



Cite this: *RSC Chem. Biol.*, 2024,
5, 372

Engineering cell-free systems by chemoproteomic-assisted phenotypic screening†

Zarina Levitskaya,^a Zheng Ser,^b Hiromi Koh,^b Wang Shi Mei,^b Sharon Chee,^a
Radoslaw Mikolaj Sobota^{*b} and John F. Ghadessy^{ib*}

Phenotypic screening is a valuable tool to both understand and engineer complex biological systems. We demonstrate the functionality of this approach in the development of cell-free protein synthesis (CFPS) technology. Phenotypic screening identified numerous compounds that enhanced protein production in yeast lysate CFPS reactions. Notably, many of these were competitive ATP kinase inhibitors, with the exploitation of their inherent substrate promiscuity redirecting ATP flux towards heterologous protein expression. Chemoproteomic-guided strain engineering partially phenocopied drug effects, with a 30% increase in protein yield observed upon deletion of the ATP-consuming SSA1 component of the HSP70 chaperone. Moreover, drug-mediated metabolic rewiring coupled with template optimization generated the highest protein yields in yeast CFPS to date using a hitherto less efficient, but more cost-effective glucose energy regeneration system. Our approach highlights the utility of target-agnostic phenotypic screening and target identification to deconvolute cell-lysate complexity, adding to the expanding repertoire of strategies for improving CFPS.

Received 5th January 2024,
Accepted 16th February 2024

DOI: 10.1039/d4cb00004h

rsc.li/rsc-chembio

Introduction

Phenotypic discovery approaches utilise compound or protein libraries to screen for desired phenotypes in living cells.^{1–3} As these are typically both mechanism and target agnostic, new biology can be discerned from hits illuminating the cellular pool of “dark biological matter”.^{1,3} Consequently, many hits from phenotypic screens would not have been readily discovered using target-based approaches.^{1,4} Here, we apply phenotypic screening towards engineering of poorly understood biologically complex cell-free protein synthesis (CFPS) systems. CFPS utilises cell-derived lysates supplemented with energy sources, amino acids, buffer components, nucleotides and nucleic acid template(s) encoding protein(s) of interest. Applications range from research-driven functional genomics, exploratory biomolecule and metabolic engineering, directed evolution, biosensing, diagnostics, and bio-circuit development, to large-scale industrial manufacturing of biopharmaceuticals and value-added compounds.^{5–15} Compared to cellular expression, the improved *in vitro* reaction tractability

has driven rapid evolution of CFPS, increasing cost-efficiency and expanding its repertoire of products.¹⁶ These now include difficult-to-synthesize proteins such as antibodies, membrane proteins, proteins incorporating unnatural amino acids, and virus-like particles.^{8,17–25} Nevertheless, efficient complex protein production and high costs of exogenous reaction components such as energy factors, amino acids, tRNAs and nucleotides remain major challenges of CFPS development.^{7,8,16,26} These have been addressed using mainly rational approaches including template and strain engineering, growth condition modulation, development of alternate energy regeneration systems and systematic reaction component optimization.^{27–38} The CFPS system with the greatest productivity to date exploits *E. coli* lysates capable of synthesizing proteins at mg mL^{−1} scale.³⁹ Prokaryotic lysates are not however inherently suitable for the production of post-translationally modified eukaryotic proteins, requiring the introduction of accessory enzymes and exogenous structures such as microsomes for complex protein synthesis.^{6,35,40,41} Whilst eukaryote-derived lysates can overcome these issues, they typically show significantly reduced overall productivity, in part associated with their complex regulatory networks and increased background ATP consumption.^{42–45} A case in point is lysates derived from *S. cerevisiae*, an organism historically known for its wide application in biocatalysis and biomanufacturing.^{45,46} Whilst yeast lysates can outperform other eukaryotic CFPS systems⁴⁷ (particularly rabbit reticulocyte lysates) they still remain amongst the least productive, and there is a need for continuing optimisation.^{8,48}

^a Protein and Peptide Engineering and Research Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Singapore, 138648. E-mail: fghadessy@imcb.a-star.edu.sg

^b Function Proteomics Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Singapore, 138648. E-mail: rmsobota@imcb.a-star.edu.sg

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4cb00004h>

Proteins made by CFPS are readily quantifiable, rendering phenotypic screening an ideal tool to interrogate small molecule libraries for modulators of protein expression. Using this approach, we identified several compounds from a library of FDA-approved drugs that increased heterologous protein expression in *S. cerevisiae* cell lysates. Notably, many of these were ATP-mimetic kinase inhibitors, suggesting that inhibition of endogenous energy consumption pathways channelled more ATP into protein synthesis. To further validate this hypothesis, we used thermal stability proteomics analysis⁴⁹ for the first time in CFPS to identify novel protein targets that potentially compete for ATP. Lysates prepared from a yeast strain deleted for one of these targets, the ATP-consuming SSA1 component of the HSP70 chaperone, were more productive in CFPS.

Results and discussion

Screening of an FDA-approved small molecule library for CFPS agonists

A high-throughput screen was performed using a library of 1443 FDA-approved drugs on yeast CFPS to identify enhancers of heterologous protein production (Fig. 1(A)). Coupled transcription and translation reactions were set up according to optimised protocols.^{48,50} Translation was initiated *via* a cap-independent mechanism by the inclusion of the TMV virus 5'UTR Ω sequence upstream of the nano-luciferase (nLuc) reporter gene, and transcripts were stabilised by the addition of a 90-nt poly-A tail at the 3' end.⁵⁰ Yeast extract fractions were pre-incubated with drugs for 15 min, allowing sufficient time

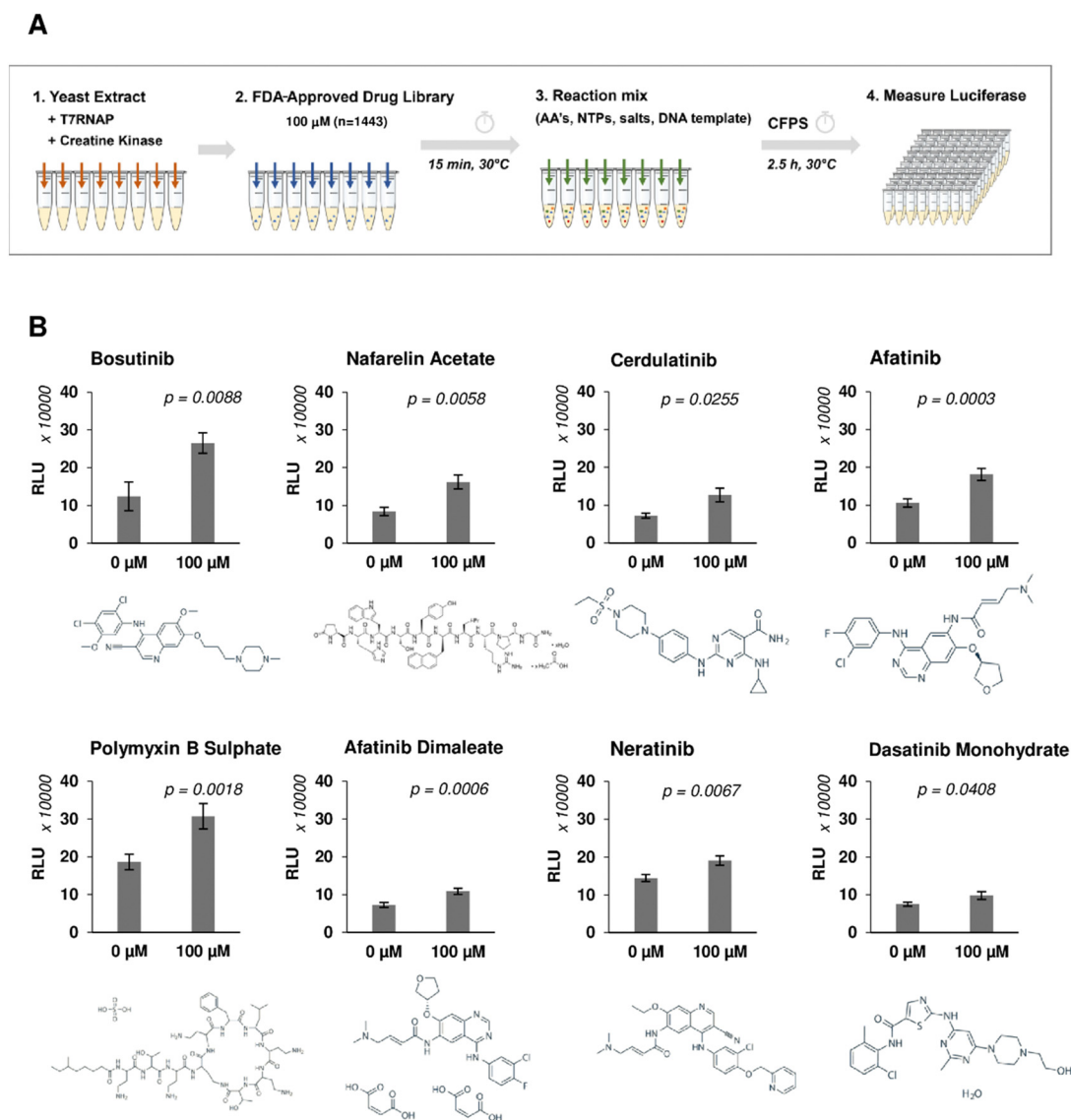


Fig. 1 Yeast lysate CFPS phenotypic screening workflow to identify compounds that increase heterologous protein expression yield. (A) Schematic of the drug screening workflow. Drugs (100 μ M, $n = 1443$) were incubated with yeast extract for 15 minutes prior to addition of nLuc DNA template and other essential components (amino acids, nucleotides, salts) to initiate expression. nLuc produced was measured after 2.5 h incubation. (B) Repeat experiment of selected drug candidates showing fold increase in nLuc production. $n = 3 \pm$ SD. Significance measured using two-tailed Student's *t*-test.



