# Reverse Engineering: Transaminase Biocatalyst Development Using Ancestral Sequence Reconstruction

Matthew Wilding,*[a] Thomas S. Peat,[b] Subha Kalyaanamoorthy,[a] Janet Newman,[b] Colin Scott,[a] and Lars S. Jermiin.[a,c]

## Supplementary Materials

### Materials and Methods

*Multiple Sequence Alignment*

An initial set of 249 amino-acids sequences, encoding putative transaminases, was obtained by BLAST analysis of KES23360 at Uniprot.[1] An MSA for the amino-acid sequences was inferred using MAFFT[2] (v. 7.043; MAFFT was chosen because of its accuracy[3, 4]), and Seaview was used to visualise the alignment (v. 4.4.1).[5] A preliminary MSA was inferred for the data using the L-iNS-I option of MAFFT. Using Seaview, the preliminary alignment was manually refined to yield the final alignment.

*Removal of Duplicate Sequences*

Before further analysis of the data, duplicate sequences were identified, and then removed from the final alignment using IQ-TREE (v. 0.9.3).[6] A reduced alignment of 192 unique sequences was produced, realigned as described above and called the "master alignment".

*Preliminary Survey of Master Alignment*

To assess whether the data could be assumed to have evolved under globally SRH conditions, we surveyed the master alignment using Homo (v. 1.0). The 18,336 *p*-values obtained (i.e., one for each pair of sequences) were plotted against the expected distribution of *p*-values (i.e., that which would be obtained if the sequences had evolved under globally SRH conditions) and finally interpreted.

*Identifying the Optimal Model for Sequence Evolution*

The optimal model of sequence evolution for the master alignment was identified by IQ-TREE (v. 0.9.3) using the -m TESTONLY option with IQ-TREE invoked. The optimal model was found to be the LG+I+G4 model.

*Phylogenetic Analysis*

The most likely tree was inferred using the LG+G4 model of sequence evolution. In addition, a non-parametric bootstrap analysis (using the UFBoot method with default settings) was done to determine the consistency of the data. Both procedures were executed using IQ-TREE. The inferred ML tree was used to infer the ancestral sequences.

*Ancestral Sequence Reconstruction*

Ancestral amino-acid sequences were inferred under the ML criterion using FastML[7], and assuming the LG+G4 model. Both marginal and joint estimates of ancestral sequences were obtained for all the ancestral nodes in the ML tree. Selected ancestral sequences were then identified and chosen for further biochemical analysis (see below).

*Subcloning, Expression and Purification*

Synthetic genes corresponding to the six ancestral proteins were cloned in the same manner as the KES23360 transaminase protein. Each gene was obtained as an *E. coli* codon-optimised synthetic gene in pMA-RQ vector[8] (Invitrogen). DNA for each gene, and the desired expression vector, pETcc2, was digested using *Nde*I and *Bam*HI restriction endonucleases (NEB), separated using a 0.8% agarose gel and purified using a Gel Extraction Kit and following the Manufacturer's Guidelines (Macherey Nagel). Linear gene and vector fragments were ligated using T4 DNA ligase (NEB) and the DNA was transformed into *E. coli* BL21 λDE3 electrocompetent cells by electroporation. Visualised transformants were used to inoculate 10 mL LB medium containing ampicillin (100 µg/mL) and grown overnight at 37 °C with shaking. DNA was isolated using a Bioline Miniprep Kit following manufacturers guidelines and sequenced to validate the final construct for each gene (Macrogen, S. Korea).

Protein overexpression was achieved by growing 500 mL of *E. coli* BL21 λDE3 containing the desired vector in LB media containing ampicillin (100 µg/mL) at 37°C. When the $OD_{600}$ reached 0.6-1.0, the cultures were induced by the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside; 1 mM final concentration) and further incubated at 15 °C for 18 hours. The cells were isolated by centrifugation (4000 x *g*; 20 minutes) and the supernatant discarded. The pellet was resuspended in imidazole/sodium chloride/potassium phosphate buffer (5 mM/500 mM/10 mM, pH 7.5) and cell lysis was achieved using an Avestin C3 Emulsiflex Homogeniser at 20 kPSI. Cellular debris was pelleted by centrifugation (40,000 x *g*, 45 minutes) and the supernatant was passed over a HiTrap Chelating HP column (GE Healthcare) on an Åkta FPLC (Fast Protein Liquid Chromatography, GE Healthcare) Protein was eluted with an increasing concentration of imidazole (5-500 mM) and the separated protein was transferred into potassium phosphate buffer (10 mM, pH 7.5), concentrated by centrifugation (GE Healthcare; 10k MWCO) and further purified by gel filtration (Superdex 200; G.E. Healthcare) in the same phosphate buffer. Purity was estimated to be >95% by SDS-PAGE.

*Crystallography*

All protein samples were analysed for buffer preference and stability using DSF.[9] Table S1 shows stability values for KES23360 and ancestral proteins, based on triplicate melt curves.

Initial crystallisation trials were performed with the KES23360 protein. The protein sample (5 mg/mL in 10 mM $KPO_4$ buffer, pH 7.5) was used in initial screens (JCSG+, PACT, PS gradient – for details of screens see http://c6.csiro.au). Drops were set up with 200 nL protein mixed with 200 nL reservoir solution and equilibrated against 50 µL of reservoir in SD-2 plates (Molecular Dimensions, UK). Yellow, blunt ended hexagonal bipyramidal crystals grew out of conditions containing 2 M ammonium sulfate at 8 °C. The initial conditions were refined by screening around the initial condition, and by additive optimisation. Each of the ancestral peptides was first screened as supplied (conditions shown in Table S2) in the JCSG screen at 20 °C. The ancestral peptides were transferred into different buffer formulations, based on the results of the DSF experiments, and the reformulated proteins were screened under the optimum conditions determined from the crystal screening, with and without seeding. Seeds from the N16 protein were used for all of the ancestral peptides. Final crystallisation conditions for each of the six ancestral peptides, along with the buffer formulations used, are shown in Table S2.

