PNAS
www.pnas.org

Supplementary Information for

Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae.*

Olena P. Ishchuk[1*], Iván Domenzain[1,2], Benjamín J. Sánchez[1,2,3,4], Facundo Muñiz-Paredes[1], José L. Martínez[1,3], Jens Nielsen[1,2,5], Dina Petranovic[1,2 *]

[1]*Department of Biology and Biological Engineering, Systems and Synthetic Biology, Chalmers University of Technology, Gothenburg, Sweden*

[2]*Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, Gothenburg, Sweden*

[3]*Department of Biotechnology and Biomedicine, Section for Synthetic Biology, Technical University of Denmark, Kgs. Lyngby, Denmark*

[4]*Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark*

[5]*BioInnovation Institute, Copenhagen, Denmark*

*Olena P. Ishchuk, Dina Petranovic

Email: ishchuk@chalmers.se and dina.petranovic@chalmers.se

This PDF file includes:

Supplementary text
Figures S1 to S15
SI References

Other supplementary materials for this manuscript include the following:

Dataset S1 to S3

Supplementary Information Text

Preliminary target selection using Yeast8

As a first model-driven tool for obtaining candidates with increased heme production, we followed a recently implemented constraint-based approach [1,2], which adapted the procedure of flux scanning based on the enforced objective flux (FSEOF) method [3]. We simulated *S. cerevisiae*'s growth on glucose at 16 different biomass yields, ranging from half of the experimental biomass yield (0.122 gDW/g) to twice that value, in uniform intervals. All unused carbon for biomass accumulations was directed towards heme production. The average simulated flux fold-change for each reaction was then calculated as a score, equal to the mean flux across all simulations divided by the reaction rate at maximum growth rate (WT) conditions. These scores were in turn used to compute gene scores, as average scores across all corresponding reactions. Gene scores above one correspond to genes associated with reaction fluxes, which increase as heme production increases and are therefore candidates for upregulation. Gene scores between zero and one represent genes connected to reaction fluxes that go down as heme production goes up and are therefore targets for down-regulation. Finally, gene scores equal to zero are genes that get turned off during production of heme and are therefore adequate deletion targets. Inconsistent reaction scores, *i.e.*, reactions with simulated flux fold-changes both above and below one at varying conditions, and inconsistent gene scores, *i.e.*, genes associated with reaction scores both above and below one, were filtered out of the analysis. The yeast consensus genome-scale model [2] version 8.0.1 [4], the parsimonious flux balance analysis (pFBA) approach [5], and the COBRA toolbox [6] were used for all simulations.

Reference flux distribution using ecYeast8

The pFBA approach approximately minimizes the active enzymatic steps (as a proxy for the total enzyme mass burden, as enzymes are ignored in standard models of metabolism) while achieving a fixed cellular objective.

In ecYeast8 [2], the enzyme requirements for a metabolic reaction are formulated as equations of the form (for all i)

$$v_i = k_{cat_{ij}} e_j$$

coupled with the inequalities

$$\sum_{j=1}^{P} MW_j e_j \leq P_{pool}$$

where the notation is standard: v_i indicates the flux for i-th reaction in the network (given in mmol/gDW h); $k_{cat_{ij}}$, is the turnover number for the j-th enzyme participating in the i-th reaction (given in h^{-1}); e_j, represents the enzyme usage for the j-th enzyme (in mmol/gDW); MW_j, the molecular weight for the j-th enzyme (in g/mmol); and P_{pool}, represents the total protein pool available for metabolic enzymes in the cell (in $g_{protein}$/gDW). These enzymatic constraints enable an explicit implementation of the pFBA approach in which the total enzyme burden of the reaction network can be minimized. Therefore, a parsimonious flux distribution is obtained as an optimal basic feasible solution of the linear program

$$\max: Z = C^T v$$

Subject to:

$$S \cdot v = 0$$
$$lb_j \leq v_j \leq ub_j$$

Then a second optimization is run by fixing the optimal value for the objective function and minimizing the total protein usage:

$$min: \sum_{j=1}^{P} MW_j e_j$$

Subject to:

$$C^T v = Z^{max}$$

where v indicates the vector of reaction fluxes (in mmol/gDW h); S represents the stoichiometric matrix with metabolites and enzymes being its rows and reactions its columns; C^T is a vector of coefficients for the objective function (Z), which is a linear combination of the reaction fluxes in the network.
A reference pFBA flux distribution for optimal heme production used in this study was calculated by maximizing heme production (reaction r_s3714_Ex in both Yeast8 and ecYeast8) subject to a fixed suboptimal biomass yield (experimental value of 0.122 g biomass/g glucose) and then minimizing for the total protein burden.

Gene target selection using ecYeast8
Gene targets for increased heme production were identified by using the FSEOF-based procedure [1,2,3], and the model ecYeast8 v8.3.4 [2]. An implementation of this procedure suitable for the characteristics and structure of enzyme-constrained models (ecModels) was added to the GECKO toolbox and is available at:
https://github.com/SysBioChalmers/heme_production_ecYeastGEM/releases/tag/v1.0.

Identification of an optimal combination of targets using ecYeast8
To reduce the number of gene target candidates for experimentation, a multi-step procedure that leverages the enzymatic information of metabolic reactions in ecModels was developed. This procedure consists of a classification of targets by enzyme usage variability ranges, sequential *in silico* validation of individual targets, and cumulative integration of top candidates into an *in silico* mutant.

1.Enzyme usage variability analysis
The genes, which are related to enzymes included as pseudo-metabolites in the ecModel, were then classified into different categories according to their utilization profile. Enzyme usage variability analysis (EVA) was run for every enzyme connected to the target gene candidates, by independently solving FBA problems using minimization and maximization of each of the enzyme usage reactions as objective functions. A suboptimal biomass production yield was fixed as 50% of the maximum yield with a glucose uptake rate of 1 mmol/gDW h, and the maximum objective production rate (at the same conditions) was used as a lower bound for the objective reaction. Enzyme usage variability ranges were obtained by subtracting the minimum flux value from the maximum value for each of the probed enzymes. Additionally, a parsimonious flux distribution was obtained by minimizing the total enzyme mass demand (protein pool usage pseudo-reaction), while subject to the same set of constraints used for EVA. These results enabled the prioritization of gene candidates linked to modeled enzymes according to the following categories:

- **Level 1:** Non-redundant targets (genes without isoenzymes) for overexpression with a minimum usage higher than 0 mmol/gDW and close to its maximum usage value (tightly-constrained essential enzymes). Non-redundant targets for deletion of knock-down with a predicted maximum usage of 0 mmol/gDW (unusable enzymes).

- **Level 2:** Isoenzyme targets for overexpression with minimum usage higher than 0 mmol/gDW (optimal isoenzymes). Suboptimal isoenzyme candidates for knock-out/down with maximum usage higher than 0 mmol/gDW (suboptimal isoenzymes).

- **Level 3:** Isoenzyme candidates for overexpression with a predicted parsimonious usage higher than 0 but lower than its maximal usage. Groups of isoenzymes or complex subunits that are not used for optimal production levels (minimum and maximum usages equal to 0 mmol/gDW). All gene target candidates that are not associated with a parameterized enzyme in the ecModel are assigned to this category.

The remaining enzyme targets that did not fit into any of the priority levels above were discarded. The isoenzymes and enzyme complex subunits were identified by finding all groups of identical vectors in a metabolite-gene matrix, which connects all genes with the metabolites involved in the

reactions catalysed by their corresponding enzymes. In an ecModel, the metabolite-gene matrix is defined as:

$$MG = logical(S) * RxnGeneMat$$

Where $logical(S)$ is the Boolean form of the model's stoichiometric matrix, representing all metabolites as rows and reactions as columns; $RxnGeneMat$, is a Boolean matrix that establishes the relationship between enzyme-encoding genes and biochemical reactions in the ecModel, in which a value of 1 indicates genes whose protein products catalyse a given reaction.

2.Mechanistic genetic modifications using ecYeast8

As ecModels incorporate enzymes as part of the metabolic network explicitly, constrained by their specific kinetic parameters, and limited by a protein pool with a limited mass, the effects of genetic modifications on metabolic capabilities can be directly computed.

The gene deletions were performed by blocking the usage of pseudo-reactions for the enzymes encoded by these genes. In this way such enzymes are not available for carrying any flux as they are not being "drawn" from the available protein pool. Moreover, as deleted sequences cannot be transcribed and then translated into proteins, the available protein mass can be used for synthesis of other enzymes relevant for production. For a gene target encoding for the enzyme E_i, with a molecular weight MW_i, an enzyme usage reaction is represented in the model as:

$$e_i: MW_i P_{pool} \rightarrow E_i$$

A deletion of the gene encoding by enzyme E_i is then introduced by setting the upper bound of e_i to zero.

