SUPPLEMENTAL MATERIAL

Experimental procedures	p. 1-3
Figures S1-S6	р. 4-6

Experimental procedures.

DNA extraction. DNA was extracted from 10^6 *T. brucei* TC221 cells by the following procedure. The trypanosomes were centrifuged at 3,000 g for 3 min and resuspended in 100 µl of 50 mM Tris-HCl pH 7.6 supplemented with 1 mM dithiothreitol (DTT). The sample was incubated for 5 min at 100°C, precipitated by adding an equal volume of isopropanol, and centrifuged for 5 min at 16,000 g. The pellet was washed with 70% (v/v) ethanol, recentrifuged and resuspended in 100 µl of 50 mM Tris-HCl pH 7.6.

Isolation of the *T. brucei* **TK gene from a cDNA library.** A 310-bp fragment of the TK gene was amplified from *T. brucei* TC221 genomic DNA with the following two primers:

- TKdir: 5'-AAT TGA TCC GCA CGA CAG -3'
- TKrev: 5'-AAT GAC GTC GTA GTC ACG -3'

The PCR was performed with a mixture of Taq and Pfu DNA polymerases (2.5 units of each) using 30 PCR cycles and an annealing temperature of 50°C. A partial cDNA from a putative TK of *T. b. rhodesiense* found in GenBankTM (accession no. AA0986595) was used to design the primers (the project was started before the sequencing of the *T. brucei* genome). After verification by dideoxy sequencing, the PCR product was purified on low-melting agarose, labeled with ³²P-dCTP using Klenow fragment and random hexanucleotide primers, and used as a probe to recover the TK ORF from a procyclic *T. brucei* 427 (TbGARP16) cDNA library.

Identification and sequencing of the two TK alleles in T. brucei TC221 cells. A PCR (25 cycles) was performed on a T. brucei TC221 lysate using the two primers, Tkdir (see above) and 5'-TTC AGT TTG CAC CAT ATT C-3' (TKr6), with a mixture of Taq and Pfu polymerases (2.5 units of each) and an annealing temperature of 52°C. The PCR product was sequenced and confirmed to contain both 89-bp fragments. At four positions, the TK sequence was ambiguous due to the presence of two different alleles (see Results). A third PCR (25 cycles, annealing temperature 50°C) was performed on the trypanosome lysate with Tkdir as the 5'-primer. The other primer was either 5'-CAG AAC GCG GCC ACA C-3' (TKrev1) or 5'-CAG AAC GCG GCC ACA T-3' (TKrev2), which are identical except for the 3'-base. These two primers could discriminate between G and A at the fourth ambiguous position. The PCR product with the primers TKdir and TKrev1 was found to have an identical sequence to the one obtained from the cDNA library whereas the PCR product with Tkdir and TKrev2 differed at the four ambiguous positions. Finally, a PCR was performed with the primers 5'-CCA CAC GCT CGC ATT-3' (TKf3) and PCR-4 (see sequence below) to also cover the region downstream of TKrev1 and TKrev2. Dideoxy sequencing of the PCR product with the allele-specific primers 5'-GCT CAA AGA TCA TGT TAT-3' and 5'-GCT CAA AGA TCA TGT TAC-3' revealed a fifth allelic difference (see Results).

Creation of a pET expression vector containing His₆-tagged *T. brucei* **TK.** The *T. brucei* **TK** open reading frame (ORF) was amplified from pUC-*Tb*TK (the plasmid is described in the main text) using

forward and reverse primers containing restriction sites for NdeI and BamH1, respectively (underlined):

- TbTK-pET: 5'-ATT GCC ATC TCC GTC GTG <u>CAT ATG</u> CAC GAC GGA GAT GGC AAT-3'
- PCR-4: 5'-TAG CT<u>G GAT CC</u>T GGC ACA ATA CGC TGC G -3'

The PCR product was digested by NdeI and BamHI and subcloned into a pET3a vector opened by the same enzymes. The resulting plasmid was called pET-*Tb*TK. In order to make a His₆-tagged *T. brucei* TK construct, two oligonucleotides were constructed:

- His-tag NdeI fwd: 5'-TAT GCA TCA CCA TCA CCA TCA CGC-3'
- His-tag NdeI rev: 5'-TAG CGT GAT GGT GAT GGT GAT GCA-3

The two oligonucleotides were annealed and ligated into pET-*Tb*TK opened by NdeI to create the plasmid pET-His*Tb*TK.

