Protein Engineering, Design & Selection vol. 26 no. 1 pp. 25–33, 2013
Published online September 25, 2012 doi:10.1093/protein/gzs065

Redesigning and characterizing the substrate specificity and activity of *Vibrio fluvialis* aminotransferase for the synthesis of imagabalin

Katarina S.Midelfort[1,4], Rajesh Kumar[1], Seungil Han[1,4], Michael J.Karmilowicz[1], Kevin McConnell[1], Daniel K.Gehlhaar[2], Anil Mistry[1], Jeanne S.Chang[1], Marie Anderson[1], Alan Villalobos[3], Jeremy Minshull[3], Sridhar Govindarajan[3] and John W.Wong[1]

[1]Pfizer Worldwide Research and Development, Eastern Point Road, Groton, CT 06340, USA, [2]Pfizer Worldwide Research and Development, 10555 Science Center Drive, San Diego, CA 92121, USA and [3]DNA 2.0, 1140 O'Brien Drive, Suite A, Menlo Park, CA 94025, USA

[4]To whom correspondence should be addressed. Midelfort@msoe.edu (K.M.); Seungil.Han@Pfizer.com (S.H.)

Received May 30, 2012; revised August 21, 2012; accepted August 23, 2012

Edited by David Ollis

Several protein engineering approaches were combined to optimize the selectivity and activity of *Vibrio fluvialis* aminotransferase (Vfat) for the synthesis of (3*S*,5*R*)-ethyl 3-amino-5-methyloctanoate; a key intermediate in the synthesis of imagabalin, an advanced candidate for the treatment of generalized anxiety disorder. Starting from wild-type Vfat, which had extremely low activity catalyzing the desired reaction, we engineered an improved enzyme with a 60-fold increase in initial reaction velocity for transamination of (*R*)-ethyl 5-methyl 3-oxooctanoate to (3*S*,5*R*)-ethyl 3-amino-5-methyloctanoate. To achieve this, <450 variants were screened, which allowed accurate assessment of enzyme performance using a low-throughput ultra performance liquid chromatography assay. During the course of this work, crystal structures of Vfat wild type and an improved variant (Vfat variant r414) were solved and they are reported here for the first time. This work also provides insight into the critical residues for substrate specificity for the transamination of (*R*)-ethyl 5-methyl 3-oxooctanoate and structurally related β-ketoesters.

Keywords: crystal structure/enzyme engineering/ β-ketoesters/protein GPS/*Vibrio fluvialis* aminotransferase

Introduction

There is increasing recognition of the potential of α and ω-aminotransferase enzymes to aid in the production of chiral amines, as well as natural and unnatural amino acids, without the use of expensive and sometimes hazardous metal catalysts (Koszelewski *et al.*, 2010; Savile *et al.*, 2010). Among various (*S*)-selective ω-aminotransferases, a *Vibrio fluvialis* aminotransferase (Vfat) is of interest for the synthesis of chiral amines, kinetic resolution of racemic amines and asymmetric transamination of ketones (Koszelewski *et al.*, 2010). Vfat is an ω-amino acid:pyruvate transaminase that has been reported to show catalytic activity toward aliphatic amines, though with limited substrate scope (Cho *et al.*, 2008). There have also been partially successful attempts to increase the substrate specificity of Vfat for aliphatic amines (Hwang and Kim, 2004; Cho *et al.*, 2008). However, the yields were inadequate and the required reaction conditions inappropriate for pharmaceutical manufacture. Progress has also been hampered by a lack of structural data for the enzyme, although homology models have provided some indications of important pockets in the protein (Cho *et al.* 2008).

Figure 1 shows a route for the synthesis of imagabalin, a late-stage candidate for generalized anxiety disorder. Several approaches have been reported previously for the synthesis of (3*S*,5*R*)-ethyl 3-amino-5-methyloctanoate (intermediate **2** in Fig. 1) using chemical catalysis (Birch *et al.*, 2011a,b). Although these routes could provide material for clinical trials, they had projected costs that were too high for commercial manufacturing. Enzymatic transamination was identified as the most direct asymmetric and potentially cost-effective route for the synthesis of intermediate **2** from (*R*)-ethyl 5-methyl 3-oxooctanoate (substrate **1** in Fig. 1). An initial screen of 44 transaminases identified Vfat and two very similar enzymes from *Rhodobacter sphaeroides* and *Paracoccus denitrificans*, as enzymes able to catalyze the desired reaction, albeit with very low activity.

The Vfat enzyme was selected as a target to be optimized for the conversion of substrate **1** to intermediate **2** (Fig. 1). Key challenges for this protein engineering were the limited literature information about productive mutational hotspots and the absence of structural data (Cho *et al.*, 2008). Furthermore, the need to obtain a stereoselective enzyme required the use of a low-throughput ultra performance liquid chromatography (UPLC) screen. To overcome these obstacles we integrated homology modeling, bioinformatics, machine learning, crystal structure analysis and site saturation mutagenesis of specific sites to identify and overcome kinetic and stability constraints on productivity. Designing and testing a total of $<$450 variants resulted in a 60-fold improvement in initial activity and an enzyme capable of the preparative transamination of substrate **1**.

We also report the crystal structure of the wild-type enzyme, which shows some similarity to the prior homology models, but also highlights some differences. The crystal structure of the improved variant provides the basis for an understanding of the link between most of the mutations and the increased rate of reaction for the desired substrate. The crystal structures and other mutational data reported herein should provide a good basis for future Vfat enzyme engineering toward other substrates. Analysis of the kinetic parameters of several variants highlights the change in specificity obtained through the engineering of Vfat.

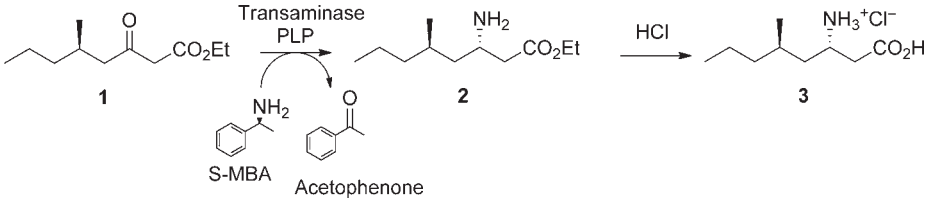

Fig. 1. Synthesis of imagabalin (**3**) via transamination of (*R*)-ethyl 5-methyl-3 oxooctanoate (**2**), with S-MBA as the amine donor, using Vfat ω-transaminase variants.

