Supplementary materials and methods

Protein kinase inhibition assay

Biochemical assays for human EGFR (Epidermal growth factor receptor) and BTK (Bruton's tyrosine kinase) were carried out using a homogenous time-resolved fluorescence (HTRF) assay as described previously [1] .

High throughput phenotypic screening for identification of growth inhibitors.

A kinase focused inhibitor library (containing approximately 10,000 compounds) was screened at a single point concentration of 10 μM for ability to inhibit growth of *T. b. brucei* Lister 427. This resulted in 2264 compounds showing > 50% growth inhibition. Chemi-informatic analysis of hits resulted in ~200 clusters, representatives of these were subjected to 10 point dose-response growth inhibition assays against *T.b. brucei* Lister 427. Further removal of non-favourable chemical structures such as nitrofuranes, cytotoxicity profiling using the HepG2 cell line (selectivity index of 10) and favourable physico-chemical properties (polar surface area <100; molecular weight <500 and lipophilicity <4.5) led to identification of an amidobenzimidazole scaffold (AB0) for further follow up.

Archive based structure activity relationship.

The Novartis library had several compounds belonging to amidobenzimidazole series in the archive, which were synthesized as a part of other protein kinase inhibitor drug discovery programs. They were further used for structure activity relationship using multiple cellular and enzyme assays as described in this manuscript.

Kill kinetics and reversibility assay.

Ability of AB1 compound to kill *T. b. brucei* Lister 427 was measured over a period of 6, 24 and 48 hr post-compound treatment. The ATP content of parasites was used as a surrogate of the viability of parasites. The assay was conducted similar to the growth inhibition assay stated above with minor modifications. Compound-containing plates were incubated with parasites at 1 x 10^5 ml^{-1} and at each time point, CellTiter Glo reagent was added to lyse the parasites and luminescence was measured using a Tecan M1000 plate reader after 30 min incubation.

Reversibility assessment to establish time and concentration required to achieve irreversible (relapse-free) growth inhibition under *in vitro* condition was carried out as described elsewhere[2]. The AC_{cure} is the absolute concentration required to achieve sterile cure under *in vitro* conditions with incubation of compound for 72 hr.

Mouse model of stage II (CNS) HAT.

The GVR35 mouse model mimics the second (CNS) stage of the disease. Female CD1 (Charles River UK; ~8 weeks old; protocol number PPL 60/4442, University of Glasgow) mice were infected by injection into the peritoneum with 3×10^4 *T. brucei* (GVR35-VSL2) bloodstream form parasites [3]. Starting on day 21, mice were dosed by oral gavage once-daily with AB1 (n= 6) at 100 mg/kg for 7 days. A group of untreated mice (n= 3) was included as control.

Mice were monitored weekly for parasitaemia from day 21 post-infection. *T. brucei* was quantified in blood samples from the tail vein by microscopy, and *in vivo* bioluminescence imaging of infected mice was performed before treatment on day 21 post-infection and in weeks following the treatment (day 27, 28, 35 post-infection). Imaging on groups of three mice was performed 10 min after i.p.injection of 150 mg D-luciferin (Promega)/kg body weight (in PBS) using an IVIS Spectrum (PerkinElmer) as described previously [4]. Mice were euthanized between days 27 and 35 by cervical dislocation. Data analysis for bioluminescence imaging was performed using Living Image Software. The same rectangular region of interest (ROI) covering the mouse body was used for each whole body image to show the bioluminescence in total flux (photons per second) within that region. Image panels of whole mouse bodies are composites of the original images with areas outside the ROI cropped out to save space.

All animal procedures were undertaken in adherence to experimental guidelines and procedures approved by The Home Office of the UK government. All work was covered by Home Office Project Licence PPL60/4442 entitled “Molecular Genetics of Trypanosomes and Leishmania”. All animal protocols received approval from the University of York and University of Glasgow Ethics Committees.

Pharmacokinetic analysis of AB1.

Blood samples from mice were collected at 1, 3, 8 and 24 hours post last dose of CNS stage mice studies. Total concentration of AB1 compound in blood was analysed using standard LC-MS/MS as described elsewhere [1]. Free concentration in brain was calculated by taking into consideration the brain to plasma ratio, plasma protein binding and brain tissue binding.

Cell cycle analysis and Cell Sorting.

