Clinical and veterinary trypanocidal benzoxaboroles target CPSF3

Wall, Richard; Rico Vidal, Eva; Lukac, Iva; Zuccotto, Fabio; Elg, Sara; Gilbert, Ian; Freund, Yvonne R.; Alley, Michael R. K.; Field, Mark; Wyllie, Susan; Hörn, David

Published in:
Proceedings of the National Academy of Sciences of the United States of America

DOI:
10.1073/pnas.1807915115

Publication date:
2018

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 11. Dec. 2018
The clinical and veterinary trypanocidal benzoxaboroles target CPSF3

Richard Wall1, Eva Rico1, Iva Lukac1, Fabio Zuccotto1, Sara Elg1, Ian Gilbert1, Yvonne Freund2, M.R.K. Alley3, Mark Field1, Susan Wyllie1, David Horn1

1University of Dundee, 2Anacor, 3Anacor Pharmaceuticals, Inc.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

African trypanosomes cause lethal and neglected tropical diseases, known as sleeping sickness in humans and nagana in animals. Current therapies are limited but, fortunately, promising new therapies are in advanced clinical and veterinary development, including acoziborole (AN5568 or SCYX-7158) and AN11736, respectively. These benzoxaboroles are new chemical entities and will likely be key to the WHO target of disease control by 2030. Their mode-of-action was previously unknown. We have developed a high-coverage overexpression library and use it here to explore drug mode-of-action in Trypanosoma brucei. Initially, an inhibitor with a known target was used to select for drug-resistance and to target, a pooled library of parasites have genome-wide mapping this effectively identified the known target and validated the approach. Subsequently, the overexpression screening approach was used to identify the target of the benzoxaboroles, Cleavage and Polyadenylation Specificity Factor 3 (CPSF3, Tb927.4.1340). We validated the CPSF3 endonuclease as the target using independent overexpression strains. Knockdown provided genetic validation of CPSF3 as essential and GFP-tagging confirmed the expected nuclear localisation. Molecular docking and CRISPR-Cas9 based editing demonstrated how acoziborole can specifically block the active site and mRNA processing by parasite, but not host CPSF3. Thus, our findings provide both genetic and chemical validation for CPSF3 as an important new drug-target in trypanosomes and reveal inhibition of mRNA maturation as the mode-of-action of the trypanocidal benzoxaboroles. Understanding the mechanism of action of new benzoxaborole-based therapies can assist development of improved therapies and also the prediction and monitoring of resistance, if or when it arises.

CPSF3 | drug discovery | genetic screening | N-myristoyltransferase | Ysh1

African trypanosomes are transmitted by tsetse flies and cause devastating and lethal diseases, sleeping sickness in humans and nagana in livestock. The closely related parasites that cause chronic and acute forms of the human disease in Eastern and Western Africa are Trypanosoma brucei rhodesiense and T. b. gambiense, respectively (1). These parasites also infect other mammals, while T. congolense, T. vivax and T. b. brucei also cause livestock disease (2). The vast majority of human cases are caused by T. b. gambiense, which progresses from a haemolymphatic first stage to a typically lethal second stage, when parasites enter the central nervous system (CNS).

There is no effective vaccine against African trypanosomes. Current drugs against the first stage disease, pentamidine and suramin, are ineffective if parasites have entered the CNS, frequently the case at the point of diagnosis. For second stage disease, the CNS-penetrant drugs, melarsoprol or nifurtimox and eflornithine, are required (3). Melarsoprol suffers from significant toxicity and drug-resistance and its use is now limited (4). Nifurtimox and eflornithine are currently used as a combination therapy (NECT). NECT is effective but expensive and administration is complex, requiring hospitalisation, trained staff and multiple drug infusions, which may not be sustainable (5). Drug resistance also threatens the usefulness of current chemotherapy and chemoprophylaxis for animal trypanosomiasis with diminazene acetate (Berenil) or isometamidium (Samorin) (2).

Despite the challenges, cases of sleeping sickness have declined in recent years and the disease is now a World Health Organisation (WHO) target for elimination as a public health problem by 2020 (defined as <1 case / 10,000 inhabitants in 90% of endemic foci). A further target is to stop disease transmission by 2030 (6). Attainment of these goals likely depends upon the success of new therapies currently in clinical and veterinary development (3), since it will also be important to control reservoirs in animals (7). Safe, affordable, oral therapies, effective against both stages of the human disease, and effective drugs against the animal disease, would be transformative. The former could also remove the need for painful and cumbersome lumbar puncture diagnosis to confirm the second stage of human infection.

