



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 12:18 pm BST

PDB ID : 4N5C  
Title : Crystal structure of Ypp1  
Authors : Wu, X.; Chi, R.J.; Baskin, J.M.; Lucast, L.; Burd, C.G.; De Camilli, P.;  
Reinisch, K.M.  
Deposited on : 2013-10-09  
Resolution : 3.25 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

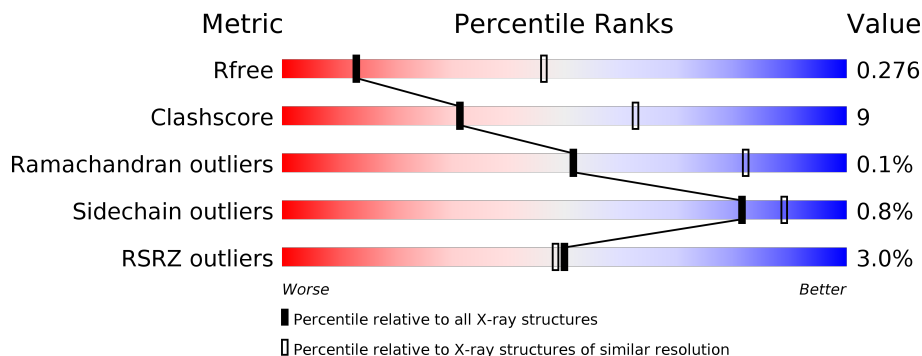
MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:  
*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.25 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



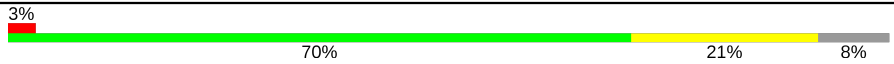

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	802	
1	B	802	
1	C	802	
1	D	802	
1	E	802	
1	F	802	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	802	 <p>3% 70% 21% 8%</p>
1	H	802	 <p>4% 68% 23% 8%</p>

## 2 Entry composition i

There is only 1 type of molecule in this entry. The entry contains 48674 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Cargo-transport protein YPP1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	745	Total 6138	C 3963	N 987	O 1160	S 11	Se 17	0	0	0
1	B	741	Total 6109	C 3948	N 982	O 1151	S 11	Se 17	0	0	0
1	C	737	Total 6080	C 3930	N 977	O 1145	S 11	Se 17	0	0	0
1	D	757	Total 6241	C 4031	N 1003	O 1179	S 11	Se 17	0	0	0
1	E	727	Total 6000	C 3884	N 967	O 1123	S 11	Se 15	0	0	0
1	F	727	Total 5993	C 3872	N 967	O 1128	S 11	Se 15	0	0	0
1	G	735	Total 6056	C 3910	N 977	O 1143	S 11	Se 15	0	0	0
1	H	734	Total 6057	C 3916	N 976	O 1138	S 11	Se 16	0	0	0

There are 120 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	6	GLY	-	EXPRESSION TAG	UNP P46951
A	7	PRO	-	EXPRESSION TAG	UNP P46951
A	8	LEU	-	EXPRESSION TAG	UNP P46951
A	9	GLY	-	EXPRESSION TAG	UNP P46951
A	10	SER	-	EXPRESSION TAG	UNP P46951
A	?	-	LEU	DELETION	UNP P46951
A	?	-	ASP	DELETION	UNP P46951
A	?	-	LYS	DELETION	UNP P46951
A	?	-	LYS	DELETION	UNP P46951
A	?	-	PRO	DELETION	UNP P46951
A	?	-	GLY	DELETION	UNP P46951
A	?	-	LYS	DELETION	UNP P46951
A	?	-	ARG	DELETION	UNP P46951