Crystals were cryoprotected and frozen in liquid nitrogen before data collection at the Australian Synchrotron. The KES23360 data set was collected using 0.5° oscillations due to the long unit cell, whereas the ancestral peptide data sets were collected using 1.0° oscillations. Full datasets were obtained by collecting either 180° or 360° depending on the space group (X-ray data is shown in Table S2). Data were indexed using XDS[10] and scaled using Aimless.[11] The KES23360 structure was solved using PDB code 3GJU as the molecular replacement model in Phaser[12] and all other structures were solved using Phaser with KES23360 as the model. The models were rebuilt manually in Coot[13] and refined using Refmac.[14]

KES23360 was cryoptotected with glycerol. N16, N43 and N48 were cryoprotected with AP/E core 150; N6, N15 and N17 were cryoprotected with the reservoir solution (which contained glycerol) and supplemented with additional glycerol to a final concentration of 25%.

*Activity Assays*

Transaminases were assayed using previously described methods.[15] Activity for each of the transaminases was assessed using enzyme-coupled dehydrogenase assays. A typical assay comprised: 6.25 mM substrate, 0.5 mM pyruvate, 1.25 mM nicotinamide adenine dinucleotide ($NAD^+$), 0.035 U of alanine dehydrogenase (ADH; where 1 U corresponds to the amount of enzyme which converts 1 µmol L-alanine per minute at pH 10.0 and 30°C), 2 - 50 nM transaminase, potassium phosphate (100 mM, pH 10). The rates of the transaminases were inferred from the coupled rate of $NAD^+$ turnover by alanine dehydrogenase, which was dependent on the production of co-product (alanine) by the transaminase. $NAD^+$ turnover was measured by the change in UV absorbance at 340 nm using a SpectraMax M2 spectrophotometer (Molecular Devices, Australia); reactions were conducted at 28 °C. Kinetic parameters were obtained using the above method, recording initial rates of activity across a range of substrate concentrations. Parameters were subsequently calculated using non-linear regression.

*Molecular Dynamics Simulations*

Models of KES23360, N6 and N43 were prepared in Accelrys Discovery Studio v 3.5 from the experimentally derived structure as a starting point. A model for the HMD:PLP external aldimine intermediate and PLP (for N43) were created in Accelrys Discovery Studio v 3.5 and prepared for MD using the Full Minimization tool in Discovery Studio v 3.5 using the default settings (CHARMm forcefield). HMD-PLP aldimine complexes were manually orientated in the active site using the electron density from native PLP as a guide. For N43 no PLP was solved in the initial X-ray structure, so this was constructed *in silico* and manually orientated based on alignment against KES23360 and N16. Atomic charges were calculated in Accelrys Materials Studio v8.0 using the QEq method and the substrates were manually docked into the active site. Ligands were prepared for MD in AMBER 14 [16] using the GAFF forcefield and the protein models were solvated in a TIP3P octahedral solvent box with a minimum 12 Å periodic boundary distance and charge-neutralised by the addition of $Na^+$ ions. Proteins were prepared using the ff99sb forcefield.

Initial minimisation of both systems was performed using AMBER 14 over 10,000 steps under a constant pressure of 1 Bar. Bonds lengths on bonds involving hydrogen were constrained in SHAKE and force evaluation on these bonds was not performed. 200 ns MD simulations with a step-size of 0.002 ps were performed at 310 K and 1 bar pressure with a 2 ps relaxation time. Trajectories were

analysed using VMD (v. 1.9.2) [17]. Analysis was conducted on 150 ns of the simulation, removing the first 50 ns to ensure the systems had equilibrated as determined through RMSD analysis.

Figure S1 – The complete phylogenetic tree constructed for KES23360. The highlighted region contains the subtree investigated further and contains the KES23360 sequence (blue). Unless shown, all calculated bootstrap values were 100.

