wwPDB X-ray Structure Validation Summary Report (i)

Aug 17, 2022 – 07:19 PM EDT

PDB ID	:	4Q7J
Title	:	Complex structure of viral RNA polymerase
Authors	:	Takeshita, D.
Deposited on	:	2014-04-25
Resolution	:	2.90 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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

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

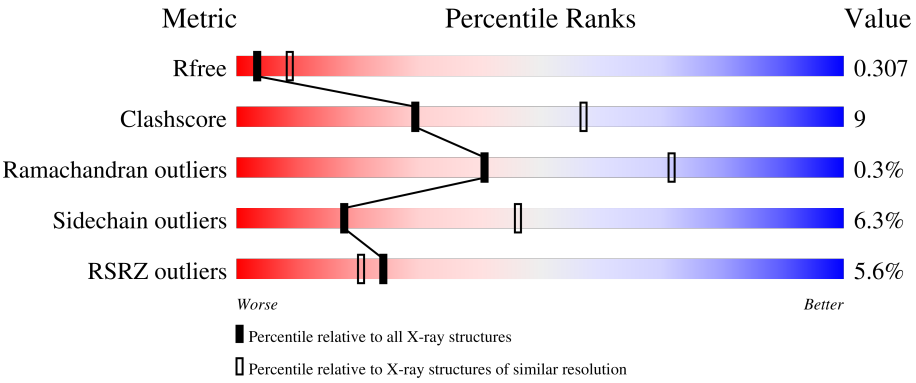
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 2.90 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.



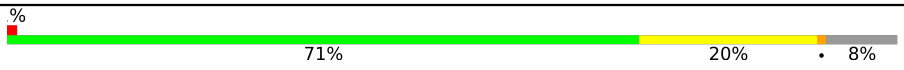
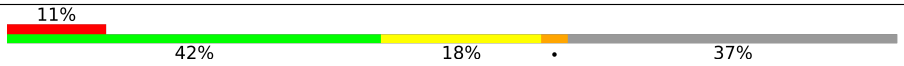
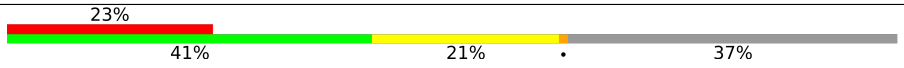
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 <=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	282	4% 76% 21%
1	E	282	8% 83% 15%
2	B	393	% 67% 24% 8%
2	F	393	2% 63% 28% 7%
3	C	594	% 71% 19% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	594	
4	D	281	
4	H	281	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 21369 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 Elongation factor Ts.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	278	Total	C	N	O	S	0	0	0
			2093	1314	359	409	11			
1	E	280	Total	C	N	O	S	0	0	0
			2107	1323	361	412	11			

- Molecule 2 is a protein called Elongation factor Tu 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	362	Total	C	N	O	S	0	0	0
			2790	1769	477	531	13			
2	F	364	Total	C	N	O	S	0	0	0
			2802	1776	479	534	13			

- Molecule 3 is a protein called Q beta replicase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	547	Total	C	N	O	S	0	0	0
			4300	2723	750	807	20			
3	G	545	Total	C	N	O	S	0	0	0
			4287	2716	747	804	20			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	589	HIS	-	expression tag	UNP Q8LTE0
C	590	HIS	-	expression tag	UNP Q8LTE0
C	591	HIS	-	expression tag	UNP Q8LTE0
C	592	HIS	-	expression tag	UNP Q8LTE0
C	593	HIS	-	expression tag	UNP Q8LTE0
C	594	HIS	-	expression tag	UNP Q8LTE0
G	589	HIS	-	expression tag	UNP Q8LTE0
G	590	HIS	-	expression tag	UNP Q8LTE0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	591	HIS	-	expression tag	UNP Q8LTE0
G	592	HIS	-	expression tag	UNP Q8LTE0
G	593	HIS	-	expression tag	UNP Q8LTE0
G	594	HIS	-	expression tag	UNP Q8LTE0

