Srivastava *et al.* – Supplemental material

Table S1. List of oligonucleotides used in quantitative DNA/RNA analysis

Application	Target	Sequence
RT-[q]PCR	*TFL1* mRNA	5′-GATCATTCTGCCTCAGTGC-3′ 5′-GATATTGCCATCCTCTGACG-3′
RT-[q]PCR	*TFL2* mRNA	5′-GATGGAAGGTCTGCTGTACC-3′ 5′-CTCAACAGCCTCTTGTCTGG-3′
RT-[q]PCR / ChIP	α tubulin coding region	5′-GTGCATTGAACGTGGATCTG-3′ 5′-GCCTACCACGAGCAACTCTC-3′
RT-[q]PCR / ChIP qPCR	*SLRNA* promoter	5′-CTACCGACACATTTCTGGC-3′ 5′-GGTATGAGAAGCTCCCAGTAGCAGC-3′
RT-[q]PCR / ChIP qPCR	*SLRNA* intergenic region	5′-ATGGCTTATACGTGCTCGTTTCTCC-3′ 5′-CACATATAGGCGCTTTAAAGTCTGCT-3′
RT-[q]PCR / ChIP qPCR	*GPEET* procyclin promoter region	5′-TGATAGGTATCTCTTATTAGTATAG-3′ 5′-GGGGTTATCGGGTGAGTAC-3′
RT-[q]PCR ChIP qPCR	U2 snRNA	5′-ATATCTTCTCGGCTATTTAGC-3′ 5′-ACAGGCAACAGTTTTGATCC-3′
RT-[q]PCR ChIP qPCR	β–α tubulin intergenic region	5′-GCTGATTTCTGACAGATCTTCAAAC-3′ 5′-GTGGATGCAGATAGCCTCACGCATG-3′
Primer extension of *in vitro* transcribed RNA	SL RNA GPEET-trm RNA	SLtag 5′-GCCTGGCGCCATACCATGG-3′ Tag_PE 5′-GAGTGAATGATGATAGATTTG-3′
Primer extension of total endogenous RNA	SL RNA U2 snRNA	SL_PE 5′-CGACCCCACCTTCCAGATTC-3′ U2f 5′-ACAGGCAACAGTTTTGATCC-3′

For each PCR primer pair, the forward primer is listed on top and the reverse primer on the bottom.

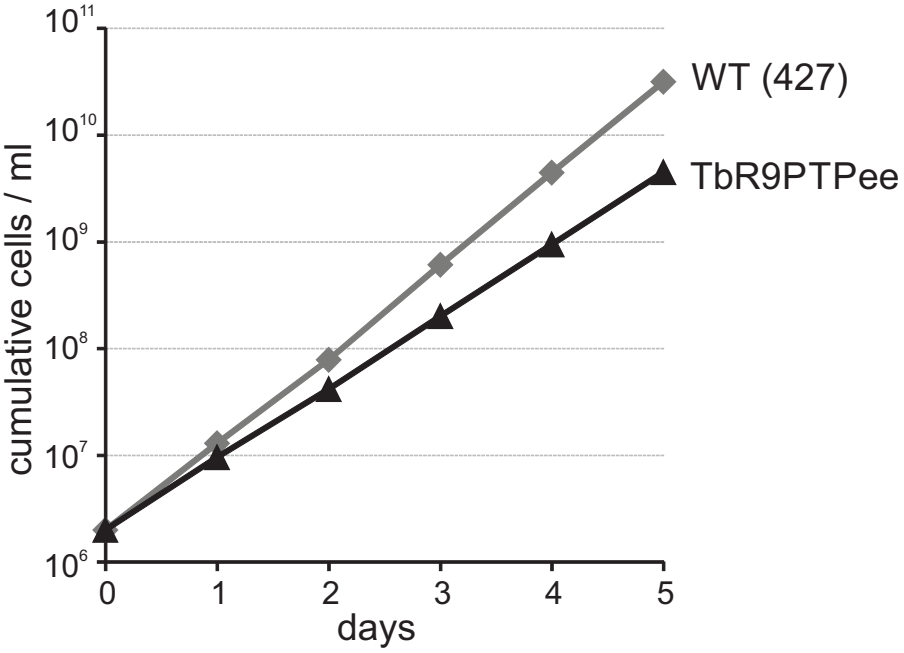


Figure S1. Reduced growth of TbR9PTPee culture. Cell densities of wild-type 427 procyclic trypanosomes (WT) and their derivative TbR9PTPee cells were determined and cultures diluted with medium to a density of $2x10^6$ cells daily. The doubling time of wild-type cells of 8.59 h was increased in TbR9PTPee cells to 10.78 h, suggesting that either the tag has a mild effect on RPB9 function or deleting one *RPB9* allele (see Figure 1A) resulted in haplo-insufficiency.

