wwPDB X-ray Structure Validation Summary Report ⓘ

May 3, 2025 – 04:29 PM EDT

PDB ID	:	4EGV / pdb_00004egv
Title	:	Crystal structure of a monomeric SCP2-thiolase like protein type 1 (STLP1) from Mycobacterium smegmatis
Authors	:	Janardan, N.; Harijan, R.K.; Wierenga, R.K.; Murthy, M.R.N.
Deposited on	:	2012-04-01
Resolution	:	2.71 Å(reported)

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

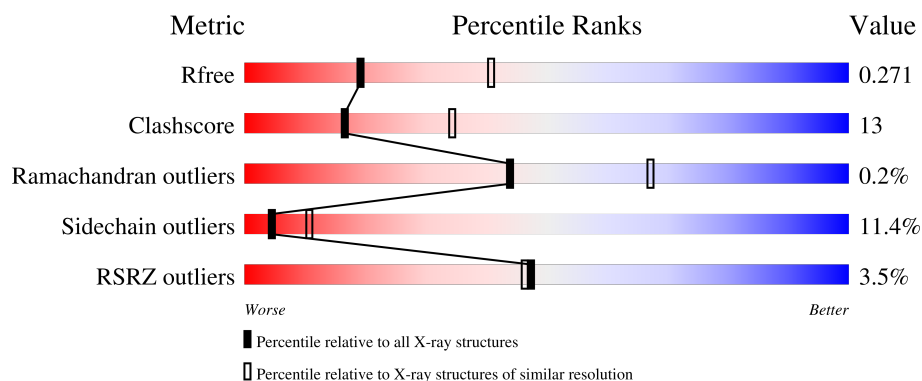|		
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 2.71 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		0.271
Clashscore		13
Ramachandran outliers		0.2%
Sidechain outliers		11.4%
RSRZ outliers		3.5%

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$_{free}$	164625	4050 (2.74-2.70)
Clashscore	180529	4439 (2.74-2.70)
Ramachandran outliers	177936	4374 (2.74-2.70)
Sidechain outliers	177891	4375 (2.74-2.70)
RSRZ outliers	164620	4050 (2.74-2.70)

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	520	3% 66% 17% • • 14%
1	B	520	4% 65% 19% • • 13%
1	C	520	4% 64% 18% • 13%
1	D	520	2% 62% 18% 5% 14%
1	E	520	2% 65% 17% • 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	520	

3%

63% 18% · · 14%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 20206 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 Acetyl-CoA acetyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	448	Total	C	N	O	S	Se	0	0	0
			3327	2086	579	650	3	9			
1	B	452	Total	C	N	O	S	Se	0	3	0
			3376	2115	588	661	3	9			
1	C	453	Total	C	N	O	S	Se	0	0	0
			3373	2116	588	656	3	10			
1	D	445	Total	C	N	O	S	Se	0	0	0
			3299	2066	576	645	3	9			
1	E	447	Total	C	N	O	S	Se	0	2	0
			3336	2090	585	648	3	10			
1	F	449	Total	C	N	O	S	Se	0	0	0
			3335	2090	581	652	3	9			

