WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 11, 2023 – 11:28 AM EDT

PDB ID	:	7RNW
Title	:	SARS-CoV-2 Main Protease in complex with a cyclic peptide inhibitor
Authors	:	Frkic, R.L.; Jackson, C.J.
Deposited on	:	2021-07-30
Resolution	:	2.35 Å(reported)

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

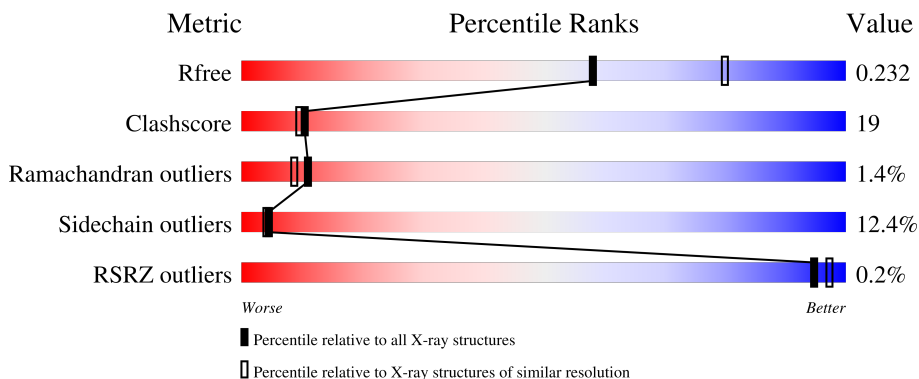|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.35 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		0.232
Clashscore		19
Ramachandran outliers		1.4%
Sidechain outliers		12.4%
RSRZ outliers		0.2%

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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	60% 33% 6% ·
1	B	306	60% 36% · ·
1	C	306	63% 33% · ·
1	D	306	% 54% 39% 6% ·
2	W	15	27% 20% 27% 27%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	X	15	7% 33% 13% 7% 40%
2	Y	15	7% 7% 40% 20% 7% 27%
2	Z	15	33% 33% 20% 7% 7%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	DTY	Z	1	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9819 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	303	Total 2340	C 1481	N 398	O 439	S 22	0	0	0
1	B	304	Total 2347	C 1485	N 399	O 441	S 22	0	0	0
1	C	304	Total 2347	C 1485	N 399	O 441	S 22	0	0	0
1	D	303	Total 2340	C 1481	N 398	O 439	S 22	0	0	0

- Molecule 2 is a protein called ACE-DTY-LEU-GLN-TYR-ALA-VAL-LEU-ARG-HIS-LYS-ARG-ARG-GLU-SEC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	9	Total 72	C 47	N 10	O 14	Se 1	0	0	0
2	Y	11	Total 101	C 64	N 19	O 17	Se 1	0	1	0
2	Z	14	Total 129	C 83	N 25	O 20	Se 1	0	1	0
2	W	11	Total 96	C 60	N 19	O 16	Se 1	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	10	Total 10	O 10	0	0
3	B	8	Total 8	O 8	0	0
3	C	8	Total 8	O 8	0	0
3	D	19	Total 19	O 19	0	0

Continued on next page...

Continued from previous page...

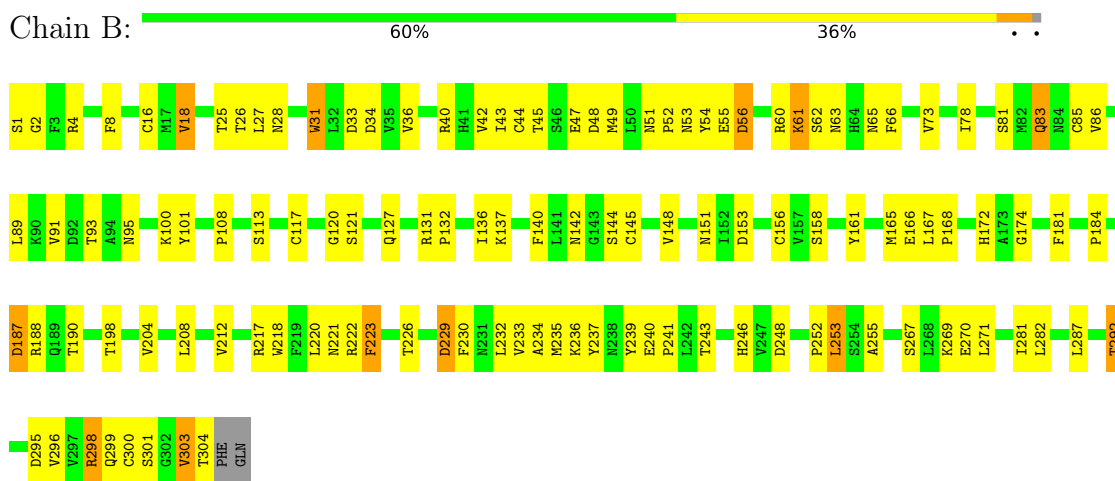| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	Z	1	Total 1	O 1	0	0
3	W	1	Total 1	O 1	0	0