At the end of the last century, there were no new drugs in the pipeline. There are now two new and promising oral drugs in clinical trials against sleeping sickness; acoziborole (AN5568 or SCYX-7158) and fexinidazole (10), and one in trials against nagana, AN11736 (11). However, the relevant cellular targets of these drugs remain unknown, hindering further development, exploitation of related molecules and understanding of toxicity or resistance mechanisms.

Both acoziborole and AN11736 are benzoxaboroles (8, 9, 11). Acoziborole, developed by Anacor and Scynexis, emerged in 2011 as an effective, safe and orally active preclinical candidate against T. b. gambiense, T. b. rhodesiense and T. b. brucei, including melarsoprol resistant strains, and against both disease and chemotherapy-resistant strains. Acoziborole (AN5568 or SCYX-7158) and AN11736, respectively. The WHO target is to stop sleeping sickness transmission by 2030. Current challenges include a shortage of safe, affordable and efficacious drugs, and veterinary reservoirs of trypanosomes, which themselves cause livestock disease. Benzoxaboroles are new chemical entities under development for these lethal diseases that show great promise in clinical and veterinary trials. We developed a new, optimised genome-scale gain-of-function library in trypanosomes and used it to identify the benzoxaborole target. These drugs bind the active site of CPSF3, an enzyme that processes messenger RNA and facilitates gene expression. Our studies validate the gain-of-function approach and reveal an important novel drug-target. These findings will facilitate development of improved therapies and prediction and monitoring of drug resistance.

Significance

The WHO target is to stop sleeping sickness transmission by 2030. Current challenges include a shortage of safe, affordable and efficacious drugs, and veterinary reservoirs of trypanosomes, which themselves cause livestock disease. Benzoxaboroles are new chemical entities under development for these lethal diseases that show great promise in clinical and veterinary trials. We developed a new, optimised genome-scale gain-of-function library in trypanosomes and used it to identify the benzoxaborole target. These drugs bind the active site of CPSF3, an enzyme that processes messenger RNA and facilitates gene expression. Our studies validate the gain-of-function approach and reveal an important novel drug-target. These findings will facilitate development of improved therapies and prediction and monitoring of drug resistance.

Reserved for Publication Footnotes
stages (8, 9). A Phase I clinical trial was successfully completed in 2015 and a Phase II/III trial was initiated in 2016 by the Drugs for Neglected Diseases initiative (DNDi) (www.dndi.org/diseases-projects/portfolio/scyx-7158/). AN11736 is effective against both T. congolense and T. vivax, shows great promise in early development studies (11), and could be the first novel drug for the treatment of animal trypanosomiasis in 50 years. Another benzoxaborole, AN4169 (SCYX-6759) displays activity against the South American T. cruzi, shows great promise in early development studies (12) and T. b. brucei Trypanosoma cruzi, which causes Chagas disease. In addition, the benzoxaborole, DNDI-6148, was recently approved for Phase II development studies (11), and could be the first novel drug for animal trypanosomiasis in 50 years. Another benzoxaborole, AN11736 is effective against both T. congolense and T. vivax, shows great promise in early development studies (11), and could be the first novel drug for animal trypanosomiasis in 50 years. Another benzoxaborole, AN4169 (SCYX-6759) displays activity against the South American T. cruzi, shows great promise in early development studies (12) and T. b. brucei.
panosomes, particularly due to defects in drug uptake or metabolism (26), but this loss-of-function approach does not typically reveal the target(s) of a drug. Indeed, RNAi screens recently revealed a pro-drug activation mechanism for a class of aminomethyl-benzoxaboroles, including AN3057, involving a T. b. brucei aldehyde dehydrogenase, but not the target (27).

A gain-of-function approach is more likely to identify drug-targets and proof-of-concept has been achieved using an overexpression library in T. b. brucei (28). However, the use of relatively short inserts (~1.2 kbp) and fusion with a common RNA-binding domain at the N-termini (29) precluded the overexpression of full-length native proteins, many proteins larger than ~50 kDa or protein targeting to the secretory pathway or mitochondrion (28). A related overexpression approach, with the potential to identify drug targets, Cos-seq, has also been described for Leishmania (30).

