wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 24, 2024 – 02:13 PM EDT

PDB ID	:	5TBZ
Title	:	E. Coli RNA Polymerase complexed with NusG
Authors	:	Liu, B.; Steitz, T.A.
Deposited on	:	2016-09-13
Resolution	:	7.00 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	:	2.37.1
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)	:	2.37.1
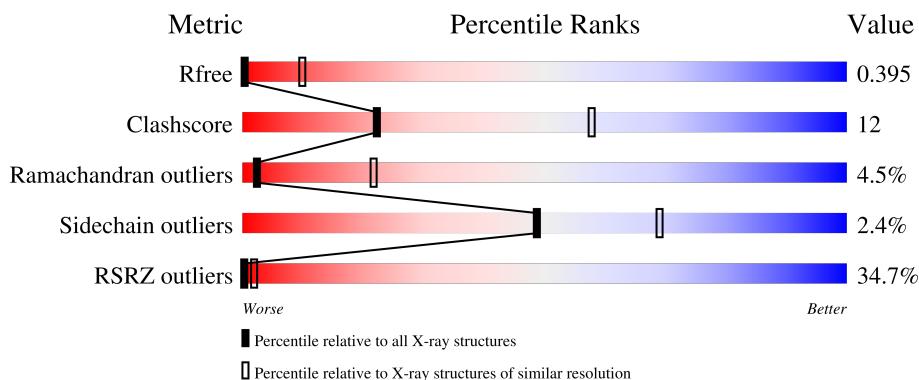
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION
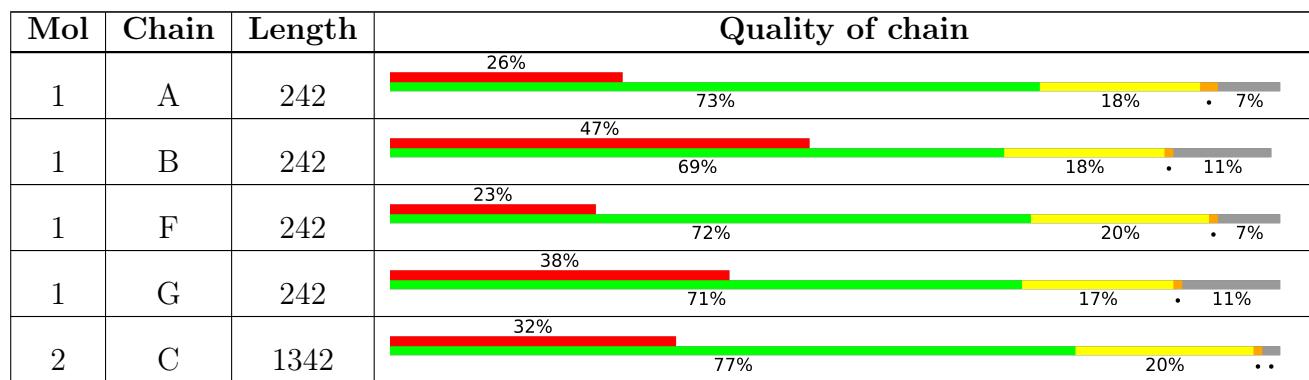
The reported resolution of this entry is 7.00 Å.

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	Percentile Ranks	Value
Rfree		0.395
Clashscore		12
Ramachandran outliers		4.5%
Sidechain outliers		2.4%
RSRZ outliers		34.7%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1004 (10.00-3.90)
Clashscore	141614	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)
Sidechain outliers	138945	1002 (10.00-3.86)
RSRZ outliers	127900	1004 (9.50-3.80)

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 <=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	242	26% / 73% 18% · 7%
1	B	242	47% / 69% 18% · 11%
1	F	242	23% / 72% 20% · 7%
1	G	242	38% / 71% 17% · 11%
2	C	1342	32% / 77% 20% ··

