

b



С

а

Supplementary Figure 1: CryoEM micrographs and class averages of PaaZ

a Micrograph of PaaZ on ice. Trimeric shape of enzyme (top and bottom view) is marked in red. The occasional dissociated dimer molecules are shown in blue. Scale bar is 800 Å.

b Micrograph of PaaZ on Ultrafoil gold grids coated with graphene oxide. Trimeric shape of PaaZ (top and bottom view) is marked in red. Elongated structures (side view) are marked in green. Scale bar is 300 Å.

c 2D class averages of PaaZ showing the views of the molecule. The box size is 416 pixels and sampled at 1.07 Å/pixel. The class averages shown here were obtained with data collected at the National CryoEM facility, Bangalore.



Supplementary Figure 2: Local resolution plots, FSC curves and density maps of substrate-free PaaZ

a the local resolution plots of substrate-free PaaZ EM map showing the whole molecule on right and a cut away section on left.

b the refinement of the PaaZ model against one of the half-maps is used to verify overfitting. On the left, the Fourier shell correlation (FSC) curves of two half-maps (blue) and the map and model (red) are shown. The resolution estimated from the half-maps at FSC 0.143 is 2.9 Å and that estimated from map and model at FSC 0.5 is 3.5 Å.

c Examples of EM potential density (grey mesh) is shown for α -helices and β -sheets in stereo view. In the dehydrogenase domain, residues 92-116 (α -helix) and 150-156,227-231 (β -sheet) are shown. In the hydratase domain, residues 548-560 (α -helix) and 623-631,647-656 (β -sheet) are shown. The maps were contoured at 7.0 σ and carved 2.0 Å around the atoms. The final B-factor sharpened map was used for making this figure with Pymol.





Dehydrogenase domain

PaaZ - blue 2VR0 - salmon Hydratase domain

PaaZ - blue 5CPG - salmon

Supplementary Figure 3: Overlay of individual dehydrogenase and hydratase domains of known structures with PaaZ

The aldehyde dehydrogenase domain from *Burkholderia xenovorans* (2VRO) is shown in salmon and the PaaZ dehydrogenase domain in blue. The RMSD between the molecules is 0.7 Å for the Ca atoms. The enoyl-Coenzyme A hydratase from *Pseudomonas spp.* (5CPG) is shown in salmon and the PaaZ hydratase domain in blue and has an overall RMSD of 1.7 Å for the Ca atoms. The structures of the dehydrogenase and the hydratase domains used for comparison have been determined by X-ray crystallography.



Supplementary Figure 4: Local resolution plots and FSC curves of PaaZ with NADPH

a the local resolution plots of PaaZ with NADPH showing the whole molecule on right and a cut away section on left.

b the refinement of the PaaZ+NADPH model against one of the half-maps is used to verify overfitting. On the left, the Fourier shell correlation (FSC) curves of two half-maps (blue) and the map and model (red) are shown. The resolution estimated from the half-maps at FSC 0.143 is 3.3 Å and that estimated from map and model at FSC 0.5 is 3.7 Å.



Supplementary Figure 5

Supplementary Figure 5: FSC curves and Local resolution plots of PaaZ OCoA and CCoA

- **a** the local resolution plots of PaaZ with NADP⁺ and octanoyl CoA showing the whole molecule on right and a cut away section on left.
- **b** the refinement of PaaZ OCoA model against one of the half-maps is used to verify overfitting. On the left, the Fourier shell correlation (FSC) curves of two half-maps (blue) and the map and model (red) are shown. The resolution estimated from the half-maps at FSC 0.143 is 3.1 Å and that estimated from map and model at FSC 0.5 is 3.5 Å.
- **c** the local resolution plots of PaaZ with NADP⁺ and crotonyl CoA showing the whole molecule on right and a cut away section on left.
- **d** the refinement of PaaZ CCoA model against one of the half-maps is used to verify overfitting. On the left, the Fourier shell correlation (FSC) curves of two half-maps (blue) and the map and model (red) are shown. The resolution estimated from the half-maps at FSC 0.143 is 3.1 Å and that estimated from map and model at FSC 0.5 is 3.5 Å.



Supplementary Figure 6: Difference maps calculated between the structures of PaaZ

Unsharpened combined maps after refinement were used for the calculation of the difference maps in Chimera. The maps were aligned before the calculation of the difference maps. The PaaZ molecule is shown in transparent gray and the difference density in red. For simplicity, the substrate-free enzyme is called 'native' enzyme.

