



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 9, 2017 – 12:34 PM EDT

PDB ID : 2H56  
Title : Crystal structure of DNA-3-methyladenine glycosidase (10174367) from *Bacillus halodurans* at 2.55 Å resolution  
Authors : Joint Center for Structural Genomics (JCSG)  
Deposited on : unknown  
Resolution : 2.55 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

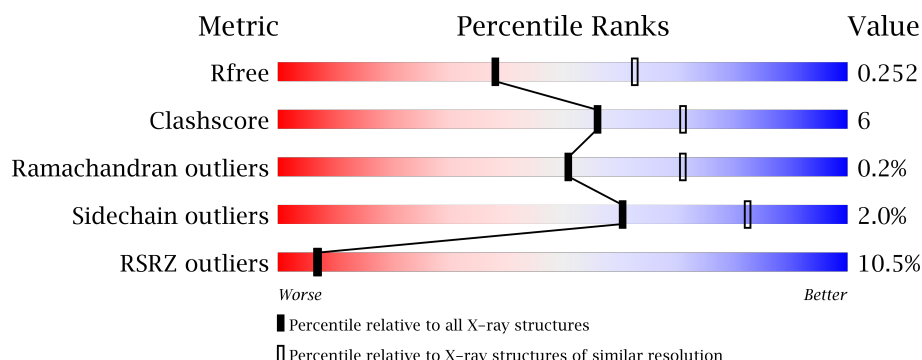
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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))
$R_{free}$	100719	4993 (2.58-2.50)
Clashscore	112137	5755 (2.58-2.50)
Ramachandran outliers	110173	5652 (2.58-2.50)
Sidechain outliers	110143	5654 (2.58-2.50)
RSRZ outliers	101464	5026 (2.58-2.50)

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 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	233	<div> <div>4%</div> <div>83%</div> <div>10%</div> <div>6%</div> </div>
1	B	233	<div> <div>5%</div> <div>80%</div> <div>12%</div> <div>7%</div> </div>
1	C	233	<div> <div>19%</div> <div>76%</div> <div>11%</div> <div>12%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4855 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 DNA-3-methyladenine glycosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	218	Total	C	N	O	S	Se	0	0	0
			1650	1062	281	301	2	4			
1	B	217	Total	C	N	O	S	Se	0	0	0
			1653	1067	282	298	2	4			
1	C	204	Total	C	N	O	S	Se	0	0	0
			1512	968	262	276	2	4			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	LEADER SEQUENCE	UNP Q9KC25
A	-10	GLY	-	LEADER SEQUENCE	UNP Q9KC25
A	-9	SER	-	LEADER SEQUENCE	UNP Q9KC25
A	-8	ASP	-	LEADER SEQUENCE	UNP Q9KC25
A	-7	LYS	-	LEADER SEQUENCE	UNP Q9KC25
A	-6	ILE	-	LEADER SEQUENCE	UNP Q9KC25
A	-5	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-4	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-3	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-2	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-1	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	0	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
A	141	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
A	143	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
A	144	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	-11	MSE	-	LEADER SEQUENCE	UNP Q9KC25
B	-10	GLY	-	LEADER SEQUENCE	UNP Q9KC25
B	-9	SER	-	LEADER SEQUENCE	UNP Q9KC25
B	-8	ASP	-	LEADER SEQUENCE	UNP Q9KC25
B	-7	LYS	-	LEADER SEQUENCE	UNP Q9KC25
B	-6	ILE	-	LEADER SEQUENCE	UNP Q9KC25
B	-5	HIS	-	LEADER SEQUENCE	UNP Q9KC25

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	-3	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	-2	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	-1	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	0	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	141	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	143	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	144	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	-11	MSE	-	LEADER SEQUENCE	UNP Q9KC25
C	-10	GLY	-	LEADER SEQUENCE	UNP Q9KC25
C	-9	SER	-	LEADER SEQUENCE	UNP Q9KC25
C	-8	ASP	-	LEADER SEQUENCE	UNP Q9KC25
C	-7	LYS	-	LEADER SEQUENCE	UNP Q9KC25
C	-6	ILE	-	LEADER SEQUENCE	UNP Q9KC25
C	-5	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-4	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-3	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-2	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-1	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	0	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	141	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	143	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	144	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25

