Supporting Information

“High-Throughput Screening of Catalytically Active Inclusion Bodies Using Laboratory Automation and Bayesian Optimization”

Laura Marie Helleckes[1,2†], Kira Küsters[1,2†], Christian Wagner[1,2], Rebecca Hamel[1,2], Ronja Saborowski[1], Jan Marienhagen[1,2], Wolfgang Wiechert[1,3], Marco Oldiges[1,2*]

[†]These authors contributed equally.

[1]Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[2]Institute of Biotechnology, RWTH Aachen University, 52074 Aachen, Germany

[3]Computational Systems Biotechnology (AVT.CSB), RWTH Aachen University, 52074 Aachen, Germany

*Corresponding author: m.oldiges@fz-juelich.de

ORCID-Identifier

Laura Marie Helleckes: https://orcid.org/0000-0001-7825-7998

Kira Küsters: https://orcid.org/0000-0002-2472-9926

Christian Wagner: https://orcid.org/ 0000-0002-8311-8412

Rebecca Hamel: https://orcid.org/0000-0002-2090-9364

Jan Marienhagen: https://orcid.org/0000-0001-5513-3730

Wolfgang Wiechert: https://orcid.org/0000-0001-8501-0694

Marco Oldiges: https://orcid.org/0000-0003-0704-5597

Supplementary Table S1: List of all constructed plasmids. The sign "/" symbolize that each of the sequences were integrated into a vector backbone respectively. "N" and "C" symbolize the enzyme terminus that is fused with the linker and aggregation-inducing tag.

Vector	Genotype
pET28a(+)::Kan	*ColE1 lacZ'* Kan^R P_{T7} P_{lac}
pET28a(+)::Kan::CcdB	*ColE1 lacZ'* Kan^R P_{T7} P_{lac} *ccdB*
pET28a(+)::Kan::*Bs*GDH_wildtype	*Bs*GDH_wildtype fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::TDoT	*Bs*GDH_N/*Bs*GDH_C::SG::TDoT fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::18AWT	*Bs*GDH_N/*Bs*GDH_C::SG::18AWT fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::L6KD	*Bs*GDH_N/*Bs*GDH_C::SG::L6KD fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::GFIL8	*Bs*GDH_N/*Bs*GDH_C::SG::GFIL8 fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::3HAMP	*Bs*GDH_N/*Bs*GDH_C::SG::3HAMP fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::TorA	*Bs*GDH_N/*Bs*GDH_C::SG::TorA fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::CBDCell	*Bs*GDH_N/*Bs*GDH_C::SG::CBDCell fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG::ELK16	*Bs*GDH_N/*Bs*GDH_C::SG::ELK16 fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::PT::TDoT	*Bs*GDH_N/*Bs*GDH_C::PT::TDoT fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::PT::18AWT	*Bs*GDH_N/*Bs*GDH_C::PT::18AWT fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::PT::L6KD	*Bs*GDH_N/*Bs*GDH_C::PT::L6KD fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_C::PT::GFIL8	*Bs*GDH_C::PT::GFIL8 fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::PT::3HAMP	*Bs*GDH_N/*Bs*GDH_C::PT::3HAMP fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::PT::TorA	*Bs*GDH_N/*Bs*GDH_C::PT::TorA fragment in pET28a(+)

pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::PT::CBDCell	*Bs*GDH_N/*Bs*GDH_C::PT::CBDCell fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N::PT::ELK16	*Bs*GDH_N::PT::ELK16 fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C:: G1/2/4/10::L6KD	*Bs*GDH_N/*Bs*GDH_C:: G1/2/4/10::L6KD fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C:: P1/2/3/4/5/10::L6KD	*Bs*GDH_N/*Bs*GDH_C:: P1/2/3/4/5/10::L6KD fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG:: L12(/24/48)KD	*Bs*GDH_N/*Bs*GDH_C::SG:: L12(/24/48)KD fragment in pET28a(+)
pET28a(+)::Kan::*Bs*GDH_N/*Bs*GDH_C::SG:: $(L6KD)_{2/(4)/6/8}$	*Bs*GDH_N/*Bs*GDH_C::SG:: $(L6KD)_{2/(4)/6/8}$ fragment in pET28a(+)

Supplementary Table S2: Recipe of M9 Autoinduction medium – 1000 mL

Salt Stock solution (5x)	200 mL
$MgSO_4*7H_2O$ solution (246.48 g L^{-1})	1 mL
$CaCl_2*5H_2O$ solution (14.702 g L^{-1})	1 mL
Trace element solution (1000x)	1 mL
Citrate/Fe solution (7.5 g L-1 $FeSO_4*7H_2O$ 113.95 g L-1 tri-NaCitrat*$2H_2O$)	2 mL
Thiamin solution (10 g L^{-1})	1 mL
2 % (w/v) Lactose solution	100 mL
5 % (w/v) Glucose solution	10 mL
Glycerin 99%	4 mL
Kanamycin solution (50 g L^{-1})	1 mL
add Milli-Q (final volume)	1000 mL
Salt Stock (5x)	1000 mL
$(NH_4)_2SO_4$	25 g
KH_2PO_4	15 g
Na_2HPO_4	33.9 g
NaCl	2.5 g
NH_4Cl	10 g
add Milli-Q (final volume)	1000 mL
Trace elements (1000x)	1000 mL
$AlCl_3*6H_2O$	0.75 g

