Supplementary Information (SI) for Green Chemistry.
This journal is © The Royal Society of Chemistry 2025

Supporting Information

Structure-guided Engineering of Benzaldehyde Lyase for Efficient Synthesis of Dihydroxyacetone and One-pot Biosynthesis of 2-Amino-1,3-propanediol

Yue Fan,[a, b, c, ‡] *Xuemeng Li,*[a, b, ‡] *Yu Li,*[b, ‡] *Jinhui Feng,*[b] *Min Wang,*[a] *Cuiying Zhang,* [a] *Peiyuan Yao,*[b, *] *Qiaqing Wu,*[b, *] *and Dunming Zhu* [b, *]

[a] School of Biotechnology, Key Lab of Industrial Fermentation Microbiology of the Ministry of Education, State Key Laboratory of Food Nutrition and Safety, Tianjin University of Science and Technology, Tianjin 300457, China

[b] State Key Laboratory of Engineering Biology for Low-Carbon Manufacturing, National Center of Technology Innovation for Synthetic Biology, National Engineering Research Center of Industrial Enzymes and Tianjin Engineering Research Center of Biocatalytic Technology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[c] Haihe Laboratory of Synthetic Biology, Tianjin 300308, China

‡ These authors contributed equally

Content

1 Experimental Procedures

1.1 Materials

Reagents for molecular biology were obtained from Thermo Scientific and Omega Biotek. Ampicillin was purchased from Shanghai Yuanye Biotechnology Co. Ltd. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Beijing Solepol Technology Co. Ltd. Other commercial chemicals were purchased from Aladdin Industrial Corporation and Bide Pharmatech Ltd. The samples were analyzed by an Agilent high performance liquid chromatography (HPLC) system equipped with a variable wavelength detector (VWD), using Eclipse Plus-C18 column (250 mm x 4.6 mm, 5 μm, Agilent). ^{1}H and ^{13}C NMR spectra were recorded on a Bruker Avance III 400 MHz NMR spectrometer. The changes in absorbance were measured by SpectraMax M5 microplate reader (Molecular Devices). And the protein thermal stability was determined by Uncle (Unchained Labs).

1.2 Expression of recombinant ThDP-dependent enzymes and ω-transaminases (ω-TAs)

The genes of ThDP-dependent enzymes and ω-TAs were all synthesized by GENEWIZ, Inc. Recombinant ThDP-dependent enzymes and ω-TAs cloned into pET21a (+) were expressed in *E. coli* BL21 (DE3) cell. *E. coli* BL21 (DE3) cells containing plasmids encoding ThDP-dependent enzymes or ω-TAs of interest were cultivated overnight at 37°C in 10 mL lysogeny broth medium, supplemented with 100 μg/mL ampicillin. The culture was then transferred into 800 mL lysogeny broth medium containing 100 μg/mL ampicillin in a 2 L baffled flask and incubated at 37°C in a shaker at 220 rpm until OD_{600} reached 0.6-0.8. Then IPTG was added to the culture with a final concentration of 0.1 mM to induce protein expressions at 20ºC (25ºC for ω-TAs) for 18 h in a shaker at 220 rpm. After centrifugation at 6,000 *g* for 15 min at 4°C, the cells were harvested and stored at -20°C.

1.3 Screening of ThDP-dependent enzymes and *He*BAL mutants

E. coli BL21(DE3) colonies carried ThDP-dependent enzymes and *He*BAL mutants' plasmids were transferred into a 96 deep-well plate with 400 μL of lysogeny broth medium, containing 100 μg/mL ampicillin. The plate was incubated at 37°C with 800

rpm in a shaker overnight. For protein expression, 50 μL seed culture was transferred to a new 96 deep-well plate with 800 μL of fresh lysogeny broth medium, containing 100 μg/mL ampicillin. When the OD_{600} of new plates reached 0.6-0.8, IPTG was added with a final concentration of 0.1 mM. After incubated at 20°C for 18 h, the cells were harvested by centrifugation at 4,000 rpm for 5 min. The resulting cell pellets were stored at -80°C for the further study, which were used directly in the screening of enzyme activities.

The colorimetric assay was used to determine the activity of the enzyme with a slight modification.[1] The reaction mixture contained (200 μL) $MgSO_4$ (2.5 mM), ThDP (0.1 mM), and different concentrations of HCHO (10 mM, 20 mM, 40 mM or 80 mM) in potassium phosphate buffer (100 mM, pH 7.0) was added to a 96 deep-well plate of wet cells for ThDP-dependent enzymes or *He*BAL mutants. The reaction mixture was incubated at 30°C for 30 min. Then, tetrazolium red solution (20 μL, 0.2% 2,3,5-triphenyltetrazolium chloride in methanol) and 3 M NaOH (10 μL) were added to the reaction mixture (100 μL) by a good mixing. Finally, the reaction was monitored by SpectraMax M5 microplate reader after 2 min at 485 nm.

1.4 Screening of ω-TAs

The sensitive colorimetric assay was used for the screening of ω-TAs with a slight modification.[2] The assay was performed in 100 mM potassium phosphate buffer (pH 7.6) with 10 mM 1,3-dihydroxyacetone (DHA), 25 mM 2-(4-nitrophenyl)ethan-1-amine and 0.2 mM PLP. The cell-free extracts of ω-TAs was added to initiate the reaction. The total volume was 200 μL. The reaction was incubated at 30°C with 800 rpm for 60 min. The activities of ω-TAs were evaluated according to the coloration observed.

1.5 Construction of *He*BAL M6 libraries

The structural model of *He*BAL M6 was constructed in PyMOL (http://www.pymol.org) through site-directed mutagenesis,[3] by introducing three targeted mutations (A99S, Y395A and A476L) into the crystal structure of *He*BAL M3 (PDB ID: 8Y8M).[4] And the product DHA was docked into the active center of *He*BAL M6 based on CDOCKER protocol in Discovery Studio V4.1 (Biovia). The amino acid residues G26, I27, L111, Q112, A392, S417, M419, L476, T477, and A480 by a sphere of 5 Å of DHA were

selected for subsequent mutations. The mutant libraries of S417Y/A480X, V105X/S147Y/A480L, D106X/S147Y/A480L, E107X/S147Y/A480L, T108X/S147Y/A480L, N109X/S147Y/A480L, T110X/S147Y/A480L, G278X/S147Y/A480L, L279X/S147Y/A480L, A280X/S147Y/A480L, S147Y/A480L/E549X, S147Y/A480L/I552X, S147Y/A480L/M553X, V105X/A280T/S147Y/A480L, D106X/A280T/S147Y/A480L, E107X/A280T/S147Y/A480L, T108X/A280T/S147Y/A480L, N109X/A280T/S147Y/A480L, T110X/A280T/S147Y/A480L, G278X/A280T/S147Y/A480L, L279X/A280T/S147Y/A480L, A280T/S147Y/A480L/E549X, A280T/S147Y/A480L/I552X, A280T/S147Y/A480L/M553X, and V105X/A280T/S147Y/A480L/I552M were generated by applying site-saturation mutagenesis (168 clones of each library). On the basis of the *He*BAL M6 or related mutants' sequences, appropriate primers were designed (**Table S2**). The mutants genes were then constructed using the two-step PCR approach with FastPfu-DNA polymerase.[5] The PCR conditions were similar to that of the literature with a slight modification;[6] For short fragment: 95°C, 2 min, (95°C, 20 sec; 50°C, 20 sec; 72°C, 12 sec) × 30 cycles, 72°C, 5 min. For mega-PCR: 95°C, 2 min, (95°C, 20 sec; 55°C, 50 sec; 72°C, 4 min) × 30 cycles, 72°C, 5 min. The PCR products were digested with DpnI and then directly transformed into *E. coli* BL21(DE3) by heat-shocked for 1 min at 42°C. The cells were then transferred on ice for 5 mins before the addition of 600 μL of lysogeny broth medium. The cells were then incubated at 37°C for 1 h. The recovered cells were spread on agar plate, containing 100 μg/mL ampicillin and grown overnight at 37°C. The quality of the mutant library was evaluated by DNA sequencing (BGI) of several individual clones.

1.6 Screening of *He*BAL M6 mutant libraries

The methods for the protein expression and the activity determined were performed as described in the section **S1.3**.

1.7 Purification of *He*BAL mutants and ω-TAs

Purification of *He*BAL mutants

According to the literature, the recombinant *He*BAL mutants have been purified using affinity chromatography.[4] The protein was eluted using elution buffer containing 0.1

mM ThDP, 2.5 mM $MgSO_4$, 1 mM dithiothreitol (DTT), 300 mM imidazole, 50 mM potassium phosphate buffer (pH 7.0) and 300 mM NaCl. The protein was analyzed by SDS-PAGE. The final yield of purified enzyme for *He*BAL M11 was 384 mg from 51 g wet cells.

Purification of ω-TAs

Recombinant ω-TA was purified using affinity chromatography, and the expressed His6-tagged proteins were applied to a Ni^{2+} chelating affinity column for purification. The protein was eluted using potassium phosphate buffer (50 mM, pH 7.0) containing 500 mM imidazole and 500 mM NaCl. The purity protein was analyzed by SDS-PAGE and then dialyzed by desalting column. The final yield of purified enzyme for CV2025 was 710 mg from 30 g wet cells.

1.8 Sample derivatization protocols of DHA, GA and serinol

The derivatization of DHA and GA

DHA and GA were derivatized with *O*-benzylhydroxylamine hydrochloride in a slight modification.[7] The reaction sample of DHA or GA catalyzed by the enzyme (10 μL) was mixed with a solution of *O*-benzylhydroxylamine hydrochloride (990 μl, 50 mM, pH 4.5) in the water. After incubation at 30ºC for 60 min, the reaction mixture was detected by HPLC analysis which was performed on an Agilent 1200 HPLC equipped with an Eclipse Plus-C18 column (4.6 × 250 mm, 5 μm, Agilent, USA). The following condition was used to elute the samples: aqueous trifluoroacetic acid (TFA, 0.1% (v/v)) : acetonitrile = 70:30, flow rate 1 mL/min, detection at 215 nm, retention 40 min.

The derivatization of serinol

Serinol was derivatized with benzyl chloroformate (Cbz-Cl) in a slight modification.[8] Benzyl chloroformate (0.010 mmol) was added into serinol (0.005 mmol) in a solution of sodium hydrogen carbonate (1 mM, 1 mL). The mixture was stirred at room temperature for 60 min and then detected by HPLC with an Eclipse Plus-C18 column (4.6 × 250 mm, 5 μm, Agilent, USA) and UV detector at 210 nm. The system was operated at 30°C with aqueous trifluoroacetic acid (TFA, 0.1% (v/v)) and acetonitrile (70:30) as the mobile phase at the flow rate of 1 mL/min and the retention time was 20 min. The calibration curve of serinol was also produced after derivatized by Cbz-Cl.

And the concentrations of serinol were 0.5 mM, 1.0 mM, 2.0 mM, 3.0 mM, 5.0 mM, 7.0 mM, 8.0 mM and 10.0 mM (**Figure S4**).

1.9 Apparent kinetic parameters of *He*BAL mutants toward HCHO

The recombinant *He*BAL mutants were purified as described in the section **S1.7**. And one unit (U) of activity is described as the amount of protein that catalyzes the formation of 1 μmol DHA from HCHO per minute at 30°C. The reaction mixture contained 0.5 mg/mL purified enzyme of mutants (expect 1 mg/mL *He*BAL M6), $MgSO_4$ (2.5 mM), ThDP (0.1 mM), 20% DMSO, various concentrations of HCHO (10 mM-450 mM) in potassium phosphate buffer (100 mM, pH 7.0). The reaction mixture was initiated by the addition of the purified enzyme and the total volume was 1 mL. After being incubated at 30°C for 20 min, the obtained DHA was determined by HPLC following derivatized with *O*-benzylhydroxylamine hydrochloride, as described in the section **S1.8**. Kinetic parameters were deduced by non-linear regression analysis based on Michaelis-Menten kinetics using the program GraphPad Prism 9.0. All activities were measured in triplicates and error bars indicated the standard deviation. The calibration curve of DHA or GA was also produced after derivatized by *O*-benzylhydroxylamine hydrochloride. And the concentrations of DHA were 0.5 mM, 1.0 mM, 2.0 mM, 4.0 mM, 6.0 mM, 8.0 mM, 10.0 mM, 15.0 mM and 20.0 mM (**Figure S5**). The concentrations of GA were 0.1 mM, 0.25 mM, 0.5 mM, 1.0 mM, 2.0 mM, 4.0 mM, 5.0 mM, 8.0 mM and 10.0 mM (**Figure S6**).

1.10 Activity of TA toward DHA

The recombinant TA (CV2025) was purified as described in the section **S1.7**. And one unit (U) of activity is described as the amount of protein that catalyzes the formation of 1 μmol serinol from DHA per minute at 37°C. The reaction mixture contained 1.5 mg/mL purified enzyme of CV2025, 0.2 mM PLP, 200 mM DHA, 1 M IPA and 10% DMSO in potassium phosphate buffer (100 mM, pH 7.0). The reaction mixture was initiated by the addition of purified enzyme or cell-free extracts of CV2025 and the total volume was 1 mL. After being incubated at 37°C for 15 min, the obtained serinol was determined by HPLC following derivatized with Cbz-Cl, as described in the section **S1.8**.

1.11 Molecular dynamics simulations (MD)

The structure of *He*BAL M6 was generated using the mutagenesis tool included in PyMOL (http://www.pymol.org),[3] based on the crystal structure of *He*BAL M3.[4] And the homology model of *He*BAL M11 was constructed by AlphaFold 3.[9] Then the covalent complex of the cofactor ThDP and DHA (ThDP_DHA) was constructed and docked into the active center of the structure of *He*BAL M6 or *He*BAL M11 based on CDOCKER protocol in Discovery Studio V4.1 (Biovia). The Sobtop[10] and Multiwfn[11] software with the general AMBER force-field and RESP charges were used to obtain force field parameters of ThDP_DHA for MD simulations. MD simulations were performed using the GROMACS 2022.6 program with the Amber14SB force field.[12] The *He*BAL M6 and *He*BAL M11 were implemented periodic boundary condition across the system using a TIP3P water box and a small amount of Na^+ ions were introduced to neutralize the system. Subsequently, energy minimization of the constructed model was performed using the steepest descent method. All systems were simulated for 50 ns at NPT ensemble at 1 atm and 298.15 K. Pressure and temperature were controlled using the Velocity-rescale and parrinello-Rahman algorithms. The RMSD values of ThDP_DHA were calculated using MD trajectories of 50 ns for the *He*BAL M6 or *He*BAL M11 system. And the volumes of the binding pockets in *He*BAL M6 and *He*BAL M11 were calculated using ParKVFinder.[13] The substrate tunnels and electrostatic potential distribution of *He*BAL M6 and *He*BAL M11 were analyzed by CAVER 3.0[14] and PyMOL,[3] respectively.

1.12 General procedure for the synthesis of DHA

General procedure

The biotransformation containing different concentrations of HCHO, 0.1 mM ThDP and 2.5 mM $MgSO_4$ in 100 mM potassium phosphate buffer (pH 7.0, 1 mL) was performed. The reaction was initiated with the addition of 50 mg/mL wet cells of *He*BAL mutants. Then the reaction was incubated at 30°C and 120 rpm for 0.5-12 h. And the analytical yield of DHA was determined by HPLC analysis as mentioned in the section **S1.8**.

Optimization of reaction conditions

The reaction mixture (1 mL) contained 300 mM HCHO, DMSO (10%), 0.1 mM ThDP, 2.5 mM $MgSO_4$ and different concentrations of purified enzyme of *He*BAL M11 (1.0 mg/mL, 3.0 mg/mL and 4.5 mg/mL). The reaction systems were carried out at 30°C for

8 h. The analytical yield of DHA was determined by HPLC analysis as mentioned in the section **S1.8**.

1.13 General procedure for the synthesis of serinol

General procedure

The biotransformation containing different concentrations of DHA (100 mM and 200 mM), 10 eq. isopropylamine hydrochloride (IPA), 10% DMSO and 0.2 mM PLP in 100 mM potassium phosphate buffer (pH 7.6, 1 mL) was performed. The reaction was initiated with the addition of 15 mg/mL cell-free extracts of ω-TAs and incubated at 37°C for 20 h. The analytical yield of serinol was determined by HPLC analysis as mentioned in the section **S1.8**.

Optimization of reaction conditions

Optimization of the concentrations of amino donor IPA. The reaction mixture (10 mL) contained 200 mM DHA, different concentrations of IPA (5 eq. or 10 eq.), 10% DMSO and 0.2 mM PLP in 100 mM potassium phosphate buffer (pH 7.6). The reaction was initiated with the addition of 50 mg/mL cell-free extracts of CV2025 and carried out at 37°C for 12 h. The analytical yield of serinol was determined by HPLC analysis as mentioned in the section **S1.8**.

Optimization of the concentrations of DMSO. The reaction mixture (10 mL) contained 200 mM DHA, 1 M IPA, different concentrations of DMSO (0% or 10%) and 0.2 mM PLP in 100 mM potassium phosphate buffer (pH 7.6). The reaction was initiated with the addition of 50 mg/mL cell-free extracts of CV2025 and carried out at 37°C for 12 h. The analytical yield of serinol was determined by HPLC analysis as mentioned in the section **S1.8**.

Optimization of the pH values in the reaction system. The reaction mixture (5 mL) contained 100 mM DHA, 500 mM IPA, 10% DMSO and 0.2 mM PLP in different buffers (100 mM, pH 4-8.5). The reaction was initiated with the addition of 50 mg/mL cell-free extracts of CV2025 and carried out at 37°C for 8 h. The analytical yield of serinol was determined by HPLC analysis as mentioned in the section **S1.8**.

Preparative scale enzymatic syntheses of serinol

The enzymatic transamination of DHA was conducted using the enzyme CV2025. The typical 100 mL reaction system was executed containing 200 mM DHA, 10% DMSO, 1 M IPA, 0.2 mM PLP and 3 mg/mL purified enzyme of CV2025 in 100 mM potassium phosphate buffer (pH 7.6). The reaction was stirred at 37°C for 8 h. The analytical yield of serinol was determined by HPLC analysis as mentioned in the section **S1.8**. After the complete conversion of DHA, the reaction was quenched by a slowly addition of HCl (6 M, 5 mL) and the precipitate was removed by centrifugation. Then, the pH value of the supernatant was adjusted to 11 using NaOH (6 M) and the supernatant was concentrated in vacuo in order to remove IPA. Finally, the crude product was purified via an ion exchange column. A strong acid cation exchange resin D001 (from Shanghai Huazhen Technology Co., Ltd., 200 g) was prepared by washing with 2 L deionised water. The crude product was applied to the resin at 1 mL/min, and then washed with deionised water until the eluent was pH 7. Then, the product was eluted using 5% v/v NH_4OH and concentrated under reduced pressure to afford the pure product (1.43 g, isolated yield 79%).