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	ALA	DELETION	UNP P46951
A	?	-	LYS	DELETION	UNP P46951
B	6	GLY	-	EXPRESSION TAG	UNP P46951
B	7	PRO	-	EXPRESSION TAG	UNP P46951
B	8	LEU	-	EXPRESSION TAG	UNP P46951
B	9	GLY	-	EXPRESSION TAG	UNP P46951
B	10	SER	-	EXPRESSION TAG	UNP P46951
B	?	-	LEU	DELETION	UNP P46951
B	?	-	ASP	DELETION	UNP P46951
B	?	-	LYS	DELETION	UNP P46951
B	?	-	LYS	DELETION	UNP P46951
B	?	-	PRO	DELETION	UNP P46951
B	?	-	GLY	DELETION	UNP P46951
B	?	-	LYS	DELETION	UNP P46951
B	?	-	ARG	DELETION	UNP P46951
B	?	-	ALA	DELETION	UNP P46951
B	?	-	LYS	DELETION	UNP P46951
C	6	GLY	-	EXPRESSION TAG	UNP P46951
C	7	PRO	-	EXPRESSION TAG	UNP P46951
C	8	LEU	-	EXPRESSION TAG	UNP P46951
C	9	GLY	-	EXPRESSION TAG	UNP P46951
C	10	SER	-	EXPRESSION TAG	UNP P46951
C	?	-	LEU	DELETION	UNP P46951
C	?	-	ASP	DELETION	UNP P46951
C	?	-	LYS	DELETION	UNP P46951
C	?	-	LYS	DELETION	UNP P46951
C	?	-	PRO	DELETION	UNP P46951
C	?	-	GLY	DELETION	UNP P46951
C	?	-	LYS	DELETION	UNP P46951
C	?	-	ARG	DELETION	UNP P46951
C	?	-	ALA	DELETION	UNP P46951
C	?	-	LYS	DELETION	UNP P46951
D	6	GLY	-	EXPRESSION TAG	UNP P46951
D	7	PRO	-	EXPRESSION TAG	UNP P46951
D	8	LEU	-	EXPRESSION TAG	UNP P46951
D	9	GLY	-	EXPRESSION TAG	UNP P46951
D	10	SER	-	EXPRESSION TAG	UNP P46951
D	?	-	LEU	DELETION	UNP P46951
D	?	-	ASP	DELETION	UNP P46951
D	?	-	LYS	DELETION	UNP P46951
D	?	-	LYS	DELETION	UNP P46951
D	?	-	PRO	DELETION	UNP P46951

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	?	-	GLY	DELETION	UNP P46951
D	?	-	LYS	DELETION	UNP P46951
D	?	-	ARG	DELETION	UNP P46951
D	?	-	ALA	DELETION	UNP P46951
D	?	-	LYS	DELETION	UNP P46951
E	6	GLY	-	EXPRESSION TAG	UNP P46951
E	7	PRO	-	EXPRESSION TAG	UNP P46951
E	8	LEU	-	EXPRESSION TAG	UNP P46951
E	9	GLY	-	EXPRESSION TAG	UNP P46951
E	10	SER	-	EXPRESSION TAG	UNP P46951
E	?	-	LEU	DELETION	UNP P46951
E	?	-	ASP	DELETION	UNP P46951
E	?	-	LYS	DELETION	UNP P46951
E	?	-	LYS	DELETION	UNP P46951
E	?	-	PRO	DELETION	UNP P46951
E	?	-	GLY	DELETION	UNP P46951
E	?	-	LYS	DELETION	UNP P46951
E	?	-	ARG	DELETION	UNP P46951
E	?	-	ALA	DELETION	UNP P46951
E	?	-	LYS	DELETION	UNP P46951
F	6	GLY	-	EXPRESSION TAG	UNP P46951
F	7	PRO	-	EXPRESSION TAG	UNP P46951
F	8	LEU	-	EXPRESSION TAG	UNP P46951
F	9	GLY	-	EXPRESSION TAG	UNP P46951
F	10	SER	-	EXPRESSION TAG	UNP P46951
F	?	-	LEU	DELETION	UNP P46951
F	?	-	ASP	DELETION	UNP P46951
F	?	-	LYS	DELETION	UNP P46951
F	?	-	LYS	DELETION	UNP P46951
F	?	-	PRO	DELETION	UNP P46951
F	?	-	GLY	DELETION	UNP P46951
F	?	-	LYS	DELETION	UNP P46951
F	?	-	ARG	DELETION	UNP P46951
F	?	-	ALA	DELETION	UNP P46951
F	?	-	LYS	DELETION	UNP P46951
G	6	GLY	-	EXPRESSION TAG	UNP P46951
G	7	PRO	-	EXPRESSION TAG	UNP P46951
G	8	LEU	-	EXPRESSION TAG	UNP P46951
G	9	GLY	-	EXPRESSION TAG	UNP P46951
G	10	SER	-	EXPRESSION TAG	UNP P46951
G	?	-	LEU	DELETION	UNP P46951
G	?	-	ASP	DELETION	UNP P46951