Bloodstream form *T. brucei* cell lines were incubated or not for 6 hr with AB compounds at a final concentration of 5x the individual EC_{50} value for each compound (averaged from viability assays). Control cultures were treated with 0.5µl DMSO. Cultures were pelleted and cells were collected and washed once in *Trypanosoma* dilution buffer (TDB) supplemented with 5 mM of EDTA and resuspended in 70% methanol. Cells were centrifuged at 1400 g for 10 min to remove methanol and washed once in TDB 1x with 5mM EDTA. Cells were resuspended in 1ml TDB 1x with 5mM EDTA, 10µg ml^{-1} of propidium iodide and 10µl of RNase A. Cell suspensions in 1.5 ml tubes were wrapped in foil to avoid bleaching by light. Cells were incubated for 30 min at 37°C in the dark until FACS analysis. Cells were analysed for FACS using a Beckman Coulter CyAn ADP flow cytometer (excitation; 535, emission; 617).

In the cell cycle analysis, CLK1 OE was induced during 18 hr with tetracycline (1 µg ml^{-1}) and later treated with 5x the individual EC_{50} value for each compound for 6 hr (maintaining tetracycline induction), and finally collected for flow cytometry as above.

Parasite cell sorting was conducted as described previously [5]. Briefly, cell lines were harvested during exponential growth by centrifugation for 10 min in a clinical centrifuge at room temperature. The parasite pellet was then resuspended at a concentration of 1×10^6 cells ml^{-1} in HMI-9 medium supplemented with 2% FCS and 10 µg ml^{-1} penicillin/streptomycin. Vybrant DyeCycle Violet (Molecular Probes, Invitrogen) was added to a final concentration of 1 µg $ml^{-1}$and the cell suspension incubated for 30 min at 37 °C, the tube being protected from light by wrapping in aluminium foil. The samples were then centrifuged and resuspended back in the staining media prior to sorting on a MoFlo XDP Sorter (Beckman Coulter Life Sciences). During and after the sorting, the samples were cooled to below 20 °C to limit cell metabolic activity. The dye was excited using a 407 nm Violet laser and emission detected via a 450/40 bandpass filter. Live parasites were gated based on FSC/SSC profiles, and the gates were set up to collect only the 2C fraction (G0/G1 cells) and 4C fraction (G2, mitotic and post-mitotic cells), to ensure efficient discrimination and selection of these cell-cycle stages. The nucleus and kinetoplast configuration of synchronized cells were monitored over time by staining cells with DAPI.

Multiple sequence alignment.

The Uniprot knowledgebase [6] and TriTryDB [7] were used to retrieve fasta sequence files for the sequences of hEGFR (P00533), hCLK1 (P49759) and CLK1 (Q382U0). The multiple sequence alignment was performed using the Needleman-Wüncsh algorithm [8] modified to include zero end gap penalties (ZEGA) [9] with the Gonnet substitution matrix [10], as implemented in ICM molecular modelling package (ICM version 3.8-6a, Molsoft LLC).

Analysis of covalent modification by LC-MS.

Two μM of WT CLK1 and CLK1 C215A were incubated with 4 μM compound at RT for 1 hr, and quenched with 80 μM DTT. Approximately 18 pmol of protein is loaded onto a 2.1 x 50 mm 1000Å PLRP-S column (Agilent) equilibrated in 5% buffer B and eluted by a 2 minute 5-65% gradient at 0.2 mL/min into an Agilent 6530 QToF equipped with an electrospray source. Buffer A is 0.1% formic acid in water, and Buffer B is 0.1% formic acid in acetonitrile. Source voltage was 5500V, drying gas was 12 L/min at 350 C, and the nebulizer was at 60 psig. Fragmentor, skimmer and octopole are at 175 V, 60 V and 750 V, respectively. Mass spectra are acquired at 1 Hz over the m/z range 400-2000. An averaged m/z spectrum representing the total ion chromatogram peak of the protein is extracted, and a zero-charge mass spectrum is generated by maximum entropy deconvolution is performed over a range of 6,000-225,000 Daltons with a 0.5 Da step.

Recoded CLK1 plasmid and site directed mutagenesis

Recoded CLK1 was synthesised by Eurofins Genomics. The recoded CLK1 sequence ($CLK1^R$) codes for the same amino acid sequence as CLK1 but only shares 95.06% nucleotide identity. All segments of identity between CLK1 and $CLK1^R$ are less than 20 base pairs long. $CLK1^R$ was inserted by Gibson assembly® (New England Biolabs) into the plasmid pGL2492 using *Xba*I and *Bam*HI restriction sites, generating pGL2832. This plasmid is designed to constitutively express CLK1 from the tubulin locus, with the addition of a C-terminal 6x HA tag. To express the cysteine 215 mutants, the cysteine 215 was changed to serine or alanine by mutating pGL2832 (recoded CLK1) and pGL2465 (CLK1 OE), carrying the coding sequence for CLK1, using site-directed mutagenic PCR as follows:

