal of Biological Chemistry* **262**, 5515-5521.
de Alencar, B. C., Araujo, A. F., Penido, M. L., Gazzinelli, R. T. & Rodrigues, M. M. (2007). *Vaccine* **25**, 6018-6027.
De Greef, C. & Hamers, R. (1994). *Molecular and Biochemical Parasitology* **68**, 277-284.
de Koning, H. P., Bridges, D. J. & Burchmore, R. J. (2005). *FEMS microbiology reviews* **29**, 987-1020.
de Koning, H. P., Gould, M. K., Sterk, G. J., Tenor, H., Kunz, S., Luginbuehl, E. & Seebeck, T. (2012). *The Journal of infectious diseases* **206**, 229-237.
de Koning, H. P. & Jarvis, S. M. (1997). *Molecular and Biochemical Parasitology* **89**, 245-258.
de Langen, A. J., Klabbers, B., Lubberink, M., Boellaard, R., Spreeuwenberg, M. D., Slotman, B. J., de Bree, R., Smit, E. F., Hoekstra, O. S. & Lammertsma, A. A. (2009). *European journal of nuclear medicine and molecular imaging* **36**, 389-395.
Dick, A. J., Matheson, A. T. & Wang, J. H. (1970). *Canadian journal of biochemistry* **48**, 1181-1188.
Doehl, J. (2013). thesis, University of York.
Dong, L., Cheng, N., Wang, M. W., Zhang, J., Shu, C. & Zhu, D. X. (2005). *Microbiology* **151**, 2017-2023.
Dou, Q. P. & Pardee, A. B. (1996). *Progress in nucleic acid research and molecular biology* **53**, 197-217.
Drew, M. E., Morris, J. C., Wang, Z., Wells, L., Sanchez, M., Landfear, S. M. & Englund, P. T. (2003). *Journal of Biological Chemistry* **278**, 46596-46600.
Duan, X., Yonemitsu, Y., Chou, B., Yoshida, K., Tanaka, S., Hasegawa, M., Tetsutani, K., Ishida, H., Himeno, K. & Hisaeda, H. (2009). *Vaccine* **27**, 6154-6159.
Dumas, C., Ouellette, M., Tovar, J., Cunningham, M. L., Fairlamb, A. H., Tamar, S., Olivier, M. & Papadopoulou, B. (1997). *The EMBO journal* **16**, 2590-2598.
Dumonteil, E., Escobedo-Ortegon, J., Reyes-Rodriguez, N., Arjona-Torres, A. & Ramirez-Sierra, M. J. (2004). *Infection and immunity* **72**, 46-53.
Durocher, Y., Perret, S. & Kamen, A. (2002). *Nucleic Acids Research* **30**, E9.
Eickhoff, C. S., Giddings, O. K., Yoshida, N. & Hoft, D. F. (2010). *Memorias do Instituto Oswaldo Cruz* **105**, 687-691.
Eickhoff, C. S., Vasconcelos, J. R., Sullivan, N. L., Blazevic, A., Bruna-Romero, O., Rodrigues, M. M. & Hoft, D. F. (2011). *PLoS Neglected Tropical Diseases* **5**, e983.

el Kouni, M. H., Diop, D. & Cha, S. (1983). *Proceedings of the National Academy of Sciences of the United States of America* **80**, 6667-6670.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501.
Engman, D. M. & Leon, J. S. (2002). *Acta tropica* **81**, 123-132.
Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N. & Overath, P. (2007). *Cell* **131**, 505-515.
Eriksson, S., Munch-Petersen, B., Johansson, K. & Eklund, H. (2002). *Cellular and Molecular Life Sciences* **59**, 1327-1346.
Esser, K. M., Schoenbechler, M. J. & Gingrich, J. B. (1982). *Journal of immunology* **129**, 1715-1718.
Estevez, A. M., Kempf, T. & Clayton, C. (2001). *The EMBO journal* **20**, 3831-3839.
Evans, K. J. & Kedzierski, L. (2012). *Journal of tropical medicine* **2012**, 892817.
Evans, P. (2006). *Acta crystallographica. Section D, Biological crystallography* **62**, 72-82.
Evans, P. R. (2011). *Acta crystallographica. Section D, Biological crystallography* **67**, 282-292.
Evans, P. R. & Murshudov, G. N. (2013). *Acta crystallographica. Section D, Biological crystallography* **69**, 1204-1214.
Feasey, N., Wansbrough-Jones, M., Mabey, D. C. & Solomon, A. W. (2010). *British medical bulletin* **93**, 179-200.
Field, M. C. & Carrington, M. (2009). *Nature reviews. Microbiology* **7**, 775-786.
Fogg, M. J., Alzari, P., Bahar, M., Bertini, I., Betton, J. M., Burmeister, W. P., Cambillau, C., Canard, B., Corrondo, M. A., Coll, M., Daenke, S., Dym, O., Egloff, M. P., Enguita, F. J., Geerlof, A., Haouz, A., Jones, T. A., Ma, Q., Manicka, S. N., Migliardi, M., Nordlund, P., Owens, R. J., Peleg, Y., Schneider, G., Schnell, R., Stuart, D. I., Tarbouriech, N., Unge, T., Wilkinson, A. J., Wilmanns, M., Wilson, K. S., Zimhony, O. & Grimes, J. M. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1196-1207.
Frank, F. M., Petray, P. B., Cazorla, S. I., Munoz, M. C., Corral, R. S. & Malchiodi, E. L. (2003). *Vaccine* **22**, 77-86.
Frick, L. & Wolfenden, R. (1985). *Biochimica et biophysica acta* **829**, 311-318.
Garg, N. & Tarleton, R. L. (2002). *Infection and immunity* **70**, 5547-5555.
Geiser, F., Luscher, A., de Koning, H. P., Seebeck, T. & Maser, P. (2005). *Molecular pharmacology* **68**, 589-595.
Gentry, G. A. (1992). *Pharmacology & Therapeutics* **54**, 319-355.
Giannousis, P. P. & Bartlett, P. A. (1987). *Journal of Medicinal Chemistry* **30**, 1603-1609.
Gibson, W. C. (1986). *Parasitology today* **2**, 255-257.
Gibson, W. C. (1989). *Parasitology* **99 Pt 3**, 391-402.
Giddings, O. K., Eickhoff, C. S., Sullivan, N. L. & Hoft, D. F. (2010). *Infection and immunity* **78**, 1333-1338.
Gilbert, R. J., Klein, R. A. & Miller, P. G. (1983). *Comparative biochemistry and physiology. B, Comparative biochemistry* **74**, 277-281.
Gillerman, I. & Fischer, B. (2011). *Journal of Medicinal Chemistry* **54**, 107-121.
Ginger, M. L., Prescott, M. C., Reynolds, D. G., Chance, M. L. & Goad, L. J. (2000). *European journal of biochemistry / FEBS* **267**, 2555-2566.

Gordon, J. J., Kelly, B. K. & Miller, G. A. (1962). *Nature* **195**, 701-702.
Gossage, S. M., Rogers, M. E. & Bates, P. A. (2003). *International journal for parasitology* **33**, 1027-1034.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Research* **31**, 3320-3323.
Gourlay, S. C. & Colloms, S. D. (2004). *Molecular microbiology* **52**, 53-65.
Gray, D. P., Grever, M. R., Siaw, M. F., Coleman, M. S. & Balcerzak, S. P. (1982). *Cancer treatment reports* **66**, 253-257.
Grellier, P., Vendeville, S., Joyeau, R., Bastos, I. M., Drobecq, H., Frappier, F., Teixeira, A. R., Schrevel, J., Davioud-Charvet, E., Sergheraert, C. & Santana, J. M. (2001). *Journal of Biological Chemistry* **276**, 47078-47086.
Gu, Y. Q. & Walling, L. L. (2002). *European journal of biochemistry / FEBS* **269**, 1630-1640.
Hammond, D. J. & Gutteridge, W. E. (1984). *Molecular and Biochemical Parasitology* **13**, 243-261.
Harbut, M. B., Velmourougane, G., Dalal, S., Reiss, G., Whisstock, J. C., Onder, O., Brisson, D., McGowan, S., Klemba, M. & Greenbaum, D. C. (2011). *Proceedings of the National Academy of Sciences of the United States of America* **108**, E526-534.
Hawkins, C. F. & Bagnara, A. S. (1987). *Biochemistry* **26**, 1982-1987.
Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hubner, C. G. & Kern, D. (2007). *Nature* **450**, 838-844.
Hidalgo-Zarco, F., Camacho, A. G., Bernier-Villamor, V., Nord, J., Ruiz-Perez, L. M. & Gonzalez-Pacanowska, D. (2001). *Protein Science* **10**, 1426-1433.
Hirumi, H. & Hirumi, K. (1994). *Parasitology today* **10**, 80-84.
Hofer, A., Crona, M., Logan, D. T. & Sjoberg, B. M. (2012). *Critical reviews in biochemistry and molecular biology* **47**, 50-63.
Hofer, A., Steverding, D., Chabes, A., Brun, R. & Thelander, L. (2001). *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6412-6416.
Hoft, D. F., Eickhoff, C. S., Giddings, O. K., Vasconcelos, J. R. & Rodrigues, M. M. (2007). *Journal of immunology* **179**, 6889-6900.
Horn, D. & McCulloch, R. (2010). *Current opinion in microbiology* **13**, 700-705.
Howarth, J. & Lloyd, D. G. (2000). *The Journal of antimicrobial chemotherapy* **46**, 625-628.
Hu, C. M. & Chang, Z. F. (2007). *Journal of biomedical science* **14**, 491-497.
Hu, C. M. & Chang, Z. F. (2008). *Cancer research* **68**, 2831-2840.
Hu, C. M., Yeh, M. T., Tsao, N., Chen, C. W., Gao, Q. Z., Chang, C. Y., Lee, M. H., Fang, J. M., Sheu, S. Y., Lin, C. J., Tseng, M. C., Chen, Y. J. & Chang, Z. F. (2012). *Cancer cell* **22**, 36-50.
Issa, V. S. & Bocchi, E. A. (2010). *Lancet* **376**, 768; author reply 768-769.
Jardim, A., Funk, V., Caprioli, R. M. & Olafson, R. W. (1995). *The Biochemical journal* **305 (Pt 1)**, 307-313.
Jia, H., Nishikawa, Y., Luo, Y., Yamagishi, J., Sugimoto, C. & Xuan, X. (2010). *Molecular and Biochemical Parasitology* **170**, 1-6.