$CoCl_2*6H_2O$	0.6 g
$CuSO_4*5H_2O$	2.5 g
H_3BO_3	0.5 g
$MnSO_4*1H_2O$	17.1 g
$Na_2MoO_4*2H_2O$	3 g
$NiCl_2*6H_2O$	1.7 g
$ZnSO_4*7H_2O$	15 g
Dissolve in 100 mL Milli-Q and 50 mL 32% HCl and add Milli-Q to final volume	

Overview of semi-automated CatIB construction workflow

To start the CatIB cloning process, plasmids containing the GGA fragments, i.e., the gene of interest, linker and aggregation-inducing tag sequence, were commercially synthesized by Synbio Technologies (Monmouth Junction, US). Competent *E. coli* DH5α and *E. coli* BL21(DE3) cells were prepared manually, because it was conducted only once in a large batch. The first transformation step of *E. coli* DH5α to multiply the copy number of the synthesized plasmid was automated using the Opentrons OT-2 liquid handling system together with an integrated thermocycler for heat-shocking and cooling of the cells. After transformation, 5 µL of the cells were spotted on an agar plate with the robotic platform. The transformed cells were used to purify the multiplied plasmid *via* an accelerated plasmid preparation by attaching the preparation filters on a vacuum station instead of performing several centrifugation steps. Automation of this step was challenging because plasmid amounts after automated purification were too low for efficient GGA. However, the acceleration of this step already led to a time-saving of approximately 20 %. The master mix for GGA was set up by mixing the required enzymes and buffer with Milli-Q® water. This short step was performed manually to ensure a remaining activity of the heat-sensitive reagents. GGA was performed again with the Opentrons OT-2 system in combination with the integrated thermocycler for the incubation of BsaI restriction enzyme or T4-ligase. The following transformation of *E. coli* DH5α with assembled GGA constructs was performed manually because the volumes for transformation in the integrated thermocycler was limited to 200 µL. However, to ensure successful transformation after GGA, a minimum of 1 mL batches were required. The following plasmid preparation and the retransformation of *E. coli* BL21(DE3) to generate the CatIB production strains could be accelerated or automated as described before.

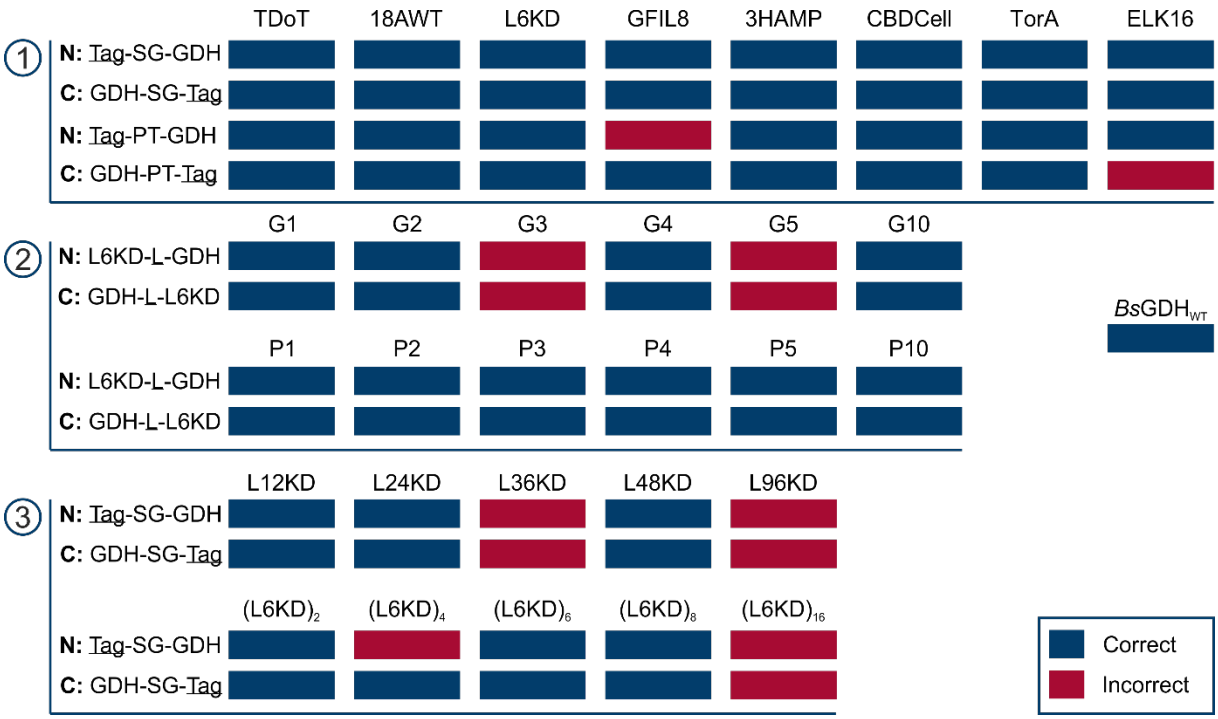

Figure S1: Overview of 76 tested *Bs*GDH-CatIB combinations and *Bs*GDH$_{WT}$ control using semi-automated cloning workflow. The blue marked constructs were verified by sequencing. The red marked constructs revealed incorrect sequencing results or no colonies were formed at the end of the cloning process.