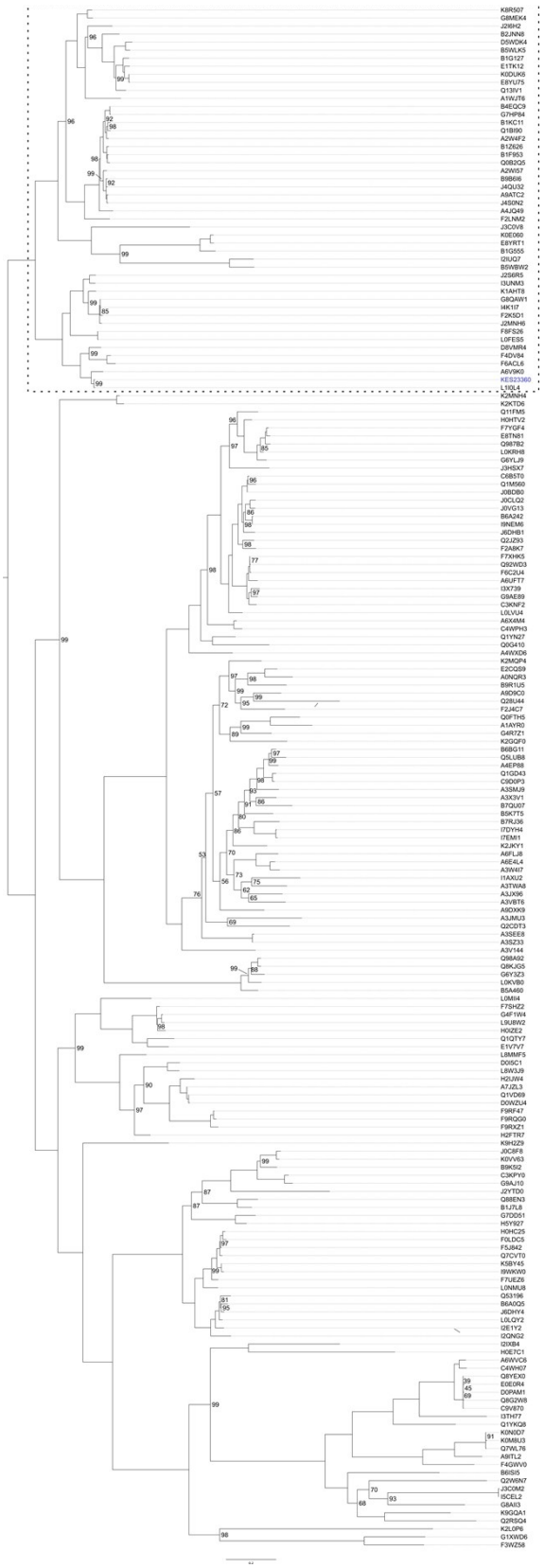
K8R507
G8ME64
J2I6H2
B2JNN6
D5W0K4
B5WLK5
B1G127
E1TK12
K0DUK5
E8YU75
Q13IV1
A1WJT6
B4EQC9
G7HP64
B1KC11
Q1BI90
A2W4F2
B1Z6Z6
B1F953
Q0B2Q5
A2WI57
B9B6I6
J4QU32
A9ATC2
J4S0N2
A4JQ49
F2LNK2
J3C0V8
K0E060
E8Y8T1
B1G555
I2IJQ7
B5WBW2
J2S6R5
I3UNM3
K1AHT8
G8QAW1
I4K1I7
F2K5D1
J2MNH6
F8FS26
L0FES5
D8VMR4
F4DVB4
F6ACL6
A6V9K0
KES23360
L1I0L4
K2MNH4
K2KTD6
Q11FM5
H0HTV2
F7YGF4
E8TN81
Q9ET82
L0KRH8
G6YLJ9
J3HSX7
C6B5T0
Q1M560
J0BDB0
J0CLQ2
J0VG13
B6A242
I9NEM6
J6DHB1
Q2JZ93
F2ABK7
F7XHK5
Q9ZWD3
F6C2U4
A6UFT7
I3X739
G9AE89
C3KNF2
L0LVJ4
A6X4M4
C4WPH3
Q1YN27
Q0G410
A4WXD6
K2MQP4
E2CQE9
A0NQR3
B9R1U5
A8QNC0
Q2BU44
F2J4C7
Q0FTH5
A1AYR0
G4RTZ1
K2GQF0
B6BG11
Q5LUB8
A4EP88
Q1GD43
C9D0P3
A3SMJ9
A3X3V1
B7QU07
B9KTT5
B7RJ36
I7DYH4
I7EMI1
K2JKY1
A6FLJ8
A6E4L4
A3WI47
I1AXU2
A3TWA8
A3JX96
A3VBT6
A9DXK9
A3JMU3
Q2CDT3
A3SEE8
A3SZ33
A3V144
Q9BA92
Q8KJG5
G6YJZ3
L0KVB0
B5A460
L0MII4
F7SHZ2
G4F1W4
L9U8W2
H0IZE2
Q1QTY7
E1V7V7
L8MMF5
D0I9C1
L8WIJ9
H2IJW4
A7JZL3
Q1VD69
D0NZU4
F9RF47
F9RQG9
F9RXZ1
H2FTR7
K9HZZ9
J0C8F8
K0VV63
B9K5I2
C3KPY0
G9AJ10
J2YTD0
Q8BEN3
B1J7L8
G7DD51
H5Y927
H0HC25
F0LDC5
F5J842
Q7CVT0
K5BY45
I9W6W0
F7UEZ6
L0MMJ8
Q53196
B6A0Q5
J6DHY4
L0LQY2
I2E1Y2
I2QNG2
I2IXB4
H0E7C1
A6WVC6
C4WH07
Q8YEX0
E0E0R4
D0PAM1
Q8G2W8
C9V870
I3TH77
Q1YKQ8
K0N0D7
K0M8U3
Q7WL76
A9ITL2
F4GWV0
B6ISI5
Q2W6N7
J3C0M2
I5CEL2
G8AI33
K9GQA1
Q2RSQ4
K2L0P6
G1XWD6
F3WZ58

