WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 09:41 AM EDT

PDB ID	:	5D4T
Title	:	SeMet-labelled HcgC from Methanocaldococcus jannaschii in space group P212121
Authors	:	Fujishiro, T.; Ermler, U.; Shima, S.
Deposited on	:	2015-08-09
Resolution	:	2.90 Å(reported)

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

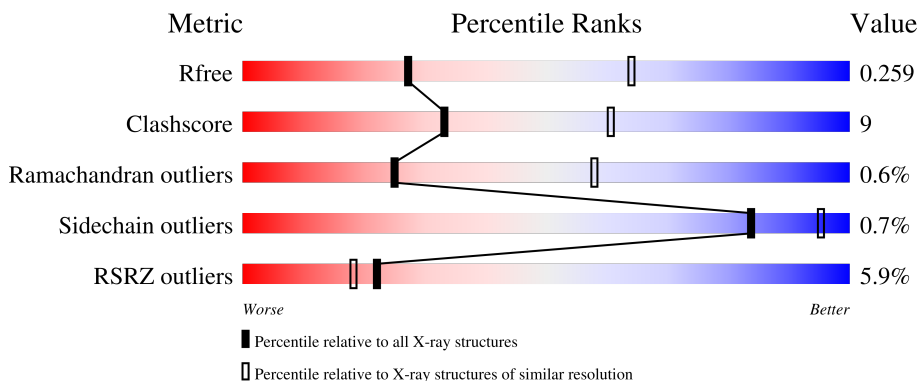|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.90 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.
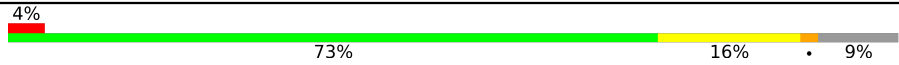
Metric	Percentile Ranks	Value
Rfree		0.259
Clashscore		9
Ramachandran outliers		0.6%
Sidechain outliers		0.7%
RSRZ outliers		5.9%

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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	A	268	9% — 69% 19% ·· 9%
1	B	268	.% — 75% 16% · 8%
1	C	268	7% — 72% 18% 9%
1	D	268	2% — 74% 17% 9%
1	E	268	11% — 68% 20% ·· 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	268	4% 73% 16% 9%
1	G	268	6% 78% 12% 9%
1	H	268	3% 74% 16% 8%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 15781 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 Uncharacterized protein MJ0489.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	243	Total	C	N	O	Se	0	0	0
			1961	1274	309	375	3			
1	B	246	Total	C	N	O	Se	0	0	0
			1983	1285	314	381	3			
1	C	243	Total	C	N	O	Se	0	0	0
			1961	1274	309	375	3			
1	D	244	Total	C	N	O	Se	0	0	0
			1967	1277	310	377	3			
1	E	243	Total	C	N	O	Se	0	1	0
			1967	1278	309	377	3			
1	F	244	Total	C	N	O	Se	0	0	0
			1967	1277	310	377	3			
1	G	244	Total	C	N	O	Se	0	1	0
			1973	1281	310	379	3			
1	H	247	Total	C	N	O	Se	0	0	0
			1991	1289	316	383	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	initiating methionine	UNP Q57913
B	1	MSE	-	initiating methionine	UNP Q57913
C	1	MSE	-	initiating methionine	UNP Q57913
D	1	MSE	-	initiating methionine	UNP Q57913
E	1	MSE	-	initiating methionine	UNP Q57913
F	1	MSE	-	initiating methionine	UNP Q57913
G	1	MSE	-	initiating methionine	UNP Q57913
H	1	MSE	-	initiating methionine	UNP Q57913

- Molecule 2 is water.

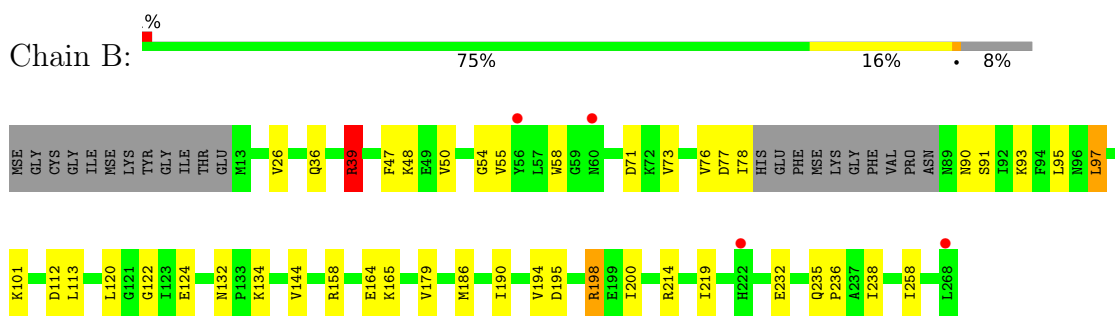| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	O 1	0	0
2	B	3	Total 3	O 3	0	0
2	C	1	Total 1	O 1	0	0
2	D	2	Total 2	O 2	0	0
2	F	1	Total 1	O 1	0	0
2	G	1	Total 1	O 1	0	0
2	H	2	Total 2	O 2	0	0