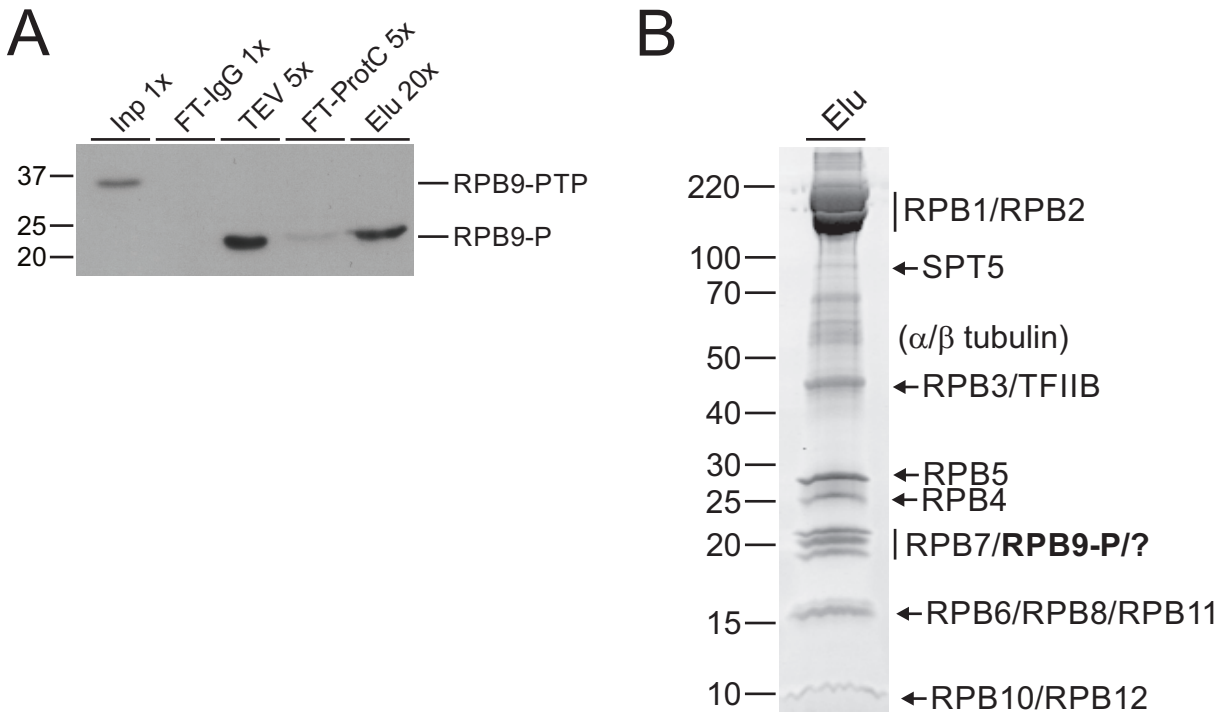
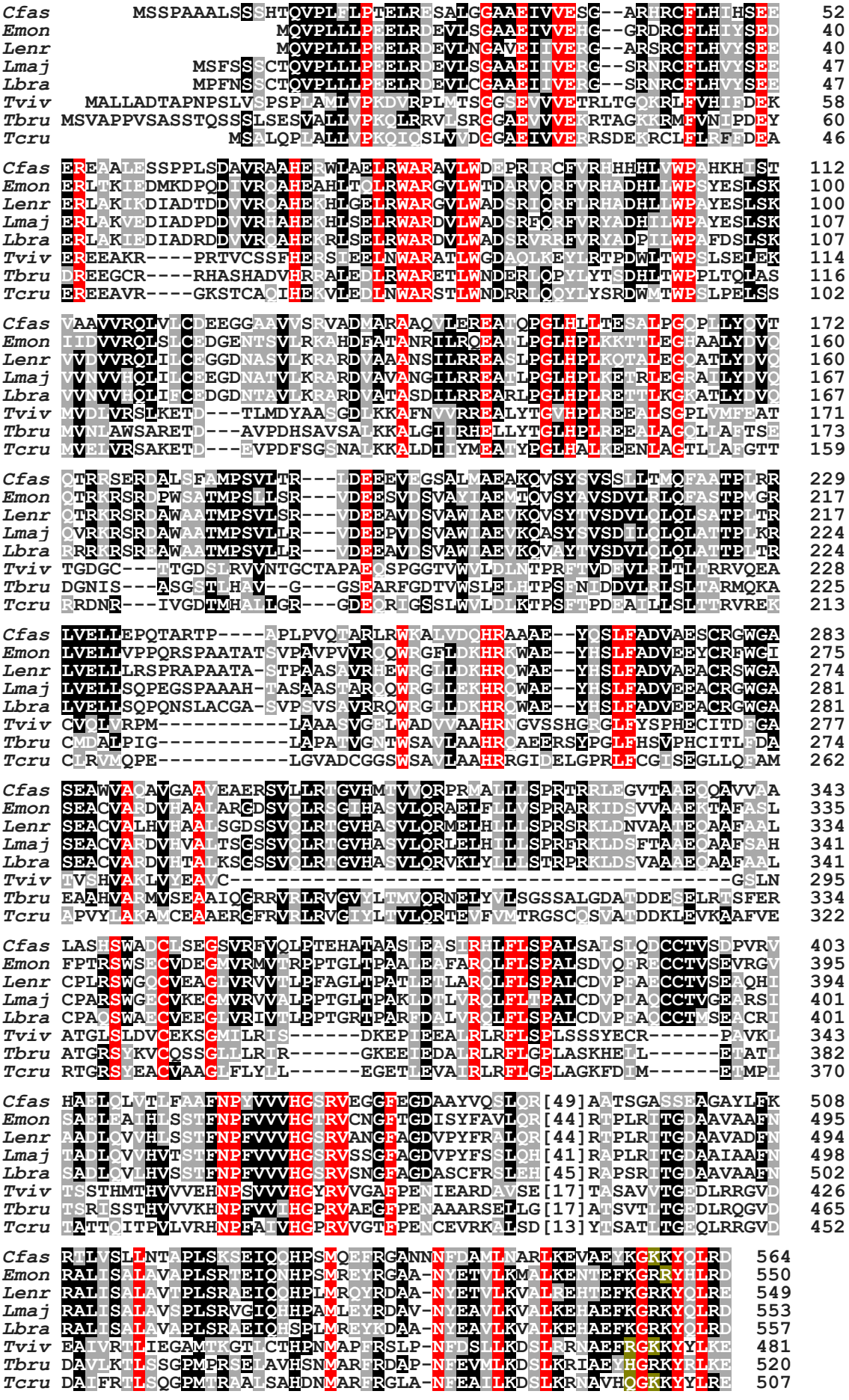


Figure S2. 2nd RPB9-PTP purification. RPB9-PTP was tandem affinity-purified from extract of TbR9PTPee cells. (A) Immunoblot analysis of PRP9-PTP and PRP9-P with anti-ProtC antibody in crude extract (Inp), flow-through of the IgG column (FT-IgG), TEV protease eluate (TEV), flow-through of the anti-ProtC immunoaffnity column (FT-ProtC), and the final eluate (Elu). x-Values indicate relative amounts loaded. (B) The final eluate (Elu) was separated on a SDS/10-20% polyacrylamide gradient gel and stained with Sypro Ruby. The gel lane was cut into several slices and analyzed by LC/MS/MS. Band annotation is according to the pattern analyzed in Figure 1C.