Figure S2 – Multiple Sequence Alignment of the seven peptides. Generated using Clustal Omega. Average Sequence Identity is 67.1%.

```
KES23360      MTDYAKLFEQDRAHFMHPSTHAHDHASGALPGRIITGASGVRIRDHQGRELLDAFAGLYC
N6            MTDFDQLFEQDRAHFMHPSTHAHDHASGALPGRIITGASGIRIRDHEGRELIDAFAGLYC
N15           MTDLDQLFEMDRAHFMHPSTHAHDHASGALPGRIITGGKGIRIQDHEGREYIDAFAGLYC
N16           MQSLDQLFEMDRAHFMHPSTHAHDHASGALPGRIITGGKGIRIEDHEGREYIDAFAGLYC
N17           MTSLDQLFEEDRAHFMHPSTHAHDHASGALPGKIVTGGKGIRIEDHQGREYIDAFAGLYC
N43           MTSLEQLLEMDRAHFMHPSTHAYDHASGALPGRIITGGKGIRIEDHEGREYIDAFAGLYC
N48           MTDTDDLLEMDRAHFFHPSTHLRDHASGELPGRIITGGKGIRIQDSEGREYIDAFAGLYC
              * .  .*:* *****:*****  ***** ***:*:**..*:**.* :*** :********

KES23360      VNIGYGRLEVADAIHEQAKQLAYYHTYVGHASEAIIELSARIIRDWAPAGMKKVYYGLSG
N6            VNIGYGRTEVADAIYKQAKELAYYHTYVGHSTEAIIELSSRIIRDWAPAGMKKVYYGLSG
N15           VNIGYGRTEVADAIYEQAKELAYYHTYVGHSTEAIIELSSRIIRDWAPAGMKKVYYGMSG
N16           VNIGYGRTEVADAIYEQAKQLAYYHTYVGHSTDAIIELSSRIIRDWAPAGMKKVYYGMSG
N17           VNIGYGRTEVADAIYEQAKQLAYYHTYVGHSTDAIIELSSRIIRDWAPAGMKKVYYGMSG
N43           VNIGYGREEVADAIYEQAKQLAYYHTYVGHSNDPVIELSSRIIEDWAPAGMKKVFYGMSG
N48           VNVGYGRTEIADAIYEQAKELAYYHTYVGHSNEPIIELSERIIREWAPAGMSKVYYGMSG
              **:**** *:****::***:**********:.: :**** ***.:******.**:**:**

KES23360      SDANETQVKLVRYYNNVLGRPQKKKIISRQRGYHGSGIVTGSLTGLASFHQHFDLPVEGV
N6            SDANETQIKLVRYYNNVLGRPQKKKIISRQRGYHGSGIMTGSLTGLPSFHQHFDLPVEGI
N15           SDANETQIKLVWYYNNVLGRPQKKKIISRQRGYHGSGIMTGSLTGLPSFHQHFDLPIERI
N16           SDANETQIKLVWYYNNVLGRPNKKKIISRERGYHGSGIVTGSLTGLPSFHQHFDLPIDRV
N17           SDANETQIKIVWYYNNVLGRPNKKKIISRERGYHGSGIVTGSLTGLPSFHQHFDLPIDRV
N43           SDANETQIKLVWYYNNVLGRPNKKKIIARERSYHGSGIVTGSLTGLPSFHQHFDLPIDRV
N48           SDANETQIKLVWYYNNVLGRPQKKKIISRQRGYHGSGIMTGSLTGLPAFHNHFDLPLEPI
              *******:*:* *********:*****:*:*.******:******* :**:*****:: :

KES23360      KHTLCPHFYKAPAGMDEAAFVRHCAQELENLILAEGPDTVAAFIGEPVMGTGGIIVPPKG
N6            KHTVCPHWYKAPAGMDEAAFVRYCADELEKLILAEGPDTVAAFIGEPVMGTGGIIVPPKG
N15           KHTVCPHWYKAPAGMSEAQFVRYCADELEKLILAEGPDTVAAFIGEPVMGTGGIIPPPQG
N16           KHTVCPHWYRAPAGMSEAQFVAYCVEELEKLIAREGADTIAAFIAEPVMGTGGIIPPPQG
N17           KHTVCPHWYKAPAGMSEAQFVAYCVEELEKLIAREGADTIAAFIAEPVMGTGGIIAPPQG
N43           KHTVCPHWYNAPPGMSEAQFVAYCVEELEKLIAREGADTIAAFIAEPVMGTGGIVPPPQG
N48           RHTTCPHYYRAPAGMSEAEFSRHCADELEKMILAEGPDTVAAFIGEPVMGTGGIVPPPEG
              :** ***:*.** **.** *  :*.:***::*  ** **:****.*********: **:*

KES23360      YWEAIQAVLAKYDVLLIADEVVCAFGRLGDKMGSQRHAMRPDLITTAKGLTSAYAPLSAV
N6            YWEAIQAVLNKYDVLLIADEVVCAFGRLGSKMGSQRYGMRPDLITTAKGLTSAYAPLSAV
N15           YWEAIQAVLNKYDILLIADEVVCGFGRLGSKMGSQHYGMKPDLITVAKGLTSAYAPLSGV
N16           YWEAIQAVLRKHDILLIADEVVCGFGRLGSKMGSQHYGIKPDLITVAKGLTSAYAPLSGV
N17           YWEAIQAVLRKHDILLISDEVVCGFGRLGSKMGAQHYGIKPDLITVAKGLTSAYAPLSGV
N43           YWEAIQAVLRKHDILLIADEVVCGFGRLGSKTGSEHYGIKPDLITVAKGLTSAYAPLSAV
N48           YWEAIQAVLNKYDILLIADEVVCGFGRTGSMFGSHHYGMKPDLITVAKGLTSAYAPLSGV
              ********* *:*:***:*****.*** *.  *:.::.::*****.************.*

KES23360      IVGEKVWDVIDSASTREGAMGHGWTYSGHPICAAAALANLDILERENITANAADVGGYLN
N6            IVGEKVWDVIEKASQKEGAMGHGWTYSGHPICAAAALANLDILERENLTANAADVGAYLN
N15           IVGEKVWDVIEKGSQEHGPMGHGWTYSGHPICAAAALANLDILERENLTGNAADVGAYLQ
N16           IVGEKVWDVIEKGSQEHGPMGHGWTYSGHPICAAAALANLDILERENLTGNAADVGAYLQ
N17           IVGEKVWDVIEKGSQEHGPMGHGWTYSGHPICAAAALANLDILERENLTGNAADVGAYLQ
N43           IVSEKVWDVIEKGSREHGVMGHGWTYSGHPVCAAAALANLDILERENLTGNAADVGAYLQ
N48           IVGEKVWKVLEQGSDQYGPIGHGWTYSGHPICAAAALANLDIIERENLTGNAADTGAYFQ
              **.****.*::..* . * :**********:***********:****:*.****.*.*::

KES23360      QQLRQAFEGHPLVGEVRGDGMLAALEFMADREARTPFDAALKVGPKVSAACLERGMIARA
N6            QRLRETFEGHPLVGEVRGDGMLAALEFMADREARTPFDPALKVGPKVSAACLEDGMIARA
N15           QRLRETFGGHPLVGEVRGVGMLAALEFMADKDARTPFDPALKVGPKVSAACLEDGMIARA
N16           QRLHEAFGAHPLVGEVRGVGMLAALEFMADKDARTPFDPALKVGPKVSAAALEDGMIARA
N17           QRLHEAFGAHPLVGEVRGVGMLAALEFMADKDARTPFDPALKVGPKVSAAALEDGVIARA
N43           QRLHEAFGAHPLVGEVRGVGMLAALEFMADKGARTPFDPALKVSQKVAAAALEDGLIVRA
N48           QRMRETFGDHPLVGEVRGVGLMAALEFVADKDKRTRFDPSLKVGPRVSAACLEDGMIARA
              *:::::*  ********* *::*****:**:  ** ** :***. :*:**.** *:*.**

KES23360      MPHGDILGFAPPLVLSRAEADEVVGIAKAAVDAVAAEVL
N6            MPHGDILGFAPPLVLTRAEADEIVGIAKAAVDEVAGEVL
N15           MPHGDILGFAPPLVITRAEVDEIVGIAKQAVDEVADEVL
N16           MPHGDILGFAPPLVTTRAEVDEIVGIAKQAVDEVADEVL
N17           MPHGDILGFAPPLVTTRAEVDEIVGIVKQAVDEVADEVL
N43           LPHGDILGFAPPLVTTRAEVDEIVAIAKEAFDEVADAVL
N48           MPHGDILGFAPPLVITRAEVDEIVDIAKQAVDAVADELV
              :************* :***.**:* *.* *.* **  ::
```

