WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 14, 2024 – 11:52 AM EST

PDB ID	:	3M5N
Title	:	Crystal structure of HCV NS3/4A protease in complex with N-terminal product 4B5A
Authors	:	Schiffer, C.A.; Romano, K.P.
Deposited on	:	2010-03-12
Resolution	:	1.90 Å(reported)

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

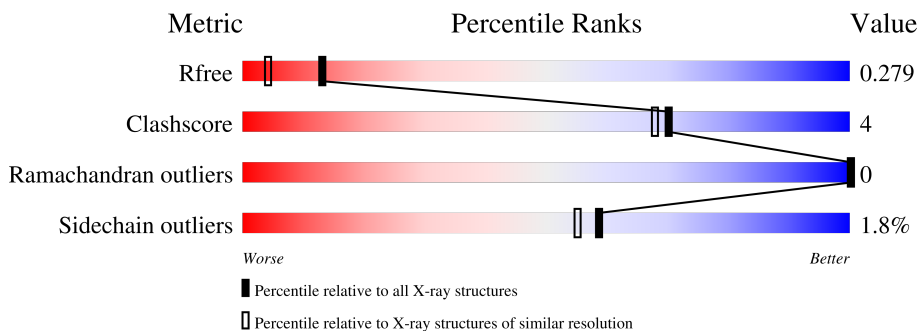|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.90 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	Percentile Ranks	Value
Rfree		0.279
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		1.8%

Worse — *Better*

▮ Percentile relative to all X-ray structures

⫴ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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%

Mol	Chain	Length	Quality of chain
1	A	203	92% · · · 6% ·
1	B	203	93% · · · 5% ·
1	C	203	86% · · · 10% ·
1	D	203	89% · · · 8% ·
2	F	7	86% · · · 14%
2	G	7	86% · · · 14%
2	H	7	100%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 6372 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 NS3/4A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	200	Total	C	N	O	S	0	0	0
			1445	894	266	277	8			
1	B	199	Total	C	N	O	S	0	0	0
			1420	880	257	275	8			
1	C	197	Total	C	N	O	S	0	0	0
			1418	881	260	269	8			
1	D	198	Total	C	N	O	S	0	0	0
			1421	881	257	275	8			

There are 108 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	981	SER	-	expression tag	UNP A8DG50
A	982	HIS	-	expression tag	UNP A8DG50
A	983	MET	-	expression tag	UNP A8DG50
A	984	ALA	-	expression tag	UNP A8DG50
A	985	SER	-	expression tag	UNP A8DG50
A	986	MET	-	engineered mutation	UNP A8DG50
A	987	LYS	-	engineered mutation	UNP A8DG50
A	988	LYS	-	engineered mutation	UNP A8DG50
A	989	LYS	-	engineered mutation	UNP A8DG50
A	991	SER	CYS	SEE REMARK 999	UNP A8DG50
A	998	ILE	VAL	SEE REMARK 999	UNP A8DG50
A	999	ASN	ILE	SEE REMARK 999	UNP A8DG50
A	1001	SER	ALA	engineered mutation	UNP A8DG50
A	1002	GLY	PRO	engineered mutation	UNP A8DG50
A	1003	ASP	ILE	engineered mutation	UNP A8DG50
A	1013	GLU	LEU	engineered mutation	UNP A8DG50
A	1014	GLU	LEU	engineered mutation	UNP A8DG50
A	1017	GLN	ILE	engineered mutation	UNP A8DG50
A	1018	GLU	ILE	engineered mutation	UNP A8DG50
A	1021	GLN	LEU	engineered mutation	UNP A8DG50
A	1040	THR	ALA	engineered mutation	UNP A8DG50