3   Residue-property plots ⓘ

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.
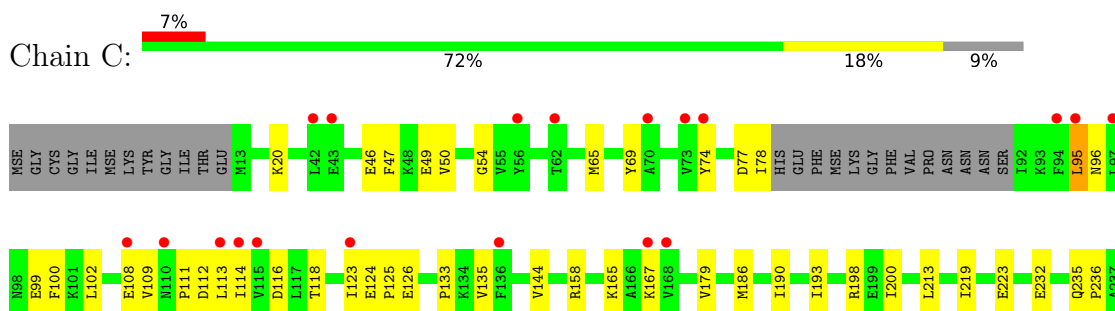
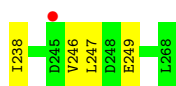- Molecule 1: Uncharacterized protein MJ0489

Chain A:

- Molecule 1: Uncharacterized protein MJ0489

Chain B:

- Molecule 1: Uncharacterized protein MJ0489

Chain C:

I238 D245 V246 L247 D248 E249 L268

• Molecule 1: Uncharacterized protein MJ0489

Chain D:

2%
74%
17%
9%

MSE GLY CYS GLY ILE MSE LYS TYR GLY ILE THR GLU M13 V23 Q36 R39 E43 G44 K48 E49 V50 V55 Y56 Y61 M65 L66 S67 D71 K72 V73 Y74 L75 V76 D77 I78 HIS GLU PHE MSE LYS GLY PHE VAL PRO ASN ASN ASN S91 I92 K93 F94

L95 L102 N110 P111 D112 L113 L120 N132 V144 I149 F162 I163 E164 K165 A166 V170 V179 M186 I190 D195 R198 E199 I200 R214 Y215 Y216 D225 E232 Q235 P236 A237 I238 L268

• Molecule 1: Uncharacterized protein MJ0489

Chain E:

11%
68%
20%
9%

MSE GLY CYS GLY ILE MSE LYS TYR GLY ILE THR GLU M13 V14 K15 E29 Y35 R39 L42 E43 G44 E45 E46 F47 K48 E49 V50 V51 I52 F53 G54 V55 Y56 M65 L66 S67 K68 Y69 A70 D71 K72 V73 Y74 L75 V76 D77 I78 HIS GLU PHE MSE LYS GLY PHE VAL

PRO ASN ASN SER I92 K93 F94 L95 N96 E99 F100 K101 L102 K103 F104 I105 E108 V109 N110 P111 D112 L113 I114 L120 I123 F127 P133 K134 V135 V138 V144 Y156 T159 I163 E164 K165 A166 K167 V170 V179 I193 R198 L213
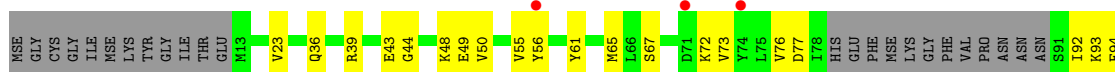
R214 L220 F221 H222 E223 N224 D225 K228 E232 Q235 I258 L268

• Molecule 1: Uncharacterized protein MJ0489

Chain F:

4%
73%
16%
9%

MSE GLY CYS GLY ILE MSE LYS TYR GLY ILE THR GLU M13 K20 Y35 Q36 R39 E43 G44 E49 V50 F53 Y56 M65 L66 S67 D71 K72 V73 Y74 L75 V76 D77 I78 HIS GLU PHE MSE LYS GLY PHE VAL PRO ASN ASN ASN S91 I92 K93 F94 L95

