Electronic supplementary information

# Understanding substrate binding and the role of gatekeeping residues in PigC access tunnels

Stefanie Brands,[a] Jarno G. Sikkens,[a] Mehdi D. Davari,[a] Hannah U. C. Brass,[b] Andreas S. Klein,[b] Jörg Pietruszka,[b,c] Anna Joëlle Ruff[a] and Ulrich Schwaneberg*[a,d]

[a] Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 3, 52074 Aachen, Germany and Bioeconomy Science Center (BioSC). E-mail: u.schwaneberg@biotec.rwth-aachen.de.

[b] Institute of Bioorganic Chemistry, Heinrich Heine University Düsseldorf located at Forschungszentrum Jülich, Stetternicher Forst, Building 15.8, 52426 Jülich, Germany and Bioeconomy Science Center (BioSC).

[c] Institute of Bio- and Geosciences (IBG-1: Biotechnology), Forschungszentrum Jülich, 52426 Jülich, Germany.

[d] DWI-Leibniz Institut für Interaktive Materialien, Forckenbeckstraße 50, 52074Aachen, Germany.

***Corresponding author:**

Prof. Dr. U. Schwaneberg

u.schwaneberg@biotec.rwth-aachen.de

Phone.: +49 241 80 24170

# Table of Contents

# I. Experimental part

## Chemical synthesis of substrates and prodiginines

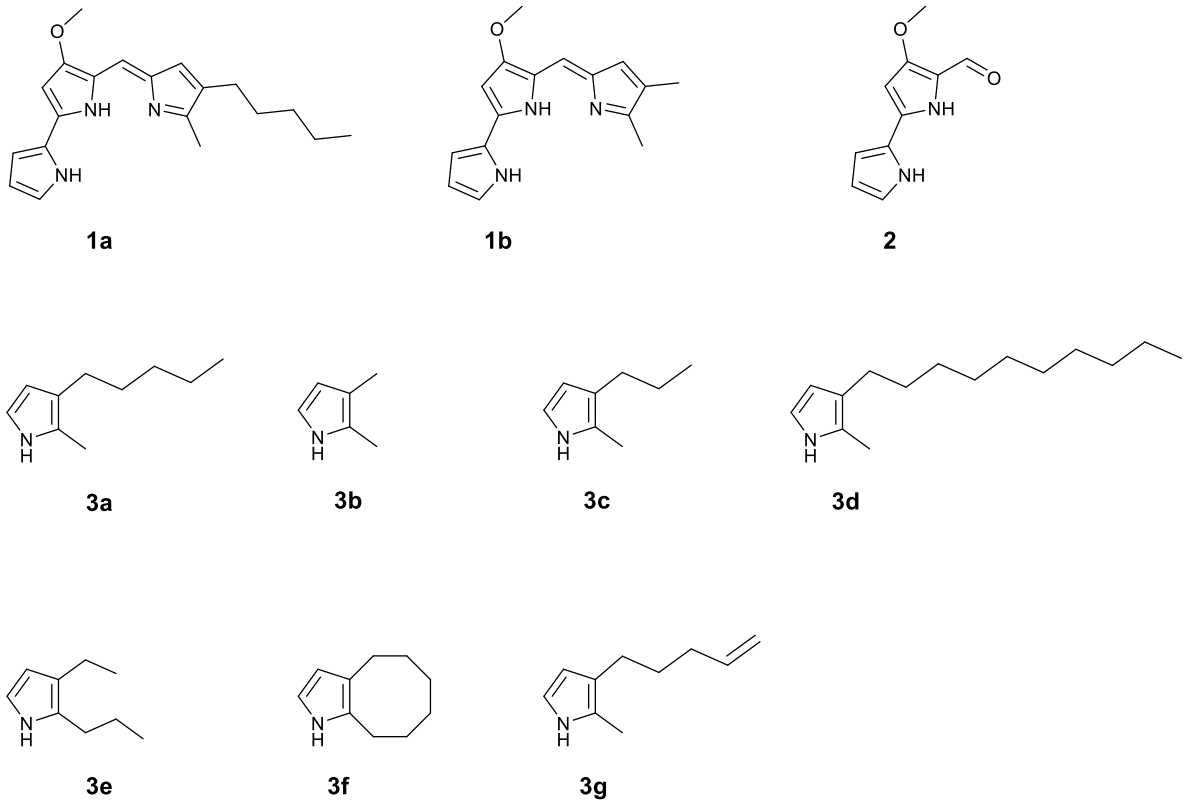



**Figure S1** Overview of pyrrolic compounds and prodiginines applied in this study. Prodigiosin (**1a**), prodiginine **1b** {4-methoxy-5-[(4,5-dimethyl-2*H*-pyrrol-2-yliden)methyl]-1*H*,1'*H*-2,2'-bipyrrole·HCl}, MBC (**2;** 4-methoxy-2,2'bipyrrole-5-carbaldehyde); MAP (**3a;** 3-amyl-2-methyl-1*H*-pyrrole); 2,3-dimethyl-1*H*-pyrrole (**3b**); 2-methyl-3-propyl-1*H*-pyrrole (**3c**); 3-decyl-2-methyl-1*H*-pyrrole (**3d**); 3-ethyl-2-propyl-1*H*-pyrrole (**3e**); 4,5,6,7,8,9-hexahydro-1*H*-cycloocta[*b*]pyrrole (**3f**), and 2-methyl-3-pentenyl-1*H*-pyrrole (**3g**) were synthesised as published.[1–4]

## Bacterial strains and plasmids

In this study, freshly prepared chemically competent *Escherichia coli* DH5α and NEB5alpha (New England Biolabs, Ipswich) were used for cloning of libraries and *pigC* constructs in the pVLT33[5] vector. *E. coli* BL21(DE3) pET28a(+) was applied for heterologous *pigC* expression for kinetic characterization of PigC variants. For screening and evaluation of the substrate profile, *Pseudomonas putida* mt-2 KT2440[6] was cultivated as published previously.[4]

## Site-saturation mutagenesis (SSM)

For saturation of the seven target amino acid positions in the PigC sequence, degenerate primers were designed that contained an NNK codon (N = A/G/T/C; K = G/T) at the respective target site. **Table S1** shows an overview of the targeted PigC positions (T329, G330, V333, T334, F603, R674, and P680) and respective NNK-primers.

**Table S1** Primers designed for site-saturation mutagenesis within PigC

| Position | | Target sequence | Primer sequence |
|---|---|---|---|
| T329 | Fwd | GCATGGAT**ACC**GGTGAAATCGTGACCGGTC | GCATGGAT**NNK**GGTGAAATCGTGACCGGTC |
| | Rev | CGATTTCACC**GGT**ATCCATGCGGCTGAAG | CGATTTCACC**MNN**ATCCATGCGGCTGAAG |
| G330 | Fwd | GGATACC**GGT**GAAATCGTGACCGGTC | GGATACC**NNK**GAAATCGTGACCGGTC |
| | Rev | CGATTTC**ACC**GGTATCCATGCGGCTG | CGATTTC**MNN**GGTATCCATGCGGCTG |
| V333 | Fwd | GAAATC**GTG**ACCGGTCTGATGACC | GAAATC**NNK**ACCGGTCTGATGACC |
| | Rev | GACCGGT**CAC**GATTTCACCGGTATCC | GACCGGT**MNN**GATTTCACCGGTATCC |
| T334 | Fwd | CGTG**ACC**GGTCTGATGACCCCAC | CGTG**NNK**GGTCTGATGACCCCAC |
| | Rev | CAGACC**GGT**CACGATTTCACC | CAGACC**MNN**CACGATTTCACC |
| F603 | Fwd | CGTCAAGAA**TTC**GAACTGAGCCTGCCACG | CGTCAAGAA**NNK**GAACTGAGCCTGCCACG |
| | Rev | CAGTTC**GAA**TTCTTGACGGCCACGGGCACC | CAGTTC**MNN**TTCTTGACGGCCACGGGCACC |
| R674 | Fwd | CGAG**CGT**CGCGAGGCGACCCGTCC | CGAG**NNK**CGCGAGGCGACCCGTCC |
| | Rev | CCTCGCG**ACG**CTCGGCCATCACG | CCTCGCG**MNN**CTCGGCCATCACG |
| P680 | Fwd | CGACCCGT**CCA**ACCTTCGTGACCGAAACC | CGACCCGT**NNK**ACCTTCGTGACCGAAACC |
| | Rev | CGAAGGT**TGG**ACGGGTCGCCTCGCGACG | CGAAGGT**MNN**ACGGGTCGCCTCGCGACG |

Site-saturation mutagenesis libraries were generated in *E. coli* NEB5alpha with a minimum of 200 clones per target position. The degenerate primers that implement a randomised NNK codon at the target site were used to amplify the complete pVLT33::*pigC* template (12.5 kb) in a 50 µL two-step PCR based on a modified QuikChange Mutagenesis protocol.[7,8] The standard two-step PCR contained 0.5 µM of both forward and reverse primer, 1x PfuS buffer (10 mM Tris-HCl pH 8.9, 50 mM KCl, 2 mM

$MgCl_2$, 0.1% Triton X-100), 20 ng linearised template DNA (pVLT33::*pigC* cut with *Hind*III), 0.2 U/µL PfuS polymerase, 0.2 mM dNTP mix and 0.8 M betaine monohydrate. Half of the final volume (25 µL) was prepared for each primer separately to prevent primer dimerization. After three PCR cycles of denaturation (30 s, 94 °C), annealing (30 s, 62-72 °C) and elongation (7 min, 72 °C), the two single primer PCRs were mixed, 1 µL of additional PfuS polymerase was added, and the reaction continued for 22 more cycles. All resulting PCR products were digested with *Dpn*I (0.8 U/µL, 2 h at 37 °C), purified or, in case of extensive side-product formation, separated by gel electrophoresis and extracted from a 1% agarose gel using a NucleoSpin Gel and PCR Clean-up Kit by Macherey-Nagel (Düren, Germany). NEB5alpha cells were transformed with the purified PCR products by heat shock to assemble the final constructs by homologous recombination,[9] following the supplier's instructions. The transformation was repeated until sufficient colony numbers were obtained, plasmids were pooled in a library, and transferred to *P. putida* KT2440 by electroporation[10] for screening.

## PigC screening

Screening of site-saturation mutagenesis (SSM) libraries in *P. putida* KT2440 was performed according to the ProdEvolve assay protocol.[4] Improved clones from the first screening were rescreened in biological triplicates, and confirmed beneficial clones were sequenced.

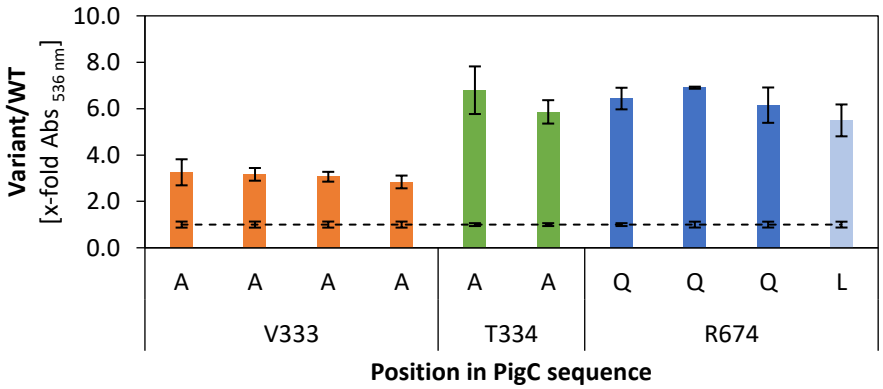



**Figure S2** Rescreening of beneficial SSM variants in biological triplicates. V333A was found 4-times, T334A 2-times, R674Q 3 times, and R674L one time by sequencing.

Recombination genes V1–V4 were synthesised by Thermo Fisher Scientific (Waltham, US-Massachusetts) implementing the following codons (**Table S2**).

