Full wwPDB X-ray Structure Validation Report ⓘ

May 30, 2020 – 02:53 am BST

PDB ID	:	6FWV
Title	:	The Bacillus anthracis TIE protein
Authors	:	Miller, O.K.; Banfield, M.J.; Schwarz-Linek, U.
Deposited on	:	2018-03-07
Resolution	:	2.58 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:

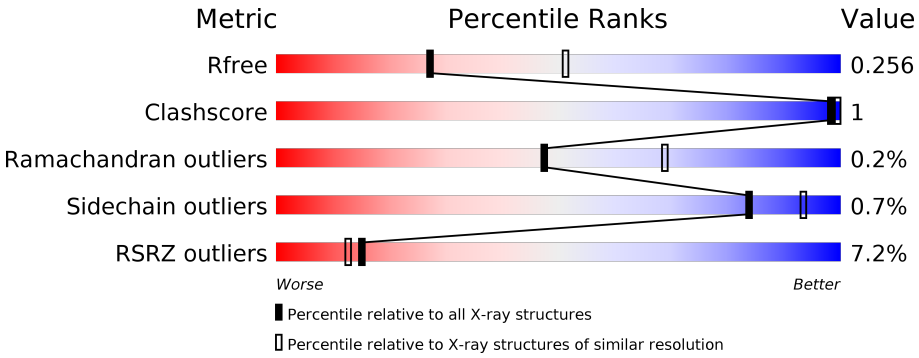
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.58 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	3676 (2.60-2.56)
Clashscore	141614	4049 (2.60-2.56)
Ramachandran outliers	138981	3979 (2.60-2.56)
Sidechain outliers	138945	3979 (2.60-2.56)
RSRZ outliers	127900	3614 (2.60-2.56)

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	526	5% 98%
1	B	526	9% 96%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8226 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 Collagen Adhesion protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	522	Total	C	N	O	S	Se	0	0	0
			4092	2593	675	817	1	6			
1	B	520	Total	C	N	O	S	Se	0	0	0
			4072	2580	671	813	1	7			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP A0A0F7RA58
A	2	ALA	-	expression tag	UNP A0A0F7RA58
A	3	MSE	-	expression tag	UNP A0A0F7RA58
B	1	GLY	-	expression tag	UNP A0A0F7RA58
B	2	ALA	-	expression tag	UNP A0A0F7RA58
B	3	MSE	-	expression tag	UNP A0A0F7RA58

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	3	Total	Zn	0	0
			3	3		
2	A	4	Total	Zn	0	0
			4	4		

- Molecule 3 is water.

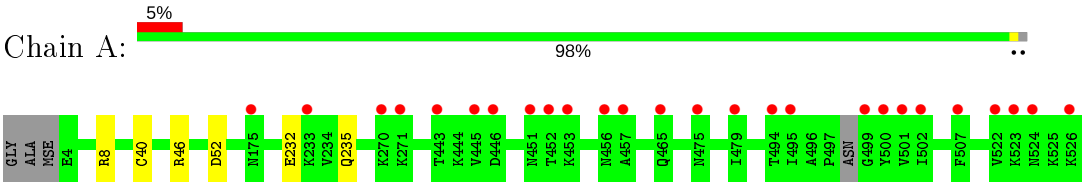
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	31	Total	O	0	0
			31	31		
3	B	24	Total	O	0	0
			24	24		

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.

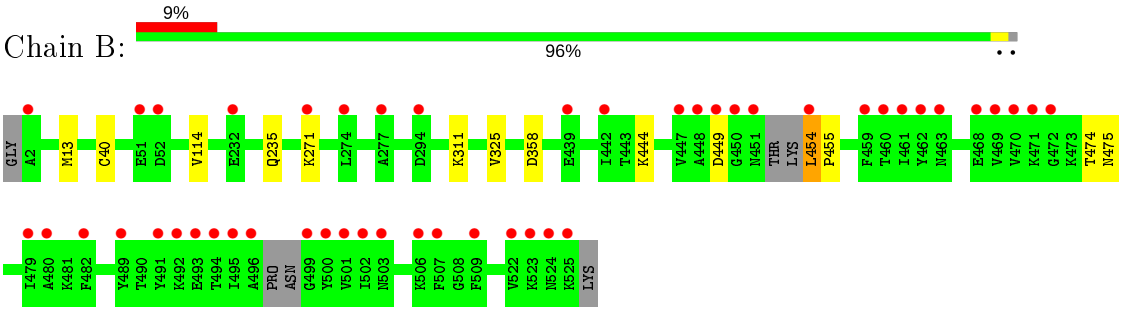
- Molecule 1: Collagen Adhesion protein

Chain A:



- Molecule 1: Collagen Adhesion protein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	208.82Å 66.50Å 106.58Å 90.00° 116.81° 90.00°	Depositor
Resolution (Å)	93.18 – 2.58 93.19 – 2.58	Depositor EDS
% Data completeness (in resolution range)	98.8 (93.18-2.58) 98.8 (93.19-2.58)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.63 (at 2.58Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.221 , 0.256 0.224 , 0.256	Depositor DCC
R_{free} test set	2028 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	58.2	Xtriage
Anisotropy	0.196	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 39.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	8226	wwPDB-VP
Average B, all atoms (Å^2)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.43% 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: ZN

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	0.40	0/4148	0.57	0/5577
1	B	0.40	0/4125	0.57	0/5544
All	All	0.40	0/8273	0.57	0/11121

There are no bond length outliers.

