

# wwPDB X-ray Structure Validation Summary Report (i)

#### Feb 27, 2024 – 09:20 AM EST

PDB ID	:	5PUG
Title	:	PanDDA analysis group deposition – Crystal Structure of BRD1 after initial
		refinement with no ligand modelled (structure 194)
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	:	2.00 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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

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

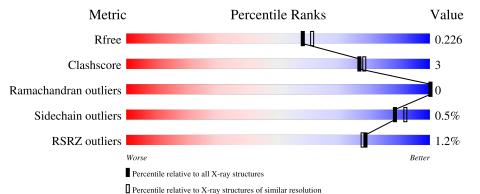
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber $(2001)$
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)

# 1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.00 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	А	156	<sup>2%</sup> <b>72%</b>	5%	22%	
1	В	156	69%	9%	• 21%	

Validation Pipeline (wwPDB-VP) : 2.36



# 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2307 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.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	Λ	121	Total	С	Ν	0	$\mathbf{S}$	0	F	0
		121	989	621	175	186	$\overline{7}$	0	5	0
1	В	123	Total	С	Ν	0	S	0	6	0
	D	123	1038	651	187	194	6	0	0	0

• Molecule 1 is a protein called Bromodomain-containing protein 1.

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

There are 50 discrepancies between the modelled and reference sequences:

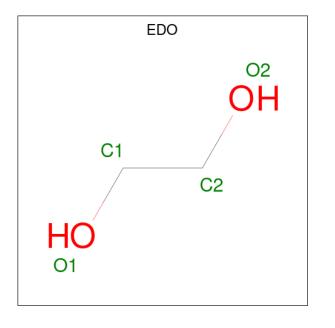
Continued on next page...



Chain	Residue	Modelled	Actual	Comment	Reference
В	1	MET	-	expression tag	UNP O95696
В	2	HIS	-	expression tag	UNP O95696
В	3	HIS	-	expression tag	UNP O95696
В	4	HIS	-	expression tag	UNP O95696
В	5	HIS	-	expression tag	UNP O95696
В	6	HIS	-	expression tag	UNP O95696
В	7	HIS	-	expression tag	UNP O95696
В	8	SER	-	expression tag	UNP O95696
В	9	SER	-	expression tag	UNP O95696
В	10	GLY	-	expression tag	UNP O95696
В	11	VAL	-	expression tag	UNP O95696
В	12	ASP	-	expression tag	UNP O95696
В	13	LEU	-	expression tag	UNP O95696
В	14	GLY	-	expression tag	UNP O95696
В	15	THR	-	expression tag	UNP O95696
В	16	GLU	-	expression tag	UNP O95696
В	17	ASN	-	expression tag	UNP O95696
В	18	LEU	-	expression tag	UNP O95696
В	19	TYR	-	expression tag	UNP 095696
В	20	PHE	-	expression tag	UNP 095696
В	21	GLN	-	expression tag	UNP O95696
В	22	SER	-	expression tag	UNP O95696
В	23	MET	VAL	engineered mutation	UNP O95696
В	34	GLU	PRO	engineered mutation	UNP O95696
В	37	ARG	VAL	engineered mutation	UNP 095696

Continued from previous page...

• Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).







Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 8 & 4 & 4 \end{array}$	0	1
2	В	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
2	В	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	1

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	В	1	Total Na 1 1	0	0

• Molecule 4 is water.

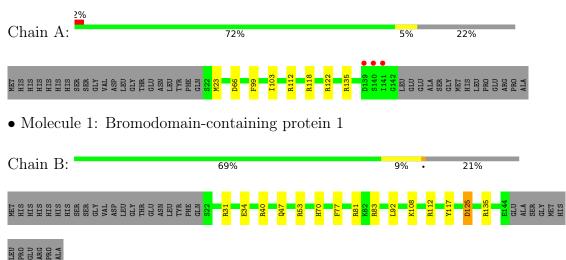
Mol	Chain	Residues	Residues Atoms		AltConf
4	А	111	Total O 111 111	0	0
4	В	148	Total O 148 148	0	1



## 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: Bromodomain-containing protein 1





## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	55.23Å $56.36$ Å $101.84$ Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	29.08 - 2.00	Depositor
Resolution (A)	29.08 - 2.00	EDS
% Data completeness	98.5 (29.08-2.00)	Depositor
(in resolution range)	98.6 (29.08-2.00)	EDS
R <sub>merge</sub>	0.14	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.18 (at 2.00 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
D D.	0.179 , $0.219$	Depositor
$R, R_{free}$	0.187 , $0.226$	DCC
$R_{free}$ test set	1014 reflections $(4.64\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	22.0	Xtriage
Anisotropy	0.059	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.33 , $45.5$	EDS
L-test for twinning <sup>2</sup>	$< L >=0.48, < L^2>=0.31$	Xtriage
Estimated twinning fraction	0.028 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2307	wwPDB-VP
Average B, all atoms $(Å^2)$	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.65% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities 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: EDO, NA

