



Full wwPDB X-ray Structure Validation Report ⓘ

Sep 25, 2017 – 02:06 PM EDT

PDB ID : 5VIV
Title : Crystal structure of monomeric near-infrared fluorescent protein miRFP670
Authors : Pletnev, S.
Deposited on : unknown
Resolution : 1.33 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

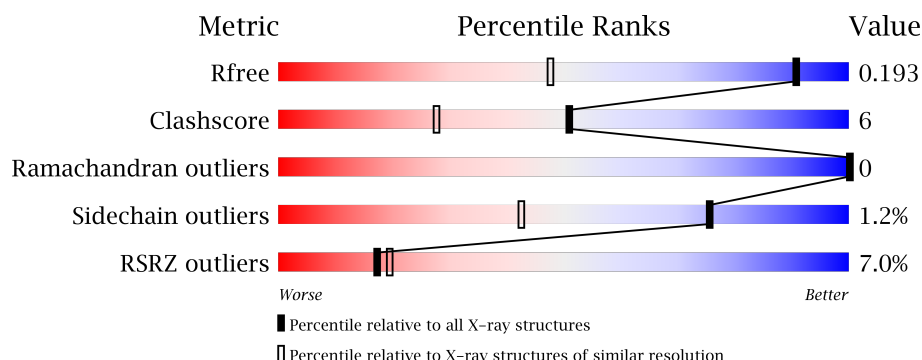
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.33 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}	100719	1990 (1.38-1.30)
Clashscore	112137	2078 (1.38-1.30)
Ramachandran outliers	110173	2021 (1.38-1.30)
Sidechain outliers	110143	2021 (1.38-1.30)
RSRZ outliers	101464	1993 (1.38-1.30)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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.

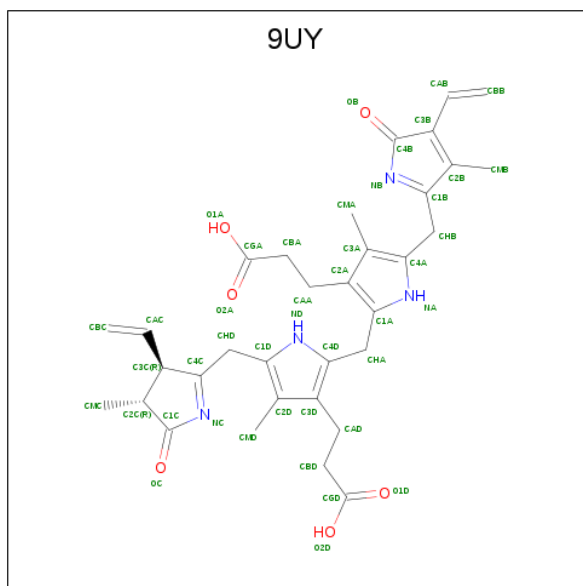
Mol	Chain	Length	Quality of chain
1	A	315	<div> <div>7%</div> <div>80%</div> <div>14%</div> <div>• •</div> </div>

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 monomeric near-infrared fluorescent protein miRFP670.

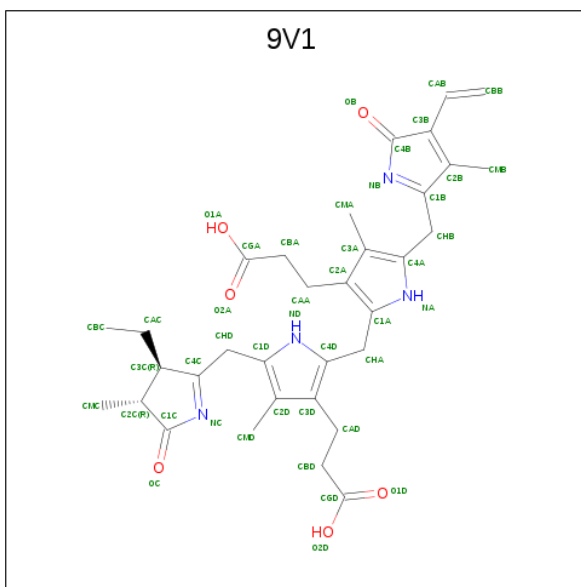
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	301	Total	C	N	O	S	0	18	0
			2476	1555	459	443	19			