For simulation of gene overexpressions the following procedure was applied. Due to the internal structure of the stoichiometric matrix in an ecModel, metabolic fluxes can be expressed as:

$$v_i^{WT} = k_{cat}^{ij} * e_j$$

A perturbed expression of the enzyme e_j will, therefore, affect the flux capacity for the corresponding reaction:

$$v_i^{mut} = k_{cat}^{ij} * e_j^{mut}$$

Where e_j^{mut} represents the usage of enzyme E_j in the mutant phenotype. The later relationship can also be represented as a function of the original enzyme usage variable (e_j, obtained from wild-type simulation):

$$v_i^{mut} = k_{cat}^{ij} * \left(f_{E_j}^{mut} * e_j\right)$$

Where $f_{E_j}^{mut}$ represents a mutant expression factor for enzyme E_j. When simulating gene over-expressions, we used this relationship and set $f_{E_j}^{mut}$ to a value of 2. In the cases that this would yield unfeasible simulations (*e.g.* for tightly controlled enzymes, where the model cannot find a solution with such an increase in enzyme usage), the kinetics where in turn increased by the same factor, as the previous equation is equivalent to:

$$v_i^{mut} = \left(f_{E_j}^{mut} * k_{cat}^{ij}\right) * e_j$$

This means that, in these cases, all the turnover numbers associated to the enzyme were multiplied by the same mutant expression factor of 2. In an ecModel this is also a straightforward procedure, due to the explicit integration of enzyme turnover numbers in the expanded stoichiometric matrix.

It should be noted that the simulation of overexpressed or knocked-down enzymes does not necessarily induce a perturbed reaction flux, since in ecModels metabolic fluxes can be constrained by either, nutrients or enzyme mass availability.

Independent FBA simulations were performed for each genetic modification, using unlimited nutrients availability (in a mineral minimal media) and a fixed growth rate. For each individual gene modification, product formation was then maximized. Fold-changes between the wild-type and each mutant strain were calculated for both product formation rate and product yield (mmol of product divided by the glucose uptake rate). An evaluation metric for the effect gene perturbation *i* over product *P* is defined as:

$$\eta_i = \frac{1}{2}\left(FC_P^{rate} + FC_P^{yield}\right) = \frac{v_P^{MAX_{mut}}}{v_P^{MAX_{WT}}} + \frac{1}{2}\left(\frac{v_P^{MAX_{mut}}}{v_P^{MAX_{WT}}} * \frac{v_{glc}^{in_{WT}}}{v_{glc}^{in_{mut}}}\right)$$

Gene modifications resulting in an average fold-change lower or equal to 1 were then discarded from the list of gene target candidates. Gene modifications resulting in an average fold-change higher than 1 were selected.

3.Generation of optimal combinations of gene targets

After we validated *in silico* all individual gene-target candidates (using the ecModel and FBA simulations), we assessed which genetic modifications can be combined in a feasible strain that maximizes heme accumulation. The list of generated gene candidate targets were sorted by priority level. At each level the gene candidates were ordered according to their η_i value, obtained from their individual *in silico* validation.

A sequential procedure is then followed, which adds each of the ranked gene targets to a mutant model in a cumulative way. This procedure uses the same constraints described above and for each newly added genetic modification an evaluation metric η_i is computed, using the same FBA setting described in the previous section. The metric η_i is then compared to the value of the previous iteration (η_{i-1}). Gene modifications that did not increase the production evaluation metric, in comparison to the previous combined mutant model, were removed from the updated strain model. This approach ensures that only modifications with a positive impact on the production of the desired chemical are kept in the mutant model, drastically reducing the number of remaining gene target candidates whilst verifying growth viability of the resulting strain at each step.

Media and growth conditions

Strains were maintained on rich medium YPD (20 g/L glucose, 10 g/L peptone, 5 g/L yeast extract). Yeast transformants were selected on the minimal medium SD without uracil (20 g/L glucose, 6.9 g/L YNB without amino acids, amino acids dropout mixture without uracil) supplemented with 20 g/L of agar. For the heme assay, BY4741 knockout strains were cultivated in liquid YPD at pH 6.0; CEN.PK.113-11c transformants were cultivated in liquid SD without uracil (20 g/L glucose, 6.9 g/L YNB without amino acids, amino acids dropout mixture without uracil) supplemented with 100 mM glycine, and 100 µM ferric citrate at pH 6.0. In addition, for the heme assay, the phosphate buffer was used to buffer the medium. For *URA3* plasmids removal, strains were maintained on SD medium with 50 mg/L uracil and 1 g/L 5-fluorotic acid (5-FOA).

Genome engineering

The CRISPR-Cas9 genome engineering was performed using the IMX581 strain carrying the CRISPR-Cas9 construct integrated into the genome [7], and it was used as a recipient for plasmids derived from the pMEL10 vector [7] carrying gRNAs targeting genes to be deleted or gRNAs targeting intergenic loci for the integration of expression cassettes.
First, we integrated the *HEM15* gene overexpression cassette (prTEF1-HEM15-terADH1) into the XI-3 integration site of XI chromosome (Fig.S1). For this purpose, the prTEF1-HEM15-terADH1

cassette was amplified from the pRS316+prTEF1+HEM15+terADH1 plasmid with primers (delXI3—1 and delXI3—2) with homology of 48 bp upstream and 49 bp downstream to the XI-3 integration region. The amplified cassette was co-transformed with the gRNA plasmid pMEL10+XI-3up_gRNA into the IMX581 strain. After verification of the integration event, the episomal plasmid pMEL10+XI-3up_gRNA was removed by the incubating of the strain on SD medium supplemented with 5-FOA, uracil and 2% glucose. The strain without gRNA plasmid but with *HEM15* expression cassette integrated into genome, IMX581-HEM15, was further transformed with the pRS316+prTEF1+HEM13+terADH1 plasmid. Heme production of the resulting IMX581-HEM15-HEM13 strain was shown to be 2-fold higher than the IMX581-HEM13 strain and 3-fold higher than that of the IMX581 control strain (Fig.S2). Next, we amplified the *HEM14* gene expression cassette from the pRS316+prTEF1+HEM14+terADH1 plasmid with primer pair (XII-5up3 and XII-5up4) and fused it by overlap PCR with 142 bp and 129 bp fragments homologous to upstream and downstream regions of the XII-5 integration site (Fig.S1). The resulting cassette was co-transformed with the pMEL10+XII-5up_gRNA plasmid into the IMX581-HEM15 strain followed by gRNA plasmid removal as described above. The strain with the *HEM14* gene expression cassette, the IMX581-HEM15-HEM14, was then transformed with the pRS316+prTEF1+HEM13+terADH1 plasmid. At 24 hours of growth, the IMX581-HEM15-HEM14-HEM13 strain produced heme at the same level as IMX581-HEM15-HEM13, however at 48 hours its heme production was 1.4 times higher, and at 72 hours it was 1.7 times higher than that of IMX581-HEM15-HEM13 (Fig.S2). Compared to the IMX581 strain carrying an empty vector, IMX581-HEM15-HEM14-HEM13 produced 2.3-, 4.3-, and 4.2-more at 24, 48, and 72 hours respectively (Fig.S2). Next, the *TEF1*-expression cassette for the *HEM3* gene was amplified with upstream (237 bp) and downstream (237 bp) homologous regions to XII-5 site and integrated using the gRNA (pMEL10+XII-5down_gRNA) plasmid for the second position at XII-5 site (Fig.S1). The resulting IMX581-HEM15-HEM14-HEM3-HEM13 strain produced slightly more heme at 48 and 72 hours when compared to the IMX581-HEM15-HEM14-HEM13 strain (Fig.S2). Next, the *SHM1* gene was deleted and the resulting strain, the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM13, produced 5-fold more heme than the IMX581 strain at 72 hours (Fig.S2), and upon buffering the cultivation medium the heme production was ~14-fold higher than the IMX581 (Fig.S3). Overexpression of the *ACH1* gene by integrating its expression cassette into the XII-2 locus of the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM13 strain background did not improve heme production (Fig.S3). The transformants (IMX581-HEM15-HEM14-HEM3-∆shm1-∆gcv2-HEM13) with deletion of the *GCV2* gene (involved in glycine catabolism) had slightly increased heme production at 24 hours, but later had a very large distribution of heme production among biological replicates (Fig.S2). The deletion of this gene in the IMX581-HEM15-HEM14-∆faa4-HEM13 background resulted in lower heme production (Fig.S4). As Gcv2 is one of subunits of the mitochondrial glycine decarboxylase complex involved in mitochondrial glycine cleavage [8], we assumed that the genes encoding other subunits of this complex, for example, the *GCV1, GCV3*, and *LPD1* must be deleted as well to have an effect. Both *GCV1* and *GCV2* deletions were found beneficial in the strains with other gene combinations (for example, as shown in Fig.S7). The subsequent integration of the *HEM2* overexpression cassette into the XII-2 locus improved production at 24 hours and resulted in 1.8 times more heme than in the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2-HEM13 strain compared to the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM13 (Fig.S3). Using hemin as a standard in the assay, the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2-HEM13 strain was found to produce ~1 µg of heme per OD of biomass. While the deletion of the *FDH1*, and *YLR446w* genes in the IMX581-HEM15-HEM14 strain background improved heme production only at 24 hours (Fig.S4), deletions of *FAA4*, *FDH1*, and *YLR446w* in the IMX581-HEM15-HEM14-HEM3-∆shm1 strain with *HEM13* overexpression did not improve heme production (Fig.S5). Expression of the *HEM1* gene in the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2 strain transformant had great distribution in heme production throughout biological replicates, however, the median heme production was ~20-33% higher when compared to IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2-HEM13 strain (Fig.S5A). Cells of the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2-HEM13 and IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2-HEM1 strains had a slightly reddish color when grown on solid medium (Fig.S5B). Among the strains tested, only the IMX581-HEM15-HEM14-HEM3-∆shm1-HEM2-HEM13 strain developed a strong red color after the addition of 5-aminolevulinic acid (5-ALA) (Fig.S5B), which