Table 1 List of drug hits increasing yeast extract productivity

Drug	Mode of action	Targets	
Bosutinib	ATP-competitive tyrosine kinase inhibitor (BCR-ABL/SRC)	Tyrosine kinases Serine/threonine kinases ATP-dependent efflux pump	BCR, ABL1, LYN, HCK, SRC CDK2, MAP2K1, MAP2K2, MAP3K2, CAMK2G ABCB1
Cerdulatinib	ATP-competitive tyrosine kinase inhibitor (SYK/JAK)	Tyrosine kinases	SYK, JAK1/2/3, TYK2
Dasatinib monohydrate	ATP-competitive tyrosine kinase inhibitor (SRC family)	Tyrosine kinases	ABL1/2, SRC, EPHA2/5, LCK, YES1, KIT, PDGFRB, FYN, BTK, BCR, CSK, EPHB4, FGR, FRK, LYN ZAK, MAPK14, STAT5B, NR4A3, HSPA8 PPAT ABCB1 ABCG2 GNRHR
Nafarelin acetate	Gonadotropin-releasing hormone (GnRH) agonist	Serine/threonine kinases Transcription factors Heat shock protein Phosphoribosyltransferase ATP-dependent efflux pump ATP-binding cassette G-protein-coupled receptors	
Afatinib	ATP-competitive tyrosine kinase inhibitor (ErbB family)	Tyrosine kinases ATP-dependent efflux pump ATP-binding cassette	EGFR, ERBB2/4 ABCB1 ABCG2
Neratinib	ATP-competitive tyrosine kinase inhibitor (HER2/EGFR)	Tyrosine kinases Serum proteins ATP-dependent efflux pump	EGFR, HER2 ALB, AAG ABCB1
Polymyxin B sulphate	Antibiotic	LPS of Gram-negative bacteria outer membrane	Displaces Ca ²⁺ and Mg ²⁺ from LPS increasing membrane permeability

for binding and onset of drug-mediated responses. Reactions were initiated by the addition of DNA template, amino acids, nucleotides, and energy factors. Drug candidates were identified based on the fold increase of luminescence signal observed from drug-treated reactions against drug-free controls and a 50% cut-off was applied to select hits for secondary analysis (Fig. S1, ESI†). Repeat experiments confirmed 8 out of 110 hits, increasing nLuc yields by 1.30–2.13-fold (Table 1 and Fig. 1(B)). None of the positive hits were shown to directly enhance nLuc activity in a counter-screen to exclude false positives (Fig. S2A, ESI†). Interestingly, six out of eight hits (bosutinib, cerdulatinib, afatinib, afatinib dimaleate, neratinib and dasatinib) are tyrosine kinase inhibitors (TKIs) known to compete for ATP-binding sites of various targets including SRC, ABL, JAK, and EGFR.^{51–55} The remaining two are the gonadotropin-releasing hormone agonist nafarelin acetate and the antibiotic polymyxin B sulphate, which disrupts Gram-negative bacterial membranes.^{56,57}

Numerous drugs which reduced the nLuc yields were also identified in the screen (Fig. S3 and Table S1, ESI†). None were direct inhibitors of nLuc (Fig. S2B, ESI†). These comprised known protein synthesis inhibitors (hygromycin B, puromycin, gentamicin, pentamidine and paramomycin), thus validating screen efficiency. Other inhibitors likely impede the transcriptional step (doxorubicin, proflavine hemisulfate), interfere with DNA synthesis, induce DNA damage (aprotonin, ethacridine lactate and mitoxantrone) or inhibit T7RNAP (peparin).^{58,59} The remaining two inhibitory compounds, calcium levofolinate (chemotherapy adjuvant) and carbenoxolone sodium (antiulcer agent), are of further potential interest.

Additive effects of drugs enhancing CFPS

We next varied the reaction conditions to see if drug activity could be further potentiated. Significant improvements in total yields were observed with lower reaction temperature (22 °C *versus* 30 °C)

and removal of the drug pre-incubation step prior to initiation of transcription/translation (Fig. 2(A)). The pre-incubation period (15 min) reduced the extract activity at both temperatures, likely due to competing background reactions and reduced viability (particularly at the higher incubation temperature) prior to the initiation of luciferase expression by the addition of template and energy factors. The observed inhibition could be rescued with drug treatment (Fig. 2(A)). Following this modified protocol, dose responsive increases (1.28–1.95-fold) in nLuc yield were observed, further validating hits (Fig. 2(B)). We next tested drug combinations for possible additive/synergistic effects. Four of the most potent drugs, bosutinib, cerdulatinib, nafarelin acetate and dasatinib, were added to CFPS reactions alone or in combination. Combining nafarelin with either cerdulatinib or bosutinib exhibited a partial additive effect, further enhancing nLuc production from the original 1.75–1.91-fold to 2.85–3.15-fold (Fig. 2(C)).

Interestingly, the minimum additive effect was observed when combining bosutinib with cerdulatinib and nafarelin acetate with dasatinib, suggesting possible overlapping functions of the drugs within pairs. A similar pattern was observed with *in vitro* expression of the p53 test protein, whereby the drugs resulted in a 1.98–2.40-fold increase in yield when added alone and as high as 3.88 and 5.47-fold when nafarelin acetate was respectively combined with either bosutinib or cerdulatinib (Fig. 2(D)).

Exploring the functional role of bosutinib in improving yeast CFPS

Given its role as an ATP-competitive kinase inhibitor, we questioned whether bosutinib (one of the best hits) increases the CFPS yield by reducing background ATP consumption. More ATP could then be funneled into heterologous protein synthesis, particularly for tRNA aminoacylation. Time-course experiments revealed that the majority of protein synthesis (> 90%) occurred within the first 30 min, reaching completion



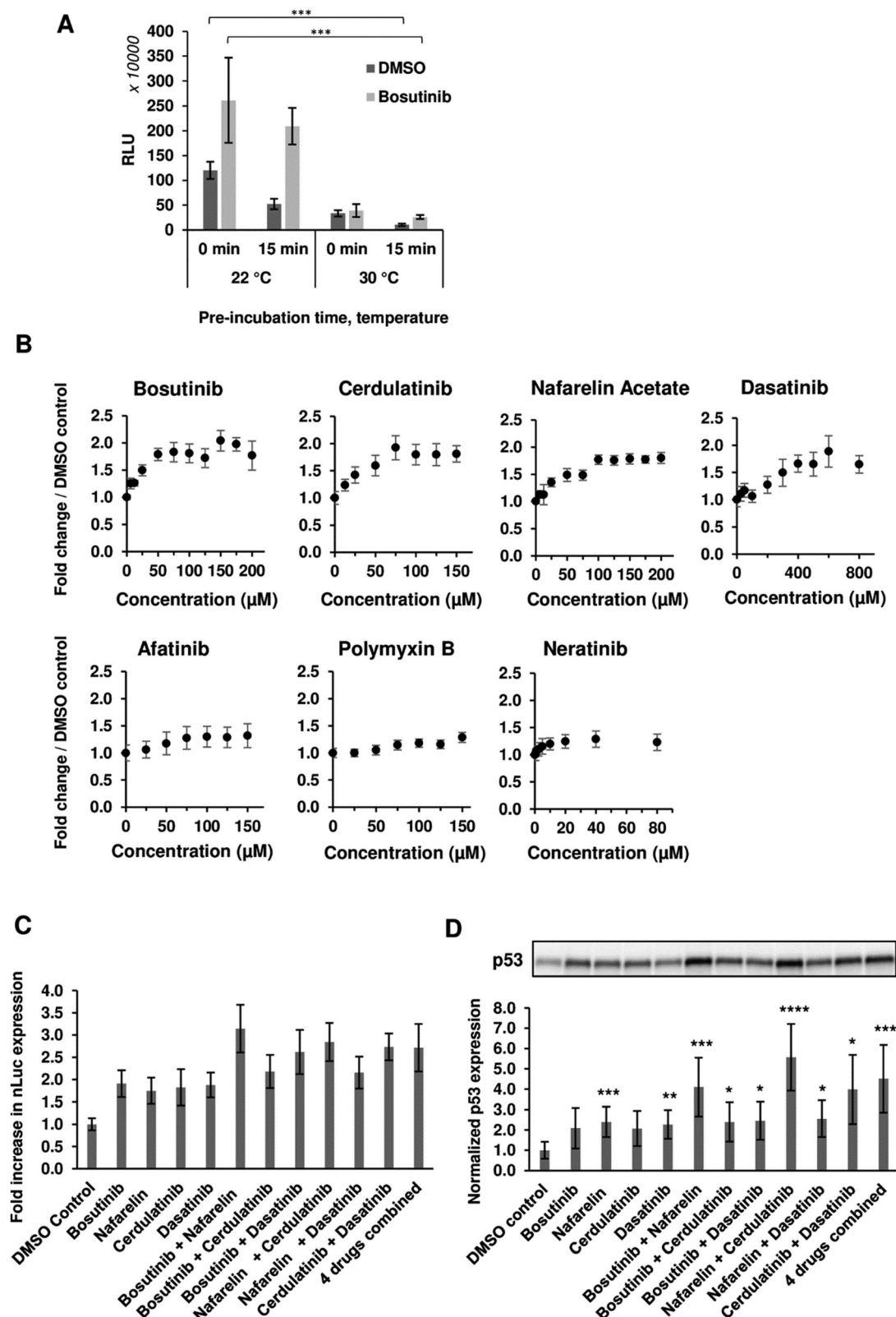


Fig. 2 Optimization of the drug incubation conditions. (A) Effect of CFPS temperature and drug pre-incubation was assessed on the final nLuc yield. A reaction mix containing nLuc template was added to the yeast extract fraction either directly with 100 μM bosutinib/vehicle or after 15 min pre-incubation of the extract. Extract pre-incubation and subsequent CFPS were performed either at 22 °C or 30 °C. $n = 3 \pm \text{SD}$. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$ (two-tailed Student's t -test). (B) nLuc expression in yeast lysate measured over the indicated drug concentrations. $n = 3 \pm \text{SD}$. (C), (D) Bosutinib, cerdulatinib, nafarelin acetate and dasatinib were added to CFPS reactions at their optimum concentrations individually or at the indicated combinations. The drug effect was assessed on the synthesis of nLuc (C) and p53 (D) by means of luminescence and western blot densitometry analysis, respectively. All readings were normalised to DMSO-treated controls. $n = 3 \pm \text{SD}$.