**Table S2** Codon usage for substitutions in recombination genes V1–V4.

| PigC | Position | | |
|---|---|---|---|
| | **333** | **334** | **674** |
| **WT** | Val<br>GTG | Thr<br>ACC | Arg<br>CGT |
| **V1** | Ala<br>GCG | Ala<br>GCT | - |
| **V2** | Ala<br>GCG | - | Gln<br>CAG |
| **V3** | - | Ala<br>GCG | Gln<br>CAG |
| **V4** | Ala<br>GCG | Ala<br>GCT | Gln<br>CAG |

For activity assessment and substrate profile determination of recombination variants V1 to V4, the agar plate prescreening step was omitted, and colonies picked directly into 96-well plate precultures (150 µL LB, 50 µg/mL kanamycin) to be screened in microtiter plates.[4]

## Expression, purification, and kinetic characterization of PigC

Expression of select variants was performed in *E. coli* BL21(DE3) with the pET28a(+) vector system. Inserts were subcloned in the pET28a(+) vector backbone by homologous recombination,[9] applying the following primers:

| | |
|---|---|
| *HR_pigC_fwd* | CGCGGCAGCCACATATGAACCCGACCCTGG |
| *HR_pigC_rev* | GGTGGTGGTGCCTCGAGTCAGCCGTCGGCACG |
| *HR_pET28a_fwd* | CGTGCCGACGGCTGACTCGAGGCACCACCACC |
| *HR_pET28a_rev* | CCAGGGTCGGGTTCATATGTGGCTGCCGCG |

The PCR contained 5 U PfuS polymerase, 500 µM forward and reverse primer, 20 ng of template DNA, 200 nM of each dNTP and 0.8 M betaine, in a final volume of 50 µL in PfuS buffer (10 mM Tris-HCl pH 8.9, 50 mM KCl, 2 mM $MgCl_2$, 0.1% Triton X-100). The PCR protocol for pET28a(+) vector amplification was i) 3 min 94 °C, ii) 25 cycles of 30 s 94 °C, 30 s 56 °C, and 3 min 72 °C, and iii) 1 min 72 °C, before iv)

cooling to 4 °C for storage. PCR templates were digested with *Dpn*I (0.8 U/µL *Dpn*I, 2 h at 37 °C) and *E. coli* BL21(DE3) competent cells were transformed by heat shock to assemble the final constructs by homologous recombination.[9]

Expression (16–18 h, 18 °C, 250 rpm) in *E. coli* BL21(DE3) was performed in 200 mL ZYM5052 autoinduction media (10 g/L yeast extract, 20 g/L tryptone, 0.5 g/L glucose, 2 g/L α-lactose, 30 g/L glycerol, 2.55 g/L $Na_2HPO_4$, 3.4 g/L $KH_2PO_4$, 2.68 g/L $NH_4Cl$, 0.71 g/L $Na_2SO_4$, 0.49 g/L $MgSO_4$ x 7 $H_2O$; 2 h 37 °C) in 1 L flasks as published.[4] Subsequent isolation of the membrane fraction that contains PigC for kinetic characterization was performed as previously described.[3,4] PigC concentrations in the membrane fractions were determined by automated gel electrophoresis (Experion, Bio-Rad Laboratories, Inc., Hercules, US-California) on Pro260 microchips. Samples were prepared according to the Experion Pro260 Analysis Kit. The PigC concentrations were calculated from the fluorescence chromatogram with an internal bovine serum albumin standard (250 ng/µL). All measurements were performed in technical triplicates.

For kinetic characterization, a 96-well assay was performed in a buffered system as published before[4] [100 mM Tris/HCl, pH 7.5, 30 °C in 200 µL final volume with 25% (*v*/*v*) diluted *E. coli* BL21(DE3) membrane fraction, 0–200 µM MBC (**2**) and 0-200 µM monopyrrole (**3b**/**3e**) gradients, respectively, and 200 µM ATP]. Reaction buffer conditions were set according to previously published PigC catalytic optimum, pH 7.5 in Tris/HCl. Kinetic curves were monitored in a microtiter plate reader (CLARIOstar Plus, BMG Labtech, Ortenberg, Germany), and analyzed with Origin Lab 9.1 to determine kinetic parameters $k_{cat}$ and $K_M$. For fitting of *Michaelis-Menten* plots, the standard *Michaelis-Menten* equation with the *Levenberg-Marquardt* iteration algorithm was employed. An adapted *Michaelis-Menten* equation was used for fitting in case of substrate inhibition.[3,4]

$$v = \frac{v_{max}[S]}{K_M + [S](1 + \frac{[S]}{K_i})}$$

**Equation S1** Fit function for Michaelis Menten kinetics with substrate excess inhibition.

## *Michaelis-Menten* plots of PigC variants with substrates 2 and 3b

Variants V333A, T334A, R674Q and V1–V4 were kinetically characterised in *E. coli* BL21(DE3) membrane fraction in triplicates [Gradients: 0–100 µM MBC (**2**) and 0–200 µM 2,3-dimethyl-1*H*-pyrrole (**3b**), respectively].

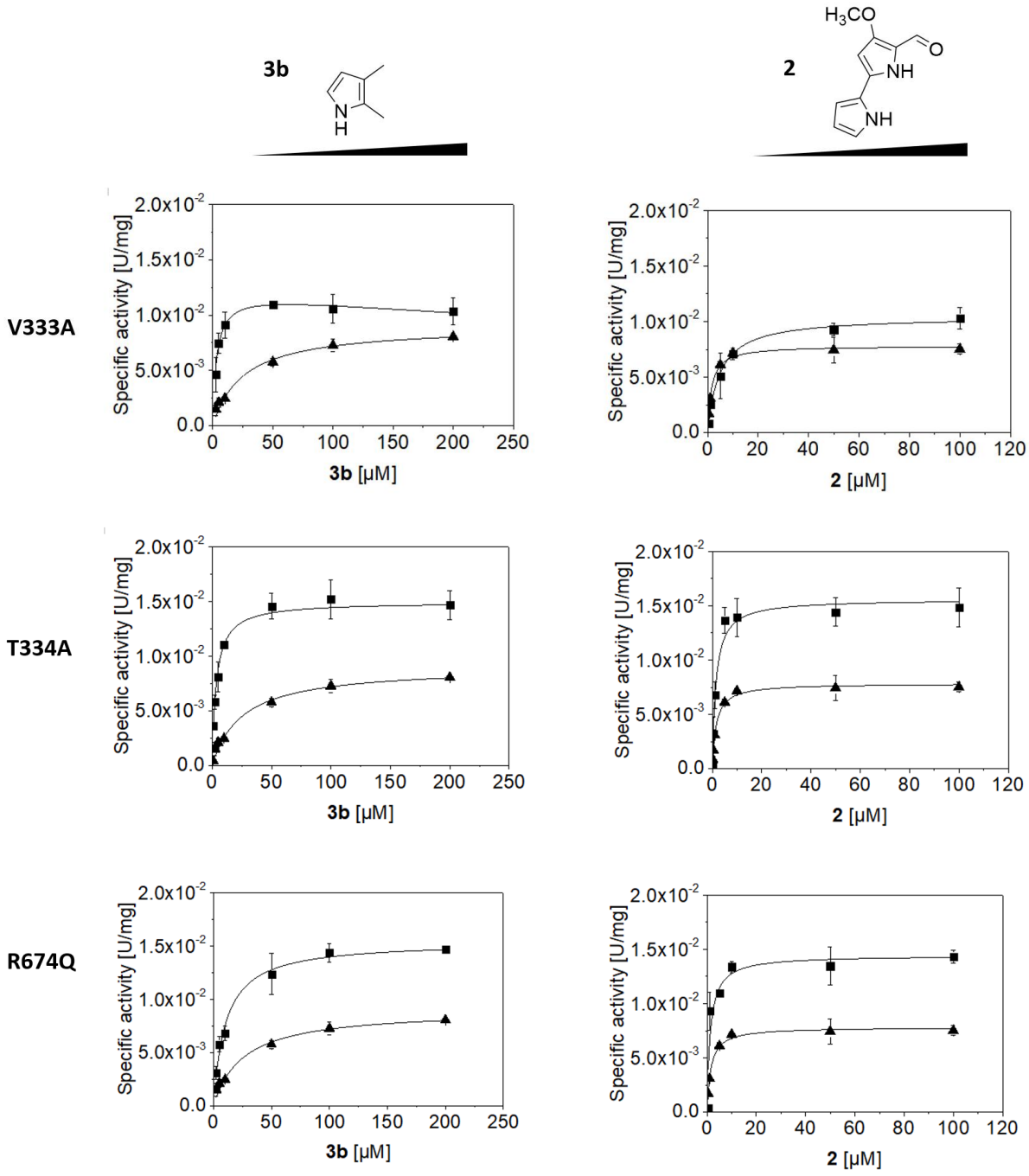



**Figure S3** *Michaelis-Menten* plots of variants V333A, T334A and R674Q with substrates **3b** and **2**. The wild type (WT, ▲) kinetic curves with the respective substrates are shown in each diagram for comparison. The error bars mark the standard deviation of triplicate PigC reactions. The error bars mark the standard deviation of triplicate PigC reactions.

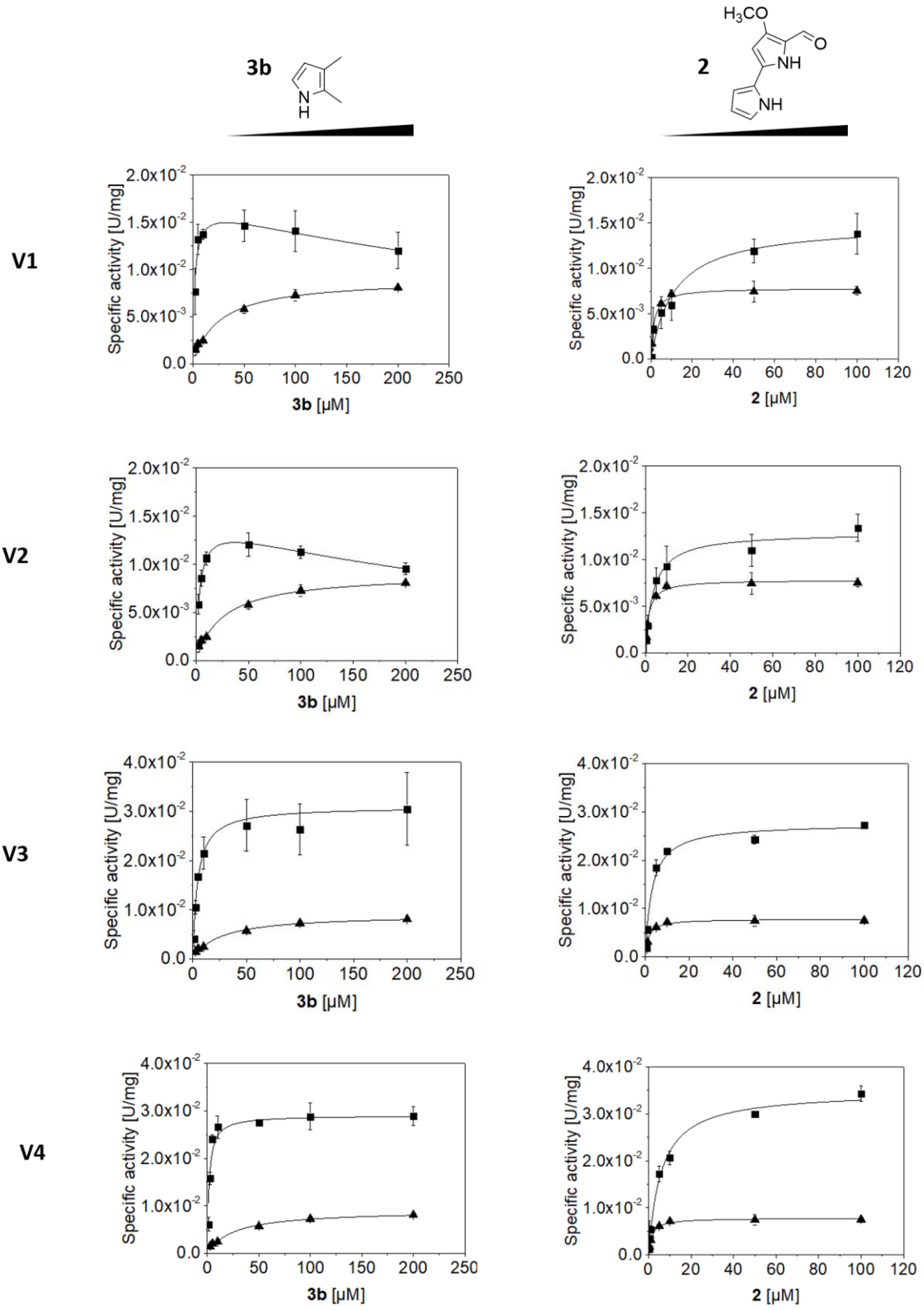



**Figure S4** *Michaelis-Menten* plots of variants V1–V4 with substrates **3b** and **2**. The wild type (WT, ▲) kinetic curves with the substrates are shown in each diagram for comparison. For fitting the V1 and V2 curves, **Equation S1** for substrate excess inhibition was applied. The error bars mark the standard deviation of triplicate PigC reactions.