1.14 One-pot, two-step cascade reaction for the synthesis of serinol

The C-C bond formation of HCHO was performed starting from 300 mM HCHO (0.577 mL, 13.0 M), 10% DMSO (2.5 mL), 0.1 mM ThDP (0.001 g), 2.5 mM $MgSO_4$ (0.007 g) and 4.5 mg/mL purified enzyme of *He*BAL M11 (112.50 mg) in potassium phosphate buffer (100 mM, pH 7.0). The total reaction volume was 25 mL and the reaction was executed at 30°C for 2 h. The analytical yield of DHA was detected by HPLC after derivatization with *O*-benzylhydroxylamine hydrochloride. When the conversion of HCHO reached >99%, the transamination of DHA was executed by the addition of 500 mM IPA (1.19 g), 0.2 mM PLP (1.24 mg), and 2.5 mg/mL purified enzyme of CV2025 (62.50 mg). After incubated at 37°C with 220 rpm for 6 h, the reaction was quenched by a slowly addition of HCl (6 M, 2 mL) and the precipitate was removed by centrifugation. The crude product was purified as mentioned in the section **S1.12**. Finally, 0.20 g serinol was obtained in 91% isolated yield. ^{1}H NMR (400 MHz, D_2O) δ = 3.61 - 3.51 (m, 2 H), 3.49 - 3.41 (m, 2 H), 2.93 - 2.81 (m, 1 H). ^{13}C NMR (100 MHz, D_2O) δ = 62.91, 53.02.

1.15 Green metrics

The green metrics calculated for the determination of the sustainability of this work and their comparison among selected references were calculated as follows:[15]

Product mass: Product obtained at the end of the reaction (g).

Waste mass: Is the product mass subtracted from the total mass in the bulk accounting for the reagents, the solvent (including water) and the catalysts.

Waste mass = (Total reagents mass (g) + total solvent mass (g) + total catalysts mass (g) – product mass (g))

E factor = Waste mass / Product mass

E factor $_{reagents}$ = Total reagents mass (g) / Product mass (g)

E factor $_{solvent}$ = Total solvent mass (g) / Product mass (g)

E factor $_{water}$ = Total water mass (g) / Product mass (g)

E factor $_{catalysts}$ = Total catalysts mass (g) / Product mass (g)

2 Results and Discussion

2.1 Supplementary Tables

Table S1. ThDP-Dependent enzymes and ω-TAs used in this study

Enzyme	Organism	Accession number
ATA	*Actinomycetia*	WP_030166319.1
CV2025	*Chromobacterium violaceum* DSM30191	WP_011135573.1
*Mi*TA	*Mycolicibacterium iranicum*	WP_024448131.1
*Mnf*TA	*Microbacterium* sp. NFIX05	WP_134125182.1
*Amb*TA	*Aeromicrobium* sp. CC-CFT486	WP_147685584.1
*Mba*TA	*Mycolicibacterium aromaticivorans*	WP_051660175.1
*Atm*TA	*Aspergillus thermomutatus*	XP_026618438.1
*Pa*TA	*Pseudonocardia ammonioxydans*	WP_093355841.1
*Ps*TA	*Prauserella*	WP_112275778.1
*Ab*TA	*Actinobacteria*	WP_030166319.1
*Sl*TA	*Shinella lacus*	WP_256119312.1

*He*BAL	*Herbiconiux* sp. SALV-R1	WP_171706903.1
BAL-2	*Polymorphobacter arshaanensis*	WP_135246357.1
BAL-3	*Streptomyces* sp. BK335	WP_133047913.1
BAL-5	*Streptomyces umbrinus*	WP_189844730.1
BAL-10	*Gordonia* sp. LAM0048	WP_064569730.1
BAL-15	*Cnuibacter physcomitrellae*	WP_085018062.1
BAL-23	*Streptomyces griseorubiginosus*	WP_208777857.1

Table S2. Primers used in site saturation mutagenesis

Mutants	Primers	Primer sequence from 5'-3'[a]
M6 G26X	Forward	GTTGCATTTGGCATTAAT**NNK**ATTCATATTG
	Reverse	CCACTCGATGCAATATACAGAACC
M6 I27X	Forward	GGCATTAATGGT**NNK**CATATTGATAGTATC
	Reverse	CCACTCGATGCAATATACAGAACC
M6 L111X	Forward	GGCATTAATGGT**NNK**CATATTGATAGTATC
	Reverse	CTGCACGTGCCAGTTCTTC
M6 Q112X	Forward	GATGAAACCAATACCCTG**NNK**GCAG
	Reverse	CTGCACGTGCCAGTTCTTC
M6 A392X	Forward	GATGGT**NNK**CTGACCGCGCTGTG
	Reverse	GAACCACAACCACAACCGGCAGAC
M6 S417X	Forward	CATGGTTATCTG**NNK**AGTATGGGCGTTGG
	Reverse	CACCACTATAACTGCCATTTTCCAGG
M6 M419X	Forward	CTGAGTAGT**NNK**GGCGTTGGCGTTG
	Reverse	CACCACTATAACTGCCATTTTCCAGG
M6 L476X	Forward	CGCACCGGTTGCAGATTATCTGTG

M6 T477X	Reverse	GGT**MNN**GCCCCATGCACGATTATTCAG
	Forward	CCGATGGATGCAGTGACCGCAATTGC
	Reverse	GTTCCTGGGCATGCAG**MNN**CAAG
M6 A480X	Forward	CCGGTTGCAGATTATCTGTGTCATGG
	Reverse	CAGTTCCTG**MNN**ATGCAGGGTCAAG
M6 S417Y/A480X	Forward	CCGGTTGCAGATTATCTGTGTCATGG
	Reverse	CAGTTCCTG**MNN**ATGCAGGGTCAAG
M6 V105X/S417Y/A480L	Forward	GCTGGGC**NNK**GATGAAACCAATAC
	Reverse	CCATCCAGAATACGTTCAAC
M6 D106X/S417Y/A480L	Forward	GGCGTT**NNK**GAAACCAATACC
	Reverse	CCATCCAGAATACGTTCAAC
M6 E107X/S417Y/A480L	Forward	GGCGTTGAT**NNK**ACCAATACC
	Reverse	CCATCCAGAATACGTTCAAC
M6 T108X/S417Y/A480L	Forward	CTGGGCGTTGATGAA**NNK**AATACCC
	Reverse	CCATCCAGAATACGTTCAAC
M6 N109X/S417Y/A480L	Forward	GATGAAACC**NNK**ACCCTGCAG
	Reverse	CCATCCAGAATACGTTCAAC
M6 T110X/S417Y/A480L	Forward	CCAAT**NNK**CTGCAGGCAGGTATTG
	Reverse	CCATCCAGAATACGTTCAAC
M6 G278X/S417Y/A480L	Forward	GGTTTACGTTTT**NNK**CTGACTACCGCACATGGC
	Reverse	CCATGACACAGATAATCTGCAACCG
M6 L279X/S417Y/A480L	Forward	CGTTTTGGC**NNK**ACTACCGCACATGG
	Reverse	CCATGACACAGATAATCTGCAACCG
M6 A280X/S417Y/A480L	Forward	GTTTTGGCCTG**NNK**ACCGCAC

	Reverse	GATCATCACGCACTGCCTG
M6 S417Y/A480L/E549X	Forward	CGTTGGTTATAGTCTGGGCGAATTTG
	Reverse	CATT**MNN**TTCCGGCGGAACC
M6 S417Y/A480L/I552X	Forward	GGTTGTGGTTCTGAATAATCGTG
	Reverse	CCACCCAT**MNN**AACATTTTCTTCCGGCG
M6 S417Y/A480L/M553X	Forward	GGTTGTGGTTCTGAATAATCGTG
	Reverse	CCACC**MNN**AATAACATTTTCTTCCGGCG
M6 V105X/A280T/S417Y/A480L	Forward	GCTGGGC**NNK**GATGAAACCAATAC
	Reverse	CCATCCAGAATACGTTCAAC
M6 D106X/A280T/S417Y/A480L	Forward	GGCGTT**NNK**GAAACCAATACC
	Reverse	CCATCCAGAATACGTTCAAC
M6 E107X/A280T/S417Y/A480L	Forward	GGCGTTGAT**NNK**ACCAATACC
	Reverse	CCATCCAGAATACGTTCAAC
M6 T108X/A280T/S417Y/A480L	Forward	CTGGGCGTTGATGAA**NNK**AATACCC
	Reverse	CCATCCAGAATACGTTCAAC
M6 N109X/A280T/S417Y/A480L	Forward	GATGAAACC**NNK**ACCCTGCAG
	Reverse	CCATCCAGAATACGTTCAAC
M6 T110X/A280T/S417Y/A480L	Forward	CCAAT**NNK**CTGCAGGCAGGTATTG
	Reverse	CCATCCAGAATACGTTCAAC
M6 G278X/A280T/S417Y/A480L	Forward	GGTTTACGTTTT**NNK**CTGACTACCGCACATGGC
	Reverse	CCATGACACAGATAATCTGCAACCG
	Forward	CGTTTTGGC**NNK**ACTACCGCACATGG

M6 L279X/A280T/S417Y/A480L	Reverse	CCATGACACAGATAATCTGCAACCG
M6 A280T/S417Y/A480L/E549X	Forward	CGTTGGTTATAGTCTGGGCGAATTTG
	Reverse	CATT**MNN**TTCCGGCGGAACC
M6 A280T/S417Y/A480L/I552X	Forward	GGTTGTGGTTCTGAATAATCGTG
	Reverse	CCACCCAT**MNN**AACATTTTCTTCCGGCG
M6 A280T/S417Y/A480L/M553X	Forward	GGTTGTGGTTCTGAATAATCGTG
	Reverse	CCACC**MNN**AATAACATTTTCTTCCGGCG
M6 V105X/A280T/S417Y/A480L/I552M	Forward	GCTGGGC**NNK**GATGAAACCAATAC
	Reverse	CCATCCAGAATACGTTCAAC

a N = A, C, G, T (equimolar amounts); K = G, T (equimolar amounts); M = A, C (equimolar amounts).

Table S3. Apparent kinetic parameters of formolase mutants reported in literatures

Mutants	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}\cdot M^{-1}$)
K7[16]	0.119±0.008	53±5	2.25 ±0.26
FLS_M9[17]	1.16±0.05	20.5±1.89	57

Table S4. The target mutants of BALs derived from analogous *He*BAL M11 mutations

Mutants	Identity%	Hotspots					
*He*BAL M11	100	A27I	V29I	Y395A	S417Y	A476L	A480L
BAL-2 M6	37	G32I	H34I	S399A	G421Y	M480L	G484L
BAL-3 M5	63	A28I	-	Y396A	G418Y	A478L	A482L
BAL-5 M5	56	A28I	-	E394A	S416Y	A475L	T479L
BAL-10 M6	37	G28I	L30I	P394A	G416Y	M475L	G479L
BAL-15 M6	85	A27I	V29I	Y394A	G416Y	A475L	A479L
BAL-23 M5	57	A28I	-	E395A	S417Y	A476L	T480L

"-" means no mutation required

Table S5. The analysis of protein purifications

Enzyme	Purification step	Total protein (mg)	Activity (U/mg)	Activity yield %	Purification fold
*He*BAL M11	cell-free extract	7548	0.06	100	1
	Ni-NTA	384	1.16	98	19
CV2025	cell-free extract	8574	0.11	100	1
	Ni-NTA	710	0.99	75	9

Table S6. The enzyme cost per gram of DHA or serinol

Product	DHA	Serinol	
Enzyme	*He*BAL M11	*He*BAL M11	CV2025
Purified enzyme concentration (g/L)	4.50	4.50	3.00
Wet-cells weight (g/L)	597.66	597.66	126.76
Enzyme cost (US$/L)	229.49	229.49	25.07
Enzyme cost per gram of product (US$/g)	26.56	30.05	

Note:

A: The synthesis of DHA was catalyzed by *He*BAL M11, and the reaction contained 9 g/L HCHO, with 96% analytical yield.

B: The synthesis of serinol was catalyzed by *He*BAL M11 and CV2025, and the reaction contained 9 g/L HCHO and 29.6 g/L IPA, with 93% analytical yield.

2.2 Supplementary Figures

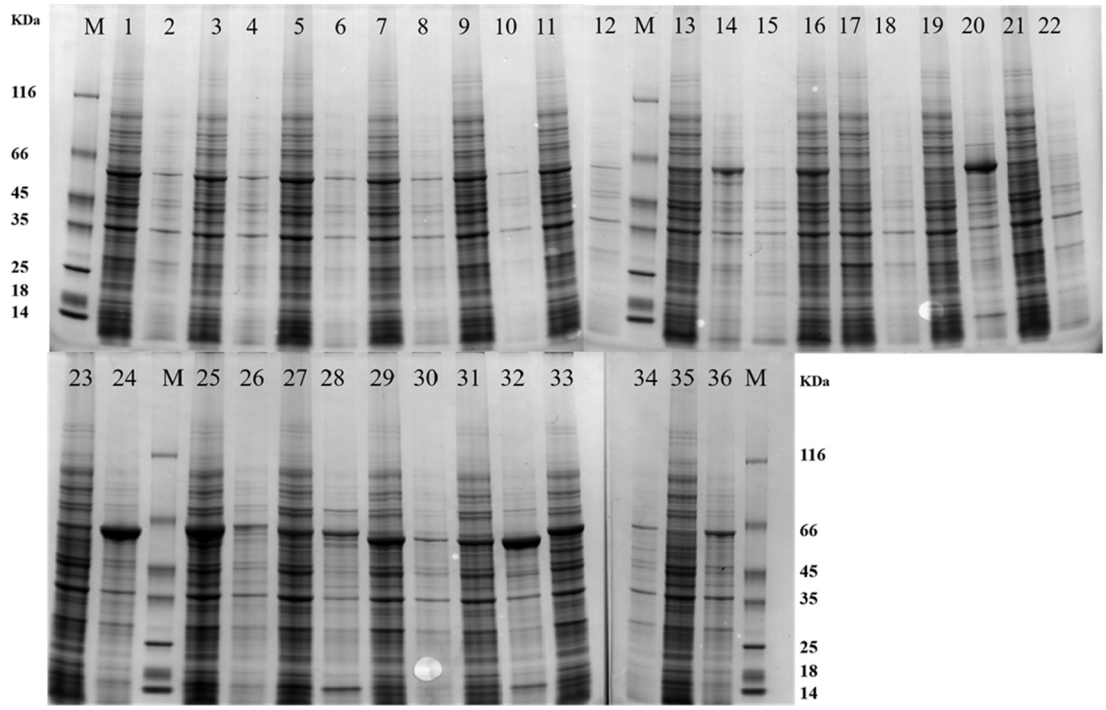


Figure S1. SDS-PAGE analysis of the mutants of BALs. Lane M: marker; Lane 1: the supernatant of *He*BAL M6; Lane 2: the precipitate of *He*BAL M6; Lane 3: the supernatant of *He*BAL M7; Lane 4: the precipitate of *He*BAL M7; Lane 5: the supernatant of *He*BAL M8; Lane 6: the precipitate of *He*BAL M8; Lane 7: the supernatant of *He*BAL M9; Lane 8: the precipitate of *He*BAL M9; Lane 9: the supernatant of *He*BAL M10; Lane 10: the precipitate of *He*BAL M10; Lane 11: the supernatant of *He*BAL M11; Lane 12: the precipitate of *He*BAL M11; Lane 13: the supernatant of BAL-2; Lane 14: the precipitate of BAL-2; Lane 15: the precipitate of BAL-2 M6; Lane 16: the supernatant of BAL-2 M6; Lane 17: the supernatant of BAL-3; Lane 18: the precipitate of BAL-3; Lane 19: the supernatant of BAL-3 M5; Lane 20: the precipitate of BAL-3 M5; Lane 21: the supernatant of BAL-5; Lane 22: the precipitate of BAL-5; Lane 23: the supernatant of BAL-5 M5; Lane 24: the precipitate of BAL-5 M5. Lane 25: the supernatant of BAL-10; Lane 26: the precipitate of BAL-10; Lane 27: the supernatant of BAL-10 M6; Lane 28: the precipitate of BAL-10 M6; Lane 29: the supernatant of BAL-15; Lane 30: the precipitate of BAL-15; Lane 31: the supernatant of BAL-15 M6; Lane 32: the precipitate of BAL-15 M6; Lane 33: the supernatant of BAL-23; Lane 34: the precipitate of BAL-23; 35: the supernatant of BAL-23 M5; Lane 36: the precipitate of BAL-23 M5;

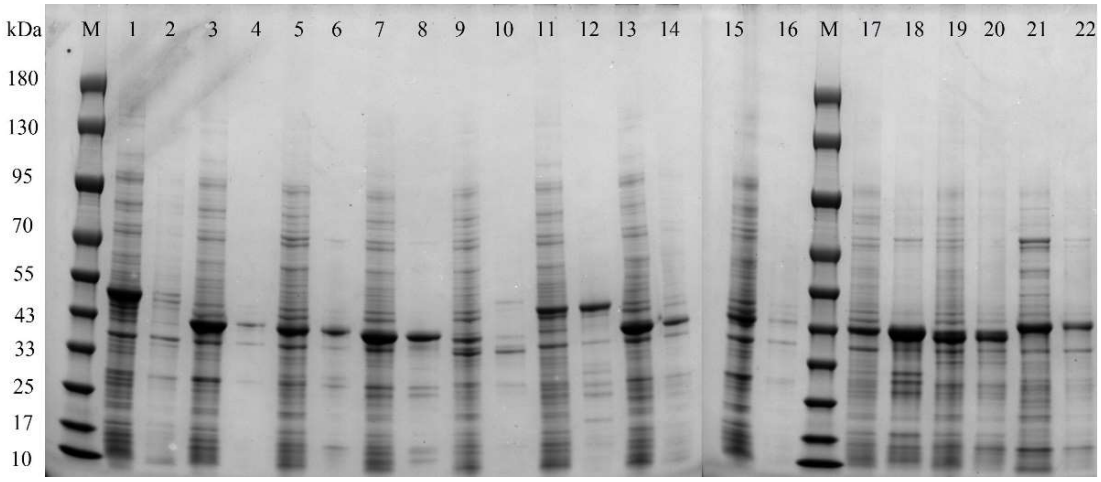


Figure S2. SDS-PAGE analysis of ω-TAs used in this study. Lane M: marker; Lane 1: the supernatant of CV2025; Lane 2: the precipitate of CV2025; Lane 3: the supernatant of *Ps*TA; Lane 4: the precipitate of *Ps*TA; Lane 5:the supernatant of *Atm*TA; Lane 6: the precipitate of *Atm*TA; Lane 7: the supernatant of *Ab*TA; Lane 8: the precipitate of *Ab*TA; Lane 9: the supernatant of *Mnf*TA; Lane 10: the precipitate of *Mnf*TA; Lane 11: the supernatant of *Sl*TA; Lane 12: the precipitate of *Sl*TA; Lane 13: the supernatant of ATA; Lane 14: the precipitate of ATA; Lane 15: the supernatant of *Amb*TA; Lane 16: the precipitate of *Amb*TA; Lane 17: the supernatant of *Mi*TA; Lane 18: the precipitate of *Mi*TA; Lane 19: the supernatant of *Mba*TA; Lane 20: the precipitate of *Mba*TA; Lane 21: the supernatant of *Pa*TA; Lane 22: the precipitate of *Pa*TA.