We assembled a T. b. brucei inducible overexpression library for full-length genes and with optimised genome coverage. Here, we describe and validate high-throughput overexpression, combined with massive parallel screening, and use the approach to identify the common target of several benzoxaboroles, including acoziborole and AN11736. The trypanocidal benzoxaborole target is the nuclear mRNA processing endonuclease, cleavage and polyadenylation specificity factor 3 (CPSF3/CPSF73/Ysh1).

**Results**

A high coverage T. b. brucei overexpression library. Our goal was to develop an unbiased bloodstream-form T. b. brucei overexpression library with optimised genome coverage. We first assembled the overexpression construct, pRPa® (30), and then constructed a genomic DNA partially digested with Sau3AI (Fig. 1A). The T. b. brucei genome comprises 46% GC content (31) so Sau3AI sites (GATC) occur once every ~256 bp. The genome has only two known introns (32), such that the vast majority of genes comprise a single protein-coding exon with an average size of 1,592 bp; the core genome is also otherwise compact with average intergenic regions of 1,279 bp (31). Partially digested 3-10 kbp genomic DNA fragments were cloned in pRP® using a novel BbsI semi-filling strategy to optimise DNA fragments under the control of a tetracycline-inducible ribosomal RNA (rRNA) promoter (Fig. 1A). The DNA polymerase I promotor that can drive high-level expression of protein coding genes in trypanosomes (33); transcription by RNA polymerase I provides forty-fold stronger transcription than RNA polymerase II, which produces the vast majority of the natural transcriptome (34). The resulting plasmid library comprised >20 million clones from which 26 of 30 analysed (87%) contained inserts in the expected size-range (SI Appendix, Fig. S1A).

To establish a T. b. brucei library, we used a system for high-efficiency transfection previously used to assemble other libraries in bloodstream-form cells (35). This involves I-SceI meganuclease-mediated induction of a break at the chromosomal integration target site for pRP® (30). I-SceI was also used to linearize the plasmid library prior to transfection (see Fig. 1A, further facilitating optimal genome coverage (see Materials and Methods). The number of T. b. brucei clones recovered was estimated to be ~1 million, equating to a library with ~8-fold genome coverage (4-fold in each direction); 17 of 22 T. b. brucei clones analysed (77%) contained inserts in the expected size-range, as determined using a long-range PCR assay (SI Appendix, Fig. S1B).

The workflow for library construction and screening is illustrated in Figure 1B. Prior to screening, we first carried out a quality-control step, both to determine the integrity and coverage of the T. b. brucei library, and to establish the screening protocol

---

Footnote Author

PNAS | Issue Date | Volume | Issue Number | 3

---

**Fig. 4.** Molecular docking and editing of the T. b. brucei CPSF3 active site. (A) The phylogenetic tree shows the relationship among CPSF3 proteins from humans and protozoal parasites, all discussed in the text. The domain structure for each protein is indicated to the right. Domains: grey, metallo-β-lactamase (residues 48-250 in T. brucei); green, β-CASP (residues 271-393 in T. brucei); blue, Zn-dependent metallo-hydrolase (residues 408-463 in T. brucei); rust, C-terminal (residues 502-746 in T. brucei). (B) Docking model for acoziborole (yellow structure) bound to T. b. brucei CPSF3. Gray spheres, zinc atoms; dotted lines, protein-ligand interactions. (C) Modified docking model illustrating a steric clash between acoziborole (red patch) and the His residue (beige) in Asn 232 in CPSF3. Other binding site residues that differ between the human and T. b. brucei proteins are indicated. (D) Modified docking model illustrating a steric clash between acoziborole (red patch) and a Tyr residue (beige) in human CPSF3, in place of Asn 132 in T. b. brucei CPSF3. The acoziborole molecular surface is otherwise represented in grey. (E) Modified docking model illustrating a steric clash between acoziborole (red patch) and the His residue (beige) in AcZb32/His substituted CPSF3. These back-up benzoxaboroles are also available if needed.

