Supplementary files for manuscript

"Canavanine utilization via homoserine and hydroxyguanidine by a PLP-dependent γ-lyase in *Pseudomonadaceae* and *Rhizobiales*"

Franziskus Hauth[a,b], Hiltrun Buck[a], Marco Stanoppi[a] and Jörg S. Hartig[a,b*]

[a.] *Department of Chemistry, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany.*
[b.] *Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany.*
** To whom correspondence should be addressed. Tel: +49 7531 88 4575; Email: joerg.hartig@uni-konstanz.de*

Table of content:

Experimental Procedure

Media

Minimal M9 salt medium contained 8.5 g/l $Na_2HPO_4 \cdot 2H_2O$, 3 g/l KH_2PO_4, 0.5 g/l NaCl, 1 g/l NH_4Cl, 2 mM $MgCl_2$, 100 µM $CaCl_2$ and was supplemented with trace elements (0.1 mM EDTA, 0.03 mM $FeCl_3$, 6.2 µM $ZnCl_2$, 0.76 µM $CuCl_2$, 0.42 µM $CoCl_2$, 1.62 µM H_3BO_3; 0.08 µM $MnCl_2$) and vitamins (0.1 mg/l cyanocobalamin, 0.08 mg/l 4-aminobenzoic acid, 0.02 mg/l D-(+)-biotin, 0.2 mg/l niacin, 0.1 mg/l Ca-D-(+)-pantothenic acid, 0.3 mg/l pyridoxamine-chloride, 0.2 mg/l thiamindichloride). As carbon source either 10 mM canavanine or 13 mM glycerol was added. Minimal M9 salt medium agar plates were made by supplementing agar. For the overexpression of proteins in Escherichia coli cells were grown in lysogeny broth supplemented with kanamycin as selection marker. Minimal salt medium M8 only differed from M9 minimal salt medium in the omission of ammonium chloride.

Enrichment of canavanine-degrading bacteria

Environmental soil samples were taken in August 2020 from the legume runner bean (*Phaseolus vulgaris*), sampled on the Isle of Reichenau (47 ° 42'08.0 "N 9 ° 02'40.1" E). The soil/root samples were dissolved in Ringers' solution, diluted 1:100 and streaked out on selective M9 salt minimal medium agar plates supplemented with 1 mM canavanine. The next day, single colonies were transferred onto a new plate and the selection was repeated until homogenous colonies were obtained.

Identification of bacteria degrading canavanine via hydroxyguanidine and homoserine

Isolated strains were grown in M9 salt minimal medium with 10 mM canavanine as sole carbon source at 30°C, while shaking at 200 rpm. After 24 h the cultivation medium as well as the bacterial pellet were analysed by mass spectrometry to identify the levels of canavanine, guanidine and hydroxyguanidine. Bacterial pellet samples were lysed by sonification (1.5 s on/off, 3 minutes, 20% amplitude) on ice with a Branson digital sonifier 450 and centrifuged to remove cell debris. Afterwards samples were analysed by mass spectrometry.

Mass spectrometry

Prominence HPLC system with LCMS-2020 single quadrupole MS (Shimadzu) was applied. Prior to analysis 2 µl of sample was mixed with 18 µl buffer B (90% acetonitrile, 0.2% formic acid, 10 mM ammonium formate), 2 µl of this solution was injected. Guanidine, hydroxyguanidine and canavanine were separated using a Nucleodur HILIC column (250 mm length x 2 mm i.d., 3 µm particle size, Macherey-Nagel), which was equilibrated with buffer B and eluted with a linear gradient of 45% buffer A (10 mM ammonium formate, pH 3.0) over an 8-min period followed by an isocratic step of 45% buffer B for 6 min. The column was operated at 20.0 ± 0.1° C with a flow rate of 0.15 ml/min. MS detection was performed by single ion monitoring (SIM) of corresponding protonated ions in positive ionization mode. LC-MS data were analysed using the LabSolutions software (Shimadzu, Release 5.93). For quantification, stable isotope-labelled guanidine (^{13}C, $^{15}N_3$) was added to the samples. Calibration curves for hydroxyguanidine and canavanine were obtained for the measuring range and used for quantification of these substances.

Genome sequencing and 16S rRNA gene analysis

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Quiagen) following the manufacturer's protocol and further purified by agarose gel electrophoresis. Whole genome sequencing was conducted by Novogene (Cambridge, United Kingdom). A detailed protocol for library construction and sequencing can be found in the supplementary material. For 16S rRNA gene analysis the gene was amplified from gDNA by Phusion (NEB) PCR using Primers 16S_27 (5'-AGAGTTTGATCCTGGCTCA-3') and 16S_1492 (5'-CGGCTACCTTGTTACGAC-3') [1]. Amplicons were sequenced by GATC (Eurofins, Constance, Germany) and analysed using the EzBioCloud Database [2]. Average nucleotide identity was calculated using the TrueBac ID server [3] and digital DNA-DNA hybridization was conducted using the Type (Strain) Genome Server [4].

Proteome data

Cells were grown in M9 salt minimal media with 10 mM canavanine or 13 mM glycerol as sole carbon source at 30°C. Same amounts of cells were harvested after app. 18 h and were lysed as described before. The total protein amount was determined with the BCA Kit from Thermo Scientific according to the manufacturer's protocol. 50 µg of protein was used for each proteome analysis sample which was carried out by the Proteomics facility of the University of Konstanz. In short, protein samples were delivered in-gel, reduced by dithiothreitol and alkylated using chloroacetamide, followed by tryptic digest. Digested proteins were analyzed on a QExactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to an Easy-nLC 1200 Nanoflow-liquid chromatography system (Thermo Fisher Scientific, Bremen, Germany). Raw data was evaluated using the software proteome discoverer 1.4 (Thermo Fisher scientific) and compared to the whole proteome of *P. canavaninivorans*.

Construction and expression of upregulated enzymes

The full length genes of CanγL (MBJ2347155.1), HD (MBJ2346999.1), AC (MBJ2347154.1) and AH (MBJ2347000).1 and AAL (MBJ2345991.1) from *P. canavaninivorans* were amplified by Phusion (NewEngland Biolabs) PCR and either Gibson overhangs or NdeI and XhoI restriction sites were introduced with primers listed in SI Table 2. *P. canavaninivorans* gDNA was used as template. Inserts were cloned into the overexpression vector pet28a (KmR) bearing a N-terminal 6x HisTag (pet28a displayed in SI item 3) by standard restriction digestions and Quick LigationTM or by Gibson assembly® (NEB) following the manufactures protocols. For the rhizobial CanγL (WP_012555980.1) the corresponding codon optimized gene sequence with NdeI and XhoI restriction sites was synthesized by Twist Biosciences and inserted into the overexpression vector used above by cloning via restriction site enzymes NdeI/XhoI and Quick LigationTM. CanγL K213A was constructed by whole plasmid PCR on the CanγL wt vector and the resulting fragment was ligated by Quick LigationTM.

The resulting plasmids were transformed by electroporation into electro competent E. coli XL10 cells and grown overnight on selective medium. Single grown colonies were used for plasmid extraction (Zyppy Plasmid MiniPrep kit), verified by sequencing (GATC, Eurofins) and transformed into *E. coli BL21 (DE3)* cells.

Enzyme expression

For protein overexpression the respective strain was grown with suitable antibiotic overnight in lysogeny broth at 37°C, 200 rpm, diluted 1:100 and cultivated at 37°C, 200 rpm, to an OD600 of approx. 0.5. After induction with 0.5 mM isopropyl-β-D-thiogalactopyranosid the cultures were incubated at 18°C for approx. 16 hours. Cells were harvested by centrifugation and stored at -20°C. Active enzyme was purified by standard nickel metal affinity chromatography. In short, frozen pellet was resuspended in lysis buffer (20 mM Tris, 20 mM imidazole, 200 mM NaCl, EDTA-free protease inhibitor (cOmplete, Roche), 0.02 mg/ml lysozyme, pH 8.0) and left on ice for 30 minutes. Cell lysis was completed by 2 cycles of sonication (1.5 s on/off, 3 minutes, 20% amplitude) on ice, after which the lysate was centrifuged to remove cell debris and insoluble material. The supernatant fraction was filtered through a 0.2 µm filter and loaded onto high performance Ni-NTA-resin at a flow rate of 1 ml per minute. The resin was washed with buffer C (20 mM Tris, 20 mM imidazole, 200 mM NaCl, pH 8.0) and the protein eluted with buffer D (20 mM Tris, 500 mM imidazole, 200 mM NaCl, pH 8.0). The protein elution was buffer exchanged into buffer E (20 mM Tris, 200 mM NaCl, pH 8.0) using PD10-desalting columns (GE Healthcare), concentrated by Amicon centrifugal filters (Merck), snap frozen in liquid nitrogen and stored at -80°C upon further usage. SDS-page samples were taken from all steps to monitor enzyme purification and to check purity. CanγL was additionally processed by TEV cleavage to remove the His-Tag, followed by reversed Ni-NTA and gel filtration chromatography on a Superdex 200 16/600 column.

