



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 05:12 PM GMT

PDB ID : 4HJI  
Title : Structure of the CooA pilin subunit from enterotoxigenic Escherichia coli  
Authors : Kolappan, S.; Zong, Z.; Craig, L.  
Deposited on : 2012-10-12  
Resolution : 1.60 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

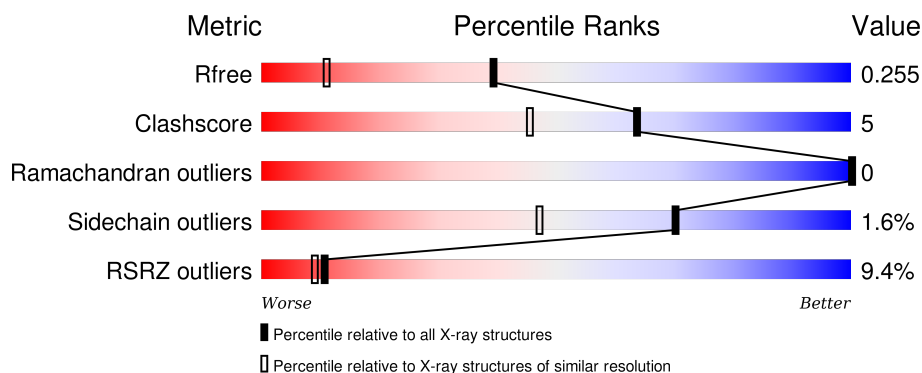
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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))
$R_{free}$	91344	2475 (1.60-1.60)
Clashscore	102246	2732 (1.60-1.60)
Ramachandran outliers	100387	2654 (1.60-1.60)
Sidechain outliers	100360	2653 (1.60-1.60)
RSRZ outliers	91569	2479 (1.60-1.60)

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	175	<div> <div>9%</div> <div>78%</div> <div>10%</div> <div>12%</div> </div>
1	B	175	<div> <div>7%</div> <div>75%</div> <div>12%</div> <div>13%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2456 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 CS1 fimbrial subunit A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	154	Total	C	N	O	0	8	0
			1140	705	193	242			
1	B	153	Total	C	N	O	0	10	0
			1154	716	194	244			

There are 76 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	MET	-	EXPRESSION TAG	UNP P0ABW7
A	-8	GLY	-	EXPRESSION TAG	UNP P0ABW7
A	-7	SER	-	EXPRESSION TAG	UNP P0ABW7
A	-6	SER	-	EXPRESSION TAG	UNP P0ABW7
A	-5	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	-4	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	-3	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	-2	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	-1	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	0	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	1	SER	-	EXPRESSION TAG	UNP P0ABW7
A	2	SER	-	EXPRESSION TAG	UNP P0ABW7
A	3	GLY	-	EXPRESSION TAG	UNP P0ABW7
A	4	LEU	-	EXPRESSION TAG	UNP P0ABW7
A	5	VAL	-	EXPRESSION TAG	UNP P0ABW7
A	6	PRO	-	EXPRESSION TAG	UNP P0ABW7
A	7	ARG	-	EXPRESSION TAG	UNP P0ABW7
A	8	GLY	-	EXPRESSION TAG	UNP P0ABW7
A	9	SER	-	EXPRESSION TAG	UNP P0ABW7
A	10	HIS	-	EXPRESSION TAG	UNP P0ABW7
A	11	MET	-	EXPRESSION TAG	UNP P0ABW7
A	149	ASP	-	EXPRESSION TAG	UNP P0ABW7
A	150	ASN	-	EXPRESSION TAG	UNP P0ABW7
A	151	LYS	-	EXPRESSION TAG	UNP P0ABW7
A	152	GLN	-	EXPRESSION TAG	UNP P0ABW7

