



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 9, 2024 – 11:54 am GMT

PDB ID : 5D98
Title : Influenza C Virus RNA-dependent RNA Polymerase - Space group P43212
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 : 3.90 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

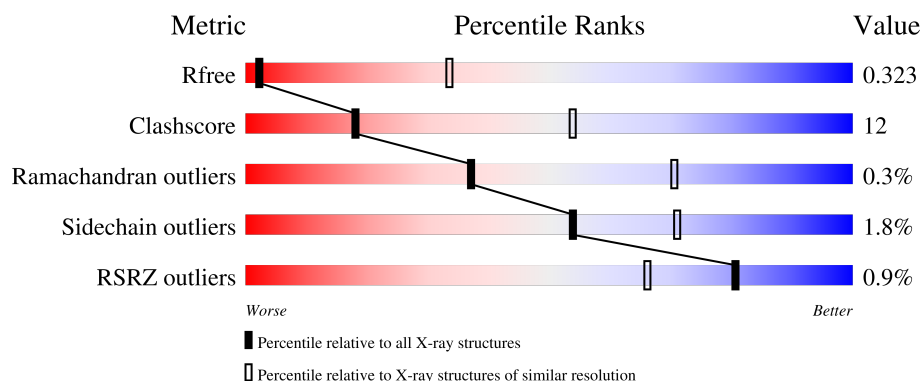
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.90 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}	164625	1157 (4.10-3.70)
Clashscore	180529	1219 (4.10-3.70)
Ramachandran outliers	177936	1177 (4.10-3.70)
Sidechain outliers	177891	1169 (4.10-3.70)
RSRZ outliers	164620	1157 (4.10-3.70)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	<div> <div></div> <div>67%</div> <div>30%</div> <div>..</div> </div>
1	D	709	<div> <div>%</div> <div>68%</div> <div>29%</div> <div>..</div> </div>
2	B	754	<div> <div>%</div> <div>74%</div> <div>19%</div> <div>6%</div> </div>
2	E	754	<div> <div>%</div> <div>73%</div> <div>20%</div> <div>6%</div> </div>
3	C	782	<div> <div>%</div> <div>70%</div> <div>26%</div> <div>..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	782	<div><div><div>%</div><div><div></div></div><div>67%</div><div>29%</div><div><div></div><div></div></div></div></div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 34720 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
1	A	693	Total	C	N	O	S	0	0	0
			5630	3589	954	1043	44			
1	D	693	Total	C	N	O	S	0	0	0
			5630	3589	954	1043	44			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	762	Total	C	N	O	S	0	0	0
			6076	3845	1066	1128	37			
3	F	762	Total	C	N	O	S	0	0	0
			6076	3845	1066	1128	37			

There are 16 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

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
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

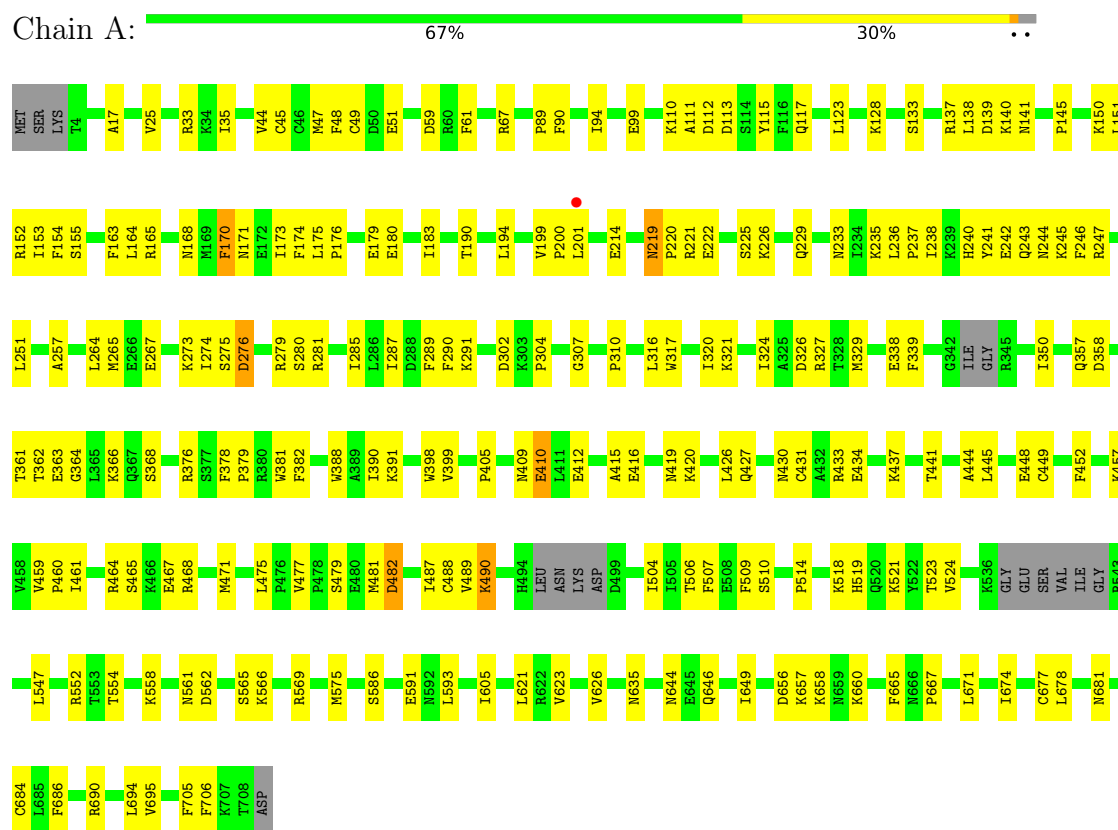
- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total Mg 2 2	0	0
4	D	2	Total Mg 2 2	0	0

