



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 11:06 pm BST

PDB ID : 6QPG  
Title : Influenza A virus Polymerase Heterotrimer A/nt/60/1968(H3N2) in complex with Nanobody NB8205  
Authors : Fan, H.T.; Keown, J.R.; Fodor, E.; Grimes, J.M.  
Deposited on : 2019-02-13  
Resolution : 3.34 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

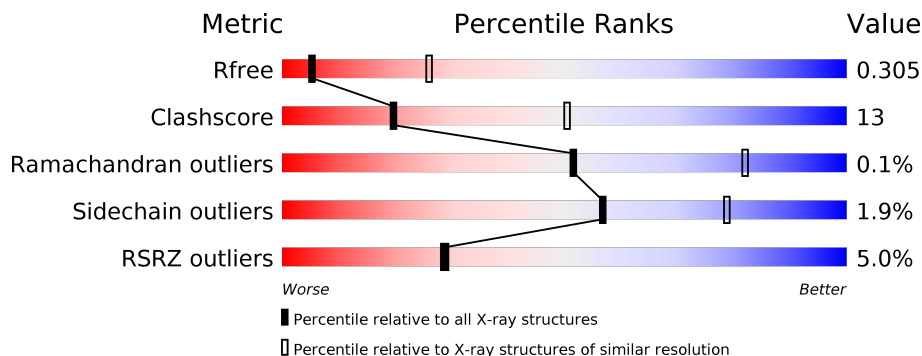
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.34 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	1060 (3.38-3.30)
Clashscore	141614	1111 (3.38-3.30)
Ramachandran outliers	138981	1090 (3.38-3.30)
Sidechain outliers	138945	1089 (3.38-3.30)
RSRZ outliers	127900	1028 (3.38-3.30)

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	716	
1	D	716	
1	G	716	
1	J	716	
2	B	757	
2	E	757	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	H	757	<p>6% 60% 28% • 12%</p>
2	K	757	<p>4% 61% 27% • 12%</p>
3	C	765	<p>6% 67% 26% • 6%</p>
3	F	765	<p>10% 65% 29% • 5%</p>
3	I	765	<p>5% 65% 29% • 5%</p>
3	L	765	<p>7% 64% 29% • 5%</p>
4	M	134	<p>0% 66% 23% • 10%</p>
4	N	134	<p>3% 70% 20% • 10%</p>
4	O	134	<p>4% 58% 32% • 10%</p>
4	P	134	<p>3% 71% 19% • 10%</p>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 70549 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	691	Total 5621	C 3570	N 952	O 1059	S 40	0	0	0
1	D	695	Total 5652	C 3586	N 957	O 1069	S 40	0	0	0
1	G	694	Total 5644	C 3582	N 955	O 1067	S 40	0	0	0
1	J	692	Total 5631	C 3572	N 953	O 1066	S 40	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	656	Total 5239	C 3305	N 903	O 991	S 40	0	0	0
2	E	666	Total 5326	C 3357	N 919	O 1008	S 42	0	0	0
2	H	667	Total 5327	C 3361	N 916	O 1009	S 41	0	0	0
2	K	664	Total 5303	C 3342	N 916	O 1004	S 41	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	577	LYS	GLU	conflict	UNP P03432
E	577	LYS	GLU	conflict	UNP P03432
H	577	LYS	GLU	conflict	UNP P03432
K	577	LYS	GLU	conflict	UNP P03432

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	717	5720	3596	1037	1047	40	0	0	0
3	F	726	5786	3637	1045	1063	41	0	0	0
3	I	728	5804	3648	1049	1066	41	0	0	0
3	L	723	5764	3623	1043	1057	41	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	760	GLU	-	expression tag	UNP P03429
C	761	ASN	-	expression tag	UNP P03429
C	762	LEU	-	expression tag	UNP P03429
C	763	TYR	-	expression tag	UNP P03429
C	764	PHE	-	expression tag	UNP P03429
C	765	GLN	-	expression tag	UNP P03429
F	760	GLU	-	expression tag	UNP P03429
F	761	ASN	-	expression tag	UNP P03429
F	762	LEU	-	expression tag	UNP P03429
F	763	TYR	-	expression tag	UNP P03429
F	764	PHE	-	expression tag	UNP P03429
F	765	GLN	-	expression tag	UNP P03429
I	760	GLU	-	expression tag	UNP P03429
I	761	ASN	-	expression tag	UNP P03429
I	762	LEU	-	expression tag	UNP P03429
I	763	TYR	-	expression tag	UNP P03429
I	764	PHE	-	expression tag	UNP P03429
I	765	GLN	-	expression tag	UNP P03429
L	760	GLU	-	expression tag	UNP P03429
L	761	ASN	-	expression tag	UNP P03429
L	762	LEU	-	expression tag	UNP P03429
L	763	TYR	-	expression tag	UNP P03429
L	764	PHE	-	expression tag	UNP P03429
L	765	GLN	-	expression tag	UNP P03429

- Molecule 4 is a protein called Nanobody NB8205.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	M	121	933	581	164	182	6	0	0	0

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>					<b>ZeroOcc</b>	<b>AltConf</b>	<b>Trace</b>
4	N	121	Total 933	C 581	N 164	O 182	S 6	0	0	0
4	O	121	Total 933	C 581	N 164	O 182	S 6	0	0	0
4	P	121	Total 933	C 581	N 164	O 182	S 6	0	0	0

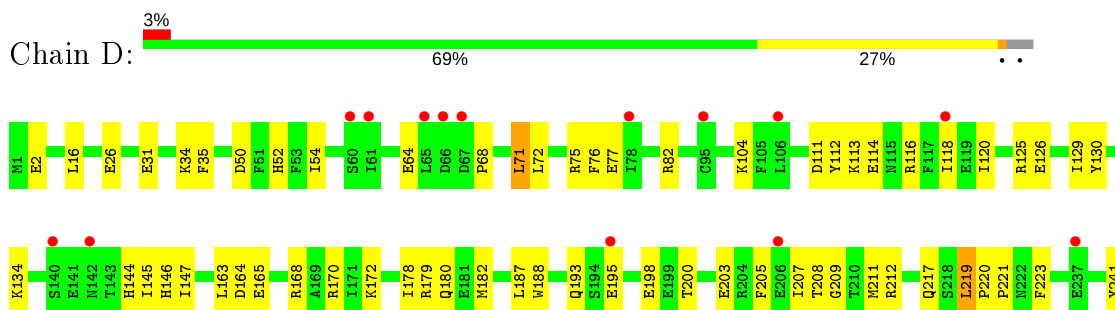
### 3 Residue-property plots [i](#)

