The clinical and veterinary trypanocidal benzoxaboroles target CPSF3

Richard J. Wall, Eva Rico, Iva Lukac, Fabio Zuccotto, Sara Elg, Ian H. Gilbert, Yvonne Freund, M.R.K. Alley, Mark C. Field, Susan Wyllie and David Horn

Supporting Information Appendix

www.pnas.org/cgi/doi/10.1073/pnas.1807915115

SI Materials and Methods

Trypanosomes. Bloodstream-form *T. b. brucei*, Lister 427, MiTat 1.2, clone 221a and derivatives, including 2T1 cells (1), were cultured at 37°C in the presence of 5% CO_2 in HMI9 medium, as described (2). These cells were transfected using a Nucleofector (Lonza) with T-cell Nucleofector solution. Transformants were cloned by limiting-dilution and selected with phleomycin (1 μg.ml^{-1}), blasticidin (10 μg.ml^{-1}; maintenance at 2 μg.ml^{-1}) or hygromycin (2.5 μg.ml^{-1}) for constructs integrated at the tagged *rRNA* locus (3). Cumulative growth curves were generated from cultures seeded at 10^4 cells.ml^{-1}, counted on a haemocytometer every 24 h and diluted as necessary. Overexpression and RNAi were induced with 1 μg.ml^{-1} tetracycline. For EC_{50} assays, overexpression strains were pre-induced for 24 h and assays were carried out using AlamarBlue as described (4). Data was processed using GRAFIT (version 5.0.4; Erithacus software) and fitted to a 2-parameter equation to calculate EC_{50}.

pRPa^{OEX} overexpression plasmid assembly. The pRPa^{OEX} construct was assembled as follows. A *lacZ* stuffer fragment, amplified using OEX5 and OEX3, was used to engineer BbsI and FseI restriction sites (italics) on either side of the stuffer in pRPa^{iSL} (3). Next, NSAS5 and NSAS3 oligonucleotides, containing I-SceI sites and an AscI site (italics) were annealed and cloned in pBluescript digested with NotI. The entire AscI cassette from the pRPa^{iSL} derivative above was then moved to the new AscI-digested vector, thereby placing I-SceI sites on either side of the assembly. Finally, a blasticidin (*BLA*) resistance cassette was synthesised (Genscript) and, following digestion with NsiI/MluI, was inserted upstream of the *EP* procyclin promoter.

OEX5:GATCGGTACC*GGCCGGCCGAAGAC*GATGATGAGATAGGGTTGAGTGTTGT
OEX3:GATCGGGCCCGCTTACGGCGTACACCCTATCAATGATAGAGT*GGCCGGCCGAAG*
*C*GATGATAAACCCTGGCGTTACCCAA
NSAS5:GGCC*TAGGGATAACAGGGTAATGGCGCGCCTAGGGATAACAGGGTAAT*
NSAS3:GGCC*ATTACCCTGTTATCCCTAGGCGCGCCATTACCCTGTTATCCCTA*

pRPa^{OEX} plasmid overexpression library assembly. pRPa^{OEX} was digested with BbsI and 'semi-filled' with Klenow and dTTP, followed by dephosphorylation with Antarctic phosphatase (New England Bioloabs). 25 µg of *T. b. brucei* genomic DNA was digested with 0.04 U Sau3AI/µg gDNA for 1 h at 37°C. The partially digested DNA was separated on an agarose gel (1x TAE, 120V, 35 min). The gel slab containing 3-10 kbp fragments was excised, set in a new gel and concentrated in the gel by reversing the direction of electrophoresis (1x TAE, 120V, 35 min). The *T. b. brucei* DNA was then purified from the gel and 'semi-filled' with Klenow and dGTP. The resulting pRPa^{OEX} plasmid and *T. b. brucei* insert DNA were then ligated overnight at 16°C, purified, electroporated in MegaX DH10B Electrocomp T1R cells (Thermo Fisher), added to 500 ml LB-amp media and grown overnight at 37°C. A series of dilutions were also grown on agar plates in order to estimate library complexity; plasmids from 30 colonies were digested with FseI to determine whether inserts were present. The plasmid library was purified from the 500 ml culture using a Qiagen Plasmid Maxi Kit.

***T. b. brucei* overexpression library assembly.** 500 µg of overexpression plasmid library was digested with I-SceI. Plasmid linearization prior to transfection increases transfection efficiency and the use of I-SceI that specifically cleaves an 18 bp sequence, not present in *T. brucei,* ensures that linearization will not reduce library coverage; a restriction enzyme that cleaves a shorter sequence would have fragmented our library constructs; approximately 10% of them in the case of an enzyme that recognises an 8 bp sequence. 2T1 *T. b. brucei* cells were transfected with pSce* as described (2). A puromycin-sensitive 2T1^{Sce*} clone was expanded (2.5 x 10^9 cells), induced with tetracycline (1 µg/ml) for 3 h and transfected with the I-SceI digested overexpression library: forty replicates with 10 µg DNA and 5 x 10^7 cells in each replicate. After 6 h, phleomycin (1 µg/ml; for the tetracycline operator cassette) and blasticidin (10 µg/ml; for the

library) selection was applied. A series of dilutions were also grown in 96-well plates in order to estimate library complexity; genomic DNA was extracted from 22 clones from these plates to assess the presence of library inserts, amplified using a long-range PCR reaction (see below). Ten inserts were then Sanger sequenced using OEseq1 (GTAAAGTCAATACAACACACAATAGG) and OEseq2 (CTGGCACAGAGAGCGAGC), revealing that they were derived from random sites across the genome, as expected.