PRIMER SEQUENCES	MUTATION	PLASMID
5'- AAGGGTCGGTGCAGCATTAAGGCCGC	CLK1 $\mathbf{C^{215S}}$	NITD001
3´- TTGGGCATCACAATACAC		
5'- AAGGGTCGGTGCAGCATTAAGGCCGC	$CLK1^{R}$ $\mathbf{C^{215S}}$	pGL2845
3´- TTGGGCATCACAATACAC		
5'- GTACGGACCAGCATTATTAGATTGGATAATG	$CLK1^{R}$ $\mathbf{C^{215A}}$	pGL2846
3´- TTGGGCATCACAATACAC		

Antibodies

Antibody	**Obtained from**
Mouse Imprint Monoclonal anti-Ty1 antibody (clone BB2)	Sigma Aldrich (SAB4800032)
Mouse Anti-HA (clone 12CA5)	Roche (11583816001)
Mouse Anti-EF1α Antibody, (clone CBP-KK1)	Merck-Milipore (05-235)
Mouse anti c-Myc Monoclonal Antibody (9E10)	Invitrogen (13-2500)
StarBright™ Blue 520 Goat anti-Mouse IgG	BIORAD (12005867)
StarBright™ Blue 700 Goat Anti-Mouse/Anti-Rabbit IgG	BIORAD (12004162)

Real-time PCR

Total RNA was extracted from $5x10^7$ parasites from parental 2T1 cell line using the RNeasy Mini Kit (Qiagen) according to the manufacture´s recommendations. The amount of total RNA was quantified by Qubit RNA HS Assay. RT-qPCR targeting TbCLK1 and TbCLK2 was performed by using Luna® Universal One-Step RT-qPCR Kit (New England Biolabs), with a starting concentration of 250 ng of RNA according to manufacturing protocol on an Applied Biosystems Viia 7 RT PCR instrument. The gene expression level was measured using the comparative cycle threshold (Ct) method. Expression of GADPH gene was used as the housekeeping control.

Bibliography

1. Lelais, G. *et al.* Discovery of (R,E)-N-(7-Chloro-1-(1-[4-(dimethylamino)but-2-enoyl]azepan-3-yl)-1H-benzo[d]imidazol-2-yl)-2-methylisonicotinamide (EGF816), a Novel, Potent, and WT Sparing Covalent Inhibitor of Oncogenic (L858R, ex19del) and Resistant (T790M) EGFR Mutants for the Treatment of EGFR Mutant Non-Small-Cell Lung Cancers. *J. Med. Chem.* **59,** 6671–6689 (2016).
2. Jacobs, R. T. *et al.* Benzoxaboroles: a new class of potential drugs for human African trypanosomiasis. *Future Med. Chem.* **3,** 1259–1278 (2011).
3. McLatchie, A. P. *et al.* Highly sensitive in vivo imaging of Trypanosoma brucei expressing "red-shifted" luciferase. *PLoS Negl. Trop. Dis.* **7,** e2571 (2013).
4. Myburgh, E. *et al.* In vivo imaging of trypanosome-brain interactions and development of a rapid screening test for drugs against CNS stage trypanosomiasis. *PLoS Negl. Trop. Dis.* **7,** e2384 (2013).
5. Kabani, S., Waterfall, M. & Matthews, K. R. Cell-cycle synchronisation of bloodstream forms of Trypanosoma brucei using Vybrant DyeCycle Violet-based sorting. *Mol. Biochem. Parasitol.* **169,** 59–62 (2010).
6. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47,** D506–D515 (2019).
7. Aslett, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38,** D457–62 (2010).
8. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443–453 (1970).
9. Abagyan, R. A. & Batalov, S. Do aligned sequences share the same fold? *J. Mol. Biol.* **273,** 355–368 (1997).
10. Gonnet, G. H., Cohen, M. A. & Benner, S. A. Exhaustive matching of the entire protein sequence database. *Science* **256,** 1443–1445 (1992).

