



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 01:49 am BST

PDB ID : 4P0S  
Title : human Mus81-Eme1-3'flap DNA complex  
Authors : Gwon, G.H.; Baek, K.; Cho, Y.  
Deposited on : 2014-02-22  
Resolution : 6.00 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	:	2.11
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

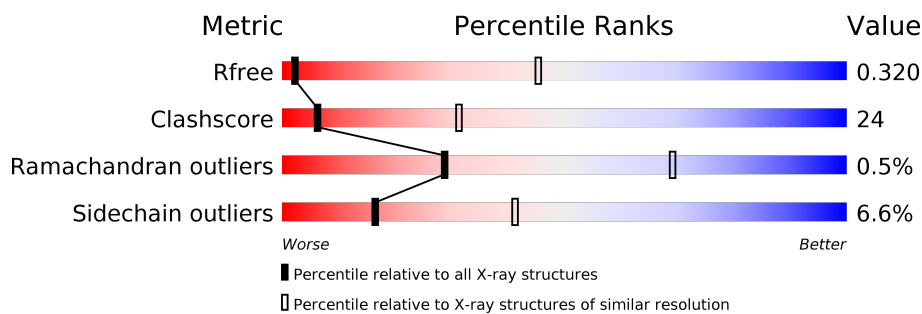
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 6.00 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.




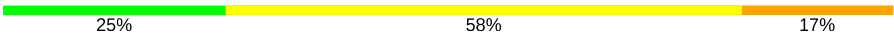
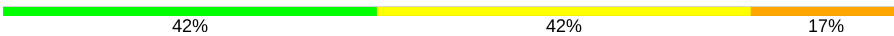
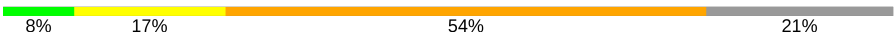


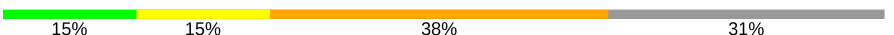

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1000 (8.00-3.88)
Clashscore	141614	1049 (8.00-3.90)
Ramachandran outliers	138981	1016 (8.00-3.86)
Sidechain outliers	138945	1017 (8.00-3.82)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	306	52% 34% • 10%
1	C	306	53% 32% • 10%
1	E	306	53% 35% • 10%
1	G	306	49% 36% 5% 10%
2	B	393	41% 27% 5% 28%
2	D	393	43% 25% • 28%
2	F	393	42% 25% 5% 28%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	H	393	
3	I	12	
3	M	12	
3	Q	12	
3	U	12	
4	J	24	
4	N	24	
4	R	24	
4	V	24	
5	L	13	
5	P	13	
5	T	13	
5	X	13	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 20884 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 Crossover junction endonuclease MUS81.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	275	Total	C	N	O	S	0	0	0
			2173	1364	403	398	8			
1	C	275	Total	C	N	O	S	0	0	0
			2173	1364	403	398	8			
1	E	275	Total	C	N	O	S	0	0	0
			2173	1364	403	398	8			
1	G	275	Total	C	N	O	S	0	0	0
			2173	1364	403	398	8			

- Molecule 2 is a protein called Crossover junction endonuclease EME1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	284	Total	C	N	O	S	0	0	0
			2227	1401	397	415	14			
2	D	284	Total	C	N	O	S	0	0	0
			2227	1401	397	415	14			
2	F	284	Total	C	N	O	S	0	0	0
			2227	1401	397	415	14			
2	H	284	Total	C	N	O	S	0	0	0
			2227	1401	397	415	14			

- Molecule 3 is a DNA chain called DNA GAATGTGTGTCT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	12	Total	C	N	O	P	0	0	0
			249	119	43	75	12			
3	M	12	Total	C	N	O	P	0	0	0
			249	119	43	75	12			
3	Q	12	Total	C	N	O	P	0	0	0
			249	119	43	75	12			
3	U	12	Total	C	N	O	P	0	0	0
			249	119	43	75	12			

