W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Nov 13, 2023 – 08:45 PM JST

PDB ID	:	5Y3D
Title	:	Structural insight into the interaction between RNA polymerase and VPg for norovirus replication
Authors	:	Kim, K.H.; Lee, J.-H.; Seok, J.H.
Deposited on	:	2017-07-28
Resolution	:	3.14 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

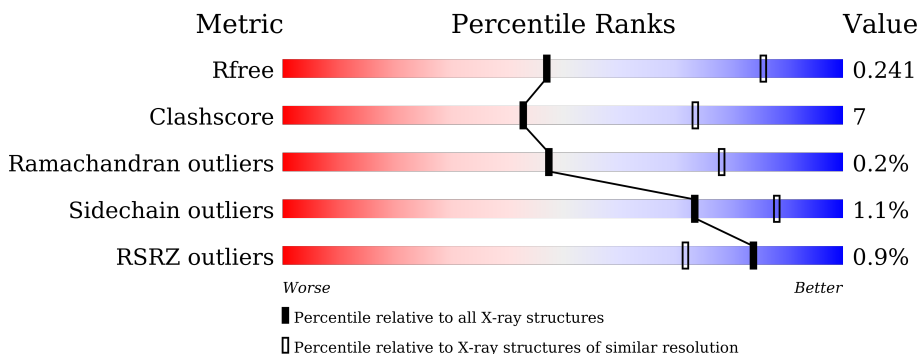|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.14 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		0.241
Clashscore		7
Ramachandran outliers		0.2%
Sidechain outliers		1.1%
RSRZ outliers		0.9%

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	1626 (3.18-3.10)
Clashscore	141614	1735 (3.18-3.10)
Ramachandran outliers	138981	1677 (3.18-3.10)
Sidechain outliers	138945	1677 (3.18-3.10)
RSRZ outliers	127900	1588 (3.18-3.10)

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	517	75% 17% 7%
1	B	517	77% 15% 7%
1	C	517	71% 21% 7%
1	D	517	75% 18% 7%
1	E	517	77% 16% · 7%
1	F	517	78% 15% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	35	89%　　　　　　　6%　6%
2	H	35	80%　　　　　　　11%　9%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 23161 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 RNA-dependent RNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	479	Total	C	N	O	S	0	0	0
			3792	2397	666	705	24			
1	B	479	Total	C	N	O	S	0	0	0
			3797	2402	663	707	25			
1	C	480	Total	C	N	O	S	0	0	0
			3804	2406	669	704	25			
1	D	479	Total	C	N	O	S	0	0	0
			3800	2403	668	705	24			
1	E	483	Total	C	N	O	S	0	0	0
			3815	2412	670	709	24			
1	F	481	Total	C	N	O	S	0	0	0
			3811	2410	669	707	25			

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	510	ALA	-	expression tag	UNP Q80J95
A	511	ALA	-	expression tag	UNP Q80J95
A	512	ALA	-	expression tag	UNP Q80J95
A	513	LEU	-	expression tag	UNP Q80J95
A	514	GLU	-	expression tag	UNP Q80J95
A	515	HIS	-	expression tag	UNP Q80J95
A	516	HIS	-	expression tag	UNP Q80J95
A	517	HIS	-	expression tag	UNP Q80J95
A	518	HIS	-	expression tag	UNP Q80J95
A	519	HIS	-	expression tag	UNP Q80J95
A	520	HIS	-	expression tag	UNP Q80J95
B	510	ALA	-	expression tag	UNP Q80J95
B	511	ALA	-	expression tag	UNP Q80J95
B	512	ALA	-	expression tag	UNP Q80J95
B	513	LEU	-	expression tag	UNP Q80J95
B	514	GLU	-	expression tag	UNP Q80J95
B	515	HIS	-	expression tag	UNP Q80J95

