

# UCSF

## UC San Francisco Previously Published Works

### Title

CRISPR-based oligo recombineering prioritizes apicomplexan cysteines for drug discovery.

### Permalink

<https://escholarship.org/uc/item/5jr82346>

### Journal

Nature Microbiology, 7(11)

### Authors

Benns, H  
Storch, M  
Falco, J  
et al.

### Publication Date

2022-11-01

### DOI

10.1038/s41564-022-01249-y

Peer reviewed

# CRISPR-based oligo recombineering prioritizes apicomplexan cysteines for drug discovery

Received: 10 June 2022

Accepted: 9 September 2022

Published online: 20 October 2022

 Check for updates

H. J. Bennis<sup>1,2</sup>, M. Storch<sup>3</sup>, J. A. Falco<sup>4</sup>, F. R. Fisher<sup>1</sup>, F. Tamaki<sup>5</sup>, E. Alves<sup>1</sup>, C. J. Wincott<sup>1</sup>, R. Milne<sup>5</sup>, N. Wiedemar<sup>5</sup>, G. Craven<sup>2</sup>, B. Baragaña<sup>5</sup>, S. Wyllie<sup>5</sup>, J. Baum<sup>1,6</sup>, G. S. Baldwin<sup>1</sup>, E. Weerapana<sup>4</sup>, E. W. Tate<sup>2</sup>✉ and M. A. Child<sup>1</sup>✉

Nucleophilic amino acids are important in covalent drug development yet underutilized as anti-microbial targets. Chemoproteomic technologies have been developed to mine chemically accessible residues via their intrinsic reactivity towards electrophilic probes but cannot discern which chemically reactive sites contribute to protein function and should therefore be prioritized for drug discovery. To address this, we have developed a CRISPR-based oligo recombineering (CORE) platform to support the rapid identification, functional prioritization and rational targeting of chemically reactive sites in haploid systems. Our approach couples protein sequence and function with biological fitness of live cells. Here we profile the electrophile sensitivity of proteinogenic cysteines in the eukaryotic pathogen *Toxoplasma gondii* and prioritize functional sites using CORE. Electrophile-sensitive cysteines decorating the ribosome were found to be critical for parasite growth, with target-based screening identifying a parasite-selective anti-malarial lead molecule and validating the apicomplexan translation machinery as a target for ongoing covalent ligand development.

Electrophilic small molecules that engage protein-encoded amino-acid nucleophiles are resurgent in drug discovery as versatile chemical probes and therapeutic agents<sup>1,2</sup>. As a result, considerable efforts are devoted to the development of chemical proteomic technologies termed ‘reactivity-based profiling’ for the identification of nucleophilic sites<sup>3</sup>. Isotopic tandem-orthogonal activity-based protein profiling (isoTOP-ABPP) is a standard method for proteome-wide profiling of intrinsic amino-acid reactivity<sup>4</sup>. Central to isoTOP-ABPP is the use of quantitative mass spectrometry (MS) to measure the extent of protein labelling with a highly reactive electrophilic probe. Initially applied to rank the reactivity of cysteines in the human proteome using an

iodoacetamide (IAA) probe<sup>4</sup>, isoTOP-ABPP has since been expanded to other amino-acid types, including lysine<sup>5</sup>, methionine<sup>6</sup> and tyrosine<sup>7</sup>. Moreover, this method has been successfully adapted for competitive screening of covalent fragments<sup>5,8–10</sup>, enabling identification of sites that can be pursued in fragment-based lead discovery programmes for ‘inverse drug discovery’<sup>11</sup>.

Despite advances in reactivity-based profiling, the prioritization of chemically reactive or ligandable amino acids as targets following their proteomic identification remains biased; target selection is typically based on the availability of existing functional information or assays for the associated protein class. This inevitably leads to proteins with

<sup>1</sup>Department of Life Sciences, Imperial College London, London, UK. <sup>2</sup>Department of Chemistry, Imperial College London, London, UK. <sup>3</sup>London Biofoundry, Imperial College Translation & Innovation Hub, London, UK. <sup>4</sup>Department of Chemistry, Boston College, Boston, MA, USA. <sup>5</sup>Wellcome Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dundee, UK. <sup>6</sup>School of Biomedical Sciences, UNSW, Sydney, NSW, Australia. ✉e-mail: [e.tate@imperial.ac.uk](mailto:e.tate@imperial.ac.uk); [m.child@imperial.ac.uk](mailto:m.child@imperial.ac.uk)

untapped therapeutic value being overlooked<sup>12</sup>. The ability to efficiently interrogate individual amino acids across the proteome at high throughput would expand our understanding of protein sequence–function relationships in complex biological systems, and solve one of the grand challenges of universal inverse drug discovery.

Over recent years, multiplexed ‘recombineering’ screens (for example, MAGE, CRMAGE and CREATE) have been developed to simultaneously map the phenotypic effects of thousands of amino-acid substitutions across genomes<sup>13–15</sup>. These platforms monitor the allelic frequency of amino-acid mutants in a population over a period of selective pressure or growth, enabling the identification of substitutions that impact cellular fitness. Such methods typically indirectly estimate mutant frequency, limiting their ability to probe sequence–function relationships. Other technologies overcome these limitations by directly sequencing the modified chromosomal loci<sup>16–18</sup>. However, their application has been restricted to single or small panels of targets (for example, isolated exons or individual loci in saturation mutagenesis) and/or a limited range of amino-acid substitution types. Therefore, a strategy for direct, quantitative assessment of the contribution of individual amino acids to protein function across diverse genomic loci (such as sites identified by reactivity-based profiling or post-translationally modified protein networks) is needed.

In this Article, we introduce CRISPR-based oligo recombineering (CORE) for proteome-wide assessment of amino-acid contribution to protein function in cells. Combined with isoTOP-ABPP, we apply CORE to the eukaryotic pathogen *Toxoplasma gondii*, functionally annotating electrophile-sensitive cysteines and prioritizing new therapeutic targets. Validating our strategy, we leverage *T. gondii*'s evolutionary relationship with the malaria parasite, *Plasmodium falciparum*, and confirm that cysteine-targeted electrophilic fragments can inhibit *P. falciparum* protein translation. Our work introduces CORE for scalable protein sequence–function studies and the expedient, unbiased prioritization of chemically reactive sites, proteins and biological processes for ligand discovery.

## Cysteine reactivity profiling in *T. gondii*

We sought to establish a platform for the prioritization of covalently druggable sites on protein targets. While we intended our final CORE target prioritization platform to be amino acid agnostic, for proof of concept we focused on electrophile-sensitive cysteines in the apicomplexan parasite *T. gondii*. *T. gondii* is an experimentally tractable eukaryotic host–pathogen model<sup>19</sup> with medical and veterinary importance<sup>20</sup>. However, current frontline therapeutics are ineffective against the chronic bradyzoite lifecycle stage and limited by toxicity<sup>21</sup>, highlighting the need for rapid identification and prioritization of new therapeutic targets.

We applied a variant of the isoTOP-ABPP platform that uses an iodoacetamide-alkyne (IA-alkyne) probe and isotopically differentiated

biotin tags with cleavable azobenzene linkers<sup>22</sup> to *T. gondii* (Fig. 1a). We identified a total of 1,097 cysteines in 691 proteins that were labelled by IA-alkyne in extracellular *T. gondii* tachyzoites (Supplementary Tables 1 and 2). Similarly to previous studies<sup>4,23,24</sup>, individual cysteines displayed a range of inherent reactivity towards the probe (Fig. 1b). Amino-acid ‘hyperreactivity’ was previously found to be a predictor of functionality in cells<sup>4,5</sup>, and we hypothesized the same might be true in *T. gondii*. We therefore partitioned cysteines by their respective isotopic ratios into high ( $R < 3$ ), medium ( $R = 3–5$ ) and low ( $R > 5$ ) reactivity groups. As reactivity is a relative concept determined by the electrophilic component of the probe, we refer to the spread of residue reactivity in terms of electrophile sensitivity from hereon. In total, 130 highly electrophile-sensitive cysteines were identified in 97 proteins with diverse biological functions (Extended Data Fig. 1a). This includes proteins with known cysteine-based catalytic mechanisms (for example, thioredoxins), well-characterized parasite proteins for which no functional role has previously been attributed to the identified cysteines (for example, myosin F) and hypothetical proteins (Supplementary Table 3). Analysis of functional annotations assigned to highly electrophile-sensitive cysteine-containing genes revealed enrichment of translation-associated proteins, including those of the ribosome (Fig. 1c, Extended Data Fig. 1a and Supplementary Table 4), which were not correlated with protein abundance (Extended Data Fig. 1b) and were absent from similar datasets obtained from other eukaryotic cell systems<sup>4,23,24</sup>.

We next assessed the association of highly electrophile-sensitive cysteines with gene essentiality according to ‘phenotype scores’ from a genome-wide CRISPR knockout (KO) screen in *T. gondii* (Fig. 1d)<sup>25</sup>. The more negative a gene's phenotype score, the more fitness conferring the gene's function is in vitro. Using a phenotype score threshold of  $-2$  or below as an indicator of gene essentiality, we observed enrichment of indispensable genes in our electrophile-sensitive cysteines dataset relative to all protein-coding genes or protein-coding genes containing at least one cysteine. No difference in the distribution of phenotype scores was observed between the low-, medium- and high-electrophile-sensitivity groups. Phenotype scoring identified a focused group of 75 highly electrophile-sensitive cysteines in 56 essential genes (Supplementary Table 3). Phylogenetic analysis of these targets indicated varying degrees of cysteine conservation across different protein classes and eukaryotes (Fig. 1e). Interestingly, several sites appeared to be widely conserved in clinically important pathogens yet absent in the human host, suggesting the potential for these cysteines to be selectively targeted with cysteine-directed drugs.

## CORE platform rationale and design

To systematically interrogate cysteines identified by isoTOP-ABPP, we designed a methodology to prioritize individual sites on the basis of demonstrated contribution to protein function in cells. We refer to our

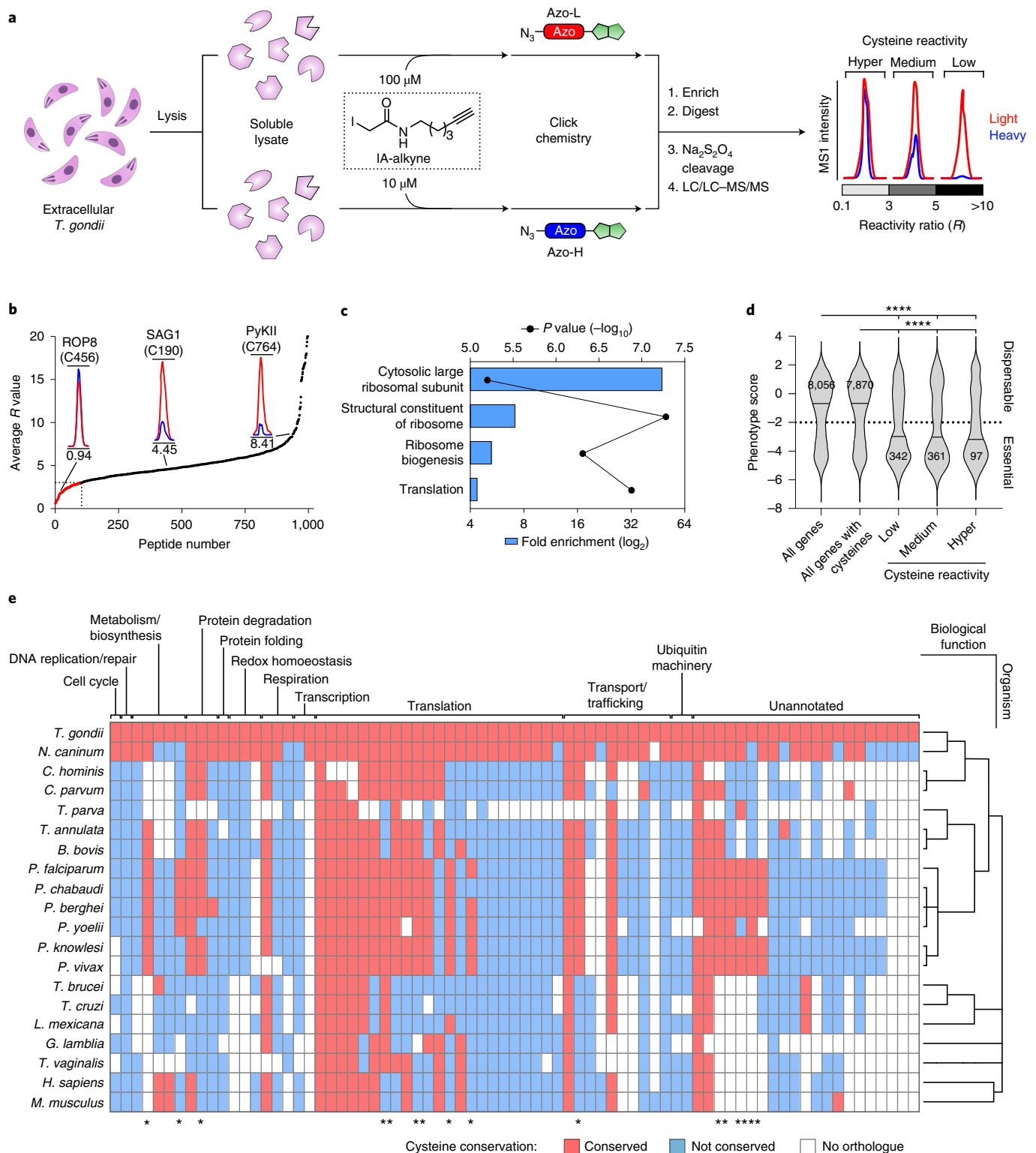
### Fig. 1 | Cysteine reactivity profiling in *T. gondii* reveals enrichment of highly electrophile-sensitive cysteines in essential and translation-associated proteins.

**a**, isoTOP-ABPP workflow for quantifying cysteine electrophile-sensitivity in *T. gondii* parasites. Soluble lysates from extracellular tachyzoites were independently labelled with high (100  $\mu$ M) and low (10  $\mu$ M) concentrations of a thiol-reactive IA-alkyne probe. Labelled samples were then click-conjugated to isotopically differentiated, reductant-cleavable biotin tags (heavy (blue) and light (red) for 10  $\mu$ M and 100  $\mu$ M treatment groups, respectively), combined and enriched on streptavidin-immobilized beads. Immobilized proteins were then subject to tandem on-bead trypsin digestion and sodium hydrosulfite treatment, eluting probe-modified peptides for LC/LC–MS/MS analysis. Cysteine electrophile sensitivity is quantified by  $R$  values, which represent the differences in MS1 peak intensities between the light- and heavy-conjugated proteomes. **b**, Ranked average  $R$  values for probe-labelled peptides from two independent experiments ( $n = 2$ ). Representative chromatograms of cysteines within three groups of reactivity (high,  $R < 3$ ; medium,  $R = 3–5$ ; low,  $R > 5$ ) are annotated. **c**, Enrichment analysis of functional annotations in annotated genes containing

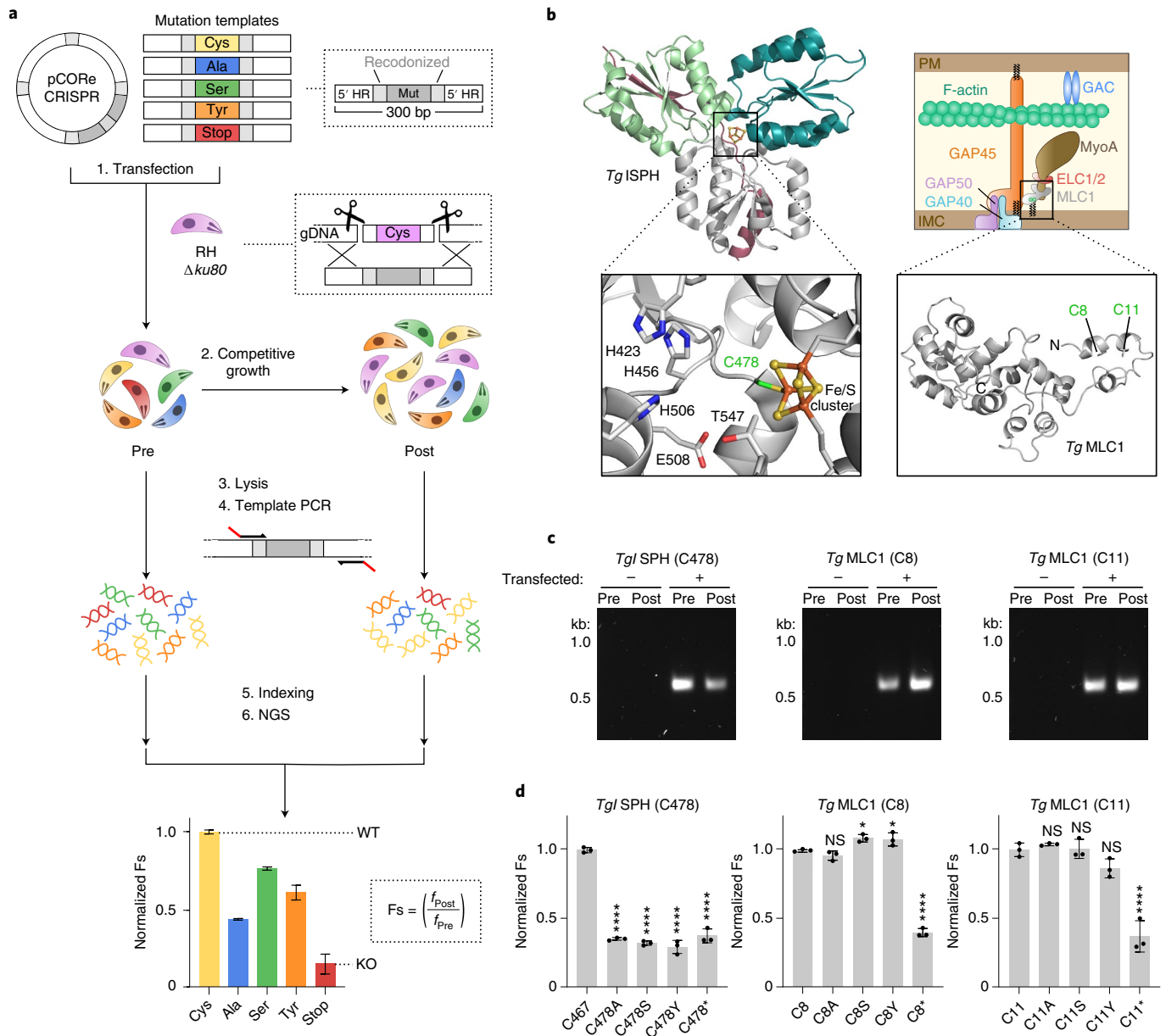
highly electrophile-sensitive cysteines relative to the *T. gondii* genome. Fold change is plotted against statistical significance determined from a two-tailed Fisher's exact test. **d**, Comparative distribution analysis of published phenotype scores<sup>25</sup> for the *T. gondii* tachyzoite genome with all cysteine- and electrophile-sensitive cysteine-containing genes. Essential genes are classified by a score of  $< -2$ . Statistical significance was assessed using two-tailed Kolmogorov–Smirnov  $t$ -test ( $****P < 0.0001$ ). **e**, Conservation of highly electrophile-sensitive cysteines identified in essential *T. gondii* genes across eukaryotic orthologues in *Neospora caninum*, *Cryptosporidium hominis*, *Cryptosporidium parvum*, *Theileria parva*, *Theileria annulata*, *Babesia bovis*, *Plasmodium chabaudi*, *Plasmodium berghei*, *Plasmodium yoelii*, *Plasmodium knowlesi*, *Plasmodium vivax*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania mexicana*, *Giardia lamblia*, *Trichomonas vaginalis*, *Homo sapiens*, *Mus musculus*. Cysteines are grouped by the predicted function of their associated genes, and organisms by their phylogenetic relationship. Asterisks indicate residues highly conserved in eukaryotic pathogens, but absent in mammalian systems.