Overexpression library screening and data analysis. The *T. b. brucei* overexpression library was maintained at or above 2 x 10^7 cells to maintain complexity, in medium containing phleomycin (1 μg/ml) and blasticidin (1 μg/ml). Overexpression was induced with tetracycline (1 μg/ml) for 24 h and 2 x 10^7 cells in 150 ml media were used to initiate each screen. Cells were passaged as required and genomic DNA was extracted after 8-9 days using a Qiagen DNeasy Blood & Tissue Kit. Overexpressed fragments were amplified using the OEseqA primer (CGGCGTACACCCTATCAATGA) in a 'long-range' PCR reaction using LongAmp polymerase (1 cycle: 30 s at 94°C; 30 cycles: 30 s at 94°C, 30 s at 60°C and 9 min at 65°C; 1 cycle: 10 min at 65°C) and purified using a QIAquick PCR Purification Kit. The products were sequenced using an Illumina HiSeq platform at the Beijing Genomics Institute. Reads were aligned to the *T. brucei* 927 reference genome (v9.0, tritrypdb.org) with Bowtie 2 software (5) using the conditions: very-sensitive-local–phred33. The subsequent alignment files were manipulated with SAMtools (6) and a custom script (2) to identify reads with barcodes (GATAGAGTGGTACCGGCCGG or CAATGATAGAGTGGCCGGCC), which also revealed insert orientation. Total and barcoded reads were then quantified using the Artemis genome browser (7) and Excel.

Construction of other plasmids. For knockdown of Tb927.4.1340 (*CPSF3*) using RNA interference, we used CPSF-RF and CPSF-RR primers, incorporating Bsp1201/Acc651 and XbaI/BamHI sites, respectively (italics) and cloned the resulting amplified 606 bp fragment in the pRPa^{iSL} construct (3). The construct was introduced into 2T1 *T. b. brucei* cells (1), which were selected with hygromycin (2.5 μg/ml) and phleomycin (1 μg/ml); only puromycin sensitive clones were analysed further. For tagging a native Tb927.4.1340 allele with GFP at the *C*-terminus, we used CPSF-TF and CPSF-TR primers, incorporating HindIII and XbaI sites, respectively (italics) and cloned the resulting amplified 654 bp fragment in the pNAT^{xGFP} construct (3). The construct was linearised with XhoI prior to transfection and introduced into 2T1 *T. b. brucei* cells (1) with the RNAi construct. These cells were selected with hygromycin and phleomycin as above and also blasticidin. For inducible overexpression of Tb927.4.1340 with GFP at the *C*-terminus, we used the CPSF-OF primer, incorporating a PacI site (italics), and CPSF-TR and cloned the resulting amplified complete coding sequence (2,332 bp) in the pRPa^{ixGFP} construct (3). The construct was introduced into 2T1 *T. b. brucei* cells (1), selected with hygromycin and phleomycin as above; again, only puromycin sensitive clones were analysed further. pRPa-based constructs were linearised with AscI prior to transfection. Cas9-based editing of *T. brucei* CPSF3 Asn^{232} was carried out as described (8). Briefly, an sgRNA construct was assembled using the cpsfgF and cpsfgR primers. The resulting construct was linearised with NotI prior to transfection into 2T1$^{T7\text{-}Cas9}$ cells. The cpsfN^{232}T oligonucleotide was used as the single-stranded repair template and acoziborole selection was applied at 2 x EC_{50} for eight days. A negative control lacking the repair template failed to yield acoziborole-resistant cells. Resistant clones were isolated, from which a segment spanning the editing site was PCR-amplified and sequenced.

CPSF-RF: GATC*GGGCCCGGTACC*ACAGACACGCGACTTCATCC
CPSF-RR: GATC*TCTAGAGGATCC*TCTCGTGCTCTGCTGAACAT
CPSF-TF: GATC*AAGCTT*ACAGACACGCGACTTCATCC
CPSF-TR: GATC*TCTAGA*TGCGCCCTCGTCAATGGG
CPSF-OF: GATC*TTAATTAA*ATGTTTTCACACGCGGCGTA

cpsfgF: AGGGTACAAATGGTATACGCGAAC
cpsfgR: AAACGTTCGCGTATACCATTTGTA
cpsfN^{232}T:
CGATATTCTCATTGCAGAGAGTACATATGGAATTAGGGAGTTAGAGTCACGTGAAGAGCGG
GAATCCC

Protein blotting. *T. brucei* parasites were grown with or without tetracycline for 24 h and harvested. Samples were then resuspended in Laemmli buffer and boiled for 5 min. Lysates (4 x 10^6 cells/well) were separated on 8% acylamide gels and transferred to nitrocellulose membrane. The top half of the membrane (above 70 kDa) was blocked with 5% BSA in TBS (GFP) and the bottom half (below 70 kDa) was blocked in 5% skimmed milk in PBS (EF1a; control) for 1 h. Blots were then incubated with either α-GFP (1:1000; D5.1XP, Cell signalling Technology) or α-EF1α (1:20000; Merck Millipore) primary antibodies overnight. Membranes were then incubated with either rabbit HRP (For GFP, 1:5000; BioRad) or mouse HRP (For EF1α, 1:5000; BioRad) secondary antibodies for 1 h. Blots were developed using ECL plus (with enhanced chemiluminescence; GE Healthcare).