Analysis of BsGDH-CatIBs via microscopy, SDS-PAGE and enzymatic assay

To analyze and compare different *Bs*GDH-CatIBs microscopic analysis (**Figure S2-S5**) and SDS-PAGE (**Figure S6**) were performed.

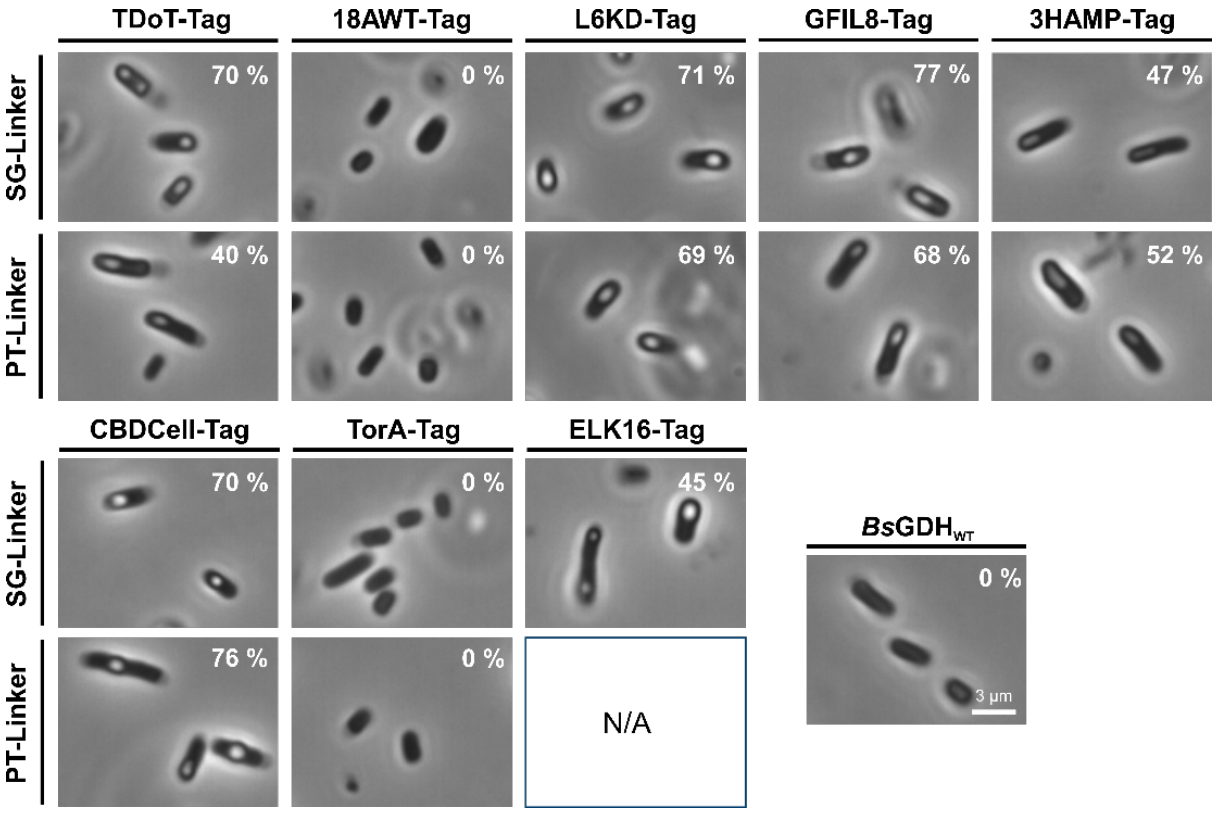

Figure S2: Microscopic images of strains producing *Bs*GDH-CatIBs with different linker/aggregation-inducing tag combinations tagged at the C-Terminus of the enzyme and *Bs*GDH$_{WT}$. Percentage of CatIB producing cells for each variant is displayed. Phase contrast microscopy was conducted with a 1000-fold magnification. All strains were cultivated for 72 h at 25 °C in M9 AI medium.