- Molecule 4 is a DNA chain called DNA TAGACACACATTTCGGGACATGCAG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	J	19	Total	C	N	O	P	0	0	0
			389	185	76	109	19			
4	N	19	Total	C	N	O	P	0	0	0
			389	185	76	109	19			
4	R	19	Total	C	N	O	P	0	0	0
			389	185	76	109	19			
4	V	19	Total	C	N	O	P	0	0	0
			389	185	76	109	19			

- Molecule 5 is a DNA chain called DNA TCTGCATGTCATT.

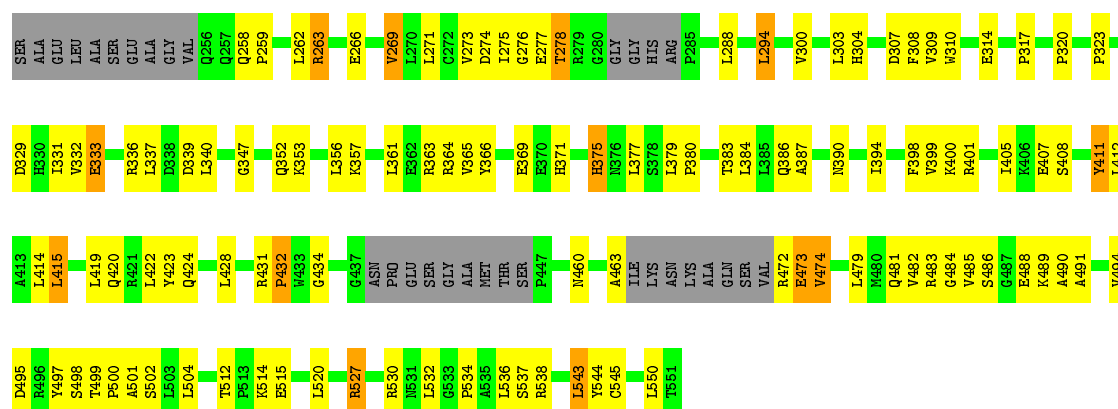
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	L	9	Total	C	N	O	P	0	0	0
			183	88	29	57	9			
5	P	9	Total	C	N	O	P	0	0	0
			183	88	29	57	9			
5	T	9	Total	C	N	O	P	0	0	0
			183	88	29	57	9			
5	X	9	Total	C	N	O	P	0	0	0
			183	88	29	57	9			

### 3 Residue-property plots

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. 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. 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.

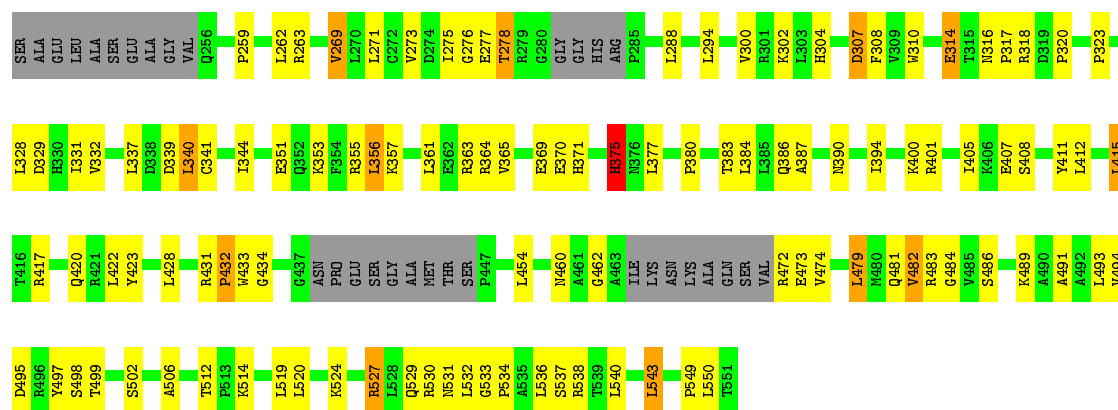
#### • Molecule 1: Crossover junction endonuclease MUS81

