



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 09:06 am BST

PDB ID : 5Y7G  
Title : Crystal structure of paFAN1 bound to 1nt 5'flap DNA with gap  
Authors : Cho, Y.; Jin, H.  
Deposited on : 2017-08-17  
Resolution : 3.40 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

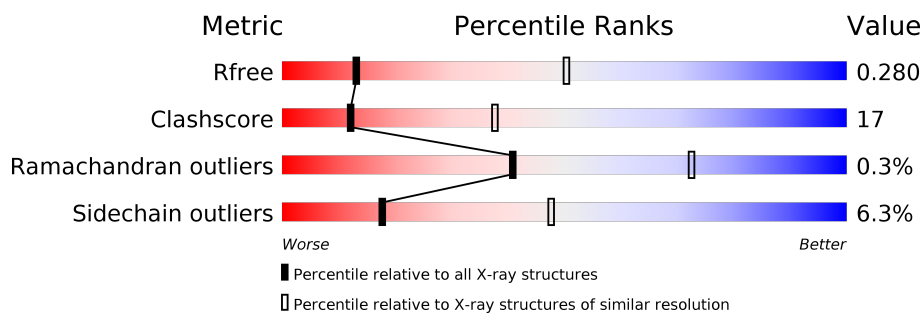
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.40 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)

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	580	
1	B	580	
1	C	580	
2	D	10	
2	F	10	
2	H	10	
3	E	24	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	24	<div><div></div><div>38%54%8%</div></div>
3	I	24	<div><div></div><div>29%71%</div></div>
4	J	14	<div><div></div><div>29%43%29%</div></div>
4	L	14	<div><div></div><div>14%71%14%</div></div>
4	N	14	<div><div></div><div>7%86%7%</div></div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 16164 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 Fanconi-associated nuclease 1 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	542	Total	C	N	O	S	0	0	0
			4406	2816	799	771	20			
1	B	543	Total	C	N	O	S	0	0	0
			4424	2827	800	777	20			
1	C	540	Total	C	N	O	S	0	0	0
			4401	2813	794	774	20			

There are 63 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	expression tag	UNP Q9I2N0
A	-19	GLY	-	expression tag	UNP Q9I2N0
A	-18	SER	-	expression tag	UNP Q9I2N0
A	-17	SER	-	expression tag	UNP Q9I2N0
A	-16	HIS	-	expression tag	UNP Q9I2N0
A	-15	HIS	-	expression tag	UNP Q9I2N0
A	-14	HIS	-	expression tag	UNP Q9I2N0
A	-13	HIS	-	expression tag	UNP Q9I2N0
A	-12	HIS	-	expression tag	UNP Q9I2N0
A	-11	HIS	-	expression tag	UNP Q9I2N0
A	-10	SER	-	expression tag	UNP Q9I2N0
A	-9	SER	-	expression tag	UNP Q9I2N0
A	-8	GLY	-	expression tag	UNP Q9I2N0
A	-7	LEU	-	expression tag	UNP Q9I2N0
A	-6	VAL	-	expression tag	UNP Q9I2N0
A	-5	PRO	-	expression tag	UNP Q9I2N0
A	-4	ARG	-	expression tag	UNP Q9I2N0
A	-3	GLY	-	expression tag	UNP Q9I2N0
A	-2	SER	-	expression tag	UNP Q9I2N0
A	-1	HIS	-	expression tag	UNP Q9I2N0
A	0	MET	-	expression tag	UNP Q9I2N0
B	-20	MET	-	expression tag	UNP Q9I2N0
B	-19	GLY	-	expression tag	UNP Q9I2N0

