Full wwPDB X-ray Structure Validation Report (i)

Jul 21, 2025 – 04:09 PM JST

PDB ID	:	9JCX / pdb_00009jcx
Title	:	Crystal structure of the HCoV-HKU1 RBD and TMPRSS2
Authors	:	Wang, H.; Li, M.; Duan, Y.; Yang, H.
Deposited on	:	2024-08-30
Resolution	:	2.75 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 (i) 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 (1)) were used in the production of this report:

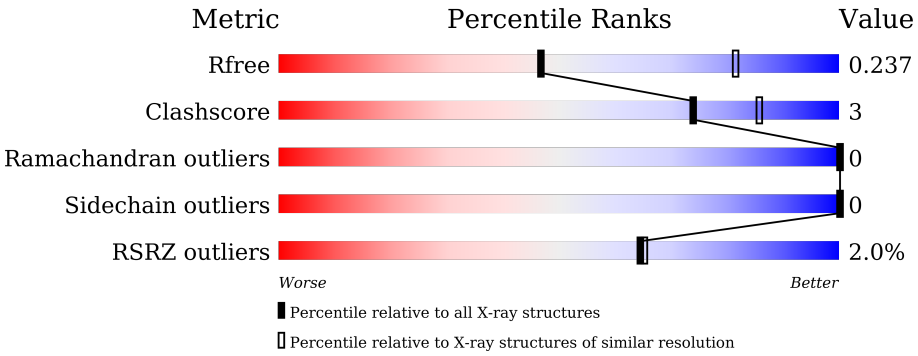
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.44

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 2.75 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	1606 (2.78-2.74)
Clashscore	180529	1689 (2.78-2.74)
Ramachandran outliers	177936	1665 (2.78-2.74)
Sidechain outliers	177891	1665 (2.78-2.74)
RSRZ outliers	164620	1606 (2.78-2.74)

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	146	8%, 88%, 7%, 5%
2	C	249	90%, 6%, .
3	B	308	85%, 8%, 7%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5156 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 Transmembrane protease serine 2 non-catalytic chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	138	Total	C	N	O	S	0	0	0
			1025	625	185	202	13			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	250	ASP	SER	engineered mutation	UNP O15393
A	251	ASP	SER	engineered mutation	UNP O15393
A	252	ASP	ARG	engineered mutation	UNP O15393
A	253	ASP	GLN	engineered mutation	UNP O15393
A	254	LYS	SER	engineered mutation	UNP O15393

- Molecule 2 is a protein called Transmembrane protease serine 2 catalytic chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	238	Total	C	N	O	S	0	1	0
			1840	1177	312	336	15			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	441	ALA	SER	engineered mutation	UNP O15393
C	493	GLU	-	expression tag	UNP O15393
C	494	PHE	-	expression tag	UNP O15393
C	495	VAL	-	expression tag	UNP O15393
C	496	GLU	-	expression tag	UNP O15393
C	497	HIS	-	expression tag	UNP O15393
C	498	HIS	-	expression tag	UNP O15393
C	499	HIS	-	expression tag	UNP O15393
C	500	HIS	-	expression tag	UNP O15393
C	501	HIS	-	expression tag	UNP O15393
C	502	HIS	-	expression tag	UNP O15393

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	503	HIS	-	expression tag	UNP O15393
C	504	HIS	-	expression tag	UNP O15393

- Molecule 3 is a protein called Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	286	Total	C	N	O	S	0	2	0
			2224	1385	379	436	24			