by 45 min (Fig. 3(A)). Bosutinib did not prolong the reaction time, but almost doubled the rate of synthesis in the period between 15–30 min. Completion of the reaction can be explained by rapid depletion of ATP to almost 0.2 mM after 45 min (Fig. 3(B)). No evident differences in ATP consumption were observed between bosutinib treated and untreated reactions. Moreover, the rates of ATP consumption in reactions actively synthesizing nLuc were comparable to controls (no DNA template encoding nLuc added), indicating a minimal effect of nLuc synthesis on overall ATP consumption (Fig. 3(B)). Nevertheless, the addition of bosutinib did result in altered ATP levels both in CFPS and NTC control reactions, but at a much later time. Bosutinib either suppressed or delayed regeneration of ATP that occurred after 2 hours (Fig. 3(B)). Moreover, the observed regeneration of ATP must have been driven by endogenous processes, as it also occurred in CFPS lacking exogenous creatine phosphate and creatine kinase added to generate ATP (Fig. 3(B)). Yeast extracts have the capacity to generate ATP *via* the glycolytic pathway, with the addition of glucose, cAMP and inorganic phosphate facilitating the synthesis of

3.64 $\mu\text{g mL}^{-1}$ of active luciferase.⁶⁰ ATP regeneration *via* the glycolytic pathway occurred with a delay after 1.5 hours, with subsequent protein synthesis initiated at around 2 h.⁶⁰ We thus tested if bosutinib's effect on ATP levels observed in the later time period would also impact protein synthesis in glucose-driven CFPS. Interestingly, bosutinib exhibited a 3.5-fold increase in nLuc yield, attaining a similar productivity level to creatine phosphate/kinase-driven CFPS (Fig. S4A, ESI†).

The absolute yield of active nLuc produced using the creatine phosphate/kinase energy system was 0.77 $\mu\text{g mL}^{-1}$ and increased to 1.65 $\mu\text{g mL}^{-1}$ upon treatment with bosutinib. Despite using a near-identical preparation protocol, the efficiency of our base yeast cell extract was far less than the reported 7 $\mu\text{g mL}^{-1}$.⁴⁸ Drastic drops in efficiency have been reported when extracts are prepared from cells harvested at the stationary phase due to metabolic shifts in response to stress and nutrient depletion.³⁰ Our use of standard Erlenmeyer flasks as opposed to Tunair flasks could have resulted in accumulated stress due to prolonged culture time to reach the desired OD₆₀₀ of 12.³⁰ For subsequent experiments, extracts were therefore prepared from yeast harvested much earlier at OD₆₀₀ of 1.2. These extracts were more efficient, and alongside optimization of template concentration, yielded 5.09 and 6.27 $\mu\text{g mL}^{-1}$ of active nLuc with the creatine phosphate/kinase and glucose energy regeneration systems, respectively (Fig. 3(C) and Fig. S4B, ESI†). Notably, bosutinib lead to a greater increase in yield when the glucose energy regeneration system was used (Fig. 3(C)). The 3-fold increased nLuc yield to 19 $\mu\text{g mL}^{-1}$ is the highest thus far reported using the cost-effective glucose energy alternative.⁶⁰ Interestingly, a time-course experiment with glucose-driven reactions supported the hypothesis of bosutinib's effect on ATP consumption. As before, bosutinib increased the rate of nLuc production (Fig. 3(D)), and reactions containing bosutinib were able to sustain higher levels of ATP in the 4–6 hour time period post initiation (Fig. 3(E)). Moreover, the significantly later onset of nLuc production in glucose-driven reactions (3 hours compared to immediate initiation for creatine phosphate driven reactions) (Fig. 3(A) and (D)) could explain differences observed in yield between the two energy systems when the DNA template concentration was increased. Similar to previous studies, saturation in yield was observed with 10 nM of template using creatine phosphate as an energy source.⁴⁸ However, with glucose-based energy regeneration, an optimal template concentration of 40 nM increased the nLuc yield by 7-fold (Fig. S4B, ESI†). Here, increased template degradation by endogenous nucleases during the considerable 3 hour lag phase is likely mitigated by increasing input DNA levels.

Exploring drug effects on HeLa cell-based CFPS

Six out of the seven positive hits identified in the screen were drugs designed to target human cells. We therefore investigated if the observed outcomes could be translated to the improvement of more complex eukaryotic CFPS systems. HeLa cell-based CFPS reactions were tested with the compounds for improved nLuc production. No significant increase in yield was observed using the original drug screening protocol

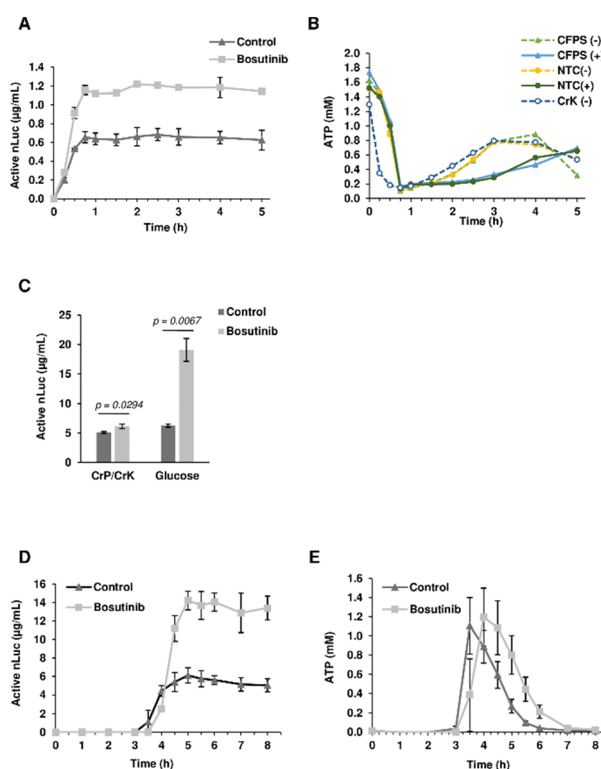


Fig. 3 Characterization of the bosutinib effect on yeast CFPS. (A) nLuc synthesis in CFPS reactions treated with 0 μM and 100 μM bosutinib was monitored at varying time intervals from 0–5 h. $n = 2 \pm \text{SD}$. (B) Corresponding ATP levels were measured from nLuc-synthesizing CFPS reactions, NTC controls (–/+ indicating addition of bosutinib) and CFPS reactions without the creatine phosphate secondary energy regeneration system. (C) Productivity of the yeast extract from cells harvested at OD₆₀₀ = 1.2 was assessed. CFPS reactions were treated with 100 μM bosutinib and set with either creatine phosphate or glucose as a secondary energy source. $n = 3 \pm \text{SD}$. (D, E) nLuc and ATP levels in glucose-driven reactions were monitored from 0–8 h. $n = 2 \pm \text{SD}$.



(15 min incubation with core extract prior to addition of the template and energy components), but following prolonged pre-incubation of the extract fraction with the drug (30 min), the yields of reactions containing bosutinib and dasatinib were 1.64 and 1.97-fold higher than the corresponding controls with DMSO (Fig. 4). However, the overall yields did not improve with the 30 minute pre-incubation time compared to non-treated lysates with 15 minute pre-incubation time. During the prolonged incubation, increased background depletion of endogenous ATP (up to 5 mM) will occur. This reduction in total ATP levels (endogenous + regenerated) is likely responsible for the observed drug-effects, as inhibition of competing ATP consumers will not be as beneficial when ATP is not limiting. A similar drug effect was observed for yeast lysates with and without pre-incubation (Fig. 2(A)), highlighting bosutinib and dasatinib as useful tools to suppress background ATP metabolism during the preparation of eukaryotic CFPS systems and improve efficiency.

Thermal proteomics approach for target identification

Thermal stability-based proteomics methods such as the cellular thermal shift assay (CETSA)^{49,61} or thermal proteome profiling (TPP)^{62–64} coupled to mass spectrometry (MS) can identify proteins whose thermal stability is affected by small molecule action. In particular, isothermal dose response (ITDR) analysis allows the identification of dose-dependent effects on protein thermal stability at elevated temperatures,⁶⁵ enabling CETSA application to a variety of cellular systems for identification of possible targets of small molecules.^{66–68} Yeast extracts were treated with bosutinib at 9 different concentrations (0.01 nM to 1 mM) along with a vehicle control prior to heat treatment at 55 °C. Proteins stabilized by bosutinib binding at this elevated temperature were then quantified by tandem mass tag (TMT) labelling, and dose response curves and pathway analysis were performed (Fig. 5(A)). A total of 545 proteins were fully quantified from the treated yeast lysates across the three replicates. Proteins with high coefficient-of-variation (%CV) greater than 20% at the lowest and highest three drug concentrations were filtered away, leaving 457 proteins which were quantified. Relative protein ratios were computed by normalizing against the lowest concentration (0.01 nM), and mean fold-changes

were obtained by dividing the mean protein abundance in the highest three concentrations against the mean of the lowest three concentrations across the three replicates (Supplementary File 1, ESI†).

Two hundred and fifty one proteins were identified with at least 10% increase in protein abundance in the presence of bosutinib (Fig. 5(B)). Of these, the cytosolic HSP70 chaperone proteins SSA1 and SSB2 were the top 2 proteins stabilised, with up to ~1.6-fold dose-responsive increases in abundance measured (Fig. 5(D) and (F)). Furthermore, known interactors of SSA1 and SSB2 (SKP1 and STI1) are present within the top 20 drug-stabilised proteins, strongly implicating the ATP-consuming HSP70 chaperones as bosutinib targets (Fig. 5(E)). Gene set enrichment analysis of stabilised proteins identified 77 significantly enriched biological pathways. The most enriched pathways included proteasomal ubiquitin-independent proteins (PRE1, PRE3-9, PUP1-3) and other metabolic pathways, such as dicarboxylic acid, glutamine, homoserine, hexose and gluconeogenesis metabolism (Fig. 5(C)). Several components of these pathways utilize ATP (e.g. proteasome complex, PCS60, THR1, PCK1 and PYC2), and non-specific inhibition by bosutinib may contribute to reducing background lysate ATP metabolism. We additionally performed a proteomics time-course experiment at 37 °C to monitor protein stability during a CFPS reaction (Fig. S5 and Supplementary File 1, ESI†). The SSA1 and SSB2 protein levels remained consistent, indicating that the dose-dependent stabilising effects of bosutinib were not confounded by intrinsic thermal instability.

