



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 02:41 am BST

PDB ID : 3G7G  
Title : Crystal structure of the protein with unknown function from *Clostridium acetobutylicum* ATCC 824  
Authors : Zhang, R.; Wu, R.; Abdullah, J.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2009-02-09  
Resolution : 1.99 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](mailto: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 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 : 2.11  
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) : 2.11

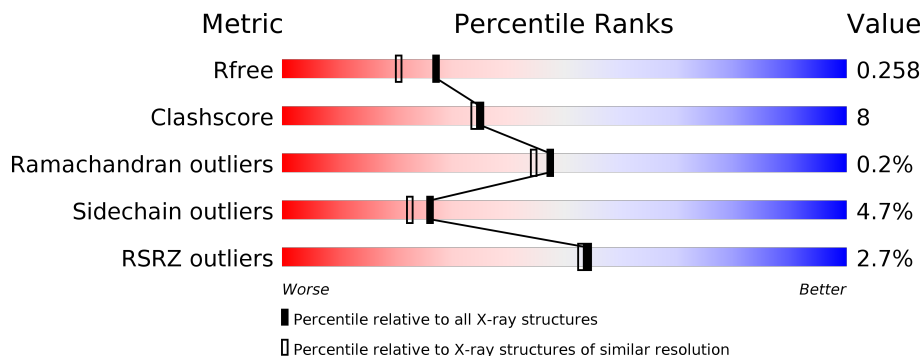
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.99 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.



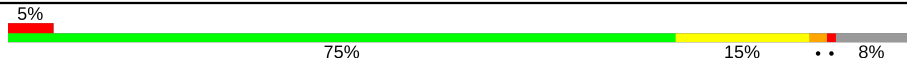

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	158	<p>71% 20% 6%</p>
1	B	158	<p>73% 15% 9%</p>
1	C	158	<p>70% 19% 9%</p>
1	D	158	<p>79% 13% 6%</p>
1	E	158	<p>73% 12% 15%</p>
1	F	158	<p>72% 20% 6%</p>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	158	
1	H	158	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 9804 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 UPF0311 protein CA\_C3321.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	149	1189	763	197	223	6	0	0	0
1	B	144	1150	741	192	211	6	0	0	0
1	D	148	1180	758	196	220	6	0	0	0
1	E	135	1076	693	182	195	6	0	0	0
1	C	144	1149	739	191	213	6	0	0	0
1	F	148	1185	764	196	219	6	0	0	0
1	G	146	1160	746	194	214	6	0	0	0
1	H	152	1209	776	202	225	6	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	48	Total 48	O 48	0	0
2	B	72	Total 72	O 72	0	0
2	D	53	Total 53	O 53	0	0
2	E	70	Total 70	O 70	0	0
2	C	63	Total 63	O 63	0	0
2	F	68	Total 68	O 68	0	0

*Continued on next page...*

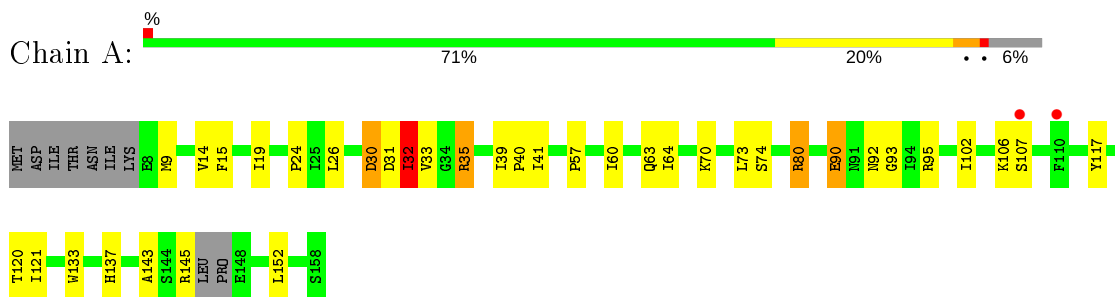
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
2	G	69	Total O 69 69	0	0
2	H	63	Total O 63 63	0	0