Chain A: 



#### • Molecule 1: Crossover junction endonuclease MUS81

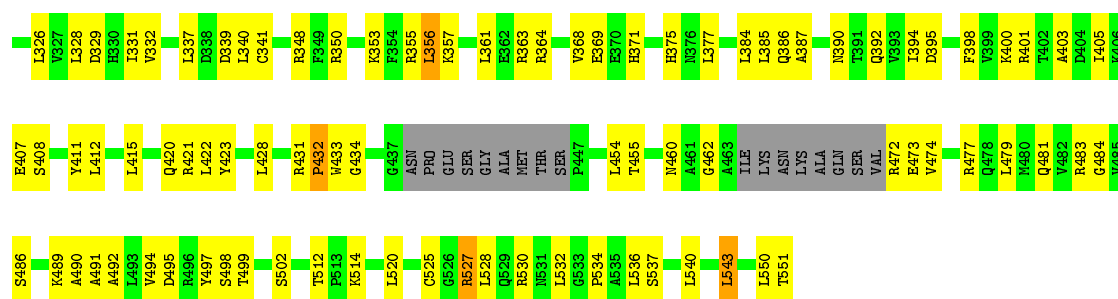
Chain C: 



#### • Molecule 1: Crossover junction endonuclease MUS81

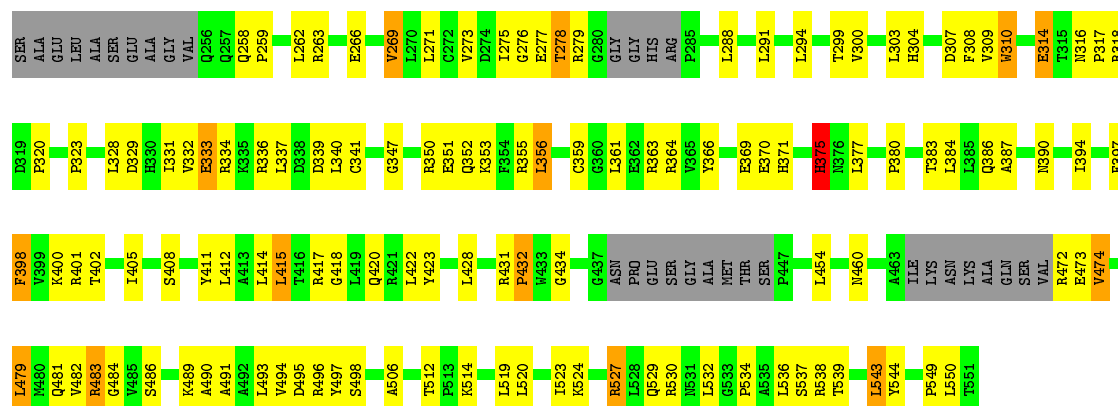
Chain E: 