approach as CORE (Fig. 2a). The underlying principle of CORE is that for essential genes there is a direct relationship between the molecular function of the encoded protein and cellular fitness; mutations that perturb protein function will similarly impact cellular fitness. By comparing the fitness of wild-type (WT) and cysteine mutants for a specific electrophile-sensitive cysteine-containing gene product, the functional contribution of the target residue can be assessed within the sequence context of the protein. This is achieved through site-specific integration of different mutations using a CRISPR-Cas9-based homology-directed

repair (HDR) strategy, and subsequent quantitative comparison of the fitness of the resulting reactive site mutant(s) with WT and KO controls. Highly electrophile-sensitive cysteines were selected as targets for their higher intrinsic nucleophilicity to covalent ligands rather than any presumed or predicted functionality. This is important for downstream fragment-based lead discovery campaigns, which typically use warheads with lower reactivity than IAA. For functional interrogation of highly electrophile-sensitive cysteines in *T. gondii*, we selected five mutation types; a recodoned cysteine (synonymous







**Fig. 2 | CORE discriminates between fitness-conferring and non-fitness-conferring chemically reactive sites.** **a**, Workflow of CORE for functional interrogation of highly electrophile-sensitive cysteines in *T. gondii*. A single pCORe CRISPR plasmid is co-transfected into *T. gondii* parasites with a panel of linear double-stranded donor templates that encode different codon switches (a recodonomized cysteine codon, alanine, serine, tyrosine and stop codon). Each plasmid encodes Cas9 nuclease and two gRNA cassettes that direct Cas9 to induce DSBs at sites 5' and 3' of a target cysteine codon. This promotes integration of templates at the excised genomic locus via HDR, substituting the endogenous cysteine for a given mutation. To increase the efficiency of HDR, a cell line deficient in NHEJ-based DNA repair is used (RH $\Delta ku80$ ) (ref. 36). Genomic DNA from the transfected parasite population is extracted before ('Pre') and after ('Post') competitive lytic growth. For each timepoint, specific amplicons are generated by targeting primers to regions of recodonomized sequence within the templates. The abundance of each mutation is quantified by NGS. The read frequency of each mutant in 'Post' ( $f_{Post}$ ) is normalized to 'Pre' ( $f_{Pre}$ ) to determine

$F_s$  that reflect the viability of parasites following amino-acid substitution.  $F_s$  values for the amino-acid substitutions are compared against the synonymous recodonomized cysteine (WT) and stop codon (KO) mutations to identify deleterious mutations (that is, functional cysteines). **b**, Structural models of CORE targets *Tg*ISP (left) and *Tg*MLC1 (right). Insets show the positions of their associated target cysteines. IMC, inner membrane complex; PM, plasma membrane; GAC, glideosome-associated connector. **c**, Amplicons generated following mutation of *Tg*ISP (C478) and *Tg*MLC1 (C8/C11). Agarose gel shown is representative of three independent experiments. **d**, Histograms showing  $F_s$  values for cysteine mutants of *Tg*ISP (C478) and *Tg*MLC1 (C8/C11), normalized to the recodonomized cysteine control. Data represent mean  $\pm$  s.d. values for three independent experiments ( $n = 3$ ). Statistical significance for each mutant was compared against the recodonomized cysteine control by one-way ANOVA with Dunnett's correction for multiple comparisons (\*\*\*\* $P < 0.0001$ ; \* $P < 0.05$ ; NS, no significance,  $P > 0.05$ ).

replacement of the target cysteine; WT), a stop codon (for disruption of the target gene; KO<sup>26</sup>) and three distinct amino-acid substitutions: alanine, serine or tyrosine. While alanine and serine are commonly used in mutagenesis studies, tyrosine was included to probe sites that

could participate in protein–protein interactions (PPIs). Meta-analysis of PPI mutation datasets obtained from cancer studies<sup>27</sup> revealed that tyrosine is the most frequent cysteine substitution that causes destabilization at PPI interfaces (Supplementary Table 6). We reasoned that

a tyrosine mutation might identify cysteine-dependent PPI hotspots, while acknowledging the caveat that a large aromatic substitution may affect protein function via folding defects. Details on the optimization of CORE are provided in Methods.

## CORE functionally prioritizes cysteine targets in live cells

We trialled CORE against electrophile-sensitive cysteines in two targets: *Tg*ISPH (TGGT1\_227420) and *Tg*MCL1 (TGGT1\_257680) (Fig. 2b). ISPH (also known as 'LytB') is an oxidoreductase essential for isoprenoid biosynthesis and an established anti-microbial drug target<sup>28,29</sup>. The electrophile-sensitive cysteine identified in *Tg*ISPH contributes to an iron-sulfur cluster that is required for enzyme catalysis, and therefore any substitutions at this site are expected to be deleterious. *Tg*MCL1 is part of the glideosome complex required for parasite motility and host-cell invasion<sup>30</sup> and contains two N-terminal electrophile-sensitive cysteines that are known to be S-acylated<sup>31,32</sup> but do not contribute *Tg*MCL1 function<sup>33</sup>. We predicted that substitutions at these residues would not affect parasite fitness. We applied CORE to these targets; integration-specific amplicons were successfully generated (Fig. 2c), and next-generation sequencing (NGS) analysis confirmed our expectations with high biological reproducibility. Any substitution of the *Tg*ISPH-associated cysteine negatively impacted parasite fitness, with all three mutations being analogous to disruption of the gene following integration of the stop codon (Fig. 2d). In agreement with published data, the cysteines on *Tg*MCL1 were permissive to all mutations, indicating that these residues (and their post-translational modification) do not contribute to the essential component of this protein's function (Fig. 2d). We next benchmarked a standard genetic analysis workflow against which CORE could be compared. For this purpose, we selected a hyperreactive cysteine associated with a hypothetical protein (TGGT1\_258070). We generated an inducible KO (iKO) line using the DiCre system (RH *Tg*Hypo<sup>iKO</sup>) (Extended Data Fig. 2a), and confirmed the expected genomic re-arrangement by polymerase chain reaction (PCR), expression of the protein by western blot, and localization by immunofluorescence microscopy (Extended Data Fig. 2b–d). Treatment of RH *Tg*Hypo<sup>iKO</sup> parasites with rapamycin resulted in efficient gene excision and protein degradation (Extended Data Fig. 2e,f), and in agreement with the gene's phenotype score (−5.24) plaque assay confirmed that KO parasites were not viable (Extended Data Fig. 2g,h). We then sought to assess the contribution of the reactive cysteine to protein function by genetic complementation in the genetic background of this iKO. We were unable to complement for the loss of this gene, precluding functional interrogation of the associated cysteine. Supplementary Fig. 1 summarizes the comparative timeline for CORE with this standard approach.

We applied CORE to our complete set of highly electrophile-sensitive cysteine-containing essential genes. Construction of 59 CRISPR plasmids was accomplished in 5 days using a linker-based DNA assembly strategy (Extended Data Fig. 3a,b), followed by parasite transfection and competitive lytic growth (8 days), integration-specific

amplicon production (achieving 100% coverage for our target cysteines; Supplementary Fig. 2), NGS library construction (7 days) and Illumina NextSeq processing (7 days). The entire CORE workflow took approximately 1 month to complete for 74 electrophile-sensitive cysteine targets, with the final dataset indicating exceptional reproducibility across independent biological replicates. These data are summarized in Fig. 3a and Supplementary Fig. 3. For ~90% of the target cysteines (66/74), the integration of the premature stop codon resulted a significant ( $P < 0.05$ ) reduction in parasite fitness (Fig. 3b). Interestingly, the relative magnitude of the effect of integrating the stop codon did not correlate with published gene phenotype scores (Extended Data Fig. 4a,d), and no deleterious growth phenotype was detected for stop codon mutants in eight targets. This may reflect the proximity of these mutagenized cysteine to the protein C-terminus as these proteins could retain functional domains. Despite a lack of overall correlation between cysteine position and stop codon disruption or phenotype score (Extended Data Fig. 4b,c), several of these are positioned C-terminally to predicted domains (Extended Data Fig. 4e).

Analysis capturing aspects of both the magnitude and statistical significance of the effect of each individual substitution identified robustly fitness conferring cysteines, prioritizing target sites according to their contribution protein function in live cells (Fig. 3b and Supplementary Table 7). The majority of substitutions were benign (−83%, 184/222), with only a small fraction of the highly electrophile-sensitive cysteines measurably contributing to the function of the protein (−17%, 38/222) (Fig. 3c). Unexpectedly, CORE identified gain- as well as loss-of-function mutations. For several targets (for example, *Tg*GAPDH2), we also recognized that individual cysteines can show different mutational tolerances (Extended Data Fig. 5), indicating that CORE can effectively discriminate between functional and non-functional sites within a single protein. Illustrating the challenge of selecting targets in the absence of an approach such as CORE, there was no association between the Fs of a given cysteine and the effect of stop codon integration (Fig. 3d), the phenotype score of the associated gene (Fig. 3e) or increased electrophile sensitivity of the cysteine itself (Fig. 3f). The challenge of target selection is highlighted by the cysteine chosen for validation via the standard genetic workflow (Extended Data Fig. 2). For this hypothetical protein, CORE indicated that the reactive cysteine does not contribute to protein function (Fig. 3a and Supplementary Fig. 3). Addressing the relationship between cysteine 'essentiality' and function, we compared the extent of conservation for fitness-conferring and non-fitness-conferring cysteines according to 'conservation scores' (Fig. 3g and Supplementary Table 5). While non-fitness-conferring cysteines appeared to be normally distributed across the analysed species, fitness-conferring cysteines displayed a bimodal distribution with higher scores. This indicated that conservation should not be taken as the sole predictor of function.

Integration of three different amino-acid substitutions enabled deeper interrogation of each site, and an increased appreciation of functionally disruptive biochemistry (Fig. 3h). In agreement

**Fig. 3 | CORE prioritizes apicomplexan protein translation as a target for covalent inhibition.** **a**, Heat map showing normalized Fs values for all target cysteines and mutation types ordered by the mutation sensitivity of the cysteines (high to low, top to bottom). **b**, Volcano plot showing the normalized Fs values of each cysteine mutation and statistical significance against the recodonized cysteine control as determined by one-way ANOVA. Data represent mean Fs values for three independent experiments ( $n = 3$ ). Significant mutations ( $P < 0.05$ ) with mean Fs values  $< 0.66$  and  $> 1.33$  represent deleterious and gain of function, respectively, and are coloured. Only cysteines (66/74) featuring a deleterious stop codon mutation are shown and used in subsequent analyses. **c**, Proportion of amino-acid substitutions causing deleterious or gain-of-function phenotypes. **d**, **e**, Distribution of normalized stop codon Fs values (**d**) and phenotype scores (**e**) between genes containing at least one ( $n = 19$ ) or no fitness-conferring cysteines ( $n = 31$ ), and isoTOP-ABPP *R* values (**f**) of the fitness

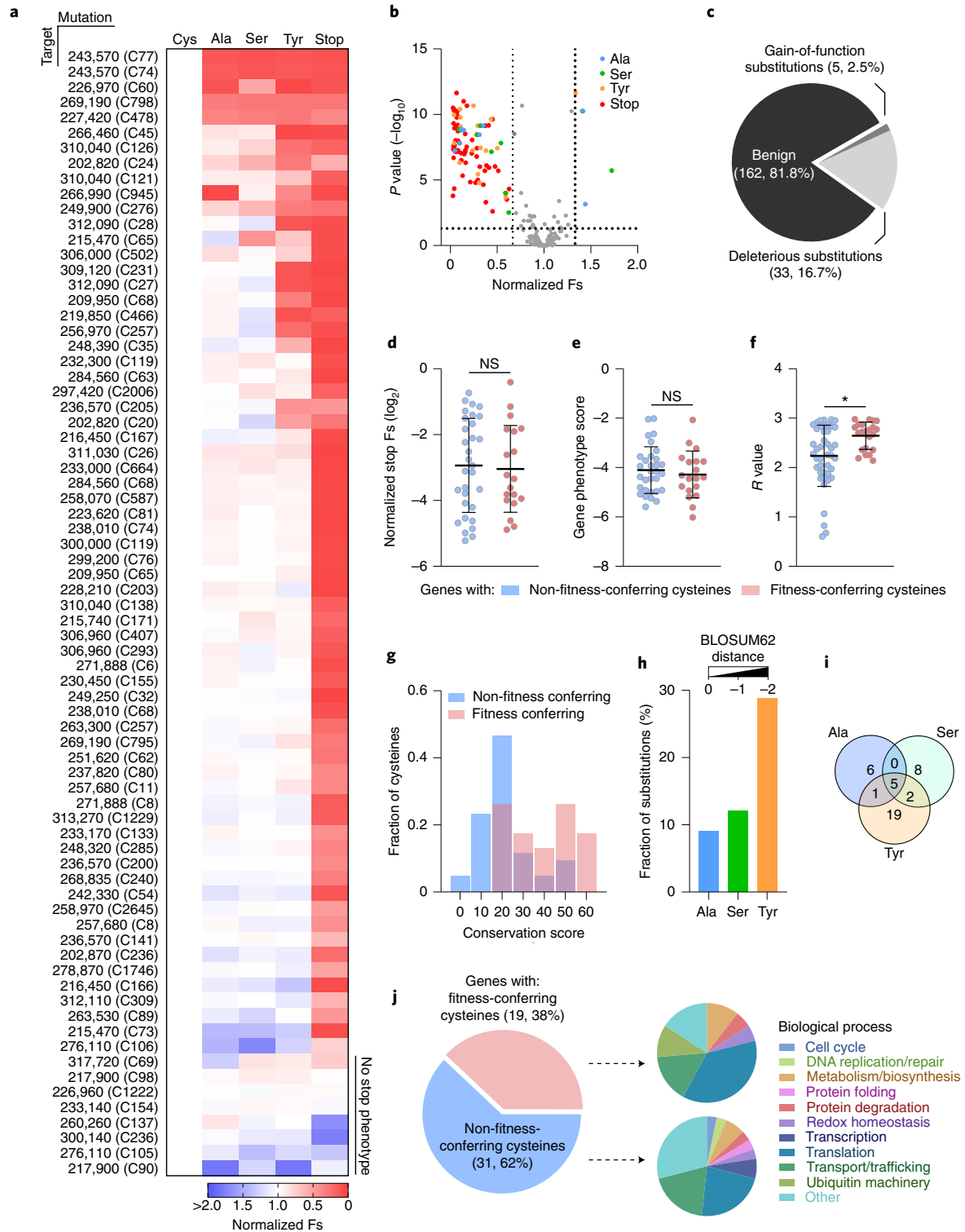
conferring ( $n = 23$ ) or non-fitness conferring cysteines ( $n = 43$ ) themselves. Bars represent mean  $\pm$  s.d. Statistical differences between group means were assessed by two-tailed Student's *t*-tests ( $*P < 0.05$ ; NS, no significance,  $P > 0.05$ ). *P* values: stop codon Fs values, 0.6702; phenotype scores, 0.3498; *R* values, 0.0150.