Materials and methods

Genes and plasmids

The Vfat gene was synthesized by DNA 2.0. The gene was cloned into the pET28b vector (EMD Biosciences, Billerica, MA, USA). Luria broth (LB) media was from Invitrogen (Grand Island, NY, USA). Overnight Express Auto Induction media was from MolBio Inc. (Boulder, CO, USA). All other reagents were from Aldrich (St Louis, MO, USA) or VWR (Radnor, PA, USA) unless noted.

Site specific mutations (either single mutations or saturation mutagenesis)

The QuikChange kit from Stratagene (Agilent Technologies, Santa Clara, CA, USA) was used for all single or saturation site-specific mutations as directed. DNA primers were ordered from Integrated DNA Technologies (Coralville, IA, USA).

Homology models

Homology models of the *V. fluvialis* protein structure were created by two methods. The first used the Accelrys Discovery Studio (San Diego, CA, USA) homology model protocols and modeled the structure on the PDB structure 1D7R. The second model used the Schrödinger Prime program (New York, NY, USA) and was based on the 2EO5 PDB structure. Both models provided structures which were within 1.06 Å root-mean-square deviation (RMSD) of each other. The desired substrate was also modeled into the binding pocket of the first homology model by inspection. The desired substrate was modeled into the binding pocket of the second homology model by alignment with the wild-type inhibitor found in the crystal structure. When compared with the wild-type Vfat crystal structure that was eventually solved, the two homology modeled structures had a heavy main chain backbone atoms (N, Ca, C and O) RMSD of 3.5 Å.

Protein expression

The pET28b vector with the aminotransferase gene was transformed into BL21(DE3) Gold cells (Stratagene, Agilent Technologies) by standard heat shock methods. The cells were plated on LB/kanamycin (50 μg/ml) plates and grown overnight in a 37°C incubator. A 3 ml LB culture was then started from a single colony and grown at 37°C, 210 rpm overnight. For induction of the protein expression, 1 ml of the LB starter culture was transferred to 50 ml of Overnight Express TB (MolBio Inc.) and put in a 250 ml baffled, vented flask and put at 30°C for 24 h. Cells were centrifuged at 3200 × *g* to pellet. The pellets were stored at −20°C.

General screening

Screening reactions were carried out at 0.5 ml scale in a 96-well plate (2 ml capacity with round bottom) containing 100 mM of ketone substrate, 100 mM of (*S*)-methylbenzylamine (S-MBA) as an amine donor (used as 1 M solution adjusted to pH 7.0 using hydrochloric acid), 2 mM pyridoxal phosphate and 80 g/l *Escherichia coli* cells of Vfat variants in 100 mM potassium phosphate buffer (pH 7.0). After adding all the components, the plate was sealed and incubated at 30°C and 900 rpm (Thermomixer R, Eppendorf, Hamburg, Germany). The reaction progress was monitored by an UPLC analysis after derivatization of the reaction samples with Marfey's reagent at various time intervals.

For an UPLC analysis, reaction samples (50 μl) were quenched by adding 450 μl of acetonitrile and mixing for 5 min at 750 rpm. The quenched reaction samples (50 μl each) were then transferred to another 96-well plate (2 ml capacity with round bottom) and treated with 1 M aqueous sodium bicarbonate (5 μl) and Marfey's reagent (*N*-α-(2,4-dinitro-5-fluorophenyl)-L-alaninamide (200 μl of a 5 g/l solution in acetonitrile) for 1 h at 40°C. The derivatization reaction was quenched by adding 5 μl of 1 M HCl, diluted with 50 μl of acetonitrile and analyzed by UPLC.

UPLC analysis was performed on an Acquity UPLC® system from Waters using a C18 column (Waters BEH C18: 2.1 × 50 mm (1.7 μm), Waters Corp., Milford, MA, USA). Analyses were run at 30°C using max plot (210–400 nm) for detection. For analysis, a gradient solvent system of 1% triethylamine (pH adjusted to 3 with phosphoric acid) from 62 to 45% in 5 min in acetonitrile with a flow rate of 0.8 ml/min was used. Between injections, the system was equilibrated for 2 min using 38% of acetonitrile in 1% triethylamine solution.

Basic bioinformatics

A simplistic bioinformatics approach used a ClustalW alignment (Chenna *et al.*, 2003) of the only nine homologous sequences with >50% sequence identity to the Vfat protein sequence identified by a BLAST search of the non-redundant protein sequences database in NCBI (Altschul *et al.*, 1997) (GenInfo Identifier (gi): 6993337, 1462769, 1264627, 7746395, 1267256, 8905361, 8613959, 1103470 and 1342881). Separately, using the structure of the *E. coli* gamma-aminobutyrate aminotransferase protein (PDB ID: 1SF2), all residue locations within 15 Å of the center atom of the pyridoxyl 5′-phosphate (PLP) in this structure were identified. This gave a list of 55 non-conserved sites at the intersection of the non-conserved positions in the alignment and the positions identified to be within 15 Å of the active site. The alternate amino acids, different from the *V. fluvialis* wild-type

protein, were noted for each site. This created a list of 70 single mutations at 55 sites, some sites having more than one alternate amino acid possible. These 70 mutations were made as point mutations in the wild-type *V. fluvialis* gene and tested with the substrate analog, ethyl 3-oxohexanoate, for activity. These mutations included: 59N, 60T, 84S, 85L, 100M, 112Y, 113F, 122E, 123S, 124V, 125I, 144I, 149S, 153A, 155V, 156A, 156M, 157A, 157T, 158S, 159A, 160L, 162A, 164D, 164A, 166V, 177I, 179A, 179V, 220M, 226Q, 226L, 232L, 232F, 232M, 254V, 255A, 260T, 268L, 268M, 280I, 280G, 282V, 283A ,286S, 286A, 287M, 287I, 288S, 291Y, 294V, 298M, 298L, 300S, 301A, 301K, 302D, 302A, 303I, 303F, 304A, 304D, 304T, 323S, 324A, 326N, 360M, 386E, 414I and 424A.