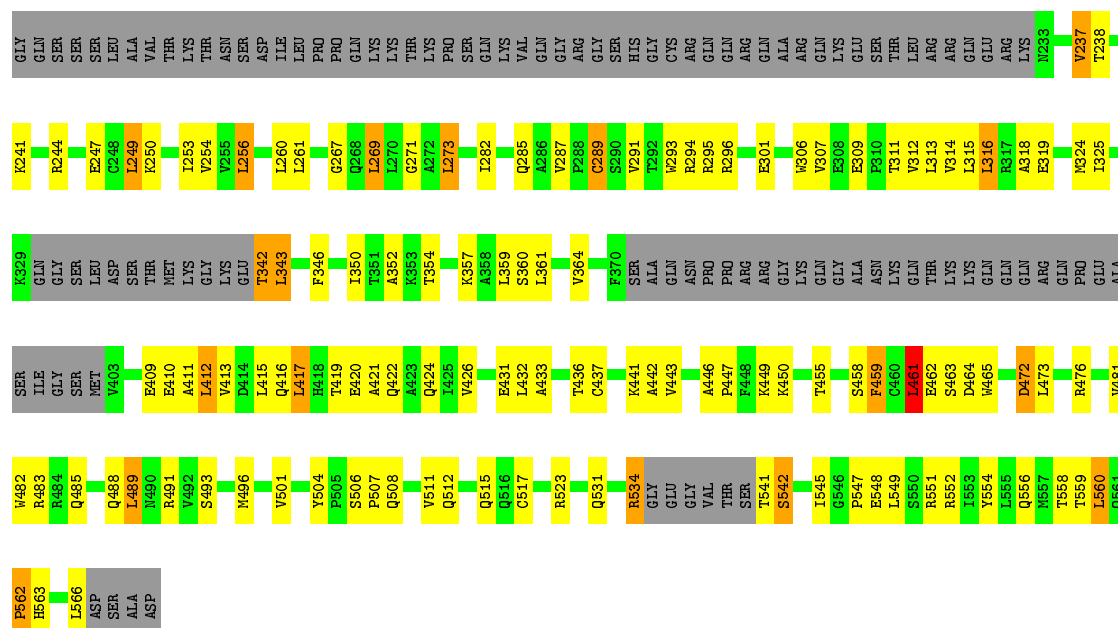
• Molecule 1: Crossover junction endonuclease MUS81

Chain G: 49% 36% 5% 10%



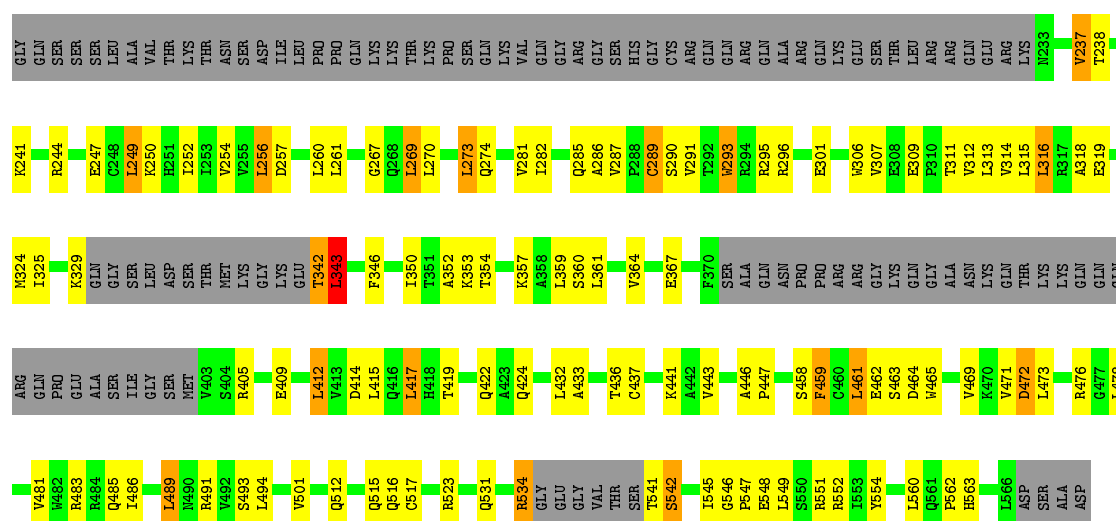
• Molecule 2: Crossover junction endonuclease EME1