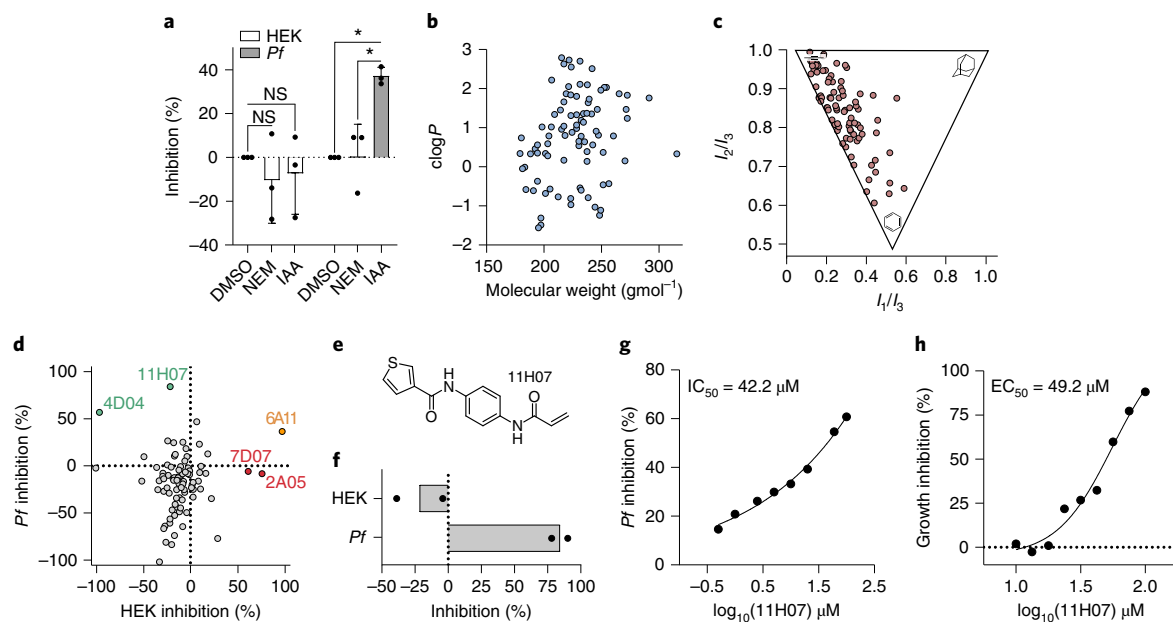
**g**, Frequency distribution of conservation scores assigned to fitness-conferring and non-fitness-conferring cysteines across 20 eukaryotic organisms; higher scores indicate wider conservation across the analysed species. **h**, Fraction of deleterious amino-acid substitutions for each mutation type. The BLOSUM62 distance scores for each substitution are annotated and organized by increasing distance from the native cysteine residue (left to right). **i**, Overlap of cysteines with deleterious alanine, serine and/or tyrosine substitutions. **j**, Proportion and functional annotations of proteins containing fitness-conferring and non-fitness-conferring cysteines.

with anticipated evolutionary mutational tolerance, the greater the BLOSUM62 matrix distance<sup>34</sup> between the individual mutation and cysteine, the more likely the mutation affected the function of the associated protein. This supports a degree of functional buffering or resistance against gradual evolutionary change of protein function as a result of changes in protein sequence, with a range of tolerance observed for each individual cysteine (Fig. 3i). We recognized that the functional annotations provided by CORE might encompass impacts upon protein folding and stability. To test this, we obtained or predicted high-confidence structures for 35/56 genes using the Protein Data

Bank, AlphaFold<sup>35</sup> or Phyre2<sup>36</sup>, and used FoldX<sup>37</sup> to predict the change in Gibbs free energy ( $\Delta\Delta G$ ) resulting from the cysteine substitutions. These  $\Delta\Delta G$  values were then compared with CORE Fs values (Extended Data Fig. 6). While the more radical tyrosine substitution was predicted to have a greater impact upon protein stability for a subset of targets, there was no clear correlation between CORE Fs,  $\Delta\Delta G$  and a particular substitution type.

To improve our chances of finding an individual druggable target, we initially sought to identify broader biological processes with sensitivity to cysteine-reactive covalent small molecules. We performed





**Fig. 4 | The apicomplexan translation machinery is selectively inhibited by thiol-reactive small molecules.** **a**, Mean  $\pm$  s.d. inhibition of *P. falciparum* (*Pf*) and HEK293 (HEK) IVT with 100  $\mu$ M IAA or *N*-ethylmaleimide (NEM). Translational output was measured from cell lysates using a luciferase-based IVT assay<sup>41</sup>. DMSO was used as a vehicle control. Statistical significance was determined from three independent experiments ( $n = 3$ ) by two-way ANOVA with Šidák's correction for multiple comparisons ( $*P < 0.05$ ; NS, no significance,  $P > 0.05$ ). HEK  $P$  values: DMSO versus NEM, 0.7124; DMSO versus IAA, 0.8754. *Pf*  $P$  values: DMSO versus NEM, 0.0114; DMSO versus IAA, 0.0122. **b**, Molecular weight versus the  $\text{clog}P$  value of 88 acrylamide-containing fragments.

**c**, Normalized PMI ratios of the acrylamide fragment library. Ratios are plotted in a triangular graph to depict the molecular shape diversity, where the vertices represent a perfect rod ( $x = 0, y = 1$ ), disc ( $x = 0.5, y = 0.5$ ) and sphere ( $x = 1, y = 1$ ). **d**, Mean percentage inhibition of HEK293 and *Pf* IVT following treatment of cell lysates with 100  $\mu$ M of each acrylamide fragment. Representative compounds inhibiting *Pf*, HEK293 or both lysates are annotated green, red and orange, respectively. **e**, Chemical structure of hit compound, 11H07. **f**, Expanded IVT profile for 11H07 from **d**. **g, h**, Concentration-dependent inhibition of *Pf* IVT (**g**) and growth (**h**) with 11H07. Data in **d** and **f–h** represent two independent experiments ( $n = 2$ ), each with three technical replicates.

an enrichment analysis of functional annotations assigned to proteins containing fitness-conferring or non-fitness-conferring cysteines (as defined by CORE) (Fig. 3j). The breakdown for the two groups was distinct, with 'translation' annotation enriched in genes containing fitness-conferring cysteines. We undertook an in-depth analysis of CORE-prioritized electrophile-sensitive cysteines present on proteins associated with translation. The majority of these sites (9/10) were encoded in proteins decorating the surface of the cytoplasmic 80S ribosome (Extended Data Fig. 7a–c). To further understand the mutational tolerance of these residues, we applied CORE for site-saturation mutagenesis of ten ribosomal cysteines. As expected, CORE identified a broad spectrum of fitness changes (Extended Data Fig. 7d, Supplementary Fig. 4 and Supplementary Table 8). This included residues tolerant to all substitutions, as well as sites intolerant to any substitution presumably where the functional contribution of the residue is dependent on the cysteine thiol. FoldX analysis indicated that, while aromatic substitutions had the greatest impact on protein folding stability (Extended Data Fig. 7e), there was no consistency in the effect of deleterious substitutions upon  $\Delta\Delta G$  across targets (Extended Data Fig. 7f). We focused on RPL4(C231) to validate these data as the mutational profile for this site suggested that specific amino-acid side chains were not tolerated. Dedicated ribosome assembly chaperones Kap104 and Acl4 have been documented to sequester Tom1-dependent polyubiquitination sites present on the surface of yeast RPL4, and thereby protect unassembled RPL4 from proteasomal degradation<sup>38</sup>. We hypothesized that, should these interactions be conserved in *T. gondii*, deleterious RPL4(C231) substitutions may not be tolerated owing to a disruptive effect on these PPIs leading to increased cellular degradation of RPL4. Our saturated substitution dataset and FoldX analysis directed our selection of conservative benign (C231A/S) and non-conservative substitutions found to be fitness conferring but

with different predicted effects on  $\Delta\Delta G$  (C231D/Y) (Extended Data Fig. 8a). We generated parasite lines ectopically overexpressing a second copy of RPL4 and C231 mutants, confirming the genomic integration, expression and expected localization of all constructs (Extended Data Fig. 8b–e). Consistent with the CORE readout, the synonymous cysteine and C231A/S mutants had similar protein expression levels and growth phenotypes (Extended Data Fig. 8f–i). In agreement with their deleterious fitness effects, expression levels of the C231Y/D mutants were notably lower. While overexpression of the C231Y mutant did not impact parasite lytic growth, a reduction in plaque area was observed for the C231D mutant (Extended Data Fig. 8g–i). Considering the relatively low predicted impact of C231D on protein folding (Extended Data Fig. 8a), this suggests that the effect of this substitution upon protein levels is not solely due to changes in stability.

## Covalent inhibition of malaria parasite protein translation

Malaria remains a cause of substantial mortality and morbidity worldwide, and owing to existing and emergent drug resistance there is a constant demand for new therapeutic targets and modalities. Our cross-disciplinary pipeline prioritized protein translation for further target-based screening. Translation has a track record as a therapeutic target, including the malaria parasite, *P. falciparum*<sup>39,40</sup>, but has not previously been targeted for covalent inhibition. The majority of fitness-conferring translation-associated cysteines in *T. gondii* were conserved in *P. falciparum*. Interestingly, not all were conserved in humans, indicating the possibility of parasite-specific functions that could be therapeutically targeted (Extended Data Fig. 7a,c). An established assay for in vitro translation (IVT) is not currently available for *T. gondii*. With that in mind, we took advantage of an IVT assay recently established for *P. falciparum*<sup>41</sup> to test the sensitivity of both



*P. falciparum* and human translation to covalent inhibition with the promiscuous cysteine alkylating molecule, IAA. Parasite, but not human, translation, was sensitive to inhibition (Fig. 4a). Interestingly, no inhibitory effect was observed with another cysteine alkylator, *N*-ethylmaleimide, suggesting IAA-mediated translation inhibition was not solely due to indirect modification of non-protein thiols (for example, glutathione). Encouraged by these data, we assembled a focused library of 88 structurally diverse cysteine-reactive fragments bearing acrylamide warheads (Fig. 4b,c and Supplementary Tables 9 and 10). Screening this library against *P. falciparum* and human IVT identified a lead parasite-selective molecule, 1IH07, with a translation half-maximal inhibitory concentration ( $IC_{50}$ ) of  $42 (\pm 9.8) \mu\text{M}$  (Fig. 4d–g and Supplementary Table 11). Our lead fragment also inhibited the growth of *P. falciparum* asexual blood stages in vitro, with a half-maximal effective concentration ( $EC_{50}$ ) of  $49 (\pm 1.0) \mu\text{M}$ . Combined, these data validated cysteine-targeted inhibition of protein translation in the malaria parasite as a new potential therapeutic modality for this biological process, paving the way for optimization and target deconvolution of our lead compound.

## Discussion

In recent years the scope, scale and speed with which chemically reactive amino acids can be profiled has accelerated dramatically; chemoselective probes are now available for cysteine<sup>4</sup>, serine<sup>42</sup>, lysine<sup>5</sup>, methionine<sup>6</sup>, tyrosine<sup>7</sup>, aspartate/glutamate<sup>9,43</sup>, tryptophan<sup>44</sup> and histidine<sup>45</sup>. Supporting this, advances in MS have greatly expanded the number and rate at which individual reactive sites can be profiled<sup>46,47</sup>, and subsequently exploited by electrophilic drug hunters. CORE provides a technology bridge between unbiased proteomic profiling of chemically reactive sites and protein sequence–function relationships, and will be a valuable tool for proteome engineers alongside other methods, including MAGE, CRMAGE and CREATE. CORE provides a simple strategy to interrogate any individual amino acid, directly assessing its contribution to protein function, which we anticipate will prove as useful as alanine scanning in traditional protein structure–function studies. CORE is similar to other saturation mutagenesis approaches, but whereas these methods mutate entire domains to saturation, we focus our substitution analyses on chemically reactive amino acids on protein surfaces. This is highly beneficial when hunting new targetable sites, as electrophile-sensitive residues are, by definition, chemically accessible. Surprisingly, we observed no clear association between the relative reactivity (as measured by isoTOP-ABPP *R* values) of the cysteines interrogated by CORE and their likelihood of being fitness conferring in the parasite (Fig. 3f). Our data indicate that, contrary to expectation from other published work, cysteine electrophile sensitivity does not predict functionality in *T. gondii*. Extended consideration of our data is presented in the supplementary discussion. While *T. gondii* was used for proof-of-concept studies with electrophile-sensitive cysteines, CORE is amino acid agnostic, with one exciting future application being the systematic profiling of all post-translationally modified proteins of a given class, such as sites of phosphorylation or *N*-myristoylation. Many advances in the target identification and validation sphere are currently achieved retrospectively following identification of a suitable ligand. In contrast, global amino-acid chemical reactivity profiling combined with CORE supports prospective strategies in target identification and validation campaigns. Our approach enables the critical concept of prioritization to be used to promote protein targets and targetable biological processes into screening platforms where an identified prospectively druggable site is already proven to contribute to protein function in intact cell systems. As such, combined with chemical reactivity profiling, CORE has the potential to focus drug discovery pipelines on functional sites on identified targets, accelerating the discovery of targets and next-generation small-molecule therapeutics. The translation of our findings to the related malaria parasite *P. falciparum* provides the

first evidence for this potential, with covalent inhibition of apicomplexan parasite translation apparatus being a tantalizing modality for new broad-spectrum anti-microbials.

## Methods

### General

Unless otherwise stated, all reagents were provided by Sigma. All primers/oligonucleotides and synthetic DNA used in this study are listed in Supplementary Tables 12 and 13, respectively. The IA-alkyne probe, Azo-L and Azo-H tags were synthesized as previously described<sup>4,22</sup>.

### Cell culture and parasite isolation

RH strain *T. gondii* tachyzoites were cultured by serial passage on confluent monolayers of human foreskin fibroblasts (HFFs; HFF-1 ATCC SCRC-1041). HFFs were grown at 37 °C and 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) heat-inactivated foetal bovine serum, 100  $\mu\text{g ml}^{-1}$  penicillin/streptomycin and 2 mM L-glutamine. Unless otherwise stated, parasites were collected for assays or transfection via mechanical syringe lysis of heavily infected HFFs through a 25-gauge needle.

Highly synchronized 3D7 strain *P. falciparum* asexual parasites were cultured in RPMI-1640 medium supplemented with 0.5% (w/v) AlbuMAX II (Life Technologies), 50  $\mu\text{g ml}^{-1}$  hypoxanthine, 25  $\mu\text{g l}^{-1}$  gentamycin and 0.3 mg  $\text{ml}^{-1}$  L-glutamine. Parasites were routinely cultured at 37 °C and 5% CO<sub>2</sub>/3% O<sub>2</sub> with 2% haematocrit blood (NHS UK Blood Transfusion Service). Medium was exchanged daily until the culture reached 10–20% parasitaemia with predominantly late trophozoites and early schizonts. Infected red blood cells (RBCs) were isolated by centrifugation (800g, 5 min) and lysed in RBC lysis buffer (45 mM HEPES pH 7.45, 100 mM potassium acetate, 1.5 mM magnesium acetate, 2 mM DTT and 0.075% (w/v) saponin) for 10 min at room temperature. The lysed RBCs were then centrifuged (2,800g and 4 °C, 10 min), and the resulting parasite pellet was suspended in cell lysis buffer (45 mM HEPES pH 7.45, 100 mM potassium acetate, 1.5 mM magnesium acetate and 2 mM DTT). This step was repeated until all RBC debris was removed.

HEK 293F cells were cultured in FreeStyle 293 Expression Medium (Life Technologies) at 37 °C and 5% CO<sub>2</sub>. Cells were collected at a density of  $2 \times 10^6 \text{ ml}^{-1}$  by centrifugation (1,000g for 10 min at 4 °C) and washed once in cell lysis buffer supplemented with 20 U of human placental RNase inhibitor and cComplete EDTA-free Protease Inhibitor Cocktail (Roche) before processing lysates.

All parasite and host cell strains were confirmed negative for the presence of *Mycoplasma* contamination by PCR.

### Plasmid design and construction

To construct pG140::TgHypo-3×HA, a recodonized TgHypo complementary DNA sequence fused to a C-terminal 3×HA tag was synthesized by GeneArt (Life Technologies). This fragment was cloned into the *Bam*HI and *Hind*III sites of a modified version of the parental plasmid p5RT70loxPKillerRedloxPYFP-HX<sup>48</sup>, in which the *TUB8* promoter had been deleted using the Q5 Site-Directed Mutagenesis Kit (NEB) protocol with primers P1/P2. Next, fragments encompassing the TgHypo 5' or 3' untranslated region (UTR) were PCR amplified from genomic DNA of RHdiCreΔku80Δhxgprt parasites using primers P3/P4 and P5/P6, respectively. The 5' UTR fragment was cloned into the *Nar*I site of the intermediate plasmid, followed by the 3' UTR fragment at the *Sac*I site to generate pG140::TgHypo-3×HA.

To construct pSAG1::Cas9-U6::sgTgHypo(×2), Cas9 sgRNA sequences targeting the TgHypo 5' or 3' UTR were first selected using the Eukaryotic Pathogen gRNA Design Tool (EuPaGDT)<sup>49</sup>. Two single gRNA vectors containing either the 5' UTR-targeting gRNA or the 3' UTR-targeting gRNA were then generated using the pSAG1::Cas9-U6::sgUPRT plasmid as a backbone (Addgene #54467) (ref. 50). Here, the parental UPRT-targeting gRNA was replaced with either TgHypo gRNA using the Q5 Site-Directed Mutagenesis Kit

protocol with primers P7/P9 (5' gRNA) and P8/P9 (3' gRNA). Next, a fragment encompassing the 5' gRNA was PCR amplified using primers P10/P11 and Gibson cloned<sup>51</sup> into the other *KpnI* and *XhoI*-digested 3' gRNA plasmid, generating pSAG1::Cas9-U6::sg*TgHypo*( $\times 2$ ).

To construct pUPRT::FLAG-RPL4, a recodonized RPL4 cDNA fragment with an N-terminal FLAG tag was cloned into the *EcoRI* and *PacI* sites of the expression vector pTUB8-roGFP<sup>52</sup>. An expression cassette encompassing both the TUB8 promoter and FLAG-RPL4 cassette (pTUB8-FLAG-RPL4) was PCR amplified from the intermediate expression plasmid using primers P12/P13 and assembled with a UPRT-targeting backbone via Gibson assembly. To generate the pUPRT::FLAG-RPL4 mutant derivatives (pUPRT::FLAG-RPL4<sup>C231A/S/D/Y</sup>), point mutations were introduced into the recodonized RPL4 sequence in pUPRT::FLAG-RPL4 using the Q5 Site-Directed Mutagenesis Kit with primers P14/18 (C231A), P15/P18 (C231S), P16/P18 (C231D) or P17/P18 (C231Y).