### Substrate profiles of PigC variants

Substrate profiles of variants V333A, T334A, R674Q, and V1–V4 were determined in a 96-well assay with a set of substrates (**3a**-**3g**; **Figure S1**) as published before.[4]

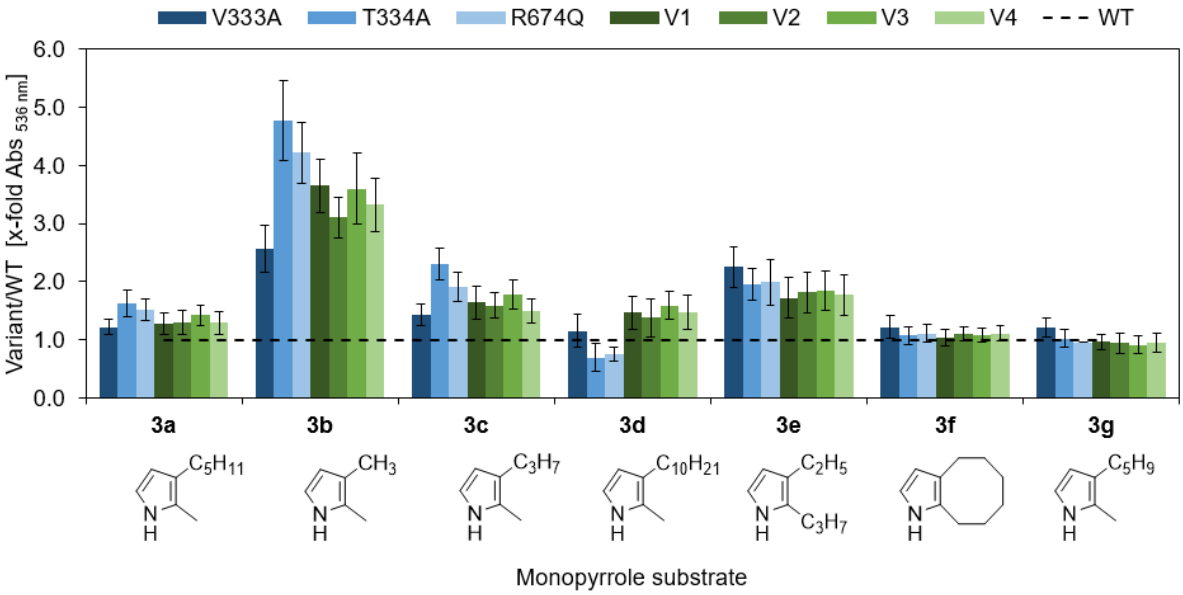



**Figure S5** Substrate acceptance profiles of variants V333A, T334A, R674Q, and V1–V4 as ratio to PigC wild type prodiginine (dashed line) with monopyrrolic substrates (50 µM **3a–g**, end-point absorbance measurement at 536 nm). The error bars indicate standard deviation of biological triplicates.

## II. Computational part

### Homology modelling

A PigC homology model (**Figure S6a**) was built on the I-TASSER web server[11–13] with the PigC amino acid sequence as input. The closest homologue on which the model was mainly built was a rifampin phosphotransferase from *Listeria monocytogenes* (PDB ID 5fbt; 2.70 Å),[14] with a coverage of 83% and a sequence identity of 22.6%. From a decoy database the best model was chosen with a c-score of 0.30, estimated TM-score of 0.75 ± 0.10, and an estimated RMSD of 7.9 ± 4.4 Å. The PDB file of the model was imported to YASARA and energy minimisation was performed (927.88 kcal/mol to 548.56 kcal/mol). Model validation was performed using the SAVES 5.0,[15–21] PRoSA-web, [22,23] and Molprobity[24] web tools on default settings. 98.6% of the amino acid residues were found in accepted regions in the Ramachandran plot (**Figure S6b**).[25] The overall model quality (Z-Score) of ProSA was -10.92, which put the model in range with proteins of similar size (**Figure S6c**). The 3D-1D score passed with 83.87% of the residues having an average score of >= 0.2. The Clashscore of all atoms was within the 99$^{th}$ percentile, and the Molprobity score was positive (2.16).

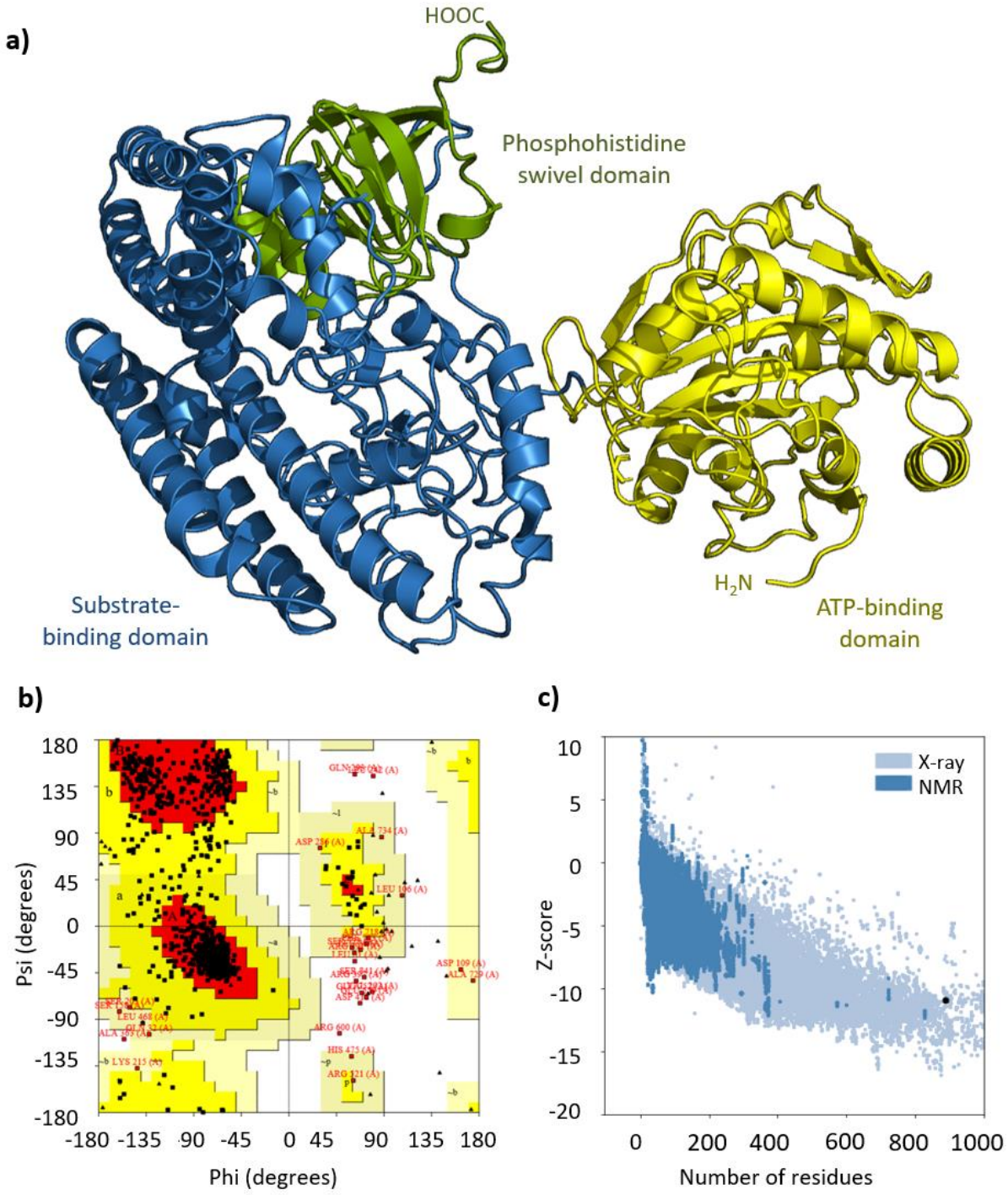



**Figure S6** PigC homology model validated by SAVES v5.0[15–21] and the PRoSA-web tool.[22,23] **a)** PigC homology model built on the I-TASSER web server.[11–13] The three PigC domains are coloured in **yellow** (ATP-binding domain, residues 1-298), **blue** (substrate-binding domain, 299-779), and **green** (phosphohistidine swivel domain, 780-888). **b)** Ramachandran plot of the PigC homology model. 81.4% of all residues are in favoured regions, 17.3% in allowed regions, and 1.4% in disallowed regions. **c)** Overall model quality of PigC homology model (•; Z-Score = -10.92). The Z-score measures the overall model quality in comparison to experimentally determined protein chains in the PDB (blue dots). The PigC homology model Z-score (•) is within the range of proteins with similar size.

## PigC *in silico* mutagenesis

The structural model of the PigC variants (V333A, T334A, R674Q and V1–V4) were constructed in YASARA Structure Version 17.4.17[26] employing the FoldX plugin[27] and using the FoldX method.[28] The starting coordinates for the FoldX *in silico* mutagenesis were taken from the homology model of PigC. Stability energies (**Table S3**) were computed with the FoldX Suite 4.0[28] using standard settings. The models were further neutralised and solvated in a periodic box containing TIP3P[29] water. All energy minimizations were carried using AMBER14[30,31] and GAFF[32] force field. Atomic partial charges were derived using the AM1/BCC procedure[33] implemented in YASARA. Electrostatics interactions was calculated using a cutoff of 7.86 Å and long-range interactions were calculated by using the particle-mesh Ewald (PME) integration. Initial energy minimizations by steepest descent were followed by simulated annealing until reaching convergence in potential energy (<0.02 kJ $mol^{-1}$ per atom during 200 steps). The refined models were used further for molecular docking simulations and cavity analysis.

**Table S3** Stability calculation (free energy change of folding; $\Delta\Delta G_{fold}$) after *in silico* mutagenesis by the FoldX method.[27]

| **Variant** | **$\Delta\Delta G_{fold}$** [kcal/mol] | **RMSD to WT** [Å] |
|---|---|---|
| V333A | +0.85 | 0.16 |
| T334A | +1.56 | 0.06 |
| R674Q | -0.06 | 0.06 |
| V1 - V333A/T334A | +2.39 | 0.17 |
| V2 - V333A/R674Q | +2.55 | 0.17 |
| V3 - T334A/R674Q | +0.52 | 0.09 |
| V4 - V333A/T334A/R674Q | +4.08 | 0.18 |

## Molecular docking

Ligands 2,3-dimethyl-1*H*-pyrrole (**3b**), MBC (**2**) and prodiginine **1b** were created and minimised in YASARA Structure Version 17.4.17.[26] For molecular docking, a grid box of 8 Å around the active site was generated by selecting the catalytic residue H840. Molecular docking simulations were performed using VINA[34] within YASARA with default parameters and a fixed protein backbone. 25 docking runs were performed and the docking poses were clustered by applying a RMSD cutoff of 0.5 Å and using the default settings provided within the YASARA dock_run macro file. The docking simulation was 10-times repeated with a randomly revolved receptor in the 8 Å simulation cell. The average binding energy was calculated of the best hits of 10 independent docking experiments with 25 runs each.