Jia, H., Terkawi, M. A., Aboge, G. O., Goo, Y. K., Luo, Y., Li, Y., Yamagishi, J., Nishikawa, Y., Igarashi, I., Sugimoto, C., Fujisaki, K. & Xuan, X. (2009). *Parasitology* **136**, 945-952.

Johansson, K., Ramaswamy, S., Ljungcrantz, C., Knecht, W., Piskur, J., Munch-Petersen, B., Eriksson, S. & Eklund, H. (2001). *Nat Struct Biol* **8**, 616-620.

Jordan, A. & Reichard, P. (1998). *Annual review of biochemistry* **67**, 71-98.

Junqueira, C., Caetano, B., Bartholomeu, D. C., Melo, M. B., Ropert, C., Rodrigues, M. M. & Gazzinelli, R. T. (2010). *Expert reviews in molecular medicine* **12**, e29.

Kabsch, W. (2010a). *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132.

Kabsch, W. (2010b). *Acta crystallographica. Section D, Biological crystallography* **66**, 133-144.

Kale, A., Pijning, T., Sonke, T., Dijkstra, B. W. & Thunnissen, A. M. (2010). *J Mol Biol* **398**, 703-714.

Kamhawi, S., Belkaid, Y., Modi, G., Rowton, E. & Sacks, D. (2000). *Science* **290**, 1351-1354.

Kandeel, M., Kato, A., Kitamura, Y. & Kitade, Y. (2009). *Nucleic acids symposium series*, 283-284.

Ke, P. Y. & Chang, Z. F. (2004). *Molecular and cellular biology* **24**, 514-526.

Kettner, C., Glover, G. I. & Prescott, J. M. (1974). *Archives of biochemistry and biophysics* **165**, 739-743.

Kierszenbaum, F. (1999). *Clinical microbiology reviews* **12**, 210-223.

Kim, H., Burley, S. K. & Lipscomb, W. N. (1993). *Journal of Molecular Biology* **230**, 722-724.

Kim, H. & Lipscomb, W. N. (1993). *Biochemistry* **32**, 8465-8478.

Kim, J. K., Natarajan, S., Park, H., Huynh, K. H., Lee, S. H., Kim, J. G., Ahn, Y. J. & Kang, L. W. (2013). *Journal of microbiology* **51**, 627-632.

Kinyua, J. K., Nguu, E. K., Mulaa, F. & Ndung'u, J. M. (2005). *Vaccine* **23**, 3824-3828.

Kline, J. P. & Larson, R. A. (2005). *Expert opinion on pharmacotherapy* **6**, 2711-2718.

Kloetzel, P. M. & Ossendorp, F. (2004). *Current opinion in immunology* **16**, 76-81.

Knecht, W., Petersen, G. E., Munch-Petersen, B. & Piskur, J. (2002). *Journal of Molecular Biology* **315**, 529-540.

Knowles, G. (1993). *The Journal of antimicrobial chemotherapy* **32**, 172-174.

Kosinska, U., Carnrot, C., Eriksson, S., Wang, L. & Eklund, H. (2005). *The FEBS journal* **272**, 6365-6372.

Kosinska, U., Carnrot, C., Sandrini, M. P., Clausen, A. R., Wang, L., Piskur, J., Eriksson, S. & Eklund, H. (2007). *The FEBS journal* **274**, 727-737.

Krassner, S. M. & Flory, B. (1972). *The Journal of protozoology* **19**, 682-685.

Krieger, S., Schwarz, W., Ariyanayagam, M. R., Fairlamb, A. H., Krauth-Siegel, R. L. & Clayton, C. (2000). *Molecular microbiology* **35**, 542-552.

Krissinel, E. & Henrick, K. (2004). *Acta crystallographica. Section D, Biological crystallography* **60**, 2256-2268.

Krissinel, E. & Henrick, K. (2007). *Journal of Molecular Biology* **372**, 774-797.

Kuettel, S., Greenwald, J., Kostrewa, D., Ahmed, S., Scapozza, L. & Perozzo, R. (2011). *PLoS Neglected Tropical Diseases* **5**, e1164.

Kuettel, S., Mosimann, M., Maser, P., Kaiser, M., Brun, R., Scapozza, L. & Perozzo, R. (2009). *PLoS Neglected Tropical Diseases* **3**, e506.

Kuettel, S., Zambon, A., Kaiser, M., Brun, R., Scapozza, L. & Perozzo, R. (2007). *Journal of Medicinal Chemistry* **50**, 5833-5839.

La Greca, F. & Magez, S. (2011). *Human vaccines* **7**, 1225-1233.

Lalmanach, G., Boulange, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F. & Authie, E. (2002). *Biological chemistry* **383**, 739-749.

Lamour, N., Riviere, L., Coustou, V., Coombs, G. H., Barrett, M. P. & Bringaud, F. (2005). *Journal of Biological Chemistry* **280**, 11902-11910.

Laskowski, R. A. & Swindells, M. B. (2011). *Journal of chemical information and modeling* **51**, 2778-2786.

Leahey, A., Kelly, K., Rorke, L. B. & Lange, B. (1997). *Journal of pediatric hematology/oncology* **19**, 304-308.

Lefurgey, A., Gannon, M., Blum, J. & Ingram, P. (2005). *The Journal of eukaryotic microbiology* **52**, 277-289.

Leibly, D. J., Abendroth, J., Bryan, C. M., Sankaran, B., Kelley, A., Barrett, L. K., Stewart, L. & Van Voorhis, W. C. (2011). *Acta crystallographica. Section F, Structural biology and crystallization communications* **67**, 1090-1094.

Li, Q., Lau, A., Morris, T. J., Guo, L., Fordyce, C. B. & Stanley, E. F. (2004). *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24**, 4070-4081.

Li, S. Q., Fung, M. C., Reid, S. A., Inoue, N. & Lun, Z. R. (2007). *Parasite immunology* **29**, 191-199.

Li, S. Q., Yang, W. B., Ma, L. J., Xi, S. M., Chen, Q. L., Song, X. W., Kang, J. & Yang, L. Z. (2009). *Parasitology research* **104**, 429-435.

Li, Y., Zhang, Y. & Yan, H. (1996). *Journal of Biological Chemistry* **271**, 28038-28044.

Limon-Flores, A. Y., Cervera-Cetina, R., Tzec-Arjona, J. L., Ek-Macias, L., Sanchez-Burgos, G., Ramirez-Sierra, M. J., Cruz-Chan, J. V., VanWynsberghe, N. R. & Dumonteil, E. (2010). *Vaccine* **28**, 7414-7419.

Lingner, J. & Keller, W. (1993). *Nucleic Acids Research* **21**, 2917-2920.

Linstead, D. J., Klein, R. A. & Cross, G. A. (1977). *Journal of general microbiology* **101**, 243-251.

Lonsdale-Eccles, J. D. & Grab, D. J. (2002). *Trends in parasitology* **18**, 17-19.

Lubega, G. W., Byarugaba, D. K. & Prichard, R. K. (2002). *Experimental parasitology* **102**, 9-22.

Lubega, G. W., Ochola, D. O. & Prichard, R. K. (2002). *Experimental parasitology* **102**, 134-142.

Luscher, A., Onal, P., Schweingruber, A. M. & Maser, P. (2007). *Antimicrobiological Agents and Chemotherapy* **51**, 3895-3901.

Lynx, M. D., Bentley, A. T. & McKee, E. E. (2006). *Biochemical pharmacology* **71**, 1342-1348.

Lynx, M. D. & McKee, E. E. (2006). *Biochemical pharmacology* **72**, 239-243.

Machado, A. V., Cardoso, J. E., Claser, C., Rodrigues, M. M., Gazzinelli, R. T. & Bruna-Romero, O. (2006). *Human gene therapy* **17**, 898-908.

Magez, S., Caljon, G., Tran, T., Stijlemans, B. & Radwanska, M. (2010). *Parasitology* **137**, 2017-2027.

Maggioli, G., Acosta, D., Silveira, F., Rossi, S., Giacaman, S., Basika, T., Gayo, V., Rosadilla, D., Roche, L., Tort, J. & Carmona, C. (2011). *Vaccine* **29**, 9057-9063.

Maggiora, L. L., Orawski, A. T. & Simmons, W. H. (1999). *Journal of Medicinal Chemistry* **42**, 2394-2402.

Mancilla, R., Naquira, C. & Lanas, C. (1967). *Experimental parasitology* **21**, 154-159.

Marcilla, A., De la Rubia, J. E., Sotillo, J., Bernal, D., Carmona, C., Villavicencio, Z., Acosta, D., Tort, J., Bornay, F. J., Esteban, J. G. & Toledo, R. (2008). *Clinical and vaccine immunology : CVI* **15**, 95-100.

Marin-Neto, J. A., Rassi, A., Jr., Morillo, C. A., Avezum, A., Connolly, S. J., Sosa-Estani, S., Rosas, F., Yusuf, S. & Investigators, B. (2008). *American heart journal* **156**, 37-43.

Maroof, A., Brown, N., Smith, B., Hodgkinson, M. R., Maxwell, A., Losch, F. O., Fritz, U., Walden, P., Lacey, C. N., Smith, D. F., Aebischer, T. & Kaye, P. M. (2012). *The Journal of infectious diseases* **205**, 853-863.

Mathew, Z., Knox, T. M. & Miller, C. G. (2000). *Journal of bacteriology* **182**, 3383-3393.