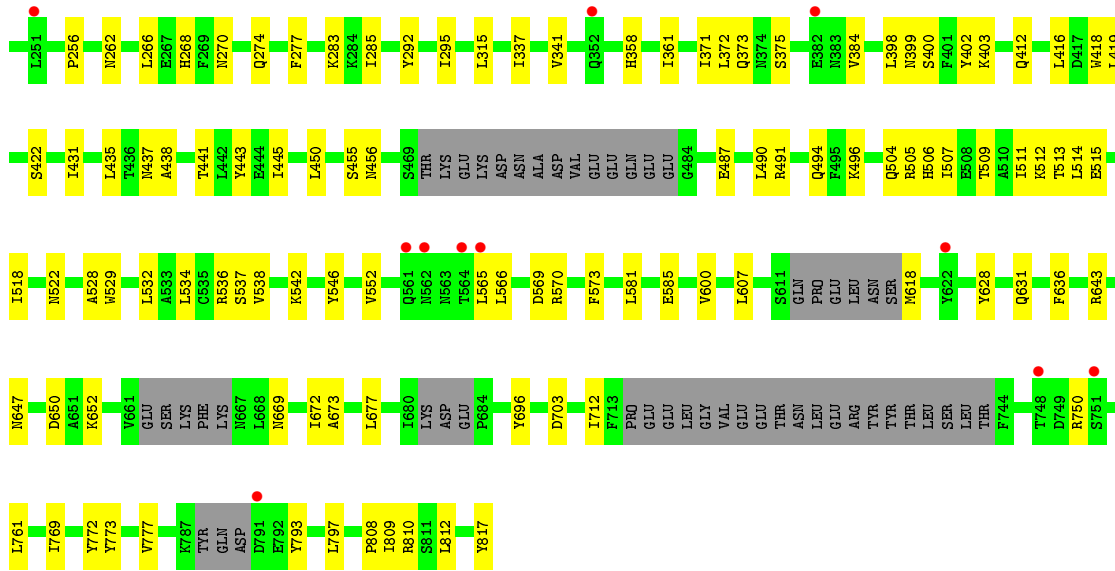
*Continued on next page...*

*Continued from previous page...*

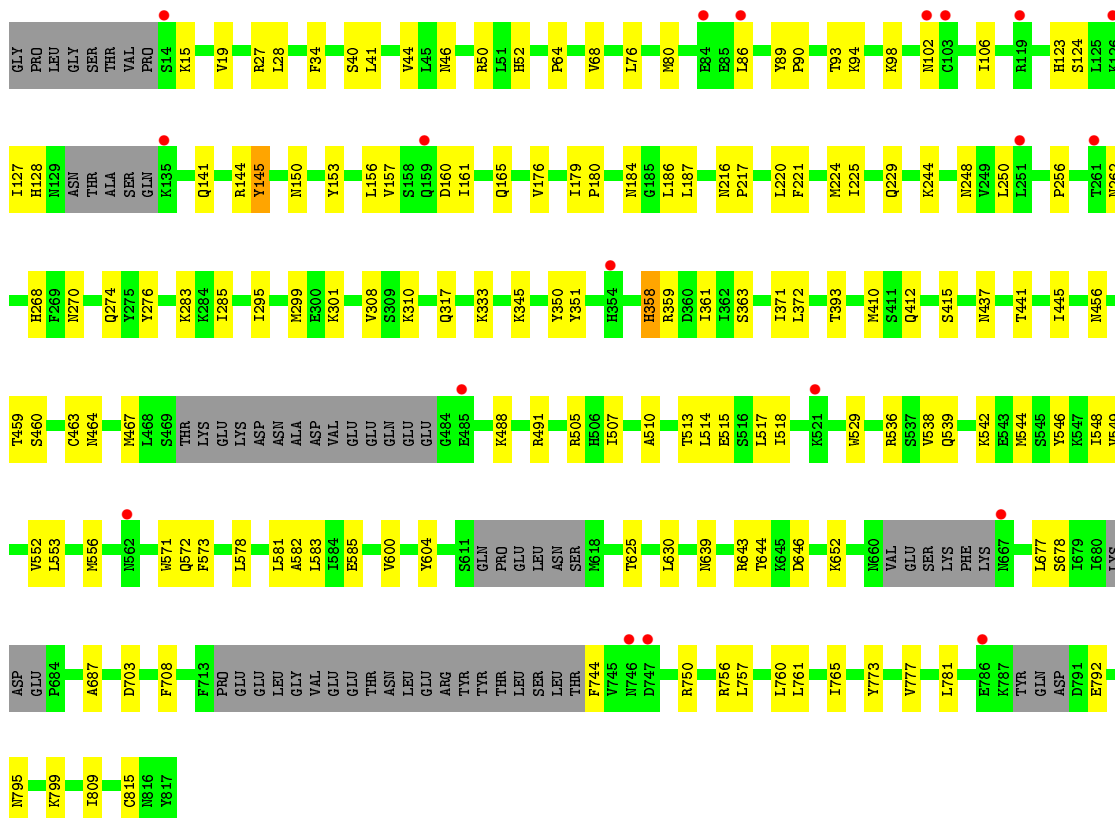
Chain	Residue	Modelled	Actual	Comment	Reference
G	?	-	LYS	DELETION	UNP P46951
G	?	-	LYS	DELETION	UNP P46951
G	?	-	PRO	DELETION	UNP P46951
G	?	-	GLY	DELETION	UNP P46951
G	?	-	LYS	DELETION	UNP P46951
G	?	-	ARG	DELETION	UNP P46951
G	?	-	ALA	DELETION	UNP P46951
G	?	-	LYS	DELETION	UNP P46951
H	6	GLY	-	EXPRESSION TAG	UNP P46951
H	7	PRO	-	EXPRESSION TAG	UNP P46951
H	8	LEU	-	EXPRESSION TAG	UNP P46951
H	9	GLY	-	EXPRESSION TAG	UNP P46951
H	10	SER	-	EXPRESSION TAG	UNP P46951
H	?	-	LEU	DELETION	UNP P46951
H	?	-	ASP	DELETION	UNP P46951
H	?	-	LYS	DELETION	UNP P46951
H	?	-	LYS	DELETION	UNP P46951
H	?	-	PRO	DELETION	UNP P46951
H	?	-	GLY	DELETION	UNP P46951
H	?	-	LYS	DELETION	UNP P46951
H	?	-	ARG	DELETION	UNP P46951
H	?	-	ALA	DELETION	UNP P46951
H	?	-	LYS	DELETION	UNP P46951



