



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2020 – 07:06 PM BST

PDB ID : 6Z6B  
Title : Structure of full-length La Crosse virus L protein (polymerase)  
Authors : Cusack, S.; Gerlach, P.; Reguera, J.  
Deposited on : 2020-05-28  
Resolution : 3.96 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.13  
buster-report : 1.1.7 (2018)  
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) : 2.13

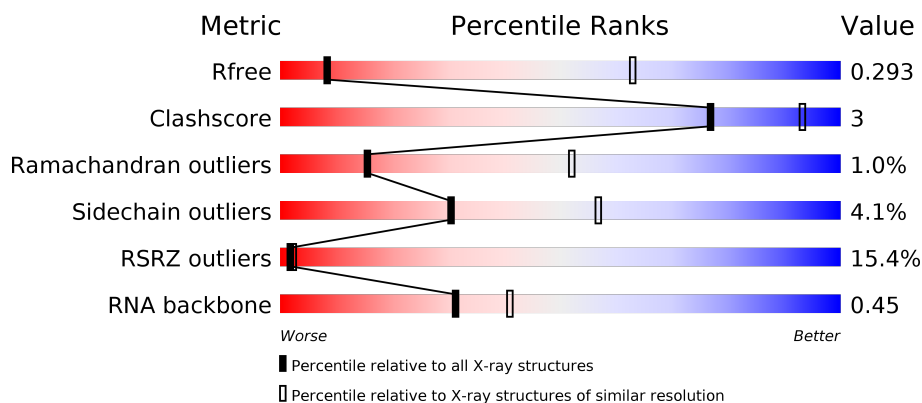
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.96 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	1025 (4.22-3.70)
Clashscore	141614	1085 (4.22-3.70)
Ramachandran outliers	138981	1047 (4.22-3.70)
Sidechain outliers	138945	1039 (4.22-3.70)
RSRZ outliers	127900	1013 (4.28-3.64)
RNA backbone	3102	1041 (4.84-3.00)

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  $\geq 3$ , 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	BBB	10	<div> <div>30%</div> <div>50%</div> <div>50%</div> </div>
1	UUU	10	<div> <div>60%</div> <div>40%</div> </div>
2	CCC	16	<div> <div>19%</div> <div>81%</div> <div>19%</div> </div>
2	HHH	16	<div> <div>13%</div> <div>75%</div> <div>19%</div> <div>6%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	DDD	8	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> <div>13%88%13%</div>
3	XXX	8	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> <div>25%75%25%</div>
4	EEE	2285	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> <div>15%78%10%•11%</div>
4	PPP	2285	<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> <div>12%79%10%•11%</div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 34640 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 RNA chain called RNA (5'-R(P\*AP\*GP\*UP\*AP\*GP\*UP\*GP\*UP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	BBB	10	Total	C	N	O	P	0	0	0
			216	96	39	71	10			
1	UUU	10	Total	C	N	O	P	0	0	0
			216	96	39	71	10			

- Molecule 2 is a RNA chain called RNA (5'-R(P\*UP\*UP\*AP\*GP\*UP\*AP\*GP\*UP\*AP\*CP\*AP\*CP\*UP\*AP\*CP\*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	CCC	16	Total	C	N	O	P	0	0	0
			336	151	56	113	16			
2	HHH	16	Total	C	N	O	P	0	0	0
			336	151	56	113	16			

- Molecule 3 is a RNA chain called RNA (5'-R(\*GP\*CP\*UP\*AP\*CP\*UP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	DDD	8	Total	C	N	O	P	0	0	0
			165	76	30	52	7			
3	XXX	8	Total	C	N	O	P	0	0	0
			165	76	30	52	7			