ProteinGPS™

The creation of an alignment of 56 homologs, with as little as 25% sequence identity, was also performed. The non-conserved sites present in this set of homologs were scored using structural, phylogenetic and substitution information as previously described and called ProteinGPS™ (Ehren *et al.*, 2008). The initial round of variants was designed using the top 59 scored mutations and the W57F mutation identified above. The mutations tested included: M22L, S24A, S24T, L25V, Q27E, R28K, I69V, D70E, M95L, Q96E, S108A, Q139E, V156A, S157A, M160L, S167A, V177I, L179V, T180A, T180D, T180G, Y187E, Y187F, E189Q, E189R, T180S, E190A, E190D, E190P, T193S, E195A, E207A, E207D, Q210A, Q210L, R211K, M226L, I245V, I254V, S255A, I259V, I282V, M294L, S304A, L307I, L307V, E308A, E308Q, T309A, I311V, I314V, I332V, M341L, K361R, E365D, I375V, K385R and N396D. Ninety-five variants were designed, based on the wild-type *V. fluvialis* sequence, each containing 3–5 mutations. Each of the 60 selected mutations was seen in the set of variants six times. These 95 round 1 variants were created through DNA synthesis (DNA 2.0, Inc., Menlo Park, CA, USA) and then transformed as noted. Owing to the low yields and detection abilities for the desired substrate, the first-round plate of over-expression cultures was screened with both the smaller model substrate, ethyl 3-oxohexanoate, and desired substrate **1**.

Three additional subsequent rounds of variants were designed using the best mutations identified from the prior round and any positive mutations identified from the other methods. Substitutions were assessed using a variety of multivariate regression and machine learning techniques described previously (Minshull *et al.*, 2004; Liao *et al.*, 2007).

Protein purification, crystallization and data collection

One liter of *E. coli* cell paste containing the overexpressed selenomethionine (Se-Met) incorporated Vfat recombinant protein (Se-Met Vfat) was resuspended in cold Buffer A (25 mM Tris-HCl pH 7.5, 5% (v/v) glycerol, 300 mM NaCl, 1 mM TCEP and 10 mM imidazole) at a volume of three times the wet weight. Additionally, one Complete™ ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor tablet (Hoffmann-La Roche, Basel, Switzerland) and 1 μl of benzonase nuclease (Sigma E-1014, Sigma-Aldrich, St Louis, MO, USA) was added per 50 ml of Buffer A. Cells were lysed with two passes in a microfluidizer (pressure at 18 000 psi). The lysate was clarified by centrifugation at 4°C for 30 min at 12 000 rpm in a Sorvall SS-34 rotor (Thermo Scientific, Waltham, MA, USA). The supernatant was extracted and filtered through a 0.22-μm low protein binding membrane. It was then applied via a superloop to a HisTrap FF crude pre-packed column (2 × 5 ml, GE Healthcare Lifesciences, Pittsburgh, PA, USA) that was pre-equilibrated in Buffer A and connected to an Äkta Purifier. The column was washed with 5–10 column volumes (CV) of Buffer A and then eluted with a steep gradient (0–500 mM imidazole in Buffer A over three CV). After pooling the peak fractions from the Ni column, the protein was dialyzed into Buffer B (25 mM Tris pH 7.5, 5% (v/v) glycerol, 20 mM NaCl and 1 mM TCEP) for ion exchange chromatography. The dialyzed sample was applied via superloop to a Q XL 16/10 column (GE Healthcare LifeSciences) that was pre-equilibrated in Buffer B. After washing with 5–10 CV of Buffer B, the protein was eluted with a gradient of 0–1000 mM NaCl in Buffer B over 20 CV. The resultant protein-containing peak fractions were pooled and concentrated to 10 mg/ml using an Amicon filtration unit (Ultrafree-15 membrane with 10K MWCO, Millipore, Billerica, MA, USA) for size-exclusion chromatography. A HiLoad Superdex 16/60 Superdex 200 prep grade column (GE Healthcare LifeSciences) were equilibrated in Buffer C (25 mM Tris pH 7.5, 200 mM NaCl and 1 mM TCEP). The protein sample was applied via superloop, and the peak fractions were pooled and further concentrated for crystallization trials. The r414 mutant protein was purified in the same manner.

Crystallization of Se-Met Vfat was achieved using a 14 mg/ml protein stock with 2 mM O-(carboxymethyl)hydroxylamine hemihydrochloride (OCMM, Fluka/Sigma-Aldrich) and 2 mM PLP (Sigma-Aldrich). Both the OCMM and PLP solutions were titrated to pH 7. Crystals were grown using sitting drop vapor diffusion in a 2-drop Medical Research Council plate (Swisssci, Zug Switzerland). 0.5 + 0.4 μl drops were set up with an 80 μl well using a Mosquito (TTP Labtech, Cambridge, MA, USA). 0.1 μl of 20% (w/v) benzamidine-HCl (Hampton Research, Aliso Viejo, CA, USA) was added to each drop. The well consisted of 20% (w/v) polyethylene glycol (PEG) 8000, 0.2 M lithium sulfate and 0.1 M bis–tris pH 6.5. Plates were incubated at 22°C, and crystals grew within 1–2 days. One round of streak seeding was needed to obtain single, diffraction quality crystals. The resultant crystals were found to be pyridoxamine-5′-phosphate-bound apo (PMP-bound apo). The r414 mutant protein was crystallized under the same condition as Se-Met Vfat and was soaked for several hours in a solution free of primary amines (23% PEG 8000, 0.2 M lithium sulfate and 0.1 M MES pH 6.5). A subsequent soak in the amine-free solution with 1 mM PLP produced a complex without catalytic turnover.

Structure determination

Diffraction data were collected from flash-frozen crystals at 100 K at beamline 17-ID at Advanced Photon Source of the Argonne National Laboratory. Data were processed using the HKL2000 suite of software (Otwinowski and Minor, 1997). Data collection statistics are summarized in Table I. The structure of PMP-bound Se-Met incorporated Vfat was solved by molecular replacement methods with the CCP4 version of PHASER (McCoy *et al.*, 2007), using a search model derived from the dialkylglycine decarboxylase structure (PDB code 1DGE). After molecular replacement, maximum likelihood-based refinement of the atomic position

Table I. Data collection and refinement statistics

Complex	PMP-bound Se-Met Vfat	PLP-bound R414 mutant Vfat
Data collection		
PDB code	4E3Q	4E3R
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å)	$a = 63.1$	$a = 63.0$
	$b = 162.2$	$b = 161.9$
	$c = 180.4$	$c = 179.3$
Resolution (Å)	1.9	1.9
Unique observations	144 735	143 186
Completeness (%)[a]	98.9 (98.3)	98.8 (92.5)
Redundancy[a]	5.9 (5.1)	6.3 (4.1)
Rmerge[a]	0.08 (0.62)	0.06 (0.41)
$I/\sigma I$[a]	21.8 (2.5)	18.9 (3.0)
Refinement		
Resolution (Å)	50–1.9	30–1.9
Reflections (Rfree)	144 629 (7258)	143 087 (7175)
Rwork/Rfree (%)	17.4 (21.0)	15.7 (19.0)
Average B-factor (Å^2)	34.4	27.8
RMSD bonds (Å)	0.010	0.010
RMSD angle (°)	1.04	1.01
Ramachandran plot (%)		
Most favorable region (outlier)	89.6 (0.3)	90.8 (0.3)

[a]Values in brackets represent statistics for highest-resolution shells.

and temperature factors were performed with a REFMAC/Buster (Murshudov *et al.*, 1997; Blanc *et al.*, 2004) and the atomic model was built with the program COOT (Emsley and Cowtan, 2004). The refined PMP-bound Vfat structure was then used as a starting model for PLP-bound Vfat structure. The stereochemical quality of the final model was assessed by the PROCHECK (Laskowski *et al.*, 1993). Crystallographic statistics for the final models are shown in Table I. Figures were prepared with PYMOL (DeLano Scientific, Schrödinger Inc., NewYork, NY, USA; http://www.pymol.org).

