



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 08:39 PM UTC

PDB ID : 4HKH / pdb_00004hkh
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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

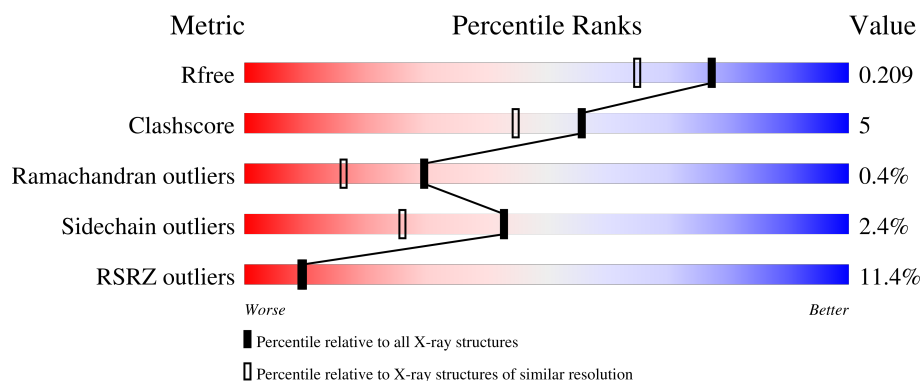
1 Overall quality at a glance

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}	180053	5551 (1.70-1.70)
Clashscore	190562	5924 (1.70-1.70)
Ramachandran outliers	187476	5846 (1.70-1.70)
Sidechain outliers	187428	5846 (1.70-1.70)
RSRZ outliers	180081	5554 (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 $\leq 5\%$. 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	169	11% 76% 11% • 12%
1	B	169	8% 70% 11% • 18%
1	D	169	12% 69% 15% 16%
1	E	169	7% 68% 12% • 19%
1	F	169	11% 70% 14% 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	169	 A horizontal bar chart showing the quality of chain G. The bar is divided into four segments: a small red segment at the beginning labeled '9%', followed by a large green segment labeled '70%', then a yellow segment labeled '14%', and finally a grey segment at the end labeled '16%'.

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

There are 48 discrepancies between the modelled and reference sequences:

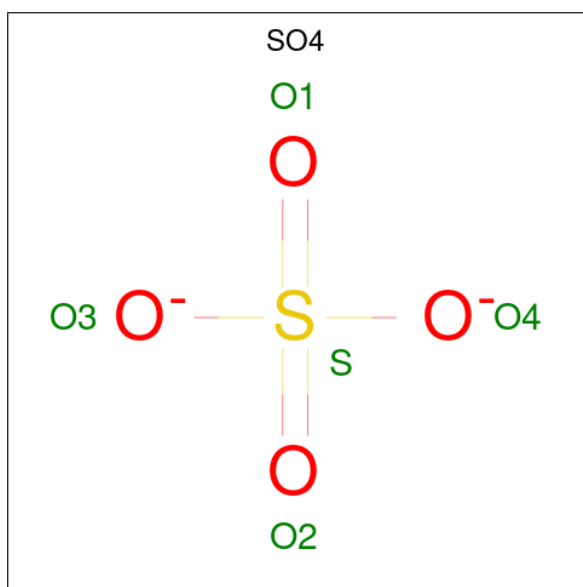
Chain	Residue	Modelled	Actual	Comment	Reference
A	93	TRP	ASN	engineered mutation	UNP D3GUW0
A	158	TRP	SER	engineered mutation	UNP D3GUW0
A	164	HIS	-	expression tag	UNP D3GUW0
A	165	HIS	-	expression tag	UNP D3GUW0
A	166	HIS	-	expression tag	UNP D3GUW0
A	167	HIS	-	expression tag	UNP D3GUW0
A	168	HIS	-	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 (CCD ID: SO4) (formula: O₄S).



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

- Molecule 3 is water.

