

A CALPAIN-LIKE MULTIGENE FAMILY IN
TRYPANOSOMA BRUCEI

A Thesis Submitted for the Degree of Doctor of Philosophy in the
University of Oxford
Hilary Term 2010

Sofia Amalia Olego-Fernandez
Lincoln College
Sir William Dunn School of Pathology

DEDICATIONS.

This thesis is especially dedicated to my dad, Ovidio Olego Corcoba, whose love has always accompanied me.

ABSTRACT.

Trypanosomatid parasites are unicellular eukaryotes characterised by the presence of a subpellicular array of microtubules, a single flagellum, and a kinetoplast (containing the condensed mitochondrial DNA). The majority of trypanosomatid species undergo complex life-cycles, alternating between a mammalian host and an insect vector. Progression through this life-cycle requires the differentiation of trypanosomatids into distinct, niche adapted developmental forms. Differentiation into each life-cycle stage involves important biochemical and morphological changes, including the remodelling of the subpellicular cytoskeleton that defines cell shape. In higher eukaryotes, proteins from the calpain superfamily are involved in developmentally- and environmentally-regulated remodelling of the cytoskeleton and the dynamic organisation of signal transduction cascades. Interestingly, trypanosomatids contain unusually large families of calpain-related proteins, but there is little knowledge about the functional roles of these molecules during the life-cycle of trypanosomatid parasites.

In this thesis, I present the results of the bioinformatic analysis of calpain-like proteins in three trypanosomatid parasites, *Trypanosoma brucei*, *Leishmania major* and *Leishmania infantum*. From this analysis, I selected several calpain-related proteins for RNAi functional analysis, on the bases of their domain composition and conservation across different species. The detailed analysis of the resulting RNAi phenotypes revealed the essential function of some calpain-like proteins for the correct morphogenesis of specific developmental forms of *T. brucei*, shedding some light on the mechanisms that regulate this parasite differentiation and cytoskeletal remodelling, and providing new putative therapeutic targets for African sleeping sickness.

DECLARATION.

The following figures included in this thesis have been modified from published data by previous and current members of the Gull laboratory:

Chapter 1, Figure 1.1: adapted from *Moreira-Leite, 2002*.

Chapter 1, Figure 1.2A: adapted from *Vaughan and Gull, 2003*.

Chapter 1, Figure 1.2.C: adapted from *Bastin et al., 2000*.

Chapter 1, Figure 1.6: adapted from *Ersfeld et al., 2005*.

No other portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

PUBLISHED WORK.

Olego-Fernandez, S., S. Vaughan, M. K. Shaw, K. Gull and M. L. Ginger, (2009). Cell morphogenesis of *Trypanosoma brucei* requires the paralogous, differentially expressed calpain-related proteins CAP5.5 and CAP5.5V. *Protist.* 160: 576-90.

ACKNOWLEDGMENTS.

First of all, I would like to thank my supervisors Professor Keith Gull and Doctor Michael L. Ginger for their invaluable help, advice and support throughout this thesis work.

I would also like to thank the people from the Gull lab for making it such a nice place to work and learn. Special thanks to Bill, for his generosity and help through these years; to Eva, for being such a nice bench neighbour; to Andrea, for being so lovely and supportive to Neil, Steve, Helen, etc.

Working in the Dunn School has been an honour and a privilege, not only for its historical importance and state of the art science, but also for the wonderful people that work there. Thanks to all those that made it possible.

I would also like to thank my friends Camilla Oxley and Tatjana Lalic, for reminding me that there was life outside the lab, being so great fun and better friends. My "Oxford Experience" would have never been so good without you, girls. I will miss you lots. Thanks also to Kimiya Minoukadeh, for her friendship and support and for making my first year so great. Especial thanks also to the Trapido family and Margaret Alice Stuart-Liberty, you have been a second family to me and I'll be indebt to you forever.

My warmest thanks to my sister Lulú (Luisa María Olego-Fernandez), for her support and generosity, I would have never finished writing without you; and to Sofia Fernandez-Rodriguez, my mum, to whom I own everything.

Thanks also to Belen Pintado, for her love and support.

Muchas gracias a mis tías Pilar y Fina, por su ánimo y cariño; a my hermanita María, a la que adoro; y a todos mis amigos en España, muchas gracias por vuestro apoyo.

TABLE OF CONTENTS.

CHAPTER 1: INTRODUCTION.....	1
1.1. Introduction to <i>Trypanosomatidae</i> and their developmental forms	2
1.2. The lifecycle of <i>T. brucei</i>	6
1.3. Cytoskeletal architecture in <i>T. brucei</i>.....	9
1.3.1. The subpellicular microtubule corset.....	10
1.3.2. The basal bodies and flagellar axoneme.....	13
1.3.3. The Flagellum Attachment Zone (FAZ).....	14
1.4. The cell-cycle of procyclic <i>T. brucei</i>.....	17
1.5. The cell-cycle of bloodstream <i>T. brucei</i>	19
1.6. RNA interference.....	20
1.7. Calpain family of cysteine proteases.....	22
1.7.1. Introduction to calpains.....	22
1.7.2. Calpain properties.....	23
1.7.3. Calpains activity and regulation.	27
1.7.4. Tissue localisation and physiological functions.	29
1.7.5. Atypical calpains.....	31
1.8. Expanded calpain-like family in trypanosomatids.....	34
1.8.1. CAP5.5: the first identified calpain-like protein in trypanosomatids.	34
1.8.2. Bioinformatic identification and analysis of a large calpain-like family in kinetoplastids.....	40
1.8.3. Localisation and function.	47
1.9. Aims and strategy	48
CHAPTER 2: MATERIALS AND METHODS.....	49
2.1. Bioinformatics: database mining and sequence analysis.....	50
2.2. Molecular biology.....	52
2.2.1. RNAi constructs.....	52
2.2.2. Polymerase Chain Reaction (PCR).....	57
2.2.3. Anchored PCR.....	57
2.2.4. Restriction digests.	57
2.2.5. Purification of DNA fragments.....	58
2.2.6. DNA ligations.....	58
2.2.7. Bacterial strains and molecular cloning.....	59
2.2.8. Plasmid DNA isolation.	59
2.2.9. Sequencing of RNAi plasmids.	59
2.2.10. Expression analysis by real-time-PCR (RT-PCR)	60

2.3. Trypanosome cell lines and cell culture	61
2.3.1. Culture of trypanosomes	61
2.3.2. Transformation of trypanosomes	61
2.3.3. Cloning by limiting dilution of procyclic trypanosomes	62
2.3.4. RNAi induction	63
2.4. Protein biochemistry	63
2.4.1. Preparation of cell extracts	63
2.4.2. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)	63
2.4.3 Western blotting and immunodetection	64
2.5. Light microscopy	65
2.5.1. Immunofluorescence	65
2.5.2. Epifluorescence microscopy	67
2.5.3. Videomicroscopy	67
2.6. Electron microscopy (EM)	67
2.6.1. Preparation of samples for thin-section Transmission Electron Microscopy (TEM)	67
2.6.2. Scanning Electron Microscopy (SEM)	68
CHAPTER 3: BIOINFORMATIC ANALYSIS OF KINETOPLASTID CALPAIN-LIKE SEQUENCES	69
3.1. Introduction	70
3.2. Analysis: New sequences identification and characterisation	72
3.2.1. Identification of trypanosomatid calpain-like domains	73
3.2.2. Identification of no-calpain-related domains	85
3.2.3 Analysis of calpain-like sequences containing 67-68 amino acid tandem-repeats	95
3.2.4. Identification of a large chromosome duplication event in <i>T. brucei</i> , containing a CAP5.5 paralogous gene, Tb907.8.8330.	114
3.3. Reclassification of kinetoplastid calpain-like sequences	125
3.4. Syntenic genome organisation	132
3.5. Kinetoplastid calpain-like proteins database	133
3.6. Selection of <i>T. brucei</i> calpain-like sequences for functional analysis by RNAi	138
3.7. Functional survey of calpain-like sequences by RNAi	141
CHAPTER 4: RNAi ANALYSIS OF TWO PARALOGOUS, DIFFERENTIALLY EXPRESSED CALPAIN-LIKE PROTEINS: CAP5.5 AND CAP5.5V	145
4.1. Introduction	146
4.2. CAP5.5 and CAP5.5V are stage-specific regulated: expression analysis by RT-PCR	148
4.3. Selection of sequence fragments for RNAi	150

4.4. Applying the RNAi phenotype test: CAP5.5 is essential in procyclic <i>T. brucei</i>.....	150
4.4.1. CAP5.5 RNAi silencing produces a slow growth phenotype and aberrant cytokinesis.....	156
4.4.2. CAP5.5 is essential for correct morphogenetic patterning during the cell division-cycle of procyclic <i>T. brucei</i>	160
4.4.3. CAP5.5 knockdown does not affect tubulin localisation	170
4.4.4. WCB localisation is unaffected by CAP5.5 silencing.....	175
4.4.5. CAP17 and CAP5.5 may have competing, antagonistic functions.....	178
4.4.6. SEM analysis of CAP5.5 mutants gives further evidence of aberrant cell morphology and cytokinesis induced by CAP5.5 silencing.....	180
4.4.7 Ultrastructural analysis of CAP5.5 RNAi phenotype reveals microtubule disorganisation induced by CAP5.5 silencing	188
4.5. RNAi against CAP5.5 targeting a different region of its sequence produced a qualitatively similar phenotype	193
4.6 CAP5.5V is essential in bloodstream <i>T. brucei</i>.....	201
4.6.1. CAP5.5V RNAi in bloodstream <i>T. brucei</i>	201
4.6.2. CAP5.5V silencing does not affect the basic components of microtubules.....	212
4.6.3. CAP5.5V is essential for the correct morphogenetic patterning during the cell division-cycle of bloodstream <i>T. brucei</i>	212
4.6.4. CAP5.5/CAP5.5V RNAi silencing phenotype in bloodstream <i>T. brucei</i>	224
CHAPTER 5: RNAi ANALYSIS OF GROUP 4 CALPAIN-RELATED PROTEINS	230
5.1. Introduction	231
5.2. Applying the RNAi test	234
5.3. RNAi functional analysis of group 4 calpain-like sequences in <i>T. brucei</i> PCF: Silencing Tb11.57.0008 and Tb11.47.0035	238
5.3.1. RNAi knockdown of group 4 calpain-like proteins produced a reduction in growth rate and cell-body size	239
5.3.2. RNAi silencing of group 4 calpain-like proteins resulted in aberrant morphology and organelle mispositioning.....	244
5.3.3. Group 4 calpain-like proteins are essential for the correct positioning of the nucleus, kinetoplast and flagellar pocket in PCF trypanosomes	253
5.3.4. Group 4 calpain-like proteins have an essential role in cell morphogenesis.....	261
5.3.5 RNAi knockdown of group 4 calpain-like sequences alters the structure of the FAZ	267
5.3.6 SEM analysis of group 4 calpain-like proteins RNAi cell lines showed aberrant morphogenesis of induced cells that resemble epimastigote-like forms	269
5.3.7. RNAi knockdown of group 4 calpain-like proteins causes aberrant assembly of the FAZ in PCF trypanosomes.....	277

5.4. RNAi functional analysis of group 4 calpain-like sequences in <i>T. brucei</i> BSF targeting the Lcr1-like repeats: Tb11.47.0035 and Tb11.v4.0001	279
5.5. RNAi functional analysis of group 4 calpain-like sequences in <i>T. brucei</i> BSF targeting the GM6-like repeats: Tb11.47.0036 and Tb11.57.0008.....	280
CHAPTER 6: DISCUSSION.....	282
6.1. Calpain-like proteins are essential for the correct morphogenesis of <i>T. brucei</i> procyclic and bloodstream forms	288
6.2. Calpain-like proteins are essential for the correct structure and length of the FAZ	298
6.3. Bioinformatic analysis of trypanosomatid calpain-like sequences: a tool for functional prediction.....	304
6.5 Comparison between different RNAi approaches	309
6.6. Concluding Remarks and Perspective	311
REFERENCES.....	312

TABLE OF FIGURES.

Figure 1.1: Main developmental forms found in the life-cycles of trypanosomatid human parasites.....	5
Figure 1.2: Overview of <i>T. brucei</i> lifecycle	7
Figure 1.3: . Scheme of a <i>T. brucei</i> procyclic cell	11
Figure 1.4: Typical and atypical calpains	24
Figure 1.5: Alignment of the catalytic domains of human calpains and CAP5.5..	36
Figure 1.6: Classification of trypanosomatid calpain-like proteins	46
Figure 2.1: The ClustalX colour code.....	51
Figure 2.2: The p2T7-177 RNAi vector	53
Figure 3.1: Multiple alignment of domain I ^K sequences.	74
Figure 3.2: Multiple alignment of domain II sequences	76
Figure 3.3: Multiple alignment of domain III sequences	81
Figure 3.4: Multiple alignment of the META domain sequences.....	87
Figure 3.5: Synteny map of newly identified calpain-like sequencess.....	88
Figure 3.6: Multiple alignment of calpain-like sequences and kinesin ARK proteins	91
Figure 3.7: Trypanosomatid ADCYs containing a calpain-like domain I ^K	94
Figure 3.8: Multiple alignment of FRA repeat units	100
Figure 3.9: Multiple alignment of Tb11.57.0008, Tb11.47.0036 and GM6 repeats	101
Figure 3.10: Detailed diagram of the <i>T. brucei</i> genome locus of tandem repeats-containing calpain-like sequences	103
Figure 3.11: Tb11.47.0035 and Tb11.V4.0001 repeats alignment	105
Figure 3.12: Alignment of the 67 amino acid repeats of LmjF27.500.....	107
Figure 3.13: Multiple alignment GM6 and FRA repeats	110
Figure 3.14: Multiple alignment of the most conserved FRA and Lcr1 repeats..	111
Figure 3.15: Multiple alignment of the most conserved GM6 and Lcr1 repeats..	112
Figure 3.16: Sequence logos of the GM6, FRA, Lcr1 and Lcr1-like repeats.....	113
Figure 3.17: Amino acid alignment of CAP5.5 and CAP5.5V CDS	115
Figure 3.18: Schematic representation of CAP5.5 and CAP5.5V genome locus.	119
Figure 3.19: Nucleotide alignment of CAP5.5 and CAP5.5V CDSs and UTRs	120

Figure 3.20: Re-classification of trypanosomatid calpain-like proteins	126
Figure 3.21: Synteny map of <i>T. brucei</i> chromosomes 4 and 8.....	134
Figure 3.22: Synteny map of newly identified <i>T. brucei</i> calpain-like sequence with its orthologous regions in <i>L. major</i> and <i>L. infantum</i>	135
Figure 3.23: Survey of <i>T. brucei</i> calpain-like sequences by RNAi	142
Figure 4.1: Detergent extracted cytoskeletons of <i>T. brucei</i> procyclic probed with CAP5.5 mAb.....	147
Figure 4.2: q-RT-PCR analysis of CAP5.5 and CAP5.5V	149
Figure 4.3: Selection of Sequence Fragments for RNAi.	151
Figure 4.4: 5.5-C and 5.5V-C RNAi cell lines cumulative growth curve and western blot	153
Figure 4.5: 5.5V-C-U RNAi cell line induction. Cumulative growth Curve and DAPI counts of the number of nuclei and kinetoplasts	154
Figure 4.6: RNAi against CAP5.5V does not reduce expression of CAP5.5.....	155
Figure 4.7: 5.5-C-U RNAi cell line induction cumulative growth curve.....	157
Figure 4.8: RNAi against CAP5.5 silenced CAP5.5 expression.....	158
Figure 4.9: RNAi against CAP5.5 results in aberrant cytokinesis.....	159
Figure 4.10: IF analysis of PCF 5.5-C-U RNAi cell line not-induced cells.....	161
Figure 4.11: IF analysis of PCF 5.5-C-U RNAi cell line. 8h post-induction	162
Figure 4.12: IF analysis of PCF 5.5-C-U RNAi cell line. 0 and 8h post-induction	163
Figure 4.13: IF analysis of PCF 5.5-C-U RNAi cell line. 24h post-induction (1) ...	164
Figure 4.14: IF analysis of PCF 5.5-C-U RNAi Cell line. 24h (2).....	165
Figure 4.15: IF analysis of PCF 5.5-C-U RNAi Cell line. 24h post-Induction (3) .	166
Figure 4.16: IF analysis of PCF 5.5-C-U RNAi cell line. 48h post-induction (1)	168
Figure 4.17: IF analysis of PCF 5.5-C-U RNAi cell line. 48h post-induction (2) ...	169
Figure 4.18: IF analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction (1) ...	171
Figure 4.19: IF analysis of 5.5-C-U RNAi cell line. 72h post-induction (2)	172
Figure 4.20: IF analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction (3) ...	173
Figure 4.21: IF analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction(4)	174
Figure 4.22: 5.5-C-U RNAi cell line. 0 and 24h post-induction. TAT-1 antibody..	176
Figure 4.23: 5.5-C-U RNAi cell line 0 and 24h post-induction, CAP5.5 and WCB mAbs	177
Figure 4.24: CAP5.5 RNAi cell line (5.5-C-U). CAP5.5 and CAP17 Antibodies	179

Figure 4.25: 29-13 and not-induced 5.5-C-U RNAi cell lines. SEM	181
Figure 4.26: CAP5.5 RNAi cell line (5.5-C-U). 18h post-induction. SEM	182
Figure 4.27: CAP5.5 RNAi cell Line (5.5-C-U). 18h post-induction. SEM (2).....	183
Figure 4.28: CAP5.5 RNAi cell line (5.5-C-U). 24h post-induction. SEM.(3)	184
Figure 4.29: CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction. SEM	185
Figure 4.30: CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction. SEM.(2)	186
Figure 4.31: CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction. SEM.(3)	187
Figure 4.32: CAP5.5 RNAi cell line (5.5-C-U). 0h controls. TEM	188
Figure 4.33: CAP5.5 RNAi cell line (5.5-C-U). 18h post-induction. TEM.....	190
Figure 4.34: CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction.TEM	191
Figure 4.35: CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction. TEM.(2)	192
Figure 4.36: CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V) growth curve.....	194
Figure 4.37: CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V) counts of the number of nuclei and kinetoplasts.....	195
Figure 4.38: CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V). Detergent extracted cytoskeletons. 0, 24 and 48h post-induction.....	197
Figure 4.39: CAP5.5/CAP5.5V RNAi cell line.SEM. 0, 12, 24 and 48h post- induction	199
Figure 4.40: CAP5.5/CAP5.5V RNAi cell line. TEM. 0 and 48h post-induction	200
Figure 4.41: BSF CAP5.5V RNAi cell line (5.5V-C-U) cumulative growth curve .	202
Figure 4.42: BSF CAP5.5V RNAi cell line (5.5V-C-U) counts of number of nuclei and kinetoplast	204
Figure 4.43: BSF CAP5.5V RNAi cell line. Not-induced.cells	205
Figure 4.44: BSF CAP5.5V RNAi cell line (5.5V-C-U). 6h post-induction (1).....	206
Figure 4.45: BSF CAP5.5V RNAi cell line (5.5V-C-U). 6h post-induction (2).....	207
Figure 4.46: BSF CAP5.5V RNAi cell line (5.5V-C-U). 12h post-induction (1).....	209
Figure 4.47: BSF CAP5.5V RNAi cell line (5.5V-C-U). 12h post-induction (2).....	210
Figure 4.48: BSF CAP5.5V RNAi cell line (5.5-C-U). 12h post induction (3)	211
Figure 4.49: BSF CAP5.5V RNAi cell line (5.5V-C-U). 18h post-induction. TAT-1 mAb.....	212
Figure 4.50: SEM of BSF CAP5.5V RNAi cell line (5.5V-C-U). 0h.....	215
Figure 4.51: SEM BSF CAP5.5V RNAi cell line (5.5V-C-U). 12h post-induction...	216
Figure 4.52: SEM BSF CAP5.5V RNAi cell line (5.5V-C-U). 18h post-induction...	217

Figure 4.53: SEM BSF CAP5.5V RNAi cell line (5.5V-C-U). 24h post-induction...	218
Figure 4.54: TEM of BSF CAP5.5V RNAi cell line (5.5V-C-U)	220
Figure 4.55: TEM of BSF CAP5.5V RNAi cell line (5.5V-C-U). 12-24h post-induction	221
Figure 4.56: TEM BSF CAP5.5V RNAi cell line (5.5V-C-U). 24h post-induction...	222
Figure 4.57: TEM of BSF CAP5.5V RNAi cell line (5.5V-C-U). 24h (2).....	223
Figure 4.58: BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V). Growth curve	225
Figure 4.59: BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V).counts of number of nuclei and kinetoplasts	226
Figure 4.60: SEM of BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V)	227
Figure 4.61:TEM of BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V)	229
Figure 5.1: Schematic representation of <i>T. brucei</i> group 4 calpain-like sequences	236
Figure 5.2: BSF CAP5.5/CAP5.5V RNAi cell line. Cumulative growth curves	240
Figure 5.3: DS-N2-RNAi cell line 18h post-induction. Abnormal 1K1N cells	243
Figure 5.4: Inductions of the DS-N2 and HP-N RNAi cell lines	243
Figure 5.5: Counts of the number of nuclei and kinetoplasts	245
Figure 5.6: HP-N RNAi cell line. 0h images 40x.....	246
Figure 5.7: DS-N2 RNAi cell line. 0h images 40x.....	247
Figure 5.8: HP-N RNAi cell line. 0h images 100x.....	248
Figure 5.9: DS-N2-RNAi first induction	250
Figure 5.10: DS-N2 RNAi second induction.....	251
Figure 5.11: HP-N RNAi induction	252
Figure 5.12: Morphometry of DS-N2 RNAi cells (1)	255
Figure 5.13: Morphometry of DS-N2 RNAi cells.(2)	256
Figure 5.14: Microscopic analysis of DS-N2-RNAi cell line, 24h induced	260
Figure 5.15: Microscopic analysis of DS-N2-RNAi cell line, 48h induced	264
Figure 5.16: Microscopic analysis of DS-N2 RNAi cell line, 72h induced.....	265
Figure 5.17: Microscopic analysis of DS-N2-RNAi cell line, 96h induced	266
Figure 5.18: IF analysis DS-N2 RNAi cell line labelled with L3B2 mAb	268
Figure 5.19: IF analysis of DS-N2 RNAi cell line labelled with CD10 mAb.....	270
Figure 5.20: SEM of DS-N2 RNAi cell line. 0 and 15h induced, low magnitude...	271

Figure 5.21: SEM of DS-N2 RNAi cell line. 0 and 15h induced, high magnitude	272
Figure 5.22: SEM of DS-N2 RNAi cell line. 24h post-induction	274
Figure 5.23: SEM of DS-N2-RNAi cell line. 48h post-induction	275
Figure 5.24: SEM of DS-N2-RNAi cell line. 48h post-induction (2)	276
Figure 5.25: TEM of DS-N2 RNAi cell line	278
Figure 5.26: BSF 08-36-Rep RNAi cell line. Growth curve	278
Figure 6.1: Typical and atypical calpains	284
Figure 6.2. Groups of calpain-like proteins lacking or possessing a calpain "catalytic" domain II	284

LIST OF TABLES

Table 2.1. List of primers used in this thesis	55
Table 2.2. Restriction enzymes used in this thesis	58
Table 2.3. Primary and secondary antibodies used in this thesis	66
Table 3.1. Newly identified calpain-like sequences	71
Table 3.2. Previous classification of trypanosomatids calpain-like sequences	127
Table 3.3. Trypanosomatid calpain-like proteins table/database	136
Table 6.1. Kinetoplastid calpain-like sequences containing a calpain "catalytic" domain II	288
Table 6.2. <i>T. brucei</i> calpain-like sequences and <i>L. major</i> orthologs	305

APPENDIX (DVD).

Appendix to Chapter 3: Appendix 3.

Appendix to Chapter 4: Appendix 4.

Appendix to Chapter 5: Appendix 5.

Electronic copy of this thesis: Sofia Olego-Fernandez DPhil Thesis.

Published work: Olego-Fernandez, et al., 2009. Cell morphogenesis of *Trypanosoma brucei* requires the paralogous, differentially expressed calpain-related proteins CAP5.5 and CAP5.5V. *Protist*. 160: 576-90. (Olego-Fernandez et al., 2009).

ABBREVIATIONS.

Ab: Antibody.

ADCY: Adenylate cyclase.

ADP: Adenosine Diphosphate.

ATP: Adenosine Triphosphate.

BB: Basal Body.

BLAST: Basic Local Alignment Search Tool.

bp: base pairs.

BSA: Bovine Serum Albumin.

BSF: Bloodstream form.

°C: Centigrade degree.

CAP: Cytoskeletal Associated Protein.

cm: centimetre.

cm²: square-centimetres.

CP: Central Pair.

DAPI: 4,6-diamidino-2-phenylindole.

ddH₂O: Double-distilled water.

DIC: Differential Interference Contrast.

DMSO: Dimethyl-sulphoxide.

DNA: Deoxyribonucleic acid.

dNTPs: deoxyribonucleoside triphosphates.

DRM: Detergent Resistant Membranes.

dsRNAs: Double stranded ribonucleic acid.

EB: Elution Buffer.

ECL: Enhanced chemiluminescence reagent

EDTA: Ethylene Diamine Tetraacetate.

EGTA: Ethylene Glycol Tetraacetate

EM: Electron Microscopy.

ER: Endoplasmic Reticulum.

F: Flagellum.

FAZ: Flagellum Attachment Zone.

FC: Flagella Connector.

FLA: Flagellar adhesion protein.

FP: Flagellar Pocket.

Fr: Flagellar.

g; Grams

g: Times gravity.

GA: Golgi Apparatus.

GDP: Guanosine Diphosphate.

GFP: Green Fluorescent Protein.

GPI: Glycosylphosphatidylinositol

GTP: Guanosine Triphosphate.

h: hour.

HIFCS: Heat Inactivated Foetal Calf Serum.

HP: Hairpin RNAi construct..

HP-N: Hairpin RNAi construct that targets the N-terminus of Tb11.47.0035 and Tb11.57.0008.

HSP: High-scoring Segment Pairs.

IF: Immunofluorescence.

IFT: Intraflagellar Transport.

K: Kinetoplast.

kb: kilo-base.

kDa: Kilo-Dalton

kDNA: Kinetoplast (mitochondrial) DNA.

l: Litre.

λ DNA: DNA from the λ phage (48502 bp) of *E. coli* λ cl857 S7.

LB: Luria Broth.

M: Molar (moles per litre)

mAb: Monoclonal Antibody.

MAFFT: Multiple Alignment based of the Fast Fourier Transform.

MAP: Microtubule Associated Protein.

MARP-1: Microtubule Associated Repetitive Protein-1

μg: Micrograms (10^{-6} grams).

μl: Microlitres.

min: Minutes.

MIT: Microtubule Interacting and Transport domain.

ml: millilitre (10^{-3} litres).

mm: millimetre.

mM: Millimolar (millimoles per litre).

mol: Mole.

ms: millisecond

MTOC: Microtubule Organising Centre.

MtQ: Microtubule Quartet.

mya: million years ago

N: Nucleus.

NF: New Flagellum.

ng: nanogram.

NMT: N-Methyl Transferase

NP-40: Nonidet P-40.

nmoles: Nanomoles (10^{-9} moles)

nM: Nanomolar (nmoles per litre)

OF: Old Flagellum.

ORF: Open Reading Frame.

PBS: Phosphate-Buffered Saline.

PCF: Procyclic form.

PCR: Polymerase Chain Reaction.

PEME: Buffer containing PIPES, EDTA, $MgCl_2$ and EGTA.

PEST: Acidic sequence rich in proline, glutamic and aspartic acids and threonine.

PFR: Paraflagellar Rod.

PIPES: Piperazine-N,N-bis (2-ethane sulphonic acid).

PPEF: Protein Phosphatase with EF-hands.

PTGS: Post-Transcriptional Gene Silencing.

RAC: Receptor-type Adenylate Cyclase.

q-RT-PCR: Quantitative Real Time PCR.

RdRP: RNA-dependant RNA polymerase.

RISC: RNA-Inducible Silencing Complex.

RNA: Ribonucleic acid.

RNAi: RNA interference.

RNAP: RNA Polymerase.

rpm: Revolutions per minute.

s: Second.

SDS: Sodium-dodecyl-sulphate.

SDS-PAGE: SDS polyacrylamide gel electrophoresis.

SEM: Scanning Electron Microscopy.

siRNA: Small interfering RNA.

SL: Splice leader.

spp: Species.

TAC: Tripartite Attachment Complex.

TBST: Tris-Buffered-Saline-Tween-20.

TEM: Transmission Electron Microscopy.

TEMED: Tetramethylethylenediamine.

Tet: Tetracycline.

Tet(R): Tet Repressor.

U: Units.

U of Phusion DNA Polymerase: Amount of enzyme that will incorporate 10 nmoles of dNTPs into acid-insoluble form at 74°C in 30 minutes under the manufacturers assay conditions.

U of Restriction Enzymes: Amount of enzyme required to digest 1 µg of λ DNA in 1 hour at 37°C in a total reaction volume of 50 µl

UTR: Untranslated Region.

V: Volt (electric potential difference).

v: Volume.

w: Weight.

WCB: Whole Cell-Body protein.

WHO: World Health Organisation.

ZPFM: Zimmerman Post-Fusion medium

CHAPTER 1

INTRODUCTION

1. CHAPTER 1: INTRODUCTION.

The work presented in this thesis describes my attempts to determine the function of different members from an unusually large and diverse family of calpain-related proteins present in the human and cattle pathogen *Trypanosoma brucei*. My analysis includes the characterisation of three calpain-like proteins that are required for cytoskeletal re-modelling and/or cell morphogenesis. To provide the context for these characterisations my introduction provides an overview of:

- Trypanosome lifecycle and developmental forms
- The trypanosome cytoskeleton.
- RNA interference (RNAi) as a tractable approach to study gene function in trypanosomes.
- Calpain function.

1.1. Introduction to *Trypanosomatidae* and their developmental forms.

The trypanosomatids are parasitic representatives from a diverse and ubiquitous collection of protozoan flagellates known as the kinetoplastids. Phylogenetic studies, based on the evolution of the cytoskeleton, suggest that the kinetoplastids belong to the eukaryotic infrakingdom Excavata, phylum Euglenozoa (*Cavalier-Smith, 2002; Hampl et al., 2009*). Euglenozoa has been recently reclassified as an independent infrakingdom, close to the root of the eukaryotic tree, as phylogenetic studies also suggest that they are rather divergent from the other eukaryotic supergroups (*Cavalier-Smith, 2010*). This finding may be consistent with the many unusual biological characteristics that have been identified in the diverse collection of protists that constitute the kinetoplastids. The defining characteristic of the kinetoplastids is the

kinetoplast, which is the name given to the unusually structured mitochondrial genome of these organisms (Honigberg, 1963). All members of the *Trypanosomatidae* family are obligatory parasites that exhibit an extraordinarily broad host-range that extends from humans to plants, demonstrating a great adaptability and capacity for life as parasites. Of its nine recognized genera, two include species that are pathogenic for humans and animals —*Leishmania* and *Trypanosoma*— which cause parasitic diseases of tremendous medical and economic importance, including sleeping sickness, Chagas' disease and several types of leishmaniases. These diseases kill and debilitate hundreds of thousands of people worldwide each year (<http://www.who.int>). *Trypanosoma brucei*, the causative agent of human sleeping sickness (Barrett, 2006), is responsible for over 60.000 deaths annually, threatening millions of people in Sub-Saharan Africa (Mathers et al., 2003; Mathers, 2005; Kennedy, 2008). Moreover, *T. brucei* subspecies that are able to transmit between vertebrates, have spread across Asia, where they cause diseases in different mammals, resulting in a huge economic problem (Lai et al., 2008).

All trypanosomatids share basic features of their cell architecture, including a cortex of microtubules bellow the plasma membrane —the subpellicular microtubule corset that defines cell shape— and, prior to entry into the cell division-cycle, a single nucleus, single flagellum, and single kinetoplast. The kinetoplast is always located at the proximal end of the flagellum, where it is physically attached to the basal body through a mitochondrial membrane-spanning complex, the tripartite attachment complex (TAC) (Ogbadoyi et al., 2003). Most trypanosomatid species are characterised by complex lifecycles, generally involving alternation between a mammalian host and an insect vector. In both vector and host, trypanosomatids differentiate into numerous, niche adapted morphological forms, each of which can exhibit unique variations in cell architecture or form. For instance, developmentally-regulated changes in form include

the position of the nucleus and kinetoplast relative to one-another and to the posterior and anterior cell-ends, as well as the extent of the flagellar apparatus

Based on different morphological appearances, Horace and Wallace introduced a unified terminology to define the main developmental forms of trypanosomatids. (*Hoare and Wallace, 1966*) (Fig.1.1). The most common of these forms are trypomastigotes, which have the kinetoplast posterior to the nucleus; epimastigotes, with the kinetoplast in a position anterior-lateral to the nucleus; promastigotes and amastigotes, both of which have the kinetoplast anterior to the nucleus. In trypomastigotes the flagellum emerges laterally from the posterior-end of the cell, where the kinetoplast is located, following a path towards the anterior-end attached to the cell-body. However, in epimastigotes and promastigotes the flagellum emerges from a position closer to the anterior cell-end and is only partially attached to the cell body. Additionally, amastigotes lack an emergent flagellum (Fig.1.1).

The morphological changes during trypanosomatids lifecycles are accompanied by biochemical alterations and alternation between proliferative and non-proliferative phases, which are vital for the parasites adaptation to different environments, increasing transmission and infection efficiency (*Vickerman, 1985*). Transcription of the vast majority of trypanosome's protein-coding genes occurs polycistronically (*Clayton, 2002*), in the absence of gene-specific promoters. Regulation of gene expression is therefore mainly achieved post-transcriptionally, through trans-splicing of nascent transcripts and specific rates of mRNA degradation. This regulation generates steady state transcript levels that can differ in abundance across several orders of magnitude. Some mRNA levels are developmentally regulated, each developmental form having a slightly different transcriptome. Variations between in vivo and in vitro cultured trypanosomes have been also shown to occur (*Koumandou et al., 2008*), adding an extra level of complexity in the search for therapeutic drug targets.

Figure 1.1.

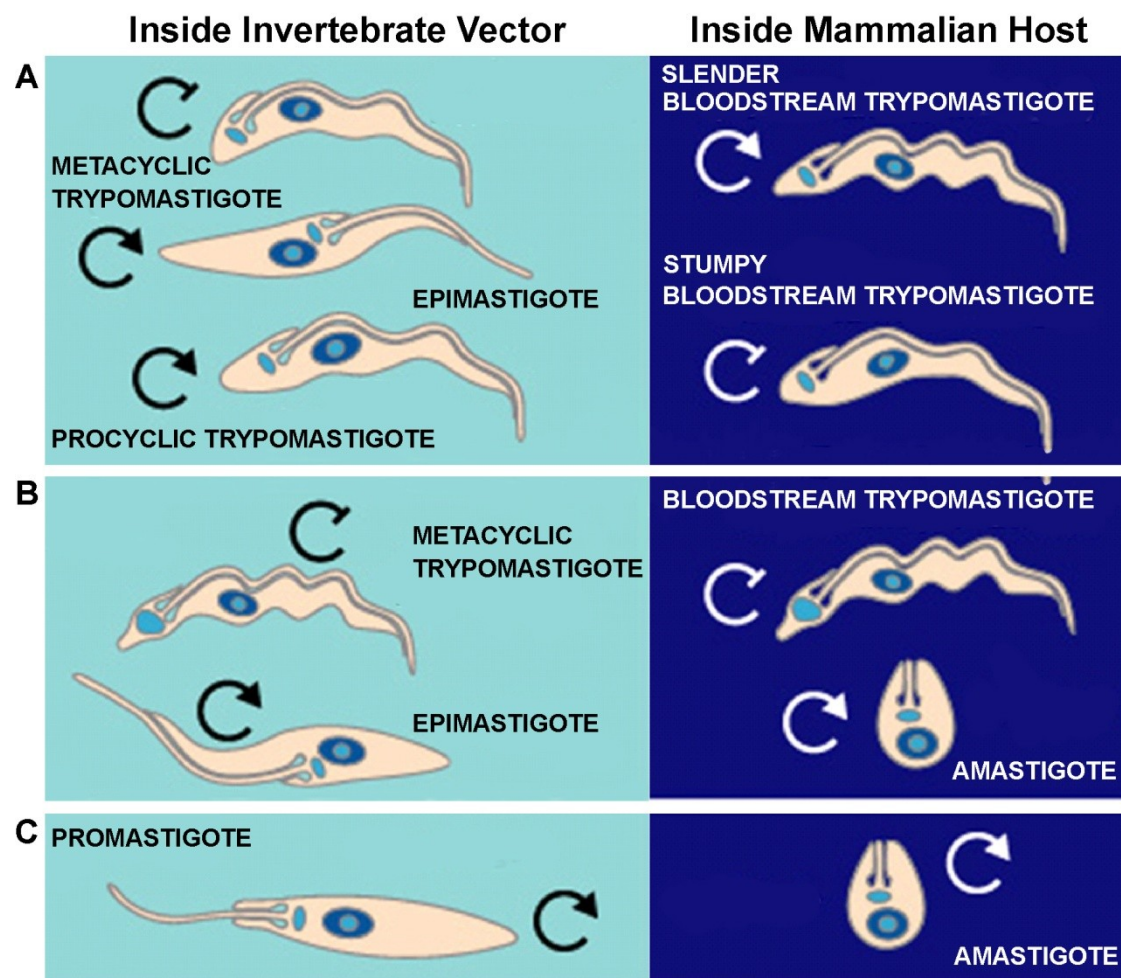


Figure 1.1. Main developmental forms found in the lifecycles of trypanosomatid human parasites, adapted from *Moreira-Leite, 2002*. A: *T. brucei*; B: *T. cruzi*; C: *Leishmania*. Circular arrows indicate proliferative forms. Interrupted circles indicate non-proliferative forms.

1.2. The lifecycle of *T. brucei*.

The complex lifecycle of *T. brucei* (Fig.1.2), its transmission from the mammalian host and the insect vector, and vice-versa, is mediated by the bite of the Tsetse fly. When an infected Tsetse fly takes a blood meal, the non-proliferative metacyclic trypomastigotes are transmitted to the mammalian host, where they rapidly differentiate into the proliferative, long slender bloodstream trypomastigotes. Dividing slender forms are adapted to the bloodstream environment, escaping the humoral immune response through antigenic variation of their surface coat (Cross, 1977). To this end, bloodstream trypomastigotes exclusively express one antigenic type of GPI anchored variant surface glycoprotein (VSG) at a time, called the homotype, stochastically switching to an antigenically different VSG, at a low frequency, from a still not fully known repertoire of more than 1000 VSG genes (Hutchinson *et al.*, 2007). Slender bloodstream forms obtain their energy exclusively from glycolysis, taking glucose from the host fluids and metabolising it in spherical glycosomes, which house the glycolytic chain enzymes (Vickerman, 1985). As parasite numbers increase, long slender trypomastigotes differentiate into stumpy, non-proliferative trypomastigotes, resulting in a heterogeneous trypanosome population (Vassella *et al.*, 1997).

Accumulation of division-arrested stumpy trypomastigotes, which can only continue the lifecycle in the fly vector, achieves two goals. First, it limits the increase in parasite numbers, prolonging host survival and, thereby, the probability of transmission. Second, the arrest of stumpy forms in G1 phase ensures the coordination of their morphological changes upon transmission, with their coordinated re-entrance into the cell-cycle, which is critically important for the correct organelle positioning required by their next developmental stage, the procyclic form, for the successful completion of their cell-cycle (Matthews, 2005). Additionally, there is a partial reactivation of the mitochondrion, repressed in slender bloodstream forms, which is most probably able to

Figure 1.2.

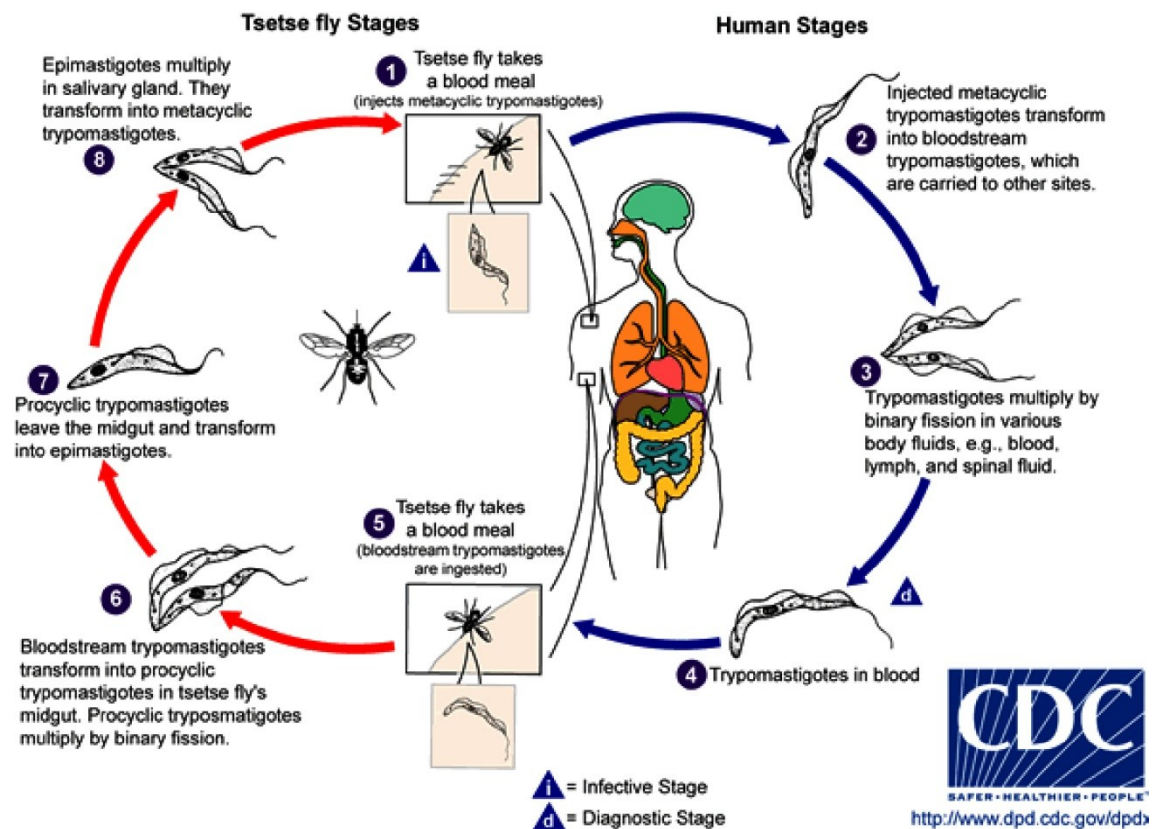


Figure 1.2. Overview of *T. brucei* lifecycle. Illustration courtesy of the Centre for Disease Control and Prevention, www.cdc.gov.

produce ATP (Bienen *et al.*, 1991). When infected blood is ingested by the Tsetse fly, the stumpy trypanosomes differentiated into proliferative procyclic trypomastigotes, in the fly's posterior mid-gut (Vickerman, 1985), over a period of 10-14 hours (h) of G1 arrest (Ziegelbauer *et al.*, 1990). This differentiation involves morphological and biochemical changes, including cell-body elongation, with a noticeable increase in the length of the post-kinetoplasmic portion of the body, and changes in the surface coat composition and energy metabolism. The VSG coat is substituted for a less-dense surface coat of GPI anchored glycoproteins, known as procyclins (Roditi and Liniger,

2002). The generation of energy also switches from glycolysis to a mitochondrion-base respiratory system, requiring the structural expansion and full metabolic activation of this organelle. The expansion of the mitochondrion, which becomes a branched structure with discoid cristae, is coupled to morphological changes of the glycosomes from spherical to bacilliform structures (*Vickerman, 1985*). In addition, procyclic forms have a different pattern of organelle movement during mitosis (*Robinson et al., 1995; Tyler et al., 2001*).

As the infection progress within the fly mid-gut, procyclic trypanosomes migrate across the peritrophic membrane into the ectoperitrophic space, and from here to the salivary glands, via the foregut, mouthparts and salivary ducts (*Lewis and Langridge, 1947; Vickerman et al., 1988; Van Den Abbeele et al., 1999*). During this migration, procyclic trypomastigotes differentiate into epimastigotes through an asymmetric cell division-cycle that starts with a trypomastigote and produces one long and one short epimastigote daughter cell (*Lewis and Langridge, 1947; Van Den Abbeele et al., 1999*). This differentiation involves reduction in cell length, achieved through the asymmetric cell division, and the apparent sacrifice of the longer daughter cell; exchange in relative position of the nucleus and kinetoplast along the anterior-posterior axis of the cell; and loss of the procyclin surface coat (*Urwyler et al., 2005*), which is probably substituted by alanine-rich proteins (*Urwyler et al., 2007*). Additionally, epimastigotes seem to be arrested in G1 phase, and do not appear to replicate until they reach the salivary glands (*Sharma et al., 2008*).

In the salivary glands, epimastigotes become attached by the flagellum to the microvilli of the epithelial cells of the gland lumen, through an elaboration of the flagellar membrane (*Tetley and Vickerman, 1985*). Attached epimastigotes become proliferative again, proliferation being probably stimulated by the attachment. A possible explanation of why trypanosomes rearrange the position of their organelles could be

that the repositioning of the basal body to the anterior half of the cell-body, nearer to the origin of the cleavage furrow, facilitates cytokinesis in attached trypanosomes (Sharma *et al.*, 2008). Dividing attached epimastigotes undergo further differentiation into non-proliferative metacyclic trypomastigotes, upon which depends their transmission to the mammalian host when the fly takes a blood meal (Tetley and Vickerman, 1985). Transition from epimastigotes to metacyclic trypomastigotes occurs through two intermediate states. First, epimastigotes transform into uncoated trypomastigotes, called premetacyclics, which remain attached and are still able to divide. Premetacyclics differentiate into still attached “nascent metacyclics” by acquiring a glycoprotein coat and ceasing division. The acquisition of the surface coat is accompanied by mitochondrion repression, which often becomes unbranched, with tubular cristae, and by transformation of bacilliform glycosomes into spherical structures. These morphological modifications suggest that surface coat acquisition is coupled to respiratory changes, pre-adapting the parasite for survival in the mammalian host (Vickerman, 1985). Mature metacyclic trypomastigotes are released into the salivary gland lumen from where they are able to establish an infection in the mammal, upon the fly's bite.

1.3. Cytoskeletal architecture in *T. brucei*.

The trypanosome cytoskeleton is dominated by microtubular structures. The reliance on a microtubule rather than actin-based cytoskeleton is evident from the genome annotation, which reveals that trypanosomes contain fewer components of a classic actin-dependent cytoskeleton than many other eukaryotes. Moreover, in trypanosomes actin has not been observed to associate with the cytoskeleton and its role seems restricted to the formation and trafficking of endocytotic vesicles, (Garcia-Salcedo *et al.*, 2004). Important microtubular structures are the subpellicular corset that accounts for the determination and maintenance of cell shape (Sherwin and Gull, 1989a; Ngo *et al.*,

1998), the flagellar axoneme, the basal bodies (Gull, 1999), and the flagellum attachment zone (FAZ), which contains four specialised microtubules, or microtubule quartet (MtQ), that are integrated in the cortical microtubular array (Fig.1.3).

T. brucei contains several members of the tubulin superfamily — α , β , γ , δ , ϵ and ζ -tubulin— (Vaughan *et al.*, 2000). Of these, α - and β -tubulins dimerise to constitute the building blocks of microtubules; γ -tubulin is present in microtubule organising centres (MTOCs) (Scott *et al.*, 1997) and essential for the nucleation of the axonemal central pair (CP) of microtubules (McKean *et al.*, 2003). RNAi depletion of δ -tubulin in *T. brucei* has been shown to produce cell paralysis, aberrant basal bodies and disruption of the CP orientation (Gadelha *et al.*, 2006); whereas ϵ -tubulin in *Paramecium* is essential for the assembly and anchorage of the centriolar microtubules (Dupuis-Williams *et al.*, 2002). Additionally, ζ -tubulins seem to be implicated in the biology of centrioles/basal bodies and the flagellum (Vaughan *et al.*, 2000).

1.3.1. The subpellicular microtubule corset.

Cell shape in *T. brucei* is determined by a monolayer of evenly arranged microtubules, of 24 nm diameter, that form an helical pattern along the longitudinal axis of the cell (Sherwin and Gull, 1989a) and are cross-linked to each other and to the plasma membrane (Sherwin and Gull, 1989a; Hemphill *et al.*, 1991a). The diameter of the *T. brucei* cell varies along its length and the number of subpellicular microtubules found in cross sections of the mid-body is higher than in cross sections of the anterior or posterior ends (Vickerman and Preston, 1976 ; Hemphill *et al.*, 1991b). However, although intermicrotubule spacing —18-22 nm— remains constant throughout the cell-body (Sherwin and Gull, 1989a), the length of individual microtubules varies within the subpellicular corset (Gull, 1999). In the early stages of the cell division-cycle, new microtubules are intercalated between old ones, suggesting a template morphogenesis of the microtubule array and semi-conservative distribution to the daughter cells

Figure 1.3.

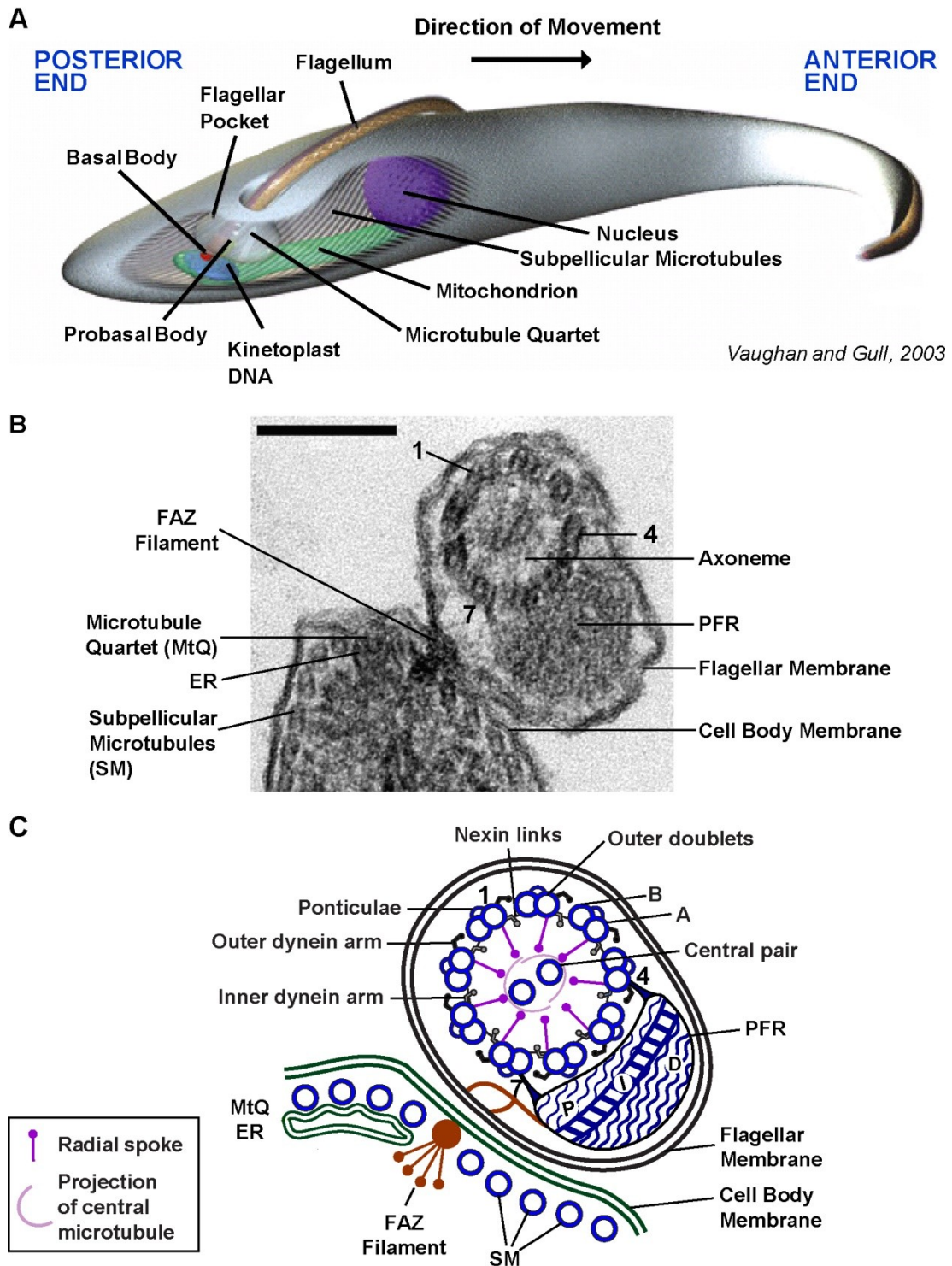


Figure 1.3. A. Scheme of a *T. brucei* procyclic cell adapted from Vaughan and Gull, 2003. **B.** Cross-section through the flagellum and flagellum attachment zone (FAZ) of a procyclic *T. brucei* cell, seen from the posterior end. Numbers 1, 4 and 7 refer to the numbering of the axonemal outer doublets according to Afzelius, 1959. **C.** Schematic representation of B modified from Bastin et al., 2000. Scale bar represents 200 nm. ER, smooth endoplasmic reticulum. PFR, paraflagellar rod. SM, subpellicular microtubules

(*Sherwin and Gull, 1989b*). The microtubules of the subpellicular corset have an intrinsic polarity, with the fast-growing or + ends at the posterior cell-end (*Robinson et al., 1995*).

The subpellicular microtubules are interconnected and linked to the plasma membrane by discrete projections of microtubule-associated proteins (MAPs). A number of MAPs have been identified from *T. brucei* and other trypanosomatids, however the molecular mechanisms which orchestrate cytoskeletal re-modelling during the life and cell-cycles, remain largely unknown. Among the so far identified MAPs are MARP-1 (microtubule-associated repetitive protein-1), CAP15 and CAP17 (cytoskeletal-associated proteins 15 and 17), WCB (whole cell-body protein) and CAP5.5 (cytoskeletal-associated protein 5.5). MARP-1 exclusively localises at the membrane-oriented face of individual microtubules and seems to have a stabilising effect (*Hemphill et al., 1992*). CAP15 and CAP17 are two low molecular weight MAPs, codified by paralogous, differentially expressed genes: CAP15 is expressed in bloodstream trypanosomes and CAP17 in procyclic trypanosomes (*Vedrenne et al., 2002*). CAP15 and CAP17 interact with the anterior, more stable portion of the cytoskeleton but, when overexpressed, both proteins localise throughout the entire length of the cell-body, leading to defects in cell morphogenesis, perhaps through hyper-stabilisation of the subpellicular microtubule array, causing problems for organelle (e.g. nucleus, kinetoplast) movement during the latest stages of the cell-cycle (*Vedrenne et al., 2002*). WCB, has been shown to be involved in the linking of microtubules to the plasma membrane, and its RNAi silencing results in aberrant cytokinesis and cell death (*Baines and Gull, 2008*). CAP5.5 (*Birkett et al., 1992; Matthews and Gull, 1994a*) is a stage specific regulated protein, expressed in procyclic trypanosomes, and evenly distributed across the cell body cytoskeleton but absent from the flagellum (*Hertz-Fowler et al., 2001*). CAP5.5 is myristoylated and palmitoylated at the N-terminus, suggesting an interaction with the plasma membrane

(Resh, 1999). In addition to its cytoskeletal association, CAP5.5 is a calpain-like protein, the first identified in trypanosomes, (Hertz-Fowler *et al.*, 2001), and later shown to be a member of an expanded calpain-like family of proteins in trypanosomatids (Ersfeld *et al.*, 2005). Recently, a CAP5.5 paralog has been identified in *T. brucei* (Jackson, 2007b). CAP5.5 stage specific regulation, cytoskeletal association and possible interaction with the plasma membrane pointed for an important role of this protein in *T. brucei*, which will be analysed in detail in the section 1.8.1 of this chapter.

1.3.2. The basal bodies and flagellar axoneme.

Cells at the beginning of the cell-cycle contain one mature basal body that subtends the flagellar axoneme, and a pro-basal body. As the cell-cycle progresses, the pro-basal body matures to a basal body, initialising the morphogenesis of the new flagellar axoneme. Additionally, two new pro-basal bodies are formed, each one next to a mature basal body (Sherwin and Gull, 1989a). The mature basal bodies have a tripartite structure, with the first 100 nm consisting of 9 microtubule triplets, followed by a transition zone of ~400 nm of 9 microtubule doublets, and a distal region termed the basal plate that exhibits the canonical 9+2 axonemal profile.

The axoneme of eukaryotic cilia and flagellar is a highly conserved and ordered structure responsible for ciliary and flagellar beating. The axoneme is constituted by nine outer-microtubule doublets and two central single microtubules —the 9+2 microtubular profile—. Additionally, the axoneme also contains a large number of associated sub-structures, such as the inner and outer dynein arms and nexin links, which connect the microtubule doublets to each other, radial spokes, which are projections that connect the microtubule doublets to the CP, and other microtubule elaborations essential for flagellar function (Fig. 1.3B-C). Axonemal microtubules have the opposite polarity to that of the subpellicular corset, with the minus ends located at the posterior cell-end, in the basal body, and the plus ends located at the flagellar tip.

During motility, flagellar bending occurs through a combination of sliding between adjacent outer doublets and constriction of this sliding by structures such as the nexin links between outer doublets and the anchoring at the basal body (*Satir, 1989, 1992*). Inter-doublet sliding is produced by the action of outer and inner dynein arms, (*King, 2000; Porter and Sale, 2000*). Additionally, outer dynein arms act by increasing beat frequency and adding power to beating (*King, 2000*). Interestingly, in addition to the conserved components that are important for the regulation of flagellar motility in trypanosomes and other eukaryotic flagellates (*Branche et al., 2006; Broadhead et al., 2006; Ralston et al., 2006*), the trypanosomes and their near relations contain another flagellar structure, the paraflagellar rod (PFR). The PFR is an elaborate, lattice-like structure that is physically attached to and runs alongside the axoneme (*Fuge, 1969; Vickerman, 1969; Bastin et al., 1996*). Although its precise function is not yet fully elucidated, the PFR is essential for motility in both bloodstream and procyclic cells (*Bastin et al., 1998; Broadhead et al., 2006*).

1.3.3. The Flagellum Attachment Zone (FAZ).

The FAZ in *T. brucei* is a complex structure that connects the flagellum to the long axis of the cell body, consisting on the FAZ filament, the MtQ, and associate membranous elements (*Sherwin and Gull, 1989a*). The FAZ filament, formed by abundant fibrous connections, is located in a gap between the microtubules of the subpellicular corset. The MtQ is located immediately to the left of the FAZ filament, when viewed from the posterior cell-end, and is associated to a specialised region of the smooth endoplasmic reticulum (ER) (Fig.1.3B-C). Interestingly, the MtQ, which is nucleated close to the basal body, has the opposite polarity to the microtubules of the subpellicular corset and distinct biochemical characteristics, such as salt-resistance (*Robinson et al., 1995*). Electron microscopy has shown that the FAZ filament is constituted by a row of punctuate electron-dense structures or “maculae” from which filaments emanate

towards the cytoplasm. Opposite to the FAZ filament, in the region between the juxtaposed membranes of the cell body and the flagellum, and between the flagellar membrane and the proximal domain of the PFR, there is a region of electron-dense material (*Sherwin and Gull, 1989a*). This region has been shown to be formed by filamentous structures, which seem to be anchored to the microtubules of the cell body cytoskeleton at one end and to interact with the paraflagellar rod (PFR) at the other end (*Sherwin and Gull, 1989a*). These filaments are often Y shaped, suggesting that they may consist of several polypeptide chains (*Hemphill et al., 1991a*).

During the cell division-cycle, the FAZ filament assembly lags slightly behind the new flagellum, as it exits the flagellar pocket, attaching the newly synthesised flagellum to the cell-body, following a left-handed helical path (*Kohl et al., 1999*) guided by the flagella connector (*Moreira-Leite et al., 2001*). The FAZ site, located between the old and new flagella in dividing cells, provide a positional link between the basal body/kinetoplast/ flagellar pocket area, at the posterior cell-end, and the site of cleavage initiation, at the anterior end (*Baum et al., 1981*), and is thought to define the position and polarity of the cleavage plane (*Robinson et al., 1995; Davidge et al., 2006*). At the FAZ, components of the flagellum and the cell body appear to be directly or indirectly linked to each other (*Sherwin and Gull, 1989a*). Moreover, these components probably depend on each other for their morphogenesis during cell duplication, with some components having to be formed before others, in a hierarchical manner (*Kohl et al., 1999*). Flagellum attachment to the cell-body has been shown to be essential for cell morphogenesis in both, procyclic and bloodstream trypanosomes (*LaCount et al., 2002*). RNAi knockdown of FLA1, a protein required for the attachment of the flagellum to the cell body, not only produced cells with a detached flagellum but also numerous anucleate zoids and multinucleate cells, suggesting the importance of flagellum attachment for the correct onset of cytokinesis (*LaCount et al., 2002*).

Despite the importance of flagellar attachment, only a few components of the *T. brucei* FAZ have been identified to date: FLA1, GM6, TbPLK and FAZ1. FLA1 is a plasma membrane glycoprotein that contains two transmembrane domains and localises preferentially at the FAZ and the flagellar pocket of *T. brucei* and *T. cruzi* (Nozaki *et al.*, 1996; Moreira-Leite *et al.*, 2001). RNAi knockdown of FLA1 in *T. brucei* results in flagellar detachment and cytokinesis defects (Moreira-Leite *et al.*, 2001). GM6 is a repetitive antigen that exclusively localises at the FAZ, associated to the FAZ filament (Muller *et al.*, 1992; Muller *et al.*, 1993). Additionally, GM6 and its orthologs in *T. cruzi* and *L. major*, FRA and Lcr1, respectively, have been found to be part of calpain-like proteins containing 67-68 amino acid repeats (Ersfeld and Gull, 2001). The functional analysis of *T. brucei* calpain-like proteins containing GM6 and Lcr1-like repeats will be presented in the Chapter 5 of this thesis. TbPLK is a *T. brucei* Polo-like kinase which has been shown to localise at the proximal region of the FAZ and may be involved in cytokinesis regulation (Umeyama and Wang, 2008). FAZ1, a recently characterised component of the FAZ, is a repeat-containing protein, associated to the FAZ filament that has been shown to be essential for the correct structure and function of the FAZ, important for flagellar attachment and cytokinesis (Vaughan *et al.*, 2008).

Additionally, several antibodies that recognise the FAZ and detect high molecular weight proteins by Western blotting, have been identified in *T. brucei* and *T. cruzi*: CD10, DOT1 (Woods *et al.*, 1989), Ag 2G4 (Ruiz-Moreno *et al.*, 1995), BS7 (Woodward *et al.*, 1995), L3B2 and L6B3 (Kohl *et al.*, 1999). However, only the protein recognised by L3B2 and L6B3, FAZ1 has been identified and characterised so far (Vaughan *et al.*, 2008).

1.4. The cell-cycle of procyclic *T. brucei*.

Although nuclear-S and kinetoplast-S phases, as well as the duplication and segregation of other organelles (e.g. new flagellum elongation, Golgi duplication, mitochondrial division) all need to occur in replicative stages of the *T. brucei* lifecycle, in the two replicative forms that are amenable to laboratory study and manipulation — the procyclic and bloodstream forms— there are some important differences in the regulation and, potentially, the timing of key cell-cycle-regulated events. The cell-cycle of the procyclic form (PCF) has been investigated to great detail, although the precise mechanisms of some of its major morphogenic events still remain to be fully elucidated (Sherwin and Gull, 1989a). The first detectable process of cell division in procyclic trypomastigotes is the initiation of kinetoplast replication, which begins immediately before the onset of nuclear S-phase. Kinetoplast S-phase is shorter than its nuclear counterpart, probably due to the smaller size of its genome (Woodward and Gull, 1990). The first detectable cytoskeletal event of the cell-cycle is the maturation of the pro-basal body into a basal body, shortly after the initiation of nuclear S-phase (Sherwin and Gull, 1989a). The assembly of the new flagellum starts in the flagellar pocket and, as its tip contact the lateral side of the old flagellum, the flagella connector (FC) is formed (Briggs et al., 2004). The new flagellum elongates at a constant speed of $\sim 3.6 \mu\text{m}$ (Bastin et al., 1999b), and, as it exits the flagellar pocket, remains attached to the old flagellum by the FC, ensuring the structural inheritance of the old flagellum left-handed helical pattern (Moreira-Leite et al., 2001). Thus, the FC is a mobile transmembrane junction, that migrates along the old flagellum as the new flagellum grows (Briggs et al., 2004). The new flagellum is also attached to the cell body by a newly assembled FAZ, which provides morphogenetic information for cytokinesis (Robinson et al., 1995; Gull, 1999; Ploubidou et al., 1999).

The initial extension of the new flagellum occurs without significant increase in inter-basal body distance. However, after the new flagellum has elongated to ~50% of the old flagellum length, there is a rapid increase in inter-basal body distance, as both, the kinetoplast and basal-body segregate. After this “critical point”, the new flagellum elongates to match the length of the old flagellum, but when it reaches ~60% of the old flagellum length the FC stops its migration. Moreover, after this point the growth of the new flagellum correlates with the increase in inter-basal body distance, suggesting that the force exerted by the new flagellum elongation, resisted by a halted FC could be the “pushing force” responsible for the separation of the basal bodies and the concomitant kinetoplast segregation (*Davidge et al., 2006*). This model is in agreement with the earlier observation that, during procyclic cell division, the kinetoplast associated to the new flagellum moves towards the posterior cell-end, whereas the one associated to the old flagellum migrates closer to the nucleus (*Sherwin and Gull, 1989a; Woodward and Gull, 1990*).

After the completion of nuclear mitosis, the posterior daughter nucleus drifts between the two segregated kinetoplasts and cytokinesis initiates with the formation of the cleavage furrow. The ingression of the cleavage furrow occurs on the longitudinal axis of the cell, from the anterior to the posterior end, in a position probably determined by the FAZ (*Moreira-Leite et al., 2001; Vaughan et al., 2008*), distributing a single nucleus, kinetoplast and flagellum in each daughter cell (*Sherwin and Gull, 1989a*). Interestingly, disruption of the mitotic spindle of procyclic trypomastigotes does not prevent cytokinesis, resulting in one daughter cell with an un-segregated nucleus and one daughter anucleate cytoplasm —termed zoid—containing a single kinetoplast and a flagellum (*Ploubidou et al., 1999*). Thus, procyclic trypanosomes do not require the completion of nuclear mitosis to undergo cytokinesis.

1.5. The cell-cycle of bloodstream *T. brucei*.

Although the cell-cycle of bloodstream form (BSF) *T. brucei* at the cytoskeletal level has not been studied with the great detail of PCF, and although many similarities exist between the cell-cycle of the two developmental forms, several differences in numerous division stages have been identified. To begin with, the cell division-cycle of BSF trypanosomes is much faster than in PCF, with a doubling time of ~6.3h (Tyler *et al.*, 2001), as oppose to the 8.5h of procyclic trypanosomes (Sherwin and Gull, 1989a). The flagellar growth rate of the BSF is also significantly higher, being more than two folds of the PCF ($8 \mu\text{m.h}^{-1}$ versus $3.6 \mu\text{m.h}^{-1}$) (Bastin *et al.*, 1999b; Tyler *et al.*, 2001). Additionally, the kinetoplast of bloodstream *T. brucei* is located in a more posterior position than in the PCF (Vickerman, 1965) and its repositioning during cell division follows a different pattern that procyclic trypanosomes, with the two segregated kinetoplasts never being anterior to any of the segregated nuclei (Tyler *et al.*, 2001). The absence of a flagella connector from bloodstream trypanosomes, which has not been detected so far (Briggs *et al.*, 2004), could explain this difference in cell architecture. In the absence of a FC that resists the force exerted by the new flagellum growth, this could never have the “pushing effect” towards the posterior cell-end over the new kinetoplast (Briggs *et al.*, 2004).

Interestingly, cell motility is essential for cytokinesis in BSF *T. brucei* (Broadhead *et al.*, 2006; Griffiths *et al.*, 2007), whereas in the PCF paralysed parasites are still able to complete cytokinesis (Bastin *et al.*, 1999a; Dawe *et al.*, 2007), although this is a controversial issue (Branche *et al.*, 2006; Ralston *et al.*, 2006). Cell-cycle regulation is also different in bloodstream and procyclic *T. brucei*. In the PCF, kinetoplast segregation and flagellar growth are sufficient to trigger cytokinesis, as shown by induced mitotic block through RNA interference of the cyclin CYC6, which results in cells with un-segregated nucleus that can complete cytokinesis producing an anucleate

zoid. However, in BSF trypanosomes, mitotic block by CYC6 RNA interference resulted in major cytokinesis defects (*Hammarton et al., 2003*).

1.6. RNA interference.

RNA silencing processes are widespread in almost all eukaryotic organisms, where they are implicated in several functions, as genome protection for unwanted nucleic acids, genome rearrangement, control of gene expression during development and heterochromatin formation (*Baulcombe, 2004; Meister and Tuschl, 2004; Ullu et al., 2004; Matzke and Birchler, 2005; Zamore and Haley, 2005*). RNA interference (RNAi) is an ATP-dependant mechanism through which double stranded RNA (dsRNA) induces the specific destruction of homologous mRNAs into 21-26 bp dsRNAs, termed short interfering RNAs (siRNAs) (*Elbashir et al., 2001*), resulting in post-translational knockdown of gene expression (*Fire et al., 1998*). This degradation is carried out by Dicer, a member of the ribonuclease III (RNase III) family of dsRNA-specific endonucleases (*Bernstein et al., 2001*). The resulting siRNAs are then incorporated into a, still inactive, enzymatic complex, known as the RNA-induced silencing complex (RISC). Unwinding of the siRNAs remodels the complex generating an active RISC. The active RISC acts by scanning RNA and targeting for destruction those sequences which are complementary to the guide strand of the siRNAs (*Nykanen et al., 2001*). The RISC complex contains RNA helicases, Argonaute (AGO) proteins and, in plants, fungi and worms, RNA-dependant RNA polymerases (*Hutvagner and Zamore, 2002*). *T. brucei* contains a single member of the AGO protein family, TbAGO1, and a Dicer-like protein, TbDcl1, both of which are required for RNAi in trypanosomes (*Durand-Dubief and Bastin, 2003; Shi et al., 2004; Shi et al., 2006*). The presence and utility of RNAi in *T. brucei* was demonstrated very shortly after its discovery by *Fire et al.* in *C. elegans* (*Bastin et al., 1998; Fire et al., 1998; Ngo et al., 1998*). Since then, RNAi has

become one of the most important tools for downregulation of gene expression in *T. brucei* bloodstream and procyclic forms.

In *T. brucei*, several vector systems have been developed to produce stable cell lines capable of inducible RNAi in only one round of transfection, as opposed to at least two for a genomic knockout (LaCount *et al.*, 2000; Wang *et al.*, 2000). The most widely used RNAi constructs are the pZJM and the p2T7-177 vectors. (LaCount *et al.*, 2000; Wang *et al.*, 2000; Wickstead *et al.*, 2002) These RNAi vectors are stably integrated into the *T. brucei* genome by homologous recombination when they are transfected into genetically engineered *T. brucei* procyclic and bloodstream cell lines, 29.13 and 90.13 respectively. These RNAi cell lines constitutively express the tetracycline-repressor and the T7 RNA polymerase, to allow induction and tight regulation of RNAi (Wirtz and Clayton, 1995; Wirtz *et al.*, 1998; Wirtz *et al.*, 1999).

The production of dsRNA directed against the target gene can be attained by two different approaches. The first method consists in flanking a region of the gene to be silenced by a pair of inducible and opposing T7 promoters, where RNAi can be “switched on” by addition of tetracycline to the culture medium. The second approach consist in inserting inverted repeats of a sequence of the target gene, separated by a DNA “stuffer”, downstream of a single inducible promoter (hairpin RNAi constructs) (Bastin *et al.*, 2000; Shi *et al.*, 2000). The presence of the stuffer fragment that separates the back to back repeats is necessary for the propagation of the plasmid in bacteria due to instability of adjacent long inverted repeats (Shi *et al.*, 2000). Hairpin RNAi constructs seems to produce a more efficient RNA silencing, what could be explain by the fact that dsRNA expressed from back to back repeats is self-folding, forming a stem-loop, that increase its stability, facilitating its export to the cytoplasm where the RNAi action is presumed to take place (Durand-Dubief *et al.*, 2003).

Comparatively, dsRNA expressed from the dual facing promoters is formed by annealing of two separate RNA molecules, sense and anti-sense, and lacks the 5' and 3' processing sequences, what would explain lower cytosolic concentrations. However, direct comparison of dsRNA levels produced by the two approaches has not been performed due to dsRNA intrinsic instability, being rapidly degraded into short fragments of low reliable detection. Additionally, hairpin RNAi constructs seem to produce less leakiness (RNAi in the absence of induction) than head-to-head RNAi systems, what has been attributed to the high strength of the T7 promoter (*Motyka and Englund, 2004*) and to the fact that induction control is easier when dsRNA is expressed from a single promoter constructs than from double, opposing promoter constructs, where two promoters need to be regulated (*Durand-Dubief et al., 2003*). Nevertheless, the T7 head to head promoters are more widely used because of the ease of cloning (one cloning step as opposed to two in the stem-loop constructs).

1.7. Calpain family of cysteine proteases.

1.7.1. Introduction to calpains.

Calpains constitute a large protein family of Ca^{2+} -regulated cysteine proteases that are widely distributed in eukaryotes. They are associated with intracellular organelles – cytoskeleton, endoplasmic reticulum, Golgi apparatus or plasma membrane– and are implicated in a wide spectrum of cellular events, including cytoskeletal remodelling, signal transduction, apoptosis, cellular differentiation, adhesion and migration. In humans, multiple pathologies are linked to the calpains:

- Genetic alterations: Limb girdle muscular dystrophy, gastric cancer and Type 2 diabetes.
- Conditions leading to loss of calcium homeostasis that alters calpain activity: Alzheimer's disease, cataract formation, neuronal ischemia or Multiple sclerosis.

The best characterised calpains are the mammalian, ubiquitously expressed μ - and m-calpains, whose names refer to the micromolar (μ) and millimolar (m) Ca^{2+} concentrations require for their optimal activation (Suzuki, 1991). However, despite extensive studies, several aspects of calpain function remain unclear –e.g. *in vivo* regulation, physiological functions and substrates (Goll *et al.*, 2003).

1.7.2. Calpain properties.

Conventional calpains —calpains in vertebrates and higher eukaryotes— are generally heterodimeric proteins, composed by a large 80 kDa and a small 28 kDa subunit (Fig.1.4). Conventional μ - and m-calpains share the same small subunit or calpain-4. Both, the large and small subunits have a highly modular structure, being constituted by well defined domains. The large subunit is classically divided into four domains on the basis of its amino acid sequence¹ (Suzuki, 1990; Hosfield *et al.*, 1999):

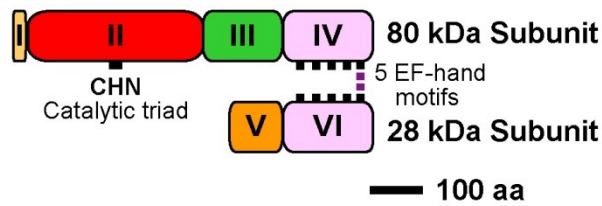
- **Domain I:** The N-terminal domain of conventional calpains is only 18-20 amino-acids long and forms a single α -helix. Conventional calpain domain I sequences are highly conserved among different higher eukaryotes (72-86% identity), but have no sequence homology with any other known proteins in the databases (Goll *et al.*, 2003). Although domain I function is still not known, the α -helix seems to be important for stability and activation of calpains, as it contains the site of autolysis cleavage that occurs prior or parallel to calpain proteolysis of external substrates (Sorimachi and Suzuki, 2001; Goll *et al.*, 2003). Additionally, some calpain proteins have unique N-terminal sequences, including Zn-fingers and transmembrane sequences (Sorimachi and Suzuki, 2001).

¹ The crystallographic structure of m-calpain has shown that its 80 kDa subunit has six “domains”, two of them constituted by only 17 and 18 amino acids each (Strobl *et al.*, 2000). However, in the literature most authors have continued to use the previous four domains division.

Figure 1.4.

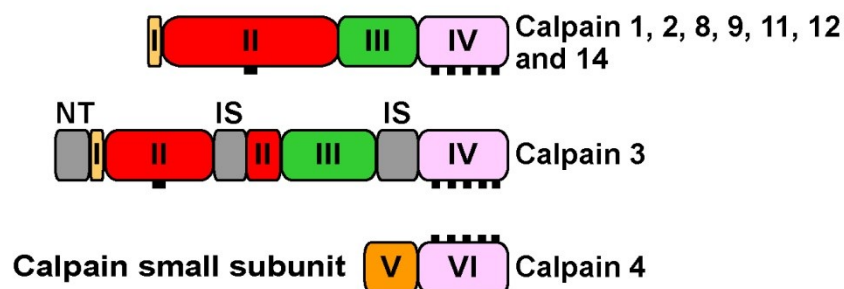
A

TYPICAL CALPAIN HETERODIMER



B

TYPICAL CALPAINS



ATYPICAL CALPAINS

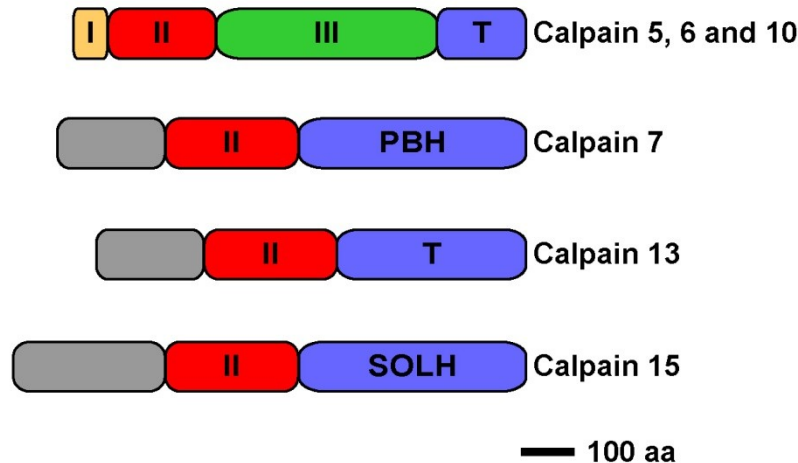


Figure 1.4.

A. Typical calpain heterodimer formed through the fifth EF-hand of the small and large subunits.

B. Domain structure of the human calpain family. Typical calpains are composed of four domains (I-IV). Calpain 1 and 2 form a heterodimer with the small subunit or calpain 4, composed of two domains (V and IV). Atypical calpains have the basic structure of the large subunit of conventional calpains but lack the EF-hand domain IV. Some calpain-like sequences have some domains replaced (domain I) or deleted (domain III). All sequences contain a calpain "catalytic" domain II, although some have deviations from the canonical catalytic triad (calpain 6). PBH: homologous to the PalB domain of AAA-ATPases; NT: N-terminal sequence of calpain 3, 20-30 aa long and rich in proline; calpain 3 IS1 and IS2: inserted sequences 1 (containing autolytic sites) and 2 (containing a nuclear translocation signal); T-domain: homologous to the C-terminal domain of *C. elegans* TRA-3 protein, similar to the C-2 domain and calpain domain 3; SOLH, homologous to the C-terminal domain of *Drosophila* SOL protein.

- **Domain II:** Calpain domain II sequences (330 amino acids) are highly conserved, with 85 to 93% amino acid identity in higher eukaryotes. Calpain domain II contains the catalytic triad typical of cysteine proteases: cysteine (positions 115 and 105 of μ - and m-calpains, respectively), histidine (positions 272 and 262) and asparagine (positions 296 and 286) —CHN—, being also known as calpain catalytic or protease domain. However, calpain domain II shares little homology with other cysteine proteases, having likely evolved from a different ancestral gene (*Goll et al., 2003*). Conventional calpain domain II also shows the conservation of amino acid residues close to the catalytic triad, aspartate and proline —D, preceding the active C, and P, next to the active N. X-ray crystallographic studies have shown that calpain domain II is constituted by two sub-domains, sub-domain IIa that contains the active C, and sub-domain IIb that contains the active H and N. Each of these sub-domains bind an atom of Ca^{2+} in an acidic peptide loop, which are essential for the correct alignment of residues within the active site (*Moldoveanu et al., 2002*). The binding of these two Ca^{2+} ions requires the presence of five conserved acidic residues (*Moldoveanu et al., 2004*). All identified calpain domains II contain a highly conserved motif unique to calpains and constituted by lysine, alanine, tyrosine, alanine and lysine —the KAYAK-motif. This motif is always located at the centre of calpain domain II, although its function is not known.

- **Domain III:** The domain III of calpain proteins (~150 amino acids) has no homology to other known protein sequences. The crystal structure of m-calpain revealed an eight-stranded antiparallel β -sandwich conformation for domain III, that has been propose to act as a linker or amplifier of conformational changes between domain II and the EF-hand containing domain IV (*Vilei et al., 1997; Hosfield et al., 1999; Strobl et al., 2000; Reverter et al., 2001*). The structure of calpain domain III also resembles that of the C-2 domain found in several Ca^{2+} -regulated proteins such

as protein kinase C and synaptotagmins (*Rizo and Sudhof, 1998; Duncan et al., 2000*). C-2 domains bind phospholipids in a Ca^{2+} -dependent manner and are thought to be responsible for orchestrating Ca^{2+} -dependant translocation to membranes and regulation of enzyme activity. The high similarity between the structure of calpain domain III and the C-2 domain strongly suggest that calpain domain III may be a lipid binding site, also responsible for Ca^{2+} -dependent translocation of calpains to the membrane and for orchestrating Ca^{2+} and lipid activation of these enzymes (*Tompa et al., 2001*). Additionally, calpain domain III contains an acidic loop constituted by a stretch of consecutive aspartic acid and glutamic acid residues. The function of this acidic loop has been proposed to be the regulation of calpain activity through an electrostatic switch induced by calcium binding, either directly to the acidic loop, to the EF-hand motifs of domain IV or to both (*Fernandez-Montalvan et al., 2004*).

- **Domain IV:** The domain IV of conventional calpains (~170 amino acids) contains five EF-hand calcium-binding motifs, which are essential for protease activity and for dimerisation with the small subunit (through the fifth EF-hand motif). Domain IV sequences are marginally homologous to calmodulin, which has four EF-hand motifs. This homology, together with the modular structure of calpains, led to the hypothesis that calpains have evolved by fusion of a calmodulin-like gene with a protease precursor domain (*Strobl et al., 2000*). However, the presence of five EF-hand motifs identified calpains as members of the penta-EF-hand protein family, known to dimerise through their fifth EF-hand and to associate with membranes (*Xie et al., 2001*).

The small subunit of conventional calpains, calpain-4, which role has not been fully established, is organised in two domains:

- **Domain V:** The N-terminal domain (~100 amino acids), also known as calpain hydrophobic domain, is rich in glycine, containing stretches of up to twenty

consecutive G.. Although the amino acid composition of this domain is not characteristic of a highly hydrophobic domain, as G is near the centre of the hydrophobicity scale, it has been suggested that domain V binds to phospholipid and has a role in interacting with cell membranes (*Gil-Parrado et al., 2003*).

- **Domain VI:** The C-terminal domain of the small subunit (~160 amino acids), resembles domain IV in that it also contains five EF-hand Ca^{2+} binding motifs, the fifth of which is involved in dimerisation with the large subunit (*Goll et al., 2003*).

1.7.3. Calpains activity and regulation.

Activation of conventional requires the presence of Ca^{2+} to induce conformational changes in the calpain structure, reducing the distance between the sub-domain IIa, containing the active C, and the sub-domain IIb, containing the active H and N, thus forming an active catalytic triad. However, *in vitro* activity assays have shown that the Ca^{2+} concentration required for calpain protease activity are much higher than the 50-300 nM that exist in living cells. Therefore, there must be some mechanisms that reduce the Ca^{2+} requirements for calpain activation *in vivo*, although they have not been conclusively demonstrated. Among the proposed mechanisms are: activation of calpains in the presence of Ca^{2+} “spikes” in localised cellular areas; interaction with the plasma membrane near a Ca^{2+} channel; and the presence of different molecules, including activator proteins, that reduce calpains Ca^{2+} requirements.

Additionally, μ - and m-calpains autolyse rapidly in the presence of Ca^{2+} , which results in the excision of several amino acids from the N-terminus of their two subunits. However, the biological significance of calpains autolysis remains controversial, as both autolysed and un-autolysed calpains can have proteolytic activity (*Goll et al., 2003*). Calpain activity *in vivo* seems to be regulated not only by altering the Ca^{2+} concentration required for proteolytic activity, but also by the intracellular location or

access to substrate and by subsite² specificity. The subsite specificity of calpains is very limited, as they selectively cleave or not very close related proteins, generally at a limited number of sites, producing large polypeptide fragments (Goll *et al.*, 2003). Typical calpains are not able to cleavage small peptides (Sasaki *et al.*, 1984), indicating that the subsite recognition by calpains uses a quite large area of the polypeptide substrate (Goll *et al.*, 2003). However, several studies have shown that the subsite specificity of calpains is not governed by the amino acid sequence, but by the conformations of the polypeptide chain (Harris *et al.*, 1988; Croall *et al.*, 1996; Stabach *et al.*, 1997), suggesting that calpain activity may be regulated not only by modifying the calpains (calpastatin) but also by altering the substrates (Goll *et al.*, 2003).

A large number of proteins (>100) are cleaved by calpains in *in vitro* assays. The majority of these proteins belong to one of four general categories: cytoskeletal proteins, especially those involved in cytoskeletal/plasma membrane interactions; kinases and phosphatases; membrane-associated proteins, including receptor and ion-channel proteins; and some transcription factors (Goll *et al.*, 2003). However, *in vitro* cleavage of a protein by calpains does not necessarily indicate that it is an *in vivo* calpain substrate. Similarly, there is no conclusive evidence that *in vitro* substrates cannot also be cleaved *in vivo*. Nevertheless, it is clear that not all potential calpain substrates in a cell will be cleaved during a single cellular event (Goll *et al.*, 2003). The preferential calpain substrates identified to date include cytoskeletal proteins —as

² Peptidases subsites were defined by Schechter and Berger in 1967 who observed that the large active sites of some proteases could be divided into different “subsites”, located on both sides of the catalytic core of the active site, each one accommodating one amino acid residue of the peptide substrate. The active site of an enzyme has the double function of binding the substrate and catalysing the reaction, which efficiency determines the enzyme overall activity (e.g. specificity) towards a particular substrate. As each subsite of an enzyme has its own stereospecificity and specific interactions with different amino acids side-chains, the reactivity of a certain bond in a protein depends not only on the two residues forming the bond but also on the nature of the residues in its neighbourhood or adjacent subsites, which explain the unexpected cleavages observed in some proteins (Schechter and Berger, 1967).

integrin and spectrin— (*Kulkarni et al., 2002; Takamure et al., 2005; Glantz et al., 2007*), membrane proteins —as plasma membrane Ca^{2+} ATPase— (*Ariyoshi et al., 1993*), components of receptor-mediated signal transduction —as cytokine receptors— (*Noguchi et al., 1997; Oda et al., 2002*) and cytosolic proteins that, under specific conditions, are targeted to membranes —as protein kinase C— (*Aragon et al., 2002*).

Two motifs frequently found in calpain substrates are PEST sequences and calmodulin-binding domains. The PEST sequence is rich in proline, glutamic and aspartic acids, and threonine; having an acidic character that could sequester Ca^{2+} , thus creating a microambient of high Ca^{2+} concentration that would activate calpains. However, deletion of the PEST sequences does not decrease substrate degradation by calpains (*Molinari et al., 1995*). On the contrary, modification of calmodulin binding domains of calpain substrates decrease their susceptibility to the enzymes (*Molinari et al., 1995*).

1.7.4. Tissue localisation and physiological functions.

The well-characterised calpains, μ -calpain and m-calpain, are exclusively intracellular, generally associated to organelles as the cytoskeleton, endoplasmic reticulum, Golgi apparatus and plasma membrane (*Watt and Molloy, 1993; Azarian et al., 1995; Di Lisa et al., 1995; Yamazaki et al., 1997; Shevchenko et al., 1998; Hood et al., 2003*). The expression of these conventional calpains seem to be ubiquitous, as one or more have been detected in all vertebrate cells examined for their presence (*Goll et al., 2003*). However, different tissues/cells differ widely in their ratios of these two proteins. For example, skeletal muscle of domestic animals contains approximately equal amounts of μ - and m-calpains, whereas human platelets and erythrocytes have no detectable m-calpain, and bovine platelets and vascular smooth muscle have no detectable μ -calpain (*Thompson and Goll, 2000*). In addition to μ -calpain (also known as calpain-1) and m-calpain (also known as calpain-2) several other calpain proteins have been

identified in mammals: calpain-3 to calpain-15. Some of these proteins have the same domain architecture than μ - and m-calpains —calpain 3, 8, 9, 11, 12 and 14—and have been classified as typical or conventional calpains. However, other calpain-related sequences identified in mammals lack an EF-hand containing domain IV—calpain-5, 6, 7, 10, 13 and 15— and have been classified as atypical calpains or calpain-like proteins.

Some of these typical and atypical calpains are also ubiquitously expressed —calpain-4 (small subunit), calpain-5, calpain-7, calpain-10, calpain-12 and calpain-13—, whereas others have been shown to have tissue specific expression. Calpain-3, also known as p94, is preferentially expressed in skeletal muscle and its lost-of-function mutation causes limb girdle muscular dystrophy type 2A (*Sorimachi et al., 1989; Sorimachi et al., 1997*). Calpain-6 is only expressed during embryogenesis, in the placenta and the embryo (*Dear and Boehm, 1999*). Calpain-8 is preferentially expressed in smooth muscle, being abundant in the gastric mucus cells (*Hata et al., 2006*). Calpain-9 is an active protease preferentially expressed in the digestive track (*Lee et al., 1998*), where it seems to play a tumour-suppressing role (*Yoshikawa et al., 2000*). Calpain-11 expression is mainly restricted to testis (*Dear et al., 1999*), where it reach high levels during spermatogenesis and localises in the spermatozoa, being probably involved in regulating Ca^{2+} -dependent signal transduction events (*Ben-Aharon et al., 2006*). Calpain-13 expression have only been detected, and at low levels, in testis and lungs (*Dear and Boehm, 2001*). Additionally, some ubiquitously expressed calpains are more abundant in determined tissues, as calpain-5, which levels are characteristically high in colon, small intestine and testis; and calpain 12, which is found at high levels in the hair follicle.

Human calpains have many isoforms and multiple splice variants that difficult the establishment of relationships between calpain gene sequences, protein isoforms and

their distinct or overlapping functions. However, many of these calpains have been experimentally shown to be implicated in cytoskeletal/membrane attachments (*Santella et al., 2000*), cell division (*Yamaguchi et al., 1994*), cell differentiation (*Simonson et al., 1985; Kwak et al., 1993; Cottin et al., 1994*), secretion (*Sato et al., 1995*), motility (*Glading et al., 2002; Perrin and Huttenlocher, 2002*), signal transduction pathways (*Inomata et al., 1996*), cell-cycle regulation (*Santella et al., 1998*), control of gene expression (*Xu and Mellgren, 2002*), apoptosis (*Squier et al., 1994; Kubbutat and Vousden, 1997; Nakagawa and Yuan, 2000*), gastric acid secretion (*Yao et al., 1993*), long-term potentiation in the central nervous system (*Hell et al., 1996*) and embryonic development (*Emori and Saigo, 1994*)

1.7.5. Atypical calpains.

During the last twenty years, genome sequencing and cloning has led to the identification of multiple calpain-related sequences in many different organisms. The usual criterion for the identification of calpain family members is sequence homology to domain II of conventional calpains, with amino acid identity over 20-25% threshold³. This inclusive approach allowed the identification of numerous calpain-related sequences in mammals, invertebrates, lower eukaryotes, plants and some bacteria: 14 in mammals, 4 in *Drosophila*, 12 in *C. elegans* and 2 in fungi and yeast, transmembrane calpains in some plants and Tpr calpain-like protease in *Porphyromonas gingivalis* (*Goll et al., 2003*). Surprisingly, this type of analysis identified much larger numbers of calpain-related proteins in unicellular organisms, as trypanosomes (*Ersfeld et al., 1999*) and ciliates (*Eisen et al., 2006*). Many of the calpain-related sequences identified in this manner lack an EF-hand containing domain

³ The percentage of proteins with the same class of function decreases drastically when sequence identity falls below 35% (*Wilson et al., 2000*).

and showed widely divergent properties outside the catalytic domain II. These sequences were termed “atypical” calpains or calpain-like proteins (Fig.1.4).

Phylogenetic analysis of known conventional and atypical calpains has shown that these highly modular proteins have evolved through gene fusion and truncation as well as tandem versus chromosomal duplication events (*Jekely and Friedrich, 1999; Croall and Ersfeld, 2007*). These analyses also revealed a clear segregation of the EF-hand-containing calpain family members from those lacking EF-hands (*Jekely and Friedrich, 1999*) and showed that the acquisition of the penta-EF-hand domain involved in Ca^{2+} binding, and the subsequent formation of calpain heterodimers, was a relatively late event in calpain evolution (*Jekely and Friedrich, 1999; Croall and Ersfeld, 2007*). Furthermore, calpain-like proteins containing N-terminal EF-hands identified in *Tetrahymena thermophila* are also phylogenetically well separated from animal calpains containing C-terminal EF-hands (*Croall and Ersfeld, 2007*). This indicates that the acquisition of EF-hands by these groups may have occurred through independent gene-fusion events.

Some calpain-like sequences identified in invertebrates and lower eukaryotes showed high homology to conventional calpains and several of them have been experimentally demonstrated to have Ca^{2+} dependent protease activity — as *Schistosoma mansoni* (*Andresen et al., 1991; Karcz et al., 1991; Siddiqui et al., 1993*), *Drosophila* (*Pinter and Friedrich, 1988; Emori and Saigo, 1994; Theopold et al., 1995*), Crustacean (*Mykles and Skinner, 1982, 1986; Beyette et al., 1997*) and *C. elegans* proteases (*Barnes and Hodgkin, 1996*). However, the majority of calpain-related sequences found in invertebrates and lower eukaryotes are atypical calpains, with little homology to conventional calpains outside the protease domain II. These atypical calpains do not have calmodulin-like EF-hand sequences in their domain IV, some even lack a domain IV altogether, and it is uncertain if they are able to bind Ca^{2+} . Many of them also have

deviations from the canonical calpain catalytic triad and may not be proteolytically active. The identification of calpain-like proteins with >35% sequence identity to conventional calpains domain II, but lacking one or more of the typical catalytic triad residues, implied that calpains have a function other than/in addition to proteolytic activity. Additionally, genetic evidence shows that many atypical calpains are involved in signal transduction cascades, tissue differentiation and sex determination (*Goll et al., 2003*).

Atypical calpains in vertebrates —calpain-5, calpain-6, calpain-7, calpain-10, calpain-13 and calpain-15 (Fig.1.4) — are monomeric, as they lack the EF-hand containing domain IV. Instead, calpains 5, 6, 10 and 13 contain C-terminal domains homologous to that present in *C.elegans* TRA-3 protein, known as the TRA-3 like T-domain (*Dear et al., 1997; Horikawa et al., 2000*). The T-domain of *C. elegans* TRA-3 and its homologues in calpain 5, 6, 10 and 13, are C2 domains, similar to the C2-like domain of calpain domain III, and represent a Ca^{2+} -binding membrane module, which is involved in signal transduction and transmembrane traffic (*Barnes and Hodgkin, 1996; Tonami et al., 2007*). Additionally, calpain-7 possess a C-terminal domain homologous to the PalB domain of AAA-ATPases, which is a microtubule interacting and transport domain, also known as MIT, involved in signal transduction (*Denison et al., 1995*) The PalB homologous domain or PBH domain of calpain-7 is also present in calpain-like proteins of *Aspergillus nidulans* and *Saccharomyces cerevisiae*, and is thought to be responsible for intracellular traffic of molecules and the association of calpain-7 with microtubules (*Franz et al., 1999*). Calpain-15 is the vertebrate homolog of *Drosophila melanogaster* small optic lobe protein, SOL, having significant homology to SOL's catalytic (II) and C-terminal domains (SOL homologous domain, SOH, of unknown function) (*Kamei et al., 1998*). Calpain-15 also contains Zn-fingers motifs at its N-terminus (*Pinter et al., 1992; Kamei et al., 1998*). Although atypical calpains lack

EF-hand Ca^{2+} binding motifs, their possible interaction with Ca^{2+} cannot be excluded, as it could be accomplished by other structures within their sequences, domain II, domain III and the C-terminal or T-domains.

1.8. Expanded calpain-like family in trypanosomatids.

The majority of protozoan genomes studied so far contain a single calpain-coding gene —apicomplexan parasites as *Plasmodium falciparum*, *Theileria annulata* and *Cryptosporidium parvum*; or the amitochondrial parasite *Entamoeba histolytica*— (Wu *et al.*, 2003; Abrahamsen *et al.*, 2004; Gardner *et al.*, 2005; Loftus *et al.*, 2005). However, kinetoplastid parasites —as *T. brucei*, *T. cruzi* and *Leishmania* species— and the ciliate *Tetrahymena thermophila* have shown a large expansion of calpain-related genes. —12 in *T. brucei*, 15 in *T. cruzi*, 17 in *Leishmania major* and 26 in *T. thermophila*— (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005; Ersfeld *et al.*, 2005; Ivens *et al.*, 2005; Eisen *et al.*, 2006).

1.8.1. CAP5.5: the first identified calpain-like protein in trypanosomatids.

As discussed earlier, *T. brucei* CAP5.5 protein was the first calpain-like sequence found in kinetoplastids (Hertz-Fowler *et al.*, 2001). CAP5.5 was first identified in 1992 through the screening of λ gt11-cDNA library with a polyspecific/polyclonal antiserum raised against *T. brucei* cytoskeletons (Birkett *et al.*, 1992). This approach allowed the isolation of several cDNA clones containing sequences that coded for cryptic cytoskeletal proteins. One of the resulting monospecific, polyclonal antisera recognised a protein of around 120 kDa which gave an immunofluorescence signal over the whole cell body but was absent from the flagellum. The cell-body staining was resistant to extraction with non-ionic detergents and followed the helical pattern of the subpellicular microtubules, showing that the antigenic protein colocalised with the cell-body

cytoskeleton. The clone was called λ 5.5 and the protein that it coded for was termed, based on its localisation, cytoskeletal associated protein 5.5 or CAP5.5 (*Birkett et al., 1992*).

The sequencing and characterisation of CAP5.5 revealed a high degree of similarity between its central domain and the catalytic domain of conventional μ and m-calpains, also known as CAPN1 and CAPN2, with an overall identity of 23% and a 41% similarity. However, although CAP5.5 central domain or domain II was termed as “catalytic” due to its similarities to the catalytic domain II of conventional calpains, it has major deviations from the canonical calpain C-H-N catalytic triad, being unknown if its doubly substituted S-Y-N triad retains any proteolytic activity. Additionally, CAP5.5 sequence only has two conserved acidic residues of the five needed by conventional calpain catalytic domains to bind the two Ca^{2+} ions that are essential for the correct alignment of the amino acids of the catalytic triad within the active site (*Hertz-Fowler et al., 2001; Moldoveanu et al., 2002; Moldoveanu et al., 2004*), (Fig. 1.5A). Although the similarity of CAP5.5 to calpain proteins initially seemed limited to its central domain, as it lacked the Ca^{2+} -binding EF-hand domain IV of typical calpains, the analysis of its sequence revealed a four-domain modular structure similar to that of human atypical calpain-6 (CAPN6), which also lacks an EF-hand containing domain and has deviations in the catalytic triad (*Hertz-Fowler et al., 2001*), (Fig. 1.5B).

CAP5.5 expression is tightly regulated in a lifecycle dependant manner, being expressed in procyclic form trypanosomes throughout all stages of the cell-cycle but absent from long-slender and short-stumpy bloodstream forms (*Matthews and Gull, 1994a; Hertz-Fowler et al., 2001*). Furthermore, the gain of CAP5.5 after the initiation of bloodstream form differentiation into procyclic forms has been shown to be irreversible (*Matthews and Gull, 1994a*). A CAP5.5 monoclonal antibody (mAb) was raised against a β -galactosidase- λ 5.5 cDNA fusion-protein, which showed an immunofluorescence

Figure 1.5.

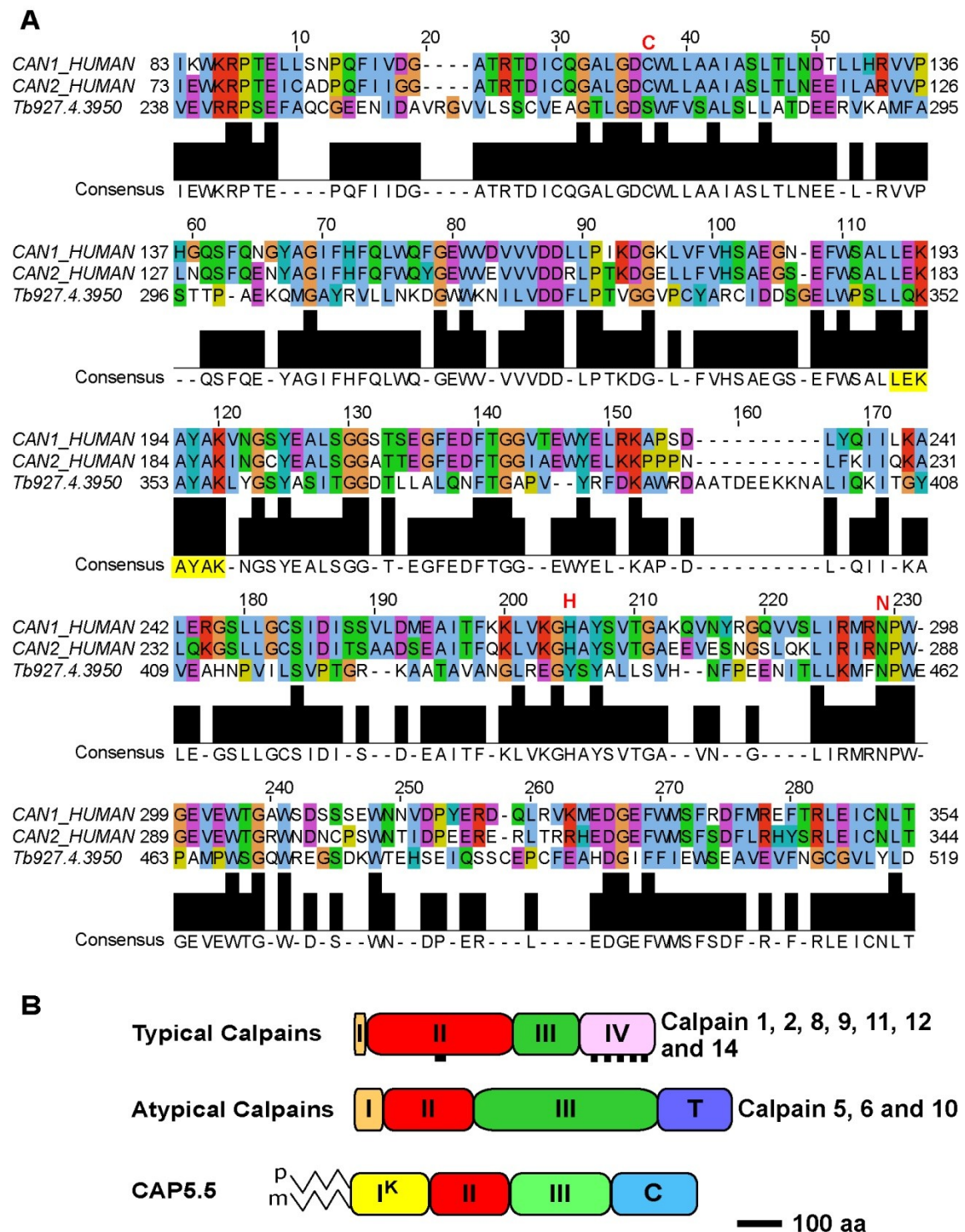


Figure 1.5. A. Alignment by the MAFFT method of the catalytic domains of human calpains CAN-1 and CAN-2 and the corresponding domain of CAP5.5. Note CAP5.5 deviations from the catalytic triad (in red font above the sequences) of conventional calpains and its conservation of the KAYAK motif (highlighted in yellow), exclusive of calpains and calpain-like sequences. The colour scheme is according to the ClustalX colour code. **B:** Schematic representation of the domain architecture of typical, atypical calpains and CAP5.5. Scale bar represents 100 amino acids (aa).

labelling pattern identical to that of the affinity purified polyclonal antiserum and identified a protein of the same size (120 kDa) by western blotting. Biochemical fractionation of *T. brucei* cells into Triton X-100 soluble and insoluble fractions followed by western blot analysis showed that CAP5.5 was exclusively present in the insoluble cytoskeletal fraction (*Hertz-Fowler et al., 2001*).

The anti-CAP5.5 mAb, when used on methanol-fixed whole cells, gave a diffuse staining of the cell body with the exception of the flagellum. However, when used on Triton X-100 extracted cells, the CAP5.5 staining followed the helical organisation of the subpellicular cytoskeleton. In addition, a single or double dot that colocalised with the basal bodies was intensely stained (*Hertz-Fowler et al., 2001*). The anti-CAP5.5 mAb was also used in electron microscopy studies on whole-mount cytoskeletons. Immunogold localisation of CAP5.5 gave the same labelling pattern as that obtained with the immunofluorescence techniques, labelling the subpellicular cytoskeleton, but being absent from the axonemal microtubules (*Hertz-Fowler et al., 2001*). Southern blot analysis was carried out to determine if CAP5.5 was a single copy gene, but restriction analysis showed the presence of additional cDNA fragments not predictable from the CAP5.5 sequence, suggesting a complex genomic organisation of the CAP5.5 gene or the presence of a partial copy of CAP5.5 in the *T. brucei* genome (*Hertz-Fowler, 2001*).

Analysis of the *T. brucei* annotated genome in geneDB (*Hertz-Fowler et al., 2004*), identified CAP5.5 coding gene, Tb927.4.4330, which is located in a subtelomeric region of chromosome 4. Tb927.4.4330 is 2558 bp long, encoding a protein of 853 amino acids with a predicted molecular mass of 94.7kDa, approximately 20kDa smaller than the observed mass of the protein by western blots and a statistically predicted pI of 4.1. The acidic pI was confirmed by two-dimensional electrophoresis followed by western blotting with CAP5.5 monoclonal antibody (*Hertz-Fowler et al., 2001*). A low pI is relevant not only because it is characteristic of many other microtubule associated

proteins (*Kreis and Vale, 1999*) but because it argues against a basic protein not-specifically attached to the exposed acidic projections of microtubules (*Hertz-Fowler et al., 2001*).

Interestingly, CAP5.5 N-terminal domain showed the presence of a double acylation consensus motif – MGCGGS – for myristoylation at the subterminal glycine residue and palmitoylation at the cysteine residue. Double acylation of the CAP5.5 protein was confirmed by metabolic labelling (*Hertz-Fowler et al., 2001*) and could explain the apparent larger molecular mass of CAP5.5 by sodium-dodecyl-sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). This type of dual acylation is exclusively found in membrane-associated proteins (*Resh, 1999*). Myristoylation targets proteins to hydrophobic lipid domains within the cell, being necessary for protein–membrane interactions, and is a prerequisite for subsequent palmitoylation, which then establishes high-affinity interactions with membranes (*Godsel and Engman, 1999; Resh, 1999*). Conversely, cellular fractionation experiments had previously shown that CAP5.5 remains into the cytoskeletal insoluble fraction, indicating a strong interaction with the microtubule corset. The close proximity of the cortical cytoskeleton to the plasma membrane (10 nm) and the uniform spacing between microtubules (18-20 nm) make it very difficult to determine if a protein is implicated in membrane-microtubule interactions, in interactions between microtubules or in other functions, even by immunogold electron microscopy. However, the presence of both N-terminus dual acylation and sequestration within the cytoskeletal fraction suggested a possible function of CAP5.5 in the cross-linking of the plasma membrane with the microtubule corset (*Hertz-Fowler et al., 2001*).

Further analysis of CAP5.5 N-terminus sequence also revealed the presence of a proline-rich region similar to the binding consensus motif found in many ligands of SH3-type domains – PXXP – (*Ren et al., 1993; Yu et al., 1994*). SH3-type domains are

found in a great variety of intracellular and membrane associated proteins, such as proteins with enzymatic activity, adaptor proteins that lack catalytic activity and in cytoskeletal associated proteins, e.g. fodrin and yeast actin binding protein 1, ABP-1 (Musacchio *et al.*, 1992; Mayer and Baltimore, 1993; Pawson, 1995). The presence of this proline-rich motif could be important in the search for CAP5.5 binding partners, within newly identified membrane and cytoskeletal associated proteins, paying special attention to SH3-type domain containing sequences. CAP5.5 domain III has a 19% overall identity to domain III of conventional calpains. Although CAP5.5 does not have an EF-hand containing domain IV and its domain III does not contain an acidic loop, secondary structure analysis predicted a high probability of β -strand conformation in CAP5.5 domain III, consistent with the C2-domain structure of conventional calpains and could function as a transmitter/amplifier in Ca^{2+} -independent regulatory mechanisms (Ersfeld *et al.*, 2005). The recent discovery of a large chromosome duplication event in *T. brucei*, involving the subtelomeric regions of chromosomes 4 and 8 (Jackson, 2007b), and the bioinformatic analysis of *T. brucei* calpain like proteins carried out during this thesis work, led to the identification of a CAP5.5 paralog on chromosome 8, Tb927.8.8330, which I have termed CAP5.5V (V for variant), which only differs significantly from CAP5.5 in its C-terminal domain.

Human calpain-6 (or CAPN6), which has the most remarkable similarities to CAP5.5, has been positively shown to be associated with microtubules (Tonami *et al.*, 2007). Calpain-6 is the only vertebrate calpain that lacks a conserved catalytic triad, with the active cysteine residue substituted by lysine in humans, and may not be proteolytically active (Dear *et al.*, 1997). Calpain-6 domain IV also lacks EF-hands Ca^{2+} binding motifs and its sequence resembles that of the terminal T-domain of *Caenorhabditis elegans* TRA-3, a calpain-like sex determination factor (Sorimachi *et al.*, 1997). However, the function of the T-domain remains unclear as it is not essential for TRA-3 activity

(Barnes and Hodgkin, 1996). Calpain-6 domain III also resembles CAP5.5 in that it lacks the sequence corresponding to the conserved acidic loop important for the calcium sensitivity of calpains. Furthermore, it has been shown that the interaction of calpain-6 with microtubules is mainly mediated by domain III, although the implicated motifs are still unknown (Tonami *et al.*, 2007). Calpain-6 over-expression in fibroblasts promotes the formation of stable cytoplasmic microtubule bundles while its knock down by RNAi produce actin reorganisation and membrane ruffling. These data suggest that calpain-6 is a microtubule-stabilising protein involved in regulation of the cytoskeletal organisation (Tonami *et al.*, 2007). Considering the *in vitro* ability of calpains to cleave tubulin and microtubule associated proteins (Goll *et al.*, 2003), the loss of proteolytic activity could have been a necessary early step in calpain 6 evolution as a microtubule-stabilising protein.

1.8.2. Bioinformatic identification and analysis of a large calpain-like family in kinetoplastids.

The release of *T. brucei*, *T. cruzi* and *L. major* genome sequences allowed a whole-genome analysis approach for the search and characterisation of calpain-related proteins in these trypanosomatid species (Ersfeld *et al.*, 2005). Ersfeld and co-workers used the “catalytic” domain II of CAP5.5 as bait for the database mining of the predicted coding sequences of the tree parasites in geneDB. They initially identified 12 calpain-related sequences in *T. brucei*, 15 in *T. cruzi* and 17 in *L. major* (Ersfeld *et al.*, 2005). However, the bioinformatic analysis and comparison of the full length of these sequences revealed the presence of a novel N-terminal sequence element outside domain II that was fairly conserved among many of the analysed sequences. This N-terminal sequence element corresponded to the location of conventional calpains domain I but was significantly longer (~100 amino acids longer). Moreover, it did not show any similarity to typical calpain domain I or indeed to any other sequences

currently in the databases, being exclusively present in kinetoplastids. Database searches within the genome of the three studied parasites showed the presence of this sequence element not only at the N-terminus of calpain-like sequences but also as the core element of several short open reading frames (ORFs), with an average length of ~200 amino acids (*Ersfeld et al., 2005*). The unique correlation between kinetoplastid calpain-like proteins and this short sequence element led Ersfeld and co-workers to include the short “stand alone” sequences in the dataset, raising the total number of calpain-related proteins to 18 sequences in *T. brucei*, 24 in *T. cruzi* and 27 in *L. major* (*Ersfeld et al., 2005*). However, the N-terminal sequence element is clearly not a calpain related domain, being only present in some calpain-like sequences that contain a calpain “catalytic” domain. The identification of such high numbers of calpain-related sequences in kinetoplastid parasites exceeding the numbers found in most other organisms, including vertebrates, was both surprising and intriguing, perhaps pointing to trypanosomatid-specific functions (*Ersfeld et al., 2005*).

Ersfeld and co-workers carried out multiple alignments and comparisons of all known kinetoplastid calpain-related sequences, identifying and characterising their different domains, some of which showed clear similarities to those of conventional calpains (*Ersfeld et al., 2005*):

1. Domain I. The domain I equivalents of kinetoplastid calpain-related sequences were grouped in two categories according to their sequence composition: domain I^K and domain I^H.

- **Domain I^K** (K for kinetoplastids): Domain I^K corresponds to the newly identified N-terminal sequence elements exclusive of kinetoplastids. As discussed above, this sequence element is present at the N-terminus of many domain II-containing calpain-like sequences and as short “stand-alone” genes that lack a calpain domain II.

Comparison between these kinetoplastid short sequences revealed a high degree of conservation, with an average amino acid identity of ~50% and ~70% similarity. These values rose between putative orthologs up to 60% and 80%, respectively. Domain I^K is characterised by the presence of three highly conserved motifs: glycine, leucine, leucine and phenylalanine or tyrosine —G-L-L-F/Y— towards the N-terminus; tryptophan, alanine, phenylalanine, tyrosine, asparagine, aspartate and threonine —W-A-F-Y-N-D-T— at the centre; and valine, tyrosine, proline, any, glutamate, threonine and glutamate —V-Y-P-x-E-T-E— towards the C-terminus (*Ersfeld et al., 2005*). These motifs are unique to kinetoplastids, lacking any similarity to other functional motifs in the databases, and their functions remain unknown.

- **Domain I^H** (H for heterogeneous): Domain I^H includes the N-terminal sequences found in some domain II-containing kinetoplastid calpain-like proteins that are heterogeneous in both intra- and interspecies comparisons, with significant similarities only found between putative orthologues. These sequences are exclusively present at the N-terminus of kinetoplastid calpain-like sequences and never as “stand-alone” genes. Domain I^H heterogeneous sequences are not related to any motifs present in the databases, with the exception of Tb.927.6.3310 and its *T. cruzi* ortholog, which showed some homology to the N-terminus of *T. brucei* cAMP-dependent kinases regulatory subunit (*Ersfeld et al., 2005*).

2. Domain II. The domain II of kinetoplastid calpain-like sequences showed an overall identity to conventional calpain catalytic domains of 25%, 45% similarity. Domain II of conventional calpains are characterised by the presence of several motifs unique to calpain proteases. Many of these motifs were conserved in domain II of kinetoplastid calpain-related sequences, although with major differences (*Ersfeld et al., 2005*):

- Deviations from the catalytic triad that constitute the active site of conventional

calpains —C-H-N—, essential for protease activity, which was only conserved in 7 of the 44 identified domain II sequences: *T. brucei* Tb11.02.2240; *L. major* LmjF04.0450, LmjF25.1480, LmjF31.0460 and LmjF33.2010; *T. cruzi* TcCALP855.40 and TcCALP227.130.

- Substitution in some sequences of the critical amino acids close to the catalytic triad and typical of conventional calpains —aspartate, preceding the active cysteine, and proline, next to the active asparagine— although they were conserved in the majority of kinetoplastid calpain-like sequences.
- Deviations from the KAYAK-motif, unique to calpains domain II. Although this motif was highly conserved in kinetoplastids, some calpain-like proteins showed deviations from its canonical sequence.
- Partial conservation of the five amino acid residues critical for the binding of two Ca^{2+} ions by calpains domain II, which are essential for the correct alignment of the catalytic triad within the active site (*Moldoveanu et al., 2002; Moldoveanu et al., 2004*). These five residues were completely conserved in only one kinetoplastid calpain-like sequence, LmjF25.1480, which also showed a completely conserved catalytic triad

3. Domain III. The majority of kinetoplastid domain II-containing sequences also contained a sequence that showed some degree of similarity to domain III of conventional calpains (15-23% identity, 25-35% similarity). As discussed earlier, conventional calpains domain III has an eight-stranded antiparallel β -sandwich conformation, structurally similar to the membrane-targeting C2-domain (*Rizo and Sudhof, 1998*). Through this spatial conformation, conventional calpains domain III is thought to act as a transmitter and amplifier of the conformational changes induced to domain IV EF-hand motifs by the binding of Ca^{2+} (*Hosfield et al., 1999; Strobl et al.,*

2000; Reverter *et al.*, 2001). Although none of the kinetoplastid calpain-like sequences have an EF-hand containing domain IV, secondary structure analysis predicted a high probability of β -strand conformation between their domains II and IV (Ersfeld *et al.*, 2005). This β -stranded conformation is comparable to conventional calpains domain III eight-stranded antiparallel β -sandwich structure. Thus, the domain III of trypanosomatid calpain-like sequences could also act as a transmitter and amplifier of conformational changes, in a Ca^{2+} independent mechanism, and be involved in protein-membrane interactions (Ersfeld *et al.*, 2005).

4. Domain IV. The C-terminal domains of trypanosomatid calpain-related proteins do not show any similarity to conventional calpains EF-hand containing domain IV or to any other C-terminal domains found in atypical calpains. Moreover, the domain IV equivalents of kinetoplastid calpain-related sequences do not show any similarities to each other, except between putative orthologues, or to any identifiable functional motif, being not known if any of them is able to bind Ca^{2+} . The domain IV equivalents of kinetoplastid calpain-like sequences were termed as domain C (C for C-terminus) to distinguish them from other calpain domains IV (Ersfeld *et al.*, 2005).

5. Tandem Repeats. Seven of the kinetoplastid calpain-related sequences identified by Ersfeld and co-workers contained long arrays of 67-68 amino acids residues tandem repeats, predicted to form a coiled-coil structure. Of these sequences, five contained three copies of domains II and III, with the second and third copies separated by the repeat arrays —Tb11.47.0035, LmjF27.0490, LmjF27.0500, LmjF27.0510 and TcCALP721.30—. However, two sequences contained only one copy of domains II and III, located after an N-terminal array of tandem repeats —Tb11.47.0036 and TcCALP441.10. (Ersfeld *et al.*, 2005). Sequence fragments containing the repeats of some of these sequences have been published as antigens FRA, GM6 and Lcr1 in

T. cruzi, *T. brucei* and *L. chagasi*, respectively (Lafaille et al., 1989; Muller et al., 1992; Wilson et al., 1995). Sequencing of FRA, GM6 and Lcr1 showed a high degree of homology between them. Additionally, the antibodies against the three antigens detected high molecular weight proteins by western blotting. Moreover, FRA and GM6 have been shown to localise at the FAZ in *T. cruzi* and *T. brucei*, correspondingly (Cotrim et al., 1990; Muller et al., 1992; Cotrim et al., 1995)

6. N-terminal acylation. Many short and long domain I^K-containing sequences, showed the presence of an N-terminus double acylation modification motif: a subterminal glycine, for myristoylation, followed by a cysteine, for palmitoylation, before the beginning of domain I^K consensus sequences. The presence of glycine and cysteine is necessary but not sufficient for acylation to occur, requiring also the presence of an extended motif within the first 10-15 N-terminal amino acids (Maurer-Stroh et al., 2002a). Bioinformatic analysis of the N-terminus of domain I^K sequences showed that acylation is highly likely to occur (Ersfeld et al., 2005). Furthermore, in CAP5.5 myristoylation and palmitoylation have been experimentally demonstrated (Hertz-Fowler et al., 2001). Acylated proteins are often associated to the cytoplasmic face of membranes and lipid rafts, where they are involved in signal transduction events (Resh, 1999).