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-18	SER	-	expression tag	UNP Q9I2N0
B	-17	SER	-	expression tag	UNP Q9I2N0
B	-16	HIS	-	expression tag	UNP Q9I2N0
B	-15	HIS	-	expression tag	UNP Q9I2N0
B	-14	HIS	-	expression tag	UNP Q9I2N0
B	-13	HIS	-	expression tag	UNP Q9I2N0
B	-12	HIS	-	expression tag	UNP Q9I2N0
B	-11	HIS	-	expression tag	UNP Q9I2N0
B	-10	SER	-	expression tag	UNP Q9I2N0
B	-9	SER	-	expression tag	UNP Q9I2N0
B	-8	GLY	-	expression tag	UNP Q9I2N0
B	-7	LEU	-	expression tag	UNP Q9I2N0
B	-6	VAL	-	expression tag	UNP Q9I2N0
B	-5	PRO	-	expression tag	UNP Q9I2N0
B	-4	ARG	-	expression tag	UNP Q9I2N0
B	-3	GLY	-	expression tag	UNP Q9I2N0
B	-2	SER	-	expression tag	UNP Q9I2N0
B	-1	HIS	-	expression tag	UNP Q9I2N0
B	0	MET	-	expression tag	UNP Q9I2N0
C	-20	MET	-	expression tag	UNP Q9I2N0
C	-19	GLY	-	expression tag	UNP Q9I2N0
C	-18	SER	-	expression tag	UNP Q9I2N0
C	-17	SER	-	expression tag	UNP Q9I2N0
C	-16	HIS	-	expression tag	UNP Q9I2N0
C	-15	HIS	-	expression tag	UNP Q9I2N0
C	-14	HIS	-	expression tag	UNP Q9I2N0
C	-13	HIS	-	expression tag	UNP Q9I2N0
C	-12	HIS	-	expression tag	UNP Q9I2N0
C	-11	HIS	-	expression tag	UNP Q9I2N0
C	-10	SER	-	expression tag	UNP Q9I2N0
C	-9	SER	-	expression tag	UNP Q9I2N0
C	-8	GLY	-	expression tag	UNP Q9I2N0
C	-7	LEU	-	expression tag	UNP Q9I2N0
C	-6	VAL	-	expression tag	UNP Q9I2N0
C	-5	PRO	-	expression tag	UNP Q9I2N0
C	-4	ARG	-	expression tag	UNP Q9I2N0
C	-3	GLY	-	expression tag	UNP Q9I2N0
C	-2	SER	-	expression tag	UNP Q9I2N0
C	-1	HIS	-	expression tag	UNP Q9I2N0
C	0	MET	-	expression tag	UNP Q9I2N0

- Molecule 2 is a DNA chain called DNA (5'-D(P\*GP\*TP\*TP\*GP\*GP\*GP\*AP\*TP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	10	Total	C	N	O	P	0	0	0
			211	100	38	63	10			
2	F	10	Total	C	N	O	P	0	0	0
			211	100	38	63	10			
2	H	10	Total	C	N	O	P	0	0	0
			211	100	38	63	10			

- Molecule 3 is a DNA chain called DNA (5'-D(P\*GP\*AP\*AP\*TP\*GP\*TP\*GP\*TP\*GP\*TP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*CP\*CP\*AP\*AP\*CP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	24	Total	C	N	O	P	0	0	0
			488	234	84	146	24			
3	G	24	Total	C	N	O	P	0	0	0
			487	233	84	146	24			
3	I	24	Total	C	N	O	P	0	0	0
			488	234	84	146	24			

- Molecule 4 is a DNA chain called DNA (5'-D(P\*TP\*GP\*AP\*CP\*AP\*CP\*AP\*CP\*AP\*TP\*TP\*CP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	J	14	Total	C	N	O	P	0	0	0
			284	136	53	81	14			
4	L	14	Total	C	N	O	P	0	0	0
			284	136	53	81	14			
4	N	13	Total	C	N	O	P	0	0	0
			264	126	51	74	13			

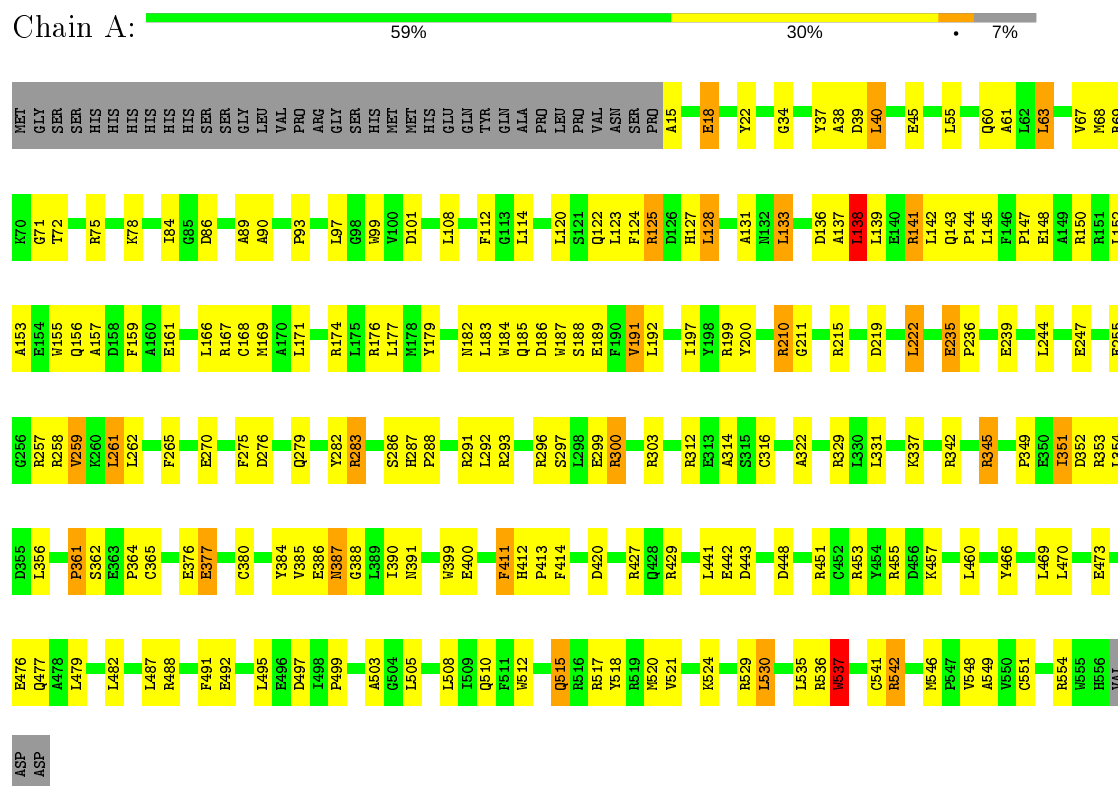
- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total	Ca	0	0
			2	2		
5	A	2	Total	Ca	0	0
			2	2		
5	C	1	Total	Ca	0	0
			1	1		

