Supplementary Information

ENGINEERING FATTY ACID SYNTHASES FOR DIRECTED POLYKETIDE PRODUCTION

Jan Gajewski[a], Floris Buelens[b], Sascha Serdjukow[c], Melanie Janßen[a], Niña Cortina[a], Helmut Grubmüller[b]* and Martin Grininger[a,c]*

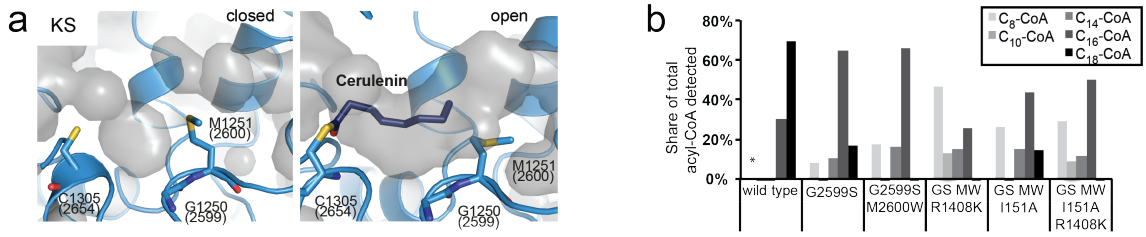
[a]*Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Cluster of Excellence „Macromolecular Complexes", Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt am Main, Germany.*

[b]*Max-Planck-Institute for Biophysical Chemistry, Department of Theoretical and Computational Biophysics, Am Fassberg 11, 37077 Goettingen, Germany.*

[c]*Max-Planck-Institute of Biochemistry, Project Group Biological Chemistry, Am Klopferspitz 18, D-82152 Martinsried near Munich, Germany.*

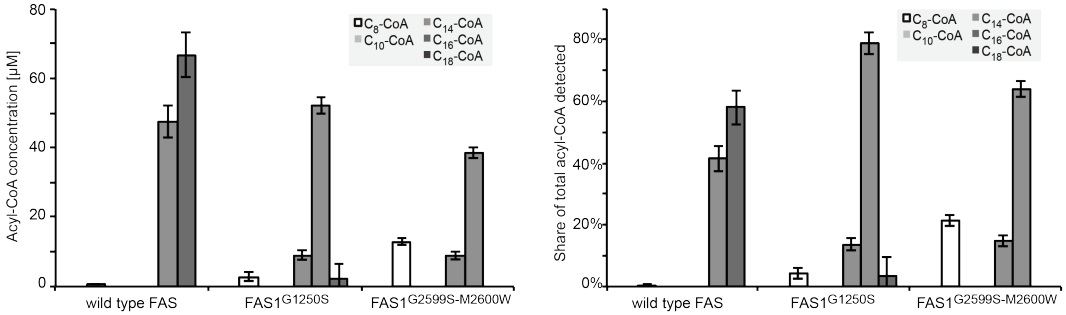
Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary Results



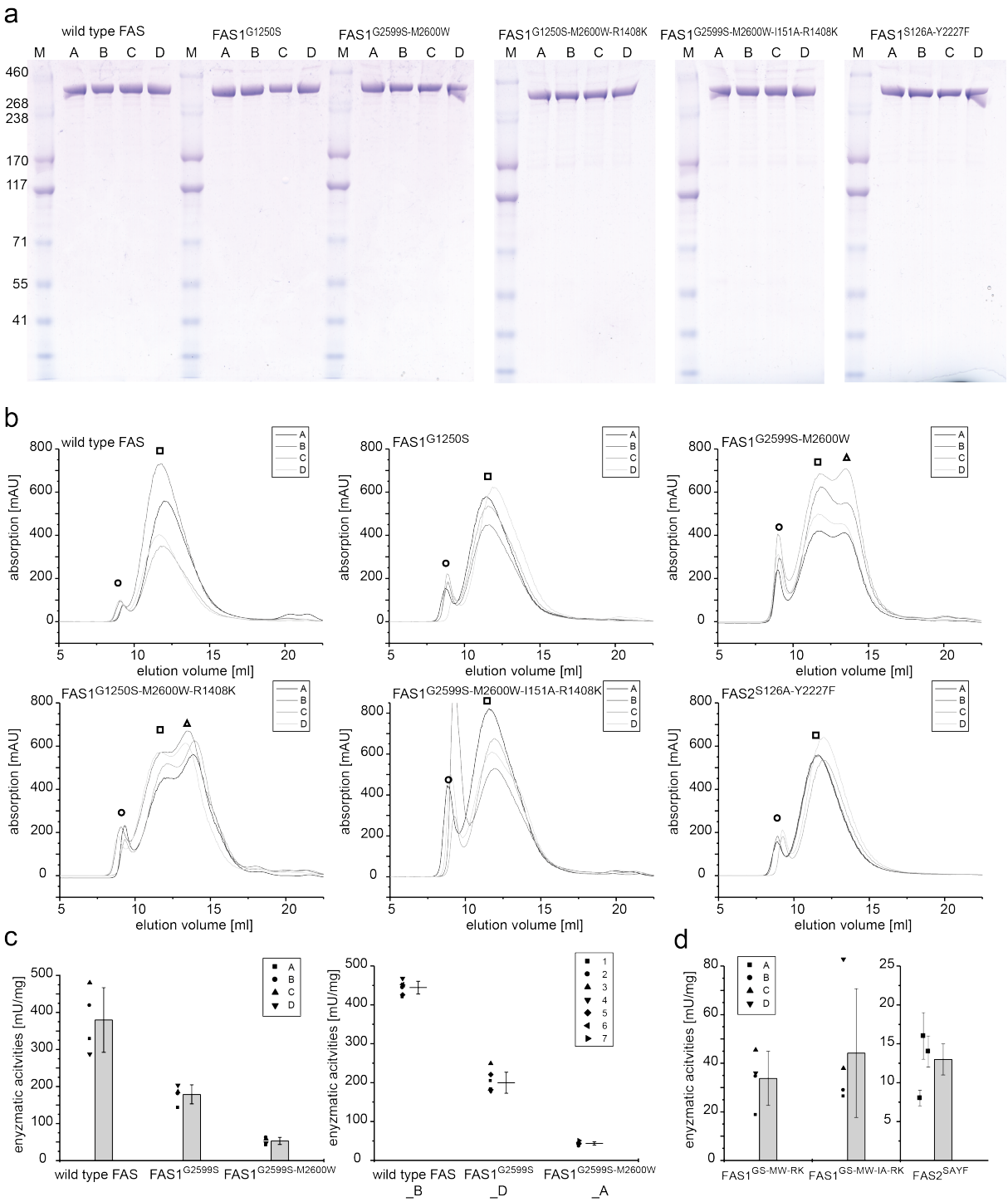
Supplementary Figure 1: Engineering of FAS in KS domain. (a) KS domain of *S. cerevisiae* FAS (*C. ammoniagenes* FAS numbering in parentheses). By essentially populating two conformations, one directed into the KS channel (closed state), leaving the outer part of the channel accessible for substrate binding, the other rotated aside (open state), giving access also the distal portion of the KS channel, residue M1251 has been proposed to act as a gatekeeper. In the closed state (PDB code 2UV8), M1251 is pointing into the KS channel, which makes only the outer part accessible (the binding channel is indicated as a gray cloud). The open state (PDB code 2VKZ) is induced by the inhibitor cerulenin mimicking a bound acyl chain. M1251 is positioned on a flexible loop, and mutation G1250S aims to restrict M1251 further in its flexibility freezing the closed state. **(b)** Product distributions of engineered FAS1. Acyl-CoA output spectra for FAS1 variants in percentages of the specific acyl-CoA in total acyl-CoA (see output in micromolar concentrations in Fig. 2a). FAS1 variants are abstracted by their specific mutation, so that for example "G2599S" equals FAS1^{G2599S}. Also, FAS1$^{G2599S\text{-}M2600W}$ with additional mutations R1408Q/A/G were tested but didn't show any products. Bars marked with (*) are based on values obtained by extrapolation (values outside of calibration range). Data refer to means of technical replicates (n = 1; three measurements). For more information on statistics, see Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314



Supplementary Figure 2: Biological repeatability of FA synthesis. Biological repeatability was determined in a separate experiment on wild type FAS and KS-mutated FAS. Each construct was prepared in 4 independent biological samples (n=4; 4 independent expressions starting with different transformations of cells). While biological errors are low, FA chain length distribution is generally shifted slightly to shorter FA, as compared to data presented in Fig. 2c and d. As the sensitivity of the FA chain length distribution on acyl-CoA substrate concentrations has been observed before[3], as well as we see an influence on substrate concentrations also on the overall lactone yield (see Fig. 2c and d), we speculate that particularly variations in substrate concentrations might account for the shift in FA chain length spectra. We emphasize that the claims made in the study do not rest upon absolute data, but the comparison of data of mutants. Acyl-CoA output spectra for FAS1 variants are depicted in molar concentrations in the test tube (left panel), and percentages of the specific acyl-CoA in total acyl-CoA (right). Means and standard deviations are given. For more information on statistics, see Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314