Figure S3 – The X-ray crystal structure of KES23360 (PDB accession number 5KQT) shown as a monomer (top left) and dimer (top right). The protein is shown in a cartoon representation with the pyridoxal-5'-phosphate (PLP) cofactor shown in stick representation. Bottom – an overlay of the seven proteins illustrating the high overall structural conservation between the proteins.

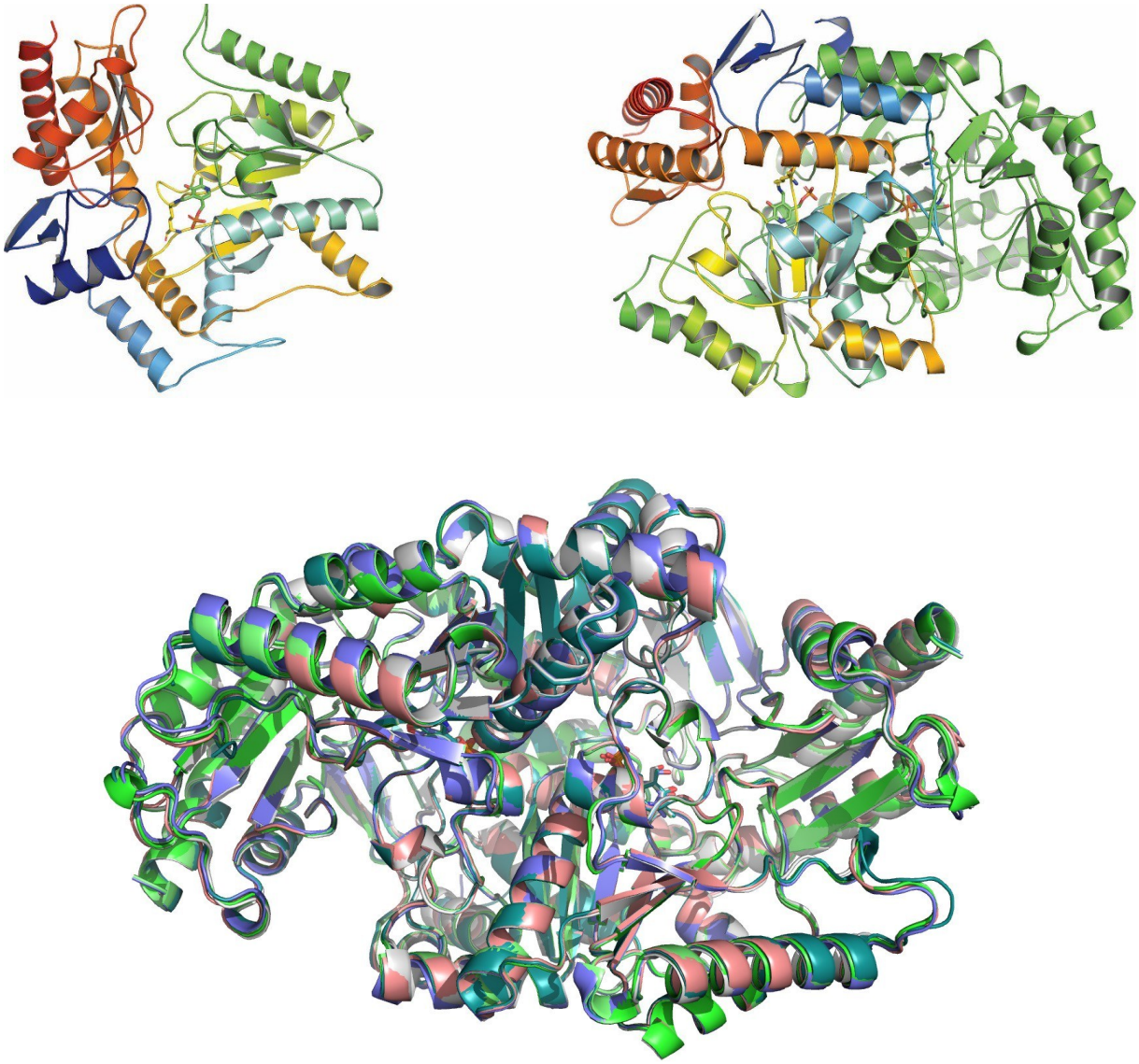

Table S1 – DSF analysis

| **Protein** | **Melt temp ($KPO_4$ buffer; °C)** | **Highest Melt temp (°C)** | **Preferred buffer** | **Comments** |
|---|---|---|---|---|
| KES23360 | 57 | 64 | 50 mM ADA pH 6.5, 200 mM NaCl | |
| N6 | 62 | 69.5 | 50 mM MOPS pH 7, 200 mM NaCl | |
| N15 | 61 | 64.5/72.5 | 50 mM ADA pH 6.5, 200 mM NaCl | Two-part unfolding. |
| N16 | 66 | 70 | 50 mM Tris pH 8, 200 mM NaCl | |
| N17 | 62.5 | 67.5 | 50 mM Tris pH 8, 200 mM NaCl | |
| N43 | 62 | 69.5 | 50 mM ADA pH 6.5, 200 mM NaCl | |
| N48 | 49/56.5 | 46/64 | 50 mM MOPS pH 7, 200 mM NaCl | Two-part unfolding. |

