



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 5, 2024 – 12:32 pm GMT

PDB ID : 9F2G
Title : Crystal structure of SARS-CoV-2 N-protein C-terminal domain (apo form)
Authors : Marquez-Monino, M.A.; Gonzalez, B.; Perez-Canadillas, J.M.
Deposited on : 2024-04-23
Resolution : 1.57 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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.003 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.39

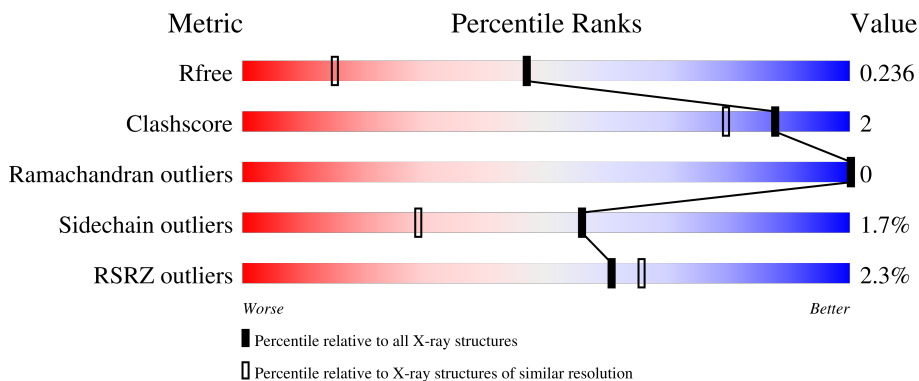
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.57 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.



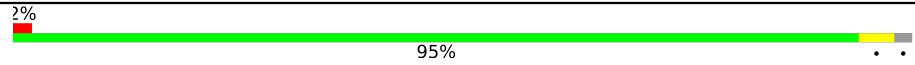
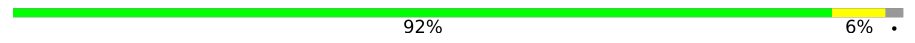

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	7165 (1.60-1.56)
Clashscore	180529	1026 (1.58-1.58)
Ramachandran outliers	177936	1005 (1.58-1.58)
Sidechain outliers	177891	1004 (1.58-1.58)
RSRZ outliers	164620	7163 (1.60-1.56)

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	A	111	 3% 94% 2% 1%
1	B	111	 2% 90% 8% 1%
1	C	111	 2% 92% 6% 1%
1	D	111	 3% 92% 6% 1%
1	E	111	 6% 89% 7% 1%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	111	 <p>2% 95% 6% . .</p>
1	G	111	 <p>92% 6% .</p>
1	H	111	 <p>% 88% 9% .</p>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8310 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 Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	108	881	559	156	163	3	0	2	0
1	B	109	915	581	164	166	4	0	5	0
1	C	109	904	575	160	165	4	0	4	0
1	D	109	921	585	167	166	3	0	6	0
1	E	108	903	575	161	163	4	0	5	0
1	F	109	901	571	162	165	3	0	3	0
1	G	109	904	575	160	165	4	0	4	0
1	H	108	886	562	156	165	3	0	3	0

There are 16 discrepancies between the modelled and reference sequences:

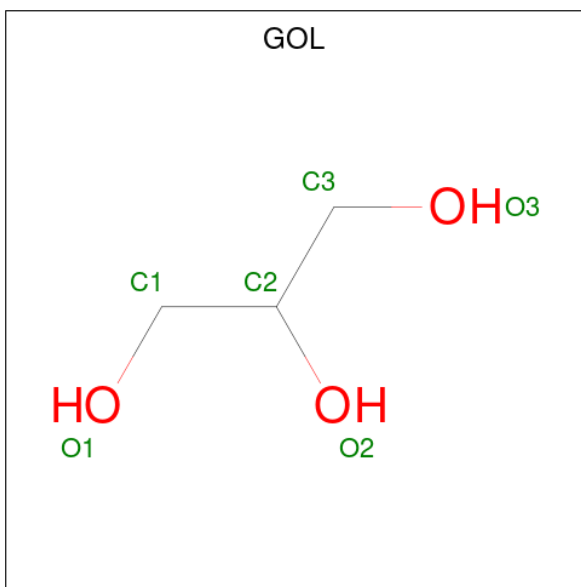
Chain	Residue	Modelled	Actual	Comment	Reference
A	254	GLY	-	expression tag	UNP P0DTC9
A	255	SER	-	expression tag	UNP P0DTC9
B	254	GLY	-	expression tag	UNP P0DTC9
B	255	SER	-	expression tag	UNP P0DTC9
C	254	GLY	-	expression tag	UNP P0DTC9
C	255	SER	-	expression tag	UNP P0DTC9
D	254	GLY	-	expression tag	UNP P0DTC9
D	255	SER	-	expression tag	UNP P0DTC9
E	254	GLY	-	expression tag	UNP P0DTC9
E	255	SER	-	expression tag	UNP P0DTC9
F	254	GLY	-	expression tag	UNP P0DTC9
F	255	SER	-	expression tag	UNP P0DTC9
G	254	GLY	-	expression tag	UNP P0DTC9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	255	SER	-	expression tag	UNP P0DTC9
H	254	GLY	-	expression tag	UNP P0DTC9
H	255	SER	-	expression tag	UNP P0DTC9

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	D	1	Total	C	O	0	0
			6	3	3		
2	D	1	Total	C	O	0	0
			6	3	3		
2	E	1	Total	C	O	0	0
			6	3	3		
2	G	1	Total	C	O	0	0
			6	3	3		

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



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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	116	Total	O	0	0
			116	116		
4	B	139	Total	O	0	0
			139	139		
4	C	148	Total	O	0	0
			148	148		
4	D	150	Total	O	0	0
			150	150		
4	E	94	Total	O	0	0
			94	94		
4	F	139	Total	O	0	0
			139	139		
4	G	116	Total	O	0	0
			116	116		

Continued on next page...

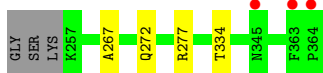
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	H	138	Total 138	O 138	0	0

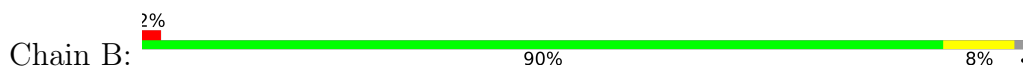
3 Residue-property plots [i](#)

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.

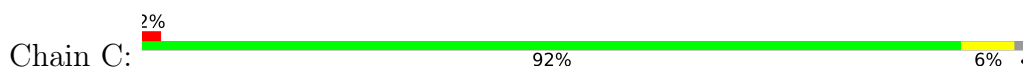
- Molecule 1: Nucleoprotein



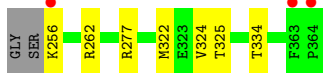
- Molecule 1: Nucleoprotein



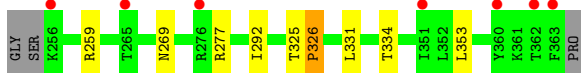
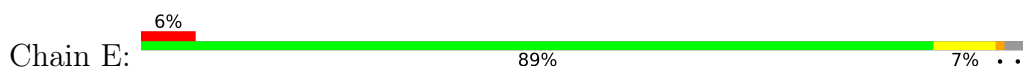
- Molecule 1: Nucleoprotein



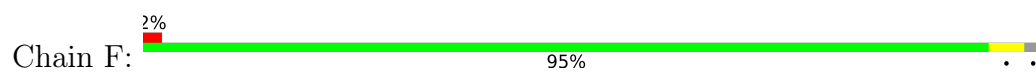
- Molecule 1: Nucleoprotein



- Molecule 1: Nucleoprotein



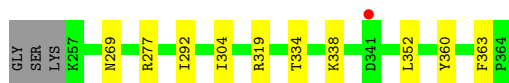
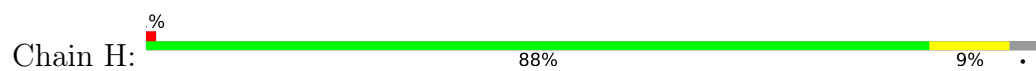
- Molecule 1: Nucleoprotein