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 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.
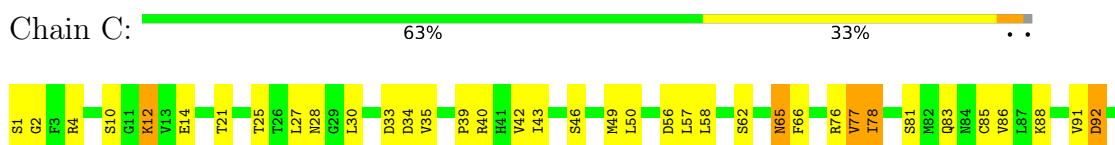
- Molecule 1: 3C-like proteinase

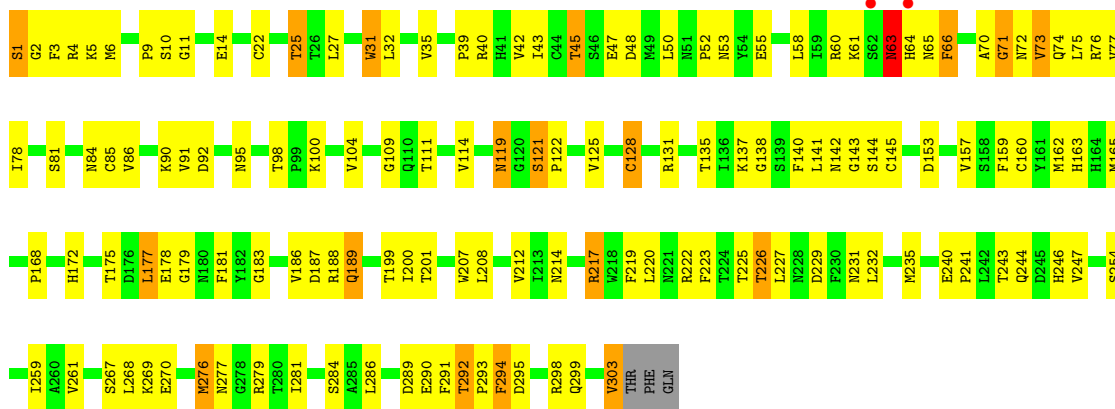Chain A:

- Molecule 1: 3C-like proteinase

Chain B:

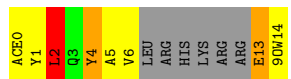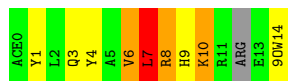- Molecule 1: 3C-like proteinase

Chain C:

N95 T98 P99 K100 I106 Q107 P108 F112 L115 A116 C117 G120 S121 P122 S123 G124 V125 Y126 Q127 C128 A129 M130 R131 P132 N133 F134 T135 I136 K137 G138 S139 F140 V148 Y154 D155 C156 F159 C160 Y161 H164 M165 E166 T169 H172 A173 L177 P184

F185 V186 D187 R188 Q189 T190 T198 N203 V204 W207 L208 Y209 N214 R217 L220 F223 T224 T225 L232 K236 T243 Q244 D245 I249 P252 V261 L262 L271 L272 Q273 N274 T280 I281 L282 L287 E288 D289 T292 P293 V296 V297 R298

Q299 C300 T304 PHE GLN

• Molecule 1: 3C-like proteinase

Chain D:

%

54%   39%   6%   •

S1 G2 F3 R4 K5 M6 P9 S10 G11 E14 C22 T25 T26 L27 W31 L32 V35 P39 R40 H41 V42 I43 C44 T45 S46 E47 D48 M49 L50 N51 P52 N53 Y54 E55 L58 I59 R60 K61 S62 N63 H64 N65 F66 A70 G71 N72 V73 Q74 L75 R76 V77

I78 S81 N84 C85 V86 K90 V91 D92 N95 T98 P99 K100 V104 G109 Q110 T111 V114 N119 G120 S121 P122 V125 C128 R131 T135 I136 K137 G138 S139 F140 L141 N142 G143 S144 C145 D153 V157 S158 F159 C160 Y161 M162 H163 H164 M165

P168 H172 T175 D176 L177 E178 G179 N180 F181 Y182 G183 V186 D187 R188 Q189 T199 I200 T201 W207 L208 V212 I213 N214 R217 W218 F219 L220 N221 R222 F223 T224 T225 T226 L227 N228 D229 N231 L232 M235 E240 P241 L242 T243 Q244 D245 H246 V247 S254

