wwPDB X-ray Structure Validation Summary Report (i)

Sep 9, 2023 – 11:09 PM EDT

PDB ID	:	4HKH
Title	:	Structure of the Hcp1 protein from E. coli EAEC 042 pathovar, mutants N93W-S158W
Authors	:	Douzi, B.; Spinelli, S.; Derrez, E.; Blangy, S.; Brunet, Y.R.; Cascales, E.; Cambillau, C.
Deposited on	:	2012-10-15
Resolution	:	1.69 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
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)	:	2.35.1

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 1.69 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.
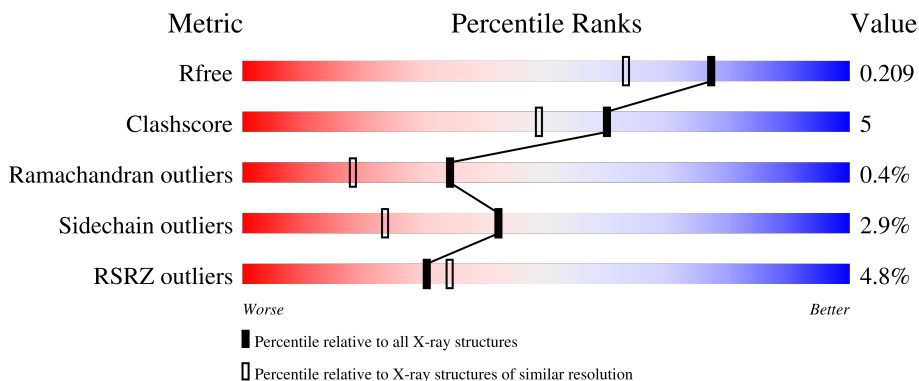
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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 >=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 <=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	169	7% / 76% / 12% · 12%
1	B	169	4% / 72% / 9% · 18%
1	D	169	4% / 71% / 13% 16%
1	E	169	4% / 69% / 12% 19%
1	F	169	4% / 72% / 12% 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	169	

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 7794 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 Putative type VI secretion protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	149	Total 1225	C 790	N 202	O 231	S 2	0	3	0
1	B	139	Total 1149	C 745	N 188	O 214	S 2	0	3	0
1	D	142	Total 1171	C 759	N 193	O 217	S 2	0	3	0
1	E	137	Total 1141	C 741	N 186	O 212	S 2	0	5	0
1	F	142	Total 1163	C 752	N 191	O 218	S 2	0	4	0
1	G	142	Total 1168	C 755	N 191	O 220	S 2	0	3	0