Enzymatic assays

To determine the kinetic parameters of CanγL (or RCanγL) 20 µl of different canavanine concentrations were placed in a 96-well transparent flat plate. The reaction was started with 180 µl of reaction buffer (0.1 µM CanγL, 2.4 µM HD, 0.5 mM NAD+, 100 mM

NaCl, 50 mM TrisHCl buffer, pH 8.0) and the absorbance at 340 nm was monitored. A calibration curve with different levels of NADH was measured simultaneously to enable the calculation of substrate turnover. To calculate KM and kcat the initial velocities were plotted over the respective canavanine concentration and fitted by Michaelis-Menten Kinetics using GraphPad Prism 6.0 and equation 1:

$$v = \frac{v_{max}[S]}{K_M + [S]} \qquad \text{equation 1}$$

where v corresponds to the reaction velocity, vmax to the maximum reaction velocity, S to the substrate concentration and KM to the substrate concentration at half vmax.

The in vitro conversion of canavanine to hydroxyguanidine measured by MS was conducted by mixing 10 mM canavanine with 0.1 µM CanγL (or RCanγL) in 50 mM TrisHCl buffer, 100 mM NaCl, pH 8.0. The reaction took place at room temperature. At different time points samples were taken and quenched with buffer B (see above). Afterwards, samples were directly injected into the MS quadrupole system.

Activity of K213A was tested by the NADH coupled assay described above with 25 mM canavanine in the reaction assay.

NMR measurements

Standard NMR samples consisted of 10 mM substrate in 50 mM TrisHCl, 100 mM NaCl, 200 µM PLP, pH 8.0. Enzymes (normally 0.1 µM) and cofactors (NAD^+ or $NADP^+$) were added as needed specifically for the experiment conducted. 7 µM trimethylsilylpropionate (TMSP) was added as an internal reference. ^{1}H-NMR measurements were conducted by the NMR facility of the University of Konstanz on a Bruker Avance III 600 MHz spectrometer with a TCI-H/C/N-triple resonance cryoprobe head. Spectra were measured at 300 K with 10% D_2O as field lock. Water suppression was achieved using the WATERGATE sequence. Spectra were analysed using TopSpin 4.0 and Mestrenova 14.2.0. For the hydrogen/deuterium exchange experiment a ^{1}H-decoupled natural abundance ^{13}C experiment was performed and the signal was locked on C_6D_6.

Homologous recombination of *P. canavaninivorans*

To generate a gene knockout mutant of *P.* canavaninivorans by homologous recombination we used the method described by Huang and Wilks[5]. In short, first the upstream and downstream sequences (500 nts on each site) of the respective gene to be deleted were cloned into the pEX18Gm sacB suicide plasmid (donation from Prof. Herbert Schweizer, University of Florida, USA) by Gibson assembly. The resulting plasmid was transformed into *P. canavaninivorans* by electroporation. Electro competent cells were obtained by repeatedly washing overnight grown *P. canavaninivorans* cells with 1 mM $MgSO_4$. After recovery in SOC medium cells were plated out on LB with gentamycin. Single colonies that integrated the chromosome by homologous recombination were identified by colony PCR using a primer that binds on the pEX18 Gm plasmid and a primer binding on the genome. Deletion mutants were obtained via sacB mediated sucrose counter selection and the recombination event was confirmed by sequencing.

UV/Vis spectra of the CanγL reaction

To monitor the characteristic bands of the PLP cofactor during the CanγL reaction we obtained UV-Vis spectra. The reaction was measured in a quartz cuvette (50 µM enzyme in 50 mM TrisHCl, 100 mM NaCl, pH 8.0) and started by the addition of 10 mM canavanine. The experiment was performed in a Cary 60 UV-Vis spectrometer (Agilent Technologies) in continuous scan mode at a rate of 24000 nm/min, over the range of 600-300 nm in 5nm steps at room temperature.

Alpha fold modelling

The primary sequence of CanγL was submitted to the alpha fold server [6–10] and the resulting first rank model was visualized and superimposed using PyMol version 2.0 [11].

Phylogenetic analysis of canavanine-γ-lyase

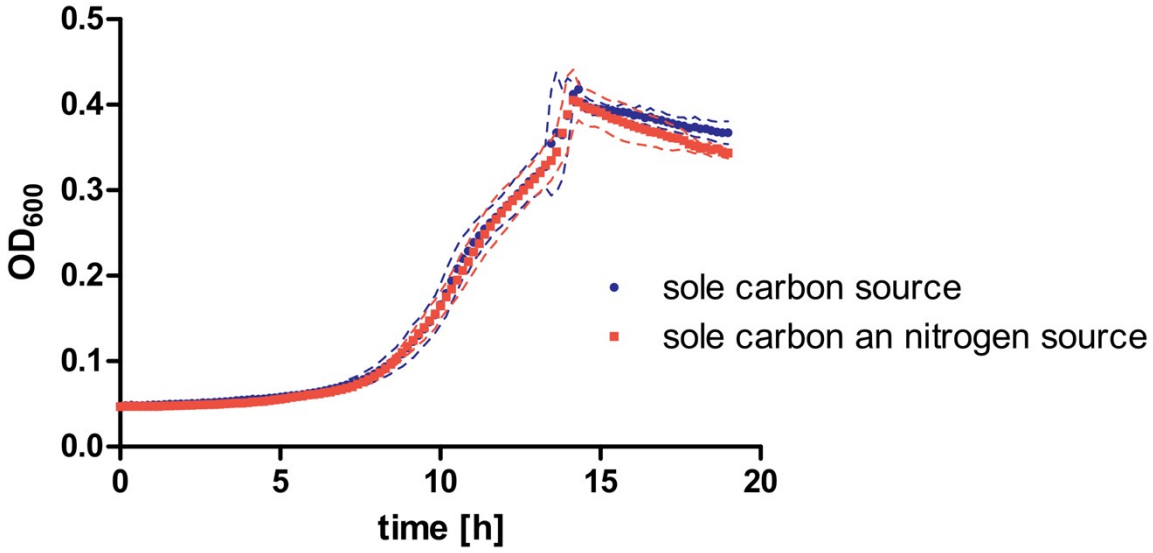
Homology search and multiple sequence alignment were performed with the Consurf platform [12]. Multiple sequence alignment was built using MAFFT on homologs collected from UNIREF90 database using BLAST/PSI-BLAST (3 iterations, E-value 0.0001). Subsequently, a phylogenetic tree was generated based on neighbour joining with BLOSUM62 [13]. The phylogenetic tree was illustrated with iTOL [14].