There are 78 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	MSE	-	expression tag	UNP A0R3M0
A	-11	ARG	-	expression tag	UNP A0R3M0
A	-10	GLY	-	expression tag	UNP A0R3M0
A	-9	SER	-	expression tag	UNP A0R3M0
A	-8	HIS	-	expression tag	UNP A0R3M0
A	-7	HIS	-	expression tag	UNP A0R3M0
A	-6	HIS	-	expression tag	UNP A0R3M0
A	-5	HIS	-	expression tag	UNP A0R3M0
A	-4	HIS	-	expression tag	UNP A0R3M0
A	-3	HIS	-	expression tag	UNP A0R3M0
A	-2	MSE	-	expression tag	UNP A0R3M0
A	-1	ALA	-	expression tag	UNP A0R3M0
A	0	SER	-	expression tag	UNP A0R3M0
B	-12	MSE	-	expression tag	UNP A0R3M0
B	-11	ARG	-	expression tag	UNP A0R3M0
B	-10	GLY	-	expression tag	UNP A0R3M0
B	-9	SER	-	expression tag	UNP A0R3M0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-8	HIS	-	expression tag	UNP A0R3M0
B	-7	HIS	-	expression tag	UNP A0R3M0
B	-6	HIS	-	expression tag	UNP A0R3M0
B	-5	HIS	-	expression tag	UNP A0R3M0
B	-4	HIS	-	expression tag	UNP A0R3M0
B	-3	HIS	-	expression tag	UNP A0R3M0
B	-2	MSE	-	expression tag	UNP A0R3M0
B	-1	ALA	-	expression tag	UNP A0R3M0
B	0	SER	-	expression tag	UNP A0R3M0
C	-12	MSE	-	expression tag	UNP A0R3M0
C	-11	ARG	-	expression tag	UNP A0R3M0
C	-10	GLY	-	expression tag	UNP A0R3M0
C	-9	SER	-	expression tag	UNP A0R3M0
C	-8	HIS	-	expression tag	UNP A0R3M0
C	-7	HIS	-	expression tag	UNP A0R3M0
C	-6	HIS	-	expression tag	UNP A0R3M0
C	-5	HIS	-	expression tag	UNP A0R3M0
C	-4	HIS	-	expression tag	UNP A0R3M0
C	-3	HIS	-	expression tag	UNP A0R3M0
C	-2	MSE	-	expression tag	UNP A0R3M0
C	-1	ALA	-	expression tag	UNP A0R3M0
C	0	SER	-	expression tag	UNP A0R3M0
D	-12	MSE	-	expression tag	UNP A0R3M0
D	-11	ARG	-	expression tag	UNP A0R3M0
D	-10	GLY	-	expression tag	UNP A0R3M0
D	-9	SER	-	expression tag	UNP A0R3M0
D	-8	HIS	-	expression tag	UNP A0R3M0
D	-7	HIS	-	expression tag	UNP A0R3M0
D	-6	HIS	-	expression tag	UNP A0R3M0
D	-5	HIS	-	expression tag	UNP A0R3M0
D	-4	HIS	-	expression tag	UNP A0R3M0
D	-3	HIS	-	expression tag	UNP A0R3M0
D	-2	MSE	-	expression tag	UNP A0R3M0
D	-1	ALA	-	expression tag	UNP A0R3M0
D	0	SER	-	expression tag	UNP A0R3M0
E	-12	MSE	-	expression tag	UNP A0R3M0
E	-11	ARG	-	expression tag	UNP A0R3M0
E	-10	GLY	-	expression tag	UNP A0R3M0
E	-9	SER	-	expression tag	UNP A0R3M0
E	-8	HIS	-	expression tag	UNP A0R3M0
E	-7	HIS	-	expression tag	UNP A0R3M0
E	-6	HIS	-	expression tag	UNP A0R3M0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-5	HIS	-	expression tag	UNP A0R3M0
E	-4	HIS	-	expression tag	UNP A0R3M0
E	-3	HIS	-	expression tag	UNP A0R3M0
E	-2	MSE	-	expression tag	UNP A0R3M0
E	-1	ALA	-	expression tag	UNP A0R3M0
E	0	SER	-	expression tag	UNP A0R3M0
F	-12	MSE	-	expression tag	UNP A0R3M0
F	-11	ARG	-	expression tag	UNP A0R3M0
F	-10	GLY	-	expression tag	UNP A0R3M0
F	-9	SER	-	expression tag	UNP A0R3M0
F	-8	HIS	-	expression tag	UNP A0R3M0
F	-7	HIS	-	expression tag	UNP A0R3M0
F	-6	HIS	-	expression tag	UNP A0R3M0
F	-5	HIS	-	expression tag	UNP A0R3M0
F	-4	HIS	-	expression tag	UNP A0R3M0
F	-3	HIS	-	expression tag	UNP A0R3M0
F	-2	MSE	-	expression tag	UNP A0R3M0
F	-1	ALA	-	expression tag	UNP A0R3M0
F	0	SER	-	expression tag	UNP A0R3M0