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	1342	34% ██ 75% ███ 21% ··
3	D	1407	29% ██ 63% ███ 26% · 7%
3	I	1407	35% ██ 64% ███ 25% · 7%
4	J	181	27% ██ 61% ███ 18% · 19%
4	K	181	33% ██ 56% ███ 22% ·· 19%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 50147 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 alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	225	Total	C	N	O	S	0	0	0
			1740	1082	309	343	6			
1	B	215	Total	C	N	O	S	0	0	0
			1657	1034	291	326	6			
1	F	225	Total	C	N	O	S	0	0	0
			1740	1082	309	343	6			
1	G	216	Total	C	N	O	S	0	0	0
			1667	1040	294	327	6			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	ALA	-	expression tag	UNP P0A7Z6
A	-5	HIS	-	expression tag	UNP P0A7Z6
A	-4	HIS	-	expression tag	UNP P0A7Z6
A	-3	HIS	-	expression tag	UNP P0A7Z6
A	-2	HIS	-	expression tag	UNP P0A7Z6
A	-1	HIS	-	expression tag	UNP P0A7Z6
A	0	HIS	-	expression tag	UNP P0A7Z6
B	-6	ALA	-	expression tag	UNP P0A7Z6
B	-5	HIS	-	expression tag	UNP P0A7Z6
B	-4	HIS	-	expression tag	UNP P0A7Z6
B	-3	HIS	-	expression tag	UNP P0A7Z6
B	-2	HIS	-	expression tag	UNP P0A7Z6
B	-1	HIS	-	expression tag	UNP P0A7Z6
B	0	HIS	-	expression tag	UNP P0A7Z6
F	-6	ALA	-	expression tag	UNP P0A7Z6
F	-5	HIS	-	expression tag	UNP P0A7Z6
F	-4	HIS	-	expression tag	UNP P0A7Z6
F	-3	HIS	-	expression tag	UNP P0A7Z6
F	-2	HIS	-	expression tag	UNP P0A7Z6
F	-1	HIS	-	expression tag	UNP P0A7Z6
F	0	HIS	-	expression tag	UNP P0A7Z6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	-6	ALA	-	expression tag	UNP P0A7Z6
G	-5	HIS	-	expression tag	UNP P0A7Z6
G	-4	HIS	-	expression tag	UNP P0A7Z6
G	-3	HIS	-	expression tag	UNP P0A7Z6
G	-2	HIS	-	expression tag	UNP P0A7Z6
G	-1	HIS	-	expression tag	UNP P0A7Z6
G	0	HIS	-	expression tag	UNP P0A7Z6

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1319	Total 10401	C 6524	N 1814	O 2020	S 43	0	1	0
2	H	1319	Total 10401	C 6524	N 1814	O 2020	S 43	0	1	0

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1302	Total 10085	C 6326	N 1800	O 1911	S 48	0	0	0
3	I	1306	Total 10126	C 6353	N 1809	O 1916	S 48	0	0	0
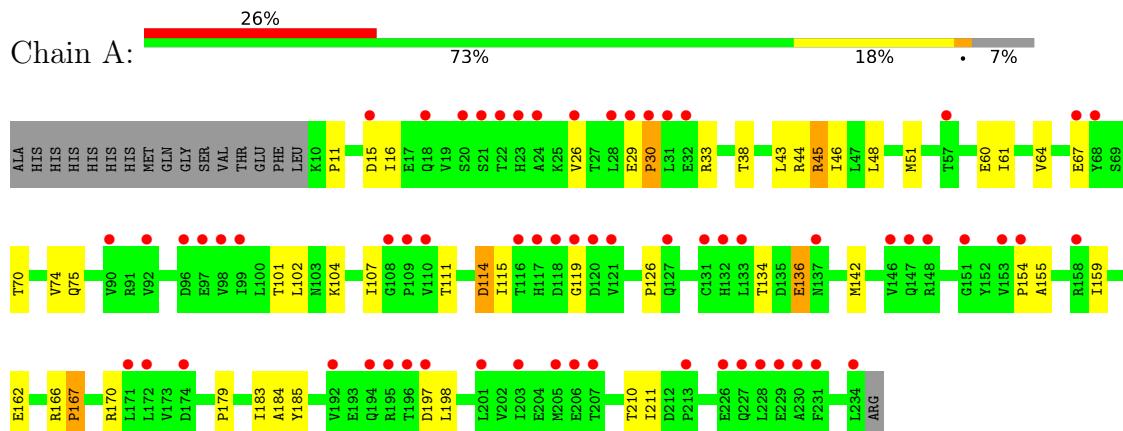
- Molecule 4 is a protein called Transcription termination/antitermination protein NusG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	J	147	Total 1165	C 740	N 201	O 217	S 7	0	0	0
4	K	147	Total 1165	C 740	N 201	O 217	S 7	0	0	0
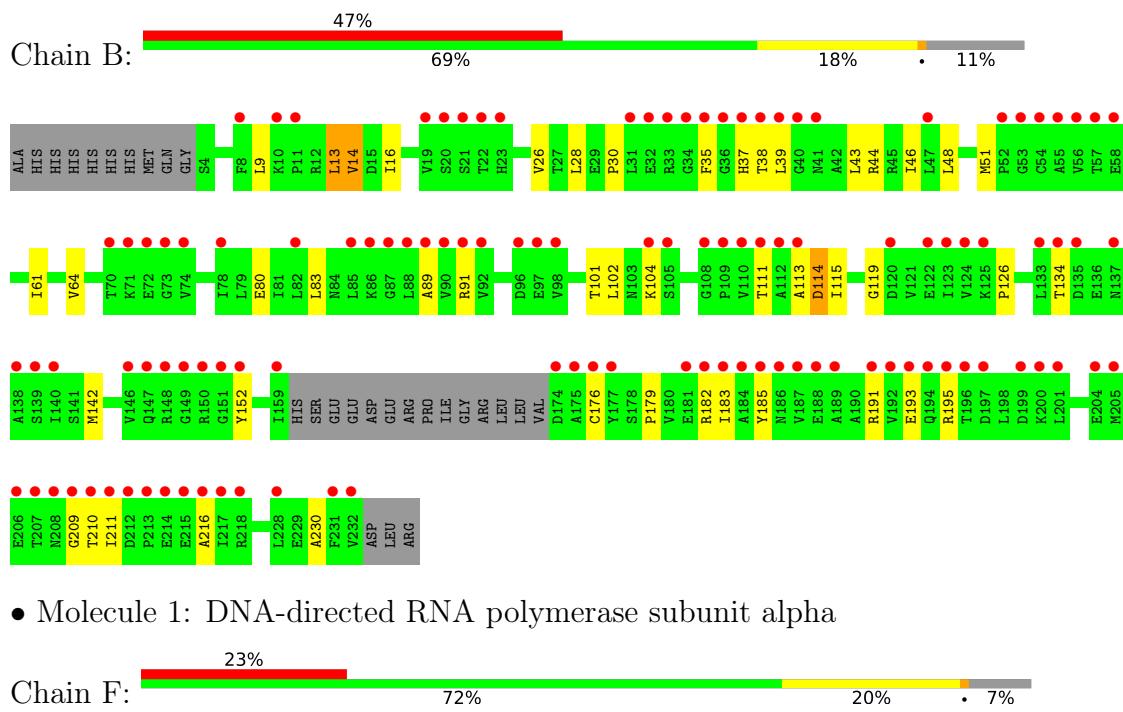
3 Residue-property plots ⓘ

