



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



## X-ray crystal structure of *Saccharomyces cerevisiae* Pdx1 provides insights into the oligomeric nature of PLP synthases

Martina Neuwirth<sup>a,1</sup>, Marco Strohmeier<sup>b,1</sup>, Volker Windeisen<sup>b</sup>, Silvia Wallner<sup>a</sup>, Sigrid Deller<sup>a</sup>, Karsten Rippe<sup>c</sup>, Irmgard Sinning<sup>b</sup>, Peter Macheroux<sup>a,\*</sup>, Ivo Tews<sup>b,\*</sup>

<sup>a</sup> Technische Universität Graz, Institut für Biochemie, Petersgasse 12/2, A-8010 Graz, Austria

<sup>b</sup> Heidelberg University Biochemistry Center (BZH), Im Neuenheimer Feld 328, 69120 Heidelberg, Germany

<sup>c</sup> Deutsches Krebsforschungszentrum and BIOQUANT, Research Group Genome Organization & Function, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

### ARTICLE INFO

#### Article history:

Received 20 October 2008

Revised 1 June 2009

Accepted 4 June 2009

Available online 11 June 2009

Edited by Hans Eklund

#### Keywords:

Glutamine amidotransferase  
Vitamin B6 (pyridoxal 5-phosphate)  
Pdx  
Snz  
Oligomerisation  
X-ray structure  
Twinning

### ABSTRACT

**The universal enzymatic cofactor vitamin B6 can be synthesized as pyridoxal 5-phosphate (PLP) by the glutamine amidotransferase Pdx1. We show that *Saccharomyces cerevisiae* Pdx1 is hexameric by analytical ultracentrifugation and by crystallographic 3D structure determination. Bacterial homologues were previously reported to exist in hexamer:dodecamer equilibrium. A small sequence insertion found in yeast Pdx1 elevates the dodecamer dissociation constant when introduced into *Bacillus subtilis* Pdx1. Further, we demonstrate that the yeast Pdx1 C-terminus contacts an adjacent subunit, and deletion of this segment decreases enzymatic activity 3.5-fold, suggesting a role in catalysis.**

#### Structured summary:

MINT-7147859: *PDX1* (uniprotkb:P16451) and *PDX1* (uniprotkb:P16451) bind (MI:0407) by cosedimentation in solution (MI:0028)  
MINT-7147899: *PDX1* (uniprotkb:P37528) and *PDX1* (uniprotkb:P37528) bind (MI:0407) by cosedimentation in solution (MI:0028)

© 2009 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

The compounds pyridoxal, pyridoxin, pyridoxamin and their phosphorylated derivatives are commonly known as vitamin B6. Being one of the most versatile molecules in prokaryotic and eukaryotic cells, vitamin B6 is involved in many enzymatic reactions such as transamination and decarboxylation in amino acid metabolism and biosynthesis of antibiotics [1], has a function in the maintenance of the nervous and immune systems in higher eukaryotes [2], and can directly act as an antioxidant [3,4].

Two exclusive pathways for *de novo* vitamin B6 biosynthesis exist [5]. The so-called deoxyxylose 5-phosphate (DXP) independent

pathway is predominant and found in archaea, most eubacteria and eukarya [6]. It involves the glutamine amidotransferase [7] PLP synthase, shown to be sufficient for production of pyridoxal 5-phosphate (PLP), a biologically active form of vitamin B6 [8,9]. A pent(ul)ose and a triose are required for PLP biosynthesis by the enzyme Pdx1 (Scheme 1). The associated glutaminase subunit Pdx2 hydrolyses glutamine to derive glutamate and the highly reactive ammonia, the source for the required ring nitrogen [7,10]. Pdx1 catalyzes a large number of enzymatic steps of different reaction types [11].

Several independent lines of research led to the discovery of the enzymes involved. Pdx1 was known as a stress-induced/ethylene responsive gene from the rubber tree *Hevea brasiliensis* [12] and the sponge *Suberites domuncula* [13]. It was independently characterized as singlet oxygen resistance mediating in *Cercospora nicotianae* [14]. Very recently, horizontal gene transfer of a Pdx1 protein from bacteria to the nematode *Heterodera glycines* has been postulated [15].

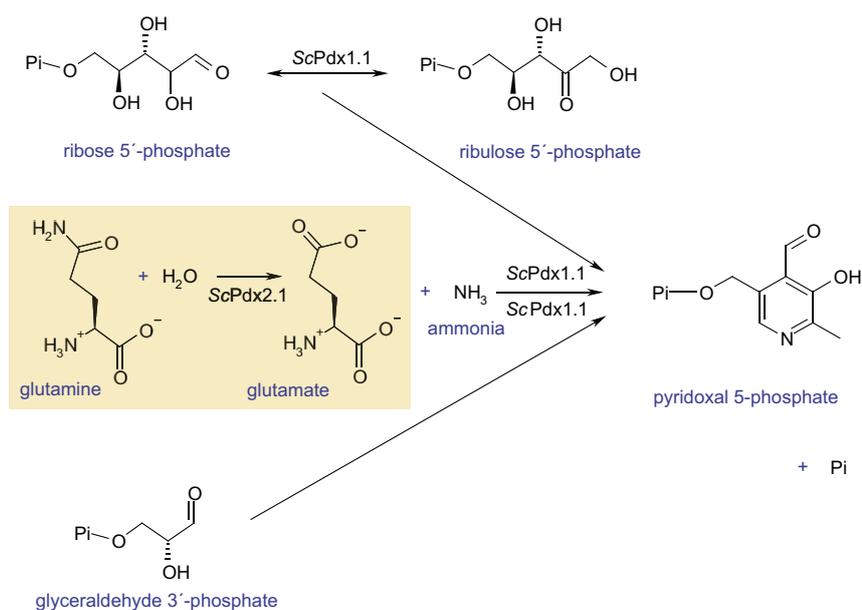
In *Saccharomyces cerevisiae*, Pdx1 was originally discovered as a stationary phase up-regulated gene [16]. The three identified Pdx1 homologues *SNZ1*, *SNZ2*, and *SNZ3* are located adjacent to the Pdx2 glutaminase homologues *SNO1*, *SNO2* and *SNO3* [17]. Genes are

**Abbreviations:** BsPdx1, Pdx1 from *Bacillus subtilis*; BsPdx1His<sub>6</sub>@K178, insertion mutant of BsPdx1 with a lysine residue in position 178; IPTG, isopropyl β-D-1-thiogalactopyranoside; *K<sub>d</sub>*, dissociation constant; MTG, 1-thioglycerol; PLP, pyridoxal 5-phosphate; ScPdx1.1, Pdx1 from *Saccharomyces cerevisiae*; Tris, 2-amino-2-(hydroxymethyl)-1,3-propanediol

\* Corresponding authors. Fax: +43 316 8736 952 (P. Macheroux); +49 6221 5447 90 (I. Tews).

E-mail addresses: [peter.macheroux@tugraz.at](mailto:peter.macheroux@tugraz.at) (P. Macheroux), [ivo.tews@bzh.uni-heidelberg.de](mailto:ivo.tews@bzh.uni-heidelberg.de) (I. Tews).

<sup>1</sup> These authors contributed equally to this work.



**Scheme 1.** Biosynthesis of PLP from a pentasaccharide, a trisaccharide and glutamine by the glutamine amidotransferase PLP synthase. The glutaminase ScPdx2.1 (Sno1) hydrolyses L-glutamine to glutamate and ammonia; the synthase ScPdx1.1 (Snz1) catalyzes isomerization of the pentose substrate ribose 5'-phosphate (R5P) to ribulose 5'-phosphate (Ru5P) and subsequent synthesis of PLP, incorporating the trisaccharide glycerinaldehyde 3'-phosphate (G3P) and ammonia produced by ScPdx2.1. Pi, inorganic phosphate.

co-regulated in tandem, and the Snz1 and Sno1 proteins have been shown to interact [17]. SNZ2 and SNZ3 are up-regulated upon thiamine suppression in the growth medium [18], suggesting a link to vitamin B1 synthesis. The transcription factor Pho4, an *in vivo* substrate of the Pho85 kinase, is involved in regulation of the timing of SNZ1 but not SNZ2/3 expression [19].

X-ray crystallographic studies revealed a dodecameric structure for the Pdx1 synthase subunits from *Bacillus subtilis* [20] and *Geobacillus stearothermophilus* [21]. Pdx1 has a  $(\beta\alpha)_8$ - or TIM-barrel fold [22] and was reported to exist in a salt dependent hexamer-dodecamer equilibrium in solution [20,21]. The dodecamer is formed by two interdigitating rings, each consisting of six Pdx1 molecules. Pdx2 subunits attach to the Pdx1 oligomer, and fully decorated complexes with an approximate molecular weight of 650 kDa contain 12 Pdx2 subunits attached to a Pdx1 dodecamer. PLP synthase complexes have been characterized structurally from *B. subtilis* [20] and from *Thermotoga maritima* [23].

We have investigated the *S. cerevisiae* Pdx1 homologue Snz1. For consistency, we use the Pdx nomenclature, and name Snz1, Snz2 and Snz3 as ScPdx1.1, ScPdx1.2 and ScPdx1.3, respectively, while Sno1, Sno2 and Sno3 are referred to as ScPdx2.1, ScPdx2.2 and ScPdx2.3 throughout the text.

