Full wwPDB X-ray Structure Validation Report (i)

Oct 12, 2024 – 09:05 AM EDT

PDB ID	:	5UKH
Title	:	Structure of TelC from Streptococcus intermedius B196
Authors	:	Whitney, J.C.; Ching, M.Q.; Bryant, D.; Mougous, J.D.
Deposited on	:	2017-01-22
Resolution	:	1.98 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

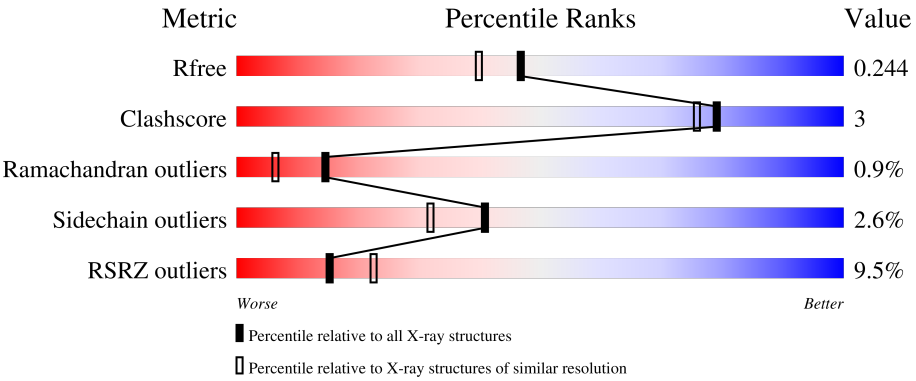
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 1.98 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}	164625	1356 (1.98-1.98)
Clashscore	180529	1437 (1.98-1.98)
Ramachandran outliers	177936	1426 (1.98-1.98)
Sidechain outliers	177891	1426 (1.98-1.98)
RSRZ outliers	164620	1356 (1.98-1.98)

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	365	8%, 81%, 5%, 12%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2691 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	321	Total	C	N	O	Se	0	0	0
			2534	1604	418	506	6			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	188	MSE	-	initiating methionine	UNP T1ZG69
A	189	GLY	-	expression tag	UNP T1ZG69
A	190	SER	-	expression tag	UNP T1ZG69
A	191	SER	-	expression tag	UNP T1ZG69
A	192	HIS	-	expression tag	UNP T1ZG69
A	193	HIS	-	expression tag	UNP T1ZG69
A	194	HIS	-	expression tag	UNP T1ZG69
A	195	HIS	-	expression tag	UNP T1ZG69
A	196	HIS	-	expression tag	UNP T1ZG69
A	197	HIS	-	expression tag	UNP T1ZG69
A	198	SER	-	expression tag	UNP T1ZG69
A	199	GLN	-	expression tag	UNP T1ZG69
A	200	ASP	-	expression tag	UNP T1ZG69

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Ca	0	0
			3	3		

- Molecule 3 is water.

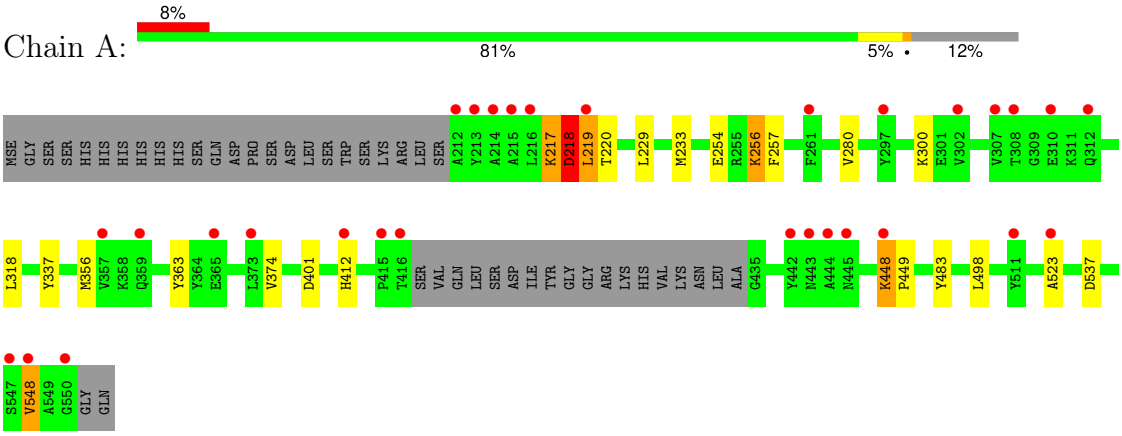
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	154	Total	O	0	0
			154	154		