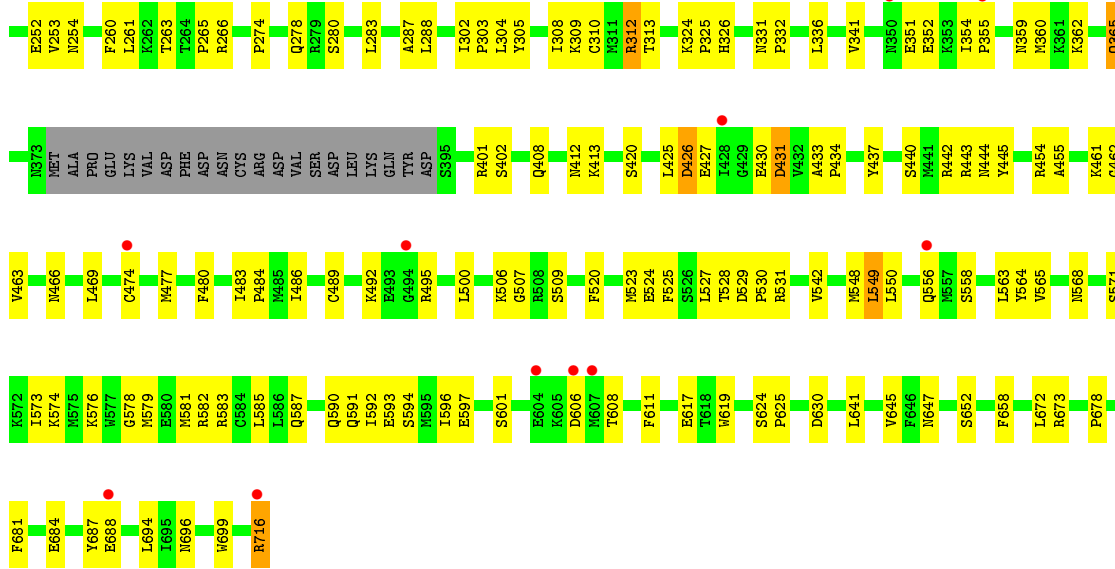
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Polymerase acidic protein