Chemoproteomic-guided strain engineering to improve CFPS

The BY4743-SSA1Δ strain from the homozygous knockout collection was selected to investigate if single gene knockout could reproduce drug-mediated enhancement of yeast CFPS. Extracts from BY4743 WT and SSA1Δ strains were prepared as previously described. As reported earlier, deletion of the SSA1 gene displayed no effect on cell growth⁶⁹ (Fig. S6, ESI†). Glucose-driven CFPS reactions produced 30% more nLuc (9.1 μg mL⁻¹) when SSA1Δ extract was used (Fig. 6(A)). Moreover, the addition of bosutinib resulted in significantly greater improvement of the WT extract (2.6-fold) than the SSA1Δ extract (2.0-fold). A similar pattern was observed with two other TKI drug hits (cerdulatinib and dasatinib), but not with nafarelin acetate GnRH agonist (Fig. 6(B)). We further treated lysates with drug combinations previously shown to improve yield (Fig. 2(C)). Up to 49.7 and 28.1 μg mL⁻¹ of nLuc was made using combined cerdulatinib and nafarelin treatment in parental and SSA1Δ strains, respectively. A very similar pattern was observed when comparing the expression yields of another test protein, the E3 ligase Mdm2, in parental and SSA1Δ extracts. In the absence of drugs, the yields were higher in the SSA1Δ extract (Fig. 6(C)). Drug addition also resulted in less improvement of the yield when added to the SSA1Δ extract (Fig. 6(D)). Notably, the expression yield of Mdm2 could be improved up to 44 times using the nafarelin and cerdulatinib combination (Fig. 6(C)).

Our study demonstrates the utility of phenotypic screening to increase the productivity of CFPS. High-throughput screens

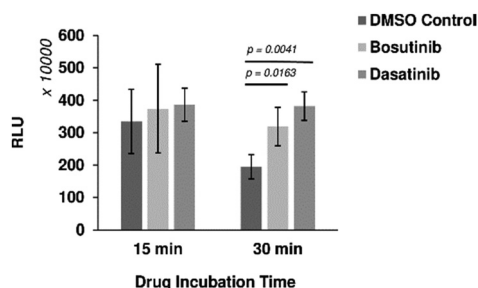


Fig. 4 Drug effects on mammalian cell-based CFPS. Bosutinib and dasatinib monohydrate were pre-incubated with the HeLa cell lysate fraction for 15 or 30 min. Subsequent nLuc production was assessed with respect to DMSO-treated controls. $n \geq 3 \pm \text{SD}$.



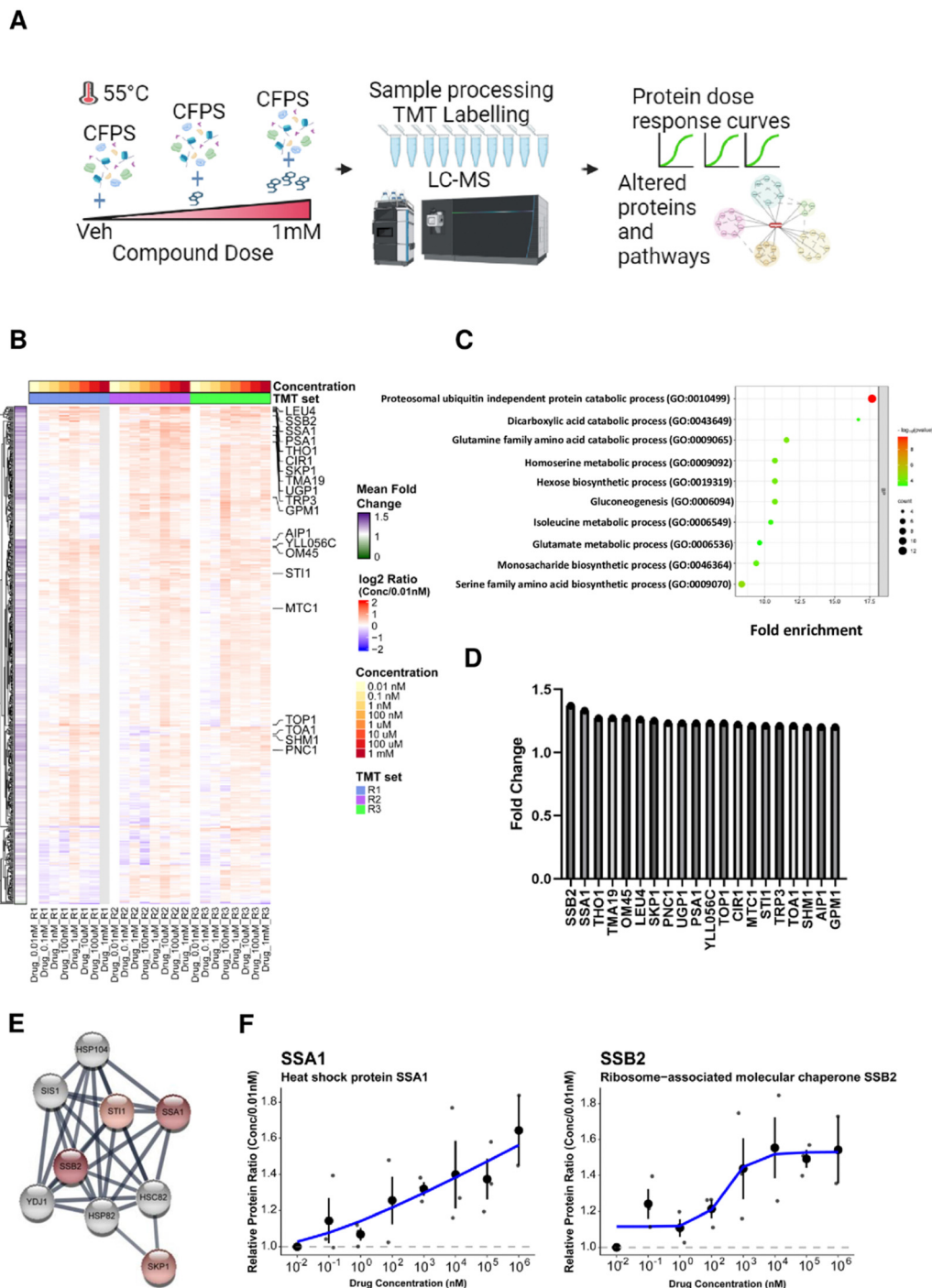


Fig. 5 Effect of bosutinib on proteins and pathways in CFPS identified by thermal proteomics. (A) Workflow for thermal proteomics on the cell free protein system. Yeast extracts were treated with different doses of bosutinib (vehicle, 0.01 nM, 0.1 nM, 1 nM, 100 nM, 1 μ M, 10 μ M, 100 μ M and 1 mM) and subjected to heat stress at 55 °C. Proteins from the different conditions were processed and peptides were labelled using tandem mass tags (TMT). Liquid chromatography coupled with tandem mass-spectrometry (LC-MS/MS) acquisition was performed to profile proteins based on the dose response relationship. Data analysis was carried out to identify proteins and pathways affected by bosutinib treatment. (B) Heat map showing the relative protein changes normalized to the lowest drug concentration (0.01 nM) across 457 proteins identified from three TMT replicate experiments. Twenty proteins with greater than 1.2-fold increase in mean fold change are annotated in the heatmap. (C) Top 10 biological processes based on gene ontology (GO) enrichment analysis using 231 proteins with mean increase $\geq 10\%$. (D) Bar plot of the twenty proteins with greater than 1.2-fold increase in decreasing order of magnitude. SSB2 and SSA1 were identified as the top 2 proteins with mean fold change of 1.37 and 1.33, respectively. (E) Protein interaction network for the top 20 hits highlighted four chaperone proteins (SSB2, SSA1, SKP1, STI1) using the STRING database (proteins are coloured in red based on mean fold change with the color intensity reflecting a higher mean fold-change). (F) Individual loess curves fitted for drug dose against protein quantity measured by mass spectrometry for SSA1 and SSB2, respectively.



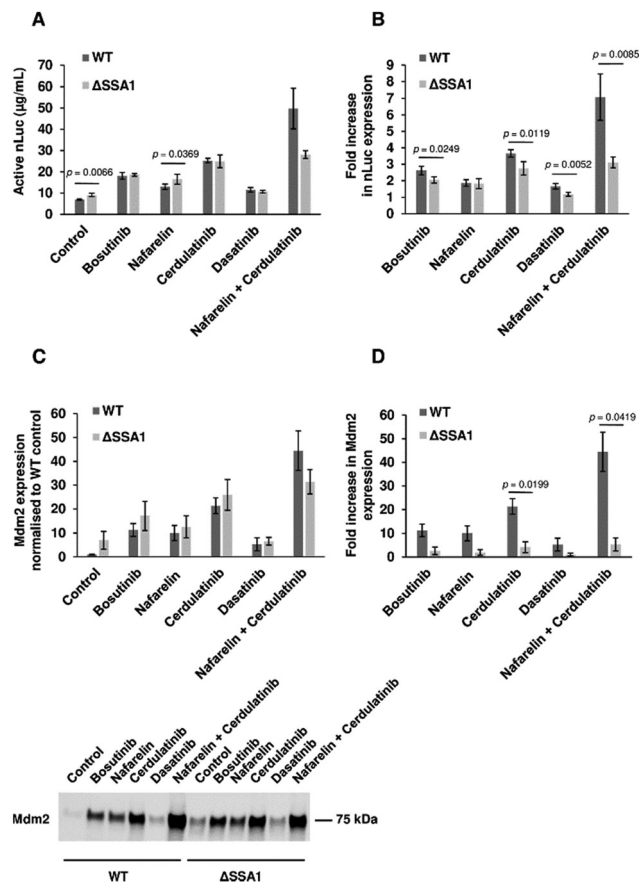


Fig. 6 Chemoproteomic-guided strain engineering. (A) Activity of extracts from BY4743-WT and BY4743- Δ SSA1 strains (without drug treatment and with bosutinib, nafarelin acetate, cerclatutinib and dasatinib) were compared in nLuc expressing CFPS reactions set using glucose as the secondary energy source. $n = 4 \pm$ SD. p values determined by two-tailed Student's t -test. (B) Drug effects on nLuc synthesis in CFPS reactions using WT and Δ SSA1 extracts were compared by determining the fold increase in yield with respect to corresponding drug-free controls. $n = 4 \pm$ SD. p values determined by two-tailed Student's t -test. (C), (D) CFPS reactions with WT and Δ SSA1 extracts were set for Mdm2 protein synthesis. Protein yields were assessed using densitometry analysis of the western blot image. The Mdm2 yield in the Δ SSA1 extract is normalised to that of the WT extract (C top). Representative blot of Mdm2 synthesized by drug treated extracts is presented (C bottom). Fold increase in Mdm2 levels is calculated with respect to the corresponding WT and Δ SSA1 controls (D). $n = 2 \pm$ SD. p values determined by two-tailed Student's t -test.

