



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 12:35 am BST

PDB ID : 4CSK  
Title : human Aquaporin  
Authors : Ruiz-Carrillo, D.; To-Yiu-Ying, J.; Darwis, D.; Soon, C.H.; Cornvik, T.; Torres, J.; Lescar, J.  
Deposited on : 2014-03-08  
Resolution : 3.28 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
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

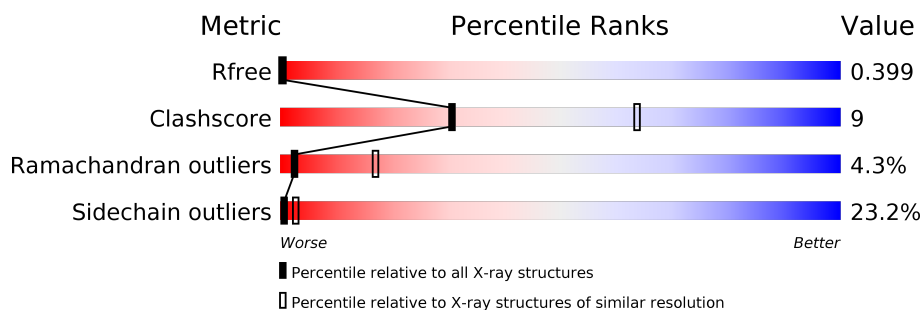
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.28 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	1177 (3.32-3.24)
Clashscore	141614	1044 (3.30-3.26)
Ramachandran outliers	138981	1026 (3.30-3.26)
Sidechain outliers	138945	1025 (3.30-3.26)

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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	292	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 1728 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 AQUAPORIN-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	233	Total	C	N	O	S	0	0	0
			1722	1115	295	307	5			

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MET	-	expression tag	UNP P29972
A	-21	GLY	-	expression tag	UNP P29972
A	-20	HIS	-	expression tag	UNP P29972
A	-19	HIS	-	expression tag	UNP P29972
A	-18	HIS	-	expression tag	UNP P29972
A	-17	HIS	-	expression tag	UNP P29972
A	-16	HIS	-	expression tag	UNP P29972
A	-15	HIS	-	expression tag	UNP P29972
A	-14	SER	-	expression tag	UNP P29972
A	-13	SER	-	expression tag	UNP P29972
A	-12	GLY	-	expression tag	UNP P29972
A	-11	VAL	-	expression tag	UNP P29972
A	-10	ASP	-	expression tag	UNP P29972
A	-9	LEU	-	expression tag	UNP P29972
A	-8	GLY	-	expression tag	UNP P29972
A	-7	THR	-	expression tag	UNP P29972
A	-6	GLU	-	expression tag	UNP P29972
A	-5	ASN	-	expression tag	UNP P29972
A	-4	LEU	-	expression tag	UNP P29972
A	-3	TYR	-	expression tag	UNP P29972
A	-2	PHE	-	expression tag	UNP P29972
A	-1	GLN	-	expression tag	UNP P29972
A	0	SER	-	expression tag	UNP P29972

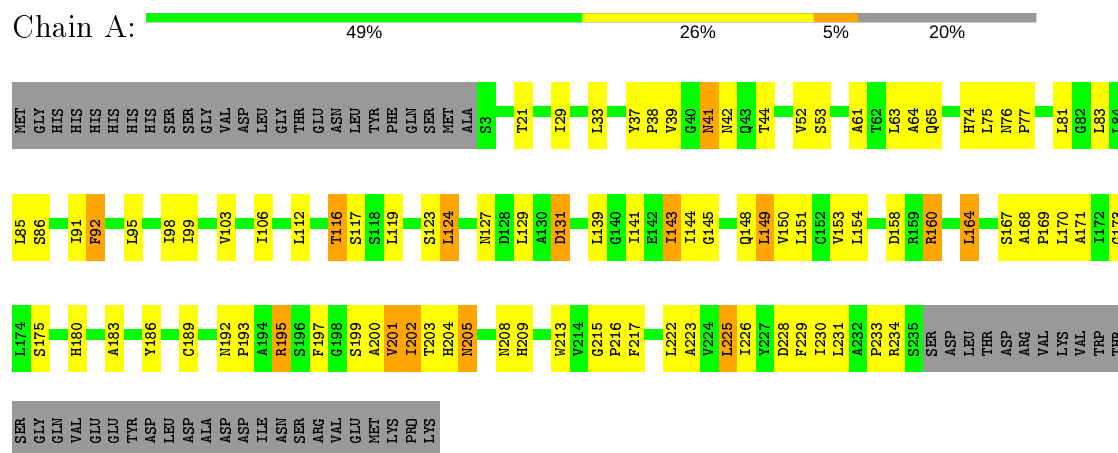
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	6	Total	O	0	0
			6	6		

### 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. 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: AQUAPORIN-1



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.28 Å   89.28 Å   174.85 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	12.49 – 3.28 12.49 – 3.28	Depositor EDS
% Data completeness (in resolution range)	99.0 (12.49-3.28) 99.2 (12.49-3.28)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 3.29 Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, $R_{free}$	0.234   ,   0.298 0.353   ,   0.399	Depositor DCC
$R_{free}$ test set	259 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	124.7	Xtriage
Anisotropy	0.111	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 176.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	1728	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	197.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.84% 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](#)

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	0.50	0/1755	0.81	0/2388

There are no bond length outliers.

