



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 11, 2017 – 12:58 PM EDT

PDB ID : 5DGY
Title : Crystal structure of rhodopsin bound to visual arrestin
Authors : Zhou, X.E.; Gao, X.; Kang, Y.; He, Y.; de Waal, P.W.; Suino-Powell, K.M.; Wang, M.; Melcher, K.; Xu, H.E.
Deposited on : unknown
Resolution : 7.70 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

<http://wwpdb.org/validation/2016/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
Xtriage (Phenix) : 1.9-1692
EDS : rb-20029824
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

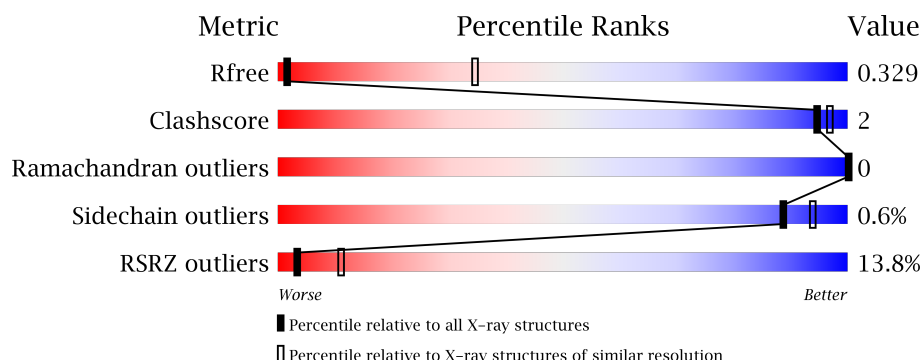
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 7.70 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}	100719	1100 (10.00-3.70)
Clashscore	112137	1036 (11.50-3.80)
Ramachandran outliers	110173	1004 (11.50-3.76)
Sidechain outliers	110143	1099 (11.50-3.70)
RSRZ outliers	101464	1004 (11.50-3.72)

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. 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	906	<div> <div>9%</div> <div>86%</div> <div>6%</div> <div>8%</div> </div>
1	B	906	<div> <div>7%</div> <div>71%</div> <div>•</div> <div>26%</div> </div>
1	C	906	<div> <div>13%</div> <div>82%</div> <div>•</div> <div>13%</div> </div>
1	D	906	<div> <div>19%</div> <div>87%</div> <div>5%</div> <div>8%</div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 24665 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 Endolysin,Rhodopsin,S-arrestin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	833	Total	C	N	O	S	0	0	0
			6565	4264	1076	1187	38			
1	B	673	Total	C	N	O	S	0	0	0
			5296	3463	847	951	35			
1	C	789	Total	C	N	O	S	0	0	0
			6231	4052	1019	1121	39			
1	D	833	Total	C	N	O	S	0	0	0
			6573	4269	1079	1186	39			

There are 104 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-149	GLY	ARG	engineered mutation	UNP P00720
A	-107	THR	CYS	engineered mutation	UNP P00720
A	-64	ALA	CYS	engineered mutation	UNP P00720
A	-24	ARG	ILE	engineered mutation	UNP P00720
A	2	CYS	ASN	engineered mutation	UNP P08100
A	113	GLN	GLU	engineered mutation	UNP P08100
A	257	TYR	MET	engineered mutation	UNP P08100
A	282	CYS	ASN	engineered mutation	UNP P08100
A	995	ALA	-	linker	UNP P08100
A	996	ALA	-	linker	UNP P08100
A	997	ALA	-	linker	UNP P08100
A	998	GLY	-	linker	UNP P08100
A	999	SER	-	linker	UNP P08100
A	1000	ALA	-	linker	UNP P08100
A	1001	GLY	-	linker	UNP P08100
A	1002	SER	-	linker	UNP P08100
A	1003	ALA	-	linker	UNP P08100
A	1004	GLY	-	linker	UNP P08100
A	1005	SER	-	linker	UNP P08100
A	1006	ALA	-	linker	UNP P08100
A	1007	GLY	-	linker	UNP P08100