Supplementary Methods Table 1

OE plasmids details

927 ID	PRIMER SEQUENCES	OE Plasmid	Gene	LOF	KDa*
Tb927.9.14430	5'- ACAAAGCTTATGTCCGTCTCTTCGCAACC	pGL2366	**CK2A1**	Yes	50
	3'-ACATCTAGACTCCGTGACAACATTCCCTC				
Tb927.11.1180	5'- ACAAAGCTTATGTCTGATACATTAAGTAGCAAAC	pGL2360	**CRK6**	Yes	44
	3'- ACATCTAGAGTCACCTTGCAGAAAAGGATG				
Tb927.10.4990	5'- TACGATATCATGACAATGCTTGGGGCGTT	pGL2380	**CRK3**	Yes	42.2
	3'- CTCGATATCTAAACATGGCATCACTAAACC				
Tb927.10.1070	5'-TGAAAGCTTATGGGGAGTCGTTACGAGC	pGL2358	**CRK1**	Yes	41.5
	3'-ACATCTAGAGAACTCGACAGAAAAGTATGG				
Tb927.11.10520	5'-ACCAAGCTTATGTTCAATGTCTCACCAGC	pGL2393	**KKT2**	Yes	146.4
	3'- ACTTCTAGATTTCCGGTAGACGTTGCGCAT				
Tb927.9.10920	5'- TGAAAGCTTATGATTGGCACAGTAGAACA	pGL2376	**PK6/KKT3**	Yes	125.3
	3'- CACTCTAGACACGCCACCGTAGAGCATC				
Tb927.11.12410	5'- TGTAAGCTTATGGATAGCATAAGCTCTGG	pGL2465	**CLK1/KKT10**	Yes	60.3
	3'- ACTTCTAGACATGACAGGTGTTGGGGGGAG				
Tb927.7.7360	5'- ACCTTAATTAAATGCAGGTGCAGGTGCAGGAAG	pGL2359	**CRK2**	Yes	46.4
	3'- ACATCTAGAAGCTCCGTTGAACCGCATCAAC				
Tb927.11.12310	5'- ACCTTAATTAAATGGGTATGGCAACACGTTCGC	pGL2362	**CRK12**	Yes	91.8
	3'- ACATCTAGACGCAGCGCTTGGATAAGGC				
Tb927.10.4940	5'- GCATTTAATTAAATGTCAGATGTGCCTAAATC	pGL2383	**PK50**	Yes	57.3
	3'- GCATTCTAGAATCATCATCTGAAAAATTCTGC				
Tb927.8.5730	5'- GCATTTAATTAAATGAGCGACGAGGGGTTTGAAG	pGL2388	**SLK1**	Yes	60.2
	3'- GCATTCTAGACAGGCGGTAGACACTCTTCAAC				
Tb927.8.7220	5'- GCATTTAATTAAATGAAAAACAACGACGCTAACC	pGL2389	**TLK2**	Yes	80.6
	3'- GCATTCTAGAGGGGATGTTACTTAACTTCTCC				
Tb927.11.4470	5'- CTGATTAATTAAATGCCATCTAATTCTGTTGTAG	pGL2395	**ULK**	Yes	130.6
	3'- ACGACCTAGGCGAGGACGAAATCCTTTGC				
Tb927.10.7780	5'- GCACGAAGCTTATGGTGTCGTTCAGCATCGATGGA	pGL2469	**KFR1**	Yes	46.6
	3'- GCACGTCTAGACAAGGCTAACTTTTGCTT				
Tb927.6.4970	5'- GCATTTAATTAAATGCCCCGACCCGACAG	pGL2403	**SRPK**	Yes	89.9
	3'- GCATTCTAGAGTCATTAGTGCAGCTGCTGTCAG				
Tb927.3.2440	5'- GCATTTAATTAAATGCTCAATAGACTATTTGG	pGL2426	**AGC**	Yes	53
	3'- GCATTCTAGAACTACCACCACCCAGATG				
Tb927.11.5340	5'- GCATTTAATTAAATGAGGGTGATGAGACACC	pGL2402	**CMGC**	Yes	80.6
	3'- GCATTCTAGACTCAAACAGCACCTTGCG				
Tb927.5.790	5'- CAGTTAATTAAATGGGTGGCGACGGGCGCA	pGL2441	**CK1**	Yes	45.1
	3'- GACTCTAGATTCGGTGTCATCACTCTCCTC				
Tb927.7.6220	5'- GCGTTAATTAAATGGACTTCCATTCGTTGC	pGL2467	**CAMK**	Yes	68
	3'- GTATCTAGATCTCTCGTTGCGGCGCTTGG				
Tb927.10.2040	5'- GGCTTAATTAAATGCGGGCTAACGACATTGTG	pGL2442	**STE**	Yes	100.1