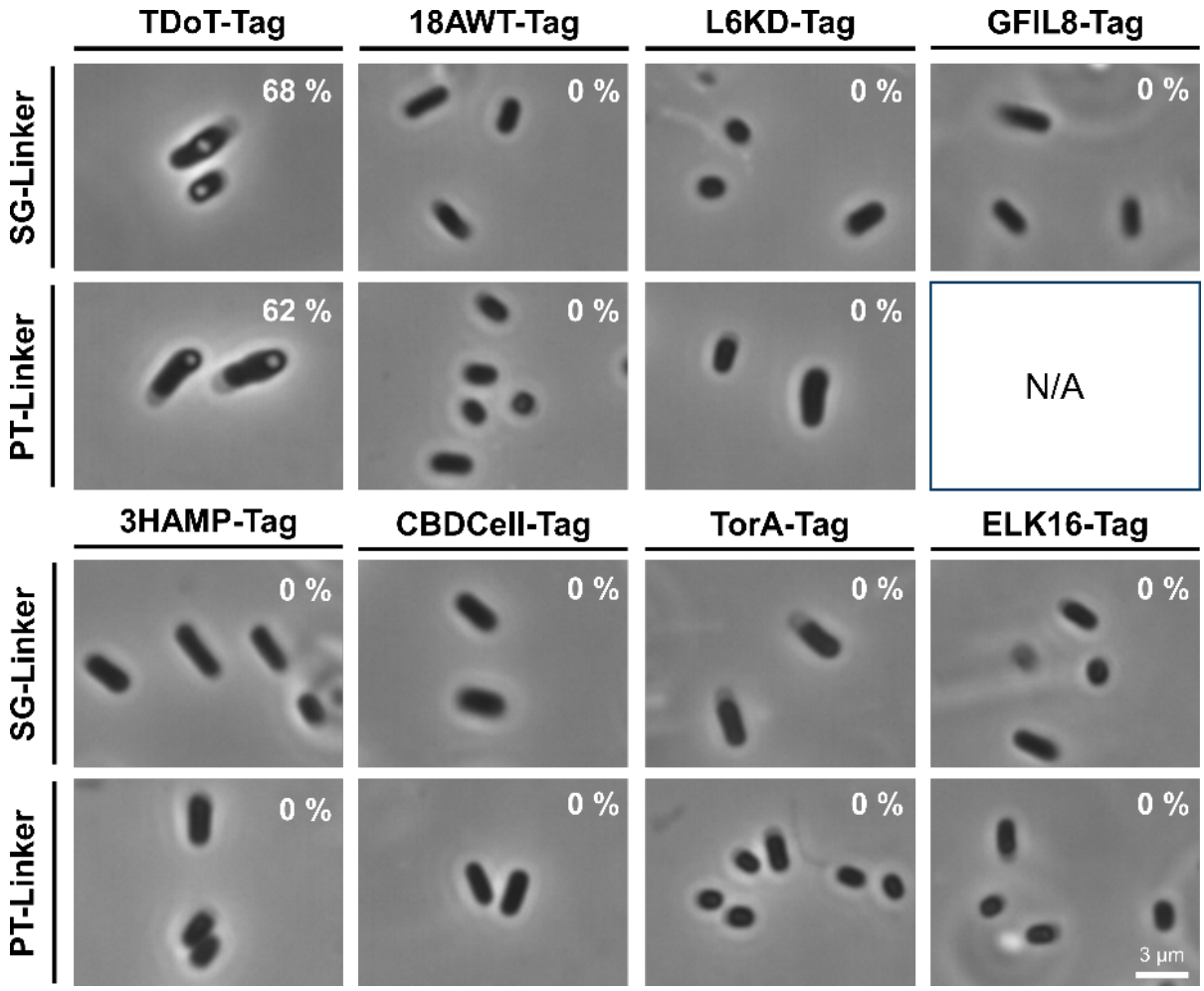

Figure S3: Microscopic images of strains producing *Bs*GDH-CatIBs with different linker/aggregation-inducing tag combinations tagged at the N-Terminus of the enzyme. Percentage of CatIB producing cells for each variant is displayed. Phase contrast microscopy was conducted with a 1000-fold magnification. All strains were cultivated for 72 h at 25 °C in M9 AI medium.

