Supplement to

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

Martina Neuwirth[a,1], Marco Strohmeier[b,1], Volker Windeisen[b], Silvia Wallner[a], Sigrid Deller[a], Karsten Rippe[c], Irmgard Sinning[b], Peter Macheroux[a,*], Ivo Tews[b,*]

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 (Ivo Tews). Fax: +43-316-8736-952; *E-mail address*: peter.macheroux@tugraz.at (Peter Macheroux).

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

Twinning analysis

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 distortion index	Best cell (symmetrized) 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

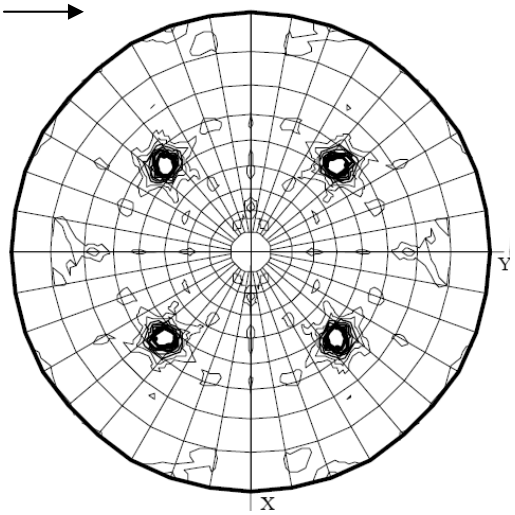
Table S1. Autoindexing table, produced by denzo (HKL Research Inc.); Data reduction with denzo/scalepack (HKL Research Inc.) indicated that only the orthorhombic cell merges with proper statistics.

Symmetry	Cell axes in Å	R^1 overall (LR bin^2 / HR bin^3)
$P2_12_12_1$	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%)
$P2_13$	a=b=c=153.656	17.0% (12.3% / 48.1%)

Table S2. Data scaling using denzo/scalepack (HKL Research Inc.); 1 Given is Rlinear = Σ(ABS(I-<I>))/Σ(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 $P2_12_12_1$ 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 $P2_12_12_1$ 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.

Figure S1. Self rotation function for space group $P2_12_12_1$ 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.



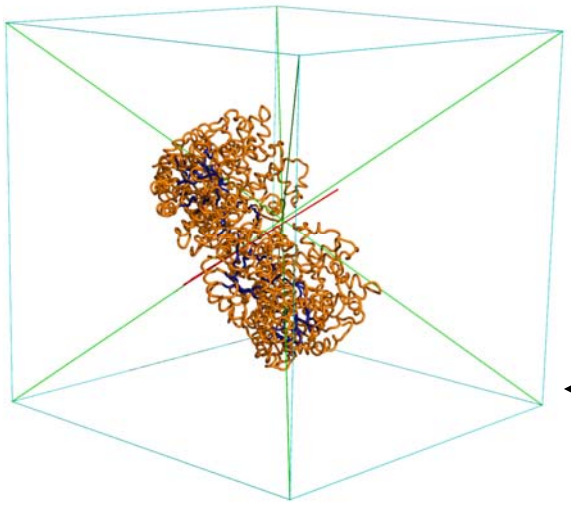


Figure S2. Crystal packing analysis in the $P2_12_12_1$ cell – cell axes are in blue, body diagonals in green; one hexameric particle is shown (compare Figure 3). The rotation axis of the particle is indicated in red. Drawn with pymol [5].

Refinement with Refmac5 [2] in $P2_12_12_1$ gave an R_{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_{free}=0.415 and a FOM=0.553 resulted, while refinement in $P2_13$ gave R_{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/, [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° ±0.5°, comparing superimposition of individual monomers onto each other around the particle rotation axis; this is probably owing to crystal packing. Thus, this is a case were 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 Xtriage from the Phenix suite [6]. Initial checks on data integrity showed highly complete data to 3.09 Å. 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 accentric reflections were detected.

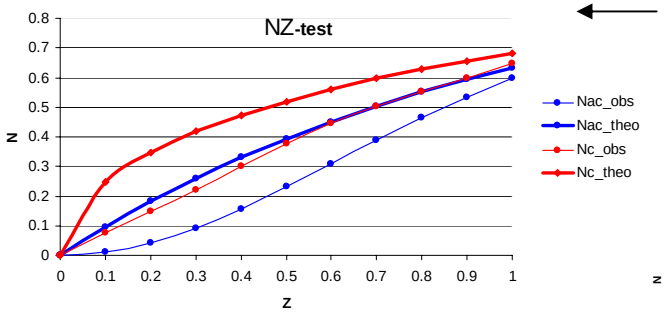


Figure S3. N(Z)-test. Theoretical and observed values are plotted for centric and accentric reflections. Statistics generated by Xtriage [6].

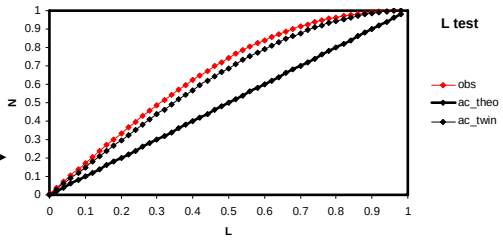


Figure S4. L-test. Theoretical values are plotted in black for untwined (bold) and twinned cases, together with observed values in red. Statistics generated by Xtriage [6].

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. Xtriage [6] reported five pseudo-merohedral twin laws, as listed in Table S3:

twin law	axis	R_{metric}(%)	H-test[1]	Britton α	R *vs* R[2]	ML[3]
k,h,-l	2-fold	0.022	0.169 / 0.149	0.166	0.347 / 0.182	0.022 / 0.027
-h,-l,-k	2-fold	0.044	0.167 / 0.149	0.166	0.347 / 0.182	0.022 / 0.024
l,-k,h	2-fold	0.023	0.167 / 0.149	0.166	0.348 / 0.183	0.022 / 0.024
k,l,h	3-fold	0.044	0.354 / 0.348	0.359	0.152 / 0.033	0.346 / 0.297
l,h,k	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 $P2_12_12_1$ cell on data statistics, as reported by Xtriage [6]. [1] First number given: Estimation of twin fraction via mean |H|; second number given: Estimation of twin fraction via cumulative distance of H; [2] First number given: R_abs_twin observed data; second number given R_sq_twin observed data; [3] First number given: estimated twin fraction; second number: same, but taking NCS into account.

twin law (twin fraction)		R (%)	R_{Free}(%)	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_{Free}, figures of merit (FOM), and bond and angle deviations.

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 $^1/_3$ / $^1/_3$ / $^1/_3$. We then carried out twin refinement to investigate this case. Refinement with phenix.refine [7] gave starting R / R_{free} values of 38.59 % / 37.87 %, and final R / R_{free} 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 α of 0.42 was determined, and starting R / R_{free} values were 29.86 % / 29.42 %, and final R / R_{free} values of 18.14 % / 23.87 %. For the twin operator (-l-hk), a twin fraction α of 0.39 was determined, and starting R / R_{free} values were 30.82 % / 30.34 %, and final R / R_{free} 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 refmac5 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.

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.Raschle T, Amrhein N, Fitzpatrick TB (2005) *J Biol Chem* 280:32291-32300.