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	1722	0	1776	32	0
2	A	6	0	0	0	0
All	All	1728	0	1776	32	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 9.

All (32) 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:119:LEU:HD23	1:A:123:SER:H	1.64	0.62
1:A:167:SER:OG	1:A:169:PRO:HD2	1.99	0.61
1:A:53:SER:HA	1:A:180:HIS:HB3	1.83	0.59
1:A:160:ARG:HH21	1:A:234:ARG:HA	1.69	0.58

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:226:ILE:O	1:A:231:LEU:HG	2.04	0.57
1:A:64:ALA:HB3	1:A:169:PRO:HB3	1.86	0.56
1:A:168:ALA:O	1:A:171:ALA:HB3	2.07	0.54
1:A:116:THR:HG23	1:A:117:SER:H	1.72	0.54
1:A:145:GLY:O	1:A:148:GLN:HG3	2.08	0.54
1:A:127:ASN:HD21	1:A:195:ARG:NH1	2.07	0.53
1:A:197:PHE:HB2	1:A:213:TRP:HE1	1.72	0.53
1:A:200:ALA:HA	1:A:203:THR:OG1	2.10	0.50
1:A:223:ALA:HA	1:A:226:ILE:HD12	1.92	0.50
1:A:103:VAL:HA	1:A:106:ILE:HD12	1.95	0.49
1:A:77:PRO:HD2	1:A:193:PRO:HD2	1.93	0.49
1:A:143:ILE:HD12	1:A:215:GLY:HA3	1.94	0.49
1:A:116:THR:HG21	1:A:124:LEU:HG	1.94	0.49
1:A:149:LEU:O	1:A:153:VAL:HG23	2.13	0.48
1:A:158:ASP:C	1:A:160:ARG:H	2.19	0.46
1:A:201:VAL:HG23	1:A:202:ILE:HD13	1.97	0.46
1:A:61:ALA:HB2	1:A:173:GLY:HA3	1.97	0.46
1:A:160:ARG:NH2	1:A:234:ARG:HA	2.30	0.46
1:A:149:LEU:HD23	1:A:150:VAL:HG23	2.00	0.43
1:A:92:PHE:HA	1:A:95:LEU:HD12	2.00	0.43
1:A:139:LEU:HG	1:A:143:ILE:HB	2.01	0.42
1:A:77:PRO:HB3	1:A:98:ILE:HG12	2.01	0.42
1:A:74:HIS:C	1:A:76:ASN:H	2.23	0.42
1:A:225:LEU:HA	1:A:229:PHE:CD1	2.55	0.42
1:A:226:ILE:O	1:A:230:ILE:HB	2.19	0.42
1:A:164:LEU:HD13	1:A:164:LEU:HA	1.91	0.41
1:A:41:ASN:HB3	1:A:42:ASN:H	1.66	0.41
1:A:183:ALA:HA	1:A:186:TYR:HD2	1.86	0.40

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	231/292 (79%)	184 (80%)	37 (16%)	10 (4%)	2	17

All (10) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	189	CYS
1	A	216	PRO
1	A	217	PHE
1	A	116	THR
1	A	205	ASN
1	A	75	LEU
1	A	129	LEU
1	A	204	HIS
1	A	131	ASP
1	A	38	PRO

### 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	181/234 (77%)	139 (77%)	42 (23%)	1	3

All (42) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	21	THR
1	A	29	ILE
1	A	33	LEU
1	A	37	TYR
1	A	39	VAL
1	A	41	ASN
1	A	44	THR
1	A	52	VAL
1	A	63	LEU
1	A	65	GLN
1	A	81	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	83	LEU
1	A	85	LEU
1	A	86	SER
1	A	91	ILE
1	A	92	PHE
1	A	99	ILE
1	A	112	LEU
1	A	124	LEU
1	A	131	ASP
1	A	141	ILE
1	A	143	ILE
1	A	144	ILE
1	A	149	LEU
1	A	151	LEU
1	A	154	LEU
1	A	160	ARG
1	A	164	LEU
1	A	170	LEU
1	A	175	SER
1	A	192	ASN
1	A	195	ARG
1	A	199	SER
1	A	201	VAL
1	A	202	ILE
1	A	205	ASN
1	A	208	ASN
1	A	209	HIS
1	A	222	LEU
1	A	225	LEU
1	A	228	ASP
1	A	233	PRO

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	47	GLN
1	A	69	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 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](#)

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.