Full wwPDB X-ray Structure Validation Report ⓘ

Aug 12, 2025 – 02:22 PM JST

PDB ID	:	9KGQ / pdb_00009kgq
Title	:	Discovery of an orally bioavailable reversible covalent SARS-CoV-2 Mpro inhibitor with pan-coronavirus activity
Authors	:	Baburajendran, N.
Deposited on	:	2024-11-08
Resolution	:	1.50 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 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:

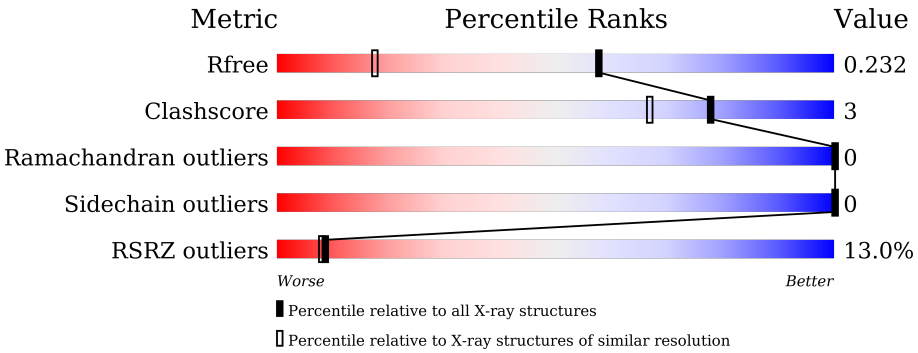
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.45.1

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 1.50 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}	164625	3717 (1.50-1.50)
Clashscore	180529	4048 (1.50-1.50)
Ramachandran outliers	177936	3970 (1.50-1.50)
Sidechain outliers	177891	3967 (1.50-1.50)
RSRZ outliers	164620	3718 (1.50-1.50)

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	311	15%, 88%, 7%, 5%
1	B	311	10%, 90%, 7%, .
2	C	4	75%, 25%
2	D	4	75%, 25%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4967 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 3C-like proteinase nsp5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	302	Total	C	N	O	S	0	0	0
			2313	1466	393	432	22			
1	A	296	Total	C	N	O	S	0	0	0
			2270	1442	381	426	21			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	GLY	-	expression tag	UNP P0DTD1
B	-3	PRO	-	expression tag	UNP P0DTD1
B	-2	LEU	-	expression tag	UNP P0DTD1
B	-1	GLY	-	expression tag	UNP P0DTD1
B	0	SER	-	expression tag	UNP P0DTD1
A	-4	GLY	-	expression tag	UNP P0DTD1
A	-3	PRO	-	expression tag	UNP P0DTD1
A	-2	LEU	-	expression tag	UNP P0DTD1
A	-1	GLY	-	expression tag	UNP P0DTD1
A	0	SER	-	expression tag	UNP P0DTD1

- Molecule 2 is a protein called compound 4.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	4	Total	C	N	O	0	0	0
			29	20	4	5			
2	D	4	Total	C	N	O	0	0	0
			29	20	4	5			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	191	Total	O	0	0
			191	191		

Continued on next page...

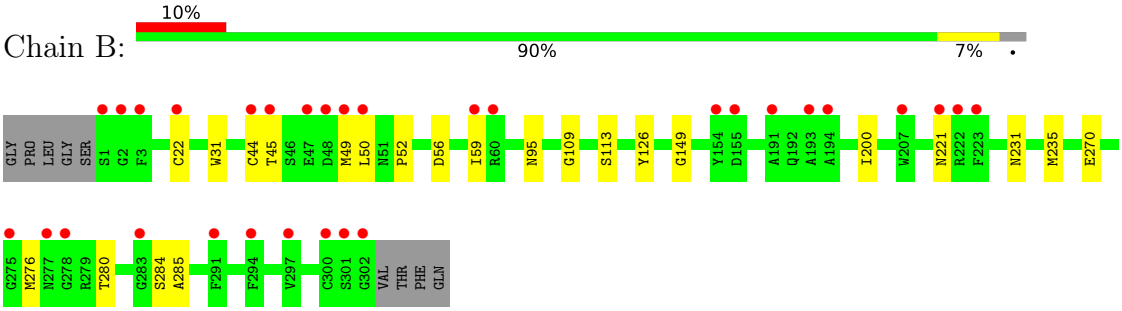
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	131	Total O 131 131	0	0
3	C	2	Total O 2 2	0	0
3	D	2	Total O 2 2	0	0