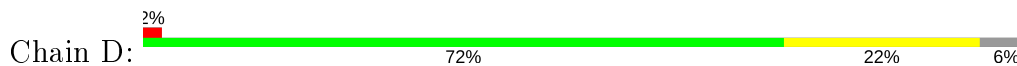




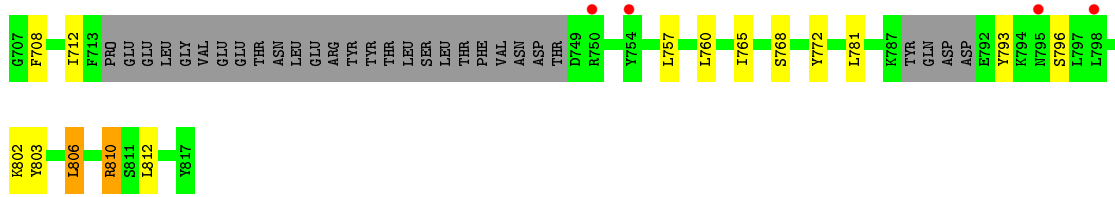
- Molecule 1: Cargo-transport protein YPP1



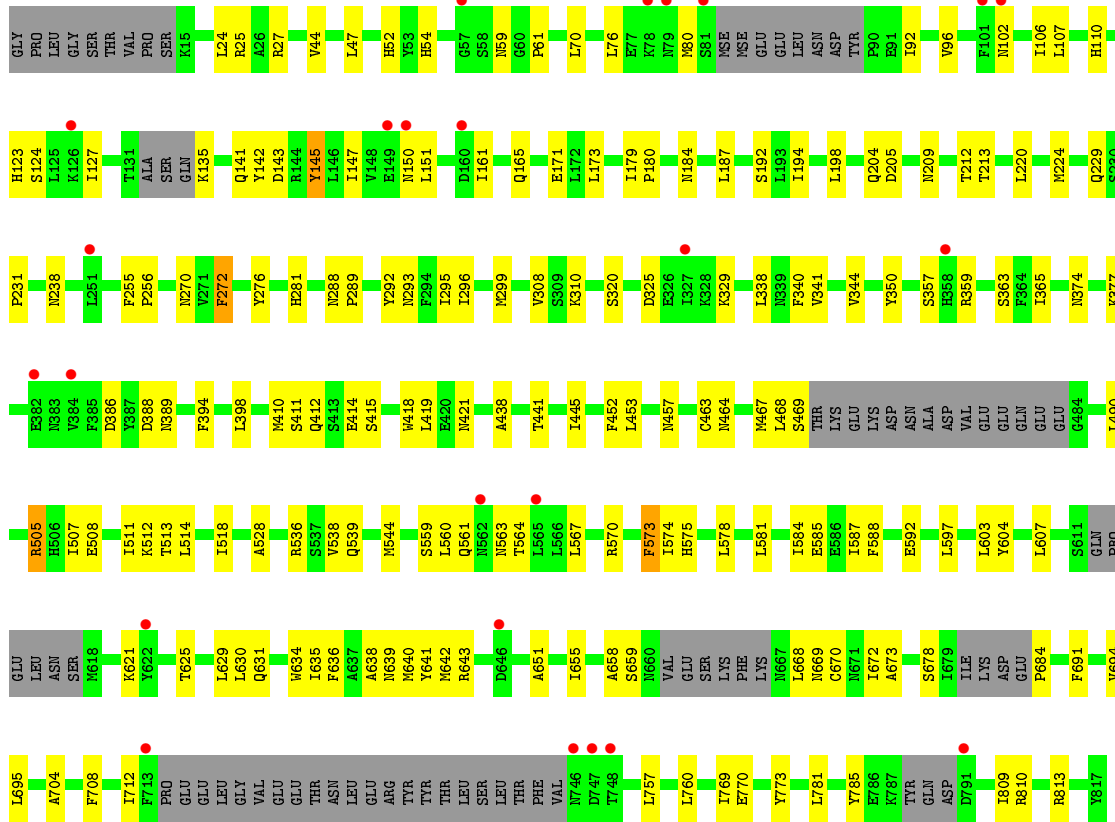
- Molecule 1: Cargo-transport protein YPP1