- Molecule 4 is a protein called RNA-directed RNA polymerase L.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	EEE	2031	Total	C	N	O	S	0	0	0
			16582	10626	2758	3090	108			
4	PPP	2036	Total	C	N	O	S	0	0	0
			16622	10651	2766	3096	109			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
EEE	-21	MET	-	initiating methionine	UNP A5HC98
EEE	-20	GLY	-	expression tag	UNP A5HC98
EEE	-19	HIS	-	expression tag	UNP A5HC98
EEE	-18	HIS	-	expression tag	UNP A5HC98
EEE	-17	HIS	-	expression tag	UNP A5HC98
EEE	-16	HIS	-	expression tag	UNP A5HC98
EEE	-15	HIS	-	expression tag	UNP A5HC98
EEE	-14	HIS	-	expression tag	UNP A5HC98
EEE	-13	ASP	-	expression tag	UNP A5HC98
EEE	-12	TYR	-	expression tag	UNP A5HC98
EEE	-11	ASP	-	expression tag	UNP A5HC98
EEE	-10	ILE	-	expression tag	UNP A5HC98
EEE	-9	PRO	-	expression tag	UNP A5HC98
EEE	-8	THR	-	expression tag	UNP A5HC98
EEE	-7	THR	-	expression tag	UNP A5HC98
EEE	-6	GLU	-	expression tag	UNP A5HC98
EEE	-5	ASN	-	expression tag	UNP A5HC98
EEE	-4	LEU	-	expression tag	UNP A5HC98
EEE	-3	TYR	-	expression tag	UNP A5HC98
EEE	-2	PHE	-	expression tag	UNP A5HC98
EEE	-1	GLN	-	expression tag	UNP A5HC98
EEE	0	GLY	-	expression tag	UNP A5HC98
PPP	-21	MET	-	initiating methionine	UNP A5HC98
PPP	-20	GLY	-	expression tag	UNP A5HC98
PPP	-19	HIS	-	expression tag	UNP A5HC98
PPP	-18	HIS	-	expression tag	UNP A5HC98
PPP	-17	HIS	-	expression tag	UNP A5HC98
PPP	-16	HIS	-	expression tag	UNP A5HC98
PPP	-15	HIS	-	expression tag	UNP A5HC98
PPP	-14	HIS	-	expression tag	UNP A5HC98
PPP	-13	ASP	-	expression tag	UNP A5HC98
PPP	-12	TYR	-	expression tag	UNP A5HC98
PPP	-11	ASP	-	expression tag	UNP A5HC98
PPP	-10	ILE	-	expression tag	UNP A5HC98
PPP	-9	PRO	-	expression tag	UNP A5HC98
PPP	-8	THR	-	expression tag	UNP A5HC98
PPP	-7	THR	-	expression tag	UNP A5HC98
PPP	-6	GLU	-	expression tag	UNP A5HC98
PPP	-5	ASN	-	expression tag	UNP A5HC98
PPP	-4	LEU	-	expression tag	UNP A5HC98
PPP	-3	TYR	-	expression tag	UNP A5HC98
PPP	-2	PHE	-	expression tag	UNP A5HC98
PPP	-1	GLN	-	expression tag	UNP A5HC98

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
PPP	0	GLY	-	expression tag	UNP A5HC98

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by author).

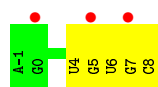
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	PPP	1	Total	Zn	0	0
			1	1		
5	EEE	1	Total	Zn	0	0
			1	1		

### 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: RNA (5'-R(P\*AP\*GP\*UP\*AP\*GP\*UP\*GP\*UP\*GP\*C)-3')

Chain BBB: 




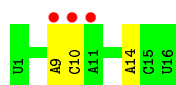
- Molecule 1: RNA (5'-R(P\*AP\*GP\*UP\*AP\*GP\*UP\*GP\*UP\*GP\*C)-3')

Chain UUU: 




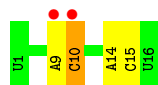
- Molecule 2: RNA (5'-R(P\*UP\*UP\*AP\*GP\*UP\*AP\*GP\*UP\*AP\*CP\*AP\*CP\*UP\*AP\*CP\*U)-3')

Chain CCC: 




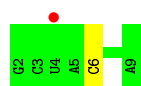
- Molecule 2: RNA (5'-R(P\*UP\*UP\*AP\*GP\*UP\*AP\*GP\*UP\*AP\*CP\*AP\*CP\*UP\*AP\*CP\*U)-3')

Chain HHH: 

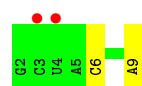
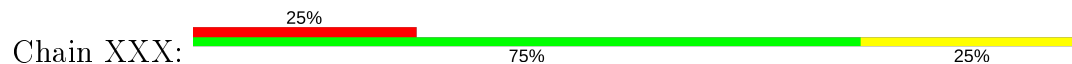