Mathews, II, Erion, M. D. & Ealick, S. E. (1998). *Biochemistry* **37**, 15607-15620.

Mathews, C. K. (2006). *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **20**, 1300-1314.

Matsushima, M., Takahashi, T., Ichinose, M., Miki, K., Kurokawa, K. & Takahashi, K. (1991). *Biochemical and biophysical research communications* **178**, 1459-1464.

Maudlin, I., Turner, M. J., Dukes, P. & Miller, N. (1984). *Acta tropica* **41**, 253-257.

Mazareb, S., Fu, Z. Y. & Zilberstein, D. (1999). *Experimental parasitology* **91**, 341-348.

McCoy, A. J. (2007). *Acta crystallographica. Section D, Biological crystallography* **63**, 32-41.

McGowan, S., Oellig, C. A., Birru, W. A., Caradoc-Davies, T. T., Stack, C. M., Lowther, J., Skinner-Adams, T., Mucha, A., Kafarski, P., Grembecka, J., Trenholme, K. R., Buckle, A. M., Gardiner, D. L., Dalton, J. P. & Whisstock, J. C. (2010). *Proceedings of the National Academy of Sciences of the United States of America* **107**, 2449-2454.

McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. (2011). *Acta crystallographica. Section D, Biological crystallography* **67**, 386-394.

Miyahira, Y., Takashima, Y., Kobayashi, S., Matsumoto, Y., Takeuchi, T., Ohyanagi-Hara, M., Yoshida, A., Ohwada, A., Akiba, H., Yagita, H., Okumura, K. & Ogawa, H. (2005). *Infection and immunity* **73**, 7356-7365.

Mkunza, F., Olaho, W. M. & Powell, C. N. (1995). *Vaccine* **13**, 151-154.

Molina-Arcas, M., Casado, F. J. & Pastor-Anglada, M. (2009). *Current vascular pharmacology* **7**, 426-434.

Morty, R. E. & Morehead, J. (2002). *Journal of Biological Chemistry* **277**, 26057-26065.

Muchmore, S. W., Smith, R. A., Stewart, A. O., Cowart, M. D., Gomtsyan, A., Matulenko, M. A., Yu, H., Severin, J. M., Bhagwat, S. S., Lee, C. H., Kowaluk, E. A., Jarvis, M. F. & Jakob, C. L. (2006). *Journal of Medicinal Chemistry* **49**, 6726-6731.

Munch-Petersen, B. (1984). *Molecular and cellular biochemistry* **64**, 173-185.
Munch-Petersen, B. (2009). *The FEBS journal* **276**, 571-580.
Munch-Petersen, B., Cloos, L., Tyrsted, G. & Eriksson, S. (1991). *Journal of Biological Chemistry* **266**, 9032-9038.
Munch-Petersen, B., Tyrsted, G. & Cloos, L. (1993). *Journal of Biological Chemistry* **268**, 15621-15625.
Murray, P. K., Jennings, F. W., Murray, M. & Urquhart, G. M. (1974a). *Immunology* **27**, 815-824.
Murray, P. K., Jennings, F. W., Murray, M. & Urquhart, G. M. (1974b). *Immunology* **27**, 825-840.
Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367.
Mutahir, Z., Clausen, A. R., Andersson, K. M., Wisen, S. M., Munch-Petersen, B. & Piskur, J. (2013). *The FEBS journal* **280**, 1531-1541.
Nagata, Y., Mizutani, S., Nomura, S., Kurauchi, O., Kasugai, M. & Tomoda, Y. (1991). *Enzyme* **45**, 165-173.
Nagill, R. & Kaur, S. (2011). *International immunopharmacology* **11**, 1464-1488.
Nakai, K. & Horton, P. (1999). *Trends in biochemical sciences* **24**, 34-36.
Nankya-Kitaka, M. F., Curley, G. P., Gavigan, C. S., Bell, A. & Dalton, J. P. (1998). *Parasitology research* **84**, 552-558.
Nantulya, V. M., Doyle, J. J. & Jenni, L. (1980). *Parasitology* **80**, 133-137.
Nicholls, R. A., Long, F. & Murshudov, G. N. (2012). *Acta crystallographica. Section D, Biological crystallography* **68**, 404-417.
Noazin, S., Modabber, F., Khamesipour, A., Smith, P. G., Moulton, L. H., Nasseri, K., Sharifi, I., Khalil, E. A., Bernal, I. D., Antunes, C. M., Kieny, M. P. & Tanner, M. (2008). *Vaccine* **26**, 6759-6767.
Nordwig, A. & Mayer, H. (1973). *Hoppe-Seyler's Zeitschrift fur physiologische Chemie* **354**, 380-383.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *Journal of Molecular Biology* **302**, 205-217.
Oberholzer, M., Marti, G., Baresic, M., Kunz, S., Hemphill, A. & Seebeck, T. (2007). *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **21**, 720-731.
Ogbunude, P. O. & Ikediobi, C. O. (1982). *Acta tropica* **39**, 219-224.
Ohuchi, S., Suda, H., Naganawa, H., Kawamura, K., Aoyagi, T. & Umezawa (1984). *The Journal of antibiotics* **37**, 1741-1743.
Okwor, I. & Uzonna, J. E. (2009). *Immunotherapy* **1**, 765-776.
Olenick, J. G., Wolff, R., Nauman, R. K. & McLaughlin, J. (1988). *Infection and immunity* **56**, 92-98.
Olliaro, P. L., Guerin, P. J., Gerstl, S., Haaskjold, A. A., Rottingen, J. A. & Sundar, S. (2005). *The Lancet. Infectious diseases* **5**, 763-774.
Ooi, C. P. & Bastin, P. (2013). *Frontiers in cellular and infection microbiology* **3**, 71.
Ostermann, N., Schlichting, I., Brundiers, R., Konrad, M., Reinstein, J., Veit, T., Goody, R. S. & Lavie, A. (2000). *Structure* **8**, 629-642.
Palella, T. D., Andres, C. M. & Fox, I. H. (1980). *Journal of Biological Chemistry* **255**, 5264-5269.

Park, J. & Gupta, R. S. (2008). *Cellular and Molecular Life Sciences* **65**, 2875-2896.
Paul, S. & Summers, D. (2004). *Plasmid* **52**, 63-68.
Peacock, L., Ferris, V., Bailey, M. & Gibson, W. (2008). *Molecular and Biochemical Parasitology* **160**, 100-106.
Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M. & Gibson, W. (2011). *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3671-3676.
Pereira, C. A., Alonso, G. D., Ivaldi, S., Silber, A., Alves, M. J., Bouvier, L. A., Flawia, M. M. & Torres, H. N. (2002). *FEBS letters* **526**, 111-114.
Pereira, C. A., Alonso, G. D., Paveto, M. C., Flawia, M. M. & Torres, H. N. (1999). *The Journal of eukaryotic microbiology* **46**, 566-570.
Pereira, C. A., Alonso, G. D., Paveto, M. C., Iribarren, A., Cabanas, M. L., Torres, H. N. & Flawia, M. M. (2000). *Journal of Biological Chemistry* **275**, 1495-1501.
Pereira, C. A., Alonso, G. D., Torres, H. N. & Flawia, M. M. (2002). *The Journal of eukaryotic microbiology* **49**, 82-85.
Perez-Moreno, G., Sealey-Cardona, M., Rodrigues-Poveda, C., Gelb, M. H., Ruiz-Perez, L. M., Castillo-Acosta, V., Urbina, J. A. & Gonzalez-Pacanowska, D. (2012). *International journal for parasitology* **42**, 975-989.
Prechel, M. M., Orawski, A. T., Maggiora, L. L. & Simmons, W. H. (1995). *The Journal of pharmacology and experimental therapeutics* **275**, 1136-1142.
Prescott, J. M. & Wilkes, S. H. (1966). *Archives of biochemistry and biophysics* **117**, 328-336.
Price, H. P., Menon, M. R., Panethymitaki, C., Goulding, D., McKean, P. G. & Smith, D. F. (2003). *Journal of Biological Chemisry* **278**, 7206-7214.
Radwanska, M., Guirnalda, P., De Trez, C., Ryffel, B., Black, S. & Magez, S. (2008). *PLoS pathogens* **4**, e1000078.
Radwanska, M., Magez, S., Dumont, N., Pays, A., Nolan, D. & Pays, E. (2000). *Parasite immunology* **22**, 639-650.
Ranjbarian, F., Vodnala, M., Vodnala, S. M., Rofougaran, R., Thelander, L. & Hofer, A. (2012). *Journal of Biological Chemistry* **287**, 17628-17636.
Raper, J., Portela, M. P., Lugli, E., Frevert, U. & Tomlinson, S. (2001). *Current opinion in microbiology* **4**, 402-408.
Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. (2014). *Nucleic Acids Research* **42**, D503-509.
Reddy, M. C., Palaninathan, S. K., Shetty, N. D., Owen, J. L., Watson, M. D. & Sacchettini, J. C. (2007). *Journal of Biological Chemistry* **282**, 27334-27342.
Redmond, S., Vadivelu, J. & Field, M. C. (2003). *Molecular and Biochemical Parasitology* **128**, 115-118.
Reynes, J. P., Tiraby, M., Baron, M., Drocourt, D. & Tiraby, G. (1996). *Journal of bacteriology* **178**, 2804-2812.
Rich, D. H., Moon, B. J. & Harbeson, S. (1984). *Journal of Medicinal Chemistry* **27**, 417-422.
Richman, D. D. (1990). *The American journal of medicine* **88**, 8S-10S.
Roditi, I. & Lehane, M. J. (2008). *Current opinion in microbiology* **11**, 345-351.
Rohloff, P., Rodrigues, C. O. & Docampo, R. (2003). *Molecular and Biochemical Parasitology* **126**, 219-230.
Rost, B., Yachdav, G. & Liu, J. (2004). *Nucleic Acids Research* **32**, W321-326.