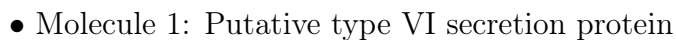
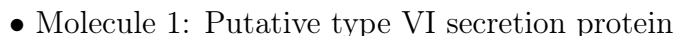
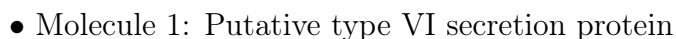
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	126	Total	O	0	0
			126	126		

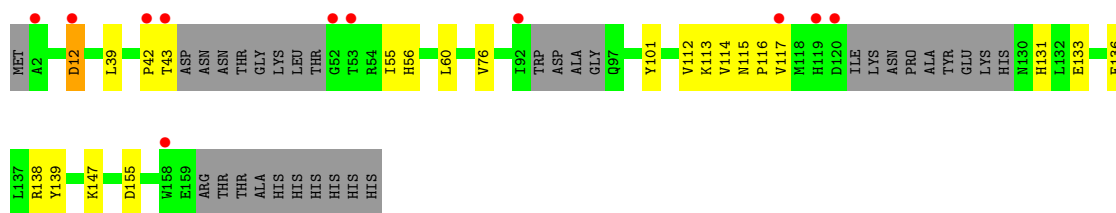
Continued on next page...

Continued from previous page...

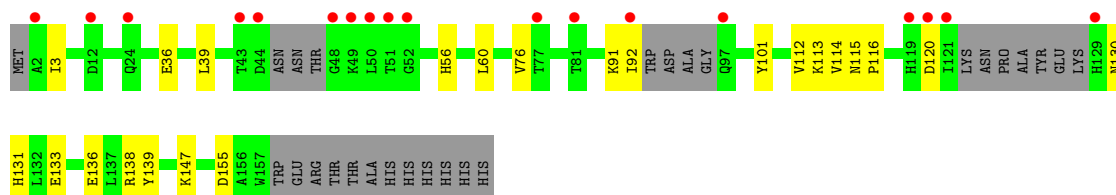
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	125	Total 125	O 125	0	0
3	D	125	Total 125	O 125	0	0
3	E	120	Total 120	O 120	0	0
3	F	112	Total 112	O 112	0	0
3	G	114	Total 114	O 114	0	0

- Molecule 1: Putative type VI secretion protein

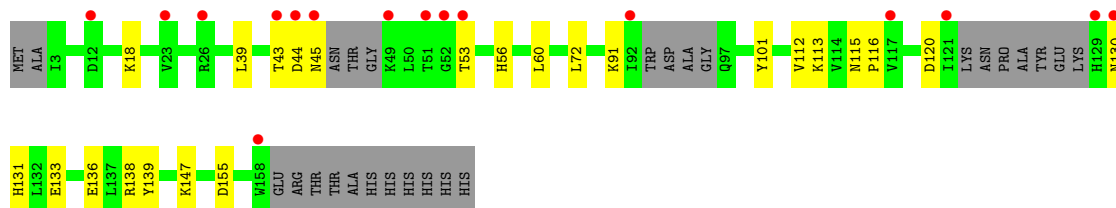




● Molecule 1: Putative type VI secretion protein



● Molecule 1: Putative type VI secretion protein



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
$\langle I/\sigma(I) \rangle$ ¹	1.91 (at 1.69Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, R_{free}	0.180 , 0.196 0.189 , 0.209	Depositor DCC
R_{free} test set	5862 reflections (4.99%)	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)	0.34 , 54.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	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.*

¹Intensities estimated from amplitudes.

²Theoretical 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: 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	0.81	0/1266	1.14	6/1721 (0.3%)
1	B	0.77	0/1186	1.06	3/1607 (0.2%)
1	D	0.80	0/1208	1.11	6/1636 (0.4%)
1	E	0.77	0/1184	1.07	3/1605 (0.2%)
1	F	0.78	0/1201	1.05	3/1625 (0.2%)
1	G	0.76	0/1205	1.10	3/1634 (0.2%)
All	All	0.78	0/7250	1.09	24/9828 (0.2%)