Creation of expression vectors encoding fusion proteins attached to full-length *T. brucei* **TK, domain 1, domain 2 or human TK1.** The full-length *T. brucei* TK and each of its two domains were amplified from pET-*Tb*TK by PCR using forward and reverse primers containing restriction sites for NcoI and Acc65I, respectively (underlined). For the amplification of human TK1, a cDNA clone (IMAGE 2905608, Source Bioscience, UK) was used as template and the forward primer contained a BspH1 site (isoschizomer of NcoI). The primer pairs used were the following:

Domain 1:

- *TKD1R1_Fwd-NcoI: 5'-AAC TGC T<u>CC ATG G</u>CA ATG CAC GAC GGA GAT GGC A-3'*
- TKD1R1_Rev-Acc651: 5'-GGA CC<u>G GTA CC</u>T TAA CCG TGT GCA CCA TTT GGC AC-3'

Domain 2:

- TKD2_TrxA Fwd-NcoI: 5'-AAC TGC T<u>CC ATG G</u>TG CCA AAT GGT GCA CAC GGT CGC-3'
- TKD2TrxA Rev-Acc65I: 5'-ACG G<u>GG TAC C</u>TA AGT AGT ATC AAC GGC CAT-3

Full length T. brucei TK:

- TbTKfull_fwd-NcoI: 5´-AAC TGC T<u>CC ATG G</u>CA ATG CAC GAC GGA GAT GGC A-3´
- TbTKfull_rev-Acc65I: 5´-ACG GGG TAC CTA AGT AGT ATC AAC GGC CAT-3´

Human TK1:

- hTK_Fwd, BspH1: 5'- GCT ACT CAT GAG CTG CAT TAA CCT GCC CAC-3'
- hTK_Fwd, Acc65I: 5'-GAT ACG GTA CCT CAG TTG GCA GGG CTG C-3'

The PCR products were subsequently digested with NcoI (or BspH1) and Acc65I and subcloned into several variants of bacterial expression vectors from European Molecular Biology (see main text). They were all verified by DNA sequencing from both directions.

Expression and purification of fusion constructs containing removable His-tagged fusion partners in frame with human TK1, *T. brucei* full-length TK, domain or domain 2. Z-tagged human TK1, Z-tagged *T. brucei* TK, MBP-tagged domain 1, and TrxA-tagged domain 2 were expressed in Rosetta (DE3) pLysS *E. coli* cells (Novagen) and grown in the presence of 34 mg/L chloramphenicol and 25 mg/L kanamycin (34 mg/L chloramphenicol and 100 mg/L carbenicillin for TrxA-tagged domain 2) in 1 L Luria-Bertani broth. Growth temperatures and IPTG induction were similar as described for the His₆-tagged *T. brucei* TK. The induced cells were harvested, washed in

