



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 24, 2024 – 07:52 AM EDT

PDB ID : 6GBH
Title : Helicobacter pylori adhesin HopQ type II bound to the N-terminal domain of human CEACAM1
Authors : Moonens, K.; Kruse, T.; Gerhard, M.; Remaut, H.
Deposited on : 2018-04-13
Resolution : 2.59 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	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
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)	:	2.37.1

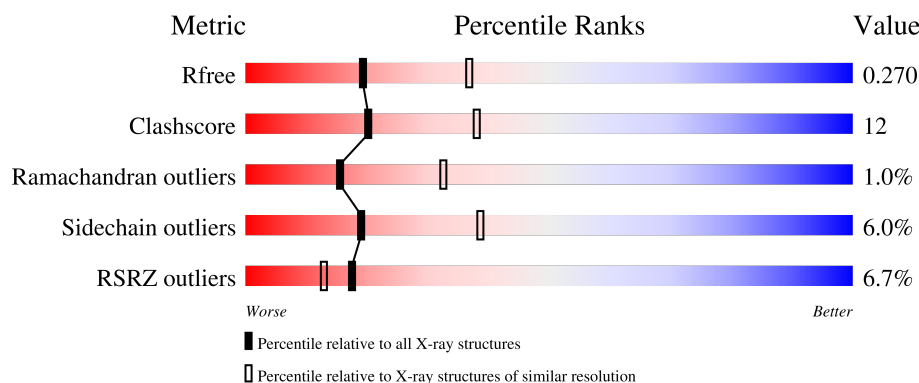
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 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	B	115	<div> <div>10%</div> <div> <div></div> <div>63%</div> <div>28%</div> <div>• 6%</div> </div> </div>
1	D	115	<div> <div>17%</div> <div> <div></div> <div>74%</div> <div>17%</div> <div>•• 6%</div> </div> </div>
2	A	425	<div> <div>3%</div> <div> <div></div> <div>54%</div> <div>19%</div> <div>• 25%</div> </div> </div>
2	C	425	<div> <div>4%</div> <div> <div></div> <div>53%</div> <div>20%</div> <div>• 25%</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6490 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 Carcinoembryonic antigen-related cell adhesion molecule 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	108	Total	C	N	O	S	0	0	0
			847	537	143	166	1			
1	D	108	Total	C	N	O	S	0	1	0
			852	540	144	167	1			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P13688
B	109	HIS	-	expression tag	UNP P13688
B	110	HIS	-	expression tag	UNP P13688
B	111	HIS	-	expression tag	UNP P13688
B	112	HIS	-	expression tag	UNP P13688
B	113	HIS	-	expression tag	UNP P13688
B	114	HIS	-	expression tag	UNP P13688
D	0	MET	-	initiating methionine	UNP P13688
D	109	HIS	-	expression tag	UNP P13688
D	110	HIS	-	expression tag	UNP P13688
D	111	HIS	-	expression tag	UNP P13688
D	112	HIS	-	expression tag	UNP P13688
D	113	HIS	-	expression tag	UNP P13688
D	114	HIS	-	expression tag	UNP P13688

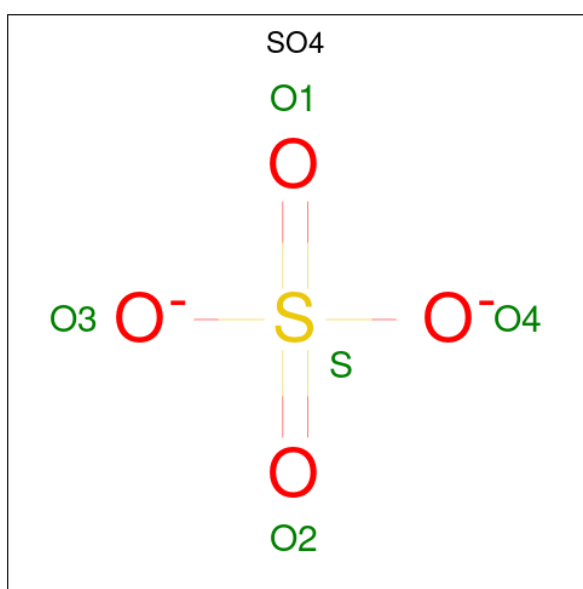
- Molecule 2 is a protein called HopQ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	318	Total	C	N	O	S	0	0	0
			2367	1453	411	495	8			
2	A	318	Total	C	N	O	S	0	1	0
			2378	1461	414	495	8			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	16	MET	-	initiating methionine	UNP Q8GDI6
C	435	HIS	-	expression tag	UNP Q8GDI6
C	436	HIS	-	expression tag	UNP Q8GDI6
C	437	HIS	-	expression tag	UNP Q8GDI6
C	438	HIS	-	expression tag	UNP Q8GDI6
C	439	HIS	-	expression tag	UNP Q8GDI6
C	440	HIS	-	expression tag	UNP Q8GDI6
A	16	MET	-	initiating methionine	UNP Q8GDI6
A	435	HIS	-	expression tag	UNP Q8GDI6
A	436	HIS	-	expression tag	UNP Q8GDI6
A	437	HIS	-	expression tag	UNP Q8GDI6
A	438	HIS	-	expression tag	UNP Q8GDI6
A	439	HIS	-	expression tag	UNP Q8GDI6
A	440	HIS	-	expression tag	UNP Q8GDI6

