WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2020 – 01:59 am BST

PDB ID	:	6FRG
Title	:	Crystal structure of G-1F mutant of Ssp DnaB Mini-Intein variant M86
Authors	:	Popp, M.A.; Blankenfeldt, W.; Friedel, K.; Mootz, H.D.
Deposited on	:	2018-02-15
Resolution	:	1.53 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 (i)) 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 1.53 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.

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
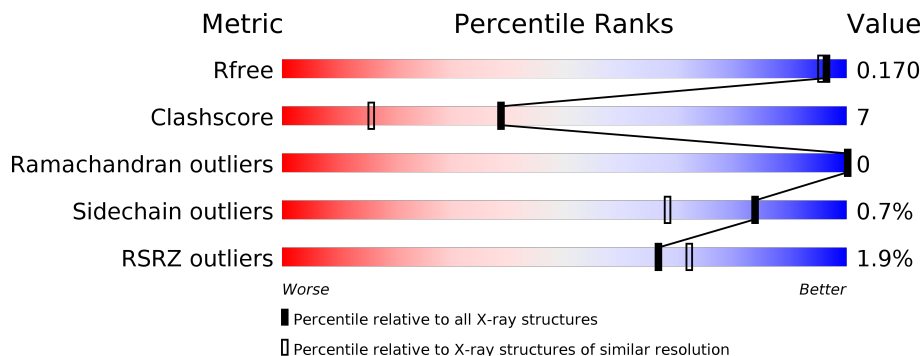
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2556 (1.56-1.52)
Clashscore	141614	2634 (1.56-1.52)
Ramachandran outliers	138981	2580 (1.56-1.52)
Sidechain outliers	138945	2577 (1.56-1.52)
RSRZ outliers	127900	2524 (1.56-1.52)

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	169	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	1PE	A	201	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3035 atoms, of which 1453 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 Replicative DNA helicase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	158	2647	832	1356	215	242	2	0	8	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	MET	-	initiating methionine	UNP Q55418
A	0	PHE	GLY	engineered mutation	UNP Q55418
A	1	ALA	CYS	engineered mutation	UNP Q55418
A	18	PRO	SER	engineered mutation	UNP Q55418
A	24	GLY	ASP	engineered mutation	UNP Q55418
A	58	THR	ILE	engineered mutation	UNP Q55418
A	107	ALA	-	linker	UNP Q55418
A	108	PRO	-	linker	UNP Q55418
A	109	GLU	-	linker	UNP Q55418
A	110	ILE	-	linker	UNP Q55418
A	111	GLU	-	linker	UNP Q55418
A	114	PRO	SER	engineered mutation	UNP Q55418
A	122	PRO	SER	engineered mutation	UNP Q55418
A	142	LEU	PRO	engineered mutation	UNP Q55418
A	143	ARG	HIS	engineered mutation	UNP Q55418
A	154	ALA	ASN	engineered mutation	UNP Q55418
A	160	LYS	-	expression tag	UNP Q55418
A	161	LEU	-	expression tag	UNP Q55418
A	162	GLY	-	expression tag	UNP Q55418
A	163	GLY	-	expression tag	UNP Q55418

- Molecule 2 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula: $C_{10}H_{22}O_6$).
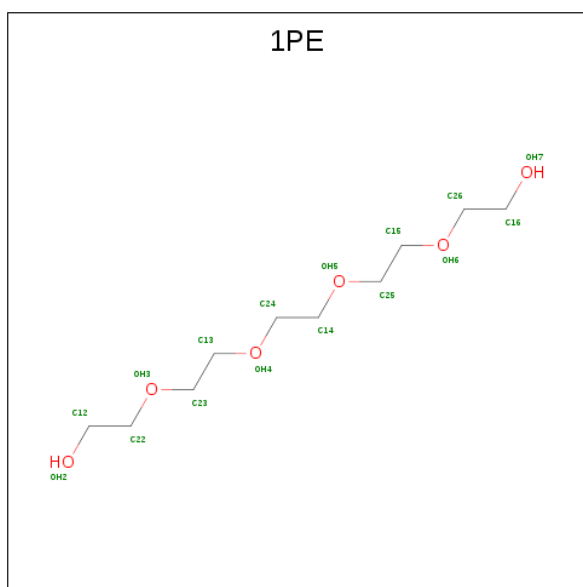
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 38	C 10	H 22	O 6	0	0
2	A	1	Total 38	C 10	H 22	O 6	0	0

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).
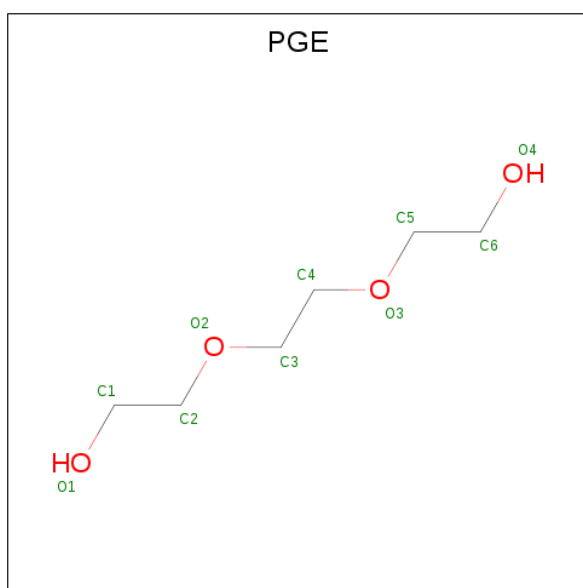
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 23	C 6	H 13	O 4	0	0