All CORE plasmids were assembled by Biopart Assembly Standard for Indempotent Cloning (BASIC)<sup>53</sup>. To construct the pCORE recipient vector, three DNA parts (a Cas9 nuclease, *hxgprt* selectable marker and an mScarlett counterselection cassette) were generated with flanking BASIC Prefix and Suffix sequences. The Cas9 part was generated via PCR amplification of pCas9/Decoy (Addgene #80324) (ref.<sup>25</sup>) using primers P19/P20. The mScarlett part was synthesized by Twist ([www.twistbioscience.com](http://www.twistbioscience.com)). The *hxgprt* part was amplified from pTUB1:YFP-mAID-3HA, DHFR-TS:HXGPR (Addgene #87259) (ref.<sup>54</sup>) using primers P21/P22. Before amplification, two internal *BsaI* sites in the DHFR UTRs of the *hxgprt* cassette were removed using the Q5 Site-Directed Mutagenesis Kit with primers P23/P24 and P25/P26. The resulting DNA parts were cloned into an ampR-p15A backbone in a four-part BASIC reaction, forming pCORE. All BASIC linkers used in the assemblies were synthesized by Biolegio and are listed in Supplementary Table 13.

### Transfections

All transfections were performed by electroporation using an Amaxa 4D-Nucleofector (Lonza) with programme 'F1-115'. Transfections were carried out using freshly collected extracellular tachyzoites in P3 buffer (5 mM KCl, 15 mM MgCl<sub>2</sub>, 120 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.2 and 50 mM D-mannitol).

### Stable parasite line generation

To generate the iKO strain for *TgHypo* (here referred to as RH *TgHypo*<sup>iKO</sup>), 10  $\mu$ g of *ScaI*-linearized pG140::*TgHypo*-3  $\times$  HA was co-transfected with 10  $\mu$ g of pSAG1::Cas9-U6::sg*TgHypo*( $\times 2$ ) into  $5 \times 10^6$  RHdiCre $\Delta ku80\Delta hxgprt$  parasites<sup>55</sup>. Transgenic parasites were selected with 25  $\mu$ g  $\mu$ l<sup>-1</sup> mycophenolic acid and 50  $\mu$ g  $\mu$ l<sup>-1</sup> xanthine 24 h post-transfection, and individual resistant clones were obtained by limiting dilution. Successful 5' and 3' integration of the DNA construct at the endogenous *TgHypo* locus was confirmed by PCR using primer P27/P28 and P29/P30, respectively. Disruption of the endogenous *TgHypo* locus was confirmed using primers P31/P32. Rapamycin-induced excision of the integrated *TgHypo* iKO construct was verified using primers P33/P34.

To generate parasite strains ectopically overexpressing RPL4 and cysteine-231 mutants of (here referred to as RH $\Delta ku80$ <sup>RPL4</sup> and RH  $\Delta ku80$ <sup>RPL4(C231X)</sup>, respectively) at the *UPRT* locus, 20  $\mu$ g of *PciI*-linearized pUPRT::FLAG-RPL4 or pUPRT::FLAG-RPL4<sup>C231A/S/D/Y</sup> was co-transfected with 10  $\mu$ g pSAG1::Cas9-U6::sgUPRT into  $5 \times 10^6$  RH $\Delta ku80\Delta hxgprt$  parasites<sup>56</sup>. Transgenic parasites were selected with 5  $\mu$ M 5-fluorodeoxyuridine, and individual resistant clones were obtained by limiting dilution. Integration at the 5' and 3' sites was validated using primers P35/P36 and P37/P38, respectively. Disruption of the endogenous UPRT was confirmed using primers P39/P40.

### iKO of *TgHypo*

Confluent HFF monolayers in T25 flasks were infected with  $\sim 2\text{--}5 \times 10^6$  parasites for 4 h before treatment with 50 nM rapamycin or an

equivalent volume of vehicle (DMSO) for 4 h. After washout, parasites were grown for at least 24 h before PCR or western blot analysis.

### SDS-PAGE and western blot analysis

Extracellular parasites were lysed RIPA buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.0), 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS and 1 mM EDTA) or mild lysis buffer (1% Triton X-100 and 0.1% SDS) supplemented with cOmplete Protease Inhibitor Cocktail (Roche) for 1 h on ice. Lysates were then centrifuged (21,000g, 30 min at 4 °C), and protein concentration in the supernatant was quantified using the Pierce BCA Protein Assay Kit (Thermo Scientific). Laemmli buffer was added to the lysate to 1 $\times$  concentration (2% SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.002% bromophenol blue and 125 mM Tris-HCl, pH 6.8) and boiled (95 °C, 5 min) before separation by SDS-PAGE on 12% polyacrylamide gels. Thirty micrograms of protein were typically loaded per lane. Proteins were transferred (20 V, 1 min; 23 V, 4 min; 25 V; 2 min) to nitrocellulose membranes using an iBlot 2 Dry Blotting System (Invitrogen). Membranes were briefly washed in PBS-T (0.1% Tween-20/PBS), blocked (5% skimmed milk/PBS-T, 1 h) and incubated with primary antibodies (1% BSA/PBS-T, overnight at 4 °C) at the following dilutions: mouse  $\alpha$ -SAG1 (1:1,000, Invitrogen, D61S), rat  $\alpha$ -HA (1:2,000, Roche, 3F10), rabbit  $\alpha$ -*T. gondii* (1:2,000, Invitrogen, PA1-7252) and rabbit  $\alpha$ -FLAG (1:2,000, Sigma, F7425). Following washing (PBS-T, 3 $\times$ ), membranes were incubated with HRP-conjugated secondary antibodies (1:5,000-10,000, Invitrogen) in 1% BSA/PBS-T for 1 h at room temperature. Protein bands were developed using the ECL Western Blotting Detection Reagent (GE Healthcare), and chemiluminescence was visualized using a ChemiDoc MP Imaging System (Bio-Rad). Relative band intensity was routinely quantified using ImageJ software. For analysis of RPL4 expression in RH $\Delta ku80$ <sup>RPL4</sup> and RH  $\Delta ku80$ <sup>RPL4(C231X)</sup> parasites, the intensity of the  $\sim 55$  kDa  $\alpha$ -FLAG band (that is, RPL4) was normalized to the  $\sim 50$  kDa band in the  $\alpha$ -*T. gondii* loading control. The mean ( $\pm$  standard deviation (s.d.)) relative band intensity of RPL4 cysteine mutants was then statistically compared with the RH $\Delta ku80$ <sup>RPL4</sup> sample by one-way analysis of variance (ANOVA).

### Immunofluorescence microscopy

Confluent HFF monolayers grown on glass coverslips were seeded with  $\sim 100,000$  parasites. Approximately 24 h post-infection, cells were fixed (4% paraformaldehyde for 15 min at room temperature) permeabilized (0.1% Triton X-100/PBS for 5–10 min) and blocked (3% BSA/PBS for 1 h at room temperature). Staining was performed for 1 h with primary antibodies at the following dilutions: mouse  $\alpha$ -SAG1 (1:1,000, Invitrogen, D61S), rabbit  $\alpha$ -HA (1:1,000, Cell Signaling Technology, C29F4), rabbit  $\alpha$ -*T. gondii* (1:1,000, Invitrogen, PA1-7252) and mouse  $\alpha$ -FLAG M2 (1:1,000, Sigma, F1804). Labelled proteins were stained for 1 h at room temperature using Alexa Fluor 488/594-conjugated goat antibodies (1:2,000, Invitrogen). Nuclei were stained using the intercalating DNA dye DAPI at 5  $\mu$ g ml<sup>-1</sup>. Stained coverslips were mounted onto glass slides using VECTASHIELD Antifade Mounting Media (Vector Labs) and imaged on a Nikon Ti-E inverted microscope using NIS-Elements software. Images were acquired using an ORCA-Flash 4.0 camera and processed using ImageJ software.

### *T. gondii* plaque formation assay

Confluent HFF monolayers grown in six-well plates were seeded with 200–400 parasites, and plaques were left to form undisturbed for 6–7 days. For plaque assays involving conditional KO of *TgHypo*, parasites were first allowed to invade overnight before treating with 50 nM rapamycin or DMSO for 4 h and replacing the drug medium with standard culture medium. Monolayers were then fixed with ice-cold methanol for 10 min and stained with crystal violet stain (2.3% crystal violet, 0.1% ammonium oxalate and 20% ethanol) for 2 h. Plaques were enumerated manually, and statistical significance in plaque counts



between rapamycin and DMSO-treated samples was tested using two-tailed unpaired Student's *t*-tests with unequal variance. The data are presented as mean ( $\pm$  s.d.) counts.

### ***P. falciparum* growth inhibition assay**

Highly synchronized *P. falciparum* 3D7 cultures were diluted to 2% parasitaemia and 1% haematocrit before culturing in 96-well plates containing the appropriate concentration of drug. Wells containing DMSO and 10  $\mu$ M cycloheximide were used as a negative and positive controls, respectively. Parasites were maintained at 37 °C under a gas mixture of 96% N<sub>2</sub>, 3% O<sub>2</sub> and 1% CO<sub>2</sub> for 72 h before freezing plates overnight at –80 °C. Thawed parasites were lysed in SYBR Green lysis buffer (20 mM Tris, 5 mM EDTA, 0.008% (w/v) saponin and 0.08% (v/v) Triton X-100, pH 7.5, 1:10,000 SYBR Green; Invitrogen) for 1 h at room temperature. Fluorescence was then measured using a Tecan M200 Infinite Pro microplate reader at 490 nm excitation and 520 nm emission. The percentage inhibition of asexual parasite growth was calculated relative to control treatments.

### **Design and optimization of the CORE platform**

The design of the CORE workflow begins with the identification and selection of paired CRISPR guide RNA (gRNA) sequences that target the Cas9 nuclease to sites 5' and 3' of a target cysteine codon. As demonstrated in *Caenorhabditis elegans*<sup>57</sup>, we reasoned that a dual gRNA strategy would provide positive selection towards HDR-mediated integration of mutational templates for our essential gene subset, as the lack of repair of two double-strand breaks (DSBs) in an essential gene should be refractory to growth. To test this hypothesis, the frequency of mutants following mutagenesis of an N-terminal proline codon in surface antigen gene1 (*SAG1*) was compared using single or dual gRNAs in combination with single- or double-stranded strand donor repair templates (Extended Data Fig. 9a). These experiments revealed that dual gRNAs in combination with double-stranded templates provided the highest integration efficiency in the absence of any selectable marker. As anticipated, in the absence of drug selection, the frequency of mutants was low (Extended Data Fig. 9b). The potential negative impact of this upon quantitation of integration events was circumvented through the inclusion of recodoned sequence within the donor template. This allowed for integration-selective priming and therefore generation of PCR amplicons of modified genomic loci for downstream NGS analyses (Fig. 2a and Extended Data Fig. 9c). The protein-centric CRISPR guide design tool, CRISPR-TAPE<sup>58</sup>, was used to simplify and accelerate the gRNA identification and selection process for target cysteines. Accommodating the need for high-throughput multiplexed vector construction, BASIC<sup>53</sup> was adapted to our sequences and used for facile, modular and scalable production of all transfection vectors, with dual gRNA cassettes and Cas9 encoded on the same vector as previously reported (Extended Data Fig. 3a)<sup>25,59</sup>. The RH $\Delta$ ku80 $\Delta$ hxgprt NHEJ-deficient parasite strain was used to further promote HDR<sup>56</sup>.

Donor repair templates were designed to (1) destroy the protospacer adjacent motif (PAM) and/or gRNA seed sequence required for Cas9 targeting and so prevent further modification of the site following integration; (2) provide a recodoned stretch of sequence proximal to the target cysteine for the generation of integration-specific amplicons at mutated sites. Transfection with the dual gRNA vector introduces DSBs 5' and 3' of the target cysteine. The excised locus is subsequently repaired using one of the donor templates, producing a mixed mutant pool, which is sampled shortly after transfection for subsequent genomic DNA extraction ('Pre' sample) (Fig. 2a). For each reactive cysteine candidate, *T. gondii* tachyzoites are co-transfected with a single cysteine-targeting dual gRNA plasmid and all five donor templates for HDR (Fig. 2a). The repair templates encoded for either a WT synonymous replacement of the target cysteine, a stop codon or one of the three amino-acid substitution options.

Following transfection, the mixed population of mutants grow competitively, and are sampled for genomic extraction ('Post' sample) (Fig. 2a). Where the DSB is repaired using the synonymous WT template, parasites are expected to grow normally. In instances where the stop codon template is integrated, the gene coding sequence is disrupted, with parasite growth anticipated to be attenuated equivalent to a KO<sup>26</sup>. After quantitative deep sequencing of integration-specific amplicons encompassing a target cysteine, the frequency of reads for a given mutant in the Post sample ( $f_{\text{Post}}$ ) is normalized to Pre ( $f_{\text{Pre}}$ ) to derive fitness scores (Fs) that reflect the viability of parasites during competitive lytic growth. The Fs for the amino-acid mutants are benchmarked against the synonymous WT and stop codon mutants. This provides a quantitative assessment of the contribution of an individual cysteine to protein function in live cells, using mutant cell fitness as a measurable phenotype and NGS reads as the readout. Multiplexing of CRISPR vector construction with BASIC, 96-well-plate-based transfections and automated NGS sample preparation workflow enables hundreds of targets to be functionally interrogated in parallel.

### **CORE plasmid and template library design and construction**

gRNAs were searched against the *T. gondii* GT1 genome (release 46; [www.toxodb.org](http://www.toxodb.org)) using the 'position-specific' function of CRISPR-TAPE (version 1.0.0) (ref. 58). Briefly, gRNAs binding in close proximity to a target cysteine codon were identified by applying a search distance threshold of  $\pm$ 200 nt. For each codon, two gRNAs binding at sites 5' and 3' of the residue were then selected. Selection criteria was based on the number of potential off-target sequences, %GC content and the ability to introduce synonymous PAM or guide blocking mutations at the target genomic sequence. gRNAs were synthesized by Twist as a fragment containing a U6 promoter and flanking BASIC Prefix and Suffix sequences, and independently cloned into *Bsa*I sites of a kan<sup>R</sup>-pMB1 storage plasmid, pTwist Kan (High Copy). For each target cysteine, the corresponding 5'- and/or 3'-binding gRNA fragment were subcloned into pCORE in a three-part BASIC reaction, replacing the mScarlett counterselection cassette and generating the pCORE-CRISPR plasmid. The sequences of all gRNA fragments are listed in Supplementary Table 13.

Donor templates for mutation of target cysteines were synthesized as 300 bp double-stranded fragments by Twist. Lyophilized templates were re-hydrated and used directly in the transfection. For the *SAG1* experiments, 70 bp single-stranded oligonucleotides were used and hybridized to generate double-stranded templates. For each cysteine codon, five templates were designed to incorporate single unique mutations: a recodoned cysteine codon, alanine, serine, tyrosine or a stop codon. Mutation sites were flanked by regions of synonymous recodoned sequence to (1) enable specific detection of cysteine mutants by PCR, and (2) introduce blocking mutations at the PAM and/or gRNA seed sequence to prevent re-excision of modified genomic loci. The recodoned regions were designed to incorporate the most frequently used codons possible (according to the codon usage bias for *T. gondii* listed on the IDT Codon Optimization Tool; [www.eu.idtdna.com/CodonOpt](http://www.eu.idtdna.com/CodonOpt)), while maintaining a minimum of 30% unique bases in the PCR priming sequence. Recodoning was avoided or minimized at intron–exon junctions to avoid interference with messenger RNA splicing. Homology regions were incorporated on either end of templates to promote genomic integration of mutational templates by HDR. The sequences of all mutational templates are listed in Supplementary Table 13.

### **CORE mutagenesis screens**

Transfections were carried out in 16-well Nucleocuvette strips using the Amaxa 4D-Nucleofector X-Unit (Lonza). For the primary optimized CORE screen, 7  $\mu$ g of pCORE-CRISPR and 0.2  $\mu$ g of each of the five corresponding mutational templates (equivalent to a -1:5 plasmid-to-template molar ratio) were co-transfected into  $1 \times 10^6$

RH $\Delta$ ku80 $\Delta$ hxgprt parasites<sup>56</sup>. For the saturation mutagenesis experiments, 7  $\mu$ g of pCORE-CRISPR and 0.5  $\mu$ g of each of the 21 mutational templates (-1:5 plasmid-to-template molar ratio) were transfected. For the *SAG1* experiments, 6  $\mu$ g of pCORE-CRISPR and 2  $\mu$ g of a single template were transfected (-1:100 plasmid-to-template molar ratio). Transfected parasites were expanded in HFF monolayers grown in 24-well plates and allowed to egress naturally 3 days after infection. Approximately  $2 \times 10^6$  of the egressed parasites were used to infect confluent HFF monolayers in six-well plates, and the remaining parasites ( $\sim 2 \times 10^6$ ) were pelleted and frozen for genomic DNA extraction as the initial 'Pre' mutant population. Parasites were allowed to egress naturally 5 days after infection and similarly collected as the 'Post' mutant population. Parasite genomic DNA from frozen cell pellets was extracted using the DNeasy Blood & Tissue Kit (Qiagen) for downstream NGS library preparation.

### Illumina library preparation, sequencing and data analysis

Genomic DNA libraries were prepared similarly to the 16S Metagenomic Sequencing Library Preparation guide (Illumina). Briefly, for each target cysteine, a -600–800 bp fragment targeting the modified genomic locus was PCR amplified from parasite DNA. For the *SAG1* experiments, the amplicons were designed to encompass the template integration site of both modified and unmodified loci. All primers were designed to include overhanging Illumina adapter sequences and are listed in Supplementary Table 12 (P41–P190). The resulting amplicon was purified using AMPure XP magnetic beads (Beckman Coulter). Dual indices and sequencing adapters were then ligated to the purified products using the Nextera XT Index Kit (Illumina). Indexed amplicons were then purified using AMPure XP beads, and quantified using the Qubit dsDNA HS/BR Assay Kits (Invitrogen), or the QuantiFluor ONE dsDNA System (Promega). Indexed amplicons were pooled at equimolar concentration, and the size and purity of the resulting library was assessed on a TapeStation 2200 with the D1000 ScreenTape System (Agilent). The transfer of reagents used for the purification and indexing of amplicons was performed using acoustic liquid handling (Echo 525, Labcyte). Pooled libraries were sequenced using an Illumina NextSeq 500 75PE Mid Output run with a PhiX spike-in of 10%. For the initial screen, indexed amplicons encompassing all 74 electrophile-sensitive cysteines encoding five substitutions/cysteine were indexed for multiplexing onto a single run, that is, 370 unique substitutions were analysed on a single NGS run. Independent biological replicates of the screen were analysed on independent NGS runs. For the saturation substitution analysis of the 10 translation-associated cysteines, all 210 substitutions (10 cysteines with 21 substitutions/cysteine) were indexed for multiplexing, and all three biological replicates analysed on a single run, that is, 630 substitutions were analysed on a single NGS run. Following acquisition, sequencing data were demultiplexed using CASAVA 2.17 and analysed using the Galaxy web server ([www.usegalaxy.org](http://www.usegalaxy.org)). For each uniquely indexed sample, the sequences were concatenated and separated by each template variant to determine the read counts of the different mutation types. Fs values, representing the change in frequency of each mutant variant, were then calculated by dividing the percent proportion of reads in the Post population sample to the Pre. The differences in Fs values of the non-synonymous mutations were normalized and statistically tested against the recodonized cysteine mutation by one-way ANOVA. Deleterious and gain-of-function mutations were classified by statistically significant normalized Fs of <0.66 and >1.33, respectively. Fitness-conferring cysteines were defined as sites containing at least one deleterious or gain-of-function substitution type.