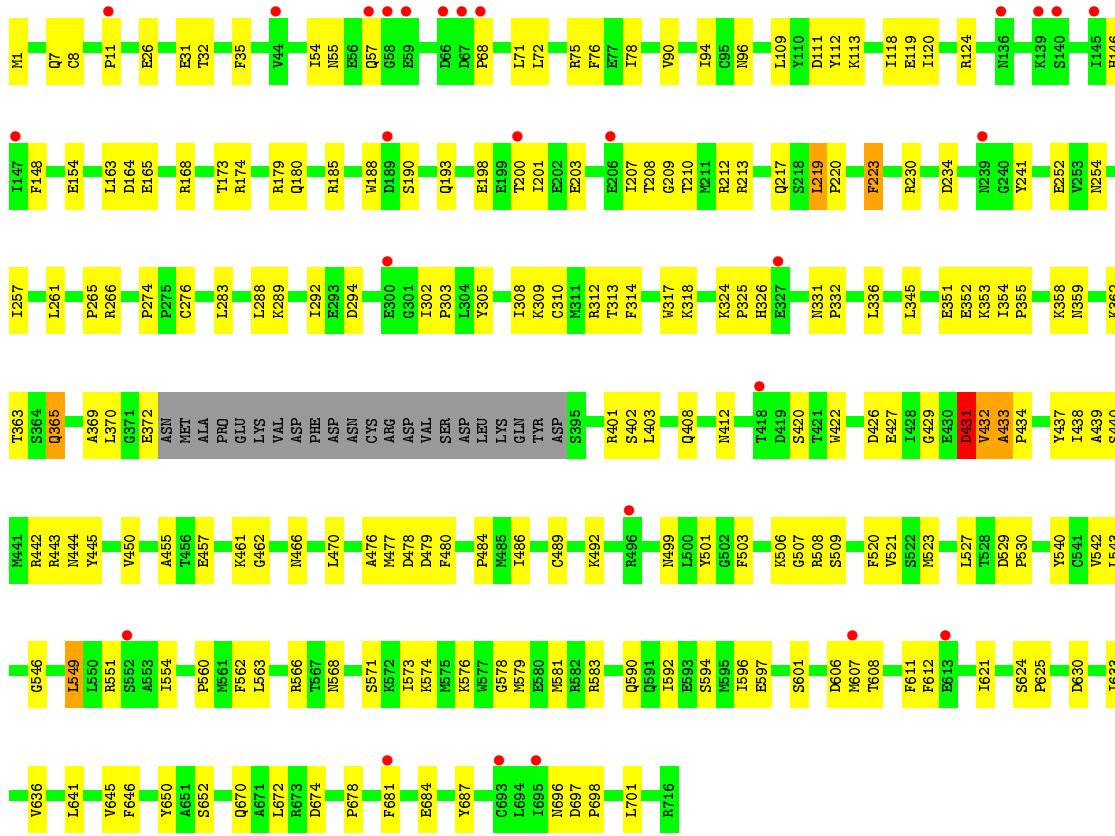


- Molecule 1: Polymerase acidic protein





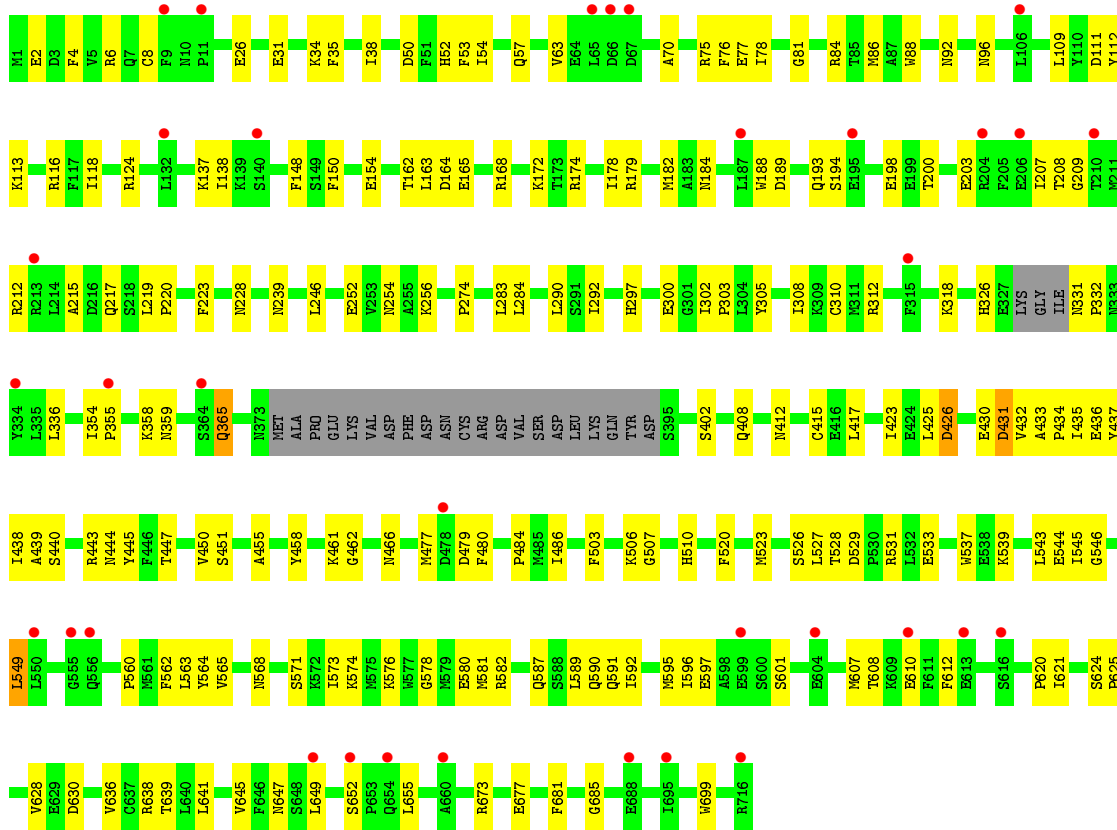
● Molecule 1: Polymerase acidic protein



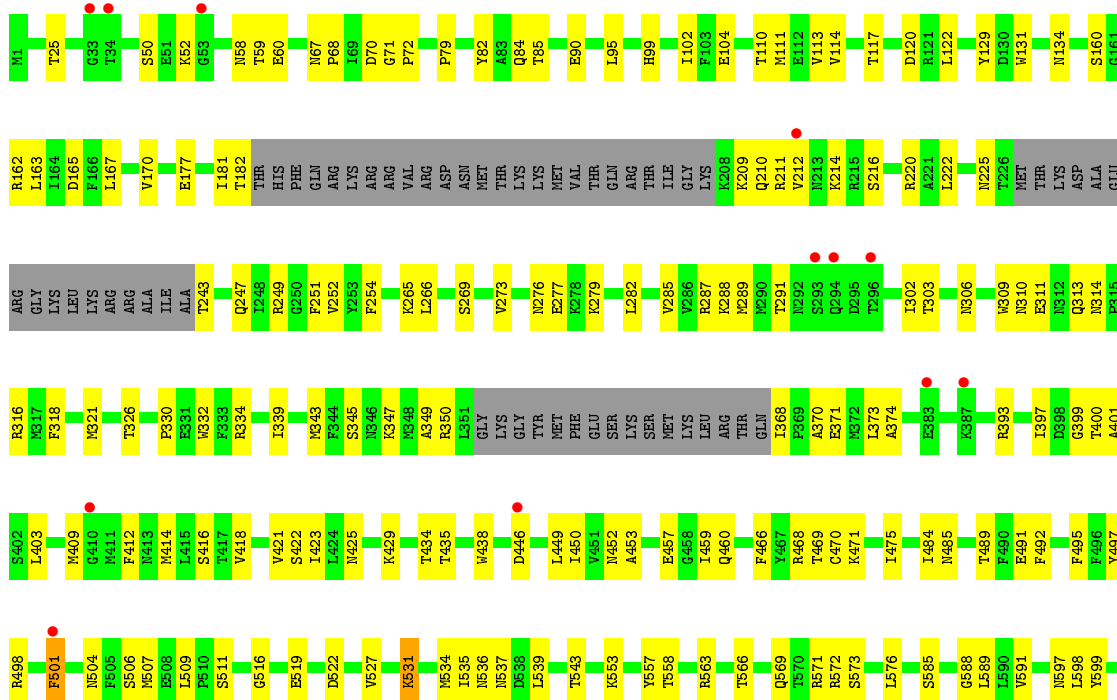
