



STRUCTURAL
BIOLOGY

Volume 76 (2020)

Supporting information for article:

Crystal structures of the two domains that constitute the *Plasmodium vivax* p43 protein

Swati Gupta, Jyoti Chhibber-Goel, Manmohan Sharma, Suhel Parvez, Karl Harlos, Amit Sharma and Manickam Yogavel

Crystal structures of the two domains that constitute the *Plasmodium vivax* p43 protein

Swati Gupta^{a+}, Jyoti Chhibber-Goel^{a+}, Manmohan Sharma^{a+}, Suhel Parvez^b, Karl Harlos^c, Amit Sharma^a and Manickam Yogavel^{a*}

⁺ Equal contribution

^a Molecular Medicine - Structural Parasitology Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Road, 110067 New Delhi, India.

^b Department of Medical Elementology and Toxicology, Jamia Hamdard University, New Delhi, India.

^c Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom.

Supplementary Tables and Figures

Supplementary Table S1a. Summary of anomalous data collection and anomalous signal analysis[#].

Data Set ^{##}	Co-x3_2	Co-x3_4	Co-x5_2	Co-x7_2
Beam line	DLS I02	DLS I02	DLS I02	DLS I02
Wavelength (λ) (Å)	1.59	1.59	1.59	1.59
Detector type	DECTRIS PILATUS 6MF	DECTRIS PILATUS 6MF	DECTRIS PILATUS 6MF	DECTRIS PILATUS 6MF
Crystal-to-detector distance (mm)	625	625	625	625
Oscillation (°)	0.1	0.1	0.1	0.1
Exposure (s)	0.050	0.050	0.050	0.050
Beam size (μm) (elliptical)	100 x 20	100 x 20	100 x 20	100 x 20
Flux (photons s^{-1})	1.16×10^{11}	1.16×10^{11}	1.16×10^{11}	1.16×10^{11}
Transmission (%)	4.94	4.94	4.94	4.94
Number of images	7200	7200	7200	9900
Cell dimensions a , b , c (Å)	86.99, 53.29	86.93, 86.93, 53.34	86.72, 86.72, 53.24	86.81, 86.81, 53.48
Space group	$P4_1 2_1 2$	$P4_1 2_1 2$	$P4_1 2_1 2$	$P4_1 2_1 2$
Resolution (Å)	86.99-2.67 (2.72-2.67)	61.47-2.71 (2.76-2.71)	61.32-3.11 (3.16-3.11)	61.39-2.75 (2.80-2.75)
R_{meas} (%)	0.111 (4.468)	0.110 (3.625)	0.266 (3.004)	0.134 (3.404)
$\langle I/\sigma(I) \rangle$	19.2 (1.0)	20.8 (1.1)	10.3 (0.7)	20.4 (1.3)
Completeness (%)	96.0 (71.1)	96.9 (73.3)	100.0 (96.2)	92.2 (47.9)
Average redundancy	44.4 (27.3)	44.9 (29.6)	42.1 (20.5)	65.4 (63.8)
$\text{CC}_{1/2}$	1.0 (0.5)	1.0 (0.5)	1.0 (0.2)	1.0 (0.6)
Number of observations	263434 (5778)	257492 (6327)	166795 (3661)	343056 (8670)
Number of unique reflections	5929 (212)	5734 (214)	3963 (179)	5245 (136)
$[\langle \Delta F/\sigma(\Delta F) \rangle]^{\$}$	1.22	1.30	1.01	1.25
$(\langle \Delta F \rangle/\langle F \rangle)^{\$\$}$	0.065	0.067	0.131	0.065
Practical resolution limit of anomalous signal (Å) [@]	4.9	4.6	8.4	4.9

[#] Several Co-SAD data sets were collected using ten different crystals (x1-x10) and multiple datasets were collected for each crystal at a different spots. The statistics of the selected data sets with $(\Delta F)/\sigma(\Delta F) > 1.0$ were listed.

^{##} In the data set name, x1 to x10 refers to the crystal number and the number after the underscore refers to

the number of the dataset that was collected on a particular crystal at different spots.

^{\$}Anomalous signal-to-noise ratio was calculated using HKL2MAP (Pape & Schneider, 2004).

^{\$\$}The Bijvoet ratio was calculated using *phenix.reflections_statistics* Xtriage (Zwart *et al.*, 2005).

[@]Practical resolution limit of anomalous signal (Å) was calculated using Xtriage (Zwart *et al.*, 2005).

Supplementary Table S1b. Summary of anomalous data collection and anomalous signal analysis[#].

