



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 07:36 PM UTC

PDB ID : 5X50 / pdb_00005x50
Title : RNA Polymerase II from Komagataella Pastoris (Type-2 crystal)
Authors : Ehara, H.; Umehara, T.; Sekine, S.; Yokoyama, S.
Deposited on : 2017-02-14
Resolution : 4.29 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

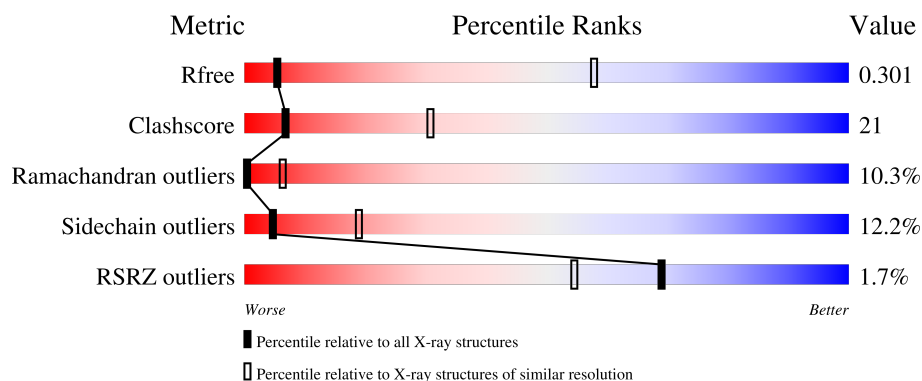
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.29 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}	180053	1052 (4.70-3.90)
Clashscore	190562	1097 (4.70-3.90)
Ramachandran outliers	187476	1001 (4.70-3.90)
Sidechain outliers	187428	1007 (4.72-3.88)
RSRZ outliers	180081	1049 (4.70-3.90)

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	1743	
2	B	1227	
3	C	304	
4	D	186	
5	E	214	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	F	155	%23%26%5%46%
7	G	171	2%50%42%8%.
8	H	145	3%50%31%8%.10%
9	I	115	3%50%40%7%..
10	J	72	%28%49%8%.11%
11	K	118	2%44%42%9%5%
12	L	73	27%26%8%.37%

2 Entry composition

There are 13 unique types of molecules in this entry. The entry contains 28437 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 DNA-directed RNA polymerase subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1409	Total	C	N	O	S	0	0	0
			10410	6594	1819	1943	54			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1079	Total	C	N	O	S	0	0	0
			8241	5236	1442	1517	46			

- Molecule 3 is a protein called RNA polymerase II third largest subunit B44, part of central core.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	263	Total	C	N	O	S	0	0	0
			1852	1181	308	355	8			

- Molecule 4 is a protein called RNA polymerase II subunit B32.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	143	Total	C	N	O	S	0	0	0
			1004	647	170	185	2			

- Molecule 5 is a protein called RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	214	Total	C	N	O	S	0	0	0
			1638	1045	288	297	8			

- Molecule 6 is a protein called RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	84	Total	C	N	O	S	0	0	0
			637	406	109	119	3			

- Molecule 7 is a protein called RNA polymerase II subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	171	Total	C	N	O	S	0	0	0
			1187	775	192	217	3			

- Molecule 8 is a protein called RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	131	Total	C	N	O	S	0	0	0
			965	616	155	191	3			

- Molecule 9 is a protein called DNA-directed RNA polymerase subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	113	Total	C	N	O	S	0	0	0
			856	537	150	158	11			

- Molecule 10 is a protein called RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	64	Total	C	N	O	S	0	0	0
			500	327	87	80	6			

- Molecule 11 is a protein called RNA polymerase II subunit B12.5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	112	Total	C	N	O	S	0	0	0
			820	535	137	146	2			

- Molecule 12 is a protein called RNA polymerase subunit, found in RNA polymerase complexes I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	46	Total	C	N	O	S	0	0	0
			319	196	64	55	4			