Molecular modeling and analysis

Molecular modeling for this project was performed using structures derived from the A and B chains (that is, one dimer) of the wild-type and r414 Vfat crystal structures. These model structures were initially created by extracting the A and B chains from both the wild-type and the r414 Vfat crystal structures, along with their associated water molecules. The r414 model structure was then superposed onto the wild-type model structure employing a proprietary algorithm based on the alignment of secondary structure elements (minor variances in the superposition would not have significant impact on the subsequent analysis, as it was largely done for convenience). The PMP of the wild-type model structure was converted into PLP, using the coordinates of the superimposed r414 PLP as a guide. Hydrogens were added to both model structures using a proprietary application that optimizes hydrogen-bonding networks using a self-consistent mean field optimization method. For the wild-type model structure, the Se-Mets were manually converted to methionines, as most commercial molecular mechanics packages lack the requisite parameters for the selenium atoms.

An approximate docking of the substrate into the wild-type model was performed manually, based on the placement of the reactive ketone adjacent to the modeled PLP, in accordance with the desired reaction stereochemistry. This pose was refined through iterative minimization of the model–substrate complex until a suitable model with no close contacts and no obviously significantly strained protein conformations was achieved. Minimization was performed using the Schrödinger's BatchMin application (Schrödinger Inc.), with the AMBER force field, GB/SA solvation and PRCG minimizer. The modeled substrate, and all protein atoms up to 8 Å (residue-based) from the modeled substrate were allowed to move freely during minimization; all other atoms were held in their crystallographically determined positions. An analogous procedure was used for the r414 model.

Biophysical characterization of the variants

The enzymes were purified by immobilized metal ion affinity chromatography. Five to ten grams of cell pellet with overexpressed enzyme, prepared as noted above, were resuspended in 50 ml of phosphate-buffered saline (PBS) with one Complete™ EDTA-free protease inhibitor tablet (Hoffmann-La Roche) and 1 mM PLP. Cells were kept on ice and then lysed with the MP Bio Fast Prep (MP Biomedicals, Solon, OH, USA) using the glass beads B for 6 m/s for 40 s for two sets with a 15 min ice bath between the sets. The lysate was then centrifuged at 30 000 × *g* for 30 min. The supernatant was poured off and diluted with 50 ml of PBS with 1 mM PLP solution and filtered through a 0.2 μm stericup filter (Millipore). A column with 3 ml of Ni-NTA superflow resin (Qiagen, Valencia, CA, USA) was prepared and 20 ml of supernatant was added to the column and allowed to flow by gravity. The column was washed with 20 ml of PBS with 1 mM PLP and 30 mM imidazole and then eluted with three times 3 ml of PBS with 1 mM PLP and 300 mM imidazole. Two and a half milliliters of the third elute fraction were put through a PD-10 gel filtration column (GE Healthcare LifeSciences), which had been equilibrated with PBS with 0.1 mM PLP, and was eluted in a final volume of 3 ml PBS with 0.1 mM PLP. This sample was used in kinetics studies and the concentration was determined by ultraviolet-visible spectroscopy at 280 nm, using a molar absorptivity (ϵ) of 61 000 l mol^{-1} cm^{-1} identified through the ProtParam tool in ExPASy (Gasteiger *et al.*, 2005). A standard sodium dodecyl sulfate-polyacrylamide gel electrophoresis gel showed predominantly a single band of the correct molecular weight protein. The kinetic constants were determined through the enzyme assays noted below.

Enzyme assays

Enzyme activity assays were all performed in duplicate on sodium pyruvate (10 mM), or other noted substrates (0–10 mM), at 30°C in potassium phosphate buffer (100 mM, pH 7.0) containing S-MBA (50 mM), PLP (2 mM) and purified enzyme solution (72–74 μg/ml). Reactions were quenched with acetonitrile and analyzed by UPLC for acetophenone. Enzyme activity assays on ethyl 3-oxohexanoate and substrate **1** were performed under similar conditions to those described for sodium pyruvate except for the addition of dimethyl sulfoxide (DMSO, 15% v/v) to alleviate the low aqueous solubility of these substrates. Kinetic parameters were determined for the transamination of sodium pyruvate with S-MBA for wild-type and W57F enzymes. Assay mixtures containing sodium pyruvate (0.2–10 mM), S-MBA

(50 mM), PLP (2 mM) and enzyme solution (72–74 μg/ml) in the potassium phosphate buffer (100 mM, pH 7.0) were incubated at 30°C for 4 min and quenched with acetonitrile. The quenched samples were analyzed for acetophenone by UPLC. Similar conditions were used to determine kinetic parameters for the transamination of ethyl 3-oxohexanoate and **1** with r414 except that reactions also contained DMSO (15% v/v) and were quenched after 2 h. Data were subjected to non-linear regression analysis (GraphPad Prism 5.0, GraphPad Software Inc., La Jolla, CA, USA) to obtain kinetic parameters. UPLC separations were carried out on a Waters BEH C18 column (2.1 × 50 mm, 1.7 μ) on a Waters Acquity UPLC® system equipped with an Aquity PDA detector (Waters Corp.). The column was eluted with 1% trifluoroacetic acid in water (v/v) : acetonitrile (75 : 25) at 0.8 ml/min with max plot detection (210–400 nm).

Scale-up methods

To a jacketed reaction vessel maintained at 30°C was added 4.0 g of wet *E. coli* cells containing Vfat mutant r414, 88 ml of the potassium phosphate buffer (0.1 M, pH 7.0) and 54 mg of pyridoxal phosphate. The contents of the reaction vessel were stirred for ~5 min after which substrate **1** (1.0 g) and S-MBA (7.5 ml of 1 M solution) were charged. The reaction mixture was stirred at 30°C (speed 12, 842 Titrando titrator, Metrohm, Riverview, FL, USA). Reaction samples (50 μl) were withdrawn over the course of the reaction, diluted with 450 μl of acetonitrile, derivatized with Marfey's reagent and analyzed by high-performance liquid chromatography on an Agilent 1100 system (Agilent Technologies).