I259 A260 V261 S267 L268 K269 E270 M276 N277 G278 R279 T280 I281 S284 A285 L286 D289 E290 F291 T292 P293 F294 D295 R298 Q299 V303 THR PHE GLN

• Molecule 2: ACE-DTY-LEU-GLN-TYR-ALA-VAL-LEU-ARG-HIS-LYS-ARG-ARG-GLU-SE
C

Chain X: 7%   33%   13%   7%   40%

ACE0 Y1 L2 Q3 Y4 A5 V6 LEU ARG HIS LYS ARG ARG E13 9OW14

• Molecule 2: ACE-DTY-LEU-GLN-TYR-ALA-VAL-LEU-ARG-HIS-LYS-ARG-ARG-GLU-SE
C

Chain Y:

7%

7%   40%   20%   7%   27%

ACE0 Y1 L2 Q3 Y4 A5 V6 LEU ARG HIS LYS R11 R12 E13 9OW14

• Molecule 2: ACE-DTY-LEU-GLN-TYR-ALA-VAL-LEU-ARG-HIS-LYS-ARG-ARG-GLU-SE
C

Chain Z: 33%   33%   20%   7% 7%

ACE0 Y1 L2 Q3 Y4 A5 V6 L7 R8 H9 K10 R11 ARG E13 9OW14

• Molecule 2: ACE-DTY-LEU-GLN-TYR-ALA-VAL-LEU-ARG-HIS-LYS-ARG-ARG-GLU-SEC

Chain W:

27% 20% 27% 27%
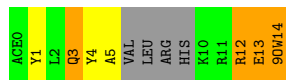
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	48.36Å 201.73Å 60.54Å 90.00° 113.49° 90.00°	Depositor
Resolution (Å)	33.62 – 2.35 33.62 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.6 (33.62-2.35) 99.6 (33.62-2.35)	Depositor EDS
R_{merge}	0.25	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.171 , 0.231 0.173 , 0.232	Depositor DCC
R_{free} test set	2306 reflections (5.24%)	wwPDB-VP
Wilson B-factor (Å^2)	32.0	Xtriage
Anisotropy	1.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 19.7	EDS
L-test for twinning[2]	$< |L| > = 0.39, < L^2 > = 0.22$	Xtriage
Estimated twinning fraction	0.438 for h,-k,-h-l	Xtriage
Reported twinning fraction	0.508 for H, K, L 0.492 for -H, -K, H+L	Depositor
Outliers	2 of 44016 reflections (0.005%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9819	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.16% 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: 9OW, ACE, DTY

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	0.70	0/2392	0.90	0/3251
1	B	0.68	0/2399	0.90	0/3261
1	C	0.71	0/2399	0.93	0/3261
1	D	0.71	0/2392	0.93	0/3251
2	W	0.74	0/73	1.24	0/93
2	X	0.67	0/49	1.04	0/64
2	Y	0.75	0/77	1.26	1/99 (1.0%)
2	Z	0.70	0/107	1.22	0/140
All	All	0.70	0/9888	0.93	1/13420 (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	D	0	1
2	W	0	4
2	X	0	1
2	Y	0	1
2	Z	0	2
All	All	0	9

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	Y	4	TYR	CB-CA-C	5.79	121.97	110.40

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	183	GLY	Peptide
2	X	13	GLU	Mainchain
2	Y	13	GLU	Mainchain
2	Z	7	LEU	Peptide
2	Z	8	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	2340	0	2290	87	0
1	B	2347	0	2297	90	0
1	C	2347	0	2297	83	0
1	D	2340	0	2290	102	0
2	W	96	0	88	10	0
2	X	72	0	57	9	0
2	Y	101	0	93	11	0
2	Z	129	0	125	17	0
3	A	10	0	0	1	0
3	B	8	0	0	2	0
3	C	8	0	0	0	0
3	D	19	0	0	2	0
3	W	1	0	0	0	0
3	Z	1	0	0	3	0
All	All	9819	0	9537	359	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 19.