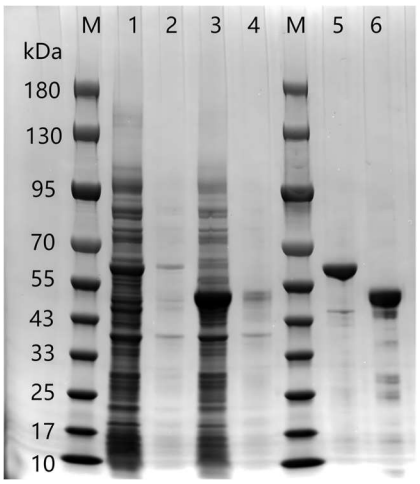


Figure S3. SDS-PAGE analysis of *He*BAL M11 and CV2025. Lane M: marker; Lane 1: the supernatant of *He*BAL M11; Lane 2: the precipitate of *He*BAL M11; Lane 3: the supernatant of CV2025; Lane 4: the precipitate of CV2025; Lane 5: the purified enzyme of *He*BAL M11; Lane 6: the purified enzyme of CV2025.

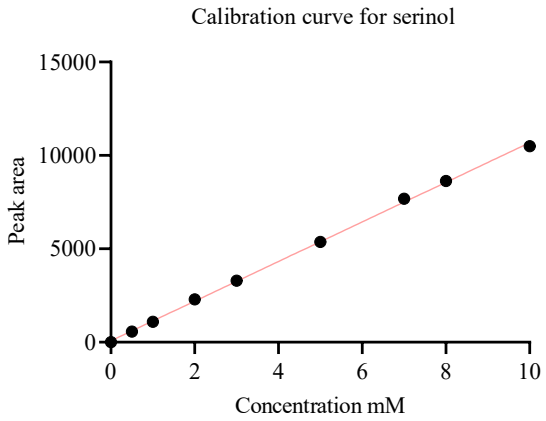


Figure S4. Calibration curve for serinol after derivatization with Cbz-Cl. The R^2 of calibration curve was 0.9985.

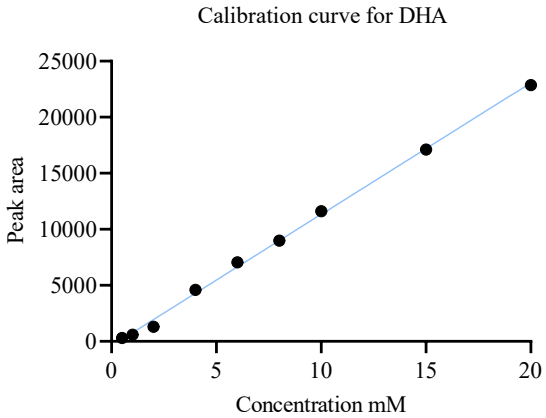


Figure S5. Calibration curve for DHA after derivatization with *O*-benzylhydroxylamine hydrochloride. The R^2 of calibration curve was 0.9982.

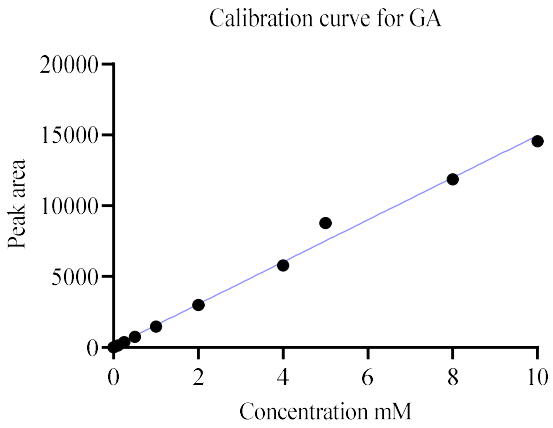


Figure S6. Calibration curve for GA after derivatization with *O*-benzylhydroxylamine hydrochloride. The R^2 of calibration curve was 0.9960.

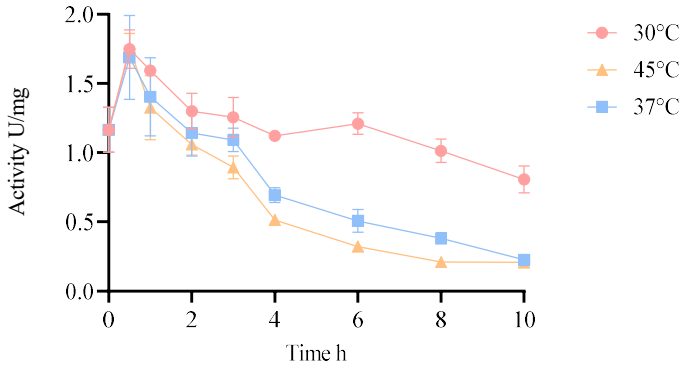


Figure S7. The half-life of the mutant *He*BAL M11 at different temperatures. Reaction conditions: Potassium phosphate buffer (100 mM, pH 7.0) containing 300 mM HCHO, 0.5 mg/mL purified enzyme of *He*BAL M11, 2.5 mM $MgSO_4$, 0.1 mM ThDP and 10% DMSO. The total volume was 1 mL and the reaction was stirred at 30°C for 20 min. The yield of reaction was determined by HPLC after derivation with *O*-benzylhydroxylamine hydrochloride.

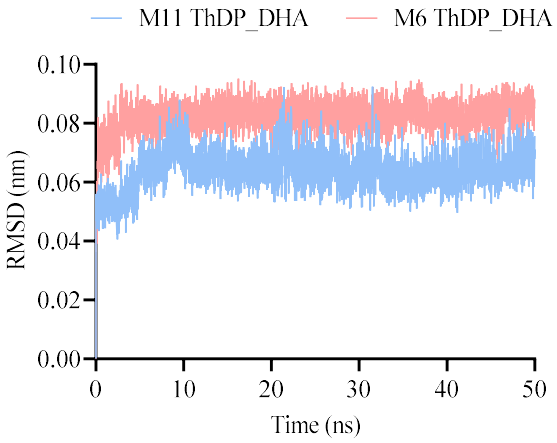


Figure S8. The RMSD values of ThDP_DHA in *He*BAL M6 and *He*BAL M11 mutants during 50 ns MD simulations.

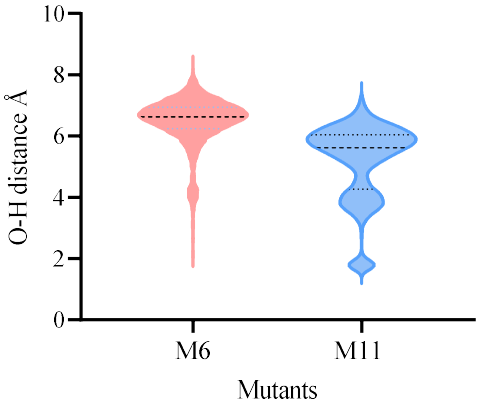
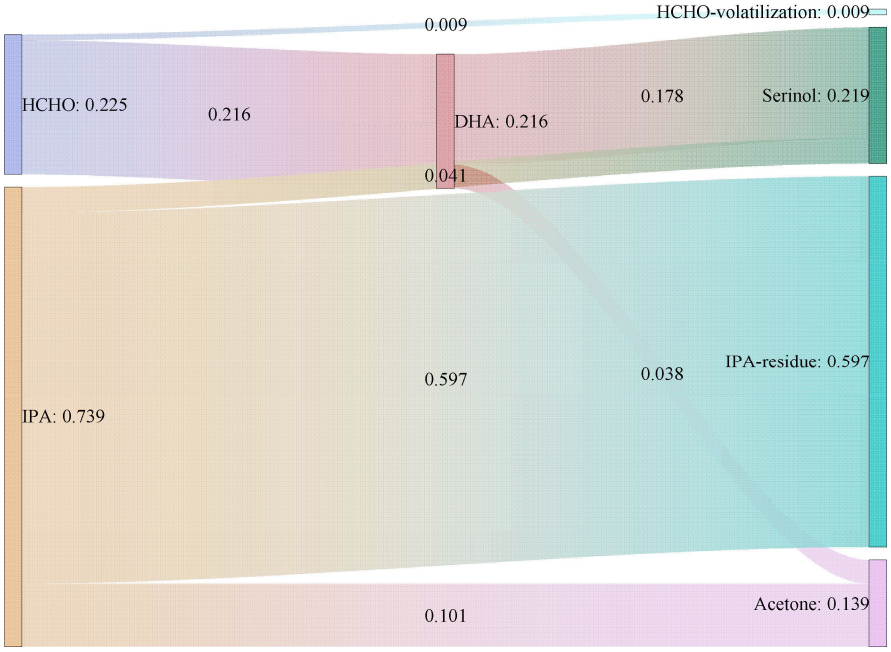


Figure S9. The O-H distances in the mutants *He*BAL M6 and *He*BAL M11 during MD simulations.



Unit: g

Figure S10. Mass flow analysis of the one-pot, two-step cascade reaction for synthesizing serinol.

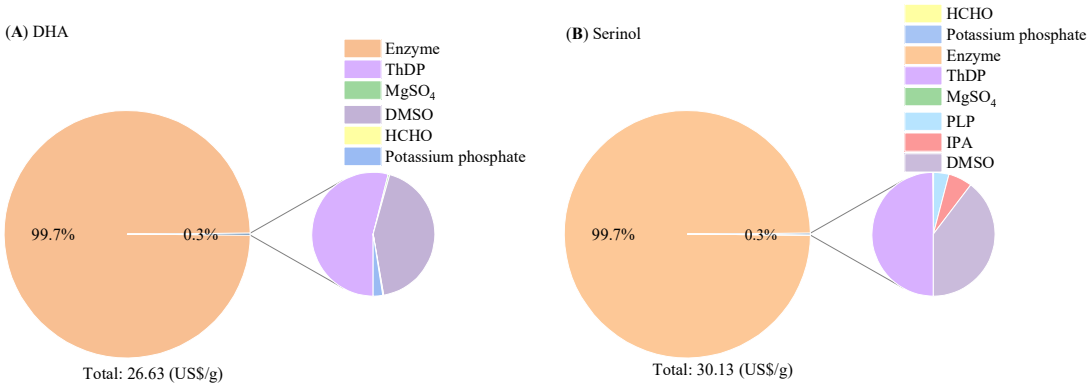


Figure S11. The total cost per gram of DHA or serinol.

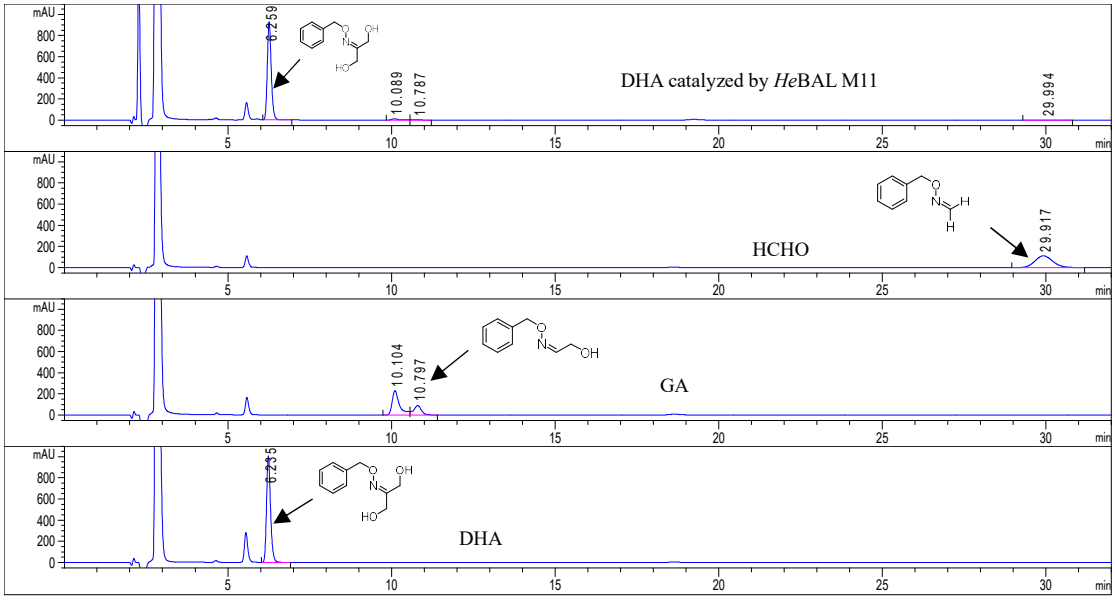


Figure S12. HPLC chromatogram of the DHA catalyzed by *He*BAL M11, the standard of HCHO, GA and DHA after derivatization with *O*-benzylhydroxylamine hydrochloride.

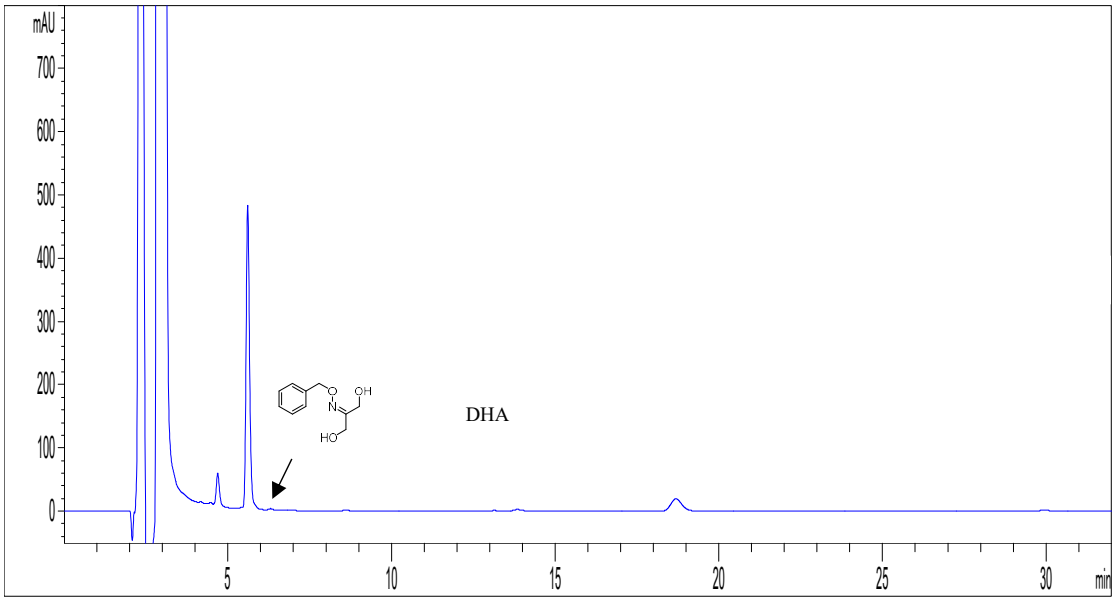


Figure S13. HPLC chromatogram of the DHA residue in the transamination reaction after derivatization with *O*-benzylhydroxylamine hydrochloride.

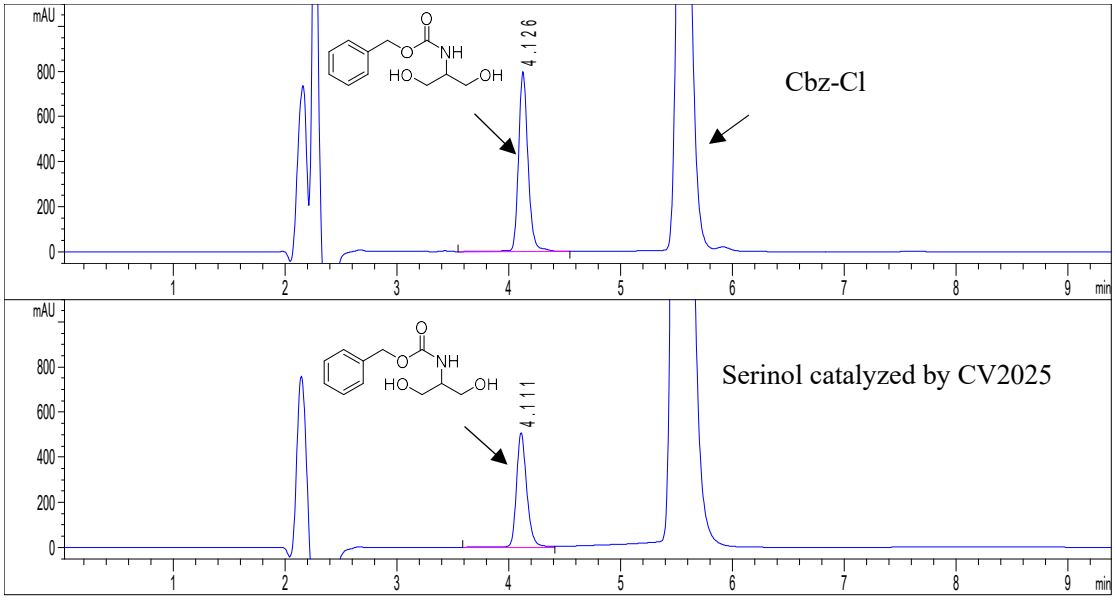


Figure S14. HPLC chromatogram of the standard of serinol and serinol catalyzed by CV2025 after derivatization with Cbz-Cl.

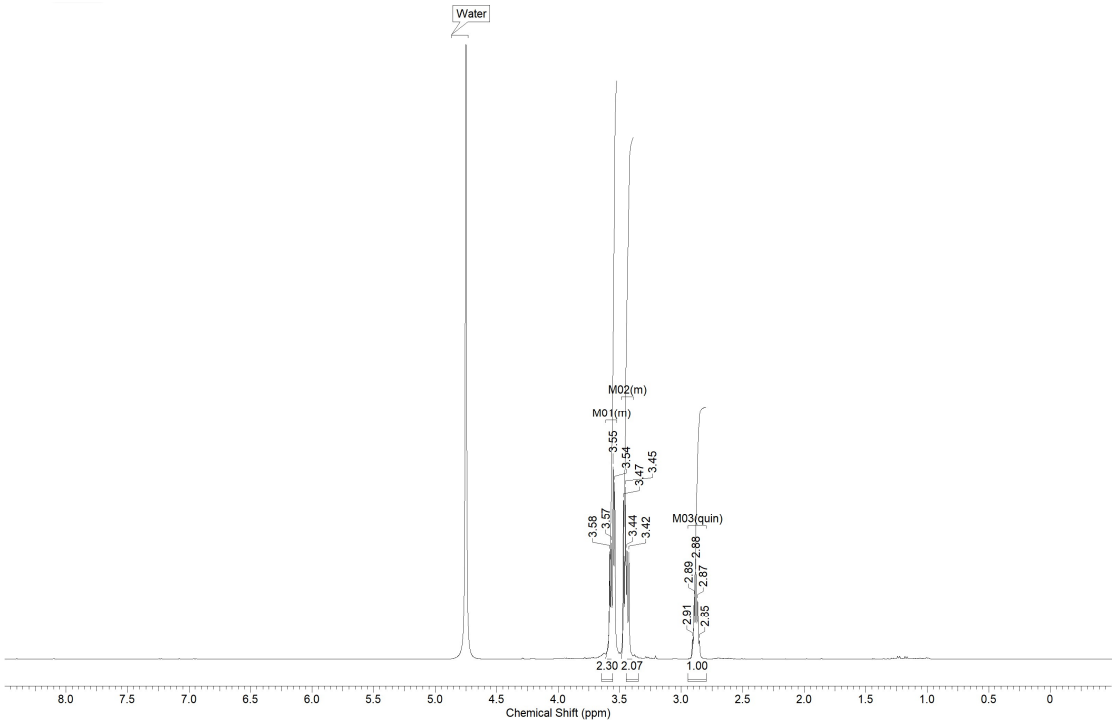


Figure S15. ^{1}H-NMR spectra of serinol.