buffer A (0.3 M NaCl, 50 mM Na₂SO₄, 6 mM β-mercaptoethanol, and 20 mM Tris-HCl pH7.8), and split into two tubes. For domain 1 and the human TK1, Na₂SO₄ was omitted from buffer A here and in subsequent steps. For each purification, the material from one of the tubes was thawed, resuspended in 10 ml buffer A containing 10 mM imidazole, and centrifuged at 45,000 rpm for 1 hour (Beckman L-90 ultracentrifuge, Ti70 rotor). The supernatant was applied to a column containing 1 ml Nickel-NTA Superflow (Qiagen), which was subsequently washed in a step-wise manner with buffer A containing higher and higher concentrations of imidazole (10 mM, 20 mM, and 50 mM). Each washing step was continued until no more detectable protein came off the column (as measured with the Bio Rad Protein Assay). TrxA-tagged domain 2 and Z-tagged T. brucei TK were eluted with 150 mM imidazole whereas Z-tagged human TK1 and MBP-tagged domain 1 were purified using an additional washing step with 100 mM imidazole before elution with 250 mM imidazole. The buffer of the purified proteins was subsequently exchanged into buffer A (supplemented with 4 mM MgCl₂) by Sephadex G-25 chromatography. The fusion partner of each of the purified proteins was removed by digestion with equal amount (in mg) of His₆-tagged TEV protease at 25°C for 2.5 hours (no loss of enzyme activity was observed during the incubation step) and subsequently applied to a second Nickel-NTA Superflow column equilibrated with buffer A. The His6-tagged fusion partners, TEV protease, and non-cleaved protein remained on the column, whereas the TK proteins/domains eluted mainly in the flow-through and to some extent in the subsequent 20 mM imidazole wash. The protein solutions were, if necessary, concentrated by centrifugation with a 30 kDa cutoff filter (Centricon or Amicon tubes, Millipore) and subsequently divided into aliquots for storage at -80°C. They were only thawed once to avoid any loss of enzymatic activity due to repeated freeze-thaw cycles.

