wwPDB X-ray Structure Validation Summary Report (i)

Feb 21, 2023 – 04:05 PM EST

PDB ID	:	8G4U
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 (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:

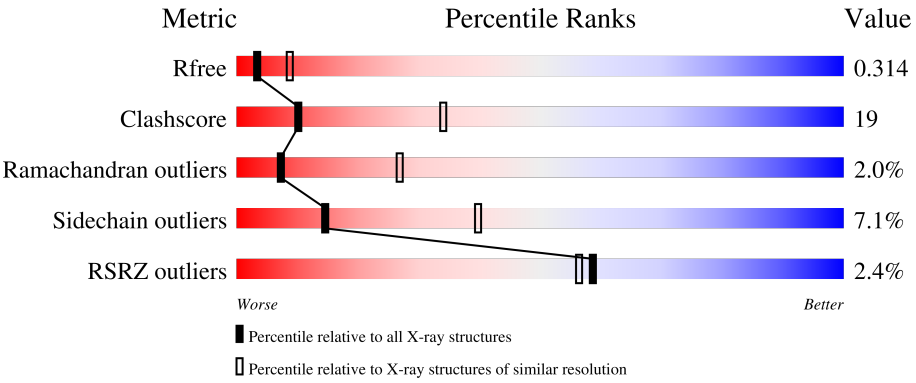
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.32.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

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	881	56% 34% · 7%
1	B	881	51% 34% 5% 11%
1	C	881	4% 52% 31% · 15%
1	D	881	2% 57% 32% · 7%

2 Entry composition (i)

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.

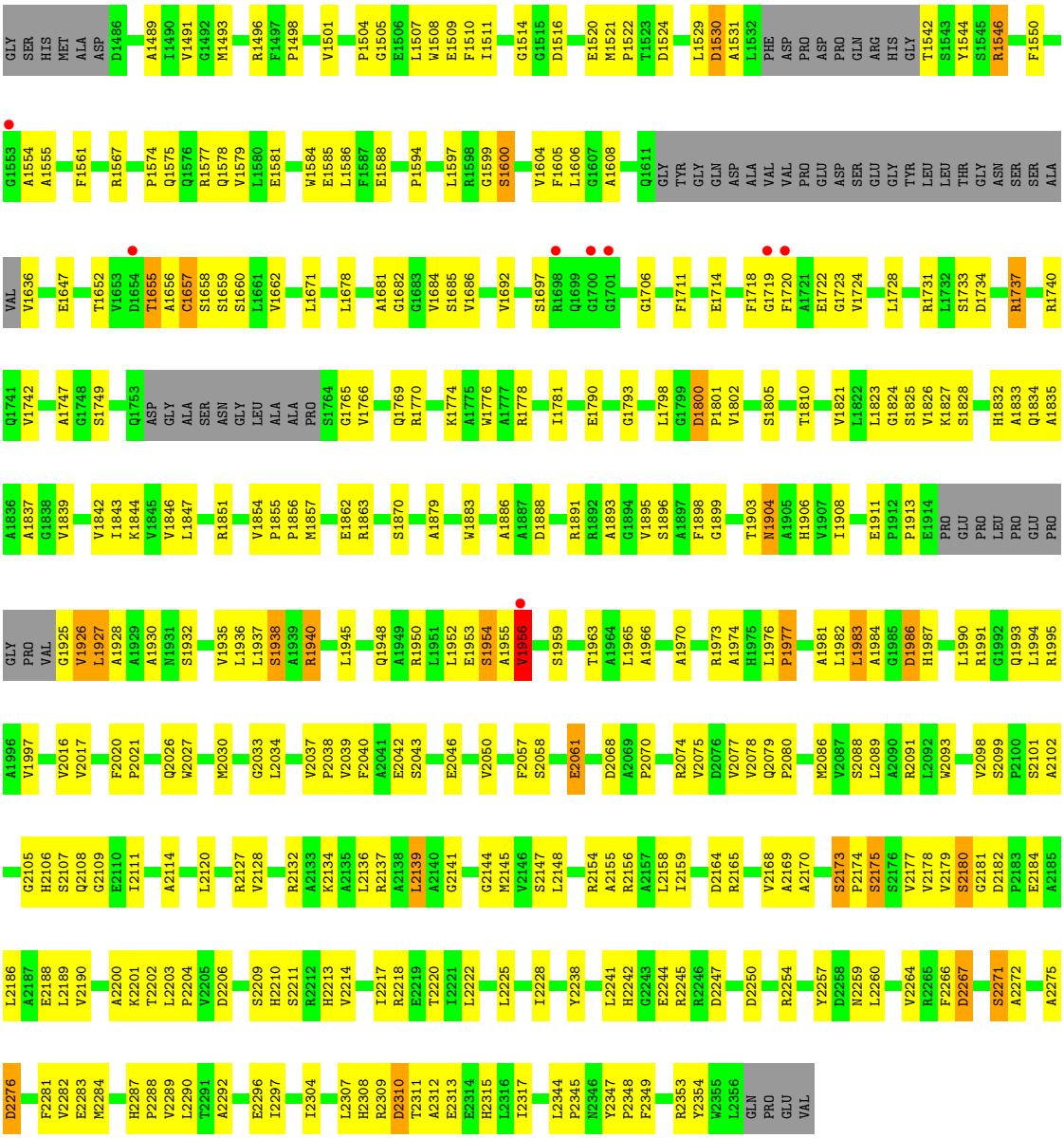
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O 1 1	0	0
2	B	3	Total O 3 3	0	0
2	C	3	Total O 3 3	0	0
2	D	1	Total O 1 1	0	0