is the product of Hem1 enzyme. We suppose that there is a limitation at the *HEM1* step, and that improving Hem1 activity through overexpression of this gene or by addition of the pyridoxal phosphate (Hem1 cofactor) supply could increase heme production further. Next, we used the *FET4*, *ADH1*, and *ARH1* genes expression cassettes to knockout the *HMX1* gene ORF, which encodes heme oxygenase that is responsible for heme degradation [9]. The expression cassettes carried 288 bp upstream and 265 bp downstream of the *HMX1* ORF, and by the cassette's integration into this locus, most of the *HMX1* gene ORF was deleted. Integration of the *FET4* and *ADH1* expression cassettes integration resulted in a slight increase in heme production at 24 hours, however the production in strains with the expression of the *ADH1* and *ARH1* gene produced lower amount of heme with prolonged cultivation time, at 48 and 72 hours correspondingly (Fig.S6). In the IMX581-HEM15-HEM14-HEM3-Δshm1-HEM2-Δhmx1-FET4 strain, we had further deleted the *GCV2* gene, which improved heme production further by 30-40% (Fig.S7). Furthermore, the integration of the *HEM1* gene expression cassette, the deletion of the *GCV1* gene, and the overexpression of the *HEM13* gene resulted in an improved heme production ranging from 42 to 70-fold during the cultivation at 72 hours (Fig.S8).

Determination of glucose concentration

The glucose concentration in the culture media was determined by Glucose Colorimetric Detection Kit, EIAGLUC (Invitrogen, Thermo Fisher Scientific) following the manufacturer recommendations.

Cell dry weight (CDW) analysis

For the CDW measurements, the cells were collected on pre-weighted 0.45 μm cellulose-acetate filter paper (Satorius Biolabs). The filter with cells was then washed with MQ, dried for 15 min in a microwave (at 325 W), and placed in a desiccator for a further 3 days of drying. The weight increase was recorded as CDW.

Determination of heme concentration

Heme was extracted from the cells by oxalic acid treatment and determined as described before [10]. For heme extraction, the cells in amount of 8 OD_{600} of cells were collected by centrifugation at 2688 *g*, washed with MQ, resuspended in 500 microliters of 20 mM oxalic acid, and further mixed with 500 microliters of 2 M oxalic acid. In addition, the 500 microliters of whole culture (cells and the medium) were collected and mixed with 500 microliters of 2 M oxalic acid. The mixtures of cells or whole culture (cells and the medium) with oxalic acid were further incubated at 98℃ for 30 min. After the incubation, the extracts were centrifuged for 5 min at 4200 *g* and the supernatant was further used for fluorescence measurement. After oxalic acid treatment heme loses iron and becomes fluorescent. The fluorescence was detected using filters with excitation at λ=400nm and emission at λ=600nm on a FLUOstar Omega plate reader spectrophotometer. Hemin (Sigma-Aldrich) was used as a standard to estimate heme concentration.

Heme biosensor

This heme biosensor, which was developed in this study, was designated Heme-LBB (heme ligand-binding biosensor). For the development of the Heme-HBB we used the synthetic fusion construct of **α**- and γ-globins [11] as a base. The synthetic fusion construct of **α**- and γ-globins was first codon optimized for expression in *S. cerevisiae*, and its N-terminus was fused with GFP and cloned into pRS316 vector under control of the copper inducible promoter *CUP1* and with the terminator *ADH1* resulting in biosensor plasmid pRS316+prCUP1-GFP-Hb-terADH1. The biosensor plasmid (pRS316+prCUP1-GFP-Hb-terADH1) was transformed in strains carrying different engineered gene targets engineered. For the *CUP1* promoter induction, 0.05 mM CuSO4 was added to the medium. The biomass (gains settings of 20) and GFP fluorescence (gains settings of 100) were measured of yeast cultures grown in s48 well MTP-48 FlowerPlates on BioLector bioreactor at 30℃ and 400 rpm.

Hemoglobin is a heme-containing protein and heme incorporation during translation determines the correct folding of hemoglobin [12, 13]. Thus, we assumed that if there is more intracellular heme, there will be more correctly folded GFP-hemoglobin protein (Heme-LBB) and its

fluorescence will be higher. The biosensor fluorescence was found to have a positive correlation with heme production in strains tested.
Very low autofluorescence (10 units) was detected in the control strain (IMX581 carrying an empty plasmid) and the biosensor plasmid in the IMX581 strain resulted in a GFP-fluorescence yield of around 18 units. Integration of the *HEM15* gene resulted in an output of 19 units yield (Fig.S9). Fluorescence of the biosensor with the *HEM14* gene integration was 20 units, 19 units with the *HEM3* gene, 20 units with the Δ*shm1* mutation, and 22 units with the Δ*gcv2* mutation (Fig.S9). However, for Δ*faa4* strain the biosensor GFP yield decreased to 19 units, and in the strain with Δ*ylr446w*, Δ*fdh1* and *ACH1* modifications, the fluorescence yield was as low as WT (Strain1) (Fig.S9). Strikingly, the integration of the *HEM2* gene overexpression cassette resulted in increased fluorescence yield of ~13-fold at 24 hours (200 units), and of ~8-fold (120 units) at 60 hours compared to the IMX581 strain (Strain 1) (Fig.S9). No overlap between native heme fluorescence and the heme biosensor fluorescence in the GFP channel was found, as deduced by analyzing heme production strains without the biosensor plasmid (Fig.S10).

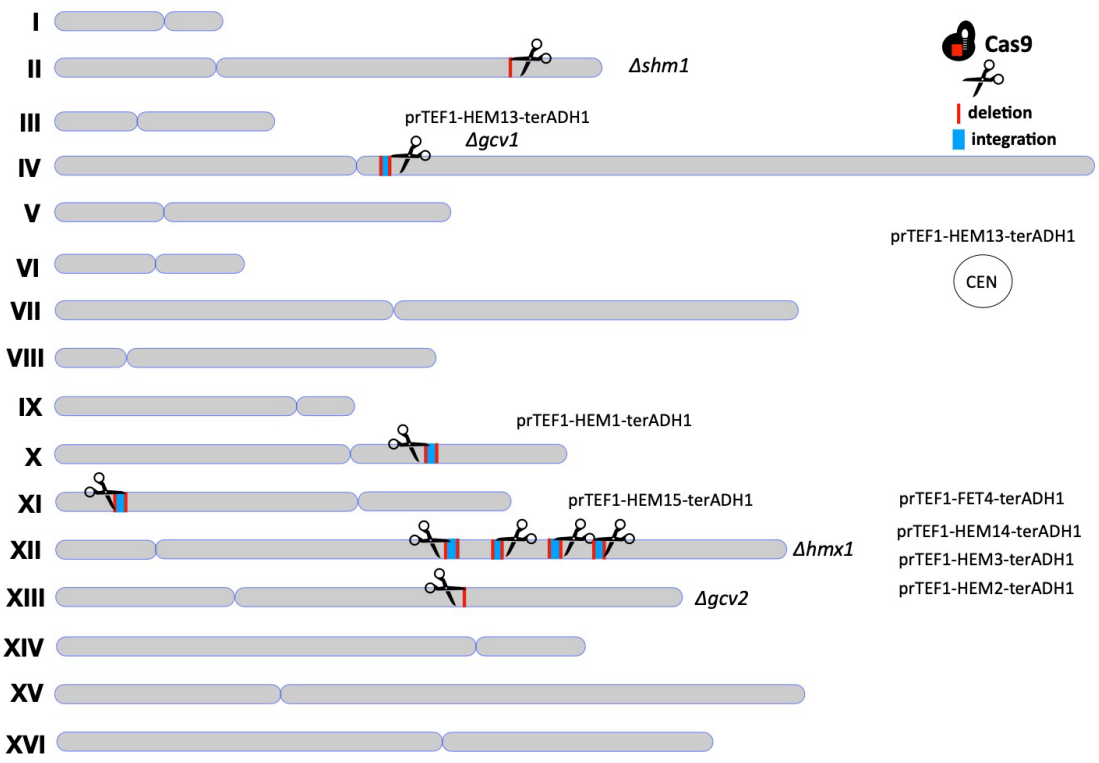


Fig. S1. Genome engineering using CRISPR-Cas9.
The expression cassettes carrying the *TEF1* promoter, target gene, and *ADH1* terminator were integrated into chromosomes: IV (one site), X (one site), XI (one site), XII (four sites). The genes *SHM1, GCV1, GCV2,* and *HMX1* were deleted. The Cas9 targeting to different sited was performed using corresponding gRNAs. The *HEM13* gene was also expressed under control of the *TEF1* promoter from centromeric plasmid.