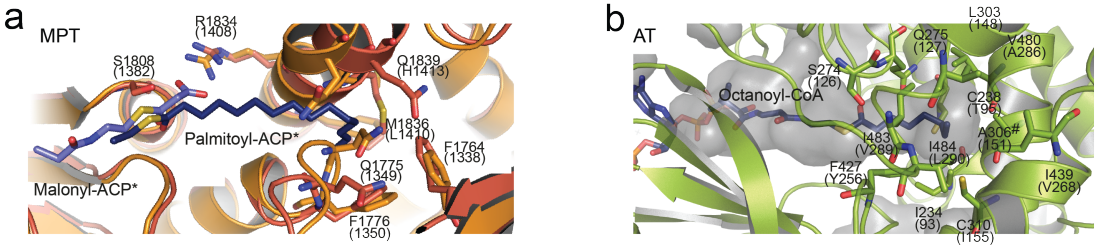


Supplementary Figure 3: Preparation and enzymatic activities of FAS1 and FAS2 constructs. Each construct was prepared from 4 independent expressions, termed A, B, C and D (n=4; 4 independent expressions starting with different transformations of cells). **(a)** 3-8% Tris/Acetate SDS-PAGE gel (NuPage, Invitrogen) of purified proteins. **(b)** SEC profiles of FAS constructs. SEC was performed on a Superose 6 Tricorn 30/100 (GE Healthcare). Peaks correspond to higher oligomeric (aggregated) (○), hexameric (□) and monomeric (△) protein. SEC was used as a last purification step and pooled fraction covering of the hexameric proteins (elution volume 10-12 ml) were used for activity assays and FA and 6-HHP synthesis. **(c)** Enzymatic activity of wild type FAS and KS-mutated FAS1 constructs. Left panel shows biological error. Individual data points (n=4), means and standard deviations are given. Technical repeatability is shown in the right panel for one sample of each construct measured seven times (error in standard deviation, each data point represents one measurement). Activities are shown in mU/mg of protein (1 U was defined as the incorporation of 1 µmol of malonyl-CoA per minute). Activities were determined by monitoring NADPH consumption. **(d)** Enzymatic activity for transferase-mutated FAS1

Nature Chemical Biology: doi:10.1038/nchembio.2314

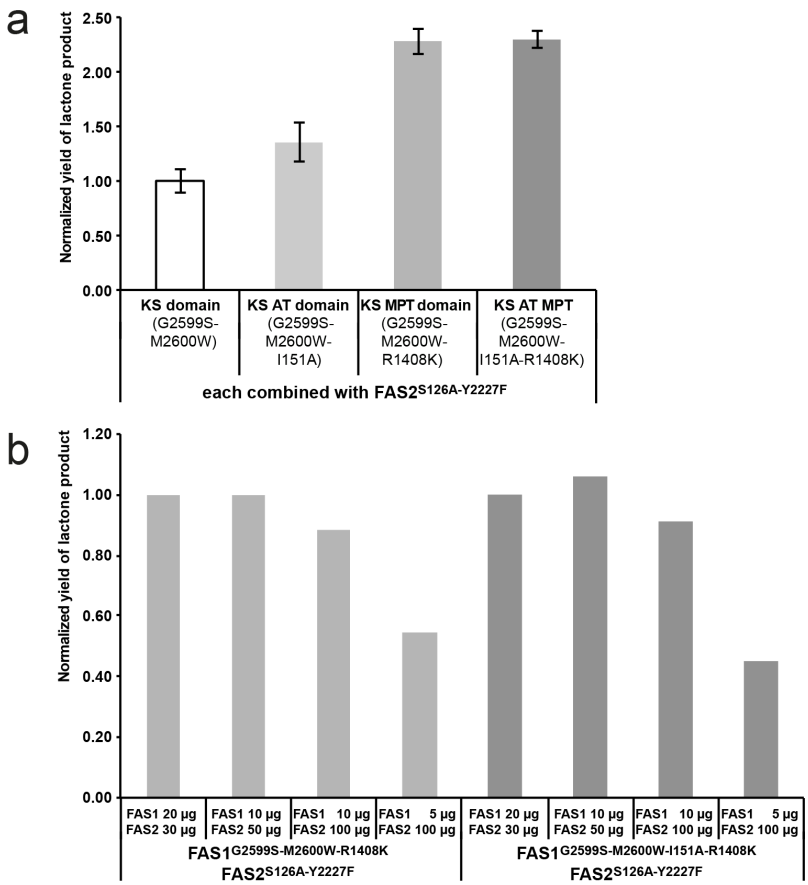
constructs, and the FAS2 construct FAS2$^{S126A-Y2227F}$. For FAS1 constructs, individual data points (n=4), means and standard deviation are given. For FAS2$^{S126A-Y2227F}$, individual data points (n=3) determined in technical replicates (error in standard deviation), means and standard deviation are given. Activities are shown in mU/mg of protein (1 U was defined as the incorporation of 1 µmol of malonyl-CoA per minute). For FAS1 constructs, activities were determined by monitoring NADPH consumption, while for FAS2$^{S126A-Y2227F}$ product 6-HHP formation was recorded. Figure labeling with mutations abbreviated as following; G2599S, GS; M2600W, MW; I151A, IA; R1408K, RK; S126A-Y2227F, SAYF. For more information on protein preparation, see Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314



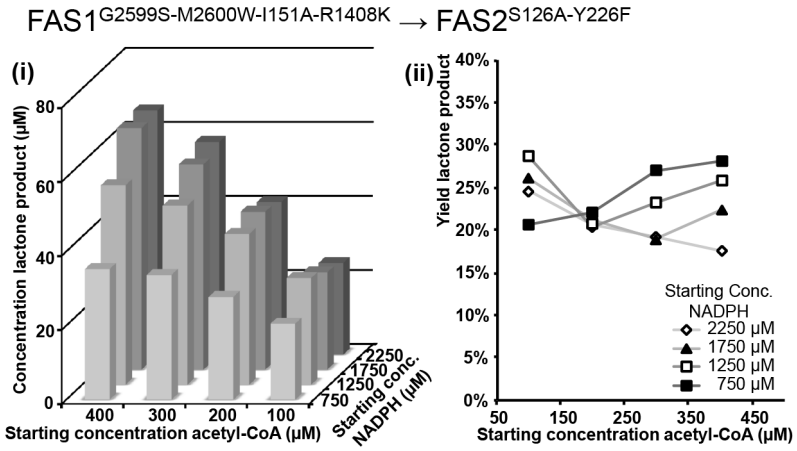
Supplementary Figure 4: Engineering of FAS in transferase domains. (a) Structures from molecular dynamics simulations of the *S. cerevisiae* FAS MPT domain (based on PDB code 2UV8) for the binding of palmitoyl (C_{16}; ligand in dark blue, structure in red) and malonyl (ligand in light blue, structure in orange). Selected side chains are shown in stick representation, such as the active site serine (S1808, *S. cerevisiae* numbering; *C. ammoniagenes* FAS numbering in parentheses), the arginine involved in malonyl stabilization (R1834), several rests (F1764, F1776, F1831, M1836) forming a hydrophobic pocket as postulated in literature, and candidate mutation sites Q1775 and Q1839 (see Fig. 2b). Free energy calculations quantifying the effect of mutation R1834K yielded an estimated weakened malonyl-CoA binding, by $\Delta\Delta G$ = +14.2 ± 1.1 kJ/mol. **(b)** Molecular dynamics simulations of the *S. cerevisiae* FAS AT domain (based on PDB code 2UV8) mutated in I306A (*S. cerevisiae* numbering; *C. ammoniagenes* FAS numbering in parentheses) for the binding of C_8-CoA. C_8-CoA is shown in stick representation in blue with its nucleotide moiety extending to the left outside of the frame. Of ten 400 ns independent unbiased molecular dynamics simulations, a stable pre-catalytic conformation for C_8-CoA was maintained in nine cases. Residues involved in forming the putative hydrophobic binding pocket are also shown as stick representation (I234, C238, Q275, L303, A306, C310, F427, I439, V480, I483, I484). The I306A mutation has little effect on acetyl-CoA binding affinity (calculated $\Delta\Delta G$ value of -0.4 ± 1.6 kJ/mol). The active site S274 is marked with (*); the mutated A306 with (#).