Rosta, E., Kamerlin, S. C. & Warshel, A. (2008). *Biochemistry* **47**, 3725-3735.
Rottenberg, M. E., Masocha, W., Ferella, M., Petitto-Assis, F., Goto, H., Kristensson, K., McCaffrey, R. & Wigzell, H. (2005). *The Journal of infectious diseases* **192**, 1658-1665.
Sadlova, J., Price, H. P., Smith, B. A., Votypka, J., Volf, P. & Smith, D. F. (2010). *Cellular microbiology* **12**, 1765-1779.
Salskov, A., Tammisetti, V. S., Grierson, J. & Vesselle, H. (2007). *Seminars in nuclear medicine* **37**, 429-439.
Sanchez-Burgos, G., Mezquita-Vega, R. G., Escobedo-Ortegon, J., Ramirez-Sierra, M. J., Arjona-Torres, A., Ouaissi, A., Rodrigues, M. M. & Dumonteil, E. (2007). *FEMS immunology and medical microbiology* **50**, 333-341.
Santana, J. M., Grellier, P., Schrevel, J. & Teixeira, A. R. (1997). *The Biochemical journal* **325 (Pt 1)**, 129-137.
Saraiva, E. M., de Figueiredo Barbosa, A., Santos, F. N., Borja-Cabrera, G. P., Nico, D., Souza, L. O., de Oliveira Mendes-Aguiar, C., de Souza, E. P., Fampa, P., Parra, L. E., Menz, I., Dias, J. G., Jr., de Oliveira, S. M. & Palatnik-de-Sousa, C. B. (2006). *Vaccine* **24**, 2423-2431.
Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990). *Trends in biochemical sciences* **15**, 430-434.
Sarid, S., Berger, A. & Katchalski, E. (1962). *Journal of Biological Chemistry* **237**, 2207-2212.
Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. (1999). *Cell* **99**, 167-177.
Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012). *Nature methods* **9**, 671-675.
Schumacher, M. A., Scott, D. M., Mathews, II, Ealick, S. E., Roos, D. S., Ullman, B. & Brennan, R. G. (2000). *J Mol Biol* **298**, 875-893.
Scranton, M. A., Yee, A., Park, S. Y. & Walling, L. L. (2012). *Journal of Biological Chemistry* **287**, 18408-18417.
Segura-Pena, D., Lichter, J., Trani, M., Konrad, M., Lavie, A. & Lutz, S. (2007). *Structure* **15**, 1555-1566.
Segura-Pena, D., Lutz, S., Monnerjahn, C., Konrad, M. & Lavie, A. (2007). *Journal of Molecular Biology* **369**, 129-141.
Shenvi, A. B. (1986). *Biochemistry* **25**, 1286-1291.
Shields, A. F., Lawhorn-Crews, J. M., Briston, D. A., Zalzala, S., Gadgeel, S., Douglas, K. A., Mangner, T. J., Heilbrun, L. K. & Muzik, O. (2008). *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 4463-4468.
Shiota, H., Inoue, S. & Matsumura, K. (1978). *Nippon Ganka Gakkai zasshi* **82**, 144-151.
Simarro, P. P., Franco, J., Diarra, A., Postigo, J. A. & Jannin, J. (2012). *Parasitology* **139**, 842-846.
Simarro, P. P., Franco, J. R., Cecchi, G., Paone, M., Diarra, A., Ruiz Postigo, J. A. & Jannin, J. G. (2012). *Journal of travel medicine* **19**, 44-53.
Singh, A. K., Mukhopadhyay, C., Biswas, S., Singh, V. K. & Mukhopadhyay, C. K. (2012). *PloS one* **7**, e38489.
Sole, V. A., Papillon, E., Cotte, M., Walter, P. & Susini, J. (2007). *Spectrochimica Acta Part B-Atomic Spectroscopy* **62**, 63-68.

Spitznagel, D., Ebikeme, C., Biran, M., Nic a' Bhaird, N., Bringaud, F., Henehan, G. T. & Nolan, D. P. (2009). *The FEBS journal* **276**, 7187-7199.

Stack, C. M., Lowther, J., Cunningham, E., Donnelly, S., Gardiner, D. L., Trenholme, K. R., Skinner-Adams, T. S., Teuscher, F., Grembecka, J., Mucha, A., Kafarski, P., Lua, L., Bell, A. & Dalton, J. P. (2007). *Journal of Biological Chemistry* **282**, 2069-2080.

Stirling, C. J., Colloms, S. D., Collins, J. F., Szatmari, G. & Sherratt, D. J. (1989). *The EMBO journal* **8**, 1623-1627.

Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. (2011). *Bioinformatics* **27**, 3315-3316.

Stoltze, L., Schirle, M., Schwarz, G., Schroter, C., Thompson, M. W., Hersh, L. B., Kalbacher, H., Stevanovic, S., Rammensee, H. G. & Schild, H. (2000). *Nature immunology* **1**, 413-418.

Strater, N. & Lipscomb, W. N. (1995a). *Biochemistry* **34**, 9200-9210.

Strater, N. & Lipscomb, W. N. (1995b). *Biochemistry* **34**, 14792-14800.

Strater, N., Sherratt, D. J. & Colloms, S. D. (1999). *The EMBO journal* **18**, 4513-4522.

Strater, N., Sun, L., Kantrowitz, E. R. & Lipscomb, W. N. (1999). *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11151-11155.

Sundar, S. & Chakravarty, J. (2008). *Expert opinion on investigational drugs* **17**, 787-794.

Susan-Resiga, D., Bentley, A. T., Lynx, M. D., LaClair, D. D. & McKee, E. E. (2007). *Antimicrob Agents Chemother* **51**, 1142-1149.

Sylvester, D. & Krassner, S. M. (1976). *Comparative biochemistry and physiology. B, Comparative biochemistry* **55**, 443-447.

Tait, A., Barry, J. D., Wink, R., Sanderson, A. & Crowe, J. S. (1985). *Parasitology* **90 (Pt 1)**, 89-100.

Tarleton, R. L. & Zhang, L. (1999). *Parasitology today* **15**, 94-99.

Tennant, B. C., Baldwin, B. H., Graham, L. A., Ascenzi, M. A., Hornbuckle, W. E., Rowland, P. H., Tochkov, I. A., Yeager, A. E., Erb, H. N., Colacino, J. M., Lopez, C., Engelhardt, J. A., Bowsher, R. R., Richardson, F. C., Lewis, W., Cote, P. J., Korba, B. E. & Gerin, J. L. (1998). *Hepatology* **28**, 179-191.

Thelander, L. & Reichard, P. (1979). *Annual review of biochemistry* **48**, 133-158.

Thiakaki, M., Rohousova, I., Volfova, V., Volf, P., Chang, K. P. & Soteriadou, K. (2005). *Microbes and infection / Institut Pasteur* **7**, 760-766.

Thiel, M., Harder, S., Wiese, M., Kroemer, M. & Bruchhaus, I. (2008). *Molecular and Biochemical Parasitology* **158**, 152-162.

Tobe, H., Kojima, F., Aoyagi, T. & Umezawa, H. (1980). *Biochimica et biophysica acta* **613**, 459-468.

Tovar, J., Wilkinson, S., Mottram, J. C. & Fairlamb, A. H. (1998). *Molecular microbiology* **29**, 653-660.

Towne, C. F., York, I. A., Neijssen, J., Karow, M. L., Murphy, A. J., Valenzuela, D. M., Yancopoulos, G. D., Neefjes, J. J. & Rock, K. L. (2005). *Journal of immunology* **175**, 6605-6614.

Turzynski, A. & Mentlein, R. (1990). *European journal of biochemistry / FEBS* **190**, 509-515.

Umezawa, H. (1980). *Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer* **75**, 115-125.
Umezawa, H. (1982). *International journal of clinical pharmacology, therapy, and toxicology* **20**, 19-23.
Umezawa, H., Aoyagi, T., Ohuchi, S., Okuyama, A., Suda, H., Takita, T., Hamada, M. & Takeuchi, T. (1983). *The Journal of antibiotics* **36**, 1572-1575.
Umezawa, H., Aoyagi, T., Suda, H., Hamada, M. & Takeuchi, T. (1976). *The Journal of antibiotics* **29**, 97-99.
Umezawa, H., Aoyagi, T., Tanaka, T., Suda, H., Okuyama, A., Naganawa, H., Hamada, M. & Takeuchi, T. (1985). *The Journal of antibiotics* **38**, 1629-1630.
Urwyler, S., Studer, E., Renggli, C. K. & Roditi, I. (2007). *Molecular microbiology* **63**, 218-228.
Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homble, F., Grelard, A., Zhendre, V., Nolan, D. P., Lins, L., Crowet, J. M., Pays, A., Felu, C., Poelvoorde, P., Vanhollebeke, B., Moestrup, S. K., Lyngso, J., Pedersen, J. S., Mottram, J. C., Dufourc, E. J., Perez-Morga, D. & Pays, E. (2013). *Nature* **501**, 430-434.
Vagin, A. & Teplyakov, A. (1997). *Journal of Applied Crystallography* **30**, 1022-1025.
van Griensven, J., Balasegaram, M., Meheus, F., Alvar, J., Lynen, L. & Boelaert, M. (2010). *The Lancet. Infectious diseases* **10**, 184-194.
van Steensel, B., van Binnendijk, E. P., Hornsby, C. D., van der Voort, H. T., Krozowski, Z. S., de Kloet, E. R. & van Driel, R. (1996). *Journal of cell science* **109 (Pt 4)**, 787-792.
Van Wart, H. E. & Lin, S. H. (1981). *Biochemistry* **20**, 5682-5689.
van Weelden, S. W., Fast, B., Vogt, A., van der Meer, P., Saas, J., van Hellemond, J. J., Tielens, A. G. & Boshart, M. (2003). *Journal of Biological Chemistry* **278**, 12854-12863.
van Weelden, S. W., van Hellemond, J. J., Opperdoes, F. R. & Tielens, A. G. (2005). *Journal of Biological Chemistry* **280**, 12451-12460.
Vanhollebeke, B., De Muylder, G., Nielsen, M. J., Pays, A., Tebabi, P., Dieu, M., Raes, M., Moestrup, S. K. & Pays, E. (2008). *Science* **320**, 677-681.
Vickerman, K. (1985). *British medical bulletin* **41**, 105-114.
Vieira, L. L., Lafuente, E., Gamarro, F. & Cabantchik, Z. (1996). *The Biochemical journal* **319 (Pt 3)**, 691-697.
Viotti, R., Vigliano, C., Lococo, B., Alvarez, M. G., Petti, M., Bertocchi, G. & Armenti, A. (2009). *Expert review of anti-infective therapy* **7**, 157-163.
Vodnala, M., Fijolek, A., Rofougaran, R., Mosimann, M., Maser, P. & Hofer, A. (2008). *Journal of Biological Chemistry* **283**, 5380-5388.
Vodnala, S. K., Ferella, M., Lunden-Miguel, H., Betha, E., van Reet, N., Amin, D. N., Oberg, B., Andersson, B., Kristensson, K., Wigzell, H. & Rottenberg, M. E. (2009). *PLoS Neglected Tropical Diseases* **3**, e495.
Vodnala, S. K., Lundback, T., Yeheskieli, E., Sjoberg, B., Gustavsson, A. L., Svensson, R., Olivera, G. C., Eze, A. A., de Koning, H. P., Hammarstrom, L. G. & Rottenberg, M. E. (2013). *Journal of Medicinal Chemistry* **56**, 9861-9873.
Vogt, J., Perozzo, R., Pautsch, A., Prota, A., Schelling, P., Pilger, B., Folkers, G., Scapozza, L. & Schulz, G. E. (2000). *Proteins* **41**, 545-553.