Data Set [#]	Co-x7_4	Co-x9_2	Co-x10_3	Co-x10_6
Beam line	DLS I02	DLS I02	DLS I02	DLS I02
Wavelength (λ) (Å)	1.59	1.59	1.59	1.59
Detector type	DECTRIS PILATUS 6MF	DECTRIS PILATUS 6MF	DECTRIS PILATUS 6MF	DECTRIS PILATUS 6MF
Crystal-to-detector distance (mm)	625	625	625	625
Oscillation (°)	0.1	0.1	0.1	0.1
Exposure (s)	0.050	0.050	0.050	0.050
Beam size (μm) (elliptical)	100 x 20	100 x 20	100 x 20	100 x 20
Flux (photons s ⁻¹)	1.12 x 10 ¹¹	9.62 x 10 ¹⁰	6.54 x 10 ¹⁰	5.57 x 10 ¹⁰
Transmission (%)	4.94	4.94	3.0	3.0
Number of images	9900	7200	9900	9900
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	87.00, 87.00, 53.29	86.66, 86.66, 53.33	87.40, 87.40, 53.41	86.97, 86.97, 53.283
Space group	<i>P</i> 4 ₁ 2 ₁ 2			
Resolution (Å)	61.52-2.67 (2.72-2.67)	86.66-2.75 (2.80-2.75)	61.80-2.59 (2.63- 2.59)	61.50-2.81 (2.86-2.81)
R _{meas} (%)	0.122 (3.883)	0.188 (3.039)	0.081 (3.353)	0.092 (2.885)
$\langle I/\sigma(I) \rangle$	23.9 (1.5)	14.6 (0.7)	27.1 (1.2)	22.7 (1.1)
Completeness (%)	88.9 (42.9)	100.0 (100.0)	95.8 (71.8)	100.0 (98.8)
Average redundancy	65.3 (61.4)	44.3 (23.4)	58.1 (28.3)	63.3 (43.3)
CC _{1/2}	1.0 (0.6)	1.0 (0.5)	1.0 (0.7)	1.0 (0.9)
Number of observations	359633 (7803)	251067 (6671)	379840 (6850)	338184 (10387)
Number of unique reflections	5504 (127)	5664 (285)	6539 (242)	5338 (240)
$[\langle \Delta F/\sigma(\Delta F) \rangle]^{\$}$	1.38	1.22	1.71	1.29
$(\langle \Delta F \rangle/\langle F \rangle)^{\$\$}$	0.056	0.118	0.062	0.052
Practical resolution limit of anomalous signal (Å) [@]	4.5	5.4	4.0	4.5

[#] Several Co-SAD data sets were collected using ten different crystals (x1-x10) and multiple datasets were collected for each crystal at a different spots. The statistics of the selected data sets with $(\Delta F)/\sigma(\Delta F) > 1.0$ were listed.

^{##} In the data set name, x1 to x10 refers to the crystal number and the number after the underscore refers to the number of the dataset that was collected on a particular crystal at different spots.

^{\$} Anomalous signal-to-noise ratio was calculated using HKL2MAP (Pape & Schneider, 2004).

^{\$\$} The Bijvoet ratio was calculated using *phenix.reflections_statistics* Xtriage (Zwart *et al.*, 2005).

[@] Practical resolution limit of anomalous signal (\AA) was calculated using Xtriage (Zwart *et al.*, 2005).

Supplementary Table S1c. Statistics of merged Co SAD data set and anomalous signal analysis

Data Set	Merged data [#]
Beam line	DLS I02
Wavelength (λ) (Å)	1.59
Detector type	DECTRIS PILATUS 6MF
Crystal-to-detector distance (mm)	625
Oscillation (°)	0.1
Exposure (s)	0.050
Beam size (μm) (elliptical)	100 x 20
Flux (photons s ⁻¹)	1.12 x 10 ¹¹
Number of images	44100
Cell dimensions a , b , c (Å)	86.99, 86.99, 53.35
Space group	$P4_1 2_1 2$
Resolution (Å)	61.51-2.67 (2.72-2.67)
R _{meas} (%)	0.107 (2.193)
$\langle I/\sigma(I) \rangle$	40.8 (1.1)
Completeness (%)	97.1 (77.9)
Average redundancy	136.0 (19.7)
CC _{1/2}	1.0 (0.6)
Number of observations	1521516 (8672)
Number of unique reflections	11188 (441)
[$\langle \Delta F/\sigma(\Delta F) \rangle$]	2.01
$(\langle \Delta F \rangle/\langle F \rangle)$	0.043
Practical resolution limit of anomalous signal (Å)	3.7

The five merged different data sets are: x3_2, x3_4, x7_2, x10_3 and x10_6

Supplementary Table S2. Statistics of anomalous signal-to-noise ratio [$\langle \Delta F/\sigma(\Delta F) \rangle$] versus resolution for x10_3 and merged data set[#]. The statistics of anomalous signal-to-noise ratio for different resolution bins was calculated using HKL2MAP (Pape & Schneider, 2004).