A


Cfas MSSPAAALSSSHTQVPLFLPTELRESALGGAAEIVVESG--ARHRCFLHIHSEE 52
Emon MQVPLLLLPEELRDEVLSGAAEIVVEFG--GRDRCFLHIYSED 40
Lenr MQVPLLLLPEELRDEVLNGAVEIIVERG--ARSRCFLHVYSEE 40
Lmaj MSFSSSCTQVPLLLPEELRDEVLSGAAEIIVERG--SRNRCFLHVYSEE 47
Lbra MPFNSSCTQVPLLLPEELRDEVLCGAAEIIVERG--SRNRCFLHVYSEE 47
Tviv MALLADTAPNPSLVSPSPLAMLVPKDVRPLMTSGGSEVVVETRLTGQKRLFVHIEDEK 58
Tbru MSVAPPVSASSTQSSSLSESVALLVPKQLRRVLSRGGAEVVVEKRTAGKKRMFVNIPDEY 60
Tcru MSALQPLALLVPKQIQSLVVDGGAEIVVERRSDEKRCLFLRFEDEA 46
Cfas EREAALESSPPLSDAVRAAHERWLAELRWARAVLWDEPRIRCFVRHHHHLVWPAHKHIST 112
Emon ERLTKIEDMKDPQDIVRQAHEAHLTQLRWARGVLWTDARVQRFVRHADHLLWPSYESLSK 100
Lenr ERLAKIKDIADTDDVVRQAHEKHLGELRWARGVLWADSRIQRFLRHADHLLWPAYESLSK 100
Lmaj ERLAKVEDIADPDDVVRGAHEKHLSELRWARDVLWADSRFQRFVRYADHILWPAYESLSK 107
Lbra ERLAKIEDIADRDDVVRQAHEKRLSELRWARDVLWADSRVRRFVRYADPILWPAFDSLSK 107
Tviv EREEAKR----PRTVCSSFHERSLEELNWARATLWGDAQLKEYIRTPDWLTWPSLSELEK 114
Tbru DREEGCR----RHASHADVHRRALEDLRWARETLWNDERLQPYLYTSDHLTWPPLTQLAS 116
Tcru EREEAVR----GKSTCAQIHEKVLEDLNWARSTLWNDRRLQQYLYSRDWMTWPSLPELSS 102
Cfas VAAVVRQLVLCDEEGGAAVVSRVADMARAAQVLEREAIQPGLHLLTESALPGQPLLYQVT 172
Emon IIDVVRQLSLCEDGENTSVLRKAHDFATANRILRQEAILPGLHPLKKTTLEGHAALYDVQ 160
Lenr VVDVVRQLILCEGGDNASVLKRARDVAAANSILRREASLPGLHPLKQTALEGQATLYDVQ 160
Lmaj VVNVVHQLILCEEGDNATVLKRARDVAVANGILRREAILPGLHPLKETRLEGRAILYDVQ 167
Lbra VVNVVHQLIFCEDGDNTAVLKRARDVATASDILRREARLPGLHPLRETTLKGKATLYTVQ 167
Tviv MVDLVRSLKETD---TLMDYAASGDLKKAFNVVRREALYTGVHPLREEALSGPLVMFEAT 171
Tbru MVNLAWSARETD---AVPDHSAVSALKKALGIIRHELLYTGLHPLREEALAGQLLAETSE 173
Tcru MVELVRSAKETD---EVPDFSGSNALKKALDIIYMEAIYPGLHALKEENLAGTLLAEGTT 159
Cfas QTRRSERDALSFAMPSVLTR---IDEEEVEGSALMAEAKQVSYSVSSLLTMQFAATPTRR 229
Emon QTRKRSRDPWSATMPSLLSR---VDEESVDSVAYIAEMTQVSYAVSDVLRLQFASTPVGR 217
Lenr QTRKRSRDAWAATMPSVLSR---VDEEAVDSVAWIAEVKQVSYTVSDVLQLQLSATPLTR 217
Lmaj QVRKRSRDAWAATMPSVLLR---VDEEPVDSVAWIAEVKQASYSVSDILQLQLATTPIKR 224
Lbra RRKKRSREAWAATMPSVLLR---VDEEAVDSVAWIAEVKQVAYTVSDVLQLQLATTPITR 224
Tviv TGDGC---TTGDSLRVVNTGCTAPAEQSPGGTVWVLDLNTPRFTVDEVLRITLTRRVQEA 228
Tbru DGNIS---ASGSTLHAV--G---GSEARFGDTVWSLELHTPSENIDDVLRISLTARMQKA 225
Tcru RRDNR---IVGDTMHAILGR---GDEQRIGSSLWVLDIKTPSFTPDEAILISLTTRVREK 213
Cfas LVELLEPQTARTP----APLPVQTARLRWKALVDQHRAAAE--YQSLFADVAESCRGWGA 283
Emon LVELLVPPQRSPAATATSVPAVPVVRQQWRGFLDKHRKWAE--YESLFADVEEYCRFWGI 275
Lenr LVELLLRSPRAPAATA-STPAASAVRHEWRGLLDKHRQWAE--YHSLFADVAEACRSWGA 274
Lmaj LVELLSQPEGSPAAAH-TASAASTARQQWRGLLEKHRQWAE--YHSLFADVEEACRGWGA 281
Lbra LVELLSQPQNSLACGA-SVPSVSAVRRQWRGLLDKHRQWAE--YHSLFADVEEACRGWGA 281
Tviv CVQLVRPM-----------LAAASVGELWADVVAAHRNGVSSHGRGLFYSPHECITDFGA 277
Tbru CMDALPIG-----------LAPATVGNTWSAVLAAHRQAEERSYPGLFHSVPHCITLEDA 274
Tcru CLRVMQPE-----------LGVADCGGSWSAVLAAHRRGIDELGPRLFCGISEGLLQEAM 262
Cfas SEAWVAQAVGAAVEAERSVLLRIGVHMTVVQRPRMALLLSPRTRRLEGVTAAEQQAVVAA 343
Emon SEACVARDVEAALARGDSVQLRSGTHASVLQRAELFLLVSPRARKIDSVVAAEKTAFASL 335
Lenr SEACVALHVEAALSGDSSVQLRIGVHASVLQRMELHLLLSPRSRKLDNVAAIEQAAFAAL 334
Lmaj SEACVARDVEVALTSGSSVQLRIGVHASVLQRLELHLLLSPRFRKLDSFTAAEQAAFSAH 341
Lbra SEACVARDVETALKSGSSVQLRIGVHASVLQRVKLYLLLSTRPRKLDSVAAAEQAAFAAL 341
Tviv TVSHVAKLVYEAVC--GSLN 295
Tbru EAAHVARMVSEAAIQGRRVRLRVGVYLTMVQRNELYVLSGSSALGDATDDESELRTSFER 334
Tcru APVYLAKAVCEAAERGFRVRLRVGIYLTVLQRTEVFVMTRGSCQSVATDDKLEVKAAFVE 322
Cfas LASHSWADCLSEGSVRFVQLPTEHATAASLEASIRHLFLSPAISALSLQDCCTVSDPVRV 403
Emon FPTRSWSECVDEGMVRMVTRPPTGLTPAALEAFARQLFLSPALSDVQFRECCTVSEVRGV 395
Lenr CPLRSWGQCVEAGLVRVVTLPFAGLTPATLETLARQLFLSPALCDVPFAECCTVSEAQHI 394
Lmaj CPARSWGECVKEGMVRVVALPPTGLTPAKLDTLVRQLFLTPAICDVPLAQCCTVGEARSI 401
Lbra CPAQSWAECVEEGLVRIVTLPPTGRTPARLDALVRQLFLSPAICDVPFAQCCTMSEACRI 401
Tviv ATGLSLDVCEKSGMILRIS------DKEPIEEALRLRFLSPLSSSYECR------PAVKL 343
Tbru ATGRSYKVCQSSGLLLRIR------GKEEIEDALRLRFLGPLASKHELL------ETATL 382
Tcru RTGRSYEACVAAGLFLYLL------EGETIEVAIRLRFLGPLAGKFDIM------EIMPL 370
Cfas HAELQLVTLFAAFNPYVVVHGSRVEGGFEGDAAYVQSLQR[49]AATSGASSEAGAYLEK 508
Emon SAELEAIHLSSTFNPFVVVHGTRVCNGFTGDISYFAVLQR[44]RTPLRITGDAAVAAFN 495
Lenr AADLQVVHLSSTFNPFVVVHGSRVANGFAGDVPYFRALQR[44]RTPLRITGDAAVADFN 494
Lmaj TADLQVVHVTSTFNPFVVVHGSRVSSGFAGDVPYFSSLQH[41]RAPLRITGDAAIAAFN 498
Lbra SADLQVLHVSSTFNPFVVVHGSRVSNGFAGDASCFRSLEH[45]RAPSRITGDAAVAAFN 502
Tviv TSSTHMTHVVVEHNPSVVVHGYRVVGAFPENIEARDAVSE[17]TASAVVTGEDLRRGVD 426
Tbru TSRISSTHVVVKHNPFVVIHGPRVAEGFPENAAARSELLG[17]ATSVTLTGEDLRQGVD 465
Tcru TATTQITPVLVRHNPEAIVHGPRVVGTEPENCEVRKALSD[13]YTSATLTGEQLRRGVD 452
Cfas RILVSLLNTAPLSKSEIQQHPSMQEFRGANNNFDAMLNARLKEVAEYKGKKYQLRD 564
Emon RALISALAVAPLSRIEIQNHPSMREYRGAA-NYETVLKMALKENTEFKGRRYHLRD 550
Lenr RALISALAVTPLSRAEIQQHPLMRQYRDAA-NYETVLKVALREHTEFKGRKYQLRE 549
Lmaj RALISALAVSPLSRVGIQHHPAMLEYRDAV-NYEAVLKVALKEHAEFKGRKYQLRD 553
Lbra RALISALAVAPLSRAEIQHSPLMREYKDAA-NYEAVLKVALKEHAEFKGRKYQLRD 557
Tviv EAIVRTLIEGAMTKGTLCTHPNMAPFRSLP-NFDSLLKDSLRRNAEFRGKKYYLKE 481
Tbru DAVLKTLSSGPMPRSELAVHSNMARFRDAP-NFEVMLKDSLKRIAEYHGRKYRLKE 520
Tcru DAIFRTLSQGPMTRAALSAHDNMARFRGLA-NFEAILKDSLKRNAVHQGKKYYLRE 507