SUPPLEMENTAL FIGURES S1-S6

				10		20	30	40	50	
T. brucei D1					- MHDGD	GNIELII	GPMFACKT	TELMRRVQRH	KHAQRSCYLINYS	SR 44
T. brucei D2					VPNGAH	GRIELII	GPMFACKT	TELMRRVQRH	KHAQRSCYLLKYA	G 39
A. suum D1					MAS	GHIQLIF	GP <mark>M</mark> FS CK T	TEMLRRLNRY	KLANRTCRIVKYF	RN 42
A. suum D2					SR	GRIEVIL	.GP <mark>M</mark> FS GK T	TEMLRRYNRH	ALAGRECRVIKYF	RG 41
T. congolense D1				N	NTQSPQS	GRIELVI	GPMFACKT	TELMSRVNRY	SLAGYSCYVIEYS	SK 46
T. congolense D2						GRIELVI	GP <mark>M</mark> FA CK T	TEL I SRVNRN	YLASHSCYVIEYS	SK 39
T. congolense D3						GRIELVI	GP <mark>M</mark> FA CK T	TEL I SRVNRN	YLARYSCYVIEYS	SK 39
T. congolense D4						GRIELVI	GP <mark>M</mark> FA CK T	TEL I SRVNRN	YLARYSCYVIEYS	SK 39
T. vaginalis D1					MQA	GRIEVIV	'GP <mark>M</mark> FA CK T	TEML RRI NRA	EHGHRRVVVMKYD	DK 42
T. vaginalis D2					FGIQTS	GMINLTV	GP IKSCKT	TELLRVLNRY	LIAGRKAICLRPE	ES 45
H. sapiens TK1			MSC	INLPTVLPO	SPSKTR	GQIQVIL	. GP <mark>M</mark> FSCKS	TELMRRVRRF	QIAQYKCLVIKYA	AK 57
L. major					MFR	GRIELII	GPMF ACKI	TELMRRVKRE	THARRSCEVIKYS	SK 42
T. Cruzi	MFASHISICILKLRY	AEDQRIFRLA	FGFCASSTQQL	FADISLQMM	MLENESH					SR 90
E. COII T. thermonhilus					M	GELEVIT		TEL LEOVEDO		1 40
C elecans			KNEMTCOSSNS			GSITVII				1 40 1 71
C parvum					M	IAKI YEYY		TVLLOSSENY		SK 40
A. thaliana		MATI	KASELLKTLDS			GAVHVIN		TSLI BRI KSE	ISDGRSVAMI KSS	SK 69
G. lamblia					M	INSETTE	GPMFACKS	TELVGI HARL	VAAHKRVI VVKHT	F 40
U. urealyticum				MAKAH	AFSKKV	GWIELIT	GPMFACKT	AELIRRLHRL	EYADVKYLVFKPF	RI 50
	60 70		80	90	100		110	120	130	
T. brucei D1	NS-YQNQRLSTHDQL	SLTAN- VS- I	AKLSEVCDEWR	DY DV I A		FP- DVVC	FCARAANE	GKTVIVSAL	DVDCRETP <mark>F</mark> DE V	C 125
T. brucei D2	DTRYSEGAITSHDQR	ALTAN- VS- V	SNL HDVGDEWR	KYDVI A	AV DEGOF	FP- GVAA	FCSKAADS	- GKV <mark>V</mark> I VSA <mark>L</mark>	DADYLQEP <mark>F</mark> EE- I	C 121
A. suum D1	DTRYAL DHVATHDLQ	MQEAIS-A	VKVADVMDELS	EAHVV A	AIDEGOF	FD- DIAE	CSENLANC	λ-GKI <mark>V</mark> IVSA <mark>L</mark>	DGDF NRKR <mark>F</mark> KN- V	'L 123
A. suum D2	DTRYDANKVVTHDQL	MHDGVI-A	THIADI FDELL	AYKVI A	AIDEGOF	FQ-DIAE	CCERLANN	1- GK I 🛛 I VAAL	DGDYSRKE <mark>F</mark> ASKV	'L 123
T. congolense D1	HERYSKK IFSPQHRL	SMITT-VS-V	TKLSEVGNAWH	DYDVIA	AV D <mark>e</mark> gQF	FP-DILE	FCSMAANA	- GKK <mark>V</mark> I VAA <mark>L</mark>	DGDYRNEP <mark>F</mark> EN- I	A 128
T. congolense D2	NMRYSREN I PTHDGR	TLPAT-VS-V	NKLSEVGNAWH	DYDVI A	AV DGG <mark>Q</mark> F	FP-DILE	FCSMAANA	- GKK <mark>V</mark> I VAA <mark>L</mark>	DGDYRNEP <mark>F</mark> EH- I	A 121
T. congolense D3	DMRYSKEIIPTYDRR	TLPAT-VS-V	NKLSEVGNAWH	DYDVI A	AV DEGOF	FP-DILE	FCSMVANA	- GKK <mark></mark> ∎I VAA <mark>L</mark>	DGDYRNEP <mark>F</mark> EH- I	A 121
T. congolense D4	DMRYSKE I I STYDRR	TLPAT-VS-V	NKLSEVGNAWH	DYDVI A	AV <mark>DE</mark> G <mark>OF</mark>	FP-DILE	FCSMAANA	- GKK <mark>V</mark> I VAA <mark>L</mark>	DGDYRNEP <mark>F</mark> EH- I	A 121
T. vaginalis D1	D <mark>QR</mark> YSVNKVSTHDEY	MHDAIP-C	NMLLPHLEECL	GYEVIG	GV <mark>DE</mark> G <mark>OF</mark>	FP- DVVE	FSEKLANF	- GRT <mark>V</mark> I VAA <mark>L</mark>	DGT F QRK P <mark>F</mark> GK - V	'L 123
T. vaginalis D2	AHDFHKSNVEIKFIQ	NLPS	IEELN	QYDIIG	GI DEAQN	FE-NIAD	DWADELANS	- GKV	DGNENHVAYPS- I	V 117