Statistics for x10_3 data set												
Resolution (Å)	∞-8.0	8.0-	6.0	6.0-5.0	5.0-	4.0-	3.8-	3.6-	3.4-	3.2-	3.0-	2.8-
N(data)	276	342	420	919	294	395	465	584	762	980	1075	
$\langle I/\sigma(I) \rangle$	105.0	70.6	65.2	57.3	40.4	34.3	27.1	16.6	9.2	4.6	1.4	
Completeness (%)	96.5	99.1	99.1	99.6	99.7	99.7	99.6	99.8	99.7	98.4	77.9	
$\langle \Delta F/\sigma(\Delta F) \rangle$	3.87	2.92	2.23	1.42	1.25	1.13	1.23	1.24	1.15	1.11	1.27	

Statistics for merged data set[#]												
Resolution (Å)	∞-8.0	8.0-	6.0	6.0-5.0	5.0-	4.1-	3.9-	3.7-	3.5-	3.3-	3.1-	2.90-
N(data)	455	596	766	1417	517	624	778	970	1243	1604	2206	
$\langle I/\sigma(I) \rangle$	177.2	107.3	95.4	80.8	52.7	44.0	33.9	23.5	12.2	6.9	2.3	
Completeness (%)	98.3	99.8	99.7	99.6	99.2	99.5	98.9	99.2	99.2	99.2	88.1	
$\langle \Delta F/\sigma(\Delta F) \rangle$	5.06	3.88	2.92	1.74	1.29	1.09	0.91	0.83	0.76	0.75	0.67	

The five merged different data sets are: x3_2, x3_4, x7_2, x10_3 and x10_6

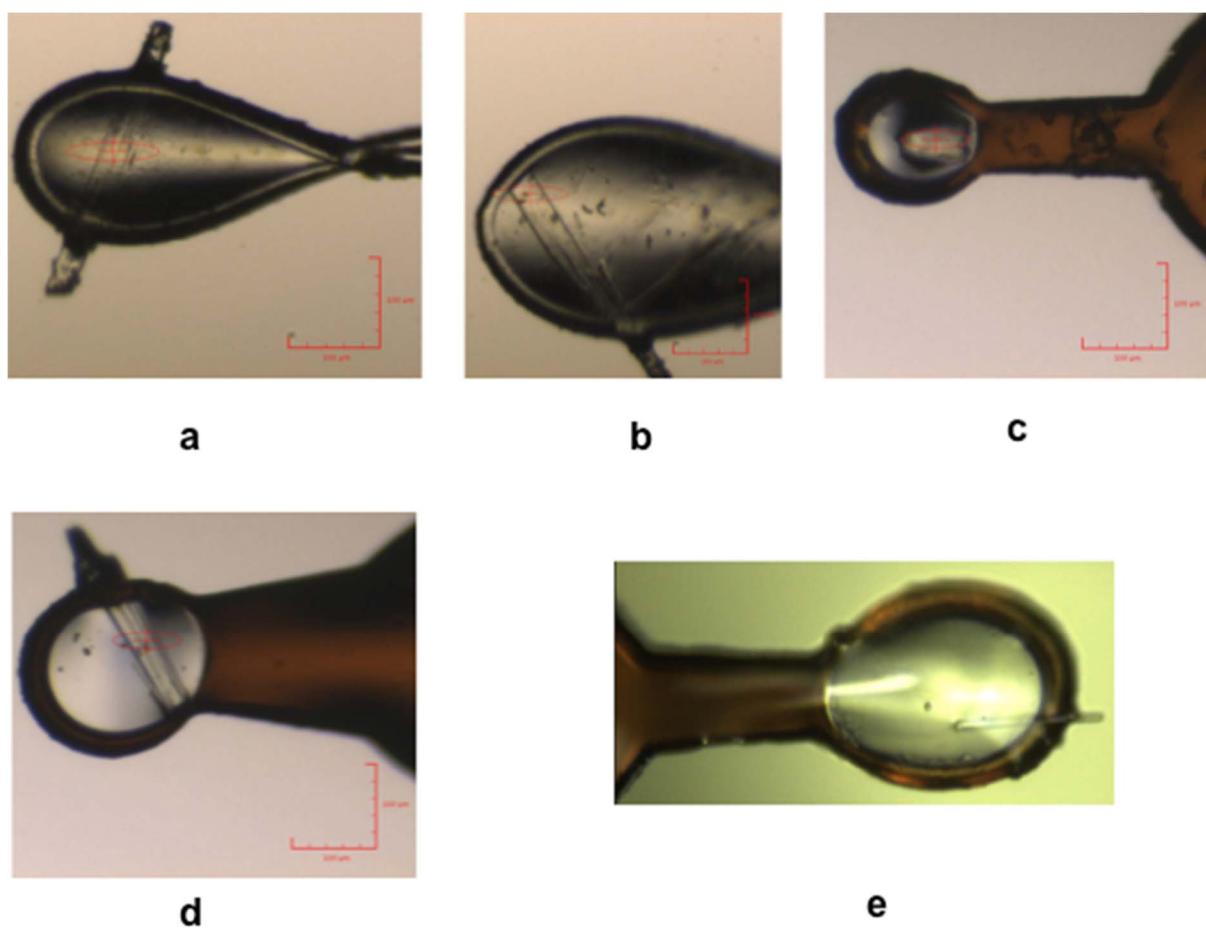
Supplementary Table S3. Results of DALI search for *Pvp43* NTD (residues 1-180) against PDB25 (a subset of Protein Data Bank). The structure was specified by concatenating string of the PDB identifier (4 characters) and a chain identifier (1 character). The match list is sorted by Z-scores (Z-scores between 8 and 20 means the two are probably homologous). The other listed parameters in Table S3 are: root mean square deviation (RMSD) of C^a atoms in superimposition; Lali (number of structurally equivalent positions); N_{res} - length of the structural neighbor and %id - percentage amino acid identity in aligned positions.

