



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 30, 2021 – 07:11 PM EST

PDB ID : 3LX6  
Title : Crystal structure of putative dna cytosine methylase from shigella flexneri 2a str. 2457T  
Authors : Ramagopal, U.A.; Toro, R.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2010-02-24  
Resolution : 2.29 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

<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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtrriage (Phenix) : 1.13  
EDS : 2.16  
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) : 2.16

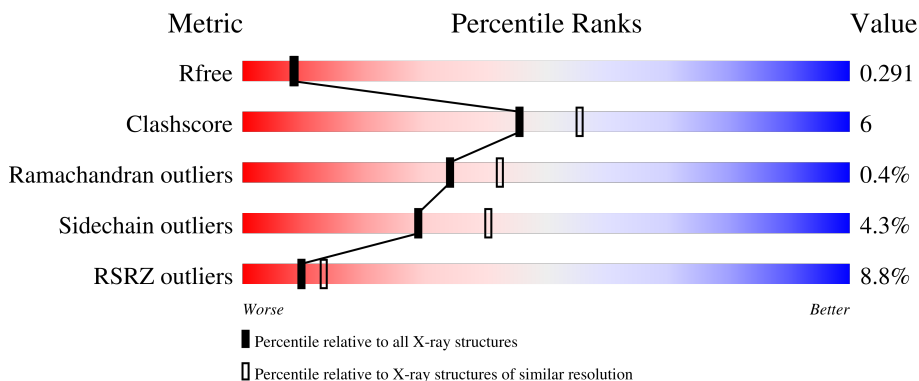
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.29 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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  $\geq 3$ , 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	410	
1	B	410	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 5835 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 Cytosine-specific methyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	368	2964	1889	531	532	6	6	0	0	0
1	B	362	2842	1800	514	516	6	6	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	MSE	-	expression tag	UNP E3Y0J2
A	62	SER	-	expression tag	UNP E3Y0J2
A	463	GLU	-	expression tag	UNP E3Y0J2
A	464	GLY	-	expression tag	UNP E3Y0J2
A	465	HIS	-	expression tag	UNP E3Y0J2
A	466	HIS	-	expression tag	UNP E3Y0J2
A	467	HIS	-	expression tag	UNP E3Y0J2
A	468	HIS	-	expression tag	UNP E3Y0J2
A	469	HIS	-	expression tag	UNP E3Y0J2
A	470	HIS	-	expression tag	UNP E3Y0J2
B	61	MSE	-	expression tag	UNP E3Y0J2
B	62	SER	-	expression tag	UNP E3Y0J2
B	463	GLU	-	expression tag	UNP E3Y0J2
B	464	GLY	-	expression tag	UNP E3Y0J2
B	465	HIS	-	expression tag	UNP E3Y0J2
B	466	HIS	-	expression tag	UNP E3Y0J2
B	467	HIS	-	expression tag	UNP E3Y0J2
B	468	HIS	-	expression tag	UNP E3Y0J2
B	469	HIS	-	expression tag	UNP E3Y0J2
B	470	HIS	-	expression tag	UNP E3Y0J2

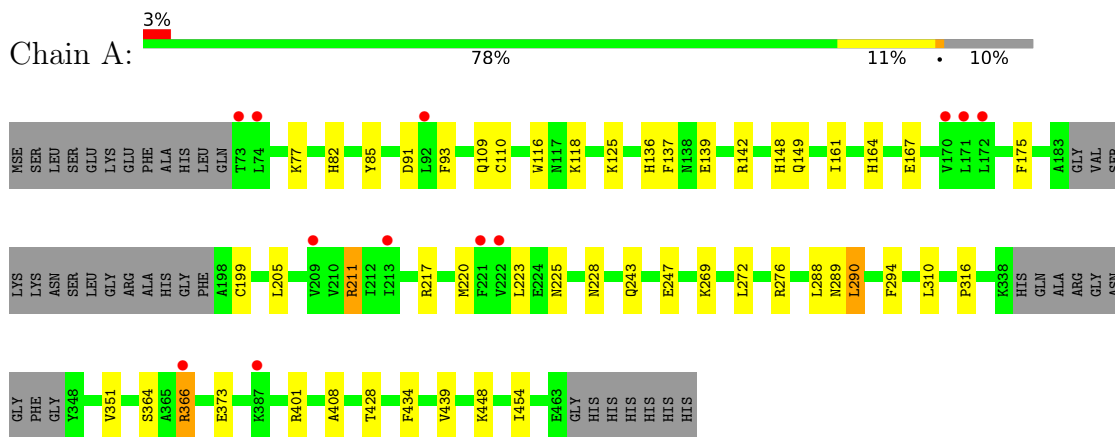
- Molecule 2 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	A	24	Total 24	O 24	0	0
2	B	5	Total 5	O 5	0	0

