Crystal structures of two monomeric triosephosphate isomerase variants identified via a directed-evolution protocol selecting for L-arabinose isomerase activity

Mirja Krause,^a Tiila-Riikka Kiema,^b Peter Neubauer^a and Rik K. Wierenga^{b*}

Received 1 March 2016

Accepted 5 May 2016

Edited by W. N. Hunter, University of Dundee, Scotland

Keywords: enzyme engineering; non-natural enzymes; triosephosphate isomerase; structure-based rational design; TIM barrel.

PDB references: Ma21 TIM, 4pc8; Ma18 TIM, 4pcf

Supporting information: this article has supporting information at journals.iucr.org/f

^aLaboratory of Bioprocess Engineering, Department of Biotechnology, Technische Universität Berlin, Ackerstrasse 76, ACK 24, Berlin, Germany, and ^bFaculty of Biochemistry and Molecular Medicine, Biocenter Oulu, University of Oulu, FIN-90014 Oulu, Finland. *Correspondence e-mail: rik.wierenga@oulu.fi

The crystal structures are described of two variants of A-TIM: Ma18 (2.7 Å resolution) and Ma21 (1.55 Å resolution). A-TIM is a monomeric loop-deletion variant of triosephosphate isomerase (TIM) which has lost the TIM catalytic properties. Ma18 and Ma21 were identified after extensive directed-evolution selection experiments using an *Escherichia coli* L-arabinose isomerase knockout strain expressing a randomly mutated A-TIM gene. These variants facilitate better growth of the *Escherichia coli* selection strain in medium supplemented with 40 mM L-arabinose. Ma18 and Ma21 differ from A-TIM by four and one point mutations, respectively. Ma18 and Ma21 are more stable proteins than A-TIM, as judged from CD melting experiments. Like A-TIM, both proteins are monomeric in solution. In the Ma18 crystal structure loop 6 is open and in the Ma21 crystal structure loop 6 is closed, being stabilized by a bound glycolate molecule. The crystal structures show only small differences in the active site compared with A-TIM. In the case of Ma21 it is observed that the point mutation (Q65L) contributes to small structural rearrangements near Asn11 of loop 1, which correlate with different ligand-binding properties such as a loss of citrate binding in the active site. The Ma21 structure also shows that its Leu65 side chain is involved in van der Waals interactions with neighbouring hydrophobic side-chain moieties, correlating with its increased stability. The experimental data suggest that the increased stability and solubility properties of Ma21 and Ma18 compared with A-TIM cause better growth of the selection strain when coexpressing Ma21 and Ma18 instead of A-TIM.

1. Introduction

Triosephosphate isomerase (TIM; EC 5.3.1.1) is a dimeric enzyme consisting of two identical subunits that adopt the TIM-barrel fold (Fig. 1). This enzyme is an aldose–ketose sugar isomerase that catalyzes the transfer of a proton between adjacent C atoms of the substrate dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate (D-GAP) via an enediolate intermediate (Richard *et al.*, 2014; Supplementary Fig. S1). This reaction is an important step in the glycolytic pathway. The classical studies of Knowles (1991) highlighted that TIM has evolved towards becoming a perfect enzyme in which the rate-limiting step is the diffusion-limited encounter of substrate and enzyme. TIM is also characterized by having a very high substrate specificity. The only known physiological substrates are D-GAP and DHAP (Wierenga *et al.*, 2010). The active site of TIM is formed by the ‘front’ loops (referred to as loop 1 to loop 8; Fig. 1 and Supplementary Fig. S2) of the classical TIM-barrel fold, which connect the β -strands to the subsequent α -helices. The ‘back’ loops

connect the α -helices to the subsequent β -strands. The front loops are important for the function, as they shape the active site, whereas the back loops are important for protein stability and folding (Kursula *et al.*, 2002; Silverman *et al.*, 2001). The

Figure 1
Ribbon drawing of the wild-type TIM dimer showing the active-site residues located in loop 1 (Asn11 and Lys13), in loop 4 (His95) and in loop 6 (Glu167). The shortened dimer-interface loops, loop 1, loop 3 and loop 8, are shown in red, orange and green, respectively. Loop 7 is shown in yellow.

Figure 2
The anchoring interactions of Asn11 OD2 in wild-type TIM (cyan; PDB entry 1n55; complexed with PGA in grey) by hydrogen bonds from Asn11 OD2 to Trp12 N and Lys13 N. In the wild-type TIM active site Asn11 ND2 functions as an oxyanion-hole donor for the aldehyde substrate (Kursula *et al.*, 2001). In A-TIM (green, PDB entry 2vel; complexed with PGA in yellow) the peptide after Trp12 is flipped, causing a different conformation of the Asn11 side chain. Hydrogen-bonding interactions are highlighted by dotted lines.

