



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 07:45 am BST

PDB ID : 5D9A  
Title : Influenza C Virus RNA-dependent RNA Polymerase - Space group P212121  
Authors : Hengrung, N.; El Omari, K.; Serna Martin, I.; Vreede, F.T.; Cusack, S.;  
Rambo, R.P.; Vonrhein, C.; Bricogne, G.; Stuart, D.I.; Grimes, J.M.; Fodor,  
E.  
Deposited on : 2015-08-18  
Resolution : 4.30 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

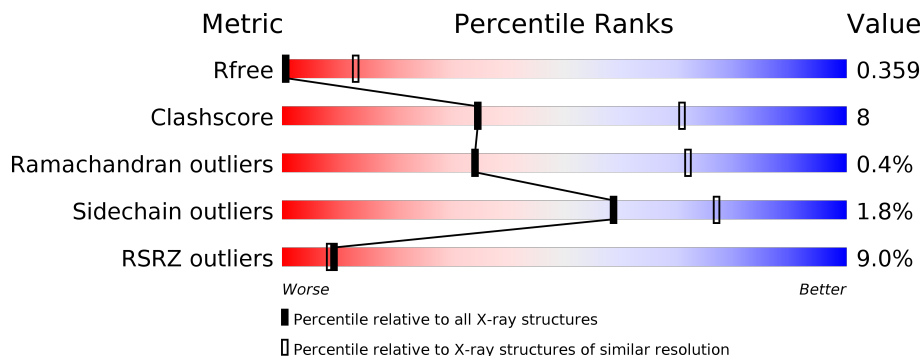
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.30 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	1014 (4.80-3.80)
Clashscore	141614	1077 (4.80-3.80)
Ramachandran outliers	138981	1029 (4.80-3.80)
Sidechain outliers	138945	1012 (4.80-3.80)
RSRZ outliers	127900	1075 (4.90-3.70)

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	A	709	
1	D	709	
1	G	709	
1	J	709	
2	B	754	
2	E	754	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	H	754	<p>6% 80% 14% 6%</p>
2	K	754	<p>8% 83% 11% 6%</p>
3	C	782	<p>21% 86% 10% •</p>
3	F	782	<p>6% 82% 15% •</p>
3	I	782	<p>11% 84% 13% ••</p>
3	L	782	<p>17% 87% 10% •</p>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 69371 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	693	Total 5630	C 3589	N 954	O 1043	S 44	0	0	0
1	D	693	Total 5630	C 3589	N 954	O 1043	S 44	0	0	0
1	G	693	Total 5630	C 3589	N 954	O 1043	S 44	0	0	0
1	J	693	Total 5630	C 3589	N 954	O 1043	S 44	0	0	0

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	711	Total 5652	C 3587	N 956	O 1056	S 53	0	0	0
2	E	711	Total 5652	C 3587	N 956	O 1056	S 53	0	0	0
2	H	711	Total 5652	C 3587	N 956	O 1056	S 53	0	0	0
2	K	711	Total 5652	C 3587	N 956	O 1056	S 53	0	0	0

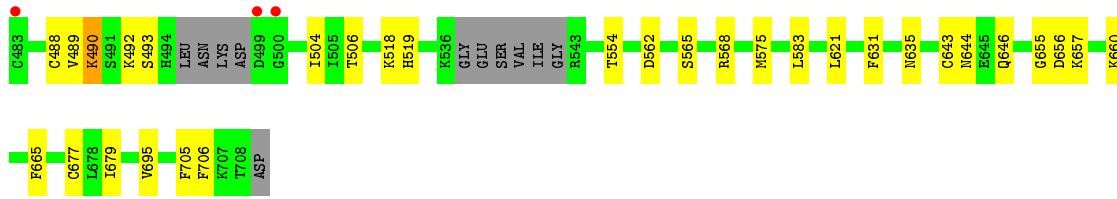
- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	754	Total 6015	C 3806	N 1056	O 1117	S 36	0	0	0
3	F	762	Total 6076	C 3845	N 1066	O 1128	S 37	0	0	0
3	I	762	Total 6076	C 3845	N 1066	O 1128	S 37	0	0	0
3	L	762	Total 6076	C 3845	N 1066	O 1128	S 37	0	0	0

