



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 14, 2020 – 02:47 PM BST

PDB ID : 6WH9
Title : Ketoreductase from module 1 of the 6-deoxyerythronolide B synthase (KR1)
in complex with antibody fragment (Fab) 1D10
Authors : Cogan, D.P.; Mathews, I.I.; Khosla, C.
Deposited on : 2020-04-07
Resolution : 2.75 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 : 2.14.4.dev1
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) : 2.14.4.dev1

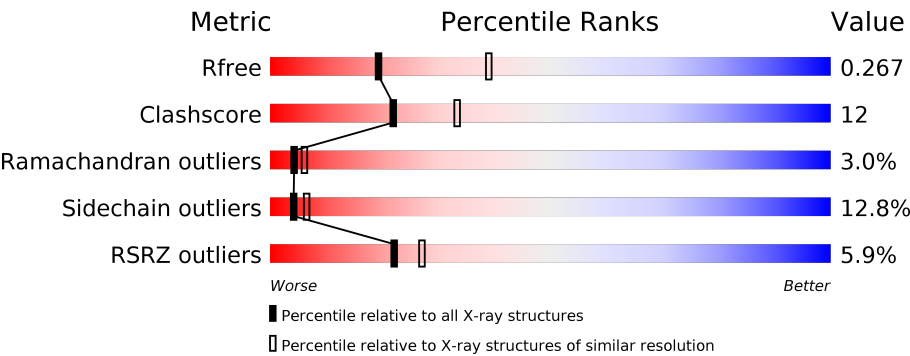
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	485	<div><div>12%</div><div><div></div><div>65%</div><div>20%</div><div>•</div><div>11%</div></div></div>
1	D	485	<div><div>6%</div><div><div></div><div>70%</div><div>19%</div><div>•</div><div>7%</div></div></div>
1	G	485	<div><div>6%</div><div><div></div><div>66%</div><div>23%</div><div>•</div><div>9%</div></div></div>
2	B	246	<div><div>2%</div><div><div></div><div>70%</div><div>17%</div><div>•</div><div>9%</div></div></div>
2	E	246	<div><div>7%</div><div><div></div><div>54%</div><div>30%</div><div>7%</div><div>9%</div></div></div>
2	H	246	<div><div>2%</div><div><div></div><div>63%</div><div>23%</div><div>•</div><div>11%</div></div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain			
3	C	231	2%	64%	23%	6% 7%
3	F	231	2%	66%	22%	5% 6%
3	I	231	2%	66%	20%	7% 6%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 19434 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 KR1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	439	Total	C	N	O	S	0	0	0
			3237	2023	590	617	7			
1	D	450	Total	C	N	O	S	0	0	0
			3283	2049	593	635	6			
1	A	431	Total	C	N	O	S	0	0	0
			3104	1940	558	599	7			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	1444	GLY	-	expression tag	UNP Q5UNP6
G	1445	SER	-	expression tag	UNP Q5UNP6
G	1446	HIS	-	expression tag	UNP Q5UNP6
G	1447	MET	-	expression tag	UNP Q5UNP6
D	1444	GLY	-	expression tag	UNP Q5UNP6
D	1445	SER	-	expression tag	UNP Q5UNP6
D	1446	HIS	-	expression tag	UNP Q5UNP6
D	1447	MET	-	expression tag	UNP Q5UNP6
A	1444	GLY	-	expression tag	UNP Q5UNP6
A	1445	SER	-	expression tag	UNP Q5UNP6
A	1446	HIS	-	expression tag	UNP Q5UNP6
A	1447	MET	-	expression tag	UNP Q5UNP6

- Molecule 2 is a protein called 1D10 (Fab heavy chain).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	224	Total	C	N	O	S	0	0	0
			1640	1027	275	330	8			
2	B	225	Total	C	N	O	S	0	0	0
			1633	1022	270	333	8			
2	H	219	Total	C	N	O	S	0	0	0
			1613	1013	271	322	7			

