

Full wwPDB X-ray Structure Validation Report (i)

May 13, 2020 – 08:03 am BST

PDB ID : 5PZB

Title : PanDDA analysis group deposition - Crystal Structure of SP100 after initial

refinement with no ligand modelled (structure 107)

Authors: Pearce, N.M.; Krojer, T.; Talon, R.; Bradley, A.R.; Fairhead, M.; Sethi, R.;

Wright, N.; MacLean, E.; Collins, P.; Brandao-Neto, J.; Douangamath, A.; Renjie, Z.; Dias, A.; Ng, J.; Brennan, P.E.; Cox, O.; Bountra, C.; Arrowsmith,

C.H.; Edwards, A.; von Delft, F.

Deposited on : 2017-02-08

Resolution : 2.05 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 following versions of software and data (see references (1)) 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)

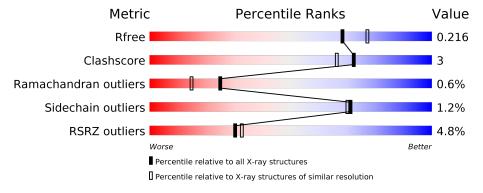
Validation Pipeline (wwPDB-VP) : 2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.05 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$egin{aligned} ext{Similar resolution} \ (\# ext{Entries}, ext{resolution range}(ext{Å})) \end{aligned}$
R_{free}	130704	$1692 \ (2.04-2.04)$
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 >=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	Quality of chain		
1	A	180	95%			
1	В	180	87%	11%		

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:



Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	EDO	A	905	_	_	_	X



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3620 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.

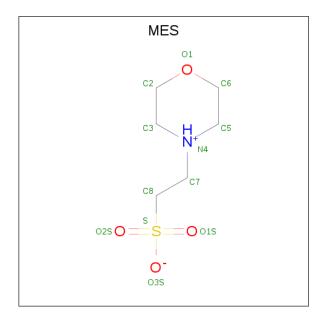
• Molecule 1 is a protein called Nuclear autoantigen Sp-100.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	Λ	178	Total C		N	О	S	0	10	0
1	1 A	110	1536	971	268	278	19	U	10	U
1	B	177	Total	С	N	О	S	0	7	0
1	D	111	1480	940	258	263	19		1	U

• Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	В	2	Total Zn 2 2	0	0
2	A	3	Total Zn 3 3	0	0

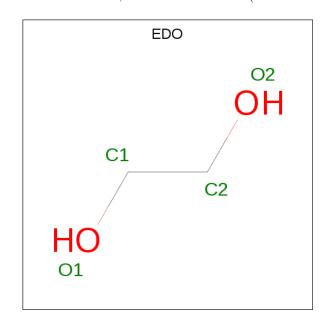
• Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).





\mathbf{Mol}	Chain	Residues	Atoms			ZeroOcc	AltConf		
9	Λ	1	Total	С	N	О	S	0	0
3	А	1	12	6	1	4	1	0	

 \bullet Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $\mathrm{C_2H_6O_2}).$



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0
4	В	1	Total C O 4 2 2	0	0

• Molecule 5 is water.

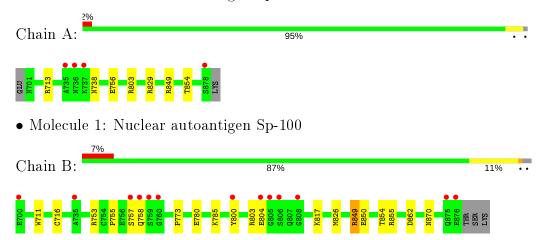
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	317	Total O 317 317	0	3
5	В	262	Total O 262 262	0	0



3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: Nuclear autoantigen Sp-100





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants	127.62Å 45.48Å 83.42Å	Depositor
a, b, c, α , β , γ	90.00° 102.15° 90.00°	Depositor
Resolution (Å)	39.02 - 2.05	Depositor
Resolution (A)	39.02 - 2.05	EDS
% Data completeness	97.8 (39.02-2.05)	Depositor
(in resolution range)	97.8 (39.02-2.05)	EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.05 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
D D.	0.160 , 0.210	Depositor
R, R_{free}	0.169 , 0.216	DCC
R_{free} test set	1420 reflections (4.88%)	wwPDB-VP
Wilson B-factor (\mathring{A}^2)	28.6	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.30 , 53.4	EDS
L-test for twinning ²	$ < L >=0.49, < L^2>=0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3620	wwPDB-VP
Average B, all atoms (Å ²)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.84% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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, EDO, MES