TIM-barrel fold is used in nature for a wide range of reactions, and this fold has therefore also been used in various enzyme-engineering approaches (Khersonsky *et al.*, 2012; Blomberg *et al.*, 2013; Huang *et al.*, 2016). These studies concern monomeric TIM-barrel proteins. It has been shown that the monomeric variant of the TIM dimer, derived from trypanosomal TIM, is still active, but its k_{cat} value is about 1000-fold lower and its K_{m} value is ten times higher (Borchert *et al.*, 1994; Schliebs *et al.*, 1996) compared with wild-type TIM. This monomeric variant was obtained by replacing the extended dimer interface loop, loop 3, with a much shorter loop (Supplementary Fig. S2). Subsequently, studies have been initiated to change the substrate specificity of this variant (Wierenga *et al.*, 2010). In these studies, the phosphate-binding loop, loop 8, was shortened, resulting in A-TIM (Alahuhta, Salin *et al.*, 2008; Salin *et al.*, 2010). In the TIM-barrel fold loop 8 is adjacent to loop 7 and loop 1 (Fig. 1). Loop 7 is involved in binding the phosphate moiety of the substrate, whereas loop 1 provides two catalytic residues, Asn11 and Lys13. Structural comparison of A-TIM and wild-type TIM shows that there are no structural changes in loop 7, whereas loop 1 adopts a different conformation, affecting in particular the properties of the Asn11 side chain, which is caused by a peptide flip of the Trp12-Lys13 peptide bond (Fig. 2). In wild-type TIM loop 8 tightly interacts with the phosphate moiety of the substrate, and indeed A-TIM no longer catalyses the TIM reaction. In order to search for new catalytic activities, the A-TIM variant was subjected to directed-evolution experiments aimed at finding new variants with L-arabinose isomerase (L-AI; EC 5.3.1.4) and D-xylose isomerase (D-XI; EC 5.3.1.5) activity using directed-evolution approaches (Krause *et al.*, 2015). These two catalytic activities are also of interest for biotechnological applications (Staudigl *et al.*, 2014; Langan *et al.*, 2014; Waltman *et al.*, 2014; Salonen *et al.*, 2013). Of these two targets, only the L-AI selection approach resulted in improved new variants of A-TIM. The D-XI directed-evolution experiments failed to find new variants (Krause *et al.*, 2015). In the end, the *in vivo* selection approach with the L-AI knockout variant resulted in the identification of three new variants with *in vivo* L-AI activity: Ma4, Ma18 and Ma21. It was also shown that

A-TIM has *in vivo* L-AI activity (Krause *et al.*, 2015). Expression and characterization studies of these proteins were subsequently initiated in order to understand which biochemical properties allowed A-TIM and the three A-TIM variants to facilitate growth of the selection strain. Here, we report the purification and characterization of the Ma18 and Ma21 variants, including the determination of their crystal structures. The most detailed studies were performed with Ma21. In Ma21 a closed conformation is observed for loop 6/loop 7, which is stabilized by a glycolic acid molecule which interacts with loop 6 and loop 7 *via* hydrogen bonds. It is also reported that the Q65L mutation of Ma21 causes a different side-chain conformation of Asn11 of loop 1, such that Ma21 no longer binds citrate in its active site, in contrast to A-TIM, which was crystallized with a bound citrate in its active site.

2. Materials and methods

2.1. Bacterial strains and plasmids

The appropriate genes were cloned *via* NdeI/HindIII digestion (FastDigest restriction endonucleases; Thermo Scientific) and subsequent ligation (T4 DNA Ligase; Roche) into a derivative of the expression vector pCTUT7 (Kraft *et al.*, 2007). The *Escherichia coli* L-AI gene was amplified by PCR from the *E. coli* chromosome *via* colony PCR using the primers GGGAATTCCATATGACGATTTTGTATAATTA (forward) and CCCAAGCTTTTAGCGACGAAACCCGT-AAT (reverse) including the appropriate restriction sites. The sequences of the cloned genes were confirmed by sequencing. Plasmids were isolated according to the protocol published by Grunzel *et al.* (2014) using the commercial Invisorb Spin Plasmid Mini Two kit (STRATEC Molecular) according to the manufacturer's instructions.

2.2. Expression and purification of A-TIM, Ma18, Ma21 and *E. coli* L-AI

The directed-evolution experiments (Krause *et al.*, 2015) identified three A-TIM variants (Ma4, Ma18 and Ma21) that allow faster growth of the L-AI knockout *E. coli* variant. The Ma4 variant could only be expressed in inclusion bodies, but for Ma18 and Ma21 a successful purification strategy was developed consisting of two chromatography steps. Recombinant *E. coli* strains expressing A-TIM, Ma18, Ma21 or *E. coli* L-AI were cultivated in Luria–Bertani (LB) medium. For expression of *E. coli* L-AI, the LB medium was supplemented with 10 mM MnCl₂ at the start of the cultivation. The cultures were grown at 37°C until the OD₆₀₀ reached 0.6–0.8. The cultivation temperature was reduced to 30°C and protein expression was induced by the addition of IPTG to a final concentration of 100 µM for TIM variants and 400 µM for *E. coli* L-AI. Cells were harvested (by centrifugation at 21 000g for 45 min at 4°C) 12–14 h after induction and stored at –20°C. The cell pellet was suspended in lysis buffer [20 mM triethanolamine (TEA) pH 7.5, 1 mM EDTA, 1 mM DTT, 1 mM NaN₃, 150 mM (NH₄)₂SO₄] supplemented with 5 mM MgCl₂, 25 µg ml^{–1} DNase, 25 µg ml^{–1} RNase and 100 µg ml^{–1}

lysozyme. Complete Protease Inhibitor Cocktail tablets (Roche) were added according to the manufacturer's instructions. The samples were incubated for 60 min on ice and centrifuged at 21 000g for 45 min at 4°C. Prior to purification, the samples were dialyzed against 5 l buffer A (20 mM TEA pH 7.5, 1 mM EDTA, 1 mM DTT, 1 mM NaN₃) overnight at 4°C. TIM variants were applied onto a HighPrep CM FF 16/10 cation-exchange column (GE Healthcare) and *E. coli* L-AI was applied onto a HighPrep DEAE FF 16/10 anion-exchange column (GE Healthcare). Bound proteins were eluted with a linear gradient of 0.5 M NaCl from 0 to 100% in buffer A. TIM variants were further purified by size-exclusion chromatography using HiLoad 16/600 Superdex 200 pg (GE Healthcare) equilibrated with 20 mM Tris pH 7.4. The protein purity was checked by SDS–PAGE and the protein quantity was determined using a Nanodrop 2000 (Thermo Scientific) with an extinction coefficient for monomeric TIM (monoTIM) of 40 575 M^{–1} cm^{–1} (Salin *et al.*, 2010). Subsequently, aliquots of the purified protein solution were rapidly frozen in liquid nitrogen and stored at –80°C.