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	36	Total 36	O 36	0	0
2	B	29	Total 29	O 29	0	0
2	C	37	Total 37	O 37	0	0
2	D	18	Total 18	O 18	0	0
2	E	25	Total 25	O 25	0	0
2	F	15	Total 15	O 15	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.
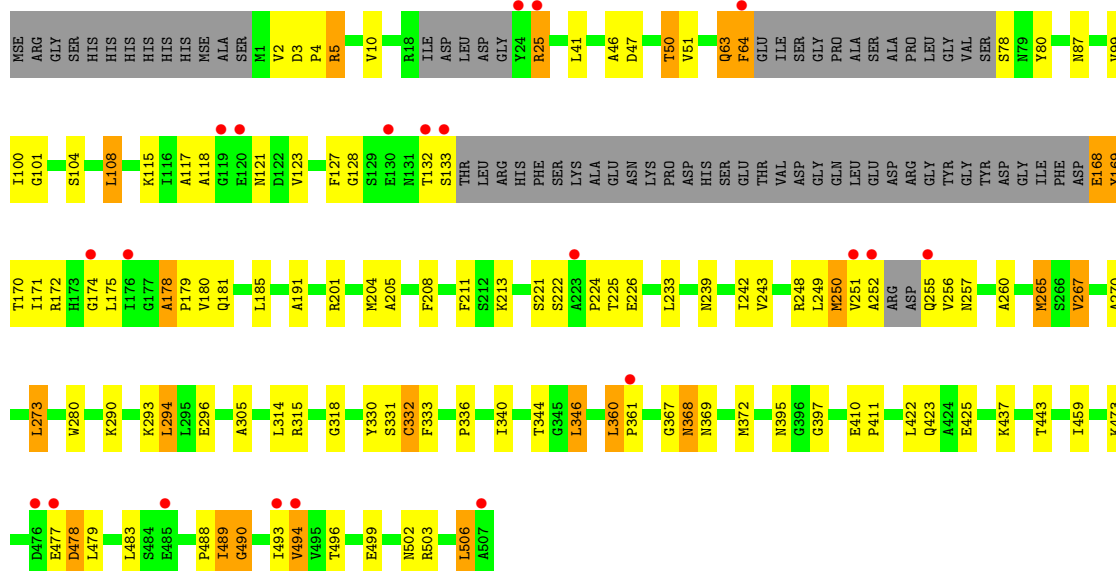
• Molecule 1: Acetyl-CoA acetyltransferase
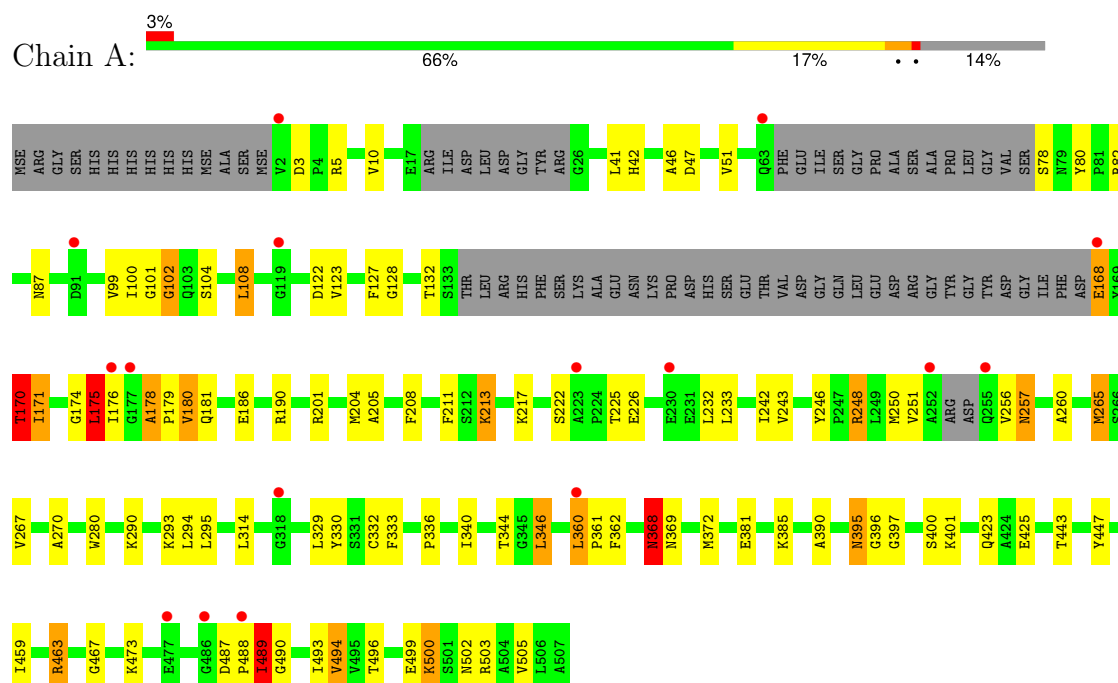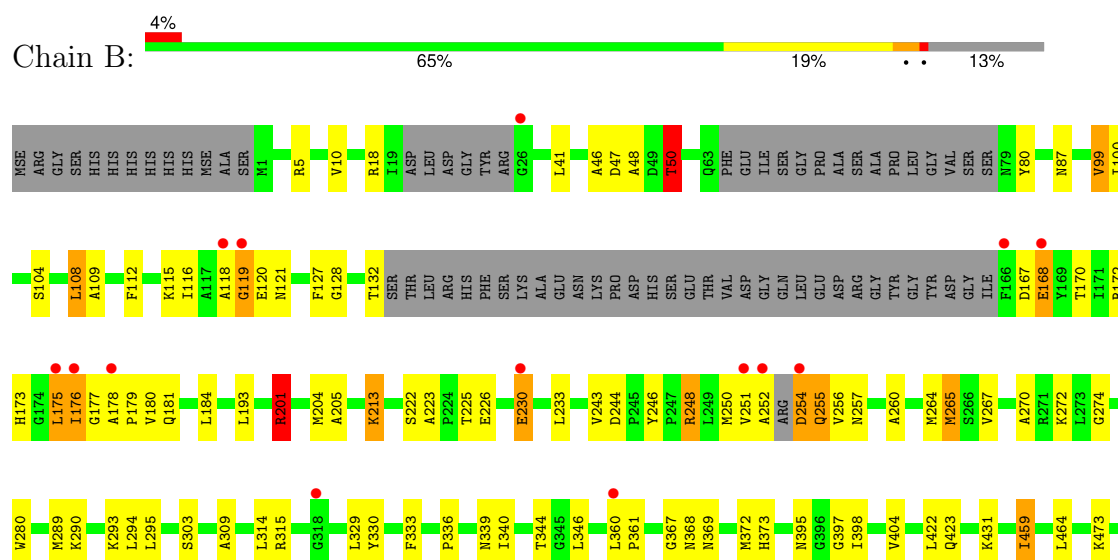