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	Boı	nd lengths	Bond angles		
Wioi Chain		RMSZ	# Z > 5	RMSZ	# Z >5	
1	A	0.91	0/1603	0.85	1/2154 (0.0%)	
1	В	0.98	1/1538 (0.1%)	0.92	4/2071 (0.2%)	
All	All	0.94	1/3141 (0.0%)	0.89	5/4225 (0.1%)	

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	В	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	$\mathbf{Ideal}(\mathbf{\AA})$
1	В	780	GLU	CD-OE2	5.07	1.31	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	${f Z}$	$Observed(^o)$	$\operatorname{Ideal}({}^o)$
1	В	855	ARG	NE-CZ-NH1	6.36	123.48	120.30
1	В	753	ARG	NE-CZ-NH2	-5.46	117.57	120.30
1	В	753	ARG	NE-CZ-NH1	5.37	122.98	120.30
1	В	849	ARG	NE-CZ-NH1	5.22	122.91	120.30
1	A	713	ARG	NE-CZ-NH2	5.05	122.82	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	В	755	PRO	Peptide

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	1536	0	1497	5	0
1	В	1480	0	1414	12	0
2	A	3	0	0	0	0
2	В	2	0	0	0	0
3	A	12	0	13	1	0
4	A	4	0	6	1	0
4	В	4	0	6	1	0
5	A	317	0	0	7	0
5	В	262	0	0	7	2
All	All	3620	0	2936	20	2

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 (20) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{array}{ll} ext{Interatomic} \ ext{distance} \ (ext{\AA}) \end{array}$	$egin{array}{c} ext{Clash} \ ext{overlap } (ext{Å}) \end{array}$	
4:B:903:EDO:O2	5:B:1001:HOH:O	1.72	1.05	
1:A:803[A]:ARG:NH2	5:A:1001:HOH:O	1.90	1.00	
1:A:829[A]:ARG:NH1	5:A:1002:HOH:O	1.99	0.95	
3:A:904:MES:H81	5:A:1239:HOH:O	1.75	0.87	
1:B:716[A]:CYS:SG	5:B:1136:HOH:O	2.45	0.73	
1:B:826[B]:MET:SD	5:B:1018:HOH:O	2.49	0.71	
1:B:826[B]:MET:HE2	5:B:1187:HOH:O	1.98	0.64	
1:A:756:GLU:HG3	5:A:1123:HOH:O	2.01	0.60	
1:A:803[B]:ARG:NH1	5:A:1005:HOH:O	2.34	0.60	
5:A:1001:HOH:O	1:B:773:PRO:HG2	2.05	0.56	
1:B:849:ARG:O	1:B:854[B]:THR:HG21	2.11	0.50	
4:A:905:EDO:C2	5:A:1120:HOH:O	2.62	0.47	
1:B:800:TYR:HE1	1:B:803:ARG:CB	2.28	0.46	
1:B:870:ASN:HB2	5:B:1194:HOH:O	2.17	0.45	

Continued on next page...



$\alpha \cdots \tau$	r	•	
Continued	trom	nromanne	naae
\circ	110116	picolous	puyc

Atom-1	Atom-2	$egin{array}{ll} ext{Interatomic} \ ext{distance} & (ext{Å}) \end{array}$	Clash overlap (Å)
1:B:800:TYR:CE1	1:B:803:ARG:CB	3.01	0.43
1:A:849:ARG:O	1:A:854[B]:THR:HG21	2.19	0.43
1:B:803:ARG:HA	1:B:804:GLU:HA	1.89	0.41
1:B:817:LYS:NZ	5:B:1022:HOH:O	2.53	0.41
1:B:785:LYS:HE3	5:B:1234:HOH:O	2.21	0.40

All (2) 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	$egin{array}{l} ext{Interatomic} \ ext{distance} \ (ext{Å}) \end{array}$	Clash overlap (Å)
5:B:1047:HOH:O	5:B:1047:HOH:O[2_556]	0.88	1.32
5:B:1197:HOH:O	5:B:1197:HOH:O[2_555]	1.82	0.38