2.3. Enzymatic assay

The formation of the ketone from the substrate L-arabinose was measured at 290 nm as the increase in absorption based on a previously described protocol (Zhang *et al.*, 2003). The extinction coefficient of the product, L-ribulose, was determined experimentally as $\epsilon = 3.7 \text{ M}^{-1} \text{ cm}^{-1}$ (Krause, 2014). The reaction mixture was pre-incubated at 30°C for at least 10 min before starting the reaction by the addition of 500 µg A-TIM, Ma18 or Ma21 (in 0.25 ml assay mixture in the cuvette). For the A-TIM, Ma18 and Ma21 assays the reaction mixture contained 100 mM L-arabinose (freshly prepared) in 10 mM TEA buffer pH 7.5. The control assay was performed using *E. coli* L-AI (5 µg in the cuvette), which requires metal ions for catalytic activity. These were added to the reaction buffer (10 mM MnCl₂ and 1 mM CoCl₂) of the L-AI assay. Additionally, 1 mM MnCl₂ was added to the L-AI stock solution prior to the assay. The initial rates were obtained from progress curves recorded for 10 min, covering a substrate range from 25 to 500 mM. The k_{cat} and K_{m} values were calculated from the Michaelis–Menten saturation curve using SigmaPlot (v.11; Systat Software Inc). All measurements were performed in triplicate.

2.4. Static light scattering (SLS)

The oligomeric state of Ma18 and Ma21 in solution was determined by the SLS method using a protein buffer consisting of 20 mM Tris pH 7.4. Approximately 50 µl purified protein solution (5 mg ml^{–1}) was loaded onto a HiLoad 16/600 Superdex 200 pg gel-filtration column (GE Healthcare) using an ÄKTApurifier (GE Healthcare) and subsequently passed through a Shodex refractive-index detector (Showa Denko) and a miniDAWN TREOS SLS instrument (Wyatt). The results were analyzed with the ASTRA 5.3 software (Wyatt).

2.5. Circular-dichroism (CD) spectroscopy

To measure the CD spectra, 600 μl 0.1 mg ml⁻¹ protein was dialyzed overnight at 4°C against 100 ml 10 mM sodium phosphate buffer pH 7.2 using Slide-A-Lyzer dialysis devices with 3.5K MWCO (Thermo Scientific). The wavelength range used was 190–250 nm at a temperature of 20°C. A 1 mm quartz cuvette was used. Thermal denaturation was carried out between 20 and 90°C with a temperature ramp of 2°C min⁻¹. The CD instrument used was an Applied Photophysics Chirascan. The corresponding buffer spectra were subtracted for all measurements. For data analysis, *GLOBAL3* (Applied Photophysics) and *Excel* (Microsoft) were used. The error estimates were calculated from the curve-fitting calculations using the same software.

2.6. Crystallization, data collection and structure refinement of Ma18 and Ma21

For the crystallization experiments, the A-TIM, Ma21 and Ma18 proteins were dissolved in 20 mM Tris pH 7.4 (Supplementary Table S1). A-TIM was crystallized using the hanging-drop vapour-diffusion method as described previously (Salin *et al.*, 2010) in 20% PEG 6000, 100 mM citrate pH 5.5. For the Ma21 crystallization experiment 10 mg ml⁻¹ protein solution was supplemented with 100 mM L-arabinose and the well solution (21% PEG 6000, 100 mM citrate pH 5.5) was the same as in the A-TIM crystallization setup, except for the presence of 5% glycerol. The A-TIM crystals were used as a seeding material to grow Ma21 crystals. The obtained Ma21 crystals were used to seed new Ma21 crystals, which were co-crystallized with 100 mM L-arabinose. 10 mg ml⁻¹ Ma18 protein solution was mixed in a 1:1 ratio with 100 mM citrate solution pH 5.5 prior to crystallization in 1.75 M (NH₄)₂HPO₄ pH 8.2. All crystallizations were set up and incubated at room temperature. For Ma18 crystals 100% vacuum-dried paraffin oil was used as a cryoprotectant. Diffraction data were collected from the Ma18 crystal using a MICROSTAR generator with a PLATINUM¹³⁵ detector (Bruker) and were processed with the *PROTEUM2* software (Bruker). Diffraction data for A-TIM and Ma21 crystals were collected on beamline I03 at Diamond Light Source (DLS), Oxford, England. Prior to data collection, the Ma21 crystals were soaked in 21% PEG 6000, 100 mM citrate pH 5.5, 5% glycerol, 15% PEG 400, 10 mM L-arabinose and the A-TIM crystals in 100% vacuum-dried paraffin oil. Crystals were flash-cooled in liquid nitrogen. These data sets were processed at DLS using the *xia2* automatic software pipeline (Winter, 2010; Winter *et al.*, 2013). Data for Ma18 and Ma21 were further processed with the *CCP4* v.6.4.0 suite of programs (Winn *et al.*, 2011). The data-processing statistics are listed in Table 1. The determination of the two structures was achieved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the A-TIM structure (PDB entry 2vek chain A, without ligands and without waters; Alahuhta, Salin *et al.*, 2008) as the search model. Structure building was performed with *Buccaneer* (Cowtan, 2012) and structure refinement was performed with *REFMAC5* (Murshudov *et al.*, 2011). Model building into