Microscopy. Fluorescence microscopy was carried out according to standard protocols. Briefly, 1 x 10^6 cells were fixed in 2% formaldehyde in PBS, washed in PBS and allowed to dry onto coverslips. Cells were mounted in Vectashield (Vector Laboratories) containing the DNA counterstain DAPI. GFP fluorescence images were captured using a Zeiss Axiovert 200M fluorescence microscope and processed in ZenPro (Zeiss).

Phylogeny and domain analysis. The sequences were retrieved from TritrypDB, PlasmoDB, ToxoDB and NCBI: *T. b. brucei* (Tb927.4.1340); *T. b. gambiense* DAL972 (XP_011772702.1, Tbg972.4.1170); *T. congolense* IL3000 (CCC90000.1, TcIL3000_4_840.1); *Leishmania donovani* (LdBPK_343210.1); *T. cruzi sylvio* X10/1 (TCSYLVIO_003775); *Trichomonas vaginalis* G3 (TVAG_437970); *Thermus thermophilus* (YP_143518.1); *Toxoplasma gondii*

(TGME49_285200); *Plasmodium falciparum* (PF3D7_1438500) and *Homo sapiens* (AAF00224.1). Alignment and tree generation were performed using Clustal X and viewed using TreeView X. Domains figures were generated using InterPro and Prosite MyDomains.

***In silico* docking studies.** To identify suitable template structures, the *T. brucei* CPSF3 sequence (UniProtKB - Q581U7) was used to query the PDB using "BLAST" as implemented in the NCBI blastp suite (https://blast.ncbi.nlm.nih.gov/). Endonucleases from three different species; human CPSF73 (PDB code 2I7T – sequence identity 56%), *Pyrococcus horikoshii* CPSF (PDB code 3AF5 – sequence identity 29%) and *T. thermophilus* TTHA0252 (several structures available exemplified by PDB code 3IEM – sequence identity 31%) were identified as close analogues that could be used as template structures. The *T. thermophilus* structure is complexed with a stable RNA analogue bound to the catalytic site located at the interface between the metallo-β-lactamase and β-CASP domains whereas both human and *P. horikoshii* structures are apo forms in which the Zn containing binding site is inaccessible. Thus, the *T. thermophilus* structure was selected as template. The alignment (Fig. S4*A*) indicates the typical sequence motifs associated to metallo-β-lactamase (9) and β-CASP (10) catalytic activity. The presence of His^{446} is associated with specificity for RNA substrates (11).

A homology model for the *Tb*CPSF3 structure was built using the knowledge-based method in Prime (12). The Zn atoms were modelled in the structure but not the substrate. Due to its empty p-orbital, the B atom in acoziborole is a strong electrophile (Lewis acid) that can react with solvent water molecules. The nucleophilic attack of an activated water molecule on the trigonal boron atom leads to the formation of a tetrahedral negatively charged boron species (Fig. S4*B*) (13). The three-dimensional structure of the hydroxylated form of acoziborole was built in Maestro (12), minimized with the OPLS3 force field and docked in the catalytic site of the *Tb*CPSF3 model using GLIDE. In the template structure, the Arg^{270} side-chain interacts with an RNA phosphate unit such that it adopts a folded conformation. In the absence of the pre-mRNA

phosphate or a phosphate mimic, we considered additional side-chain conformations. The acoziborole binding mode was further investigated by induced-fit docking (12) that accounts for the flexibility of the binding site residues.

SI REFERENCES

1. Alsford S, Kawahara T, Glover L, & Horn D (2005) Tagging a *T. brucei RRNA* locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol Biochem Parasitol* 144(2):142-148.
2. Glover L*, et al.* (2015) Genome-scale RNAi screens for high-throughput phenotyping in bloodstream-form African trypanosomes. *Nat Protoc* 10(1):106-133.
3. Alsford S & Horn D (2008) Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei*. *Mol Biochem Parasitol* 161(1):76-79.
4. Raz B, Iten M, Grether-Buhler Y, Kaminsky R, & Brun R (1997) The Alamar Blue assay to determine drug sensitivity of African trypanosomes (*T.b. rhodesiense* and *T.b. gambiense*) *in vitro*. *Acta Trop* 68(2):139-147.
5. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.
6. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
7. Carver T, Harris SR, Berriman M, Parkhill J, & McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28(4):464-469.
8. Rico E, Jeacock L, Kovarova J, & Horn D (2018) Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. *Sci Rep* 8(1):7960.
9. Ishikawa H, Nakagawa N, Kuramitsu S, & Masui R (2006) Crystal structure of TTHA0252 from *Thermus thermophilus* HB8, a RNA degradation protein of the metallo-β-lactamase superfamily. *J Biochem* 140(4):535-542.
10. Mandel CR*, et al.* (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444(7121):953-956.
11. Callebaut I, Moshous D, Mornon JP, & de Villartay JP (2002) Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. *Nucleic Acids Res* 30(16):3592-3601.
12. Schrödinger (2017) Small Molecule Drug Discovery suite. *Release 2017-4*.
13. Baker SJ, Tomsho JW, & Benkovic SJ (2011) Boron-containing inhibitors of synthetases. *Chem Soc Rev* 40(8):4279-4285.
14. Brinkman EK*, et al.* (2018) Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol Cell* 70(5):801-813 e806.