- Molecule 3: RNA (5'-R(\*GP\*CP\*UP\*AP\*CP\*UP\*AP\*A)-3')

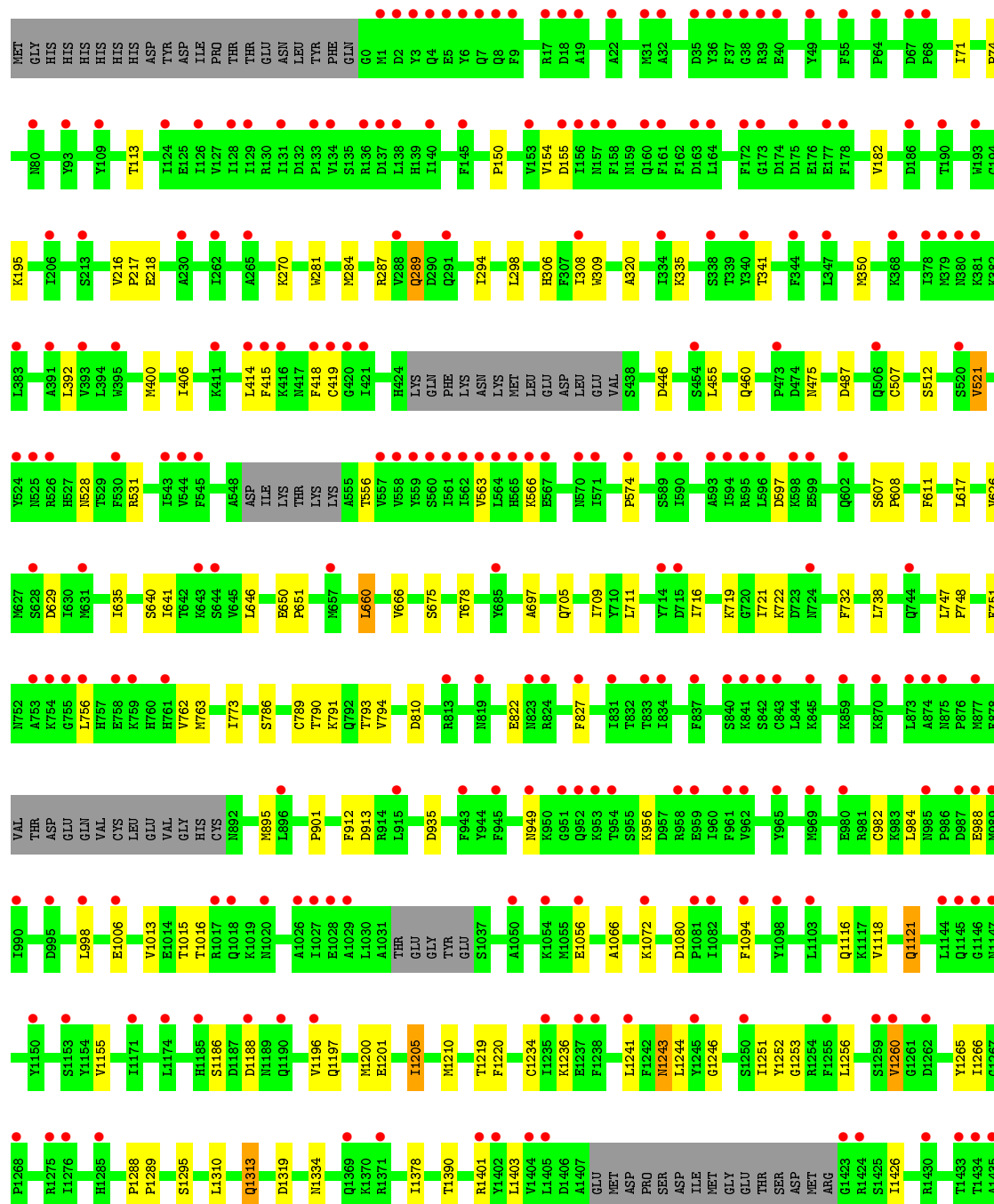
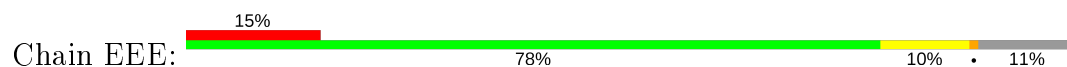
Chain DDD: 