Chain B: 41% 27% 5% 28%



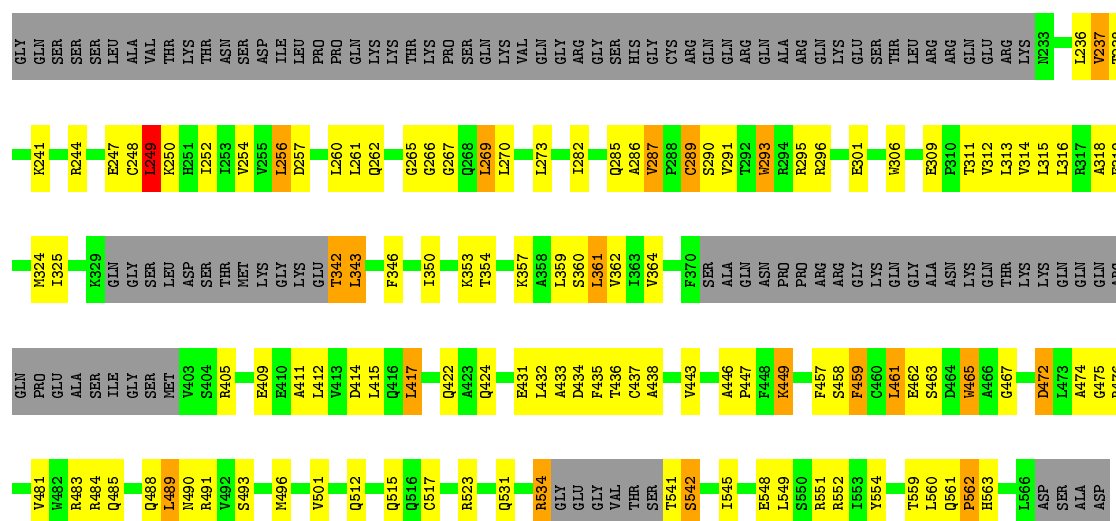
• Molecule 2: Crossover junction endonuclease EME1

Chain D: 



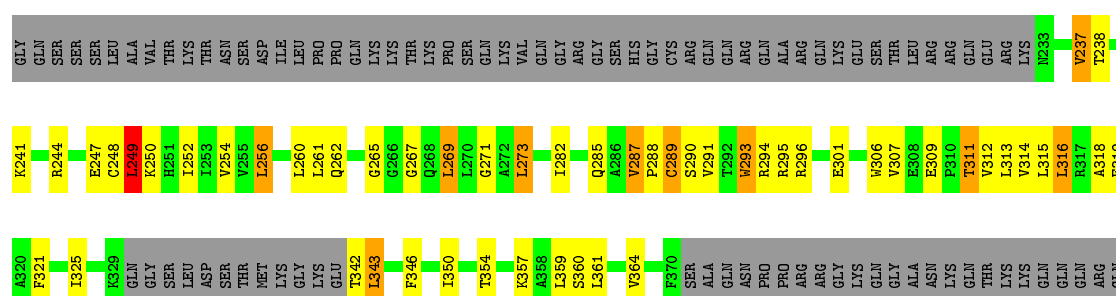
• Molecule 2: Crossover junction endonuclease EME1

Chain F: 

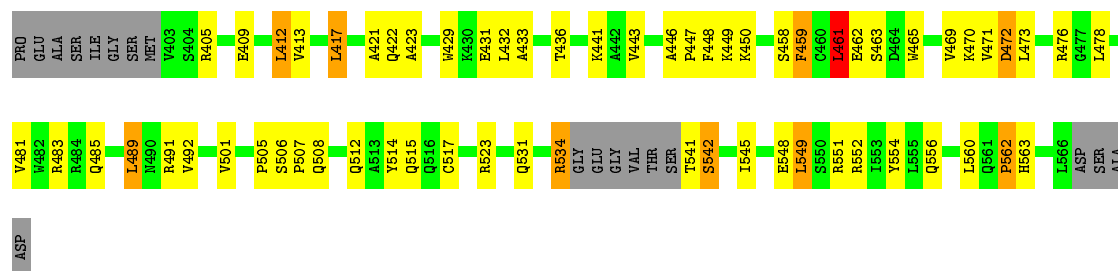


• Molecule 2: Crossover junction endonuclease EME1

Chain H: 