● Molecule 1: Polymerase acidic protein

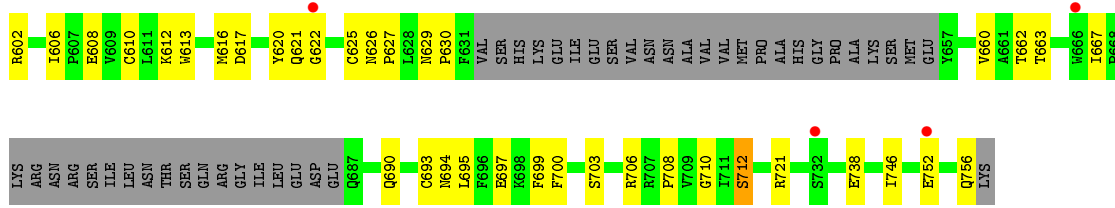




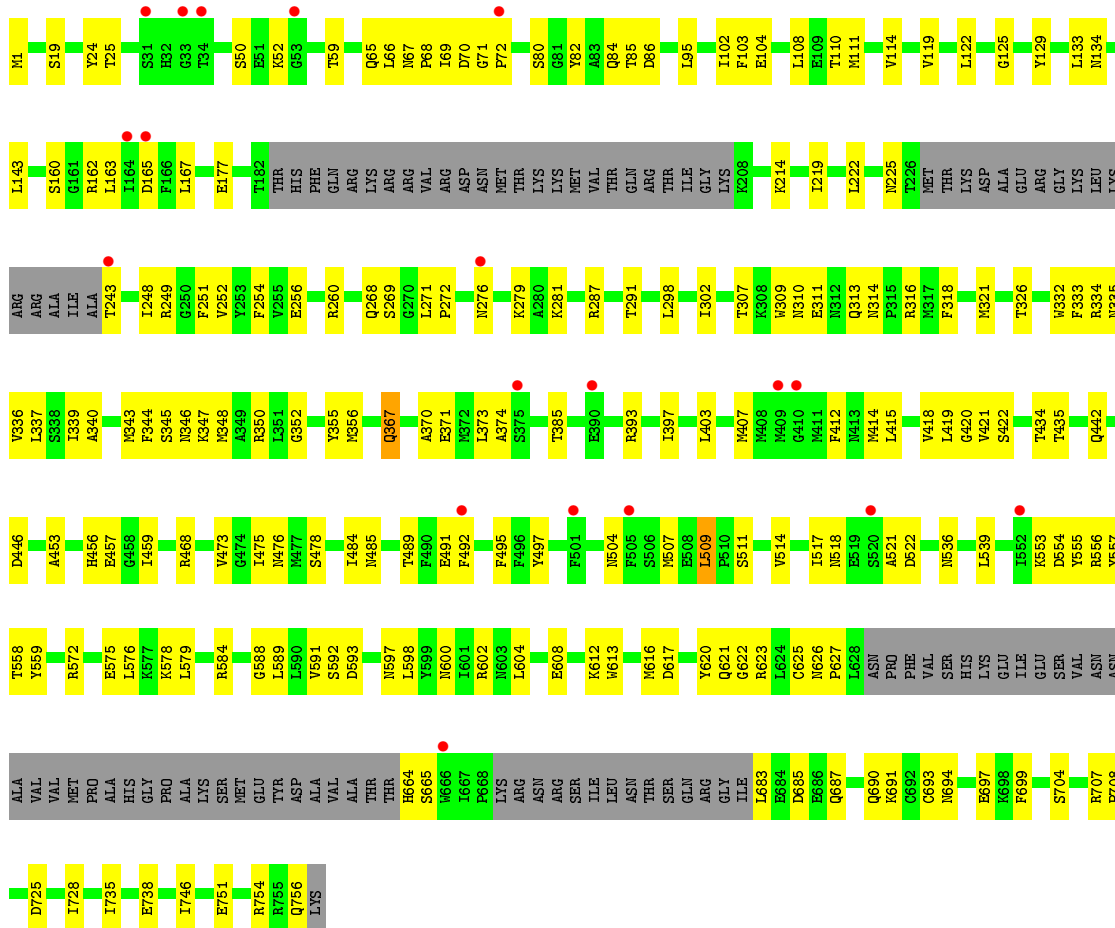


• Molecule 2: RNA-directed RNA polymerase catalytic subunit

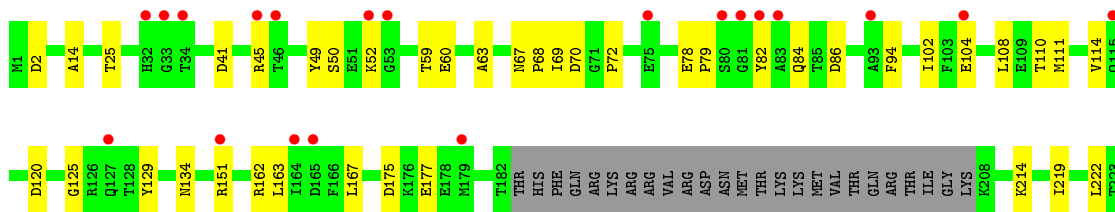