	3'- GCATCTAGATCCCACATTTGTCTGTGCCTG				
Tb927.11.2040	5'- GCGTTAATTAAATGAAGCTGAGGGCTGACTCC	pGL2443	**STE/STE11**	Yes	87.9
	3'- CGCTCTAGATTTGATCAACCACTTGATCAC				
Tb927.7.5770	5'- CCGTTAATTAAATGTCCACGCGGAAAGAGC	pGL2468	**PK53**	Yes	60
	3'- GGCTCTAGACCGTTTGAGGTTGAATTTG				
Tb927.10.10350	5'- CGGTTAATTAAATGAGTACACAAGAATCTC	pGL2464	**STE/STE11**	Yes	73.9
	3'- GCGTCTAGAAAACGTAAGAAATGGATGT				
Tb927.3.690	SYNTHESISED ORF	pGL2466	**CMGC/ RCK**	Yes	71
	SYNTHESISED ORF				
Tb927.10.10870	5'- TTCACAAGCTTAATTAATGTATGCACAATGTCCCGTTAC	pGL2722	**MAPKLK1**	Yes	73.1
	3'- TCAGAGATCAGTTTCTGTTCACGAGGTGATGAAAAGTG				
	3'- TCAGAGATCAGTTTCTGTTCTATGGGACCCATAGCCCTTG				
Tb927.7.3210	5'- TTCACAAGCTTAATTAATGTATGTTACTTCATTAC AGGACGC	pGL2725	**Orphan**	U	78.8
	3'- TCAGAGATCAGTTTCTGTTCATCAGAGTCCGCA ACAGTTTTAATG				
Tb927.4.5310	SYNTHESISED ORF	pGL2323	**RDK2**	Yes	57
	SYNTHESISED ORF				
Tb927.7.6310	5'- GATCTCTGAAGAAGACCTGCCTAGGATGCACG CAACCGCTGAG	pGL2726	**PLK1**	Yes	98.9
	3'- TAAATGGGCAGGATCCTCACCTAGGTAAATATCACGGTTTTG TATGAGCTCG				

* Predicted size including 6c-myc tag.
LOF: Loss of fitness

RNAi constructs and qPCR

927 ID	PRIMER SEQUENCES	USE	Gene
Tb927.11.12410	5'- GGTCTCCAGCAATACCCCTC	RNAi	**CLK1**
	3'- TTTTGCAGTACTCGTCGTG		
Tb927.11.12410	5'- TGGATAGCATAAGCTCTG	qPCR	**CLK1**
	3'- GTCGTCAGCAGTGTCG		
Tb927.11.12420	5'- TAACTCATCCACCATTGCCA	qPCR	**CLK2**
	3'- AGATGCATATGACGTGACGC		
Tb927.6.4280	5'- AGATTGATGTCGTTGC	qPCR	**GADPH**
	3'- ATGGCTTGCTCTTCGT		

Ty-mNG Endogenous tagging

927 ID	PRIMER SEQUENCES	Gene
Tb927.11.12410	5'- GGAGAGAAAGTACGGCACTTGCTTCGTCTGATAGATCAC TAAAACAAGAAACAAAGAGTATTTACTACAAAACAACCACAGTATAATGCAGACCTGCTGC	CLK1
	3'- TTTACGTGGACATTTACTAGTCCTGAGGGGTATTGCTGGAGACCAGCTCTG TGGGTTAGACCAGAGCTTATGCTATCCATACTACCCGATCCTGATCC	
Tb927.11.12420	5'- TCTCGTCTGACTACATTTTTCTGTCAGCAATTTGTCGGACCATTTGTTTTTGGAGGAACA TTTATAGAACCACGTATAATGCAGACCTGCTGC	CLK2
	3'- CCGTAGAAACCCGCAGTGTCTCCGCTCATGCCCTTGGCAATGGTGGATGAGTTAACG TTTGACGCTATGTATGAAGCCATACTACCCGATCCTGATCC	

Supplementary Table 1 Data collection and refinement statistics

	TbCLK1/AB1 complex
Data collection	
Space group	$P3_1$
Cell dimensions	
a, *b*, *c* (Å)	147.15, 147.15, 264.8
α, β, γ (°)	90.00, 90.00, 120
Resolution (Å)	47.39 - 2.60 (2.69 - 2.60) *
R_{sym} or R_{merge}	0.24 (1.46)
$I / \sigma I$	98.56 (99.32)
Completeness (%)	98.56 (99.32)
Redundancy	5.1 (4.9)
Refinement	
Resolution (Å)	2.70
No. reflections	194424
R_{work} / R_{free}	0.2611 / 0.3079
No. atoms	
Protein	44352
Ligand/ion	725
Water	511
B-factors	
Protein	53.98
Ligand/ion	46.85
Water	38.58
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.69

*A single crystal was used for this structure. *Values in parentheses are for highest-resolution shell.