



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 11, 2023 – 12:08 PM JST

PDB ID : 8JWU
Title : PHD Finger Protein 7 (PHF7) fused to UBE2D2 via a (GSGG)₃ linker
Authors : Lee, H.S.; Bang, I.; Choi, H.-J.
Deposited on : 2023-06-29
Resolution : 3.58 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
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)
Validation Pipeline (wwPDB-VP) : 2.36

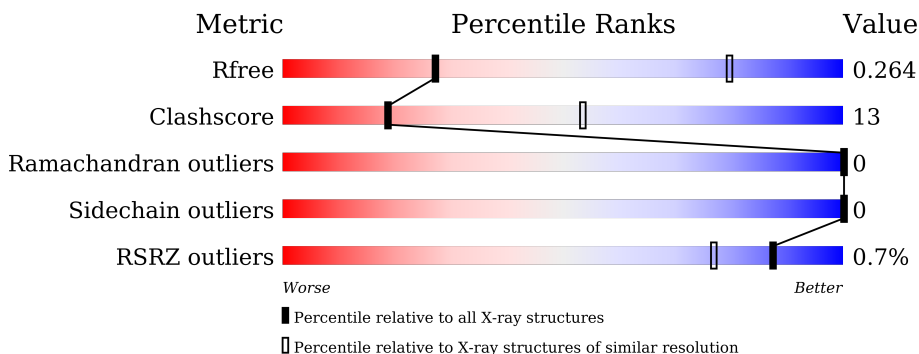
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.58 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	1094 (3.66-3.50)
Clashscore	141614	1181 (3.66-3.50)
Ramachandran outliers	138981	1143 (3.66-3.50)
Sidechain outliers	138945	1143 (3.66-3.50)
RSRZ outliers	127900	1012 (3.66-3.50)

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 $\leq 5\%$. 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	458	 67% 21% 12%
1	B	458	 63% 27% 11%
1	C	458	 63% 27% 10%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9348 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 PHD finger protein 7, Ubiquitin-conjugating enzyme E2 D2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	404	Total 3072	C 1939	N 526	O 572	S 35	0	0	0
1	B	409	Total 3090	C 1942	N 535	O 577	S 36	0	0	0
1	C	412	Total 3163	C 1988	N 550	O 590	S 35	0	0	0