have been coupled with CFPS technology for protein interaction studies, prototyping, directed evolution and identification of novel therapeutics.^{27,70–72} To date, none have been designed to improve the translation efficiency of the CFPS system itself. The majority of hit compounds identified in the screen were tyrosine kinase inhibitors, suggesting redirection of ATP flux towards the protein translation machinery, particularly ATP-dependent tRNA synthetases. This assumption was verified for the ATP-competitive kinase inhibitor bosutinib, where heterologous protein expression increased with no overall change in extract ATP consumption kinetics. Thermal stability-based proteomic analysis of yeast extracts treated with bosutinib identified proteins involved in metabolic activity, some of

which are ATP-dependent enzymes as possible targets. Whilst proteomics approaches have been used to characterize protein composition and quantities in CFPS for evaluation purposes,^{73–75} they have not previously been employed to delineate competing endogenous pathways and guide strain engineering. The most likely drug target candidates identified are the HSP70 molecular chaperones SSB2 and SSA1. HSP70 chaperones bear an ATPase domain involved in the regulation of substrate binding and are highly expressed.^{76–78} Deletion of the yeast *SSA1* gene resulted in 30% improvement of active nLuc synthesis, partially phenocopying bosutinib treatment of parental lysate. Furthermore, bosutinib showed reduced potentiation of *SSA1* Δ extracts, further validating *SSA1* as a target. Notably, two other TKI hits (cerclatutinib and dasatinib) also showed reduced efficacy in *SSA1* Δ extracts, suggesting *SSA1* to be a promiscuous substrate for ATP analogues. Nevertheless, *SSA1* is not the sole target of identified TKI's given the high homology among the 4 *SSA* isoforms (*SSA1–4*) and 2 *SSB* isoforms (*SSB1–2*) of HSP70 chaperones.⁷⁶ *SSA1* is orthologous to several human heat shock proteins (*HSPA1L*, *HSPA1B*, *HSPA8*), highlighting these as potential 'off-targets' mediating the increased expression yield in HeLa cell lysates treated with bosutinib. Other potential ATP-consuming off-targets of bosutinib include *SRC*, *ERK1/2* and *JNK*.⁷⁹ Mass-spectrometry-based phosphoproteomics is an ideal complementary tool to further explore both of these and delineate further competitors and associated pathways.⁸⁰ Strain engineering to delete further candidate ATP-metabolising targets will likely pose host cell viability issues that can be circumvented by our drug supplementation approach and/or lysate depletion of genomically tagged target proteins.⁸¹

The most efficient cell extracts to date are derived from *E. coli*, typically yielding $0.1–3 \text{ mg mL}^{-1}$ of heterologous proteins.⁸ However, these are largely incapable of complex protein synthesis and post-translational processing (e.g. glycosylation) required for functional eukaryotic proteins. Whilst the use of alternative cell extracts derived from higher eukaryotes can address this deficit, these typically generate lower protein yields. Furthermore, they are associated with increased cost of production due to additional processing steps and overall less robust cell growth conditions.⁸² Microbial eukaryotes such as *S. cerevisiae* and *P. pastoris* are potentially ideal organisms to derive cell-free extracts from. They convey advantages of easy genetic manipulation and propagation.⁸³ However, CFPS systems derived from these organisms are much less productive than *E. coli*-based lysates. *P. pastoris* CFPS can yield $50–300 \text{ µg mL}^{-1}$ of protein, with maximal yields in *S. cerevisiae* trailing at 20 µg mL^{-1} .^{84–86} The latter has been obtained after significant efforts to boost yields and overall cost efficiency using numerous approaches.^{30,50,60,87} Key strategies involved fine-tuning of extract metabolic processes through optimization of growth and harvesting conditions and elimination of proteins with inhibitory effects on CFPS such as *PEP4* protease, *XRN1* nuclease and *STM1* ribosome inactivation factor.^{30,44,50,84} Our approach further increased nLuc yield 7-fold to 50 µg mL^{-1} , the highest reported so far in yeast, and adds to the growing

compendium of CFPS components and parameters that can be tuned to improve performance.

Recent advances in sugar-based energy regeneration systems hold great promise for widespread adoption of CFPS in academia and industry.^{26,33,60,88} These have driven down costs of yeast CFPS, despite the reduced productivity ($3.6 \mu\text{g mL}^{-1}$ nLuc) compared to the widely-adopted and more costly creatine phosphate/creatine kinase system for ATP regeneration ($20 \mu\text{g mL}^{-1}$ nLuc).^{60,84} In our study, final yields were increased up to 7.2-fold to $50 \mu\text{g mL}^{-1}$ for the production of nLuc, exceeding creatine phosphate powered CFPS. Drug treatment contributed to 0.42–39.87% increase in cost of CFPS (Table S2, ESI†). Given the increased productivity of extracts, this translates up to a considerable 9.9-fold increased cost-efficiency from a previous $3.7 \mu\text{g}$ protein per \$ reagent cost to $35.5 \mu\text{g}$ protein per \$ reagent cost using the highly effective nafarelin and cerdulatinib combination (Fig. 6).⁶⁰ Furthermore, drug-induced extract rewiring gave a 44-fold increased yield of the larger Mdm2 (55 kDa) test protein in glucose driven reactions, pointing to even higher cost savings.

In summary, chemoproteomic-assisted phenotypic screening has enabled the enhancement of protein expression yields in yeast and human CFPS. The approach outlined also highlights the utility of CFPS as a surrogate platform to investigate off-target drug activities and toxicity.

Conclusions

Advances in CFPS technology have principally been guided by rational design and systematic optimization of known user-defined components such as nucleic acid template(s) encoding protein(s) of interest and energy regeneration systems. In this study, we applied a novel target-agnostic phenotypic screening approach to identify/repurpose FDA-approved drugs that improve heterologous protein expression in yeast (*S. cerevisiae*) lysates, one of the least productive CFPS platforms. The top hits increased protein expression up to 44-fold, and yielded the highest levels reported for a model enzyme in *S. cerevisiae* lysates. This approach, coupled with proteomic-enabled target deconvolution, uncovered both novel CFPS agonists and targets, highlighting deleterious background pathways that are now more tractable. The methodology developed aids in the deconvolution of complex biology and complements existing approaches to make CFPS technology both cheaper and more accessible.

Materials and methods

Generation of CFPS constructs

pJL1-sfGFP plasmid was purchased from Addgene (#102634), and modified to include 5'UTR and 3'UTR sequences for efficient cap-independent translation in yeast CFPS.⁴⁸ The ribosome binding site downstream of the T7 promoter was replaced with a tobacco mosaic virus 5'UTR fragment of Ω sequence by inverse PCR, retaining an NdeI restriction site.

Similarly using inverse PCR, a stretch of 90-nt poly(A) sequence was inserted immediately after the sfGFP stop codon with a BamHI restriction site in between. The sfGFP coding region was subsequently replaced by In-Fusion cloning (Takara) whereby the vector was digested with restriction enzymes NdeI and BamHI (New England BioLabs Inc.) and an insert was generated by standard PCR. All primers were purchased from integrated DNA technologies (IDT). For primer sequences, refer to Supplementary File 1 (ESI†).

Yeast extract preparation

S30 yeast extract was prepared with reference to established protocols.^{48,50} The S288c yeast strain was a kind gift from A/Prof. Yew Wen Shan from Yong Loo Lin School of Medicine, NUS. BY4743 WT and BY4743 *SSA1Δ* knockout strains were a kind gift from Dr Prakash Arumugam from the Institute of Food and Biotechnology Innovation (SIFBI) in A*STAR. The S288c yeast strain was used for drug screening and subsequent experiments unless stated otherwise. Briefly, yeast was grown in YPD broth (ForMedium™) supplemented with 50 mM Potassium Phosphate (pH 5.5) at 30 °C with shaking (250 rpm) and harvested at OD 10–12, unless specified otherwise. Frozen pellets were resuspended in Lysis Buffer containing 20 mM HEPES-KOH pH 7.4, 100 mM potassium glutamate, 2 mM magnesium glutamate, 2 mM DTT, and 0.5 mM PMSF at 1 mL mg^{-1} of pellet. Cells were lysed by means of high-pressure homogenization using an EmulsiFlex[®]-C3 high pressure homogenizer (25 000 psi, single pass) and dialyzed against the same lysis buffer. The final yeast extract concentration was determined using a Bradford Assay (Bio-Rad). The aliquots of the extract were flash frozen in liquid nitrogen and stored at –80 °C.

Cell free protein synthesis (CFPS)

CFPS reactions were set up with respect to optimised protocols.⁵⁰ Briefly, yeast extract was used at 50% (v/v) in a 15 μL reaction mix. The reactions were set in 1.5 mL Eppendorf tubes at 22 °C for 2.5 h for creatine phosphate-driven CFPS and for 6 h for glucose-driven CFPS. T7RNAP was purified according to Brodiazhenko, and added to the reaction mix at 0.027 mg mL^{-1} .⁴⁴ pJL1 plasmids with target genes, purified with Maxi kit (Thermo Fisher Scientific), were used as CFPS templates. Reactions were supplemented with 0.08 mM of each of 20 amino acids, 1.5 mM each of ATP, UTP, GTP and CTP, 1.7 mM DTT, 2 mM Putrescine and 0.5 mM Spermidine. The Mg^{2+} and K^{+} concentrations were optimized for each extract and template concentration was optimized for creatine phosphate and glucose energy systems. Creatine phosphate-driven CFPS included 0.27 mg mL^{-1} creatine kinase (Roche), 25 mM creatine phosphate (Roche) and 1.7 mM DTT (Fig. S7, ESI†). Glucose-driven CFPS was set following the established protocol and included 25 mM glucose, 0.3 mM cAMP, 10 mM potassium phosphate and 4 mM DTT.⁶⁰

Nano-luciferase assay

Nano-luciferase activity was determined using a Nano-Glo[®] Luciferase Assay (Promega). CFPS reactions were diluted in



Nano-Glo[®] Buffer. Luminescence was read using an EnVision 2104 Multilabel Reader every 20 s for at least 15 min, and the maximum reading was recorded. The absolute nano-luciferase yield was determined by comparing the luminescence reads against the standard curve generated with nano-luciferase recombinant protein purchased from Promega (Nluc-HT Protein) (Fig. S8, ESI[†]).

