



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:58 AM GMT

PDB ID : 2YW7
Title : Crystal structure of C-terminal deletion mutant of Mycobacterium smegmatis Dps
Authors : Roy, S.; Saraswathi, R.; Gupta, S.; Sekar, K.; Chatterji, D.; Vijayan, M.
Deposited on : 2007-04-19
Resolution : 3.30 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
<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 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

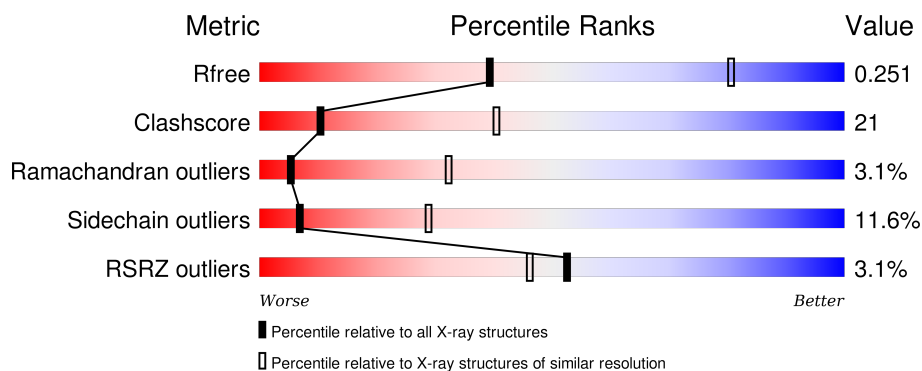
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.30 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	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

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. 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	183	<div> <div>2%</div> <div>45% 30% 5% 20%</div> </div>
1	B	183	<div> <div>1%</div> <div>40% 31% 8% 21%</div> </div>
1	C	183	<div> <div>2%</div> <div>43% 33% 5% 20%</div> </div>
1	D	183	<div> <div>4%</div> <div>41% 34% • • 21%</div> </div>
1	E	183	<div> <div>2%</div> <div>39% 34% 6% 21%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	183	<div><div><div>%</div><div><div></div><div>44%</div><div>29%</div><div>7%</div><div>20%</div></div></div></div>
1	G	183	<div><div><div>%</div><div><div></div><div>47%</div><div>29%</div><div></div><div>21%</div></div></div></div>
1	H	183	<div><div><div>4%</div><div><div></div><div>33%</div><div>42%</div><div></div><div>20%</div></div></div></div>
1	I	183	<div><div><div>4%</div><div><div></div><div>41%</div><div>32%</div><div>6%</div><div>21%</div></div></div></div>
1	J	183	<div><div><div>5%</div><div><div></div><div>48%</div><div>27%</div><div>5%</div><div>20%</div></div></div></div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 11597 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.

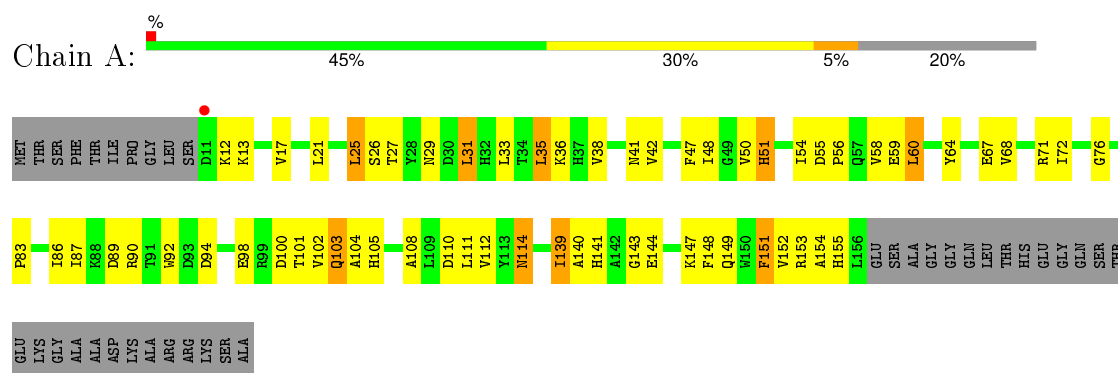
- Molecule 1 is a protein called Starvation-induced DNA protecting protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	146	Total	C	N	O	S	0	0	0
			1164	739	201	223	1			
1	B	145	Total	C	N	O	S	0	0	0
			1156	733	200	222	1			
1	C	147	Total	C	N	O	S	0	0	0
			1173	744	202	226	1			
1	D	145	Total	C	N	O	S	0	0	0
			1155	733	200	221	1			
1	E	145	Total	C	N	O	S	0	0	0
			1156	733	200	222	1			
1	F	146	Total	C	N	O	S	0	0	0
			1160	737	201	221	1			
1	G	144	Total	C	N	O	S	0	0	0
			1147	729	198	219	1			
1	H	146	Total	C	N	O	S	0	0	0
			1165	740	201	223	1			
1	I	145	Total	C	N	O	S	0	0	0
			1156	735	200	220	1			
1	J	146	Total	C	N	O	S	0	0	0
			1165	740	201	223	1			

