wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 03:01 am BST

PDB ID	:	2QZI
Title	:	The crystal structure of a conserved protein of unknown function from Streptococcus thermophilus LMG 18311.
Authors	:	Tan, K.; Zhou, M.; Freeman, L.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2007-08-16
Resolution	:	2.20 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:

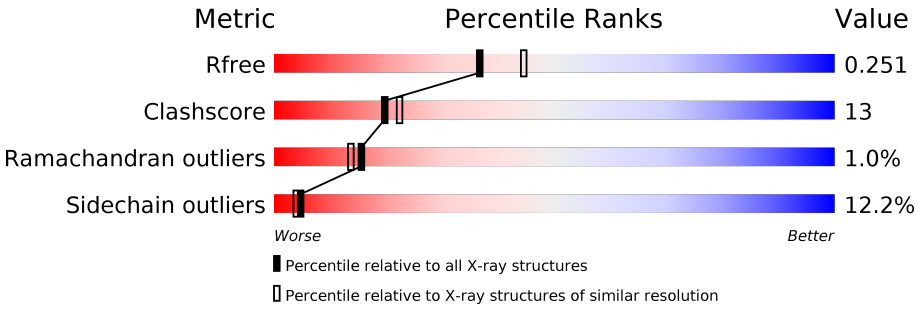
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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	103	73% 18% 6% • •
1	B	103	71% 24% • •
1	C	103	78% 17% • • •
1	D	103	51% 36% 7% • 5%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3526 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 Uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	101	Total	C	N	O	Se	0	3	0
			850	544	146	158	2			
1	B	101	Total	C	N	O	Se	0	2	0
			840	535	142	161	2			
1	C	101	Total	C	N	O	Se	0	3	0
			847	542	144	159	2			
1	D	98	Total	C	N	O	Se	0	1	0
			815	520	139	155	1			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	EXPRESSION TAG	UNP Q5M594
A	-1	ASN	-	EXPRESSION TAG	UNP Q5M594
A	0	ALA	-	EXPRESSION TAG	UNP Q5M594
B	-2	SER	-	EXPRESSION TAG	UNP Q5M594
B	-1	ASN	-	EXPRESSION TAG	UNP Q5M594
B	0	ALA	-	EXPRESSION TAG	UNP Q5M594
C	-2	SER	-	EXPRESSION TAG	UNP Q5M594
C	-1	ASN	-	EXPRESSION TAG	UNP Q5M594
C	0	ALA	-	EXPRESSION TAG	UNP Q5M594
D	-2	SER	-	EXPRESSION TAG	UNP Q5M594
D	-1	ASN	-	EXPRESSION TAG	UNP Q5M594
D	0	ALA	-	EXPRESSION TAG	UNP Q5M594

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

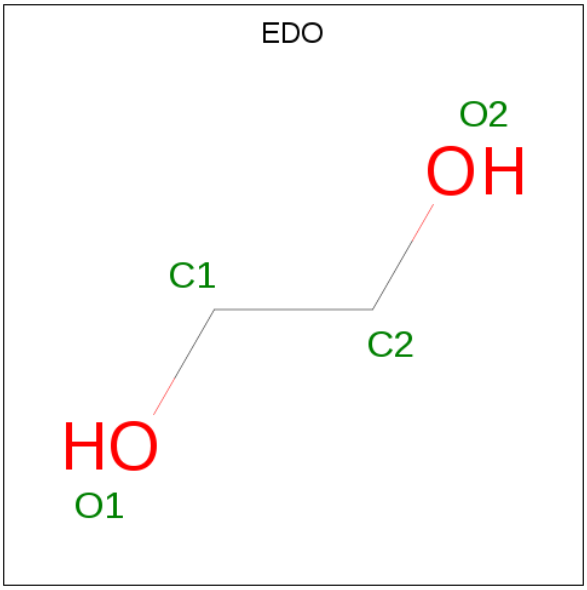
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	3	Total	Na	0	0
			3	3		
2	A	2	Total	Na	0	0
			2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	2	Total Na 2 2	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C O 4 2 2	0	0

- Molecule 4 is water.

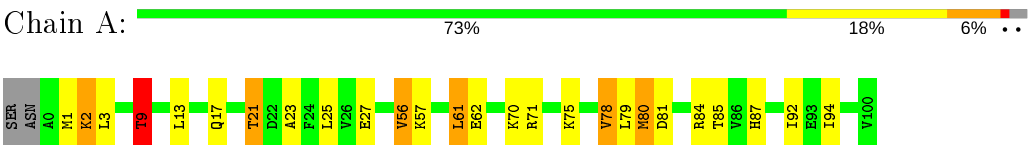
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	50	Total O 50 50	0	0
4	B	67	Total O 67 67	0	0
4	C	38	Total O 38 38	0	0
4	D	8	Total O 8 8	0	0

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.

