Full wwPDB X-ray Structure Validation Report (i)

May 13, 2020 – 09:06 am BST

PDB ID	:	4XL5
Title	:	X-ray structure of bGFP-A / EGFP complex
Authors	:	Chevrel, A.; Urvoas, A.; Li de la Sierra-Gallay, I.; Van Tilbeurgh, H.; Minard, P.; Valerio-Lepiniec, M.
Deposited on	:	2015-01-13
Resolution	:	2.00 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 following versions of software and data (see references (1)) were used in the production of this report:

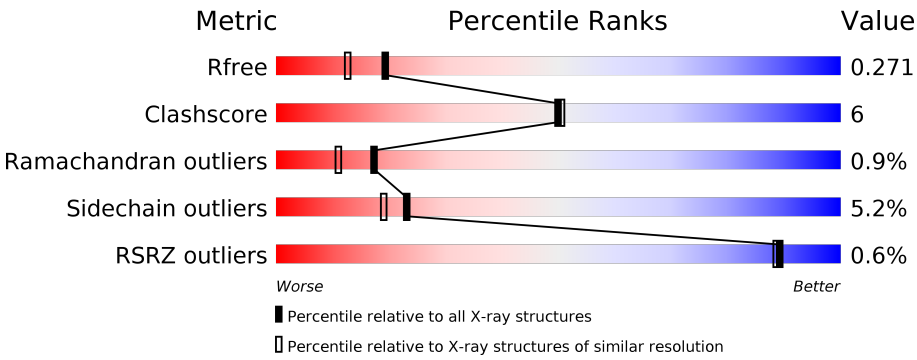
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 on 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	272	% 71% 11% • 17%
2	C	263	77% 14% • 7%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3790 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
			Total	C	N	O	S			
1	A	227	1818	1158	306	348	6	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	initiating methionine	UNP P42212
A	-19	ARG	-	expression tag	UNP P42212
A	-18	GLY	-	expression tag	UNP P42212
A	-17	SER	-	expression tag	UNP P42212
A	-16	HIS	-	expression tag	UNP P42212
A	-15	HIS	-	expression tag	UNP P42212
A	-14	HIS	-	expression tag	UNP P42212
A	-13	HIS	-	expression tag	UNP P42212
A	-12	HIS	-	expression tag	UNP P42212
A	-11	HIS	-	expression tag	UNP P42212
A	-10	THR	-	expression tag	UNP P42212
A	-9	ASP	-	expression tag	UNP P42212
A	-8	PRO	-	expression tag	UNP P42212
A	-7	HIS	-	expression tag	UNP P42212
A	-6	ALA	-	expression tag	UNP P42212
A	-5	SER	-	expression tag	UNP P42212
A	-4	SER	-	expression tag	UNP P42212
A	-3	ALA	-	expression tag	UNP P42212
A	-2	GLY	-	expression tag	UNP P42212
A	-1	SER	-	expression tag	UNP P42212
A	0	PRO	-	expression tag	UNP P42212
A	1	VAL	-	expression tag	UNP P42212
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	231	LEU	HIS	engineered mutation	UNP P42212

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	239	GLY	-	expression tag	UNP P42212
A	240	SER	-	expression tag	UNP P42212
A	241	GLY	-	expression tag	UNP P42212
A	242	ALA	-	expression tag	UNP P42212
A	243	SER	-	expression tag	UNP P42212
A	244	GLU	-	expression tag	UNP P42212
A	245	GLN	-	expression tag	UNP P42212
A	246	LYS	-	expression tag	UNP P42212
A	247	LEU	-	expression tag	UNP P42212
A	248	ILE	-	expression tag	UNP P42212
A	249	SER	-	expression tag	UNP P42212
A	250	GLU	-	expression tag	UNP P42212
A	251	GLU	-	expression tag	UNP P42212
A	252	ASP	-	expression tag	UNP P42212
A	253	LEU	-	expression tag	UNP P42212

- Molecule 2 is a protein called bGFP-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	245	1869	1173	334	360	2	0	0	0

- Molecule 3 is water.

