wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 02:32 pm BST

PDB ID	:	5PPR
Title	:	PanDDA analysis group deposition – Crystal Structure of BRD1 after initial refinement with no ligand modelled (structure 28)
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	:	2017-02-07
Resolution	:	2.69 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 (i) symbol.

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

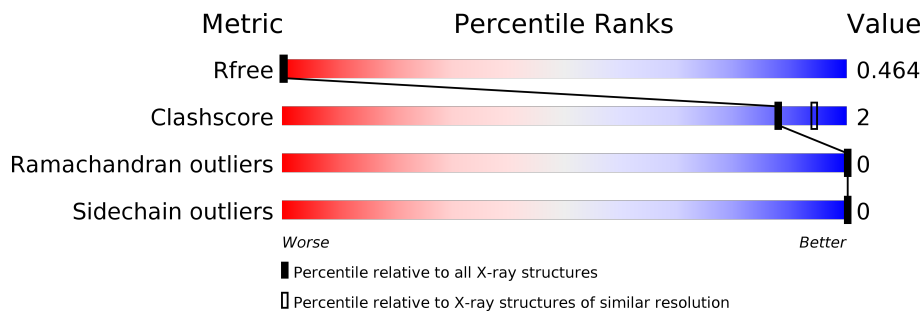|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.69 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	Percentile Ranks	Value
Rfree		0.464
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		0

Worse ← → *Better*

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)

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%

Mol	Chain	Length	Quality of chain
1	A	156	70% 8% 22%
1	B	156	73% 6% 21%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2048 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 Bromodomain-containing protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	121	Total	C	N	O	S	0	5	0
			989	621	175	186	7			
1	B	123	Total	C	N	O	S	0	6	0
			1038	651	187	194	6			

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP O95696
A	2	HIS	-	expression tag	UNP O95696
A	3	HIS	-	expression tag	UNP O95696
A	4	HIS	-	expression tag	UNP O95696
A	5	HIS	-	expression tag	UNP O95696
A	6	HIS	-	expression tag	UNP O95696
A	7	HIS	-	expression tag	UNP O95696
A	8	SER	-	expression tag	UNP O95696
A	9	SER	-	expression tag	UNP O95696
A	10	GLY	-	expression tag	UNP O95696
A	11	VAL	-	expression tag	UNP O95696
A	12	ASP	-	expression tag	UNP O95696
A	13	LEU	-	expression tag	UNP O95696
A	14	GLY	-	expression tag	UNP O95696
A	15	THR	-	expression tag	UNP O95696
A	16	GLU	-	expression tag	UNP O95696
A	17	ASN	-	expression tag	UNP O95696
A	18	LEU	-	expression tag	UNP O95696
A	19	TYR	-	expression tag	UNP O95696
A	20	PHE	-	expression tag	UNP O95696
A	21	GLN	-	expression tag	UNP O95696
A	22	SER	-	expression tag	UNP O95696
A	23	MET	VAL	engineered mutation	UNP O95696
A	34	GLU	PRO	engineered mutation	UNP O95696
A	37	ARG	VAL	engineered mutation	UNP O95696

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	expression tag	UNP O95696
B	2	HIS	-	expression tag	UNP O95696
B	3	HIS	-	expression tag	UNP O95696
B	4	HIS	-	expression tag	UNP O95696
B	5	HIS	-	expression tag	UNP O95696
B	6	HIS	-	expression tag	UNP O95696
B	7	HIS	-	expression tag	UNP O95696
B	8	SER	-	expression tag	UNP O95696
B	9	SER	-	expression tag	UNP O95696
B	10	GLY	-	expression tag	UNP O95696
B	11	VAL	-	expression tag	UNP O95696
B	12	ASP	-	expression tag	UNP O95696
B	13	LEU	-	expression tag	UNP O95696
B	14	GLY	-	expression tag	UNP O95696
B	15	THR	-	expression tag	UNP O95696
B	16	GLU	-	expression tag	UNP O95696
B	17	ASN	-	expression tag	UNP O95696
B	18	LEU	-	expression tag	UNP O95696
B	19	TYR	-	expression tag	UNP O95696
B	20	PHE	-	expression tag	UNP O95696
B	21	GLN	-	expression tag	UNP O95696
B	22	SER	-	expression tag	UNP O95696
B	23	MET	VAL	engineered mutation	UNP O95696
B	34	GLU	PRO	engineered mutation	UNP O95696
B	37	ARG	VAL	engineered mutation	UNP O95696

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).
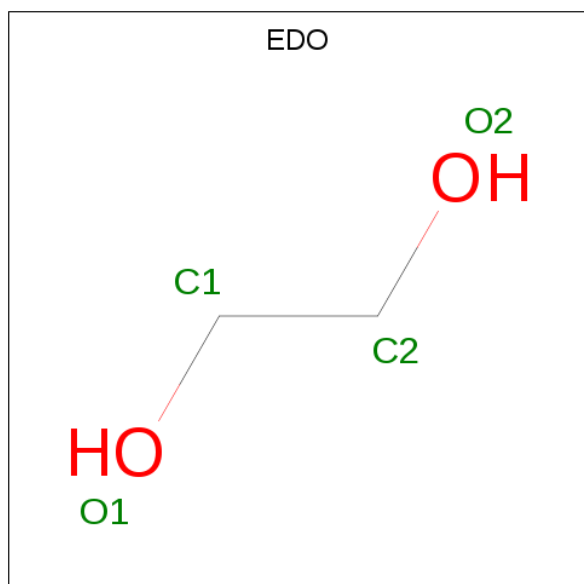
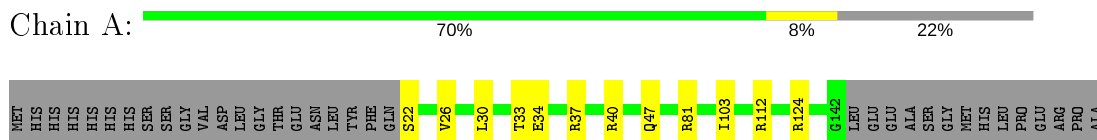| Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 2	4	0
2	A	1	Total 8	C 4	O 4	8	1
2	B	1	Total 4	C 2	O 2	4	0
2	B	1	Total 4	C 2	O 2	4	1