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.

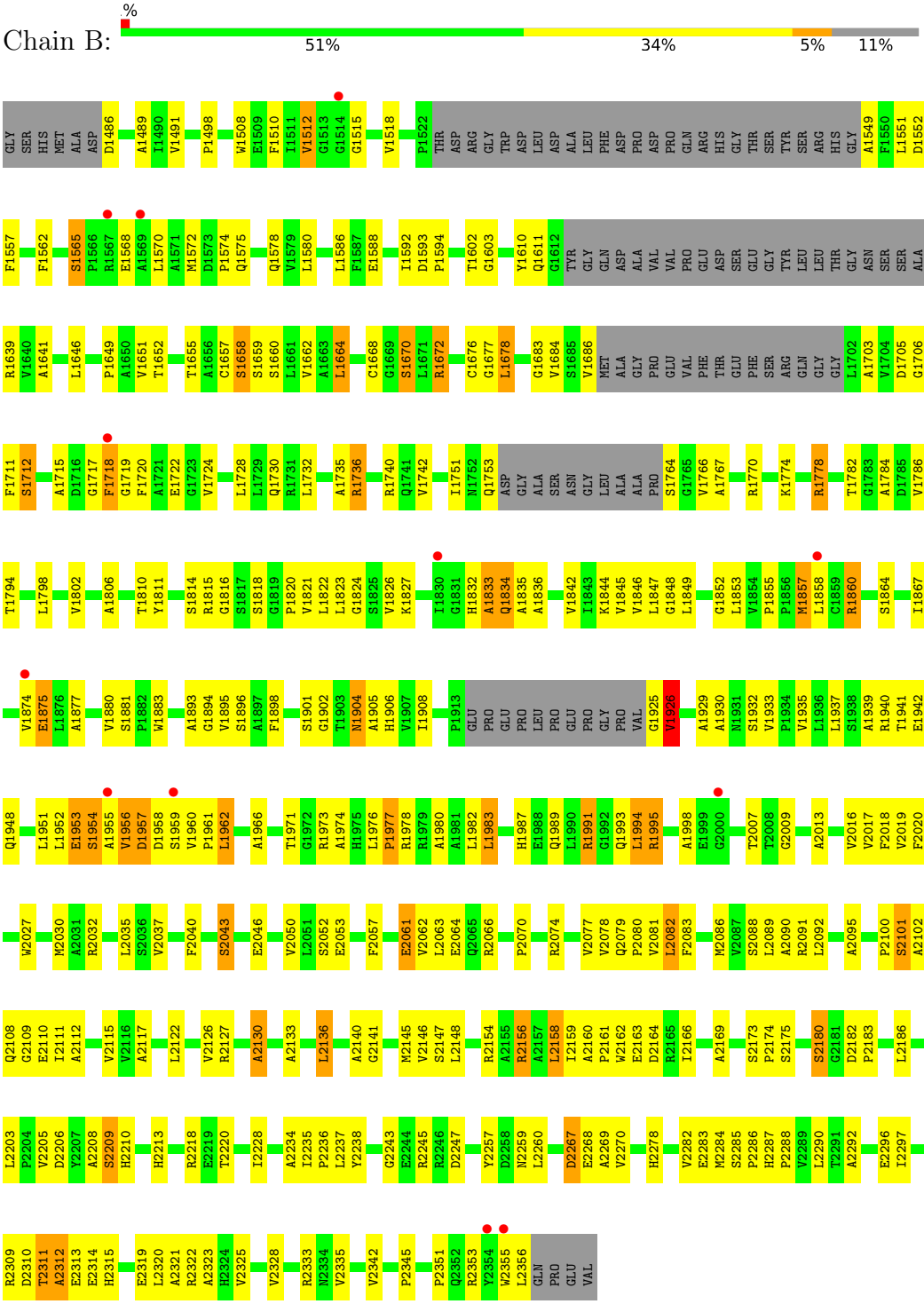
- Molecule 1: EryAIII

