



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 18, 2024 – 09:41 PM EDT

PDB ID : 4GY5  
Title : Crystal structure of the tandem tudor domain and plant homeodomain of UHRF1 with Histone H3K9me3  
Authors : Cheng, J.; Yang, Y.; Fang, J.; Xiao, J.; Zhu, T.; Chen, F.; Wang, P.; Xu, Y.  
Deposited on : 2012-09-05  
Resolution : 2.96 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](mailto: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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

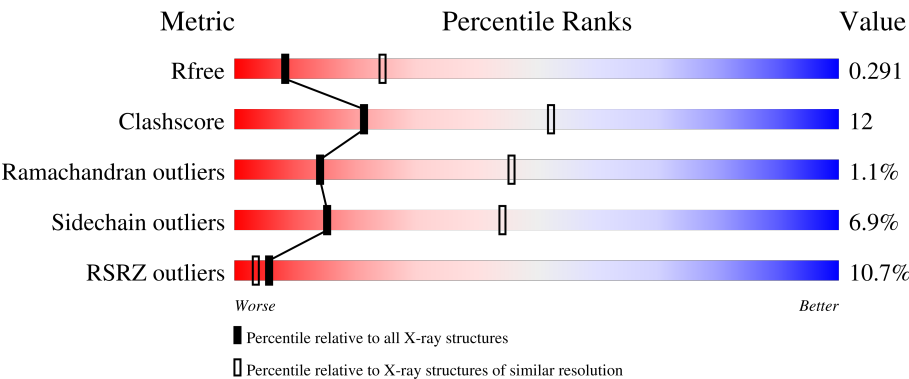
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

# 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 2.96 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	241	<div><div>8%</div><div>58%</div><div>28%</div><div>••</div><div>11%</div></div>
1	B	241	<div><div>15%</div><div>60%</div><div>25%</div><div>•</div><div>13%</div></div>
1	C	241	<div><div>2%</div><div>49%</div><div>13%</div><div>•</div><div>36%</div></div>
1	D	241	<div><div>6%</div><div>45%</div><div>15%</div><div>•</div><div>38%</div></div>
2	E	17	<div><div>6%</div><div>24%</div><div>29%</div><div>6%</div><div>41%</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	17	<div><div><div></div><div></div><div></div><div></div><div></div></div><div><div>29%</div><div>18%</div><div>29%</div><div>6%</div><div>47%</div></div></div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6138 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 E3 ubiquitin-protein ligase UHRF1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	215	Total	C	N	O	S	4	0	0
			1753	1087	315	332	19			
1	B	210	Total	C	N	O	S	0	0	0
			1708	1057	307	325	19			
1	C	154	Total	C	N	O	S	30	0	0
			1267	790	229	241	7			
1	D	150	Total	C	N	O	S	32	0	0
			1235	770	224	235	6			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	126	GLY	-	expression tag	UNP Q96T88
A	127	PRO	-	expression tag	UNP Q96T88
A	128	LEU	-	expression tag	UNP Q96T88
A	129	GLY	-	expression tag	UNP Q96T88
A	130	SER	-	expression tag	UNP Q96T88
A	131	PRO	-	expression tag	UNP Q96T88
A	132	GLU	-	expression tag	UNP Q96T88
A	133	PHE	-	expression tag	UNP Q96T88
B	126	GLY	-	expression tag	UNP Q96T88
B	127	PRO	-	expression tag	UNP Q96T88
B	128	LEU	-	expression tag	UNP Q96T88
B	129	GLY	-	expression tag	UNP Q96T88
B	130	SER	-	expression tag	UNP Q96T88
B	131	PRO	-	expression tag	UNP Q96T88
B	132	GLU	-	expression tag	UNP Q96T88
B	133	PHE	-	expression tag	UNP Q96T88
C	126	GLY	-	expression tag	UNP Q96T88
C	127	PRO	-	expression tag	UNP Q96T88
C	128	LEU	-	expression tag	UNP Q96T88
C	129	GLY	-	expression tag	UNP Q96T88
C	130	SER	-	expression tag	UNP Q96T88

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	131	PRO	-	expression tag	UNP Q96T88
C	132	GLU	-	expression tag	UNP Q96T88
C	133	PHE	-	expression tag	UNP Q96T88
D	126	GLY	-	expression tag	UNP Q96T88
D	127	PRO	-	expression tag	UNP Q96T88
D	128	LEU	-	expression tag	UNP Q96T88
D	129	GLY	-	expression tag	UNP Q96T88
D	130	SER	-	expression tag	UNP Q96T88
D	131	PRO	-	expression tag	UNP Q96T88
D	132	GLU	-	expression tag	UNP Q96T88
D	133	PHE	-	expression tag	UNP Q96T88