## 2. Materials and methods

### 2.1. Reagents

Ampicillin, isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), imidazole, 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris), NaCl, and  $(\text{NH}_4)_2\text{SO}_4$  were from Roth (Carl Roth, Karlsruhe, Germany).  $\text{Na}_2\text{HPO}_4$ ,  $\text{NaH}_2\text{PO}_4$  and KCl were from Merck (Merck KGA, Darmstadt, Germany). 1-Thioglycerol (MTG), DL-glyceraldehyde 3'-phosphate, and ribose 5'-phosphate were from Sigma (Sigma-Aldrich, Vienna, Austria).

### 2.2. Molecular biology

Genomic DNA from *S. cerevisiae* strain BY4741 was isolated as described [24]. ScPdx1.1 (yeastgenome.org) was amplified by stan-

dard PCR, using custom primers (VBC-Biotech Service GmbH, Vienna, Austria). The fragment was NdeI/EcoRI (New England Biolabs GmbH, Frankfurt, Germany) cloned into pET21a(+) (Novagen Merck, Nottingham, UK) to generate C-terminally tagged ScPdx1.1His<sub>6</sub>.

Site directed mutagenesis was carried out using the QuickChange<sup>®</sup> XL Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA): native ScPdx1.1 – introduction of a stop-codon after Trp297; ScPdx1.1 $\Delta$ C<sub>268–297</sub> – introduction of a stop-codon after Val267; ScPdx1.1 $\Delta$ C<sub>276–297</sub> – introduction of a stop-codon after Met275; BsPdx1His<sub>6</sub>@K178 – introduction of a lysine codon at position 178 into BsPdx1His<sub>6</sub> pET21a(+) [9].

### 2.3. Protein expression and purification

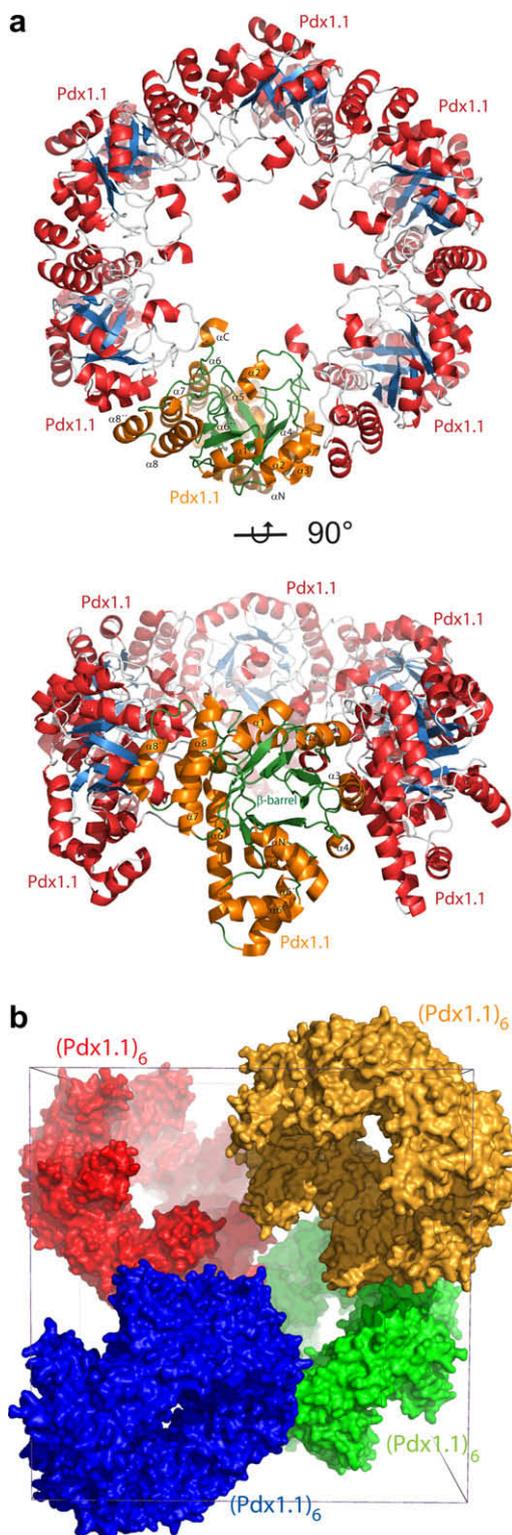
Chemically competent *Escherichia coli* BL21 (DE3) cells (Stratagene) were transformed with respective plasmids. Cells were grown at 37 °C in LB media containing ampicillin (100  $\mu$ g/ml) at 37 °C, induced with 0.1 mM IPTG (final concentration) at an OD<sub>600</sub> of 0.6, harvested after an additional 3 h, washed with 0.9% NaCl, and stored at –20 °C.

For purification of hexahistidine tagged proteins, cells were thawed (lysis buffer 50 mM  $\text{NaH}_2\text{PO}_4$ , pH 8.0, containing 300 mM NaCl, 10 mM imidazole), lysed by sonication, cleared by ultracentrifugation (40 000 $\times$ g, 30 min, 4 °C), and applied onto a Ni-NTA HP column (GE Healthcare, Freiburg, Germany). After a wash step (lysis buffer containing 50 mM imidazole) and elution (lysis buffer containing 250 mM imidazole), the protein containing fractions were concentrated using 10 kDa molecular weight cut-off Centrprep<sup>®</sup> centrifugal filter devices (Millipore, Vienna, Austria) and applied onto a size exclusion column (S200 26/60, GE Healthcare), equilibrated with 20 mM Tris-HCl, pH 8.0, containing 300 mM KCl, 0.02% MTG.

For purification of native proteins, cells were thawed (lysis buffer 50 mM  $\text{KH}_2\text{PO}_4$ , pH 7.0, containing 10 mM NaCl, 0.02% MTG), lysed by sonication, cleared by ultracentrifugation (40 000 $\times$ g, 30 min, 4 °C), and applied onto a DEAE-Sephacel column (GE Healthcare). A linear gradient from 10 mM NaCl to 300 mM NaCl in lysis buffer in 8 column volumes was used for elution. Desired

fractions (checked by SDS–PAGE) were combined and concentrated before application onto a size exclusion column (S200 26/60, see above). Fractions containing ScPdx1.1 were brought to 750 mM ammonium salt concentration and applied to a phenyl sepharose

FF column (GE Healthcare), equilibrated with 50 mM Tris–HCl buffered at pH 8.0, containing 10 mM NaCl, 1.3 mM EDTA, 0.02% MTG, 10% glycerol and 750 mM ammonium sulfate. A linear gradient from 750 mM to 0 mM ammonium sulfate in 8 column volumes was used for elution.



**Fig. 1.** 3D crystallographic structure of ScPdx1.1. (a) Cartoon representation of the hexameric ScPdx1.1 structure in two orthogonal views. Helices and strands are colored in red and blue, respectively.  $\alpha$ -Helices have been labeled for the monomer highlighted in orange and green. (b) Four ScPdx1.1 hexamers, shown here in surface representation, pack into the unit cell (axes  $\sim 154$  Å, compare Table 1).

#### 2.4. 3D structure determination

ScPdx1.1 or ScPdx1.1His<sub>6</sub> (20 mg ml<sup>-1</sup>) were crystallised by sitting drop vapour diffusion using 4 M Na-formate as crystallisation buffer. ScPdx1.1 crystals of a size of 50  $\mu$ m  $\times$  50  $\mu$ m  $\times$  50  $\mu$ m were flash frozen in liquid nitrogen using mother liquor containing 25% (v/v) PEG 400 as cryo-protectant. Data to 3.0 Å were collected from a single crystal (beamline ID14-3, European Synchrotron Radiation Facility, Grenoble, France) with high redundancy to improve  $I/\sigma(I)$ . Data reduction with denzo/scalepack (HKL Research Inc.) and truncate [25] was followed by molecular replacement with MOLREP [26] using Pdx1 from *B. subtilis* as replacement model (PDB code: 2NV1 [20]). Model correction and refinement were performed in iterative steps with coot [27] and Refmac5 [28], first with overall and in the final steps with atom *B*-factor refinement. NCS averaging with tight main- and side-chain restraints was used for all chains but excluded helices  $\alpha$ N (residues 3–16) and  $\alpha$ 1 (residues 28–35) that make crystal contacts; two further separate NCS restraints were set up for chains A, C and E and chains B, D and F.

The ScPdx1.1 dataset was plagued by pseudosymmetry and crystal twinning. Although the crystal unit cell appeared cubic and merged relatively well in P2<sub>1</sub>3, the 3-fold turned out to be non-crystallographic, and the structure had thus to be refined in P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. However, the metric of the unit cell and placement of molecules within the cell (compare Fig. 1b) allowed for pseudo-morohedral twinning in two directions, as detected by Xtriage [30] and refined with Refmac5. An account on crystal pathologies is given in Supplementary data; crystallographic statistics are given in Table 1. Coordinates have been deposited with the Protein

**Table 1**  
Crystallographic analysis.

Data statistics	ScPdx1.1
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell <i>a</i> , <i>b</i> , <i>c</i> (Å)	154.1 $\times$ 154.2 $\times$ 154.9
Resolution (Å)/high resolution shell (Å)	20.0–3.02/3.05–3.02
<i>R</i> <sub>merge</sub> (%) <sup>a</sup> /high resolution shell	7.5/41.6
Completeness (%) /high resolution shell	99.8/96.2
$I/\sigma(I)$ /high resolution shell	23.4/3.4
Redundancy/high resolution shell	7.9/6.5
Mosaicity (°)	0.65
Unique reflections	72 026
Average <i>B</i> (Å <sup>2</sup> )	80
Solvent content (%)	74
Monomers in asymmetric unit	6
<b>Refinement statistics</b>	
NCS groups	3 (see text)
Twin target, fraction	( <i>h k l</i> ), 0.47 ( <i>k -l -h</i> ), 0.30 ( <i>-l -h k</i> ), 0.23
Amino acids per asymmetric unit	1686
Protein atoms per asymmetric unit	12 594
Root mean square deviation bonds (Å)/angles (°)	0.011/1.28
Ramachandran plot most favoured (residues/%)	1274/86.3
Ramachandran plot additional favoured (residues/%)	201/13.6
Ramachandran plot generously allowed (residues/%)	1/0.1
Ramachandran plot disallowed (residues/%)	0/0.0
<i>R</i> <sub>work</sub> (%) / <i>R</i> <sub>free</sub> (%) <sup>b</sup>	16.3/18.1

<sup>a</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)}$ .