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

PEG

C4

OH O4

O2

O

C3

C1

C2

HO

O1

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 17	C 4	H 10	O 3	0	0
4	A	1	Total 17	C 4	H 10	O 3	0	0
4	A	1	Total 17	C 4	H 10	O 3	0	0
4	A	1	Total 17	C 4	H 10	O 3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	221	Total 221	O 221	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 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: Replicative DNA helicase
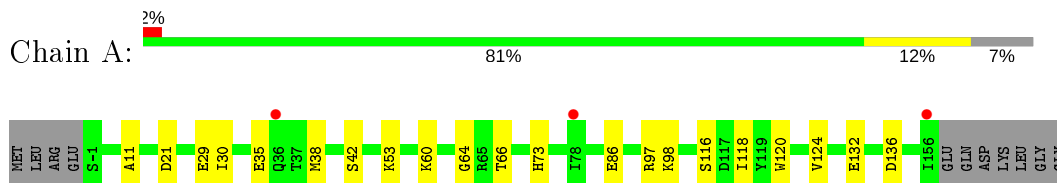
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 42 2 2	Depositor
Cell constants a, b, c, α, β, γ	71.47Å 71.47Å 87.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.94 – 1.53 43.94 – 1.54	Depositor EDS
% Data completeness (in resolution range)	100.0 (43.94-1.53) 100.0 (43.94-1.54)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.11 (at 1.54Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.143 , 0.170 0.144 , 0.170	Depositor DCC
R_{free} test set	1735 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	20.9	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 57.4	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	0.98	EDS
Total number of atoms	3035	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.66% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: PEG, PGE, 1PE

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	0.88	6/1338 (0.4%)	0.94	2/1808 (0.1%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	97	ARG	CB-CG	-7.75	1.31	1.52
1	A	86	GLU	CD-OE2	5.84	1.32	1.25
1	A	132	GLU	CD-OE2	-5.25	1.19	1.25
1	A	116	SER	CB-OG	-5.14	1.35	1.42
1	A	132	GLU	CB-CG	-5.13	1.42	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	21	ASP	CB-CG-OD1	5.19	122.97	118.30
1	A	97	ARG	CG-CD-NE	-5.04	101.22	111.80

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	1291	1356	1358	20	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	32	44	44	8	0
3	A	10	13	14	2	0
4	A	28	40	40	3	0
5	A	221	0	0	3	2
All	All	1582	1453	1456	21	2

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 7.

The worst 5 of 21 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:60:LYS:HZ3	2:A:201:1PE:H251	1.32	0.94
1:A:73[B]:HIS:NE2	5:A:301:HOH:O	2.16	0.79
1:A:73[B]:HIS:CE1	5:A:301:HOH:O	2.45	0.68
1:A:60:LYS:HZ1	2:A:201:1PE:H231	1.63	0.63
1:A:60:LYS:NZ	2:A:201:1PE:H242	2.26	0.51

All (2) 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)
5:A:303:HOH:O	5:A:350:HOH:O[7_556]	1.90	0.30
5:A:309:HOH:O	5:A:440:HOH:O[7_556]	1.94	0.26

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	164/169 (97%)	160 (98%)	4 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	148/149 (99%)	147 (99%)	1 (1%)	84	68

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	136	ASP

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)

7 ligands are modelled in this entry.

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	$\#|Z| > 2$
2	1PE	A	201	-	15,15,15	0.49	0	14,14,14	0.75	1 (7%)
3	PGE	A	202	-	9,9,9	0.49	0	8,8,8	1.38	1 (12%)
4	PEG	A	206	-	6,6,6	0.47	0	5,5,5	0.34	0
2	1PE	A	203	-	15,15,15	0.51	0	14,14,14	0.30	0
4	PEG	A	205	-	6,6,6	0.53	0	5,5,5	0.37	0
4	PEG	A	204	-	6,6,6	0.55	0	5,5,5	0.67	0
4	PEG	A	207	-	6,6,6	0.34	0	5,5,5	1.07	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
2	1PE	A	201	-	-	9/13/13/13	-
3	PGE	A	202	-	-	3/7/7/7	-
4	PEG	A	206	-	-	3/4/4/4	-
2	1PE	A	203	-	-	5/13/13/13	-
4	PEG	A	205	-	-	1/4/4/4	-
4	PEG	A	204	-	-	3/4/4/4	-
4	PEG	A	207	-	-	3/4/4/4	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	202	PGE	C3-O2-C2	-2.68	101.68	113.29
2	A	201	1PE	OH6-C26-C16	2.18	119.64	110.07

There are no chirality outliers.

5 of 27 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	207	PEG	O2-C3-C4-O4
2	A	201	1PE	OH2-C12-C22-OH3
2	A	203	1PE	OH2-C12-C22-OH3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	A	203	1PE	OH7-C16-C26-OH6
4	A	207	PEG	O1-C1-C2-O2

There are no ring outliers.

5 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	201	1PE	7	0
3	A	202	PGE	2	0
2	A	203	1PE	1	0
4	A	204	PEG	1	0
4	A	207	PEG	2	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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	158/169 (93%)	-0.37	3 (1%)	66	72	16, 23, 40, 64	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	156	ILE	8.0
1	A	78	ILE	2.6
1	A	36	GLN	2.6

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, 95th 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	1PE	A	203	16/16	0.70	0.18	72,99,121,128	0
4	PEG	A	206	7/7	0.71	0.22	89,107,120,120	0
4	PEG	A	205	7/7	0.81	0.20	53,67,86,86	0
3	PGE	A	202	10/10	0.87	0.22	34,62,77,93	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
4	PEG	A	207	7/7	0.90	0.11	32,50,69,82	0
4	PEG	A	204	7/7	0.92	0.14	39,47,58,58	17
2	1PE	A	201	16/16	0.93	0.14	21,39,96,97	38

6.5 Other polymers (i)

There are no such residues in this entry.