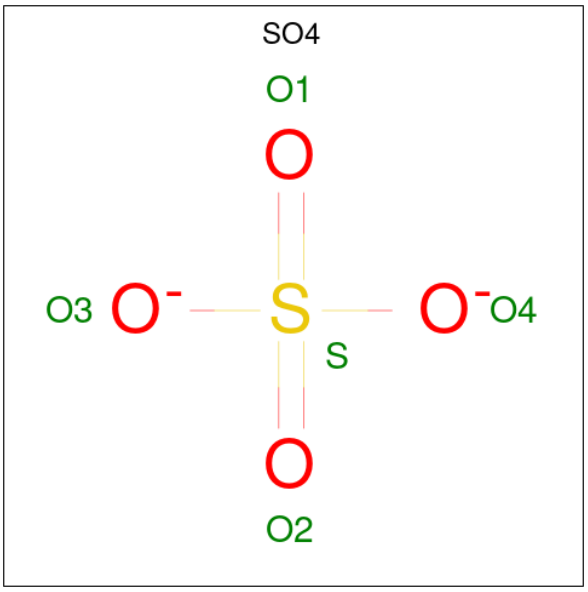
- Molecule 4 is a protein called 30S ribosomal protein S1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	D	177	Total	C	N	O	0	0	0
			1359	855	233	271			
4	H	177	Total	C	N	O	0	0	0
			1358	855	232	271			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	274	LEU	-	expression tag	UNP P0AG67
D	275	GLU	-	expression tag	UNP P0AG67
D	276	HIS	-	expression tag	UNP P0AG67
D	277	HIS	-	expression tag	UNP P0AG67
D	278	HIS	-	expression tag	UNP P0AG67
D	279	HIS	-	expression tag	UNP P0AG67
D	280	HIS	-	expression tag	UNP P0AG67
D	281	HIS	-	expression tag	UNP P0AG67
H	274	LEU	-	expression tag	UNP P0AG67
H	275	GLU	-	expression tag	UNP P0AG67
H	276	HIS	-	expression tag	UNP P0AG67
H	277	HIS	-	expression tag	UNP P0AG67
H	278	HIS	-	expression tag	UNP P0AG67
H	279	HIS	-	expression tag	UNP P0AG67
H	280	HIS	-	expression tag	UNP P0AG67
H	281	HIS	-	expression tag	UNP P0AG67

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total O S 5 4 1	0	0
5	C	1	Total O S 5 4 1	0	0
5	G	1	Total O S 5 4 1	0	0
5	G	1	Total O S 5 4 1	0	0

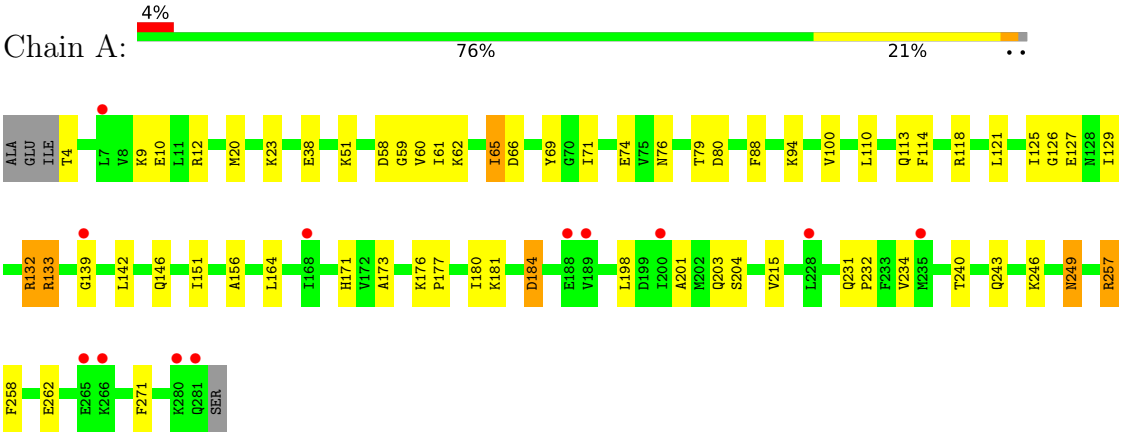
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	30	Total O 30 30	0	0
6	B	36	Total O 36 36	0	0
6	C	65	Total O 65 65	0	0
6	D	9	Total O 9 9	0	0
6	E	28	Total O 28 28	0	0
6	F	27	Total O 27 27	0	0
6	G	52	Total O 52 52	0	0
6	H	6	Total O 6 6	0	0