<sup>b</sup> 2.9% of the data were excluded using Phenix [29] to calculate *R*<sub>free</sub>, taking NCS into account, but not the presence of twinning, which may have lowered reported *R*<sub>free</sub> values.

Data Bank (accession code 3FEM). All structural diagrams were drawn using PyMOL [31].

### 2.5. Biophysical and kinetic analysis

For analytical ultracentrifugation, ScPdx1.1His<sub>6</sub> was dialysed against 20 mM Tris–HCl pH 7.5, then supplemented with 0 mM, 200 mM, 1 M Na-formate, 200 mM NaH<sub>2</sub>PO<sub>4</sub> or 200 mM KCl. BsPdx1His<sub>6</sub>@K178 was assayed in 20 mM Tris–HCl, pH 7.5, containing 200 mM KCl. Sedimentation velocity runs using a Beckman Optima XL-A ultracentrifuge (Beckman Coulter, Fullerton, CA) were performed at 35 000 rpm, 20 °C,  $\lambda = 280$  nm, with  $\sim 80$   $\mu$ M protein. Programs SEDNTERP version 1.09 (University of New Hampshire), SEDFIT [32] and HYDROPRO [33] were used for data analysis (see legend to Table 2).

PLP synthase activity was determined as described in [9]. Purified proteins were generally dialyzed against 50 mM Tris–HCl, pH 8.0. PLP formation was monitored at a wavelength of 414 nm at 37 °C, using 40  $\mu$ M protein, 1 mM ribose 5'-phosphate and 2 mM DL-glyceraldehyde 3'-phosphate with 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as nitrogen source. The data are listed in Table 3 and were fitted using Michaelis–Menten kinetics [9].

## 3. Results and discussion

### 3.1. ScPdx1.1. is a hexamer in the 3D structure and in solution

The crystal structure of ScPdx1.1 has been determined by molecular replacement. The Matthews parameter suggested that one dodecamer could fit the asymmetric unit ( $V_M = 2.4$ ); however the final 3D structure of ScPdx1.1 shows only a hexamer per asymmetric unit ( $V_M = 4.7$ , Fig. 1a). The structure of Pdx1 from *B. subtilis* (BsPdx1) [20] used for molecular replacement crystallised in an orthorhombic space group with a single hexamer per AU; crystal symmetry then generates the BsPdx1 dodecamer. In fact, all other Pdx1 homologues determined by X-ray crystallography to date assume a dodecameric quaternary structure [20,21,23]. Surprisingly, the crystal packing for ScPdx1.1 is such that the hexamers do not pack into dodecamers (Fig. 1b). This leads to the observed high solvent content of 74% and probably is the reason for the low diffraction limit of 3.0 Å and the high overall *B* value of 80 Å<sup>2</sup> (Table 1).

ScPdx1.1 has a ( $\beta\alpha$ )<sub>8</sub>-barrel fold, and a number of deviations from the regular motif are present (Fig. 1a), like with other PLP synthases [20,21,23]. Firstly, nearly the full N-terminus is resolved in electron density, including  $\alpha$ -helix  $\alpha$ N, residues 4–17. This helix

**Table 2**  
Analysis of ScPdx1.1, BsPdx1 and BsPdx1@K178 proteins by analytical ultracentrifugation.<sup>a</sup>

Species	Tris–HCl, pH 7.5, mM	Salt	$\rho$ , g/cm <sup>3b</sup>	$\eta$ , cP <sup>b</sup>	$S_{\text{exp}}$ , S <sup>c</sup>	$S_{\text{calc}}$ , S <sup>d</sup>	$M_{\text{exp}}$ , kDa <sup>e</sup>	$M_{\text{calc}}$ , kDa <sup>f</sup>	$f/f_0$ <sup>g</sup>
(ScPdx1.1His <sub>6</sub> ) <sub>6</sub>	20	–	0.99880	0.01007	8.8 ± 0.1	8.2	185 ± 3	204	1.27
	20	0.2 M Na-formate	1.00740 <sup>h</sup>	0.01049 <sup>h</sup>	8.9 ± 0.1		190 ± 4		
	20	1.0 M Na-formate	1.03985 <sup>h</sup>	0.01212 <sup>h</sup>	8.6 ± 0.1		178 ± 6		
	50	–	0.99966	0.01017	8.5 <sup>i</sup>		175 <sup>i</sup>		
	50	0.2 M NaPO <sub>4</sub>	1.02775	0.01141	8.5 ± 0.2		171 ± 5		
	20	0.2 M KCl	1.00824	0.01005	8.6 ± 0.1		188 ± 3		
(BsPdx1His <sub>6</sub> ) <sub>6</sub>	20	0.2 M KCl	1.00824	0.01005	8.8 <sup>u</sup>	8.5	199 <sup>i</sup>	196	1.29
(BsPdx1His <sub>6</sub> @K178) <sub>6</sub>	20	0.2 M KCl	1.00824	0.01005	8.7 ± 0.1 <sup>j</sup>	8.5	179 ± 1 <sup>j</sup>	196	1.29
(BsPdx1His <sub>6</sub> ) <sub>12</sub>	20	0.2 M KCl	1.00824	0.01005	13.5 <sup>i</sup>	14.4	346 <sup>i</sup>	392	1.25
(BsPdx1His <sub>6</sub> @K178) <sub>12</sub>	20	0.2 M KCl	1.00824	0.01005	12.2 ± 0.1 <sup>j</sup>	14.4	299 ± 2 <sup>j</sup>	392	1.25

<sup>a</sup> Experimental values are means of at least three independent experiments (unless otherwise noted).

<sup>b</sup> Buffer density  $\rho$ , viscosity  $\eta$ , partial specific volumes  $\bar{v}$  and extinction coefficients calculated using the program SEDNTERP version 1.09 (University of New Hampshire).

<sup>c</sup> Sedimentation coefficients were determined under standard conditions (20 °C, H<sub>2</sub>O) from the  $c(s)$  distribution calculated from the raw velocity sedimentation data using SEDFIT [32].

<sup>d</sup> Calculated sedimentation coefficients using HYDROPRO and 3.1-Å bead size [33].

<sup>e</sup> Molecular mass derived from molar mass distributions  $c(M)$ .

<sup>f</sup> Calculated molecular mass.

<sup>g</sup> Ratio of the friction coefficient ( $f$ ) to the friction coefficient ( $f_0$ ) of a sphere. Translational diffusion coefficient and volume of the protein calculated with HYDROPRO [33].

<sup>h</sup> Buffer density  $\rho$  and viscosity  $\eta$  experimentally determined using a SVM 3000 Viscosimeter (Anton Paar, Ostfildern, Germany).

<sup>i</sup> Single measurement.

<sup>j</sup> Average of two measurements.

**Table 3**  
Kinetic analysis of ScPdx1.1, ScPdx1.1 $\Delta$ C<sub>276–297</sub> and BsPdx1 proteins.<sup>a</sup>

Protein	PLP synthase activity <sup>b</sup>		Ribose 5-phosphate <sup>c</sup>		Glyceraldehyde 3-phosphate <sup>d</sup>	
	Activity (nmol mg <sup>-1</sup> min <sup>-1</sup> )	Activity (%) <sup>e</sup>	$K_m$ (mM)	$v_{\text{max}}$ (nmol mg <sup>-1</sup> min <sup>-1</sup> )	$K_m$ (mM)	$v_{\text{max}}$ (nmol mg <sup>-1</sup> min <sup>-1</sup> )
ScPdx1.1	1.2 ± 0.14	100	0.11 ± 0.05	1.45 ± 0.01	0.30 ± 0.09	1.54 ± 0.12
ScPdx1.1 $\Delta$ C <sub>276–297</sub>	0.4 ± 0.09	33	0.15 ± 0.01	0.36 ± 0.03	0.37 ± 0.08	0.41 ± 0.04
BsPdx1	0.3 ± 0.07	25	0.068 ± 0.002 <sup>f</sup>	0.63 <sup>g</sup>	0.077 ± 0.002 <sup>f</sup>	n.d. <sup>h</sup>

<sup>a</sup> All values are means of at least three independent experiments.

<sup>b</sup> 40  $\mu$ M protein, 1 mM ribose 5'-phosphate, 2 mM DL-glyceraldehyde 3'-phosphate, 10 mM ammonium sulfate, 50 mM Tris–HCl, pH 8.0, 37 °C.

<sup>c</sup> As (b), but varying the concentration of ribose 5'-phosphate.

<sup>d</sup> As (b), but varying the concentration of glyceraldehyde 3'-phosphate.