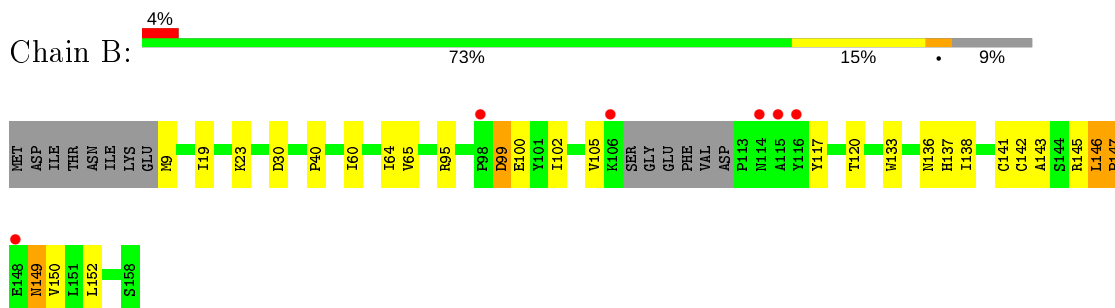
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

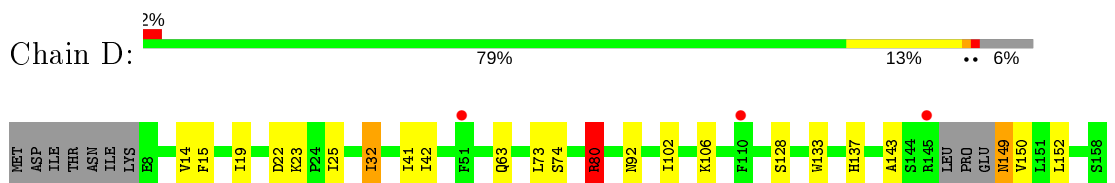
- Molecule 1: UPF0311 protein CA\_C3321



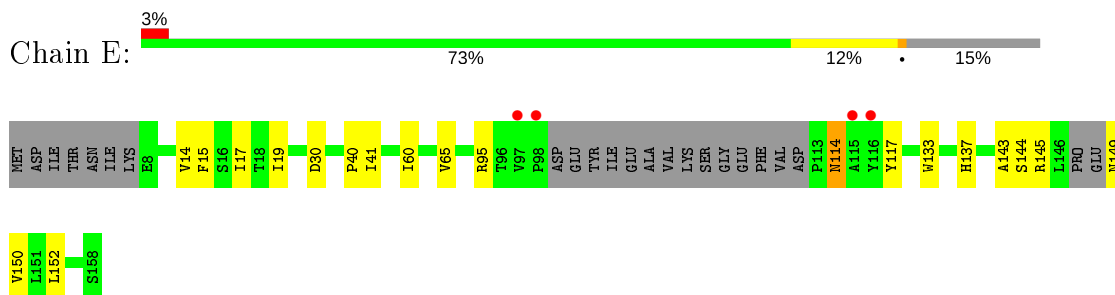
- Molecule 1: UPF0311 protein CA\_C3321



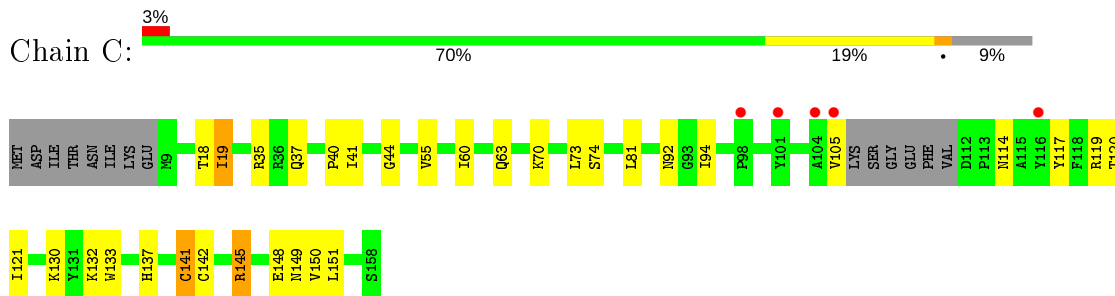
- Molecule 1: UPF0311 protein CA\_C3321



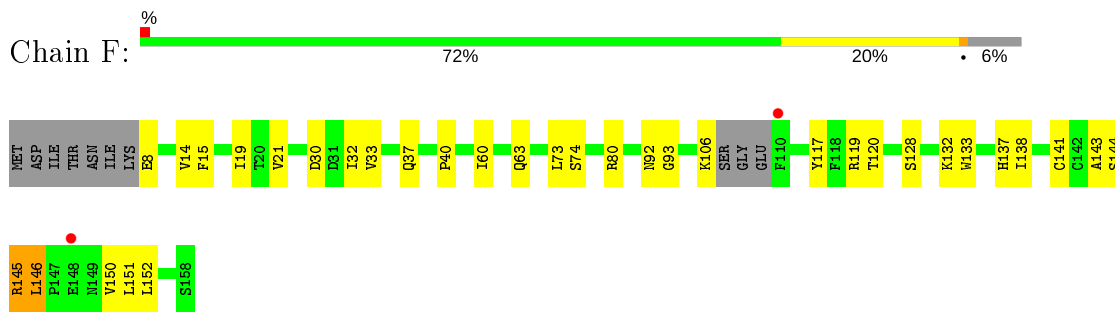
- Molecule 1: UPF0311 protein CA\_C3321



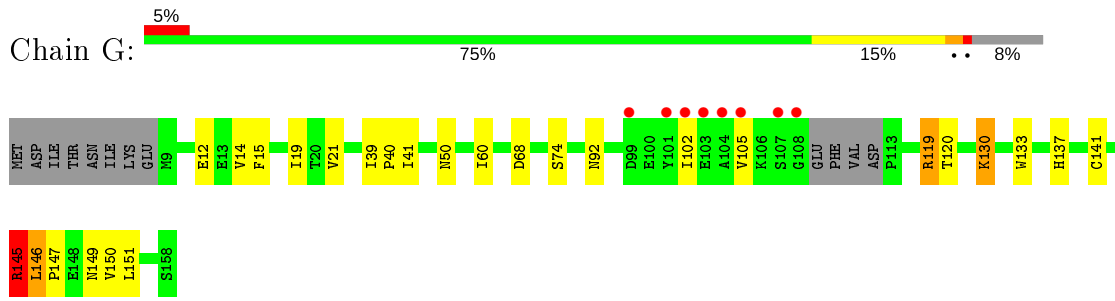
• Molecule 1: UPF0311 protein CA\_C3321



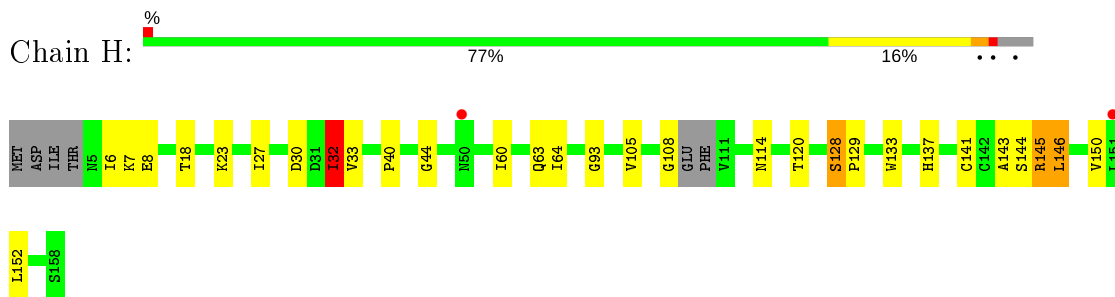
• Molecule 1: UPF0311 protein CA\_C3321



• Molecule 1: UPF0311 protein CA\_C3321



• Molecule 1: UPF0311 protein CA\_C3321



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.05Å 166.84Å 66.97Å 90.00° 119.06° 90.00°	Depositor
Resolution (Å)	58.62 – 1.99 33.98 – 1.99	Depositor EDS
% Data completeness (in resolution range)	97.9 (58.62-1.99) 97.6 (33.98-1.99)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.56 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.193 , 0.254 0.203 , 0.258	Depositor DCC