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	C	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		

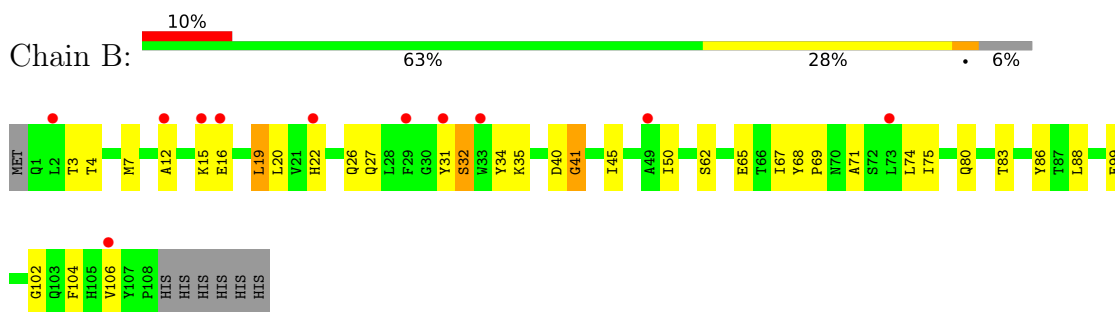
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	O	0	0
			1	1		
4	D	1	Total	O	0	0
			1	1		
4	C	4	Total	O	0	0
			4	4		

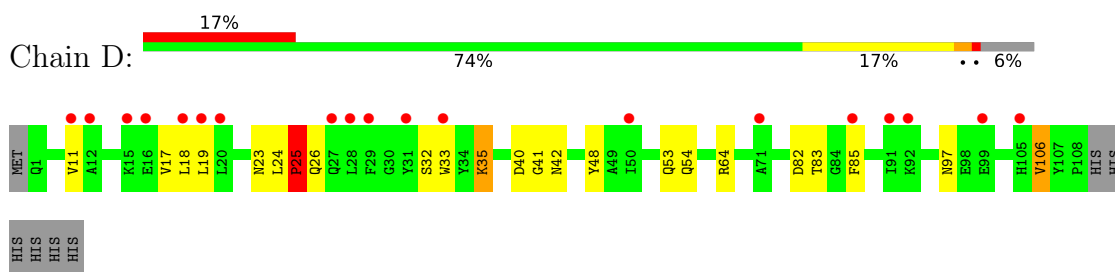
3 Residue-property plots [i](#)

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.

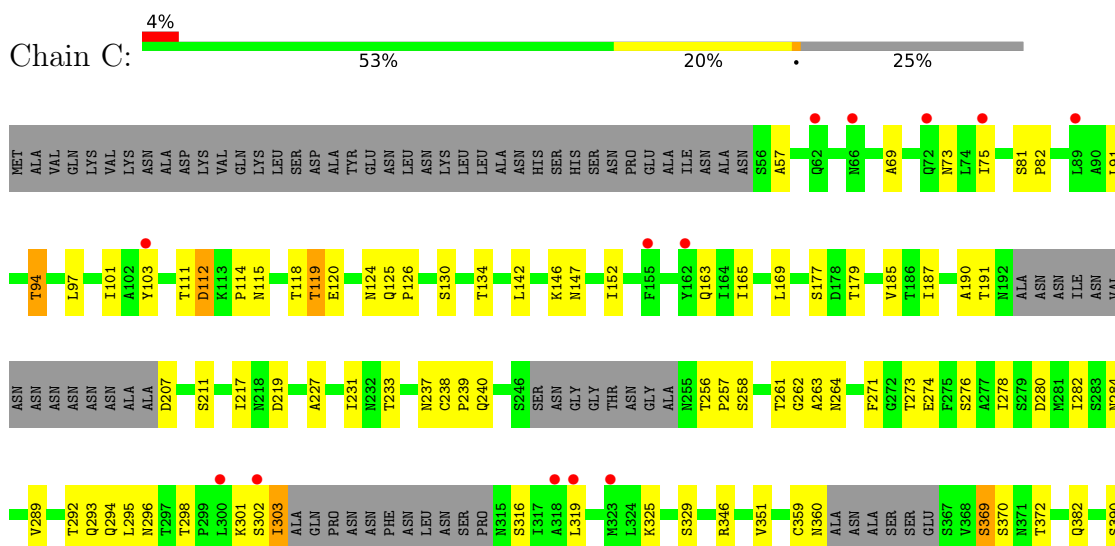
- Molecule 1: Carcinoembryonic antigen-related cell adhesion molecule 1