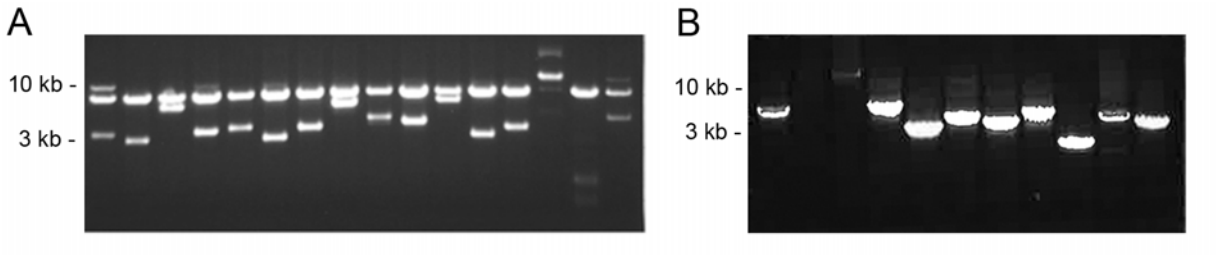


Fig. S1. Assessment of *T. b. brucei* genomic DNA inserts in pRPa^{OEX} library and in the *T. b. brucei* library. (*A*) Representative gel showing DNA from sixteen bacterial colonies from the plasmid library digested with FseI; 30 colonies were analysed in total. Backbone, ~6.5 kbp; inserts, 3-10 kbp. (*B*) Representative gel showing PCR products between 3-10 kb from eleven *T. b. brucei* clones from the library; 22 clones were analysed in total.

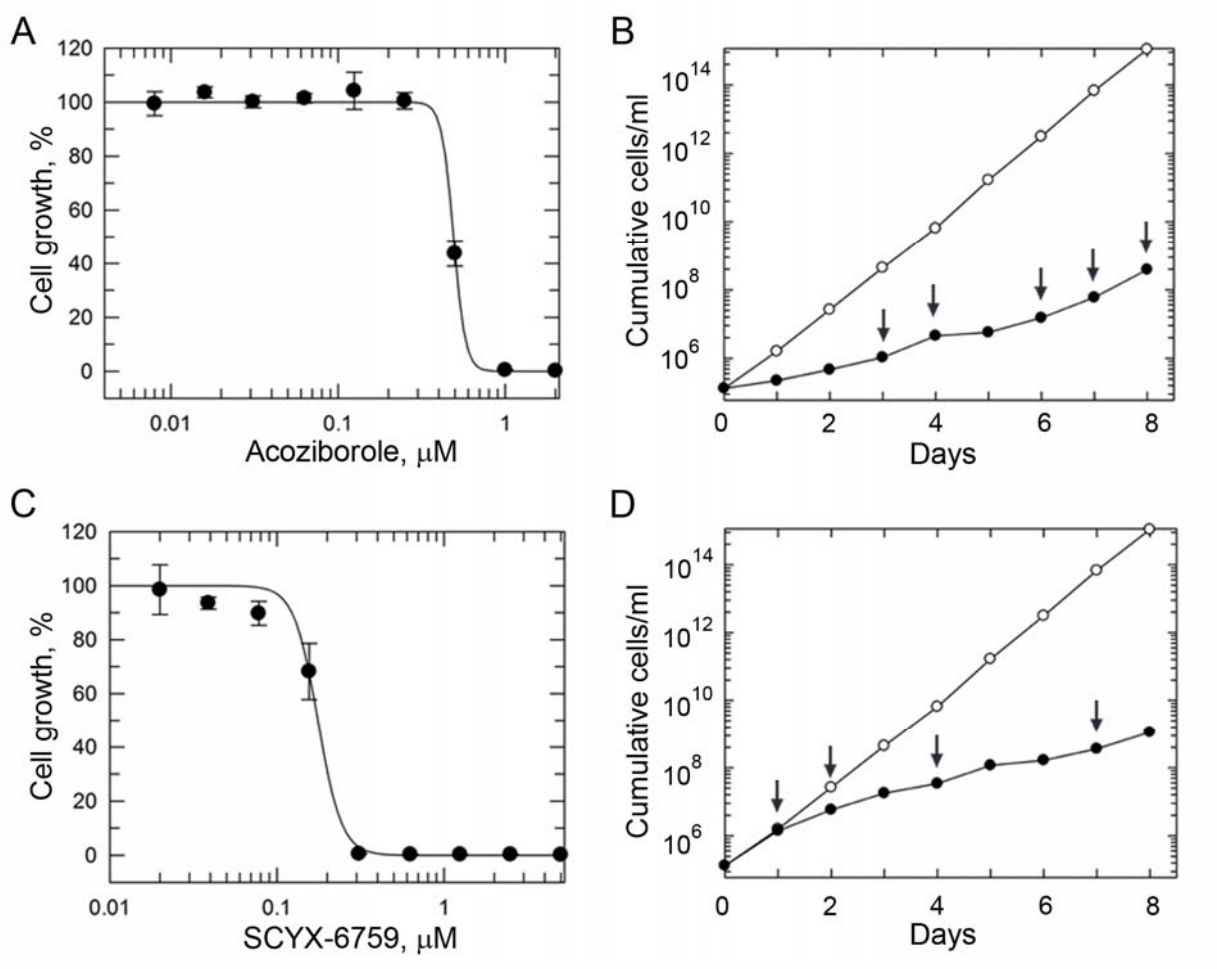


Fig. S2. Acoziborole and SCYX-6759 EC_{50} and library screening. (*A*) Dose-response curve for acoziborole: EC_{50} = 0.49 ± 0.02 μM. (*B*) The growth curves indicate expansion of the induced library in the absence (open circles) or presence (closed circles) of acoziborole. For the drug-treated culture, black arrows indicate the addition of Tet plus drug. (*C*) Dose-response curve for SCYX-6759: EC_{50} = 0.18 ± 0.01 μM. (*D*) As in B but for SCYX-6759.