$R_{free}$ test set	4251 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.8	Xtriage
Anisotropy	0.543	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 26.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.014 for l,k,-h-l 0.014 for -h-l,k,h 0.017 for -h-l,-k,l 0.019 for h,-k,-h-l 0.477 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	9804	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.36% 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, 0.333 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	0.77	0/1214	0.81	3/1642 (0.2%)
1	B	0.75	1/1175 (0.1%)	0.78	0/1590
1	C	0.77	1/1174 (0.1%)	0.78	1/1591 (0.1%)
1	D	0.77	0/1205	0.79	2/1630 (0.1%)
1	E	0.78	0/1098	0.79	0/1483
1	F	0.82	1/1211 (0.1%)	0.82	1/1640 (0.1%)
1	G	0.76	1/1185 (0.1%)	0.82	2/1603 (0.1%)
1	H	0.82	1/1234 (0.1%)	0.86	2/1670 (0.1%)
All	All	0.78	5/9496 (0.1%)	0.81	11/12849 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	141	CYS	CB-SG	-8.00	1.68	1.82
1	F	141	CYS	CB-SG	-6.32	1.71	1.82
1	H	105	VAL	CB-CG1	-5.53	1.41	1.52
1	G	130	LYS	CE-NZ	5.20	1.62	1.49
1	B	141	CYS	CB-SG	-5.20	1.73	1.81