**Table S4** Binding energies determined in the best of 25 molecular docking runs with substrate 2,3-dimethyl-1*H*-pyrrole (**3b**).

| **PigC** | **Binding energy** [kcal/mol] | **Involved residues** |
|---|---|---|
| WT | -4.001 | M327 E331 Y348 Q349 W368 F489 F603 R675 R679 F682 |
| V333A | -4.402 | E331 I332 A333 R600 Q601 E602 F603 R675 R679 D817 A818 H840 |
| T334A | -4.018 | M327 E331 C345 Y348 Q349 W368 F603 R675 R679 F682 |
| R674Q | -4.016 | M327 E331 C345 Y348 Q349 W368 F603 R675 R679 F682 |
| V1 | -4.340 | E331 I332 A333 R600 Q601 E602 F603 R679 D817 A818 H840 |
| V2 | -4.396 | E331 I332 A333 R600 Q601 E602 F603 R675 R679 D817 A818 H840 |
| V3 | -4.013 | M327 E331 C345 Y348 Q349 W368 F603 R675 R679 F682 |
| V4 | -4.359 | E331 I332 A333 R600 Q601 E602 F603 R679 D817 A818 H840 |

**Table S5** Binding energies determined in the best of 25 molecular docking runs with substrate MBC (**2**).

| **PigC** | **Binding energy** [kcal/mol] | **Involved residues** |
|---|---|---|
| WT | -6.040 | E331 C345 Y348 Q349 W368 C493 F499 Q502 F603 R675 T678 R679 F682 |
| V333A | -6.331 | E331 I332 A333 C345 Y348 Q349 R600 Q601 E602 F603 R675 R679 F682 D817 A818 H840 |
| T334A | -6.067 | E331 C345 Y348 Q349 W368 C493 Q502 F603 R675 T678 R679 F682 |
| R674Q | -6.050 | E331 C345 Y348 Q349 W368 C493 F499 Q502 F603 R675 T678 R679 F682 |
| V1 | -5.684 | E331 Y348 Q349 W368 C493 Q502 F603 R675 T678 R679 F682 |
| V2 | -6.301 | E331 I332 A333 C345 Y348 Q349 R600 Q601 E602 F603 R675 R679 F682 D817 A818 H840 |
| V3 | -6.057 | E331 C345 Y348 Q349 W368 C493 F499 Q502 F603 R675 T678 R679 F682 |
| V4 | -5.714 | E331 Y348 W368 C493 Q502 F603 R675 T678 R679 F682 |

## Evolutionary conservation analysis

### ConSurf Analysis

The evolutionary conservation of PigC residues was analysed by using the ConSurf Server (https://consurf.tau.ac.il/) with the PigC sequence as input.[35–37]

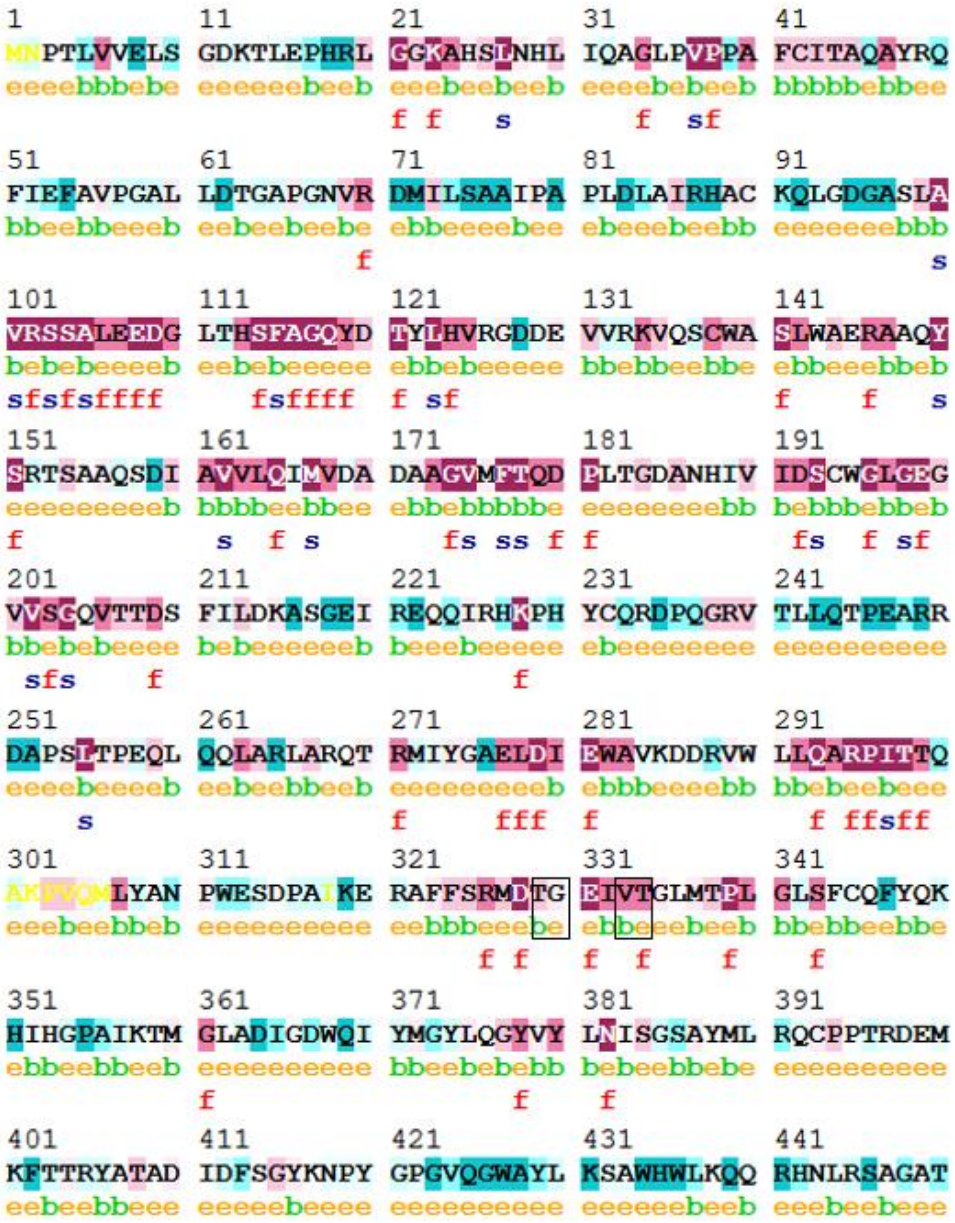


**Figure S7** ConSurf analysis of PigC sequence (Residues 1-450; continued on the next page).

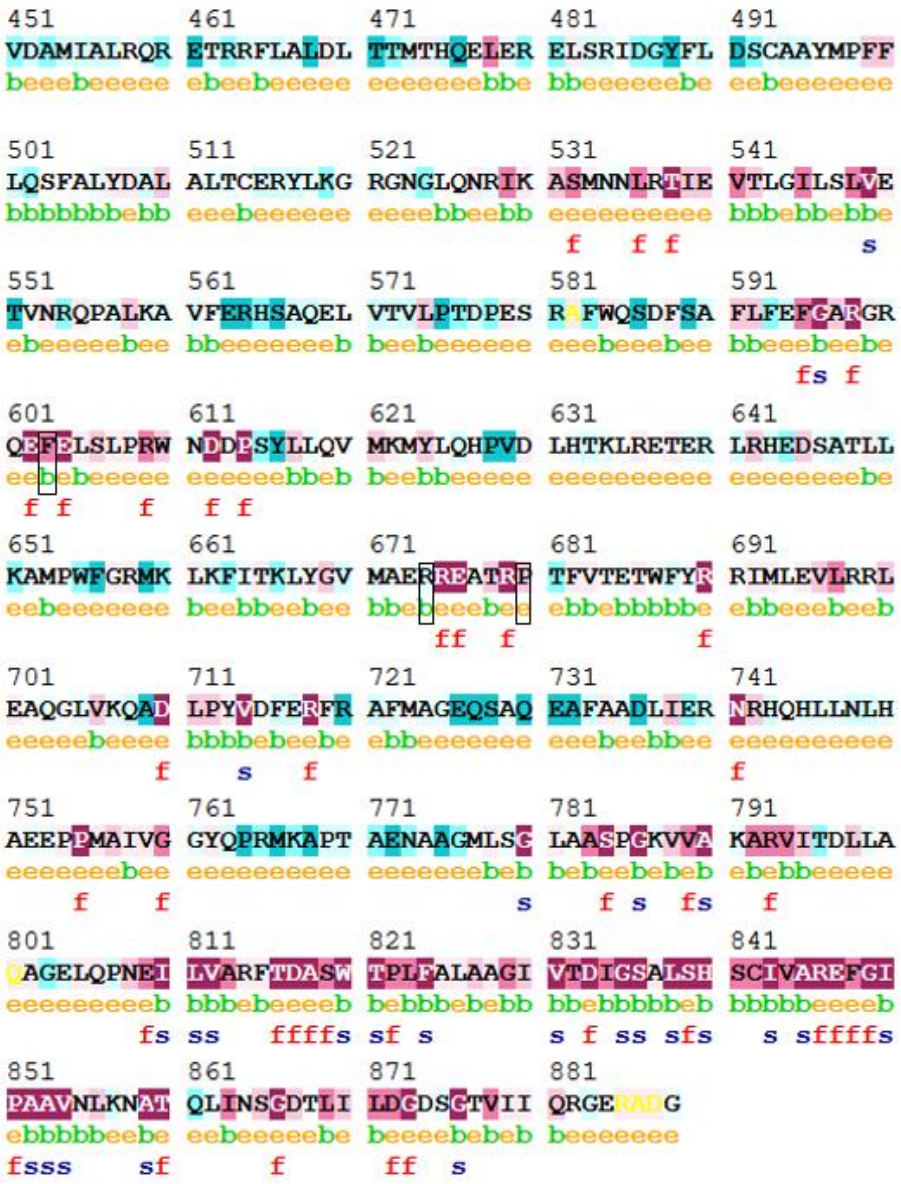
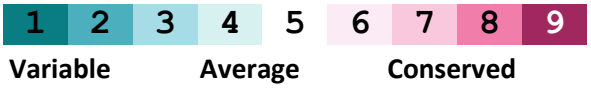


**Figure S7 (continued)** ConSurf analysis of PigC sequence. Residues selected for site-saturation mutagenesis are framed. The residues are colour-coded as indicated by the conservation scale:




e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

**Evolutionary Trace Analysis**

The evolutionary trace has been analysed on the Evolutionary Trace Server developed in the Lichtarge lab with standard settings.[38,39] The results are based on a multiple sequence alignment with 77 sequences. The output is the colour-coded PigC sequence by evolutionary importance (**Figure S8**).