No.	PDB_Chain	Z-score	RMSD	Lali	N _{res}	% id	Description
1	5A34_C	15.0	2.6	153	170	17	Bifunctional Glutamate/Proline-tRNA Ligase
2	3MDK_B	14.9	2.7	159	206	10	Stringent Starvation Protein A
3	1NHY_A	14.8	2.4	158	219	13	Elongation Factor 1-Gamma 1
4	4ZB7_A	14.8	2.7	152	171	12	Pcure2p6
5	3VK9_B	14.7	2.7	156	213	13	Glutathione S-Transferase Delta
6	5G5A_A	14.6	2.8	157	219	13	Glutathione S-Transferase U25
7	5Y6L_B	14.5	3.2	150	173	14	Methionine-tRNA Ligase, Cytoplasmic
8	5NR1_A	14.3	2.5	156	226	10	Ftsz-Binding Protein Fzla
9	4KDU_A	14.0	2.7	155	207	12	Glutathione S-Transferase Domain
10	4GLT_A	13.9	2.6	153	208	8	Glutathione S-Transferase-Like Protein
11	5FT3_B	13.7	2.7	154	221	16	Glutathione S-Transferase Epsilon 2
12	4O92_A	13.7	2.8	147	196	12	Glutathione S-Transferase
13	5U51_C	13.6	2.5	151	195	11	Stringent Starvation Protein A
14	4MPG_B	13.6	2.6	156	252	12	Glutathione S-Transferase Theta-2
15	4XT0_A	13.5	2.3	152	243	14	Protein Ligf
16	6F70_A	13.4	2.6	157	242	12	Glutathione Transferase
17	1OYJ_C	13.4	2.8	157	227	11	Glutathione S-Transferase
18	3RBT_D	13.2	2.6	155	242	14	Glutathione Transferase O1
19	5O84_A	13.1	2.9	153	216	14	Glutathione S-Transferase U23
20	4MDC_B	13.0	2.7	153	233	11	Putative Glutathione S-Transferase
21	4IW9_B	13.0	2.8	159	215	12	Glutathione Transferase
22	3M8N_D	12.9	2.7	149	222	12	Possible Glutathione S-Transferase
23	4AGS_A	12.9	2.9	157	445	8	Thiol-Dependent Reductase 1
24	4EXJ_B	12.9	2.9	153	222	17	Uncharacterized Protein
25	3LXT_D	12.3	3.2	152	205	7	Glutathione S-Transferase
26	4JBB_A	12.3	2.5	145	208	10	Putative Glutathione S-Transferase
27	3LG6_A	12.3	2.7	152	229	11	Putative Glutathione Transferase