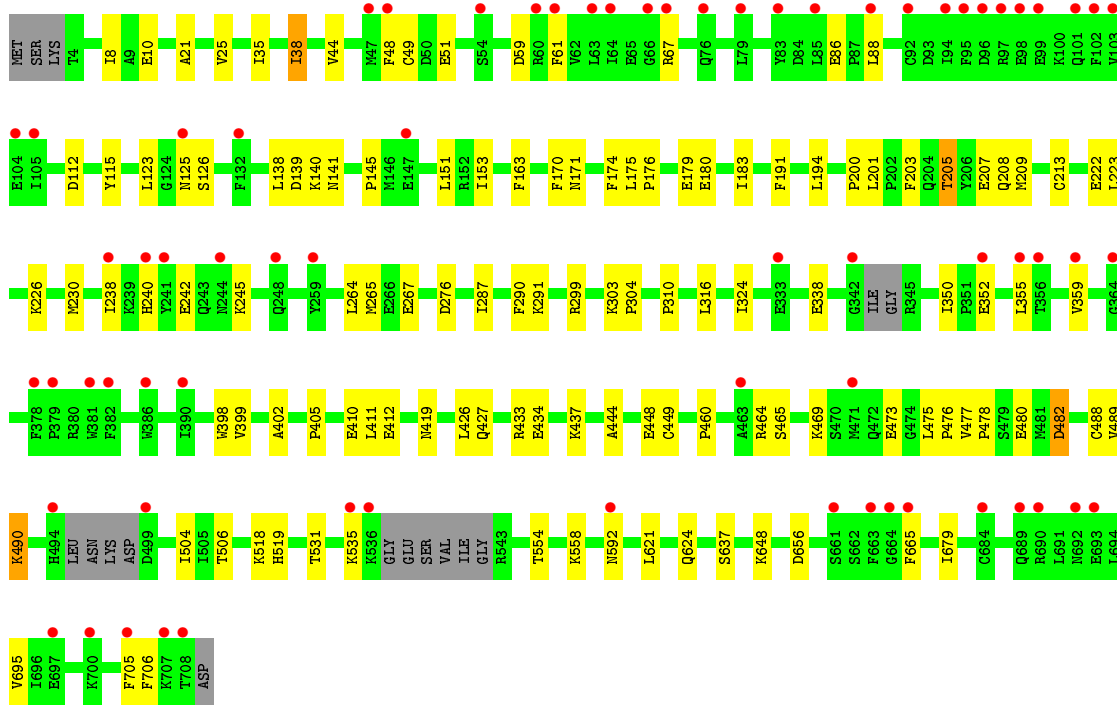
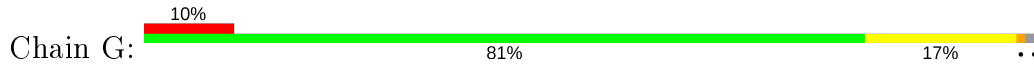
There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	775	ALA	-	expression tag	UNP Q9IMP3
C	776	ARG	-	expression tag	UNP Q9IMP3
C	777	GLU	-	expression tag	UNP Q9IMP3
C	778	ASN	-	expression tag	UNP Q9IMP3
C	779	LEU	-	expression tag	UNP Q9IMP3
C	780	TYR	-	expression tag	UNP Q9IMP3
C	781	PHE	-	expression tag	UNP Q9IMP3
C	782	GLN	-	expression tag	UNP Q9IMP3
F	775	ALA	-	expression tag	UNP Q9IMP3
F	776	ARG	-	expression tag	UNP Q9IMP3
F	777	GLU	-	expression tag	UNP Q9IMP3
F	778	ASN	-	expression tag	UNP Q9IMP3
F	779	LEU	-	expression tag	UNP Q9IMP3
F	780	TYR	-	expression tag	UNP Q9IMP3
F	781	PHE	-	expression tag	UNP Q9IMP3
F	782	GLN	-	expression tag	UNP Q9IMP3
I	775	ALA	-	expression tag	UNP Q9IMP3
I	776	ARG	-	expression tag	UNP Q9IMP3
I	777	GLU	-	expression tag	UNP Q9IMP3
I	778	ASN	-	expression tag	UNP Q9IMP3
I	779	LEU	-	expression tag	UNP Q9IMP3
I	780	TYR	-	expression tag	UNP Q9IMP3
I	781	PHE	-	expression tag	UNP Q9IMP3
I	782	GLN	-	expression tag	UNP Q9IMP3
L	775	ALA	-	expression tag	UNP Q9IMP3
L	776	ARG	-	expression tag	UNP Q9IMP3
L	777	GLU	-	expression tag	UNP Q9IMP3
L	778	ASN	-	expression tag	UNP Q9IMP3
L	779	LEU	-	expression tag	UNP Q9IMP3
L	780	TYR	-	expression tag	UNP Q9IMP3
L	781	PHE	-	expression tag	UNP Q9IMP3
L	782	GLN	-	expression tag	UNP Q9IMP3