In panel **a**, the difference between PaaZ NADPH-native is shown and the difference density located in the dehydrogenase domain. In panel **b**, the difference between the NADP⁺ and octanoyl coA (OCoA) bound PaaZ and the native structure is shown. Note that the two peaks are in distinct domains, the NADP⁺ in the dehydrogenase domain and the octanoyl CoA in the hydratase domain. The shape of the octanoyl CoA adopted in binding of PaaZ is clearly visible.

In panel **c**, the difference between the NADP⁺ and crotonoyl coA (CCoA) bound PaaZ and the native structure is shown. As the maps of the native enzyme and the CCoA were obtained from different microscopes, the maps were aligned and pixel size adjusted to get the higest correlation and then difference were calculated. Perhaps due to lower occupancy of the CCoA, the difference density is not as clear as those observed for NADP⁺ or OCoA but strong peak close to the tunnel (marked with CCoA) is visualised. In Panel **d**, the difference between the PaaZ with NADP⁺-octanoyl CoA and PaaZ with NADPH only is shown. Here only the peak from OCoA is observed.



Supplementary Figure 7: Density of ligands and cofactor in PaaZ reconstructions

a, b & c Density of NADPH/NADP+ in the maps of PaaZ mixed with NADPH alone or with OCoA and CCoA respectively. The density for active site residues (C295, E256, H472) are also shown. The density in OCoA and CCoA maps show a continuous density from C295 to the NADP+ ligand. In the maps with NADPH alone there is some residual density around the active site residues. d Density of OCoA in a monomer where no break in density is observed.
e OCoA density in a monomer where the break in density is observed (marked with arrow). It is not clear why the density is fragmented in a region closer to the active site in three of the monomers in the PaaZ-OCoA maps. One possibility could be the small differences in the B-factor sharpening between the monomers in the density is continuous till the active site H566 and a short stretch of density is also found downstream of H566.
f Density of CCoA. The final sharpened maps were used to make the figures with Pymol and the maps have been carved around 2 Å of atoms.



Supplementary Figure 8: The active site of PaaZ with CCoA and NADP+

a An overlay of the NADP⁺/NADPH and the active site of PaaZ+NADPH with the model of CCoA+NADP⁺. The active site residues from both the models are shown in stick representation (carbon atoms in green). The NADP⁺ and NADPH molecules in the models of PaaZ CCoA+NADP⁺ and PaaZ+NADPH are coloured in yellow and magenta respectively. The displacement of the E256 side chain as a result of nicotinamide ring is clearly visible. There are no significant changes in other residues in the active site (C295 and H472) or other residues not involved in catalysis such as N155 and W162.

b The interaction of NADP⁺ with the enzyme in the PaaZ-CCoA+NADP⁺ model. The ligand and key residues are shown in stick representation; the carbon atoms of the ligand and enzyme are shown in yellow and green respectively. Much of the interactions involving the phosphates of the ligand is the same as in the structure of NADPH alone. Due to the position of the ribose and nicotinamide ring, the C4 atom is now placed closed to C295. The ligand is now stabilized by the hydrogen bonds between the carbonyl and nitrogen atoms of the nicotinamide ring to E256 and H472.



Supplementary Figure 9: Small-angle X-ray scattering analysis of PaaZ

a The cryoEM derived model of PaaZ fits the experimental SAXS profile with a chi value of 4.3. A single state model obtained from BILBOMD fits with a chi value of 2.9, whereas as two state model (A and B) fits slightly better with a chi value of 2.5.

b Superposition of cryoEM derived model (shown in grey) and model A (shown in magenta) and Model B (shown in cyan) obtained from BILBOMD.





The conservation scale:

1 2 3 4 5 6 7 8 9

Variable Average Conserved

e - An exposed residue according to the neural-network algorithm.

