



# Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2016 – 05:00 AM EDT

PDB ID : 5FH8  
Title : Crystal structure of the fifth bromodomain of human PB1 in complex with compound 28  
Authors : Tallant, C.; Sutherell, C.L.; Siejka, P.; Sorrell, F.J.; Krojer, T.; Picaud, S.; Fedorov, O.; von Delft, F.; Arrowsmith, C.H.; Edwards, A.M.; Bountra, C.; Brennan, P.E.; Ley, S.V.; Knapp, S.  
Deposited on : 2015-12-21  
Resolution : 1.55 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027674  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027674

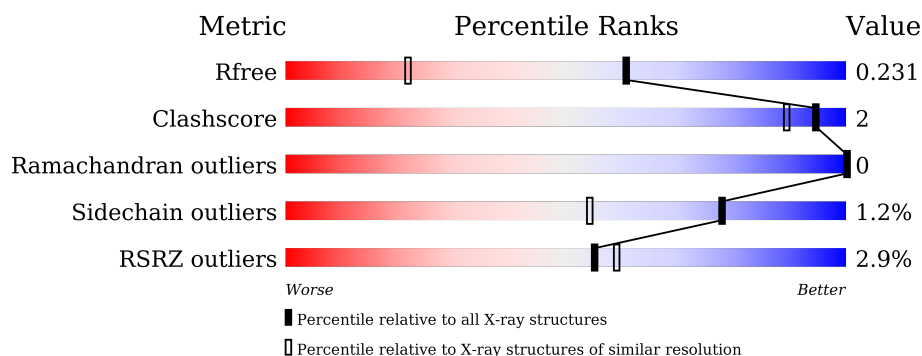
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*





The reported resolution of this entry is 1.55 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))
$R_{free}$	91344	1665 (1.58-1.54)
Clashscore	102246	1014 (1.56-1.56)
Ramachandran outliers	100387	1704 (1.58-1.54)
Sidechain outliers	100360	1702 (1.58-1.54)
RSRZ outliers	91569	1668 (1.58-1.54)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	124	
1	B	124	
1	C	124	
1	D	124	

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	DMS	A	802	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4054 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.

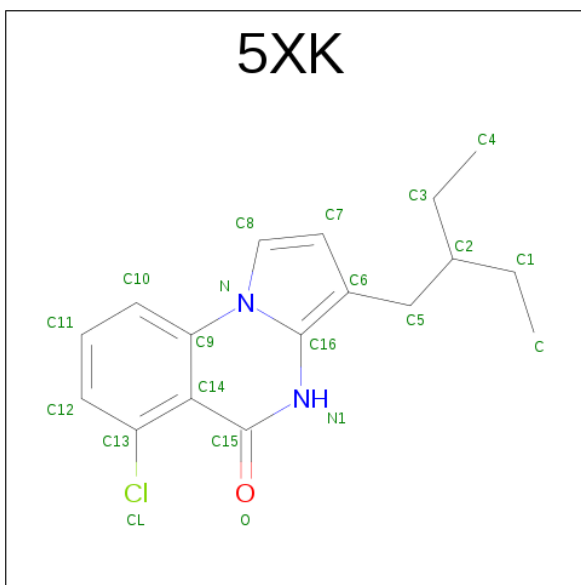
- Molecule 1 is a protein called Protein polybromo-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	112	Total	C	N	O	S	0	3	0
			949	602	162	175	10			
1	B	111	Total	C	N	O	S	0	0	0
			929	590	158	171	10			
1	C	109	Total	C	N	O	S	0	0	0
			912	578	155	169	10			
1	D	109	Total	C	N	O	S	0	1	0
			915	580	155	170	10			

There are 8 discrepancies between the modelled and reference sequences:

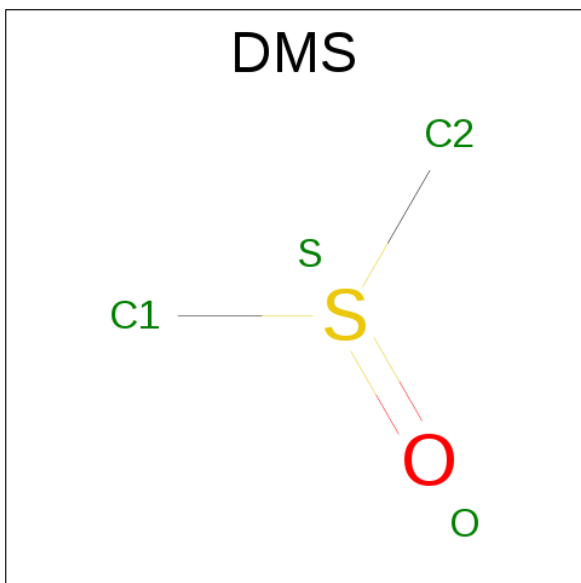
Chain	Residue	Modelled	Actual	Comment	Reference
A	643	SER	-	expression tag	UNP Q86U86
A	644	MET	-	expression tag	UNP Q86U86
B	643	SER	-	expression tag	UNP Q86U86
B	644	MET	-	expression tag	UNP Q86U86
C	643	SER	-	expression tag	UNP Q86U86
C	644	MET	-	expression tag	UNP Q86U86
D	643	SER	-	expression tag	UNP Q86U86
D	644	MET	-	expression tag	UNP Q86U86

- Molecule 2 is 6-chloranyl-3-(2-ethylbutyl)-4 {H}-pyrrolo[1,2-a]quinazolin-5-one (three-letter code: 5XK) (formula: C<sub>17</sub>H<sub>19</sub>ClN<sub>2</sub>O).



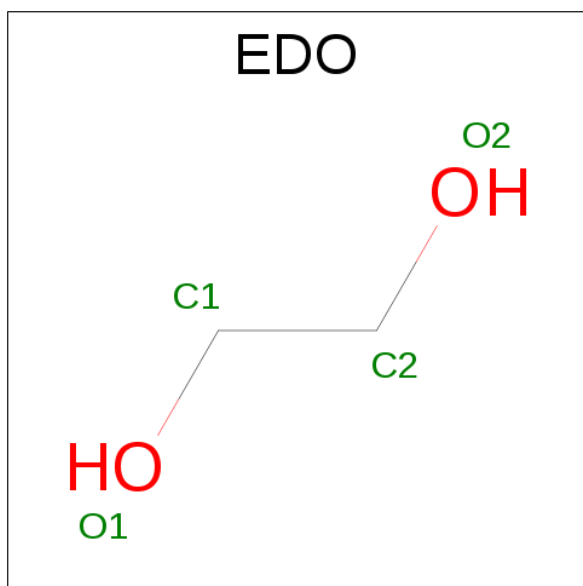
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	0	0
			21	17	1	2	1		
2	B	1	Total	C	Cl	N	O	0	0
			21	17	1	2	1		
2	C	1	Total	C	Cl	N	O	0	0
			21	17	1	2	1		
2	D	1	Total	C	Cl	N	O	0	0
			21	17	1	2	1		

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula:  $C_2H_6OS$ ).



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

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



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


- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	92	Total	O	0	0
			92	92		
5	B	65	Total	O	0	0
			65	65		
5	C	53	Total	O	0	0
			53	53		
5	D	47	Total	O	0	0
			47	47		

### 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 errors 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: Protein polybromo-1

Chain A:  85% . . 10%




- Molecule 1: Protein polybromo-1

Chain B:  3% 82% 7% 10%




- Molecule 1: Protein polybromo-1

Chain C:  2% 82% 6% 12%



- Molecule 1: Protein polybromo-1

Chain D:  5% 82% 6% 12%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.58Å 62.42Å 65.82Å 90.00° 91.75° 90.00°	Depositor
Resolution (Å)	64.55 – 1.55 64.55 – 1.55	Depositor EDS
% Data completeness (in resolution range)	90.4 (64.55-1.55) 90.4 (64.55-1.55)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.65 (at 1.55Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, $R_{free}$	0.187 , 0.224 0.197 , 0.231	Depositor DCC
$R_{free}$ test set	3423 reflections (5.24%)	DCC
Wilson B-factor (Å <sup>2</sup> )	23.1	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 40.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.002 for l,k,-h 0.029 for h,-k,-l 0.017 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4054	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.73% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.



## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: DMS, EDO, 5XK

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	0.92	1/975 (0.1%)	0.95	2/1308 (0.2%)
1	B	0.87	0/946	0.94	3/1270 (0.2%)
1	C	0.76	0/929	0.87	0/1248
1	D	0.78	0/935	0.89	0/1256
All	All	0.84	1/3785 (0.0%)	0.91	5/5082 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	740	GLU	CD-OE1	6.94	1.33	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	658	MET	CG-SD-CE	6.38	110.40	100.20
1	A	679	ARG	NE-CZ-NH1	5.94	123.27	120.30
1	B	757	LEU	CB-CG-CD2	5.49	120.33	111.00
1	A	679	ARG	NE-CZ-NH2	-5.25	117.67	120.30
1	B	748	ASP	CB-CG-OD1	5.08	122.87	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts

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	949	0	969	2	0
1	B	929	0	941	3	0
1	C	912	0	917	4	0
1	D	915	0	922	4	0
2	A	21	0	0	0	0
2	B	21	0	0	0	0
2	C	21	0	0	0	0
2	D	21	0	0	0	0
3	A	4	0	6	0	0
4	D	4	0	6	1	0
5	A	92	0	0	0	0
5	B	65	0	0	1	0
5	C	53	0	0	0	0
5	D	47	0	0	2	0
All	All	4054	0	3761	13	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 2.

All (13) 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:C:707:GLU:OE1	1:C:710:ARG:NH1	2.12	0.83
1:A:712:HIS:CD2	1:A:717:LYS:HE3	2.28	0.69
1:D:662:LEU:HD23	1:D:721:ILE:HD13	1.85	0.58
1:C:721:ILE:HD12	1:C:722:ASP:N	2.23	0.52
1:A:671:ASN:OD1	1:A:679:ARG:NH2	2.43	0.51
1:B:671:ASN:O	1:B:673:THR:HG23	2.12	0.50
1:B:679:ARG:HD3	5:B:960:HOH:O	2.14	0.47
1:B:746:TYR:CE1	1:B:750:LEU:HD11	2.49	0.47
1:C:714:MET:CE	1:C:714:MET:HA	2.48	0.44
1:D:710:ARG:NH2	5:D:902:HOH:O	2.51	0.43
1:D:733:ASN:ND2	5:D:904:HOH:O	2.52	0.41
1:D:658:MET:HG3	4:D:802:EDO:C1	2.51	0.41
1:C:713:MET:HG2	1:C:714:MET:HE3	2.03	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	113/124 (91%)	113 (100%)	0	0	100	100
1	B	109/124 (88%)	109 (100%)	0	0	100	100
1	C	107/124 (86%)	107 (100%)	0	0	100	100
1	D	108/124 (87%)	107 (99%)	1 (1%)	0	100	100
All	All	437/496 (88%)	436 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains

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	109/116 (94%)	108 (99%)	1 (1%)	84	66
1	B	105/116 (90%)	104 (99%)	1 (1%)	82	62
1	C	103/116 (89%)	102 (99%)	1 (1%)	82	62
1	D	104/116 (90%)	102 (98%)	2 (2%)	65	33
All	All	421/464 (91%)	416 (99%)	5 (1%)	78	54

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	695	ASP
1	B	695	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	717	LYS
1	D	691	SER
1	D	701	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	712	HIS
1	C	712	HIS
1	D	733	ASN

### 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](#)

6 ligands are modelled in this entry.