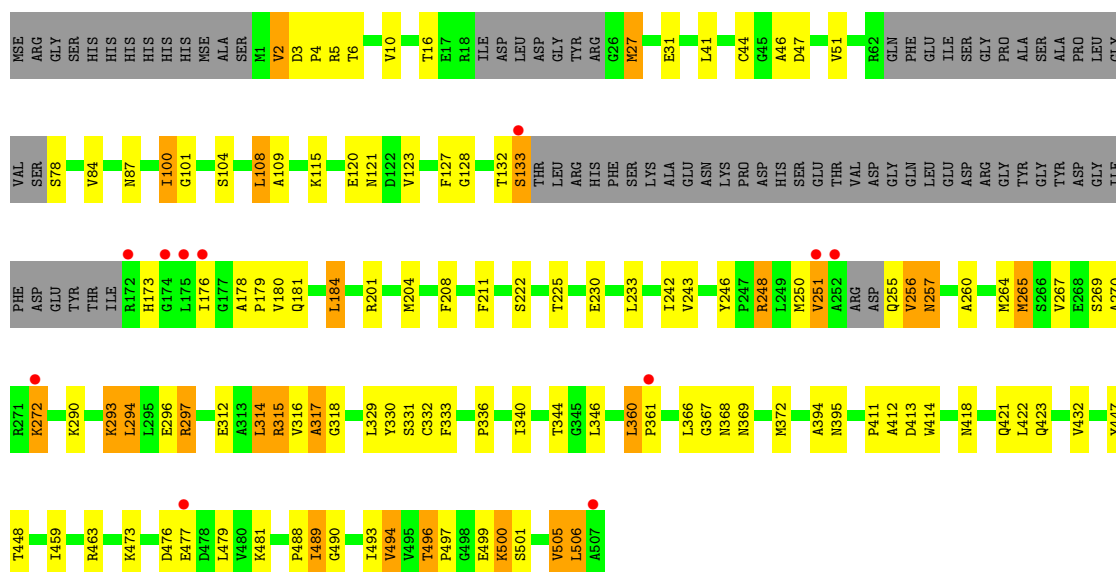• Molecule 1: Acetyl-CoA acetyltransferase
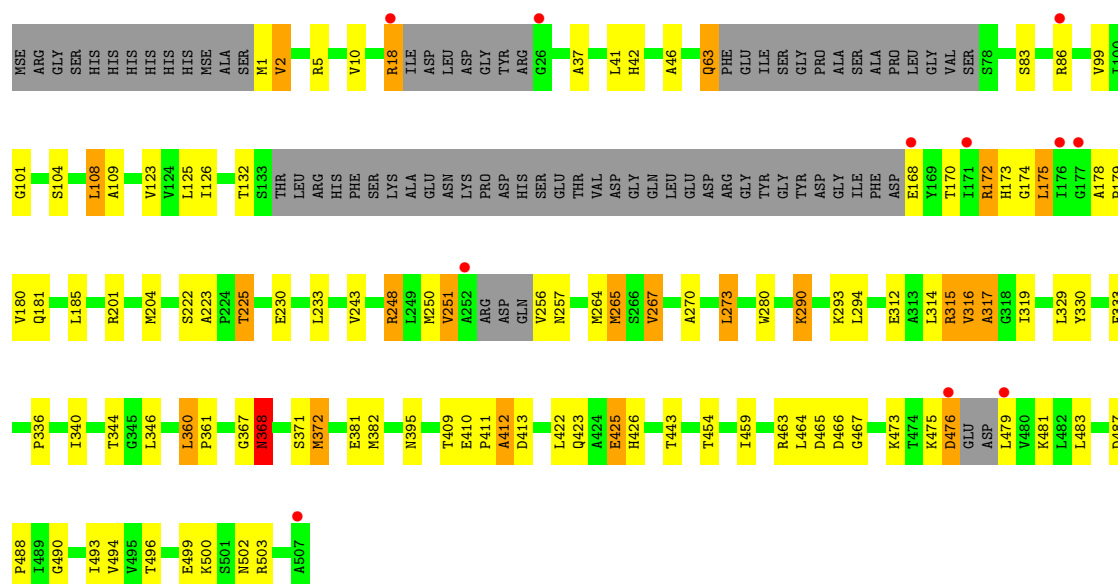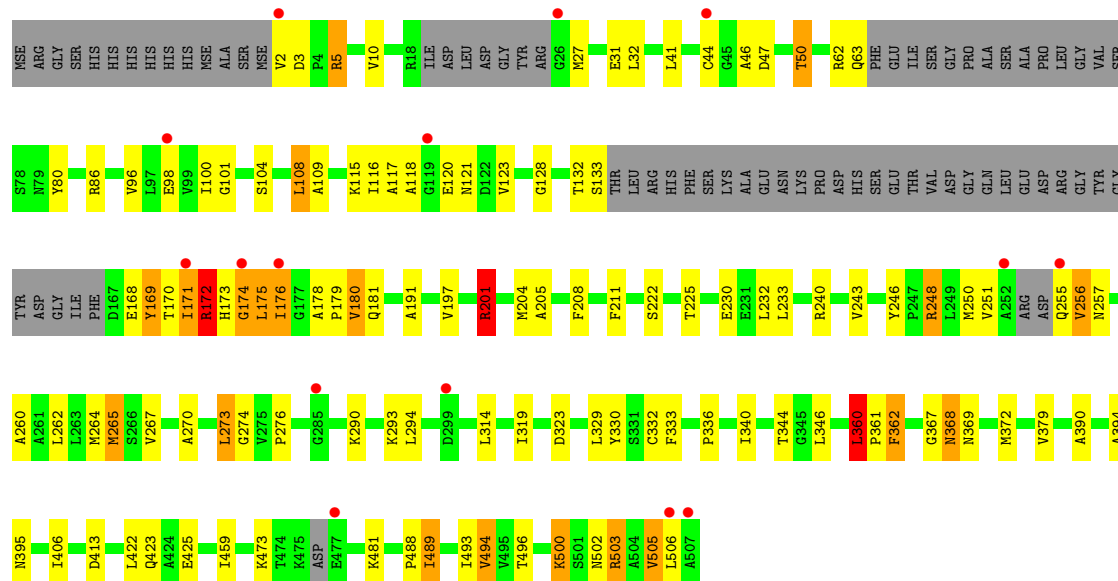
• Molecule 1: Acetyl-CoA acetyltransferase

