



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2020 – 03:33 PM BST

PDB ID : 5GQ1
Title : Crystal structure of 2C helicase from enterovirus 71 (EV71)
Authors : Guan, H.X.; Tian, J.; Qin, B.; Wojdyla, J.; Wang, M.T.; Cui, S.
Deposited on : 2016-08-05
Resolution : 2.49 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.13.1
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) : 2.13.1

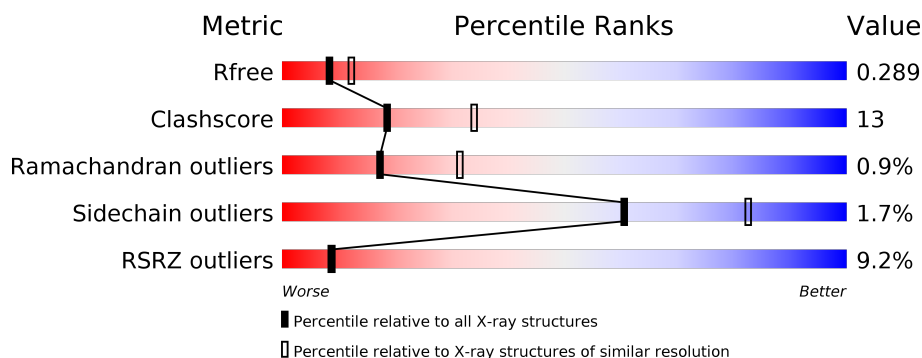
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.49 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 ≥ 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	214	<div> <div>3%</div> <div> <div></div> <div>69%</div> <div>25%</div> <div>• •</div> </div> </div>
1	B	214	<div> <div>4%</div> <div> <div></div> <div>67%</div> <div>25%</div> <div>7%</div> </div> </div>
1	C	214	<div> <div>9%</div> <div> <div></div> <div>59%</div> <div>30%</div> <div>• 8%</div> </div> </div>
1	D	214	<div> <div>4%</div> <div> <div></div> <div>58%</div> <div>34%</div> <div>8%</div> </div> </div>
1	E	214	<div> <div>12%</div> <div> <div></div> <div>59%</div> <div>27%</div> <div>14%</div> </div> </div>
1	F	214	<div> <div>17%</div> <div> <div></div> <div>58%</div> <div>22%</div> <div>19%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ZN	B	401	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17865 atoms, of which 8891 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.

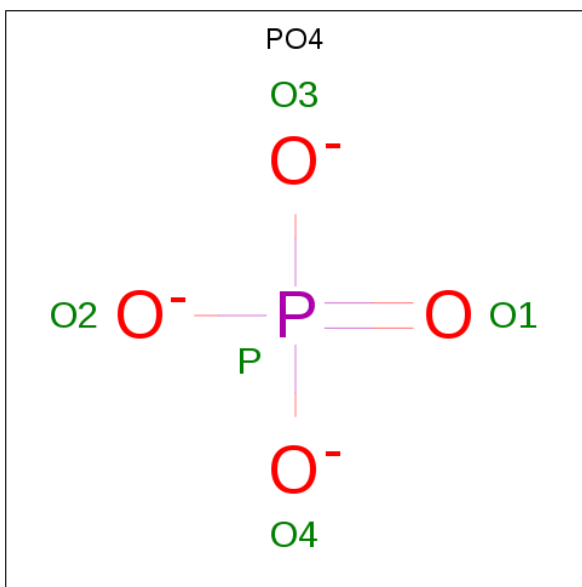
- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	207	Total	C	H	N	O	S	0	0	0
			3166	989	1575	280	310	12			
1	B	198	Total	C	H	N	O	S	0	0	0
			3043	951	1518	266	296	12			
1	C	197	Total	C	H	N	O	S	0	0	0
			3032	952	1509	268	291	12			
1	D	197	Total	C	H	N	O	S	0	0	0
			3025	947	1509	266	291	12			
1	E	185	Total	C	H	N	O	S	0	0	0
			2879	901	1441	254	272	11			
1	F	173	Total	C	H	N	O	S	0	0	0
			2669	833	1339	229	256	12			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total	Zn	0	0
			1	1		
2	E	1	Total	Zn	0	0
			1	1		
2	B	1	Total	Zn	0	0
			1	1		
2	C	1	Total	Zn	0	0
			1	1		
2	A	1	Total	Zn	0	0
			1	1		
2	F	1	Total	Zn	0	0
			1	1		