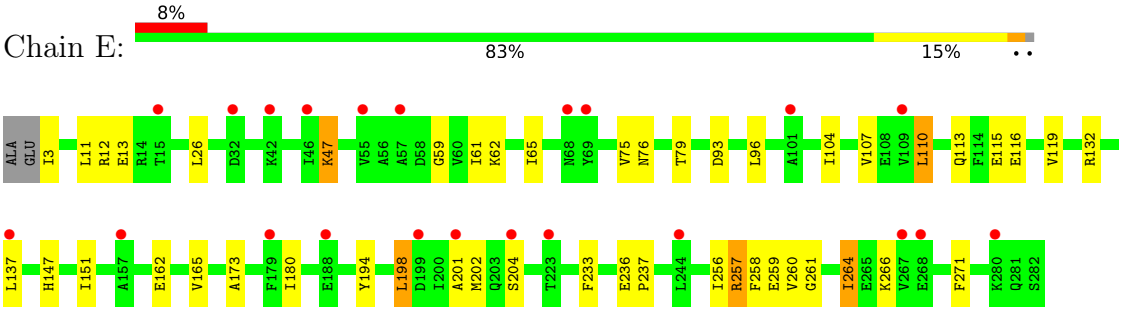
3 Residue-property plots (i)

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.

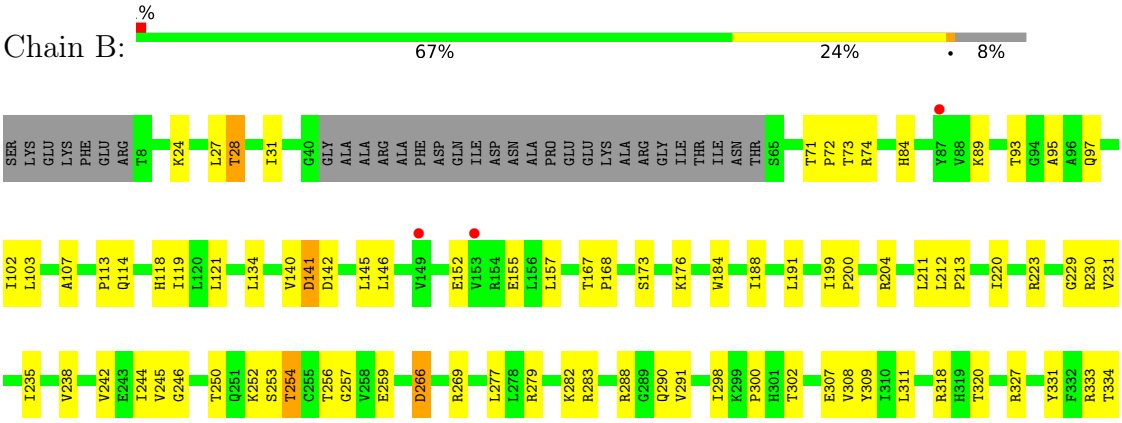
• Molecule 1: Elongation factor Ts

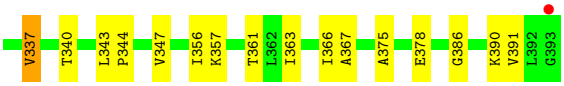


