Full wwPDB X-ray Structure Validation Report (i)

Oct 11, 2023 – 01:25 PM EDT

PDB ID	:	7S68
Title	:	Structure of human PARP1 domains (Zn1, Zn3, WGR and HD) bound to a DNA double strand break.
Authors	:	Rouleau-Turcotte, E.; Pascal, J.M.
Deposited on	:	2021-09-13
Resolution	:	3.30 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i) 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 (1)) were used in the production of this report:

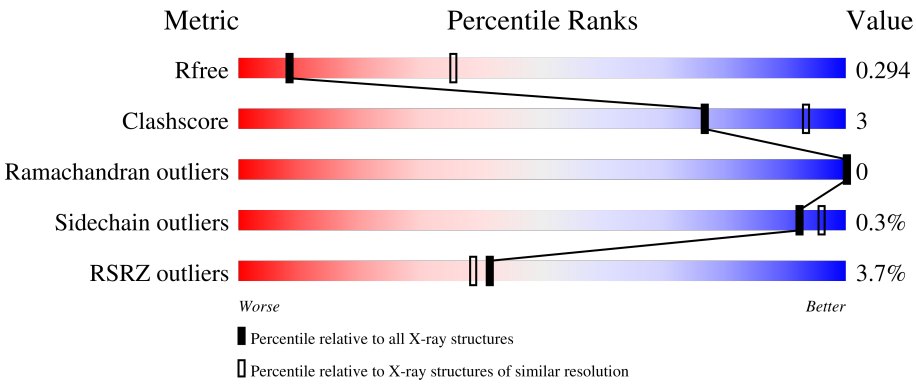
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 <=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	M	10	70% 30%
1	N	10	90% 10%
2	B	266	3% 73% 7% 21%
3	A	276	3% 74% 7% 19%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7428 atoms, of which 3590 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 DNA chain called DNA (5'-D(*GP*CP*CP*TP*GP*CP*AP*GP*GP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
1	M	10	315	96	113	39	58	9	0	0	0
1	N	10	315	96	113	39	58	9	0	0	0

- Molecule 2 is a protein called Poly [ADP-ribose] polymerase 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	211	3364	1076	1677	279	325	7	0	0	0

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	507	MET	-	initiating methionine	UNP P09874
B	508	GLY	-	expression tag	UNP P09874
B	509	SER	-	expression tag	UNP P09874
B	510	SER	-	expression tag	UNP P09874
B	511	HIS	-	expression tag	UNP P09874
B	512	HIS	-	expression tag	UNP P09874
B	513	HIS	-	expression tag	UNP P09874
B	514	HIS	-	expression tag	UNP P09874
B	515	HIS	-	expression tag	UNP P09874
B	516	HIS	-	expression tag	UNP P09874
B	517	SER	-	expression tag	UNP P09874
B	518	SER	-	expression tag	UNP P09874
B	519	GLY	-	expression tag	UNP P09874
B	520	LEU	-	expression tag	UNP P09874
B	521	VAL	-	expression tag	UNP P09874
B	522	PRO	-	expression tag	UNP P09874
B	523	ARG	-	expression tag	UNP P09874

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	524	GLY	-	expression tag	UNP P09874
B	525	SER	-	expression tag	UNP P09874
B	526	HIS	-	expression tag	UNP P09874
B	?	-	SER	deletion	UNP P09874
B	?	-	LYS	deletion	UNP P09874
B	?	-	LEU	deletion	UNP P09874
B	?	-	PRO	deletion	UNP P09874
B	?	-	LYS	deletion	UNP P09874
B	?	-	PRO	deletion	UNP P09874
B	?	-	VAL	deletion	UNP P09874
B	?	-	GLN	deletion	UNP P09874
B	?	-	ASP	deletion	UNP P09874
B	?	-	LEU	deletion	UNP P09874
B	?	-	ILE	deletion	UNP P09874
B	?	-	LYS	deletion	UNP P09874
B	?	-	MET	deletion	UNP P09874
B	?	-	ILE	deletion	UNP P09874
B	762	ALA	VAL	variant	UNP P09874