Microbe strains

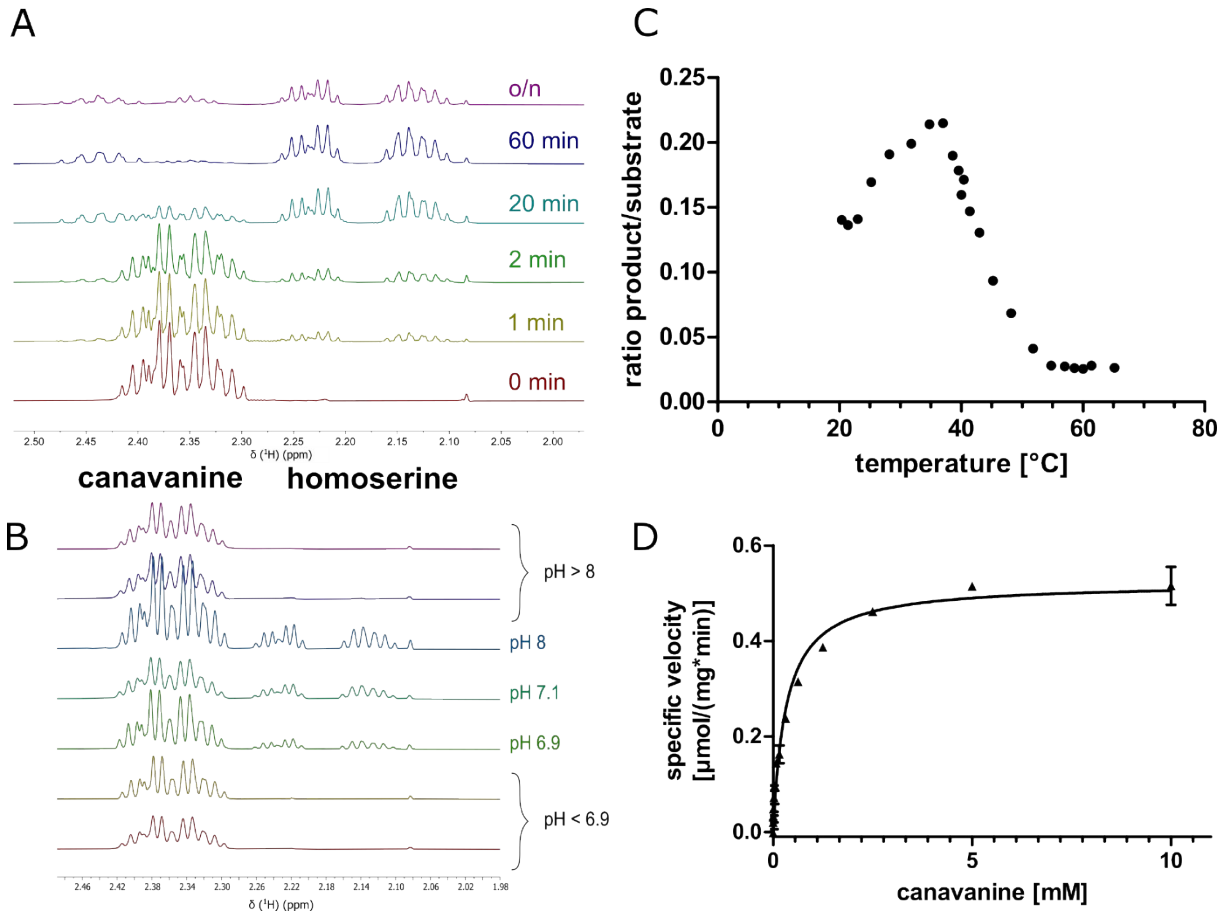
E.coli XL10 and *E.coli BL21(DE3)* cells were maintained in lysogeny broth medium at 37°C.

P. canavaninivorans (DSM No.: 112525) was maintained in lysogeny broth at 30°C. More detailed information on the bacterium can be found in its taxonomic and phenotypic description [15].

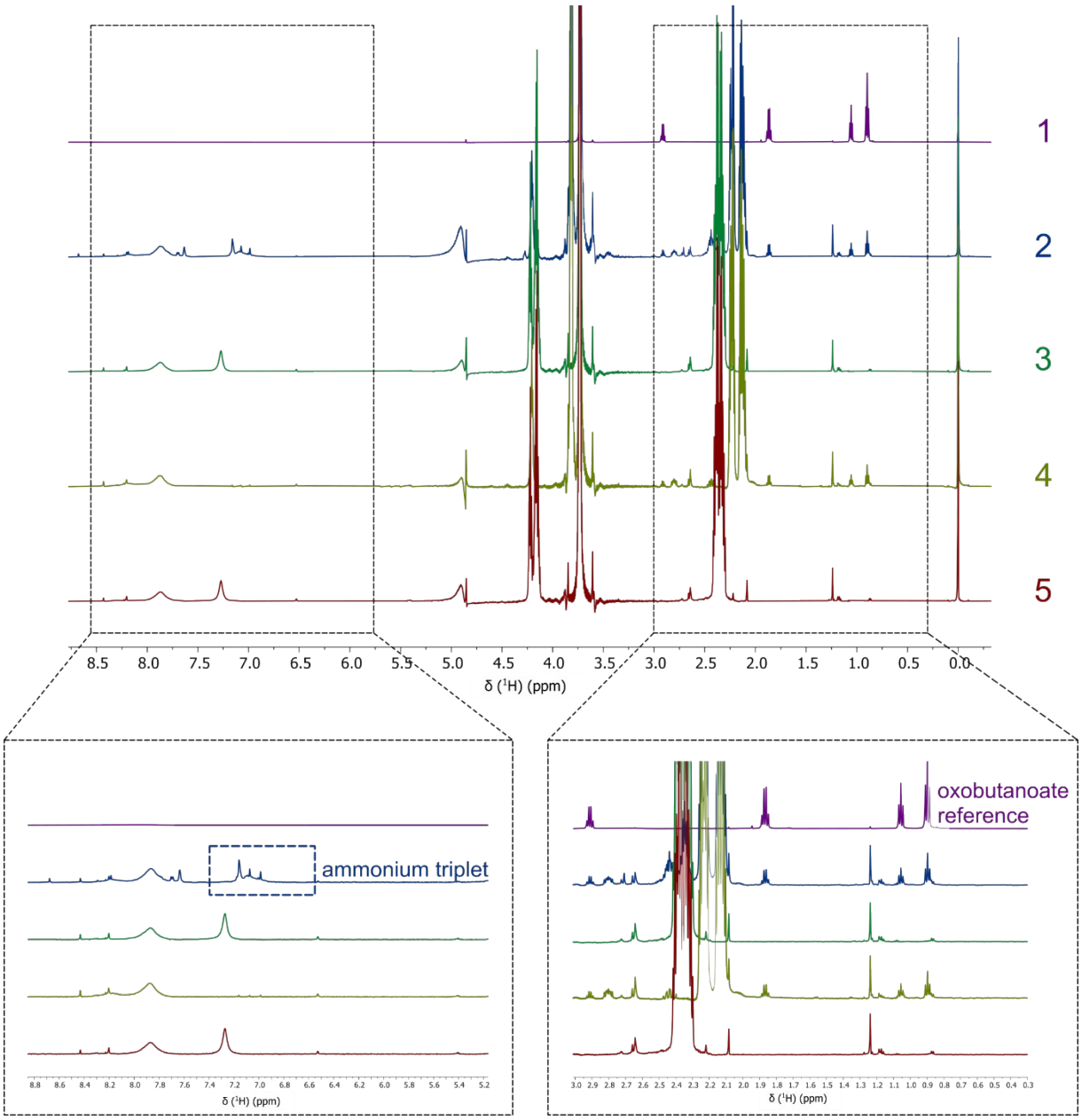
Rhizobium leguminosarum (DSM No.: 30141) was grown in Medium 98 (DSMZ) at 26°C.



SI figure 1: growth of *P. canavaninivorans* in minimal medium with 10 mM canavanine as sole carbon and nitrogen (red) or sole carbon (blue) source, n=3, dashed line = SD.

