



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 09:08 PM UTC

PDB ID : 8G4U / pdb_00008g4u
Title : Final ketosynthase+acyltransferase of the erythromycin modular polyketide synthase
Authors : Keatinge-Clay, A.T.
Deposited on : 2023-02-10
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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

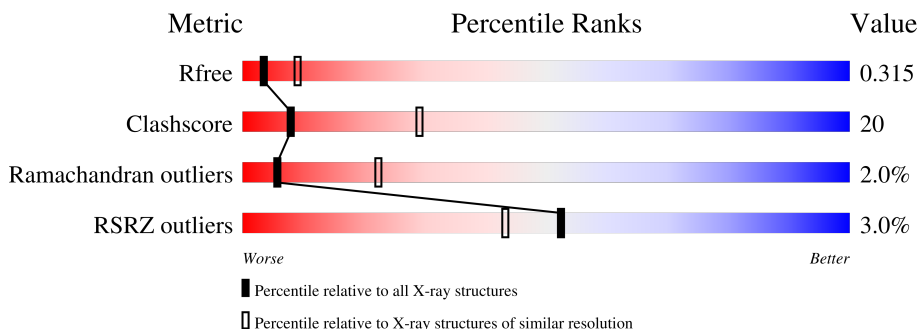
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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}	180053	2481 (2.90-2.90)
Clashscore	190562	2690 (2.90-2.90)
Ramachandran outliers	187476	2623 (2.90-2.90)
RSRZ outliers	180081	2481 (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 $\leq 5\%$. 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	881	2% 57% 34% 7%
1	B	881	2% 52% 35% 11%
1	C	881	5% 54% 30% 15%
1	D	881	2% 59% 33% 7%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 22987 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 EryAIII.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	818	Total	C	N	O	S	0	0	0
			5950	3723	1077	1131	19			
1	B	786	Total	C	N	O	S	0	0	0
			5696	3567	1032	1079	18			
1	C	750	Total	C	N	O	S	0	0	0
			5423	3396	990	1021	16			
1	D	815	Total	C	N	O	S	0	0	0
			5910	3696	1072	1123	19			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1480	GLY	-	expression tag	UNP Q5UNP4
A	1481	SER	-	expression tag	UNP Q5UNP4
A	1482	HIS	-	expression tag	UNP Q5UNP4
A	1483	MET	-	expression tag	UNP Q5UNP4
A	2357	GLN	ALA	engineered mutation	UNP Q5UNP4
B	1480	GLY	-	expression tag	UNP Q5UNP4
B	1481	SER	-	expression tag	UNP Q5UNP4
B	1482	HIS	-	expression tag	UNP Q5UNP4
B	1483	MET	-	expression tag	UNP Q5UNP4
B	2357	GLN	ALA	engineered mutation	UNP Q5UNP4
C	1480	GLY	-	expression tag	UNP Q5UNP4
C	1481	SER	-	expression tag	UNP Q5UNP4
C	1482	HIS	-	expression tag	UNP Q5UNP4
C	1483	MET	-	expression tag	UNP Q5UNP4
C	2357	GLN	ALA	engineered mutation	UNP Q5UNP4
D	1480	GLY	-	expression tag	UNP Q5UNP4
D	1481	SER	-	expression tag	UNP Q5UNP4
D	1482	HIS	-	expression tag	UNP Q5UNP4
D	1483	MET	-	expression tag	UNP Q5UNP4
D	2357	GLN	ALA	engineered mutation	UNP Q5UNP4