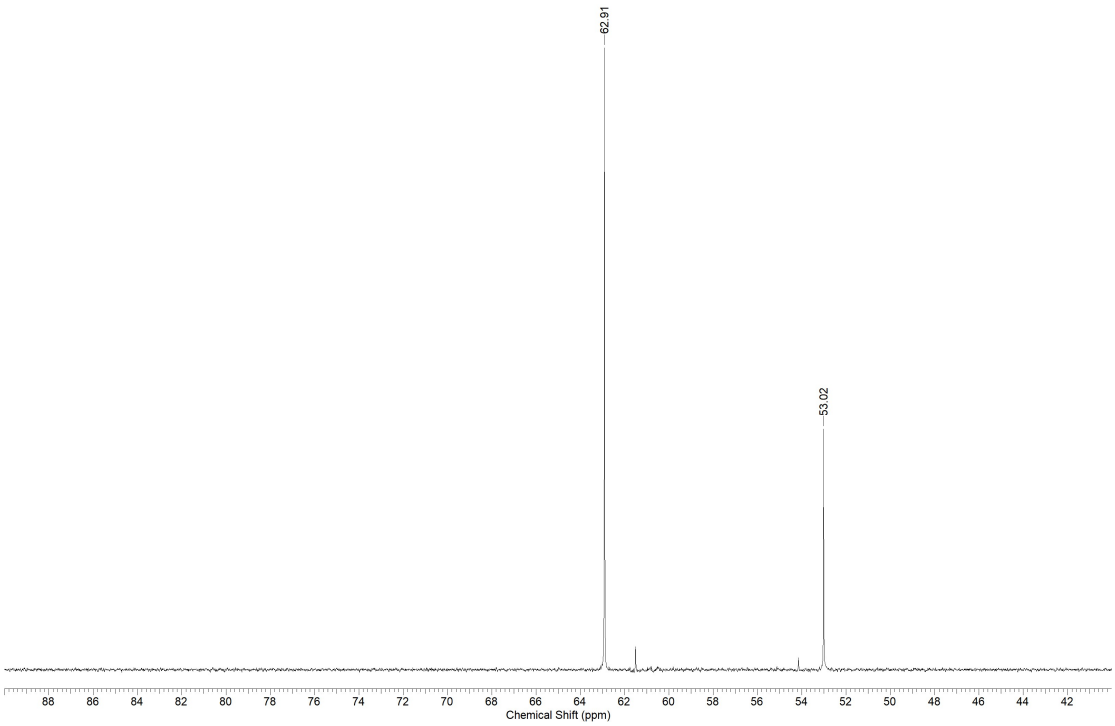


Figure S16. ^{13}C-NMR spectra of serinol.

3 Sequence Information

>*He*BAL M6

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAG
CGTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCA
GGCAGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGA
ATGCAGGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGC
GTGGCCCTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATT
GCAAATGCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGT
CCGCTGGGCGTTGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGT
GGCAATTGCCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGG
AACTGCTGCCGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCC
CGCGTGGTCCGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCG
TTGATGATGCACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTG
ACCGCAGCCCTGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGC
AGGTGCAGAACGCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTG
ACGGCGGTGCCGCATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTG

TTTAGTGATACCGAAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTT
TGGCCTGCTGCAGGGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATC
GCGTGGTGCTGTTTGGTTTACGTTTTGGCCTGGCCACCGCACATGGCAGTG
GTATTCTGATTCCGCGTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCC
GCGAACTGGGTCGTCTGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCG
GCAGCCGCTGCTGAAGAACTGGCACGTGCAGCACTGACCTGGGCAGCAG
GCTGGCCGGATCGTAGCCGCTGGCAGGAACGTCTGCGCGAACTGGTGGAT
GGTCGCTTTGAAAGTGTTACCGCACAGGCAGTGCGTGATGATCGTATTCA
TCCGATGGATGCAGTGACCGCAATTGCCGAAACCGTTCCGGCAGGCAGTG
TTGTTGTGGCCGATGGTGCACTGACCGCGCTGTGGCTGAGCGAAACCATT
AGCCGCGCACCGGTTGCAGATTATCTGTGTCATGGTTATCTGAGTAGTATG
GGCGTTGGCGTTGGTACCGCACTGGGTGCCCAGGCAGCAGATGTGACCCG
CCCGGTTGTTCTGGTGACCGGTGACGGTGCCGTTGGTTATAGTCTGGGCGA
ATTTGATAGCATGGTTCGTGCAGGTCTGCCGGTTGTGGTTGTGGTTCTGAA
TAATCGTGCATGGGGCTTGACCCTGCATGCCCAGGAACTGATTCTGGGTC
CGGATCGCGTTGTTAATAATCGCCTGGAAAATGGCAGTTATAGTGGTGTT
GCCCGCGCCCTGGGCGCAGATAGTATTGATGTTAGCGATATTGCAGATCT
GGCCCCGACCCTGCGCGAAGCACTGGCAAGCGGCCGTCCGACCTGCATTG
AAGTTCATGTGAGTCTGGCACCGGTTCCGCCGGAAGAAAATGTTATTATG
GGTGGTAAACCGTTT

>*He*BAL M7

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAG
CGTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCA
GGCAGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGA
ATGCAGGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGC
GTGGCCCTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATT
GCAAATGCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGT
CCGCTGGGCGTTGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGT
GGCAATTGCCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGG
AACTGCTGCCGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCC
CGCGTGGTCCGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCG

TTGATGATGCACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTG
ACCGCAGCCCTGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGC
AGGTGCAGAACGCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTG
ACGGCGGTGCCGCATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTG
TTTAGTGATACCGAAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTT
TGGCCTGCTGCAGGGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATC
GCGTGGTGCTGTTTGGTTTACGTTTTGGCCTGGCCACCGCACATGGCAGTG
GTATTCTGATTCCGCGTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCC
GCGAACTGGGTCGTCTGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCG
GCAGCCGCTGCTGAAGAACTGGCACGTGCAGCACTGACCTGGGCAGCAG
GCTGGCCGGATCGTAGCCGCTGGCAGGAACGTCTGCGCGAACTGGTGGAT
GGTCGCTTTGAAAGTGTTACCGCACAGGCAGTGCGTGATGATCGTATTCA
TCCGATGGATGCAGTGACCGCAATTGCCGAAACCGTTCCGGCAGGCAGTG
TTGTTGTGGCCGATGGTGCACTGACCGCGCTGTGGCTGAGCGAAACCATT
AGCCGCGCACCGGTTGCAGATTATCTGTGTCATGGTTATCTGTATAGTATG
GGCGTTGGCGTTGGTACCGCACTGGGTGCCCAGGCAGCAGATGTGACCCG
CCCGGTTGTTCTGGTGACCGGTGACGGTGCCGTTGGTTATAGTCTGGGCGA
ATTTGATAGCATGGTTCGTGCAGGTCTGCCGGTTGTGGTTGTGGTTCTGAA
TAATCGTGCATGGGGCTTGACCCTGCATGCCCAGGAACTGATTCTGGGTC
CGGATCGCGTTGTTAATAATCGCCTGGAAAATGGCAGTTATAGTGGTGTT
GCCCGCGCCCTGGGCGCAGATAGTATTGATGTTAGCGATATTGCAGATCT
GGCCCCGACCCTGCGCGAAGCACTGGCAAGCGGCCGTCCGACCTGCATTG
AAGTTCATGTGAGTCTGGCACCGGTTCCGCCGGAAGAAAATGTTATTATG
GGTGGTAAACCGTTT

>*He*BAL M8

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAG
CGTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCA
GGCAGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGA
ATGCAGGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGC
GTGGCCCTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATT
GCAAATGCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGT

CCGCTGGGCGTTGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGT
GGCAATTGCCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGG
AACTGCTGCCGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCC
CGCGTGGTCCGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCG
TTGATGATGCACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTG
ACCGCAGCCCTGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGC
AGGTGCAGAACGCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTG
ACGGCGGTGCCGCATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTG
TTTAGTGATACCGAAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTT
TGGCCTGCTGCAGGGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATC
GCGTGGTGCTGTTTGGTTTACGTTTTGGCCTGGCCACCGCACATGGCAGTG
GTATTCTGATTCCGCGTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCC
GCGAACTGGGTCGTCTGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCG
GCAGCCGCTGCTGAAGAACTGGCACGTGCAGCACTGACCTGGGCAGCAG
GCTGGCCGGATCGTAGCCGCTGGCAGGAACGTCTGCGCGAACTGGTGGAT
GGTCGCTTTGAAAGTGTTACCGCACAGGCAGTGCGTGATGATCGTATTCA
TCCGATGGATGCAGTGACCGCAATTGCCGAAACCGTTCCGGCAGGCAGTG
TTGTTGTGGCCGATGGTGCACTGACCGCGCTGTGGCTGAGCGAAACCATT
AGCCGCGCACCGGTTGCAGATTATCTGTGTCATGGTTATCTGTATAGTATG
GGCGTTGGCGTTGGTACCGCACTGGGTGCCCAGGCAGCAGATGTGACCCG
CCCGGTTGTTCTGGTGACCGGTGACGGTGCCGTTGGTTATAGTCTGGGCGA
ATTTGATAGCATGGTTCGTGCAGGTCTGCCGGTTGTGGTTGTGGTTCTGAA
TAATCGTGCATGGGGCTTGACCCTGCATCTTCAGGAACTGATTCTGGGTCC
GGATCGCGTTGTTAATAATCGCCTGGAAAATGGCAGTTATAGTGGTGTTG
CCCGCGCCCTGGGCGCAGATAGTATTGATGTTAGCGATATTGCAGATCTG
GCCCCGACCCTGCGCGAAGCACTGGCAAGCGGCCGTCCGACCTGCATTGA
AGTTCATGTGAGTCTGGCACCGGTTCCGCCGGAAGAAAATGTTATTATGG
GTGGTAAACCGTTT

>HeBAL M9

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAG
CGTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCA

GGCAGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGA
ATGCAGGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGC
GTGGCCCTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATT
GCAAATGCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGT
CCGCTGGGCGTTGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGT
GGCAATTGCCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGG
AACTGCTGCCGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCC
CGCGTGGTCCGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCG
TTGATGATGCACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTG
ACCGCAGCCCTGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGC
AGGTGCAGAACGCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTG
ACGGCGGTGCCGCATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTG
TTTAGTGATACCGAAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTT
TGGCCTGCTGCAGGGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATC
GCGTGGTGCTGTTTGGTTTACGTTTTGGCCTGACTACCGCACATGGCAGTG
GTATTCTGATTCCGCGTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCC
GCGAACTGGGTCGTCTGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCG
GCAGCCGCTGCTGAAGAACTGGCACGTGCAGCACTGACCTGGGCAGCAG
GCTGGCCGGATCGTAGCCGCTGGCAGGAACGTCTGCGCGAACTGGTGGAT
GGTCGCTTTGAAAGTGTTACCGCACAGGCAGTGCGTGATGATCGTATTCA
TCCGATGGATGCAGTGACCGCAATTGCCGAAACCGTTCCGGCAGGCAGTG
TTGTTGTGGCCGATGGTGCACTGACCGCGCTGTGGCTGAGCGAAACCATT
AGCCGCGCACCGGTTGCAGATTATCTGTGTCATGGTTATCTGTATAGTATG
GGCGTTGGCGTTGGTACCGCACTGGGTGCCCAGGCAGCAGATGTGACCCG
CCCGGTTGTTCTGGTGACCGGTGACGGTGCCGTTGGTTATAGTCTGGGCGA
ATTTGATAGCATGGTTCGTGCAGGTCTGCCGGTTGTGGTTGTGGTTCTGAA
TAATCGTGCATGGGGCTTGACCCTGCATCTTCAGGAACTGATTCTGGGTCC
GGATCGCGTTGTTAATAATCGCCTGGAAAATGGCAGTTATAGTGGTGTTG
CCCGCGCCCTGGGCGCAGATAGTATTGATGTTAGCGATATTGCAGATCTG
GCCCCGACCCTGCGCGAAGCACTGGCAAGCGGCCGTCCGACCTGCATTGA
AGTTCATGTGAGTCTGGCACCGGTTCCGCCGGAAGAAAATGTTATTATGG
GTGGTAAACCGTTT

>*He*BAL M10

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAG
CGTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCA
GGCAGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGA
ATGCAGGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGC
GTGGCCCTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATT
GCAAATGCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGT
CCGCTGGGCGTTGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGT
GGCAATTGCCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGG
AACTGCTGCCGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCC
CGCGTGGTCCGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCG
TTGATGATGCACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTG
ACCGCAGCCCTGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGC
AGGTGCAGAACGCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTG
ACGGCGGTGCCGCATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTG
TTTAGTGATACCGAAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTT
TGGCCTGCTGCAGGGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATC
GCGTGGTGCTGTTTGGTTTACGTTTTGGCCTGACTACCGCACATGGCAGTG
GTATTCTGATTCCGCGTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCC
GCGAACTGGGTCGTCTGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCG
GCAGCCGCTGCTGAAGAACTGGCACGTGCAGCACTGACCTGGGCAGCAG
GCTGGCCGGATCGTAGCCGCTGGCAGGAACGTCTGCGCGAACTGGTGGAT
GGTCGCTTTGAAAGTGTTACCGCACAGGCAGTGCGTGATGATCGTATTCA
TCCGATGGATGCAGTGACCGCAATTGCCGAAACCGTTCCGGCAGGCAGTG
TTGTTGTGGCCGATGGTGCACTGACCGCGCTGTGGCTGAGCGAAACCATT
AGCCGCGCACCGGTTGCAGATTATCTGTGTCATGGTTATCTGTATAGTATG
GGCGTTGGCGTTGGTACCGCACTGGGTGCCCAGGCAGCAGATGTGACCCG
CCCGGTTGTTCTGGTGACCGGTGACGGTGCCGTTGGTTATAGTCTGGGCGA
ATTTGATAGCATGGTTCGTGCAGGTCTGCCGGTTGTGGTTGTGGTTCTGAA
TAATCGTGCATGGGGCTTGACCCTGCATCTTCAGGAACTGATTCTGGGTCC
GGATCGCGTTGTTAATAATCGCCTGGAAAATGGCAGTTATAGTGGTGTTG
CCCGCGCCCTGGGCGCAGATAGTATTGATGTTAGCGATATTGCAGATCTG
GCCCCGACCCTGCGCGAAGCACTGGCAAGCGGCCGTCCGACCTGCATTGA

AGTTCATGTGAGTCTGGCACCGGTTCCGCCGGAAGAAAATGTTATGATGG
GTGGTAAACCGTTT

>*He*BAL M11

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAG
CGTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCA
GGCAGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGA
ATGCAGGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGC
GTGGCCCTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATT
GCAAATGCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGT
CCGCTGGGCTCGGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGT
GGCAATTGCCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGG
AACTGCTGCCGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCC
CGCGTGGTCCGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCG
TTGATGATGCACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTG
ACCGCAGCCCTGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGC
AGGTGCAGAACGCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTG
ACGGCGGTGCCGCATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTG
TTTAGTGATACCGAAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTT
TGGCCTGCTGCAGGGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATC
GCGTGGTGCTGTTTGGTTTACGTTTTGGCCTGACTACCGCACATGGCAGTG
GTATTCTGATTCCGCGTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCC
GCGAACTGGGTCGTCTGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCG
GCAGCCGCTGCTGAAGAACTGGCACGTGCAGCACTGACCTGGGCAGCAG
GCTGGCCGGATCGTAGCCGCTGGCAGGAACGTCTGCGCGAACTGGTGGAT
GGTCGCTTTGAAAGTGTTACCGCACAGGCAGTGCGTGATGATCGTATTCA
TCCGATGGATGCAGTGACCGCAATTGCCGAAACCGTTCCGGCAGGCAGTG
TTGTTGTGGCCGATGGTGCACTGACCGCGCTGTGGCTGAGCGAAACCATT
AGCCGCGCACCGGTTGCAGATTATCTGTGTCATGGTTATCTGTATAGTATG
GGCGTTGGCGTTGGTACCGCACTGGGTGCCCAGGCAGCAGATGTGACCCG
CCCGGTTGTTCTGGTGACCGGTGACGGTGCCGTTGGTTATAGTCTGGGCGA
ATTTGATAGCATGGTTCGTGCAGGTCTGCCGGTTGTGGTTGTGGTTCTGAA

TAATCGTGCATGGGGCTTGACCCTGCATCTTCAGGAACTGATTCTGGGTCC
GGATCGCGTTGTTAATAATCGCCTGGAAAATGGCAGTTATAGTGGTGTTG
CCCGCGCCCTGGGCGCAGATAGTATTGATGTTAGCGATATTGCAGATCTG
GCCCCGACCCTGCGCGAAGCACTGGCAAGCGGCCGTCCGACCTGCATTGA
AGTTCATGTGAGTCTGGCACCGGTTCCGCCGGAAGAAAATGTTATGATGG
GTGGTAAACCGTTT

>*He*BAL M12

Gene sequence:

ATGACCGCCACCGGTGGTGAACTGGTTATTCGTACCCTGGAACGCGCAAGC
GTTGATGTTGCATTTGGCATTAATGGTATTCATATTGATAGTATCTATCAGGC
AGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCAATGAAATGAATGCA
GGCCATGCAGCCGAAGGTTATGCACGTGCAGGTCATCGCCTGGGCGTGGCC
CTGCTGACCGCAGGTGGTGGTTTTACCAATGCCGTTACCAGTATTGCAAAT
GCACATCTGGATCGCACCCCGGTTCTGTATATTGCATCGAGTGGTCCGCTGG
GCACTGATGAAACCAATACCCTGCAGGCAGGTATTGATCAGGTGGCAATTG
CCACCCCGATTACCAAATGGGCACATCGCGTGACCCGCGTGGAACTGCTGC
CGCGTCTGATTGCCCAGGCAATTCGTATTGCAACCCATGGCCCGCGTGGTC
CGGTGCTGCTGGATATTCCGTGGGATGTTCTGACCGCCACCGTTGATGATGC
ACTGGCAGATGGTGTGGAAGAACTGGGTGCACATGCCCTGACCGCAGCCC
TGGGCGCTGATGCAGTTGAACGTATTCTGGATGGTCTGGCAGGTGCAGAAC
GCCCGGTGTTTATTGCCGGTAGTGAACTGACCCGCGGTGACGGCGGTGCCG
CATTACGTCGCCTGGCAGAAATTACCGGCACCCCGCTGTTTAGTGATACCG
AAGCCCTGGGCGCCATTCGCGAAAGCCCGCTGAGTTTTGGCCTGCTGCAG
GGTCTGTTTGGTCTGGATGAAGCCGAACGTCCGGATCGCGTGGTGCTGTTT
GGTTTACGTTTTGGCCTGACTACCGCACATGGCAGTGGTATTCTGATTCCGC
GTGATGCAGCAGTGGTTCAGATTGATAGTGATGCCCGCGAACTGGGTCGTC
TGCAGCCGATTGAACTGGGTGCCGTTGGCGATCCGGCAGCCGCTGCTGAA
GAACTGGCACGTGCAGCACTGACCTGGGCAGCAGGCTGGCCGGATCGTAG
CCGCTGGCAGGAACGTCTGCGCGAACTGGTGGATGGTCGCTTTGAAAGTG
TTACCGCACAGGCAGTGCGTGATGATCGTATTCATCCGATGGATGCAGTGA
CCGCAATTGCCGAAACCGTTCCGGCAGGCAGTGTTGTTGTGGCCGATGGTG
CACTGACCGCGCTGTGGCTGAGCGAAACCATTAGCCGCGCACCGGTTGCA