After the sequence analysis and characterisation summarised above, Ersfeld and co-workers classified the identified kinetoplastid calpain-like proteins in five different groups, according to their domain architecture and composition (Ersfeld et al., 2005), (Fig.1.6):

- **Group1: I^K-II-III-C.** Constituted by a kinetoplastid-specific domain I^K, some containing single or double acylation modification motifs, followed by a calpain domain II, a kinetoplastid calpain domain III and a C-terminal domain lacking EF-hands motifs.

Figure 1.6.

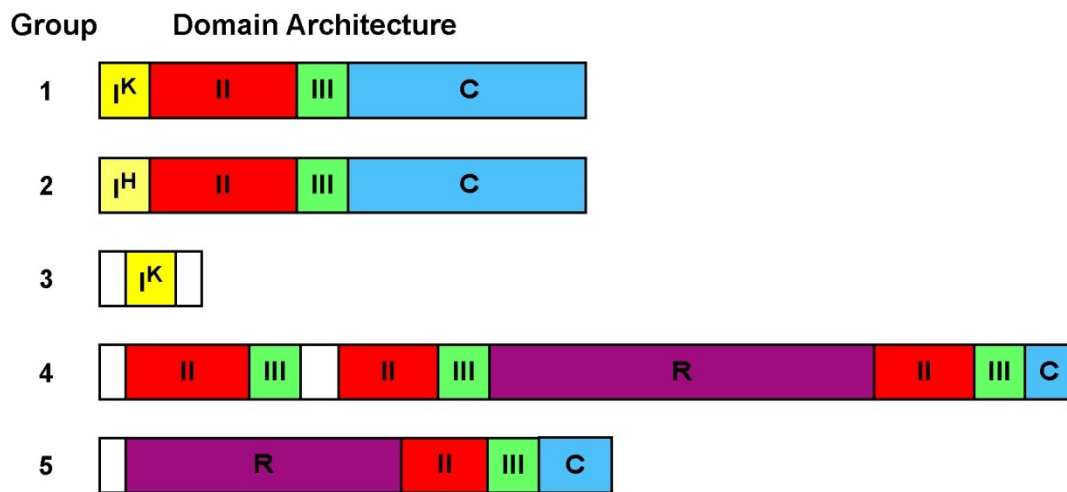


Figure 1.6.

Classification of trypanosomatid calpain-like proteins by *Ersfeld et al., 2005* according to their domain architecture and composition. Cartoon modified from *Ersfeld et al., 2005*, with the exception of group 5, which was corrected to match their description.

-
- **Group 2: I^H-II-III-C.** Constituted by sequences similar to those of Group 1 but containing a heterogeneous domain I^H instead of a typical kinetoplastid calpain-like domain I^K.
 - **Group 3: I^K.** Constituted by short ORFs containing only a kinetoplastid-specific domain I^K. These sequences often contain single or double acylation modification motifs. Additionally, one member of this group —Tb10.70.5060— contained four degenerate tandem repeats of domain I^K (4xI^K).
 - **Group 4: II-III-II-III-Repeats-II-III.** Constituted by sequences containing three copies of kinetoplastid calpain-like domains II and III, the second and third copies separated by tandem repeats of 65-68 amino acids residues unit length.
 - **Group 5: Repeats-II-III.** Includes sequences with N-terminal tandem repeats similar to those of Group 4 and followed by a single copy of calpain-like domains II and III.

1.8.3. Localisation and function.

Several kinetoplastid calpain-like proteins have been shown to have specific cellular localisations (*Cotrim et al., 1990; Muller et al., 1992; Hertz-Fowler et al., 2001; Tull et al., 2004; Liu et al., 2009*): cytoskeleton, cell-body and flagellar membranes and/or cytosol, which most relevant examples I summarise below: CAP5.5, small myristoylated proteins and repeats-containing calpain-like sequences.

CAP5.5.

CAP5.5 is a group 1 calpain-like protein that specifically localises at the cell-body cytoskeleton but is absent from the flagellum (*Hertz-Fowler et al., 2001*).

Flagellar Small Myristoylated Proteins.

SMP-1, a group 3 calpain-like protein containing only a domain I^K, specifically localises at the inner leaflet of the flagellar membrane (*Tull et al., 2004*). SMP-1 is also double acylated at the N-terminus, palmitoylated and myristoylated. Dual acylation of SMP-1 has been shown to be essential for this protein localisation to the detergent-resistant flagellar membrane. In addition, the flagellar targeting of SMP-1 is not dependent on interactions with the axoneme or other cytoskeletal proteins of the flagellum (*Tull et al., 2004*). However, the N-terminal double acylation of CAP5.5 is not sufficient to target this protein to the flagellum. As discussed above, CAP5.5 is associated to the subpellicular cytoskeleton and absent from the flagellum. However, the double acylation of CAP5.5 N-terminus could mediate an interaction with the plasma membrane, suggesting a role for CAP5.5 in cross-linking the plasma membrane to the microtubule corset (*Hertz-Fowler et al., 2001*).

Repeat-containing calpain-like proteins localise at the FAZ.

Some of the identified repeat-containing calpain-like sequences (groups 4 and 5) have been previously characterised as cytoskeletal associated proteins GM6, in *T. brucei*,

and FRA, in *T. cruzi*, that specifically localise at the FAZ, (Lafaille *et al.*, 1989; Muller *et al.*, 1992). Considering that several cytoskeletal associated proteins in Kinetoplastids have been shown to contain internal repetitive structures (Gull, 1999), suggests that the repeats of these calpain-like proteins are the structures that mediate the interaction with microtubules, whereas the presence of calpain-like domains II acts as an additional element of functionality.

1.9. Aims and strategy.

The objective of this thesis was to investigate the function of representative members of the calpain-like family in *T. brucei*. To this end I used the most recent genome annotation of *T. brucei* (version 4 update), *L. major* and *L. infantum* on the geneDB database at <http://www.geneDB.org/tryp/> (Hertz-Fowler *et al.*, 2004) to update the bioinformatics survey of kinetoplastid calpain-like sequences carried out by Ersfeld and co-workers and then I exploited the power of RNAi to survey the function of representative calpain-family members in *T. brucei*. The key findings from my work were:

- The identification of several new calpain-like sequences in the databases —10 in *T. brucei*, 10 in *L. major* and 11 in *L. infantum*— and reclassification of the repertoire of calpain-like proteins in these organisms into five tailored groups (Chapter 3).
- The identification of CAP5.5 and its paralog CAP5.5V as essential for cell morphogenesis in procyclic and bloodstream form trypanosomes, respectively (Chapter 4).
- The identification of *T. brucei* calpain-like sequences containing 68 amino-acid tandem repeats as essential for the correct morphogenetic pattern of procyclic trypanosomes (Chapter 5).

CHAPTER 2

MATERIALS AND METHODS

2. MATERIALS AND METHODS.

2.1. Bioinformatics: database mining and sequence analysis.

Genome databases of *T. brucei*, *L. major* and *L. infantum* in geneDB, <http://www.genedb.org>, were searched using BLASTp, an homology search program based on string matching algorithms (Altschul *et al.*, 1997; Camacho *et al.*, 2009) using different domains of known kinetoplastid calpain-like sequences as baits: CAP5.5 domains I, II, and III, and the different I^H and C domains from all kinetoplastid calpain-like sequences previously described by Ersfeld *et al.*, 2005. The highest scoring sequences from each BLASTp search were compared to known kinetoplastid calpain-like proteins to identify new calpain-like sequences in the dataset. Multiple alignments were performed using the MAFFT tool, a multiple sequence alignment based on the fast Fourier transform that allows rapid detection of homologous segments, reducing the central processing unit (CPU) time and has an improved and simplified scoring system, in which the similarity matrix is fixed for any input sequences, the extension gap penalty is not explicitly contained on the dynamic programming algorithm and parameters such as sequence divergence, length, presence of large insertions and deletions, are carefully adjusted to improve accuracy (Kato *et al.*, 2002). Multiple alignments were manually edited on the Jalview platform (Waterhouse *et al.*, 2009). The colour scheme of protein alignments is according to the ClustalX colour code, which follows the rules explained in Figure 2.1. Sequence domains not related to calpains were investigated by multi-organisms BLASTp searches within the geneDB and NCBI databases at <http://www.genedb.org/tryp> and <http://www.ncbi.nlm.nih.gov>. All sequences were scanned for motifs using InterPro, at <http://www.ebi.ac.uk>, SMART, at <http://www.smart.embl-heidelberg.de> or Pfam, at <http://www.pfam.sanger.ac.uk>.

Figure 2.1.

RULES	
BLUE	(W,L,V,I,M,F): {50%, p}{60%, wlvimafcyhp} (A): {50%, p}{60%, wlvimafcyhp}{85%, t,s,g} (C): {50%, p}{60%, wlvimafcyhp}{85%, s}
RED	(K,R): {60%, kr}{85%, q}
GREEN	(T): {50%, ts}{60%, wlvimafcyhp} (S): {50%, ts}{80%, wlvimafcyhp} (N): {50%, n}{85%, d} (Q): {50%, qe}{60%, kr}
PINK	(C): {85%, c}
MAGENTA	(D): {50%, de,n} (E): {50%, de,qe}
ORANGE	(G): {always}
CYAN	(H,Y): {50%, p}{60%, wlvimafcyhp}
YELLOW	(P): {always}

Figure 2.1. The ClustalX colour code. (Thompson *et al.*, 1997)

The colours shown above are applied according to the rules next to each colour box, which are specified in the following way:

(A,C,D): {50%, p,q,rstv} {85%, w,y}

The column residue is given first in the round brackets; more than one may be specified, in which case the rules apply to each of these residues.

The rule or rules are given in curly braces; only one rule has to be met for the colour to be applied. The minimum percentage is given first, followed by the residue or residues which must meet or exceed this percentage within the column.

If a group of residues is concatenated together, such as 'rstv', then any combination of these residues in total must meet or exceed the given percentage for the colour to be applied.

For residues or residue groups separated by commas, at least one of these must by itself exceed the percentage.

Secondary structure predictions were performed using the JNet method (Cuff and Barton, 2000; Cole et al., 2008). Reciprocal BLASTp was used for the identification of putative orthologous ($p < 1e^{-10}$ and reciprocal BLAST hit) and paralogous sequences within the *T. brucei*, *T. cruzi*, *L. major* and *L. infantum* genome databases, employed to build synteny maps.

2.2. Molecular biology.

2.2.1. RNAi constructs.

To make RNAi constructs, fragments of 300-1000 base pairs (bp) of the gene of interest were amplified by a polymerase-chain-reaction (PCR) from *T. brucei* genomic DNA, with DNA primers containing restriction sites added at their 5'-ends to allow ligation into the p2T7-177 vector (Wickstead et al., 2002). Restriction sites were always preceded by 2-3 random nucleotides to increase digestion effectiveness. In this vector the target sequence is flanked by opposing T7 promoters to produce dsRNA of the target gene. (Fig. 2.2A). For CAP5.5 and CAP5.5V specific silencing, I PCR amplified 500-600 bp fragments comprising the 3'-end of each gene coding sequence (CDS) and ~200 bp of their 3'-untranslated regions (UTRs). For the RNAi functional analysis of repeats-containing calpain-like sequences I performed different RNAi approaches:

1. Targeting the N-terminus of the two larger sequences, Tb11.57.0008 and Tb11.47.0035, which have a 99.9% identity over 1031 bp (Tb11.57.0008 49-1079 bp and Tb11.47.0036 46-1076 bp) of their CDSs (Appendix Chapter 5, Group 4 seq.). From these regions, two DNA fragments were PCR amplified from genomic DNA, having the same start point (same forward primer) but different length (different reverse primers): 601 bp and 921 bp respectively (Tb11.57.0008 85-685 bp and 85-1005 bp, and Tb11.47.0035 82-682 bp and 82-1002 bp). These fragments were used to make two different types of RNAi constructs (Fig. 2.2B):

Figure 2.2.

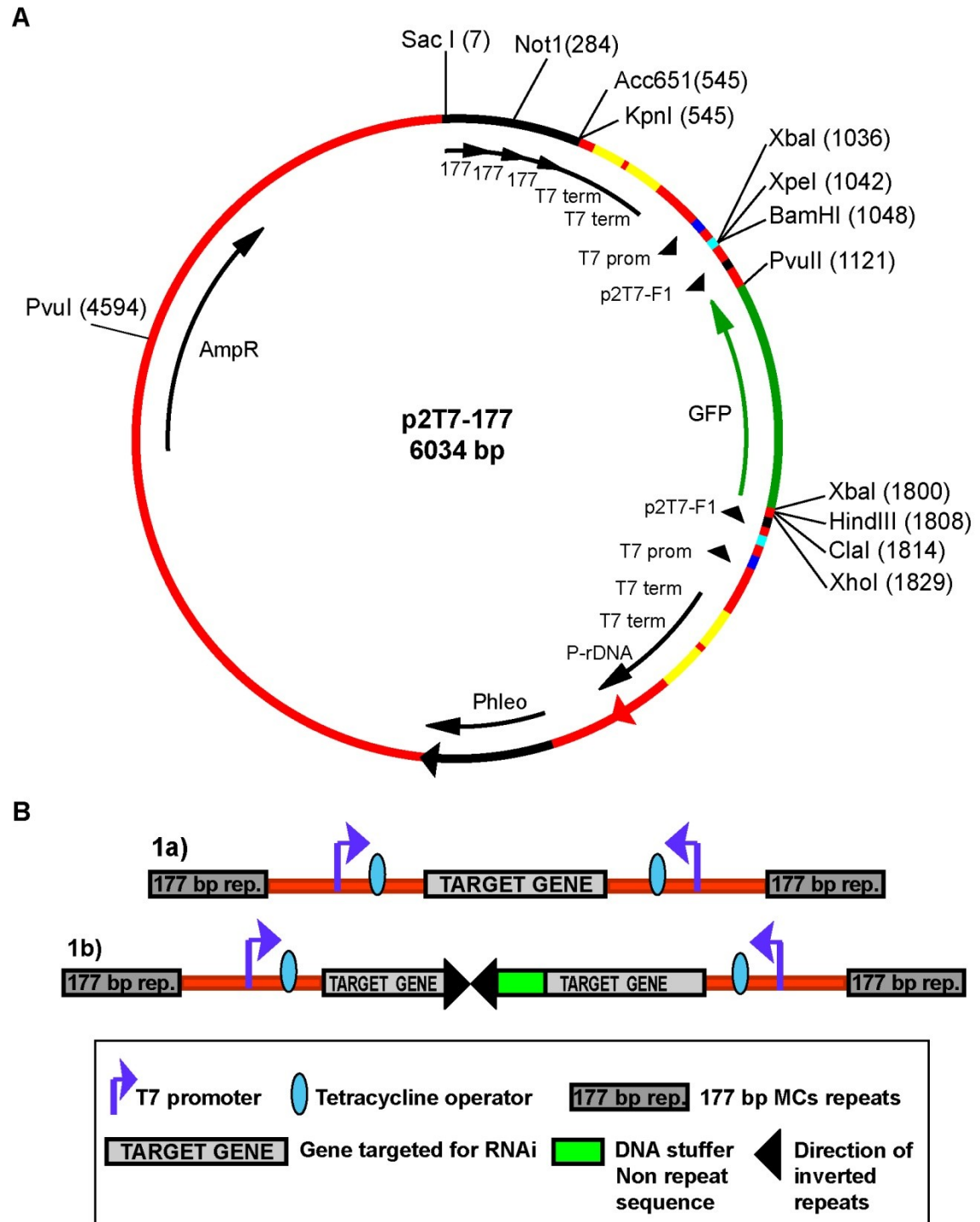


Figure 2.2. A. p2T7-177 RNAi vector (*Wickstead et al., 2002*) used for this thesis work. B. The two RNAi approaches used in this thesis: 1a) Double stranded construct (DS) where a single fragment of the target gene is flanked by two opposite T7 promoters. 1b) Hairpin construct (HP), where two back to back repeats of the target gene are flanked by opposite T7 promoters. The target gene sequences have a different length of 200-300 bp. This not repeat sequence acts as DNA "stuffer" (green box).

1a) DS-N2. Normal double stranded constructs with opposing promoters. The largest amplified fragment, 921bp, was cloned into the p2T7-177 vector, thus forming a normal double stranded RNAi construct with opposite promoters.

1b) HP-N. Hairpin constructs with opposing promoters. The two fragments were cloned into the p2T7-177 RNAi vector, in opposite orientations, so that their difference in size acts as DNA stuffer, hence forming a hairpin RNAi construct with opposite promoters.

2. Targeting the 204 bp repeats common to Tb11.57.0008 and Tb11.47.0036, which are probably parts of the same ORF and have a 97.5% identity over 1025 bp (Tb11.57.0008 5287-6311 bp and Tb11.47.0036 1-1025 bp, Appendix 5, Group 4 seq.). A DNA fragment of 384 bp was amplified from genomic DNA using an anchored step-up PCR approach, which allowed the amplification of one fragment containing almost two complete repeats (384 bp). This fragment was then cloned into the p2T7-177 RNAi vector, rendering the 08-36-Rep RNAi construct (normal double stranded construct).

3. Targeting the repeats common to both Tb11.47.0035 and Tb11.V4.0001, which are also probably parts of the same ORF. The N-terminus of Tb11.V4.0001 has a 100% identity to a repeats region of Tb11.47.0035 (Tb11.47.0035 11254-11815 bp and Tb11.V4.0001 1-562 bp). A 472 bp fragment of this sequence was PCR amplified from genomic DNA and cloned into the p2T7-177 RNAi vector, rendering the 35-01-Rep-RNAi construct (normal double stranded construct). Primer sequences are shown in Table 2.1.

2.2.2. Polymerase Chain Reaction (PCR).

Primers were designed using the OLIGO[®] 4 software (*Rychlik and Rhoads, 1989*), selecting forward and reverse primers of similar melting temperatures (T_m) calculated by the Nearest Neighbour thermodynamic method (*SantaLucia, 1998*). Restriction sites

Table 2.1. Page 1 of 2.

GENE	Pri.	SEQUENCE	Nt	Tm	GC%	Prod.	NAME
CAP5.5-3'-UTR	F	gca ggatcc ACGAAAAACAATGAAGAAGA TG	21	52	43,3	747bp	5.5-UTR-F
	R	gca aagctt TTCCTTACAGTTCATTTCATAC AA	23	53,5	34,4		5.5-UTR-R
CAP5.5V-3'-UTR	F	gca ggatcc CAGAACCCGCAACATACA	18	53,7	55,6	703bp	V-UTR-F
	R	gca aagctt GAAAACACCAACACCAACA C	20	55,2	44,8		V-UTR-R
CAP5.5/ CAP5.5V	F	tta ggatcc AGGAAGTTGAGGAGGTGC	18	56	51,9	934bp	5.5-F
	R	ccg aagctt TAACCTTCCCTGAGACCATTTC G	22	60,3	51,6		5.5-R
CAP5.5 qPCR	F	GCCGGAAGCTGAGGA	15	53,3	66,7	78bp	5.5-RT-F
	R	TTCATCGGCTTGGGTCTCT	19	56,7	52,6		5.5-RT-R
CAP5.5V qPCR	F	GATGAGAAGCCGCACGA	17	55,2	58,8	100bp	V-RT-F
	R	GTTCCGGTTGCGCTTC	16	54,3	62,5		V-RT-REV
Tb11.47.0035 Tb11.57.0008 Targets N-terminus	F	ttag gatccaagctt GTCTCGGAGCAATGGA T	17	56,8	52	765bp	N-08-F
	R-1	cgt tctaga TGCGTATCATCCACGTTA	18	57,3	44		N-08-R-1
	R-2	cgt tctaga CCCTTCCACTCATCATCA	18	57,6	50	875bp	N-08-R-2
Tb11.47.0036 Tb11.57.0008 Targets Repeats	F	tta ctcgag GGATCGTGAAGGTAATGC	18	57,1	48,1	1099bp	REP-08-36-F
	R-1	cgt tctaga TAGGATCAAGGAAGCTGC	18	57,2	50		REP-08-36-R-1
	R-2	cgt tctaga CTTACGACGCTCCTCTTC	18	56,9	53,8	1337bp	REP-08-36-R-2
Tb11.v4.0001 Tb11.47.0035 Targets Repeats	F	tta ggatccaagctt CGATCGTGACTACTTGC A	18	56,7	50	473bp	REP-01-35-F
	R-1	cgt tctaga CTGTGTCCAAAGATAGCAAG	20	57,4	45		REP-01-35-R-1
	R-2	cgt tctaga GTTTTGCTCCCCTTCTTT	18	58,4	44,4	331bp	REP-01-35-R-2
Tb11.v4.0001 Tb11.47.0035 Tb11.47.036 C-Terminus	F	tta ggatccaagctt GAAGTGTGAGCCCCAAG TT	18	56,5	50	614bp	01-35-36-F
	R	cgt tctaga GGTTTCCTCCACAGCTTA	18	57	50		01-35-36-R
Tb11.0035	F	tta actagt CACTGGACAATGACGATG	18	56	50	1176bp	35-F
	R-1	cga tcatg CCAAGAACATTCGATCCT	18	56,6	44,4		35-R-1
	R-2	cga tcatg CGTTAAGCATTTGTCCA	18	56,6	38,9	918bp	35-R2
Tb11.0036	F	tta ggatccaagctt TATTTGTTACCGTTGGG G	18	58,2	44,4	573bp	36-F
	R-1	cgt tctaga AAAAGCACGTTCTCACT	18	56,6	44,4		36-R-1
	R-2	cgt tctaga TGTTCCCTCATGCATTCT	18	57,6	44,4	728bp	36-R-2
Tb11.57.0008	F	tta ggatccaagctt GAAAGAATTGCTGAAC TGG	19	57	42,1	256bp	08-F
	R-1	cgt tctaga ATAATTGTCGTGTTTCGTCC	18	57	44,4		08-R-1
	R-2	cgt tctaga TGGCAACAGCGTATTCTA	19	57	42,1	147bp	08-R-2

Table 2.1. Page 1 of 2. List of primers used in this thesis work. Restriction sites and the 2-3 random nucleotides added at the primers' 5'-ends to improve digestion efficiency are in lower case. Restriction sites are also shown in blue font. Pri (primers), R (reverse), F (forward), Nt (nucleotides), Tm (melting temperature), Prod. (product).

ggatcc=BamHI; **aagctt**=HindIII; **tctaga**=XbaI; **actagt**=SpeI; **atcgat**=ClaI; **ctcgag**=XhoI.

Table 2.1. Page 2 of 2.

Tb11.02.2240	F	tta ggatccaagctt GTGTGGAGTTCTGCATC AT	19	57	47,4	1164bp	22-F
	R-1	cg tctaga CTATTTTACGCCCAACTC	19	57,1	42,1		22-R-1
	R-2	cg tctaga CAGATTCCAACCGACTAAC	19	56,7	47,4	1272bp	22-R-2
Tb09.160.5550	F	tta ggatccaagctt GCTGATATTTCCGACCC	17	56,9	52,9	1053bp	5-F
	R-1	cg tctaga CACGAGTTCTTCAGGCA	17	56,1	52,9		5-R-1
	R-2	cg tctaga CAACAGACGGGAGGTAAT	18	56,5	50	1183bp	5-R-2
Tb11.01.6870	F	tta ggatccaagctt TTGAGGAGGAGGCGTA T	17	57	52,9	630bp	11.68-F
	R-1	cg tctaga TTGTGTGGTTATGGCGTA	18	57,3	52,9		11.68-R-1
	R-2	cg tctaga CATAACCGACACCTCATTC	19	57,2	47,4	848bp	11.68-R-2
Tb927.5.2190	F	tta ggatccaagctt GGTGGTGTGAGGTTG TT	18	56,7	50	654bp	5-F
	R-1	cg tctaga GAACATCACACGCCTCTC	18	57,3	55,6		5-R-1
	R-2	cg tctaga GCATCTTGAAACCTTATCAC	20	56,4	40	969bp	5-R-2

Table 2.1. Page 2 of 2. List of primers used in this thesis work.

were added to the 5' end of each primer preceded by 3 random nucleotides to increase effective digestion. PCR amplifications were performed in 50 microlitres (µl) ddH₂O containing 0.2 millimoles (mM) of each primer, 0.2 mM of each deoxyribonucleoside triphosphate (dNTPs), 2 units (U) of Phusion High-Fidelity DNA polymerase, New England BioLabs (2 U.µl⁻¹), 10 µl of 5xHF-buffer (7.5 mM.µl⁻¹ MgCl₂) and 50-250 nanograms (ng) of trypanosome genomic DNA. Amplification was carried out using a Perkin Elmer GeneAmp 2400 PCR machine. The thermal cycling conditions were:

30 cycles	Initial denaturation step: 5 minutes (min) at 94 Centigrade degrees (°C).
	Denaturation step: 30 seconds (s) at 94°C.
	Annealing step: 30 s at 55-57°C (T _m).
	Elongation step: 30 s at 72°C.
	Polishing step: 5 min at 72°C.

2.2.3. Anchored PCR.

The use of long double-stranded RNAs (dsRNAs) for RNAi is thought to increase silencing specificity, reducing the probability of off-target effects (*Echeverri and Perrimon, 2006*). Thus, to make RNAi constructs that target the 204 bp tandem repeats of *T. brucei* group 4 calpain-like sequences Tb11.47.0036 and Tb11.57.0008, I used a step-up anchored PCR approach consisting in 5 cycles of normal PCR thermal conditions followed by 25 cycles with a higher annealing temperature. The higher annealing temperature increases the annealing at the primers ends, thus producing fragments of different sizes containing one or more tandem repeats. A 384 bp fragment was obtained by this PCR approach, containing almost two complete 204 bp repeats. Thermal conditions are described below.

Initial denaturation step: 5 min at 94 °C.	
5 cycles	Denaturation step: 10 s at 94°C.
	Annealing step: 10 s at 55°C.
	Elongation step: 1 min at 72°C.
25 cycles	Denaturation step: 10 s at 94°C.
	Annealing step: 10 s 62°C.
	Elongation step: 1 min at 72°C.
Polishing step: 5 min at 72°C.	

2.2.4. Restriction digests.

Restriction enzymes used to build RNAi constructs (Table 2.2) were from New England BioLabs or Roche diagnostics. Digestion reactions were performed using 10 U of enzyme to cut 1 microgram (µg) of DNA for 1 hour (h) at 37°C.

Table 2.2.

Restriction Enzymes	RNAi Construct
BamHI / HindIII	CAP5.5
	CAP5.5V
	CAP5.5/CAP5.5V
BamHI / XbaI / HindIII	HP-N- 35-08
	HP-Rep-01-35
	HP-36
	HP-08
	HP-22
	HP-5
	HP-9
Hind III / XbaI	DS-N2 (35-08)
	DS-Rep-01-35
	DS-36
	DS-08
	DS-22
	DS-5
	DS-9
Spe I / Cla I	DS-35
Xho I / Xba I	HP-Rep-36-08

Table 2.2. Restriction enzymes used to build the RNAi constructs designed for this thesis work.

2.2.5. Purification of DNA fragments.

PCR products and DNA fragments from restriction enzyme digestions were resolved by gel electrophoresis on 0.8-1% weight/volume (w/v) agarose gels containing 0.1 $\mu\text{g}.\text{ml}^{-1}$ of ethidium bromide. PCR products and DNA fragments excised from agarose gels were purified using the Wizard SV Gel and PCR Clean-Up System (Promega) or the QIAquick PCR Purification Kit (Qiagen) in accordance with the manufacturers' instructions.

2.2.6. DNA ligations.

For functional RNAi of *T. brucei* calpain-like proteins, PCR amplified fragments of selected sequences were ligated into the p2T7-177 RNAi vector (Wickstead *et al.*, 2002). Vector and DNA fragments were digested with the appropriate restriction enzymes (Table 2.2). The digested vector was dephosphorilated using Alkaline Phosphatase (Roche) to prevent self-ligation (2 µg of digested and purified plasmid was treated with 20 U of Alkaline Phosphatase, for 1h at 37°C) and purified. Ligation was performed using 50 ng of backbone plasmid and an adequate amount of DNA fragment to maintain a 1:3 ratio of plasmid versus DNA fragment molecules. For a three way ligation, a 1:3:3 ratio was used. DNA ligations were carried out by adding 1 U of T4 DNA ligase (Roche Diagnostics) to a 10 µl final volume reaction.

2.2.7. Bacterial strains and molecular cloning.

Ligated plasmids were transformed into *E. coli* XL1-blue strain and Sure[®] Competent cells (Stratagene). Bacteria were transformed by incubation of 100 µl of competent cells with recombinant plasmids on ice for 30 min. Cells were then subjected to heat shock for 30 s at 42°C and then placed back on ice for 2 min and spread onto agar plates containing 100 mg.ml⁻¹ ampicillin and incubated overnight at 37°C. Single colonies from the agar plates were used to inoculate 3 ml of Luria Broth (LB). Liquid cultures were grown in a shaking incubator overnight at 37°C and 200 revolutions-per-minute (rpm). Large-scale liquid cultures of bacteria were made by a 1:1000 dilution of the overnight 3 ml culture in 50 ml of LB in an autoclaved conical flask and incubated overnight at 37 °C and 200 rpm.

2.2.8. Plasmid DNA isolation.

Small-scale plasmid DNA isolation was carried out from 3 ml overnight bacterial cultures using the QIAprep[®] spin miniprep kit (Qiagen). Large-scale plasmid DNA

isolation was carried out using the Qiagen Midiprep© kit (Qiagen). All kits were used according to manufacturer's instructions. The final pellet was re-suspended in elution buffer (EB: 10 mM Tris-HCl buffer, pH 8.5).

2.2.9. Sequencing of RNAi constructs.

All RNAi plasmid constructs were verified by sequencing at the Geneservice DNA Sequencing Facility, Department of Biochemistry, University of Oxford, using an ABI 3730 DNA Analyser from Applied Biosystems™

2.2.10. Expression analysis by real-time-PCR (RT-PCR).

RNA was isolated from bloodstream and procyclic 427 trypomastigotes using a High Pure RNA isolation kit (Roche) and cDNA was synthesised using an Omniscript reverse transcriptase kit (Qiagen), 2 µg RNA template and oligo dT primers (final concentration 1 µM) in a final volume of 20 µl. The accumulation of CAP5.5 and CAP5.5V transcripts relative to that of either γ -tubulin or QM10 (a 60S ribosomal protein) was then determined by quantitative RT-PCR as described previously (*Dawe et al., 2005*). For these PCR reactions 1 µl of cDNA product was used as the starting template in a final volume of 25 µl, which also included 12.5 µl of Brilliant®SYBRGreen QPCR master mix (Stratagene), gene-specific PCR primers (5 µM), and the Expand HighFidelityPlus polymerase (Roche). Two reverse transcriptase reactions were carried out per RNA sample and cDNA samples were analysed in duplicate for the PCR reactions. Amplification was performed using the Mx3000P Real-time PCR System (Stratagene) under the following cycling conditions: 94°C for 10 min and then 35 cycles of 94°C for 15 s, 57°C for 30 s and 72°C for 15 s. In each cycle the fluorescence was recorded at the end of the annealing phase. The Ct value for each reaction was measured, and the absolute DNA amount and the relative ratios from the means of four individual values $\pm$ standard error (SE) were calculated. Probes used for amplification

of cDNA corresponding to γ -tubulin or QM10 transcripts were γ -tubulin: CAAATATACAGTCTTCTGTGC (forward-primer), TCAAAAGTTTCGTATATATC (reverse-primer) and QM10: GCGTGCCAACAAGGAAT (forward-primer), GCGAAGTACGTGGAACGGA (reverse-primer), as described previously (McKean *et al.*, 2003). Primer combinations used for the detection of calpain-related protein transcripts (CAP5.5 and CAP5.5V) are shown in Table 2.1.

2.3. Trypanosome cell lines and cell culture.

2.3.1. Culture of trypanosomes.

Trypanosoma brucei procyclic form, strains 427 and 29-13 (Wirtz *et al.*, 1999), were cultured in SDM-79 medium (Brun and Jenni, 1977) supplemented with 10% volume/volume (v/v) heat inactivated foetal calf serum (HIFCS, Gibco). Procyclic trypanosomes were cultured in 25 square-centimetres (cm²) tissue culture flasks in a 28°C incubator, and were maintained by subculture at a cell density consistent with the log-phase of the procyclic trypanosomes growth curve (10^5 - 2×10^7 cells.ml⁻¹). *T. brucei* bloodstream form, strains 427 and 90-13 (Wirtz *et al.*, 1999), were cultured in HMI-9 medium (Hirumi and Hirumi, 1989) supplemented with 15% v/v HIFCS, in 25 cm² tissue culture flasks with vented caps, in a 37°C incubator with 5% carbon dioxide (CO₂). Bloodstream trypanosomes were cultured to a maximum density of 10^6 cells.ml⁻¹. Cell density was determined using either a Neubauer haemocytometer or a CASY®1 cell counter (Schärfe systems GmbH).

2.3.2. Transformation of trypanosomes.

For stable RNAi transfections of procyclic *T. brucei*, 29-13 cells co-expressing T7 RNA polymerase (RNAP) and Tetracycline Repressor (TetR) (Wirtz *et al.*, 1999) were harvested from log phase cultures (3 - 8×10^6 cells.ml⁻¹) with a gentle spin of 500 times

gravity (g) in 15 ml polypropylene conical tubes (BD Falcon™), washed once and resuspended at 6×10^7 cells.ml⁻¹ in Zimmerman Post-Fusion medium (ZPFM, 132 mM NaCl, 8 mM KCl, 8 mM Na₂HPO₄, 1.5 mM Mg(CH₃COO)₂, 90 µM Ca(CH₃COO)₂), with 10-20 µg of linearised plasmid (*Bellofatto and Cross, 1989; Beverley and Clayton, 1993*). Cells were electroporated twice with three consecutive 100 milliseconds (ms) square pulses of 1700 volts (V) at 200 ms intervals in an Electro Square Porator ECM830, BTX. Cells were then transferred to 10 ml SDM-79 medium and incubated overnight. After 24 h of transfection, cells were diluted to 1:20 into SDM-79 with 3 µg.ml⁻¹ phleomycin, for selection of stable transfectants.

For stable RNAi transfections of bloodstream *T. brucei*, 90.13 cells were harvested from log-phase cultures ($1-5 \times 10^5$ cells.ml⁻¹) and resuspended at 5×10^7 cells.ml⁻¹ in ZPFM with 20 µg of linearised plasmid. Cells were electroporated once with three consecutive 100 ms square pulses of 1700 V at 200 ms intervals (Electro Square Porator ECM830, BTX), diluted in 25 ml of HMI-9 medium and incubated at 37°C. After 12 hours of transfection, cells were diluted to 1:20 into HMI-9 with 3 µg.ml⁻¹ phleomycin and aliquoted in 1 ml into 24 well-plates.

2.3.3. Cloning by limiting dilution of procyclic trypanosomes.

Conditioned medium was prepared as follows: procyclic form cells were grown as described above to a density of approximately 5×10^7 cells.ml⁻¹ (stationary-phase). This culture was pelleted and the supernatant was filter sterilised through a 0.22 µm filter (Millipore). Transfected cultures were diluted to 1 cell.ml⁻¹ in SDM-79 medium with a 15% v/v conditioned medium. Cloning by limiting dilution was achieved by aliquoting this cell culture in 200 µl into 96 well plates (0.2 cells per well) in the presence of a selective drug (3 µg.ml⁻¹ phleomycin).

2.3.4. RNAi induction.

RNAi was induced by the addition of doxycycline to a final concentration of $1 \mu\text{g}.\text{ml}^{-1}$; all cultures were maintained in the absence of selectable markers for 48h prior to the start of an RNAi induction. After RNAi induction, procyclic form not-induced and induced cells were passaged to $10^6 \text{cells}.\text{ml}^{-1}$ every 48h by addition of cells to SDM-79 medium pre-warmed at 28°C . In order to determine the growth rate, cell counts were taken before and after dilution and a cumulative growth curve was inferred using the growth rate data.

2.4. Protein biochemistry.

2.4.1. Preparation of cell extracts.

For whole cell extracts, 2×10^7 cells from mid-log phase trypanosome cultures were centrifuged at 500 g for 10 min in 1.5 ml Eppendorf tubes. The supernatants were removed and the pellets immediately solubilised in boiling Laemmli sample buffer (2% w/v sodium-dodecyl-sulphate (SDS), 0.002% w/v bromophenol blue, 10% v/v glycerol, 50 mM Tris-HCl (pH 6.8) and 400 mM β -Mercaptoethanol) to obtain a final concentration of 2.5×10^5 cell-equivalents per μl . The samples were boiled for 5 min and stored at -80°C .

2.4.2. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE).

Protein samples from $2.5 \times 10^5 \text{cells}.\mu\text{l}^{-1}$ cell equivalents (prepared as explained above) were separated by SDS-PAGE in mini gels 0.75 millimetres (mm) thick and 6 cm long (BioRad). These gels contained either a 10% or a 8% v/v acrylamide resolving solution (10% or 8% v/v acrylamide from a 30% acrylamide stock (BioRad), 150 mM Tris-HCl (pH 8.8), 0.1% v/v SDS, 0.1% w/v of ammonium persulphate (APS) and 0.04% v/v of Tetramethylethylenediamine. (TEMED, Sigma Aldrich) in ddH₂O to a final volume of 12

ml) In all cases the gel also contained a 5% v/v acrylamide stacking gel (5% v/v acrylamide from a 30% acrylamide stock (BioRad), 125 mM Tris-HCl (pH 6.8), 0.1% v/v SDS, 0.1% w/v of APS and 0.04% v/v of TEMED). Gels were run in running buffer (25 mM Tris-HCl, 250 mM glycine, 0.1% w/v SDS, pH 8.3) at a constant current of 20 milliamperes (mA) per gel for ~1h. Visualisation of proteins on SDS-PAGE gels was carried out using the Coomassie blue stain (2.5 grams (g) Coomassie brilliant blue R250 in 45% methanol and 10% acetic acid). Gels were incubated for 1h in Coomassie blue stain and then incubated twice for 15 min in a destaining solution (45% methanol and 7% acetic acid).

2.4.3 Western blotting and immunodetection.

Proteins samples run on SDS-polyacrylamide gels were transferred onto nitrocellulose membranes for 14h with a constant current of 120 mA, in transfer buffer (50 mM Tris-HCl, 380 mM glycine, 0.1% w/v SDS and 20% methanol). After transfer, membranes were stained with Ponceau-S stain (0.1% w/v Ponceau S in 5% acetic acid v/v) for 5 min, to assess protein loading and transfer. Membranes were washed twice in Tris-Buffered Saline Tween-20 (TBST: 140 Mm NaCl, 10 mM Tris-HCl (Ph 7,4), Tween-20 0,05% v/v) and blocked with 5% w/v skimmed milk powder in TBST for 1 h. After blocking, the membranes were incubated with the appropriated dilution of primary antibody in TBST with 1% w/v skimmed milk for 1h and then washed 5 times for 5 min in TBST. The membranes were then incubated for 1h with a 1:10⁴ dilution of goat anti-mouse horseradish peroxidase-conjugated secondary antibody (Sigma Aldrich) in 1% skimmed milk in TBST. The membrane was again washed 5 times for 5 min in TBST. Final detection was performed by adding enhanced chemiluminescence reagent (ECL) according to manufacturer's instructions (Perkin Elmer). The membrane was exposed to Kodak BIOMAX MR film (Sigma Aldrich) for an appropriate time and developed in a Kodak M35-M X-OMAT processor.

2.5. Light microscopy.

2.5.1. Immunofluorescence.

For cell imaging, live cells were harvested by centrifugation at 500 *g* for 5 min in a microfuge. The pellet was washed once with Phosphate-Buffered-Saline (PBS) and then resuspended to approximately 10^7 cells.ml⁻¹. Cells were then allowed to settle and stick onto poly-lysine coated glass slides and fixed with 3.7% para-formaldehyde for 20 min in a humid chamber. After rinsing with PBS, the cells were permeabilised with ice-cold methanol for 2-5 min at -20°C and washed with PBS again. For cytoskeletons, 1% Nonidet (NP-40) was added to cells settled onto glass slides in the same way and incubated for 30 s. Following detergent extraction cells were fixed in ice-cold methanol for 30 min at -20°C.

Whole cells and detergent extracted cytoskeletons were washed three times in PBS for 5 min and pre-blocked with PBS containing 0.1% Tween-20, 1% Bovine-Serum-Albumin (BSA) for 30 min. Slides were incubated in the primary antibody (Ab) diluted in PBS for 45 min in a humid chamber and then washed 4 times with PBS to remove unbound antibody. The secondary Ab was added and the cells incubated for a further 45 min before being washed 4 times with PBS. Cells were then mounted with Vectorshield containing the DNA-binding dye 4,6-diamidino-2-phenylindole (DAPI) and sealed with nail varnish. Primary and secondary antibodies used in this thesis are shown in Table 2.3.

Table 2.3.

<u>Primary antibodies</u>	<u>Dilution</u>
Anti-CAP5.5 mouse IgG mAb	1:5 in 1% BSA-PBS
TAT-1 (mouse IgG), the anti-tubulin mAb	1:10 in 1% BSA-PBS
BB2 (mouse IgG), the anti-TY tag mAb	Neat and 1:20 in 1% BSA-PBS
L8C4 (mouse IgG ₁), a mAb against PFR2	1:50 in 1% BSA-PBS
Rod-1 (mouse IgM), the anti-PFR mAb,	1:10 in 1% BSA-PBS
L6B3 (mouse IgM), the anti-FAZ1 mAb,	1:2 in 1% BSA-PBS
CD-10 ⁴ (mouse IgG ₃)	Neat
Dot1 (mouse IgM)	Neat
Anti-CAP15 rabbit polyclonal Ab.	1:500 in 1% Triton X-100, 0.1% BSA-PBS
Anti-CAP17 rabbit polyclonal Ab.	1:500 in 1% Triton X-100, 0.1% BSA-PBS

<u>Secondary antibodies</u>	<u>Dilution</u>
Anti-mouse IgG FITC conjugate F5887	1:2400 in 1% BSA-PBS
Anti-mouse IgM Rhodamine AP128R	1:10 in 1% BSA-PBS
Anti-mouse IgG TRITC conjugate T-7657	1:2400 in 1% BSA-PBS
DAKO anti-mouse Igs FITC conjugate	1:50 in 1% BSA-PBS

⁴ CD-10 (mouse IgG₃) and Dot1 (mouse IgM) mAbs label the FAZ and recognise unidentified high molecular weight proteins by Western blotting (Woods *et al.*, 1989).

2.5.2. Epifluorescence microscopy.

For accurate counts of cells in different positions of the cell-cycle (counts of the number of nuclei and kinetoplasts), cells were fixed in culture medium by the addition of paraformaldehyde (3.7% v/v final concentration), spread onto slides, and air-dried prior to processing. Nuclear and mitochondrial DNA were stained using DAPI and cells were imaged using a Zeiss Axioplan 2 microscope.

2.5.3. Videomicroscopy.

Not-induced and induced cells were placed onto microscope slide wells in 10 μ m aliquots. Videos were recorded for 1-5 min at 10x and 35x magnification using phase-contrast illumination.

2.6. Electron Microscopy (EM).

2.6.1. Preparation of samples for thin-section Transmission Electron Microscopy (TEM).

For TEM, cells were initially fixed by adding 1 ml of 25% v/v glutaraldehyde to 9 ml of cells in culture medium at ambient temperature. After ~5 min cells were gently centrifuged (~500 *g* for 3 min) and then transferred into a buffered fixative consisting of 2.5% v/v glutaraldehyde, 2% w/v paraformaldehyde and 0.1% w/v picric acid in 100 mM phosphate (pH 7.0). Cells were pelleted into a 1.5 ml Eppendorf at 14,000 *g* for 3 min. Fixation continued for 2-24 h at 4°C. Samples were post-fixed in 1% w/v osmium-tetroxide in 100 mM phosphate (pH 7.0) for 1.5 h at 4°C, en bloc stained with 2% w/v aqueous uranyl acetate for 2 h at 4°C in the dark, dehydrated and embedded in epoxy resin. Ultrathin (~70 nm thick) sections were double stained with uranyl acetate and lead citrate and examined in a FEI Tecnai 12 electron microscope.

2.6.2. Scanning Electron Microscopy (SEM)

For SEM, trypanosomes were washed in PBS and aliquots of live cells deposited on clean 13 mm coverslips. The cells were allowed for ~10 min to attach to the glass and then were fixed with a mixture of 2.5% glutaraldehyde in 100 mM phosphate (pH 7.0). After fixation the samples were rinsed several times with ddH₂O, dehydrated through series of 5 min in different v/v solutions of ethanol in ddH₂O (from 30% to 90% v/v), followed by 3x5 min in 100% ethanol. Samples were critically point-dried and sputter-coating with gold. Samples were examined in a JOEL JSM 6390 scanning electron microscope.

CHAPTER 3

BIOINFORMATIC ANALYSIS OF KINETOPLASTID CALPAIN-LIKE SEQUENCES

3. BIOINFORMATIC ANALYSIS OF KINETOPLASTID CALPAIN-LIKE SEQUENCES.

3.1. Introduction.

In this chapter, I present the results of the bioinformatic survey of kinetoplastid calpain-like sequences that I carried out using the most recent genome annotation of *T. brucei* (version 4 update), *L. major* and *L. infantum* (Hertz-Fowler *et al.*, 2004) in geneDB (<http://www.genedb.org/trypan>). During my analysis, database searches and sequence comparisons led to the identification of several new trypanosomatid calpain-like proteins that have not been previously described: 10 in *T. brucei*, 10 in *L. major* and 11 in *L. infantum* (Table 3.1). Additionally, I found multiple new calpain-related sequences in *T. cruzi*. However, the nature of the currently available *T. cruzi* genome sequence, obtained from a hybrid of two different strains, and its unfinished state in unassembled contigs, made it very difficult to positively distinguish between real sequences, sequence fragments and gene copies in different alleles. Thus, I decided not to include *T. cruzi* calpain-like sequences in the final dataset.

The data presented here describe the new sequences identification, bioinformatic analysis and comparison to known proteins, related and not related to calpains. The analyses revealed an unusual domain composition and architecture in some of the newly identified sequences, which could not be classified into any of the five groups of kinetoplastid calpain-like proteins previously proposed by Ersfeld and co-workers (Ersfeld *et al.*, 2005). Thus, I modified their classification, categorising the whole repertoire of calpain-like proteins of the three studied trypanosomatid species into five tailored groups. Ultimately, based on the results of these analyses, I selected several *T. brucei* calpain-like proteins for RNAi functional analysis.

Table 3.1.

<i>T. brucei</i> 10 New seq.	<i>L. major</i> 10 New seq.	<i>L. infantum</i> 11 New seq.
Tb927.5.2190	LmjF17.1110	LinJ17.1230
Tb927.5.2220	LmjF20.1185	LinJ20.1220
Tb927.5.2250	LmjF21.0120	LinJ21.0010
Tb927.5.2410	LmjF30.2040	LinJ21.1770
Tb927.7.4070	LmjF32.0970	LinJ30_v3.2040
Tb927.8.8330	LmjF32.2010	LinJ32.1370
Tb11.01.5800	LmjF34.0200	LinJ32_v3.2140
Tb11.01.6870	LmjF34.0280	LinJ34.0220
Tb11.57.0008	LmjF36.0780	LinJ34_v3.0300
Tb11.v4.0001	LmjF36.3180	LinJ36_v3.0840
		LinJ36.3180

Table 3.1. New calpain-like sequences in *T. brucei*, *L. major* and *L. infantum*, found after the release of the version 4 update of the *T. brucei* genome annotation in geneDB. BLASTp searches were performed on the *T. brucei*, *L. major* and *L. infantum* geneDB databases, using different domains of previously described kinetoplastid calpain-like proteins as baits. Analysis of the high scoring hits allowed the identification of several new calpain-like sequences: 10 in *T. brucei*, 10 in *L. major* and 11 in *L. infantum*.

3.2. Analysis: New sequences identification and characterisation.

I started my analysis by performing BLASTp searches for calpain-like proteins in *T. brucei*, *T. cruzi* and *L. major* completed genome sequences (*El-Sayed et al., 2000*; *Berriman et al., 2005*; *Ivens et al., 2005*) and the draft genome sequence of *L. infantum*, available on geneDB, an interface of the Wellcome Trust Sanger Institute (<http://www.genedb.org/>). To identify all putative calpain-like sequence, BLASTp searches on the databases mentioned above were carried out using the sequences of different kinetoplastid calpain-like domains as baits. The sequences that I used as baits in this analysis were:

- CAP5.5 domain I^K: to search for sequences containing the short sequence element identified by Ersfeld and co-workers and defined by them as kinetoplastid calpain-like domain I^K—exclusively present at the N-terminus of domain II containing kinetoplastid calpain-like sequences or as stand along short ORFs— (*Ersfeld et al., 2005*). However, this sequence element cannot be strictly considered a calpain-like domain, as it does not have any motif related to calpains and is not shared by all groups of kinetoplastid calpain-related proteins. Moreover, this short sequence element was most probably acquired through gene fusion after the gene duplication events that gave rise to the different calpain-like proteins in kinetoplastids. Nevertheless, I have included this sequence element or domain I^K in my analysis for consistency with Ersfeld and co-workers classification and nomenclature.
- CAP5.5 domain II: to identify any protein bearing a calpain “catalytic” domain II.
- CAP5.5 domain III, to search for sequences containing a kinetoplastid calpain-related domain III.
- N-terminal domain I^H and C-terminal domain C sequences identified by Ersfeld and co-workers in different kinetoplastid calpain-related proteins (*Ersfeld et al., 2005*), to

check if any of them were present in any other trypanosomatid protein related or unrelated to calpains.

The highest scoring sequences from each BLASTp search were then compared to known calpain-like protein sequences. Multiple alignments were performed using the MAFFT tool, a multiple sequence alignment method based on the fast Fourier transform (Kato *et al.*, 2002), and manually edited on the Jalview platform (Waterhouse *et al.*, 2009). In this manner, I identified 31 new calpain-related sequences: 10 in *T. brucei*, 10 in *L. major* and 11 in *L. infantum* (Table 3.1 and Appendix 3, BLASTp file). To determine the domain composition and architecture of these new trypanosomatid calpain-like sequences, each one was analysed for the presence of previously described kinetoplastid calpain-like sequence motifs as well as for any other domain or motif, not related to calpains, present in the Pfam and InterPro databases.

3.2.1. Identification of trypanosomatid calpain-like domains.

Comparison of each sequence to the alignment of each kinetoplastid calpain-like domain allowed the identification of sequence motifs characteristic of both conventional calpains and trypanosomatid calpain-related proteins:

- **Domain I^K.** Domain I^K identification was based on the presence of its characteristic motifs G-L-LF/Y at the N-terminus, W-A-F-Y-N-D-T at the centre and V-Y-P-x-ET-E towards its C-terminus (Fig. 3.1). Kinetoplastid calpain-like domain I^K was found in 14 out of 31 newly identified calpain-related sequences, 6 in *T. brucei*, 4 in *L. major* and 4 in *L. infantum*:

T. brucei:

Tb927.5.2190, Tb927.5.2220, Tb927.5.2250 (identical gene copies); Tb927.7.4050, Tb927.8.8330 and Tb11.01.6870.

Figure 3.1. Page 1 of 2.

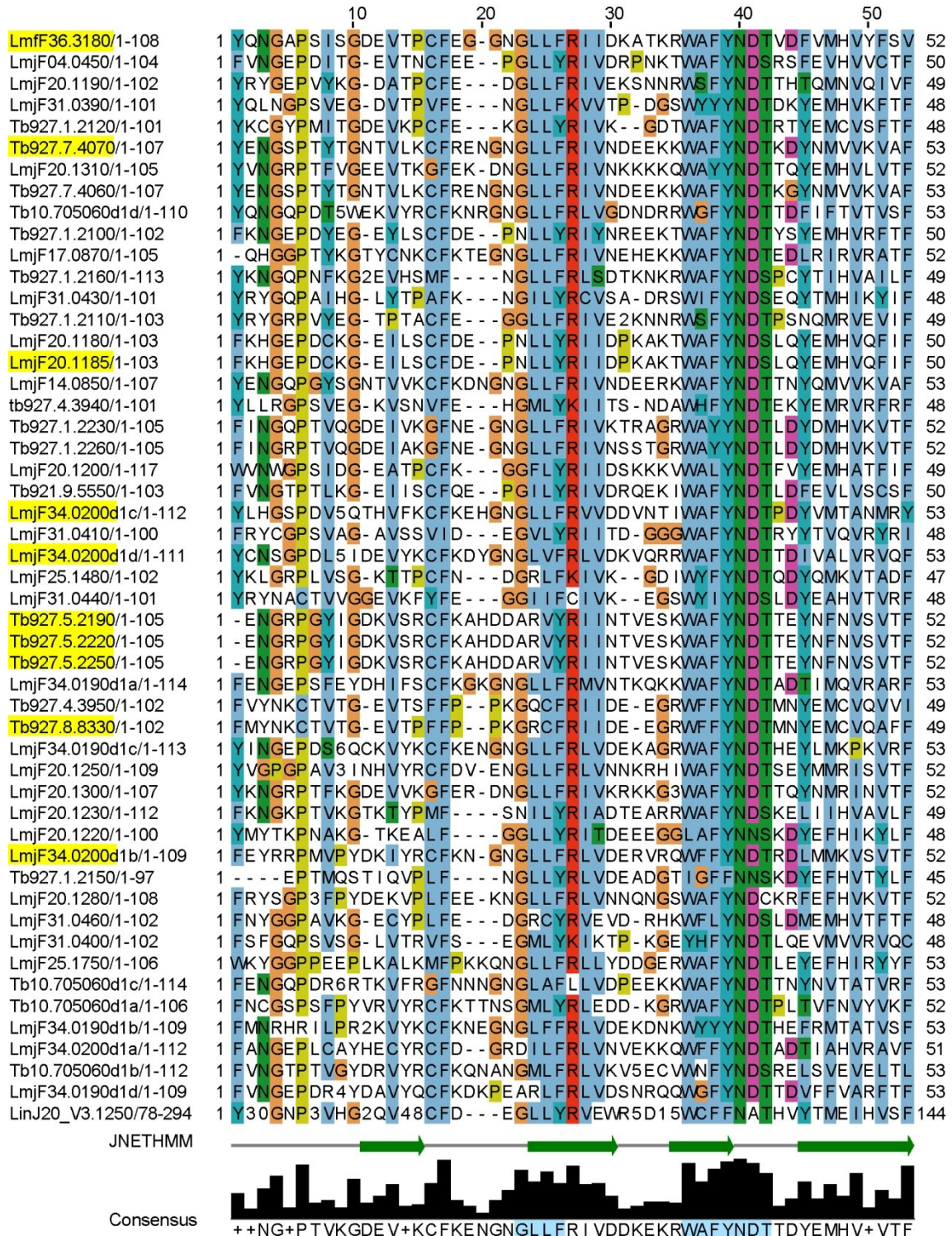


Figure 3.1. Page 1 of 2. Multiple alignment of domain I^K sequences of trypanosomatid calpain-like proteins by the MAFFT method. Colour scheme is according to the ClustalX colour code. Conserved motifs are highlighted in light-blue in the consensus sequence: G-L-L-F/Y at the N-terminus, W-A-F-Y-N-D-T at the centre and V-Y-P-x-ET-E at the C-terminus. The systematic name of newly identified sequences are highlighted in yellow. Secondary structure prediction is by the JNet method, β -strand conformations are indicated by green arrows.

Figure 3.1. Page 2 of 2.

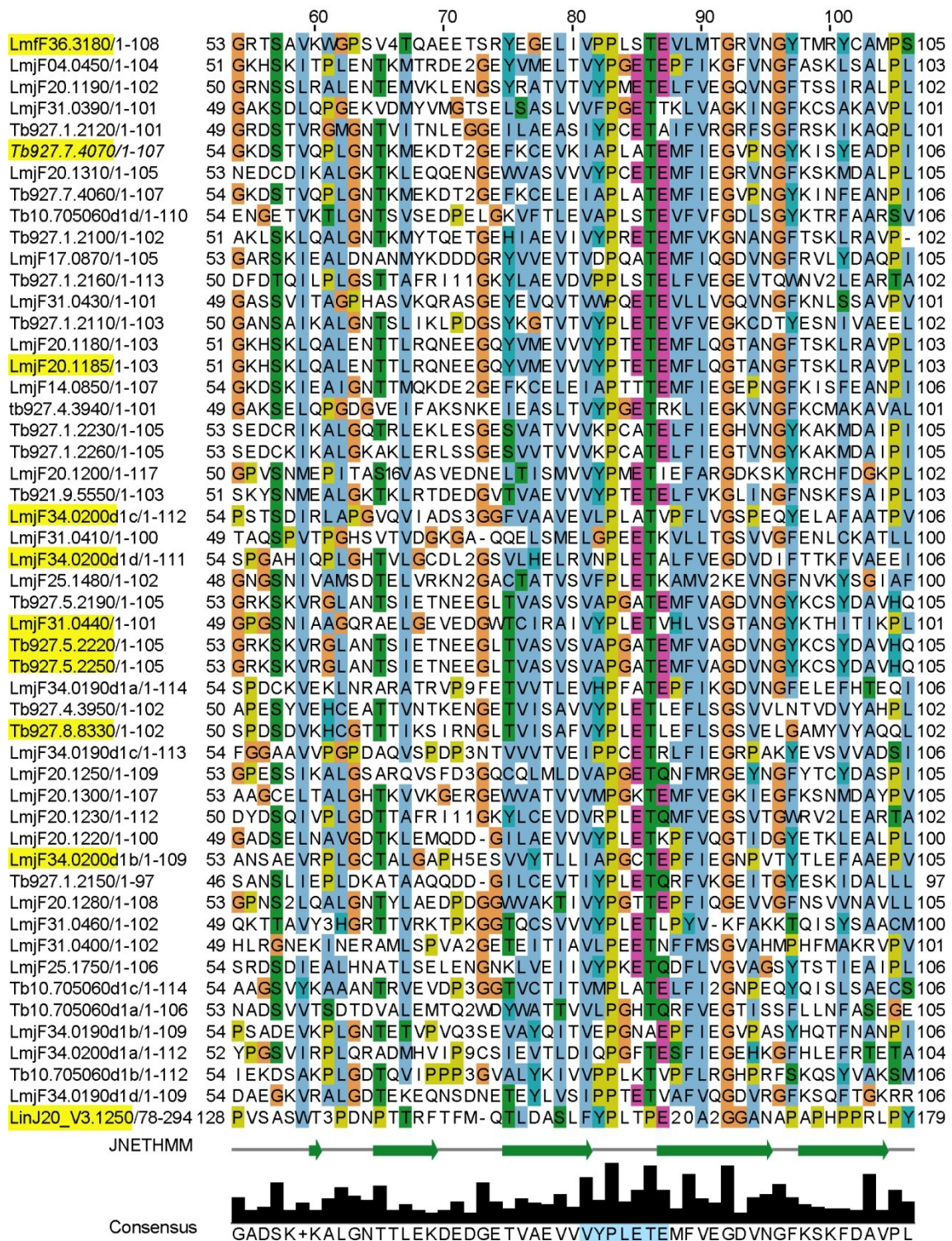


Figure 3.1. Page 2 of 2. Multiple alignment of domain I^K sequences of trypanosomatid calpain-like proteins by the MAFFT method.

L. major:

LmjF20.1185, LmjF32.2010, LmjF34.0200 and LmjF36.3180.

L. infantum:

LinJ20.1210, LinJ32_v3.2140, LinJ34.0220 and LinJ36_v3.3330.

However, of the newly identified 14 domain I^K containing sequences, 8 did not have any other calpain-related domain, and cannot be strictly considered calpain-like proteins (*T. brucei* Tb927.7.4070, Tb927.5.2190, Tb927.5.2220 and Tb927.5.2250; *L. major* LmjF34.0200 and LmjF36.3180; *L. infantum* LinJ34.0220 and LinJ36_v3.3330).

Domain II. Calpain-like domain II sequences were mainly identified by the presence of the K-A-Y-A-K-motif, unique to calpain protease domains, but of unknown function. The newly found domain II sequences were also analysed for the conservation of calpain domain II characteristic motifs: the catalytic triad —C-H-N— with its preceding D and following P, and the presence of the five Ca²⁺-binding residues essential for the correct alignment of the catalytic triad within the active site (Fig. 3.2). Of the 31 newly found calpain-like sequences 20 contain a calpain-like domain II, 5 in *T. brucei*, 7 in *L. major* and 8 in *L. infantum*:

T. brucei:

Tb927.8.8330, Tb11.01.5800, Tb11.01.6870, Tb11.v4.0001 and Tb11.57.0008.

L. major:

LmjF20.1185, LmjF21.0120, LmjF30.2040, LmjF32.2010, LmjF32.0970, LmjF34.0280 and LmjF36.0780.

L. infantum:

LinJ20.1220, LinJ21_v3.0160, LinJ21_v3.0170, LinJ32_v3.1020, LinJ32_v3.2140, LinJ30_v3.2040, LinJ34_v3.0300 and LinJ36_v3.0840.

Figure 3.2. Page 1 of 4.

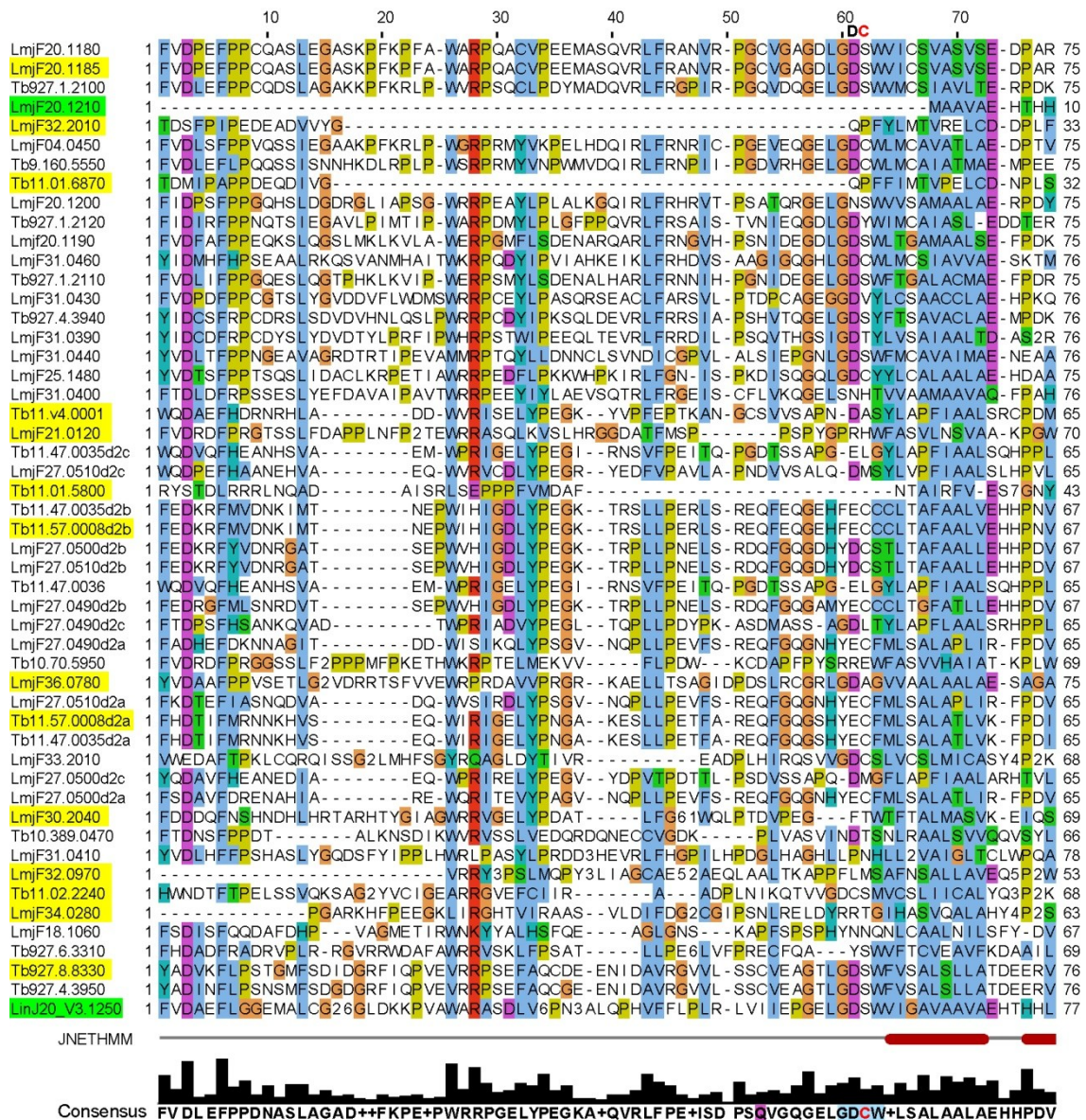


Figure 3.2. Page 1 of 4. Multiple alignment of domain II sequences of trypanosomatid calpain-like proteins by the MAFFT method. The colour scheme is according to the ClustalX colour code. The catalytic triad equivalent residues are in red font in the consensus sequence, motifs characteristic of calpain domain II are highlighted in blue, Ca^{2+} binding residues are highlighted in purple. The systematic names of newly identified calpain-like sequences are highlighted in yellow. The sequence previously classified as group 1 (domains I^K-II-III-C), *L. major* LmjF20.1210, which analysis showed that lacks a domain I^K, is highlighted in green (its ortholog in *L. infantum*, LinJ20_v3.1250, also highlighted in green, does have a domain I^K, although poorly conserved, Fig. 3.1). Secondary structure prediction is by the JNet method, α -helices are indicated by red bars and β -strands by green arrows.

Figure 3.2. Page 2 of 4.

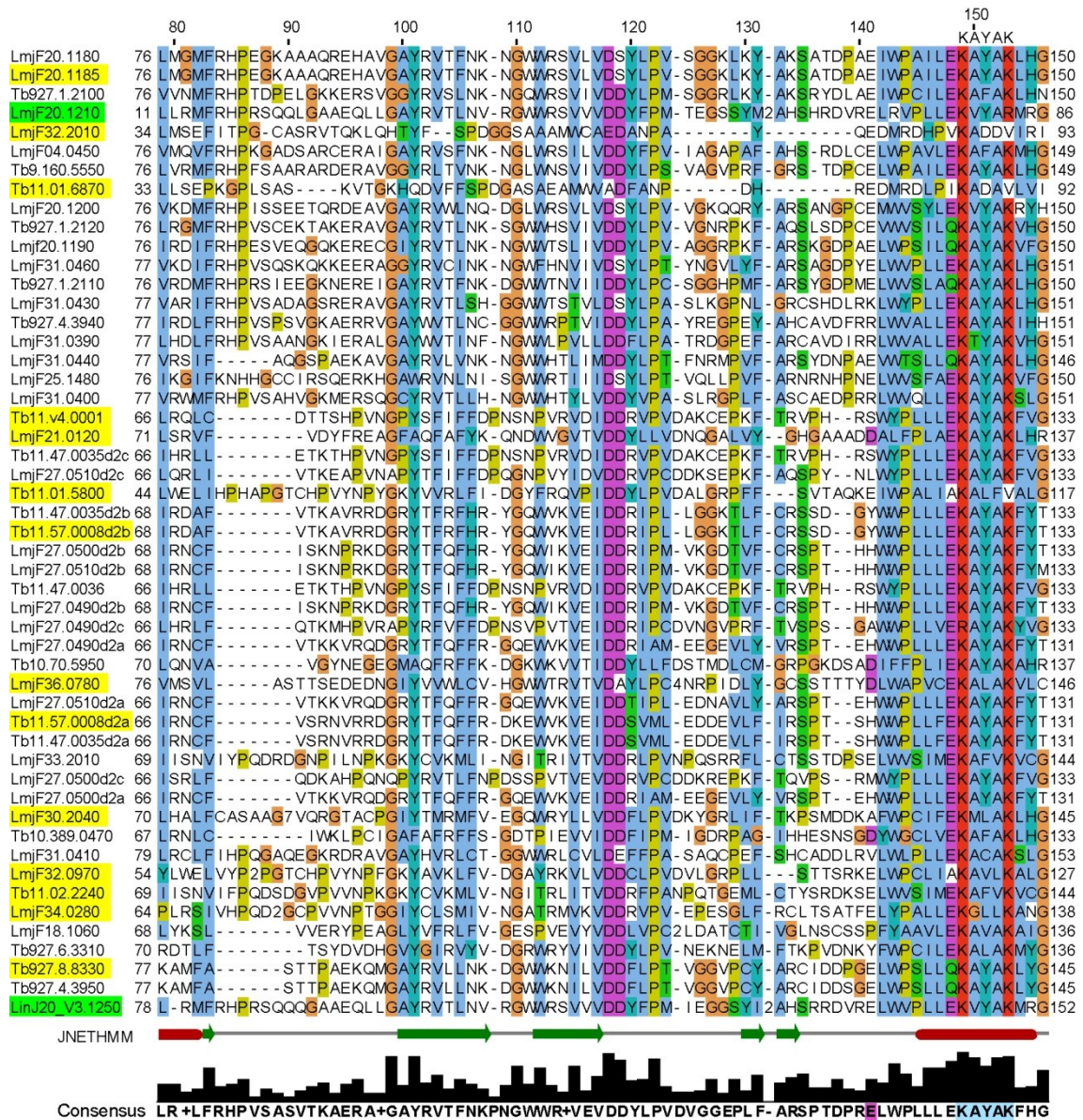


Figure 3.2. Page 2 of 4. Multiple alignment of domain II sequences of trypanosomatid calpain-like proteins by the MAFFT method.

Note the high conservation of the KAYAK motif, highlighted in blue in the consensus sequence, and the partial conservation of one of the amino acid residues required for calpain domain II binding of two Ca²⁺ ions, essential for the correct alignment of the catalytic triad, highlighted in purple in the consensus sequence.

Figure 3.2. Page 3 of 4.

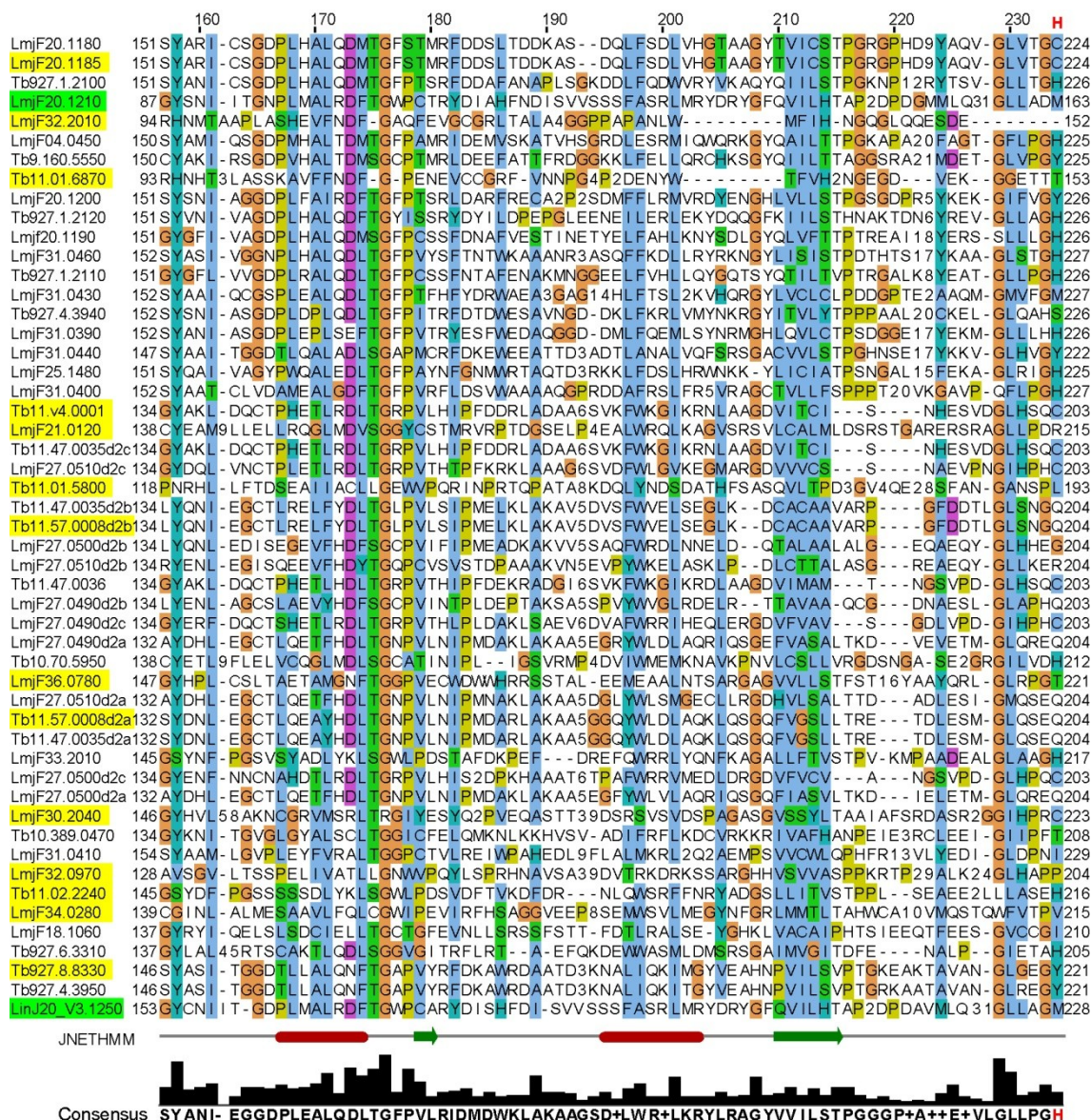


Figure 3.2. Page 3 of 4. Multiple alignment of domain II sequences of trypanosomatid calpain-like proteins by the MAFFT method. Note the conservation of the active site Histidine, in red font in the consensus sequence.

Figure 3.2. Page 4 of 4.

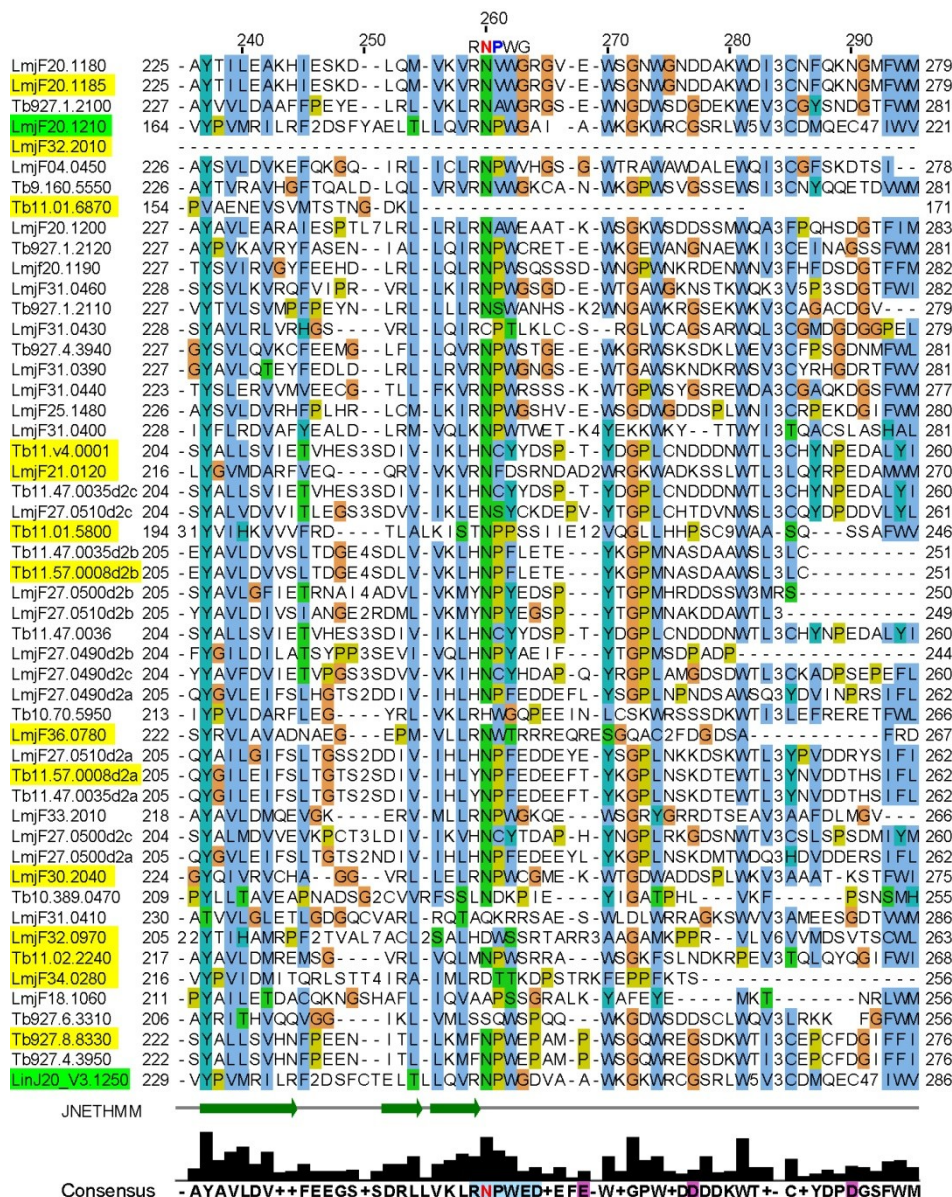


Figure 3.2. Page 4 of 4. Multiple alignment of domain II sequences of trypanosomatid calpain-like proteins by the MAFFT method.

Note the conservation of the active Asparagine residue, in red font in the consensus sequence. Sequence motifs characteristic of calpain domain II are highlighted in blue, Ca²⁺ binding residues are highlighted in purple.

Multiple alignment of kinetoplastid calpain-like domain II sequences also showed that the previously described *L. major* LmjF20.1210, classified by Ersfeld and co-workers as a group 1 (I^K-II-III-C) calpain-like protein (Ersfeld *et al.*, 2005), lacks a domain I^K, containing an N-terminal calpain “catalytic” domain II (Fig. 3.2).

- **Domain III.** Calpain-like domain III sequences identification was carried out by comparison to both conventional (human calpain-1, CAN1) and unconventional calpains (human calpain-6, CAN6 and kinetoplastid calpain-like proteins) domain III sequences (Fig. 3.3). Of the newly identified sequences, 20 contained a calpain-like domain III, 5 in *T. brucei*, 7 in *L. major* and 8 in *L. infantum*:

T. brucei:

Tb927.5.2410, Tb927.8330, Tb11.01.5800, Tb11.v4.0001 and Tb11.57.0008.

L. major:

LmjF17.1110, LmjF20.1185, LmjF21.0120, LmjF30.2040, LmjF32.0970, LmjF34.0280 and LmjF36.0780.

L. infantum:

LinJ17.1230, LinJ20.1220, LinJ21_v3.0160, LinJ21_v3.0170, LinJ30_v3.2040, LinJ32_v3.1020, LinJ34_v3.0300 and LinJ36_v3.0840.

Secondary structure prediction of the aligned kinetoplastid calpain-like domain III sequences by the Jnet algorithm (Cole *et al.*, 2008) on the Jalview platform (Waterhouse *et al.*, 2009) showed a high probability of β -strand conformation for the most conserved motifs of these sequences. This secondary structure, also found in previously described kinetoplastid calpain-like domain III sequences (Ersfeld *et al.*, 2005), is consistent with the presence of an eight-stranded antiparallel β -sandwich in domain III of typical calpains, which is thought to transmit/amplify Ca^{2+} induced conformational changes between domains II and IV. The similar spatial conformation found in trypanosomatid calpain-like domain III sequences could also act as a

Figure 3.3. Page 1 of 3.

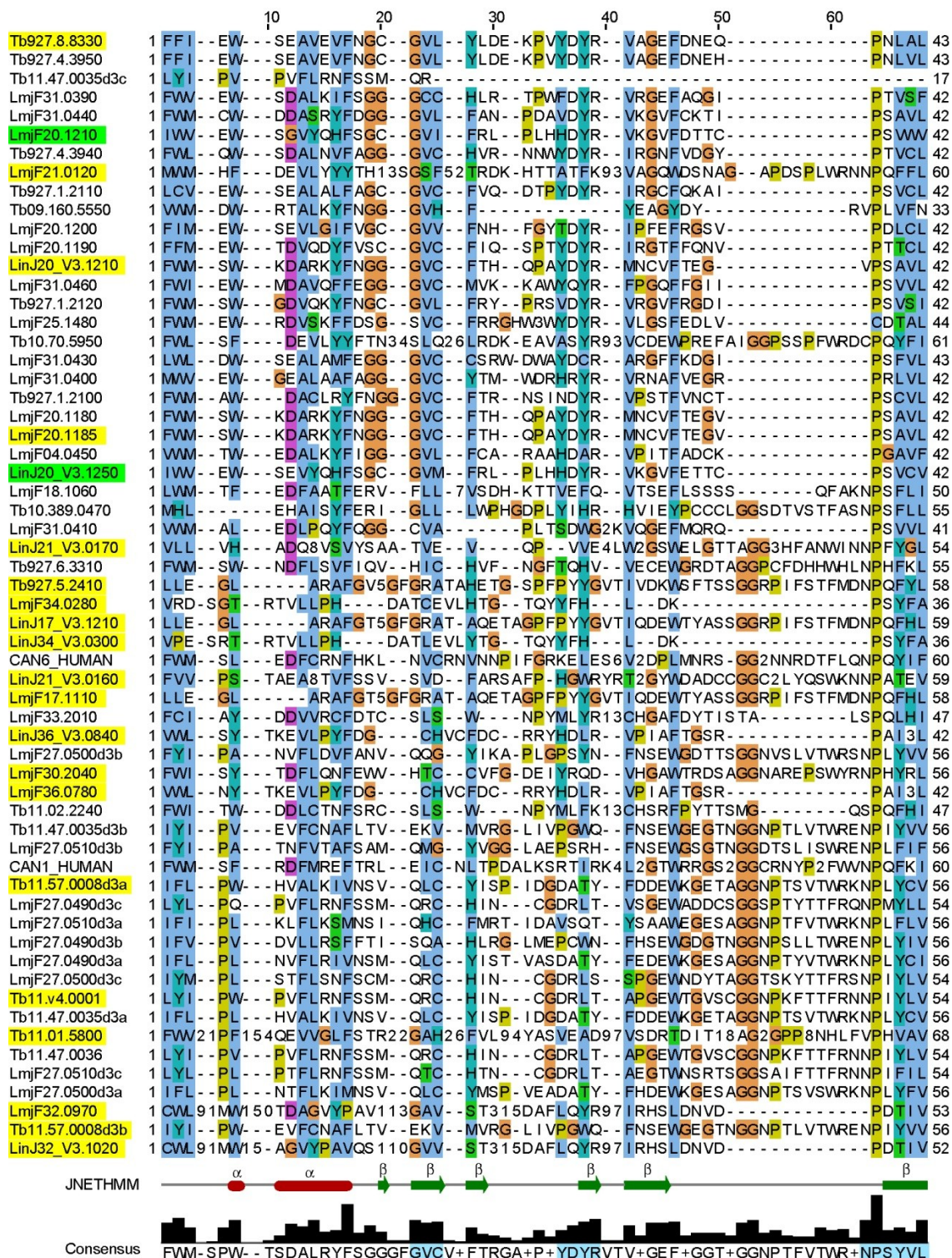


Figure 3.3. Page 2 of 3.

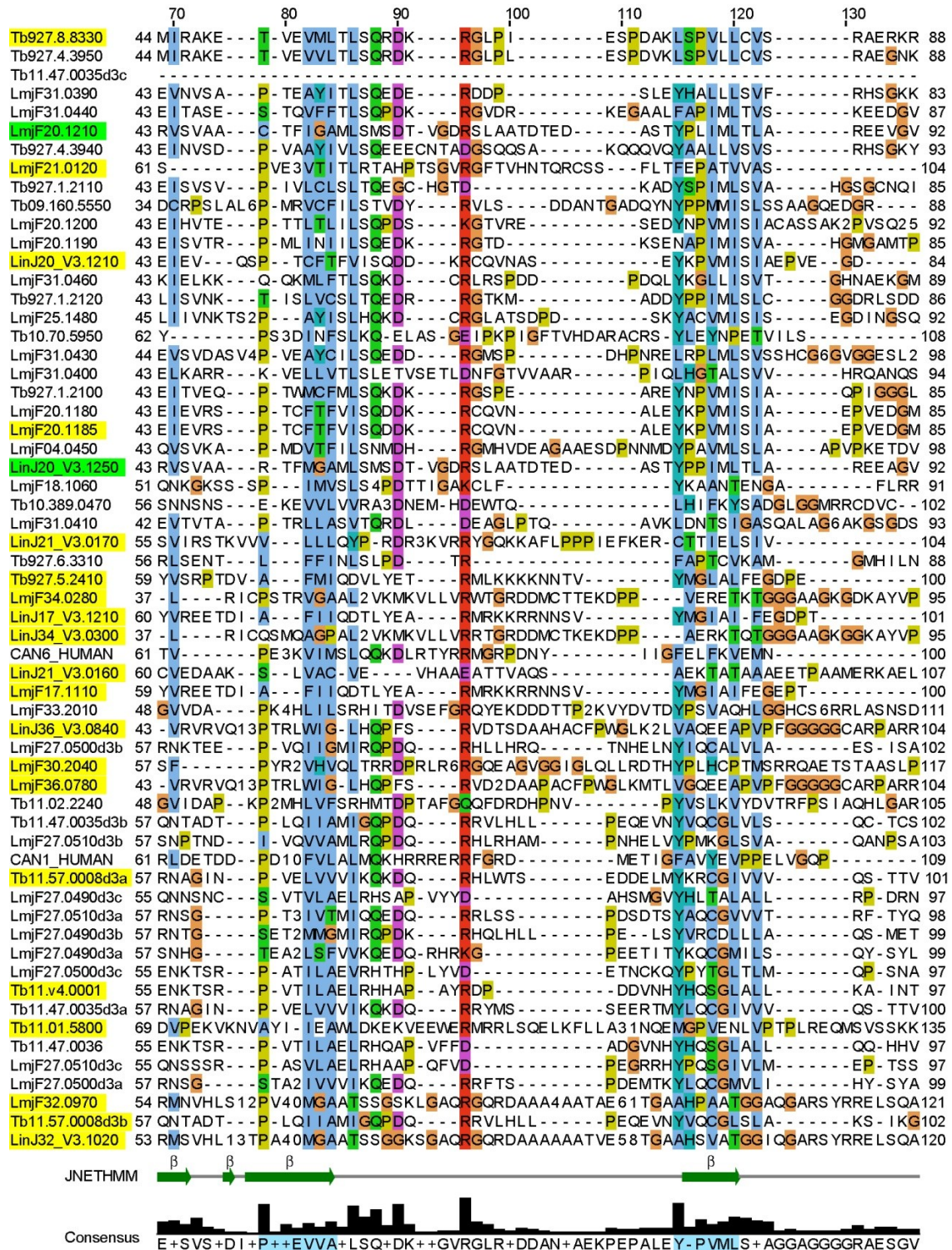


Figure 3.3. Page 2 of 3. Multiple alignment of domain III sequences of trypanosomatid calpain-like proteins by the MAFFT method.

Figure 3.3. Page 3 of 3.

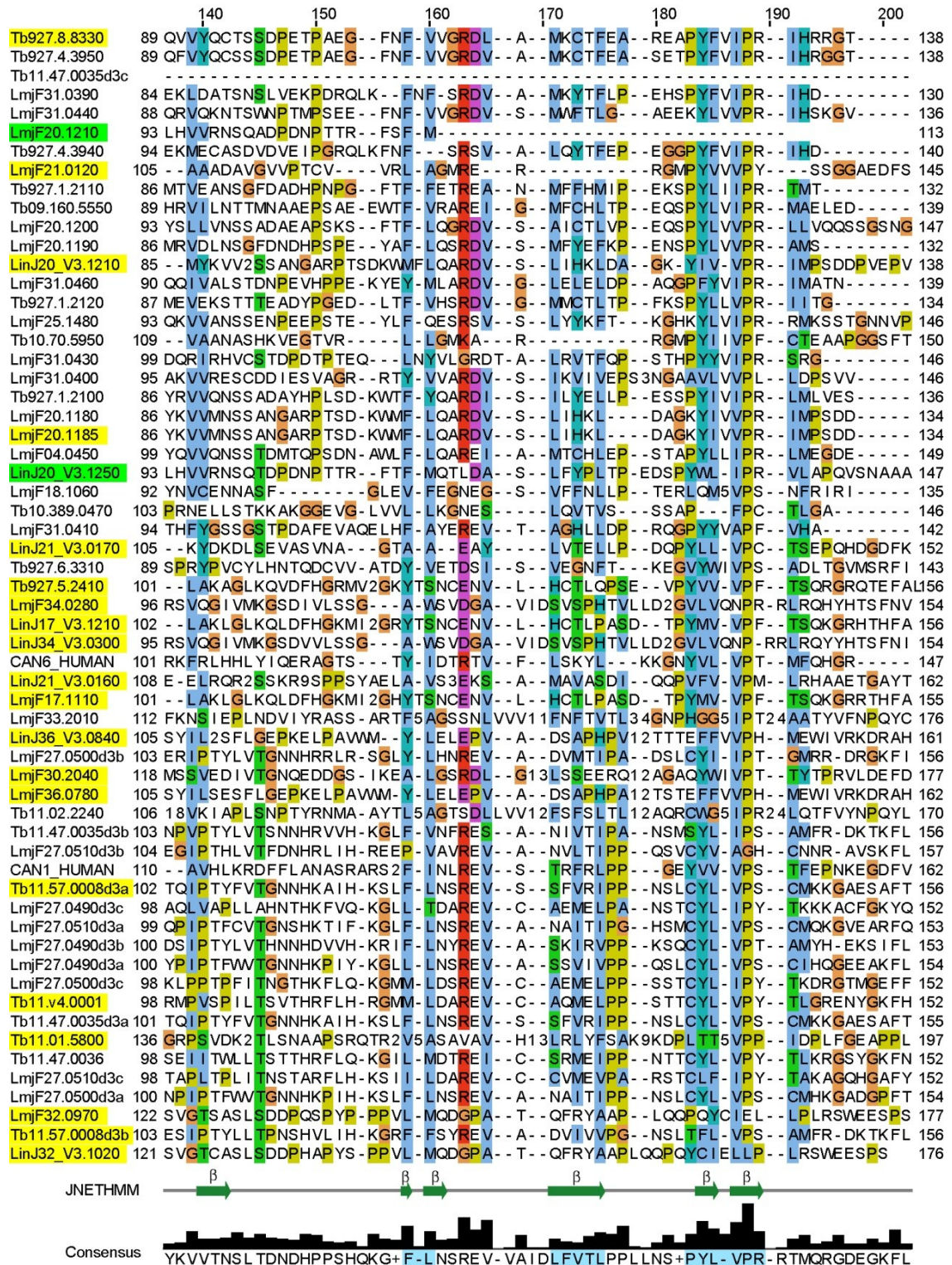


Figure 3.3. Page 3 of 3. Multiple alignment of domain III sequences of trypanosomatid calpain-like proteins by the MAFFT method.

transmitter or amplifier of conformational changes between kinetoplastid calpain-like domains II and IV, although in Ca^{2+} -independent regulatory mechanisms (*Ersfeld et al.*, 2005).

- **Domains I^H and C (C-terminal domain).** These domains are heterogeneous in both intra- and interspecies comparisons, having significant similarities only between putative orthologs. Both domains are exclusively found in sequences containing a calpain-like domain II (*Ersfeld et al.*, 2005). With the exception of domain I^H of Tb927.6.3310 and its newly found orthologs in *L. major* and *L. infantum*, LmjF30.2040 and LinJ30_v3.2040 (Fig. 3.2 and 3.3), which have a significant homology to the N-terminus of cAMP-dependant kinases regulatory subunit (*Ersfeld et al.*, 2005), these domains lack any similarity to known motifs. Thus, BLASTp searches within the genome of the three studied parasites, using known kinetoplastids I^H and C-domains as baits, did not produce hits with significant similarity, except between putative orthologs (Appendix 3, BLASTp).

3.2.2. Identification of domains not related to calpains.

All sequences were also analysed for the presence of any other known motifs present in the Pfam and InterProt databases, which could hint to their potential functions. Unexpectedly, several known domains, not related to calpains, were found in some of the newly and previously described sequences: META domain, kinesin-like motor domain, armadillo tandem repeats and adenylate cyclase-A receptor-type domain. Additionally, some of the newly identified sequences containing calpain-like domains II and III also contained long arrays of 67-68 amino acid tandem repeats, highly similar to those found by Ersfeld and co-workers in several calpain-like proteins and which they classified as groups 4 and 5 (*Ersfeld et al.*, 2005). The analysis of these repeat-containing sequences and their genome loci during this thesis work revealed important

insights about their nature and domain architecture and have been included in a different section of this chapter (Section 3.2.3).

The META-domain.

The META-domain was found in the previously described *L. major* calpain-like sequence, LmjF17.0870 (META2 protein), and in its newly identified orthologs in *T. brucei* —Tb927.5.2190, Tb927.5.2220 and Tb927.5.2250 (identical gene copies) — and *L. infantum* —LinJ17.1000. Multiple alignment of these sequences revealed that all of them contain three degenerate copies of the META-domain (Fig. 3.4). The META domain, although of unknown function, has been found as part of motility and heat shock proteins in bacteria (Yeats *et al.*, 2003). Moreover, *L. major* META1 protein, LmjF17.0890, a META-domain containing protein not related to calpains, has been shown to be essential in *L. major* promastigotes, whereas its overexpression in *L. amazonensis* increased virulence (Uliana *et al.*, 1999). Both *L. major* META domain-containing proteins, META1 and META2, have been shown to be upregulated in the insect infective stage, metacyclic promastigotes (Ramos *et al.*, 2004). *T. brucei*, in addition to its 3 META2 orthologs (100% identical), contains 5 orthologs of the META1 protein, which have 100 to 89% identity to each other (Fig. 3.5). Interestingly, *T. brucei* META1 proteins are upregulated in procyclic forms (Jones *et al.*, 2006; Jensen *et al.*, 2009), whereas the META2 protein is preferentially expressed in bloodstream trypanosomes (Liu *et al.*, 2009). In both, *Leishmania* and *T. brucei* META1 and META2 genes are located in the same gene cluster, only separated by one gene not related to calpains, which is also conserved in the two species (Fig. 3.5). Analysis of the genome locus strongly suggests that *T. brucei* META1 and META2 genes were duplicated several times after *Leishmania* divergence, indicating a possibly important function of these proteins in *T. brucei*.

Figure 3.4.

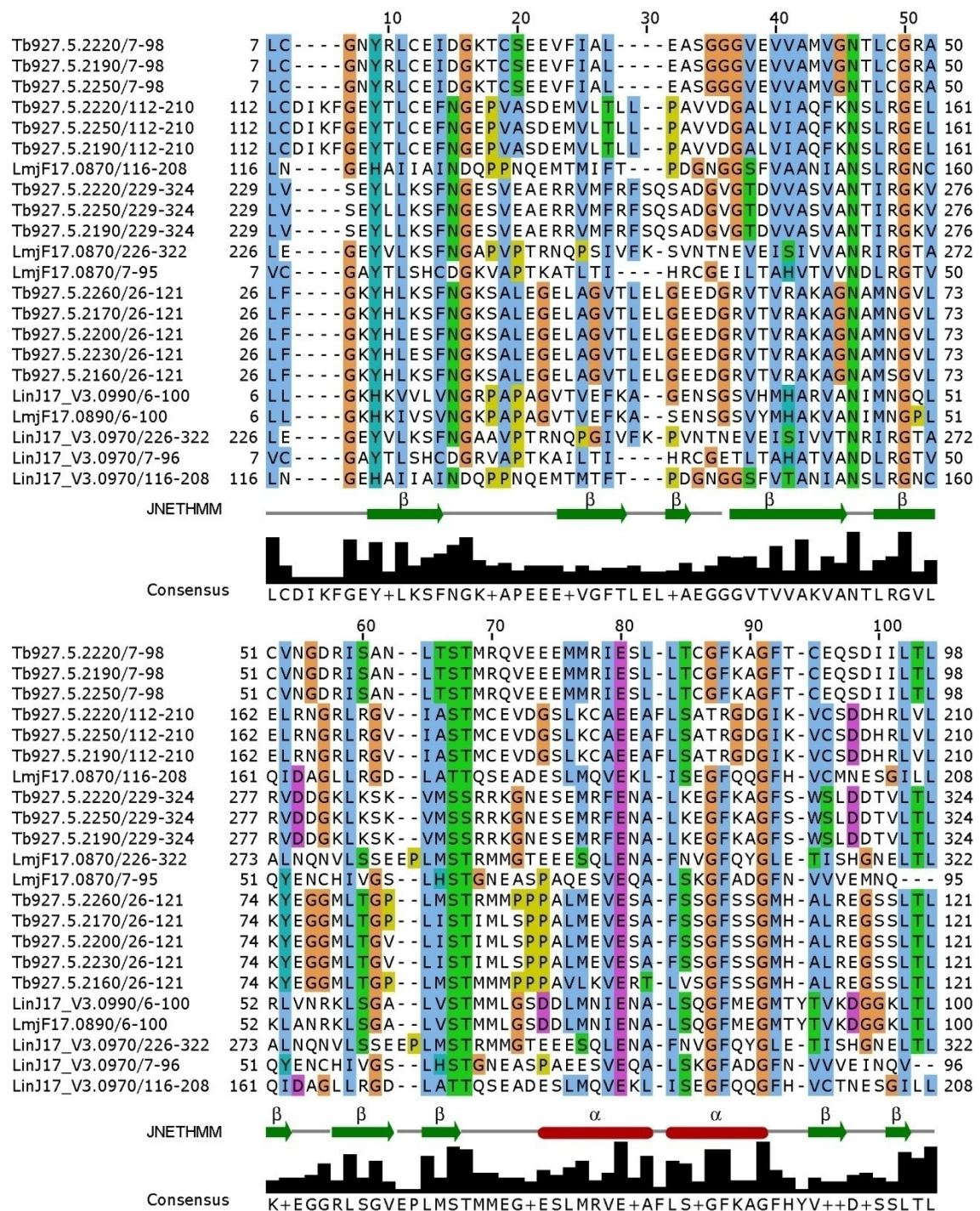


Figure 3.4. Multiple alignment of the META domain sequences of trypanosomatid calpain-like proteins by the MAFFT method.

Figure 3.5.

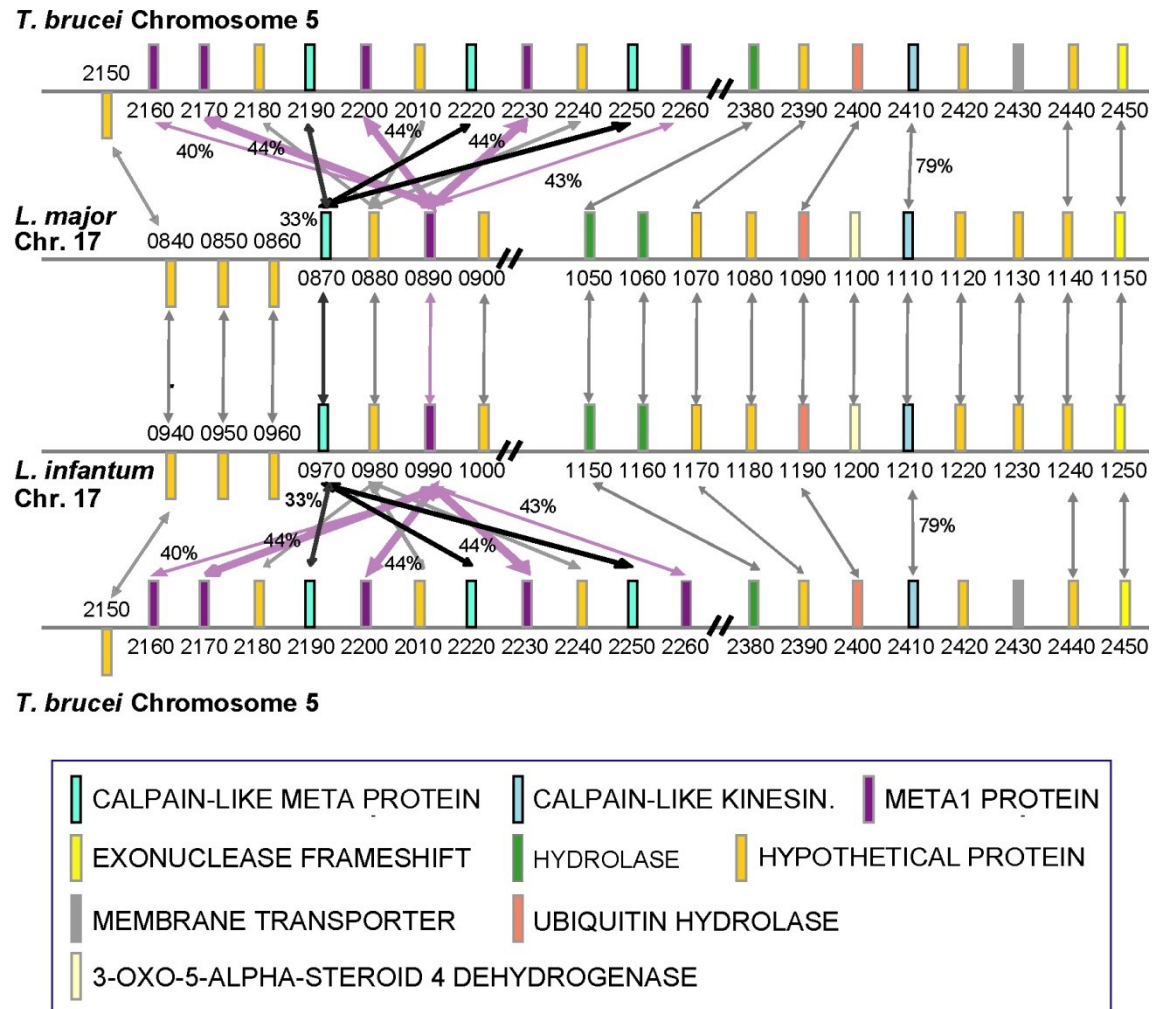


Figure 3.5. Synteny map of kinetoplastid META proteins: the newly identified *T. brucei* calpain-like sequences Tb927.5.2190, Tb927.5.2220, Tb927.5.2250 and Tb927.5.2410 genome loci with their orthologous regions in *L. major* and *L. infantum*. Tb927.5.2490, Tb927.5.2220 and Tb927.5.2250 are identical gene copies and contain three degenerate META domains. The three sequences are located in neighbouring positions on chromosome 5, forming a gene cluster which also contains five orthologs of the *Leishmania* META1 protein. Each META-domain containing calpain-like sequence is located between two META1 orthologs and preceded by a hypothetical protein of unknown function. The amino acid sequences of the three hypothetical proteins preceding each META-containing calpain-like sequence are 99-100% identical to each other and are the orthologs of *L. major* LmjF17.0880, an hypothetical protein also located between the META1 gene, LmjF17.0890, and the META-containing calpain-like sequence, LmjF17.0870. META-domain containing calpain-like sequences orthologs are linked by black arrows and META1 protein orthologs are linked by violet arrows. All other orthologs are linked by grey arrows. Tb927.5.2410, LmjF30.2040 and LinJ21.1770 are group 3, subgroup 3d calpain-like sequences, which contain a kinesin-like motor domain. These genes are also orthologs located on syntenic regions.

The Kinesin motor domain and Armadillo tandem repeats.

An N-terminal kinesin-like motor domain was found in three calpain-related sequences: *T. brucei* Tb927.5.2410, *L. major* LmjF17.1110 and *L. infantum* LinJ17.1220. The three sequences contain a C-terminal calpain-like domain III, but lack a calpain “protease” domain II. Although these sequences have not been previously described as calpain-like proteins by Ersfeld and co-workers, Tb927.5.2410 and LmjF17.1110 have been identified as kinesin-like ungrouped proteins by phylogenetic analysis of the kinesin superfamily (Richardson *et al.*, 2006; Wickstead and Gull, 2006). Kinesins are motor proteins that employ energy derived from ATP hydrolysis to transport cargoes unidirectionally, using microtubules as tracks for motility (Richardson *et al.*, 2006). They are involved in many essential processes, as microtubule dynamics, segregation of organelles, intracellular trafficking, motility, intraflagellar transport and ciliary functions (Verhey and Hammond, 2009). All kinesins have a conserved motor domain of ~350 amino acid residues —located at the N-terminus, C-terminus or internally—, which contains a highly conserved ATP binding motif and a motor or microtubule binding region (Yang *et al.*, 1989). The motor domain is followed by a coiled coil region important for dimerisation and a non-conserved tail domain, which is thought to interact with specific cargoes (Richardson *et al.*, 2006).

Multiple alignment of the three trypanosomatid calpain-like sequences containing a kinesin-like motor domain by the MAFFT method (Kato *et al.*, 2002), showed a high degree of conservation between the three sequences (Appendix 3, Fig. A.3.1) —95% amino acid identity between LmjF17.1110 and LinJ17.1220, 61% between these and Tb927.5.2410. Analysis of their domain composition and architecture, based on SMART predictions (Schultz *et al.*, 1998), showed that the N-terminal kinesin-like motor domain was followed by a small coiled-coil region (37 amino acids long) and 4 armadillo/ β -catenin-like tandem repeats (ARM-repeats), before the C-terminal

kinetoplastid calpain-like domain III (Fig. 3.6). Interestingly, the same phylogenetic analyses that identified these proteins as kinesin-like sequences, (*Richardson et al., 2006; Wickstead and Gull, 2006*), also found ten ungrouped plant-specific kinesins containing an N-terminal kinesin-like motor domain followed by a coiled-coil region and 3 or 4 ARM-repeats. Thus, the ARM-repeats-containing kinetoplastid and plant ungrouped kinesin-like proteins form a clade that includes members from two of the eight main eukaryotic lineages (*Richards and Cavalier-Smith, 2005*). This clade could represent a novel kinesin family and/or early derived members of the so far recognised kinesin families (*Richardson et al., 2006*). However, the low conservation of these sequences' kinesin motor domains argued against the possibility that they had kinesin-motor activity (*Wickstead and Gull, 2006*). However, latter studies have demonstrated that some of these kinesin-like proteins depolymerised microtubules in plants (*Sakai et al., 2008*).

In general, kinesins with an N-terminal motor domain undergo directed motility towards the plus (rapidly growing) end of microtubules, whereas kinesins with a C-terminal motor domain move in the opposite direction, towards the minus (slow growing) end of microtubules. Additionally, kinesins with a central motor domain destabilise microtubules rather than move along their surface (*Verhey and Hammond, 2009*). However, there are important exceptions to this generalisation, among which are the ARM-repeat-containing kinesins or ARK proteins of *Arabidopsis thaliana*. The ARK1 protein, which contains an N-terminal kinesin-like motor domain followed by four ARM repeats, is involved in root-hair morphogenesis, controlling tip growth by promoting microtubule depolymerisation (*Sakai et al., 2008*). Additionally, the ARM-repeats, which typically form a superhelix, are known to mediate protein-protein interactions (*Huber et al., 1997; Conti et al., 1998*). Moreover, diverse proteins containing ARM-repeats function in intracellular signalling and cytoskeletal regulation (*Vale, 2003*). Thus, the

Figure 3.6.

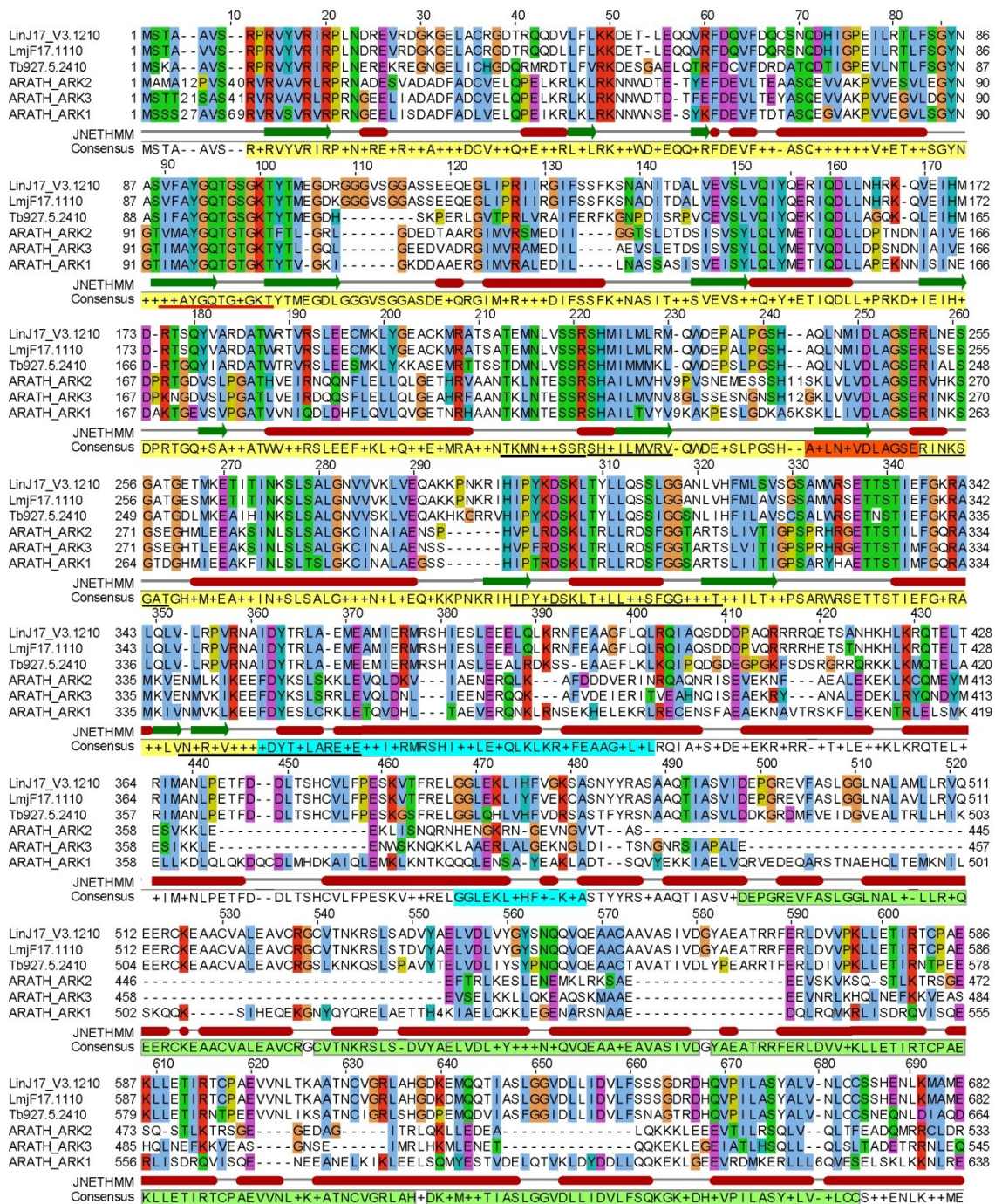


Figure 3.6. Multiple alignment by the MAFFT method of *Arabidopsis thaliana* kinesin ARK proteins and trypanosomatid calpain-like sequences containing an N-terminal kinesin-like motor domain. Both groups of sequences contain ARM-repeats. Kinesin motor domain is highlighted in yellow, with the ATP binding motif underlined in red, the kinesin heavy chain signature is underline in black and the motor region is highlighted in red. The coiled coil region is highlighted in turquoise. ARM-repeats are highlighted in green. Calpain-like domain III starts at Tb927.5.2410 685 bp, outside the alignment (for calpain-like domain III of these kinetoplastid sequences see Figure 3.3). Secondary structure prediction by the Jnet method, α -helices are indicated by red bars and β -strands by green arrows.

ARM-repeats of ARK1 may also be important for protein-protein interactions (*Sakai et al., 2008*). Multiple alignment of *Arabidopsis thaliana* ARK proteins and trypanosomatid calpain-like sequences containing a kinesin-like motor domain showed a high degree of similarity between these sequences' kinesin motor domains, including their ATP and microtubule binding motifs (Fig. 3.6). However, the subpellicular microtubules of trypanosomes do not seem to depolymerise. On the contrary, they remain stable throughout the cell-cycle and cell-growth occurs by the intercalation of new microtubules into the existing subpellicular corset (*Sherwin and Gull, 1989a*). Nevertheless, the similarity of these trypanosomatid calpain-like proteins to *Arabidopsis* ARK proteins, including their kinesin motor domains and ARM-tandem-repeats, together with the structural similarity of the calpain-like domain III to the C2 lipid-binding domain of membrane associated proteins (*Rizo and Sudhof, 1998*), suggest that these kinetoplastid calpain-like proteins might be involved in the microtubule dynamics of trypanosomes, perhaps remodelling microtubule-membrane attachments.

Adenylate cyclase domain.

The adenylate cyclase (ADCY) catalytic domain was found in two newly identified calpain-like sequences specific of *Leishmania* species: LmjF36.3180 in *L. major* and LinJ36_v3.3330 in *L. infantum*. Multiple alignment and comparison of these sequences to other kinetoplastid calpain-like proteins by the MAFFT method (*Katoh et al., 2002*), showed that they contain a C-terminal calpain-like domain I^K (Fig. 3.1), but lack any other calpain-related domain. BLASTp searches on the genome annotation database of both *L. major* and *L. infantum* on geneDB (<http://www.genedb.org/trypan>), using the full length of these sequences, showed a high degree of similarity to known *L. major* and *L. infantum* Receptor-type Adenylate Cyclases (RACs), (Appendix 3, BLAST searches). Multiple alignment and comparison of the two calpain-like sequences containing an ADCY catalytic domain to *Leishmania* RACs revealed that the two

calpain-like sequences, LmjF36.3180 and LinJ36_v3.3330, contain an N-terminal signal peptide, a transmembrane domain and an ADCY catalytic domain before their calpain-like domain I^K, similarly to RACs (Fig. 3.7).

ADCYs are the effector enzymes responsible for converting adenosine-triphosphate (ATP) to cyclic-adenosine-3,5-monophosphate (cAMP), an ubiquitous second messenger that mediates multiple cellular responses, mainly by activating cAMP-dependant kinases. ADCYs have been classified into different families according to their biochemical and structural properties (*Hurley, 1999; Patel et al., 2001*). The best-characterised members of these families are mammalian ADCYs, which have a conserved structure: two domains of six transmembrane helices and two intracellular pseudosymmetrical catalytic domains, which are separated by the second transmembrane domain (*Taussig and Zimmermann, 1998*). Trypanosomatid RACs differ extensively from their mammalian counterparts, with a predicted molecular architecture of a large hydrophilic N-terminal domain, presumably extracellular, followed by a single transmembrane domain and a highly conserved catalytic domain (*Alexandre et al., 1990; Sanchez et al., 1995; Taylor et al., 1999*). This structure resembles that of guanylyl cyclase receptors, suggesting that they may function as sensory receptors (*Lucas et al., 2000*). Furthermore, several data suggest that the ADCY/cAMP pathway may be involved in the regulation of *T. cruzi* differentiation in response to exogenous peptide ligands (*Gonzales-Perdomo et al., 1988; Ouaisi et al., 1992; Fraidenraich et al., 1993*). During the differentiation of *T. brucei* from bloodstream to procyclic form there is also transient adenylate cyclase activation (*Rolin et al., 1993*). Moreover, despite the different structures and low overall sequence conservation between trypanosomatid and mammalian ADCYs, the catalytic domain of trypanosomatid RACs conserves all the strategic residues of mammalian cyclases (*Taylor et al., 1999; Naula and Seebeck, 2000*).

Figure 3.7.

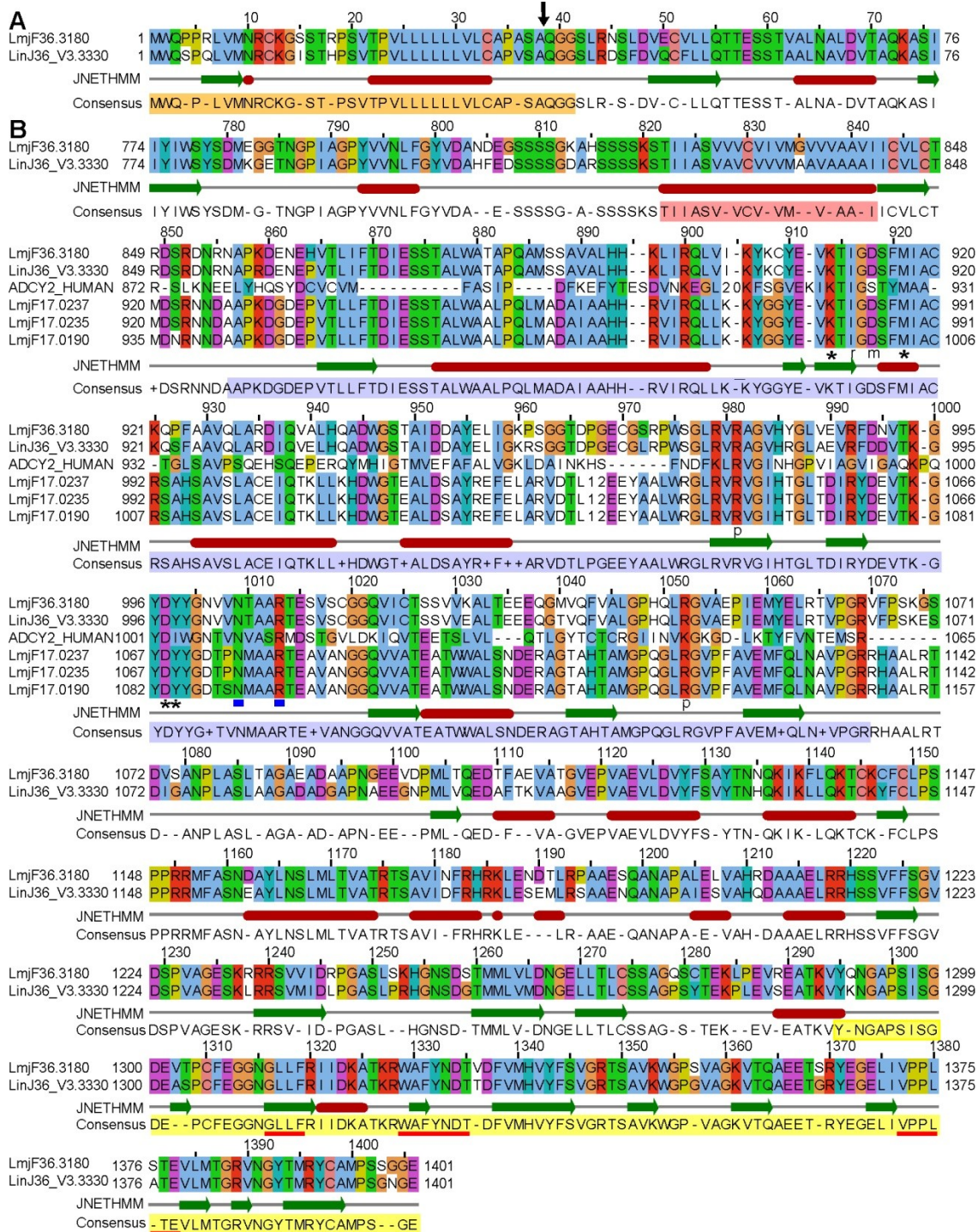


Figure 3.7. Trypanosomatid ADCYs containing a calpain-like domain I^K (*Leishmania* spp. only). **A:** N-terminal signal peptide is highlighted in orange, arrow indicates the predicted cleavage site. **B:** Transmembrane domain, highlighted in pink. Adenylate cyclase catalytic domain, highlighted in grey. C-terminal kinetoplastid calpain-like domain I^K, highlighted in yellow, typical motifs underlined in red. The C2 catalytic domain of human type II ADCYs and of related *L. major* RAC proteins have been included for comparison. Residues essential for catalytic activity are blue underlined (N and R). Other functional residues are indicated by symbols: *, adenine binding, the first one, K (LmjF36.3180, LmjF36_v3.3330 K₉₁₀) confers adenosine specificity (substituted by E in guanylyl cyclases); m, magnesium binding; p, pyrophosphate binding; and r, ribose binding.