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	O	P	0	0
			5	4	1		
3	D	1	Total	O	P	0	0
			5	4	1		

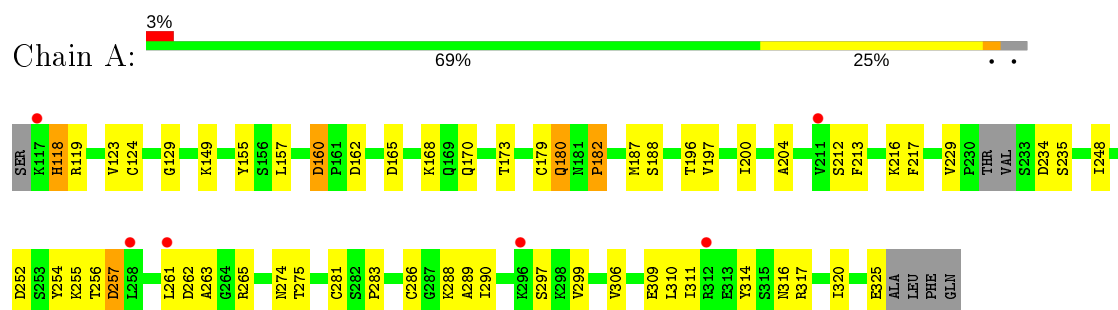
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	11	Total	O	0	0
			11	11		
4	B	3	Total	O	0	0
			3	3		
4	C	6	Total	O	0	0
			6	6		
4	D	9	Total	O	0	0
			9	9		
4	E	3	Total	O	0	0
			3	3		
4	F	3	Total	O	0	0
			3	3		

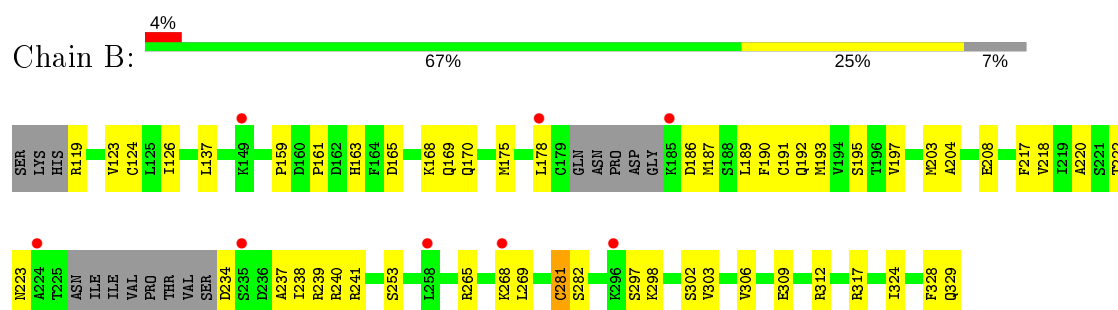
3 Residue-property plots [i](#)

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.

