Full wwPDB X-ray Structure Validation Report ⓘ

May 23, 2022 – 01:13 pm BST

PDB ID	:	7ZB6
Title	:	Crystal Structure of SARS-CoV-2 Main Protease (Mpro) variant C44S at 2.12 A resolution
Authors	:	Paknia, E.; Rabe von Pappenheim, F.; Funk, L.-M.; Tittmann, K.; Chari, A.
Deposited on	:	2022-03-23
Resolution	:	2.12 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/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:

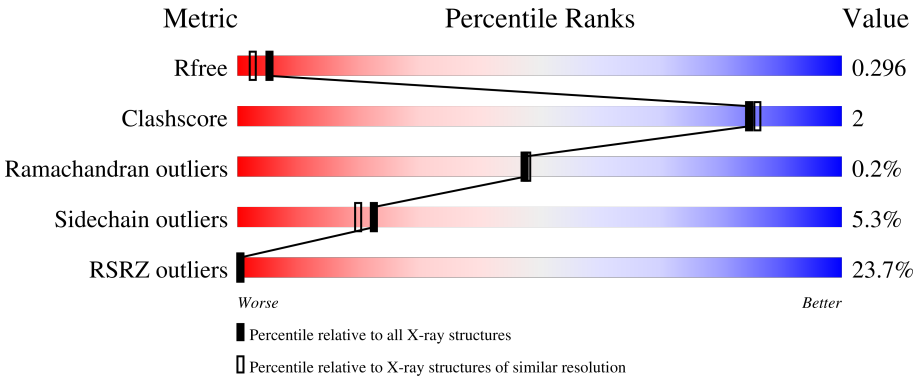
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.12 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	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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	306	24% 89% 10%
1	C	306	23% 91% 8%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9466 atoms, of which 4636 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.

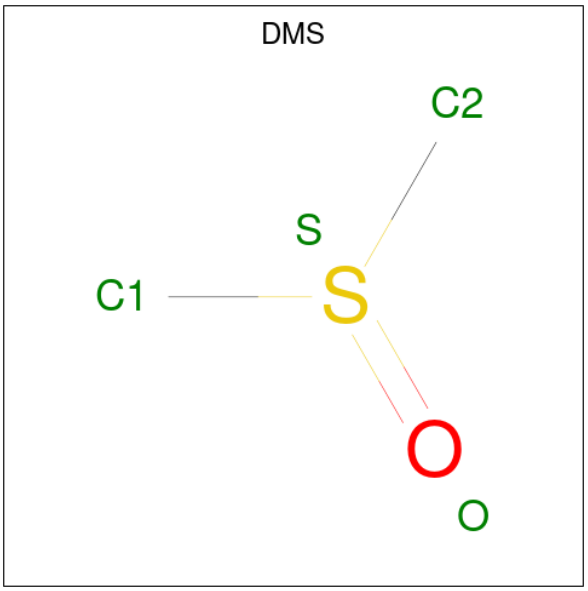
- Molecule 1 is a protein called 3C-like proteinase nsp5.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	306	Total	C	H	N	O	S	0	2	0
			4696	1507	2318	402	447	22			
1	C	306	Total	C	H	N	O	S	0	2	0
			4687	1504	2312	403	446	22			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	SER	CYS	engineered mutation	UNP P0DTD1
C	44	SER	CYS	engineered mutation	UNP P0DTD1

- Molecule 2 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	H	O	S	0	0
			10	2	6	1	1		

