wwPDB X-ray Structure Validation Summary Report (i)

Jan 4, 2022 – 12:09 PM JST

PDB ID	:	7F8E
Title	:	Crystal structure of YggS from Fusobacterium nucleatum
Authors	:	Wang, L.; Chen, Y.; Bu, T.; Bai, X.
Deposited on	:	2021-07-02
Resolution	:	2.08 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 (1)) were used in the production of this report:

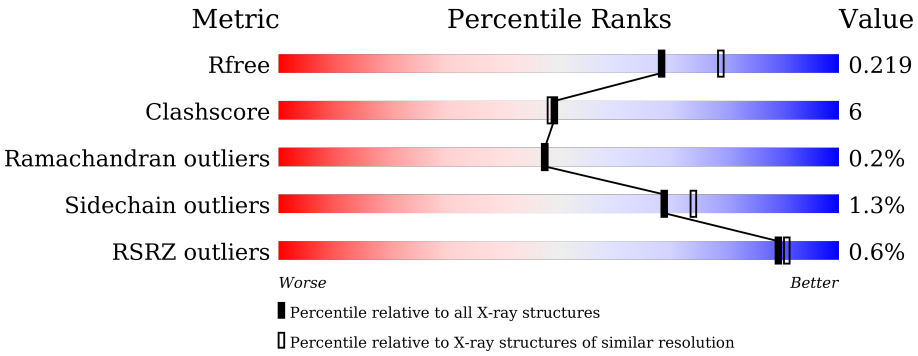
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.25
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.25

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 2.08 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	6189 (2.10-2.06)
Clashscore	141614	6738 (2.10-2.06)
Ramachandran outliers	138981	6663 (2.10-2.06)
Sidechain outliers	138945	6664 (2.10-2.06)
RSRZ outliers	127900	6057 (2.10-2.06)

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% 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	223	87% 13%
1	B	223	85% 13%
1	C	223	90% 9%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6056 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.

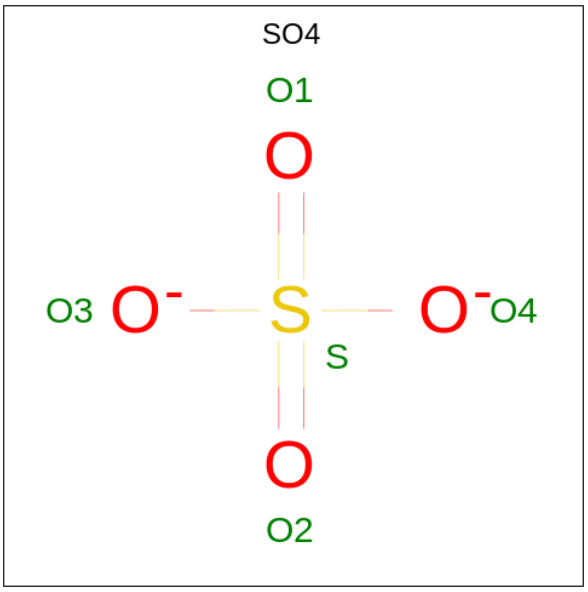
- Molecule 1 is a protein called Pyridoxal phosphate homeostasis protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	223	Total	C	N	O	S	Se	0	2	0
			1832	1175	295	352	2	8			
1	B	219	Total	C	N	O	S	Se	0	2	0
			1799	1156	289	344	2	8			
1	C	223	Total	C	N	O	S	Se	0	2	0
			1832	1175	295	352	2	8			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	5	ALA	THR	conflict	UNP Q8RFW9
A	202	SER	ASN	conflict	UNP Q8RFW9
B	5	ALA	THR	conflict	UNP Q8RFW9
B	202	SER	ASN	conflict	UNP Q8RFW9
C	5	ALA	THR	conflict	UNP Q8RFW9
C	202	SER	ASN	conflict	UNP Q8RFW9

- 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	B	1	Total O S 5 4 1	0	0
2	C	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

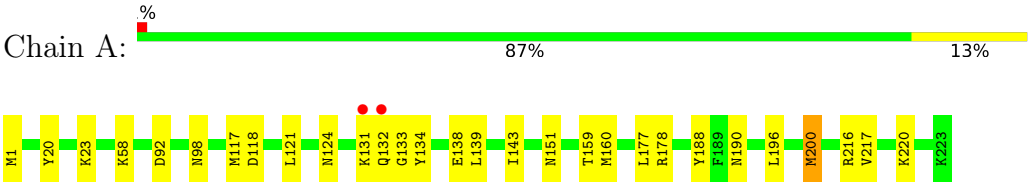
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	214	Total O 214 214	0	0
3	B	166	Total O 166 166	0	0
3	C	198	Total O 198 198	0	0

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 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.

