



Full wwPDB X-ray Structure Validation Report ⓘ

Apr 4, 2019 – 04:16 PM EDT

PDB ID : 6CF7
Title : Crystal structure of the A/Solomon Islands/3/2006(H1N1) influenza virus hemagglutinin in complex with small molecule JNJ4796
Authors : Kadam, R.U.; Wilson, I.A.
Deposited on : 2018-02-13
Resolution : 2.72 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix) : 1.13
EDS : rb-20031633
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20031633

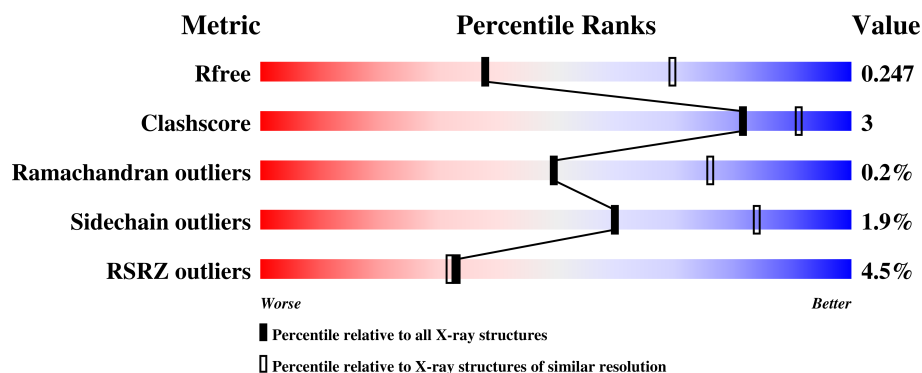
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.72 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}	111664	2885 (2.74-2.70)
Clashscore	122126	3205 (2.74-2.70)
Ramachandran outliers	120053	3157 (2.74-2.70)
Sidechain outliers	120020	3158 (2.74-2.70)
RSRZ outliers	108989	2802 (2.74-2.70)

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	326	 5% 88% 10% ..
2	B	174	 4% 92% 6% •

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
3	NAG	A	412	-	-	-	X

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 4157 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 Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	322	Total	C	N	O	S	0	0	0
			2529	1593	442	483	11			

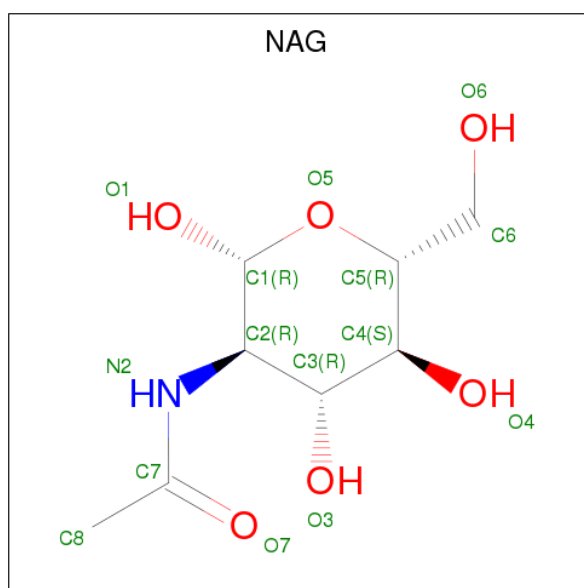
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	53	ARG	LEU	engineered mutation	UNP A7Y8I1

- Molecule 2 is a protein called Hemagglutinin.