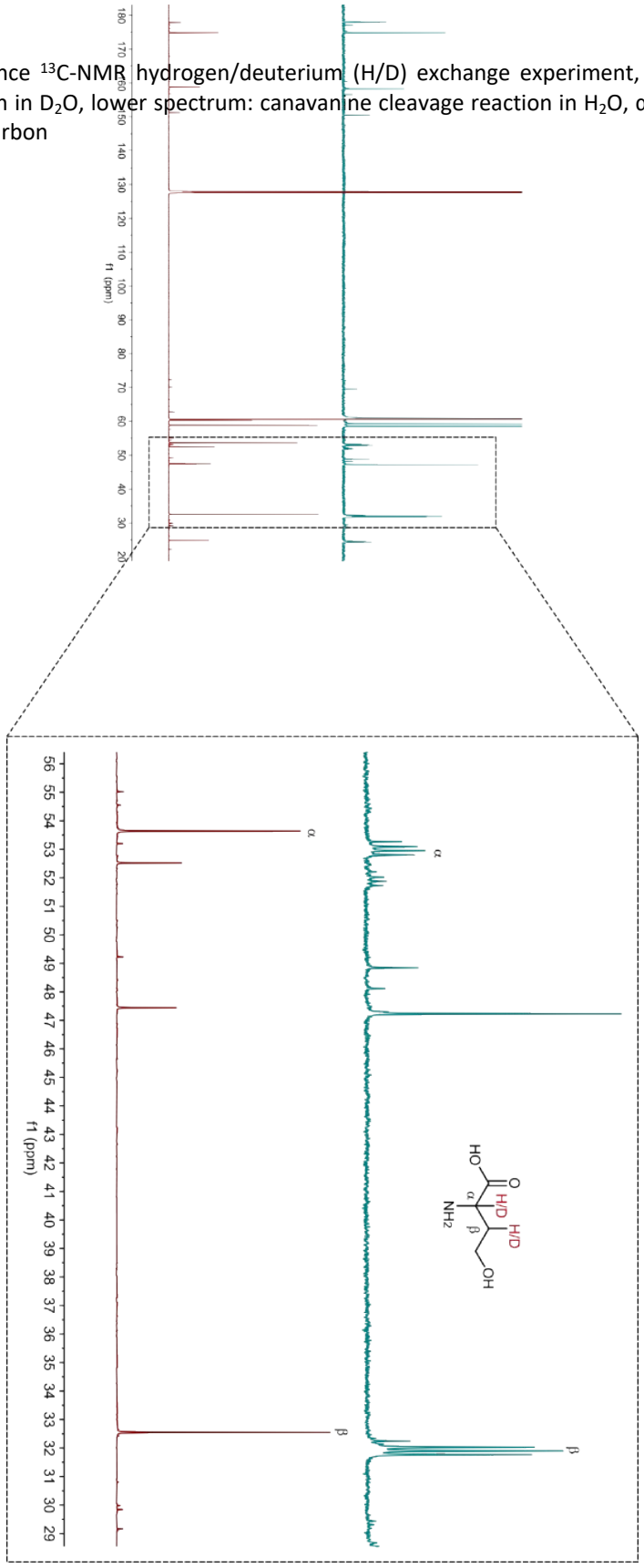


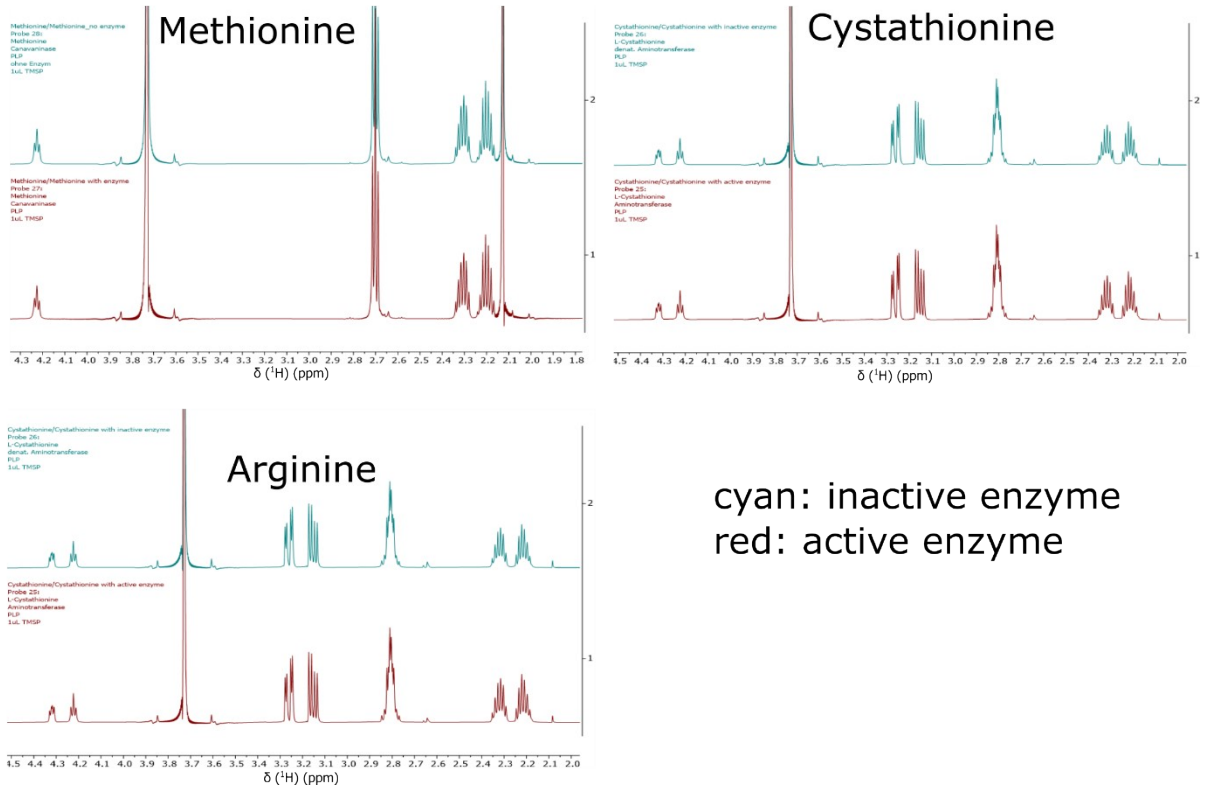
SI figure 2: CanγL characteristics: A: ^{1}H NMR spectra overlay of CanγL reactions stopped at different time points. **B:** ^{1}H NMR spectra overlay of CanγL reactions at different pH values, canavanine (2.42-2.30 ppm), homoserine (2.26 – 2.10 ppm). **C:** ratio of product/substrate conversion at different temperatures, CanγL reactions were stopped after a certain time by the addition of perchloric acid and measured by MS. **D:** specific velocities of CanγL at different substrate concentrations and Michaelis Menten kinetics fit.



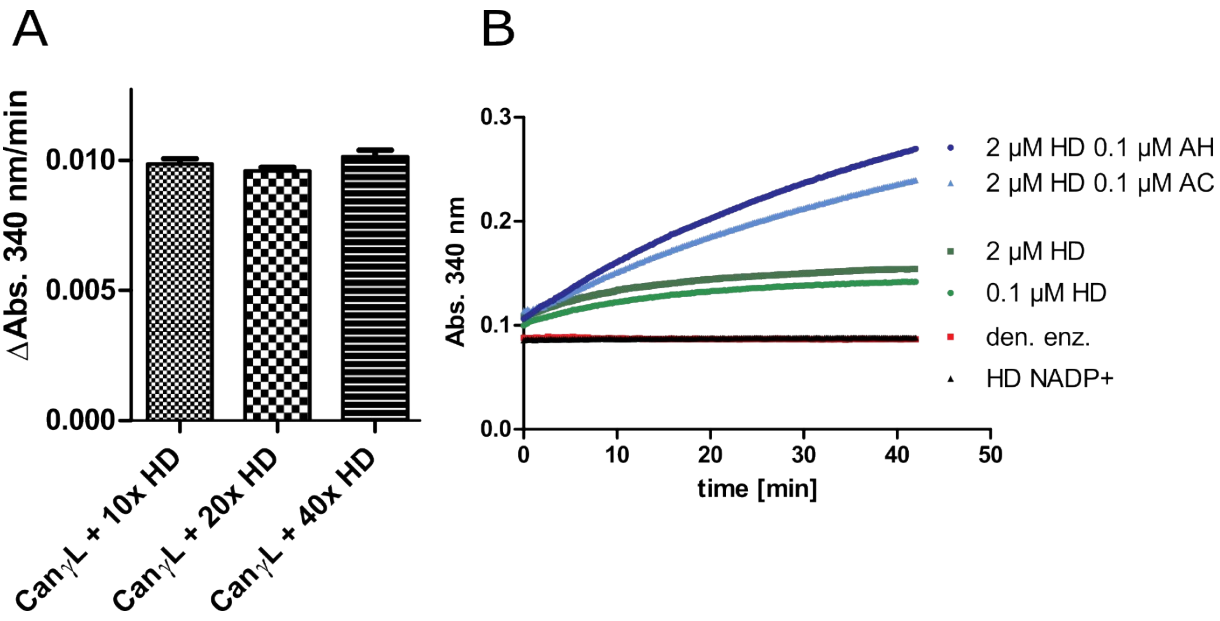
SI figure 3: ^{1}H-NMR spectra of CanγL reaction. **1:** oxobutanoate reference, **2:** canavanine with active CanγL **3:** canavanine with inactivated CanγL, **4:** homoserine with active enzyme **5:** canavanine reference, left bottom panel: formation of ammonium triplet, right bottom panel: formation of oxobutanoate

SI figure 4: natural abundance ^{13}C-NMR hydrogen/deuterium (H/D) exchange experiment, upper spectrum: canavanine cleavage reaction in D_2O, lower spectrum: canavanine cleavage reaction in H_2O, α/β correspond to the respective canavanine carbon