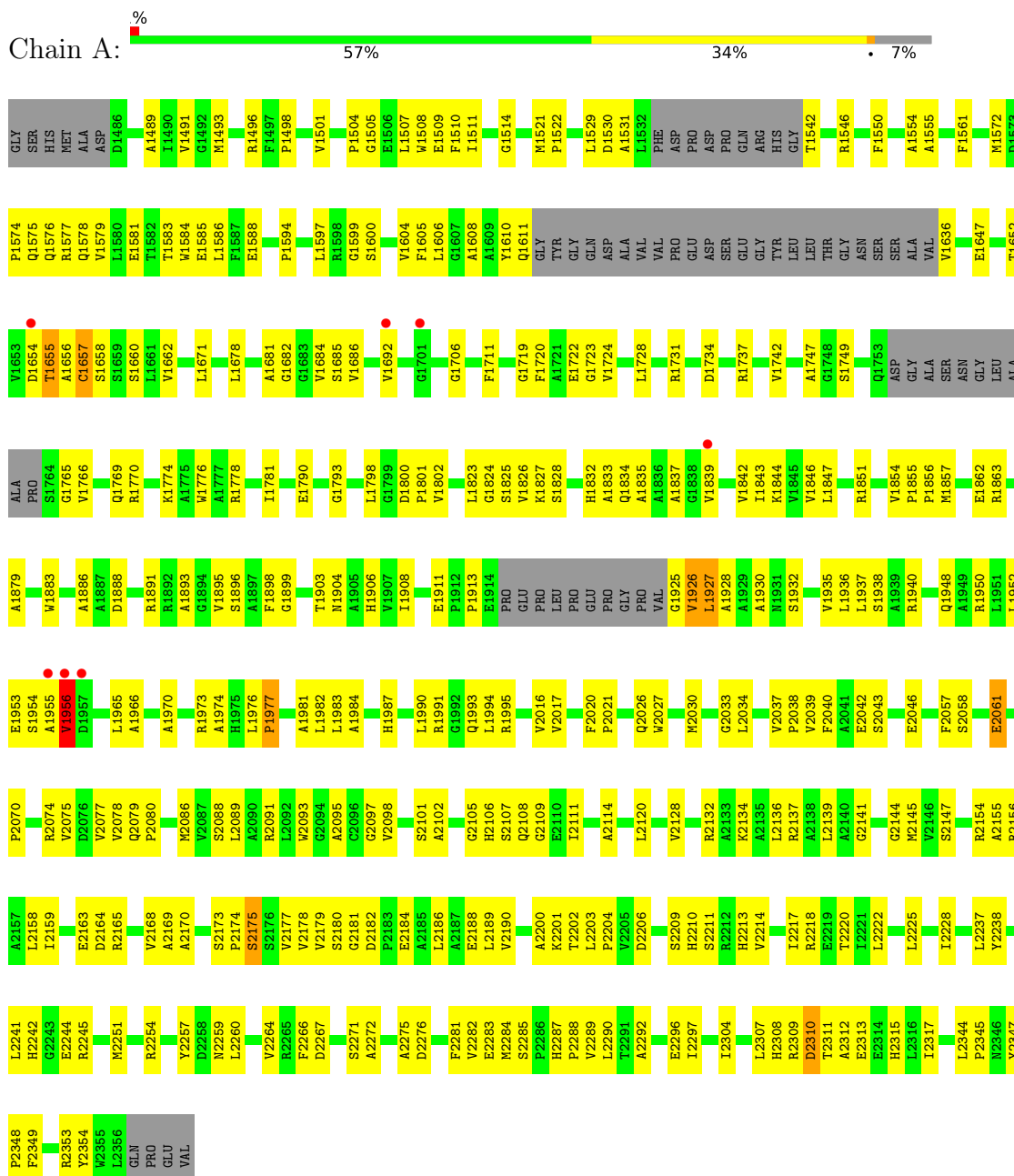
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	O 1	0	0
2	B	3	Total 3	O 3	0	0
2	C	3	Total 3	O 3	0	0
2	D	1	Total 1	O 1	0	0