Work-up: After 24 h, the reaction mixture pH was adjusted to 6.0 using 1 N HCl, centrifuged (8000 rpm × 20 min) and the supernatant was collected. The pellet was washed with water (10 ml), and again centrifuged (8000 rpm × 20 min) to collect the supernatant. The supernatants from both centrifugations were combined and extracted with *tert*-butyl methyl ether (TBME) (2 × 75 ml). Concentration of this TBME extract under vacuum gave 475 mg of a light-yellow oil. The aqueous fraction was then adjusted to pH 12 with 4 N aqueous sodium hydroxide and extracted with TBME (2 × 75 ml). Concentration of this extract gave 693 mg of a colorless liquid. The proton nuclear magnetic resonance analysis of the crude product (693 mg) showed 40% potency for **2**, which corresponds to 277 mg of **2** (28% yield).

Protein data bank accession codes

The coordinates and structure factors have been deposited with the Protein Data Bank with accession codes, 4E3Q and 4E3R.

Results

Three related enzymes initially indentified with very low desired activity

Preliminary feasibility for enzymatic transamination of substrate **1** was determined by screening transaminases in reverse direction, i.e. for resolution of (5*R*)-ethyl 3-amino-5-methyloctanoate, a diastereomeric mixture of **2**. This screen identified three enzymes, Vfat, *R. sphaeroides* and *P. denitrificans* aminotransferases, that demonstrated desired selectivities by giving product mixtures enriched in (3*R*,5*R*)-ethyl 3-amino-5-methyloctanoate, the undesired diastereomer of **2**. The activities of these enzymes, however, were very low, and preliminary attempts to prepare intermediate **2** by transamination of substrate **1** with these enzymes were unsuccessful, as might be expected from the other literature work (Hwang and Kim, 2004). Although the three enzyme hits showed similar activities in the preliminary resolution screen, Vfat was selected as the starting enzyme for optimization because there was slightly more literature information available (Cho *et al.*, 2008).

The engineering comprised two components. The first of these was the identification of mutations that were potentially beneficial using a variety of sources. The second was the creation of sets of variants containing various combinations of the mutations identified in the first component, measuring of their activity and machine learning analysis to determine the changes that contributed most to the desired activity. These processes were interdependent: new potentially beneficial mutations could be identified at any point during the engineering process and incorporated into the next set of variants, and the machine learning determined which of these changes should be maintained and which should be eliminated.

Potentially beneficial mutations were identified

We first created all possible amino acid substitutions at predicted active site residues 21, 57, 147, 233, 297 and 415 in the wild-type Vfat protein (Cho *et al.*, 2008). Since the low activity of Vfat for transamination of substrate **1** made detection of improved variants difficult, a substrate analog, ethyl 3-oxohexanoate, was also used for initial activity studies as transamination of this substrate was readily detected. This work showed that a mutation at W57F was a very beneficial mutation for activity, providing a 4-fold activity increase over wild-type Vfat for transamination of ethyl 3-oxohexanoate (Fig. 2a). They also suggested that R415F and D21Y were beneficial, each providing ~2-fold improvements in activity.

The second mutation identification strategy employed a basic bioinformatics approach to identify 55 non-conserved sites, with a total of 70 mutations, within an area of ~15 Å of the active site. From this set of single mutant variants, the following mutations were seen to give improvements of 1.5- to 3-fold in activity for substrate **1**: A84S, V153A, T288S, M294V and A323S.

A third mutation identification strategy, using the ProteinGPS™ method, involved the identification of 60 substitutions combined into a set of 95 variants each with 3–5 mutations. Reactions with most of the variants showed no detectable conversions after 4 h for the substrate **1** and very low conversions after 24 h. However, activity for the model substrate, ethyl 3-oxohexanoate, could be measured at both time points. We employed ProteinGPS™ statistical analysis of the data (Minshull *et al.*, 2004; Liao *et al.*, 2007) to identify the subset of mutations, which appeared to improve activity for both substrates, and a few additional mutations which appeared to only improve activity for ethyl 3-oxohexanoate or substrate **1**, but not both. This analysis was used to select 37 mutations that were beneficial or neutral for conversion of the model substrate, ethyl 3-oxohexanoate and 35 mutations that were beneficial or neutral for conversion of substrate **1**. From these data, two complementary second round sets, each of 35 combinatorial variants were created.