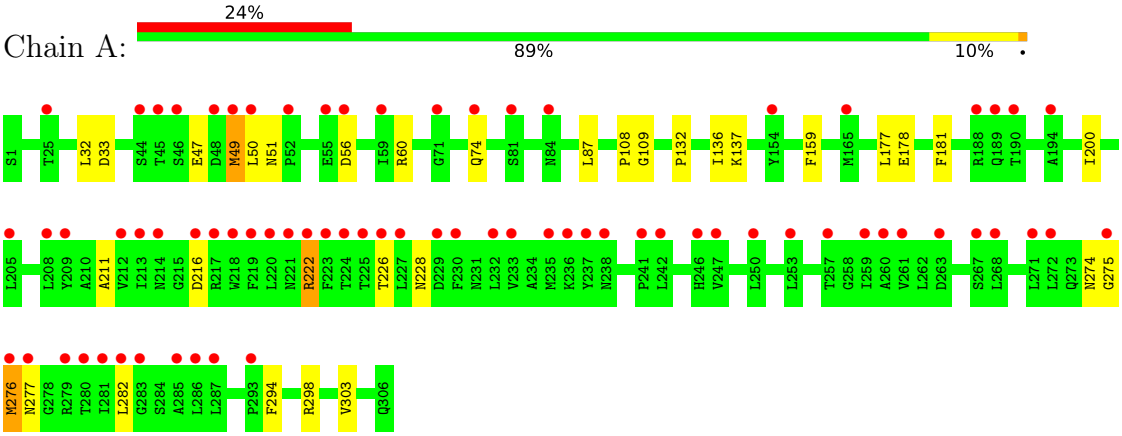
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	34	Total O 34 34	0	0
3	C	39	Total O 39 39	0	0

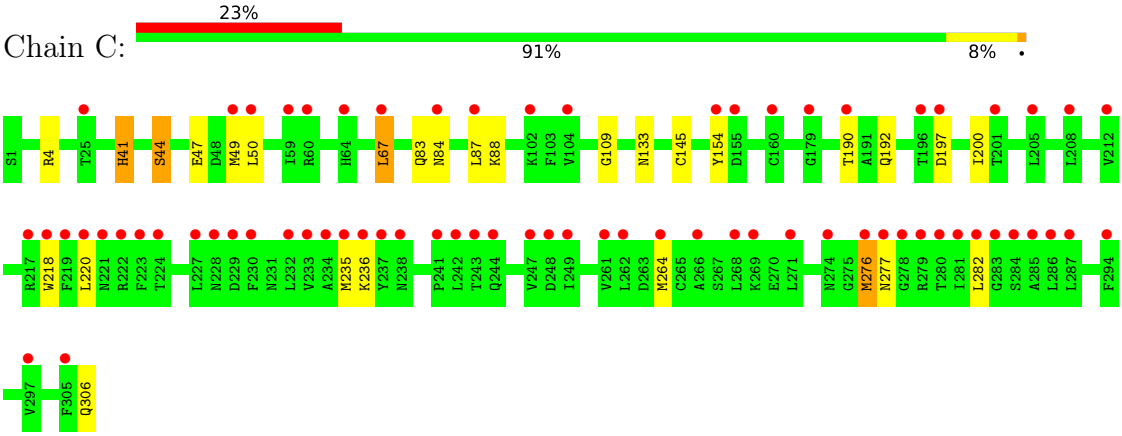
3 Residue-property plots (i)

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: 3C-like proteinase nsp5



• Molecule 1: 3C-like proteinase nsp5



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	122.88Å 81.36Å 63.56Å 90.00° 90.04° 90.00°	Depositor
Resolution (Å)	46.39 – 2.12 67.84 – 2.12	Depositor EDS
% Data completeness (in resolution range)	63.5 (46.39-2.12) 63.5 (67.84-2.12)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.53 (at 2.12Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.261 , 0.298 0.260 , 0.296	Depositor DCC
R_{free} test set	1153 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	47.6	Xtriage
Anisotropy	0.027	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.020 for -h,-k,l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	9466	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.38% 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

5.1 Standard geometry

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

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	0.26	0/2437	0.50	0/3311
1	C	0.31	0/2433	0.53	1/3306 (0.0%)
All	All	0.29	0/4870	0.52	1/6617 (0.0%)

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	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	41	HIS	N-CA-CB	7.35	123.83	110.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	222	ARG	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	A	2378	2318	2329	14	0
1	C	2375	2312	2318	9	1
2	C	4	6	6	0	0
3	A	34	0	0	1	0
3	C	39	0	0	1	0
All	All	4830	4636	4653	22	1

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 2.