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1047	SER	CYS	engineered mutation	UNP A8DG50
A	1052	LEU	CYS	engineered mutation	UNP A8DG50
A	1072	THR	ILE	engineered mutation	UNP A8DG50
A	1086	GLN	PRO	engineered mutation	UNP A8DG50
A	1139	ALA	SER	engineered mutation	UNP A8DG50
A	1159	SER	CYS	engineered mutation	UNP A8DG50
B	981	SER	-	expression tag	UNP A8DG50
B	982	HIS	-	expression tag	UNP A8DG50
B	983	MET	-	expression tag	UNP A8DG50
B	984	ALA	-	expression tag	UNP A8DG50
B	985	SER	-	expression tag	UNP A8DG50
B	986	MET	-	engineered mutation	UNP A8DG50
B	987	LYS	-	engineered mutation	UNP A8DG50
B	988	LYS	-	engineered mutation	UNP A8DG50
B	989	LYS	-	engineered mutation	UNP A8DG50
B	991	SER	CYS	SEE REMARK 999	UNP A8DG50
B	998	ILE	VAL	SEE REMARK 999	UNP A8DG50
B	999	ASN	ILE	SEE REMARK 999	UNP A8DG50
B	1001	SER	ALA	engineered mutation	UNP A8DG50
B	1002	GLY	PRO	engineered mutation	UNP A8DG50
B	1003	ASP	ILE	engineered mutation	UNP A8DG50
B	1013	GLU	LEU	engineered mutation	UNP A8DG50
B	1014	GLU	LEU	engineered mutation	UNP A8DG50
B	1017	GLN	ILE	engineered mutation	UNP A8DG50
B	1018	GLU	ILE	engineered mutation	UNP A8DG50
B	1021	GLN	LEU	engineered mutation	UNP A8DG50
B	1040	THR	ALA	engineered mutation	UNP A8DG50
B	1047	SER	CYS	engineered mutation	UNP A8DG50
B	1052	LEU	CYS	engineered mutation	UNP A8DG50
B	1072	THR	ILE	engineered mutation	UNP A8DG50
B	1086	GLN	PRO	engineered mutation	UNP A8DG50
B	1139	ALA	SER	engineered mutation	UNP A8DG50
B	1159	SER	CYS	engineered mutation	UNP A8DG50
C	981	SER	-	expression tag	UNP A8DG50
C	982	HIS	-	expression tag	UNP A8DG50
C	983	MET	-	expression tag	UNP A8DG50
C	984	ALA	-	expression tag	UNP A8DG50
C	985	SER	-	expression tag	UNP A8DG50
C	986	MET	-	engineered mutation	UNP A8DG50
C	987	LYS	-	engineered mutation	UNP A8DG50
C	988	LYS	-	engineered mutation	UNP A8DG50
C	989	LYS	-	engineered mutation	UNP A8DG50

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	991	SER	CYS	SEE REMARK 999	UNP A8DG50
C	998	ILE	VAL	SEE REMARK 999	UNP A8DG50
C	999	ASN	ILE	SEE REMARK 999	UNP A8DG50
C	1001	SER	ALA	engineered mutation	UNP A8DG50
C	1002	GLY	PRO	engineered mutation	UNP A8DG50
C	1003	ASP	ILE	engineered mutation	UNP A8DG50
C	1013	GLU	LEU	engineered mutation	UNP A8DG50
C	1014	GLU	LEU	engineered mutation	UNP A8DG50
C	1017	GLN	ILE	engineered mutation	UNP A8DG50
C	1018	GLU	ILE	engineered mutation	UNP A8DG50
C	1021	GLN	LEU	engineered mutation	UNP A8DG50
C	1040	THR	ALA	engineered mutation	UNP A8DG50
C	1047	SER	CYS	engineered mutation	UNP A8DG50
C	1052	LEU	CYS	engineered mutation	UNP A8DG50
C	1072	THR	ILE	engineered mutation	UNP A8DG50
C	1086	GLN	PRO	engineered mutation	UNP A8DG50
C	1139	ALA	SER	engineered mutation	UNP A8DG50
C	1159	SER	CYS	engineered mutation	UNP A8DG50
D	981	SER	-	expression tag	UNP A8DG50
D	982	HIS	-	expression tag	UNP A8DG50
D	983	MET	-	expression tag	UNP A8DG50
D	984	ALA	-	expression tag	UNP A8DG50
D	985	SER	-	expression tag	UNP A8DG50
D	986	MET	-	engineered mutation	UNP A8DG50
D	987	LYS	-	engineered mutation	UNP A8DG50
D	988	LYS	-	engineered mutation	UNP A8DG50
D	989	LYS	-	engineered mutation	UNP A8DG50
D	991	SER	CYS	SEE REMARK 999	UNP A8DG50
D	998	ILE	VAL	SEE REMARK 999	UNP A8DG50
D	999	ASN	ILE	SEE REMARK 999	UNP A8DG50
D	1001	SER	ALA	engineered mutation	UNP A8DG50
D	1002	GLY	PRO	engineered mutation	UNP A8DG50
D	1003	ASP	ILE	engineered mutation	UNP A8DG50
D	1013	GLU	LEU	engineered mutation	UNP A8DG50
D	1014	GLU	LEU	engineered mutation	UNP A8DG50
D	1017	GLN	ILE	engineered mutation	UNP A8DG50
D	1018	GLU	ILE	engineered mutation	UNP A8DG50
D	1021	GLN	LEU	engineered mutation	UNP A8DG50
D	1040	THR	ALA	engineered mutation	UNP A8DG50
D	1047	SER	CYS	engineered mutation	UNP A8DG50
D	1052	LEU	CYS	engineered mutation	UNP A8DG50
D	1072	THR	ILE	engineered mutation	UNP A8DG50

