Full wwPDB X-ray Structure Validation Report (i)

Jan 31, 2016 – 06:28 PM GMT

PDB ID	:	1AY2
Title	:	STRUCTURE OF THE FIBER-FORMING PROTEIN PILIN AT 2.6 ANGSTROMS RESOLUTION
Authors	:	Forest, K.T.; Parge, H.E.; Tainer, J.A.
Deposited on	:	1997-11-13
Resolution	:	2.60 Å(reported)

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

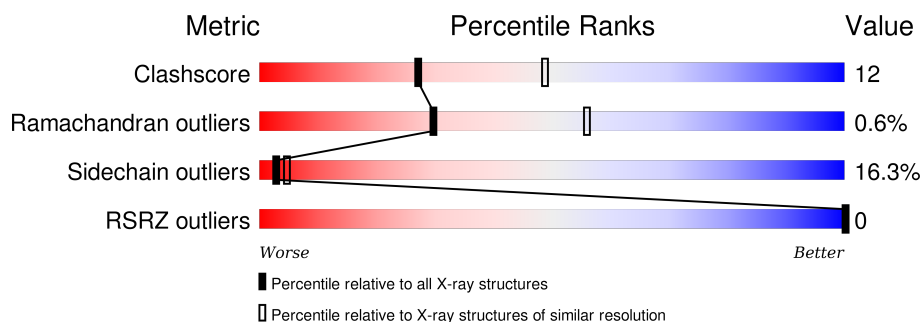|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.60 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
Clashscore		12
Ramachandran outliers		0.6%
Sidechain outliers		16.3%
RSRZ outliers		0

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))
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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 <=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	158	57% 34% 9%

Full wwPDB X-ray Structure Validation Report

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 1373 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 TYPE 4 PILIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	158	Total	C	N	O	S	0	0	0
			1208	761	206	237	4			

- Molecule 2 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	2	Total	C	N	O	0	0
			25	14	1	10		

- Molecule 3 is PLATINUM (II) ION (three-letter code: PT) (formula: Pt).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Pt	0	0
			1	1		

- Molecule 4 is HEPTANE-1,2,3-TRIOL (three-letter code: HTO) (formula: $C_7H_{16}O_3$).
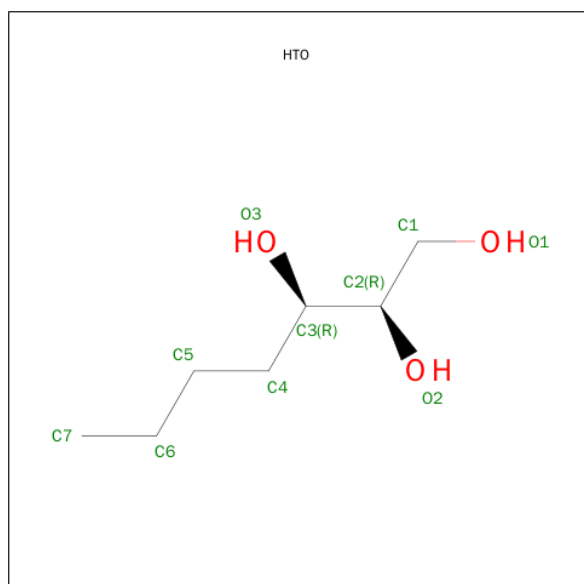
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 10	C 7	O 3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	129	Total 129	O 129	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 errors 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: TYPE 4 PILIN

Chain A:
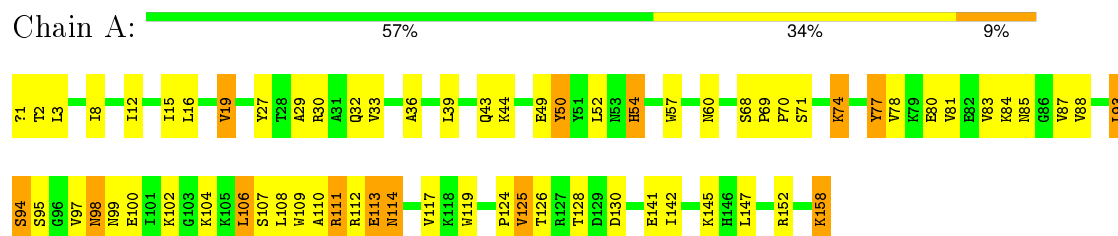
4   Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	127.58Å 121.08Å 26.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 2.60 9.97 – 2.60	Depositor EDS
% Data completeness (in resolution range)	95.0 (10.00-2.60) 96.2 (9.97-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$<I/\sigma(I)>$	-	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.194 , (Not available) 0.213 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	39.9	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.08 , 70.3	EDS
Estimated twinning fraction	0.024 for -k,-h,-l	Xtriage
L-test for twinning[1]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Outliers	1 of 6430 reflections (0.016%)	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	1373	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 7.10% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: HTO, 0A9, NAG, PT, GLA

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	0.94	0/1217	1.78	32/1650 (1.9%)

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	A	0	1

There are no bond length outliers.