Nature Chemical Biology: doi:10.1038/nchembio.2314



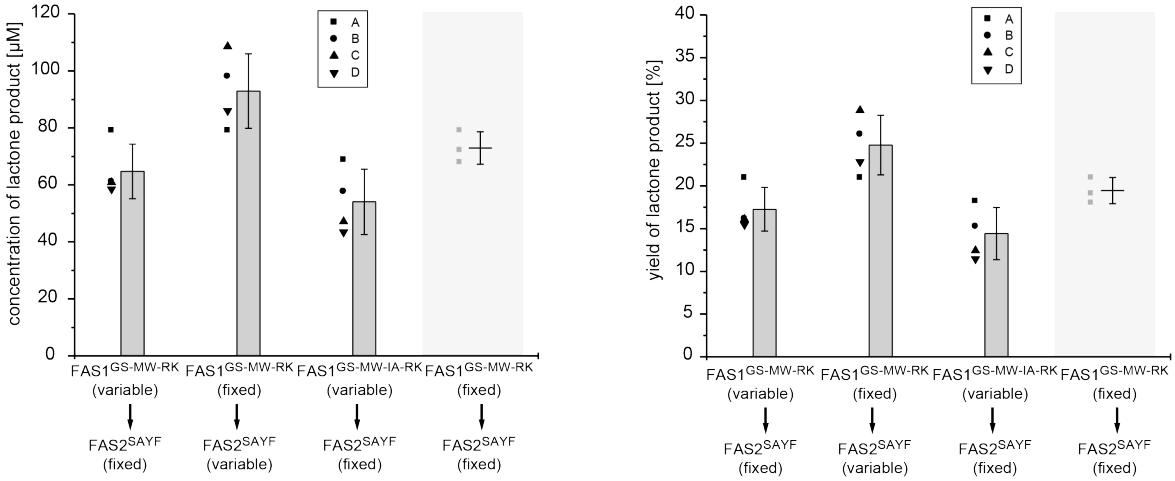
Supplementary Figure 5: 6-HHP out output under variation of construct FAS1 as well as concentrations of FAS1 and FAS2. (a) 6-HHP output with varied constructs used as module 1 ($FAS1^{G2599S-M2600W}$, $FAS1^{G2599S-M2600W-I151A}$, $FAS1^{G2599S-M2600W-R1408K}$ and $FAS1^{G2599S-M2600W-I151A-R1408K}$). Yields are normalized to the output of the coupled reaction using $FAS1^{G2599S-M2600W}$. Data were recorded in technical triplets (n=1; three measurements), and the error bars are indicating standard deviations of technical repeatability. **(b)** Scan of 6-HHP output for varying amounts of module 1 and module 2 (standard 100 µL scale). Output for each FAS1 / FAS2 combination was normalized to its yield under standard conditions of 20 µg of FAS1 and 30 µg of FAS2. No technical replicates were determined. For more information on statistics, see Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314



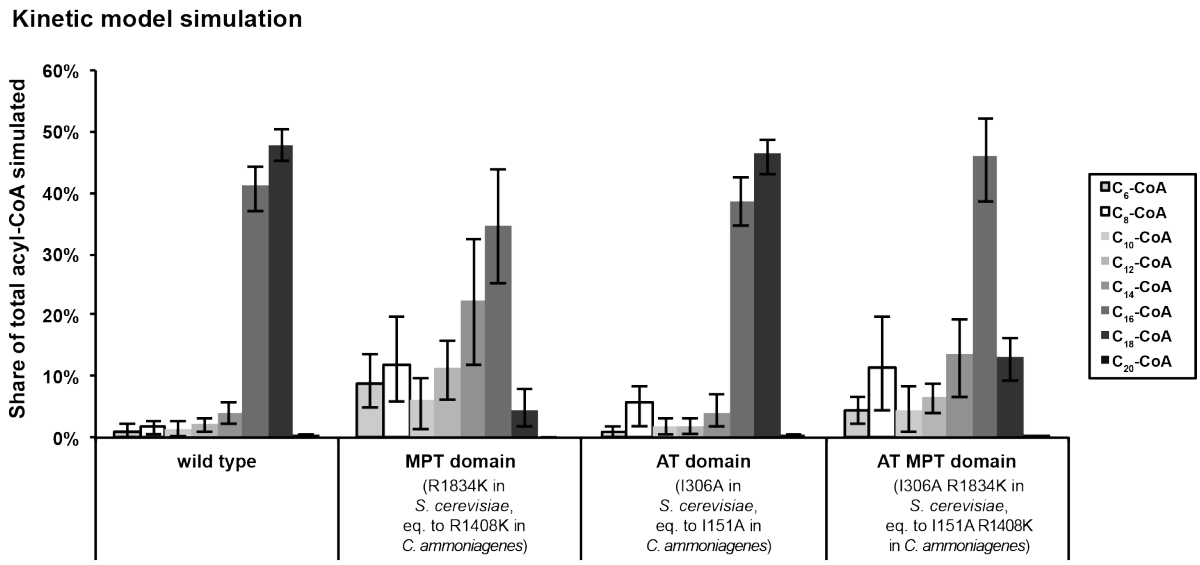
Supplementary Figure 6: Coupled assay with the FAS1$^{G2599S-M2600W-I151A-R1408K}$ → FAS2$^{S126A-Y2227F}$ reaction sequence. Sequential synthesis of 6-HHP with FAS1$^{G2599S-M2600W-I151A-R1408K}$ → FAS2$^{S126A-Y2227F}$ as influenced by substrate concentrations. Synthesis follows similar trends as observed in the better performing FAS1$^{G2599S-M2600W-R1408K}$ → FAS2$^{S126A-Y2227F}$ reaction sequence (see Fig. 2c and d). No technical replicates were determined. For more information on statistics, see Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314



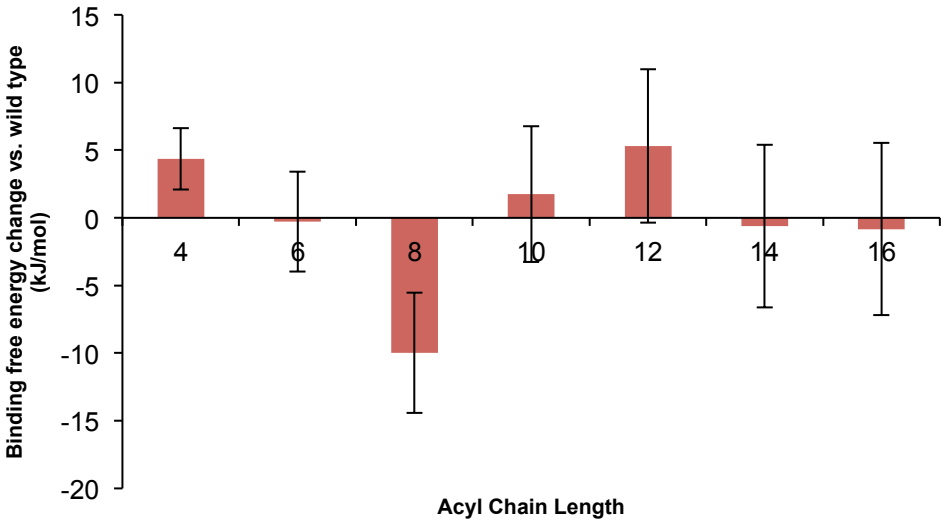
Supplementary Figure 7: Biological repeatability of lactone synthesis. Biological repeatability of the coupled reaction was determined in a separate experiment, probing the influence of the first (FAS1) and the second (FAS2) module on 6-HHP output. See labeling of the figure, for details on the specific combinations tested. Each constructs was prepared in 4 independent biological samples (n=4; 4 independent expressions starting with different transformations of cells). Outputs of coupled synthesis are displayed in final product concentrations in the reaction tube (left panel) and in yields relative to the limiting substrate NADPH (right). The technical repeatability was covered in an additional experiment (grey background). Means and standard deviations are given. Individual data points are shown for clarity. Mutations are abbreviated as following; G2599S, GS; M2600W, MW; I151A, IA; R1408K, RK; S126A-Y2227F, SAYF. For more information on statistics, see Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314



