



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 1, 2014 – 03:51 AM GMT

PDB ID : 2HQK  
Title : Crystal structure of a monomeric cyan fluorescent protein derived from *Clavul*  
laria  
Authors : Henderson, J.N.; Campbell, R.E.; Ai, H.; Remington, S.J.  
Deposited on : 2006-07-18  
Resolution : 1.19 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

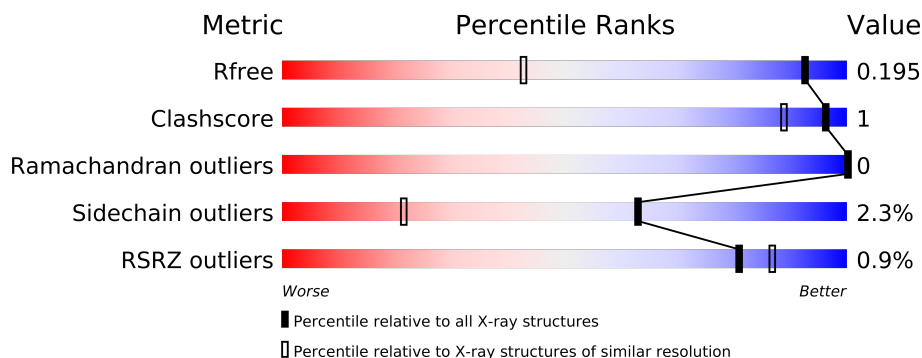
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 1.19 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))
$R_{free}$	66092	1038 (1.26-1.14)
Clashscore	79885	1158 (1.26-1.14)
Ramachandran outliers	78287	1106 (1.26-1.14)
Sidechain outliers	78261	1104 (1.26-1.14)
RSRZ outliers	66119	1038 (1.26-1.14)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	219	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
4	ACT	A	405	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 2154 atoms, of which 0 are hydrogen and 0 are deuterium.

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 chromoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	214	Total	C	N	O	S	0	22	0
			1810	1177	289	335	9			

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	42	ASN	HIS	ENGINEERED MUTATION	UNP Q9U6Y3
A	44	ILE	LEU	ENGINEERED MUTATION	UNP Q9U6Y3
A	62	THR	SER	ENGINEERED MUTATION	UNP Q9U6Y3
A	63	THR	ASN	ENGINEERED MUTATION	UNP Q9U6Y3
A	66	PIA	GLN	CHROMOPHORE	UNP Q9U6Y3
A	66	PIA	TYR	CHROMOPHORE	UNP Q9U6Y3
A	66	PIA	GLY	CHROMOPHORE	UNP Q9U6Y3
A	72	PHE	LEU	ENGINEERED MUTATION	UNP Q9U6Y3
A	80	PRO	ALA	ENGINEERED MUTATION	UNP Q9U6Y3
A	81	ASN	ASP	ENGINEERED MUTATION	UNP Q9U6Y3
A	123	HIS	ARG	ENGINEERED MUTATION	UNP Q9U6Y3
A	124	LEU	PHE	ENGINEERED MUTATION	UNP Q9U6Y3
A	125	LYS	ASP	ENGINEERED MUTATION	UNP Q9U6Y3
A	127	GLU	MET	ENGINEERED MUTATION	UNP Q9U6Y3
A	141	THR	LEU	ENGINEERED MUTATION	UNP Q9U6Y3
A	142	GLY	LYS	ENGINEERED MUTATION	UNP Q9U6Y3
A	144	ASP	GLU	ENGINEERED MUTATION	UNP Q9U6Y3
A	145	ALA	PRO	ENGINEERED MUTATION	UNP Q9U6Y3
A	149	ARG	ILE	ENGINEERED MUTATION	UNP Q9U6Y3
A	158	LYS	VAL	ENGINEERED MUTATION	UNP Q9U6Y3
A	161	VAL	ILE	ENGINEERED MUTATION	UNP Q9U6Y3
A	162	LYS	SER	ENGINEERED MUTATION	UNP Q9U6Y3
A	164	LYS	SER	ENGINEERED MUTATION	UNP Q9U6Y3
A	173	HIS	TYR	ENGINEERED MUTATION	UNP Q9U6Y3
A	175	VAL	CYS	ENGINEERED MUTATION	UNP Q9U6Y3
A	179	THR	SER	ENGINEERED MUTATION	UNP Q9U6Y3
A	182	ARG	LYS	ENGINEERED MUTATION	UNP Q9U6Y3

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	186	ALA	VAL	ENGINEERED MUTATION	UNP Q9U6Y3
A	213	VAL	LEU	ENGINEERED MUTATION	UNP Q9U6Y3
A	216	SER	ASN	ENGINEERED MUTATION	UNP Q9U6Y3
A	221	ASN	TYR	ENGINEERED MUTATION	UNP Q9U6Y3
A	223	THR	LEU	ENGINEERED MUTATION	UNP Q9U6Y3
A	224	ASP	LEU	ENGINEERED MUTATION	UNP Q9U6Y3
A	225	GLY	PRO	ENGINEERED MUTATION	UNP Q9U6Y3
A	226	MET	SER	ENGINEERED MUTATION	UNP Q9U6Y3

