wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 06:10 am BST

PDB ID	:	5I6G
Title	:	Crystal structure of C-terminal variant 2 of Chaetomium thermophilum acetyl-CoA carboxylase
Authors	:	Hunkeler, M.; Stuttfeld, E.; Hagmann, A.; Imseng, S.; Maier, T.
Deposited on	:	2016-02-16
Resolution	:	4.50 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 following versions of software and data (see references (1)) were used in the production of this report:

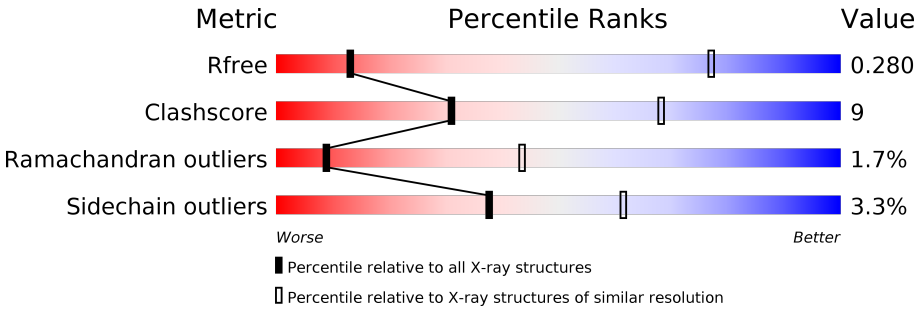
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 4.50 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	1055 (5.20-3.80)
Clashscore	141614	1123 (5.20-3.80)
Ramachandran outliers	138981	1069 (5.20-3.80)
Sidechain outliers	138945	1050 (5.20-3.80)

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	1161	66% 21% • 11%
1	B	1161	68% 20% • 11%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 16405 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 carboxylase-like protein,Acetyl-CoA carboxylase-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1028	Total	C	N	O	S	0	0	0
			8193	5204	1434	1526	29			
1	B	1030	Total	C	N	O	S	0	0	0
			8212	5216	1439	1528	29			

There are 2 discrepancies between the modelled and reference sequences:

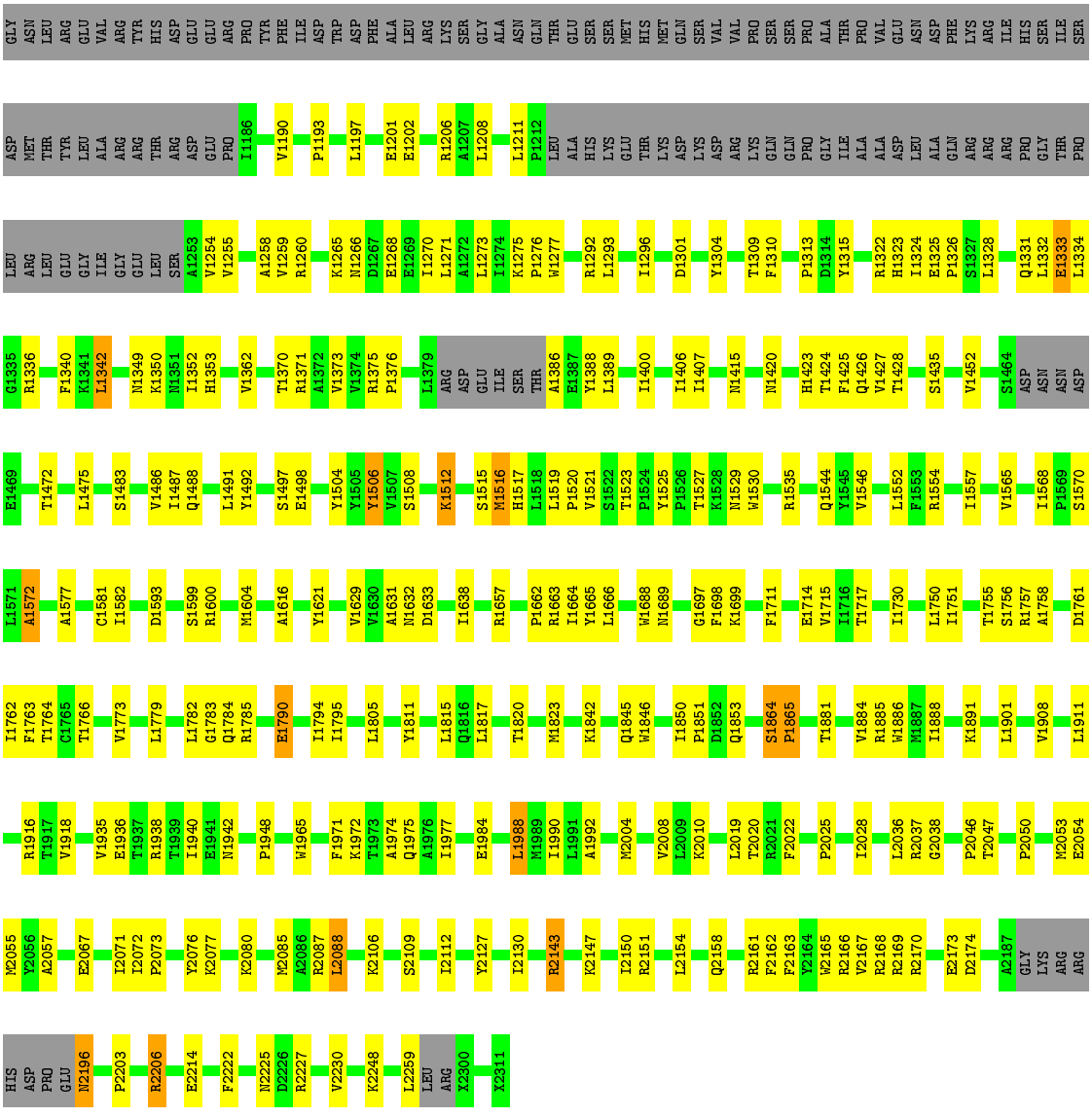
Chain	Residue	Modelled	Actual	Comment	Reference
A	1113	GLY	-	expression tag	UNP G0S3L5
B	1113	GLY	-	expression tag	UNP G0S3L5

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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. 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. 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.

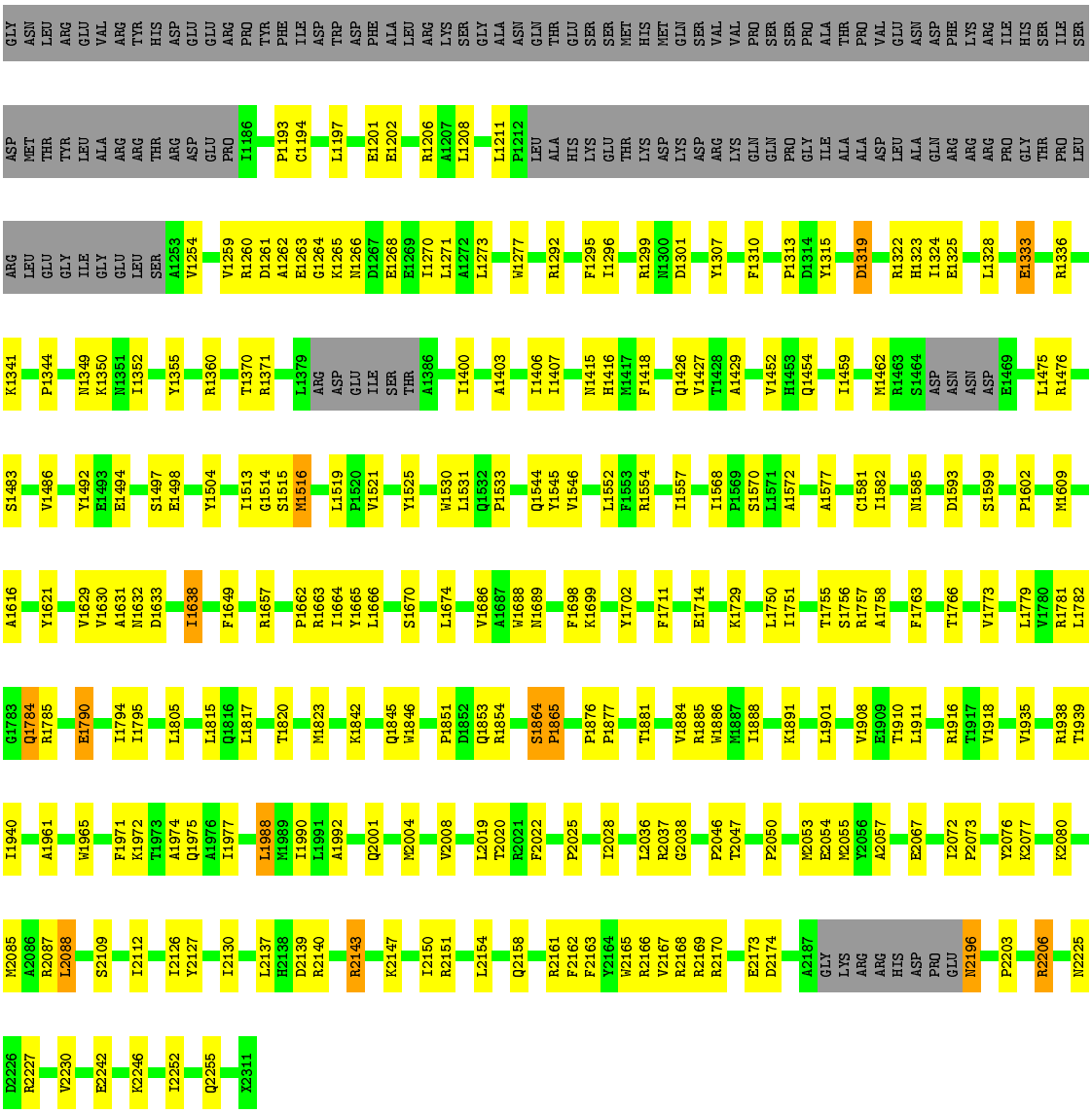
- Molecule 1: Acetyl-CoA carboxylase-like protein,Acetyl-CoA carboxylase-like protein