Table 1

Data-collection and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

	Ma18	Ma21
Data-collection statistics		
Source	Bruker MICROSTAR	I03, DLS
Detector	PLATINUM ¹³⁵	Pilatus3 6M
Wavelength (Å)	1.542	0.920
Space group	<i>P</i> 2 ₁	<i>P</i> 4 ₃ 2 ₁ 2
Unit-cell parameters (Å, °)	<i>a</i> = 92.92, <i>b</i> = 49.42, <i>c</i> = 95.51, α = 90, β = 118.5, γ = 90	<i>a</i> = <i>b</i> = 62.62, <i>c</i> = 136.01, α = β = γ = 90
No. of molecules per asymmetric unit	3	1
Resolution range (Å)	83.94–2.71 (2.73–2.71)	56.88–1.55 (1.58–1.55)
Completeness (%)	99.5 (99.1)	99.1 (99.7)
$\langle I/\sigma(I) \rangle$	30.4 (12.5)	52.6 (2.9)
<i>R</i> _{p.i.m.} † (%)	1.7 (17.0)	1.6 (42.4)
No. of unique reflections	19927 (1058)	40159 (1953)
Multiplicity	7.0 (1.2)	12.3 (12.9)
Matthews coefficient (Å ³ Da ⁻¹)	2.5	2.2
Solvent content (%)	51	44
Wilson <i>B</i> factor (Å ²)	29.9	33.3
Refinement statistics		
Resolution (Å)	48.8–2.71	56.88–1.55
<i>R</i> _{work} (%)	23.5	17.3
<i>R</i> _{free} (%)	28.3	21.7
No. of atoms	5549	2019
No. of waters	146	202
No. of glycolic acid (GOA) molecules	—	1
Geometry statistics		
R.m.s.d., bond lengths (Å)	0.011	0.018
R.m.s.d., angles (°)	1.5	1.9
Average <i>B</i> factor (Å ²)		
Protein atoms	18.1	28.1
Waters	10.5	43.0
Glycolic acid (GOA)	—	39.8
Ramachandran plot‡ (%)		
Favoured	95.7	99.2
Allowed	3.2	0.8
Outliers§	1.1	0.0
PDB code	4pcf	4pc8

† Calculated with *AIMLESS* (Evans & Murshudov, 2013). ‡ The Ramachandran percentages were calculated by *MolProbity* (Chen *et al.*, 2010). § These outliers are mostly in flexible regions. The outliers of the reference molecule (chain C) are very near the allowed regions.

electron-density maps was performed with *Coot* v.0.7.2 (Emsley & Cowtan, 2004). In the Ma18 crystal structure there are three chains per asymmetric unit. During refinement the noncrystallographic symmetry was restrained using the default protocol of *REFMAC5*. In Ma21 a double conformation was introduced for three residues (Asn11, Glu167 and Ser192). The occupancies of these double conformations were set in *Coot*. The final crystallographic statistics are summarized in Table 1.

2.7. Sequence and structure analysis

A sequence alignment of the different TIM structures was obtained using *ClustalW2* (Larkin *et al.*, 2007) and the secondary-structure assignment was performed using *DSSP* (Kabsch & Sander, 1983) from the CMBI server (<http://www.cmbi.ru.nl/dssp.html>) using PDB entry 5tim (Wierenga *et*

al., 1991). The structures were analysed with *MolProbity* (Chen *et al.*, 2010) and *Coot*. The structures were compared with several A-TIM structures, in particular with the structures of A-TIM complexed with citrate (PDB entry 2vek chain B, 1.6 Å resolution; Alahuhta, Salin *et al.*, 2008) and with the transition-state analogue 2-phosphoglycolate (PGA; PDB entry 2vel chain B, 2.3 Å resolution; Alahuhta, Salin *et al.*, 2008) and the unliganded A-TIM structure (PDB entry 2x16 chain B, 2.1 Å resolution; Salin *et al.*, 2010). Also used for comparisons were the structures of *Plasmodium falciparum* TIM (PDB entry 1ydv chain A, unliganded, 2.2 Å resolution; Velanker *et al.*, 1997), unliganded *Trypanosoma brucei* TIM (PDB entry 5tim chain A, open; Wierenga *et al.*, 1991) and leishmanial TIM complexed with PGA (PDB entry 1n55 chain A, closed, 0.83 Å resolution; Kursula & Wierenga, 2003). The graphical figures were produced with *PyMOL* (<http://www.pymol.org/>).

3. Results

3.1. Solution properties of Ma18 and Ma21

A-TIM and its two variants, Ma18 and Ma21, were purified to homogeneity as judged by SDS-PAGE. Ma18 and Ma21 are expressed better in *E. coli* compared with A-TIM as judged

Table 2

Comparison of some solution properties of A-TIM, Ma18 and Ma21.

	A-TIM	Ma18	Ma21
Point mutations with respect to A-TIM	—	E23G, A70T, S96F, A178V	Q65L
Yield of protein† (mg)	25	50	60
Precipitation at 4°C‡	After 2 d	After one week	After one week
T_m § (°C)	49.5 ± 0.1	51.3 ± 0.1	53.7 ± 0.2

† Total yield of purified protein per litre of culture. ‡ Dissolved in the size-exclusion buffer. The protein concentration was approximately 15 mg ml⁻¹. § Determined by CD measurements (Supplementary Fig. S4).

from the higher yields of purified protein (Table 2). Both Ma18 and Ma21 behaved as monomeric proteins in a size-exclusion chromatography experiment (Supplementary Fig. S3) and both proteins have a higher stability compared with A-TIM as judged from the CD melting curves (Supplementary Fig. S4). The enzymatic properties of A-TIM and its variants were studied by a spectrophotometric assay. Owing to the low absorption coefficient of the keto product, this assay is rather insensitive. As a control, this protocol was also used to measure the Michaelis–Menten parameters of *E. coli* L-AI, resulting in $k_{cat} = 193 \pm 20 \text{ s}^{-1}$ and $K_m = 79 \pm 23 \text{ mM}$. The k_{cat} value corresponds to a specific activity of $210 \mu\text{mol min}^{-1} \text{ mg}^{-1}$. For *E. coli* L-AI the reported K_m value is 60 mM and the specific activity is $200 \mu\text{mol min}^{-1} \text{ mg}^{-1}$ (Patrick & Lee, 1968,