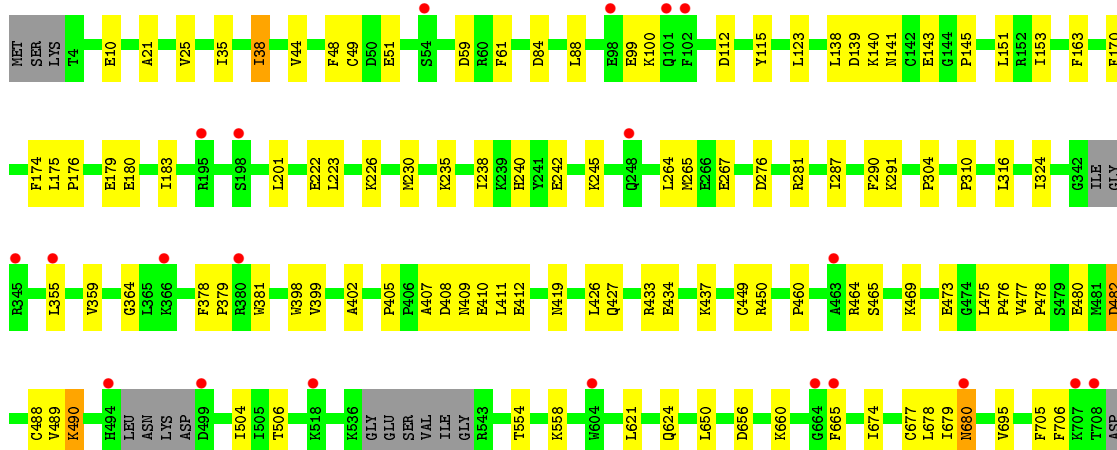
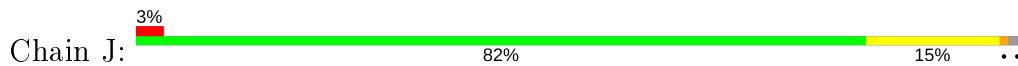




