



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 05:59 pm GMT

PDB ID : 4G6V  
Title : CdiA-CT/CdiI toxin and immunity complex from Burkholderia pseudomallei  
Authors : Morse, R.P.; Nikolakakis, K.; Willet, J.; Gerrick, E.; Low, D.A.; Hayes, C.S.;  
Goulding, C.W.  
Deposited on : 2012-07-19  
Resolution : 2.64 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

<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	:	trunk28620
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)	:	recalc28949

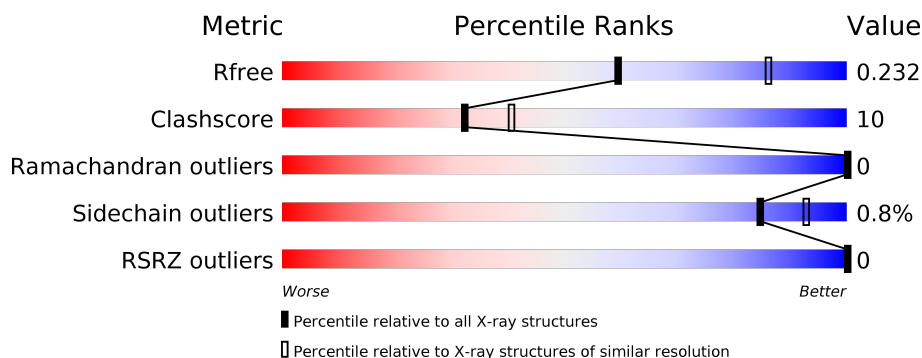
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.64 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	1044 (2.66-2.62)
Clashscore	112137	1092 (2.66-2.62)
Ramachandran outliers	110173	1077 (2.66-2.62)
Sidechain outliers	110143	1077 (2.66-2.62)
RSRZ outliers	101464	1047 (2.66-2.62)

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	176	<div> <div>65%</div> <div>10%</div> <div>25%</div> </div>
1	C	176	<div> <div>67%</div> <div>8%</div> <div>25%</div> </div>
1	E	176	<div> <div>58%</div> <div>17%</div> <div>25%</div> </div>
1	G	176	<div> <div>56%</div> <div>19%</div> <div>25%</div> </div>
2	B	111	<div> <div>76%</div> <div>16%</div> <div>7%</div> </div>
2	D	111	<div> <div>77%</div> <div>14%</div> <div>7%</div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	111	 75% 15% • 9%
2	H	111	 72% 18% • 9%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6969 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 Adhesin/hemolysin.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	132	Total	C	N	O	0	0	0
			949	590	164	195			
1	C	132	Total	C	N	O	0	0	0
			949	590	164	195			
1	E	132	Total	C	N	O	0	0	0
			949	590	164	195			
1	G	132	Total	C	N	O	0	0	0
			949	590	164	195			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	122	MET	-	EXPRESSION TAG	UNP I2KQ03
C	122	MET	-	EXPRESSION TAG	UNP I2KQ03
E	122	MET	-	EXPRESSION TAG	UNP I2KQ03
G	122	MET	-	EXPRESSION TAG	UNP I2KQ03