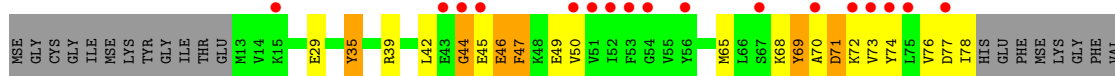N96 E99 F100 K103 E108 V109 N110 P111 L120 E124 P125 V135 V144 I149 R158 E164 K165 A166 K167 V179 D191 T192 I193 I200 L207 L213 R214 H222 Q235 P236 A237 I238 D245 V246 L247 L268

• Molecule 1: Uncharacterized protein MJ0489

Chain G:

6%
78%
12%
9%

MSE GLY CYS GLY ILE MSE LYS TYR GLY ILE THR GLU M13 I17 E43 E49 V50 V51 I52 Y56 N60 L66 S67 Y74 L75 V76 D77 I78 HIS GLU PHE MSE LYS GLY PHE VAL PRO ASN ASN ASN S91 I92 L95 N96 E99 F100 K101 L102 K103 F104 I105

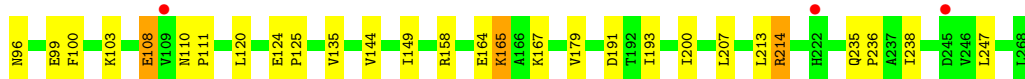R106 G107 E108 V109 N110 P111 D112 L113 L120 E124 P125 F136 V144 I149 R158 T159 A160 P161 F162 I163 E164 V179 M186 I190 I193 V195 D198 L213 R214 F221 D245 L268
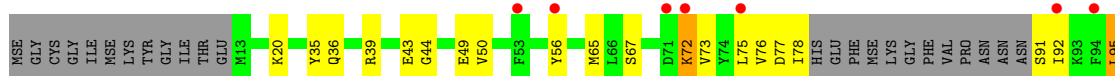
• Molecule 1: Uncharacterized protein MJ0489

Chain H:

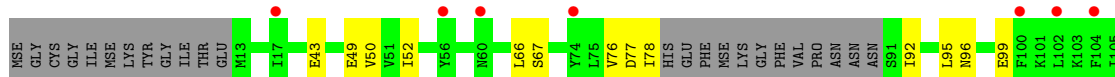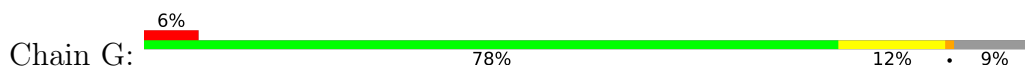# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	141.43Å 145.00Å 150.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.56 – 2.90 48.56 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.56-2.90) 100.0 (48.56-2.90)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.59 (at 2.91Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.228 , 0.259 0.231 , 0.259	Depositor DCC
R_{free} test set	3501 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	81.2	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 51.2	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.000 for k,h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	15781	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.55 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.4662e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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)

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	0.74	9/1994 (0.5%)	1.17	28/2694 (1.0%)
1	B	0.36	0/2016	0.73	4/2724 (0.1%)
1	C	0.38	0/1994	0.78	4/2694 (0.1%)
1	D	0.35	0/2000	0.68	0/2702
1	E	0.47	1/2003 (0.0%)	0.94	13/2706 (0.5%)
1	F	0.42	0/2000	0.86	5/2702 (0.2%)
1	G	0.34	0/2009	0.67	2/2714 (0.1%)
1	H	0.52	1/2024 (0.0%)	0.96	12/2735 (0.4%)
All	All	0.46	11/16040 (0.1%)	0.86	68/21671 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2
1	E	0	1
1	F	0	1
All	All	0	4

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	198	ARG	CZ-NH2	-14.91	1.13	1.33
1	A	198	ARG	CG-CD	-10.39	1.25	1.51
1	H	43	GLU	CD-OE2	-9.48	1.15	1.25
1	A	39	ARG	CG-CD	-6.81	1.34	1.51
1	A	69	TYR	C-O	6.38	1.35	1.23

The worst 5 of 68 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	40	ASP	CB-CG-OD1	-16.38	103.56	118.30
1	A	198	ARG	CG-CD-NE	-16.28	77.61	111.80
1	E	214	ARG	NE-CZ-NH2	-14.80	112.90	120.30
1	A	71	ASP	CB-CG-OD2	-13.92	105.77	118.30
1	A	39	ARG	NE-CZ-NH2	12.89	126.75	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	46	GLU	Sidechain
1	A	71	ASP	Sidechain
1	E	93	LYS	Peptide
1	F	108	GLU	Sidechain