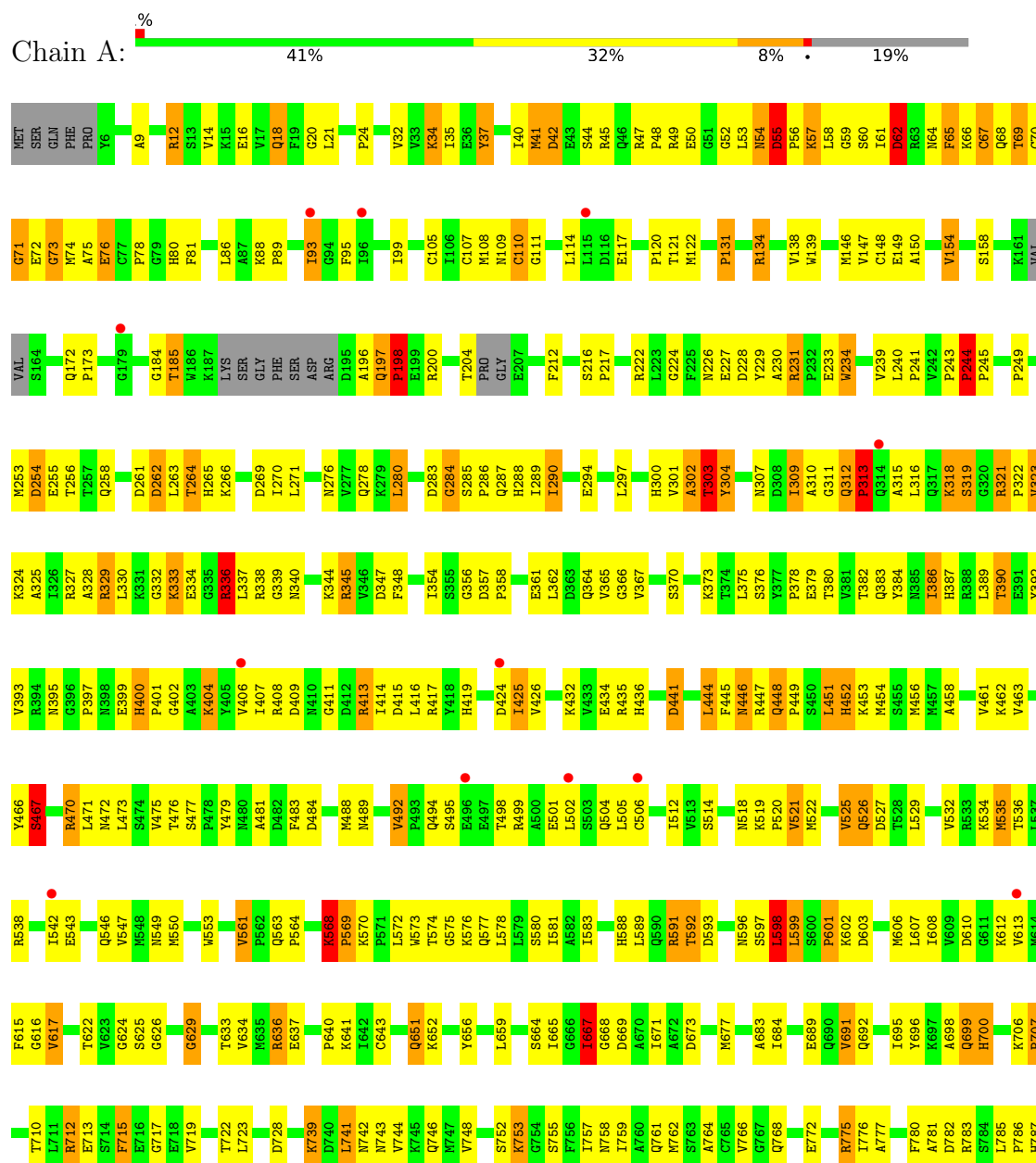
- Molecule 13 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	A	2	Total 2	Zn 2	0	0
13	B	1	Total 1	Zn 1	0	0
13	C	1	Total 1	Zn 1	0	0
13	I	2	Total 2	Zn 2	0	0
13	J	1	Total 1	Zn 1	0	0
13	L	1	Total 1	Zn 1	0	0

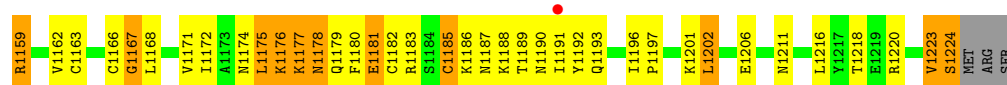
3 Residue-property plots

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.