<sup>e</sup> Setting ScPdx1.1 activity as reference 100%.

<sup>f</sup> See Ref. [9].

<sup>g</sup> Calculated from the value  $k_{\text{cat}} = 0.02$  min<sup>-1</sup> given in Ref. [9], where  $V_{\text{max}} = k_{\text{cat}}E_t$  with total enzyme concentration  $E_t = 40$   $\mu$ M.

<sup>h</sup> n.d., not determined.

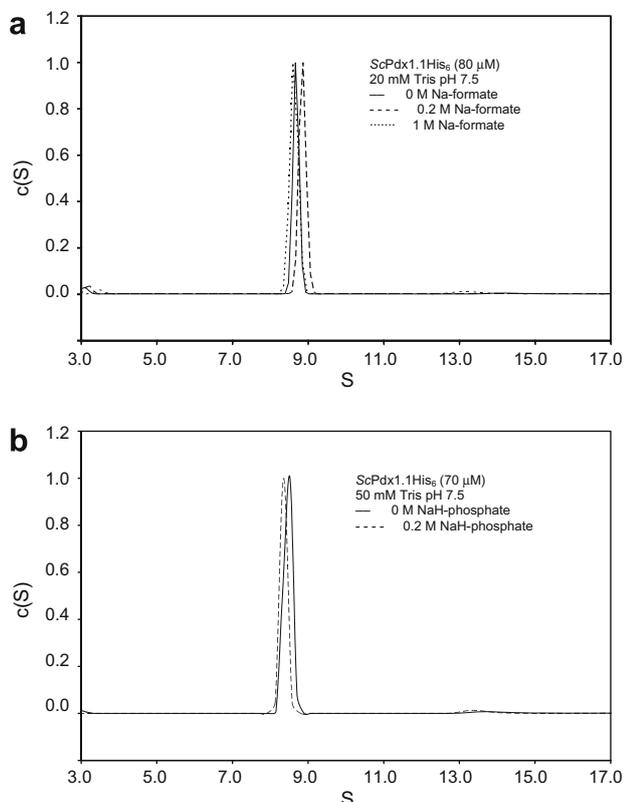
has been shown to be indispensable for complex formation with the Pdx2 subunit in *B. subtilis* and *Plasmodium falciparum* [20,34,35]. An ordered  $\alpha$ N helix was previously only observed in Pdx1/2 complexes [20,23], not the isolated Pdx1 subunits [20,21].

Secondly, helix  $\alpha 2'$  inserted between  $\beta 1$  and  $\alpha 2$  is implied in enzyme activation ([20], discussed in Section 3.3). Thirdly, helix  $\alpha 8''$  inserted after helix  $\alpha 8$  forms the hexamer interface. Finally, elongation of  $\alpha$ -helix  $\alpha 6$  and insertion of helices  $\alpha 6'$  and  $\alpha 6''$  are required for dodecamer formation [20,21,23].

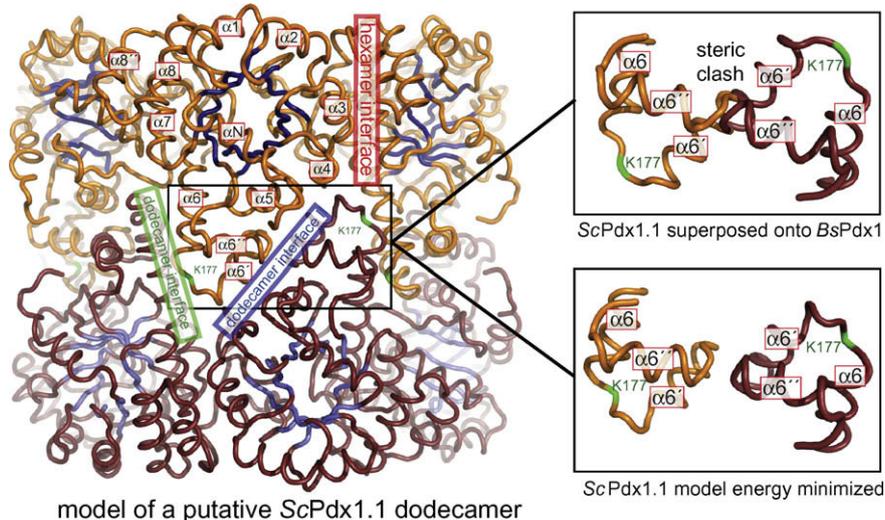
It is puzzling that the yeast homologues possess the latter insertions while “loosing” the apparent propensity to oligomerise into dodecamers. It was thus imperative to investigate by an in-solution technique whether the high salt concentration used in crystallisation (4 M sodium formate) had interfered with dodecamer formation. Protein sedimentation behaviour was analyzed by analytical ultracentrifugation (previously described for *BsPdx1*, [20]). ScPdx1.1 in Tris-buffer at pH 7.5 with varied salt strength, 0, 0.2 and 1 M sodium formate has average sedimentation coefficients normalised to standard conditions (20 °C, water) of  $8.8 \pm 0.1$ ,  $8.9 \pm 0.1$  and  $8.6 \pm 0.1$  S (Fig. 2a and Table 2). The data are in keeping with earlier experimental data using non-denaturing gradient polyacrylamide gel electrophoresis that determined a molecular weight of 230 kDa for the ScPdx1.1 complex, consistent with a hexamer [17]. Minor peaks are observed at approximately 3 S, possibly indicating presence of monomeric proteins, and around 13.6–14.8 S, suggesting ScPdx1.1 can convert into a higher molecular species.

Buffer conditions more favourable for oligomerisation of ScPdx1.1 into dodecamers might exist. For example, for Pdx1 from *G. stearothersophilus* the hexamer:dodecamer equilibrium shifts toward the dodecamer in the presence of 200 mM phosphate [21], but the enzyme is almost exclusively hexameric in the absence of phosphate ions. In the presence of 200 mM phosphate ions ScPdx1.1 is hexameric with a sedimentation coefficient of  $8.5 \pm 0.2$  S (Fig. 2b and Table 2).

The observed differences in quaternary structure may have an effect on catalytic activity. Interestingly, we found that ScPdx1.1 has 4-fold higher specific catalytic activity than *BsPdx1*, employing a glutaminase independent enzyme assay (Table 3). This could suggest that dodecamer formation is not a prerequisite to enzymatic activity. However, transient interaction of hexamers during PLP biosynthesis cannot be excluded. The nearly cubic unit cell can orient in three possible ways, leading to an unusual case of pseudo-merohedral twinning, described in Section 2 and Supplementary data.



**Fig. 2.** Oligomerization behaviour of ScPdx1.1 proteins in solution, as analyzed by analytical ultracentrifugation under various conditions: (a) ScPdx1.1 with 0 M (solid line), 0.2 M (dashed line) and 1 M sodium formate (dotted line). (b) ScPdx1.1 in the absence of sodium phosphate (solid line) and in the presence of 0.2 M sodium phosphate (dashed line). Compare Table 2.



**Fig. 3.** *In silico* modeling of a ScPdx1.1 dodecamer generated by superposition of two ScPdx1.1 hexamers onto the *BsPdx1* dodecamer (PDB code 2NV1 [20]) and subsequent energy minimization using CNS [36]. One ScPdx1.1 hexamer is shown in orange-blue, the second one in purple-blue, the position of Lys177 is highlighted in green.  $\alpha$ -Helices have been annotated for one ScPdx1.1 subunit. Hexamer and dodecamer interfaces are marked. The zoom on the right-hand side shows the dodecamer interface before and after energy minimization. In the energy minimized model, clashes between helices  $\alpha 6'$  and  $\alpha 6''$  of two monomers resulting from the superposition are avoided.

### 3.2. Disruption of the dodecameric structure of PLP synthase

Structural elements required for dodecamer formation are present in ScPdx1.1, despite its hexameric nature. We superposed the hexameric ScPdx1.1 onto two hexameric halves of the BsPdx1 dodecamer to understand which structural properties hinder oligomerisation (Fig. 3). This resulted in overlapping residues and main chain clashes, as depicted in Fig. 3 (upper inset). However, dodecamer formation might induce or coincide with structural rearrangements. Manual adjustment of helices  $\alpha 6$ ,  $\alpha 6'$  and  $\alpha 6''$  to obviate clashes followed by energy minimization using CNS [36] leads to the altered structure depicted in Fig. 3 (lower inset). From this exercise and from analytical ultracentrifugation we do not rule out that ScPdx1.1 hexamers may form into dodecamers, but restructuring of helices  $\alpha 6$ ,  $\alpha 6'$  and  $\alpha 6''$  would have to be accounted for in terms of energy.