- Molecule 2 is a protein called Peptide from Histone H3.3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	10	Total	C	N	O	0	0	0
			82	49	19	14			
2	F	9	Total	C	N	O	0	0	0
			76	46	18	12			

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

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

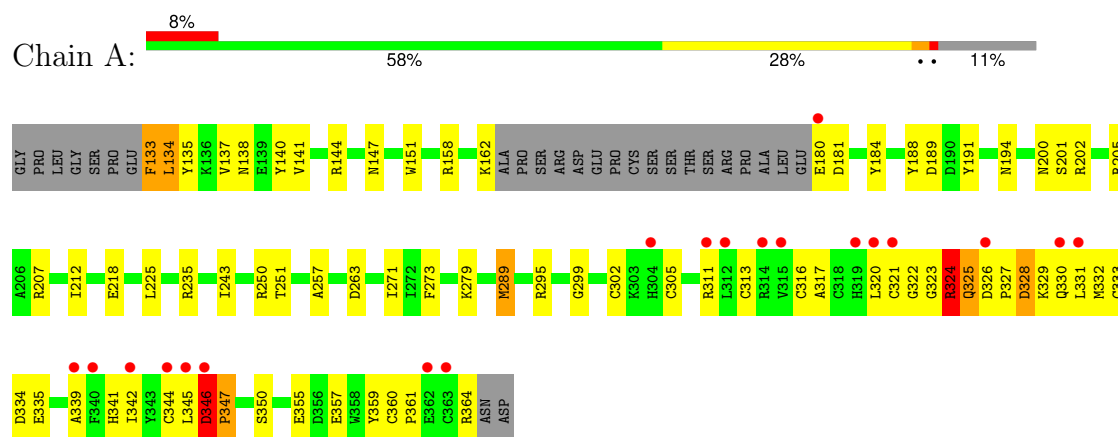
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total	O	0	0
			5	5		
4	B	2	Total	O	0	0
			2	2		
4	C	2	Total	O	0	0
			2	2		
4	D	1	Total	O	0	0
			1	1		

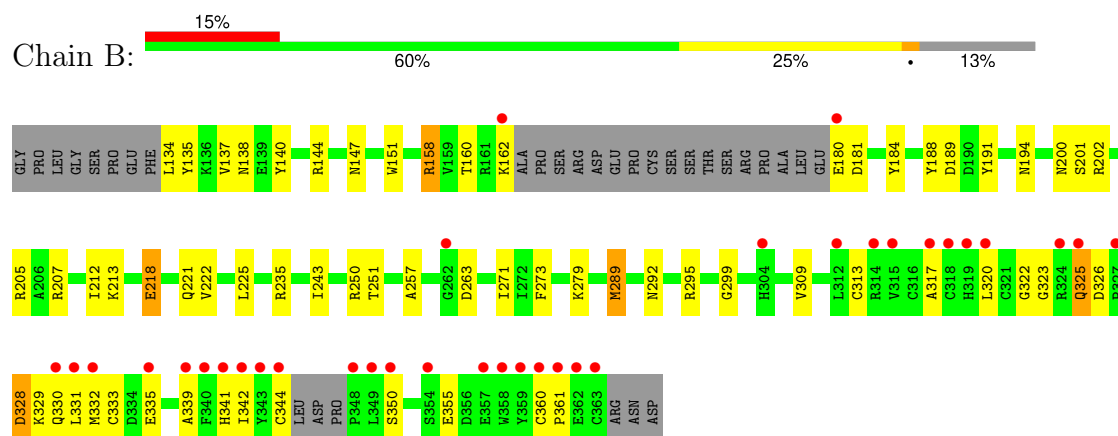
### 3 Residue-property plots [i](#)

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.

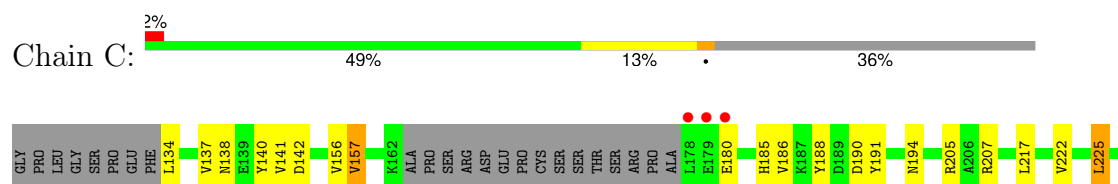
#### • Molecule 1: E3 ubiquitin-protein ligase UHRF1