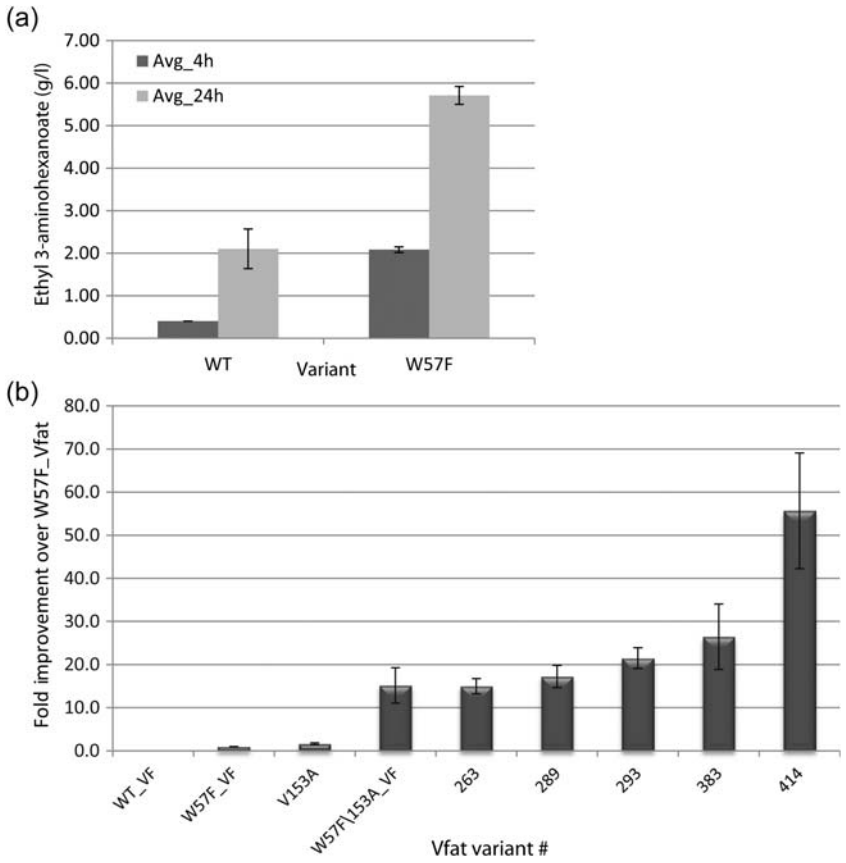

Fig. 2. Fold improvement in enzyme performance for Vfat variants. (**a**) Comparison of wild-type Vfat with W57F point mutant of Vfat (W57F) on model substrate (ethyl 3-oxohexanoate) at 4 h (dark bars) and 24 h (light bars). (**b**) Fold improvement in enzyme performance for the transamination of **1** to **2** with Vfat variants. Variants with numbers are defined as Vfat with the following mutations: #263: 57F, 153A and 163L; #289: 57F, 153A, 163L and 415F; #293: 57F, 84S, 153A, 163L and 415F; #383: 57F, 84S, 153A, 163L, 259V and 415F; #414: 19W, 57F, 85A, 88K, 153A, 163F, 259V and 415F.

The frequency with which a substitution was incorporated was calculated based on the confidence in its contribution to activity (Liao *et al.*, 2007) and each substitution was observed 3–20 times in the variant set. In addition, each set contained four additional mutations, M294V, R415F, D21Y and K163L, identified as described in the previous sections. These mutations were each incorporated three times in each set of 35 variants. These 70 s round variants were screened against substrate **1** at a 5 h time point. Analysis of the results suggested that several mutations improved activity. The specific mutations noted at this point were I69V, V177I and I259V.

In parallel with the sequence bioinformatics work, two homology structure models were created and efforts were undertaken to crystallize the wild-type Vfat enzyme. There were no crystal structures of enzymes with >30% sequence identity to the Vfat sequence available in the PDB database (Berman *et al.*, 2000). Two structures with 26% (PDB ID: 1D7R) and 30% (PDB ID: 2EO5) identity and 7% gaps were used for the homology modeling. The homology work identified 15 mutation sites that surround the proposed substrate-binding site for mutational studies. The sites were: T20, P23, L25, L56, Y150, K163, Y165, N166, S167, A228, G229, P233, R415, P416 and L417. Substitutions for these sites were chosen based on the ProteinGPS^{TM} bioinformatics work.

A third round of 95 variants were created based on the round 2 data, substitutions included 36 mutations identified from round 2, and 6 additional sites identified as described in the previous section (F19W, A84S, V153A, A228G, T288S and A323S). In addition, 45 new mutations were identified from further analysis of the ProteinGPS^{TM} alignment of the same 56 homologs used earlier; these new sites were also added into the variant design. The results of the screening of this set of variants identified 14 hotspots and indicated further work to recombine these into the best combinations. The 14 hotspot mutations identified were: F19W, D21Y, W57F, I69V, A84S, V153A, K163L, V177I, A228G, I259V, T288S, M294V, A323S and R415F. A fourth round of variants was created to study combinations of these and identified the best variant as r414 with the following mutations: F19W, W57F, F85A, R88K, V153A, K163F, I259V and R415F.

Structural determination of the wild-type and mutant variants

The crystallization work provided a structure of the Se-Met Vfat wild-type enzyme with a bound PMP molecule at the active site cleft (Fig. 3 and Table I). The Vfat is biologically active as a dimer with overall dimensions of $\sim 56 \times 88 \times 97$ Å. The two PMP cofactors are located in the subunit interface of the dimer, and are ~ 15 Å apart. This wild-type structure highlighted 11 amino acid sites that would most likely be in contact with the substrate, including F19, L56, W57, F85, F86, V153, E223, A228, I259, R415 and L417. No substrate or inhibitor was visible in the structure.

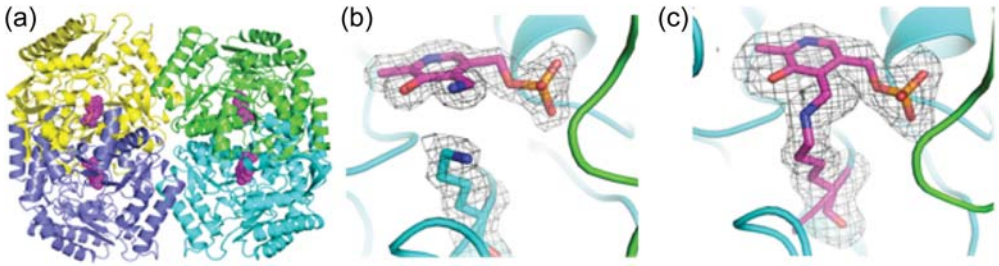

Fig. 3. Crystal structure of Vfat and r414 mutant Vfat. (**a**) Overall structure of Vfat tetramer with the PMP cofactors shown as spheres in magenta. (**b**) The (Fo−Fc) omit map contoured at 3.0 σ for Lys285 and the PMP cofactor in the active site of Se-Met incorporated Vfat structure. (**c**) The (Fo−Fc) omit map contoured at 3.0 σ for Lys285 and the PLP cofactor in the active site of Se-Met incorporated Vfat structure

As positive mutations were identified, they were combined into the current best variant and entered into the ProteinGPS^{TM} analysis. The variant progression shown in Fig. 2b shows that the successive addition of positive mutations improved both selectivity and yield. However, not all identified positive mutations were synergistic, as some combinations of positive mutations were found to decrease activity.

Improved variants were analyzed

Individual mutations identified using the various described methods were combined to yield multi-mutation variants. The progression of improvement is shown in Fig. 2b. The final improved variant, r414, contained eight mutations (F19W/W57F/F85A/R88K/V153A/K163F/I259V/R415F) and had a 60-fold improvement in initial activity over wild-type Vfat enzyme. Although the activity of wild-type Vfat for **1** was difficult to measure due to low activity, conversions, as determined by measurements of the by-product acetophenone, and **2**, after 24 h with wild-type and r414 were in agreement. In addition, the selectivity of the reaction with variant r414 was quite high and showed a diastereomeric excess (de) of 95%. There was one variant (variant r409 containing substitutions W57F, F85A, V153A, K163L, I259V and R415F) which had higher activity than r414 but lower selectivity (92% de) and was therefore not studied further. In addition, the mutation at position A228G appeared to contribute to the decrease in selectivity for transamination of **1** with the more active variant.