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	93	TRP	ASN	engineered mutation	UNP D3GUW0
A	158	TRP	SER	engineered mutation	UNP D3GUW0
A	164	HIS	-	expression tag	UNP D3GUW0
A	165	HIS	-	expression tag	UNP D3GUW0
A	166	HIS	-	expression tag	UNP D3GUW0
A	167	HIS	-	expression tag	UNP D3GUW0
A	168	HIS	-	expression tag	UNP D3GUW0
A	169	HIS	-	expression tag	UNP D3GUW0
B	93	TRP	ASN	engineered mutation	UNP D3GUW0
B	158	TRP	SER	engineered mutation	UNP D3GUW0
B	164	HIS	-	expression tag	UNP D3GUW0
B	165	HIS	-	expression tag	UNP D3GUW0
B	166	HIS	-	expression tag	UNP D3GUW0
B	167	HIS	-	expression tag	UNP D3GUW0
B	168	HIS	-	expression tag	UNP D3GUW0
B	169	HIS	-	expression tag	UNP D3GUW0
D	93	TRP	ASN	engineered mutation	UNP D3GUW0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	158	TRP	SER	engineered mutation	UNP D3GUW0
D	164	HIS	-	expression tag	UNP D3GUW0
D	165	HIS	-	expression tag	UNP D3GUW0
D	166	HIS	-	expression tag	UNP D3GUW0
D	167	HIS	-	expression tag	UNP D3GUW0
D	168	HIS	-	expression tag	UNP D3GUW0
D	169	HIS	-	expression tag	UNP D3GUW0
E	93	TRP	ASN	engineered mutation	UNP D3GUW0
E	158	TRP	SER	engineered mutation	UNP D3GUW0
E	164	HIS	-	expression tag	UNP D3GUW0
E	165	HIS	-	expression tag	UNP D3GUW0
E	166	HIS	-	expression tag	UNP D3GUW0
E	167	HIS	-	expression tag	UNP D3GUW0
E	168	HIS	-	expression tag	UNP D3GUW0
E	169	HIS	-	expression tag	UNP D3GUW0
F	93	TRP	ASN	engineered mutation	UNP D3GUW0
F	158	TRP	SER	engineered mutation	UNP D3GUW0
F	164	HIS	-	expression tag	UNP D3GUW0
F	165	HIS	-	expression tag	UNP D3GUW0
F	166	HIS	-	expression tag	UNP D3GUW0
F	167	HIS	-	expression tag	UNP D3GUW0
F	168	HIS	-	expression tag	UNP D3GUW0
F	169	HIS	-	expression tag	UNP D3GUW0
G	93	TRP	ASN	engineered mutation	UNP D3GUW0
G	158	TRP	SER	engineered mutation	UNP D3GUW0
G	164	HIS	-	expression tag	UNP D3GUW0
G	165	HIS	-	expression tag	UNP D3GUW0
G	166	HIS	-	expression tag	UNP D3GUW0
G	167	HIS	-	expression tag	UNP D3GUW0
G	168	HIS	-	expression tag	UNP D3GUW0
G	169	HIS	-	expression tag	UNP D3GUW0

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O3   O⁻ — S — O⁻ O4

S

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0
2	E	1	Total 5	O 4	S 1	0	0
2	E	1	Total 5	O 4	S 1	0	0
2	G	1	Total 5	O 4	S 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	126	Total 126	O 126	0	0

Continued on next page...

Continued from previous page...

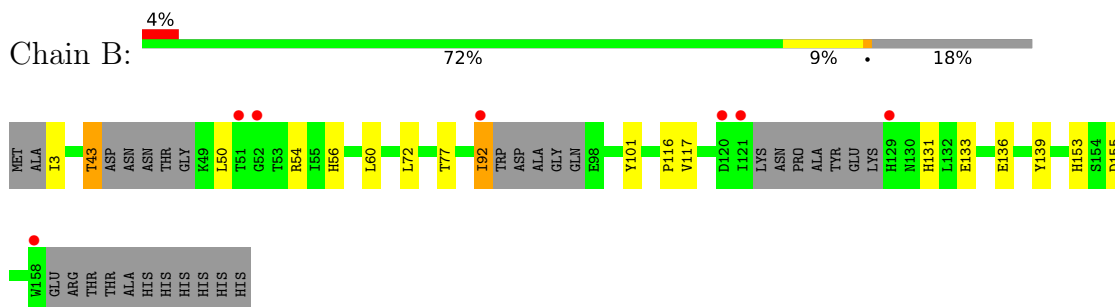| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	125	Total O 125 125		0	0
3	D	125	Total O 125 125		0	0
3	E	120	Total O 120 120		0	0
3	F	112	Total O 112 112		0	0
3	G	114	Total O 114 114		0	0