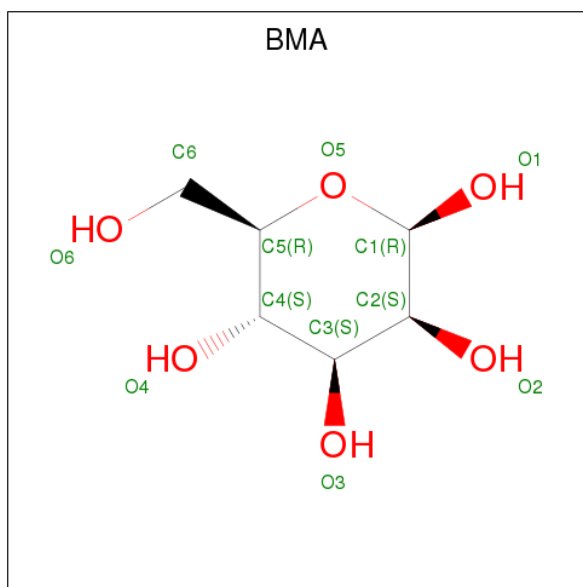
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	171	Total	C	N	O	S	0	0	0
			1377	862	235	273	7			

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



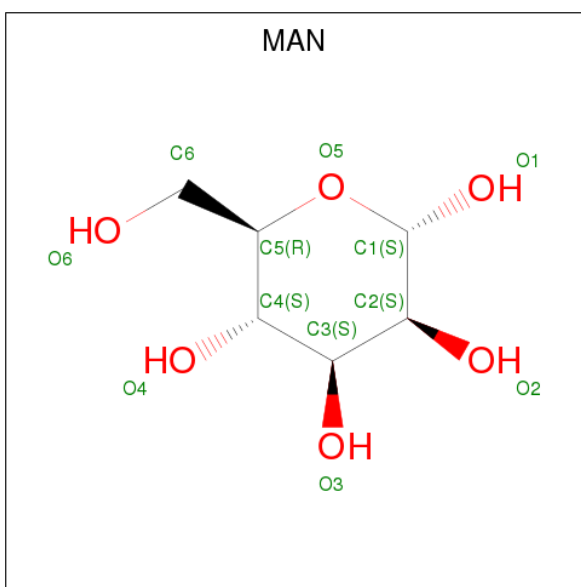
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1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

- Molecule 4 is BETA-D-MANNOSE (three-letter code: BMA) (formula: $C_6H_{12}O_6$).



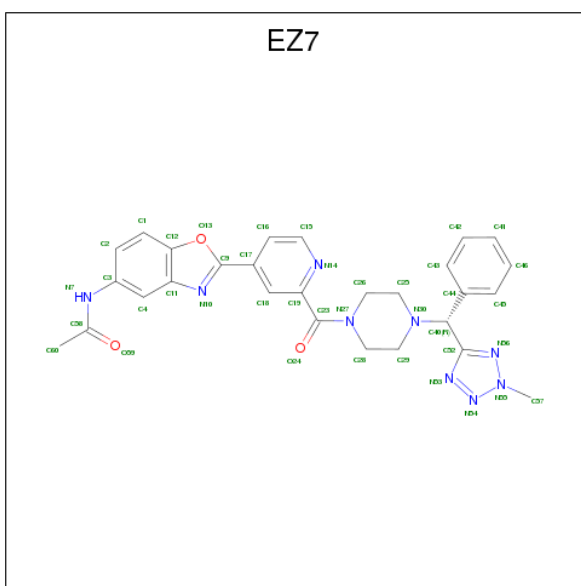
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	6	5		
4	A	1	Total	C	O	0	0
			11	6	5		

- Molecule 5 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			11	6	5		

- Molecule 6 is N-[2-(2-{4-[(R)-(2-methyl-2H-tetrazol-5-yl)(phenyl)methyl]piperazine-1-carbonyl}pyridin-4-yl)-1,3-benzoxazol-5-yl]acetamide (three-letter code: EZ7) (formula: C₂₈H₂₇N₉O₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	B	1	Total	C	N	O	0	0
			40	28	9	3		