• Molecule 1: Polymerase acidic protein



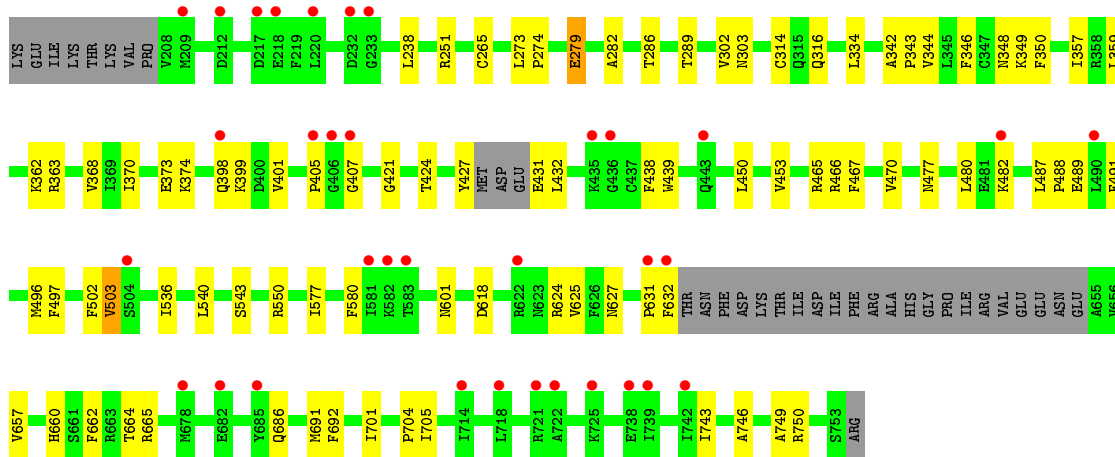
• Molecule 1: Polymerase acidic protein



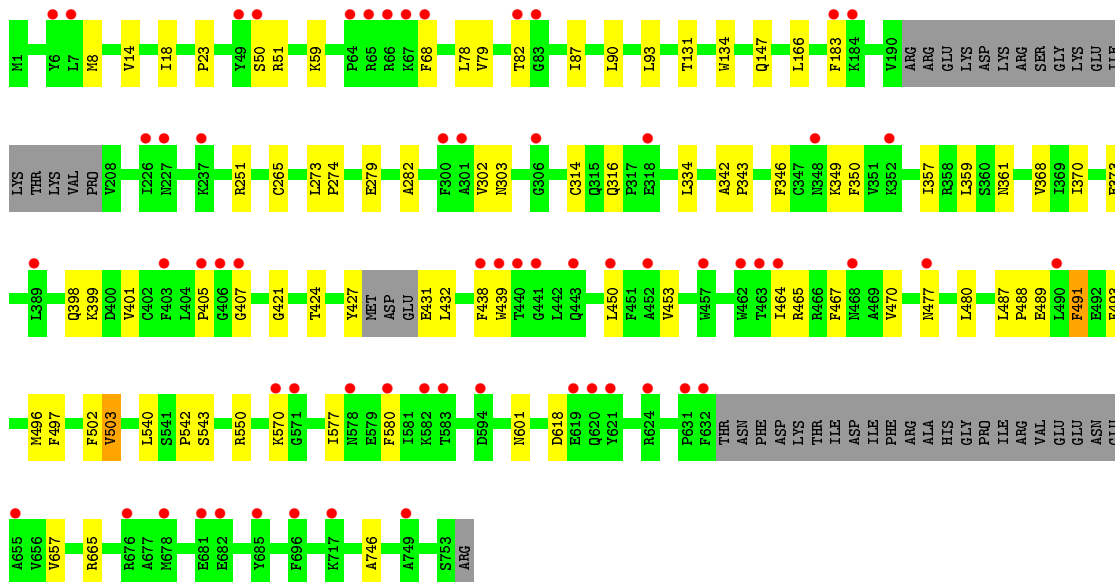
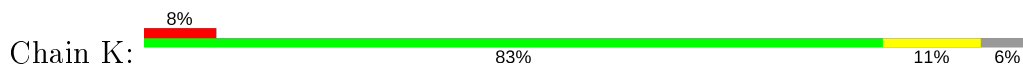
• Molecule 2: RNA-directed RNA polymerase catalytic subunit



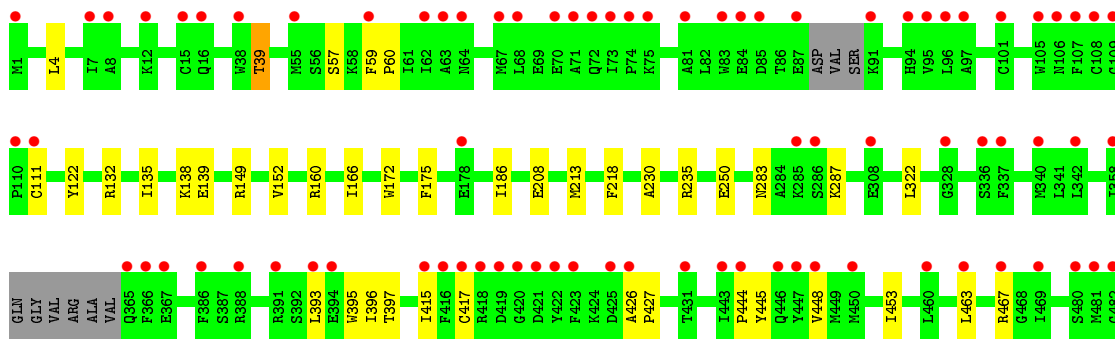
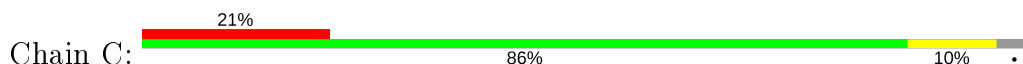