Table S2 – Crystallisation Conditions and X-ray data

| **Protein** | **Protein concentration (mg/mL)** | **Crystallisation Conditions** |
|---|---|---|
| KES23360 | 5 | 1.6 M ammonium sulfate, 1% dioxane, 50 mM MES pH 6.8 |
| N6 | 10 | Tris Chloride (0.1 M, pH 8.53), magnesium chloride (0.155 M), PEG 8000 (15.2 % w/v) and glycerol (12.8 % v/v) |
| N15 | 10 | Tris Chloride (0.1 M, pH 7.15), magnesium chloride (0.152 M), PEG 8000 (14.9 % w/v) and glycerol (17.1 % v/v) |
| N16 | 10 | Ammonium formate (0.135 M) and PEG 3350 (20.1 % w/v) |
| N17 | 10 | Tris Chloride (0.1 M, pH 7.14), magnesium chloride (0.152 M), PEG 8000 (11.9 % w/v) and glycerol (14.8 % v/v) |
| N43 | 4 | Ammonium formate (0.216 M) and PEG 3350 (16.1 % w/v) |
| N48 | 10 | Magnesium nitrate (0.03 M) and PEG 3350 ( 15.6 % w/v) |

| **PDB** | 5KQT (KES23360) | 5KQU (N6) | 5KQW (N15) |
|---|---|---|---|
| Space group | P6522 | P21 | P21 |
| Cell dimensions | | | |
| *a*, *b*, *c* (Å) | 66.4, 66.4, 343.3 | 74.1, 146.9, 131.3 | 66.1, 125.2, 110.5 |
| α, β, γ (°) | 90, 90, 120 | 90, 97.9, 90 | 90, 102.0, 90 |
| Resolution (Å) | 47.8-1.99 (2.10-1.99) | 49.0-2.62 (2.67-2.62) | 47.2-2.23 (2.27-2.23) |
| $R_{merge}$ | 0.136 (0.720) | 0.138 (0.742) | 0.130 (0.739) |
| $R_{pim}$ | 0.030 (0.160) | 0.082 (0.445) | 0.125 (0.707) |
| *CC1/2* | 0.998 (0.910) | 0.995 (0.740) | 0.988 (0.654) |
| $I / \sigma I$ | 17.6 (4.9) | 11.3 (2.5) | 9.2 (2.3) |
| Completeness (%) | 99.9 (99.3) | 99.7 (95.1) | 99.8 (99.1) |
| Redundancy | 20.7 (20.7) | 7.5 (6.9) | 3.8 (3.6) |
| | | | |
| **Refinement** | | | |
| Resolution (Å) | 47.8 - 1.99 | 49.0 - 2.62 | 47.2 - 2.23 |
| Unique Reflections | 30,450 | 79,422 | 81,119 |
| $R_{work}$ / $R_{free}$ (%) | 19.6/ 23.7 | 18.8/ 22.0 | 16.4/ 19.2 |
| No. atoms | 3,682 | 21,243 | 14,665 |
| Protein | 3,473 | 21,007 | 14,071 |
| PLP / Other | 15 | 90 | na |
| Water | 182 | 146 | 564 |
| *B*-factors (Å$^2$) | 23.9 | 41.2 | 23.0 |
| Protein | 24.0 | 41.3 | 23.1 |
| PLP / Other | 22.8 | 51.1 | na |
| Water | 26.5 | 26.4 | 21.3 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.009 | 0.010 | 0.013 |
| Bond angles (°) | 1.389 | 1.504 | 1.560 |

| **PDB** | 5KR3 (N16) | 5KR4 (N17) | 5KR5 (N43) | 5 KR6 (N48) |
|---|---|---|---|---|
| Space group | P21 | P212121 | C2221 | P21 |
| Cell dimensions | | | | |
| *a*, *b*, *c* (Å) | 63.5, 122.2, 64.3 | 67.1, 121.8, 125.6 | 98.5, 102.5, 199.8 | 61.1, 123.2, 63.3 |
| α, β, γ (°) | 90, 116.5, 90 | 90, 90, 90 | 90, 90, 90 | 90, 117.6, 90 |
| Resolution (Å) | 41.9-1.95 (1.99-1.95) | 45.8-2.00 (2.04-2.00) | 48.6-2.10 (2.15-2.10) | 48.8-1.99 (2.04-1.99) |
| $R_{merge}$ | 0.077 (0.665) | 0.156 (0.726) | 0.170 (0.899) | 0.075 (0.320) |
| $R_{pim}$ | 0.047 (0.404) | 0.092 (0.437) | 0.101 (0.532) | 0.072 (0.309) |
| *CC1/2* | 0.998 (0.965) | 0.993 (0.756) | 0.994 (0.670) | 0.995 (0.859) |
| $I / \sigma I$ | 20.4 (3.4) | 10.4 (2.4) | 9.5 (2.2) | 13.0 (3.4) |
| Completeness (%) | 98.9 (96.3) | 99.8 (97.1) | 99.6 (97.6) | 98.5 (98.1) |
| Redundancy | 7.3 (7.0) | 7.3 (7.0) | 7.3 (7.3) | 3.7 (3.6) |
| **Refinement** | | | | |
| Resolution (Å) | 41.9 - 1.95 | 45.8 - 2.00 | 48.6 - 2.10 | 48.8 - 1.99 |
| Unique Reflections | 58,359 | 66,296 | 56,252 | 53,036 |
| $R_{work}$ / $R_{free}$ (%) | 21.9/ 25.7 | 19.2/ 22.4 | 20.4/ 24.0 | 19.5/ 23.9 |
| No. atoms | 7,527 | 7,504 | 7,403 | 7,336 |
| Protein | 7,051 | 7,015 | 6,991 | 7,088 |
| PLP / Other | 30 | 30 | 30 | 30 |
| Water | 446 | 447 | 358 | 218 |
| *B*-factors (Å$^2$) | 18.2 | 21.1 | 29.3 | 22.7 |
| Protein | 18.2 | 21.2 | 29.6 | 22.8 |
| PLP / Other | 31.1 | 11.5 | 26.1 | 22.6 |
| Water | 21.2 | 23.7 | 28.7 | 22.7 |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.013 | 0.013 | 0.009 | 0.013 |
| Bond angles (°) | 1.563 | 1.565 | 1.424 | 1.533 |