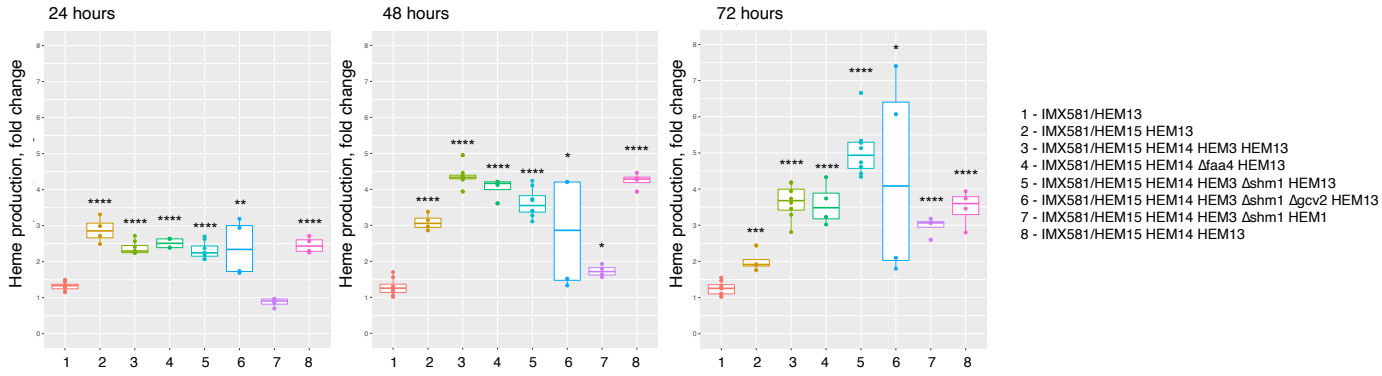


Fig. S2. Heme production of *S. cerevisiae* strains carrying expression cassettes and gene deletions of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four to six biological replicates (transformants) were used in the experiment. Statistical analysis was performed using one-way ANOVA (****$p \leq 0$, ***$p \leq 0.00029$, **$p \leq 0.00272$, *$p \leq 0.01997$) and the control strain IMX581/HEM13. IMX581 carrying an empty vector (pRS316) was used to normalize data.

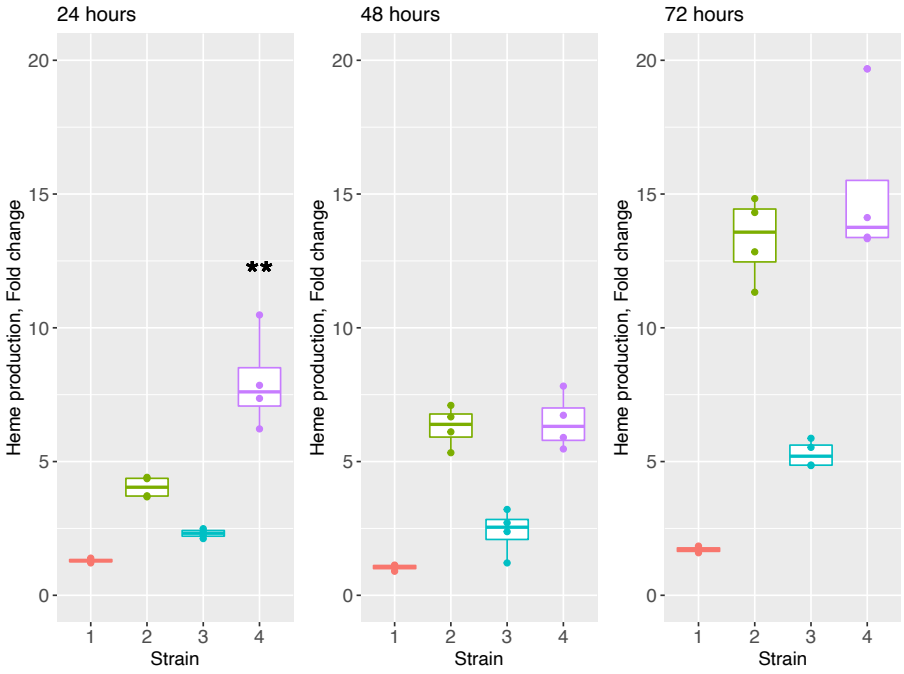


Fig. S3. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment. Strains: 1 – IMX581/HEM13; 2 – IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM13; 3 – IMX581/HEM15 HEM14 HEM3 Δshm1 ACH1 HEM13; 4 – IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM2 HEM13. Statistical analysis was performed using one-way ANOVA (**$p \leq 0.0053$) and the control strain IMX581/HEM15 HEM14 HEM3 Δshm1 HEM13.

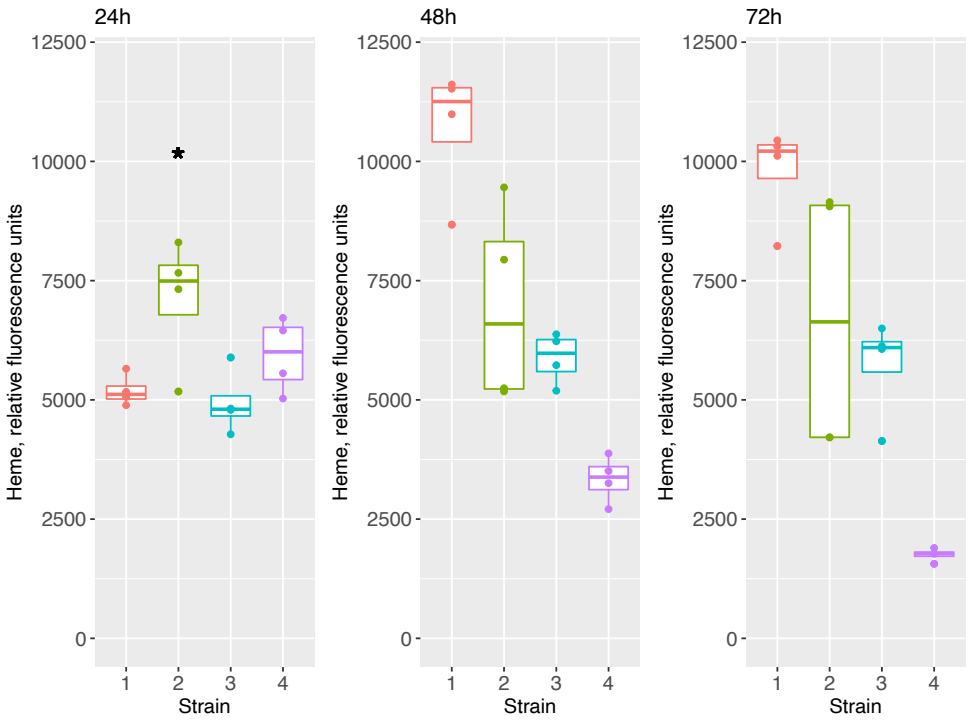


Fig. S4. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment. Strains: 1 – IMX581/HEM15 HEM14 Δfaa4 HEM13; 2 – IMX581/HEM15 HEM14 Δfaa4 Δfdh1 HEM13; 3 – IMX581/HEM15 HEM14 Δfaa4 Δgcv2 HEM13; 4 – IMX581/HEM15 HEM14 Δfaa4 Δylr446w HEM13. Statistical analysis was performed using one-way ANOVA (*$p \leq 0.03305$) and the control strain IMX581/HEM15 HEM14 Δfaa4 HEM13.

