



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 17, 2017 – 07:58 AM EDT

PDB ID : 4W6N  
Title : Crystal Structure of Full-Length Split GFP Mutant D117C Disulfide Dimer,  
C 1 2 1 Space Group  
Authors : Leibly, D.J.; Waldo, G.S.; Yeates, T.O.  
Deposited on : unknown  
Resolution : 3.38 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029824

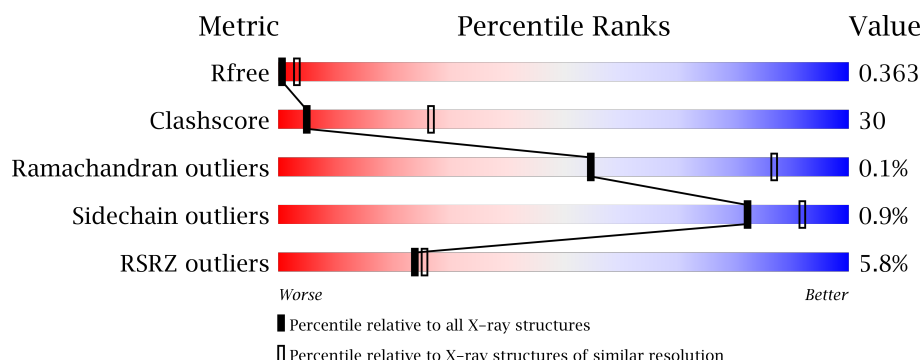
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.38 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))
$R_{free}$	100719	1241 (3.46-3.30)
Clashscore	112137	1319 (3.46-3.30)
Ramachandran outliers	110173	1298 (3.46-3.30)
Sidechain outliers	110143	1297 (3.46-3.30)
RSRZ outliers	101464	1251 (3.46-3.30)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . 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	234	
1	B	234	
1	C	234	
1	D	234	
1	E	234	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	234	<div><div></div><div>6%</div><div>53%</div><div>35%</div><div>• 10%</div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 10419 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.

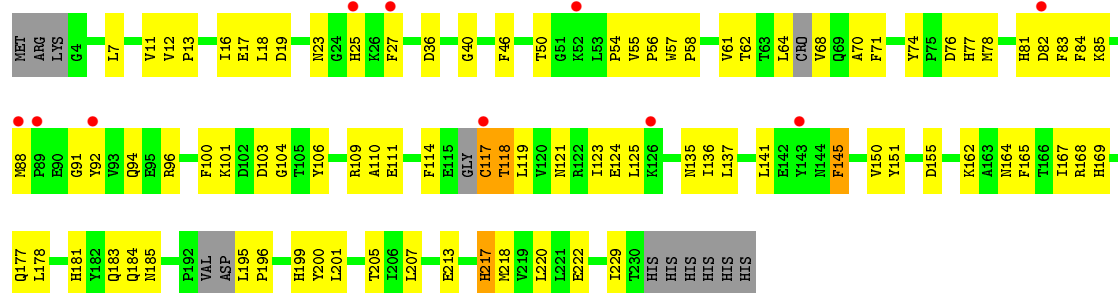
- Molecule 1 is a protein called fluorescent protein D117C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	217	Total	C	N	O	S	0	0	0
			1711	1085	291	331	4			
1	B	218	Total	C	N	O	S	0	0	0
			1758	1121	298	336	3			
1	C	219	Total	C	N	O	S	0	0	0
			1743	1109	296	334	4			
1	D	221	Total	C	N	O	S	0	0	0
			1759	1121	299	335	4			
1	E	221	Total	C	N	O	S	0	0	0
			1759	1120	299	337	3			
1	F	210	Total	C	N	O	S	0	0	0
			1689	1072	289	324	4			