Table S3 - Specific Activity data with errors. Specific activity data in nmol/min/mg. Errors are shown to one decimal place or significant figure. N.D. – Activity not detectable

| Substrate | KES23360 | N6 | N15 | N16 | N17 | N43 | N48 |
|---|---|---|---|---|---|---|---|
| β-Alanine | 3.2 ± 0.08 | 7.8 ± 0.01 | 4.3 ± 0.02 | 0.4 ± 0.01 | 1.2 ± 0.04 | 1.8 ± 0.07 | 0.1 ± 0.002 |
| 4-Aminobutyrate | 48.1 ± 0.7 | 60.9 ± 2.0 | 21.9 ± 0.4 | 4.2 ± 0.06 | 9.0 ± 0.05 | 49.7 ± 0.3 | 1.4 ± 0.004 |
| 5-Aminopentanoate | 81.4 ± 1.8 | 56.3 ± 2.2 | 19.5 ± 0.6 | 5.5 ± 0.03 | 9.6 ± 0.1 | 76.5 ± 0.5 | 2.1 ± 0.01 |
| 6-Aminohexanoate | 81.4 ± 1.8 | 57.7 ± 2.4 | 20.5 ± 0.6 | 4.5 ± 0.01 | 9.1 ± 0.1 | 74.7 ± 0.5 | 1.7 ± 0.01 |
| 7-Aminoheptanoate | 78.6 ± 1.4 | 56.5 ± 2.9 | 18.0 ± 0.06 | 4.2 ± 0.02 | 8.5 ± 0.2 | 70.6 ± 0.5 | 1.9 ± 0.01 |
| 8-Aminooctanoate | 64.9 ± 1.1 | 37.9 ± 2.4 | 17.5 ± 0.6 | 3.5 ± 0.06 | 7.6 ± 0.2 | 61.0 ± 0.3 | 1.9 ± 0.01 |
| 12-Aminododecanoate | 8.2 ± 0.3 | 10.2 ± 0.08 | 4.6 ± 0.2 | 0.6 ± 0.01 | 1.8 ± 0.05 | 18.7 ± 0.3 | 0.2 ± 0.002 |
| 4-$NH_2$-(*S*)-2-hydroxybutyrate | 95.1 ± 0.7 | 27.0 ± 1.0 | 14.9 ± 0.5 | 4.6 ± 0.03 | 7.5 ± 0.04 | 50.6 ± 0.1 | 2.2 ± 0.01 |
| 2,4-(*S*)-Diaminobutyrate | 24.0 ± 0.3 | 6.8 ± 0.03 | 5.3 ± 0.03 | 1.5 ± 0.007 | 3.7 ± 0.04 | 15.9 ± 0.1 | 0.5 ± 0.003 |
| 1,3-Diaminopropane | 0.7 ± 0.07 | 1.4 ± 0.05 | 1.1 ± 0.006 | 0.1 ± 0.003 | 0.5 ± 0.01 | 1.8 ± 0.04 | 0.02 ± 0.002 |
| Putrescine | 15.1 ± 0.3 | 38.8 ± 0.2 | 17.9 ± 0.09 | 2.8 ± 0.08 | 7.9 ± 0.04 | 32.3 ± 0.4 | 0.3 ± 0.003 |
| Cadaverine | 4.7 ± 0.3 | 48.6 ± 0.5 | 21.5 ± 0.6 | 2.6 ± 0.1 | 3.5 ± 0.1 | 62.0 ± 0.5 | 0.3 ± 0.002 |
| 1,6-Hexamethylenediamine | 5.4 ± 0.3 | 17.9 ± 0.3 | 13.8 ± 0.2 | 2.0 ± 0.02 | 6.7 ± 0.05 | 37.8 ± 0.9 | 0.5 ± 0.01 |
| 1,7-Heptamethylenediamine | 5.7 ± 0.2 | 29.6 ± 0.3 | 14.0 ± 0.2 | 2.1 ± 0.02 | 6.8 ± 0.2 | 48.7 ± 0.2 | 0.6 ± 0.01 |
| 1,8-Octamethylenediamine | 6.8 ± 0.3 | 24.2 ± 0.5 | 13.9 ± 0.2 | 1.7 ± 0.03 | 4.6 ± 0.1 | 37.4 ± 0.4 | 0.3 ± 0.01 |
| 1,9-Nonamethylenediamine | 12.2 ± 0.4 | 20.6 ± 0.7 | 12.1 ± 0.2 | 1.7 ± 0.03 | 6.0 ± 0.2 | 43.7 ± 0.2 | 0.5 ± 0.005 |
| 1,10-Decamethylenediamine | 7.2 ± 0.2 | 47.3 ± 0.3 | 13.7 ± 0.2 | 2.0 ± 0.02 | 6.1 ± 0.1 | 38.3 ± 0.3 | 0.6 ± 0.006 |
| Glycine | N.D. | N.D. | N.D. | N.D. | N.D. | 0.2 ± 0.02 | 0.2 ± 0.002 |
| L-Lysine | 2.2 ± 0.1 | N.D. | N.D. | 0.1 ± 0.009 | 2.2 ± 0.05 | 1.8 ± 0.05 | N.D. |
| D-Lysine | 3.9 ± 0.03 | 10.0 ± 0.09 | 6.2 ± 0.08 | 1.0 ± 0.01 | 3.2 ± 0.05 | 11.4 ± 0.07 | 0.1 ± 0.001 |
| L-Ornithine | 28.3 ± 0.3 | 29.1 ± 0.7 | 14.8 ± 0.06 | 1.5 ± 0.03 | 6.1 ± 0.1 | 19.1 ± 0.4 | 0.3 ± 0.002 |
| D-Ornithine | 2.5 ± 0.2 | 4.9 ± 0.03 | 3.3 ± 0.02 | 0.5 ± 0.0007 | 1.4 ± 0.01 | 5.0 ± 0.1 | 0.03 ± 0.002 |
| L-Arginine | 0.7 ± 0.05 | 1.7 ± 0.04 | 1.0 ± 0.02 | 0.2 ± 0.003 | 0.5 ± 0.007 | 1.4 ± 0.03 | N.D. |
| D-Arginine | N.D. | N.D. | 0.2 ± 0.005 | 0.1 ± 0.001 | 0.1 ± 0.002 | N.D. | N.D. |
| L-Glutamate | N.D. | N.D. | 0.4 ± 0.01 | 0.03 ± 0.001 | N.D. | N.D. | N.D. |
| Citrulline | N.D. | 0.9 ± 0.01 | 0.8 ± 0.007 | 0.1 ± 0.002 | 0.2 ± 0.006 | 1.8 ± 0.1 | N.D. |
| N-Acetyl-L-Ornithine | 0.7 ± 0.03 | 1.2 ± 0.05 | 1.1 ± 0.03 | 0.1 ± 0.001 | 0.3 ± 0.01 | 0.9 ± 0.1 | N.D. |
| Creatine | N.D. | N.D. | N.D. | N.D. | N.D. | N.D. | N.D. |
| Taurine | N.D. | N.D. | 0.3 ± 0.004 | N.D. | N.D. | N.D. | N.D. |
| 3-Aminoheptanoate | N.D. | 0.7 ± 0.02 | 0.5 ± 0.02 | 0.1 ± 0.002 | 0.4 ± 0.01 | 0.5 ± 0.02 | 0.03 ± 0.002 |
| 3-Aminocyclohexanoate | 0.7 ± 0.04 | 2.0 ± 0.01 | 1.0 ± 0.006 | 0.1 ± 0.002 | 0.4 ± 0.003 | 7.3 ± 0.1 | 0.04 ± 0.002 |
| 3-D/L-Aminoisobutyrate | N.D. | N.D. | 0.5 ± 0.009 | 0.1 ± 0.002 | 0.1 ± 0.003 | 0.5 ± 0.02 | N.D. |
| Gabapentin | N.D. | N.D. | N.D. | N.D. | N.D. | N.D. | N.D. |
| Serinol | N.D. | 0.5 ± 0.01 | 0.6 ± 0.02 | 0.1 ± 0.002 | 0.2 ± 0.01 | 0.9 ± 0.04 | N.D. |
| β-Homoleucine | N.D. | N.D. | 0.4 ± 0.009 | 0.1 ± 0.0009 | 0.1 ± 0.006 | N.D. | 0.03 ± 0.002 |
| α-Methylbenzylamine | 1.8 ± 0.1 | 10.2 ± 0.09 | 3.0 ± 0.02 | 0.4 ± 0.005 | 3.0 ± 0.04 | 9.6 ± 0.1 | 0.1 ± 0.002 |
| 2-Aminoindan | 0.7 ± 0.1 | 9.7 ± 0.06 | 4.7 ± 0.06 | 0.8 ± 0.01 | 3.2 ± 0.03 | 14.6 ± 0.2 | 0.1 ± 0.003 |
| Isopropylamine | N.D. | 1.2 ± 0.02 | 0.6 ± 0.01 | 0.1 ± 0.001 | 0.5 ± 0.03 | 4.1 ± 0.07 | 0.04 ± 0.003 |
| Cyclohexylamine | N.D. | 2.4 ± 0.04 | 1.2 ± 0.009 | 0.2 ± 0.002 | 0.9 ± 0.03 | 5.5 ± 0.2 | 0.1 ± 0.003 |
| 6-Aminohexanol | 4.3 ± 0.2 | 13.8 ± 0.1 | 6.4 ± 0.2 | 1.1 ± 0.01 | 4.9 ± 0.05 | 21.9 ± 0.1 | 0.3 ± 0.01 |

