

# Full wwPDB X-ray Structure Validation Report (i)

#### Oct 13, 2022 - 01:42 am BST

PDB ID	:	707W
Title	:	Crystal structure of rsEGFP2 mutant V151L in the non-fluorescent off-state
		the determined by serial femtosecond crystallography at room temperature
Authors	:	Hadjidemetriou, K.; Coquelle, N.; Barends, T.R.M.; Schlichting, I.; Colletier,
		JP.; Weik, M.
Deposited on		
Resolution	:	2.10  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:

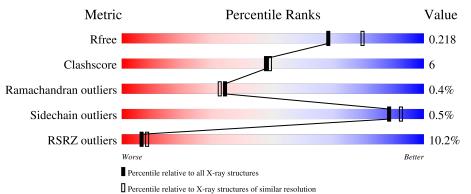
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as $541$ be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

## 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 2.10 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	5197(2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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				
			10%				
1	А	250	83%	11% • 5%			



#### 707W

## 2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2289 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 Green fluorescent protein.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	А	237	Total 2233	C 1416	N 376	0 434	${ m S} 7$	0	40	0

Chain	Residue	Modelled	Actual	Comment	Reference
А	-12	MET	-	initiating methionine	UNP P42212
А	-11	ARG	-	expression tag	UNP P42212
А	-10	GLY	-	expression tag	UNP P42212
А	-9	SER	-	expression tag	UNP P42212
А	-8	HIS	-	expression tag	UNP P42212
А	-7	HIS	-	expression tag	UNP P42212
А	-6	HIS	-	expression tag	UNP P42212
А	-5	HIS	-	expression tag	UNP P42212
А	-4	HIS	-	expression tag	UNP P42212
А	-3	HIS	-	expression tag	UNP P42212
А	-2	THR	-	expression tag	UNP P42212
А	-1	ASP	-	expression tag	UNP P42212
А	0	PRO	-	expression tag	UNP P42212
А	2	VAL	-	insertion	UNP P42212
А	65	LEU	PHE	engineered mutation	UNP P42212
А	68	PIA	SER	chromophore	UNP P42212
А	68	PIA	TYR	chromophore	UNP P42212
А	68	PIA	GLY	chromophore	UNP P42212
А	70	LEU	GLN	engineered mutation	UNP P42212
А	151	LEU	VAL	engineered mutation	UNP P42212
А	164	SER	VAL	engineered mutation	UNP P42212
А	207	LYS	ALA	engineered mutation	UNP P42212
А	232	LEU	HIS	engineered mutation	UNP P42212

There are 23 discrepancies between the modelled and reference sequences:

• Molecule 2 is water.

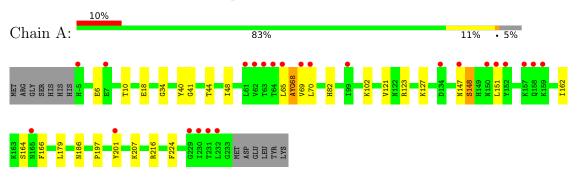


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	56	$\begin{array}{cc} {\rm Total} & {\rm O} \\ 56 & 56 \end{array}$	0	7



## 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: Green fluorescent protein



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	51.85Å 62.93Å 72.04Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	31.28 – 2.10	Depositor
Resolution (A)	31.26 - 2.10	EDS
% Data completeness	99.9 (31.28-2.10)	Depositor
(in resolution range)	100.0 (31.26-2.10)	EDS
R <sub>merge</sub>	(Not available)	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.84 (at 2.10 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
D D.	0.173 , $0.213$	Depositor
$R, R_{free}$	0.179 , $0.218$	DCC
$R_{free}$ test set	702 reflections $(4.91\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	39.4	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	(Not available), (Not available)	EDS
L-test for twinning <sup>2</sup>	$ L  > = 0.45, < L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2289	wwPDB-VP
Average B, all atoms $(Å^2)$	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.43% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: PIA

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		
	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.66	0/2249	0.82	0/3041	

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 (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	А	2233	0	2140	27	1
2	А	56	0	0	5	0
All	All	2289	0	2140	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 6.