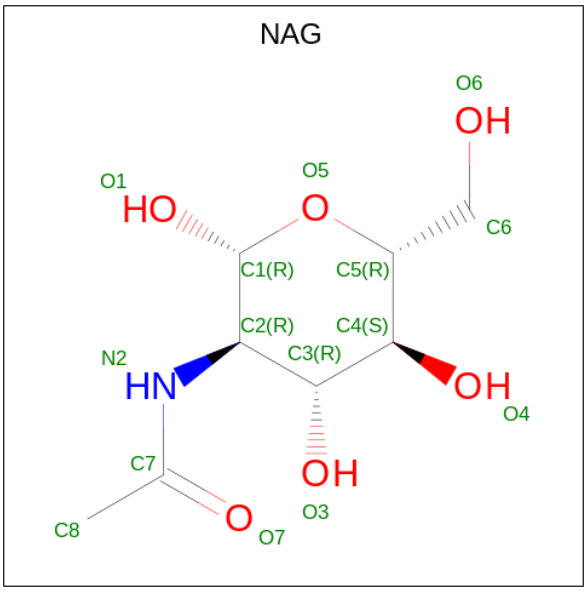
There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	610	SER	-	expression tag	UNP Q5MQD0
B	611	GLY	-	expression tag	UNP Q5MQD0
B	612	LEU	-	expression tag	UNP Q5MQD0
B	613	GLU	-	expression tag	UNP Q5MQD0
B	614	VAL	-	expression tag	UNP Q5MQD0
B	615	LEU	-	expression tag	UNP Q5MQD0
B	616	PHE	-	expression tag	UNP Q5MQD0
B	617	GLN	-	expression tag	UNP Q5MQD0
B	618	GLY	-	expression tag	UNP Q5MQD0
B	619	PRO	-	expression tag	UNP Q5MQD0
B	620	GLY	-	expression tag	UNP Q5MQD0
B	621	GLY	-	expression tag	UNP Q5MQD0
B	622	SER	-	expression tag	UNP Q5MQD0
B	623	HIS	-	expression tag	UNP Q5MQD0
B	624	HIS	-	expression tag	UNP Q5MQD0
B	625	HIS	-	expression tag	UNP Q5MQD0
B	626	HIS	-	expression tag	UNP Q5MQD0
B	627	HIS	-	expression tag	UNP Q5MQD0
B	628	HIS	-	expression tag	UNP Q5MQD0
B	629	HIS	-	expression tag	UNP Q5MQD0
B	630	HIS	-	expression tag	UNP Q5MQD0

- Molecule 4 is CALCIUM ION (CCD ID: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Ca	0	0
			1	1		

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (CCD ID: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C N O 14 8 1 5	0	0
5	B	1	Total C N O 14 8 1 5	0	0
5	B	1	Total C N O 14 8 1 5	0	0

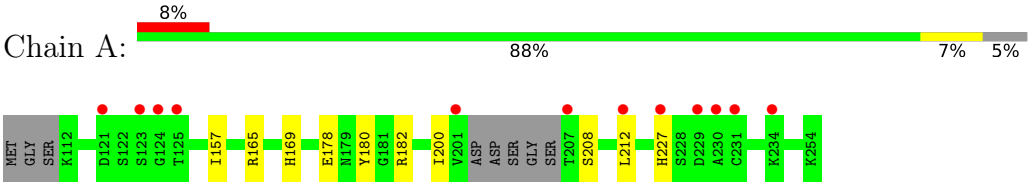
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	4	Total O 4 4	0	0
6	C	8	Total O 8 8	0	0
6	B	12	Total O 12 12	0	0

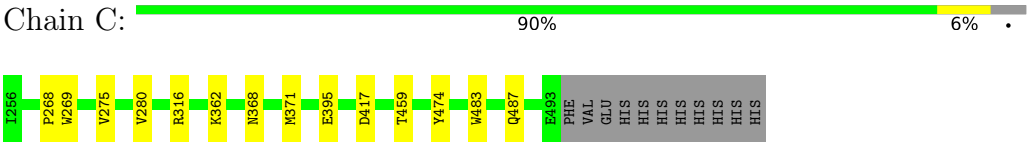
3 Residue-property plots (i)

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.