### Cysteine labelling and click chemistry

Cell pellets of *T. gondii* RH $\Delta$ ku80 $\Delta$ hxgprt parasites were lysed by sonication in PBS (pH 7.4) and soluble fractions separated by centrifugation at 3,500g for 5 min. Protein concentrations were determined using the

DC Protein Assay Kit (Bio-Rad) and a NanoDrop 2000c Spectrophotometer (Thermo Scientific). Proteome samples diluted to 2 mg ml<sup>-1</sup> were treated with 10 or 100  $\mu$ M IA-alkyne (from 1 mM and 10 mM stocks in DMSO, respectively) and incubated for 1 h at room temperature with rotation. The labelled proteins were then subject to click chemistry by addition of 100  $\mu$ M Azo-L or Azo-H, 1 mM TCEP, 100  $\mu$ M TBTA and 1 mM CuSO<sub>4</sub> (final concentrations). Click reactions were incubated for 1 h at room temperature with shaking. The Azo-L/H-labelled protein samples were then precipitated by adding trichloroacetic acid to 10% (v/v) concentration. After overnight storage at -80 °C, precipitated proteins were pelleted by centrifugation (17,000g, 10 min), washed three times with chilled MeOH and resolubilized in 1.2% SDS in PBS by gentle sonication and heating (80 °C, 10 min).

### Enrichment and on-bead digestion

Labelled proteome samples were diluted to 0.2% SDS with PBS. The resulting samples were then added to 100  $\mu$ l of Pierce streptavidin beaded agarose resin (Thermo Scientific) and incubated overnight at 4 °C followed by a further 2 h at room temperature. Protein-bound beads were washed with 1 $\times$  0.2% SDS in PBS, 3 $\times$  PBS and 3 $\times$  H<sub>2</sub>O before resuspending in 6 M urea in PBS + 10 mM DTT and incubating at 65 °C for 15 min. Reduced samples were then alkylated by adding IAA to a final concentration of 20 mM and incubating for 30 min at 37 °C with rotation. Samples were diluted three-fold with PBS and centrifuged (1,400g, 2 min) to pellet the beads. The beads were resuspended in a mixture of 200  $\mu$ l of 2 M urea in PBS, 1 mM CaCl<sub>2</sub> and 2  $\mu$ g trypsin (Promega) and incubated overnight at 37 °C. The beads were separated from the digest by centrifugation and washed three times with PBS and three times with H<sub>2</sub>O. Azo-labelled peptides were then cleaved by adding 50 mM sodium hydrosulfite (Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>) and rotating at room temperature for 1 h. Eluted peptides were then collected from the supernatant, and Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub> cleavage was repeated twice more to fractionate the sample. Between each cleavage, the beads were washed with 2 $\times$  H<sub>2</sub>O and combined with the previous elution. Formic acid was added to the sample to 20% (v/v) concentration before storing at -20 °C until MS analysis.

### LC-MS/MS analysis, peptide identification and quantification

Liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis was performed on an LTQ-Orbitrap Discovery mass spectrometer (Thermo Scientific) coupled to an Agilent 1200 Series HPLC. Azo digests were pressure loaded onto 250  $\mu$ m fused silica desalting columns (Agilent) packed with 4 cm Aqua C18 reverse phase resin (Phenomenex). Peptides were then eluted onto a biphasic column consisting of 100  $\mu$ m fused silica packed with 10 cm C18 and 4 cm Partisphere SCX resin (Whatman) following a five-step multi-dimensional LC/LC-MS/MS protocol (MudPIT). Each step used a salt push (0%, 50%, 80%, 100% and 100%) followed by an elution gradient of 5–100% Buffer B in Buffer A (Buffer A: 95% H<sub>2</sub>O, 5% MeCN and 0.1% formic acid; Buffer B: 20% H<sub>2</sub>O, 80% MeCN and 0.1% formic acid) at a flow rate of 250 nl min<sup>-1</sup>. Eluted peptides were injected into the mass spectrometer by electrospray ionization (spray voltage set at 2.75 kV). For every MS1 survey scan (400–1800 *m/z*), eight data-dependent scans were run for the *n*th most intense ions with dynamic exclusion enabled.

The generated tandem MS data were searched using the SEQUEST algorithm<sup>60</sup> against the *T. gondii* database (GT1 proteome), ToxoDB ([www.toxodb.org](http://www.toxodb.org)). A static modification of +57.02146 on cysteine was specified to account for alkylation with IAA. Variable modifications of +456.2849 and +462.2987 were further assigned on cysteine to account for the probe modification with the isotopically light (Azo-L) and heavy (Azo-H) variant of the IA-alkyne-Azo adduct, respectively. Output files from SEQUEST were filtered using DTASelect 2.0. Quantification of isotopic light:heavy ratios was performed using the CIMAGE quantification package as previously described<sup>4</sup>. Overlapping tryptic peptides containing the same labelled cysteine

(but different charge states or tryptic termini) were grouped and the median reported as the final light:heavy ratio ( $R$ ).  $R$  values were averaged across biological replicates, and peptides with relative s.d. of the  $\geq 50\%$   $R$  value were removed.

### Bioinformatics analysis of electrophile-sensitive cysteine dataset

Functional annotation of electrophile-sensitive cysteine-containing proteins was carried out using BLASTP, Gene Ontology and InterPro searches within Blast2GO 5 PRO software<sup>61</sup>. Consensus protein sequences were BLASTP searched against the non-redundant (nr) NCBI protein database using an  $E$ -value cut-off of  $10^{-6}$ . Gene Ontology terms (molecular function, biological process and subcellular localization) were then mapped from the top 20 hits and merged with annotations derived from the InterPro database ([www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro)). Assignments were further optimized using Annex augmentation. Enrichment of annotations was assessed using Fisher's exact test against the *T. gondii* proteome (strain GT1; UniProt Taxonomy ID 507601) at  $<0.05$  false discovery rate. All hypothetical proteins with no functional annotations were removed from data before performing enrichment analyses.

Assessments of gene essentiality in targets containing electrophile-sensitive cysteines were carried out using published phenotype scores<sup>25</sup>. The distribution of phenotype scores of gene products between datasets was statistically compared by two-tailed  $t$ -tests. To ensure comparison of only relevant tachyzoite genes, all genes upregulated in bradyzoite stage (displaying  $\geq 5$ -fold higher expression in chronic versus acute infection in vivo)<sup>62</sup> were removed before analysis.

For conservation analyses of electrophile-sensitive cysteines, orthologues of the associated protein were identified from orthologue groups classified on OrthoMCL<sup>63</sup>. Conservation of a given residue was assessed following BLASTP alignment of the orthologous protein sequence against the *T. gondii* template sequence. Scores were assigned to each alignment on the basis of the presence or absence of a matched cysteine; a score of 3 was assigned to conserved cysteines, 1 for no conservation and 0 if no protein was identified in the orthologue group for a given species. Conservation scores were determined for each cysteine by summing of the scores across the analysed species.

### Design and physico-chemical assessment of the acrylamide fragment library

One-hundred compounds were initially filtered from a large collection of  $>1,000$  cysteine-reactive fragments using the FragFp diversity algorithm in DataWarrior (version 5.5.0, OpenMolecules). This subset was further filtered to 88 acrylamide electrophiles on the basis of the following physico-chemical properties, derived from their SMILES strings using Molecular Operating Environment (Chemical Computing Group):

- Molecular weight ( $\text{g mol}^{-1}$ ): 150–320
- Number of aromatic rings:  $\leq 2$
- Hydrogen bond acceptors: 2–5
- Hydrogen bond donors:  $\leq 2$
- $\text{clogP}$ :  $-2$  to 3

The principal moment of inertia (PMI) was calculated for each fragment to assess the molecular shape diversity of the library. The SMILES strings of the library were imported into Molecular Operating Environment, and a conformational search was performed using the following parameters: force field, MMFF94x; method, stochastic; rejection limit, 200; iteration limit, 10,000; RMS gradient, 0.005; MM iteration limit, 500; allow amide bond rotation, yes; allow unconstrained double bond rotation, yes; enforce chair conformations, yes; refine with QM, no; RMSD limit, 0.15; energy window, 7; conformation limit, 1. normalized PMI values ( $\text{npr1}, I_1/I_3$ ;  $\text{npr2}, I_2/I_3$ ) were visualized in triangular graphs, where  $x (I_1/I_3)$  and  $y (I_2/I_3)$  values of 0, 1, 0.5, 0.5 and 1, 1 represent a perfect rod, disc and sphere, respectively.

### IVT assay

Pellets of *P. falciparum* 3D7 or HEK 293F cells were suspended in  $1\times$  pellet volume of lysis buffer supplemented with 20 U of human placental RNase inhibitor and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). Resuspended cells were then transferred to a pre-chilled nitrogen cavitation chamber (Parr Instrument Company) and incubated on ice at 1,500 PSI for 60 min. Following release from the chamber, the crude lysate was clarified by differential centrifugation (15 min at 10,000g and 4 °C, followed by 15 min at 30,000g and 4 °C). Protein concentration was determined using a NanoDrop (Thermo Scientific) at 280 nm and adjusted to  $12 \text{ mg ml}^{-1}$  before storage at  $-80$  °C.

Before performing IVT assays, low-bind 384-well plates (Corning) were printed (D300e Digital Dispenser, Tecan) with compounds dissolved in DMSO to be assayed at 0.5% of the total assay volume. Five microlitres of clarified *P. falciparum* or HEK 293F lysate was then added to each well, followed by 4.5  $\mu\text{l}$  L-amino acids (each at 200  $\mu\text{M}$  in 45 mM HEPES pH 7.45, 100 mM potassium acetate, 1.5 mM magnesium acetate, 1 mM DTT, 20 U human placental RNase inhibitor, 15  $\mu\text{M}$  leupeptin, 1.5 mM ATP, 0.15 mM GTP, 40 U  $\text{ml}^{-1}$  creatine phosphokinase and 4 mM creatine phosphate (Thermo Scientific), 2% (w/w) PEG3000, 1 mM spermidine and 0.5 mM folinic acid) and 0.45  $\mu\text{l}$  of purified red click-beetle luciferase (CBG99) mRNA ( $1 \mu\text{g } \mu\text{l}^{-1}$ ). CBG99 mRNA was transcribed from expression plasmids pH-CBG99-H (for use in *P. falciparum* assays) or pT7CFECBG99 (HEK 293F assays) as previously described<sup>41</sup>. Prepared plates were incubated at 32 °C for 1 h 40 min before adding 10  $\mu\text{l}$  of 45 mM HEPES pH 7.45, 1 mM magnesium chloride, 1 mM ATP, 5 mM DTT, 1% (v/v) Triton-X, 10  $\text{mg ml}^{-1}$  BSA,  $1\times$  Reaction Enhancer (Thermo Scientific), 1  $\text{mg ml}^{-1}$  D-luciferin (Thermo Scientific) and 0.5 mM cycloheximide. Luminescence was measured across each well using a Tecan M200 Infinite Pro microplate reader heated to 37 °C. Each fragment screen was run as two independent assays with DMSO and cycloheximide (10  $\mu\text{M}$ ) used as negative and positive controls, respectively. The percent of translation inhibition ( $\%I$ ) for each test compound was calculated from raw luminescence values as in equation (1) (where  $V$  is the measured value,  $\mu$  is the mean value, and  $d$ ,  $p$  and  $n$  represent the test compound, positive control and negative control, respectively).

$$\%I = 100 - \left( \frac{(V_d - \mu_p)}{(\mu_n - \mu_p)} \times 100 \right) \quad (1)$$

The robustness of plate-based screens were determined by  $Z$ -factors ( $Z'$ ) from the positive and negative controls as in equation (2) (where  $\sigma$  is the s.d., and  $0.5 \leq Z'$  reflects acceptable assay quality).

$$Z' = 1 - \left( 3 \times \frac{(\sigma_p + \sigma_n)}{(\mu_p - \mu_n)} \right) \quad (2)$$

### Protein structures, homology modelling and folding analyses

Solved protein structures were downloaded from the RCSB Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)). Homology models were predicted from primary protein sequences using AlphaFold 2.0 (ref. 35) or the Phyre2 web portal<sup>36</sup>; only models constructed with 100% confidence and  $\geq 40\%$  sequence identity (Phyre<sup>2</sup>) or  $\geq 70$  pLDDT (AlphaFold) across  $\geq 70\%$  of the sequence were used for downstream analyses. Structural analyses of the impact of cysteine substitutions on protein folding stability in silico were predicted using FoldX 5.0 (<http://foldxsuite.crg.eu>)<sup>37</sup>. For each cysteine mutant, differences in the free energy of folding ( $\Delta\Delta G$ ) relative to the WT structure was determined using the 'PositionScan' command as in equation (3) (where  $\Delta G_{\text{mutant}}$  and  $\Delta G_{\text{WT}}$  reflect the folding free energy of the mutant and WT protein models, respectively).

$$\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{WT}} \quad (3)$$



### Statistical analysis

Statistical tests were performed using GraphPad Prism 8.0 as described in the individual experimental sections above. *P* value significance thresholds were set at: \*\*\*\**P* < 0.0001, \*\*\**P* < 0.001, \*\**P* < 0.01 and \**P* < 0.05. All statistically significant results are annotated with a line and asterisk(s) in the graphs.

### General software

Schematics were created using Adobe Illustrator (version 22.1) and Inkscape (version 0.92.3). Chemical structures were drawn in ChemDraw Professional (version 18.0). PyMOL (version 2.1.1) was used to generate images of 3D protein structures.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

As far as possible, all raw data supporting the findings in this study are available within the Article and its Supplementary Information files. Data obtained from sequencing have been deposited in the Sequence Read Archive ([www.ncbi.nlm.nih.gov/sra](http://www.ncbi.nlm.nih.gov/sra)) under accession number [PRJNA860585](https://doi.org/10.1038/s41564-022-01249-y). The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PX035658](https://doi.org/10.1038/s41564-022-01249-y). Existing data associated with the gene IDs in this project are available from ToxoDB (<https://toxodb.org/toxo/app>). Additional unprocessed data are available from the corresponding author upon request. Source data are provided with this paper.

### References

- Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
- de Vita, E. 10 years into the resurgence of covalent drugs. *Future Med. Chem.* **13**, 193–210 (2021).
- Benns, H. J., Wincott, C. J., Tate, E. W. & Child, M. A. Activity- and reactivity-based proteomics: recent technological advances and applications in drug discovery. *Curr. Opin. Chem. Biol.* **60**, 20–29 (2020).
- Weerapana, E. et al. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).
- Hacker, S. M. et al. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
- Lin, S. et al. Redox-based reagents for chemoselective methionine bioconjugation. *Science* **355**, 597–602 (2017).
- Hahm, H. S. et al. Global targeting of functional tyrosines using sulfur–triazole exchange chemistry. *Nat. Chem. Biol.* **16**, 150–159 (2020).
- Backus, K. M. et al. Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–574 (2016).
- Bach, K., Beerkens, B. L. H., Zanon, P. R. A. & Hacker, S. M. Light-activatable, 2,5-disubstituted tetrazoles for the proteome-wide profiling of aspartates and glutamates in living bacteria. *ACS Cent. Sci.* **6**, 546–554 (2020).
- Brulet, J. W., Borne, A. L., Yuan, K., Libby, A. H. & Hsu, K. L. Liganding functional tyrosine sites on proteins using sulfur–triazole exchange chemistry. *J. Am. Chem. Soc.* **142**, 8270–8280 (2020).
- Mortenson, D. E. et al. ‘Inverse drug discovery’ strategy to identify proteins that are targeted by latent electrophiles as exemplified by aryl fluorosulfates. *J. Am. Chem. Soc.* **140**, 200–210 (2018).
- Carter, A. J. et al. Target 2035: probing the human proteome. *Drug Discov. Today* **24**, 2111–2115 (2019).
- Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
- Ronda, C., Pedersen, L. E., Sommer, M. O. & Nielsen, A. T. CRIMAGE: CRISPR optimized MAGE recombineering. *Sci. Rep.* **6**, 19452 (2016).
- Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2017).
- Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
- Jakočiūnas, T., Pedersen, L. E., Lis, A. V., Jensen, M. K. & Keasling, J. D. CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metab. Eng.* **48**, 288–296 (2018).
- Choudhury, A. et al. CRISPR/Cas9 recombineering-mediated deep mutational scanning of essential genes in *Escherichia coli*. *Mol. Syst. Biol.* **16**, e9265 (2020).
- Kim, K. & Weiss, L. M. *Toxoplasma gondii*: the model apicomplexan. *Int. J. Parasitol.* **34**, 423–432 (2004).
- Dunay, I. R., Gajurel, K., Dhakal, R., Liesenfeld, O. & Montoya, J. G. Treatment of toxoplasmosis: historical perspective, animal models, and current clinical practice. *Clin. Microbiol. Rev.* **31**, e00057–17 (2018).
- Ben-Harari, R. R., Goodwin, E. & Casoy, J. Adverse event profile of pyrimethamine-based therapy in toxoplasmosis: a systematic review. *Drugs Res. Dev.* **17**, 523 (2017).
- Qian, Y. et al. An isotopically tagged azobenzene-based cleavable linker for quantitative proteomics. *ChemBioChem* **14**, 1410–1414 (2013).
- Martell, J. et al. Global cysteine-reactivity profiling during impaired insulin/IGF-1 signaling in *C. elegans* identifies uncharacterized mediators of longevity. *Cell Chem. Biol.* **23**, 955–966 (2016).
- Vinogradova, E. V. et al. An activity-guided map of electrophile-cysteine interactions in primary human T cells. *Cell* **182**, 1009–1026 e29 (2020).
- Sidik, S. M. et al. A genome-wide CRISPR screen in toxoplasma identifies essential apicomplexan genes. *Cell* **166**, 1423–1435 (2016).
- Billon, P. et al. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell* **67**, 1068–1079 e4 (2017).
- Engin, H. B., Kreisberg, J. F. & Carter, H. Structure-based analysis reveals cancer missense mutations target protein interaction interfaces. *PLoS ONE* **11**, e0152929 (2016).
- Hunter, W. N. Isoprenoid precursor biosynthesis offers potential targets for drug discovery against diseases caused by apicomplexan parasites. *Curr. Top. Med. Chem.* **11**, 2048–2059 (2011).
- Imlay, L. & Odom, A. R. Isoprenoid metabolism in apicomplexan parasites. *Curr. Clin. Microbiol. Rep.* **1**, 37–50 (2014).
- Frenal, K., Dubremetz, J. F., Lebrun, M. & Soldati-Favre, D. Gliding motility powers invasion and egress in Apicomplexa. *Nat. Rev. Microbiol.* **15**, 645–660 (2017).
- Foe, I. T. et al. Global analysis of palmitoylated proteins in *Toxoplasma gondii*. *Cell Host Microbe* **18**, 501–511 (2015).
- Caballero, M. C. et al. Identification of new palmitoylated proteins in *Toxoplasma gondii*. *Biochim. Biophys. Acta* **1864**, 400–408 (2016).
- Rompikuntal, P. K. et al. Blocking palmitoylation of toxoplasma gondii myosin light chain 1 disrupts glideosome composition but has little impact on parasite motility. *mSphere* **6**, e00823–20 (2021).

34. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA* **89**, 10915–10919 (1992).
35. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
36. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
37. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, 382–388 (2005).
38. Huber, F. M. & Hoelz, A. Molecular basis for protection of ribosomal protein L4 from cellular degradation. *Nat. Commun.* **8**, 14354 (2017).
39. Baragaña, B. et al. A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature* **522**, 315–320 (2015).
40. Wong, W. et al. Mefloquine targets the *Plasmodium falciparum* 80S ribosome to inhibit protein synthesis. *Nat. Microbiol.* **2**, 1–9 (2017).
41. Fabio, T. et al. High-throughput screening platform to identify inhibitors of protein synthesis with potential for the treatment of malaria. *Antimicrob. Agents Chemother.* **66**, e00237–22 (2022).
42. Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl Acad. Sci. USA* **96**, 14694–14699 (1999).
43. Ma, N. et al. 2-H-azirine-based reagents for chemoselective bioconjugation at carboxyl residues inside live cells. *J. Am. Chem. Soc.* **142**, 6051–6059 (2020).
44. Seki, Y. et al. Transition metal-free tryptophan-selective bioconjugation of proteins. *J. Am. Chem. Soc.* **138**, 10798–10801 (2016).
45. Jia, S., He, D. & Chang, C. J. Bioinspired thiophosphorodichloridate reagents for chemoselective histidine bioconjugation. *J. Am. Chem. Soc.* **141**, 7294–7301 (2019).
46. Kuljanin, M. et al. Reimagining high-throughput profiling of reactive cysteines for cell-based screening of large electrophile libraries. *Nat. Biotechnol.* **39**, 630–641 (2021).
47. Yan, T. et al. SP3-FAIMS chemoproteomics for high coverage profiling of the human cysteinome. *ChemBioChem* <https://doi.org/10.1002/cbic.202000870> (2021).
48. Andenmatten, N. et al. Conditional genome engineering in *Toxoplasma gondii* uncovers alternative invasion mechanisms. *Nat. Methods* **10**, 125–127 (2013).
49. Peng, D. & Tarleton, R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb. Genomics* **1**, e000033 (2015).
50. Shen, B., Brown, K. M., Lee, T. D. & Sibley, L. D. Efficient gene disruption in diverse strains of *Toxoplasma gondii* using CRISPR/CAS9. *mBio* **5**, e01114–e01114 (2014).
51. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
52. Biddau, M. et al. Two essential thioredoxins mediate apicoplast biogenesis, protein import, and gene expression in *Toxoplasma gondii*. *PLoS Pathog.* **14**, e1006836 (2018).
53. Storch, M. et al. BASIC: a new biopart assembly standard for idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth. Biol.* **4**, 781–787 (2015).
54. Brown, K. M., Long, S. & Sibley, L. D. Plasma membrane association by N-acylation governs PKG function in *Toxoplasma gondii*. *mBio* **8**, e00375–17 (2017).
55. Hunt, A. et al. Differential requirements for cyclase-associated protein (CAP) in actin-dependent processes of *Toxoplasma gondii*. *Elife* **8**, e50598 (2019).
56. Huynh, M. H. & Carruthers, V. B. Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. *Eukaryot. Cell* **8**, 530–539 (2009).
57. Chen, X. et al. Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in *Caenorhabditis elegans*. *Sci. Rep.* **4**, 7581 (2014).
58. Anderson, D. P., Bennis, H. J., Tate, E. W. & Child, M. A. CRISPR-TAPE: protein-centric CRISPR guide design for targeted proteome engineering. *Mol. Syst. Biol.* **16**, e9475 (2020).
59. Sidik, S. M., Hackett, C. G., Tran, F., Westwood, N. J. & Lourido, S. Efficient genome engineering of *Toxoplasma gondii* using CRISPR/Cas9. *PLoS ONE* **9**, e100450 (2014).
60. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **5**, 976–989 (1994).
61. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
62. Pittman, K. J., Aliota, M. T. & Knoll, L. J. Dual transcriptional profiling of mice and *Toxoplasma gondii* during acute and chronic infection. *BMC Genomics* **15**, 1–19 (2014).
63. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
64. Xia, D. et al. The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol.* **9**, 1–18 (2008).

## Acknowledgements

This work was supported by grants BB/M011178/1 from the BBSRC (to H.J.B., E.W.T. and M.A.C.), 202553/Z/16/Z from the Wellcome Trust & Royal Society (to M.A.C.), NIH grant R35GM134964 (to E.W.) and 100993/Z/13/Z from the Wellcome Trust (to J.B.). Studies in Dundee were supported by funding from Medicines for Malaria Venture (RD-09-0043) and also from the Wellcome Trust (203134/Z/16/Z), which supports The Wellcome Centre for Anti-Infectives Research. This funding covers all Dundee participants. We acknowledge and thank I. Andrew and L. Game at the UKRI London Institute of Medical Sciences Genomics Laboratory for providing resources and support that contributed to the research results reported within this paper. We also thank M. Bogyo and M. Treeck for early guidance and support for this work.

## Author contributions

Investigation: H.J.B., M.S., J.A.F., F.R.F., F.T., E.A., C.J.W., R.M., N.W., G.C. and M.A.C. Formal analysis: H.J.B., F.R.F., F.T., E.W. and M.A.C. Visualization: H.J.B. and M.A.C. Conceptualization: M.A.C. Writing—original draft: H.J.B., E.W.T. and M.A.C. Writing—review and editing: H.J.B., M.S., F.R.F., E.A., C.J.W., J.B., G.S.B., E.W., E.W.T. and M.A.C. Supervision: B.B., S.W., J.B., G.S.B., E.W., E.W.T. and M.A.C. Funding acquisition: B.B., S.W., J.B., G.S.B., E.W., E.W.T. and M.A.C.

## Competing interests

The authors declare no conflicts or competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41564-022-01249-y>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41564-022-01249-y>.

**Correspondence and requests for materials** should be addressed to E. W. Tate or M. A. Child.

**Peer review information** *Nature Microbiology* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

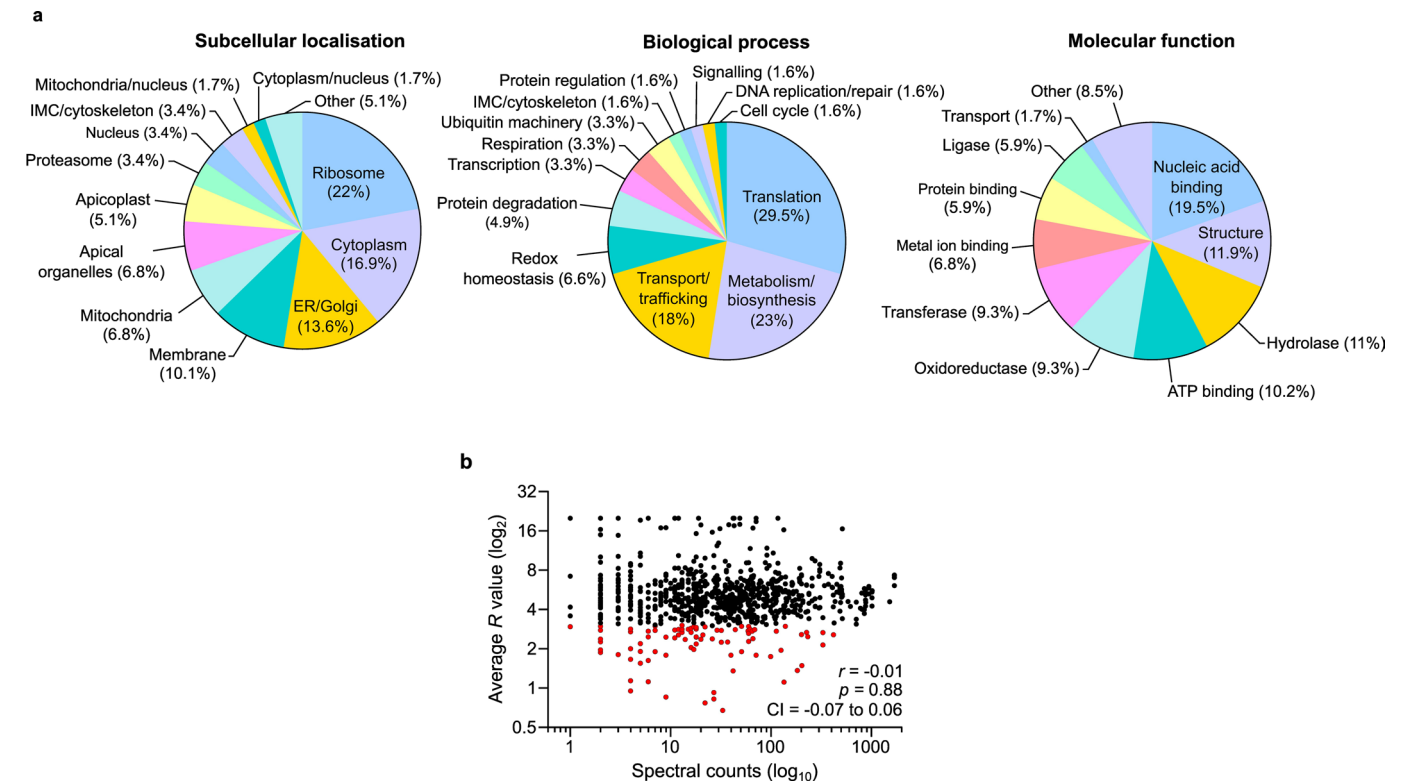
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and

the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

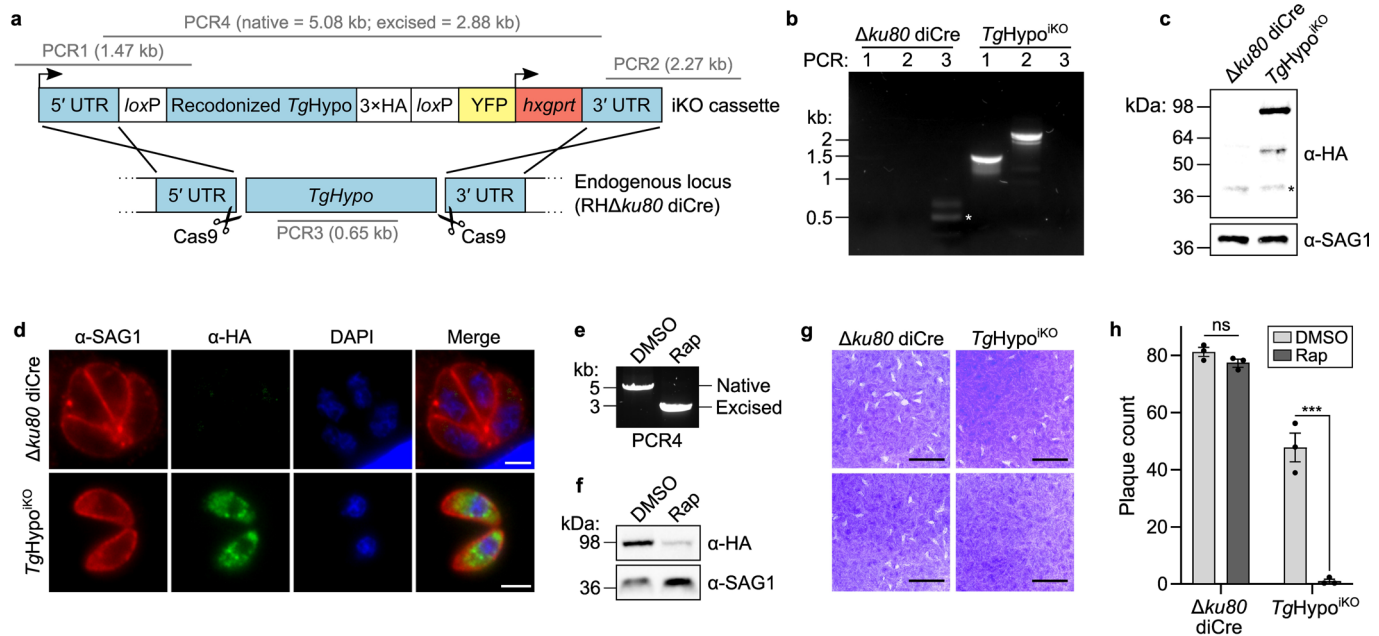
© The Author(s) 2022





**Extended Data Fig. 1 | Enrichment hyperreactive cysteines in ribosomal proteins is independent of protein abundance.** **a**, Proportions of hyperreactive cysteine-containing genes with functional annotations for three gene ontology categories; subcellular localization, biological process and molecular function. Pie charts depict overrepresentation of naturally abundant proteins, such as ribosome components. **b**, Correlation of isoTOP-ABPP  $R$  values against total spectral counts of the associated proteins (a semi-quantitative

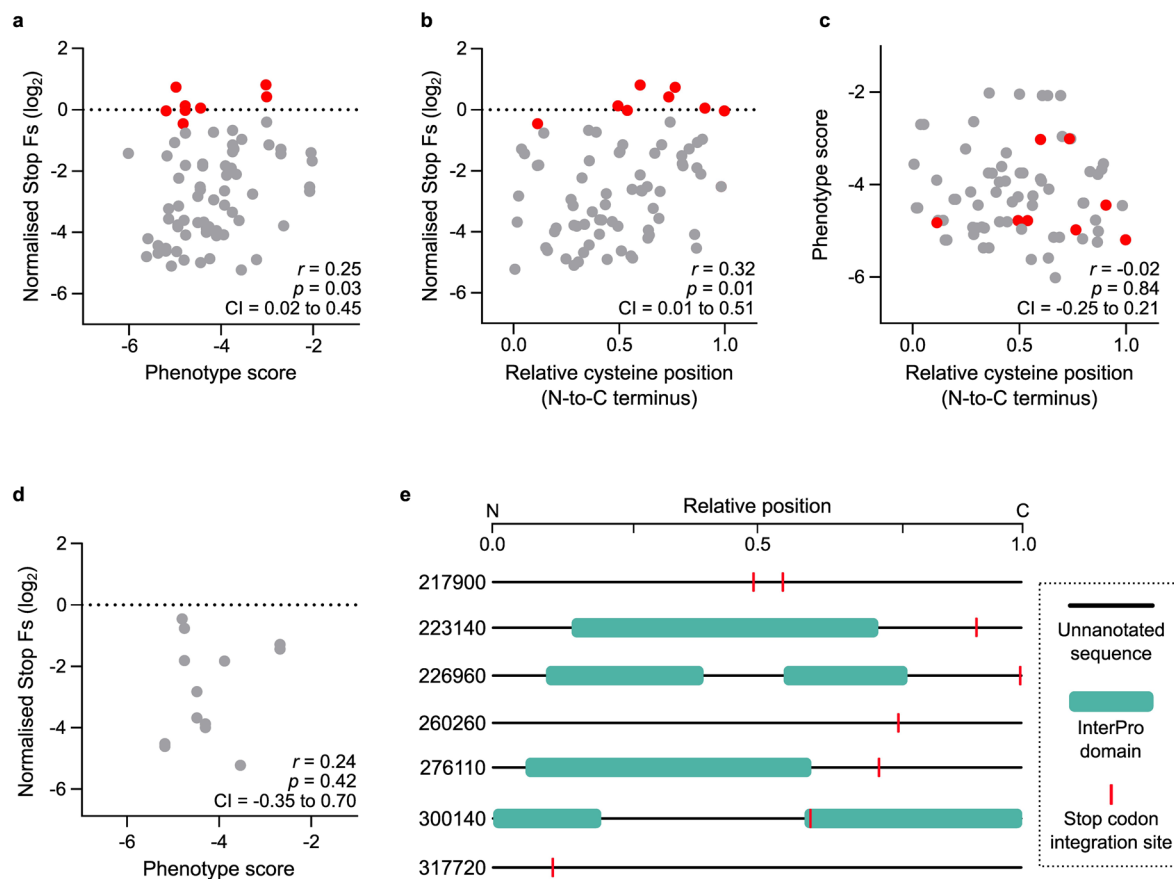
measure of protein abundance). Spectral counts were obtained from a published proteomic dataset for extracellular *T. gondii* parasites<sup>64</sup>. Annotated  $r$  values indicate the degree of correlation between the datasets following a two-tailed Pearson's correlation analysis. Annotated  $p$  and CI values reflect the statistical significance and 95% confidence intervals of the correlation analysis, respectively. Note that proteins with low  $R$  values ( $< 3$ , highlighted in red) span a broad range of spectral counts.