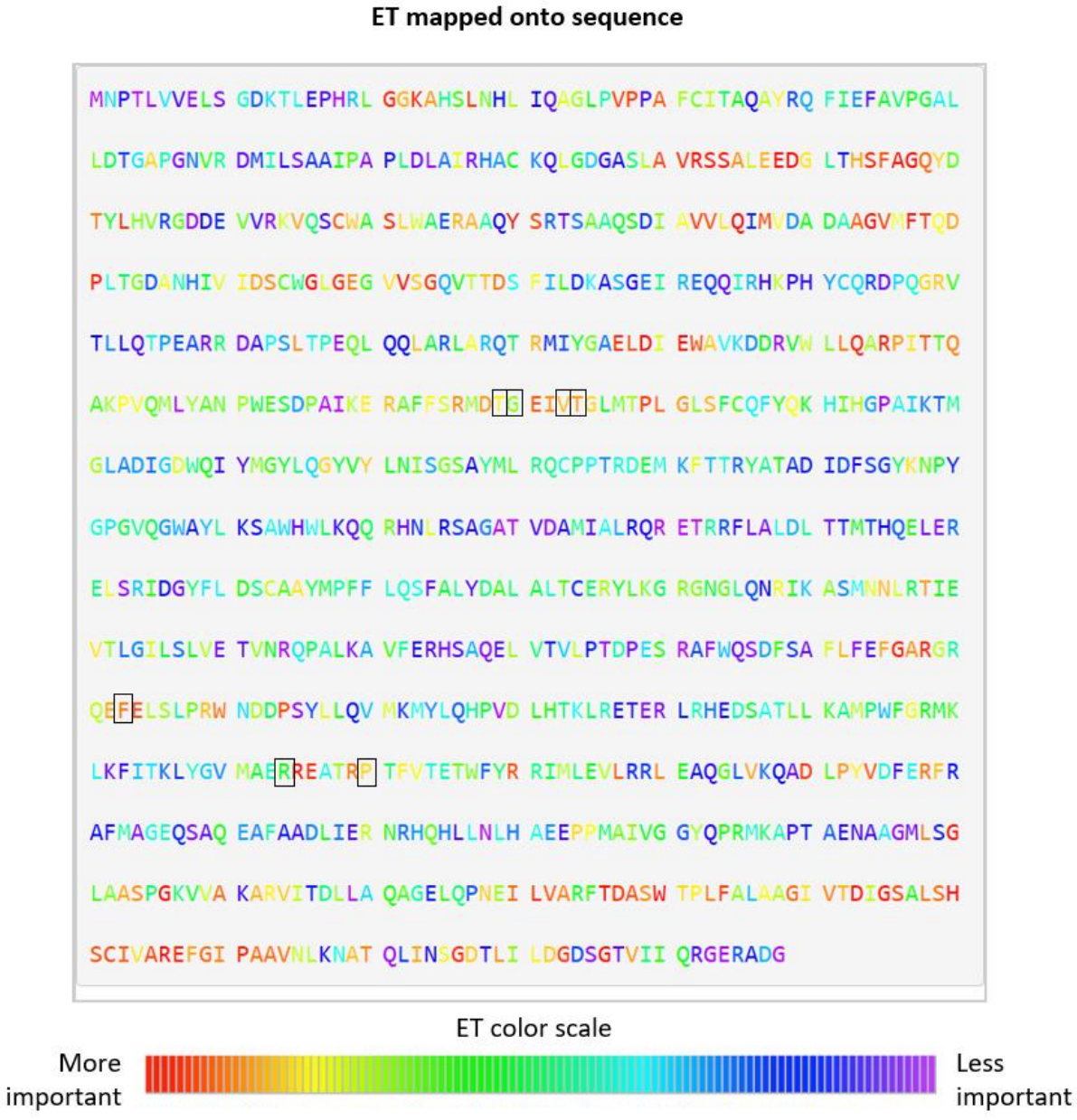



**Figure S8** Evolutionary trace (ET) mapped onto the PigC amino acid sequence. The evolutionary importance of residues is colour coded from red (more important) to violet (less important). Selected residues are highlighted in frames.

The reciprocal real-value evolutionary trace (rvET) of the above evolutionary trace analysis is plotted against the PigC sequence in **Figure S9** to highlight evolutionary important regions of the PigC substrate-binding domain (residues 299-779, conserved regions indicated by arrows).

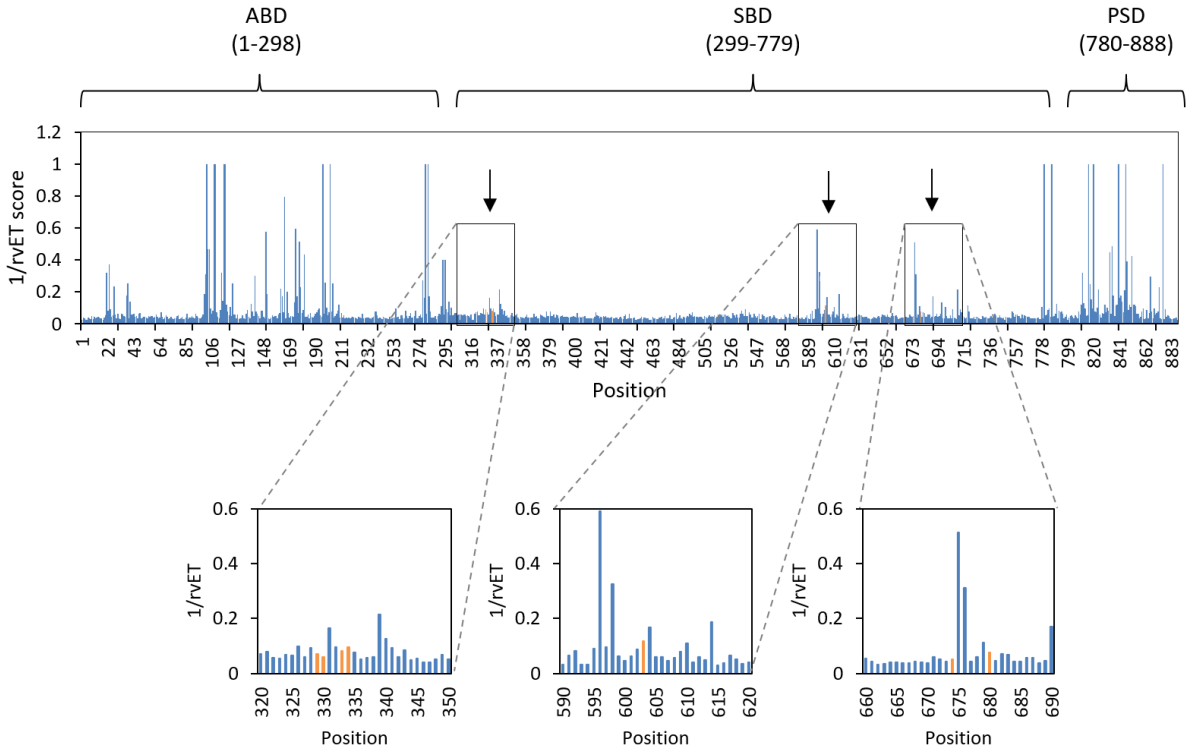



**Figure S9** Reciprocal rvET score of PigC residues. High values report high evolutionary importance. Target residues for semi-rational design in this study are marked by arrows and coloured orange. While they are neighbours to conserved sites, the target residues are not highly conserved themselves.[40]

**Multiple sequence alignment**

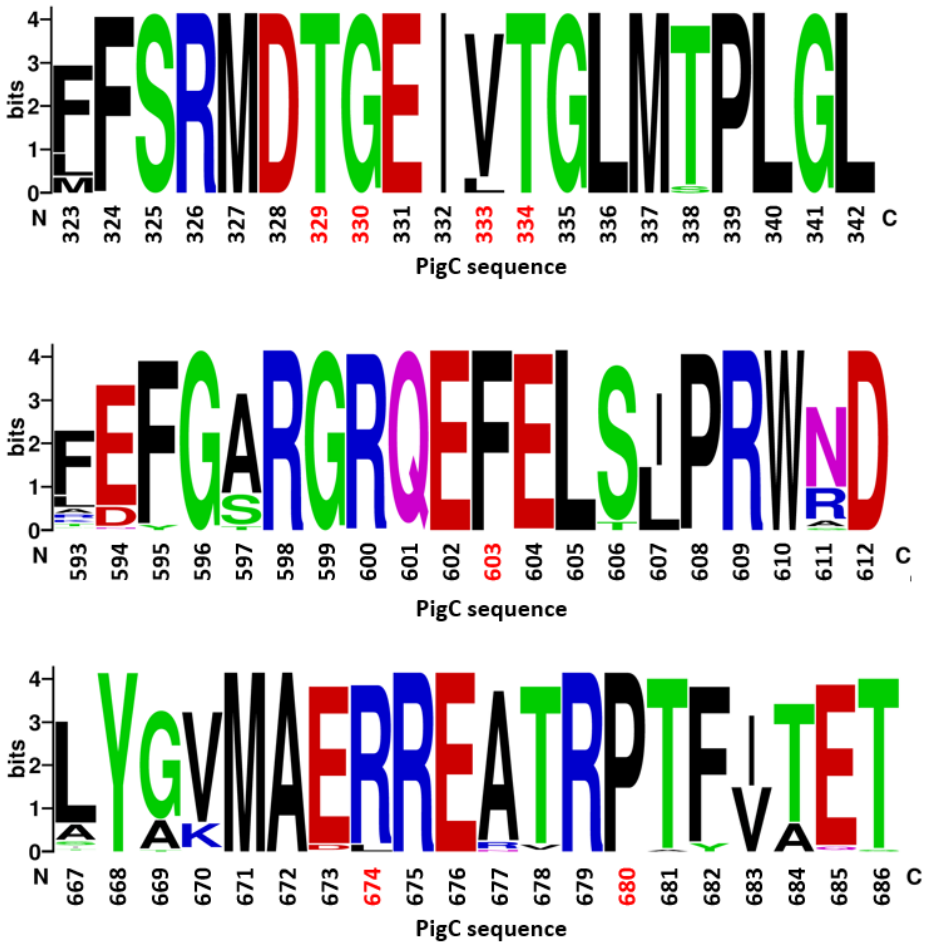



**Figure S10** Multiple sequence alignment of semi-rational design positions and flanking residues. PigC and 99 homologous sequences from a BLASTp search. The Figure was created with WebLogo.[41] The target position numbers are marked in red. **Black**: Hydrophobic amino acids. **Green**: Polar amino acids. **Red**: Amino acids with negative charge. **Blue**: amino acids with positive charge.

**FuncLib Analysis**

**Table S6** Highest scoring FuncLib[40] outputs on positions identified through ConSurf and Evolutionary Trace analysis. Serial numbers indicate the amino acid for each position; 01 codes for the wild type residue. One letter amino acid codes are given for each position. Positions showing variation (highlighted grey/bold numbers) were chosen as targets for semi-rational design.

| PigC | Serial number | Position in PigC sequence | | | | | | | | | | Total score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 327 | **329** | **330** | **333** | **334** | **603** | **674** | 677 | 678 | **680** | |
| WT | '01010101010101010101 | M | T | G | V | T | F | R | A | T | P | -1353.7 |
| | '01050601010301010104 | M | M | T | V | T | L | R | A | T | M | -1377.2 |
| | '01010401020401010105 | M | T | Q | V | A | R | R | A | T | Q | -1376.2 |
| | '01010601010304010102 | M | T | T | V | T | L | L | A | T | F | -1376.1 |
| | '01010602010104010102 | M | T | T | F | T | F | L | A | T | F | -1376.0 |
| | '01020602010101010102 | M | A | T | F | T | F | R | A | T | F | -1375.8 |
| | '01020101010304010104 | M | A | G | V | T | L | L | A | T | M | -1375.8 |
| | '01050601010103010102 | M | M | T | V | T | F | F | A | T | F | -1375.7 |
| | '01050101020104010102 | M | M | G | V | A | F | L | A | T | F | -1375.6 |
| | '01040401010301010104 | M | L | Q | V | T | L | R | A | T | M | -1375.4 |
| | '01050101020301010102 | M | M | G | V | A | L | R | A | T | F | -1375.2 |

## Cavity and tunnel analysis of PigC

The PigC active cavity and tunnels were localised and computed by Caver Web 1.0[42] with phosphohistidine 840 in the catalytic pocket as starting point and default settings (0.9 Å minimum probe radius, 4 Å shell depth, 3 Å shell radius, 3.5 Å clustering threshold, 3 Å maximal distance, and 5 Å desired radius). The identified PigC tunnels were subjected to ligand transport analysis in Caver Web 1.0.[42,43] Ligands were drawn in the web tool and their transport simulated through tunnels 2 and 3, which connected the active pocket with the protein surface. For substrates 2,3-dimethyl-1*H*-pyrrole (**3b**) and MBC (**2**), the entrance through the tunnel towards H840 was simulated, for the prodiginine product (**1b**) the release from the active pocket was simulated.

**Table S7** List of residues involved in predicted active cavity identified by the Caver Web tool 1.0.[42] Residues that were selected for site-saturation mutagenesis are marked red.