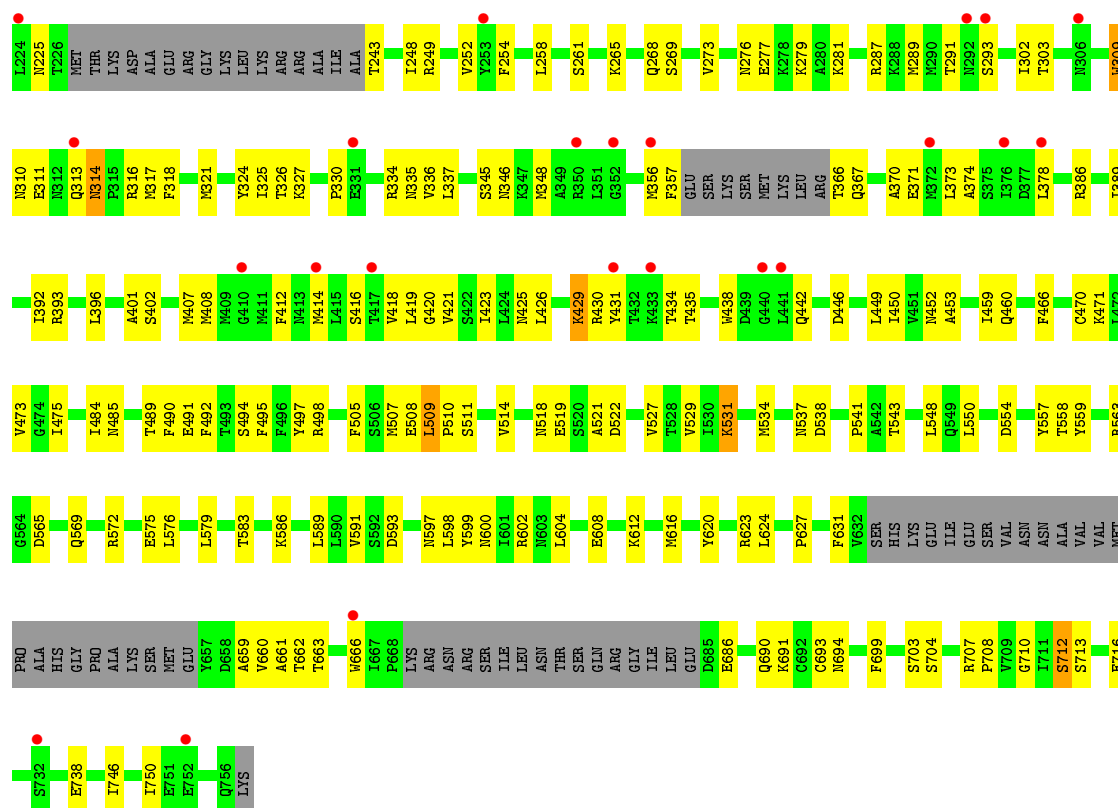


● Molecule 2: RNA-directed RNA polymerase catalytic subunit

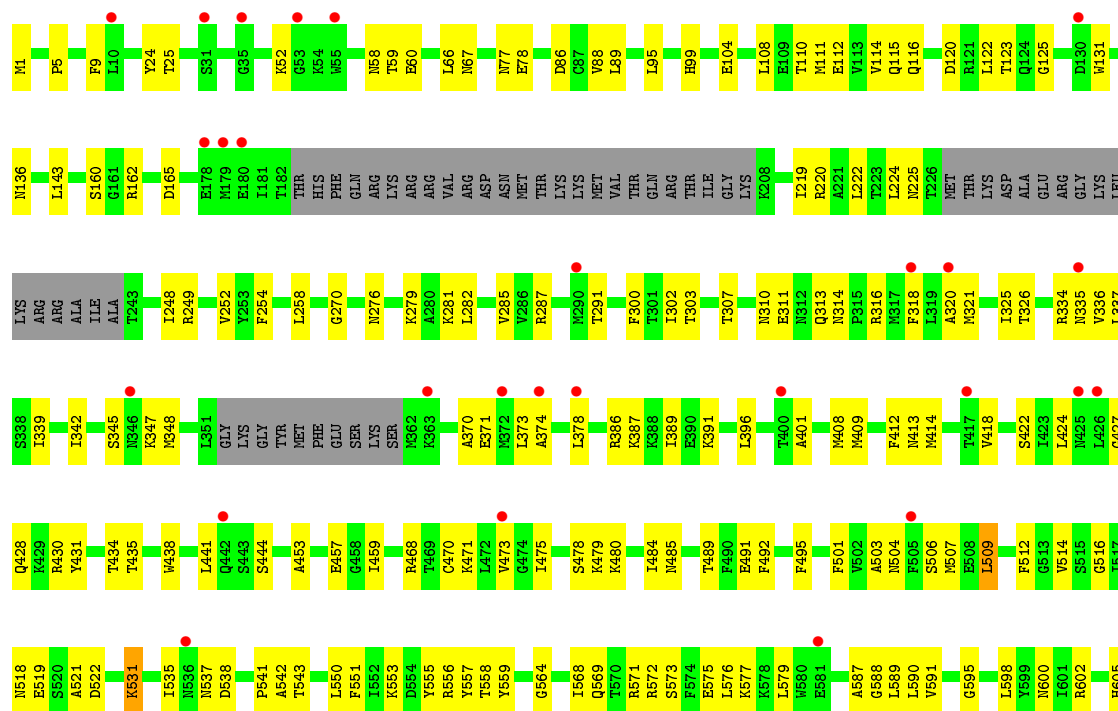


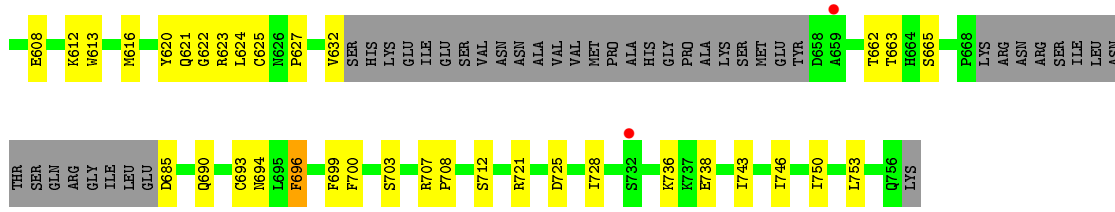
● Molecule 2: RNA-directed RNA polymerase catalytic subunit