There are 99 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q9DAG9
A	2	PRO	-	expression tag	UNP Q9DAG9
A	3	THR	-	expression tag	UNP Q9DAG9
A	4	GLU	-	expression tag	UNP Q9DAG9
A	5	THR	-	expression tag	UNP Q9DAG9
A	6	SER	-	expression tag	UNP Q9DAG9
A	7	GLN	-	expression tag	UNP Q9DAG9
A	8	VAL	-	expression tag	UNP Q9DAG9
A	9	ALA	-	expression tag	UNP Q9DAG9
A	10	PRO	-	expression tag	UNP Q9DAG9
A	11	ALA	-	expression tag	UNP Q9DAG9
A	12	GLY	-	expression tag	UNP Q9DAG9
A	13	ILE	-	expression tag	UNP Q9DAG9
A	14	GLN	-	expression tag	UNP Q9DAG9
A	984	GLY	-	linker	UNP Q9DAG9
A	985	SER	-	linker	UNP Q9DAG9
A	986	THR	-	linker	UNP Q9DAG9
A	987	GLY	-	linker	UNP Q9DAG9
A	988	SER	-	linker	UNP Q9DAG9
A	989	GLY	-	linker	UNP Q9DAG9
A	990	GLY	-	linker	UNP Q9DAG9
A	991	GLY	-	linker	UNP Q9DAG9
A	992	SER	-	linker	UNP Q9DAG9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	993	GLY	-	linker	UNP Q9DAG9
A	994	GLY	-	linker	UNP Q9DAG9
A	995	GLY	-	linker	UNP Q9DAG9
A	996	SER	-	linker	UNP Q9DAG9
A	997	GLY	-	linker	UNP Q9DAG9
A	998	GLY	-	linker	UNP Q9DAG9
A	999	ALA	-	linker	UNP Q9DAG9
A	1000	SER	-	linker	UNP Q9DAG9
A	1022	ARG	SER	engineered mutation	UNP P62837
A	1085	LYS	CYS	engineered mutation	UNP P62837
B	1	GLY	-	expression tag	UNP Q9DAG9
B	2	PRO	-	expression tag	UNP Q9DAG9
B	3	THR	-	expression tag	UNP Q9DAG9
B	4	GLU	-	expression tag	UNP Q9DAG9
B	5	THR	-	expression tag	UNP Q9DAG9
B	6	SER	-	expression tag	UNP Q9DAG9
B	7	GLN	-	expression tag	UNP Q9DAG9
B	8	VAL	-	expression tag	UNP Q9DAG9
B	9	ALA	-	expression tag	UNP Q9DAG9
B	10	PRO	-	expression tag	UNP Q9DAG9
B	11	ALA	-	expression tag	UNP Q9DAG9
B	12	GLY	-	expression tag	UNP Q9DAG9
B	13	ILE	-	expression tag	UNP Q9DAG9
B	14	GLN	-	expression tag	UNP Q9DAG9
B	984	GLY	-	linker	UNP Q9DAG9
B	985	SER	-	linker	UNP Q9DAG9
B	986	THR	-	linker	UNP Q9DAG9
B	987	GLY	-	linker	UNP Q9DAG9
B	988	SER	-	linker	UNP Q9DAG9
B	989	GLY	-	linker	UNP Q9DAG9
B	990	GLY	-	linker	UNP Q9DAG9
B	991	GLY	-	linker	UNP Q9DAG9
B	992	SER	-	linker	UNP Q9DAG9
B	993	GLY	-	linker	UNP Q9DAG9
B	994	GLY	-	linker	UNP Q9DAG9
B	995	GLY	-	linker	UNP Q9DAG9
B	996	SER	-	linker	UNP Q9DAG9
B	997	GLY	-	linker	UNP Q9DAG9
B	998	GLY	-	linker	UNP Q9DAG9
B	999	ALA	-	linker	UNP Q9DAG9
B	1000	SER	-	linker	UNP Q9DAG9
B	1022	ARG	SER	engineered mutation	UNP P62837

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1085	LYS	CYS	engineered mutation	UNP P62837
C	1	GLY	-	expression tag	UNP Q9DAG9
C	2	PRO	-	expression tag	UNP Q9DAG9
C	3	THR	-	expression tag	UNP Q9DAG9
C	4	GLU	-	expression tag	UNP Q9DAG9
C	5	THR	-	expression tag	UNP Q9DAG9
C	6	SER	-	expression tag	UNP Q9DAG9
C	7	GLN	-	expression tag	UNP Q9DAG9
C	8	VAL	-	expression tag	UNP Q9DAG9
C	9	ALA	-	expression tag	UNP Q9DAG9
C	10	PRO	-	expression tag	UNP Q9DAG9
C	11	ALA	-	expression tag	UNP Q9DAG9
C	12	GLY	-	expression tag	UNP Q9DAG9
C	13	ILE	-	expression tag	UNP Q9DAG9
C	14	GLN	-	expression tag	UNP Q9DAG9
C	295	GLY	-	linker	UNP Q9DAG9
C	296	SER	-	linker	UNP Q9DAG9
C	297	THR	-	linker	UNP Q9DAG9
C	298	GLY	-	linker	UNP Q9DAG9
C	299	SER	-	linker	UNP Q9DAG9
C	300	GLY	-	linker	UNP Q9DAG9
C	301	GLY	-	linker	UNP Q9DAG9
C	302	GLY	-	linker	UNP Q9DAG9
C	303	SER	-	linker	UNP Q9DAG9
C	304	GLY	-	linker	UNP Q9DAG9
C	305	GLY	-	linker	UNP Q9DAG9
C	306	GLY	-	linker	UNP Q9DAG9
C	307	SER	-	linker	UNP Q9DAG9
C	308	GLY	-	linker	UNP Q9DAG9
C	309	GLY	-	linker	UNP Q9DAG9
C	310	ALA	-	linker	UNP Q9DAG9
C	311	SER	-	linker	UNP Q9DAG9
C	333	ARG	SER	engineered mutation	UNP P62837
C	396	LYS	CYS	engineered mutation	UNP P62837

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	8	Total Zn 8 8	0	0
2	B	7	Total Zn 7 7	0	0

Continued on next page...