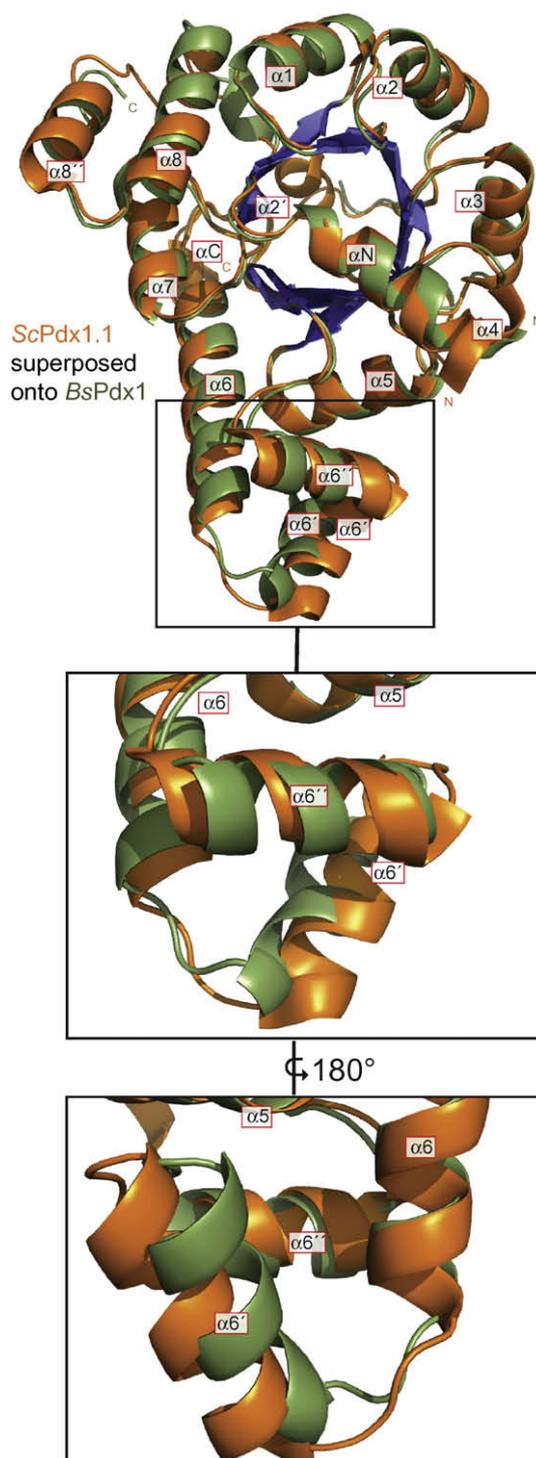
From comparison of ScPdx1.1 and BsPdx1 3D-structures (Fig. 4) differences are seen at the C-terminal end of helix  $\alpha 6$  and helices  $\alpha 6'$  and  $\alpha 6''$ . A multiple sequence alignment shows an insertion of a single lysine residue in this region in ScPdx1.1 (marked green in Fig. 3) and insertion of an asparagine and a lysine in ScPdx1.2 and ScPdx1.3 (not shown). The insertion in the yeast proteins is unique and not found in bacterial and plant homologues.

Simple deletion of Lys177 in ScPdx1.1 did not convert the hexamer into dodecamer in analytical ultracentrifugation (data not shown). Therefore the Lys177 insertion is not the sole factor determining the quaternary organisation of PLP synthase. The reverse experiment on a dodecameric enzyme allows quantifying the interference of the lysine insertion with dodecamer formation. For this purpose we inserted a lysine in an equivalent position into BsPdx1, generating BsPdx1His<sub>6</sub>@K178 that can be heterologously expressed in *E. coli* with similar yields, stability and enzymatic activity as wild-type protein (data not shown). Size exclusion chromatography indicated conversion of BsPdx1 dodecamers into hexameric species (data not shown).

Analytical ultracentrifugation of BsPdx1wt and BsPdx1His<sub>6</sub>@K178 (Fig. 5) showed that the hexamer:dodecamer ratio changes, and the dissociation constant of dodecamer is increased from 4.2  $\mu$ M for the wild-type protein to 16.9  $\mu$ M for the insertion variant. The hexameric species have similar *S*-values with  $8.7 \pm 0.1$  S for BsPdx1His<sub>6</sub>@K178 and 8.8 S for BsPdx1 (Table 2). In contrast, the dodecameric species differ from each other and show *S*-values of 13.5 S for BsPdx1 and of  $12.2 \pm 0.1$  S for BsPdx1His<sub>6</sub>@K178. In sedimentation velocity runs, a shift towards a lower sedimentation coefficient can result from a destabilization and partial dissociation of the higher molecular weight oligomer, or from a change in shape leading to an increased friction coefficient. While hexameric and dodecameric species are well separated, we conclude that insertion of the single lysine leads to a conformational change that is not fully compatible with the dodecamer, consequently changing the quaternary structure of BsPdx1. Thus, we predict that presence of the insertion largely determines the quaternary structure of PLP synthases; sequence comparison suggests that a number of fungal proteins sharing insertions (e.g. *Ashbya*, *Apergillus*, *Candida*, *Debaryomyces*, *Kluyveromyces*, *Lodderomyces*, *Pichia*, and *Vanderwaltozyma* sp.) all are (mainly) hexameric in nature.

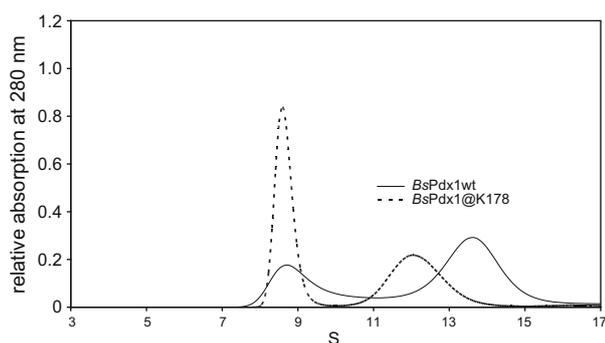
### 3.3. Involvement of the C-terminus in catalysis

A large part of the C-terminus up to residue 283 is ordered in the structure of ScPdx1.1. The structure resolves an additional 15 residues compared with BsPdx1, and four further amino acids when compared with the *T. maritima* Pdx1/Pdx2 structure [23].



**Fig. 4.** Superposition of the 3D-structures of one ScPdx1.1 monomer, orange and blue, with one BsPdx1 monomer (PDB code 2NV1, [20]), green and blue.  $\alpha$ -Helices have been annotated; the zoom shows differences in the region of helices  $\alpha 6$ ,  $\alpha 6'$  and  $\alpha 6''$ .

The C-terminus is positioned in the interface between two ScPdx1.1 monomers of a hexameric ring in contact distance to the loop between  $\alpha 2'$  and  $\alpha 2$  of an adjacent monomer, i.e. in close proximity to the catalytic centre. This segment is mostly extended, leading into an unpredicted 3–10  $\alpha$ -helix, named here  $\alpha C$ . Like  $\alpha 2'$ ,  $\alpha C$  points into the centre of the ScPdx1.1 hexamer, as is  $\alpha 2'$  (Fig. 1a). The helical segment  $\alpha 2'$  itself is observed in the ScPdx1.1



**Fig. 5.** Oligomerisation behaviour of wild-type BsPdx1 (solid line) and of BsPdx1His<sub>6</sub>@K178 variant protein (dashed line), investigated by analytical ultracentrifugation. Compare Table 2.

structure but was previously only resolved when Pdx1 was in complex with Pdx2 [20,23]. Thus, formation of  $\alpha 2'$  on the C-terminal face of the ( $\beta\alpha$ )<sub>8</sub>-barrel was described as a priming of the synthase for catalysis [20]. Observation of this helix possibly explains why ScPdx1.1 has higher PLP synthase activity than BsPdx1, determined in the absence of glutaminase (with ammonium ions as nitrogen donor, Table 3). These observations suggest that stabilisation of the C-terminus and closing off the active site by helix  $\alpha 2'$  and the C-terminus is necessary for catalysis.

We constructed two C-terminal deletion mutants to understand the involvement of the C-terminal segment in catalysis. While the shorter of the two deletion variants, ScPdx1.1 $\Delta$ C<sub>268–297</sub>, was exclusively found in inclusion bodies after heterologous expression in *E. coli* BL21 (DE3), ScPdx1.1 $\Delta$ C<sub>276–297</sub> could be purified in soluble form. No change with respect to oligomerisation properties has been detected during purification using size exclusion chromatography (not shown). However, PLP synthase activity of ScPdx1.1 $\Delta$ C<sub>276–297</sub> is 3-fold lower compared to wild-type ScPdx1.1, and comparable to BsPdx1 using the same glutaminase independent assay (Table 3). Variation of substrate concentration to determine Michaelis–Menten kinetic parameters shows that ScPdx1.1 and ScPdx1.1 $\Delta$ C<sub>276–297</sub> have similar  $K_m$  values, but  $V_{max}$  drops 3.5-fold in the truncation variant. This indicates that the C-terminus of ScPdx1.1 influences enzyme activity, but is not directly involved in substrate binding. Since the C-terminus of one monomer is interacting with a second monomer in the hexamer, hexamer formation has a role in catalysis in ScPdx1.1 and possibly in Pdx1 synthases in general. It will be interesting to investigate whether ScPdx1 dodecamer formation represents another regulatory layer to PLP synthase activity, possibly primed by interaction with ScPdx2, and we have now initiated a complete characterization of PLP production by ScPdx1.1 in complex with ScPdx2.1.

#### Note added in proof

While this manuscript was under review, another report on the involvement of the C-terminus in catalysis was published (J. Biol. Chem. 284 (2009) 7706–7718).

#### Acknowledgements

We thank staff at the ESRF synchrotron for help with data collection and Julia Hupfeld for experimental support. This work was supported by the European Commission to P.M. and I.T. (VIT-BIOMAL-012158), by the Fonds zur Förderung der wissenschaftli-

chen Forschung (FWF) to P.M. (P17215), and by the DFG to I.T. (TE368/4-1).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2009.06.009.