Sequence analysis of the newly identified kinetoplastid calpain-like sequences containing an ADCY catalytic domain also showed the conservation of most of these strategic residues (Fig. 3.7):

- Residues essential for catalytic activity —N and R— that stabilise the transition state during ATP hydrolysis. N(1004) and R(1008) in both sequences.
- Residues essential for nucleotide-binding, including the K that confers adenosine specificity (substitute by E in guanylate cyclases): K(910), M(917), D(997) and Y(998) in both sequences
- Residues involved in Mg²⁺-binding: D(914) in both sequences.
- Residues involved in ribose-binding: I(912) in both sequences..
- Residues involved in phosphate-binding: R(977) in both sequences.

3.2.3 Analysis of calpain-like sequences containing 67-68 amino acid tandem-repeats.

As discussed earlier, short fragments containing the repeats of some calpain-like sequences identified by Ersfeld and co-workers (group 4 and 5 calpain-like proteins, *Ersfeld et al., 2005*) have been published as antigens FRA, in *T. cruzi* (TcCLP441.10); GM6 in *T. brucei* (Tb11.47.0036) and Lcr1 in *L. chagasi* (*Lafaille et al., 1989; Muller et al., 1992; Wilson et al., 1995*). Antibodies against FRA, GM6 and Lcr1 detect high molecular weight proteins by western blotting. Additionally, FRA and GM6 have been shown to localise at the flagellum attachment zone (FAZ) in *T. cruzi* and *T. brucei*, respectively (*Cotrim et al., 1990; Muller et al., 1992; Cotrim et al., 1995*). In this section, I present the bioinformatic analysis of the newly found *T. brucei* repeat-containing calpain-like sequences: Tb11.57.0008 and Tb11.v4.0001, whose repeats are homologous to those of GM6 and Lcr1, respectively. Additionally, database mining of the *T. cruzi* genome project in GeneDB (<http://www.genedb.org/tryp/>) with the FRA sequence, identified two calpain-like sequences containing FRA repeat units:

Tc00.1047053505985.9 and Tc00.1047053511441.10 (Appendix 3, *T. cruzi*-FRA-sequences file), whose analysis is also presented here.

The basic repeat unit of all identified trypanosomatid repeat-containing calpain-like sequences is 204 bp (68 amino acids) long. This unit was first described in different strains of *T. cruzi*, through a series of λ gt11 DNA libraries, as part of highly immunogenic clones bearing 204 bp repetitive epitopes: clone 1 (*Ibanez et al., 1988*), FRA clones (*Lafaille et al., 1989*), JL7 clone (*Levin et al., 1989*) and H49 clone (*Cotrim et al., 1990*). The repeat units of these clones showed a high degree of similarity to each other (88% average amino acid identity), indicating that they probably encode parts of the same ORF (*Cotrim et al., 1995*). Specifically, FRA and H49 repeats, the best studied of these clones, are 94% identical, and their few amino acid substitutions are mainly conservative (*Lafaille et al., 1989; Cotrim et al., 1995*). Immuno-electron microscopy studies have shown that these clones encode part of a protein that localises at the FAZ of *T. cruzi* epimastigotes and trypomastigotes (*Lafaille et al., 1989; Souto-Padron et al., 1989; Cotrim et al., 1990*), associated to the flagellar cytoskeleton, parallel to the axoneme and concentrated on the side of the flagellum that faces the parasite body (*Lafaille et al., 1989; Souto-Padron et al., 1989*). However, ultrastructural localisation studies of *T. cruzi* repetitive antigen H49, which contains four identical 204 bp repeat units (94% identical to the FRA repeats), showed that H49 is neither a component of the flagellar axoneme nor of the paraflagellar rod (PFR), but localises at the FAZ filament that connects the flagellum to the cell body (*Cotrim et al., 1990*). Furthermore, the claim by *Lafaille et al 1989* that FRA is located in the flagellar side of the FAZ is not enough justified, as their published data (*Lafaille et al., 1989*) does not show if FRA localises at the flagellar or cytoplasmic side of the FAZ complex or at the fibrous connections that extend between both. Additionally, immunoelectron microscopy has shown that during the differentiation of *T. cruzi* amastigotes into

trypomastigotes, H49 concentrates at the flagellum-cell body contact region, close to the point where the flagellum emerges from the body, suggesting that H49 could be involved in the assembly of the growing FAZ (*Cotrim et al., 1995*).

Sequencing of the FRA clone revealed that its 204 bp repeat units contain a unique HindIII site, in the middle, and two TaqI sites, near the ends (*Lafaille et al., 1989*). The presence of these restriction sites was used in Southern blot analysis to determine the number of repeat units in the FRA gene, carrying out partial digestions of *T. cruzi* genomic DNA. The results indicated that FRA contains at least fourteen repeat units, but also suggested that FRA is probably not a single copy gene (*Lafaille et al., 1989*). Western blot analysis showed that the anti-FRA antibodies recognise a high molecular weight protein, of about 300 kDa, which, although weakly detected in epimastigotes, is abundant in both metacyclic and tissue culture trypomastigotes (*Lafaille et al., 1989*). A similar approach was carried out with the H49 clone, where Northern blot analysis revealed that the H49 gene also encodes a protein of about 300 kDa, further supporting the hypothesis that both antigens are part of the same ORF. Additionally, Southern blot analysis showed that the H49 gene contains at least fifteen tandemly arranged 204 bp repeat units.

Computer-assisted secondary structure predictions suggested that the repeat region of FRA/H49 protein sequence consist of alpha-helical regions of 34 and 18 amino acids separated by a turn and coiled regions (*Cotrim et al., 1995*). Thus, the protein might assume a highly regular configuration of alternated helical and non-helical segments, which is unaffected by the amino acid substitutions found in some repeat units (*Cotrim et al., 1995*). The highly conserved repetitive structure of the FRA/H49 containing protein, its unusually high content of alpha helices and its cellular localisation at the FAZ filament, suggest that it may have a structural role in maintaining the cell body connected to the flagellum (*Cotrim et al., 1995*). The FRA antigen has been later

identified as part of a *T. cruzi* calpain-like sequence, TcCLP441.10 (Ersfeld *et al.*, 2005). TcCLP441.10 is constituted by an N-terminal array of FRA repeats followed by a calpain-like domain II and III, and was classified as a kinetoplastid group 5 calpain-like sequence (Ersfeld *et al.*, 2005).

GM6, the *T. brucei* repetitive antigen homologous to FRA, was also characterised through a cDNA expression library (Muller *et al.*, 1992). GM6 contains one incomplete and two complete copies of highly conserved, tandemly arranged 204 bp repeat units (99% amino acid identity between the two complete repeats), that have a 51% amino acid identity with the FRA repeats (Appendix 3, Fig. A.3.2). Immunoelectron microscopy studies have shown that GM6 is associated with the network of fibres that extend between the microtubules of the cell body and the PFR, being most likely a component of the FAZ filament (Muller *et al.*, 1992). Western blot analysis using the anti-GM6 antibody showed that it also recognises a high molecular weight protein of more than 300 kDa, in both bloodstream and procyclic form trypanosomes. The completion of the *T. brucei* genome project (Berriman *et al.*, 2005) allowed the identification of GM6 as part of a group 5 calpain-like protein, Tb11.47.0036, containing N-terminal repeats followed by one copy of calpain-like domains II and III (Ersfeld *et al.*, 2005).

Lcr1, another repetitive antigen originally identified in *L. chagasi* (Wilson *et al.*, 1995) contains 4 divergent 67 amino acid repeats (201 bp) homologous to FRA 68 amino acid repeats (49% identity, Appendix 3, Fig. A.3.2). Analyses of Lcr1 by Southern and western blotting showed that Lcr1 is a single copy gene encoding a protein of more than 200 kDa (Wilson *et al.*, 1995). Lcr1 was later identified as part of a *L. major* calpain-like sequence, LmjF27.0500 (Ersfeld *et al.*, 2005). LmjF27.0500 and its *T. brucei* ortholog, Tb11.47.0035, contain three copies of the calpain-like domain II, the second and third copies separated by a long array of 67-68 amino acid tandem

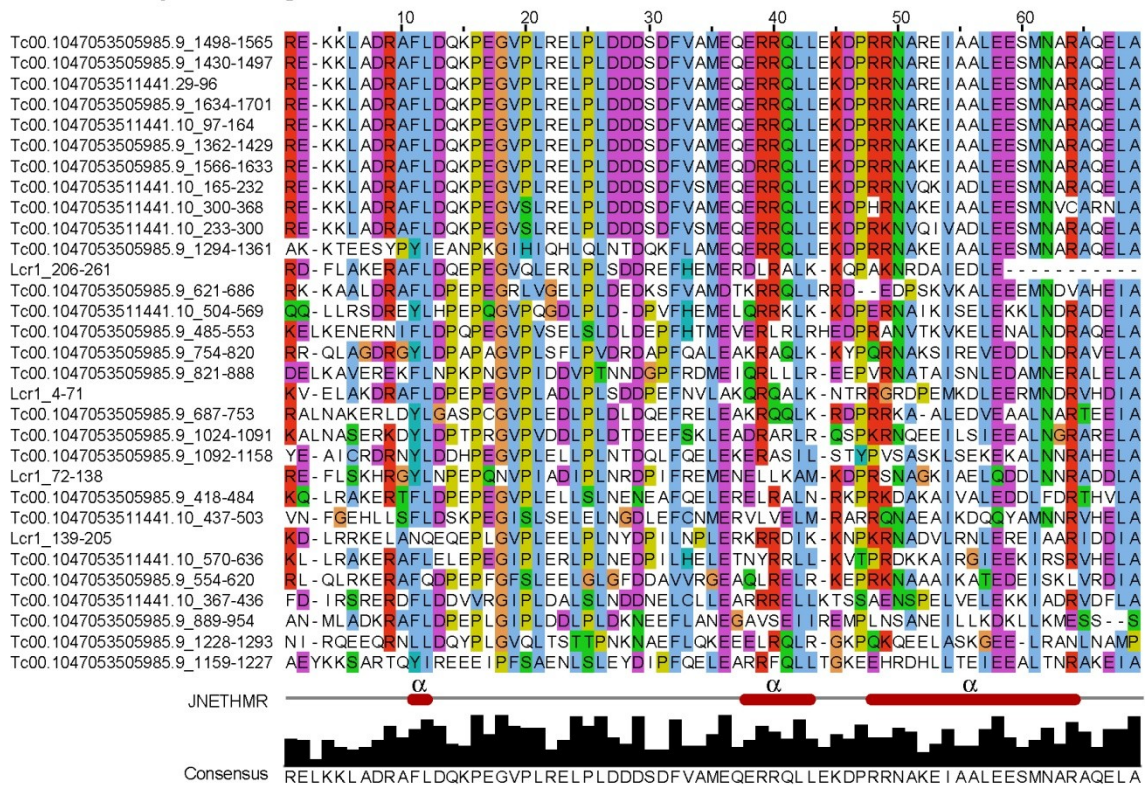
repeats, and 2 (my analysis) or 3 (*Ersfeld et al., 2005*) copies of the calpain-like domain III. Both sequences were classified as group 4 calpain-like proteins by Ersfeld and co-workers (*Ersfeld et al., 2005*).

During this thesis work, database mining of *T. cruzi*, *T. brucei*, *L. major* and *L. infantum* genome sequences in GeneDB (<http://www.genedb.org/tryp/>) with the sequences of FRA, GM6 and Lcr1 repetitive antigens, rendered some interesting information. In *T. cruzi*, two sequences were found to contain the FRA repeat units: the previously described TcCLP441.10 calpain-like sequence (Tc00.1047053511441.10) and the novel Tc00.1047053505985.9, which, following the nomenclature proposed by Ersfeld *et al., 2005*, I have named TcCLP985.9 (Appendix 3, FRAseq-T.cruzi.txt file). TcCLP441.10 contains 10 FRA repeat units, two of them 99% identical to FRA, whereas the novel TcCLP5985.9 contains 22 FRA repeat units, five of them 99% identical to FRA (Fig. 3.8 and Appendix 3, Fig. A.3.3). The two sequences are located in different contigs and both lack start and stop codons, suggesting that they may be parts of the same ORF. However, the nature of the *T. cruzi* genome sequence (obtained from a hybrid of two different strains) and its unfinished state make very difficult to positively distinguish between fragments of the same gene, gene copies or alleles in different contigs.

In *T. brucei*, database searches using FRA, GM6 and Lcr1 sequences as baits (Appendix 3, FRAseq-T.brucei.txt, GM6seq-T.brucei.txt and Lcr1seq-T.brucei.txt files) identified two novel calpain-like sequences bearing 68 amino acid tandem repeats: Tb11.57.0008 and Tb11.v4.0001. These two sequences are located in chromosome 11, as the previously described *T. brucei* group 4 and 5 sequences, Tb11.47.0035 and Tb11.47.0036. The domain architecture of Tb11.57.0008 resembles that of group 4 kinetoplastid calpain-like sequences, as it contains two copies of domains II and III followed by 68 amino acid tandem repeats (Fig. 3.2 and 3.3), although it lacks the third

Figure 3.8.

A. FRA Repeats Alignment.



B. Alignment of the most conserved FRA repeats.

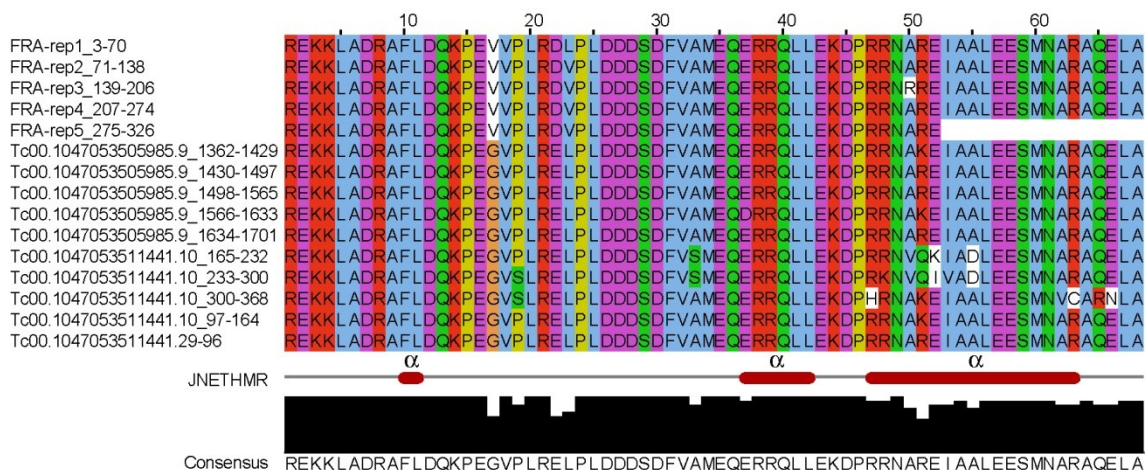


Figure 3.8. A. Multiple alignment by the MAFFT method of the repeat units of the FRA antigen and *T. cruzi* FRA containing sequences: Tc00.1047053505985.9 (TcCLP985.9) and Tc00.1047053511441.10 (TcCLP441.10). **B.** Alignment of only the most conserved repeats of TcCLP441.10 and TcCLP985.9 with the repeats of the FRA antigen. The colour scheme is according to the ClustalX colour code. Secondary structure prediction is by the JNet method.

copy of domains II and III characteristic of group 4 proteins. However, multiple alignment of Tb11.57.0008 to known *T. brucei* repeat-containing calpain-like sequences revealed a high degree of similarity to Tb11.47.0036, a group 5 sequence containing GM6 repeats. Tb11.57.0008 (5793 amino acids long) contains 67 repeat units (positions 1250-5795) homologous to the 68 amino acid repeats of GM6 and Tb11.47.0036 (Fig. 3.9 and Appendix 3, Fig. A.3.4). Of these 67 repeats, 59 are highly conserved —32 are 100% identical to the repeats of the GM6 antigen (*Muller et al., 1992*) and 14 have only one amino acid substitution (Fig.3.9 and Appendix 3, Fig. A.3.4). Tb11.47.0036 (1409 amino acids long) contains 10 complete and one incomplete repeats (positions 1-773), which range from 100% to 32% amino acid identity to the GM6 repeats (Fig.3.9 and Appendix 3, Fig. A.3.4). Due to their similarity to GM6, I termed Tb11.47.0008 and Tb11.47.0036 repeats as GM6-like repeats.

Analysis of the genome locus of Tb11.57.0008 and Tb11.47.0036 revealed that the two sequences are located in different contigs of chromosome 11, in adjacent positions of their provisional assembly (Fig. 3.10). Furthermore, Tb11.57.0008 and Tb11.47.0036 annotated ORFs lack a stop and a start codon, respectively: Tb11.57.0008 ends in an incomplete repeat of 63 amino acids (positions 5733-5795) and Tb11.47.0036 starts with a truncated repeat of 29 amino acids (Appendix 3, Fig. A.3.5). These data suggest that both sequences may be parts of a unique ORF, although, so far, I do not have further data to support this hypothesis. Secondary structure prediction of Tb11.57.0008 and Tb11.47.0036 repeats by the JNet algorithm (*Cole et al., 2008*) suggests that they consist of two alpha-helix regions of 9 and 12 amino acids (positions 35-43 and 52-63 of the 68 amino acids GM6 repeats, Fig. 3.9) separated by a turn region (position 48-49, Fig. 3.9). This structure is highly similar to that predicted for the FRA/H49 repeats (*Cotrim et al., 1995*) and, as FRA, it might assume a greatly regular conformation of alternated helical and not-helical segments.

Figure 3.9.

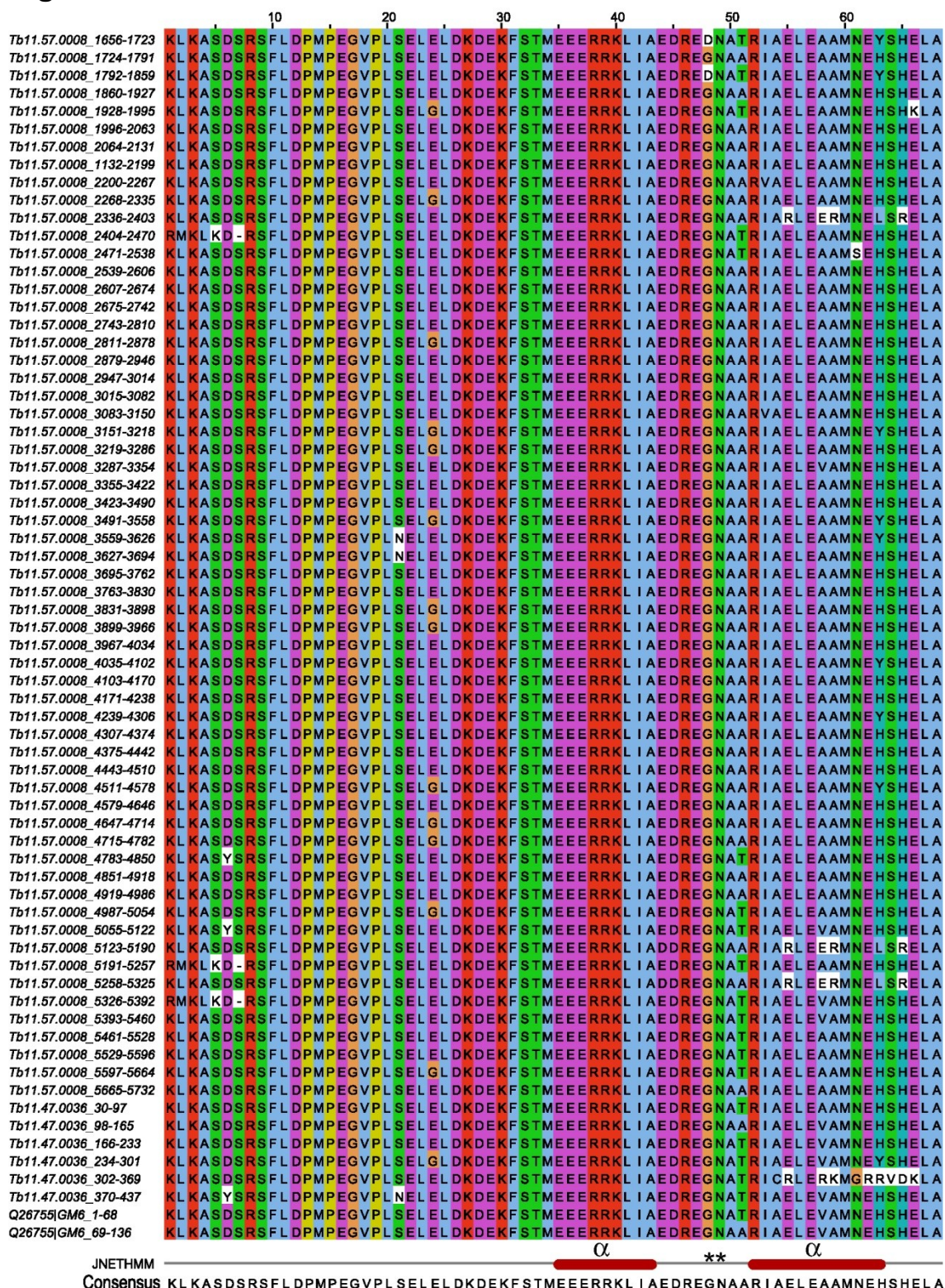


Figure 3.9. Multiple alignment by the MAFFT method of Tb11.57.0008, Tb11.47.0036 and GM6 68 amino acids repeats. Only the most conserved repeat units of Tb11.57.0008 and Tb11.47.0036 have been included. The colour scheme is according to the ClustalX colour code. Secondary structure prediction is by the JNet method: α -helices are indicated by red bars and the turn region by asterisks. For an alignment of all repeat units see Appendix 3, Fig. A.3.4.

Figure 3.10.

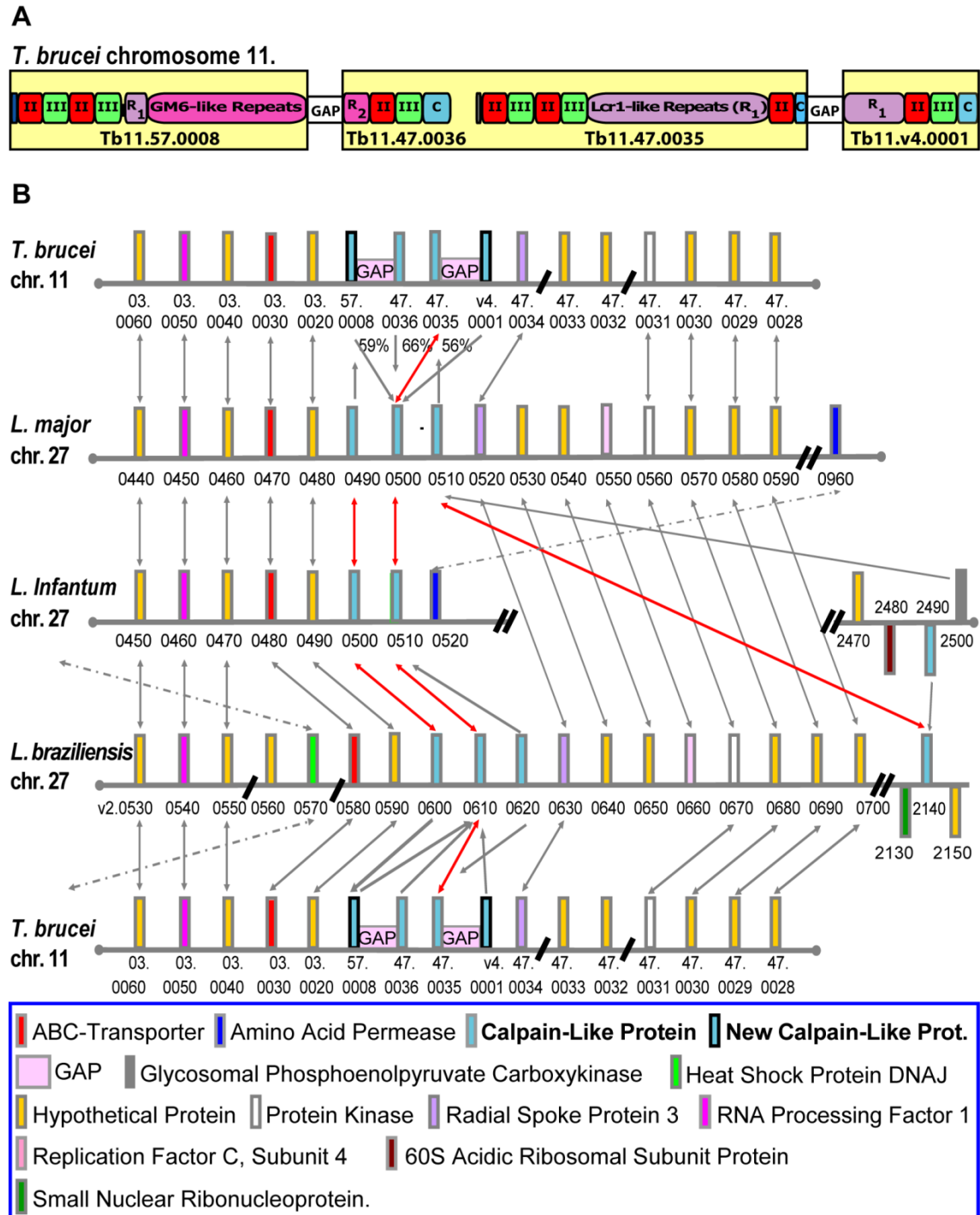


Figure 3.10. A. Detailed diagram of the *T. brucei* genome locus of tandem repeats-containing calpain-like sequences. Contigs are indicated by yellow boxes and gaps between contigs by white boxes. Predicted protein domain architecture is indicated for each sequence. The two different kinds of repeats are indicated by two different colours: Purple for Lcr1-like repeats (R_1) and fuchsia for GM6-like repeats (R_2). **B.** Genomic organisation of tandem repeats containing calpain-like sequences in different trypanosomatid species based on reciprocal BLAST results. Orthologs across different species are indicated by grey connection arrows, red arrows for calpain-like sequences. Gaps in the genome sequences are indicated by black solid boxes. Newly identified calpain-like sequences are indicated by bold-blue boxes.

The cellular localisation of GM6 at the FAZ, its association with cytoskeleton —being most probably a component of the FAZ filament— and the highly regular structure of its tandem repeats suggest a structural role for Tb11.57.0008 and Tb11.47.0036 GM6-containing protein/s, which could be implicated in maintaining the cell body attached to the flagellum. However, the presence of several calpain-like domains II and III hint at a more important and complex role for these protein/s.

Tb11.v4.0001, the other *T. brucei* new calpain-like repeat-containing sequence identified during this thesis work, has the same domain architecture as group 5 calpain-like proteins: an N-terminal array of tandem repeats followed by one copy of calpain-like domains II and III (Fig. 3.2 and 3.3). However, multiple alignment and comparison of Tb11.v4.0001 to known *T. brucei* repeat-containing calpain-like sequences revealed a high degree of similarity to Tb11.47.0035, a group 4 sequence (Fig. 3.11). Tb11.47.0035 (5799 amino acids long) contains 62 tandem repeats of 68 amino acids (positions 1211 to 5799), which are homologous to those of the Lcr1 antigen. Tb11.v4.0001 (1649 amino acids long) contains 15 complete and one incomplete 68 amino acid tandem repeats (positions 1 to 1014) similar to those of Tb11.47.0035 (17 to 100% identity, Fig. 3.11 and Appendix 3, Tb01-35_rep_align.txt file). The 100% identical repeat region between both sequences corresponds to the N-terminal region of Tb11.v4.0001 (positions 4 to 187) and a central region of Tb11.47.0035 (positions 3752 to 3938, Appendix 3, Fig. A.3.6). Analysis of both genes genome loci showed that they are present in two different contigs of chromosome 11, in adjacent positions of their provisional assembly, similarly to Tb11.57.0008 and Tb11.47.0036 (Fig. 3.10A). Although Tb11.47.0035 lacks a stop codon, Tb11.v4.0001 has a start one, a closer look to Tb11.v4.0001 contig revealed that its coding region starts 84 nucleotides (28 amino acids) downstream of its proposed start codon (Appendix 3, Fig. A.3.6). These 84 nucleotides code for almost the same amino acids (25 out of 28) as Tb11.47.0035 in

Figure 3.11.

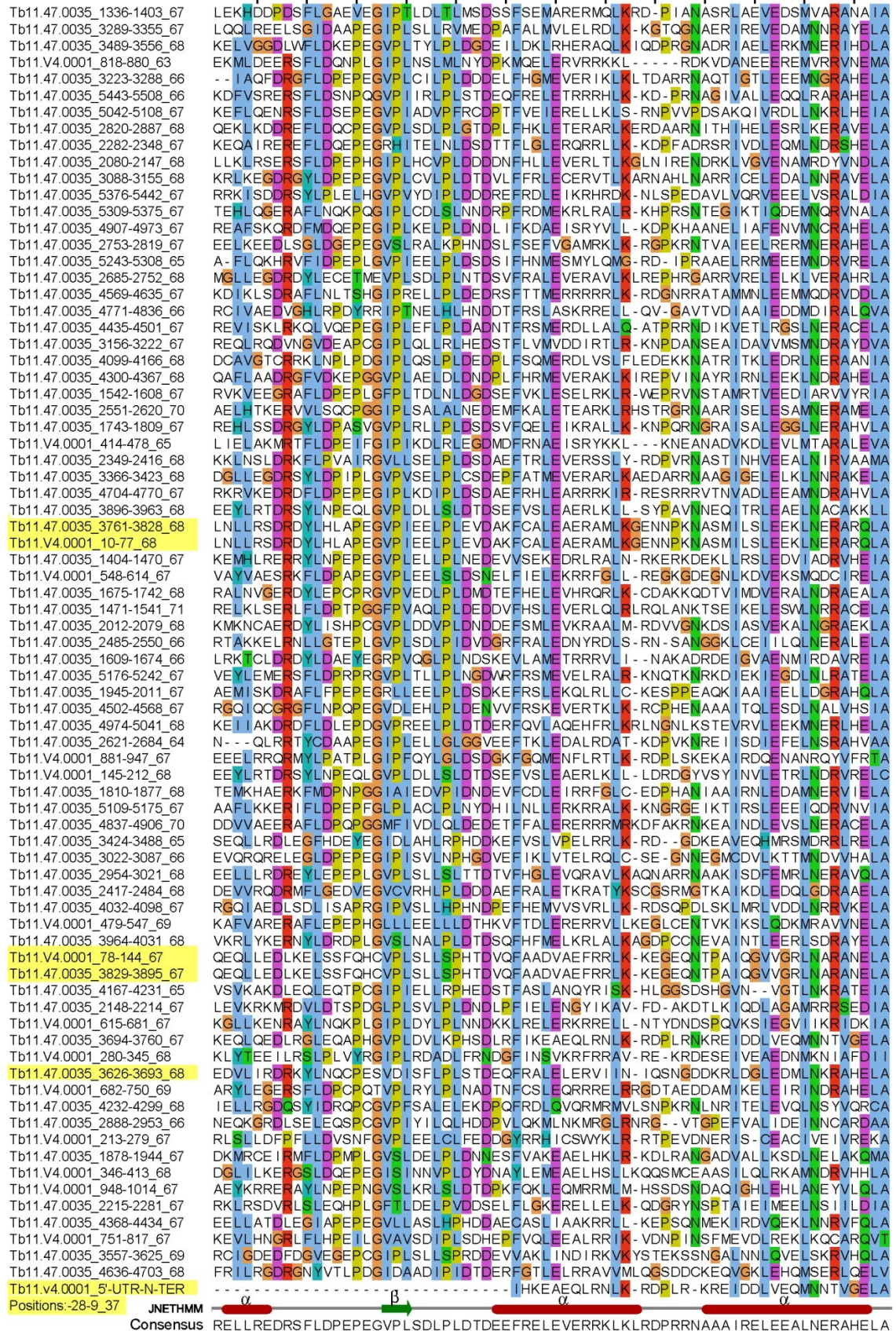


Figure 3.11. Tb11.47.0035 and Tb11.V4.0001 repeats alignment by the MAFFT method. Repeats that are 100% identical are highlighted in yellow.

the region immediately downstream of the sequence 100% identical to Tb11.v4.0001 N-terminus (Tb11.47.0035 positions 3724 to 3751, Fig. 3.11 and Appendix 3, Fig. A.3.6). Thus, the start codon of Tb11.v4.0001 in geneDB *T. brucei* current genome sequence (<http://www.genedb.org/tryp/>) corresponds to the methionine of the 36 repeat of Tb11.47.0035 (position 3572). Therefore, the identical repeat region between both sequences codes for three 68 amino acid repeat units. In addition, another region of Tb11.v4.0001 (positions 1091 to 1305) is also 100% identical to the last 215 amino acids of Tb11.47.0035 (positions 5588 to 5799, Appendix 3, Fig. A.3.6) and correspond to Tb11.v4.0001 and Tb11.47.0035 calpain-like domain II (third copy in Tb11.47.0035, Fig. 3.2). These data strongly suggest that Tb11.47.0035 and Tb11.v4.0001 are also parts of the same ORF.

Additionally, reciprocal BLAST searches of Tb11.47.0035 within the genome sequences of *T. brucei*, *T. cruzi* and *L. major* in geneDB (<http://www.genedb.org/tryp/>), showed that Tb11.47.0035 is the ortholog of *L. major* Lcr1 containing sequence, LmjF27.0500 (Appendix 3, Reciprocal_BLAST). The two genes are syntenic (Fig. 3.10B) and have the same domain architecture (Appendix 3, Fig. A.3.7). Moreover, although *T. brucei* Tb11.47.0035 and Tb11.v4.0001 contain 68 amino acids repeats (Fig. 3.11) and *L. major* LmjF27.500 contains 67 amino acids repeats (Fig. 3.12), their repeats have a high degree of similarity (Appendix 3, Reciprocal_BLAST). Thus, I termed Tb11.47.0035 and Tb11.v4.0001 repeats as Lcr1-like repeats. Reciprocal BLAST also showed that *L. major* LmjF27.0490, the sequence preceding LmjF27.500, is the ortholog of GM6 and FRA containing sequences, whereas LmjF25.500 and its following gene, LmjF27.0510, are closely related paralogs (Appendix 3, Reciprocal_BLAST). Interestingly, the three *L. major* repeat-containing sequences are located in the same contig and all of them have start and stop codons (Fig. 3.10B).

Figure 3.12.

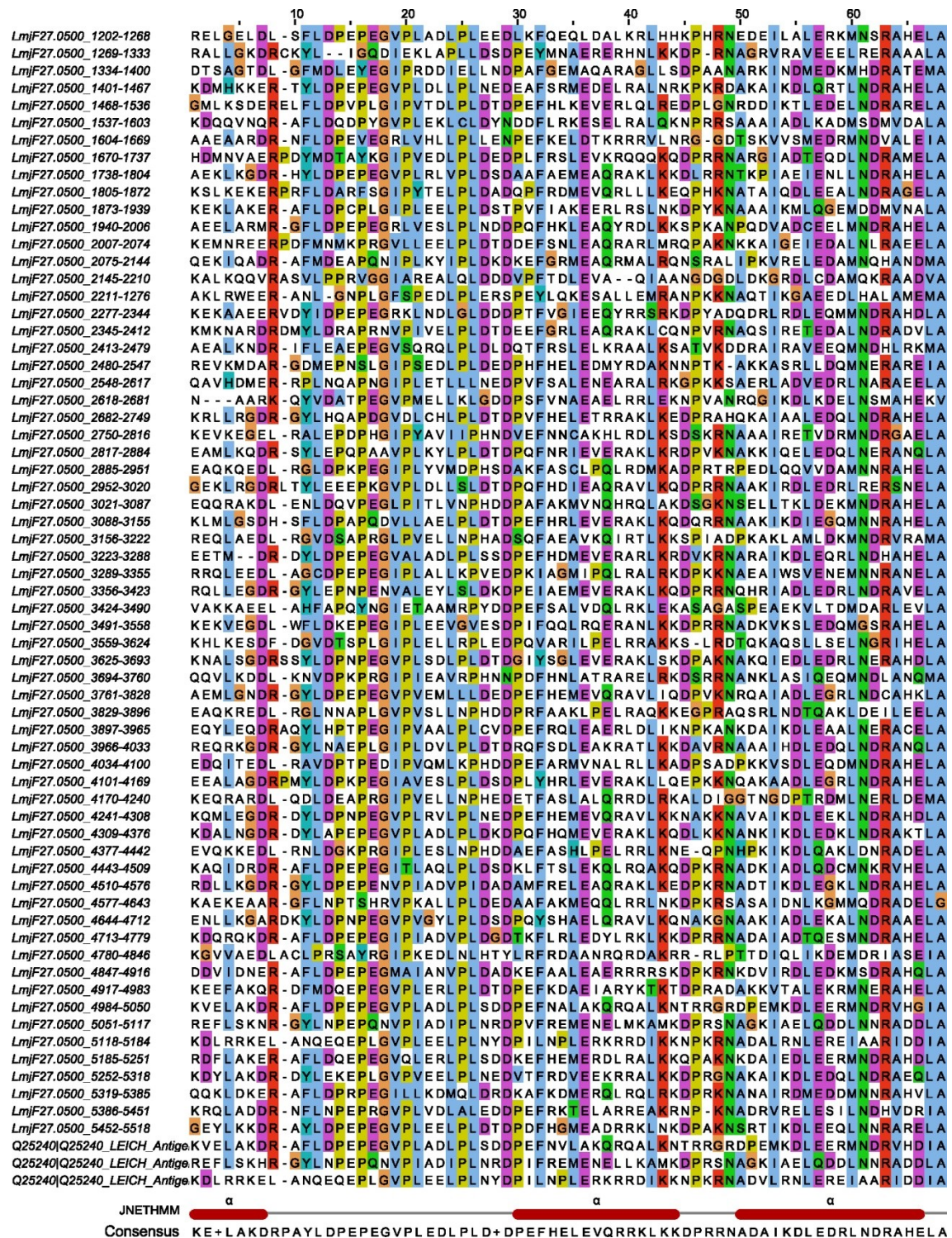


Figure 3.12. Alignment of the 67 amino acid repeats of LmjF27.500 by the MAFFT method. Lcr1 antigen repeats are placed at the bottom of the alignment for comparison. Secondary structure prediction is by the JNet method, α -helices are indicated by red bars.

The detailed analysis of the complete set of tandem-repeats-containing calpain-like sequences —comprising their genome locus, domain architecture, conservation across different species and level of synteny— rendered some intriguing and interesting information. In the first instance, all the studied kinetoplastid species have a minimum of three tandem-repeat-containing calpain-like sequences, always situated on a single chromosome and often in adjacent positions (Fig. 3.10B). The sequences are either very large, with an average size of about 5000-6000 amino acids, or truncated sequences of about 1500 amino acids, which lack the start and/or the stop codons. In addition to the large segments encoding 67-68 amino acid tandem repeats, these sequences contain one to three copies of calpain-like domains II and III. The tandem repeats are at the N-terminus (shorter sequences), between the second and third copy of calpain-like domains II and III (longer sequences) or at the C-terminus (sequences containing two copies of domains II and III, Fig 3.10A). Multiple alignment of the different repeat sequences showed diverse degrees of conservation intra and interspecies (Appendix 3, Fig. A.3.8) and reciprocal BLASTp allowed the identification of orthologous and paralogous sequences in *T. brucei*, *L. major*, *L. infantum* and *L. braziliensis* (Fig 3.10B and Appendix 3, Reciprocal_BLAST). The presence of gene clusters containing several copies of identical or paralogous genes (Myler *et al.*, 1999; El-Sayed *et al.*, 2003; Hall *et al.*, 2003) and the conservation of syntenic groups across different trypanosomatid species (Bringaud *et al.*, 1998) are typical of kinetoplastids. However, tandem-repeats-containing calpain-like sequences are in many cases more conserved between orthologs than between adjacent genes (Fig. 3.10B).

To analyse the conservation of residues between the different tandem-repeats found in kinetoplastid calpain-like proteins, I carried out several sequence alignments and comparisons of the different types of repeats in pairs: FRA with GM6, FRA with Lcr1 and GM6 with Lcr1 repeats. For comparison of FRA and GM6 repeats I performed the

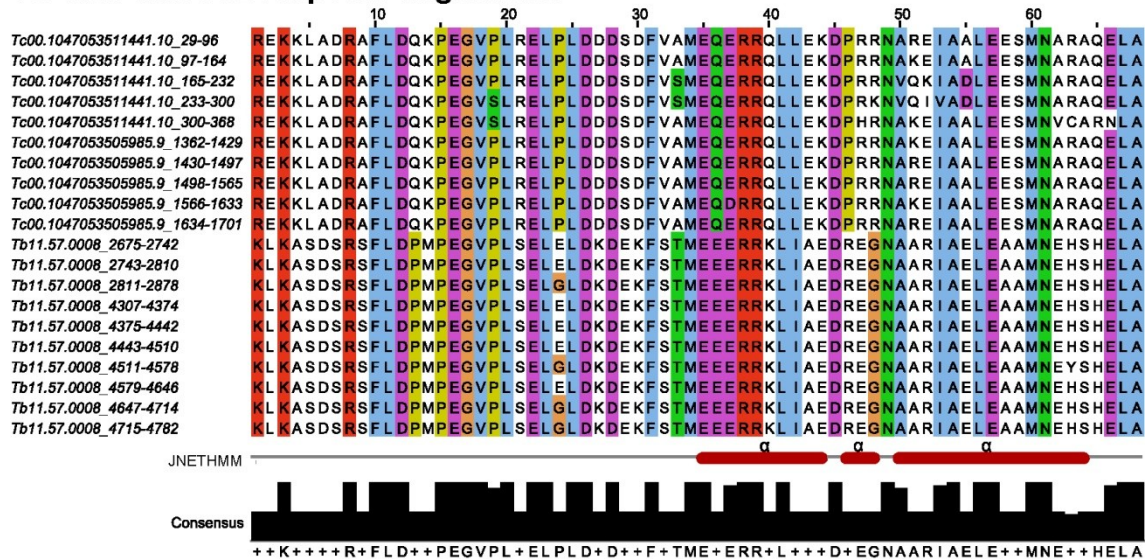
alignment of equal numbers of the most conserved FRA and GM6 repeat sequences of TcCALP985.9, TcCALP441.10 (FRA) and Tb11.57.0008 (GM6), by the MAFFT method (Kato *et al.*, 2002), (Fig. 3.13A). Alignment of FRA and GM6 repeats consensus sequences showed a 51% identity between them (Fig. 3.13B). These data is in agreement with the great similarity between FRA and GM6 secondary structures (Fig. 3.8 and 3.9) predicted by the Jnet method (Cuff and Barton, 2000). From the alignment of FRA and GM6 repeats (Fig. 3.13A), I also built a sequence logo (Fig. 3.13C), using the Weblogo software (Crooks *et al.*, 2004). The FRA-GM6 alignment and sequence logo illustrate the residue conservation between homologous repeats in *T. brucei* and *T. cruzi*, with the majority of conserved residues being either hydrophobic or acidic (Fig. 3.13B-C). The same analysis was performed for FRA (TcCALP985.9, TcCALP441.10) and Lcr1 repeats (LmjF27.0500, Fig. 3.14), and for GM6 (Tb11.57.0008) and Lcr1 repeats (Fig. 3.15). Additionally, I used the alignments of each individual kind of repeat to obtain their sequence logos, using the Weblogo software (Crooks *et al.*, 2004):

- The FRA repeats sequence logo, obtained from the alignment of TcCLP441.10 and TcCLP985.9 repeats (Fig. 3.16A).
- The GM6 repeats sequence logo, obtained from the alignment of Tb11.47.0036 and Tb11.57.0008 repeats (Fig. 3.16B).
- The Lcr1 repeats sequence logo, obtained from the alignment of LmjF27.0500 (Fig. 3.16C).
- The Lcr1-like repeats sequence logo, obtained from the alignment of Tb11.v1.0001 and Tb11.47.0035 repeats (Fig. 3.16D).

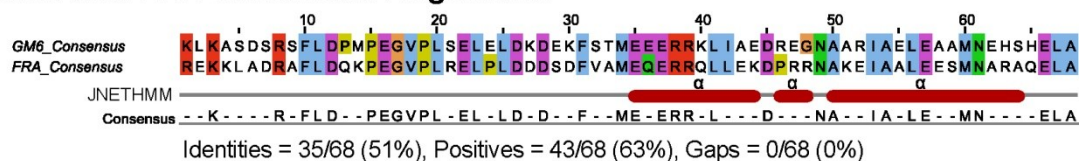
Comparison of these sequence logos showed a high degree of conservation between FRA and GM6 repeats (encoded by orthologous genes), with 37 out of 68 conserved amino acids (51% identity, 63% similarity). Lcr1 and Lcr1-like repeats (also encoded by orthologous genes) are highly divergent, both from each other and from the FRA/GM6

Figure 3.13.

A. GM6 and FRA repeats alignment.



B. GM6 and FRA consensus Alignment.



C. GM6-FRA Repeats Logo.

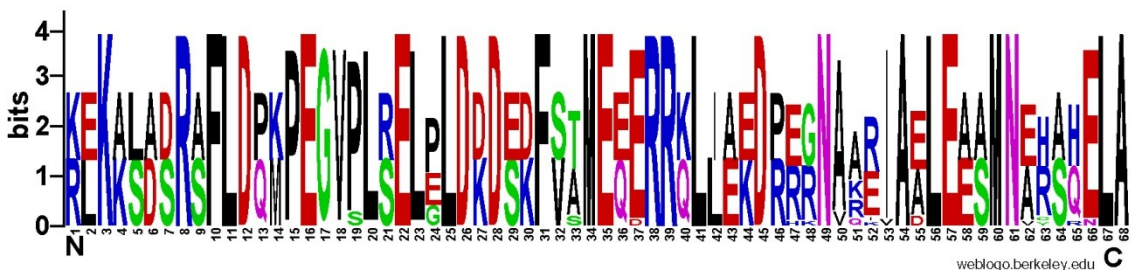
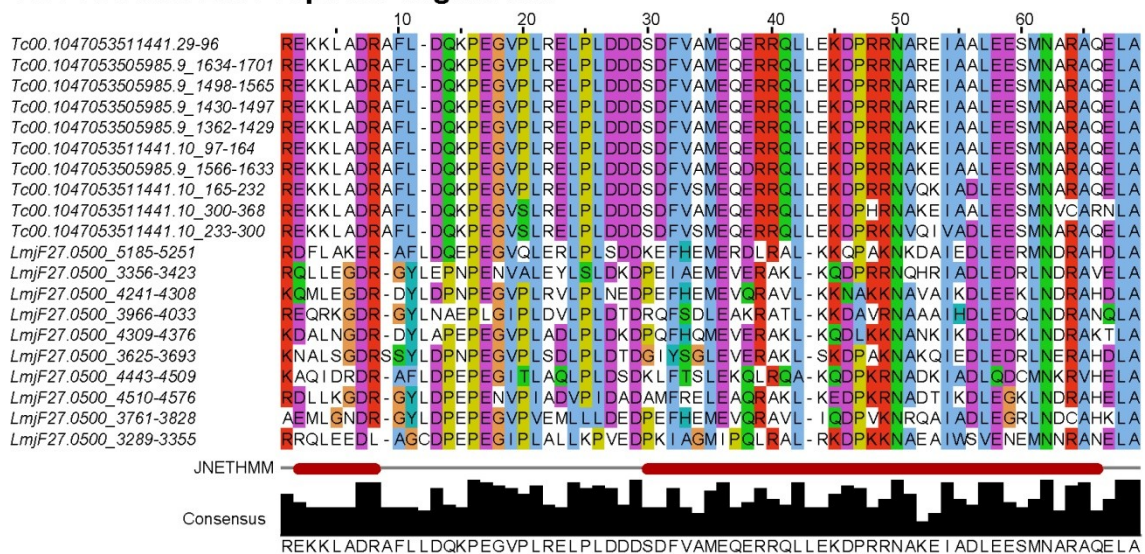


Figure 3.13.

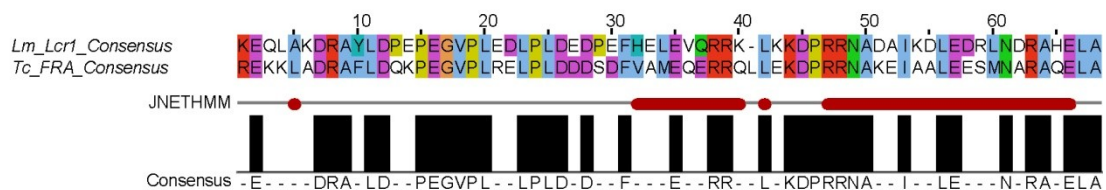
A. Multiple alignment by the MAFFT method of equal numbers of GM6 and FRA repeats units. Secondary structure prediction is by the JNet method, α -helices are indicated by red bars.
B. GM6 and FRA consensus sequences alignment by the MAFFT method.
C. Sequence logo of the aligned GM6 and FRA repeats (A). Amino acids are colored according to their chemical properties: polar amino acids -G,S,T,Y and C- are green, Q and N are fuchsia; basic -K,R,H- are blue, acidic -D,E- are red, and hydrophobic -A,V,L,I,P,W,F,M- are black. Note that the majority of conserved amino acids are hydrophobic and acidic.

Figure 3.14.

A. FRA and Lcr1 repeats alignment.



B. FRA and Lcr1 Consensus Alignment.



C. FRA-Lcr1 Repeats Logo.

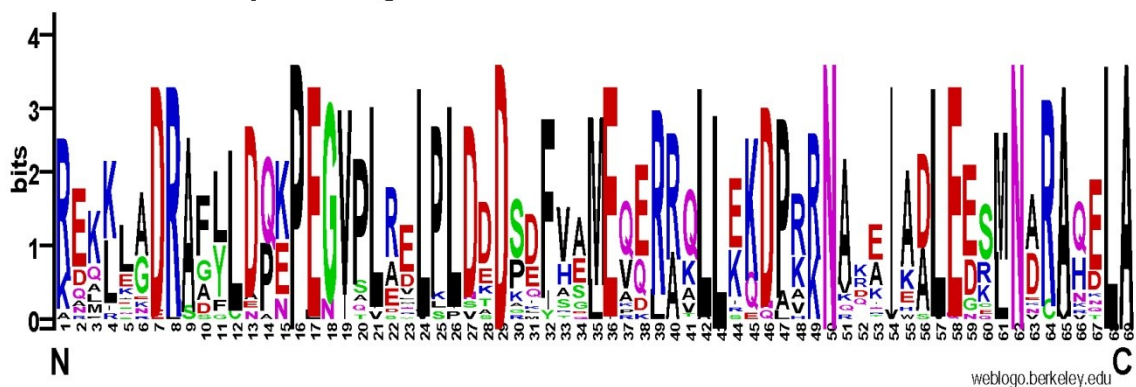


Figure 3.14.

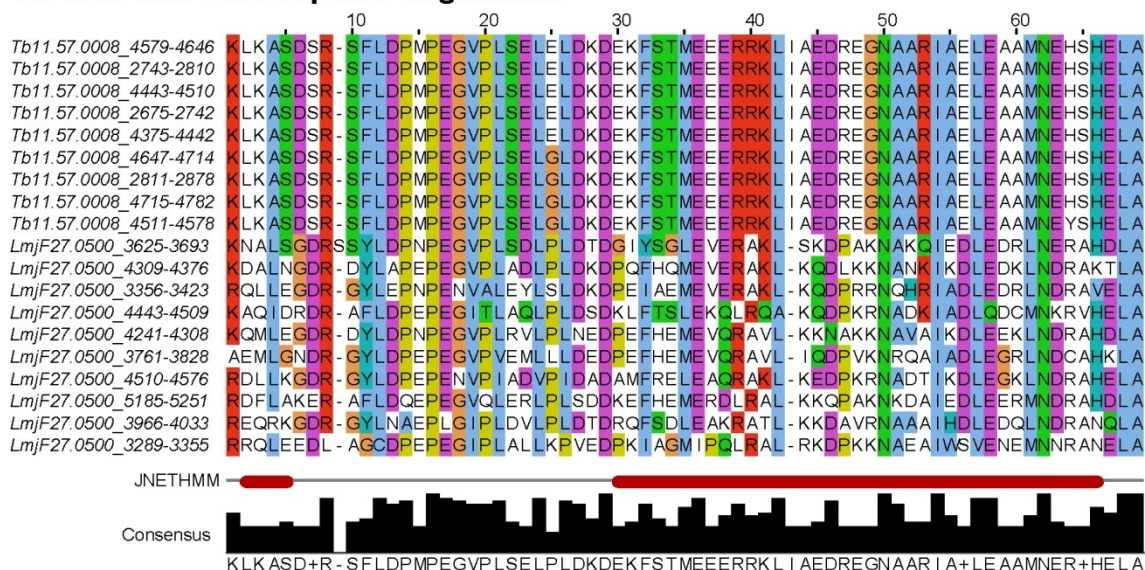
A. Multiple alignment by the MAFFT method of an equal number of the most conserved FRA and Lcr1 repeats.

B. FRA and Lcr1 consensus sequences alignment. Alignments colour scheme is according to the ClustalX colour code. Secondary structure prediction is by the JNet method, α -helices are indicated by red bars.

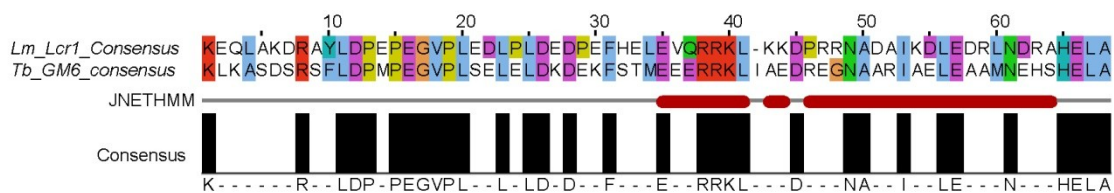
C. Sequence logo of the aligned FRA and Lcr1 repeats (A). Amino acids are coloured according to their chemical properties: polar amino acids -G,S,T,Y and C- are green, Q and N are fuchsia; basic -K,R,H- are blue, acidic -D,E- are red, and hydrophobic -A,V,L,I,P,W,F,M- are black. Note that the majority of conserved amino acids are hydrophobic, acidic and basic.

Figure 3.15.

A. GM6 and Lcr1 repeats alignment.



B. GM6 and Lcr1 Consensus Alignment.



C. GM6-Lcr1 Repeats Logo.

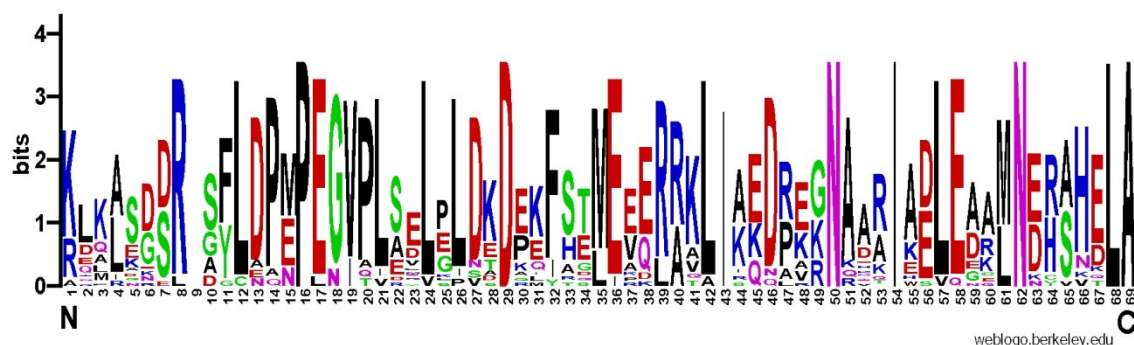


Figure 3.15.

A. Multiple alignment by the MAFFT method of equal numbers of the most conserved GM6 and Lcr1 repeats.

B. GM6 and Lcr1 consensus sequences alignment.

Alignment colour scheme is according to the ClustalX colour code. Secondary structure prediction is by the JNet method, α -helices are indicated by red bars.

C. Sequence logo of the aligned GM6 and Lcr1 repeats (A). Amino acids are coloured according to their chemical properties: polar amino acids -G,S,T,Y and C- are green, Q and N are fuchsia; basic -K,R,H- are blue, acidic -D,E- are red, and hydrophobic -A,V,L,I,P,W,F,M- are black. Note that the majority of conserved amino acids are hydrophobic, acidic and basic.

T. brucei GM6 aa Repeats Logo: Tb11.57.0008 and Tb11.47.0036

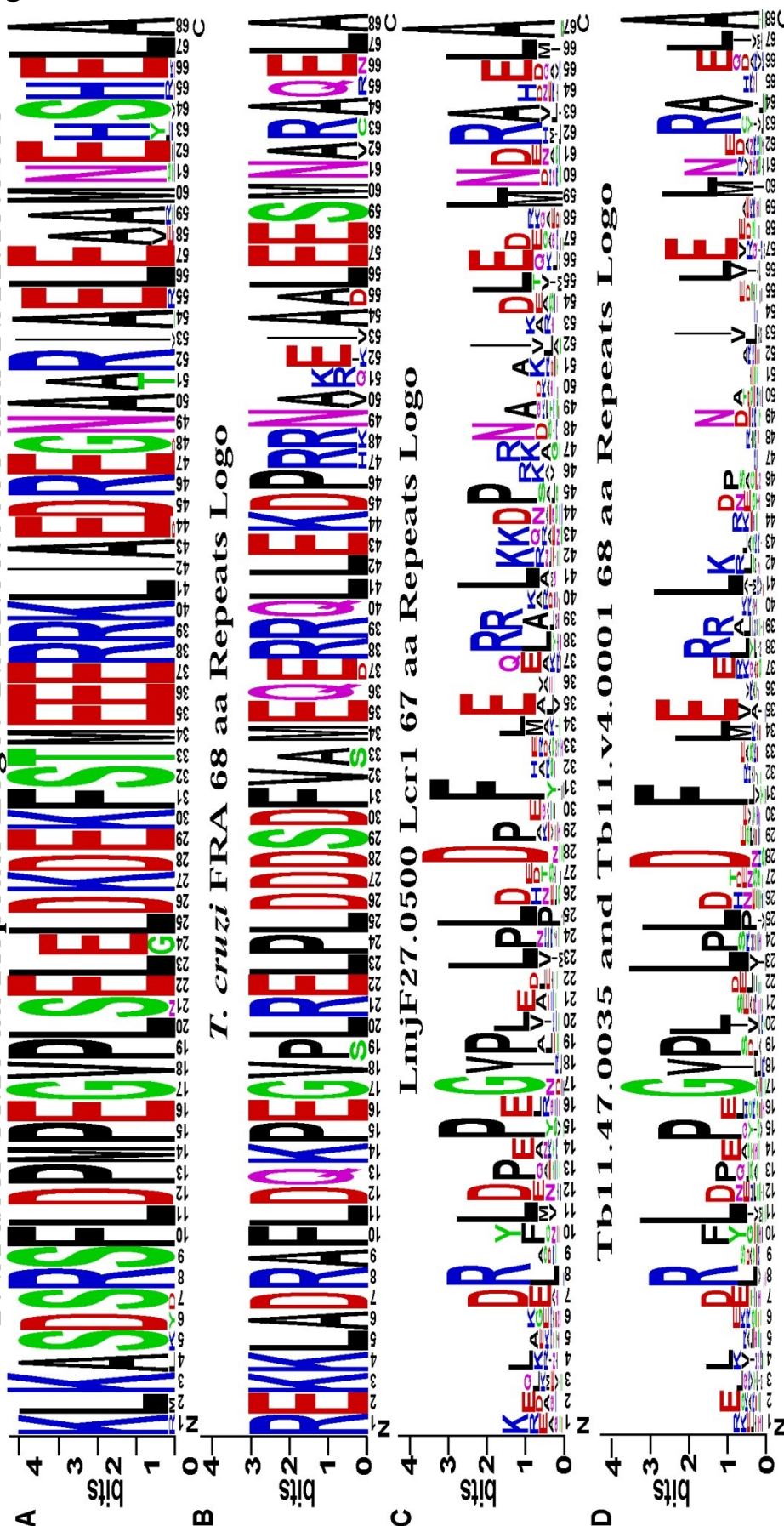


Figure. 3.16. Sequence logos of the 67-68 amino acid repeats of *T. brucei* GM6, *T. cruzi* FRA, *L. major* Lcr1 and *T. brucei* Lcr1 orthologs, Tb11.47.0035 and Tb11.47.0001 (Lcr1-like repeats). Note the high conservation of GM6 (A) and FRA (B) repeats and the high degree of similarity between FRA and GM6 repeats and between Lcr1 and Lcr1-like repeats.

repeats (Fig. 3.14 and 3.15). However, Lcr1 and Lcr1-like repeats contain the same conserved amino acids (38 out of 67/68 amino acid residues), which were also highly conserved in the GM6 repeats (Fig.3.16). The majority of the conserved residues between the different repeats are hydrophobic (26), acidic (10) and basic (7). In addition, two asparagines and one glycine are also highly conserved. These results indicate that the different repeats evolved from the same ancestor sequence and that the most conserved amino acids may be critical for these proteins function.

3.2.4. Identification of a large chromosome duplication event in *T. brucei* containing a CAP5.5 paralogous gene, Tb907.8.8330.

One of the newly identified calpain-like sequences in *T. brucei*, Tb927.8.8330, which contains domains I^K, II, III and C (Fig. 3.1 to 3.3) showed a 75% of amino acid identity to CAP5.5 (Fig. 3.17), the first calpain-like sequences described in kinetoplastids (*Hertz-Fowler et al., 2001*). Furthermore, when comparing only the first ~720 amino acid residues of both proteins, their percentage of identity raised to 84% (Fig. 3.17 and Appendix 3, Fig. A.3.9). Reciprocal BLAST within the *T. brucei* genome sequence identified CAP5.5 and Tb927.8.8330 genes as putative paralogs (Appendix 3, Reciprocal_BLAST). The high similarity between these two genes, located in different chromosomes, was both surprising and intriguing. Analysis of their genomic context revealed that they are located in sub-telomeric regions of their respective chromosomes, Tb927.4.3950 or CAP5.5 in chromosome 4 and its paralog, Tb927.8.83330, which I termed as ⁵CAP5.5V, in chromosome 8.

⁵ CAP5.5 (Tb927.4.3950; encoding cytoskeletal-associated protein 5.5) was previously named CALP1 by *Hertz-Fowler et al. (2001)*, and then revised to TbCALP4.1CAP5.5 by *Ersfeld et al (2005)*. However, in the publicly available *T. brucei* genome sequence the original name of CAP5.5 is retained. Applying the nomenclature suggested by *Ersfeld et al. (2005)* CAP5.5V (Tb927.8.8330) would formally be recognised as TbCALP8.1CAP5.5V.

Figure 1 Schematic representation of the domain organization of the *Tb* JNETHMM protein. The protein is composed of several domains: Domain I (yellow), Domain II (red), and Domain III (green). The JNETHMM protein is shown in black bars, and the consensus sequence is shown in white bars. The protein is 600 amino acids long. The domains are: Domain I (1-154), Domain II (155-368), Domain III (369-599), and C-Terminal Domain (600-693). The protein is shown in black bars, and the consensus sequence is shown in white bars. The protein is 600 amino acids long. The domains are: Domain I (1-154), Domain II (155-368), Domain III (369-599), and C-Terminal Domain (600-693).

115

Figure 3.17. Page 2 of 2.

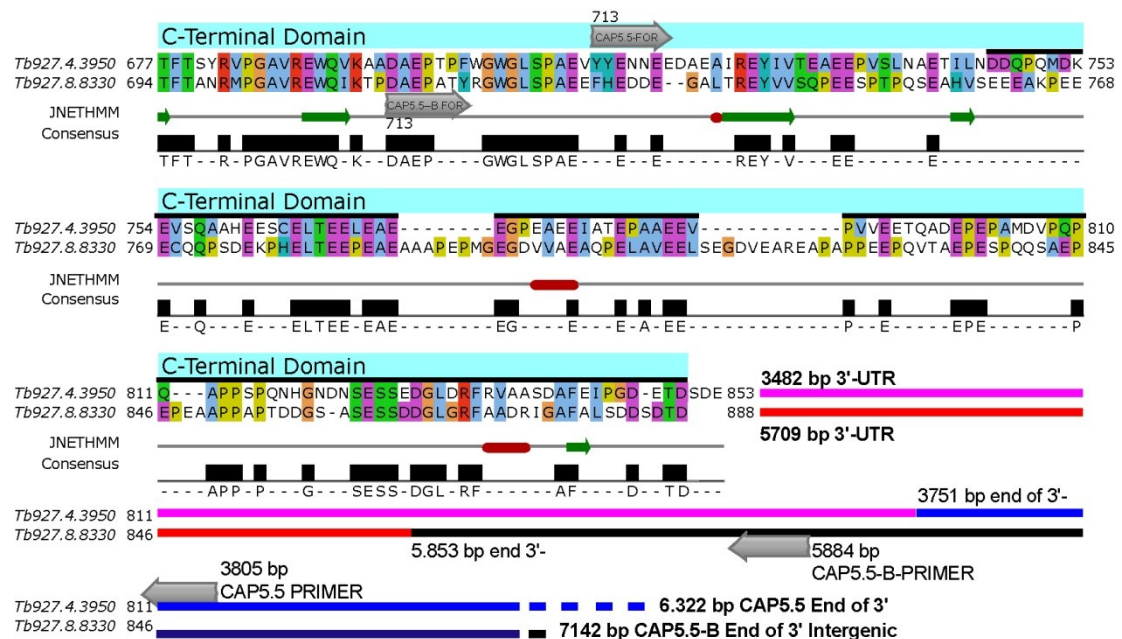


Figure 3.17 Page 2 of 2. Individual domains are marked by different colour bars. Domain I (yellow bar), characteristic motifs are indicated by blue bars underlining the consensus sequence. Domain II (red bar): potential active-site residues are denoted above the sequences by their equivalent residues in conventional calpains; other characteristic residues of calpain domains II are also marked above the sequences and underlined when conserved, the KAYAK-motif is indicated by a black bar; the presence and absence of the five amino acid residues essential for the binding of two Ca^{+2} ions within domain II are denoted by plus (+) and minus (-) symbols, respectively. Domain III (green bar) contains eight β -strand predicted conformations. In the C-terminal domain (light blue bar), the extension in CAP5.5, originally noted by *Hertz-Fowler et al. (2001)* is marked by a black bar above the sequence. Equivalent positions of the nucleotide primers are represented by arrows: black for the CAP5.5/CAP5.5V N-terminus RNAi construct, grey for CAP5.5 and blue for CAP5.5V C-terminus RNAi constructs.

Comparison of chromosomes 4 and 8 using the DOTTER program (*Sonnhammer and Durbin, 1995*) showed a high degree of similarity across, at least, 0.2 Mb of their sub-telomeric regions (Appendix 3, DOTTER 4+8 file). This high similarity indicated that the two sub-telomeric regions could represent a large chromosome duplication event, the first to be found in trypanosomes. My results were soon confirmed by the publication of the discovery of a 0.5 Mb duplication event in *T. brucei* comprising the 3' regions of chromosomes 4 and 8 (*Jackson, 2007b*). Gene duplication is one of the main mechanisms of genomic evolution (*Taylor and Raes, 2005; Van der Peer and Meyer,*

2005). In the process through which gene duplication influences evolution, sensible changes in structure correlate with changes in function by altering the coding sequences (CDS) or the non-coding, regulatory sequences (NCS). The final outcome of gene duplication is likely to reflect both selective pressures (adaptation) and the relevance of the original gene —its original function, indispensability and connectivity (Roth *et al.*, 2007). In this context, the identified 0.5 Mb duplicons in *T. brucei* chromosomes 4 and 8 were compared with homologous regions in the genome of different trypanosomatids species, showing widespread colinearity of gene order within and between species (Jackson, 2007b). Comparison with homologous regions in related species also showed that the duplicated region of *T. brucei* corresponds to the entire *Leishmania* chromosome 31. The absence of duplicons in *T. congolense* showed that the duplication event was unique to *T. brucei* and occurred prior to speciation (Jackson, 2007b), about 250 Mya (Douzery *et al.*, 2004). Additionally, PCR assays have shown that these chromosome duplication was present in different *T. brucei* subspecies (*T. b. gambiense*, *T. b. rhodiense* and *T. evansi*), confirming that the duplication was common to all members of the *T. brucei* clade (Jackson, 2007b).

Comparison between duplicons and an outgroup (*T. congolense*) allowed the classification of the duplicated genes as shared (present in both duplicons and an outgroup), lost (present in on one duplicon and an outgroup) or gained (not present in an outgroup). This classification rendered 74 loci shared by both duplicons and an outgroup, 27 and 58 genes lost by chromosome 4 and 8, respectively; 7 loci gained by chromosome 4 and 18 loci gained by Chromosome 8. Therefore, 47% of all duplicated loci were retained as paralogs on both duplicons, a surprisingly large number for this type of event. Furthermore, gene loss was selective, as shown by the over-representation of genes with low “connectivity” within the cell (i.e. surface expressed genes, transmembrane helices and signal peptides). This is consistent with a dosage

balance hypothesis, in which low connectivity genes are more likely to be preserved due to their dose-insensitivity (*Jackson, 2007b*). Additionally, sequence divergence of conserved paralogs followed different evolutionary trajectories, often with radical structural changes among their coding and/or regulatory regions, strongly indicating that many of these divergences are involved functional changes (*Jackson, 2007b*).

Among the 74 loci shared by both duplicons is the pair constituted by CAP5.5, on chromosome 4, and its paralog CAP5.5V, on chromosome 8 (*Jackson, 2007b*). During this thesis work, sequence conservation and divergence patterns between CAP5.5 and CAP5.5V genes were analysed by comparison of their coding and non-coding regions. CAP5.5 and CAP5.5V coding and intergenic regions were obtained from the version 4 update of the *T. brucei* genome project (*Hertz-Fowler et al., 2004*) in GeneDB (<http://www.genedb.org/tryp/>). Untranslated regulatory regions of both genes were identified manually, analysing the untranslated regions for the presence of the sequence motifs of the spliced leader and poly-A tail additions, as previously described (*Benz et al., 2005*). Visual comparison of CAP5.5 and CAP5.5V intergenic regions showed great differences in size between the 5'- and 3'-NCS (Fig. 3.18). Whereas CAP5.5 has a short 5'-intergenic region (919 bp), CAP5.5V has a large one (3041 bp). The opposite is true for their 3'-intergenic regions, CAP5.5 having a large one (2840 bp) and CAP5.5V a short one (1439 bp). Alignment of both 5'- and 3'-intergenic regions using a global alignment algorithm (*Gotoh, 1982*) showed little identity between them. The 5'-intergenic regions showed a 17.5% overall identity and a 51.6% identity over the last 1000bp of CAP5.5V 5' intergenic region (Appendix 3, Fig. A.3.10). Similarly, the 3- intergenic regions also showed high diversity between them, with a 10.4% overall identity and a 50.1% identity over the first 744 bp of CAP5.5 and the last 709 bp of CAP5.5V 3' intergenic regions (Appendix 3, Fig. A.3.11). Coding and UTR sequences of both genes were compared by translation alignment using the MUSCLE

Figure 3.18.

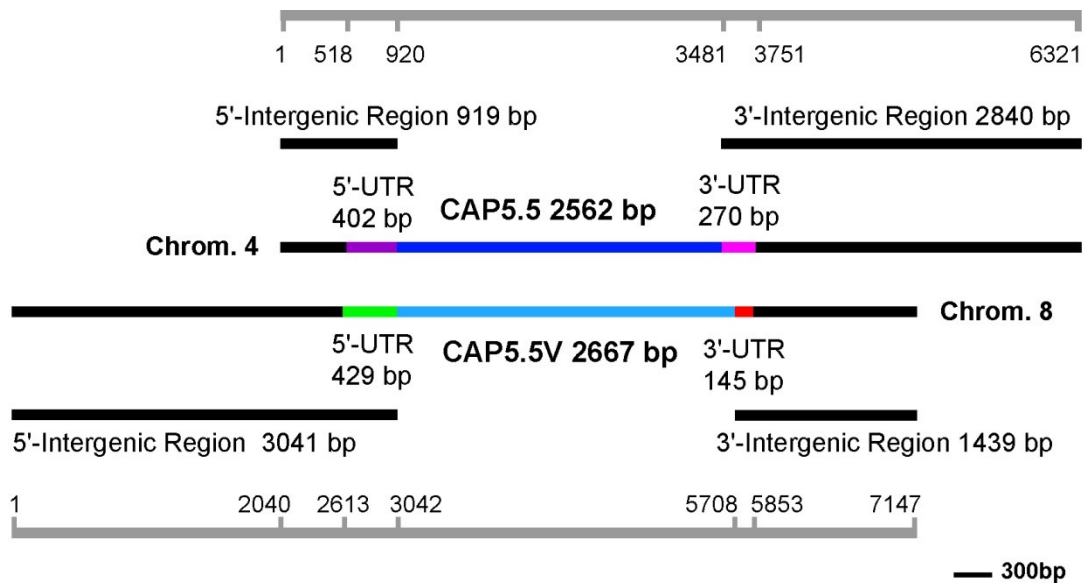


Figure. 3.18. Schematic representation of the genome locus of CAP5.5 and CAP5.5V on chromosome 4 and 8 respectively. CAP5.5 CDS is represented by a dark blue bar and CAP5.5V CDS by a light blue bar. Intergenic regions are represented by black bars. 5' and 3' UTRs of each gene are represented by bars of different colours to highlight their divergences: CAP5.5 5' UTR and 3' UTR are represented by purple and fuchsia bars, respectively; CAP5.5V 5' UTR and 3' UTR are represented by green and red bars, respectively. Gray rulers are placed above and below CAP5.5 and CAP5.5V genome locus. Note the different sizes between the NCS and particularly between the 3' regulatory regions. Scale bar represents 300 bp.

alignment algorithm (*Edgar, 2004*). Divergence between the UTRs was greater than between the CDS. CAP5.5 and CAP5.5V 5'- and 3'-UTRs are very short and highly divergent, showing only 35% and 25% identity, respectively (Appendix 3, Fig. A.3.12 and A.3.13). These values are similar to those obtained by Jackson (*Jackson, 2007b*).

CAP5.5 and CAP5.5V nucleotide coding and regulatory sequences were compared using a global alignment algorithm with open gaps (*Gotoh, 1982*), showing a 65.5% nucleotide pairwise identity (Fig. 3.19). Comparison of CAP5.5 and CAP5.5V showed a high degree of conservation of their CDS, with 74.8% nucleotide identity, and highly divergent 5'- and 3'-UTRs. Indeed, their UTRs were unalignable, with only 29.8% and 25.8% identity between their 5'- and 3'-UTRs, respectively, very low values considering

Figure 3.19. Page 1 of 2.



Figure 3.19. Page 1 of 2. Nucleotide alignment of CAP5.5 and CAP5.5V coding and regulatory regions (CDSs and UTRs) using a global alignment with open gaps algorithm (Gotoh, 1982). Overall identity 65.5%. Score = 9012.5, Identities = 3404/8979 (37%), Positives = 3422/8979 (38%), Gaps = 4490/8979 (50%). The start of CAP5.5 and CAP5.5V 5' UTRs are indicated by black arrows. ORF are represented by yellow bars above CAP5.5 and bellow CAP5.5V. Primers used for the CAP5.5/CAP5.5V RNAi constructs are indicated by dark grey arrows

Figure 1 displays the genomic organization and sequence of the *Cap5* gene, showing the 5' and 3' UTRs, exons, and introns. The gene structure is shown as a series of boxes representing exons and lines representing introns. The sequence is provided for each exon and intron, with the 5' and 3' UTRs indicated. The sequence is color-coded to highlight specific motifs and features. The 5' UTR contains a 5' cap (m7G) and a 5' splice site. The 3' UTR contains a 3' splice site and a poly(A) tail. The exons are numbered 1 through 10. The introns are numbered 1 through 9. The sequence is shown in the 5' to 3' direction. The 5' UTR is 2381 bp long. The 3' UTR is 4554 bp long. The exons are 2461, 2463, 2541, 2543, 2621, 2623, 2701, 2703, 2781, 2783, 2861, 2863, 2941, 2943, 3021, 3023, 3101, 3103, 3181, 3183, 3261, 3263, 3341, 3343, 3421, 3423, 3501, 3503, 3581, 3583, 3661, 3663, 3741, 3743, 3821, 3823, 3901, 3903, 3981, 3983, 4061, 4063, 4141, 4143, 4221, 4223, 4301, 4303, 4381, 4383, 4461, 4463, 4541, 4543, 4621, 4623, 4701, 4703, 4781, 4783, 4861, 4863, 4941, 4943, 5021, 5023, 5101, 5103, 5181, 5183, 5261, 5263, 5341, 5343, 5421, 5423, 5501, 5503, 5581, 5583, 5661, 5663, 5741, 5743, 5821, 5823, 5901, 5903, 5981, 5983, 6061, 6063, 6141, 6143, 6221, 6223, 6301, 6303, 6381, 6383, 6461, 6463, 6541, 6543, 6621, 6623, 6701, 6703, 6781, 6783, 6861, 6863, 6941, 6943, 7021, 7023, 7101, 7103, 7181, 7183, 7261, 7263, 7341, 7343, 7421, 7423, 7501, 7503, 7581, 7583, 7661, 7663, 7741, 7743, 7821, 7823, 7901, 7903, 7981, 7983, 8061, 8063, 8141, 8143, 8221, 8223, 8301, 8303, 8381, 8383, 8461, 8463, 8541, 8543, 8621, 8623, 8701, 8703, 8781, 8783, 8861, 8863, 8941, 8943, 9021, 9023, 9101, 9103, 9181, 9183, 9261, 9263, 9341, 9343, 9421, 9423, 9501, 9503, 9581, 9583, 9661, 9663, 9741, 9743, 9821, 9823, 9901, 9903, 9981, 9983, 10061, 10063, 10141, 10143, 10221, 10223, 10301, 10303, 10381, 10383, 10461, 10463, 10541, 10543, 10621, 10623, 10701, 10703, 10781, 10783, 10861, 10863, 10941, 10943, 11021, 11023, 11101, 11103, 11181, 11183, 11261, 11263, 11341, 11343, 11421, 11423, 11501, 11503, 11581, 11583, 11661, 11663, 11741, 11743, 11821, 11823, 11901, 11903, 11981, 11983, 12061, 12063, 12141, 12143, 12221, 12223, 12301, 12303, 12381, 12383, 12461, 12463, 12541, 12543, 12621, 12623, 12701, 12703, 12781, 12783, 12861, 12863, 12941, 12943, 13021, 13023, 13101, 13103, 13181, 13183, 13261, 13263, 13341, 13343, 13421, 13423, 13501, 13503, 13581, 13583, 13661, 13663, 13741, 13743, 13821, 13823, 13901, 13903, 13981, 13983, 14061, 14063, 14141, 14143, 14221, 14223, 14301, 14303, 14381, 14383, 14461, 14463, 14541, 14543, 14621, 14623, 14701, 14703, 14781, 14783, 14861, 14863, 14941, 14943, 15021, 15023, 15101, 15103, 15181, 15183, 15261, 15263, 15341, 15343, 15421, 15423, 15501, 15503, 15581, 15583, 15661, 15663, 15741, 15743, 15821, 15823, 15901, 15903, 15981, 15983, 16061, 16063, 16141, 16143, 16221, 16223, 16301, 16303, 16381, 16383, 16461, 16463, 16541, 16543, 16621, 16623, 16701, 16703, 16781, 16783, 16861, 16863, 16941, 16943, 17021, 17023, 17101, 17103, 17181, 17183, 17261, 17263, 17341, 17343, 17421, 17423, 17501, 17503, 17581, 17583, 17661, 17663, 17741, 17743, 17821, 17823, 17901, 17903, 17981, 17983, 18061, 18063, 18141, 18143, 18221, 18223, 18301, 18303, 18381, 18383, 18461, 18463, 18541, 18543, 18621, 18623, 18701, 18703, 18781, 18783, 18861, 18863, 18941, 18943, 19021, 19023, 19101, 19103, 19181, 19183, 19261, 19263, 19341, 19343, 19421, 19423, 19501, 19503, 19581, 19583, 19661, 19663, 19741, 19743, 19821, 19823, 19901, 19903, 19981, 19983, 20061, 20063, 20141, 20143, 20221, 20223, 20301, 20303, 20381, 20383, 20461, 20463, 20541, 20543, 20621, 20623, 20701, 20703, 20781, 20783, 20861, 20863, 20941, 20943, 21021, 21023, 21101, 21103, 21181, 21183, 21261, 21263, 21341, 21343, 21421, 21423, 21501, 21503, 21581, 21583, 21661, 21663, 21741, 21743, 21821, 21823, 21901, 21903, 21981, 21983, 22061, 22063, 22141, 22143, 22221, 22223, 22301, 22303, 22381, 22383, 22461, 22463, 22541, 22543, 22621, 22623, 22701, 22703, 22781, 22783, 22861, 22863, 22941, 22943, 23021, 23023, 23101, 23103, 23181, 23183, 23261, 23263, 23341, 23343, 23421, 23423, 23501, 23503, 23581, 23583, 23661, 23663, 23741, 23743, 23821, 23823, 23901, 23903, 23981, 23983, 24061, 24063, 24141, 24143, 24221, 24223, 24301, 24303, 24381, 24383, 24461, 24463, 24541, 24543, 24621, 24623, 24701, 24703, 24781, 24783, 24861, 24863, 24941, 24943, 25021, 25023, 25101, 25103, 25181, 25183, 25261, 25263, 25341, 25343, 25421, 25423, 25501, 25503, 25581, 25583, 25661, 25

that 25% nucleotide identity is the value expected by chance between two unrelated sequences (*Jackson, 2007b*). The only exception in these dissimilar UTRs was found in CAP5.5V, which seems to have evolved a new stop codon, making the beginning of its 3'-UTR briefly align with the 3'-terminus of CAP5.5 CDS (Fig. 3.19). Additionally, relative rate test analysis (*Tajima, 1993*), have shown that the two genes evolved at a similar pace after duplication (*Jackson, 2007b*). Likelihood algorithms, used to assess non-neutral evolution in the selection of individual codons, showed that both paralogs have a high number of positive (27) and negative (41) selected codons (*Jackson, 2007b*). Thus, the sequence divergence between both paralogs might be the consequence of neo or subfunctionalisation, that is, they may have different functions or different expression profiles. In summary, the data indicate that these genes have evolved without much structural change, perhaps resulting in a novel expression profile (*Jackson, 2007b*), as has been reported for other genes in different organisms (*Hua et al., 2003; Bhushan et al., 2005*). However, the concentration of non-synonymous substitutions at the 3' terminus of their CDS also point to the possibility of functional specialisation.

The analysis of CAP5.5 and CAP5.5V amino acid alignment allowed the identification of interesting features (Fig. 3.17). To begin with, both CAP5.5 and CAP5.5V contain an N-terminal dual acylation modification motif, formed by a subterminal glycine followed by a cysteine, previously described in CAP5.5 (*Hertz-Fowler et al., 2001; Ersfeld et al., 2005*). Both glycine and cysteine are necessary but not sufficient for acylation to occur. The enzymes that catalyse the modification by fatty acids require an extended sequence motif (*Maurer-Stroh et al., 2002a*). Specifically, the first six N-terminal amino acids have to fit deep into the N-myristoyltransferases (NMTs) binding pocket, and are therefore restricted by volume, backbone flexibility and other side chain properties. NMTs have different substrate specificity and their acylation modification motifs vary

between taxons (*Maurer-Stroh et al., 2002b*). The N-terminal double acylation motif found in CAP5.5 and completely conserved in CAP5.5V —MGCGGS— possesses all the required characteristics to fit into the NMTs pocket. Furthermore, CAP5.5 has been experimentally shown to be double acylated, both myristoylated and palmitoylated (*Hertz-Fowler et al., 2001*). N-terminal myristoylation of a protein increases its hydrophobicity, targeting the protein to hydrophobic lipid domains within the cell and thus triggering membrane-protein association. Myristoylation is also important because is a prerequisite for subsequent palmitoylation, which then establishes a high affinity interaction with membranes (*Resh, 1999*). Dual acylation with myristic and palmitic acid is exclusively found in membrane-associated proteins (*Resh, 1999*), and therefore strongly suggest that CAP5.5 and CAP5.5V interact with membranes. However, it is also known that CAP5.5 and CAP5.5V completely sequester into the Triton X-100 insoluble fraction in cellular fractioning experiments (*Hertz-Fowler et al., 2001; Liu et al., 2009*), indicating that both proteins interact strongly with the subpellicular microtubule cytoskeleton. CAP5.5 and CAP5.5V have acidic statistically predicted pI, 4.1 and 4.2 respectively, which not only are characteristic of microtubule-associated proteins (*Kreis and Vale, 1999*), but also rule out the possibility for CAP5.5 and CAP5.5V of being basic proteins, non-specifically attached to the exposed acidic projections of microtubules (*Hertz-Fowler et al., 2001*).

Both CAP5.5 and CAP5.5V N-terminus sequences contain a proline-rich region similar to the consensus motif found in the binding sites of many ligands of SH3-type domains —PXXP— (*Ren et al., 1993; Yu et al., 1994*). SH3-type domains are found in a wide range of intracellular and membrane associated proteins, such as proteins with enzymatic activity, adaptor proteins that lack catalytic sequences and in cytoskeletal associated proteins —e.g. fodrin and yeast Actin Binding Protein 1, ABP-1 (*Musacchio et al., 1992; Mayer and Baltimore, 1993; Pawson, 1995*). The presence of this proline-

rich motif could be important in the search for CAP5.5 and CAP5.5V binding partners. CAP5.5V contains a much larger proline-rich sequence than CAP5.5, with 23 proline residues out of a 30 amino acid long sequence (9 out of 11 residues in CAP5.5). The proline rich sequences are flanked by conserved regions and could represent binding partner specificities between both proteins (Fig. 3.17).

The catalytic domains of conventional calpains are able to bind two Ca^{2+} ions in a coordinated mechanism, which is essential to align the catalytic residues within the active site (Moldoveanu *et al.*, 2002; Moldoveanu *et al.*, 2004). In mammals, five amino acids have been shown to be critical for calpain domain II Ca^{2+} binding. Analysis of CAP5.5 and CAP5.5V domain II sequences revealed the conservation in both proteins of three of the five amino acid residues critical for Ca^{2+} binding in mammalian calpain catalytic domains. The conservation of these residues suggests the possibility that CAP5.5 and CAP5.5V could also bind calcium (Fig. 3.17). Additionally, CAP5.5 and CAP5.5V central domains have the same deviations from the conventional calpain catalytic triad. In the two sequences, the conventional triad —C-H-N— triad is doubly substituted, rendering a S-Y-N triad, of unknown proteolytic activity. Several other residues and motifs characteristic of conventional calpain domains II are also conserved in CAP5.5 and CAP5.5V. Noticeably, some of these conserved residues are those that typically surround the amino acids of the catalytic triad. In example, the GD and W residues that precede and follow the active C, respectively, are conserved in both paralogs, even though the C residue has been substituted by S (Fig. 3.17). Secondary structure analysis for CAP5.5 and CAP5.5V domain III by the JNet method (Cuff and Barton, 2000) predicted eight β -strand conformations along both sequences, which is consistent with the presence of an eight-stranded antiparallel β -sandwich in domain III of typical calpains (Fig. 3.17). CAP5.5 and CAP5.5V C-terminal domains lack the EF-hand motifs characteristic of conventional calpains and are the most

divergent parts of the proteins. After the first 40 residues of their domain IV, CAP5.5 and CAP5.5V C-terminal sequences are very divergent, mainly due to the presence in CAP5.5V of residues not shared by CAP5.5, that alternate with more conserved sections (Fig.3.17).

3.3. Reclassification of kinetoplastid calpain-like sequences.

After the analysis and characterisation the newly found kinetoplastid calpain-like sequences, it was obvious that some of these sequences could not be included in any of the classification groups defined by Ersfeld *et al.*, 2005. Thus, I redefined their classification⁶ of kinetoplastid calpain-like proteins, also according to the proteins' domain composition and architecture, as follows (Fig. 3.20):

- **Group 1: I^K-II-III-C.**

The group 1 of kinetoplastid calpain-like proteins was defined by Ersfeld and co-workers as that including sequences which domain composition and architecture most closely resembles those of conventional calpains (Ersfeld *et al.*, 2005). This is, sequences constituted by a domain I^K —often containing N-terminal palmitoylation and/or myristoylation motifs— followed by a calpain “catalytic” domain, a calpain domain III homolog and a C-terminal domain. The definition of group 1 was not modified here, although I have included in it some of the new (4) and old (1) sequences that do not exactly match this group domain architecture:

T. brucei Tb11.01.6870; *L. major* LmjF32.2010; and *L. infantum* LmjF32_v3.2140.

These newly identified sequences only contain a calpain-like domain I^K and a

⁶ Ersfeld and co-workers classification and nomenclature of kinetoplastid calpain-like proteins is shown in Table 3.2, next to the reclassification and nomenclature presented and used in this thesis.

Figure 3.20.

Classification of kinetoplastid calpain-like proteins according to their domain architecture:

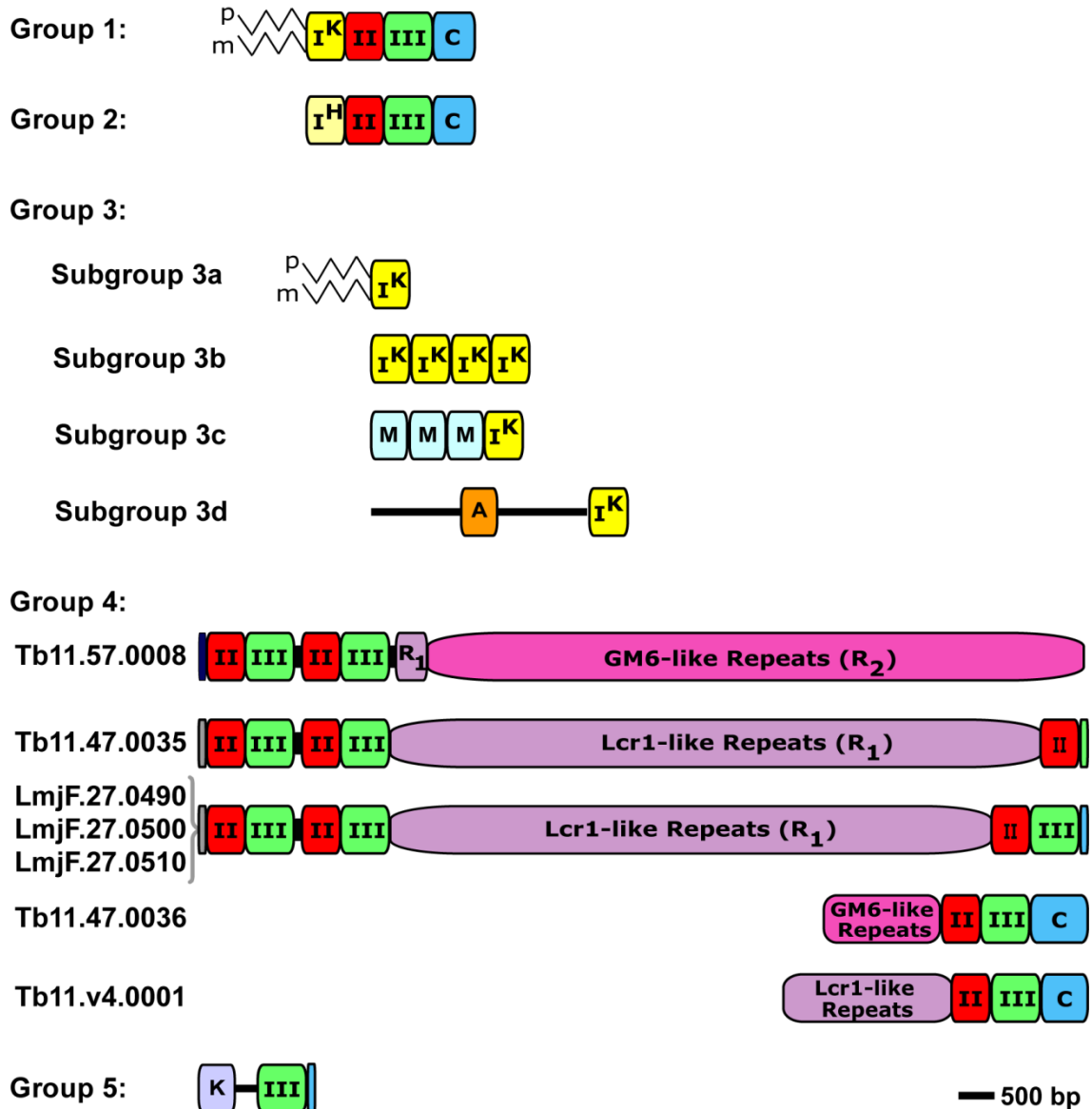


Figure 3.20. Classification of trypanosomatid calpain-like proteins according to their domain architecture and composition.

I^K: Short sequence element present at the N-terminus of trypanosomatid calpain-like sequences or as short stand alone ORFs.

I^H: Heterogeneous domain I.

M: META domain.

A: Adenylate cyclase catalytic domain.

K: Kinesin Motor domain.

Scale bar represents 500 bp

Table 3.2. Page 1 of 2.

Ersfeld et al. Classification					Reclassification	
<i>T. burecei</i>					<i>T. brucei</i>	
Name	geneDB ID	Domains	G	geneDB ID/Name	Domains	G
TbCALP1.1	Tb927.1.2100	I ^K , II, III, C	1	Tb927.1.2100	I ^K , II, III, C	1
TbCALP1.2	Tb927.1.2110	I ^K , II, III, C	1	Tb927.1.2110	I ^K , II, III, C	1
TbCALP1.3	Tb927.1.2120	I ^K , II, III, C	1	Tb927.1.2120	I ^K , II, III, C	1
TbSKCRP1.4	Tb927.1.2150	I ^K	3	Tb927.1.2150	I ^K	3, Sg. 3a
TbSKCRP1.5	Tb927.1.2160	I ^K	3	Tb927.1.2160	I ^K	3, Sbg. 3a
TbSKCRP1.6	Tb927.1.2230	I ^K	3	Tb927.1.2230/SMP-1a	I ^K	3, Sg. 3a
TbSKCRP1.7	Tb927.1.2260	I ^K	3	Tb927.1.2260/SMP1b	I ^K	3, Sg. 3a
TbCALP4.1	Tb04.ID20.740	I ^K , II, III, C	1	Tb927.4.3950/CAP5.5	I ^K , II, III, C	1
TbCAP4.2	Tb04.ID20.770	I ^K , II, III, C	1	Tb927.4.3940	I ^K , II, III, C	1
				Tb927.5.2190	META-I ^K	3, Sg 3c
				Tb927.5.2220	META-I ^K	3, Sg 3c
				Tb927.5.2250	META-I ^K	3, Sg 3c
				Tb927.5.2410	Kines-III-C	5
TbCAP6.1	Tb06.2N9.670	I ^H , II, III, C	2	Tb927.6.3310	I ^H , II, III, C	2
TbSKCRP7.1	Tb07.5F10.560	I ^K	3	Tb927.7..4060	I ^K	3, Sg 3a
				Tb927.7.4070	I ^K	3, Sg 3a
				Tb927.8.8330/CAP5.5V	I ^K , II, III, C	1
TbCALP9.1	Tb09.160.5550	I ^K , II, III, C	1	Tb09.160.5550	I ^K , II, III, C	1
TbCALP10.1	Tb10.389.0470	I ^H , II, III, C	2	Tb10.389.0470	I ^H , II, III, C	2
TbCALP10.2	Tb10.70.5950	I ^H , II, III, C	2	Tb10.70.5950	I ^H , II, III, C	2
TbSKCRP10.3	Tb10.70.5060			Tb10.70.5060	4xI ^K	3, Sg 3b
				Tb11.01.5800	I ^H , II, III, C	2
				Tb11.01.6870	I ^K -II only	1
TbCALP11.3	Tb11.02.2240	I ^H , II, III, C	2	Tb11.02.2240	I ^H , II, III, C	2
TbCALP11.1	Tb11.47.0035	3xII-2xIII-C	4	Tb11.47.0035	II-III-II-III-R ₁ -II	4
TbCALP11.2	Tb11.47.0036	I-R-II-III-C	5	Tb11.47.0036	R ₂ -II-III-C	4
				Tb11.v4.0001	R ₁ -II-III-C	4
				Tb11.57.0008	II-III-II-III-R ₁ -R ₂	4
<i>L. major</i>				<i>L. major</i>		
LmCALP4.1	LmjF04.0450	I ^K , II, III, C	1	LmjF04.0450	I ^K , II, III, C	1
LmSKCRP14.1	LmjF14.0850	I ^K	3	LmjF14.0850	I ^K	3, Sg 3a
LmSKCRP17.1	LmjF17.0870	I ^K		LmjF17.0870	I ^K	3, Sg 3a
				LmjF17.1110	Kines-III-C	5
LmCALP18.1	LmjF18.1060	I ^H , II, III, C	2	LmjF18.1060	I ^H , II, III, C	2
LmCALP20.1	LmjF20.1180	I ^K , II, III, C		LmjF20.1180	I ^K , II, III, C	1
				LmjF20.1185	I ^K , II, III, C	1
LmCALP20.2	LmjF20.1190	I ^K , II, III, C	1	LmjF20.1190	I ^K , II, III, C	1
LmCALP20.3	LmjF20.1200	I ^K , II, III, C	1	LmjF20.1200	I ^K , II, III, C	1
LmCALP20.4	LmjF20.1210	I ^K , II, III, C	1	LmjF20.1210	II, III only	1
LmSKCRP20.5	LmjF20.1220	I ^K	3	LmjF20.1220	I ^K	3, Sg 3a
LmSKCRP20.6	LmjF20.1230	I ^K	3	LmjF20.1230	I ^K	3, Sg 3a
LmSKCRP20.7	LmjF20.1250	I ^K	3	LmjF20.1250	I ^K	3, Sg 3a
LmSKCRP20.8	LmjF20.1280	I ^K	3	LmjF20.1280/SMP-3	I ^K	3, Sg 3a
LmSKCRP20.9	LmjF20.1300	I ^K	3	LmjF20.1300/SMP-2	I ^K	3, Sg 3a
LmSKCRP20.10	LmjF20.1310	I ^K	3	LmjF20.1310/SMP-1	I ^K	3, Sg 3a
				LmjF21.0120	I ^H , II, III, C	2
LmCALP25.1	LmjF25.1480	I ^K , II, III, C	1	LmjF25.1480	I ^K , II, III, C	1
LmSKCRP25.2	LmjF25.1750	I ^K	3	LmjF25.1750	I ^K	3, Sg 3a
LmCALP27.1	LmjF27.0490	2xII-2xIII-R-II-III	4	LmjF27.0490	II-III-II-III-R ₁ -II-III-C	4
LmCALP27.2	LmjF27.0500	2xII-2xIII-R-II-III	4	LmjF27.0500	II-III-II-III-R ₁ -II-III-C	4
LmCALP27.3	LmjF27.0510	2xII-2xIII-R	4	LmjF27.0510	II-III-II-III-R ₁ -II-III-C	4
				LmjF30.2040	I ^H , II, III, C	2

Table 3.2. Page 1 of 2. Full list of Ersfeld and co-workers classification and nomenclature of kinetoplastid calpain-like proteins (Ersfeld et al., 2005) and the corresponding re-classification and nomenclature presented and used in this thesis. G: group, Sg: subgroup, Kines: kinesin-like motor domain, ADCY: adenylate cyclase catalytic domain.

Table 3.2. Page 2 of 2.

<i>L. major</i>				<i>L. major</i>		
Name	geneDB ID	Domains	G	geneDB/Name	Domains	G
LmCALP31.1	LmjF31.0390	I ^K , II, III, C	1	LmjF31.0390	I ^K , II, III, C	1
LmCALP31.2	LmjF31.0400	I ^K , II, III, C	1	LmjF31.0400	I ^K , II, III, C	1
LmCALP31.3	LmjF31.0410	I ^K , II, III, C	1	LmjF31.0410	I ^K , II, III, C	1
LmCALP31.4	LmjF31.0430	I ^K , II, III, C	1	LmjF31.0430	I ^K , II, III, C	1
LmCALP31.5	LmjF31.0440	I ^K , II, III, C	1	LmjF31.0440/CAP5.5-like	I ^K , II, III, C	1
LmCALP31.6	LmjF31.0460	I ^K , II, III, C	1	LmjF31.0460	I ^K , II, III, C	1
				LmjF32.0970	I ^H , II, III, C	2
				LmjF32.2010	I ^K , II, III, C	1
LmCALP33.1	LmjF33.2010	I ^H , II, III, C	2	LmjF33.2010	I ^H , II, III, C	2
LmSKCRP34.1	LmjF34.0190	I ^K	3	LmjF34.0190	4xI ^K	3, Sg 3b
				LmjF34.0200	4xI ^K	3, Sg 3b
				LmjF34.0280	I ^H , II, III, C	2
				LmjF36.0780	I ^H , II, III, C	2
				LmjF36.3180	ADCY-I ^K	3, Sg 3b
<i>T. cruzi</i>				<i>T. cruzi</i>		
TcCALP721.30	Tc00.1047053506721.30	II-III-II-III-R-II-III-C	4	TcCALP721.30	II-III-II-III-R ₂ -II-C	4
TcCALP441.10	Tc00.1047053511441.10	R-II-III-C		TcCALP441.10	R ₂ -II-III-C	4

“catalytic” domain II, lacking the domains III and C characteristic of group 1 calpain-like proteins. Reciprocal BLAST searches indicate that the three genes are orthologs across these different kinetoplastid species (Appendix 3, Reciprocal_BLAST).

L. major LmjF20.1210 and *L. infantum* LinJ20_v3.1250.

L. major LmjF20.1210, previously described by Ersfeld and co-workers as a group 1 calpain-like sequence, contains an N-terminal calpain “catalytic” domain II followed by a calpain-like domain III, lacking the I^K and C domains stated by *Ersfeld et al., 2005*. Interestingly, the *L. infantum* ortholog of this sequence, LinJ20_v3.1250, (Appendix 3, Reciprocal_BLAST) does have the domain architecture characteristic of kinetoplastid group 1 calpain-like sequences defined by Ersfeld and co-workers —domains I^K, II, III and C—. Furthermore, reciprocal BLAST searches within the *L. major* and *L. infantum* genome sequences identified LmjF20.1210 and LinJ20_v3.1250 as the putative paralogs of LmjF32.2010 and LinJ32_3.2140, respectively (domains I^K and II only), despite their different domain composition (Appendix 3, Reciprocal_BLAST) and

chromosome localisation. The high similarity between the domain II sequences of these putative paralogs and LmjF32.2010 and LinJ32_3.2140 ortholog in *T. brucei*, Tb11.01.6870, suggest that they may have evolved from a common ancestor gene containing only domains II and III, losing domain III or acquiring domains I^K and C throughout evolution.

T. brucei Tb927.8.8330, CAP5.5V.

The newly identified CAP5.5V, Tb927.8.8330, has the same group 1 domain architecture as CAP5.5, its paralogous gene (Fig. 3.16).

- **Group 2: I^H-II-III-C.**

This group was defined by Ersfeld *et al.* as that similar to group 1 but containing a heterogeneous domain I^H, instead of a kinetoplastid domain I^K. As discussed in the Introduction chapter (Section 1.8.2., page 42), the I^H domain is unrelated to domain I^K and does not contain acylation modification motifs (Ersfeld *et al.*, 2005). In my re-classification, I kept the group 2 of kinetoplastid calpain-like proteins as it was defined by Ersfeld *et al.*, including in it 7 of the newly identified calpain-related sequences:

T. brucei Tb11.01.5800; *L. major* LmjF32.0970; and *L. infantum* LinJ32_v3.1020, which are orthologous genes (Appendix 3, Reciprocal_BLAST).

L. major LmjF34.0280, LmjF36.0780; and *L. infantum* LinF34_v3.0300, LinJ36_v3.08.

These *Leishmania* specific sequence pairs are orthologs, with the members of each pair being in-paralogs (Appendix 3, Reciprocal_BLAST).

- **Group 3.**

The group 3 of trypanosomatid calpain-like proteins was defined by Ersfeld *et al.* as that constituted by sequences containing a domain I^K —the sequence element exclusively found in kinetoplastids, either at the N-terminus of group 1 calpain-like sequences or as standalone ORFs— but lacking a calpain “catalytic” domain. As

discussed earlier (Section 3.2, page 71), the analysis of kinetoplastid calpain-like sequences performed during this thesis work showed that domain I^K cannot be considered a calpain-related domain, as it does not contain any motif related to calpains and is only found in the group 1 of kinetoplastid calpain-like proteins. Nevertheless, I have included the group 3 of kinetoplastid calpain-related proteins defined by Ersfeld *et al.* in my re-classification of trypanosomatid calpain-like proteins for consistency with their published classification (Ersfeld *et al.*, 2005), based on the unique correlation between domain I^K and group 1 kinetoplastid calpain-related proteins. Therefore, during my analysis I also searched for new trypanosomatid sequences that match Ersfeld and co-workers definition of group 3 calpain-like proteins: sequences containing a domain I^K but lacking any calpain-related domain. This search rendered the identification of 8 new sequences —4 in *T. brucei*, 2 in *L. major* and 2 in *L. infantum*—, some of which containing several copies of domain I^K (similar to the previously described *T. brucei* Tb10.70.5060) or additional domains not related to calpains: META and Adenylate Cyclase catalytic domains. Thus, I subdivided the previously described group 3 into four subgroups according to their domain composition —subgroups 3a to 3d—, although none of this sequences can be strictly considered a calpain-like protein.

Subgroup 3a, I^K. Includes sequences that contain a domain I^K followed by very short C-terminal extensions, similar to *L. major* Small Myristoylated Proteins (SMP1-3) (Tull *et al.*, 2004), often containing single or double N-terminal acylation motifs:

T. brucei Tb927.1.2150, Tb927.1.2160, Tb927.1.2230, Tb927.1.2260, Tb927.7.4060 (previously described) and Tb927.7.4070 (newly identified in my analysis).

L. major LmjF20.1220, LmjF20.1230, LmjF20.1310 SMP-1, LmjF20.1300 SMP-2, LmjF20.1280 SMP-3, LmjF14.0850, orthologs of *T. brucei* group 3a sequences; but also the *Leishmania* species-specific sequences LmjF20.1250 and LmjF25.1750.

L. infantum LinJ20.1260, LinJ20.1270, LinJ20.1350, LinJ20.1340, LinJ20.1320, LinJ14.0900, LinJ14.0890, orthologs of *T. brucei* group 3a sequences; and the *Leishmania* species-specific sequences LinJ20_v3.1290 and LinJ25_v3.2830, which are in turn orthologs of *L. major* LmjF20.1250 and LmjF25.1750 (Appendix 3, Reciprocal_BLAST).

Subgroup 3b, I^K-I^K-I^K-I^K. Includes sequences containing four degenerate tandem repeats of domain I^K:

T. brucei Tb10.70.5060, *L. major* LmjF34.0190, *L. infantum* LinJ34.0210 and the new identified *L. major* LmjF34.0200 and *L. infantum* LinJ34.0220. *L. major* and *L. infantum* sequences are encoded by paralogous genes, all of them homologs of *T. brucei* Tb10.70.5060 (Appendix 3, Reciprocal_BLAST).

Subgroup 3c, META-META-META-I^K. Includes sequences containing three degenerate tandem-repeats of the META-domain followed by a domain I^K:

T. brucei Tb927.5.2190, Tb927.5.2220 and Tb927.5.2250, three newly identified sequences that are identical gene copies; and their orthologs in *L. major*, LmjF17.0870, and *L. infantum*, LinJ17.1000 (Appendix 3, Reciprocal_BLAST).

Subgroup 3d, ADCY-I^K. Includes the newly identified *Leishmania* specific sequences containing an adenylate cyclase catalytic domain (ADCY) and a C-terminal domain I^K:

L. major LmjF36.3180 and *L. infantum* LinJ36_v3.3330, which are orthologous genes (Appendix 3, Reciprocal_BLAST).

- **Group 4.**

I redefined the group 4 of kinetoplastid calpain-like proteins to include all calpain-related sequences containing 67-68 amino acid tandem repeats. These sequences have been previously classified as groups 4 and 5, according to the number of copies

of calpain-like domains II and III that they contained. However, the analysis of these sequences during this thesis work has shown that the shorter tandem repeats-containing sequences, containing only one copy of domains II and III (classified as group 5 by Ersfeld *et al.*) are most probably fragments of the larger sequences, which contains two copies of these domains (previously classified as group 4). Thus, the redefined group 4 of kinetoplastid calpain-like sequences includes:

T. brucei Tb11.47.0035, Tb11.47.0036 and the newly identified Tb11.57.0008 and Tb11.v4.0001; *L. major* LmjF27.0490, LmjF27.0500, LmjF27.0510; and *L. infantum* LinJ27_v3.2490, LinJ_v3.0500 and LinJ_v3.0510.

- **Group 5.**

A new group 5 of kinetoplastid calpain-like proteins was defined to include the newly identified sequences containing a kinesin-like motor domain and a kinetoplastid calpain-like domain III:

T. brucei Tb927.5.2410, *L. major* LmjF17.1110, and *L. infantum* LinJ17.1230, which are orthologous genes (Appendix 3, Reciprocal_BLAST).

3.4. Syntenic genome organisation.

Reciprocal blast of the whole repertoire of kinetoplastid calpain-like proteins in *T. brucei*, *L. major* and *L. infantum*, including the newly found sequences and those previously described by Ersfeld and co-workers, allowed the identification of orthologous and paralogous genes, but also of species-specific sequences (Appendix 3, Reciprocal_BLAST). The majority of kinetoplastid calpain-related proteins are clustered on particular chromosomes. Some of these clusters contained identical or paralogous genes as tandem repeats, a typical feature of trypanosomatids gene organisation (Myler *et al.*, 1999; El-Sayed *et al.*, 2003; Hall *et al.*, 2003; Worthey *et al.*, 2003).

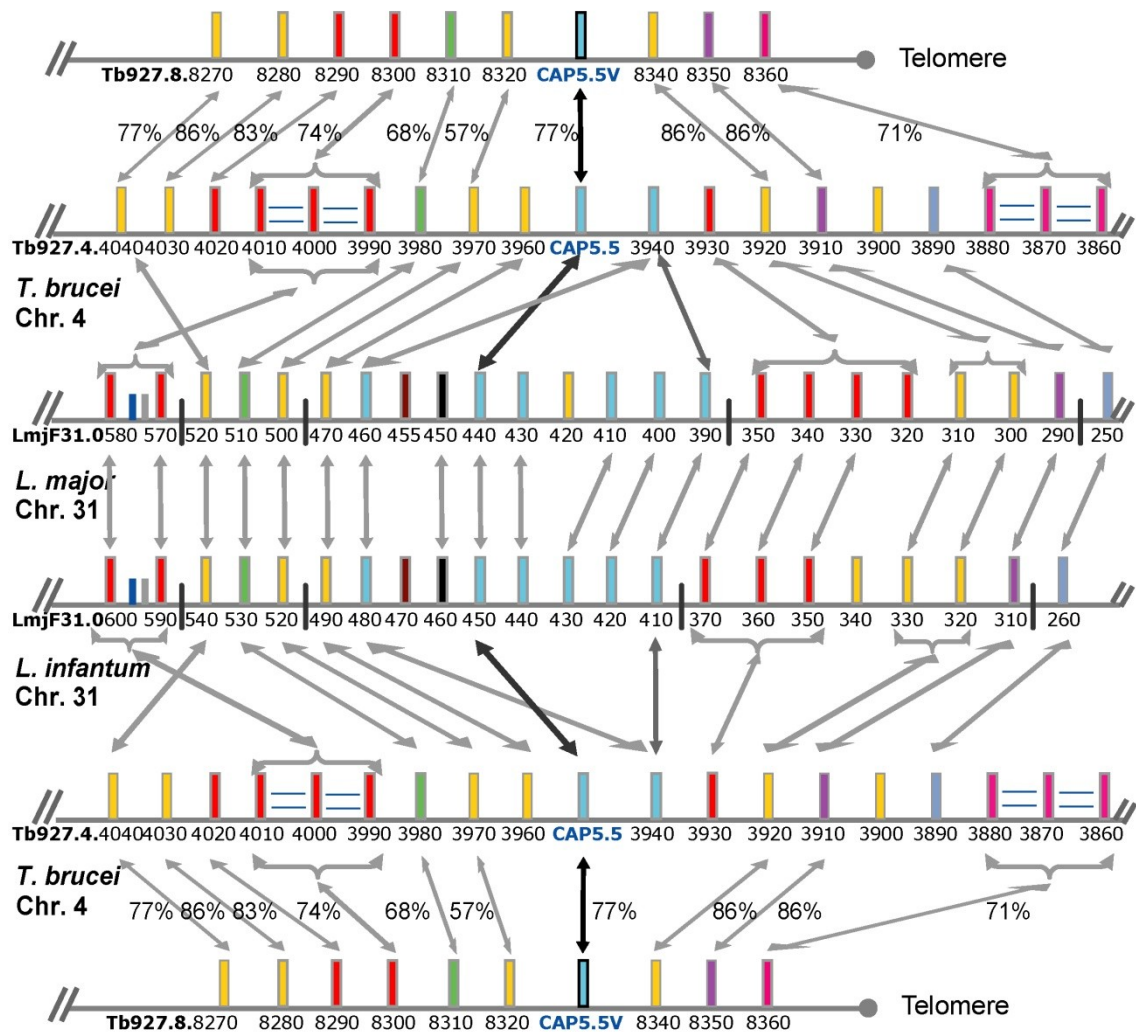
The presence of kinetoplastid calpain-like sequences clustered on particular chromosomes have already been described (*Ersfeld et al., 2005*). The gene clusters found by Ersfeld and co-workers on *T. brucei* chromosomes 1 and 4 were all syntenic in *L. major* and *T. cruzi* (*Ersfeld et al., 2005*). These clusters were constituted by calpain-like proteins belonging to groups 1 and 3 —domain I^K and II-containing sequences and domain I^K only genes, respectively— but also by ORFs not related to calpains. Sequences from both groups were not interspersed and similarities between orthologous genes were often greater than between adjacent sequences (*Ersfeld et al., 2005*). Analysis of the genome locus of the newly identified orthologous genes showed that the majority of calpain-like sequences were located on syntenic regions between the three studied parasites (Fig. 3.5, 3.10, 3.21 and 3.22). Reciprocal BLAST also identified several species-specific calpain-like sequences in *Leishmania* (Appendix 3, Fig. A.3.14).

3.5. Kinetoplastid calpain-like proteins table/database.

The data obtained from the sequence analysis presented above was used to build a database of the whole repertoire of calpain-like proteins in the three analysed organisms (Table 3.3). This database allows the easy identification of orthologs, paralogs and species-specific sequences. The database also includes information about the sequences location and length, conservation of the catalytic triad, presence of N-terminal acylation modification motifs and of the five Ca²⁺ binding residues of domain II (Table 3.3). Analysis of the whole repertoire of calpain-like sequences in the three studied kinetoplastid species showed that, from a total of 62 domain II containing sequences, only 9 have a complete conservation of the catalytic triad characteristic of conventional calpains: C-H-N. Additionally, all sequences lacked the EF-hand calcium-binding motifs of a typical calpains domain IV.

Figure 3.21.

***T. brucei* Chromosome 8**



***T. brucei* Chromosome 8**

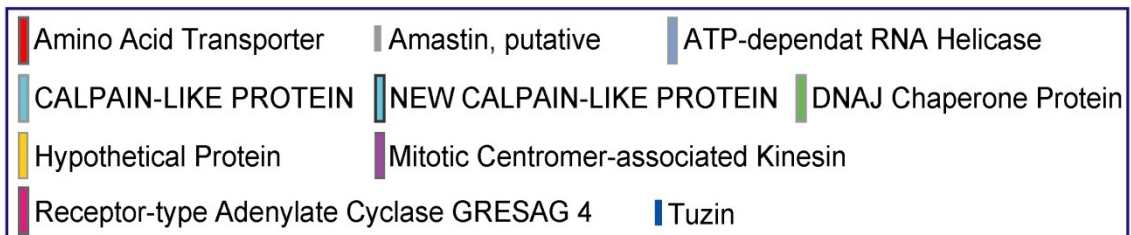


Figure 3.21. Synteny map and percentage of amino acid identity between the subtelomeric regions of *T. brucei* chromosomes 4 and 8 and their syntenic regions in *L. major* and *L. infantum*, chromosome 31. Reciprocal blast within the *T. brucei* genome sequence identified paralogous genes in chromosomes 4 and 8 subtelomeric regions. The percentage of amino acid identity between paralogs ranged from 57 to 86%. Comparison and reciprocal blast of these *T. brucei* regions with *Leishmania* species showed the presence of only one syntenic region in *L. major* and *L. infantum*, chromosome 31, containing orthologous genes to those on *T. brucei* chromosomes 4 and 8 subtelomeric regions. These data strongly indicate a large chromosome duplication event in *T. brucei* that occurred after the divergence of *Trypanosoma* and *Leishmania*, about 250 Mya (Douzery et al., 2004).

Figure 3.22.