Chain A:



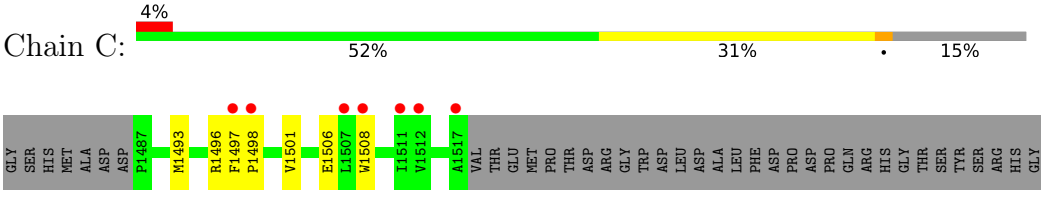
• Molecule 1: EryAIII

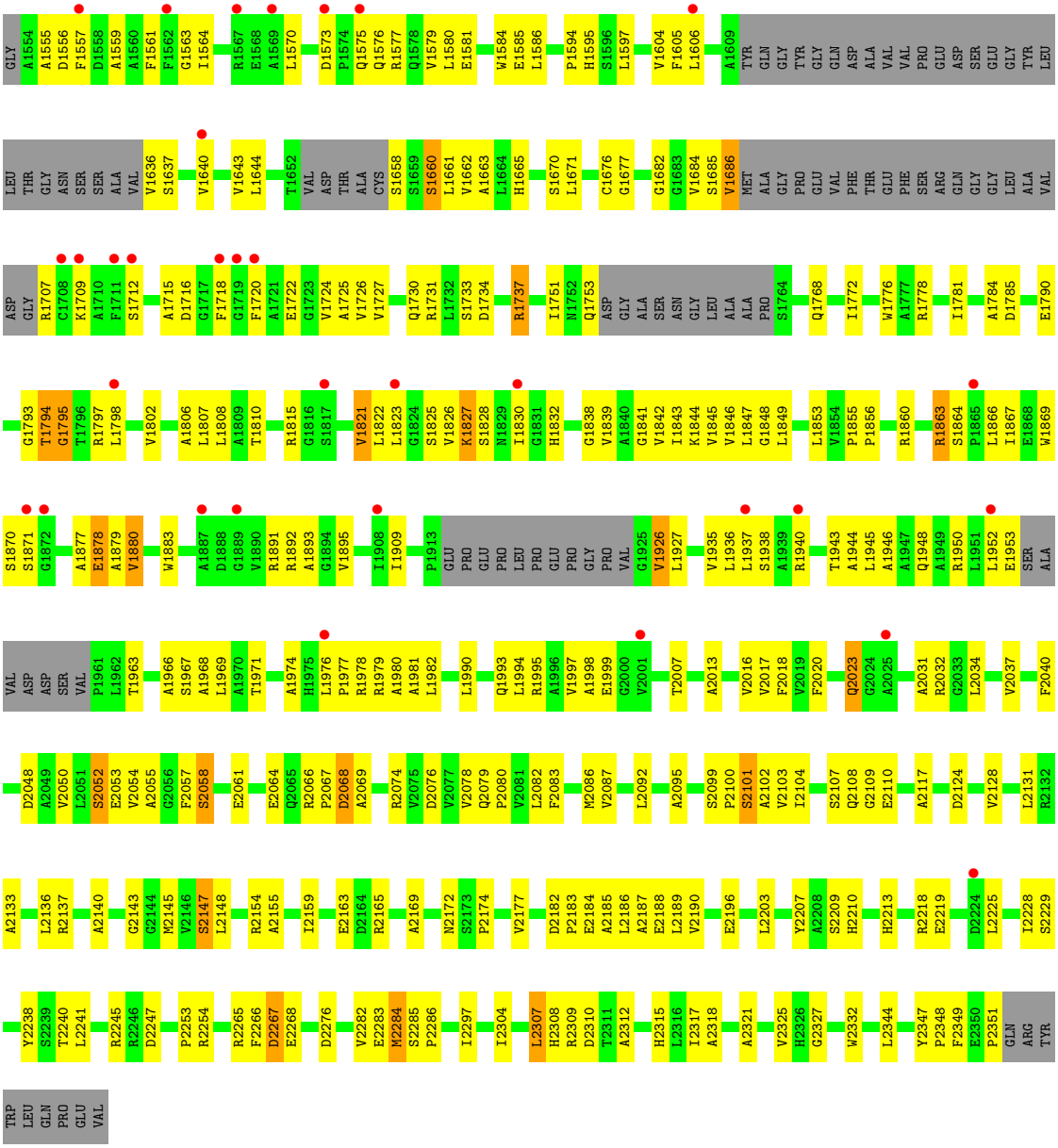
Chain B:



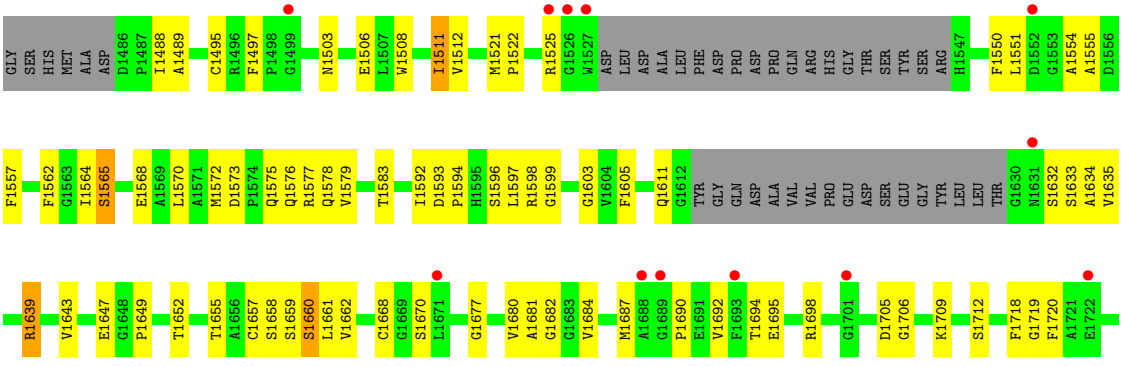
• Molecule 1: EryAIII

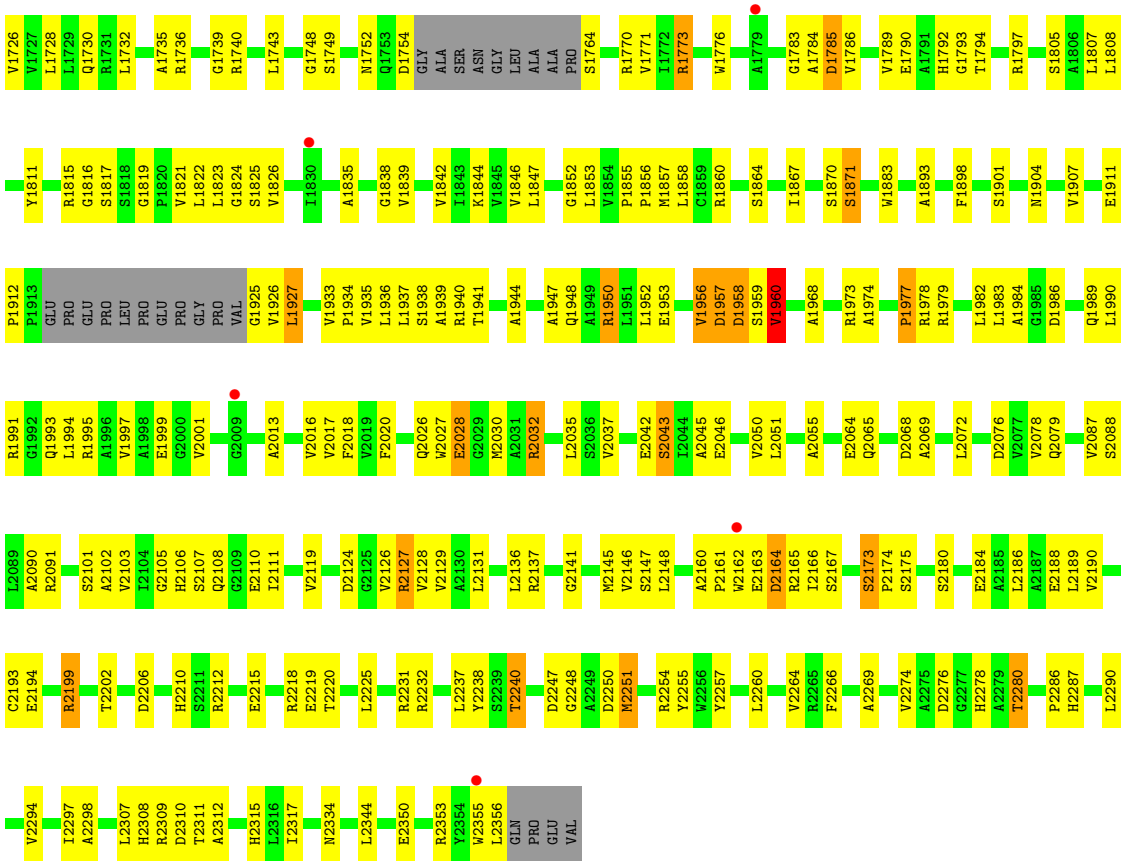
Chain C:





• Molecule 1: EryAIII




V1726 V1727 L1728 L1729 Q1730 R1731 L1732 A1735 R1736 G1739 R1740 L1743 G1748 S1749 N1752 Q1753 D1754 GLY ALA SER ASN GLY LEU ALA ALA PRO S1764 R1770 V1771 I1772 R1773 W1776 A1779 G1783 A1784 D1785 V1786 V1789 E1790 A1791 H1792 G1793 T1794 R1797 S1805 A1806 L1807 L1808
Y1811 R1815 G1816 S1817 S1818 G1819 P1820 V1821 L1822 L1823 G1824 S1825 V1826 I1830 A1835 G1838 V1839 V1842 I1843 K1844 V1845 V1846 L1847 G1852 L1853 V1854 P1855 P1856 M1857 L1858 C1859 R1860 S1864 I1867 S1870 S1871 W1883 A1893 F1898 S1901 N1904 V1907 E1911
P1912 P1913 GLU PRO GLU PRO LEU PRO GLU PRO GLY PRO VAL G1925 V1926 L1927 V1933 P1934 V1935 L1936 L1937 S1938 A1939 R1940 T1941 A1944 A1947 Q1948 A1949 R1950 L1951 L1952 E1953 V1956 D1957 D1958 S1959 V1960 A1968 R1973 A1974 P1977 R1978 R1979 L1982 L1983 A1984 G1985 D1986 Q1989 L1990