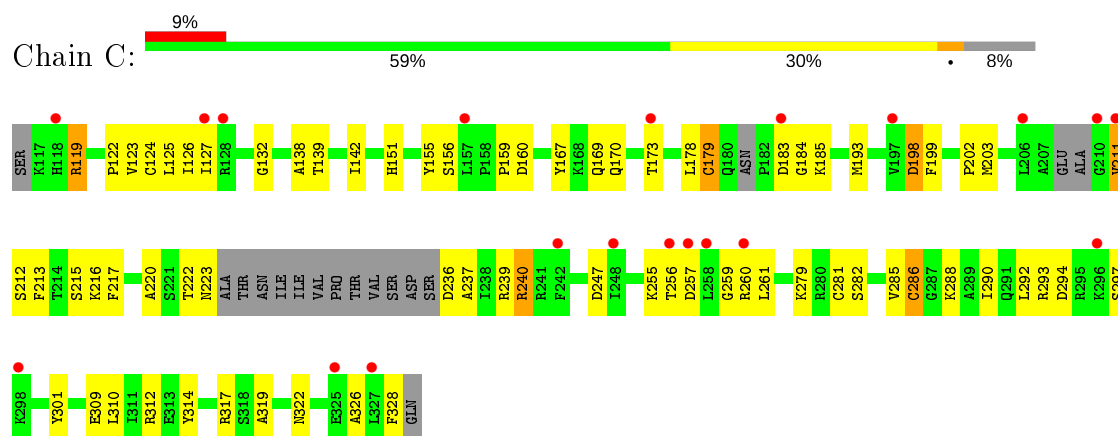
• Molecule 1: Genome polyprotein



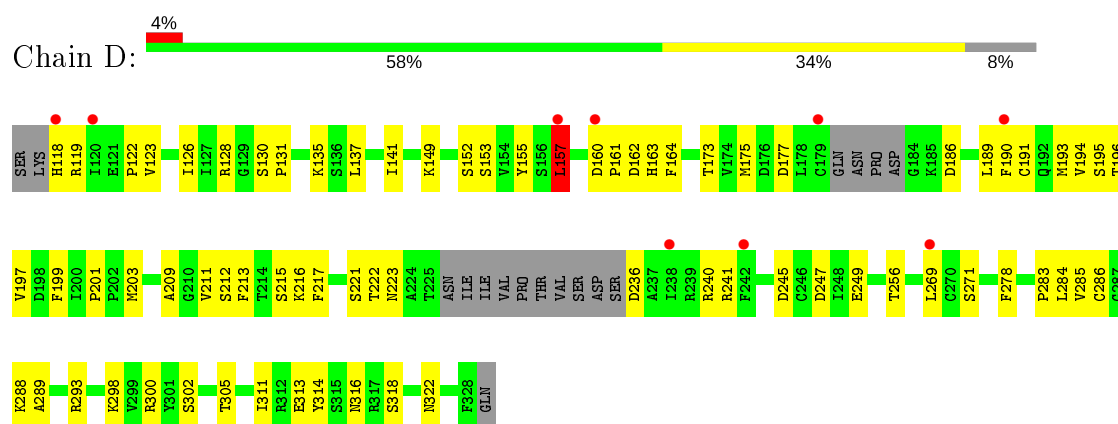
• Molecule 1: Genome polyprotein



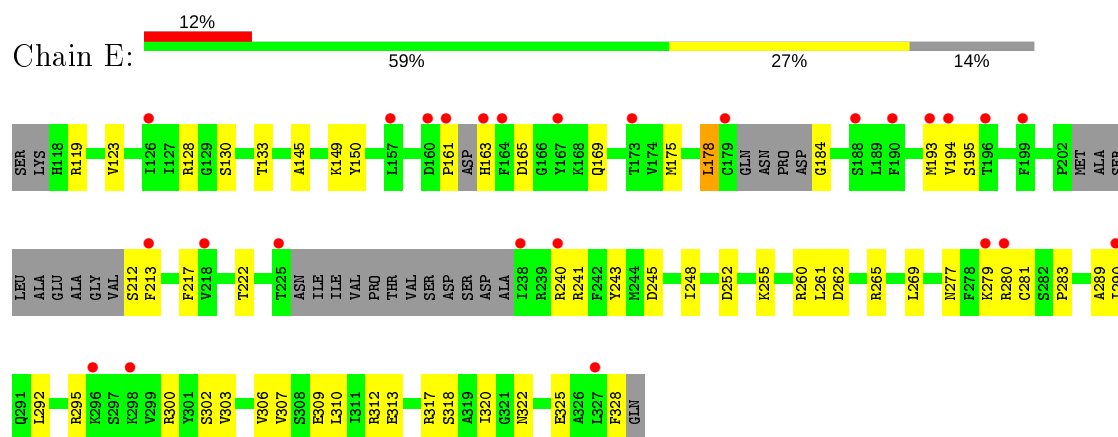
• Molecule 1: Genome polyprotein



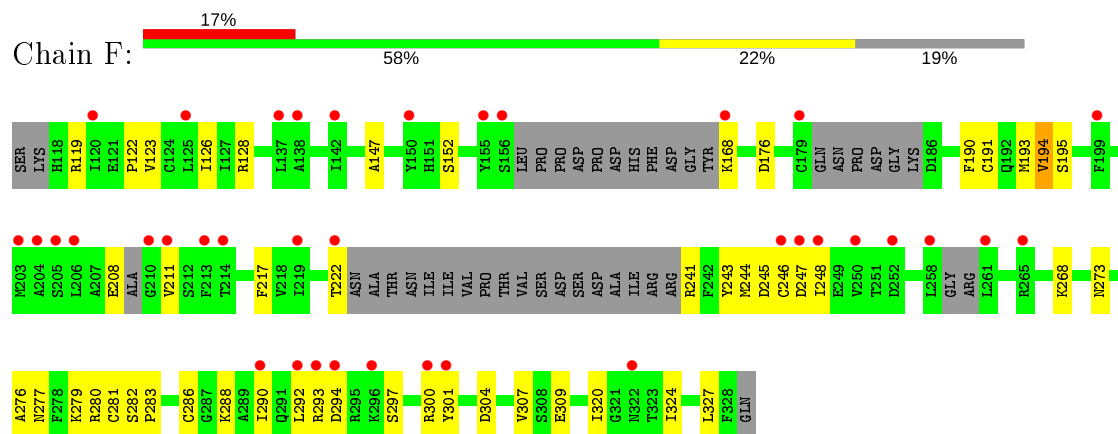
• Molecule 1: Genome polyprotein



• Molecule 1: Genome polyprotein



• Molecule 1: Genome polyprotein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.11Å 159.65Å 75.61Å 90.00° 90.16° 90.00°	Depositor
Resolution (Å)	43.63 – 2.49 43.63 – 2.49	Depositor EDS
% Data completeness (in resolution range)	98.8 (43.63-2.49) 99.0 (43.63-2.49)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.11 (at 2.48Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.229 , 0.286 0.233 , 0.289	Depositor DCC
R_{free} test set	2149 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	62.3	Xtriage
Anisotropy	0.393	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 41.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.41$, $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	0.248 for h,-k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	17865	wwPDB-VP
Average B, all atoms (Å ²)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.74% 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: ZN, PO4