B

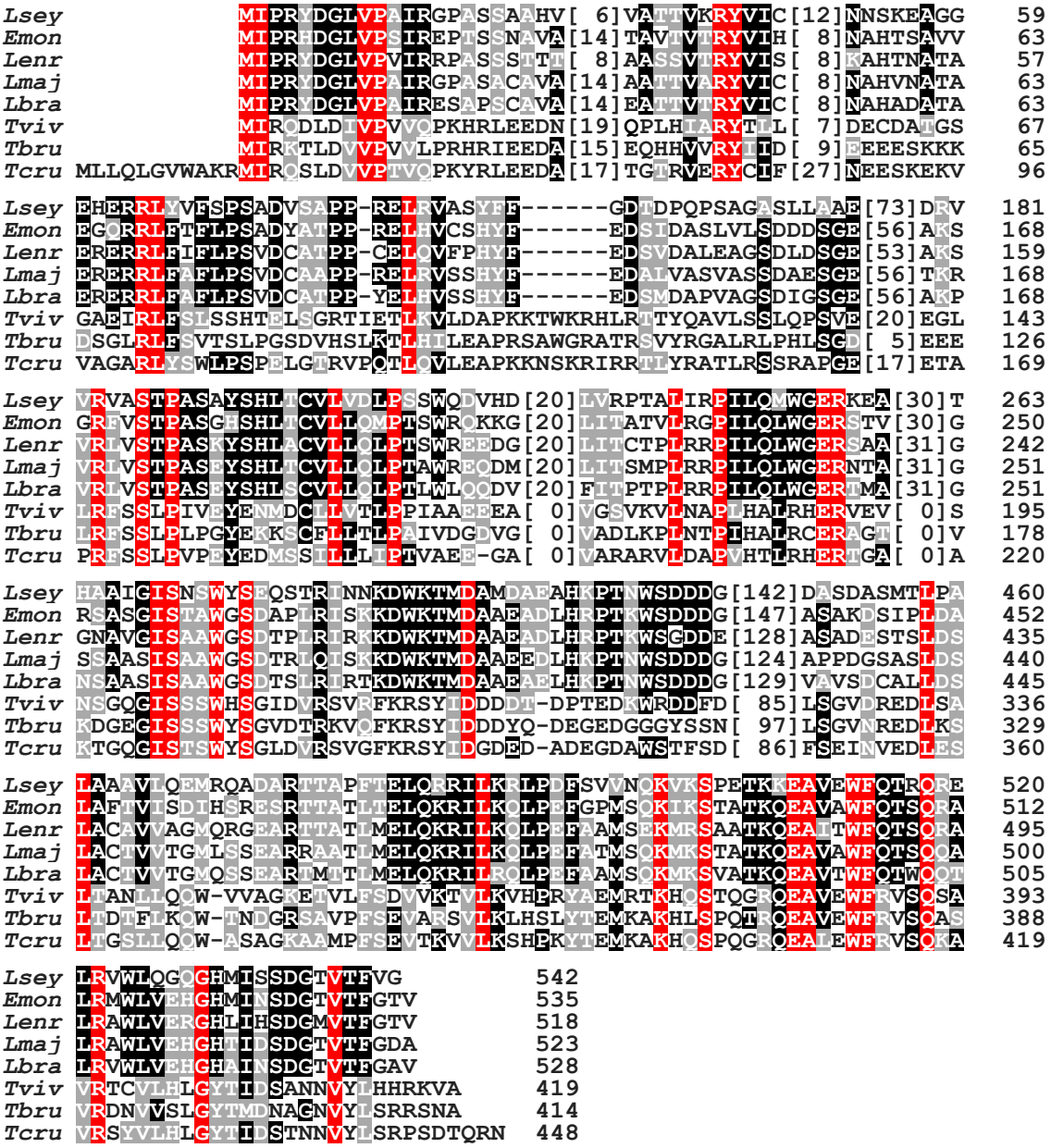


Figure S3. Multiple sequence alignment of trypanosomatid TFL1 (**A**) and TFL2 (**B**) orthologs. Amino acid sequences were aligned using the Clustal Omega server of the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=clustalo) at default parameters (1). Positions with more than 50% identity or similarity are highlighted in black or gray, respectively, and completely identical positions are highlighted in red. Dashes indicate missing residues. Numbers in parentheses specify number of residues without obvious sequence similarity. TFL1-ortholgous sequences are from *Crithdia fasciculata* (*Cfas*, accession number CFAC1_230048500), *Endotrypanum monterogeii* strain LV88 (*Emon*, EMOLV88_270008900), *Leishmania enriettii* (*Lenr*, LENLEM3045_270009800), *Leishmania major* (*Lmaj*, LmjF27.0440), *Leishmania braziliensis* (*Lbra*, LbrM.27.0530), *Trypanosoma vivax* (*Tviv*, TvY486_1100930), and *Trypanosoma cruzi* (*Tcru*, TcCLB.506925.500) whereas TFL2-orthologous sequences are from *Leptomonas seymouri* (*Lsey*; Lsey_0142_0040), *E. monterogeii* (EMOLV88_320027800), *L. enriettii* (LENLEM3045_ 320031900), *L. major* (LmjF32.2390), *L. braziliensis* (LbrM.32.2620), *T. vivax* (tviv1020g12.q1k_6), and *T. cruzi* (TcCLB.504159.60).