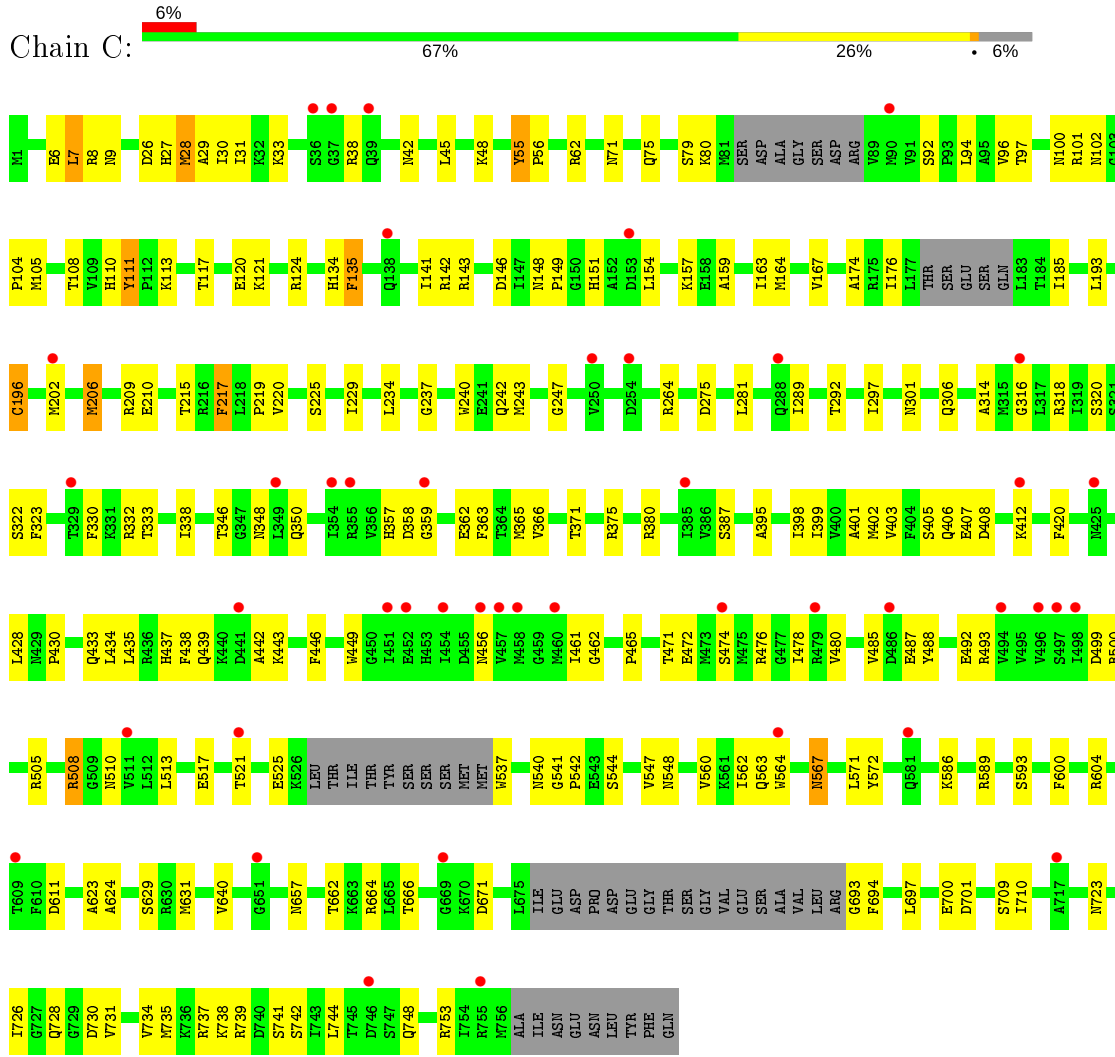


● Molecule 2: RNA-directed RNA polymerase catalytic subunit

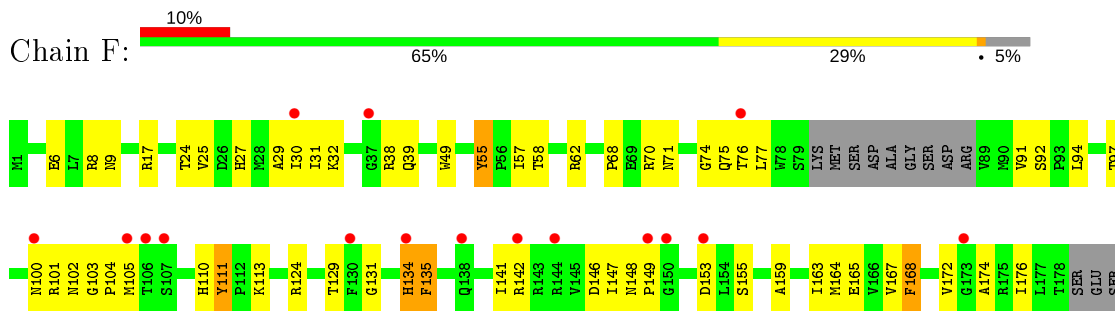




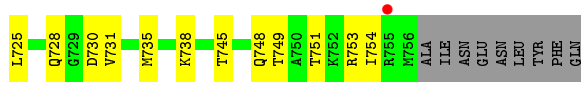
• Molecule 3: Polymerase basic protein 2



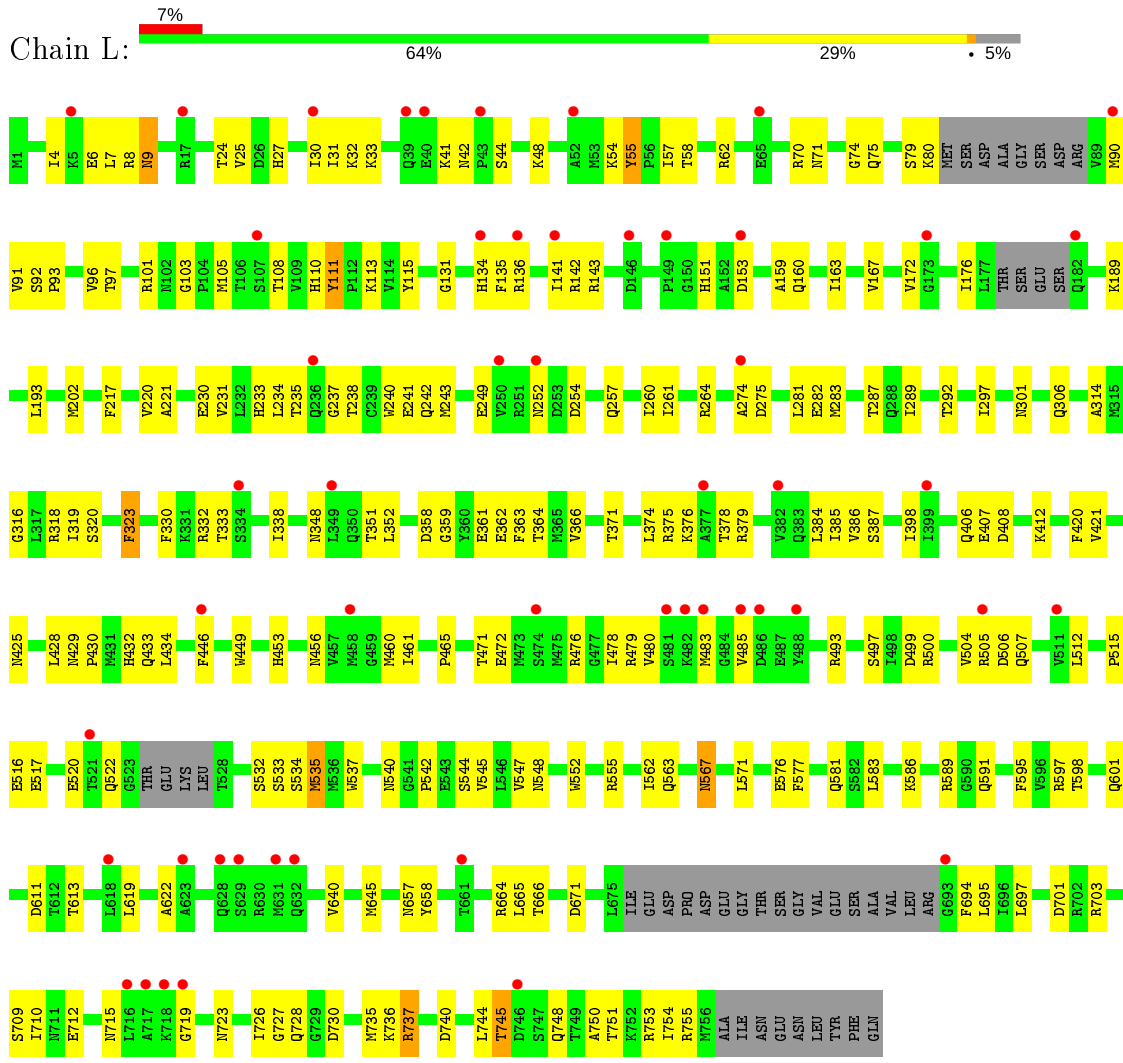
• Molecule 3: Polymerase basic protein 2