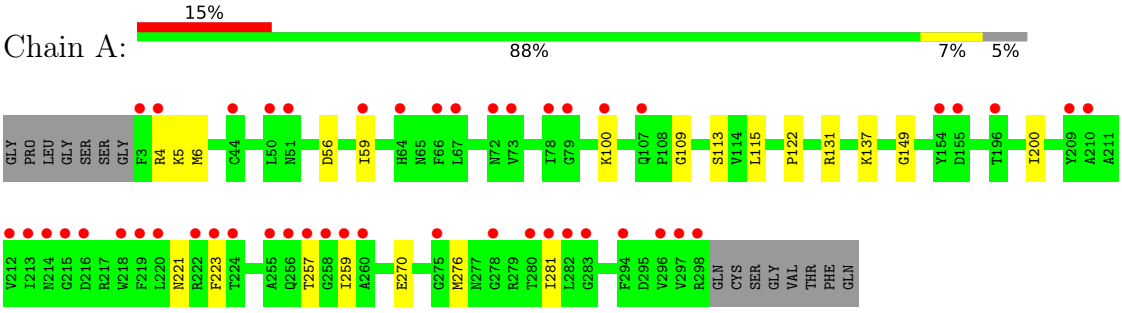
3 Residue-property plots (i)

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.

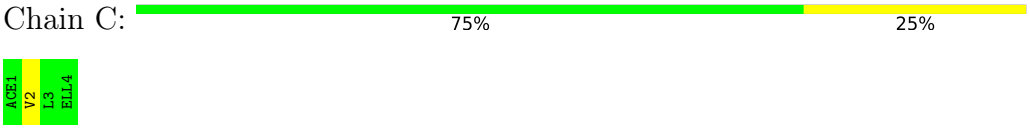
• Molecule 1: 3C-like proteinase nsp5



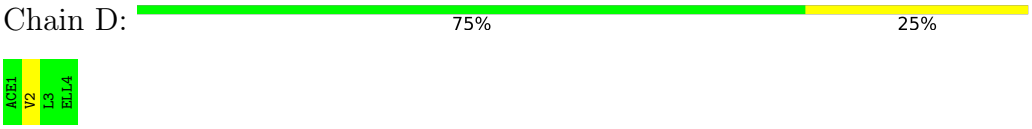
• Molecule 1: 3C-like proteinase nsp5



• Molecule 2: compound 4



• Molecule 2: compound 4



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.66Å 99.29Å 59.93Å 90.00° 108.66° 90.00°	Depositor
Resolution (Å)	46.57 – 1.50 46.57 – 1.50	Depositor EDS
% Data completeness (in resolution range)	98.0 (46.57-1.50) 98.3 (46.57-1.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 1.50Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.200 , 0.233 0.201 , 0.232	Depositor DCC
R_{free} test set	2008 reflections (2.03%)	wwPDB-VP
Wilson B-factor (Å^2)	23.8	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 29.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4967	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.15% 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: ELL, DVA, ACE

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	0.52	0/2321	0.74	2/3159 (0.1%)
1	B	0.54	1/2365 (0.0%)	0.73	0/3217
2	C	0.81	0/7	1.31	0/8
2	D	0.86	0/7	0.98	0/8
All	All	0.53	1/4700 (0.0%)	0.73	2/6392 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	22	CYS	CB-SG	-5.01	1.64	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	5	LYS	CD-CE-NZ	5.11	128.24	111.90
1	A	4	ARG	CG-CD-NE	-5.10	100.78	112.00

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	2270	0	2206	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2313	0	2250	14	0
2	C	29	0	23	0	0
2	D	29	0	22	0	0
3	A	131	0	0	2	0
3	B	191	0	0	1	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
All	All	4967	0	4501	25	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 3.

All (25) 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:A:109:GLY:HA2	1:A:200:ILE:HD13	1.76	0.66
1:A:257:THR:HG23	1:A:259:ILE:H	1.65	0.62
1:B:49:MET:HA	1:B:52:PRO:HG3	1.86	0.56
1:A:100:LYS:HE3	3:A:498:HOH:O	2.04	0.56
1:B:280:THR:OG1	1:B:284:SER:N	2.35	0.56
1:B:56:ASP:HA	1:B:59:ILE:HD12	1.89	0.54
1:A:131:ARG:HD3	1:A:137:LYS:HE2	1.90	0.53
1:B:280:THR:HG23	3:B:496:HOH:O	2.09	0.53
1:B:44:CYS:HB3	1:B:49:MET:HG2	1.92	0.51
1:B:221:ASN:HD22	1:B:270:GLU:HG3	1.75	0.51
1:A:56:ASP:O	1:A:59:ILE:HG22	2.10	0.50
1:A:276:MET:HE3	1:A:281:ILE:HD12	1.94	0.49
1:B:50:LEU:HG	1:B:50:LEU:O	2.15	0.47
1:B:231:ASN:O	1:B:235:MET:HG3	2.14	0.47
1:A:100:LYS:HE2	3:A:505:HOH:O	2.15	0.46
1:B:276:MET:HE3	1:B:285:ALA:O	2.16	0.46
1:B:109:GLY:HA2	1:B:200:ILE:HD13	1.97	0.46
1:B:31:TRP:CE2	1:B:95:ASN:HB2	2.51	0.45
1:A:115:LEU:HD11	1:A:122:PRO:HB3	2.00	0.44
1:B:113:SER:O	1:B:149:GLY:HA2	2.19	0.42
1:A:221:ASN:ND2	1:A:270:GLU:HG3	2.35	0.41
1:A:113:SER:O	1:A:149:GLY:HA2	2.20	0.41
1:B:45:THR:O	1:B:49:MET:HG3	2.19	0.41
1:A:223:PHE:CD2	1:A:223:PHE:N	2.89	0.40
1:B:126:TYR:CD1	1:A:6:MET:HE2	2.57	0.40

