



# Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 04:40 pm BST

PDB ID : 2H5Q  
Title : Crystal structure of mCherry  
Authors : Shu, X.; Remington, S.J.  
Deposited on : 2006-05-26  
Resolution : 1.36 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

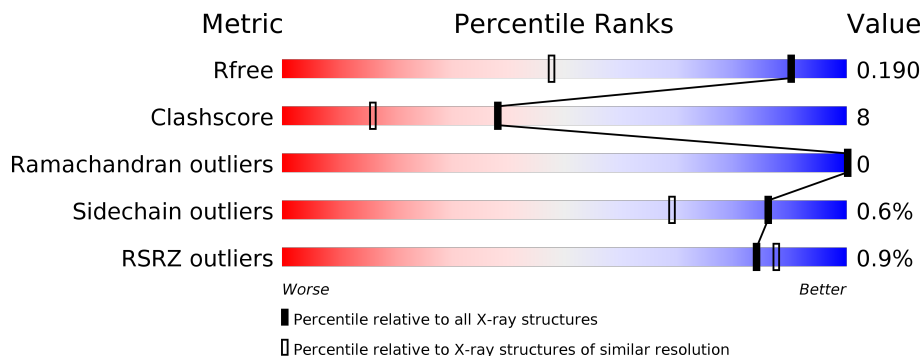
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.36 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	1509 (1.38-1.34)
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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 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  $\leq 5\%$ . 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	234	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	CH6	A	66	X	-	-	-

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2121 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 mCherry.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	218	1743	1113	293	330	7	0	4	0

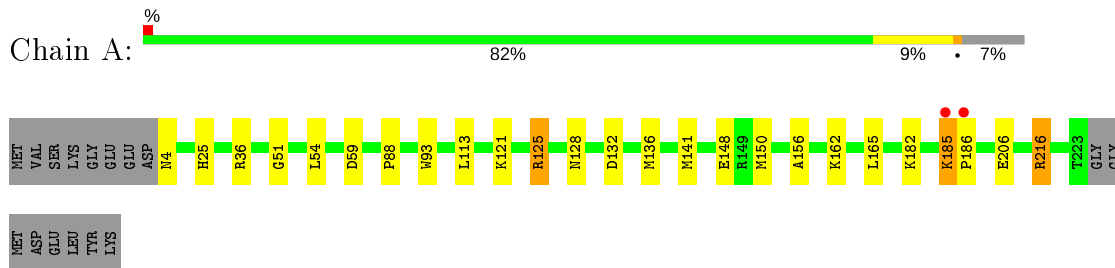
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	378	Total	O	0	2
			378	378		

### 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: mCherry



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	48.76Å 42.85Å 61.06Å 90.00° 112.31° 90.00°	Depositor
Resolution (Å)	10.00 – 1.36 20.34 – 1.36	Depositor EDS
% Data completeness (in resolution range)	94.3 (10.00-1.36) 96.4 (20.34-1.36)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 1.36Å)	Xtrriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.147 , 0.191 0.129 , 0.190	Depositor DCC
$R_{free}$ test set	2485 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.1	Xtrriage
Anisotropy	0.274	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 75.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.022 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	2121	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	18.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.89% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: CH6

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	0.66	0/1780	1.22	8/2399 (0.3%)

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	1	0

There are no bond length outliers.

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	125	ARG	CD-NE-CZ	15.96	145.95	123.60
1	A	125	ARG	NE-CZ-NH1	-8.54	116.03	120.30
1	A	125	ARG	NH1-CZ-NH2	7.01	127.11	119.40
1	A	125	ARG	NE-CZ-NH2	-6.89	116.85	120.30
1	A	93	TRP	CA-CB-CG	5.92	124.95	113.70
1	A	216	ARG	NE-CZ-NH2	-5.57	117.52	120.30
1	A	132	ASP	CB-CG-OD1	5.39	123.15	118.30
1	A	36	ARG	NE-CZ-NH1	5.21	122.90	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	66	CH6	CA1

There are no planarity outliers.

## 5.2 Too-close contacts

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	1743	0	1684	26	0
2	A	378	0	0	18	0
All	All	2121	0	1684	26	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 8.