GATTATCTGTGTCATGGTTATCTGTATAGTATGGGCGTTGGCGTTGGTACCGC
ACTGGGTGCCCAGGCAGCAGATGTGACCCGCCCGGTTGTTCTGGTGACCG
GTGACGGTGCCGTTGGTTATAGTCTGGGCGAATTTGATAGCATGGTTCGTGC
AGGTCTGCCGGTTGTGGTTGTGGTTCTGAATAATCGTGCATGGGGCTTGAC
CCTGCATCTTCAGGAACTGATTCTGGGTCCGGATCGCGTTGTTAATAATCGC
CTGGAAAATGGCAGTTATAGTGGTGTTGCCCGCGCCCTGGGCGCAGATAGT
ATTGATGTTAGCGATATTGCAGATCTGGCCCCGACCCTGCGCGAAGCACTG
GCAAGCGGCCGTCCGACCTGCATTGAAGTTCATGTGAGTCTGGCACCGGTT
CCGCCGGAAGAAAATGTTATGATGGGTGGTAAACCGTTT

>BAL-2

Gene sequence:

CATATGAGCAGTCCGGAAGCCCGTTATACCGGTGGCGATCTGCTGGCACAG
ACCCTGCATGATGCAGGTGTGACCAAAATTTTTGCACTGCATGGCGGCCAT
CATGAAGCACTGTTTAAAGGCTGCATTGATCAGGGCATTGATCTGATTGATT
TTCGTCATGAAGCCGCCGCCGGCCATGCAGCCGATGCATACGCTCGCACCA
CCGGCAAACTGGGTGTGTGTATTATTACCGCCGGCCCGGGTTTTACCAATGC
CATTAGCGCAATTGCCAATGCACAGCTGGATGCAAGCCCGGTGCTGTTTCT
GATTGGCGCACCGCCGCTGCGTGAAGTGGAAACCAATCCGCTGCAGGGTG
GTATTGATCAGATTGCCATGGCCCGCCCGGCAGCCAAATGGGCTCTGAGCA
TTCCGAGCACCGAACGCGTTCGTGATCTGACCGCAATGGCAATTCGCAAAG
CCATGACCGGTCGCAAAGGCCCGGTGGTTCTGGAAATTCCGATTGATATTC
TGCACATGAGCGTTACCGGTGCACAGGCCACCCCGAGTGCCGGTCTGGCA
GTTCGCCCGCAGCCGGCACCTGCTCCTGAAGAAGTTGCAGCACTGGCCGA
ACTGCTGCTGCGTGCAGAACGTCCGGTTATTGTGGCAGGCCTGGAAAGCG
CCAGCGCCGCTACCGCAGTTGCCCTGCGTGCACTGGTTGCCAAACTGCCGC
TGCCGGTTTTTGCCAAACCGCAGGCATACGGTCTGCTGCCGGCCGGTCATG
CCTGTGATGCAGGTGCAGCCGGTAATCTGGCAGTTCTGCCGATTATTGGCG
CCGGTGCACCGGATCTGGTTATTCTGCTGGGTGCCCGCCTGGGCCTGATGC
TGGGTGGTCGTAGTGGTGCCCTGGTTCCGCATGATGCCCATGTTGTGCAGA
TCTATAGCGATGCCAGCGAAATTGGTCGTCTGCGTGATATTGATCTGCCGAT
TGCCGCCGATTGCGCACAGACCTTAACCGCACTGACCAAAGCACTGGCCG
CAGTTGATCTGCCGGATACCAGCGCATGGACCGCCCGTGCCGCCGGTGCAA

AAGCACTGGCAGCAAGTGCATGGCCGGATGCAGAAGTGGCCGGTGGTATT
CATCCGTATCATGCCGCAAAAGCAGTTGCAAATGCCGCAGGTCAGGATGCC
GCCTATGTGTTTGATGGCGGCGAAAGTAGCAGCTGGGGCACCGCAACCGTT
GCCGTTGATGCACCGGCCCGCGTTCTGAGTCATGGCTATCTGGGTTGTCTG
GGCATTGGTCCGGGTTTTGCAATTGGTATGCAGATTGCCCATCCGGATCGTC
GCGTTGTGCAGGTTACCGGTGACGGCGCAATGGGCTTTCATATTCAGGAAT
TTGATACCATGGTTCGCCATCGCCTGCCGATTGTGACCGTGATTCTGAATAA
TCAGGTTTGGGGCATGAGTATTCATGGTCAGCAGATGATGTATGGCGCCAAT
TATAATGTGATTACCAAACTGGGCAGTACCCAGTATGCAAGCATTGCCGCCG
CATTTGGCTGCCATGCAGAACGCGTGACCGCATTTGCAGAAATTGCCCCGG
CAATGGCCCGCGCATTTGCAAGTGGTAAACCGGCACTGGTGGAAATTATGA
CCGATGCCGATGTTGTTCATCCGGCCACCGTTGCAATGCTGGGTCAGCTGG
CAGAAGGTAGCCGCGATATTATGATTCCGTATTATGAAAATATCGCGGCCAG
CCTCGAG

>BAL-3

Gene sequence:

CATATGCCGCTGGTTAGTGGCGGCGAACTGGTTGTTCGCCTGCTGCAGCA
GGCCGGCGTGGATCGTATGTTTGGCATTAATGGCGCCCATATTGATGCAG
TTTATCAGGCCGCCCTGGATCATCGTCTGCCGATTGTTGATACCCGTCATG
AAATGAATGCAGGTCATGCCGCCGAAGGCTATGCCCGTGTGCGTAATGCA
CTGGGTGTGGCACTGCTGACCGCCGGTGGTGGCTTTACCAATGCCGTGAC
CAGTATGGCAAATGCCCATCTGGATCGCACCCCGGTTCTGTATCTGGCCGC
AAGTGGTCCGCTGGCAGATGATCAGACCAATACCCTGCAGGCAGGCCTGG
ATCAGGTGGCCATTGCAACCCCGGTGACCAAATGGGCACATCGTGTGACC
CGTGCAAGCCTGATTCCGCGTATTCTGGCCCGCGCCATTCGTACCGCCCTG
AGCGCACCGCGCGGTCCTGTTCTGGTTGATATTCCGTGGGATGTTCTGACC
GAACAGGTGGATCTGCCGGATCTGGCCACCGAACTGCTGCATGGCGCCGA
TCATGGTCCGGGCCTGAGCCAGGCAGGCGCAGATCGTATTCTGGCACTGC
TGAGCGAAGCCCGCCGTCCGGTTGTGGCCGTGGGTTCAGAAGTGCCGCGC
AGCGGCGCAGGTGGCGAAGTTCGTGCATTTGCAGAACGTCTGGGCGTGCC
GCTGCTGAGCGATTATGAAGGTCTGGGCGCCATTAGCCCGAGTCCGATGA
ATTTTGGTCTGGTGCAGAGTCTGCATGGCCTGCCGGCCGGCGCAGAACCT

GATCTGGTTATTATGGCAGGTCTGCGTTTTGGTCTGACCACCGCCCATGGC
AGTGGTGCACTGATTCCGCGCACCGCACGTGTTCTGCAGATTGATCCGAG
TGGTGCAGAACTGGGCCATCTGCAGCCGGTTGCCCTGAGCGTTGCCGCAG
ATCCGCTGCCGGTTTTTCAGGATCTGAATCGTGTGCTGGCCGTGCGTCAGA
GCGCAGCACGTCCGGATCTGGCAGATTGGCGTGATACCCTGCATGCACAT
CTGCGCCGCCGTCGCGAAGCAGTTGCAGCACAGGTTGTTCAGGATGAACG
TGTTCATCCGTATGATGCAGTGTGCGTTATTGCAGAAGCACTGCCGGATGC
CAGTGTGGTGGTGGCAGATGGTGCCCTGACCTATCTGTGGCTGAGTGAAA
CCATTAGCAGCGCCCCGGTGAGTGCCTTTCTGTGCCATGGTTATCTGGGTA
GTATGGGTGTGGGCATGGGCATTGCCCTGGGTGCACAGGCAGCAGTTGAA
CCGGGTACCAATGTTGTGCTGGTTACCGGCGATGGTAGTGTGGGTTATAGT
CTGGGTGAATTTGATAGTATGGTTCGCGCAGGTCTGGGTGCCGTTGTTGTG
GTGCTGAATAATCGCGCATGGGGCGCCACCCTGCATGCGCAAGAAATTAT
GCATGGCAAAGATCGTGTGGTTAATAATCGCCTGGAAAATGGTAGCTATG
CCGCAGCAGCAATTGCACTGGGTGCAGATGGTTATCAGGTGCGCACCCTG
GATGAACTGGCCCCGGCCCTGGCAAAAGCACTGCGCAGCGGCCGCCCGA
GTTGCGTTGAAGTGCATGTGAGTCTGGACCCTATTCCGCCGGAAGAACGT
GTTATTATGGGCGGTAGTCCGTTTCTCGAG

>BAL-5

Gene sequence:

CATATGGCCACCGTTACCGGTAGTGGCCTGGTTGTTCGTACCCTGCAGCGC
GCCGGTGTGCATGTGGCCTTTGGCCTGCCGGGCGCCCATATTGATCCGATT
CTGCAGGAAGCACTGGATGCCAAACTGCGCGTTGTTGATGTTCGTCATGA
AATGAATGCCGGCCATGCAGCAGAAGGTTATGCACGTGTTACCGGCGATC
TGGGTGTGGCAGTGGTTACCGCCGGCGGTGGCTTTACCAATGTTCTGACC
AGCATTACCAATGCCTATCTGGATCGTACCCCGGTTCTGTATATTGCAGGT
AGCGGCCCGCTGGATCGCGATCAGATTAATGATCAGCAGGCAGGTTTTGA
TCAGGTGGCCATGGCCGCCCCGGTGACCAGATTTGCCCATCGTGTTACCC
GTGCCGATCTGATTCCGCGTCTGGTTGCCCAGGCAATTCGTACCGCCCAGA
GTGAACCGAAAGGCCCGGTGCTGCTGGATATTCCGTGGGATGTGCTGCGC
CAGACCGTGGATGTTGATGAAGTGGAAGATTATCGCGTGCAGATTGATGG
TACCGGCATTGCCCCGGCAGGTGCCGTTGATCGCATTCTGGAAGCACTGA

GCGCAGCCCGCCGTCCGATTGCCCTGGTGGGTAAAAGCTTTGTGACCGCA
GAAGCACGCGCCCAGCTGCATGAATTTTCAGCCCGTACCGGTGTTCCGCT
GTTTAGTGATTGGGAAGGTCTGGGCGCCATTGTGGGTAGTGAACAGCATG
CCGGCCTGGTGCAGAGCCTGGCAACCGTGCCGGAAGATGAACGTCCGGAT
CTGGTTCTGCTGCTGGGTCTGCGTTTTGGCATGATGACCCAGTTTGGCACC
GGTGGTCTGATGCCGAAAAGCGCCCGTATTTTTCAGATTGATCCGGATGG
TCGTGAACTGGGTCGCCTGCAGGAAGTGGAACTGGGTATTCAGGCAGATC
CGATTGGCACCATTGGTCTGCTGAATGAACGCCTGGGCGATAGCCCGGTG
CCGCCTGGTCGTGATGATTGGCTGAAAACCCTGCGCGATGTGAGCGCAGC
ACGTCGTAGCGCCCTGGCAGCCGAAACCCAGCATCATGAAGGCGCAATTC
ATCCGTATCTGGCCGTTCGCACCATTGCCGAAAGCGTGCCGGATGGCGCC
ACCGTGCTGGTGGATGGCGCACTGACCGAACTGTGGCTGAGCGAAACCAT
TGCACTGGCACCGCTGGCACATTATCTGAATCATGGCTATCTGAGCAGTAT
GGGTAGCAATTTTGGTGTGGCCCTGGGTGCCCAGTATGCAACCCCGGATA
AAGCAACCATTCTGGTTACCGGTGACGGCGCAGTGGGTTATAGCCTGGCA
GAATTTGATACCCTGGTGCGTGCACGTCTGCCGGTTATTGTTATTGTTCTG
AATAATCGTGCCTGGGGCGCCACCCTGCATACCCAGCAGTTTCTGTTTGGC
CAGGATCGTGTTACCAATAATCGCCTGGAAAATGGCAGTTATAGCGGTGT
GGCACGTGCACTGGGTGCAGATGGCGTTGATGTTACCGAACTGGATCAGC
TGCGCCCGGCAATTGAAGCAGCACTGGCAGCACGCCGCCCGACCTGTATT
GATCTGCGCGTTAGCCTGGAACCGATTCCGCCGGAAGAACGTGTGATGAG
TGGCAAAGCCCCGTTTGATGTGGTGCAGACCGATGCCCTCGAG

>BAL-10

Gene sequence:

CATATGGCAACCGTGAATGGTGGTCAGCTGCTGGCCCGTGCCCTGGCACA
GGCTGGTACCACCGAAGTTTTTACCCTGCATGGTGGTCATCTGGATGCATT
TCTGATTGCATGTGCCGGTGAAGGTATTCGCCTGACCGATACCCGTCATGA
AGCAAGCGCAGGTCATGCAGCAGATGCATACGCTCGTGTTACCGGCGGTT
TTGGTGTTTGTGTTGTGACCAGCGGTCCGGGCTTTACCAATGTTTATACCG
CCCTGGCCAATGCCTATCTGGATCGTAGCCCGACCCTGTTTGTGGTTGGTG
CCCCGCCGCTGCGCGAAACCGAAACCAATCCGCTGCAGGGTGGCTTTGAT
CAGATTGCCGCAGCAGATCCGGTTACCAAATGGGCCTATCGTATTACCGA

TGCAGCCCGTGTGCCGGAAATTGTGGCACTGGCCATTCGCAAAACCACCA
GCGGCGTGCCGGGTCCGGTTCTGCTGGAACTGCCGATTGATGTGATGTTTG
GTGAAGCAGATGATGATCAGGTTCGCTTTCCGACCAATTATCGCGTTAGC
ACCCGCAGTGGTGCAGATCCGGATGCAGTTACCCTGGCACTGGATCTGCT
GCAGACCGCAAGTAATCCGGCAATTGTTATTGGTGGTGGTATTACCTTTAG
CCGCGCCGAAGAAGCCCTGGTGGCATTTGCAGAAACCGTGGGCGTGCCGG
TGTTTTATCCGGGCAAAGCAGATGGTGCCATTCCGGCAGATCATCCGCTG
GCCGGTGGCGGTCTGCTGAGTATGGGCACCATTCCGGCACTGGGTGCCCC
GACCCCGGATGTTGTTGTGATGGCAGGTACCCGCGCCGGTATGTTTACCG
GTGGCCGCGCCAGTATGTTTCCGGGCGCAAAAATTATTCAGATTGATATT
GATCCGGCAGAAATTGGTCGCATCTATGATGTGGAAGTGCCGATTGTGGC
CGATTGTCGTGCCGCCCTGGAACAGCTGACCGCCGCTGCTGCCGGCCGTA
CATGGCCTGATTGGACCGAATGGGCCAGCACCGTTAAAGCAGCAAAAGGT
GCCCATGCAGCCGAATTTCCGGATGCCACCACCGATACCGGTAAAATGCA
TCCGTATTTTGCAGCCAAAGCCATTGTGGAAGCATGCCCGCCGGATACCA
TTTTTGTTCTGGATGGCGCAGAAGCACCGAGCTGGGCAGAATTTTTCGTGG
CCGTTGGTCGCCCGGGTAGTGTTCTGCGCCTGGGTTATCTGGGCTGCCTGG
GTGTGGGCCCGGGTTTTGCAATTGGCGCCGCACGTGCCCGTCCGGGTGCT
CCTGTTGTGCTGATTACCGGTGACGGCGCAGCAGGTTTTCATCCGCAGGA
ATTTGATACCATGGCCCGTCATCATCTGCCGGTGACCACCGTGGTGTTTAA
TAATGCCGTTTGGGGCATGAGCATTCATGGTCAGGAAGCAGTTTTTGGTCC
GGAAGGCGTTGTGGTGAGCGAACTGGCCGATAGCGATTATGAAAAAATTG
CAGAAGCCTTTGGTGGCATTGGCATGCGCGTTAAACATCTGGATGATCTG
GCCCCGGCCGTTGAAAAAGCCCTGGGCGCCGATGTTCCGGCCTGTATTAA
TGCAGAAATTGAACCGGGTGTTGTTCATCCGATTACCACCATGATGCTGG
GCGATGTGACCAGCACCGATGAAATTGTTGTGCCGTATTATGAAAATCTG
CCGCGCCTCGAG

>BAL-15

Gene sequence:

CATATGGCAGTTACCGGCGGTGAACTGGTGATTCGTGCCCTGGAACGTGC
AGGCGTTGATGTGGCCTTTGGTATTAATGGTGCACATGTGGATAGTATGTA
TCAGGCCGCCCTGGATCGCAGCTTTCGTATTGTTGATACCCGCCATGAAAT

GAATGCCGGTCATGCCGCCGAAGGCTATGCACGTGCAGGTCATCGTCTGG
GCGTTGCCCTGCTGACCGCCGGTGGTGGTTTTACCAATGCCGTTACCAGTA
TTGCCAATGCATATCTGGATCGCACACCTGTTCTGTATATTGCCGCAAGCG
GCCCGCTGGCCGTGGATGAAACCAATACCCTGCAGGCCGGCATTGATCAG
GTTGCCATTGCAACACCTATTACCCGTTGGGCCCATCGCATTACCCGTGCA
GAACTGATTCCGCGCCTGCTGGCACAGGCAATTCGTATTGCAACCCAGGG
TCCGCGTGGCCCGGTTCTGATTGATATTCCGTGGGATGTGCTGACCGCAGC
AGTTGATGATGATCTGGCAGATGAAGCCCAGGAACTGGGCGCCGATAGCG
TTCTGGCAGCACCTGGCGCCGATGCCGTTAGCCGTATTCTGGATGGTCTGG
CCGCAGCCGAACGCCCGGTTCTGATCGCCGGTAGTGAACTGGTTCGCAGT
GAAGGTGGTGCAGCCCTGCGTCGCCTGGCAGAACGCACCGGTACACCTCT
GTTTAGCGATACCGAAGCACTGGGTGCCATTCGCGAAAGCCCGCTGAGTT
ATGGCCTGCTGCAGGGTCTGTTTGGCCTGGATGAAGGCGAACGTCCGGAT
CGCGTTATTCTGCTGGGTCTGCGTTTTGGTCTGGCAACCGCCCATGGTAGT
GGTATTCTGATTCCGACCGGCAGTGCCGTGGTGCAGATTGATAGTGATGC
ACGCGAACTGGGTCGCCTGCAGCCGATTGAACTGGGTGTGCTGGCCGATG
CCGGCGCAGCTGCTGCAGAACTGGCCGAAGCAGCAGCACGTCGCGGTGAT
TGGCCGGATCGCAGCGCCTGGCAGCGCAGACTGCGCGATCTGGTGGATGG
CCGCTTTGATGCAGTGACCGCACAGGCAGTTCGCGATGATCGTATTCATCC
GATGGATGCAGTTACCGCCATTGCAGAAACCGTTCCGGCCGGCAGTGTGG
TGGTGGCCGATGGTGCACTGACCTATCTGTGGCTGAGCGAAACCATTAGC
CGTGCACCGGTTGCAGATTATCTGTGTCATGGCTATCTGGGTAGCATGGGC
GTGGGCGTGGGCACCGCTCTGGGTGCACAGGCTGCCGATACCGGCCGCCC
TGTTGTTCTGGTGACCGGCGATGGTGCCGTTGGCTATAGCCTGGCAGAATT
TGATAGCATGGTTCGCGCAGGCCTGCCGGTGGTTGTTGTTGTTCTGAATAA
TCGCGCCTGGGGTGCCACCCTGCATGCACAGGAACTGCTGCTGGGCCCGG
ATCGTGTGGTTAATAATCGCCTGGAAAATGGCAGTTATAGCGCAGTGGCA
CGCGCCCTGGGTGCCGATGGTATTAATGTTACCGAACTGGCAGATCTGGT
GCCGGCCCTGGAAAGTGCCCTGGTGAGCGGCCGCCCGACCTGTATTGATG
TGCAGGTGAGCCTGGCACCGGTTCCGCCGGAAGAAAATGTTATTATGGGT
GGCAAACCGTTTCTCGAG