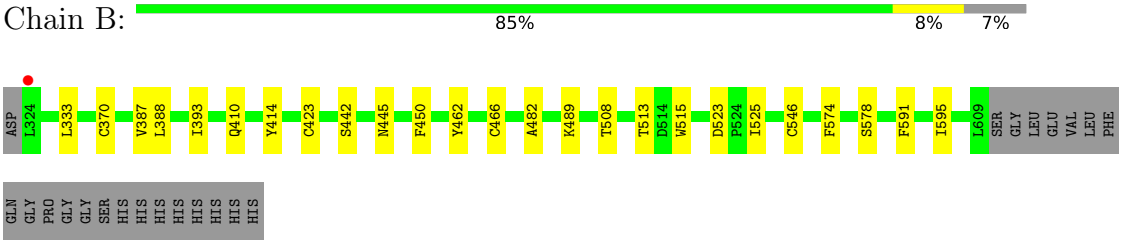
• Molecule 1: Transmembrane protease serine 2 non-catalytic chain



• Molecule 2: Transmembrane protease serine 2 catalytic chain



• Molecule 3: Spike protein S1



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	63.35Å 120.72Å 126.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.95 – 2.75 19.95 – 2.75	Depositor EDS
% Data completeness (in resolution range)	99.9 (19.95-2.75) 99.6 (19.95-2.75)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.63 (at 2.75Å)	Xtriage
Refinement program	PHENIX (1.21.2_5419: ???)	Depositor
R, R_{free}	0.205 , 0.237 0.205 , 0.237	Depositor DCC
R_{free} test set	1296 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	53.2	Xtriage
Anisotropy	0.720	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 39.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.024 for -h,l,k	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5156	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.12% 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: NAG, CA

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	0.11	0/1045	0.30	0/1418
2	C	0.11	0/1894	0.28	0/2590
3	B	0.11	0/2287	0.29	0/3111
All	All	0.11	0/5226	0.29	0/7119

There are no bond length outliers.

There are no bond angle outliers.

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	1025	0	893	9	0
2	C	1840	0	1755	9	0
3	B	2224	0	2064	13	0
4	A	1	0	0	0	0
5	A	14	0	13	0	0
5	B	28	0	26	0	0
6	A	4	0	0	0	0
6	B	12	0	0	0	0
6	C	8	0	0	0	0
All	All	5156	0	4751	30	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 (30) 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)
3:B:387:VAL:HB	3:B:595:ILE:HB	1.81	0.61
1:A:180:TYR:CD2	1:A:212:LEU:HD21	2.39	0.58
2:C:316[B]:ARG:HB3	2:C:395:GLU:OE2	2.06	0.55
3:B:442:SER:HB3	3:B:445:ASN:HB2	1.88	0.54
3:B:508:THR:HG22	3:B:513:THR:HA	1.90	0.52
2:C:368:ASN:O	2:C:371:MET:HG2	2.10	0.52
3:B:466[B]:CYS:HB3	3:B:546:CYS:SG	2.51	0.49
3:B:333:LEU:HD21	3:B:388:LEU:HD12	1.95	0.49
2:C:417:ASP:H	3:B:515:TRP:HH2	1.61	0.48
1:A:157:ILE:HD11	1:A:200:ILE:HD13	1.96	0.48
2:C:316[A]:ARG:HB3	2:C:395:GLU:OE2	2.12	0.48
1:A:178:GLU:O	1:A:182:ARG:HG3	2.16	0.46
2:C:459:THR:HA	2:C:474:TYR:CD2	2.51	0.46
3:B:370:CYS:HA	3:B:423:CYS:HA	1.97	0.46
1:A:169:HIS:CE1	1:A:208:SER:HB2	2.51	0.45
3:B:462:TYR:CZ	3:B:578:SER:HB3	2.53	0.44
3:B:482:ALA:O	3:B:489:LYS:HD3	2.17	0.44
1:A:165:ARG:HD2	1:A:227:HIS:CG	2.53	0.43
2:C:275:VAL:HG12	2:C:280:VAL:HG21	1.99	0.43
1:A:180:TYR:CE2	1:A:212:LEU:HD21	2.54	0.42
3:B:393:ILE:HD11	3:B:591:PHE:HB2	2.00	0.42
3:B:450:PHE:HB2	3:B:574:PHE:CZ	2.55	0.42
1:A:180:TYR:HD2	1:A:212:LEU:HD21	1.83	0.42
2:C:268:PRO:HD2	2:C:269:TRP:CE3	2.55	0.41
2:C:268:PRO:HB2	2:C:362:LYS:N	2.36	0.41
3:B:523:ASP:OD1	3:B:525:ILE:HB	2.20	0.41
1:A:165:ARG:HD2	1:A:227:HIS:CD2	2.56	0.41
1:A:208:SER:HB3	1:A:227:HIS:NE2	2.36	0.41
2:C:483:TRP:O	2:C:487:GLN:HG2	2.21	0.40
3:B:410:GLN:HA	3:B:414:TYR:O	2.22	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	134/146 (92%)	122 (91%)	12 (9%)	0	100	100
2	C	237/249 (95%)	225 (95%)	12 (5%)	0	100	100
3	B	286/308 (93%)	278 (97%)	8 (3%)	0	100	100
All	All	657/703 (94%)	625 (95%)	32 (5%)	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	106/127 (84%)	106 (100%)	0	100	100
2	C	194/209 (93%)	194 (100%)	0	100	100
3	B	264/282 (94%)	264 (100%)	0	100	100
All	All	564/618 (91%)	564 (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 (6) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	327	GLN
2	C	398	ASN
3	B	471	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	488	HIS
3	B	589	ASN
3	B	597	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 oligosaccharides in this entry.