Oxaboroles display antiviral, antibacterial, antifungal, antiprotozoal and anti-inflammatory activity. Known targets include a viral protease (15), bacterial β-lactamase (16), bacterial leucyl-tRNA synthetase (17) and apicomplexan CPSF3, a metallo-β-lactamase (18, 19). Tavaborone (AN2690 or Kerdyin®) is an FDA-approved topical treatment for toenail onychomycosis (20) that targets the fungal cytoplasmic leucyl-tRNA synthetase (21). Oxaboroles are also potential antihypertensive rho-activated kinase inhibitors (22) and anti-inflammatory phosphodiesterase inhibitors for the treatment of psoriasis (23). Genomic and proteomic approaches, using oxaborole-1, a close analogue of acoziborole, previously yielded lists of candidate genes implicated in the mode-of-action (24), while metabolomic analysis following acoziborole exposure indicated perturbation of S-adenosyl-L-methionine metabolism (25), but the trypanocidal targets of acoziborole and AN11736 remained unknown.

RNA interference (RNAi) library screens have been very effective for identifying drug resistance mechanisms in trypanosomes, particularly due to defects in drug uptake or metabolism (26), but this loss-of-function approach does not typically reveal the target(s) of a drug. Indeed, RNAi screens recently revealed a pro-drug activation mechanism for a class of aminomethyl-benzoxaboroles, including AN3057, involving a T. b. brucei aldehyde dehydrogenase, but not the target (27).

A gain-of-function approach is more likely to identify drug-targets and proof-of-concept has been achieved using an overexpression library in T. b. brucei (28). However, the use of relatively short inserts (~1.2 kbp) and fusion with a common RNA-binding domain at the N-termini (29) precluded the overexpression of full-length native proteins, many proteins larger than ~50 kDa or protein targeting to the secretory pathway or mitochondrion (28). A related overexpression approach, with the potential to identify drug targets, Cos-seq, has also been described for Leishmania (30).

We assembled a T. b. brucei inducible overexpression library for full-length genes and with optimised genome coverage. Here, we describe and validate high-throughput overexpression, combined with massive parallel screening, and use the approach to identify the common target of several benzoxaboroles, including acoziborole and AN11736. The trypanocidal benzoxaborole target is the nuclear mRNA processing endonuclease, cleavage and polyadenylation specificity factor 3 (CPSF3/CPSF73/Ysh1).

Results

A high coverage T. b. brucei overexpression library. Our goal was to develop an unbiased bloodstream-form T. b. brucei overexpression library with optimised genome coverage. We first assembled the overexpression construct, pRPa® (30), and then constructed a library using T. b. brucei genomic DNA partially digested with Sau3AI (Fig. 1A). The T. b. brucei genome comprises 46% GC content (31) so Sau3AI sites (GATC) occur once every ~256 bp. The genome has only two known introns (32), such that the vast majority of genes comprise a single protein-coding exon with an average size of 1,592 bp; the core genome is also otherwise compact with average intergenic regions of 1,279 bp (31). Partially digested 3-10 kbp genomic DNA fragments were cloned in pRP® using a novel BbsI semi-filling strategy to optimise DNA fragments under the control of a tetracycline-inducible ribosomal RNA (rRNA) promoter (Fig. 1A). The DNA polymerase I promotor that can drive high-level expression of protein coding genes in trypanosomes (33); transcription by RNA polymerase I provides forty-fold stronger transcription than RNA polymerase II, which produces the vast majority of the natural transcriptome (34). The resulting plasmid library comprised >20 million clones from which 26 of 30 analysed (87%) contained inserts in the expected size-range (SI Appendix, Fig. S1A).

To establish a T. b. brucei library, we used a system for high-efficiency transfection previously used to assemble other libraries in bloodstream-form cells (35). This involves I-SceI meganuclease-mediated induction of a break at the chromosomal integration target site for pRP® (30). I-SceI was also used to linearize the plasmid library prior to transfection (see Fig. 1A, further facilitating optimal genome coverage (see Materials and Methods). The number of T. b. brucei clones recovered was estimated to be ~1 million, equating to a library with ~8-fold genome coverage (4-fold in each direction); 17 of 22 T. b. brucei clones analysed (77%) contained inserts in the expected size-range, as determined using a long-range PCR assay (SI Appendix, Fig. S1B).