Figure 3
2F_o - F_c OMIT maps obtained after ten cycles of additional OMIT refinement of the two models, leaving out the mutated side chains and the ligand glycolic acid (GOA). (a) Ma18, near E23G (contour level 1σ). (b) Ma18, near A70T (contour level 0.8σ). (c) Ma18, near S96F (contour level 0.8σ). (d) Ma18, near A178V (contour level 0.8σ). (e) Ma21, near Q65L (contour level 0.8σ). (f) Ma21, near the bound glycolic acid (GOA) (contour level 1σ). The refined models are shown in pink and the A-TIM reference model (PDB entry 2vek) is shown in green.

1975). For A-TIM the specific activity in a buffer with 100 mM L-arabinose and 500 µg enzyme in the cuvette was found to be at the detection limit of this assay. The specific activity could also not be quantified precisely for Ma21 and Ma18.

3.2. The crystal structure of Ma18

Ma18 has four sequence differences with respect to A-TIM (Supplementary Fig. S2). In the crystal structure of Ma18 refined at 2.7 Å resolution there are three molecules in the asymmetric unit. There is no ligand bound in the active sites. The three molecules are assembled as a trimer, and in this assembly the intertrimer contacts stabilize an open loop 6 conformation. In each of the subunits both loop 6 and loop 7 are found to be in the open conformation, which agrees with the notion that these two loops move in a concerted conformational change (Casteleijn *et al.*, 2006). In each of the molecules loop 1 has the highest *B* factors, and residues 16–19 of loop 1 are ill-defined in the map in molecules *A* and *B*. Only loop 1 of molecule *C* could be built completely and it follows a similar trace as observed in A-TIM. Superposition of chain *C* of Ma18 on the A-TIM-PGA structure (chain *B*) reveals r.m.s.d.s for C α atoms of 0.9 Å (excluding loop 6) and 0.6 Å (excluding loop 1, loop 3 and loop 6). The structures of the other loops of molecule *C* are also clear from the electron-density map, and superposition calculations reveal that the three chains mutually superimpose with an r.m.s.d. of 0.5 Å for the common C α atoms. The conformation of the active-site residues of molecule *C* is well defined, except for the side chain of Lys13. The side chain of Lys13 is also poorly defined in the unliganded A-TIM structure (Salin *et al.*, 2010).

The four point mutations of Ma18 with respect to A-TIM are clearly visible in the electron-density map (Fig. 3). These mutations cause only minor perturbations of the structure and are located in or near the 'front' loops (Fig. 4). The E23G mutation occurs at the beginning of helix 1. The A70T mutation occurs at the beginning of helix 3. The Glu23 and Ala70 side chains point into bulk solvent. The S96F and

A178V mutations are closer to functional regions of the TIM active site. The S96F point mutation occurs at the beginning of the helical part of loop 4 and the A178V mutation is at the end of loop 6, near the beginning of helix 6. The Ser96 side chain points outwards and the Ala178 side chain points inwards. In a previous study (Alahuhta, Casteleijn *et al.*, 2008) the A178L variants of wild-type TIM and the mutated catalytically competent monomeric variant ml1TIM were studied. It was noted that a more bulky side chain at this position favours the closure of loop 6. It was also observed that the A178L mutation decreases the stability as well as the catalytic efficiency.

The S96F mutation involves the highly conserved Ser96 (Gayathri *et al.*, 2009). The Ser96 side chain points outwards towards loop 6 and bulk solvent. In the unliganded active site it is hydrogen-bonded to the Glu167 side chain, adopting a swung-out conformation. In some sequences a phenylalanine is observed at this position, as in *P. falciparum* TIM (Gayathri *et al.*, 2009; Velanker *et al.*, 1997; Parthasarathy *et al.*, 2002). It has been proposed that the presence of this phenylalanine favours the open conformation of loop 6. In any case, in Ma18 loop 6 is observed in the open state (Supplementary Fig. S5), like loop 7. Loop 4 of Ma18 adopts the same conformation as observed in trypanosomal TIM (Supplementary Fig. S5). The Phe96 side chain of Ma18 adopts the same conformation as in *P. falciparum* TIM (Supplementary Fig. S6). Comparison of the active sites of A-TIM and Ma18 near Asn11 of loop 1 (Fig. 5a) shows that its Asn11 side chain has adopted a different conformation, which is correlated with the different position of Trp12 O. There are no further significant structural differences in the active-site region between A-TIM and Ma18. Gln65 NE2 is hydrogen-bonded to Asn11 OD1, as in A-TIM (Fig. 5a).