Drug screening

A library of 1443 FDA-approved inhibitors was used for drug screening (Selleck). Drugs (100 μ M) were pre-incubated with a yeast extract mixture containing T7RNAP and creatine kinase for 15 min at 30 °C. Reactions were set in PCR tubes in 96-well formats with no-drug and no-template controls included in the last column. No-drug controls had a final DMSO concentration (1%) matched to the test groups. Each plate was tested once, and an average of two luciferase readings were obtained. Drug candidates were determined based on the fold change of the luminescence signal of drug-treated reactions with respect to that of untreated controls. A cut-off value of at least 50% increase was used to select candidates for second round testing following the same method. Finalised drug candidates were further tested in a luciferase counter-screen test to eliminate false-positives whereby drugs (100 μ M) were added only after completion of CFPS reactions. The synthesized nano-luciferase was incubated with drugs for 5 min at 30 °C and assayed following the method described earlier.

Western blot assay for p53 and Mdm2

Equal volumes of CFPS reactions were boiled and resolved by SDS-PAGE. CFPS products were blotted with anti-p53 (DO1, 1:1000) or anti-Mdm2 (2A9, 1:1000) antibodies. Both antibodies were a kind gift from Borek Vojtesek. Western blots from at least two independent experiments were performed and one representative image was selected. Densitometry analysis was performed using Image Lab software (Bio-Rad) to determine the fold increase in yields of drug-treated CFPS reactions with respect to drug-free controls.

HeLa CFPS

HeLa cell lysate-based CFPS reactions from 1-Step Human Coupled IVT Kit (ThermoFisher Scientific) were set up according to the manufacturer's protocol with the addition of a drug pre-incubation step. Bosutinib and dasatinib monohydrate at 100 μ M concentrations were added to the HeLa lysate fraction. After 30 min, accessory proteins, reaction mix and pT7-nLuc template (40 ng μ L⁻¹) were added. The reaction was run at 30 °C in a 1.5 mL Eppendorf tube for 6 h.

CFPS time-course experiment

CFPS reactions were set as described earlier. The reactions were aliquoted into multiple individual 1.5 mL Eppendorf tubes in 15 μ L fractions. At given time points (0–5 hours for creatine phosphate CFPS and 0–8 hours for glucose CFPS) the entire reaction was snap frozen in liquid nitrogen and kept at –20 °C at a given time point. Upon thawing, 1 μ L of reaction mix was

sampled out in duplicate for nLuc and ATP analysis. nLuc yield was determined as described earlier. ATP levels were measured using The CellTiter-Glo[®] 2.0 Assay. CFPS samples were diluted 10 times and mixed with an equal volume of 10% TCA solution. Reactions were spun down at 12 000 g for 10 min to remove precipitated proteins. 25 times diluted supernatant was mixed with an equal volume of CellTiter-Glo[®] 2.0 reagent, incubated in the dark for 10 min and the luminescence signal was read. The ATP levels were determined by comparing the luminescence signal against the ATP standard curve.

Drug dose and thermal treatment of yeast extract

150 μ g of yeast extract was diluted to 45 μ L with dilution buffer of 50 mM HEPES pH 7.5, and 10 mM MgCl₂ with protease inhibitors. A stock solution of 180 mM bosutinib in DMSO was serially diluted to 100 \times solution with DMSO and a subsequent dilution to 10 \times working solution with dilution buffer. 5 μ L of 10 \times working solution was added to 45 μ L of yeast extract, resulting in a 1% final DMSO concentration and final drug concentrations: vehicle, 0.01 nM, 0.1 nM, 1 nM, 10 nM, 100 nM, 1 μ M, 10 μ M, 100 μ M and 1 mM. This is repeated over three technical replicates where each replicate was run on a separate TMT-10 set. Lysates were incubated with drug for 3 min at room temperature before performing heat treatment at 55 °C in a PCR thermocycler for 3 min, followed by cooling at 4 °C for 3 min. Samples were kept on ice and centrifuged at 20 000g for 30 min at 4 °C. The supernatant containing soluble proteins after thermal treatment was transferred to a new tube with care taken to avoid disrupting pelleted proteins.

Time course treatment of the yeast extract

30 μ g of yeast extract was diluted to 50 μ L with dilution buffer of 50 mM HEPES pH 7.5, and 10 mM MgCl₂ with protease inhibitors. The lysates were incubated in a pre-heated 37 °C water bath for 0, 10, 30, 60 and 120 min. Samples taken out of the water bath were placed on ice for 3 min before centrifugation at 20 000g for 20 min at 4 °C. The supernatant containing soluble proteins after thermal treatment was transferred to a new tube with care taken to avoid disrupting pelleted proteins.

Data processing and statistical analysis of CFPS yields

Bar graph results were obtained from at least 3 independent experiments presenting mean \pm SD. Statistical analysis between two groups was performed using two-tailed Student's *t*-test with *p* < 0.05 considered significant. **p* < 0.05; ***p* < 0.01; ****p* < 0.005; NS, not significant.

Sample preparation of thermally treated proteins for mass spectrometry proteomics

Proteins were subjected to acetone precipitation by adding 4 \times volume of ice-cold acetone, incubated at –20 °C for 12 h, then centrifuged at 14 000g for 20 min at 4 °C. The supernatant was removed and the protein pellet was air-dried. The protein pellet was reconstituted in 50 μ L of 8 M urea in 50 mM HEPES, pH 7.5. Proteins were reduced, alkylated and digested using a slightly modified version of a previously reported protocol. In brief,



TCEP was added to 10 mM and incubated at 25 °C for 30 min. CAA was then added to 55 mM and incubated in the dark for 30 min. TEAB buffer was added to dilute 8 M urea to below 2 M urea, and 1 µg lysC was added with incubation for 4 h at 25 °C. TEAB buffer was then added to dilute urea to below 1 M concentration and 1 µg trypsin was added with incubation for 16 h at 25 °C. Digestion was quenched by the addition of TFA to a final 1% (v/v) concentration. Peptides were then desalted using self-packed Empore C18 stage tips. Stage tips were activated with acetonitrile, equilibrated twice with 0.1% formic acid in water. Samples were loaded on the stage-tip twice and then washed twice with 0.1% formic acid in water. Peptides were eluted with 65% acetonitrile, 0.1% formic acid in water. Eluted peptides were dried by vacuum centrifugation. 5 µg of dried peptides were then resolubilized with 7 µl of 100 mM TEAB pH 8.5. 3 µl of TMT-10 labelling reagent was added and incubated at room temperature for 16 h. Labelling was quenched by the addition of 5% hydroxylamine and incubated for 15 min. The samples were pooled together and diluted with 100 µl ammonium formate pH 10 in water. Desalting at high pH was performed using self-packed spin columns (MoBiTec) fritted with a 10 µm pore size filter and loaded with solid phase, ReproSil Pur Basic resin 10 µm particle size (Dr Maisch), in acetonitrile. Spin columns were then equilibrated with 100% acetonitrile and conditioned by passing 10 mM ammonium formate pH 10 through twice. The samples were loaded and then eluted with 50% acetonitrile in 10 mM ammonium formate pH 10. Fractions were dried by vacuum centrifugation and stored at −20 °C before mass spectrometry analysis.

Liquid chromatography mass spectrometry proteomics

TMT-10 labelled peptides were resuspended in water with 2% acetonitrile, 0.5% acetic acid and 0.06% trifluoroacetic acid and loaded on a heated (50 °C) Easy-Spray 75 µm × 50 cm column on a Vanquish Neo (Thermo Scientific) liquid chromatography system coupled to an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Scientific) with an EASY-Spray source. Peptides were resolved at a flow rate of 300 ml min^{−1}, with pre-column equilibration by 100% mobile phase A (0.1% formic acid in water) and resolved by increasing mobile phase B (80% acetonitrile in water with 0.1% formic acid) with a gradient as follows: 0–35% B for 75 min, 35–50% B for 8 min, 50–100% B for 3 min, 100% B for 5 min. The column was equilibrated with 100% A post-run. Mass spectra were collected in Data-Dependent mode with a cycle time of 3 s between master scans. MS1 scans were performed in the Orbitrap with 60 K resolution, AGC target of 400 000 and maximum injection time of 100 ms. MS2 scans collected by Orbitrap with 50 K resolution, 42% HCD collision energy, first mass set at 110, AGC target of 75 000, and maximum injection time of 100 ms.

Mass spectra raw files were searched with SequestHT in Proteome Discoverer 3.0 against a *Saccharomyces cerevisiae* database (retrieved Mar 2017), with the following parameters: precursor mass tolerance of 10 ppm, fragment mass tolerance of 0.06 Da, and trypsin as the enzyme with a maximum of 3 missed cleavages. TMT 10plex was set as the quantification method,

with static modification for carbamidomethyl (C) and TMT modification (protein N-terminus, K), and dynamic modifications for acetyl (protein N-terminus), oxidation (M) and deamidation (N,Q). A strict 1% false discovery rate was set using the percolator node, and reporter ions quantifier nodes were added to the workflow for TMT10 quantification. Output search files were exported as separate.txt files for each of the three replicates.