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	1086	GLN	PRO	engineered mutation	UNP A8DG50
D	1139	ALA	SER	engineered mutation	UNP A8DG50
D	1159	SER	CYS	engineered mutation	UNP A8DG50

• Molecule 2 is a protein called SECTTPC peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	7	Total	C	N	O	S	0	0	0
			49	27	7	13	2			
2	G	7	Total	C	N	O	S	0	0	0
			49	27	7	13	2			
2	H	7	Total	C	N	O	S	0	0	0
			49	27	7	13	2			

• Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
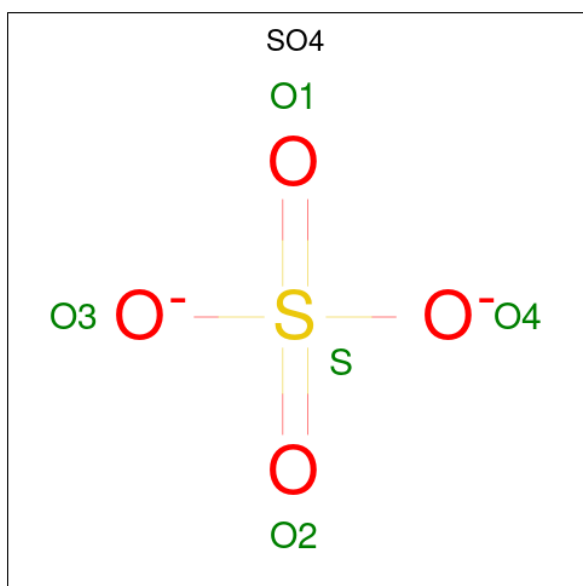
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	C	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		

• Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

• Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	131	Total 135	O 135	0	4
5	B	141	Total 142	O 142	0	1
5	C	103	Total 107	O 107	0	4
5	D	105	Total 107	O 107	0	2
5	F	3	Total 3	O 3	0	0
5	G	2	Total 2	O 2	0	0
5	H	1	Total 1	O 1	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. 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. 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: NS3/4A

Chain A:

• Molecule 1: NS3/4A

Chain B:

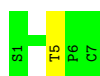• Molecule 1: NS3/4A

Chain C:

• Molecule 1: NS3/4A

Chain D:

• Molecule 2: SECTTPC peptide

Chain F:

• Molecule 2: SECTTPC peptide

Chain G:

• Molecule 2: SECTTPC peptide

Chain H:

There are no outlier residues recorded for this chain.