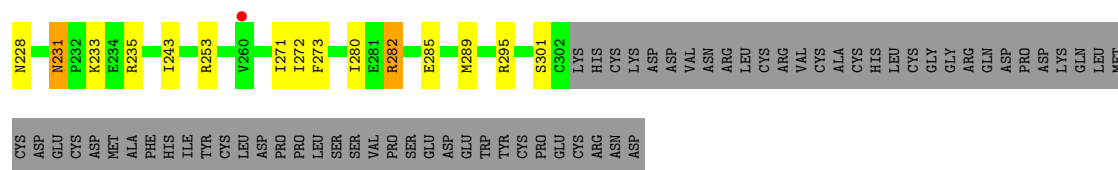


#### • Molecule 1: E3 ubiquitin-protein ligase UHRF1

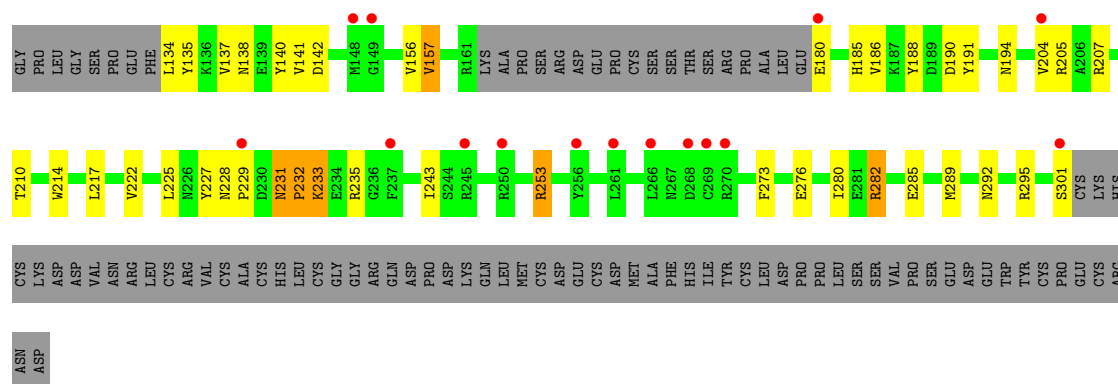


#### • Molecule 1: E3 ubiquitin-protein ligase UHRF1

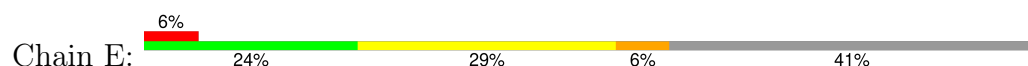




• Molecule 1: E3 ubiquitin-protein ligase UHRF1



• Molecule 2: Peptide from Histone H3.3



• Molecule 2: Peptide from Histone H3.3



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	145.94Å 145.94Å 125.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.38 – 2.96 45.38 – 2.96	Depositor EDS
% Data completeness (in resolution range)	95.9 (45.38-2.96) 99.6 (45.38-2.96)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.39 (at 2.96Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.1_353)	Depositor
R, $R_{free}$	0.255 , 0.288 0.255 , 0.291	Depositor DCC
$R_{free}$ test set	1471 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	86.6	Xtriage
Anisotropy	0.010	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 48.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	6138	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	97.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.82% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: M3L, ZN

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	0.21	0/1791	0.71	9/2421 (0.4%)
1	B	0.25	0/1743	0.65	6/2353 (0.3%)
1	C	0.19	0/1291	0.37	0/1744
1	D	0.24	0/1259	0.39	0/1702
2	E	0.19	0/69	0.41	0/90
2	F	0.19	0/63	0.42	0/82
All	All	0.23	0/6216	0.57	15/8392 (0.2%)

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	2

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	207	ARG	NE-CZ-NH2	-12.40	114.10	120.30
1	B	207	ARG	NE-CZ-NH1	-12.39	114.10	120.30
1	B	207	ARG	NE-CZ-NH2	12.29	126.45	120.30
1	A	202	ARG	NE-CZ-NH2	-11.94	114.33	120.30
1	B	202	ARG	NE-CZ-NH1	-11.80	114.40	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	324	ARG	Sidechain
1	A	346	ASP	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	1753	0	1665	51	1
1	B	1708	0	1621	42	0
1	C	1267	0	1230	18	0
1	D	1235	0	1195	29	0
2	E	82	0	97	6	0
2	F	76	0	91	6	0
3	A	3	0	0	0	0
3	B	3	0	0	0	0
3	C	1	0	0	0	0
4	A	5	0	0	0	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
4	D	1	0	0	0	0
All	All	6138	0	5899	142	1

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.