b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

b

150

PaaZ_Escherichia_coli	α1 <u>00000000</u> 50	α2 20200 6	α3 2020202020	2000000 29	α4 202020 ΒΟ	η1 222 90	٩. ٩٩
PaaZ_Escherichia_coli	AARQFAIEK.	. GAPALRAM	TFIERAAML	AVAKHLLS	SEKERFYAL.	SAQTGAT	RADS.WV
Aldh+MaoC_Methylobacterium_nodulans	DRALQAASRA	FEG <mark>W</mark> KRT	SPYERGEILF	RRAANLMRE	ERLDHIATVI	TLEEGKT	LAEA.KG
Aldh+MaoC_Escherichia_coli	RAAIDAANRA	LPAWRAL	TAKERANILF	RNWFNLMME	HLDDLARLM	ITLEQGKP	LAEA.KG
Aldh+MaoC_Afipia_broomeae	RRAIELANAA	LPA <mark>W</mark> RAK	TAKERSAILF	RKWFDLMIA	ANQDDLAAIM	ITAEQ <mark>G</mark> KP	LAES.KG
Aldh+MaoC_Deinococcus_deserti	RRAIDAAETA	LRE <mark>W</mark> RQT	TGYDRGQVLF	RRWHDLMLQ	HKEELARLM	ITLEMGKP	ITET.RG
Aldh+MaoC_Thermus_thermophilus	REALEEAVLA	FPA <mark>W</mark> SRA	TAYERAQVLF	RRWYER <mark>I</mark> LE	EHQEPLARLM	IALEMGKP	LKEG.RA
Aldh+MaoC_Marinobacter_sp.	DVAVAAARLA	LPA <mark>W</mark> RRL	SGTRRADYLI) GFAGA <mark>L</mark> TF	(RREALIRLS	STNNGKA	IAEA.AI
Aldh+MaoC_Roseobacter_sp.	ERAVAAARRA	FEDGR <mark>W</mark> AKQ	SPAARKKILH	IRIADQ I EA	AEAVELAVLO	VRDNGTE	FNMALKA
Aldh+MaoC_Gordonia_polyisoprenivorans	TAAIAAARTT	FDDGR <mark>W</mark> PAT	PVAERAALLF	RRIADL <mark>L</mark> VF	RDKAELARIE	TLDTGKT	LAES.EI
Aldh+MaoC_Pseudomonas_fluorescens	IA <mark>AV</mark> DAAAS.	. AQKTWGKL	TSIQRAEHLF	RAFAAA <mark>L</mark> EI	[CAESIGRAI	ASESGKS	LDDA.SN
Aldh+MaoC_Pseudomonas_putida	EA <mark>AL</mark> QRAKV.	. GYRQWRQV	SLGQRSEYLI	LALAATLEA	AKA <mark>E</mark> AF <mark>A</mark> HMI	SREIGKP	IAQA.RG
Aldh+MaoC_Paracoccus_denitrificans	DRATAAAVA.	.SFERWSQS	SLDERKAVLN	JRLADLIDF	RDREPLLRNM	IVAESGLP	ITPA.GR
Aldh+MaoC_Herbaspirillum_sp.	EEAIVGADKA	FRTSG <mark>W</mark> AQK	K P H E <mark>R</mark> A I V L J	(RVAQL <mark>I</mark> RE	ERG <mark>E</mark> ALAQRC	RLDNGKP	ITET.RN
Aldh+MaoC_Nitratireductor_pacificus	ERAIERAKRA	QADPA <mark>W</mark> RNL	K L H E <mark>R</mark> A R F L J	AIADGIEA	ANAQRIAHIC	TRDTGKT	LRET.GA
Aldh+MaoC_Alcanivorax_dieselolei	RQAVSAARR.	.AFPGWSRC	GIQARADALI	DFIGSE <mark>L</mark> RA	ARREELGRLI	SREEGKT	LPEG.IA
Aldh+MaoC_Amycolatopsis_azurea	DRAVAAARKA	FTEGPWPKL	LPRERSRVL	JRIADL <mark>V</mark> ES	GRDKRLAALE	SFDSGLP	IAQA.LG
Aldh+MaoC_Sphingomonas_wittichii	DRVVAVARHA	FDHGPWPRM	TGAERAAVLF	RRIAGE <mark>M</mark> GF	(RADDFARAW	ISLQVGMP	YAQS.SM
Aldh+MaoC_Rhodococcus_jostii	DEAVLAARAA	LDNGPWGRS	T P H E R A E T M F	RRFADSLEF	RGESTARLV	SQQNGMI	RPLS.II
Aldh+MaoC_Saccharomonospora_marina	DQAVRAAR	LAFPG <mark>W</mark> AAT	SRQERATYLF	RRLHDALAA	ARVDSVATTI	ATDVGTP	MRIA.SR
PaaZ_Escherichia_coli	α5 <u>00.000000</u> 100	000000000	β6 120	β7 130	β8	→ 150	β9 →
PaaZ_Escherichia_coli	α5 <u>00-000000</u> 100	000000000 119 *	Q β6 120	β7 → 130	β8 140	150	β9
PaaZ_Escherichia_coli PaaZ_Escherichia_coli	α5 <u>00-0000000</u> 100 DI.EGGIGIL	000000000 110 * FTYASLGSR	Ω 120 ELPDDTLWPE	$\beta7$ 130 EDELIPLSP	β8 140 KeggfaarhI	LISKSGV	β9 AVHINAF