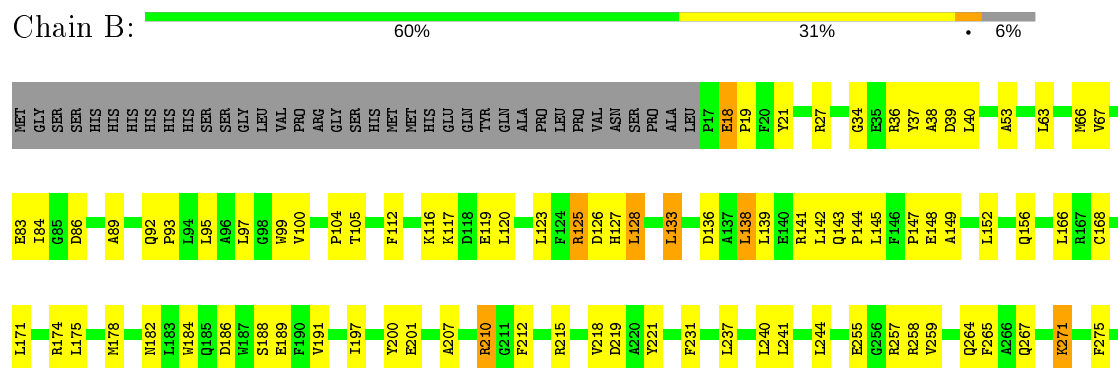
### 3 Residue-property plots

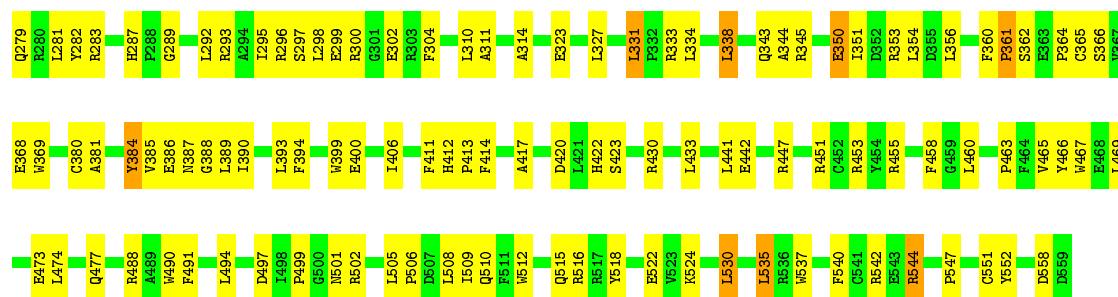
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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: Fanconi-associated nuclease 1 homolog