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	1961	0	1990	54	1
1	B	1983	0	2007	35	0
1	C	1961	0	1990	31	0
1	D	1967	0	1995	26	1
1	E	1967	0	1996	52	1
1	F	1967	0	1995	49	0
1	G	1973	0	2001	22	0
1	H	1991	0	2013	32	1
2	A	1	0	0	0	0
2	B	3	0	0	0	0
2	C	1	0	0	0	0
2	D	2	0	0	0	0
2	F	1	0	0	0	0
2	G	1	0	0	0	0
2	H	2	0	0	1	0
All	All	15781	0	15987	273	2

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 9.

The worst 5 of 273 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:39:ARG:NE	1:A:69:TYR:OH	1.77	1.17
1:A:46:GLU:N	1:A:46:GLU:OE1	1.91	1.02
1:A:198:ARG:HH21	1:B:198:ARG:NH1	1.59	0.97
1:G:67:SER:HB3	1:G:92:ILE:HD11	1.51	0.92
1:A:214:ARG:NH2	1:B:232:GLU:OE1	2.03	0.92

All (2) 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:D:48:LYS:NZ	1:E:71:ASP:OD1[4_465]	2.00	0.20
1:A:46:GLU:OE2	1:H:39:ARG:NH2[3_655]	2.03	0.17

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	239/268 (89%)	233 (98%)	3 (1%)	3 (1%)	12	37
1	B	242/268 (90%)	240 (99%)	1 (0%)	1 (0%)	34	66
1	C	239/268 (89%)	233 (98%)	6 (2%)	0	100	100
1	D	240/268 (90%)	233 (97%)	4 (2%)	3 (1%)	12	37
1	E	240/268 (90%)	232 (97%)	7 (3%)	1 (0%)	34	66
1	F	240/268 (90%)	231 (96%)	8 (3%)	1 (0%)	34	66
1	G	241/268 (90%)	238 (99%)	3 (1%)	0	100	100
1	H	243/268 (91%)	237 (98%)	4 (2%)	2 (1%)	19	51
All	All	1924/2144 (90%)	1877 (98%)	36 (2%)	11 (1%)	25	58

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	43	GLU
1	A	46	GLU
1	D	164	GLU
1	H	90	ASN
1	B	164	GLU

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	218/233 (94%)	216 (99%)	2 (1%)	78	93
1	B	221/233 (95%)	220 (100%)	1 (0%)	88	96
1	C	218/233 (94%)	216 (99%)	2 (1%)	78	93
1	D	219/233 (94%)	218 (100%)	1 (0%)	88	96
1	E	219/233 (94%)	216 (99%)	3 (1%)	67	89
1	F	219/233 (94%)	219 (100%)	0	100	100
1	G	220/233 (94%)	218 (99%)	2 (1%)	78	93
1	H	222/233 (95%)	220 (99%)	2 (1%)	78	93
All	All	1756/1864 (94%)	1743 (99%)	13 (1%)	84	95

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	93	LYS
1	E	95	LEU
1	H	43	GLU
1	G	112	ASP
1	H	40	ASP

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

Mol	Chain	Res	Type
1	B	90	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	132	ASN
1	E	36	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

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(Å²)	Q<0.9
1	A	240/268 (89%)	0.54	23 (9%)	8	6	54, 82, 112, 134	0
1	B	243/268 (90%)	0.13	4 (1%)	72	71	55, 72, 98, 117	0
1	C	240/268 (89%)	0.47	20 (8%)	11	8	61, 92, 141, 166	0
1	D	241/268 (89%)	0.14	5 (2%)	63	61	57, 78, 108, 133	0
1	E	240/268 (89%)	0.69	29 (12%)	4	3	62, 97, 161, 180	0
1	F	241/268 (89%)	0.26	10 (4%)	37	32	60, 82, 108, 133	0
1	G	241/268 (89%)	0.42	15 (6%)	20	16	49, 78, 112, 129	0
1	H	244/268 (91%)	0.28	7 (2%)	51	47	53, 76, 107, 135	0
All	All	1930/2144 (90%)	0.37	113 (5%)	22	18	49, 81, 128, 180	0

The worst 5 of 113 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	56	TYR	9.3
1	E	74	TYR	6.5
1	E	92	ILE	5.4
1	H	268	LEU	5.4
1	E	100	PHE	5.3

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 ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.