- Molecule 3 is a protein called 1D10 (Fab light chain).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	216	Total	C	N	O	S	0	0	0
			1573	997	258	313	5			
3	C	215	Total	C	N	O	S	0	0	0
			1577	999	257	316	5			
3	I	216	Total	C	N	O	S	0	1	0
			1593	1009	259	320	5			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	G	1	Total	O	S	0	0
			5	4	1		
4	E	1	Total	O	S	0	0
			5	4	1		
4	F	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		
4	C	1	Total	O	S	0	0
			5	4	1		
4	D	1	Total	O	S	0	0
			5	4	1		
4	A	1	Total	O	S	0	0
			5	4	1		

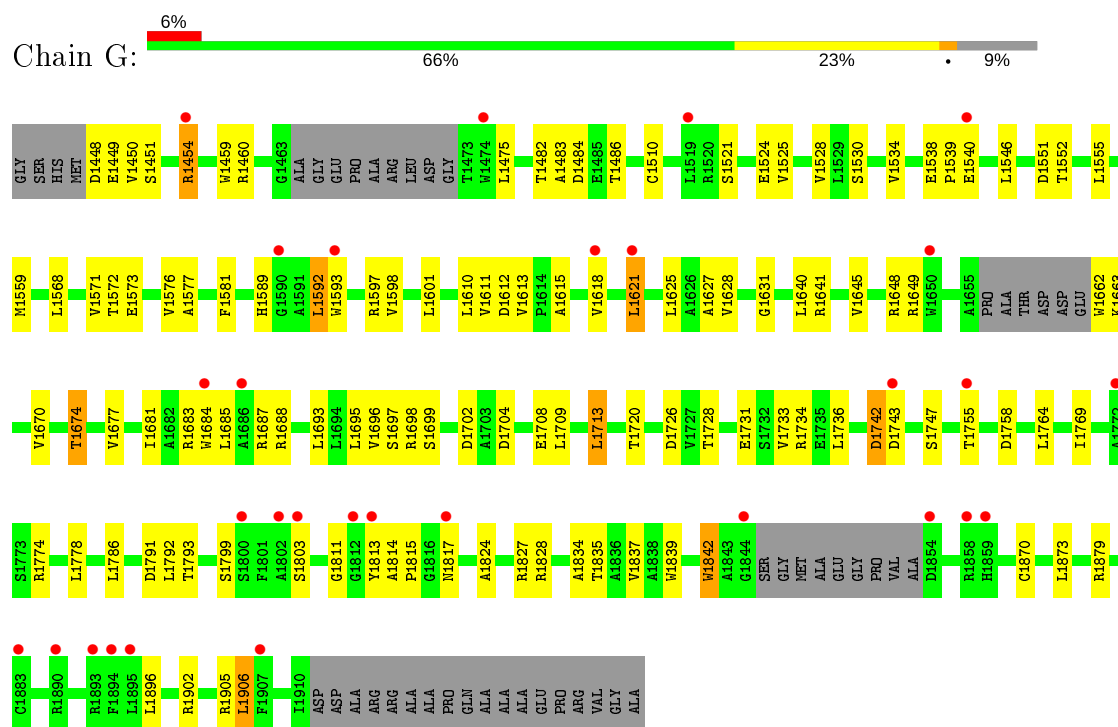
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	G	14	Total 14	O 14	0	0
5	E	14	Total 14	O 14	0	0
5	F	20	Total 20	O 20	0	0
5	B	13	Total 13	O 13	0	0
5	C	26	Total 26	O 26	0	0
5	H	16	Total 16	O 16	0	0
5	I	23	Total 23	O 23	0	0
5	D	9	Total 9	O 9	0	0
5	A	11	Total 11	O 11	0	0