>BAL-23

Gene sequence:

CATATGGCGACGGCGACGGGTAGCGCGTTAGTTGTTCGCACCCTGCAGCG
CGCGGGCGTGACCGTGGCGTTTGGCCTGCCGGGCGCGCATATTGATGGCA
TTCTGCAAGATGCGCTGGATGCGAAACTGCGCGTGGTGGATGTGCGCCAT
GAAGTGAACGCGGGCCATGCCGCGGAAGGCTATGCGCGCGTGACGGGTG
AACTGGGCGTGGCGGTGGTGACCGCGGGCGGTGGCTTTACCAACGTGCTG
ACGAGCGTGGCGAACGCGCATCTGGATCGCAGCCCGGTGCTGTATCTGGC
GGGCAGCGGCCCGCTGGGCACCGATCAGATTAACGATCAGCAAGCGGGC
TTTGATCAAGTGGCGATGGCGGCGCCGGTGACCAAATTTGCGCATCGCGT
GACCCGCACCGAACTGATTCCGCGCCTGGTGGCGCAAGCGATTCGCATTG
CGCGCAGCGAACCGAAAGGCCCGGTGCTGCTGGATATTCCGTGGGATGTG
CTGCGTCAGAGCGTGGATGTGGATGAAGTGGATGATTATCGCATTGAAGT
GGATGGTACCGGCGTGAGTCCGGCGGGCGCGGTTGACCGCATTCTGGAAG
CGTTAAGTGCGGCCCGCCGCCCGATTGCGCTGGTGGGCAAAAGCTTTGTG
ACCGCGGAAGCGCGCGCGCGCCTGCATGAATTTGCGGCGCGCACCGGCGT
GCCGCTGTTTAGCGATTGGGAAGGCCTGGGCGCGATTGTGGGCAGCGAAC
ATCATGTGGGCCTGCTGCAGACCCTGGCGACCGTGCCGGAAGATCAGCGC
CCGGATCTGGTGCTGATGCTGGGCCTGCGCTTTGGCATGGCGACGCAGTTT
GGCACCGGCCGCCTGCTGCCGAAAACGAGCCGCATTTTTCAGATTGATCC
GGATGGCCGCGAACTGGGCCGCCTGCAAGAAGTGGAACTGGGCATTCAA
GCGGATCCGGTGGGCACCATTGGCCTGCTGAACGAACGCCTGAGCGAACG
CGCGGCGCCGAGCGGCCGCGATGATTGGCTGGCGACCCTGCGCGATATTA
GCGCGACCCGCCGCAGCGCGCTGGCGGCGGAGACCGAACAGCATGAAGA
TGCGGCGATTCATCCGTATCTGGCGGTGCGCACCATTGCGGAAAGCGTGC
CGGATCAAGCGACCGTGGTTGTGGATGGCGCGCTGACCGAACTGTGGCTG
AGCGAAACCATTGCGCTGGCGCCGCTGAGCCATTATCTGGGCCATGGCTA
TCTGAGCAGCATGGGCAGCAACTTTGGCGTGGCGCTGGGTGCGCAGTACG
CGGCGCCGGATAAAGCGACCGTGCTGGTGACCGGCGATGGCGCGGTGGG
CTATAGCCTGGCGGAATTTGATAGCCTGGTGCGCGCGGGCCTGCCGGTGG
TTGTGATTGTGCTGAACAACCGCGCGTGGGGCGCGACCCTGCATACGCAG
CAGTTTTTCTTTGGCCAAGATCGCGTGACCAACAACCGCCTGGAAAACGG
CAGCTATAGCGCGGTGGCGCGCGCGCTGGGCGCGGATGGCGTGGATGTGA
CCGAACTGGACCAACTGCGTCCGGCCATTGAGGCGGCGCTGGCCGCGCGC
CGCCCGACCTGCATTGATGTGCGCGTGAGCCTGGCGCCGATTCCGCCGGA

AGAACGCGTGCTGAACGGTGGCGCGCCGTTTGGTGGCATCGAGGTGGATG
CGCTCGAG

>BAL-2 M6

Gene sequence:

CATATGTCAAGCCCAGAAGCAAGGTACACAGGAGGAGATTTACTAGCTCA
AACACTACACGATGCAGGAGTAACAAAAATATTTGCACTACACGGAATAC
ATATCGAAGCATTATTTAAAGGTTGTATAGACCAGGGTATTGATTTGATTG
ACTTTCGTCATGAAGCTGCGGCTGGGCACGCGGCTGACGCGTATGCTCGA
ACCACGGGCAAACTGGGTGTTTGCATCATCACGGCTGGCCCGGGTTTTAC
GAACGCGATTAGCGCGATTGCTAATGCACAACTTGATGCCTCACCGGTAC
TGTTTCTGATAGGTGCGCCGCCGCTGCGTGAAGTGGAAACGAACCCGCTC
CAAGGCGGCATCGACCAGATTGCGATGGCTCGTCCCGCCGCCAAGTGGGC
ACTGTCAATCCCATCCACCGAGCGCGTGCGTGATCTGACCGCGATGGCCA
TACGCAAAGCCATGACCGGACGCAAAGGACCCGTCGTCTTGGAGATTCCA
ATTGATATTCTGCACATGTCCGTGACGGGCGCGCAGGCAACTCCGTCGGC
CGGTCTGGCAGTTCGTCCGCAACCTGCACCGGCACCGGAAGAGGTGGCGG
CGCTGGCCGAATTACTCCTTCGTGCCGAACGCCCTGTCATTGTGGCGGGCT
TAGAATCGGCTTCTGCCGCTACCGCCGTGGCTCTGCGCGCGCTGGTAGCC
AAACTGCCGCTACCAGTCTTCGCTAAACCGCAGGCGTACGGGCTCCTGCC
AGCTGGCCACGCATGTGATGCAGGCGCGGCTGGTAACCTGGCGGTACTTC
CTATAATTGGCGCGGGCGCGCCGGACCTGGTCATCCTGCTGGGGGCGCGC
CTGGGGCTGATGCTGGGTGGTCGCTCCGGTGCCCTGGTGCCTCATGACGCT
CATGTTGTTCAGATCTACAGCGATGCGAGTGAAATCGGCCGTCTTCGCGA
TATCGACCTGCCCATTGCCGCAGATTGCGCTCAAACCTTGACTGCCTTAAC
AAAAGCGCTAGCCGCCGTTGACTTGCCGGACACCTCGGCGTGGACTGCCC
GTGCAGCAGGGGCCAAAGCATTGGCCGCCAGCGCGTGGCCGGATGCAGA
GGTAGCTGGTGGAATTCATCCTTATCATGCCGCGAAAGCGGTTGCAAACG
CTGCGGGCCAGGATGCCGCTTACGTCTTTGATGGCGGTGAATCCGCGAGC
TGGGGTACCGCAACCGTGGCCGTAGACGCTCCGGCCCGGGTGCTGTCACA
TGGCTACCTTTATTGTCTCGGCATCGGCCCGGGCTTTGCCATCGGTATGCA
GATCGCTCACCCCGATCGGCGCGTCGTACAGGTGACTGGCGACGGTGCGA
TGGGATTTCACATCCAGGAATTCGATACTATGGTTCGTCACCGCTTACCGA

TTGTGACAGTGATCCTGAATAATCAAGTATGGGGTCTGTCTATTCACCTAC
AGCAGATGATGTATGGTGCTAACTATAACGTTATTACGAAGCTGGGCTCT
ACGCAGTATGCGTCTATCGCGGCAGCCTTCGGTTGCCATGCTGAGCGAGT
CACTGCATTTGCGGAAATCGCACCTGCGATGGCGCGTGCGTTCGCCAGTG
GCAAACCGGCGCTGGTGGAAATTATGACCGATGCCGACGTTGTCCATCCT
GCCACTGTTGCCATGCTCGGTCAATTAGCTGAAGGTTCTAGAGATATTATG
ATACCCTATTACGAAAACATTGCCGCTTCCCTCGAG

>BAL-3 M5

Gene sequence:

CATATGCCATTAGTATCAGGAGGAGAATTAGTTGTAAGATTATTACAACA
AGCAGGAGTAGACCGAATGTTTGGAATAAACGGAATACATATCGATGCA
GTATATCAGGCAGCATTGGACCATAGATTACCAATTGTTGACACACGCCA
TGAGATGAACGCTGGGCACGCGGCTGAGGGCTACGCGAGGGTGCGCAAC
GCTCTGGGCGTGGCCTTGCTGACGGCTGGCGGTGGTTTTACGAACGCTGT
GACTTCCATGGCCAACGCGCATCTTGATCGCACACCGGTACTGTATCTGGC
TGCAAGTGGCCCGTTAGCCGACGATCAGACGAACACTTTGCAGGCGGGTT
TGGACCAGGTCGCCATTGCTACCCCAGTAACCAAATGGGCCCACCGTGTT
ACTCGCGCCAGCCTGATTCCGCGCATTCTGGCCCGTGCGATACGGACAGC
GCTTTCAGCCCCGAGGGGACCCGTCCTTGTAGATATCCCATGGGATGTGCT
GACCGAACAGGTCGACCTCCCGGACCTGGCGACTGAGCTACTGCACGGTG
CCGATCACGGTCCAGGTTTGTCCCAGGCGGGCGCTGATCGCATCCTGGCT
CTTTTATCGGAGGCGCGCCGACCAGTGGTGGCAGTTGGCAGCGAAGTACC
ACGTAGCGGTGCAGGGGGTGAGGTCCGCGCATTTGCAGAACGCTTAGGCG
TTCCGCTTCTGTCAGATTATGAAGGTCTGGGTGCAATATCTCCGAGCCCAA
TGAATTTCGGCCTGGTGCAGTCGCTGCATGGCCTGCCGGCGGGTGCGGAG
CCGGATCTGGTAATTATGGCAGGCTTGCGCTTTGGTTTGACCACGGCCCAC
GGTTCTGGCGCTTTAATTCCGCGTACAGCAAGGGTGCTGCAGATTGATCCT
AGTGGCGCAGAACTGGGCCACCTGCAACCTGTGGCGTTATCCGTCGCCGC
AGACCCTCTGCCGGTTTTCCAAGATCTCAATCGCGTTTTGGCGGTACGTCA
ATCTGCCGCGCGTCCAGACTTAGCCGATTGGCGTGATACATTACACGCTC
ATTTGCGCCGTCGTCGCGAGGCTGTTGCAGCACAGGTGGTTCAGGACGAA
CGAGTACATCCGTACGATGCGGTTTGCGTAATCGCCGAAGCCCTACCGGA

TGCAAGTGTCGTGGTCGCCGATGGCGCGCTGACCGCATTGTGGCTTAGTG
AAACCATTTCTTCAGCTCCGGTGAGCGCCTTTCTCTGCCATGGCTACCTTT
ATTCCATGGGCGTGGGTATGGGTATCGCGTTGGGCGCCCAGGCCGCGGTG
GAACCAGGCACTAACGTAGTTTTAGTGACTGGCGACGGTTCTGTGGGTTA
CAGCTTGGGGGAATTTGATAGCATGGTACGCGCCGGTCTGGGCGCGGTCG
TGGTGGTGCTGAACAATCGCGCATGGGGCTTAACCTTGCACCTACAGGAA
ATCATGCACGGCAAAGATCGCGTTGTGAATAACAGACTAGAAAATGGTTC
CTACGCAGCCGCGGCGATTGCTCTGGGAGCAGACGGATACCAGGTGCGTA
CCTTGGACGAACTGGCTCCGGCCCTGGCAAAAGCATTACGTTCAGGTCGT
CCCAGCTGTGTCGAAGTCCATGTATCCTTAGATCCTATCCCCCCAGAAGAA
CGCGTAATTATGGGTGGTTCACCGTTCCTCGAG

>BAL-5 M5

Gene sequence:

CATATGGCAACAGTAACAGGATCAGGATTAGTCGTACGAACATTACAAAG
AGCAGGAGTACACGTGGCATTTGGATTACCGGGAATACATATCGATCCCA
TTTTACAAGAAGCATTAGACGCAAAACTACGCGTTGTTGACGTACGTCAC
GAGATGAACGCTGGGCACGCGGCTGAGGGCTACGCGAGGGTAACCGGTG
ATCTGGGCGTGGCCGTGGTGACGGCTGGCGGTGGTTTTACGAACGTCTTA
ACTAGCATCACGAATGCATATCTGGACCGCACCCCGGTACTGTATATTGC
GGGGAGTGGCCCGTTAGATCGCGATCAGATTAATGACCAGCAAGCGGGCT
TCGACCAGGTCGCCATGGCAGCGCCAGTAACCCGCTTCGCACACCGTGTA
ACTCGTGCTGATCTGATTCCGCGCCTGGTGGCGCAGGCGATTCGGACTGC
GCAGAGCGAACCAAAAGGACCCGTCCTTTTAGATATCCCATGGGATGTGC
TGCGCCAGACGGTTGACGTTGATGAGGTAGAAGACTACCGTGTCCAGATT
GATGGTACAGGCATCGCTCCAGCGGGTGCAGTGGACCGTATCCTGGAGGC
CCTGTCCGCCGCACGTCGTCCTATTGCGCTGGTTGGCAAAAGTTTTGTAAC
TGCGGAAGCCCGTGCTCAACTGCATGAATTCAGCGCACGTACCGGCGTTC
CGCTCTTCTCCGATTGGGAGGGTCTGGGTGCAATAGTCGGTTCTGAACAA
CACGCCGGCCTGGTACAGTCGCTGGCTACTGTGCCGGAAGATGAGCGTCC
GGACCTGGTCTTGTTACTAGGCTTGCGCTTTGGTATGATGACTCAGTTCGG
TACGGGCGGTTTGATGCCGAAAAGCGCGCGCATCTTTCAAATCGATCCGG
ATGGGCGCGAACTCGGTCGCTTGCAGGAAGTTGAGCTGGGCATCCAGGCT

GATCCGATTGGCACCATTGGCCTGTTAAACGAACGTTTAGGCGACTCACC
GGTACCGCCAGGAAGAGATGATTGGCTGAAAACACTGCGTGACGTGTCGG
CTGCACGACGGTCGGCACTTGCTGCTGAAACGCAACACCATGAAGGCGCG
ATTCATCCTTATCTGGCCGTTCGCACCATCGCAGAATCAGTCCCCGATGGA
GCCACTGTACTAGTCGATGGCGCGCTGACCGCATTGTGGCTTAGTGAAAC
CATTGCTCTGGCACCTCTGGCACATTATTTGAATCATGGCTACCTTTATTC
CATGGGCTCAAACTTTGGTGTAGCGTTGGGCGCCCAGTATGCCACGCCGG
ATAAGGCGACGATTTTAGTGACTGGCGACGGTGCGGTGGGTTACAGCTTG
GCCGAATTTGATACTCTGGTACGCGCCCGGCTGCCCGTGATCGTGATCGTG
CTGAACAATCGCGCATGGGGCTTAACCTTGCACCTACAGCAGTTCCTGTTT
GGCCAAGATCGAGTTACCAATAACAGACTAGAAAATGGTTCCTACTCGGG
CGTCGCGCGCGCTCTGGGAGCAGACGGAGTGGATGTGACCGAGTTGGACC
AACTGCGCCCGGCTATCGAAGCTGCGCTGGCAGCCCGCCGTCCCACCTGT
ATAGACCTCCGAGTATCACTAGAACCAATCCCCCCAGAAGAAAGAGTAAT
GTCCGGAAAAGCCCCATTTGATGTAGTGCAGACAGACGCCCTCGAG

>BAL-10 M6

Gene sequence:

CATATGGCTACAGTAAACGGGGGGCAACTTTTAGCAAGAGCATTAGCACA
AGCAGGGACCACAGAAGTATTCACACTACACGGAATACATATCGACGCAT
TTTTAATAGCGTGCGCAGGAGAAGGAATCAGGCTGACGGATACCCGTCAT
GAAGCTAGTGCTGGGCACGCGGCTGACGCGTATGCTCGTGTGACCGGGGG
CTTCGGCGTTTGCGTGGTTACCTCGGGCCCGGGTTTTACGAACGTCTACAC
CGCCTTGGCTAATGCTTATCTCGATCGCTCCCCGACCCTCTTTGTGGTAGG
TGCGCCGCCGCTGCGTGAAACTGAAACGAACCCGCTCCAAGGCGGCTTTG
ACCAGATTGCGGCAGCGGATCCGGTAACCAAGTGGGCGTACCGCATTACC
GACGCCGCCAGGGTGCCGGAAATTGTTGCCCTGGCCATACGCAAAACCAC
GAGCGGCGTACCGGGCCCAGTGTTACTGGAGCTGCCCATTGATGTTATGT
TTGGGGAGGCAGACGACGACCAGGTGCGGTTCCCGACAAACTATCGTGTT
TCAACGCGTAGTGGGGCTGATCCAGATGCGGTCACCTTAGCGCTGGATCT
GTTACAAACGGCCTCTAATCCAGCGATTGTGATTGGCGGTGGAATTACCTT
CTCTAGAGCAGAGGAAGCCCTGGTGGCGTTTGCTGAAACGGTCGGCGTTC
CGGTGTTCTACCCGGGTAAAGCGGATGGTGCGATACCAGCTGACCACCCG