*Continued on next page...*

*Continued from previous page...*

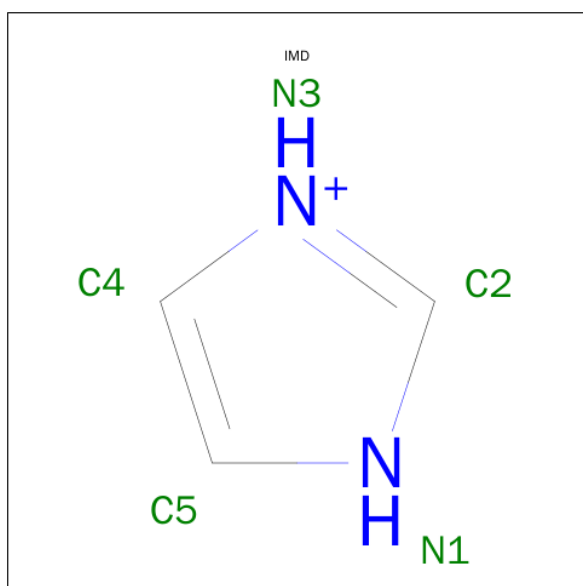
Chain	Residue	Modelled	Actual	Comment	Reference
A	153	VAL	-	EXPRESSION TAG	UNP P0ABW7
A	154	GLU	-	EXPRESSION TAG	UNP P0ABW7
A	155	LYS	-	EXPRESSION TAG	UNP P0ABW7
A	156	THR	-	EXPRESSION TAG	UNP P0ABW7
A	157	ILE	-	EXPRESSION TAG	UNP P0ABW7
A	158	SER	-	EXPRESSION TAG	UNP P0ABW7
A	159	VAL	-	EXPRESSION TAG	UNP P0ABW7
A	160	THR	-	EXPRESSION TAG	UNP P0ABW7
A	161	ALA	-	EXPRESSION TAG	UNP P0ABW7
A	162	SER	-	EXPRESSION TAG	UNP P0ABW7
A	163	VAL	-	EXPRESSION TAG	UNP P0ABW7
A	164	ASP	-	EXPRESSION TAG	UNP P0ABW7
A	165	PRO	-	EXPRESSION TAG	UNP P0ABW7
B	-9	MET	-	EXPRESSION TAG	UNP P0ABW7
B	-8	GLY	-	EXPRESSION TAG	UNP P0ABW7
B	-7	SER	-	EXPRESSION TAG	UNP P0ABW7
B	-6	SER	-	EXPRESSION TAG	UNP P0ABW7
B	-5	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	-4	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	-3	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	-2	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	-1	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	0	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	1	SER	-	EXPRESSION TAG	UNP P0ABW7
B	2	SER	-	EXPRESSION TAG	UNP P0ABW7
B	3	GLY	-	EXPRESSION TAG	UNP P0ABW7
B	4	LEU	-	EXPRESSION TAG	UNP P0ABW7
B	5	VAL	-	EXPRESSION TAG	UNP P0ABW7
B	6	PRO	-	EXPRESSION TAG	UNP P0ABW7
B	7	ARG	-	EXPRESSION TAG	UNP P0ABW7
B	8	GLY	-	EXPRESSION TAG	UNP P0ABW7
B	9	SER	-	EXPRESSION TAG	UNP P0ABW7
B	10	HIS	-	EXPRESSION TAG	UNP P0ABW7
B	11	MET	-	EXPRESSION TAG	UNP P0ABW7
B	149	ASP	-	EXPRESSION TAG	UNP P0ABW7
B	150	ASN	-	EXPRESSION TAG	UNP P0ABW7
B	151	LYS	-	EXPRESSION TAG	UNP P0ABW7
B	152	GLN	-	EXPRESSION TAG	UNP P0ABW7
B	153	VAL	-	EXPRESSION TAG	UNP P0ABW7
B	154	GLU	-	EXPRESSION TAG	UNP P0ABW7
B	155	LYS	-	EXPRESSION TAG	UNP P0ABW7
B	156	THR	-	EXPRESSION TAG	UNP P0ABW7

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	157	ILE	-	EXPRESSION TAG	UNP P0ABW7
B	158	SER	-	EXPRESSION TAG	UNP P0ABW7
B	159	VAL	-	EXPRESSION TAG	UNP P0ABW7
B	160	THR	-	EXPRESSION TAG	UNP P0ABW7
B	161	ALA	-	EXPRESSION TAG	UNP P0ABW7
B	162	SER	-	EXPRESSION TAG	UNP P0ABW7
B	163	VAL	-	EXPRESSION TAG	UNP P0ABW7
B	164	ASP	-	EXPRESSION TAG	UNP P0ABW7
B	165	PRO	-	EXPRESSION TAG	UNP P0ABW7