Wainberg, M. A., Sawyer, J. P., Montaner, J. S., Murphy, R. L., Kuritzkes, D. R. & Raffi, F. (2005). *Antiviral therapy* **10**, 13-28.
Walter, R. D. & Ebert, F. (1977). *Hoppe-Seyler's Zeitschrift fur physiologische Chemie* **358**, 23-26.
Wang, L., Sun, R. & Eriksson, S. (2011). *Antimicrobial Agents and Chemotherapy* **55**, 2552-2558.
Wang, Q., Melzer, I. M., Kruse, M., Sander-Juelch, C. & Wiese, M. (2005). *Kinetoplastid biology and disease* **4**, 6.
Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). *Bioinformatics* **20**, 2138-2139.
Welin, M., Kosinska, U., Mikkelsen, N. E., Carnrot, C., Zhu, C., Wang, L., Eriksson, S., Munch-Petersen, B. & Eklund, H. (2004). *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17970-17975.
Whittingham, J. L., Carrero-Lerida, J., Brannigan, J. A., Ruiz-Perez, L. M., Silva, A. P., Fogg, M. J., Wilkinson, A. J., Gilbert, I. H., Wilson, K. S. & Gonzalez-Pacanowska, D. (2010). *The Biochemical journal* **428**, 499-509.
Wild, K., Bohner, T., Folkers, G. & Schulz, G. E. (1997). *Protein Science* **6**, 2097-2106.
Wilkes, S. H. & Prescott, J. M. (1985). *Journal of Biological Chemistry* **260**, 13154-13162.
Williamson, J. (1972). *Transactions of the Royal Society of Tropical Medicine & Hygiene* **66**, 354-355.
Williamson, J. & Macadam, R. F. (1976). *Transactions of the Royal Society of Tropical Medicine & Hygiene* **70**, 130-137.
Wilson, D. K., Rudolph, F. B. & Quiocho, F. A. (1991). *Science* **252**, 1278-1284.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242.
Winter, G. (2010). *Journal of Applied Crystallography* **43**, 186-190.
Wirtz, E., Hoek, M. & Cross, G. A. (1998). *Nucleic Acids Research* **26**, 4626-4634.
Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. (1999). *Molecular and Biochemical Parasitology* **99**, 89-101.
Wittinghofer, A. (1997). *Current biology : CB* **7**, R682-685.
Wolfrum, S., Richardt, G., Dominiak, P., Katus, H. A. & Dendorfer, A. (2001). *British Journal of Pharmacology* **134**, 370-374.
Woolwine, S. C., Sprinkle, A. B. & Wozniak, D. J. (2001). *Journal of bacteriology* **183**, 4674-4679.
World Health Organisation (2013). *Reproduced with permission of the publisher, from Map Production: Control of Neglected tropical diseases.,* http://www.who.int/gho/map_gallery/en/.
World Health Organisation (2014a). *Human African trypanosomiasis* http://www.who.int/entity/trypanosomiasis_african/en/.
World Health Organisation (2014b). *Neglected tropical diseases,* http://www.who.int/neglected_diseases/en/.
World Health Organisation (2014c). *Chagas disease (American trypanosomiasis),* http://www.who.int/neglected_diseases/diseases/chagas/en/.

World Health Organisation (2014d). *Leishmaniasis fact sheet,* http://www.who.int/mediacentre/factsheets/fs375/en/.

Xong, H. V., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P. & Pays, E. (1998). *Cell* **95**, 839-846.

Xu, H., Faber, C., Uchiki, T., Racca, J. & Dealwis, C. (2006). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4028-4033.

Young, J. D., Yao, S. Y., Sun, L., Cass, C. E. & Baldwin, S. A. (2008). *Xenobiotica; the fate of foreign compounds in biological systems* **38**, 995-1021.

Zapata-Estrella, H., Hummel-Newell, C., Sanchez-Burgos, G., Escobedo-Ortegon, J., Ramirez-Sierra, M. J., Arjona-Torres, A. & Dumonteil, E. (2006). *Immunology letters* **103**, 186-191.

Zelada, C., Montemartini, M., Cazzulo, J. J. & Nowicki, C. (1996). *Molecular and Biochemical Parasitology* **79**, 225-228.

Zeledon, R. (1960). *The Journal of parasitology* **46**, 541-551.

Zhang, Y., El Kouni, M. H. & Ealick, S. E. (2006). *Acta crystallographica. Section D, Biological crystallography* **62**, 140-145.

Zhang, Y., El Kouni, M. H. & Ealick, S. E. (2007). *Acta crystallographica. Section D, Biological crystallography* **63**, 126-134.

Zingales, B., Miles, M. A., Campbell, D. A., Tibayrenc, M., Macedo, A. M., Teixeira, M. M., Schijman, A. G., Llewellyn, M. S., Lages-Silva, E., Machado, C. R., Andrade, S. G. & Sturm, N. R. (2012). *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **12**, 240-253.

Bibliography

Timm J., Gonzalez-Pacanowska D. & Wilson K. S. (2014), Structures of adenosine kinase from *Trypanosoma brucei brucei*. Acta Cryst. F70, 34-39.

Acta Crystallographica Section F
Structural Biology Communications
ISSN 2053-230X

Structures of adenosine kinase from *Trypanosoma brucei brucei*

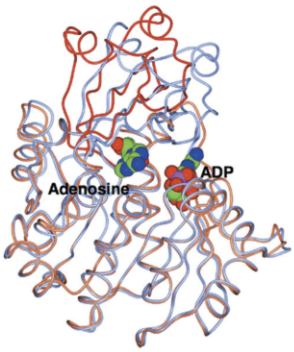
Jennifer Timm,[a] Dolores González-Pacanowska[b] and Keith S. Wilson[a]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, England, and [b]Instituto de Parasitología y Biomedicina 'López-Neyra', Consejo Superior de Investigaciones Científicas, 18100 Armilla, Granada, Spain

Correspondence e-mail: keith.wilson@york.ac.uk

Received 7 November 2013
Accepted 12 December 2013

PDB references: apo *Tb*AK, 4n08; *Tb*AK–adenosine–AMPPNP, 4n09



Trypanosoma brucei is a single-cellular parasite of the genus Kinetoplastida and is the causative agent of African sleeping sickness in humans. Adenosine kinase is a key enzyme in the purine-salvage pathway, phosphorylating adenosine to AMP, and also activates cytotoxic analogues such as cordycepin and Ara-A by their phosphorylation. The structures of *T. brucei brucei* adenosine kinase (*Tb*AK) in its unliganded open conformation and complexed with adenosine and ADP in the closed conformation are both reported to 2.6 Å resolution. The structures give insight into the binding mode of the substrates and the conformational change induced upon substrate binding. This information can be used to guide the improvement of cytotoxic substrate analogues as potential antitrypanosomal drugs.

1. Introduction

Parasites of the genus *Trypanosoma* cause a variety of devastating diseases, mainly in developing countries. *T. brucei* is the causative agent of African sleeping sickness (World Health Organization, 2011). All four current drugs show severe toxicity, and in addition resistance is emerging, so new chemotherapeutic agents are urgently needed and novel targets for drug development need to be identified (Barrett *et al.*, 2007). A key requirement is to identify major differences in the biochemical pathways between trypanosomes and their human hosts. One important difference lies in purine metabolism, since while mammalian cells are able to synthesize purines *de novo*, trypanosomes lack this ability and depend entirely on the purine-salvage pathway (el Kouni, 2003; James & Born, 1980). A key enzyme in the purine-salvage pathway is adenosine kinase (AK; Vodnala *et al.*, 2008), which catalyses the phosphorylation of adenosine to AMP. AK appears to be non-essential for the growth of the parasites, but it activates several cytotoxic adenosine analogues such as cordycepin and Ara-A (Iovannisci & Ullman, 1984). However, a major problem with these compounds as drug candidates is their degradation by adenosine deaminase *in vivo* (Mäser *et al.*, 2001; Williamson, 1972). If this limitation could be evaded by modifications of these compounds that hamper deamination yet do not alter their phosphorylation by AK, they could become candidates for the design of effective antitrypanosomal chemotherapeutics.