28	3LXZ_A	12.2	3.2	153	218	10	Glutathione S-Transferase Family Protein
29	5UUN_A	12.1	2.9	161	286	10	Glutathione S-Transferase-Like Protein
30	4BVY_A	12.1	2.6	144	211	10	Methionine-tRNA ligase, Cytoplasmic
31	2YCD_A	12.0	2.6	147	213	13	Glutathione S-Transferase
32	2HQT_J	11.9	2.5	98	121	19	Gu4 Nucleic-Binding Protein 1
33	2HRA_A	11.8	2.9	144	180	9	Glutamyl-tRNA Synthetase, Cytoplasmic
34	4KF9_A	11.7	2.6	151	318	10	Glutathione S-Transferase Protein
35	4G10_A	11.7	2.9	154	263	7	Glutathione S-Transferase Homolog
36	4AI6_A	11.7	3.1	150	2650	13	Glutathione S-Transferase Class-Mu 26 kDa Isozyme
37	2WB9_A	11.5	2.8	147	211	12	Glutathione Transferase Sigma Class
38	5LKD_A	11.4	3.0	160	352	9	Glutathione S-Transferase Omega-Like 2
39	4PTS_B	11.3	3.1	158	333	11	Glutathione S-Transferase
40	5A5H_D	11.3	3.3	151	213	7	Bifunctional Glutamate/Proline-tRNA Ligase
41	5XFT_A	11.1	2.7	142	215	8	Dehydroascorbate Reductase
42	3H1N_B	11.0	3.4	154	242	11	Probable Glutathione S-Transferase
43	5U51_A	10.7	2.9	142	203	13	Stringent Starvation Protein A
44	4W66_A	10.7	3.0	153	236	8	Glutathione S-Transferase Domain Protein
45	2D2Z_B	10.6	2.9	145	240	12	Chloride Intracellular Channel Protein 4
46	1Z9H_A	10.2	2.8	145	274	10	Membrane-Associated Prostaglandin E Synthase-2
47	4LMV_A	10.0	3.0	148	252	9	Glutathione Transferase
48	2YV9_A	10.0	3.1	154	285	11	Chloride Intracellular Channel Exc-4
49	2FNO_A	9.9	3.5	153	239	6	Agr_Pat_752p
50	3IR4_A	9.2	3.5	141	218	11	Glutaredoxin 2

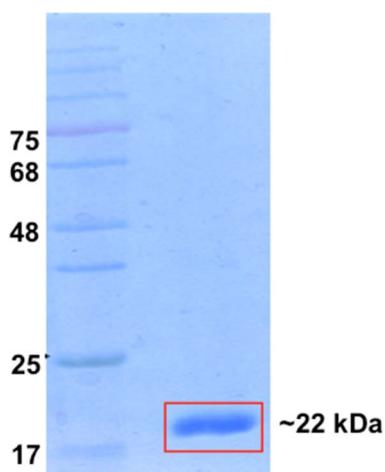
Supplementary Table S4. Results of DALI search for *Pvp43* CTD (residues 241-405) against PDB25 (a subset of Protein Data Bank). The structure was specified by concatenating string of the PDB identifier (4 characters) and a chain identifier (1 character). The match list is sorted by Z-scores (Z-scores between 8 and 20 means the two are probably homologous). The other listed parameters in the Table S3 are: root-mean-square deviation (RMSD) of C^a atoms in superimposition; Lali (number of structurally equivalent positions; and N_{res} - length of the structural neighbor and %id - percentage amino acid identity in aligned positions.

No	PDB_Chain	Z-Score	RMSD	Lali	N _{res}	%	Description
1	1NTG_A	26.5	1.3	162	171	38	Tyrosyl-tRNA Synthetase
2	4R1J_A	25.5	1.4	162	174	31	Gu4 Nucleic-Binding Protein 1
3	5H34_B	14.0	1.7	103	112	36	Methionine-tRNA Ligase
4	2Q2H_A	12.1	2.3	104	118	23	Secretion Chaperone, Phage-display derived
5	3PCO_B	9.7	3.0	102	795	28	Phenylalanyl-tRNA synthetase, Alpha Subunit
6	3BU2_A	9.3	3.4	96	195	28	Putative tRNA-binding Protein