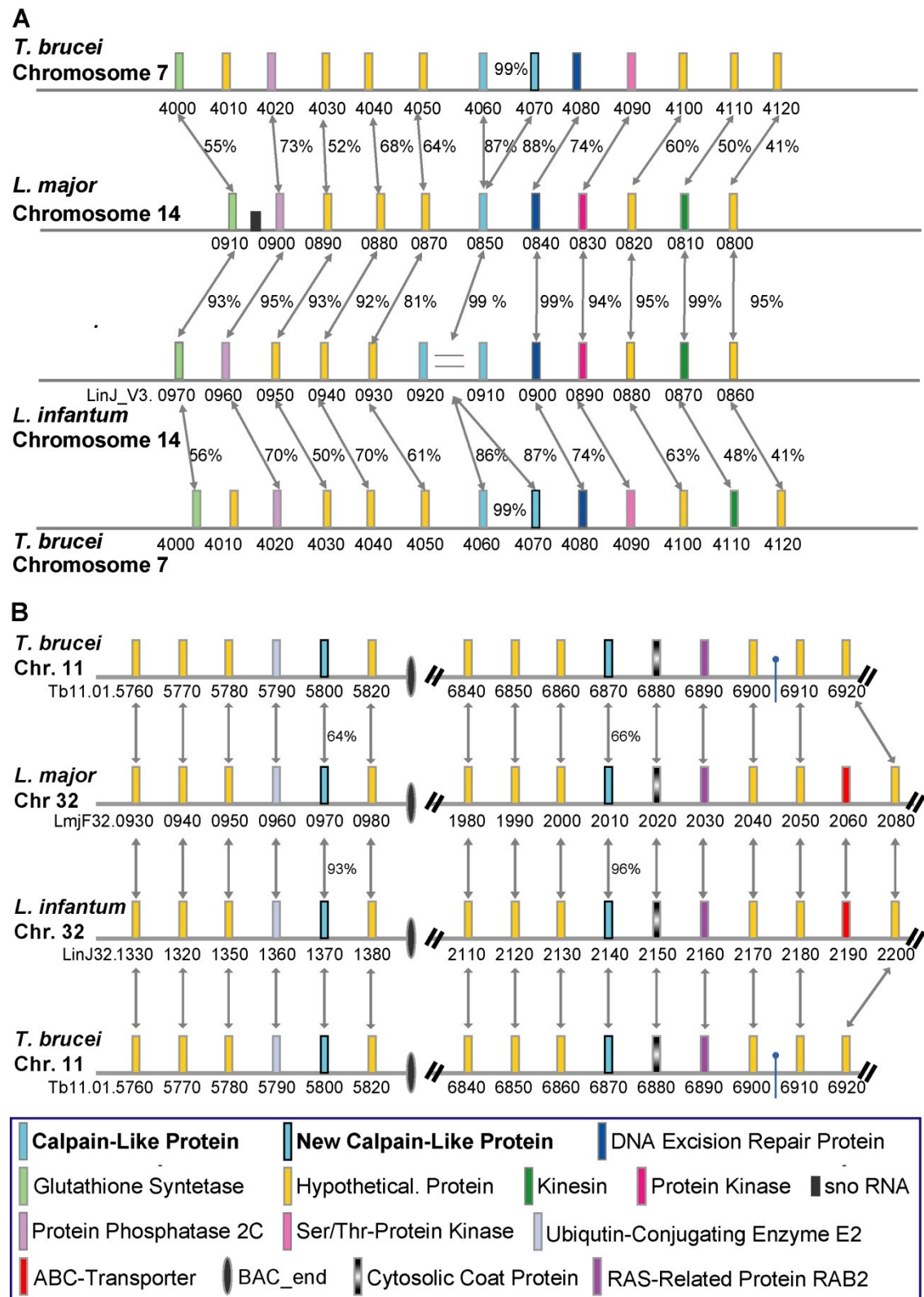


Figure 3.22. Synteny map of the genome locus of newly identified *T. brucei* calpain-like sequence with its orthologous regions in *L. major* and *L. infantum*. **A.** Tb927.7.4070 showed a 99% amino acid identity with the adjacent calpain-like sequence Tb927.7.4060, previously identified by Ersfeld *et al.*, 2005. **B.** Tb11.01.5800 and Tb11.01.6870 with their orthologous regions in *L. major* and *L. infantum*.

Table 3.3. Page 1 of 2.

	<i>T. brucei</i>	Chro.	Catalytic Triad	DII Ca ²⁺ binding sites D E E D D	Length (aa)	<i>L. major</i>
GROUP 1 I ^K -II-III-C	Tb927.1.2100	1	SHN	-E776E906D913D936	1146	LmjF20.1180
	Tb927.1.2110	1	SHN	-E314--D472	682	LmjF20.1185
	Tb927.1.2120	1	YHN	-E357--D510	732	LmjF20.1190
	Tb927.4.3940	4	SSN	--E537	784	LmjF31.0390
						LmjF31.0400
						LmjF31.0410
						LmjF31.0430
						LmjF31.0460
	Tb927.4.3950	4	SYN	-E344--D496	853	LmjF31.0440
	Tb927.8.8330	8	SYN	-E361----	888	
	Tb09.160.5550	9	CYN	D345E428--D599	856	LmjF04.0450
						LmjF20.1200
GROUP 2 I ^H -II-III-C						LmjF25.1480
	Tb11.01.6870 (I ^K -II only)	11	PT-	-E199----	301	LmjF32.2010
	G 1 Total = 8					LmjF20.1210 (II-III)
						G 1 Total = 14
	Tb927.6.3310	6	SHS	E311--D546--	1010	LmjF30.2040
	Tb10.389.0470	10	KT-	E170D254----	678	LmjF18.1060
	Tb10.70.5950	10	HH-	D738----D971	1594	LmjF21.0120
	Tb11.01.5800	11	IDH	-E370----	1504	LmjF32.0970
	Tb11.02.2240	11	CHN	-E359----	1053	LmjF33.2010
						LmjF34.0280
						LmjF36.0780
GROUP 3 I ^K						G 2 Total = 7
	Tb927.1.2150	1	--	--	130	LmjF20.1220
	Tb927.1.2160	1	--	--	140	LmjF30.2040
	Tb927.1.2230	1	--	--	123	LmjF20.1310 SMP-1
	Tb927.1.2260	1	--	--	130	LmjF20.1300 SMP-2
						LmjF20.1280 SMP-3
	Tb927.7.4060	7	--	--	115	LmjF14.0850
	Tb927.7.4070	7	--	--	120	
						LmjF20.1250
						LmjF25.1750
	Tb10.70.5060	10	--	--	484	LmjF34.0190
						LmjF34.0200
GROUP 4 Repeats containing sequences IIa-IIIa-IIb-IIIb-R-IIc-IIId IIa IIIa IIb IIIb R R-II-III-C						LmjF17.0870
						LmjF36.3180
	G 3 Total = 10					G3 Total = 12
	Tb11.47.0035	11	C65 Q223 N248 C682 Q839 N866 G5563 C5721 N5747	E55 _ E254 _ D279 E672 _ _ D884 _ D5554 E5681 _ D5781	6145	LmjF27.0500
						LmjF27.0510
	Tb11.57.0008	11	C66 Q224 N249 C685 Q842 N869	E56 _ E255 _ D280 E675 _ _ D884 _	5795	LmjF27.0490
	Tb11.47.0036	11	GCN	D819 E886 _ _ _	1409	
	Tb11.v4.0001	11	SCN	-E1127 _ _ _	1649	
	G 4 Total = 4					G 4 Total = 3
	Tb927.5.2410	5	--	--	1339	LmjF17.1110
	G 5 Total = 1					G 5 Total = 1
G 5 Kinesin-III-C						
<i>T. brucei</i> Calpain-Like Proteins Total Number = 28					<i>L. major</i> = 37	
10 New Calpain-Like Sequences					10 New	

Table key

Double acylated
 Single acylated
 Paralogs
 Specie specific
 CHN Conserved catalytic triad

Table 3.3. Page 1 of 2. Trypanosomatid calpain-like proteins database: *T. brucei*, *L. major* and *L. infantum*. Orthologous genes are displayed in the same row. Newly identified sequences are shown in blue font.

Table 3.3. Page 2 of 2.

	<i>L. major</i>	Chro.	Catalytic Triad	DII Ca ²⁺ -binding sites D E E D D	Length (aa)	<i>L. infantum</i>	Chro.	Cat. Triad	Synteny	Length (aa)
G1	LmjF20.1180	20	SHN	— E613 E738 —	976	LinJ20.1210	20	SHN	Yes	1014
	LmjF20.1185	20	SHN	— E459 E584 —	823	LinJ20.1220	20	SHN	Yes	821
	LmjF20.1190	20	SHN	— E309 — D475	687	LinJ20.1230	20	SHN	Yes	687
	LmjF31.0390	31	THN	— — E607 D640	855	LinJ31.1140	31	THN	Yes	784
	LmjF31.0400	31	HHN	— — — —	932	LinJ31.1150	31	HHN	No	936
	LmjF31.0410	31	HIS	— — E469 —	754	LinJ31.1160	31	HIS		754
	LmjF31.0430	31	VMC	— — — — D703	938	LinJ31.0590	31	VMC		939
	LmjF31.0460	31	CHN	— E341 — D513	732	LinJ31.0550	31	CHN		777
	LmjF31.0440	31	SYN	— E46 E478 — D513	723	LinJ31.v3.0450	31	SYN	Yes	724
	LmjF04.0450	4	CHN	— E425 — —	854	LinJ04.0440	4	CHN	Yes	855
	LmjF20.1200	20	-MK	— E322 — D463 D485	743	LinJ20.1240	20	-MK	No	743
	LmjF25.1480	25	CHN	D224 E308 E442 D449 D471	705	LinJ25.1550	20	CHN		706
LmjF32.2010	32	P--	— E199 — —	294	LinJ32.v3.2140	32	P--	Yes	295	
LmjF20.1210(II-III only)	20	-MN	— E72 — —	414	LinJ20.v3.1250	20	SVN	Yes	999	
	G 1 Total = 14					G 1 Total = 14				
G2	LmjF30.2040	30	SHS	D379 — E677 D686 —	1354	LinJ30.v3.2040	30	SHS	Yes	1354
	LmjF18.1060	18	-IA	E245 — — —	774	LinJ18.1070	18		Yes	763
	LmjF21.0120	21		— — D978 E1006	1628	LinJ21.v3.0160	21		Yes	982
						LinJ21.v3.0170	21		Yes	471
	LmjF32.0970	32		— E359 D530 — —	1894	LinJ32.v3.1020	32			1888
	LmjF33.2010	33	CHN	— E483 E597 — —	1379	LinJ33.2020	33	CHN	No	1045
	LmjF34.0280	34	TVD	— E613 — — —	1555	LinJ34.v3.0300	34	TVD	No	1555
LmjF36.0780	36	ATN	— D537 E668 — —	1071	LinJ36.v3.0840	36	ATN	No	1119	
	G 2 Total = 7					G 2 Total = 8				
G3a	LmjF20.1220	20	—		146	LinJ20.1260	20		Yes	146
	LmjF20.1230	20	—		133	LinJ20.1270	20		Yes	133
	LmjF20.1310 SMP-1	20	—		131	LinJ20.1350	20		Yes	131
	LmjF20.1300 SMP-2	20	—		136	LinJ20.1340	20			136
	LmjF20.1280 SMP-3	20	—		152	LinJ20.1320	20			151
	LmjF14.0850	14	—		115	LinJ14.0900	14		Yes	115
						LinJ14.0890	14		Yes	115
LmjF20.1250	20	—		236	LinJ20.v3.1290	20			235	
LmjF25.1750	25	—		142	LinJ25.v3.2830	25			142	
G3b	LmjF34.0190	34	—		491	LinJ34.0210	34		Yes	491
	LmjF34.0200	34	—		517	LinJ34.0220	34			517
G3c	LmjF17.0870	17	—	—	444	LinJ17.1000	17	—	Yes	444
G3d	LmjF36.3180	36	—	—	1401	LinJ36.v3.3330	36	—		1401
	G 3 Total = 11					G 3 Total = 12				
G4	LmjF27.0500	27	C62 Q220 N245 C679 G836 N863 G5573 C5732 N5758	E52 E127 E251 E276 D669 — — D878 D5564 — — — D5793	6164	LinJ27.v3.0510	27		Yes	6168
	LmjF27.0510	27	C64 Q222 N247 C681 R838 N863 S4764 C4922 N4948	E54 E129 E253 E278 D671 — — D878 D4755 — — — D4983	5358	LinJ27.v3.2490	27			662
	LmjF27.0490	27	C65 Q223 N248 C684 Q840 N866 T3974 C4132 N4158	E55 E130 E254 — — D674 — — D882 — D3965 — — — —	4553	LinJ27.v3.0500	27			5982
	G 4 Total = 3					G 4 Total = 3				
G5	LmjF17.1110	17	—	—	1430	LinJ17.1230	17			1430
	G 5 Total = 1					G 5 Total = 1				
L. major Total Calpain-Like Proteins = 37						L. infantum Total Calpain-Like proteins=39				
10 New Calpain-Like Sequences						11 New Calpain-Like Sequences				

Double acylated

Single acylated

Paralogs

Specie specific

CHN Conserved catalytic triad

Double acylated
 Single acylated
 Paralogs
 Specie specific
 CHN Conserved catalytic triad

Table 3.3. Page 2 of 2. Trypanosomatid calpain-like proteins database: *T. brucei*, *L. major* and *L. infantum*. Orthologous genes are displayed in the same row. Newly identified sequences are shown in blue font.

3.6. Selection of *T. brucei* calpain-like sequences for functional analysis by RNAi.

After the analysis and classification of trypanosomatid calpain-like proteins, I selected 10 *T. brucei* sequences as the most interesting candidates for performing functional analysis by RNAi, based on the presence of a completely or partially conserved calpain catalytic triad, conservation across different trypanosomatid species, and/or the presence of domains not related to calpains that could hint to their biological functions.

1-Tb9.160.5550.

Tb9.160.5550 is a group 1 calpain-like sequence, with only one deviation from the canonical catalytic triad: substitution of H by Y (Table. 3.3). The hydroxyl group of Y is able to form hydrogen bonds that could collaborate in interactions with a substrate or binding partner and increase the nucleophilic activity of cysteine. Tb9.160.5550 orthologs in *L. major*, LmjF04.0450, *L. infantum*, LinJ04.0440, and *T. cruzi*, have a completely conserved catalytic triad. The conservation of this protein across different species suggests a common and important role in kinetoplastids. All these make Tb9.160.5580 a possible candidate to have protease activity.

2-CAP5.5

3-CAP5.5V.

These two group 1 paralogous genes, which are the product of a large chromosome duplication event, only differ significantly at their C-terminal domains. Both are cytoskeletal associated proteins and CAP5.5 is stage specific regulated, being only expressed in procyclic form trypanosomes (*Matthews and Gull, 1994a; Hertz-Fowler et al., 2001*). Additionally, CAP5.5 has been experimentally shown to be myristoylated and palmitoylated at the N-terminus, suggesting an interaction with the plasma membrane.

The cytoskeletal-association and possibly interaction with the plasma membrane of these two paralogous genes, together with the stage specificity of CAP5.5 expression, suggest an important function for CAP5.5 and CAP5.5V. Thus, they were selected for RNAi functional analysis.

4-Tb11.01.6870.

This group 1 calpain like protein has an unusual domain architecture, containing only a domain I^K and a calpain-like “catalytic” domain II, which has major deviations from the catalytic triad of conventional calpains. However, reciprocal BLAST searches have shown the conservation of this protein across different kinetoplastid species (Appendix 3, Reciprocal_BLAST). Furthermore, its orthologs in *L. major* and *L. infantum*, LmjF32.2010 and LinJ32_v3.2140, also have putative paralogs with different domain architecture: LmjF20.1210 (calpain-like domains II and III only) and LinJ20_v3.1250 (domains I^K, II, III and C). Reciprocal BLAST also suggest that these sequences have evolved from a common ancestor gene, probably containing calpain-like domains II, and III, which lost some of its domains and/or gain domains I^K and C after duplication. The unusual domain architecture of Tb11.01.6870 and the dissimilar domain architecture of some of its orthologs could relate to species-specific functions. Thus, I selected Tb11.01.6870 for RNAi functional analysis.

5-Tb11.02.2240

This group 2 sequence is the only calpain-like protein in *T. brucei* that has a completely conserved catalytic triad, C-H-N (Table. 3.3), being the most suitable candidate to have protease activity. Therefore, I selected Tb11.02.2240 for RNAi functional analysis.

6-Tb927.5.2190 and its identical gene copies: Tb927.5.2220 and Tb927.5.2250.

These calpain-like sequences contain 3 degenerated copies of the META domain, followed by a domain I^K sequence. The three identical gene copies are interspersed with the META1 proteins (Fig. 3.5). The META1 protein has been shown to be essential in *L. major* promastigotes, whereas its overexpression in *L. amazonensis* increased virulence (Uliana *et al.*, 1999). The presence of the META and I^K domains in these conserved sequences is both surprising and intriguing. Thus, I decided to investigate their function by RNAi.

7-Tb11.57.0008.

8-Tb11.47.0035.

9-Tb11.47.0036.

10-Tb11.v4.0001.

These group 4 sequences contain very long arrays of 67-68 amino acid tandem repeats, which are highly conserved across different kinetoplastid species (Fig.3.10). All studied kinetoplastid species contain at least three tandem repeats-containing sequences. The high conservation and numbers of these sequences suggest an important role in kinetoplastids. Although all analysed group 4 sequences contain a poorly conserved catalytic triad (Table. 3.3), their predicted highly regular structure, with alternated alpha-helical and not helical regions, together with their cellular localisation at the FAZ filament, points to a role for these proteins in cell-body-flagellum attachment. Furthermore, the presence of very similar group 4 sequences in *T. brucei* is compatible with stage-specific regulation. All these data make of group 4 calpain-like sequences good candidates for RNAi functional analysis.

Whereas Tb11.57.0008 and Tb11.47.0035 contain two and three copies of calpain-like “catalytic” domain II, respectively, Tb11.47.0036 and Tb11.v4.0001 only have one domain II copy. Moreover, Tb11.47.0036 and Tb11.v4.0001 are probably sequence fragments of Tb11.57.0008 and Tb11.47.0035, correspondingly. Tb11.47.0036 contains regions that are 100% identical to Tb11.57.0008 repeat sequences; whereas Tb11.v4.0001 shares some identical regions with Tb11.47.0035. Additionally, the N-terminus of Tb11.57.0008 and Tb11.47.0035 also contain identical regions. Exploiting the presence of these sequences identical regions, I built RNAi constructs with the capability of knocking down the expression of more than one of these proteins, if they were simultaneously expressed:

Tb11.57.0008 and Tb11.47.0036.

Tb11.47.0035 and Tb11.v4.0001.

Each pair of sequences, which are probably part of the same ORF, was targeted at the repeats. Additionally, Tb11.57.0008 and Tb11.47.0035, were targeted at their identical N-terminal region, whereas Tb11.v4.0001, Tb11.47.0035 and Tb11.47.0036 were targeted at their identical C-terminus.

3.7. Functional survey of calpain-like sequences by RNAi.

To probe the selected proteins function, inducible RNAi constructs (described in the Methods chapter) were transfected into procyclic 29.13 and bloodstream 90.13 cell lines (Wirtz *et al.*, 1999). After selection of stable RNAi transfectants, induction of RNAi was carried out by addition of doxycycline to the culture medium, obtaining the following results (Fig.3.23):

Figure 3.23.

GROUP	NAME	PCF	BSF
Group 1	CAP5.5	+	X
	CAP5.5V	X	+
	Tb9.160.5550 (CYN)	X	X
	Tb11.01.6870	X	X
Group 2	Tb11.02.2240 (CHN)	X	X
Group 3	Tb927.5.2190 (META protein)	N/O	N/O
Group 4	Tb11.47.0035/Tb11.47.0008	+	N/O
	Tb11.57.0008/Tb11.47.0036	N/O	+
	Tb11.47.0035/Tb11.V4.0001	N/O	X

Figure 3.23. Survey of *T. brucei* calpain-like sequences by RNAi.

Appearance of RNAi phenotype after induction is indicated by a plus (+) symbol in a red box. Absence of phenotype after induction is indicated by a X in a grey box. Transfectants that could not be obtained (possibly due to a lethal effect of RNAi leakiness) are indicated by N/O.

Tb9.160.5550.

Tb9.160.5550 RNAi cell lines were obtained in both, procyclic and bloodstream forms, where it seems to be similarly expressed (*Liu et al., 2009*). However, induction of these cell lines did not produce any discernible phenotype (Appendix 3, Tb9.160.5550 RNAi).

CAP5.5 and CAP5.5V (Tb927.8.8330).

CAP5.5 and CAP5.5V RNAi cell lines were obtained in both procyclic and bloodstream form trypanosomes. Induction of CAP5.5 PCF RNAi cell line produced a strong growth and morphology phenotype. As expected, CAP5.5 RNAi in BSF did not produce any noticeable phenotype. Interestingly, the contrary happened with CAP5.5V RNAi cell lines: RNAi of CAP5.5V in PCF did not produce any discernible phenotype, whereas in BSF it caused strong growth and morphology phenotypes. These results strongly suggest a bloodstream specificity of CAP5.5V expression and an important role of the two paralogs in the cell morphogenesis of procyclic —CAP5.5— and bloodstream —

CAP5.5V— forms. The RNAi functional analysis of both proteins is detailed in chapter 4 of this thesis and in *Olego-Fernandez et al., 2009*.

Tb11.01.6870.

Tb11.01.6870 RNAi cell lines were obtained in both procyclic and bloodstream form *T. brucei*. However, the RNAi induction of these cell lines did not produce any discernible phenotype (Appendix 3, Tb11.01.6870 RNAi).

Tb11.02.2240.

Tb11.02.2240 RNAi cell lines were obtained in both procyclic and bloodstream trypanosomes. However, induction of neither RNAi cell line produced any noticeable phenotype, even though this sequence is the only *T. brucei* calpain-like protein with a completely conserved catalytic triad (Appendix 3, Tb11.02.2240 RNAi).

Tb927.5.2190, Tb927.5.2220 and Tb927.5.2250 (identical gene copies).

Despite several attempts, I could not obtain any Tb927.5.2190 RNAi cell line in procyclic or bloodstream form trypanosomes. A possible explanation could be an essential role of this protein in both procyclic and bloodstream forms, with a lethal effect of RNAi leakiness. However, Tb927.5.2190 has been recently shown to be preferentially expressed in BSF trypanosomes by qRT-PCR (*Liu et al., 2009*).

Tb11.57.0008, Tb11.47.0035, Tb11.47.0036 and Tb11.v4.0001.

I obtained the following group 4 RNAi cell lines, with the capability of knocking-down the expression of more than one sequence:

- Tb11.57.0008/Tb11.47.0035 PCF RNAi cell line, which induction produced a strong growth and morphology phenotype.
- Tb11.57.0008/Tb11.47.0036 BSF RNAi cell line, which induction also produced a strong growth and morphology phenotype.

- Tb11.47.0035/Tb11.v4.0001 BSF RNAi cell line, which induction did not produce any noticeable phenotype. Despite several attempts, I could not obtain a Tb11.47.0035/Tb11.v4.0001 RNAi cell line in procyclic trypanosomes, possibly due to the essential role of this protein or proteins in PCF, which could cause a lethal effect of RNAi leakiness.

These results suggest an essential role for Tb11.57.0008/Tb11.47.0036 and Tb11.47.0035/Tb11.v4.0001 in the cell morphogenesis of BSF and PCF trypanosomes, respectively. The RNAi functional analysis of these proteins is detailed in the chapter 5 of this thesis.

CHAPTER 4

RNAi ANALYSIS OF TWO PARALOGOUS

CALPAIN-LIKE PROTEINS:

CAP5.5 AND CAP5.5V

4. CAP5.5 AND CAP5.5V FUNCTIONAL ANALYSIS

4.1. Introduction.

CAP5.5 was the first calpain-like family member to be biochemically characterised in *Trypanosoma brucei* (Hertz-Fowler *et al.*, 2001). CAP5.5 is a cytoskeletal-associated protein evenly distributed across the sub-pellicular microtubule corset of procyclic trypomastigotes throughout all stages of the cell-cycle (Hertz-Fowler *et al.*, 2001) in detergent-extracted cells (Fig. 4.1). CAP5.5 expression is tightly regulated in a lifecycle dependant manner, being only expressed in procyclic forms and classically providing a late stage marker for bloodstream-procyclic trypomastigote differentiation, detected by the anti-CAP5.5 mAb (Matthews and Gull, 1994a). Interestingly, CAP5.5 is double acylated at the N-terminus, myristoylated and palmitoylated, suggesting an interaction with the plasma membrane, as well as with the subpellicular microtubules (Hertz-Fowler *et al.*, 2001). CAP5.5 calpain-related protease domain has a degenerated calpain catalytic triad—S-H-N—, which is doubly substituted from the canonical catalytic triad of conventional calpains (C-H-N) raising uncertainty as to whether CAP5.5 is an active protease (Hertz-Fowler *et al.*, 2001) and recombinant expression of native protein has been so far unsuccessful. CAP5.5V, the newly identified CAP5.5 paralog described in Chapter 3, only differs significantly from CAP5.5 in its 3'-coding and 5'- and 3'-regulatory regions, having the same doubly substituted catalytic triad and N-terminal double acylation modification motif than CAP5.5. Moreover, CAP5.5V has been recently shown to be cytoskeletal-associated (Liu *et al.*, 2009).

In this chapter I show the stage specificity expression of CAP5.5V in bloodstream trypanosomes, indicated by quantitative RT-PCR, and describe the results from RNAi studies that selectively targeted CAP5.5 or/and CAP5.5V in both, procyclic and bloodstream trypanosomes. These studies were performed with the aim of identifying

Figure 4.1.

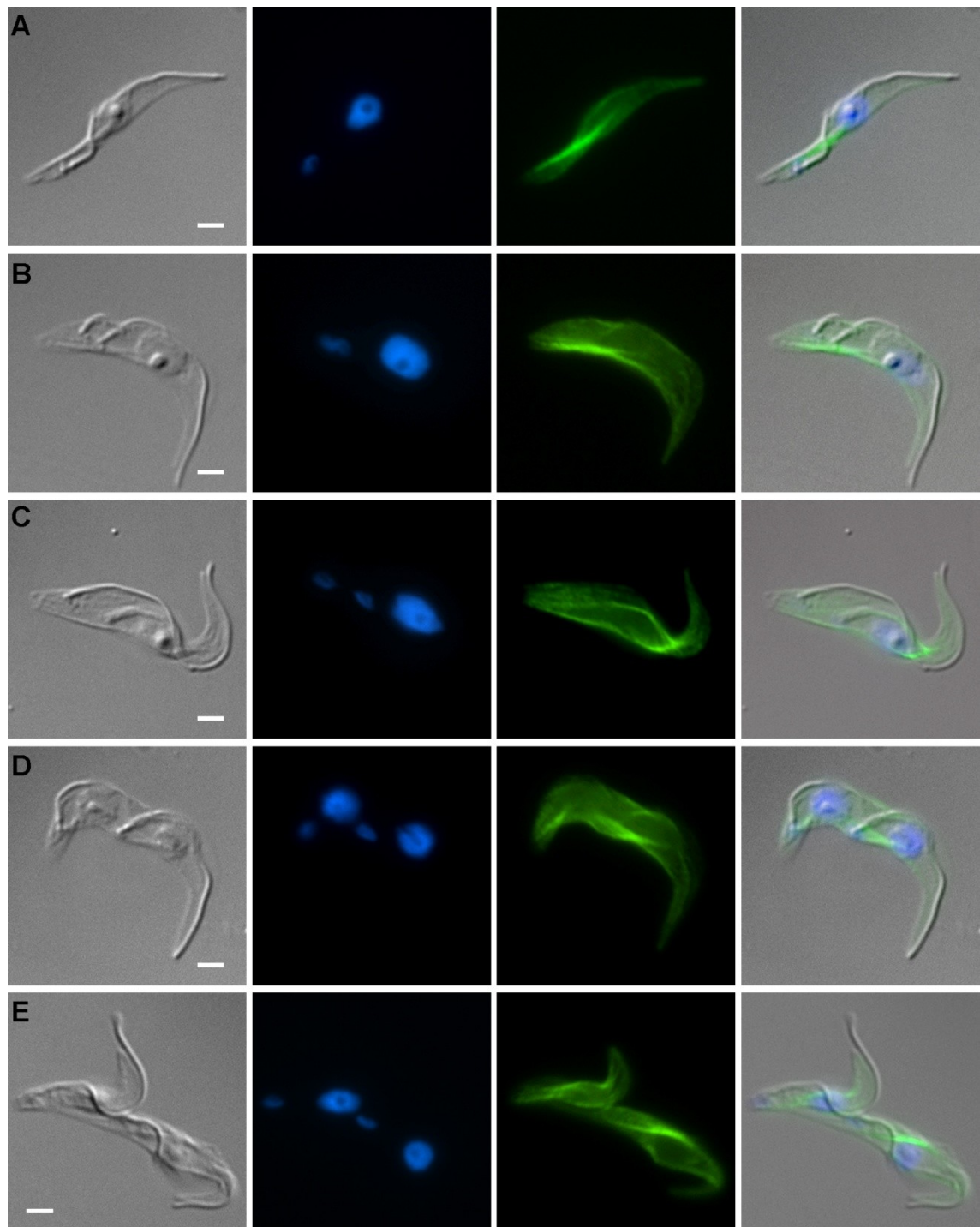


Figure 4.1. Detergent extracted cytoskeletons of *T. brucei* procyclic form, 427 strain, probed with CAP5.5 mAb. DIC, DAPI, CAP5.5 and merged 100x images. CAP5.5 localises across the whole cell-body cytoskeleton during all stages of the cell-cycle, but is always absent from the flagellum. Scale bars represent 2 µm.

the function of these paralogous, differentially expressed calpain-like proteins in *T. brucei*. The consequences of CAP5.5 and CAP5.5V RNAi in procyclic and bloodstream forms, respectively, were organelle mis-positioning, aberrant cytokinesis and disorganisation of the sub-pellicular cytoskeleton that defines cell shape. Additionally, immunocytochemistry of CAP5.5 RNAi-induced procyclic cells, using the anti-CAP5.5 mAb (which is CAP5.5 specific), showed the loss of CAP5.5 from the posterior cell-end shortly after induction, preceding aberrant organelle segregation and cytokinesis, suggesting that the cytoskeletal association of CAP5.5 at the posterior cell-end is highly dynamic during the cell division-cycle. Ultimately, defective organisation of the subpellicular microtubule corset was observed in both procyclic and bloodstream form trypanosomes after CAP5.5 and CAP5.5V RNAi induction, respectively. These data reveal an essential role for these paralogous, differentially expressed proteins for the correct morphogenetic patterning of *T. brucei* during the cell division-cycle of procyclic (CAP5.5) and bloodstream (CAP5.5V) trypomastigotes, as well as for the correct organisation of their subpellicular microtubule corset.

4.2. CAP5.5 and CAP5.5V are stage-specific regulated: expression analysis by RT-PCR.

I used the non-homologous regions between CAP5.5 and CAP5.5V at the C-terminus of their CDS and their 3'-untranslated regions to design gene-specific oligonucleotides for quantitative RT-PCR analysis (Fig. 4.2). The accumulation of transcripts was determined by normalisation against γ -tubulin mRNA (expressed in both cell forms). The analysis of CAP5.5 expression, relative to that of γ -tubulin, revealed an approximately five-fold greater accumulation of mRNA in procyclic trypomastigotes than that observed in bloodstream cells (Fig.4.2), correlating with previous studies that revealed CAP5.5 to be a procyclic-specific protein (*Hertz-Fowler et al., 2001*).

Figure 4.2.

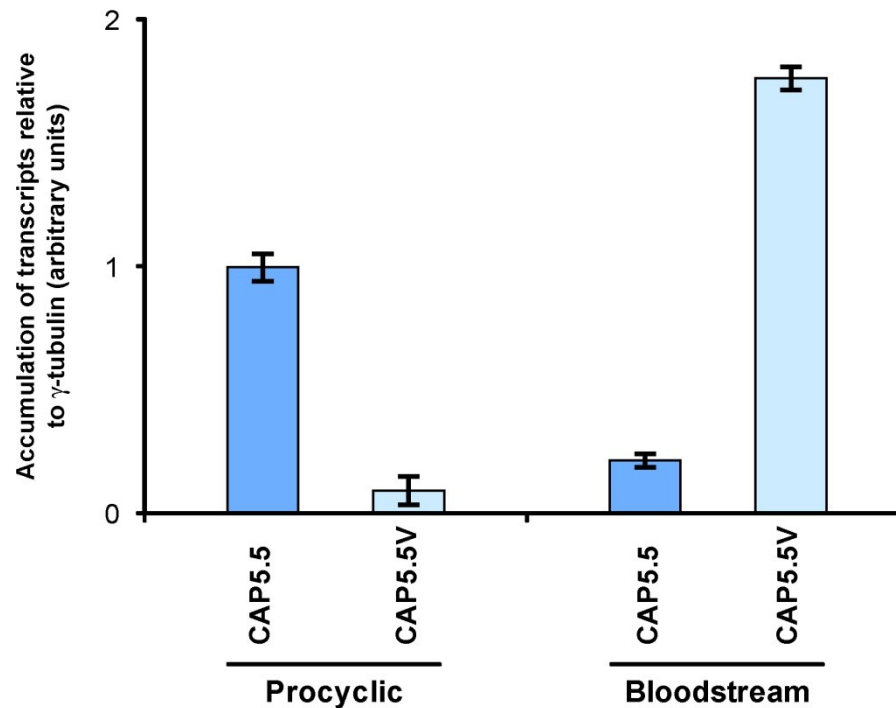


Figure 4.2. Quantitative real-time PCR analysis revealed the accumulation of CAP5.5 and CAP5.5V transcripts in procyclic or bloodstream trypomastigotes, respectively. Accumulation of transcripts was determined by normalisation to γ -tubulin expression. Relative expression of CAP5.5 in procyclic cells was arbitrarily set to 1. Values and errors are as described in the Methods chapter. The same pattern of stage-specific expression was observed when transcript abundance was normalised against QM10 expression (data not shown).

In contrast, the accumulation of CAP5.5V mRNA was approximately nineteen-fold greater in bloodstream *T. brucei* relative to procyclic cells (Fig.4.2). The same pattern of stage-specific expression was observed when transcript abundance was normalised against the ribosomal protein QM10 and against the geometric mean of γ -tubulin and QM10. Although mRNAs do not necessarily represent protein levels, the differences observed between CAP5.5 and CAP5.5V mRNA levels in two different developmental forms strongly indicate that CAP5.5V is a bloodstream-specific protein. The close similarity between CAP5.5 and CAP5.5V, coupled to their stage-specific regulation, suggests that the two paralogs could provide similar functions in different lifecycle stages.

4.3. Selection of sequence fragments for RNAi.

CAP5.5 and CAP5.5V DNA coding sequences differ extensively only in the last 550-600 bp of their 3' regions (from the 2917 and 2049 bp positions of their CDS, respectively). However, the regulatory regions of both genes are highly divergent (Chapter 3, Fig. 3.18). I exploited these differences to design different inducible RNAi constructs for gene-specific silencing of CAP5.5 and CAP5.5V (Fig. 4.3):

- RNAi constructs that target unique sequences of the 3'-coding regions of CAP5.5 and CAP5.5V —5.5-C and 5.5V-C RNAi constructs.
- RNAi constructs that target the 3'-coding and intergenic regions unique to each gene —5.5-C-U and 5.5V-C-U RNAi constructs.
- RNAi construct where the insert corresponded to the 1041-1979 bp positions of the CAP5.5 gene and which, due to extensive sequence identity with CAP5.5V, would be likely to give RNAi cross-talk between both sequences —5.5/5.5V RNAi construct.

4.4. Applying the RNAi phenotype test: CAP5.5 is essential in procyclic *T. brucei*.

To probe CAP5.5 and CAP5.5V function, the three different types of inducible RNAi constructs described above were initially transfected into procyclic 29.13 *T. brucei* cells. Upon induction, RNAi efficiency to silence CAP5.5 expression was assessed by Western blotting using the CAP5.5 mAb. Additionally, cultures were monitored to evaluate the appearance of changes in both cell growth and cell morphology. The temporal order of key events in the procyclic cell division-cycle (e.g. the relative timing of kinetoplast replication, nuclear S-phase, kinetoplast division, mitosis, and microtubule mediated events, such as new flagellum growth or cleavage furrow ingression) has long been described and understood (Sherwin and Gull, 1989a; Woodward and Gull, 1990). In asynchronous cultures of procyclic trypanosomes, the

Figure 4.3.

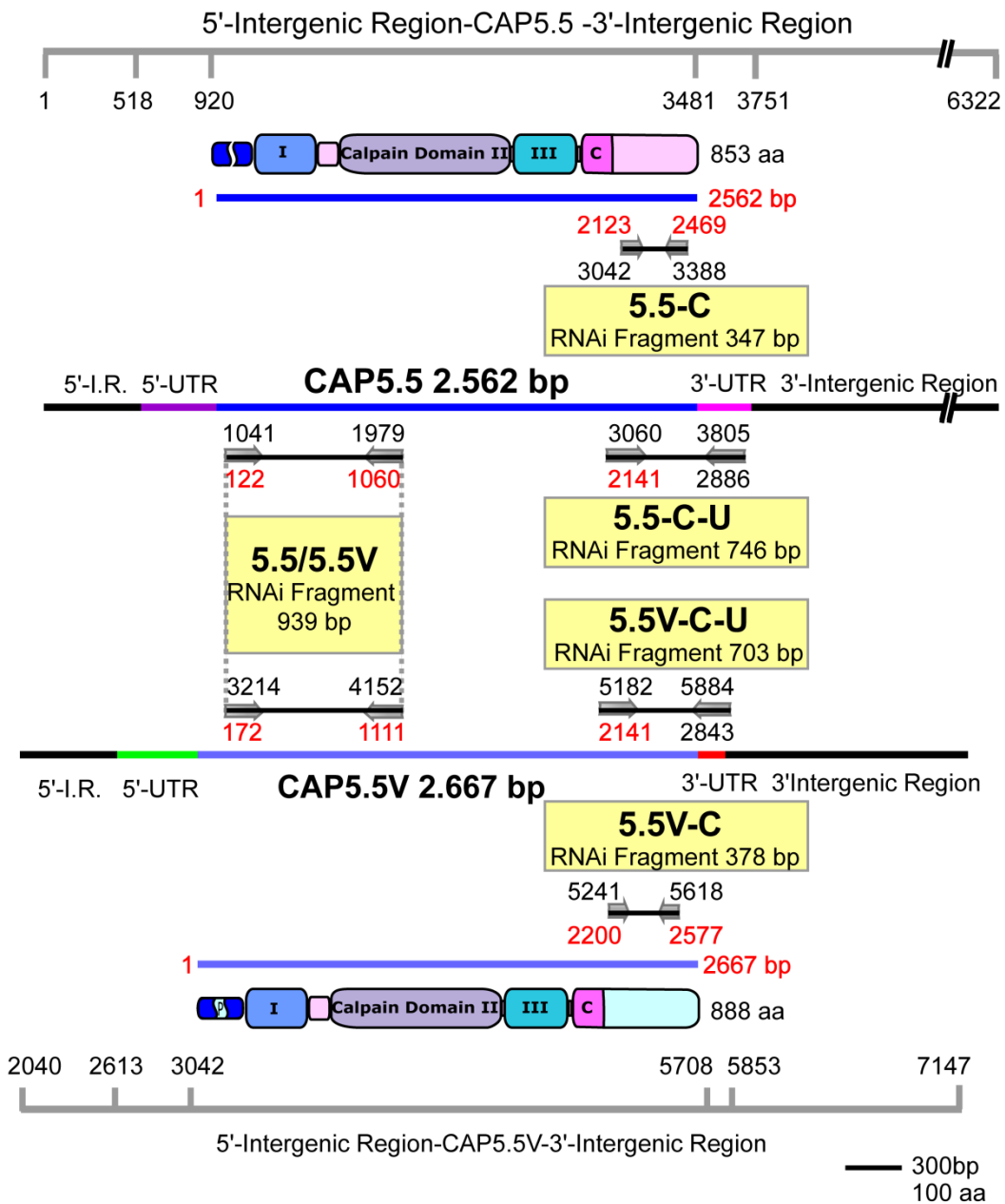


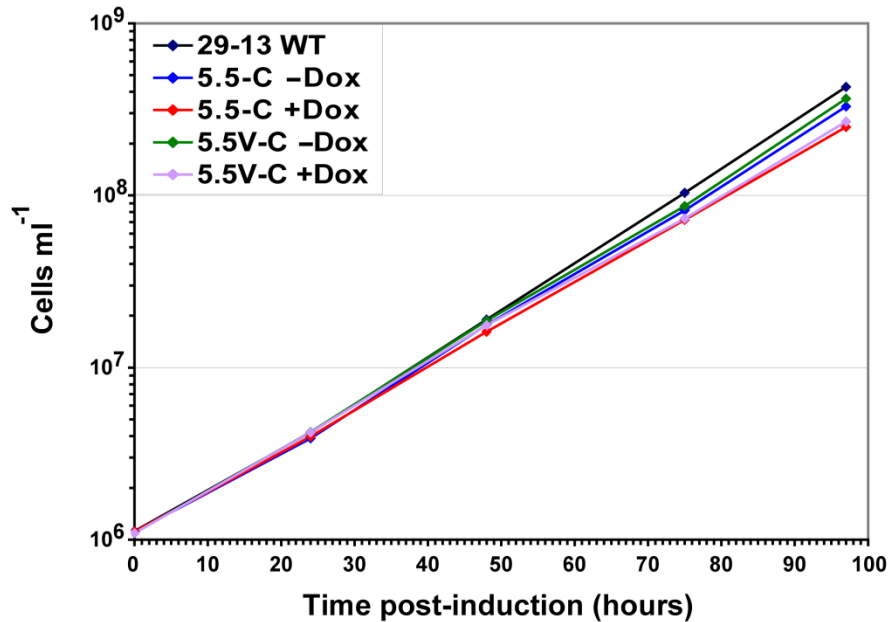
Figure 4.3. Selection of sequence fragments for RNAi functional analysis of CAP5.5 and CAP5.5V. 5.5-C and 5.5V-C RNAi constructs that target the different 3'-terminal coding sequences of each gene; 5.5-C-U and 5.5V-C-U RNAi constructs that target the different 3'-terminus of the CDS and the 3' regulatory regions of each gene; 5.5/5.5V RNAi construct that targets the N-terminus sequence shared by both genes. The bp positions on the genome locus of each DNA fragment used for RNAi are indicated in grey font; the bp positions of each fragment on the gene CDS are indicated in red font. DNA sequences are represented by solid bars and their corresponding amino acid sequences, including their domain architecture, are represented by box diagrams. CDS: coding sequences; I.R.: Intergenic Region; UTR: Untranslated Region; aa: amino acids.

approximate position of individual cells within their cell division-cycles can be readily determined by DAPI staining of mitochondrial and nuclear DNA, with normally three evident types of cells, corresponding to 1K1N, 2K1N and 2K2N configurations, in relatively constant proportions throughout the cell-cycle.

Induction of the RNAi cell line obtained using the insert that corresponds to the 3'-terminal coding region of CAP5.5, did not produce any discernible phenotype, not affecting the growth rate (Fig. 4.4A) or cell morphology. Western blot analysis using the anti CAP5.5 mAb (*Hertz-Fowler et al., 2001*) revealed no variation in the expression level of CAP5.5 between induced, not-induced cells and the 29.13 parental cell line (Fig. 4.4B). This was possibly due to the short length of the insert in this RNAi construct (375 bp, Fig. 4.3), which was probably not long enough to silence CAP5.5 expression. Similarly, induction of the RNAi cell line obtained using the insert that corresponds to CAP5.5V 3'-terminal coding region, did not produce any discernible growth (Fig. 4.5A) or morphology phenotype (Fig. 4.5B), nor did it alter the expression levels of CAP5.5 in the procyclic form (Fig. 4.6).

Figure 4.4.

A- Cumulative Growth Curve.



B- Western Blotting.

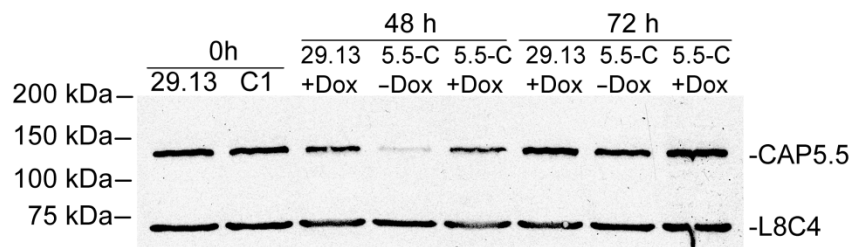


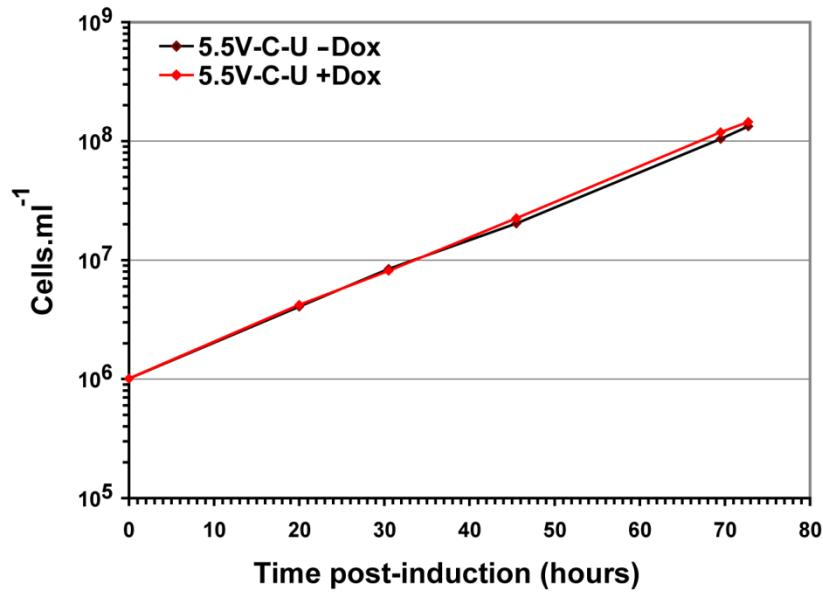
Figure 4.4. 5.5-C and 5.5V-C PCF RNAi cell lines induction: growth curve and western blot.

A. Cumulative Growth Curve: RNAi against CAP5.5 and CAP5.5V in procyclic *T. brucei* using the 5.5-C and 5.5V-C RNAi constructs that target their distinct CDS did not produce any significant growth phenotype. All cultures —29-13 (parental cell line), 5.5-C and 5.5V-C, both non-induced (-Dox) and induced (+Dox)— were re-inoculated at a cell density of 10⁶ cells.ml⁻¹ after 48h of induction, and their cumulative cell density was plotted.

B. Western blotting: Immunoblot analysis using the monoclonal CAP5.5 antibody showed no differences in the expression of CAP5.5 between induced and not-induced cultures. 2 x 10⁷ cell equivalents were loaded in each lane. Loading control for the immunoblot analysis was provided by double labelling of the nitrocellulose membrane with L8C4 monoclonal antibody (Paraflagellar Rod Protein 2) and CAP5.5 antibody. The little amount of CAP5.5 in the 48h line of the not-induced control is due to damage of the gel.

Figure 4.5.

A



B

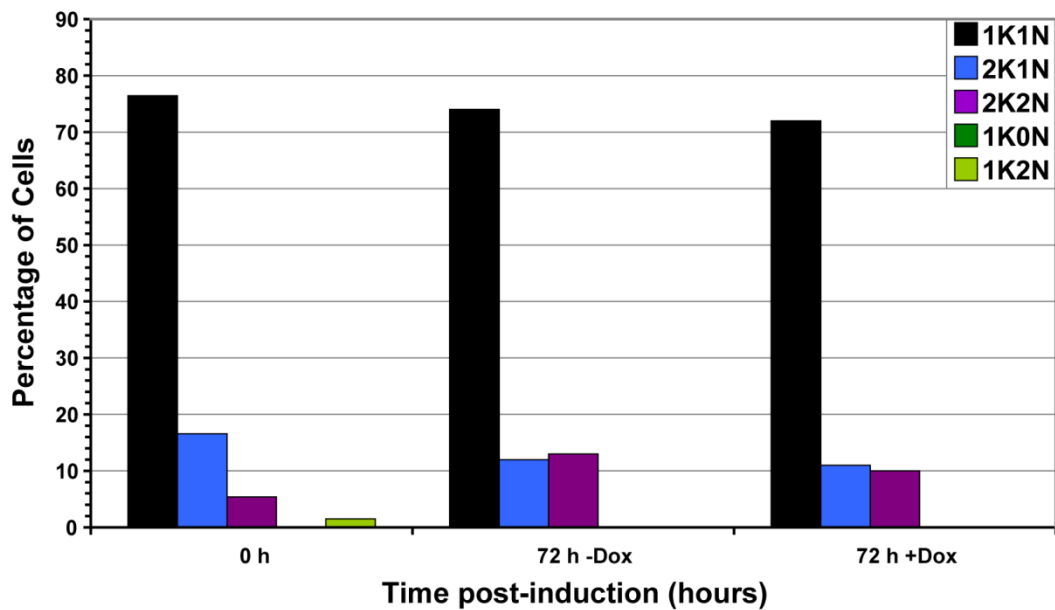


Figure 4.5. 5.5V-C-U PCF RNAi cell line induction: growth curve and percentage of cells in different position of the division-cycle.

A. Cumulative growth Curve. RNAi against CAP5.5V using the insert that corresponded to CAP5.5V 3'-terminal coding region and 3'-intergenic region did not affect the growth rate of procyclic trypanosomes. The cumulative growth curve illustrates the lack of effect of CAP5.5V RNAi induction on cell density. Both cultures, 5.5V-C-U -Dox and 5.5V-C-U +Dox, were reinoculated at 2×10^6 cells ml⁻¹ after 20 and 45.5 hours of RNAi induction and the cumulative cell densities were plotted.

B. The number of nuclei and kinetoplasts were count in DAPI-stained populations of uninduced (-Dox) and induced (+Dox) cultures at 0 and 72 hours post-induction of RNAi. No significant differences in the proportion of cell types were found between uninduced and induced cultures.

Figure 4.6.

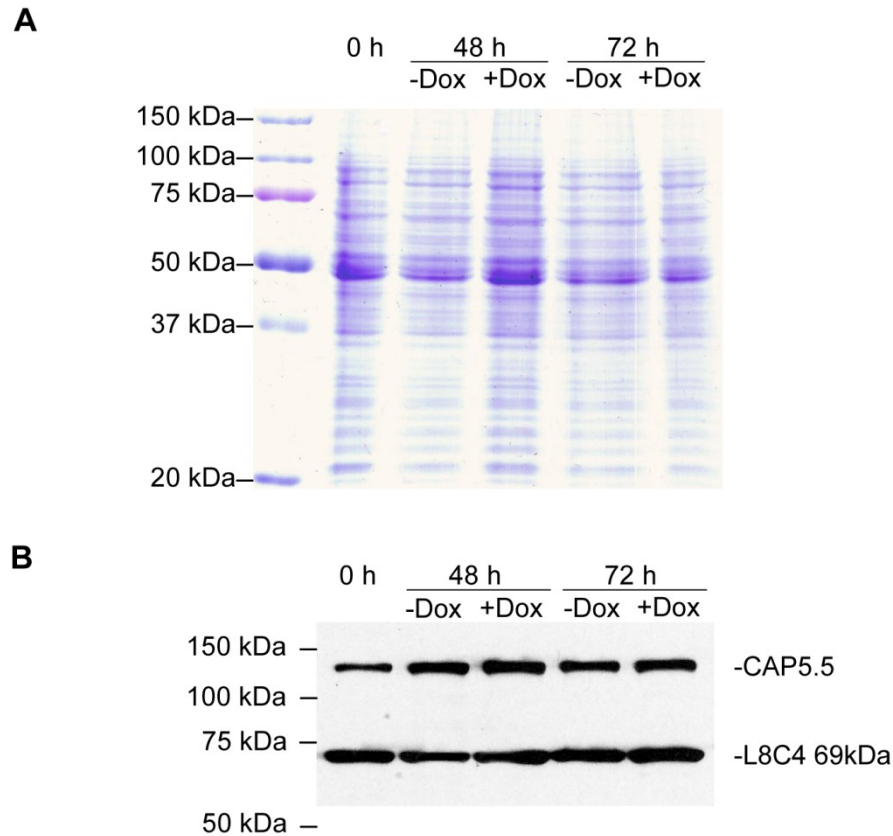


Figure 4.6. RNAi against CAP5.5V did not reduce the expression of CAP5.5 in procyclic form trypanosomes.

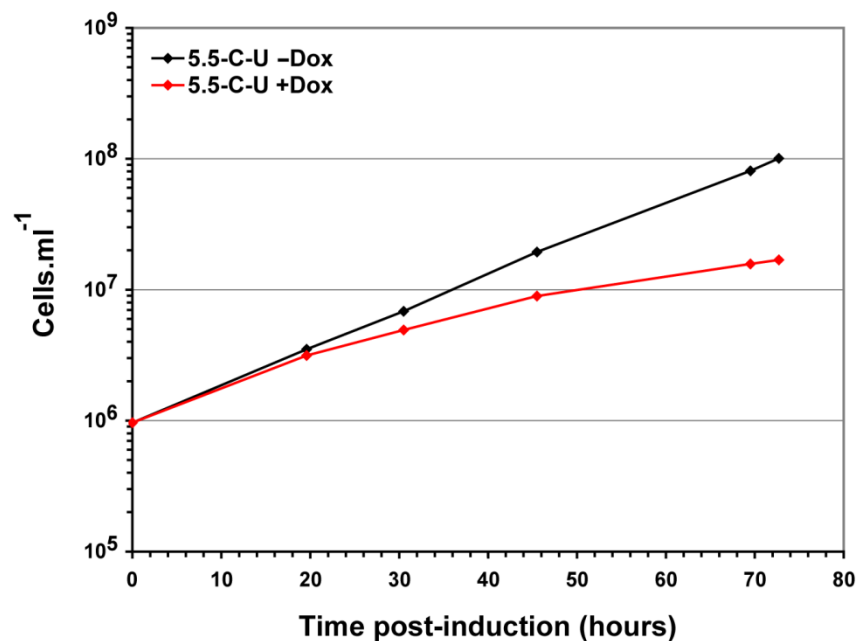
A: Protein samples from induced and not-induced 5.5V-C-U RNAi cultures were separated on a 10% SDS-PAGE gel and visualised with Coomassie blue staining. 2×10^6 cell equivalents were loaded into each lane. A similar SDS-PAGE gel was used for western blotting against CAP5.5.

B: Western blotting against CAP5.5 in protein samples from the CAP5.5.5V RNAi cell line (5.5V-C-U RNAi construct) showed no differences in CAP5.5 expression between induced and not-induced cells. 2×10^6 cell equivalents were loaded into each lane. Loading control was provided by double labelling of the nitrocellulose membrane with L8C4 (Paraflagellar Rod Protein 2) and CAP5.5 monoclonal antibodies.

4.4.1. CAP5.5 RNAi silencing in PCF produces a slow growth phenotype and aberrant cytokinesis.

In contrast to the cell line obtained with the RNAi construct that targets only the 3'-coding sequence of CAP5.5, induction of the RNAi cell line obtained with the construct that targets CAP5.5 3'-coding and regulatory regions (5.5-C-U construct) resulted in a strong reduction in the growth rate, first apparent after 24h of induction (Fig. 4.7). The loss of CAP5.5 protein in induced cells was revealed by Western blotting using the CAP5.5 mAb, with no band visible after 48h of induction (Fig. 4.8). CAP5.5 silencing also produced a strong morphology phenotype, consisting in mis-positioning of nuclei following mitosis and aberrant cytokinesis, which, by 24h post-induction, gave rise to abnormal 1K2N cells and anucleate 1K0N zoids (Fig. 4.9). Accumulation of multinucleate cells, characteristically containing a higher number of nuclei than kinetoplast ($N_n > K_n$), at later time points after induction (48 and 72h) indicates that the 1K2N progeny of aberrant cytokinesis were able to enter into further cell division-cycles. Zoids also continued to accumulate as a function of time in RNAi-induced cultures and were the most prevalent "cell" type by 72h post-induction, without a parallel increase in the number of 1K2N cells (Fig. 4.9). Since cell density always increased between the 48 and 72h time-points, albeit at a slower rate than in not-induced cultures, the distribution of cell types present in RNAi-induced cultures points towards the longevity of zoids (with presumably cannot replicate) and/or the ability of multinucleate cells to undergo cytokinesis and abscission, perhaps producing multiple zoids.

Figure 4.7.



5.5-C-U RNAi cell line		
Hours post-induction	Doubling Time	
	-Dox	+Dox
10	10,7	11,7
32,75	10,3	16,8
60,25	12,4	32,2

Average doubling time of not-induced cells: 11.1 h.

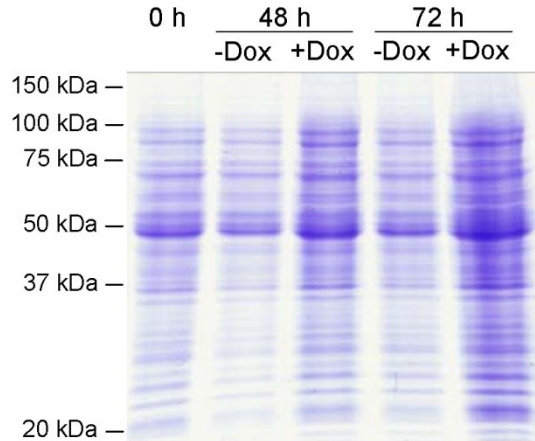
Average doubling time of induced cells: ~20 h.

Figure 4.7. 5.5-C-U PCF RNAi cell line induction. Cumulative growth curve.

Silencing of CAP5.5 expression produced a slow growth phenotype first apparent after 24 hours of induction. Both induced (+Dox) and uninduced (-Dox) cultures were re-inoculated at a cell density of 10⁶ cell ml⁻¹ after 46 hours of RNAi induction and the cumulative cell density was plotted.

Figure 4.8.

A



B

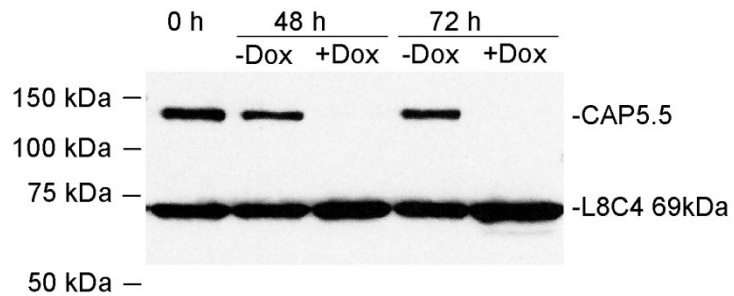


Figure 4.8. RNAi against CAP5.5 using the 5.5-C-U RNAi construct in PCF silences CAP5.5 expression.

A: Protein samples from induced and not-induced 5.5-C-U RNAi cultures were separated on a 10% SDS-PAGE gel and visualised with Coomassie blue staining.

2×10^6 cell equivalents were loaded into each lane. A similar SDS-PAGE gel was used for western blotting against CAP5.5. mAb.

B: Western blotting against CAP5.5 mAb on protein samples from the 5.5-C-U RNAi cell line showed a strong reduction of CAP5.5 expression in induced cells, with no band visible after 48h of induction. 2×10^6 cell equivalents were loaded in each lane. Loading control was provided by double labelling of the nitrocellulose membrane with L8C4 (Paraflagellar Rod Protein 2) and CAP5.5 mAbs.

Figure 4.9.

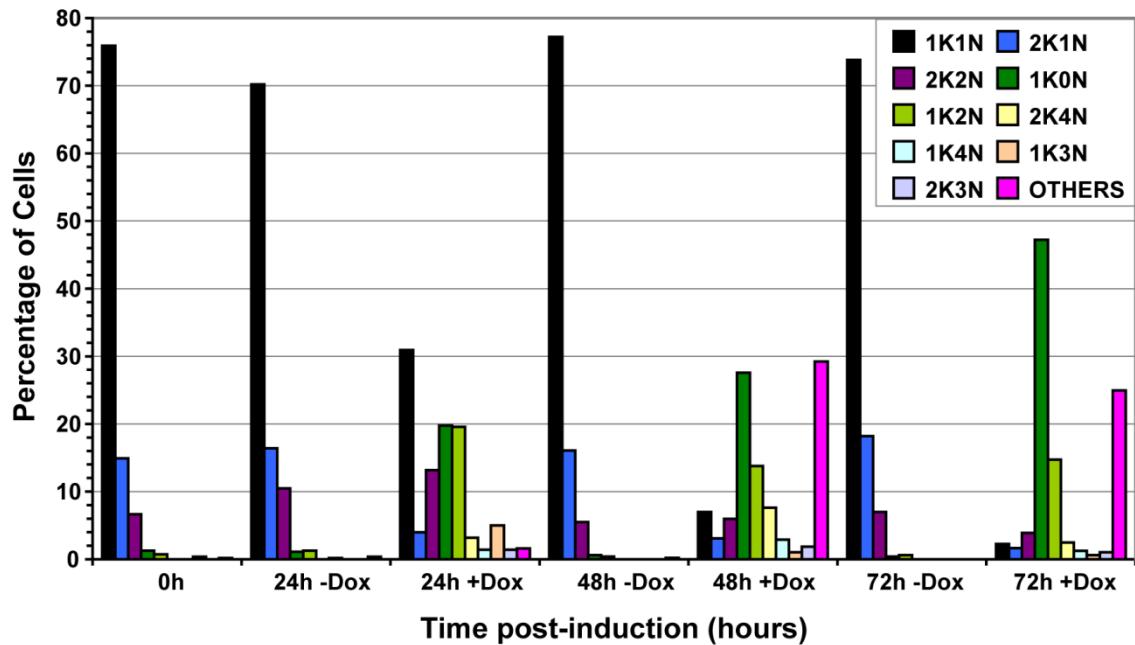


Figure 4.9. CAP5.5silencing in PCF causes aberrant cytokinesis.

Counts of the number of nuclei and kinetoplasts in DAPI-stained populations of uninduced (0h, -Dox) and induced (+Dox) cultures after 0, 24, 48 and 72h of induction of RNAi. The number of abnormal 1K2N cells and 1K0N "zoids", products of aberrant cytokinesis, rose dramatically and proportionally after 24 hours of induction. Numbers of multinuclear cells, named here as "others", which characteristically contained a higher number of nuclei than kinetoplasts, increased remarkably after 48 hours of induction, suggesting that the 1K2N progeny of aberrant cell division were able to enter into further cell division-cycles but failed to complete normal cytokinesis. Zoids continued to accumulate over time and by 72 hours of induction they were the most prevalent "cell" type (~48%). These data, coupled with the slight increase in cell density between 48 and 72 hours of induction, point to zoid longevity and/or the ability of multinuclear cells to undergo cytokinesis and abscission, giving rise to multiple zoids.

4.4.2. CAP5.5 is essential for procyclic *T. brucei* correct morphogenic patterning during the cell division-cycle.

Immunofluorescence (IF) analysis of the CAP5.5-3'-UTR RNAi cell line using the CAP5.5 mAb showed the maintenance of the regular CAP5.5 staining in uninduced cells, evenly distributed across the subpellicular microtubules of the cell body, but absent from the flagellum, (Fig. 4.10), (*Matthews and Gull, 1994a; Hertz-Fowler et al., 2001*). However, upon induction, IF analysis of induced cells, using the CAP5.5 mAb, revealed the loss of CAP5.5 from their posterior-ends after only 8h of induction (Fig. 4.11 and 4.12C-G). The cell-cycle time for not-induced cultures was ~11h (Fig. 4.7), thus 2K1N and 2K2N 8h induced cells with posterior loss of CAP5.5 (Fig. 4.11 and Fig. 4.12F-G) would have been at an earlier stage in the same division-cycle (either 1K1N or 2K1N, respectively) at the point when RNAi was induced. At latter points in the cell-cycle, the loss of CAP5.5 from the posterior cell-end correlated with the aberrant positioning of the nuclei following mitosis (Fig. 4.12F-G). In normal 2K2N cells (Fig. 4.12B) the posterior nucleus sits between the divided and segregated kinetoplasts, whereas in 2K2N 8h CAP5.5 RNAi induced cells the ingression of the cleavage furrow indicate mis-positioning of the nuclei, which are situated in more anterior locations than in normal cells. Despite nuclear mis-positioning, cytokinesis still occurred in RNAi induced cells, giving rise to 1K2N and anucleate zoids (Fig. 4.9).

In 24h induced cultures, 2K2N cells with aberrant organelle positioning and undergoing asymmetric cytokinesis were readily observed (Fig. 4.13 and 4.14). As these cells continued their progression through the division-cycle, CAP5.5 loss became more evident, with disappearance of CAP5.5 protein from the posterior half of the cell body (Fig. 4.14A and Fig.4.15A). When these cells enter cytokinesis, the position of the cleavage furrow indicates that the posterior daughter cell, which almost completely lacks CAP5.5 protein, would not contain a nucleus (Fig. 4.14B-C), whilst the anterior

Figure 4.10.

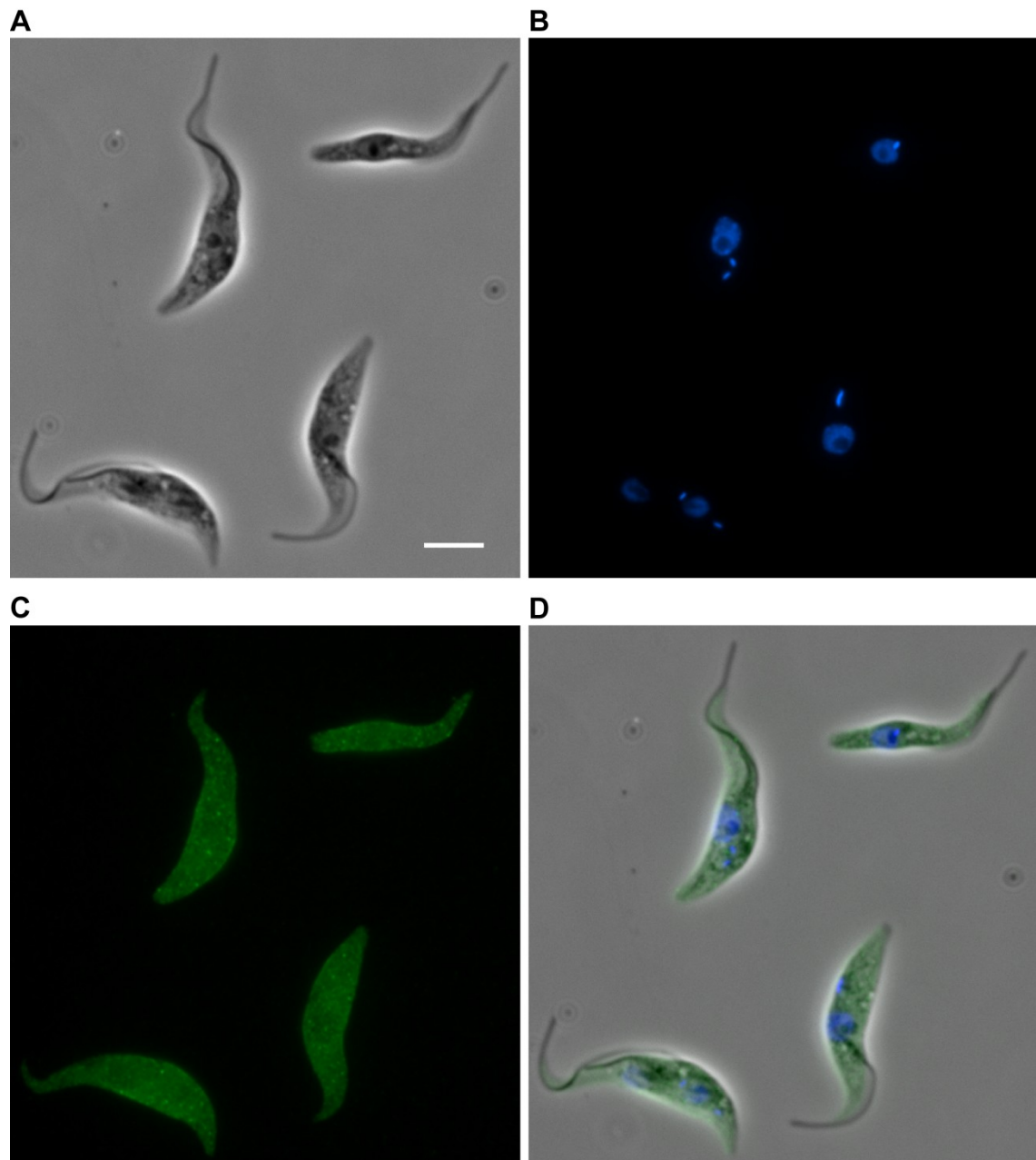


Figure 4.10. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. Not-induced.

Phase contrast (A), DAPI (B), CAP5.5 mAb (C) and merged (D) 100x images.

CAP5.5 is evenly distributed across the sub-pellicular microtubules of whole cells during the different stages of procyclic *T. brucei* cell division-cycle: 1K1N cells (top and bottom right), 2K1N cell (top left) and 2K2N cell (bottom left).

Scale bar represents 5 μ m.

Figure 4.11.

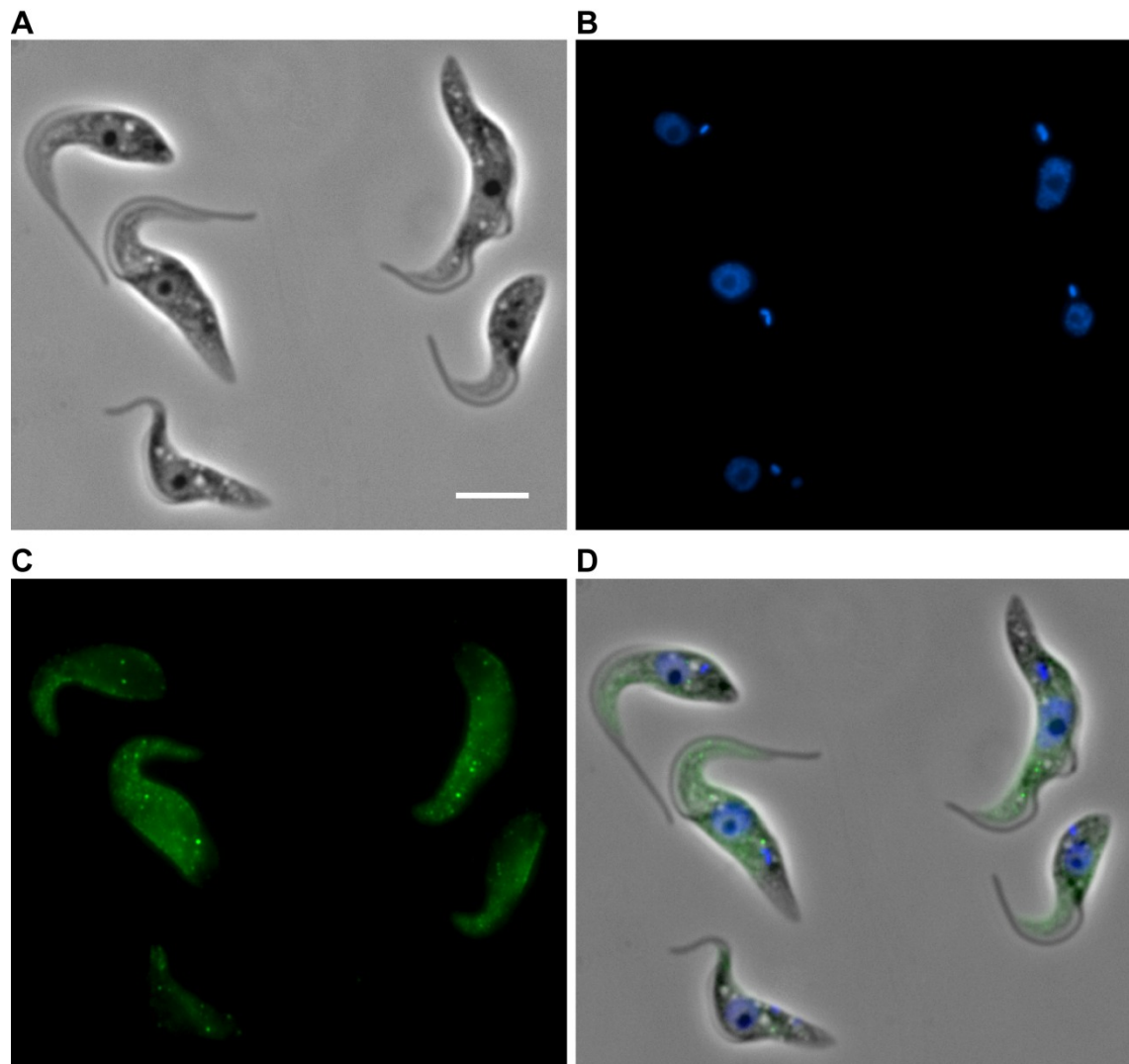


Figure 4.11. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 8h post-induction.

Phase contrast (**A**), DAPI (**B**), CAP5.5 mAb (**C**) and merged (**D**) 100x images.

After 8 hours of CAP5.5 RNAi induction, 1K1N and 2K1N (bottom) cells exhibited loss of CAP5.5 from their posterior ends.

Scale bar represents 5 μ m.

Figure 4.12.

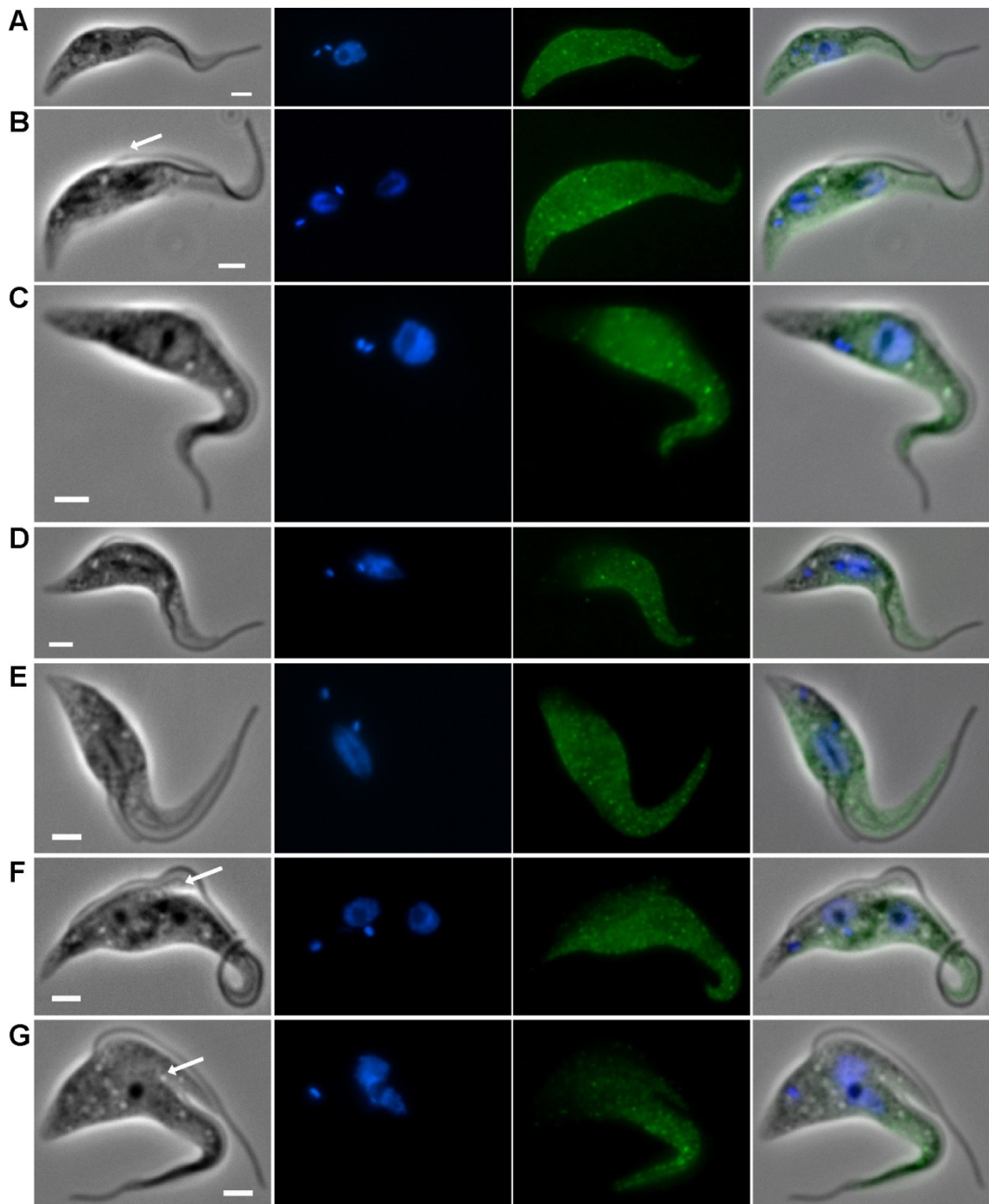


Figure 4.12. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 0 and 8h post-induction.

Phase contrast, DAPI, CAP5.5 mAb and merged 100x images.

A and B, 0h. Normal 2K1N and 2K2N cells with CAP5.5 evenly distributed across the sub-pellicular cytoskeleton. Note the position of the posterior nucleus in the post-mitotic 2K2N cell (B) where it sits between the divided and segregated kinetoplasts.

C to G, 8h. Cells after 8h of induction in different stages of the cell division cycle: 1K1N (C) with a dividing kinetoplast; 2K1N cells (D and E) with mitotic nuclei; and 2K2N cells (F and G) with mis-positioning of the dividing nuclei, as evidenced by the ingression of the cleavage furrow (arrows). All induced cells show loss of CAP5.5 at their posterior ends. In cells at late points in the cell cycle (2K2N) CAP5.5 loss correlates with aberrant positioning of the nuclei.

Scale bars represent 2 μ m.

Figure 4.13.

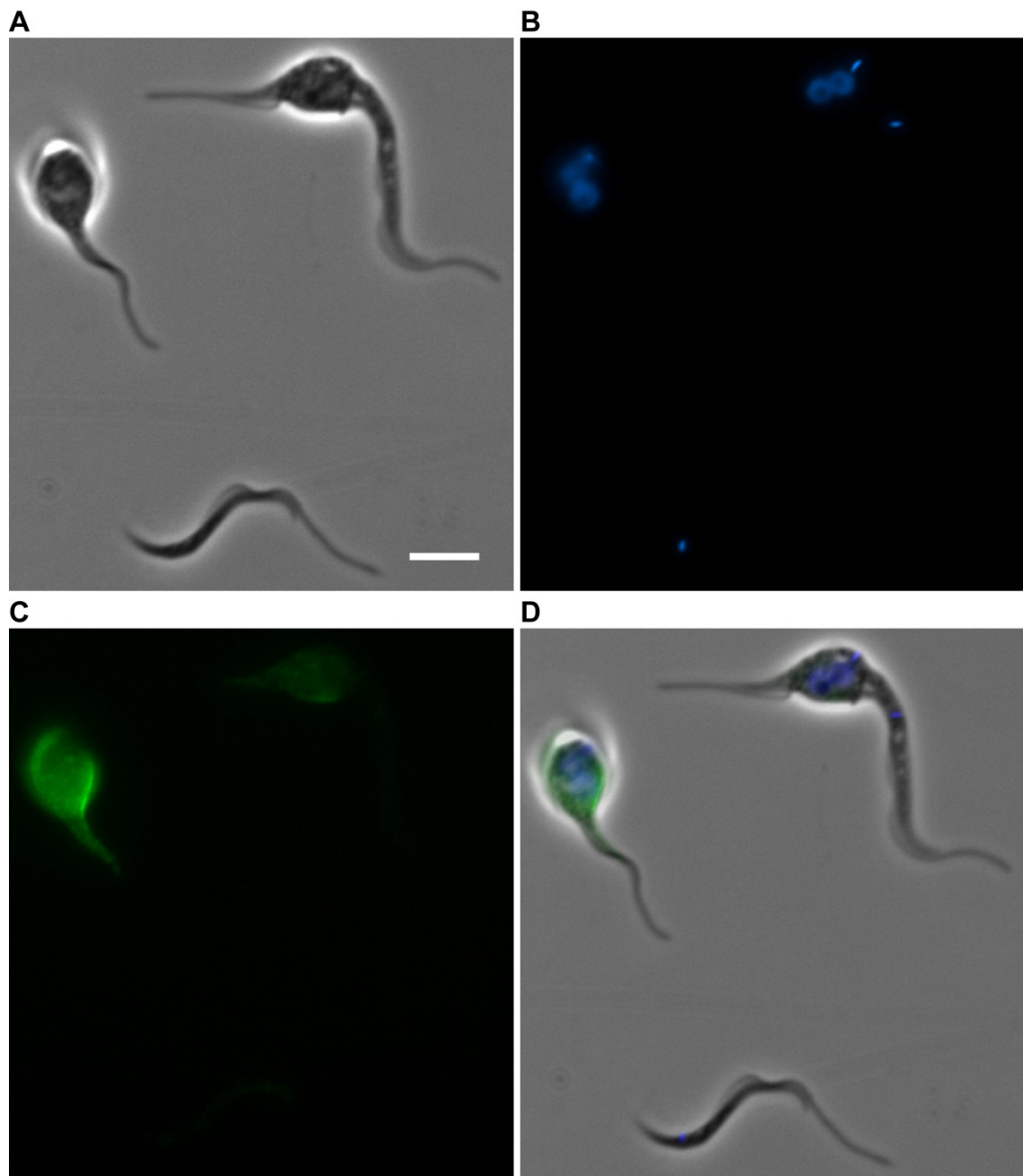


Figure 4.13. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 24h post-induction (1).

Phase contrast (**A**), DAPI (**B**), CAP5.5 mAb (**C**) and merged (**D**) 100x images.

Cells with aberrant organelle positioning and undergoing anomalous cytokinesis were readily observed in 24h induced cultures (2K2N cell, top right). Abnormal cells showed great loss of CAP5.5 from the side that presumably possesses the new flagellum, the original posterior end. The predicted cytokinesis progeny of these cells, which would include an anucleate zoid (bottom) and a 1K2N cell (top left) were also abundant. Zoids contained little or no CAP5.5 protein, suggesting that they are the posterior product of asymmetric cell division. The 1K2N progeny still contained CAP5.5 at the anterior end. Scale bar represents 5 μ m.

Figure 4.14.

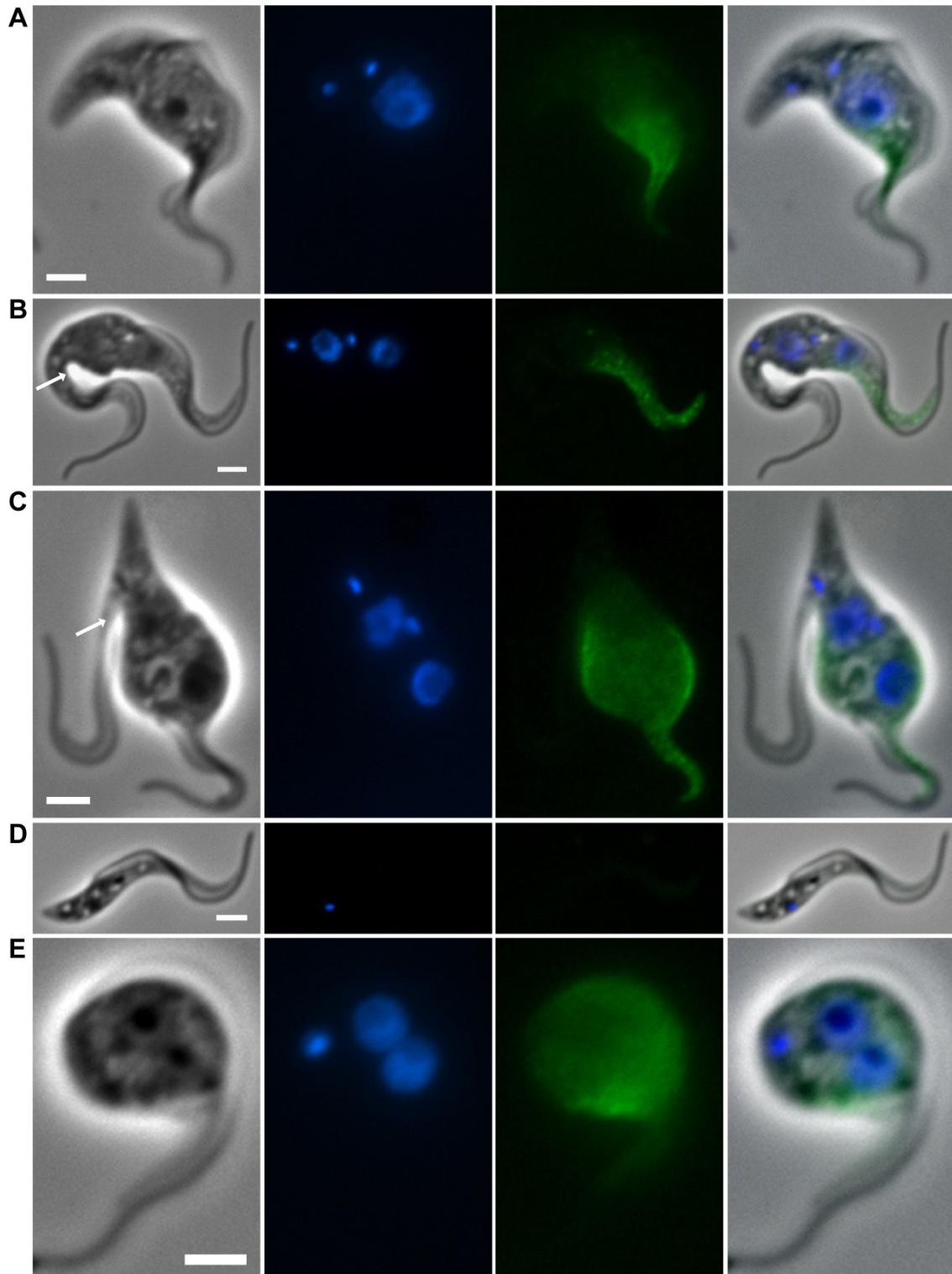


Figure 4.14. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 24h post-induction (2).

Phase contrast, DAPI, CAP5.5 mAb and merged 100x images.

A. Abnormal 2K2N cell with aberrant positioning of the nuclei and loss of CAP5.5 from the posterior half of the cell body. **B-C.** 2K2N cells undergoing asymmetric cytokinesis, with mispositioning of the nuclei and/or cleavage furrow (arrows), and with the budding posterior sibling showing a complete loss of CAP5.5. **D.** 1K0N, anucleate zoid with marginal CAP5.5 staining at its anterior-end. **E.** 1K2N cell with aberrant morphology and a more intense CAP5.5 staining at its anterior-end. Scale bars represent 2 μ m.

Figure 4.15.

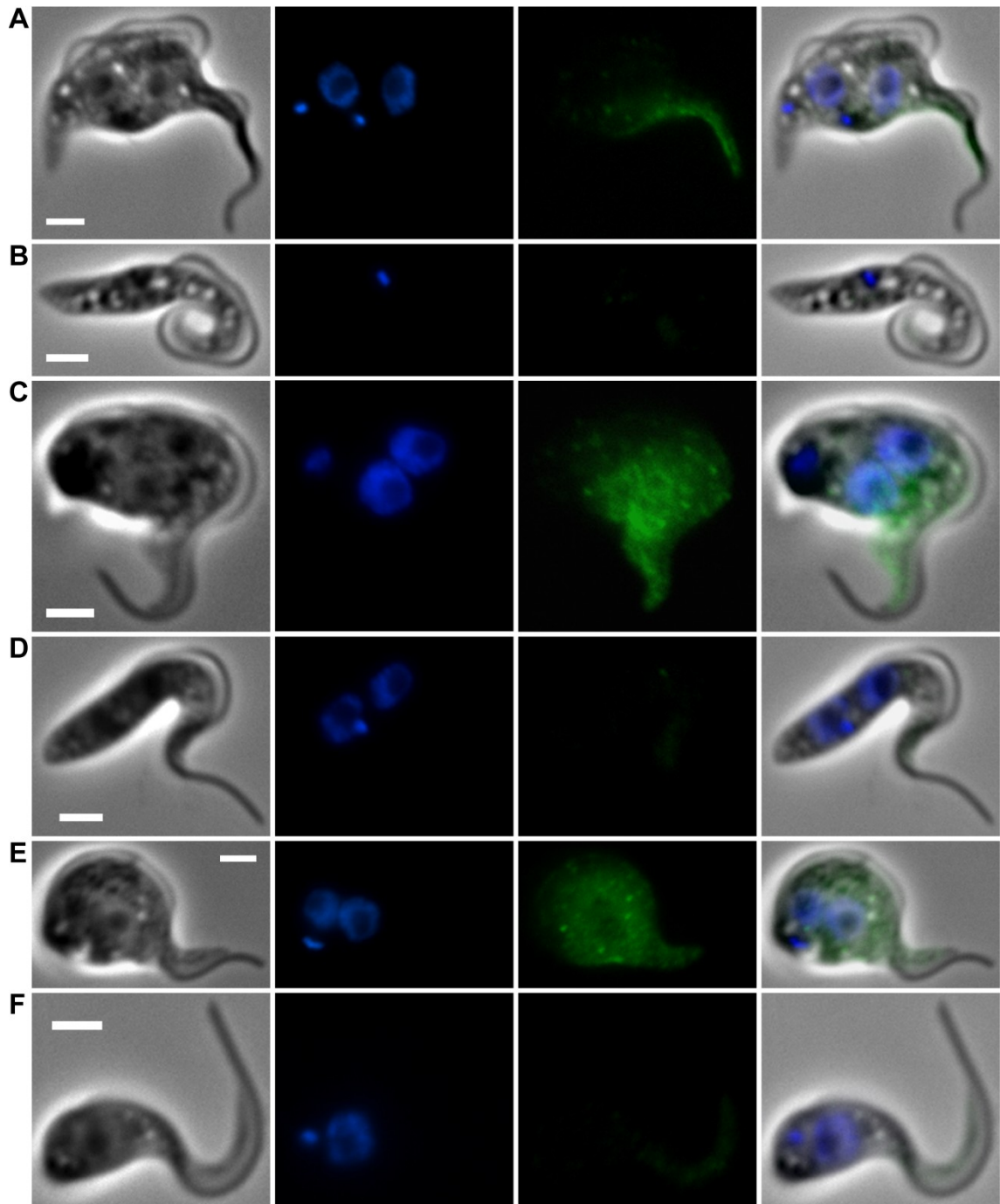


Figure 4.15. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 24h post-induction (3).

Phase contrast, DAPI, CAP5.5 mAb and merged 100x images.

A. Abnormal 2K2N cell with aberrant positioning of the nuclei and loss of CAP5.5 from the posterior half of the cell body. **B.** 1K0N anucleate zoid lacking CAP5.5 staining. **C-E.** 1K2N cells with aberrant morphology and loss of CAP5.5 from the posterior half of the cell body.

F. Abnormal 1K1N with marginal CAP5.5 staining at its anterior-end and aberrant cell-body and nucleus morphology. Scale bars represent 2 μ m.

daughter cell (which still contains CAP5.5 at its anterior end) would have an abnormal 1K2N composition. This hypothesis concurs with the observed paired increase in 1K2N and anucleate zoids in 24h induced cultures (Fig. 4.9). As expected, zoids viewed by IF using the CAP5.5 mAb contained little or no CAP5.5 protein (Fig. 4.14D and Fig. 4.15B) whereas 1K2N cells still possessed CAP5.5 at their anterior ends (Fig. 4.14E and Fig. 4.15C-E). 1K1N cells continued to be the most prevalent cell type after 24h of RNAi induction, but in some cases (5% of total 1K1N), they showed loss of CAP5.5 protein and abnormal morphology (Fig. 4.15F). These abnormal 1K1N cells are probably the progeny of 2K2N cells that were at a late stage of the division-cycle when RNAi was induced, giving rise to 1K1N cells with little loss of CAP5.5 protein. These 1K1N cells would have been able to progress through the next division-cycle and complete “normal” cytokinesis, despite CAP5.5 loss.

After 48h of induction the most prevalent cell type in induced cultures were multinucleate cells (29-35%), closely followed by zoids (28-32%, Fig. 4.9 and Appendix 4, Fig. A.4.7). Accumulation of multinucleate cells was good evidence that the 1K2N progeny of aberrant cytokinesis were able to enter into further cell division-cycles. These multinucleate cells have, in their majority, a higher number of nuclei than kinetoplasts ($n^N > n^K$) and many presented thin flagellate extensions, which did not normally contain nuclei (Fig. 4.16 and 4.17). The nuclei and kinetoplast of these anomalous cells often showed abnormal morphology, size and/or intensity of DNA staining, suggesting that the loss of CAP5.5 also affects organelle segregation (Fig. 4.17). Additionally, the CAP5.5 content of multinucleate cells was much reduced and often concentrated at one cell side (Fig. 4.16). The increase in the number of 1K0N zoids was not paired to a proportional increase in the number of their putative 1K2N siblings (28% and 14% of “cell” types, respectively Fig. 4.9). This disparity, coupled with the growth, albeit slow, of induced cultures between the 48 and 72h time-points

Figure 4.16.

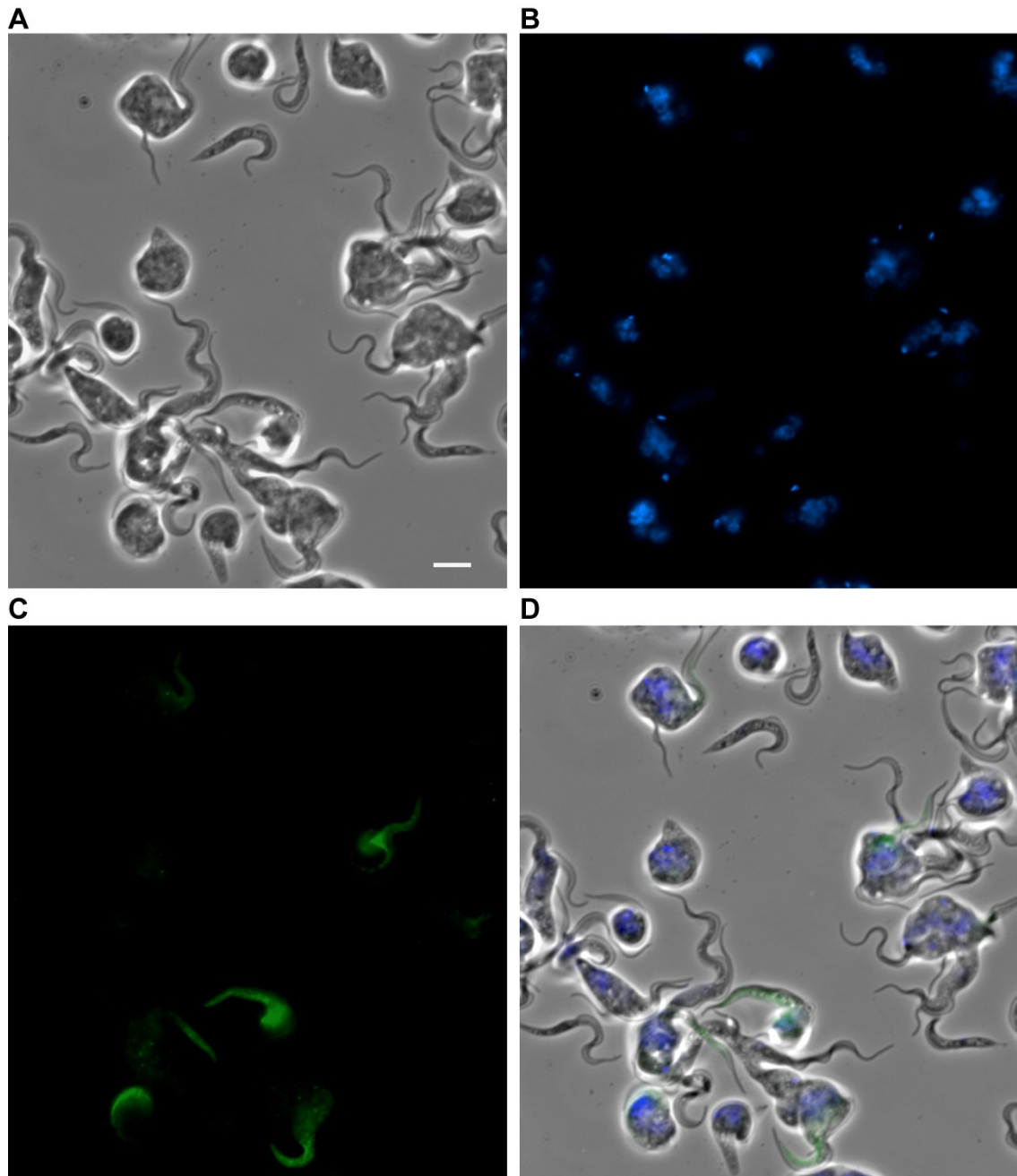


Figure 4.16. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 48h post-induction (1).

Phase contrast (A), DAPI (B), CAP5.5 antibody (C) and merged (D) 100x images.

In 48h induced CAP5.5 RNAi cultures the predominant cell types are multinuclear cells closely followed by zoids. Multinuclear cells contain, in their majority, a higher number of nuclei than kinetoplasts and often have thin flagellate extensions. The rise in the number of zoids is not accompanied by a proportional rise in the number of their 1K2N siblings. These data points either towards the longevity of zoids, which presumably cannot replicate, and/or towards the ability of multinuclear cells to complete cytokinesis and abscission, giving rise to zoids. CAP5.5 contain is greatly reduced in both zoids and multinuclear cells, but whereas multinuclear cells still present a discernible CAP5.5 signal, this is completely lost in zoids.

Scale bar represents 5 μ m.

Figure 4.17.

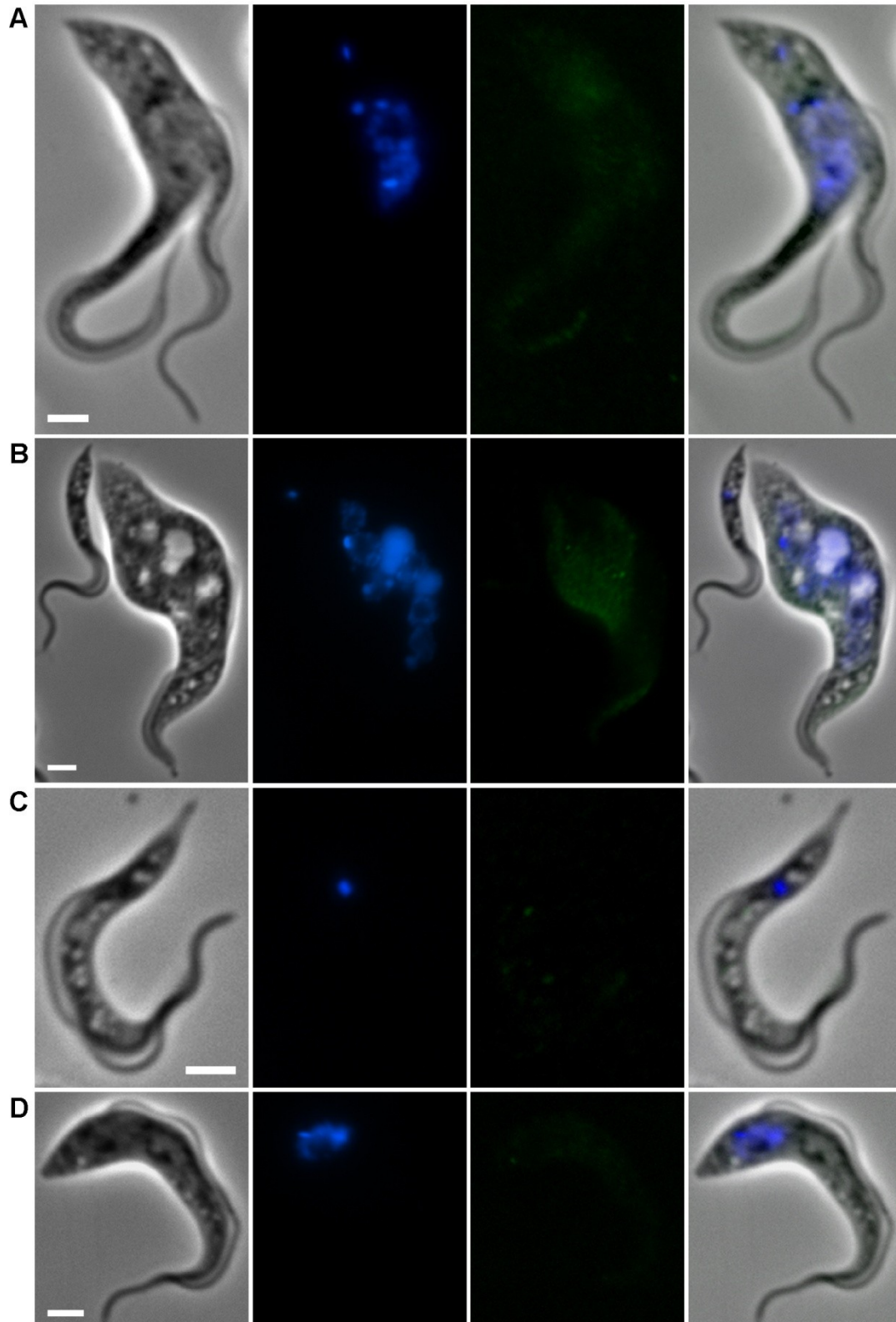


Figure 4.17. immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 48h post-induction (2). Phase contrast, DAPI, CAP5.5 mAb and merged 100x images.

A: 4K2N cell, probably the result of a 2K2N cell, which has undergone mitosis but has been unable to complete cytokinesis, note the abnormal structure of the nuclei. **B:** Multinuclear cell (right) with a higher number of nuclei than kinetoplasts, zoid (left). **C:** Zoid. **D:** 2K1N. Zoids and 2K1N cells contain little or no CAP5.5 protein. Scale bars represent 2 μ m.

(Fig. 4.7), could indicate longevity of zoids (which presumably cannot replicate) and/or the ability of multinucleate cells to undergo cytokinesis and abscission of their thin flagellate extensions, producing multiple zoids.

By 72h of induction, anucleate zoids, which have continued to accumulate as a function of time, were the most prevalent “cell” type (47%, Fig 4.9), widening the differences in number with 1K2N cells (15%, Fig. 4.9) Multinucleate cells (25%, Fig. 4.10) became larger and their number of nuclei and kinetoplasts higher. The increasing variety in nuclear and kinetoplast size, shape and DAPI staining intensity, coupled with their high numbers, made impossible to accurately quantify them by normal fluorescence microscopy (Fig. 4.18, 4.19 and 4.20). Of the much reduced number of 1K1N, 2K1N and 2K2N cells (2, 1.6 and 4% respectively, Fig. 4.9) more than half have abnormal organelle positioning, aberrant morphology and lack of CAP5.5 protein (Fig. 4.21).

4.4.3. CAP5.5 knockdown does not affect tubulin localisation.

The overall structure of eukaryotic microtubules in general and trypanosomatids in particular is that of hollow cylindrical polymers formed by 13 protofilaments of head to tail arrangements of α/β tubulin heterodimers (*Erickson, 1974*). Polymerisation of the α/β tubulin subunits confers microtubules with an intrinsic polarity. Thus, although addition of subunits is possible at both ends of a microtubule, it will always be more time efficient at the + end (*Bergen and Borisy, 1980; Margolis and Wilson, 1981*). *T. brucei* microtubules have been shown to have a defined polarity, with all microtubule plus ends orientated towards the posterior end of the cell (*Robinson et al., 1995*). Therefore, trypanosomes cell growth occurs by polymerisation of individual microtubules at the posterior cell end and also by intercalation of individual new microtubules between the existing ones (*Robinson et al., 1995*). Treatment of trypanosomes with the antimicrotubule agent rhizosin has shown to produce defects in nuclei segregation and positioning, but does not inhibit cytokinesis, causing the

Figure 4.18.

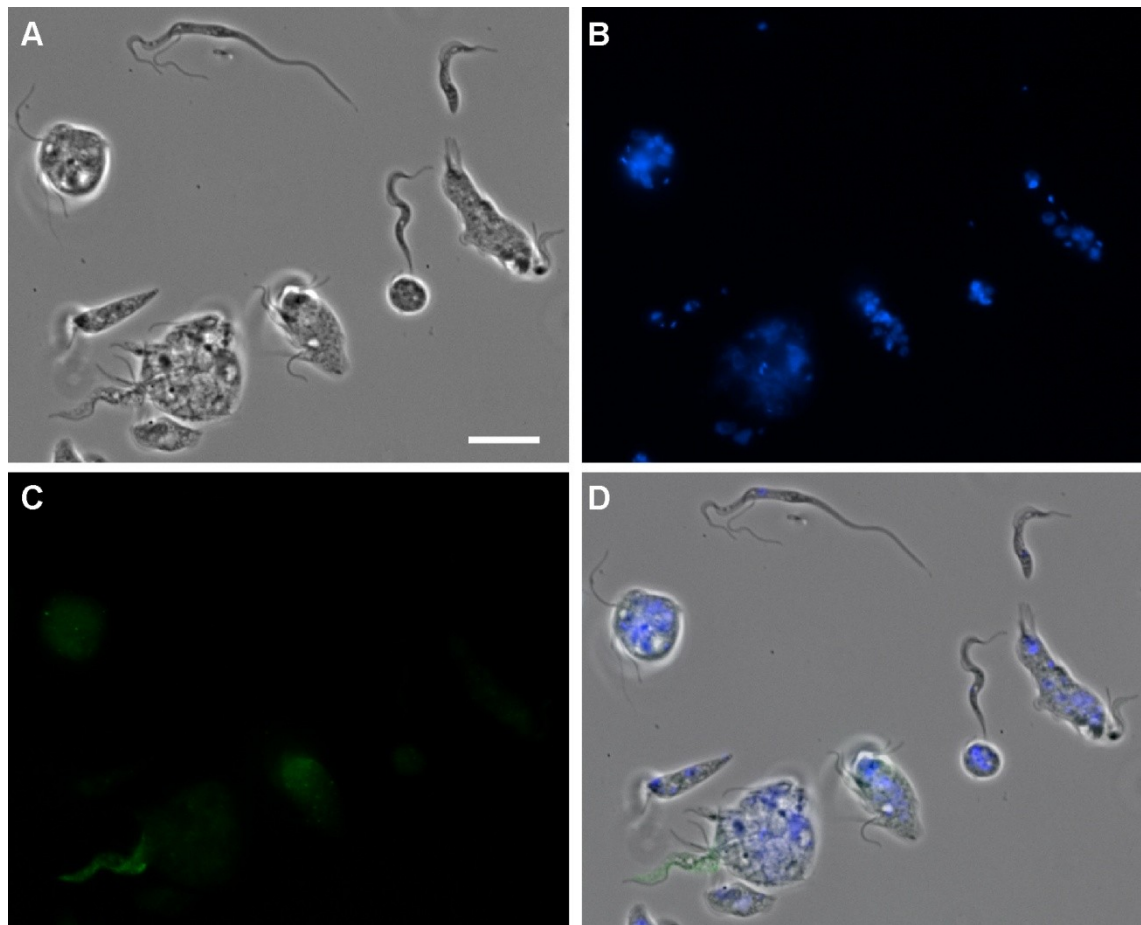


Figure 4.18. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction (1).

Phase contrast (A), DAPI (B), CAP5.5 mAb (C) and merged (D) 40x images.

Zoids and multinuclear cells are the predominant “cell” types in 72h induced CAP5.5 RNAi PCF cultures. Multinuclear cells often contain abnormal nuclei and kinetoplasts (in size, shape and intensity of DNA staining), which make very difficult to assess their numbers with certainty. CAP5.5 contain is greatly reduced in both zoids and multinuclear cells.

Scale bar represents 5 μ m.

Figure 4.19.

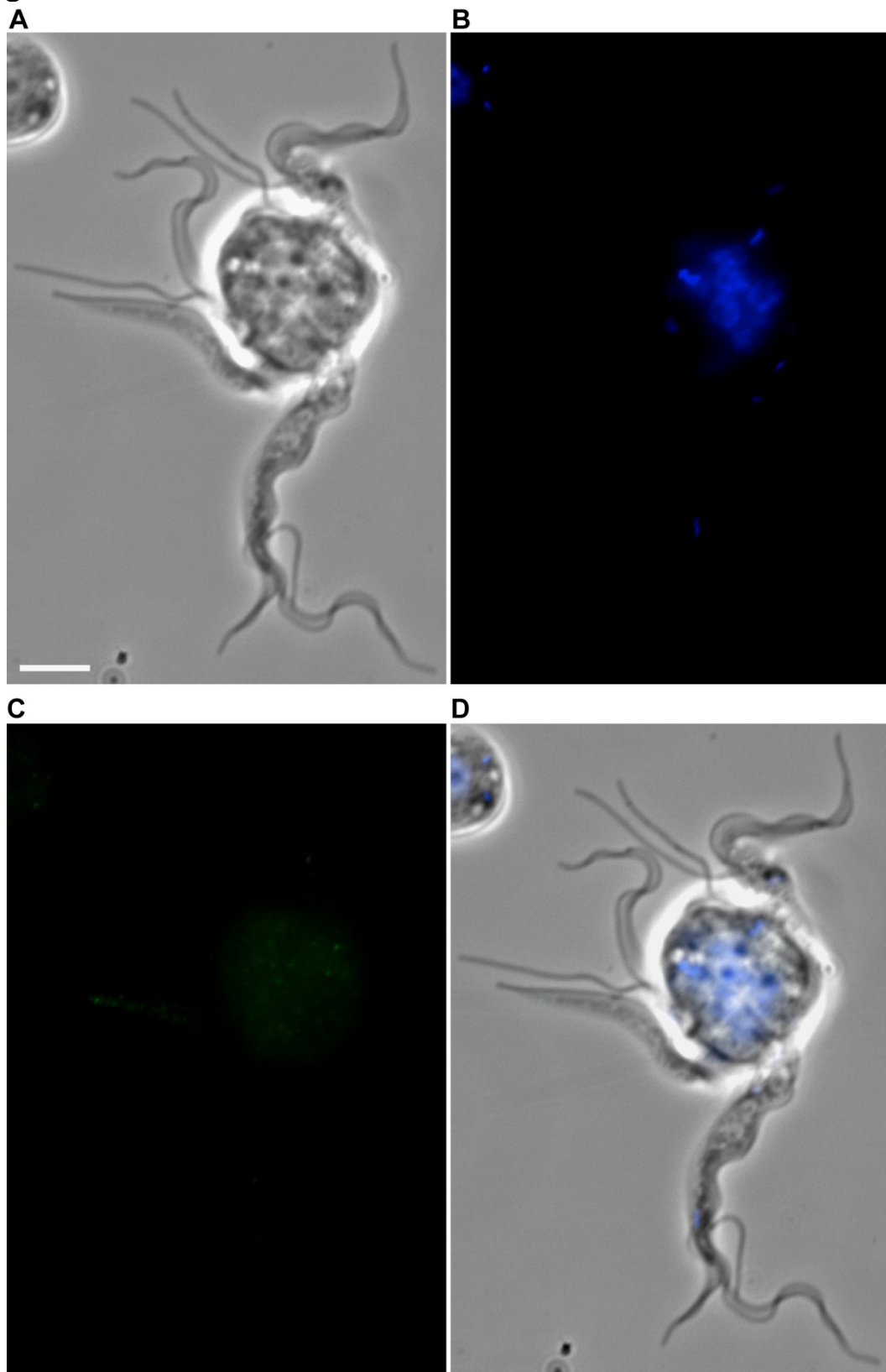


Figure 4.19. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction (2). Phase contrast (**A**), DAPI (**B**), CAP5.5 mAb (**C**) and merged (**D**) 100x images. Multinuclear cells showing the presence of multiple thin, flagellate extensions, almost completely lacking CAP5.5, and which, if able to split free, could become zoids. Scale bar represents 5 μ m.

Figure 4.20.

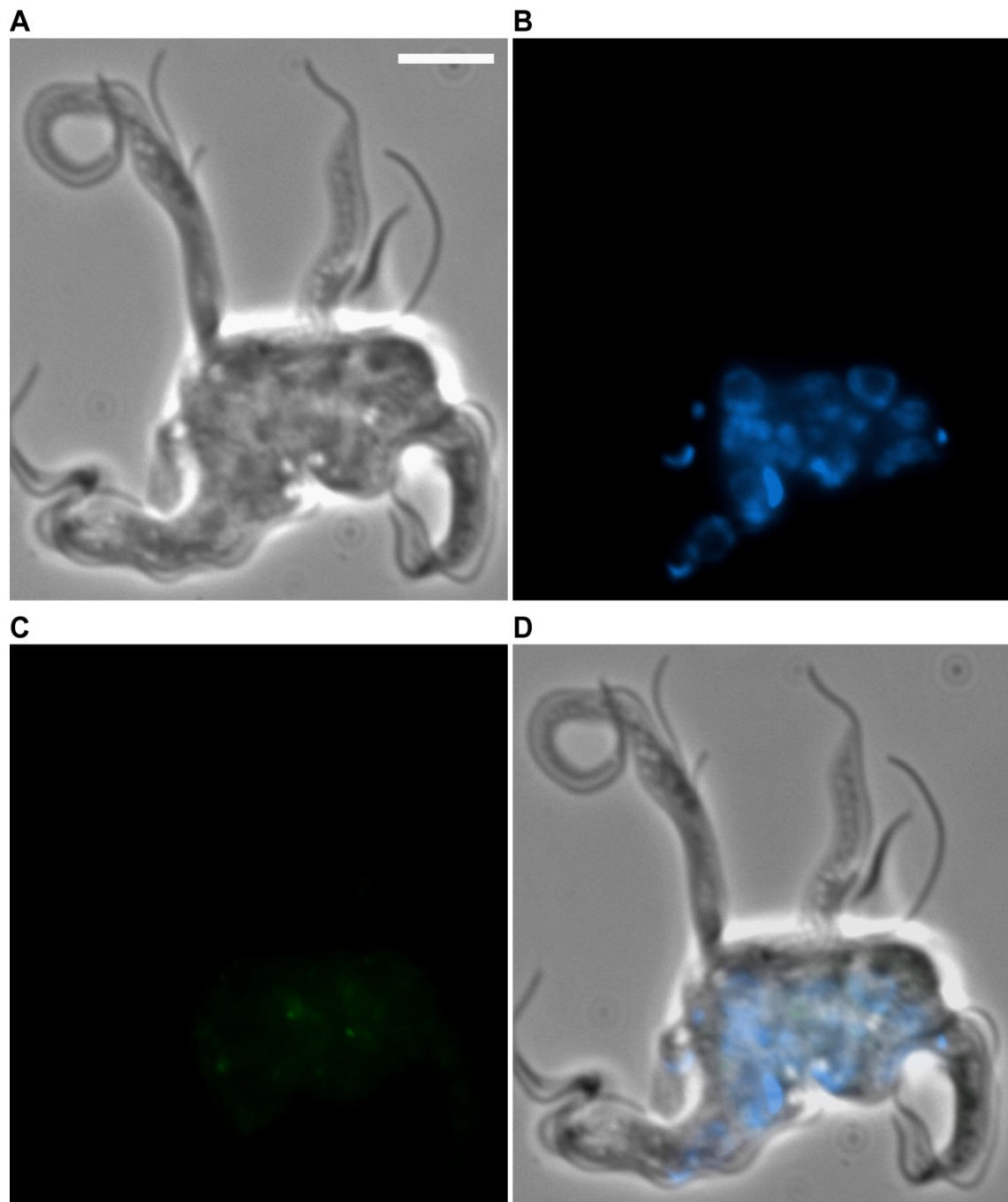


Figure 4.20. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction (3).

Phase contrast (A), DAPI (B), CAP5.5 mAb (C) and merged (D) 100x images.

72h induced multinuclear cells contain abnormal nuclei and kinetoplast, which vary in size, shape and intensity of DNA staining, being almost impossible to accurately assess their number by normal fluorescence microscopy.

Scale bar represents 5 μ m.

Figure 4.21.

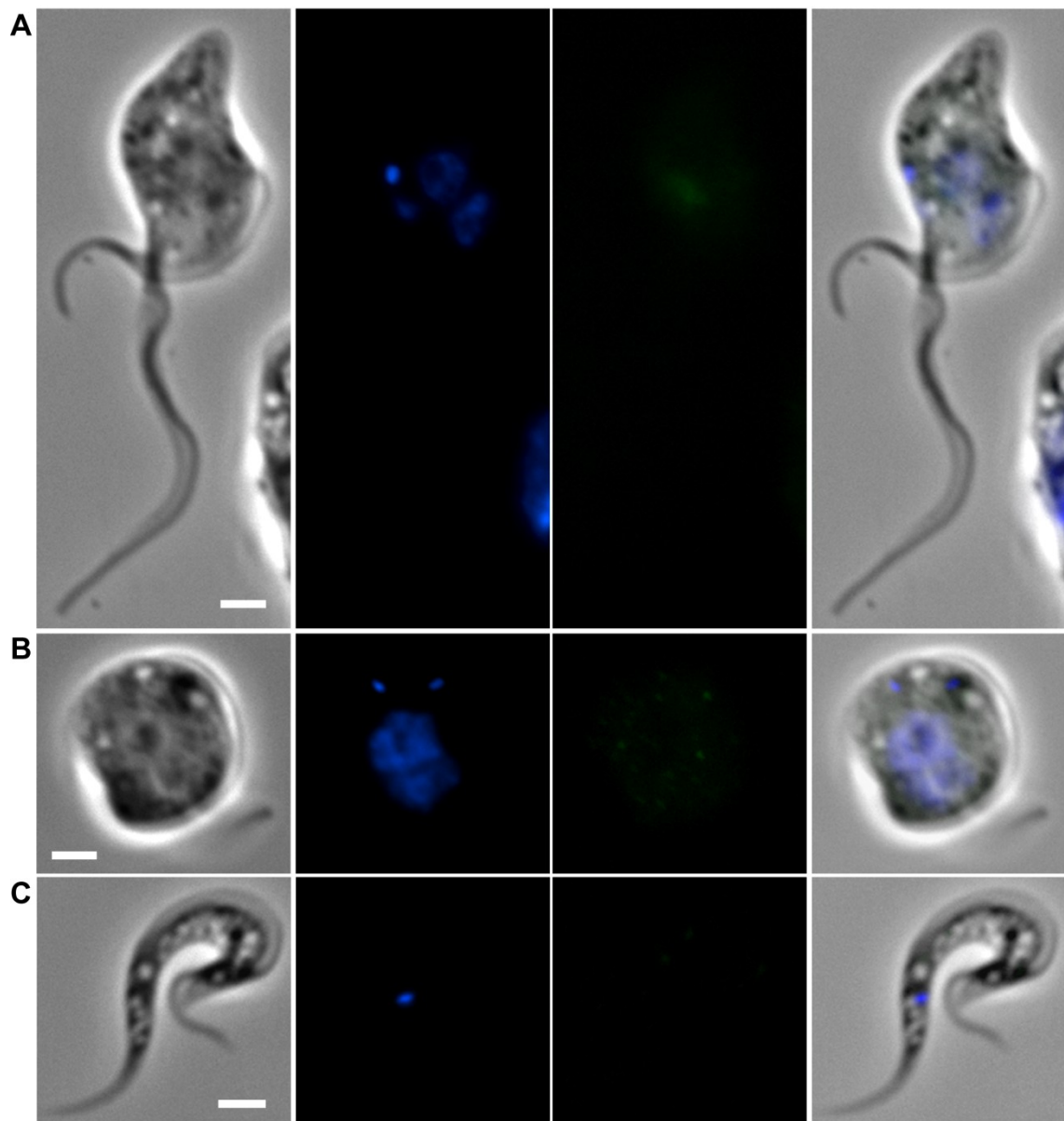


Figure 4.21. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. 72h post-induction (4).

Phase contrast, DAPI, CAP5.5 antibody and merged 100x. images.

A-B. Abnormal 2K2N cells with organelle mis-positioning and aberrant morphology.

C. 1K0N (zoid).

All "cells" contain little or no CAP5.5 protein.

Scale bars represent 5 μm.

appearance of anucleate zoids and their siblings: 1K2N and abnormal 1K1N cells (1K1N cells with elevated levels of nuclear DNA) (Robinson *et al.*, 1995). These effects are very similar to those caused by CAP5.5 silencing. Analysis of RNAi induced cultures with the anti α -tubulin antibody TAT-1 (Woods *et al.*, 1989) showed that the microtubule network basic structure was not affected by CAP5.5 loss, although it most likely affected microtubule dynamics (Fig. 4.22). In many ways, this result was not unexpected since cell morphology of RNAi induced cells was not similar to the “fat cell” phenotype that results from tubulin RNAi (Ngo *et al.*, 1998).