**59 residues of active pocket by Caver Web 1.0 analysis.**

| | | | | | |
|---|---|---|---|---|---|
| M327 | Y348 | L457 | P498 | R675 | W820 |
| D328 | Q349 | R458 | F499 | E676 | T821 |
| **T329** | W368 | R460 | L501 | A677 | F824 |
| **G330** | G377 | E461 | Q502 | T678 | L838 |
| E331 | Y378 | F489 | L536 | R679 | S839 |
| **V333** | V379 | D491 | T538 | **P680** | H840 |
| **T334** | T450 | S492 | Q601 | F682 | S841 |
| L336 | V451 | C493 | **F603** | T816 | I843 |
| L342 | A453 | A495 | A672 | D817 | V844 |
| C345 | M454 | M497 | **R674** | A818 | |

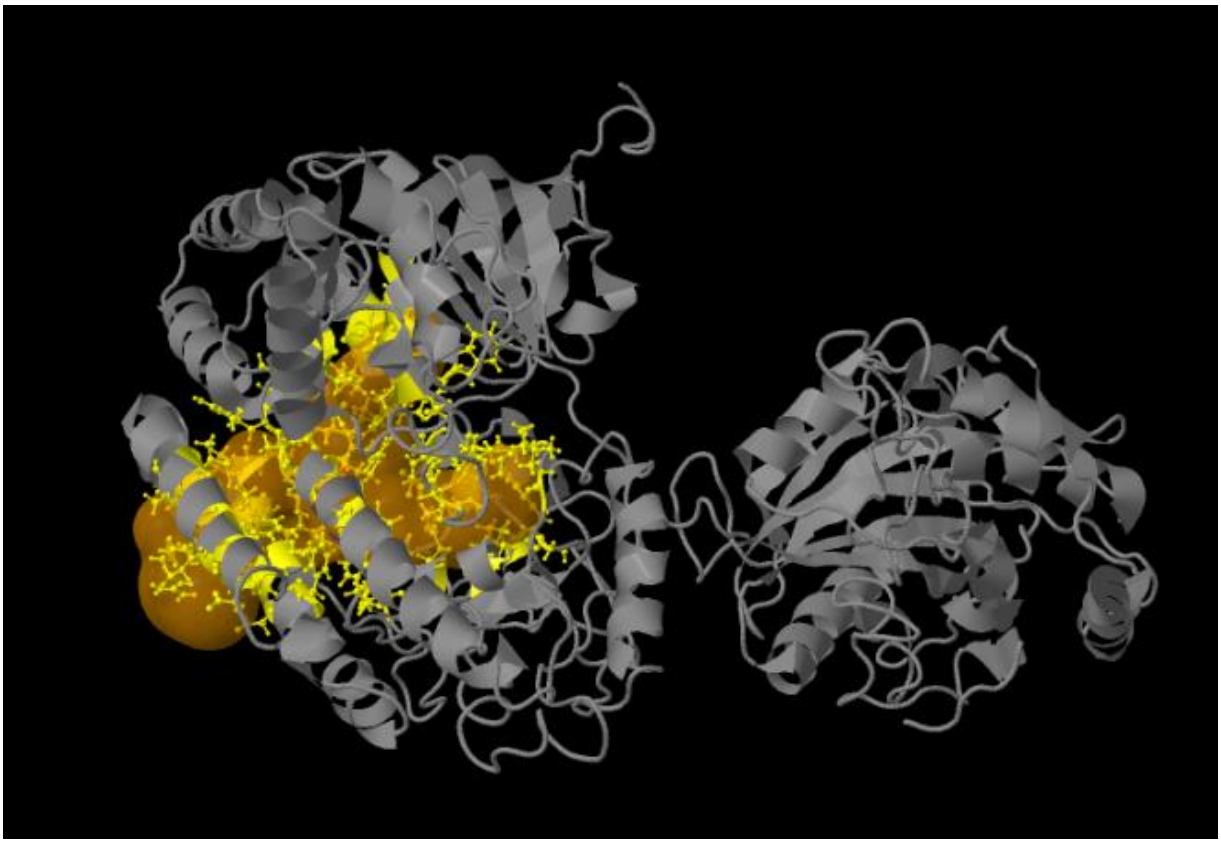


**Figure S11** Active cavity identified in the substrate-binding domain of the PigC wild type structural model. The cavity with 59 lining residues (**Table S7**) had a volume of 1612 $Å^3$, a relevance score of 100% and a druggability score of 0.3.[42]

**Table S8** CaverDock results of ligand transport analysis through PigC access tunnels 1-3.

| Tunnel | Bottleneck radius [Å] | Length [Å] | Curvature | Throughput |
|---|---|---|---|---|
| 1 | 1.3 | 12.6 | 1.4 | 0.6 |
| 2 | 1.3 | 28.8 | 1.4 | 0.3 |
| 3 | 1 | 42.3 | 2 | 0.2 |

Bottleneck radius – radius of the narrowest part of the tunnel; Length – length of the tunnel; Curvature – the curvature of the tunnel; Throughtput – troughput of the tunnel.

**Table S9** CaverDock results of ligand transport analysis through PigC access tunnels 1-3.

| Ligand | Tunnel | Direction | $E_{bound}$ [kcal/mol] | $E_{max}$ [kcal/mol] | $E_{surface}$ [kcal/mol] | $E_a$ [kcal/mol] | $\Delta E_{BS}$ [kcal/mol] |
|---|---|---|---|---|---|---|---|
| **1b** | 1 | OUT | 16.4 | 16.4 | −6.4 | 22.8 | 22.8 |
| **1b** | 2 | OUT | 15.5 | 15.5 | −5.5 | 21 | 21 |
| **1b** | 3 | OUT | 16 | 16 | −6.4 | 22.4 | 22.4 |
| **2** | 1 | IN | 5.7 | 11.6 | −5.0 | 16.6 | 10.7 |
| **2** | 2 | IN | 5.5 | 11.1 | −1.1 | 12.2 | 6.6 |
| **2** | 3 | IN | 5.8 | 11 | −5.5 | 16.5 | 11.3 |
| **3b** | 1 | IN | 4.2 | 7.4 | −3.3 | 10.7 | 7.5 |
| **3b** | 2 | IN | 4.3 | 8.3 | −2.9 | 11.2 | 7.2 |
| **3b** | 3 | IN | 4.1 | 8.2 | −4.0 | 12.2 | 8.1 |

**Ligand** – number of the used molecule (**Figure S1**); **Tunnel** – selected protein tunnel; **Direction** – direction of the CaverDock trajectory (into or out of active site pocket); $E_{bound}$ – the binding energy of the ligand located in the binding site; $E_{max}$ – a the highest binding energy in the trajectory; $E_{surface}$ – the binding energy of the ligand lo at the protein surface; $E_a$ – activation energy of association or dissociation ($E_{max}$ - $E_{surface}$; describes the difficulty of getting through the tunnel); $\Delta E_{BS}$ – difference of the binding energies of the ligand in the active site and at the surface (corresponds to enthalpy; thermodynamics).

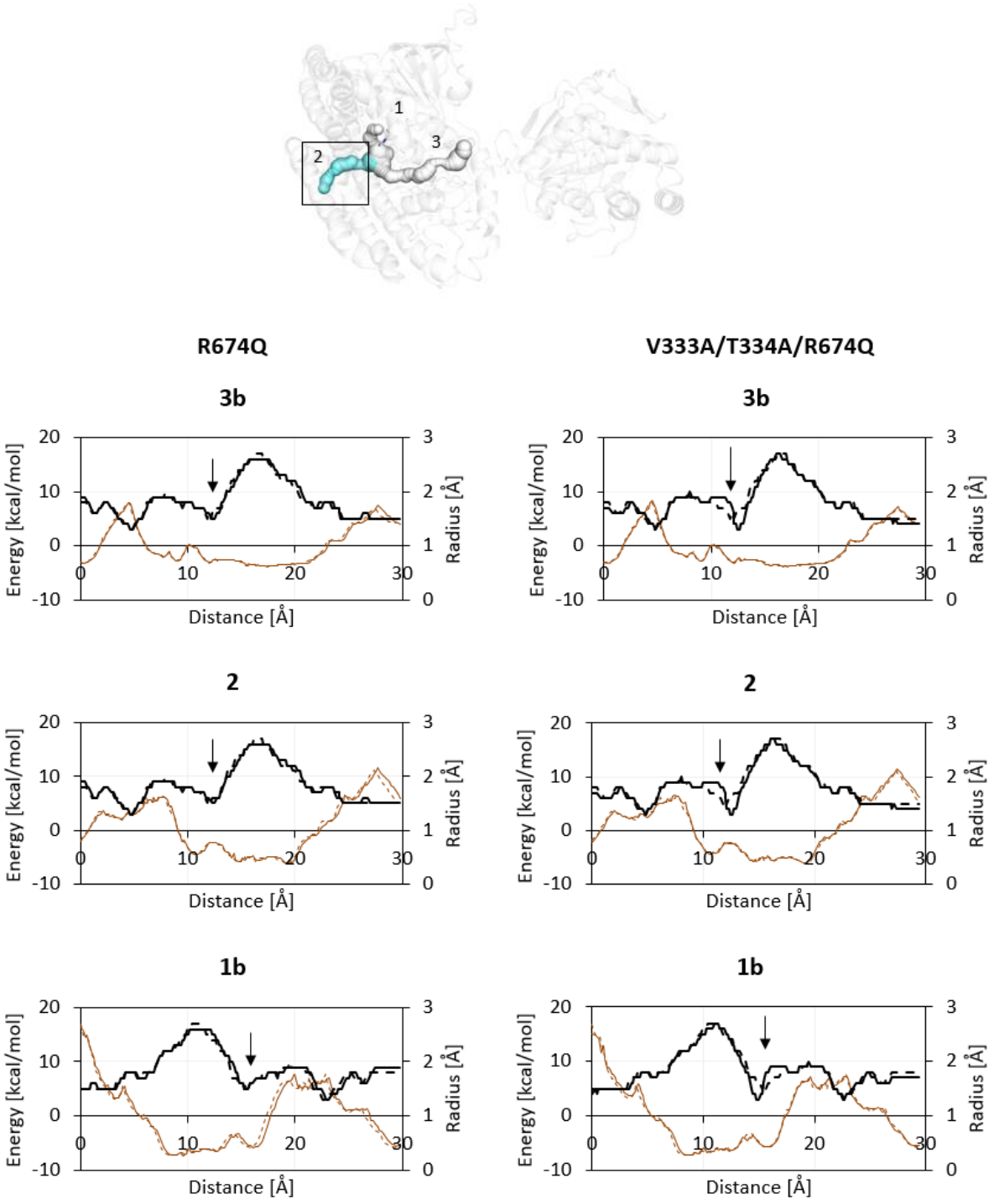



**Figure S12** Influence of R674Q on ligand transport (Caver Web 1.0)[42] through tunnel **#2** (shown in cyan in PigC structural model) in variants R674Q and V4 (V333A/T334A/R674Q) with substrates 2,3-dimethyl-1*H*-pyrrole (**3b**) and MBC (**2**) from the surface to the active site (*in*), and the product prodiginine **1b** from the active site towards the protein surface (*out*). The ligand binding energy is plotted in orange, the tunnel radius in black. The wild type trajectory is implemented in the plots as dashed lines.