Data processing and analysis of mass spectrometry data

The three data sets were first individually filtered for missing values and only proteins that were not missing across all samples were retained. The protein abundance at 10 nM for all three replicates and at 1 mM for one of the replicates (set 1) showed systematic shifts in the measurements and were very different from the rest of the samples that cannot be corrected numerically. Thus, we decided to drop these 4 samples from the downstream analysis and used 0.01 nM as the lowest drug concentration value. Next, they were combined into a single data set using 542 common proteins across the three sets. We computed the coefficient-of-variation (%CV) using the lowest three concentrations (0.01 nM, 0.1 nM and 1 nM) and the highest three concentrations (10 µM, 100 µM and 1 mM) for each protein in each replicate. Keeping only proteins with CV below 20% at the highest and lowest three concentrations across all three replicate sets, we retained 457 proteins for the downstream analysis. To remove the differences across the TMT sets, the data was normalized at each concentration against the lowest concentration value to derive the relative protein ratios with respect to 0.01 nM. Then for each replicate, we calculated the mean protein relative ratio at the three highest and three lowest drug concentrations. The values were averaged across replicates and the mean fold change is computed by dividing the average at the higher concentrations by the average at the lower concentrations. Proteins were ranked and prioritized by decreasing magnitude of mean fold changes and those with greater than 1.1 fold increase were subjected to gene ontology (GO) enrichment analysis for biological processes,⁸⁹ and plotted using SRplot.⁹⁰ Protein interactors were plotted using Cytoscape v3.10.0,⁹¹ using StringApp⁹² to retrieve and map STRING protein interactors. A loess curve was used to fit the changes in the protein abundances across the nine concentrations to model the dose response curves. All analyses were carried out using R studio version 4.2.3.⁹³

Data availability

Raw mass spectrometry spectra and search data were uploaded to the jPost repository⁹⁴ with the following accession numbers: JPST002427 (jPOST) and PXD047986 (ProteomeXchange).

(For reviewers, access key 1584, URL: <https://repository.jpostdb.org/preview/11885796746582ad707f418>).

Author contributions

Conceptualization, J. G.; methodology, J. G., Z. L. and R. S.; investigation – Z. L., Z. S., H. K., W. S. and S. C.; writing –



original draft, Z. L. and J. G.; writing – review & editing, Z. L., Z. S., H. K., W. S., R. S. and J. G. Funding acquisition, R. S. and J. G.; supervision, R. S. and J. G.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the Agency for Science, Technology and Research (A*STAR)(Singapore) SIBER grant. We thank A/Prof. Yew Wen Shan from Yong Loo Lin School of Medicine, NUS, for providing the S288c strain and Dr Prakash Arumugam from Institute of Food and Biotechnology Innovation (SIFBI) in A*STAR for providing the BY4743 WT and BY4743 *SSA1Δ* strains. Z. S. is supported by an A*STAR career development award and by an A*STAR Young Achiever award. R. M. S. is supported by awarded A*STAR Core funding and by the Singapore National Research Foundation under its NRF-SIS “Sing-Mass” scheme.

Notes and references

- 1 F. Vincent, A. Nueda, J. Lee, M. Schenone, M. Prunotto and M. Mercola, *Nat. Rev. Drug Discovery*, 2022, **21**, 899–914.
- 2 J. A. Lee and E. L. Berg, *J. Biomol. Screening*, 2013, **18**, 1143–1155.
- 3 L. P. Conway, W. Li and C. G. Parker, *Cell Chem. Biol.*, 2021, **28**, 371–393.
- 4 J. Eder, R. Sedrani and C. Wiesmann, *Nat. Rev. Drug Discovery*, 2014, **13**, 577–587.
- 5 E. D. Carlson, R. Gan, C. E. Hodgman and M. C. Jewett, *Biotechnol. Adv.*, 2012, **30**, 1185–1194.
- 6 N. E. Gregorio, M. Z. Levine and J. P. Oza, *Methods Protoc.*, 2019, **2**, 24.
- 7 A. D. Silverman, A. S. Karim and M. C. Jewett, *Nat. Rev. Genet.*, 2020, **21**, 151–170.
- 8 S. K. Dondapati, M. Stech, A. Zemella and S. Kubick, *Bio-Drugs*, 2020, **34**, 327–348.
- 9 A. Brookwell, J. P. Oza and F. Caschera, *Life*, 2021, **11**, 1367.
- 10 S. J. Moore, J. T. MacDonald and P. S. Freemont, *Biochem. Soc. Trans.*, 2017, **45**, 785–791.
- 11 Y. Tay, C. Ho, P. Droge and F. J. Ghadessy, *Nucleic Acids Res.*, 2010, **38**, e25.
- 12 S. J. Wei, T. Joseph, A. Y. Sim, L. Yurlova, K. Zolghadr, D. Lane, C. Verma and F. Ghadessy, *PLoS One*, 2013, **8**, e62564.
- 13 W. Goh, D. Lane and F. Ghadessy, *Cell Cycle*, 2010, **9**, 3030–3038.
- 14 T. Tabuchi and Y. Yokobayashi, *RSC Chem. Biol.*, 2021, **2**, 1430–1440.
- 15 R. Maini, S. Umemoto and H. Suga, *Curr. Opin. Chem. Biol.*, 2016, **34**, 44–52.
- 16 C. Meyer, Y. Nakamura, B. J. Rasor, A. S. Karim, M. C. Jewett and C. Tan, *Life*, 2021, **11**, 551.
- 17 G. Yin, E. D. Garces, J. Yang, J. Zhang, C. Tran, A. R. Steiner, C. Roos, S. Bajad, S. Hudak, K. Penta, J. Zawada, S. Pollitt and C. J. Murray, *MAbs*, 2012, **4**, 217–225.
- 18 R. W. Martin, N. I. Majewska, C. X. Chen, T. E. Albanetti, R. B. C. Jimenez, A. E. Schmelzer, M. C. Jewett and V. Roy, *ACS Synth. Biol.*, 2017, **6**, 1370–1379.
- 19 J. P. Yang, T. Cirico, F. Katzen, T. C. Peterson and W. Kudlicki, *BMC Biotechnol.*, 2011, **11**, 57.
- 20 A. Zemella, S. Grossmann, R. Sachse, A. Sonabend, M. Schaefer and S. Kubick, *Sci. Rep.*, 2017, **7**, 3740.
- 21 J. J. Wu and J. R. Swartz, *Biochim. Biophys. Acta*, 2008, **1778**, 1237–1250.
- 22 A. R. Goerke and J. R. Swartz, *Biotechnol. Bioeng.*, 2009, **102**, 400–416.
- 23 B. C. Bundy, M. J. Franciszkowicz and J. R. Swartz, *Biotechnol. Bioeng.*, 2008, **100**, 28–37.
- 24 H. Furukawa, H. Inaba, Y. Sasaki, K. Akiyoshi and K. Matsuura, *RSC Chem. Biol.*, 2022, **3**, 231–241.
- 25 Y. Goto, Y. Ito, Y. Kato, S. Tsunoda and H. Suga, *Chem. Biol.*, 2014, **21**, 766–774.
- 26 T.-W. Kim, J.-W. Keum, I.-S. Oh, C.-Y. Choi, H.-C. Kim and D.-M. Kim, *J. Biotechnol.*, 2007, **130**, 389–393.
- 27 R. Gan, M. D. Cabezas, M. Pan, H. Zhang, G. Hu, L. G. Clark, M. C. Jewett and R. Nicol, *ACS Synth. Biol.*, 2022, **11**, 2108–2120.
- 28 M. C. Jewett and J. R. Swartz, *Biotechnol. Bioeng.*, 2004, **86**, 19–26.
- 29 L. R. Burrington, K. R. Watts and J. P. Oza, *ACS Synth. Biol.*, 2021, **10**, 1821–1829.
- 30 A. Choudhury, C. E. Hodgman, M. J. Anderson and M. C. Jewett, *Biochem. Eng. J.*, 2014, **91**, 140–148.
- 31 S. H. Hong, Y.-C. Kwon, R. W. Martin, B. J. Des Soye, A. M. de Paz, K. N. Swonger, I. Ntai, N. L. Kelleher and M. C. Jewett, *ChemBioChem*, 2015, **16**, 844–853.
- 32 L. E. Contreras-Llano, C. Meyer, Y. Liu, M. Sarker, S. Lim, M. L. Longo and C. Tan, *Nat. Commun.*, 2020, **11**, 3138.
- 33 K. A. Calhoun and J. R. Swartz, *Biotechnol. Bioeng.*, 2005, **90**, 606–613.
- 34 P. H. Wang, K. Fujishima, S. Berhanu, Y. Kuruma, T. Z. Jia, A. N. Khusnutdinova, A. F. Yakunin and S. E. McGlynn, *ACS Synth. Biol.*, 2020, **9**, 36–42.
- 35 T. Jaroentomeechai, J. C. Stark, A. Natarajan, C. J. Glasscock, L. E. Yates, K. J. Hsu, M. Mrksich, M. C. Jewett and M. P. DeLisa, *Nat. Commun.*, 2018, **9**, 2686.
- 36 R. Sachse, S. K. Dondapati, S. F. Fenz, T. Schmidt and S. Kubick, *FEBS Lett.*, 2014, **588**, 2774–2781.
- 37 X. Jin, W. Kightlinger and S. H. Hong, *Methods Protoc.*, 2019, **2**, 28.
- 38 R. W. Martin, B. J. Des Soye, Y.-C. Kwon, J. Kay, R. G. Davis, P. M. Thomas, N. I. Majewska, C. X. Chen, R. D. Marcum, M. G. Weiss, A. E. Stoddart, M. Amiram, A. K. Ranji Charna, J. R. Patel, F. J. Isaacs, N. L. Kelleher, S. H. Hong and M. C. Jewett, *Nat. Commun.*, 2018, **9**, 1203.
- 39 F. Caschera and V. Noireaux, *Biochimie*, 2014, **99**, 162–168.