• Molecule 1: Nucleoprotein



• Molecule 1: Nucleoprotein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	92.70Å 45.14Å 109.06Å 90.00° 93.96° 90.00°	Depositor
Resolution (Å)	48.42 – 1.57 48.42 – 1.57	Depositor EDS
% Data completeness (in resolution range)	97.4 (48.42-1.57) 97.4 (48.42-1.57)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.74 (at 1.57Å)	Xtrriage
Refinement program	REFMAC 5.8.0411	Depositor
R, R_{free}	0.189 , 0.229 0.200 , 0.236	Depositor DCC
R_{free} test set	6517 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å ²)	23.8	Xtrriage
Anisotropy	0.282	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 32.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	8310	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 30.99 % 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 1.1919e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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, GOL

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	0.42	0/904	0.77	0/1220
1	B	0.41	0/941	0.79	0/1266
1	C	0.44	0/930	0.81	0/1252
1	D	0.43	0/953	0.79	0/1282
1	E	0.39	0/931	0.71	0/1254
1	F	0.40	0/924	0.73	0/1245
1	G	0.38	0/930	0.72	0/1252
1	H	0.44	0/912	0.77	0/1231
All	All	0.41	0/7425	0.76	0/10002

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	E	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	259	ARG	Sidechain

5.2 Too-close contacts [i](#)

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	881	0	860	4	0
1	B	915	0	906	5	0
1	C	904	0	894	5	0
1	D	921	0	919	5	0
1	E	903	0	898	5	0
1	F	901	0	885	2	0
1	G	904	0	894	6	0
1	H	886	0	864	5	0
2	A	6	0	8	0	0
2	D	12	0	16	1	0
2	E	6	0	8	0	0
2	G	6	0	8	1	0
3	D	5	0	0	0	0
3	F	20	0	0	0	0
4	A	116	0	0	0	0
4	B	139	0	0	0	0
4	C	148	0	0	0	0
4	D	150	0	0	0	0
4	E	94	0	0	0	0
4	F	139	0	0	0	0
4	G	116	0	0	0	0
4	H	138	0	0	0	0
All	All	8310	0	7160	28	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 2.

The worst 5 of 28 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:334[B]:THR:HG23	1:D:334[B]:THR:HG22	1.60	0.82
1:H:319:ARG:HB2	1:H:334[A]:THR:HG22	1.67	0.76
1:D:262:ARG:HH22	2:D:403:GOL:H2	1.61	0.65
1:C:282:THR:HG22	1:C:282:THR:O	1.98	0.62
1:A:267:ALA:HB2	1:G:334[B]:THR:CG2	2.31	0.60

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	108/111 (97%)	107 (99%)	1 (1%)	0	100	100
1	B	112/111 (101%)	108 (96%)	4 (4%)	0	100	100
1	C	111/111 (100%)	109 (98%)	2 (2%)	0	100	100
1	D	113/111 (102%)	113 (100%)	0	0	100	100
1	E	111/111 (100%)	109 (98%)	2 (2%)	0	100	100
1	F	110/111 (99%)	107 (97%)	3 (3%)	0	100	100
1	G	111/111 (100%)	110 (99%)	1 (1%)	0	100	100
1	H	109/111 (98%)	108 (99%)	1 (1%)	0	100	100
All	All	885/888 (100%)	871 (98%)	14 (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	92/92 (100%)	91 (99%)	1 (1%)	70	50
1	B	96/92 (104%)	95 (99%)	1 (1%)	73	55
1	C	95/92 (103%)	92 (97%)	3 (3%)	34	7
1	D	97/92 (105%)	95 (98%)	2 (2%)	48	20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	95/92 (103%)	93 (98%)	2 (2%)	48	20
1	F	94/92 (102%)	92 (98%)	2 (2%)	48	20
1	G	95/92 (103%)	94 (99%)	1 (1%)	70	50
1	H	93/92 (101%)	92 (99%)	1 (1%)	70	50
All	All	757/736 (103%)	744 (98%)	13 (2%)	56	29