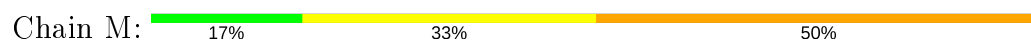




- Molecule 3: DNA GAATGTGTGTCT



- Molecule 3: DNA GAATGTGTGTCT



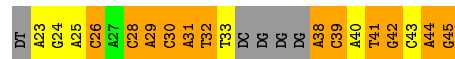
- Molecule 3: DNA GAATGTGTGTCT



- Molecule 4: DNA TAGACACACATTCGGGACATGCAG



- Molecule 4: DNA TAGACACACATTCGGGACATGCAG



- Molecule 4: DNA TAGACACACATTCGGGACATGCAG



- Molecule 4: DNA TAGACACACATTCGGGACATGCAG

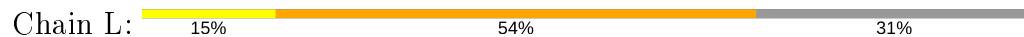




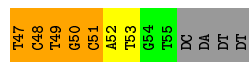
- Molecule 4: DNA TAGACACACATTCGGGACATGCAG



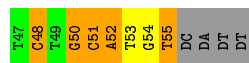
- Molecule 5: DNA TCTGCATGTCATT



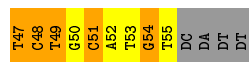
- Molecule 5: DNA TCTGCATGTCATT



- Molecule 5: DNA TCTGCATGTCATT



- Molecule 5: DNA TCTGCATGTCATT



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.44Å 250.76Å 430.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.98 – 6.00 49.43 – 5.48	Depositor EDS
% Data completeness (in resolution range)	99.4 (29.98-6.00) 99.2 (49.43-5.48)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.09 (at 5.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.250 , 0.308 0.268 , 0.320	Depositor DCC
$R_{free}$ test set	853 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	323.6	Xtriage
Anisotropy	0.352	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 207.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	20884	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	152.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.41% 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 ⓘ

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	0.89	2/2211 (0.1%)	1.13	8/2989 (0.3%)
1	C	0.83	1/2211 (0.0%)	1.11	10/2989 (0.3%)
1	E	0.82	0/2211	1.13	10/2989 (0.3%)
1	G	0.93	6/2211 (0.3%)	1.13	11/2989 (0.4%)
2	B	0.85	3/2258 (0.1%)	1.13	13/3054 (0.4%)
2	D	0.88	4/2258 (0.2%)	1.13	15/3054 (0.5%)
2	F	0.83	4/2258 (0.2%)	1.10	14/3054 (0.5%)
2	H	0.87	3/2258 (0.1%)	1.13	14/3054 (0.5%)
3	I	1.10	0/278	1.82	8/428 (1.9%)
3	M	1.17	1/278 (0.4%)	1.91	12/428 (2.8%)
3	Q	0.84	0/278	1.50	4/428 (0.9%)
3	U	1.07	0/278	1.62	6/428 (1.4%)
4	J	1.46	2/436 (0.5%)	2.72	45/667 (6.7%)
4	N	1.35	3/436 (0.7%)	2.40	30/667 (4.5%)
4	R	1.05	0/436	2.05	18/667 (2.7%)
4	V	1.37	4/436 (0.9%)	2.55	35/667 (5.2%)
5	L	1.58	3/203 (1.5%)	2.93	28/311 (9.0%)
5	P	1.82	6/203 (3.0%)	2.63	19/311 (6.1%)
5	T	1.48	3/203 (1.5%)	2.43	13/311 (4.2%)
5	X	1.38	2/203 (1.0%)	2.40	15/311 (4.8%)
All	All	0.95	47/21544 (0.2%)	1.42	328/29796 (1.1%)

The worst 5 of 47 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	359	CYS	CB-SG	-8.35	1.68	1.82
2	F	409	GLU	CB-CG	8.26	1.67	1.52
5	X	47	DT	C1'-N1	8.22	1.59	1.49
5	T	55	DT	C1'-N1	7.82	1.59	1.49
5	P	49	DT	C5-C7	7.78	1.54	1.50