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans	α5 00.0000000 100 DI.EGGIGTL EI.TARAEFF	QQQQQQQQQ 119 * FTYASLGSR DWFAEEGRR	β6 120 ELPDDTLWPE	β7 130 EDELIPLSF (GRTIPARI	β8 140 (EGGFAARHI PDLRQTV	LISKSGV TQEPVGP	β9 AVHINAF VACFTPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli	00.0000000 100.0000000 DI.EGGIGTL EI.TAAAEFF EI.SYAASFI	QQQQQQQQ 110, * FTYASLGSR DWFAEEGRR EWFAEEGKR	β6 2 → 120 ELPDDTLWPE 	β7 130 EDELIPLSP (GRTIPARI (GDTIPGH0	β8 140 KEGGFAARHI L.PDLRQTV 2.ADKRLIV	LISKSGV TQEPVGP IKQPIGV	β9 AVHINAF VACFTPW TAAITPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Afipia_broomeae	00.0000000 100 DI.EGGIGTL EI.TAAAEFF EI.SYAASFI EI.AYAASFI	000000000 110 * FTYASLGSR EWFAEEGKR EWFAEEGKR EWFAEEGKR	β6 120 ELPDDTLWPE 	β7 130 EDELTPLSK (GRTIPARI (GDTIPGHC (GDTIPFT)	β8 140 KEGGFAARHI LPDLRQTV 2ADKRLIV CSDRRIV CSDRRIV	LISKSGV TQEPVGP IKQPIGV LKQPIGV	β9 AVHINAF VACFTPW TAAITPW CAAITPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Deinococcus_deserti	00.0000000 100 DI.EGGIGTL EI.TAAAEFF EI.SYAASFI EI.AYAASFI EV.HYAASFI	QQQQQQQQQ 119 * FTYASLGSR DWFAEEGRR EWFAEEGRR EWFAEEGRR EWFAEEASR	β6 2 → 120 ELPDDTLWPF 	β7 130 EDELIPLSK (GRTIPARI (GDTIPGHC (GDTIPFT GERISMRI GERISMRI	β8 140 KEGGFAARHI LPDLRQU 2ADKRLIV LSDRRIV LSDRRIV	LISKSGV TQEPVGP IKQPIGV LKQPIGV SAEPVGI	β9 AVHINAF VACFTPW TAAITPW CAAITPW VYAVTPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Thermus_thermophilus	US OQ.OCOOOOO DI.EGGIGIL EI.TAAAEFF EI.SYAASFI EI.AYAASFI EV.HYAASFI EV.HYAASFI EV.AYAAGFV	QQQQQQQQQ TIQ FTYASLGSR DWFAEEGRR EWFAEEGRR EWFAEEGKR EWCAEEASR EWQAEEASR	β6 120 ELPDDTLWPF 	β7 130 EDELIPLSP (GRTIPARI (GDTIPGHC (GDTIPF) GERINRI (GETVPSQE	β8 140 KEGGFAARHI L.PDLRQIV 2.ADKRLIV C.SDRRIVV C.SDRRIVV 5.GNKRGFS F.PHKRILV	150 LISKSGV TQEPVGP IKQPIGV SAEPVGI RYEPVGP	β9 AVHINAF VACFTPW TAAITPW CAAITPW VYAVTPW VYAUTPW VYGITPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Marinobacter_sp.	CS QQ.QCQQQQQ D C C C C C C C C C C C C C	QQQQQQQQQ 11 Q * FTYASLGSR FTYASLGSR EWFAEEGRR EWFAEEGKR EWFAEEGKR EWFAEEAKR EWYAEEAKR RYYAGQARA	β6 120 ELPDDTLWPE 	B7 130 EDELITPISF (GDTIPPF) (GDTIPPF) (GDTIPPF) (GETVPSQE (GETVPSQE) (GETVPSQE) (GETVPSQE)	β8 140 LEGGFAARHI L. PDLRQTV 2. ADKRLIV 2. ADKRLIV 3. GNKRGS 5. PHKRILV 4. KGVVARC	LISKSGV TQEPVGP IKQPIGV SAEPVGI XHDPVGV YHDPVGV	β9 AVHINAF VACFTPW TAAITPW CAAITPW VYAUTPW VYGITPW VGLITPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Roseobacter_sp. Aldh+MaoC_Roseobacter_sp.	α5 00.0000000 DT.EGGIGTL EI.TAAAEFF EI.SYAASFI EI.AYAASFI EV.HYAASFI EV.AYAASFI DL.DDAIACY EP.GSAAGTF	QQQQQQQQQ TIQ FTYASLGSR DWFAEEGRR EWFAEEGRR EWFAEEGRR EWCAEEASR EWYAEQARA RYYAQQARA RYYAQARAUFT	β6 120 ELPDDTLWPF 	β7 130 EDELTPISF (GRTIPARI (GDTIPFT GERISMRI GERVALEN (GE, VAPTR (GE) (GE)	β8 14 φ KEGGFAARHI L. PDLRQTV Q. ADKRLIV C. SDRRIVV L. GNKRGFS F. PHKRILV Δ. KGVVARC Δ. KGVVARC Δ. PDVLGLV	150 LITSKSGV TQEPVGP IKQPIGV LKQPIGV SAEPVGI XHDPVGV HRAPVGV	β9 AVHINAF VACFTPW TAAITPW CAAITPW VYAVTPW VYGITPW VGLITPW VGLITPW VGLITPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Gordonia_polyisoprenivorans Aldh+MaoC_Gordonia_polyisoprenivorans	US OQ. DI.EGGIGIL EI.TAAAEFF EI.SYAASFI EV.HYAASFI EV.HYAASFI EV.AYAAGFV DL.DDAIACY EP.GSAAGTF DI.DDVESVF DI.DDVESVF	QQQQQQQQQ TIQ FTYASIGSR DWFAEEGRR EWFAEEGRR EWFAEEGRR EWYAEEASR EWYAEEASR RYYAEALDK RYYAEALDK RYYAEALDK	β6 2 120 ELPDDTLWPE 	β7 → 130 EDELIPISE GENTIPARI (GDTIPPF) GERLISMRI (GETVPS0E QESVALEN (GE.VAPTA ADRAVDIGV	β8 140 KEGGFARHI Δ. PDLRQIV Δ. ADKRLIV Δ. GNKRGV Δ. GNKRGV Δ. KGVVARC Δ. PDVLGLV Δ. PGVISRV Δ. PGVISRV	LISKSGV TQEPVGP IKQPIGV LKQPIGV SAEPVGI YHDPVGV YHDPVGV VREPIGV OPPEVGV	β9 VACFTPW TAATTPW CAATTPW VYAVTPW VYGITPW VGLITPW VGLITPW CVLIAPW CVLIAPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Escherichia_coli Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Marinobacter_sp. Aldh+MaoC_Roseobacter_sp. Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_nutide	C5 OC.OCOOQOO DI.EGGIGTL EI.TAAAEFF EI.SYAASFI EV.AYAASFI EV.AYAASFI EV.AYAAGFV DL.DDAIACY EP.GSAAGFV DI.DDVESVF EA.RYAAQIT EV.AYAAQIT	200000000 110 FTYASLGSR FTYASLGSR EWFAEEGRR EWFAEEGRR EWFAEEGRR EWYAEEARR RYYAEALDKR RYYAEALDKR RYYAEALVET RYHAEWARR	β6 120 ELPDDTLWPF 	B7 130 EDELIPISE (GDTIPARI (GDTIPFFI (GDTIPFFI (GETVPSQF QGESVALEM (GETVPSQF QGESVALEM 100 100 100 100 100 100 100 10	β8 140 CEGGFAARHI C.PDLRQIV 2.ADKRLIV 3.GNKRGFS 5.PHKRILV 4.KGVVARC 4.PDVLGLV 7.PGVLSRV 7.PDENLFI 0.00000000000000000000000000000000000	LISKSCV TQEPVGP IKQPIGV SAEPVGP YHDPVGV HRAPVGV VREPIGV QREPIGV QREPIGV	β9 VHINAF VACFTPW TAAITPW CAAITPW VYGITPW VYGITPW VGAIVPW VGAIVPW CVLIAPW VACLIPF
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Escherichia_coli Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Roseobacter_sp. Aldh+MaoC_Gordonia_polyisoprenivorans Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_fluorescens	α5 QQ.QQQQQQQQ DT.EGGIGTL EI.TAAAEFF EI.SYAASFI EV.HYAASFI EV.HYAASFI EV.AYAASFI EV.AYAAGFV DL.DDAIACY EP.GSAAGTF DI.DDAISVF EA.RYAAQIT EV.AKCVGLC	QQQQQQQQQ IIQ FTYASLGSR DWFAEEGRR EWFAEEGRR EWFAEEGRR EWCAEEASR EWCAEEASR EWYAEQARA RYYAEQARA RYYAEALDK RYYASLWAR RYYAEWARR RWYAEHQAA	$\begin{array}{c} & & \beta 6 \\ 2 \\ 120 \\ ELPDDTLWPF \\ & & S3 \\ & & & S3 \\ & & & S3 \\ & & & & S3 \\ & & & & S3 \\ & & & & & S3 \\ & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & S3 \\ & & & & & & & & & S3 \\ & & & & & & & & & S3 \\ & & & & & & & & & & S3 \\ & & & & & & & & & & & S3 \\ & & & & & & & & & & & S3 \\ & & & & & & & & & & & & S3 \\ & & & & & & & & & & & & & & S3 \\ & & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & & & & & \\ &$	β7 130 EDELLPLSF (GRTIPARI (GDTIPFT) GERISMRI GETIVPSQE (GE., VAPTA JORAVDIG CEPTIPSDI LAAREP	β8 14 φ KEGGFAARHI L. PDLRQTV Q. ADKRLIV C. SDRRIVV L. GNKRGFS F. PHKRILV Δ. KGVVARC Δ. PDVLGLV V. PGVISRV C. PDCVLSRV C. PDCVLSRV C. PCVEKARI C. QVEKARI	LISKSCV TQEPVGP IKQPIGV SAEPVGI SAEPVGI SAEPVGV WREPIGV VREPIGV QREPIGV URCPIGV	β9 AVHINAF VACFTPW TAAITPW CAAITPW VYAVTPW VYGLITPW VGLITPW VGLITPW VGLITPF ILAVMPW VACLIPF ILAVMPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Gordonia_polyisoprenivorans Aldh+MaoC_Gordonia_polyisoprenivorans Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_putida Aldh+MaoC_Paracoccus_denitrificans Aldh+MaoC_Paracoccus_denitrificans	US OQ DI.EGGIGIL EI.TAAAEFF EI.SYAASFI EV.HYAASFI EV.HYAASFI EV.AYAAGFV DL.DDAIACY EP.GSAAGTF DI.DDVESVF EA.RYAAQIT EV.AKCVGLC SQVDGASDEL UCC SASTE	200000000 FTYASLGSR DWFAEEGRR EWFAEEGRR EWFAEEGKR EWCAEEASR EWCAEEASR EWCAEALDK RYYAEALDK RYYASLVET RYHAEWADE RYHAEWADE RYHAEWGPA RUMAEVVDQ	β6 2 120 ELPDDTLWPE 	β7 130 EDELIPLSE GERTIPARI (GDTIPPF) GERTSMRI (GDTVPS0F GERTSMRI (GDTVPS0F