wwPDB X-ray Structure Validation Summary Report (i)

Jan 23, 2021 – 02:50 PM EST

PDB ID	:	1K47
Title	:	Crystal Structure of the Streptococcus pneumoniae Phosphomevalonate Kinase (PMK)
Authors	:	Romanowski, M.J.; Bonanno, J.B.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2001-10-05
Resolution	:	2.42 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

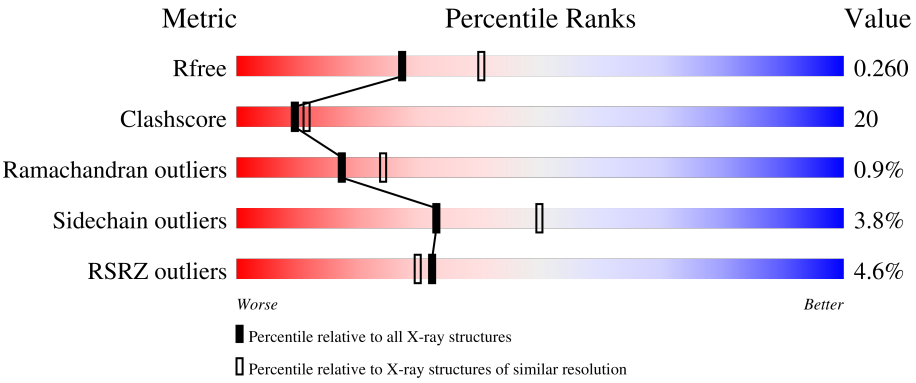
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.42 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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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	337	5% 64% 31%
1	B	337	4% 65% 27%
1	C	337	5% 66% 27%
1	D	337	5% 63% 30%
1	E	337	3% 61% 33%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	337	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15779 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 phosphomevalonate kinase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	322	Total	C	N	O	S	Se	0	0	0
			2500	1595	413	480	4	8			
1	B	322	Total	C	N	O	S	Se	0	0	0
			2500	1595	413	480	4	8			
1	C	322	Total	C	N	O	S	Se	0	0	0
			2500	1595	413	480	4	8			
1	D	322	Total	C	N	O	S	Se	0	0	0
			2500	1595	413	480	4	8			
1	E	322	Total	C	N	O	S	Se	0	0	0
			2500	1595	413	480	4	8			
1	F	322	Total	C	N	O	S	Se	0	0	0
			2500	1595	413	480	4	8			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1001	HIS	-	cloning artifact	UNP Q8DR49
A	1002	MSE	-	cloning artifact	UNP Q8DR49
A	1	MSE	MET	modified residue	UNP Q8DR49
A	34	MSE	MET	modified residue	UNP Q8DR49
A	50	MSE	MET	modified residue	UNP Q8DR49
A	73	MSE	MET	modified residue	UNP Q8DR49
A	95	MSE	MET	modified residue	UNP Q8DR49
A	148	MSE	MET	modified residue	UNP Q8DR49
A	216	MSE	MET	modified residue	UNP Q8DR49
B	1001	HIS	-	cloning artifact	UNP Q8DR49
B	1002	MSE	-	cloning artifact	UNP Q8DR49
B	1	MSE	MET	modified residue	UNP Q8DR49
B	34	MSE	MET	modified residue	UNP Q8DR49
B	50	MSE	MET	modified residue	UNP Q8DR49
B	73	MSE	MET	modified residue	UNP Q8DR49
B	95	MSE	MET	modified residue	UNP Q8DR49
B	148	MSE	MET	modified residue	UNP Q8DR49