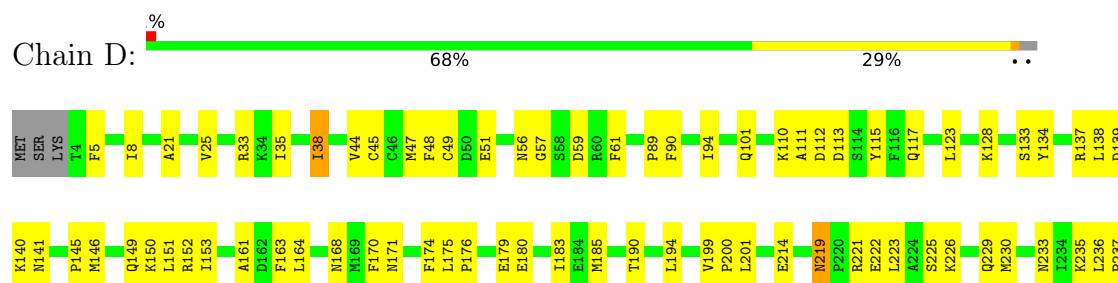
3 Residue-property plots

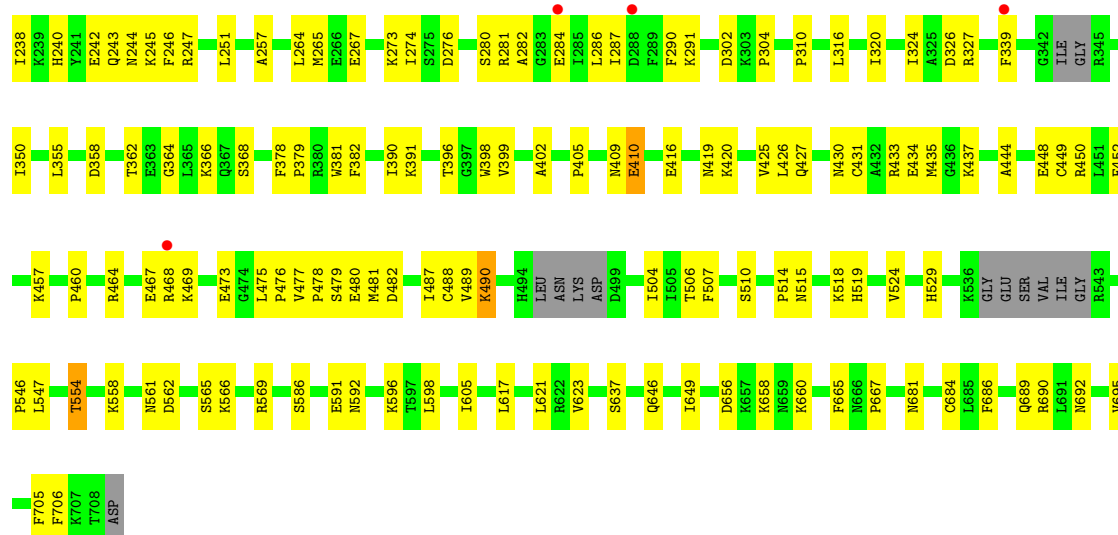
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density.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