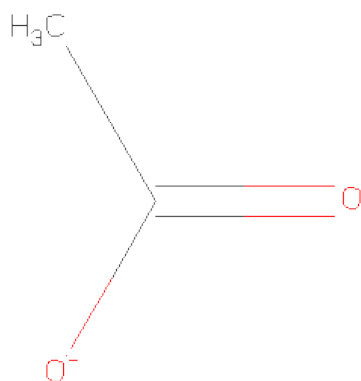
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	3	Total Zn 3 3	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Cl 1 1	0	0

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	336	Total 336	O 336	0	7

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Cyan fluorescent chromoprotein

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.83Å 38.02Å 61.12Å 90.00° 90.81° 90.00°	Depositor
Resolution (Å)	10.00 – 1.19 35.95 – 1.19	Depositor EDS
% Data completeness (in resolution range)	93.2 (10.00-1.19) 91.1 (35.95-1.19)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.09 (at 1.19Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.137 , 0.206 0.144 , 0.195	Depositor DCC
$R_{free}$ test set	3060 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	15.1	Xtriage
Anisotropy	0.068	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 66.6	EDS
Estimated twinning fraction	0.016 for -h,-k,l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Outliers	0 of 65320 reflections	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	2154	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.82% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities 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: PIA, ZN, CL, ACT

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	0.68	0/1887	1.23	14/2546 (0.5%)

There are no bond length outliers.

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	177	PHE	CB-CG-CD2	-6.95	115.93	120.80
1	A	192	TYR	CB-CG-CD1	6.82	125.09	121.00
1	A	77	ASP	CB-CG-OD2	6.71	124.34	118.30
1	A	221	ASN	CA-CB-CG	-6.59	98.91	113.40
1	A	177	PHE	CB-CG-CD1	6.47	125.33	120.80
1	A	220	ARG	NE-CZ-NH2	-5.96	117.32	120.30
1	A	192	TYR	CB-CG-CD2	-5.94	117.44	121.00
1	A	200	GLU	OE1-CD-OE2	5.82	130.29	123.30
1	A	75	TYR	CG-CD2-CE2	5.72	125.88	121.30
1	A	196	ASP	CB-CG-OD2	-5.70	113.17	118.30
1	A	198	ARG	NE-CZ-NH1	5.50	123.05	120.30
1	A	87	PHE	CB-CG-CD2	-5.41	117.01	120.80
1	A	215	GLU	OE1-CD-OE2	5.15	129.48	123.30
1	A	220	ARG	CG-CD-NE	-5.12	101.05	111.80

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit,



and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1810	0	0	2	0
2	A	3	0	0	0	0
3	A	1	0	0	0	0
4	A	4	0	0	0	0
5	A	336	0	0	2	0
All	All	2154	0	0	2	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (2) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:178[A]:LYS:NZ	5:A:571:HOH:O	2.50	0.45
1:A:203:ASN:ND2	5:A:661:HOH:O	2.48	0.45

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	228/219 (104%)	224 (98%)	4 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains ⓘ

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	194/190 (102%)	189 (97%)	5 (3%)	59 16

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	113[A]	MET
1	A	113[B]	MET
1	A	120	TYR
1	A	163	HIS
1	A	210	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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	PIA	A	66	1	21,21,22	4.00	9 (42%)	27,29,31	1.64	5 (18%)

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	PIA	A	66	1	-	0/8/27/28	0/2/2/2

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	PIA	O3-C3	15.81	1.22	1.11
1	A	66	PIA	CA3-N3	-3.83	1.44	1.47
1	A	66	PIA	CG2-CB2	-3.62	1.39	1.46
1	A	66	PIA	CA2-C2	-3.09	1.45	1.48
1	A	66	PIA	CE1-CZ	3.07	1.45	1.38
1	A	66	PIA	OH-CZ	-2.64	1.30	1.37
1	A	66	PIA	CA1-C1	-2.60	1.47	1.51
1	A	66	PIA	CB2-CA2	2.22	1.36	1.35
1	A	66	PIA	C1-N3	2.16	1.41	1.37

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	66	PIA	O2-C2-CA2	4.79	133.73	130.96
1	A	66	PIA	CA3-N3-C2	2.87	125.25	123.46
1	A	66	PIA	CB1-CA1-C1	2.76	114.41	110.34
1	A	66	PIA	CA2-C2-N3	2.34	104.77	103.44
1	A	66	PIA	CD2-CE2-CZ	2.26	122.59	119.87

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 for Mogul analysis.

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
4	ACT	A	405	-	1,3,3	3.07	1 (100%)	0,3,3	0.00	-

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	ACT	A	405	-	-	0/0/0/0	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	405	ACT	CH3-C	3.07	1.53	1.48

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	214/219 (97%)	-0.11	2 (0%) 81 87	13, 19, 30, 46	8 (3%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	130	PRO	2.6
1	A	104[A]	ILE	2.4

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
1	PIA	A	66	20/21	0.06	-0.51	12,14,16,18	0

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	ACT	A	405	4/4	0.15	5.98	29,31,31,39	4
2	ZN	A	401	1/1	0.04	-1.28	18,18,18,18	0
2	ZN	A	403	1/1	0.06	-2.19	38,38,38,38	1
2	ZN	A	402	1/1	0.05	-2.25	27,27,27,27	0
3	CL	A	404	1/1	0.03	-2.59	24,24,24,24	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.