The workflow for library construction and screening is illustrated in Figure 1B. Prior to screening, we first carried out a quality-control step, both to determine the integrity and coverage of the T. b. brucei library, and to establish the screening protocol.
Briefly, genomic DNA was extracted from the pooled library, fragments were amplified from the integrated overexpression constructs using long-range PCR, these products were deep-sequenced and the reads were mapped to the reference genome (Fig. 1B); pRpa^D8K-derived ‘barcodes’ were included to precisely map overexpressed fragments and to determine their orientation. This indicated that >95% of genes are included in the library and a representative 35 kbp region of the genome reveals excellent coverage, with fragment-junctions (barcoded reads) corresponding to the expected locations of Sau3AI sites (Fig. 1C). The barcodes also effectively reveal fragment orientation with respect to the RNA4 promoter in the overexpression construct. As expected, half of the fragments are in the ‘forward’ direction and half are in the ‘reverse’ direction (Fig. 1C, colour coding).

**Drug-target identification provides validation for the overexpression library.** Overexpression or ‘gain-of-function’ of a drug target can produce drug resistant cells by increasing the pool of functional protein, possibly by reducing intracellular free drug concentration through binding, or through both mechanisms simultaneously. Other resistance mechanisms are also possible (36). In the first instance, we wanted to determine whether our screening approach could identify a known drug-target. There are very few validated drug targets in Trypanosoma brucei, and therefore we used the well-characterised, experimental N-myristoyltransferase inhibitor, DDD85646 (37). The library was induced with tetracycline for 24 h prior to selection for nine days with 5.2 nM DDD85646 (~2 x EC₅₀, effective concentration of drug to reduce cell growth by 50%). Drug-selected cells displayed the expected reduced growth relative to a non-selected, but induced, control population (Fig. 1D). Since integrated, over-expressed fragments are replicated within the T. brucei genome, those that confer resistance will be over-represented following drug-selection.

Following the scheme above (Fig. 1B), a genome-scale map of hits was generated from the DDD85646 screen (Fig. 1E); the map includes the full non-redundant set of ~7,500 genes. This screen revealed a single major hit on chromosome 10 (Fig. 1E) comprising >66% of all mapped reads. A closer inspection revealed the barcoded junctions of a 3.982 kbp fragment. This fragment (>2.7 million mapped reads) encompassed one complete protein coding sequence (SI Appendix, Table S1) that encodes N-myristoyltransferase (Fig. 1E, inset; Tb927.10.14240, NMT), the known target of DDD85646 (37). The barcodes (green and black peaks) reveal the orientation of the gene and indeed, as expected, show a sense orientation of the coding sequence with respect to the RNA4 promoter driving overexpression.

Our approach also presents an opportunity to uncover novel biology beyond an immediate drug target. In the case of the NMT-inhibitor, we asked whether overexpression of essential NMT-substrate(s) would confer drug-resistance. Global profiling revealed 53 high-confidence candidate N-myristoylated proteins in T. brucei (38). Only four additional fragments with intact CDSs registered on the genome-scale map for the DDD85646-screen, and one of these (>54,000 mapped reads) encompassed an array of almost identical genes on chromosome 9 (Fig. 1E, SI Appendix, Table S1). These genes encode the N-myristoylated ADP-ribo-sylation factor 1 (ARF1), an essential Golgi protein required for endocytosis (39). Notably, both ARF1 knockdown (39) and DDD85646 exposure / NMT-inhibition (37) yield the same endocytosis-defective ‘BigEye’ phenotype (40).

Thus, the results of this screen support the view that DDD85646 is a specific inhibitor of T. brucei NMT and also that ARF1 is a major NMT-substrate that contributes to lethality when N-myristoylation is compromised. As a known target in ARF1 protein, and endocytosis are defective, Taken together, this provides excellent validation for the high-throughput overexpression library and for the genome-wide screening approach.