- Molecule 2 is 3-[5-[[[3 {R},4 {R}]-3-ethenyl-4-methyl-5-oxidanylidene-3,4-dihydropyrrol-2-yl)methyl]-2-[[5-[4-ethenyl-3-methyl-5-oxidanylidene-pyrrol-2-yl)methyl]-3-(3-hydroxy-3-oxopropyl)-4-methyl-1 {H}-pyrrol-2-yl)methyl]-4-methyl-1 {H}-pyrrol-3-yl]propanoic acid (three-letter code: 9UY) (formula: C₃₃H₃₈N₄O₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			43	33	4	6		

- Molecule 3 is 3-[2-[[5-[(4-ethenyl-3-methyl-5-oxidanylidene-pyrrol-2-yl)methyl]-3-(3-hydroxy-3-oxopropyl)-4-methyl-1 {H}-pyrrol-2-yl)methyl]-5-[(3 {R},4 {R})-3-ethyl-4-methyl-5-oxidanylidene-3,4-dihydropyrrol-2-yl)methyl]-4-methyl-1 {H}-pyrrol-3-yl]propanoic acid (three-letter code: 9V1) (formula: C₃₃H₄₀N₄O₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			43	33	4	6		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total Cl 2 2	0	0

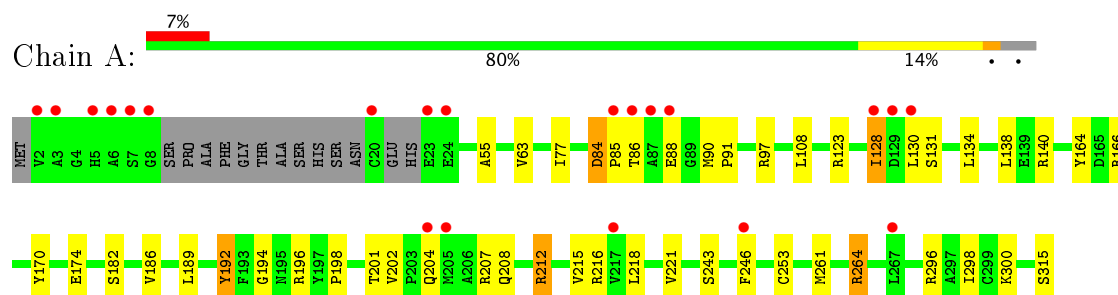
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	314	Total O 314 314	0	0

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: monomeric near-infrared fluorescent protein miRFP670



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	52.54Å 53.28Å 107.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.60 – 1.33 29.58 – 1.33	Depositor EDS
% Data completeness (in resolution range)	99.5 (29.60-1.33) 99.5 (29.58-1.33)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.96 (at 1.33Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.150 , 0.189 0.158 , 0.193	Depositor DCC
R_{free} test set	2105 reflections (3.12%)	DCC
Wilson B-factor (Å ²)	19.4	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 47.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.021 for k,h,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	2878	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.21% of the height of the origin peak. No significant pseudotranslation is detected.*

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: 9UY, 9V1, CL

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	1.15	8/2524 (0.3%)	1.08	13/3415 (0.4%)

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	164	TYR	CE1-CZ	-7.99	1.28	1.38
1	A	182	SER	CB-OG	-6.22	1.34	1.42
1	A	192	TYR	CE1-CZ	-6.04	1.30	1.38
1	A	85	PRO	N-CD	5.67	1.55	1.47
1	A	315	SER	C-O	5.38	1.33	1.23
1	A	243	SER	CA-CB	5.14	1.60	1.52
1	A	194	GLY	N-CA	-5.11	1.38	1.46
1	A	192	TYR	CG-CD2	-5.07	1.32	1.39

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	97	ARG	NE-CZ-NH2	10.05	125.33	120.30
1	A	296	ARG	NE-CZ-NH1	6.98	123.79	120.30
1	A	253[B]	CYS	CA-CB-SG	-6.93	101.52	114.00
1	A	253[C]	CYS	CA-CB-SG	-6.93	101.52	114.00
1	A	97	ARG	NE-CZ-NH1	-6.81	116.90	120.30
1	A	166[A]	ARG	NE-CZ-NH1	-6.12	117.24	120.30
1	A	166[B]	ARG	NE-CZ-NH1	-6.12	117.24	120.30
1	A	264[A]	ARG	NE-CZ-NH2	-5.51	117.55	120.30
1	A	264[B]	ARG	NE-CZ-NH2	-5.51	117.55	120.30
1	A	170	TYR	CB-CG-CD2	5.50	124.30	121.00
1	A	123[A]	ARG	NE-CZ-NH1	5.18	122.89	120.30
1	A	123[B]	ARG	NE-CZ-NH1	5.18	122.89	120.30
1	A	84	ASP	C-N-CD	5.02	138.93	128.40

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	2476	0	2485	33	2
2	A	43	0	0	1	0
3	A	43	0	0	0	0
4	A	2	0	0	1	0
5	A	314	0	0	8	2
All	All	2878	0	2485	33	2

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 6.