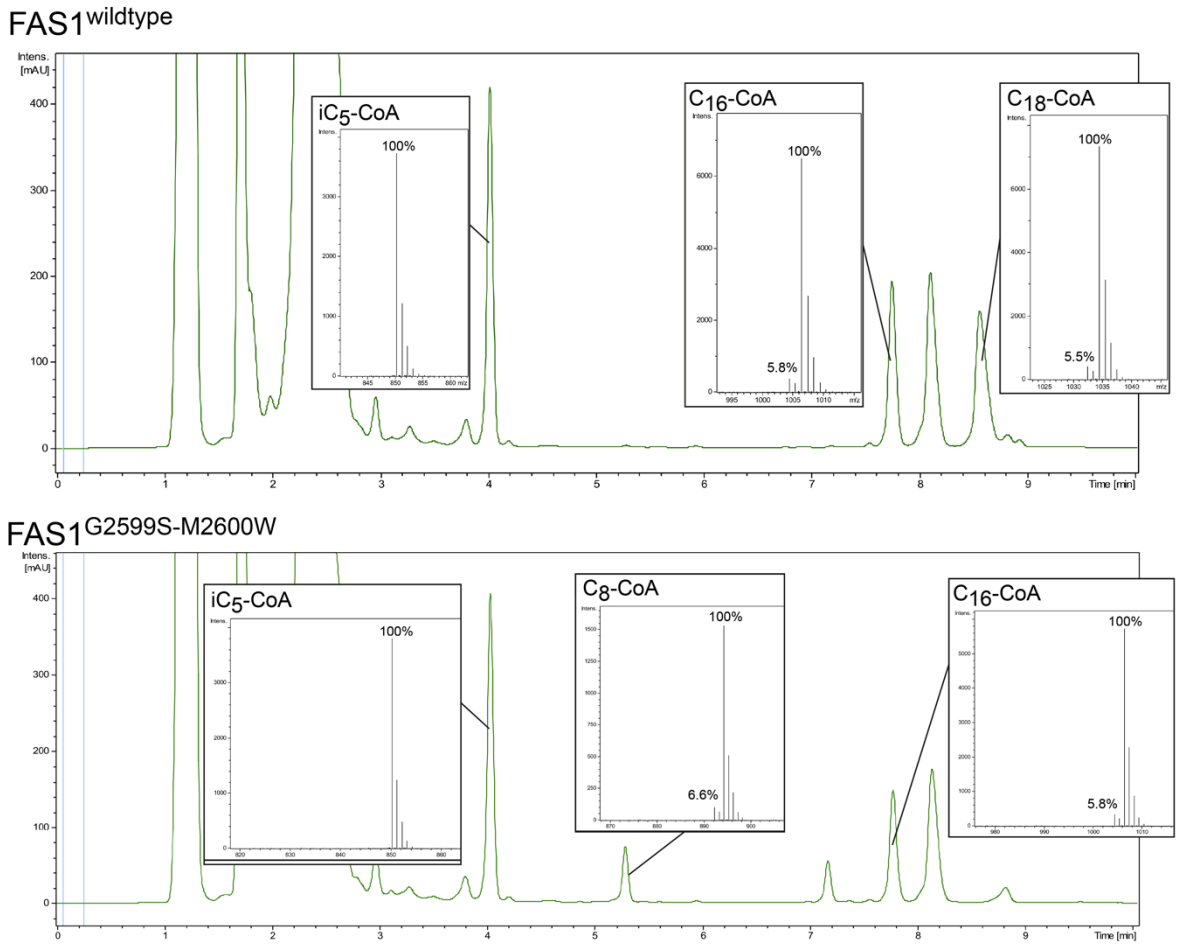
Supplementary Figure 8: Computational output spectra of engineered FAS1 from the initial kinetic model. FAS1 variants are abstracted by the domain names and mutations. Simulation of product spectra with the kinetic model for wild type FAS, MPT-mutated FAS (R1834K, *S. cerevisiae* FAS numbering), AT-mutated FAS (I306A, *S. cerevisiae* FAS numbering) and AT-MPT double mutated FAS (I306A-R1834K). Average output spectra from the kinetic model with at least 3000 model runs for each chart. In applying the computational model to results from the transferase mutations in *C. ammoniagenes* FAS, our model correctly reproduced experimental data that were not used in the model construction; namely (i) the broad shift towards shorter products observed as a result of a malonyl throughput-restricting MPT mutation, (ii) the increased C_8-CoA yield as a result of the AT mutation shaping a novel binding channel, (iii) the non-additive nature of the MPT and AT mutations, and (iv) the slight shift towards longer products that accompanies the AT mutation (see Fig. 2a).

Nature Chemical Biology: doi:10.1038/nchembio.2314



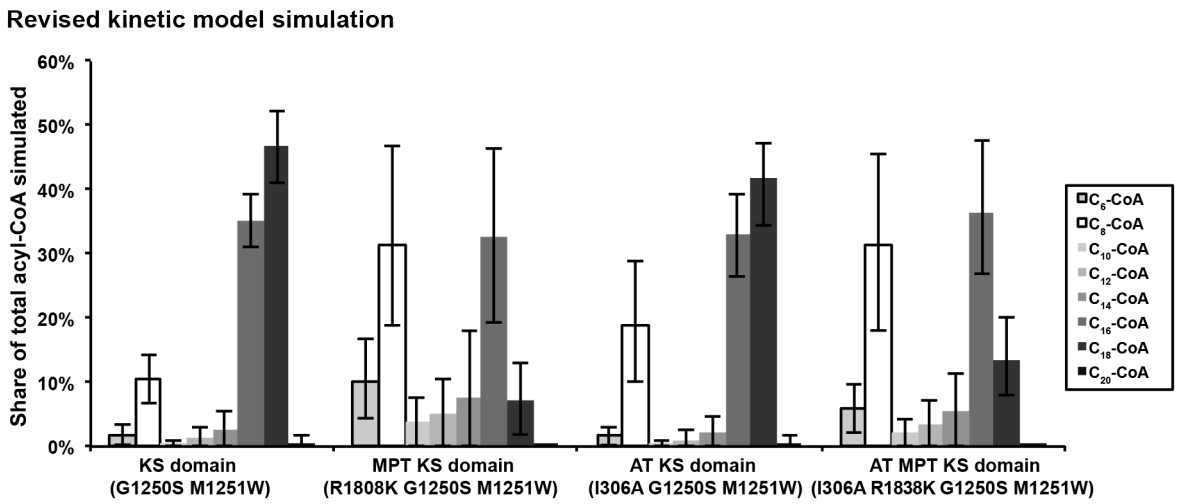
Supplementary Figure 9: Binding free energy change for acyl chains C_4-C_{16} in the mutant FAS1$^{G1250S\text{-}M1251W}$ with respect to wild type, as calculated from molecular dynamics simulations. Alternative representation of the data shown in Fig. 2f, with relative binding free energy calculated for each chain length as a sum over the preceding reactions: $\Delta\Delta G_n = \sum_{i=4,6,\ldots}^{n}\left(\Delta G_{mut}^{i-2\rightarrow i} - \Delta G_{WT}^{i-2\rightarrow i}\right)$. This representation shows that binding of the C_8 intermediate to the KS site is calculated to be enhanced by -10.0 kJ/mol as a result of the double mutation, while the effect on other acyl chain lengths is modest. The increase in uncertainty with increasing chain length reflects summation of errors.

Nature Chemical Biology: doi:10.1038/nchembio.2314



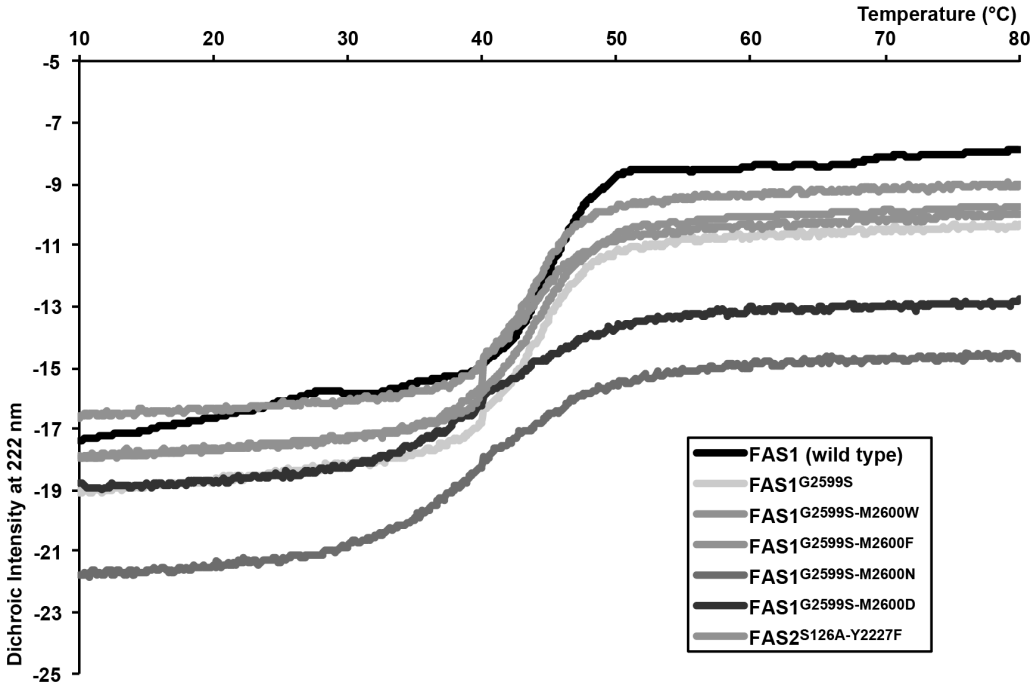
Supplementary Figure 10: Testing increased decarboxylation as a reason for shorter fatty acids. HPLC chromatograms (UV absorption at 260 nm) of product assays with wild type FAS and FAS1$^{G2599S-M2600W}$, in which ^{13}C-labelled acetyl-CoA was used as a starter. The exemplary mass spectra of the product peaks of C_8-CoA, C_{16}-CoA and C_{18}-CoA indicate that products mostly derive from the labeled ^{13}C-acetyl-CoA having a molecular mass of M+2 in comparison to the expected ^{12}C-acetyl-CoA reactions (intensities of these highest peaks were set to 100%). A small fraction, however, shows regular mass M, which indicates that in these cases the acyl-CoA synthesis started from an unlabeled acetyl-CoA. It can be assumed that either a low basal decarboxylation activity of the enzymes or an acetyl-CoA contamination of substrates is underlying this observation. The fractions of unlabeled products in FAS1$^{G2599S-M2600W}$ are, however, not significantly different from those found in wild type FAS, demonstrating that mutations in FAS1$^{G2599S-M2600W}$ do not induce malonyl decarboxylation. Mass spectra are also shown for iC_5-CoA, an internal standard (missing a M+2 peak).