There are no bond angle outliers.

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	4092	0	4133	3	0
1	B	4072	0	4106	7	0
2	A	4	0	0	0	0
2	B	3	0	0	0	0
3	A	31	0	0	0	0
3	B	24	0	0	0	0
All	All	8226	0	8239	10	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 1.

All (10) 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:40:CYS:SG	1:B:235:GLN:CD	2.13	1.27
1:A:40:CYS:SG	1:A:235:GLN:CD	2.13	1.27
1:A:232:GLU:HA	1:A:232:GLU:OE1	1.86	0.76
1:B:13:MSE:HE1	1:B:114:VAL:HG21	1.90	0.53
1:B:13:MSE:CE	1:B:114:VAL:HG21	2.38	0.52
1:B:40:CYS:SG	1:B:235:GLN:OE1	2.68	0.48
1:B:455:PRO:HA	1:B:474:THR:HG23	1.96	0.47
1:B:311:LYS:HD3	1:B:325:VAL:HG12	2.01	0.41
1:B:444:LYS:HG3	1:B:454:LEU:HD11	2.02	0.41
1:A:40:CYS:SG	1:A:235:GLN:OE1	2.68	0.40

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	518/526 (98%)	507 (98%)	11 (2%)	0	100	100
1	B	514/526 (98%)	503 (98%)	9 (2%)	2 (0%)	34	55
All	All	1032/1052 (98%)	1010 (98%)	20 (2%)	2 (0%)	47	69

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	271	LYS
1	B	475	ASN

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	452/447 (101%)	449 (99%)	3 (1%)	84	93
1	B	449/447 (100%)	446 (99%)	3 (1%)	84	93
All	All	901/894 (101%)	895 (99%)	6 (1%)	84	93

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	8	ARG
1	A	46	ARG
1	A	52	ASP
1	B	358	ASP
1	B	449	ASP
1	B	454	LEU

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

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

Of 7 ligands modelled in this entry, 7 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

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	516/526 (98%)	0.29	26 (5%)	28	25	40, 64, 153, 211	0
1	B	513/526 (97%)	0.54	48 (9%)	8	7	35, 69, 179, 233	0
All	All	1029/1052 (97%)	0.42	74 (7%)	15	13	35, 66, 168, 233	0

All (74) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	499	GLY	12.1
1	B	502	ILE	8.9
1	B	495	ILE	7.1
1	B	448	ALA	7.0
1	B	470	VAL	6.5
1	B	459	PHE	5.5
1	A	495	ILE	5.2
1	B	525	LYS	4.7
1	B	523	LYS	4.4
1	A	475	ASN	4.3
1	B	460	THR	4.1
1	B	449	ASP	4.1
1	B	522	VAL	4.1
1	A	522	VAL	4.1
1	B	492	LYS	4.1
1	A	479	ILE	3.9
1	B	454	LEU	3.8
1	B	439	GLU	3.6
1	B	493	GLU	3.5
1	B	500	TYR	3.5
1	B	482	PHE	3.5
1	A	494	THR	3.5
1	A	446	ASP	3.5
1	A	445	VAL	3.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	468	GLU	3.4
1	B	501	VAL	3.3
1	B	506	LYS	3.3
1	B	524	ASN	3.3
1	A	457	ALA	3.3
1	B	2	ALA	3.2
1	B	480	ALA	3.1
1	B	463	ASN	3.1
1	A	270	LYS	3.1
1	A	452	THR	3.1
1	B	462	TYR	3.0
1	B	274	LEU	3.0
1	B	450	GLY	3.0
1	A	465	GLN	3.0
1	A	500	TYR	2.9
1	A	451	ASN	2.9
1	B	469	VAL	2.9
1	A	502	ILE	2.8
1	B	479	ILE	2.7
1	B	461	ILE	2.7
1	B	277	ALA	2.6
1	A	499	GLY	2.6
1	A	524	ASN	2.6
1	B	271	LYS	2.6
1	B	494	THR	2.6
1	B	491	TYR	2.6
1	B	472	GLY	2.5
1	A	501	VAL	2.5
1	A	523	LYS	2.5
1	A	443	THR	2.5
1	B	442	ILE	2.5
1	B	52	ASP	2.5
1	B	489	TYR	2.4
1	B	451	ASN	2.4
1	A	507	PHE	2.4
1	B	294	ASP	2.4
1	A	233	LYS	2.4
1	A	456	ASN	2.3
1	A	271	LYS	2.2
1	B	447	VAL	2.2
1	A	526	LYS	2.2
1	B	51	GLU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	232	GLU	2.2
1	B	496	ALA	2.2
1	B	507	PHE	2.2
1	B	509	PHE	2.1
1	B	503	ASN	2.1
1	A	175	ASN	2.1
1	B	471	LYS	2.1
1	A	453	LYS	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)

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	ZN	B	602	1/1	0.90	0.05	104,104,104,104	0
2	ZN	A	603	1/1	0.91	0.09	104,104,104,104	0
2	ZN	A	604	1/1	0.92	0.05	115,115,115,115	0
2	ZN	A	602	1/1	0.97	0.12	70,70,70,70	0
2	ZN	B	603	1/1	0.98	0.19	48,48,48,48	0
2	ZN	B	601	1/1	0.98	0.07	72,72,72,72	0
2	ZN	A	601	1/1	0.99	0.19	55,55,55,55	0

6.5 Other polymers (i)

There are no such residues in this entry.