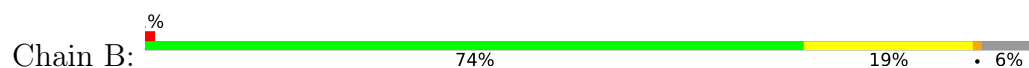


• Molecule 1: Polymerase acidic protein

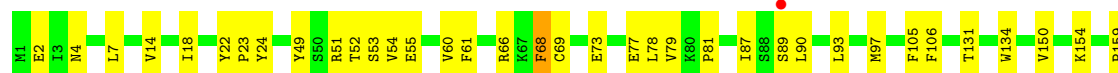
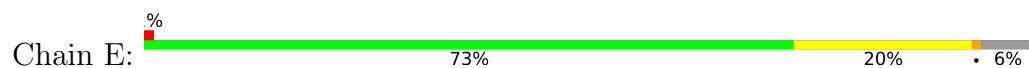


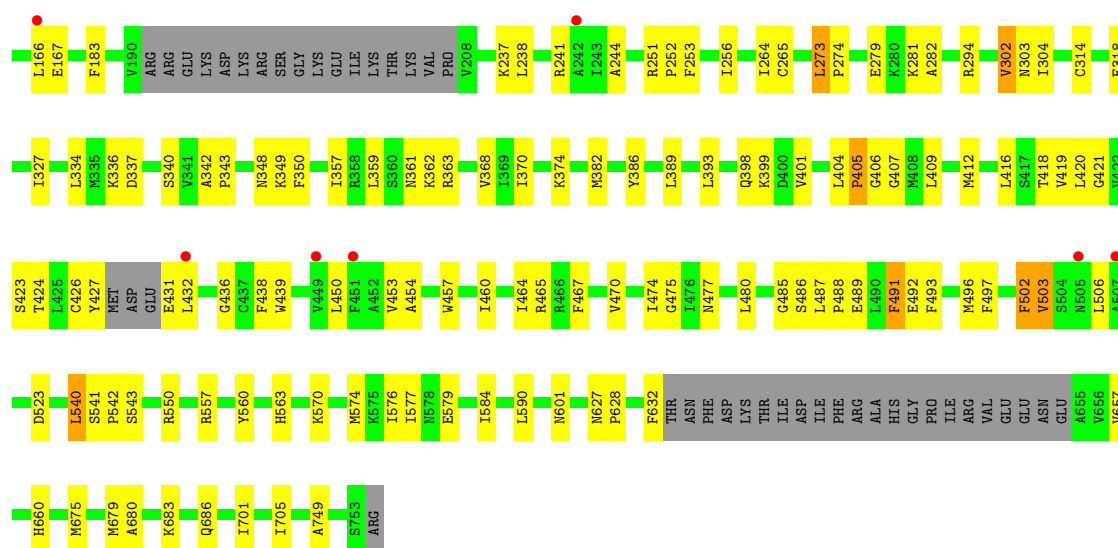


• Molecule 2: RNA-directed RNA polymerase catalytic subunit

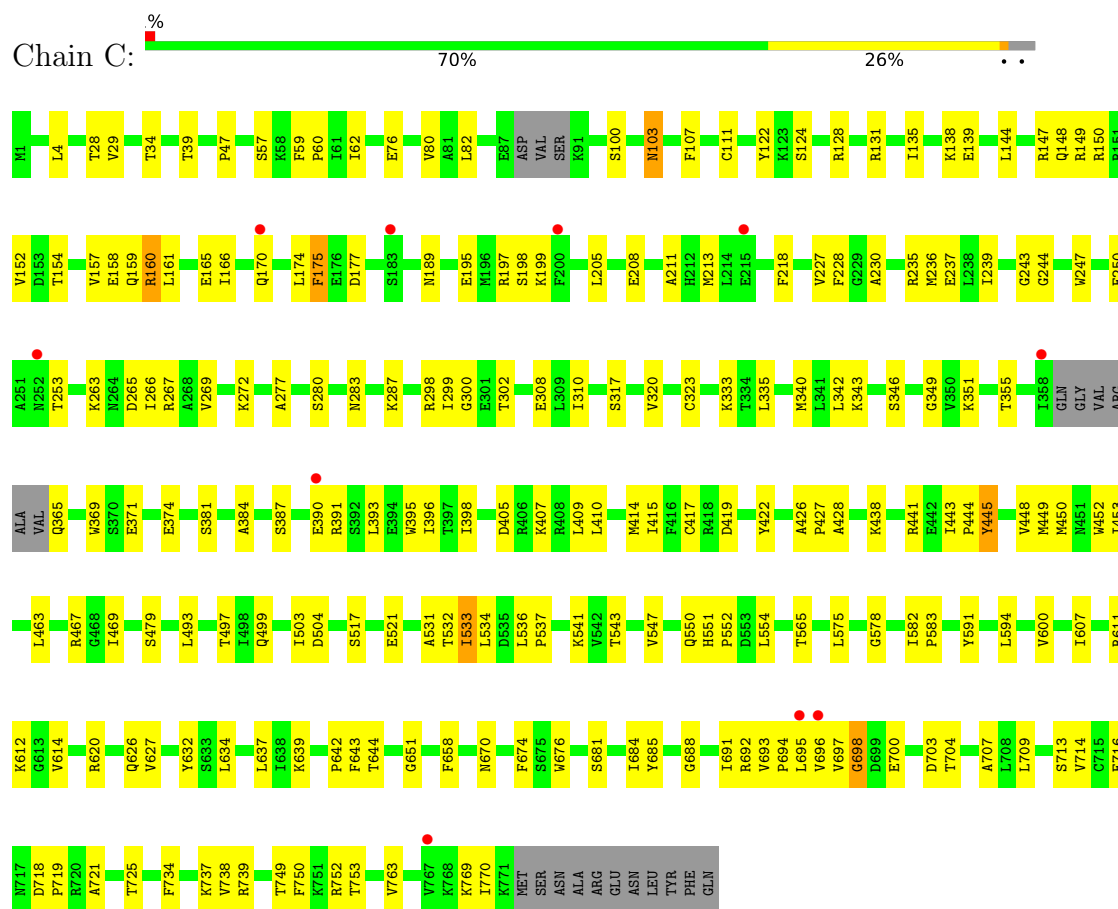


• Molecule 2: RNA-directed RNA polymerase catalytic subunit





• Molecule 3: Polymerase basic protein 2



• Molecule 3: Polymerase basic protein 2



GLN	G688	Q550	P427	K343	I239	H1
	I691	H551	A428	S346	W247	L4
	R692	P552	T429	G349	I248	Q16
	V693	D553	K438	V550	E250	T28
	P694	L554	R441	K351	T253	V29
	L695	T665	P444	T355	A254	K37
	Y696	L575	Y445	I358	K263	W38
	D703	G578	V448	GLN	W264	T39
	T704	M449	M450	GLY	D265	L49
	S713	I582	M451	VAL	I266	L57
	V714	P583	W452	ARG	R267	E158
	G715	I586	I453	ALA	A268	R52
	E716	Y591	L463	VAL	V269	W53
	H717	L594	S466	Q365	C270	S56
	D718	Y595	R467	W369	R271	S57
	P719	V600	I469	S370	K272	R58
	R720	I607	T472	E371	V273	F59
	A721	R611	K476	E374	C274	P60
	T725	K612	S479	E374	L275	E76
	F734	G613	S480	S381	A276	H77
	K737	V614	S492	A384	S280	V80
	V738	R620	L493	S387	N283	A81
	R739	Q626	T497	E390	K287	L82
	L740	V627	I498	R391	L288	E87
	F741	Y632	Q499	S392	T292	ASP
	Q744	L634	I503	L393	R298	VAL
	V747	L637	D504	E394	T299	SER
	R748	P642	I512	W395	G300	K91
