Full wwPDB X-ray Structure Validation Report (i)

May 15, 2020 – 02:45 pm BST

PDB ID	:	6H2L
Title	:	Receptor-binding domain of Proteus mirabilis Uroepithelial Cell Adhesin UcaD21-217
Authors	:	Wangshu, J.; Knight, S.D.
Deposited on	:	2018-07-13
Resolution	:	1.50 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*
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 (1)) were used in the production of this report:

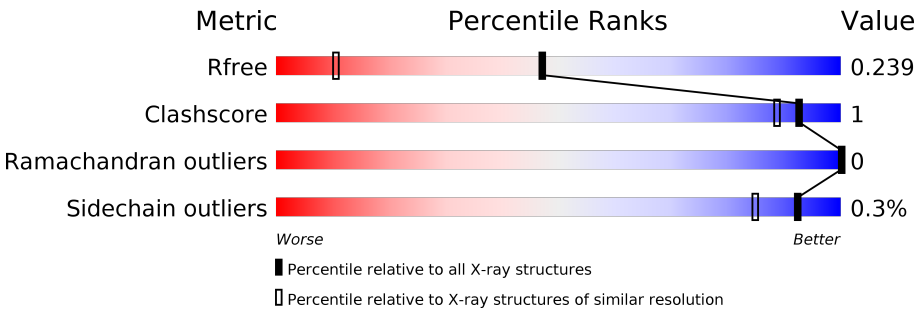
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)

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 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	204	90% • 7%
1	B	204	96% • •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6524 atoms, of which 3101 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.

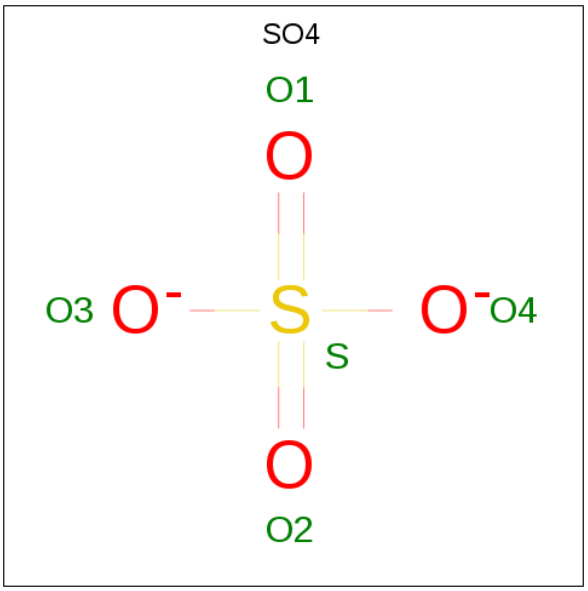
- Molecule 1 is a protein called Putative fimbrial adhesin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	190	3023	955	1526	261	278	3	0	17	0
1	B	201	3140	998	1575	275	289	3	0	12	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	25	HIS	TYR	engineered mutation	UNP B4EV65
A	218	ALA	-	expression tag	UNP B4EV65
A	219	HIS	-	expression tag	UNP B4EV65
A	220	HIS	-	expression tag	UNP B4EV65
A	221	HIS	-	expression tag	UNP B4EV65
A	222	HIS	-	expression tag	UNP B4EV65
A	223	HIS	-	expression tag	UNP B4EV65
A	224	HIS	-	expression tag	UNP B4EV65
B	25	HIS	TYR	engineered mutation	UNP B4EV65
B	218	ALA	-	expression tag	UNP B4EV65
B	219	HIS	-	expression tag	UNP B4EV65
B	220	HIS	-	expression tag	UNP B4EV65
B	221	HIS	-	expression tag	UNP B4EV65
B	222	HIS	-	expression tag	UNP B4EV65
B	223	HIS	-	expression tag	UNP B4EV65
B	224	HIS	-	expression tag	UNP B4EV65

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

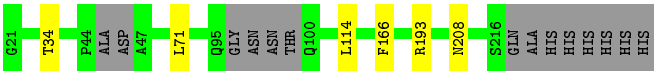
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	154	Total O 154 154	0	0
3	B	162	Total O 162 162	0	0

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.

