wwPDB X-ray Structure Validation Summary Report (i)

Nov 6, 2023 – 07:36 PM EST

PDB ID	:	1EMC
Title	:	GREEN FLUORESCENT PROTEIN FROM AEQUOREA VICTORIA, MUTANT
Authors	:	Palm, G.; Zdanov, A.; Wlodawer, A.
Deposited on	:	1997-03-31
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 (i)) were used in the production of this report:

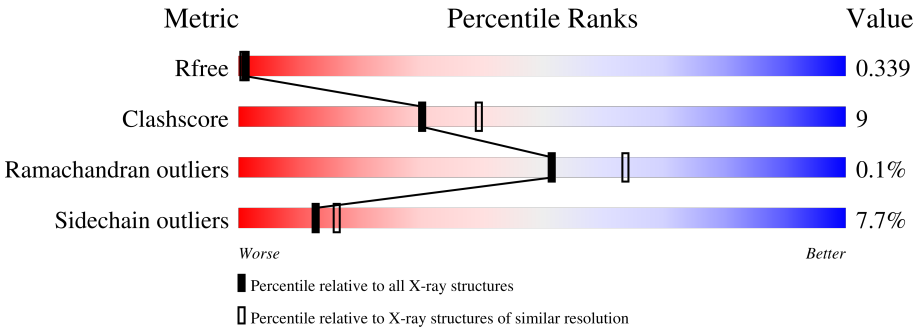
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 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)

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%

Mol	Chain	Length	Quality of chain
1	A	237	70% 22% • 5%
1	B	237	71% 22% • 5%
1	C	237	64% 29% • 5%
1	D	237	69% 25% • 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7354 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	1	0
			1795	1142	302	345	6			
1	B	225	Total	C	N	O	S	0	1	0
			1795	1142	302	345	6			
1	C	226	Total	C	N	O	S	0	1	0
			1811	1151	308	346	6			
1	D	225	Total	C	N	O	S	0	1	0
			1795	1142	302	345	6			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	64	LEU	PHE	engineered mutation	UNP P42212
A	66	GYS	SER	chromophore	UNP P42212
A	66	GYS	TYR	chromophore	UNP P42212
A	66	GYS	GLY	chromophore	UNP P42212
A	80	ARG	GLN	conflict	UNP P42212
A	167	THR	ILE	engineered mutation	UNP P42212
B	64	LEU	PHE	engineered mutation	UNP P42212
B	66	GYS	SER	chromophore	UNP P42212
B	66	GYS	TYR	chromophore	UNP P42212
B	66	GYS	GLY	chromophore	UNP P42212
B	80	ARG	GLN	conflict	UNP P42212
B	167	THR	ILE	engineered mutation	UNP P42212
C	64	LEU	PHE	engineered mutation	UNP P42212
C	66	GYS	SER	chromophore	UNP P42212
C	66	GYS	TYR	chromophore	UNP P42212
C	66	GYS	GLY	chromophore	UNP P42212
C	80	ARG	GLN	conflict	UNP P42212
C	167	THR	ILE	engineered mutation	UNP P42212
D	64	LEU	PHE	engineered mutation	UNP P42212
D	66	GYS	SER	chromophore	UNP P42212
D	66	GYS	TYR	chromophore	UNP P42212

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	66	GYS	GLY	chromophore	UNP P42212
D	80	ARG	GLN	conflict	UNP P42212
D	167	THR	ILE	engineered mutation	UNP P42212

- Molecule 2 is water.

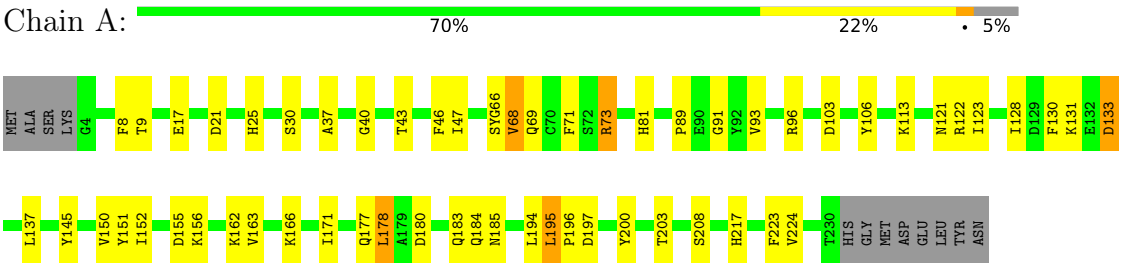
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	40	Total O 40 40	0	0
2	B	42	Total O 42 42	0	0
2	C	38	Total O 38 38	0	0
2	D	38	Total O 38 38	0	0

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. 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. 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.