• Molecule 1: Elongation factor Ts



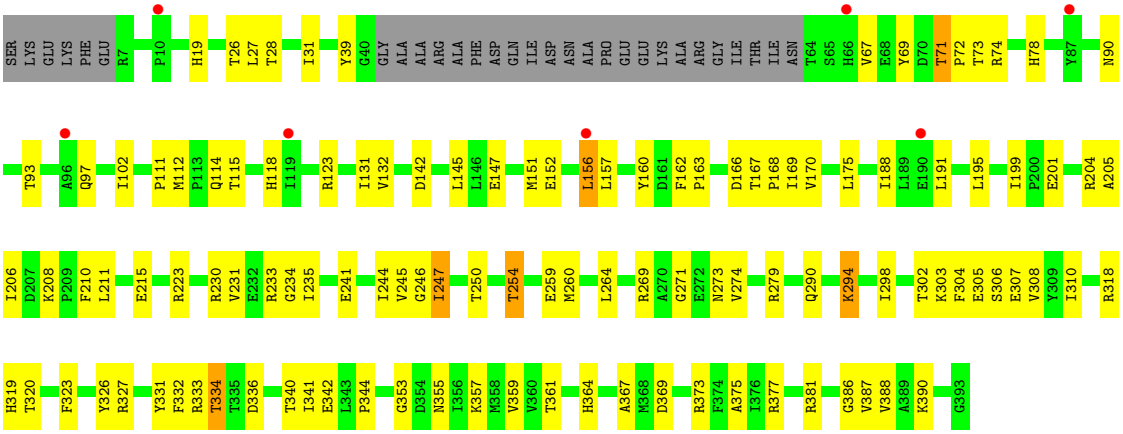
• Molecule 2: Elongation factor Tu 1





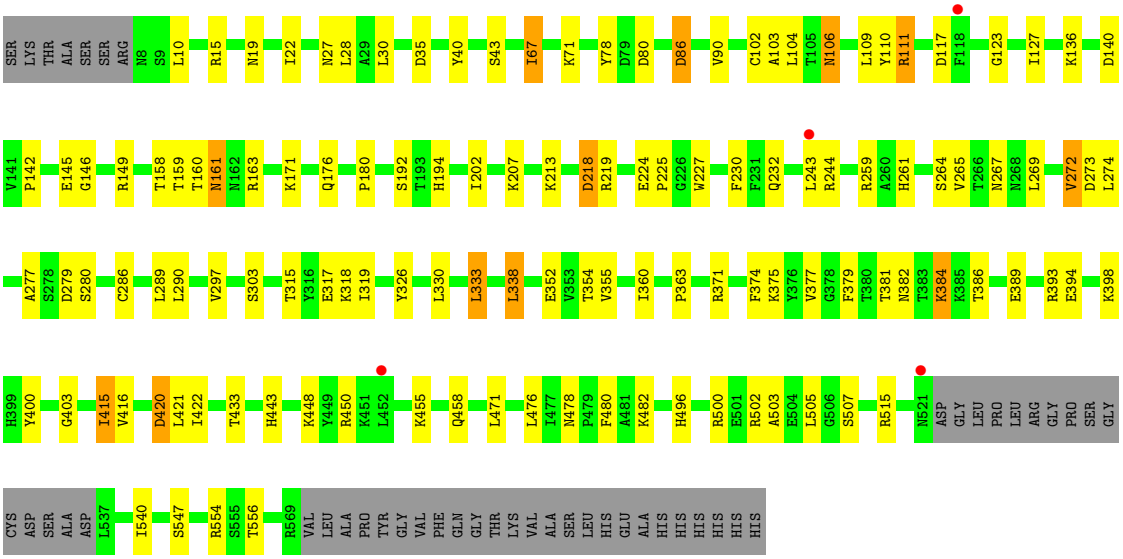
• Molecule 2: Elongation factor Tu 1

Chain F:



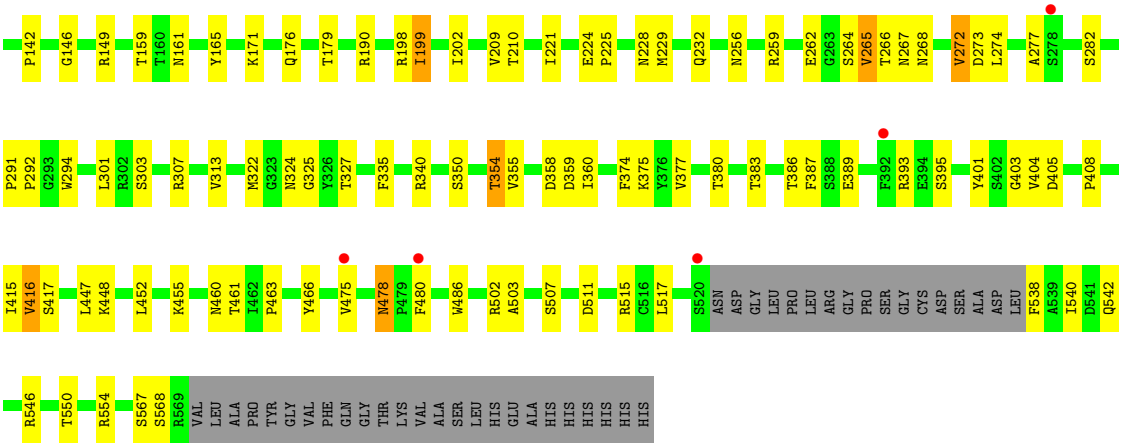
• Molecule 3: Q beta replicase

Chain C:



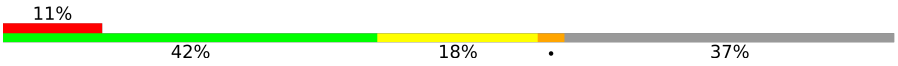
• Molecule 3: Q beta replicase

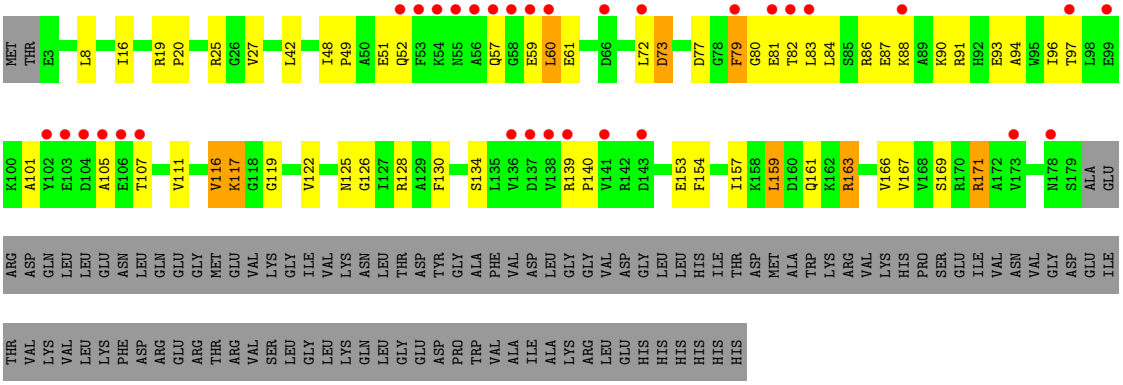
Chain G:



• Molecule 4: 30S ribosomal protein S1

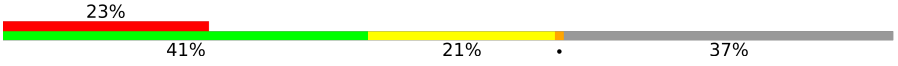
Chain D:

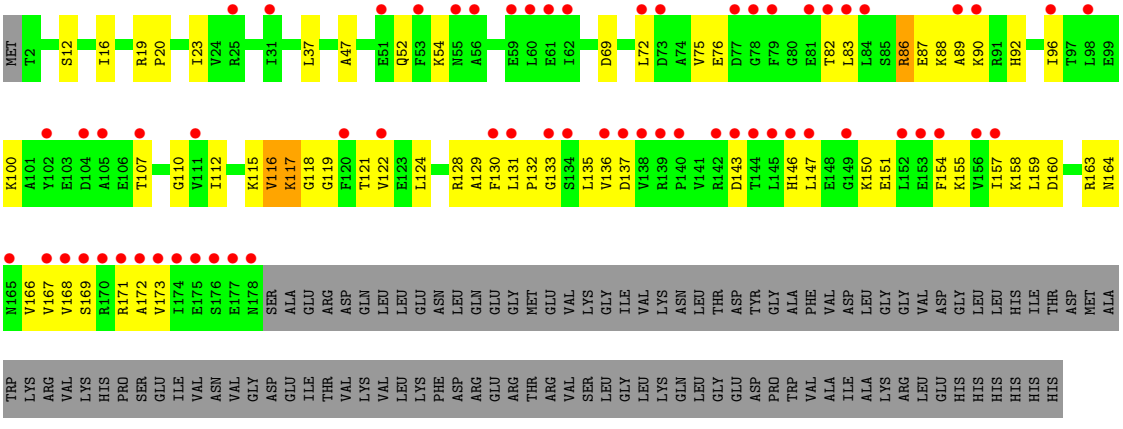




• Molecule 4: 30S ribosomal protein S1

Chain H:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	132.19Å 150.83Å 189.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 2.90 46.99 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.1 (19.98-2.90) 99.1 (46.99-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.48 (at 2.91Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.255 , 0.309 0.254 , 0.307	Depositor DCC
R_{free} test set	4184 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	52.6	Xtriage
Anisotropy	0.212	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 57.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.36$, $< L^2 > = 0.19$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	21369	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.91% 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

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.23	0/2114	0.39	0/2837
1	E	0.23	0/2128	0.40	0/2856
2	B	0.23	0/2842	0.44	0/3848
2	F	0.23	0/2854	0.43	0/3865
3	C	0.24	0/4395	0.41	0/5960
3	G	0.24	0/4382	0.41	0/5942
4	D	0.23	0/1374	0.45	0/1855
4	H	0.24	0/1373	0.48	0/1854
All	All	0.23	0/21462	0.42	0/29017

There are no bond length outliers.

There are no bond angle outliers.

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	2093	0	2140	42	0
1	E	2107	0	2156	23	0
2	B	2790	0	2808	62	0
2	F	2802	0	2817	78	0
3	C	4300	0	4241	76	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	4287	0	4233	80	0
4	D	1359	0	1366	31	0
4	H	1358	0	1367	40	0
5	C	10	0	0	0	0
5	G	10	0	0	1	0
6	A	30	0	0	6	0
6	B	36	0	0	5	0
6	C	65	0	0	7	0
6	D	9	0	0	0	0
6	E	28	0	0	2	0
6	F	27	0	0	6	0
6	G	52	0	0	5	0
6	H	6	0	0	0	0
All	All	21369	0	21128	399	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 9.