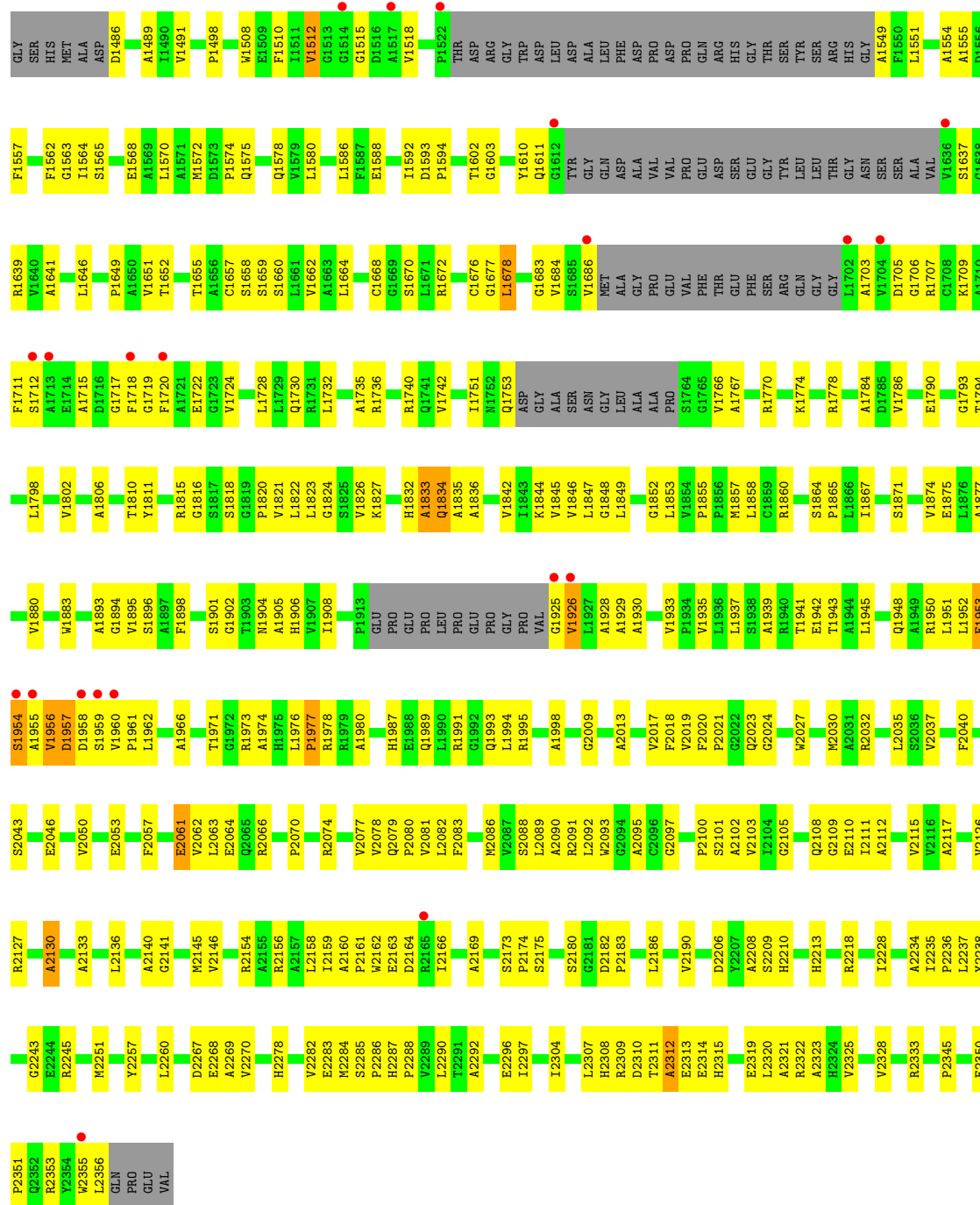
3 Residue-property plots

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: EryAIII

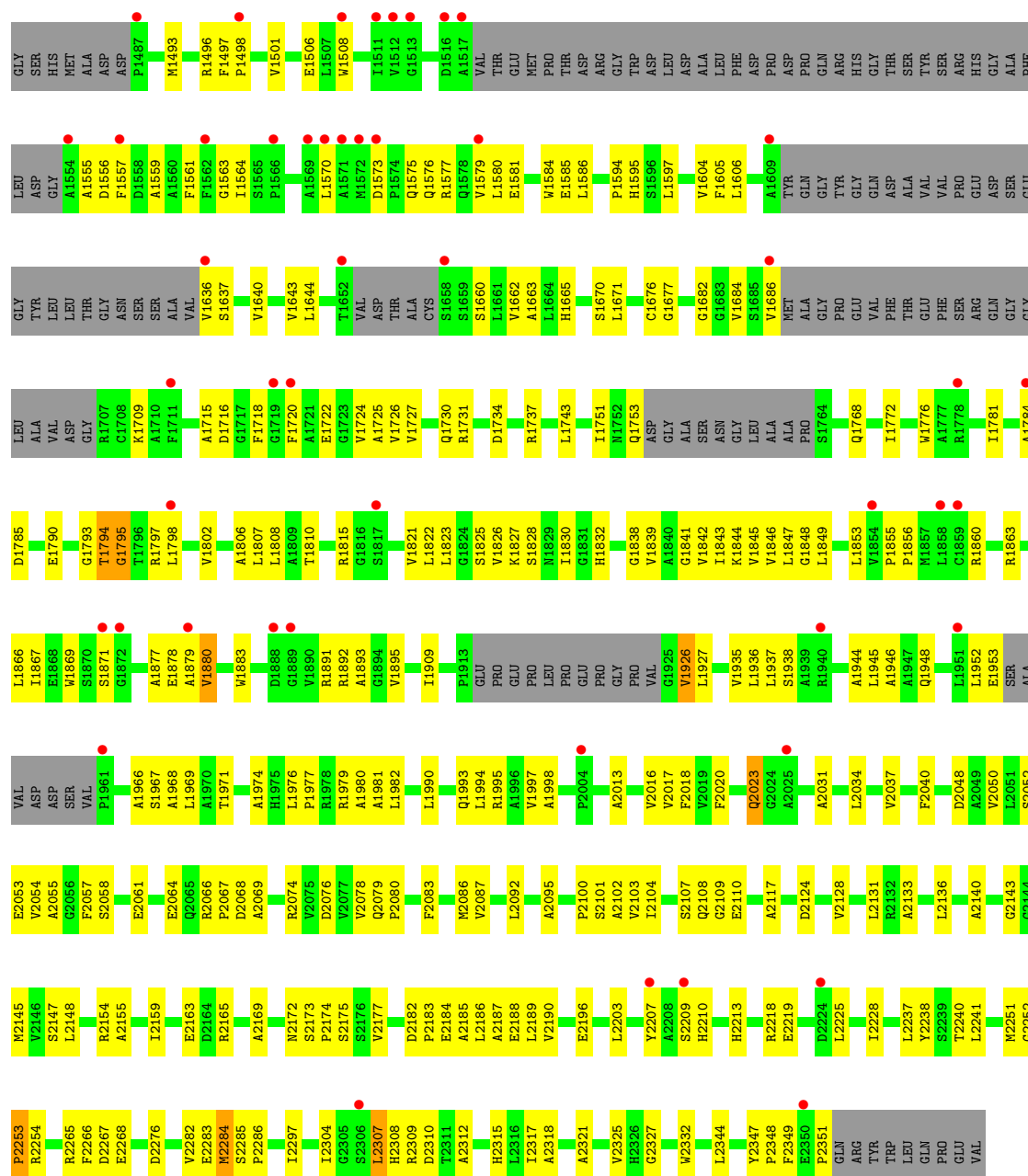


• Molecule 1: EryAIII

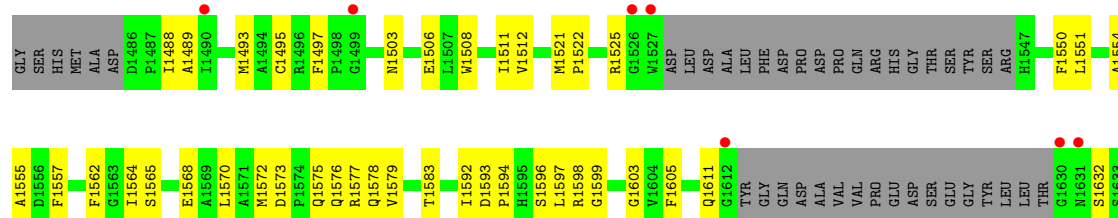


• Molecule 1: EryAIII





• Molecule 1: EryAIII





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	71.21Å 167.94Å 184.23Å 90.00° 99.43° 90.00°	Depositor
Resolution (Å)	76.23 – 2.90 76.23 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.2 (76.23-2.90) 85.5 (76.23-2.90)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.53 (at 2.77Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.239 , 0.314 0.239 , 0.315	Depositor DCC
R_{free} test set	2006 reflections (1.84%)	wwPDB-VP
Wilson B-factor (Å ²)	41.5	Xtriage
Anisotropy	0.649	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 57.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.035 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	22987	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 39.64 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.0818e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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](#)