**Screens with benzoxaboroles identify CPSF3 as the probable target.** The overexpression library was induced as described above and selected for eight days with 1 µM acoxiborole (~2 x EC₅₀, SI Appendix, Fig. S2A and B). Sequencing and mapping of overexpressed inserts revealed another remarkably specific and dominant hit on the genome-scale map; >72% of all reads mapped to a single region on chromosome 4 (Fig. 2A, SI Appendix, Table S2; >6.1 million mapped reads). Only one additional fragment that registered on this map encompassed an intact CDS, encoding a hypothetical protein (Tb927.10.5630) on chromosome 10 (Fig. 2A, SI Appendix, Table S2), but this fragment registered <1% of the reads mapped to the major hit fragment, so was not investigated further. Closer inspection of the mapped reads on chromosome 4 revealed a single 4.276 kbp fragment encompassing one complete 2.725 kbp protein coding sequence; Tb927.4.1340 (Fig. 2A, SI Appendix, Table S2), Tb927.4.1340 encodes a subunit of the cleavage and polyadenylation specificity factor complex (CPSF3, Fig. 2A, inset). As above, the barcodes reveal the expected sense orientation of the fragment with respect to the promoter.

A screen with a second benzoxaborole, AN4169 / SCYX-6759, active against both T. brucei (12) and T. cruzi (13, 14), for eight days at 0.3 µM (~2 x EC₅₀, SI Appendix, Fig. S2C and D) revealed the same major hit as acoxiborole, Tb927.4.1340 / CPSF3 (Fig. 2B). Once again, this hit was remarkably specific and dominant, with no additional fragments registered on the genome-scale map in this case (Fig. 2B, SI Appendix, Table S3).

**Genetic validation of CPSF3 as the target of the trypanocidal benzoxaboroles.** The results above suggest that acoxiborole and SCYX-6759 are specific inhibitors of T. b. brucei CPSF3. To validate this hypothesis, we first generated independent T. b. brucei strains for inducible overexpression of a C-terminal GFP-tagged version of CPSF3; inducible overexpression of CPSF3 gps94 was confirmed by protein blotting (Fig. 3A). The EC₅₀ values for both acoxiborole and SCYX-6759 were determined; revealing that cells induced to overexpress CPSF3 gps94 displayed 5.7-fold (Fig. 3B) and 4.2-fold resistance (Fig. 3C), relative to uninduced cells, respectively. AN11176 is another benzoxaborole (see SI Appendix, Fig. S3A) that shows great promise for the treatment of animal African trypanosomiasis (11). We also determined the AN11176 EC₅₀ and found that cells induced to overexpress CPSF3 gps94 displayed 3.6-fold resistance (Fig. 3D). Thus, CPSF3 is likely the target of both the human and animal trypanocidal benzoxaboroles. Two additional benzoxaboroles (24, 27) (SI Appendix, Fig. S3A), and the NMT-inhibitor (37) as a control, were also tested. CPSF3 gps94 overexpressing cells displayed 4-fold resistance to oxaborole-1 (SI Appendix, Fig. S3B); 2.6-fold resistance to AN3057 (SI Appendix, Fig. S3C); and, as expected, no resistance to the NMT-inhibitor (Fig. 3E). Thus, CPSF3 appears to be the target of all five benzoxaboroles tested; acoxiborole, SCYX-6759, AN11176, oxaborole-1 and AN3057.

Inducible RNA interference was used to further genetically validate CPSF3 as a drug target. We suspected that CPSF3 knockdown would result in loss of viability, both because of its role in RNA processing and because our genetic screens above indicate that the trypanocidal benzoxaboroles act by specifically targeting CPSF3. In addition, prior genome-wide knockdown profiling indicated that CPSF3 was among only 16% of genes that registered a significant loss-of-fitness in multiple experiments (41). Inducible knockdown strains were assembled with a GFP-tagged native allele of CPSF3, and inducible knockdown of CPSF3 gps94 was confirmed by protein blotting (Fig. 3F). We observed a major loss-of-fitness following CPSF3 knockdown (Fig. 3G). Thus, T. b. brucei CPSF3 function is essential for viability.

The subcellular localisation of CPSF3 gps94 was also examined in cells expressing a native copy of CPSF3 gps94 (Fig. 3H) or over-expressing CPSF3 gps94 (Fig. 3I) with different functions; the former reveals an intermediate growth phenotype (40).
expressing CPSF3 SEP (Fig. 3). Consistent with its role in mRNA maturation, we observed the expected nuclear localisation for the protein expressed from the native locus (Fig. 3f). In contrast, and consistent with increased expression, we observed a more intense nuclear CPSF3 SEP signal and an additional, ectopic, cytoplasmic signal in cells overexpressing the protein (Fig. 3f).