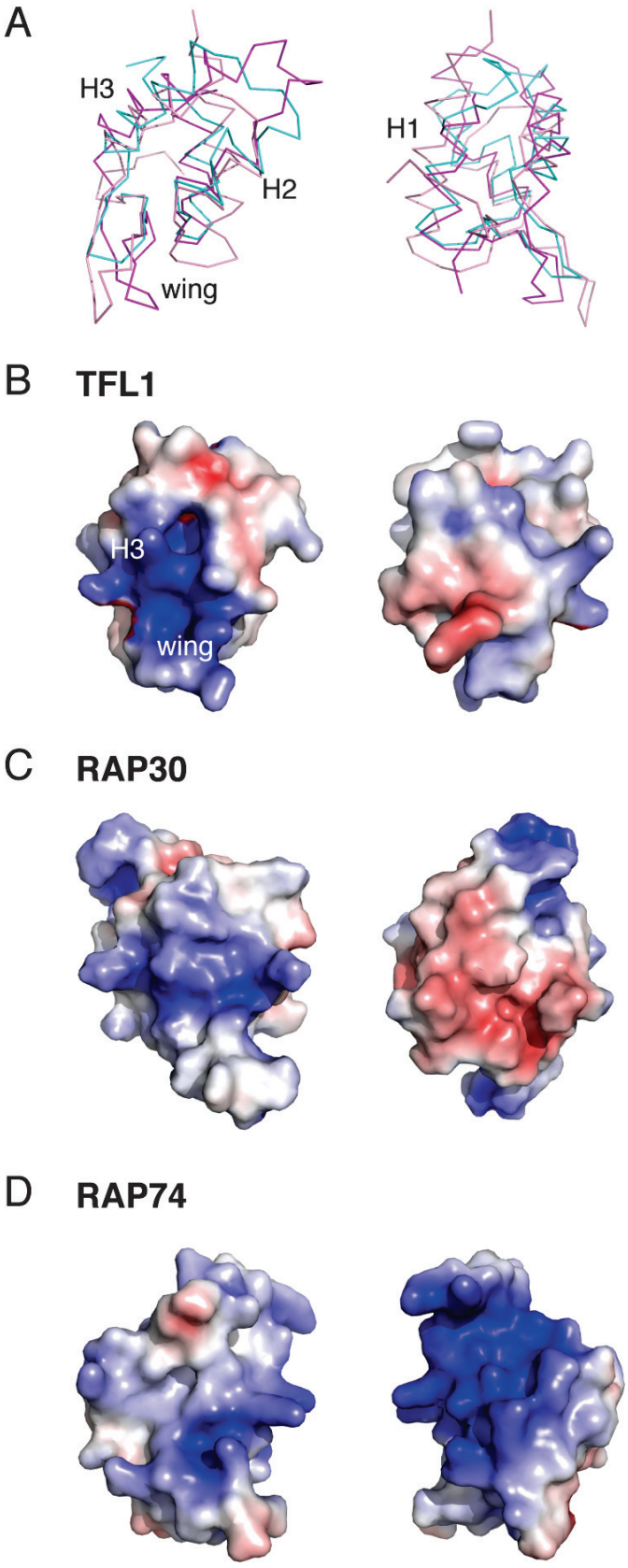


Figure S4. Surface electrostatic properties of the WH domains in TFL1, RAP30 and RAP74. (**A**) α-carbon superposition of TFL1 (cyan) with RAP30 (pink) and RAP74 (magenta). (**B-D**) Molecular surface representation of the WH domains, colored according to the local electrostatic potential (positive, blue; negative, red), in an orientation similar to the one above. The left and right images are related by 180º rotations about the vertical.