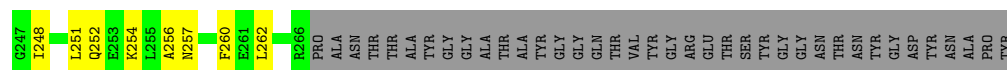
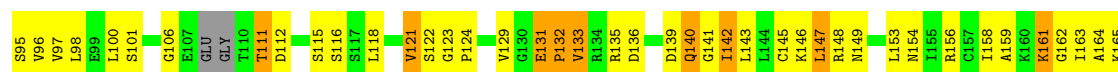
• Molecule 1: DNA-directed RNA polymerase subunit



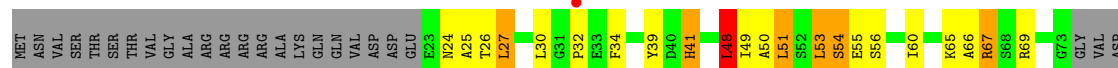
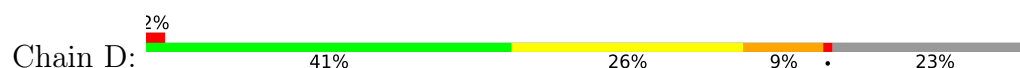




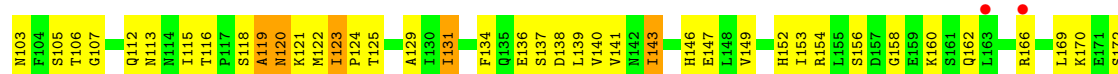
- Molecule 3: RNA polymerase II third largest subunit B44, part of central core



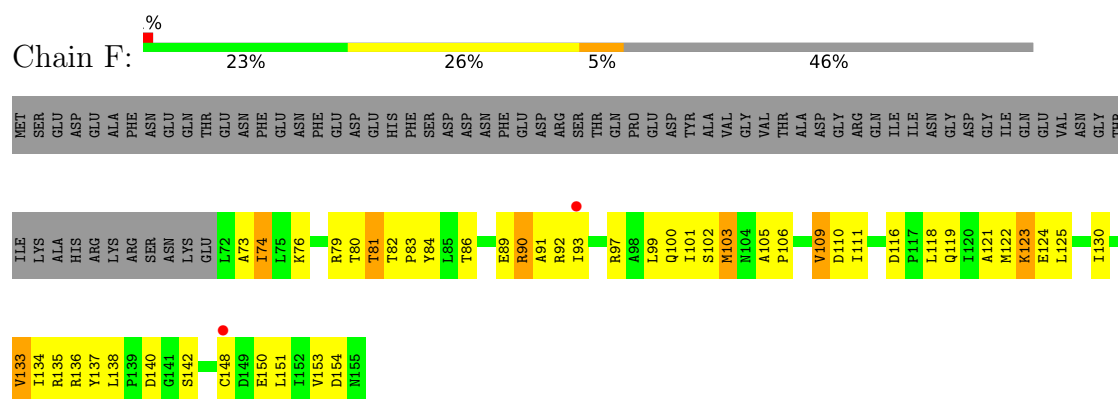
- Molecule 4: RNA polymerase II subunit B32



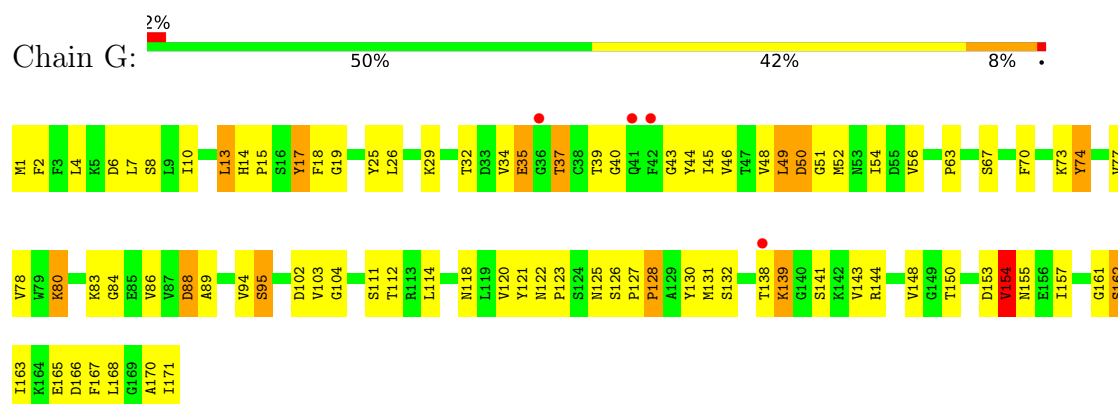
- Molecule 5: RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III