#### References

- [1] Eliot, A.C. and Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* 73, 383–415.
- [2] Bender, D.A. (1989) Vitamin B6 requirements and recommendations. *Eur. J. Clin. Nutr.* 43, 289–309.
- [3] Ehrenshaft, M., Bilski, P., Li, M.Y., Chignell, C.F. and Daub, M.E. (1999) A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proc. Natl. Acad. Sci. USA* 96, 9374–9378.
- [4] Osmani, A.H., May, G.S. and Osmani, S.A. (1999) The extremely conserved pyroA gene of *Aspergillus nidulans* is required for pyridoxine synthesis and is required indirectly for resistance to photosensitizers. *J. Biol. Chem.* 274, 23565–23569.
- [5] Fitzpatrick, T.B., Amrhein, N., Kappes, B., Macheroux, P., Tews, I. and Raschle, T. (2007) Two independent routes of de novo vitamin B6 biosynthesis: not that different after all. *Biochem. J.* 407, 1–13.
- [6] Mittenhuber, G. (2001) Phylogenetic analyses and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *J. Mol. Microbiol. Biotechnol.* 3, 1–20.
- [7] Zalkin, H. and Smith, J.L. (1998) Enzymes utilizing glutamine as an amide donor. *Adv. Enzymol. Relat. Areas Mol. Biol.* 72, 87–144.
- [8] Burns, K.E., Xiang, Y., Kinsland, C.L., McLafferty, F.W. and Begley, T.P. (2005) Reconstitution and biochemical characterization of a new pyridoxal-5'-phosphate biosynthetic pathway. *J. Am. Chem. Soc.* 127, 3682–3683.
- [9] Raschle, T., Amrhein, N. and Fitzpatrick, T.B. (2005) On the two components of pyridoxal 5'-phosphate synthase from *Bacillus subtilis*. *J. Biol. Chem.* 280, 32291–32300.
- [10] Tazuya, K., Adachi, Y., Masuda, K., Yamada, K. and Kumaoka, H. (1995) Origin of the nitrogen atom of pyridoxine in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 1244, 113–116.
- [11] Hanes, J.W., Keresztes, I. and Begley, T.P. (2008) 13C NMR snapshots of the complex reaction coordinate of pyridoxal phosphate synthase. *Nat. Chem. Biol.* 4, 425–430.
- [12] Sivasubramanian, S., Vanniasingham, V.M., Tan, C.T. and Chua, N.H. (1995) Characterisation of HEVER, a novel stress-induced gene from *Hevea brasiliensis*. *Plant Mol. Biol.* 29, 173–178.
- [13] Krasko, A., Schroder, H.C., Perovic, S., Steffen, R., Kruse, M., Reichert, W., Muller, I.M. and Muller, W.E. (1999) Ethylene modulates gene expression in cells of the marine sponge *Suberites domuncula* and reduces the degree of apoptosis. *J. Biol. Chem.* 274, 31524–31530.
- [14] Ehrenshaft, M., Jenns, A.E., Chung, K.R. and Daub, M.E. (1998) SOR1, a gene required for photosensitizer and singlet oxygen resistance in *Cercospora* fungi, is highly conserved in divergent organisms. *Mol. Cell.* 1, 603–609.
- [15] Craig, J.P., Bekal, S., Hudson, M., Domier, L., Niblack, T. and Lambert, K.N. (2008) Analysis of a horizontally transferred pathway involved in vitamin B6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. *Mol. Biol. Evol.* 25, 2085–2098.
- [16] Fuge, E.K., Braun, E.L. and Werner-Washburne, M. (1994) Protein synthesis in long-term stationary-phase cultures of *Saccharomyces cerevisiae*. *J. Bacteriol.* 176, 5802–5813.
- [17] Padilla, P.A., Fuge, E.K., Crawford, M.E., Errett, A. and Werner-Washburne, M. (1998) The highly conserved, coregulated SNO and SNZ gene families in *Saccharomyces cerevisiae* respond to nutrient limitation. *J. Bacteriol.* 180, 5718–5726.
- [18] Rodriguez-Navarro, S. et al. (2002) Functional analysis of yeast gene families involved in metabolism of vitamins B1 and B6. *Yeast* 19, 1261–1276.
- [19] Nishizawa, M., Komai, T., Morohashi, N., Shimizu, M. and Toh-e, A. (2008) Transcriptional repression by the Pho4 transcription factor controls the timing of SNZ1 expression. *Eukaryot. Cell* 7, 949–957.
- [20] Strohmeier, M., Raschle, T., Mazurkiewicz, J., Rippe, K., Sinning, I., Fitzpatrick, T.B. and Tews, I. (2006) Structure of a bacterial pyridoxal 5'-phosphate synthase complex. *Proc. Natl. Acad. Sci. USA* 103, 19284–19289.
- [21] Zhu, J., Burgner, J.W., Harms, E., Belitsky, B.R. and Smith, J.L. (2005) A new arrangement of ( $\beta$ / $\alpha$ )<sub>8</sub> barrels in the synthase subunit of PLP synthase. *J. Biol. Chem.* 280, 27914–27923.
- [22] Sterner, R. and Hocker, B. (2005) Catalytic versatility, stability, and evolution of the ( $\beta$ / $\alpha$ )<sub>8</sub>-barrel enzyme fold. *Chem. Rev.* 105, 4038–4055.
- [23] Zein, F., Zhang, Y., Kang, Y.N., Burns, K., Begley, T.P. and Ealick, S.E. (2006) Structural insights into the mechanism of the PLP synthase holoenzyme from *Thermotoga maritima*. *Biochemistry* 45, 14609–14620.
- [24] Hoffmann, C. S. (1997) Preparation of yeast DNA, RNA, and proteins. In F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl (Eds.), *Current Protocols in Molecular Biology* (pp. 13.11.1–13.11.4). New York: John Wiley & Sons.

- [25] Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.
- [26] Vagin, A.A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* 30, 1022–1025.
- [27] Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- [28] Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.
- [29] Adams, P.D. et al. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1948–1954.
- [30] Zwart, P.H., Grosse-Kunstleve, R.W., Lebedev, A.A., Murshudov, G.N. and Adams, P.D. (2008) Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr. D Biol. Crystallogr.* 64, 99–107.
- [31] DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA. 2002.
- [32] Dam, J. and Schuck, P. (2004) Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. *Meth. Enzymol.* 384, 185–212.
- [33] Garcia De La Torre, J., Huertas, M.L. and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78, 719–730.
- [34] Flicker, K., Neuwirth, M., Strohmeier, M., Kappes, B., Tews, I. and Macheroux, P. (2007) Structural and thermodynamic insights into the assembly of the heteromeric pyridoxal phosphate synthase from *Plasmodium falciparum*. *J. Mol. Biol.* 374, 732–748.
- [35] Neuwirth, M., Flicker, K., Strohmeier, M., Tews, I. and Macheroux, P. (2007) Thermodynamic characterization of the protein–protein interaction in the heteromeric *Bacillus subtilis* pyridoxal phosphate synthase. *Biochemistry* 46, 5131–5139.
- [36] Brunger, A.T. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.

## Supplement to

# X-Ray Crystal Structure of *Saccharomyces cerevisiae* Pdx1 Provides Insights into the Oligomeric Nature of PLP Synthases

Martina Neuwirth<sup>a,1</sup>, Marco Strohmeier<sup>b,1</sup>, Volker Windeisen<sup>b</sup>, Silvia Wallner<sup>a</sup>, Sigrid Deller<sup>a</sup>, Karsten Rippe<sup>c</sup>, Irmgard Sinning<sup>b</sup>, Peter Macheroux<sup>a,\*</sup>, Ivo Tews<sup>b,\*</sup>

- a. Technische Universität Graz, Institut für Biochemie, Petersgasse 12/2, A-8010 Graz, Austria
- b. Heidelberg University Biochemistry Center (BZH), Im Neuenheimer Feld 328, 69120 Heidelberg, Germany
- c. Deutsches Krebsforschungszentrum and BIOQUANT, Research Group Genome Organization & Function, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

1. These authors contributed equally to this work.

\* Corresponding authors. Fax: +49-6221-5447-90; *E-mail address*: [ivo.tews@bzh.uni-heidelberg.de](mailto:ivo.tews@bzh.uni-heidelberg.de) (Ivo Tews). Fax: +43-316-8736-952; *E-mail address*: [peter.macheroux@tugraz.at](mailto:peter.macheroux@tugraz.at) (Peter Macheroux).