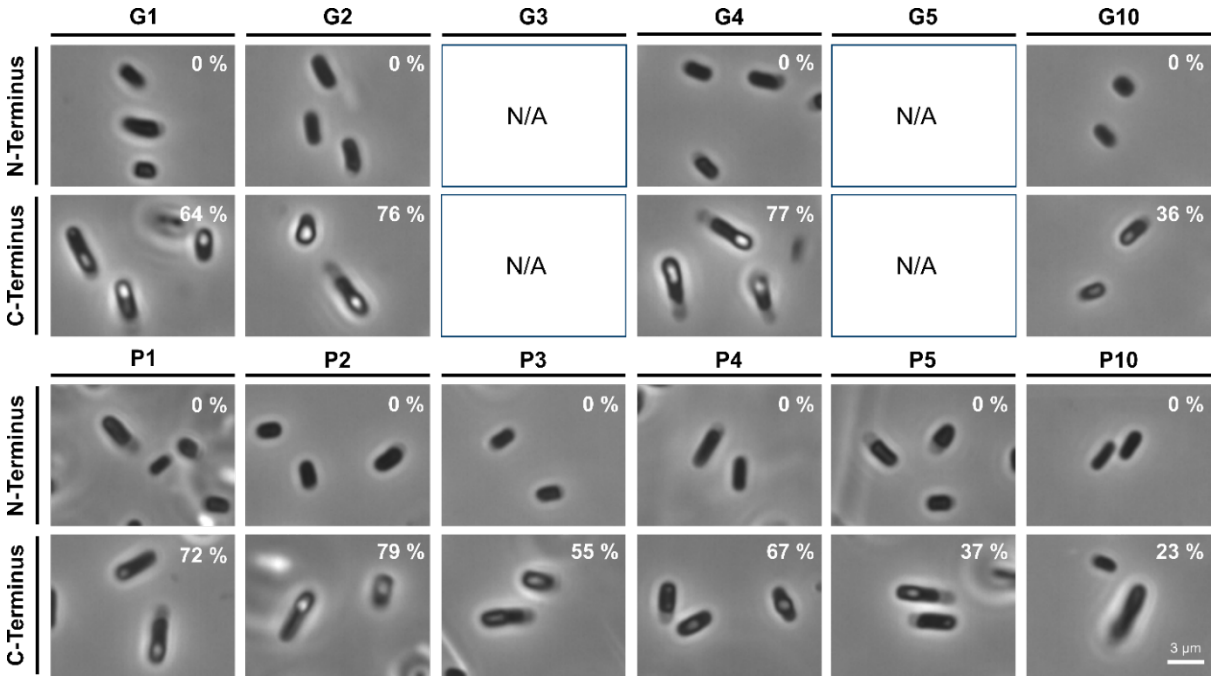

Figure S4: Microscopic images of strains producing *Bs*GDH-CatIBs with different lengths of glycine or proline linker tagged at the C- and N-Terminus of the enzyme. Percentage of CatIB producing cells for each variant is displayed. Phase contrast microscopy was conducted with a 1000-fold magnification. All strains were cultivated for 72 h at 25 °C in M9 AI medium.

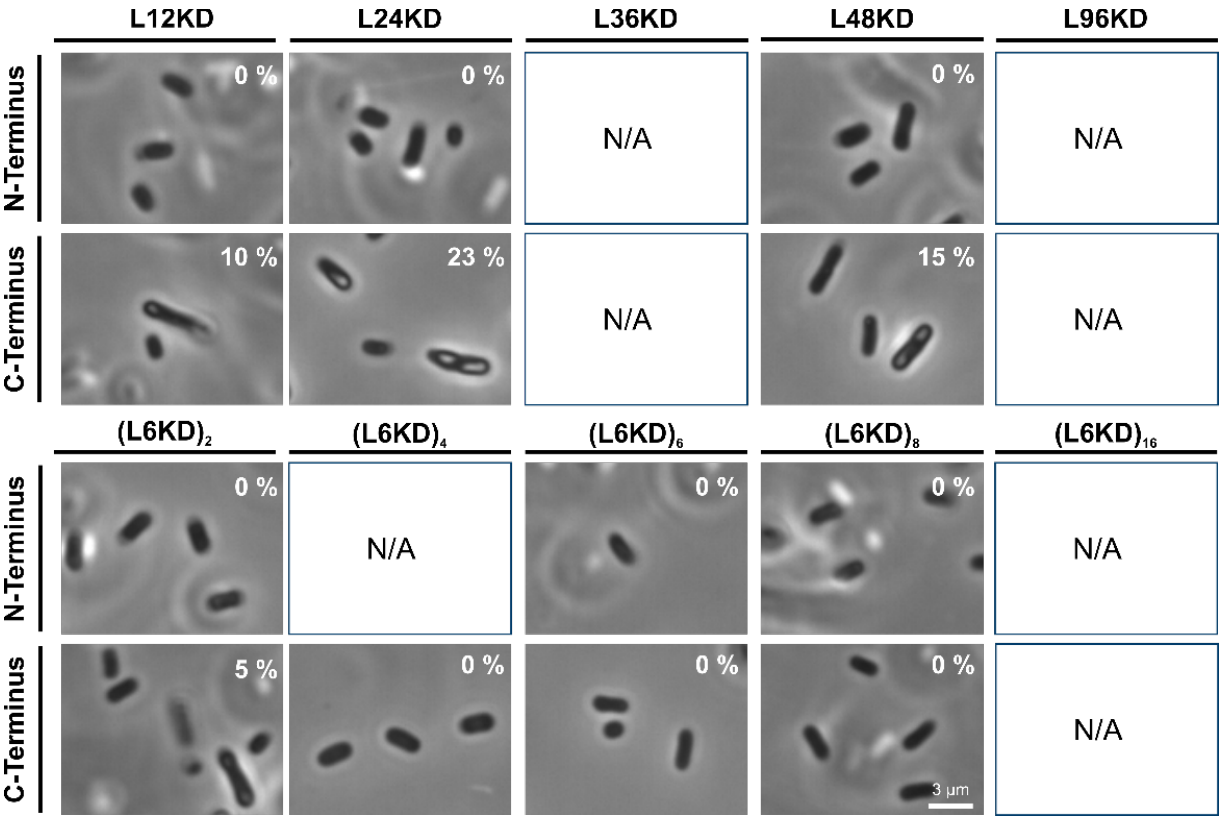

Figure S5: Microscopic images of strains producing *Bs*GDH-CatIBs with different lengths of L6KD-Tag linker tagged at the C- and N-Terminus of the enzyme. Percentage of CatIB producing cells for each variant is displayed. Phase contrast microscopy was conducted with a 1000-fold magnification. All strains were cultivated for 72 h at 25 °C in M9 AI medium.