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	0.49	0/6065	0.73	2/8264 (0.0%)
1	B	0.50	0/5803	0.74	0/7908
1	C	0.40	0/5522	0.63	0/7517
1	D	0.40	0/6024	0.65	0/8208
All	All	0.45	0/23414	0.69	2/31897 (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	A	0	2
1	B	0	2
1	D	0	1
All	All	0	5

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1956	VAL	CA-C-N	-6.54	112.42	122.62
1	A	1956	VAL	C-N-CA	-6.54	112.42	122.62

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1657	CYS	Peptide
1	A	1954	SER	Peptide
1	B	1953	GLU	Peptide
1	B	1956	VAL	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	D	1957	ASP	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	5950	0	5885	226	0
1	B	5696	0	5658	238	0
1	C	5423	0	5410	221	1
1	D	5910	0	5845	227	1
2	A	1	0	0	0	0
2	B	3	0	0	0	0
2	C	3	0	0	0	0
2	D	1	0	0	0	0
All	All	22987	0	22798	903	1

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 903 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:1934:PRO:HD3	1:D:2317:ILE:HD11	1.42	1.00
1:B:1926:VAL:HG11	1:B:2037:VAL:HG22	1.48	0.96
1:B:1824:GLY:HA2	1:B:1857:MET:HE2	1.48	0.94
1:D:1687:MET:HE1	1:D:1720:PHE:HB3	1.49	0.94
1:D:1824:GLY:HA2	1:D:1857:MET:HE2	1.52	0.91

All (1) 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:2054:VAL:O	1:D:1773:ARG:NH1[2_656]	1.85	0.35