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	Bo	nd lengths	Bond angles		
	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.88	0/1016	0.90	3/1371~(0.2%)	
1	В	0.96	1/1068~(0.1%)	0.99	4/1437~(0.3%)	
All	All	0.92	1/2084~(0.0%)	0.95	7/2808~(0.2%)	

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	В	125	ASP	CB-CG	7.05	1.66	1.51

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$\mathbf{Ideal}(^{o})$
1	А	66	ASP	CB-CG-OD1	7.40	124.96	118.30
1	А	122	ARG	NE-CZ-NH2	-5.57	117.52	120.30
1	В	31	ARG	NE-CZ-NH1	5.33	122.97	120.30
1	В	83	ARG	NE-CZ-NH1	5.31	122.95	120.30
1	В	40	ARG	NE-CZ-NH1	-5.19	117.70	120.30

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	А	989	0	959	3	1
1	В	1038	0	1033	9	0
2	А	12	0	18	0	0
2	В	8	0	12	3	0
3	В	1	0	0	0	0
4	А	111	0	0	0	0
4	В	148	0	0	5	1
All	All	2307	0	2022	12	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 3.

The worst 5 of 12 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:112:ARG:HD2	1:B:70:HIS:CE1	2.27	0.69
1:B:92:LEU:HD21	2:B:203[B]:EDO:H11	1.88	0.54
1:B:92:LEU:CD2	2:B:203[B]:EDO:H11	2.39	0.52
1:B:47:GLN:HG2	1:B:77:PHE:CD2	2.45	0.52
1:B:108[B]:LYS:HD2	4:B:399:HOH:O	2.12	0.49

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:A:23[A]:MET:SD	4:B:309:HOH:O[1_655]	1.92	0.28

#### 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	А	124/156~(80%)	122 (98%)	2(2%)	0	100 100

Continued on next page...



	Chain	Analysed	Favoured	Allowed	Outliers	Perce	ntiles
1	В	127/156~(81%)	127 (100%)	0	0	100	100
All	All	251/312~(80%)	249 (99%)	2(1%)	0	100	100

Continued from previous page...

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	А	103/138~(75%)	103 (100%)	0	100 100		
1	В	112/138~(81%)	111~(99%)	1 (1%)	78 83		
All	All	215/276~(78%)	214 (100%)	1 (0%)	88 92		

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

Mol	Chain	Res	Type	
1	В	125	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 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	Turne	Chain	Dec	Res Link		Bog Link Bond lengths			Bond angles		
	Type	Unam	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2	
2	EDO	А	201	-	3,3,3	0.56	0	$2,\!2,\!2$	0.23	0	
2	EDO	А	202[B]	-	3,3,3	0.38	0	2,2,2	0.71	0	
2	EDO	В	203[B]	-	$3,\!3,\!3$	0.53	0	$2,\!2,\!2$	1.08	0	
2	EDO	В	202	-	3,3,3	0.36	0	2,2,2	0.24	0	
2	EDO	А	202[A]	-	3,3,3	0.41	0	$2,\!2,\!2$	0.49	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	А	201	-	-	0/1/1/1	-
2	EDO	А	202[B]	-	-	1/1/1/1	-
2	EDO	В	203[B]	-	-	1/1/1/1	-
2	EDO	В	202	-	-	1/1/1/1	-
2	EDO	А	202[A]	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	В	202	EDO	O1-C1-C2-O2
2	А	202[B]	EDO	O1-C1-C2-O2
2	А	202[A]	EDO	O1-C1-C2-O2
2	В	203[B]	EDO	O1-C1-C2-O2

There are no ring outliers.



Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	В	203[B]	EDO	2	0
2	В	202	EDO	1	0

2 monomers are involved in 3 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,  $95^{th}$  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	$\langle RSRZ \rangle$	#RSRZ>2	$\mathbf{OWAB}(\mathbf{\AA}^2)$	Q < 0.9
1	А	121/156~(77%)	-0.14	3 (2%) 57 56	15, 27, 44, 65	0
1	В	123/156~(78%)	-0.39	0 100 100	12, 19, 30, 51	0
All	All	244/312~(78%)	-0.26	3 (1%) 79 78	12, 22, 39, 65	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	139	ASP	3.1
1	А	140	SER	2.9
1	А	141	ILE	2.6

### 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,  $95^{th}$  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	$\mathbf{B} ext{-factors}(\mathrm{\AA}^2)$	Q < 0.9
2	EDO	В	203[B]	4/4	0.90	0.18	24,24,26,26	4
2	EDO	А	201	4/4	0.94	0.13	34,36,36,37	0

Continued on next page...



Mol	Type	Chain	Res	Atoms	RSCC	RSR	$B-factors(Å^2)$	Q<0.9
2	EDO	В	202	4/4	0.96	0.14	$31,\!33,\!35,\!39$	0
2	EDO	А	202[A]	4/4	0.97	0.13	35,36,36,39	4
2	EDO	А	202[B]	4/4	0.97	0.13	$23,\!23,\!23,\!25$	4
3	NA	В	201	1/1	0.99	0.10	$17,\!17,\!17,\!17$	0

Continued from previous page...

## 6.5 Other polymers (i)

There are no such residues in this entry.