References


1. The UniProt Consortium, *Nucleic Acids Research*, 2014, **42**, D191-D198.
2. K. Katoh and D. M. Standley, *Molecular Biology and Evolution*, 2013, **30**, 772-780.
3. B. P. Blackburne and S. Whelan, *Bioinformatics*, 2012, **28**, 495-502.
4. T. Golubchik, M. J. Wise, S. Easteal and L. S. Jermiin, *Molecular Biology and Evolution*, 2007, **24**, 2433-2442.
5. M. Gouy, S. Guindon and O. Gascuel, *Molecular Biology and Evolution*, 2010, **27**, 221-224.
6. B. Q. Minh, M. A. T. Nguyen and A. von Haeseler, *Molecular Biology and Evolution*, 2013, **30**, 1188-1195.
7. H. Ashkenazy, O. Penn, A. Doron-Faigenboim, O. Cohen, G. Cannarozzi, O. Zomer and T. Pupko, *Nucleic Acids Research*, 2012, **40**, W580-W584.
8. Mammalian Gene Collection, pMA Vector Map, http://mgc.nci.nih.gov/Vectors/prot_pMA.
9. S. A. Seabrook and J. Newman, *ACS Combinatorial Science*, 2013, **15**, 387-392.
10. W. Kabsch, *Acta Crystallographica Section D*, 2010, **66**, 125-132.
11. P. Evans, *Acta Crystallographica Section D*, 2011, **67**, 282-292.
12. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *Journal of Applied Crystallography*, 2007, **40**, 658-674.
13. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallographica Section D*, 2010, **66**, 486-501.
14. G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallographica Section D*, 2011, **67**, 355-367.
15. M. Wilding, E. F. A. Walsh, S. J. Dorrian and C. Scott, *Microbial Biotechnology*, 2015, **8**, 665-672.
16. D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, I. Cheatham, T. E., T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, University of California, San Francisco, 2014.
17. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33-38.