4.4.4. WCB localisation is unaffected by CAP5.5 silencing.

WCB is a *T. brucei* MAP that locates between the inner-leaflet of the plasma membrane and the subpellicular cytoskeleton. Recent studies (Baines and Gull, 2008) have shown that silencing of WCB by RNAi also causes aberrant cytokinesis and produce anucleate zoids and binucleated 1K2N cells. WCB RNAi induction also results in the loss of the targeted protein from the posterior end of the cells (Baines and Gull, 2008). When I probed CAP5.5 RNAi induced cells with anti-WCB-1 mAb (Woods *et al.*, 1992) there was no alteration in WCB localisation (Fig. 4.23). One of the major differences between CAP5.5 and WCB RNAi phenotypes is that RNAi against WCB, but not CAP5.5, results in the “blebbing” of plasma membrane vesicles to the extracellular environment (Baines and Gull, 2008). This suggests that an important property of WCB, but not CAP5.5, is to maintain the local plasma-membrane-to-microtubule-corset integrity. The mutant-specificity of membrane blebbing could relate to the different motifs present in each protein, which could mediate the interaction with the plasma membrane: WCB contains an N-terminal phospholipid-binding C2-like domain, whereas CAP5.5 is subject to N-terminal acylation. C2 domains are often Ca^{2+} dependant membrane targeting domains, present in proteins implicated in signal transduction and membrane trafficking (Davletov and Sudhof, 1993).

Figure 4.22.

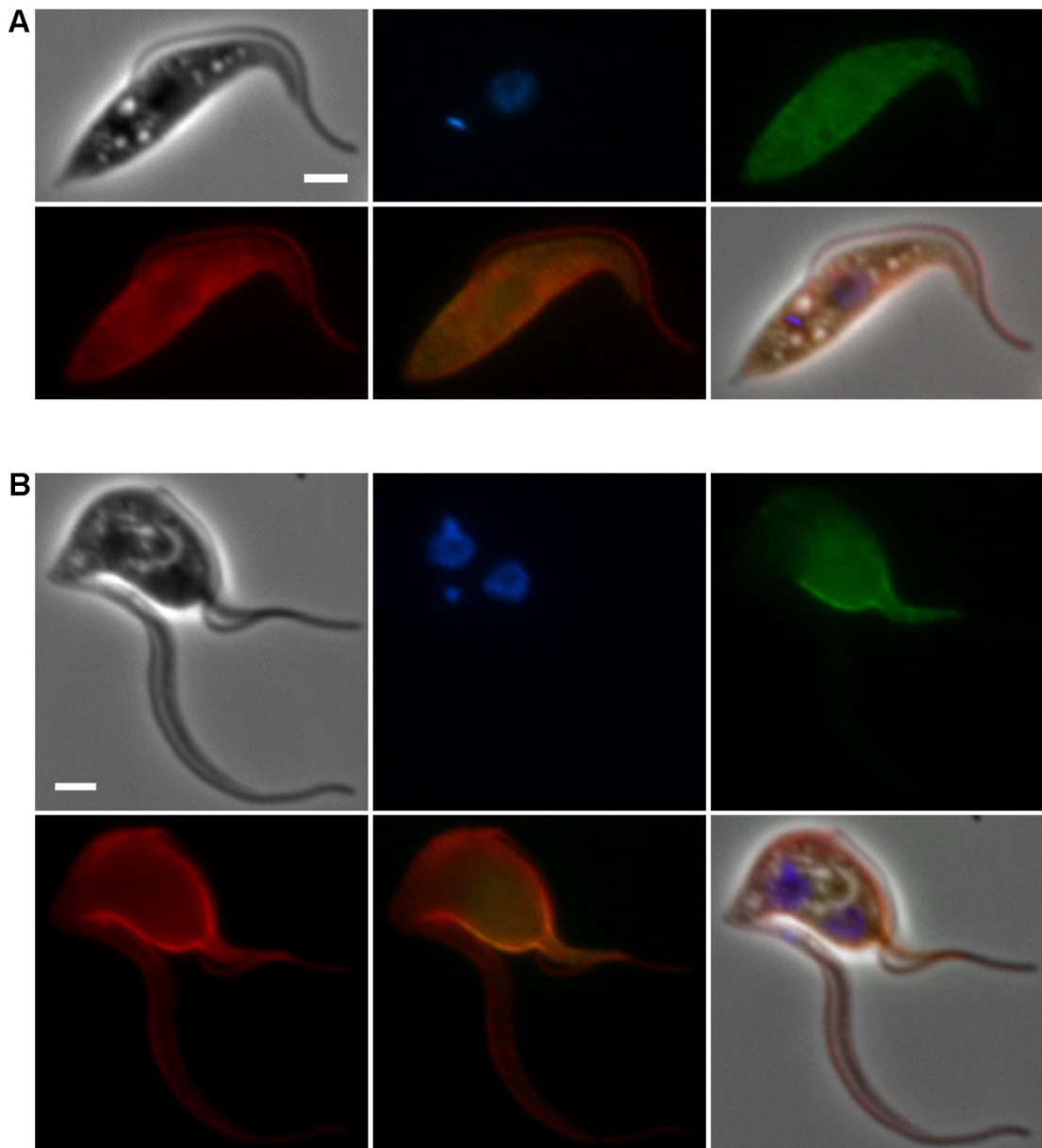


Figure 4.22. Immunofluorescence analysis of 5.5-C-U RNAi cell line. 0 and 24h post-induction. TAT-1 and CAP5.5 mAbs.

Phase contrast, DAPI, CAP5.5 mAb, TAT-1 mAb, CAP5.5 and TAT-1 superimposed images and phase contrast, DAPI, CAP5.5 and TAT-1 merged 100x images.

A. Not-induced CAP5.5 RNAi cells showed CAP5.5 characteristic labelling of the cell body but not the flagellum, and TAT-1 typical staining of the cell body and the flagellum.

B. 24h induced CAP5.5 RNAi cells showed the predicted lost of CAP5.5 from the posterior-end and abnormal morphology, but no alteration in the labelling pattern of TAT-1, which is evenly distributed across the flagellum and the cell body.

Scale bars represent 2 μ m.

Figure 4.23.

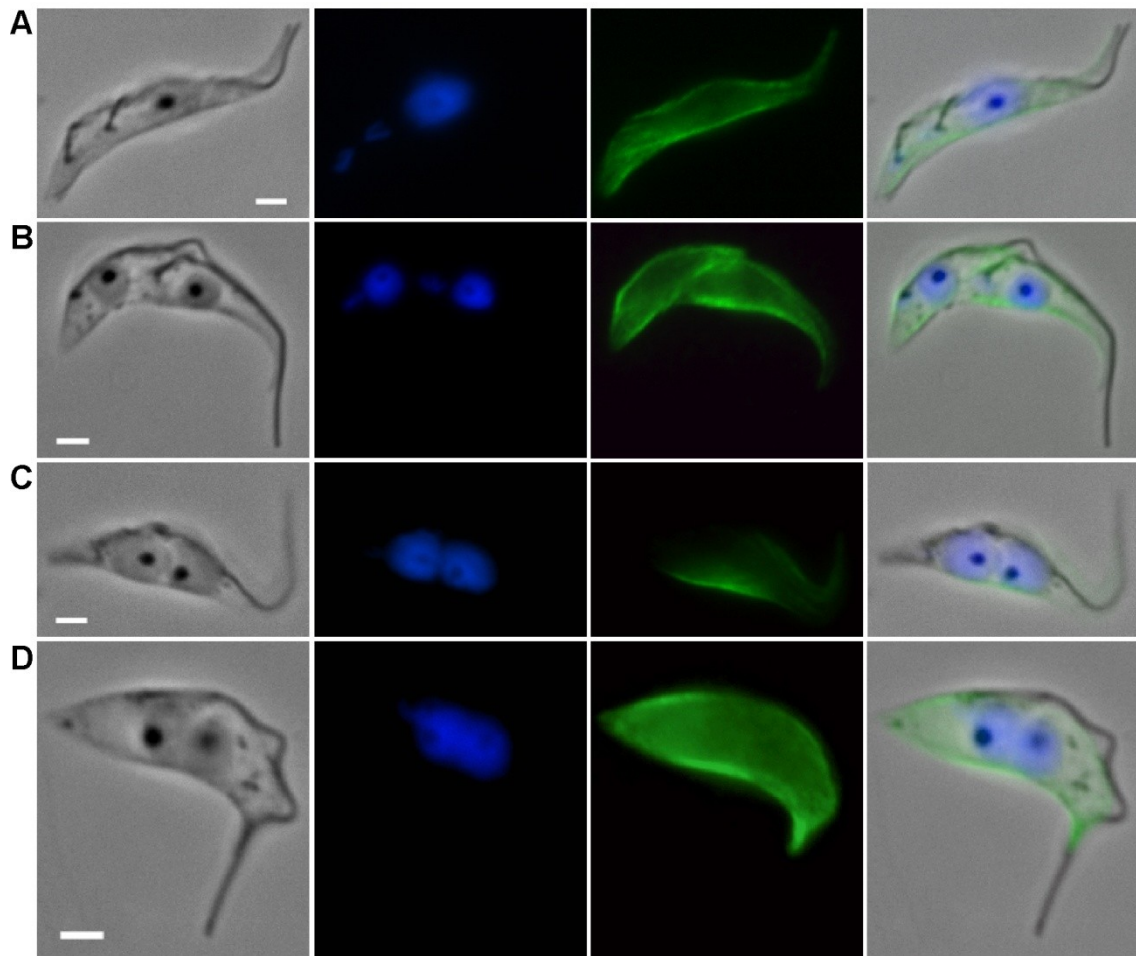


Figure 4.23. Immunofluorescence analysis of PCF 5.5-C-UR RNAi cell line. Detergent extracted cytoskeletons labelled with CAP5.5 and WCB mAbs. 0 and 24h post-induction. Phase contrast, DAPI, CAP5.5/WCB mAb and superimposed 100x images.

A-B. Not-induced. CAP5.5 RNAi cytoskeletons.

A. CAP5.5 mAb staining follows the helical pattern of the subpellicular microtubule corset.

B. WCB mAb has a similar staining pattern to CAP5.5, being evenly distributed across the cell-body and absent from the flagellum.

C-D. 24h post-induction. CAP5.5 RNAi cytoskeletons.

C. CAP5.5 loss from the posterior end of a 1K2N cell.

D. WCB localisation is not affected by the loss of CAP5.5 in induced cells.

Scale bars represent 2 μ m.

4.4.5. CAP17 and CAP5.5 may have competing, antagonistic functions.

T. brucei CAP15 and CAP17 are closely related stage-specific, cytoskeletal-associated proteins (Vedrenne *et al.*, 2002), encoded by paralogous genes located in different contigs of chromosome 11. CAP17 is only expressed in procyclic form trypanosomes, whereas CAP15 is preferentially expressed in bloodstream forms (Vedrenne *et al.*, 2002). CAP15 and CAP17 localise at the anterior cell-end, the less dynamic half of the subpellicular corset, and are absent from the flagellum, basal body and mitotic spindle. Both proteins stabilise microtubules. CAP15 and CAP17 possess subterminal hydrophobic domains but required their entire structure for microtubule binding. When overexpressed in *T. brucei*, both CAP15 and CAP17 localised across the whole subpellicular corset and induced morphological disorders, cell-cycle based abnormalities and asymmetric division (Vedrenne *et al.*, 2002). Analysis of the aberrant bloodstream and procyclic populations overexpressing CAP15 and CAP17, respectively, showed that their two major effects were inhibition of kinetoplast segregation and loss of coordination between mitosis, kinetoplast duplication and cytokinesis (Vedrenne *et al.*, 2002).

Given that CAP15/17 overexpression and CAP5.5 RNAi both result in cell-morphology phenotypes, I decided to look by immunofluorescence at CAP5.5 RNAi induced cells with both the anti-CAP17 polyclonal antibody, CAP17.A (Vedrenne *et al.*, 2002), and the anti-CAP5.5 mAb (Matthews and Gull, 1994a). In not-induced cells, these antibodies showed the expected localisation of CAP5.5 across the whole cell-body and CAP17 confinement to the anterior half of the cell (Fig. 4.24A). However, after 48h of induction, CAP5.5 RNAi induced population showed alterations in CAP17 localisation (Fig. 4.24B). CAP5.5 depletion did not affect the ability of CAP17 to associate to the microtubule corset. On the contrary, the lost of CAP5.5 seemed to cause the spreading

Figure 4.24.

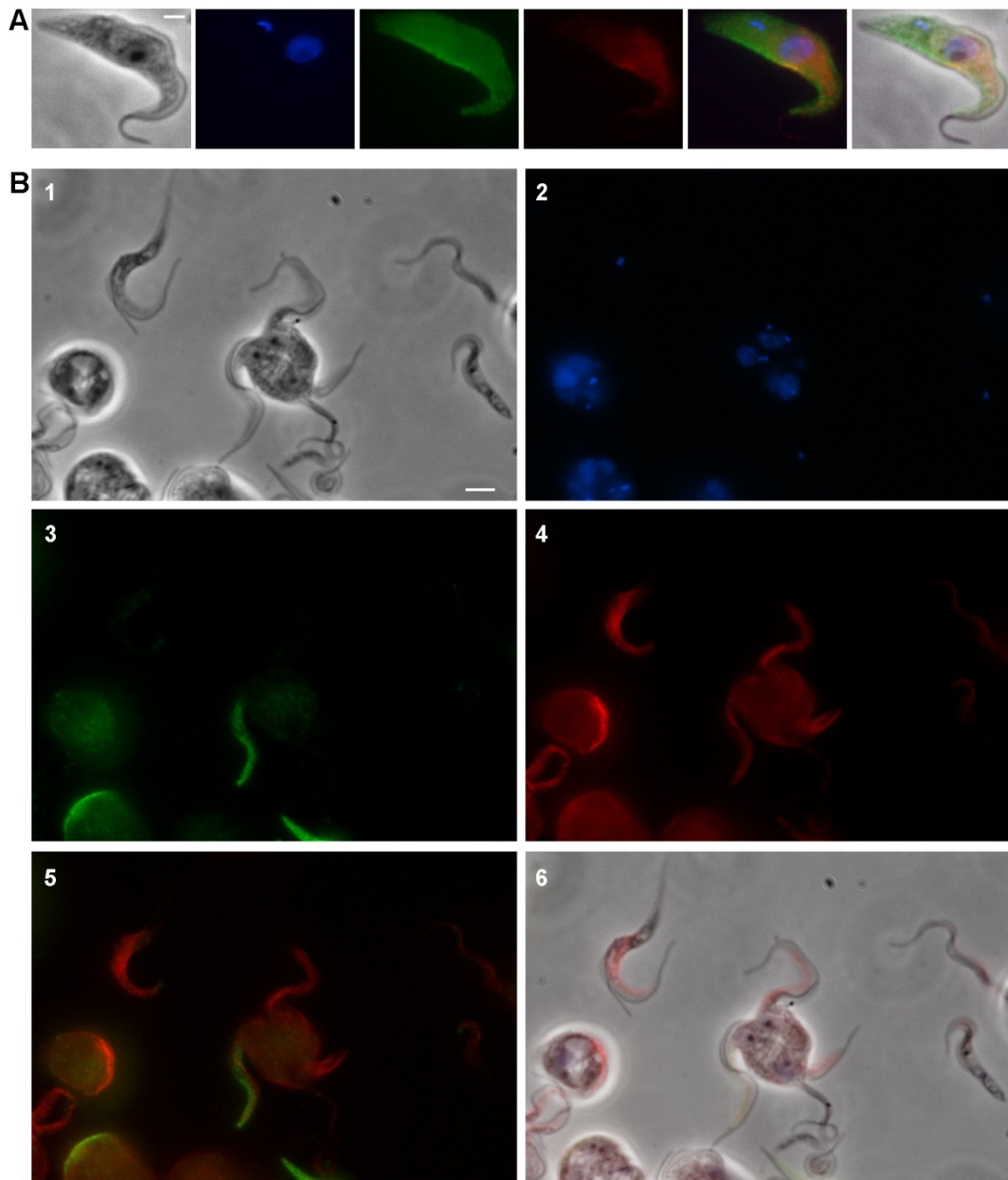


Figure 4.24. Immunofluorescence analysis of PCF 5.5-C-U RNAi cell line. Whole cells labelled with CAP5.5 mAb and CAP17.A polyclonal antibody. 0 and 48h post-induction. Phase contrast (1), DAPI (2), CAP5.5 mAb (3), CAP17.A antibody (4), CAP5.5/CAP17 superimposed (5) and merged (6) 100x images.

A. 0h, non-induced. CAP5.5 antibody labels the whole cell body whereas CAP17 localises only at the anterior half of the cells.

B. 48h post-induction. Loss of CAP5.5 protein does not disrupt CAP17 association to the subpellicular cytoskeleton. On the contrary, the loss of CAP5.5 seems to induce mislocalisation of CAP17, which is not longer confined to the anterior half of the cell but spreads across the whole subpellicular cytoskeleton.

Scale bars represent 5 μ m.

of CAP17 localisation from the anterior half of the cell to the whole cell body (Fig. 4.24B). Why CAP5.5 silencing effects CAP17 localisation is not known, although it could be speculated that they may compete for microtubule binding. A future experiment to investigate this hypothesis will be electron microscopy immunolocalisation of CAP5.5 and CAP17 on whole-mount cytoskeletons of CAP5.5 RNAi uninduced and induced cells (5.5-C-U RNAi cell line).

4.4.6. SEM analysis of CAP5.5 mutants gives further evidence of aberrant cell morphology and cytokinesis induced by CAP5.5 silencing.

To look further at the effect of CAP5.5 RNAi in cell morphology, uninduced and induced cultures were prepared for SEM at 0, 18, 24 and 48h post-induction of RNAi. Figure 4.25 shows examples of not-induced cells with the normal morphology of procyclic *T. brucei*. At 18h of induction, after the induced population would have completed at least one division-cycle, aberrant morphology was evidenced by the appearance of cells with abnormal posterior-ends and thin, zoid-like shaped “cells”. (Fig. 4.26). Closer examination of fixed 18 and 24h induced cultures also showed bi-flagellate cells with enlarged mid-bodies undergoing asymmetric division (Fig. 4.27-28). After 48h of induction CAP5.5 RNAi population was mainly formed by multi-flagellate cells with slender flagellate extensions, which abscission would likely yield zoids (Fig. 4.29-31).

Figure 4.25.

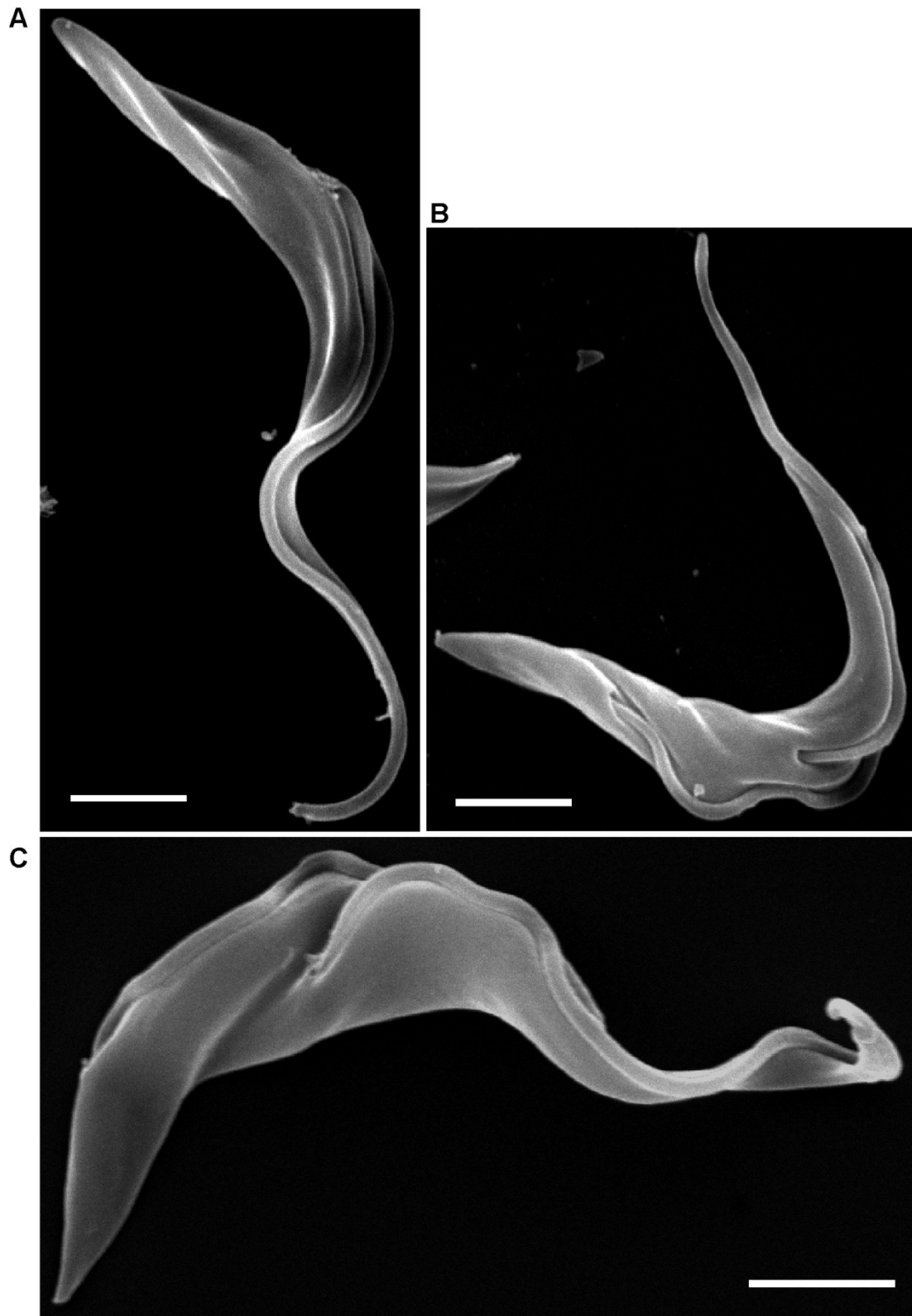


Figure 4.25. SEM of 29-13 and non-induced 5.5-C-U RNAi cell lines.

A-B. 29.13 cells in different stages of the cell division cycle. **A.** Uni-flagellate cell likely to be a 1K1N cell. **B.** Bi-flagellate cell that could be in the 2K1N or 2K2N stages. **C.** Uninduced, bi-flagellate CAP5.5 RNAi cell (5.5-C-U cell line) with the normal morphology of 2K1N or 2K2N stages. Scale bars represent 2µm.

Figure 4.26.

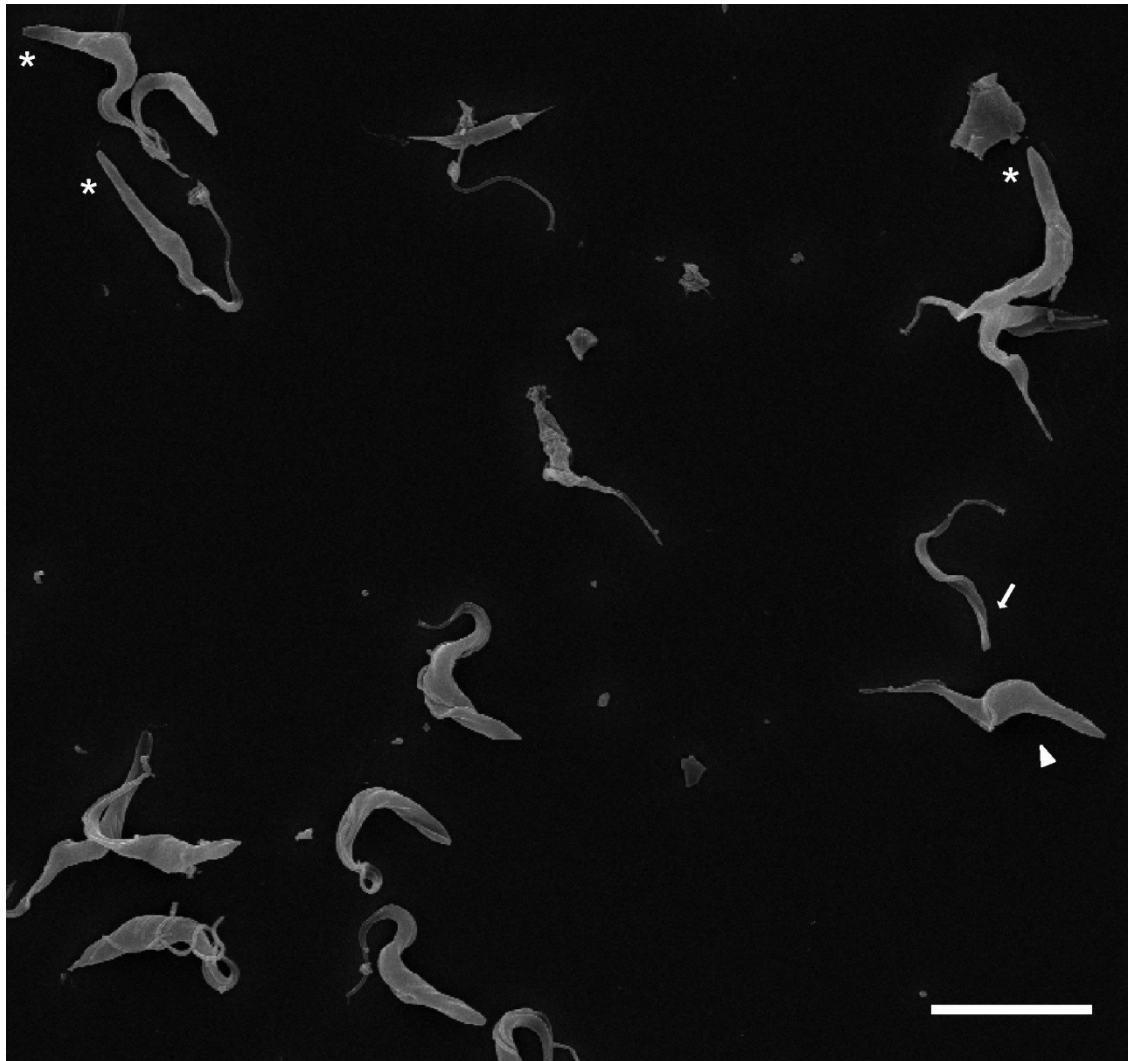


Figure 4.26. SEM of CAP5.5 RNAi cell line (5.5-C-U). 18h post-induction (1).

Scanning electron micrograph showing a representative population of CAP5.5 RNAi induced cells after 18h of induction. Note the presence of cells with abnormal posterior ends (*), enlarged mid-body (arrowhead) and "zoid-like" shape (arrow). Scale bar represents 10 μ m.

Figure 4.27.



Figure 4.27. SEM of CAP5.5 RNAi cell line (5.5-C-U). 18h post-induction (2).

A. Bi-flagellate cell undergoing asymmetric division.

B. Bi-flagellate cell, with an abnormally large posterior end and an engrossed mid-body. Inset: note the persistence of the flagella connector.

C. "Zoid like" uniflagellate, thin form.

Scale bars represent 2µm. Inset scale bar represents 500 nm.

Figure 4.28.

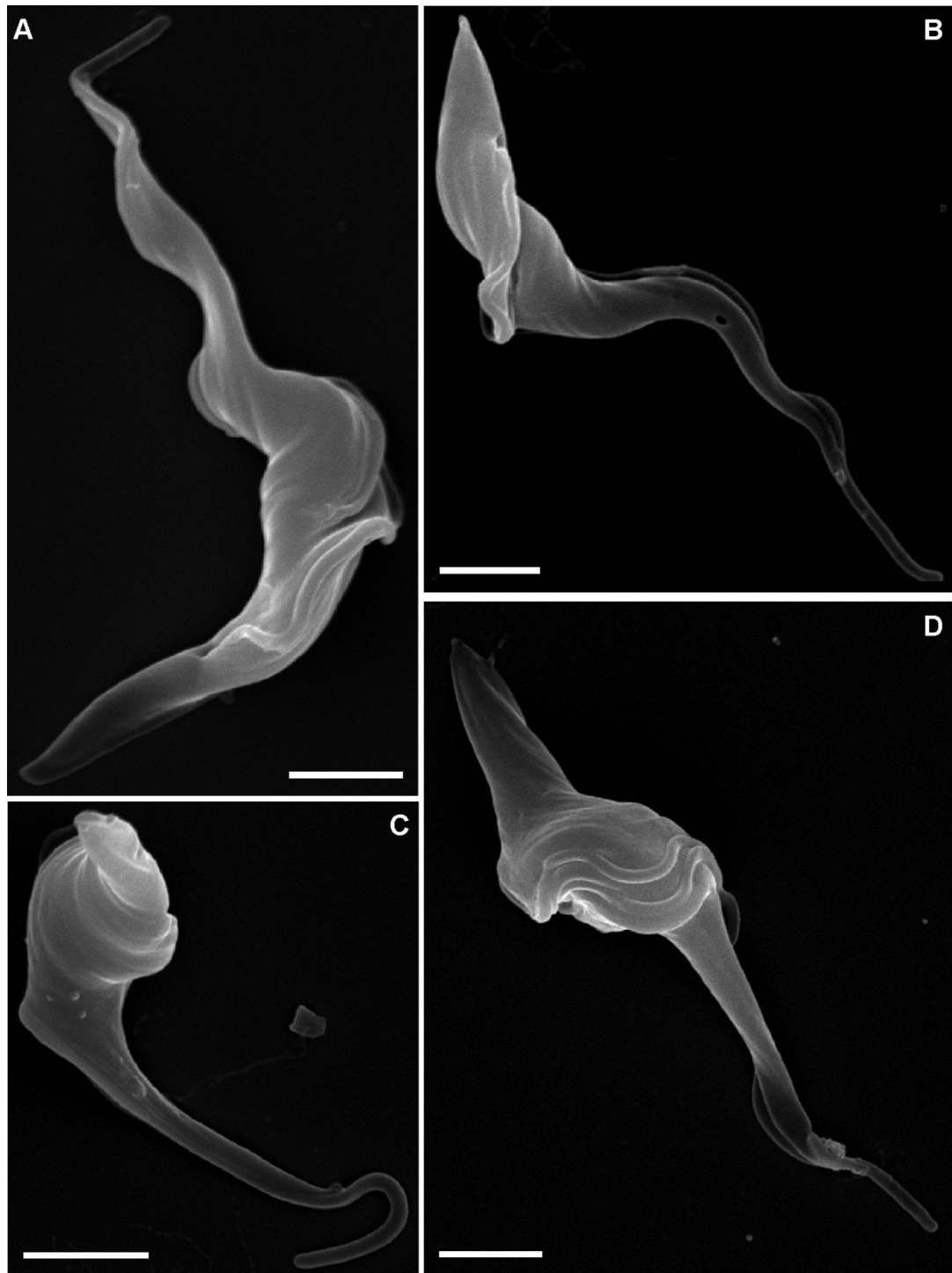


Figure 4.28. SEM of CAP5.5 RNAi cell line (5.5-C-U). 24h post-induction.

After 24h of induction, the majority of observed cells have abnormal morphology and/or are undergoing asymmetric division.

A. Bi-flagellated cell with an abnormally large posterior end. **B.** Bi-flagellated cell undergoing asymmetric cytokinesis. **C.** Cell with an abnormally wide posterior end. **D.** Bi-flagellated cell with an abnormally engorged mid-body.

Scale bars represent 2 μm .

Figure 4.29.

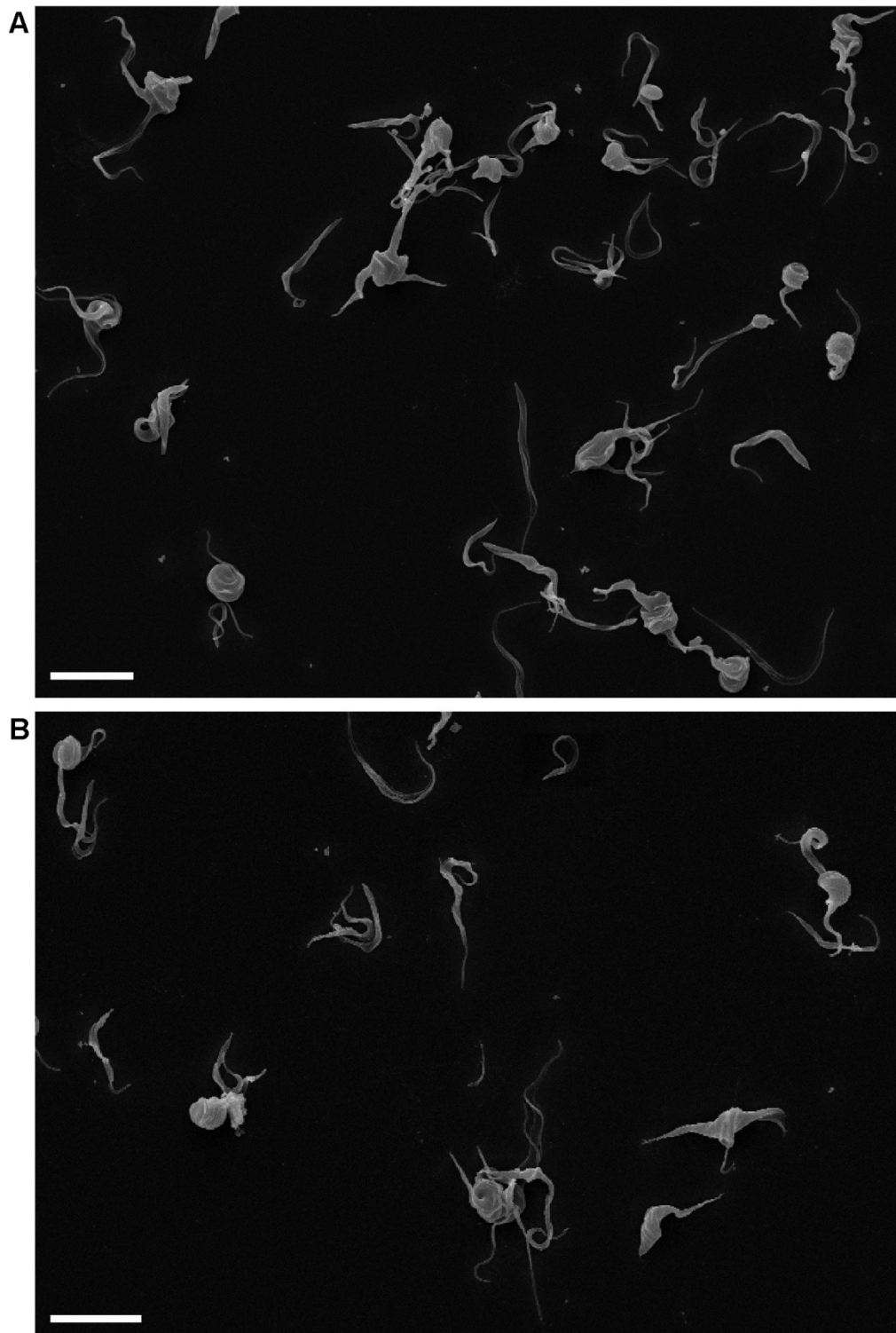


Figure 4.29. SEM of CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction (1).

Low magnitude scanning electron micrographs of CAP5.5 RNAi cultures after 48h of induction. The majority of observed cells were multi-flagellate giant cells and very thing, flagellate "zoid like" forms. Scale bars represent 10 μ m.

Figure 4.30.

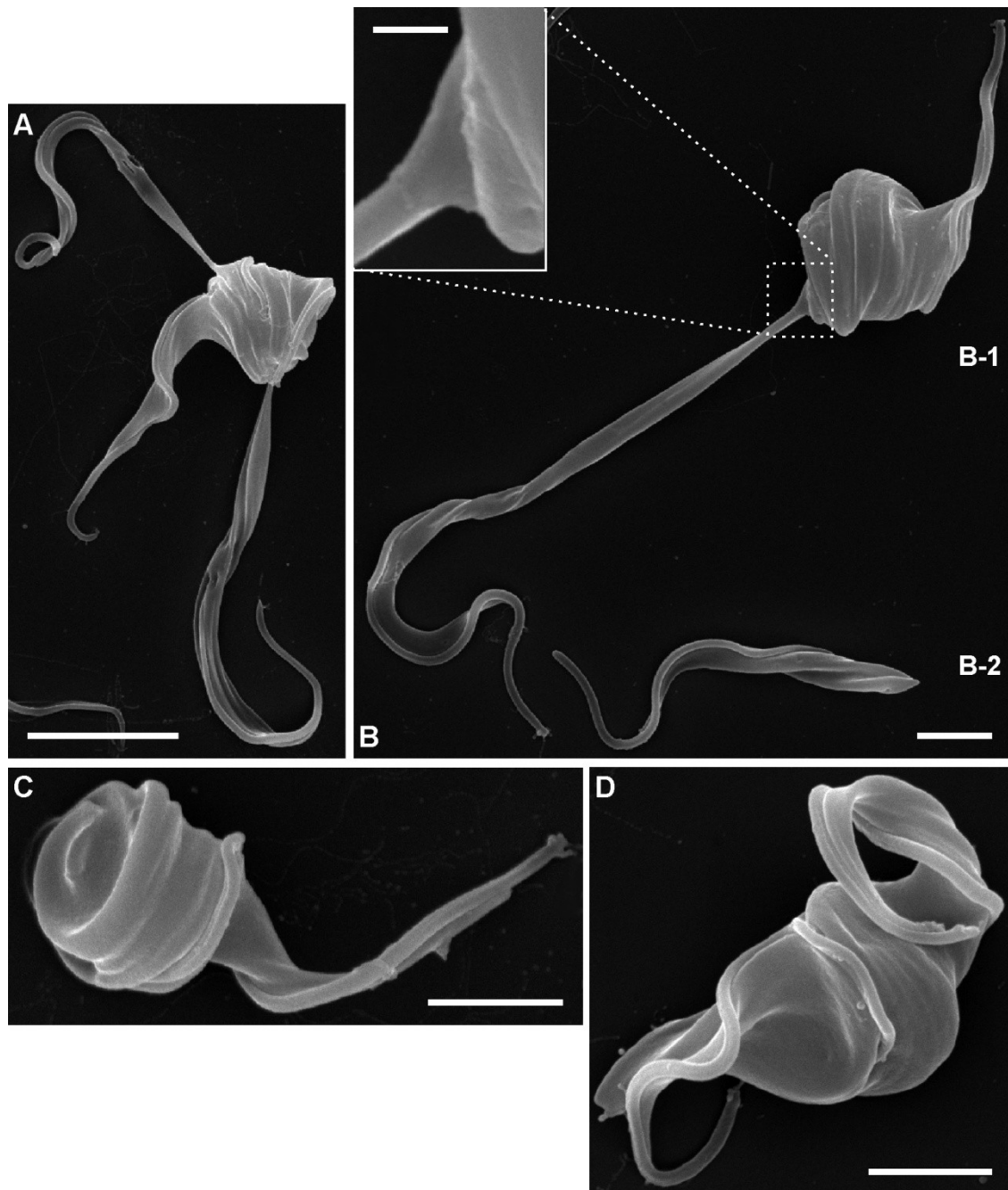


Figure 4.30. SEM of CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction (2).

A. Asymmetrically dividing cell which is likely to be giving rise to two zoids.

B. B-1. Asymmetrically dividing cell, possibly giving rise to one zoid.

B-2. Zoid-like "cell".

C. Cell with an abnormally wide posterior end.

D. Bi-flagellate cell with an abnormally wide mid-body.

Scale bars represent 2µm. Inset scale bar represents 400 nm.

Figure 4.31.

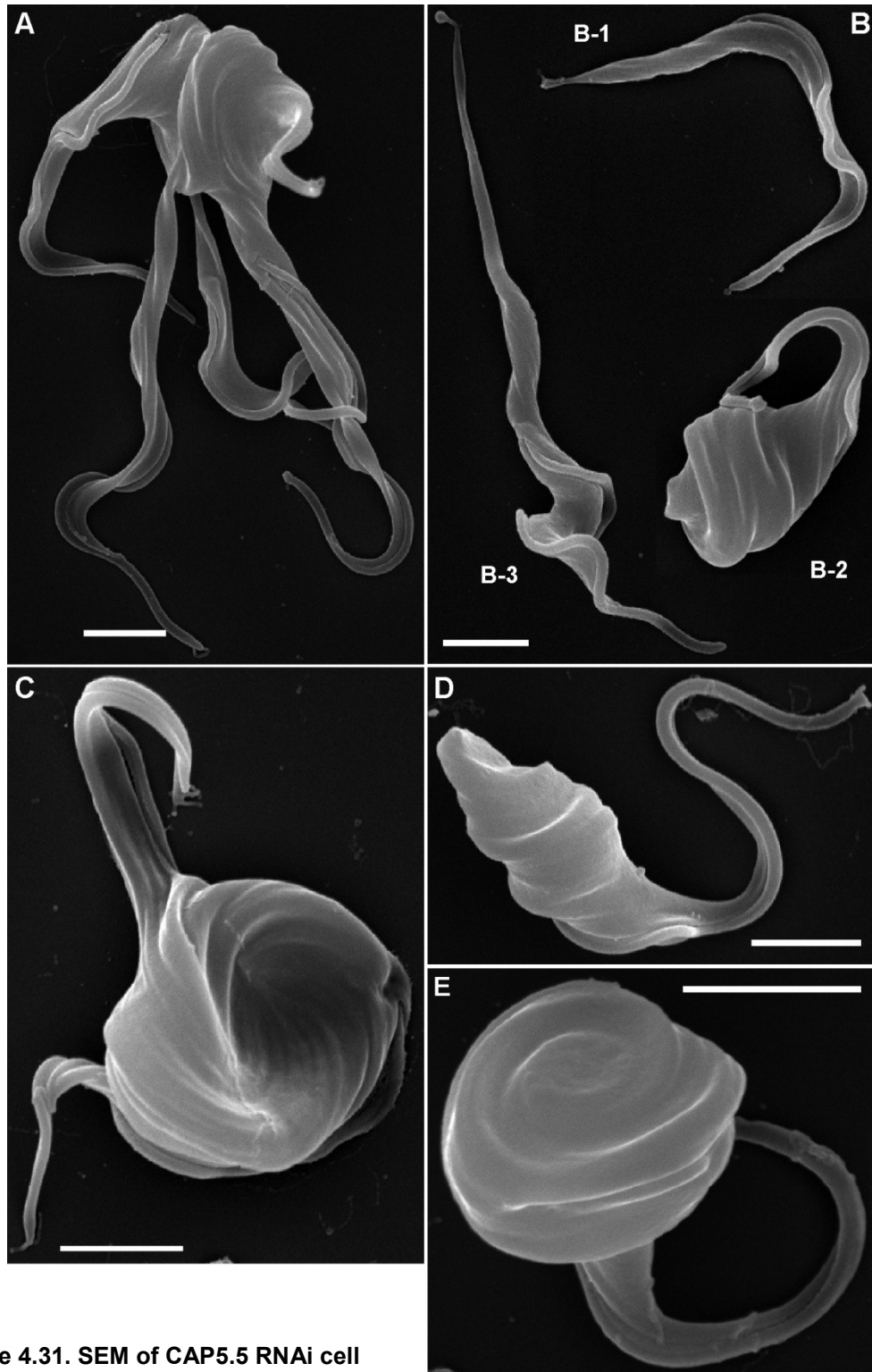


Figure 4.31. SEM of CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction (3).

A. Multi-flagellated cell that is likely to be giving rise to zoid-like "cells".

B. B-1. Zoid like "cell". **B-2 and B-3.** Uni-flagellated cells with abnormal posterior ends: B-2 has an abnormally wide posterior end and B-3 has an abnormally long one. **C.** Bi-flagellated cell with an abnormally enlarged mid-body. **D and E.** Uni-flagellated cells with abnormally wide posterior ends. Scale bars represent 2 μ m.

4.4.7 Ultrastructural analysis of CAP5.5 RNAi phenotype reveals microtubule disorganisation induced by CAP5.5 silencing.

To study the ultrastructural consequences of CAP5.5 silencing, cells were fixed for thin section electron microscopy at 0, 18 and 48h post-induction. Transverse sections of cells fixed at the 0h time-point exhibited the characteristic, evenly spaced microtubule array that lies beneath the plasma membrane of a normal trypanosome cell (Fig. 4.32). However, by 18h post-induction some defects in microtubule organisation were evident in ~8% of cell sections (N=350), most notably in sections corresponding to the posterior end or flagellar pocket areas of cells (Fig. 4.33). At 48h of induction many cell sections were characterised by multinucleate profiles and an abnormal shape. About 15% of cell sections after 48h of induction (N=300) revealed some form of microtubule abnormality (Fig. 4.34 and 4.35), most notably aberrations in inter-microtubule spacing and/or the bundling of cytoplasmic microtubules beneath the plasma membrane. Additionally, at the 48h time point, microtubules-related defects were observed in both the posterior and anterior region of cells, as judged by the width of the cell-body and the presence or absence of characteristic posterior-end features (flagellar pocket, basal bodies, kinetoplast and/or Golgi stacks profiles). Apparent irregularities in anterior-end microtubule organisation potentially represent cross-sections through zoids or cells such as that shown in figures 4.30A, 4.30B and 4.31A, in which multiple, partially ingressed cleavage furrows are evident. Defects in microtubule organisation also correlate with difficulties in flagellar pocket morphogenesis, as illustrated by figure 4.34D where four flagellar sections are evident in three flagellar pocket profiles, including two flagellar sections in the same flagellar pocket. Transverse sections where a paraflagellar rod (PFR) is abnormally present within the flagellar pocket were also observed (in normal cells the PFR is present and attached to the axoneme only from

Figure 4.32.

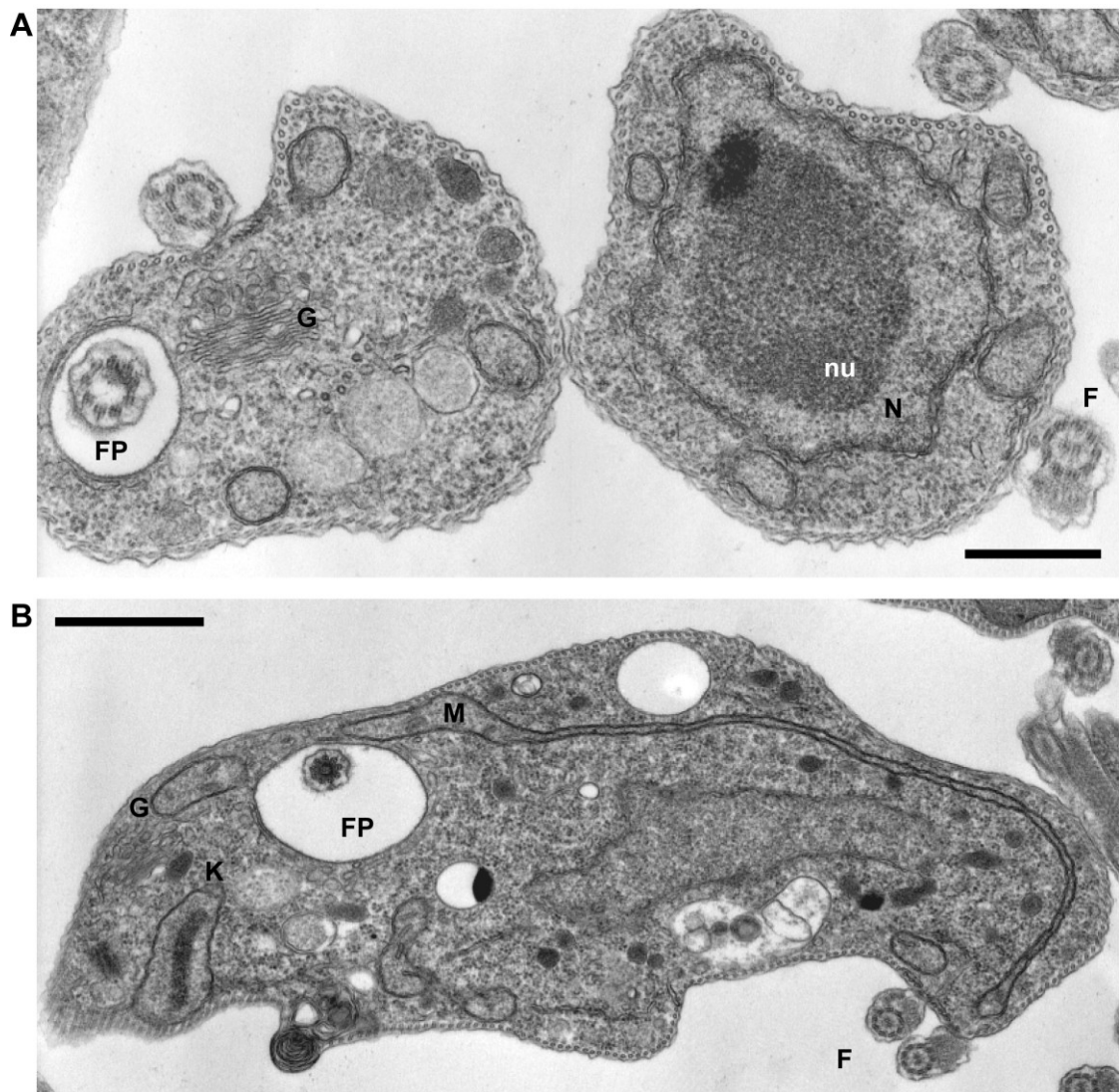


Figure 4.32. TEM of PCF CAP5.5 RNAi cell line (5.5-C-U). 0h controls.

Transmission electron micrographs of uninduced CAP5.5 RNAi cells showing the presence of the characteristic subpellicular cytoskeleton, with evenly spaced microtubules underneath the plasma membrane.

A. Transverse section through two uninduced procyclic cells. Profiles through the flagellar pocket (FP) and Golgi stacks (G) in the left-hand cell identify a transverse section through the posterior end of a cell; the nuclear (N) profile in the cell on the right identifies a likely transverse section through the mid-point of a cell-body. In both sections an evenly spaced monolayer of microtubules lies under the plasma membrane. nu, nucleolus.

B. Glancing section across the anterior-posterior axis of a not-induced, dividing cell. Profiles through the kinetoplast (K), Golgi stacks (G) and flagellar pocket (FP) on the left side of the cell identify it as the posterior cell-end. Profile through the nucleus (N) at the centre of the cell indicates the mid-point of the cell-body. Profiles through attached flagella (F) on the right side of the cell indicate the glancing angle of the section. A monolayer of microtubules is visible underneath the plasma membrane.

Scale bars represent 500 nm.

Figure 4.33.

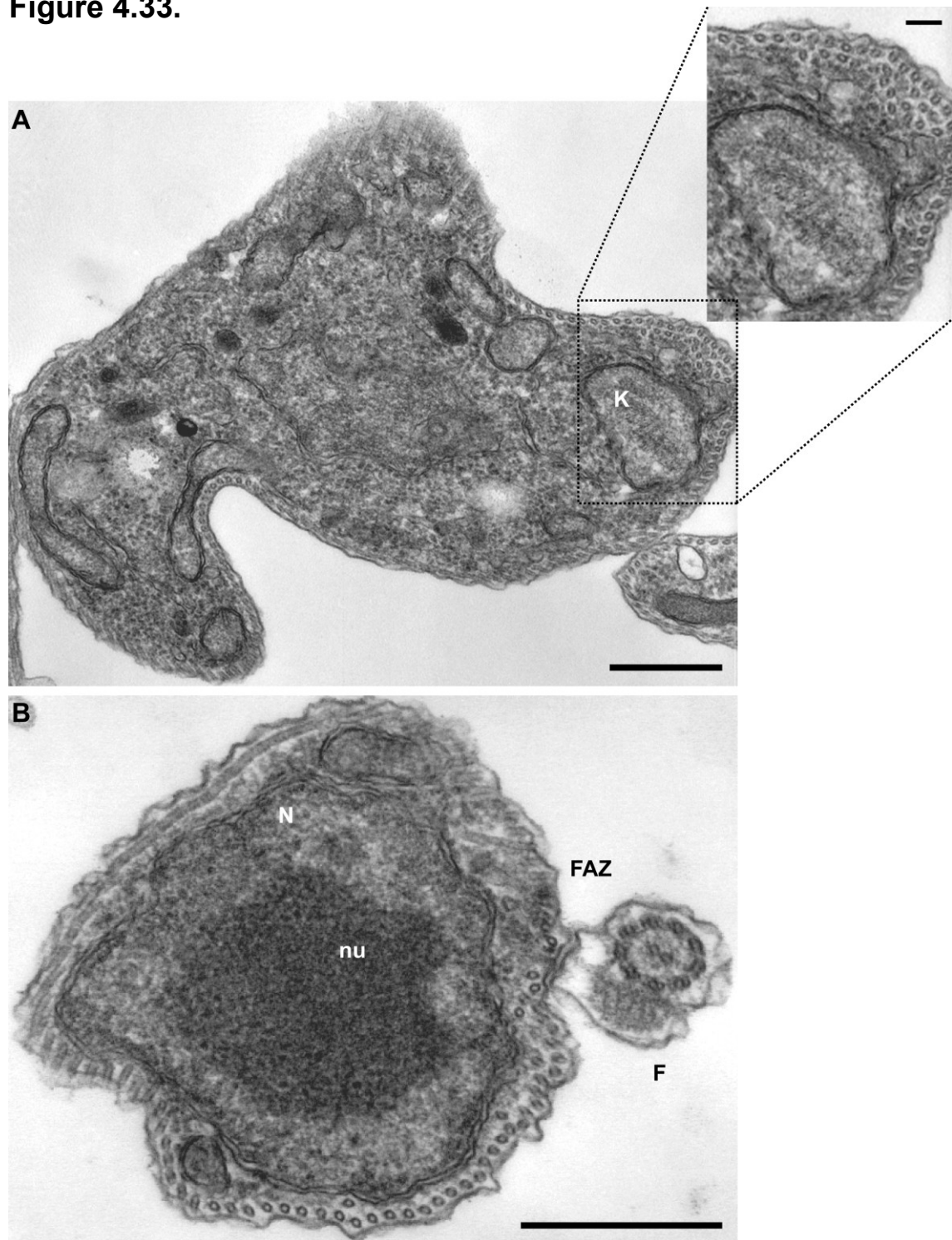


Figure 4.33. TEM of PCF CAP5.5 RNAi cell line (5.5-C-U). 18h post-induction.

A: Transverse section through a posterior cell-end, as evidenced by a kinetoplast (K) profile, reveals an example of microtubule disorganisation. The boxed region is displayed at a higher magnification in the inset and shows the abnormal location of microtubules lying between the subpellicular monolayer and the K.

B: Transverse section through the mid-cell body of a trypanosome as evidenced by the nuclear (N) and nucleolar (nu) profiles, showing another example of microtubule disorganisation after 18h of induction. Note the aberrant inter-microtubule spacing near the FAZ, as well as the presence of additional microtubules beneath the subpellicular monolayer. (F, flagellar profile). Scale bars represent 500 nm, inset scale bar represent 100 nm.

Figure 4.34.

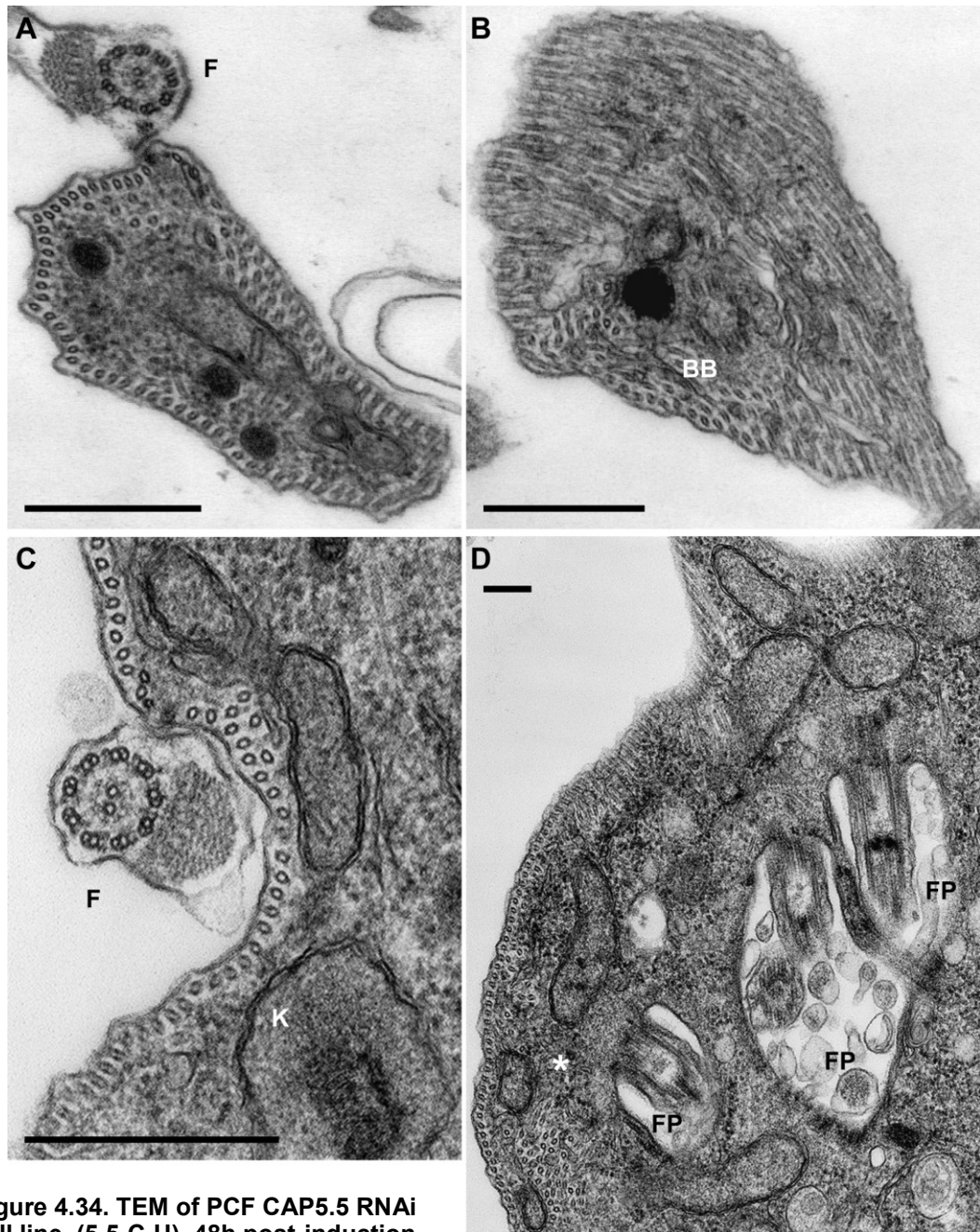


Figure 4.34. TEM of PCF CAP5.5 RNAi cell line. (5.5-C-U). 48h post-induction.

A. Transversal section through a small cell profile showing multiple additional microtubules beneath the plasma membrane. The absence of posterior end organelle profiles (flagellar pocket, Golgi stacks or kinetoplast) could identify this section as the anterior end of a cell but also as a section through a zoid, which are slender alongside their cell-body length. (F, flagellar profile).

B. Glancing section through a posterior cell end (as evidenced by a basal body (BB) profile) showing a high number of additional microtubules across the whole section.

C. Transversal section through a multi-nuclear cell (not shown) with aberrant inter-microtubule spacing and additional microtubules lying beneath the sub-pellicular monolayer and near a kinetoplast (K) profile.

D. Difficulties in flagellar pocket biogenesis as evidenced by two flagellar cross-sections in the middle flagellar pocket profile (FP). Note the high number of additional microtubules between the left-hand flagellar pocket and the plasma membrane (*).

Scale bars represent 500 nm.

Figure 4.35.

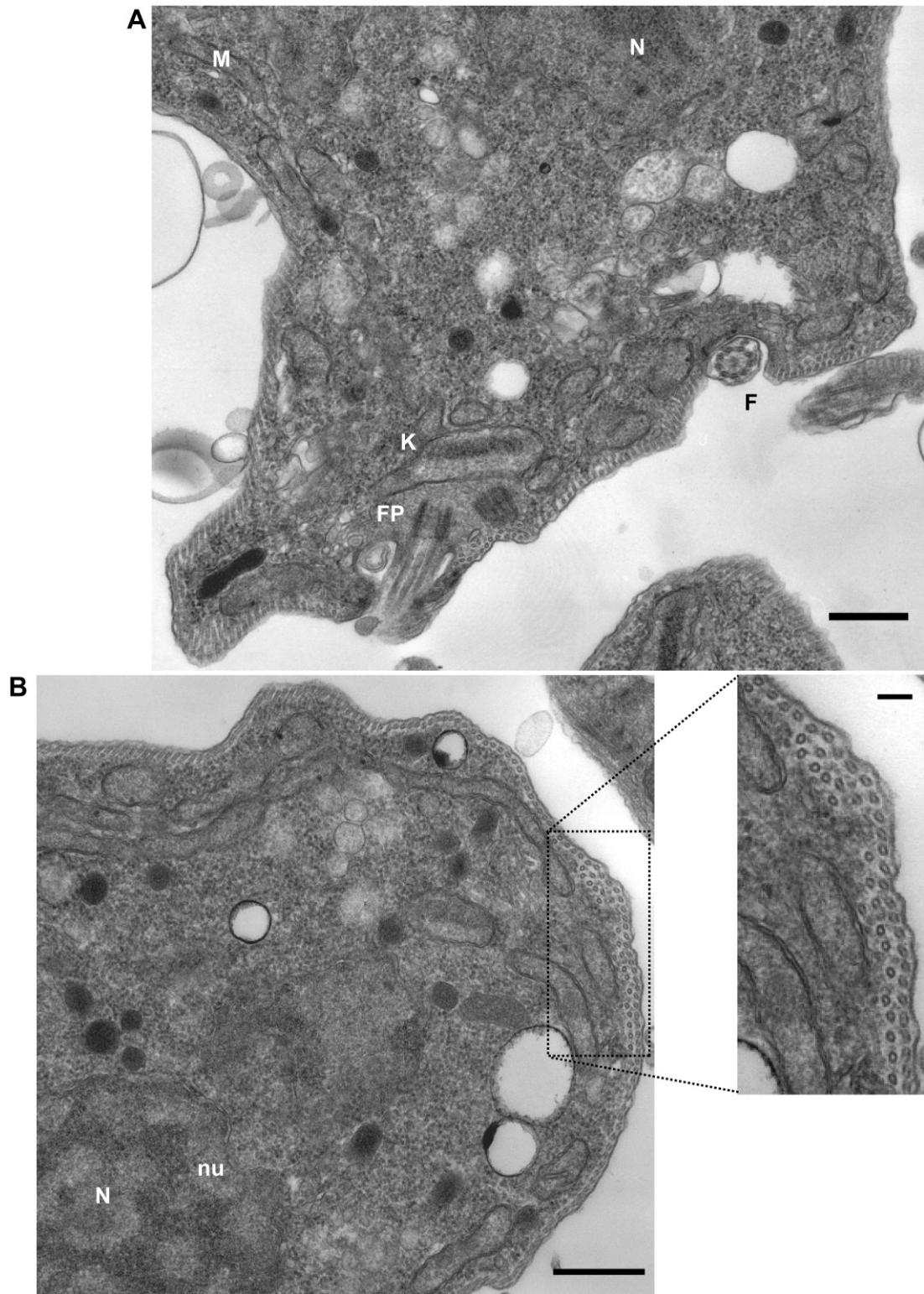


Figure 4.35. TEM of PCF CAP5.5 RNAi cell line (5.5-C-U). 48h post-induction (2).

A. Section through a multinucleate (N), multi-flagellated (F) cell profile showing the abnormal state of the cytoplasm, invaded by multiple vacuoles and ramified mitochondrion (M). (FP and K, flagellar pocket and kinetoplast profiles).

B. Transversal section of a multinucleate (N) cell profile showing additional microtubules (enlarged in the inset) and abnormal nucleolar structure (nu).

Scale bars represent 500 nm, inset scale bar represents 100 nm.

the point where the flagellum exits its flagellar pocket). Longitudinal views of microtubules in finger-like cell projections provided further evidence of microtubule disorganisation, conceivably corresponding to cell regions where the cleavage furrow ingression and subsequent abscission would give rise to a zoid (Fig. 4.35B).

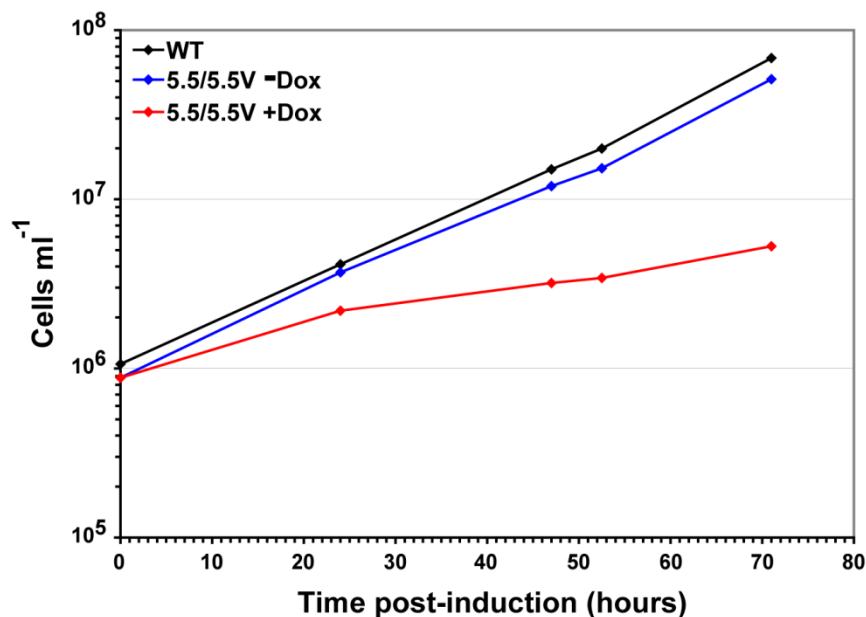
4.5. RNAi against CAP5.5 targeting a different region of its sequence produced a qualitatively similar phenotype.

To rule out the possibility of off-target effects in the induced CAP5.5 RNAi cell line, I designed a different RNAi construct, targeting a different region of the CAP5.5 gene, providing a qualitative control for CAP5.5 silencing. Thus, I used a different DNA insert, which corresponded to the N-terminus the CAP5.5 CDS (Fig. 4.3). However, due to the extensive identity between CAP5.5 and CAP5.5V CDS, this construct has the potential capability to produce RNAi cross-talk with CAP5.5V, and I termed it as 5.5/5.5V. Induction of procyclic trypanosomes transfected with the 5.5/5.5V RNAi construct resulted in a slow cell growth phenotype, highly similar to that produced by the 5.5-C-U RNAi construct, although it appeared slightly earlier (after ~20h as opposed to 24h in the CAP5.5 specific construct, Fig. 4.36A and Fig. 4.7). CAP5.5 silencing was assessed by Western blotting, which showed a great reduction of CAP5.5 expression after 48h of induction (Fig. 4.36B). However, even after 72h of induction a very faint band was still discernible, showing that CAP5.5 knockdown was not as efficient as with the 5.5-C-U construct, where no band was visible after 48h of induction (Fig. 4.8).

RNAi induction of the 5.5/5.5V RNAi cell line also produced a morphology phenotype comparable to that of the 5.5-C-U RNAi cell line, as evidenced by the counts of DAPI stained nuclei and kinetoplast (Fig. 4.37). These counts showed the appearance of abnormal anucleate zoids and their 1K2N siblings after 24h of induction, in the same

Figure 4.36.

A. Cumulative Growth Curve.



5.5/5.5V RNAi cell line

Doubling Time

Hours post-induction	WT	-Dox	+Dox
12	12.2	11.6	18.2
35,5	12.3	13,6	42.3
49,75	13.4	15.7	56.9
61,75	10.4	10.6	29.7

Average doubling time of 29.13 WT cells: 12.1h.

Average doubling time of not-induced cells: 12.9h.

Average doubling time of induced cells: 36.8h.

B. Western Blotting.

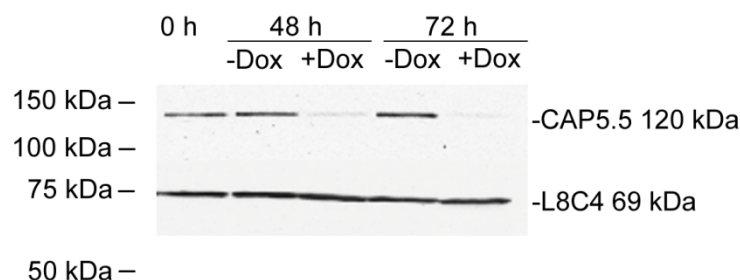


Figure 4.36. CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V).

A. Cumulative growth curve: Induction of RNAi cultures produced a phenotype of slow growth, first apparent after ~20h of induction, slightly earlier than that produced by the 5.5-C-U construct (~24h). Both cultures, induced (+Dox) and uninduced (-Dox) were re-inoculated at a cell density of 10^6 cell.ml⁻¹ after 48h of RNAi induction and the cumulative cell density was plotted.

B. Western blot: Western blotting against CAP5.5 using protein samples from the 5.5/5.5V RNAi cell line showed great reduction of CAP5.5 expression in induced cells after 48h of induction. However, a very faint band is still discernible even after 72h of induction. 2×10^6 cell equivalents were loaded in each lane. Loading control was provided by double labelling of the nitrocellulose membrane with L8C4 (Paraflagellar Rod Protein 2) and CAP5.5 mAbs.

Figure 4.37.

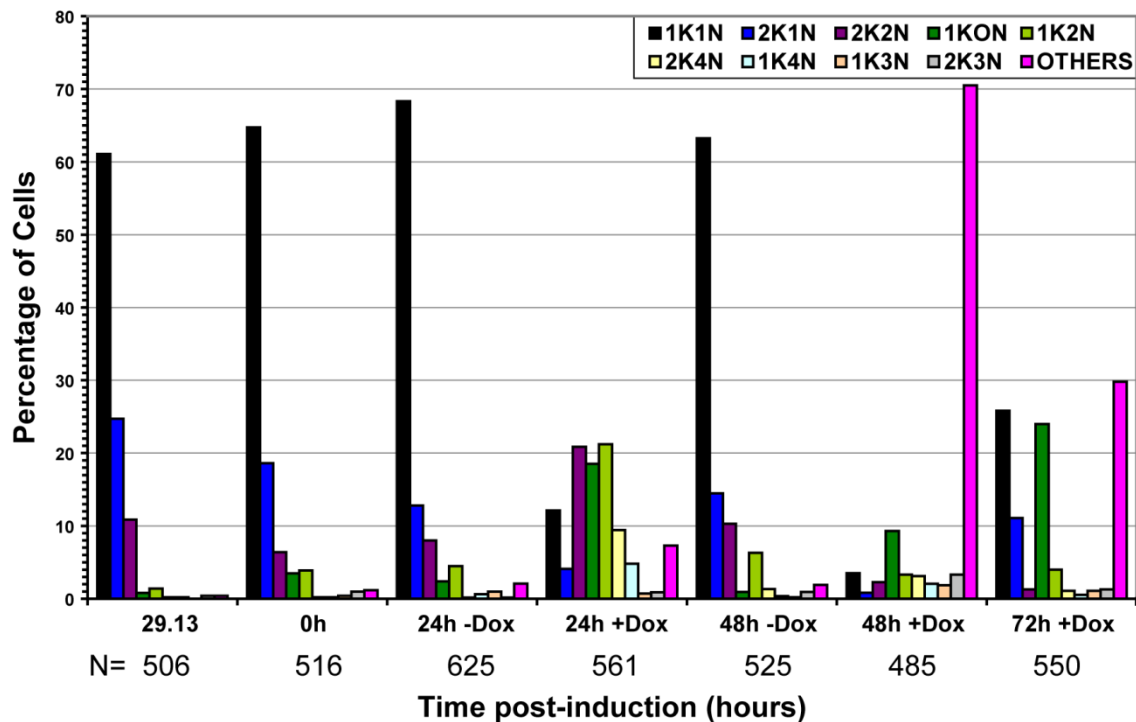


Figure 4.37. CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V).

Counts of the number of nuclei and kinetoplasts in DAPI-stained populations of the procyclic 29.13 and 5.5/5.5V RNAi cell lines.

Uninduced cultures, 0h, -Dox. Induced 5.5/5.5V cultures after 0, 24, 48 and 72h of RNAi induction, +Dox.

The number of abnormal 1K2N cells and their siblings, 1K0N "zoids", the products of aberrant cytokinesis, increased proportionally after 24h of induction. 2K2N cells also increased in absolute numbers, reflecting the slow growth rate, but also the defects in cytokinesis of induced cultures (abnormal 2K2N cells).

The number of multinucleate cells, named here as "others", which characteristically contained a higher number of nuclei than kinetoplasts, increased dramatically after 48h of induction, suggesting that the 1K2N progeny of aberrant cell division were able to enter into further cell division-cycles, but failed to complete normal cytokinesis.

Zoids continued to accumulate over time, but in smaller numbers than in the 5.5-C-U RNAi cell line.

proportion as in CAP5.5 specific silencing (~20%). Indeed, the main differences between both inductions appeared at 48h of induction, when the 5.5/5.5V RNAi population was mainly formed by multinucleate cells (71%), with only a relatively small percentage of zoids (9%). This is in contrast with 5.5-C-U RNAi population, which by 48h had almost as many zoids as multinucleate cells (28 and 29% respectively, Fig. 4.9). By 72h of induction, both induced populations look more similar, although the proportion of zoids was far higher in the 5.5-C-U RNAi population (47%) than in the 5.5/5.5V RNAi population (24%). The percentage of multinuclear cells was also dissimilar after 72h of induction, with values of 25% and 30% in the 5.5-C-U and 5.5/5.5V RNAi populations, respectively. Thus, it appears that the 5.5-C-U RNAi cell line has a higher ability to complete aberrant cytokinesis, giving rise to more zoids than the 5.5/5.5V RNAi cell line, which, upon induction, gets richer in multinucleate cells.

Immunofluorescence analysis of the 5.5/5.5V RNAi cell line, using the monoclonal CAP5.5 antibody, also evidenced the same kind of morphological phenotype as in the 5.5-C-U RNAi population. Uninduced cells displayed the typical even distribution of CAP5.5 protein throughout the cell-body (Fig. 4.38A-C). After 24h of induction, the majority of induced cells showed disappearance of CAP5.5 from the posterior end. The loss of CAP5.5 correlated with the appearance of aberrant positioning of the nuclei following mitosis, evidenced by high numbers of 2K2N cells with nuclei mis-positioning and/or undergoing asymmetric division (Fig. 4.38D-E). Moreover, 24h induced cultures showed the presence of numerous zoids, with little or no CAP5.5 protein, and their 1K2N siblings, with CAP5.5 staining only at the posterior cell-end (Fig. 4.38F-G). After 48h of induction, the induced population consisted mainly of multinucleate, multi-flagellate cells (Fig 4.38H-I), which contained little CAP5.5 protein and generally have a higher number of nuclei than kinetoplasts; and zoids, which almost completely lack CAP5.5.

Figure 4.38.

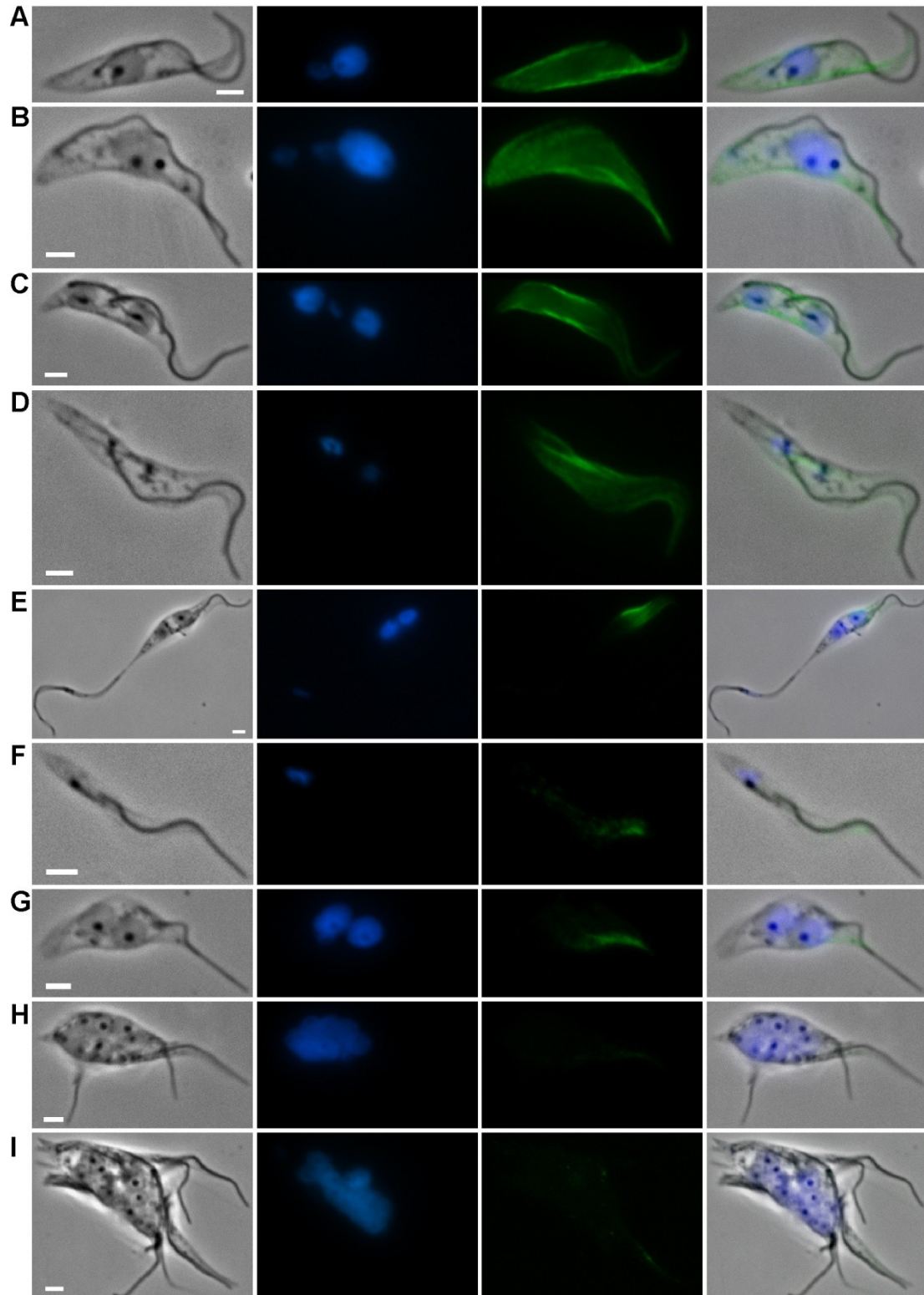


Figure 4.38. CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V). Detergent extracted cytoskeletons of uninduced (A-C) and induced cells (D-I). **A-C.** 1K1N, 2K1N and 2K2N respectively showing the typical morphology of procyclic trypanosomes. **D-G:** 24h induced cells; **D.** 1K1N with loss of CAP5.5 at the posterior end; **E.** 2K2N cell undergoing asymmetric division and giving rise to a zoid; **F.** Zoid showing very little CAP5.5 protein. **G.** 1K2N cell with CAP5.5 staining only at its anterior end. **H-I.** 48 and 72h, respectively, induced cells with multiple nuclei and flagella, which are unable to complete cytokinesis. Scale bars represent 2 μ m.

Scanning electron microscopy of induced cultures allowed the visualisation of cells undergoing asymmetric division likely giving rise to zoids, as early as 12h post-induction (Fig 4.39C). The general features of 24, 48 and 72h induced cells were also very similar to the ones showed by the 5.5-C-U RNAi cell line. However, a striking difference in the 5.5/5.5V RNAi induced cells was the appearance of cell-body twisting (Fig. 4.39D-H). Whether this difference is due to RNAi cross-talk with CAP5.5V or if it is a consequence of the two RNAi constructs different efficiency to silence CAP5.5 expression remains to be elucidated. However, Western blotting analysis and comparison of CAP5.5 silencing by the two RNAi constructs points towards a lesser capacity to efficiently silence CAP5.5 by the 5.5/5.5V RNAi construct (Fig. 4.36).

Ultrastructural analysis of 5.5/5.5V RNAi induced cells by TEM revealed the same kind of phenotype as in the induced 5.5-C-U RNAi cell line. By 24h of induction, ~6% of cell sections (N=400) revealed microtubule abnormalities, consisting on additional microtubules beside the subpellicular monolayer of the microtubular corset. These abnormalities were mainly observed in sections corresponding to posterior or flagellar pocket areas. By 48h post-induction, the majority of cell sections have multinucleate profiles, abnormal shape and size. About 11% of 48h sections (N=500) revealed microtubule disorganisation (Fig. 4.40C-D). These microtubule abnormalities were present in both the posterior and anterior regions of the cells and consisted mainly on aberrations in inter-microtubule spacing and/or the bundling of cytoplasmic microtubules beneath the plasma membrane.

Figure 4.39.

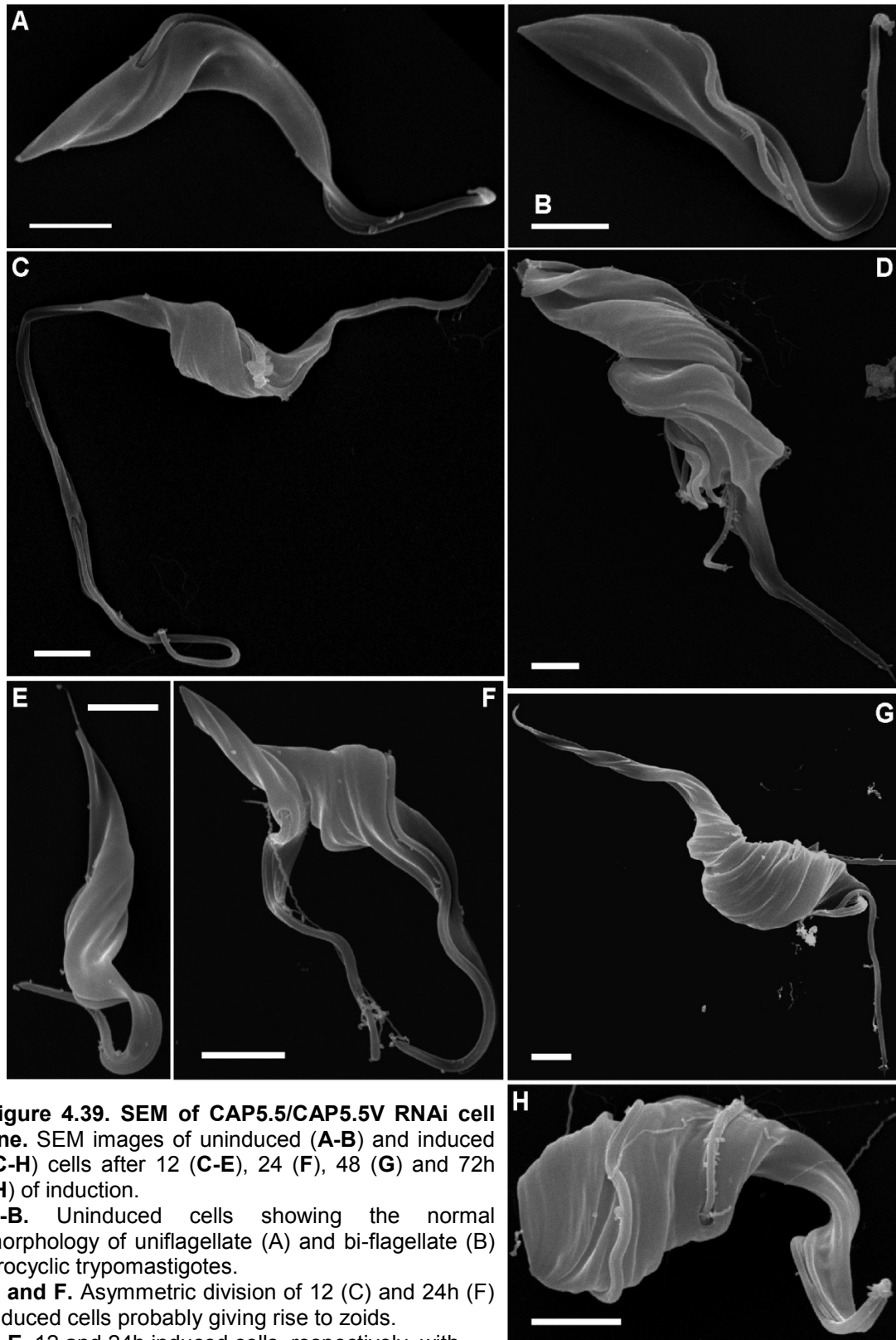


Figure 4.39. SEM of CAP5.5/CAP5.5V RNAi cell line. SEM images of uninduced (**A-B**) and induced (**C-H**) cells after 12 (**C-E**), 24 (**F**), 48 (**G**) and 72h (**H**) of induction.

A-B. Uninduced cells showing the normal morphology of uniflagellate (**A**) and bi-flagellate (**B**) procyclic trypomastigotes.

C and F. Asymmetric division of 12 (**C**) and 24h (**F**) induced cells probably giving rise to zoids.

D-E. 12 and 24h induced cells, respectively, with abnormal, twisted morphology.

G. 48h induced cell with an abnormally wide mid-body and twisted morphology.

H. 72h induced, bi-flagellate cell with an abnormally wide posterior end and twisted cell body.

Scale bars represent 2 μm.

Figure 4.40.

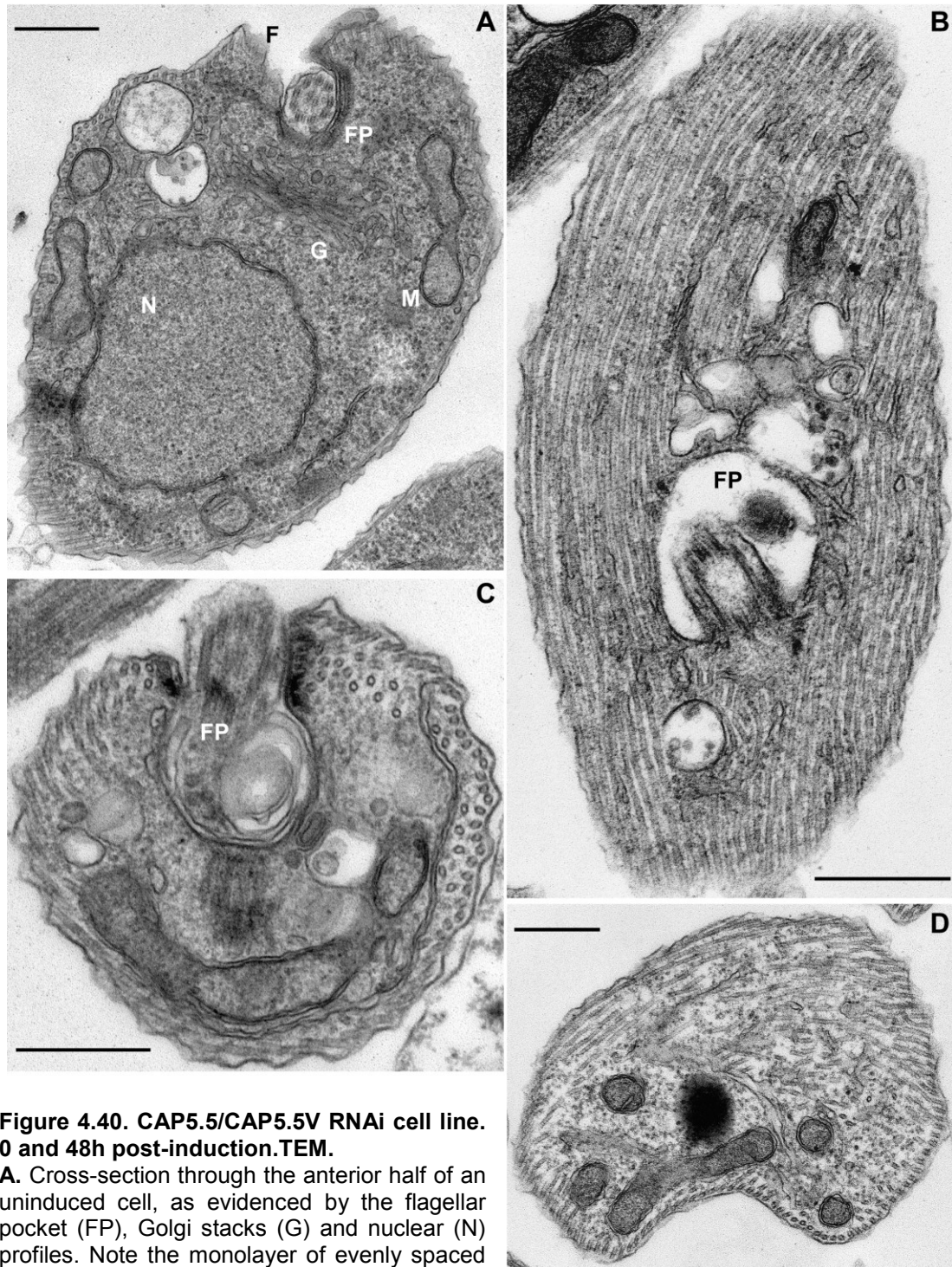


Figure 4.40. CAP5.5/CAP5.5V RNAi cell line. 0 and 48h post-induction. TEM.

A. Cross-section through the anterior half of an uninduced cell, as evidenced by the flagellar pocket (FP), Golgi stacks (G) and nuclear (N) profiles. Note the monolayer of evenly spaced microtubules underneath the plasma membrane (M and F, mitochondrial and flagellar profiles).

B-D. Sections of 48h induced cells. **B.** Glancing section through a posterior cell-end, as evidenced by FP profile, showing. **C.** Cross-section through a posterior cell-end, also evidenced by a FP profile, **D.** Glancing section, the lack of organelle profiles, especially the F, could identify it as a posterior cell-end or an aberrant cell with a detached F. B-D all show multiple microtubules across the cytoplasm and underneath the plasma membrane. Scale bars represent 500 nm.

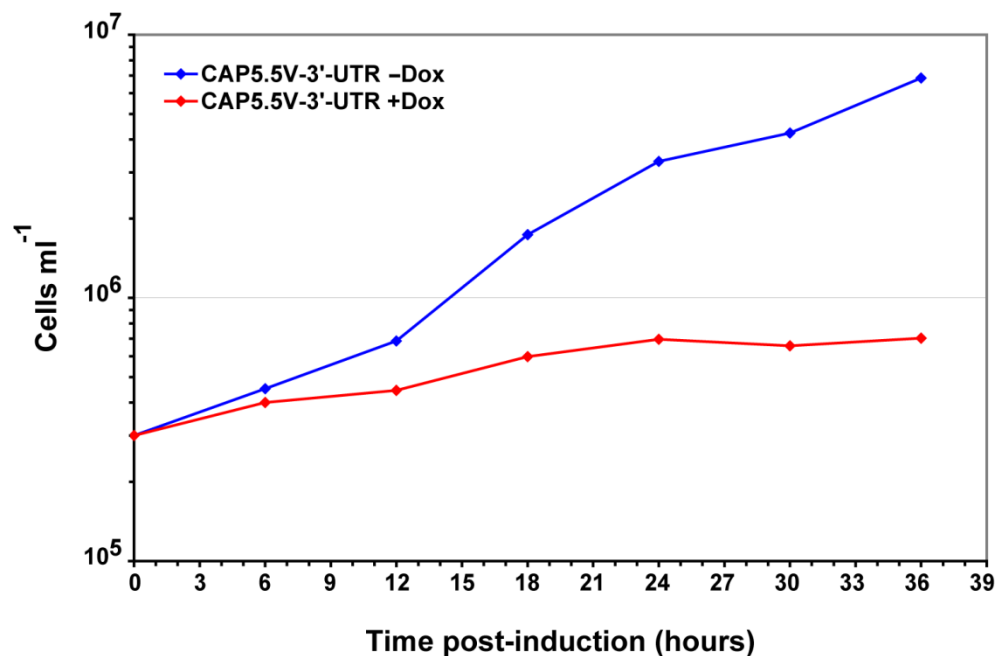
4.6. CAP5.5V is essential in bloodstream *T. brucei*.

After discovering the essential role of CAP5.5 for cell morphogenesis in procyclic *T. brucei*, an obvious next step was to analyse the function of CAP5.5 paralog, CAP5.5V, in the bloodstream stage, where its mRNA is present at a higher level than in procyclic forms (Fig. 4.2). With this aim, I used the previously described RNAi constructs, 5.5-C-U, 5.5V-C-U and 5.5/5.5V (Fig. 4.3), for experiments with bloodstream trypomastigotes. As expected, RNAi against CAP5.5V but not CAP5.5, using the 5.5V-C-U construct, resulted in strong growth and morphology phenotypes; CAP5.5 specific silencing, using the 5.5-C-U construct, did not produce any noticeable effect in bloodstream trypanosomes and RNAi against both CAP5.5 and CAP5.5V, using the 5.5/5.5V construct, phenocopied the CAP5.5V specific RNAi mutant. The constructs that contain CAP5.5 and CAP5.5V C-terminal specific coding sequences (5.5-C and 5.5V-C) were not used here, due to the inability shown by the 5.5-C construct to knockdown CAP5.5 expression in procyclic form trypanosomes (Fig. 4.6). Thus, this analysis and that of the 5.5-C-U RNAi construct, which did not produce any discernible phenotype in bloodstream forms, will not be described here, (relevant data can be found in the Appendix section: Appendix 4, Fig. A.4.8). Therefore, in this section I described the RNAi analysis of CAP5.5V in BSF trypanosomes using the 5.5V-C-U 5.5/5.5V RNAi constructs (Fig. 4.3).

4.6.1. CAP5.5V specific RNAi in bloodstream *T. brucei*.

RNAi specific silencing of CAP5.5V in bloodstream trypomastigotes produced a slow growth phenotype, with a marked decline of the growth rate by ~6-8h of induction (Fig. 4.41 and Appendix 4, Fig. A.4.9). The cell-cycle time for the not-induced RNAi cultures was ~8.5h (~6h for the parental 90.13 bloodstream cell line, (Wirtz *et al.*, 1999)). Counts of the number of nuclei and kinetoplasts stained with DAPI revealed the appearance of morphological abnormalities as early as 6h post-induction, although

Figure 4.41.



CAP5.5V-3-UTR		
RNAi cell line Doubling Time		
Hours post-induction	-Dox	+Dox
12	10.1	21.1
24	5.3	18.6
36	11.4	882.8

Average doubling time of uninduced cells: 8.9h.
Average doubling time of induced cells: 307.2h

Figure 4.41. BSF CAP5.5V RNAi cell line (5.5V-C-U). Cumulative growth curve. Silencing of CAP5.5V expression produced a phenotype of slow growth, first apparent after 6 to 8h of induction.

some abnormal cells were already detected by 3h post-induction (Fig. 4.42 and Appendix 4, Fig. A.4.10 and A.4.11). Taking into account the differences between the cell-cycle times of procyclic (~11h) and bloodstream (~8.5h) RNAi cell lines, the kinetics of phenotype onset, relative to the cell-cycle time, were similarly fast for CAP5.5 RNAi in procyclic cells and CAP5.5V RNAi in bloodstream trypanosomes. After 6h of induction, cultures showed an increase in the number of 2K2N cells, reflecting the slower growth rate by this time-point. Many of these 2K2N cells have abnormal morphology and mispositioning of the nuclei and/or kinetoplasts. Similarly to the procyclic CAP5.5 RNAi mutant, zoids and multinucleate cells accumulated in 6h induced RNAi cultures (Fig. 4.42). These results, resembling the ones obtained with CAP5.5 RNAi induction in procyclic form trypanosomes, are very significant, considering that the patterns of kinetoplasts and nuclei positioning, prior to cell division, are distinct in procyclic and bloodstream form trypanosomes. In bloodstream 2K2N trypomastigotes the replicated and segregated kinetoplasts lie posterior to the divided nuclei (Fig. 4.43), being subject to a more limited separation during segregation than in procyclic trypomastigotes, where one daughter kinetoplast sits between the divided nuclei. However, CAP5.5 and CAP5.5V RNAi in procyclic and bloodstream forms, respectively, give rise to the same products of asymmetric division —zoids and 1K2N. In 6h induced bloodstream CAP5.5V RNAi cell line, zoids and 1K2n cells increased proportionally to each other (Fig. 4.42, 4.44 and 4.45).

Zoids continued to accumulate as a function of time and by 9h post-induction they doubled the amount of 1K2N cells (Fig. 4.42). Multinucleate cells also increased in induced cultures, becoming the most prevalent cell type by 16-24h of induction (Appendix 4, Fig. A.4.10 and A.4.11). Accumulation of multinucleate cells evidenced that the 1K2N progeny of aberrant cytokinesis were able to enter into further cell division-cycles. Additionally, by 9h of induction, about 50% of these multinucleate cells

Figure 4.42.

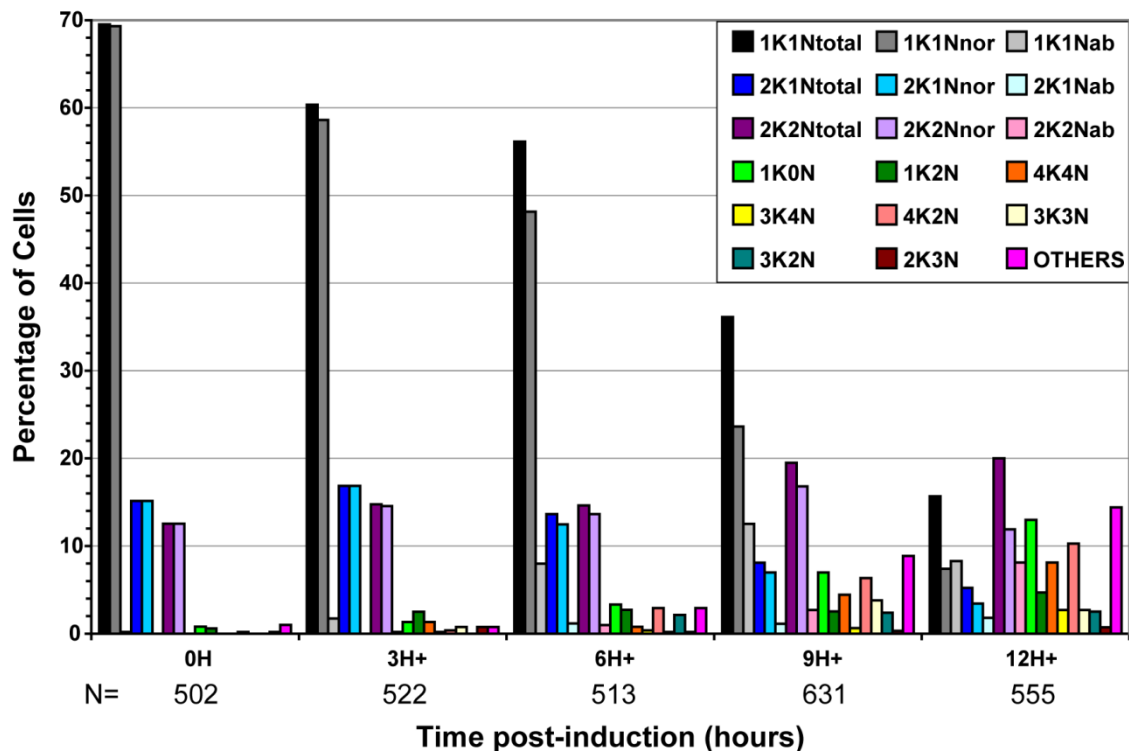


Figure 4.42. BSF CAP5.5V RNAi cell line (5.5V-C-U). Number of nuclei and kinetoplasts.

Counts of number of nuclei and kinetoplast in DAPI-stained populations of uninduced (0h, -Dox) and induced (+Dox) cultures at 0, 3, 6, 9 and 12h post-induction of CAP5.5V BSF RNAi cell line. The number of abnormal 1K2N cells and 1K0N "zoids", products of aberrant cytokinesis, increased noticeably and proportionally after 6h of induction.

Numbers of multinuclear cells, named here as 4K4N, 3K4N, 3K3N, 2K3N and "others", many containing a higher number of nuclei than kinetoplasts, increased remarkably after 9h of induction, suggesting that the 1K2N progeny of aberrant cell division were able to enter into further cell division-cycles, but failed to complete normal cytokinesis.

Zoids continue to accumulate over time, suggesting zoid longevity and/or the ability of multinuclear cells to undergo cytokinesis and abscission, giving rise to multiple zoids. Note also the increase in 2K2N cells after 9h of induction, reflecting the reduction in growth rate.

Figure 4.43.

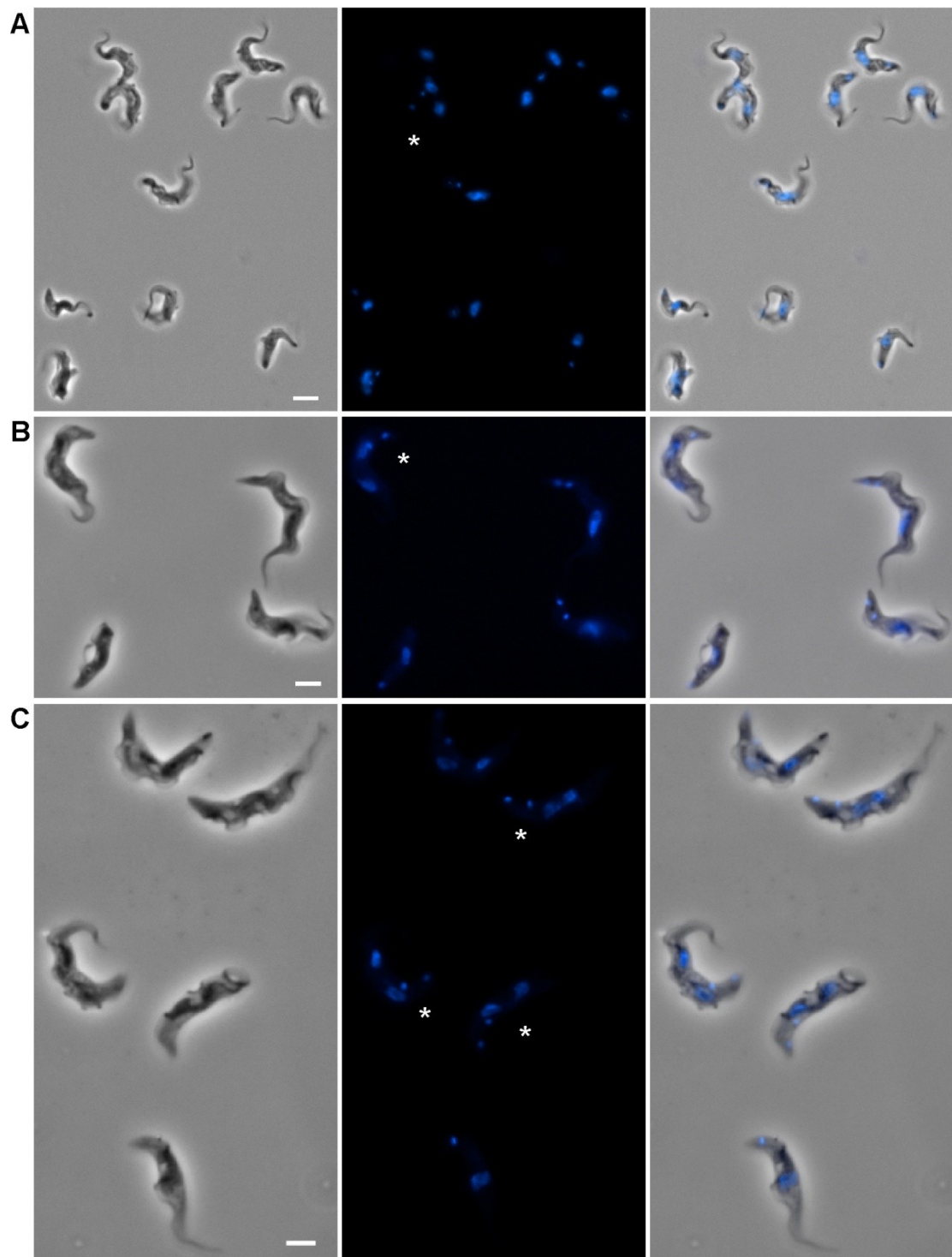


Figure 4.43. BSF CAP5.5V RNAi cell line. Not-induced.

Phase contrast, DAPI and merged images of BSF CAP5.5V RNAi populations before induction. Note the position of segregated kinetoplast at the posterior cell end (*).

A scale bar represent 10 μm (40x images).

B and C scale bars represent 5 μm (100x images).

Figure 4.44.

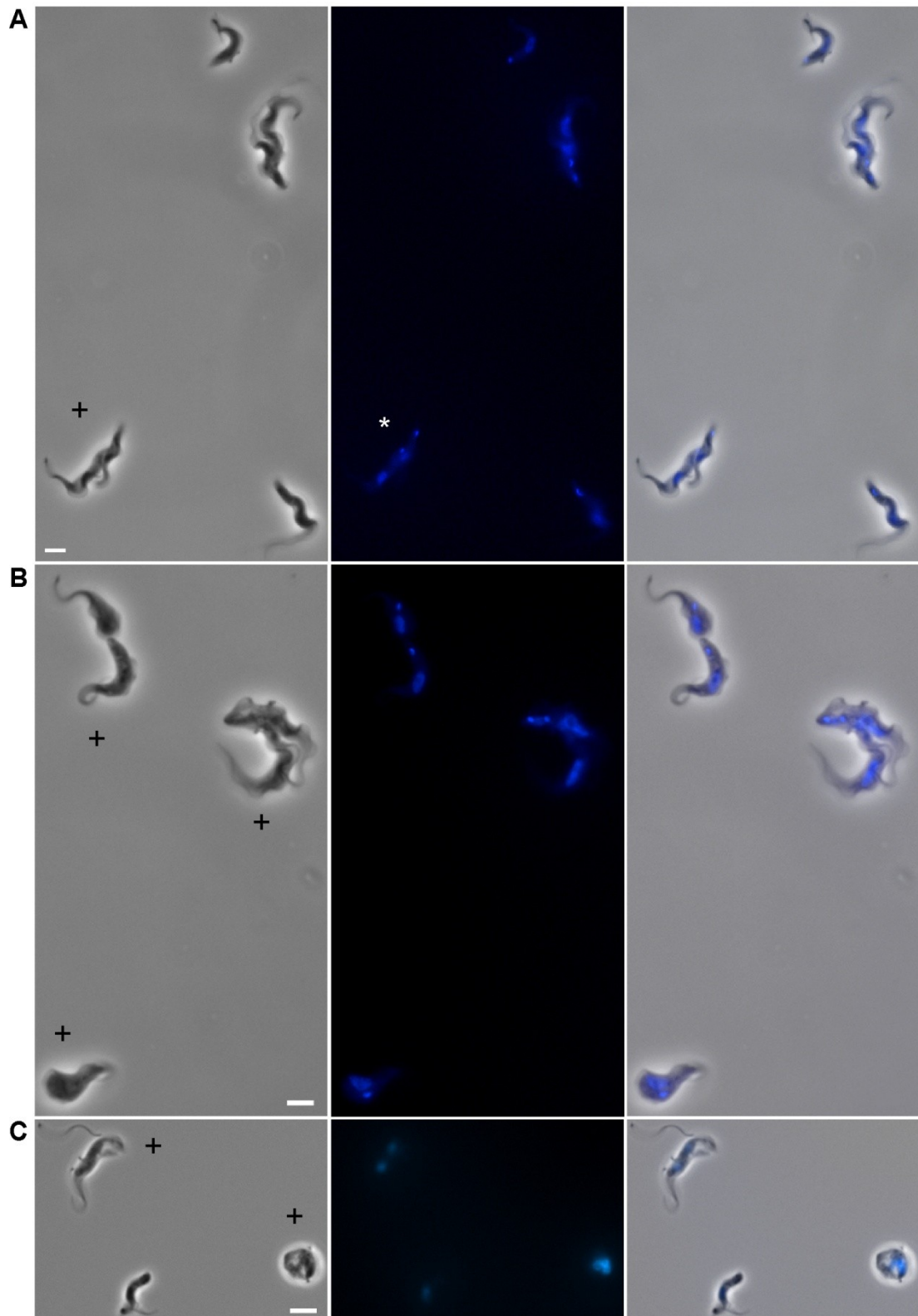


Figure 4.44. BSF CAP5.5V RNAi cell line (5.5V-C-U). 6h post-induction. Phase contrast, DAPI and merged images of CAP5.5V RNAi 6h induced cells showing abnormal cell shape of dividing cells (+) and kinetoplast/nuclei mis-positioning (*). A and B scale bars represent 2 μ m (100X images). C scale bar represents 5 μ m (40X images).

Figure 4.45.

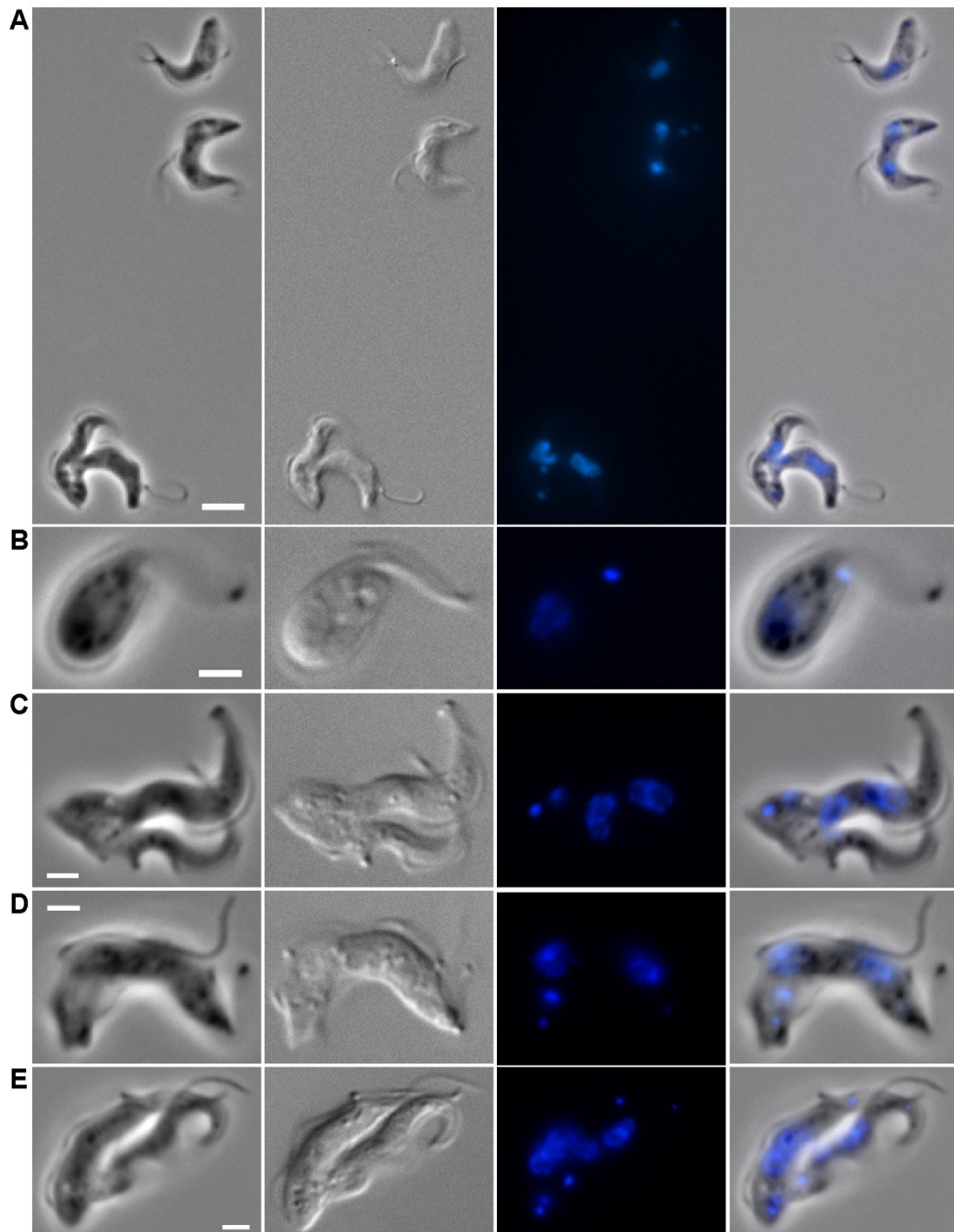


Figure 4.45. BSF CAP5.5V RNAi cell line (5.5V-C-U). 6h post-induction.

After 6h of induction, CAP5.5V RNAi BSF cultures showed an increasing proportion of cells with abnormal shape and/or organelle (nuclei and/or kinetoplasts) mis-positioning.

A. 6h induced cells with slightly abnormal morphology. **B.** Abnormal 1K1N cell with an anterior kinetoplast and a mitotic nucleus. **C.** Abnormal 2K2N cell with mitotic nuclei. The position of the cleavage furrow indicates asymmetric division, probably giving rise to a zoid. **D.** Abnormal 4K2N cell. Note the position of the two extra kinetoplasts above each nucleus. **E.** Multinucleate cell with a higher number of nuclei than kinetoplasts. A scale bar represents 5 μm (40X images).

B-D scale bars represent 2 μm (100X images).

contained a higher number of nuclei than kinetoplasts. At later time points after induction, the increasing contorted appearance of multinucleate, multiflagellate cells, containing numerous nuclei and kinetoplasts, made difficult to ascertain numbers of these organelles with confidence. Multinucleate cells often presented several cleavage furrows, which ingression could give rise to multiple zoids, thus explaining the increasing number of zoids in the absence of a parallel increase of their 1K2N siblings after 9h of induction. By 12h of induction, multinucleate cells were the most prevalent cell type, constituting ~40-50% of the induced population, followed by zoids, which constituted ~10-15% (Fig. 4.42). The slight increase in “cell” density of RNAi induced cultures beyond 12-15h post-induction correlates with a steady increase in the number of zoids, which rose to ~25-30% of the population by 24h of induction, probably as a result of abscission from multinucleate cells (Appendix 4, Fig. A.4.10-11).

From 12h post-induction onwards, multiple types of abnormal cells were found in RNAi induced cultures (Fig. 4.46-4.48). The most relevant of them were:

- Multinucleate cells with all kinetoplasts grouped together at a wide posterior end.
- Abnormal 1K1N cells with enlarged posterior ends and often with the nucleus posterior to the kinetoplast.
- Abnormal 2K2N cells, with misplaced nuclei or kinetoplast: both kinetoplasts lying together at the posterior end, both above the nuclei; or one kinetoplast at the posterior end and the other between the segregated nuclei, as in procyclic trypanosomes.
- Slender zoids.
- 1K2N cells.
- Round multinucleate cells, sometimes containing an enlarged flagellar pocket.

After 18 to 24h post-induction the proportion of abnormal cells continued to increase in induced cultures, specially the number of multinucleate cells and zoids (Fig. 4.42 and Appendix 4, Fig. A.4.10 and A.4.11).

Figure 4.46.

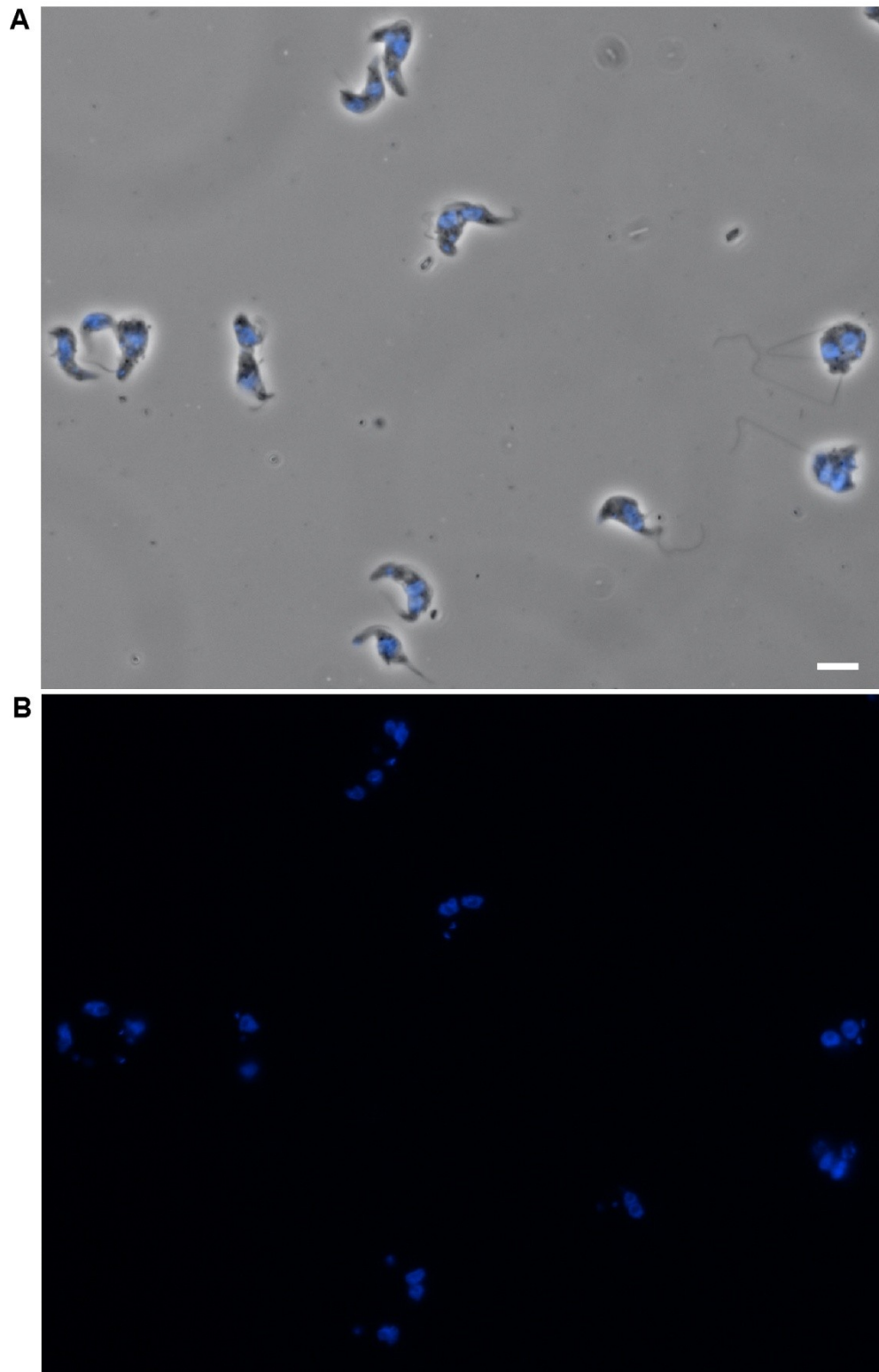


Figure 4.46. BSF CAP5.5V RNAi cell line (5.5V-C-U). 12h post-induction (1). Phase contrast/DAPI merged (A) and DAPI (B) 40X images of 12h induced CAP5.5V RNAi cultures showing a high proportion of abnormal cells, with alterations in shape, organelle positioning and/or number of nuclei and kinetoplasts. Scale bar represents 10 μm .

Figure 4.47.