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	516	HIS	-	expression tag	UNP Q80J95
B	517	HIS	-	expression tag	UNP Q80J95
B	518	HIS	-	expression tag	UNP Q80J95
B	519	HIS	-	expression tag	UNP Q80J95
B	520	HIS	-	expression tag	UNP Q80J95
C	510	ALA	-	expression tag	UNP Q80J95
C	511	ALA	-	expression tag	UNP Q80J95
C	512	ALA	-	expression tag	UNP Q80J95
C	513	LEU	-	expression tag	UNP Q80J95
C	514	GLU	-	expression tag	UNP Q80J95
C	515	HIS	-	expression tag	UNP Q80J95
C	516	HIS	-	expression tag	UNP Q80J95
C	517	HIS	-	expression tag	UNP Q80J95
C	518	HIS	-	expression tag	UNP Q80J95
C	519	HIS	-	expression tag	UNP Q80J95
C	520	HIS	-	expression tag	UNP Q80J95
D	510	ALA	-	expression tag	UNP Q80J95
D	511	ALA	-	expression tag	UNP Q80J95
D	512	ALA	-	expression tag	UNP Q80J95
D	513	LEU	-	expression tag	UNP Q80J95
D	514	GLU	-	expression tag	UNP Q80J95
D	515	HIS	-	expression tag	UNP Q80J95
D	516	HIS	-	expression tag	UNP Q80J95
D	517	HIS	-	expression tag	UNP Q80J95
D	518	HIS	-	expression tag	UNP Q80J95
D	519	HIS	-	expression tag	UNP Q80J95
D	520	HIS	-	expression tag	UNP Q80J95
E	510	ALA	-	expression tag	UNP Q80J95
E	511	ALA	-	expression tag	UNP Q80J95
E	512	ALA	-	expression tag	UNP Q80J95
E	513	LEU	-	expression tag	UNP Q80J95
E	514	GLU	-	expression tag	UNP Q80J95
E	515	HIS	-	expression tag	UNP Q80J95
E	516	HIS	-	expression tag	UNP Q80J95
E	517	HIS	-	expression tag	UNP Q80J95
E	518	HIS	-	expression tag	UNP Q80J95
E	519	HIS	-	expression tag	UNP Q80J95
E	520	HIS	-	expression tag	UNP Q80J95
F	510	ALA	-	expression tag	UNP Q80J95
F	511	ALA	-	expression tag	UNP Q80J95
F	512	ALA	-	expression tag	UNP Q80J95
F	513	LEU	-	expression tag	UNP Q80J95

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	514	GLU	-	expression tag	UNP Q80J95
F	515	HIS	-	expression tag	UNP Q80J95
F	516	HIS	-	expression tag	UNP Q80J95
F	517	HIS	-	expression tag	UNP Q80J95
F	518	HIS	-	expression tag	UNP Q80J95
F	519	HIS	-	expression tag	UNP Q80J95
F	520	HIS	-	expression tag	UNP Q80J95

• Molecule 2 is a protein called viral protein genome-linked (VPg).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	G	33	Total 164	C 98	N 33	O 33	0	0	0
2	H	32	Total 160	C 96	N 32	O 32	0	0	0

• Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total 2	O 2	0	0
3	B	2	Total 2	O 2	0	0
3	C	2	Total 2	O 2	0	0
3	D	4	Total 4	O 4	0	0
3	E	6	Total 6	O 6	0	0
3	F	2	Total 2	O 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.

