

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

#### Aug 19, 2023 – 09:24 PM EDT

PDB ID : 2HFC

Title : Structure of S65T Y66F R96A GFP variant in precursor state Authors : Barondeau, D.P.; Kassmann, C.J.; Tainer, J.A.; Getzoff, E.D.

Deposited on : 2006-06-23

Resolution : 1.20 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
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:

 $\begin{array}{ccc} & Mol Probity & : & 4.02b\text{-}467 \\ & Xtriage \text{ (Phenix)} & : & 1.13 \end{array}$ 

EDS: 2.35

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

Refmac : 5.8.0158

CCP4 : 7.0.044 (Gargrove)
roteins) : Engh & Huber (2001

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

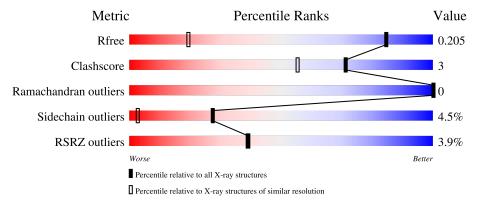
Validation Pipeline (wwPDB-VP) : 2.35

## 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 1.20 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\#  ext{Entries},  ext{ resolution range}( ext{Å}))$
$R_{free}$	130704	1223 (1.22-1.18)
Clashscore	141614	1286 (1.22-1.18)
Ramachandran outliers	138981	1240 (1.22-1.18)
Sidechain outliers	138945	1239 (1.22-1.18)
RSRZ outliers	127900	1200 (1.22-1.18)

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			
			4%			
1	A	239	85%	9%	٠	-



# 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2234 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.

$\mathbf{Mol}$	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace			
1	A	232	Total 1876	C 1195	N 310	O 366	S 5	0	12	0	

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual Comment		Reference
A	0	MET	-	cloning artifact	UNP P42212
A	1	ALA	-	cloning artifact	UNP P42212
A	64	LEU	PHE	engineered mutation	UNP P42212
A	65	THR	SER	engineered mutation	UNP P42212
A	66	PHE	TYR	engineered mutation	UNP P42212
A	96	ALA	ARG	engineered mutation	UNP P42212
A	99	SER	PHE	engineered mutation	UNP P42212
A	153	THR	MET	engineered mutation	UNP P42212
A	163	ALA	VAL	engineered mutation	UNP P42212

• Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mo	Chain	Residues	Atoms		Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Mg 1	0	0		

• Molecule 3 is water.

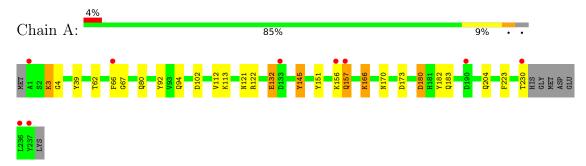
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	357	Total O 357 357	0	0



# 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	51.41Å 62.79Å 70.50Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 - 1.20	Depositor
rtesolution (A)	18.81 - 1.20	EDS
% Data completeness	(Not available) (20.00-1.20)	Depositor
(in resolution range)	79.6 (18.81-1.20)	EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.03	Depositor
$< I/\sigma(I) > 1$	2.48 (at 1.20Å)	Xtriage
Refinement program	SHELXL-97	Depositor
D D.	0.154 , 0.203	Depositor
$R, R_{free}$	0.180 , 0.205	DCC
$R_{free}$ test set	3079 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	11.4	Xtriage
Anisotropy	0.779	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.32, 62.0	EDS
L-test for twinning <sup>2</sup>	$ < L > = 0.49, < L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2234	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 7.85% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<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: MG

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).

Mal	Chain		lengths		ond angles
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5
1	A	0.62	0/1968	1.29	17/2661 (0.6%)

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
1	A	92	TYR	CD1-CE1-CZ	9.03	127.92	119.80
1	A	92	TYR	CB-CG-CD2	8.22	125.93	121.00
1	A	102	ASP	CB-CG-OD2	8.05	125.55	118.30
1	A	92	TYR	CG-CD1-CE1	-7.03	115.67	121.30
1	A	92	TYR	CG-CD2-CE2	6.94	126.86	121.30

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	1876	0	1824	13	0
2	A	1	0	0	0	0
3	A	357	0	0	7	0
All	All	2234	0	1824	13	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 3.

The worst 5 of 13 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	Clash overlap (Å)
1:A:157:GLN:HG2	3:A:1346:HOH:O	1.90	0.72
1:A:3:LYS:HE2	3:A:1317:HOH:O	2.01	0.60
1:A:132:GLU:HG2	3:A:1300:HOH:O	2.03	0.59
1:A:3:LYS:HG3	1:A:4:GLY:N	2.22	0.53
1:A:204[B]:GLN:NE2	3:A:1310:HOH:O	2.31	0.46

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	240/239 (100%)	237 (99%)	3 (1%)	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.

M	ol	Chain	Analysed	Rotameric	Outliers	Percentiles	
		A	212/206 (103%)	201 (95%)	11 (5%)	23 2	

5 of 11 residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	A	157	GLN
1	A	166[A]	LYS
1	A	230	THR
1	A	166[B]	LYS
1	A	121[B]	ASN

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

Mol	Chain	Res	Type
1	A	25	HIS
1	A	94	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)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

#### 5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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 (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 $>$	$\# \mathrm{RSRZ}{>}2$		$OWAB(A^2)$	Q<0.9	
1	A	232/239 (97%)	0.20	9 (3%)	39	39	12, 16, 30, 50	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	237	TYR	5.1
1	A	1	ALA	4.8
1	A	236	LEU	4.7
1	A	230	THR	4.5
1	A	157	GLN	4.3

#### 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 monosaccharides 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
2	MG	A	1000	1/1	0.98	0.16	18,18,18,18	0



# 6.5 Other polymers (i)

There are no such residues in this entry.