R1991 G1992 Q1993 L1994 R1995 A1996 V1997 A1998 E1999 G2000 V2001 G2009 A2013 V2016 V2017 F2018 V2019 F2020 Q2026 W2027 E2028 G2029 M2030 A2031 R2032 L2035 S2036 V2037 E2042 S2043 I2044 A2045 E2046 V2050 L2051 A2055 E2064 Q2065 D2068 A2069 L2072 D2076 V2077 V2078 Q2079 V2087 S2088
L2089 A2090 R2091 S2101 A2102 V2103 I2104 G2105 H2106 S2107 Q2108 G2109 E2110 I2111 V2119 D2124 G2125 V2126 R2127 V2128 V2129 A2130 L2131 L2136 R2137 G2141 M2145 V2146 S2147 L2148 A2160 P2161 W2162 E2163 D2164 R2165 I2166 S2167 S2173 P2174 S2175 S2180 E2184 A2185 L2186 A2187 E2188 L2189 V2190
C2193 E2194 R2199 T2202 D2206 H2210 S2211 R2212 E2215 R2218 E2219 T2220 L2225 R2231 R2232 L2237 Y2238 S2239 T2240 D2247 G2248 A2249 D2250 M2251 R2254 Y2255 W2256 Y2257 L2260 V2264 R2265 F2266 A2269 V2274 A2275 D2276 G2277 H2278 A2279 T2280 P2286 H2287 L2290
V2294 I2297 A2298 L2307 H2308 R2309 D2310 T2311 A2312 H2315 L2316 I2317 N2334 L2344 E2350 R2353 Y2354 W2355 L2356 GLN PRO GLU VAL