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	808/881 (92%)	725 (90%)	68 (8%)	15 (2%)	6	23
1	B	774/881 (88%)	687 (89%)	67 (9%)	20 (3%)	4	17
1	C	734/881 (83%)	666 (91%)	54 (7%)	14 (2%)	6	23
1	D	805/881 (91%)	733 (91%)	59 (7%)	13 (2%)	7	27
All	All	3121/3524 (89%)	2811 (90%)	248 (8%)	62 (2%)	6	22

5 of 62 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1656	ALA
1	A	1913	PRO
1	A	1926	VAL
1	A	2026	GLN
1	A	2310	ASP

5.3.2 Protein sidechains [i](#)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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 oligosaccharides 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 [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	818/881 (92%)	0.28	7 (0%) 81 75	25, 48, 77, 99	0
1	B	786/881 (89%)	0.31	21 (2%) 56 47	24, 49, 81, 98	0
1	C	750/881 (85%)	0.86	48 (6%) 25 20	41, 74, 103, 119	0
1	D	815/881 (92%)	0.53	18 (2%) 62 53	40, 63, 85, 109	0
All	All	3169/3524 (89%)	0.49	94 (2%) 52 43	24, 59, 91, 119	0

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1511	ILE	6.3
1	A	1956	VAL	5.4
1	C	1720	PHE	5.4
1	C	1570	LEU	4.5
1	C	1569	ALA	4.3

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 oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.