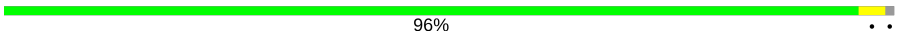
- Molecule 1: Putative fimbrial adhesin

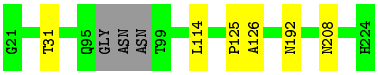
Chain A:



- Molecule 1: Putative fimbrial adhesin

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	30.47Å 49.83Å 61.83Å 86.45° 74.51° 75.52°	Depositor
Resolution (Å)	37.40 – 1.50 37.59 – 1.20	Depositor EDS
% Data completeness (in resolution range)	96.6 (37.40-1.50) 87.5 (37.59-1.20)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 1.20Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, R_{free}	0.214 , 0.238 0.215 , 0.239	Depositor DCC
R_{free} test set	4473 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	6.0	Xtriage
Anisotropy	0.354	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.45 , 47.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,-h+l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6524	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.00% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.38	0/1601	0.52	0/2185
1	B	0.39	0/1638	0.53	0/2240
All	All	0.38	0/3239	0.53	0/4425

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	1497	1526	1455	6	0
1	B	1565	1575	1550	3	0
2	A	20	0	0	0	0
2	B	25	0	0	0	0
3	A	154	0	0	0	0
3	B	162	0	0	1	0
All	All	3423	3101	3005	9	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 1.

All (9) 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:114:LEU:HD11	1:A:166:PHE:HD2	1.63	0.64
1:A:71:LEU:HD11	1:A:193:ARG:HH11	1.75	0.52
1:A:34[B]:THR:HG23	1:A:208:ASN:O	2.11	0.50
1:A:114:LEU:HD21	1:A:166:PHE:CE2	2.48	0.49
1:B:125:PRO:O	1:B:126:ALA:HB3	2.15	0.46
1:B:192[A]:ASN:OD1	3:B:401:HOH:O	2.21	0.41
1:A:71:LEU:HD11	1:A:193:ARG:NH1	2.34	0.40
1:A:114:LEU:HD11	1:A:166:PHE:CD2	2.49	0.40
1:B:31:THR:HG21	1:B:208:ASN:ND2	2.36	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	203/204 (100%)	200 (98%)	3 (2%)	0	100	100
1	B	210/204 (103%)	205 (98%)	5 (2%)	0	100	100
All	All	413/408 (101%)	405 (98%)	8 (2%)	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	173/166 (104%)	173 (100%)	0	100	100
1	B	173/166 (104%)	171 (99%)	2 (1%)	71	48
All	All	346/332 (104%)	344 (99%)	2 (1%)	92	74

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	114[A]	LEU
1	B	114[B]	LEU

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

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

9 ligands are 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	$\#\|Z\| > 2$
2	SO4	B	301	-	4,4,4	0.16	0	6,6,6	0.13	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	SO4	A	304	-	4,4,4	0.15	0	6,6,6	0.08	0
2	SO4	A	303	-	4,4,4	0.15	0	6,6,6	0.07	0
2	SO4	B	303	-	4,4,4	0.16	0	6,6,6	0.10	0
2	SO4	B	304	-	4,4,4	0.14	0	6,6,6	0.06	0
2	SO4	A	301	-	4,4,4	0.15	0	6,6,6	0.06	0
2	SO4	B	305	-	4,4,4	0.16	0	6,6,6	0.04	0
2	SO4	A	302	-	4,4,4	0.13	0	6,6,6	0.06	0
2	SO4	B	302	-	4,4,4	0.15	0	6,6,6	0.10	0

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 (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.