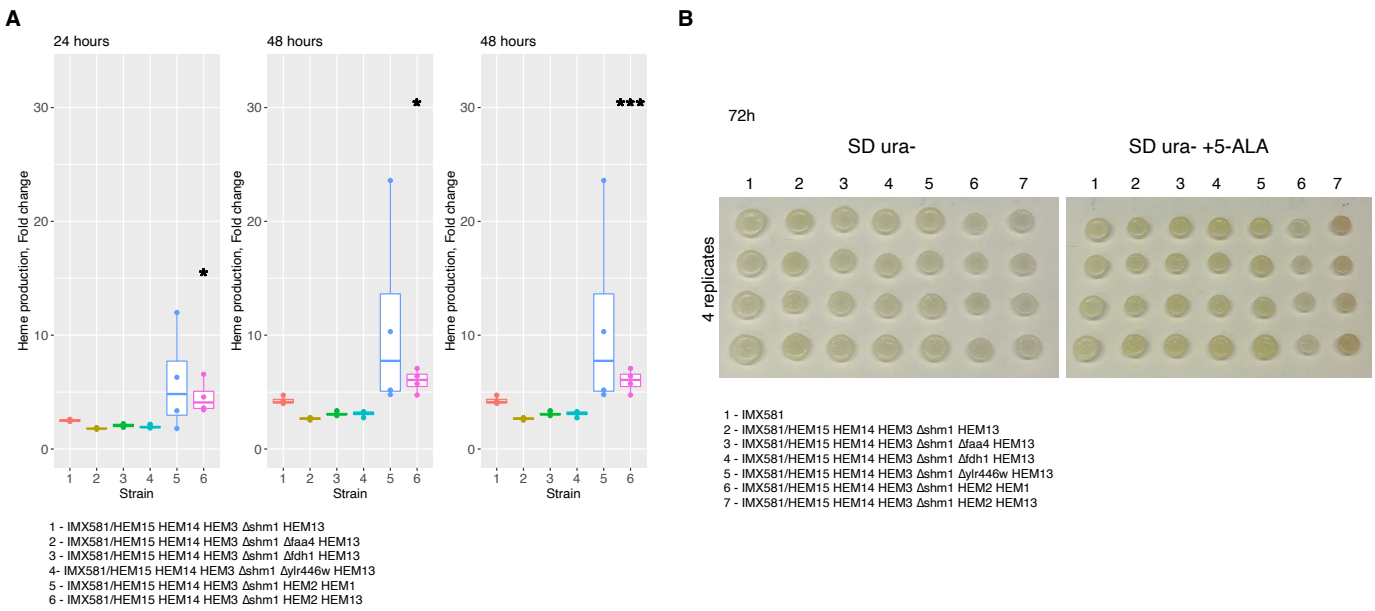


Fig. S5. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
A) Heme production was analyzed at 24, 48 and 72 hours in buffered SD ura- supplemented with 2% glucose, 100 mM glycine and 100 µM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment. IMX581 carrying empty vector (pRS316) was used to normalize data. B) Growth on solid SD-ura medium with and without the addition of 5 mM 5-ALA. Statistical analysis was performed using one-way ANOVA (*$p \leq 0.03001$, ***$p \leq 0.00022$) and the control strain IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM13.

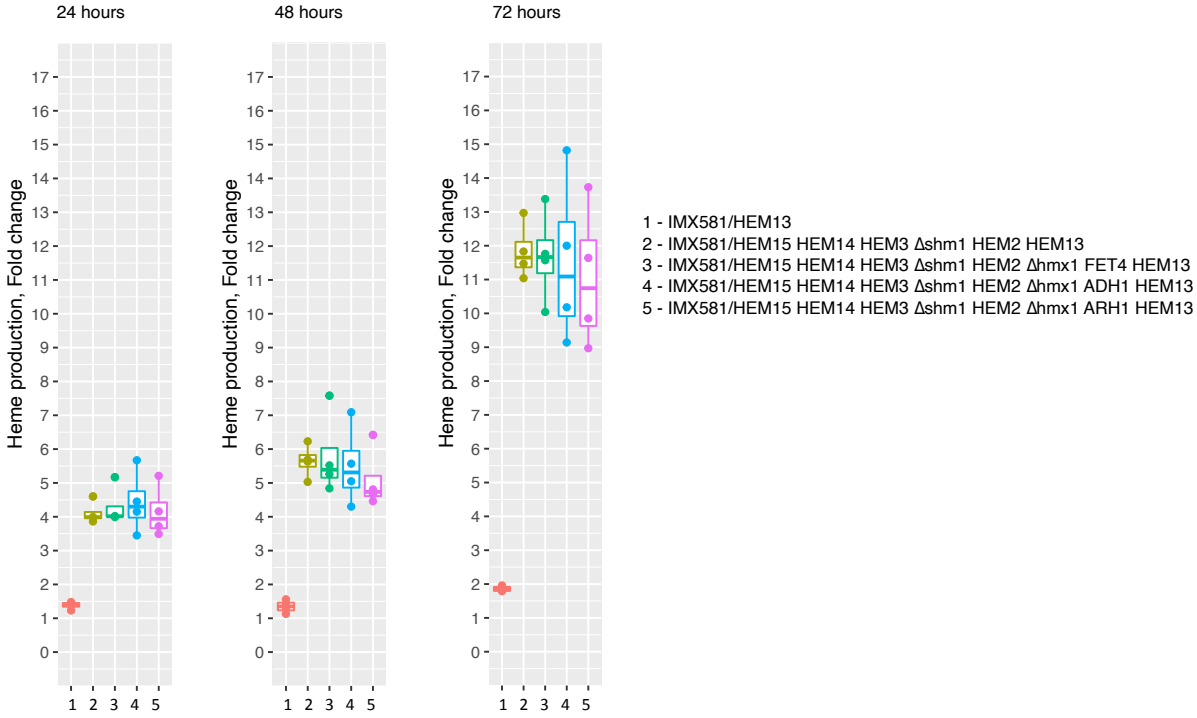


Fig. S6. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment.

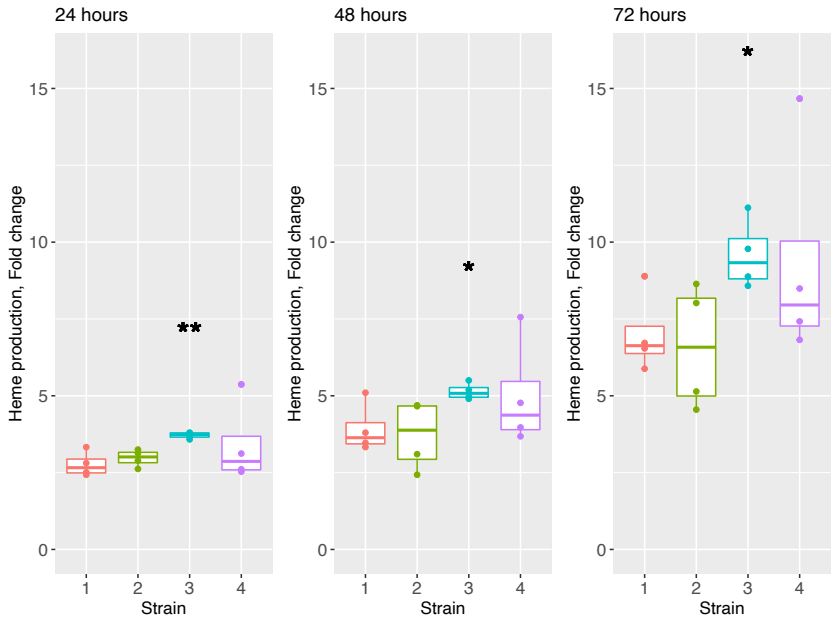


Fig. S7. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment. Strains: 1 – IMX581/HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 FET4 HEM13; 2 – IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 ADH1 HEM13; 3 – IMX581/HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 FET4 Δgcv2 HEM13; 4 – IMX581/HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 ADH1 Δgcv2 HEM13. Statistical analysis was performed using one-way ANOVA (**$p \leq 0.00418$, *$p \leq 0.02905$) and the control strains IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 FET4 HEM13 and strains IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 ADH1 HEM13.