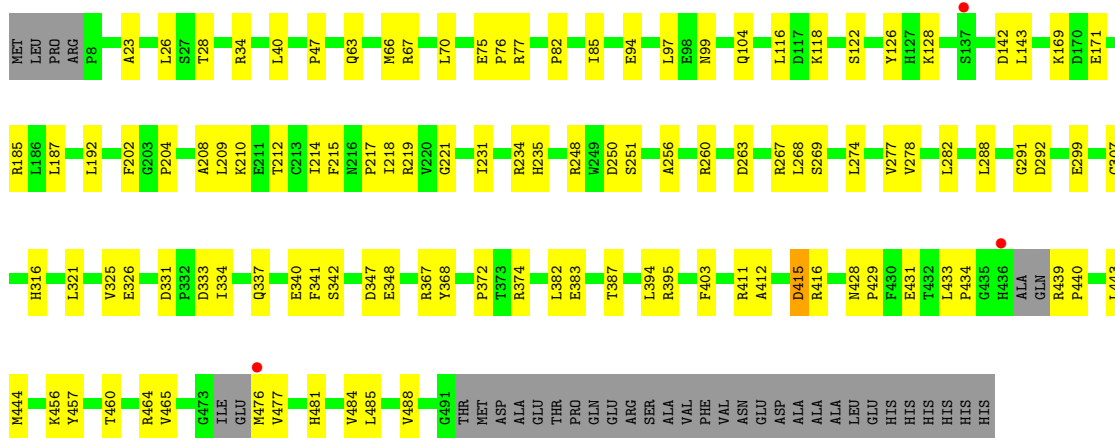- Molecule 1: RNA-dependent RNA polymerase
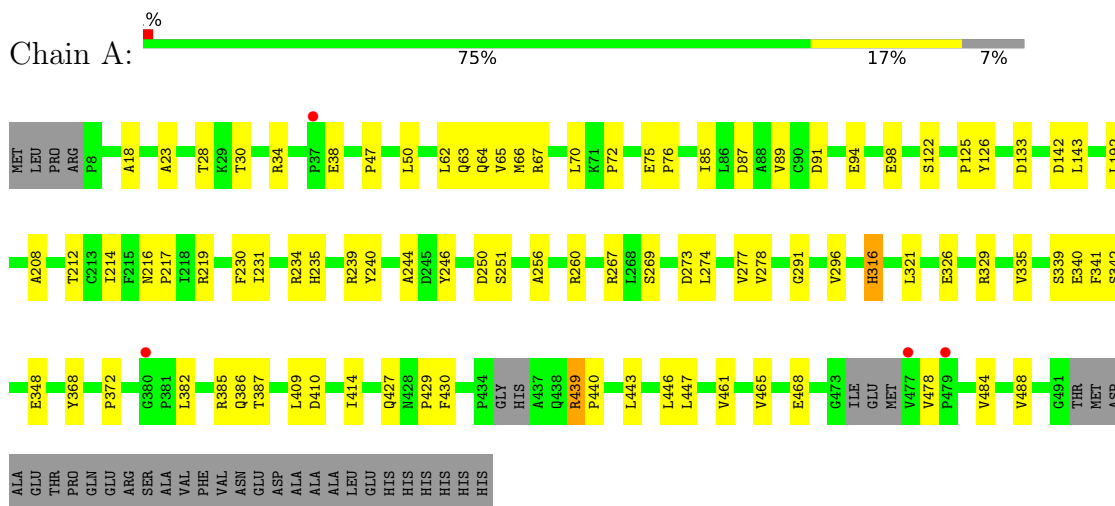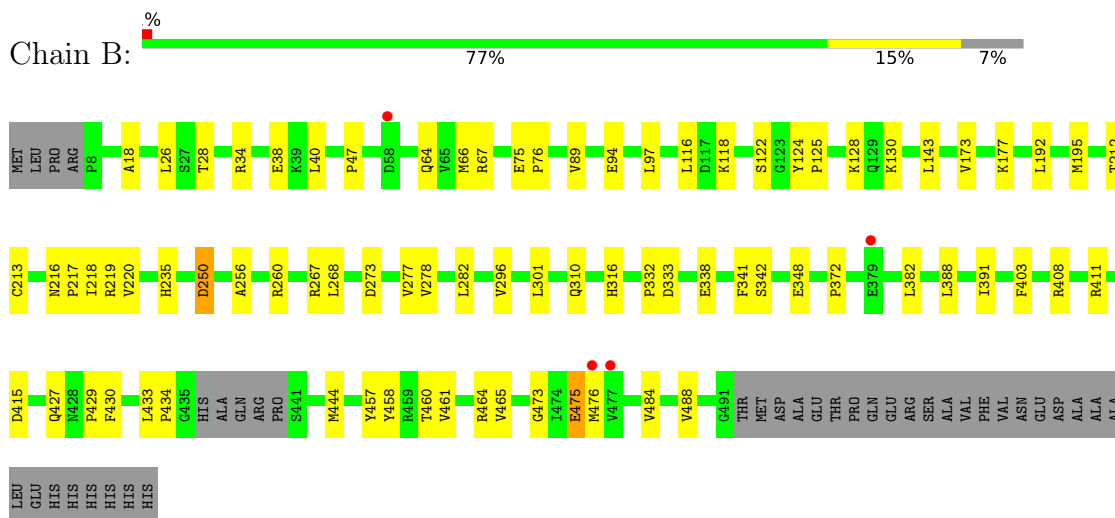
Chain A:

- Molecule 1: RNA-dependent RNA polymerase

Chain B:

- Molecule 1: RNA-dependent RNA polymerase

Chain C:

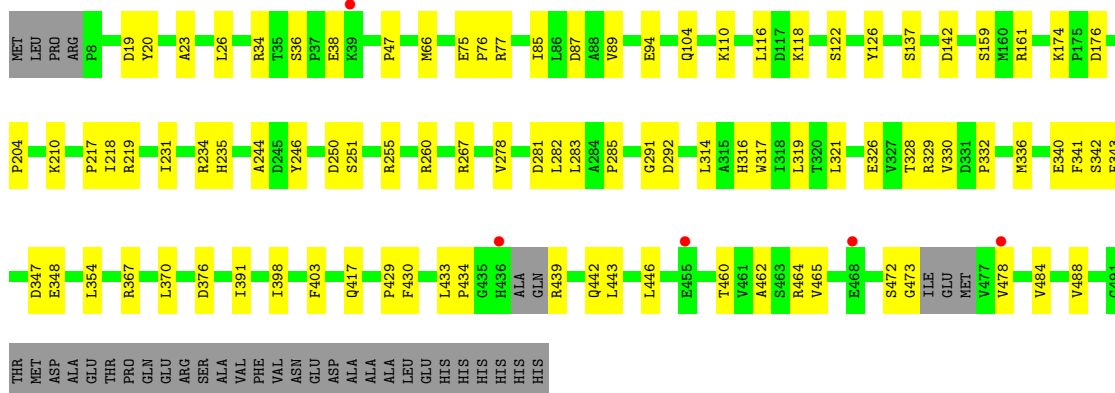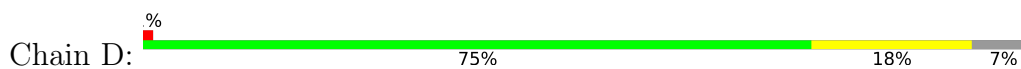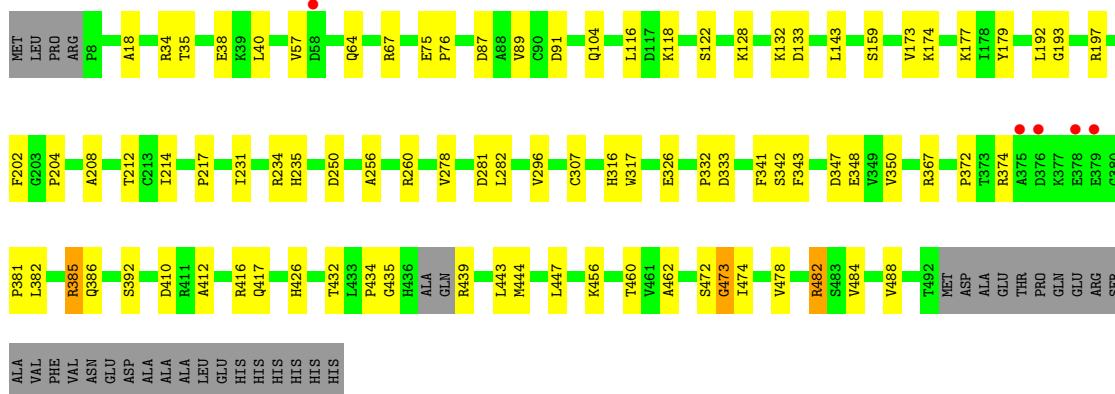• Molecule 1: RNA-dependent RNA polymerase

Chain D:

• Molecule 1: RNA-dependent RNA polymerase
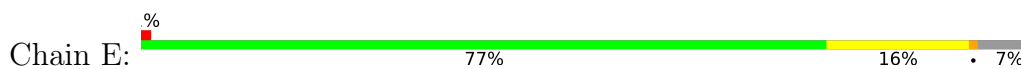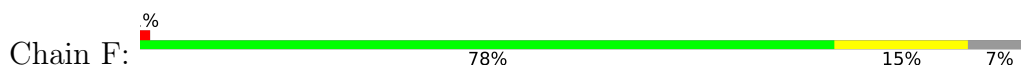
Chain E:

• Molecule 1: RNA-dependent RNA polymerase

Chain F:

MET LEU PRO ARG P8 Y20 L26 R34 T35 S36 P37 E38 P47 M66 E75 P76 R77 D87 A88 V89 Q104 W107 K128 L143 R161 V173 K174 K177 L192 R200 P204 T212 C213 I214 F215 N216 P217 I218 R219 I231

R234 D250 S251 R260 D273 L274 V277 V278 P285 S303 T308 H316 L319 E326 P332 D333 F341 E348 V359 R367 Y368 G369 L370 L371 P372 E378 E379 G380 P381 L382 F403 L409 D410 S413 I414 D415 L418 N428 P429

F430 E431 T432 L433 P434 G435 H436 ALA GLN R439 P440 L443 M444 A445 L446 V461 V465 E468 S472 G473 ILE GLU M476 G491 T492 MET ASP ALA GLU THR PRO GLN GLU ARG SER ALA VAL PHE VAL ASN GLU ASP ALA ALA LEU GLU HIS HIS HIS HIS HIS HIS

• Molecule 2: viral protein genome-linked (VPg)
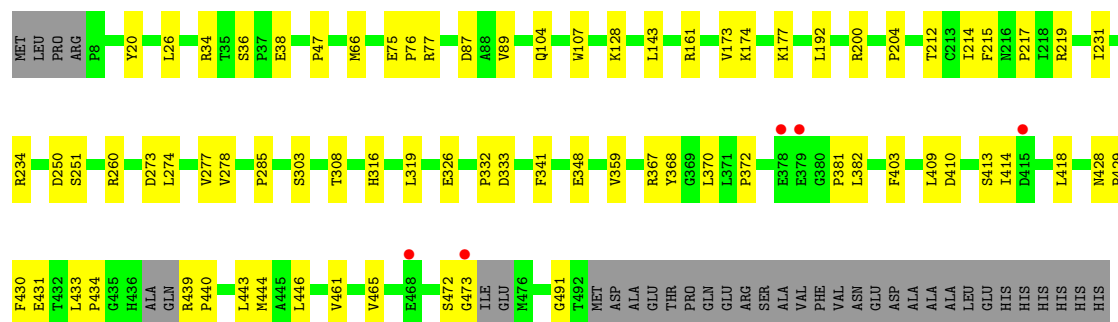
Chain G:

89%　　　　　　6%　　6%

X22 X37 UNK UNK X40 X52 X56
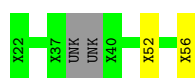
• Molecule 2: viral protein genome-linked (VPg)

Chain H:

80%　　　　　　11%　　　9%

X22 X23 X24 X25 X36 UNK UNK UNK X40 X52 X56
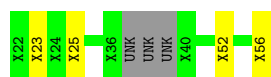
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	109.17Å 159.73Å 121.70Å 90.00° 97.24° 90.00°	Depositor
Resolution (Å)	47.23 – 3.14 47.24 – 3.14	Depositor EDS
% Data completeness (in resolution range)	96.3 (47.23-3.14) 96.3 (47.24-3.14)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.55 (at 3.12Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.203 , 0.241 0.203 , 0.241	Depositor DCC
R_{free} test set	3518 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	54.5	Xtriage
Anisotropy	0.486	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 46.8	EDS
L-test for twinning[2]	$< |L| > = 0.41, < L^2 > = 0.23$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	23161	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.28% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.27	0/3883	0.44	1/5258 (0.0%)
1	B	0.27	0/3889	0.43	0/5267
1	C	0.27	0/3897	0.44	0/5277
1	D	0.27	0/3893	0.43	0/5272
1	E	0.27	0/3908	0.43	0/5295
1	F	0.27	0/3904	0.44	0/5287
All	All	0.27	0/23374	0.43	1/31656 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	125	PRO	N-CA-CB	5.52	109.92	103.30

There are no chirality outliers.

