Full wwPDB X-ray Structure Validation Report (i)

Oct 16, 2021 – 11:10 PM EDT

PDB ID	:	1OXD
Title	:	Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins
Authors	:	Hyun Bae, J.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M.H.; Azim, M.K.; Kim, J.S.; Zumbusch, A.; Holak, T.A.; Moroder, L.; Huber, R.; Budisa, N.
Deposited on	:	2003-04-02
Resolution	:	1.15 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 (i)) were used in the production of this report:

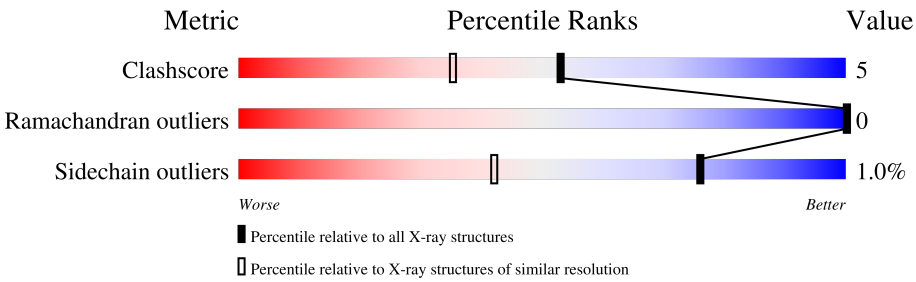
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 1.15 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))
Clashscore	141614	1537 (1.18-1.10)
Ramachandran outliers	138981	1483 (1.18-1.10)
Sidechain outliers	138945	1480 (1.18-1.10)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	227	90% 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1966 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 cyan fluorescent protein cfp.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	227	1800	1147	305	342	6	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	66	CRF	THR	chromophore	GB 11321072
A	66	CRF	TRP	chromophore	GB 11321072
A	66	CRF	GLY	chromophore	GB 11321072
A	80	ARG	GLN	engineered mutation	GB 11321072

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	166	166	166	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.

Note EDS was not executed.

- Molecule 1: cyan fluorescent protein cfp

Chain A:

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	50.95Å 62.77Å 69.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	7.98 – 1.15	Depositor
% Data completeness (in resolution range)	86.6 (7.98-1.15)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.214 , 0.227	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1966	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

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: CRF

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	0.36	1/1816 (0.1%)	0.70	0/2455

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	64	LEU	C-O	5.22	1.33	1.23

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	1800	0	1740	19	0
2	A	166	0	0	0	0
All	All	1966	0	1740	19	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 5.

All (19) 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:45:LYS:HE2	1:A:47:ILE:HD11	1.66	0.76
1:A:33:GLY:HA3	1:A:44:LEU:HD13	1.73	0.69
1:A:168:ARG:HB3	1:A:176:VAL:HG21	1.76	0.67
1:A:43:THR:C	1:A:44:LEU:HD22	2.18	0.63
1:A:203:THR:HG23	1:A:224:VAL:HG22	1.87	0.56
1:A:168:ARG:HH11	1:A:168:ARG:HG2	1.74	0.53
1:A:90:GLU:H	1:A:90:GLU:CD	2.13	0.51
1:A:42:LEU:HB3	1:A:44:LEU:HD21	1.92	0.51
1:A:43:THR:O	1:A:44:LEU:HD13	2.11	0.51
1:A:168:ARG:HB3	1:A:176:VAL:CG2	2.41	0.50
1:A:136:ILE:HD12	1:A:136:ILE:N	2.28	0.49
1:A:18:LEU:HD23	1:A:18:LEU:C	2.33	0.48
1:A:44:LEU:HD22	1:A:44:LEU:N	2.30	0.46
1:A:47:ILE:HD13	1:A:217:HIS:CB	2.48	0.44
1:A:168:ARG:NH1	1:A:178:LEU:HD13	2.33	0.43
1:A:42:LEU:HD21	1:A:71:PHE:CG	2.54	0.42
1:A:47:ILE:HD13	1:A:217:HIS:HB3	2.02	0.42
1:A:33:GLY:HA3	1:A:44:LEU:CD1	2.46	0.41
1:A:205:SER:HA	1:A:222:GLU:HG2	2.03	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	222/227 (98%)	217 (98%)	5 (2%)	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	Rotameric	Outliers	Percentiles
1	A	192/197 (98%)	190 (99%)	2 (1%)	76 42

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	145	TYR
1	A	196	PRO

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	149	ASN
1	A	170	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)

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	CRF	A	66	1	25,26,27	2.27	9 (36%)	32,37,39	2.83	10 (31%)

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	CRF	A	66	1	-	0/10/31/32	0/3/3/3

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	CRF	CB2-CA2	7.64	1.41	1.35
1	A	66	CRF	C1-N2	-3.50	1.27	1.32
1	A	66	CRF	CZ3-CE3	3.30	1.44	1.36
1	A	66	CRF	CA2-N2	-2.93	1.32	1.38
1	A	66	CRF	CA2-C2	-2.56	1.46	1.48
1	A	66	CRF	O3-C3	2.42	1.33	1.19
1	A	66	CRF	CG2-CB2	-2.30	1.42	1.46
1	A	66	CRF	O2-C2	2.19	1.27	1.23
1	A	66	CRF	CH2-CZ2	2.07	1.41	1.36

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	CRF	CA2-N2-C1	9.16	112.52	105.77
1	A	66	CRF	C2-CA2-N2	-5.79	104.88	108.93
1	A	66	CRF	CA2-C2-N3	5.04	105.75	103.37
1	A	66	CRF	O2-C2-CA2	-4.89	128.22	130.96
1	A	66	CRF	CB2-CA2-C2	4.62	127.79	122.28
1	A	66	CRF	C2-N3-C1	-4.18	105.85	107.97
1	A	66	CRF	C1-CA1-N1	-3.77	103.84	109.96
1	A	66	CRF	O3-C3-CA3	-2.98	117.39	126.39
1	A	66	CRF	CG2-CB2-CA2	-2.25	126.43	130.81
1	A	66	CRF	CA3-N3-C2	2.22	128.89	123.80

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 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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.