**Extended Data Fig. 2 | TgHypo is indispensable for T. gondii in vitro. a,**

Schematic of the CRISPR-based HDR strategy used for generating a TgHypo inducible knockout line (TgHypo<sup>iKO</sup>) using the diCre system. The predicted sizes of the PCR amplicons used for validating genomic integration and excision of loxP-flanked gene constructs are annotated. **b**, PCR products confirming correct integration of the floxed TgHypo construct at the 5' and 3' UTRs, and loss of the wildtype TgHypo at its endogenous locus. Agarose gel shown is representative of two independent experiments. **c**, Western blot showing expression of the 3 $\times$ HA-tagged TgHypo construct in TgHypo<sup>iKO</sup> parasites using an  $\alpha$ -HA antibody; equal protein loading was verified using an  $\alpha$ -SAG1 antibody. Image shown is representative of two independent experiments. **d**, Immunofluorescence micrographs of TgHypo<sup>iKO</sup> parasites following staining with  $\alpha$ -HA antibodies, showing correct cytosolic localization of the TgHypo-3 $\times$ HA construct. SAG1 and DAPI were used as parasite surface and nuclear markers, respectively. Scale bar

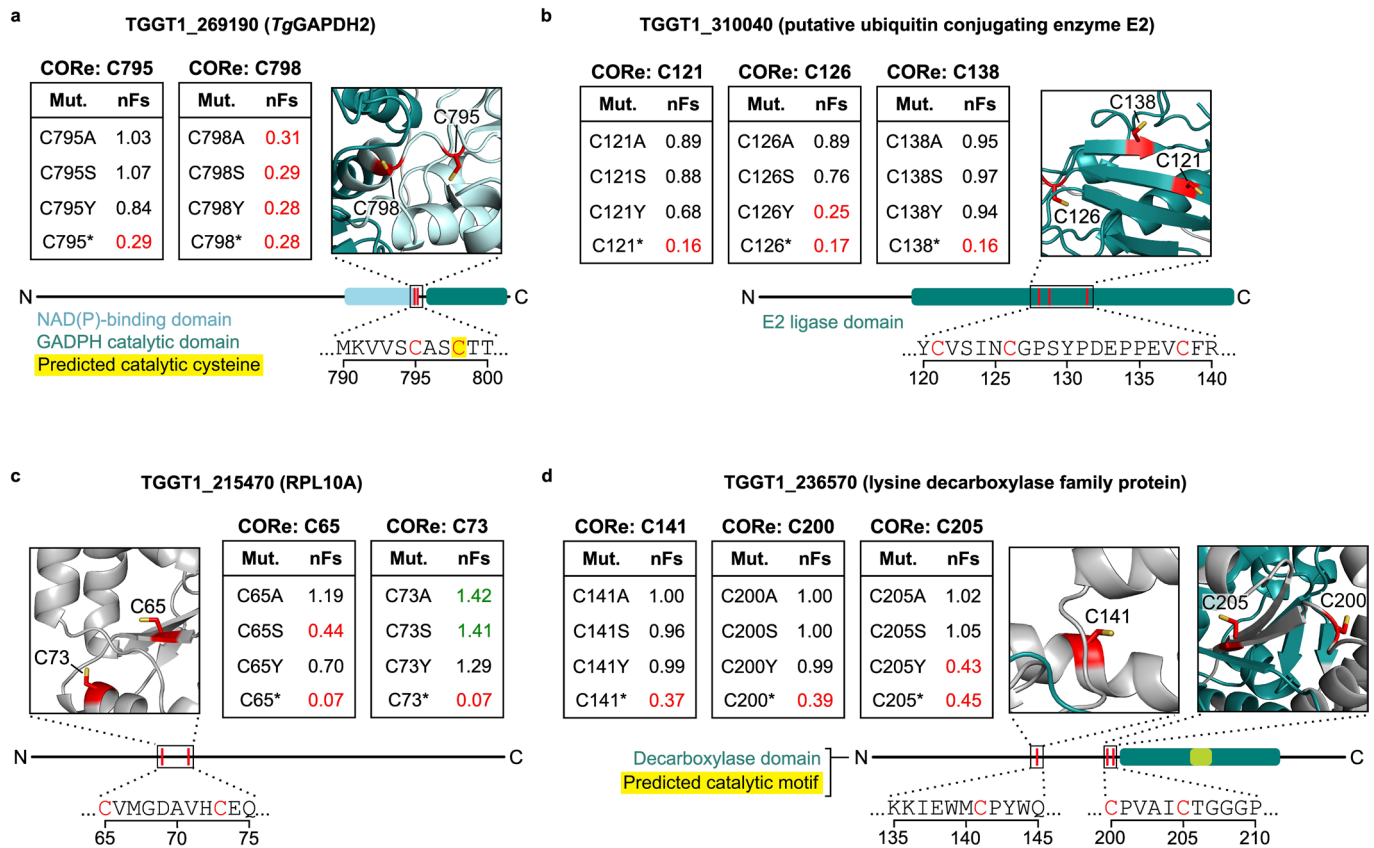
= 3  $\mu$ m. Micrographs are representative of images captured from multiple fields of view across two independent experiments. **e, f**, Analytical PCR (**e**) and western blot (**f**) showing excision and degradation of the floxed TgHypo-3 $\times$ HA construct in TgHypo<sup>iKO</sup> parasites upon rapamycin treatment. Protein loading was assessed using an  $\alpha$ -SAG1 antibody. Image scans are representative of results from two independent experiments. **g**, Representative images of plaques formed on HFF monolayers by the indicated strains in the presence of rapamycin or DMSO. Scale bar = 0.5 cm. **h**, Plaque counts for each strain determined from (g) showing loss of plaquing capacity in TgHypo<sup>iKO</sup> parasites upon rapamycin treatment. Data are presented as mean  $\pm$  SEM across three independent experiments (n = 3), with each point representing mean counts from three technical replicates. Statistical significance was determined by one-way analysis of variance (\*\*\*)  $p < 0.001$ ; ns = no significance,  $p > 0.05$ .  $p$  values:  $\Delta ku80$  diCre DMSO vs Rap, 0.13; TgHypo<sup>iKO</sup> DMSO vs Rap, 0.0007.





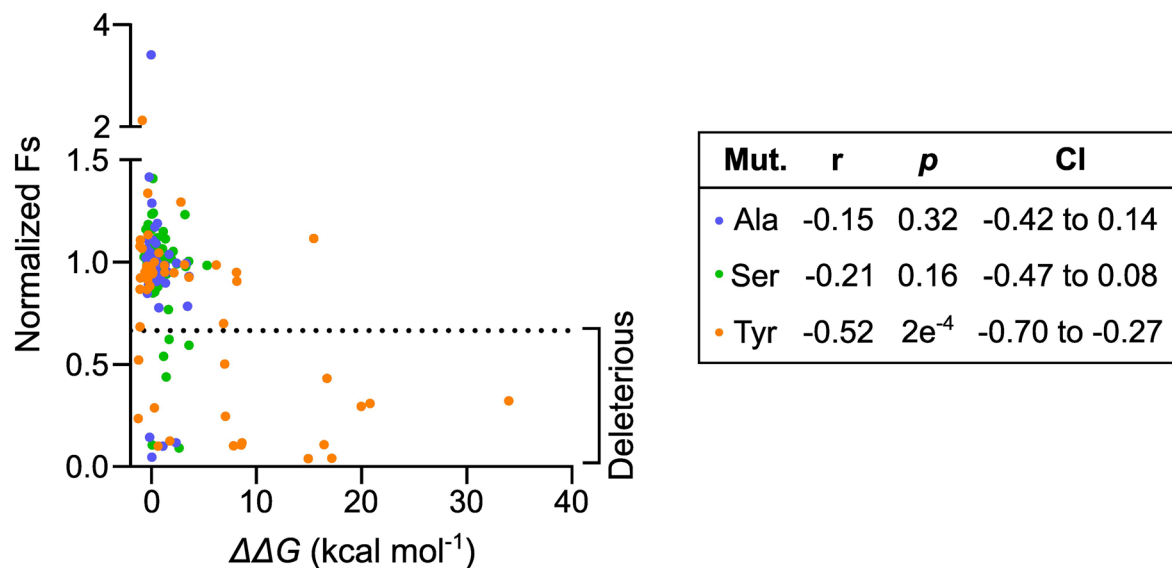
**Extended Data Fig. 4 | CORE stop codon mutant fitness does not correlate with cysteine position or gene phenotype scores.** **a, b.** Correlation of normalised mean stop codon Fs values (**a**) and phenotype scores (**b**) against the relative position of the mutagenized cysteine in the associated protein sequence. While no overall correlation is observed, targets without a statistically significant stop codon phenotype (colored red) generally cluster toward the C terminus. **c, d.** Comparison of statistically significant stop codon mutant phenotypes against published phenotype scores<sup>25</sup> for all interrogated genes (**c**) or interrogated genes with cysteines close to the protein N-terminus (relative

cysteine position  $\leq 0.2$ , **d**); no overall relationship is observed. Annotated  $r$  values indicate the degree of correlation between datasets being compared following a two-tailed Pearson's correlation analysis. Annotated  $p$  and CI values represent the statistical significance and 95% confidence intervals of the correlation analysis, respectively. **e.** Relative positions of functional domains (predicted by InterPro; [www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro)) and mutagenized cysteines across the primary sequences of targets undisrupted by stop codons. The numerical *ToxoDB* gene IDs for each target are listed.



**Extended Data Fig. 5 | Electrophile-sensitive cysteines within individual proteins have diverse mutational tolerances. a-d.** CORE mutagenesis profiles for target cysteines in TGGT1\_269190 (a), TGGT1\_310040 (b), TGGT1\_215470 (c) and TGGT1\_236570 (d). Inset tables show the average normalized Fs (nFs) for each cysteine upon alanine, serine, tyrosine and stop codon substitution; deleterious and gain-of-function mutations are colored in red and green, respectively. The

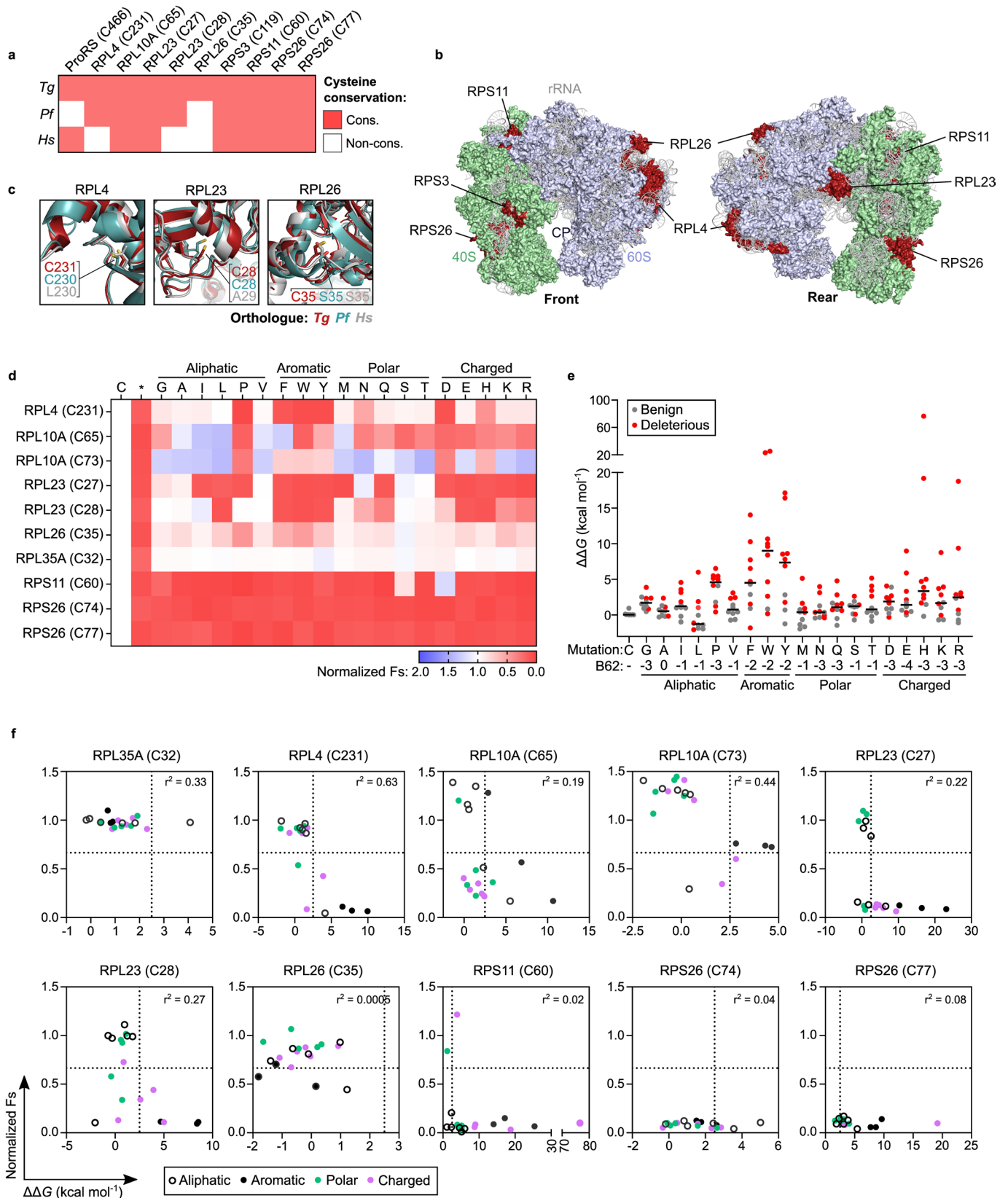
relative positions of the mutagenized cysteines across the associated protein sequence are highlighted red in the domain architecture schematic alongside domains and motifs predicted by InterPro ([www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro)). The sequence context of these sites and their locations within the tertiary protein structure (predicted using AlphaFold 2.0<sup>35</sup>) are highlighted in the inset.



**Extended Data Fig. 6 | Protein folding stability is not substantially impacted by specific substitution type(s).** Protein structures for 35 genes interrogated by CORE were obtained from the RCSB PDB, or from high-confidence models predicted by Phyre2 or AlphaFold 2.0<sup>35,36</sup>. The impact of cysteine substitutions (alanine, serine and tyrosine) on the free energy of folding ( $\Delta\Delta G$ ) was assessed using the PositionScan function of FoldX<sup>37</sup>, where negative and positive  $\Delta\Delta G$  values reflect an increase and decrease in folding stability relative to the wildtype

model, respectively. Scatter plot shows the relationship between  $\Delta\Delta G$  and the normalized Fs value for the associated cysteine mutation. Annotated  $r$  values indicate the degree of correlation between  $\Delta\Delta G$  and Fs for each mutation type following a two-tailed Pearson's correlation analysis. The statistical significance ( $p$ ) and 95% confidence (CI) values are noted. While a subset of tyrosine substitutions have a destabilizing effect, there is no overall correlation.

