Full wwPDB X-ray Structure Validation Report ⓘ

May 26, 2020 – 07:37 pm BST

PDB ID	:	5G23
Title	:	Type IV-like pilin TTHA1219 from Thermus thermophilus
Authors	:	Karuppiah, V.; Derrick, J.P.
Deposited on	:	2016-04-06
Resolution	:	1.85 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 ⓘ symbol.

The following versions of software and data (see references ⓘ) were used in the production of this report:

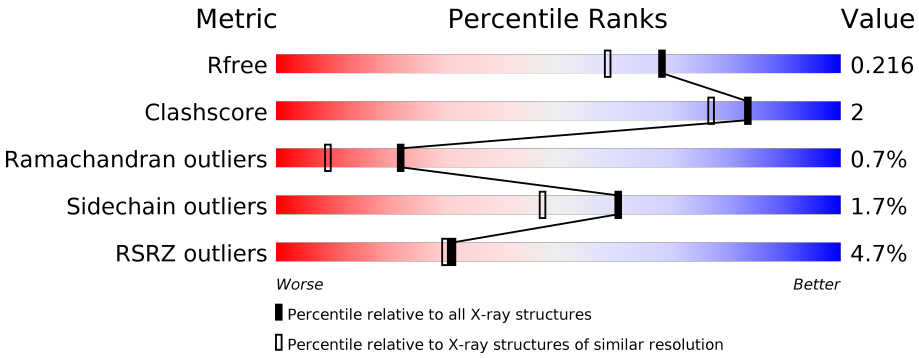
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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% 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	236	5% 77% • 19%
1	B	236	% 77% 5% • 18%
1	C	236	3% 76% 5% 19%
1	D	236	6% 75% 6% • 18%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6264 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 TYPE-IV LIKE PILIN TTHA1219.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	191	Total	C	N	O	S	0	0	0
			1415	892	249	268	6			
1	B	194	Total	C	N	O	S	0	0	0
			1434	902	252	274	6			
1	C	191	Total	C	N	O	S	0	0	0
			1418	893	252	268	5			
1	D	194	Total	C	N	O	S	0	0	0
			1441	905	258	272	6			

- Molecule 2 is water.

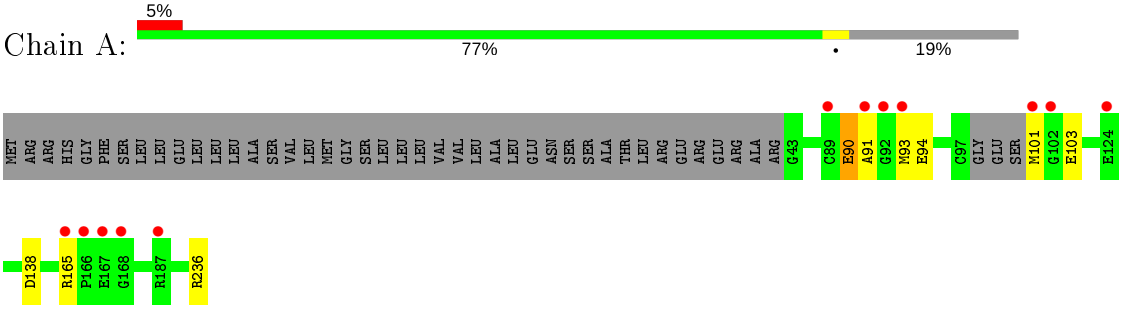
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	145	Total	O	0	0
			145	145		
2	B	139	Total	O	0	0
			139	139		
2	C	171	Total	O	0	0
			171	171		
2	D	101	Total	O	0	0
			101	101		

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 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.

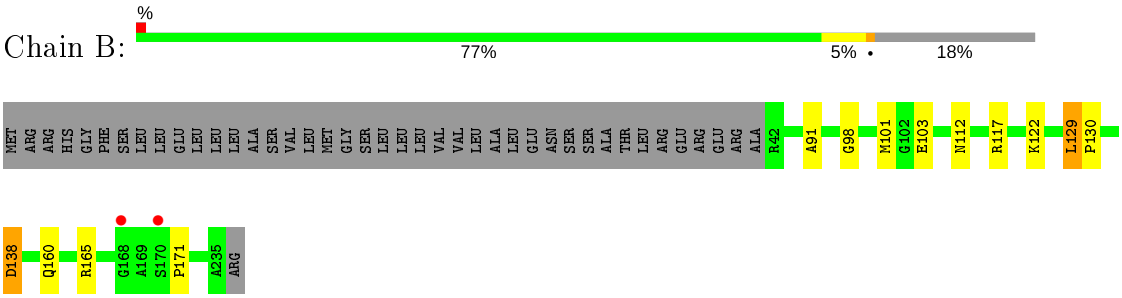
- Molecule 1: TYPE-IV LIKE PILIN TTHA1219

Chain A:



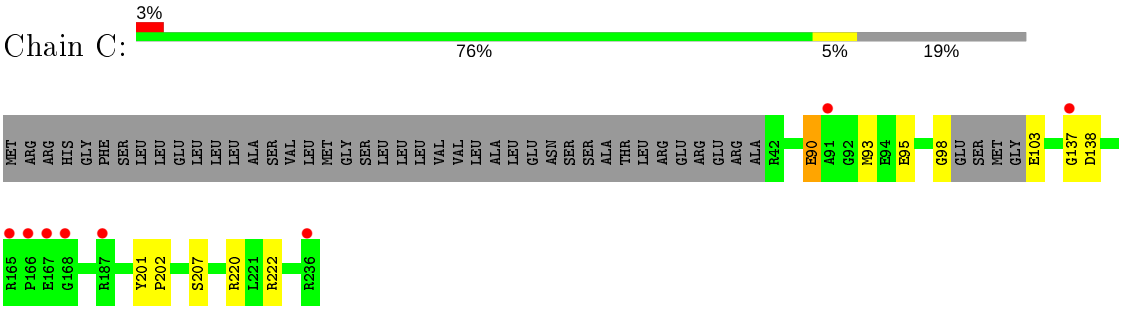
- Molecule 1: TYPE-IV LIKE PILIN TTHA1219

Chain B:



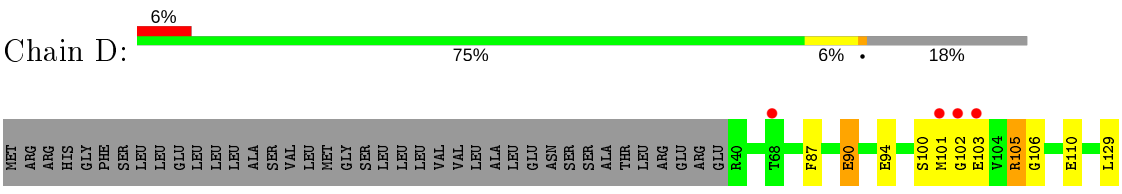
- Molecule 1: TYPE-IV LIKE PILIN TTHA1219

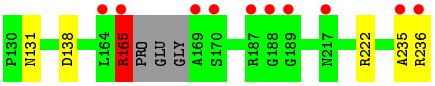
Chain C:



- Molecule 1: TYPE-IV LIKE PILIN TTHA1219

Chain D:




P130
N131
D138
L164
R165
PRO
GLU
GLY
A169
S170
R187
G188
G189
N217
R222
A235
R236

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	65.40Å 66.95Å 97.47Å 90.00° 99.14° 90.00°	Depositor
Resolution (Å)	96.23 – 1.85 35.94 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.2 (96.23-1.85) 99.2 (35.94-1.85)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.21 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.180 , 0.209 0.188 , 0.216	Depositor DCC
R_{free} test set	3564 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	23.4	Xtriage
Anisotropy	0.089	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 42.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6264	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.96% 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	0.93	1/1444 (0.1%)	0.89	0/1966
1	B	0.99	1/1464 (0.1%)	0.92	2/1994 (0.1%)
1	C	0.98	1/1447 (0.1%)	0.97	4/1970 (0.2%)
1	D	0.98	2/1469 (0.1%)	0.96	3/1997 (0.2%)
All	All	0.97	5/5824 (0.1%)	0.94	9/7927 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	90	GLU	CD-OE2	9.24	1.35	1.25
1	C	90	GLU	CD-OE2	7.13	1.33	1.25
1	D	110	GLU	CD-OE1	6.93	1.33	1.25
1	B	138	ASP	CB-CG	6.57	1.65	1.51
1	A	90	GLU	CD-OE2	5.42	1.31	1.25

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	138	ASP	CB-CG-OD1	8.79	126.21	118.30
1	D	102	GLY	N-CA-C	-7.53	94.28	113.10
1	C	222	ARG	NE-CZ-NH1	7.27	123.93	120.30
1	D	165	ARG	NE-CZ-NH1	6.92	123.76	120.30
1	B	129	LEU	CA-CB-CG	5.75	128.52	115.30
1	D	222	ARG	NE-CZ-NH1	5.65	123.13	120.30
1	C	222	ARG	NE-CZ-NH2	-5.57	117.51	120.30
1	C	220	ARG	NE-CZ-NH1	5.30	122.95	120.30
1	C	93	MET	CA-CB-CG	5.18	122.11	113.30

There are no chirality outliers.

There are no planarity outliers.

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	1415	0	1407	4	0
1	B	1434	0	1422	11	0
1	C	1418	0	1411	7	0
1	D	1441	0	1436	7	0
2	A	145	0	0	0	0
2	B	139	0	0	0	1
2	C	171	0	0	1	0
2	D	101	0	0	0	1
All	All	6264	0	5676	27	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 2.