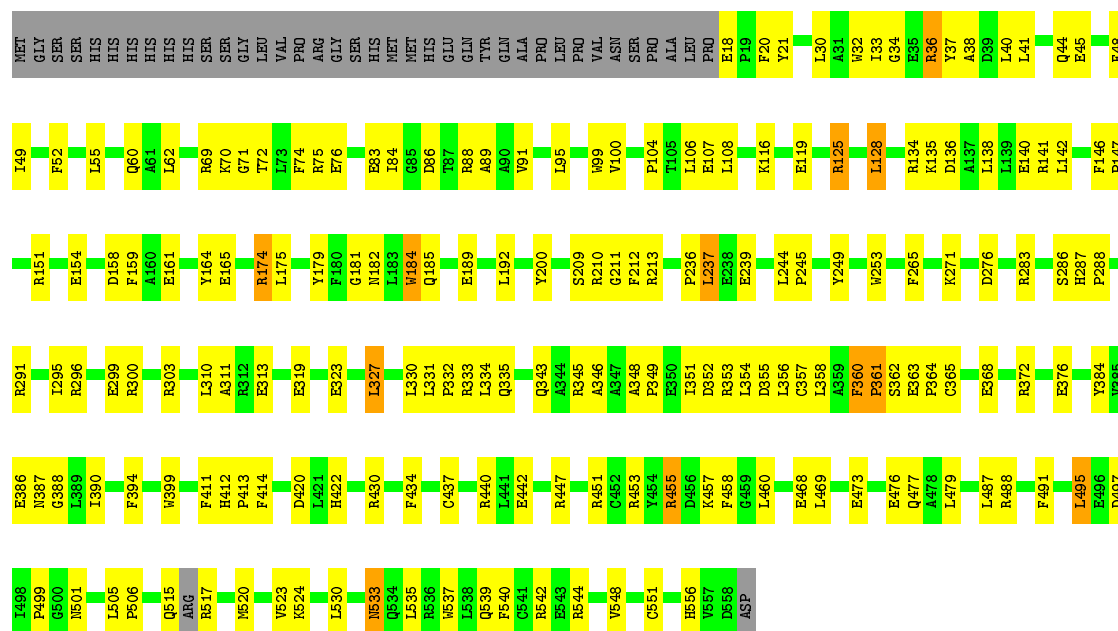
- Molecule 1: Fanconi-associated nuclease 1 homolog





- Molecule 1: Fanconi-associated nuclease 1 homolog

Chain C: 61% 30% 7%



- Molecule 2: DNA (5'-D(P\*GP\*TP\*TP\*GP\*GP\*GP\*AP\*TP\*TP\*G)-3')

Chain D: 20% 50% 30%



- Molecule 2: DNA (5'-D(P\*GP\*TP\*TP\*GP\*GP\*GP\*AP\*TP\*TP\*G)-3')

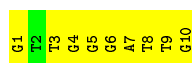
Chain F: 40% 50% 10%



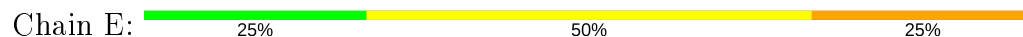
- Molecule 2: DNA (5'-D(P\*GP\*TP\*TP\*GP\*GP\*GP\*AP\*TP\*TP\*G)-3')

Chain H: 10% 90%

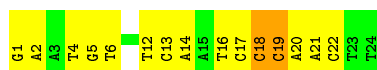




- Molecule 3: DNA (5'-D(P\*GP\*AP\*AP\*TP\*GP\*TP\*GP\*TP\*GP\*TP\*CP\*TP\*CP\*AP\*AP\*T  
P\*CP\*CP\*CP\*AP\*AP\*CP\*TP\*T)-3')



- Molecule 3: DNA (5'-D(P\*GP\*AP\*AP\*TP\*GP\*TP\*GP\*TP\*GP\*TP\*CP\*TP\*CP\*AP\*AP\*T  
P\*CP\*CP\*CP\*AP\*AP\*CP\*TP\*T)-3')



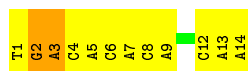
- Molecule 3: DNA (5'-D(P\*GP\*AP\*AP\*TP\*GP\*TP\*GP\*TP\*GP\*TP\*CP\*TP\*CP\*AP\*AP\*T  
P\*CP\*CP\*CP\*AP\*AP\*CP\*TP\*T)-3')



- Molecule 4: DNA (5'-D(P\*TP\*GP\*AP\*CP\*AP\*CP\*AP\*CP\*AP\*TP\*TP\*CP\*AP\*A)-3'  
)



- Molecule 4: DNA (5'-D(P\*TP\*GP\*AP\*CP\*AP\*CP\*AP\*CP\*AP\*TP\*TP\*CP\*AP\*A)-3'  
)