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:A:148[B]:SER:HA	2:A:301[B]:HOH:O	1.52	1.08
1:A:147[B]:ASN:O	2:A:301[B]:HOH:O	1.75	1.04

Continued on next page...



7	Ο	7	W	
•	$\sim$	•	••	

		Interatomic	Clash
Atom-1	Atom-2	distance $(\text{\AA})$	overlap (Å)
1:A:68[C]:PIA:HE2	1:A:151[C]:LEU:HD21	1.72	0.72
1:A:148[B]:SER:CA	2:A:301[B]:HOH:O	2.23	0.70
1:A:68[B]:PIA:N2	1:A:68[B]:PIA:HD1	2.11	0.65
1:A:207[B]:LYS:HE2	2:A:341:HOH:O	1.96	0.64
1:A:151[B]:LEU:CD1	1:A:166:PHE:CD1	2.81	0.64
1:A:151[B]:LEU:HD13	1:A:166:PHE:CD1	2.33	0.63
1:A:147[B]:ASN:C	2:A:301[B]:HOH:O	2.34	0.56
1:A:68[C]:PIA:HD2	1:A:68[C]:PIA:O2	2.07	0.54
1:A:68[B]:PIA:HB2	1:A:70:LEU:HD21	1.91	0.51
1:A:68[C]:PIA:CE2	1:A:151[C]:LEU:HD21	2.40	0.51
1:A:151[C]:LEU:HD23	1:A:166:PHE:CD1	2.46	0.50
1:A:121:VAL:HG12	1:A:123[A]:ARG:HG3	1.95	0.48
1:A:151[B]:LEU:HD11	1:A:164:SER:OG	2.14	0.47
1:A:102:LYS:HD2	1:A:179[A]:LEU:HD23	1.96	0.47
1:A:68[C]:PIA:CE2	1:A:151[C]:LEU:HD11	2.45	0.46
1:A:48:ILE:HD13	1:A:216:ARG:CZ	2.47	0.44
1:A:41:GLY:O	1:A:224:PHE:HA	2.18	0.43
1:A:6:GLU:O	1:A:10[A]:THR:HG23	2.18	0.43
1:A:151[B]:LEU:CD1	1:A:164:SER:OG	2.66	0.43
1:A:65[B]:LEU:O	1:A:68[B]:PIA:HA31	2.19	0.42
1:A:151[C]:LEU:O	1:A:201:TYR:HA	2.21	0.41
1:A:69[B]:VAL:O	1:A:69[B]:VAL:HG23	2.21	0.41
1:A:82:HIS:O	1:A:197:PRO:HB3	2.21	0.41
1:A:34:GLY:HA3	1:A:44:THR:O	2.21	0.40
1:A:162[A]:ILE:HG13	1:A:186:ASN:HB2	2.02	0.40

Continued from previous page...

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)
1:A:18[B]:GLU:OE1	1:A:40:TYR:OH[4_545]	1.88	0.32

### 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	Favoured Allowed		Percentiles
1	А	270/250~(108%)	265~(98%)	3 (1%)	2(1%)	22 18

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	148[B]	SER
1	А	148[C]	SER

#### 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	А	248/220~(113%)	247 (100%)	1 (0%)	91 94		

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

Mol	Chain	Res	Type
1	А	127	LYS

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such side chains are listed below:

Mol	Chain	Res	Type
1	А	171	ASN
1	А	205	GLN

#### 5.3.3 RNA (i)

There are no RNA molecules in this entry.



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

2 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	Turne	Chain	Res	Link	Bo	ond leng	$\mathbf{ths}$	В	ond ang	les
INIOI	Type	Unam	nes	LIIIK	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
1	PIA	А	68[C]	1	21,21,22	2.84	5 (23%)	27,29,31	4.37	9 (33%)
1	PIA	А	68[B]	1	21,21,22	3.69	4 (19%)	27,29,31	3.98	6 (22%)

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	PIA	А	68[C]	1	-	2/8/27/28	0/2/2/2
1	PIA	А	68[B]	1	-	4/8/27/28	0/2/2/2