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	Favoured Allowed		Percentiles		
1	A	186/180 (103%)	182 (98%)	4 (2%)	0	100 100		
1	В	182/180 (101%)	172 (94%)	8 (4%)	2 (1%)	14 5		
All	All	$368/360 \ (102\%)$	354 (96%)	12 (3%)	2 (0%)	25 18		

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	В	757	SER
1	В	758	GLN

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	178/171 (104%)	177 (99%)	1 (1%)	86 87		
1	В	165/171~(96%)	160 (97%)	5 (3%)	41 34		
All	All	343/342 (100%)	337 (98%)	6 (2%)	71 57		

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	738	ASN
1	В	711	TRP
1	В	850[A]	GLU
1	В	850[B]	GLU
1	В	862[A]	ASP
1	В	862[B]	ASP

Some 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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 5 are monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The



Link column lists molecule types, if any, to which the group is linked. 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| > 2 is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol Type Chain		ain Res		Link Bond lengths			В	ond ang	les	
Moi Type	Chain	res	Lilik	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2	
4	EDO	A	905	_	3,3,3	0.47	0	2,2,2	0.16	0
4	EDO	В	903	-	3,3,3	0.45	0	2,2,2	0.22	0
3	MES	A	904	-	12,12,12	1.84	1 (8%)	14,16,16	2.20	3 (21%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	EDO	A	905	_	-	1/1/1/1	-
4	EDO	В	903	_	-	0/1/1/1	-
3	MES	A	904	-	-	5/6/14/14	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	${f Observed(\AA)}$	$[Ideal(\AA)]$
3	A	904	MES	C8-S	-5.93	1.69	1.77

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^o)$
3	A	904	MES	O1S-S-C8	5.14	113.10	106.92
3	A	904	MES	C6-C5-N4	4.39	116.76	110.10
3	A	904	MES	O2S-S-C8	2.34	109.74	106.92

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	904	MES	C8-C7-N4-C3
4	A	905	EDO	O1-C1-C2-O2
3	A	904	MES	C7-C8-S-O1S
3	A	904	MES	C7-C8-S-O2S
3	A	904	MES	C7-C8-S-O3S
3	A	904	MES	N4-C7-C8-S



There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	905	EDO	1	0
4	В	903	EDO	1	0
3	A	904	MES	1	0

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^{th} 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 $>$	$\#\mathrm{RSRZ}{>}2$		$OWAB(A^2)$	Q < 0.9	
1	A	178/180 (98%)	-0.42	4 (2%)	62	66	17, 30, 67, 110	0
1	В	177/180 (98%)	-0.03	13 (7%)	15	16	19, 33, 90, 143	0
All	All	355/360 (98%)	-0.22	17 (4%)	30	33	17, 32, 83, 143	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	757	SER	6.8
1	В	806	SER	6.8
1	A	735	ALA	5.6
1	В	805	GLY	5.5
1	В	876	GLU	4.3
1	В	759	SER	3.3
1	В	758	GLN	3.2
1	В	804	GLU	3.2
1	A	736	ASN	3.0
1	В	735	ALA	2.9
1	A	737	LYS	2.8
1	В	700	GLU	2.4
1	В	875	GLN	2.4
1	В	760	GLY	2.3
1	A	878	SER	2.2
1	В	808	GLY	2.0
1	В	800	TYR	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 carbohydrates 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, 95^{th} 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	${f B\text{-factors}}({f \AA}^2)$	Q < 0.9
4	EDO	A	905	4/4	0.78	0.43	61,72,77,83	0
2	ZN	A	903	1/1	0.85	0.21	91,91,91,91	1
4	EDO	В	903	4/4	0.95	0.13	53,55,58,59	0
3	MES	A	904	12/12	0.96	0.25	30,34,36,36	12
2	ZN	A	902	1/1	0.98	0.05	36,36,36,36	0
2	ZN	В	902	1/1	0.99	0.04	30,30,30,30	0
2	ZN	A	901	1/1	0.99	0.05	30,30,30,30	0
2	ZN	В	901	1/1	1.00	0.07	25,25,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.