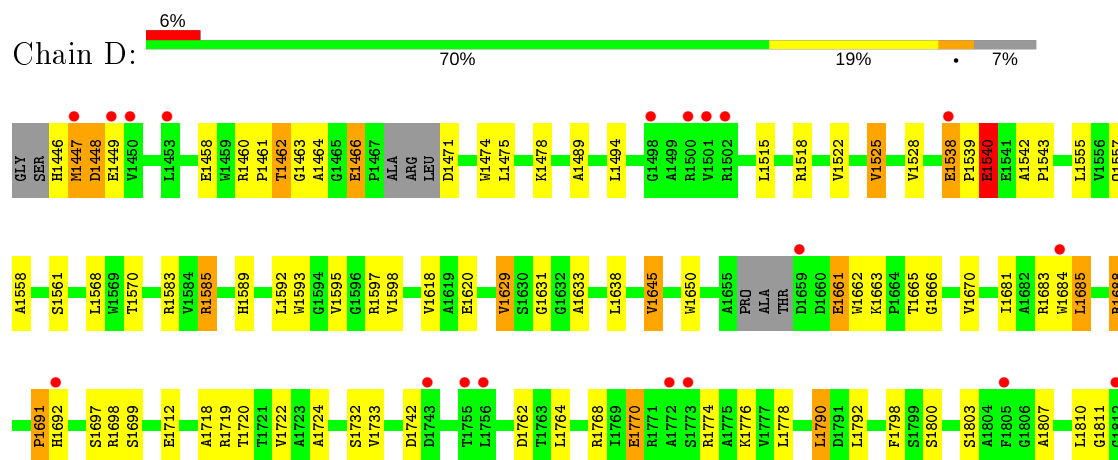
3 Residue-property plots

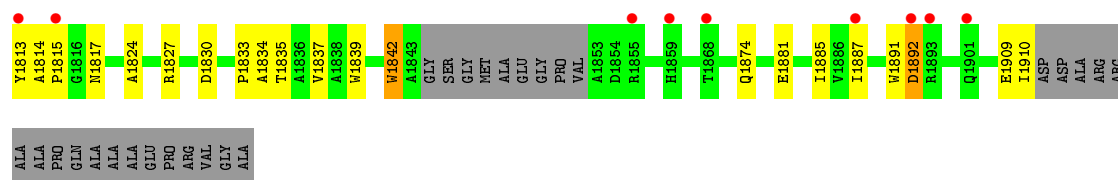
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: KR1