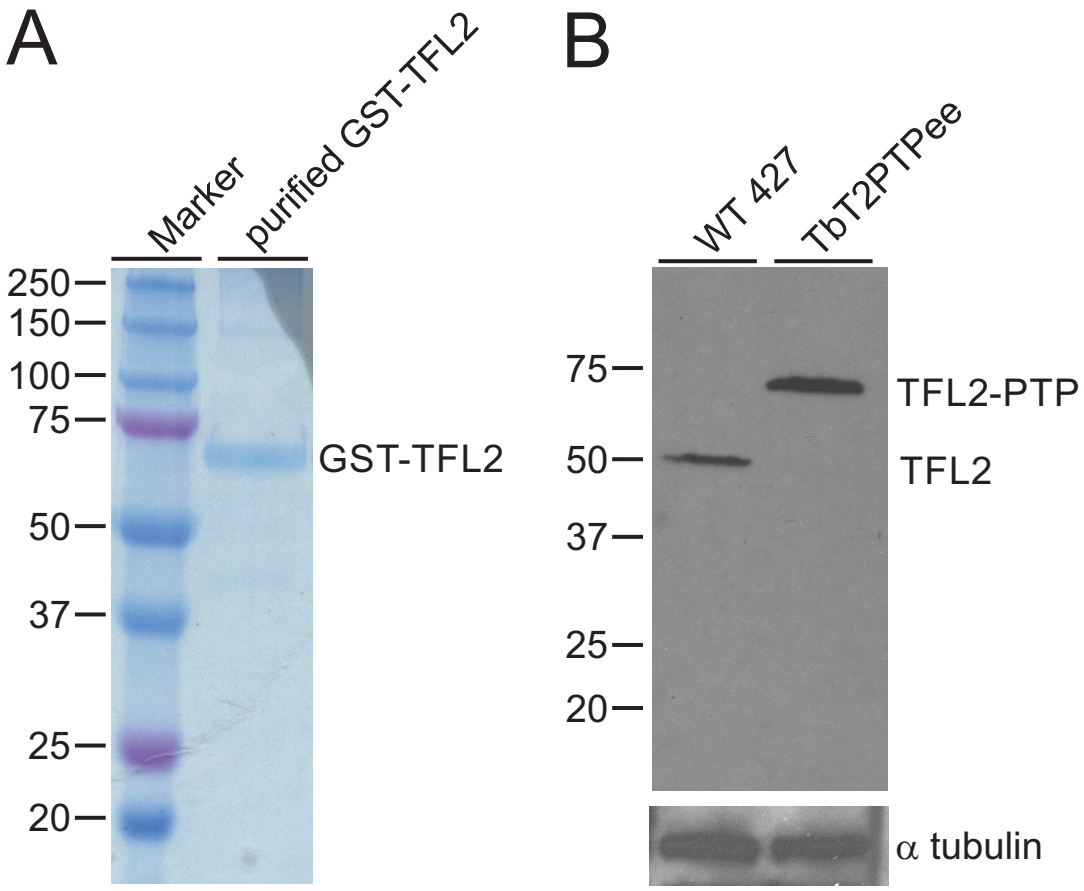


Figure S5. Rat anti-TFL2 immune serum. (**A**) The coding sequence of *Trypanosoma brucei brucei strain 427* TLF2 was cloned into pGEX-6P-1 (GE Healthcare Life Sciences) downstream of the gluthathione S-tranferase (GST) coding sequence and transformed into the BL21 strain of *E. coli* (MilliporeSigma). The recombinant GST-TLF2 protein was purified by gluthathione affinity chromatography using Glutathione Sepharose 4 Fast Flow beads (GE Healthcare Life Sciences) according to the manufacturer's specifications and by concentration of the eluate in an Amicon Ultracel-30 centrifugal filter (MilliporeSigma). (**B**) The purified protein was injected into the rat bloodstream, and immune serum was raised and obtained according to a published protocol (2). The immune serum was used to detect TLF2 in whole cell lysates of wild-type trypanosomes (WT 427) and in TbT2PTPee cells that exclusively express tagged TFL2-PTP and no untagged protein. As a loading control, α tubulin was detected on the same blot.

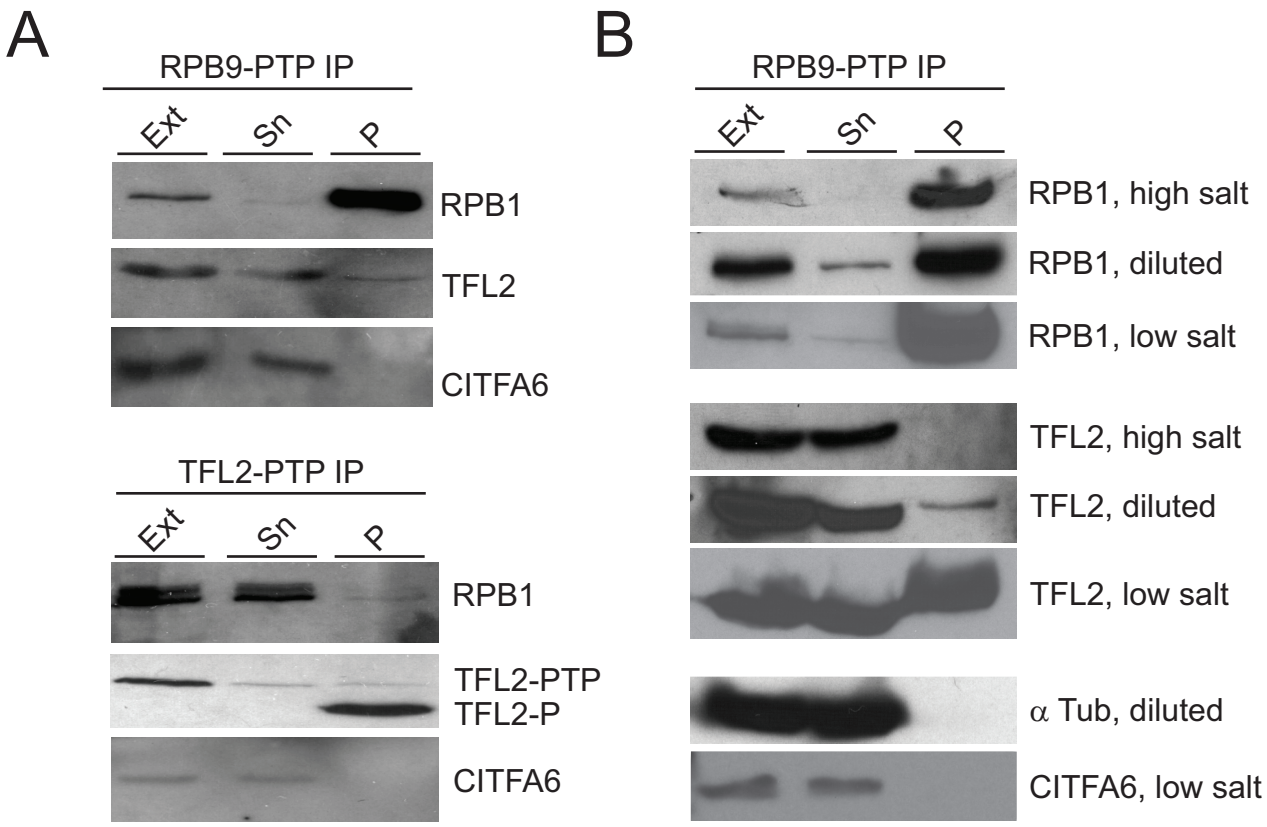
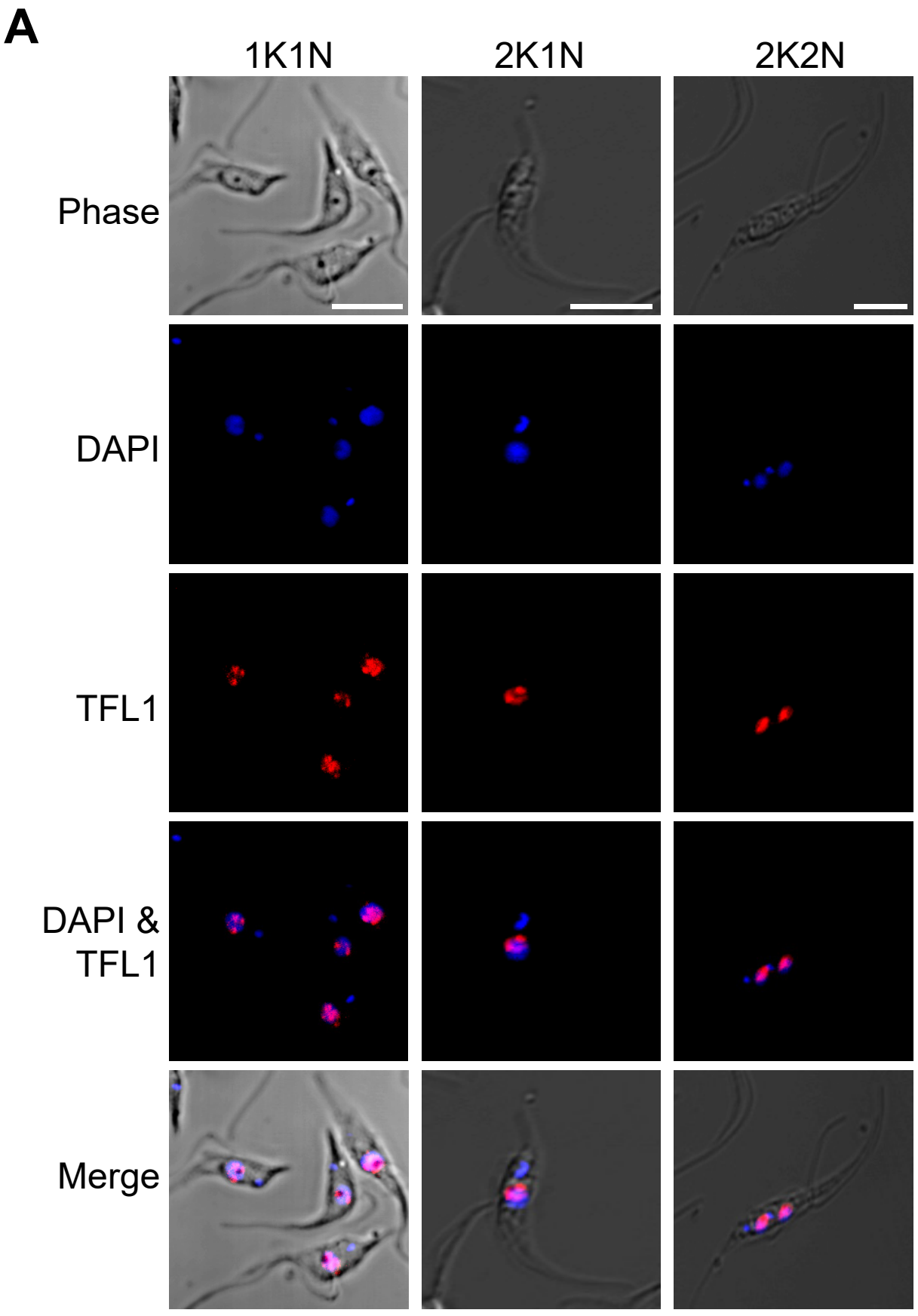