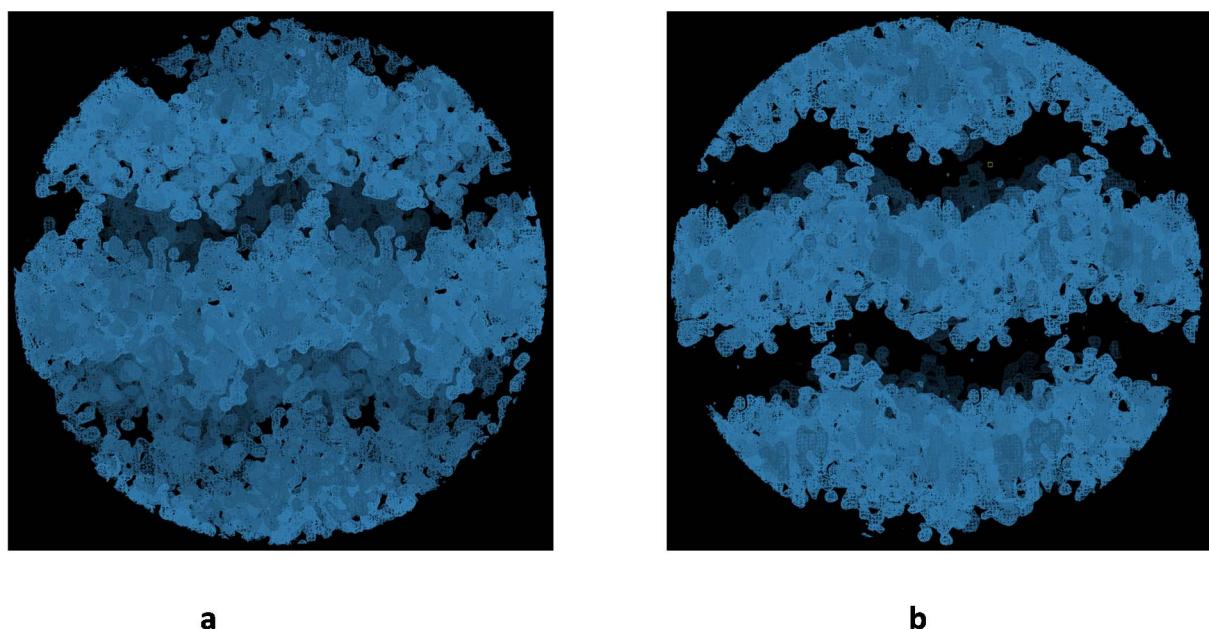


Supplementary Fig. S1. Snapshots of mounted *Pvp43* NTD and *Pvp43* CTD crystals.

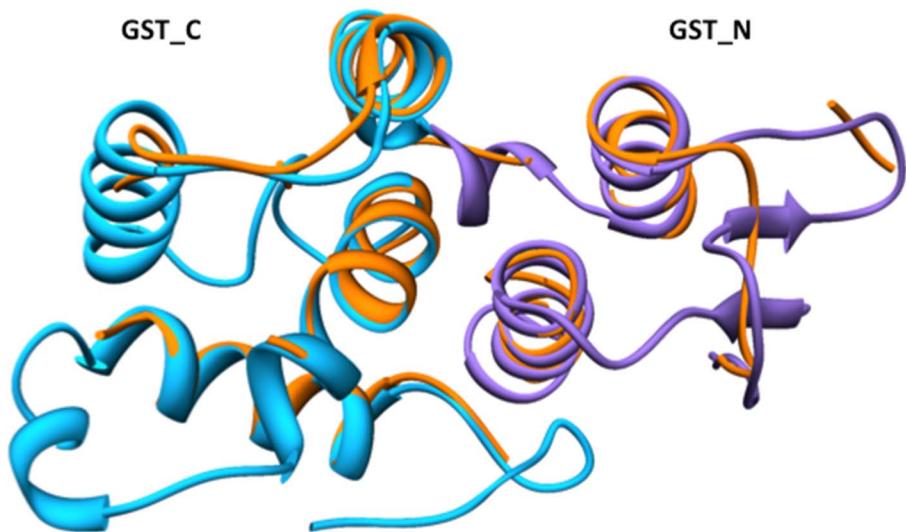
The beam shape x is marked as red elliptical circle and beam size (μm) is marked at lower right and values of the beam size is listed in Table 2. (a) Crystal obtained for the *Pvp43* full-length protein was cryo-soaked in 200 mM CoCl₂ salt for 30 s (Form 1 SG *P4*₁*2*₁₂). (b) Native crystal of the *Pvp43* NTD (Form 2 SG *P2*₁) and (c) The *Pvp43* NTD in complex with GSH (Form 3 SG *P2*₁*2*₁₂₁). (d) Form 1 (SG *P4*₃) and (e) Form 2 (SG *P2*₁*2*₁₂₁) crystals of the *Pvp43* CTD protein.