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 alpha

• Molecule 1: DNA-directed RNA polymerase subunit alpha

• Molecule 1: DNA-directed RNA polymerase subunit alpha

• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain G:

• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C:

• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain H:

• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:

29%

63%　　　　　　　　　　　　　　26%　　　·　7%

• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain I:

• Molecule 4: Transcription termination/antitermination protein NusG

Chain J:

• Molecule 4: Transcription termination/antitermination protein NusG

Chain K:
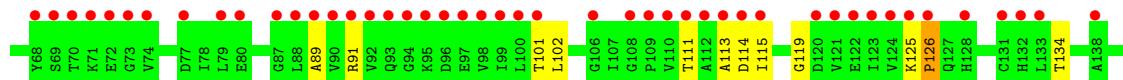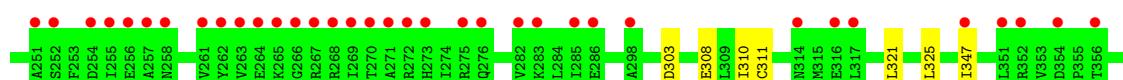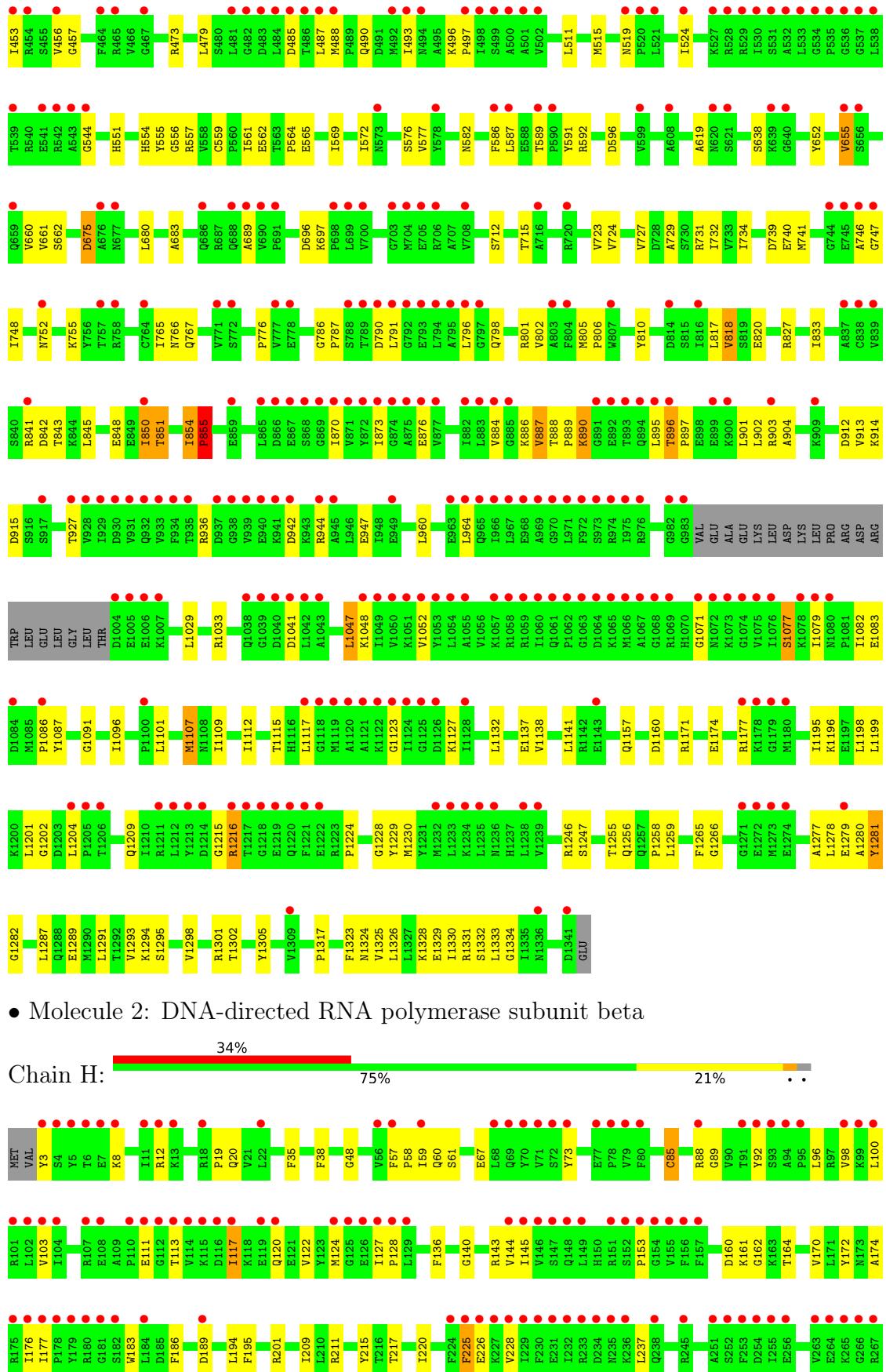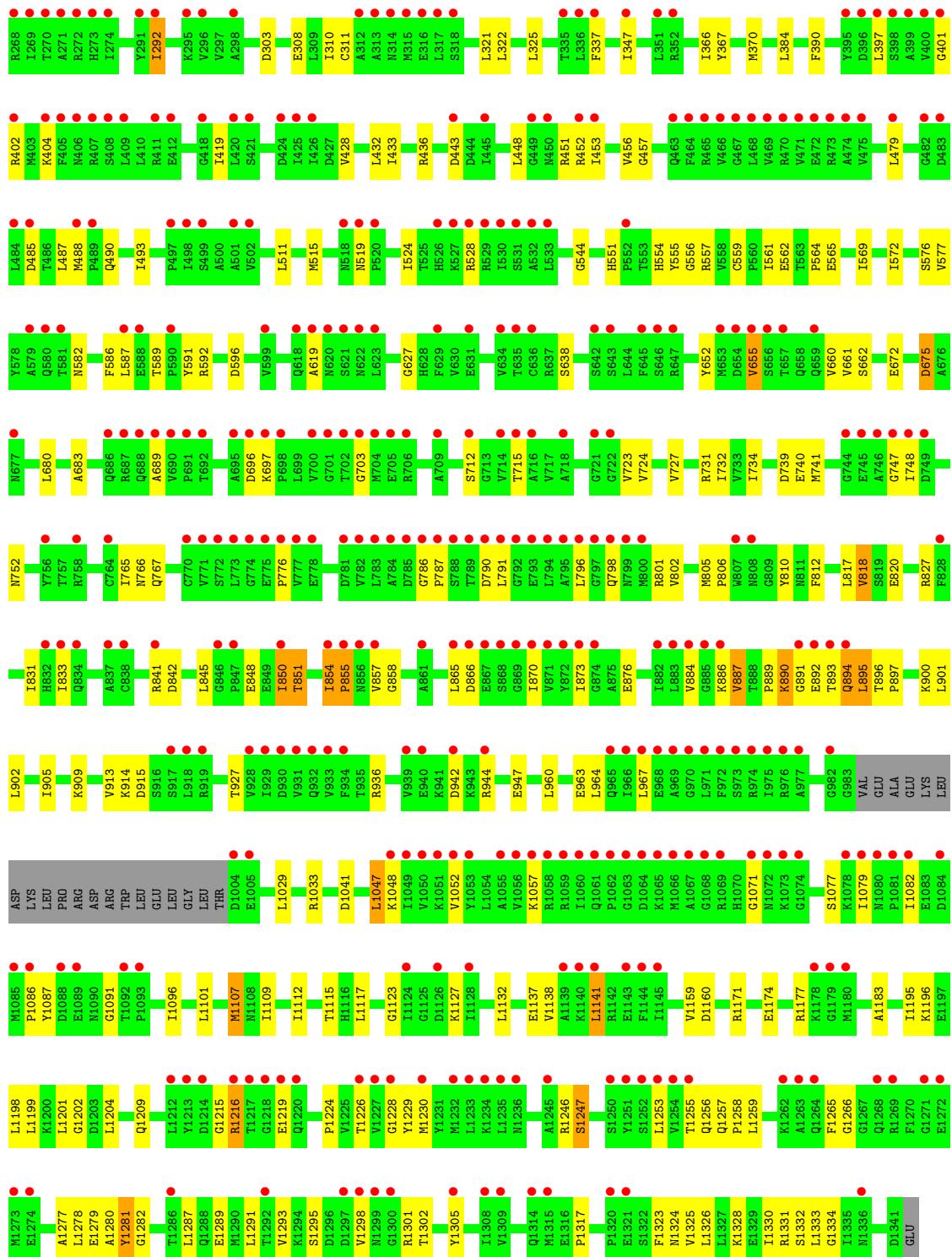