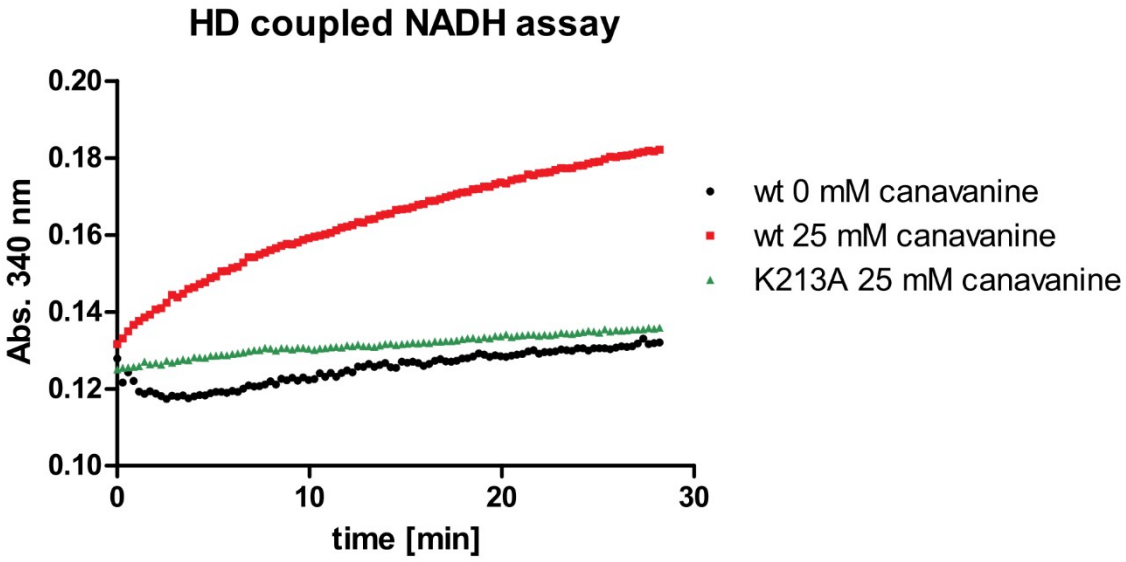




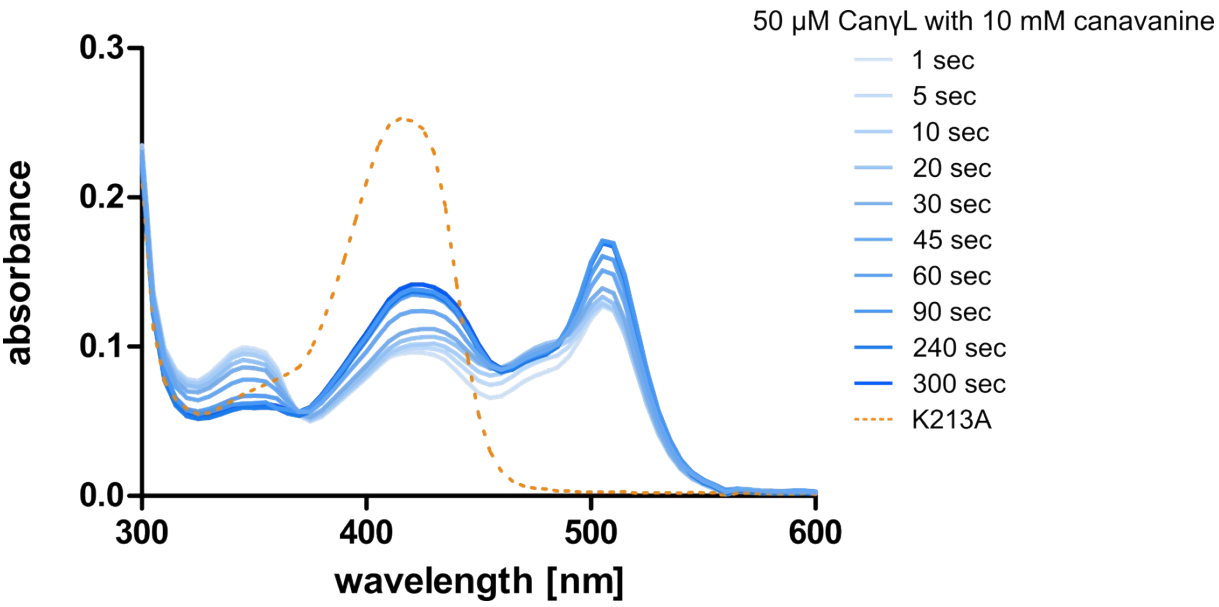
SI figure 5: ^{1}H-NMR spectra indicating substrate specificity of CanγL, spectra of methionine, cystathionine and arginine with active (red) or heat-inactivated (cyan), no differences in peak intensities or any additional or shifting peaks could be observed comparing active or heat-inactivated CanγL.



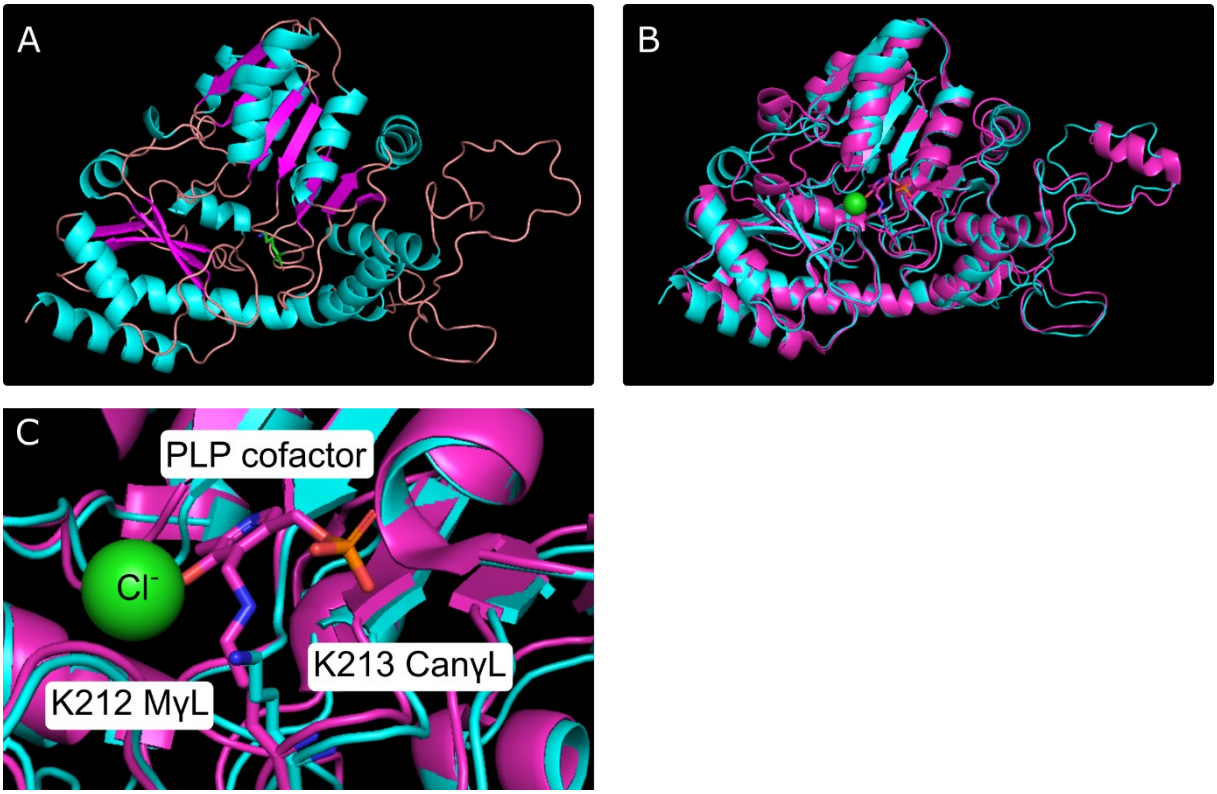
SI figure 6: A: determination of HD excess over CanγL for *in vitro* assay. **B:** NADH dependency of HD and further coupled *in vitro* reaction of HD with either of the aldehyde dehydrogenases AC or AH. Each data point represents the mean of a technical triplicate. Den. enz. corresponds to a heat inactivated HD sample.