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	0.62	1/1616 (0.1%)	0.72	0/2184
1	B	0.61	1/1547 (0.1%)	0.72	0/2087
1	C	0.70	2/1546 (0.1%)	0.71	0/2082
1	D	0.62	0/1539	0.78	2/2076 (0.1%)
1	E	0.43	0/1459	0.61	0/1964
1	F	0.40	0/1344	0.64	0/1806
All	All	0.58	4/9051 (0.0%)	0.70	2/12199 (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
1	C	0	2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	286	CYS	CB-SG	-11.78	1.62	1.82
1	B	281	CYS	CB-SG	5.97	1.92	1.82
1	C	281	CYS	CB-SG	-5.54	1.72	1.81
1	A	124	CYS	CB-SG	-5.34	1.73	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	160	ASP	CB-CG-OD2	5.17	122.95	118.30
1	D	157	LEU	C-N-CD	5.06	139.03	128.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	211	VAL	Peptide
1	C	285	VAL	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	1591	1575	1586	42	2
1	B	1525	1518	1522	43	2
1	C	1523	1509	1523	42	6
1	D	1516	1509	1514	46	4
1	E	1438	1441	1443	39	2
1	F	1330	1339	1339	28	0
2	A	1	0	0	0	0
2	B	1	0	0	2	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	B	5	0	0	0	0
3	D	5	0	0	0	0
4	A	11	0	0	4	0
4	B	3	0	0	1	0
4	C	6	0	0	6	0
4	D	9	0	0	3	0
4	E	3	0	0	2	0
4	F	3	0	0	1	0
All	All	8974	8891	8927	236	8