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

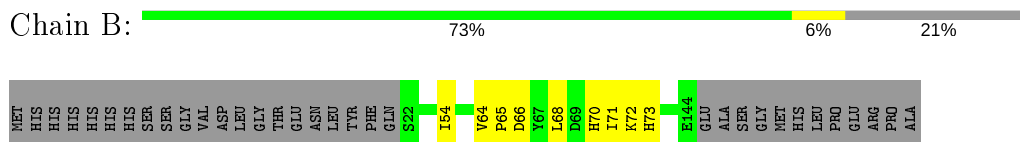| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Na 1	1	0

3 Residue-property plots ⓘ

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. 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. 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: Bromodomain-containing protein 1

Chain A:

70%   8%   22%

MET HIS HIS HIS HIS HIS SER SER GLY VAL ASP LEU GLY THR GLU ASN LEU TYR PHE GLN S22 V26 L30 T33 E34 R37 R40 Q47 R81 I103 R112 R124 G142 LEU GLU GLU ALA SER GLY MET HIS LEU PRO GLU ARG PRO ALA

• Molecule 1: Bromodomain-containing protein 1

Chain B:

73%   6%   21%

MET HIS HIS HIS HIS HIS SER SER GLY VAL ASP LEU GLY THR GLU ASN LEU TYR PHE GLN S22 I54 V64 P65 D66 Y67 L68 D69 H70 I71 K72 H73 E144 GLU ALA SER GLY MET HIS LEU PRO GLU ARG PRO ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	55.73Å　　56.32Å　　102.50Å 90.00°　　90.00°　　　90.00°	Depositor
Resolution (Å)	29.21　–　2.69 29.21　–　2.69	Depositor EDS
% Data completeness (in resolution range)	99.0 (29.21-2.69) 97.1 (29.21-2.69)	Depositor EDS
R_{merge}	1.39	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.06 (at 2.68Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.464　,　0.470 0.462　,　0.464	Depositor DCC
R_{free} test set	460 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	12.6	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 3.3	EDS
L-test for twinning [2]	$<|L|> = 0.30, <L^2> = 0.14$	Xtriage
Estimated twinning fraction	0.098 for k,h,-l	Xtriage
F_o,F_c correlation	0.25	EDS
Total number of atoms	2048	wwPDB-VP
Average B, all atoms (Å^2)	2.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.97% 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: NA, EDO

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	0.50	0/1016	0.66	0/1371
1	B	0.47	0/1068	0.69	0/1437
All	All	0.49	0/2084	0.67	0/2808

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	989	0	959	7	18
1	B	1038	0	1033	2	18
2	A	12	0	18	0	0
2	B	8	0	12	0	0
3	A	1	0	0	0	0
All	All	2048	0	2022	8	18

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 2.

The worst 5 of 8 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:26:VAL:O	1:A:30:LEU:N	2.31	0.61
1:A:112:ARG:HD2	1:B:70:HIS:CE1	2.38	0.59
1:A:103[B]:ILE:HD13	1:A:124[B]:ARG:HB2	1.92	0.51
1:A:47:GLN:OE1	1:A:81:ARG:NH2	2.47	0.47
1:A:37:ARG:O	1:A:40:ARG:HB3	2.18	0.43

The worst 5 of 18 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:A:40:ARG:NH2	1:B:71:ILE:O[4_445]	1.60	0.60
1:A:40:ARG:NE	1:B:72:LYS:C[4_445]	1.63	0.57
1:A:40:ARG:NE	1:B:72:LYS:CA[4_445]	1.65	0.55
1:A:37:ARG:CB	1:B:68:LEU:CD2[4_445]	1.69	0.51
1:A:33:THR:O	1:B:68:LEU:N[4_445]	1.71	0.49

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	124/156 (80%)	123 (99%)	1 (1%)	0	100	100
1	B	127/156 (81%)	125 (98%)	2 (2%)	0	100	100
All	All	251/312 (80%)	248 (99%)	3 (1%)	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	103/138 (75%)	103 (100%)	0	100	100
1	B	112/138 (81%)	112 (100%)	0	100	100
All	All	215/276 (78%)	215 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 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	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z|>2$	Counts	RMSZ	$\#|Z|>2$
2	EDO	B	202[B]	-	3,3,3	0.47	0	2,2,2	0.31	0
2	EDO	B	201	-	3,3,3	0.45	0	2,2,2	0.30	0
2	EDO	A	201	-	3,3,3	0.48	0	2,2,2	0.43	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z|$ > 2	Counts	RMSZ	#$|Z|$ > 2
2	EDO	A	202[A]	-	3,3,3	0.45	0	2,2,2	0.40	0
2	EDO	A	202[B]	-	3,3,3	0.48	0	2,2,2	0.38	0

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
2	EDO	B	202[B]	-	-	0/1/1/1	-
2	EDO	B	201	-	-	1/1/1/1	-
2	EDO	A	201	-	-	0/1/1/1	-
2	EDO	A	202[A]	-	-	1/1/1/1	-
2	EDO	A	202[B]	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	201	EDO	O1-C1-C2-O2
2	A	202[B]	EDO	O1-C1-C2-O2
2	A	202[A]	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

5.7　Other polymers ⓘ

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.