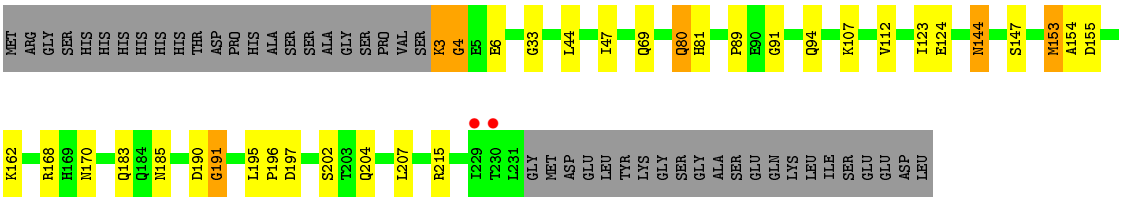
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	57	57	57	0	0
			Total	O		
3	C	46	46	46	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

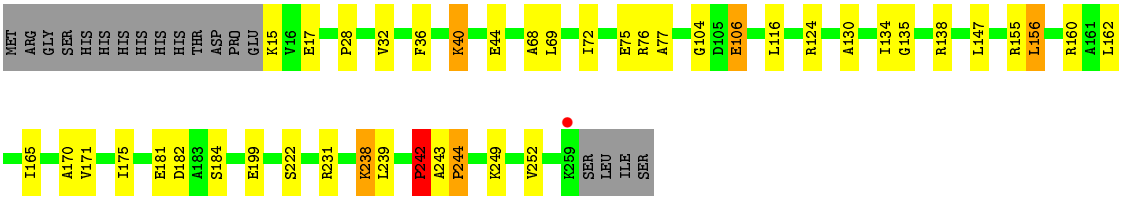
- Molecule 1: Green fluorescent protein

Chain A:



- Molecule 2: bGFP-A

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	72.28Å 45.98Å 74.89Å 90.00° 90.76° 90.00°	Depositor
Resolution (Å)	39.18 – 2.00 39.18 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.5 (39.18-2.00) 98.5 (39.18-2.00)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.76 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, R_{free}	0.205 , 0.268 0.213 , 0.271	Depositor DCC
R_{free} test set	1657 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.711	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 45.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.003 for l,k,-h 0.022 for h,-k,-l 0.017 for l,-k,h	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3790	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.38% 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: 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	0.72	0/1836	0.83	0/2481
2	C	0.69	0/1893	0.83	2/2557 (0.1%)
All	All	0.70	0/3729	0.83	2/5038 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	156	LEU	CA-CB-CG	7.17	131.78	115.30
2	C	160	ARG	NE-CZ-NH1	5.05	122.82	120.30

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	1818	0	1770	20	0
2	C	1869	0	1916	23	0
3	A	57	0	0	2	0
3	C	46	0	0	1	0
All	All	3790	0	3686	43	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 6.

All (43) 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)
2:C:238:LYS:O	2:C:242:PRO:CD	2.05	1.04
2:C:238:LYS:O	2:C:242:PRO:HD3	1.64	0.97
2:C:238:LYS:O	2:C:242:PRO:HD2	1.69	0.93
1:A:183:GLN:HE21	1:A:185:ASN:HD21	1.19	0.90
2:C:239:LEU:HD12	2:C:243:ALA:HB2	1.70	0.73
2:C:239:LEU:CD1	2:C:243:ALA:HB2	2.25	0.66
1:A:3:LYS:HD3	1:A:89:PRO:HD3	1.84	0.60
1:A:47:ILE:HG21	1:A:215:ARG:HE	1.66	0.59
1:A:3:LYS:HG3	1:A:4:GLY:N	2.18	0.58
1:A:81:HIS:HD2	1:A:197:ASP:H	1.52	0.58
2:C:182:ASP:OD1	2:C:184:SER:HB2	2.04	0.57
1:A:147:SER:O	1:A:168:ARG:NH2	2.40	0.55
1:A:3:LYS:O	1:A:6:GLU:HB2	2.06	0.55
1:A:80:GLN:HE21	1:A:80:GLN:H	1.54	0.55
1:A:94:GLN:NE2	3:A:313:HOH:O	2.31	0.54
2:C:130:ALA:O	2:C:134:ILE:HG12	2.07	0.54
2:C:44:GLU:OE2	2:C:76:ARG:CZ	2.55	0.54
1:A:183:GLN:HE21	1:A:185:ASN:ND2	1.97	0.53
1:A:33:GLY:HA3	1:A:44:LEU:HD23	1.91	0.53
2:C:69:LEU:HD22	2:C:77:ALA:HB2	1.89	0.53
2:C:68:ALA:O	2:C:72:ILE:HG12	2.09	0.53
2:C:106:GLU:HG2	2:C:138:ARG:NH2	2.24	0.52
2:C:171:VAL:O	2:C:175:ILE:HG12	2.10	0.51
1:A:144:ASN:C	1:A:144:ASN:HD22	2.13	0.51
1:A:47:ILE:HD13	1:A:215:ARG:HH21	1.75	0.51
2:C:239:LEU:O	2:C:243:ALA:HB3	2.14	0.47
2:C:28:PRO:O	2:C:32:VAL:HG23	2.15	0.47
1:A:155:ASP:CG	1:A:162:LYS:HE3	2.36	0.46
2:C:116:LEU:O	2:C:124:ARG:HD2	2.15	0.46
1:A:144:ASN:HD21	1:A:170:ASN:HB2	1.80	0.45
2:C:104:GLY:HA2	2:C:134:ILE:HD13	1.97	0.45
1:A:91:GLY:HA3	1:A:112:VAL:O	2.17	0.44
2:C:147:LEU:O	2:C:155:ARG:HD2	2.17	0.43
2:C:199:GLU:HB3	2:C:231:ARG:HD3	2.00	0.43
1:A:153:MET:HG2	1:A:154:ALA:O	2.19	0.42
2:C:181:GLU:HG2	3:C:333:HOH:O	2.20	0.42
1:A:190:ASP:O	1:A:191:GLY:C	2.57	0.41
2:C:162:LEU:HD22	2:C:170:ALA:HB2	2.02	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:123:ILE:HG22	1:A:124:GLU:N	2.35	0.41
2:C:36:PHE:O	2:C:40:LYS:HG2	2.20	0.41
2:C:135:GLY:HA2	2:C:165:ILE:CD1	2.50	0.41
1:A:69:GLN:HG2	3:A:339:HOH:O	2.20	0.41
2:C:249:LYS:HA	2:C:252:VAL:HG22	2.02	0.40