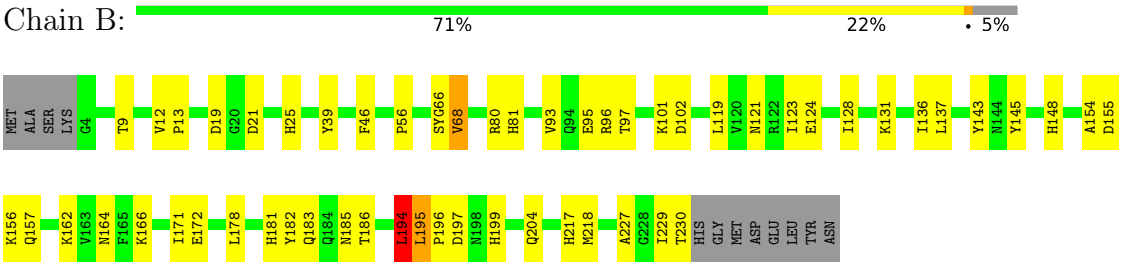
• Molecule 1: GREEN FLUORESCENT PROTEIN

Chain A:



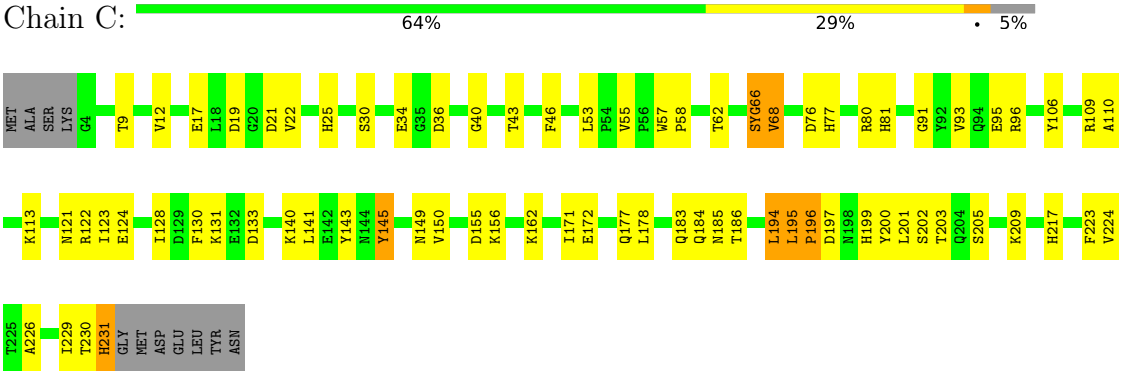
• Molecule 1: GREEN FLUORESCENT PROTEIN

Chain B:



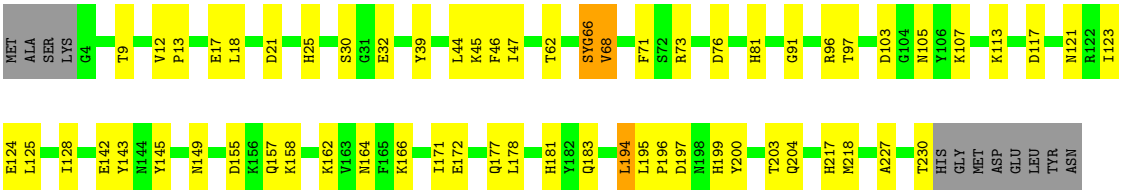
• Molecule 1: GREEN FLUORESCENT PROTEIN

Chain C:



• Molecule 1: GREEN FLUORESCENT PROTEIN

Chain D:


MET ALA SER LYS G4 T9 V12 P13 E17 L18 D21 H25 S30 G31 E32 Y39 L44 K45 F46 I47 T62 SYG66 V68 F71 S72 R73 D76 H81 G91 R96 T97 D103 G104 N105 Y106 K107 K113 D117 N121 R122 I123
E124 L125 I128 E142 Y143 N144 Y145 N149 D155 K156 Q157 K158 K162 V163 N164 F165 K166 I171 E172 Q177 L178 H181 Y182 Q183 L194 L195 P196 D197 N198 H199 Y200 T203 Q204 H217 M218 A227 T230 HIS GLY MET ASP GLU LEU TYR ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	93.00Å 52.00Å 103.00Å 90.00° 101.80° 90.00°	Depositor
Resolution (Å)	10.00 – 2.30 10.05 – 2.29	Depositor EDS
% Data completeness (in resolution range)	72.5 (10.00-2.30) 71.5 (10.05-2.29)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	2.12 (at 2.28Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.205 , 0.288 0.298 , 0.339	Depositor DCC
R_{free} test set	2211 reflections (7.13%)	wwPDB-VP
Wilson B-factor (Å^2)	15.8	Xtriage
Anisotropy	0.488	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 83.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.82	EDS
Total number of atoms	7354	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 50.40 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.5076e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GYS