All (32) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	106	LEU	CA-CB-CG	9.36	136.82	115.30
1	A	2	THR	N-CA-C	8.94	135.12	111.00
1	A	77	TYR	CB-CG-CD2	-8.50	115.90	121.00
1	A	109	TRP	CD1-CG-CD2	8.16	112.83	106.30
1	A	119	TRP	CD1-CG-CD2	8.12	112.80	106.30
1	A	57	TRP	CD1-CG-CD2	7.91	112.63	106.30
1	A	152	ARG	NE-CZ-NH2	-7.86	116.37	120.30
1	A	57	TRP	CE2-CD2-CG	-7.52	101.29	107.30
1	A	119	TRP	CE2-CD2-CG	-7.49	101.31	107.30
1	A	109	TRP	CE2-CD2-CG	-7.19	101.55	107.30
1	A	27	TYR	CB-CG-CD1	-7.12	116.73	121.00
1	A	57	TRP	CG-CD2-CE3	7.04	140.24	133.90
1	A	19	VAL	CA-CB-CG2	-6.82	100.67	110.90
1	A	119	TRP	CG-CD2-CE3	6.80	140.02	133.90
1	A	30	ARG	NE-CZ-NH2	-6.69	116.95	120.30
1	A	57	TRP	CB-CG-CD1	-6.35	118.75	127.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	68	SER	CA-CB-OG	6.28	128.15	111.20
1	A	152	ARG	NE-CZ-NH1	6.22	123.41	120.30
1	A	125	VAL	N-CA-CB	-6.08	98.12	111.50
1	A	104	LYS	CA-CB-CG	5.93	126.44	113.40
1	A	32	GLN	O-C-N	-5.89	113.28	122.70
1	A	19	VAL	CA-CB-CG1	5.63	119.34	110.90
1	A	119	TRP	CB-CG-CD1	-5.62	119.69	127.00
1	A	2	THR	OG1-CB-CG2	5.57	122.80	110.00
1	A	2	THR	CB-CA-C	-5.50	96.76	111.60
1	A	109	TRP	CG-CD1-NE1	-5.45	104.66	110.10
1	A	32	GLN	CA-C-N	5.44	129.17	117.20
1	A	142	ILE	CB-CA-C	-5.34	100.91	111.60
1	A	113	GLU	N-CA-C	-5.32	96.63	111.00
1	A	113	GLU	CA-C-N	-5.24	105.67	117.20
1	A	119	TRP	CG-CD1-NE1	-5.23	104.87	110.10
1	A	87	VAL	CG1-CB-CG2	-5.12	102.72	110.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	50	TYR	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	1208	0	1211	30	0
2	A	25	0	22	0	0
3	A	1	0	0	0	0
4	A	10	0	16	1	0
5	A	129	0	0	0	0
All	All	1373	0	1249	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 12.