- Molecule 4: DNA (5'-D(P\*TP\*GP\*AP\*CP\*AP\*CP\*AP\*CP\*AP\*TP\*TP\*CP\*AP\*A)-3'  
)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	140.29Å 143.26Å 172.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.58 – 3.40 46.03 – 3.17	Depositor EDS
% Data completeness (in resolution range)	95.6 (33.58-3.40) 94.4 (46.03-3.17)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	0.16	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.42 (at 3.19Å)	Xtriage
Refinement program	PHENIX (1.10.1-2155_1069: ???)	Depositor
R, $R_{free}$	0.221 , 0.280 0.221 , 0.280	Depositor DCC
$R_{free}$ test set	2841 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	105.5	Xtriage
Anisotropy	0.109	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 64.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.004 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	16164	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	163.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.87% 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: CA

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	0.62	1/4522 (0.0%)	0.87	7/6125 (0.1%)
1	B	0.55	1/4541 (0.0%)	0.79	4/6150 (0.1%)
1	C	0.47	0/4516	0.72	1/6115 (0.0%)
2	D	1.26	2/236 (0.8%)	1.25	2/364 (0.5%)
2	F	1.09	1/236 (0.4%)	1.10	0/364
2	H	0.76	0/236	1.11	0/364
3	E	1.43	6/545 (1.1%)	1.18	5/838 (0.6%)
3	G	1.17	3/544 (0.6%)	1.15	1/836 (0.1%)
3	I	1.00	0/545	1.12	0/838
4	J	1.42	3/318 (0.9%)	1.17	2/487 (0.4%)
4	L	1.08	1/318 (0.3%)	1.05	1/487 (0.2%)
4	N	0.82	0/296	0.94	0/453
All	All	0.71	18/16853 (0.1%)	0.88	23/23421 (0.1%)

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
1	B	0	1
All	All	0	3

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	12	DT	C1'-N1	8.56	1.60	1.49
3	E	4	DT	C3'-O3'	-8.22	1.33	1.44
2	F	7	DA	C3'-O3'	-7.67	1.33	1.44
2	D	7	DA	C3'-O3'	-7.26	1.34	1.44

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	J	4	DC	C3'-O3'	-6.70	1.35	1.44

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	7	DA	O5'-P-OP2	-8.03	98.47	105.70
4	J	3	DA	O5'-P-OP2	-7.43	99.01	105.70
1	A	377	GLU	C-N-CD	-7.32	104.49	120.60
4	L	2	DG	O4'-C4'-C3'	-6.84	101.76	104.50
3	E	4	DT	O4'-C4'-C3'	-6.81	101.78	104.50

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	361	PRO	Peptide
1	A	71	GLY	Peptide
1	B	361	PRO	Peptide

## 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	A	4406	0	4306	149	1
1	B	4424	0	4317	139	0
1	C	4401	0	4293	129	1
2	D	211	0	115	16	0
2	F	211	0	115	5	0
2	H	211	0	115	14	0
3	E	488	0	273	16	0
3	G	487	0	270	15	0
3	I	488	0	273	20	0
4	J	284	0	158	11	0
4	L	284	0	158	16	0
4	N	264	0	146	11	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	1	0	0	0	0
All	All	16164	0	14539	524	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 17.

The worst 5 of 524 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:342:ARG:NH1	4:J:14:DA:OP1	1.91	1.03
1:B:387:ASN:HB3	1:B:505:LEU:HB2	1.48	0.96
3:E:21:DA:H2"	3:E:22:DC:H5"	1.49	0.94
3:G:21:DA:H2"	3:G:22:DC:H5"	1.49	0.94
4:N:9:DA:H2"	4:N:10:DT:H5"	1.50	0.93

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:376:GLU:O	1:C:447:ARG:NH2[3_545]	2.18	0.02

## 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	540/580 (93%)	520 (96%)	18 (3%)	2 (0%)	34	67
1	B	541/580 (93%)	523 (97%)	16 (3%)	2 (0%)	34	67
1	C	536/580 (92%)	511 (95%)	24 (4%)	1 (0%)	47	78
All	All	1617/1740 (93%)	1554 (96%)	58 (4%)	5 (0%)	41	72

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	361	PRO
1	A	18	GLU
1	B	344	ALA
1	B	19	PRO
1	A	235	GLU

### 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	441/478 (92%)	406 (92%)	35 (8%)	12	39
1	B	444/478 (93%)	422 (95%)	22 (5%)	24	54
1	C	442/478 (92%)	416 (94%)	26 (6%)	19	49
All	All	1327/1434 (92%)	1244 (94%)	83 (6%)	18	47

5 of 83 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	133	LEU
1	B	350	GLU
1	C	453	ARG
1	B	138	LEU
1	B	271	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	182	ASN
1	A	279	GLN
1	C	343	GLN
1	C	531	GLN
1	C	533	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](#)

Of 5 ligands modelled in this entry, 5 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 ⓘ

### 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.