The best variant, r414, was also studied by crystallography and a structure has been determined at 1.9Å resolution (Fig. 3 and Table I). Importantly, the r414 variant Vfat enzyme crystals, which were incubated in a solution free of primary amines, shows the covalent modification of PLP to the ε-amino group of Lys285 forming the internal aldimine (Schiff base) linkage.

Although we do not have any substrate-bound structures, it is clear that several of the mutations near the binding pocket are providing more room and a better fit for **1** (Fig. 4). Substrate modeling experiments for both the wild-type and r414 variant structures reveal putative substrate binding modes, and provide strong structure-based reasoning for the benefit of several key mutations. The W57F and I259V mutations appear to allow more room for the ethyl group of the ester to fit better, with less protein strain required to fit the substrate into the binding site (Fig. 4a). In particular, the modeled ethyl group in the r414 structure seems to occupy the same space as the removed CD1 atom in the wild-type

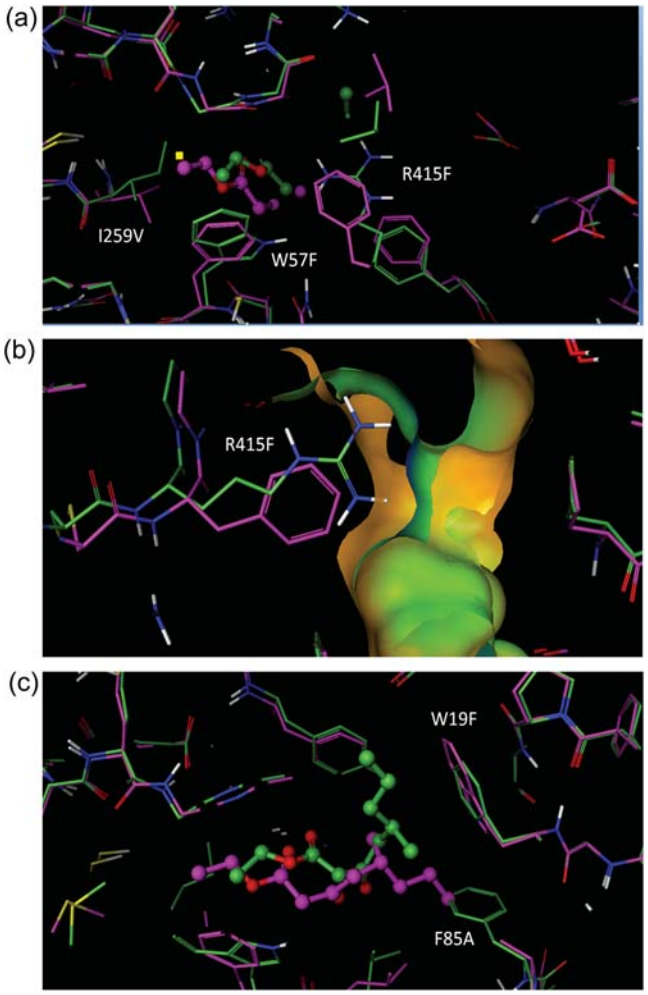

Fig. 4. Close-up of the modeled substrates (ball and stick models) in the wild-type (green, wireframe) and r414 (purple, wireframe) enzymes. (**a**) The I259V and W57F mutations allow significantly more room for the ethyl tail of the substrate to sit deeply into the non-polar pocket. The original position of the wild-type ILE-259:CD1 atom is noted with a yellow square. (**b**) Close-up of the modeled positions of the wild-type ARG-415 (green, wireframe) and r414 PHE-415 residues (purple, wireframe), with associated Connelly surfaces colored by hydrophobicity (brown is hydrophobic, green/blue is hydrophilic). The r414 binding site is considerably larger and more hydrophobic as a result of this mutation. (**c**) The wild-type PHE-85 forced the substrate into a bent conformation, while the r414 ALA-85 allows the substrate to exist in an extended conformation that occupies the location of the wild-type PHE side chain.

Ile. With regard to the R415F mutation, Arg did not appear to be involved in any notable polar interactions with the modeled substrate, and was thus likely incurring a substantial

desolvation penalty upon ligand binding. The mutation to Phe makes this part of the binding pocket much less polar (saving the desolvation penalty), and somewhat larger, better accommodating the desired substrate (Fig. 4b). Finally, it appears that F85A opens up a significant side pocket relative to the wild-type structure, allowing a much more extended (less strained, more favorable) conformation of the modeled substrate in the r414 structure relative to the wild-type structure (Fig. 4c). The remaining three mutations (R88K, V153A and K163F) do not interact significantly with the substrate, and it is likely that their effects are more subtle, for example allowing easier protein conformational changes to bind the substrate.

Kinetic analysis of improved variant showed substantial substrate specificity changes

Wild-type Vfat and variants W57F and r414 were purified in order to obtain kinetic constants (Table II). As noted earlier, the wild-type enzyme had very limited activity for the desired substrate. This meant that accurate kinetic constants could not be determined for **1** with the wild-type enzyme, so kinetic constants are shown for the substrate sodium pyruvate, which was a good substrate for the wild-type Vfat. Interestingly, the r414 variant did not have measurable activity for pyruvate. Thus, enzyme engineering of Vfat has optimized activity for **1**, but resulted in loss of activity for pyruvate. It is also notable that the r414 variant has a much higher K_m value for the model substrate ethyl 3-oxohexanoate than **1**; however, it also has a higher k_{cat} for this substrate leaving an ~2-fold difference in k_{cat}/K_m for these susbtrates.

Selection of conditions related to process considerations

Due to the low activity of Vfat, extensive chemical process development was not undertaken in the early stages of this project. However, selection of an amino donor for use in screening reactions, and ultimately for process use, was considered a critical requirement for enzyme engineering as improvements would likely be specific for ketone **1** and the amino donor. On the basis of reported substrate scope of Vfat and process criteria, including cost and ease of removal, S-MBA, isopropylamine and *sec*-butylamine were selected as potentially suitable amino donors. Studies on the transamination of ethylacetoacetate and ethyl 3-oxohexanoate with S-MBA, isopropylamine and *sec*-butylamine using Vfat indicated that all three amines were potentially suitable amino donors, and S-MBA was selected based on higher reactivity compared to the aliphatic amines. The removal of S-MBA from the transamination reaction was a potential concern. This concern was alleviated by demonstrating that **2** could be hydrolyzed to **3** and isolated in suitable purity as its hydrochloride salt in the presence of S-MBA. Reaction conditions used for screening transamination of **1** with S-MBA were optimized for temperature, pH and PLP concentration.