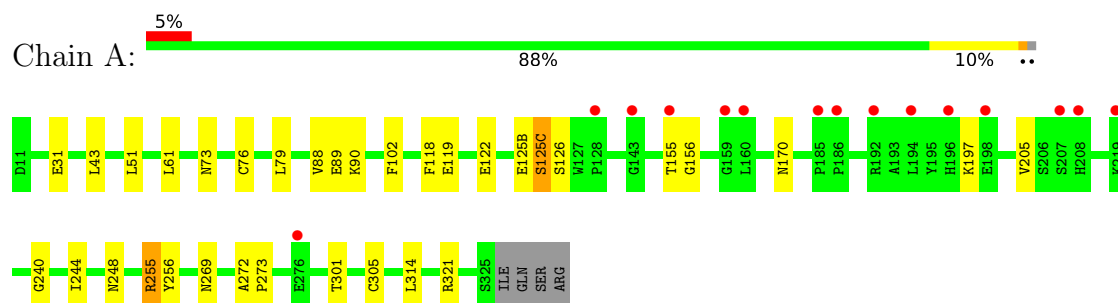
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	28	Total 28	O 28	0	0
7	B	24	Total 24	O 24	0	0

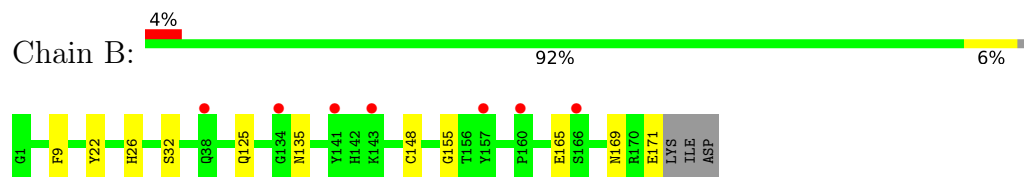
3 Residue-property plots [i](#)

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: Hemagglutinin



• Molecule 2: Hemagglutinin



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 3 2	Depositor
Cell constants a, b, c, α , β , γ	161.44Å 161.44Å 161.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.78 – 2.72 44.78 – 2.72	Depositor EDS
% Data completeness (in resolution range)	99.6 (44.78-2.72) 99.6 (44.78-2.72)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.32 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.10_2155	Depositor
R, R_{free}	0.226 , 0.249 0.226 , 0.247	Depositor DCC
R_{free} test set	996 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	59.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 42.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	4157	wwPDB-VP
Average B, all atoms (Å ²)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.26% 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: BMA, EZ7, NAG, MAN

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	0.26	0/2595	0.46	0/3530
2	B	0.29	0/1404	0.41	0/1887
All	All	0.27	0/3999	0.44	0/5417

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	2529	0	2437	18	0
2	B	1377	0	1302	4	0
3	A	126	0	111	0	0
4	A	22	0	19	0	0
5	A	11	0	10	0	0
6	B	40	0	0	0	0
7	A	28	0	0	1	0
7	B	24	0	0	0	0
All	All	4157	0	3879	22	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 (22) 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:170:ASN:HB3	1:A:240:GLY:H	1.54	0.73
1:A:155:THR:HG22	1:A:156:GLY:H	1.60	0.67
1:A:197:LYS:HE3	1:A:248:ASN:HB2	1.84	0.59
1:A:51:LEU:HD13	1:A:88:VAL:HG21	1.86	0.56
2:B:26:HIS:O	2:B:32:SER:HA	2.08	0.53
1:A:61:LEU:HD12	1:A:89:GLU:HB2	1.95	0.49
1:A:122:GLU:HA	1:A:256:TYR:HA	1.95	0.49
2:B:165:GLU:O	2:B:169:ASN:ND2	2.37	0.49
1:A:255:ARG:HG2	1:A:256:TYR:CD1	2.49	0.48
1:A:31:GLU:OE2	1:A:321:ARG:NH1	2.47	0.47
1:A:43:LEU:HB2	1:A:314:LEU:HB2	1.97	0.47
1:A:205:VAL:HG23	1:A:244:ILE:HB	1.98	0.45
1:A:61:LEU:HA	1:A:79:LEU:HD21	1.97	0.45
1:A:119:GLU:HB3	7:A:501:HOH:O	2.17	0.45
1:A:89:GLU:O	1:A:269:ASN:HA	2.17	0.44
1:A:73:ASN:HB3	1:A:76:CYS:SG	2.59	0.43
1:A:301:THR:HB	1:A:305:CYS:SG	2.59	0.43
2:B:125:GLN:NE2	2:B:155:GLY:HA2	2.33	0.43
2:B:9:PHE:O	2:B:135:ASN:HA	2.20	0.42
1:A:90:LYS:HA	1:A:90:LYS:HD3	1.83	0.41
1:A:272:ALA:HA	1:A:273:PRO:HD3	1.93	0.41
1:A:125(C):SER:O	1:A:125(C):SER:OG	2.36	0.41