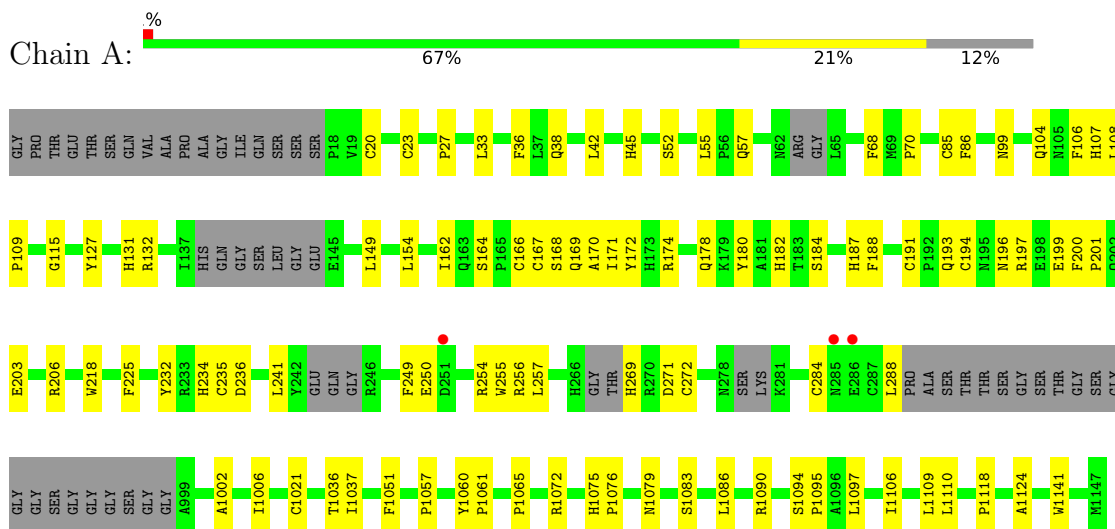
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	8	Total 8	Zn 8	0	0

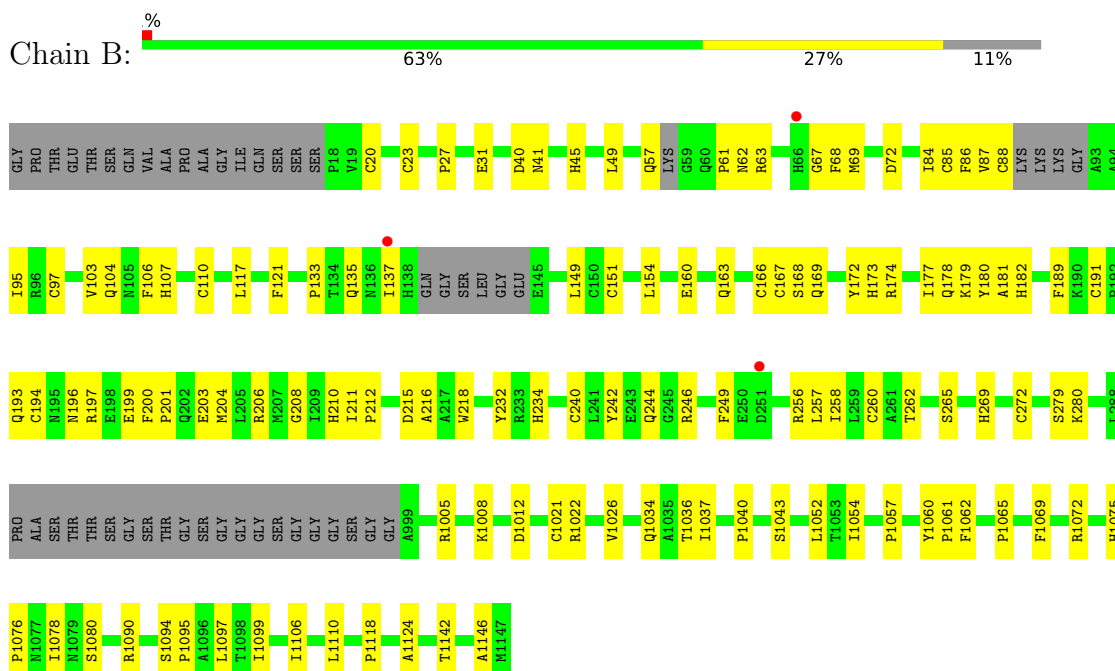
3 Residue-property plots [i](#)