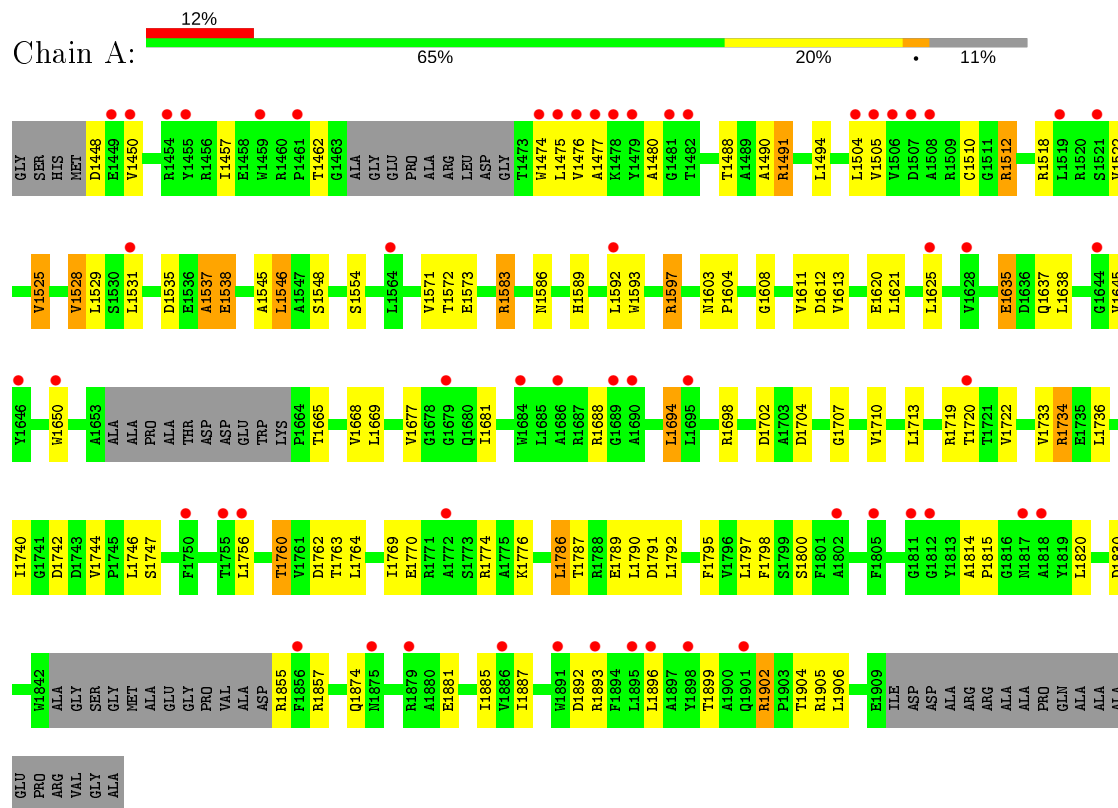


• Molecule 1: KR1

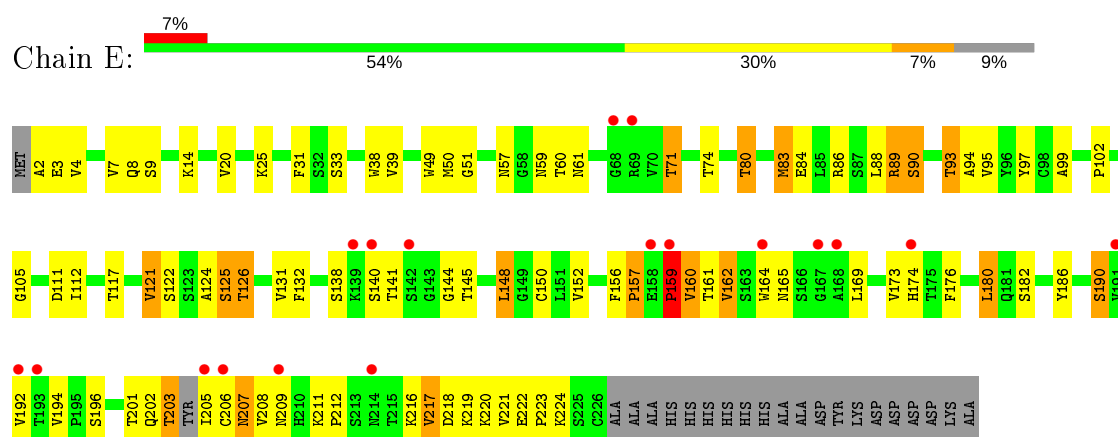




• Molecule 1: KR1

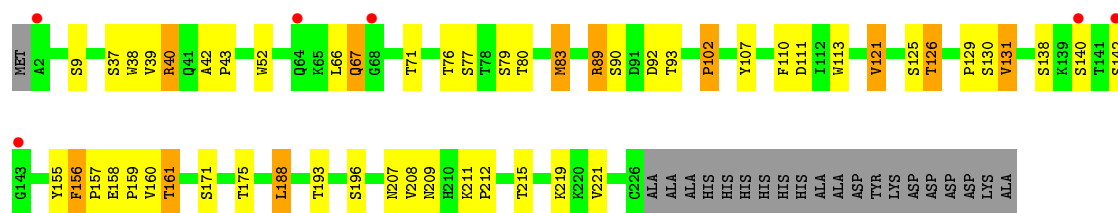


• Molecule 2: 1D10 (Fab heavy chain)