3.3. The crystal structure of Ma21

Ma21 differs by one point mutation from A-TIM: Q65L (Supplementary Fig. S2). This point mutation was frequently

Figure 4
Ribbon trace of A-TIM (same view as in Fig. 1; PDB entry 2vel). Highlighted are the five A-TIM residues that were mutated in these experiments: E23G, A70T, S96F and A178V (in Ma18) and Q65L (in Ma21). The active site is identified by the bound active-site ligand PGA, which is shown as a skeleton model.

found in these directed-evolution studies (Krause *et al.*, 2015). In the crystal structure of Ma21 there is one molecule per asymmetric unit and this structure was refined at 1.55 Å resolution. The polypeptide chain of Ma21 could be completely built and the conformation of each of the loops is well defined. The side chains of the active-site residues are in good density, except for Lys13 which is disordered, as in Ma18 and in the unliganded A-TIM structure. The side chains of the active-site residues Asn11 and Glu167 adopt double conformations with approximately the same occupancy. Loop 4 of this structure adopts the out conformation, as previously described for monoTIM (Borchert *et al.*, 1993). With respect to the structures of Ma18 and of the A-TIM-PGA complex the front loops loop 1, loop 3, loop 4 and loop 6 adopt a different conformation. Superposition on the Ma18 structure (chain C) reveals an r.m.s.d. of 0.6 Å (excluding loop 1, loop 3, loop 4 and loop 6) for C α atoms and for the superposition on the A-TIM-PGA structure (chain B) the r.m.s.d. is 0.8 Å (excluding loop 1, loop 3, loop 4 and loop 6).

Loop 6 and loop 7 are in the closed conformation in Ma21 (Supplementary Fig. S5). The closed conformation of these two TIM loops is only observed in the presence of a substrate-analogue ligand such as sulfate, phosphate or citrate or a transition-state analogue such as 2-phosphoglycolate (PGA). Indeed, in the Ma21 crystal form a ligand is also bound between loop 6 and loop 7 (Figs. 3 and 6), which could be identified as glycolate. Although L-arabinose was present in the crystallization and cryo buffers (Supplementary Table S1) it is not bound in the active site. Glycolate is apparently present as an oxidation product of ethylene glycol, which is the

monomer of PEG. It has been found in a few other protein structures deposited in the PDB. The carboxylate group of glycolate mimics the phosphate moiety of an A-TIM-bound 2-phosphoglycolate and the carboxylate of citrate bound to the A-TIM active site, as shown in Fig. 6. The carboxylate O atom of glycolate forms a hydrogen bond to Gly173 N (loop 6) and to Ser213 N (loop 7). The distance between Gly173 N and Ser213 N in this structure is 4.9 Å.

The Q65L sequence change is, like the Ma18 mutations, in a front loop, in this case loop 3, which is just after β -strand 3. The Leu65 side chain of Ma21 is clearly visible in the electron-density map (Fig. 3). As discussed below, this mutation causes some structural changes in its immediate neighbourhood but no gross structural rearrangements. Gln65 is rather well conserved in TIM. In wild-type TIM it is buried at the dimer interface and its Gln65 NE2 side-chain atom is hydrogen-bonded to loop 1, in particular to Lys13 O (Fig. 2). In monomeric TIMs Gln65 also forms a hydrogen-bond to loop 1 but in a different way, being hydrogen-bonded to Asn11 OD1. This is also observed in Ma18 (Fig. 5a).

The point mutation Q65L causes small rearrangements in loop 1 and loop 3. Leu65 of Ma21 protrudes out of loop 3 towards loop 1. It pushes away the side chain of Asn11 (Fig. 5b). In A-TIM, the Asn11 side chain stabilizes the mode of binding of citrate to A-TIM and, although Ma21 crystallizes in the same mother liquor as was used in the A-TIM crystallization protocol, a citrate molecule (present in the A-TIM active site) is not bound in its active site. Apparently owing to the Q65L mutation of Ma21, the affinity of A-TIM for citrate has been lost. Leu65 also pushes away Trp12 O in loop 1 from

Figure 5
The loop 1 structure near Asn11 of Ma18 and Ma21. (a) Superposition of Ma18 (blue) and A-TIM (green, complexed with PGA; PDB entry 2vel). (b) Superposition of Ma21 (blue) and A-TIM (pink, complexed with citrate; PDB entry 2vek). In the latter complex the mode of binding of citrate is stabilized by a hydrogen bond to the Asn11 side chain, which is not possible in the Ma21 active site as this side chain is being pushed away by the Leu65 side chain.

its position in A-TIM (Fig. 5*b*) and it makes hydrophobic contacts with Ala42 (loop 2), Thr44 (loop 2) and also Val92 (loop 4), as well as with the C^β atoms of Asn11 and Lys13 (Fig. 7). Owing to the rearrangements in loop 1 near Asn11 in Ma21 the Asn11 O–Leu238 N hydrogen-bond interaction becomes somewhat stronger (3.0 Å; in the A-TIM–PGA structure this distance is 3.4 Å).

4. Discussion

The crystal structures of Ma18 and Ma21 have been determined to determine whether the structural data can provide further clues towards understanding the *in vivo* properties of Ma18 and Ma21. Ma21 is the best-studied variant both *in vivo* and *in vitro*. The structure of Ma21 reveals that loop 6/loop 7 adopt the closed, active conformation in the presence of a bound glycolate ligand (Supplementary Fig. S5), whereas in Ma18 the active site is unliganded and loop 6/loop 7 adopt the open conformation. The crystal structures show that for both Ma18 and Ma21 there are small structural changes near the active site owing to the point mutations. Most notable are the S96F mutation of Ma18 and the Q65L mutation of Ma21. The Ma18 structure is at a relatively low resolution (2.7 Å), whereas the Ma21 structure is at near-atomic resolution (1.55 Å). Overall, from the comparison of the structures of Ma18 and A-TIM it can be noted that the A178V and S96F mutations are not predicted to be stabilizing mutations. Nevertheless, together with the E23G and A70T mutations,