- Molecule 3 is a protein called Fusion of PARP1 zinc fingers 1 and 3 (Zn1, Zn3).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	A	223	3432	1103	1687	301	328	13	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP P09874
A	-18	GLY	-	expression tag	UNP P09874
A	-17	SER	-	expression tag	UNP P09874
A	-16	SER	-	expression tag	UNP P09874
A	-15	HIS	-	expression tag	UNP P09874
A	-14	HIS	-	expression tag	UNP P09874
A	-13	HIS	-	expression tag	UNP P09874
A	-12	HIS	-	expression tag	UNP P09874
A	-11	HIS	-	expression tag	UNP P09874
A	-10	HIS	-	expression tag	UNP P09874
A	-9	SER	-	expression tag	UNP P09874
A	-8	SER	-	expression tag	UNP P09874
A	-7	GLY	-	expression tag	UNP P09874
A	-6	LEU	-	expression tag	UNP P09874

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	VAL	-	expression tag	UNP P09874
A	-4	PRO	-	expression tag	UNP P09874
A	-3	ARG	-	expression tag	UNP P09874
A	-2	GLY	-	expression tag	UNP P09874
A	-1	SER	-	expression tag	UNP P09874
A	0	HIS	-	expression tag	UNP P09874

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

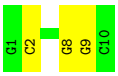
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
4	A	2	2	2	0	0

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: DNA (5'-D(*GP*CP*CP*TP*GP*CP*AP*GP*GP*C)-3')

Chain M:

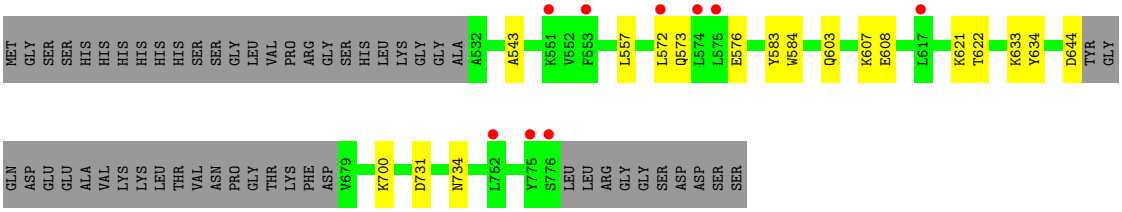


• Molecule 1: DNA (5'-D(*GP*CP*CP*TP*GP*CP*AP*GP*GP*C)-3')

Chain N:

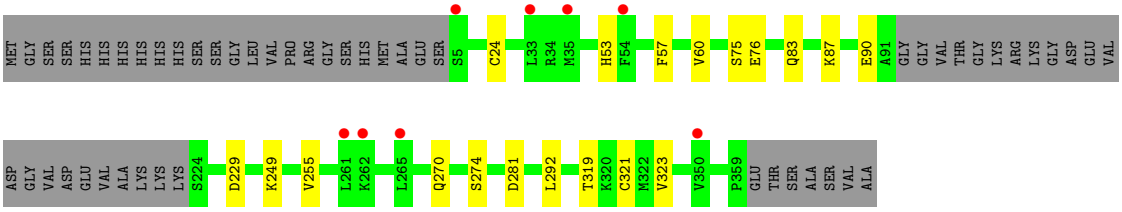
• Molecule 2: Poly [ADP-ribose] polymerase 1

Chain B:



• Molecule 3: Fusion of PARP1 zinc fingers 1 and 3 (Zn1, Zn3)

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	87.21Å 93.65Å 119.26Å 90.00° 106.29° 90.00°	Depositor
Resolution (Å)	47.41 – 3.30 47.41 – 3.30	Depositor EDS
% Data completeness (in resolution range)	90.2 (47.41-3.30) 90.2 (47.41-3.30)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.29 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.251 , 0.292 0.251 , 0.294	Depositor DCC
R_{free} test set	592 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	118.8	Xtriage
Anisotropy	0.073	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 95.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7428	wwPDB-VP
Average B, all atoms (Å^2)	169.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.11% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	M	0.49	0/226	0.79	0/347
1	N	0.53	0/226	0.82	0/347
2	B	0.25	0/1721	0.40	0/2318
3	A	0.24	0/1782	0.39	0/2401
All	All	0.29	0/3955	0.47	0/5413

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	M	202	113	113	2	0
1	N	202	113	113	1	0
2	B	1687	1677	1677	11	0
3	A	1745	1687	1687	12	0
4	A	2	0	0	0	0
All	All	3838	3590	3590	22	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.