Aciborole specifically blocks the catalytic site of trypanosome CPSF3. Analysis of CPSF3 sequences (Fig. 4A; SI Appendix, Fig. S4A) revealed that the T. brucei and T. gambiensesequences have identical metallo-β-lactamase and β-CASP domains, which form the catalytic site, with only two amino acid changes in the C-terminal domain. The livestock trypanosome, T. congolense, also displays a high degree of conservation within the metallo-β-lactamase and β-CASP domains, with only ten conservative amino acid changes. Phylogeny reveals a cluster of trypanosomatid sequences that is distinct from the human, apicomplexan and Trichomonas vaginalis sequences (Fig. 4A). This reflects divergence of trypanosomatids from the eukaryotic lineage but may also be partly explained by the unusual mechanism of coupled polyadenylation and trans-splicing in trypanosomatids (42). Without the benzoxaborole moiety, CPSF3 is conserved from trypanosomatids to humans (Fig. 4A).

To further explore the interactions between the benzoxaboroles and CPSF3, molecular docking studies were carried out (Fig. 4B). A T. brucei CPSF3 homology model was built using the Thermus thermophilus TTHA0252 structure complexed with an RNA substrate analogue [PDB: 3IE3 (43)] as a template. The catalytic site is located at the interface of the metallo-β-lactamase and β-CASP domains and comprises two zinc atoms coordinated by a network of histidine and aspartic acid residues (SI Appendix, Fig. S4A). In the model, the benzoxaborole moiety occupies the same area as the phosphate of the RNA where the two zinc atoms are present. The boron atom reacts with an activated water molecule located at the bi-metal centre leading to a tetrahedral negatively charged species that mimics the transition state of the phosphate of the RNA substrate (SI Appendix, Fig. S4B). Arg279, which in the template is involved in the recognition of a second phosphate group, extends towards the ligand to establish a hydrogen bond interaction with the amide oxygen and a cation-π interaction with the aromatic six-membered ring of the benzoxaborole. The amide in position 6 of the benzoxaborole ring directs the o-CF3, p-F phenyl ring of aciborole towards the area occupied by the terminal uracil base of the RNA substrate and establishes a face-to-face π-stacking interaction with Tyr265. The trifluoro methyl group and the fluorine atom on the phenyl ring do not appear to establish a specific interaction.

The proposed mode of binding is consistent with the established structure-activity relationships for this series (8, 9); specifically, the boron atom in the heterocyclic core is essential for trypanocidal activity and variation on the five-membered ring of the benzoxaborole had a pronounced (> 10-fold) effect, whereas variation on the phenyl ring distal to the benzoxaborole core had relatively little effect. The o-CF3 and p-F however provide a favourable pharmacokinetic/dynamic profile that results in an optimal response in the trypanosomiasis in vivo model. The docking model is also consistent with structural data available for other bi-metal systems, such as benzoxaboroles binding to phosphodiesterase 4 (23), cyclic boronate binding to β-lactamases (44), and the mechanism of action proposed for other RNA degrading proteins of the metallo-β-lactamase family (43). Thus, the proposed mode of aciborole binding indicates a steric block at the active site of trypanosome CPSF3 that perturbs the pre-mRNA endonuclease activity.

Finally, we asked why aciborole is well tolerated in animal models and in humans, given the similarity between T. brucei CPSF3 and human CPSF3 (SI Appendix, Fig. S4A). There are 26 residues within 5 Å of aciborole in the model shown in Figure 4B (SI Appendix, Fig. S4I) and only four of these residues differ between the T. brucei and human enzymes (Fig. 4C); all 26 residues are conserved amongst the trypanosomatids. The bulky Tyr side-chain in the human enzyme, in place of Asn232 in particular, presents a steric clash with the o-CF3,p-F phenyl moiety of aciborole (Fig. 4C). A Tyr residue is also present at this position in the apicomplexan sequences, while a His residue with a bulky side-chain is present at this position in the T. vaginalis sequence (SI Appendix, Fig. S5). A clash with aciborole in this position would likely prevent binding or disrupt the geometry required for covalent bond formation. This hypothesis was explored using CRISPR-Cas9 based editing (45) to mutate T. brucei CPSF Asn232. The results suggested that an Asn232 Tyr edit was not tolerated, and we found that T. brucei expressing an Asn232His substituted form of CPSF3 (SI Appendix, Fig. S5) were 5-fold resistant to aciborole (Fig. 4D). As expected, a His residue in place of Asn232 in our docking model, like the Tyr residue in the human enzyme, presents a steric clash with aciborole (Fig. 4E). Thus, specific structural differences between trypanosome and host CPSF3 offer an explanation for the safety profile and selective activity of aciborole.