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	216	MSE	MET	modified residue	UNP Q8DR49
C	1001	HIS	-	cloning artifact	UNP Q8DR49
C	1002	MSE	-	cloning artifact	UNP Q8DR49
C	1	MSE	MET	modified residue	UNP Q8DR49
C	34	MSE	MET	modified residue	UNP Q8DR49
C	50	MSE	MET	modified residue	UNP Q8DR49
C	73	MSE	MET	modified residue	UNP Q8DR49
C	95	MSE	MET	modified residue	UNP Q8DR49
C	148	MSE	MET	modified residue	UNP Q8DR49
C	216	MSE	MET	modified residue	UNP Q8DR49
D	1001	HIS	-	cloning artifact	UNP Q8DR49
D	1002	MSE	-	cloning artifact	UNP Q8DR49
D	1	MSE	MET	modified residue	UNP Q8DR49
D	34	MSE	MET	modified residue	UNP Q8DR49
D	50	MSE	MET	modified residue	UNP Q8DR49
D	73	MSE	MET	modified residue	UNP Q8DR49
D	95	MSE	MET	modified residue	UNP Q8DR49
D	148	MSE	MET	modified residue	UNP Q8DR49
D	216	MSE	MET	modified residue	UNP Q8DR49
E	1001	HIS	-	cloning artifact	UNP Q8DR49
E	1002	MSE	-	cloning artifact	UNP Q8DR49
E	1	MSE	MET	modified residue	UNP Q8DR49
E	34	MSE	MET	modified residue	UNP Q8DR49
E	50	MSE	MET	modified residue	UNP Q8DR49
E	73	MSE	MET	modified residue	UNP Q8DR49
E	95	MSE	MET	modified residue	UNP Q8DR49
E	148	MSE	MET	modified residue	UNP Q8DR49
E	216	MSE	MET	modified residue	UNP Q8DR49
F	1001	HIS	-	cloning artifact	UNP Q8DR49
F	1002	MSE	-	cloning artifact	UNP Q8DR49
F	1	MSE	MET	modified residue	UNP Q8DR49
F	34	MSE	MET	modified residue	UNP Q8DR49
F	50	MSE	MET	modified residue	UNP Q8DR49
F	73	MSE	MET	modified residue	UNP Q8DR49
F	95	MSE	MET	modified residue	UNP Q8DR49
F	148	MSE	MET	modified residue	UNP Q8DR49
F	216	MSE	MET	modified residue	UNP Q8DR49

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	128	Total O 128 128	0	0

Continued on next page...

Continued from previous page...

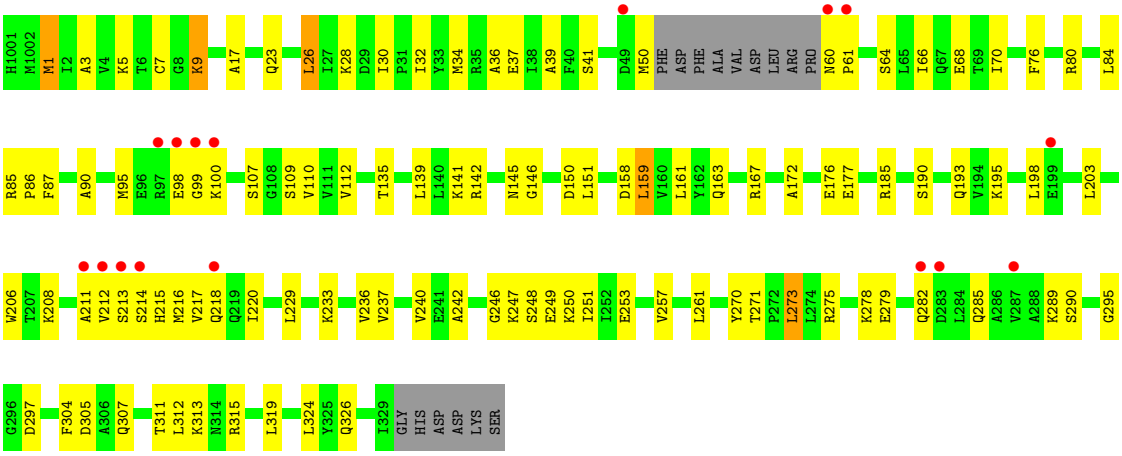
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	133	Total O 133 133	0	0
2	C	122	Total O 122 122	0	0
2	D	136	Total O 136 136	0	0
2	E	131	Total O 131 131	0	0
2	F	129	Total O 129 129	0	0

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.