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1008	SER	-	linker	UNP P08100
A	1009	ALA	-	linker	UNP P08100
A	1374	ALA	LEU	engineered mutation	UNP P20443
A	1375	ALA	VAL	engineered mutation	UNP P20443
A	1376	ALA	PHE	engineered mutation	UNP P20443
B	-149	GLY	ARG	engineered mutation	UNP P00720
B	-107	THR	CYS	engineered mutation	UNP P00720
B	-64	ALA	CYS	engineered mutation	UNP P00720
B	-24	ARG	ILE	engineered mutation	UNP P00720
B	2	CYS	ASN	engineered mutation	UNP P08100
B	113	GLN	GLU	engineered mutation	UNP P08100
B	257	TYR	MET	engineered mutation	UNP P08100
B	282	CYS	ASN	engineered mutation	UNP P08100
B	995	ALA	-	linker	UNP P08100
B	996	ALA	-	linker	UNP P08100
B	997	ALA	-	linker	UNP P08100
B	998	GLY	-	linker	UNP P08100
B	999	SER	-	linker	UNP P08100
B	1000	ALA	-	linker	UNP P08100
B	1001	GLY	-	linker	UNP P08100
B	1002	SER	-	linker	UNP P08100
B	1003	ALA	-	linker	UNP P08100
B	1004	GLY	-	linker	UNP P08100
B	1005	SER	-	linker	UNP P08100
B	1006	ALA	-	linker	UNP P08100
B	1007	GLY	-	linker	UNP P08100
B	1008	SER	-	linker	UNP P08100
B	1009	ALA	-	linker	UNP P08100
B	1374	ALA	LEU	engineered mutation	UNP P20443
B	1375	ALA	VAL	engineered mutation	UNP P20443
B	1376	ALA	PHE	engineered mutation	UNP P20443
C	-149	GLY	ARG	engineered mutation	UNP P00720
C	-107	THR	CYS	engineered mutation	UNP P00720
C	-64	ALA	CYS	engineered mutation	UNP P00720
C	-24	ARG	ILE	engineered mutation	UNP P00720
C	2	CYS	ASN	engineered mutation	UNP P08100
C	113	GLN	GLU	engineered mutation	UNP P08100
C	257	TYR	MET	engineered mutation	UNP P08100
C	282	CYS	ASN	engineered mutation	UNP P08100
C	995	ALA	-	linker	UNP P08100
C	996	ALA	-	linker	UNP P08100
C	997	ALA	-	linker	UNP P08100

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	998	GLY	-	linker	UNP P08100
C	999	SER	-	linker	UNP P08100
C	1000	ALA	-	linker	UNP P08100
C	1001	GLY	-	linker	UNP P08100
C	1002	SER	-	linker	UNP P08100
C	1003	ALA	-	linker	UNP P08100
C	1004	GLY	-	linker	UNP P08100
C	1005	SER	-	linker	UNP P08100
C	1006	ALA	-	linker	UNP P08100
C	1007	GLY	-	linker	UNP P08100
C	1008	SER	-	linker	UNP P08100
C	1009	ALA	-	linker	UNP P08100
C	1374	ALA	LEU	engineered mutation	UNP P20443
C	1375	ALA	VAL	engineered mutation	UNP P20443
C	1376	ALA	PHE	engineered mutation	UNP P20443
D	-149	GLY	ARG	engineered mutation	UNP P00720
D	-107	THR	CYS	engineered mutation	UNP P00720
D	-64	ALA	CYS	engineered mutation	UNP P00720
D	-24	ARG	ILE	engineered mutation	UNP P00720
D	2	CYS	ASN	engineered mutation	UNP P08100
D	113	GLN	GLU	engineered mutation	UNP P08100
D	257	TYR	MET	engineered mutation	UNP P08100
D	282	CYS	ASN	engineered mutation	UNP P08100
D	995	ALA	-	linker	UNP P08100
D	996	ALA	-	linker	UNP P08100
D	997	ALA	-	linker	UNP P08100
D	998	GLY	-	linker	UNP P08100
D	999	SER	-	linker	UNP P08100
D	1000	ALA	-	linker	UNP P08100
D	1001	GLY	-	linker	UNP P08100
D	1002	SER	-	linker	UNP P08100
D	1003	ALA	-	linker	UNP P08100
D	1004	GLY	-	linker	UNP P08100
D	1005	SER	-	linker	UNP P08100
D	1006	ALA	-	linker	UNP P08100
D	1007	GLY	-	linker	UNP P08100
D	1008	SER	-	linker	UNP P08100
D	1009	ALA	-	linker	UNP P08100
D	1374	ALA	LEU	engineered mutation	UNP P20443
D	1375	ALA	VAL	engineered mutation	UNP P20443
D	1376	ALA	PHE	engineered mutation	UNP P20443

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 ($\text{RSRZ} > 2$). 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.