4 Data and refinement statistics (i)

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 – 2.76	Depositor EDS
% Data completeness (in resolution range)	99.2 (76.23-2.90) 82.7 (76.23-2.76)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.53 (at 2.77Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.239 , 0.314 0.239 , 0.314	Depositor DCC
R_{free} test set	2006 reflections (1.85%)	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 , 59.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.035 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.90	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.*

[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	0.53	0/6065	0.77	7/8264 (0.1%)
1	B	0.54	0/5803	0.77	8/7908 (0.1%)
1	C	0.44	0/5522	0.70	0/7517
1	D	0.45	0/6024	0.70	0/8208
All	All	0.49	0/23414	0.74	15/31897 (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.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	2148	LEU	CA-CB-CG	6.56	130.39	115.30
1	A	1986	ASP	CB-CG-OD2	6.48	124.13	118.30
1	B	2335	VAL	C-N-CA	6.14	137.05	121.70
1	A	2139	LEU	CA-CB-CG	6.09	129.31	115.30
1	A	1520	GLU	C-N-CA	-6.07	106.53	121.70

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1657	CYS	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	1954	SER	Peptide
1	B	1953	GLU	Peptide
1	B	1956	VAL	Peptide
1	D	1957	ASP	Peptide

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	5950	0	5885	217	0
1	B	5696	0	5658	227	0
1	C	5423	0	5410	214	1
1	D	5910	0	5845	215	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	865	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 19.

The worst 5 of 865 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:C:2308:HIS:ND1	1:C:2309:ARG:O	2.05	0.89
1:B:1826:VAL:HB	1:B:1844:LYS:HD2	1.53	0.88
1:C:2064:GLU:OE1	1:C:2066:ARG:NH1	2.06	0.88

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%)	8 28
1	B	774/881 (88%)	687 (89%)	67 (9%)	20 (3%)	5 20
1	C	734/881 (83%)	666 (91%)	54 (7%)	14 (2%)	8 28
1	D	805/881 (91%)	733 (91%)	59 (7%)	13 (2%)	9 32
All	All	3121/3524 (89%)	2811 (90%)	248 (8%)	62 (2%)	7 27

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)

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	595/643 (92%)	558 (94%)	37 (6%)	18 47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	569/643 (88%)	522 (92%)	47 (8%)	11 32
1	C	540/643 (84%)	504 (93%)	36 (7%)	16 43
1	D	590/643 (92%)	547 (93%)	43 (7%)	14 38
All	All	2294/2572 (89%)	2131 (93%)	163 (7%)	14 40

5 of 163 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	2147	SER
1	D	1983	LEU
1	C	2267	ASP
1	D	1808	LEU
1	D	2127	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	1832	HIS
1	C	2108	GLN
1	D	1989	GLN
1	D	1575	GLN
1	D	1578	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	818/881 (92%)	0.10	8 (0%)	82	82	25, 48, 77, 99	0
1	B	786/881 (89%)	0.10	12 (1%)	73	73	24, 49, 81, 98	0
1	C	750/881 (85%)	0.44	39 (5%)	27	23	41, 74, 103, 119	0
1	D	815/881 (92%)	0.22	17 (2%)	63	61	40, 63, 85, 109	0
All	All	3169/3524 (89%)	0.21	76 (2%)	59	56	24, 59, 91, 119	0

The worst 5 of 76 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1720	PHE	9.9
1	C	1511	ILE	8.4
1	C	1817	SER	5.5
1	C	1872	GLY	5.4
1	B	2000	GLY	4.6

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.