

Full wwPDB X-ray Structure Validation Report (i)

Nov 15, 2023 – 12:23 AM JST

PDB ID : 6IR7

Title: Green fluorescent protein variant GFPuv with the modification to 6-

hydroxynorleucine at the C-terminus

Authors : Nakatani, T.; Yasui, N.; Yamashita, A.

Deposited on : 2018-11-12

Resolution : 1.28 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:

 $Mol Probity \quad : \quad 4.02b\text{--}467$

Mogul: 1.8.5 (274361), CSD as541be (2020)

Xtriage (Phenix) : 1.13 EDS : 2.36

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

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove)

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

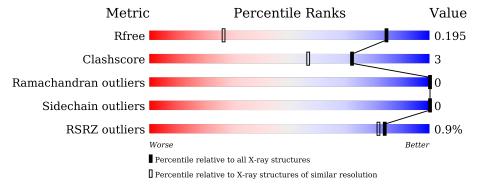
Validation Pipeline (wwPDB-VP) : 2.36

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 1.28 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries},\ {\rm resolution\ range}({\rm \AA})) \end{array}$
R_{free}	130704	1850 (1.30-1.26)
Clashscore	141614	1926 (1.30-1.26)
Ramachandran outliers	138981	1860 (1.30-1.26)
Sidechain outliers	138945	1859 (1.30-1.26)
RSRZ outliers	127900	1807 (1.30-1.26)

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	
1	A	235	92%	6% •



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2292 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.

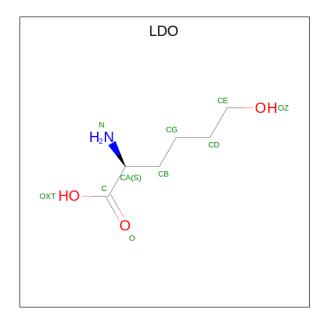
• Molecule 1 is a protein called Green fluorescent protein.

\mathbf{Mol}	Chain	Residues		Ato	oms			ZeroOcc	AltConf	Trace
1	A	232	Total 1897	C 1210	N 313	O 367	S 7	0	9	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	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	99	SER	PHE	engineered mutation	UNP P42212
A	153	THR	MET	engineered mutation	UNP P42212
A	163	ALA	VAL	engineered mutation	UNP P42212
A	206	LYS	ALA	engineered mutation	UNP P42212

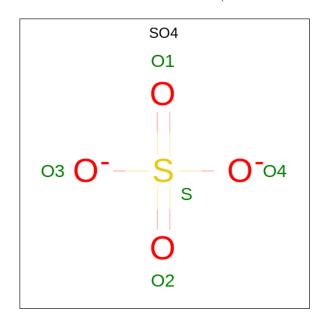
• Molecule 2 is 6-HYDROXY-L-NORLEUCINE (three-letter code: LDO) (formula: $C_6H_{13}NO_3$).





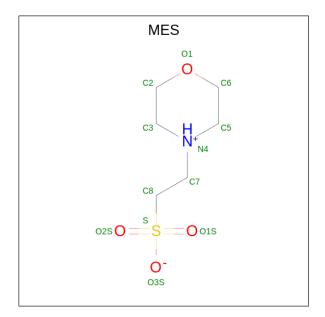
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C N O 10 6 1 3	0	0

 \bullet Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	
3	A	1	Total 5	O 4	S 1	0	0

• Molecule 4 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf		
4	A	1	Total	С	N	О	S	0	0
	11		12	6	1	4	1		9

• Molecule 5 is water.

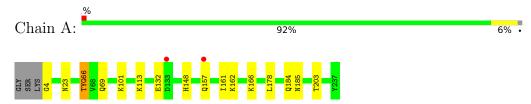
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	368	Total O 368 368	0	0



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 1 21 1	Depositor
Cell constants	47.30Å 51.15Å 48.51Å	Depositor
a, b, c, α , β , γ	90.00° 100.79° 90.00°	Depositor
Resolution (Å)	47.65 - 1.28	Depositor
resolution (A)	47.65 - 1.28	EDS
% Data completeness	96.3 (47.65-1.28)	Depositor
(in resolution range)	96.3 (47.65-1.28)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) > 1$	$1.50 \; (at \; 1.28 \text{Å})$	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.175 , 0.195	Depositor
it, it free	0.175 , 0.195	DCC
R_{free} test set	2889 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	11.6	Xtriage
Anisotropy	0.145	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.30, 39.7	EDS
L-test for twinning ²	$< L >=0.49, < L^2>=0.32$	Xtriage
Estimated twinning fraction	0.019 for l,-k,h	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2292	wwPDB-VP
Average B, all atoms (Å ²)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 9.40% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities 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: MES, CRO, LDO, SO4

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	Bond	lengths	Bond	angles
IVIOI	Chain	RMSZ	# Z > 5	RMSZ	# Z > 5
1	A	0.32	0/1948	0.53	0/2633

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	A	1897	0	1856	13	0
2	A	10	0	10	0	0
3	A	5	0	0	0	0
4	A	12	0	12	0	0
5	A	368	0	0	6	0
All	All	2292	0	1878	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.

All (13) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.



Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} (\mathring{\rm A}) \end{array}$	Clash overlap (Å)
1:A:113[B]:LYS:NZ	5:A:405:HOH:O	2.29	0.65
1:A:69:GLN:NE2	5:A:404:HOH:O	2.29	0.57
1:A:101:LYS:NZ	5:A:410:HOH:O	2.37	0.57
1:A:23:ASN:ND2	1:A:132:GLU:OE2	2.31	0.55
1:A:4:GLY:N	5:A:413:HOH:O	2.43	0.51
1:A:162:LYS:HE2	1:A:184:GLN:HG2	1.91	0.51
1:A:66:CRO:CZ	1:A:203[B]:THR:HG21	2.42	0.50
1:A:166:LYS:HD3	1:A:178:LEU:HD13	1.94	0.48
1:A:157:GLN:NE2	5:A:418:HOH:O	2.46	0.48
1:A:161[A]:ILE:HG13	1:A:185:ASN:HB2	1.95	0.48
1:A:66:CRO:CE2	1:A:203[B]:THR:HG21	2.45	0.46
1:A:113[A]:LYS:NZ	5:A:405:HOH:O	2.50	0.44
1:A:148:HIS:O	1:A:203[B]:THR:HG22	2.19	0.42

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	238/235 (101%)	235 (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.



Mol	Chain	Analysed	alysed Rotameric		Percentiles		
1	A	211/204 (103%)	211 (100%)	0	100	100	

There are no protein residues with a non-rotameric sidechain to report.

Sometimes 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)

1 non-standard protein/DNA/RNA residue is 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	Dog	Link	Bo	ond leng	$ ag{ths}$	В	ond ang	eles
MIOI	туре	Chain	rtes	Lilik	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
1	CRO	A	66	1	22,22,24	1.68	5 (22%)	27,30,34	1.87	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
ſ	1	CRO	A	66	1	-	0/9/29/32	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	Observed(A)	$\operatorname{Ideal}(ext{\AA})$
1	A	66	CRO	C1-N2	4.53	1.38	1.32
1	A	66	CRO	CG2-CB2	3.35	1.53	1.46
1	A	66	CRO	C1-N3	2.76	1.41	1.37
1	A	66	CRO	CA2-C2	2.68	1.51	1.48

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\text{\AA})$	$Ideal(\AA)$
1	A	66	CRO	CA3-N3	-2.05	1.43	1.47

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\mathbf{Observed}(^o)$	$\mathrm{Ideal}(^{o})$
1	A	66	CRO	CA2-C2-N3	4.83	105.66	103.37
1	A	66	CRO	O3-C3-CA3	-3.63	115.42	126.39
1	A	66	CRO	O2-C2-CA2	-3.40	129.05	130.96
1	A	66	CRO	CG2-CB2-CA2	-3.36	125.82	129.94
1	A	66	CRO	CA1-C1-N3	-2.26	121.90	124.85
1	A	66	CRO	CD2-CG2-CD1	2.20	120.90	117.64
1	A	66	CRO	CA2-N2-C1	2.18	107.38	105.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	66	CRO	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

3 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).

Mal	Type	Chain	Res Link		Bo	ond leng	$ ag{ths}$	В	ond ang	les
MIOI	туре	Chain	rtes	Lilik	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	MES	A	303	-	12,12,12	1.99	1 (8%)	14,16,16	1.76	3 (21%)



Mol	Iol Type Chain Res Link		Bo	ond leng	ths	Bond angles				
MIOI	Type	Chain	nes	Lilik	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	SO4	A	302	-	4,4,4	0.15	0	6,6,6	0.33	0
2	LDO	A	301	1	8,9,9	0.86	0	9,10,10	1.13	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
4	MES	A	303	-	-	4/6/14/14	0/1/1/1
2	LDO	A	301	1	-	2/9/9/9	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\text{\AA})$	$\operatorname{Ideal}(ext{\AA})$
4	A	303	MES	C8-S	-6.66	1.68	1.77

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$Ideal(^{o})$
4	A	303	MES	C5-N4-C3	4.46	118.88	108.83
4	A	303	MES	O1S-S-C8	2.91	110.42	106.92
4	A	303	MES	O2S-S-C8	2.68	110.14	106.92

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	303	MES	C7-C8-S-O1S
4	A	303	MES	C7-C8-S-O2S
4	A	303	MES	C7-C8-S-O3S
2	A	301	LDO	CE-CD-CG-CB
4	A	303	MES	C8-C7-N4-C5
2	A	301	LDO	CA-CB-CG-CD

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></rsrz>	$\#\mathrm{RSRZ}{>}2$		$\mathbf{OWAB}(\mathbf{\mathring{A}}^2)$	Q<0.9	
1	A	231/235 (98%)	0.18	2 (0%)	84	82	8, 13, 24, 35	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	133	ASP	3.3
1	A	157	GLN	3.1

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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
1	CRO	A	66	21/23	0.96	0.07	7,9,12,14	0

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	$\mathbf{B} ext{-}\mathbf{factors}(\mathring{\mathbf{A}}^2)$	Q<0.9
2	LDO	A	301	10/10	0.81	0.20	18,41,45,48	0
3	SO4	A	302	5/5	0.87	0.16	16,17,31,33	0
4	MES	A	303	12/12	0.88	0.14	14,20,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.