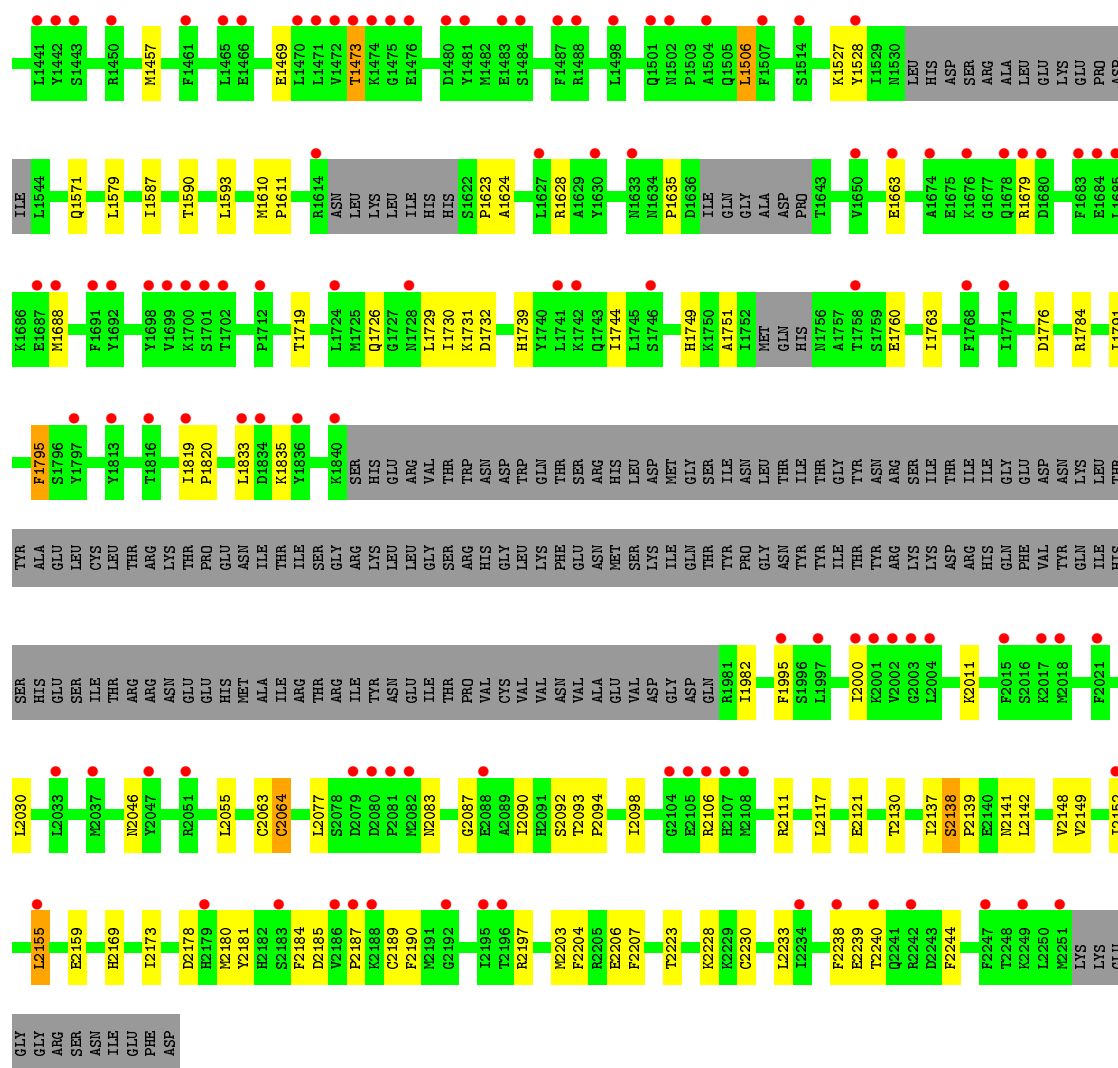
- Molecule 3: RNA (5'-R(\*GP\*CP\*UP\*AP\*CP\*UP\*AP\*A)-3')