TTGGCAGGCGGTGGCCTGCTCAGCATGGGAACGATCCCGGCCCTAGGTGC
GCCGACTCCGGATGTCGTAGTAATGGCCGGCACGCGTGCAGGCATGTTTA
CTGGCGGTCGCGCCTCAATGTTTCCAGGTGCTAAAATTATCCAAATCGATA
TCGACCCGGCGGAAATCGGCCGTATCTATGACGTTGAAGTGCCGATCGTT
GCGGACTGCCGCGCTGCGCTGGAACAGCTTACCGCGGCGGCCGCGGGTCG
TACATGGCCGGACTGGACAGAATGGGCCTCCACTGTGAAAGCAGCGAAG
GGTGCGCATGCGGCGGAATTCCCTGATGCTACTACGGACACCGGCAAAAT
GCACCCTTATTTCGCCGCGAAAGCGATCGTCGAAGCGTGTCCGCCTGATA
CGATTTTTGTGCTTGACGGAGCTGAAGCCGCGAGCTGGGCAGAATTCTTT
GTAGCAGTGGGGCGTCCAGGCTCGGTGCTGCGCCTGGGCTACCTTTATTGT
CTCGGCGTGGGCCCGGGCTTTGCCATCGGTGCTGCACGCGCGCGTCCCGG
CGCACCGGTTGTTCTGATTACTGGCGACGGTGCGGCAGGTTTCCATCCGCA
GGAATTCGATACTATGGCGCGCCATCACCTTCCGGTCACTACCGTCGTTTT
TAACAATGCAGTATGGGGTCTGTCTATTCACCTACAGGAAGCTGTGTTCG
GCCCGGAAGGGGTCGTTGTTAGCGAGCTGGCTGATTCCGATTACGAGAAA
ATCGCAGAAGCATTTGGTGGGATCGGAATGCGTGTCAAGCATCTGGATGA
CTTAGCCCCTGCCGTGGAAAAAGCCTTAGGCGCCGATGTGCCCGCGTGTA
TTAACGCGGAAATTGAACCAGGCGTTGTCCATCCTATTACCACCATGATG
CTAGGAGATGTGACTTCCACTGACGAAATTGTTGTTCCCTATTACGAAAAC
TTACCCCGCCTCGAG

>BAL-15 M6

Gene sequence:

CATATGGCAGTTACGGGAGGAGAATTAGTTATAAGAGCATTAGAAAGAGC
AGGAGTAGACGTGGCATTTGGAATAAACGGAATACATATCGATAGCATGT
ATCAGGCAGCATTGGACAGATCATTTAGAATTGTTGACACACGCCATGAG
ATGAACGCTGGGCACGCGGCTGAGGGCTACGCGAGGGCGGGACATCGTC
TGGGCGTGGCCTTGCTGACGGCTGGCGGTGGTTTTACGAACGCTGTGACTT
CCATTGCCAACGCCTACCTTGATCGCACACCGGTACTGTATATCGCTGCAA
GTGGCCCGTTAGCCGTCGATGAAACGAACACTTTGCAGGCGGGTATCGAC
CAGGTCGCCATTGCTACCCCAATTACAAGATGGGCCCACCGTATTACGCG
TGCTGAACTGATTCCGCGCCTGCTGGCGCAGGCGATTCGGATTGCCACCC
AAGGCCCGAGGGGACCCGTCCTTATCGATATCCCATGGGATGTGCTGACC

GCGGCGGTTGATGATGACTTGGCGGATGAAGCACAGGAGTTAGGCGCAG
ACTCTGTCCTAGCAGCCCCGGGCGCTGACGCGGTGAGCCGCATTCTGGAT
GGCCTGGCCGCAGCTGAGCGTCCGGTGTTAATTGCCGGCAGCGAGCTGGT
CCGAAGCGAAGGTGGTGCCGCGCTCCGTCGCTTAGCGGAACGTACGGGTA
CACCGCTCTTCTCCGATACCGAAGCGCTGGGTGCAATACGTGAAAGCCCG
TTGAGCTATGGGCTGCTCCAAGGTCTGTTTGGCCTTGATGAAGGTGAAAG
GCCAGACCGGGTTATCCTGCTAGGCTTGCGCTTTGGTTTGGCAACGGCCCA
CGGTTCTGGCATCCTGATTCCGACTGGTAGCGCCGTAGTACAGATTGATA
GCGACGCCCGCGAACTCGGTCGCTTGCAGCCTATCGAACTGGGCGTGCTG
GCGGATGCTGGCGCAGCGGCTGCGGAACTGGCGGAGGCGGCTGCGCGTC
GTGGCGATTGGCCGGATCGTTCGGCTTGGCAGCGCCGCCTGCGCGATCTG
GTAGATGGTCGGTTCGATGCCGTTACCGCGCAAGCGGTTCGTGACGACCG
AATTCATCCAATGGACGCGGTAACCGCCATCGCAGAAACGGTGCCGGCCG
GAAGTGTCGTGGTCGCCGATGGCGCGCTGACCGCATTGTGGCTTAGTGAA
ACCATTTCTCGCGCGCCCGTTGCAGATTATCTCTGCCATGGCTACCTTTAT
TCCATGGGCGTGGGTGTTGGTACCGCGTTGGGCGCCCAGGCCGCGGATAC
CGGCCGTCCGGTAGTTTTAGTGACTGGCGACGGTGCGGTGGGTTACAGCT
TGGCCGAATTTGATAGCATGGTACGCGCCGGTCTGCCGGTTGTCGTGGTG
GTGCTGAACAATCGCGCATGGGGCTTAACCTTGCACCTACAGGAACTGTT
ACTTGGTCCCGATCGCGTTGTGAATAACAGACTAGAAAATGGTTCCTACT
CGGCGGTCGCGCGCGCTCTGGGAGCAGACGGAATCAACGTGACCGAGTTG
GCCGATCTGGTGCCTGCCCTGGAGTCAGCTTTAGTTTCAGGTCGTCCCACC
TGTATAGACGTTCAAGTATCACTAGCACCAGTTCCCCCAGAAGAAAATGT
AATTATGGGTGGTAAACCGTTTCTCGAG

>BAL-23 M6

Gene sequence:

CATATGGCAACAGCAACGGGATCAGCATTAGTCGTACGAACATTACAAAG
AGCAGGAGTAACAGTGGCATTTGGATTACCGGGAATACATATCGATGGAA
TATTACAGGACGCATTAGACGCAAAACTACGCGTTGTTGACGTACGTCAC
GAGGTTAACGCTGGGCACGCGGCTGAGGGCTACGCGAGGGTAACCGGTG
AACTGGGCGTGGCCGTGGTGACGGCTGGCGGTGGTTTTACGAACGTCTTA
ACTAGCGTGGCAAACGCCCATCTCGATCGCTCACCGGTACTGTATCTTGCG

GGGAGTGGCCCGTTAGGTACCGATCAGATTAATGACCAGCAAGCGGGCTT
CGACCAGGTCGCCATGGCAGCGCCAGTAACCAAATTCGCACACCGTGTAA
CTCGTACCGAACTGATTCCGCGCCTGGTGGCGCAGGCGATTCGGATTGCC
CGTAGCGAACCAAAAGGACCCGTCCTTTTAGATATCCCATGGGATGTGCT
GCGCCAGAGCGTTGACGTTGATGAGGTAGATGACTACCGCATTGAAGTCG
ATGGTACAGGCGTGAGCCCAGCGGGTGCAGTGGACCGTATCCTGGAGGCC
CTGTCCGCCGCACGTCGTCCTATTGCGCTGGTTGGCAAAAGTTTTGTAACT
GCGGAAGCCCGTGCTCGTCTGCATGAATTCGCGGCACGTACCGGCGTTCC
GCTCTTCTCCGATTGGGAGGGTCTGGGTGCAATAGTCGGTTCTGAACACC
ACGTGGGGCTGCTCCAAACGCTGGCTACTGTGCCGGAAGATCAGCGTCCG
GACCTGGTCTTGATGCTAGGCTTGCGCTTTGGTATGGCCACTCAGTTCGGT
ACGGGCCGTCTGCTGCCTAAAACCTCTCGCATCTTTCAAATCGATCCGGAT
GGGCGCGAACTCGGTCGCTTGCAGGAAGTTGAGCTGGGCATCCAGGCTGA
TCCGGTCGGCACCATTGGCCTGTTAAACGAACGTTTAAGCGAGCGTGCCG
CTCCGTCGGGAAGAGATGATTGGCTGGCGACACTGCGTGACATCTCTGCA
ACGCGACGGTCGGCACTTGCTGCTGAAACGGAACAACATGAAGACGCCG
CCATTCATCCTTATCTGGCCGTTCGCACCATCGCAGAATCAGTCCCCGATC
AGGCGACCGTGGTTGTCGATGGCGCGCTGACCGCATTGTGGCTTAGTGAA
ACCATTGCTCTGGCACCTCTGTCTCATTATCTGGGTCATGGCTACCTTTATT
CCATGGGCTCAAACTTTGGTGTAGCGTTGGGCGCCCAGTATGCCGCGCCG
GATAAGGCGACGGTTTTAGTGACTGGCGACGGTGCGGTGGGTTACAGCTT
GGCCGAATTTGATAGCCTGGTACGCGCCGGTCTGCCGGTTGTCGTGATCGT
GCTGAACAATCGCGCATGGGGCTTAACCTTGCACCTACAGCAGTTCTTCTT
TGGCCAAGATCGAGTTACCAATAACAGACTAGAAAATGGTTCCTACTCGG
CGGTCGCGCGCGCTCTGGGAGCAGACGGAGTGGATGTGACCGAGTTGGAC
CAACTGCGCCCGGCTATCGAAGCTGCGCTGGCAGCCCGCCGTCCCACCTG
TATAGACGTTCGAGTATCACTAGCACCAATCCCCCCAGAAGAAAGAGTAT
TAAACGGCGGCGCACCATTTGGCGGAATAGAAGTAGACGCGCTCGAG

>CV2025

Gene sequence:

ATGCAGAAACAGCGTACCACCTCTCAGTGGCGTGAACTGGACGCTGCTCA
CCACCTGCACCCGTTCACCGACACCGCTTCTCTGAACCAGGCTGGTGCTCG

TGTTATGACCCGTGGTGAAGGTGTTTACCTGTGGGACTCTGAAGGTAACAA
AATCATCGACGGTATGGCTGGTCTGTGGTGCGTTAACGTTGGTTACGGTCGT
AAAGACTTCGCTGAAGCTGCTCGTCGTCAGATGGAAGAACTGCCGTTCTAC
AACACCTTCTTCAAAACCACCCACCCGGCTGTTGTTGAACTGTCTTCTCTG
CTGGCTGAAGTTACCCCGGCTGGTTTCGACAGGGTATTCTACACCAACTCT
GGTTCTGAATCTGTTGACACCATGATCCGTATGGTTCGTCGTTACTGGGACG
TTCAGGGTAAACCGGAAAAAAAAACCCTGATCGGTCGTTGGAACGGTTAC
CACGGTTCTACCATCGGTGGTGCTTCTCTGGGTGGTATGAAATACATGCACG
AACAGGGTGACCTGCCGATCCCGGGTATGGCTCACATCGAACAGCCGTGGT
GGTACAAACACGGTAAAGACATGACCCCGGACGAATTCGGTGTTGTTGCT
GCTCGTTGGCTGGAAGAAAAAATCCTGGAAATCGGTGCTGACAAAGTTGC
TGCTTTCGTTGGTGAACCGATCCAGGGTGCTGGTGGTGTTATCGTTCCGCC
GGCTACCTACTGGCCGGAAATCGAACGTATCTGCCGTAAATACGACGTTCT
GCTGGTTGCTGACGAAGTTATATGCGGCTTCGGTCGTACCGGCGAATGGTT
CGGTCACCAGCACTTCGGTTTCCAGCCGGACCTGTTCACCGCTGCTAAAGG
TCTGTCTTCTGGTTACCTGCCGATCGGTGCTGTTTTCGTTGGTAAACGTGTT
GCTGAAGGTCTGATAGCTGGCGGCGACTTCAACCACGGGTTCACCTACTCG
GGCCATCCGGTCTGCGCAGCTGTTGCTCACGCTAACGTTGCTGCTCTGCGT
GACGAAGGTATCGTTCAGCGTGTTAAAGACGACATCGGTCCGTACATGCAG
AAACGTTGGCGTGAAACCTTCTCTCGTTTCGAACACGTTGACGACGTTCGT
GGTGTTGGTATGGTTCAGGCTTTCACCCTGGTTAAAAACAAAGCTAAACGT
GAACTGTTCCCGGACTTCGGTGAAATCGGTACCCTGTGCCGTGACATCTTC
TTCCGTAACAACCTGATCATGCGTGCTTGCGGTGACCATATCGTAAGCGCTC
CGCCGCTGGTTATGACCCGTGCTGAAGTTGACGAAATGCTGGCTGTTGCTG
AACGTTGCCTGGAAGAATTCGAACAGACCCTGAAAGCTCGTGGTCTGGCT

>ATA

Gene sequence:

ATGACCATCTCTAAAGACATCGACTACTCTACCTCTAACCTGGTTTCTGTTG
CTCCGGGTGCTATCCGTGAACCGACCCCGGCTGGTTCTGTTATCCAGTACTC
TGACTACGAACTGGACGAATCTTCTCCGTTCGCTGGTGGTGCTGCTTGGAT
CGAAGGTGAATACGTTCCGGCTGCTGAAGCTCGTATCTCTCTGTTCGACAC
CGGTTTCGGTCACTCTGACCTGACCTACACCGTTGCTCACGTTTGGCACGG

TAACATCTTCCGTCTGAAAGACCACATCGACCGTGTTTTCGACGGTGCTCA
GAAACTGCGTCTGCAGTCTCCGCTGACCAAAGCTGAAGTTGAAGACATCA
CCAAACGTTGCGTTTCTCTGTCTCAGCTGCGTGAATCTTTCGTTAACATCAC
CATCACCCGTGGTTACGGTGCTCGTAAAGGTGAAAAAGACCTGTCTAAACT
GACCTCTCAGATCTACATCTACGCTATCCCGTACCTGTGGGCTTTCCCGCCG
GAAGAACAGATCTTCGGTACCTCTGCTATCGTTCCGCGTCACGTTCGTCGT
GCTGGTCGTAACACCGTTGACCCGACCGTTAAAAACTACCAGTGGGGTGA
CCTGACCGCTGCTTCTTTCGAAGCTAAAGACCGTGGTGCTCGTACCGCTAT
CCTGCTGGACGCTGACAACTGCGTTGCTGAAGGTCCGGGTTTCAACGTTGT
TATGGTTAAAGACGGTAAACTGTCTTCTCCGTCTCGTAACGCTCTGCCGGG
TATCACCCGTCTGACCGTTATGGAAATGGCTGACGAAATGGGTATCGAATTC
ACCCTGCGTGACATCACCTCTCGTGAACTGTACGAAGCTGACGAACTGATC
GCTGTTACCACCGCTGGTGGTATCACCCCGATCACCTCTCTGGACGGTGAA
CCGCTGGGTGACGGTACCCCGGGTCCGGTTACCGTTGCTATCCGTGACAGG
TTCTGGGCTATGATGGACGAACCGTCTTCTCTGGTTGAAGCTATCGAATAC

>*Mi*TA

Gene sequence:

ATGACCGTTACCGATCAGGGCACCAGTAATCTGGTGGCAGTTGAACCGGGT
GCCATTCGCGAAGATACCCCGCCGGGCAGTGTGATTCAGTATAGTGATTATG
AACTGGATACCAGCAGCCCGTTTGCAGGCGGTGTGGCATGGATTGAAGGC
GAATATATGCCGGCCGAAGAAGCCAAAATTAGTATTTTTGATACCGGCTTTG
GCCATAGTGATCTGACCTATACCGTGGCACATGTGTGGCATGGCAATATTTT
TCGCCTGGGCGATCATCTGGATCGCCTGCTGGATGGTGCCCGTAAACTGCG
CCTGGATGCAGGTTATAGTAAAGATGAACTGGCCGAAATTACCAAAAAATG
CGTTAGTCTGAGCCAGCTGCGTGAAAGCTTTGTGAATCTGACCGTTACCCG
TGGCTATGGTAAACGTAAAGGTGAAAAAGATCTGAGTAAACTGACCCATCA
GGTGTATATTTATGCAATTCCGTATCTGTGGGCATTTCCGCCGGCCGAACAG
ATTTTTGGCACCACCGCCATTGTTCCGCGTCATGTGCGCCGTGCCGGTCGC
AATACCGTGGATCCGACCATTAAAAATTATCAGTGGGGCGATCTGACCGCC
GCAAGCTTTGAAGCCAAAGATCGTGGTGCACGTACCGCCATTCTGCTGGAT
AGTGATAATTGTGTGGCCGAAGGTCCGGGCTTTAATGTTTGTATTGTGAAAG
ATGGTAAGCTGGCCAGCCCGAGCCGTAATGCCCTGCCGGGTATTACCCGCA

AAACCGTTTTTGAACTGGCCGATCAGATGGGTATTGAAGCCACCCTGCGTG
ATGTGACCAGCCATGAACTGTATGATGCAGATGAAATTATGGCCGTGACCA
CCGCAGGCGGTGTTACCCCGATTAATACCCTGGATGGCGAAGCCATTGGCA
GCGGCGAACCGGGTCCGATGACCGTTGCCATTCGCGATCGCTTTTGGGCAC
TGATGGATGAACCGAGTCCGCTGATTGAAGCAATTGAATATTAA*

>*Sl*TA

Gene sequence:

ATGAACAAACATGTGGGCAACAGCTGGGAACAGCGCGATATGGAAAGCGT
GTTTCATGGCTTTACGGATCTGGAAACCCTGAACAAAAACGGCCCGGTGGT
GCTGACCCATGGCGAAGGCATTCATGTGTTTGATGTGCATGGCAAAAGCTA
TATGGATGCGAACAGCGGCCTGTGGAACAACGTGGCGGGCTTTAACCATCC
GGGCCTGATTGAAGCGATTTGCGAACAAGCGCGCCGCTTTCCGGGCTATCA
TGCGTTTTTTGGCCGCGTGGCGGATACCACCGTGGCGCTGAGCGAAAAACT
GGTGAAACTGAGCCCGTTTGAACGCGGCAAAGTGTATTATACCAACAGCG
GCAGCGAAGCGAACGATACCGTGGTGAAAATGCTGTGGATGCTGCATCGC
CGCAACGGTCAGCCGCAGCGCCGCAAAATTATTACCCGCGTGAACGCGTAT
CATGGCGTGACCGTGGCGACCGCGAGCATGACCGGCAAAGCGTATAACGC
GGAATTTGGCCTGCCGCTGCCGGGCTTTCTGCATGCGGATTGCCCGCATTAT
TGGCGCTTTGCGAAACCGGGCGAAAGCGAACTGGAATTTAGTCAGCGCCT
GGCGCGCAACCTGGAAGAACTGATTATTAAAGAAGGCGCGGATACCATTGC
GGGCATGTTTTGCGAACCGGTGCAAGGCGCGGGCGGCGTGATTCCGCCGA
GCGAAGGCTATTTTCAAGCGATTCGCCCGATTCTGCGCAAATATGGCATTCC
GCTGATTGCGGATGAAGTGATTACCGGCTTTGGCCGCACCGGCGAAATGTG
GGGCAGCGTGAAATATGATGTGCAGCCGGATGCGATTGTGGCGAGCAAATG
CATTACCGCGGGCTATTTTCCGATGGGTGCCGTGATTCTGGGCCCGGATCTG
TGCGATGCGCTGACCCGCGTGAGCGAAGAAGCGGAAGAATTTCCGCATGG
CTTTACCGCGGGCGGCAACCCGCTGGGCAGCGCGGTGGCGCTGAAAGCGC
TGGATATTCTGGAAACCGAAGGCCTGCTGGAAAACGTGCGCCGCGTGAGC
ACCCGCTTTCTGGCGGGCCTGAACAGCCTGGCGGAACATAAATATGCGGGC
GAAGCGCGCGGCGTGGGCCTGATGGGTGCGGTGGAGCTGGTGGCGGATAA
AAAAACCAAAGCGCCGCTGGATGGCAAACTGCAAGTGAGCGAACGCATTG
CGAACAAAGCGCTGCAGAAAGGCCTGATTTGCCGCCCGCTGGGCCAAGCG