Nature Chemical Biology: doi:10.1038/nchembio.2314



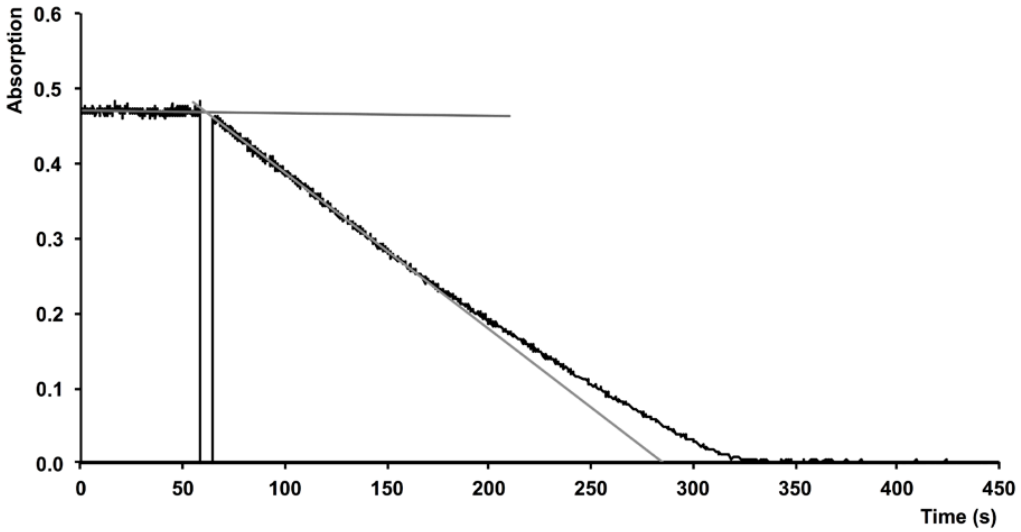
Supplementary Figure 11: Computational output spectra of engineered FAS1 with combined transferase KS mutants from the revised kinetic model. FAS1 variants are abstracted by the domain names and mutations. Data are given in percentages of molar concentrations of specific acyl-CoA in total acyl-CoAs. Each chart shows average output spectra from the kinetic model with at least 3000 model runs. We explicitly note that the agreement between experimental data and model results for the KS domain mutation arises by construction, by enforcing *a posteriori* agreement with experimental data, as discussed in the supplementary text. For more information see Supplementary Note 2.

Nature Chemical Biology: doi:10.1038/nchembio.2314



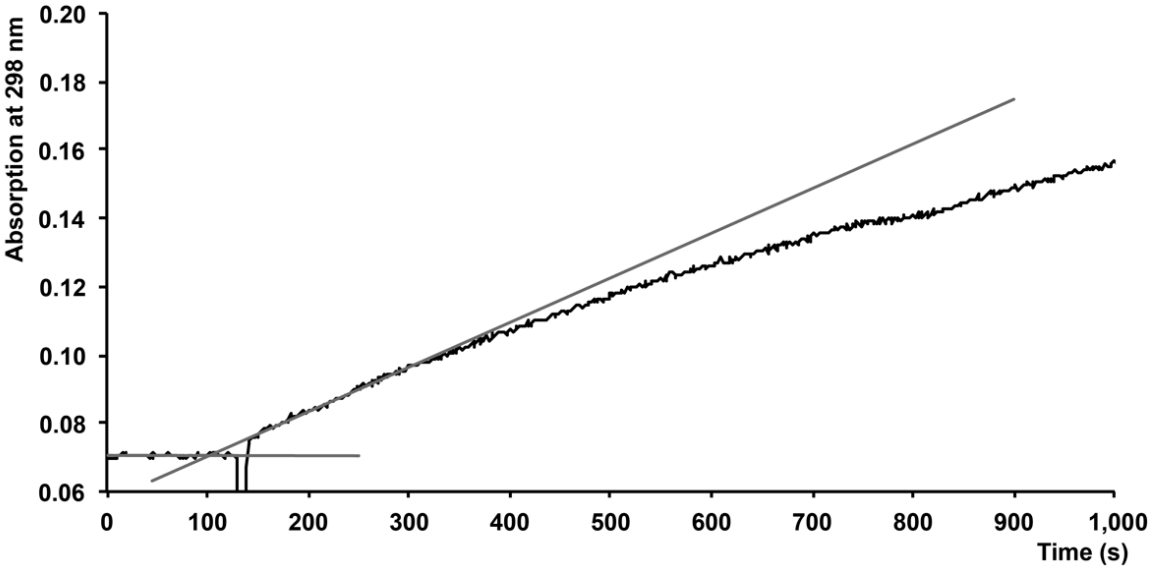
Supplementary Figure 12: Melting curves of selected FAS constructs. Melting curves were determined by UV-circular dichroism measurements at 222 nm (α-helical range) from 10 °C to 80 °C (with an increase of 60 °C per hour) using a dichrograph Jasco J-715 equipped with a Jasco Peltier element (PFD 350S/350L) for temperature control. Protein solutions of around 500 nM in buffer (200 mM NaH2PO4/Na2HPO4, pH = 7.2) were probed in quartz cells with a 0.1 cm path length. The melting temperatures and concentrations of the protein solution are listed in Supplementary Table 2. For more information on the integrity of the protein and the repeatability of the protein purifications, see also Supplementary Fig. 3 and Online Methods.

Nature Chemical Biology: doi:10.1038/nchembio.2314



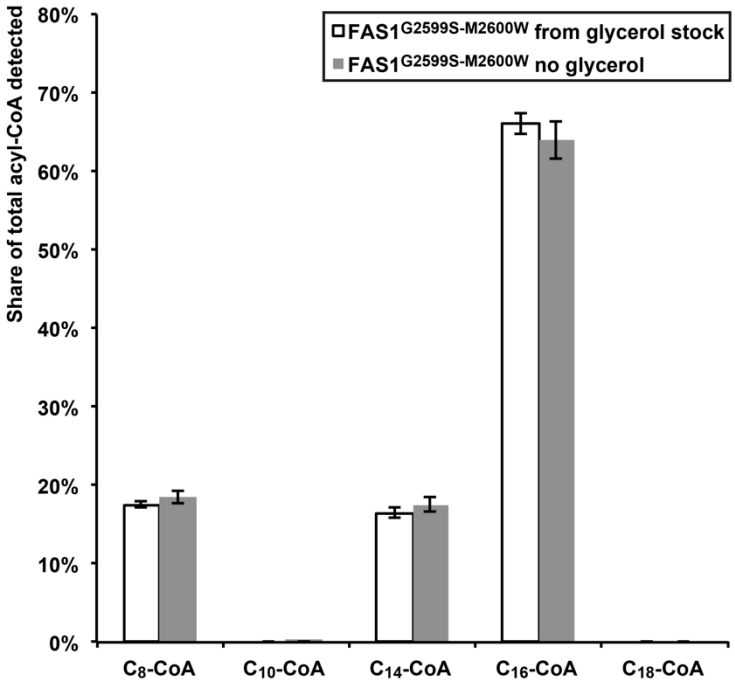
Supplementary Figure 13: Exemplary result from activity assay of module 1. NADPH decrease was monitored at 334 nm over time. After a period of constant absorption, the reaction is started by the addition of malonyl-CoA. Linear trends for both slopes, before and after the reaction, are used for are used for the determination of enzymatic activity. A value of 6.0 mM^{-1} cm^{-1} was used as extinction coefficient $\varepsilon_{334\ nm}$[6].

Nature Chemical Biology: doi:10.1038/nchembio.2314



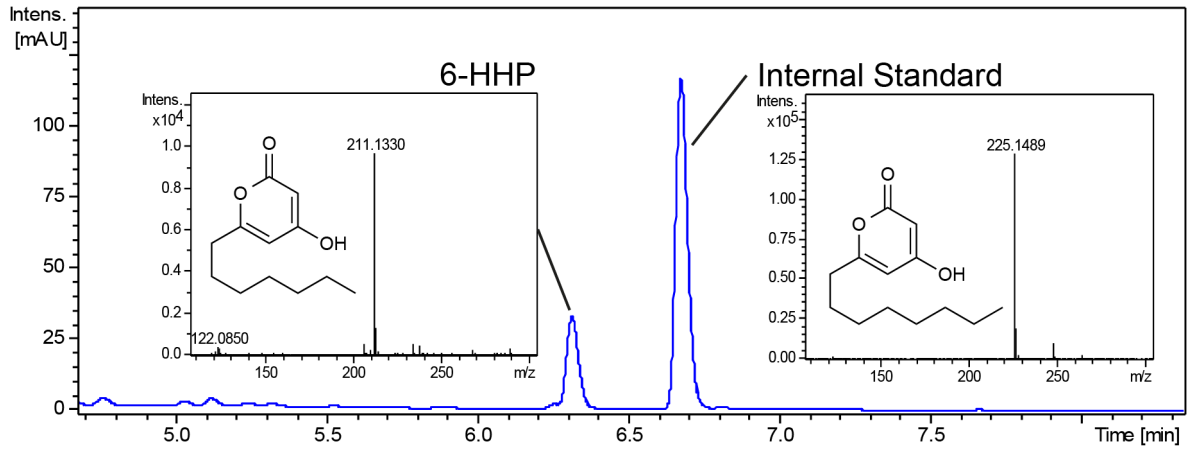
Supplementary Figure 14: Exemplary result of activity assay of module 2. In this assay, the formation of the product 6-HHP was recorded at 298 nm over time. After a period of constant absorption, the reaction is started by the addition of malonyl-CoA. Linear trends for both slopes, before and after the reaction, are used for the determination of enzymatic activity.