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	222/272 (82%)	215 (97%)	5 (2%)	2 (1%)	17	11
2	C	243/263 (92%)	237 (98%)	4 (2%)	2 (1%)	19	13
All	All	465/535 (87%)	452 (97%)	9 (2%)	4 (1%)	17	11

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	244	PRO
1	A	191	GLY
2	C	242	PRO
1	A	4	GLY

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	198/235 (84%)	188 (95%)	10 (5%)	24	19
2	C	184/201 (92%)	174 (95%)	10 (5%)	22	18
All	All	382/436 (88%)	362 (95%)	20 (5%)	23	19

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3	LYS
1	A	80	GLN
1	A	107	LYS
1	A	144	ASN
1	A	153	MET
1	A	195	LEU
1	A	196	PRO
1	A	202	SER
1	A	204	GLN
1	A	207	LEU
2	C	15	LYS
2	C	17	GLU
2	C	40	LYS
2	C	75	GLU
2	C	106	GLU
2	C	156	LEU
2	C	222	SER
2	C	238	LYS
2	C	242	PRO
2	C	244	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (11) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	80	GLN
1	A	81	HIS
1	A	94	GLN
1	A	144	ASN
1	A	146	ASN
1	A	159	ASN
1	A	164	ASN
1	A	170	ASN
1	A	177	GLN
1	A	185	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	C	258	HIS

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	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
1	CRO	A	66	1	23,23,24	3.88	6 (26%)	30,32,34	4.13	7 (23%)

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	A	66	1	-	0/12/31/32	0/2/2/2

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	CRO	CB2-CA2	17.20	1.49	1.35
1	A	66	CRO	C2-N3	-3.44	1.31	1.39
1	A	66	CRO	C1-N2	3.14	1.36	1.32
1	A	66	CRO	CA1-C1	-2.32	1.48	1.51
1	A	66	CRO	O2-C2	2.30	1.28	1.23
1	A	66	CRO	CA2-N2	-2.22	1.33	1.38

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	CRO	O2-C2-CA2	-14.43	122.86	130.96
1	A	66	CRO	CA2-C2-N3	13.56	109.78	103.37
1	A	66	CRO	C2-N3-C1	-5.50	105.18	107.97
1	A	66	CRO	O3-C3-CA3	-4.53	112.71	126.39
1	A	66	CRO	C2-CA2-N2	-4.52	105.77	108.93
1	A	66	CRO	CB2-CA2-C2	4.08	127.14	122.28
1	A	66	CRO	C1-CA1-N1	-2.05	106.63	109.96

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates 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	226/272 (83%)	-0.06	2 (0%)	84	83	33, 46, 70, 81	0
2	C	245/263 (93%)	-0.04	1 (0%)	92	92	35, 46, 67, 89	0
All	All	471/535 (88%)	-0.05	3 (0%)	89	88	33, 46, 68, 89	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	230	THR	3.7
1	A	229	ILE	3.6
2	C	259	LYS	2.3

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	CRO	A	66	22/23	0.97	0.10	31,34,37,43	0

6.3 Carbohydrates (i)

There are no carbohydrates 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.