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	12	ASP	CA-CB-CG	7.85	120.45	112.60
1	D	92	ILE	CA-C-N	7.45	135.11	121.70
1	D	92	ILE	C-N-CA	7.45	135.11	121.70
1	A	12	ASP	CA-CB-CG	7.29	119.89	112.60
1	A	157	TRP	CA-C-N	7.09	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	17	0
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	65	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 65 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 [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	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%)	18	7
1	E	134/169 (79%)	132 (98%)	2 (2%)	0	100	100
1	F	138/169 (82%)	136 (99%)	1 (1%)	1 (1%)	18	7
1	G	137/169 (81%)	134 (98%)	2 (2%)	1 (1%)	18	7
All	All	828/1014 (82%)	811 (98%)	14 (2%)	3 (0%)	30	16

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%)	130 (98%)	3 (2%)	44	27
1	B	127/151 (84%)	123 (97%)	4 (3%)	35	18
1	D	128/151 (85%)	127 (99%)	1 (1%)	73	65
1	E	125/151 (83%)	123 (98%)	2 (2%)	55	41
1	F	127/151 (84%)	125 (98%)	2 (2%)	55	41
1	G	127/151 (84%)	121 (95%)	6 (5%)	23	9
All	All	767/906 (85%)	749 (98%)	18 (2%)	43	27

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	43	THR
1	G	91	LYS
1	G	72	LEU
1	E	12	ASP
1	G	39	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	56	HIS
1	G	103	ASN
1	G	153	HIS
1	D	103	ASN
1	D	80	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

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	A	203	-	4,4,4	0.36	0	6,6,6	0.14	0
2	SO4	B	203	-	4,4,4	0.34	0	6,6,6	0.06	0
2	SO4	A	202	-	4,4,4	0.22	0	6,6,6	0.14	0
2	SO4	A	201	-	4,4,4	0.22	0	6,6,6	0.10	0
2	SO4	B	201	-	4,4,4	0.34	0	6,6,6	0.15	0
2	SO4	E	201	-	4,4,4	0.28	0	6,6,6	0.14	0
2	SO4	D	201	-	4,4,4	0.28	0	6,6,6	0.19	0
2	SO4	G	201	-	4,4,4	0.32	0	6,6,6	0.12	0
2	SO4	B	202	-	4,4,4	0.29	0	6,6,6	0.06	0
2	SO4	E	202	-	4,4,4	0.35	0	6,6,6	0.11	0
2	SO4	D	202	-	4,4,4	0.34	0	6,6,6	0.16	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 [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, 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%)	0.47	18 (12%) 8 8	11, 21, 66, 81	3 (2%)
1	B	139/169 (82%)	0.37	13 (9%) 14 14	11, 22, 51, 78	3 (2%)
1	D	142/169 (84%)	0.49	21 (14%) 5 5	11, 22, 58, 82	3 (2%)
1	E	137/169 (81%)	0.25	11 (8%) 18 19	10, 19, 46, 84	5 (3%)
1	F	142/169 (84%)	0.44	18 (12%) 8 7	11, 22, 53, 83	4 (2%)
1	G	142/169 (84%)	0.45	16 (11%) 10 10	10, 21, 48, 80	3 (2%)
All	All	851/1014 (83%)	0.41	97 (11%) 10 9	10, 21, 57, 84	21 (2%)

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	93	TRP	9.2
1	D	158	TRP	7.4
1	F	121	ILE	7.3
1	E	43	THR	7.3
1	G	121	ILE	7.0

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 oligosaccharides 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	0.77	0.14	81,82,82,82	0
2	SO4	B	203	5/5	0.78	0.17	67,68,69,70	0
2	SO4	B	201	5/5	0.81	0.14	65,67,68,68	0
2	SO4	A	203	5/5	0.83	0.14	70,70,72,73	0
2	SO4	E	202	5/5	0.85	0.13	69,69,71,71	0
2	SO4	G	201	5/5	0.87	0.12	70,70,71,71	0
2	SO4	D	202	5/5	0.88	0.14	55,55,56,57	0
2	SO4	A	201	5/5	0.89	0.13	59,59,61,63	0
2	SO4	D	201	5/5	0.89	0.12	52,55,56,56	0
2	SO4	B	202	5/5	0.91	0.11	55,56,58,59	0
2	SO4	A	202	5/5	0.92	0.11	58,58,59,59	0

6.5 Other polymers [i](#)

There are no such residues in this entry.