



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 26, 2024 – 02:09 PM JST

PDB ID : 8JKG  
Title : Crystal structure of the Green fluorescent protein SE\_A277 variant at pH 5.5  
Authors : Shin, S.C.  
Deposited on : 2023-06-01  
Resolution : 2.20 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

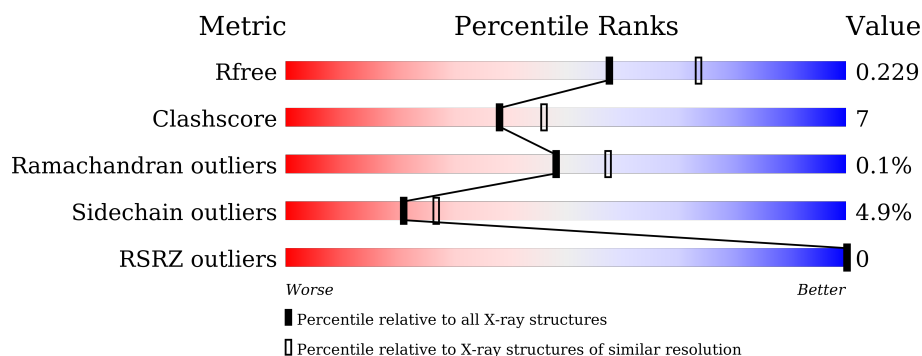
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
$R_{free}$	164625	5791 (2.20-2.20)
Clashscore	180529	6634 (2.20-2.20)
Ramachandran outliers	177936	6560 (2.20-2.20)
Sidechain outliers	177891	6561 (2.20-2.20)
RSRZ outliers	164620	5791 (2.20-2.20)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	236	<div> <div>74%</div> <div>19%</div> <div>• 5%</div> </div>
1	B	236	<div> <div>84%</div> <div>9%</div> <div>• 5%</div> </div>
1	C	236	<div> <div>81%</div> <div>11%</div> <div>• 5%</div> </div>
1	D	236	<div> <div>82%</div> <div>12%</div> <div>• 5%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7749 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	A	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			
1	B	225	Total	C	N	O	S	0	0	0
			1806	1152	304	344	6			
1	C	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			
1	D	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	64	LEU	PHE	engineered mutation	UNP P42212
A	66	CRO	SER	chromophore	UNP P42212
A	66	CRO	TYR	chromophore	UNP P42212
A	66	CRO	GLY	chromophore	UNP P42212
A	80	ARG	GLN	engineered mutation	UNP P42212
A	147	ASP	SER	engineered mutation	UNP P42212
A	149	GLN	ASN	engineered mutation	UNP P42212
A	163	ALA	VAL	engineered mutation	UNP P42212
A	175	GLY	SER	engineered mutation	UNP P42212
A	202	PHE	SER	engineered mutation	UNP P42212
A	204	THR	GLN	engineered mutation	UNP P42212
A	206	THR	ALA	engineered mutation	UNP P42212
B	64	LEU	PHE	engineered mutation	UNP P42212
B	66	CRO	SER	chromophore	UNP P42212
B	66	CRO	TYR	chromophore	UNP P42212
B	66	CRO	GLY	chromophore	UNP P42212
B	80	ARG	GLN	engineered mutation	UNP P42212
B	147	ASP	SER	engineered mutation	UNP P42212
B	149	GLN	ASN	engineered mutation	UNP P42212
B	163	ALA	VAL	engineered mutation	UNP P42212
B	175	GLY	SER	engineered mutation	UNP P42212

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	202	PHE	SER	engineered mutation	UNP P42212
B	204	THR	GLN	engineered mutation	UNP P42212
B	206	THR	ALA	engineered mutation	UNP P42212
C	64	LEU	PHE	engineered mutation	UNP P42212
C	66	CRO	SER	chromophore	UNP P42212
C	66	CRO	TYR	chromophore	UNP P42212
C	66	CRO	GLY	chromophore	UNP P42212
C	80	ARG	GLN	engineered mutation	UNP P42212
C	147	ASP	SER	engineered mutation	UNP P42212
C	149	GLN	ASN	engineered mutation	UNP P42212
C	163	ALA	VAL	engineered mutation	UNP P42212
C	175	GLY	SER	engineered mutation	UNP P42212
C	202	PHE	SER	engineered mutation	UNP P42212
C	204	THR	GLN	engineered mutation	UNP P42212
C	206	THR	ALA	engineered mutation	UNP P42212
D	64	LEU	PHE	engineered mutation	UNP P42212
D	66	CRO	SER	chromophore	UNP P42212
D	66	CRO	TYR	chromophore	UNP P42212
D	66	CRO	GLY	chromophore	UNP P42212
D	80	ARG	GLN	engineered mutation	UNP P42212
D	147	ASP	SER	engineered mutation	UNP P42212
D	149	GLN	ASN	engineered mutation	UNP P42212
D	163	ALA	VAL	engineered mutation	UNP P42212
D	175	GLY	SER	engineered mutation	UNP P42212
D	202	PHE	SER	engineered mutation	UNP P42212
D	204	THR	GLN	engineered mutation	UNP P42212
D	206	THR	ALA	engineered mutation	UNP P42212