- Chain A:
-

- [illegible]

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	107.48Å 107.26Å 460.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.98 – 7.70 29.98 – 7.25	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.98-7.70) 98.8 (29.98-7.25)	Depositor EDS
R_{merge}	0.31	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.44 (at 7.23Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.285 , 0.335 0.281 , 0.329	Depositor DCC
R_{free} test set	466 reflections (7.03%)	DCC
Wilson B-factor (Å ²)	312.8	Xtriage
Anisotropy	0.296	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 289.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.31$, $\langle L^2 \rangle = 0.14$	Xtriage
Estimated twinning fraction	0.266 for k,h,-l	Xtriage
Reported twinning fraction	0.390 for k,h,-l	Depositor
Outliers	3 of 7833 reflections (0.038%)	Xtriage
F_o, F_c correlation	0.78	EDS
Total number of atoms	24665	wwPDB-VP
Average B, all atoms (Å ²)	420.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.58% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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	0.42	1/6723 (0.0%)	0.52	1/9138 (0.0%)
1	B	0.44	2/5433 (0.0%)	0.54	1/7395 (0.0%)
1	C	0.42	1/6383 (0.0%)	0.54	2/8676 (0.0%)
1	D	0.42	3/6731 (0.0%)	0.53	1/9146 (0.0%)
All	All	0.42	7/25270 (0.0%)	0.53	5/34355 (0.0%)

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	6
1	B	0	3
1	C	0	6
1	D	0	4
All	All	0	19

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	1086	PHE	C-O	-7.99	1.08	1.23
1	B	1086	PHE	C-O	-7.42	1.09	1.23
1	A	1086	PHE	C-O	-6.99	1.10	1.23
1	B	1088	ARG	C-O	-6.78	1.10	1.23
1	C	1086	PHE	C-O	-6.34	1.11	1.23

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1088	ARG	NE-CZ-NH1	-8.23	116.19	120.30
1	A	1032	TYR	O-C-N	5.05	130.78	122.70
1	D	1032	TYR	O-C-N	5.04	130.77	122.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1032	TYR	O-C-N	5.04	130.77	122.70
1	B	1032	TYR	O-C-N	5.00	130.71	122.70

There are no chirality outliers.

5 of 19 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	0	TYR	Mainchain
1	A	1088	ARG	Mainchain
1	A	1096	GLY	Mainchain
1	A	1255	TYR	Mainchain
1	A	1257	VAL	Mainchain

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	6565	0	6626	27	13
1	B	5296	0	5340	14	6
1	C	6231	0	6287	21	7
1	D	6573	0	6647	17	22
All	All	24665	0	24900	76	24

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 2.

The worst 5 of 76 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:1184:GLU:N	1:A:1184:GLU:OE1	1.99	0.94
1:A:1246:ALA:O	1:A:1256:TYR:N	2.29	0.66
1:A:1064:CYS:N	1:A:1085:TYR:O	2.33	0.61
1:A:1183:PRO:HG2	1:A:1184:GLU:OE1	2.00	0.61
1:A:1246:ALA:N	1:A:1256:TYR:O	2.31	0.59

The worst 5 of 24 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:A:-45:ASN:N	1:D:1184:GLU:OE2[3_545]	1.15	1.05
1:C:199:ASN:ND2	1:D:1343:LEU:CA[4_446]	1.63	0.57
1:B:1184:GLU:OE2	1:D:-46:THR:N[4_456]	1.67	0.53
1:A:-45:ASN:N	1:D:1184:GLU:CD[3_545]	1.73	0.47
1:B:1184:GLU:OE1	1:D:-46:THR:CB[4_456]	1.74	0.46

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	827/906 (91%)	796 (96%)	31 (4%)	0	100	100
1	B	667/906 (74%)	641 (96%)	26 (4%)	0	100	100
1	C	781/906 (86%)	753 (96%)	28 (4%)	0	100	100
1	D	827/906 (91%)	797 (96%)	30 (4%)	0	100	100
All	All	3102/3624 (86%)	2987 (96%)	115 (4%)	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	721/777 (93%)	716 (99%)	5 (1%)	87	93
1	B	589/777 (76%)	586 (100%)	3 (0%)	91	95

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	686/777 (88%)	682 (99%)	4 (1%)	89	94
1	D	723/777 (93%)	719 (99%)	4 (1%)	89	94
All	All	2719/3108 (88%)	2703 (99%)	16 (1%)	89	94

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1174	LEU
1	C	-57	PHE
1	D	-57	PHE
1	B	268	TYR
1	D	268	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	-29	ASN
1	D	-29	ASN
1	D	1287	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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	833/906 (91%)	0.55	79 (9%) 9 16	211, 344, 728, 951	0
1	B	673/906 (74%)	0.47	65 (9%) 8 15	192, 330, 600, 858	0
1	C	789/906 (87%)	0.70	117 (14%) 3 10	226, 403, 694, 840	0
1	D	833/906 (91%)	0.93	172 (20%) 1 8	244, 456, 758, 955	0
All	All	3128/3624 (86%)	0.67	433 (13%) 3 11	192, 393, 718, 955	0

The worst 5 of 433 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1335	THR	15.7
1	A	1337	SER	14.9
1	B	1346	SER	14.8
1	A	1335	THR	14.6
1	B	1336	VAL	14.4

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

There are no non-standard protein/DNA/RNA residues in this entry.

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.