AK belongs to the phosphofructokinase B family of carbohydrate kinases, with two characteristic sequence motifs: the di-glycine switch close to the N-terminus and the sequence D*X*NGAGD close to the C-terminus (Park & Gupta, 2008). AKs of several species, including that of *T. brucei rhodesiense* (*Tb*AK; Kuettel *et al.*, 2011), have been well characterized. A number of crystal structures are known, which confirm that the enzymes have an open apoenzyme form but a closed form in complex with ligands and substrates (Cook *et al.*, 2000; Kuettel *et al.*, 2011; Mathews *et al.*, 1998; Muchmore *et al.*, 2006; Reddy *et al.*, 2007; Schumacher *et al.*, 2000; Zhang *et al.*, 2006, 2007). Here, we report the structures of *Tb*AK in the apo form and in complex with adenosine and adenosine diphosphate (*Tb*AK–adenosine–ADP). We show these to be very similar to the published structures of complexes with the activator (*Tb*AK–KRM) and the AP5A bi-substrate mimic (*Tb*AK–AP5A), respectively (Kuettel *et al.*, 2011).

2. Methods

2.1. Cloning, expression and purification

The gene coding for *Tb*AK (SwissProt entry Q584S0) was inserted into the YSBLIC3C vector using ligation-independent cloning (Fogg & Wilkinson, 2008). The construct included an N-terminal hexahistidine tag together with a cleavage site for HRV 3C protease. The cloning was confirmed by colony PCR and sequencing, and the construct was transformed into C41 cells and plated on Luria–Bertani plates containing 35 µg ml^{-1} kanamycin. The next day, the colonies were washed off with 5 ml Luria–Bertani medium and used for the inoculation of 8 × 500 ml medium. Cells were grown at 37°C with shaking at 180 rev min^{-1} (Innova 44, New Brunswick Scientific) until they reached an optical density at 600 nm of 0.2, and were then transferred to 20°C with shaking at 251 rev min^{-1}. Protein expression was induced by the addition of 1 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG) at an optical density at 600 nm of 0.6 for 6 h. The cells were harvested by centrifugation at 5000*g* for 30 min and washed with 100 ml PBS prior to freezing.

The cells were thawed on ice and were lysed in 80 ml lysis buffer *A* (20 m*M* Tris pH 7.5, 200 m*M* NaCl, 30 m*M* imidazole, 10 m*M* *β*-mercaptoethanol, 100 µ*M* PMSF) or buffer *B* (20 m*M* Tris pH 7.5, 200 m*M* NaCl, 30 m*M* imidazole, 1 m*M* DTT) using sonication. Cell debris was removed by centrifugation at 5000*g* and 4°C for 40 min. The supernatant was loaded onto a pre-equilibrated 5 ml HisTrap FF crude column (GE Healthcare), which was then washed with three column volumes of lysis buffer. The protein was eluted using an imidazole gradient. Protein-containing fractions were pooled and concentrated to ∼2 mg ml^{-1} before the addition of HRV 3C protease at a ratio of 1:100 protease:*Tb*AK. Proteolysis was monitored by 12% SDS–PAGE and had reached completion overnight at 4°C.

To remove the histidine tag and noncleaved *Tb*AK, the protein buffer was exchanged into lysis buffer using centrifugal filters (Amicon Ultra Centrifugal Filter from Millipore or Vivaspin 6 from Sartorium Stedim Biotech; 30 kDa cutoff) and nickel-affinity purification was performed using the same buffers as before. The resulting protein was concentrated and loaded onto a Superdex 75 column (GE Healthcare) in buffer consisting of 10 m*M* Tris pH 7.5, 200 m*M* NaCl with 10 m*M* *β*-mercaptoethanol or 1 m*M* DTT as reducing agent. The protein was concentrated to 96 mg ml^{-1}, dialyzed against the size-exclusion chromatography buffer, frozen in liquid nitrogen and stored at −80°C at a concentration of 48 mg ml^{-1}. Its identity was confirmed by electrospray ionization mass spectrometry.

Dynamic light-scattering experiments were carried out using a DynaPro (Protein Solutions) to determine the polydispersity of the samples before crystallization. For this, the protein was diluted to a concentration of 3 mg ml^{-1} at 20°C with a sample volume of 20 µl.

2.2. Crystallization

Crystallization proved to be challenging, with severe problems with regard to reproducibility. Screening was carried out using a Mosquito robot (TTP LabTech, UK) with sitting drops composed of 150 nl protein solution at a concentration of 30 mg ml^{-1} plus an equal volume of precipitant. For the complex, protein purified with buffers containing *β*-mercaptoethanol was screened in the presence of 1 m*M* adenosine and 1 m*M* *β*,*γ*-imidoadenosine 5′-triphosphate (AMPPNP). Crystals were optimized using hanging-drop vapour diffusion and the best crystals grew in 0.1 *M* bis-tris propane pH 8.5, 0.2 *M* sodium citrate, 24%(*w*/*v*) polyethylene glycol 3350 within three weeks to maximum dimensions of ∼600 × 100 × 50 µm.

For the apoenzyme, extensive screening failed to produce any crystals. Following the publication of the structure in the open form (Kuettel *et al.*, 2011), we focused around the conditions reported therein, using a protein sample purified with buffers containing DTT. This led to a large number of crystals, of which 200 were tested for diffraction. However, very few showed any diffraction, and finally only one provided a usable diffraction pattern, with the rest showing very restricted resolution or substantial splitting and disorder. The single usable crystal grew at 4°C in 0.1 *M* Tris pH 8.8, 60%(*v*/*v*) MPD, and grew to a maximum size of approximately 80 × 20 × 20 µm after one week.

Thus, the reducing agent appeared to affect the crystallization behaviour, as the protein purified with *β*-mercaptoethanol crystallized only in the closed conformation (with ligands added), whereas that purified in the presence of DTT crystallized only in the open conformation (without ligands). Furthermore, repeating the purification with the same or with modified buffer conditions failed to yield diffracting crystals even after extensive screening. It is relevant that Kuettel and coworkers were unable to obtain a well diffracting crystal of the apoenzyme (Kuettel *et al.*, 2011). In addition, extensive attempts to crystallize *Tb*AK in the presence of cordycepin or the transition-state analogues ADP–AlF_4–adenosine and ADP–MgF_4–adenosine were also unsuccessful.

2.3. Data collection, structure solution and refinement

Computations were performed using programs from the *CCP*4 suite (Winn *et al.*, 2011) unless otherwise stated. Data were processed using *XDS* (Kabsch, 2010*a*,*b*), *POINTLESS* (Evans, 2011) and *SCALA* (Evans, 2006). The crystal grown in the presence of adenosine and AMPPNP was cryoprotected in mother liquor containing 5% ethylene glycol and data were collected on beamline I24 at the Diamond Light Source (DLS). The crystal belonged to space group *C*2 with four molecules in the asymmetric unit. The unit-cell parameters do not allow twinning.

The apo crystal was not cryoprotected before vitrification and data were collected on beamline I03 at DLS. The crystal appeared to

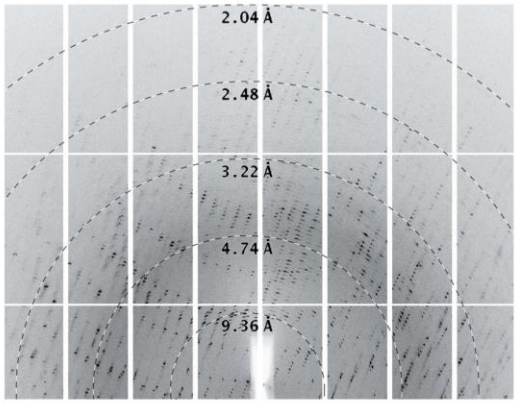


Figure 1
Diffraction pattern from the apo crystal of *Tb*AK. The figure shows a close-up of one of the images used for structure solution, with the rings annotating the resolution created using the *adxv* software (Arvai, 2012). The crystal is clearly split, with more than one lattice evident in the image. While the data appear to extend beyond 2.6 Å resolution on this image, the diffraction was anisotropic and the integration was restricted to 2.6 Å resolution overall.

Table 1
Crystallographic data and refinement statistics.

	Apo *Tb*AK	*Tb*AK–adenosine–AMPPNP
Data collection		
Diffraction source	DLS beamline I03	DLS beamline I24
Wavelength (Å)	0.9763	0.9686
Temperature (K)	100	100
Detector	Pilatus 6M-F	Pilatus 6M Prosport+
Rotation range per image (°)	0.2	0.2
Total rotation range (°)	180	220
Crystal data		
Space group	$P4_12_12$	$C2$
Unit-cell parameters		
a (Å)	60.08	184.53
b (Å)	60.08	90.22
c (Å)	192.86	98.04
α (°)	90.00	90.00
β (°)	90.00	103.78
γ (°)	90.00	90.00
Resolution (Å)	57.36–2.60 (2.74–2.60)	95.21–2.60 (2.74–2.60)
Total reflections	132048 (20863)	160133 (23793)
Unique reflections	11649 (1664)	47962 (6973)
Completeness (%)	99.90 (100)	99.5 (99.4)
Multiplicity	11.3 (12.5)	3.3 (3.4)
R_{merge}	0.120 (0.395)	0.176 (0.991)
$\langle I/\sigma(I)\rangle$	15.8 (6.3)	6.6 (2.3)
Matthews coefficient (Å^3 Da^{-1})	2.28	2.59
Refinement statistics		
Reflections, working set	10989	45507
Reflections, test set	660	2455
Resolution range (Å)	57.36–2.60	95.21–2.60
R factor	0.170	0.186
R_{free}	0.231	0.221
No. of non-H atoms		
Protein	2463	10276
Ligand	—	192
Water	157	148
Mean B factors (Å^2)		
Protein	38.0	45.2
Ligand	—	44.4
Water	40.5	38.3
R.m.s. deviation from ideal		
Bond lengths (Å)	0.013	0.012
Bond angles (°)	1.466	1.531
Ramachandran plot† (%)		
Residues in favoured region	93.6	96.05
Residues in allowed region	5.8	3.58
Outliers	0.6	0.37
MolProbity score	1.35	1.45
Poor rotamers‡ (%)	1.61	2.01
PDB code	4n08	4n09

† The Ramachandran plot was generated with *Coot* (Emsley *et al.*, 2010). ‡ Poor rotamer analysis was performed using *MolProbity* (Chen *et al.*, 2010).

consist of several independent lattices (Fig. 1), which limited the data quality. *XDS* selected one of these lattices and the data were integrated to 2.6 Å resolution in space group $P4_12_12$ with one molecule per asymmetric unit. As expected, the unit-cell parameters are closely similar to those reported for the KRM-soaked crystal by Kuettel and coworkers (PDB entry 2xtb; a = 61.24, c = 193.85 Å, also in space group $P4_12_12$). Analysis during data reduction clearly indicated that the crystal was not twinned, which was confirmed by the low R factors achieved during subsequent refinement in *REFMAC* (Murshudov *et al.*, 2011). The data and statistics for both crystals are summarized in Table 1.