GERTSMRI (GDTVPS0F GERTSMRI (GDTIPF) GESTALEN (GDTIPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) GESTALEN (GDTIPPF) (GDTIPPF) GESTALEN (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPPF) (GDTIPF) (GDTIPPF) (GDTIPF) (GDTIPPF) (GDTIPF) (GDTIPF) (GDTIPF) (GDTIPPF) (GDTIFF) (GDTIF	β8 140 KEGGFARHI Δ. PDLRQIV Δ. ADKRLIV Δ. GNKRGFS F. PHKRILV Δ. KGVVARC Δ. PDVLGLV Γ. PGVISRV Γ. PDENLF Γ. QVEKARI Γ. HYGSSRV C. DVLKARI	LISKSGV TQEPVGP IKQPIGV LKQPIGV SAEPVGI YHDPVGV YHDPVGV VREPIGV QREPIGV QREPIGV VRQAAGV	β9 AVHINAF VACFTPW TAAITPW CAAITPW CAAITPW VYGLITPW VGLITPW VGLITPW CVLIAPW CVLIAPW VACLIPF ILAVMPW TGAITAW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Scherichia_coli Aldh+MaoC_Afipia_broomeae Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Thermus_thermophilus Aldh+MaoC_Marinobacter_sp. Aldh+MaoC_Gordonia_polyisoprenivorans Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_putida Aldh+MaoC_Pseudomonas_putida Aldh+MaoC_Berbaspirillum_sp. Aldh+MaoC_Herbaspirillum_sp.	C5 OQ.OCOOOOO DI.EGGIGTL EI.TAAAEFF EI.SYAASFI EV.HYAASFI EV.HYAASFI EV.AYAAGFV DL.DDAIACY EP.GSAAGTF DI.DDVESVF EA.RYAAQIT EV.AYAAQIT EV.AXCVGLC SQVDGASDEL LVG.SAAATF	20000000 110 FTYASLGSR DWFAEEGRR EWFAEEGRR EWFAEEGKR EWCAEEASR EWYAEEAKR RYYAGQARA RYYAEALDK RYYASLVET RYHAEWARR RYYASLVET RYHAEWARR RYYAEHGPA RLMAEVVDQ QFFAAACET VFGAMIET	β6 2 120 ELPDDTLWPE 	β7 → 130 EDELTPLSF GENTIPARI GENTIPARI GETTPARI GETVPSQF GERSTSMRI GETVPSQF GESVALEM ADRAVDIGV GETTPSDI DWA1 EETITPME	β8 140 CEGGFAARHI ADKRIV ADKRIV C.SDRRIV C.SDRRIV C.GNKRGFS F.PHKRIV Δ.KGVVARC Δ.PDVLGIV V.PGVISRV C.PDENLFI C.QVEKARI C.HYGSSRV C.GDNLTMS C.NAITMON	LISKSGV TQEPVGP TQEPVGP IKQPIGV SAEPVGI RYEPVGP YHDPVGV VREPIGV VREPIGV VREPIGV VREPLGP VRQAAGV VYEPMGV VYEPMGV	β9 VACFTPW TAAITPW CAAITPW CAAITPW VYGITPW VYGITPW VGLITPW VGLITPW CVLIAPW VACLIPF ILAVMPW TGAITAW VAAITPW
PaaZ_Escherichia_coli PaaZ_Escherichia_coli Aldh+MaoC_Methylobacterium_nodulans Aldh+MaoC_Escherichia_coli Aldh+MaoC_Escherichia_coli Aldh+MaoC_Deinococcus_deserti Aldh+MaoC_Marinobacter_sp. Aldh+MaoC_Marinobacter_sp. Aldh+MaoC_Cordonia_polyisoprenivorans Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_fluorescens Aldh+MaoC_Pseudomonas_putida Aldh+MaoC_Herbaspirillum_sp. Aldh+MaoC_Mitratireductor_pacificus Aldh+MaoC_Nitratireductor_pacificus	US OQ.OQOQOQO DI EI.EGGIGTL EI.TAAAEFF EI.SYAASFI EV.AYAASFI EV.AYAAGFV DI.DDAIACY EP.GSAAGFV DI.DDVESVF EA.RYAAQIT EV.AKCVGLC SQVDGASDEL LVG.SAAATF LAA.SAAGFF	200000000 110 ** FTYASLGSR EWFAEEGRR EWFAEEGKR EWFAEEGKR EWCAEEASK EWCAEEASK RYYASLVET RYYASLVET RYYASLVET RYHAEWARB RWYAEHGPA RWYAEHGPA RLMAEVVDQ QFFAAACET RYFGAMLET	β6 120 ELPDDTLWPF 	B7 130 EDELTPLSF (GDTIPPFI (GDTIPPFI GETVPSQE QESVALEN DRAVDIGV EGEIIPSDI LAAEPI DWAL1 EETITPMF DEALTAQE	β8 140 KEGGFAARHI L. PDLRQTV 2. ADKRLIV 2. ADKRLIV 4. KGVVARC 4. PDVLGLV 4. PDVLGLV 4. PDVLGLV 5. PDENLFL 5. QVEKARI 5. GDNLTMS 8. GDNLTMS 2. GNALTMS 3. PDVHV	LISKSGV TQEPVGP IKQPIGV SAEPVGP YHDPVGV YHDPVGV QREPIGV QREPIGV VYEPIGV VYEPMGV VYEPMGV VYEPMGV TBEPVGV	β9 VHINAF VACFTPW CAAITPW CAAITPW VACITPW VGITPW VGITPW VGITPW VGITPW VACLIPF ILAVMPW VACLIPF ILAVMPW VACITPW VAAITPW VGAITPW