All (27) 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:160:GLN:HE22	1:B:171:PRO:HG3	1.40	0.86
1:C:90:GLU:OE2	1:D:90:GLU:OE2	1.94	0.85
1:B:91:ALA:HB2	1:B:101:MET:HG2	1.75	0.69
1:A:91:ALA:HB2	1:A:101:MET:HG2	1.75	0.67
1:B:160:GLN:HE22	1:B:171:PRO:CG	2.11	0.63
1:C:95:GLU:HG3	2:C:2059:HOH:O	1.98	0.63
1:C:98:GLY:O	1:C:103:GLU:HG2	1.99	0.63
1:B:160:GLN:NE2	1:B:171:PRO:HG3	2.13	0.62
1:B:129:LEU:CD1	1:C:202:PRO:HD3	2.34	0.57
1:A:90:GLU:H	1:A:93:MET:HE2	1.73	0.54
1:B:160:GLN:NE2	1:B:171:PRO:CG	2.71	0.54
1:D:105:ARG:HG3	1:D:106:GLY:N	2.22	0.52
1:A:90:GLU:H	1:A:93:MET:CE	2.23	0.51
1:C:137:GLY:O	1:C:138:ASP:O	2.29	0.50
1:C:137:GLY:O	1:C:138:ASP:C	2.49	0.49
1:D:165:ARG:HG2	1:D:165:ARG:HH11	1.77	0.49
1:A:93:MET:HG2	1:A:94:GLU:HG3	1.94	0.49
1:D:87:PHE:CZ	1:D:105:ARG:HG2	2.48	0.48
1:B:112:ASN:HB3	1:B:117:ARG:NH1	2.30	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:103:GLU:OE1	1:B:122:LYS:CE	2.64	0.45
1:D:235:ALA:O	1:D:236:ARG:C	2.55	0.43
1:B:103:GLU:OE1	1:B:122:LYS:HE3	2.20	0.42
1:B:165:ARG:HG2	1:B:165:ARG:HH11	1.83	0.41
1:B:129:LEU:HA	1:B:130:PRO:HD3	1.90	0.41
1:D:165:ARG:HH11	1:D:165:ARG:CG	2.33	0.41
1:C:201:TYR:OH	1:C:207:SER:OG	2.09	0.40
1:D:236:ARG:HD2	1:D:236:ARG:C	2.41	0.40

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)
2:B:2001:HOH:O	2:D:2081:HOH:O[1_545]	2.09	0.11

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	187/236 (79%)	185 (99%)	1 (0%)	1 (0%)	29 15
1	B	192/236 (81%)	189 (98%)	1 (0%)	2 (1%)	15 5
1	C	187/236 (79%)	185 (99%)	2 (1%)	0	100 100
1	D	190/236 (80%)	184 (97%)	4 (2%)	2 (1%)	14 4
All	All	756/944 (80%)	743 (98%)	8 (1%)	5 (1%)	22 9

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	98	GLY
1	A	138	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	138	ASP
1	D	103	GLU
1	D	138	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	145/183 (79%)	142 (98%)	3 (2%)	53	38
1	B	147/183 (80%)	147 (100%)	0	100	100
1	C	145/183 (79%)	145 (100%)	0	100	100
1	D	147/183 (80%)	140 (95%)	7 (5%)	25	10
All	All	584/732 (80%)	574 (98%)	10 (2%)	60	47

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	103	GLU
1	A	165	ARG
1	A	236	ARG
1	D	94	GLU
1	D	100	SER
1	D	101	MET
1	D	105	ARG
1	D	129	LEU
1	D	131	ASN
1	D	165	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	132	ASN
1	B	160	GLN
1	B	163	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	217	ASN
1	D	79	GLN
1	D	131	ASN
1	D	132	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	191/236 (80%)	-0.02	12 (6%)	20	19	15, 24, 56, 96	0
1	B	194/236 (82%)	-0.22	2 (1%)	82	82	14, 23, 45, 57	0
1	C	191/236 (80%)	-0.11	8 (4%)	36	34	13, 21, 54, 70	0
1	D	194/236 (82%)	0.10	14 (7%)	15	15	15, 28, 51, 83	0
All	All	770/944 (81%)	-0.06	36 (4%)	31	30	13, 24, 52, 96	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	102	GLY	5.1
1	A	91	ALA	5.0
1	C	166	PRO	5.0
1	A	101	MET	5.0
1	A	166	PRO	5.0
1	A	168	GLY	4.6
1	D	68	THR	4.3
1	B	168	GLY	3.9
1	D	165	ARG	3.8
1	D	101	MET	3.7
1	C	165	ARG	3.5
1	A	167	GLU	3.4
1	A	93	MET	3.3
1	D	236	ARG	3.3
1	A	124	GLU	3.3
1	D	235	ALA	3.0
1	A	165	ARG	2.9
1	D	169	ALA	2.9
1	D	164	LEU	2.8
1	D	187	ARG	2.8
1	A	89	CYS	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	217	ASN	2.7
1	D	102	GLY	2.7
1	D	189	GLY	2.6
1	A	92	GLY	2.6
1	A	187	ARG	2.5
1	C	168	GLY	2.5
1	C	187	ARG	2.5
1	C	167	GLU	2.5
1	D	188	GLY	2.5
1	C	236	ARG	2.5
1	C	91	ALA	2.4
1	D	170	SER	2.4
1	D	103	GLU	2.3
1	C	137	GLY	2.2
1	B	170	SER	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 carbohydrates 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.