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: PHD finger protein 7, Ubiquitin-conjugating enzyme E2 D2

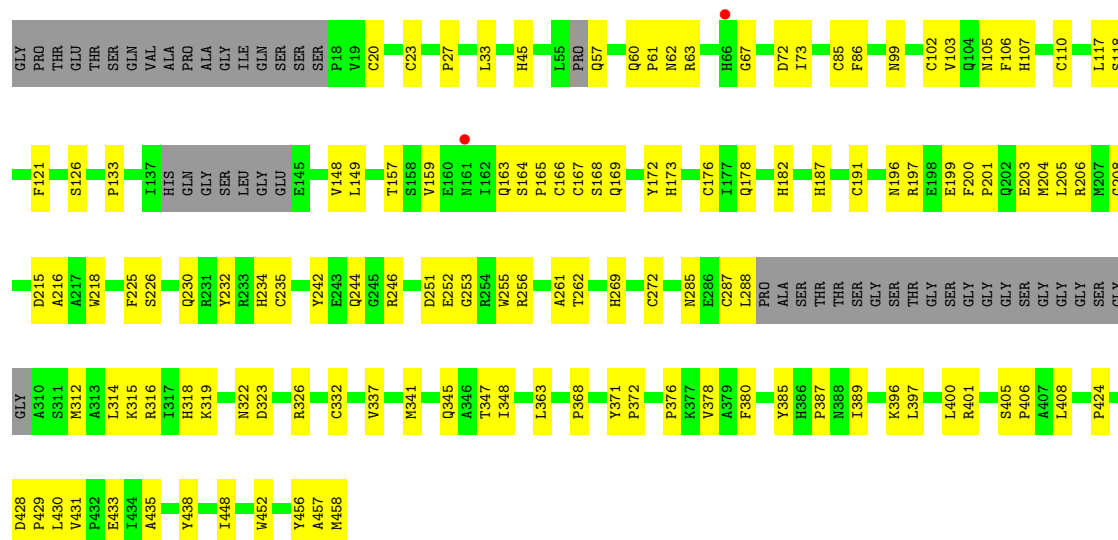


- Molecule 1: PHD finger protein 7, Ubiquitin-conjugating enzyme E2 D2



- Molecule 1: PHD finger protein 7, Ubiquitin-conjugating enzyme E2 D2

Chain C:  63% 27% 10%



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	210.54Å 102.05Å 78.64Å 90.00° 95.92° 90.00°	Depositor
Resolution (Å)	29.89 – 3.58 29.89 – 3.58	Depositor EDS
% Data completeness (in resolution range)	99.5 (29.89-3.58) 91.7 (29.89-3.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.78 (at 3.56Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.216 , 0.265 0.216 , 0.264	Depositor DCC
R_{free} test set	1951 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å ²)	104.5	Xtrriage
Anisotropy	0.630	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 81.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	9348	wwPDB-VP
Average B, all atoms (Å ²)	149.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.74% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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.