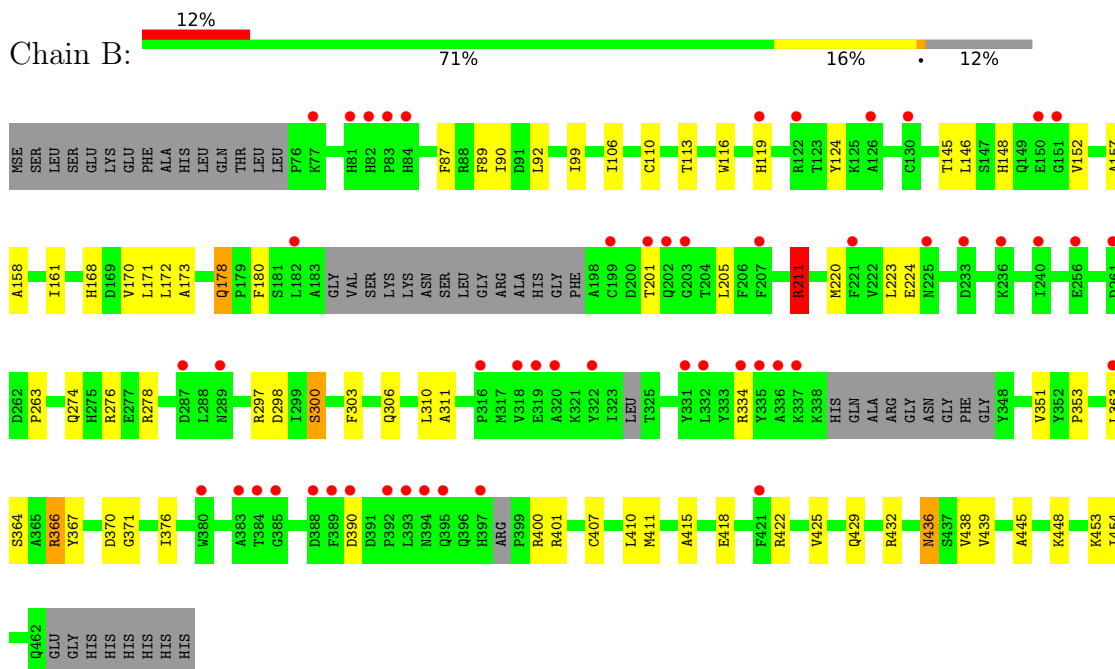
### 3 Residue-property plots [i](#)

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: Cytosine-specific methyltransferase



- Molecule 1: Cytosine-specific methyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	122.96Å 83.81Å 113.24Å 90.00° 118.51° 90.00°	Depositor
Resolution (Å)	31.26 – 2.29 31.26 – 2.29	Depositor EDS
% Data completeness (in resolution range)	96.5 (31.26-2.29) 96.5 (31.26-2.29)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.24 (at 2.29Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.243 , 0.288 0.246 , 0.291	Depositor DCC
$R_{free}$ test set	2214 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.5	Xtrriage
Anisotropy	0.355	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 47.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	5835	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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	0.76	1/3034 (0.0%)	0.77	1/4100 (0.0%)
1	B	0.62	0/2903	0.68	1/3919 (0.0%)
All	All	0.70	1/5937 (0.0%)	0.73	2/8019 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	110	CYS	CB-SG	-5.04	1.73	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	91	ASP	CB-CG-OD1	-5.14	113.67	118.30
1	B	211	ARG	NE-CZ-NH1	5.01	122.80	120.30