All (22) 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:C:44:SER:OG	1:C:49:MET:CE	2.51	0.59
1:A:74:GLN:NE2	3:A:402:HOH:O	2.35	0.59
1:A:276:MET:SD	1:A:276:MET:N	2.73	0.58
1:C:276:MET:SD	1:C:277:ASN:N	2.75	0.55
1:A:274:ASN:OD1	1:A:275:GLY:N	2.43	0.51
1:C:44:SER:OG	1:C:49:MET:SD	2.70	0.49
1:A:211:ALA:O	1:A:216:ASP:N	2.47	0.48
1:A:49:MET:O	1:A:50:LEU:HB3	2.13	0.48
1:A:51:ASN:ND2	1:A:51:ASN:O	2.48	0.47
1:A:109:GLY:HA2	1:A:200:ILE:HD13	1.98	0.46
1:C:133:ASN:ND2	1:C:197:ASP:OD1	2.49	0.45
1:C:67:LEU:O	3:C:501:HOH:O	2.20	0.45
1:A:137:LYS:O	1:C:4:ARG:NE	2.40	0.45
1:C:109:GLY:HA2	1:C:200:ILE:HD13	2.00	0.43
1:A:108:PRO:HB3	1:A:132:PRO:HA	2.00	0.43
1:A:32:LEU:O	1:A:33:ASP:C	2.57	0.43
1:C:83:GLN:HB2	1:C:88:LYS:HE2	2.01	0.43
1:A:276:MET:HG2	1:A:277:ASN:H	1.84	0.42
1:A:159:PHE:HB3	1:A:177:LEU:HD13	2.02	0.42
1:C:190:THR:HG23	1:C:192:GLN:HG3	2.03	0.40
1:A:49:MET:SD	1:A:49:MET:N	2.95	0.40
1:A:298:ARG:HG2	1:A:303:VAL:HB	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:145:CYS:SG	1:C:306:GLN:C[4_545]	1.79	0.41

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	306/306 (100%)	279 (91%)	27 (9%)	0	100 100
1	C	306/306 (100%)	281 (92%)	24 (8%)	1 (0%)	41 40
All	All	612/612 (100%)	560 (92%)	51 (8%)	1 (0%)	47 48

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	41	HIS

5.3.2 Protein sidechains (i)

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	265/263 (101%)	251 (95%)	14 (5%)	22 20
1	C	264/263 (100%)	250 (95%)	14 (5%)	22 20
All	All	529/526 (101%)	501 (95%)	28 (5%)	22 20

All (28) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	47	GLU
1	A	49	MET
1	A	56	ASP
1	A	60	ARG
1	A	87	LEU
1	A	136	ILE
1	A	178	GLU
1	A	181	PHE
1	A	222	ARG
1	A	226	THR
1	A	228	ASN
1	A	276	MET
1	A	282	LEU
1	A	294	PHE
1	C	44	SER
1	C	47	GLU
1	C	50	LEU
1	C	67	LEU
1	C	84	ASN
1	C	87	LEU
1	C	154	TYR
1	C	218	TRP
1	C	220	LEU
1	C	235	MET
1	C	236	LYS
1	C	264	MET
1	C	276	MET
1	C	282	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

1 ligand is 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	DMS	C	401	-	3,3,3	0.56	0	3,3,3	0.70	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

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^{th} 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	306/306 (100%)	1.53	74 (24%)	0	0	27, 59, 138, 204	0
1	C	306/306 (100%)	1.55	71 (23%)	0	0	28, 59, 128, 189	0
All	All	612/612 (100%)	1.54	145 (23%)	0	0	27, 59, 135, 204	0