ATTGTGCTGGGCCCACCGTTTATTATTACCGAAGCGCAGATTGATGAAATGT
TTGATATTCTGCGCGAAACCATGGCGGAAGTGTTTGCGGATGTGGGCCTGT
AA*

>*Pa*TA

Gene sequence:

ATGACCCTGGCAGATAGTGGTACCGATTTTAGTACCAGTAATCTGGTTGCAG
TGGAACCGGGCGCCATTCGTGAAGATACCCCGCCGGGCAGCGTTATTCAGT
ATAGCGATTATGAACTGGATACCAGTAGCCCGTATGCAGGTGGTGCAGCCT
GGATTGAAGGCGAATATGTTCCGGCAAGTGAAGCACGTATTAGTATTTTCG
ATACCGGCTTTGGCCATAGCGATCTGACCTATACCGTTGCACATGTTTGGCA
TGGTAATATTTTCCGTCTGGCAGATCATATTGAACGTCTGCTGGATGGCGCC
CGCAAACTGCGTCTGGCCAGTCCGTATGATGAAACCGAAATTGCCGAAATT
GCCAAACGCTGTGTTGGCCTGAGCCAGCTGCGCGAAGCCTATGTGAATATT
ACCCTGACCCGTGGTTATGGCAAACGTAAAGGTGAAAAGGATCTGAGTAA
ACTGACCAGCCAGATCTATGTTTATGCAATTCCGTATCTGTGGGCATTTCCG
CCGTATGAACAGATTTTCGGTACCAGCGCCGTTGTTCCGCGCCATGTTCAG
CGCGCAGGTCGCAATACCATTGATCCGACCATTAAGAATTATCAGTGGGGT
GACCTGACCGCAGCCAGCTTTGAAGCAAAAGATCGTGGTGCCCGCACCGG
CATTCTGCTGGATGCAGATGGTTGCGTTGCAGAAGGCCCGGGTTTTAATGT
TGTGGTGGTTAAAGATGGCGCCCTGGCCAGCCCGAGTCGTAATGCACTGCC
GGGCATTACCCGTAAAACCGTTTTCGAAATTGCACATGCACGCGGTATTAG
CGCAGAACTGCGCGATGTTACCAGTCGTGAACTGTATGATGCCGATGAACT
GATGGCAGTTACCACCGCAGGTGGCGTTACCCCGATTACCAGCCTGGATGG
CGCGGCAGTTGGTGACGGTGAACCGGGTCCGATTACCGTTGCAATTCGTGA
TCGTTTCTGGGCCCTGATGGATGAACCGAGCGATCTGATTGATACCATTCGT
TATGATGTGGATCGC*

>*Ps*TA

Gene sequence:

ATGGCCATTACCGATAGTGCCAGTGATAGCGCCACCACCAATCTGGTTAGTG
TGGAACCGGGCGCCATTCGCGAAGATACCCCGGCCGGTAGTGTGATTCAGT
ATAGCGATTATGCCCTGGATACCAGCAGCCGTTTTGCCGGTGGCGCCGCCT

GGATTGAAGGCGAATATGTTCCGGCAAGCGAAGCACGTATTAGCATTTTCG
ATACCGGTTTTGGTCATAGCGATCTGACCTATACCGTTGCCCATGTTTGGCAT
GGCAATATTTTCCGCCTGGGTGACCATATTGATCGCCTGCTGGATGGCGCCC
GTAAACTGCGCCTGCCGAGTCCGCTGACCAAAGCAGAACTGGCCAGCATT
GCCAAACGTTGCGTTGCCCTGAGCCAGCTGCGTGAAAGCTATGTGAATATT
ACCCTGACCCGTGGCTTTGGCGCCCGTAAGGGCGAAAAGGATCTGAGTAA
ACTGACCACCCAGGTTTATGTGTATGCCATTCCGTATCTGTGGGCCTTTCCG
CCGCATGAACAGATTTTCGGCACCAGCGCAATTGTTCCGCGTCATGTGCGC
CGTGCCGGCCGTAATACCATTGATCCGACCGTGAAGAATTATCAGTGGGGT
GACCTGACCGCAGCCAGCTTTGAAGCAAAAGATCGTGGTGCACGCACCGC
AATTCTGCTGGATGCCGATGGCTGCGTTGCAGAAGGTCCGGGTTTTAATGT
TGTTATTGTTAAAGATGGCCGCCTGGTTAGTCCGAGCCGCAATGCACTGCC
GGGCATTACCCGCCGCACCGTTTTCGAAATTGCCAATGCAATGGGTATTGAT
GCCGAACTGCGTGATGTGACCAGTCAGGAACTGTATGATGCAGATGAACTG
ATTGCAGTTACCACCGCAGGCGGTGTGACCCCGATTACCAGTCTGGATGGC
GCACCGGTTAGCGATGGTACCCCGGGTAGCATTACCGTGGCACTGCGCGAT
CGCTTTTGGGCCCTGATGGATGAACCGAGTCCGTTAGTTGAAGCCATTGAT
TAT*

>*Ab*TA

Gene sequence:

ATGACCATTAGCAAAGATATCGATTACAGTACCAGTAATCTGGTGAGTGTTG
CCCCGGGTGCAATTCGTGAACCGACCCCGGCCGGCAGTGTTATTCAGTATA
GCGATTATGAACTGGATGAAAGCAGTCCGTTTGCAGGCGGCGCAGCATGG
ATTGAAGGTGAATATGTTCCGGCAGCCGAAGCCCGTATTAGTCTGTTTGATA
CCGGCTTTGGCCATAGTGATCTGACCTATACCGTTGCACATGTGTGGCATGG
CAATATTTTCCGTCTGAAAGATCATATTGACCGTGTTTTCGATGGCGCACAG
AAACTGCGTCTGCAGAGCCCGCTGACCAAAGCAGAAGTTGAAGATATTAC
CAAACGTTGCGTTAGTCTGAGTCAGCTGCGTGAAAGTTTTGTGAATATTAC
CATTACCCGCGGCTATGGTGCACGTAAAGGTGAAAAGGATCTGAGTAAACT
GACCAGCCAGATCTATATCTATGCCATTCCGTATCTGTGGGCCTTTCCGCCG
GAAGAACAGATTTTCGGTACCAGCGCAATTGTGCCGCGTCATGTTCGTCGT
GCAGGTCGCAATACCGTTGATCCGACCGTGAAGAATTATCAGTGGGGTGAC

CTGACCGCAGCAAGTTTTGAAGCAAAAGATCGTGGTGCACGTACCGCAAT
TCTGCTGGATGCAGATAATTGCGTTGCAGAAGGTCCGGGCTTTAATGTGGT
TATGGTTAAAGATGGTAAACTGAGCAGCCCGAGTCGTAATGCCCTGCCGGG
CATTACCCGTCTGACCGTGATGGAAATGGCAGATGAAATGGGTATTGAGTT
TACTCTGCGTGATATTACCAGTCGCGAACTGTATGAAGCAGATGAACTGATT
GCCGTGACCACCGCCGGTGGTATTACCCCGATTACCAGTCTGGATGGTGAA
CCGCTGGGCGATGGTACCCCGGGTCCGGTTACCGTGGCCATTCGCGATCGT
TTCTGGGCAATGATGGATGAACCGAGCAGCCTGGTTGAAGCCATTGAATAT
*

>*Mnf*TA

Gene sequence:

ATGTCAGATGCTCCTATGTTTCATCGCAAAGAAGCTATGTCAGTTCCTATTC
CTACACTTCCTGCTTATCGCGCTGAAACAGGCCTTGATTTCTTGTCGTATTC
AGATTATGAACTTGATGCTGCTGATCCTTTCGCCGGTGGTGCTGCGTATGTT
GATGGCGAATTTGTTCCTCTTGTTGATGCTCGCATTCCTCTTTATGATCAGGG
CTTTCTTCATTCACATCTTACATATACAGTCTTCCACGTGTGGAACGGCCTT
GCTTTCCGTGTAGAAGATCATCTTGATCGCCTTCTTGCTAACGCTGCTGCTA
TTCGCCTTGAAGCTCCTATTTCACGCGATGAAATTCGCGCTGCTTCACTTGA
AATGGTTGCTCGCTCAGGCCTTCGCGAAGCTTATGTTAACATTACAATTACA
GGCGGCTATGCTGATGGCGATCCTGGCCAGCGCCGCTGGGCTCTTCATCGC
CCTCGCGTTTATATGTATGCTGTTCCTTATCTGTTCTTACCGGGCTTTGATGG
CATTCGCGATGGTCTGGCCGCGACCGTGGTTCGCGGCGTTCGCCGCTCACC
TCGCAACGTTATTGATCCTCAGGTTAAGAATTTCCAGTGGGGGGACCTTATT
CGCGCTTCACAGGAAGCTGGCGATCGCGGCTATTCATCAGCTATTCTTCTTG
ATTCAGATGGCCTTGTTGCTGAAGGCCCTGGCTTTAACGTTGCTATTATTCG
CGGCGGCGAACTTCTTACACCTGCTCGCAACGCTCTTCCTGGCATTACACG
CAAGACGTTGCTTGAAATTGGCGAAGAACGCGGGCTCGCCGCGCGTGCCA
CAGATATTACAGCTGCTGAACTTTATGCTGCTGATGAAGTTCTTGGCTGTAC
AACAGCTGGCGGCATTTGGCCTTTCGTGCGAATTGATGATACAGTTATTGGC
GATGGCACACCTGGCCCTCATACACGCGGCATTATTGAACGCTATTGGCAG
CGCTGTGTTGAACCTTCACCTCTTGTTACACCTGTTGATTATCGC

>*Mba*TA

Gene sequence:

ATGACAATTCATACTGAGTTCGCGACCTCCAACCTTGTTTCAGTTGAACCTG
GCGCTATTCGCGAAGATACACCTCCTGGCTCAGTTATTCAGTATTCAGATTAT
GAACTTGATTATTCAACGGAGTTCGCGGGTGGTGCTGCTTGGATTGAAGGC
GAATATGTTCCTGCTTCAGAAGCTCGCATTTCAATATTCGACACGGGCTTTG
GCCATTCAGATCTTACATATACAGTTGCTCATGTTTGGCATGGCAACATCTTC
CGATTGGGCGATCATCTTGATCGCCTTCTTGATGGCGCTCGCAAACTTCGCC
TTGATCCTGGCATGTCAAAGGAGGAGCTTGCTGATATTACAAAGAGATGCG
TTGCTCTTTCAGGCCTTCGCGAATCATTTGTTAACCTTACAGTTACACGCGG
CTATGGCAAACGCAAAGGCGAGAAGGACCTGTCAAAGCTCAATCATCAGG
TTTATATTTATGCTATTCCTTATCTTTGGGCTTTCCCACCGGAAGAACAGATC
TTCGGGACCACAGCTATTGTTCCTCGCCATGTTCGCCGCGCTGGCCGCAAC
ACAGTTGATCCTACAATTAAGAATTACCAGTGGGGAGACTTGACAGCTGCT
TCATTTGAAGCTAAAGATCGCGGCGCTCGCACAGCTATTCTTCTTGATGCTG
ATAACTGTGTTGCTGAAGGCCCTGGCTTTAACGTTGTTATTGTTAAAGATGG
CAAATTATTCAGCCCATCACGCAACGCTCTTCCTGGCATTACACGCAAGAC
TGTGTTCGAGATCGCTGATGCTATTGGCATTGAAGCTACACTTTGTGATGTT
ACATCACGCGAACTTTATGATGCTGATGAACTTATGGCTGTTACAACAGCTG
GCGGCGTTACACCTATTGTTTCACTTGATGGCGAAGCTTTCGGTGATGGGC
GCCCTGGCCCTATTACAGTTGCTATTCGCGATCGCTTCTGGGCGCTTATGGA
TGAACCTTCACCTCTTATTGAAGCTATTGATTAT

>*Atm*TA

Gene sequence:

ATGACATCAATGGGCAAAGTCTTCGCGGGATATGGCGCTCGCCAGGCTGTT
CTTGAAGCCTCCAAGAATCCTTATACACAGGGCATTGCTTGGGTTGAAGGC
GAACTTGTTCCTCTTGCTGAAGCTCGCATTCCTCTTCTTGATCAGGGCTTTC
TTCATTCAGATCTTACTTATGATGTTCCTTCAGTTTGGGATGGCCGCTTCTTC
CGGCTGGATGATCATCTTAAACGCTTAGAGGCCTCCTGCGCCAAACTTCGC
CTTAAACTTGCTCTTCCTCGCGATCAGGTTAAACAGATTCTTGTTGATATGG
TTGTTAAATCAGGCCTTCGCGATGCTTTCGTAGAGGTTATGGTTACACGCGG
CCTTAAATGGATTCGCGGCGCTCGCTCAGAAGAAATTGCTAACAACCTTTAT
ATGTTTGTTCAGCCTTATATTTGGGTTTCAGAACCTGAAGAACAGCGCGGC

GGCGGCGGCCGCGCTGTTATTGCTCGCACAGTTCGCCGCGTTCCTCCTGGC
GCTATTGATCCTACAATTAAGAATTTGCAGTGGGGAGACCTTATGCGCGGCA
TGTTTGAAGCTACAGATCGCGGCGCTATGTATCCTTTCTTAACGGATGGCGA
TGCTCATCTTACAGAAGGCTCAGGCTTTAACATTGTTCTTGTTAAAGATGGC
AAACTTTATACACCTGATCGCGGCGTTCTTGAAGGCATTACACGCCGCTCA
GTTATGGAAGTTGCTAAAGCTAAAGGCATTGAAGTTCGCCTTGAAATGGTT
CCTGTTGAACTTGCTTATAACTGTGATGAAATCTTCATGTGCACAACAGCTG
GCGGCGTTATGCCTATTACATGGCTTGATGGCAAACCTGTTAACGGCGGCG
AAGTTGGCCCTGTTACAAAGGCCATCTGGGATGGCTATTGGGATATGCATTA
TGATGATCGCTATTCATTTGAAATTAAATATAACGAT

>*Amb*TA

Gene sequence:

ATGACAGCTTCAATTGATCATTCAATGTCAAACCTTGTTTCAGTTGAACCTG
GCGCTATTCGCGAAGCTACACCTGAAGGCTCAGTTATTCAGTATTCAGATTA
TGAACTTGATACATCATCAGAATTTGCAGGCGGTGTTGCATGGATTGAAGG
CGAATATGTTCCTGCTTCAGAAGCTCGCATTTCAATCTTCGACACGGGCTTT
GGCCATTCAGATCTTACATATACAGTTCCTCATGTTTGGCATGGCAACATATT
CCGTCTCGGCGATCATCTTGATCGCCTTCTTGATGGCGCTCGCCGCCTTTAT
CTTGATCCTGGCTATACAAAGGAGGAACTGGCCGACGTCGTGAAGAAATGT
GTATCACTTTCACAGCTTCGCGAATCATTTGTTAACATTACAATTACACGCG
GCTTTGGCCAGCGCCGCGGCGAGAAGGACTTATCAAAGCTAACGCATCAG
GTTTATATTTATGCTATTCCTTATCTTTGGGCTTTCCCGCCAGATGAACAGAT
CTTCGGAACGTCAGTTATTGTTCCTCGCCATGTTCGCCGCGCTTCACGCTCA
ACAATTGATCCTACAATTAAGAATTACCAGTGGGGCGACCTGACAGCTGCT
TCATTTGAAGCTAAAGAACGCGGCGCTCGCTCAGCTATTCTTCTTGATGCTG
ATAACTGTGTTGCTGAAGGCCCTGGCTTTAACGTTGTTCTTGTTAAAGATGG
CGAATTATTCAGCCCGTCTAGAAACGCTCTTCCTGGCATTACACGCAAGAC
TGTGTTCGAACTAGCGGACGTGATGGGCGTACCGGCTACACTTCGCGATGT
TACATCACGCGAACTTTATGAAGCTGATGAACTTATGGCTGTTACAACAGC
AGGTGGAGTCACTCCGATTATTGAACTTGATGGCGAACCTCAGTCAGGCGG
CAAACCTGGCCCTCTTACAACAGAAATTCGCGATCGCTTCTGGGCCCTTAT
GGATGAACCTTCAGATCTTATTGAAGCTATTCAGTAT

4 References

1. M. E. Smith, U. Kaulmann, J. M. Ward and H. C. Hailes, *Bioorg. Med. Chem.,* 2006, **14**, 7062-7065.

2. D. Baud, N. Ladkau, T. S. Moody, J. M. Ward and H. C. Hailes, *Chem. Commun.,* 2015, **51**, 17225-17228.

3. The PyMOL Molecular Graphics System, Version v 2.5.2 (Schrödinger LLC).

4. Y. Zhang, Y. Li, Y. Chen, W. Liu, Q. Zhao, J. Feng, P. Yao, Q. Wu and D. Zhu, *ACS Catal.,* 2024, **14**, 9687-9700.

5. A. Li, C. G. Acevedo-Rocha and M. T. Reetz, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 6095-6103.

6. Y. Li, P. Yao, S. Zhang, J. Feng, H. Su, X. Liu, X. Sheng, Q. Wu, D. Zhu and Y. Ma, *Chem Catal.,* 2023, **3**, 100467.

7. D. Güclü, A. Szekrenyi, X. Garrabou, M. Kickstein, S. Junker, P. Clapés and W.-D. Fessner, *ACS Catalysis*, 2016, **6**, 1848−1852.

8. H. Teller and A. Furstner, *Chemistry,* 2011, **17**, 7764-7767.

9. E. Callaway, *Nature,* 2024, **635**, 531-532.

10. Lu, T., Sobtop Version 1.0 (dev3.1). http://sobereva.com/soft/Sobtop (accessed on 17-Aug-2022).

11. T. Lu and F. Chen, *J. Comput. Chem.,* 2011, **33**, 580-592.

12. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.

13. J. V. d. S. Guerra, H. V. Ribeiro Filho, L. O. Bortot, R. V. Honorato, J. G. d. C. Pereira and P. S. Lopes-de-Oliveira, *SoftwareX*, 2020, **12**.

14. 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.

15. M. Ripoll, S. Velasco-Lozano, E. Jackson, E. Diamanti, L. Betancor and F. López-Gallego, *Green Chem.*, 2021, **23**, 1140-1146.

16. Z. Tan, Z. Tang, H. Wei, R. Zhang, L. Sun, W. Liu, H. Liu, L. Zhu and Y. Ma, *ACS Catal.,* 2025, **15**, 1586-1595.

17. Y. Wu, G. Lu, R. Zeng, Z. Li, X. Xu, Y. Gui, C. Guo, L. Deng, Y. Bie, D. Zhang, Y. He, Y. He, Y. Zhu, C. Fu and L. Yu, *Chem. Eng. J.* , 2025, **505**, 159573.