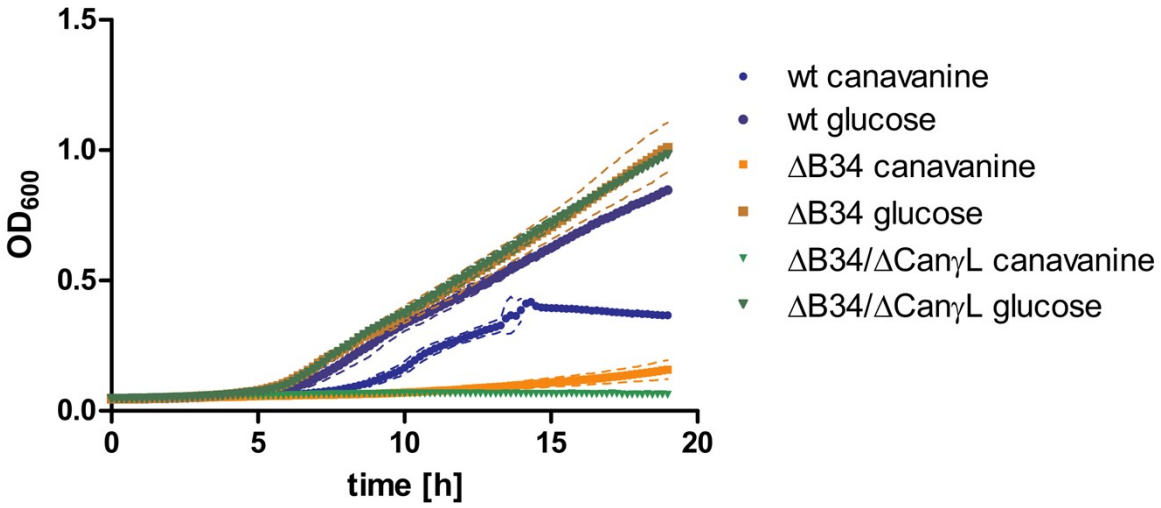
SI figure 7: NADH production by either wt or K213A CanγL in the HD coupled assay with 25 or 0 mM canavanine substrate. Each data point represents the mean of a technical triplicate.



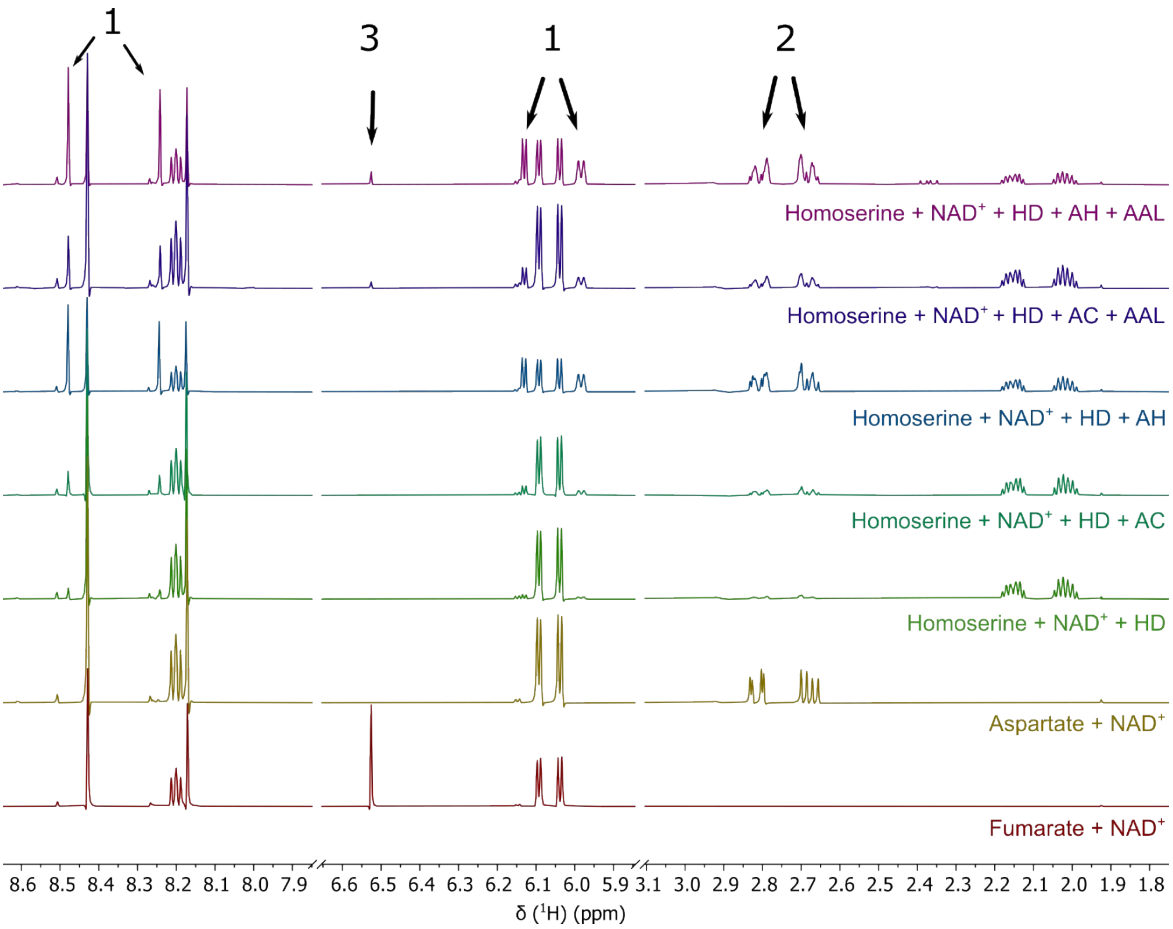
SI figure 8: UV/Vis spectra of 50 µM wt or K213A CanγL after the addition of 10 mM canavanine. For mutant K213A a representative spectrum is shown since no changes over reaction time were observable. The spectra display absorption peaks in accordance to characteristic PLP intermediates reported in the literature [16,17]: The peak at 420 nm presumably corresponds to the PLP aldimine, while the peak at 510 nm shows the quinonoid intermediate. The assignment of the peak at 350 nm is less straightforward but it could represent the canavanine ketamine. In general, UV/Vis spectra of PLP-dependent enzymes and their many reaction intermediates are often complicated to interpret and further mechanistic studies would be necessary in order to confirm the proposed assignments.



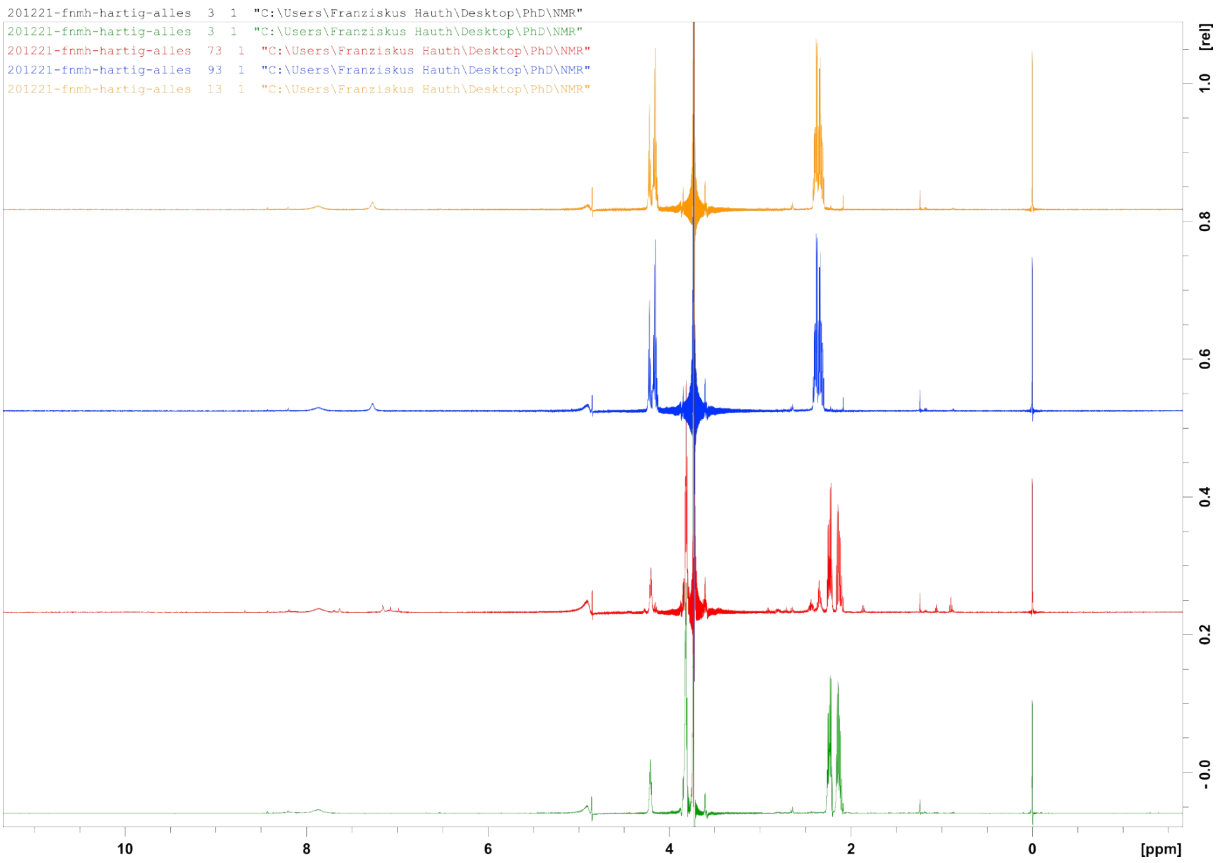
SI figure 9: Fold type identification of CanγL. **A:** Alphafold [6–10] model visualized using PyMol [11]. Colours represent secondary structures, bronze: loop, turquois: α-helices, pink: β-sheets. **B:** CanγL model superimposed with methionine-γ-lyase from *Clostridium sporonges* (PBD code 5DX5 chain A), green ball: Cl^- found in the crystal structure of MγL. **C:** zoom on the PLP cofactor with the PLP binding lysines of CanγL (K213) and MγL (K212).