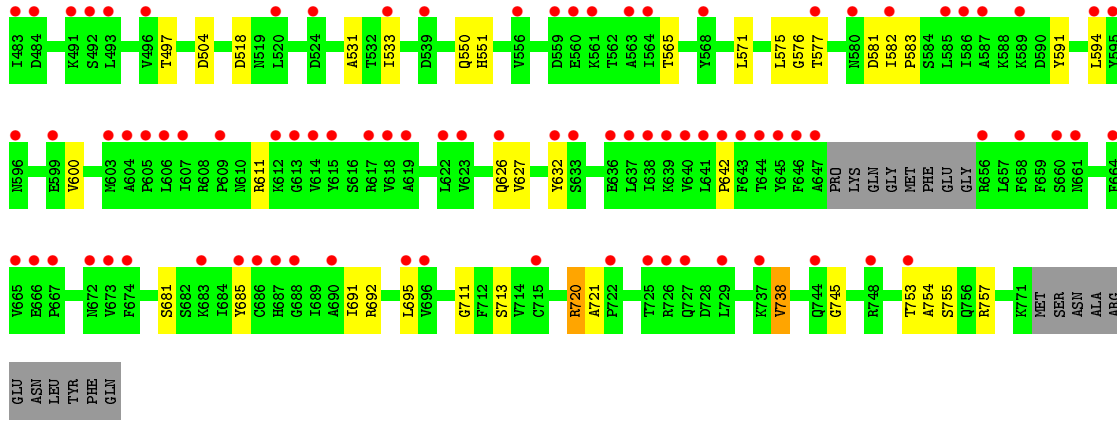


• Molecule 2: RNA-directed RNA polymerase catalytic subunit

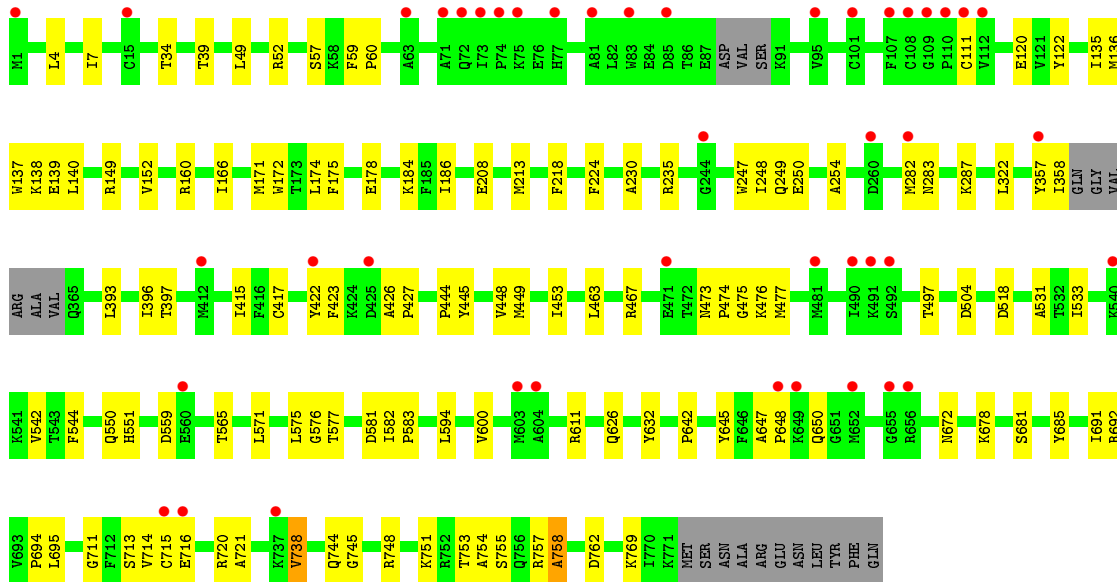
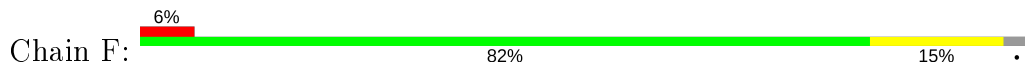