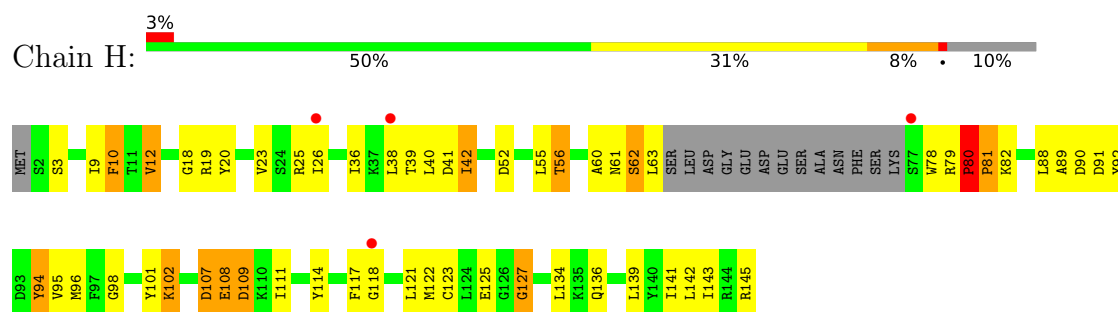
- Molecule 6: RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III



• Molecule 7: RNA polymerase II subunit



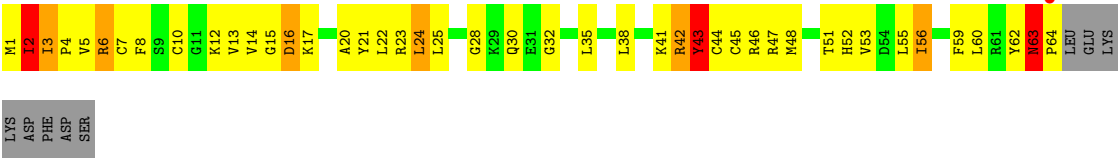
• Molecule 8: RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III



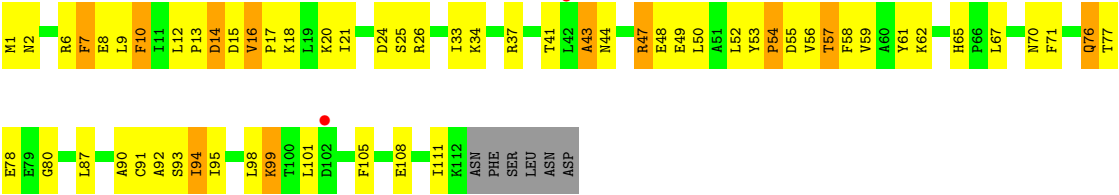
• Molecule 9: DNA-directed RNA polymerase subunit



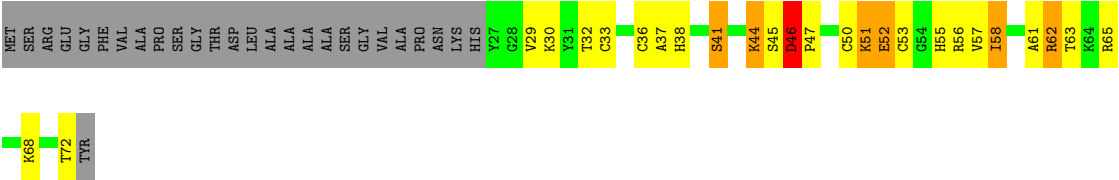
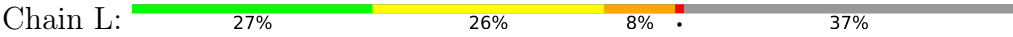
• Molecule 10: RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III



• Molecule 11: RNA polymerase II subunit B12.5



• Molecule 12: RNA polymerase subunit, found in RNA polymerase complexes I, II, and III