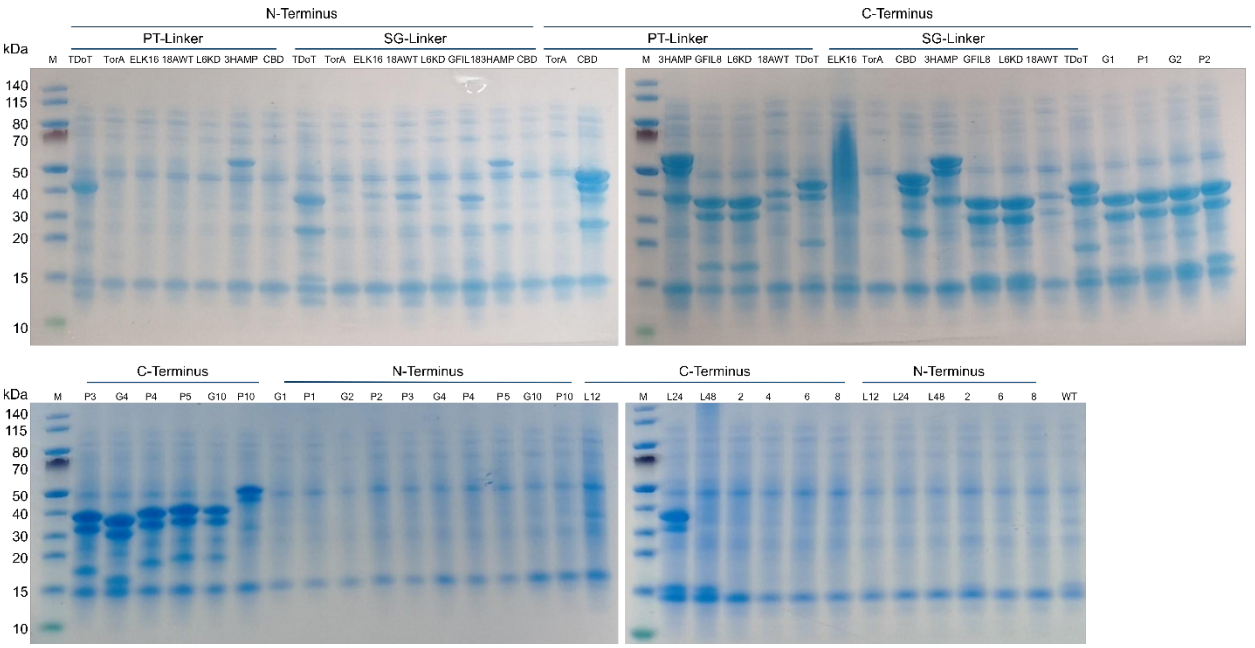

Figure S6: Evaluation of 63 *Bs*GDH-CatIB formation and *Bs*GDH$_{WT}$ by SDS-PAGE analysis. After cultivation, the cells were disrupted and the crude cell extract was separated by centrifugation into the soluble protein containing supernatant and the insoluble CatIB-containing pellet fractions. The pellet fraction was washed once with Milli-Q® water. The samples were diluted 1:1 with SDS sample buffer and 15 µL of each sample was loaded onto the gel and stained with SimplyBlue™ SafeStain. The molecular mass of the wildtype *Bs*GDH is 28 kDa.

Comparison of cultivation conditions with microscopic analysis

To further investigate the influence of standard CatIB cultivation conditions (3 h at 37 °C and 69 h at 15 °C) compared to the novel cultivation conditions (72 h at 25 °C), microscopic images were generated using a 1,000x magnification. Cells with successful CatIB production were counted and compared to cells without CatIB formation (**Figure S7).**

	Cells with CatIBs (%)			Cells with CatIBs (%)	
CatIB variant	**25 °C**	**37 - 15 °C**	CatIB variant	**25 °C**	**37 - 15 °C**
TDoT-PT-GDH	62	11	GDH-G1-L6KD	64	56
TDoT-SG-GDH	68	36	GDH-P1-L6KD	72	65
GDH-PT-CBDCell	76	52	GDH-G2-L6KD	76	53
GDH-PT-3HAMP	52	6	GDH-P2-L6KD	79	36
GDH-PT-GFIL8	68	75	GDH-P3-L6KD	55	38
GDH-PT-L6KD	69	63	GDH-G4-L6KD	77	72
GDH-PT-TDoT	40	13	GDH-P4-L6KD	67	57
GDH-SG-ELK16	45	75	GDH-P5-L6KD	37	22
GDH-SG-TorA	0	32	GDH-G10-L6KD	36	20
GDH-SG-CBDCell	70	75	GDH-P10-L6KD	23	0
GDH-SG-3HAMP	47	16	GDH-SG-L12KD	10	0
GDH-SG-GFIL8	77	63	GDH-SG-L24KD	23	86
GDH-SG-L6KD	71	79	GDH-SG-L48KD	15	71
GDH-SG-TDoT	70	0	GDH-SG-$(L6KD)_2$	5	0

Figure S7: Influence of cultivation temperature on CatIB formation analyzed via microscopy. The cultivations were performed at 25 °C (72 h) or 37 °C (3 h) + 15 °C (69 h) with M9 AI medium in a FlowerPlate. Microscopic images of all 63 strains were generated using a 1,000x magnification. The numbers indicate the percentage of CatIB producing cells given the overall number of counted cells.