The worst 5 of 359 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:165:MET:HE3	2:Z:1:DTY:HE1	1.28	1.12
1:D:247:VAL:HG13	1:D:261:VAL:HG11	1.43	0.98
2:Z:6:VAL:HG11	2:Z:9:HIS:CE1	2.05	0.91

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1:SER:OG	1:B:166:GLU:OE2	1.89	0.90
1:B:165:MET:HE3	2:Z:1:DTY:CE1	2.04	0.88

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	301/306 (98%)	280 (93%)	16 (5%)	5 (2%)	9	7
1	B	302/306 (99%)	284 (94%)	16 (5%)	2 (1%)	22	23
1	C	302/306 (99%)	286 (95%)	14 (5%)	2 (1%)	22	23
1	D	301/306 (98%)	280 (93%)	19 (6%)	2 (1%)	22	23
2	W	6/15 (40%)	3 (50%)	3 (50%)	0	100	100
2	X	4/15 (27%)	2 (50%)	0	2 (50%)	0	0
2	Y	6/15 (40%)	3 (50%)	1 (17%)	2 (33%)	0	0
2	Z	10/15 (67%)	5 (50%)	3 (30%)	2 (20%)	0	0
All	All	1232/1284 (96%)	1143 (93%)	72 (6%)	17 (1%)	11	9

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	61	LYS
1	D	71	GLY
2	Y	4	TYR
1	A	14	GLU
1	A	302	GLY

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	260/263 (99%)	228 (88%)	32 (12%)	4	4
1	B	261/263 (99%)	233 (89%)	28 (11%)	6	6
1	C	261/263 (99%)	231 (88%)	30 (12%)	5	5
1	D	260/263 (99%)	223 (86%)	37 (14%)	3	3
2	W	7/11 (64%)	6 (86%)	1 (14%)	3	3
2	X	5/11 (46%)	3 (60%)	2 (40%)	0	0
2	Y	8/11 (73%)	5 (62%)	3 (38%)	0	0
2	Z	11/11 (100%)	10 (91%)	1 (9%)	9	8
All	All	1073/1096 (98%)	939 (88%)	134 (12%)	4	4

5 of 134 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	217	ARG
1	D	232	LEU
2	Y	6[B]	VAL
1	B	187	ASP
1	B	181	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	228	ASN
1	D	214	ASN
1	C	256	GLN
1	D	41	HIS
2	Z	3	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

8 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	9OW	Y	14	2	4,6,6	0.42	0	5,7,7	1.39	1 (20%)
2	9OW	W	14	2	4,6,6	0.45	0	5,7,7	2.90	3 (60%)
2	9OW	X	14	2	4,6,6	0.61	0	5,7,7	2.15	3 (60%)
2	9OW	Z	14	2	4,6,6	0.94	0	5,7,7	2.49	3 (60%)

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	9OW	Y	14	2	-	0/4/6/6	-
2	9OW	W	14	2	-	2/4/6/6	-
2	9OW	X	14	2	-	2/4/6/6	-
2	9OW	Z	14	2	-	2/4/6/6	-

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	W	14	9OW	CA-C-NB	4.68	124.71	116.68
2	Z	14	9OW	CA-C-NB	4.00	123.54	116.68
2	W	14	9OW	O-C-NB	-3.91	116.21	123.00
2	X	14	9OW	CA-C-NB	2.86	121.59	116.68
2	X	14	9OW	CB-CA-C	2.76	115.37	109.49

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	X	14	9OW	O-C-CA-N
2	X	14	9OW	NB-C-CA-N
2	Z	14	9OW	O-C-CA-CB
2	Z	14	9OW	NB-C-CA-CB
2	W	14	9OW	O-C-CA-CB

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	W	14	9OW	1	0

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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, 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	303/306 (99%)	-0.53	0	100	100	21, 39, 56, 79	0
1	B	304/306 (99%)	-0.53	0	100	100	23, 38, 56, 79	0
1	C	304/306 (99%)	-0.67	0	100	100	21, 32, 45, 78	0
1	D	303/306 (99%)	-0.57	2 (0%)	87	92	21, 34, 56, 107	0
2	W	8/15 (53%)	-0.46	0	100	100	29, 45, 50, 52	0
2	X	6/15 (40%)	-0.34	0	100	100	32, 39, 40, 40	0
2	Y	8/15 (53%)	0.19	1 (12%)	3	6	34, 44, 53, 53	0
2	Z	11/15 (73%)	-0.37	0	100	100	28, 38, 63, 76	0
All	All	1247/1284 (97%)	-0.57	3 (0%)	95	97	21, 36, 56, 107	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	62	SER	2.5
2	Y	6[A]	VAL	2.2
1	D	64	HIS	2.1

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

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
2	DTY	Y	1	12/13	0.92	0.15	53,60,64,70	0
2	DTY	Z	1	12/13	0.93	0.14	36,44,47,50	0
2	DTY	W	1	12/13	0.96	0.11	31,37,41,42	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	DTY	X	1	12/13	0.97	0.11	30,34,37,37	0
2	9OW	X	14	7/7	0.97	0.11	32,37,50,56	0
2	9OW	Z	14	7/7	0.97	0.07	42,49,58,71	0
2	9OW	Y	14	7/7	0.99	0.08	37,42,51,61	0
2	9OW	W	14	7/7	0.99	0.09	43,52,66,71	0

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.