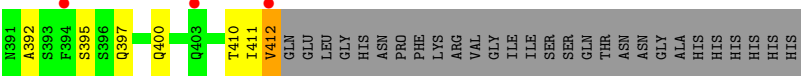


- Molecule 1: Carcinoembryonic antigen-related cell adhesion molecule 1

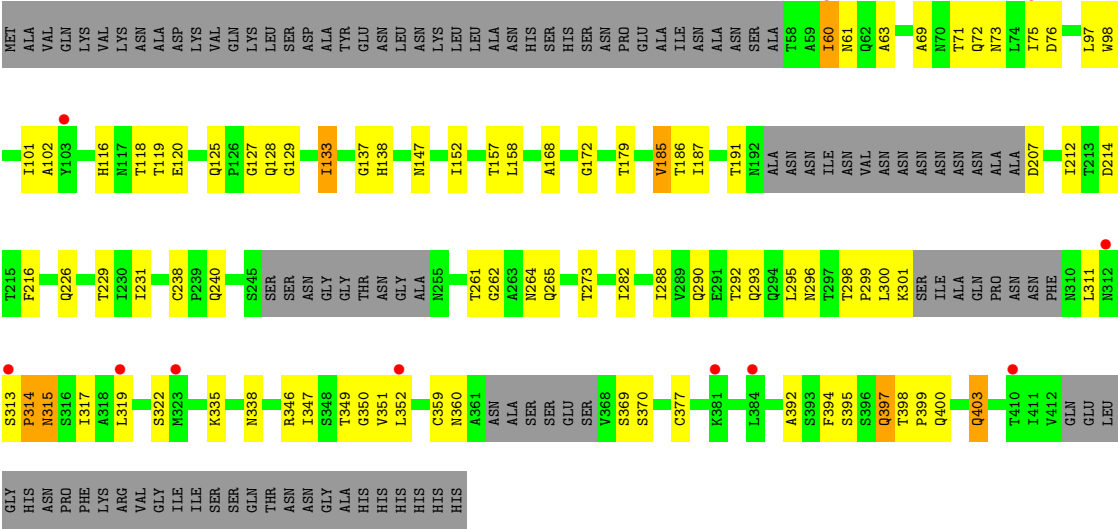


- Molecule 2: HopQ





• Molecule 2: HopQ



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	79.82Å 169.67Å 198.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	99.05 – 2.59 48.73 – 2.59	Depositor EDS
% Data completeness (in resolution range)	98.3 (99.05-2.59) 98.4 (48.73-2.59)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.09 (at 2.58Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.218 , 0.274 0.222 , 0.270	Depositor DCC
R_{free} test set	2071 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	70.4	Xtriage
Anisotropy	0.842	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 65.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6490	wwPDB-VP
Average B, all atoms (Å ²)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.03% 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

5.1 Standard geometry

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	B	0.61	0/866	0.82	1/1181 (0.1%)
1	D	0.61	1/874 (0.1%)	0.75	0/1192
2	A	0.71	0/2406	0.90	6/3272 (0.2%)
2	C	0.70	2/2391 (0.1%)	0.86	3/3250 (0.1%)
All	All	0.68	3/6537 (0.0%)	0.86	10/8895 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	112	ASP	CB-CG	5.32	1.62	1.51
2	C	274	GLU	CD-OE2	5.28	1.31	1.25
1	D	54	GLN	N-CA	-5.15	1.36	1.46

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	214	ASP	CB-CG-OD2	-7.28	111.75	118.30
2	C	303	ILE	N-CA-CB	6.16	124.97	110.80
2	A	214	ASP	CB-CG-OD1	6.13	123.82	118.30
2	A	129	GLY	N-CA-C	-5.83	98.53	113.10
2	A	128	GLN	N-CA-C	5.72	126.46	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	25	PRO	Peptide

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	B	847	0	821	23	0
1	D	852	0	827	15	0
2	A	2378	0	2341	60	0
2	C	2367	0	2329	65	0
3	A	10	0	0	1	0
3	B	10	0	0	1	0
3	C	5	0	0	1	0
3	D	15	0	0	1	0
4	B	1	0	0	0	0
4	C	4	0	0	0	0
4	D	1	0	0	0	0
All	All	6490	0	6318	160	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 12.