- Molecule 2 is water.

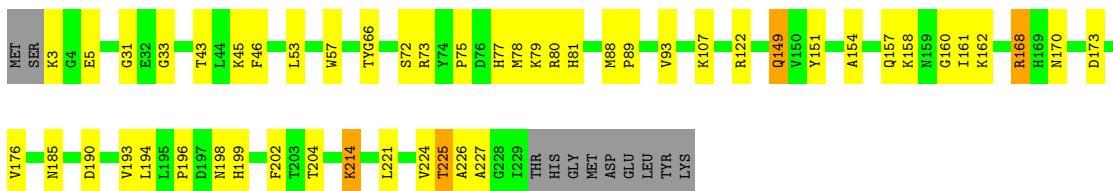
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	124	Total O 124 124	0	0
2	B	138	Total O 138 138	0	0
2	C	147	Total O 147 147	0	0
2	D	113	Total O 113 113	0	0

### 3 Residue-property plots


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

Chain A: 




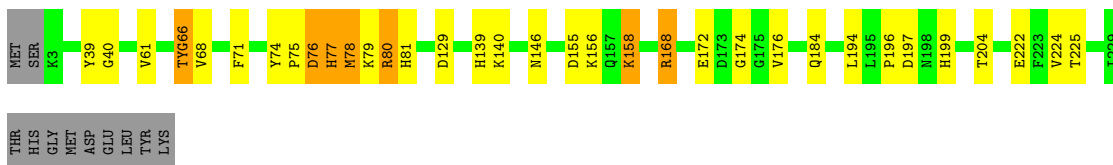
- Molecule 1: Green fluorescent protein

Chain B: 




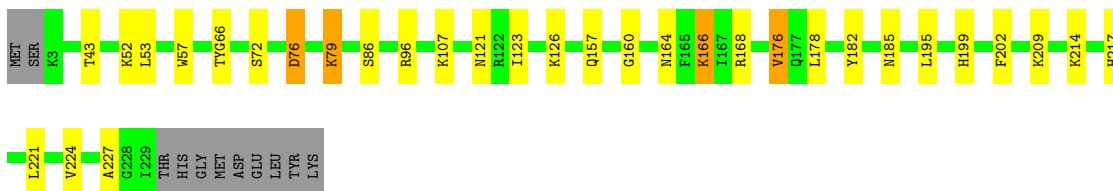
- Molecule 1: Green fluorescent protein

Chain C: 



- Molecule 1: Green fluorescent protein

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.38Å 79.98Å 179.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.57 – 2.20 40.57 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.1 (40.57-2.20) 97.1 (40.57-2.20)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.18 (at 2.20Å)	Xtriage
Refinement program	PHENIX (1.19.2_4158: ???)	Depositor
R, $R_{free}$	0.187 , 0.230 0.190 , 0.229	Depositor DCC
$R_{free}$ test set	52958 reflections (3.64%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	24.1	Xtriage
Anisotropy	0.383	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 22.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.114 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7749	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.01% 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	0.45	0/1826	0.64	0/2467
1	B	0.47	0/1825	0.63	0/2465
1	C	0.52	1/1826 (0.1%)	0.65	0/2467
1	D	0.45	0/1826	0.64	0/2467
All	All	0.47	1/7303 (0.0%)	0.64	0/9866

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	39	TYR	CB-CG	-5.14	1.44	1.51

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	A	1807	0	1745	37	0
1	B	1806	0	1741	16	0
1	C	1807	0	1746	29	0
1	D	1807	0	1745	20	0
2	A	124	0	0	2	0
2	B	138	0	0	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	147	0	0	4	0
2	D	113	0	0	3	0
All	All	7749	0	6977	97	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 7.

