wwPDB X-ray Structure Validation Summary Report (i)

Mar 24, 2022 – 03:49 pm GMT

PDB ID	:	6QF5
Title	:	X-Ray structure of human Aquaporin 2 crystallized on a silicon chip
Authors	:	Lieske, J.; Cerv, M.; Kreida, S.; Barthelmess, M.; Fischer, P.; Pakendorf, T.; Yefanov, O.; Mariani, V.; Seine, T.; Ross, B.H.; Crosas, E.; Lorbeer, O.; Burkhardt, A.; Lane, T.J.; Guenther, S.; Bergtholdt, J.; Schoen, S.; Tornroth-Horsefield, S.; Chapman, H.N.; Meents, A.
Deposited on	:	2019-01-09
Resolution	:	3.70 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
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:

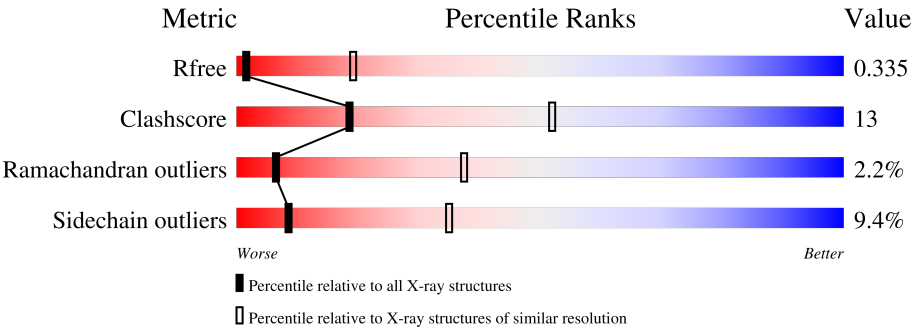
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.27
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 3.70 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}	130704	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)

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%

Mol	Chain	Length	Quality of chain
1	A	242	66% 25% · 6%
1	B	242	60% 30% 5% 5%
1	C	242	61% 27% · 8%
1	D	242	59% 31% · 9%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6616 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-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	228	Total	C	N	O	S	0	0	0
			1678	1096	281	295	6			
1	B	230	Total	C	N	O	S	0	0	0
			1687	1105	283	293	6			
1	C	222	Total	C	N	O	S	0	0	0
			1629	1067	271	285	6			
1	D	220	Total	C	N	O	S	0	0	0
			1614	1059	269	280	6			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP P41181
A	2	SER	-	cloning artifact	UNP P41181
B	1	GLY	-	cloning artifact	UNP P41181
B	2	SER	-	cloning artifact	UNP P41181
C	1	GLY	-	cloning artifact	UNP P41181
C	2	SER	-	cloning artifact	UNP P41181
D	1	GLY	-	cloning artifact	UNP P41181
D	2	SER	-	cloning artifact	UNP P41181

- Molecule 2 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Cd	0	1
			2	2		
2	C	1	Total	Cd	0	0
			1	1		
2	D	1	Total	Cd	0	0
			1	1		

- Molecule 3 is water.

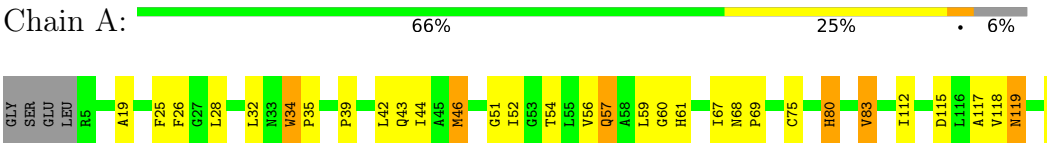
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total O 2 2	0	0
3	B	1	Total O 1 1	0	0
3	C	1	Total O 1 1	0	0

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.

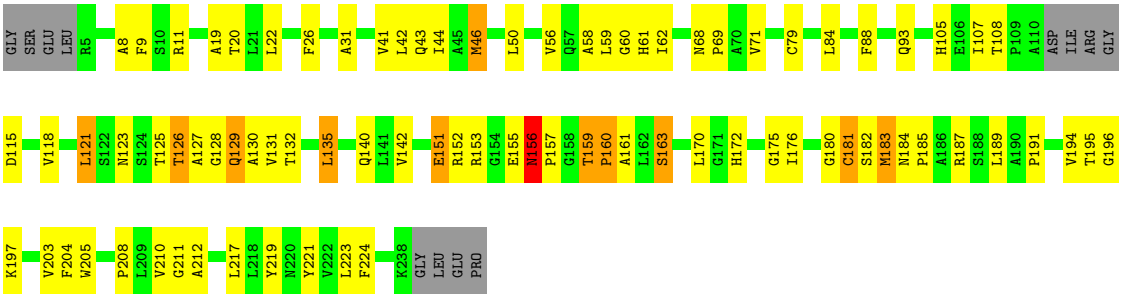
- Molecule 1: Aquaporin-2

Chain A:



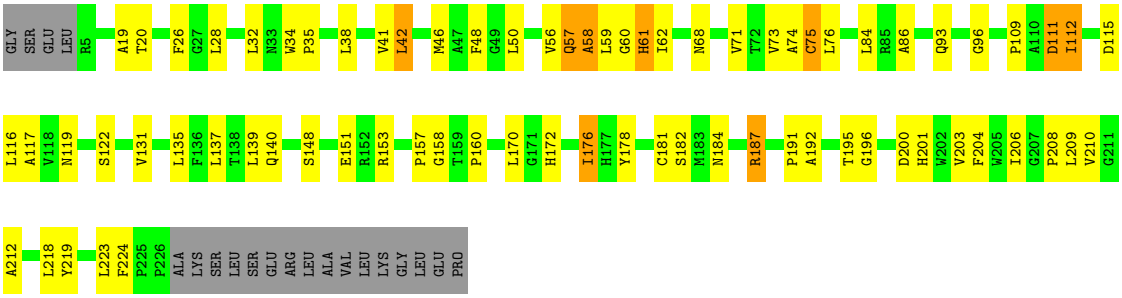
- Molecule 1: Aquaporin-2

Chain B:



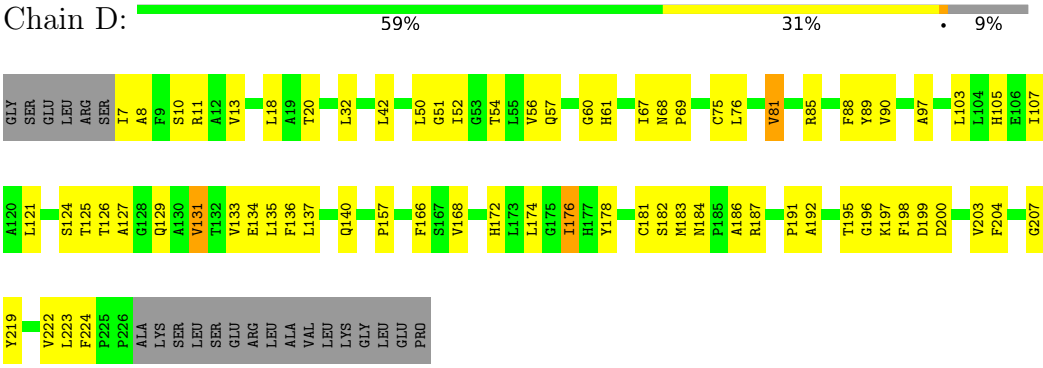
- Molecule 1: Aquaporin-2

Chain C:



• Molecule 1: Aquaporin-2

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 42	Depositor
Cell constants a, b, c, α, β, γ	122.20Å 122.20Å 94.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.89 – 3.70 37.38 – 3.70	Depositor EDS
% Data completeness (in resolution range)	99.1 (33.89-3.70) 90.0 (37.38-3.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	-0.27 (at 3.66Å)	Xtriage
Refinement program	PHENIX 1.10.1-2155_9999	Depositor
R, R_{free}	0.282 , 0.336 0.282 , 0.335	Depositor DCC
R_{free} test set	1493 reflections (10.04%)	wwPDB-VP
Wilson B-factor (Å^2)	93.4	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.060 for h,-k,-l	Xtriage
F_o,F_c correlation	0.65	EDS
Total number of atoms	6616	wwPDB-VP
Average B, all atoms (Å^2)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.02% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CD

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	0.26	0/1719	0.42	0/2352
1	B	0.31	0/1727	0.50	0/2363
1	C	0.28	0/1670	0.44	0/2288
1	D	0.28	0/1655	0.44	0/2268
All	All	0.28	0/6771	0.45	0/9271

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

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	1678	0	1706	45	0
1	B	1687	0	1721	60	0
1	C	1629	0	1650	53	0
1	D	1614	0	1641	44	0
2	B	2	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	1	0	0	0	0
All	All	6616	0	6718	176	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 13.

The worst 5 of 176 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:B:56:VAL:O	1:B:60:GLY:HA3	1.41	1.21
1:C:56:VAL:O	1:C:60:GLY:HA3	1.43	1.16
1:B:160:PRO:O	1:B:163:SER:N	1.87	1.07
1:C:172:HIS:O	1:C:176:ILE:HB	1.56	1.06
1:D:172:HIS:O	1:D:176:ILE:HB	1.55	1.06

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	226/242 (93%)	189 (84%)	32 (14%)	5 (2%)	6 37
1	B	226/242 (93%)	186 (82%)	33 (15%)	7 (3%)	4 32
1	C	220/242 (91%)	190 (86%)	27 (12%)	3 (1%)	11 45
1	D	218/242 (90%)	193 (88%)	20 (9%)	5 (2%)	6 36
All	All	890/968 (92%)	758 (85%)	112 (13%)	20 (2%)	6 37

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	156	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	159	THR
1	D	157	PRO
1	A	119	ASN
1	B	58	ALA

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	173/184 (94%)	160 (92%)	13 (8%)	13 43
1	B	173/184 (94%)	154 (89%)	19 (11%)	6 29
1	C	167/184 (91%)	150 (90%)	17 (10%)	7 31
1	D	165/184 (90%)	150 (91%)	15 (9%)	9 36
All	All	678/736 (92%)	614 (91%)	64 (9%)	8 35

5 of 64 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	136	PHE
1	D	178	TYR
1	B	152	ARG
1	B	151	GLU
1	D	187	ARG

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	57	GLN
1	B	119	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)

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.