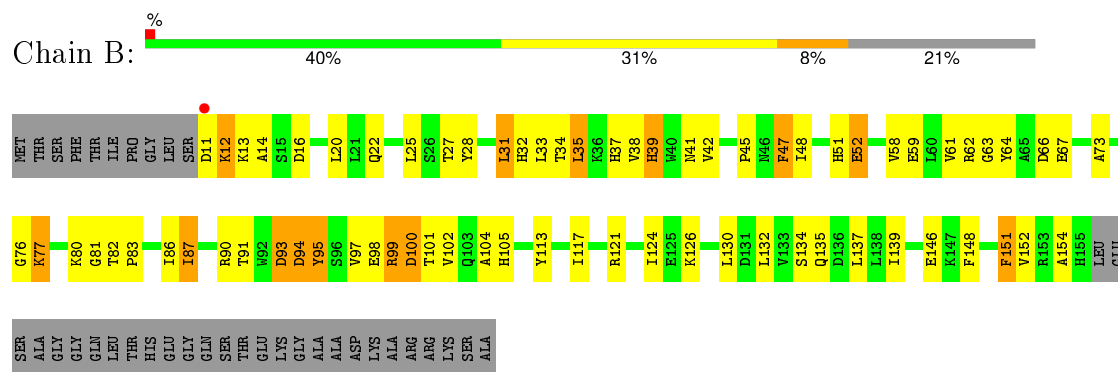
3 Residue-property plots [i](#)

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.

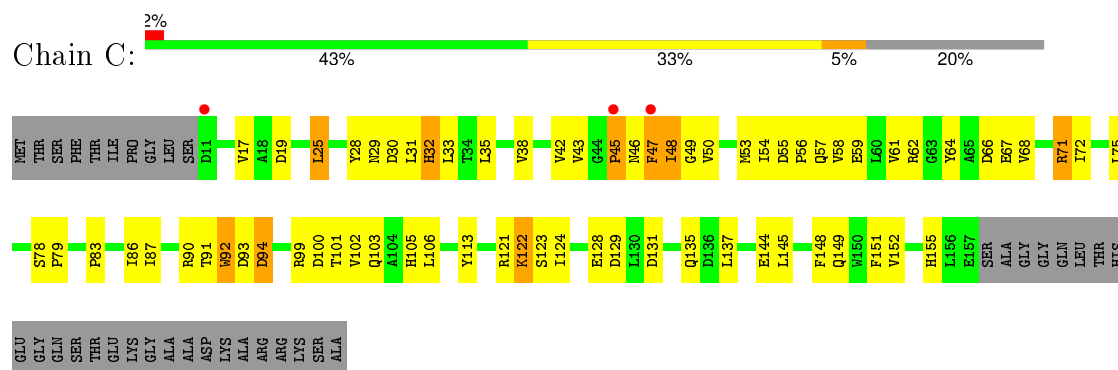
• Molecule 1: Starvation-induced DNA protecting protein



• Molecule 1: Starvation-induced DNA protecting protein

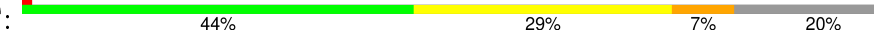


• Molecule 1: Starvation-induced DNA protecting protein



- [illegible]

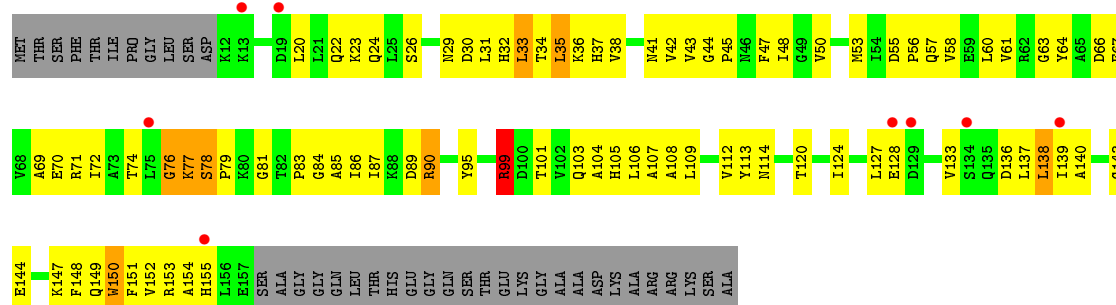
- Chain E:
-
- Sequence logo for Chain E showing amino acid conservation across 100 positions. The y-axis represents information content in bits. The x-axis shows positions 1 to 100. The sequence is: MET, THR, SER, PHE, THR, ILE, PRO, GLY, LEU, SER, D11, K12, K13, D16, V17, L20, L21, Q24, L25, L26, T27, L31, H32, L33, T34, L35, K36, K37, V38, H39, V42, P45, L48, T49, V49, V50, H51, M53, L54, D55, P56, Q57, V58, E59, L60, V61, V64, V68, R71. Notable conservation is seen at positions 11-13 (D11, K12, K13) and position 49 (T49).