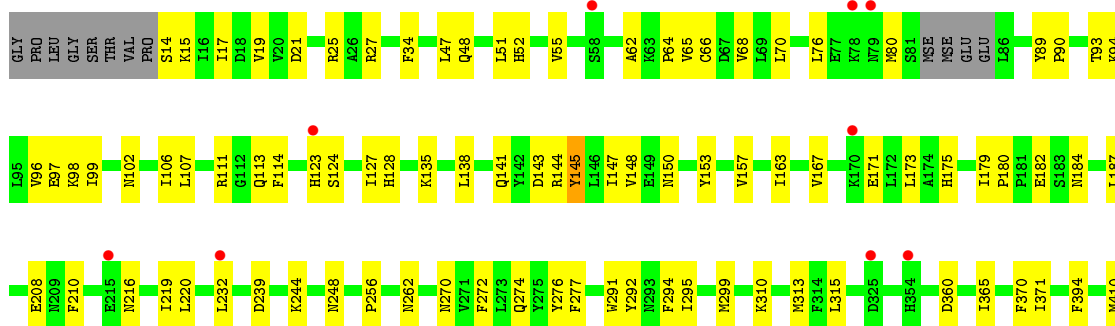


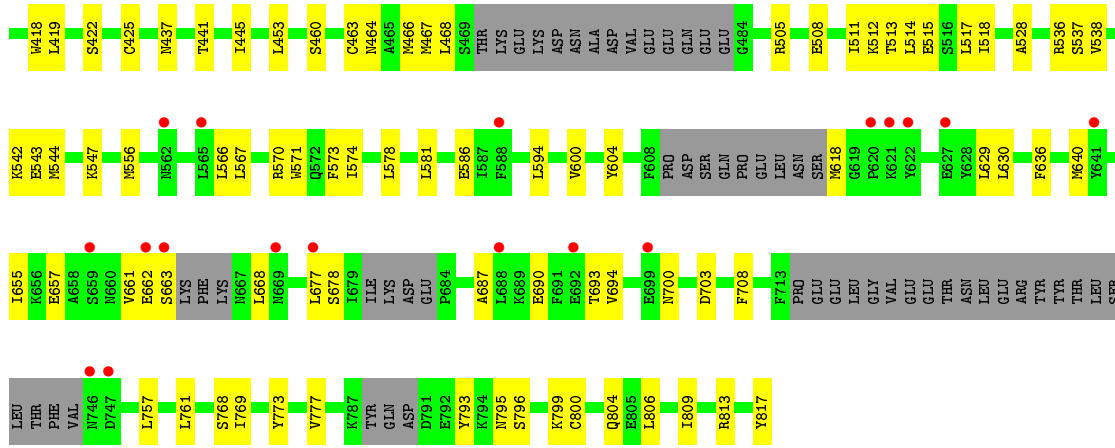


• Molecule 1: Cargo-transport protein YPP1

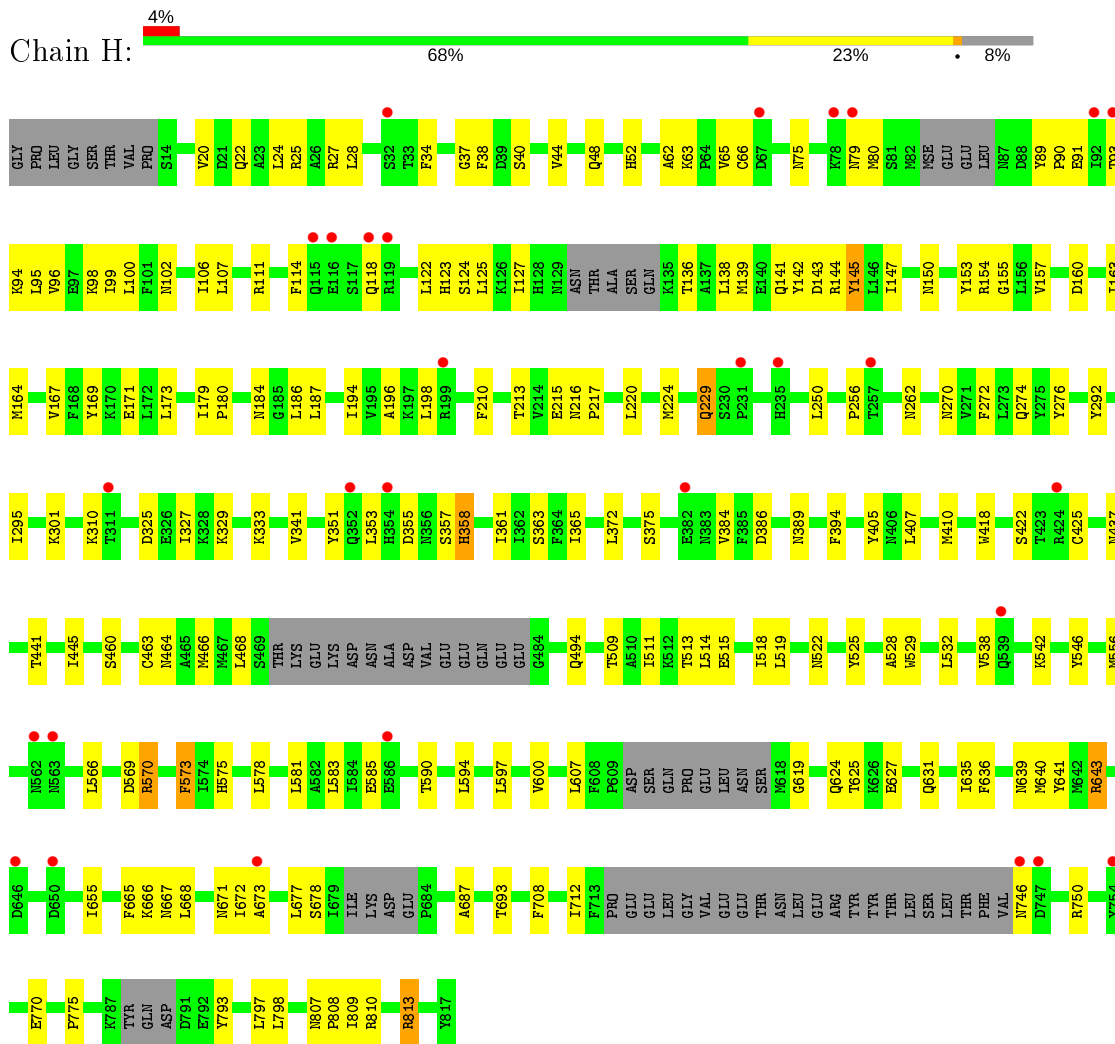


• Molecule 1: Cargo-transport protein YPP1





● Molecule 1: Cargo-transport protein YPP1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	97.55Å 136.65Å 154.12Å 76.69° 85.67° 72.74°	Depositor
Resolution (Å)	25.00 – 3.25 25.12 – 3.25	Depositor EDS
% Data completeness (in resolution range)	98.7 (25.00-3.25) 92.1 (25.12-3.25)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.44 (at 3.23Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.240 , 0.277 0.239 , 0.276	Depositor DCC
$R_{free}$ test set	1973 reflections (1.72%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	71.2	Xtrriage
Anisotropy	0.224	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 20.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	48674	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 32.33 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.5695e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.28	0/6252	0.49	1/8423 (0.0%)
1	B	0.28	0/6223	0.49	0/8382
1	C	0.28	0/6193	0.49	1/8340 (0.0%)
1	D	0.29	0/6359	0.51	1/8569 (0.0%)
1	E	0.27	0/6114	0.48	0/8234
1	F	0.27	0/6105	0.49	0/8223
1	G	0.29	0/6169	0.49	0/8312
1	H	0.29	0/6171	0.52	1/8310 (0.0%)
All	All	0.28	0/49586	0.49	4/66793 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	663	SER	N-CA-C	-5.50	96.14	111.00
1	C	756	ARG	NE-CZ-NH1	5.33	122.96	120.30
1	A	565	LEU	CA-CB-CG	5.31	127.51	115.30
1	H	353	LEU	CA-CB-CG	5.23	127.33	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	234	LYS	Peptide

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	6138	0	6060	112	0
1	B	6109	0	6038	101	1
1	C	6080	0	6008	99	0
1	D	6241	0	6171	115	0
1	E	6000	0	5943	104	1
1	F	5993	0	5927	114	0
1	G	6056	0	5984	115	0
1	H	6057	0	5993	124	0
All	All	48674	0	48124	873	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 9.