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	2964	0	2911	26	0
1	B	2842	0	2733	50	0
2	A	24	0	0	0	0
2	B	5	0	0	0	0
All	All	5835	0	5644	73	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 (73) 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:178:GLN:HE22	1:B:205:LEU:H	0.96	0.91
1:A:205:LEU:HD21	1:B:306:GLN:NE2	1.91	0.86
1:B:415:ALA:HB3	1:B:418:GLU:HG3	1.59	0.84
1:B:90:ILE:HD11	1:B:171:LEU:HD12	1.65	0.77
1:B:274:GLN:HB2	1:B:410:LEU:HD11	1.68	0.75
1:B:180:PHE:CD1	1:B:201:THR:HA	2.22	0.74
1:B:178:GLN:HE22	1:B:205:LEU:N	1.80	0.74
1:A:310:LEU:CD2	1:A:351:VAL:HG11	2.22	0.70
1:B:178:GLN:NE2	1:B:205:LEU:H	1.82	0.67
1:A:243:GLN:O	1:A:247:GLU:HG3	1.99	0.62
1:A:310:LEU:HD22	1:A:351:VAL:HG11	1.79	0.62
1:A:82:HIS:HD2	1:A:85:TYR:OH	1.82	0.61
1:B:300:SER:O	1:B:303:PHE:HB2	2.01	0.61
1:A:428:THR:HG21	1:B:298:ASP:OD2	2.00	0.61
1:B:371:GLY:HA3	1:B:401:ARG:HD3	1.83	0.60
1:B:310:LEU:CD2	1:B:351:VAL:HG11	2.32	0.60
1:B:276:ARG:HD3	1:B:439:VAL:CG2	2.32	0.60
1:A:220:MSE:HE1	1:A:454:ILE:HA	1.85	0.58
1:B:146:LEU:CD1	1:B:152:VAL:HG11	2.34	0.58
1:B:445:ALA:HA	1:B:448:LYS:HE2	1.86	0.57
1:B:263:PRO:HB3	1:B:297:ARG:HG3	1.87	0.57
1:B:311:ALA:HB2	1:B:353:PRO:O	2.05	0.56
1:B:220:MSE:HE1	1:B:454:ILE:HA	1.87	0.56
1:B:90:ILE:HD11	1:B:171:LEU:CD1	2.35	0.55
1:A:408:ALA:HB2	1:A:434:PHE:CZ	2.42	0.54
1:B:376:ILE:HB	1:B:400:ARG:HG2	1.90	0.54
1:B:99:ILE:HG22	1:B:172:LEU:HD13	1.90	0.54
1:A:276:ARG:HD3	1:A:439:VAL:HG23	1.91	0.53
1:B:363:LEU:HD11	1:B:407:CYS:SG	2.48	0.53
1:B:276:ARG:HD3	1:B:439:VAL:HG23	1.91	0.53
1:B:276:ARG:NE	1:B:411:MSE:HE1	2.25	0.52
1:A:93:PHE:CD1	1:A:116:TRP:HZ3	2.29	0.51
1:A:93:PHE:CZ	1:A:205:LEU:HD22	2.46	0.50
1:A:136:HIS:NE2	1:A:164:HIS:O	2.38	0.50
1:B:146:LEU:HD13	1:B:152:VAL:HG11	1.94	0.49
1:A:161:ILE:HD11	1:A:211:ARG:HG2	1.94	0.49
1:B:87:PHE:HE2	1:B:106:ILE:HG13	1.77	0.49

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:367:TYR:OH	1:B:401:ARG:HG2	2.13	0.48
1:A:93:PHE:HD1	1:A:116:TRP:HZ3	1.61	0.48
1:B:119:HIS:N	1:B:119:HIS:CD2	2.82	0.48
1:A:93:PHE:CE1	1:A:205:LEU:HD22	2.49	0.48
1:B:180:PHE:HD1	1:B:201:THR:HA	1.74	0.48
1:B:145:THR:O	1:B:211:ARG:NH2	2.39	0.47
1:B:90:ILE:HD13	1:B:168:HIS:CE1	2.50	0.47
1:A:364:SER:CB	1:A:366:ARG:HE	2.28	0.46
1:B:89:PHE:CZ	1:B:110:CYS:HB2	2.50	0.46
1:B:92:LEU:HB2	1:B:173:ALA:HB2	1.98	0.46
1:B:310:LEU:HD22	1:B:351:VAL:HG11	1.98	0.46
1:A:175:PHE:HE1	1:A:223:LEU:HB3	1.83	0.44
1:B:173:ALA:O	1:B:223:LEU:HA	2.17	0.44
1:B:425:VAL:HB	1:B:429:GLN:HB2	1.99	0.44
1:B:432:ARG:O	1:B:436:ASN:HB2	2.17	0.44
1:A:142:ARG:HD3	1:B:306:GLN:OE1	2.18	0.44
1:A:85:TYR:CG	1:A:109:GLN:HB2	2.54	0.43
1:B:146:LEU:HD12	1:B:152:VAL:HG11	1.99	0.43
1:B:170:VAL:HA	1:B:220:MSE:O	2.19	0.43
1:B:367:TYR:CD1	1:B:371:GLY:HA2	2.54	0.43
1:B:113:THR:HG21	1:B:124:TYR:CZ	2.54	0.42
1:B:274:GLN:HB2	1:B:410:LEU:CD1	2.44	0.42
1:B:158:ALA:HA	1:B:161:ILE:HD12	2.02	0.42
1:A:118:LYS:HE3	1:A:139:GLU:OE2	2.20	0.41
1:A:276:ARG:HD3	1:A:439:VAL:CG2	2.49	0.41
1:A:294:PHE:C	1:A:294:PHE:CD2	2.93	0.41
1:B:92:LEU:HB2	1:B:173:ALA:CB	2.50	0.41
1:A:125:LYS:HE2	1:A:137:PHE:CE2	2.56	0.41
1:B:310:LEU:HD23	1:B:351:VAL:HG11	2.02	0.41
1:A:269:LYS:HA	1:A:272:LEU:O	2.21	0.41
1:B:274:GLN:HB3	1:B:274:GLN:HE21	1.67	0.40
1:B:364:SER:HB2	1:B:366:ARG:HH21	1.86	0.40
1:B:99:ILE:HG13	1:B:438:VAL:HG21	2.03	0.40
1:A:288:LEU:HB3	1:A:290:LEU:HD12	2.03	0.40
1:B:146:LEU:HD12	1:B:157:ALA:HA	2.03	0.40
1:A:401:ARG:HH21	1:A:401:ARG:HD3	1.75	0.40