H. sapiens TK1	DTRYSS-SFCTHDRN	TMEALP-A	CLLRDVAQEAL	GVAVI G	GI DEGOF	FP-DIVE	FCEAMANA	- GKTVI VAAL	DGTFQRKP <mark>F</mark> GA- I	L 137
L. major	DIRYDEHNVASHDQL	MLRAQ-AA-V	SQLIEVRDIWK	RF DVL A		FS-DLVN	IF CNTAADA	- GKVVMVSAL	DGDYRRKP <mark>F</mark> GQ- I	C 124
I. cruzi	DARYSRESVSSHDKL	LLGAI - AA- V	AELREVGDAWR	PEDVV A	AV DEGOF	FP-DIVO			DGDYRRQPFDG-I	C 1/2
E. COII T. thormonbiluo		SSPAKLENQN	SSLFDEI KAEH						RIDERGELFIG-S	Q 128
C elecans							TCEELAOR		NCTEERKPERO- I	C 154
C. narvum		SEKAHTETPD							RTDEKGNI EEG. S	K 120
A. thaliana		GEPOWAL P- D		AYNKI DVI G		EG-DIYE	ECCKVADD		DGDYL RRSEGA-V	1 156
G. lamblia		RIAAK R- A	VLLGELATEFG	DYDAL I		FA-DIVS	GVQDALSK	GLYNYVSAL	SGNFKREPFEL-I	P 121
U. urealyticum	DTR-STQNIKSRTGT	SLPSIEVENA	PEILSYIMSDN	FDNEIKVIG		FD-DRIC	EVANILAE		DKNFKGEP <mark>F</mark> GP-I	A 137
	140 150	160	170		180		190	200	210	220
T. (Ξ.
T. brucei D1	RLVPRAESVLKLSAV		ETVETVKSD	EDKIV				OTEKVIVSCV		204
			FTKPLTSNT	CRRLV	GCRDV			MPPHS	GINEGSTSECSFG	187
		CTECG SDAS	ESERTTAHK-	NOFII		RAMCRAC		NSHNYPTRSP		31 206
T. congolense D1	KI I PAAESVTKI SAV	CMSCRVSNAY	YTHRTVOSE	ORFLI	GGADK	FATCRAC	YNRI NNOG	GSI VENPOLO	OS	200
T congolense D2	KLIPAAESVTKLSAV	MSCG-CDAY	YTHRTVOSE	QRELI	GGAEKY	FATCRAC	YNRLNNOG	GSL VENPOLO	QS	192
T. congolense D3	KLIPAAESVTKLSAV	CMSCRLSNAY	YTHRIVQSE	QREL I	GGADMY	TAACRAC	YNRLNSQG	GSLAENPQSP	QS	193
T. congolense D4	KLIPAAESVTKLSAV	CMSCG-CDAH	YTHRTVQSE	QREL I	GGADI	IATCRAC	YNRLNNQG	GSLVENPKGP	QSG VAVS HTV EDA	N 204
T. vaginalis D1	DLMSKCESI TKLSAV	CSQTG-SEAA	FSK <mark>R</mark> IVNST	DVELI	GSESY	vaas <mark>r</mark> aa	١F			175
T. vaginalis D2	ELFPRCEKVTKLDSI	CPLTG-LPAP	FTAVVDGLK	F P S	RLGIL	RASMGHV	LNSTSTN			175
H. sapiens TK1	NLVPLAESVVKLTAV	CMECF-REAA	YTK <mark>R</mark> LGTEK	EVEV I	<mark>G</mark> GADK	HSVC <mark>R</mark> LC	YFKKASGQ	PAGPDNKENC	PVP GKPG EAV AAR	K 220
L. major	ELVPYCEAVDKLTAV	CMMCHEQPAC	FTR <mark>R</mark> TVNVE	QQEL I	<mark>G</mark> GADM <mark>Y</mark>	IATCREC	YSKQQLPS	I EEMRTQQMA	KEVEKRYLGMSD	K 208
T. cruzi	RLIPLAESVKKLTAV	CMECHCRSAS	FTY <mark>R</mark> TVSSE	KREL I	<mark>G</mark> GADM <mark>Y</mark>	I AAC <mark>R</mark> T (FVTKSKKR	AELEAKGSVQ	TADKCSDTPAAEK	Q 256
E. coli	YLLAWSDKLVELKTI	EF-CG-RKAS	MV L <mark>R</mark> L D QA GRP	YNEGEQVVI	<mark>G</mark> GNER	VSVC <mark>R</mark> KF	IYKEALQVD	SLTAIQERHR	HD	205
T. thermophilus	ELL VRADNI VYL TAV	CTVCG KPAT	RSQ <mark>R</mark> L I DGKPA	PRNSPVILV	′ <mark>G</mark> GRES <mark>Y</mark>	EARC <mark>R</mark> EF	HLVPDE			197
C. elegans	LLLPYANEIKQVTAV	CVECG SQAN	FSF <mark>R</mark> STLDK	KVEVI	<mark>G</mark> GSDT	TAL C <mark>R</mark> EC	YVQKSEEK	DAEEQMKTGC	DKNENDITGIFLA	K 239
C. parvum	YLLAWADKLTEIKTI	CR-CG-KKAT	MT I <mark>R</mark> LNSNGEP	VFSGEQILI	<mark>G</mark> DNS I Y	TSVC <mark>R</mark> KF	HI INCEEY	NF		195
A. thaliana	DIIPIADSVTKLTAR	EVCG HKAF	FTLRKNCDT	RTEL I	GGADV	MPVCRKH	IYITNHIVI	KASKKVLEDS	DKA RAES CVA AT I	238
G. lamblia	RLFPLASATYLRSAT	CAICH APAP	FSARFSAQT	EEIVI	GGAEL	APTCRTC	WKSIAHQR	SIKKARQLTD	SEIKSFVSDAAAI	L 204
U. urealyticum	KLFAYADKITKLTAI	CNECG AEAT	hsl <mark>r</mark> k i DGKYA	NYDDEIVKI	<mark>G</mark> CQEFY	SAVC <mark>R</mark> HF	IHKVPNRPY	LNANSEEFIR	F F K NK KR NKN V	223