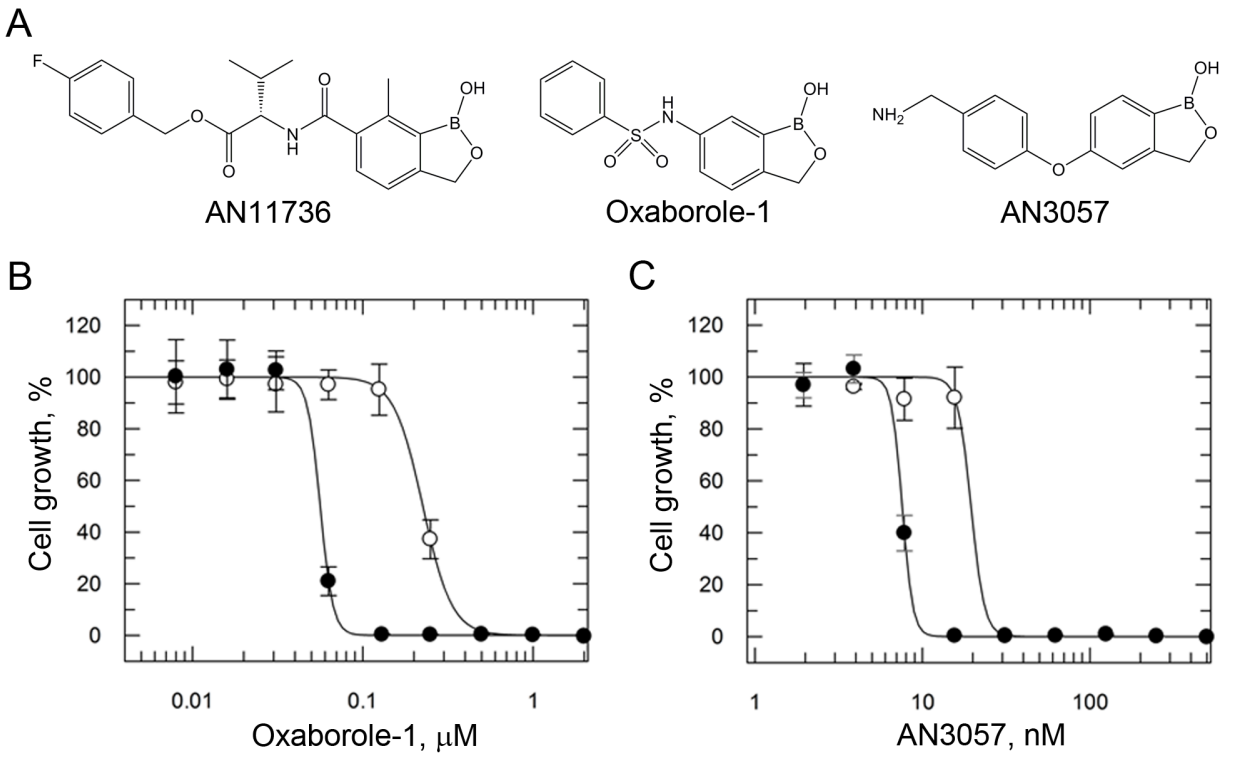
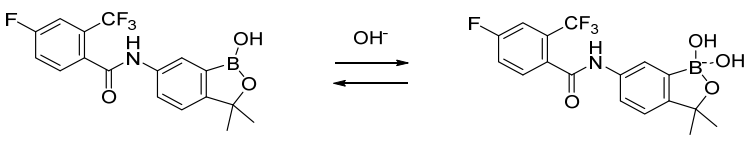


Fig. S3. Cells overexpressing CPSF3^{GFP} are resistant to oxaborole-1 and AN3057. (*A*) Chemical structures of AN11736, oxaborole-1 and AN3057. (*B*) Dose-response curves for oxaborole-1 with (open circles) and without (closed circles) CPSF3^{GFP} overexpression: EC_{50} no Tet, 0.06 ± 0.001 μM; + Tet, 0.23 ± 0.004 μM; 4-fold shift. (*C*) As in B but for AN3057: EC_{50} no Tet, 7.58 ± 0.09 nM; + Tet, 19.53 ± 2.28 nM; 2.6-fold shift.