Chain A:



- Molecule 1: Acetyl-CoA carboxylase-like protein,Acetyl-CoA carboxylase-like protein

Chain B:


GLY ASN LEU ARG GLU VAL ARG TYR HIS ASP GLU GLU ARG PRO TYR PHE ILE ASP TRP ASP PHE ALA LEU ARG LYS SER GLY ALA ASN GLN THR GLU SER SER MET HIS MET GLN SER VAL VAL PRO SER SER PRO ALA THR PRO VAL GLU ASN ASP PHE LYS ARG ILE HIS SER ILE SER
ASP MET THR TYR LEU ALA ARG ARG THR ARG ASP GLU PRO I1186 P1193 C1194 L1197 E1201 E1202 R1206 A1207 L1208 L1211 P1212 LEU ALA HIS LYS GLU THR LYS ASP LYS ASP ARG LYS GLN GLN PRO GLY ILE ALA ALA ASP LEU ALA GLN ARG ARG ARG PRO GLY THR PRO LEU
ARG LEU GLU GLY ILE GLY GLU LEU SER A1253 Y1254 V1259 R1260 D1261 A1262 E1263 G1264 K1265 N1266 D1267 E1268 E1269 I1270 L1271 A1272 L1273 W1277 R1292 F1295 I1296 R1299 N1300 D1301 Y1307 F1310 P1313 D1314 Y1315 D1319 R1322 H1323 I1324 E1325 L1328 E1333 R1336
K1341 P1344 N1349 K1350 N1351 I1352 Y1355 R1360 T1370 R1371 L1379 ARG ASP GLU ILE SER THR A1386 I1400 A1403 I1406 I1407 N1415 H1416 M1417 F1418 Q1426 V1427 T1428 A1429 V1452 H1453 Q1454 I1459 M1462 R1463 S1464 ASP ASN ASN ASP E1469 L1475 R1476
S1483 V1486 Y1492 E1493 E1494 S1497 E1498 Y1504 I1513 G1514 S1515 M1516 L1519 P1520 V1521 Y1525 W1530 L1531 Q1532 P1533 Q1544 Y1545 V1546 L1552 F1553 R1554 I1557 I1568 P1569 S1570 L1571 A1572 A1577 C1581 I1582 N1585 D1593 S1599 P1602 M1609
A1616 Y1621 V1629 V1630 A1631 N1632 D1633 I1638 F1649 R1657 P1662 R1663 I1664 Y1665 L1666 S1670 L1674 V1686 A1687 W1688 N1689 F1698 K1699 Y1702 F1711 E1714 K1729 L1750 I1751 T1755 S1756 R1757 A1758 F1763 T1766 V1773 L1779 V1780 R1781 L1782
G1783 Q1784 R1785 E1790 I1794 I1795 L1805 L1815 Q1816 L1817 T1820 M1823 K1842 Q1845 W1846 P1851 D1852 Q1853 R1854 S1864 P1865 P1876 P1877 T1881 V1884 R1885 W1886 M1887 I1888 K1891 L1901 V1908 E1909 T1910 L1911 R1916 T1917 V1918 V1935 R1938 T1939
I1940 A1961 W1965 F1971 K1972 T1973 A1974 Q1975 A1976 I1977 L1988 M1989 I1990 L1991 A1992 Q2001 M2004 V2008 L2019 T2020 R2021 F2022 P2025 I2028 L2036 R2037 G2038 P2046 T2047 P2050 M2053 E2054 M2055 Y2056 A2057 E2067 I2072 P2073 Y2076 K2077 K2080
M2085 A2086 R2087 L2088 S2109 I2112 I2126 Y2127 I2130 L2137 H2138 D2139 R2140 R2143 K2147 I2150 R2151 L2154 Q2158 R2161 F2162 F2163 Y2164 W2165 R2166 V2167 R2168 R2169 R2170 E2173 D2174 A2187 GLY LYS ARG ARG HIS ASP PRO GLU M2196 P2203 R2206 N2225
D2226 R2227 V2230 E2242 K2246 I2252 Q2255 X2311