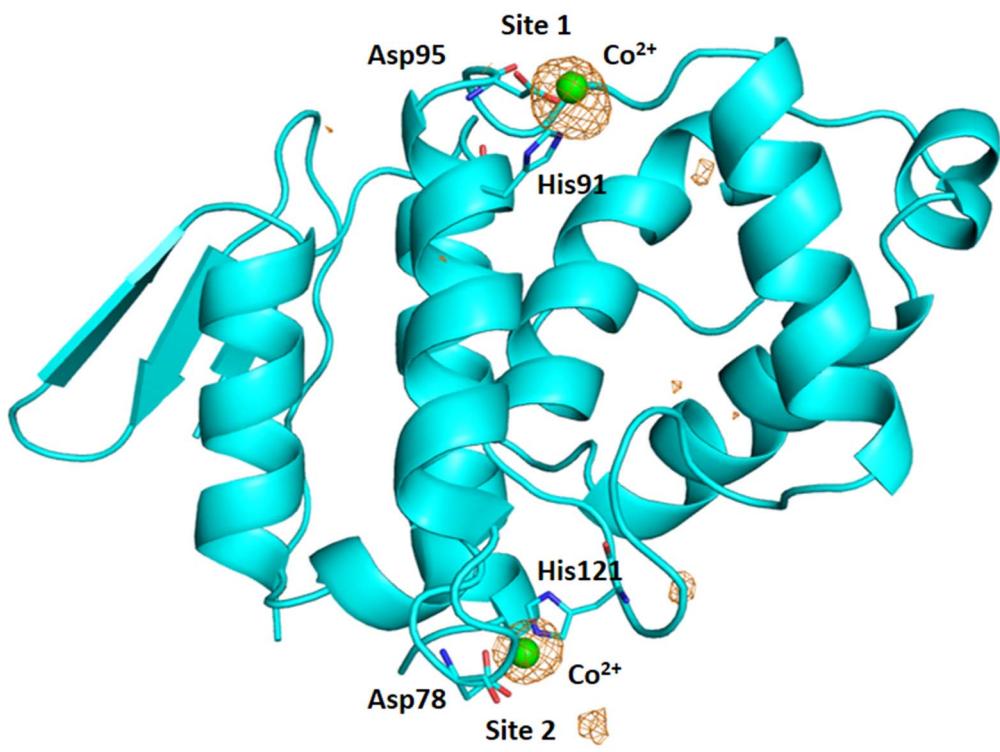
Supplementary Fig. S2. SDS PAGE analysis of the crystallized fragment size of the full-length *Pvp43* protein.



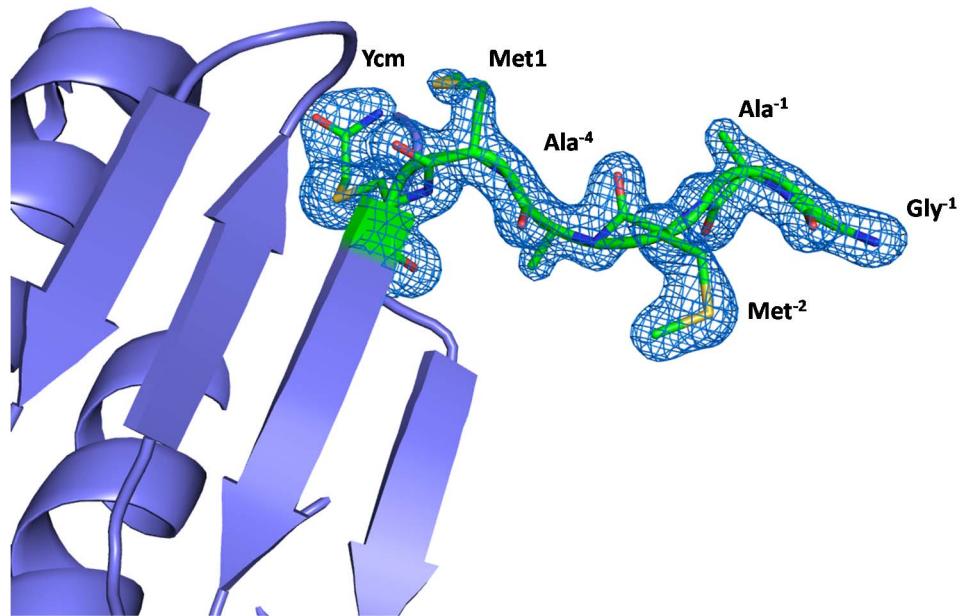
Supplementary Fig. S3. Map contrast after phasing and density modification for (a) Co-SAD data set x10_3 and (b) merged Co SAD data set. The map was contoured at 1σ level.