Chain C:

• Molecule 1: Acetyl-CoA acetyltransferase

Chain D:

• Molecule 1: Acetyl-CoA acetyltransferase

Chain E:

• Molecule 1: Acetyl-CoA acetyltransferase

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	96.31Å 102.00Å 102.42Å 116.46° 101.28° 97.44°	Depositor
Resolution (Å)	47.95 – 2.71 47.95 – 2.71	Depositor EDS
% Data completeness (in resolution range)	96.9 (47.95-2.71) 97.1 (47.95-2.71)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.27 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.234 , 0.261 0.244 , 0.271	Depositor DCC
R_{free} test set	4512 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	49.0	Xtriage
Anisotropy	0.038	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 36.6	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.018 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	20206	wwPDB-VP
Average B, all atoms (Å^2)	46.0	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	1.29	15/3377 (0.4%)	1.17	13/4572 (0.3%)
1	B	1.24	8/3435 (0.2%)	1.18	18/4649 (0.4%)
1	C	1.23	7/3424 (0.2%)	1.17	15/4634 (0.3%)
1	D	1.19	5/3348 (0.1%)	1.14	12/4532 (0.3%)
1	E	1.21	4/3388 (0.1%)	1.17	9/4582 (0.2%)
1	F	1.18	7/3384 (0.2%)	1.13	11/4580 (0.2%)
All	All	1.23	46/20356 (0.2%)	1.16	78/27549 (0.3%)

The worst 5 of 46 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	294	LEU	C-O	-9.13	1.13	1.24
1	F	294	LEU	C-O	-9.11	1.13	1.24
1	A	294	LEU	C-O	-8.54	1.14	1.24
1	C	99	VAL	CA-CB	8.37	1.62	1.54
1	C	294	LEU	C-O	-7.68	1.15	1.24