- Molecule 4: RNA-directed RNA polymerase L









## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	371.19Å 145.32Å 234.19Å 90.00° 116.33° 90.00°	Depositor
Resolution (Å)	209.89 – 3.96 209.89 – 3.96	Depositor EDS
% Data completeness (in resolution range)	60.8 (209.89-3.96) 60.8 (209.89-3.96)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.79 (at 4.02Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, $R_{free}$	0.264 , 0.299 0.261 , 0.293	Depositor DCC
$R_{free}$ test set	1514 reflections (2.56%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	167.7	Xtriage
Anisotropy	0.066	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 209.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	34640	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	205.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.77% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	BBB	0.24	0/241	0.67	0/374
1	UUU	0.23	0/241	0.63	0/374
2	CCC	0.22	0/374	0.64	0/579
2	HHH	0.23	0/374	0.64	0/579
3	DDD	0.18	0/184	0.61	0/285
3	XXX	0.18	0/184	0.62	0/285
4	EEE	0.67	0/16918	0.70	0/22801
4	PPP	0.67	0/16960	0.70	0/22858
All	All	0.66	0/35476	0.70	0/48135

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 [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	BBB	216	0	108	0	0
1	UUU	216	0	108	1	0
2	CCC	336	0	171	0	0
2	HHH	336	0	171	2	0
3	DDD	165	0	85	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	XXX	165	0	85	0	0
4	EEE	16582	0	16651	104	0
4	PPP	16622	0	16692	105	0
5	EEE	1	0	0	0	0
5	PPP	1	0	0	0	0
All	All	34640	0	34071	203	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.

The worst 5 of 203 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)
4:EEE:716:ILE:HB	4:EEE:719:LYS:HE3	1.40	1.01
4:EEE:716:ILE:HB	4:EEE:719:LYS:CE	1.96	0.94
4:EEE:721:ILE:HG21	4:EEE:738:LEU:HD23	1.62	0.82
4:PPP:2152:ILE:HD11	4:PPP:2226:PHE:CZ	2.18	0.79
4:PPP:1522:SER:O	4:PPP:1526:ASP:HB2	1.85	0.76

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	
4	EEE	2009/2285 (88%)	1808 (90%)	180 (9%)	21 (1%)	15	52
4	PPP	2016/2285 (88%)	1812 (90%)	183 (9%)	21 (1%)	15	52
All	All	4025/4570 (88%)	3620 (90%)	363 (9%)	42 (1%)	15	52

5 of 42 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	EEE	1623	PRO
4	PPP	1623	PRO
4	EEE	270	LYS
4	EEE	678	THR
4	EEE	1234	CYS

### 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	
4	EEE	1873/2107 (89%)	1795 (96%)	78 (4%)	30	56
4	PPP	1877/2107 (89%)	1801 (96%)	76 (4%)	31	57
All	All	3750/4214 (89%)	3596 (96%)	154 (4%)	30	57

5 of 154 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	EEE	2093	THR
4	PPP	455	LEU
4	PPP	1835	LYS
4	EEE	2117	LEU
4	PPP	71	ILE

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

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	BBB	9/10 (90%)	5 (55%)	1 (11%)
1	UUU	9/10 (90%)	3 (33%)	0
2	CCC	15/16 (93%)	3 (20%)	0
2	HHH	15/16 (93%)	3 (20%)	0
3	DDD	7/8 (87%)	1 (14%)	0
3	XXX	7/8 (87%)	2 (28%)	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
All	All	62/68 (91%)	17 (27%)	1 (1%)

5 of 17 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	BBB	4	U
1	BBB	5	G
1	BBB	6	U
1	BBB	7	G
1	BBB	8	C

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	BBB	5	G

## 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 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 ⓘ

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<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	BBB	10/10 (100%)	1.54	3 (30%) 0 0	153, 182, 218, 228	0
1	UUU	10/10 (100%)	1.19	0 100 100	122, 141, 171, 201	0
2	CCC	16/16 (100%)	1.11	3 (18%) 1 1	113, 200, 291, 292	0
2	HHH	16/16 (100%)	0.99	2 (12%) 3 5	93, 168, 206, 212	0
3	DDD	8/8 (100%)	0.46	1 (12%) 3 5	190, 224, 238, 244	0
3	XXX	8/8 (100%)	0.94	2 (25%) 0 0	145, 176, 201, 214	0
4	EEE	2031/2285 (88%)	0.96	346 (17%) 1 2	87, 211, 334, 415	0
4	PPP	2036/2285 (89%)	0.89	279 (13%) 3 3	80, 181, 312, 403	0
All	All	4135/4638 (89%)	0.93	636 (15%) 2 2	80, 193, 325, 415	0