Nature Chemical Biology: doi:10.1038/nchembio.2314



Supplementary Figure 15: Glycerol influence on product spectrum of module 1. Protein used in FA and 6-HHP was stored in glycerol leading to final glycerol concentrations of up to 1.2% in the reactions solutions. Product spectra of FAS1$^{G2599-M2600W}$ as freshly prepared protein and as received from a glycerol stock do not show differences in FA output indicating that glycerol does not influence product output. Data were recorded in technical triplets (n=1, three measurements), and error bars reflect technical repeatability.

Nature Chemical Biology: doi:10.1038/nchembio.2314



Supplementary Figure 16: Exemplary raw data of coupled assay. Shown is the HPLC chromatogram at 298 nm from a coupled assay of FAS1$^{G2599S\text{-}M2600W\text{-}I151A\text{-}R1408K}$ → FAS2$^{S126A\text{-}Y2227F}$. For the product, 6-HHP, and the internal standard, the structures and mass spectra are shown.

Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary Table 1: Reactions included in the computational model of the FAS reaction network. Kinetic parameters were assigned to forward and reverse reactions within constraints as described in the text. We defined a standard reaction rate under saturating conditions of 10 units malonyl-CoA consumed per second. "Priming" reactions are those that must occur once per product release cycle; reaction rates were chosen to ensure a throughput no slower than 1/s at standard substrate concentrations. "Main sequence" reactions are those that would result in a slower overall reaction rate than the experimentally observed ca. 10/s if they were to proceed any slower than this; therefore reaction rates were chosen to ensure a throughput no slower than 10/s at standard substrate concentrations.

Reaction formula	**Number of reactions**	**Class**	**Restrictions imposed**
ACP:C_2 + AT^a <-> ACP + AT^a:C_2	**1**	Priming	>1/s; Michaelis constant: 28 μM
AT^a + AT^b + acetyl:CoA <-> AT^a:C_2 + AT^b:CoA	**1**	Priming	>1/s
AT^b:CoA <-> AT^b + CoA	**1**	Priming	>1/s
mal:CoA + MPT^a + MPTb <-> MPT^a:mal + MPT^b:CoA	**1**	Main sequence	>10/s; Michaelis constant: 8 μM
MPT^a:mal + ACP <-> MPT^a + ACP:mal	**1**	Main sequence	>10/s
MPT^b:CoA <-> MPT^b + CoA	**1**	Main sequence	>10/s
ACP:C_n + KS <-> ACP + KS:C_n	**11**	Main sequence for C_2-C_{18}	>10/s for C_2-C_{18}
ACP:C_n + MPT^a <-> ACP + MPT^a:C_n	**11**	Main sequence for C_{16} and C_{18}	>10/s for C_{16}-C_{18}
KS:C_n + ACP:mal <-> KS + ACP:C_{n+2} + CO_2	**11**	Main sequence for C_2-C_{18}	>10/s for C_2-C_{18}
MPT^a:C_n + MPT^b:CoA <-> MPT^a + MPT^b + C_n:CoA	**11**	Main sequence for C_{16} and C_{18}	>10/s for C_{16}-C_{18}

Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary Table 2: Melting temperatures of different FAS constructs. Values were determined by dichroism measurements as described in Supplementary Fig. 12. Data show that mutations slightly affect protein stability. For more data on the stability of FAS mutants, see also Supplementary Fig. 3 and Online Methods. Mutations FAS1$^{G2599S\text{-}M2600F}$, FAS1$^{G2599S\text{-}M2600N}$ and FAS1$^{G2599S\text{-}M2600D}$ have been tested for their ability to influence the FA chain length spectrum at the beginning of this study. Owing to minor effects, these mutants have not been further employed.

Construct	c (nM)	T_M (°C)
FAS (wild type)	562	45.1
FAS1^{G2599S}	496	43.7
FAS1$^{G2599S\text{-}M2600W}$	485	42.1
FAS1$^{G2599S\text{-}M2600F}$	424	43.4
FAS1$^{G2599S\text{-}M2600N}$	570	40.3
FAS1$^{G2599S\text{-}M2600D}$	540	41.0
FAS2$^{S126A\text{-}Y2227F}$	456	43.9

Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary Table 3: Characteristics of the calibration for acyl-CoA esters. UV detection at 260 nm was used for quantification. For each calibration point three replicates were recorded.

Compound	slope (m)	intercept (c)	Correlation coefficient (R^2)	Calibration range
iC_5-CoA	25.94	-23.83	1.0000	5 - 200 µM
C_{17}-CoA	19.74	-104.31	0.9981	5 - 200 µM
C_6-CoA	20.52	76.38	0.9981	10 - 200 µM
C_8-CoA	27.11	1.77	1.0000	5 - 200 µM
C_{10}-CoA	18.29	-0.45	0.9998	5 - 100 µM
C_{14}-CoA	27.06	-37.69	0.9993	5 - 150 µM
C_{16}-CoA	27.64	-44.87	0.9997	5 - 200 µM
C_{18}-CoA	24.62	-182.83	0.9992	15 - 150 µM

Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary Note 1: Calibration of Model Parameters

Our computational model of the reaction network of *S. cerevisiae* FAS represents 51 distinct reactions, enumerated in (Supplementary Table 1). The core catalytic cycle was simplified by including only the canonical condensation and acetyl-, malonyl- and acyltransferase reactions, as catalyzed by the enzymatic activities of the KS, MPT and AT domains, and mediated through the ACP shuttle domain. The three catalytic activities responsible for the conversion of the post-condensation β-keto intermediate to a fully saturated acyl chain were not explicitly represented, under the assumption that these reactions proceed considerably faster than the overall time scale of the reaction. This assumption is justified by the observation that no unsaturated products are detected under normal conditions.

The 51 included reactions, each with a forward and reverse rate constant, yield a total of 102 model parameters. In the absence of an exhaustive experimental characterization of each of the individual reactions that comprise the network, we proceeded to narrow down the open parameter space in accordance with known properties of yeast FAS, as follows.

As a first step, we defined overall restrictions on the range of reaction rates to represent. Yeast FAS is known to process malonyl-CoA at a rate of about 2000 mU/mg FAS[1,2], which corresponds to 10-18 units consumed per second per synthetic unit, where a synthetic unit is defined as a grouping of one ACP shuttle domain and the single copy of each catalytic domain which is physically accessible to each ACP. Choosing a reference speed of 10/s for convenience, we assume that all relevant phenomena can be described by rates no more than three orders of magnitude slower or faster, yielding lower and upper bounds of $10^{-2}\,s^{-1}$ and $10^{4}\,s^{-1}$, thus representing a search range spanning six orders of magnitude.

A 'main sequence' of reactions, defined as essential steps, which must proceed forward at a certain rate to match the observed overall synthesis rate, was defined, consisting of the following reactions, which were constrained to proceed at rates of no less than 10/s:

- malonyl-CoA binding to the active site of MPT (given near-saturating concentrations of malonyl-CoA);
- CoA release from MPT;
- malonyl transfer from MPT to ACP;
- transfer of acyl intermediates of 4 to 14-carbon length from ACP to KS;
- condensation of ACP-bound malonyl with KS-bound acyl intermediate;
- transfer of favored products C_{16} and C_{18} from ACP to MPT;
- transfer of favored products C_{16} and C_{18} from MPT to CoA (product release).

Analogously, the reactions involved in acetyl-CoA priming (acetyl-CoA binding to AT, CoA release from AT, and acetyl transfer to ACP), which under saturating conditions occurs once for every 7 or 8 malonyl residues incorporated, were constrained to rates no slower than 1/s.

Acyl products in the range C_6-C_{14} can be generated in vitro in small but detectable quantities even under standard conditions[3]; a minimum rate for reactions required for these exit events was set 100-fold lower at 0.1/s.

The values of the Michaelis constant for substrates malonyl-CoA (8 μM) and acetyl-CoA (28 μM) have been reported[4]. Taken as a proxy for the binding affinity of these substrates for their respective canonical active sites, these values determine the ratio of the forward (k_1) and reverse (k_2) rates for the respective binding and unbinding reactions by the relationship $K_m=k_2/k_1$. For these reactions, the forward rate was picked randomly according to the general restrictions described above, and the reverse rate was calculated according to the experimentally determined ratio.

To sample relevant parameter sets from the large parameter space that remains after applying the constraints described above, a multi-level filtering process was implemented, to exclude the vast majority of candidate parameter sets, which were not descriptive of measured properties of yeast FAS described by Sumper et al.[3] Five sets of conditions were applied as filters, each corresponding to a scenario described by Sumper et al:

1) Yeast fatty acid synthase with iodoacetamide-induced condensation deficiency. Iodoacetamide-treated yeast FAS was shown to be deficient in the condensation reaction, with the gain of an ACP-

Nature Chemical Biology: doi:10.1038/nchembio.2314

and KS-dependent malonyl-CoA decarboxylase activity[3,5], at a rate of 300 mU/mg under saturating conditions, corresponding to 2 s^{-1} units of malonyl-CoA per synthetic unit. Candidate parameter sets were filtered to exclude decarboxylation rates outside the range of 0.4-4 s^{-1}. This filter has the advantage of invoking only the small number of reactions involved in acetyl and malonyl import, export and transfer, bypassing the much larger number of chain length-dependent reactions which depend on the condensation reaction.
2) Standard reaction. Initial filtering was for an output of >80% C_{14} and longer acyl products.
3) High-acetyl-CoA condition, with 9:1 ratio of acetyl-CoA to malonyl-CoA.
4) Acetyl-CoA input throttled: acetyl-CoA is initially not present, and is generated at a known rate by a separate malonyl-CoA decarboxylase enzyme.
5) Malonyl-CoA input throttled: malonyl-CoA is initially not present, and is generated at a known rate by a separate acetyl-CoA carboxylase enzyme.