	T749	F643	H513	I396	E308	I102
	F750	T644	S517	T397	L309	M103
	R751	G651	E521	I398	P204	N106
	R752	F658	A531	D405	L205	C111
	T753	N670	T532	R406	E208	V117
	A754	W676	I533	K407	A211	V121
	A758	L534	L534	R408	H212	Y122
	D762	W676	D535	L409	W213	K123
	K769	L536	L536	L410	F218	S124
	I770	P537	P537	A411	K324	R125
	K771	S681	T543	M412	C323	F126
	HET	I684	V547	C413	M340	G127
	SER	Y685		M414		R128
	ASN			I415		L129
	ALA			F416		F130
	ARG			C417		R131
	GLU			R418		I135
	ASN			D419		M136
	LEU			Y422		
	TYR			A426		
	PHE					

4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	185.66Å 185.66Å 598.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.01 – 3.90 50.01 – 3.90	Depositor EDS
% Data completeness (in resolution range)	98.8 (50.01-3.90) 98.8 (50.01-3.90)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.92 (at 3.89Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.286 , 0.326 0.286 , 0.323	Depositor DCC
R_{free} test set	4770 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	161.2	Xtriage
Anisotropy	0.644	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 156.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	34720	wwPDB-VP
Average B, all atoms (Å ²)	207.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.59% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: MG