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	277	ARG
1	E	326	PRO
1	H	277	ARG
1	F	341	ASP
1	G	277	ARG

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

10 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
2	GOL	E	401	-	5,5,5	0.11	0	5,5,5	0.31	0
3	SO4	F	403	-	4,4,4	0.38	0	6,6,6	0.06	0
3	SO4	F	402	-	4,4,4	0.36	0	6,6,6	0.07	0
2	GOL	G	401	-	5,5,5	0.12	0	5,5,5	0.43	0
3	SO4	D	401	-	4,4,4	0.43	0	6,6,6	0.09	0
3	SO4	F	401	-	4,4,4	0.36	0	6,6,6	0.06	0
2	GOL	D	402	-	5,5,5	0.20	0	5,5,5	0.63	0
2	GOL	A	401	-	5,5,5	0.11	0	5,5,5	0.27	0
3	SO4	F	404	-	4,4,4	0.46	0	6,6,6	0.23	0
2	GOL	D	403	-	5,5,5	0.11	0	5,5,5	0.28	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
2	GOL	E	401	-	-	2/4/4/4	-
2	GOL	G	401	-	-	2/4/4/4	-
2	GOL	D	402	-	-	4/4/4/4	-
2	GOL	A	401	-	-	2/4/4/4	-
2	GOL	D	403	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	401	GOL	C1-C2-C3-O3
2	D	402	GOL	C1-C2-C3-O3
2	E	401	GOL	C1-C2-C3-O3
2	D	402	GOL	O1-C1-C2-C3
2	D	403	GOL	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	G	401	GOL	1	0
2	D	403	GOL	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	A	108/111 (97%)	0.16	3 (2%) 55 59	8, 21, 39, 73	2 (1%)
1	B	109/111 (98%)	0.10	2 (1%) 67 72	9, 23, 37, 66	5 (4%)
1	C	109/111 (98%)	0.04	2 (1%) 67 72	8, 21, 36, 61	4 (3%)
1	D	109/111 (98%)	0.15	3 (2%) 55 59	9, 20, 34, 49	6 (5%)
1	E	108/111 (97%)	0.90	7 (6%) 26 28	11, 29, 50, 65	5 (4%)
1	F	109/111 (98%)	0.33	2 (1%) 67 72	11, 24, 40, 62	3 (2%)
1	G	109/111 (98%)	0.14	0 100 100	9, 25, 40, 64	4 (3%)
1	H	108/111 (97%)	0.12	1 (0%) 81 84	12, 23, 34, 52	3 (2%)
All	All	869/888 (97%)	0.24	20 (2%) 61 66	8, 23, 41, 73	32 (3%)

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	364	PRO	5.6
1	C	282	THR	5.3
1	F	256	LYS	4.5
1	E	363	PHE	3.5
1	D	364	PRO	3.2

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(Å ²)	Q<0.9
3	SO4	F	402	5/5	0.62	0.13	83,86,91,92	0
3	SO4	F	403	5/5	0.67	0.12	99,101,104,107	0
2	GOL	A	401	6/6	0.70	0.21	50,52,54,62	0
3	SO4	F	401	5/5	0.71	0.14	87,88,95,95	0
2	GOL	D	403	6/6	0.75	0.15	58,63,64,65	0
2	GOL	E	401	6/6	0.76	0.14	49,50,55,58	0
2	GOL	G	401	6/6	0.77	0.16	52,54,55,57	0
2	GOL	D	402	6/6	0.80	0.18	29,41,48,49	0
3	SO4	D	401	5/5	0.91	0.11	29,31,35,35	5
3	SO4	F	404	5/5	0.91	0.12	33,36,38,40	5

6.5 Other polymers [i](#)

There are no such residues in this entry.