The worst 5 of 160 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)
2:C:261:THR:O	2:C:263:ALA:N	1.77	1.16
1:B:83:THR:HB	1:B:106:VAL:HG12	1.45	0.98
2:C:316:SER:HA	2:C:319:LEU:HD23	1.53	0.88
2:A:346:ARG:O	2:A:349:THR:HB	1.75	0.87
2:A:120:GLU:HG3	2:A:152:ILE:HG21	1.54	0.87

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	B	106/115 (92%)	99 (93%)	6 (6%)	1 (1%)	17	35
1	D	107/115 (93%)	97 (91%)	9 (8%)	1 (1%)	17	35
2	A	309/425 (73%)	273 (88%)	33 (11%)	3 (1%)	15	32
2	C	308/425 (72%)	285 (92%)	20 (6%)	3 (1%)	15	32
All	All	830/1080 (77%)	754 (91%)	68 (8%)	8 (1%)	15	32

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	57	ALA
2	C	119	THR
2	C	262	GLY
1	B	41	GLY
2	A	314	PRO

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	B	92/99 (93%)	88 (96%)	4 (4%)	29	54
1	D	93/99 (94%)	89 (96%)	4 (4%)	29	54
2	A	269/355 (76%)	255 (95%)	14 (5%)	23	46
2	C	268/355 (76%)	247 (92%)	21 (8%)	12	25
All	All	722/908 (80%)	679 (94%)	43 (6%)	19	39

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	412	VAL
2	A	295	LEU
2	A	118	THR
2	A	158	LEU
2	A	351	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	A	360	ASN
2	A	402	ASN
2	C	265	GLN
2	C	389	ASN
2	A	166	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](#)

8 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
3	SO4	A	501	-	4,4,4	0.40	0	6,6,6	0.50	0
3	SO4	D	202	-	4,4,4	0.39	0	6,6,6	0.16	0
3	SO4	C	501	-	4,4,4	0.31	0	6,6,6	0.23	0
3	SO4	B	202	-	4,4,4	0.38	0	6,6,6	0.55	0
3	SO4	B	201	-	4,4,4	0.43	0	6,6,6	0.73	0
3	SO4	D	201	-	4,4,4	0.29	0	6,6,6	0.34	0
3	SO4	D	203	-	4,4,4	0.31	0	6,6,6	0.15	0
3	SO4	A	502	-	4,4,4	0.35	0	6,6,6	0.22	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.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	SO4	1	0
3	D	202	SO4	1	0
3	C	501	SO4	1	0
3	B	201	SO4	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 [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	B	108/115 (93%)	0.91	11 (10%) 6 4	71, 100, 133, 150	0
1	D	108/115 (93%)	1.07	19 (17%) 1 0	72, 101, 128, 144	0
2	A	318/425 (74%)	0.63	11 (3%) 44 36	65, 87, 132, 197	0
2	C	318/425 (74%)	0.68	16 (5%) 28 23	66, 86, 120, 154	0
All	All	852/1080 (78%)	0.74	57 (6%) 17 13	65, 90, 128, 197	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	312	ASN	8.2
1	D	105	HIS	4.3
1	B	16	GLU	4.2
2	C	62	GLN	3.8
1	D	11	VAL	3.6

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](#)

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(\AA^2)	Q<0.9
3	SO4	B	202	5/5	0.81	0.25	121,129,137,137	0
3	SO4	D	203	5/5	0.82	0.27	145,160,164,169	0
3	SO4	C	501	5/5	0.83	0.17	112,131,141,148	0
3	SO4	D	201	5/5	0.86	0.13	116,121,142,142	0
3	SO4	D	202	5/5	0.86	0.12	136,140,156,164	0
3	SO4	B	201	5/5	0.89	0.13	93,107,128,140	0
3	SO4	A	502	5/5	0.89	0.18	109,117,133,155	0
3	SO4	A	501	5/5	0.90	0.18	98,130,136,152	0

6.5 Other polymers [i](#)

There are no such residues in this entry.