All (26) 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:136:MET:HE1	2:A:676:HOH:O	1.85	0.76
1:A:182:LYS:HE2	2:A:433:HOH:O	1.88	0.73
1:A:141:MET:HE2	2:A:538:HOH:O	1.88	0.73
1:A:206:GLU:HG2	2:A:675:HOH:O	1.92	0.67
1:A:54:LEU:HD12	2:A:592:HOH:O	1.99	0.63
1:A:59:ASP:HB3	1:A:165:LEU:HD11	1.83	0.59
1:A:121:LYS:HG3	2:A:413:HOH:O	2.04	0.58
1:A:185:LYS:HE3	2:A:444:HOH:O	2.04	0.57
1:A:25:HIS:HE1	1:A:51:GLY:O	1.89	0.56
1:A:185:LYS:HG3	1:A:186:PRO:HD2	1.88	0.55
1:A:113:LEU:HD23	2:A:670:HOH:O	2.06	0.54
1:A:156:ALA:HB2	1:A:182:LYS:HD3	1.89	0.53
1:A:88:PRO:HB3	2:A:670:HOH:O	2.10	0.52
1:A:125:ARG:HD3	2:A:640:HOH:O	2.10	0.51
1:A:162:LYS:HE3	2:A:516:HOH:O	2.13	0.47
1:A:4:ASN:N	2:A:445:HOH:O	2.48	0.47
1:A:128:ASN:ND2	2:A:429:HOH:O	2.48	0.45
1:A:141:MET:HG3	2:A:538:HOH:O	2.17	0.43
1:A:59:ASP:CB	1:A:165:LEU:HD11	2.49	0.43
1:A:162:LYS:HE3	1:A:162:LYS:HB3	1.84	0.43
1:A:148:GLU:HG2	1:A:150:MET:SD	2.59	0.42
1:A:162:LYS:NZ	2:A:567:HOH:O	2.52	0.42
1:A:185:LYS:HG3	1:A:186:PRO:CD	2.49	0.42
1:A:216:ARG:NE	2:A:612:HOH:O	2.50	0.41
1:A:216:ARG:HD2	2:A:655:HOH:O	2.21	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:186:PRO:HB2	2:A:574:HOH:O	2.21	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	217/234 (93%)	215 (99%)	2 (1%)	0	<a href="#">100</a> <a href="#">100</a>

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	182/198 (92%)	181 (100%)	1 (0%)	<a href="#">88</a> <a href="#">74</a>

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	185	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:



Mol	Chain	Res	Type
1	A	25	HIS
1	A	188	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](#)

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	CH6	A	66	1	24,24,25	2.41	10 (41%)	28,32,34	3.80	13 (46%)

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	CH6	A	66	1	1/1/5/7	4/12/31/32	0/2/2/2

All (10) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	CH6	OH-CZ	-4.53	1.26	1.37
1	A	66	CH6	CG2-CB2	-4.28	1.38	1.46
1	A	66	CH6	C1-N3	4.06	1.44	1.37
1	A	66	CH6	CE1-CZ	3.91	1.46	1.38
1	A	66	CH6	CE2-CZ	3.51	1.45	1.38
1	A	66	CH6	CA1-N1	-3.45	1.30	1.48
1	A	66	CH6	CB2-CA2	3.21	1.37	1.35
1	A	66	CH6	C1-N2	2.99	1.36	1.32

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	CH6	CA3-N3	-2.17	1.43	1.47
1	A	66	CH6	CD2-CG2	2.00	1.43	1.39

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	66	CH6	O2-C2-CA2	14.19	138.93	130.96
1	A	66	CH6	CA2-C2-N3	6.93	106.65	103.37
1	A	66	CH6	C2-N3-C1	-6.28	104.79	107.97
1	A	66	CH6	O2-C2-N3	-5.73	112.96	124.35
1	A	66	CH6	CB1-CA1-N1	4.30	121.44	110.17
1	A	66	CH6	CG2-CB2-CA2	3.56	134.31	129.94
1	A	66	CH6	N3-C1-N2	2.74	113.35	111.45
1	A	66	CH6	CB2-CA2-C2	-2.73	119.01	122.28
1	A	66	CH6	CA1-C1-N3	-2.34	121.80	124.85
1	A	66	CH6	CB2-CA2-N2	2.21	131.90	128.83
1	A	66	CH6	CE1-CZ-CE2	-2.20	116.06	119.77
1	A	66	CH6	CA3-N3-C1	2.00	129.57	127.16
1	A	66	CH6	O3-C3-CA3	-2.00	120.35	126.39

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	66	CH6	CA1

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	66	CH6	C1-CA1-CB1-CG1
1	A	66	CH6	N2-CA2-CB2-CG2
1	A	66	CH6	C2-CA2-CB2-CG2
1	A	66	CH6	CB1-CG1-SD-CE

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	65:PHE	C	66:CH6	N1	1.18

## 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, 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	217/234 (92%)	-0.25	2 (0%) 84 87	10, 14, 26, 37	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	186	PRO	2.9
1	A	185	LYS	2.2

### 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, 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	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
1	CH6	A	66	23/24	0.98	0.06	14,17,20,40	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.