Having established (i) initial constraints for the parameter space based on known properties of the system, and (ii) a filtering procedure designed to identify parameter sets producing fatty acid product spectra matching those observed experimentally, we proceeded, using high-performance computing resources, to generate and evaluate large numbers of candidate parameter sets. For each candidate parameter set, the procedure is as follows:

- populate the array of forward and reverse rate constants with randomly chosen values, chosen within the constraints defined above;
- launch successive stochastic simulations of the FAS reaction network under the five filtering conditions given above;
- reject and discard any parameter set when the output product spectrum fails to match any of the five filtering conditions.

In this step, we tested hundreds of millions of candidate parameter sets, and retained 31735 that matched, within a given tolerance, product output spectra for all five of the filter conditions outlined above.

The five-level filtering round was followed by a Monte Carlo optimization procedure. Candidate parameter sets were assigned a score *S* consisting of the root mean square deviation of the fractional representation of acyl chain lengths observed in the model (p_m) compared to those measured experimentally (p_e), with additional penalties for deviations for the rates of the iodoacetamide-blocked and 'standard' reactions 1) and 2) (r_1, r_2) from the reference values of 2 s^{-1} and 10 s^{-1}:

$$S = \sqrt{\frac{1}{9}\sum_{i=6,8,10,\ldots}^{22}(p_m - p_e)^2} + \frac{|r_1 - 2\mathrm{s}^{-1}| + |r_2 - 10\mathrm{s}^{-1}|}{100}$$

Parameters of the input models were then varied randomly, with any changes leading to a reduction in score being accepted.
We subjected each of the 31735 candidate parameter sets from the filtering stage to the described Monte Carlo optimization procedure. The best-scoring 1000 parameter sets were then chosen to compose the overall ensemble model for yeast FAS. The scores *S* for these 1000 sets ranged from 0.025 to 0.052.

Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary Note 2: Cross-validation of kinetic model against FAS mutations

Physical basis for representation of mutations in kinetic model

As described in the main text, we represented the effects of active site mutations in the MPT, KS and AT domains in the kinetic model, using free energy quantities calculated from atomistic simulation. These were calculated using truncated representations of the area surrounding each mutation, as described under Methods, necessarily excluding whole-protein scale effects in this megadalton complex, but accessing the relevant molecular recognition steps in atomic detail.

The relationship between the free energy change ΔG of a reaction, the equilibrium rate constant K, and the forward and reverse rate constants k_1 and k_2, is given by

$$\Delta G = -RT \ln K = -RT \ln \frac{k_1}{k_2}$$

From this relationship, a calculated change in the free energy difference of the reaction (ΔΔG) dictates a change in the ratio of the forward and reverse rate constants k_1 and k_2. We accordingly sampled the range of possible thermodynamic outcomes consistent with the calculated change for each mutation by running tens of thousands of simulations of the kinetic network, each based on a single parameter set selected from the 1000-member kinetic model ensemble, with the respective rate constants affected by the mutation modified in accordance with the calculated ratio. To interpret the result, the product spectra of these tens of thousands of simulations of kinetic scenarios consistent with the calculated free energy change were aggregated and averaged, with population standard deviations (shown in the respective figures as error bars) giving an indication of the spread of likely outcomes of the mutation.

Cross-validation against mutations

Following model training against known experimental conditions for *S. cerevisiae*, we applied the framework described above to represent FAS mutations tested in our *C. ammoniagenes* experimental system. In so doing, we pursued the dual goals of cross-validating the model against untrained data, and of supporting our interpretation of the experimental results.

As presented in the main text, while was not initially possible to reconcile modeling results for the double G1250S/M1251W mutation to the KS domain with experimental output and the “gatekeeper” hypothesis that guided the design of the mutations, an updated conception of the “gatekeeper” rationale allowed us to suggest a mutually consistent interpretation of the model and results. Moreover, the transferase mutations were tested experimentally only in combination with the double KS mutation, thus complicating our interpretation of the transferase mutation results.

For this reason, we here divide the cross-validation against experimental data into two parts. In the first, we consider the effects of the tested transferase mutations in *S. cerevisiae* in isolation, without changes applied to the KS domain, and compare to the respective experimental effects in our *C. ammoniagenes* system for each of these mutations, with the caveat that the experimental results were obtained in combination with the double KS mutation. In the second part, we describe our updated conception of the KS mutation, and again consider each transferase mutation in this updated context, yielding a more direct comparison with the experimental data, albeit subject to greater uncertainty.

Cross-validation without KS modification

To represent MPT mutation R1834K mutation in the *S. cerevisiae* FAS-based theoretical model, we first calculated, using atomistic simulation-based alchemical free energy calculations, the effect of the mutation on the binding affinity of malonyl-CoA in the active site. This calculation yielded a prediction of weakened binding by ΔΔG = +14.2 ± 1.1 kJ/mol upon introduction of the mutation. Consistent with our measurements on *C. ammoniagenes* FAS product spectra, the kinetic model responded to the incorporation of the calculated effect with a broad shift towards shorter products (see Supplementary Fig. 8). Intriguingly, this single perturbation led to a greater relative increase in the release of C_8-CoA than for the neighboring C_6 and C_{10} short chain products.

Nature Chemical Biology: doi:10.1038/nchembio.2314

For the AT domain mutation I151A (*C. ammoniagenes*) / I306A (*S. cerevisiae*), two hypotheses were considered as an explanation for increased C_8-CoA output: increased acetyl throughput on the one hand, and the creation of a novel exit pathway for short-length acyl chains on the other. For *S. cerevisiae* mutation I306A, in the AT domain, we calculated a ΔΔG value of -0.4 ± 1.6 kJ/mol for the effect of the mutation on binding of acetyl-CoA, indicating little to no effect on binding affinity. This speaks against the first hypothesis, that C_8-CoA output is enhanced by virtue of a favored loading of acetyl-CoA.
We probed the second hypothesis, which proposes a novel short-acyl binding channel, with atomistic simulations of C_8-CoA in complex with *S. cerevisiae* mutant I306A. C_8-CoA could not be plausibly fitted in alignment with the catalytic residues in the wild type active site, but was accommodated in the I306A mutant with minor local rearrangement (see Supplementary Fig. 4). The resulting structure was used as a starting point for ten unbiased molecular dynamics simulations, each exceeding 400 ns in length. The simulations maintained a plausible pre-catalytic conformation, in alignment with the catalytic serine side chain and the oxyanion hole, in nine cases, suggesting a complex that is stable at least in the microsecond range.
Additionally, we note that the first hypothesis, which suggested that an increase in acetyl availability as a result of the mutation would yield increased C_8-CoA production by virtue a general shift towards shorter products, as observed under conditions of high relative acetyl-CoA availability, would predict a reduction in the proportion of longer (C_{16} and C_{18} CoA esters) released. Conversely, were the relative availability of the AT active site to be reduced through competition with a non-canonical substrate in the form of short acyl chains, a relative reduction in acetyl availability could be expected, which could shift the balance towards longer products.
Linking this line of reasoning with our experimental findings, we note that the *C. ammoniagenes* mutation I151A yielded an increase in the detected level of C_{18}-CoA (see Fig. 2a), representing further evidence against the hypothesis that the tested AT mutation enhances C_8-CoA production through an increase in acetyl-CoA throughput, and in favor of a non-canonical AT-mediated C_8-CoA exit pathway.

Taken together, these results favor the hypothesis that I306A-mutated *S. cerevisiae* AT is a competent acyltransferase in this acyl chain length range. Introducing the possibility of C_8 exit through the AT domain into the kinetic model with a forward rate of $3s^{-1}$ for both the transfer of C_8 from ACP to AT domain and its exit as CoA ester, led to increased release of C_8-CoA product at the expense of marginal reductions in C_{16}-CoA and C_{18}-CoA output (see Supplementary Fig. 8). The rate of $3s^{-1}$ was chosen because, while we have no direct experimental or theoretical thermodynamic reference point, the hypothesis as formulated requires *a priori* a reaction rate on the same order as the overall synthesis reaction (ca. 10 s^{-1}).

C. ammoniagenes MPT mutation R1408K and AT mutation I151A, each of which individually brought significant increases in the relative yield of C_8-CoA, were further tested in combination (in addition to the KS domain mutations, yielding quadruply-mutated $FAS1^{G2599S\text{-}M2600W\text{-}R1408K\text{-}I151A}$). The combination of these two beneficial mutations might be expected to be superior to each in isolation; however, their effects on relative C_8-CoA yield were not additive. Rather, the *C. ammoniagenes* result showed the KS/MPT combination remained the highest-yielding for specific production of C_8-CoA (see Fig. 2a).