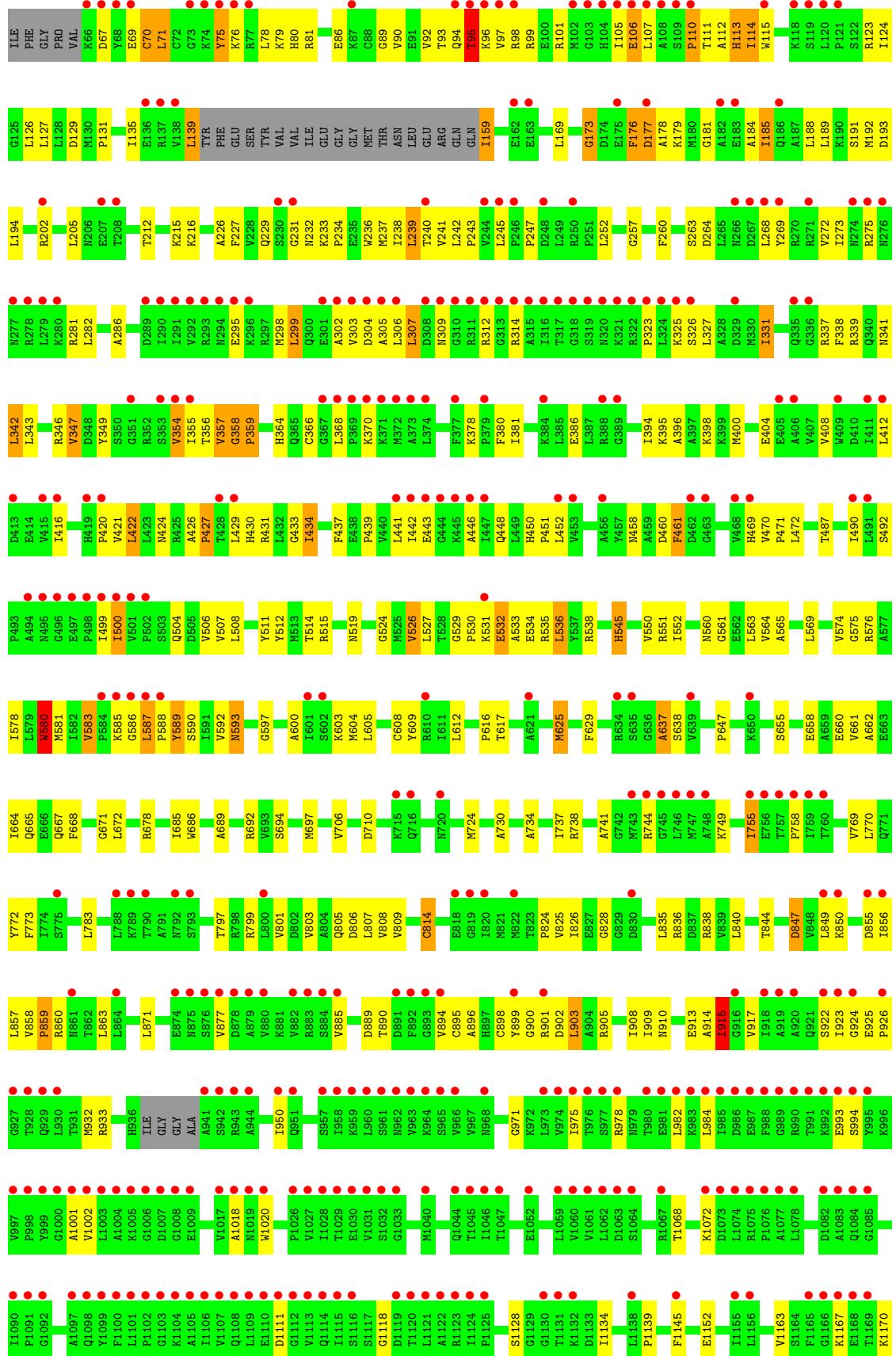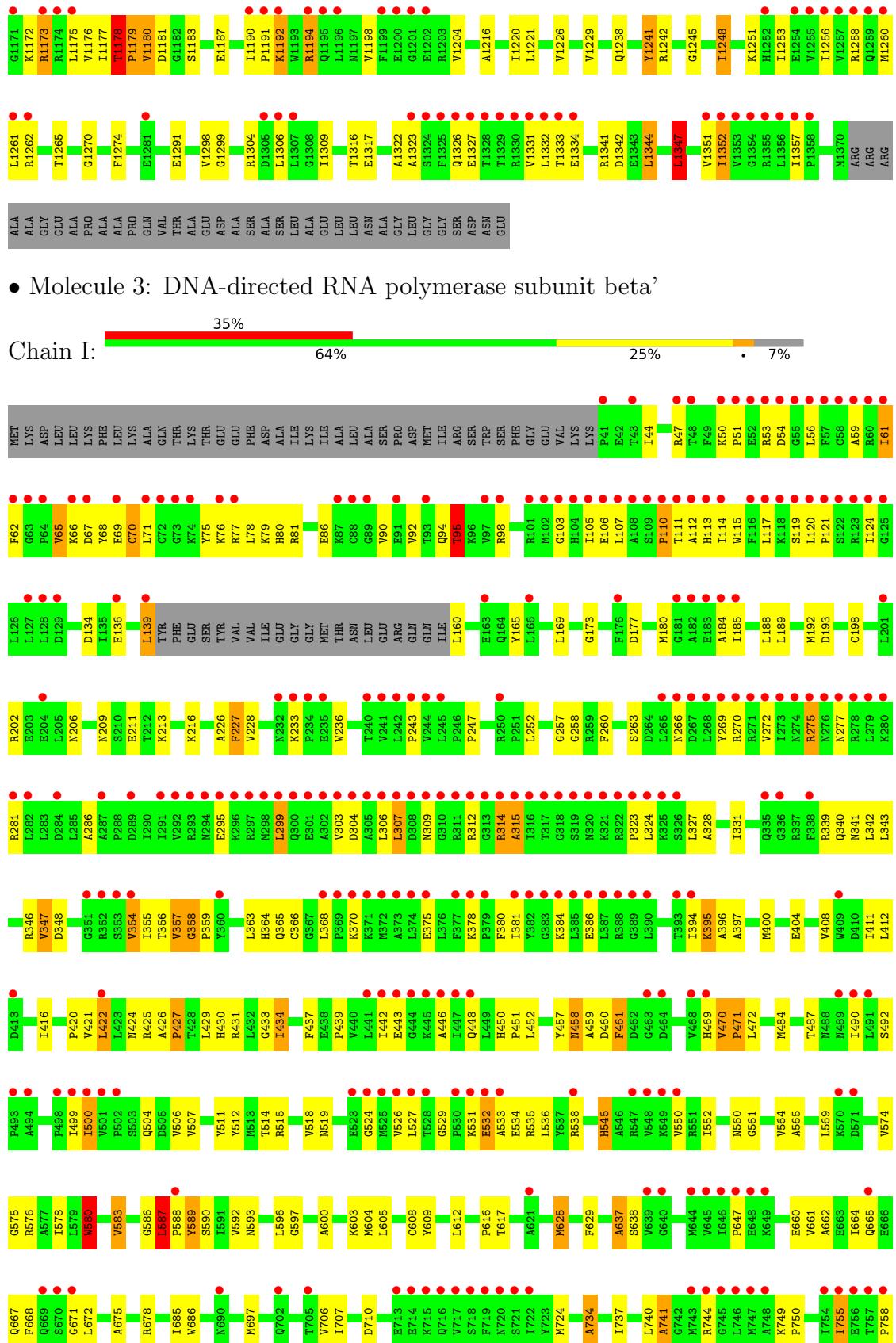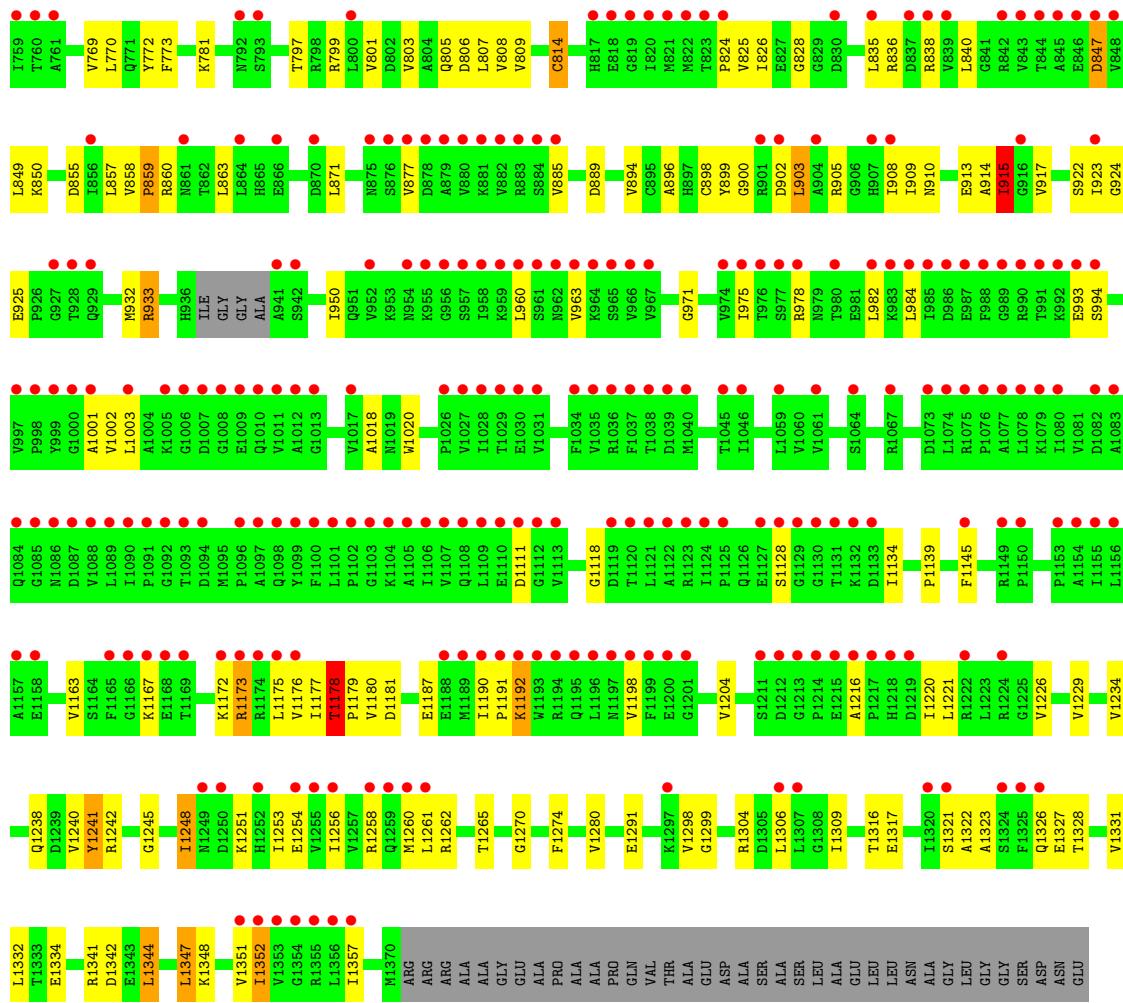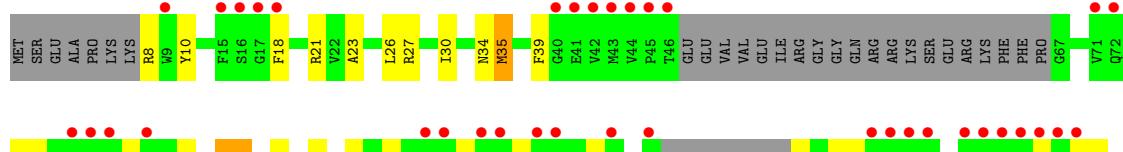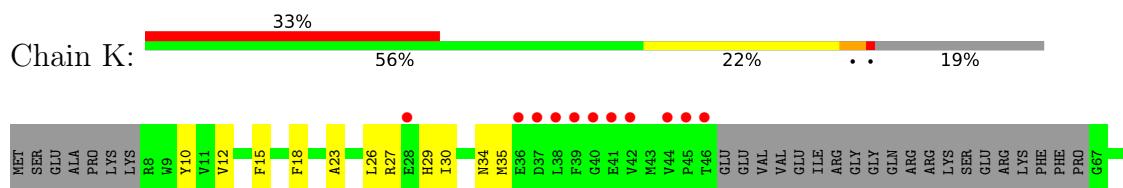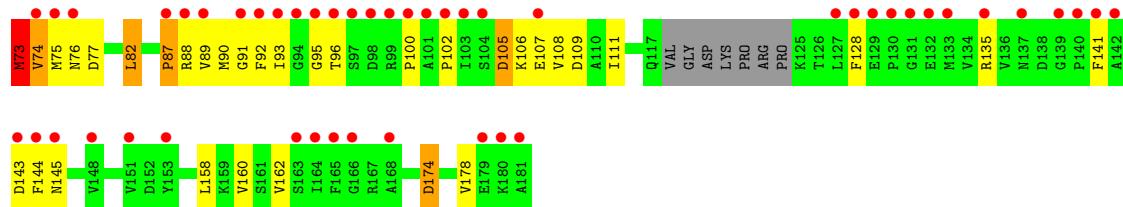
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	248.16Å 313.78Å 162.79Å 90.00° 130.23° 90.00°	Depositor
Resolution (Å)	162.19 – 7.00 49.91 – 7.00	Depositor EDS
% Data completeness (in resolution range)	96.0 (162.19-7.00) 93.3 (49.91-7.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.04 (at 6.68Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.330 , 0.395 0.330 , 0.395	Depositor DCC
R_{free} test set	707 reflections (4.90%)	wwPDB-VP
Wilson B-factor (Å^2)	314.0	Xtriage
Anisotropy	0.296	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 500.0	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.35, < L^2 > = 0.18$	Xtriage
Estimated twinning fraction	0.185 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.78	EDS
Total number of atoms	50147	wwPDB-VP
Average B, all atoms (Å^2)	330.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.98% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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 ⓘ