- 40 C. Guarino and M. P. DeLisa, *Glycobiology*, 2012, **22**, 596–601.
- 41 X. Jiang, Y. Ookubo, I. Fujii, H. Nakano and T. Yamane, *FEBS Lett.*, 2002, **514**, 290–294.
- 42 B. Panthu, T. Ohlmann, J. Perrier, U. Schlattner, P. Jalinot, B. Elena-Herrmann and G. J. P. Rautureau, *ACS Synth. Biol.*, 2018, **7**, 218–226.
- 43 N. A. Aleksashin, S. T. Chang and J. H. D. Cate, *RNA*, 2023, **29**, 1960–1972.
- 44 T. Brodiazhenko, M. J. O. Johansson, H. Takada, T. Nissan, V. Hauryliuk and V. Murina, *Front. Microbiol.*, 2018, **9**, 3041.
- 45 D. Mattanovich, M. Sauer and B. Gasser, *Microb. Cell Fact.*, 2014, **13**, 34.
- 46 J. Nielsen, *Bioengineered*, 2013, **4**, 207–211.
- 47 X. Wang, J. Liu, Y. Zheng, J. Li, H. Wang, Y. Zhou, M. Qi, H. Yu, W. Tang and W. M. Zhao, *J. Biosci. Bioeng.*, 2008, **106**, 8–15.
- 48 R. Gan and M. C. Jewett, *Biotechnol. J.*, 2014, **9**, 641–651.
- 49 D. Martinez Molina, R. Jafari, M. Ignatushchenko, T. Seki, E. A. Larsson, C. Dan, L. Sreekumar, Y. Cao and P. Nordlund, *Science*, 2013, **341**, 84–87.
- 50 C. E. Hodgman and M. C. Jewett, *Biotechnol. Bioeng.*, 2013, **110**, 2643–2654.
- 51 B. Ricciuti, S. Baglivo, A. De Giglio and R. Chiari, *Ther. Adv. Respir. Dis.*, 2018, **12**, 1753466618808659.
- 52 L. L. Remsing Rix, U. Rix, J. Colinge, O. Hantschel, K. L. Bennett, T. Stranzl, A. Muller, C. Baumgartner, P. Valent, M. Augustin, J. H. Till and G. Superti-Furga, *Leukemia*, 2009, **23**, 477–485.
- 53 G. Coffey, A. Betz, F. DeGuzman, Y. Pak, M. Inagaki, D. C. Baker, S. J. Hollenbach, A. Pandey and U. Sinha, *J. Pharmacol. Exp. Ther.*, 2014, **351**, 538–548.
- 54 A. Wissner and T. S. Mansour, *Arch. Pharm.*, 2008, **341**, 465–477.
- 55 D. G. Aguilera and A. M. Tsimberidou, *Ther. Clin. Risk Manage.*, 2009, **5**, 281–289.
- 56 N. A. Letassy, D. F. Thompson, M. L. Britton and R. R. Suda, Sr., *DICP, Ann. Pharmacother.*, 1990, **24**, 1204–1209.
- 57 S. S. Mohapatra, S. K. Dwibedy and I. Padhy, *J. Biosci.*, 2021, **46**, 85.
- 58 M. R. Islam, K. Ihenacho, J. W. Park and I. S. Islam, *Sci. Rep.*, 2019, **9**, 11596.
- 59 M. Chamberlin and J. Ring, *J. Biol. Chem.*, 1973, **248**, 2245–2250.
- 60 M. J. Anderson, J. C. Stark, C. E. Hodgman and M. C. Jewett, *FEBS Lett.*, 2015, **589**, 1723–1727.
- 61 L. Dai, T. Zhao, X. Bisteau, W. Sun, N. Prabhu, Y. T. Lim, R. M. Sobota, P. Kaldis and P. Nordlund, *Cell*, 2018, **173**, 1481–1494.
- 62 H. Franken, T. Mathieson, D. Childs, G. M. Sweetman, T. Werner, I. Togel, C. Doce, S. Gade, M. Bantscheff, G. Drewes, F. B. Reinhard, W. Huber and M. M. Savitski, *Nat. Protoc.*, 2015, **10**, 1567–1593.
- 63 F. B. Reinhard, D. Eberhard, T. Werner, H. Franken, D. Childs, C. Doce, M. F. Savitski, W. Huber, M. Bantscheff, M. M. Savitski and G. Drewes, *Nat. Methods*, 2015, **12**, 1129–1131.
- 64 M. M. Savitski, F. B. Reinhard, H. Franken, T. Werner, M. F. Savitski, D. Eberhard, D. Martinez Molina, R. Jafari, R. B. Dovega, S. Klaeger, B. Kuster, P. Nordlund, M. Bantscheff and G. Drewes, *Science*, 2014, **346**, 1255784.
- 65 H. Almqvist, H. Axelsson, R. Jafari, C. Dan, A. Mateus, M. Haraldsson, A. Larsson, D. Martinez Molina, P. Artursson, T. Lundback and P. Nordlund, *Nat. Commun.*, 2016, **7**, 11040.
- 66 Y. Y. Liang, S. Bacanu, L. Sreekumar, A. D. Ramos, L. Dai, M. Michaelis, J. Cinatl, T. Seki, Y. Cao, C. R. Coffill, D. P. Lane, N. Prabhu and P. Nordlund, *Cell Chem. Biol.*, 2022, **29**, 572–585.
- 67 Q. Lu, Y. Zhang, J. Hellner, C. Giannini, X. Xu, J. Pauwels, Q. Ma, W. Dejonghe, H. Han, B. Van de Cotte, F. Impens, K. Gevaert, I. De Smet, J. Friml, D. M. Molina and E. Russinova, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2118220119.
- 68 A. Mateus, J. Bobonis, N. Kurzawa, F. Stein, D. Helm, J. Hevler, A. Typas and M. M. Savitski, *Mol. Syst. Biol.*, 2018, **14**, e8242.
- 69 E. A. Craig and K. Jacobsen, *Cell*, 1984, **38**, 841–849.
- 70 L. E. Contreras-Llano and C. Tan, *Synth. Biol.*, 2018, **3**, ysy012.
- 71 J. Chappell, K. Jensen and P. S. Freemont, *Nucleic Acids Res.*, 2013, **41**, 3471–3481.
- 72 F. Tamaki, F. Fisher, R. Milne, F. S.-R. Terán, N. Wiedemar, K. Wrobel, D. Edwards, H. Baumann, I. H. Gilbert, B. Baragana, J. Baum and S. Wyllie, *Antimicrob. Agents Chemother.*, 2022, **66**, e0023.
- 73 G. B. Hurst, K. G. Asano, C. J. Doktycz, E. J. Consoli, W. L. Doktycz, C. M. Foster, J. L. Morrell-Falvey, R. F. Standaert and M. J. Doktycz, *Anal. Chem.*, 2017, **89**, 11443–11451.
- 74 B. J. Rasor, P. Chirania, G. A. Rybnicky, R. J. Giannone, N. L. Engle, T. J. Tschaplinski, A. S. Karim, R. L. Hettich and M. C. Jewett, *ACS Synth. Biol.*, 2023, **12**, 405–418.
- 75 D. Foshag, E. Henrich, E. Hiller, M. Schafer, C. Kerger, A. Burger-Kentischer, I. Diaz-Moreno, S. M. Garcia-Maurino, V. Dotsch, S. Rupp and F. Bernhard, *N Biotechnol.*, 2018, **40**, 245–260.
- 76 D. Sharma and D. C. Masison, *Protein Pept. Lett.*, 2009, **16**, 571–581.
- 77 M. Werner-Washburne, J. Becker, J. Kosic-Smithers and E. A. Craig, *J. Bacteriol.*, 1989, **171**, 2680–2688.
- 78 S. K. Lotz, L. E. Knighton, Nitika, G. W. Jones and A. W. Truman, *Curr. Genet.*, 2019, **65**, 1127–1134.
- 79 P. B. Kirmizibayrak, R. Ilhan, S. Yilmaz, S. Gunal and B. E. Tepedelen, *J. Biochem.*, 2018, **43**, 101–109.
- 80 M. G. Leeming, S. O'Callaghan, L. Licata, M. Iannuccelli, P. Lo Surdo, E. Micarelli, C. S. Ang, S. Nie, S. Varshney, S. Ameen, H. C. Cheng and N. A. Williamson, *Bioinformatics*, 2021, **37**, 1635–1636.
- 81 D. C. Garcia, J. L. N. Dinglasan, H. Shrestha, P. E. Abraham, R. L. Hettich and M. J. Doktycz, *Metab. Eng. Commun.*, 2021, **12**, e00162.
- 82 A. Zemella, L. Thoring, C. Hoffmeister and S. Kubick, *ChemBioChem*, 2015, **16**, 2420–2431.



- 83 A. M. Vieira Gomes, T. Souza Carmo, L. Silva Carvalho, F. Mendonça Bahia and N. S. Parachin, *Microorganisms*, 2018, **6**, 38.
- 84 J. A. Schoborg, L. G. Clark, A. Choudhury, C. E. Hodgman and M. C. Jewett, *Synth. Syst. Biotechnol.*, 2016, **1**, 2–6.
- 85 L. Zhang, W. Q. Liu and J. Li, *Front. Bioeng. Biotechnol.*, 2020, **8**, 536.
- 86 Y. Wang, T. Wang, X. Chen and Y. Lu, *Bioresour. Bioprocess*, 2023, **10**, 35.
- 87 J. A. Schoborg, C. E. Hodgman, M. J. Anderson and M. C. Jewett, *Biotechnol. J.*, 2014, **9**, 630–640.
- 88 F. Guzman-Chavez, A. Arce, A. Adhikari, S. Vadhin, J. A. Pedroza-Garcia, C. Gandini, J. W. Ajioka, J. Molloy, S. Sanchez-Nieto, J. D. Varner, F. Federici and J. Haseloff, *ACS Synth. Biol.*, 2022, **11**, 1114–1128.
- 89 P. D. Thomas, D. Ebert, A. Muruganujan, T. Mushayahama, L. P. Albou and H. Mi, *Protein Sci.*, 2022, **31**, 8–22.
- 90 D. Tang, M. Chen, X. Huang, G. Zhang, L. Zeng, G. Zhang, S. Wu and Y. Wang, *PLoS One*, 2023, **18**, e0294236.
- 91 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498–2504.
- 92 N. T. Doncheva, J. H. Morris, J. Gorodkin and L. J. Jensen, *J. Proteome Res.*, 2019, **18**, 623–632.
- 93 R. C. Team, *Journal*, 2023.
- 94 S. Okuda, Y. Watanabe, Y. Moriya, S. Kawano, T. Yamamoto, M. Matsumoto, T. Takami, D. Kobayashi, N. Araki, A. C. Yoshizawa, T. Tabata, N. Sugiyama, S. Goto and Y. Ishihama, *Nucleic Acids Res.*, 2017, **45**, D1107.