The worst 5 of 399 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:G:274:LEU:HB2	3:G:277:ALA:HB2	1.64	0.80
2:B:235:ILE:HD12	2:B:269:ARG:HD2	1.65	0.77
4:H:116:VAL:HG12	4:H:117:LYS:H	1.47	0.77
4:D:87:GLU:HG3	4:D:91:ARG:HD3	1.67	0.76
3:C:382:ASN:HD21	3:C:384:LYS:HG2	1.50	0.76

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	276/282 (98%)	268 (97%)	8 (3%)	0	100 100
1	E	278/282 (99%)	265 (95%)	12 (4%)	1 (0%)	34 66
2	B	358/393 (91%)	341 (95%)	17 (5%)	0	100 100
2	F	360/393 (92%)	344 (96%)	14 (4%)	2 (1%)	25 58
3	C	543/594 (91%)	521 (96%)	22 (4%)	0	100 100
3	G	541/594 (91%)	523 (97%)	18 (3%)	0	100 100
4	D	175/281 (62%)	160 (91%)	14 (8%)	1 (1%)	25 58
4	H	175/281 (62%)	155 (89%)	17 (10%)	3 (2%)	9 31
All	All	2706/3100 (87%)	2577 (95%)	122 (4%)	7 (0%)	41 71

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	264	ILE
4	H	117	LYS
4	H	118	GLY
2	F	333	ARG
4	H	116	VAL

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	216/219 (99%)	200 (93%)	16 (7%)	13 38
1	E	218/219 (100%)	202 (93%)	16 (7%)	14 38
2	B	301/325 (93%)	283 (94%)	18 (6%)	19 49
2	F	302/325 (93%)	288 (95%)	14 (5%)	27 60
3	C	468/508 (92%)	438 (94%)	30 (6%)	17 45
3	G	467/508 (92%)	448 (96%)	19 (4%)	30 64
4	D	146/239 (61%)	128 (88%)	18 (12%)	4 14
4	H	146/239 (61%)	134 (92%)	12 (8%)	11 32

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2264/2582 (88%)	2121 (94%)	143 (6%)	18 46

5 of 143 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	90	VAL
3	G	265	VAL
3	G	540	ILE
3	C	272	VAL
3	C	265	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	197	GLN
3	G	206	ASN
3	G	478	ASN
3	G	460	ASN
3	C	382	ASN

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)

4 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
5	SO4	G	602	-	4,4,4	0.13	0	6,6,6	0.15	0
5	SO4	C	602	-	4,4,4	0.13	0	6,6,6	0.16	0
5	SO4	G	601	-	4,4,4	0.15	0	6,6,6	0.13	0
5	SO4	C	601	-	4,4,4	0.13	0	6,6,6	0.12	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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	G	602	SO4	1	0

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	278/282 (98%)	0.24	12 (4%)	35	31	44, 71, 97, 112	0
1	E	280/282 (99%)	0.42	22 (7%)	12	10	53, 78, 108, 132	0
2	B	362/393 (92%)	0.04	4 (1%)	80	80	18, 49, 74, 92	0
2	F	364/393 (92%)	0.14	7 (1%)	66	65	25, 52, 78, 98	0
3	C	547/594 (92%)	0.13	4 (0%)	87	87	23, 48, 82, 113	0
3	G	545/594 (91%)	0.21	7 (1%)	77	77	22, 54, 82, 101	0
4	D	177/281 (62%)	1.17	32 (18%)	1	1	52, 103, 143, 167	0
4	H	177/281 (62%)	1.62	64 (36%)	0	0	49, 111, 149, 163	0
All	All	2730/3100 (88%)	0.34	152 (5%)	24	20	18, 60, 116, 167	0

The worst 5 of 152 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	82	THR	8.5
4	H	60	LEU	7.3
4	H	143	ASP	7.0
4	H	152	LEU	6.5
4	H	82	THR	6.4

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

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, 95^{th} 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
5	SO4	G	602	5/5	0.85	0.18	89,92,98,111	0
5	SO4	C	602	5/5	0.92	0.16	86,90,99,121	0
5	SO4	G	601	5/5	0.97	0.11	48,48,50,70	0
5	SO4	C	601	5/5	0.97	0.10	50,51,66,67	0

6.5 Other polymers

There are no such residues in this entry.