• Molecule 3: Polymerase basic protein 2

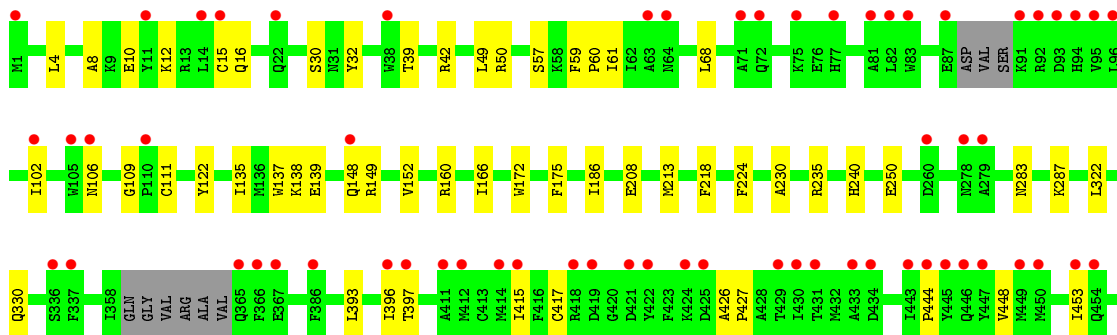
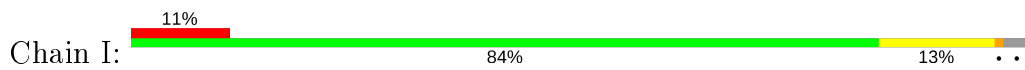


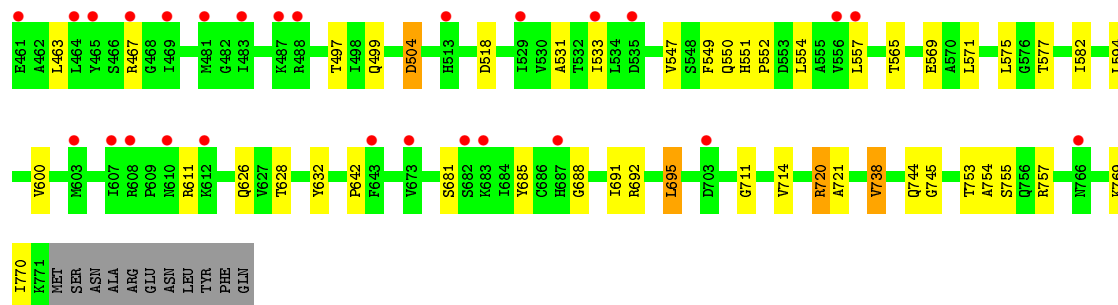


• Molecule 3: Polymerase basic protein 2

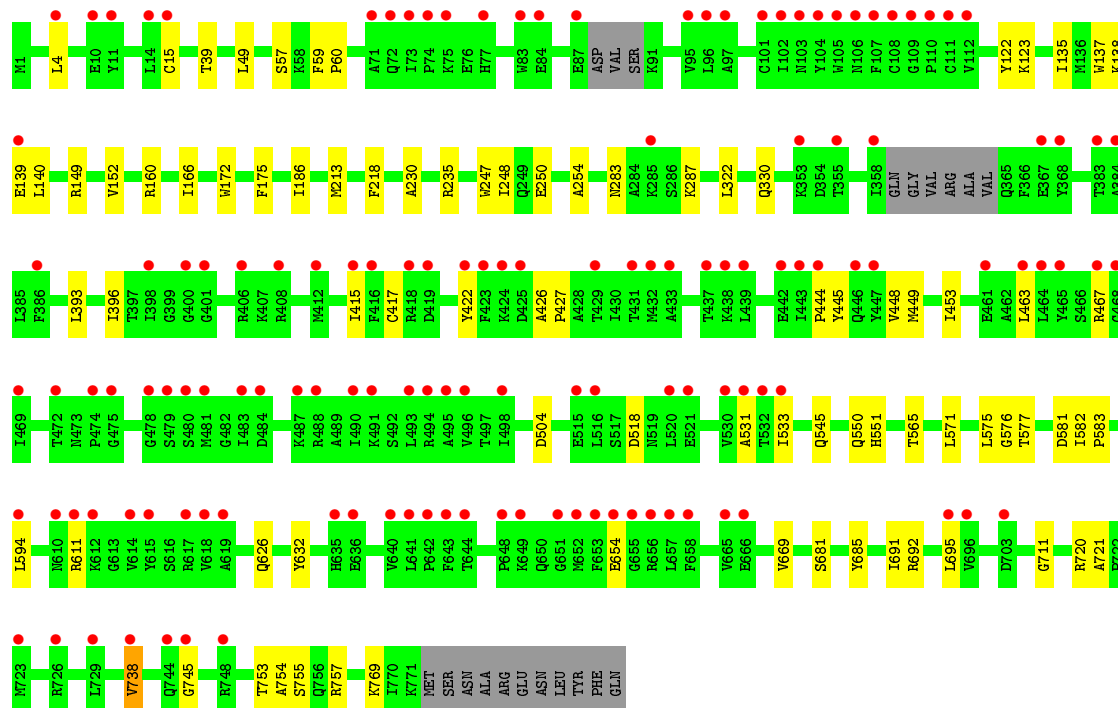
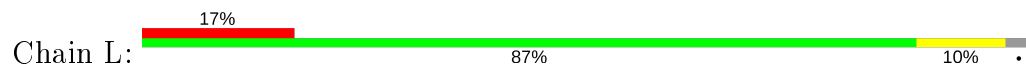


