wwPDB X-ray Structure Validation Summary Report (i)

Mar 23, 2024 – 01:03 PM EDT

PDB ID	:	4PPJ
Title	:	Crystal structure of Phanta, a weakly fluorescent photochromic GFP-like protein. ON state
Authors	:	Don Paul, C.; Traore, D.A.K.; Devenish, R.J.; Close, D.; Bell, T.; Bradbury, A.; Wilce, M.C.J.; Prescott, M.
Deposited on	:	2014-02-27
Resolution	:	2.30 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:

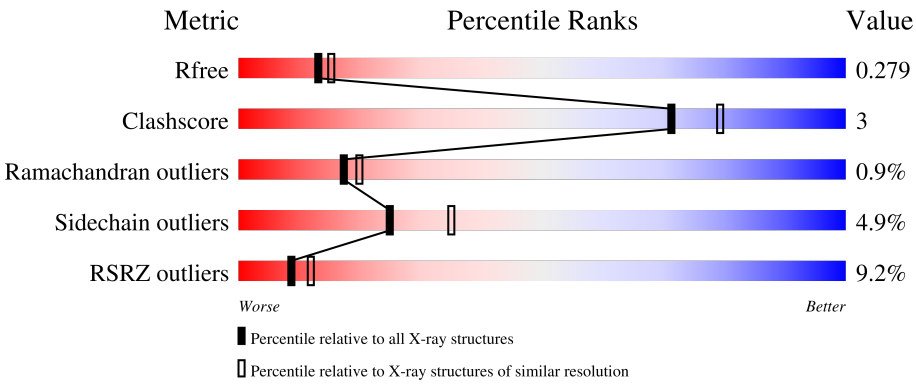
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.1

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 2.30 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}	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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 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	245	% 77% 9% • 12%
1	B	245	3% 77% 9% • 12%
1	C	245	13% 78% 9% • 12%
1	D	245	16% 76% 12% • 11%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7291 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 Monomeric Azami Green.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	216	1756	1130	289	325	12	0	0	0
1	B	216	1759	1128	291	328	12	0	0	0
1	C	216	1753	1125	288	328	12	0	0	0
1	D	217	1764	1134	293	325	12	0	0	0

- Molecule 2 is water.

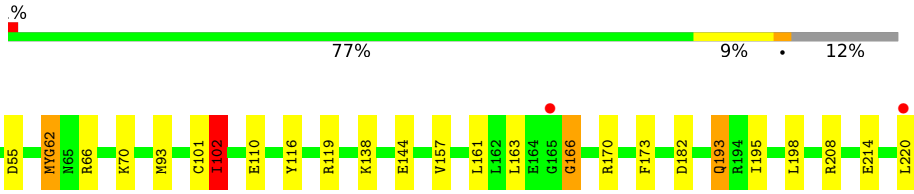
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	72	72	72	0	0
2	B	61	61	61	0	0
2	C	57	57	57	0	0
2	D	69	69	69	0	0

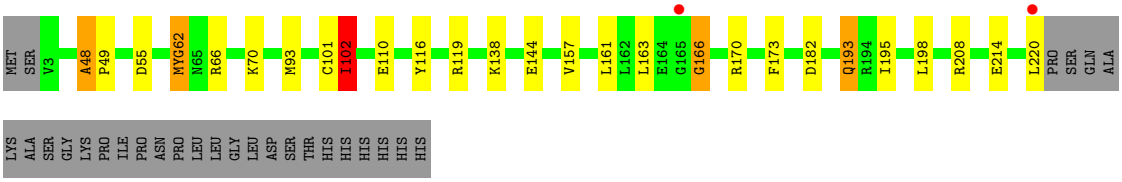
3 Residue-property plots (i)

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: Monomeric Azami Green

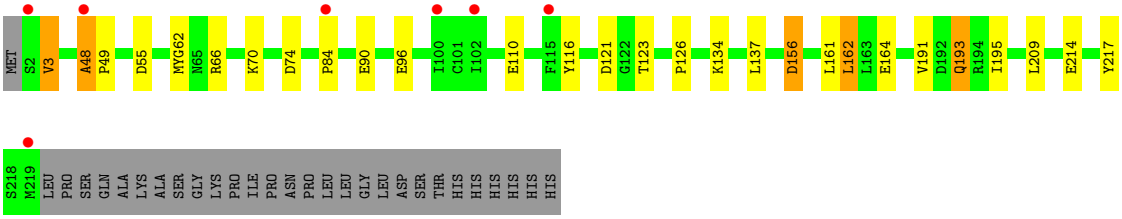
Chain A:





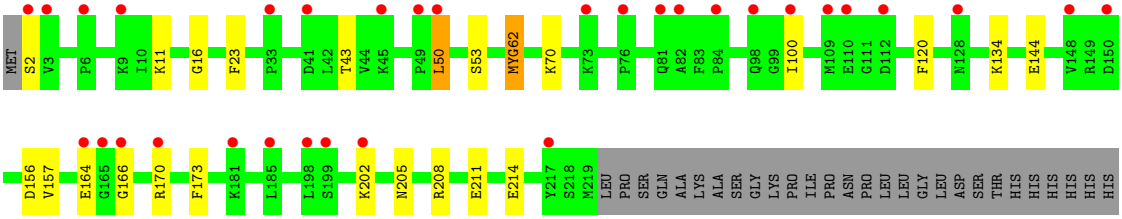
• Molecule 1: Monomeric Azami Green

Chain B:



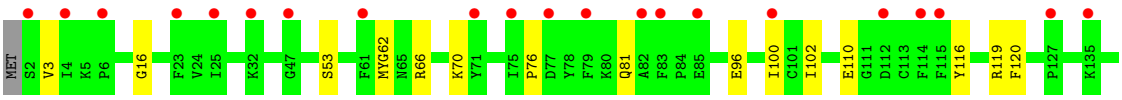
• Molecule 1: Monomeric Azami Green

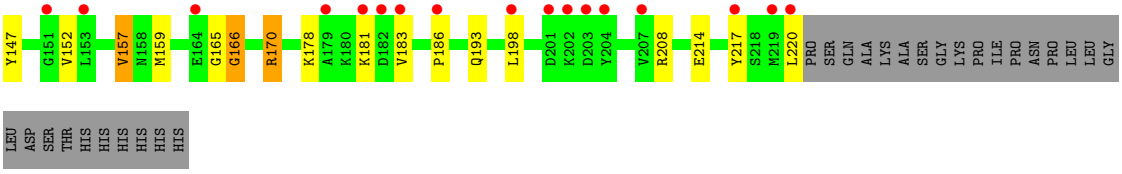
Chain C:



• Molecule 1: Monomeric Azami Green

Chain D:




Y147
G151
V152
L153
V157
N158
M159
E164
G165
G166
R170
K178
A179
K180
K181
D182
V183
P186
Q193
L198
D201
K202
D203
Y204
V207
R208
E214
Y217
S218
M219
L220
PRO
SER
GLN
ALA
LYS
ALA
SER
GLY
LYS
PRO
ILE
PRO
ASN
PRO
LEU
LEU
GLY
LEU
ASP
SER
THR
HIS
HIS
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	69.81Å 79.37Å 71.00Å 90.00° 102.40° 90.00°	Depositor
Resolution (Å)	30.68 – 2.30 34.67 – 2.30	Depositor EDS
% Data completeness (in resolution range)	98.2 (30.68-2.30) 98.5 (34.67-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.28 (at 2.31Å)	Xtriage
Refinement program	REFMAC 5.5.0088, BUSTER 2.10.0	Depositor
R, R_{free}	0.197 , 0.263 0.206 , 0.279	Depositor DCC
R_{free} test set	1669 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	49.3	Xtriage
Anisotropy	0.080	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 65.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.039 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7291	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]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.

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: NRQ

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	0.44	0/1777	0.73	1/2395 (0.0%)
1	B	0.43	0/1780	0.71	0/2399
1	C	0.39	0/1774	0.66	0/2392
1	D	0.42	0/1785	0.69	0/2405
All	All	0.42	0/7116	0.70	1/9591 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	102	ILE	N-CA-CB	5.01	122.33	110.80

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	1756	0	1712	15	0
1	B	1759	0	1709	14	0
1	C	1753	0	1698	11	0
1	D	1764	0	1724	14	0
2	A	72	0	0	0	0
2	B	61	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	57	0	0	2	0
2	D	69	0	0	0	0
All	All	7291	0	6843	48	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 48 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:170:ARG:HD3	1:D:147:TYR:CZ	2.17	0.80
1:A:48:ALA:HB1	1:A:49:PRO:HD3	1.63	0.79
1:A:70:LYS:HB3	1:A:214:GLU:HG2	1.72	0.69
1:C:16:GLY:HA3	1:C:120:PHE:O	2.00	0.61
1:D:70:LYS:HB3	1:D:214:GLU:HG2	1.83	0.60