Figure S6. RNA pol II and TFL2 co-precipitate in a salt-sensitive manner. (**A**) From extract of TbR9PTPee cells and from extract of TbT2PTPee cells, RPB9-PTP and TFL2-PTP were precipitated with IgG beads, respectively. RNA pol II and TFL2 [co-]precipitation was analyzed by immunoblotting with rat-derived specific immune sera [(3,4); this study], detecting the largest RNA pol II subunit RPB1, tagged or untagged TFL2, and, as a negative control, the RNA pol I transcription factor subunit CITFA6 in extract (Ext), supernatant (Sn) and precipitate (P). Five times more precipitate was loaded relative to extract and supernatant. (**B**) The RPB9-PTP precipitation was repeated with undiluted extract (high salt, ~300 mM KCl), with extract that was diluted 1:1 with transcription buffer (150 mM sucrose, 20 mM HEPES-KOH, pH 7.7, 20 mM potassium L-glutamate, 3 mM $MgCl_2$, 1 mM DTT, 10 µg/ml aprotinin, 10 µg/ml leupeptin) just prior to precipitation, and with extract that was generated by adding only 1/20 instead of 1/10 the volume of extraction buffer (150 sucrose, 20 mM HEPES-KOH, pH 7.7, 1500 mM KCl, 3 mM $MgCl_2$) to broken cell suspension (low salt). In addition to CITFA6, the stronger expressed α tubulin (α Tub) was detected as a negative control. Tenfold more precipitate relative to extract and supernatant was loaded.

The nearly undetectable and strong co-precipitation of TFL2 with RPB9-PTP in high and low-salt extracts, respectively, suggests that the RPB9-PTP/TLF complex was disrupted by high salt extraction and partially reformed upon diluting the extract with transcription buffer whereas it remained largely intact when only half the salt was used for extraction.


A
1K1N
2K1N
2K2N
Phase
DAPI
TFL1
DAPI &
TFL1
Merge

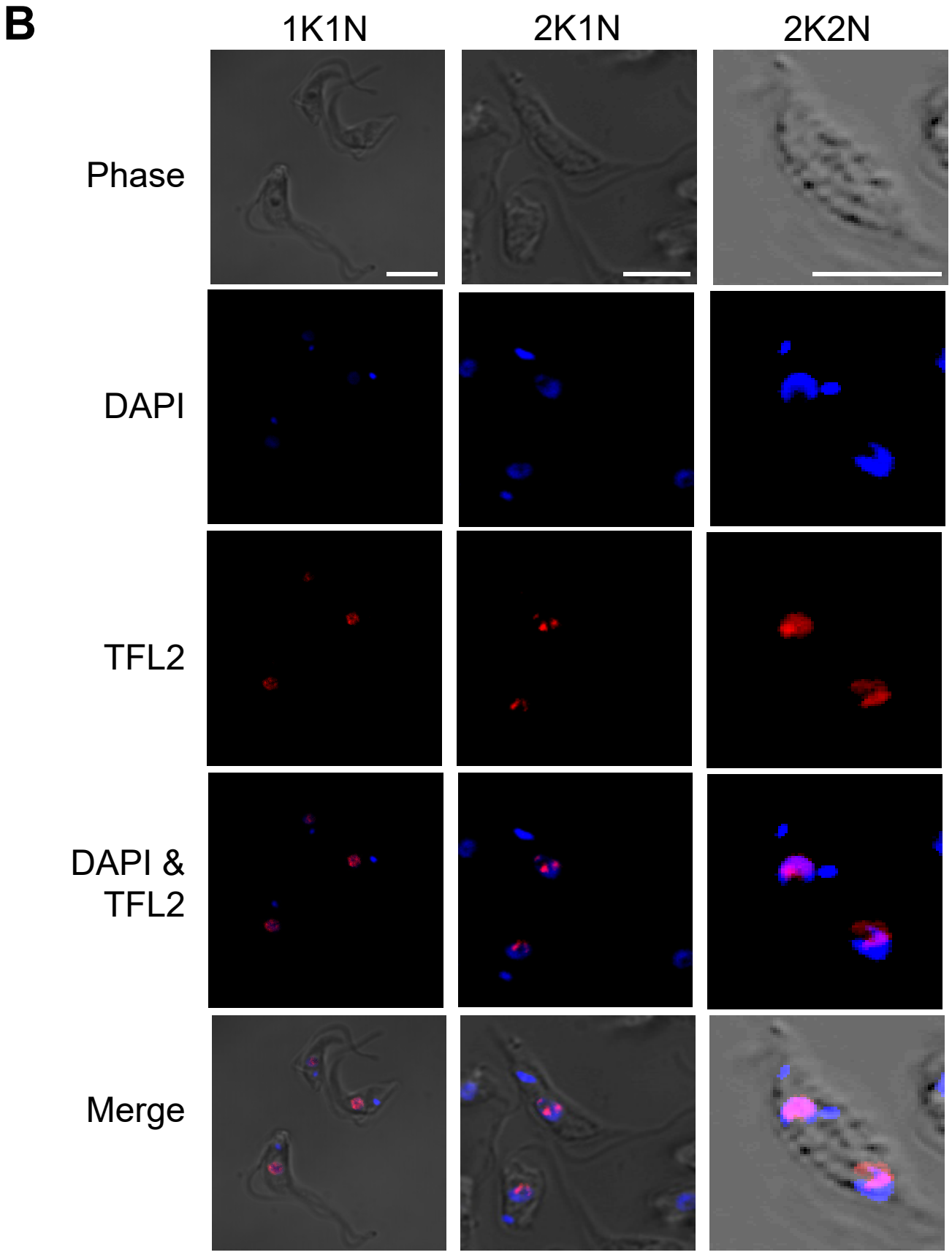


Figure S7. Localization of TLF1-PTP (**A**) and TLF2-PTP (**B**) in procyclic trypanosomes (red) which are in different cell cycle stages, having one nucleus and one kinetoplast (1K1N; G1 phase), 2K1N (late S phase) and 2K2N (postmitotic cells). DNA was stained with DAPI (blue) showing nuclei and smaller kinetoplasts. The nucleolus can be recognized within a nucleus as a spherical structure of low DNA density. White bars in top panels represent 5 μm.