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	294/311 (94%)	287 (98%)	7 (2%)	0	100	100
1	B	300/311 (96%)	295 (98%)	5 (2%)	0	100	100
2	C	1/4 (25%)	1 (100%)	0	0	100	100
2	D	1/4 (25%)	1 (100%)	0	0	100	100
All	All	596/630 (95%)	584 (98%)	12 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	250/266 (94%)	250 (100%)	0	100	100
1	B	254/266 (96%)	254 (100%)	0	100	100
2	C	1/1 (100%)	1 (100%)	0	100	100
2	D	1/1 (100%)	1 (100%)	0	100	100
All	All	506/534 (95%)	506 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	274	ASN
1	A	65	ASN
1	A	274	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

2 non-standard protein/DNA/RNA residues 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	DVA	D	2	2	4,6,7	0.66	0	6,7,9	1.24	1 (16%)
2	DVA	C	2	2	4,6,7	0.66	0	6,7,9	1.17	1 (16%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	DVA	D	2	2	-	0/5/6/8	-
2	DVA	C	2	2	-	0/5/6/8	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	2	DVA	CB-CA-C	-2.94	108.97	112.94
2	C	2	DVA	CB-CA-C	-2.74	109.25	112.94

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	296/311 (95%)	0.98	47 (15%) 6 5	18, 32, 53, 70	0
1	B	302/311 (97%)	0.62	31 (10%) 13 13	18, 27, 51, 66	0
2	C	1/4 (25%)	0.91	0 100 100	29, 29, 29, 29	0
2	D	1/4 (25%)	0.43	0 100 100	27, 27, 27, 27	0
All	All	600/630 (95%)	0.80	78 (13%) 9 8	18, 29, 53, 70	0

All (78) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	223	PHE	6.5
1	A	3	PHE	5.0
1	B	302	GLY	5.0
1	B	3	PHE	4.8
1	B	154	TYR	4.5
1	A	212	VAL	4.2
1	B	50	LEU	4.2
1	A	294	PHE	4.2
1	A	282	LEU	4.1
1	A	213	ILE	3.9
1	A	258	GLY	3.9
1	B	223	PHE	3.9
1	B	48	ASP	3.5
1	A	298	ARG	3.4
1	B	294	PHE	3.3
1	A	224	THR	3.3
1	B	59	ILE	3.2
1	B	193	ALA	3.1
1	B	222	ARG	3.1
1	A	218	TRP	3.1
1	A	259	ILE	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	222	ARG	3.1
1	B	300	CYS	3.1
1	A	220	LEU	3.0
1	A	64	HIS	3.0
1	B	2	GLY	3.0
1	A	219	PHE	3.0
1	A	154	TYR	2.9
1	A	196	THR	2.9
1	A	257	THR	2.9
1	A	73	VAL	2.8
1	B	301	SER	2.8
1	A	216	ASP	2.8
1	A	210	ALA	2.8
1	B	60	ARG	2.8
1	A	50	LEU	2.8
1	B	49	MET	2.7
1	A	255	ALA	2.7
1	A	59	ILE	2.7
1	A	214	ASN	2.6
1	A	215	GLY	2.5
1	B	277	ASN	2.5
1	B	155	ASP	2.5
1	A	72	ASN	2.5
1	A	297	VAL	2.5
1	B	221	ASN	2.5
1	A	4	ARG	2.4
1	B	47	GLU	2.4
1	B	194	ALA	2.4
1	B	1	SER	2.4
1	A	78	ILE	2.4
1	A	67	LEU	2.4
1	A	283	GLY	2.4
1	A	66	PHE	2.4
1	A	275	GLY	2.4
1	B	191	ALA	2.4
1	A	260	ALA	2.4
1	B	45	THR	2.3
1	A	278	GLY	2.3
1	A	100	LYS	2.3
1	B	297	VAL	2.3
1	A	280	THR	2.3
1	A	209	TYR	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	283	GLY	2.3
1	A	79	GLY	2.3
1	A	281	ILE	2.2
1	B	22	CYS	2.2
1	B	44	CYS	2.2
1	A	44	CYS	2.2
1	A	256	GLN	2.2
1	A	51	ASN	2.2
1	B	278	GLY	2.1
1	B	291	PHE	2.1
1	B	275	GLY	2.1
1	A	107	GLN	2.1
1	A	155	ASP	2.0
1	B	207	TRP	2.0
1	A	296	VAL	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (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	DVA	D	2	7/8	0.94	0.11	25,26,32,37	0
2	DVA	C	2	7/8	0.95	0.08	28,29,32,34	0

6.3 Carbohydrates (i)

There are no oligosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.