All (33) 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:212[B]:ARG:NH2	5:A:1001:HOH:O	1.73	1.08
1:A:212[B]:ARG:NH1	5:A:1002:HOH:O	2.06	0.88
1:A:174[B]:GLU:H	1:A:174[B]:GLU:CD	1.91	0.74
1:A:86:THR:HG22	1:A:88:GLU:H	1.53	0.73
1:A:77:ILE:HG12	4:A:904:CL:CL	2.29	0.69
1:A:84:ASP:HB3	5:A:1009:HOH:O	1.94	0.67
1:A:218:LEU:HD21	1:A:221[A]:VAL:HG22	1.81	0.63
1:A:140:ARG:NE	5:A:1005:HOH:O	2.33	0.61
1:A:207[A]:ARG:NH2	5:A:1004:HOH:O	2.33	0.61
1:A:215[A]:VAL:HG21	1:A:300:LYS:HG3	1.84	0.59
1:A:264[B]:ARG:HG3	1:A:264[B]:ARG:O	2.03	0.58
1:A:128:ILE:HD12	1:A:128:ILE:C	2.26	0.57
1:A:128:ILE:HD11	1:A:298:ILE:CD1	2.36	0.55
1:A:221[A]:VAL:HG23	1:A:264[A]:ARG:O	2.08	0.53
1:A:108:LEU:HD22	1:A:246[A]:PHE:CD2	2.46	0.50
1:A:198:PRO:HG2	1:A:201:THR:HG23	1.95	0.49
1:A:218:LEU:CD2	1:A:221[A]:VAL:HG22	2.45	0.47
1:A:186:VAL:HG22	1:A:189:LEU:CD1	2.45	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:186:VAL:HG22	1:A:189:LEU:HD11	1.97	0.46
1:A:215[A]:VAL:CG2	1:A:300:LYS:CG	2.93	0.46
1:A:215[B]:VAL:HG11	1:A:300:LYS:HG3	1.99	0.45
1:A:215[A]:VAL:CG2	1:A:300:LYS:HG3	2.46	0.45
1:A:140:ARG:HD3	5:A:1005:HOH:O	2.17	0.43
1:A:140:ARG:CD	5:A:1005:HOH:O	2.67	0.43
1:A:90:MET:HA	1:A:91:PRO:C	2.38	0.42
1:A:55:ALA:HB2	1:A:63:VAL:HG23	2.00	0.42
1:A:130:LEU:O	1:A:134:LEU:HD13	2.19	0.42
1:A:192:TYR:CZ	1:A:261:MET:CE	3.03	0.42
1:A:86:THR:HG22	1:A:88:GLU:N	2.28	0.41
1:A:134:LEU:O	1:A:138:LEU:HG	2.21	0.41
1:A:204:GLN:O	1:A:208[A]:GLN:HG3	2.20	0.40
1:A:88:GLU:HG3	5:A:1262:HOH:O	2.21	0.40
1:A:202:VAL:HG22	2:A:901:9UY:C4A	2.51	0.40

All (2) 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:A:196:ARG:NH2	5:A:1108:HOH:O[4_445]	2.14	0.06
1:A:131[B]:SER:OG	5:A:1111:HOH:O[3_644]	2.18	0.02

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	314/315 (100%)	309 (98%)	5 (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	270/263 (103%)	266 (98%)	4 (2%)	70 33

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	128	ILE
1	A	212[A]	ARG
1	A	212[B]	ARG
1	A	216	ARG

Some 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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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
2	9UY	A	901	1,3	31,46,46	2.58	13 (41%)	27,67,67	2.05	9 (33%)
3	9V1	A	902	1,2	31,46,46	2.50	15 (48%)	30,67,67	1.94	9 (30%)

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	9UY	A	901	1,3	-	0/18/58/58	0/4/4/4
3	9V1	A	902	1,2	-	0/18/58/58	0/4/4/4