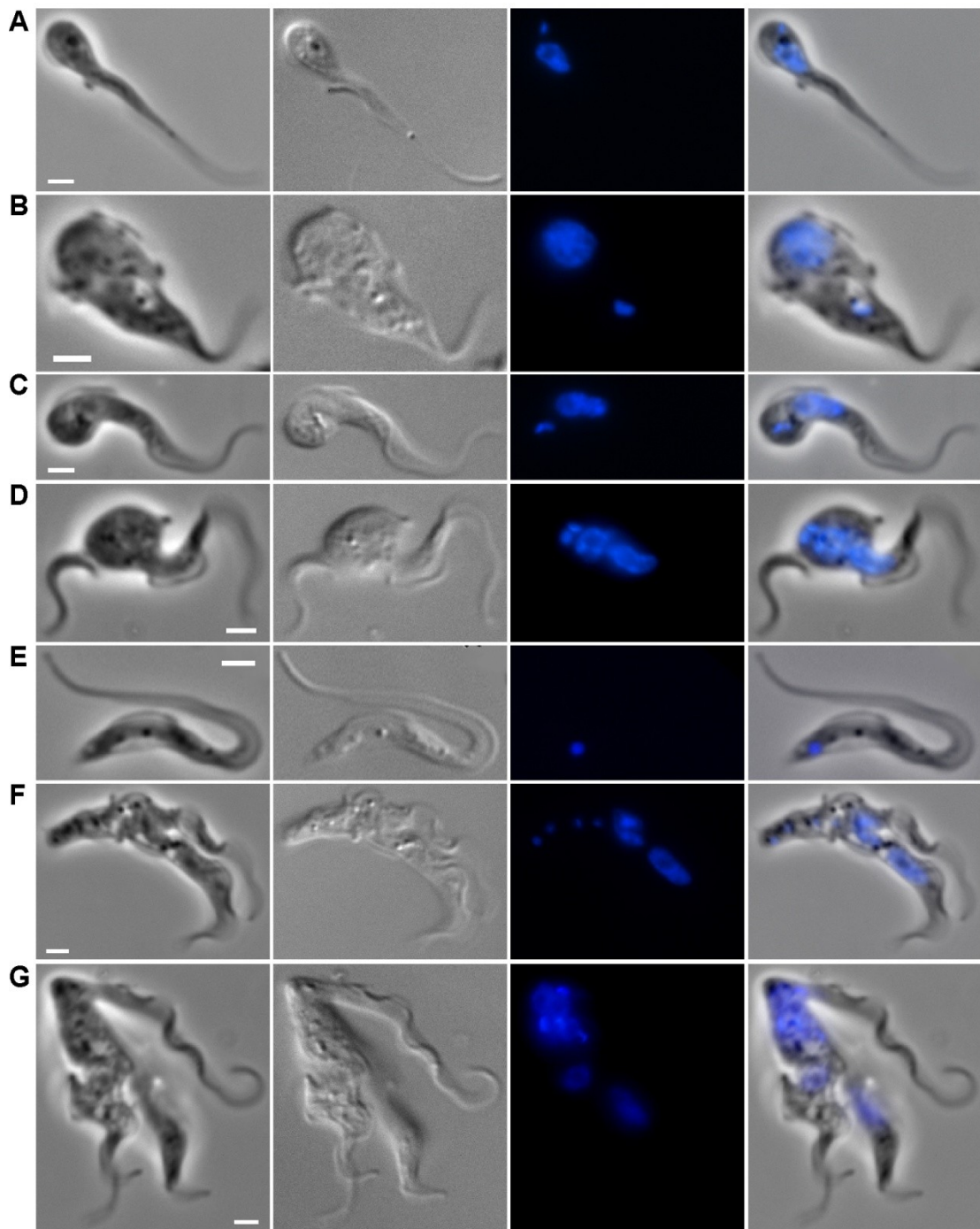


Figure 4.47. BSF CAP5.5V RNAi cell line (5.5V-C-U). 12h post-induction (2).

CAP5.5V RNAi 12h induced cells showed abnormal morphology and/or abnormal nuclei and kinetoplasts, which appeared either in misplaced locations, abnormal numbers or both.

A. Abnormal 1K1N cell with an aberrantly wide posterior end. **B.** Abnormal 1K1N cell with the kinetoplast in a position anterior to the nucleus. The nucleus large size could indicate an un-segregated but replicated nucleus. **C.** 1K2N cell, probably the result of aberrant cytokinesis of a 2K2N cell. **D.** Abnormal 2K2N cell with misplaced kinetoplasts and two mitotic nuclei. **E.** Anucleated zoid, progeny of aberrant cytokinesis. **F.** Abnormal 4K2N or 4K3N cell, with the four kinetoplasts located at the posterior end and with the anterior nucleus undergoing mitosis. **G.** 4K6N cell, with four posterior nuclei and at least two ingressing cleavage furrows, the right side one probably giving rise to a zoid. Scale bars represent 2 μ m.

Figure 4.48.

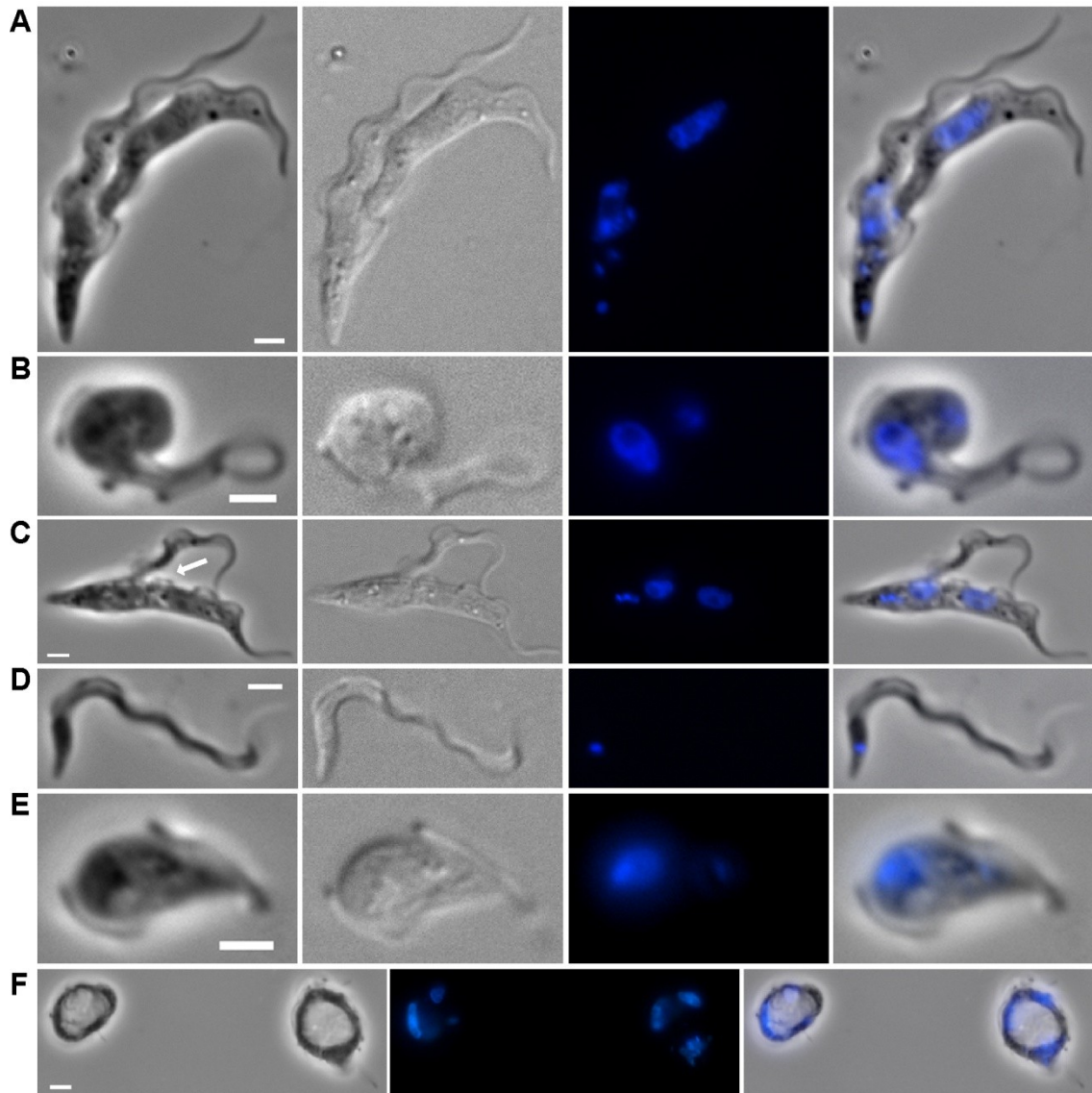


Figure 4.48. BSF CAP5.5V RNAi cell line (5.5-C-U). 12h post induction (3).

Phase contrast, DIC, DAPI and merged 100X images (A-E) and Phase contrast, DAPI and merged 40X images (F) of 12h induced CAP5.5V RNAi cells showing abnormal shape, position and/or number of organelles (nuclei and kinetoplasts).

A. Multinucleate cell (4K3N) with all kinetoplasts located at the posterior end.

B. Abnormal 1K1N cell with a very wide posterior end.

C. Abnormal 2K2N cell with the kinetoplasts close together at the posterior end. The cell is undergoing asymmetric division, probably giving rise a zoid, as indicated by the position of the cleavage furrow (arrow).

D. Anucleate zoid, with a slender "cell" body.

E. Abnormal 2K1N cell with mis-positioning of the nuclei and/or the kinetoplasts.

F. Multinucleate cells with a large flagellar pocket.

A-E scale bars represent 2 μ m. F scale bar represents 5 μ m.

4.6.2. CAP5.5V silencing does not affect the basic components of microtubules: anti α -tubulin antibody TAT-1.

To analyse the potential effect of CAP5.5V ablation on the main components of the subpellicular cytoskeleton of bloodstream trypanosomes, I performed the immunofluorescence analysis of CAP5.5V RNAi mutants with the anti α -tubulin mAb TAT-1 (Woods *et al.*, 1989), which showed no alteration in the localisation of the basic component of microtubules (Fig. 4.49).

4.6.3. CAP5.5V is essential for the correct morphogenic patterning during the cell division-cycle of bloodstream *T. brucei*.

CAP5.5V specific RNAi silencing in bloodstream trypomastigotes resulted in nuclei and/or kinetoplasts mispositioning and segregation difficulties, abnormal cell morphology and aberrant cytokinesis. The similarity between CAP5.5 and CAP5.5V RNAi phenotypes, high percentage of sequence identity, similar domain architecture and cytoskeletal association, and their different stage-specific expression strongly suggest analogous roles for these paralogous genes in the two lifecycle stages in which they are expressed. The recent confirmation of CAP5.5 cytoskeletal association is very significant (Liu *et al.*, 2009), as the inheritance of subpellicular microtubules at cytokinesis is semi-conservative (Sherwin and Gull, 1989b) and trypanosome microtubule dynamics occurs mainly at the posterior end of the cells (Robinson *et al.*, 1995). As explained above, the earliest manifestations of both CAP5.5 and CAP5.5V silencing, in procyclic and bloodstream forms respectively, take place at the posterior end of the cells shortly after induction. As in CAP5.5 silencing, this could indicate an almost immediate loss of CAP5.5V from the posterior cytoskeleton of bloodstream trypanosomes. This would imply that new microtubules growth and intercalation during cell elongation is accompanied by extensive remodelling that sees CAP5.5V stripped

Figure 4.49.

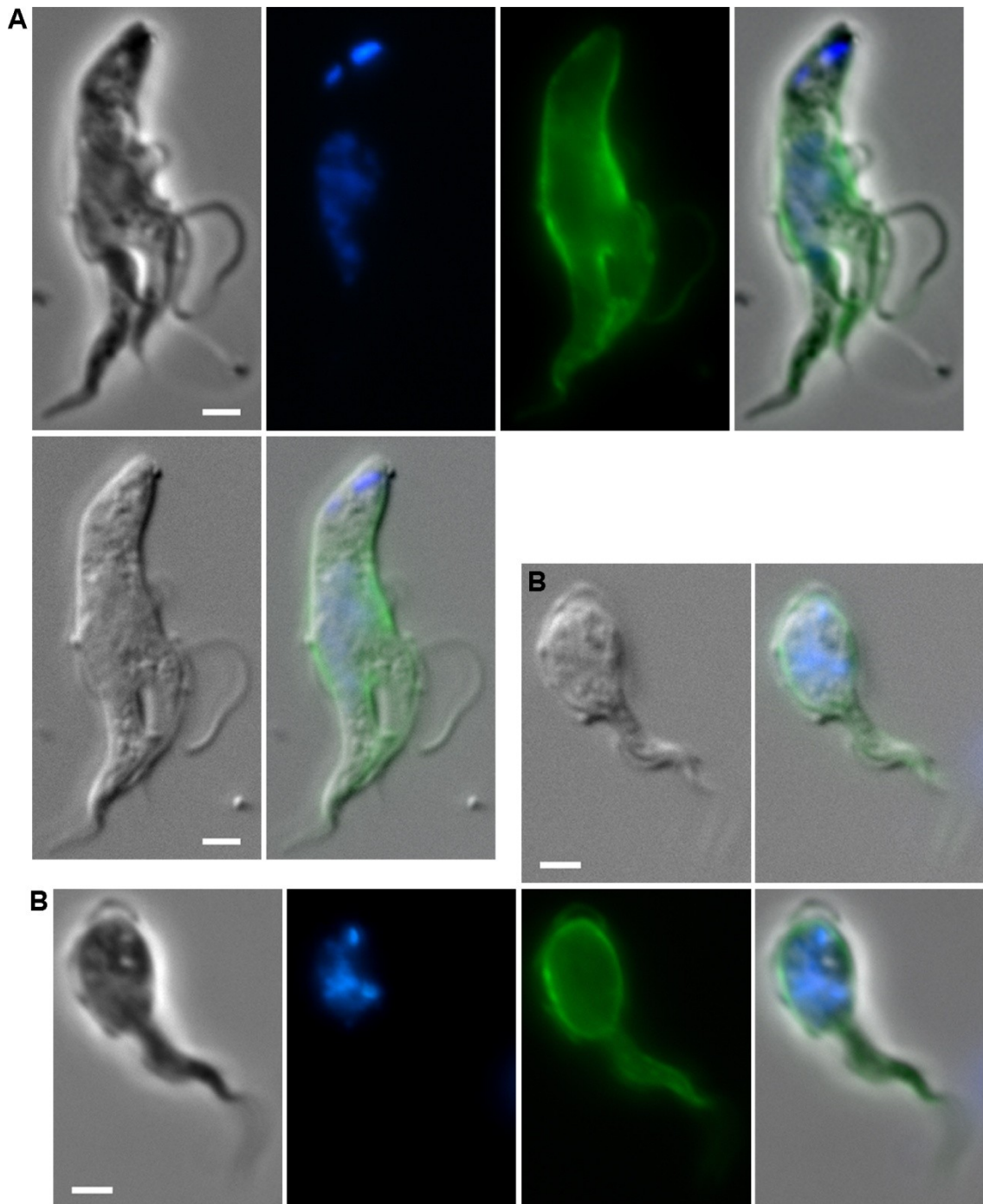


Figure 4.49. BSF CAP5.5V RNAi cell line (5.5V-C-U). 18h post-induction. TAT-1 mAb.

Phase contrast, DIC, DAPI, TAT-1 and merged 100x images of 18h induced CAP5.5V RNAi bloodstream cell line. Induced cells stained with the anti α -tubulin mAb TAT-1 showed no alterations of the basic structure of the microtubule network, formed by α/β -tubulin dimmers.

A. 2K4N cell, with an abnormally large kinetoplast, which has probably replicate but not segregate. **B.** Abnormal 2K2N cell with the kinetoplast above the nuclei and a wide posterior end. Both A and B images show an even distribution of α -tubulin across the whole cell-body and the flagellum. Scale bars represent 2 μ m.

from the elongating subpellicular microtubule corset. To investigate the effects of CAP5.5V silencing on the subpellicular cytoskeleton at an ultrastructural level, I fixed bloodstream CAP5.5V RNAi cultures (5.5V-C-U RNAi cell line) for transmission electron microscopy. Additionally, I performed scanning electron microscopy studies of the uninduced and induced CAP5.5V bloodstream populations for the detailed analysis of the morphological effects of CAP5.5V depletion.

SEM of CAP5.5V mutants provides evidence of aberrant cytokinesis in RNAi induced bloodstream cultures.

For the detailed analysis of the cell morphology of uninduced and induced (12, 18 and 24h) bloodstream CAP5.5V RNAi mutants (5.5V-C-U cell line) cells were fixed for SEM. Uninduced cells showed the typical morphology of bloodstream trypomastigotes (Fig. 4.50) whereas induced cultures showed abnormalities in morphology after only 12h of induction (Fig. 4.51). Aberrant cell morphologies included cells with aberrant cleavage furrows, cells with wide posterior ends and multi-flagellate cells where abscission is likely to produce zoids. By 18h of induction, multi-flagellate cells showed contorted cell bodies and increasing numbers of flagella and cleavage furrows (Fig. 4.52). An intriguing kind of abnormal cell were uniflagellate cells with very small and wide cell-body encircled by the flagellum, which were increasingly found in induced cultures as a function of time (Fig. 4.53). These abnormal cells could represent the rising numbers of abnormal 1K1N cells observed by DAPI staining (Fig. 4.42).

TEM of BSF CAP5.5V mutants revealed disorganisation of the microtubule corset.

To study the effects of CAP5.5V silencing in bloodstream form trypanosomes at the ultrastructural level, cultures were fixed for thin section electron microscopy after 0, 12 and 24h of induction. Transverse sections of not-induced cells exhibited the

Figure 4.50.

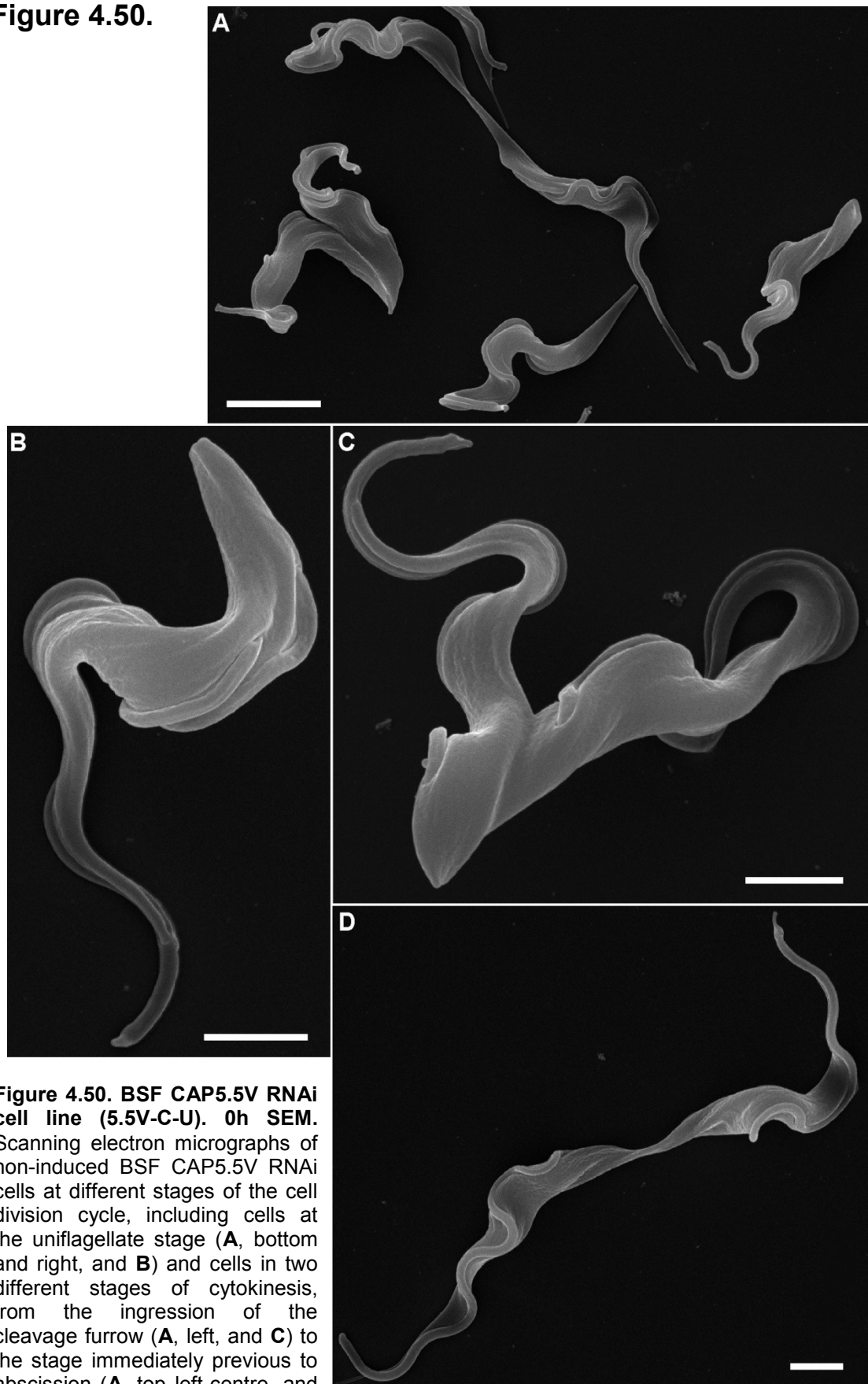


Figure 4.50. BSF CAP5.5V RNAi cell line (5.5V-C-U). 0h SEM. Scanning electron micrographs of non-induced BSF CAP5.5V RNAi cells at different stages of the cell division cycle, including cells at the uniflagellate stage (**A**, bottom and right, and **B**) and cells in two different stages of cytokinesis, from the ingression of the cleavage furrow (**A**, left, and **C**) to the stage immediately previous to abscission (**A**, top left-centre, and **D**). A scale bar represents 5 μ m. B to D scale bars represent 2 μ m.

Figure 4.51.

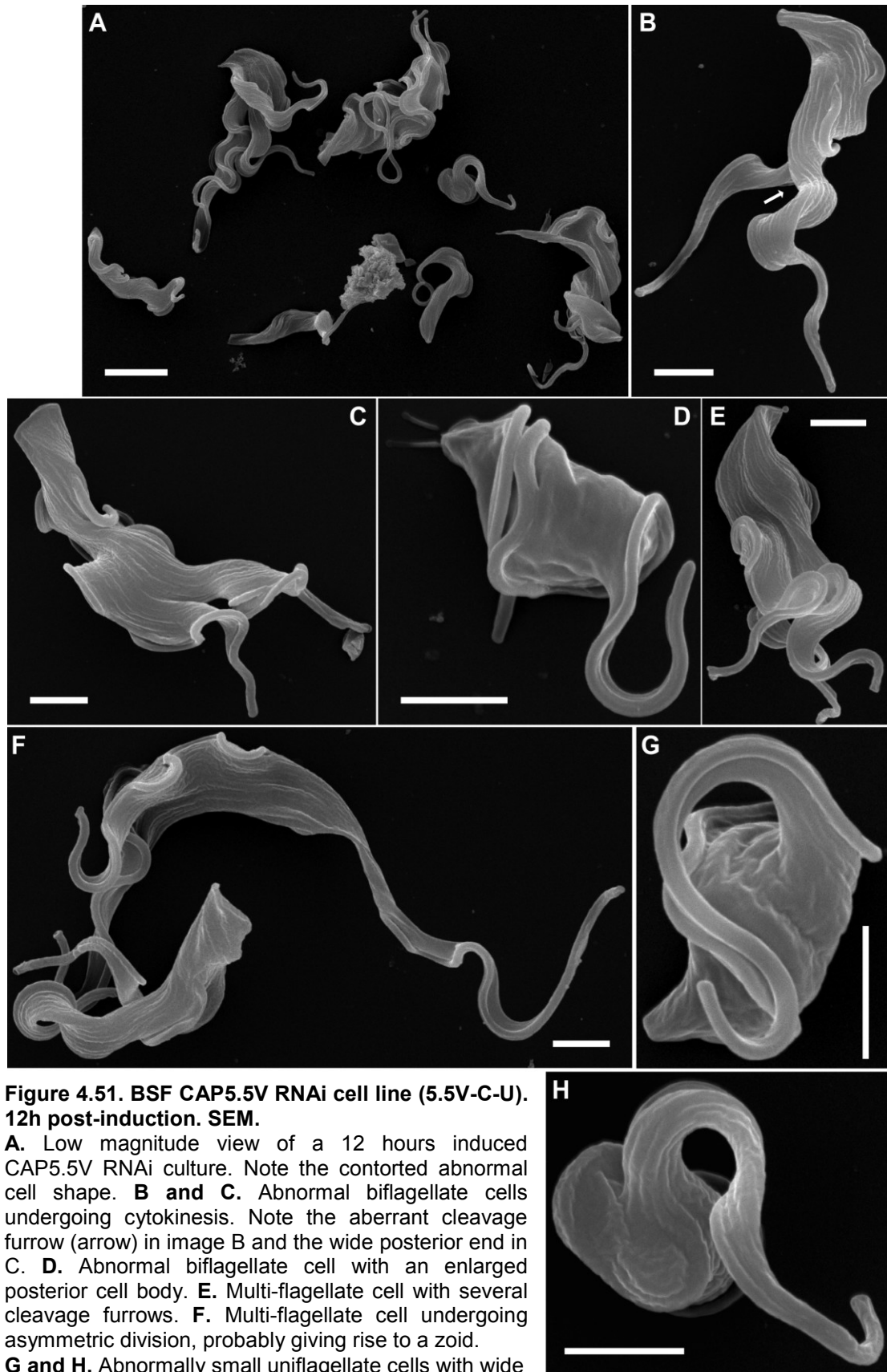


Figure 4.51. BSF CAP5.5V RNAi cell line (5.5V-C-U). 12h post-induction. SEM.

A. Low magnitude view of a 12 hours induced CAP5.5V RNAi culture. Note the contorted abnormal cell shape. **B and C.** Abnormal biflagellate cells undergoing cytokinesis. Note the aberrant cleavage furrow (arrow) in image B and the wide posterior end in C. **D.** Abnormal biflagellate cell with an enlarged posterior cell body. **E.** Multi-flagellate cell with several cleavage furrows. **F.** Multi-flagellate cell undergoing asymmetric division, probably giving rise to a zoid.

G and H. Abnormally small uniflagellate cells with wide cell-bodies encircled by the flagellum.

A scale bar represents 5 μm . B to H scale bars represent 2 μm .

Figure 4.52.

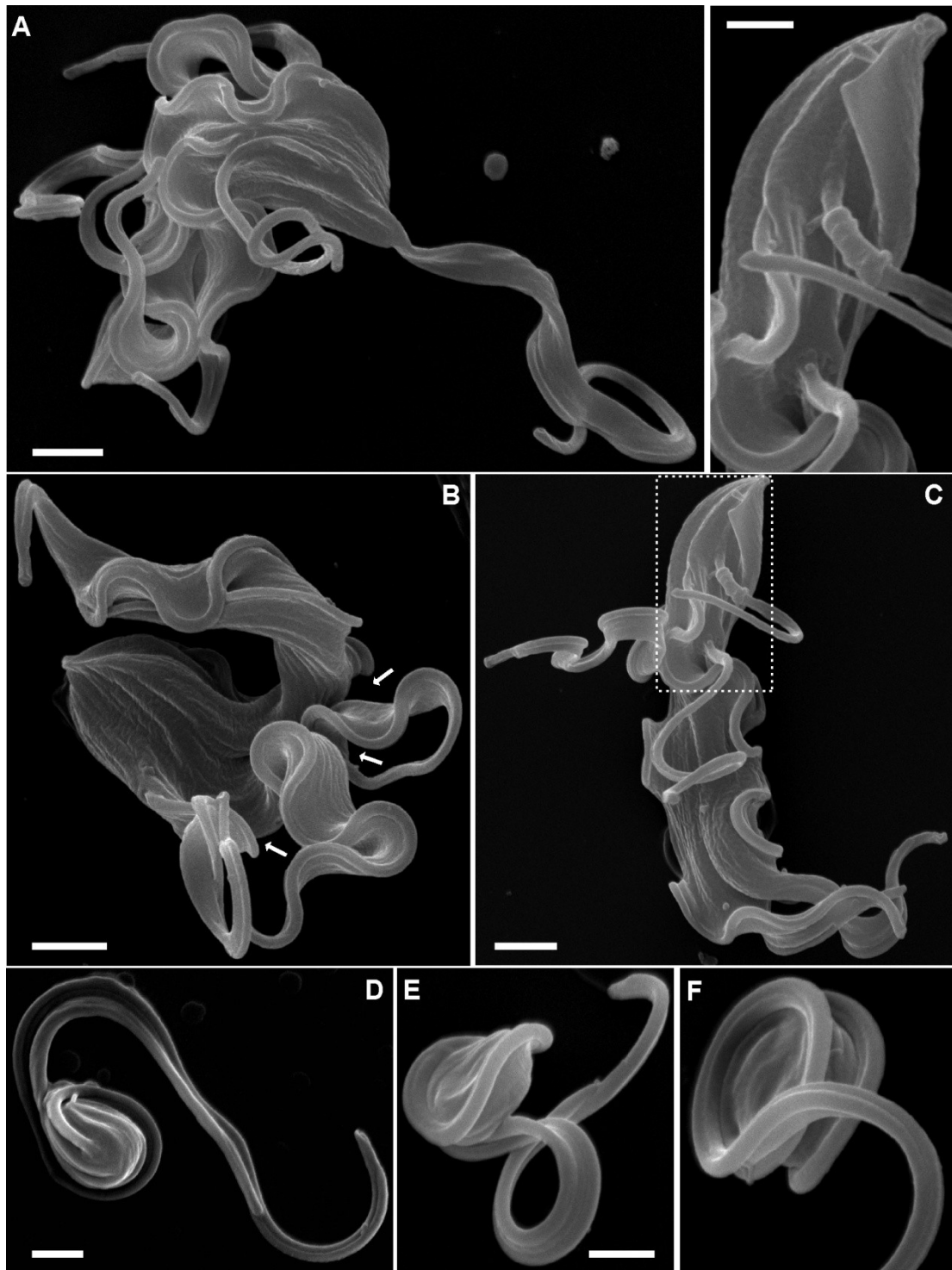


Figure 4.52. BSF CAP5.5V RNAi cell line (5.5V-C-U). 18h post-induction. SEM.

A. Multi-flagellate cell undergoing asymmetric cytokinesis, probably giving rise to a zoid.

B. Multi-flagellate cell with several cleavage furrows (arrows).

C. Multi-flagellate cell, with all flagella emerging from the posterior (enlarged and inset).

D-F. Abnormal uniflagellate cells with wide and small cell-bodies. A-C scale bars represent 2 μm . D-F and inset scale bars represent 1 μm .

Figure 4.53.

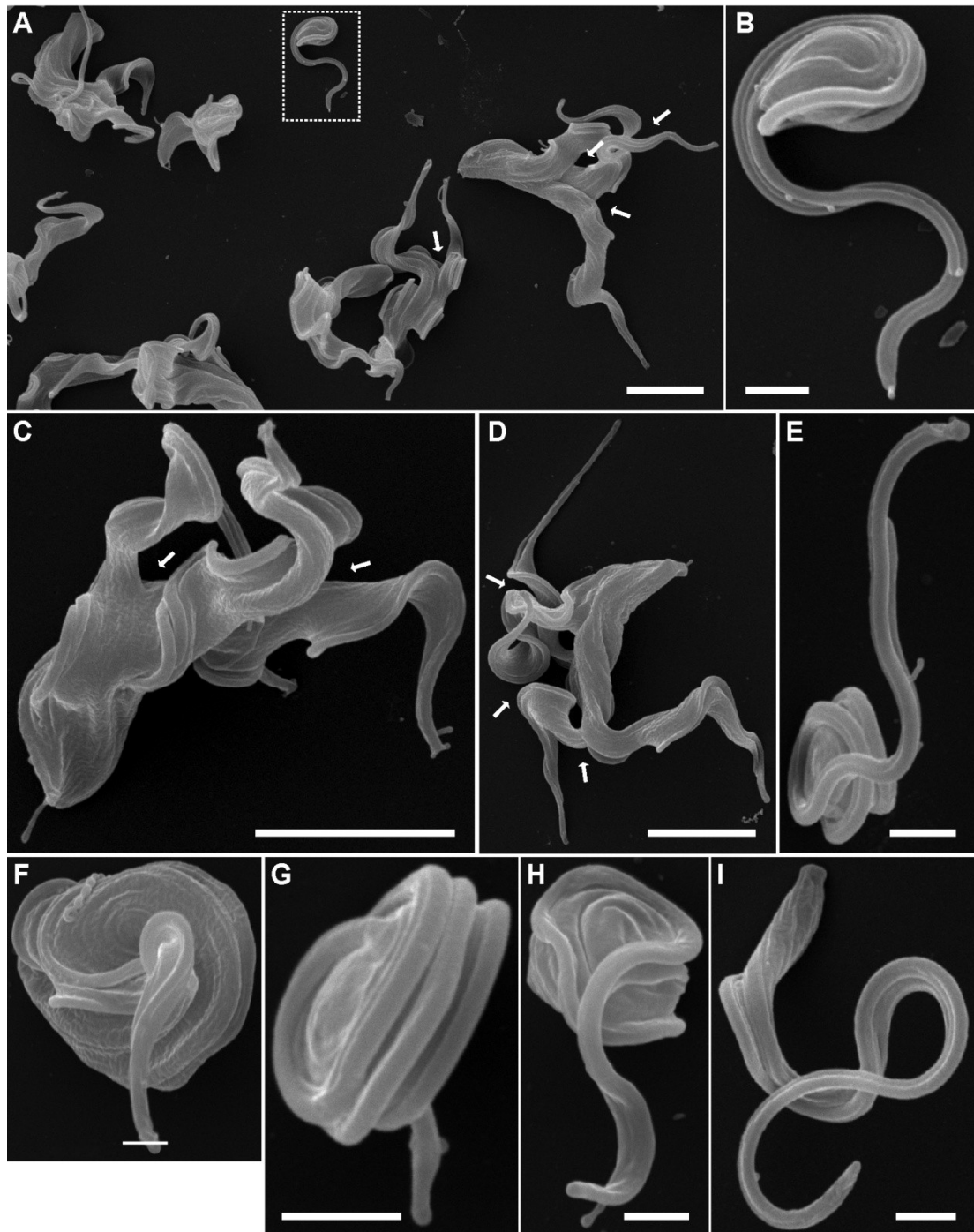


Figure 4.53. SEM of BSF CAP5.5V RNAi cell line (5.5V-C-U). 24h post-induction.

A. Low magnitude image of a 24h induced CAP5.5V RNAi population. Note the absence of normal-looking cells and the difference in size between uniflagellate (boxed and enlarged as image B) and multi-flagellate cells (aberrant cleavage furrows are indicated by arrows).

B. Enlarged view of a uniflagellate cell from image A (boxed). Note the wide and small cell-body. **C-D.** Multi-flagellate cells with several cleavage furrows (arrows), which abscission would probably give rise to zoids. **E-H.** Uniflagellate cells with wide and small cell-bodies encircled by the flagellum. **I.** Slender uniflagellate cell that could possibly be a zoid.

A, C and D scale bars represent 5 µm. B, E, and F to I scale bars represent 2 µm.

characteristic, evenly spaced microtubule monolayer that lies underneath the plasma membrane of normal trypanosome cells (Fig. 4.54A). However, by 12h of RNAi induction, defects in microtubule organisation were found in ~10% of transverse cell sections (N=350) (Fig. 4.54B-D and 4.55A-B). Abnormalities consisted in aberrant inter-microtubule spacing and/or the bundling of additional microtubules into the cytoplasm, both often found near the FAZ. Posterior cell sections frequently showed the presence of abnormally large flagellar pockets (Fig. 4.55A), whereas sections anterior to the flagellar pocket (evidenced by the presence of an attached flagellum) often showed microtubule disorganisation, sometimes involving the microtubule quartet (MtQ). Judging by the width of the cell-body, many of these “anterior” sections would correspond to anterior cell ends or, alternatively, to sections through slender zoids (Fig. 4.54B-D and 4.55B). After 24h of induction, many cell sections were characterised by multinucleate profiles and abnormal cell shape. About 30% of sections (N=400) revealed some form of microtubule abnormality (Fig. 4.55C-D, Fig. 4.56 and 4.57). Additional microtubules beneath the plasma membrane were often found near the FAZ, frequently accompanied by disorganisation of the MtQ (Fig. 4.55C-D, 4.56A and C). Sections through multi-nucleate cells also showed finger-like projections, in which abundance of microtubule bundling and abnormalities in the MtQ were frequent (Fig. 4.55C-D and 4.57A). These cell projections could correspond to examples where cleavage furrow ingression and subsequent abscission would give rise to zoids. Additionally, multinucleate cell sections containing flagellar pocket profiles often showed the presence of the paraflagellar rod (PFR) within the flagellar pocket, whereas in normal cells the PFR is only present and attached to the axoneme from the point where the flagellum exits the pocket. Thus, defects in microtubule organisation correlated with difficulties in flagellar pocket biogenesis (Fig. 4.56B-C and 4.57C-D).

Figure 4.54.

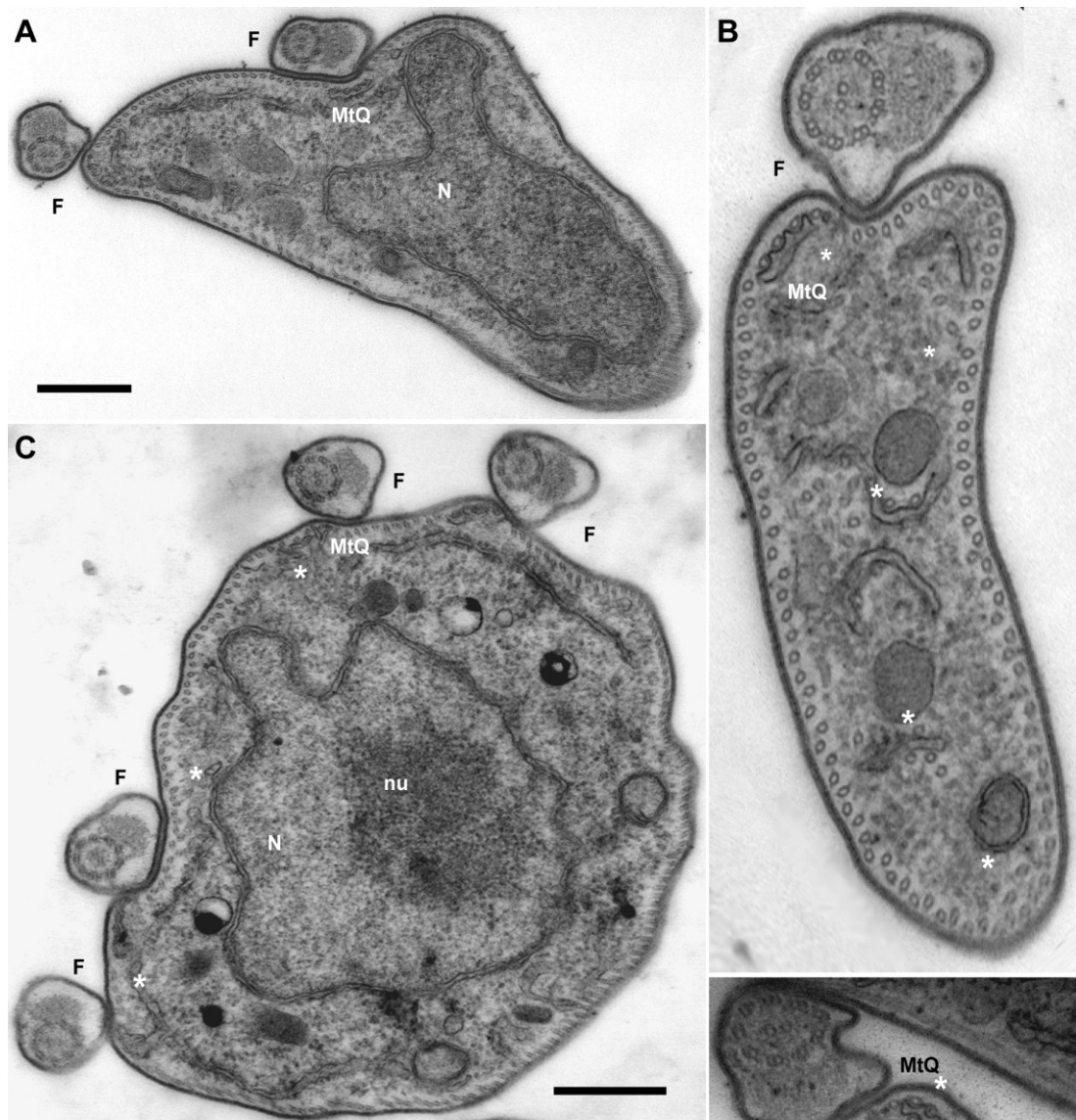


Figure 4.54. BSF CAP5.5V RNAi cell line (5.5V-C-U). TEM.

A. 0h transverse section through the mid-body, as evidenced by a nuclear (N) and two flagellar (F) profiles, of a non-induced CAP5.5V RNAi cell. Note the evenly spaced monolayer of microtubules underneath the plasma membrane. **B, C and D:** 12h induced CAP5.5V RNAi cell sections. **B.** Transverse section through an anterior cell end (the attached F profile identifies this as a section anterior to the flagellar pocket). Note the presence of additional microtubules across the cytoplasm (*).

C. Transverse section through a cell mid-body, as evidenced by the N, nucleolar (nu) and F profiles. Note the disorganisation of the sub-pellicular monolayer of microtubules (*), especially near the FAZ. **D.** Transversal section through an anterior cell end or a zoid (the attached flagellum identifies this as a section anterior to the flagellar pocket, its reduced cell width and size as a cell end or a slender zoid). Note the disorganisation of the subpellicular cytoskeleton, especially the MtQ (evidenced by its association to endoplasmic reticulum membranes) compared to the control cell (A) and the presence of additional microtubules in the cytoplasm (*). A, B and C scale bars represent 500 nm. D scale bar represents 200 nm.

Figure 4.55.

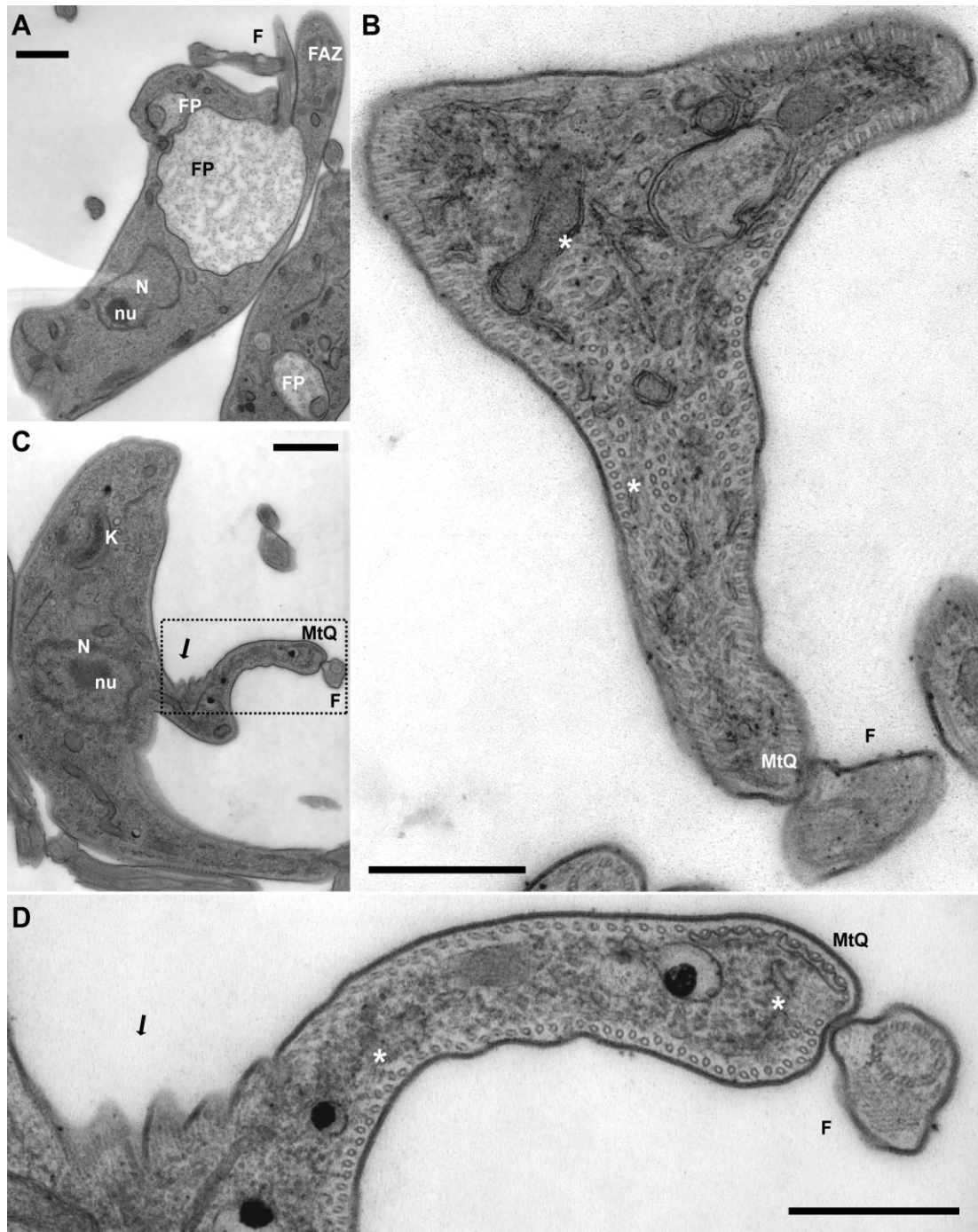


Figure 4.55. BSF CAP5.5V RNAi cell line (5.5V-C-U). 12 and 24h post-induction. TEM.

A. Transverse section through a posterior cell end, as evidenced by the flagellar pocket (FP) profiles, after 12h of induction. Note the enlarged size of the FP on the right. **B.** Transverse section through a 12h induced cell at a position anterior to the FP, as evidenced by the attached flagellum (F), revealing numerous additional microtubules across the cytoplasm (*). **C.** Longitudinal section through a 24h induced cell, as evidence by the kinetoplast (K), nuclear (N) and nucleolar (nu) profiles. The boxed finger-like projection on the right and enlarged as image **D**, is a probable example of a cleavage furrow ingression, which subsequent abscission would give rise to a zoid. Note the abnormal MtQ, which involves 9 subpellicular microtubules instead of 4 and the disorganisation of the microtubule monolayer (*), specially near the FAZ. A and C scale bars represent 1 μ m. B and D scale bars represent 500 nm.

Figure 4.56.

BSF CAP5.5V RNAi cell line (5.5V-C-U). 24h post-induction. TEM.

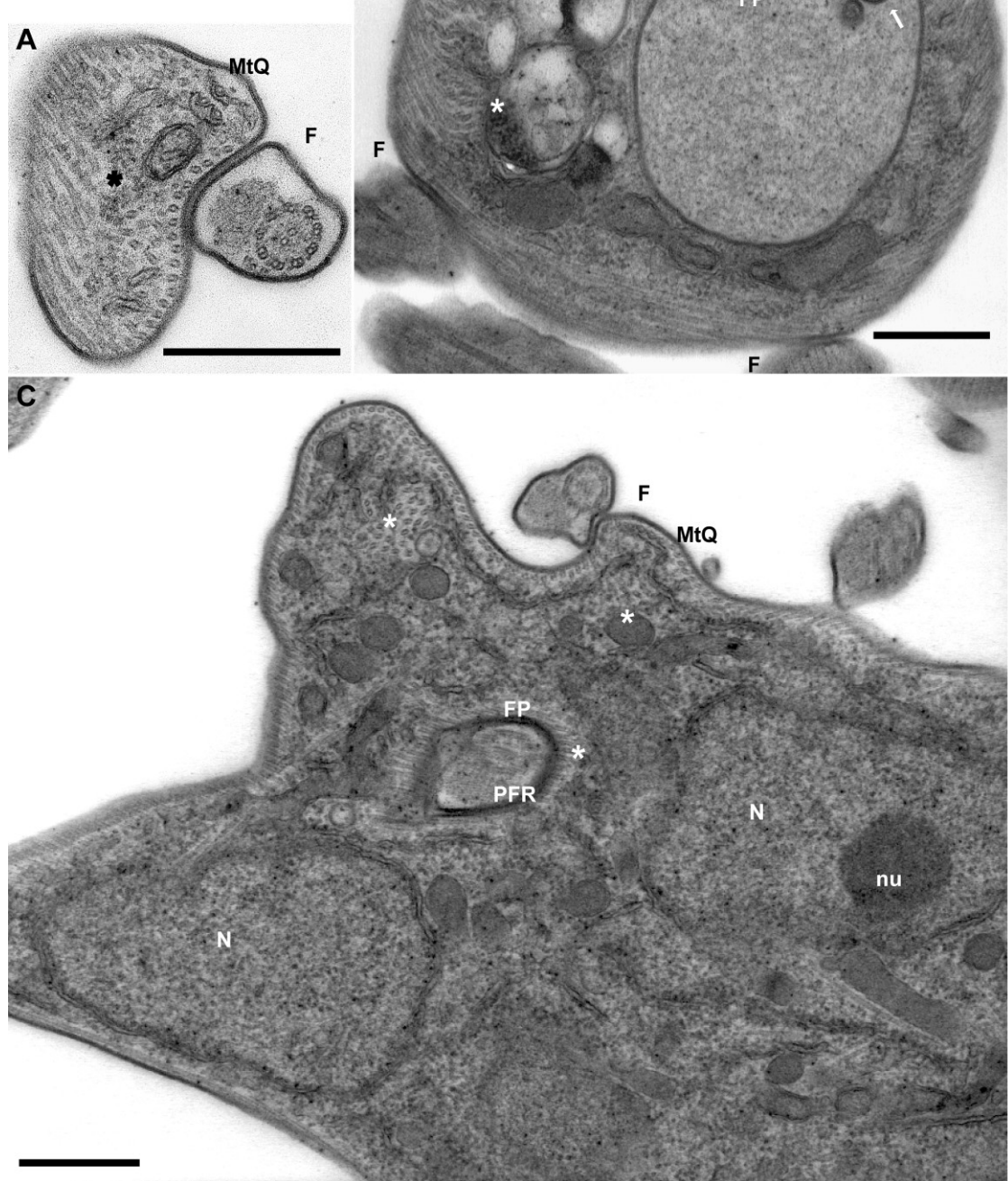


Figure 4.56. BSF CAP5.5V RNAi cell line. TEM. 24h. A. Transverse section through a cell at a position anterior to the flagellar pocket (FP), evidenced by the attached flagellum (F). The reduced cell width could identify it as an anterior cell-end or a zoid. Note the aberrant MtQ and the additional microtubules (Mtb) across the cytoplasm (*). **B.** Transverse/glancing section through a posterior cell-end, evidenced by the FP profile. Note the abundance of additional Mtb lying beneath the plasma membrane (*). The enlarged FP contains an axoneme with an incipient PFR (arrow), indicating abnormalities in FP biogenesis. **C.** Transverse /glancing section through an aberrant cell, evidenced by the presence of a FP, an attached F and two nuclear (N) profiles. Note the sectioning of an axoneme plus a PFR inside the FP and the high number of additional Mtb in the cytoplasm (*). Scale bars represent 500 nm.

Figure 4.57.

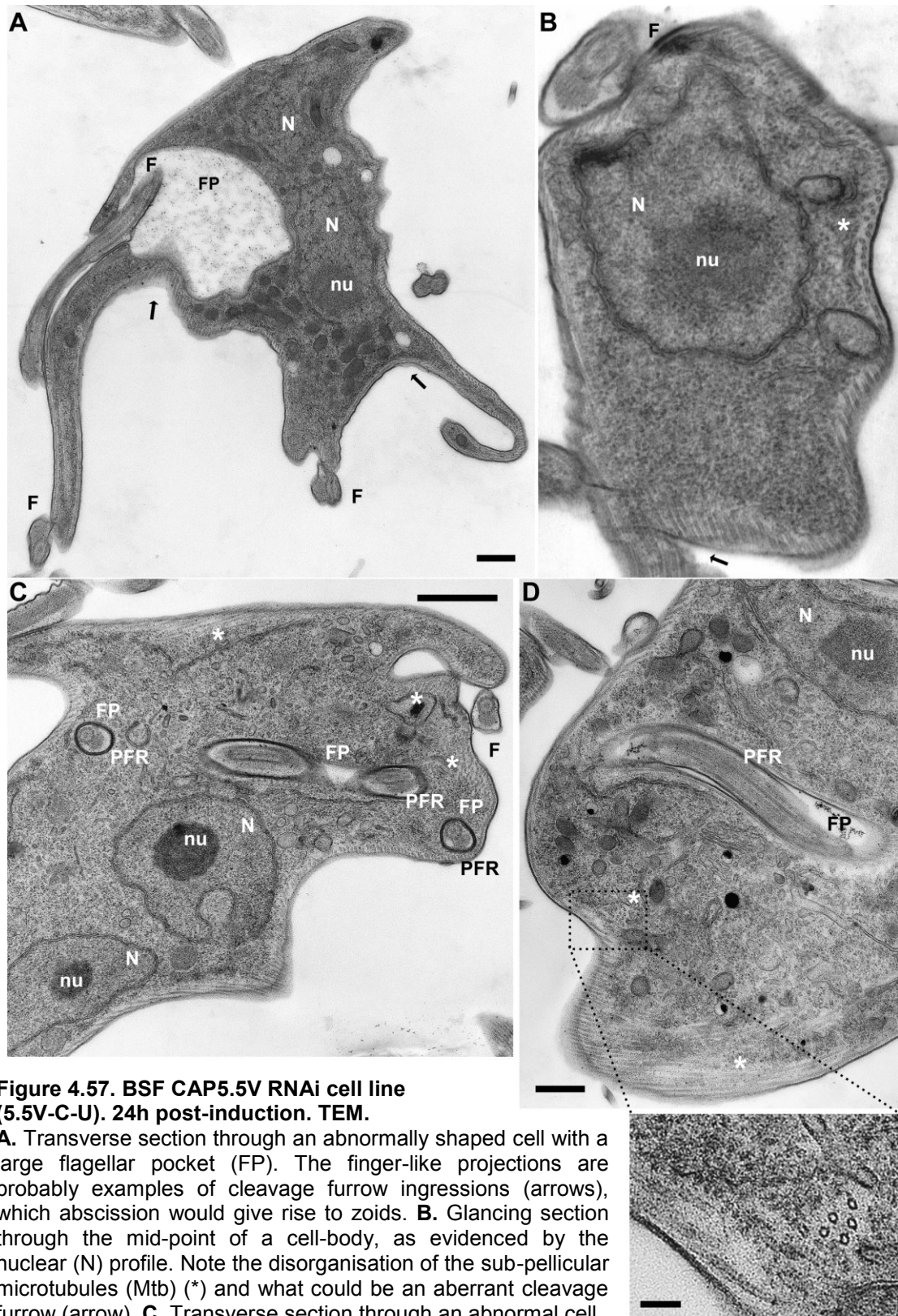


Figure 4.57. BSF CAP5.5V RNAi cell line (5.5V-C-U). 24h post-induction. TEM.

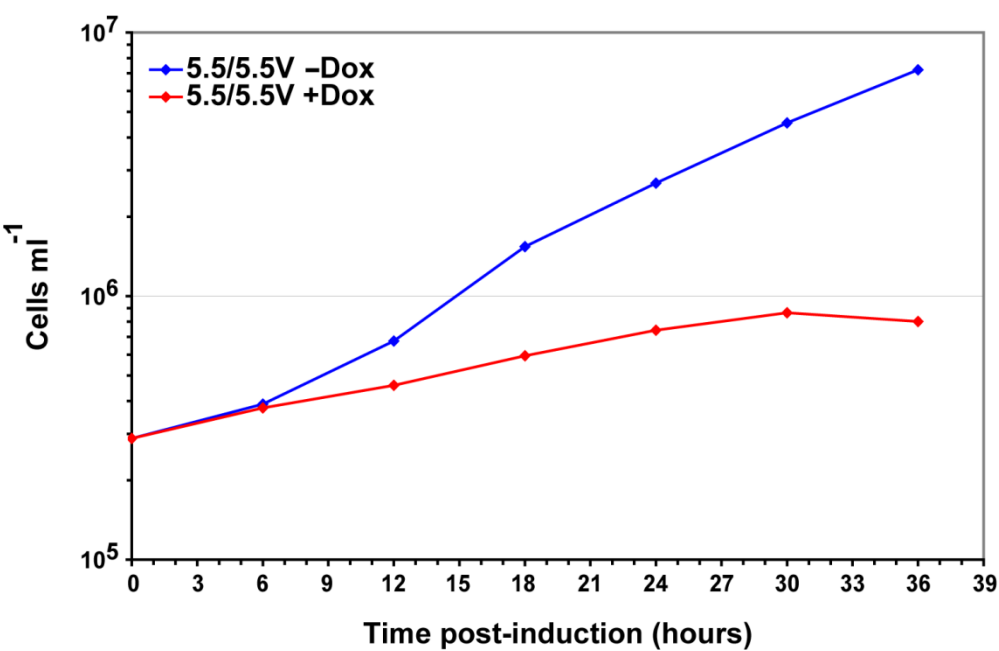
A. Transverse section through an abnormally shaped cell with a large flagellar pocket (FP). The finger-like projections are probably examples of cleavage furrow ingressions (arrows), which abscission would give rise to zoids. **B.** Glancing section through the mid-point of a cell-body, as evidenced by the nuclear (N) profile. Note the disorganisation of the sub-pellicular microtubules (Mtb) (*) and what could be an aberrant cleavage furrow (arrow). **C.** Transverse section through an abnormal cell with three FP profiles. The left hand side FP abnormally contains two flagella; all FPs contain a F with an axoneme plus a PFR, indicating difficulties in FP biogenesis. **D.** Section through an abnormal cell containing a FP with an axoneme plus PFR, additional Mtb across the cytoplasm (*, enlarged inset) and disorganisation of the sub-pellicular Mtb corset. Scale bars represent 500 nm. Inset scale bar represents 100 nm.

4.6.4. CAP5.5/CAP5.5V RNAi silencing phenotype in bloodstream *T. brucei*.

To test the effect of simultaneous silencing of both CAP5.5V and CAP5.5 in bloodstream trypanosomes, I used the RNAi construct that targets the highly similar N-terminus sequences of CAP5.5V and CAP5.5 (5.5/5.5V RNAi construct, Fig. 4.3). Induction of RNAi cultures transfected with this construct yielded a phenotype highly comparable to that produced by CAP5.5V RNAi specific silencing. Induced bloodstream 5.5/5.5V RNAi cultures showed a decrease in their growth rate after 9h of induction (Fig. 4.58 and Appendix 4, Fig. A.4.12), a slightly later time-point than in CAP5.5V specific RNAi cultures (~6h post-induction). Analysis of the number of nuclei and kinetoplast in DAPI stained cells revealed the increase in the same kind of abnormal cells —zoids, 1K2N and multinucleate cells— than in CAP5.5V specific silencing, also at a slightly later time ~9h post-induction (as opposed to 6h for CAP5.5V, Fig. 4.59 and Appendix 4, Fig. A.4.13). Thus, the onset of morphological abnormalities correlated with the reduction of the growth rate. Fluorescence microscopy of DAPI stained induced cells also illustrated aberrancies in cell morphology and division. Not-induced cells showed the characteristic morphology of bloodstream trypanosomes. After 9h of induction, 5.5/5.5V RNAi cultures showed increasing amounts of cells with nuclei and/or kinetoplasts mispositioning and wide posterior ends, often undergoing asymmetric cytokinesis. By 24h of induction, the most prevalent cell types were multinucleate cells, followed by abnormal 1K1N cells (with kinetoplast mispositioning and wide anterior cell ends). Abundant zoids were also observed in 24h induced RNAi cultures (Fig. 4.59).

Scanning electron microscopy experiments were performed with uninduced and induced (6, 12 and 24h) 5.5/5.5V RNAi cultures. Uninduced populations showed the typical morphology of bloodstream trypomastigotes (Fig 4.60A), whereas induced

Figure 4.58.



CAP5.5V-3-UTR		
RNAi cell line Doubling Time		
Hours post-induction	-Dox	+Dox
12	7.6	21.1
18	5.0	16.1
24	7.5	18.6
30	7.9	27.3
36	9.0	-54.4

Average doubling time of not-induced cells: 7.4h.

Figure 4.58. BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V). Cumulative growth curve. Induction of RNAi cultures produced a phenotype of slow growth first apparent after 9h of induction.

Figure 4.59.

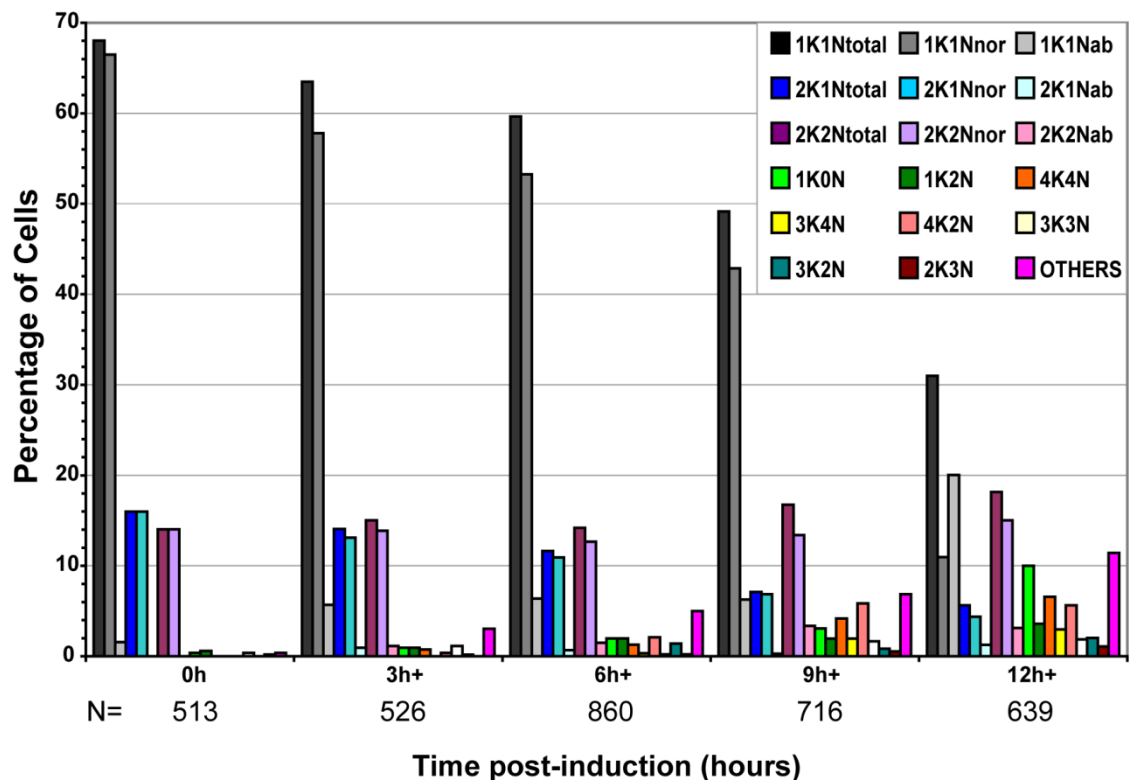


Figure 4.59. BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V).

Counts of number of nuclei and kinetoplast in DAPI-stained populations of 5.5/5.5V uninduced (0h, -Dox) and induced (+Dox) cultures after 0, 3, 6, 9 and 12h of RNAi induction.

The number of abnormal 1K2N cells and 1K0N "zoids", products of aberrant cytokinesis, increased noticeably and proportionally after 6-9h of induction. Numbers of multinucleate cells, named here as "others", many containing a higher number of nuclei than kinetoplasts, increased remarkably after 9-12h of induction, suggesting that the 1K2N progeny of aberrant cell division were able to enter into further cell division-cycles but failed to complete normal cytokinesis.

Zoids continued to accumulate over time, suggesting zoid longevity and/or the ability of multinucleate cells to undergo cytokinesis and abscission, giving rise to multiple zoids. Note the increase in the number of 2K2N after 9h of induction, reflecting the reduction in growth rate of induced cultures.

Figure 4.60.

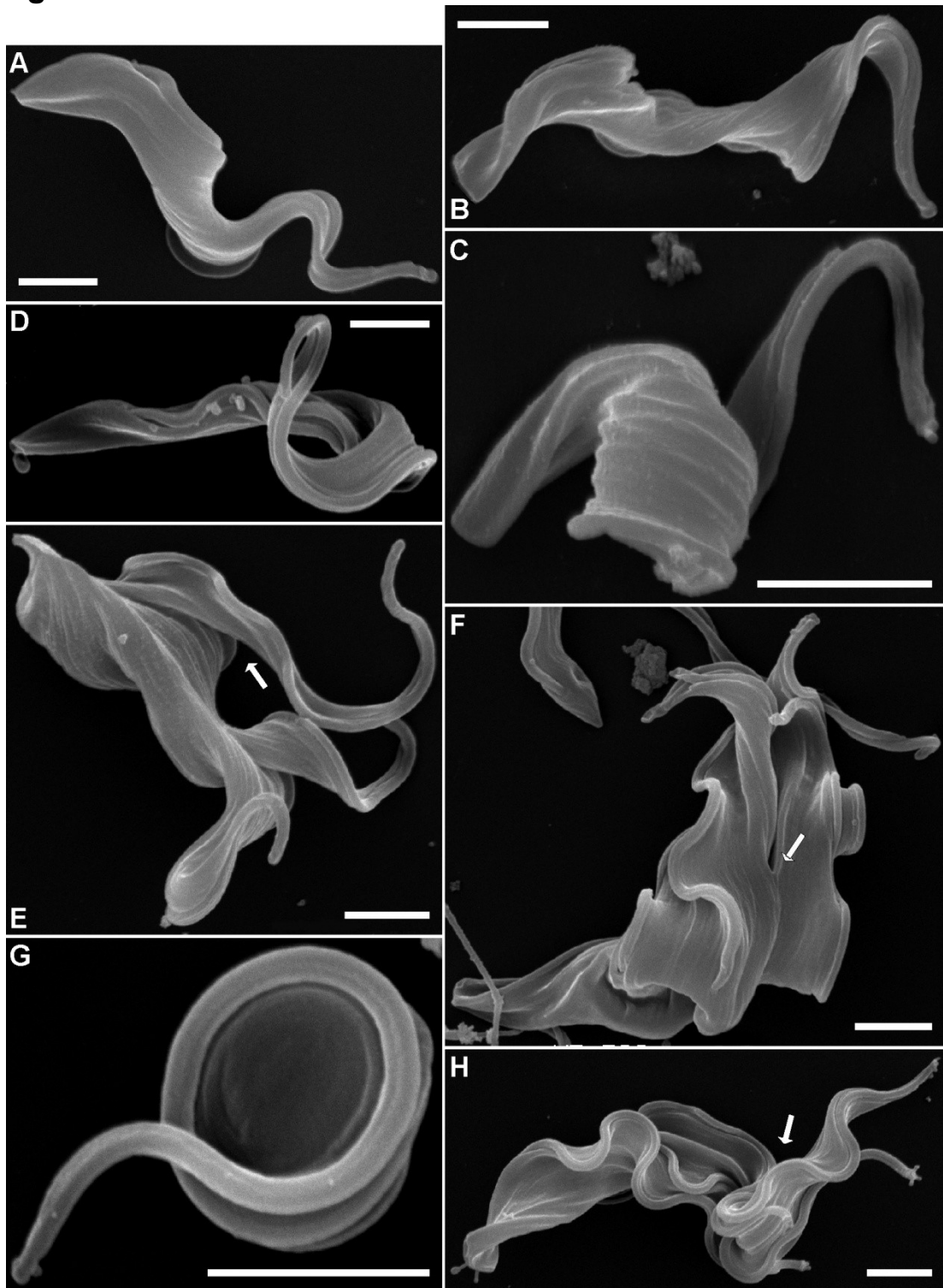


Figure 4.60. SEM of BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V).

A. Non-induced cell exhibiting the normal morphology of BSF trypomastigotes.

B-C. 6h induced cells showing abnormal morphology (wide posterior end and mid-body, respectively).

D and E. 12h induced cells. **D.** Uniflagellate "cell" with the typical morphology of a zoid.

E. Multi-flagellate cell exhibiting a cleavage furrow which ingression would likely give rise to a zoid. **F-H.** 24h induced multi-flagellate cells with ingressing cleavage furrows (arrows).

G. Small cell with a wide cell-body.

Scale bars represent 2 μm .

cultures revealed morphological abnormalities by 6h of induction (Fig. 4.60B-C), showing cells with wide posterior ends, slender zoid-like “cells” and cells with enlarged mid-bodies (which could represent 1K2N cells). By 12 and 24h post-induction many multi-flagellate cells showed several cleavage furrows which ingression could give rise to multiple zoid-like “cells” (Fig. 4.60D-H).

CAP5.5/CAP5.5V RNAi cultures were also fixed for thin section microscopy after 0, 12 and 24h of induction. Uninduced cells showed the typical, evenly spaced microtubule monolayer underneath the plasma membrane (Fig. 4.61A). However, by 12h of induction, ~9% of cell sections (N=250) showed abnormalities in microtubule organisation, most noticeably aberrant microtubule inter-spacing and microtubule bundling across the cytoplasm (Fig. 4.61B). After 24h of induction, many cell sections were characterised by abnormal shape and multinucleate profiles, with a high proportion of cell sections (~20%, N=300) revealing microtubule abnormalities. The main difference from CAP5.5V induced cultures was the much lower proportion of sections from induced cells revealing paraflagellar rod profiles within the flagellar pocket.

Figure 4.61.

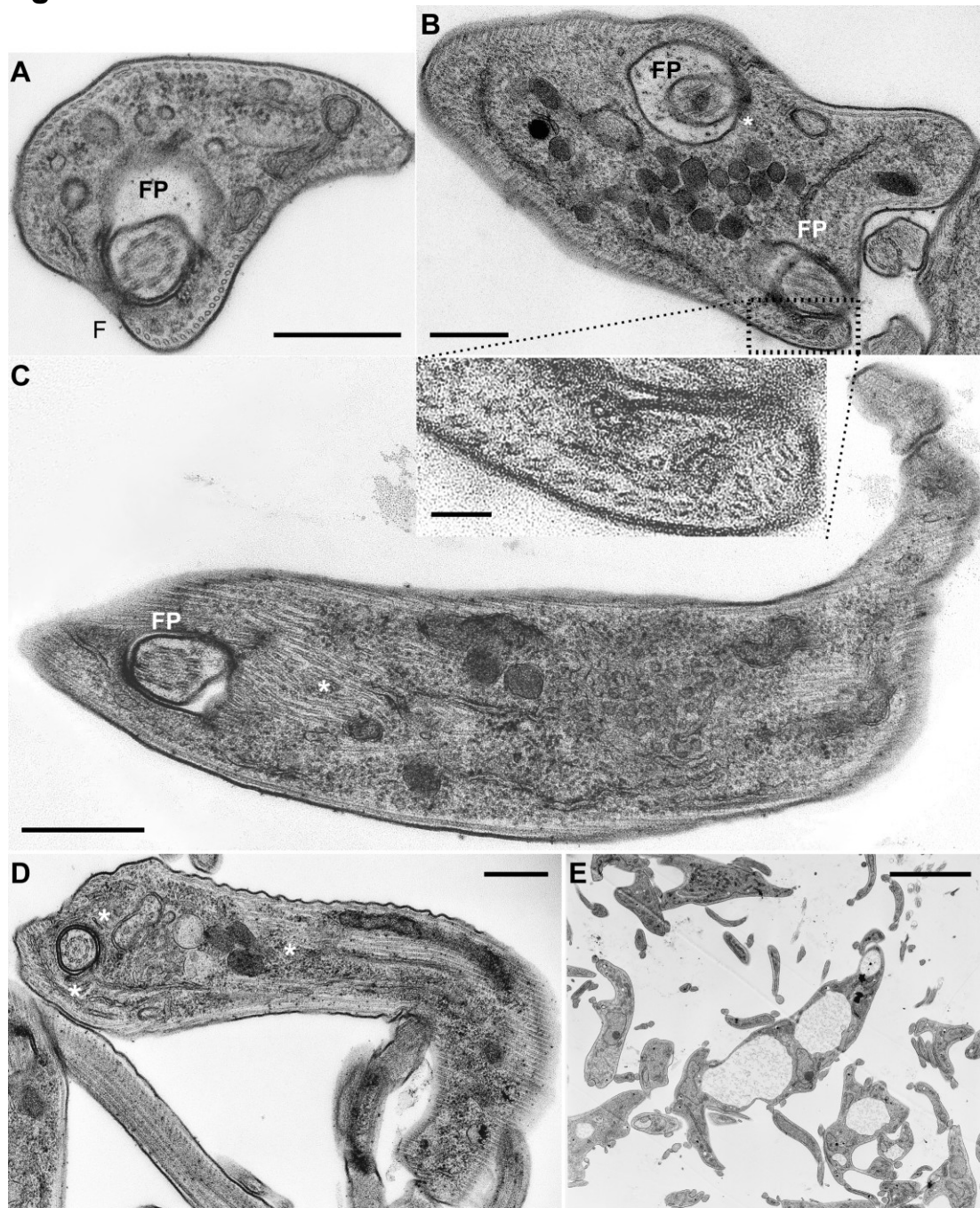


Figure 4.61. TEM of BSF CAP5.5/CAP5.5V RNAi cell line (5.5/5.5V).

A. Transversal section through a posterior end, as evidenced by the flagellar pocket (FP) profile, of a non-induced cell, exhibiting the characteristic, evenly spaced monolayer of microtubules underneath the plasma membrane. **B.** Section through the posterior end of a 12h induced cell, as evidenced by the FP profiles. Note the additional microtubules (*) and boxed area enlarged underneath) in the cytoplasm, close to the FPs. **C-E.** 24h induced cells. **C.** Glancing section through a posterior cell-end, evidenced by the FP profile. Note the longitudinal section of microtubules showing aberrant inter-microtubule spacing and additional microtubules (*) near the FP. **D.** Section through a cell-end, as evidenced by the FP profile, showing additional microtubules across the cytoplasm and besides the FP. **E.** Low magnification image of a 24h induced population. Note the presence of large FPs. A-D scale bars represent 500 nm. E scale bar represents 5 μ m. Inset scale bar represents 100 nm.

CHAPTER 5

RNAI ANALYSIS OF GROUP 4 CALPAIN-RELATED PROTEINS

5. RNAI ANALYSIS OF GROUP 4 CALPAIN-RELATED PROTEINS.

5.1. Introduction.

Ersfeld and co-workers bioinformatic analysis of trypanosomatid calpain-like sequences identified several open reading frames encoding proteins with long repetitive regions of 67-68 amino acids and classified them as group 4 and 5 calpain-like sequences, according to their domain architecture (*Ersfeld et al., 2005*). My analysis of the latter released *T. brucei* genome annotation version 4, resulted in the identification of several new sequences containing long arrays of tandem repeats that have not been previously described. I classified all trypanosomatid calpain-like sequences containing 67-68 amino acid tandem repeats as group 4 calpain-related sequences (Fig. 3.20). Short fragments containing the repeats of several group IV calpain-like sequences have been published as antigens FRA (*Lafaille et al., 1989*) and H49 (*Cotrim et al., 1990; Cotrim et al., 1995*) in *T. cruzi*, GM6 (*Muller et al., 1992; Muller et al., 1993*) in *T. brucei* and Lcr1 (*Wilson et al., 1995*), in *L. chagasi* (*Robinson et al., 1995; Davidge et al., 2006*). Furthermore, the bioinformatic analysis of *T. brucei* group 4 calpain-like sequences (Chapter 3, Section 3.2.3) revealed that their repeats regions are either homologous to GM6 (Tb11.57.0008 and Tb11.47.0036) or to Lcr1 repetitive antigens (Tb11.47.0035 and Tb11.V4.0001). Accordingly, the two different kinds of repeats were called GM6-like and Lcr1-like repeats.

FRA, GM6 and Lcr1 are homologs and all detect high molecular weight proteins by western blotting. Additionally, immuno-electron localisation analysis of FRA and GM6 showed that they are exclusively located at the FAZ of *T. cruzi* and *T. brucei* respectively, associated to the FAZ filament, the network of fibres extending between the microtubules of the cell-body and the flagellum (*Cotrim et al., 1990; Muller et al.,*

1992; Cotrim *et al.*, 1995). During cytokinesis, as the cleavage furrow progresses from the anterior end of the cells separating the flagella and the segregated kinetoplasts and nuclei, the FAZ is positioned between the new and old flagella. The position of the FAZ area facing the cytoplasm (FAZ filament and MtQ) of a dividing cell is thought to be implicated in defining the point of initiation and direction of the cleavage furrow (Robinson *et al.*, 1995).

All the studied trypanosomatid species have a minimum of three paralogous group 4 calpain-like sequences, always situated on a single chromosome and often in adjacent positions. The conservation of these sequences across different trypanosomatid species, the presence of paralogous genes and their specific localisation at the FAZ filament could indicate an important role for these proteins in kinetoplastids. In this chapter, I report the RNAi functional analysis of *T. brucei* group 4 calpain-like sequences, in both procyclic and bloodstream form parasites, aiming to elucidate their function within the FAZ but also to clarify the role of the FAZ complex (FAZ filament and MtQ) in cell morphogenesis and cytokinesis. However, it must be point out that the results of these RNAi experiments have to be validated by qRT-PCR data, to demonstrate the reduction in the expression of the RNAi targeted sequences, which were not performed due to a lack of time. Interestingly, upon induction, group 4 calpain-like proteins RNAi cell lines showed the appearance of epimastigote-like cells, with the kinetoplasts located in anterior or juxta-nuclear positions, short FAZs and long flagellar overhangs.

T. brucei epimastigotes are a developmental stage only found in the proventriculus and salivary glands of the tsetse fly (Vickerman, 1985). The main morphological differences between procyclic trypomastigotes and epimastigotes are the relative position of their nucleus, kinetoplast and flagellum along the anterior-posterior axis of the cells (Hoare, 1966). In the trypomastigote form, the kinetoplast is positioned between the posterior

end of the cell and the nucleus whereas in the epimastigote form it lies between the nucleus and the anterior end of the cell (*Vickerman, 1985*). The flagellar pocket, from which the flagellum emerges from the cell-body, is always close to the kinetoplast. Consequently, the flagellar pocket of trypomastigotes is posterior to the nucleus while the flagellar pocket of epimastigotes is in a pre-nuclear position. As the flagellum runs along the longitudinal axis of the cell and overhangs its anterior end, epimastigotes have a shorter FAZ than trypomastigotes (*Vickerman, 1985*). Additionally, *T. brucei* epimastigotes are significantly smaller than procyclic trypomastigotes (*Van Den Abbeele et al., 1999*)

Differentiation of procyclic trypomastigotes to the epimastigote morphology occurs during the parasite migration from the tsetse midgut to the salivary glands (*Lewis and Langridge, 1947; Vickerman et al., 1988; Van Den Abbeele et al., 1999; Aksoy et al., 2003*), through an asymmetric cell division (*Lewis and Langridge, 1947; Van Den Abbeele et al., 1999*). This differentiation involves the rearrangement of kinetoplast and nuclear positions along the anterior posterior axis of the cell and a remarkable reduction in cell size. The exchange in the relative positions of the nucleus and kinetoplast is achieved by the apparent movement of the nucleus towards the posterior end of the cell, with the consequent reduction of the distance between the nucleus and the posterior cell end. However, the distance between the kinetoplast and the posterior end of the cell remains relatively constant (*Sharma et al., 2008*). Once in the salivary glands, epimastigotes attach to the epithelium via an elaboration of the flagellar membrane (*Vickerman et al., 1988*). Attached epimastigotes are able to proliferate whilst remaining attached to the epithelium (*Vickerman et al., 1988*).

The results of the RNAi functional analysis of group 4 calpain-like proteins presented here, including the analysis of the morphology and morphogenesis of induced epimastigote-like cells, were addressed to the dependency relationships between the

FAZ components (FAZ filament and MtQ) during cell duplication, as well as the importance of such relationships in the correct cell morphogenesis of procyclic trypanosomes. These results strongly suggest an essential role for group 4 calpain-like proteins in cell morphogenesis and point towards the possibility of stage-specific regulation of their expression, in a similar fashion to CAP5.5 and CAP5.5V. Hence, the group 4 calpain-like RNAi phenotypes are also discussed in the context of a possible role for these proteins in the morphological variations between the lifecycle stages of *T. brucei* parasites.

Additionally, several antibodies recognising components of the FAZ have been identified in both *T. brucei* and *T. cruzi*: CD10, DOT1 (Woods *et al.*, 1989), Ag 2G4 (Ruiz-Moreno *et al.*, 1995), BS7 (Woodward *et al.*, 1995), L3B2 and L6B3 (Kohl *et al.*, 1999). However, the identities of the proteins recognised by two of these antibodies, CD10 and DOT1 (Woods *et al.*, 1989), are still not known. The immunofluorescence staining patterns of CD10 and DOT1 in *T. brucei* group 4 calpain-like proteins RNAi cell lines upon induction are also described in this chapter, showing that neither of these antibodies seems to recognise group 4 calpain-like proteins as RNAi induction only affected their localisation but not the intensity of their immunofluorescence signal. These data also suggest that group 4 calpain-like proteins have an important role in maintaining the structure of the FAZ.

5.2. Applying the RNAi test.

The current unfinished state of the calpain-like group 4 sequences genome locus in *T. brucei* (Fig 3.9A) does not allow one to positively distinguish between real ORFs and sequence fragments. The four group 4 calpain-like sequences identified in *T. brucei* (Tb11.47.0035, Tb11.47.0036 (Ersfeld *et al.*, 2005), Tb11.57.0008 and Tb11.V4.0001 (Chapter 3, Section 3.2.3) are present in three different contigs of Chromosome 11,

located in neighbouring positions of their provisional assembly (<http://www.genedb.org/tryp/>). The two larger sequences, Tb11.57.0008 and Tb11.47.0035 have a 95.3% nucleotide identity across 2940/2934 bp of their 5' subterminal regions, which include their domains IIa, IIIa, IIb and part of domain IIIb (Tb11.57.0008 49-2988 bp and Tb11.47.0035 46-2979 bp). Additionally, Tb11.57.0008 and Tb11.47.0035 have distinct long repeat regions, GM6-like and Lcr1-like repeats, respectively (Fig. 5.1A). Both sequences lack stop codons, whereas their two upstream adjacent sequences, Tb11.47.0036 and Tb11.V4.0001 respectively, start with 5'-repeats regions highly similar to the repeats of their downstream neighbours (Tb11.57.0008 5287-6796 bp and Tb11.47.0036 1-1493 bp have a 94.9% identity, whereas Tb11.47.0035 11254-11815 bp and Tb11.V4.0001 1-562 bp have a 99.8% identity, Fig. 3.10, 5.1B-C). Furthermore, Tb11.47.0036 lacks a start codon, strongly suggesting that Tb11.57.0008 and Tb11.47.0036 are part of the same ORF. Similarly, the 84 nucleotides of Tb11.V4.0001 hypothetical 5'-UTR are 97.6% identical to those preceding the Tb11.47.0035 region 99.8% identical to Tb11.V4.0001 5'-terminus (Tb11.47.0035 11169-11253 bp), suggesting that both sequences are also part of the same ORF. To elucidate if these four repeat-containing sequences are independent genes or parts of two distinct ORFs, Southern blot analysis should be carried out.

The presence of highly similar regions in *T. brucei* group 4 calpain-like sequences provides the opportunity to design RNAi constructs with the capability to knockdown the expression of more than one of these hypothetical proteins at a time. This approach was especially useful here, as, in the absence of Southern blot data, it is not known if these four sequences are or are not parts of two different ORFs. Consequently, the primers for the RNAi functional analysis of these hypothetical proteins were designed within those parts of their coding sequences shared by more than one putative ORF. Thus, four RNAi constructs were made in this fashion: HP-N, DS-N2, 08-36-Rep and

Figure 5.1

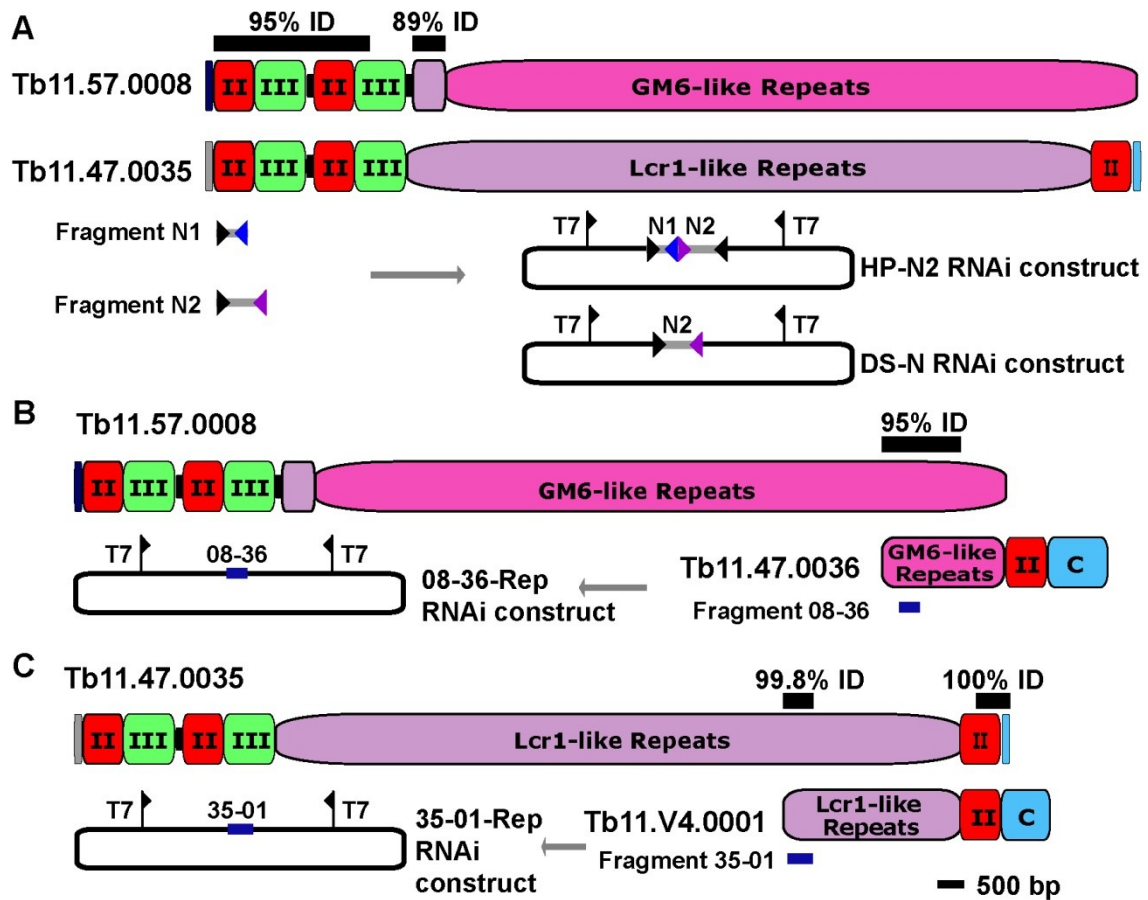


Figure 5.1. Schematic representation of *T. brucei* group 4 calpain-like sequences. Hypothetical domains are represented by boxes of different colour shades: red- calpain domain II, green- domain III and blue- C-terminal domain. The two different kinds of 68 amino acid (aa) repeats are also represented by different coloured boxes: pink- GM6-like repeats and purple- Lcr1-like repeats.

The DNA fragments cloned into the p2T7-177 vector are represented by bars underneath the sequence cartoons. Nucleotide (nt) sequence identities are represented by black bars above the sequence representations.

A. Tb11.57.0008 and Tb11.47.0035. 95.3% nt identity from 49 and 46 to 2988 and 2979 bp respectively, including domains IIa, IIIa, IIb and part of domain IIIb. Two fragments of different sizes (same forward primer, different reverse primer) were cloned in inverted directions into the p2T7-177 vector, rendering the HP-N RNAi construct. The longer fragment (N2) was also cloned into the RNAi vector, rendering the DS-N2 RNAi construct. Both HP-N and DS-N2 have the capability to knockdown the expression of Tb11.47.0008 and Tb11.47.0035 by targeting their domains IIa and IIIa.

B. Tb11.57.0008 and Tb11.47.0036. 95% nt identity from positions 5287 to 6769 and 1 to 1483 bp respectively, a region corresponding to GM6-like repeats. A 386 bp fragment of this region was amplified by anchor-PCR and cloned into the p2T7-177 RNAi vector, rendering the 08-36 Rep RNAi construct, which has the capability to knockdown the expression of Tb11.57.0008 and Tb11.47.0036 by targeting their repeats regions.

C. Tb11.47.0035 and Tb11.V4.0001. 99.8% nt identity (100% aa ID) from positions 11254 to 11815 and 1 to 562 bp, corresponding to a Lcr1-like repeats region. A 473 bp fragment of this sequence was cloned into the RNAi vector rendering the 35-01 Rep RNAi construct, which has the capability to silence the expression of both Tb11.47.0035 and Tb11.V4.0001 by targeting their repeats regions.

35-01-Rep. HP-N and DS-N2 have the capability to knockdown the two larger sequences, *Tb11.57.0008* and *Tb11.47.0035*, targeting their 95% identical 5'-subterminal regions (Fig. 5.1A). However, these two constructs differ in their approach to produce dsRNA. HP-N is a stem-loop RNAi construct, which contains two PCR amplified fragments: fragment N1 (601 bp) and fragment N2 (921 bp), Fig.5.1A). These two DNA fragments have the same starting point (same forward primer) but different lengths (different reverse primers). When the two fragments were cloned in opposite orientations into the p2T7-177 vector, flanked by the two opposing T7 promoters, the last 320 bp of the larger fragment acted as DNA “stuffer”, separating the inverted repeat fragments (Fig. 5.1A). DS-N2 is a conventional RNAi construct with opposing T7 promoters that flank a single fragment of the target gene (Fragment N2, 921 bp. Fig. 5.1A.). Thus, these two RNAi constructs were used for the direct comparison between the main approaches to produce dsRNA in *T. brucei*: stem-loop and opposing promoters RNAi constructs.

The 08-36-Rep RNAi construct has the capability to knockdown the expression of both *Tb11.57.0008* and *Tb11.47.0036*, by targeting their GM6-like repeats regions (Fig. 5.1B). These two sequences, which could be part of the same ORF, contain highly conserved GM6-like repeats (95% nucleotide identity). Thus, to obtain a DNA fragment for RNAi longer than one repeat unit (204 bp), I used an anchor PCR approach to amplify an initial fragment of 193bp, contained within one repeat unit. As a result, I obtained a 386 bp fragment, which contains almost two complete repeat units. This fragment was then cloned between the T7 opposing promoters of the p2T7-177 RNAi vector, rendering the 08-36-Rep RNAi construct. The 35-01-Rep RNAi construct has the capability to knockdown the expression of *Tb11.47.0035* and *Tb.V4.0001* hypothetical proteins by also targeting their repeats regions (Fig. 5.1C). Although the two sequences contain highly divergent Lcr1-like repeats, they have 562 bp 99.8%

identical (Tb11.47.0035 11254-11815 bp and Tb11.V4.0001 1-562 bp), indicating that they could be part of the same ORF. The divergence between their Lcr1-like repeats allowed the amplification of a 473 bp fragment of their similar regions by a conventional PCR approach (Tb11.47.0035 11298-11770 bp and Tb11.V4.0001 45-517 bp). This fragment was also cloned between the head-to-head T7 promoters of the p2T7-177 plasmid, rendering the 35-01-Rep RNAi construct. As discussed in Chapter 3, Tb11.57.0008 contains three Lcr1-like repeat units, 89.2% identical to the first Lcr1-like repeat units of Tb11.47.0035, and which are located in equivalent positions, upstream the highly similar 5'-subterminal region (Tb11.57.0008 3750-4362 bp and Tb11.47.0035 4008-4623 bp, Fig. 5.1A). However, the DNA sequence of these Lcr1-like repeats is divergent enough not to produce cross-reaction with the 473 bp fragment used to make the 35-01-Rep construct.

5.3. RNAi functional analysis of group 4 calpain-like sequences in PCF *T. brucei*: silencing Tb11.57.0008 and Tb11.47.0035.

The present section describes the results of the RNAi functional analysis of group 4 calpain-like proteins carried out in procyclic *T. brucei*. As discussed earlier, the HP-N and DS-N2-RNAi cell lines have been genetically engineered to allow the knockdown of the two large group 4 calpain-like sequences, Tb11.57.0008 and Tb11.47.0035, if they were simultaneously expressed in procyclic trypanosomes. Both the HP-N and DS-N2 RNAi constructs target the 95% identical region of Tb11.47.0035 and Tb11.57.0008, which encodes these proteins calpain-like domains IIa and IIIa. Although this region contains the motifs characteristic of kinetoplastid calpain-like domains II and III, its nucleotide sequence differs enough from any other *T. brucei* calpain-like protein to prevent RNAi cross-reaction. Procyclic and bloodstream

trypanosomes were transfected with these two RNAi constructs, but selected transfectants were only obtained in the procyclic form. A possible explanation for the lack of success in obtaining transfectants in the bloodstream form (despite several attempts) could be a loose downregulation of RNAi in the absence of induction, which, if these proteins are essential, could have a lethal effect in emerging transfectants.

5.3.1. RNAi knockdown of group 4 calpain-like proteins produces a reduction in growth rate and cell-body size.

Not-induced DS-N2 and HP-N RNAi cell lines grew well in culture, with a doubling time of approximately 12h, which is comparable to that of the parental 29.13 cell line (*Wirtz et al., 1999*). Simultaneous inductions of both cell lines were carried out in two different occasions elapsed by a 30-day time period. During this time, cell lines were kept in culture under selective drug pressure until 48h prior to RNAi induction. Addition of doxycycline to the culture medium produced a growth defect in both DS-N2 and HP-N RNAi cell lines, which was first apparent after 40-50h of induction (Fig. 5.2A-B). The growth of neither cell line was halted after this time; instead, they continued to grow, albeit at a slower rate (doubling time of induced cells was ~14.5h, as opposed to 12h of not-induced controls). However, the growth rate of HP-N RNAi induced cultures started to increase again after ~125h of induction (Fig. 5.2B). This was not observed in the DS-N2 RNAi cell line, which slow growth rate was maintained over time (160h of induction, Fig. 5.2A-B).

Although cell growth was not compromised up to 40-50h post-induction in neither induced cell line, important defects in cell morphology and morphogenesis were detected as early as 18h of induction (Fig. 5.3). Changes in cell size, monitored using an aperture-impedance CASY[®] cell counter, were detected after 24h of induction. However, changes in cell size distribution varied between cell lines. The two inductions of the DS-N2 RNAi cell line produced very similar results, both showing a progressive

Figure 5.2.

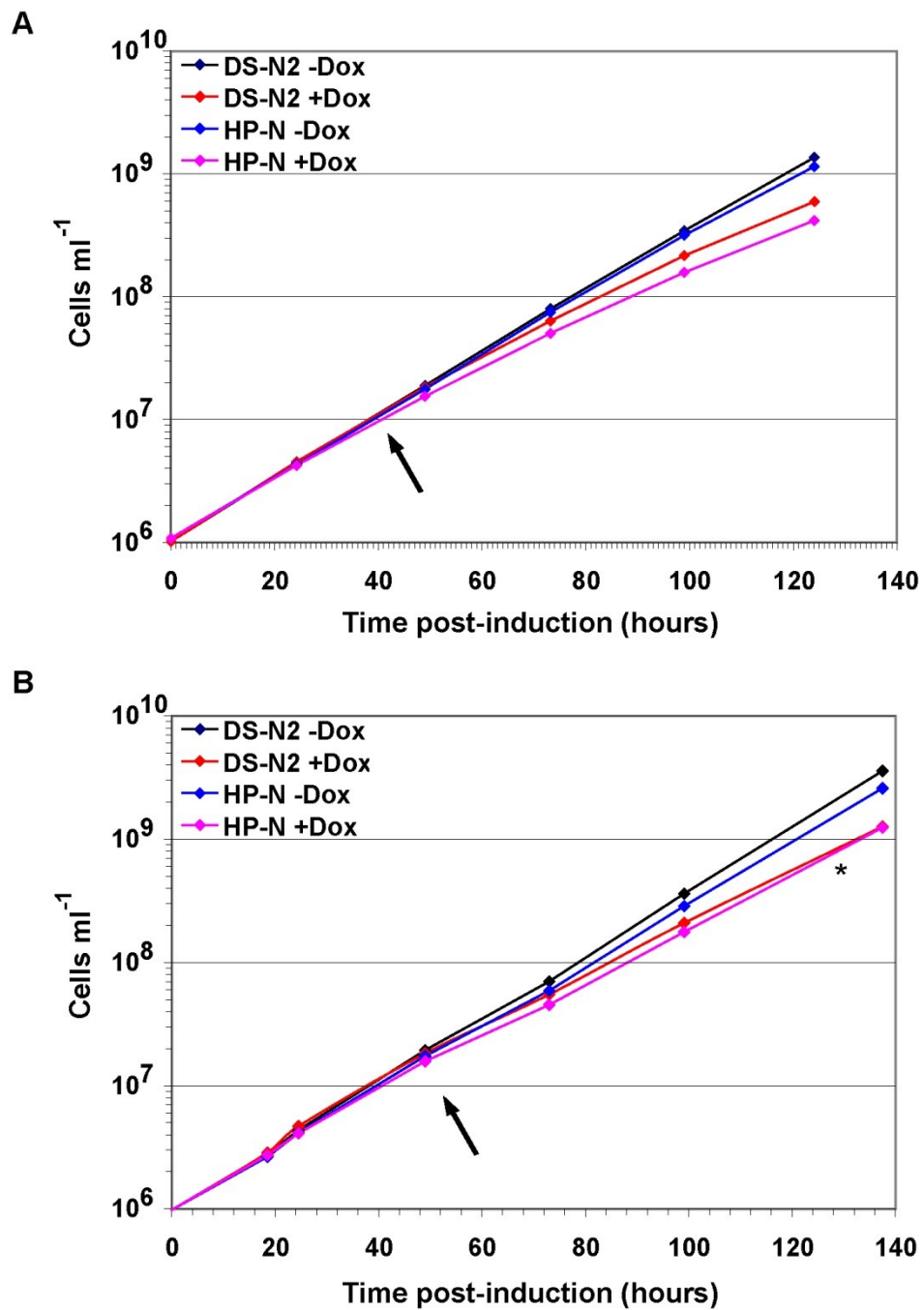


Figure 5.2. Group 4 RNAi cell lines induction. Cumulative growth curves.

A. DS-N2 and HP-N RNAi cell lines induction for 124h. Non-induced (-Dox) and induced cultures (+Dox) were reinoculated after 48h of induction at 1×10^6 cell.ml⁻¹.

B. DS-N2 and HP-N RNAi cell lines induction for 137h. Non-induced (-Dox) and induced cultures (+Dox) were reinoculated after 48h of induction at 1×10^6 cell.ml⁻¹.

Induced cell lines grew at a similar rate as the non-induced during the first 40-50h of induction (arrows). After this period the induced cultures grew slower than the non-induced ones. The growth rate of the HP-N RNAi cell line started to increase again after 120-125 hours of induction (B*), what was not observed with the DS-N2 RNAi cell line.

Figure 5.3.

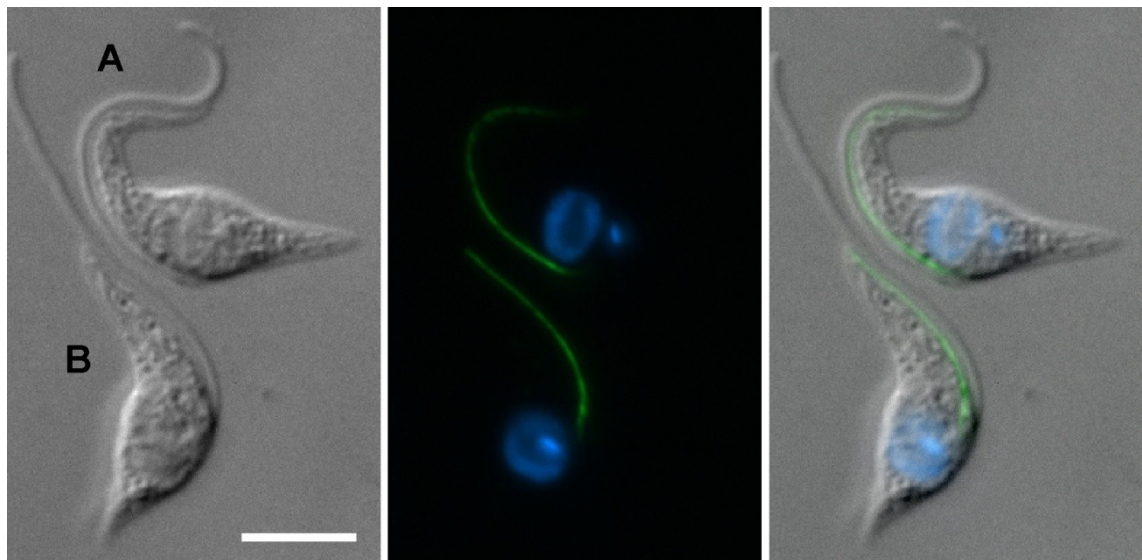


Figure 5.3. PCF DS-N2 RNAi cell line. Appearance of abnormal 1K1N cells after 18h of induction. The FAZ is stained with the L3B2 mAb, which labels the FAZ1 protein.

DIC, DAPI-L3B2 overlays and merged 100x images.

A. 1K1N cell with a reduced distance between the nucleus and kinetoplast, both positioned towards the middle of the cell-body instead of at its posterior end. The flagellar overhang is longer than in normal uninduced cells ($4.4\ \mu\text{m}$ as opposed to controls $2.63 \pm 0.14\ \mu\text{m}$ —mean $\pm$ SE).

B. 1K1N cell with the kinetoplast in a juxta-nuclear position and a very long flagellar overhang ($9\ \mu\text{m}$). Both kinetoplast and flagellar pocket are misplaced towards the anterior cell-end.

Scale bar represents $5\ \mu\text{m}$.

and remarkable reduction in cell size from 0 to 72h of induction (Fig. 5.4A,C). The reduction in cell size was maintained over time in the two induced cultures monitored for up to 124 and 168h, respectively (Fig. 5.4A, C). In contrast, the two inductions of the HP-N RNAi cell line produced different results, both from each other and from the DS-N2 RNAi cell line. The first HP-N RNAi induction (Fig. 5.4B) resulted in cell size reduction from 0 to 72h of doxycycline addition to the culture medium, although less pronounced than that observed in DS-N2 RNAi inductions (Fig. 5.4B) and which started to increase again after ~96h of induction (Fig. 5.4B). The second HP-N RNAi induction (Fig. 5.4D) showed a smaller change in cell size distribution than the first induction, with only slight decrease in cell size up to ~72h of doxycycline addition. After this time, the cell size distribution of induced cultures started to raise and, by 168h of induction, it was comparable to that of the not-induced controls (Fig. 5.4D). However, it must be point out that cell size measurements by aperture-impedance methods, as the one used by the CASY[®] cell counter, are poorly accurate. Consequently, the reduction in cell-size of DS-N2 and HP-N induced cells should be confirmed by performing experiments with more accurate cytometers, as those using light-scattering methods. Notwithstanding the low reliability of the cell-size data shown in Figure 5.4, the morphometric analysis of induced cells performed on microscopy images also suggest a reduction in cell size. Moreover, the morphometric analysis of DS-N2 induced cells, which is discussed in Section 5.3.3, showed that the decrease in cell size detected by the CASY[®] cell counter was due to a reduction in the cell-body length whereas the length of the flagellum remained relatively constant.

Figure 5.4.

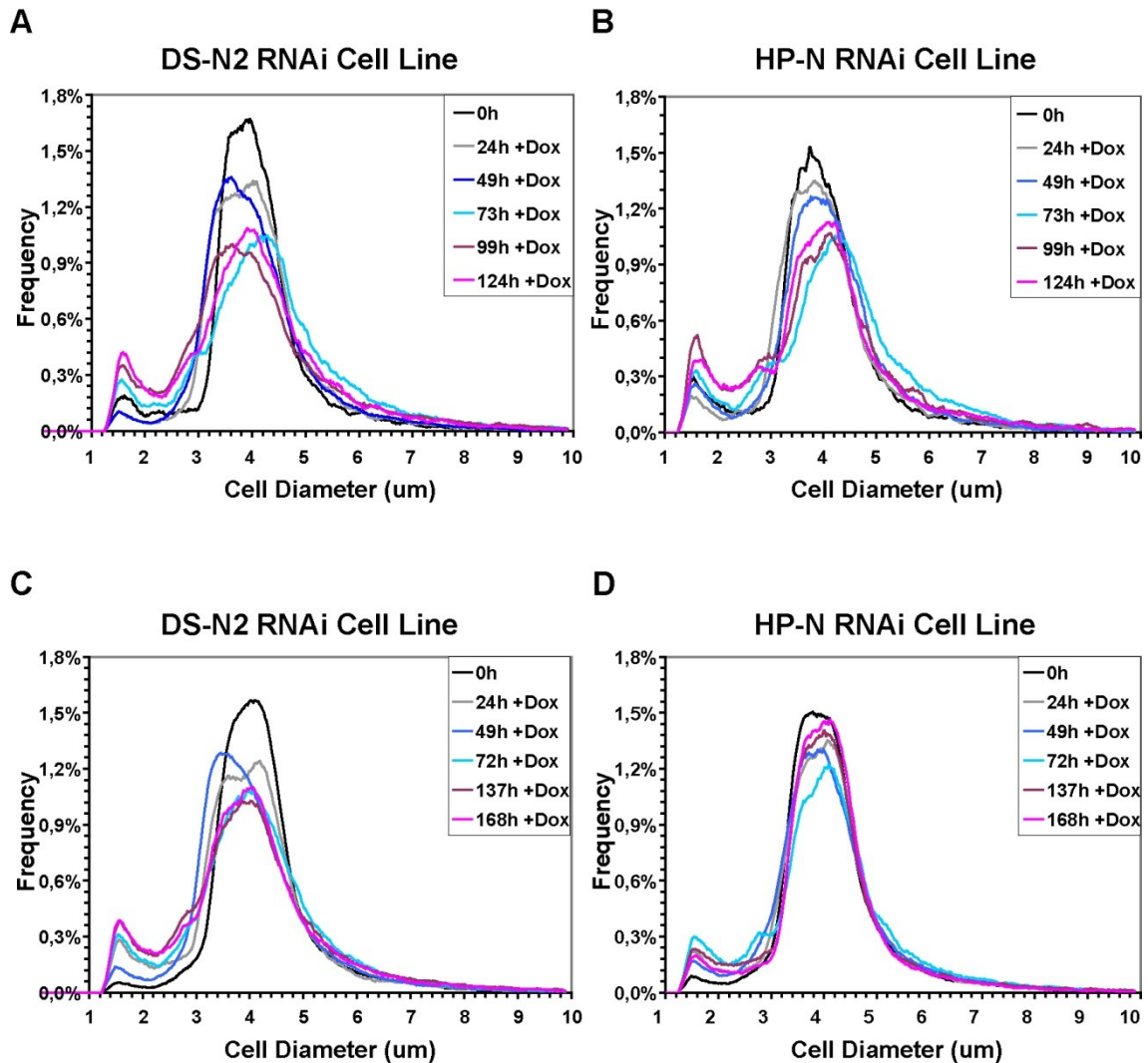


Figure 5.4. Cell size distribution in induced group 4 RNAi cell lines.

Inductions of the DS-N2 and HP-N RNAi cell lines were carried out simultaneously in two different occasions: **A and B** (first induction) and **C and D** (second induction). Cell size distribution was recorded for each different time point using a CASY[®] cell counter, which measures the cell diameter.

Reduction in cell size and appearance of cell debris was observed in both cell lines after 24h of induction. However, the DS-N2 RNAi induced cell line showed a more remarkable reduction in cell size than the HP-N RNAi induced cell line.

A and C. DS-N2 RNAi cell line inductions, both showing a dramatic reduction in cell size from 0 to 72h of induction. The reduced cell size was maintained over time (A, 124h, and C, 168h of induction).

B and D. HP-N RNAi cell line inductions, both showing a less marked decrease in cell size than the DS-N2 cell line inductions. The second HP-N induction showed little change in cell size distribution after induction (**D**). In both HP-N RNAi inductions, the cell size started to increase again after 72h of induction and, by 164h post-induction (**D**), the cell size distribution of induced cultures was similar to those of the non-induced controls.

5.3.2. RNAi silencing of group 4 calpain-like proteins results in aberrant morphology and organelle mispositioning.

Counts of the number of nuclei and kinetoplast by light microscopy of DAPI-stained cells from simultaneous inductions of both RNAi cell lines, revealed that the HP-N RNAi cell line has a higher percentage of abnormal cells in the absence of doxycycline (10%, N=513, Fig. 5.5B) than the DS-N2 RNAi cell line (2%, N=507, Fig. 5.5A, detailed data on Appendix 5, Fig. A.5.1). The higher proportion of abnormal cells in the HP-N RNAi cell line was readily observable by direct comparison of conventional light microscopy images from both RNAi cell lines (HP-N, Fig. 5.6; DS-N2, Fig. 5.7); abnormal uninduced cells have the same aberrant morphology and organelle mispositioning than those found in the induced RNAi cell lines: reduced cell-body length, juxta-nuclear or anterior kinetoplast and a long flagellar overhang (Fig. 5.8). Thus, the HP-N RNAi cell line obtained by the combination of two RNAi approaches—inverted repeats of the target sequence flanked by two facing promoters—has a looser RNAi downregulation than the DS-N2 RNAi cell lines, obtained with a typical double promoter RNAi construct, where head-to-head promoters flank a single fragment of the target gene. The “leakiness” or looser RNAi downregulation of not-induced cultures has been observed more often in RNAi cell lines obtained with opposing promoters RNAi constructs and, as discussed earlier, is thought to be due to the increased difficulty of controlling two promoters instead of one (*Durand-Dubief et al., 2003*). The higher leakiness of the HP-N RNAi cell line could therefore be due to the combinations of two factors, the presence of two promoters and the production of a “higher” amount of dsRNA, as RNA molecules containing inverted repeats are self annealing.

Figure 5.5.

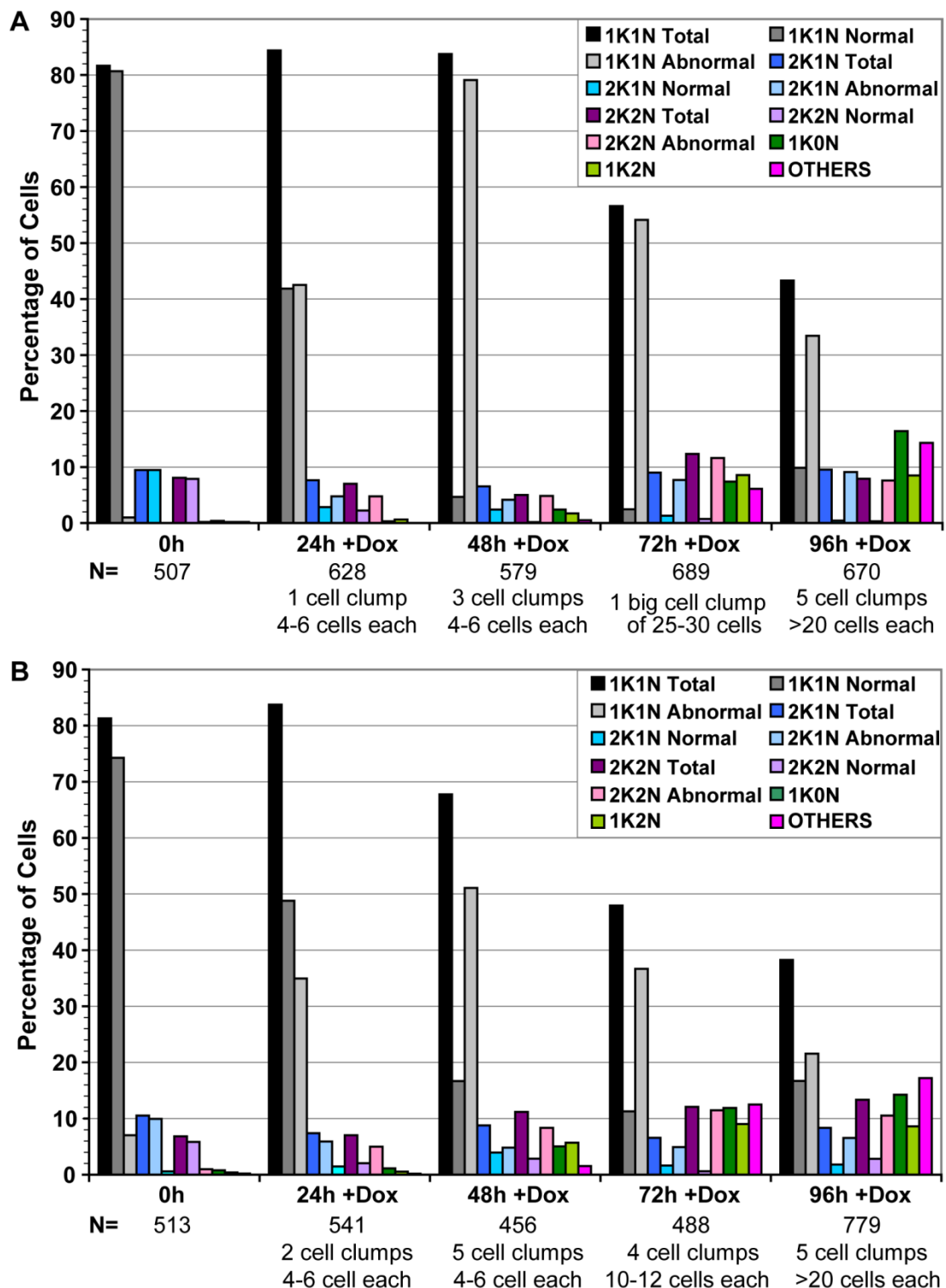


Figure 5.5. Simultaneous inductions of DS-N2 (A) and HP-N (B) RNAi cell lines. Percentage of cells in the different stages of the division cycle by DAPI staining of nuclei and kinetoplasts.

At 0h, the HP-N cell line has a higher percentage of abnormal cells than the DS-N2 (10% as opposed to 2%). Both induced cell lines showed a clear increase in the number of cells with aberrant organelle positioning, as a function of time. The percentage of abnormal cells in the HP-N RNAi cell line at 48h (77%), 72h (86%) and 96h (79%) of induction was lower than in the DS-N2 (48h-93%, 72h-96% and 96h-89%). Presence of cell clumps is indicated underneath each time-point. For detailed data see Appendix 5. Fig. A.5.1.

Figure 5.6.

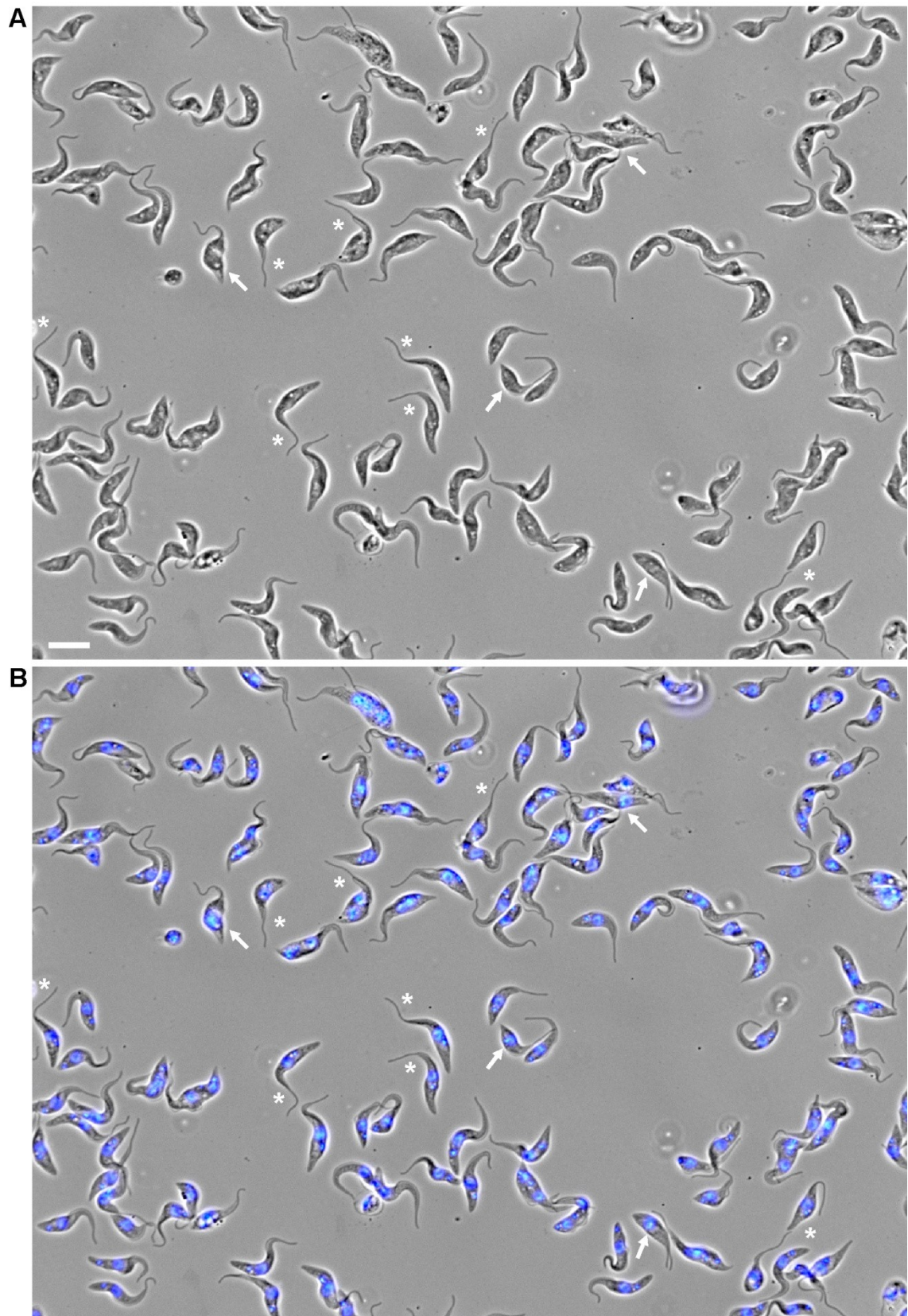


Figure 5.6. PCF HP-N RNAi cell line. 0h. Phase contrast (A) and phase/DAPI merged (B) 40x images showing the presence of abnormal cells in the non-induced state: cells with a long flagellar overhang (*) and mispositioning of the nucleus and/or kinetoplast (arrows). Scale bar represents 10 μm .

Figure 5.7.