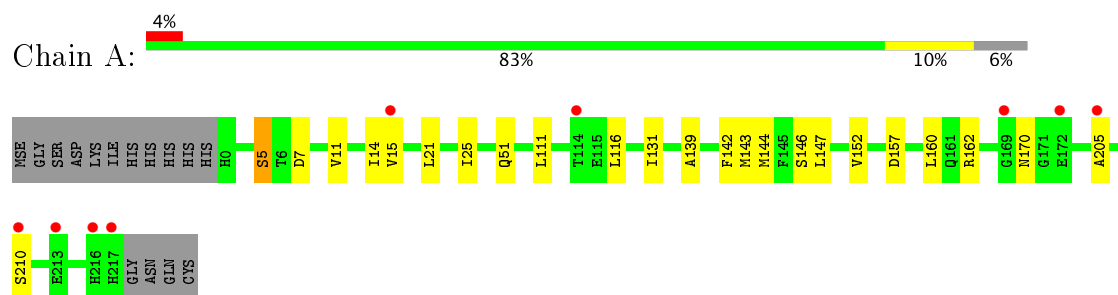
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	19	Total O 19 19	0	0
2	B	13	Total O 13 13	0	0
2	C	8	Total O 8 8	0	0

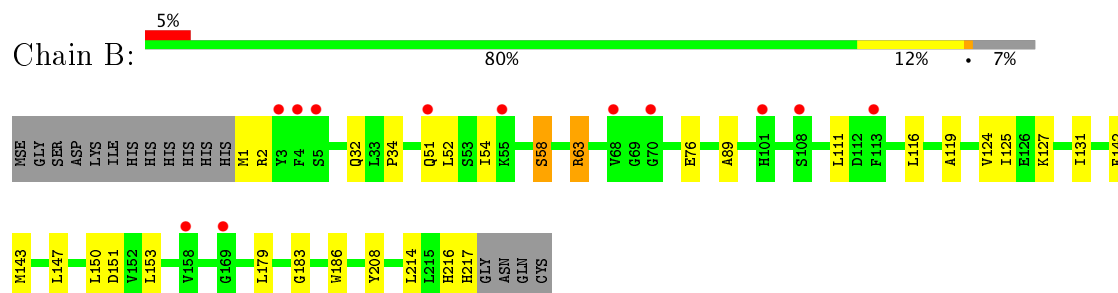
### 3 Residue-property plots [i](#)

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.

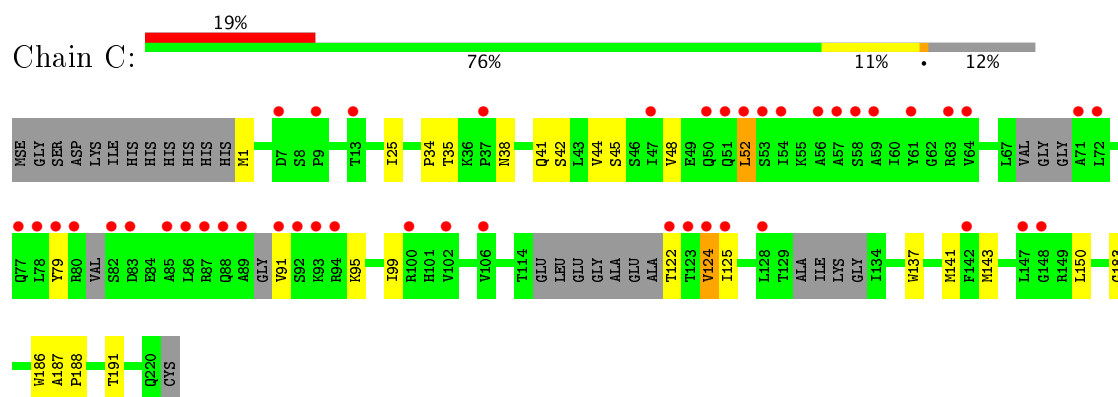
#### • Molecule 1: DNA-3-methyladenine glycosidase



#### • Molecule 1: DNA-3-methyladenine glycosidase



#### • Molecule 1: DNA-3-methyladenine glycosidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	141.75Å 141.75Å 85.95Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.40 – 2.55 46.40 – 2.55	Depositor EDS
% Data completeness (in resolution range)	98.0 (46.40-2.55) 98.0 (46.40-2.55)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.92 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.196 , 0.246 0.203 , 0.252	Depositor DCC
$R_{free}$ test set	1604 reflections (5.35%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.1	Xtriage
Anisotropy	0.289	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 69.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.044 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4855	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.49% 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	0.71	0/1679	0.72	0/2268
1	B	0.68	0/1685	0.69	1/2277 (0.0%)
1	C	0.66	0/1537	0.65	0/2073
All	All	0.69	0/4901	0.69	1/6618 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	151	ASP	CB-CG-OD1	5.29	123.06	118.30

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	1650	0	1612	17	0
1	B	1653	0	1609	23	0
1	C	1512	0	1394	16	0
2	A	19	0	0	0	0
2	B	13	0	0	0	0
2	C	8	0	0	0	0
All	All	4855	0	4615	55	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 6.