The worst 5 of 636 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	EEE	1679	ARG	13.4
4	EEE	2188	LYS	9.2
4	EEE	1471	LEU	9.0
4	EEE	1501	GLN	9.0
4	EEE	157	ASN	8.7

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

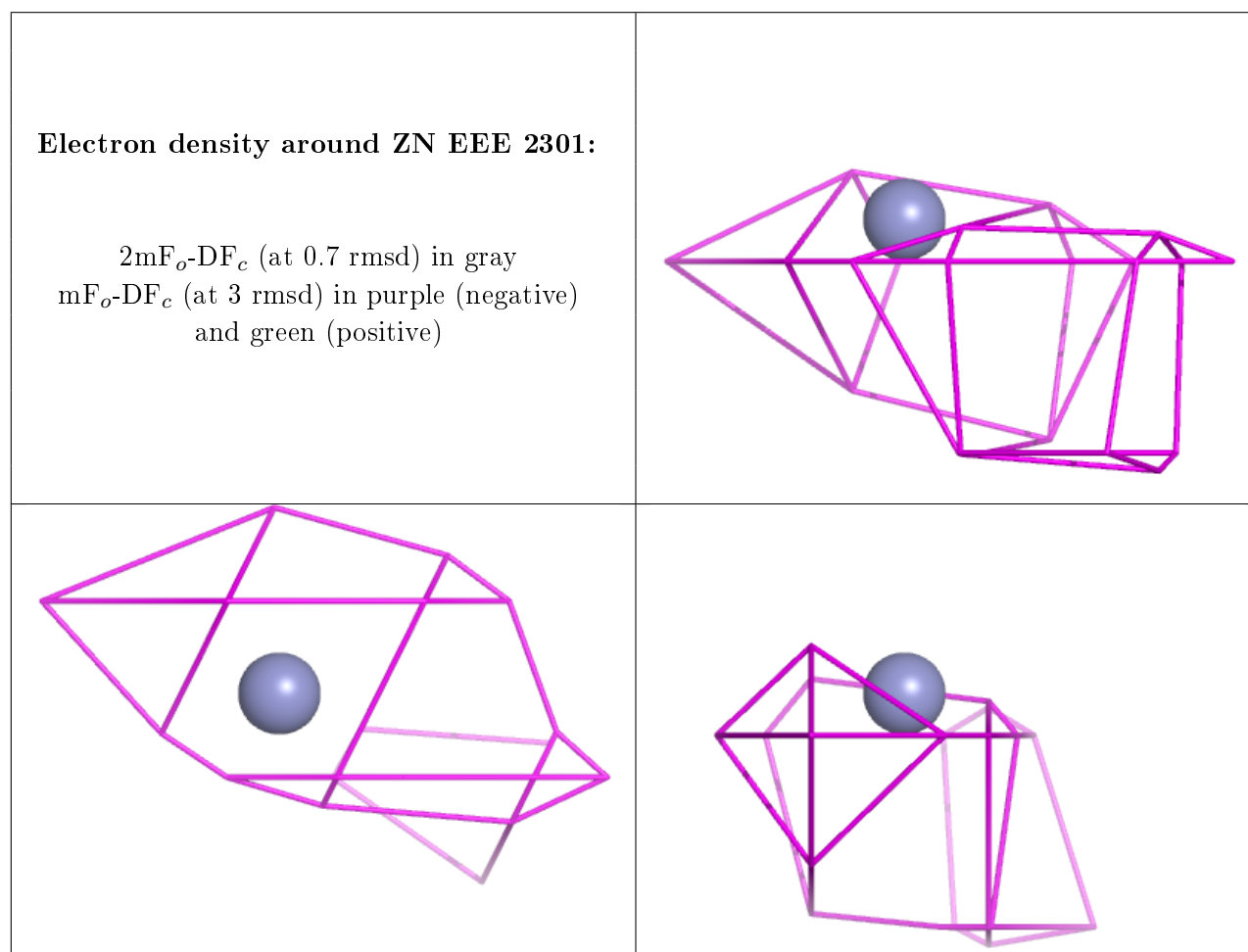
There are no monosaccharides in this entry.