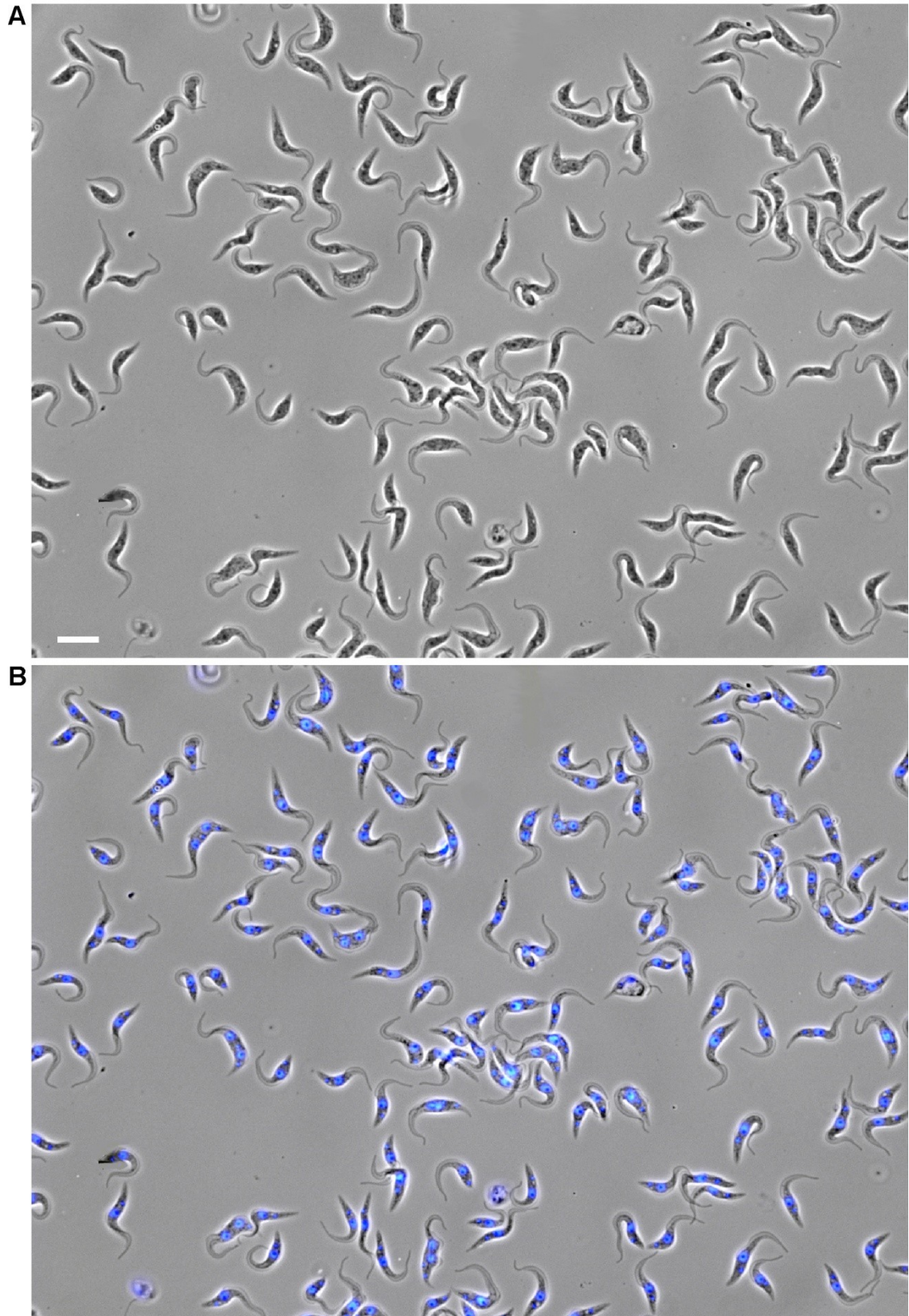


Figure 5.7. PCF DS-N2 RNAi cell line. 0h. Phase (A) and phase/DAPI merged (B) 40x images. In contrast to the HP-N-RNAi cell line, the DS-N2 non-induced population showed little presence of abnormal cells (2%). Scale bar represents 10 μ m.

Figure 5.8.

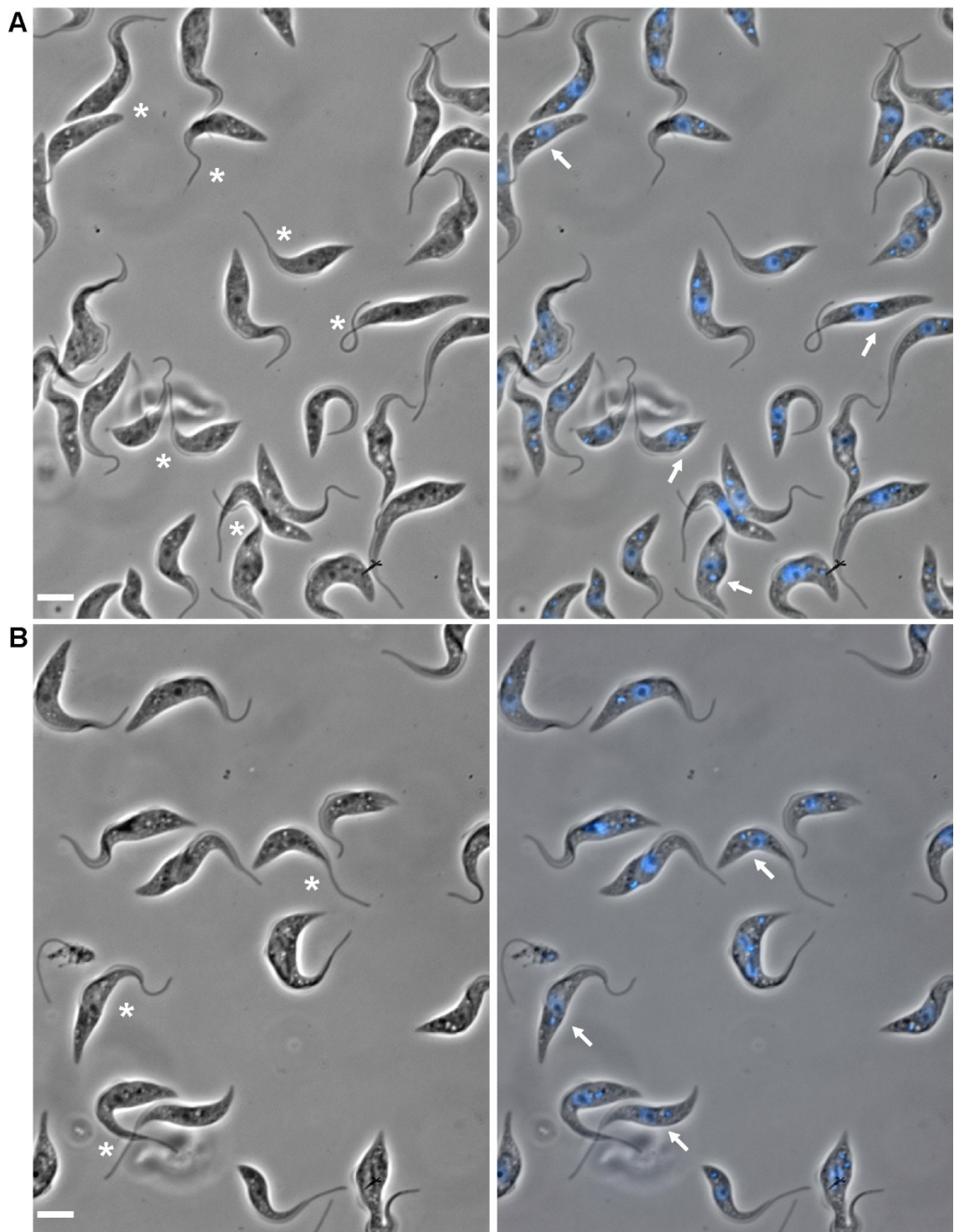


Figure 5.8. PCF HP-N RNAi cell line. 0h.

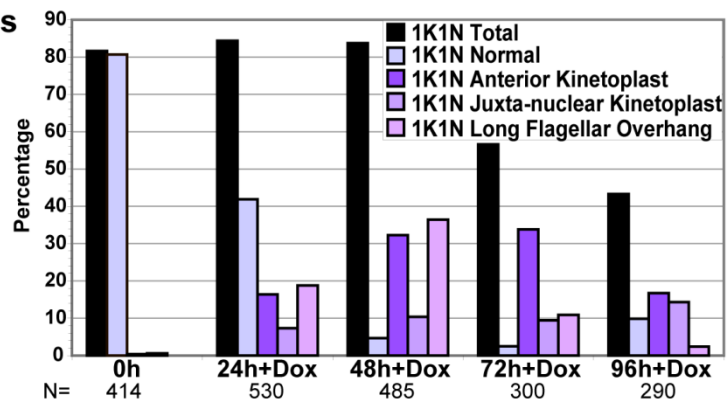
Phase and phase/DAPI merged 100x images of 0h HP-N RNAi cell line showing the presence of abnormal cells in the non-induced state: long flagellar overhang (*) and central position of nucleus and/or kinetoplast (arrows). Scale bars represent 5 μm.

The detailed analysis of DNA-stained induced cells from both RNAi cell lines (DS-N2 first and second inductions: Fig. 5.9 and 5.10; HP-N first induction: Fig. 5.11), showed increasing proportions of cells with aberrant organelle positioning. Most noticeably, there was an abrupt reduction in the number of normal 1K1N cells from 0h (~80%) to 48h of induction (2-5%, Fig. 5.9A, 5.10A and 5.11A), but also in the number of normal 2K1N (Fig. 5.9B, 5.10B and 5.11B)) and 2K2N cells (Fig. 5.9C, 5.10C and 5.11C). Additionally, there was a parallel increase in the number of abnormal 1K1N cells, which consisted mainly on epimastigote-like cells, with anterior or juxta-nuclear kinetoplasts, and on trypomastigote cells, with long flagellar overhangs (Fig. 5.9A 5.10A and 5.11A). There was also a marked increase in the number of abnormal 2K1N and 2K2N cells. Abnormal 2K1N cells have either one anterior and one juxta-nuclear kinetoplast or both kinetoplasts located in a juxta-nuclear position (Fig. 5.9B, 5.10B and 5.11B), whereas abnormal 2K2N cells have both kinetoplast lying between the segregated nuclei (Fig 5.9C, 5.10C and 5.11C). After 72h of induction, there was a dramatic reduction in the total number of 1K1N cells, coupled with a steep increase in the number of abnormal 2K2N cells, reflecting the reduction in cell growth rate. After 72h of induction there was also an increase in the number of anucleate zoids, their 1K2N siblings and multinucleate cells, indicating an increasing difficulty of induced cells to complete "correct" cytokinesis (Fig 5.9D, 5.10D and 5.11D).

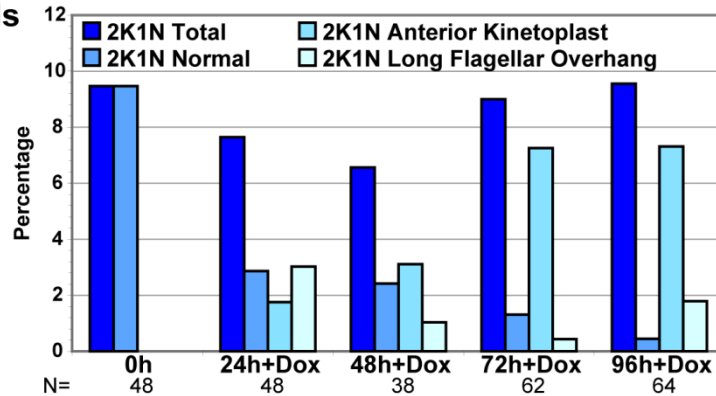
In addition, induced cultures showed the appearance of cell clumps that were resistant to shaking of the culture flasks. The number of cells present in these clumps increased as a function of time, from clumps containing 4-6 cells after 24h of induction to large clumps containing more than 20-30 cells by 96h post-induction, albeit the difficulty of asserting reliable numbers (Fig. 5.5). At later time points, cells clumps were visible to the naked eye. A possible explanation for the presence of these cell clumps could be that some induced cells that undergo cytokinesis are unable to complete abscission,

Figure 5.9.

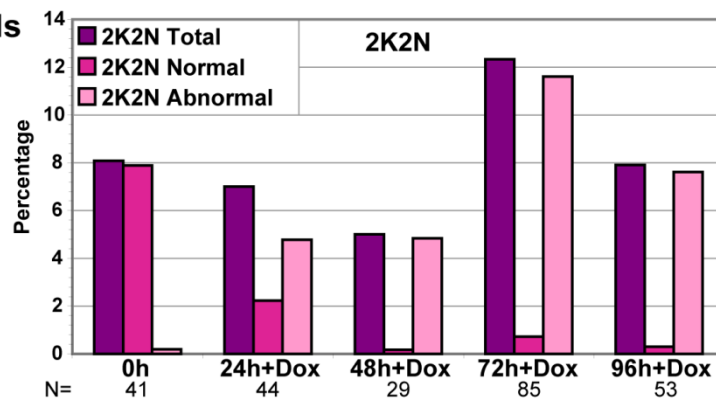
A- 1K1N cells



B- 2K1N cells



C- 2K2N cells



D- 1K0N Zoids, 1K2N cells and Others

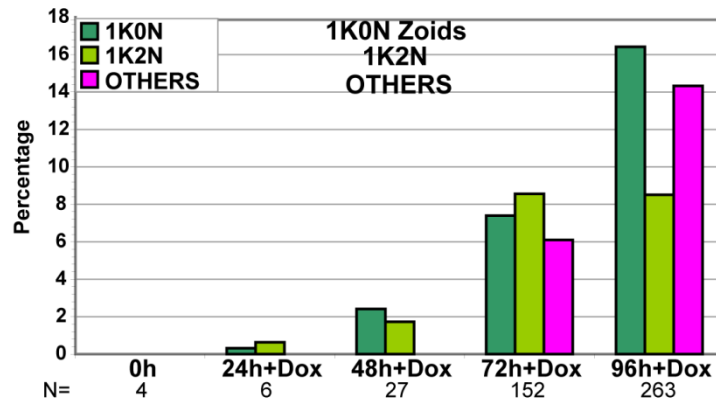
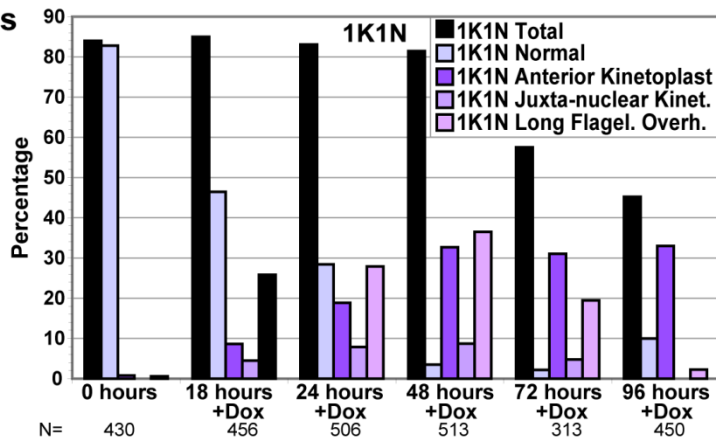


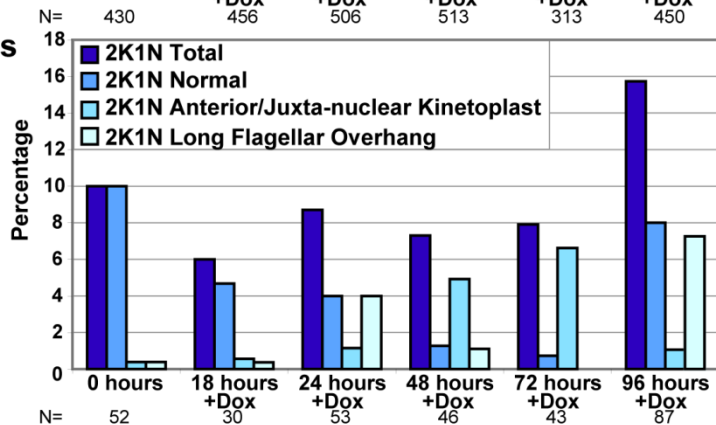
Figure 5.9. DS-N2-RNAi first induction showing a progressive increase in the number of abnormal cells after 18h of induction: 1K1N (A) and 2K1N (B) cells with anterior kinetoplasts and 2K2N cells (C) with both kinetoplasts positioned between the nuclei. Zoids, 1K2N and multinuclear cells (D) increased steeply after 72h of induction.

Figure 5.10.

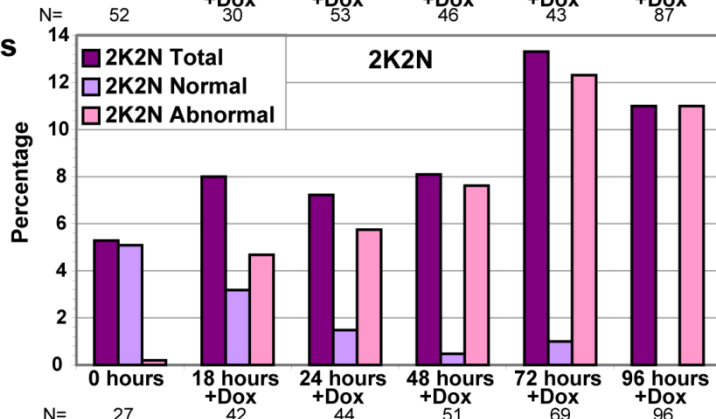
A- 1K1N cells



B- 2K1N cells



C- 2K2N cells



D- 1K0N Zoids, 1K2N cells and Others

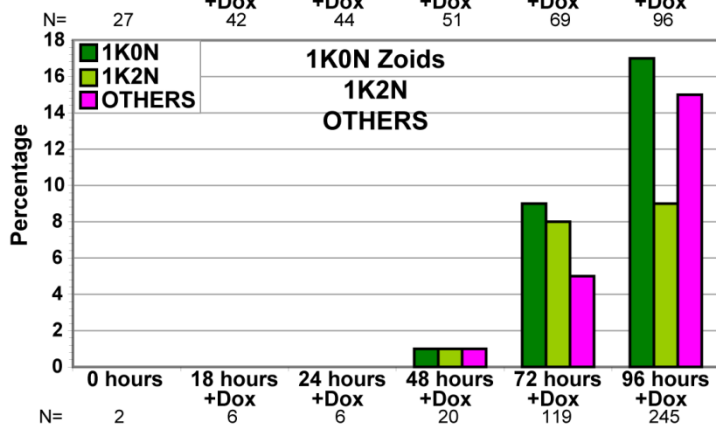
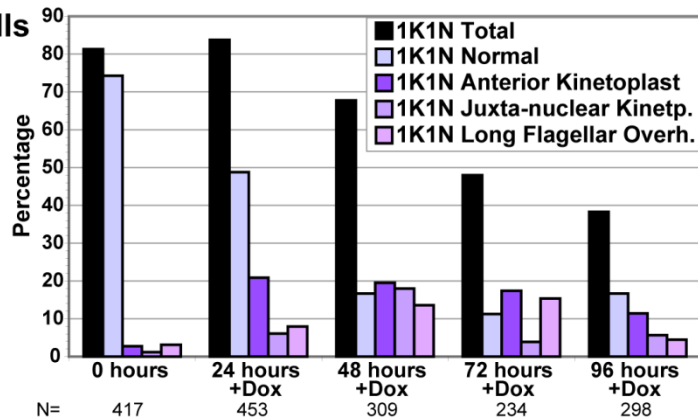


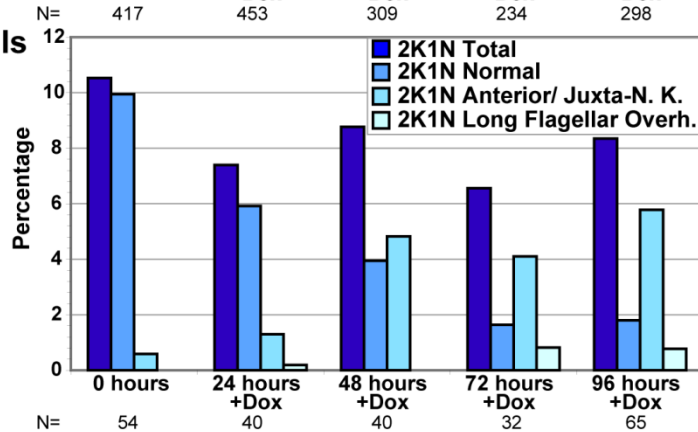
Figure 5.10. PCF DS-N2 RNAi cell line second induction. Note the appearance of abnormal cells after 18h of induction: 1K1N (A) and 2K1N cells (B) with anterior kinetoplasts and 2K2N cells (C) with both kinetoplasts positioned between the nuclei. After 72h of induction there was a steep increase in the number of zoids and multinucleate cells (D), reflecting the increasing difficulties of induced cells to complete correct cytokinesis.

Figure 5.11.

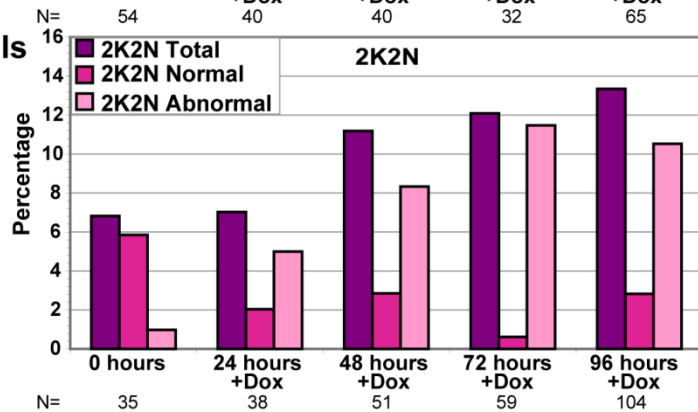
A- 1K1N cells



B- 2K1N cells



C- 2K2N cells



**D- 1K0N
Zoids, 1K2N
cells and
Others**

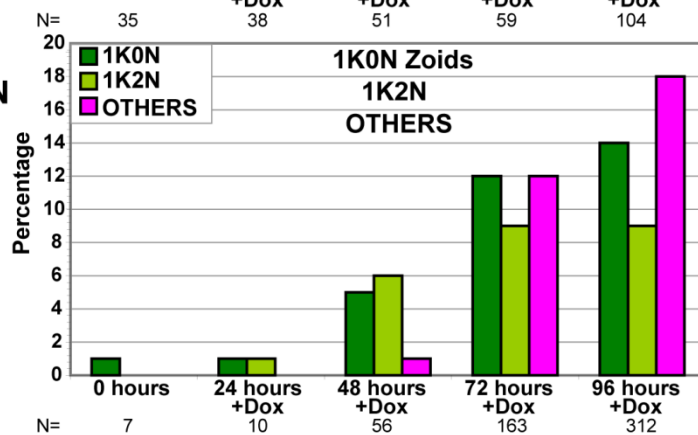


Figure 5.11. PCF HP-N RNAi induction.

Note the presence of abnormal cells in the non-induced stage (A-D). Induction of the HP-N RNAi cell line produced a similar phenotype to the DS-N2 cell line, with appearance of 1K1N and 2K1N cells with anterior or juxta-nuclear kinetoplasts (A and B) and 2K2N cells with the two kinetoplasts between the nuclei (C). Conversely, after 96h of induction the number of normal 2K2N cells started to increase again (C), in contrast to the DS-N2 RNAi cell line.

entering into the next round of cell division while still linked to their siblings. However, analysis of microscope slides containing DAPI-stained cells revealed that from 0 to 96h of induction cell clumps represent a small proportion of induced cells (Fig. 5.5). Again, whereas the analysis of the two DS-N2 RNAi cell line inductions showed very similar results (Fig. 5.9 and 5.10), the analysis of the HP-N RNAi cell line induction reflected both the higher leakiness of this RNAi cell line and its loss of RNAi phenotype after ~96h of induction, as evidenced by the increase in the number of normal 2K2N cells (Fig. 5.11C), which was not observed in induced DS-N2 RNAi cultures (Fig. 5.9C and 5.10C). Considering both the looser downregulation of the HP-N2 RNAi cell line and its gradual loss of the RNAi phenotype after ~96h of induction, I focused on the DS-N2 RNAi cell line to complete the functional analysis of group 4 calpain-like proteins in *T. brucei*.

5.3.3. Group 4 calpain-like proteins are essential for the correct positioning of the nucleus, kinetoplast and flagellar pocket in PCF trypanosomes.

Abnormal cells with a flagellar pocket located towards the centre of the cell-body (instead of at the posterior end) and an attached flagellum with a long overhang were observed in DS-N2 RNAi induced cultures as early as 18h post-induction by phase contrast or DIC light microscopy (Fig. 5.3) and in live cultures from 24h of induction onwards (Appendix 5, Video 5.1 and 5.2). Morphometric analysis of induced and uninduced cells was performed on phase/DIC microscopy images combined with fluorescence overlays of DAPI-stained cells (Fig. 5.3). The extension of the flagellum beyond the cell-body was also measured from scanning electron microscope images. Cell dimensions of not-induced 1K1N DS-N2 RNAi cells (N=11) were comparable to those of normal *T. brucei* procyclic trypomastigotes (Robinson *et al.*, 1995), with overage values (mean $\pm$ SE) of: 21.28 $\pm$ 0.36 μ m total cell length, 18.66 $\pm$ 0.3 μ m cell-

body length, $20.36 \pm 0.23 \mu\text{m}$ flagellar length and $2.63 \pm 0.14 \mu\text{m}$ flagellar overhang length (Fig. 5.12A). The flagellar-overhang or extension of the flagellum beyond the cell-body constituted $\sim 12.3\%$ of the total cell length (9.4-15.5% range, Fig. 5.12A). The distances from the kinetoplast and nucleus to the posterior cell-end were relatively constant in measured cells, with $3.86 \pm 0.15 \mu\text{m}$ and $7.8 \pm 0.15 \mu\text{m}$ overage values, respectively (Fig. 5.13A). Additionally, the distances between the kinetoplast and the nucleus were also remarkably invariable in uninduced cells, with overage values of $3.62 \pm 0.07 \mu\text{m}$ (Fig. 5.13A).

In contrast, DS-N2 RNAi induced cells showed progressive morphological changes from 18 to 96h of induction, which consisted on a decrease in cell-body length and an increase in the length of the flagellar overhang, but not in the total length of the flagellum, which remain relatively constant (Fig. 5.12A). The decrease in cell-body length occurred gradually from 0h ($18.66 \pm 0.3 \mu\text{m}$) to 48h of-induction ($13.63 \pm 0.64 \mu\text{m}$), when it reached values similar to those of salivary glands epimastigotes ($13.15 \pm 0.59 \mu\text{m}$) (*Sharma et al., 2008*). However, the extension of the flagellum beyond the cell-body increased rapidly in induced cells and, by 24h of induction, its overage length was almost five-folds the uninduced value, constituting $\sim 45\text{-}50\%$ of the total cell length (Fig. 5.12A and 5.13B). The increase in the length of the flagellar overhang was not only caused by the reduction in cell-body length but also by the repositioning of the kinetoplast — and thus of the flagellar pocket — from the posterior end of the cell to the mid-cell-body (Fig. 5.12 and 5.13). In induced cells, the distance between the kinetoplast and the posterior cell-end increased considerably after 18h of induction (Fig. 5.13A), after which time-point it started to decrease slightly but consistently until 96h of induction (Fig. 5.13A-B). However, when the reduction in cell-body length of induced cells was taken into account, the distance between the kinetoplast and the posterior cell-end constituted an increasing percentage of the

Figure 5.12.

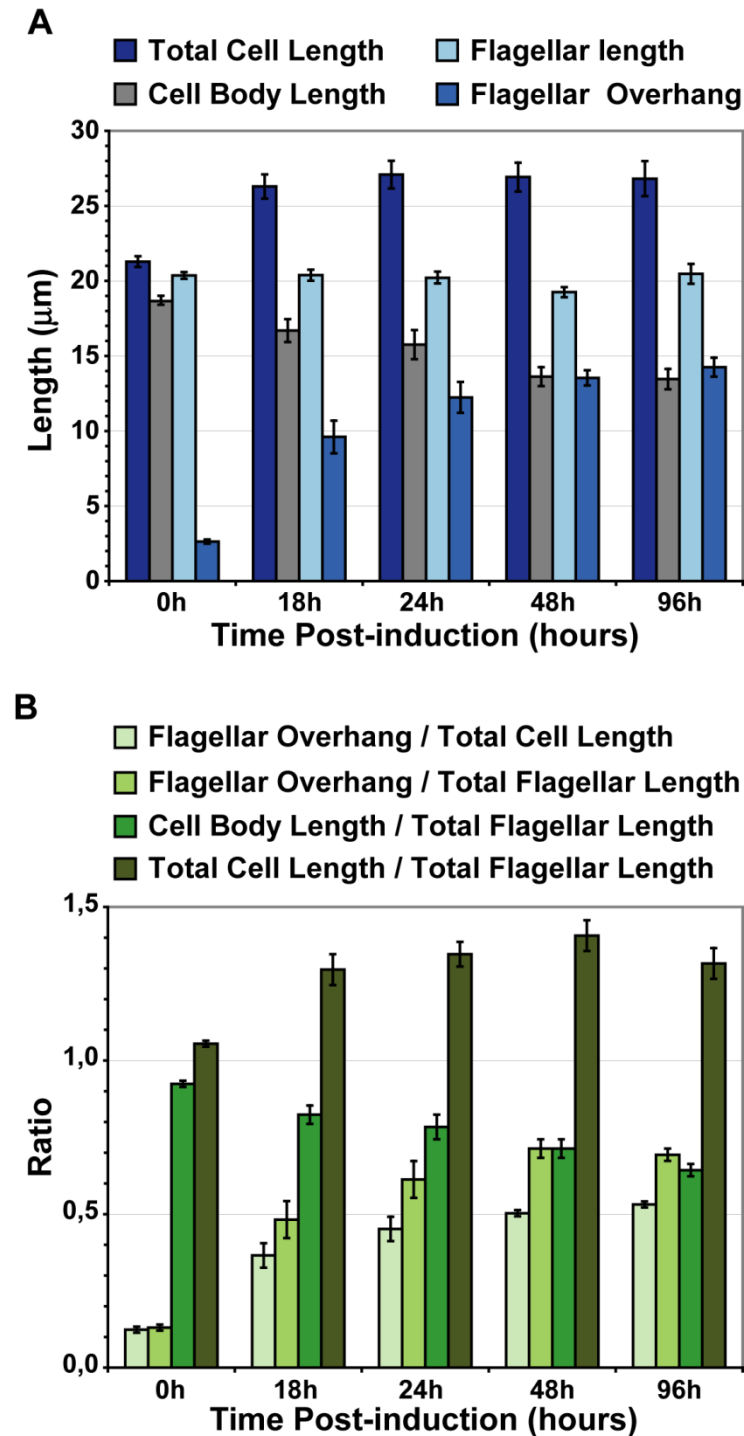


Figure 5.12. Morphometric analysis of DS-N2 RNAi cells (1). Cell dimensions were derived from images of DAPI stained cells combined with phase contrast /DIC overlays (N=11).

A. Increase of the total cell length after 18h of induction was coupled to both, a decrease in the cell-body length and an increase in the length of the flagellar overhang. In contrast, the flagellar length remained relatively constant after induction.

B. Ratios between different cell-body markers. Note the increase in the length of the flagellar overhang in relation to the total cell length and to the flagellar length. Conversely, the ratio between the cell-body length and the flagellar length decreased gradually after induction. In addition, the ratio between the total cell length and the flagellar length increased markedly after 18h of induction and then moderately until 48h after induction.

Figure 5.13.

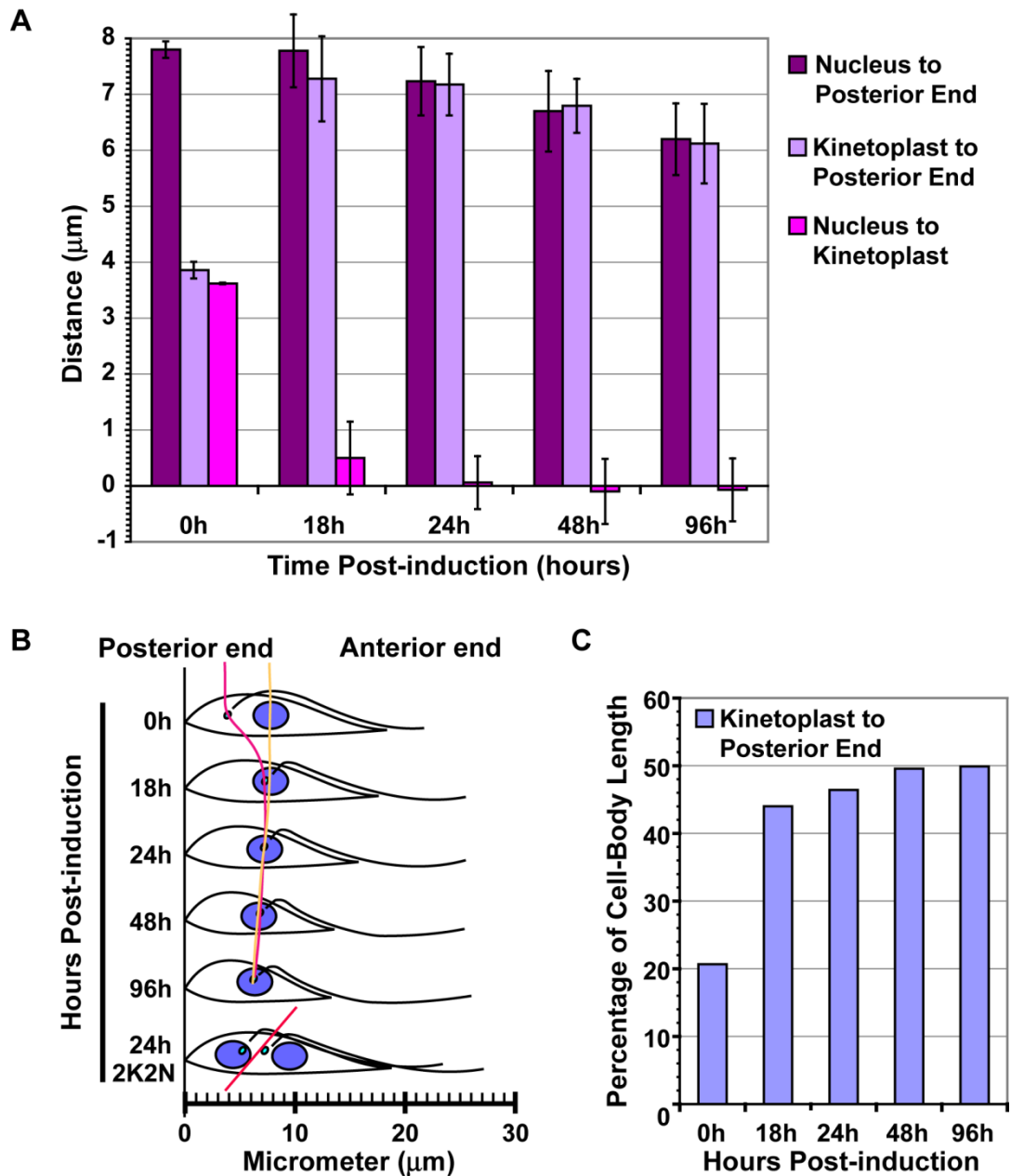


Figure 5.13. Morphometric analysis of DS-N2 RNAi cells (2).

A. Overage distances (mean $\pm$ SE in μm) from the nucleus (N) to the posterior cell-end decreased gradually after induction. Distances between the kinetoplast (K) and the posterior cell-end increased drastically after 18h of induction and then started to decrease slightly from 24 to 96h of induction. Overage distances between N and K decreased dramatically after 18h of induction and then continued to decrease slightly until 48h of induction.

B. Diagram representing the changes in cell dimensions and location of organelles of 1K1N cells from 0 to 96h of induction. Note the organelle disposition of representative induced 2K2N cells, where the most posterior K is located in a position anterior to the most posterior N, whereas the most anterior K is posterior to the most anterior N, thus giving rise to a 1K1N epimastigote-like daughter cell containing the new flagellum and a relatively "normal" 1K1N daughter cell containing the old flagellum. Therefore, organelle distances after induction reflect the overage values of these two kinds of 1K1N cells.

C. Increase, after induction, in the distance from the kinetoplast to the posterior cell-end as a percentage of the cell-body length.

cell-body length from 0 to 96h of induction: ~20% in uninduced cells to ~50% in 96h induced cells (Fig. 5.13C). Conversely, the absolute distance between the nucleus and the posterior pole of the cell decreased gradually from 0 to 96h of induction (Fig. 5.13A), but it constituted a relatively constant percentage of the cell-body length (40-45%, Appendix 5, Fig A.5.2) The distance between the kinetoplast and the nucleus decreased dramatically after 18h of induction and, by 48h, it reached a negative value, reflecting the exchange in the relative positions of the nucleus and kinetoplast in induced cells (Fig. 5.13A-B). Interestingly, the reduction in cell-body length was not proportional to the increase in the distance between the kinetoplast and the posterior-end, which, comparatively, augmented at a much higher rate (Fig. 5.12A, 5.13A, C).

Thus, many RNAi induced DS-N2 cells acquired an epimastigote-like morphology, although with very significant differences from the morphological events that characterise the differentiation of trypomastigotes into epimastigotes in the tsetse fly proventriculus. In the tsetse fly, the exchange in relative positions of the nucleus and kinetoplast on the anterior-posterior axis of the cell is primarily a result of a reduction in the distance between the nucleus and the posterior end of the cell— the distance between the kinetoplast and the posterior cell-end remaining relatively constant—. The nucleus and kinetoplast exchange in relative positions occurs during mitosis, producing a 2K2N epimastigote cell where one kinetoplast is positioned between the dividing nuclei and the other is located in a position anterior to the most anterior nucleus The 2K2N epimastigote cell then undergoes asymmetric division, giving rise to one short and one long epimastigote cell. However, in RNAi induced DS-N2 cells the exchange in the nucleus and kinetoplast relative positions, which also occurred during mitosis, was primarily caused by an increase in the distance between the kinetoplast and the posterior end of the cell. The resulting 2K2N cells have the two kinetoplasts located between the segregating nuclei. Moreover, the kinetoplast connected to the new

flagellum is located in a position anterior to the most posterior nucleus and the kinetoplast associated to the old flagellum is located in a position posterior to the most anterior nucleus (Fig. 5.13B). Cytokinesis of these abnormal 2K2N cells gives rise to an epimastigote-like cell and a “trypomastigote” cell with a reduced distance between nucleus and kinetoplast and a long flagellar overhang. The epimastigote-like daughter cell is ~6 μm shorter than its trypomastigote sibling (Fig. 5.12A).

Further differences between tsetse fly differentiating trypomastigotes and epimastigotes and RNAi induced DS-N2 cells concern the length of the flagellar overhang and the total flagellar length. During the differentiation of tsetse fly trypomastigotes the length of the flagellar overhang remains relatively unchanged and, independently of the total cell length, constitutes ~15-20% of the total flagellar length (*Sharma et al., 2008*). In contrast, both “trypomastigotes” and epimastigote-like RNAi DS-N2 induced cells have an unusually long flagellar overhang, which by 48h of induction constituted ~70% of the total flagellar length (Fig. 5.12B, 5.13B). Besides, in differentiating trypomastigotes the old flagellum of 2K2N cells elongates considerably, reaching ~35 μm in length, whereas the new flagellum is very short (*Sharma et al., 2008*). However, the flagellar length of RNAi induced DS-N2 cells remains relatively invariable after induction ($19.3 \pm 20.3 \mu\text{m}$, Fig. 5.12A). Moreover, in 2K2N DS-N2 induced cells at the stage prior to completion of cytokinesis, the length of the new flagellum was only slightly shorter than the length of old flagellum, similarly to normal procyclic trypomastigotes (*Sherwin and Gull, 1989a*).

The increase in cell density of induced cultures (Fig. 5.2), albeit at a lower rate than in the not-induced, and the rise in the percentage of abnormal cells (Fig. 5.5) suggest that the abnormal 1K1N cells of induced cultures are able to enter into further rounds of cell division and complete cytokinesis. Indeed, 2K2N cells in the last stages of cytokinesis, giving rise to an epimastigote-like cell and an abnormal trypomastigote cell, were

readily observed by DAPI/Phase contrast or DAPI/DIC light microscopy (Fig. 5.14E) and in life cultures, at every time-point after induction (Appendix 5, Video 3). The detailed analysis of DAPI-stained DS-N2 RNAi induced cells, used to determine the cell dimensions presented above, suggests a sequential ordering of the events occurring after induction. This hypothetical ordering is based on the progression of morphological changes (Fig. 5.12 and 5.13) and different DNA staining patterns (Fig. 5.9-5.11) observed upon induction, but also on the comparison of these events to those observed in tsetse fly trypomastigotes during their differentiation into epimastigotes (*Sharma et al., 2008*). The ordering of events suggested here includes the following five points:

1. Reduction in the distance between nucleus and kinetoplast (Fig. 5.14A).
2. Initiation of nuclear mitosis and kinetoplast segregation (Fig. 5.14B).
3. Progression of the segregating kinetoplasts towards the anterior cell end (one kinetoplast posterior and one juxtaposed to the mitotic nucleus, Fig. 5.14C).
4. Completion of nuclear mitosis (the two kinetoplasts located between the two daughter nuclei) and initiation of cleavage furrow ingression between the two kinetoplasts, towards the posterior end, along the longitudinal axis of the cell (Fig. 5.14D).
5. Completion of cytokinesis (Fig. 5.14E, two daughter cells only connected at their posterior ends), giving rise to an abnormal trypomastigote cell —with reduced distance between the nucleus and the kinetoplast and a long flagellar overhang (Fig. 5.14F) — and an epimastigote-like cell, which contains the new flagellum —with the kinetoplast in an anterior or juxta-nuclear position and a long flagellar overhang (Fig. 5.14G).

Figure 5.14.

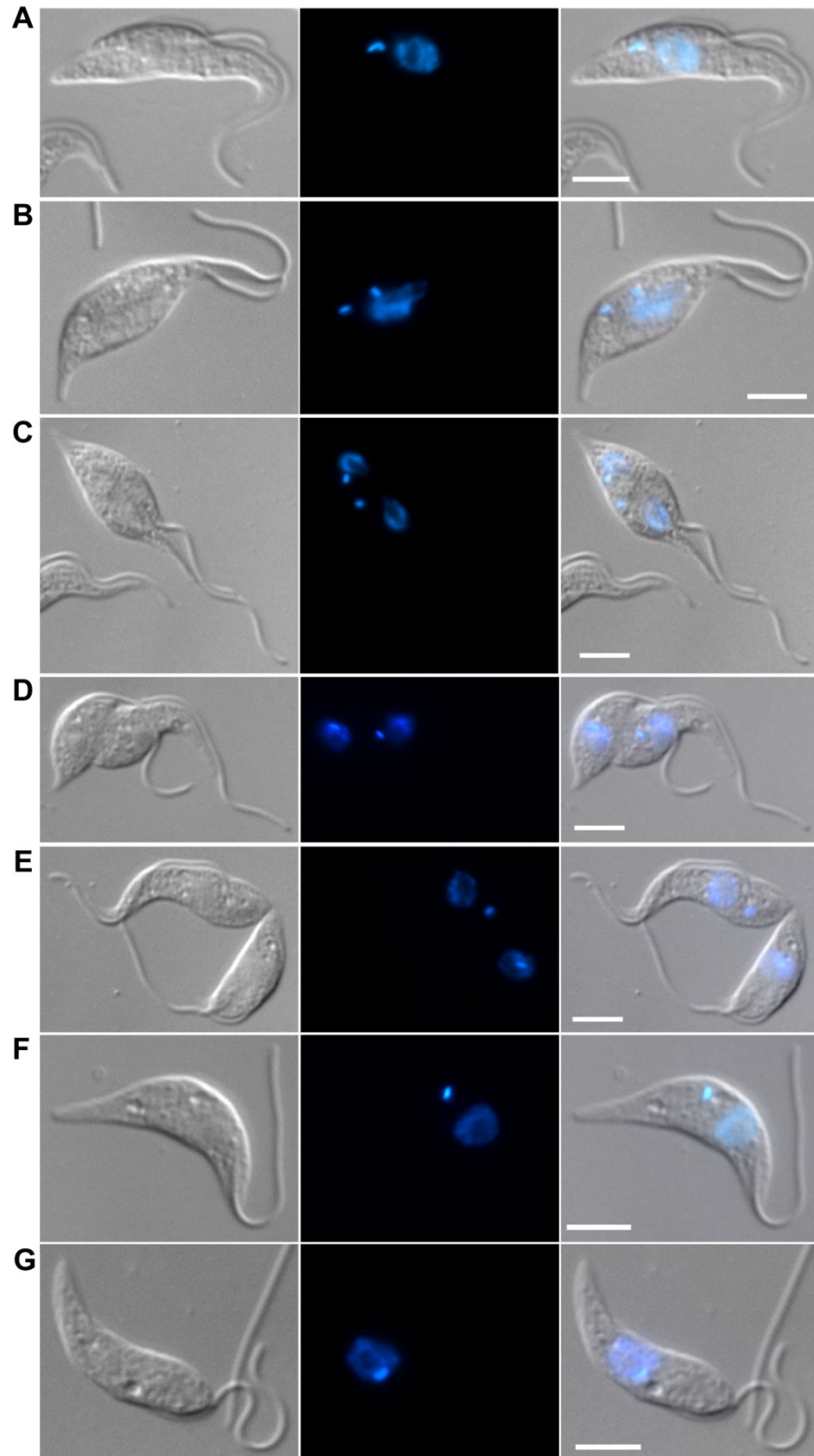


Figure 5.14. Microscopic analysis of PCF DS-N2-RNAi cell line. 24h post-induction. DIC, DAPI and merged 100x images.. **A.** 1K1N cell with reduced distance between the nucleus (N) and the kinetoplast (K). **B.** 2K1N cell with a long flagellar (Fr) overhang and a reduced distance between the posterior K and the N. **C-D.** 2K2N cells with the two Ks between the segregating nuclei. **E.** 2K2N cell in the last stage of anisomorphic division. **F.** 1K1N cell with the K and the N misplaced towards the anterior cell-end and a long Fr overhang. **G.** 1K1N cell with anterior K and FP and a long Fr overhang (epimastigote-like cell). Scale bars represent 2 μ m.

5.3.4. Group 4 calpain-like proteins have an essential role in cell morphogenesis.

The aberrations in the positioning of the kinetoplast and flagellar pocket observed upon RNAi induction of *T. brucei* group 4 calpain-like proteins, one of which has been shown to localise at the FAZ filament (*Muller et al., 1992*), indicate that these proteins are implicated in more complex roles than the mere attachment of the flagellum to the cell-body, being essential for the correct positioning of these important organelles. RNAi knockdown of other *T. brucei* FAZ proteins, as FLA1 (flagellum-adhesion glycoprotein 1) and FAZ1 (a FAZ repeat-containing protein), have been reported to produce phenotypes of flagellar detachment, aberrant morphology, mitosis and/or cytokinesis inhibition, with the consequent appearance of zoids, their 1K2N siblings and multinuclear cells (*LaCount et al., 2002; Vaughan et al., 2008*). However, the phenotype caused by the silencing of group 4 calpain-like proteins is not that of flagellar detachment, as in all analysed RNAi induced cells the portion of the flagellum that comprises the area between the flagellar pocket and the anterior cell-end seemed attached to the cell-body. Group 4 calpain-like proteins RNAi phenotype consists mainly in the appearance of epimastigote-like morphology and long flagellar overhangs, through a reduction of the cell-body length, shortening of the FAZ along the longitudinal cell axis and a mispositioning of the kinetoplast, basal bodies and flagellar-pocket, which are all placed in a more anterior position than in wild type PCF trypomastigotes (Fig. 5.12 and 5.13).

Therefore, group 4 calpain-like proteins must be essential for the correct morphogenesis of PCF trypomastigotes and it would not be unreasonable to speculate that, at least some of these proteins are down regulated during the differentiation of procyclic trypomastigotes into epimastigotes in the tsetse fly. As discussed above, the morphological changes detected upon induction were sequentially ordered by the

combined analysis of DIC/phase contrast images and DNA staining patterns. After 18-24h of induction, 1K1N cells with a reasonably normal external morphology were found to have a reduced distance between the nucleus and the kinetoplast (from $\sim 3.6 \mu\text{m}$ at 0h to $0.5 \mu\text{m}$ at 18h of induction, Fig 5.13A). Additionally, 18h induced 1K1N cells have a longer flagellar overhang than not-induced 1K1N cells (from $-2.6 \mu\text{m}$ at 0h to $9.6 \mu\text{m}$ at 18h of induction, Fig 5.12A-B). Also, 18h induced 2K1N cells often have either one anterior and one posterior kinetoplast, or the two kinetoplast located in a juxta-nuclear position, whereas 2K2N cells have the two kinetoplast positioned between the segregating nuclei (Fig. 5.14C-D).

In the majority of dividing cells observed after induction, the new flagellum seemed to remain linked to the old flagellum by the flagella connector (Fig. 5.14C and E) although in some cases it was seen detached (Fig. 5.14D). Besides, in dividing cells with a cleavage furrow and in those cells in the last stages of cytokinesis, the position of the cleavage plane indicates that the posterior cell product of division, containing the new flagellum, would be slightly shorter than its sibling ($\sim 6 \mu\text{m}$ shorter, Fig. 5.12), would have a longer flagellar overhang, (Fig. 5.14C-E) and the kinetoplast in a more anterior position than the anterior product of cell division —juxtaposed or anterior to the nucleus— acquiring an epimastigote-like morphology (Fig 5.14E and G). In addition, the anterior product of division, containing the old flagellum, often has a longer flagellar overhang and a shorter distance between the nucleus and kinetoplast than uninduced cells, with both nucleus and kinetoplast frequently placed at the mid cell-body (Fig. 5.14F).

From 48h of induction onwards, after the completion of more than two cell division-cycles, the vast majority of uniflagellate cells ($\sim 95\%$) presented abnormal morphologies. The most abundant abnormal cell types were the two likely products of abnormal 2K2N cells cytokinesis: abnormal trypomastigote cells ($\sim 37\%$, Fig. 5.5, 5.9A

and 5.10A), with a reduced distance between the nucleus and kinetoplast (from 0 to 2.3 μm , Fig. 5.13A) and a long flagellar overhang (from ~ 11 to 15 μm , Fig. 5.12A) and epimastigote-like cells ($\sim 33\%$, Fig. 5.5, 5.9A and 5.10A) with the kinetoplast and flagellar pocket juxtaposed or anterior to the nucleus (from -0.4 to -1.8 μm , Fig. 5.13A) and an even longer flagellar overhang (from ~ 12.7 to 16.8 μm , Fig. 5.12A, 5.15A-B, 5.16A and 5.17A-B). Additionally, $\sim 95\%$ of 48 and 72h induced 2K2N cells have misplaced kinetoplast (Fig. 5.5, 5.9C and 5.10C), generally located between the segregating nuclei (Fig. 5.15D-F, 5.16B and D and Fig. 5.17D) and occasionally at the anterior and posterior sides of the nuclei (Fig. 5.16C). Almost all observed 2K2N cells after 96h of induction showed kinetoplast mispositioning (99%, Fig 5.5, 5.9C, 5.10C, and 5.17D-E)

After 72h of induction there was also a marked increase in the number of anucleate zoids, which by 96h of induction constituted $\sim 17\%$ of the induced populations (Fig. 5.5, 5.9D, 5.10D, 5.16E and H and 5.17F-G). The rise in the number of zoids was initially accompanied by a proportional increase in the number of their 1K2N siblings (72h, 1K0N $\sim 9\%$, 1K2N $\sim 8\%$) and a lower increase in the number of multinuclear cells ($\sim 5\%$, Fig. 5.5, 5.9D and 5.10D). However, by 96h of induction the number of 1K2N cells was approximately half the amount of zoids (1K1N $\sim 16\%$, 1K2N $\sim 9\%$), whereas the number of multinuclear cells increased to $\sim 15\%$ of the induced populations (Fig. 5.5, 5.9D, 5.10D, 5.16F-G and 5.17H), indicating the increasing difficulty of induced cells to complete correct cytokinesis.

Figure 5.15.

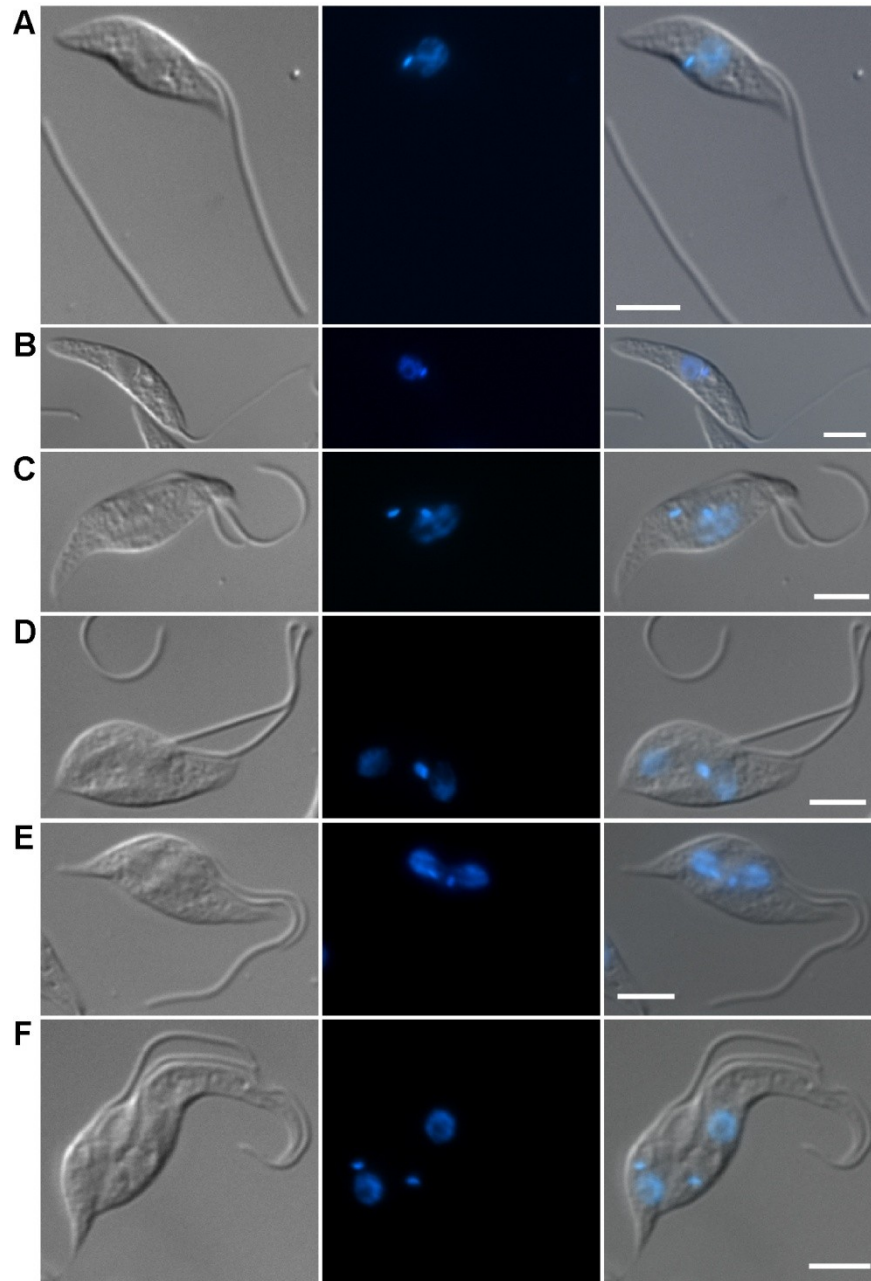


Figure 5.15. Microscopic analysis of DS-N2-RNAi cell line. 48h post-induction.

DIC, DAPI and merged 100x images.

A. 1K1N cell with a long flagellar overhang and the kinetoplast in a juxta-nuclear position.

B. 1K1N cell with an anterior kinetoplast and a long flagellar overhang.

C. 2K1N or 2K2N cell, with the kinetoplast and nucleus/nuclei are mispositioned towards the mid cell-body and with long flagellar overhangs.

D. 1K2N cell, which could be either the product of aberrant cytokinesis (2K2N giving rise to a 1K0N zoid and a 1K2N cell) or of a 1K1N cell undergoing division where the completion of nuclear mitosis precedes the kinetoplast segregation.

E-F. Abnormal 2K2N cells with misplaced kinetoplasts, located between the segregating nuclei.

Scale bars represent 2µm.

Figure 5.16.

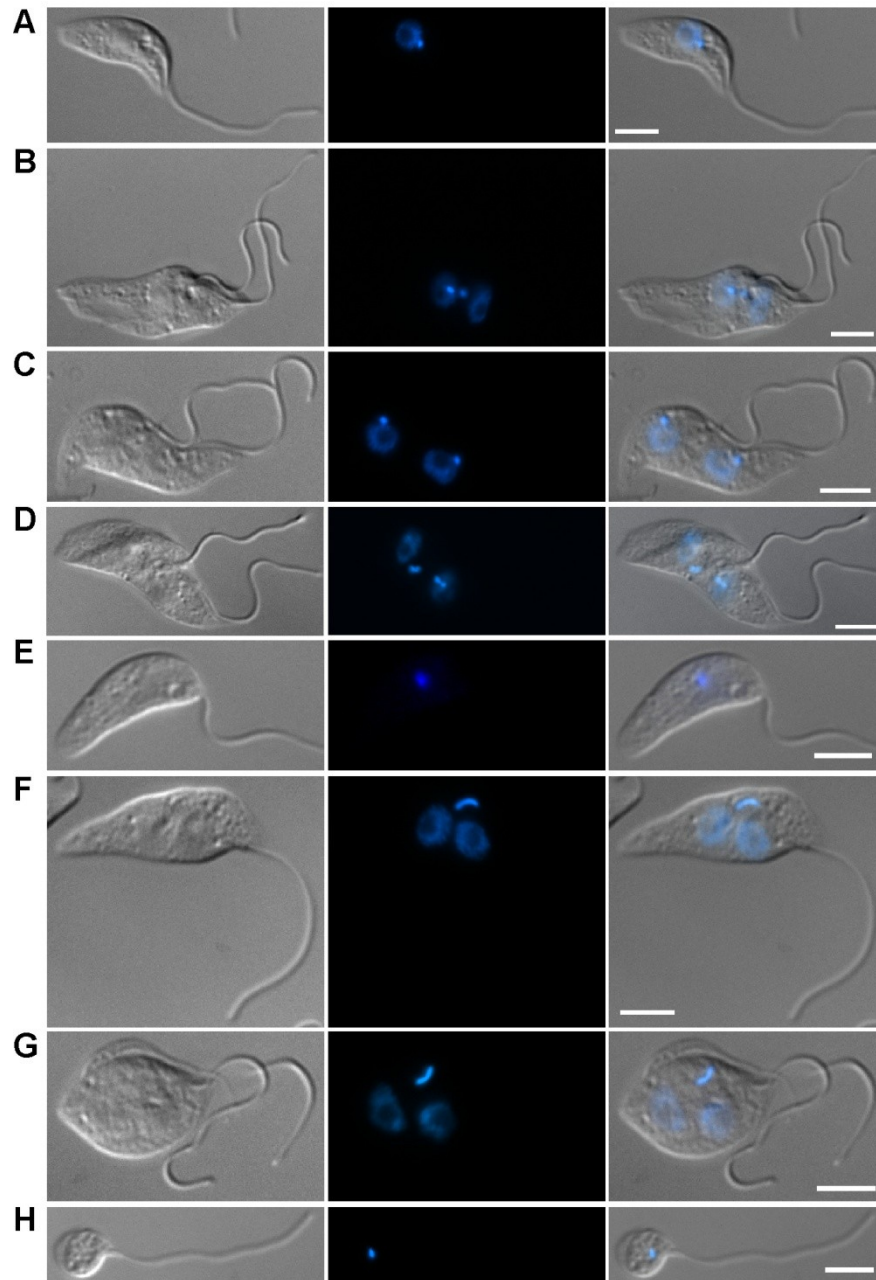


Figure 5.16. Microscopic analysis of PCF DS-N2 RNAi cell line, 72h post-induction.

DIC, DAPI and merged 100x images.

A. Abnormal 1K1N cell with anterior kinetoplast and flagellar pocket and a long flagellar overhang (epimastigote-like cell). **B.** Abnormal 2K2N cell with both kinetoplasts located between the segregating nuclei. **C.** Abnormal 2K2N cell where the kinetoplasts and nuclei have moved apart but none of the kinetoplasts is posterior to its corresponding nuclei. **D.** Abnormal mitotic 2K2N cell with aberrant positioning of the kinetoplasts at both sides of the forming cleavage furrow. **E.** Anucleate zoid (1K0N) with an anterior flagellar pocket. **F-G.** 1K2N cells with replicating kinetoplasts. Note also the appearance of cytoskeletal projections in G). **H.** 1K0N zoid with a short "cell" body. Scale bars represent 2µm.

Figure 5.17.

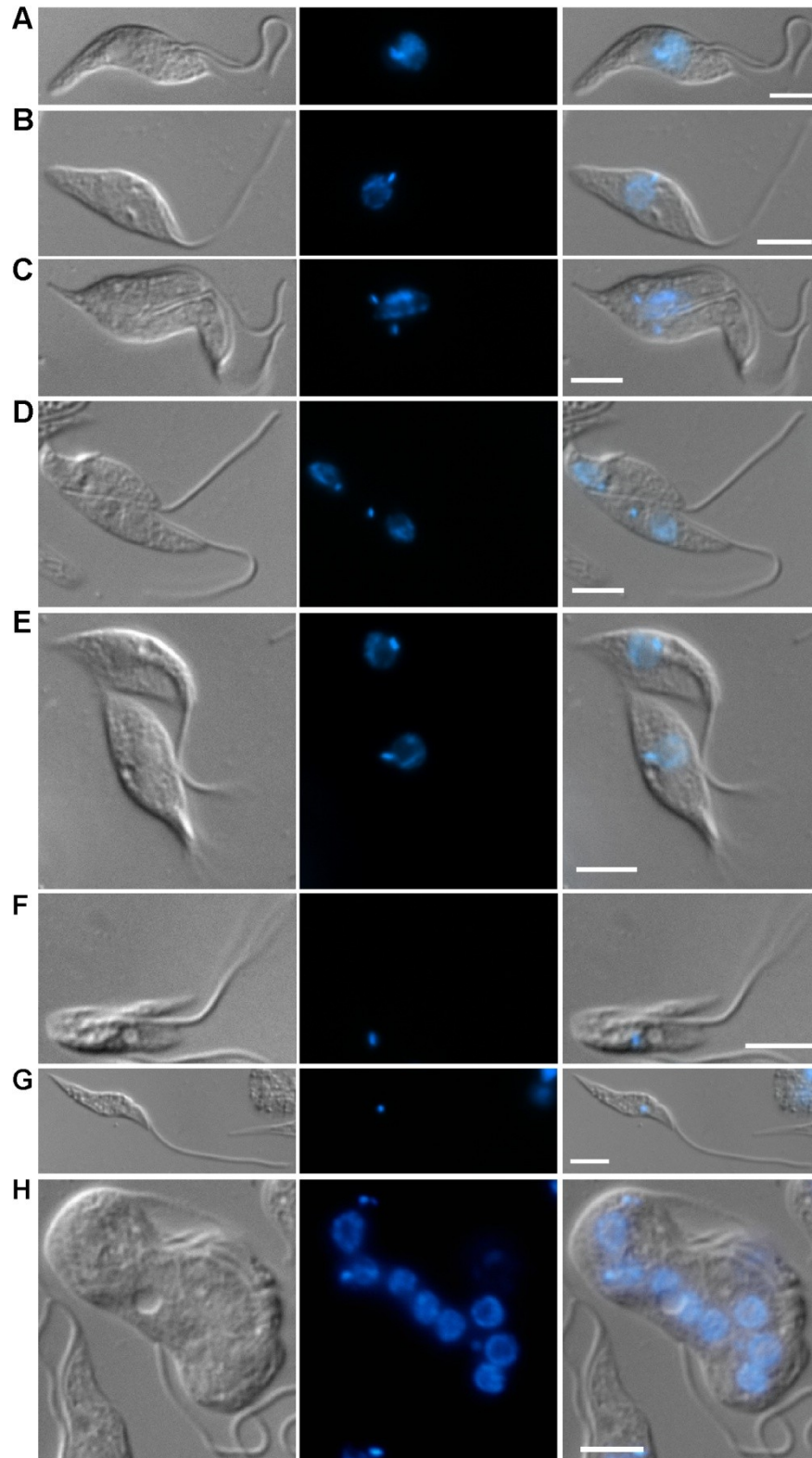


Figure 5.17. Microscopic analysis of PCF DS-N2-RNAi cell line. 96h post-induction. DIC, DAPI and merged 100x images. A-B. 1K1N cells with juxta-nuclear kinetoplast (K) and long flagellar overhangs. C. 2K1N cell with both Ks in a juxta-nuclear position. D-E. 2K2N cells in late stages of division, showing that they are able to complete cytokinesis, giving rise to an epimastigote-like cell (posterior product) and an abnormal trypomastigote cell, (anterior product, with a reduced distance between N and K. F-G. 1K0N zoids. H. Multinucleate cell. Scale bars represent 2µm.

5.3.5 RNAi knockdown of group 4 calpain-like sequences alters the structure of the FAZ.

As discussed earlier, some kinetoplastid group 4 calpain-like proteins have been shown to localise at the FAZ filament (*Muller et al., 1992; Cotrim et al., 1995*). To investigate the effect of these proteins RNAi knockdown on the FAZ filament structure and on the localisation of other cytoplasmic components of the FAZ in *T. brucei*, DS-N2 RNAi induced and uninduced cells were labelled with several mAbs that recognise the FAZ filament: L3B2, which labels the FAZ1 protein (*Kohl et al., 1999*) and two antibodies that stain the FAZ area but which antigens have not been yet identified: CD10 and DOT1 (*Woods et al., 1989*). L3B2 typically gives a linear staining of the FAZ filament, following the longitudinal axis of the cells, whereas CD10 and DOT1 give a punctuated labelling of the same structure.

Accordingly, not-induced DS-N2-RNAi cells labelled with L3B2 mAb showed the presence of a long FAZ filament that extended from the posterior end of the cell, close to the kinetoplast and immediately after the flagellar pocket, to the anterior cell-end, following the path of the attached flagellum (Fig. 5.18A). In DS-N2 RNAi induced cells that maintained a relatively normal trypomastigote morphology (albeit with a shorter cell-body and a longer flagellar overhang) the L3B2 mAb labelled a structure with the shape of a filament, although shorter than a normal FAZ, which also run from the flagellar pocket to the anterior end of the cells (Fig. 5.18B-D). In DS-N2 induced cells with an epimastigote-like morphology, which have the kinetoplast and flagellar pocket in anterior or juxta-nuclear positions, the L3B2 mAb labelled a very short structure that also extended from the flagellar pocket to the anterior cell-end (Fig. 5.19E and 5.19F). However, from 48h of induction onwards, epimastigote like-cells stained with L3B2 often showed a labelled structure not only shorter than a normal FAZ but also wider and more diffuse, indicating alterations of the FAZ typical filamentous structure.

Figure 5.18.

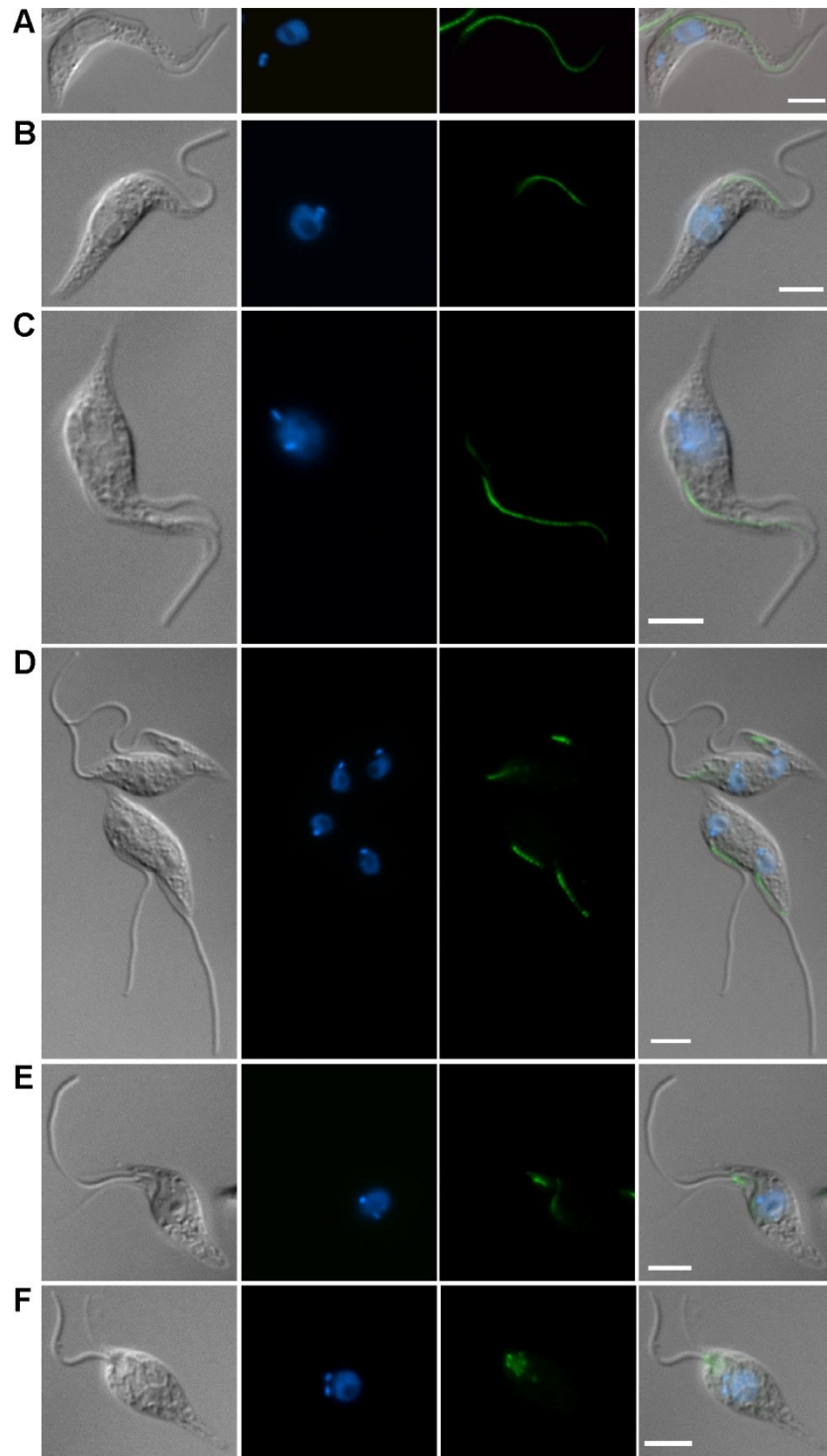


Figure 5.18. Immunofluorescence analysis of PCF DS-N2 RNAi cell line. L3B2 mAb (anti-FAZ1 Ab). DIC, DAPI, L3B2 and merged 100x images.

A. 0h normal 1K1N cell showing L3B2 typical linear labelling of the FAZ.

B-C. 18h induced cells with a short FAZ. **D-E.** 24h induced cells showing an even shorter FAZ.

F. 48h induced, epimastigote-like 2K1N cell, showing a very short FAZ at the anterior end, which has a diffused structure. In all cases L3B2 staining of the FAZ extends from the position anterior to the flagella pocket (FP) to the anterior end of the cells. Note the reduction in the FAZ length as the kinetoplast and FP progress towards the anterior cell ends. As the FAZ becomes shorter, L3B2 signal seems more diffused, suggesting disorganization in the FAZ structure. Scale bars represent 2µm.

In addition, many of these epimastigote-like cells also have a bifurcated anterior end and/or microtubule projections (Fig. 5.18E-F).

DS-N2 uninduced cells stained with CD10 and DOT1 mAb showed a punctuated labelling of the FAZ comparable to that of normal procyclic trypomastigotes. However, in DS-N2 RNAi induced cells, both CD10 and DOT1 staining resulted in a labelling pattern similar to that obtained with the L3B2 mAb: a short FAZ filament that in many cases has a diffuse structure after 48h of induction (Fig. 5.19). Neither CD10 nor DOT1 showed any reduction in the intensity of their immunofluorescence labelling. Thus, although I could not demonstrate RNAi efficiency and the consequent disappearance of group 4 calpain-like proteins, as I lack the *T. brucei* anti-GM6 antibody used by Muller and co-workers (Muller *et al.*, 1992), it is unlikely that these antibodies recognise any of the RNAi induced group 4 calpain-like proteins. Moreover, my results suggest that the knockdown of group 4 calpain-like sequences in *T. brucei* does not prevent localisation of certain other components of the FAZ, although it alters the FAZ length and structure, further supporting the FAZ localisation of these proteins (Muller *et al.*, 1992; Cotrim *et al.*, 1995).

5.3.6 SEM analysis of group 4 calpain-like proteins RNAi cell lines showed aberrant morphogenesis of induced cells that resemble epimastigote-like forms.

Analysis of DS-N2 RNAi induced cells by SEM revealed that epimastigote-like forms appeared as early as 15h after induction (Fig. 5.20B). Upon induction, the new flagellum of cells in the early stages of the cell division-cycle always appeared close to the old flagellum and linked to it by the flagella connector (Fig. 5.21D), making it difficult to establish the length of the new flagellum overhang. However, in induced cells more advanced in the cell division-cycle, the overhang of the new flagellum was clearly longer than in not-induced cells (0h, $2.1 \pm 0.51 \mu\text{m}$ (mean $\pm$ SD), 15h, $5.97 \pm 1.94 \mu\text{m}$;

Figure 5.19.

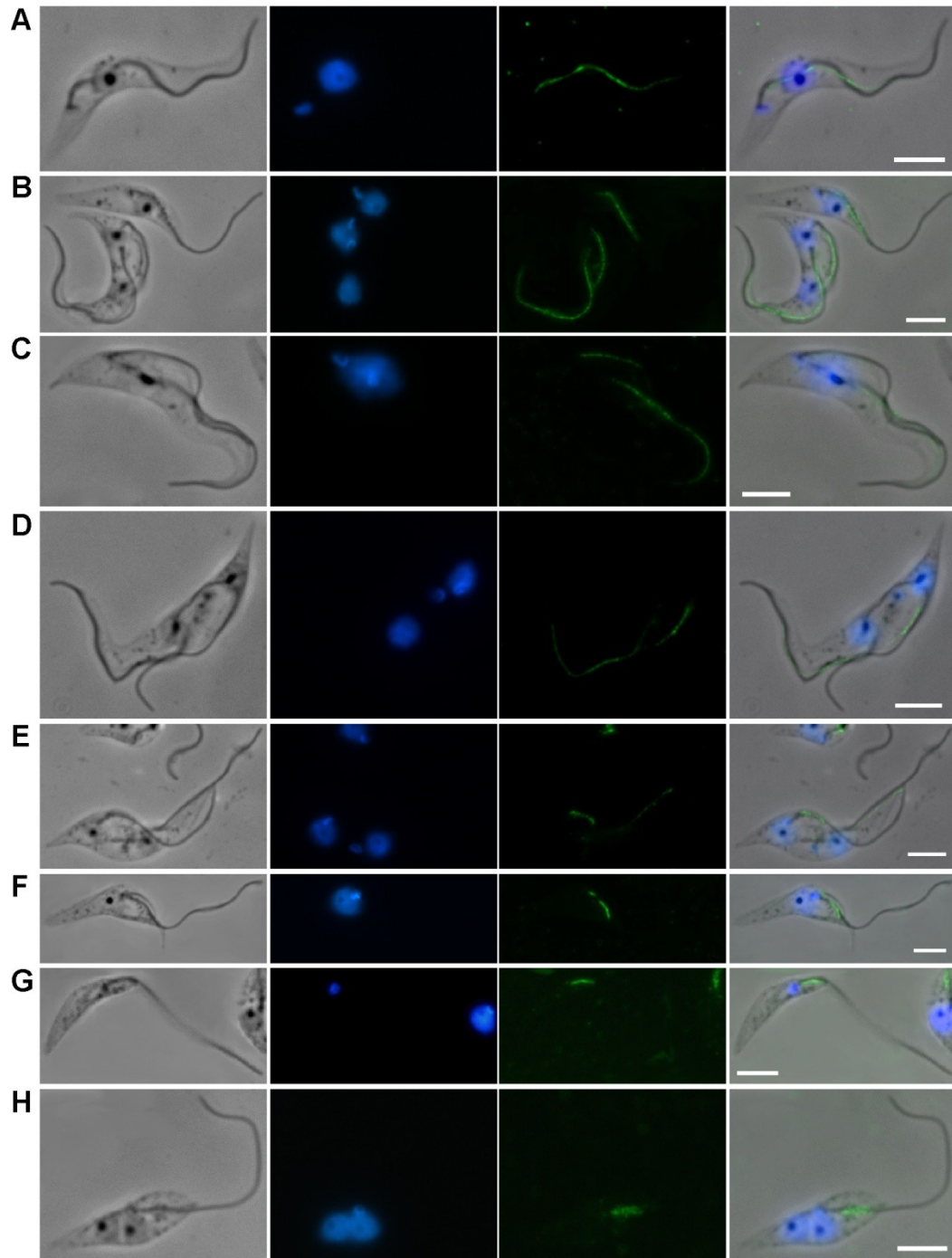


Figure 5.19. Immunofluorescence analysis of PCF of DS-N2 RNAi cell line. Detergent-extracted cytoskeletons. CD10 mAb.

Phase contrast, DAPI, CD10 mAb and merged 100x images.

A. Uninduced 1K1N cell showing the typical punctuated CD10 labelling alongside the FAZ.

B-D. 18h induced cells in different stages of the division cycle.

E-F. 24h induced 2K2N and 1K1N DS-N2 cells.

G-H. 48h induced 1K0N and 1K2N cells.

All induced cells show the progressive shortening of the FAZ labelled with CD10 mAb. The fluorescence intensity of CD10 labelling does not decrease after induction, instead the FAZ labelling pattern becomes more diffused, indicating a disorganisation of the FAZ structure. Scale bars represent 2µm.

Figure 5.20.

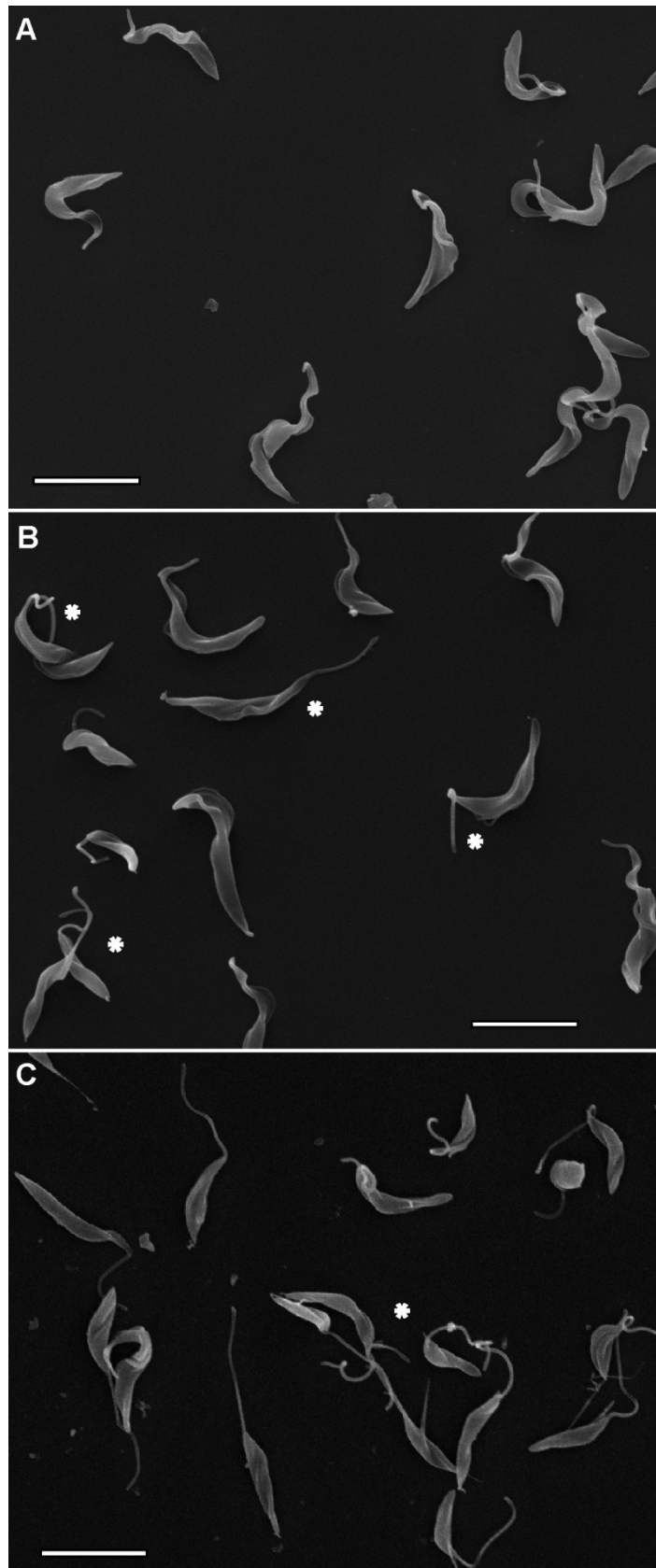


Figure 5.20. SEM analysis of PCF DS-N2 RNAi cell line. Low magnitude. 0-48h of induction. A. 0h controls. **B.** 15h induced culture showing the appearance of epimastigote-like cells (*). **C.** 48h induced culture with the majority of cells showing an epimastigote-like morphology. Scale bars represent 10 μm .

Figure 5.21.

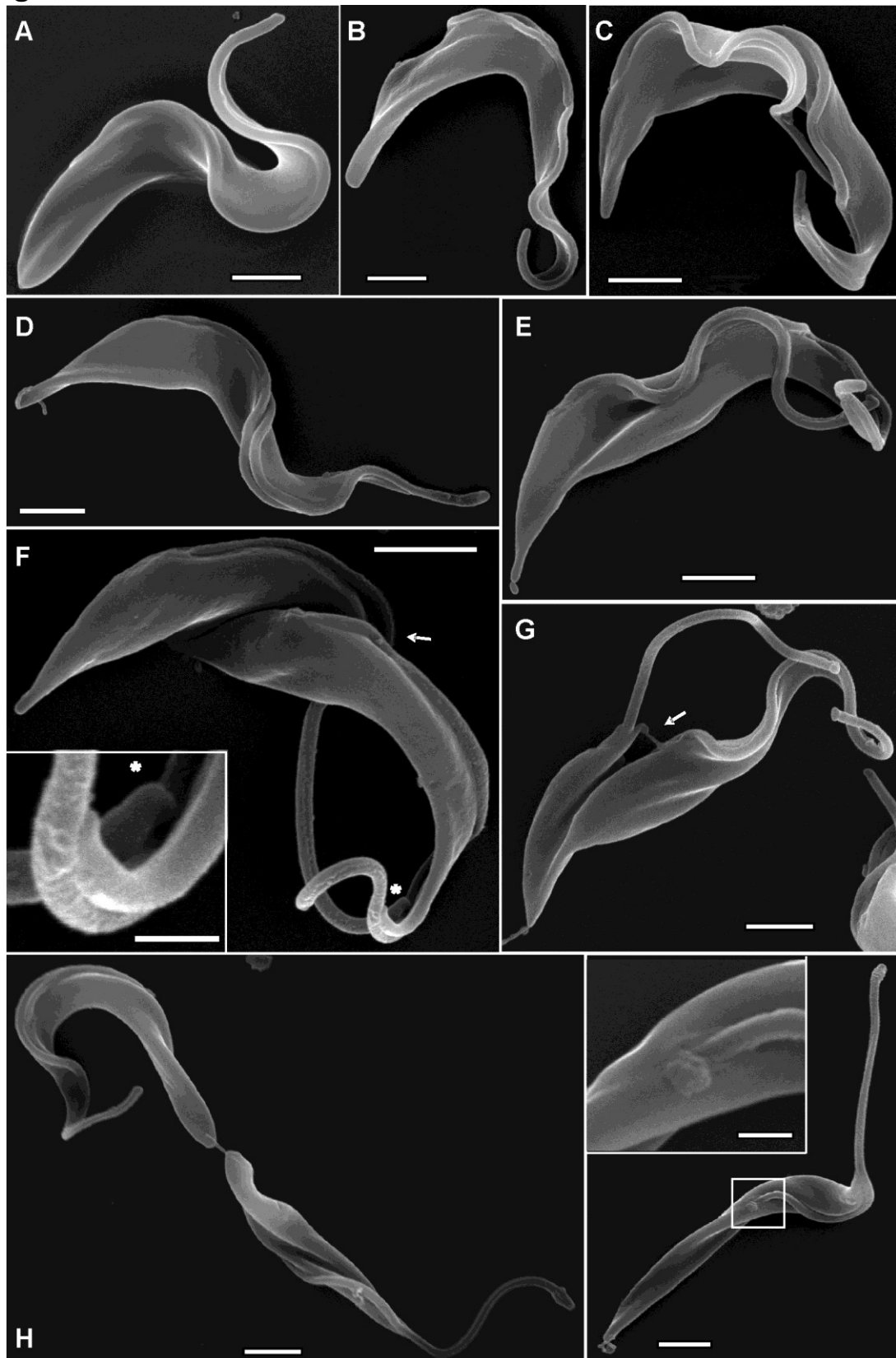


Figure 5.21. SEM analysis of DS-N2 RNAi cell line. High magnitude. 0-15h post-induction.
A-C. 0h, normal procyclic trypomastigotes. **D- I.** 15h induced cells with epimastigote-like morphology: anterior flagellar pocket and long flagellar overhangs. The flagellar connector (*) is present in cells with an advanced cleavage furrow (arrows). Scale bars represent 2µm, inset 500nm.

Appendix 5, Fig A.5.2B and Fig. 5.21C and 5.21E-H). When the cleavage furrow was visible, it clearly showed that the posterior cell product of division, containing the new flagellum, would be shorter than the anterior product but also that its morphology would resemble that of the epimastigote forms found in the proventriculus and salivary glands of the tsetse fly —flagellar pocket placed at the middle or anterior end of the cell (Fig. 5.21F-G). However, the anterior cell product of division would have trypomastigote morphology, albeit with a longer flagellar overhang than normal PCF ($4.12 \pm 3.3 \mu\text{m}$, Appendix 5, Fig.A.5.2B). The same phenomenon was observed after 24h of induction (Fig. 5.22B-E). However, at 48h of induction, after more than two cell division-cycles have been completed, dividing cells with a cleavage furrow showed that both products of cell division —anterior and posterior— would have an epimastigote-like morphology with predominantly anterior flagellar pockets and long flagellar overhangs (Fig. 5.23B-D).

The flagella of these cells seem to be attached to the cell-body from their exit of the flagellar pocket to the anterior cell-end (Fig. 5.23B-E). Moreover, the new flagellum remains associated to the old one by the flagella connector (Fig. 5.23C-E), as it has been reported to occur in tsetse fly epimastigotes (*Sharma et al., 2008*). These data suggest that the RNAi knockdown of group 4 calpain-like proteins did not caused aberrant morphology of the existing cells at the moment of induction, but prevented the normal morphogenesis of the posterior product of cell division. Eventually, after 48h of induction, abnormal cells with epimastigote-like morphology outnumbered those with a normal morphology (Fig. 5.20C). More importantly, the SEM analysis of 48h induced DS-N2 RNAi cells showed the presence of dividing epimastigote-like cells in the last stages of cytokinesis, indicating that some of these cells are able to complete cytokinesis (Fig. 5.24D).

Figure 5.22.

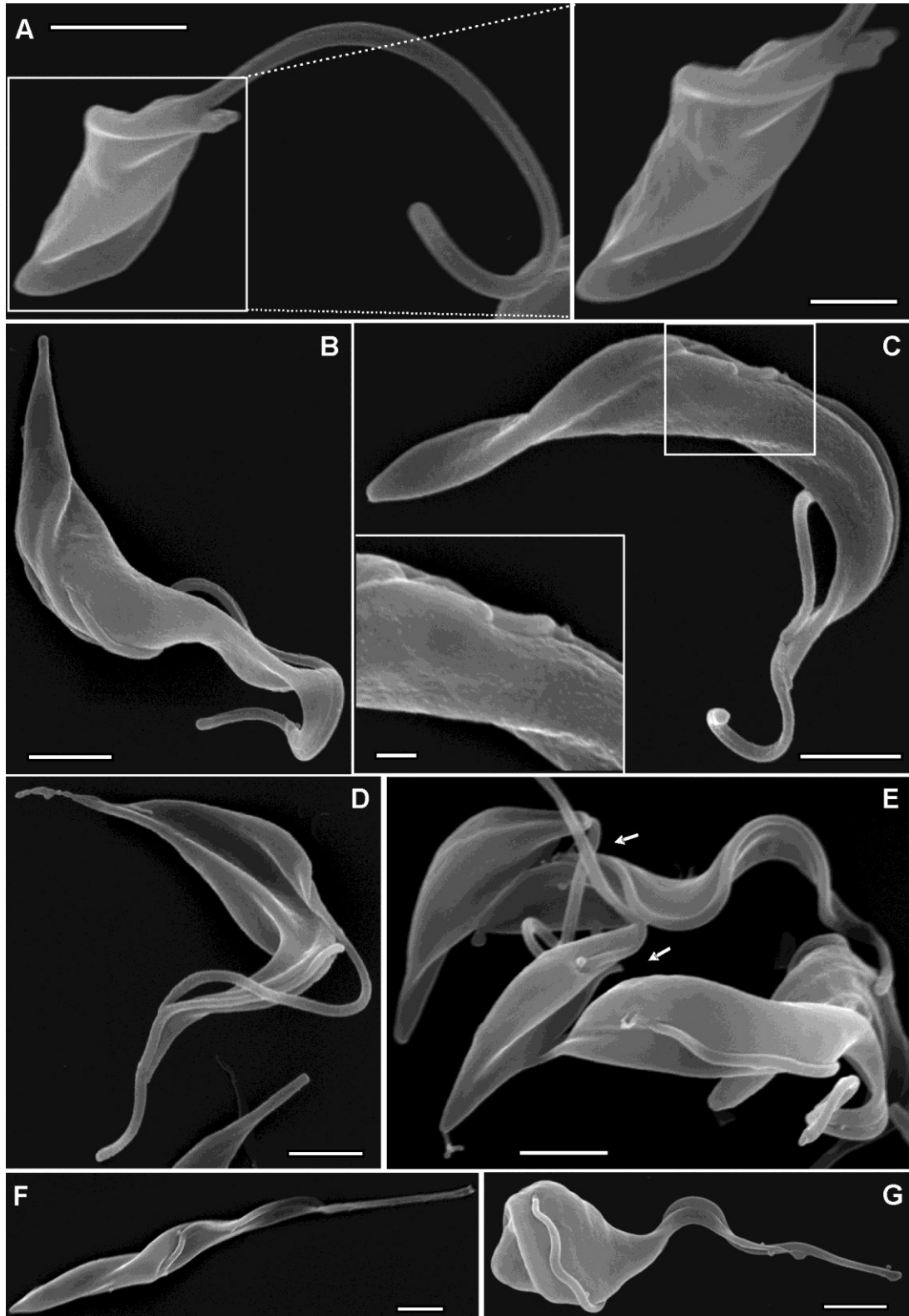


Figure 5.22. SEM analysis of PCF DS-N2 RNAi cell line. 24h post-induction.

A. Cell with short cell-body, anterior FP and long flagellar (Fr) overhang. **B.** Cell in an early stage of the division cycle with a spike-like posterior protrusion. **C-E.** Epimastigote-like cells undergoing cytokinesis, with anterior FPs (C, inset) and ingressing cleavage furrows (E, arrows). Note the smaller size of the epimastigote-like posterior product of division. **F.** Epimastigote-like cell, long Fr overhang. **G.** Cell with an enlarged posterior end. Scale bars represent 2 μ m, insets 500 nm.

Figure 5.23.

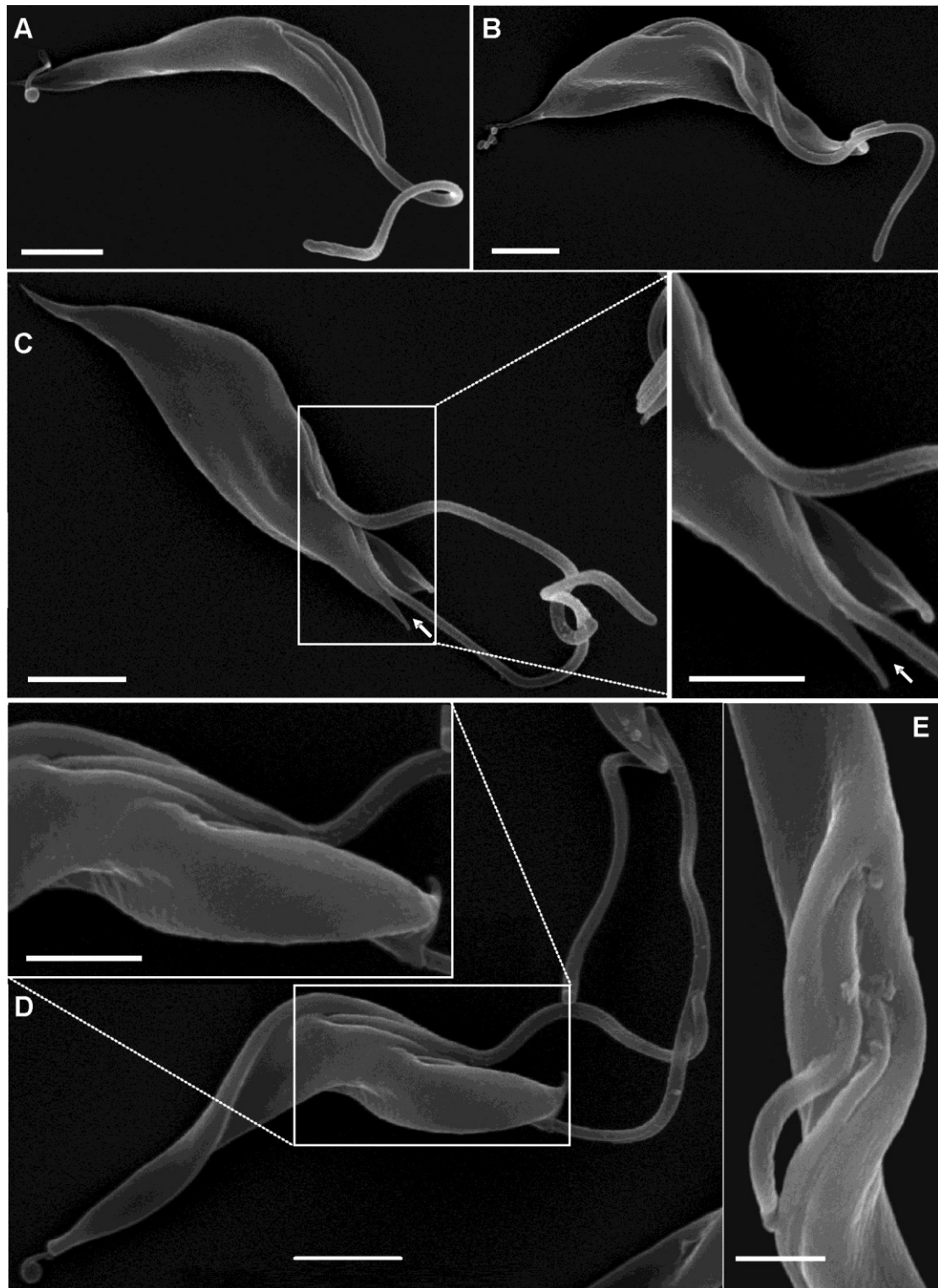


Figure 5.23. SEM analysis of PCF DS-N2 RNAi cell line. 48h post-induction.

A. Epimastigote-like cell with a long flagellar overhang. **B-D.** Dividing epimastigote-like cells with anterior new and old flagellar pockets (insets) and long flagellar overhangs. **E.** Dividing cell where the new flagellum is clearly associated to the old one by the flagella connector. Note the enlarged anterior cell-ends or cytoskeletal projections in C and D (arrows) and the presence of a spike-like protrusion at the posterior end of epimastigote-like cells (A-D). Scale bars represent 2 μ m, insets scale bars represent 500 nm.

Figure 5.24.

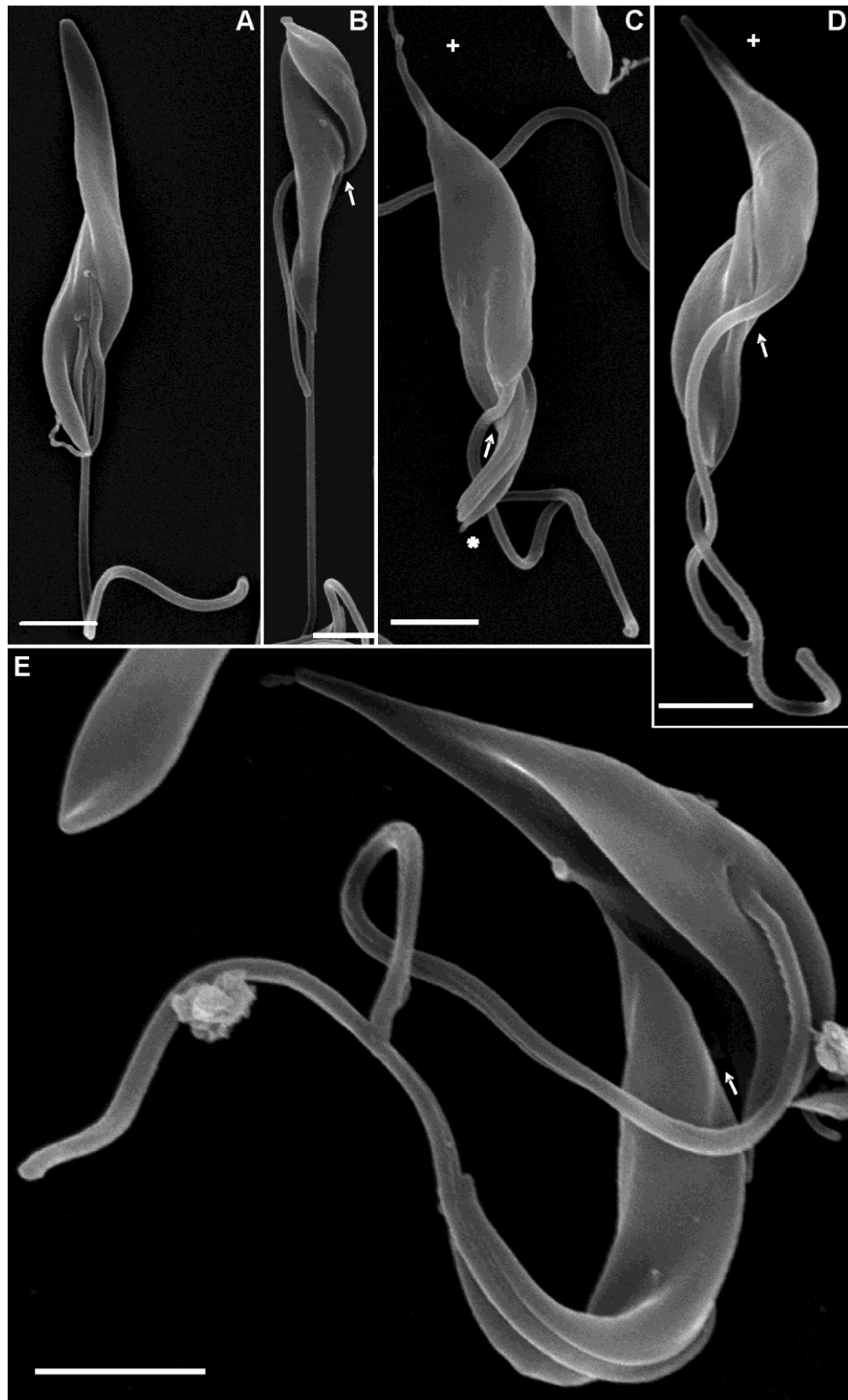


Figure 5.24. SEM analysis of PCF DS-N2-RNAi cell line. 48h post-induction (2).
A-E. Dividing epimastigote-like cells in different stages of the division cycle. Note the presence of a FC even late at cytokinesis; arrows indicate the position and direction of the cleavage furrow; (*) anterior cytoskeletal projections and (+) posterior spike-like protrusions. Scale bars represent 2 μm.

Additionally, SEM analysis showed that many induced cells have a spike-like protrusion at the posterior end (Fig. 5.21D, E, G and I; Fig. 5.22C-E; Fig. 5.23A-D and Fig. 5.24B). This structure is similar to that observed in *T. brucei* tsetse epimastigotes, which has been speculated that to be involved in host-parasite interactions, as it increases the surface area of the plasma membrane exposed to the environment (Sharma *et al.*, 2008).

5.3.7. RNAi knockdown of group 4 calpain-like proteins causes aberrant assembly of the FAZ in PCF trypanosomes.

Examination of the ultrastructure of the FAZ by TEM in 48h induced DS-N2 RNAi cells revealed several abnormalities in the FAZ assembly. Cross sections of uninduced cells showed the typical appearance of the FAZ in wild type PCF, with a FAZ filament of the appropriate size and position near the MtQ (Fig. 5.25A). However, cross sections of 48h induced cells showed an increased distance between the two sub-pellicular microtubules where the FAZ filament is located (Fig. 5.25B-D, brackets) and some areas of more electron dense material juxtaposed to the cell-body membrane than in a normal FAZ. Cross sections also showed abnormal positioning or absence of the endomembrane associated to the MtQ (Fig. 5.25B-D). This ultrastructure phenotype resembles that obtained after the RNAi knockdown of the FAZ protein FAZ1 (Vaughan *et al.*, 2008). However, RNAi silencing of FAZ1 produced a very different cell morphology and morphogenesis phenotype. FAZ1 RNAi knockdown in procyclic trypanosomes results in detached flagella and mispositioning of the cleavage plane, giving rise to zoids and its 1K2N siblings. In addition, FAZ1 RNAi causes cytokinesis impairment, evidenced by the appearance of multinuclear cells (Vaughan *et al.*, 2008). The RNAi phenotype of group 4 calpain-like sequences presented here consists in an alteration of the FAZ structure and length, noticeable at the cellular and ultrastructural levels, without altering its function of flagellar/cell-body attachment. The mechanisms

Figure 5.25.

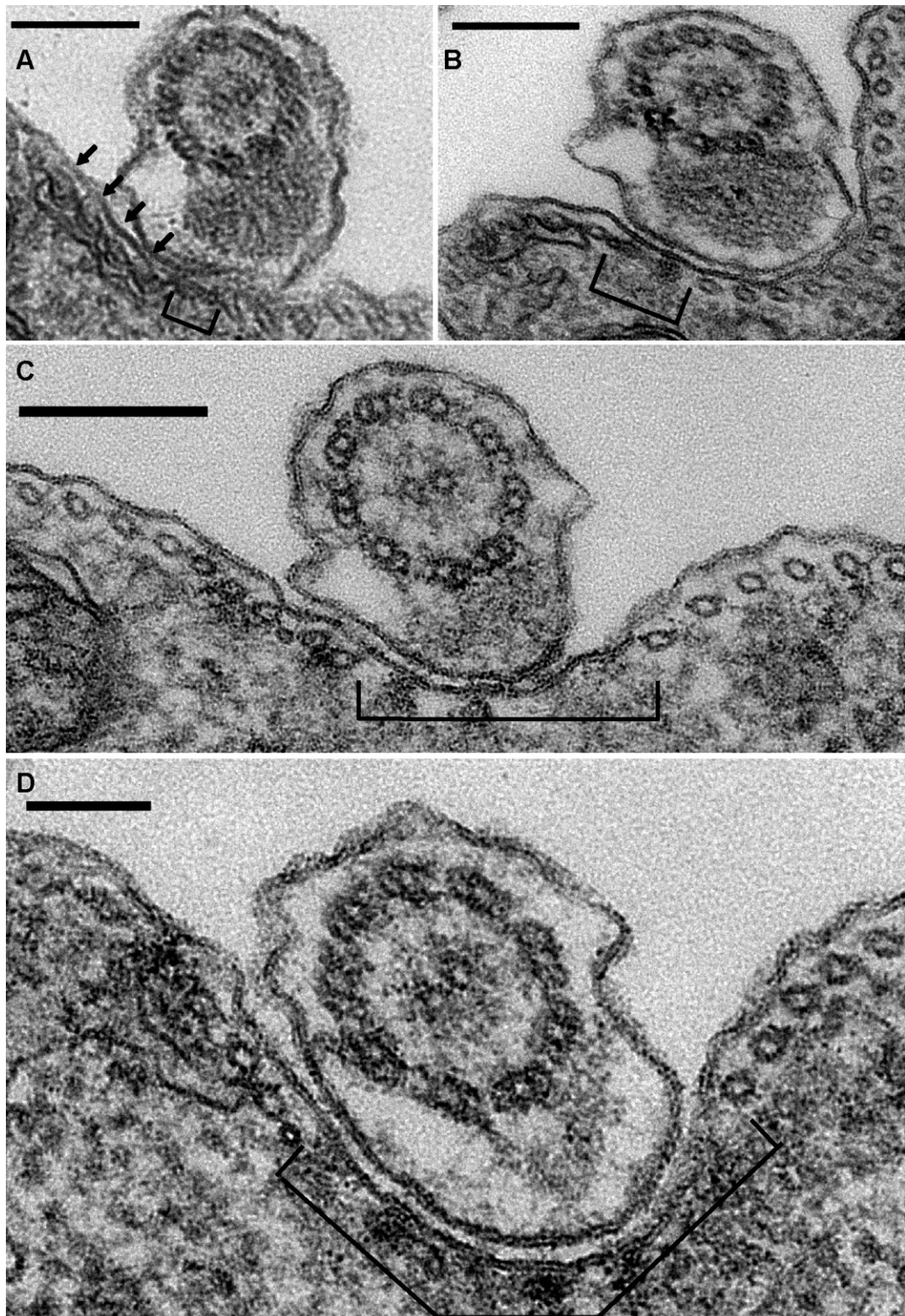


Figure 5.25. TEM analysis of PCF DS-N2 RNAi cell line. 0- 48h post-induction.

A. Non-induced control. The MtQ is indicated by arrows and the FAZ filament area is marked with a bracket. **B-D.** Cross sections of 48h induced cells showing abnormally wide FAZ filament areas (brackets), with regions of very electron-dense material. The MtQs have errors in the positioning of the endomembrane (normally located across the whole MtQ), which appear associated only to two or three specialised microtubules (B and D) or is absent altogether (C). Scale bars represent 200 nm.

through which group 4 calpain-like proteins influence the position and length of the FAZ remain to be elucidated. Thus, an examination of their binding partners at the FAZ would be the logical next step in the investigation of their function.

5.4. RNAi functional analysis of group 4 calpain-like sequences in *T. brucei* BSF targeting the Lcr1-like repeats: Tb11.47.0035 and Tb11.v4.0001.

I used the 35-01-Rep RNAi construct to selectively silence the expression of proteins containing Lcr1-like repeats, Tb11.47.0035 and Tb11.v4.0001 (Fig. 5.1). As discussed earlier, Tb11.47.0035 ORF lacks a stop codon and Tb11.v4.0001 ORF starts with a region 99.8% identical to a repeat region of Tb11.47.0035, strongly suggesting that both sequences may be part of a unique ORF. The 35-01-Rep RNAi construct targets these sequences 99.8% identical repeat region, having the capability to knockdown the expression of both Tb11.47.0035 and Tb11.v4.0001 if they were different, simultaneously expressed proteins. *T. brucei* procyclic and bloodstream forms were transfected with the 35-01 RNAi construct. However, stable transfectants were only obtained in the BSF. A possible explanation could be an essential role of these proteins in procyclic trypanosomes, with a lethal effect of RNAi leakiness in emerging transfectants.

Induction of the bloodstream 35-01-Rep RNAi cell line did not produce any discernible growth or morphology phenotype (Appendix 5, Fig. A.5.3). These data, together with the strong effect in PCF caused by the RNAi induction of cell lines obtained with the DS-N2 and HP-N constructs in PCF, which have the capability to knockdown the expression of Tb11.47.0035 and Tb11.57.0008, suggest that these proteins are either differently expressed in PCF and BSF trypanosomes —Tb11.47.0035/ Tb11.v4.0001 in PCF and Tb11.47.0036/Tb11.57.0008 in BSF— or that the phenotype obtained in PCF

with the DS-N2 and HP-N constructs was exclusively due to the silencing of Tb11.57.0008, which is not target by the 35-01 RNAi construct.

5.5. RNAi functional analysis of group 4 calpain-like sequences in *T. brucei* BSF targeting the GM6-like repeats: Tb11.47.0036 and Tb11.57.0008.

T. brucei BSF cell line was transfected with the 08-36-Rep RNAi construct, which targets a GM6-like repeat region of Tb11.47.0036 and Tb11.57.0008 (94.9% identity, Fig. 5.1). Induction of stable transfectants produced a strong growth (Fig. 5.26) and morphology phenotype. Preliminary analysis of these phenotypes suggests that RNAi of Tb11.47.0036 and Tb11.57.0008 have a lethal effect in bloodstream trypomastigotes, possibly causing aberrant morphogenesis and cytokinesis. This result is compatible with the two hypotheses discussed above. Namely, that the two putative ORF, Tb11.47.0035/Tb11.v4.0001 and Tb11.47.0036/Tb11.57.0008, are stage-specific regulated, in PCF —the former— and BSF —the later— trypanosomes; or that the two RNAi phenotypes obtained by silencing group 4 calpain-like sequences in procyclic and bloodstream forms were exclusively due to the knockdown of Tb11.57.0008 or Tb11.47.0036/Tb11.57.0008 expression. In any case, these experiments indicate that some *T. brucei* group 4 calpain-like proteins are essential for the correct morphogenesis of both, procyclic and bloodstream trypanosomes.

Figure 5.26.

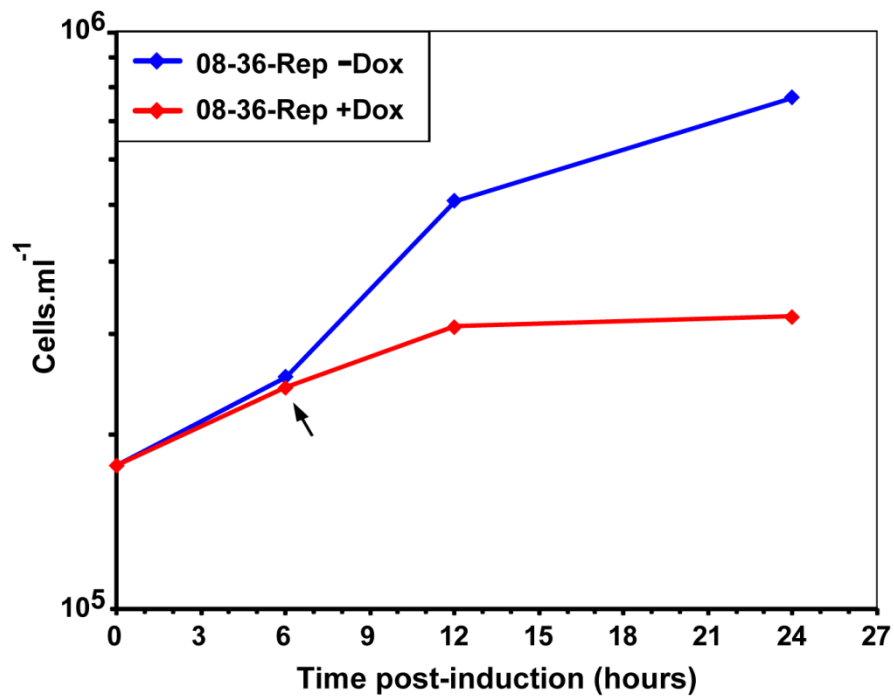


Figure 5.26. BSF 08-36-Rep RNAi cell line induction. Growth curve.

Cultures were grown in presence or absence of doxycycline for 24h. The induced culture grew at a similar rate as the non-induced during the first 6h of induction (arrows). After this time, the induced culture grew at significant slower rate than the non-induced.

CHAPTER 6

DISCUSSION

6. DISCUSSION.

The calpains constitute a diverse superfamily of cysteine proteases expanded in animals, ciliates and flagellated kinetoplastid parasites (*Croall and Ersfeld, 2007*). The majority of other organisms, sampled so far, have only one or two calpain or calpain-related genes, including plants. Calpains and calpain-like proteins are implicated in multiple cellular events, including cytoskeletal remodelling, signal transduction, apoptosis, cellular differentiation, adhesion and migration (*Goll et al., 2003*). Since the evolutionary distance between trypanosomatids and animals is very large and streamlining of functional complexity is a trait commonly associated with parasites, the presence of large calpain-related families in trypanosomatids is particularly intriguing.

Calpains and calpain-related proteins have a modular structure and diverse domain architecture, which allow their classification into two categories. Firstly, typical calpains are defined as four domain molecules, including a calpain catalytic domain II with a conserved cysteine protease catalytic triad —CHN— and a Ca^{2+} binding EF-hand-containing domain IV, essential for enzyme activity. Secondly, atypical calpains are defined as molecules containing a calpain “catalytic” domain, which often has deviation from the canonical catalytic triad, but lacking an EF-hand Ca^{2+} binding domain (Fig. 6.1). All trypanosomatid calpain-related proteins are atypical calpains — the majority of them have a domain II with a degenerated catalytic triad and all lack a Ca^{2+} binding EF-hand domain. In this thesis, I have performed a bioinformatic survey of the large and diverse group of calpain-like sequences present in the trypanosomatid parasites *T. brucei*, *L. major* —previously studied by Ersfeld and co-workers (*Ersfeld et al., 2005*)— and *L. infantum*. This survey led to the identification of 31 new trypanosomatid calpain-like proteins: 10 in *T. brucei*, 10 in *L. major* and 11 in *L. infantum*. There are two main reasons that could explain why I have managed to identified so many new members of the calpain-like family in kinetoplastids that were not previously detected

Figure 6.1.

TYPICAL CALPAINS

Human calpain 1, 2, 8, 9 and 11
Drosophila CALP A, B and C

Human Calpain 3

ATYPICAL CALPAINS

Human Calpain 4 Small subunit

Human Calpain 12 and 14

Human Calpain 5 and 6
C. elegans TRA-3

Human Calpain 10

Human Calpain 7
Aspergillus nidulans PalB
Saccharomyces cerevisiae RIM13

Human Calpain 13

Human Calpain 15
Drosophila SOL

Tetrahymena thermophila EAR84910

T. thermophila EAR94203

T. thermophila EAS05791

T. brucei CAP5.5*

T. brucei Group 2*

T. brucei Group 2 Tb927.6.3310*

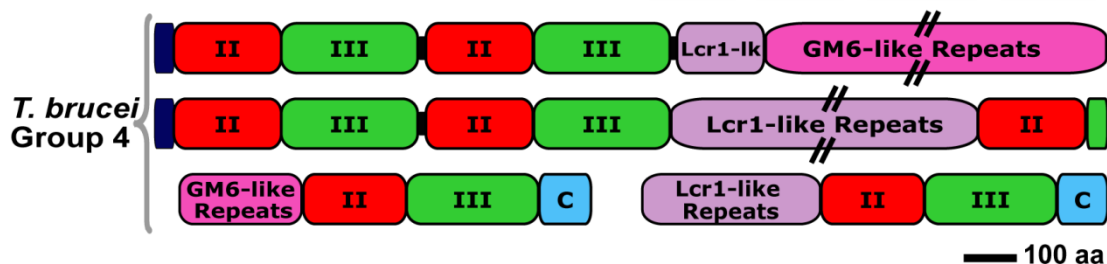


Figure 6.1. Domain structure of known typical and atypical calpains, including domain II-containing kinetoplastid calpain-like sequences. Domain abbreviations: C2, protein kinase C conserved region 2 (involved in calcium-dependent phospholipid binding); IV^{dEF}, domain IV with degenerate EF-hand motifs that are unlikely to bind calcium; EF, domain with EF-hand motifs distinct from domain IV; I^K, kinetoplastid acylated domain I (myristic acid and palmitic acid chains are indicated by zigzag lines); MIT, microtubule interacting and trafficking molecule domain; PBH, palB-homologous domain; PKA, protein kinase A regulatory subunit domain; SOLH, small optic lobe domain; TMD, transmembrane domain; Zn, zinc finger domain. Except *T. brucei* group 4 sequences (which repeats are truncated due to their long length), all sequences are drawn to scale. Scale bar represents 100 amino acids.

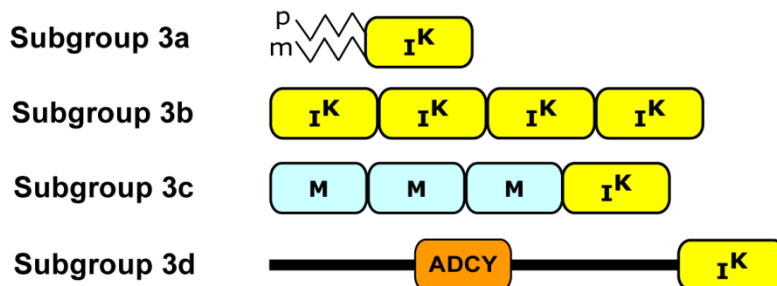
by Ersfeld and co-workers (*Ersfeld et al., 2005*). The first one is the inclusion of new genes in the version 4 update of *T. brucei* genome sequence on geneDB (<http://www.genedb.org/tryp>), which was released after Ersfeld *et al.* publication (*Ersfeld et al., 2005*) —for instance, the inclusion of the newly identified CAP5.5 paralog, CAP5.5V. The second reason is the detailed analysis, performed during this thesis work, of the highest scoring sequences from the BLASTp searches for calpain-like sequences in the three studied kinetoplastid species, which allowed the identification of additional domains in some newly identified sequences and the modification of the domain architecture of some of the previously described calpain-related proteins — for example, the lack of domains I^K and C in *L. major* LmjF20.1210. This detailed analysis consisted on the comparison of the BLASTp highest scoring sequences to known calpain-like proteins, numerous multiple alignments and thorough manual editing, as shown in chapter 3 (Figures 3.1 to 3.16), which not only identified new members of the calpain-like family but also revealed the novel domain composition and architecture of some newly and previously described calpain-related sequences. These sequences could not be classified into the five groups of kinetoplastid calpain-like proteins proposed by Ersfeld and co-workers (*Ersfeld et al., 2005*), and, therefore, I reclassified them into five tailored groups (Chapter 3, Fig. 3.20).

Two of these groups, group 3 —3a: domain I^K; 3b: four copies of domain I^K; 3c: three degenerate copies of the META domain and a domain I^K; 3d: an adenylate cyclase catalytic domain (ADCY) and a domain I^K— and group 5 — a kinesin-like motor domain and calpain-like domain III— do not match the definition of typical or atypical calpain proteins, as they include sequences that lack a calpain “catalytic” domain II (Fig. 6.2A). In spite of this, both groups of proteins were included in my classification of kinetoplastid calpain-like proteins. Group 3, comprising sequences with a domain I^K but lacking any calpain-related domain, was included for consistency with the previously

Figure 6.2.

A. Groups of kinetoplastid calpain-like proteins constituted by sequences lacking a calpain "catalytic" domain II.

Group 3:



Group 5:



B. Classification of kinetoplastid calpain-like proteins strictly applying the definition of typical and atypical calpains.