Discussion

Using a multi-pronged enzyme engineering approach, Vfat was engineered for transamination of **1** to **2**. The initial velocity of the wild-type enzyme was improved 60-fold. The wild-type enzyme had very low initial activity and was not useful in preparative reactions. The resulting variant (r414) was used for the preparation of **2** in 95% de, although further work will be necessary to make this enzyme commercially useful. This work has also opened this class of enzyme to further engineering through the solution of the high-resolution crystal structures of both wild-type Vfat and the best variant r414.

Overall, this work involved the creation and screening of <450 variants to obtain a greatly improved variant for the desired reaction. This was especially important because the assay used required derivatization of the product followed by UPLC and was not amenable to high-throughput screening.

The structure shows that the eight mutations in the variant are all near the binding pocket. The binding pocket is created near the interface of the two dimers in the structure. Five of the mutations are clearly providing additional space and interactions with the desired substrate. The structures also highlight several mutations noted in earlier homology structure work (Cho *et al.*, 2008). The residues at positions 21, 57 and 415 were found to be in the active site in our crystal structures; however, position 147 was neither found in, nor pointing towards, the active site cavity. In addition, mutations at position 147 were not found to improve activity for **1**, while mutations at the other noted positions did provide enhancements.

Often a single method of enzyme engineering is used to improve an enzyme. However, we found here that none of the methods quickly illuminated all eight sites in r414, and the complementary nature of the methods was beneficial. In addition, these methods kept the number of variants, which had to be created or synthesized, low, allowing for a detailed analytical analysis for screening for improvement.

Stereoselectivity is a critical requirement for the desired reaction and, therefore, enzymes with improved activity must also display high selectivity. Several mutations were

Table II. Kinetic results

Kinetic parameters of wild-type Vfat and variants

Enzyme	Substrate[a]	K_m (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	K_{cat}/K_m ($mM^{-1}\ s^{-1}$)
Wild-type Vfat	Sodium pyruvate	3.98 ± 0.05	19.95 ± 0.34	16.67 ± 0.28	4.19
W57F	Sodium pyruvate	6.87 ± 0.004	18.28 ± 0.06	15.27 ± 0.05	2.23
r414	Ethyl 3-oxohexanoate	8.59 ± 0.18	0.134 ± 0.001	0.112 ± 0.001	0.013
	1	0.83 ± 0.01	0.0226 ± 0.0002	0.0189 ± 0.0002	0.023

All reactions are carried out in duplicates. Average values with standard deviation. are shown.
[a]Reactions with ethyl 3-oxohexanoate and **1** carried out with 15% DMSO (v/v).

identified which improved activity for transamination of **1** and also greatly affected the selectivity of the reaction. A mutation at position 228 provided an increase in activity but significantly lowered the de value.

The results obtained here show that the omega-aminotransferase class of enzymes is amenable to enzyme engineering for the synthesis of β-aminoesters from the corresponding β-ketoesters. Further engineering has been enhanced by the discovery of important residues in this study and through the solution of high-resolution crystal structures of wild-type Vfat and the improved variant r414. This will allow further computational studies of the binding site for other desirable substrates.

Author contributions

K.S.M., R.K., S.H., J.M., S.G. and J.W.W. designed and analyzed the experiments. M.J.K. and K.S.M. performed molecular biology, cell culture and protein purification, K.M. and D.K.G. performed homology modeling, substrate docking and structural analysis. A.M., J.S.C., M.A. and S.G. performed crystallography and structural analysis. A.V., J.M., and S.G. performed ProteinGPS^{TM} methods and data analysis. J.W.W. and R.K. performed the enzyme assays, scale-up reactions and kinetic biophysical characterizations. K.S.M., S.H., R.K., J.M., S.G., J.W.W. and D.K.G. wrote and edited the manuscript.

Acknowledgements

We thank V.M., M.B., J.S. and N.W. for helpful discussions on this work.

References

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402. First published on 01 September 1997.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242. First published on 11 December 1999.

Birch,M., Challenger,S., Crochard,J.-P., *et al*. (2011a) *Org. Process Res. Dev.*, **15**, 1358–1364.

Birch,M., Challenger,S., Crochard,J.-P., *et al*. (2011b) *Org. Process Res. Dev.*, **15**, 1172–1177.

Blanc,E., Roversi,P., Vonrhein,C., Flensburg,C., Lea,S.M. and Bricogne,G. (2004) *Acta Crystallogr.*, **60**, 2210–2221.

Chenna,R., Sugawara,H., Koike,T., Lopez,R., Gibson,T.J., Higgins,D.G. and Thompson,J.D. (2003) *Nucleic Acids Res.*, **31**, 3497–3500. First published on 26 June 2003.

Cho,B.K., Park,H.Y., Seo,J.H., Kim,J., Kang,T.J., Lee,B.S. and Kim,B.G. (2008) *Biotechnol. Bioeng.*, **99**, 275–284. First published on 08 August 2007.

Ehren,J., Govindarajan,S., Moron,B., Minshull,J. and Khosla,C. (2008) *Protein Eng. Des. Sel.*, **21**, 699–707. First published on 07 October 2008.

Emsley,P. and Cowtan,K. (2004) *Acta Crystallogr.*, **60**, 2126–2132.

Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) In Walker,J.M. (ed), *The Proteomics Protocols Handbook*. New York, NY: Humana Press, pp. 571–607.

Hwang,B.-Y. and Kim,B.-G. (2004) *Enzyme Microb. Technol.*, **34**, 429–436.

Koszelewski,D., Tauber,K., Faber,K. and Kroutil,W. (2010) *Trends Biotechnol.*, **28**, 324–332. First published on 01 May 2010.

Laskowski,R.A., Moss,D.S. and Thornton,J.M. (1993) *J. Mol. Biol.*, **231**, 1049–1067.

Liao,J., Warmuth,M.K., Govindarajan,S., Ness,J.E., Wang,R.P., Gustafsson,C. and Minshull,J. (2007) *BMC Biotechnol.*, **7**, 16.

McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) *J. Appl. Crystallogr.*, **40**, 658–674.

Minshull,J., Govindarajan,S., Cox,T., Ness,J.E. and Gustafsson,C. (2004) *Methods (San Diego, Calif*, **32**, 416–427.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) *Acta Crystallogr.*, **53**, 240–255.

Otwinowski,Z. and Minor,W. (1997) *Methods Enzymol.*, **276**, 307–326.

Savile,C.K., Janey,J.M., Mundorff,E.C., *et al*. (2010) *Science (New York, NY*, **329**, 305–309. First published on 19 June 2010.