The structures were solved by molecular replacement using *Phaser* (McCoy, 2007). At the time, human AK (PDB entry 1bx4; Mathews *et al.*, 1998), with a protein sequence identity of 40%, showed the highest homology to *Tb*AK and was used as the search model for *Tb*AK–adenosine–ADP. For apo *Tb*AK, the structure of *Tb*AK–KRM was used. Refinement was carried out using *REFMAC5* (Murshudov *et al.*, 2011) with automatically generated local NCS restraints, and the model was rebuilt with *Coot* (Emsley *et al.*, 2010) and validated using *MolProbity* (Chen *et al.*, 2010).

3. Results and discussion

3.1. Characterization and crystallization

*Tb*AK was initially purified using buffers containing β-mercaptoethanol (BME) and PMSF for protease inhibition, but characterization by dynamic light scattering (DLS) revealed a polydispersity of approximately 75%. Analysis by ESI-MS showed peaks of 80 and 156 Da higher than that expected, which could be the result of phosphorylation (80 Da), a BME adduct (76 Da) or a PMSF adduct (156 Da). This batch led to crystals of *Tb*AK in the presence of adenosine and AMPPNP that diffracted to 2.6 Å resolution with four *Tb*AK molecules in the asymmetric unit. All four chains had clear density for both adenosine moieties and the α- and β-phosphates, but no density for a γ-phosphate. This could reflect high flexibility of the γ-phosphate owing to the lack of a coordinating Mg^{2+} ion, but is most likely to be the result of slow hydrolysis. We therefore modelled the ligand as ADP. No molecules of BME or PMSF were visible in the structure.

In a second protein preparation, the buffers were changed to avoid possible BME and PMSF adducts. DLS showed the protein to be monodisperse (10% polydispersity) and ESI-MS indicated a single species with a molecular weight of 38 187 Da (expected value of 38 184 Da). However, attempts to obtain new *Tb*AK–adenosine–AMPPNP crystals failed, as did attempts to obtain co-crystals with cordycepin or with magnesium and aluminium fluoride plus nucleotides. At this point, the structures of *Tb*AK in complex with the bi-substrate analogue diadenosine pentaphosphate (AP5A) and the activator 4-[5-(4-phenoxyphenyl)-2*H*-pyrazol-3-yl]morpholine (KRM) were published (Kuettel *et al.*, 2011). Using these conditions, we obtained a single apo crystal which diffracted to 2.6 Å resolution in space group $P4_12_12$ with one *Tb*AK per asymmetric unit, although the data quality was limited by the presence of more than one lattice.

3.2. Overall fold

The *T. brucei brucei* AK structure is highly similar to that of the enzyme from a closely related subspecies, *T. brucei rhodesiense* (Kuettel *et al.*, 2011), which has an identical amino-acid sequence, with the gene only differing by two silent mutations. It has only minor differences from those from *Anopheles gambiae* (*Ag*AK), human (*Hsa*AK), *Toxoplasma gondii* (*Tgo*AK) and *Mycobacterium tuberculosis* (*Mtu*AK) (Cook *et al.*, 2000; Mathews *et al.*, 1998; Muchmore *et al.*, 2006; Schumacher *et al.*, 2000; Zhang *et al.*, 2006, 2007). In brief, *Tb*AK consists of two domains: a large ATP-binding domain and a small lid domain. As in other eukaryotic AK structures, the larger domain consists of a typical α/β structure with a central nine-stranded β-sheet (β1, β5, β6, β9, β10, β11, β12, β13 and β14) enclosed by ten α-helices (α3–α12). The lid domain consists of a five-stranded β-sheet (β2, β3, β4, β7 and β8) at the interface to the large domain and two α-helices (α1 and α2) on top of it. The active site is located at the interface of the two domains.

The apo structure is in the open conformation, essentially identical to that reported for *Tb*AK soaked with the activator KRM (PDB entry 2xtb; Kuettel *et al.*, 2011) and that of apo *Tgo*AK (Schumacher *et al.*, 2000). Comparison with 2xtb showed that the overall structures are effectively the same, with the only differences being in the flexible loop region (residues 281–297). In addition, the activator complex is slightly more open, but this may not directly reflect the binding of the activator but may rather be an artefact of crystallization. The side chain of Leu138, close to the activator, has a different position in the two structures but makes no direct or indirect interactions with the activator itself.

3.3. Substrate binding

There are four independent protein molecules in the asymmetric unit, with well defined density for both the adenosine and for ADP (Fig. 2*a*). The four chains are closely similar: the r.m.s. between all C^{α} atoms is 0.37, 0.39 and 0.29 Å for chains *B*, *C* and *D* superposed on chain *A* using *SSM* (Krissinel & Henrick, 2004). Chain *A* is therefore used in the rest of the discussion and in the other panels of the figure.

*Tb*AK adopts a closed conformation similar to all other structures of complexes of AK with substrates or substrate analogues in both of the binding pockets of the active site. Superposition of the structures

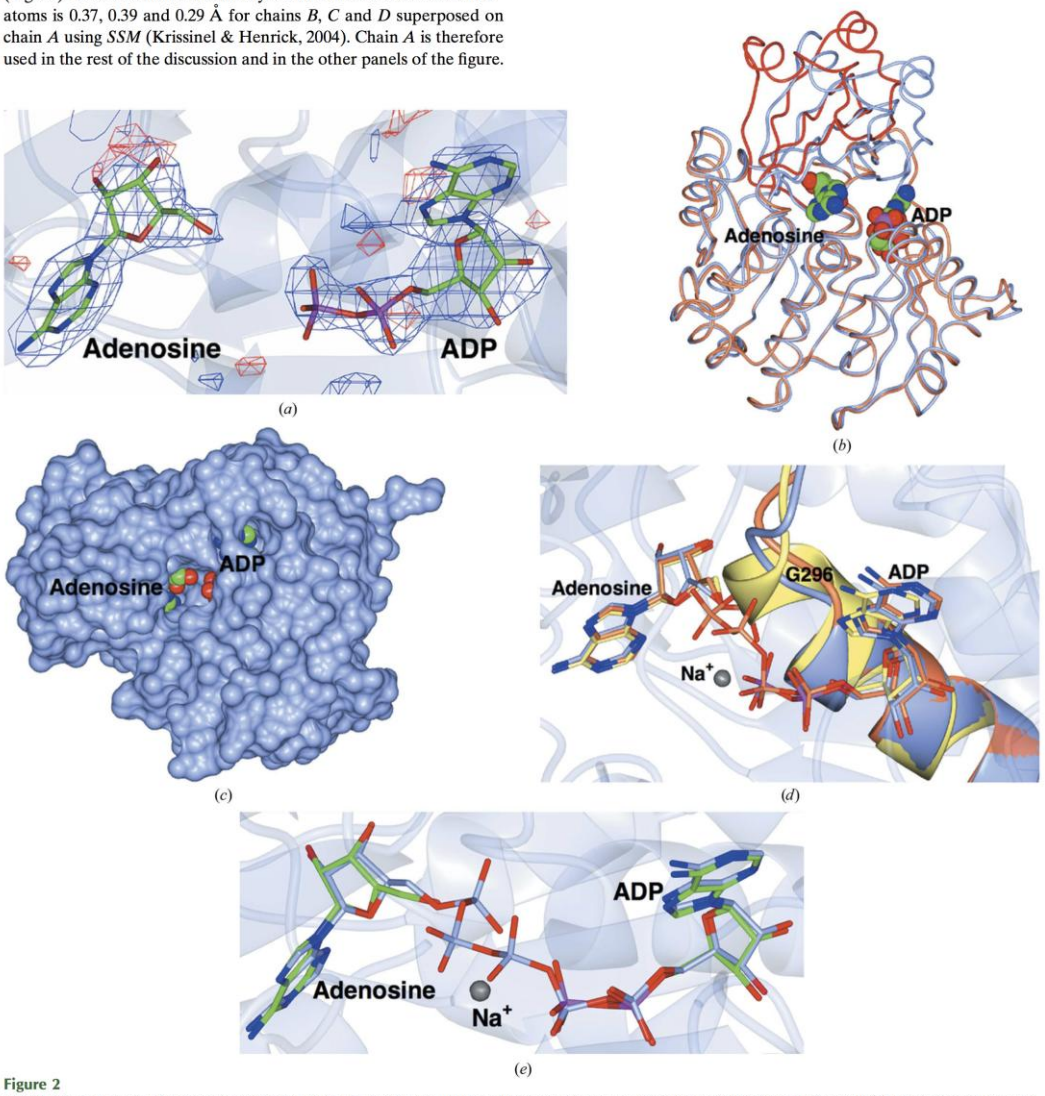


Figure 2
The structure of *Tb*AK. The views are all of the *A* chain, but there is essentially no difference between the four molecules in the asymmetric unit. (*a*) The unbiased maximum-likelihood difference map calculated with *REFMAC*5 output coefficients contoured at 3σ before the ligands were added to the model. The positive contour is in blue and the negative contour is in red. The protein chain is shown as a semi-transparent ribbon. (*b*) Worm representations of the fold of the open apo form (coral core subunit and red lid) superposed on the closed form (ice blue). The structures were superposed using *SSM* (Krissinel & Henrick, 2004) based on the core subunit. The adenosine and ADP ligands are shown as spheres. (*c*) Surface view of the *A* chain, in which the two adenines can be seen to be buried in deep pockets. (*d*) Superposition of the anion-hole region of the *Tb*AK–adenosine–ADP complex (ice blue; the protein was co-crystallized with AMPPNP, which is presumed to have hydrolysed) reported here with those of *Tb*AK–AP5A (PDB entry 3otx, coral; Kuettel *et al.*, 2011) and human AK in complex with two adenosines (PDB entry 1bx4, lemon; Mathews *et al.*, 1998). The structures were superposed using *SSM* based on the C^{α} atoms in chain *A* of both. The ligands are shown as cylinders, with the C and P atoms coloured according to chain for clarity. (*e*) Superposition of *Tb*AK–adenosine–ADP with *Tb*AK–AP5A (PDB entry 3otx). The ribbon shows the *Tb*AK–adenosine–ADP trace. The ligands are shown as cylinders coloured as in (*d*). This figure was made with *CCP4mg* (McNicholas *et al.*, 2011).