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	145	ARG	NE-CZ-NH1	-7.28	116.66	120.30
1	A	35	ARG	NE-CZ-NH1	6.30	123.45	120.30
1	C	145	ARG	NE-CZ-NH1	6.25	123.43	120.30
1	A	32	ILE	CB-CA-C	-6.23	99.14	111.60
1	D	80	ARG	NE-CZ-NH2	-6.09	117.25	120.30

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	1189	0	1181	26	0
1	B	1150	0	1158	21	0
1	C	1149	0	1148	30	0
1	D	1180	0	1175	12	0
1	E	1076	0	1087	13	0
1	F	1185	0	1185	19	0
1	G	1160	0	1166	20	0
1	H	1209	0	1214	25	0
2	A	48	0	0	2	0
2	B	72	0	0	0	0
2	C	63	0	0	7	0
2	D	53	0	0	1	0
2	E	70	0	0	3	0
2	F	68	0	0	1	0
2	G	69	0	0	4	0
2	H	63	0	0	1	0
All	All	9804	0	9314	152	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 8.

The worst 5 of 152 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:G:151:LEU:HB2	2:G:215:HOH:O	1.80	0.80
1:G:151:LEU:HD23	1:H:6:ILE:HG13	1.63	0.80
1:D:22:ASP:HB2	1:D:42:ILE:HD11	1.68	0.75
1:F:143:ALA:HB2	1:F:152:LEU:HD23	1.68	0.74
1:F:93:GLY:HA3	1:F:120:THR:HG22	1.74	0.70

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	145/158 (92%)	144 (99%)	1 (1%)	0	100	100
1	B	140/158 (89%)	134 (96%)	4 (3%)	2 (1%)	11	5
1	C	140/158 (89%)	138 (99%)	2 (1%)	0	100	100
1	D	144/158 (91%)	141 (98%)	3 (2%)	0	100	100
1	E	129/158 (82%)	128 (99%)	1 (1%)	0	100	100
1	F	144/158 (91%)	140 (97%)	4 (3%)	0	100	100
1	G	142/158 (90%)	140 (99%)	2 (1%)	0	100	100
1	H	148/158 (94%)	144 (97%)	4 (3%)	0	100	100
All	All	1132/1264 (90%)	1109 (98%)	21 (2%)	2 (0%)	47	44

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	147	PRO
1	B	146	LEU

### 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	131/140 (94%)	125 (95%)	6 (5%)	27	23
1	B	127/140 (91%)	120 (94%)	7 (6%)	21	17
1	C	127/140 (91%)	122 (96%)	5 (4%)	32	30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	130/140 (93%)	124 (95%)	6 (5%)	27	23
1	E	119/140 (85%)	115 (97%)	4 (3%)	37	36
1	F	131/140 (94%)	125 (95%)	6 (5%)	27	23
1	G	128/140 (91%)	123 (96%)	5 (4%)	32	30
1	H	134/140 (96%)	125 (93%)	9 (7%)	16	11
All	All	1027/1120 (92%)	979 (95%)	48 (5%)	26	22

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	145	ARG
1	C	130	LYS
1	H	128	SER
1	E	150	VAL
1	C	81	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	137	HIS
1	C	92	ASN
1	H	92	ASN
1	C	37	GLN
1	C	136	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](#)

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 [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<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	149/158 (94%)	-0.08	2 (1%) 77 76	23, 31, 53, 61	0
1	B	144/158 (91%)	-0.02	6 (4%) 36 35	22, 30, 56, 71	0
1	C	144/158 (91%)	-0.06	5 (3%) 44 43	24, 32, 54, 70	0
1	D	148/158 (93%)	-0.05	3 (2%) 65 63	24, 31, 52, 58	0
1	E	135/158 (85%)	-0.17	4 (2%) 50 49	23, 30, 46, 56	0
1	F	148/158 (93%)	-0.07	2 (1%) 75 74	22, 30, 47, 58	0
1	G	146/158 (92%)	0.02	8 (5%) 25 24	24, 32, 63, 75	0
1	H	152/158 (96%)	-0.03	2 (1%) 77 76	22, 30, 50, 63	0
All	All	1166/1264 (92%)	-0.06	32 (2%) 54 53	22, 31, 53, 75	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	105	VAL	4.8
1	C	105	VAL	4.6
1	D	110	PHE	4.3
1	G	101	TYR	4.2
1	G	103	GLU	4.0

### 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 carbohydrates 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.