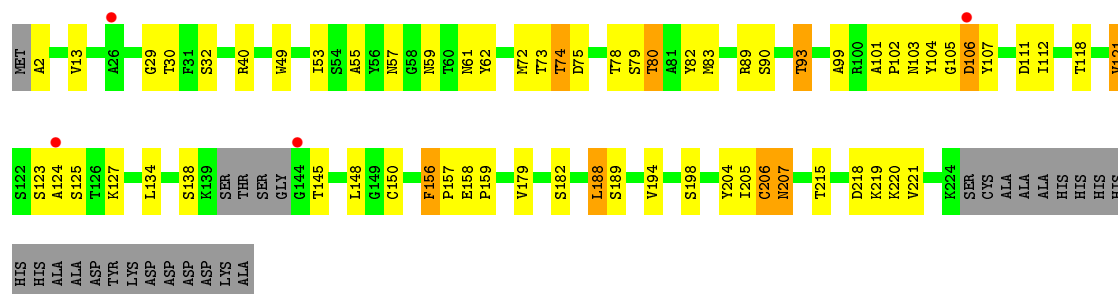


• Molecule 2: 1D10 (Fab heavy chain)

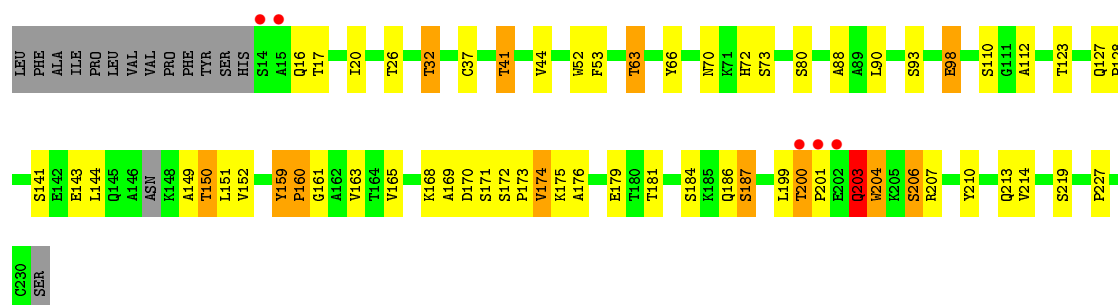




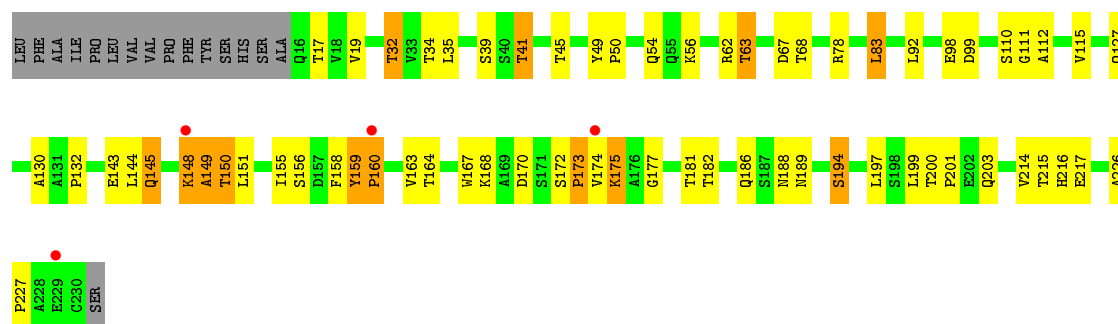
• Molecule 2: 1D10 (Fab heavy chain)



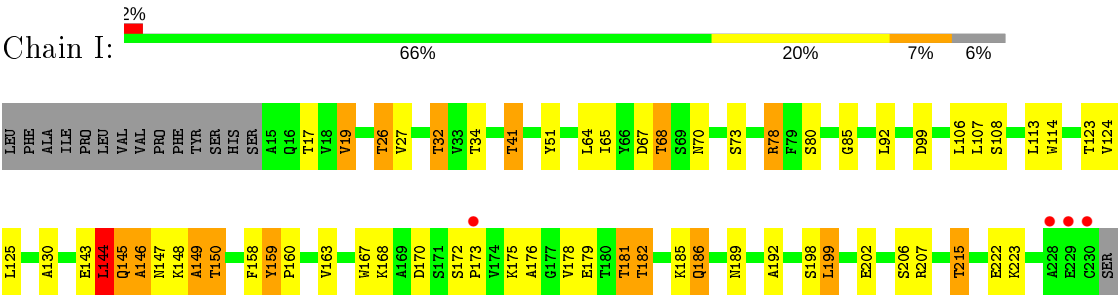
• Molecule 3: 1D10 (Fab light chain)