3 Residue-property plots ⓘ

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.
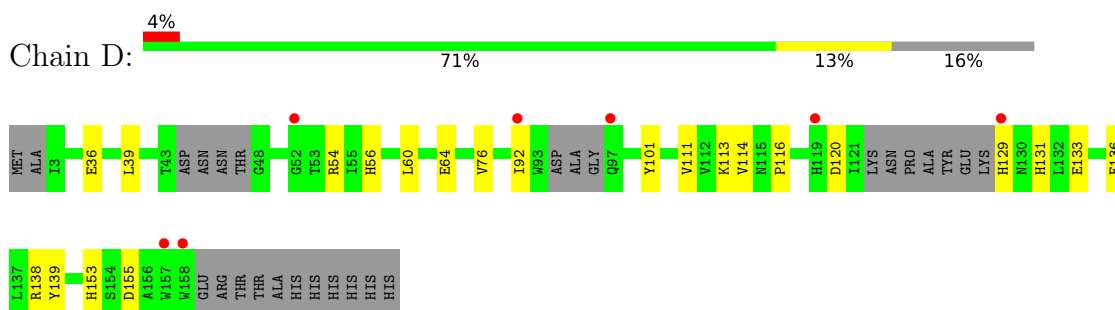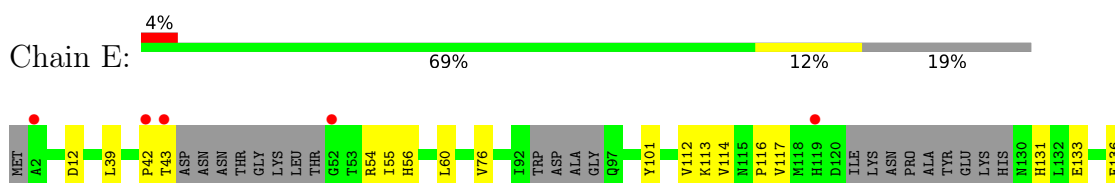
• Molecule 1: Putative type VI secretion protein
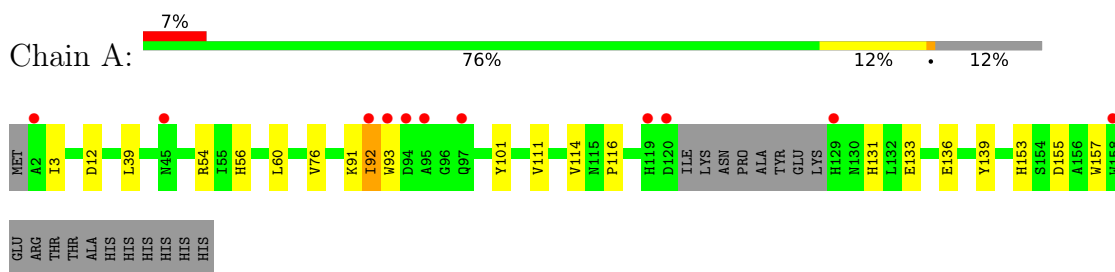
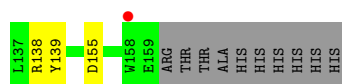Chain A:

• Molecule 1: Putative type VI secretion protein

Chain B:

• Molecule 1: Putative type VI secretion protein

Chain D:

• Molecule 1: Putative type VI secretion protein

Chain E:

L137 R138 Y139 D155 W158 E159 ARG THR THR ALA HIS HIS HIS HIS HIS HIS

• Molecule 1: Putative type VI secretion protein

Chain F:

4%

72% 12% 16%

MET A2 I3 E36 D44 ASN ASN THR G48 T51 G52 T53 R54 I55 H56 L60 V76 K91 I92 TRP ASP ALA GLY Q97 Y101 V112 K113 V114 N115 P116 D120 I121 LYS ASN PRO ALA TYR GLU LYS H129 N130 H131 L132 E133 E136 L137 R138 Y139 D155
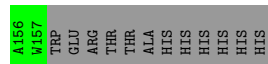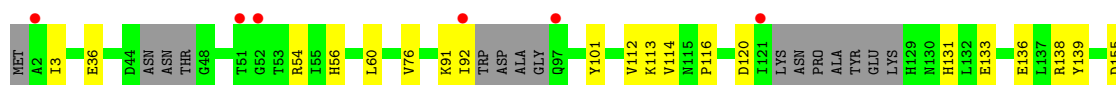
A156 W157 TRP GLU ARG THR THR ALA HIS HIS HIS HIS HIS HIS
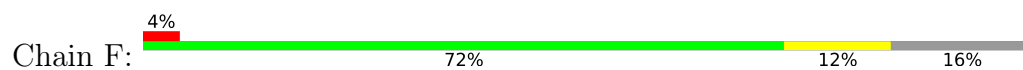
• Molecule 1: Putative type VI secretion protein