All (55) 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:125:ILE:CG1	1:B:143:MSE:HE1	1.49	1.40
1:A:142:PHE:CE2	1:A:147:LEU:HD11	2.00	0.96
1:B:125:ILE:CG1	1:B:143:MSE:CE	2.42	0.96
1:B:1:MSE:HE3	1:B:34:PRO:N	2.07	0.69
1:C:52:LEU:HD22	1:C:95:LYS:NZ	2.08	0.69
1:B:111:LEU:HD22	1:B:131:ILE:HD11	1.75	0.67
1:B:214:LEU:O	1:B:214:LEU:HD12	1.97	0.64
1:B:153:LEU:HD11	1:B:179:LEU:HD22	1.79	0.64
1:A:142:PHE:CE2	1:A:147:LEU:CD1	2.78	0.62
1:B:119:ALA:HB3	1:B:124:VAL:HG23	1.82	0.61
1:C:95:LYS:O	1:C:99:ILE:HD12	2.01	0.60
1:A:51:GLN:HE21	1:B:216:HIS:CD2	2.24	0.56
1:C:52:LEU:HD22	1:C:95:LYS:HZ1	1.71	0.55
1:A:170:ASN:OD1	1:A:210:SER:CB	2.56	0.52
1:B:2:ARG:NH1	1:B:150:LEU:HD11	2.24	0.52
1:C:44:VAL:O	1:C:48:VAL:HG23	2.09	0.52
1:A:142:PHE:HE2	1:A:147:LEU:HD11	1.71	0.51
1:B:54:ILE:O	1:B:58:SER:OG	2.28	0.51
1:C:183:GLY:HA2	1:C:186:TRP:CE3	2.47	0.50
1:B:76:GLU:N	1:B:76:GLU:OE1	2.36	0.50
1:C:52:LEU:HD22	1:C:95:LYS:HZ2	1.75	0.50
1:A:5:SER:OG	1:A:7:ASP:OD1	2.29	0.50
1:C:122:THR:N	1:C:124:VAL:HG23	2.27	0.50
1:B:214:LEU:O	1:B:217:HIS:HB2	2.12	0.49
1:A:111:LEU:HD22	1:A:131:ILE:HD11	1.94	0.49
1:C:1:MSE:HE1	1:C:34:PRO:HB3	1.94	0.49
1:C:48:VAL:HG13	1:C:91:VAL:HG22	1.94	0.49
1:A:14:ILE:HD12	1:A:21:LEU:HD21	1.94	0.48
1:C:125:ILE:CG1	1:C:143:MSE:HE1	2.44	0.48
1:C:137:TRP:CZ2	1:C:141:MSE:HE2	2.49	0.47
1:B:63:ARG:HB3	1:B:89:ALA:HB1	1.97	0.47
1:B:116:LEU:HD22	1:B:124:VAL:HA	1.96	0.47
1:A:162:ARG:NH2	1:A:205:ALA:HB1	2.29	0.46
1:B:1:MSE:HE2	1:B:32:GLN:HB2	1.98	0.46
1:A:147:LEU:HD12	1:A:147:LEU:N	2.31	0.45
1:B:142:PHE:CE2	1:B:147:LEU:HD11	2.52	0.45
1:B:111:LEU:HD12	1:B:116:LEU:HD11	1.98	0.44
1:A:111:LEU:HD12	1:A:116:LEU:HD11	2.00	0.44