• Molecule 3: 1D10 (Fab light chain)



• Molecule 3: 1D10 (Fab light chain)



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	186.03Å 114.81Å 186.14Å 90.00° 96.53° 90.00°	Depositor
Resolution (Å)	39.59 – 2.75 39.59 – 2.75	Depositor EDS
% Data completeness (in resolution range)	99.6 (39.59-2.75) 99.6 (39.59-2.75)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.85 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.228 , 0.271 0.229 , 0.267	Depositor DCC
R_{free} test set	5043 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	83.7	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 59.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	19434	wwPDB-VP
Average B, all atoms (Å ²)	107.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% 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: 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	0.69	0/3156	0.81	0/4311
1	D	0.68	0/3340	0.81	0/4558
1	G	0.67	0/3294	0.80	0/4490
2	B	0.69	0/1672	0.88	0/2291
2	E	0.70	0/1676	0.87	0/2286
2	H	0.70	0/1650	0.89	0/2253
3	C	0.74	1/1620 (0.1%)	0.95	1/2223 (0.0%)
3	F	0.69	0/1615	0.91	0/2215
3	I	0.72	0/1636	0.92	0/2244
All	All	0.69	1/19659 (0.0%)	0.86	1/26871 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	159	TYR	C-N	8.23	1.49	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	159	TYR	N-CA-CB	5.98	121.37	110.60

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	3104	0	3008	54	0
1	D	3283	0	3191	59	0
1	G	3237	0	3184	65	0
2	B	1633	0	1543	32	0
2	E	1640	0	1589	78	0
2	H	1613	0	1561	49	0
3	C	1577	0	1506	60	0
3	F	1573	0	1500	36	0
3	I	1593	0	1525	37	0
4	A	5	0	0	0	0
4	B	5	0	0	0	0
4	C	5	0	0	0	0
4	D	5	0	0	0	0
4	E	5	0	0	0	0
4	F	5	0	0	0	0
4	G	5	0	0	0	0
5	A	11	0	0	1	0
5	B	13	0	0	0	0
5	C	26	0	0	2	0
5	D	9	0	0	0	0
5	E	14	0	0	0	0
5	F	20	0	0	3	0
5	G	14	0	0	3	0
5	H	16	0	0	0	0
5	I	23	0	0	1	0
All	All	19434	0	18607	446	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 12.

The worst 5 of 446 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:A:1760:THR:HB	1:A:1762:ASP:OD1	1.60	1.00
2:E:90:SER:O	2:E:93:THR:HG23	1.66	0.96
2:B:89:ARG:HD2	2:B:90:SER:H	1.38	0.89
2:E:20:VAL:HG23	2:E:88:LEU:HD11	1.53	0.89
3:C:186:GLN:NE2	3:C:188:ASN:HD22	1.75	0.83

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	423/485 (87%)	376 (89%)	41 (10%)	6 (1%)	11	19
1	D	442/485 (91%)	402 (91%)	28 (6%)	12 (3%)	5	7
1	G	431/485 (89%)	380 (88%)	43 (10%)	8 (2%)	8	14
2	B	223/246 (91%)	200 (90%)	16 (7%)	7 (3%)	4	6
2	E	220/246 (89%)	181 (82%)	28 (13%)	11 (5%)	2	2
2	H	215/246 (87%)	193 (90%)	20 (9%)	2 (1%)	17	31
3	C	213/231 (92%)	186 (87%)	18 (8%)	9 (4%)	3	3
3	F	212/231 (92%)	183 (86%)	18 (8%)	11 (5%)	2	2
3	I	215/231 (93%)	185 (86%)	19 (9%)	11 (5%)	2	2
All	All	2594/2886 (90%)	2286 (88%)	231 (9%)	77 (3%)	4	6