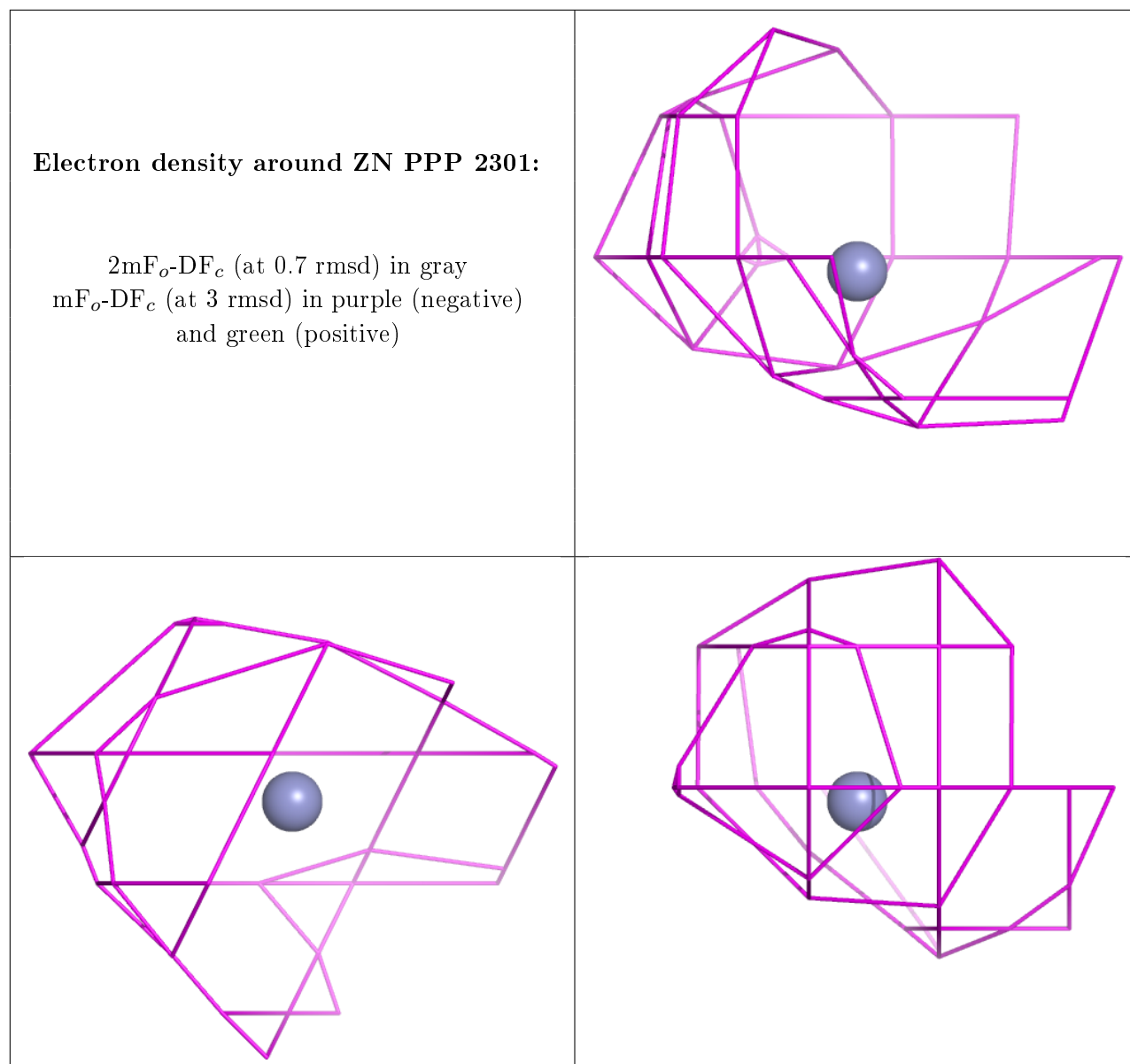
## 6.4 Ligands ⓘ

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<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
5	ZN	EEE	2301	1/1	0.99	0.15	189,189,189,189	0
5	ZN	PPP	2301	1/1	0.99	0.23	117,117,117,117	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.





## 6.5 Other polymers [i](#)

There are no such residues in this entry.