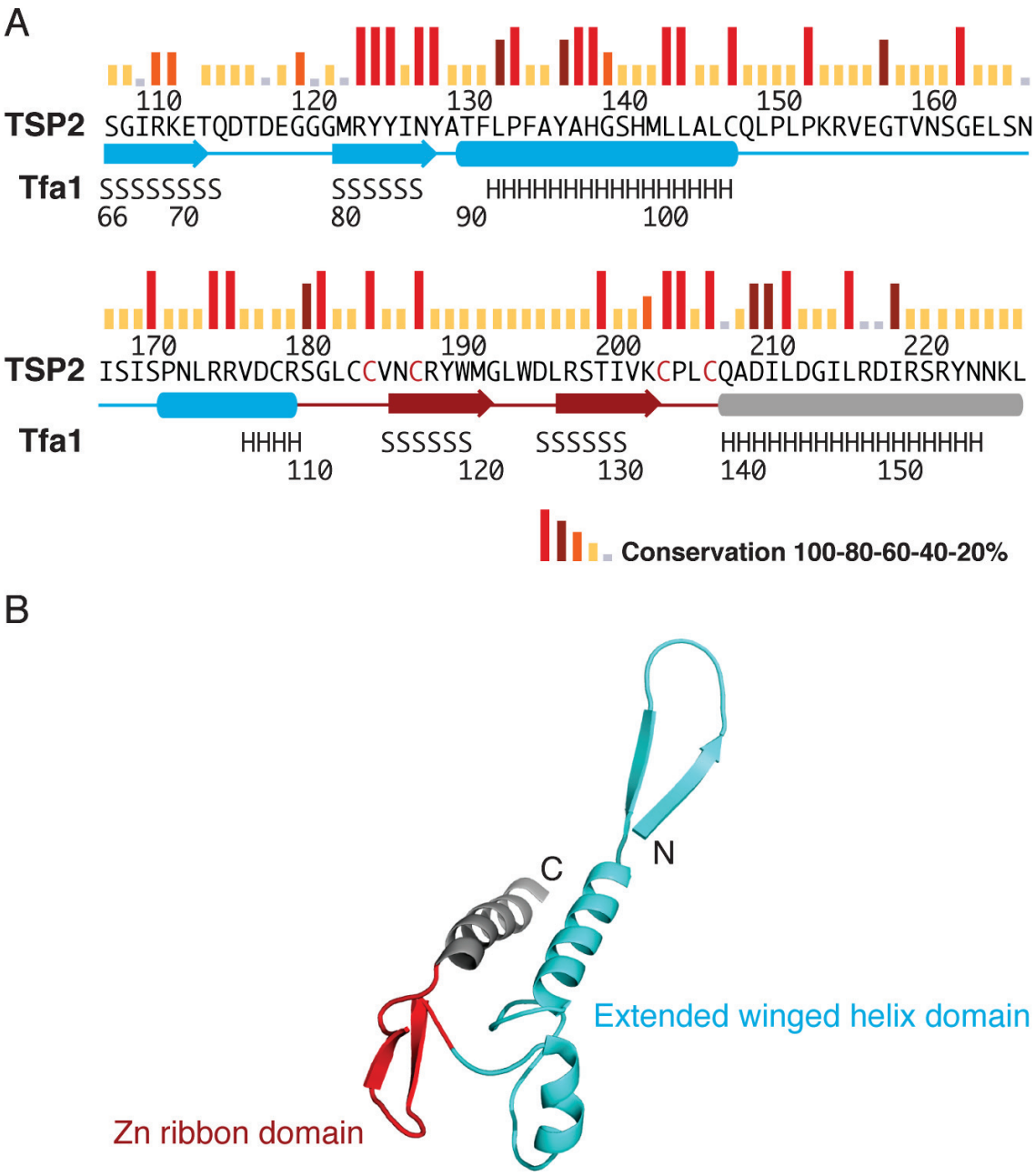


Figure S8. Sequence conservation and homology model of the N-terminal domain of TSP2. (**A**) Sequence conservation and predicted secondary-structure elements for the N-terminal domain of TSP2. Sequence conservation is shown as a bar graph, with red bars indicting identity among six trypanosomatid orthologs from *Trypanosoma brucei* (Tb927.11.14110), *Trypanosoma congolense* (TcIL3000.11.14400), *Trypanosoma cruzi* (TcCLB.511727.150), *Leishmania major* (LmjF32.0860), *Leishmania infantum* (LinJ.32.0910), and *Leishmania brazilinsis* (LbrM.32.0950). Secondary-structure assignments predicted by PHYRE 2 are shown as cylinders (α helices) and arrows (β strands). Secondary-structure assignments from the high-resolution cryo-EM structure of the N-terminal domain of Tfa1 (TFIIEα) in a yeast initiation complex (PDB ID: 5FYW) are indicated by letters ("H": α helix; "S": β strand). (**B**) Ribbon diagram of top scoring PHYRE homology model of the N-terminal domain of TSP2, with the extended winged helix domain colored in cyan and the zinc ribbon domain in red.

Supplemental References

1. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
2. Schimanski, B., Brandenburg, J., Nguyen, T.N., Caimano, M.J. and Günzl, A. (2006) A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. *Nucleic Acids Res.*, **34**, 1676-1684.
3. Badjatia, N., Ambrósio, D.L., Lee, J.H. and Günzl, A. (2013) Trypanosome cdc2-related kinase 9 controls spliced leader RNA cap4 methylation and phosphorylation of RNA polymerase II subunit RPB1. *Mol. Cell. Biol.*, **33**, 1965-1975.
4. Nguyen, T.N., Nguyen, B.N., Lee, J.H., Panigrahi, A.K. and Günzl, A. (2012) Characterization of a novel class I transcription factor A (CITFA) subunit that is indispensable for transcription by the multifunctional RNA polymerase I of *Trypanosoma brucei*. *Eukaryot. Cell*, **11**, 1573-1581.