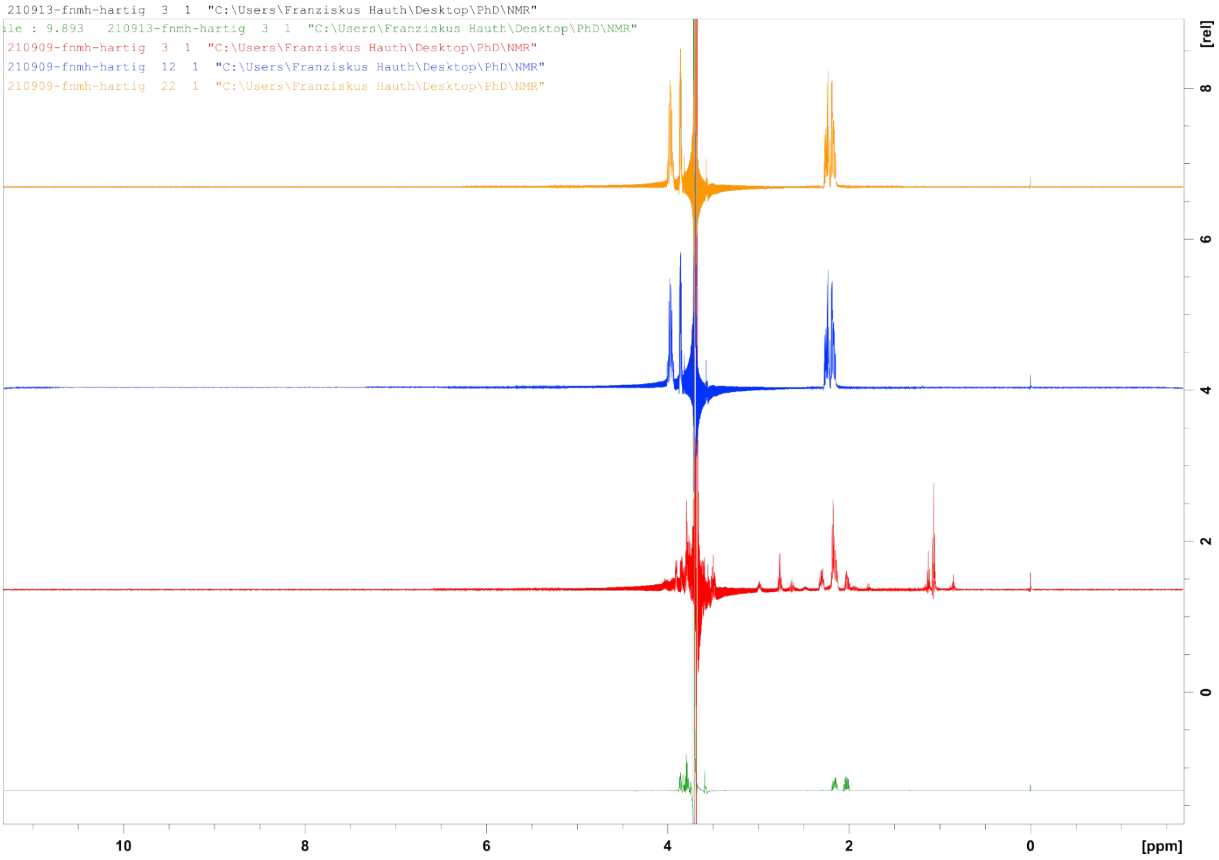
SI figure 10: growth curve of wt, ΔB3/4 and ΔB3/4Δ CanγL strains of *P. canavaninivorans*. The knockout strains were generated by homologous recombination as described in the methods part of this manuscript. Strains were grown on minimal medium with either canavanine or glucose as sole carbon source. Dashed line = SD, n = 3.



SI figure 11: full ^{1}H-NMR of the reaction of homoserine to fumarate via homoserine dehydrogenase (HD), either of the two aldehyde dehydrogenases (AC or AH) and ammonium-aspartate-lyase (AAL). 1: peaks corresponding to NADH, 2: peaks corresponding to aspartate, 3: peak corresponding to fumarate.



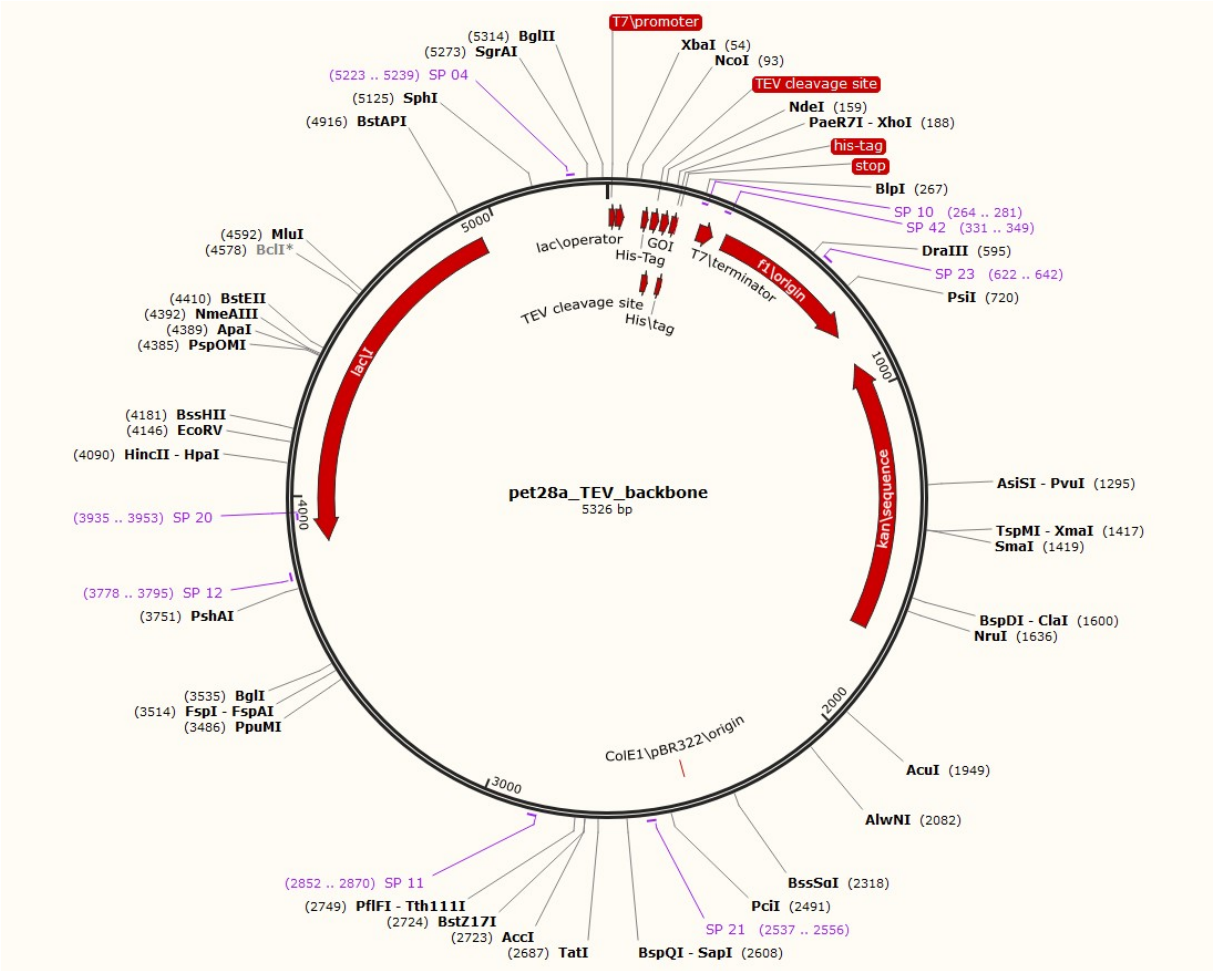
SI item 1: full ^{1}H NMR spectra of canavanine degradation by CanγL: **green:** 10 mM homoserine reference, **red:** reaction with 10 mM canavanine and active CanγL enzyme, **blue:** reaction with 10 mM canavanine and heat inactivated CanγL enzyme, **orange:** 10 mM canavanine reference. Peak at 0 ppm correspondes to the internal standard TSMP, signals around 3.75 ppm correspond to Tris buffer.