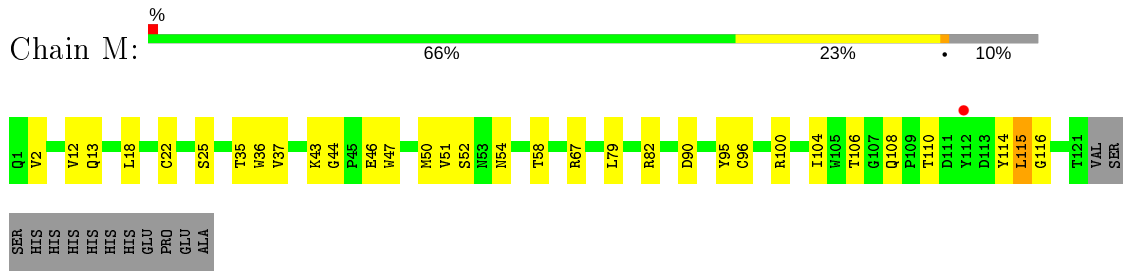




• Molecule 3: Polymerase basic protein 2

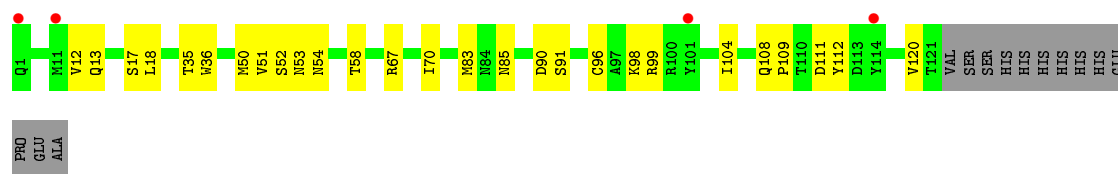


• Molecule 4: Nanobody NB8205

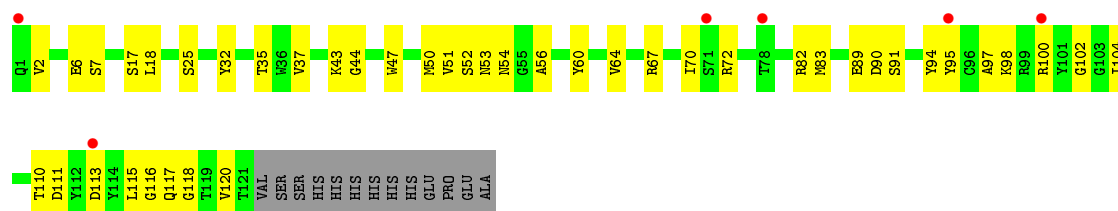


• Molecule 4: Nanobody NB8205

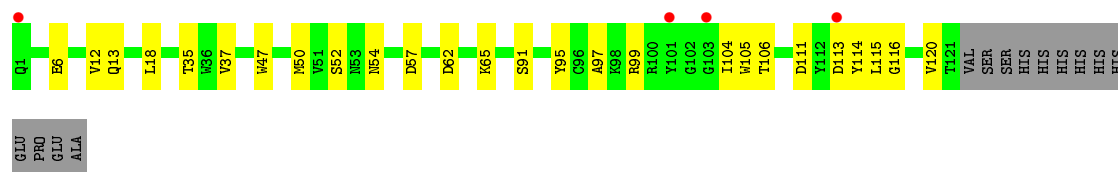




● Molecule 4: Nanobody NB8205



● Molecule 4: Nanobody NB8205



## 4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	335.18Å 192.90Å 235.10Å 90.00° 91.78° 90.00°	Depositor
Resolution (Å)	234.98 – 3.34 234.98 – 3.34	Depositor EDS
% Data completeness (in resolution range)	41.5 (234.98-3.34) 37.7 (234.98-3.34)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.41 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.13_2998	Depositor
R, $R_{free}$	0.251 , 0.305 0.251 , 0.305	Depositor DCC
$R_{free}$ test set	4533 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	69.1	Xtriage
Anisotropy	0.429	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 73.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.015 for -1/2*h-3/2*k,-1/2*h+1/2*k,-l 0.016 for -1/2*h+3/2*k,1/2*h+1/2*k,-l 0.017 for 1/2*h-3/2*k,-1/2*h-1/2*k,-l 0.016 for 1/2*h+3/2*k,1/2*h-1/2*k,-l 0.017 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.83	EDS
Total number of atoms	70549	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.48% 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	0.26	0/5742	0.46	0/7740
1	D	0.25	0/5773	0.45	1/7783 (0.0%)
1	G	0.25	0/5765	0.46	0/7772
1	J	0.25	0/5751	0.45	0/7753
2	B	0.26	0/5343	0.46	0/7214
2	E	0.26	0/5430	0.45	0/7323
2	H	0.26	0/5433	0.45	0/7334
2	K	0.25	0/5406	0.46	0/7297
3	C	0.25	0/5812	0.47	0/7825
3	F	0.25	0/5880	0.47	0/7921
3	I	0.25	0/5898	0.48	0/7944
3	L	0.25	0/5857	0.48	0/7887
4	M	0.26	0/954	0.52	1/1293 (0.1%)
4	N	0.25	0/954	0.47	0/1293
4	O	0.26	0/954	0.48	0/1293
4	P	0.26	0/954	0.48	0/1293
All	All	0.25	0/71906	0.46	2/96965 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	71	LEU	CA-CB-CG	-6.02	101.45	115.30
4	M	115	LEU	CA-CB-CG	5.42	127.76	115.30

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	5621	0	5563	161	0
1	D	5652	0	5583	146	0
1	G	5644	0	5577	152	0
1	J	5631	0	5555	144	0
2	B	5239	0	5190	155	0
2	E	5326	0	5286	167	0
2	H	5327	0	5270	175	0
2	K	5303	0	5261	166	0
3	C	5720	0	5867	147	0
3	F	5786	0	5931	188	0
3	I	5804	0	5952	174	0
3	L	5764	0	5907	174	0
4	M	933	0	893	22	0
4	N	933	0	893	18	0
4	O	933	0	893	34	0
4	P	933	0	893	16	0
All	All	70549	0	70514	1784	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 13.

