wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 12:20 am BST

PDB ID	:	2P5B
Title	:	The complex structure of JMJD2A and trimethylated H3K36 peptide
Authors	:	Zhang, G.; Chen, Z.; Zang, J.; Hong, X.; Shi, Y.
Deposited on	:	2007-03-14
Resolution	:	1.99 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

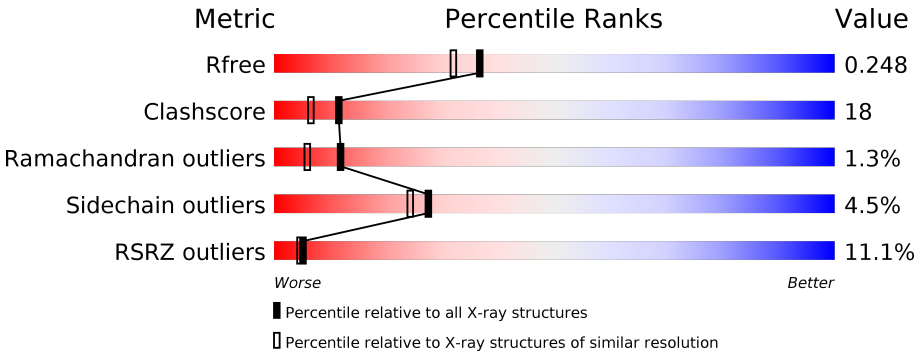
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.99 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}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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	352	9% 68% 28%
1	B	352	7% 67% 26%
2	I	22	59% 32% 45% 23%
2	J	22	55% 32% 27% 5% 36%

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	OXY	A	502	-	-	X	-

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 6164 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 JmjC domain-containing histone demethylation protein 3A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	347	Total	C	N	O	S	0	0	0
			2848	1838	480	515	15			
1	B	337	Total	C	N	O	S	0	0	0
			2776	1796	469	496	15			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	CLONING ARTIFACT	UNP O75164
A	0	SER	-	CLONING ARTIFACT	UNP O75164
A	1	MET	-	CLONING ARTIFACT	UNP O75164
B	-1	GLY	-	CLONING ARTIFACT	UNP O75164
B	0	SER	-	CLONING ARTIFACT	UNP O75164
B	1	MET	-	CLONING ARTIFACT	UNP O75164

- Molecule 2 is a protein called Histone H3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	I	17	Total	C	N	O	0	0	0
			128	82	27	19			
2	J	14	Total	C	N	O	0	0	0
			98	62	21	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	10	M3L	LYS	MODIFIED RESIDUE	UNP P84239
I	21	LEU	-	CLONING ARTIFACT	UNP P84239
J	10	M3L	LYS	MODIFIED RESIDUE	UNP P84239
J	21	LEU	-	CLONING ARTIFACT	UNP P84239

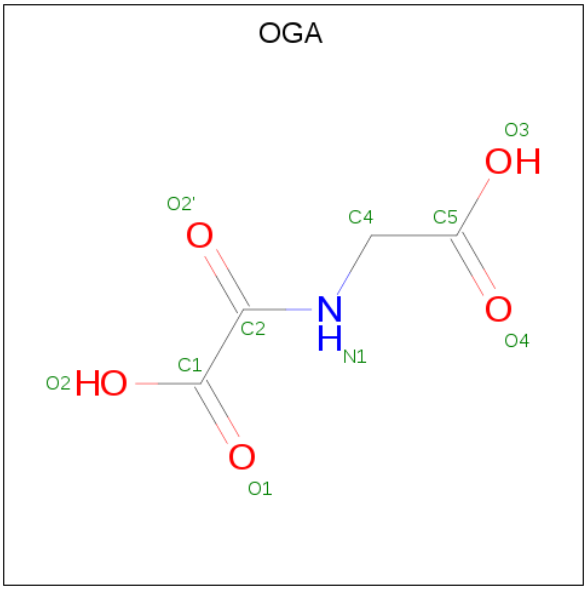
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Zn 1 1	0	0
3	A	1	Total Zn 1 1	0	0

- Molecule 4 is FE (II) ION (three-letter code: FE2) (formula: Fe).