- Chain F: 

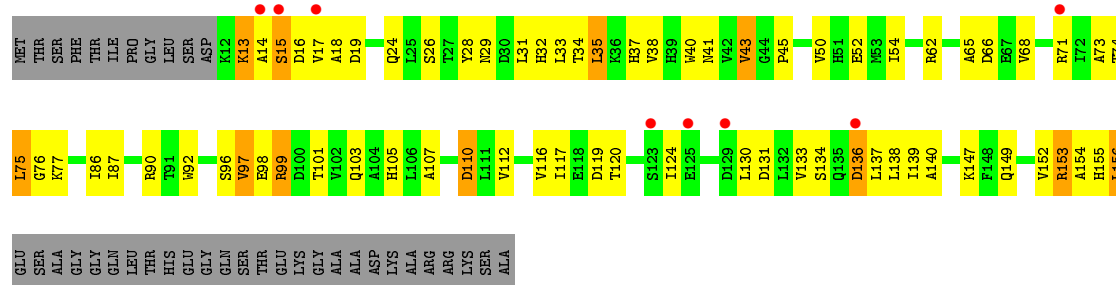
- Chain G:
-
- 47% 29% 21%
- MET THR SER PHE THR ILE PRO GLY LEU SER ASP LYS K13 A14 S15 D16 V17 L21 Q22 K23 Q24 L25 Y28 L31 H32 L33 T34 L35 K36 H37 V38 H39 W40 M41 V42 V43 G44 P45 M46 F47 V50 H51 F52 M53 I54 D55 P56 Q57 V58 F59 L60 V61 R62 G63 Y64 M65 D66 E67 V68 R71 I72 A73 T74 L75 G76 K77 S78 F79 K80 I86 L87 R90 Y95 R99 Q103 L106 A107 A108 L109 D110 L111 V112 T120 I124 L132 Q135 I139 F151 H155 L156 GLU SER ALA GLY GLN LEU THR HIS GLU GLY TYR

SER
THR
GLU
LYS
GLY
ALA
ALA
ASP
LYS
ARG
ARG
LYS
SER
ALA

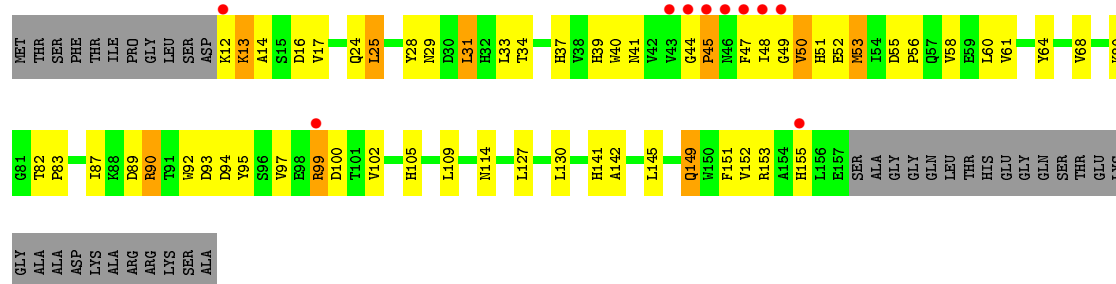
• Molecule 1: Starvation-induced DNA protecting protein



• Molecule 1: Starvation-induced DNA protecting protein



• Molecule 1: Starvation-induced DNA protecting protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	96.52Å 83.94Å 111.91Å 90.00° 105.89° 90.00°	Depositor
Resolution (Å)	28.54 – 3.30 29.39 – 3.30	Depositor EDS
% Data completeness (in resolution range)	100.0 (28.54-3.30) 99.2 (29.39-3.30)	Depositor EDS
R_{merge}	0.01	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.61 (at 3.31Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.222 , 0.259 0.216 , 0.251	Depositor DCC
R_{free} test set	1318 reflections (5.35%)	DCC
Wilson B-factor (Å ²)	55.2	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 46.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 25935 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	11597	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.57% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.36	0/1187	0.57	0/1612
1	B	0.33	0/1179	0.58	0/1601
1	C	0.33	0/1196	0.56	0/1624
1	D	0.34	0/1178	0.54	0/1600
1	E	0.34	0/1179	0.56	0/1601
1	F	0.35	0/1183	0.52	0/1607
1	G	0.34	0/1170	0.56	0/1590
1	H	0.35	0/1188	0.55	0/1613
1	I	0.34	0/1179	0.54	0/1601
1	J	0.35	0/1188	0.57	0/1613
All	All	0.34	0/11827	0.56	0/16062

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	1164	0	1153	61	0
1	B	1156	0	1142	58	0
1	C	1173	0	1159	49	0
1	D	1155	0	1139	52	0
1	E	1156	0	1142	65	0
1	F	1160	0	1149	50	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	1147	0	1136	47	0
1	H	1165	0	1155	80	0
1	I	1156	0	1149	46	0
1	J	1165	0	1155	40	0
All	All	11597	0	11479	496	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 21.