All (28) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	902	9V1	CHA-C1A	-5.15	1.47	1.51
2	A	901	9UY	CHA-C1A	-5.12	1.47	1.51
3	A	902	9V1	CHA-C4D	-4.54	1.47	1.51
2	A	901	9UY	CHB-C4A	-4.52	1.46	1.51
2	A	901	9UY	CHA-C4D	-4.32	1.48	1.51
3	A	902	9V1	CHB-C4A	-3.80	1.47	1.51
3	A	902	9V1	CHD-C1D	-2.69	1.48	1.51
3	A	902	9V1	CHD-C4C	-2.61	1.34	1.48
2	A	901	9UY	C2C-C3C	-2.51	1.52	1.56
2	A	901	9UY	CHD-C4C	-2.39	1.35	1.48
3	A	902	9V1	OC-C1C	-2.38	1.18	1.23
3	A	902	9V1	C1C-NC	-2.38	1.31	1.36
2	A	901	9UY	CHD-C1D	-2.31	1.49	1.51
3	A	902	9V1	C4B-NB	-2.26	1.33	1.37
3	A	902	9V1	C2A-C3A	2.19	1.44	1.37
2	A	901	9UY	C2A-C3A	2.39	1.44	1.37
3	A	902	9V1	C3D-C2D	3.03	1.46	1.37
2	A	901	9UY	C3D-C2D	3.03	1.46	1.37
2	A	901	9UY	C3A-C4A	3.10	1.46	1.42
3	A	902	9V1	C3A-C4A	3.37	1.47	1.42
3	A	902	9V1	C2D-C1D	3.40	1.47	1.42
2	A	901	9UY	C2D-C1D	3.42	1.47	1.42

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	901	9UY	C3D-C4D	4.15	1.46	1.39
3	A	902	9V1	C3B-C2B	4.17	1.45	1.37
3	A	902	9V1	C3D-C4D	4.28	1.46	1.39
3	A	902	9V1	C2A-C1A	4.35	1.46	1.39
2	A	901	9UY	C2A-C1A	4.62	1.47	1.39
2	A	901	9UY	C3B-C2B	5.68	1.47	1.37

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	902	9V1	CBA-CAA-C2A	-4.22	104.41	112.48
2	A	901	9UY	CBA-CAA-C2A	-3.57	105.66	112.48
2	A	901	9UY	C3C-CAC-CBC	-3.56	119.03	124.48
3	A	902	9V1	C3B-C2B-C1B	-2.98	102.94	107.38
3	A	902	9V1	CBC-CAC-C3C	-2.60	107.63	113.51
3	A	902	9V1	C2A-C3A-C4A	-2.19	102.60	111.33
2	A	901	9UY	CAD-CBD-CGD	-2.11	109.06	112.66
3	A	902	9V1	C3C-C2C-C1C	-2.08	102.74	104.25
2	A	901	9UY	C3B-C2B-C1B	-2.04	104.34	107.38
2	A	901	9UY	C2A-C3A-C4A	-2.04	103.20	111.33
3	A	902	9V1	CAA-CBA-CGA	2.47	116.89	112.66
3	A	902	9V1	CHD-C4C-NC	2.53	128.22	124.36
2	A	901	9UY	CHD-C4C-NC	2.68	128.46	124.36
2	A	901	9UY	CAD-C3D-C4D	2.88	129.32	127.30
3	A	902	9V1	C3B-C4B-NB	3.41	114.78	109.57
2	A	901	9UY	C3B-C4B-NB	3.68	115.19	109.57
2	A	901	9UY	CAA-C2A-C1A	3.79	129.96	127.30
3	A	902	9V1	CAA-C2A-C1A	4.45	130.42	127.30

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	901	9UY	1	0

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 [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	301/315 (95%)	0.29	21 (6%) 17 19	16, 23, 41, 68	1 (0%)

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	20	CYS	21.5
1	A	87	ALA	8.6
1	A	85	PRO	7.0
1	A	23	GLU	6.3
1	A	8	GLY	6.1
1	A	7	SER	5.0
1	A	128	ILE	4.7
1	A	129	ASP	4.7
1	A	130	LEU	4.4
1	A	86	THR	4.3
1	A	2	VAL	4.1
1	A	88	GLU	3.7
1	A	6	ALA	3.1
1	A	5	HIS	3.0
1	A	267	LEU	2.8
1	A	217	VAL	2.6
1	A	246[A]	PHE	2.3
1	A	3	ALA	2.2
1	A	24	GLU	2.1
1	A	205	MET	2.1
1	A	204	GLN	2.1

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
4	CL	A	904	1/1	0.99	0.07	0.41	37,37,37,37	0
3	9V1	A	902	43/43	0.80	0.13	0.16	20,24,39,55	43
2	9UY	A	901	43/43	0.83	0.13	0.14	20,24,39,55	43
4	CL	A	903	1/1	1.00	0.04	-1.14	23,23,23,23	0

6.5 Other polymers [i](#)

There are no such residues in this entry.