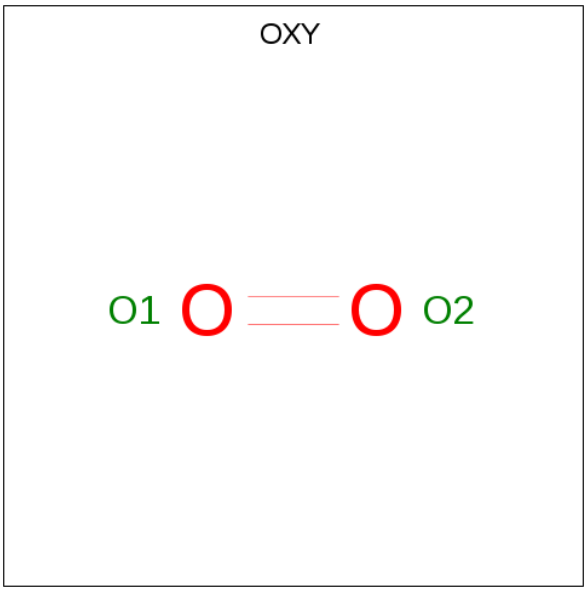
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Fe 1 1	0	0
4	A	1	Total Fe 1 1	0	0

- Molecule 5 is N-OXALYLGLYCINE (three-letter code: OGA) (formula: $C_4H_5NO_5$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C N O 10 4 1 5	0	0
5	B	1	Total C N O 10 4 1 5	0	0

- Molecule 6 is OXYGEN MOLECULE (three-letter code: OXY) (formula: O_2).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total O 2 2	0	0
6	A	1	Total O 2 2	0	0
6	B	1	Total O 2 2	0	0

- Molecule 7 is water.

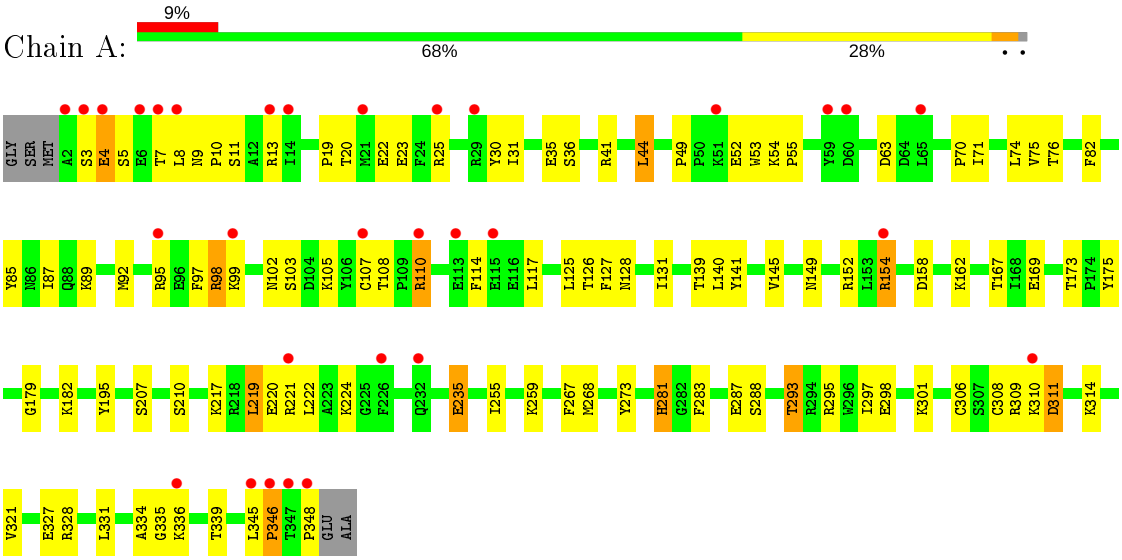
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	139	Total O 139 139	0	0
7	B	140	Total O 140 140	0	0
7	I	4	Total O 4 4	0	0
7	J	1	Total O 1 1	0	0

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.