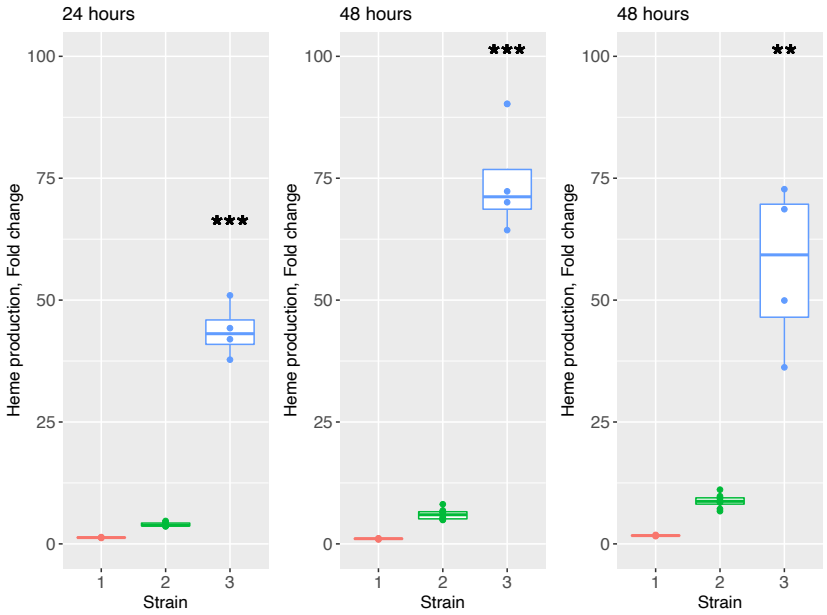


Fig. S8. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment. Strains: 1 – IMX581/HEM13; 2 – IMX581/HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 FET4 Δgcv2 HEM13; 3 – IMX581/HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 FET4 Δgcv2 HEM1 HEM13. Statistical analysis was performed using one-way ANOVA (***$p \leq 0.0002$, **$p \leq 0.00174$) and the control strain IMX581/ HEM15 HEM14 HEM3 Δshm1 HEM2 Δhmx1 FET4 Δgcv2 HEM13.

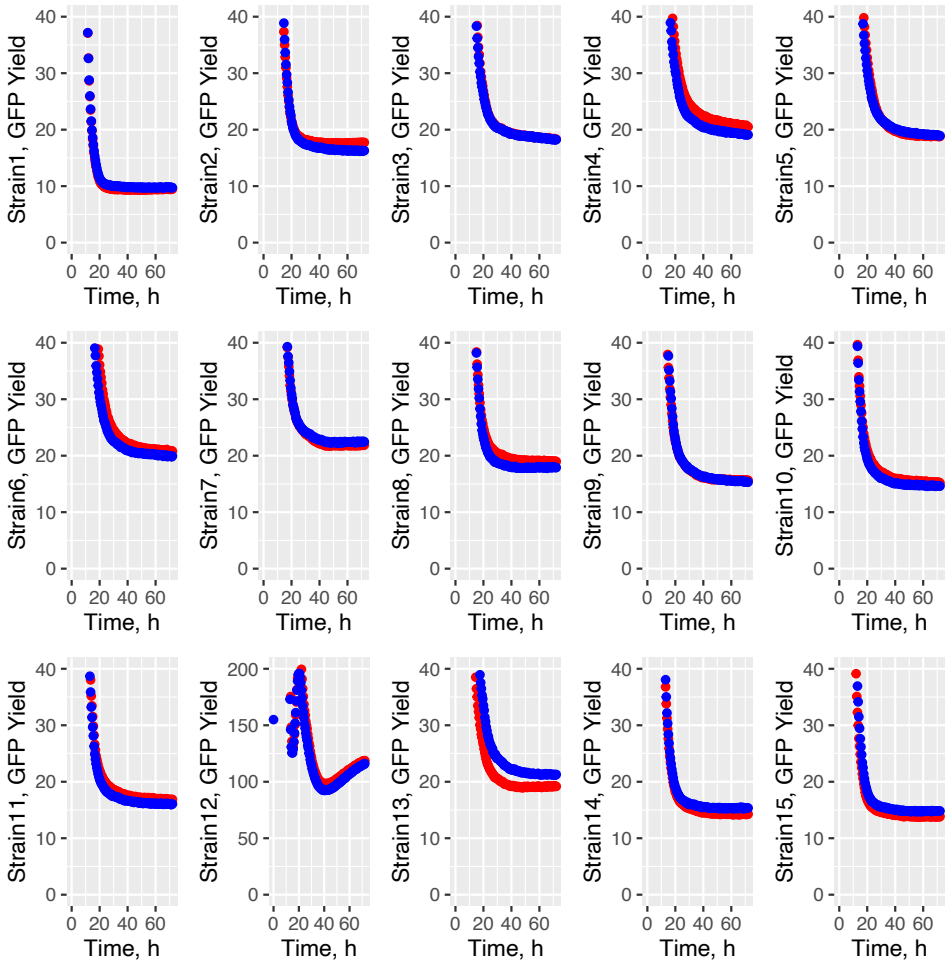


Fig. S9. GFP-fluorescence yield of the heme biosensor in heme production strains. Yield of GFP-Hb fluorescence per biomass in engineered strains. Strains: Strain1 - IMX581/empty vector; Strain2 - IMX581/GFP-Hbfusion; Strain3 - IMX581 *HEM15*/GFP-Hbfusion; Strain4 - IMX581 *HEM15 HEM14*/GFP-Hbfusion; Strain5 - IMX581 *HEM15 HEM14 HEM3*/GFP-Hbfusion; Strain6 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1* /GFP-Hbfusion; Strain7 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1* Δ*gcv2*/GFP-Hbfusion; Strain8 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1* Δ*faa4*/GFP-Hbfusion; Strain9 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1* Δ*ylr446w*/GFP-Hbfusion; Strain10 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1* Δ*fdh1*/GFP-Hbfusion; Strain11 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 ACH1*/GFP-Hbfusion; Strain12 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2*/GFP-Hbfusion; Strain13 - IMX581 *HEM15 HEM14* Δ*faa4*/GFP-Hbfusion; Strain14 - IMX581 Δ*ale1*/GFP-Hbfusion; Strain15 - IMX581 Δ*sfc1*/GFP-Hbfusion. Two transformants (replicates, blue and red) were used in the experiment and GFP-Hb fluorescence was monitored using a BioLector bioreactor.

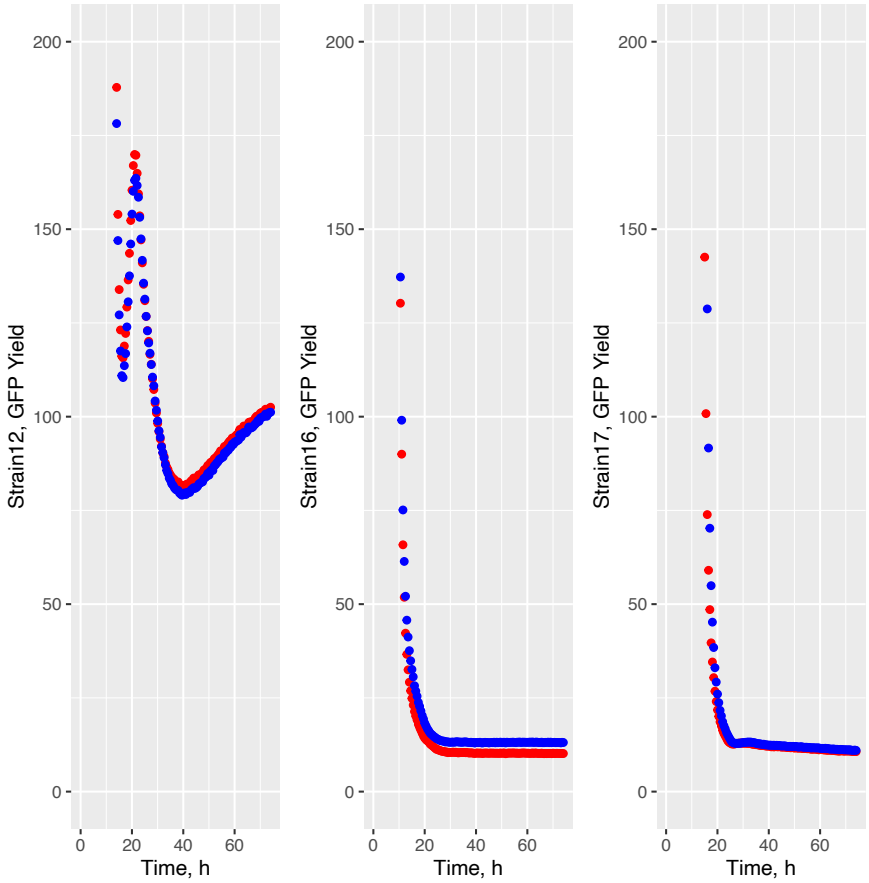


Fig. S10. GFP-fluorescence of the heme biosensor does not overlap with heme fluorescence. Strain12 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2*/GFP-Hbfusion; Strain16 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM13*; Strain17 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2 HEM13*. Two transformants (replicates, blue and red) were used in the experiment, and GFP-Hb fluorescence was monitored using a BioLector bioreactor.