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)
1:A:1:0A9:HE2	4:A:162:HTO:H41	1.49	0.92
1:A:111:ARG:NH2	1:A:158:LYS:HG2	2.08	0.68
1:A:70:PRO:HB2	1:A:80:GLU:HB2	1.81	0.62
1:A:39:LEU:HB3	1:A:81:VAL:HG21	1.83	0.59
1:A:8:ILE:O	1:A:12:ILE:HG23	2.04	0.57
1:A:93:LEU:HD12	1:A:94:SER:H	1.69	0.56
1:A:112:ARG:HH11	1:A:114:ASN:HA	1.71	0.55
1:A:124:PRO:HD2	1:A:141:GLU:HB3	1.88	0.55
1:A:98:ASN:OD1	1:A:100:GLU:HG2	2.07	0.54
1:A:36:ALA:HA	1:A:78:VAL:HG21	1.88	0.53
1:A:60:ASN:HA	1:A:83:VAL:O	2.09	0.53
1:A:112:ARG:NH1	1:A:114:ASN:HA	2.23	0.53
1:A:107:SER:OG	1:A:124:PRO:HA	2.09	0.52
1:A:44:LYS:HD3	1:A:108:LEU:HD23	1.92	0.51
1:A:71:SER:O	1:A:74:LYS:HE3	2.11	0.51
1:A:77:TYR:HB3	1:A:97:VAL:HG22	1.92	0.50
1:A:111:ARG:HH21	1:A:158:LYS:HG2	1.78	0.48
1:A:94:SER:HA	1:A:102:LYS:HB2	1.95	0.47
1:A:15:ILE:HG13	1:A:16:LEU:N	2.30	0.46
1:A:110:ALA:HB1	1:A:117:VAL:HB	1.98	0.45
1:A:69:PRO:HA	1:A:70:PRO:HD2	1.62	0.43
1:A:29:ALA:HB2	1:A:98:ASN:ND2	2.34	0.43
1:A:84:LYS:HB3	1:A:84:LYS:HE2	1.91	0.43
1:A:99:ASN:HA	1:A:102:LYS:HE3	2.01	0.43
1:A:39:LEU:HA	1:A:39:LEU:HD23	1.78	0.42
1:A:43:GLN:HE21	1:A:81:VAL:HG11	1.85	0.42
1:A:3:LEU:HD12	1:A:3:LEU:HA	1.92	0.42
1:A:113:GLU:HG3	1:A:113:GLU:O	2.20	0.41
1:A:112:ARG:HD3	1:A:114:ASN:ND2	2.36	0.41
1:A:50:TYR:O	1:A:54:HIS:ND1	2.54	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	156/158 (99%)	145 (93%)	10 (6%)	1 (1%)	30 56

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	114	ASN

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	129/129 (100%)	108 (84%)	21 (16%)	3 5

All (21) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	19	VAL
1	A	33	VAL
1	A	49	GLU
1	A	52	LEU
1	A	54	HIS
1	A	74	LYS
1	A	85	ASN
1	A	88	VAL
1	A	93	LEU
1	A	94	SER
1	A	95	SER
1	A	98	ASN
1	A	106	LEU
1	A	111	ARG
1	A	125	VAL
1	A	126	THR
1	A	128	THR
1	A	130	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	145	LYS
1	A	147	LEU
1	A	158	LYS

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

Mol	Chain	Res	Type
1	A	43	GLN
1	A	53	ASN
1	A	114	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	#$|Z| > 2$
1	0A9	A	1	1	10,11,13	1.40	1 (10%)	10,13,16	1.32	1 (10%)

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
1	0A9	A	1	1	-	0/4/6/10	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1	0A9	CE1-CD1	2.08	1.43	1.38

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1	0A9	O-C-CA	-2.49	119.01	125.49

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
1	A	1	0A9	1	0

5.5 Carbohydrates (i)

2 carbohydrates 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	GLA	A	160	2	11,11,12	0.27	0	14,15,17	0.74	1 (7%)
2	NAG	A	161	1,2	14,14,15	0.48	0	15,19,21	1.02	2 (13%)

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	GLA	A	160	2	-	0/2/19/22	0/1/1/1
2	NAG	A	161	1,2	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	161	NAG	C2-N2-C7	-2.59	119.71	123.04
2	A	161	NAG	C4-C3-C2	-2.53	107.30	111.23
2	A	160	GLA	C1-O5-C5	2.26	115.11	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	#$|Z| > 2$
4	HTO	A	162	-	9,9,9	0.85	1 (11%)	8,10,10	1.39	1 (12%)

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
4	HTO	A	162	-	-	0/10/10/10	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	162	HTO	C3-C2	2.13	1.58	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	162	HTO	O2-C2-C3	-2.58	104.52	109.79

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
4	A	162	HTO	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	157/158 (99%)	-0.77	0	100	100	5, 29, 47, 61	0

There are no RSRZ outliers to report.

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. 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(Å^2)	Q<0.9
1	0A9	A	1	11/13	0.81	0.19	-	25,38,52,54	0

6.3 Carbohydrates (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(Å^2)	Q<0.9
2	NAG	A	161	14/15	0.95	0.12	0.07	24,28,36,43	0
2	GLA	A	160	11/12	0.92	0.15	-	31,38,47,50	0

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(Å^2)	Q<0.9
3	PT	A	200	1/1	0.98	0.34	-	73,73,73,73	0
4	HTO	A	162	10/10	0.84	0.18	-	88,94,98,101	0

6.5 Other polymers (i)

There are no such residues in this entry.