- Molecule 2 is a protein called CdiI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	103	Total	C	N	O	S	0	0	0
			785	501	132	151	1			
2	D	103	Total	C	N	O	S	0	0	0
			785	501	132	151	1			
2	F	101	Total	C	N	O	S	0	0	0
			781	499	131	150	1			
2	H	101	Total	C	N	O	S	0	0	0
			781	499	131	150	1			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	102	THR	-	EXPRESSION TAG	UNP H9T8H3
B	103	SER	-	EXPRESSION TAG	UNP H9T8H3
B	104	LEU	-	EXPRESSION TAG	UNP H9T8H3
B	105	GLU	-	EXPRESSION TAG	UNP H9T8H3
B	106	HIS	-	EXPRESSION TAG	UNP H9T8H3
B	107	HIS	-	EXPRESSION TAG	UNP H9T8H3
B	108	HIS	-	EXPRESSION TAG	UNP H9T8H3
B	109	HIS	-	EXPRESSION TAG	UNP H9T8H3
B	110	HIS	-	EXPRESSION TAG	UNP H9T8H3
B	111	HIS	-	EXPRESSION TAG	UNP H9T8H3
D	102	THR	-	EXPRESSION TAG	UNP H9T8H3
D	103	SER	-	EXPRESSION TAG	UNP H9T8H3
D	104	LEU	-	EXPRESSION TAG	UNP H9T8H3
D	105	GLU	-	EXPRESSION TAG	UNP H9T8H3
D	106	HIS	-	EXPRESSION TAG	UNP H9T8H3
D	107	HIS	-	EXPRESSION TAG	UNP H9T8H3
D	108	HIS	-	EXPRESSION TAG	UNP H9T8H3
D	109	HIS	-	EXPRESSION TAG	UNP H9T8H3
D	110	HIS	-	EXPRESSION TAG	UNP H9T8H3
D	111	HIS	-	EXPRESSION TAG	UNP H9T8H3
F	102	THR	-	EXPRESSION TAG	UNP H9T8H3
F	103	SER	-	EXPRESSION TAG	UNP H9T8H3
F	104	LEU	-	EXPRESSION TAG	UNP H9T8H3
F	105	GLU	-	EXPRESSION TAG	UNP H9T8H3
F	106	HIS	-	EXPRESSION TAG	UNP H9T8H3
F	107	HIS	-	EXPRESSION TAG	UNP H9T8H3
F	108	HIS	-	EXPRESSION TAG	UNP H9T8H3
F	109	HIS	-	EXPRESSION TAG	UNP H9T8H3
F	110	HIS	-	EXPRESSION TAG	UNP H9T8H3
F	111	HIS	-	EXPRESSION TAG	UNP H9T8H3
H	102	THR	-	EXPRESSION TAG	UNP H9T8H3
H	103	SER	-	EXPRESSION TAG	UNP H9T8H3
H	104	LEU	-	EXPRESSION TAG	UNP H9T8H3
H	105	GLU	-	EXPRESSION TAG	UNP H9T8H3
H	106	HIS	-	EXPRESSION TAG	UNP H9T8H3
H	107	HIS	-	EXPRESSION TAG	UNP H9T8H3
H	108	HIS	-	EXPRESSION TAG	UNP H9T8H3
H	109	HIS	-	EXPRESSION TAG	UNP H9T8H3
H	110	HIS	-	EXPRESSION TAG	UNP H9T8H3
H	111	HIS	-	EXPRESSION TAG	UNP H9T8H3

- Molecule 3 is BROMIDE ION (three-letter code: BR) (formula: Br).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	G	1	Total Br 1 1	0	0
3	E	1	Total Br 1 1	0	0
3	H	1	Total Br 1 1	0	0
3	C	2	Total Br 2 2	0	0
3	A	2	Total Br 2 2	0	0
3	F	1	Total Br 1 1	0	0

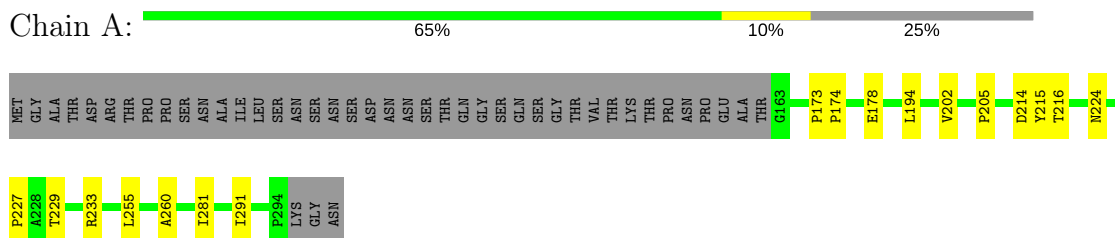
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	6	Total O 6 6	0	0
4	B	4	Total O 4 4	0	0
4	C	5	Total O 5 5	0	0
4	D	3	Total O 3 3	0	0
4	E	5	Total O 5 5	0	0
4	F	3	Total O 3 3	0	0
4	G	5	Total O 5 5	0	0
4	H	2	Total O 2 2	0	0

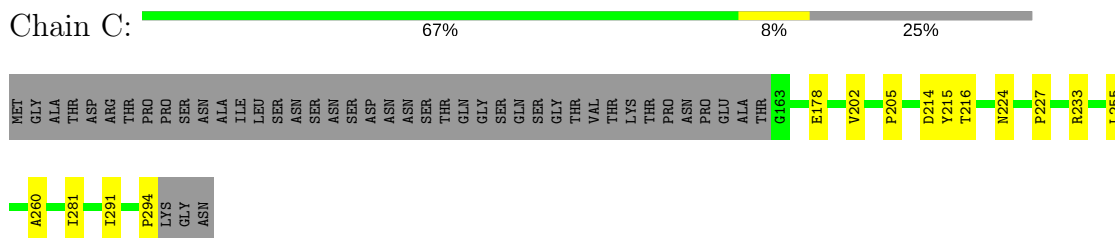
### 3 Residue-property plots [i](#)

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.