The worst 5 of 142 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:D:235:ARG:NH1	1:D:273:PHE:CE2	2.05	1.25
1:B:158:ARG:HH21	1:B:158:ARG:CG	1.57	1.14
1:B:158:ARG:HG3	1:B:158:ARG:NH2	1.43	1.12
1:A:360:CYS:HB2	1:A:361:PRO:HD2	1.69	0.74
1:D:229:PRO:O	1:D:232:PRO:HD3	1.88	0.73

All (1) 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:158:ARG:NH2	1:A:364:ARG:NH1[7_555]	2.01	0.19

### 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	211/241 (88%)	186 (88%)	20 (10%)	5 (2%)	6	26
1	B	204/241 (85%)	183 (90%)	19 (9%)	2 (1%)	15	48
1	C	150/241 (62%)	136 (91%)	14 (9%)	0	100	100
1	D	146/241 (61%)	131 (90%)	14 (10%)	1 (1%)	22	56
2	E	7/17 (41%)	7 (100%)	0	0	100	100
2	F	7/17 (41%)	6 (86%)	1 (14%)	0	100	100
All	All	725/998 (73%)	649 (90%)	68 (9%)	8 (1%)	14	46

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	134	LEU
1	A	346	ASP
1	A	347	PRO
1	D	232	PRO
1	A	263	ASP

#### 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	195/217 (90%)	181 (93%)	14 (7%)	14	42
1	B	190/217 (88%)	176 (93%)	14 (7%)	13	41
1	C	138/217 (64%)	130 (94%)	8 (6%)	20	51
1	D	134/217 (62%)	125 (93%)	9 (7%)	16	45
2	E	7/11 (64%)	7 (100%)	0	100	100
2	F	6/11 (54%)	5 (83%)	1 (17%)	2	9
All	All	670/890 (75%)	624 (93%)	46 (7%)	15	44

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	134	LEU
1	C	301	SER
1	C	157	VAL
1	C	231	ASN
1	D	157	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	198	GLN
2	F	5	GLN
1	D	267	ASN
1	B	325	GLN
1	D	194	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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	F	9	2	10,11,12	0.98	0	9,14,16	1.16	1 (11%)
2	M3L	E	9	2	10,11,12	0.96	0	9,14,16	1.16	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	F	9	2	-	1/9/10/12	-
2	M3L	E	9	2	-	3/9/10/12	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	9	M3L	CM1-NZ-CE	2.28	118.98	109.91
2	F	9	M3L	CM1-NZ-CE	2.28	118.96	109.91

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	9	M3L	C-CA-CB-CG
2	F	9	M3L	C-CA-CB-CG
2	E	9	M3L	N-CA-CB-CG
2	E	9	M3L	CE-CD-CG-CB

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	9	M3L	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	215/241 (89%)	0.70	20 (9%) 8 5	51, 84, 147, 170	1 (0%)
1	B	210/241 (87%)	1.04	35 (16%) 1 1	51, 87, 189, 199	0
1	C	154/241 (63%)	0.34	4 (2%) 56 39	58, 87, 136, 170	7 (4%)
1	D	150/241 (62%)	0.58	15 (10%) 7 4	58, 94, 131, 184	8 (5%)
2	E	9/17 (52%)	0.47	1 (11%) 5 3	87, 90, 114, 120	0
2	F	8/17 (47%)	3.51	5 (62%) 0 0	115, 138, 173, 176	0
All	All	746/998 (74%)	0.72	80 (10%) 6 3	51, 89, 173, 199	16 (2%)

The worst 5 of 80 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	330	GLN	11.4
1	A	320	LEU	8.3
1	B	314	ARG	7.8
2	F	1	ALA	7.7
1	B	320	LEU	7.1

### 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. The B-factors column lists the minimum, median, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	M3L	F	9	12/13	0.88	0.33	60,101,114,115	0
2	M3L	E	9	12/13	0.95	0.24	55,70,111,115	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	ZN	B	401	1/1	0.70	0.07	190,190,190,190	0
3	ZN	A	402	1/1	0.74	0.08	143,143,143,143	0
3	ZN	A	401	1/1	0.81	0.15	108,108,108,108	0
3	ZN	A	403	1/1	0.86	0.12	121,121,121,121	0
3	ZN	B	402	1/1	0.93	0.09	181,181,181,181	0
3	ZN	B	403	1/1	0.95	0.12	202,202,202,202	0
3	ZN	C	401	1/1	0.95	0.18	115,115,115,115	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.