			834		825	
PaaZ Escherichia coli	000000		µ24	· TT	μ25	•
5	90	600	610	620	630+	+ 640
PaaZ_Escherichia_coli	AAAGLFV	DAGVGP	. VIANYGLESLRF	IEPVKPGDTIQ	VRLTCKRKTI	KKQRSAEE
Aldh+MaoC_Methylobacterium_nodulans	LISAVLO	TRLPG.	. PGAVYISQSLNF	RAPVRIGDTVD	VTVEVAELVE	?E
Aldh+MaoC_Escherichia_coli	FFSALFG	TKLPG.	. P G C V Y V N Q S L K F	LRPVYINDTVT	ARVVLTDIDV	/v
Aldh+MaoC_Afipia_broomeae	LVAALAM	KLGLAPKDAS.	. AVINYGLDKVRF	ITPVKAGERVR	LQLALIGIEF	?К
Aldh+MaoC_Deinococcus_deserti	MLAGEFM	INTGGAPHIEGAN	RMVVNYGLNRVRF	ITPVRSGTRLR	NRAVLVSAE.	P
Aldh+MaoC_Thermus_thermophilus	MLTGLRC	RTGHFEGT	.VIAWLEIRNYKF	LK <mark>PV</mark> LIGDTVR	GESEILEKRF	CTSKP
Aldh+MaoC_Marinobacter_sp.	LISAVLO	TRLPG.	. PGAIYIDQQLRF	KAPVFIGDAVV	ASAEVQEIDI	
Aldh+MaoC_Roseobacter_sp.	LISAVIG	EQLPG.	. HGTIYMSQS <mark>LKF</mark>	LAPVRPGDMVL	AEVEVTDIVI	[D
Aldh+MaoC_Gordonia_polyisoprenivorans	IAQRVAV	SSVLTSWS	SVIAGRRLREMEF	LAPVRPGDTLS	GSMRIESVDI	DPRG
Aldh+MaoC_Pseudomonas_fluorescens	LIPKLIE	DILIMPEGLE	KMAVNYGLDSVRF	IQPVKVDSRVR	LAVTLTDATF	CKKPG
Aldh+MaoC_Pseudomonas_putida	LIPKLIE	DILVLPQGL	KMVVNYGLDSVRF	IQPVKVDSRVR	LKVTLGEVTF	EKKPG
Aldh+MaoC_Paracoccus_denitrificans	LMIGISV	HDTTLG	ITVANLGMIDIVF	PNPVFHGDTIR	VETEVKSVRF	ESKSK
Aldh+MaoC_Herbaspirillum_sp.	LTMRLLV	KTLPIKO	GGVIGAGFEEFRW	P R P T R P G D V L R	VEVEVLELKE	SQSK
Aldh+MaoC_Nitratireductor_pacificus	LMIGISV	HDTTLG	TTVANLGMTDTVF	PNPVFHGDTIR	VETEVKSVRF	ESKSK
Aldh+MaoC_Alcanivorax_dieselolei	LIPMFKD	QIYA.IEGV	KAGINYGLNGCRF	LAPVPAGARVR	ARFVMKSVTÇ	QKGEG
Aldh+MaoC_Amycolatopsis_azurea	WGIPMLA	ELLH.VTGVO	GMALNYGLNKVRF	P A <mark>P V</mark> P V G S R V R	LHASVLEVTP	PVPRD
Aldh+MaoC_Sphingomonas_wittichii	LVNYFLP	QMID.VRGFS	SMGVNVGADRLRF	LNPVRTGSRVR	GTGEIVGVEF	GV.KG
Aldh+MaoC_Rhodococcus_jostii	LVPMLTW	QVYK.VEGV	IMSVNYGADKLRF	P S P V P V G S R V R	AGVELTSVTP	PN.KL
Aldh+MaoC_Saccharomonospora_marina	LIPMLGK	DVYH.VEGLI	RMGINYGLNK <mark>VRF</mark>	PQPVKVGSRIR	AGAELVSVSD	DA.SQ