Chain G:
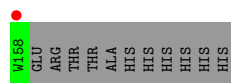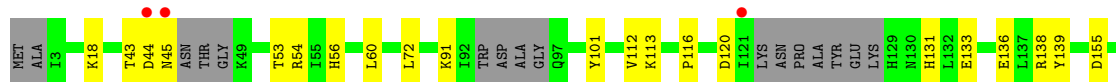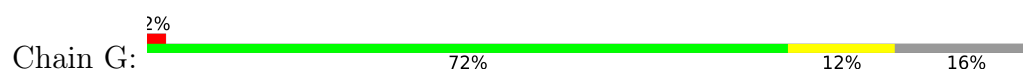
2%

72% 12% 16%

MET ALA I3 K18 T43 D44 N45 ASN THR GLY K49 T53 R54 I55 H56 L60 L72 K91 I92 TRP ASP ALA GLY Q97 Y101 V112 K113 P116 D120 I121 LYS ASN PRO ALA TYR GLU LYS H129 N130 H131 L132 E133 E136 L137 R138 Y139 D155

W158 GLU ARG THR THR ALA HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	84.21Å 145.89Å 89.85Å 90.00° 103.42° 90.00°	Depositor
Resolution (Å)	43.70 – 1.69 43.70 – 1.69	Depositor EDS
% Data completeness (in resolution range)	99.4 (43.70-1.69) 99.4 (43.70-1.69)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 1.69Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, R_{free}	0.180 , 0.196 0.190 , 0.209	Depositor DCC
R_{free} test set	5862 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	17.2	Xtriage
Anisotropy	0.774	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 55.5	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7794	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.72% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: SO4

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	0.51	0/1266	0.73	1/1721 (0.1%)
1	B	0.49	0/1186	0.69	0/1607
1	D	0.50	0/1208	0.70	1/1636 (0.1%)
1	E	0.49	0/1184	0.70	0/1605
1	F	0.48	0/1201	0.69	0/1625
1	G	0.49	0/1205	0.72	0/1634
All	All	0.49	0/7250	0.71	2/9828 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	92	ILE	C-N-CA	5.36	135.11	121.70
1	A	157	TRP	C-N-CA	5.11	134.47	121.70

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	1225	0	1190	13	0
1	B	1149	0	1129	11	0
1	D	1171	0	1151	16	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1141	0	1118	14	0
1	F	1163	0	1145	11	0
1	G	1168	0	1130	12	0
2	A	15	0	0	0	0
2	B	15	0	0	0	0
2	D	10	0	0	0	0
2	E	10	0	0	0	0
2	G	5	0	0	0	0
3	A	126	0	0	2	0
3	B	125	0	0	1	0
3	D	125	0	0	1	0
3	E	120	0	0	0	0
3	F	112	0	0	0	0
3	G	114	0	0	0	0
All	All	7794	0	6863	64	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.

The worst 5 of 64 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:3:ILE:HD11	1:B:92:ILE:HD12	1.60	0.82
1:D:138[B]:ARG:HH21	1:E:43:THR:HG21	1.47	0.79
1:A:60:LEU:HD22	1:A:136[A]:GLU:HG2	1.69	0.74
1:E:60:LEU:HD22	1:E:136[A]:GLU:HG2	1.69	0.74
1:D:60:LEU:HD22	1:D:136:GLU:HG2	1.70	0.74

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	148/169 (88%)	144 (97%)	4 (3%)	0	100	100
1	B	134/169 (79%)	131 (98%)	3 (2%)	0	100	100
1	D	137/169 (81%)	134 (98%)	2 (2%)	1 (1%)	22	8
1	E	134/169 (79%)	132 (98%)	2 (2%)	0	100	100
1	F	138/169 (82%)	136 (99%)	1 (1%)	1 (1%)	22	8
1	G	137/169 (81%)	134 (98%)	2 (2%)	1 (1%)	22	8
All	All	828/1014 (82%)	811 (98%)	14 (2%)	3 (0%)	34	18