The worst 5 of 97 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:73:ARG:HB3	1:A:225:THR:HB	1.58	0.85
1:B:214:LYS:HD3	1:B:214:LYS:H	1.40	0.84
1:D:76:ASP:HA	1:D:79:LYS:HG3	1.58	0.82
1:A:168:ARG:HG3	1:A:176:VAL:HG11	1.69	0.74
1:C:77:HIS:O	1:C:80:ARG:NH2	2.21	0.73

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	220/236 (93%)	213 (97%)	7 (3%)	0	100	100
1	B	220/236 (93%)	216 (98%)	4 (2%)	0	100	100
1	C	220/236 (93%)	214 (97%)	6 (3%)	0	100	100
1	D	220/236 (93%)	215 (98%)	4 (2%)	1 (0%)	25	28
All	All	880/944 (93%)	858 (98%)	21 (2%)	1 (0%)	48	57

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	D	79	LYS

### 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	193/205 (94%)	184 (95%)	9 (5%)	22	29
1	B	192/205 (94%)	182 (95%)	10 (5%)	19	24
1	C	193/205 (94%)	181 (94%)	12 (6%)	15	18
1	D	193/205 (94%)	186 (96%)	7 (4%)	30	40
All	All	771/820 (94%)	733 (95%)	38 (5%)	21	27

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	184	GLN
1	D	166	LYS
1	C	224	VAL
1	D	86	SER
1	D	214	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	149	GLN
1	A	170	ASN
1	B	81	HIS
1	C	81	HIS
1	C	184	GLN

### 5.3.3 RNA [i](#)

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	C	66	1	23,23,24	6.75	8 (34%)	30,32,34	5.00	11 (36%)
1	CRO	B	66	1	23,23,24	6.09	11 (47%)	30,32,34	3.14	9 (30%)
1	CRO	A	66	1	23,23,24	6.04	10 (43%)	30,32,34	3.17	9 (30%)
1	CRO	D	66	1	23,23,24	6.08	10 (43%)	30,32,34	3.15	10 (33%)

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	C	66	1	-	5/12/31/32	0/2/2/2
1	CRO	B	66	1	-	1/12/31/32	0/2/2/2
1	CRO	A	66	1	-	1/12/31/32	0/2/2/2
1	CRO	D	66	1	-	1/12/31/32	0/2/2/2

The worst 5 of 39 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	66	CRO	CB2-CA2	24.41	1.55	1.35
1	B	66	CRO	CB2-CA2	20.88	1.52	1.35
1	D	66	CRO	CB2-CA2	20.79	1.52	1.35
1	A	66	CRO	CB2-CA2	20.69	1.52	1.35
1	C	66	CRO	CA2-C2	-13.37	1.35	1.48

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	66	CRO	CA2-C2-N3	14.40	110.18	103.37
1	C	66	CRO	O2-C2-CA2	-14.39	122.88	130.96

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	66	CRO	CA2-N2-C1	9.39	112.69	105.77
1	A	66	CRO	CA2-C2-N3	9.30	107.77	103.37
1	D	66	CRO	CA2-C2-N3	9.25	107.75	103.37

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	C	66	CRO	N1-CA1-CB1-CG1
1	C	66	CRO	N1-CA1-CB1-OG1
1	C	66	CRO	C1-CA1-CB1-CG1
1	C	66	CRO	C1-CA1-CB1-OG1
1	C	66	CRO	N2-CA2-CB2-CG2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	66	CRO	1	0
1	B	66	CRO	1	0

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides 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	224/236 (94%)	-1.68	0 100 100	16, 25, 65, 84	0
1	B	224/236 (94%)	-1.73	0 100 100	14, 23, 57, 76	0
1	C	224/236 (94%)	-1.72	0 100 100	16, 24, 62, 78	0
1	D	224/236 (94%)	-1.69	0 100 100	15, 27, 64, 80	0
All	All	896/944 (94%)	-1.70	0 100 100	14, 24, 64, 84	0

There are no RSRZ outliers to report.

### 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<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	B-factors(Å <sup>2</sup> )	Q<0.9
1	CRO	A	66	22/23	1.00	0.02	17,22,24,25	0
1	CRO	B	66	22/23	1.00	0.02	17,22,24,25	0
1	CRO	C	66	22/23	1.00	0.02	17,22,24,25	0
1	CRO	D	66	22/23	1.00	0.02	17,22,24,25	0

### 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.