these four point mutations cause Ma18 to have a somewhat higher T_m value than that of A-TIM (Supplementary Fig. S4). All five mutations are at or near the catalytic site of the TIM barrel (Fig. 4). In the active-site region only a few structural changes have been noted near loop 1 of Ma18 (Fig. 5*a*). In the Ma21 structure it can be clearly seen that the Q65L mutation causes structural differences near loop 1 (Fig. 5*b*), providing different active-site citrate-binding properties. The Ma21 crystallization condition is the same as that used for A-TIM (Supplementary Table S1), including the citrate buffer, and the Ma21 crystals could only be obtained by seeding with a suspension of A-TIM crystals; nevertheless, in the A-TIM structure the active site is complexed with citrate, but in the Ma21 structure citrate is not bound in the active site. The high-resolution structure of Ma21 shows that in A-TIM, owing to the Q65L mutation, the Asn11 side chain adopts a different position (Fig. 5*b*), disfavoured its interactions with the bound citrate as present in the A-TIM–citrate complex. Therefore, the small structural changes near Asn11 of loop 1 of Ma21 cause the active sites of Ma21 and A-TIM to have different binding properties. Importantly, the Leu65 side chain makes hydrophobic contacts with nearby side chains of loop 1, loop 2 and loop 4 (Fig. 7). These favourable interactions correlate with the higher T_m value of Ma21 (Supplementary Fig. S4).

In the directed-evolution experiments the A-TIM gene was inserted in its expression plasmid with a strong promoter, favouring strong expression during the growth of the selection *E. coli* strain (Krause *et al.*, 2015). The selection experiments

Figure 6

Comparison of the modes of binding of glycolate to Ma21 (blue; PDB entry 4pc8), citrate to A-TIM (green, PDB entry 2vek) and PGA to A-TIM (grey; PDB entry 2vel) in the closed loop 6/loop 7 active site. The hydrogen bonds between the carboxylate O atom of glycolic acid (GOA) and Gly173 N (loop 6) and Ser213 N (loop 7) are shown as dotted lines.

Figure 7

The favourable hydrophobic interactions of the Leu65 side chain of Ma21 (PDB entry 4pc8). The dotted lines visualize the carbon–carbon van der Waals contacts of the Leu65 side-chain atoms with protein atoms within 4.4 Å: Asn11 C^β, Lys13 C^β, Ala42 C^β, Thr44 C^{γ2}, Val92 C^β, Val92 C^{γ1} and Val92 C^{γ2}.

were performed at 30°C and the concentration of L-arabinose in the growth medium was 40 mM. Through this experimental approach it was found that A-TIM already possesses L-arabinose isomerase activity *in vivo* and the expression of the A-TIM variants Ma18 and Ma21 resulted in more efficient growth when expressed in the L-AI knockout selection strain. For example, the Ma18 and Ma21 colonies appear earlier than the A-TIM colonies. The L-AI activity of A-TIM was unexpected. In any case, the chirality of the hydroxyl group of the aldehyde substrate that is converted to a keto group in the isomerization reactions of L-AI and TIM is the same (Supplementary Fig. S1). However, the L-AI substrate differs in several other respects from the natural substrate of TIM, such as (i) it does not have a phosphate group and (ii) it occurs mostly in a cyclic form in solution (Matthews *et al.*, 2000). The L-arabinose isomerase activity of A-TIM is very specific as the expression of A-TIM does not rescue the growth of the D-XI knockout selection strain (Krause *et al.*, 2015). Apparently, A-TIM does not have *in vivo* D-XI activity, although the only structural difference between D-xylose and L-arabinose is the chirality of the C4 atom (Supplementary Fig. S1). From the characterization studies of A-TIM, Ma18 and Ma21 it emerges that Ma18 and Ma21 are ‘better’ proteins than A-TIM, being expressed at higher levels, being more stable and being more soluble than A-TIM (Table 2). It is also possible that the Ma18 and Ma21 mutations allow more efficient folding of these proteins (Kursula *et al.*, 2002). In this respect, it can be noted that the four mutations in Ma18 (E23G, S96F, A70T and A178V) are found at or very near the N-terminus of an α -helix (Fig. 4), and residues near the N-cap of an α -helix can be important for facilitating the formation of the α -helix (elMasry & Fersht, 1994; Richardson & Richardson, 1989). The characterization of the biophysical properties of Ma18 and Ma21 rationalizes to some extent the observation that they provide better growth properties to the selection strain than A-TIM. It is puzzling that these variants have at most residual L-arabinose isomerase activity, although the *in vivo* data clearly suggest the presence of such activity. Further quantitative structural enzymological studies of A-TIM, Ma21 and Ma18 are required for the characterization of their catalytic L-arabinose isomerase properties. These results indicate that when planning directed-evolution selection experiments careful thought has to be given to the selection pressure of the experimental system. Given the importance of the potential biotechnology applications of the L-arabinose isomerase conversion, further directed-evolution studies of this system at higher evolutionary pressure will be initiated.

Acknowledgements

This work was supported by the Academy of Finland (grant No. 89211) as well as through Short-term Fellowships from the European Molecular Biology Organization (EMBO) and the Federation of European Biochemical Societies (FEBS) and through travel grants from the Boehringer Ingelheim Fonds and the DAAD. This work is part of the Cluster of Excellence ‘Unifying Concepts in Catalysis’ coordinated by the

Technische Universität Berlin. It was supported by the Deutsche Forschungsgemeinschaft (DFG) within the framework of the German Initiative for Excellence, which is gratefully acknowledged (EXC 314). The research leading to these results received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under BioStruct-X (grant agreement No. 283570). We thank Ville Ratas for technical assistance with protein purification and crystallization and the staff of beamline I03 at DLS for excellent support. The coordinates and structure factors have been deposited in the RCSB PDB for immediate release. The PDB codes are listed in Table 2. The use of the facilities and the expertise of the Biocenter Oulu core facility, a member of Biocenter Finland and Instruct-FI, are gratefully acknowledged.