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 236 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:C:203:MET:SD	4:C:501:HOH:O	2.08	1.09
1:B:208:GLU:OE1	1:E:240:ARG:NH2	1.94	1.00
1:C:169:GLN:NE2	1:C:212:SER:O	1.97	0.98
1:E:128:ARG:NH1	1:E:245:ASP:OD1	1.98	0.95
1:B:309:GLU:OE1	1:B:312:ARG:NH2	2.01	0.94

The worst 5 of 8 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:C:288:LYS:HZ1	1:D:201:PRO:O[1_556]	1.30	0.30
1:C:288:LYS:NZ	1:D:201:PRO:O[1_556]	2.03	0.17
1:B:240:ARG:HE	1:C:160:ASP:O[1_655]	1.44	0.16
1:C:326:ALA:O	1:E:280:ARG:NH2[2_648]	2.09	0.11
1:C:328:PHE:O	1:E:279:LYS:HZ1[2_648]	1.50	0.10

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	203/214 (95%)	186 (92%)	14 (7%)	3 (2%)	10	18
1	B	192/214 (90%)	179 (93%)	12 (6%)	1 (0%)	29	48
1	C	189/214 (88%)	170 (90%)	16 (8%)	3 (2%)	9	17
1	D	191/214 (89%)	175 (92%)	14 (7%)	2 (1%)	15	28
1	E	175/214 (82%)	164 (94%)	11 (6%)	0	100	100
1	F	161/214 (75%)	144 (89%)	16 (10%)	1 (1%)	25	43
All	All	1111/1284 (86%)	1018 (92%)	83 (8%)	10 (1%)	17	31

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	180	GLN
1	A	182	PRO
1	A	257	ASP
1	B	239	ARG
1	C	184	GLY

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	180/186 (97%)	177 (98%)	3 (2%)	60	82
1	B	171/186 (92%)	171 (100%)	0	100	100
1	C	171/186 (92%)	167 (98%)	4 (2%)	50	76
1	D	169/186 (91%)	165 (98%)	4 (2%)	49	74
1	E	162/186 (87%)	160 (99%)	2 (1%)	71	88
1	F	151/186 (81%)	147 (97%)	4 (3%)	46	72
All	All	1004/1116 (90%)	987 (98%)	17 (2%)	60	82

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	157	LEU
1	D	177	ASP
1	F	119	ARG
1	C	240	ARG
1	F	176	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	163	HIS
1	E	322	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 ⓘ

Of 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	PO4	D	402	-	4,4,4	0.76	0	6,6,6	0.71	0
3	PO4	B	402	-	4,4,4	1.05	0	6,6,6	0.34	0

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 ⓘ

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	207/214 (96%)	0.42	6 (2%) 51 55	47, 65, 90, 156	0
1	B	198/214 (92%)	0.52	8 (4%) 38 41	42, 66, 97, 116	0
1	C	197/214 (92%)	0.74	20 (10%) 6 6	41, 73, 99, 106	0
1	D	197/214 (92%)	0.50	9 (4%) 32 34	30, 71, 94, 113	0
1	E	185/214 (86%)	0.84	26 (14%) 2 2	65, 88, 116, 126	0
1	F	173/214 (80%)	1.33	37 (21%) 0 0	64, 102, 120, 125	0
All	All	1157/1284 (90%)	0.71	106 (9%) 9 9	30, 75, 113, 156	0

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	210	GLY	15.6
1	F	292	LEU	6.9
1	F	258	LEU	6.2
1	F	211	VAL	6.2
1	F	125	LEU	5.8

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 [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
3	PO4	B	402	5/5	0.86	0.34	86,110,120,230	0
2	ZN	F	401	1/1	0.86	0.29	118,118,118,118	0
2	ZN	D	401	1/1	0.91	0.33	76,76,76,76	0
3	PO4	D	402	5/5	0.93	0.11	79,87,96,101	0
2	ZN	E	401	1/1	0.97	0.31	107,107,107,107	0
2	ZN	C	401	1/1	0.97	0.33	75,75,75,75	0
2	ZN	A	401	1/1	0.98	0.29	68,68,68,68	0
2	ZN	B	401	1/1	0.98	0.30	79,79,79,79	0

6.5 Other polymers [i](#)

There are no such residues in this entry.