The worst 5 of 496 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:79:PRO:HD2	1:F:41:ASN:HD21	1.14	1.08
1:D:153:ARG:HB3	1:D:155:HIS:CE1	1.95	1.01
1:I:13:LYS:HE3	1:I:13:LYS:H	1.26	0.98
1:H:147:LYS:HA	1:H:150:TRP:HE3	1.34	0.93
1:H:147:LYS:HA	1:H:150:TRP:CE3	2.04	0.93

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	144/183 (79%)	127 (88%)	16 (11%)	1 (1%)	26	66
1	B	143/183 (78%)	118 (82%)	20 (14%)	5 (4%)	4	29
1	C	145/183 (79%)	121 (83%)	22 (15%)	2 (1%)	14	50
1	D	143/183 (78%)	120 (84%)	15 (10%)	8 (6%)	2	16
1	E	143/183 (78%)	121 (85%)	17 (12%)	5 (4%)	4	29
1	F	144/183 (79%)	120 (83%)	21 (15%)	3 (2%)	9	42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	G	142/183 (78%)	124 (87%)	18 (13%)	0	100	100
1	H	144/183 (79%)	114 (79%)	24 (17%)	6 (4%)	3	23
1	I	143/183 (78%)	110 (77%)	24 (17%)	9 (6%)	2	13
1	J	144/183 (79%)	113 (78%)	25 (17%)	6 (4%)	3	23
All	All	1435/1830 (78%)	1188 (83%)	202 (14%)	45 (3%)	5	32

5 of 45 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	77	LYS
1	D	47	PHE
1	D	105	HIS
1	E	45	PRO
1	E	91	THR

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	126/153 (82%)	110 (87%)	16 (13%)	5	24
1	B	125/153 (82%)	111 (89%)	14 (11%)	7	30
1	C	127/153 (83%)	107 (84%)	20 (16%)	3	15
1	D	124/153 (81%)	115 (93%)	9 (7%)	17	53
1	E	125/153 (82%)	112 (90%)	13 (10%)	9	34
1	F	125/153 (82%)	109 (87%)	16 (13%)	5	23
1	G	124/153 (81%)	112 (90%)	12 (10%)	10	38
1	H	126/153 (82%)	111 (88%)	15 (12%)	6	27
1	I	125/153 (82%)	111 (89%)	14 (11%)	7	30
1	J	126/153 (82%)	110 (87%)	16 (13%)	5	24
All	All	1253/1530 (82%)	1108 (88%)	145 (12%)	7	29

5 of 145 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	86	ILE
1	F	99	ARG
1	J	50	VAL
1	E	106	LEU
1	F	38	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	57	GLN
1	F	103	GLN
1	J	41	ASN
1	E	149	GLN
1	F	24	GLN

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](#)

There are no ligands in this entry.

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 ⓘ

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, 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	146/183 (79%)	-0.63	1 (0%) 89 86	8, 17, 34, 94	0
1	B	145/183 (79%)	-0.42	1 (0%) 89 86	11, 24, 75, 96	0
1	C	147/183 (80%)	-0.21	3 (2%) 68 62	13, 32, 86, 126	0
1	D	145/183 (79%)	-0.13	7 (4%) 34 28	21, 39, 102, 138	0
1	E	145/183 (79%)	-0.50	3 (2%) 67 60	9, 21, 56, 100	0
1	F	146/183 (79%)	-0.23	2 (1%) 78 73	13, 36, 67, 84	0
1	G	144/183 (78%)	-0.35	2 (1%) 78 73	10, 24, 54, 96	0
1	H	146/183 (79%)	0.22	8 (5%) 29 23	17, 69, 119, 154	0
1	I	145/183 (79%)	0.23	8 (5%) 29 23	33, 73, 115, 136	0
1	J	146/183 (79%)	0.04	10 (6%) 20 17	19, 48, 109, 158	0
All	All	1455/1830 (79%)	-0.20	45 (3%) 52 46	8, 34, 101, 158	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	11	ASP	5.3
1	G	14	ALA	4.9
1	C	11	ASP	4.7
1	J	12	LYS	4.1
1	I	129	ASP	3.6

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.