All (145) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	232	LEU	14.6
1	A	45	THR	11.9
1	C	278	GLY	9.8
1	A	221	ASN	9.7
1	C	217	ARG	9.7
1	A	271	LEU	9.2
1	C	244	GLN	9.2
1	C	234	ALA	9.1
1	A	222	ARG	9.1
1	C	220	LEU	9.1
1	C	268	LEU	8.9
1	C	219	PHE	8.8
1	A	217	ARG	8.8
1	C	222	ARG	8.3
1	A	50	LEU	8.1
1	A	242	LEU	7.6
1	C	281	ILE	7.6
1	A	230	PHE	7.3
1	C	285	ALA	7.3
1	C	227	LEU	7.1
1	C	84	ASN	7.0
1	C	286	LEU	7.0
1	A	46	SER	6.8
1	A	49	MET	6.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	242	LEU	6.7
1	C	208	LEU	6.5
1	A	237	TYR	6.5
1	A	220	LEU	6.4
1	C	249	ILE	6.3
1	A	219	PHE	6.3
1	C	221	ASN	6.3
1	C	230	PHE	6.2
1	A	272	LEU	6.2
1	A	84	ASN	6.0
1	A	259	ILE	5.9
1	C	205	LEU	5.9
1	C	154	TYR	5.7
1	A	268	LEU	5.5
1	C	283	GLY	5.5
1	C	218	TRP	5.4
1	A	209	TYR	5.4
1	A	216	ASP	5.3
1	A	253	LEU	5.3
1	C	262	LEU	5.3
1	A	283	GLY	5.2
1	C	271	LEU	5.1
1	A	44	SER	5.0
1	C	223	PHE	4.8
1	A	285	ALA	4.8
1	A	232	LEU	4.8
1	C	277	ASN	4.8
1	A	281	ILE	4.7
1	A	260	ALA	4.7
1	A	236	LYS	4.7
1	C	264	MET	4.7
1	C	274	ASN	4.6
1	A	218	TRP	4.5
1	A	279	ARG	4.4
1	A	214	ASN	4.4
1	A	52	PRO	4.4
1	A	154	TYR	4.4
1	C	261	VAL	4.2
1	A	55[A]	GLU	4.2
1	C	282	LEU	4.2
1	A	276	MET	4.1
1	A	226	THR	4.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	59	ILE	4.1
1	C	279	ARG	4.0
1	C	276	MET	3.9
1	C	236	LYS	3.9
1	A	208	LEU	3.9
1	C	280	THR	3.8
1	C	233	VAL	3.8
1	A	257	THR	3.7
1	C	269	LYS	3.6
1	A	223	PHE	3.6
1	A	189	GLN	3.6
1	A	205	LEU	3.6
1	C	237	TYR	3.6
1	C	287	LEU	3.6
1	A	56	ASP	3.6
1	C	235	MET	3.6
1	A	233	VAL	3.5
1	C	284	SER	3.5
1	A	250	LEU	3.5
1	A	213	ILE	3.4
1	A	227	LEU	3.4
1	A	282	LEU	3.4
1	A	261	VAL	3.4
1	C	266	ALA	3.3
1	A	247	VAL	3.2
1	A	225	THR	3.2
1	C	196	THR	3.1
1	A	246	HIS	3.0
1	A	48	ASP	3.0
1	A	277	ASN	3.0
1	A	286	LEU	3.0
1	A	238	ASN	3.0
1	C	229	ASP	2.9
1	C	190	THR	2.9
1	C	155	ASP	2.9
1	C	64	HIS	2.9
1	A	71	GLY	2.8
1	A	212	VAL	2.8
1	C	87	LEU	2.8
1	C	49	MET	2.8
1	A	229	ASP	2.7
1	C	60	ARG	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	212	VAL	2.7
1	A	165[A]	MET	2.7
1	A	188	ARG	2.6
1	C	179	GLY	2.6
1	A	25	THR	2.6
1	A	275	GLY	2.5
1	C	224	THR	2.5
1	A	241	PRO	2.5
1	A	267	SER	2.4
1	A	280	THR	2.4
1	C	294	PHE	2.4
1	C	297	VAL	2.3
1	A	81	SER	2.3
1	C	238	ASN	2.3
1	C	247	VAL	2.3
1	A	235	MET	2.3
1	A	287	LEU	2.3
1	C	305	PHE	2.3
1	A	293	PRO	2.3
1	C	241	PRO	2.3
1	C	228	ASN	2.2
1	C	201	THR	2.2
1	C	102	LYS	2.2
1	C	59	ILE	2.2
1	A	190	THR	2.1
1	A	194	ALA	2.1
1	C	243	THR	2.1
1	A	263	ASP	2.1
1	C	25	THR	2.1
1	C	197	ASP	2.1
1	C	67	LEU	2.1
1	C	160	CYS	2.1
1	A	224	THR	2.1
1	A	74	GLN	2.0
1	C	248	ASP	2.0
1	C	104	VAL	2.0
1	C	50	LEU	2.0

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 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, 95^{th} 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	DMS	C	401	4/4	0.67	0.18	98,118,120,120	0

6.5 Other polymers (i)

There are no such residues in this entry.