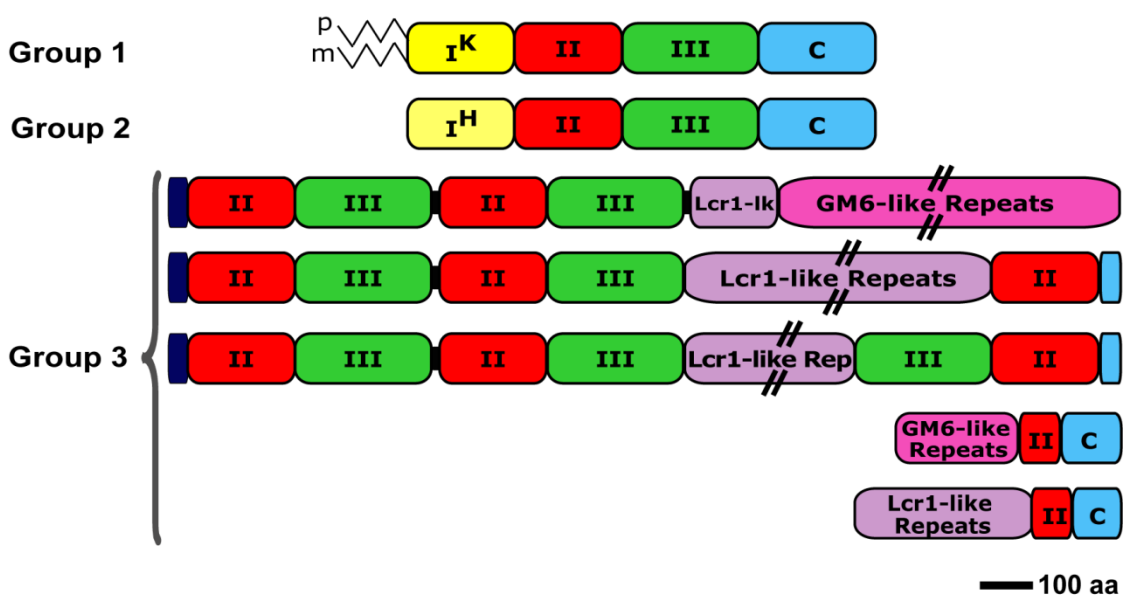


Figure 6.2.
A. Groups of kinetoplastid calpain-like sequences that do not strictly match the definition of typical or atypical calpains, as they are constituted by sequences lacking a calpain "catalytic" domain II. M, META domain; ADCY, adenylate cyclase catalytic domain; KINESIN, kinesin-like motor domain.
B. Classification of kinetoplastid calpain-like proteins strictly applying the definition of typical and atypical calpains.
Scale bar represents 100 amino acids (aa).

published classification, (Ersfeld *et al.*, 2005), which was based on the unique correlation between domain I^K and kinetoplastid calpains-like proteins. Similarly, the inclusion of group 5 sequences in my classification of trypanosomatid calpain-related proteins was done on the rationale that these proteins are the only trypanosomatid sequences that contain a calpain-related domain III whereas lacking a calpain “catalytic” domain II. Therefore, strictly applying the definition of typical and atypical calpains, kinetoplastid calpain-like proteins should be classified into only three groups: group 1: I^K-II-III-C; group 2: I^H-II-III-C and group 3: sequences containing 1 to 2 copies of domain II and 67-68 amino acids tandem repeats (Fig. 6.2B). According to this alternative classification, the number of newly identified calpain-like sequences in the three studied kinetoplastids would fall to 20 —5 in *T. brucei*, 7 in *L. major* and 8 in *L. infantum* (Table 6.1). However, the classification used throughout this thesis is the one comprising five groups of kinetoplastid “calpain-like” sequences (Fig. 3.20), as the function or characteristics of the I^K, META, ADCY and kinesin-like motor domains could help to understand evolution and function of some kinetoplastid calpain-like proteins.

Based on the results of the bioinformatic survey discussed above, I applied the RNAi phenotype test to selected *T. brucei* calpain-like proteins in both, procyclic and bloodstream trypomastigotes. Although some of these sequences did not produce any discernible phenotype, the RNAi knockdown of several calpain-like proteins resulted in strong growth and morphology phenotypes: CAP5.5, CAP5.5V and group 4, repeat-containing calpain-like sequences. Here I summarise these findings and discuss them in the context of our current knowledge of calpains and the biology of trypanosomatid parasites.

Table 6.1.

	<i>T. brucei</i>	<i>L. major</i>	<i>L. infantum</i>
GROUP 1 I^K-II-III-C	Tb927.1.2100	LmjF20.1180	LinJ20.1210
		LmjF20.1185	LinJ20.1220
	Tb927.1.2110	LmjF20.1190	LinJ20.1230
	Tb927.1.2120		
	Tb927.4.3940	LmjF31.0390	LinJ31.1140
		LmjF31.0400	LinJ31.1150
		LmjF31.0410	LinJ31.1160
		LmjF31.0430	LinJ31.0590
		LmjF31.0460	LinJ31.0550
	Tb927.4.3950 CAP5.5	LmjF31.0440	LinJ31_V3.0450
	Tb927.8.8330 CAP5.5V		
	Tb9.160.5550	LmjF04.0450	LinJ04.0440
		LmjF20.1200	LinJ20.1240
		LmjF25.1480	LinJ25.1550
	Tb11.01.68.70 (I ^K -II only)	LmjF32.2010	LinJ32_V3.2140
		LmjF20.1210 (II-III only)	LinJ20_V3.1250
Group 1 number (old + new)	8 = 6 old + 2 new	14 = 12 old + 2 new	14 = 12 old + 2 new
GROUP 2 I^H-II-III-C	Tb927.6.3310	LmjF30.2040	LinJ30_V3.2040
	Tb10.389.0470	LmjF18.1060	LinJ18.1070
	Tb10.70.5950	LmjF21.0120	LinJ21.1770
			LinJ21.0010
	Tb11.01.5800	LmjF32.0970	LinJ32.1370
	Tb11.02.2240	LmjF33.2010	LinJ33.2020
		LmjF34.0280	LinJ34_V3.0300
		LmjF36.0780	LinJ36_V3.0840
Group 2 number (old + new)	5 = 4 old + 1 new	7 = 2 old + 5 new	8 = 2 old + 6 new
GROUP 3 Ila-Illa-lib-IIIb-R-Ilc- IIc Ila IIIa IIb IIIb R R-II-III-C	Tb11.47.0035	LmjF27.0500	LinJ27_V3.0510
		LmjF27.0510	LinJ27_V3.2490
	Tb11.57.0008	LmjF27.0490	LinJ27.V3.0500
	Tb11.47.0036		
	Tb11.v4.0001		
Group 3 number (old + new)	4 = 2 old + 2 new	3 old	3 old
TOTAL NUMBER	17= 12 old + 5 new	24 = 17 old + 7 new	25 = 17 old + 8 new

Table 6.1. Kinetoplastid calpain-like sequences containing a calpain “catalytic” domain II. Calpain-like proteins in the kinetoplastids *T. brucei*, *L. major* and *L. infantum* strictly applying the definition of typical and atypical calpains. Newly identified sequences are in red font.

6.1. Calpain-like proteins are essential for the correct morphogenesis of *T. brucei* procyclic and bloodstream forms.

T. brucei CAP5.5 and CAP5.5V are paralogous, differentially expressed calpain-related proteins (Chapter 3, Fig.3.16-3.19). Both proteins are associated to the subpellicular cytoskeleton (*Hertz-Fowler et al., 2001; Liu et al., 2009*) and contain N-terminal myristoylation and palmitoylation motifs, suggesting an interaction with the plasma membrane (*Resh, 1999*). Moreover, CAP5.5 double acylation has been experimentally demonstrated (*Hertz-Fowler et al., 2001*). Thus, prior to the work presented in this thesis, the likeliest predicted function for CAP5.5 was a role in cytoskeletal remodelling or vector-parasite mediated signal transduction. However, the RNAi functional analysis of CAP5.5 and CAP5.5V presented in chapter 4 shows that these paralogous proteins are essential for the correct organisation of the microtubule corset and morphogenetic patterning during the cell division-cycle of procyclic and bloodstream trypomastigotes, respectively (Chapter 4 and *Olego-Fernandez et al., 2009*). Thus, paralogous genes seem to provide analogous roles in the different life-cycle stages in which they are expressed. Additionally, CAP5.5 functional analysis evidences the essential role of a protein that has been used as a late stage marker of bloodstream-procyclic trypomastigote differentiation for the last 16 years (*Matthews and Gull, 1994a*). The key issues to resolve in the future are how CAP5.5 and CAP5.5V contribute to cytoskeletal remodelling, whether the function of either protein is dependent upon an intrinsic proteolytic activity and if inhibition of CAP5.5V function might provide a realistic new drug target for sleeping sickness. These would require:

- Immunoprecipitation analysis of CAP5.5 and CAP5.5V —upon obtaining an anti-CAP5.5V Ab— to identify their binding partners.
- Recombinant expression of CAP5.5 and CAP5.5V, and assay of the calpain

protease of the recombinant proteins.

- Search for drugs to inhibit CAP5.5V protease activity, if any, or to block CAP5.5V interaction with the subpellicular cytoskeleton of *T. brucei* through *in vitro* assays and *in vivo* experiments in mice.

The earliest visible phenotype following CAP5.5 RNAi induction in procyclic trypomastigotes was the loss of CAP5.5 from the posterior cell-ends, followed by a profound effect on both cell growth and cell morphology. It is from the posterior cell-end that procyclic cells elongate during the early stages of the cell-cycle, that involves both microtubule extension and the intercalation of new microtubules into the subpellicular microtubule array (Sherwin *et al.*, 1987). CAP5.5 loss from the posterior end of the cells was evident after only 8h of induction. Considering that the cell-cycle time for not-induced cultures was ~11h, 2K1N and 2K2N induced cells showing a loss of CAP5.5 from their posterior ends after only 8h of induction, would have been at an earlier stage in the same division-cycle at the point when RNAi was induced. At later points after induction, loss of CAP5.5 from the posterior cell-ends correlated with the aberrant positioning of nuclei following mitosis. Whereas in normal 2K2N cells the posterior nucleus sits between the divided and segregated kinetoplasts, in CAP5.5 RNAi induced 2K2N cells the posterior nucleus is located in a position anterior to the most posterior kinetoplast (Fig. 4.12 and 4.14). Despite nuclear mispositioning, cytokinesis still occurred in RNAi induced cells, giving rise to a 1K2N cell and an anucleate zoid, 1K0N (Fig. 4.9). Zoids contained either little or no CAP5.5 protein, thus, the older observation that inheritance of subpellicular microtubules at cytokinesis is semi-conservative (Sherwin and Gull, 1989b) is particularly significant, as the loss of CAP5.5 from the posterior cytoskeleton after only 8h of induction of RNAi points towards a dynamic kinetics of cytoskeletal remodelling prior to cell division.

The fast loss of CAP5.5 from the posterior cell-end mirrors the initial RNAi phenotype of another trypanosome MAP, WCB (*Baines and Gull, 2008*). The near-immediate loss (i.e. within the same cell-cycle in which RNAi was induced) of CAP5.5 from the posterior cytoskeleton implies that new microtubule growth and intercalation during cell elongation is accompanied by extensive re-modelling that sees CAP5.5, and presumably other MAPs, stripped entirely from the elongating subpellicular microtubule corset. However, since CAP5.5 is evenly distributed across the detergent-extracted cytoskeleton of wild type procyclic trypanosomes, irrespective of their position in the cell-cycle (Fig. 4.1), (*Hertz-Fowler et al., 2001*). Thus, my observations point towards fast, dynamic kinetics for key events in cytoskeletal remodelling: removal of MAPs, microtubule growth and intercalation and addition of MAPs to a newly re-modelled subpellicular microtubule array.

Additionally, the accumulation of multinucleate cells at later time-points after induction provides evidence that the 1K2N progeny of aberrant cell division were able to enter into further cell division-cycles. Multinucleate, multi-flagellate cells showed the ingression of several cleavage furrows, which abscission would likely produce zoids (Fig. 4.40 and 4.41). Indeed, zoids accumulated in RNAi-induced cultures as a function of time, becoming the most prevalent “cell” type by 72h post-induction (Fig. 4.9). As the cell density continued to increase between the 48 and 72h time-points, albeit at a slower rate than in not-induced cultures, the increase in the number of zoids points towards zoid longevity, which presumably cannot replicate, and/or to the ability of multinucleate cells to undergo cytokinesis and abscission, perhaps producing multiple zoids. Interestingly, the RNAi-induced loss of WCB (*Baines and Gull, 2008*) from procyclic posterior cell-ends also resulted in aberrant cytokinesis, giving rise to one zoid and one 1K2N cell. The distinctive phenomenon of zoid formation by procyclic

T. brucei has been reported previously on various occasions and occurs as a consequence of different processes, such as:

- Pharmacological inhibition of mitosis using the anti-microtubule agent rhizoxin (*Robinson et al., 1995; Ploubidou et al., 1999*).
- Pharmacological inhibition of nuclear S phase using aphidicolin as an inhibitor of DNA polymerase α (*Ploubidou et al., 1999*).
- RNAi against various components involved in cell-cycle control or organelle segregation. For example, the mitotic cyclin CYC6 (*Hammarton et al., 2003*) and centrin 4 (*Umeyama and Wang, 2008*), respectively.
- Dominant-negative expression of mutant proteins interfering with cell-cycle control/organelle segregation, like cohesin (*Gluenz et al., 2008*).
- Overexpression of the normally anterior-cytoskeleton associated proteins CAP15 and CAP17. Their accumulation at the posterior cell-end results in variable defects in organelle segregation prior to cytokinesis (*Vedrenne et al., 2002*).

Zoid formation following CAP5.5 or WCB RNAi could occur either directly as a result of aberrant cleavage furrow ingression into a defective subpellicular corset or as the inevitable consequence of incorrect organelle (nuclear) positioning in an aberrantly remodelled cytoskeleton. Ultrastructural analysis of procyclic CAP5.5 RNAi-induced cells by thin section electron microscopy showed the appearance of defects in microtubule organisation after 18h of induction in a small proportion of sections corresponding to posterior or flagellar pocket areas of cells (Fig. 4.33). However, after 48h of induction cell sections were characterised by multinucleate profiles and abnormal shape (Fig. 4.34). About 15% of cell sections also showed some form of microtubule abnormality, most notably aberrations in inter-microtubule spacing and/or the bundling of cytoplasmic microtubules beneath the plasma membrane at both the posterior cell-end/ flagellar pocket area and at more anterior regions of cells (Fig. 4.34).

Defects in microtubule organisation correlated with difficulties in flagellar pocket biogenesis, as evidenced by the presence of multiple cross-sections of flagellar areas inside pocket profiles, which occasionally showed the presence of PFR within the flagellar pocket (Fig. 4.34). Apparent irregularities in anterior-end microtubule organisation potentially represented cross-sections through zoids or through the finger-like extensions observed in large cells fixed for SEM (Fig. 4.30 and 4.31) that could correspond to partially ingressed cleavage furrows, which subsequent abscission would give rise to zoids.

Although RNAi induction of either CAP5.5 and WCB in procyclic trypanosomes results in the production of zoids, RNAi against WCB, but not CAP5.5, also results in the “blebbing” of plasma membrane vesicles to the extracellular environment (*Baines and Gull, 2008*). This suggests that an important property of WCB, but not CAP5.5, is to maintain local plasma-membrane-to-microtubule corset integrity. In Chapter 4, I have also shown that WCB localisation is unaffected in CAP5.5 RNAi induced cells, which is consistent with this hypothesis. It could be speculated that mutant-specificity of membrane blebbing is linked to the different motifs present in each protein, which can mediate an interaction with the inner-leaflet of the plasma membrane —WCB contains an N-terminal phospholipid-binding C2-like domain, whereas CAP5.5 is subject to N-terminal acylation (*Hertz-Fowler et al., 2001; Baines and Gull, 2008*).

Another significant difference between CAP5.5 and WCB RNAi phenotypes is that the environment associated with intercalation of new microtubules into the subpellicular array is strongly affected in the CAP5.5 RNAi mutant. However, in neither WCB nor other RNAi mutants in which multinucleate cells accumulate as a consequence of cytoskeletal perturbations (*Hertz-Fowler et al., 2001; Moreira-Leite et al., 2001; Davidge et al., 2006; Baines and Gull, 2008*), does the subpellicular microtubule corset lose its typical organisation in intact cells. This variation in phenotypes points to

different functional roles for CAP5.5 and WCB. In the absence of membrane “blebbing”, the examples of microtubule bundling beneath the plasma membrane and aberrant inter-microtubule spacing described here more likely reflect a requirement for CAP5.5 to either break or form appropriate inter-microtubule cross-links, than to represent difficulties in cross-linked microtubules maintaining an association with the inner-leaflet of the plasma membrane. The relatively small number of sections through RNAi-induced cells (~15%) exhibiting microtubule abnormalities could reflect the relatively small area of the cytoskeleton within which defects in microtubule organisation can occur. For example, CAP5.5 is not so readily lost from the anterior region of the subpellicular microtubule cytoskeleton. Alternatively, these data can also be explained as a consequence of localised accumulation of numerous microtubule abnormalities at relatively few points on the cytoskeleton, with many more isolated examples of cytoskeletal abnormalities beneath the level of detection using conventional electron microscopy.

RNAi induction against CAP5.5V, but not CAP5.5 in bloodstream form trypanosomes yielded growth and morphology phenotypes highly similar to those obtained by the RNAi knockdown of CAP5.5 in procyclic forms. CAP5.5V RNAi-induced cultures of bloodstream trypanosomes showed a marked reduction in their growth rate after only ~6-8h of induction (Fig. 4.41), which coincided with the onset of obvious morphological difficulties (Fig. 4.42). Considering the difference between the cell-cycle times of procyclic (~11h) and bloodstream (~8h) cell lines, the kinetics of phenotype onset, relative to cell-cycle time, were similarly fast for CAP5.5 RNAi in procyclic cells and CAP5.5V RNAi in bloodstream trypanosomes. CAP5.5V RNAi induced cultures also showed the accumulation of zoids and multinucleate cells (Fig. 4.42), a significant result given that the patterns of kinetoplast and nuclei positioning prior to cell division are distinct in procyclic and bloodstream trypomastigotes. This is in contrast to

procyclic trypomastigotes where one daughter kinetoplast sits between the divided nuclei. However, in bloodstream trypomastigotes the replicated and segregated kinetoplasts lie posterior to the divided nuclei, being subject to a more limited movement during segregation.

BSF CAP5.5V RNAi-induced cultures showed a slight increase in cell density beyond 12-15h post-induction, which was consistent with a steady increase in the zoid population that reached a 25-30% of the “cell” types present (Fig. 4.41 and 4.42), probably as a result of abscission from multinucleate cells (Fig. 4.52). Defects in microtubule organisation similar to those found in procyclic CAP5.5 RNAi mutants were also observed in CAP5.5V RNAi-induced bloodstream trypanosomes (Fig. 4.54-4.57). Thus, I have shown that CAP5.5V is essential for the correct morphogenesis and cytokinesis of bloodstream trypanosomes.

The determination of whether CAP5.5 and/or CAP5.5V are bona fide proteases is crucial for the further understanding of the role(s) played by these proteins in trypanosome microtubule biology. For instance, the observation that tubulins and various MAPs provide *in vitro* substrates for “classic” mammalian calpains (Goll *et al.*, 2003) could indicate that the aberrant cytoskeletal organisation arises because the ability of new microtubules to invade the existing subpellicular array is lost following CAP5.5 RNAi. However, the bioinformatics of *T. brucei* calpain-related family argues against this possibility, as CAP5.5 and CAP5.5V both lack the C-H-N catalytic triad that characterises the cysteine protease family. Instead they contain a S-Y-N catalytic triad. Thus, it is unlikely that CAP5.5 and CAP5.5V are proteolytically active. However, there are two alternative scenarios. Firstly, the possibility that Ca^{2+} -dependent conformational changes may influence the proteolytic activity of calpain-like proteins, as it does in typical calpains and secondly, the recent description of an active cysteine protease family member with an atypical catalytic triad (Hsu *et al.*, 2008). Although, without

further data it is hard to envisage how cleavage of a peptide bond could arise from either a Ser₂₇₄-Tyr₄₃₇ or Ser₂₉₁-Tyr₄₅₄ interaction within the degenerated catalytic triads of CAP5.5 and CAP5.5V, respectively.

Additionally, the loss of catalytic activity may have also been an important step in the functional evolution of these cytoskeletal proteins. In mammals, Capn6 provides an example of an atypical calpain that lacks the essential Ca²⁺-sensitive regulatory domain IV and a canonical catalytic triad (Tonami *et al.*, 2007). Capn6 binds microtubules, and its over-expression promotes formation of stable cytoplasmic microtubule bundles in fibroblasts. Moreover, Capn6 RNAi phenotypes consist of actin reorganisation and membrane ruffling. Thus, the available data suggest that Capn6 is a microtubule-stabilising protein involved in the regulation of cytoskeletal organisation (Tonami *et al.*, 2007). Considering the dynamic association of some conventional calpains with the cytoskeleton (Gil-Parrado *et al.*, 2003), the function of the atypical Capn6 in stabilising cytoskeletal organisation represents an innovative and logical adaptation. However, the *in vitro* cleavage of tubulin and other MAP substrates by mammalian calpains (Goll *et al.*, 2003) suggests that loss of proteolytic activity would have been a necessary early step in the evolution of capn6 as a microtubule-stabilising protein. A similar loss in proteolytic activity may equally have been a necessary step in the evolution of CAP5.5 function.

The observations described in Chapter 4 provide a strong indication that *T. brucei* paralogous calpain-related proteins perform analogous roles in procyclic and bloodstream trypomastigotes. In addition, these data identify a second example of differentially expressed paralogous genes encoding cytoskeletal-associated proteins. The first paralogous genes found were CAP17 and CAP15 (Vedrenne *et al.*, 2002). Additionally, in Chapter 5, the RNAi phenotypes of different group 4 calpain-like sequences, which localise at the FAZ filament, suggest that these cytoskeletal-

associated proteins might also be subject to stage-specific regulation. It will be interesting to learn how many other cytoskeletal components are subject to stage-specific regulation in African trypanosomes and the reason(s) for this stage-specific expression. In the case of CAP5.5 and CAP5.5V, one possibility is that the differential expression of paralogous genes provides a mechanism to fine tune absolute protein requirements in two morphologically distinct parasite forms.

CAP5.5V lies within the 0.5 Mb region of *T. brucei* chromosome 4 that was duplicated and now forms a sub-telomeric arm on chromosome 8 (El-Sayed *et al.*, 2005; Jackson, 2007b). Within the African trypanosome family (i.e. including the livestock pathogens *T. congolense* and *T. vivax*), the duplication of over 70 genes from chromosome 4 is specific to *T. brucei*, and represents a relatively recent evolutionary event. Among the physiological consequences proposed for this duplication event is the introduction of functional novelty in response to selective pressures (Jackson, 2007b). The stage specificity and essential function of CAP5.5 and CAP5.5V paralogous proteins provide strong evidence that at least some retained paralogous genes after *T. brucei* large chromosome duplication event were functionally innovative: neo- or subfunctionalisation of conserved duplicated genes. Comparison of CAP5.5 and CAP5.5V genome loci to other trypanosomatid species by reciprocal BLASTp illustrates *T. brucei* specificity of chromosome 4 duplication (Fig. 3.2). All other trypanosomatid species only have one CAP5.5/CAP5.5V ortholog (Table 3.3): *L. major* LmjF31.0440 and *L. infantum* LinJ31_v3.0450. A possible explanation for this discrepancy between closely related organisms could be that CAP5.5/CAP5.5V orthologs in other trypanosomatid species are constitutively expressed throughout their entire lifecycles, carrying out equivalent functions to CAP5.5 and CAP5.5V in cytoskeletal remodelling. In support of this hypothesis are the published expression profiles of *Leishmania* species developmental forms, which indicate that LmjF31.0440

is a constitutively expressed gene (Leifso *et al.*, 2007; Rochette *et al.*, 2008). In this context, it would be interesting to obtain the expression profile of CAP5.5/CAP5.5V orthologs in the closely related trypanosomatids *T. congolense* and *T. vivax*.

Additionally, although CAP5.5 up-regulation during late bloodstream-procyclic trypomastigote differentiation has long been known (Matthews and Gull, 1994a), CAP5.5 down-regulation and switch to CAP5.5V expression remain to be elucidated. CAP5.5 up-regulation takes place in the not-proliferative stumpy bloodstream form, after only 4-6h under differentiation conditions (Matthews and Gull, 1994a). Thus, CAP5.5 to CAP5.5V expression switch would likely take place in a similar scenario; this is, shortly after the transmission of cell-cycle arrested metacyclic trypomastigotes from the tsetse fly salivary glands to the bloodstream of the mammalian host (Introduction chapter, Fig. 1.1). However, it could also be possible that CAP5.5 to CAP5.5V expression switch occurs earlier on in the tsetse fly —during the differentiation of procyclic trypomastigotes to free epimastigotes, after the attachment of epimastigotes to the microvilli of the salivary glands epithelial cells or after differentiation of attached epimastigotes to intermediate premetacyclics trypomastigotes. To positively establish when CAP5.5 to CAP5.5V expression switch takes place, it will be interesting to perform CAP5.5 immuno-detection studies, using the anti-CAP5.5 mAb, on the different tsetse fly developmental forms.

6.2. Calpain-like proteins are essential for the correct structure and length of the FAZ.

The FAZ complex runs alongside the longitudinal axis of the cell, from a position close to the basal bodies to the anterior cell-end. As the microtubule corset remains intact during the cell and lifecycles, it requires a very precise coordination between its duplication and the duplication and spatial positioning of trypanosome single copy

organelles (*Sherwin and Gull, 1989b*), in which the FAZ has been shown to have a central role (*Kohl et al., 1999; Davidge et al., 2006; Vaughan et al., 2008*). During the cell division-cycle of procyclic trypomastigotes the position of the anterior nucleus with respect to the anterior cell-end and the anterior kinetoplast/basal body remains relatively constant, whereas the mitotic spindle straddles the axis of the old flagellum. An anchoring system that maintains the position of the anterior nucleus whereas the posterior basal body and nucleus move away in concert with cell elongation, would explain the spatial positioning of these organelles previous to cytokinesis (*Robinson et al., 1995*). However, although the presence of nuclear remnants in detergent extracted cytoskeleton suggests the existence of some form of linkage between the nucleus and the cytoskeleton (*Sherwin and Gull, 1989a*), there is little more evidence of the presence of such an anchoring system.

Additionally, during PCF cell division-cycle there is an overall increase in cell-body length of $\sim 6 \mu\text{m}$ (*Sherwin and Gull, 1989a*), which is thought to be due to the increase in inter-basal body distance, as all other parameters remain relatively constant (*Robinson et al., 1995*). Consequently, the growth of PCF trypanosomes depends on the separation of the basal bodies or, more specifically, on the movement of the new basal body towards the posterior cell end. The production of anucleate zoids from cells with impaired mitosis or defects in cytoskeletal growth and remodelling indicates that the inclusion of a segregated kinetoplast/basal body is more influential in determining the position of the cleavage furrow than a daughter nucleus. Moreover, in both zoids and tsetse fly epimastigotes of different sizes, the length of the cell-body is always proportional to the length of the flagellum (*Robinson et al., 1995; Sharma et al., 2008*). During cell division, the new FAZ filament is located between the new and old flagella and its growth could be responsible for the separation of the basal bodies, providing the structural correlation between the length of the flagellum and the growth of the

cell-body and determining the position and direction of the cleavage furrow (*Robinson et al., 1995; Davidge et al., 2006*).

Trypanosomatid group 4 calpain-like sequences are high molecular weight proteins containing 2-3 copies of the calpain “protease” domain II, 1-2 copies of the C2-like calpain domain III and long sequences of 67-68 amino-acid-tandem repeats. Short sequences containing the repeats of some trypanosomatid group 4 calpain-like sequences have been published as antigens FRA, GM6 and Lcr1, in *T. cruzi*, *T. brucei* and *L. chagasi*, respectively (*Lafaille et al., 1989; Muller et al., 1992; Wilson et al., 1995*). Of these, FRA and GM6 have been shown to localise at the FAZ filament (*Cotrim et al., 1990; Muller et al., 1992; Cotrim et al., 1995*). Additionally, Lcr1 containing protein in *L. major*, LmjF27.0500, is the ortholog of *T. cruzi* FRA containing sequences (TcCALP985.9 and TcCALP441.10; Appendix 3, Reciprocal_BLAST), suggesting that it may also localise at the FAZ filament. The group 4 of *T. brucei* calpain-related proteins currently comprises 4 ORFs located in 3 different contigs of chromosome 11, in adjacent positions of its provisional assembly (Chapter 3, Fig. 3.10). However, the highly conserved repeat units shared by pairs of these sequences—which, due to their similarity to GM6 and Lcr1 repeats units, I have named GM6-like and Lcr1-like repeats—and their lack of start and/or stop codons have not allowed to positively distinguish between real ORFs and sequence fragments. The two putative proteins containing GM6-like repeats are Tb11.47.0036 and Tb11.57.0008 and those containing Lcr1-like repeats are Tb11.47.0035 and Tb11.v4.0001.

The RNAi functional analysis of *T. brucei* group 4 calpain-like proteins presented in Chapter 5 suggests an essential role for these proteins in the correct morphogenesis of PCF and BSF trypanosomes. In PCF, induction of RNAi cell lines obtained using the HP-N and DS-N2 constructs caused aberrant morphogenesis in dividing trypomastigotes, giving rise to daughter cells with mispositioning of the kinetoplast and

flagellar pocket towards the mid cell-body or anterior cell-end. These abnormal cells resemble the epimastigote forms present in the tsetse fly (Chapter 5: Fig. 5.3, 5.13-5.17). Induced cells also have long flagellar overhangs, although the total length of the flagellum remained relatively constant (Chapter 5, Fig 5.12 and Fig. 5.21-5.24). Induced cells acquired the epimastigote-like morphology through an increase in the distance between the kinetoplast and the posterior cell-end (Chapter 5, Fig. 5.13), whereas the distance between the nucleus and the posterior end remains relatively constant (~45% of the cell-body length throughout induction, Appendix 5, Fig. A.5.1), the opposite of what happens in the tsetse fly (*Sharma et al., 2008*).

Immunofluorescence staining of the FAZ filament with monoclonal antibodies against some of its components (L3B2 that recognise FAZ1 protein, and CD10 and DOT1 that label high molecular weight proteins of unknown identity) showed that RNAi induced cells have a shorter FAZ filament than wild type procyclic trypomastigotes. This shorter FAZ filament extended from the mispositioned flagellar pocket to the anterior-end of the cells. The reduced length of the FAZ in induced cells could alter the separation of the basal bodies, which is responsible for kinetoplast segregation (*Robinson and Gull, 1991; Robinson et al., 1995*), to which they are linked through the tripartite attachment complex (*Ogbadoyi et al., 2003*). Therefore, the short FAZ length would result in a reduction of the distance between the segregated kinetoplasts, causing the mispositioning of the new kinetoplast and flagella pocket towards the anterior cell-end observed after induction. The effect of group 4 calpain-like proteins RNAi silencing in the length of the FAZ supports the localisation of these proteins at the FAZ filament indicated by the anti-GM6 and FRA crude antisera (*Cotrim et al., 1990; Muller et al., 1992*). From 48h of induction onwards, many epimastigote-like cells containing a very short FAZ showed a widening of the area labelled by the anti-FAZ antibodies, suggesting that depletion of group 4 calpain-like proteins caused an alteration in the

structure of the FAZ filament (Fig. 5.18 and 5.19). This was confirmed by TEM analysis of the FAZ in 48h induced cells, which showed the presence of abnormalities in the FAZ ultrastructure (Fig 5.25). However, the absence of flagellar detachment in induced cells indicates that the FAZ, although shorter and with an altered structure, maintained its flagellum-cell-body attachment functionality. This differs significantly from other FAZ mutants published so far, all of which cause flagellar detachment (*LaCount et al., 2002; Hammarton et al., 2007; Li and Wang, 2008; Vaughan et al., 2008*). Immunofluorescence analysis of the FAZ in induced cells also showed that silencing of group 4 calpain-like proteins did not prevent the localisation of other components of the FAZ, as FLA1 and FAZ1.

Consequently, the function of group 4 calpain-like proteins within the FAZ filament must be one involved in determining its length and structure. The hypothesis that the FAZ determines the separation of the basal bodies and the position of the cleavage furrow (*Robinson et al., 1995; Davidge et al., 2006*) is supported by the concurrency of the FAZ shortening with the mispositioning of the kinetoplast and flagellar pocket towards the anterior end in induced cells. That is, this hypothesis could explain the appearance of epimastigote-like cells after induction of group 4 calpain-like sequences RNAi. A short FAZ that does not drive the basal bodies/ kinetoplasts sufficiently apart could explicate the increase in the distance between the kinetoplast and the posterior cell-end, the flagellar pocket mispositioning towards the anterior cell-end and the reduction in cell-body length observed in induced cells. Thus, group 4 calpain-like proteins are involved in determining the structure and length of the FAZ, which in turn is involved in the morphogenetic processes that determine cell morphology in the different lifecycle stages of *T. brucei*.

Additionally, preliminary results of the RNAi functional analysis of group 4 calpain-like sequences in bloodstream trypomastigotes revealed that Tb11.47.0036/Tb11.57.0008

are essential for the viability of BSF (Fig. 5.26). However, Tb11.47.0035/Tb11.v4.0001 RNAi silencing did not produce any discernible phenotype in bloodstream trypomastigotes (Appendix 5, Fig. A.5.3). As the RNAi phenotype in procyclic forms was obtained using an RNAi construct with the capability to knockdown the expression of Tb11.57.0008 and Tb11.47.0035, provided they were simultaneously expressed, the appearance of a phenotype in bloodstream forms after the RNAi silencing of Tb11.57.0008 but not of Tb11.47.0035 suggests that the two pairs of sequences are subject to stage-specific regulation. That is, Tb11.57.0008/Tb11.47.0036 may be preferentially expressed in bloodstream trypanosomes, whereas Tb11.47.0035/Tb11.v4.0001 may be upregulated in procyclic form trypomastigotes. Although the complex domain composition and architecture of repeat-containing group 4 calpain-like sequences hinder their expression analysis by Northern blot or quantitative RT-PCR, and the expression profiles, by hybridisation (Diehl *et al.*, 2002; Brems *et al.*, 2005; Jensen *et al.*, 2009; Kabani *et al.*, 2009) and digital methods (Veitch *et al.*, 2010) published so far show contradictory results, the RNAi phenotypes obtained in procyclic and bloodstream forms indicate that group 4 proteins are stage-specific regulated.

The stage-specific expression of different members of *T. brucei* large calpain-like family in two of its multiple life-cycle developmental forms —the only ones that can be cultured *in vitro*— and the appearance of epimastigote-like cells after the RNAi knockdown of some of these proteins (group 4) in PCF, raise the possibility that the appearance of *T. brucei* developmental forms may be regulated by changing the expression of relatively few genes. In this context, the unusual expansion of trypanosomatids calpain-like family could correlate with these organisms multiple developmental forms, whose regulation must be essential for progression through their complex life-cycles.

6.3. Bioinformatic analysis of trypanosomatid calpain-like sequences: a tool for functional prediction.

A typical feature of the gene organisation of trypanosomatids is the presence of syntenic homologous regions across species (*Bringaud et al., 1998*), despite the numerous biological and genetic differences that characterise each species. Syntenic regions most probably represent genes whose development was complete in trypanosomatids common ancestor, prior to speciation, whereas species-specific genes likely reflect the evolutionary constraints undergone by different species. For instance, the dissimilar, niche adapted developmental forms into which *T. brucei* and *L. major* differentiate during their complex life-cycles (Chapter 1, Fig. 1.1), possibly relate to the distinct evolutionary pressures experienced by each parasites. I have earlier hypothesised that the unusually expanded calpain-like family of trypanosomatids could relate to the multiple developmental forms undergone by these organisms. In this context, *T. brucei* and *Leishmania* calpain-like species-specific genes could relate to the different developmental forms undergone by each parasite, whereas orthologous and paralogous genes could reflect common biology. Trypanosomatids progression through the life-cycle requires biochemical and morphological adaptations for each developmental form, achieved by differential or stage-specific gene expression, which is regulated at a post-translational level.

The presence of syntenic homologous calpain-like genes in *T. brucei* and *Leishmania* spp. provided by the bioinformatic analysis presented in chapter 3 —which identified species-specific, orthologous and paralogous calpain-like proteins—, combined with the currently available data of these proteins expression profile, cellular localisation and function, in one given trypanosomatid species, could be used to predict the function of its orthologs in other kinetoplastid parasites. In Table 6.2, I summarise these combined data. As I have discussed earlier, the functional analysis of *T. brucei* CAP5.5 and

Table 6.2.

<i>T. brucei</i> Gene ID	Gr.	<i>L. major</i> syntenic homologs	<i>T. brucei</i> protein localisation	<i>T. brucei</i> expression profile
Tb927.1.2100	1	LmjF20.1180*	Unknown	*Metacyclic promastigote PCF & BSF/
		LmjF20.1185		
Tb927.1.2110	1	LmjF20.1190	Unknown	PCF/ Promastigote
Tb927.1.2120 ¹	1	X	Flagellar Tip	PCF & BSF
Tb927.4.3940	1	LmjF31.0390	Cell-body (Membrane?)	PCF & BSF
Tb927.4.3950 CAP5.5 ¹	1	LmjF31.0440	Cytoskeleton	PCF
Tb927.8.8330 CAP5.5V	1		Cytoskeleton	BSF
Tb09.160.5550	1	LmjF04.0450	Unknown	PCF & BSF
Tb11.01.6870	1	LmjF32.2010	Unknown	Unknown
		LmjF20.1210		
Tb927.6.3310	2	LmjF30.2040	Cell-body (Membrane?)	PCF & BSF
Tb10.389.0470	2	LmjF18.1060	Unknown	BSF>PCF
Tb10.70.5950	2	LmjF21.0120	Unknown	BSF>PCF
Tb11.01.5800	2	LmjF32.0970	Unknown	Unknown
Tb11.02.2240	2	LmjF33.2010	Unknown	Unknown
Tb927.1.2150	3a	LmjF20.1220	Cell-body (Membrane?)	PCF & BSF
Tb927.1.2160	3a	LmjF20.1230	Flagellar Membrane	PCF & BSF
Tb927.1.2230	3a	LmjF20.1310 SMP-1 ¹	Cell-body Membrane	PCF/ Promastigote
Tb927.1.2260	3a		Flagellar Membrane	PCF & BSF
Tb927.7.4060	3a	LmjF14.0850	Cell-body (Cytosol)	PCF & BSF
Tb927.7.4070	3a		Cell-body & Flagellar Tip	PCF
Tb10.70.5060	3b	LmjF34.0190	Unknown	BSF>PCF
		LmjF34.0200		
Tb927.5.2190	3c	LmjF17.0870	Plasma Membrane	BSF/ Metacyclic promastig
X	3d	LmjF36.3180	Unknown	Unknown
Tb11.47.0035/ Tb11.v4.0001	4	LmjF27.0500	FAZ	PCF
		LmjF27.0510		
Tb11.47.0036/ Tb11.57.0008	4	LmjF27.0490	FAZ	BSF
Tb927.5.2410	5	LmjF17.1110	Unknown	Unknown

Table 6.2. Summary of *T. brucei* calpain-like sequences and their corresponding *L. major* syntenic homologs, with their currently known cellular localisation and/or expression profiles. Blue font indicates *L. major* expression; (*) indicates the expression profile of one specific paralog; (¹) indicates the proteins where single or double N-terminal acylation have been experimentally demonstrated.

CAP5.5 strongly indicates that paralogous, differentially expressed genes provide analogous roles in cytoskeletal remodelling, essential for the two life-cycle stages in which they are expressed. However, *L. major* only has one CAP5.5/CAP5.5V ortholog, LmjF31.0440, which seems to be constitutively expressed throughout its life-cycle (Leifso et al., 2007; Rochette et al., 2008; Depledge et al., 2009), possibly performing an analogous role to CAP5.5/CAP5.5V in cytoskeletal remodelling.

In the same context, the data from *T. brucei* group 4 calpain-like proteins bioinformatic and functional analysis could be useful to predict the function of their homologs in *L. major* and *L. infantum*. Both, *L. major* and *L. infantum* have three group 4 calpain-like proteins, located in adjacent positions, on a unique contig of their chromosome 27 (Fig. 3.10). All of them have start and stop codons and seem to be encoded by real ORFs. Interestingly, one of the two orthologs of Tb11.47.0035 in *L. major*, LmjF27.0510, has been reported to be up-regulated in the promastigote form in the host. In *L. infantum* both Tb11.47.0035 orthologs, LinJ27.0510 and LinJ27_v3.2490, are up-regulated in promastigotes (Leifso et al., 2007; Rochette et al., 2008; Depledge et al., 2009). This is significant as the promastigote is the only *Leishmania* developmental form with a free flagellum and a short FAZ. The FAZ localisation of *T. brucei* group 4 calpain-like proteins and their essential function in cell morphogenesis suggest that their *Leishmania* orthologs could also have an important role in the flagellate form where they are up-regulated, maybe determining the position of the cleavage furrow.

Additionally, RNAi silencing of *T. brucei* Tb927.1.2110, a group 1 calpain-like protein results in a severe growth phenotype in bloodstream trypanosomes (Subramaniam et al., 2006). Interestingly, Tb927.1.2110 orthologs in *T. cruzi* and *L. major* have been also shown to be up-regulated in the host forms —epimastigotes under stress conditions in *T. cruzi* (Giese et al., 2008) and promastigotes in *L. major*. (Saxena et al.,

2003). These data suggest an important function for these proteins in the pathogenic form of these parasites and highlight them as potential therapeutic targets.

The data from the sequence analysis of trypanosomatid calpain-like proteins can also be combined with their known cellular localisation to infer which sequence elements are involved in targeting these proteins to specific cellular locations. For instance, many kinetoplastid calpain-like proteins from groups 1 and 3a contain an N-terminal single or double acylation modification motif (*Hertz-Fowler et al., 2001; Tull et al., 2004; Ersfeld et al., 2005*). Acylation of proteins with myristic and palmitic acid has been shown to mediate protein association to membranes (*Resh, 1999*). Myristoylation increases the hydrophobicity of proteins, targeting them to the cell's hydrophobic lipid domains and is a prerequisite for palmitoylation, which then establishes high affinity interactions with membranes (*Resh, 1999*). Some kinetoplastid calpain-like proteins have been experimentally demonstrated to be myristoylated and/or palmitoylated at the N-terminus —*L. major* small myristoylated protein SMP-1 (LmjF20.1310), *T. brucei* CAP5.5 (Tb927.4.3950) and Tb927.1.2120 (*Hertz-Fowler et al., 2001; Tull et al., 2004; Liu et al., 2009*). However, these proteins have different cellular localisations: SMP-1 is exclusively present at the flagellar membrane, CAP5.5 localises at the subpellicular microtubule corset but is absent from the flagellum, and Tb927.1.2120 is exclusively found at the flagellar tip. Dual acylation has been previously shown to be essential for the flagellar localisation of *L. major* SMP-1 (*Tull et al., 2004*). Moreover, flagellar targeting has been proposed to be the default pathway for dually acylated proteins in kinetoplastids (*Mills et al., 2007; Liu et al., 2009*). However, the different cellular locations of double acylated proteins indicate that N-terminal acylation cannot be the only factor that determines protein targeting to the flagellar membrane. Trypanosome cell membranes have three domains with different composition and chemistry: the cell-body, the flagellar and the flagellar pocket membranes.

The flagellar membrane contains high concentrations of lipid rafts —discrete regions rich in sterols and sphingolipids that serve to sequester and organise specialised protein complexes. In detergent-extracted trypanosomes, lipid rafts are visualised as discrete membrane patches on the surface of the flagellar axoneme (*Tyler et al., 2009*). Interestingly, DRMs in higher eukaryotes have been shown to be implicated in the sorting of proteins to the secretory pathway, functioning as platforms for the attachment of proteins when membranes are moved around inside the cell (*Simons and Ikonen, 1997*). Protein modification by saturated acyl groups has been shown to target proteins to the DRM rafts in mammalian cells (*Melkonian et al., 1999*). However, the targeting of proteins to the DRM of the flagellum of trypanosomes is still poorly understood.

As acylation is most likely involved in relatively unspecific association with cellular membranes (*Resh, 1999*), factors such as pH, phosphorylation, GTP-binding or oligomerisation have been shown to also regulate membrane localisation (*Tang et al., 2004*). In this context, CAP5.5 retention in the cell-body or SMP-1 selective transport to the flagellum must be regulated by other motifs within their sequences. CAP5.5 sequence contains a C2-like domain III that is thought to mediate interactions with membrane phospholipids, whereas SMP-1 lacks a domain III. The presence of the C2-like phospholipid-binding domain could modify CAP5.5 hydrophobicity, resulting in a different membrane association from SMP-1. However, CAP5.5 interaction with the plasma membrane has not yet been demonstrated and its localisation to the cell-body could be exclusively mediated by its cytoskeletal association. Additionally, SMP-1 localisation to the flagellar membrane has been shown to be mainly due to hydrophobic interactions, being not dependant on associations to the flagellar cytoskeleton or IFT mechanisms (*Ralton et al., 2002; Tull et al., 2004*). Interestingly, the recently demonstrated localization of *T. brucei* calpain-like group 1 protein Tb927.1.2120 to the flagellar tip has been shown to require, in addition to its double acylation motif, the

presence of the last 35 residues of its C-terminal domain (*Liu et al., 2009*). Although this C-terminal sequence—including the putative flagellar localization signal— does not have any other recognisable motifs, it contains an accumulation of basic amino acids (*Liu et al., 2009*). Basic residues are known to play a critical role in the attachment of protein domains to membrane interfaces through not-specific electrostatic interactions (*Hurley and Misra, 2000*). Therefore, it would not be unreasonable to hypothesise that these kind of electrostatic interactions could be the modulating factor that determines the targeting of N-terminal acylated proteins to the different domains of kinetoplastid membranes.

6.5 Comparison between different RNAi approaches.

In this section I discuss the direct comparison between two different RNAi approaches most widely used in *T. brucei*: hairpin and double promoters RNAi constructs. The comparison was carried out through an innovative third approach, in which inverted repeats of the target gene are cloned between two head-to-head T7 promoters (Fig. 5.1). In order to perform this direct comparison, RNAi experiments using the new approach and the conventional head-to-head promoters approach were carried out using the same sequence fragment of the target gene and the same RNAi vector, the p2T7-177 plasmid (*Wickstead et al., 2002*). In this way, I tried to exclude the variability that would be otherwise introduced by using two different constructs. The p2T7-177 RNAi vector has two opposing T7 promoters and contains 177 bp repeats that target its integration to *T. brucei* minichromosomes (MCs), which have been shown to yield a good down-regulation of inducible vectors in the absence of induction (*Wickstead et al., 2002*). Thus, single regions and inverted repeats of the target gene were both cloned between the opposing T7 promoters of the p2T7-177 plasmid. Upon induction of the inverted repeats RNAi construct, the resulting complementary RNA molecules could have different hypothetical fates. They could fold into themselves forming stem-loop,

intra-molecular dsRNA or could anneal to each other, forming either conventional dsRNA or inter-molecular dsRNA stem-loops. In the first case, hairpin constructs with opposing promoters could produce the double amount of dsRNA than a no hairpin construct.

The phenotype analysis and comparison of the different RNAi cell lines obtained by the two approaches showed that the appearance of RNAi phenotypes was faster and stronger in the cell lines obtained with hairpin constructs than in those obtained with the normal opposing promoters constructs (Fig. 5.2). This result adds weight to the evidence that stem-loop RNAi constructs gives more potent RNAi effects, although the higher potency of the hairpin constructs used here could also be due to the higher amount of dsRNA produced from opposing promoters flanking inverted repeats. The hairpin RNAi cell lines obtained with this type of construct were “leakier” in the not-induced state than those obtained with the single fragment approach, possibly due to the need to regulate the two opposing promoters combined with the higher stability of stem-loop dsRNA. Additionally, the RNAi cell lines obtained with the hairpin constructs started to lose the RNAi phenotype after 96h of induction whereas the RNAi cell line obtained from a single DNA fragment flanked by opposing promoters maintained the RNAi phenotype for at least 7 days after induction. A possible explanation for the loss of the RNAi phenotype by the hairpin construct could be its production of higher amounts of dsRNA resulting in lower RNAi specificity. This deleterious effect could provoke enough selective pressure to favour the survival of cells that have somehow been able to inhibit the effects of the RNAi machinery.

6.6. Concluding Remarks and Perspective.

The high numbers of calpain-like proteins in kinetoplastids, outnumbering those found in mammals, points towards important functions of these proteins in trypanosomatid parasites. Although the majority of these sequences have deviations from the catalytic triad of typical cysteine proteases and may not be proteolytically active, I have shown in this thesis that some of them have important functions, being essential for the correct morphogenesis and cytoskeletal remodelling of procyclic and bloodstream *T. brucei*. Future challenges would be to determine if any of these proteins have protease activity and to identify their binding partners. The essential role of some of these proteins in bloodstream trypanosomes highlights them as potential drug targets for the treatment of sleeping sickness.

7. REFERENCES.

- Abrahamsen, M. S., T. J. Templeton, S. Enomoto, J. E. Abrahante, G. Zhu, C. A. Lancto, M. Deng, C. Liu, G. Widmer, S. Tzipori, G. A. Buck, P. Xu, A. T. Bankier, P. H. Dear, B. A. Konfortov, H. F. Spriggs, L. Iyer, V. Anantharaman, L. Aravind and V. Kapur,** (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science (New York, N.Y.). 304: 441-5.
- Aksoy, S., W. C. Gibson and M. J. Lehane,** (2003). Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. Advances in parasitology. 53: 1-83.
- Alexandre, S., P. Paindavoine, P. Tebabi, A. Pays, S. Halleux, M. Steinert and E. Pays,** (1990). Differential expression of a family of putative adenylate/guanylate cyclase genes in *Trypanosoma brucei*. Molecular and biochemical parasitology. 43: 279-88.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman,** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25: 3389-402.
- Andresen, K., T. D. Tom and M. Strand,** (1991). Characterization of cDNA clones encoding a novel calcium-activated neutral proteinase from *Schistosoma mansoni*. The Journal of biological chemistry. 266: 15085-90.
- Aragon, B., S. Poussard, S. Dulong, K. Touyarot, E. Dargelos, J. J. Brustis, D. Leveux, A. Ducastaing and P. Cottin,** (2002). Protein kinase Calpha is a calpain target in cultured embryonic muscle cells. Molecular and cellular biochemistry. 231: 97-106.
- Ariyoshi, H., E. Shiba, J. Kambayashi, M. Sakon, T. Kawasaki, K. Yoshida and T. Mori,** (1993). Stimulation of human platelet Ca(2+)-ATPase and Ca2+ restoration by calpain. Cell Calcium. 14: 455-63.
- Azarian, S. M., A. J. King, M. A. Hallett and D. S. Williams,** (1995). Selective proteolysis of arrestin by calpain. Molecular characteristics and its effect on rhodopsin dephosphorylation. The Journal of biological chemistry. 270: 24375-84.
- Baines, A. and K. Gull,** (2008). WCB is a C2 domain protein defining the plasma membrane - sub-pellicular microtubule corset of kinetoplastid parasites. Protist. 159: 115-25.
- Barnes, T. M. and J. Hodgkin,** (1996). The tra-3 sex determination gene of *Caenorhabditis elegans* encodes a member of the calpain regulatory protease family. The EMBO journal. 15: 4477-84.
- Barrett, M. P.,** (2006). The rise and fall of sleeping sickness. Lancet. 367: 1377-8.
- Bastin, P., K. Ellis, L. Kohl and K. Gull,** (2000). Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. Journal of cell science. 113 (Pt 18): 3321-8.

Bastin, P., T. H. MacRae, S. B. Francis, K. R. Matthews and K. Gull, (1999b). Flagellar morphogenesis: protein targeting and assembly in the paraflagellar rod of trypanosomes. *Molecular and cellular biology*. 19: 8191-200.

Bastin, P., K. R. Matthews and K. Gull, (1996). The paraflagellar rod of kinetoplastida: solved and unsolved questions. *Parasitology today (Personal ed.)*. 12: 302-7.

Bastin, P., T. J. Pullen, T. Sherwin and K. Gull, (1999a). Protein transport and flagellum assembly dynamics revealed by analysis of the paralysed trypanosome mutant *snl-1*. *Journal of cell science*. 112 (Pt 21): 3769-77.

Bastin, P., T. Sherwin and K. Gull, (1998). Paraflagellar rod is vital for trypanosome motility. *Nature*. 391: 548.

Baulcombe, D., (2004). RNA silencing in plants. *Nature*. 431: 356-63.

Baum, S. G., M. Wittner, J. P. Nadler, S. B. Horwitz, J. E. Dennis, P. B. Schiff and H. B. Tanowitz, (1981). Taxol, a microtubule stabilizing agent, blocks the replication of *Trypanosoma cruzi*. *Proceedings of the National Academy of Sciences of the United States of America*. 78: 4571-5.

Bellofatto, V. and G. A. Cross, (1989). Expression of a bacterial gene in a trypanosomatid protozoan. *Science (New York, N.Y.)*. 244: 1167-9.

Ben-Aharon, I., P. R. Brown, R. Shalgi and E. M. Eddy, (2006). Calpain 11 is unique to mouse spermatogenic cells. *Molecular reproduction and development*. 73: 767-73.

Benz, C., D. Nilsson, B. Andersson, C. Clayton and D. L. Guilbride, (2005). Messenger RNA processing sites in *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 143: 125-34.

Bergen, L. G. and G. G. Borisy, (1980). Head-to-tail polymerization of microtubules in vitro. Electron microscope analysis of seeded assembly. *The Journal of cell biology*. 84: 141-50.

Bernstein, E., A. A. Caudy, S. M. Hammond and G. J. Hannon, (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. 409: 363-6.

Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D. C. Bartholomeu, N. J. Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Bohme, L. Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T. J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P. J. Mooney, S. Moule, D. M. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C. S. Peacock, J. Peterson, M. A. Quail, E. Rabinowitsch, M. A. Rajandream, C. Reitter, S. L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. Turner, A. Tait, A. R. Tivey, S. Van

- Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, S. E. Melville and N. M. El-Sayed, (2005).** The genome of the African trypanosome *Trypanosoma brucei*. *Science* (New York, N.Y.). 309: 416-22.
- Beverley, S. M. and C. E. Clayton, (1993).** Transfection of *Leishmania* and *Trypanosoma brucei* by electroporation. *Methods in molecular biology* (Clifton, N.J.). 21: 333-48.
- Beyette, J. R., Y. Emori and D. L. Mykles, (1997).** Immunological analysis of two calpain-like Ca²⁺-dependent proteinases from lobster striated muscles: relationship to mammalian and *Drosophila* calpains. *Archives of biochemistry and biophysics*. 337: 232-8.
- Bhushan, S., A. Stahl, S. Nilsson, B. Lefebvre, M. Seki, C. Roth, D. McWilliam, S. J. Wright, D. A. Liberles, K. Shinozaki, B. D. Bruce, M. Boutry and E. Glaser, (2005).** Catalysis, subcellular localization, expression and evolution of the targeting peptides degrading protease, AtPreP2. *Plant & cell physiology*. 46: 985-96.
- Bienen, E. J., M. Saric, G. Pollakis, R. W. Grady and A. B. Clarkson, Jr., (1991).** Mitochondrial development in *Trypanosoma brucei brucei* transitional bloodstream forms. *Molecular and biochemical parasitology*. 45: 185-92.
- Birkett, C. R., A. E. Parma, R. Gerke-Bonet, R. Woodward and K. Gull, (1992).** Isolation of cDNA clones encoding proteins of complex structures: analysis of the *Trypanosoma brucei* cytoskeleton. *Gene*. 110: 65-70.
- Branche, C., L. Kohl, G. Toutirais, J. Buisson, J. Cosson and P. Bastin, (2006).** Conserved and specific functions of axoneme components in trypanosome motility. *Journal of cell science*. 119: 3443-55.
- Brems, S., D. L. Guilbride, D. Gundlesdodjir-Planck, C. Busold, V. D. Luu, M. Schanne, J. Hoheisel and C. Clayton, (2005).** The transcriptomes of *Trypanosoma brucei* Lister 427 and TREU927 bloodstream and procyclic trypomastigotes. *Mol Biochem Parasitol*. 139: 163-72.
- Briggs, L. J., P. G. McKean, A. Baines, F. Moreira-Leite, J. Davidge, S. Vaughan and K. Gull, (2004).** The flagella connector of *Trypanosoma brucei*: an unusual mobile transmembrane junction. *Journal of cell science*. 117: 1641-51.
- Bringaud, F., C. Vedrenne, A. Cuvillier, D. Parzy, D. Baltz, E. Tetaud, E. Pays, J. Venegas, G. Merlin and T. Baltz, (1998).** Conserved organization of genes in *trypanosomatids*. *Molecular and biochemical parasitology*. 94: 249-64.
- Broadhead, R., H. R. Dawe, H. Farr, S. Griffiths, S. R. Hart, N. Portman, M. K. Shaw, M. L. Ginger, S. J. Gaskell, P. G. McKean and K. Gull, (2006).** Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature*. 440: 224-7.
- Brun, R. and L. Jenni, (1977).** A new semi-defined medium for *Trypanosoma brucei* spp. *Acta Tropica*. 34: 21-33.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, (2009). BLAST+: architecture and applications. BMC Bioinformatics. 10: 421.

Cavalier-Smith, T., (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of *Protozoa*. International journal of systematic and evolutionary microbiology. 52: 297-354.

Cavalier-Smith, T., (2010). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol Lett. 6: 342-5.

Clayton, C. E., (2002). Life without transcriptional control? From fly to man and back again. EMBO J. 21: 1881-8.

Cole, C., J. D. Barber and G. J. Barton, (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Research. 36: W197-201.

Conti, E., M. Uy, L. Leighton, G. Blobel and J. Kuriyan, (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell. 94: 193-204.

Cotrim, P. C., G. Paranhos-Baccala, M. R. Santos, C. Mortensen, M. I. Cano, M. Jolivet, M. E. Camargo, R. A. Mortara and J. F. Da Silveira, (1995). Organization and expression of the gene encoding an immunodominant repetitive antigen associated to the cytoskeleton of *Trypanosoma cruzi*. Molecular and biochemical parasitology. 71: 89-98.

Cotrim, P. C., G. S. Paranhos, R. A. Mortara, J. Wanderley, A. Rassi, M. E. Camargo and J. F. da Silveira, (1990). Expression in *Escherichia coli* of a dominant immunogen of *Trypanosoma cruzi* recognized by human chagasic sera. Journal of clinical microbiology. 28: 519-24.

Cottin, P., J. J. Brustis, S. Poussard, N. Elamrani, S. Broncard and A. Ducastaing, (1994). Ca(2+)-dependent proteinases (calpains) and muscle cell differentiation. Biochimica et biophysica acta. 1223: 170-8.

Croall, D. E., S. Chacko and Z. Wang, (1996). Cleavage of caldesmon and calponin by calpain: substrate recognition is not dependent on calmodulin binding domains. Biochimica et biophysica acta. 1298: 276-84.

Croall, D. E. and K. Ersfeld, (2007). The calpains: modular designs and functional diversity. Genome biology. 8: 218.

Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner, (2004). WebLogo: a sequence logo generator. Genome research. 14: 1188-90.

Cross, G. A., (1977). Antigenic variation in trypanosomes. The American journal of tropical medicine and hygiene. 26: 240-4.

Cuff, J. A. and G. J. Barton, (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins. 40: 502-11.

Davidge, J. A., E. Chambers, H. A. Dickinson, K. Towers, M. L. Ginger, P. G. McKean and K. Gull, (2006). Trypanosome IFT mutants provide insight into the motor location for mobility of the flagella connector and flagellar membrane formation. *Journal of cell science.* 119: 3935-43.

Davletov, B. A. and T. C. Sudhof, (1993). A single C2 domain from synaptotagmin I is sufficient for high affinity Ca^{2+} /phospholipid binding. *The Journal of biological chemistry.* 268: 26386-90.

Dawe, H. R., M. K. Shaw, H. Farr and K. Gull, (2007). The hydrocephalus inducing gene product, Hydin, positions axonemal central pair microtubules. *BMC biology.* 5: 33.

Dear, N., K. Matena, M. Vingron and T. Boehm, (1997). A new subfamily of vertebrate calpains lacking a calmodulin-like domain: implications for calpain regulation and evolution. *Genomics.* 45: 175-84.

Dear, T. N. and T. Boehm, (1999). Diverse mRNA expression patterns of the mouse calpain genes *Capn5*, *Capn6* and *Capn11* during development. *Mechanisms of development.* 89: 201-9.

Dear, T. N. and T. Boehm, (2001). Identification and characterization of two novel calpain large subunit genes. *Gene.* 274: 245-52.

Dear, T. N., A. Moller and T. Boehm, (1999). CAPN11: A calpain with high mRNA levels in testis and located on chromosome 6. *Genomics.* 59: 243-7.

Denison, S. H., M. Orejas and H. N. Arst, Jr., (1995). Signaling of ambient pH in *Aspergillus* involves a cysteine protease. *The Journal of biological chemistry.* 270: 28519-22.

Depledge, D. P., K. J. Evans, A. C. Ivens, N. Aziz, A. Maroof, P. M. Kaye and D. F. Smith, (2009). Comparative expression profiling of *Leishmania*: modulation in gene expression between species and in different host genetic backgrounds. *PLoS Negl Trop Dis.* 3: e476.

Di Lisa, F., R. De Tullio, F. Salamino, R. Barbato, E. Melloni, N. Siliprandi, S. Schiaffino and S. Pontremoli, (1995). Specific degradation of troponin T and I by mu-calpain and its modulation by substrate phosphorylation. *The Biochemical journal.* 308 (Pt 1): 57-61.

Diehl, S., F. Diehl, N. M. El-Sayed, C. Clayton and J. D. Hoheisel, (2002). Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA-microarray. *Mol Biochem Parasitol.* 123: 115-23.

Douzery, E. J., E. A. Snell, E. Baptiste, F. Delsuc and H. Philippe, (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences of the United States of America.* 101: 15386-91.

Duncan, R. R., M. J. Shipston and R. H. Chow, (2000). Double C2 protein. A review. *Biochimie.* 82: 421-6.

Dupuis-Williams, P., A. Fleury-Aubusson, N. G. de Loubresse, H. Geoffroy, L. Vayssie, A. Galvani, A. Espigat and J. Rossier, (2002). Functional role of epsilon-tubulin in the assembly of the centriolar microtubule scaffold. *The Journal of cell biology.* 158: 1183-93.

Durand-Dubief, M. and P. Bastin, (2003). TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. *BMC biology.* 1: 2.

Durand-Dubief, M., L. Kohl and P. Bastin, (2003). Efficiency and specificity of RNA interference generated by intra- and intermolecular double stranded RNA in *Trypanosoma brucei*. *Molecular and biochemical parasitology.* 129: 11-21.

Echeverri, C. J. and N. Perrimon, (2006). High-throughput RNAi screening in cultured cells: a user's guide. *Nature reviews. Genetics.* 7: 373-84.

Edgar, R. C., (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research.* 32: 1792-7.

Eisen, J. A., R. S. Coyne, M. Wu, D. Wu, M. Thiagarajan, J. R. Wortman, J. H. Badger, Q. Ren, P. Amedeo, K. M. Jones, L. J. Tallon, A. L. Delcher, S. L. Salzberg, J. C. Silva, B. J. Haas, W. H. Majoros, M. Farzad, J. M. Carlton, R. K. Smith, Jr., J. Garg, R. E. Pearlman, K. M. Karrer, L. Sun, G. Manning, N. C. Elde, A. P. Turkewitz, D. J. Asai, D. E. Wilkes, Y. Wang, H. Cai, K. Collins, B. A. Stewart, S. R. Lee, K. Wilamowska, Z. Weinberg, W. L. Ruzzo, D. Wloga, J. Gaertig, J. Frankel, C. C. Tsao, M. A. Gorovsky, P. J. Keeling, R. F. Waller, N. J. Patron, J. M. Cherry, N. A. Stover, C. J. Krieger, C. del Toro, H. F. Ryder, S. C. Williamson, R. A. Barbeau, E. P. Hamilton and E. Orias, (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS biology.* 4: e286.

El-Sayed, N. M., E. Ghedin, J. Song, A. MacLeod, F. Bringaud, C. Larkin, D. Wanless, J. Peterson, L. Hou, S. Taylor, A. Tweedie, N. Biteau, H. G. Khalak, X. Lin, T. Mason, L. Hannick, E. Caler, G. Blandin, D. Bartholomeu, A. J. Simpson, S. Kaul, H. Zhao, G. Pai, S. Van Aken, T. Utterback, B. Haas, H. L. Koo, L. Umayam, B. Suh, C. Gerrard, V. Leech, R. Qi, S. Zhou, D. Schwartz, T. Feldblyum, S. Salzberg, A. Tait, C. M. Turner, E. Ullu, O. White, S. Melville, M. D. Adams, C. M. Fraser and J. E. Donelson, (2003). The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic Acids Res.* 31: 4856-63.

El-Sayed, N. M., E. Ghedin, J. Song, A. MacLeod, F. Bringaud, C. Larkin, D. Wanless, J. Peterson, L. Hou, S. Taylor, A. Tweedie, N. Biteau, H. G. Khalak, X. Lin, T. Mason, L. Hannick, E. Caler, G. Blandin, D. Bartholomeu, A. J. Simpson, S. Kaul, H. Zhao, G. Pai, S. Van Aken, T. Utterback, B. Haas, H. L. Koo, L. Umayam, B. Suh, C. Gerrard, V. Leech, R. Qi, S. Zhou, D. Schwartz, T. Feldblyum, S. Salzberg, A. Tait, C. M. Turner, E. Ullu, O. White, S. Melville, M. D. Adams, C. M. Fraser and J. E. Donelson, (2003). The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic Acids Research.* 31: 4856-63.

El-Sayed, N. M., P. Hegde, J. Quackenbush, S. E. Melville and J. E. Donelson, (2000). The African trypanosome genome. *International journal for parasitology.* 30: 329-45.

El-Sayed, N. M., P. J. Myler, D. C. Bartholomeu, D. Nilsson, G. Aggarwal, A. N. Tran, E. Ghedin, E. A. Worthey, A. L. Delcher, G. Blandin, S. J. Westenberger, E. Caler, G. C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Aslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D. A. Campbell, M. Carrington, J. Crabtree, H. Darban, J. F. da Silveira, P. de Jong, K. Edwards, P. T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A. C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M. Klingbeil, S. Kluge, H. Koo, D. Lacerda, M. J. Levin, H. Lorenzi, T. Louie, C. R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J. C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J. L. Ramirez, J. Rinta, L. Robertson, S. L. Salzberg, D. O. Sanchez, A. Seyler, R. Sharma, J. Shetty, A. J. Simpson, E. Sisk, M. T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P. N. Ward, B. Wickstead, J. Wortman, O. White, C. M. Fraser, K. D. Stuart and B. Andersson, (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science (New York, N.Y.). 309: 409-15.

Elbashir, S. M., W. Lendeckel and T. Tuschl, (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes & development. 15: 188-200.

Emori, Y. and K. Saigo, (1994). Calpain localization changes in coordination with actin-related cytoskeletal changes during early embryonic development of *Drosophila*. The Journal of biological chemistry. 269: 25137-42.

Erickson, H. P., (1974). Assembly of microtubules from preformed, ring-shaped protofilaments and 6-S tubulin. Journal of supramolecular structure. 2: 393-411.

Ersfeld, K., H. Barraclough and K. Gull, (2005). Evolutionary relationships and protein domain architecture in an expanded calpain superfamily in kinetoplastid parasites. Journal of molecular evolution. 61: 742-57.

Ersfeld, K., H. Barraclough and K. Gull, (2005). Evolutionary relationships and protein domain architecture in an expanded calpain superfamily in kinetoplastid parasites. J Mol Evol. 61: 742-57.

Ersfeld, K. and K. Gull, (2001). Targeting of cytoskeletal proteins to the flagellum of *Trypanosoma brucei*. Journal of cell science. 114: 141-148.

Ersfeld, K., S. E. Melville and K. Gull, (1999). Nuclear and genome organization of *Trypanosoma brucei*. Parasitology today (Personal ed.). 15: 58-63.

Fernandez-Montalvan, A., I. Assfalg-Machleidt, D. Pfeiler, H. Fritz, M. Jochum and W. Machleidt, (2004). Electrostatic interactions of domain III stabilize the inactive conformation of mu-calpain. The Biochemical journal. 382: 607-17.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello, (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature. 391: 806-11.

Fraidenraich, D., C. Pena, E. L. Isola, E. M. Lammel, O. Coso, A. D. Anel, S. Pongor, F. Baralle, H. N. Torres and M. M. Flawia, (1993). Stimulation of *Trypanosoma cruzi* adenylyl cyclase by an alpha D-globin fragment from *Triatoma* hindgut: effect on differentiation of epimastigote to trypomastigote forms. Proceedings of the National Academy of Sciences of the United States of America. 90: 10140-4.

Franz, T., M. Vingron, T. Boehm and T. N. Dear, (1999). Capn7: a highly divergent vertebrate calpain with a novel C-terminal domain. *Mammalian genome : official journal of the International Mammalian Genome Society.* 10: 318-21.

Fuge, H., (1969). Electron microscopic studies on the intra-flagellar structures of trypanosomes. *The Journal of protozoology.* 16: 460-6.

Gadelha, C., B. Wickstead, P. G. McKean and K. Gull, (2006). Basal body and flagellum mutants reveal a rotational constraint of the central pair microtubules in the axonemes of trypanosomes. *J Cell Sci.* 119: 2405-13.

Garcia-Salcedo, J. A., D. Perez-Morga, P. Gijon, V. Dilbeck, E. Pays and D. P. Nolan, (2004). A differential role for actin during the life cycle of *Trypanosoma brucei*. *EMBO J.* 23: 780-9.

Gardner, M. J., R. Bishop, T. Shah, E. P. de Villiers, J. M. Carlton, N. Hall, Q. Ren, I. T. Paulsen, A. Pain, M. Berriman, R. J. Wilson, S. Sato, S. A. Ralph, D. J. Mann, Z. Xiong, S. J. Shallom, J. Weidman, L. Jiang, J. Lynn, B. Weaver, A. Shoaibi, A. R. Domingo, D. Wasawo, J. Crabtree, J. R. Wortman, B. Haas, S. V. Angiuoli, T. H. Creasy, C. Lu, B. Suh, J. C. Silva, T. R. Utterback, T. V. Feldblyum, M. Perte, J. Allen, W. C. Nierman, E. L. Taracha, S. L. Salzberg, O. R. White, H. A. Fitzhugh, S. Morzaria, J. C. Venter, C. M. Fraser and V. Nene, (2005). Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science (New York, N.Y.).* 309: 134-7.

Giese, V., B. Dallagiovanna, F. K. Marchini, D. P. Pavoni, M. A. Krieger and S. Goldenberg, (2008). *Trypanosoma cruzi*: a stage-specific calpain-like protein is induced after various kinds of stress. *Memorias do Instituto Oswaldo Cruz.* 103: 598-601.

Gil-Parrado, S., O. Popp, T. A. Knoch, S. Zahler, F. Bestvater, M. Felgentrager, A. Holloschi, A. Fernandez-Montalvan, E. A. Auerswald, H. Fritz, P. Fuentes-Prior, W. Machleidt and E. Spiess, (2003). Subcellular localization and in vivo subunit interactions of ubiquitous mu-calpain. *The Journal of biological chemistry.* 278: 16336-46.

Glading, A., D. A. Lauffenburger and A. Wells, (2002). Cutting to the chase: calpain proteases in cell motility. *Trends in cell biology.* 12: 46-54.

Glantz, S. B., C. D. Cianci, R. Iyer, D. Pradhan, K. K. Wang and J. S. Morrow, (2007). Sequential degradation of alphaII and betaII spectrin by calpain in glutamate or maitotoxin-stimulated cells. *Biochemistry.* 46: 502-13.

Gluenz, E., R. Sharma, M. Carrington and K. Gull, (2008). Functional characterization of cohesin subunit SCC1 in *Trypanosoma brucei* and dissection of mutant phenotypes in two life cycle stages. *Molecular microbiology.* 69: 666-80.

Godsel, L. M. and D. M. Engman, (1999). Flagellar protein localization mediated by a calcium-myristoyl/palmitoyl switch mechanism. *The EMBO journal.* 18: 2057-65.

Goll, D. E., V. F. Thompson, H. Li, W. Wei and J. Cong, (2003). The calpain system. *Physiological reviews.* 83: 731-801.

Gonzales-Perdomo, M., P. Romero and S. Goldenberg, (1988). Cyclic AMP and adenylate cyclase activators stimulate *Trypanosoma cruzi* differentiation. *Experimental Parasitology*. 66: 205-12.

Gotoh, O., (1982). An improved algorithm for matching biological sequences. *Journal of molecular biology*. 162: 705-8.

Griffiths, S., N. Portman, P. R. Taylor, S. Gordon, M. L. Ginger and K. Gull, (2007). RNA interference mutant induction in vivo demonstrates the essential nature of trypanosome flagellar function during mammalian infection. *Eukaryotic cell*. 6: 1248-50.

Gull, K., (1999). The cytoskeleton of trypanosomatid parasites. *Annual Review of Microbiology*. 53: 629-55.

Hall, N., M. Berriman, N. J. Lennard, B. R. Harris, C. Hertz-Fowler, E. N. Bart-Delabesse, C. S. Gerrard, R. J. Atkin, A. J. Barron, S. Bowman, S. P. Bray-Allen, F. Bringaud, L. N. Clark, C. H. Corton, A. Cronin, R. Davies, J. Doggett, A. Fraser, E. Gruter, S. Hall, A. D. Harper, M. P. Kay, V. Leech, R. Mayes, C. Price, M. A. Quail, E. Rabinowitsch, C. Reitter, K. Rutherford, J. Sasse, S. Sharp, R. Shownkeen, A. MacLeod, S. Taylor, A. Tweedie, C. M. Turner, A. Tait, K. Gull, B. Barrell and S. E. Melville, (2003). The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. *Nucleic Acids Research*. 31: 4864-73.

Hammarton, T. C., J. Clark, F. Douglas, M. Boshart and J. C. Mottram, (2003). Stage-specific differences in cell cycle control in *Trypanosoma brucei* revealed by RNA interference of a mitotic cyclin. *The Journal of biological chemistry*. 278: 22877-86.

Hammarton, T. C., S. Kramer, L. Tetley, M. Boshart and J. C. Mottram, (2007). *Trypanosoma brucei* Polo-like kinase is essential for basal body duplication, kDNA segregation and cytokinesis. *Molecular microbiology*. 65: 1229-48.

Hampl, V., L. Hug, J. W. Leigh, J. B. Dacks, B. F. Lang, A. G. Simpson and A. J. Roger, (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci U S A*. 106: 3859-64.

Harris, A. S., D. E. Croall and J. S. Morrow, (1988). The calmodulin-binding site in alpha-fodrin is near the calcium-dependent protease-I cleavage site. *The Journal of biological chemistry*. 263: 15754-61.

Hata, S., S. Koyama, H. Kawahara, N. Doi, T. Maeda, N. Toyama-Sorimachi, K. Abe, K. Suzuki and H. Sorimachi, (2006). Stomach-specific calpain, nCL-2, localizes in mucus cells and proteolyzes the beta-subunit of coatamer complex, beta-COP. *The Journal of biological chemistry*. 281: 11214-24.

Hell, J. W., R. E. Westenbroek, L. J. Breeze, K. K. Wang, C. Chavkin and W. A. Catterall, (1996). N-methyl-D-aspartate receptor-induced proteolytic conversion of postsynaptic class C L-type calcium channels in hippocampal neurons. *Proceedings of the National Academy of Sciences of the United States of America*. 93: 3362-7.

Hemphill, A., M. Affolter and T. Seebeck, (1992). A novel microtubule-binding motif identified in a high molecular weight microtubule-associated protein from *Trypanosoma brucei*. The Journal of cell biology. 117: 95-103.

Hemphill, A., D. Lawson and T. Seebeck, (1991a). The cytoskeletal architecture of *Trypanosoma brucei*. The Journal of parasitology. 77: 603-12.

Hemphill, A., T. Seebeck and D. Lawson, (1991b). The *Trypanosoma brucei* cytoskeleton: ultrastructure and localization of microtubule-associated and spectrin-like proteins using quick-freeze, deep-etch, immunogold electron microscopy. Journal of structural biology. 107: 211-20.

Hertz-Fowler, C., (2001) PhD Thesis
School of Biological Sciences University of Manchester

Hertz-Fowler, C., K. Ersfeld and K. Gull, (2001). CAP5.5, a life-cycle-regulated, cytoskeleton-associated protein is a member of a novel family of calpain-related proteins in *Trypanosoma brucei*. Molecular and biochemical parasitology. 116: 25-34.

Hertz-Fowler, C., C. S. Peacock, V. Wood, M. Aslett, A. Kerhornou, P. Mooney, A. Tivey, M. Berriman, N. Hall, K. Rutherford, J. Parkhill, A. C. Ivens, M. A. Rajandream and B. Barrell, (2004). GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucleic Acids Research. 32: D339-43.

Hirumi, H. and K. Hirumi, (1989). Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. The Journal of parasitology. 75: 985-9.

Hoare, C. A., (1966). The classification of mammalian trypanosomes. Ergebnisse der Mikrobiologie, Immunitätsforschung und experimentellen Therapie. 39: 43-57.

Hoare, C. A. and F. G. Wallace, (1966). Developmental stages of trypanosomatid flagellates: a new terminology. Nature. 212: 1385-1386.

Honigberg, B. M., (1963). Evolutionary and systematic relationships in the flagellate order *Trichomonadida* Kirby. The Journal of protozoology. 10: 20-63.

Hood, J. L., B. B. Logan, A. P. Sinai, W. H. Brooks and T. L. Roszman, (2003). Association of the calpain/calpastatin network with subcellular organelles. Biochemical and biophysical research communications. 310: 1200-12.

Horikawa, Y., N. Oda, N. J. Cox, X. Li, M. Orho-Melander, M. Hara, Y. Hinokio, T. H. Lindner, H. Mashima, P. E. Schwarz, L. del Bosque-Plata, Y. Oda, I. Yoshiuchi, S. Colilla, K. S. Polonsky, S. Wei, P. Concannon, N. Iwasaki, J. Schulze, L. J. Baier, C. Bogardus, L. Groop, E. Boerwinkle, C. L. Hanis and G. I. Bell, (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nature genetics. 26: 163-75.

Hosfield, C. M., J. S. Elce, P. L. Davies and Z. Jia, (1999). Crystal structure of calpain reveals the structural basis for Ca(2+)-dependent protease activity and a novel mode of enzyme activation. The EMBO journal. 18: 6880-9.

Hsu, Y., G. Jubelin, F. Taieb, J. P. Nougayrede, E. Oswald and C. E. Stebbins, (2008). Structure of the cyclomodulin Cif from pathogenic *Escherichia coli*. *Journal of molecular biology*. 384: 465-77.

<http://www.genedb.org/tryp>. "[<http://www.genedb.org/tryp>."

<http://www.genedb.org/tryp/>. "[<http://www.genedb.org/tryp/>]."

Hua, L. V., K. Hidaka, X. Pesesse, L. D. Barnes and S. B. Shears, (2003). Paralogous murine Nudt10 and Nudt11 genes have differential expression patterns but encode identical proteins that are physiologically competent diphosphoinositol polyphosphate phosphohydrolases. *The Biochemical journal*. 373: 81-9.

Huber, A. H., W. J. Nelson and W. I. Weis, (1997). Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell*. 90: 871-82.

Hurley, J. H., (1999). Structure, mechanism, and regulation of mammalian adenylyl cyclase. *The Journal of biological chemistry*. 274: 7599-602.

Hurley, J. H. and S. Misra, (2000). Signaling and subcellular targeting by membrane-binding domains. *Annual review of biophysics and biomolecular structure*. 29: 49-79.

Hutchinson, O. C., K. Picozzi, N. G. Jones, H. Mott, R. Sharma, S. C. Welburn and M. Carrington, (2007). Variant Surface Glycoprotein gene repertoires in *Trypanosoma brucei* have diverged to become strain-specific. *BMC Genomics*. 8: 234.

Hutvagner, G. and P. D. Zamore, (2002). RNAi: nature abhors a double-strand. *Current opinion in genetics & development*. 12: 225-32.

Ibanez, C. F., J. L. Affranchino, R. A. Macina, M. B. Reyes, S. Leguizamon, M. E. Camargo, L. Aslund, U. Pettersson and A. C. Frasch, (1988). Multiple *Trypanosoma cruzi* antigens containing tandemly repeated amino acid sequence motifs. *Molecular and biochemical parasitology*. 30: 27-33.

Inomata, M., M. Hayashi, Y. Ohno-Iwashita, S. Tsubuki, T. C. Saido and S. Kawashima, (1996). Involvement of calpain in integrin-mediated signal transduction. *Arch Biochem Biophys*. 328: 129-34.

Ivens, A. C., C. S. Peacock, E. A. Worthey, L. Murphy, G. Aggarwal, M. Berriman, E. Sisk, M. A. Rajandream, E. Adlem, R. Aert, A. Anupama, Z. Apostolou, P. Attipoe, N. Bason, C. Bauser, A. Beck, S. M. Beverley, G. Bianchetti, K. Borzym, G. Bothe, C. V. Bruschi, M. Collins, E. Cadag, L. Ciarloni, C. Clayton, R. M. Coulson, A. Cronin, A. K. Cruz, R. M. Davies, J. De Gaudenzi, D. E. Dobson, A. Duesterhoeft, G. Fazelina, N. Fosker, A. C. Frasch, A. Fraser, M. Fuchs, C. Gabel, A. Goble, A. Goffeau, D. Harris, C. Hertz-Fowler, H. Hilbert, D. Horn, Y. Huang, S. Klages, A. Knights, M. Kube, N. Larke, L. Litvin, A. Lord, T. Louie, M. Marra, D. Masuy, K. Matthews, S. Michaeli, J. C. Mottram, S. Muller-Auer, H. Munden, S. Nelson, H. Norbertczak, K. Oliver, S. O'Neil, M. Pentony, T. M. Pohl, C. Price, B. Purnelle, M. A. Quail, E. Rabinowitsch, R. Reinhardt, M. Rieger, J. Rinta, J. Robben, L. Robertson, J. C. Ruiz, S. Rutter, D. Saunders, M. Schafer, J. Schein, D. C. Schwartz, K. Seeger, A. Seyler, S. Sharp, H. Shin, D. Sivam, R. Squares, S. Squares, V. Tosato, C. Vogt, G. Volckaert, R. Wambutt, T. Warren, H. Wedler, J. Woodward, S. Zhou, W. Zimmermann, D. F. Smith, J. M. Blackwell, K. D. Stuart, B. Barrell and P. J. Myler, (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science (New York, N.Y.)*. 309: 436-42.

Jackson, A. P., (2007b). Evolutionary consequences of a large duplication event in *Trypanosoma brucei*: chromosomes 4 and 8 are partial duplicons. BMC Genomics. 8: 432.

Jekely, G. and P. Friedrich, (1999). The evolution of the calpain family as reflected in paralogous chromosome regions. Journal of molecular evolution. 49: 272-81.

Jensen, B. C., D. Sivam, C. T. Kifer, P. J. Myler and M. Parsons, (2009). Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. BMC Genomics. 10: 482.

Jones, A., A. Faldas, A. Foucher, E. Hunt, A. Tait, J. M. Wastling and C. M. Turner, (2006). Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*. Proteomics. 6: 259-67.

Kabani, S., K. Fenn, A. Ross, A. Ivens, T. K. Smith, P. Ghazal and K. Matthews, (2009). Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*. BMC Genomics. 10: 427.

Kamei, M., G. C. Webb, I. G. Young and H. D. Campbell, (1998). SOLH, a human homologue of the *Drosophila melanogaster* small optic lobes gene is a member of the calpain and zinc-finger gene families and maps to human chromosome 16p13.3 near CATM (cataract with microphthalmia). Genomics. 51: 197-206.

Karcz, S. R., R. B. Podesta, A. A. Siddiqui, G. A. Dekaban, G. H. Strejan and M. W. Clarke, (1991). Molecular cloning and sequence analysis of a calcium-activated neutral protease (calpain) from *Schistosoma mansoni*. Molecular and biochemical parasitology. 49: 333-6.

Katoh, K., K. Misawa, K. Kuma and T. Miyata, (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research. 30: 3059-66.

Kennedy, P. G., (2008). The continuing problem of human African trypanosomiasis (sleeping sickness). Annals of neurology. 64: 116-26.

King, S. M., (2000). The dynein microtubule motor. Biochimica et biophysica acta. 1496: 60-75.

Kohl, L., T. Sherwin and K. Gull, (1999). Assembly of the paraflagellar rod and the flagellum attachment zone complex during the *Trypanosoma brucei* cell cycle. The Journal of eukaryotic microbiology. 46: 105-9.

Koumandou, V. L., S. K. Natesan, T. Sergeenko and M. C. Field, (2008). The trypanosome transcriptome is remodelled during differentiation but displays limited responsiveness within life stages. BMC Genomics. 9: 298.

Kreis, T. and R. Vale, (1999). Guide book to the Cytoskeletal and Motor Proteins. Oxford, Smbrook and Tooze Pub. at Oxford University Press.

Kubbutat, M. H. and K. H. Vousden, (1997). Proteolytic cleavage of human p53 by calpain: a potential regulator of protein stability. *Molecular and cellular biology*. 17: 460-8.

Kulkarni, S., D. E. Goll and J. E. Fox, (2002). Calpain cleaves RhoA generating a dominant-negative form that inhibits integrin-induced actin filament assembly and cell spreading. *The Journal of biological chemistry*. 277: 24435-41.

Kwak, K. B., S. S. Chung, O. M. Kim, M. S. Kang, D. B. Ha and C. H. Chung, (1993). Increase in the level of m-calpain correlates with the elevated cleavage of filamin during myogenic differentiation of embryonic muscle cells. *Biochimica et biophysica acta*. 1175: 243-9.

LaCount, D. J., B. Barrett and J. E. Donelson, (2002). *Trypanosoma brucei* FLA1 is required for flagellum attachment and cytokinesis. *The Journal of biological chemistry*. 277: 17580-8.

LaCount, D. J., S. Bruse, K. L. Hill and J. E. Donelson, (2000). Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. *Molecular and biochemical parasitology*. 111: 67-76.

Lafaille, J. J., J. Linss, M. A. Krieger, T. Souto-Padron, W. de Souza and S. Goldenberg, (1989). Structure and expression of two *Trypanosoma cruzi* genes encoding antigenic proteins bearing repetitive epitopes. *Molecular and biochemical parasitology*. 35: 127-36.

Lai, D. H., H. Hashimi, Z. R. Lun, F. J. Ayala and J. Lukes, (2008). Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. *Proceedings of the National Academy of Sciences of the United States of America*. 105: 1999-2004.

Lee, H. J., H. Sorimachi, S. Y. Jeong, S. Ishiura and K. Suzuki, (1998). Molecular cloning and characterization of a novel tissue-specific calpain predominantly expressed in the digestive tract. *Biological chemistry*. 379: 175-83.

Leifso, K., G. Cohen-Freue, N. Dogra, A. Murray and W. R. McMaster, (2007). Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: the *Leishmania* genome is constitutively expressed. *Mol Biochem Parasitol*. 152: 35-46.

Levin, M. J., E. Mesri, R. Benarous, G. Levitus, A. Schijman, P. Levy-Yeyati, P. A. Chiale, A. M. Ruiz, A. Kahn, M. B. Rosenbaum and et al., (1989). Identification of major *Trypanosoma cruzi* antigenic determinants in chronic Chagas' heart disease. *The American journal of tropical medicine and hygiene*. 41: 530-8.

Lewis, E. A. and W. P. Langridge, (1947). Developmental forms of *Trypanosoma brucei* in the saliva of *Glossina pallipedes* and *Glossina austeni*. *Annals of Tropical Medicine and Parasitology*. 41: 6-13.

Li, Z. and C. C. Wang, (2008). KMP-11, a basal body and flagellar protein, is required for cell division in *Trypanosoma brucei*. *Eukaryotic cell*. 7: 1941-50.

Liu, W., K. Apagyi, L. McLeavy and K. Ersfeld, (2009). Expression and cellular localisation of calpain-like proteins in *Trypanosoma brucei*. Molecular and biochemical parasitology. 169: 20-6.

Loftus, B., I. Anderson, R. Davies, U. C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri, Jr., C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser and N. Hall, (2005). The genome of the protist parasite *Entamoeba histolytica*. Nature. 433: 865-8.

Lucas, K. A., G. M. Pitari, S. Kazerounian, I. Ruiz-Stewart, J. Park, S. Schulz, K. P. Chepenik and S. A. Waldman, (2000). Guanylyl cyclases and signaling by cyclic GMP. Pharmacological reviews. 52: 375-414.

Margolis, R. L. and L. Wilson, (1981). Microtubule treadmills--possible molecular machinery. Nature. 293: 705-11.

Mathers, C. D., (2005). Uncertainty and data availability for the global burden of disease estimates 2000-2002. . Evidence and Information for Policy Working Paper. Geneva, World Health Organization.

Mathers, C. D., C. Bernard, K. Iburg, M. Inoue, D. Ma Fat, K. Shibuya, C. Stein and N. Tomijima, (2003). The Global Burden of Disease in 2002: data sources, methods and results. . Geneva, World Health Organization. (GPE Discussion Paper No. 54). .

Matthews, K. R., (2005). The developmental cell biology of *Trypanosoma brucei*. Journal of cell science. 118: 283-90.

Matthews, K. R. and K. Gull, (1994a). Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. The Journal of cell biology. 125: 1147-56.

Matzke, M. A. and J. A. Birchler, (2005). RNAi-mediated pathways in the nucleus. Nature reviews. Genetics. 6: 24-35.

Maurer-Stroh, S., B. Eisenhaber and F. Eisenhaber, (2002a). N-terminal N-myristoylation of proteins: prediction of substrate proteins from amino acid sequence. Journal of molecular biology. 317: 541-57.

Maurer-Stroh, S., B. Eisenhaber and F. Eisenhaber, (2002b). N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. Journal of molecular biology. 317: 523-40.

Mayer, B. J. and D. Baltimore, (1993). Signalling through SH2 and SH3 domains. Trends in cell biology. 3: 8-13.

McKean, P. G., A. Baines, S. Vaughan and K. Gull, (2003). Gamma-tubulin functions in the nucleation of a discrete subset of microtubules in the eukaryotic flagellum. *Current biology* : CB. 13: 598-602.

Meister, G. and T. Tuschl, (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature*. 431: 343-9.

Melkonian, K. A., A. G. Ostermeyer, J. Z. Chen, M. G. Roth and D. A. Brown, (1999). Role of lipid modifications in targeting proteins to detergent-resistant membrane rafts. Many raft proteins are acylated, while few are prenylated. *The Journal of biological chemistry*. 274: 3910-7.

Mills, E., H. P. Price, A. Johner, J. E. Emerson and D. F. Smith, (2007). Kinetoplastid PPEF phosphatases: dual acylated proteins expressed in the endomembrane system of *Leishmania*. *Molecular and biochemical parasitology*. 152: 22-34.

Moldoveanu, T., C. M. Hosfield, D. Lim, J. S. Elce, Z. Jia and P. L. Davies, (2002). A Ca(2+) switch aligns the active site of calpain. *Cell*. 108: 649-60.

Moldoveanu, T., Z. Jia and P. L. Davies, (2004). Calpain activation by cooperative Ca2+ binding at two non-EF-hand sites. *The Journal of biological chemistry*. 279: 6106-14.

Molinari, M., J. Anagli and E. Carafoli, (1995). PEST sequences do not influence substrate susceptibility to calpain proteolysis. *The Journal of biological chemistry*. 270: 2032-5.

Moreira-Leite, F. F., T. Sherwin, L. Kohl and K. Gull, (2001). A trypanosome structure involved in transmitting cytoplasmic information during cell division. *Science (New York, N.Y.)*. 294: 610-2.

Motyka, S. A. and P. T. Englund, (2004). RNA interference for analysis of gene function in *Trypanosomatids*. *Current opinion in microbiology*. 7: 362-8.

Muller, N., A. Hemphill, M. Imboden, G. Duvallet, R. H. Dwinger and T. Seebeck, (1992). Identification and characterization of two repetitive non-variable antigens from African trypanosomes which are recognized early during infection. *Parasitology*. 104 Pt 1: 111-20.

Muller, N., M. Imboden, E. Detmer, J. M. Mansfield and T. Seebeck, (1993). Cytoskeleton-associated antigens from African trypanosomes are recognized by self-reactive antibodies of uninfected mice. *Parasitology*. 107 (Pt 4): 411-7.

Musacchio, A., T. Gibson, V. P. Lehto and M. Saraste, (1992). SH3--an abundant protein domain in search of a function. *FEBS letters*. 307: 55-61.

Mykles, D. L. and D. M. Skinner, (1982). Molt cycle-associated changes in calcium-dependent proteinase activity that degrades actin and myosin in crustacean muscle. *Developmental biology*. 92: 386-97.

Mykles, D. L. and D. M. Skinner, (1986). Four Ca²⁺-dependent proteinase activities isolated from crustacean muscle differ in size, net charge, and sensitivity to Ca²⁺ and inhibitors. *The Journal of biological chemistry*. 261: 9865-71.

Myler, P. J., L. Audleman, T. deVos, G. Hixson, P. Kiser, C. Lemley, C. Magness, E. Rickel, E. Sisk, S. Sunkin, S. Swartzell, T. Westlake, P. Bastien, G. Fu, A. Ivens and K. Stuart, (1999). *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proceedings of the National Academy of Sciences of the United States of America*. 96: 2902-6.

Nakagawa, T. and J. Yuan, (2000). Cross-talk between two cysteine protease families. Activation of caspase-12 by calpain in apoptosis. *The Journal of cell biology*. 150: 887-94.

Naula, C. and T. Seebeck, (2000). Cyclic AMP signaling in *Trypanosomatids*. *Parasitology today* (Personal ed.). 16: 35-8.

Ngo, H., C. Tschudi, K. Gull and E. Ullu, (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America*. 95: 14687-92.

Ngo, H., C. Tschudi, K. Gull and E. Ullu, (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A*. 95: 14687-92.

Noguchi, M., A. Sarin, M. J. Aman, H. Nakajima, E. W. Shores, P. A. Henkart and W. J. Leonard, (1997). Functional cleavage of the common cytokine receptor gamma chain (gamma_c) by calpain. *Proceedings of the National Academy of Sciences of the United States of America*. 94: 11534-9.

Nozaki, T., P. A. Haynes and G. A. Cross, (1996). Characterization of the *Trypanosoma brucei* homologue of a *Trypanosoma cruzi* flagellum-adhesion glycoprotein. *Molecular and biochemical parasitology*. 82: 245-55.

Nykanen, A., B. Haley and P. D. Zamore, (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*. 107: 309-21.

Oda, A., H. Wakao and H. Fujita, (2002). Calpain is a signal transducer and activator of transcription (STAT) 3 and STAT5 protease. *Blood*. 99: 1850-2.

Ogbadoyi, E. O., D. R. Robinson and K. Gull, (2003). A high-order trans-membrane structural linkage is responsible for mitochondrial genome positioning and segregation by flagellar basal bodies in trypanosomes. *Molecular biology of the cell*. 14: 1769-79.

Olego-Fernandez, S., S. Vaughan, M. K. Shaw, K. Gull and M. L. Ginger, (2009). Cell morphogenesis of *Trypanosoma brucei* requires the paralogous, differentially expressed calpain-related proteins CAP5.5 and CAP5.5V. *Protist*. 160: 576-90.

Ouaissi, A., J. Cornette, R. Schoneck, B. Plumas-Marty, A. Taibi, M. Loyens and A. Capron, (1992). Fibronectin cleavage fragments provide a growth factor-like activity for the differentiation of *Trypanosoma cruzi* trypomastigotes to amastigotes. *European journal of cell biology*. 59: 68-79.

- Patel, T. B., Z. Du, S. Pierre, L. Cartin and K. Scholich**, (2001). Molecular biological approaches to unravel adenylyl cyclase signaling and function. *Gene*. 269: 13-25.
- Pawson, T.**, (1995). Protein modules and signalling networks. *Nature*. 373: 573-580.
- Perrin, B. J. and A. Huttenlocher**, (2002). Calpain. *The international journal of biochemistry & cell biology*. 34: 722-5.
- Pinter, M. and P. Friedrich**, (1988). The calcium-dependent proteolytic system calpain-calpastatin in *Drosophila melanogaster*. *The Biochemical journal*. 253: 467-73.
- Pinter, M., A. Stierandova and P. Friedrich**, (1992). Purification and characterization of a Ca(2+)-activated thiol protease from *Drosophila melanogaster*. *Biochemistry*. 31: 8201-6.
- Ploubidou, A., D. R. Robinson, R. C. Docherty, E. O. Ogbadoyi and K. Gull**, (1999). Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation and cytokinesis in the absence of mitosis. *Journal of cell science*. 112 (Pt 24): 4641-50.
- Porter, M. E. and W. S. Sale**, (2000). The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility. *The Journal of cell biology*. 151: F37-42.
- Ralston, K. S., A. G. Lerner, D. R. Diener and K. L. Hill**, (2006). Flagellar motility contributes to cytokinesis in *Trypanosoma brucei* and is modulated by an evolutionarily conserved dynein regulatory system. *Eukaryotic cell*. 5: 696-711.
- Ralton, J. E., K. A. Mullin and M. J. McConville**, (2002). Intracellular trafficking of glycosylphosphatidylinositol (GPI)-anchored proteins and free GPIs in *Leishmania mexicana*. *The Biochemical journal*. 363: 365-75.
- Ramos, C. S., F. A. Franco, D. F. Smith and S. R. Uliana**, (2004). Characterisation of a new *Leishmania* META gene and genomic analysis of the META cluster. *FEMS microbiology letters*. 238: 213-9.
- Ren, R., B. J. Mayer, P. Cicchetti and D. Baltimore**, (1993). Identification of a ten-amino acid proline-rich SH3 binding site. *Science (New York, N.Y.)*. 259: 1157-61.
- Resh, M. D.**, (1999). Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica et biophysica acta*. 1451: 1-16.
- Reverter, D., S. Strobl, C. Fernandez-Catalan, H. Sorimachi, K. Suzuki and W. Bode**, (2001). Structural basis for possible calcium-induced activation mechanisms of calpains. *Biological chemistry*. 382: 753-66.
- Richards, T. A. and T. Cavalier-Smith**, (2005). Myosin domain evolution and the primary divergence of eukaryotes. *Nature*. 436: 1113-8.
- Richardson, D. N., M. P. Simmons and A. S. Reddy**, (2006). Comprehensive comparative analysis of kinesins in photosynthetic eukaryotes. *BMC Genomics*. 7: 18.
- Rizo, J. and T. C. Sudhof**, (1998). C2-domains, structure and function of a universal Ca²⁺-binding domain. *The Journal of biological chemistry*. 273: 15879-82.

Robinson, D. R. and K. Gull, (1991). Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. *Nature*. 352: 731-3.

Robinson, D. R., T. Sherwin, A. Ploubidou, E. H. Byard and K. Gull, (1995). Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. *The Journal of cell biology*. 128: 1163-72.

Rochette, A., F. Raymond, J. M. Ubeda, M. Smith, N. Messier, S. Boisvert, P. Rigault, J. Corbeil, M. Ouellette and B. Papadopolou, (2008). Genome-wide gene expression profiling analysis of *Leishmania major* and *Leishmania infantum* developmental stages reveals substantial differences between the two species. *BMC Genomics*. 9: 255.

Roditi, I. and M. Liniger, (2002). Dressed for success: the surface coats of insect-borne protozoan parasites. *Trends in microbiology*. 10: 128-34.

Rolin, S., P. Paindavoine, J. Hanocq-Quertier, F. Hanocq, Y. Claes, D. Le Ray, P. Overath and E. Pays, (1993). Transient adenylate cyclase activation accompanies differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms. *Molecular and biochemical parasitology*. 61: 115-25.

Roth, C., S. Rastogi, L. Arvestad, K. Dittmar, S. Light, D. Ekman and D. A. Liberles, (2007). Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *Journal of experimental zoology. Part B. Molecular and developmental evolution*. 308: 58-73.

Ruiz-Moreno, L., A. T. Bijovsky, J. Pudles, M. J. M. Alves and W. Colli, (1995). *Trypanosoma cruzi*: Monoclonal Antibody to Cytoskeleton Recognizes Giant Proteins of the Flagellar Attachment Zone. *Experimental Parasitology*. 80: 605-615.

Rychlik, W. and R. E. Rhoads, (1989). A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Research*. 17: 8543-51.

Sakai, T., H. Honing, M. Nishioka, Y. Uehara, M. Takahashi, N. Fujisawa, K. Saji, M. Seki, K. Shinozaki, M. A. Jones, N. Smirnov, K. Okada and G. O. Wasteneys, (2008). Armadillo repeat-containing kinesins and a NIMA-related kinase are required for epidermal-cell morphogenesis in *Arabidopsis*. *The Plant journal : for cell and molecular biology*. 53: 157-71.

Sanchez, M. A., D. Zeoli, E. M. Klamo, M. P. Kavanaugh and S. M. Landfear, (1995). A family of putative receptor-adenylate cyclases from *Leishmania donovani*. *The Journal of biological chemistry*. 270: 17551-8.

SantaLucia, J., Jr., (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*. 95: 1460-5.

Santella, L., K. Kyojuka, L. De Riso and E. Carafoli, (1998). Calcium, protease action, and the regulation of the cell cycle. *Cell Calcium*. 23: 123-30.

Santella, L., K. Kyojuka, S. Hoving, M. Munchbach, M. Quadroni, P. Dainese, C. Zamparelli, P. James and E. Carafoli, (2000). Breakdown of cytoskeletal proteins during meiosis of starfish oocytes and proteolysis induced by calpain. *Experimental Cell Research*. 259: 117-26.

Sasaki, T., T. Kikuchi, N. Yumoto, N. Yoshimura and T. Murachi, (1984). Comparative specificity and kinetic studies on porcine calpain I and calpain II with naturally occurring peptides and synthetic fluorogenic substrates. *The Journal of biological chemistry*. 259: 12489-94.

Satir, P., (1989). The role of axonemal components in ciliary motility. *Comparative biochemistry and physiology. A, Comparative physiology*. 94: 351-7.

Satir, P., (1992). Mechanisms of ciliary movement: contributions from electron microscopy. *Scanning microscopy*. 6: 573-9.

Sato, K., Y. Saito and S. Kawashima, (1995). Identification and characterization of membrane-bound calpains in clathrin-coated vesicles from bovine brain. *European journal of biochemistry / FEBS*. 230: 25-31.

Saxena, A., E. A. Worthey, S. Yan, A. Leland, K. D. Stuart and P. J. Myler, (2003). Evaluation of differential gene expression in *Leishmania major* Friedlin procyclics and metacyclics using DNA microarray analysis. *Molecular and biochemical parasitology*. 129: 103-14.

Scott, V., T. Sherwin and K. Gull, (1997). gamma-tubulin in trypanosomes: molecular characterisation and localisation to multiple and diverse microtubule organising centres. *Journal of cell science*. 110 (Pt 2): 157-68.

Schechter, I. and A. Berger, (1967). On the size of the active site in proteases. I. Papain. *Biochemical and biophysical research communications*. 27: 157-62.

Schultz, J., F. Milpetz, P. Bork and C. P. Ponting, (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*. 95: 5857-64.

Sharma, R., L. Peacock, E. Gluenz, K. Gull, W. Gibson and M. Carrington, (2008). Asymmetric cell division as a route to reduction in cell length and change in cell morphology in trypanosomes. *Protist*. 159: 137-51.

Sherwin, T. and K. Gull, (1989a). The cell division cycle of *Trypanosoma brucei brucei*: timing of event markers and cytoskeletal modulations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 323: 573-88.

Sherwin, T. and K. Gull, (1989a). The cell division cycle of *Trypanosoma brucei brucei*: timing of event markers and cytoskeletal modulations. *Philos Trans R Soc Lond B Biol Sci*. 323: 573-88.

Sherwin, T. and K. Gull, (1989b). Visualization of detyrosination along single microtubules reveals novel mechanisms of assembly during cytoskeletal duplication in trypanosomes. *Cell*. 57: 211-21.

Sherwin, T., A. Schneider, R. Sasse, T. Seebeck and K. Gull, (1987). Distinct localization and cell cycle dependence of COOH terminally tyrosinolated alpha-tubulin in the microtubules of *Trypanosoma brucei brucei*. The Journal of cell biology. 104: 439-46.

Shevchenko, S., W. Feng, M. Varsanyi and V. Shoshan-Barmatz, (1998). Identification, characterization and partial purification of a thiol-protease which cleaves specifically the skeletal muscle ryanodine receptor/Ca²⁺ release channel. The Journal of membrane biology. 161: 33-43.

Shi, H., A. Djikeng, T. Mark, E. Wirtz, C. Tschudi and E. Ullu, (2000). Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. RNA (New York, N.Y.). 6: 1069-76.

Shi, H., A. Djikeng, C. Tschudi and E. Ullu, (2004). Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. Molecular and cellular biology. 24: 420-7.

Shi, H., C. Tschudi and E. Ullu, (2006). An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. RNA (New York, N.Y.). 12: 2063-72.

Siddiqui, A. A., Y. Zhou, R. B. Podesta, S. R. Karcz, C. E. Tognon, G. H. Strejan, G. A. Dekaban and M. W. Clarke, (1993). Characterization of Ca(2+)-dependent neutral protease (calpain) from human blood flukes, *Schistosoma mansoni*. Biochimica et biophysica acta. 1181: 37-44.

Simons, K. and E. Ikonen, (1997). Functional rafts in cell membranes. Nature. 387: 569-72.

Simonson, L., M. Baudry, R. Siman and G. Lynch, (1985). Regional distribution of soluble calcium activated proteinase activity in neonatal and adult rat brain. Brain research. 327: 153-9.

Sonnhammer, E. L. and R. Durbin, (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 167: GC1-10.

Sorimachi, H., S. Imajoh-Ohmi, Y. Emori, H. Kawasaki, S. Ohno, Y. Minami and K. Suzuki, (1989). Molecular cloning of a novel mammalian calcium-dependent protease distinct from both m- and mu-types. Specific expression of the mRNA in skeletal muscle. The Journal of biological chemistry. 264: 20106-11.

Sorimachi, H., S. Ishiura and K. Suzuki, (1997). Structure and physiological function of calpains. The Biochemical journal. 328 (Pt 3): 721-32.

Sorimachi, H. and K. Suzuki, (2001). The structure of calpain. Journal of biochemistry. 129: 653-64.

Souto-Padron, T., M. B. Reyes, S. Leguizamon, O. E. Campetella, A. C. Frasch and W. de Souza, (1989). *Trypanosoma cruzi* proteins which are antigenic during human infections are located in defined regions of the parasite. European journal of cell biology. 50: 272-8.

Squier, M. K., A. C. Miller, A. M. Malkinson and J. J. Cohen, (1994). Calpain activation in apoptosis. *Journal of cellular physiology.* 159: 229-37.

Stabach, P. R., C. D. Cianci, S. B. Glantz, Z. Zhang and J. S. Morrow, (1997). Site-directed mutagenesis of alpha II spectrin at codon 1175 modulates its mu-calpain susceptibility. *Biochemistry.* 36: 57-65.

Strobl, S., C. Fernandez-Catalan, M. Braun, R. Huber, H. Masumoto, K. Nakagawa, A. Irie, H. Sorimachi, G. Bourenkow, H. Bartunik, K. Suzuki and W. Bode, (2000). The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proceedings of the National Academy of Sciences of the United States of America.* 97: 588-92.

Subramaniam, C., P. Veazey, S. Redmond, J. Hayes-Sinclair, E. Chambers, M. Carrington, K. Gull, K. Matthews, D. Horn and M. C. Field, (2006). Chromosome-wide analysis of gene function by RNA interference in the african trypanosome. *Eukaryotic cell.* 5: 1539-49.

Suzuki, K., (1990). The structure of the calpains and calpain gene. Boca Raton, Fl.

Suzuki, K., (1991). Nomenclature of calcium dependent proteinase. *Biomedica biochimica acta.* 50: 483-4.

Tajima, F., (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics.* 135: 599-607.

Takamure, M., K. Y. Murata, Y. Tamada, M. Azuma and S. Ueno, (2005). Calpain-dependent alpha-fodrin cleavage at the sarcolemma in muscle diseases. *Muscle & Nerve.* 32: 303-9.

Tang, C., E. Loeliger, P. Luncsford, I. Kinde, D. Beckett and M. F. Summers, (2004). Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proceedings of the National Academy of Sciences of the United States of America.* 101: 517-22.

Taussig, R. and G. Zimmermann, (1998). Type-specific regulation of mammalian adenylyl cyclases by G protein pathways. *Advances in second messenger and phosphoprotein research.* 32: 81-98.

Taylor, J. S. and J. Raes, (2005). Small-scale gene duplications. *The Evolution of the Genome.* T. R. Gregory and M. A. Burlington, Elsevier: 290-329.

Taylor, M. C., D. K. Muhia, D. A. Baker, A. Mondragon, P. B. Schaap and J. M. Kelly, (1999). *Trypanosoma cruzi* adenylyl cyclase is encoded by a complex multigene family. *Molecular and biochemical parasitology.* 104: 205-17.

Tetley, L. and K. Vickerman, (1985). Differentiation in *Trypanosoma brucei*: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat. *Journal of cell science.* 74: 1-19.

Theopold, U., M. Pinter, S. Daffre, Y. Tryselius, P. Friedrich, D. R. Nassel and D. Hultmark, (1995). CalpA, a *Drosophila* calpain homolog specifically expressed in a small set of nerve, midgut, and blood cells. *Molecular and cellular biology.* 15: 824-34.

- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins,** (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 25: 4876-82.
- Thompson, V. F. and D. E. Goll,** (2000). Purification of mu-calpain, m-calpain, and calpastatin from animal tissues. *Methods in molecular biology* (Clifton, N.J.). 144: 3-16.
- Tompa, P., Y. Emori, H. Sorimachi, K. Suzuki and P. Friedrich,** (2001). Domain III of calpain is a Ca^{2+} -regulated phospholipid-binding domain. *Biochemical and biophysical research communications*. 280: 1333-9.
- Tonami, K., Y. Kurihara, H. Aburatani, Y. Uchijima, T. Asano and H. Kurihara,** (2007). Calpain 6 is involved in microtubule stabilization and cytoskeletal organization. *Molecular and cellular biology*. 27: 2548-61.
- Tull, D., J. E. Vince, J. M. Callaghan, T. Naderer, T. Spurck, G. I. McFadden, G. Currie, K. Ferguson, A. Bacic and M. J. McConville,** (2004). SMP-1, a member of a new family of small myristoylated proteins in kinetoplastid parasites, is targeted to the flagellum membrane in *Leishmania*. *Molecular biology of the cell*. 15: 4775-86.
- Tyler, K. M., A. Fridberg, K. M. Toriello, C. L. Olson, J. A. Cieslak, T. L. Hazlett and D. M. Engman,** (2009). Flagellar membrane localization via association with lipid rafts. *Journal of cell science*. 122: 859-66.
- Tyler, K. M., P. G. Higgs, K. R. Matthews and K. Gull,** (2001). Limitation of *Trypanosoma brucei* parasitaemia results from density-dependent parasite differentiation and parasite killing by the host immune response. *Proceedings. Biological sciences / The Royal Society*. 268: 2235-43.
- Tyler, K. M., K. R. Matthews and K. Gull,** (2001). Anisomorphic cell division by African trypanosomes. *Protist*. 152: 367-78.
- Uliana, S. R., N. Goyal, E. Freymuller and D. F. Smith,** (1999). *Leishmania*: overexpression and comparative structural analysis of the stage-regulated meta 1 gene. *Experimental Parasitology*. 92: 183-91.
- Ullu, E., C. Tschudi and T. Chakraborty,** (2004). RNA interference in protozoan parasites. *Cellular microbiology*. 6: 509-19.
- Umeyama, T. and C. C. Wang,** (2008). Polo-like kinase is expressed in S/G2/M phase and associated with the flagellum attachment zone in both procyclic and bloodstream forms of *Trypanosoma brucei*. *Eukaryotic cell*. 7: 1582-90.
- Urwyler, S., E. Studer, C. K. Renggli and I. Roditi,** (2007). A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. *Molecular microbiology*. 63: 218-28.
- Urwyler, S., E. Vassella, J. Van Den Abbeele, C. K. Renggli, P. Blundell, J. D. Barry and I. Roditi,** (2005). Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. *PLoS pathogens*. 1: e22.
- Vale, R. D.,** (2003). The molecular motor toolbox for intracellular transport. *Cell*. 112: 467-80.

- Van Den Abbeele, J., Y. Claes, D. van Bockstaele, D. Le Ray and M. Coosemans,** (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology*. 118 (Pt 5): 469-78.
- Van der Peer, Y. and A. Meyer,** (2005). Large-scale gene and ancient genome duplications. *The Evolution of the Genome*. T. R. Gregory and M. A. Burlington, Elsevier: 330-371.
- Vassella, E., B. Reuner, B. Yutzy and M. Boshart,** (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *Journal of cell science*. 110 (Pt 21): 2661-71.
- Vaughan, S., T. Attwood, M. Navarro, V. Scott, P. McKean and K. Gull,** (2000). New tubulins in protozoal parasites. *Current biology* : CB. 10: R258-9.
- Vaughan, S., L. Kohl, I. Ngai, R. J. Wheeler and K. Gull,** (2008). A repetitive protein essential for the flagellum attachment zone filament structure and function in *Trypanosoma brucei*. *Protist*. 159: 127-36.
- Vedrenne, C., C. Giroud, D. R. Robinson, S. Besteiro, C. Bosc, F. Bringaud and T. Baltz,** (2002). Two related subpellicular cytoskeleton-associated proteins in *Trypanosoma brucei* stabilize microtubules. *Molecular biology of the cell*. 13: 1058-70.
- Veitch, N. J., P. C. Johnson, U. Trivedi, S. Terry, D. Wildridge and A. Macleod,** (2010). Digital gene expression analysis of two life cycle stages of the human-infective parasite, *Trypanosoma brucei gambiense* reveals differentially expressed clusters of co-regulated genes. *BMC Genomics*. 11: 124.
- Verhey, K. J. and J. W. Hammond,** (2009). Traffic control: regulation of kinesin motors. *Nature reviews. Molecular cell biology*. 10: 765-77.
- Vickerman, K.,** (1965). Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature*. 208: 762-6.
- Vickerman, K.,** (1969). On the surface coat and flagellar adhesion in trypanosomes. *Journal of cell science*. 5: 163-93.
- Vickerman, K.,** (1985). Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin*. 41: 105-14.
- Vickerman, K. and T. M. Preston,** (1976). *Biology of the kinetoplastida* W. H. R. a. E. Lumsden, D.A. London, Academic Press. 1: 35-130.
- Vickerman, K., L. Tetley, K. A. Hendry and C. M. Turner,** (1988). Biology of African trypanosomes in the tsetse fly. *Biology of the cell*. 64: 109-19.
- Vilei, E. M., S. Calderara, J. Anagli, S. Berardi, K. Hitomi, M. Maki and E. Carafoli,** (1997). Functional properties of recombinant calpain I and of mutants lacking domains III and IV of the catalytic subunit. *The Journal of biological chemistry*. 272: 25802-8.
- Wang, Z., J. C. Morris, M. E. Drew and P. T. Englund,** (2000). Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *The Journal of biological chemistry*. 275: 40174-9.

- Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton**, (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 25: 1189-91.
- Watt, F. and P. L. Molloy**, (1993). Specific cleavage of transcription factors by the thiol protease, m-calpain. *Nucleic Acids Research*. 21: 5092-100.
- Wickstead, B., K. Ersfeld and K. Gull**, (2002). Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 125: 211-6.
- Wickstead, B. and K. Gull**, (2006). A "holistic" kinesin phylogeny reveals new kinesin families and predicts protein functions. *Molecular biology of the cell*. 17: 1734-43.
- Wilson, C. A., J. Kreychman and M. Gerstein**, (2000). Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. *Journal of molecular biology*. 297: 233-49.
- Wilson, M. E., B. M. Young, K. P. Andersen, J. V. Weinstock, A. Metwali, K. M. Ali and J. E. Donelson**, (1995). A recombinant *Leishmania chagasi* antigen that stimulates cellular immune responses in infected mice. *Infection and immunity*. 63: 2062-9.
- Wirtz, E. and C. Clayton**, (1995). Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science (New York, N.Y.)*. 268: 1179-83.
- Wirtz, E., M. Hoek and G. A. Cross**, (1998). Regulated processive transcription of chromatin by T7 RNA polymerase in *Trypanosoma brucei*. *Nucleic Acids Research*. 26: 4626-34.
- Wirtz, E., S. Leal, C. Ochatt and G. A. Cross**, (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Molecular and biochemical parasitology*. 99: 89-101.
- Woods, A., A. J. Baines and K. Gull**, (1992). A high molecular mass phosphoprotein defined by a novel monoclonal antibody is closely associated with the intermicrotubule cross bridges in the *Trypanosoma brucei* cytoskeleton. *Journal of cell science*. 103 (Pt 3): 665-75.
- Woods, A., T. Sherwin, R. Sasse, T. H. MacRae, A. J. Baines and K. Gull**, (1989). Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *Journal of cell science*. 93 (Pt 3): 491-500.
- Woodward, R., M. J. Carden and K. Gull**, (1995). Immunological characterization of cytoskeletal proteins associated with the basal body, axoneme and flagellum attachment zone of *Trypanosoma brucei*. *Parasitology*. 111 (Pt 1): 77-85.
- Woodward, R. and K. Gull**, (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *Journal of cell science*. 95 (Pt 1): 49-57.

- Worthey, E. A., S. Martinez-Calvillo, A. Schnauffer, G. Aggarwal, J. Cawthra, G. Fazelinia, C. Fong, G. Fu, M. Hassebrock, G. Hixson, A. C. Ivens, P. Kiser, F. Marsolini, E. Rickel, R. Salavati, E. Sisk, S. M. Sunkin, K. D. Stuart and P. J. Myler,** (2003). *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Research*. 31: 4201-10.
- Wu, Y., X. Wang, X. Liu and Y. Wang,** (2003). Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite. *Genome research*. 13: 601-16.
- Xie, X., M. D. Dwyer, L. Swenson, M. H. Parker and M. C. Botfield,** (2001). Crystal structure of calcium-free human sorcin: a member of the penta-EF-hand protein family. *Protein science : a publication of the Protein Society*. 10: 2419-25.
- Xu, Y. and R. L. Mellgren,** (2002). Calpain inhibition decreases the growth rate of mammalian cell colonies. *The Journal of biological chemistry*. 277: 21474-9.
- Yamaguchi, R., M. Maki, M. Hatanaka and H. Sabe,** (1994). Unphosphorylated and tyrosine-phosphorylated forms of a focal adhesion protein, paxillin, are substrates for calpain II in vitro: implications for the possible involvement of calpain II in mitosis-specific degradation of paxillin. *FEBS letters*. 356: 114-6.
- Yamazaki, M., K. Ishidoh, Y. Suga, T. C. Saïdo, S. Kawashima, K. Suzuki, E. Kominami and H. Ogawa,** (1997). Cytoplasmic processing of human profilaggrin by active mu-calpain. *Biochemical and biophysical research communications*. 235: 652-6.
- Yang, J. T., R. A. Laymon and L. S. Goldstein,** (1989). A three-domain structure of kinesin heavy chain revealed by DNA sequence and microtubule binding analyses. *Cell*. 56: 879-89.
- Yao, X., A. Thibodeau and J. G. Forte,** (1993). Ezrin-calpain I interactions in gastric parietal cells. *The American journal of physiology*. 265: C36-46.
- Yeats, C., S. Bentley and A. Bateman,** (2003). New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiology*. 3: 3.
- Yoshikawa, Y., H. Mukai, F. Hino, K. Asada and I. Kato,** (2000). Isolation of two novel genes, down-regulated in gastric cancer. *Japanese journal of cancer research : Gann*. 91: 459-63.
- Yu, H., J. K. Chen, S. Feng, D. C. Dalgarno, A. W. Brauer and S. L. Schreiber,** (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*. 76: 933-45.
- Zamore, P. D. and B. Haley,** (2005). Ribo-gnome: the big world of small RNAs. *Science (New York, N.Y.)*. 309: 1519-24.
- Ziegelbauer, K., M. Quinten, H. Schwarz, T. W. Pearson and P. Overath,** (1990). Synchronous differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms in vitro. *European journal of biochemistry / FEBS*. 192: 373-8.