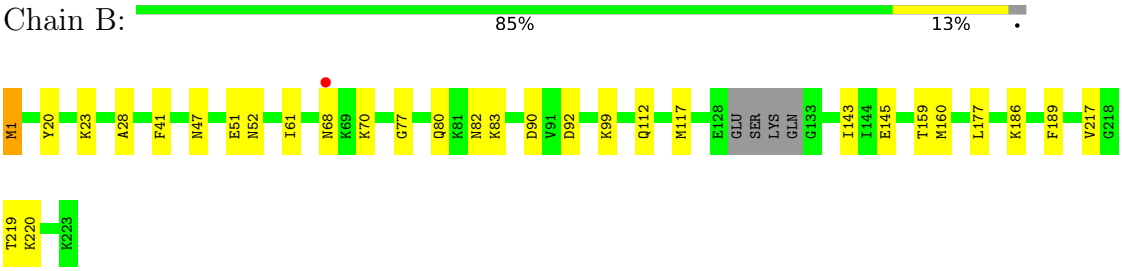
• Molecule 1: Pyridoxal phosphate homeostasis protein

Chain A:



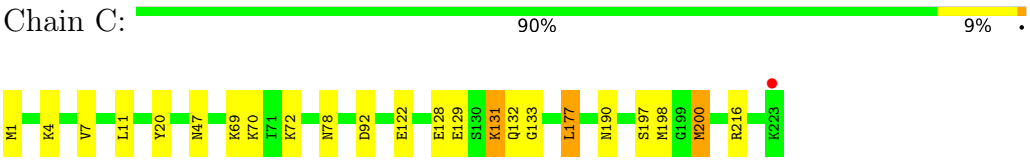
• Molecule 1: Pyridoxal phosphate homeostasis protein

Chain B:



• Molecule 1: Pyridoxal phosphate homeostasis protein

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	37.93Å 146.38Å 74.13Å 90.00° 93.36° 90.00°	Depositor
Resolution (Å)	36.59 – 2.08 40.73 – 2.08	Depositor EDS
% Data completeness (in resolution range)	96.1 (36.59-2.08) 96.1 (40.73-2.08)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.20	Depositor
$< I/\sigma(I) >$ [1]	4.13 (at 2.08Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.172 , 0.219 0.172 , 0.219	Depositor DCC
R_{free} test set	2008 reflections (4.27%)	wwPDB-VP
Wilson B-factor (Å^2)	24.5	Xtriage
Anisotropy	0.192	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 54.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6056	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 26.60 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.5573e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: 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	0.41	0/1850	0.58	0/2466
1	B	0.36	0/1816	0.54	0/2420
1	C	0.38	0/1850	0.57	1/2466 (0.0%)
All	All	0.38	0/5516	0.56	1/7352 (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	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	177	LEU	CB-CG-CD2	6.04	121.27	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	134	TYR	Peptide

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	1832	0	1894	25	1
1	B	1799	0	1861	21	1
1	C	1832	0	1894	21	0
2	A	5	0	0	0	0
2	B	5	0	0	1	0
2	C	5	0	0	0	0
3	A	214	0	0	4	0
3	B	166	0	0	6	0
3	C	198	0	0	5	0
All	All	6056	0	5649	66	1

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 6.

The worst 5 of 66 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:C:177:LEU:HB3	1:C:198:MSE:HE1	1.49	0.93
1:A:200[B]:MSE:HE3	1:A:216:ARG:HH21	1.36	0.90
1:C:131:LYS:HD3	1:C:133:GLY:H	1.36	0.89
1:C:177:LEU:HD23	1:C:198:MSE:HE3	1.54	0.88
1:C:78:ASN:HB2	1:C:132:GLN:HE22	1.52	0.74

All (1) 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:138:GLU:OE2	1:B:186:LYS:NZ[2_546]	2.09	0.11

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	223/223 (100%)	217 (97%)	6 (3%)	0	100 100
1	B	217/223 (97%)	212 (98%)	4 (2%)	1 (0%)	29 25
1	C	223/223 (100%)	221 (99%)	2 (1%)	0	100 100
All	All	663/669 (99%)	650 (98%)	12 (2%)	1 (0%)	47 47

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	68	ASN

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	210/201 (104%)	206 (98%)	4 (2%)	57 61
1	B	206/201 (102%)	204 (99%)	2 (1%)	76 81
1	C	210/201 (104%)	206 (98%)	4 (2%)	57 61
All	All	626/603 (104%)	616 (98%)	10 (2%)	69 67

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	131	LYS
1	C	200[A]	MSE
1	C	200[B]	MSE
1	A	200[B]	MSE
1	B	1	MSE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	78	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	80	GLN
1	C	132	GLN
1	B	82	ASN
1	B	68	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)

3 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	C	301	-	4,4,4	0.17	0	6,6,6	0.18	0
2	SO4	B	301	-	4,4,4	0.11	0	6,6,6	0.19	0
2	SO4	A	301	-	4,4,4	0.10	0	6,6,6	0.11	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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	301	SO4	1	0

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

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^{th} 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	216/223 (96%)	-0.28	2 (0%)	84	86	12, 24, 41, 66	0
1	B	212/223 (95%)	-0.09	1 (0%)	91	92	15, 30, 49, 71	0
1	C	216/223 (96%)	-0.23	1 (0%)	91	92	14, 25, 43, 71	0
All	All	644/669 (96%)	-0.20	4 (0%)	89	91	12, 26, 46, 71	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	68	ASN	4.1
1	C	223	LYS	3.4
1	A	131	LYS	2.9
1	A	132	GLN	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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^{th} 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^2)	Q<0.9
2	SO4	B	301	5/5	0.95	0.08	33,42,46,47	0
2	SO4	A	301	5/5	0.98	0.10	24,34,37,39	0
2	SO4	C	301	5/5	0.98	0.10	28,36,39,40	0

6.5 Other polymers (i)

There are no such residues in this entry.