Incorporating the MPT and AT mutations simultaneously, the kinetic model reproduced the finding of non-additivity, with C_8-CoA output for the double MPT/AT mutant no higher than that for the MPT single mutant (see Supplementary Fig. 8).

In summary, the representation of the tested transferase mutations in the kinetic model yielded correct reproduction of (i) the broad shift towards shorter products observed as a result of a malonyl throughput-restricting MPT mutation, (ii) the increased C_8-CoA yield as a result of the AT mutation, (iii) the non-additive nature of the MPT and AT mutations, and (iv) the slight shift towards longer products that accompanies the AT mutation (see Fig. 2a and Supplementary Fig. 8).

Nature Chemical Biology: doi:10.1038/nchembio.2314

KS-mediated chain length control revisited

In the light of the convincing cross-validation results from transferase engineering, we revisited the unexplained result we obtained in our attempt to determine the molecular basis of KS-mediated chain length control. We identified two hypotheses, both potentially reconciling the apparent contradiction.

Underlying the first hypothesis, we note that a prominent side reaction of FAS/PKS protein family is the unproductive KS-mediated decarboxylation, without subsequent condensation. of ACP-bound malonyl. For FAS, this side reaction has previously been reported as occurring at a negligible basal rate under normal conditions, but at a considerably higher rate when key catalytic residues are blocked through covalent modification [5]. We therefore hypothesized that an unusually strongly KS-bound C_8 intermediate could approximate a stalled state and induce decarboxylation, conceivably leading to elevated levels of C_8-CoA product. To test this alternate hypothesis *in silico*, we asked whether a kinetic model extended by the reaction KS:C_8 + ACP:malonyl → KS:C_8 + ACP:acetyl + CO_2 might account for the increase in C_8-CoA release. Although we were able to identify members of our model ensemble for which the introduction of the decarboxylation reaction yielded an increase in C_8-CoA output, the minimum reaction rate for the decarboxylation reaction was at least an order of magnitude faster than the rate reported [5], thus appearing to contradict this interpretation. We concluded our exploration of this hypothesis with further experimental measurements. *In vitro* reaction with ^{13}C-labelled acetyl-CoA (but unlabeled malonyl-CoA) did not show priming of FA synthesis with unlabeled acetyl, as would be expected in the case that significant quantities of unlabeled malonyl-CoA were undergoing decarboxylation in the KS-engineered FAS (see Supplementary Fig. 10).

Invoking a second hypothesis, we tested whether a less restrictive description of the influence of the double KS mutation on C_8 throughput might reconcile our findings. In our initial representation of the calculated enhanced binding affinity of the C_8:KS complex, we assumed that the calculated free energy change, considerably lower for the mutated C_8:KS state, would not yield a reduction in the forward rate of the reaction C_8:ACP + KS → ACP + C_8:KS. We relaxed this restriction to allow reductions in this reaction rate up to a factor of ten (while still fixing the ratio of forward and reverse rate constants in accordance with the calculated free energy difference). In doing so, we established that for 100 members of the 1000-member model ensemble, a slowdown in the reaction ACP:C_8+KS → ACP+KS:C_8 by less than a factor of ten yielded an increase of at least 5% in C_8-CoA output. We note that in these model scenarios, the reduced forward rates for this reaction are compensated by greater slowdowns in the reverse rate of the same reaction, in order to enforce agreement with the calculated negative binding free energy change for the KS:C_8 complex.

The results of our reevaluation of the two tested transferase mutations in conjunction with our latter interpretation of the double KS mutation are summarized in Supplementary Fig. 11.

We note that the agreement between experiment and model for the G1250S/M1251W combination arises by construction, since we selected only those parameter sets that matched the known C_8-enriched *C. ammoniagenes* product spectrum. Conversely, we may evaluate the transferase mutations in combination with the updated KS-mutated model, since our representations of the transferase mutations are based only on calculated thermodynamic quantities.

Combining MPT mutation R1834K with the G1250S/M1251W double KS mutation, we observe increased C_8-CoA output, to 31±13%, compared to 46.6% observed experimentally. As with the MPT mutation in isolation, the model reproduced the general shift towards shorter products, including a considerable reduction in C_{18}-CoA release.

Similarly, introducing the I151A mutation in addition to the double KS mutation yielded a prediction of modestly increased C_8-CoA release, at the cost of a modest decrease in the native $C1_6$-CoA and $C1_8$-CoA products, again in good agreement with the experimental finding.

Applying all four mutations in combination, the notable experimental finding that the AT and MPT mutations, while each separately advantageous, do not in combination yield greater C_8-CoA production, is again reproduced in the model.

Summarizing our combined theoretical and experimental exploration, we conclude that the studied KS double mutation does not represent a “gate” in the KS substrate tunnel, in the narrow sense of a steric barrier counteracting binding of long acyl chains as we initially conceived. Rather, our findings lead us to interpret that the double mutation introduces a kinetic barrier that steers C_8-CoA away from KS-mediated elongation and thus towards release. A reevaluation of the cross-validation results for the

Nature Chemical Biology: doi:10.1038/nchembio.2314

transferase mutations did not contradict this implementation. The fact that the findings from this second round of cross-validation remain consistent with the experimental findings serves as a further confirmation of the usefulness of the model in representing the FAS catalytic cycle.

Supplementary Note 3: Biotechnological potential of FAS mediated product synthesis

In the light of the technological relevance of C_8-CoA and 6-HHP, we are now able to translate mutations into cellular systems for the production of these compounds in fermentative processes. Given the turnover number of engineered module 1 (3.7 min^{-1} over the three step synthesis; calculated from specific activity of 34 ± 11 mU/mg for $FAS1^{G2599S\text{-}M2600W\text{-}R1408K}$ (see Supplementary Fig. 3)) and of engineered module 2 (2.1 min^{-1} over the two step synthesis; calculated from specific activity of 13 ± 4 mU/mg (see Supplementary Fig. 3)), which are comparable to iterative PKS (4.2 min^{-1} over the three cycles of iterative synthesis by the PKS methylsalicylic acid synthase (MSAS) *in vitro*[7]), as well as expected high expression levels of FAS in microbial systems, the technological usability might mainly depend on the cellular environment of the production host (i.e. the innate acyl-CoA substrate levels, conflicting/competing peripheral metabolic pathways and sensitivity of the host to the produced compounds) as well as downstream purification processes. Nevertheless, the presented synthesis can be further optimized in its protein design. Note that wild type *C. ammoniagenes* FAS (specific activity of 380 ± 87 mU/mg, see Supplementary Fig. 3) is about 10-times faster than the module 1 modified in chain length regulation. Evaluated on the basis of reported activities (see Supplementary Fig. 3), the drop in activity is accounted by both, the KS and the transferase design. A first improvement of the presented synthesis might be the optimization towards increased affinities of MPT for short chain acyl-CoAs in order to increase acyl transfer rates. Such a strategy might also allow releasing the throttled import of malonyl in module 1 by the R1408K mutation.

Lastly, we would like the emphasize that the in spite of the potential for improving the presented synthesis in some aspects, as explained above for the MPT design, FAS-mediated synthesis will, finally, be limited by the confined iterative reaction mode. A framework for a more versatile synthesis of chemical compounds will have to be based on systems for vectorial natural compound biosynthesis, such as performed by modular PKS[8].

Nature Chemical Biology: doi:10.1038/nchembio.2314

Supplementary References

1 Fichtlscherer, F., Wellein, C., Mittag, M. & Schweizer, E. A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation. *Eur. J. Biochem.* **267**, 2666-2671 (2000).

2 Wieland, F., Renner, L., Verfürth, C. & Lynen, F. Studies on the Multi-Enzyme Complex of Yeast Fatty-Acid Synthetase. *Eur. J. Biochem.* **94**, 189-197 (1979).

3 Sumper, M., Oesterhelt, D., Riepertinger, C. & Lynen, F. Synthesis of various carboxylic acids by the multienzyme complex of fatty acid synthesis from yeast, and clarification of their structure. *Eur. J. Biochem.* **10**, 377-387 (1969).

4 Lynen, F. Yeast fatty acid synthetase *Methods Enzymol* **14**, 17-33 (1969).

5 Kresze, G.-B., Streber, L., Oesterhelt, D. & Lynen, F. Malonyl-Coenzyme A Decarboxylase as Product of the Reaction of Fatty Acid Synthetase with Iodoacetamide. *Eur. J. Biochem.* **79**, 191–199 (1977).

6 Ziegenhorn, J., Senn, M. & Bucher, T. Molar absorptivities of beta-NADH and beta-NADPH. *Clin. Chem.* **22**, 151-160 (1976).

7 Moriguchi, T. *et al.* Hidden function of catalytic domain in 6-methylsalicylic acid synthase for product release. *J. Biol. Chem.* **285**, 15637-15643 (2010).

8 Robbins, T., Liu, Y.-C., Cane, D. E. & Khosla, C. Structure and mechanism of assembly line polyketide synthases. *Curr. Opin. Struct. Biol.* **41**, 10-18 (2016).

Nature Chemical Biology: doi:10.1038/nchembio.2314