4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α , β , γ	211.75Å 211.75Å 132.28Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.70 – 4.29 40.70 – 4.29	Depositor EDS
% Data completeness (in resolution range)	99.6 (40.70-4.29) 99.5 (40.70-4.29)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.51 (at 4.29Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.235 , 0.296 0.237 , 0.301	Depositor DCC
R_{free} test set	1992 reflections (4.40%)	wwPDB-VP
Wilson B-factor (Å ²)	191.6	Xtriage
Anisotropy	0.254	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 195.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.37$, $\langle L^2 \rangle = 0.20$	Xtriage
Estimated twinning fraction	0.124 for -h,-k,l 0.128 for h,-h-k,-l 0.125 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	28437	wwPDB-VP
Average B, all atoms (Å ²)	223.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.87% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section:
ZN

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	0.42	1/10569 (0.0%)	1.19	94/14276 (0.7%)
2	B	0.40	0/8386	1.08	41/11301 (0.4%)
3	C	0.44	0/1877	1.28	21/2543 (0.8%)
4	D	0.39	0/1010	1.30	10/1357 (0.7%)
5	E	0.39	0/1668	1.05	11/2245 (0.5%)
6	F	0.39	0/646	1.16	6/873 (0.7%)
7	G	0.40	0/1207	1.14	9/1629 (0.6%)
8	H	0.41	1/980 (0.1%)	0.98	1/1324 (0.1%)
9	I	0.40	0/871	1.06	4/1175 (0.3%)
10	J	0.48	0/509	1.16	2/684 (0.3%)
11	K	0.37	0/836	1.08	5/1131 (0.4%)
12	L	0.45	0/321	1.23	2/425 (0.5%)
All	All	0.41	2/28880 (0.0%)	1.15	206/38963 (0.5%)

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
10	J	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	887	ILE	CA-CB	7.80	1.58	1.54
8	H	80	PRO	CA-C	5.16	1.56	1.52

The worst 5 of 206 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	123	GLY	CA-C-N	15.15	134.67	119.82
3	C	123	GLY	C-N-CA	15.15	134.67	119.82
1	A	1279	ILE	CA-C-N	14.68	134.87	119.89
1	A	1279	ILE	C-N-CA	14.68	134.87	119.89
4	D	183	ASP	CA-C-N	14.42	134.31	120.03

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
10	J	30	GLN	Peptide

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	10410	0	9856	480	0
2	B	8241	0	7943	401	0
3	C	1852	0	1639	79	0
4	D	1004	0	913	42	0
5	E	1638	0	1551	49	0
6	F	637	0	620	35	0
7	G	1187	0	1092	56	0
8	H	965	0	864	39	0
9	I	856	0	774	46	0
10	J	500	0	503	55	0
11	K	820	0	719	39	0
12	L	319	0	288	14	0
13	A	2	0	0	0	0
13	B	1	0	0	0	0
13	C	1	0	0	0	0
13	I	2	0	0	0	0
13	J	1	0	0	0	0
13	L	1	0	0	0	0
All	All	28437	0	26762	1179	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 21.

The worst 5 of 1179 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)
2:B:285:PRO:HB2	9:I:11:ASN:HB3	1.30	1.13
1:A:256:THR:H	2:B:935:ARG:HH12	1.10	0.97
1:A:67:CYS:SG	1:A:80:HIS:NE2	2.41	0.93
2:B:777:ALA:HA	2:B:1095:LEU:HA	1.56	0.87
1:A:226:ASN:HD22	1:A:229:TYR:H	1.23	0.86

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	1393/1743 (80%)	972 (70%)	291 (21%)	130 (9%)	0	8
2	B	1055/1227 (86%)	723 (68%)	214 (20%)	118 (11%)	0	6
3	C	259/304 (85%)	186 (72%)	50 (19%)	23 (9%)	0	8
4	D	135/186 (73%)	98 (73%)	23 (17%)	14 (10%)	0	6
5	E	212/214 (99%)	150 (71%)	44 (21%)	18 (8%)	0	9
6	F	82/155 (53%)	58 (71%)	14 (17%)	10 (12%)	0	4
7	G	169/171 (99%)	129 (76%)	27 (16%)	13 (8%)	1	10
8	H	127/145 (88%)	95 (75%)	17 (13%)	15 (12%)	0	4
9	I	111/115 (96%)	71 (64%)	26 (23%)	14 (13%)	0	4
10	J	62/72 (86%)	37 (60%)	15 (24%)	10 (16%)	0	2
11	K	110/118 (93%)	83 (76%)	16 (14%)	11 (10%)	0	7
12	L	44/73 (60%)	20 (46%)	11 (25%)	13 (30%)	0	0
All	All	3759/4523 (83%)	2622 (70%)	748 (20%)	389 (10%)	0	6