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	320/326 (98%)	307 (96%)	12 (4%)	1 (0%)	43	69
2	B	169/174 (97%)	165 (98%)	4 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	489/500 (98%)	472 (96%)	16 (3%)	1 (0%)	49 76

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	125(C)	SER

5.3.2 Protein sidechains ⓘ

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	282/286 (99%)	277 (98%)	5 (2%)	62 85
2	B	147/150 (98%)	144 (98%)	3 (2%)	58 83
All	All	429/436 (98%)	421 (98%)	8 (2%)	60 84

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	102	PHE
1	A	118	PHE
1	A	125(B)	GLU
1	A	126	SER
1	A	255	ARG
2	B	22	TYR
2	B	148	CYS
2	B	171	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

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](#)

13 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	$\# Z > 2$
3	NAG	A	401	1,3	14,14,15	0.44	0	17,19,21	0.53	0
3	NAG	A	402	3,4	14,14,15	0.22	0	17,19,21	0.41	0
4	BMA	A	403	3,5	11,11,12	1.22	2 (18%)	15,15,17	1.48	4 (26%)
5	MAN	A	404	4	11,11,12	1.20	2 (18%)	15,15,17	1.86	4 (26%)
3	NAG	A	405	1,3	14,14,15	0.27	0	17,19,21	0.43	0
3	NAG	A	406	3,4	14,14,15	0.31	0	17,19,21	0.48	0
4	BMA	A	407	3	11,11,12	0.65	0	15,15,17	0.96	0
3	NAG	A	408	1,3	14,14,15	0.29	0	17,19,21	0.58	0
3	NAG	A	409	3	14,14,15	0.20	0	17,19,21	0.43	0
3	NAG	A	410	1,3	14,14,15	0.17	0	17,19,21	0.43	0
3	NAG	A	411	3	14,14,15	0.17	0	17,19,21	0.47	0
3	NAG	A	412	1	14,14,15	0.21	0	17,19,21	0.41	0
6	EZ7	B	201	-	36,45,45	3.19	8 (22%)	50,64,64	1.41	9 (18%)

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
3	NAG	A	401	1,3	-	0/6/23/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	A	402	3,4	-	0/6/23/26	0/1/1/1
4	BMA	A	403	3,5	-	0/2/19/22	0/1/1/1
5	MAN	A	404	4	-	0/2/19/22	0/1/1/1
3	NAG	A	405	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	406	3,4	-	0/6/23/26	0/1/1/1
4	BMA	A	407	3	-	0/2/19/22	0/1/1/1
3	NAG	A	408	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	409	3	-	0/6/23/26	0/1/1/1
3	NAG	A	410	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	411	3	-	0/6/23/26	0/1/1/1
3	NAG	A	412	1	-	0/6/23/26	0/1/1/1
6	EZ7	B	201	-	-	0/22/38/38	0/6/6/6

All (12) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	B	201	EZ7	N53-N54	-12.40	1.11	1.34
6	B	201	EZ7	C44-C40	-10.16	1.39	1.52
6	B	201	EZ7	C19-C23	-5.81	1.40	1.50
6	B	201	EZ7	C4-C11	-5.05	1.34	1.41
6	B	201	EZ7	N54-N55	-3.94	1.27	1.34
6	B	201	EZ7	C3-N7	-3.47	1.34	1.41
4	A	403	BMA	O5-C1	-2.53	1.39	1.43
6	B	201	EZ7	C12-C11	-2.52	1.34	1.42
4	A	403	BMA	C2-C3	2.22	1.55	1.52
6	B	201	EZ7	C15-N14	2.42	1.39	1.34
5	A	404	MAN	O5-C5	2.52	1.48	1.43
5	A	404	MAN	C1-C2	2.91	1.58	1.52