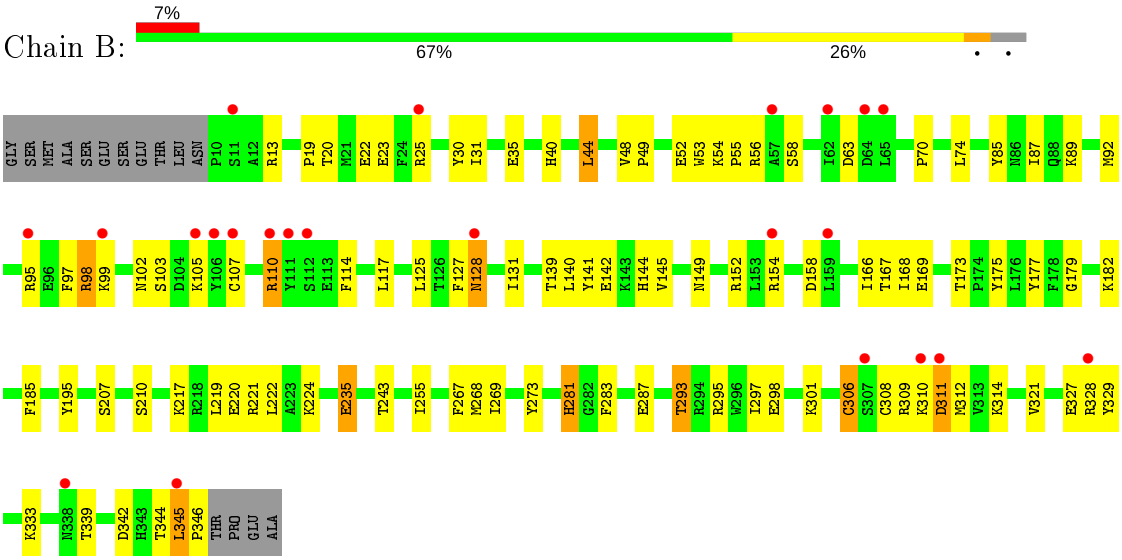
- Molecule 1: JmjC domain-containing histone demethylation protein 3A

Chain A:



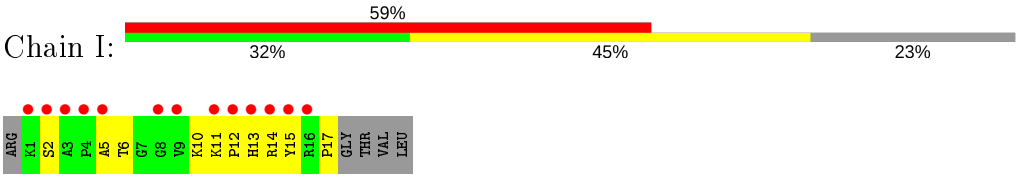
- Molecule 1: JmjC domain-containing histone demethylation protein 3A

Chain B:



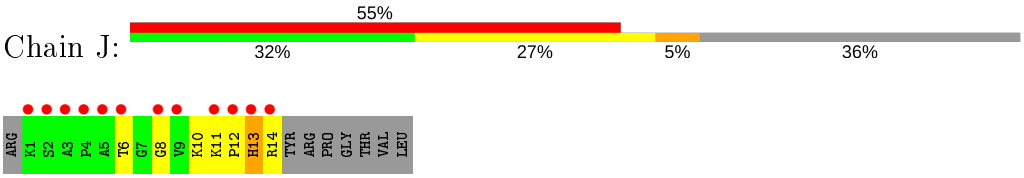
- Molecule 2: Histone H3

Chain I:



- Molecule 2: Histone H3

Chain J:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	100.92Å 150.84Å 57.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.34 – 1.99 47.34 – 1.99	Depositor EDS
% Data completeness (in resolution range)	81.0 (47.34-1.99) 81.1 (47.34-1.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.95 (at 2.00Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.238 , 0.263 0.226 , 0.248	Depositor DCC
R_{free} test set	1557 reflections (2.98%)	wwPDB-VP
Wilson B-factor (Å^2)	34.0	Xtriage
Anisotropy	0.243	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6164	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.95% 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: OGA, ZN, M3L, FE2, OXY

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	0.35	0/2934	0.59	0/3972
1	B	0.36	0/2861	0.58	0/3870
2	I	0.30	0/120	0.51	0/163
2	J	0.31	0/88	0.55	0/119
All	All	0.35	0/6003	0.58	0/8124

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	2848	0	2774	99	0
1	B	2776	0	2710	99	0
2	I	128	0	133	13	0
2	J	98	0	104	7	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	10	0	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	10	0	3	1	0
6	A	4	0	0	2	0
6	B	2	0	0	1	0
7	A	139	0	0	9	0
7	B	140	0	0	7	0
7	I	4	0	0	1	0
7	J	1	0	0	0	0
All	All	6164	0	5727	204	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 18.