A

```
                                 **  *. ::    : ::**** . . .      . :. :   ::* :*::* *:**   :*
T. brucei                        EVGRSCVVVRYKGRSVMLDCGNHPAKSGLDSLPFFDSIRCDEIDLVLITHFHLDHCGALP  108
T. congolense                    EVGRSCIVVRYKGRSVMLDCGNHPAKSGLDSLPFFDSIRCEEIDVVLITHFHLDHCGALP  108
T. cruzi                         EVGRSCVILRYKGRSVMLDCGNHPAKSGLDSLPFFDSIRCDEIDLVLITHFHLDHCGALP  108
L. donovani                      EVGRSCVVVRYKGRGVMLDCGNHPAKSGLDSLPFFDSIKCDEIDVVLITHFHLDHCGALP  108
H. sapiens                       EVGRSCIILEFKGRKIMLDCGIHPGLEGMDALPYIDLIDPAEIDLLLISHFHLDHCGALP  108
T. vaginalis                     EVGRSCIILKYHRKRVMLDCGIHPAYENFGGLPFIDAIDPAKIDVLLITHFHIDHITAVP  108
T. thermophilus                  EVTGSAHLLLAGGRRVLLDCGMFQGKEEARNHAPFG-FDPKEVDAVLLTHAHLDHVGRLP  107

                                  :  :  : *  : * :*  : . :: * : :         :     : .    : .:::
T. brucei                        YFCEQTSFRGRIFMTSATKAFYKMVMNDFLRIGA---SAEDIVNNEWLQSTIEKIETVEY  165
T. congolense                    YFCEQTAFKGRIFMTSATKAFYKMVMNDFLRVGA---SAEDIVNNEWLQSTIEKIETVEY  165
T. cruzi                         YFCEQTAFKGRVFMTSATKAFYKMVMNDFLRVGA---SANDIVTNEWLQSTIEKIETVEY  165
L. donovani                      YFCNQTSFKGRIFMTSATKAFYKMVMNDFLRIGA---GASDLVTSEWLQSTIDRIETVEY  165
H. sapiens                       WFLQKTSFKGRTFMTHATKAIYRWLLSDYVKVSN-ISADDMLYTETDLEESMDKIETINF  167
T. vaginalis                     WFLTQTNFSGPCFMTHTTKTISKTLLVDYVGVSGRGSEEPNLFTRADVANVQNMITAVNY  168
T. thermophilus                  KLFREG-YRGPVYATRATVLLMEIVLEDALKVMD-----EPFFGPEDVEEALGHLRPLEY  161

                                  :     .: :    ***: *:.: : :  *   :*:**:.    .  *  ..:*
T. brucei                        HEEVTVNGIHFQPFNAGHVLGAALFMVDIAGMKLLYTGDFSRVPDRHLLGAEVPP-----  220
T. congolense                    HEEVTVNGIHFQPFNAGHVLGAALFMVDIAGMKVLYTGDFSRVPDRHLLGAEVPP-----  220
T. cruzi                         HEEVTVNGIRFQPFNAGHVLGAALFMVDIAGMKTLYTGDFSRVPDRHLLGAEVPS-----  220
L. donovani                      HEEVTVNGISFQPFNAGHVLGAAMFMVDIAGMRALYTGDFSRVPDRHLLGAEVPP-----  220
H. sapiens                       HEVKEVAGIKFWCYHAGHVLGAAMFMIEIAGVKLLYTGDFSRQEDRHLMAAEIPN-----  220
T. vaginalis                     HQTVTHQGIKMTCYPAGHVLGACMWLVEIDGVKVLYTGDFSLENERHLQGAEIPKSLSGE  228
T. thermophilus                  GEWLRLGALSLAFGQAGHLPGSAFVVAQGEGRTLVYSGDLGNREKDVLPDPSLPP-----  216

                                    .*::: *.* *     .   :    *   : . :  **: *:* **: ****:* :*
T. brucei                        -YSPDILIAESTNGIRELESREERESLFTTWVHDVVKGGGRCLVPVFALGRAQELLLILE  279
T. congolense                    -YSPDILIAESTNGIRELESREERETLFTTWVHDVVKGGGRCLIPVFALGRAQELLLILE  279
T. cruzi                         -YSPDILIAESTNGIRELESREERETLFTTWVHDVVKGGGRCLVPVFALGRAQELLLILE  279