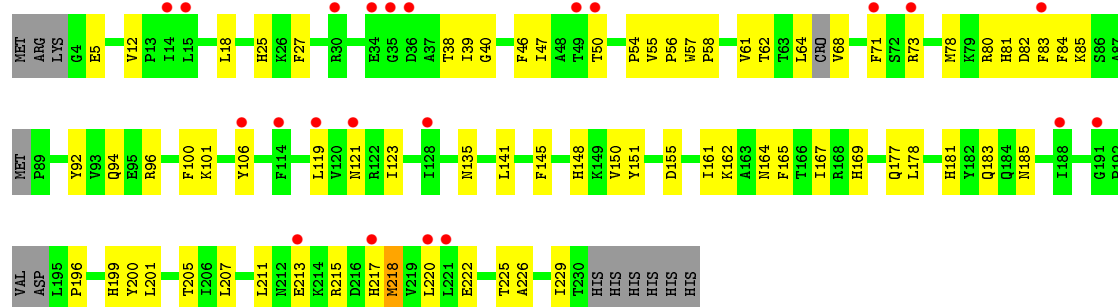




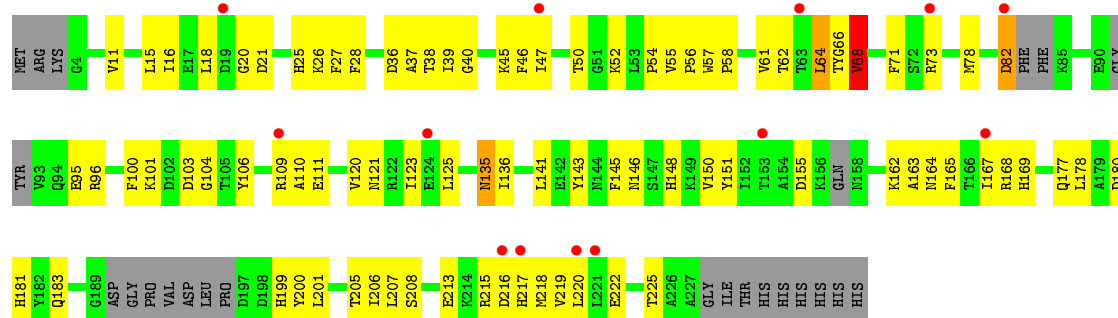
• Molecule 1: fluorescent protein D117C



• Molecule 1: fluorescent protein D117C



• Molecule 1: fluorescent protein D117C



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	181.21Å 102.68Å 84.13Å 90.00° 101.44° 90.00°	Depositor
Resolution (Å)	88.89 – 3.38 88.89 – 3.38	Depositor EDS
% Data completeness (in resolution range)	97.8 (88.89-3.38) 97.9 (88.89-3.38)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.20	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.56 (at 3.41Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.316 , 0.363 0.316 , 0.363	Depositor DCC
$R_{free}$ test set	2095 reflections (10.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	83.9	Xtrriage
Anisotropy	0.623	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 26.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.84	EDS
Total number of atoms	10419	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	99.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.27% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: CRO

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	0.42	1/1720 (0.1%)	0.71	2/2323 (0.1%)
1	B	0.40	1/1772 (0.1%)	0.67	0/2391
1	C	0.43	2/1757 (0.1%)	0.68	1/2373 (0.0%)
1	D	0.36	1/1796 (0.1%)	0.67	0/2426
1	E	0.34	0/1796	0.66	1/2425 (0.0%)
1	F	0.45	2/1697 (0.1%)	0.84	5/2286 (0.2%)
All	All	0.40	7/10538 (0.1%)	0.71	9/14224 (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	0	2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	64	LEU	C-O	9.41	1.41	1.23
1	F	68	VAL	N-CA	-7.48	1.31	1.46
1	C	68	VAL	N-CA	-6.96	1.32	1.46
1	B	68	VAL	N-CA	-6.77	1.32	1.46
1	D	117	CYS	CB-SG	-6.02	1.72	1.82

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	64	LEU	CA-C-O	-18.68	80.86	120.10
1	F	64	LEU	CB-CA-C	-14.55	82.55	110.20

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	68	VAL	N-CA-CB	8.92	131.12	111.50
1	A	68	VAL	N-CA-CB	8.36	129.89	111.50
1	F	82	ASP	CB-CA-C	-5.75	98.91	110.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	116	GLY	Mainchain,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	1711	0	1650	106	0
1	B	1758	0	1705	88	0
1	C	1743	0	1686	128	0
1	D	1759	0	1719	108	0
1	E	1759	0	1718	85	0
1	F	1689	0	1646	116	0
All	All	10419	0	10124	617	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 30.