The worst 5 of 204 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:139:THR:HG22	1:B:141:TYR:H	1.23	0.99
1:A:139:THR:HG22	1:A:141:TYR:H	1.29	0.95
1:B:40:HIS:NE2	1:B:346:PRO:HG2	1.89	0.88
1:A:75:VAL:HG12	7:A:515:HOH:O	1.78	0.82
1:A:139:THR:HG23	1:A:287:GLU:OE1	1.79	0.81

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	345/352 (98%)	328 (95%)	14 (4%)	3 (1%)	17 11
1	B	335/352 (95%)	318 (95%)	15 (4%)	2 (1%)	25 19
2	I	14/22 (64%)	12 (86%)	0	2 (14%)	0 0
2	J	11/22 (50%)	8 (73%)	1 (9%)	2 (18%)	0 0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	705/748 (94%)	666 (94%)	30 (4%)	9 (1%)	12 6

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	311	ASP
1	B	311	ASP
1	B	345	LEU
1	A	346	PRO
2	I	2	SER

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	305/308 (99%)	291 (95%)	14 (5%)	27 23
1	B	296/308 (96%)	282 (95%)	14 (5%)	26 22
2	I	10/16 (62%)	10 (100%)	0	100 100
2	J	7/16 (44%)	7 (100%)	0	100 100
All	All	618/648 (95%)	590 (96%)	28 (4%)	27 24

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	306	CYS
1	B	52	GLU
1	B	293	THR
1	A	327	GLU
1	B	44	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	84	GLN
1	B	102	ASN
1	B	262	GLN
1	A	281	HIS
1	B	144	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

2 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	M3L	I	10	2	10,11,12	0.44	0	9,14,16	1.01	1 (11%)
2	M3L	J	10	2	10,11,12	0.48	0	9,14,16	1.04	1 (11%)

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	M3L	I	10	2	-	1/9/10/12	-
2	M3L	J	10	2	-	1/9/10/12	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	10	M3L	CM2-NZ-CM1	-2.26	103.16	108.97
2	J	10	M3L	CM2-NZ-CM1	-2.23	103.25	108.97

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	J	10	M3L	CE-CD-CG-CB
2	I	10	M3L	CE-CD-CG-CB

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	I	10	M3L	2	0
2	J	10	M3L	2	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 4 are monoatomic - leaving 5 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	$\#\|Z\| > 2$
5	OGA	B	402	4	3,9,9	0.64	0	4,11,11	2.10	1 (25%)
6	OXY	B	503	4	1,1,1	1.17	0	-		
5	OGA	A	401	4	3,9,9	0.76	0	4,11,11	2.09	1 (25%)
6	OXY	A	502	4	1,1,1	1.21	0	-		
6	OXY	A	501	-	1,1,1	1.19	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
5	OGA	B	402	4	-	0/3/9/9	-
5	OGA	A	401	4	-	0/3/9/9	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	401	OGA	C1-C2-N1	3.99	119.57	115.60
5	B	402	OGA	C1-C2-N1	3.99	119.56	115.60

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	402	OGA	1	0
6	B	503	OXY	1	0
6	A	502	OXY	2	0
6	A	501	OXY	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

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, 95^{th} 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	347/352 (98%)	0.74	31 (8%)	9	8	22, 40, 69, 104	0
1	B	337/352 (95%)	0.61	23 (6%)	17	16	23, 41, 70, 79	0
2	I	16/22 (72%)	3.86	13 (81%)	0	0	80, 91, 109, 110	0
2	J	13/22 (59%)	5.89	12 (92%)	0	0	98, 104, 119, 120	0
All	All	713/748 (95%)	0.84	79 (11%)	5	4	22, 42, 78, 120	0

The worst 5 of 79 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	J	1	LYS	15.6
2	J	3	ALA	13.2
2	J	2	SER	11.9
1	A	2	ALA	10.1
2	I	13	HIS	9.2

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, 95^{th} 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	M3L	J	10	12/13	0.75	0.28	85,89,98,98	0
2	M3L	I	10	12/13	0.80	0.28	72,75,84,84	0

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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, 95^{th} 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
6	OXY	A	501	2/2	0.86	0.34	34,34,34,39	2
5	OGA	B	402	10/10	0.87	0.17	44,50,56,58	0
5	OGA	A	401	10/10	0.89	0.13	50,53,59,61	0
6	OXY	B	503	2/2	0.90	0.24	41,41,41,41	2
6	OXY	A	502	2/2	0.91	0.28	42,42,42,43	2
3	ZN	B	351	1/1	0.99	0.13	45,45,45,45	0
4	FE2	B	352	1/1	0.99	0.10	34,34,34,34	0
3	ZN	A	351	1/1	0.99	0.21	26,26,26,26	1
4	FE2	A	352	1/1	0.99	0.12	34,34,34,34	0

6.5 Other polymers

There are no such residues in this entry.