The worst 5 of 78 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	317	ALA	N-CA-C	-11.02	96.12	110.33
1	D	173	HIS	N-CA-C	-9.51	96.38	110.46
1	D	120	GLU	N-CA-C	9.13	121.23	111.28
1	F	169	TYR	N-CA-C	-8.28	101.90	114.16
1	E	316	VAL	N-CA-C	-8.00	103.10	111.58

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	3327	0	3278	79	0
1	B	3376	0	3320	84	0
1	C	3373	0	3329	85	0
1	D	3299	0	3252	92	0
1	E	3336	0	3296	95	0
1	F	3335	0	3283	85	0
2	A	36	0	0	3	0
2	B	29	0	0	2	0
2	C	37	0	0	3	0
2	D	18	0	0	2	0
2	E	25	0	0	3	0
2	F	15	0	0	1	0
All	All	20206	0	19758	505	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 13.

The worst 5 of 505 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:176:ILE:HD13	1:B:176:ILE:C	1.51	1.30
1:F:176:ILE:HD12	1:F:176:ILE:O	1.34	1.22
1:A:489:ILE:HD12	1:A:489:ILE:N	1.57	1.13
1:A:489:ILE:H	1:A:489:ILE:CD1	1.50	1.12
1:A:175:LEU:N	1:A:175:LEU:HD23	1.58	1.10

There are no symmetry-related clashes.

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	438/520 (84%)	417 (95%)	20 (5%)	1 (0%)	44	67
1	B	445/520 (86%)	425 (96%)	19 (4%)	1 (0%)	44	67
1	C	443/520 (85%)	423 (96%)	20 (4%)	0	100	100
1	D	435/520 (84%)	414 (95%)	21 (5%)	0	100	100
1	E	437/520 (84%)	420 (96%)	15 (3%)	2 (0%)	25	47
1	F	437/520 (84%)	420 (96%)	17 (4%)	0	100	100
All	All	2635/3120 (84%)	2519 (96%)	112 (4%)	4 (0%)	44	67

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	119	GLY
1	E	412	ALA
1	E	466	ASP
1	A	489	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	343/394 (87%)	307 (90%)	36 (10%)	5	13
1	B	347/394 (88%)	304 (88%)	43 (12%)	4	8
1	C	348/394 (88%)	311 (89%)	37 (11%)	5	13
1	D	340/394 (86%)	299 (88%)	41 (12%)	4	9
1	E	344/394 (87%)	304 (88%)	40 (12%)	4	10
1	F	344/394 (87%)	305 (89%)	39 (11%)	4	11
All	All	2066/2364 (87%)	1830 (89%)	236 (11%)	4	11

5 of 236 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	506	LEU
1	F	360	LEU
1	D	432	VAL
1	F	314	LEU
1	F	32	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	368	ASN
1	F	14	GLN
1	D	388	GLN
1	E	257	ASN
1	F	173	HIS

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 oligosaccharides 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	439/520 (84%)	0.44	16 (3%)	46	45	23, 46, 64, 69	0
1	B	442/520 (85%)	0.41	19 (4%)	40	39	25, 46, 64, 69	3 (0%)
1	C	443/520 (85%)	0.43	21 (4%)	37	36	25, 46, 64, 69	0
1	D	435/520 (83%)	0.29	11 (2%)	58	57	25, 46, 64, 69	0
1	E	437/520 (84%)	0.31	11 (2%)	58	57	18, 46, 64, 69	2 (0%)
1	F	440/520 (84%)	0.37	15 (3%)	48	47	25, 46, 64, 76	0
All	All	2636/3120 (84%)	0.38	93 (3%)	47	46	18, 46, 64, 76	5 (0%)

The worst 5 of 93 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	119	GLY	8.5
1	C	119	GLY	7.1
1	C	176	ILE	5.5
1	D	252	ALA	5.1
1	E	252	ALA	4.8

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 ⓘ

There are no ligands in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.5 Other polymers (i)

There are no such residues in this entry.