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	211/245 (86%)	205 (97%)	4 (2%)	2 (1%)	17	20
1	B	211/245 (86%)	202 (96%)	7 (3%)	2 (1%)	17	20
1	C	211/245 (86%)	203 (96%)	6 (3%)	2 (1%)	17	20
1	D	212/245 (86%)	204 (96%)	6 (3%)	2 (1%)	17	20
All	All	845/980 (86%)	814 (96%)	23 (3%)	8 (1%)	17	20

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	48	ALA
1	B	48	ALA
1	A	166	GLY
1	C	164	GLU
1	D	3	VAL

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	185/212 (87%)	177 (96%)	8 (4%)	29	40
1	B	186/212 (88%)	177 (95%)	9 (5%)	25	36
1	C	185/212 (87%)	176 (95%)	9 (5%)	25	35
1	D	186/212 (88%)	176 (95%)	10 (5%)	22	30
All	All	742/848 (88%)	706 (95%)	36 (5%)	25	35

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	102	ILE
1	D	220	LEU
1	D	110	GLU
1	D	170	ARG
1	B	134	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	124	ASN
1	C	205	ASN
1	D	212	HIS
1	D	193	GLN
1	B	193	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)

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	NRQ	A	62	1	23,24,25	3.43	5 (21%)	23,32,34	4.72	8 (34%)
1	NRQ	D	62	1	23,24,25	3.66	4 (17%)	23,32,34	4.68	7 (30%)
1	NRQ	B	62	1	23,24,25	3.55	5 (21%)	23,32,34	4.79	8 (34%)
1	NRQ	C	62	1	23,24,25	3.41	6 (26%)	23,32,34	4.93	7 (30%)

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	NRQ	A	62	1	-	2/9/31/32	0/2/2/2
1	NRQ	D	62	1	-	3/9/31/32	0/2/2/2
1	NRQ	B	62	1	-	4/9/31/32	0/2/2/2
1	NRQ	C	62	1	-	3/9/31/32	0/2/2/2

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	62	NRQ	CB2-CA2	16.62	1.49	1.35
1	B	62	NRQ	CB2-CA2	15.78	1.48	1.35
1	C	62	NRQ	CB2-CA2	15.17	1.47	1.35
1	A	62	NRQ	CB2-CA2	15.15	1.47	1.35
1	B	62	NRQ	CA2-C2	-3.42	1.45	1.48

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	62	NRQ	O2-C2-CA2	-19.39	120.07	130.96
1	B	62	NRQ	O2-C2-CA2	-18.82	120.39	130.96
1	A	62	NRQ	O2-C2-CA2	-18.70	120.46	130.96
1	D	62	NRQ	O2-C2-CA2	-17.73	121.00	130.96
1	D	62	NRQ	CA2-C2-N3	9.61	107.92	103.37

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	62	NRQ	N2-CA2-CB2-CG2
1	A	62	NRQ	C3-CA3-N3-C2
1	B	62	NRQ	C1-CA1-CB1-CG1
1	B	62	NRQ	N2-CA2-CB2-CG2
1	B	62	NRQ	C3-CA3-N3-C2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	62	NRQ	1	0
1	C	62	NRQ	2	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, 95th 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	A	215/245 (87%)	0.19	2 (0%)	84	88	34, 53, 84, 126	0
1	B	215/245 (87%)	0.30	7 (3%)	46	53	35, 63, 102, 151	0
1	C	215/245 (87%)	0.92	32 (14%)	2	3	43, 75, 114, 151	0
1	D	216/245 (88%)	0.94	38 (17%)	1	1	42, 68, 105, 143	0
All	All	861/980 (87%)	0.59	79 (9%)	9	12	34, 64, 104, 151	0

The worst 5 of 79 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	2	SER	15.8
1	D	112	ASP	7.1
1	C	3	VAL	7.1
1	D	220	LEU	6.4
1	C	81	GLN	5.7

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, 95th 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^2)	Q<0.9
1	NRQ	C	62	23/24	0.91	0.14	45,50,63,69	0
1	NRQ	D	62	23/24	0.91	0.16	55,57,63,67	0
1	NRQ	A	62	23/24	0.94	0.15	32,38,44,47	0
1	NRQ	B	62	23/24	0.96	0.13	40,44,52,56	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.