This supplement includes:  
Description of Crystal Pathologies  
Supplementary Figures S1 to S5  
Supplementary Tables S1 to S4

## Description of Crystal Pathologies

Data to 3.02 Å were collected at a wavelength of 0.931 Å at beamline ID14-3 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Autoindexing using denzo (HKL Research Inc.) indicated possible cubic symmetry (Table S1).

```

Volume of the primitive cell  3631656.

Lattice           Metric tensor      Best cell (symmetrized)
                   distortion index    Best cell (without symmetry restrains)

primitive cubic      0.10% 153.61 153.78 153.74  89.87  89.81  89.99
                   153.71 153.71 153.71  90.00  90.00  90.00

I centred cubic     20.32% 217.37 216.97 217.21 119.92 119.88  60.16
                   217.18 217.18 217.18  90.00  90.00  90.00

F centred cubic     20.32% 266.32 265.76 266.13 109.35 109.45 109.47
                   266.07 266.07 266.07  90.00  90.00  90.00

primitive rhombohedral 0.06% 153.78 153.74 153.61  89.81  89.99  89.87
                   153.71 153.71 153.71  89.89  89.89  89.89
                   217.28 217.28 266.73  90.00  90.00 120.00

primitive hexagonal 13.53% 153.74 153.61 153.78  90.01  89.87  90.19
                   153.68 153.68 153.78  90.00  90.00 120.00

primitive tetragonal 0.09% 153.74 153.78 153.61  89.99  90.19  90.13
                   153.76 153.76 153.61  90.00  90.00  90.00

I centred tetragonal 16.54% 265.76 217.35 153.74  90.04 125.12  89.92
                   241.55 241.55 153.74  90.00  90.00  90.00

primitive orthorhombic 0.09% 153.61 153.74 153.78  89.87  90.01  90.19
                   153.61 153.74 153.78  90.00  90.00  90.00

C centred orthorhombic 0.06% 216.97 217.69 153.78  89.91  89.92  89.95
                   216.97 217.69 153.78  90.00  90.00  90.00

I centred orthorhombic 16.57% 153.61 153.78 375.87  65.96 113.96  90.01
                   153.61 153.78 375.87  90.00  90.00  90.00

F centred orthorhombic 16.57% 217.35 217.37 375.87  54.89  89.93  89.94
                   217.35 217.37 375.87  90.00  90.00  90.00

primitive monoclinic 0.05% 153.61 153.78 153.74  90.13  90.19  89.99
                   153.61 153.78 153.74  90.00  90.19  90.00

C centred monoclinic 0.03% 217.37 217.35 153.74  89.96  90.22  89.94
                   217.37 217.35 153.74  90.00  90.22  90.00

primitive triclinic 0.00% 153.61 153.74 153.78  89.87  89.99  89.81

autoindex unit cell 153.61 153.61 153.61  90.00  90.00  90.00

```

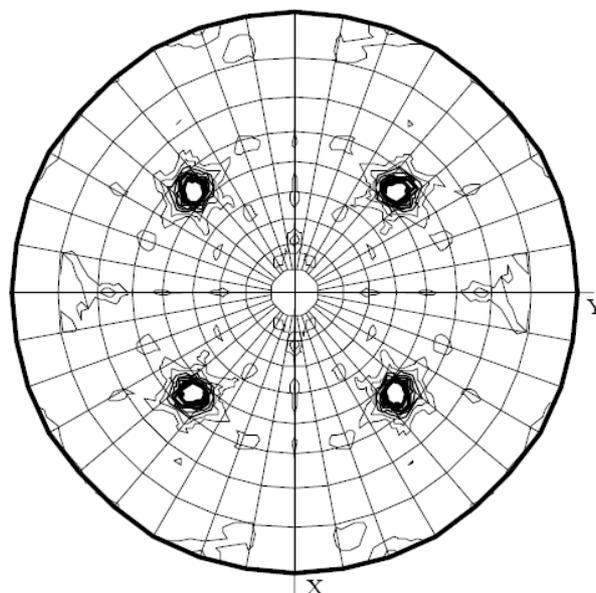
**Table S1: Autoindexing table, produced by denzo (HKL Research Inc.).**

Data reduction with denzo/scalepack (HKL Research Inc.) indicated that only the orthorhombic cell merges with proper statistics (Table S2).

Symmetry	Cell axes in Å	R <sup>1</sup> overall (LR bin <sup>2</sup> / HR bin <sup>3</sup> )
P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	a=153.646 b=153.671 c=153.620	7.5% ( 3.5% / 41.6%)
P4, l unique	a=b=153.623 c=153.676	31.4% (33.5% / 51.9%)
P4, k unique	a=b=153.660 c=153.619	28.8% (26.7% / 52.1%)
P4, h unique	a=b=153.648 c=153.656	29.4% (23.6% / 52.4%)
R3	a=b=217.221 c=266.257	16.9% (14.1% / 45.3%)
P2 <sub>1</sub> 3	a=b=c=153.656	17.0% (12.3% / 48.1%)

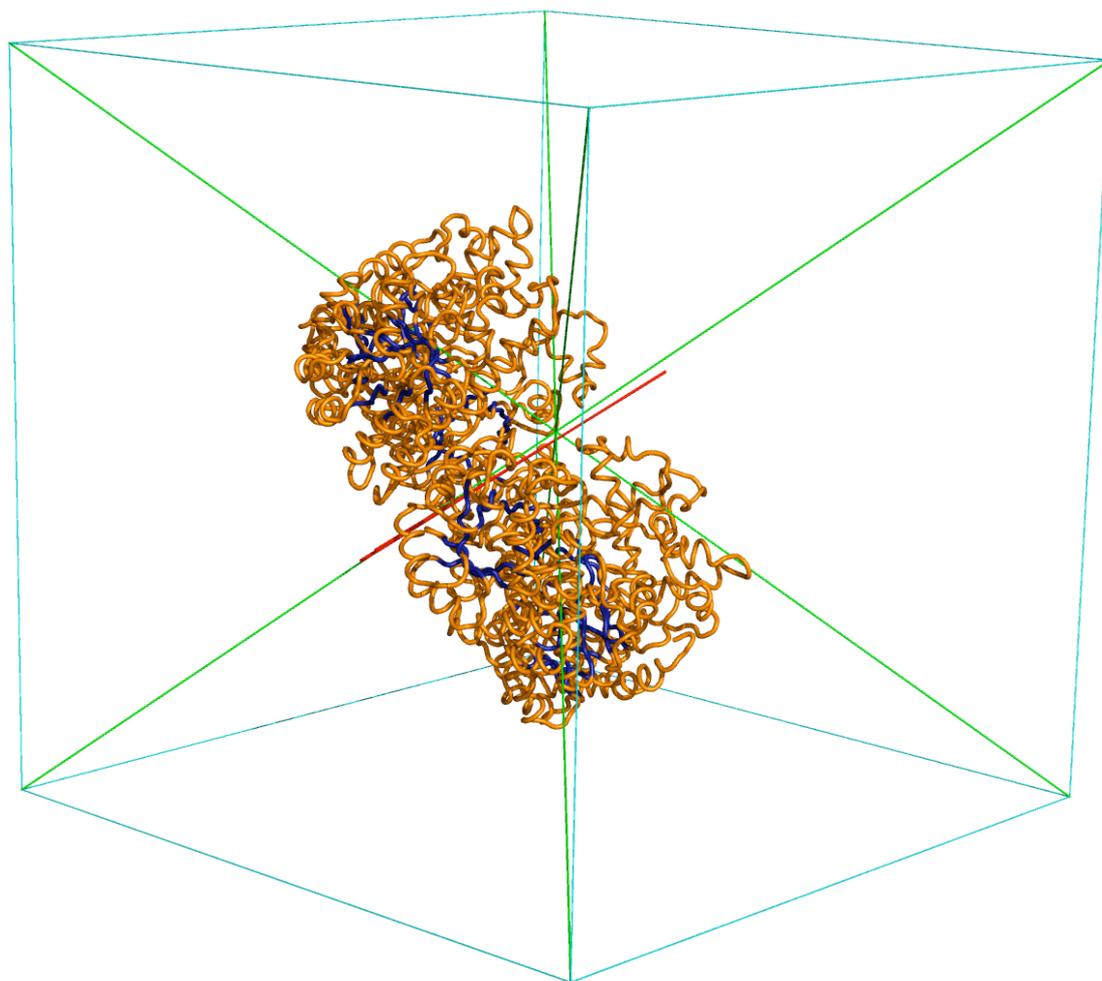
**Table S2: Data scaling using denzo/scalepack (HKL Research Inc.); 1 Given is  $R_{linear} = \Sigma(ABS(I - \langle I \rangle)) / \Sigma(I)$ ; 2 LR bin = low resolution bin 9.90Å–20Å, except for 8.03Å–20Å in P213; 3 HR bin = high resolution bin 3.02Å–3.05Å, except for 3.02Å–3.07 Å in P213**

Structure determination using molecular replacement with MOLREP [1] and data reduced in the orthorhombic P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> cell and subsequent refinement with Refmac5 [2] were carried out as described in the main text. We then tested for presence of higher symmetry and possible twinning. The self rotation function of the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> data is seen in Fig. S1. We repeated molecular replacement in all possible space-groups (except P4, because of poor merging statistics), using the partially refined coordinates as starting model.