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	0.86	0/1817	0.95	1/2458 (0.0%)
1	B	0.85	0/1817	0.92	2/2458 (0.1%)
1	C	0.81	0/1834	0.97	2/2480 (0.1%)
1	D	0.82	0/1817	0.94	3/2458 (0.1%)
All	All	0.83	0/7285	0.95	8/9854 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	2
1	C	0	1
1	D	0	1
All	All	0	5

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	231	HIS	N-CA-C	6.51	128.57	111.00
1	C	194	LEU	CA-CB-CG	6.45	130.14	115.30
1	B	97	THR	N-CA-C	-6.44	93.61	111.00
1	D	44	LEU	CA-CB-CG	5.85	128.75	115.30
1	A	73	ARG	NE-CZ-NH2	-5.35	117.63	120.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	151	TYR	Sidechain
1	B	182	TYR	Sidechain
1	B	39	TYR	Sidechain
1	C	145	TYR	Sidechain
1	D	39	TYR	Sidechain

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	1795	0	1733	29	0
1	B	1795	0	1734	32	1
1	C	1811	0	1751	40	1
1	D	1795	0	1734	34	2
2	A	40	0	0	3	0
2	B	42	0	0	4	0
2	C	38	0	0	0	0
2	D	38	0	0	2	1
All	All	7354	0	6952	132	3

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 9.

The worst 5 of 132 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:203[B]:THR:HG22	1:A:224:VAL:HG13	1.44	0.99
1:A:81:HIS:HD2	1:A:197:ASP:H	1.21	0.89
1:A:81:HIS:CD2	1:A:197:ASP:H	1.91	0.89
1:C:81:HIS:CD2	1:C:197:ASP:H	2.04	0.75
1:C:203[B]:THR:HG22	1:C:224:VAL:HG13	1.67	0.74

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:32:GLU:OE2	1:D:76:ASP:N[2_646]	1.91	0.29
1:D:117:ASP:OD1	2:D:260:HOH:O[2_646]	2.06	0.14
1:B:101:LYS:NZ	1:C:131:LYS:NZ[1_564]	2.10	0.10

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	221/237 (93%)	214 (97%)	7 (3%)	0	100 100
1	B	221/237 (93%)	211 (96%)	9 (4%)	1 (0%)	29 35
1	C	222/237 (94%)	215 (97%)	7 (3%)	0	100 100
1	D	221/237 (93%)	215 (97%)	6 (3%)	0	100 100
All	All	885/948 (93%)	855 (97%)	29 (3%)	1 (0%)	51 64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	136	ILE

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	195/206 (95%)	178 (91%)	17 (9%)	10 12
1	B	195/206 (95%)	182 (93%)	13 (7%)	16 21

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	197/206 (96%)	179 (91%)	18 (9%)	9 11
1	D	195/206 (95%)	183 (94%)	12 (6%)	18 25
All	All	782/824 (95%)	722 (92%)	60 (8%)	13 16

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	196	PRO
1	D	145	TYR
1	C	68	VAL
1	D	128	ILE
1	D	230	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	81	HIS
1	D	157	GLN
1	C	149	ASN
1	D	204	GLN
1	D	121	ASN

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	GYS	A	66	1	22,22,23	1.20	3 (13%)	27,30,32	1.25	4 (14%)
1	GYS	C	66	1	22,22,23	1.40	3 (13%)	27,30,32	1.34	5 (18%)
1	GYS	B	66	1	22,22,23	1.22	2 (9%)	27,30,32	1.81	6 (22%)
1	GYS	D	66	1	22,22,23	1.25	2 (9%)	27,30,32	1.49	6 (22%)

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	GYS	A	66	1	-	0/9/29/30	0/2/2/2
1	GYS	C	66	1	-	0/9/29/30	0/2/2/2
1	GYS	B	66	1	-	0/9/29/30	0/2/2/2
1	GYS	D	66	1	-	0/9/29/30	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	66	GYS	CE1-CD1	3.71	1.45	1.38
1	B	66	GYS	C2-N3	-2.92	1.33	1.39
1	C	66	GYS	CE2-CD2	2.89	1.44	1.38
1	B	66	GYS	CB2-CA2	-2.87	1.32	1.35
1	A	66	GYS	CB2-CA2	-2.86	1.32	1.35

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	66	GYS	O2-C2-CA2	4.65	133.57	130.96
1	C	66	GYS	N3-C1-N2	3.95	114.19	111.45
1	D	66	GYS	O2-C2-CA2	3.83	133.11	130.96
1	B	66	GYS	N3-C1-N2	3.59	113.94	111.45
1	A	66	GYS	N3-C1-N2	3.58	113.93	111.45

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	66	GYS	1	0
1	D	66	GYS	1	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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.