5 of 77 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	1484	ASP
1	G	1539	PRO
1	G	1546	LEU
1	G	1615	ALA
2	E	125	SER

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	295/348 (85%)	246 (83%)	49 (17%)	2	3
1	D	315/348 (90%)	282 (90%)	33 (10%)	7	11
1	G	314/348 (90%)	278 (88%)	36 (12%)	5	9
2	B	179/202 (89%)	154 (86%)	25 (14%)	3	5
2	E	182/202 (90%)	155 (85%)	27 (15%)	3	4
2	H	178/202 (88%)	155 (87%)	23 (13%)	4	6
3	C	170/191 (89%)	157 (92%)	13 (8%)	13	23
3	F	168/191 (88%)	144 (86%)	24 (14%)	3	4
3	I	172/191 (90%)	148 (86%)	24 (14%)	3	5
All	All	1973/2223 (89%)	1719 (87%)	254 (13%)	4	6

5 of 254 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	39	SER
2	H	188	LEU
1	A	1722	VAL
3	C	150	THR
2	H	61	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	174	HIS
2	B	181	GLN
1	D	1874	GLN
3	F	96	GLN
2	B	59	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 [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

7 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	$\# Z > 2$
4	SO4	F	301	-	4,4,4	0.38	0	6,6,6	0.05	0
4	SO4	D	2001	-	4,4,4	0.38	0	6,6,6	0.07	0
4	SO4	G	2001	-	4,4,4	0.40	0	6,6,6	0.09	0
4	SO4	A	2001	-	4,4,4	0.41	0	6,6,6	0.04	0
4	SO4	E	301	-	4,4,4	0.37	0	6,6,6	0.09	0
4	SO4	B	301	-	4,4,4	0.37	0	6,6,6	0.08	0
4	SO4	C	301	-	4,4,4	0.39	0	6,6,6	0.05	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 [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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	431/485 (88%)	0.68	56 (12%) 3 4	89, 144, 181, 212	0
1	D	450/485 (92%)	0.40	28 (6%) 20 25	76, 121, 160, 211	0
1	G	439/485 (90%)	0.45	30 (6%) 17 20	76, 119, 164, 209	0
2	B	225/246 (91%)	0.00	6 (2%) 54 63	61, 86, 111, 162	0
2	E	224/246 (91%)	0.50	18 (8%) 12 15	75, 104, 159, 232	0
2	H	219/246 (89%)	0.02	4 (1%) 68 76	57, 83, 109, 149	0
3	C	215/231 (93%)	-0.23	4 (1%) 66 75	55, 69, 101, 152	0
3	F	216/231 (93%)	-0.03	5 (2%) 60 69	57, 76, 111, 153	0
3	I	216/231 (93%)	-0.10	4 (1%) 66 75	57, 75, 113, 142	0
All	All	2635/2886 (91%)	0.27	155 (5%) 22 27	55, 104, 164, 232	0

The worst 5 of 155 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	201	PRO	9.7
3	F	200	THR	4.8
1	A	1895	LEU	4.7
2	E	206	CYS	4.5
1	A	1650	TRP	4.5

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

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(\AA^2)	Q<0.9
4	SO4	B	301	5/5	0.86	0.19	109,116,127,128	0
4	SO4	E	301	5/5	0.87	0.19	118,123,130,136	0
4	SO4	F	301	5/5	0.93	0.10	123,128,132,135	0
4	SO4	C	301	5/5	0.94	0.09	127,129,132,132	0
4	SO4	A	2001	5/5	0.96	0.24	88,94,101,105	0
4	SO4	G	2001	5/5	0.96	0.31	86,89,101,104	0
4	SO4	D	2001	5/5	0.97	0.30	87,90,96,96	0

6.5 Other polymers [i](#)

There are no such residues in this entry.