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: Uncharacterized protein

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	127.38Å 132.69Å 58.26Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.20 – 1.98 49.20 – 1.98	Depositor EDS
% Data completeness (in resolution range)	87.8 (49.20-1.98) 86.3 (49.20-1.98)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.04 (at 1.98Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.222 , 0.245 0.219 , 0.244	Depositor DCC
R_{free} test set	32814 reflections (5.70%)	wwPDB-VP
Wilson B-factor (Å^2)	31.6	Xtriage
Anisotropy	0.626	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.020 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2691	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.29% 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: CA

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	0.43	0/2580	0.56	1/3474 (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	1

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	401	ASP	CB-CG-OD1	5.15	122.94	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	218	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	2534	0	2412	13	0
2	A	3	0	0	0	0
3	A	154	0	0	1	0
All	All	2691	0	2412	13	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 3.

All (13) 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:A:229:LEU:HB3	1:A:233:MSE:HE3	1.83	0.59
1:A:483:TYR:OH	3:A:701:HOH:O	2.19	0.51
1:A:280:VAL:HG22	1:A:318:LEU:HD21	1.93	0.51
1:A:219:LEU:HA	1:A:220:THR:HB	1.96	0.48
1:A:337:TYR:HB2	1:A:374:VAL:HG21	1.97	0.46
1:A:537:ASP:OD1	1:A:537:ASP:N	2.50	0.45
1:A:217:LYS:HD3	1:A:218:ASP:H	1.82	0.44
1:A:448:LYS:CB	1:A:449:PRO:HD2	2.47	0.43
1:A:218:ASP:O	1:A:219:LEU:HB2	2.19	0.42
1:A:356:MSE:HE3	1:A:363:TYR:CD1	2.54	0.42
1:A:256:LYS:HE3	1:A:257:PHE:CZ	2.54	0.41
1:A:219:LEU:HA	1:A:220:THR:CB	2.51	0.41
1:A:498:LEU:HD12	1:A:548:VAL:HG12	2.02	0.41

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	317/365 (87%)	306 (96%)	8 (2%)	3 (1%)	14 6

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	219	LEU
1	A	448	LYS
1	A	523	ALA

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	268/313 (86%)	261 (97%)	7 (3%)	41 32

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	217	LYS
1	A	218	ASP
1	A	254	GLU
1	A	256	LYS
1	A	300	LYS
1	A	412	HIS
1	A	548	VAL

Sometimes 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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

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	315/365 (86%)	0.54	30 (9%)	15	22	31, 50, 96, 120	0

All (30) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	443	ASN	4.9
1	A	212	ALA	4.5
1	A	219	LEU	4.5
1	A	445	ASN	4.5
1	A	216	LEU	4.2
1	A	412	HIS	4.1
1	A	213	TYR	3.9
1	A	448	LYS	3.6
1	A	215	ALA	3.5
1	A	416	THR	3.5
1	A	302	VAL	3.4
1	A	261	PHE	2.9
1	A	312	GLN	2.8
1	A	415	PRO	2.8
1	A	307	VAL	2.7
1	A	214	ALA	2.7
1	A	444	ALA	2.7
1	A	548	VAL	2.5
1	A	359	GLN	2.4
1	A	550	GLY	2.3
1	A	310	GLU	2.3
1	A	357	VAL	2.3
1	A	547	SER	2.3
1	A	511	TYR	2.2
1	A	308	THR	2.1
1	A	523	ALA	2.1
1	A	373	LEU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	365	GLU	2.1
1	A	442	TYR	2.1
1	A	297	TYR	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	CA	A	603	1/1	0.76	0.10	108,108,108,108	1
2	CA	A	601	1/1	0.98	0.04	38,38,38,38	0
2	CA	A	602	1/1	0.99	0.05	47,47,47,47	1

6.5 Other polymers (i)

There are no such residues in this entry.