The worst 5 of 873 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:581:LEU:HD11	1:A:600:VAL:HG21	1.63	0.79
1:C:514:LEU:HD12	1:C:518:ILE:HD11	1.64	0.79
1:D:189:GLU:OE2	1:D:429:GLY:N	2.15	0.78
1:D:402:TYR:HB3	1:D:409:LEU:HD11	1.66	0.77
1:C:581:LEU:HD11	1:C:600:VAL:HG21	1.66	0.75

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:696:TYR:OH	1:E:671:ASN:OD1[1_655]	2.17	0.03

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	731/802 (91%)	714 (98%)	17 (2%)	0	100	100
1	B	725/802 (90%)	704 (97%)	21 (3%)	0	100	100
1	C	721/802 (90%)	701 (97%)	20 (3%)	0	100	100
1	D	747/802 (93%)	727 (97%)	19 (2%)	1 (0%)	51	82
1	E	709/802 (88%)	691 (98%)	18 (2%)	0	100	100
1	F	709/802 (88%)	688 (97%)	20 (3%)	1 (0%)	51	82
1	G	719/802 (90%)	699 (97%)	19 (3%)	1 (0%)	51	82
1	H	718/802 (90%)	700 (98%)	18 (2%)	0	100	100
All	All	5779/6416 (90%)	5624 (97%)	152 (3%)	3 (0%)	51	82

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	662	GLU
1	D	90	PRO
1	F	231	PRO

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	688/724 (95%)	683 (99%)	5 (1%)	84	90
1	B	685/724 (95%)	682 (100%)	3 (0%)	91	94

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	681/724 (94%)	677 (99%)	4 (1%)	86	91
1	D	701/724 (97%)	695 (99%)	6 (1%)	78	87
1	E	671/724 (93%)	664 (99%)	7 (1%)	76	85
1	F	671/724 (93%)	667 (99%)	4 (1%)	86	91
1	G	678/724 (94%)	672 (99%)	6 (1%)	78	87
1	H	678/724 (94%)	670 (99%)	8 (1%)	71	83
All	All	5453/5792 (94%)	5410 (99%)	43 (1%)	81	89

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	351	TYR
1	E	810	ARG
1	H	570	ARG
1	E	358	HIS
1	E	505	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	110	HIS
1	F	202	GLN
1	H	150	ASN
1	C	150	ASN
1	H	141	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	728/802 (90%)	-0.03	20 (2%) 54 51	26, 69, 112, 156	0
1	B	724/802 (90%)	-0.03	18 (2%) 57 53	35, 67, 107, 142	0
1	C	720/802 (89%)	-0.01	19 (2%) 56 52	30, 68, 106, 140	0
1	D	740/802 (92%)	-0.07	13 (1%) 68 65	27, 61, 102, 144	0
1	E	712/802 (88%)	-0.01	24 (3%) 45 42	30, 62, 101, 131	0
1	F	712/802 (88%)	0.05	24 (3%) 45 42	25, 69, 117, 155	0
1	G	720/802 (89%)	0.09	27 (3%) 40 37	27, 75, 116, 144	0
1	H	718/802 (89%)	0.14	29 (4%) 38 35	34, 77, 126, 152	0
All	All	5774/6416 (89%)	0.02	174 (3%) 50 48	25, 68, 112, 156	0

The worst 5 of 174 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	14	SER	5.9
1	D	130	ASN	5.6
1	G	746	ASN	5.4
1	F	160	ASP	5.3
1	E	79	ASN	5.0

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.