- Molecule 2 is IMIDAZOLE (three-letter code: IMD) (formula:  $C_3H_5N_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	N	0	0
			5	3	2		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Na	0	0
			1	1		

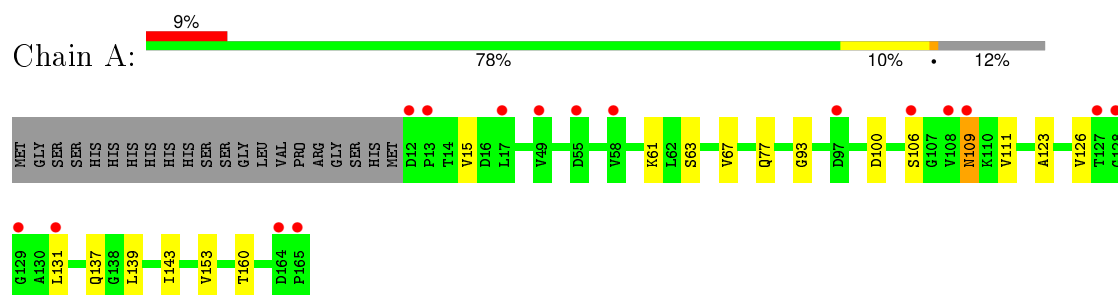
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	80	Total 80	O 80	0	0
4	B	76	Total 76	O 76	0	0

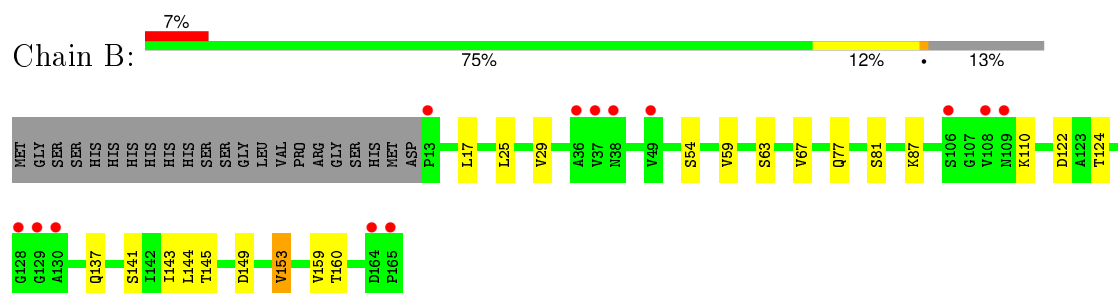
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: CS1 fimbrial subunit A



#### • Molecule 1: CS1 fimbrial subunit A



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	29.99 Å 46.87 Å 51.27 Å 83.72° 89.95° 74.58°	Depositor
Resolution (Å)	19.67 – 1.60 19.67 – 1.60	Depositor EDS
% Data completeness (in resolution range)	94.9 (19.67-1.60) 91.8 (19.67-1.60)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.72 (at 1.60 Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.222 , 0.251 0.227 , 0.255	Depositor DCC
$R_{free}$ test set	1676 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.4	Xtriage
Anisotropy	0.243	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.45 , 46.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 33502 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2456	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.94% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NA, IMD

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	0.43	0/1168	0.64	0/1598
1	B	0.41	0/1188	0.71	2/1627 (0.1%)
All	All	0.42	0/2356	0.68	2/3225 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	153[A]	VAL	CG1-CB-CG2	6.84	121.85	110.90
1	B	153[B]	VAL	CG1-CB-CG2	6.84	121.85	110.90

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	1140	0	1135	13	0
1	B	1154	0	1166	11	0
2	A	5	0	5	1	0
3	B	1	0	0	0	0
4	A	80	0	0	1	0
4	B	76	0	0	0	0
All	All	2456	0	2306	24	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 5.