L. donovani                      -YSPDILIAESTNGIRELESREEREHLFTSSVHDVVRRGGRCLVPVFALGRAQELLLILE  279
H. sapiens                       -IKPDILIIESTYGTHIHEKREEREARFCNTVHDIVNRGGRGLIPVFALGRAQELLLILD  279
T. vaginalis                     IIRPDVLIMESTHGLARELESRVDREYRFIDNVTKIIKRGGRCLIPIFALGRAQELLIILD  288
T. thermophilus                  --LADLVLAEGTYGDRPHRPYRETVREFLEILEKTLSQGGKVLIPTFAVERAQEILYVLY  274

                                  . .         ***  * :* : :  *       ...:       .. *    :  :
T. brucei                        EYWEAHKELQHIPIYYASSLAQRCMKLYQTFVSAMNDRVKKQHENHRNPFVFKYIQSLLD  339
T. congolense                    EYWEAHKELQHIPIYYASSLAQRCMKLYQTFVSAMNDRVKEQHENHRNPFVFKYIQSLLD  339
T. cruzi                         EYWEAHKELQHIPIYYASSLAQRCMKLYQTFVSAMNDRVKQQHANHRNPFVFKYIHSLME  339
L. donovani                      EFWDAHKELQNIPIYYASSLAQRCMKLYQTFVSAMNDRVKQQHANHHNPFVFKYIHSLMD  339
H. sapiens                       EYWQNHPELHDIPIYYASSLAKKCMAVYQTYVNAMNDKIRKQ-ININNPFVFKHISNLKS  338
T. vaginalis                     EYWESHPEYNGVPIYYGSNLAKQAIAAYNAFYQDHNSRV----VTAKGKFEFSYVKYIRD  344
T. thermophilus                  THGHR---LPRAPIYLDSPMAGRVLSLYPRLVRYFSEEVQAHFLQGKNPFRPAGLEVVEH  331

                                     .       * **:....*: .*    . ::    :  *.::. **  :* :   ::
T. brucei                        TRSFEDTG----PCVVLASPGMLQSGISLELFERWCGDKRNGIIVAGYCVDGTIAKDILS  395
T. congolense                    TRSFEDTG----PCVVLASPGMLQSGISLELFERWCGDKRNGIIVAGYCVDGTIAKEILS  395
T. cruzi                         TRSFEDTG----PCVVLASPGMLQSGISLELFERWCGDRRNGIIIAGYCVDGTIAKDILT  395
L. donovani                      TKSFEDNG----PCVVLASPGMLQSGISLELFERWCGDRRNGIIMAGYCVDGTIAKDVLA  395
H. sapiens                       MDHFDDIG----PSVVMASPGMMQSGLSRELFESWCTDKRNGVIIAGYCVEGTLAKHIMS  394
T. vaginalis                     YD-FDDSL----PCVVLCSPAMLQNGMSRKIFEAWCSNSVNGLIIPGYIVDGTLPQVLMK  399
T. thermophilus                  TEASKALNRAPGPMVVLAGSGMLAGGRILHHLKHGLSDPRNALVFVGYQPQGGLGAEIIA  391

                                  .*  :    *: :* :        .**.*:.  :   ::        : :*:**    :  *
T. brucei                        KPREITKPDGKVLPLRMRTIQSVSFSAHSDGRQTRDFIQALPKTKHVILVHGNVGAMGQL  455
T. congolense                    KPKEITKPDGKVLPLRMRTVQSVSFSAHSDGRQTRDFIQALPNTKHVILVHGNIGAMGQL  455
T. cruzi                         KPKEVTKPDGKVLPLRMRTIQSVSFSAHSDGRQTRDFIQALPKTQHVILVHGNVGAMGQL  455
L. donovani                      KPKEVAKPDGKVLPLRMSTIEAVSFSAHSDGRQTRDFIQSLTKVKHTILVHGNPGAMGQL  455
H. sapiens                       EPEEITTMSGQKLPLKMS-VDYISFSAHTDYQQTSEFIRAL-KPPHVILVHGEQNEMARL  452
T. vaginalis                     NPAEITTLSGKIIPRKIS-IDYVSFSGHADFNQTSRFITEL-KPKRIVLIHGVCGLMMQL  457
T. thermophilus                  RPPAVR-ILGEEVPLRASVHTLGGFSGHAGQDELLDWLQGE---PRVVLVHGEEEKLLAL  447
```