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	100.17Å 153.45Å 249.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.10 – 4.50 49.10 – 4.50	Depositor EDS
% Data completeness (in resolution range)	99.4 (49.10-4.50) 99.4 (49.10-4.50)	Depositor EDS
R_{merge}	0.27	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.38 (at 4.45Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.236 , 0.240 0.279 , 0.280	Depositor DCC
R_{free} test set	1112 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å^2)	202.9	Xtriage
Anisotropy	0.442	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 256.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	16405	wwPDB-VP
Average B, all atoms (Å^2)	275.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.64% 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	0.51	0/8311	0.70	1/11255 (0.0%)
1	B	0.49	0/8330	0.68	0/11280
All	All	0.50	0/16641	0.69	1/22535 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1506	TYR	CA-CB-CG	5.16	123.20	113.40

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	8193	0	8096	152	0
1	B	8212	0	8120	157	0
All	All	16405	0	16216	296	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 296 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:1515:SER:HA	1:B:1599:SER:O	1.25	1.28
1:B:1319:ASP:HB2	1:B:1344:PRO:HG2	1.39	1.02
1:B:1515:SER:CA	1:B:1599:SER:O	2.19	0.89
1:A:1336:ARG:HB3	1:A:1525:TYR:CD2	2.09	0.88
1:B:1609:MET:HE1	1:B:1630:VAL:CG1	2.06	0.86

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	1006/1161 (87%)	898 (89%)	89 (9%)	19 (2%)	8 41
1	B	1008/1161 (87%)	897 (89%)	96 (10%)	15 (2%)	10 46
All	All	2014/2322 (87%)	1795 (89%)	185 (9%)	34 (2%)	9 43

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1864	SER
1	A	2225	ASN
1	B	1864	SER
1	A	1333	GLU
1	A	1483	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	872/990 (88%)	838 (96%)	34 (4%)	32 57
1	B	874/990 (88%)	851 (97%)	23 (3%)	46 67
All	All	1746/1980 (88%)	1689 (97%)	57 (3%)	38 61

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2088	LEU
1	A	2196	ASN
1	B	2151	ARG
1	A	2106	LYS
1	A	2151	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 carbohydrates 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	2261:ARG	C	2300:UNK	N	11.76

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.