



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 16, 2024 – 11:39 AM EDT

PDB ID : 4WSI
Title : Crystal Structure of PALS1/Crb complex
Authors : Wei, Z.; Li, Y.; Zhang, M.
Deposited on : 2014-10-28
Resolution : 2.95 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

A user guide is available at

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

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
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)	:	2.37.1

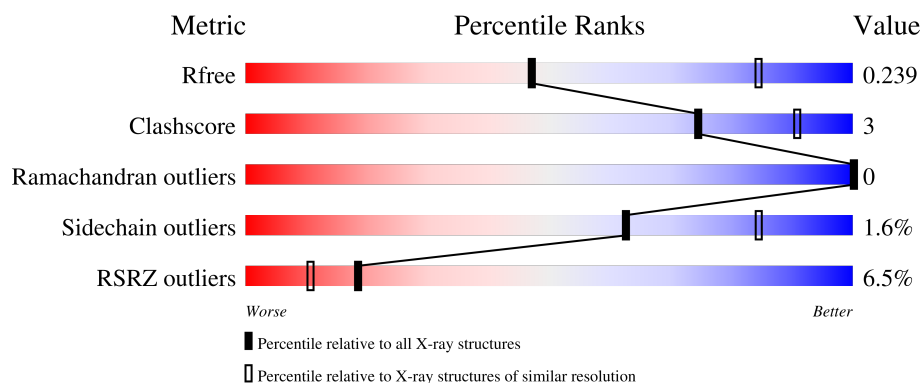
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	394	<div> <div>4%</div> <div>82%</div> <div>12%</div> <div>6%</div> </div>
1	B	394	<div> <div>8%</div> <div>85%</div> <div>6%</div> <div>9%</div> </div>
2	X	43	<div> <div>35%</div> <div>7%</div> <div>58%</div> </div>
2	Y	43	<div> <div>42%</div> <div>58%</div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6042 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 MAGUK p55 subfamily member 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	372	Total	C	N	O	S	0	1	0
			2947	1853	519	567	8			
1	B	359	Total	C	N	O	S	0	0	0
			2785	1753	484	540	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	232	GLY	-	expression tag	UNP Q8N3R9
A	233	PRO	-	expression tag	UNP Q8N3R9
A	234	GLY	-	expression tag	UNP Q8N3R9
A	235	SER	-	expression tag	UNP Q8N3R9
B	232	GLY	-	expression tag	UNP Q8N3R9
B	233	PRO	-	expression tag	UNP Q8N3R9
B	234	GLY	-	expression tag	UNP Q8N3R9
B	235	SER	-	expression tag	UNP Q8N3R9

- Molecule 2 is a protein called Protein crumbs.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	18	Total	C	N	O	S	0	0	0
			150	95	26	28	1			
2	Y	18	Total	C	N	O	S	0	0	0
			140	88	25	26	1			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	2104	GLY	-	expression tag	UNP P10040
X	2105	PRO	-	expression tag	UNP P10040
X	2106	GLY	-	expression tag	UNP P10040
X	2107	SER	-	expression tag	UNP P10040

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
X	2108	GLU	-	expression tag	UNP P10040
X	2109	PHE	-	expression tag	UNP P10040
Y	2104	GLY	-	expression tag	UNP P10040
Y	2105	PRO	-	expression tag	UNP P10040
Y	2106	GLY	-	expression tag	UNP P10040
Y	2107	SER	-	expression tag	UNP P10040
Y	2108	GLU	-	expression tag	UNP P10040
Y	2109	PHE	-	expression tag	UNP P10040

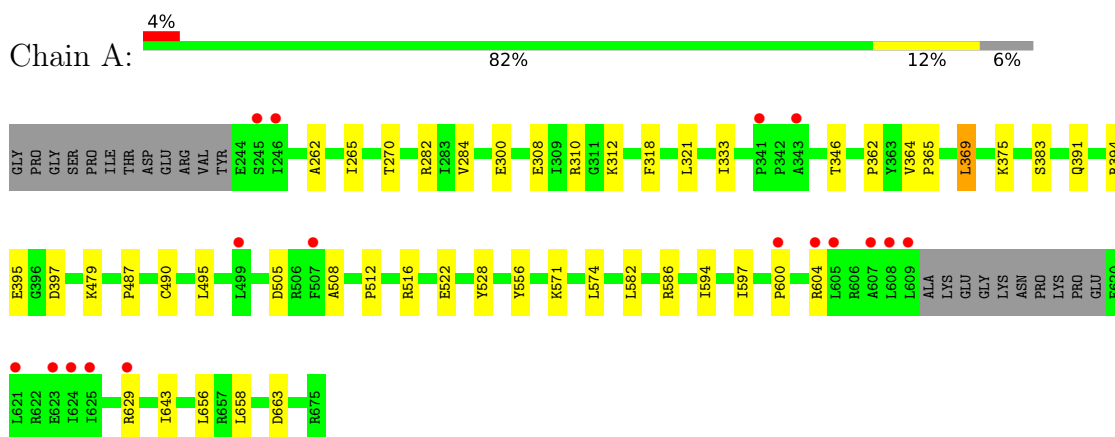
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	15	Total	O	0	0
			15	15		
3	B	4	Total	O	0	0
			4	4		
3	X	1	Total	O	0	0
			1	1		