5 of 389 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	42	ASP
1	A	48	PRO
1	A	54	ASN
1	A	57	LYS
1	A	67	CYS

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	988/1528 (65%)	875 (89%)	113 (11%)	5	20
2	B	822/1077 (76%)	724 (88%)	98 (12%)	5	19
3	C	155/264 (59%)	131 (84%)	24 (16%)	2	13
4	D	78/160 (49%)	64 (82%)	14 (18%)	2	11
5	E	155/197 (79%)	140 (90%)	15 (10%)	8	25
6	F	60/137 (44%)	53 (88%)	7 (12%)	5	19
7	G	102/148 (69%)	85 (83%)	17 (17%)	2	12
8	H	91/130 (70%)	82 (90%)	9 (10%)	7	24
9	I	82/109 (75%)	74 (90%)	8 (10%)	7	25
10	J	48/66 (73%)	39 (81%)	9 (19%)	1	9
11	K	67/109 (62%)	56 (84%)	11 (16%)	2	12
12	L	26/58 (45%)	24 (92%)	2 (8%)	12	33
All	All	2674/3983 (67%)	2347 (88%)	327 (12%)	5	19

5 of 327 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	234	THR
7	G	163	ILE
4	D	48	LEU
5	E	113	ASN
9	I	50	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 73 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	167	HIS
11	K	76	GLN
3	C	252	GLN
5	E	142	ASN
1	A	1213	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 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 ⓘ

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, 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	1409/1743 (80%)	-0.04	26 (1%) 67 53	110, 206, 318, 455	0
2	B	1079/1227 (87%)	-0.01	15 (1%) 73 58	115, 226, 316, 418	0
3	C	263/304 (86%)	0.04	2 (0%) 82 68	148, 216, 311, 349	0
4	D	143/186 (76%)	0.11	4 (2%) 55 42	167, 263, 321, 343	0
5	E	214/214 (100%)	-0.03	3 (1%) 73 58	122, 197, 260, 300	0
6	F	84/155 (54%)	-0.04	2 (2%) 59 47	160, 222, 270, 312	0
7	G	171/171 (100%)	0.14	4 (2%) 61 48	156, 213, 267, 311	0
8	H	131/145 (90%)	0.13	4 (3%) 51 40	161, 214, 298, 344	0
9	I	113/115 (98%)	0.08	4 (3%) 47 37	211, 276, 343, 386	0
10	J	64/72 (88%)	0.05	1 (1%) 70 55	154, 214, 272, 288	0
11	K	112/118 (94%)	-0.03	2 (1%) 67 53	149, 202, 274, 293	0
12	L	46/73 (63%)	-0.01	0 100 100	223, 275, 329, 350	0
All	All	3829/4523 (84%)	-0.00	67 (1%) 69 54	110, 219, 314, 455	0

The worst 5 of 67 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	115	LEU	4.0
8	H	118	GLY	3.9
4	D	32	PRO	3.7
7	G	42	PHE	3.5
2	B	1191	ILE	3.4

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

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

6.3 Carbohydrates [i](#)

There are no oligosaccharides 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($\text{\AA}^2$)	Q<0.9
13	ZN	A	1801	1/1	0.94	0.06	299,299,299,299	0
13	ZN	A	1802	1/1	0.95	0.06	270,270,270,270	0
13	ZN	C	401	1/1	0.95	0.06	283,283,283,283	0
13	ZN	I	201	1/1	0.96	0.05	239,239,239,239	0
13	ZN	L	101	1/1	0.97	0.06	318,318,318,318	0
13	ZN	J	101	1/1	0.98	0.07	158,158,158,158	0
13	ZN	I	202	1/1	0.98	0.05	242,242,242,242	0
13	ZN	B	1301	1/1	0.99	0.03	248,248,248,248	0

6.5 Other polymers [i](#)

There are no such residues in this entry.