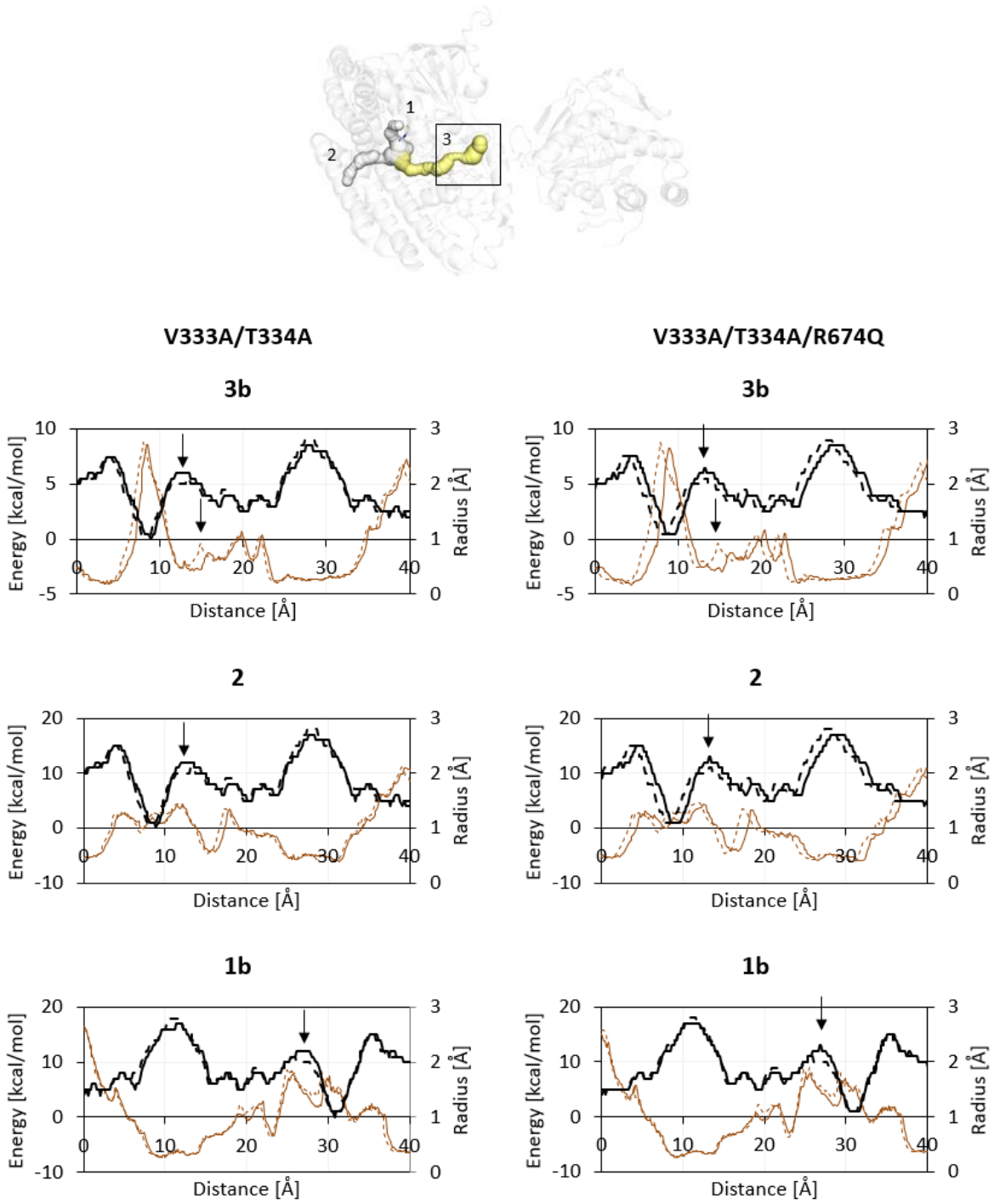



**Figure S13** Influence of V333A and T334A on ligand transport (Caver Web 1.0)[42] through tunnel **#3** (shown in yellow) in variants V1 (V333A/T334A) and V4 (V333A/T334A/R674Q) with substrates 2,3-dimethyl-1*H*-pyrrole (**3b**) and MBC (**2**) from the surface to the active site (*in*), and the product prodiginine **1b** from the active site towards the protein surface (*out*). The ligand binding energy is plotted in orange, the tunnel radius in black. The wild type trajectory is implemented in the plots as dashed lines.

Tunnel cross sections were computed by Caver analyst 2.0[43,44] at the position of V333/A333 and T334/A334 in tunnel 3. The tunnel areas were estimated by the ImageJ software.

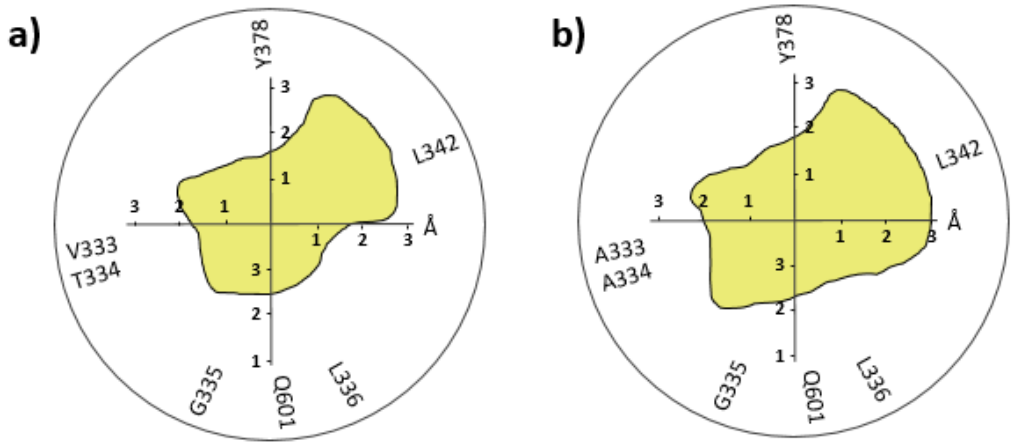



**Figure S14** Cross sections of **a)** WT tunnel 3 and **b)** Variant V1 (V333A/T334A) tunnel 3 computed by Caver Analyst 2.0;[43,44] The cross sections were taken close to the tunnel bottleneck, indicated by the red arrow in **Figure 4b**) in the main article.

## Proposed PigC catalytic mechanism

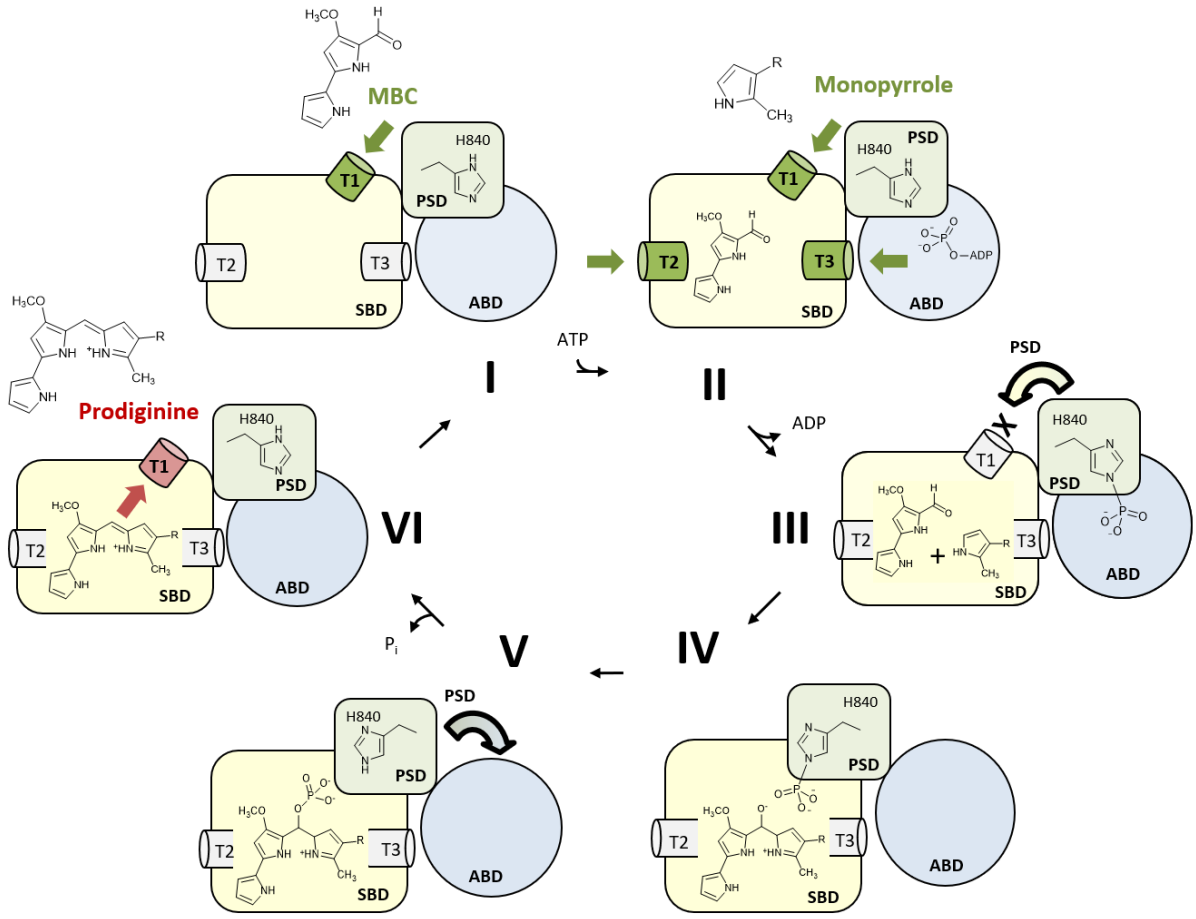



**Figure S15** Proposed PigC catalytic reaction mechanism. The ATP-binding domain (**ABD**; residues 1-298) is shown as blue circle; the substrate-binding domain (**SBD**; residues 299-779) as yellow rectangle, and the phosphohistidine swivel domain (**PSD**; residues 780-888) as green rectangle. **SBD** access tunnels **T1**-**T3** are shown as cylinders. (**I**) The substrate MBC enters through tunnel 1 (**T1**) when the **PSD** is orientated towards the **ABD**. Upon (**II**) the entrance of the monopyrrole substrate into the substrate binding pocket, both substrates react spontaneously under formation of the oxyanion intermediate (**III**-**IV**), as suggested by Picott *et al* (2020).[45] After the **PSD** (including the phosphorylated H840) performed the conformational flip towards the **SBD**, it delivers the ATP-derived phosphate group that is added at the oxyanion atom (**IV**-**V**). Finally, the phosphate group is released, resulting in the product formation. Upon the final flip of the **PSD** towards its starting position at the **ABD** (**V**), the prodiginine product is released through tunnel 1 (**VI**). The catalytic scheme is adapted based on proposed mechanisms by Chawrai *et al.* (2012) and Picott *et al.* (2020).[45,46]

## III. PigC sequences

>PigC_DNA (codon optimised for *Pseudomonas putida*)