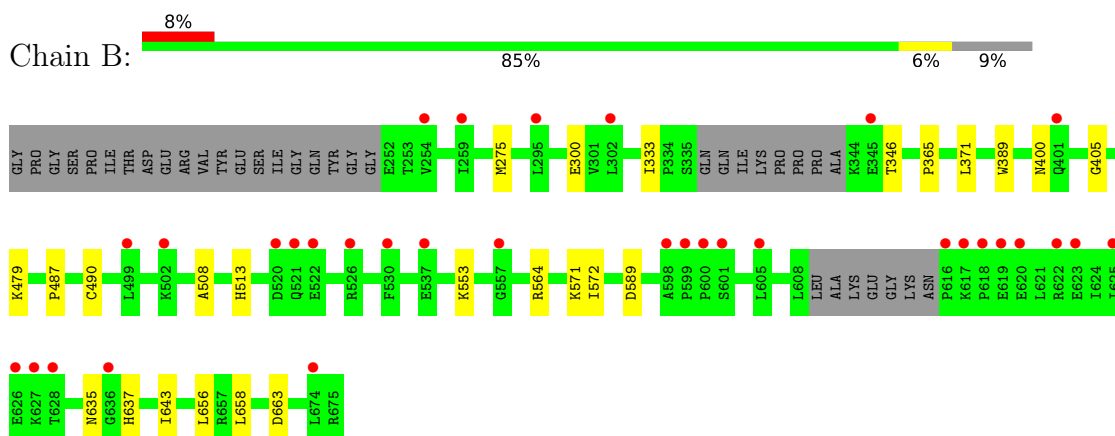
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

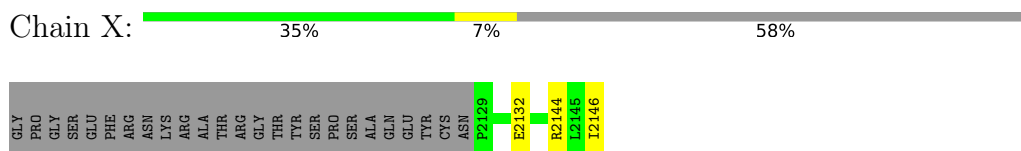
- Molecule 1: MAGUK p55 subfamily member 5



- Molecule 1: MAGUK p55 subfamily member 5



- Molecule 2: Protein crumbs



- Molecule 2: Protein crumbs

Chain Y:

42%

58%

GLY	PRO	GLY	SER	GLU	PHE	ARG	ASN	LYS	ARG	ALA	THR	ARG	GLY	THR	TYR	SER	PRO	SER	ALA	GLN	GLU	TYR	CYS	ASN	P2129	12146
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------	-------

4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	111.93Å 111.93Å 223.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.52 – 2.95 49.91 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.2 (41.52-2.95) 94.2 (49.91-2.95)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.57 (at 2.96Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.207 , 0.235 0.211 , 0.239	Depositor DCC
R_{free} test set	1523 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	87.7	Xtriage
Anisotropy	0.255	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 43.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6042	wwPDB-VP
Average B, all atoms (Å ²)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.23	0/3005	0.42	0/4071
1	B	0.22	0/2839	0.40	0/3853
2	X	0.22	0/153	0.41	0/205
2	Y	0.21	0/143	0.41	0/193
All	All	0.22	0/6140	0.41	0/8322

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	2947	0	2875	26	0
1	B	2785	0	2671	12	0
2	X	150	0	157	4	0
2	Y	140	0	135	0	0
3	A	15	0	0	1	0
3	B	4	0	0	0	0
3	X	1	0	0	1	0
All	All	6042	0	5838	38	0

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 3.

The worst 5 of 38 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:318:PHE:HB2	2:X:2144:ARG:HG2	1.74	0.69
1:A:586:ARG:NH1	3:A:714:HOH:O	2.36	0.59
1:B:479:LYS:NZ	1:B:663:ASP:OD1	2.32	0.57
1:B:371:LEU:HB2	1:B:405:GLY:HA3	1.88	0.56
1:A:508:ALA:HB2	1:A:571:LYS:HE3	1.87	0.55

There are no symmetry-related clashes.

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	369/394 (94%)	354 (96%)	15 (4%)	0	100	100
1	B	353/394 (90%)	339 (96%)	14 (4%)	0	100	100
2	X	16/43 (37%)	16 (100%)	0	0	100	100
2	Y	16/43 (37%)	16 (100%)	0	0	100	100
All	All	754/874 (86%)	725 (96%)	29 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	317/346 (92%)	311 (98%)	6 (2%)	57	81
1	B	295/346 (85%)	291 (99%)	4 (1%)	67	86
2	X	18/38 (47%)	18 (100%)	0	100	100
2	Y	15/38 (40%)	15 (100%)	0	100	100
All	All	645/768 (84%)	635 (98%)	10 (2%)	62	84

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	513	HIS
1	B	637	HIS
1	B	656	LEU
1	A	395	GLU
1	A	574	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides 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 ⓘ

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, 95th 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 ²)	Q < 0.9
1	A	372/394 (94%)	0.39	17 (4%) 32 20	44, 70, 153, 175	0
1	B	359/394 (91%)	0.53	33 (9%) 9 5	54, 93, 173, 200	0
2	X	18/43 (41%)	0.24	0 100 100	60, 72, 89, 110	0
2	Y	18/43 (41%)	-0.02	0 100 100	71, 126, 152, 157	0
All	All	767/874 (87%)	0.44	50 (6%) 18 11	44, 80, 160, 200	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	600	PRO	5.7
1	A	609	LEU	5.4
1	A	608	LEU	5.2
1	B	521	GLN	4.8
1	B	598	ALA	4.7

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 monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.