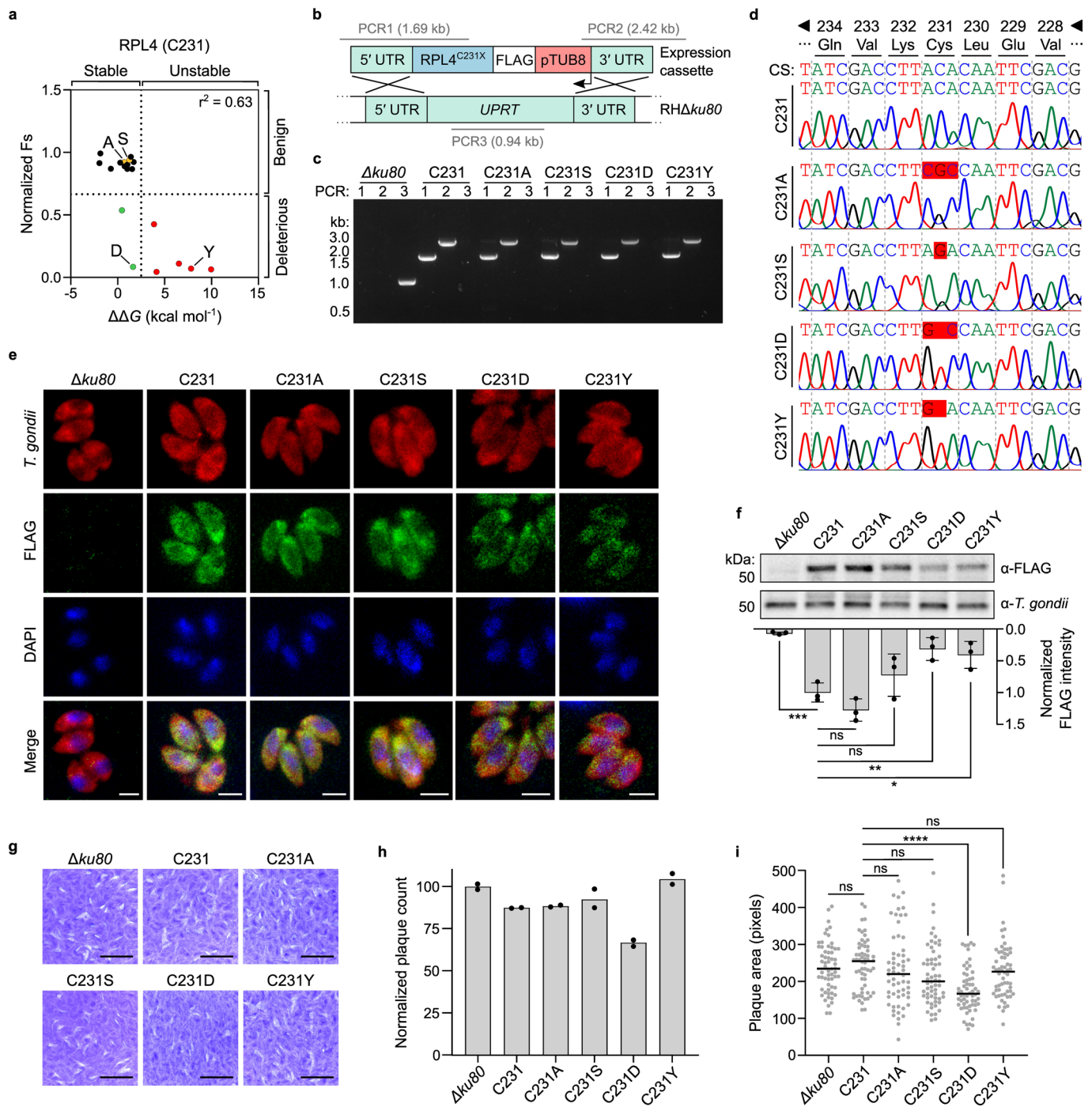




Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Saturation mutagenesis of electrophile-sensitive cysteines in ribosomal proteins in *T. gondii* has diverse impact on parasite fitness and protein folding stability.** **a**, Conservation of fitness-conferring cysteines identified in translation-associated proteins of *T. gondii* (*Tg*) in orthologues of *P. falciparum* (*Pf*) and *H. sapiens* (*Hs*). **b**, Front (left) and rear (right) views of the cytoplasmic *Tg* 80 S ribosome (PDB 5XXU/5XXB). Ribosomal subunits, RNA and proteins containing fitness-conferring cysteines are colored and annotated. **c**, Structural alignment of selected ribosomal proteins with orthologues in the *Pf* (6OKK/3J79) and *Hs* (4UG0) 80 S ribosomes. The cysteines and their positional equivalents in *Pf* and *Hs* are represented in stick form and annotated. **d**, Heatmap displaying normalized Fs values following saturation mutagenesis of 10 electrophile-sensitive cysteines in ribosomal proteins by CORE. Amino acid substitutions are grouped by the biochemical properties

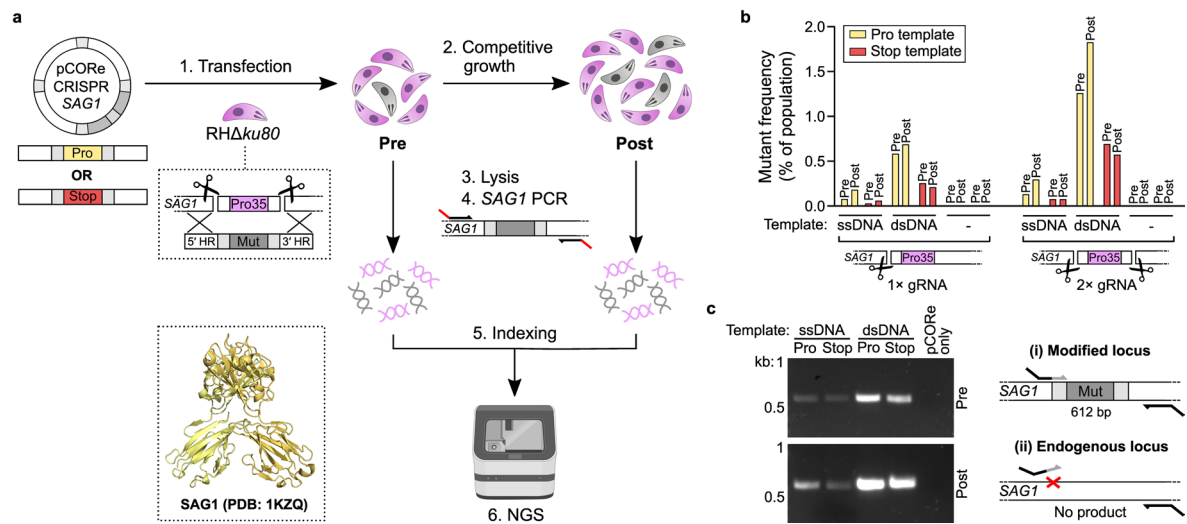
of the associated side chain. Data represents mean values from 3 independent experiments ( $n = 3$ ). **e**, Effects of different amino acid substitutions from **d** on predicted protein folding stability. Changes in the free energy of folding ( $\Delta\Delta G$ ) was assessed for each substitution using FoldX<sup>37</sup>, where negative and positive  $\Delta\Delta G$  values reflect increasing and decreasing stability relative to the wildtype protein, respectively. Data mean  $\Delta\Delta G$  values for each substitution type across 10 distinct targets ( $n = 10$ ). The BLOSUM62 (B62) distances for each amino acid from the wildtype cysteine residue are annotated. **f**, Scatter plots comparing Fs and  $\Delta\Delta G$  values for 19 amino acid substitutions in 10 electrophile-sensitive cysteines. Substitutions are colored according to their side chain chemistries; horizontal and vertical lines reflect thresholds for deleterious ( $0.66 > Fs$ ) and destabilizing ( $2.5 < \text{kcal/mol}$ ) substitutions, respectively.



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Identification of physiologically relevant cysteine substitutions in RPL4.** **a**, Normalized  $F_s$  values for amino acid mutants of RPL4 (C231) plotted their predicted folding free energies ( $\Delta\Delta G$ ), as determined by CORE and FoldX<sup>37</sup>, respectively. **b**, Genetic engineering strategy used to generate *T. gondii* parasite lines expressing FLAG-tagged C231 mutants of RPL4 at the UPRT locus ( $RH\Delta ku80^{RPL4(C231X)}$ ). The sizes of the PCR amplicons used to assess integration of the expression constructs are annotated. **c**, PCR products confirming correct integration of the wildtype and C231A/S/D/Y RPL4 expression constructs at the UPRT 5' and 3' UTRs, and disruption of endogenous UPRT locus. Image shown is representative of results from two independent experiments. **d**, Sequence chromatograms of the integrated RPL4 constructs confirming appropriate mutagenesis of C231. **e**, Immunofluorescence micrographs of  $RH\Delta ku80^{RPL4(C231X)}$  parasites following staining with an  $\alpha$ -FLAG antibody confirming expression and the expected cytosolic localisation of all RPL4 constructs. Parasite surfaces and nuclei were stained with an  $\alpha$ -*T. gondii* antibody and DAPI, respectively. Scale bar = 3  $\mu$ m. Micrographs are representative of images captured across three technical replicates of a single experiment. **f**, Western blot confirming expression of the RPL4 constructs in  $RH\Delta ku80^{RPL4(C231X)}$  parasites using an  $\alpha$ -FLAG antibody; protein loading was

assessed with an  $\alpha$ -*T. gondii* antibody (top). The FLAG staining intensities for each mutant were quantified, normalized and statistically compared to the  $RH\Delta ku80^{RPL4(C231)}$  control (bottom) by one-way analysis of variance (\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ ; ns = no significance,  $p > 0.05$ ). Data represents mean  $\pm$  s.d. intensity from three independent experiments ( $n = 3$ ), revealing notably reduced expression for deleterious C231D and C231Y variants.  $p$  values: C231 vs  $\Delta ku80$ , 0.0305; C231 vs C231A, 0.3557; C231 vs C231S, 0.3625; C231 vs C231D, 0.0055; C231 vs C231Y, 0.0147. **g**, Representative images of plaques formed on HFF monolayers by the indicated strains. Scale bar = 0.5 cm. **h**, Plaque counts (**h**) and area (**i**) for each strain depicted in **g**, showing a reduction in plaquing ability in the RPL4 C231D mutants. Data represents mean counts from two independent experiments ( $n = 2$ ), with each point indicating the average counts from three technical replicates. For plaque area, the size of sixty randomly selected plaques ( $n = 60$ ) are shown, with the mean area indicated by black bars. Differences in plaque area of each mutant against the  $RH\Delta ku80^{RPL4(C231)}$  control line were statistically assessed by one-way analysis of variance (\*\*\*\* $p < 0.0001$ ; ns,  $p > 0.05$ ).  $p$  values: C231 vs  $\Delta ku80$ , 0.9984; C231 vs C231A, 0.9784; C231 vs C231S, 0.9784; C231 vs C231D,  $< 0.0001$ ; C231 vs C231Y, 0.8502.



**Extended Data Fig. 9 | Optimal editing efficiency at *SAG1* is achieved using a dual gRNA strategy with double-stranded template conformation.**

**a**, Workflow used for assessing the integration efficiency of templates for site-directed mutagenesis of *SAG1* (P35) with a synonymous recodonized proline (wildtype) or stop codon (knockout). **b**, Frequency of *SAG1*-modified parasites before ('Pre') and after ('Post') a period of competitive lytic growth.

Templates were provided in either single (ssDNA) or double-stranded (dsDNA) conformation and transfected with single or dual gRNA-containing pCORE CRISPR plasmids. For each mutation type, a maximum integration frequency (1–2%) is achieved following transfection of dual gRNA plasmids with dsDNA templates. **c**, Template-specific PCR from dual gRNA samples showing selective amplification of proline mutant DNA. Data represents a single experiment ( $n = 1$ ).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

NIS-Elements (version 4.60.00) - Visualisation of immunofluorescence micrographs.  
Image Lab (version 6.1) - Visualisation of western blots and agarose gels.  
Tecan i-control (version 1.11) - Acquisition of luminescence signal in plate-based *in vitro* translation assays.  
AlphaFold (version 2.0) - Generation of tertiary protein structures.  
Phyre2 (version 2.0) - Generation of tertiary protein structures.  
FoldX (version 5.0) - Acquisition *in silico* protein folding stability data.  
CRISPR-TAPE (version 1.0) - Design of CRISPR sgRNA sequences.

Data analysis

SEQUEST - Searching tandem mass spectrometry data against the *Toxoplasma gondii* protein database.  
DTASelect (version 2.0) - Assembling and filtering output peptide sequences from SEQUEST.  
CIMAGE (version 1.0) - Quantification of isotopic light:heavy ratios from isoTOP-ABPP experiments.  
BLAST2GO PRO (version 5.0) - Enrichment analysis of functional annotations in the *Toxoplasma gondii* genome.  
CASAVA (version 2.17) - Demultiplexing of Illumina indices for next-generation sequencing data analysis.  
Galaxy (version 21.01.rc1) - Concatenation and splitting of demultiplexed next-generation sequencing data.  
PyMOL (version 2.1.1) - Alignment and visualisation of tertiary protein structures.  
ImageJ (version 1.52q) - Processing of immunofluorescence microscopy images.  
ChemDraw Professional (version 18.0) - Generation of chemical structures.  
Adobe Illustrator (version 22.1) - Drawing schematics and assembly of final figures.  
Inkscape (version 0.92.3) - Drawing schematics and assembly of final figures.



GraphPad Prism (version 8.0) - Producing graphs and statistical analyses.  
 DataWarrior (version 5.5.0) - Filtering chemical compounds by diversity.  
 Molecular Operating Environment (release 2020.09) - Physicochemical assessment of chemical compounds.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

As far as possible all raw data supporting the findings in this study are available within the Article and its Supplementary Information files. Data obtained from sequencing have been deposited in the Sequence Read Archive (SRA; [www.ncbi.nlm.nih.gov/sra](http://www.ncbi.nlm.nih.gov/sra)) under accession number PRJNA860585. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035658. Source data are provided with this article. Existing data associated with the gene IDs in this project is available from ToxODB (<https://toxodb.org/toxo/app>). Additional unprocessed data is available from the corresponding author upon request.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

For experiments with transgenic mutant parasites, sample sizes were determined based on effect size from previous studies examining mutant phenotypes. For the proteomic profiling of cysteine electrophile-sensitivity sample size was determined following the original application of the isoTOP-ABPP methodology (doi: 10.1038/nature09472). The intention of CORE is to remove bias from sample selection. To that end, CORE was performed for all electrophile-sensitive cysteines that were classified as being 'highly-sensitive' to the iodoacetamide probe with a reactivity ratio <3, and satisfied the criteria of having a published phenotype score <-2 (are fitness-conferring during parasite in vitro growth).

Data exclusions

A single replicate was excluded in the CORE data output for one cysteine mutation (C171Y of TGGT1\_215470). The fitness score for this sample (~5.32) was significantly larger than in the two corresponding replicates of the same mutation (~1.17 and ~1.05), and was confirmed by re-sequencing the associated PCR amplicons. Due to the discrepancy of this phenotype between the remaining biological replicates, this data likely reflected a technical issue rather than a genuine physiological response. This data point was therefore considered an outlier and omitted from subsequent statistical analyses.

Replication

Throughout all results in the manuscript were replicated at least twice. Where there was variability between replicates this has been recorded. To ensure biological reproducibility of the phenotypes observed in the CORE mutagenesis screen, transfections and growth assays were performed on distinct days between the three replicates. In addition, the parasites and primary mammalian cells used for different experiments were split and cultured independently. All attempts at replication of the CORE screen were successful.

Randomization

Samples were not randomized but were compared side-by-side with appropriate controls. In CORE experiments, cysteine mutants were generated and assayed on a per-target basis by performing independent transfections and growth assays for each mutagenesis target, avoiding any unanticipated variables that may occur with pooled competitive growth assays using pre-selected target panels.

Blinding

The experiments presented in this study are not subject to experimenter bias and were therefore not blinded.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- | n/a                                 | Included in the study                                     |
|-------------------------------------|---|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Antibodies            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms      |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern     |

### Methods

- | n/a                                 | Included in the study                           |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Antibodies

Antibodies used

Anti-HA High Affinity Rat Monoclonal Antibody (3F10) - Roche, CAT # 11867423001  
 HA-Tag Rabbit mAb (C29F4) - Cell Signalling Technology, CAT # 3724S, LOT 9  
 Monoclonal ANTI-FLAG® M2 antibody produced in mouse - Sigma, CAT # F1804  
 ANTI-FLAG® antibody produced in rabbit - Sigma, CAT # F7425  
 Toxoplasma gondii SAG1 Monoclonal Antibody (D61S) - Invitrogen, CAT # MA5-18268, LOT TB2525582  
 Toxoplasma gondii Rabbit Polyclonal Antibody (PA1-7252) - Invitrogen, CAT # AB\_561769  
 Alexa Fluor™ 488 Goat anti-Mouse IgG (H+L) - Invitrogen, CAT # A11001  
 Alexa Fluor™ 594 Goat anti-Mouse IgG (H+L) - Invitrogen, CAT # A11032, LOT 1985399  
 Alexa Fluor™ 488 Goat anti-Rabbit IgG (H+L) - Invitrogen, CAT # A11034, LOT 1885241  
 Alexa Fluor™ 594 Goat anti-Rabbit IgG (H+L) - Invitrogen, CAT # A11012  
 Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP (62-6520) - Invitrogen, CAT # 31430  
 Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP - Invitrogen, CAT # 31460

## Validation

Anti-HA High Affinity Rat Monoclonal Antibody (3F10) is a commercial antibody used for immunodetection of proteins tagged with the hemagglutinin epitope YPYDVPDYA. In their application note, the manufacturers validated this antibody for labeling against a HA-tagged GST protein. Using western blot, we have demonstrated labeling of a HA-tagged hypothetical protein expressed in *T. gondii* parasites, which is not observed in parental (untransfected) cell lines.

HA-Tag Rabbit mAb (C29F4) is a commercial antibody for immunodetection of exogenously expressed proteins containing the hemagglutinin epitope (YPYDVPDYA). The manufacturers have validated this antibody against a range of HA-tagged proteins across a broad range of applications including western blot, immunohistochemistry and flow cytometry. Using immunofluorescence microscopy, we have demonstrated labeling of HA-tagged hypothetical protein expressed in *T. gondii* parasites, which is not observed in parental (untransfected) cell lines.

Monoclonal anti-FLAG M2 produced in mouse (F1804) is a commercial antibody used for immunodetection of proteins containing the FLAG epitope (DYKDDDDK). Using immunofluorescence, we have demonstrated labelling of FLAG-tagged RPL4 constructs in *T. gondii*, which is not seen in the parental cell line. This antibody has been previously validated for western blots, immunofluorescence and flow cytometry applications in many peer-review articles and organisms, including *T. gondii* (e.g. doi 10.1038/srep03199 and 10.1371/journal.pbio.1001358).

ANTI-FLAG® antibody produced in rabbit (F7425) is a polyclonal antibody used for immunodetection of proteins tagged with the FLAG epitope (DYKDDDDK). The manufacturers have validated this antibody against a range of FLAG-tagged proteins and applications including immunoprecipitation of FLAG-BAP fusion protein from *E. coli* crude lysate, indirect immunofluorescence in HEK293T cells expressing FLAG-JNK, and western blot in *E. coli* periplasmic extract. This antibody has also previously been validated in *T. gondii*, with peer-reviewed publications reporting its use (e.g. doi 10.1371/journal.ppat.1001232).

*Toxoplasma gondii* SAG1 Monoclonal Antibody (D61S) is a commercially available antibody routinely used as a surface marker stain for *T. gondii* tachyzoites and has been validated in several peer-review articles (e.g. doi 10.1128/mBio.01678-17 and 10.3389/fcimb.2017.00259).

*Toxoplasma gondii* Polyclonal Antibody (PA1-7252) is produced in rabbit for specific immunodetection of RH stain *Toxoplasma gondii* parasites. It is routinely raised and used as a control antibody for parasite surface marker staining in immunofluorescence applications or loading controls in western blot (e.g. doi 10.7554/eLife.36491).

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HFF-1 ATCC® SCRC-1041™, HEK293F (Thermo Scientific CAT R79007), <i>Toxoplasma gondii</i> and <i>Plasmodium falciparum</i> strains (various)
Authentication	ATCC certified (HFF), Thermo Scientific certified (HEK293F), in-house tested ( <i>Toxoplasma</i> and <i>Plasmodium</i> )
Mycoplasma contamination	Confirmed free via monthly PCR testing in-house
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified lines were used.