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	0.41	0/5746	0.57	1/7717 (0.0%)
1	D	0.43	0/5746	0.57	1/7717 (0.0%)
2	B	0.40	0/5749	0.56	0/7723
2	E	0.41	0/5749	0.57	0/7723
3	C	0.41	0/6185	0.59	0/8322
3	F	0.42	0/6185	0.59	0/8322
All	All	0.41	0/35360	0.58	2/47524 (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	A	219	ASN	C-N-CD	5.11	139.14	128.40
1	D	219	ASN	C-N-CD	5.11	139.12	128.40

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	147	0
1	D	5630	0	5632	167	0
2	B	5652	0	5749	127	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	5652	0	5749	126	0
3	C	6076	0	6183	182	0
3	F	6076	0	6183	198	0
4	A	2	0	0	0	0
4	D	2	0	0	0	0
All	All	34720	0	35128	840	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 12.

The worst 5 of 840 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:F:138:LYS:HE2	3:F:250:GLU:CG	1.22	1.59
3:F:138:LYS:CE	3:F:250:GLU:HG2	1.34	1.48
3:F:138:LYS:CE	3:F:250:GLU:CG	1.83	1.46
3:F:138:LYS:CE	3:F:250:GLU:CD	1.86	1.41
3:F:138:LYS:NZ	3:F:250:GLU:HG2	1.33	1.36

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	685/709 (97%)	643 (94%)	42 (6%)	0	100	100
1	D	685/709 (97%)	645 (94%)	40 (6%)	0	100	100
2	B	703/754 (93%)	666 (95%)	35 (5%)	2 (0%)	37	70
2	E	703/754 (93%)	673 (96%)	27 (4%)	3 (0%)	30	65
3	C	756/782 (97%)	688 (91%)	63 (8%)	5 (1%)	19	54

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	F	756/782 (97%)	686 (91%)	67 (9%)	3 (0%)	30	65
All	All	4288/4490 (96%)	4001 (93%)	274 (6%)	13 (0%)	37	70

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	405	PRO
3	C	698	GLY
2	E	405	PRO
2	B	503	VAL
2	E	503	VAL

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	618/631 (98%)	603 (98%)	15 (2%)	44	63
1	D	618/631 (98%)	608 (98%)	10 (2%)	58	73
2	B	629/669 (94%)	620 (99%)	9 (1%)	62	75
2	E	629/669 (94%)	618 (98%)	11 (2%)	56	72
3	C	669/686 (98%)	660 (99%)	9 (1%)	65	76
3	F	669/686 (98%)	655 (98%)	14 (2%)	48	67
All	All	3832/3972 (96%)	3764 (98%)	68 (2%)	54	71

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	152	VAL
3	F	175	PHE
3	F	611	ARG
3	C	4	LEU
2	B	540	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	316	GLN
3	F	550	GLN
2	E	477	ASN
2	E	660	HIS
3	F	744	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 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 ⓘ

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 [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, 95th 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 ²)	Q<0.9
1	A	693/709 (97%)	-0.46	1 (0%) 92 89	132, 195, 269, 299	0
1	D	693/709 (97%)	-0.36	4 (0%) 85 73	129, 191, 258, 295	0
2	B	711/754 (94%)	-0.45	7 (0%) 79 64	131, 192, 256, 317	0
2	E	711/754 (94%)	-0.37	8 (1%) 77 62	130, 184, 266, 315	0
3	C	762/782 (97%)	-0.23	10 (1%) 74 58	149, 222, 285, 358	0
3	F	762/782 (97%)	-0.14	10 (1%) 74 58	134, 224, 287, 343	0
All	All	4332/4490 (96%)	-0.33	40 (0%) 81 66	129, 202, 275, 358	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	695	LEU	5.0
3	C	696	VAL	4.0
2	E	242	ALA	3.7
2	E	166	LEU	3.4
3	F	586	ILE	3.1

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 monosaccharides in this entry.

6.4 Ligands [i](#)

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, 95th 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(Å ²)	Q<0.9
4	MG	D	800	1/1	0.72	0.15	200,200,200,200	0
4	MG	A	800	1/1	0.74	0.10	204,204,204,204	0
4	MG	D	801	1/1	0.74	0.14	182,182,182,182	0
4	MG	A	801	1/1	0.93	0.06	185,185,185,185	0

6.5 Other polymers [i](#)

There are no such residues in this entry.