Analysis of volumetric activity for selected BsGDH variants

Three CatIB variants (*Bs*GDH-PT-3HAMP, *Bs*GDH-PT-CBDCell, *Bs*GDH-PT-18AWT) were tested for their specific volumetric activity P_v at the two different cultivation conditions (shift from 37 °C to 15 °C or constant 25 °C) to test the influence of the temperature on the specific P_v (**Figure S8**). To determine the productivity of the CatIBs, the tested strains were produced in a larger scale in shake flasks. A manual purification and enzyme assay were performed and the CatIB dry weight was determined [1].

The results for all three CatIB variants showed that the cultivation at 25 °C positively influenced the productivity of the CatIBs. Even if the specific P_v of *Bs*GDH-PT-18AWT was lower than 10, this variant in general only showed activity after 25 °C cultivation. Moreover, with a 25 °C cultivation, the specific P_v of *Bs*GDH-PT-3HAMP was increased by approx. 40 % and *Bs*GDH-PT-CBDCell was increased by approx. 18 %.

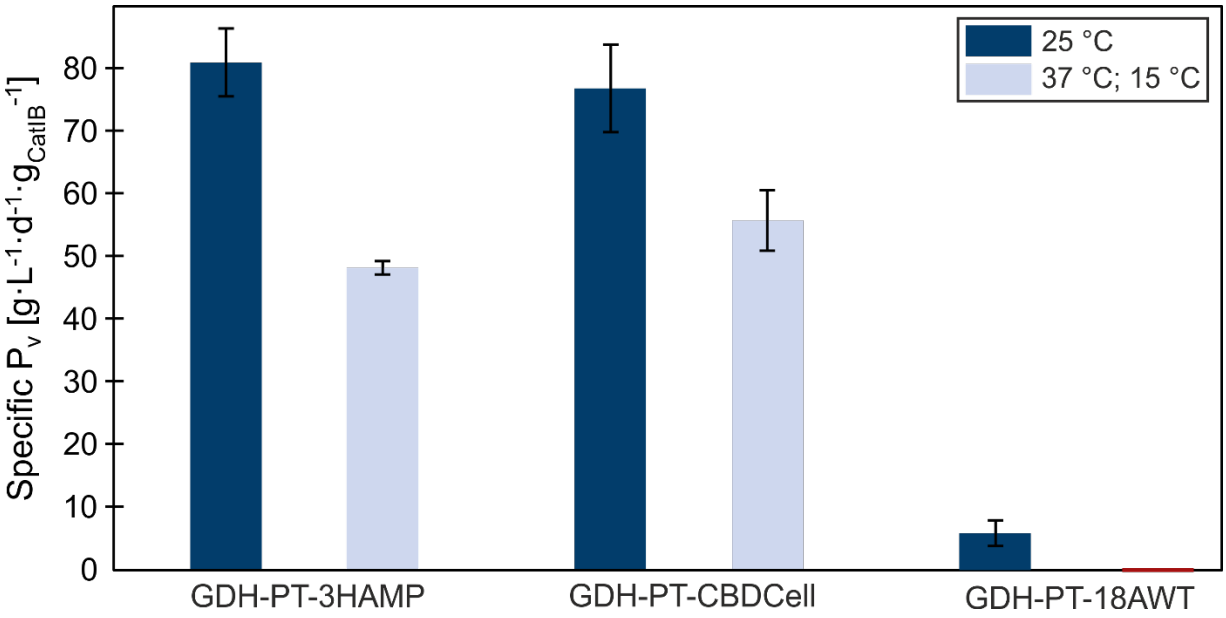

Figure S8: Influence of cultivation temperature on specific volumetric productivity of *Bs*GDH-CatIBs. The standard deviation of the technical triplicates for all CatIB variants were calculated.