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	$\mathrm{Ideal}(\mathrm{\AA})$
1	А	68[B]	PIA	CA2-C2	-15.75	1.33	1.48
1	А	68[C]	PIA	CA2-C2	-11.05	1.37	1.48
1	А	68[C]	PIA	CA3-N3	3.73	1.54	1.47
1	А	68[B]	PIA	CG2-CB2	-3.39	1.40	1.46
1	А	68[B]	PIA	C2-N3	-2.91	1.33	1.39
1	А	68[C]	PIA	CG2-CB2	-2.69	1.41	1.46
1	А	68[C]	PIA	C2-N3	-2.68	1.33	1.39
1	А	68[B]	PIA	CA2-N2	-2.16	1.33	1.38
1	А	68[C]	PIA	CB2-CA2	2.15	1.36	1.35

All (9) bond length outliers are listed below:

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
1	А	68[B]	PIA	CA2-C2-N3	13.31	109.67	103.37
1	А	68[C]	PIA	CA2-C2-N3	12.70	109.38	103.37

Continued on next page...



Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
1	А	68[C]	PIA	CB2-CA2-C2	12.41	137.09	122.28
1	А	68[B]	PIA	O2-C2-CA2	-12.08	124.18	130.96
1	А	68[C]	PIA	CB2-CA2-N2	-9.68	115.41	128.83
1	А	68[C]	PIA	C2-N3-C1	-7.23	104.31	107.97
1	А	68[B]	PIA	C2-N3-C1	-6.25	104.80	107.97
1	А	68[B]	PIA	CG2-CB2-CA2	-6.12	122.44	129.94
1	А	68[C]	PIA	O3-C3-CA3	-4.48	112.88	126.39
1	А	68[B]	PIA	O3-C3-CA3	-3.21	116.71	126.39
1	А	68[B]	PIA	CA3-N3-C2	2.64	129.86	123.80
1	А	68[C]	PIA	CA3-N3-C2	2.37	129.23	123.80
1	А	68[C]	PIA	O2-C2-CA2	-2.17	129.74	130.96
1	А	68[C]	PIA	N3-C1-N2	2.12	112.92	111.45
1	А	68[C]	PIA	C2-CA2-N2	-2.09	107.47	108.93

Continued from previous page...

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	А	68[B]	PIA	N2-CA2-CB2-CG2
1	А	68[B]	PIA	C2-CA2-CB2-CG2
1	А	68[C]	PIA	C2-CA2-CB2-CG2
1	А	68[C]	PIA	N2-CA2-CB2-CG2
1	А	68[B]	PIA	C3-CA3-N3-C2
1	А	68[B]	PIA	N2-C1-CA1-CB1

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	А	68[C]	PIA	4	0
1	А	68[B]	PIA	3	0

### 5.5 Carbohydrates (i)

There are no monosaccharides 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,  $95^{th}$  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	А	236/250~(94%)	0.59	24 (10%)	6	8	26, 37, 55, 93	1 (0%)

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ	
1	А	232	LEU	5.2	
1	А	65[B]	LEU	5.1	
1	А	152[B]	TYR	5.1	
1	А	231	THR	5.1	
1	А	201	TYR	4.2	
1	А	62	VAL	4.1	
1	А	69[B]	VAL	3.9	
1	А	147[B]	ASN	3.8	
1	А	134[A]	ASP	3.7	
1	А	158	GLN	3.6	
1	А	-5	HIS	3.5	
1	А	70	LEU	3.3	
1	А	63	THR	2.8	
1	А	61	LEU	2.8	
1	А	99	ILE	2.7	
1	А	151[B]	LEU	2.7	
1	А	64	THR	2.5	
1	А	229	GLY	2.3	
1	А	150[B]	ASN	2.3	
1	А	159	LYS	2.3	
1	А	165[A]	ASN	2.2	
1	А	7	GLU	2.1	
1	А	230	ILE	2.1	
1	А	157[A]	LYS	2.0	



### 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. 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	$\mathbf{B} ext{-factors}(\mathrm{\AA}^2)$	Q < 0.9
1	PIA	А	68[B]	20/21	0.95	0.28	28,31,36,39	20
1	PIA	А	68[C]	20/21	0.95	0.28	25,28,31,32	20

### 6.3 Carbohydrates (i)

There are no monosaccharides 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.