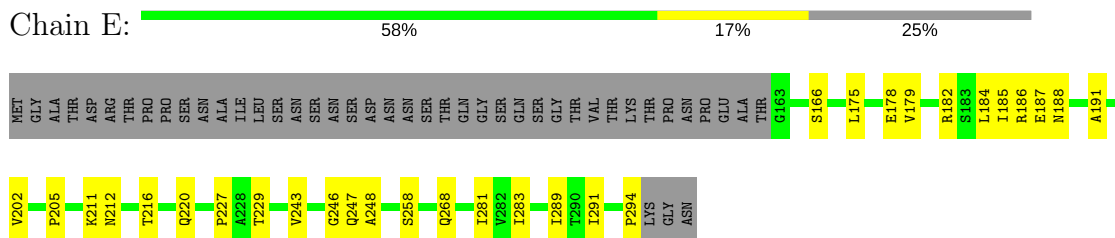
- Molecule 1: Adhesin/hemolysin



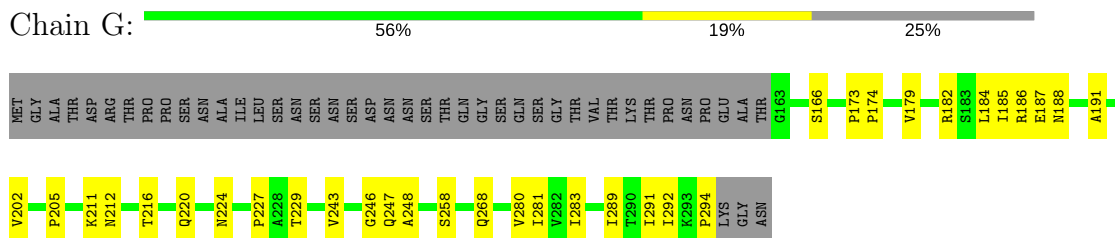
- Molecule 1: Adhesin/hemolysin




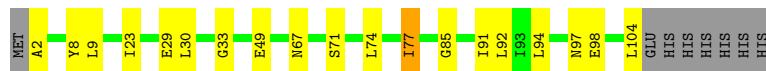
- Molecule 1: Adhesin/hemolysin




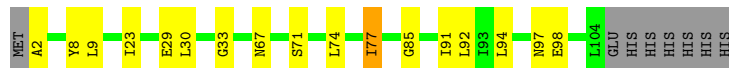
- Molecule 1: Adhesin/hemolysin




## • Molecule 2: CdiI

Chain B:  76% 16% • 7%

## • Molecule 2: CdiI

Chain D:  77% 14% • 7%

## • Molecule 2: CdiI

Chain F:  75% 15% • 9%

## • Molecule 2: CdiI

Chain H:  72% 18% • 9%



## 4 Data and refinement statistics

Property	Value	Source
Space group	F 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	151.96Å 173.65Å 174.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.40 – 2.64 46.85 – 2.64	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.40-2.64) 95.6 (46.85-2.64)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	14.93 (at 2.65Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.1_743)	Depositor
R, $R_{free}$	0.204 , 0.245 0.187 , 0.232	Depositor DCC
$R_{free}$ test set	1713 reflections (5.05%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.6	Xtriage
Anisotropy	1.472	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 9.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.399 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6969	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: BR

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	0.47	0/963	0.63	0/1322
1	C	0.46	0/963	0.63	1/1322 (0.1%)
1	E	0.53	2/963 (0.2%)	0.66	1/1322 (0.1%)
1	G	0.52	2/963 (0.2%)	0.68	1/1322 (0.1%)
2	B	0.49	0/797	0.59	0/1077
2	D	0.48	0/797	0.58	0/1077
2	F	0.48	0/793	0.61	0/1070
2	H	0.48	0/793	0.61	0/1070
All	All	0.49	4/7032 (0.1%)	0.63	3/9582 (0.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	268	GLN	CD-NE2	-5.64	1.18	1.32
1	G	268	GLN	CD-OE1	-5.40	1.12	1.24
1	E	268	GLN	CD-OE1	-5.09	1.12	1.24
1	G	268	GLN	CD-NE2	-5.01	1.20	1.32