The worst 5 of 1784 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)
3:I:290:GLY:H	3:I:529:ILE:HG12	1.32	0.95
3:F:134:HIS:HE1	3:F:241:GLU:HG2	1.35	0.91
2:B:667:ILE:HG13	3:C:56:PRO:O	1.71	0.90
1:A:714:ALA:O	1:A:716:ARG:NH1	2.07	0.88
1:A:353:LYS:HB3	1:G:353:LYS:HB3	1.57	0.87

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	
1	A	687/716 (96%)	638 (93%)	49 (7%)	0	100	100
1	D	691/716 (96%)	638 (92%)	53 (8%)	0	100	100
1	G	690/716 (96%)	633 (92%)	54 (8%)	3 (0%)	34	68
1	J	686/716 (96%)	629 (92%)	56 (8%)	1 (0%)	51	82
2	B	644/757 (85%)	600 (93%)	43 (7%)	1 (0%)	47	78
2	E	656/757 (87%)	616 (94%)	40 (6%)	0	100	100
2	H	655/757 (86%)	608 (93%)	46 (7%)	1 (0%)	47	78
2	K	652/757 (86%)	607 (93%)	45 (7%)	0	100	100
3	C	707/765 (92%)	644 (91%)	61 (9%)	2 (0%)	41	72
3	F	718/765 (94%)	655 (91%)	63 (9%)	0	100	100
3	I	720/765 (94%)	660 (92%)	60 (8%)	0	100	100
3	L	713/765 (93%)	644 (90%)	67 (9%)	2 (0%)	41	72
4	M	119/134 (89%)	115 (97%)	4 (3%)	0	100	100
4	N	119/134 (89%)	114 (96%)	5 (4%)	0	100	100
4	O	119/134 (89%)	116 (98%)	3 (2%)	0	100	100
4	P	119/134 (89%)	114 (96%)	5 (4%)	0	100	100
All	All	8695/9488 (92%)	8031 (92%)	654 (8%)	10 (0%)	51	82

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	431	ASP
3	L	516	GLU
2	B	429	LYS
1	J	70	ALA
3	L	745	THR

### 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	620/644 (96%)	606 (98%)	14 (2%)	50	75
1	D	624/644 (97%)	614 (98%)	10 (2%)	62	81
1	G	623/644 (97%)	613 (98%)	10 (2%)	62	81
1	J	622/644 (97%)	613 (99%)	9 (1%)	67	83
2	B	580/669 (87%)	570 (98%)	10 (2%)	60	80
2	E	590/669 (88%)	582 (99%)	8 (1%)	67	83
2	H	589/669 (88%)	580 (98%)	9 (2%)	65	82
2	K	588/669 (88%)	580 (99%)	8 (1%)	67	83
3	C	634/676 (94%)	613 (97%)	21 (3%)	38	68
3	F	643/676 (95%)	629 (98%)	14 (2%)	52	76
3	I	645/676 (95%)	630 (98%)	15 (2%)	50	75
3	L	640/676 (95%)	625 (98%)	15 (2%)	50	75
4	M	99/111 (89%)	98 (99%)	1 (1%)	76	87
4	N	99/111 (89%)	97 (98%)	2 (2%)	55	78
4	O	99/111 (89%)	98 (99%)	1 (1%)	76	87
4	P	99/111 (89%)	97 (98%)	2 (2%)	55	78
All	All	7794/8400 (93%)	7645 (98%)	149 (2%)	57	79

5 of 149 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	134	HIS
1	G	426	ASP
3	L	348	ASN
3	F	168	PHE
3	F	725	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
3	F	134	HIS
1	G	96	ASN
3	L	252	ASN
3	F	233	HIS
3	F	614	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

### 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	A	691/716 (96%)	0.52	20 (2%) 51 51	23, 62, 108, 163	0
1	D	695/716 (97%)	0.58	25 (3%) 42 41	37, 74, 119, 183	0
1	G	694/716 (96%)	0.59	27 (3%) 39 38	24, 72, 114, 173	0
1	J	692/716 (96%)	0.64	34 (4%) 29 29	39, 81, 124, 177	0
2	B	656/757 (86%)	0.50	16 (2%) 59 57	24, 61, 102, 159	0
2	E	666/757 (87%)	0.54	19 (2%) 51 51	32, 71, 108, 191	0
2	H	667/757 (88%)	0.68	43 (6%) 19 20	28, 75, 115, 151	0
2	K	664/757 (87%)	0.63	29 (4%) 34 35	35, 81, 116, 169	0
3	C	717/765 (93%)	0.62	44 (6%) 21 22	33, 79, 128, 168	0
3	F	726/765 (94%)	0.88	79 (10%) 5 5	38, 91, 144, 186	0
3	I	728/765 (95%)	0.63	39 (5%) 25 26	34, 77, 120, 159	0
3	L	723/765 (94%)	0.70	52 (7%) 15 16	33, 82, 127, 183	0
4	M	121/134 (90%)	0.39	1 (0%) 86 87	38, 63, 111, 130	0
4	N	121/134 (90%)	0.62	4 (3%) 46 45	54, 88, 126, 142	0
4	O	121/134 (90%)	0.51	6 (4%) 28 29	34, 66, 101, 132	0
4	P	121/134 (90%)	0.72	4 (3%) 46 45	56, 86, 118, 145	0
All	All	8803/9488 (92%)	0.62	442 (5%) 28 29	23, 76, 123, 191	0

The worst 5 of 442 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	L	717	ALA	7.9
1	D	65	LEU	7.1
3	F	717	ALA	6.6
2	E	31	SER	6.6
2	H	33	GLY	6.3

## 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 [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.