• Molecule 3: Polymerase basic protein 2





- Molecule 3: Polymerase basic protein 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.28Å 217.50Å 597.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.01 – 4.30 80.90 – 4.30	Depositor EDS
% Data completeness (in resolution range)	98.8 (50.01-4.30) 98.9 (80.90-4.30)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.54 (at 4.30Å)	Xtrriage
Refinement program	REFMAC 5.8.0124	Depositor
R, $R_{free}$	0.316 , 0.368 0.312 , 0.359	Depositor DCC
$R_{free}$ test set	4744 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	146.0	Xtrriage
Anisotropy	0.501	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 108.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.23$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	69371	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	193.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.51% 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](#)

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	0.46	0/5746	0.65	0/7717
1	D	0.47	0/5746	0.66	0/7717
1	G	0.46	0/5746	0.64	0/7717
1	J	0.46	0/5746	0.64	0/7717
2	B	0.45	0/5749	0.65	1/7723 (0.0%)
2	E	0.44	0/5749	0.65	0/7723
2	H	0.44	0/5749	0.64	0/7723
2	K	0.43	0/5749	0.64	0/7723
3	C	0.45	0/6121	0.69	1/8236 (0.0%)
3	F	0.46	0/6185	0.69	0/8322
3	I	0.45	0/6185	0.68	1/8322 (0.0%)
3	L	0.44	0/6185	0.68	0/8322
All	All	0.45	0/70656	0.66	3/94962 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	I	720	ARG	NE-CZ-NH1	6.60	123.60	120.30
3	C	720	ARG	NE-CZ-NH1	6.55	123.58	120.30
2	B	180	GLU	OE1-CD-OE2	-6.36	115.67	123.30

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	A	5630	0	5632	144	1
1	D	5630	0	5632	192	5
1	G	5630	0	5632	167	2
1	J	5630	0	5632	116	6
2	B	5652	0	5749	122	6
2	E	5652	0	5749	104	4
2	H	5652	0	5749	129	0
2	K	5652	0	5749	77	0
3	C	6015	0	6124	54	0
3	F	6076	0	6183	171	8
3	I	6076	0	6183	119	4
3	L	6076	0	6183	80	0
All	All	69371	0	70197	1164	18

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 8.

The worst 5 of 1164 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:69:ARG:CZ	1:D:91:LEU:HD11	1.14	1.55
2:B:180:GLU:HG2	1:G:203:PHE:CE1	1.47	1.50
3:I:577:THR:HG23	3:I:754:ALA:CB	1.43	1.49
2:H:363:ARG:HA	1:J:409:ASN:ND2	1.20	1.42
1:D:69:ARG:NE	1:D:91:LEU:HD11	1.33	1.41

The worst 5 of 18 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)
2:E:215:LYS:CE	3:I:628:THR:OG1[3_554]	0.85	1.35
2:E:215:LYS:NZ	3:I:628:THR:OG1[3_554]	0.92	1.28
2:E:215:LYS:CE	3:I:628:THR:CB[3_554]	1.49	0.71
2:B:457:TRP:N	1:J:84:ASP:OD2[1_565]	1.72	0.48
2:B:362:LYS:CE	1:D:81:LYS:NZ[1_655]	1.75	0.45