There are no planarity outliers.

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	3792	0	3737	56	0
1	B	3797	0	3753	47	0
1	C	3804	0	3758	73	0
1	D	3800	0	3753	55	1
1	E	3815	0	3757	50	0
1	F	3811	0	3763	43	1

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	G	164	0	37	1	0
2	H	160	0	37	2	0
3	A	2	0	0	0	0
3	B	2	0	0	1	0
3	C	2	0	0	1	0
3	D	4	0	0	2	0
3	E	6	0	0	2	0
3	F	2	0	0	0	0
All	All	23161	0	22595	313	1

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 7.

The worst 5 of 313 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:B:235:HIS:HE2	1:B:342:SER:HG	1.18	0.92
1:E:235:HIS:HE2	1:E:342:SER:HG	1.11	0.90
1:D:235:HIS:HE2	1:D:342:SER:HG	1.19	0.84
1:C:235:HIS:HE2	1:C:342:SER:HG	1.27	0.79
1:C:219:ARG:HD3	1:C:231:ILE:HD13	1.70	0.73

All (1) 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)
1:D:292:ASP:OD1	1:F:20:TYR:OH[1_455]	2.16	0.04

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	473/517 (92%)	463 (98%)	9 (2%)	1 (0%)	47	78
1	B	475/517 (92%)	463 (98%)	11 (2%)	1 (0%)	47	78
1	C	474/517 (92%)	461 (97%)	12 (2%)	1 (0%)	47	78
1	D	473/517 (92%)	462 (98%)	11 (2%)	0	100	100
1	E	479/517 (93%)	465 (97%)	11 (2%)	3 (1%)	25	59
1	F	475/517 (92%)	463 (98%)	11 (2%)	1 (0%)	47	78
All	All	2849/3102 (92%)	2777 (98%)	65 (2%)	7 (0%)	47	78

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	473	GLY
1	E	473	GLY
1	F	491	GLY
1	A	440	PRO
1	E	474	ILE

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	405/440 (92%)	400 (99%)	5 (1%)	71	87
1	B	408/440 (93%)	404 (99%)	4 (1%)	76	89
1	C	408/440 (93%)	403 (99%)	5 (1%)	71	87
1	D	408/440 (93%)	405 (99%)	3 (1%)	84	93
1	E	407/440 (92%)	400 (98%)	7 (2%)	60	82
1	F	409/440 (93%)	406 (99%)	3 (1%)	84	93
All	All	2445/2640 (93%)	2418 (99%)	27 (1%)	73	88

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	122	SER

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type
1	E	122	SER
1	F	250	ASP
1	D	376	ASP
1	E	250	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	471	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 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(Å²)	Q<0.9
1	A	479/517 (92%)	-0.27	4 (0%)	86	74	21, 47, 91, 133	0
1	B	479/517 (92%)	-0.26	4 (0%)	86	74	23, 46, 79, 151	0
1	C	480/517 (92%)	-0.13	3 (0%)	89	80	26, 54, 93, 137	0
1	D	479/517 (92%)	-0.21	5 (1%)	82	70	23, 49, 92, 162	0
1	E	483/517 (93%)	-0.21	5 (1%)	82	70	24, 52, 100, 160	0
1	F	481/517 (93%)	-0.12	5 (1%)	82	70	24, 54, 95, 122	0
2	G	0/35	-	-			-	-
2	H	0/35	-	-			-	-
All	All	2881/3172 (90%)	-0.20	26 (0%)	84	72	21, 51, 94, 162	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	476	MET	4.4
1	B	476	MET	3.4
1	B	58	ASP	3.1
1	B	477	VAL	2.7
1	A	477	VAL	2.6

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.