SI item 2: ^{1}H NMR spectra of canavanine degradation by RCanγL: **green:** 10 mM homoserine reference, **red:** reaction with 50 mM canavanine and active RCanγL enzyme, **blue:** reaction with 50 mM canavanine and heat inactivated RCanγL enzyme, **orange:** 50 mM canavanine reference. Peak at 0 ppm correspondes to the internal standard TSMP, signals around 3.75 ppm correspond to Tris buffer.

>SI item 3: pet28a_TEV_backbone (5326 bp) sequence and overview plasmid map

GCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCA
GCAGCCATCATCATCATCATCACAGCAGCGGCGGTGAGAATCTTTATTTTCAGGGCCATATGTTTTTTTTTTTTTTTTTTTTTTCTCGAGCACCACCACCACC
ACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACG
GGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTT
ACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCT
AAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCT
GATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT
GATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTT
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACT
CATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGC
AGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAA
GTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCA
TCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAAT
CGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGC
AGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTG
TAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATT
ATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCT
TGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA
AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA
TAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATG
AGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGA
AACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGC
CAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCT
TTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTA
CGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTG
ACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTC
TCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGA
TGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTT
GGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAA
CATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTT
AATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTA
CGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTG
CTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGC
TTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCG
CTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGA
TAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATT
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGT
TTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCAC
CAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCAT
TTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAG
ACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCT
TCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCT
GGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATC
GACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCA
ATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAA
ACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCT
GAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCA
TTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGG
GGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAAC
CGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCC


T7\promoter
XbaI (54)
NcoI (93)
TEV cleavage site
NdeI (159)
PaeR7I - XhoI (188)
his-tag
stop
BlpI (267)
SP 10 (264 .. 281)
SP 42 (331 .. 349)
DraIII (595)
SP 23 (622 .. 642)
PsiI (720)
(5314) BglII
(5273) SgrAI
(5223 .. 5239) SP 04
(5125) SphI
(4916) BstAPI
(4592) MluI
(4578) BclI*
(4410) BstEII
(4392) NmeAIII
(4389) ApaI
(4385) PspOMI
(4181) BssHII
(4146) EcoRV
(4090) HincII - HpaI
(3935 .. 3953) SP 20
(3778 .. 3795) SP 12
(3751) PshAI
(3535) BglI
(3514) FspI - FspAI
(3486) PpuMI
(2852 .. 2870) SP 11
(2749) PflFI - Tth111I
(2724) BstZ17I
(2723) AccI
(2687) TatI
BspQI - SapI (2608)
PciI (2491)
SP 21 (2537 .. 2556)
BssSaI (2318)
AlwNI (2082)
AcuI (1949)
BspDI - ClaI (1600)
NruI (1636)
TspMI - XmaI (1417)
SmaI (1419)
AsiSI - PvuI (1295)
lac\operator
His-Tag
GOI
T7\terminator
TEV cleavage site
His\tag
f1\origin
kan\sequence
lacI
ColE1\pBR322\origin
pet28a_TEV_backbone
5326 bp
1000
2000
3000
4000
5000

References

1 W. G. Weisburg, S. M. Barns, D. A. Pelletier and D. J. Lane, 16S ribosomal DNA amplification for phylogenetic study, *Journal of bacteriology*, 1991, **173**, 697–703.

2 S.-H. Yoon, S.-M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo and J. Chun, Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies, *International journal of systematic and evolutionary microbiology*, 2017, **67**, 1613–1617.

3 S. M. Ha, C. K. Kim, J. Roh, J. H. Byun, S. J. Yang, S. B. Choi, J. Chun and D. Yong, Application of the Whole Genome-Based Bacterial Identification System, TrueBac ID, Using Clinical Isolates That Were Not Identified With Three Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) Systems, *Annals of laboratory medicine*, 2019, **39**, 530–536.

4 J. P. Meier-Kolthoff and M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, *Nature communications*, 2019, **10**, 2182.

5 W. Huang and A. Wilks, A rapid seamless method for gene knockout in Pseudomonas aeruginosa, *BMC microbiology*, 2017, **17**, 199.

6 M. Mirdita, L. von den Driesch, C. Galiez, M. J. Martin, J. S"oding and M. Steinegger, Uniclust databases of clustered and deeply annotated protein sequences and alignments, *Nucleic acids research*, 2017, **45**, D170-D176.

7 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, ColabFold: Making Protein folding accessible to all, *Nature Methods*, 2022. DOI: 10.1038/s41592-022-01488-1.

8 M. Mirdita, M. Steinegger and J. Söding, MMseqs2 desktop and local web server app for fast, interactive sequence searches, *Bioinformatics*, 2019, **35**, 2856–2858.

9 A. L. Mitchell, A. Almeida, M. Beracochea, M. Boland, J. Burgin, G. Cochrane, M. R. Crusoe, V. Kale, S. C. Potter, L. J. Richardson, E. Sakharova, M. Scheremetjew, A. Korobeynikov, A. Shlemov, O. Kunyavskaya, A. Lapidus and R. D. Finn, MGnify: the microbiome analysis resource in 2020, *Nucleic acids research*, 2019. DOI: 10.1093/nar/gkz1035.

10 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Z'idek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021. DOI: 10.1038/s41586-021-03819-2.

11 L. L. Schrödinger, *The PyMOL Molecular Graphics System, Version 1.8*, 2015.

12 M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures, *Nucleic acids research*, 2005, **33**, W299-302.

13 S. Henikoff and J. G. Henikoff, Amino acid substitution matrices from protein blocks, *Proceedings of the National Academy of Sciences of the United States of America*, 1992, **89**, 10915–10919.

14 I. Letunic and P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic acids research*, 2019, **47**, W256-W259.

15 F. Hauth, H. Buck and J. S. Hartig, Pseudomonas canavaninivorans sp. nov., isolated from bean rhizosphere, *International journal of systematic and evolutionary microbiology*, 2022, **72**. DOI: 10.1099/ijsem.0.005203.

16 B. G. Caulkins, B. Bastin, C. Yang, T. J. Neubauer, R. P. Young, E. Hilario, Y. M. Huang, C. A. Chang, L. Fan, M. F. Dunn, M. J. Marsella and L. J. Mueller, Protonation states of the tryptophan synthase internal aldimine active site from solid-state NMR spectroscopy: direct observation of the protonated Schiff base linkage to pyridoxal-5'-phosphate, *Journal of the American Chemical Society*, 2014, **136**, 12824–12827.

17 L. Barra, T. Awakawa, K. Shirai, Z. Hu, G. Bashiri and I. Abe, β-NAD as a building block in natural product biosynthesis, *Nature*, 2021. DOI: 10.1038/s41586-021-04214-7.