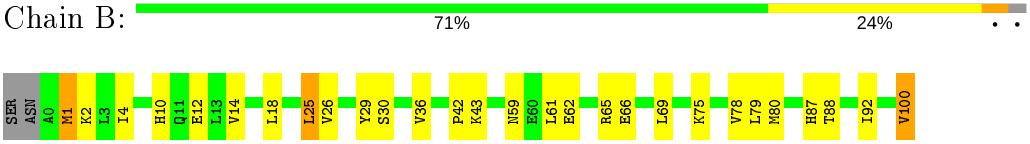
- Molecule 1: Uncharacterized protein

Chain A:



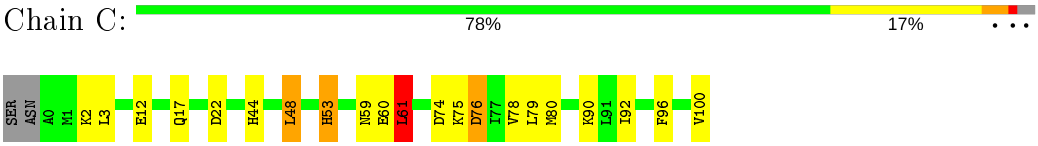
- Molecule 1: Uncharacterized protein

Chain B:



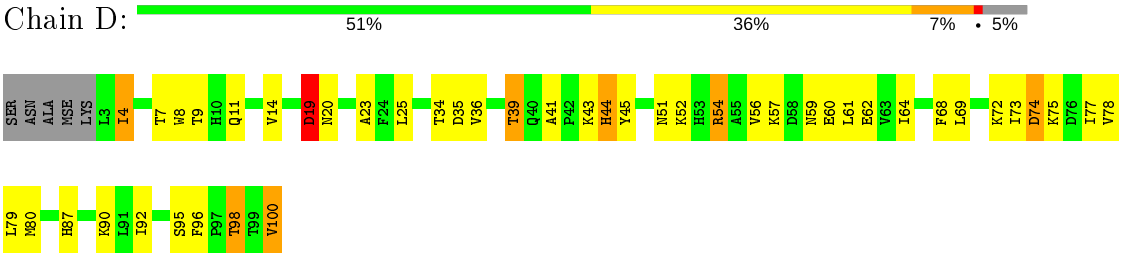
- Molecule 1: Uncharacterized protein

Chain C:



- Molecule 1: Uncharacterized protein

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 64 2 2	Depositor
Cell constants a, b, c, α, β, γ	150.78Å 150.78Å 103.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	36.23 – 2.20 36.22 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.8 (36.23-2.20) 98.8 (36.22-2.20)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.68 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.210 , 0.251 0.207 , 0.251	Depositor DCC
R_{free} test set	1755 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	44.6	Xtriage
Anisotropy	0.502	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 50.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	3526	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.33% 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

Bond lengths and bond angles in the following residue types are not validated in this section: NA, EDO

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	1.07	0/872	1.02	2/1176 (0.2%)
1	B	1.05	0/858	0.98	0/1158
1	C	1.02	0/869	0.95	3/1173 (0.3%)
1	D	0.72	0/832	0.86	0/1126
All	All	0.98	0/3431	0.95	5/4633 (0.1%)

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	1	0
1	D	0	1
All	All	1	1

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	9	THR	CA-CB-CG2	8.14	123.79	112.40
1	C	48	LEU	CA-CB-CG	6.17	129.48	115.30
1	C	61	LEU	CA-CB-CG	5.65	128.29	115.30
1	C	22	ASP	CB-CG-OD1	5.51	123.26	118.30
1	A	80	MSE	CG-SE-CE	-5.27	87.31	98.90

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	9	THR	CB

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	90	LYS	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	850	0	873	22	0
1	B	840	0	853	20	0
1	C	847	0	867	11	0
1	D	815	0	820	36	0
2	A	2	0	0	0	0
2	B	3	0	0	0	0
2	C	2	0	0	0	0
3	B	4	0	6	3	0
4	A	50	0	0	4	0
4	B	67	0	0	0	0
4	C	38	0	0	0	0
4	D	8	0	0	0	0
All	All	3526	0	3419	86	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 86 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:78:VAL:HG11	1:D:80:MSE:HE3	1.27	1.10
1:D:78:VAL:CG1	1:D:80:MSE:HE3	1.93	0.96
1:D:34:THR:HG22	1:D:51:ASN:HB2	1.47	0.95
1:B:62:GLU:OE2	1:B:65:ARG:NH1	2.17	0.77
1:A:2:LYS:HD3	1:A:3:LEU:H	1.50	0.77

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	102/103 (99%)	100 (98%)	2 (2%)	0	100 100
1	B	101/103 (98%)	99 (98%)	2 (2%)	0	100 100
1	C	102/103 (99%)	98 (96%)	4 (4%)	0	100 100
1	D	97/103 (94%)	83 (86%)	10 (10%)	4 (4%)	3 1
All	All	402/412 (98%)	380 (94%)	18 (4%)	4 (1%)	15 14

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	74	ASP
1	D	19	ASP
1	D	59	ASN
1	D	62	GLU

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	98/95 (103%)	87 (89%)	11 (11%)	6 5
1	B	97/95 (102%)	86 (89%)	11 (11%)	6 5
1	C	98/95 (103%)	87 (89%)	11 (11%)	6 5
1	D	94/95 (99%)	78 (83%)	16 (17%)	2 1
All	All	387/380 (102%)	338 (87%)	49 (13%)	5 3

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	12	GLU
1	C	60	GLU
1	D	79	LEU
1	C	53	HIS
1	C	61	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	87	HIS
1	C	53	HIS
1	D	15	ASN
1	D	20	ASN

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)

Of 8 ligands modelled in this entry, 7 are monoatomic - leaving 1 for Mogul analysis.

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	$\#\|Z\| > 2$
3	EDO	B	104	-	3,3,3	0.63	0	2,2,2	0.38	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EDO	B	104	-	-	0/1/1/1	-

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.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	104	EDO	3	0

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

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.