The worst 5 of 328 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	V	38	DA	O4'-C1'-N9	22.46	123.72	108.00
4	V	45	DG	O4'-C4'-C3'	-19.19	94.49	106.00
4	J	44	DA	O4'-C1'-N9	18.59	121.01	108.00
4	N	45	DG	O4'-C4'-C3'	-16.41	96.16	106.00
5	L	48	DC	O4'-C1'-N1	14.54	118.18	108.00

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	2173	0	2203	143	1
1	C	2173	0	2203	87	0
1	E	2173	0	2203	101	0
1	G	2173	0	2203	122	0
2	B	2227	0	2265	160	0
2	D	2227	0	2265	112	1
2	F	2227	0	2265	121	0
2	H	2227	0	2265	144	0
3	I	249	0	138	14	0
3	M	249	0	138	19	0
3	Q	249	0	138	17	0
3	U	249	0	138	10	0
4	J	389	0	214	33	0
4	N	389	0	214	34	0
4	R	389	0	214	33	0
4	V	389	0	214	31	0
5	L	183	0	104	26	0
5	P	183	0	104	12	0
5	T	183	0	104	21	0
5	X	183	0	104	28	0
All	All	20884	0	19696	971	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 24.

The worst 5 of 971 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:G:538:ARG:NH2	2:H:473:LEU:O	1.87	1.05
1:A:499:THR:HG21	2:B:560:LEU:HA	1.33	1.05
1:E:401:ARG:NH1	2:F:424:GLN:OE1	1.91	1.03
2:H:269:LEU:HD21	2:H:436:THR:HG21	1.41	1.02
1:C:276:GLY:H	1:C:278:THR:HG22	1.19	1.02

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:424:GLN:OE1	2:D:516:GLN:NE2[8_555]	2.11	0.09

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	267/306 (87%)	258 (97%)	6 (2%)	3 (1%)	14	52
1	C	267/306 (87%)	258 (97%)	6 (2%)	3 (1%)	14	52
1	E	267/306 (87%)	259 (97%)	6 (2%)	2 (1%)	22	63
1	G	267/306 (87%)	257 (96%)	7 (3%)	3 (1%)	14	52
2	B	276/393 (70%)	265 (96%)	11 (4%)	0	100	100
2	D	276/393 (70%)	265 (96%)	11 (4%)	0	100	100
2	F	276/393 (70%)	264 (96%)	12 (4%)	0	100	100
2	H	276/393 (70%)	264 (96%)	12 (4%)	0	100	100
All	All	2172/2796 (78%)	2090 (96%)	71 (3%)	11 (0%)	29	69

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	432	PRO
1	C	432	PRO
1	E	432	PRO
1	G	432	PRO
1	C	549	PRO

### 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	237/259 (92%)	223 (94%)	14 (6%)	19	45
1	C	237/259 (92%)	223 (94%)	14 (6%)	19	45
1	E	237/259 (92%)	222 (94%)	15 (6%)	18	43
1	G	237/259 (92%)	221 (93%)	16 (7%)	16	41
2	B	242/334 (72%)	225 (93%)	17 (7%)	15	40
2	D	242/334 (72%)	227 (94%)	15 (6%)	18	43
2	F	242/334 (72%)	225 (93%)	17 (7%)	15	40
2	H	242/334 (72%)	224 (93%)	18 (7%)	13	38
All	All	1916/2372 (81%)	1790 (93%)	126 (7%)	16	41

5 of 126 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	461	LEU
1	E	474	VAL
2	H	319	GLU
2	D	531	GLN
1	E	299	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	316	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	392	GLN
2	H	488	GLN
1	E	352	GLN
1	E	448	ASN

### 5.3.3 RNA [i](#)

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](#)

There are no ligands in this entry.

### 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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.