The worst 5 of 617 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:119:LEU:CD2	1:A:120:VAL:H	1.49	1.25
1:A:113:LYS:O	1:A:119:LEU:HD23	1.29	1.25
1:C:218:MET:HE1	1:C:220:LEU:CD1	1.66	1.24
1:C:28:PHE:CE1	1:C:30:ARG:HG3	1.75	1.19
1:F:143:TYR:CZ	1:F:207:LEU:HB3	1.80	1.15

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	204/234 (87%)	192 (94%)	11 (5%)	1 (0%)	32	70
1	B	207/234 (88%)	195 (94%)	12 (6%)	0	100	100
1	C	212/234 (91%)	203 (96%)	9 (4%)	0	100	100
1	D	213/234 (91%)	203 (95%)	10 (5%)	0	100	100
1	E	213/234 (91%)	203 (95%)	10 (5%)	0	100	100
1	F	197/234 (84%)	191 (97%)	6 (3%)	0	100	100
All	All	1246/1404 (89%)	1187 (95%)	58 (5%)	1 (0%)	55	87

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	117	CYS

### 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	182/203 (90%)	179 (98%)	3 (2%)	68	85
1	B	189/203 (93%)	189 (100%)	0	100	100
1	C	185/203 (91%)	183 (99%)	2 (1%)	78	89
1	D	191/203 (94%)	188 (98%)	3 (2%)	68	85
1	E	191/203 (94%)	191 (100%)	0	100	100
1	F	182/203 (90%)	180 (99%)	2 (1%)	78	89

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	1120/1218 (92%)	1110 (99%)	10 (1%)	82	92

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	165	PHE
1	D	118	THR
1	D	217	HIS
1	C	28	PHE
1	D	145	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	199	HIS
1	D	94	GLN
1	F	148	HIS
1	C	148	HIS
1	F	81	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues 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$
1	CRO	A	66	1	23,23,24	3.57	5 (21%)	27,32,34	4.01	7 (25%)
1	CRO	B	66	1	23,23,24	3.58	5 (21%)	27,32,34	3.99	6 (22%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	CRO	C	66	1	23,23,24	3.46	4 (17%)	27,32,34	4.00	7 (25%)
1	CRO	F	66	1	23,23,24	3.61	5 (21%)	27,32,34	4.00	7 (25%)

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
1	CRO	A	66	1	-	0/12/31/32	0/2/2/2
1	CRO	B	66	1	-	0/12/31/32	0/2/2/2
1	CRO	C	66	1	-	0/12/31/32	0/2/2/2
1	CRO	F	66	1	-	0/12/31/32	0/2/2/2

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	66	CRO	CA1-N1	-4.60	1.32	1.47
1	B	66	CRO	CA1-N1	-4.29	1.33	1.47
1	A	66	CRO	CA1-N1	-3.84	1.35	1.47
1	B	66	CRO	C2-N3	-2.75	1.33	1.39
1	A	66	CRO	C2-N3	-2.75	1.33	1.39

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	66	CRO	O2-C2-CA2	-10.10	125.53	130.97
1	A	66	CRO	O2-C2-CA2	-10.09	125.53	130.97
1	C	66	CRO	O2-C2-CA2	-10.07	125.54	130.97
1	F	66	CRO	O2-C2-CA2	-9.99	125.59	130.97
1	F	66	CRO	CG2-CB2-CA2	-8.81	119.95	130.19

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 20 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	66	CRO	6	0
1	B	66	CRO	5	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	66	CRO	6	0
1	F	66	CRO	3	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	216/234 (92%)	0.37	13 (6%) 23 24	41, 89, 147, 201	0
1	B	217/234 (92%)	0.23	7 (3%) 48 50	38, 84, 147, 178	0
1	C	218/234 (93%)	0.42	10 (4%) 33 34	51, 88, 153, 201	0
1	D	221/234 (94%)	0.40	10 (4%) 34 35	55, 99, 162, 200	0
1	E	221/234 (94%)	0.63	22 (9%) 8 8	48, 96, 157, 197	0
1	F	209/234 (89%)	0.51	13 (6%) 21 23	65, 106, 158, 183	0
All	All	1302/1404 (92%)	0.43	75 (5%) 24 25	38, 94, 155, 201	0

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	117	CYS	4.3
1	E	36	ASP	3.8
1	F	73	ARG	3.5
1	E	213	GLU	3.4
1	E	121	ASN	3.4

### 6.2 Non-standard residues in protein, DNA, RNA chains [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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
1	CRO	C	66	22/23	0.82	0.25	-	65,66,68,71	0
1	CRO	B	66	22/23	0.91	0.25	-	60,65,80,81	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
1	CRO	A	66	22/23	0.91	0.30	-	84,89,94,95	0
1	CRO	F	66	22/23	0.91	0.19	-	63,64,67,67	0

### 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.