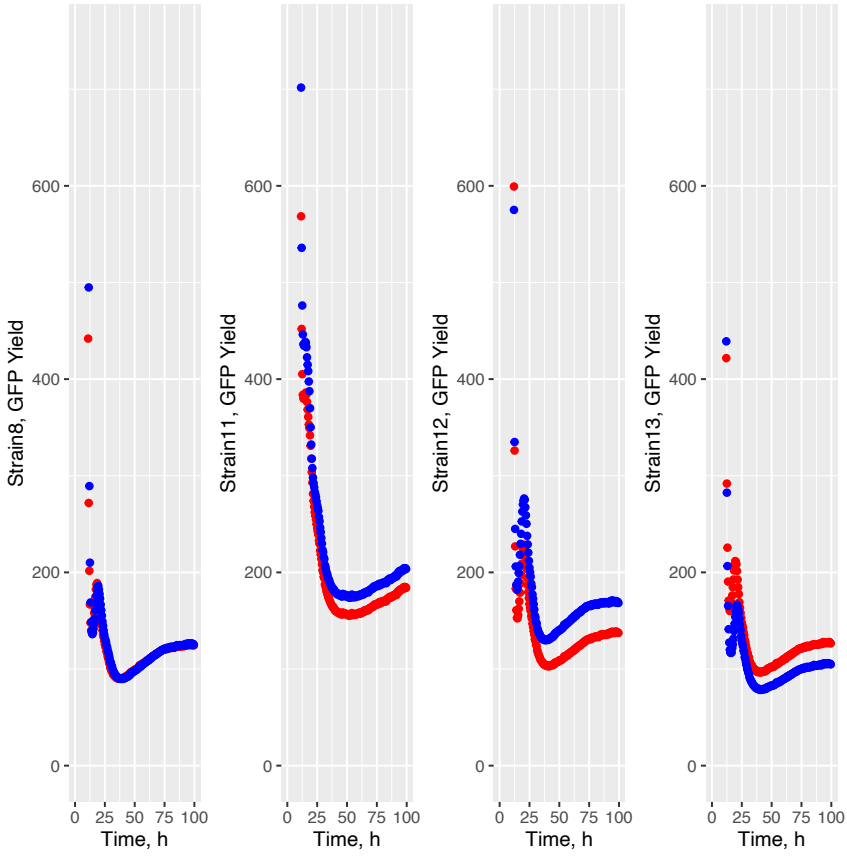


Fig. S11. GFP-fluorescence yield of the heme biosensor in heme production strains. Strain8 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2*/GFP-Hbfusion; Strain11 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2* Δ*hmx1 FET4*/GFP-Hbfusion; Strain12 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2* Δ*hmx1 ADH1*/GFP-Hbfusion; Strain13 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2* Δ*hmx1 ARH1*/GFP-Hbfusion. Two transformants (replicates, blue and red) were used in the experiment, and GFP-Hb fluorescence was monitored using a BioLector bioreactor.

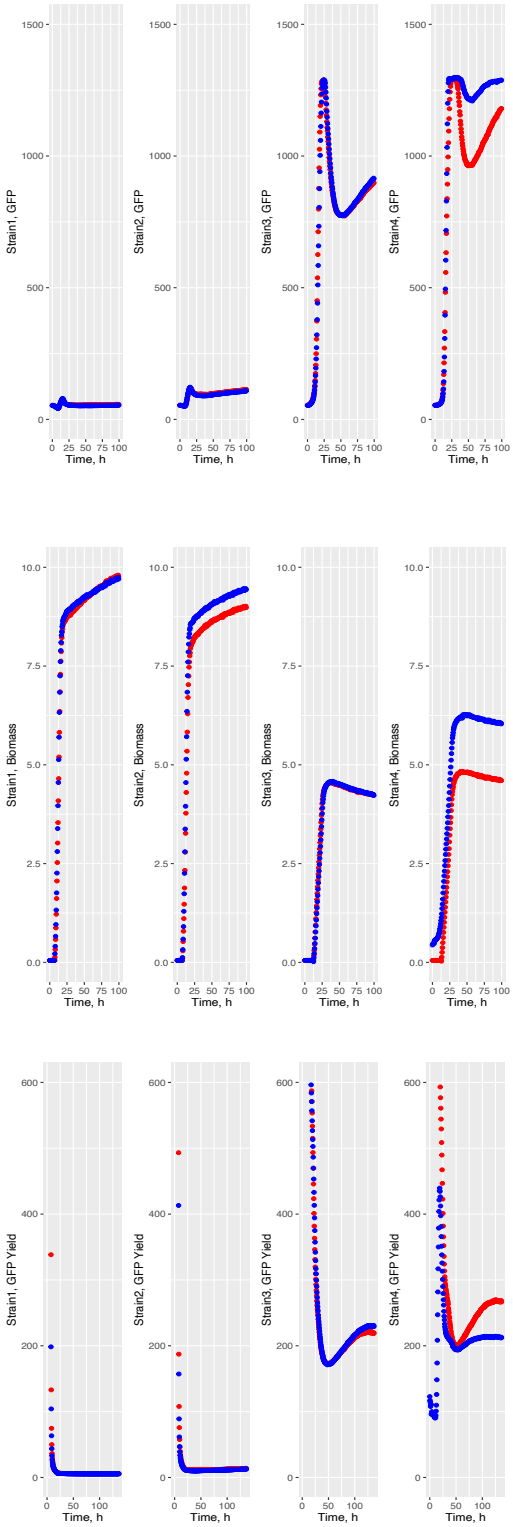
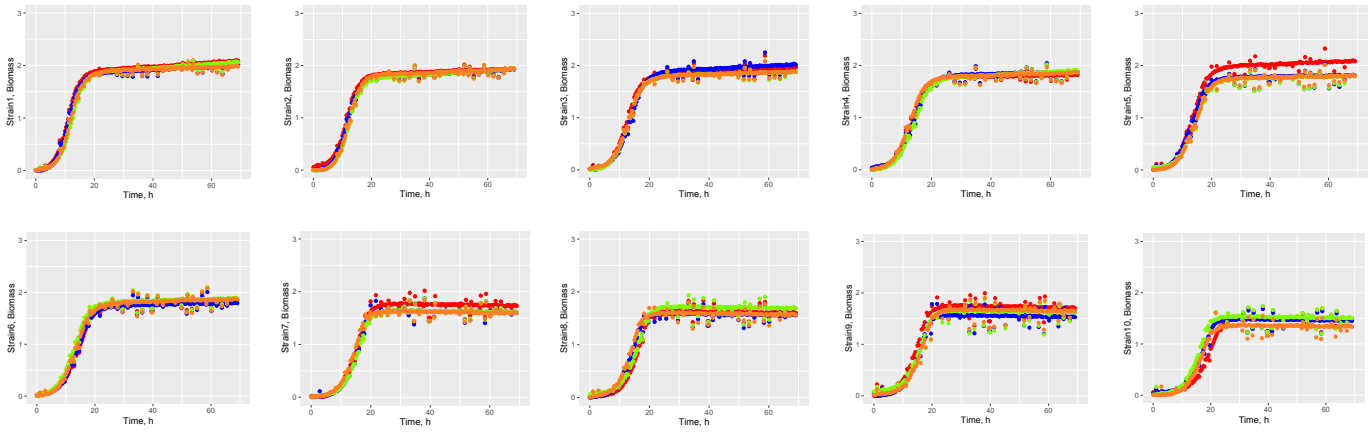


Fig. S12. Heme biosensor in heme production strains.
Strain1 - IMX581/empty vector; Strain2 - IMX581/GFP-Hbfusion; Strain3 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2* Δ*hmx1 FET4* Δ*gcv2 HEM1* carrying *HEM13* centromeric plasmid /GFP-Hbfusion; Strain4 - IMX581 *HEM15 HEM14 HEM3* Δ*shm1 HEM2* Δ*hmx1 FET4* Δ*gcv2 HEM1* Δ*gcv1* carrying *HEM13* expression cassette integrated into genome. Two transformants

(replicates, blue and red) were used in the experiment, and GFP-Hb fluorescence, biomass were monitored using a BioLector bioreactor. The biosensor GFP fluorescence yield was calculated as a ratio of GFP fluorescence to biomass.



Strain	Alternative name	Description
Strain 1	A	IMX581 [empty vector]
Strain 2	B	IMX581-HEM13
Strain 3	C	IMX581-HEM15-HEM13
Strain 4	D	IMX581-HEM15-HEM14-HEM13
Strain 5	E	IMX581-HEM15-HEM14-HEM3-HEM13
Strain 6	F	IMX581-HEM15-HEM14-HEM3-Δshm1-HEM13
Strain 7	G	IMX581-HEM15-HEM14-HEM3-Δshm1-HEM2-HEM13
Strain 8	H	IMX581-HEM15-HEM14-HEM3-Δshm1-HEM2-Δhmx1-FET4-HEM13
Strain 9	I	IMX581-HEM15-HEM14-HEM3-Δshm1-HEM2-Δhmx1-FET4-Δgcv2-HEM13
Strain 10	J	IMX581-HEM15-HEM14-HEM3-Δshm1-HEM2-Δhmx1-FET4-Δgcv2-HEM1-HEM13

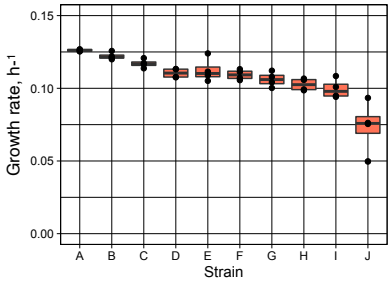


Fig. S13. Biomass and growth rate of heme production strains estimated in a BioLector. Four replicates of each strain were used in the experiment. Biomass was measured every 30 min. The growth rate, h^{-1} was estimated for all strains in the exponential growth phase (between 5 and 18 hours of growth).