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	67.31Å 58.89Å 98.27Å 90.00° 101.01° 90.00°	Depositor
Resolution (Å)	50.00 – 1.90 43.07 – 1.90	Depositor EDS
% Data completeness (in resolution range)	98.9 (50.00-1.90) 98.9 (43.07-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	3.02 (at 1.89Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.188 , 0.240 0.234 , 0.279	Depositor DCC
R_{free} test set	3010 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	20.5	Xtriage
Anisotropy	0.166	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 44.0	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6372	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 36.41 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.0213e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, ZN

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	0.46	0/1470	0.80	2/2001 (0.1%)
1	B	0.48	0/1445	0.82	2/1971 (0.1%)
1	C	0.41	0/1442	0.76	1/1962 (0.1%)
1	D	0.42	0/1446	0.76	0/1971
2	F	0.50	0/49	0.78	0/65
2	G	0.54	0/49	0.77	0/65
2	H	0.41	0/49	0.70	0/65
All	All	0.44	0/5950	0.78	5/8100 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1123	ARG	NE-CZ-NH1	6.19	123.40	120.30
1	B	1011	ARG	NE-CZ-NH2	-5.76	117.42	120.30
1	B	1011	ARG	NE-CZ-NH1	5.74	123.17	120.30
1	C	1143	LEU	CA-CB-CG	5.11	127.05	115.30
1	A	1123	ARG	NE-CZ-NH2	-5.04	117.78	120.30

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	1445	0	1423	5	0
1	B	1420	0	1382	7	0
1	C	1418	0	1412	16	0
1	D	1421	0	1390	14	0
2	F	49	0	44	1	0
2	G	49	0	44	0	0
2	H	49	0	44	0	0
3	A	5	0	0	0	0
3	B	5	0	0	0	0
3	C	5	0	0	0	0
3	D	5	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	135	0	0	1	0
5	B	142	0	0	0	0
5	C	107	0	0	1	0
5	D	107	0	0	1	0
5	F	3	0	0	0	0
5	G	2	0	0	0	0
5	H	1	0	0	0	0
All	All	6372	0	5739	42	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 4.

The worst 5 of 42 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:1132:ILE:HD11	2:F:5:THR:HB	1.61	0.82
1:C:1042:THR:HG21	1:C:1109:ARG:HH22	1.44	0.82
1:B:1042:THR:HG21	1:B:1109:ARG:HH22	1.45	0.81
1:D:1114:ILE:HD11	1:D:1134:TYR:HE2	1.52	0.75
1:C:1042:THR:CG2	1:C:1109:ARG:HH22	2.01	0.72

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	198/203 (98%)	197 (100%)	1 (0%)	0	100	100
1	B	197/203 (97%)	194 (98%)	3 (2%)	0	100	100
1	C	195/203 (96%)	194 (100%)	1 (0%)	0	100	100
1	D	196/203 (97%)	195 (100%)	1 (0%)	0	100	100
2	F	5/7 (71%)	5 (100%)	0	0	100	100
2	G	5/7 (71%)	5 (100%)	0	0	100	100
2	H	5/7 (71%)	5 (100%)	0	0	100	100
All	All	801/833 (96%)	795 (99%)	6 (1%)	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	152/164 (93%)	149 (98%)	3 (2%)	55	51
1	B	148/164 (90%)	146 (99%)	2 (1%)	67	65
1	C	149/164 (91%)	146 (98%)	3 (2%)	55	51
1	D	149/164 (91%)	147 (99%)	2 (1%)	69	68
2	F	7/7 (100%)	7 (100%)	0	100	100
2	G	7/7 (100%)	6 (86%)	1 (14%)	3	1
2	H	7/7 (100%)	7 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	619/677 (91%)	608 (98%)	11 (2%)	59 55

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1165	LYS
1	D	1127	LEU
2	G	3	CYS
1	D	1165	LYS
1	B	1117	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	1034	GLN
1	D	1027	ASN
1	D	1034	GLN
1	B	1034	GLN
1	A	1027	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 ⓘ

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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	#$|Z| > 2$
3	SO4	C	3	-	4,4,4	0.21	0	6,6,6	0.25	0
3	SO4	A	1	-	4,4,4	0.21	0	6,6,6	0.31	0
3	SO4	B	2	-	4,4,4	0.13	0	6,6,6	0.19	0
3	SO4	D	4	-	4,4,4	0.15	0	6,6,6	0.45	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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.