Mathematical description of process and calibration model

The following equations summarize the process model, which is shown as a computation graph in Error! Reference source not found. in the main manuscript:

$$k_{\text{std}} \sim \text{HalfNormal}(\sigma = 0.2)$$

$$k_{\text{mean}} \sim \text{HalfNormal}(\sigma = 0.1)$$

$$t_{\text{offset}} \sim \text{LogNormal}(\mu = \log_{10}(0.283), \sigma = 0.1)$$

$$t_{\text{plate\_to\_reader}} \sim \text{LogNormal}(\mu = \log_{10}(0.5), \sigma = 0.05)$$

$$\text{batch\_effect}(i) \sim \text{LogNormal}(\mu = 0, \sigma = 0.1)$$

$$\text{cf\_nadh\_assay}\ (j) \sim \text{LogNormal}(\mu = 0, \sigma = 0.1)$$

$$S_0 \sim \text{LogNormal}(\mu = \log_{10}(0.2), \sigma = 0.1)$$

$$k_{\text{variant}}(v) = \text{LogNormal}(\mu = \log_{10}(k_{\text{mean}}), \sigma = k_{\text{std}})$$

$$k_{\text{batch}}(i, v) = k_{\text{variant}}(v) \cdot \text{batch\_effect}(i)$$

$$k_{\text{assay}}(i, j, v) = \frac{\text{cf\_nadh}_{\text{assay}}(j) \cdot k_{\text{batch}}(i, v)}{50}$$

$$t(c) = c \cdot t_{\text{offset}} + t_{\text{plate\_to\_reader}}$$

$$Y_{\text{pred}} = P(c, i, j, s) = S_0 \cdot \left(1 - e^{\left(k_{\text{assay}}(i,j,s) \cdot t(c)\right)}\right)$$

$$\mathcal{L}\left(\theta_{\text{pm}} | Y_{\text{obs}}\right) = \phi_{\text{cm}}(Y_{\text{obs}}, Y_{\text{pred}})$$

Indices c, i, j and v indicate the column in the assay microtiter plate, the cultivation well, the well in the microtiter plate of the assay and the CatIB variant respectively. The prior for k_{variant}

is defined by the mean and the standard deviation of the population, which are expressed as hyperpriors k_{mean} and k_{std}. Since the reaction is started column-wise by addition of the reaction substrate, a column-specific time offset t_{offset} is modelled, which was measured experimentally to be around 17 seconds per columns. Similarly, the time to put the plate to the reader was measured as 20 seconds. These measurements were chosen as the mean of the respective prior distributions. $\mathrm{cf\_nadh\_assay}$ reflects the pipetting error during dilution. The batch effect between biological replicates is represented by the respective parameter and is dependent on the cultivation well i. In the prediction of the product concentration in a specific well j in the microtiter plate of the assay, the effect of the priors for the variants, the time offset, the pipetting error and the batch effect are combined.

In order to derive distributions for all mentioned parameters of the process model θ_{pm} in Markov chain Monte Carlo Sampling, a likelihood function $\mathcal{L}$ needs to be defined. Here, we use a calibration model ϕ_{cm} that describes the exponential relationship between fluorescence observation Y_{obs} and predicted concentration Y_{pred} obtained from the process model. The calibration model is shown in **Figure S9**. More details on the methodology of process and calibration models can be found in [2].

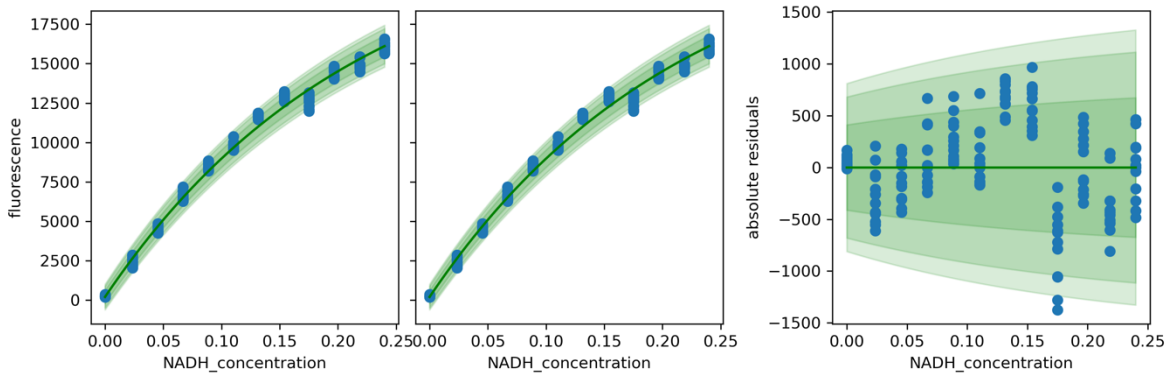

Figure S9: Calibration model that describes a normally distributed measurement error for measured NADH fluorescence in the assay. The mean of the normal distribution is described by an exponential trend between true concentration and fluorescence. The standard deviation is linearly increasing with the NADH concentrations, reflecting higher errors for increased NADH concentrations towards the upper detection limit of the plate reader. More details on calibration models can be found in [2].

References

[1] K. Küsters *et al.*, "Construction and characterization of *Bs*GDH-CatIB variants and application as robust and highly active redox cofactor regeneration module for biocatalysis," *Microb. Cell Factories*, vol. 21, no. 1, p. 108, Jun. 2022, doi: 10.1186/s12934-022-01816-2.

[2] L. M. Helleckes, M. Osthege, W. Wiechert, E. von Lieres, and M. Oldiges, "Bayesian calibration, process modeling and uncertainty quantification in biotechnology," *PLoS Comput. Biol.*, vol. 18, no. 3, p. e1009223, 2022, doi: 10.1371/journal.pcbi.1009223.