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	0.45	0/1761	0.72	0/2387
1	B	0.46	0/1676	0.71	0/2271
1	F	0.45	0/1761	0.74	0/2387
1	G	0.47	0/1687	0.70	0/2286
2	C	0.43	0/10569	0.67	0/14258
2	H	0.43	0/10569	0.67	0/14258
3	D	0.44	0/10233	0.76	5/13816 (0.0%)
3	I	0.44	0/10277	0.74	1/13877 (0.0%)
4	J	0.50	0/1188	0.70	0/1603
4	K	0.50	0/1188	0.74	1/1603 (0.1%)
All	All	0.44	0/50909	0.71	7/68746 (0.0%)

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
2	C	0	2
2	H	0	2
3	D	0	1
3	I	0	1
4	K	0	1
All	All	0	7

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	1194	ARG	NE-CZ-NH2	-6.82	116.89	120.30
3	D	239	LEU	CA-CB-CG	5.33	127.56	115.30
3	D	173	GLY	N-CA-C	5.30	126.36	113.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	I	117	LEU	CA-CB-CG	5.29	127.46	115.30
4	K	73	MET	C-N-CA	5.27	134.87	121.70

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	57	PHE	Peptide
2	C	855	PRO	Peptide
3	D	1178	THR	Peptide
2	H	57	PHE	Peptide
2	H	855	PRO	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	1740	0	1767	40	0
1	B	1657	0	1686	40	0
1	F	1740	0	1767	38	0
1	G	1667	0	1693	35	0
2	C	10401	0	10414	216	0
2	H	10401	0	10414	232	0
3	D	10085	0	10303	380	2
3	I	10126	0	10341	316	0
4	J	1165	0	1145	26	0
4	K	1165	0	1145	30	0
All	All	50147	0	50675	1182	2

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 1182 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:1001:ALA:HA	3:I:1020:TRP:HE1	1.12	1.13

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:K:29:HIS:HB3	4:K:82:LEU:HG	1.35	1.06
3:D:226:ALA:HB1	3:D:227:PHE:HA	1.31	1.06
3:D:247:PRO:HB3	3:I:53:ARG:HH12	1.18	1.06
3:D:1001:ALA:HA	3:D:1020:TRP:HE1	1.12	1.06

All (2) 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)
3:D:184:ALA:O	3:D:191:SER:OG[2_957]	2.03	0.17
3:D:1183:SER:OG	3:D:1183:SER:OG[2_957]	2.12	0.08

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	223/242 (92%)	182 (82%)	29 (13%)	12 (5%)	2	19
1	B	211/242 (87%)	175 (83%)	27 (13%)	9 (4%)	2	22
1	F	223/242 (92%)	183 (82%)	28 (13%)	12 (5%)	2	19
1	G	212/242 (88%)	177 (84%)	26 (12%)	9 (4%)	3	22
2	C	1316/1342 (98%)	1108 (84%)	172 (13%)	36 (3%)	5	31
2	H	1316/1342 (98%)	1110 (84%)	168 (13%)	38 (3%)	4	29
3	D	1294/1407 (92%)	1023 (79%)	190 (15%)	81 (6%)	1	17
3	I	1300/1407 (92%)	1024 (79%)	197 (15%)	79 (6%)	1	17
4	J	141/181 (78%)	123 (87%)	15 (11%)	3 (2%)	7	36
4	K	141/181 (78%)	117 (83%)	18 (13%)	6 (4%)	2	22
All	All	6377/6828 (93%)	5222 (82%)	870 (14%)	285 (4%)	2	22

5 of 285 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	15	ASP
1	A	67	GLU
1	A	114	ASP
1	A	155	ALA
1	B	114	ASP

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	193/208 (93%)	191 (99%)	2 (1%)	76	86
1	B	184/208 (88%)	182 (99%)	2 (1%)	73	84
1	F	193/208 (93%)	192 (100%)	1 (0%)	88	93
1	G	185/208 (89%)	183 (99%)	2 (1%)	73	84
2	C	1137/1157 (98%)	1121 (99%)	16 (1%)	67	80
2	H	1137/1157 (98%)	1120 (98%)	17 (2%)	65	80
3	D	1087/1168 (93%)	1048 (96%)	39 (4%)	35	59
3	I	1091/1168 (93%)	1054 (97%)	37 (3%)	37	60
4	J	128/158 (81%)	122 (95%)	6 (5%)	26	51
4	K	128/158 (81%)	121 (94%)	7 (6%)	21	47
All	All	5463/5798 (94%)	5334 (98%)	129 (2%)	49	69

5 of 129 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	I	1347	LEU
4	J	109	ASP
3	D	889	ASP
3	D	847	ASP
4	J	174	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 61 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1249	ASN
3	I	341	ASN
2	H	120	GLN
3	I	309	ASN
3	I	1268	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

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	225/242 (92%)	1.34	64 (28%)	0	2	267, 330, 380, 416	0
1	B	215/242 (88%)	2.69	114 (53%)	0	1	266, 335, 388, 430	0
1	F	225/242 (92%)	1.26	55 (24%)	0	2	248, 320, 383, 403	0
1	G	216/242 (89%)	2.11	91 (42%)	0	1	279, 337, 380, 412	0
2	C	1319/1342 (98%)	1.62	431 (32%)	0	2	232, 313, 384, 510	0
2	H	1319/1342 (98%)	1.75	450 (34%)	0	2	231, 315, 383, 482	0
3	D	1302/1407 (92%)	1.65	415 (31%)	0	2	230, 326, 424, 473	0
3	I	1306/1407 (92%)	2.06	498 (38%)	0	1	231, 332, 431, 492	0
4	J	147/181 (81%)	1.64	49 (33%)	0	2	300, 384, 446, 483	0
4	K	147/181 (81%)	2.13	59 (40%)	0	1	324, 413, 469, 491	0
All	All	6421/6828 (94%)	1.78	2226 (34%)	0	2	230, 324, 419, 510	0

The worst 5 of 2226 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	314	ARG	18.1
3	I	989	GLY	17.0
2	H	788	SER	16.3
3	I	315	ALA	15.7
3	D	315	ALA	15.4

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 ⓘ

There are no monosaccharides 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.