5 Model quality

5.1 Standard geometry

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	A	0.37	0/3153	0.58	0/4293
1	B	0.29	0/3172	0.54	0/4320
1	C	0.38	0/3247	0.61	0/4415
All	All	0.35	0/9572	0.58	0/13028

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

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	3072	0	2780	66	0
1	B	3090	0	2788	87	0
1	C	3163	0	2892	82	0
2	A	8	0	0	0	0
2	B	7	0	0	0	0
2	C	8	0	0	0	0
All	All	9348	0	8460	234	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 13.

The worst 5 of 234 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:B:85:CYS:SG	1:B:107:HIS:CG	2.72	0.82
1:A:149:LEU:HD22	1:A:1095:PRO:HB2	1.69	0.75
1:B:189:PHE:HZ	1:B:204:MET:HG3	1.55	0.72
1:A:235:CYS:SG	1:A:269:HIS:N	2.62	0.71
1:B:1061:PRO:HG2	1:B:1095:PRO:HB3	1.75	0.69

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	
1	A	390/458 (85%)	369 (95%)	21 (5%)	0	100	100
1	B	399/458 (87%)	368 (92%)	31 (8%)	0	100	100
1	C	404/458 (88%)	375 (93%)	29 (7%)	0	100	100
All	All	1193/1374 (87%)	1112 (93%)	81 (7%)	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	323/397 (81%)	323 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	322/397 (81%)	322 (100%)	0	100	100
1	C	335/397 (84%)	335 (100%)	0	100	100
All	All	980/1191 (82%)	980 (100%)	0	100	100

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

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 23 ligands modelled in this entry, 23 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

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	A	404/458 (88%)	-0.32	3 (0%) 87 77	80, 136, 228, 281	0
1	B	409/458 (89%)	-0.29	3 (0%) 87 77	73, 147, 239, 337	0
1	C	412/458 (89%)	-0.29	2 (0%) 91 82	91, 150, 215, 291	0
All	All	1225/1374 (89%)	-0.30	8 (0%) 87 77	73, 143, 228, 337	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	251	ASP	3.0
1	A	251	ASP	2.6
1	A	285	ASN	2.3
1	C	66	HIS	2.3
1	B	137	ILE	2.1

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(\AA^2)	Q<0.9
2	ZN	B	503	1/1	0.59	0.10	254,254,254,254	0
2	ZN	A	507	1/1	0.72	0.10	275,275,275,275	0
2	ZN	B	502	1/1	0.83	0.06	289,289,289,289	0
2	ZN	A	504	1/1	0.90	0.17	139,139,139,139	0
2	ZN	B	507	1/1	0.90	0.07	158,158,158,158	0
2	ZN	C	502	1/1	0.91	0.04	203,203,203,203	0
2	ZN	B	501	1/1	0.93	0.10	203,203,203,203	0
2	ZN	C	503	1/1	0.94	0.06	158,158,158,158	0
2	ZN	C	507	1/1	0.94	0.08	145,145,145,145	0
2	ZN	C	504	1/1	0.95	0.12	120,120,120,120	0
2	ZN	B	506	1/1	0.95	0.17	135,135,135,135	0
2	ZN	A	503	1/1	0.96	0.11	110,110,110,110	0
2	ZN	C	505	1/1	0.96	0.09	133,133,133,133	0
2	ZN	B	505	1/1	0.96	0.12	123,123,123,123	0
2	ZN	A	506	1/1	0.97	0.13	180,180,180,180	0
2	ZN	B	504	1/1	0.97	0.15	125,125,125,125	0
2	ZN	C	501	1/1	0.97	0.09	187,187,187,187	0
2	ZN	A	502	1/1	0.97	0.19	121,121,121,121	0
2	ZN	C	506	1/1	0.98	0.08	144,144,144,144	0
2	ZN	A	508	1/1	0.98	0.14	111,111,111,111	0
2	ZN	C	508	1/1	0.98	0.14	104,104,104,104	1
2	ZN	A	501	1/1	0.99	0.14	106,106,106,106	0
2	ZN	A	505	1/1	0.99	0.13	117,117,117,117	0

6.5 Other polymers [i](#)

There are no such residues in this entry.