5.6 Ligand geometry

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	$\#\|Z\| > 2$
5	NAG	A	302	1	14,14,15	0.70	0	17,19,21	0.90	0
5	NAG	B	702	3	14,14,15	0.76	0	17,19,21	1.03	1 (5%)
5	NAG	B	701	3	14,14,15	0.74	0	17,19,21	0.90	0

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
5	NAG	A	302	1	-	0/6/23/26	0/1/1/1
5	NAG	B	702	3	-	2/6/23/26	0/1/1/1
5	NAG	B	701	3	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	B	702	NAG	C2-N2-C7	2.19	126.02	122.90

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	B	702	NAG	O5-C5-C6-O6
5	B	701	NAG	C4-C5-C6-O6
5	B	701	NAG	O5-C5-C6-O6
5	B	702	NAG	C3-C2-N2-C7

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

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	138/146 (94%)	0.60	12 (8%) 17 20	44, 74, 109, 123	0
2	C	238/249 (95%)	-0.19	0 100 100	27, 50, 66, 95	1 (0%)
3	B	286/308 (92%)	-0.06	1 (0%) 90 91	31, 53, 81, 97	2 (0%)
All	All	662/703 (94%)	0.03	13 (1%) 64 65	27, 54, 95, 123	3 (0%)

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	124	GLY	4.2
3	B	324	LEU	3.9
1	A	229	ASP	3.4
1	A	201	VAL	3.3
1	A	125	THR	2.9
1	A	212	LEU	2.4
1	A	234	LYS	2.4
1	A	123	SER	2.4
1	A	207	THR	2.3
1	A	227	HIS	2.2
1	A	230	ALA	2.2
1	A	121	ASP	2.1
1	A	231	CYS	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

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
5	NAG	B	702	14/15	0.55	0.17	82,97,105,106	0
5	NAG	B	701	14/15	0.64	0.13	68,89,98,98	0
5	NAG	A	302	14/15	0.72	0.12	77,97,113,117	0
4	CA	A	301	1/1	0.98	0.06	61,61,61,61	0

6.5 Other polymers

There are no such residues in this entry.