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	120	ASP
1	G	120	ASP
1	F	120	ASP

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	133/151 (88%)	128 (96%)	5 (4%)	33	14
1	B	127/151 (84%)	123 (97%)	4 (3%)	40	21
1	D	128/151 (85%)	126 (98%)	2 (2%)	62	48
1	E	125/151 (83%)	122 (98%)	3 (2%)	49	31
1	F	127/151 (84%)	125 (98%)	2 (2%)	62	48
1	G	127/151 (84%)	121 (95%)	6 (5%)	26	10
All	All	767/906 (85%)	745 (97%)	22 (3%)	42	23

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	54	ARG
1	G	43	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	G	18	LYS
1	G	53	THR
1	B	54	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	134	GLN
1	G	56	HIS
1	E	80	GLN
1	G	131	HIS
1	F	115	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

11 ligands are modelled in this entry.

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	SO4	D	201	-	4,4,4	0.19	0	6,6,6	0.20	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
2	SO4	G	201	-	4,4,4	0.21	0	6,6,6	0.10	0
2	SO4	E	202	-	4,4,4	0.24	0	6,6,6	0.09	0
2	SO4	A	202	-	4,4,4	0.14	0	6,6,6	0.14	0
2	SO4	E	201	-	4,4,4	0.16	0	6,6,6	0.11	0
2	SO4	A	201	-	4,4,4	0.12	0	6,6,6	0.10	0
2	SO4	B	203	-	4,4,4	0.22	0	6,6,6	0.05	0
2	SO4	D	202	-	4,4,4	0.23	0	6,6,6	0.16	0
2	SO4	B	201	-	4,4,4	0.23	0	6,6,6	0.13	0
2	SO4	B	202	-	4,4,4	0.17	0	6,6,6	0.09	0
2	SO4	A	203	-	4,4,4	0.25	0	6,6,6	0.13	0

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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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^2)	Q<0.9
1	A	149/169 (88%)	0.12	11 (7%)	14	16	14, 21, 66, 81	0
1	B	139/169 (82%)	-0.11	7 (5%)	28	32	15, 22, 51, 78	0
1	D	142/169 (84%)	0.01	7 (4%)	29	33	15, 22, 58, 82	0
1	E	137/169 (81%)	-0.05	6 (4%)	34	38	13, 20, 46, 84	0
1	F	142/169 (84%)	0.00	6 (4%)	36	40	15, 23, 53, 83	0
1	G	142/169 (84%)	-0.03	4 (2%)	53	57	14, 22, 48, 80	0
All	All	851/1014 (83%)	-0.01	41 (4%)	30	34	13, 21, 57, 84	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	158	TRP	7.2
1	D	158	TRP	7.2
1	E	43	THR	4.7
1	A	2	ALA	4.5
1	E	158	TRP	4.5

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides 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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
2	SO4	E	201	5/5	0.86	0.24	81,82,82,82	0
2	SO4	A	203	5/5	0.87	0.22	70,70,72,73	0
2	SO4	B	201	5/5	0.89	0.23	65,67,68,68	0
2	SO4	G	201	5/5	0.92	0.17	70,70,71,71	0
2	SO4	E	202	5/5	0.93	0.18	69,69,71,71	0
2	SO4	D	202	5/5	0.94	0.16	55,55,56,57	0
2	SO4	B	203	5/5	0.94	0.22	67,68,69,70	0
2	SO4	B	202	5/5	0.95	0.18	55,56,58,59	0
2	SO4	D	201	5/5	0.95	0.18	52,55,56,56	0
2	SO4	A	202	5/5	0.96	0.14	58,58,59,59	0
2	SO4	A	201	5/5	0.97	0.21	59,59,61,63	0

6.5 Other polymers (i)

There are no such residues in this entry.