References

- Alahuhta, M., Casteleijn, M. G., Neubauer, P. & Wierenga, R. K. (2008). *Acta Cryst.* **D64**, 178–188.
- Alahuhta, M., Salin, M., Casteleijn, M. G., Kemmer, C., El-Sayed, I., Augustyns, K., Neubauer, P. & Wierenga, R. K. (2008). *Protein Eng. Des. Sel.* **21**, 257–266.
- Blomberg, R., Kries, H., Pinkas, D. M., Mittl, P. R., Grütter, M. G., Privett, H. K., Mayo, S. L. & Hilvert, D. (2013). *Nature (London)*, **503**, 418–421.
- Borchert, T. V., Abagyan, R., Jaenicke, R. & Wierenga, R. K. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1515–1518.
- Borchert, T. V., Abagyan, R., Kishan, K. R., Zeelen, J. & Wierenga, R. (1993). *Structure*, **1**, 205–213.
- Casteleijn, M. G., Alahuhta, M., Groebel, K., El-Sayed, I., Augustyns, K., Lambeir, A., Neubauer, P. & Wierenga, R. K. (2006). *Biochemistry*, **45**, 15483–15494.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Cowtan, K. (2012). *Acta Cryst.* **D68**, 328–335.
- elMasry, N. F. & Fersht, A. R. (1994). *Protein Eng.* **7**, 777–782.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Gayathri, P., Banerjee, M., Vijayalakshmi, A., Balaram, H., Balaram, P. & Murthy, M. R. N. (2009). *Acta Cryst.* **D65**, 847–857.
- Grunzel, P., Pilarek, M., Steinbrück, D., Neubauer, A., Brand, E., Kumke, M. U., Neubauer, P. & Krause, M. (2014). *Biotechnol. J.* **9**, 128–136.
- Huang, P.-S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D. A., Höcker, B. & Baker, D. (2016). *Nature Chem. Biol.* **12**, 29–34.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Khersonsky, O., Kiss, G., Röthlisberger, D., Dym, O., Albeck, S., Houk, K. N., Baker, D. & Tawfik, D. S. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 10358–10363.
- Knowles, J. R. (1991). *Nature (London)*, **350**, 121–124.
- Kraft, M., Knüpfer, U., Wenderoth, R., Kacholdt, A., Pietschmann, P., Hock, B. & Horn, U. (2007). *Appl. Microbiol. Biotechnol.* **76**, 1413–1422.
- Krause, M. (2014). PhD thesis. Technical University of Berlin, Germany.
- Krause, M., Neubauer, P. & Wierenga, R. K. (2015). *Protein Eng. Des. Sel.* **28**, 187–197.
- Kursula, I., Partanen, S., Lambeir, A., Antonov, D. M., Augustyns, K. & Wierenga, R. K. (2001). *Eur. J. Biochem.* **268**, 5189–5196.
- Kursula, I., Partanen, S., Lambeir, A. & Wierenga, R. K. (2002). *FEBS Lett.* **518**, 39–42.
- Kursula, I. & Wierenga, R. K. (2003). *J. Biol. Chem.* **278**, 9544–9551.
- Langan, P. *et al.* (2014). *Structure*, **22**, 1287–1300.

- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- Matthews, C. K., van Holde, K. E. & Ahern, K. G. (2000). *Biochemistry*, 3rd ed., pp. 284–286. San Francisco: Addison Wesley Longman.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Parthasarathy, S., Balaram, H., Balaram, P. & Murthy, M. R. N. (2002). *Acta Cryst.* **D58**, 1992–2000.
- Patrick, J. & Lee, N. (1968). *J. Biol. Chem.* **243**, 4312–4318.
- Patrick, J. & Lee, N. (1975). *Methods Enzymol.* **41**, 453–458.
- Richard, J. P., Zhai, X. & Malabanan, M. M. (2014). *Bioorg. Chem.* **57**, 206–212.
- Richardson, J. S. & Richardson, D. C. (1989). *Principles and Patterns of Protein Conformation*, edited by G. D. Fasman, pp. 1–98. New York: Plenum Press.
- Salin, M., Kapetaniou, E. G., Vaismaa, M., Lajunen, M., Casteleijn, M. G., Neubauer, P., Salmon, L. & Wierenga, R. K. (2010). *Acta Cryst.* **D66**, 934–944.
- Salonen, N., Salonen, K., Leisola, M. & Nyssölä, A. (2013). *Bioprocess Biosyst. Eng.* **36**, 489–497.
- Schliebs, W., Thanki, N., Eritja, R. & Wierenga, R. (1996). *Protein Sci.* **5**, 229–239.
- Silverman, J., Balakrishnan, R. & Harbury, P. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 3092–3097.
- Staudigl, P., Haltrich, D. & Peterbauer, C. K. (2014). *J. Agric. Food Chem.* **62**, 1617–1624.
- Velanker, S. S., Ray, S. S., Gokhale, R. S., Suma, S., Balaram, H., Balaram, P. & Murthy, M. R. N. (1997). *Structure*, **5**, 751–761.
- Waltman, M. J., Yang, Z. K., Langan, P., Graham, D. E. & Kovalevsky, A. (2014). *Protein Eng. Des. Sel.* **27**, 59–64.
- Wierenga, R., Kapetaniou, E. & Venkatesan, R. (2010). *Cell. Mol. Life Sci.* **67**, 3961–3982.
- Wierenga, R. K., Noble, M. E. M., Vriend, G., Nauche, S. & Hol, W. G. J. (1991). *J. Mol. Biol.* **220**, 995–1015.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* **D69**, 1260–1273.
- Zhang, R., Andersson, C. E., Savchenko, A., Skarina, T., Evdokimova, E., Beasley, S., Arrowsmith, C. H., Edwards, A. M., Joachimiak, A. & Mowbray, S. L. (2003). *Structure*, **11**, 31–42.