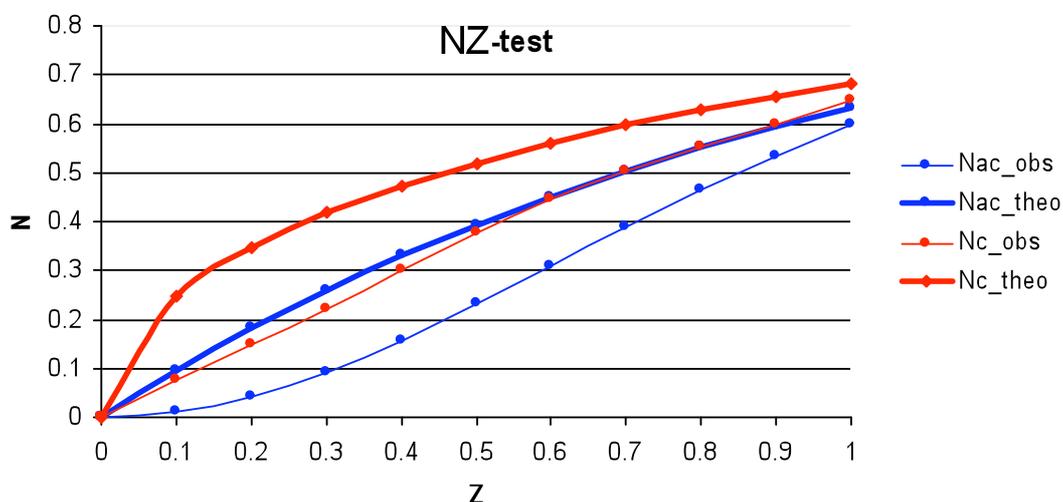
**Fig. S1: Self rotation function for space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> at chi=120 (using MOLREP [1]). The origin peak has a height of 19.3 sigma, the features seen at theta=54.74 and phi=45 or phi=135 have respective heights of 17.6 sigma.**

Refinement with Refmac5 [2] in  $P2_12_12_1$  gave an  $R_{\text{free}}=0.330$  and a figure of merit (FOM) of 0.754. Refinement statistics were much worse for space groups containing a 3-fold axis: in  $R3$ , an  $R_{\text{free}}=0.415$  and a FOM=0.553 resulted, while refinement in  $P2_13$  gave  $R_{\text{free}}=0.423$  and a FOM=0.593. This suggested the 3-fold axis is non-crystallographic. Using the programs LSQKAB [3] and DynDom ([www.cmp.uea.ac.uk/dyndom/](http://www.cmp.uea.ac.uk/dyndom/), [4]), we found that the rotation axis of the particle is indeed offset from the body diagonal of the unit cell (Fig. S2) – a strict coincidence would be required for the  $P2_13$  symmetry. Analysis of the Pdx1 six-fold also demonstrates a deformation of the hexameric particle, with a slight offset of about  $60^\circ \pm 0.5^\circ$ , comparing superimposition of individual monomers onto each other around the particle rotation axis; this is probably owing to crystal packing.

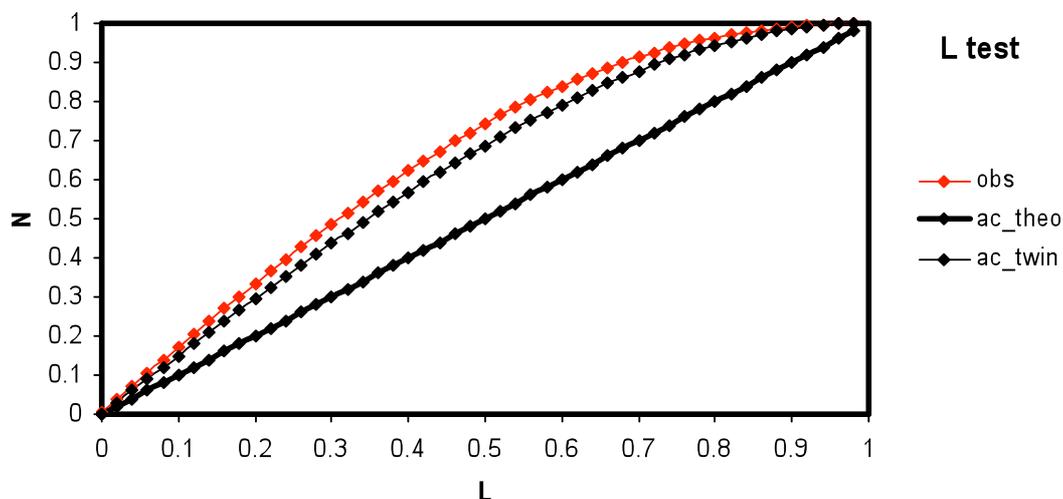


**Fig. S2:** Crystal packing analysis in the  $P2_12_12_1$  cell – cell axes are in blue, body diagonals in green; one hexameric particle is shown (compare Figure 3). The rotation axis of the particle is indicated in red. Drawn with pymol [5].

Thus, this is a case where a non crystallographic symmetry operator (NCS) is close to, but does not exactly match, a crystallographic axis. This situation can be described as pseudo-symmetry [6]. In this particular case, the symmetry then breaks down to an orthorhombic  $P2_12_12_1$  cell with a metric of three similar axes that is highly suspicious to pseudo-merohedral twinning. Analysis was carried out with Xtrriage from the Phenix suite [6]. Initial checks on data integrity showed highly complete data to 3.09 Å (**Error! Reference source not found.**). In the resolution range from 20-5Å only four reflexes were missing, there were no detectable ice ring problems, the data have low anisotropy, and no outliers in either centric or acentric reflections were detected.



*Fig. S3: N(Z)-test. Theoretical and observed values are plotted for centric and acentric reflections. Statistics generated by Xtrriage [6]*



**Fig. S4: L-test. Theoretical values are plotted in black for untwined (bold) and twinned cases, together with observed values in red. Statistics generated by Xtrriage [6]**

Twinning analysis carried out in the resolution range from 10 Å to 3.09 Å showed that intensity distributions differ significantly from the theoretical expected values, as seen in the Wilson ratio, the N(Z) test (Fig. S3), or the L-Test (Fig. S4). No pseudo-translational symmetry was detected. Xtrriage [6] reported five pseudo-merohedral twin laws, as listed in Table S3:

<b>twin law</b>	<b>axis</b>	<b>R<sub>metric</sub>(%)</b>	<b>H-test<sup>(1)</sup></b>	<b>Britton <math>\alpha</math></b>	<b>R vs R<sup>(2)</sup></b>	<b>ML<sup>(3)</sup></b>
<b>k,h,-l</b>	2-fold	0.022	0.169 / 0.149	0.166	0.347 / 0.182	0.022 / 0.027
<b>-h,-l,-k</b>	2-fold	0.044	0.167 / 0.149	0.166	0.347 / 0.182	0.022 / 0.024
<b>l,-k,h</b>	2-fold	0.023	0.167 / 0.149	0.166	0.348 / 0.183	0.022 / 0.024
<b>k,l,h</b>	3-fold	0.044	0.354 / 0.348	0.359	0.152 / 0.033	0.346 / 0.297
<b>l,h,k</b>	3-fold	0.044	0.354 / 0.348	0.345	0.152 / 0.033	0.326 / 0.293

**Table S3: Suspected twin laws of the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> cell on data statistics, as reported by Xtrriage [6]. <sup>(1)</sup> First number given: Estimation of twin fraction via mean |H|; second number given: Estimation of twin fraction via cumulative distance of H; <sup>(2)</sup> First number given: R<sub>abs\_twin</sub> observed data; second number given R<sub>sq\_twin</sub> observed data; <sup>(3)</sup> First number given: estimated twin fraction; second number: same, but taking NCS into account.**

It would appear that the crystals are composed of randomly oriented unit cells, and that the three principal orientations are responsible for the apparent twin operators (h,k,l), (k,l,h) and (l,h,k), which might be represented in an equal ratio of about  $\square / \square / \square$ . We then carried out twin refinement to investigate this case. Refinement with phenix.refine [7] gave starting R / R<sub>free</sub> values of 38.59 % / 37.87 %, and final R / R<sub>free</sub> values of 26.54 % / 32.43 %. Refinement against a single twin domain would usually give about seven to eight percent lower R values. For the twin operator (k-l-h), a twin fraction  $\alpha$  of 0.42 was determined, and starting R / R<sub>free</sub> values were 29.86 % / 29.42 %, and final R / R<sub>free</sub> values of 18.14 % / 23.87 %. For the twin operator (-l-hk), a twin fraction  $\alpha$  of 0.39 was determined, and starting R / R<sub>free</sub> values were 30.82 % / 30.34 %, and final R / R<sub>free</sub> values of 20.06 % / 25.60 %. Since refinement against two twin domains is currently not possible in phenix.refine, the automated refinement and twin determination in reftmac5 was used, which assigned three twin domains with respective fractions of 0.47, 0.298 and 0.231, and gave much improved refinement

statistics, as seen from Table S4. Refmac refinement used a two step protocol, where seven cycles of conjugant gradient refinement (CG) were carried out with a matrix parameter of 0.1, followed by combined CG / TLS-tensor refinement (translation, libration and screw-rotation displacements) with five cycles of TLS refinement and seven cycles of CG refinement, this time using very restrained geometry with a matrix parameter of 0.02. NCS definitions as given in the main text apply in all these refinement scenarios.

twin law (twin fraction)		R (%)	R <sub>Free</sub> (%)	FOM	rmsBOND	rmsANGL
hkl 0.470 k-l-h 0.298 -l-h-k 0.231	refinement start	22.35	23.39	0.700	0.0464	3.830
	refinement final	16.22	18.10	0.826	0.0107	1.284
none	refinement start	36.05	36.25	0.675	0.0464	3.830
	refinement final	31.11	32.56	0.737	0.0121	1.350

**Table S4: Refinement in Refmac5. Starting and final values are given for R/R<sub>Free</sub> figures of merit (FOM), and bond and angle deviations.**

## Acknowledgements

We thank Klemens Wild for helpful discussions. Advice by Peter Zwart and Garib Murshudov is gratefully acknowledged.

## References

1. Vagin AA & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography* **30**, 1022-1025.
2. Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica* **53**, 240-255.
3. Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* **32**, 922-923.
4. Hayward S & Berendsen HJ (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144-154.
5. DeLano WL (2002) The PyMOL Molecular Graphics System. *DeLano Scientific, San Carlos, California, USA, 2002*.
6. Zwart PH, Grosse-Kunstleve RW, Lebedev AA, Murshudov GN & Adams PD (2008) Surprises and pitfalls arising from (pseudo)symmetry. *Acta crystallographica* **64**, 99-107.
7. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK & Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica* **58**, 1948-1954.