Discussion
Phenotypic screening is a powerful strategy in drug discovery that yields many anti-infective lead compounds for potential development. The targets and mode-of-action of those compounds typically remain unknown, however. Although clinical development is possible without this knowledge, it is nonetheless highly desirable, and can facilitate further optimisation of therapeutics as well as the prediction and monitoring of drug-resistance.

Genomic and proteomic studies previously yielded a list of genes implicated in the mode-of-action of oxaborole-1; this list included CPSF3 and an adjacent glyoxalase, amplifiable in the genome in one of three resistant clones, but no candidate was identified using both approaches and the list of candidates was too extensive for a systematic appraisal (24). Our results now support an interaction between oxaborole-1 and CPSF3. CPSF3 orthologues were also recently identified as targets of another benzoxaborole (AN3661) active against malaria parasite, Plasmodium falciparum (19), and Toxoplasma gondii (18). Thus, CPSF3 is a common target of benzoxaboroles in the phylogenetically distant Apicomplexa and trypanosomatids. Our identification of T. brucei CPSF3 as the target of several benzoxaboroles now suggests that CPSF3 is also likely to be a promising drug-target, and the target of benzoxaboroles, in other protozoal parasites, including the other trypanosomatids, T. cruzi and Leishmania spp; all CPSF3 catalytic site residues are identical among these trypanosomatid sequences and the T. brucei sequence (SI Appendix, Fig. S4A). Notably, the complex is unlikely to couple polyadenylation to transcription termination in trypanosomatids since transcription is almost exclusively polycistrionic in these cells (48). Polyadenylation and trans-splicing of adjacent genes are coupled in trypanosomatids, however (42), and consistent with this, CPSF3 knockdown disrupted both polyadenylation and trans-splicing in T. brucei (49). Accordingly, inhibition of CPSF3 may explain perturbed methyl donor metabolism following aciborole exposure (25), which may be due to reduced mRNA methylation as a result of a splicing defect. CPSF3 also associates with the U1A snRNP splicing complex in trypanosomatids (50).

Molecular docking studies and evidence from the Asn232His mutant indicate how benzoxaboroles occupy the T. brucei CPSF3 active site, blocking pre-mRNA endonuclease activity. The presence of a bulky Tyr side-chain in the human CPSF3 suggests a
Unbiased genetic screens are remarkably powerful and can sample thousands of genes for specific phenotypes. We assembled a T. brucei overexpression library with optimised genome coverage. Massive parallel screens confirmed the mode-of-action of a known NMT-inhibitor as inhibition of endocytosis and revealed the mode-of-action of the benzoxaboroles as inhibition of pre-mRNA maturation. CPSF3 can now be considered genetically, chemically and clinically validated as an important novel antitypanosomal drug target. Thus, our studies indicate that both clinical and veterinary benzoxaboroles kill trypanosomes by specifically inhibiting the parasite mRNA processing endonuclease, CPSF3.

Materials and Methods

For details of T. brucei growth and manipulation, pPLaHAP plasmid library assembly, T. brucei library assembly, overexpression screening, NGS data analysis, plasmid construction, EC50 assays, western blotting, microscopy, docking studies and Cas9-based editing; see SI Appendix, Materials and Methods.

ACKNOWLEDGMENTS. We thank N. Zhang, A. Fairlamb, S. Patterson, P. Wyatt and M. De Rycker for providing samples of antitypanosomal compounds; and S. Hutchison for advice on sequence-mapping. This work was supported by a Senior Investigator Award to D.H. (100320/Z/12/Z), a Strategic Award (105021/Z/14/Z) and a Centre Award (20134/Z/16/Z), all from the Wellcome Trust; and an award from the UK Medical Research Council (MR/K008749/1).

Footline Author