## 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	685/709 (97%)	631 (92%)	51 (7%)	3 (0%)	34	72
1	D	685/709 (97%)	634 (93%)	49 (7%)	2 (0%)	41	76
1	G	685/709 (97%)	635 (93%)	48 (7%)	2 (0%)	41	76
1	J	685/709 (97%)	633 (92%)	50 (7%)	2 (0%)	41	76
2	B	703/754 (93%)	670 (95%)	30 (4%)	3 (0%)	34	72
2	E	703/754 (93%)	669 (95%)	31 (4%)	3 (0%)	34	72
2	H	703/754 (93%)	670 (95%)	30 (4%)	3 (0%)	34	72
2	K	703/754 (93%)	669 (95%)	31 (4%)	3 (0%)	34	72
3	C	746/782 (95%)	679 (91%)	64 (9%)	3 (0%)	34	72
3	F	756/782 (97%)	683 (90%)	69 (9%)	4 (0%)	29	68
3	I	756/782 (97%)	685 (91%)	68 (9%)	3 (0%)	34	72
3	L	756/782 (97%)	686 (91%)	67 (9%)	3 (0%)	34	72
All	All	8566/8980 (95%)	7944 (93%)	588 (7%)	34 (0%)	34	72

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	503	VAL
2	E	503	VAL
3	F	758	ALA
2	H	503	VAL
2	K	503	VAL

### 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	618/631 (98%)	610 (99%)	8 (1%)	69	82
1	D	618/631 (98%)	610 (99%)	8 (1%)	69	82
1	G	618/631 (98%)	608 (98%)	10 (2%)	62	79
1	J	618/631 (98%)	608 (98%)	10 (2%)	62	79
2	B	629/669 (94%)	617 (98%)	12 (2%)	57	75
2	E	629/669 (94%)	617 (98%)	12 (2%)	57	75
2	H	629/669 (94%)	617 (98%)	12 (2%)	57	75
2	K	629/669 (94%)	617 (98%)	12 (2%)	57	75
3	C	663/686 (97%)	649 (98%)	14 (2%)	53	72
3	F	669/686 (98%)	656 (98%)	13 (2%)	57	75
3	I	669/686 (98%)	656 (98%)	13 (2%)	57	75
3	L	669/686 (98%)	656 (98%)	13 (2%)	57	75
All	All	7658/7944 (96%)	7521 (98%)	137 (2%)	59	77

5 of 137 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	518	ASP
1	G	554	THR
3	L	135	ILE
3	F	611	ARG
1	G	163	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
3	F	550	GLN
1	G	519	HIS
2	K	477	ASN
1	G	31	HIS
1	G	208	GLN

### 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 [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	693/709 (97%)	0.36	47 (6%) 17 14	66, 164, 326, 490	0
1	D	693/709 (97%)	0.13	20 (2%) 51 41	72, 156, 252, 382	0
1	G	693/709 (97%)	0.43	69 (9%) 7 7	90, 180, 290, 407	0
1	J	693/709 (97%)	0.18	21 (3%) 50 39	105, 181, 273, 414	0
2	B	711/754 (94%)	0.20	41 (5%) 23 19	70, 147, 290, 437	0
2	E	711/754 (94%)	0.21	34 (4%) 30 26	71, 143, 270, 490	0
2	H	711/754 (94%)	0.33	44 (6%) 20 17	82, 170, 287, 498	0
2	K	711/754 (94%)	0.49	63 (8%) 9 8	83, 173, 284, 424	0
3	C	754/782 (96%)	0.99	168 (22%) 0 1	99, 240, 349, 416	0
3	F	762/782 (97%)	0.30	44 (5%) 23 19	87, 174, 295, 481	0
3	I	762/782 (97%)	0.66	89 (11%) 4 5	95, 224, 337, 500	0
3	L	762/782 (97%)	0.84	136 (17%) 1 2	108, 226, 332, 476	0
All	All	8656/8980 (96%)	0.43	776 (8%) 9 8	66, 181, 315, 500	0

The worst 5 of 776 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	I	95	VAL	15.1
3	L	656	ARG	10.7
3	C	665	VAL	10.4
3	L	655	GLY	9.5
3	I	96	LEU	9.4

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates

There are no carbohydrates in this entry.

### 6.4 Ligands

There are no ligands in this entry.

### 6.5 Other polymers

There are no such residues in this entry.