Fig. S1. Sequence alignment of the TK1 domains and proteins used for constructing the phylogenetic tree in Fig. 3B. The numbering is based on the human TK1 sequence. Conserved amino acid residues in all aligned species are highlighted in black whereas the yellow color indicates residues that differ in some domains but are conserved in all single-domain proteins. C-terminal ends of *T. brucei* (D2), *A. suum* (D2), *T. congolense* (D2), *H. sapiens, L. major, T. cruzi, C. elegans* and *G. lamblia* TKs are truncated in order to fit the alignment onto a single page.



Fig. S2. Western blot analysis of different *T. brucei* TK preparations with a mouse α -His-antibody (Invitrogen) shows a major full-length TK band (52 kDa) and additional C-terminal degradation products (the His₆-tag is in the N-terminus). The upper band is the full-length TK band and was the major band in all lanes at shorter exposure times. Corresponding positions of pre-stained reference protein bands on the blotting membrane are shown to the left.



Fig. S3. Effects of sulfate on *T. brucei* TK activity. (A) Enzyme activity with 5 μ M thymidine and 2 mM ATP in the presence of different concentrations of Na₂SO₄. The experiment was made in duplicate with the standard deviations indicated. (B) Lineweaver-Burk diagram showing that sulfate inhibits the TK enzymatic reaction by acting as a competitive inhibitor with respect to ATP. The experiments were performed in the absence (•) or presence of 10 mM (•) or 20 mM (\blacktriangle) ammonium sulfate. The thymidine concentration was 0.5 mM in B in order to make ATP to be the limiting substrate.



Fig. S4. Effects of salts and DTT on *T. brucei* TK activity. (A) TK enzyme activity in the absence (ctrl) or presence of different potassium, sodium or ammonium salts at 0.1 and 0.2 M concentration. (B) Effect of DTT concentration on enzyme activity. The experiments were made in duplicate with the standard deviations indicated. The substrates were 5 μ M thymidine and 2 mM ATP.



Fig. S5. Dependence of *T. brucei* TK activity on pH. The buffer used in the assays was Hepes-KOH between pH 6-7.8 (+) and Tris-HCl between pH 7.8-9 (\Box). The experiments were made in duplicate with the standard deviations indicated. The substrates were 5 μ M thymidine and 2 mM ATP.



Fig. S6. Standard curve made from gel filtration analysis of myoglobin (16.9 kDa), ovalbumin (43 kDa), transferrin (78 kDa), IgG (150 kDa), ferritin (440 kDa) and thyroglobulin (670 kDa).