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	362/410 (88%)	353 (98%)	8 (2%)	1 (0%)	41	50
1	B	352/410 (86%)	334 (95%)	16 (4%)	2 (1%)	25	31
All	All	714/820 (87%)	687 (96%)	24 (3%)	3 (0%)	34	42

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	300	SER
1	B	370	ASP
1	A	316	PRO

### 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	313/340 (92%)	299 (96%)	14 (4%)	27	39
1	B	290/340 (85%)	278 (96%)	12 (4%)	30	43
All	All	603/680 (89%)	577 (96%)	26 (4%)	29	40

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	77	LYS
1	A	148	HIS
1	A	149	GLN

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>
1	A	167	GLU
1	A	199	CYS
1	A	211	ARG
1	A	217	ARG
1	A	225	ASN
1	A	228	ASN
1	A	289	ASN
1	A	290	LEU
1	A	366	ARG
1	A	373	GLU
1	A	448	LYS
1	B	116	TRP
1	B	148	HIS
1	B	178	GLN
1	B	211	ARG
1	B	224	GLU
1	B	278	ARG
1	B	334	ARG
1	B	366	ARG
1	B	390	ASP
1	B	422	ARG
1	B	436	ASN
1	B	453	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>
1	A	82	HIS
1	A	228	ASN
1	A	258	ASN
1	A	354	ASN
1	A	436	ASN
1	B	82	HIS
1	B	178	GLN
1	B	228	ASN
1	B	258	ASN
1	B	357	GLN
1	B	433	GLN
1	B	456	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 monosaccharides 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	362/410 (88%)	0.12	12 (3%) 46 53	19, 30, 43, 77	0
1	B	356/410 (86%)	0.85	51 (14%) 2 3	19, 33, 48, 80	0
All	All	718/820 (87%)	0.49	63 (8%) 10 13	19, 32, 46, 80	0

All (63) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	335	TYR	7.3
1	B	182	LEU	6.7
1	B	199	CYS	6.1
1	B	331	TYR	6.1
1	B	383	ALA	5.8
1	B	316	PRO	4.9
1	B	336	ALA	4.8
1	B	332	LEU	4.7
1	B	319	GLU	4.5
1	B	421	PHE	4.0
1	B	83	PRO	3.9
1	B	390	ASP	3.9
1	B	203	GLY	3.8
1	B	389	PHE	3.6
1	A	387	LYS	3.4
1	B	318	VAL	3.3
1	B	150	GLU	3.3
1	B	77	LYS	3.3
1	B	261	ASP	3.0
1	B	207	PHE	3.0
1	B	82	HIS	3.0
1	A	74	LEU	3.0
1	B	240	ILE	2.9
1	B	130	CYS	2.9

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	B	202	GLN	2.8
1	A	170	VAL	2.7
1	B	84	HIS	2.7
1	B	119	HIS	2.7
1	A	92	LEU	2.7
1	A	171	LEU	2.7
1	A	73	THR	2.6
1	B	322	TYR	2.6
1	B	320	ALA	2.6
1	B	337	LYS	2.5
1	B	397	HIS	2.5
1	B	151	GLY	2.5
1	B	81	HIS	2.5
1	B	287	ASP	2.5
1	A	221	PHE	2.4
1	B	394	ASN	2.4
1	B	363	LEU	2.4
1	B	126	ALA	2.4
1	B	385	GLY	2.4
1	B	395	GLN	2.4
1	B	380	TRP	2.4
1	A	222	VAL	2.4
1	A	366	ARG	2.3
1	B	221	PHE	2.3
1	A	172	LEU	2.3
1	B	201	THR	2.3
1	A	213	ILE	2.3
1	B	393	LEU	2.2
1	B	256	GLU	2.2
1	B	334	ARG	2.2
1	B	233	ASP	2.2
1	B	236	LYS	2.1
1	B	122	ARG	2.1
1	B	225	ASN	2.1
1	B	384	THR	2.1
1	B	289	ASN	2.1
1	B	388	ASP	2.1
1	A	209	VAL	2.1
1	B	392	PRO	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 monosaccharides 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.