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:116:LEU:HD21	1:B:127:LYS:HG3	1.99	0.44
1:A:15:VAL:CG1	1:A:25:ILE:HD12	2.48	0.44
1:C:25:ILE:O	1:C:25:ILE:HG22	2.17	0.44
1:C:48:VAL:HG13	1:C:91:VAL:CG2	2.47	0.43
1:A:143:MSE:HA	1:A:147:LEU:HD13	2.01	0.43
1:B:1:MSE:HE3	1:B:34:PRO:CA	2.50	0.42
1:C:150:LEU:HA	1:C:191:THR:HB	2.01	0.42
1:B:1:MSE:HE3	1:B:34:PRO:CD	2.48	0.42
1:B:51:GLN:O	1:B:52:LEU:HD23	2.19	0.42
1:A:11:VAL:HA	1:A:14:ILE:HG12	2.02	0.42
1:A:157:ASP:HB3	1:A:160:LEU:HD12	2.01	0.42
1:B:208:TYR:HE2	1:B:217:HIS:CD2	2.38	0.41
1:B:183:GLY:HA2	1:B:186:TRP:CE3	2.55	0.41
1:C:38:ASN:HB3	1:C:41:GLN:HB3	2.02	0.41
1:A:144:MSE:HE3	1:A:152:VAL:HG12	2.01	0.41
1:C:187:ALA:HA	1:C:188:PRO:HA	1.93	0.40
1:A:139:ALA:O	1:A:143:MSE:HG3	2.21	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	216/233 (93%)	208 (96%)	8 (4%)	0	100	100
1	B	215/233 (92%)	209 (97%)	6 (3%)	0	100	100
1	C	192/233 (82%)	180 (94%)	11 (6%)	1 (0%)	32	52
All	All	623/699 (89%)	597 (96%)	25 (4%)	1 (0%)	51	71

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	79	TYR

### 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	159/188 (85%)	157 (99%)	2 (1%)	73	89
1	B	158/188 (84%)	156 (99%)	2 (1%)	73	89
1	C	138/188 (73%)	133 (96%)	5 (4%)	40	64
All	All	455/564 (81%)	446 (98%)	9 (2%)	60	82

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	5	SER
1	A	146	SER
1	B	58	SER
1	B	63	ARG
1	C	35	THR
1	C	42	SER
1	C	45	SER
1	C	52	LEU
1	C	124	VAL

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	51	GLN
1	B	216	HIS
1	B	217	HIS

### 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

### 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	214/233 (91%)	0.69	9 (4%) 37 40	61, 69, 81, 91	0
1	B	213/233 (91%)	0.84	12 (5%) 25 27	54, 70, 82, 91	0
1	C	200/233 (85%)	1.42	45 (22%) 1 0	53, 71, 84, 91	0
All	All	627/699 (89%)	0.98	66 (10%) 7 7	53, 70, 83, 91	0

All (66) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	53	SER	9.5
1	C	52	LEU	9.2
1	C	89	ALA	7.2
1	C	123	THR	7.2
1	C	80	ARG	6.8
1	C	64	VAL	6.5
1	C	57	ALA	6.1
1	C	91	VAL	5.6
1	A	213	GLU	5.5
1	C	88	GLN	5.5
1	C	51	GLN	5.5
1	C	87	ARG	5.0
1	C	86	LEU	4.9
1	C	78	LEU	4.9
1	C	59	ALA	4.8
1	C	58	SER	4.6
1	C	125	ILE	4.6
1	B	3	TYR	4.4
1	C	83	ASP	4.2
1	C	122	THR	4.2
1	C	106	VAL	4.0
1	C	54	ILE	3.6
1	C	82	SER	3.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	71	ALA	3.6
1	C	47	ILE	3.5
1	C	13	THR	3.5
1	C	63	ARG	3.4
1	C	148	GLY	3.3
1	C	92	SER	3.3
1	C	94	ARG	3.3
1	C	61	TYR	3.3
1	C	50	GLN	3.2
1	C	142	PHE	3.1
1	A	217	HIS	3.1
1	C	7	ASP	3.0
1	C	79	TYR	3.0
1	A	205	ALA	2.9
1	C	72	LEU	2.9
1	C	77	GLN	2.8
1	B	70	GLY	2.8
1	C	85	ALA	2.8
1	B	5	SER	2.8
1	C	100	ARG	2.8
1	C	124	VAL	2.7
1	A	210	SER	2.6
1	A	172	GLU	2.6
1	A	216	HIS	2.6
1	C	147	LEU	2.5
1	C	128	LEU	2.5
1	B	51	GLN	2.4
1	C	102	VAL	2.4
1	C	9	PRO	2.4
1	B	108	SER	2.3
1	A	169	GLY	2.3
1	B	169	GLY	2.3
1	C	37	PRO	2.3
1	B	158	VAL	2.3
1	B	55	LYS	2.3
1	C	93	LYS	2.2
1	A	15	VAL	2.2
1	A	114	THR	2.2
1	B	113	PHE	2.2
1	B	68	VAL	2.2
1	C	56	ALA	2.0
1	B	101	HIS	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	4	PHE	2.0

## 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 [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.