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	B	201	EZ7	C16-C15-N14	-3.54	119.71	123.94
6	B	201	EZ7	C18-C19-N14	-3.15	119.61	123.23
5	A	404	MAN	O2-C2-C3	-2.36	105.55	110.16
6	B	201	EZ7	C3-N7-C58	-2.12	124.11	128.00
4	A	403	BMA	O2-C2-C3	-2.08	106.10	110.16
6	B	201	EZ7	C16-C17-C9	-2.08	117.27	120.44
4	A	403	BMA	C2-C3-C4	2.04	114.42	110.89
6	B	201	EZ7	C18-C17-C9	2.05	123.11	120.10
6	B	201	EZ7	C28-N27-C26	2.06	116.58	112.60
5	A	404	MAN	C1-C2-C3	2.14	112.29	109.66

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	403	BMA	C3-C4-C5	2.17	114.14	110.23
6	B	201	EZ7	O59-C58-N7	2.77	126.75	123.05
6	B	201	EZ7	C17-C9-N10	2.90	127.45	123.56
6	B	201	EZ7	C15-N14-C19	3.42	121.41	116.94
4	A	403	BMA	C1-C2-C3	3.71	114.22	109.66
5	A	404	MAN	C1-O5-C5	3.93	117.55	112.20
5	A	404	MAN	O5-C1-C2	4.39	117.56	110.79

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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	322/326 (98%)	0.24	15 (4%) 31 30	28, 69, 121, 157	0
2	B	171/174 (98%)	0.40	7 (4%) 37 36	25, 63, 131, 162	0
All	All	493/500 (98%)	0.30	22 (4%) 33 32	25, 67, 128, 162	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	143	GLY	4.1
1	A	155	THR	3.8
1	A	208	HIS	3.5
1	A	160	LEU	3.2
1	A	192	ARG	3.0
2	B	143	LYS	2.9
1	A	194	LEU	2.7
2	B	141	TYR	2.6
2	B	166	SER	2.5
2	B	160	PRO	2.5
2	B	38	GLN	2.4
2	B	134	GLY	2.3
1	A	196	HIS	2.2
1	A	198	GLU	2.2
1	A	276	GLU	2.2
1	A	207	SER	2.1
1	A	128	PRO	2.1
1	A	186	PRO	2.1
2	B	157	TYR	2.1
1	A	185	PRO	2.1
1	A	219	LYS	2.1
1	A	159	GLY	2.0

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. 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
4	BMA	A	407	11/12	0.48	0.40	159,167,170,171	0
3	NAG	A	410	14/15	0.69	0.22	73,96,104,122	0
3	NAG	A	411	14/15	0.73	0.38	143,152,156,159	0
3	NAG	A	406	14/15	0.75	0.30	139,151,160,167	0
3	NAG	A	412	14/15	0.78	0.44	132,151,156,158	0
4	BMA	A	403	11/12	0.80	0.18	69,81,95,99	0
5	MAN	A	404	11/12	0.80	0.29	84,92,104,108	0
3	NAG	A	409	14/15	0.84	0.33	100,124,133,134	0
3	NAG	A	405	14/15	0.86	0.25	123,129,135,142	0
3	NAG	A	402	14/15	0.91	0.20	64,75,86,86	0
3	NAG	A	408	14/15	0.91	0.30	69,84,100,108	0
3	NAG	A	401	14/15	0.93	0.18	51,57,65,67	0
6	EZ7	B	201	40/40	0.96	0.18	35,46,74,77	0

6.5 Other polymers [i](#)

There are no such residues in this entry.