of *Tb*AK in its closed and open forms using the *ProSMART* program in *REFMAC* (Murshudov *et al.*, 2011) confirmed that the smaller lid domain closes over the active site and the large domain contracts slightly (Fig. 2*b*). This results from a rigid-body rotation of 28° with hinge motions of the amino acids connecting the two domains. In addition, *Tb*AK undergoes smaller structural rearrangements while binding to and orienting its substrates. The adenosine makes hydrogen bonds to residues Asp13, Asp17, Gly62, Asn67, Asp299 (adenosine) and Asn222, Thr264, Thr270, His266, Gly298 and Gln288 (ADP), and stacking interactions of its adenine ring with Phe169 which align the helix adjacent to it. The interaction with the substrates causes a slight bending of the α12 helix (residues 318–330) and a closing of the loop (residues 266–268) near the ATP-binding site towards the ADP. These smaller movements orient the substrates in the binding pockets and bury the two adenine moieties (Fig. 2*c*).

Important for catalysis in nucleotide kinases is the anion hole (Schumacher *et al.*, 2000), a conserved sequence of residues essential for appropriate binding of the phosphates of the substrate, the formation of which requires the P-loop (residues 290–300) to be pushed outwards. The superposition of our structure of *Tb*AK–adenosine–ADP on those of *Tb*AK–AP5A and *Hsa*AK in complex with two adenosines shows it to be in an intermediate state between the latter two conformations, reflecting the importance of this motif (Fig. 2*d*). In our complex, the γ-phosphate of AMPPNP (which mimics the phosphate donor ATP) is missing, so that its interactions cannot help to arrange the loop completely, which is therefore not bent outwards as far as it would be with all three phosphates present. The structure of *Tb*AK–AP5A (Kuettel *et al.*, 2011) shows a fully developed anion hole; the complex of *Hsa*AK with two adenosines (Mathews *et al.*, 1998) lacks the anion-hole motif completely and instead has an extended α11 helix, as there are no phosphates in the active site. In particular, residues 294–298 (centred around Gly296) adopt a helical conformation in the absence of phosphate ligands, but change to a loop producing the anion hole when phosphates are present.

The structure of *Tb*AK–adenosine–ADP in the closed conformation was very similar to that of the complex with the bi-substrate analogue, *Tb*AK–AP5A (Kuettel *et al.*, 2011), with an r.m.s. on C^{α} positions of 0.57 Å (Fig. 2*e*). The adenines in both structures occupy essentially identical positions, as does the ribose of ADP and AP5A. In the *Tb*AK–AP5A complex there are five phosphates between the two adenosine moieties, two more than must be present in the transition state, so the phosphates would need to be in different conformations in the Michaelis complex and the transition state. Structures of adenylate kinase from *Aquifex aeolicus* have been determined in complex with both AP5A (PDB entry 2rgx; Henzler-Wildman *et al.*, 2007) and with the phosphoryl-transfer transition-state mimic ADP–AlF_4–AMP in the active site (PDB entry 3sr0; Y.-J. Cho & D. Kern, unpublished work). The enzyme was in a very similar closed conformation in both, suggesting that the enzyme conformation and the adenine-binding sites in the AP5A analogues were indeed representative of the true ligand-binding state. We therefore carried out extensive crystallization screens with nucleotides and magnesium and aluminium fluorides with the aim of obtaining a better mimic of the transition state. Unfortunately, none of these led to crystals of diffraction quality.

4. Conclusions

Crystallization of *Tb*AK proved to be surprisingly challenging given the substantial number of homologous structures in the PDB. Only one crystal, of limited quality, of the apoenzyme was identified and this provided diffraction data to 2.6 Å resolution, which were sufficient to allow a description of the enzyme in its open conformation. It is impressive that modern integration software can extract useful data from such a crystal.

The structure of *Tb*AK–adenosine–ADP was in the closed conformation and was very similar to that reported for the complex with the transition-state mimic, *Tb*AK–AP5A (Kuettel *et al.*, 2011). The two structures were superposed with an r.m.s.d. on C^{α} positions of 0.57 Å using *SSM* and the ligands are shown for chain *A* of both structures in Fig. 2(*e*). It is evident that the adenines in both structures occupy essentially identical positions, as does the ribose of ADP and AP5A. What is most notable about the *Tb*AK–AP5A complex is that there are five phosphates between two adenosine moieties, two more than must be present in the transition state. This indicates that the phosphates would need to be in different conformations in the Michaelis complex and the transition state, perhaps in part reflecting the absence of chelating metal in the present complex. Structural studies of adenylate kinase from *A. aeolicus* have yielded complexes of the enzyme with both AP5A (PDB entry 2rgx; Henzler-Wildman *et al.*, 2007) and with the phosphoryl-transfer transition-state mimic ADP–AlF_4–AMP in the active site (PDB entry 3sr0; Y.-J. Cho & D. Kern, unpublished work). In this system, the enzyme was in a very similar closed conformation in both structures, suggesting that the enzyme conformation and the adenine-binding sites in the AP5A analogues were indeed representative of the true ligand-binding state. However, extensive crystallization screens with nucleotides and magnesium and aluminium fluorides failed to produce useful crystals. One problem may lie in the fact that the aluminium fluoride in *Tb*AK would be coordinated by a phosphate O atom on one side and a ribose O atom on the other, unlike the adenylate kinase system, in which the aluminium is coordinated by phosphate O atoms on both sides. This requires further work.

We thank the Diamond Light Source (beamlines I24 and I03) for the provision of X-ray data-collection facilities. This work was funded by the European Union (grant agreement No. 223238).

References

Arvai, A. (2012). *adxv: A Program to Display X-ray Diffraction Images*. http://www.scripps.edu/~arvai/adxv.html.
Barrett, M. P., Boykin, D. W., Brun, R. & Tidwell, R. R. (2007). *Br. J. Pharmacol.* **152**, 1155–1171.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cook, W. J., DeLucas, L. J. & Chattopadhyay, D. (2000). *Protein Sci.* **9**, 704–712.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Fogg, M. J. & Wilkinson, A. J. (2008). *Biochem. Soc. Trans.* **36**, 771–775.
Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G. & Kern, D. (2007). *Nature (London)*, **450**, 838–844.
Iovannisci, D. M. & Ullman, B. (1984). *Mol. Biochem. Parasitol.* **12**, 139–151.
James, D. M. & Born, G. V. (1980). *Parasitology*, **81**, 383–393.
Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.
Kouni, M. H. el (2003). *Pharmacol. Ther.* **99**, 283–309.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Kuettel, S., Greenwald, J., Kostrewa, D., Ahmed, S., Scapozza, L. & Perozzo, R. (2011). *PLoS Negl. Trop. Dis.* **5**, e1164.
Mäser, P., Vogel, D., Schmid, C., Räz, B. & Kaminsky, R. (2001). *J. Mol. Med.* **79**, 121–127.

Mathews, I. I., Erion, M. D. & Ealick, S. E. (1998). *Biochemistry*, **37**, 15607–15620.
McCoy, A. J. (2007). *Acta Cryst.* D**63**, 32–41.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Muchmore, S. W., Smith, R. A., Stewart, A. O., Cowart, M. D., Gomtsyan, A., Matulenko, M. A., Yu, H., Severin, J. M., Bhagwat, S. S., Lee, C. H., Kowaluk, E. A., Jarvis, M. F. & Jakob, C. L. (2006). *J. Med. Chem.* **49**, 6726–6731.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Park, J. & Gupta, R. S. (2008). *Cell. Mol. Life Sci.* **65**, 2875–2896.
Reddy, M. C., Palaninathan, S. K., Shetty, N. D., Owen, J. L., Watson, M. D. & Sacchettini, J. C. (2007). *J. Biol. Chem.* **282**, 27334–27342.
Schumacher, M. A., Scott, D. M., Mathews, I. I., Ealick, S. E., Roos, D. S., Ullman, B. & Brennan, R. G. (2000). *J. Mol. Biol.* **298**, 875–893.
Vodnala, M., Fijolek, A., Rofougaran, R., Mosimann, M., Mäser, P. & Hofer, A. (2008). *J. Biol. Chem.* **283**, 5380–5388.
Williamson, J. (1972). *Trans. R. Soc. Trop. Med. Hyg.* **66**, 354–355.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
World Health Organisation (2011). *Trypanosomiasis Fact Sheet*. http://www.who.int/trypanosomiasis_african/en/.
Zhang, Y., el Kouni, M. H. & Ealick, S. E. (2006). *Acta Cryst.* D**62**, 140–145.
Zhang, Y., el Kouni, M. H. & Ealick, S. E. (2007). *Acta Cryst.* D**63**, 126–134.