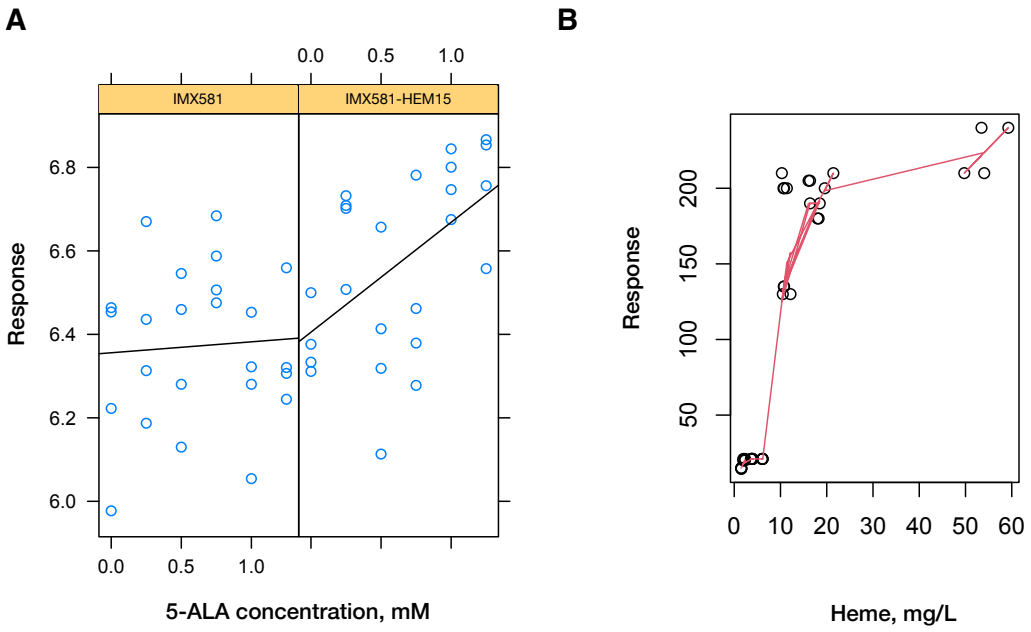


Fig. S14. Regression analysis (quantile regression with shape constraints) of the Heme-LBB biosensor to external 5-ALA supplementation and internal heme supply through overexpression of modeling gene targets. A) Heme-LBB in the IMX581 strain showed no response (Multiple R-squared: 0.09172, *p*=0.5785); Heme-LBB in the IMX581-HEM15 strain showed the response (Multiple R-squared: 0.323, *p*=0.04631; B) Heme-LBB response was plotted against heme concentration obtained by the extraction from the cultures of constructed strains. The sigmoid-like response was observed.

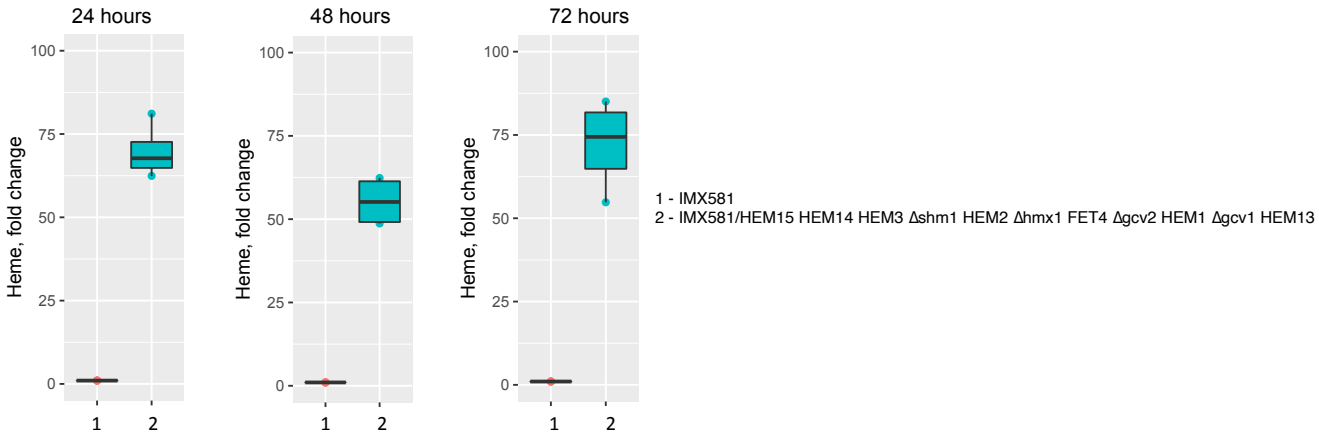


Fig. S15. Heme production of *S. cerevisiae* strains carrying the expression cassettes and deletion of modeling gene targets.
Heme production was analyzed at 24, 48 and 72 hours in SD ura- supplemented with 2% glucose, 100 mM glycine and 100 μM Fe^{3+}. Heme was extracted from 8 OD_{600} of cells. Four biological replicates (transformants) were used in the experiment.

SI References

1. R. Ferreira *et al.*, Model-Assisted Fine-Tuning of Central Carbon Metabolism in Yeast Through dCas9-Based Regulation. *ACS Synth. Biol.* **8(11)**, 2457-2463 (2019).
2. H. Lu *et al.*, A Consensus *S. cerevisiae* Metabolic Model Yeast8 and Its Ecosystem for Comprehensively Probing Cellular Metabolism. *Nat. Commun.* **10(1)**, 3586 (2019).
3. H.S. Choi, S.Y. Lee, T.Y. Kim, H.M. Woo, In silico identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* **76(10)**, 3097-105 (2010).
4. B. Sánchez, F. Li, H. Lu, E. Kerkhoven, J. Nielsen, (2018, April 16). SysBioChalmers/yeast-GEM: yeast 8.0.1 (Version v8.0.1). Zenodo.
5. N.E. Lewis *et al.*, Omic Data From Evolved *E. Coli* Are Consistent With Computed Optimal Growth From Genome-Scale Models. *Mol. Syst. Biol.* **6**, 390 (2010).
6. L. Heirendt *et al.*, Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14(3)**, 639-702 (2019).
7. R. Mans *et al.*, CRISPR/Cas9: A Molecular Swiss Army Knife for Simultaneous Introduction of Multiple Genetic Modifications in *Saccharomyces Cerevisiae*. *FEMS Yeast Res.* **15(2)**, fov004 (2015).
8. M.D.W. Piper, S.P. Hong, T. Eissing, P. Sealey, I.W. Dawes, Regulation of the Yeast Glycine Cleavage Genes Is Responsive to the Availability of Multiple Nutrients. *FEMS Yeast Res.* **2(1)**, 59-71 (2002).
9. O. Protchenko, C.C. Philpott, Regulation of intracellular heme levels by *HMX1*, a homologue of heme oxygenase, in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278(38)**, 36582-7 (2003).
10. J.K. Michener, J. Nielsen, C.D. Smolke, Identification and treatment of heme depletion attributed to overexpression of lineage of evolved P450 monooxygenases. *Proc. Natl. Acad. Sci. U S A.* 109(47), 19504-9 (2012).
11. S. Chakane, Towards New Generation of Hemoglobin-Based Blood Substitutes. Department of Chemistry, Lund University (2017).
12. A.A. Komar, A. Kommer, I.A. Krasheninnikov, A.S. Spirin, Cotranslational heme binding to nascent globin chains. *FEBS Lett.* **326(1-3)**, 261-263 (1993).
13. A.A. Komar, A. Kommer, I.A. Krasheninnikov, A.S. Spirin, Cotranslational folding of globin. *J. Biol. Chem.* **272(16)**, 10646-10651 (1997).

Dataset S1. Metabolic reactions and genes selected for modification to increase heme biosynthesis by using the genome-scale model Yeast8.

Dataset S2. Genes selected for modification to increase heme biosynthesis by using the enzyme-constrained genome-scale model ecYeast8.

Dataset S3. A list of genes predicted to preserve strain viability and to increase heme biosynthesis, sequentially accumulated by our algorithm using the enzyme-constrained genome-scale model ecYeast8.