All (24) 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:59:VAL:HB	1:B:145[B]:THR:HG22	1.66	0.78
1:B:63[B]:SER:OG	1:B:143:ILE:HD11	1.84	0.77
1:B:29[B]:VAL:CG2	1:B:159:VAL:HG22	2.19	0.73
1:A:67:VAL:HG11	1:A:77:GLN:HB3	1.78	0.66
1:B:137:GLN:HG2	1:B:160[A]:THR:HG22	1.86	0.56
1:A:67:VAL:CG1	1:A:77:GLN:HB3	2.37	0.54
1:A:139:LEU:HD21	4:A:372:HOH:O	2.08	0.53
1:A:100[A]:ASP:OD2	1:A:100[A]:ASP:O	2.27	0.52
1:A:106:SER:OG	1:A:111[A]:VAL:HG23	2.12	0.49
1:B:54:SER:HB2	1:B:110:LYS:HG3	1.94	0.48
1:A:61:LYS:HB3	1:A:93:GLY:HA2	1.94	0.48
1:B:145[A]:THR:HG21	1:B:149:ASP:O	2.14	0.47
1:A:106:SER:OG	1:A:111[B]:VAL:HG12	2.14	0.47
1:A:109:ASN:HB2	2:A:201:IMD:H4	1.96	0.47
1:A:126:VAL:HG21	1:A:131:LEU:HG	1.97	0.46
1:A:15:VAL:O	1:A:153:VAL:HG21	2.16	0.45
1:B:17:LEU:HD23	1:B:25:LEU:HD11	1.99	0.43
1:B:29[B]:VAL:HG22	1:B:159:VAL:HG22	1.99	0.43
1:B:144:LEU:HB2	1:B:153[B]:VAL:HG13	2.01	0.42
1:A:123:ALA:O	1:A:126:VAL:HG23	2.20	0.41
1:A:137[B]:GLN:HG2	1:A:160:THR:HG22	2.02	0.41
1:B:122:ASP:OD1	1:B:124:THR:OG1	2.38	0.41
1:A:63[B]:SER:OG	1:A:143:ILE:HD11	2.21	0.41
1:B:67[B]:VAL:CG1	1:B:77:GLN:HB3	2.51	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	159/175 (91%)	159 (100%)	0	0	100	100
1	B	161/175 (92%)	160 (99%)	1 (1%)	0	100	100
All	All	320/350 (91%)	319 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

### 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	135/151 (89%)	134 (99%)	1 (1%)	88	78
1	B	140/151 (93%)	134 (96%)	6 (4%)	35	11
All	All	275/302 (91%)	268 (98%)	7 (2%)	70	26

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	109	ASN
1	B	81[A]	SER
1	B	81[B]	SER
1	B	87[A]	LYS
1	B	87[B]	LYS
1	B	141[A]	SER
1	B	141[B]	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	19	GLN
1	A	109	ASN
1	A	115	GLN
1	A	135	GLN
1	B	115	GLN
1	B	120	HIS

### 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 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	IMD	A	201	-	3,5,5	0.32	0	4,5,5	0.57	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	IMD	A	201	-	-	0/0/0/0	0/1/1/1

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	201	IMD	1	0

## 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	154/175 (88%)	0.67	16 (10%) 8 7	13, 20, 34, 41	2 (1%)
1	B	153/175 (87%)	0.63	13 (8%) 13 11	13, 20, 34, 46	2 (1%)
All	All	307/350 (87%)	0.65	29 (9%) 11 9	13, 20, 34, 46	4 (1%)

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	36	ALA	6.0
1	A	165	PRO	5.6
1	A	128	GLY	5.3
1	A	109	ASN	5.3
1	A	12	ASP	4.2
1	B	13	PRO	3.8
1	B	165	PRO	3.7
1	A	129	GLY	3.7
1	A	131	LEU	3.6
1	B	164	ASP	3.5
1	B	49	VAL	3.4
1	B	109	ASN	3.3
1	A	164	ASP	3.2
1	A	106	SER	3.2
1	A	49	VAL	3.2
1	B	106	SER	3.1
1	B	130	ALA	2.9
1	B	37	VAL	2.9
1	B	129	GLY	2.8
1	B	108	VAL	2.8
1	B	128	GLY	2.7
1	A	17	LEU	2.7
1	A	127	THR	2.5
1	A	13	PRO	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	55	ASP	2.3
1	B	38	ASN	2.2
1	A	58	VAL	2.1
1	A	108	VAL	2.1
1	A	97	ASP	2.1

## 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(Å <sup>2</sup> )	Q<0.9
3	NA	B	201	1/1	0.97	0.12	0.55	21,21,21,21	0
2	IMD	A	201	5/5	0.79	0.18	-0.32	26,30,30,30	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.