QA.RRAADIFRFFADLIVA....QADDTYKV.P.GRQVNYVNRKPIGVAGLITPW TA.PYMAAYLTYFADLAEK.....GFEEVRQP.M.MGGHCIVVHEPVGVVAAVVPW SNAAAPVAVLRSMAELAEA....RTFETRRP.S.KKGHTIVRQEPVGVVGAIVAW IQAQLPLTDIATYADLLDQ......PEIE.E.KIGNTLLVREPIGVVAAITPW

С

Supplementary Figure 10: Sequence analysis and gel filtration profile of PaaZ

a Evolutionary conservation of PaaZ homologues as analysed by Consurf

Aldh+MaoC_Amycolatopsis_azurea Aldh+MaoC_Sphingomonas_wittichii Aldh+MaoC_Rhodococcus_jostii Aldh+MaoC_Saccharomonospora_marina

b Gel filtration profiles of PaaZ WT and K69A mutant show similar elution profiles.

c Sequence analysis of individual dehydrogenase and hydratase domains against PaaZ.

Alignment of 20 representative homologues of PaaZ containing individual gene products of aldehyde dehydrogenase, enoyl-CoA hydratase domains. K69, R116, K631 and K636 are conserved only in fused PaaZ and its homologues while these amino acids are seen not to be under the evolutionary pressure in individual gene products of aldehyde dehydrogenase, enoyl-CoA hydratase domains.