B



i.Trigonal planar neutral B atom ii. Tetrahedral negatively charged B

Fig. S4. CPSF3 alignment and benzoxaborole species to support docking studies. (*A*) Alignment of CPSF3 orthologues. Domains: grey, metallo-β-lactamase; green, β-CASP; blue, Zn-dependent metallo-hydrolase see Fig. 4*A*; Arg^{270} and Tyr^{383} are highlighted in red, see Fig. 4*B*; other residues defining the binding site are in pink and yellow (Asn^{232} and equivalent residue in other species), see Fig. 4B-E and text for more details. (*B*) The equilibrium between the neutral non-hydroxylated form of acoziborole (i) and the tetrahedral negatively charged hydroxylated form (ii).

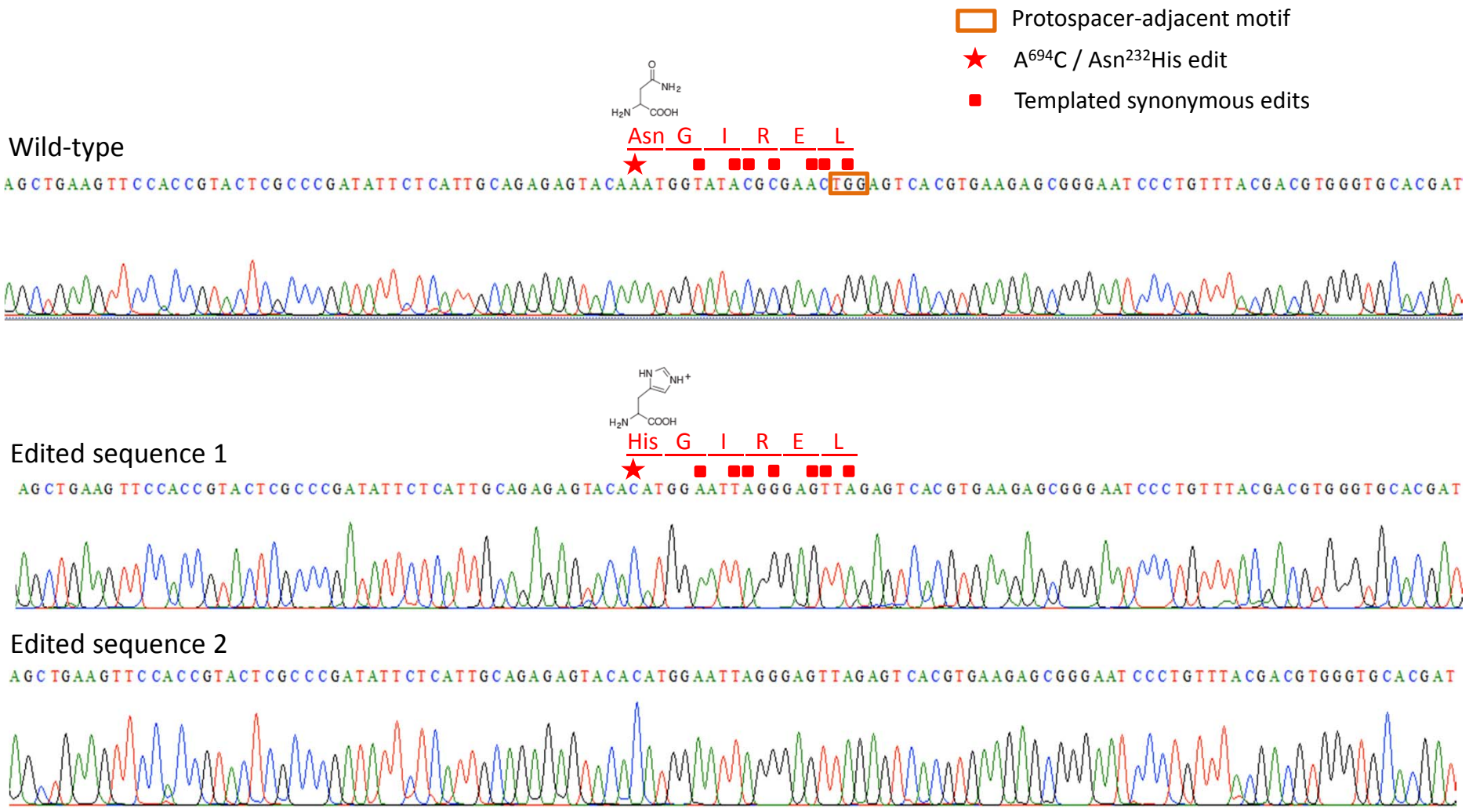


Fig. S5. Cas9-based editing of *T. brucei* CPSF3. The repair template included seven synonymous changes, including one that disrupted the protospacer-adjacent motif, and a non-synonymous change designed to replace Asn^{232} with Tyr. Both acoziborole-resistant cultures analysed (three clones from each) incorporated all of the synonymous changes and an Asn^{232}His edit, suggesting that the Tyr edit was not tolerated. The non-templated A^{694}C edit may be readily obtained due to error-prone repair (14). The Asn and His side-chains are also shown.

Supplemental Table S1: Fragment hits from the DDD85646 screen (containing genes at >1000 RPKM)

Fragment	Gene ID	Gene name	RPKM	Reads	Fragment size (kbp)	Total reads
1	Tb927.10.14230	hypothetical protein, conserved	23244	391682	4	**2712855**
	Tb927.10.14240	N-myristoyltransferase (NMT)	150001	823520		
	Tb927.10.14250	hypothetical protein, conserved	56036	199589		
2	Tb927.9.4000	hypothetical protein, conserved	8020	35660	3.1	**88131**
	Tb927.9.4040	nicotinamidase, putative	2648	6667		
3	Tb927.9.13620	hypothetical protein, conserved	29	707	6.1 [may comprise overlapping fragments]	**54217**
	Tb927.9.13630	hypothetical protein, conserved	1295	5932		
	Tb927.9.13650	ADP-ribosylation factor, putative (ARF1)	1919	4314		
	Tb927.9.13680	ADP-ribosylation factor, putative (ARF1)	4332	9736		
	Tb927.9.13710	ADP-ribosylation factor, putative (ARF1)	4320	9709		
	Tb927.9.13740	ADP-ribosylation factor, putative (ARF1)	2510	5641		
4	Tb927.7.6340	hypothetical protein, conserved	2742	14920	3.7	**39770**
	Tb927.7.6350	NADH-ubiquinone oxidoreductase, mitochondrial	1948	6555		
5	Tb927.11.3040	hypothetical protein	3011	2848	3.3	**32045**
	Tb927.11.3050	hypothetical protein	2417	10597		
	Tb927.11.3060	Phosphate transport (Pho88), putative	2696	5994		

Total reads mapped = 4,094,002 (97% of all reads).

Supplemental Table S2: Fragment hits from the Acoziborole screen (containing genes at >1000 RPKM)

Fragment	Gene ID	Gene name	RPKM	Reads	Fragment size (kbp)	Total reads
1	Tb927.4.1350	glyoxalase II	125221	1509896	4.3	**6100600**
	Tb927.4.1340	cleavage and polyadenylation specificity factor subunit 3	170185	3309917		
	Tb927.4.1330	DNA topoisomerase IB, large subunit	18019	316811		
2	Tb927.10.5620	fructose-bisphosphate aldolase, glycosomal	721	6780	3.7	**57735**
	Tb927.10.5630	hypothetical protein, conserved	1868	8766		
	Tb927.10.5640	Gem-associated protein 2, putative	2296	28725		

Total reads mapped = 8,397,885 (97% of all reads).

Supplemental Table S3: Fragment hits from the SCYX-6759 screen (containing genes at >1000 RPKM)

Fragment	Gene ID	Gene name	RPKM	Reads	Fragment size (kbp)	Total reads
1	Tb927.4.1330	DNA topoisomerase IB, large subunit	5612	64134	4.3	**1104848**
	Tb927.4.1340	cleavage and polyadenylation specificity factor subunit 3	46291	585176		
	Tb927.4.1350	glyoxalase II	36415	285391		

Total reads mapped = 5,449,729; (93% of all reads).

Overexpressed gene
Adjacent/not full CDS

Barcoded, non-redundant fragments containing at least one full-length CDS are shown.
Fragments are ranked according to total read-count.