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	294	PRO	N-CA-CB	5.77	110.22	103.30
1	E	294	PRO	N-CA-CB	5.73	110.18	103.30
1	C	294	PRO	N-CA-CB	5.02	109.32	103.30

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	949	0	936	31	0
1	C	949	0	936	27	0
1	E	949	0	936	21	0
1	G	949	0	936	22	0
2	B	785	0	769	16	0
2	D	785	0	769	13	0
2	F	781	0	775	19	0
2	H	781	0	775	22	0
3	A	2	0	0	0	0
3	C	2	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
3	H	1	0	0	0	0
4	A	6	0	0	1	0
4	B	4	0	0	0	0
4	C	5	0	0	0	0
4	D	3	0	0	0	0
4	E	5	0	0	0	0
4	F	3	0	0	0	0
4	G	5	0	0	0	0
4	H	2	0	0	0	0
All	All	6969	0	6832	138	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 10.

The worst 5 of 138 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:233:ARG:HD2	1:C:233:ARG:HH11	1.04	1.10
1:A:205:PRO:HG3	1:A:216:THR:HG23	1.35	1.07
1:C:205:PRO:HG3	1:C:216:THR:HG23	1.33	1.05
1:A:233:ARG:HH11	1:C:233:ARG:HD2	1.21	1.02
1:G:205:PRO:HG3	1:G:216:THR:HG23	1.47	0.96

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	130/176 (74%)	128 (98%)	2 (2%)	0	100	100
1	C	130/176 (74%)	128 (98%)	2 (2%)	0	100	100
1	E	130/176 (74%)	126 (97%)	4 (3%)	0	100	100
1	G	130/176 (74%)	126 (97%)	4 (3%)	0	100	100
2	B	101/111 (91%)	97 (96%)	4 (4%)	0	100	100
2	D	101/111 (91%)	96 (95%)	5 (5%)	0	100	100
2	F	99/111 (89%)	95 (96%)	4 (4%)	0	100	100
2	H	99/111 (89%)	94 (95%)	5 (5%)	0	100	100
All	All	920/1148 (80%)	890 (97%)	30 (3%)	0	100	100

There are no Ramachandran outliers to report.

#### 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	105/148 (71%)	105 (100%)	0	100	100
1	C	105/148 (71%)	105 (100%)	0	100	100
1	E	105/148 (71%)	104 (99%)	1 (1%)	80	90
1	G	105/148 (71%)	104 (99%)	1 (1%)	80	90

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	83/96 (86%)	82 (99%)	1 (1%)	75	88
2	D	83/96 (86%)	82 (99%)	1 (1%)	75	88
2	F	84/96 (88%)	83 (99%)	1 (1%)	75	88
2	H	84/96 (88%)	83 (99%)	1 (1%)	75	88
All	All	754/976 (77%)	748 (99%)	6 (1%)	85	93

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	166	SER
2	H	12	ASP
2	F	12	ASP
2	D	77	ILE
1	G	166	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	G	268	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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	132/176 (75%)	-0.24	0 100 100	27, 38, 56, 65	0
1	C	132/176 (75%)	-0.31	0 100 100	27, 39, 56, 65	0
1	E	132/176 (75%)	-0.22	0 100 100	26, 37, 54, 70	0
1	G	132/176 (75%)	-0.25	0 100 100	25, 37, 54, 71	0
2	B	103/111 (92%)	-0.29	0 100 100	24, 35, 58, 73	0
2	D	103/111 (92%)	-0.28	0 100 100	23, 35, 58, 70	0
2	F	101/111 (90%)	-0.28	0 100 100	26, 40, 57, 78	0
2	H	101/111 (90%)	-0.18	0 100 100	27, 40, 58, 79	0
All	All	936/1148 (81%)	-0.26	0 100 100	23, 38, 58, 79	0

There are no RSRZ outliers to report.

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.



Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	BR	A	301	1/1	0.98	0.07	-1.90	31,31,31,31	1
3	BR	C	302	1/1	0.96	0.11	-2.40	39,39,39,39	1
3	BR	E	301	1/1	0.97	0.08	-2.77	39,39,39,39	1
3	BR	G	301	1/1	0.99	0.04	-6.42	34,34,34,34	1
3	BR	A	302	1/1	0.92	0.06	-	97,97,97,97	1
3	BR	F	201	1/1	0.87	0.11	-	61,61,61,61	1
3	BR	H	201	1/1	0.90	0.09	-	69,69,69,69	1
3	BR	C	301	1/1	0.81	0.07	-	102,102,102,102	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.