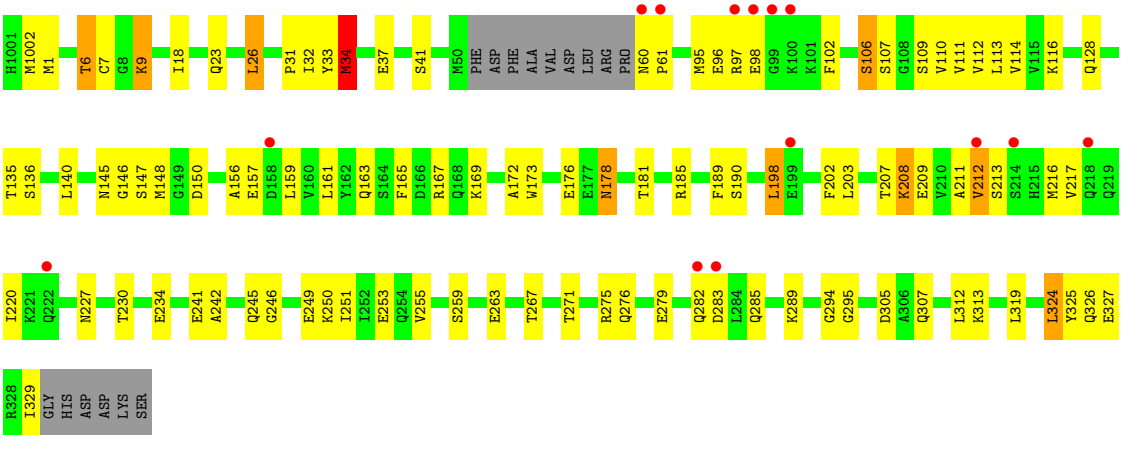
• Molecule 1: phosphomevalonate kinase

Chain A:



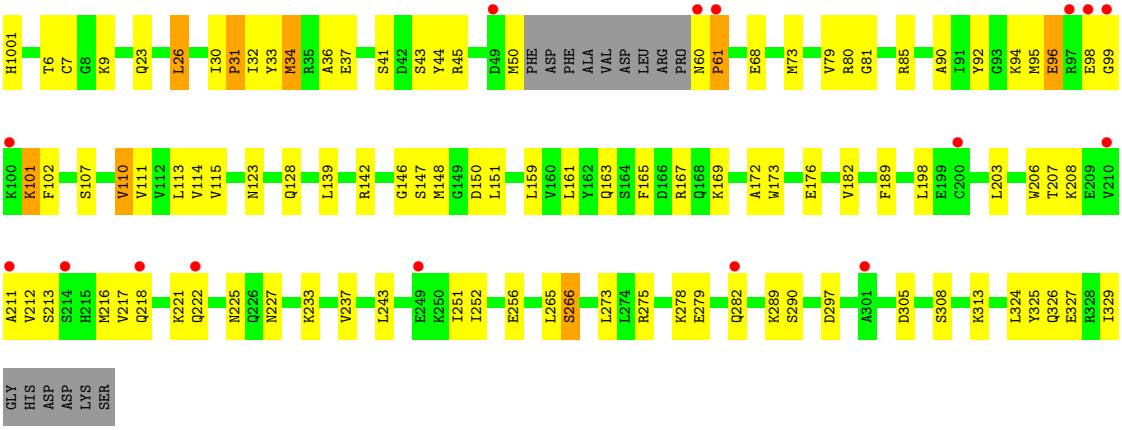
• Molecule 1: phosphomevalonate kinase

Chain B:



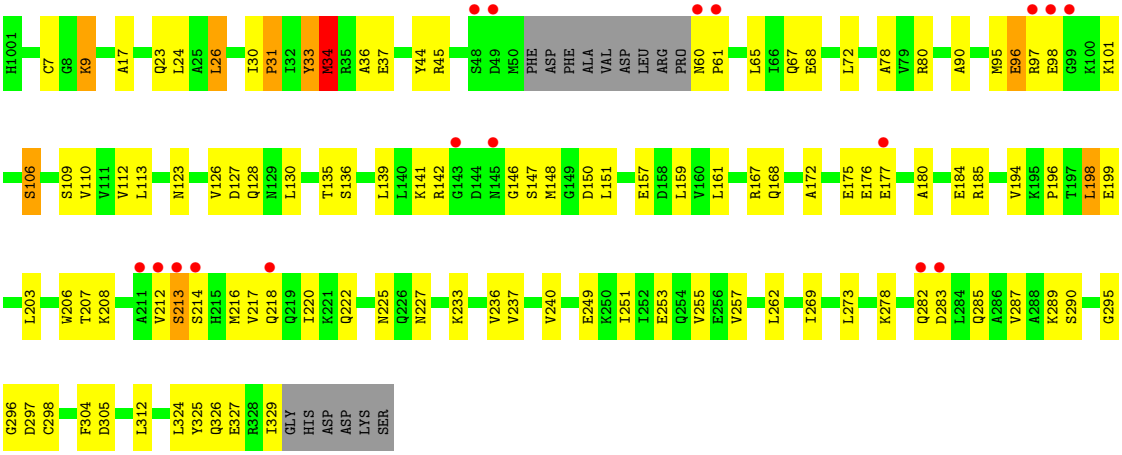
• Molecule 1: phosphomevalonate kinase

Chain C:



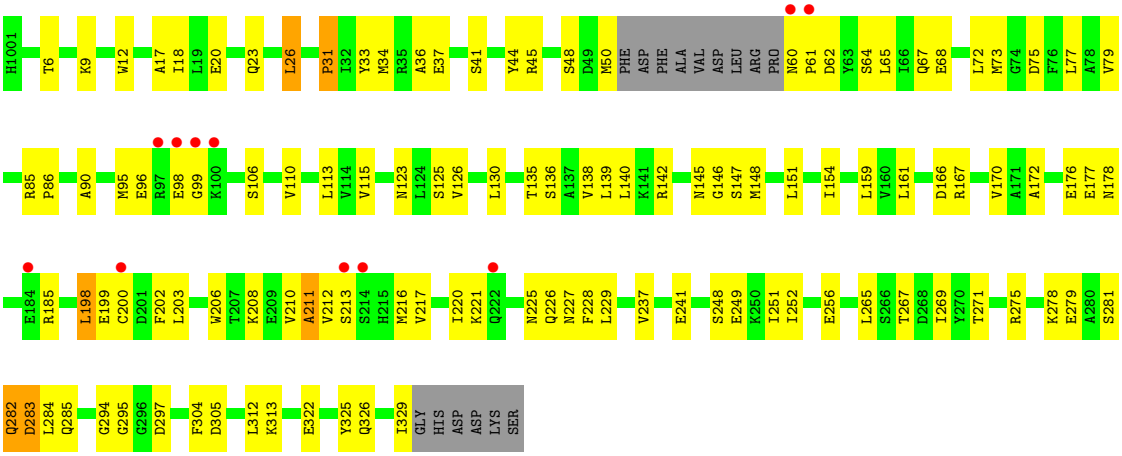
• Molecule 1: phosphomevalonate kinase

Chain D:



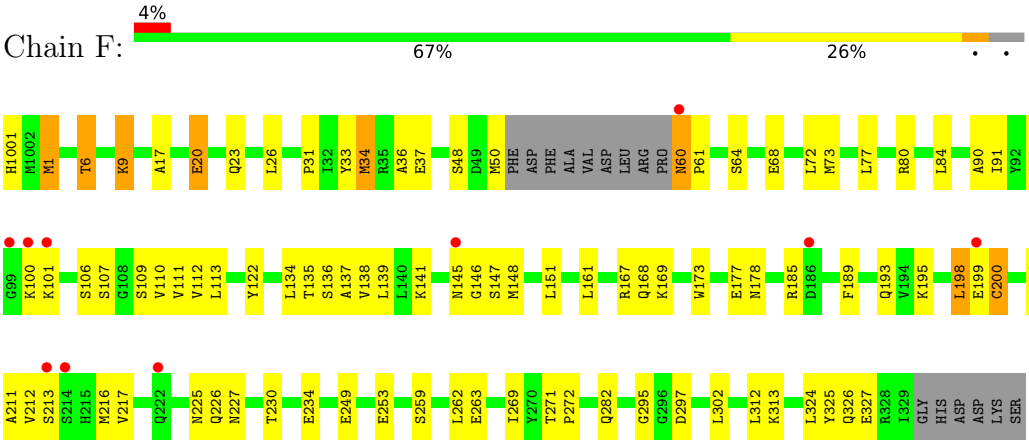
• Molecule 1: phosphomevalonate kinase

Chain E:



• Molecule 1: phosphomevalonate kinase

Chain F:


4%
67%
26%
H1001 M1002 M1 T6 K9 A17 E20 Q23 L26 P31 I32 Y33 M34 R35 A36 E37 S48 D49 M50 PHE ASP PHE ALA VAL ASP LEU ARG PRO N60 P61 S64 E68 L72 M73 L77 R80 L84 A90 I91 Y92 G93 E96 R97 E98
G99 K100 K101 S106 S107 G108 S109 V110 V111 V112 L113 Y122 L134 T135 S136 A137 V138 L139 L140 K141 N145 G146 S147 M148 L151 L161 R167 Q168 K169 W173 E177 N178 R185 D186 F189 Q193 V194 K195 L198 E199 C200 W206 T207 K208 E209 V210
A211 V212 S213 S214 H215 M216 V217 Q222 N225 Q226 N227 T230 E234 E249 E253 S259 L262 E263 I269 Y270 T271 P272 Q282 G295 G296 D297 L302 L312 K313 L324 Y325 Q326 E327 R328 I329 GLY HIS ASP ASP LYS SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α, β, γ	231.61Å 231.61Å 88.59Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 2.42 20.11 – 2.42	Depositor EDS
% Data completeness (in resolution range)	95.0 (20.00-2.42) 97.7 (20.11-2.42)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	2.95 (at 2.41Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.215 , 0.257 0.218 , 0.260	Depositor DCC
R_{free} test set	19213 reflections (9.66%)	wwPDB-VP
Wilson B-factor (Å^2)	35.6	Xtriage
Anisotropy	0.074	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 41.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.017 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	15779	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.01% 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	0.37	0/2529	0.62	2/3401 (0.1%)
1	B	0.38	0/2529	0.63	2/3401 (0.1%)
1	C	0.38	0/2529	0.63	1/3401 (0.0%)
1	D	0.36	0/2529	0.61	2/3401 (0.1%)
1	E	0.38	0/2529	0.63	1/3401 (0.0%)
1	F	0.38	0/2529	0.64	2/3401 (0.1%)
All	All	0.37	0/15174	0.63	10/20406 (0.0%)