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	5XK	A	801	-	18,23,23	3.02	5 (27%)	16,33,33	2.43	2 (12%)
3	DMS	A	802	-	3,3,3	0.52	0	3,3,3	0.70	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	5XK	B	801	-	18,23,23	2.83	5 (27%)	16,33,33	2.41	2 (12%)
2	5XK	C	801	-	18,23,23	2.89	5 (27%)	16,33,33	2.44	2 (12%)
2	5XK	D	801	-	18,23,23	3.01	5 (27%)	16,33,33	2.42	2 (12%)
4	EDO	D	802	-	3,3,3	0.32	0	2,2,2	0.80	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	5XK	A	801	-	-	0/7/8/8	0/3/3/3
3	DMS	A	802	-	-	0/0/0/0	0/0/0/0
2	5XK	B	801	-	-	0/7/8/8	0/3/3/3
2	5XK	C	801	-	-	0/7/8/8	0/3/3/3
2	5XK	D	801	-	-	0/7/8/8	0/3/3/3
4	EDO	D	802	-	-	0/1/1/1	0/0/0/0

All (20) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	801	5XK	C5-C6	-5.55	1.33	1.50
2	D	801	5XK	C5-C6	-5.52	1.33	1.50
2	C	801	5XK	C5-C6	-5.31	1.34	1.50
2	B	801	5XK	C5-C6	-5.31	1.34	1.50
2	C	801	5XK	C16-N1	-2.88	1.30	1.35
2	A	801	5XK	C16-N1	-2.85	1.30	1.35
2	D	801	5XK	C16-N1	-2.85	1.30	1.35
2	B	801	5XK	C16-N1	-2.45	1.31	1.35
2	D	801	5XK	C15-N1	3.35	1.39	1.33
2	B	801	5XK	C15-C14	3.42	1.47	1.41
2	C	801	5XK	C15-N1	3.52	1.39	1.33
2	A	801	5XK	C15-N1	3.56	1.39	1.33
2	B	801	5XK	C15-N1	3.77	1.39	1.33
2	D	801	5XK	C15-C14	4.42	1.48	1.41
2	A	801	5XK	C15-C14	4.51	1.48	1.41
2	C	801	5XK	C15-C14	4.71	1.49	1.41
2	C	801	5XK	C8-C7	8.71	1.51	1.37
2	B	801	5XK	C8-C7	9.02	1.51	1.37
2	A	801	5XK	C8-C7	9.43	1.52	1.37
2	D	801	5XK	C8-C7	9.48	1.52	1.37

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	801	5XK	C14-C15-N1	-7.88	118.10	124.15
2	A	801	5XK	C14-C15-N1	-7.88	118.10	124.15
2	D	801	5XK	C14-C15-N1	-7.87	118.10	124.15
2	B	801	5XK	C14-C15-N1	-7.87	118.10	124.15
2	B	801	5XK	C6-C5-C2	5.20	127.10	115.64
2	D	801	5XK	C6-C5-C2	5.26	127.24	115.64
2	C	801	5XK	C6-C5-C2	5.44	127.64	115.64
2	A	801	5XK	C6-C5-C2	5.47	127.70	115.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	802	EDO	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	112/124 (90%)	-0.17	0 100 100	15, 21, 35, 47	0
1	B	111/124 (89%)	-0.04	4 (3%) 46 49	17, 25, 45, 73	0
1	C	109/124 (87%)	0.03	3 (2%) 56 60	17, 29, 51, 67	0
1	D	109/124 (87%)	0.19	6 (5%) 29 29	16, 30, 64, 73	0
All	All	441/496 (88%)	0.00	13 (2%) 55 59	15, 25, 54, 73	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	658	MET	4.7
1	C	654	TYR	4.6
1	D	761	ARG	3.9
1	D	654	TYR	3.9
1	C	721	ILE	2.6
1	D	719	GLN	2.6
1	B	763	LEU	2.6
1	B	719	GLN	2.3
1	D	716	ASN	2.3
1	B	761	ARG	2.3
1	B	762	ASP	2.2
1	D	656	THR	2.1
1	C	761	ARG	2.0

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	DMS	A	802	4/4	0.72	0.22	6.13	41,42,45,54	0
4	EDO	D	802	4/4	0.80	0.31	1.93	47,56,59,63	0
2	5XK	B	801	21/21	0.98	0.08	0.09	15,17,25,25	0
2	5XK	C	801	21/21	0.98	0.07	-0.64	17,19,24,28	0
2	5XK	A	801	21/21	0.98	0.07	-0.82	14,16,22,22	0
2	5XK	D	801	21/21	0.98	0.07	-1.49	15,17,21,24	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.