Supplementary Fig. S4. The partially built model (orange) from Co-SAD data was overlapped with the final refined model of *Pvp43* NTD. The N-terminal thioredoxin fold (GST-N, purple) and C-terminal all α -helical domain (GST-C, blue) are labeled.



Supplementary Fig. S5. The anomalous difference Fourier map contoured at 4σ level is superimposed on final refined structure of *Pvp43* NTD from Co SAD data. The green sphere represents the bound Co^{2+} ions and coordinating residues are labeled.



Supplementary Fig. S6. The ordered linker residues (GAMA) between the 6X-His tag and the *Pvp43* NTD (orthorhombic form, PDB ID: 5ZKE_B). The final $2F_o - F_c$ map was contoured at 1.0σ level. The N-terminal tag residues are: MKHHHHHHHPMSDYDIPTTENLYFEGAMA-*Pvp43* NTD.

EMAPII-YRS-C_1NTG	EPEEVIPSRLDIRVGKIIITVEKHPDADSLYVEKIDVG-EAEPRTVVSGLVQFVPKEELQD
AIMP1/P43_4R3Z	DSKPIDVSRLLDRIGCIITARKHPDADSLYVEEVDVG-EIAPIRTVVSGLVNVHVPLEQMQN
Arc1p-C_4R1J	MPEPKPKPSAIDFRVGFIQKAIKHPDADSLYVSTIDVGDEEGPRTVCSGLVKHFPLDAMQE
CTD-Pvp43_6IPA	---LDDITRLNIIIVGYVEEVEIHPDADTLYCLKINVG-EEKSRDICSGLRLKKNSEDLLH : : : : * : . * * * : ** : : ** * . * : *** : : .
EMAPII-YRS-C_1NTG	RLVVVLCNLKPQKMRGVESQGMILLCASIEGINRQVEPLDPPAGSAPGEHVFKGYEKQQ-
AIMP1/P43_4R3Z	RMVILLCNLKPAKMRGVLSQAMVMCASSP---EKIEILAPPNGSVPGDRITFDAFP-GE-
Arc1p-C_4R1J	RYVVVVCNLKPVNMRGIKSTAMVLCGSND---DKVEFVEPPKDSKAGDKVFFEGFGDEA-
CTD-Pvp43_6IPA	KYVLVLANLKEKSILRGRKSHGMVLCGSFG---EQIELLAPPDGVNVERIICENMDVNKL : * : : . * * . : * : ; * . * : * : * . * : : .
EMAPII-YRS-C_1NTG	PDEE--LKPKKKVFEKLQADFKISEECIAQWKQ-----TNFMTKLGSISCKSLKGNNI
AIMP1/P43_4R3Z	PDKE--LNPKKKIWEQIQPDLHTNDECVATYKG-----VPFEVKG-KGVCRAQTMSNS
Arc1p-C_4R1J	PMKQ--LNPKKKIWEHLQPHFTTNDGLEVIFKDEEEKDHPVRKLTNAKGERFKVASIANA
CTD-Pvp43_6IPA	PDKTLSDKEKNPFFHIQPHLLVKNG-VAHYKD-----AKWLSSKGEITCPLEQGTIS * . : . : * : : * . : . : . : *
EMAPII-YRS-C_1NTG	SLE-----TNFMTKLGSISCKSLKGNNISLE
AIMP1/P43_4R3Z	GIK-----VPFEVKG-KGVCRAQTMSNSGIK
Arc1p-C_4R1J	QVREKDHPVRKLTNAKGERFKVASIANAQVR
CTD-Pvp43_6IPA	-----AKWLSSKGEITCPLEQGTIS---

Supplementary Fig. S7. Sequence alignment of C-terminal tRNA-binding domains. Proteins are: EMAP II-like domain of tyrosyl-tRNA synthetase from *H. sapiens*, the C-terminal tRNA binding domain of AIMP1/p43 from *H. sapiens*, the C-terminal tRNA binding domain of Arc1p-C from *S. cerevisiae* and the C-terminal tRNA-biding domain of Pvp43. Identical/well conserved residues; conserved residues and semi-conserved residues are marked with asterisks, semicolons and dots, respectively.

References

- Pape, T. & Schneider, T. R. (2004). *Journal of Applied Crystallography*. 37, 843–844.
 Zwart, P. H., Grosse-Kunstleve, R. W., Adams, P. D. & Zwart, P. H., Grosse-Kunstleve, R. W., Adams, P. D. (2005). *CCP4 News*.