All (22) 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)
2:B:633:LYS:NZ	2:B:634:TYR:O	2.23	0.72
3:A:75:SER:OG	3:A:76:GLU:OE2	2.15	0.64
3:A:24:CYS:SG	3:A:53:HIS:CE1	2.94	0.61
3:A:249:LYS:NZ	3:A:281:ASP:OD1	2.29	0.59
2:B:608:GLU:N	2:B:608:GLU:OE1	2.36	0.58
2:B:543:ALA:HB1	2:B:576:GLU:O	2.04	0.58
2:B:734:ASN:ND2	3:A:319:THR:HG22	2.22	0.55
1:N:23:DA:O5'	3:A:274:SER:OG	2.24	0.54
3:A:90:GLU:N	3:A:90:GLU:OE1	2.41	0.54
3:A:270:GLN:NE2	3:A:292:LEU:O	2.39	0.53
1:M:2:DC:OP2	2:B:621:LYS:NZ	2.29	0.51
3:A:255:VAL:HG12	3:A:255:VAL:O	2.12	0.50
2:B:557:LEU:HD13	2:B:622:THR:CG2	2.42	0.49
3:A:57:PHE:O	3:A:60:VAL:HG22	2.13	0.48
2:B:572:LEU:HD23	2:B:573:GLN:N	2.29	0.48
2:B:584:TRP:NE1	2:B:603:GLN:OE1	2.48	0.47
3:A:83:GLN:O	3:A:87:LYS:HG2	2.15	0.46
3:A:321:CYS:SG	3:A:323:VAL:HG23	2.56	0.46
1:M:8:DG:H4'	1:M:9:DG:OP1	2.18	0.44
2:B:731:ASP:HA	3:A:319:THR:HG21	2.01	0.42
2:B:583:TYR:CE2	2:B:607:LYS:HB3	2.55	0.42
2:B:644:ASP:HB2	2:B:700:LYS:HD2	2.04	0.40

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
2	B	207/266 (78%)	189 (91%)	18 (9%)	0	100 100
3	A	219/276 (79%)	203 (93%)	16 (7%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	426/542 (79%)	392 (92%)	34 (8%)	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	
2	B	186/230 (81%)	186 (100%)	0	100	100
3	A	187/243 (77%)	186 (100%)	1 (0%)	88	93
All	All	373/473 (79%)	372 (100%)	1 (0%)	92	96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	229	ASP

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 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	M	10/10 (100%)	-0.46	0	100	100	128, 135, 157, 161	0
1	N	10/10 (100%)	-0.20	0	100	100	117, 153, 160, 162	0
2	B	211/266 (79%)	0.14	9 (4%)	35	34	102, 169, 213, 249	0
3	A	223/276 (80%)	0.10	8 (3%)	42	40	95, 140, 185, 222	0
All	All	454/562 (80%)	0.10	17 (3%)	41	38	95, 150, 209, 249	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	574	LEU	4.3
2	B	775	TYR	4.0
3	A	5	SER	4.0
2	B	752	LEU	3.1
2	B	551	LYS	2.8
3	A	350	VAL	2.7
2	B	553	PHE	2.5
2	B	617	LEU	2.5
3	A	261	LEU	2.5
3	A	33	LEU	2.4
3	A	262	LYS	2.4
3	A	54	PHE	2.3
2	B	776	SER	2.2
2	B	572	LEU	2.2
3	A	265	LEU	2.1
2	B	575	LEU	2.0
3	A	35	MET	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	ZN	A	401	1/1	0.97	0.33	220,220,220,220	0
4	ZN	A	402	1/1	0.99	0.29	155,155,155,155	0

6.5 Other polymers (i)

There are no such residues in this entry.