ATGAACCCGACCCTGGTGGTGGAACTGAGCGGTGACAAGACCCTGGAACCGCACCGCCTGGGTGGCAAGGCCCAC
AGCCTGAACCACCTGATCCAAGCCGGTCTGCCGGTGCCGCCAGCCTTCTGCATCACCGCGCAGGCCTACCGCCAG
TTCATCGAGTTCGCCGTGCCAGGTGCGCTGCTGGACACCGGTGCGCCAGGCAACGTGCGCGACATGATCCTGAGC
GCGGCGATCCCAGCGCCACTGGATCTGGCCATCCGCCACGCCTGCAAGCAGCTGGGTGATGGTGCCAGCCTGGCC
GTGCGTAGCTCGGCCCTGGAAGAGGACGGTCTGACCCACAGCTTCGCCGGTCAGTACGACACCTACCTGCACGTG
CGTGGCGACGACGAGGTGGTGCGCAAGGTGCAAAGCTGCTGGGCCAGCCTGTGGGCCGAACGTGCCGCGCAGTAT
AGCCGCACCAGCGCCGCGCAATCGGATATCGCCGTGGTGCTGCAGATCATGGTGGACGCCGATGCCGCCGGTGTG
ATGTTCACCCAGGATCCGCTGACCGGTGACGCCAACCACATCGTGATCGACTCGTGCTGGGGTCTGGGCGAAGGC
GTGGTGAGCGGTCAGGTGACCACCGACAGCTTCATCCTGGACAAGGCCAGCGGTGAGATCCGCGAGCAGCAGATC
CGCCACAAGCCGCACTACTGCCAGCGCGATCCGCAGGGTCGTGTGACCCTGCTGCAAACCCCAGAAGCGCGTCGC
GACGCCCCATCGCTGACCCCAGAACAGCTGCAGCAACTGGCCCGTCTGGCGCGTCAGACCCGCATGATCTACGGT
GCCGAGCTGGACATCGAGTGGGCCGTGAAGGACGACCGCGTGTGGCTGCTGCAAGCCCGTCCGATCACCACCCAA
GCGAAGCCGGTGCAGATGCTGTACGCGAACCCGTGGGAGAGCGACCCAGCGATCAAAGAGCGTGCCTTCTTCAGC
CGCATGGATACCGGTGAAATCGTGACCGGTCTGATGACCCCACTGGGCCTGTCGTTCTGCCAGTTCTACCAGAAG
CACATCCACGGTCCAGCCATCAAGACCATGGGCCTGGCCGACATCGGCGACTGGCAGATCTACATGGGCTACCTG
CAGGGCTACGTGTACCTGAACATCAGCGGCAGCGCCTACATGCTGCGTCAGTGCCCACCGACCCGTGACGAGATG
AAGTTCACCACCCGCTACGCCACCGCGGACATCGACTTCAGCGGCTACAAGAACCCGTATGGCCCAGGCGTGCAA
GGCTGGGCCTACCTGAAAAGCGCCTGGCACTGGCTGAAGCAGCAGCGCCACAACCTGCGTAGCGCCGGTGCCACC
GTGGACGCCATGATCGCCCTGCGCCAGCGTGAAACCCGTCGCTTCCTGGCGCTGGACCTGACCACCATGACCCAC
CAAGAGCTGGAACGCGAGCTGAGCCGTATCGACGGCTACTTCCTGGACAGCTGCGCCGCCTATATGCCGTTCTTC
CTGCAGAGCTTCGCCCTGTACGACGCGCTGGCCCTGACCTGCGAGCGCTACCTGAAAGGCCGTGGCAACGGTCTG
CAGAACCGCATCAAGGCGAGCATGAACAACCTGCGCACCATCGAGGTGACCCTGGGCATCCTGAGCCTGGTGGAA
ACCGTGAACCGCCAGCCAGCGCTGAAGGCCGTGTTCGAACGCCACAGCGCCCAAGAACTGGTGACCGTGCTGCCG
ACCGATCCGGAAAGCCGTGCGTTCTGGCAGAGCGACTTCTCGGCCTTCCTGTTCGAGTTCGGTGCCCGTGGCCGT
CAAGAATTCGAACTGAGCCTGCCACGCTGGAACGACGACCCGAGCTACCTGCTGCAGGTCATGAAGATGTACCTG
CAACACCCGGTGGACCTGCACACCAAGCTGCGCGAAACCGAGCGCCTGCGCCATGAGGATAGCGCGACCCTGCTG
AAGGCCATGCCGTGGTTCGGTCGCATGAAGCTGAAATTCATCACCAAACTGTACGGCGTGATGGCCGAGCGTCGC
GAGGCGACCCGTCCAACCTTCGTGACCGAAACCTGGTTCTACCGTCGCATCATGCTGGAAGTGCTGCGTCGCCTG
GAAGCCCAAGGCCTGGTGAAGCAGGCCGACCTGCCGTACGTGGACTTCGAGCGCTTCCGTGCCTTCATGGCCGGT
GAGCAGTCGGCCCAAGAAGCCTTCGCCGCCGACCTGATCGAGCGCAACCGCCACCAACATCTGCTGAACCTGCAC
GCCGAGGAACCGCCAATGGCCATCGTCGGTGGCTACCAGCCACGCATGAAAGCCCCAACCGCCGAGAACGCCGCC
GGTATGCTGAGCGGTCTGGCCGCCTCGCCAGGTAAGGTGGTGGCCAAAGCGCGTGTGATCACCGACCTGCTGGCC
CAAGCGGGTGAGCTGCAGCCGAACGAGATCCTGGTGGCCCGTTTCACCGACGCCAGCTGGACCCCACTGTTCGCC
CTGGCCGCGGGTATCGTGACCGACATCGGTAGCGCCCTGAGCCACAGCTGCATCGTGGCCCGTGAGTTCGGCATC
CCAGCCGCCGTGAACCTGAAGAACGCGACCCAACTGATCAACTCGGGTGACACCCTGATCCTGGACGGCGACAGC
GGCACCGTCATCATCCAACGTGGCGAGCGTGCCGACGGCTGA

>PigC_Amino acids

MNPTLVVELSGDKTLEPHRLGGKAHSLNHLIQAGLPVPPAFCITAQAYRQFIEFAVPGALLDTGAPGNVRDMILS
AAIPAPLDLAIRHACKQLGDGASLAVRSSALEEDGLTHSFAGQYDTYLHVRGDDEVVRKVQSCWASLWAERAAQY
SRTSAAQSDIAVVLQIMVDADAAGVMFTQDPLTGDANHIVIDSCWGLGEGVVSGQVTTDSFILDKASGEIREQQI
RHKPHYCQRDPQGRVTLLQTPEARRDAPSLTPEQLQQLARLARQTRMIYGAELDIEWAVKDDRVWLLQARPITTQ
AKPVQMLYANPWESDPAIKERAFFSRMDTGEIVTGLMTPLGLSFCQFYQKHIHGPAIKTMGLADIGDWQIYMGYL
QGYVYLNISGSAYMLRQCPPTRDEMKFTTRYATADIDFSGYKNPYGPGVQGWAYLKSAWHWLKQQRHNLRSAGAT
VDAMIALRQRETRRFLALDLTTMTHQELERELSRIDGYFLDSCAAYMPFFLQSFALYDALALTCERYLKGRGNGL
QNRIKASMNNLRTIEVTLGILSLVETVNRQPALKAVFERHSAQELVTVLPTDPESRAFWQSDFSAFLFEFGARGR
QEFELSLPRWNDDPSYLLQVMKMYLQHPVDLHTKLRETERLRHEDSATLLKAMPWFGRMKLKFITKLYGVMAERR
EATRPTFVTETWFYRRIMLEVLRRLEAQGLVKQADLPYVDFERFRAFMAGEQSAQEAFAADLIERNRHQHLLNLH
AEEPPMAIVGGYQPRMKAPTAENAAGMLSGLAASPGKVVAKARVITDLLAQAGELQPNEILVARFTDASWTPLFA
LAAGIVTDIGSALSHSCIVAREFGIPAAVNLKNATQLINSGDTLILDGDSGTVIIQRGERADG

# IV. References


1 A. S. Klein, A. Domröse, P. Bongen, H. U. C. Brass, T. Classen, A. Loeschcke, T. Drepper, L. Laraia, S. Sievers, K.-E. Jaeger and J. Pietruszka, *ACS Synth. Biol.*, 2017, **6**, 1757–1765.

2 A. S. Klein, H. U. C. Brass, D. P. Klebl, T. Classen, A. Loeschcke, T. Drepper, S. Sievers, K.-E. Jaeger and J. Pietruszka, *ChemBioChem*, 2018, **19**, 1545–1552.

3 H. U. C. Brass, A. S. Klein, S. Nyholt, T. Classen and J. Pietruszka, *Adv. Synth. Catal.*, 2019, **361**, 2659–2667.

4 S. Brands, H. U. C. Brass, A. S. Klein, J. Pietruszka, A. J. Ruff and U. Schwaneberg, *Chem. Commun.*, 2020, **56**, 8631–8634.

5 V. de Lorenzo, L. Eltis, B. Kessler and K. N. Timmis, *Gene*, 1993, **123**, 17–24.

6 K. E. Nelson, C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. a. P. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. C. Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzez, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. Eisen, K. N. Timmis, A. Düsterhöft, B. Tümmler and C. M. Fraser, *Environ. Microbiol.*, 2002, **4**, 799–808.

7 W. Wang and B. A. Malcolm, *BioTechniques*, 1999, **26**, 680–682.

8 H. H. Hogrefe, J. Cline, G. L. Youngblood and R. M. Allen, *BioTechniques*, 2002, **33**, 1158–1165.

9 A. P. Jacobus and J. Gross, *PLOS ONE*, 2015, **10**, e0119221.

10 K.-H. Choi, A. Kumar and H. P. Schweizer, *J. Microbiol. Methods*, 2006, **64**, 391–397.

11 J. Yang and Y. Zhang, *Nucleic Acids Res.*, 2015, **43**, W174–W181.

12 A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, 725–738.

13 J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nat. Methods*, 2015, **12**, 7–8.

14 P. J. Stogios, G. Cox, P. Spanogiannopoulos, M. C. Pillon, N. Waglechner, T. Skarina, K. Koteva, A. Guarné, A. Savchenko and G. D. Wright, *Nat. Commun.*, 2016, **7**, 11343.

15 D. Eisenberg, R. Lüthy and J. U. Bowie, *Methods Enzymol.*, 1997, **277**, 396–404.

16 C. Colovos and T. O. Yeates, *Protein Sci. Publ. Protein Soc.*, 1993, **2**, 1511–1519.

17 R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.

18 R. W. W. Hooft, G. Vriend, C. Sander and E. E. Abola, *Nature*, 1996, **381**, 272–272.

19 J. Pontius, J. Richelle and S. J. Wodak, *J. Mol. Biol.*, 1996, **264**, 121–136.

20 R. Lüthy, J. U. Bowie and D. Eisenberg, *Nature*, 1992, **356**, 83–85.

21 J. U. Bowie, R. Lüthy and D. Eisenberg, *Science*, 1991, **253**, 164–170.

22 M. Wiederstein and M. J. Sippl, *Nucleic Acids Res.*, 2007, **35**, W407–W410.

23 M. J. Sippl, *Proteins Struct. Funct. Bioinforma.*, 1993, **17**, 355–362.

24 C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci.*, 2018, **27**, 293–315.

25 G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99.

26 E. Krieger and G. Vriend, *Bioinformatics*, 2014, **30**, 2981–2982.

27 O. Buß, J. Rudat and K. Ochsenreither, *Comput. Struct. Biotechnol. J.*, 2018, **16**, 25–33.

28 R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369–387.

29 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

30 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, N. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollmann (2014), AMBER 14, University of California, San Francisco.

31 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

32 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
33 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.
34 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
35 H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2016, **44**, W344–W350.
36 H. Ashkenazy, E. Erez, E. Martz, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2010, **38**, W529–W533.
37 G. Celniker, G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko and N. Ben-Tal, *Isr. J. Chem.*, 2013, **53**, 199–206.
38 I. Mihalek, I. Reš and O. Lichtarge, *J. Mol. Biol.*, 2004, **336**, 1265–1282.
39 O. Lichtarge, H. R. Bourne and F. E. Cohen, *J. Mol. Biol.*, 1996, **257**, 342–358.
40 O. Khersonsky, R. Lipsh, Z. Avizemer, Y. Ashani, M. Goldsmith, H. Leader, O. Dym, S. Rogotner, D. L. Trudeau, J. Prilusky, P. Amengual-Rigo, V. Guallar, D. S. Tawfik and S. J. Fleishman, *Mol. Cell*, 2018, **72**, 178-186.e5.
41 G. E. Crooks, G. Hon, J.-M. Chandonia and S. E. Brenner, *Genome Res.*, 2004, **14**, 1188–1190.
42 J. Stourac, O. Vavra, P. Kokkonen, J. Filipovic, G. Pinto, J. Brezovsky, J. Damborsky and D. Bednar, *Nucleic Acids Res.*, 2019, **47**, W414–W422.
43 A. Jurcik, D. Bednar, J. Byska, S. M. Marques, K. Furmanova, L. Daniel, P. Kokkonen, J. Brezovsky, O. Strnad, J. Stourac, A. Pavelka, M. Manak, J. Damborsky and B. Kozlikova, *Bioinformatics*, 2018, **34**, 3586–3588.
44 E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLOS Comput. Biol.*, 2012, **8**, e1002708.
45 K. J. Picott, J. A. Deichert, E. M. deKemp, V. Snieckus and A. C. Ross, *ChemBioChem*, 2020, **20**, 1036–1042.
46 S. Chawrai, N. Williamson, T. Mahendiran, G. P. Salmond and F. Leeper, *Chem. Sci.*, 2012, **3**, 447–454.