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	33	TYR	N-CA-C	6.23	127.83	111.00
1	C	33	TYR	N-CA-C	6.04	127.30	111.00
1	A	159	LEU	N-CA-C	-5.61	95.86	111.00
1	B	33	TYR	N-CA-C	5.60	126.12	111.00
1	B	34	MSE	N-CA-C	-5.30	96.69	111.00

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	2500	0	2539	93	0
1	B	2500	0	2539	99	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2500	0	2539	90	0
1	D	2500	0	2539	104	0
1	E	2500	0	2539	118	0
1	F	2500	0	2539	95	0
2	A	128	0	0	16	0
2	B	133	0	0	18	0
2	C	122	0	0	16	0
2	D	136	0	0	18	0
2	E	131	0	0	19	0
2	F	129	0	0	16	0
All	All	15779	0	15234	590	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 20.

The worst 5 of 590 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:D:225:ASN:HD21	1:D:227:ASN:HB2	1.05	1.13
1:A:213:SER:H	1:A:216:MSE:HE3	1.20	1.06
1:E:60:ASN:HB2	1:E:61:PRO:HD3	1.38	1.05
1:F:148:MSE:HE1	1:F:151:LEU:HD12	1.34	1.04
1:F:148:MSE:CE	1:F:151:LEU:HD12	1.88	1.03

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	318/337 (94%)	299 (94%)	18 (6%)	1 (0%)	41 54
1	B	318/337 (94%)	304 (96%)	11 (4%)	3 (1%)	17 24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	318/337 (94%)	302 (95%)	11 (4%)	5 (2%)	9 12
1	D	318/337 (94%)	300 (94%)	15 (5%)	3 (1%)	17 24
1	E	318/337 (94%)	303 (95%)	10 (3%)	5 (2%)	9 12
1	F	318/337 (94%)	306 (96%)	11 (4%)	1 (0%)	41 54
All	All	1908/2022 (94%)	1814 (95%)	76 (4%)	18 (1%)	17 24

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	99	GLY
1	B	208	LYS
1	E	282	GLN
1	E	283	ASP
1	C	99	GLY

5.3.2 Protein sidechains (i)

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	271/276 (98%)	261 (96%)	10 (4%)	34 51
1	B	271/276 (98%)	255 (94%)	16 (6%)	19 30
1	C	271/276 (98%)	264 (97%)	7 (3%)	46 64
1	D	271/276 (98%)	261 (96%)	10 (4%)	34 51
1	E	271/276 (98%)	263 (97%)	8 (3%)	41 59
1	F	271/276 (98%)	260 (96%)	11 (4%)	30 47
All	All	1626/1656 (98%)	1564 (96%)	62 (4%)	33 50

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	96	GLU
1	D	26	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	101	LYS
1	C	161	LEU
1	D	72	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	223	ASN
1	D	60	ASN
1	F	219	GLN
1	C	326	GLN
1	D	67	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

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, 95^{th} 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	314/337 (93%)	0.07	16 (5%)	28	26	21, 38, 61, 75	0
1	B	314/337 (93%)	0.03	14 (4%)	33	31	21, 37, 59, 75	0
1	C	314/337 (93%)	0.06	16 (5%)	28	26	22, 38, 58, 75	0
1	D	314/337 (93%)	0.18	17 (5%)	25	23	20, 41, 64, 73	0
1	E	314/337 (93%)	0.09	11 (3%)	44	42	20, 35, 59, 77	0
1	F	314/337 (93%)	0.07	12 (3%)	40	38	22, 36, 60, 78	0
All	All	1884/2022 (93%)	0.08	86 (4%)	32	30	20, 37, 61, 78	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	99	GLY	10.3
1	E	98	GLU	6.3
1	F	97	ARG	6.1
1	F	99	GLY	6.1
1	F	98	GLU	5.4

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 (i)

There are no such residues in this entry.