wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 01:56 pm BST

PDB ID	:	1XE8
Title	:	Crystal structure of the YML079w protein from Saccharomyces cerevisiae reveals a new sequence family of the jelly roll fold.
Authors	:	Zhou, C.-Z.; Meyer, P.; Quevillon-Cheruel, S.; Li de La Sierra-Gallay, I.; Collinet, B.; Graille, M.; Leulliot, N.; Sorel, I.; Janin, J.; Van Tilbeurgh, H.
Deposited on	:	2004-09-09
Resolution	:	2.80 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:

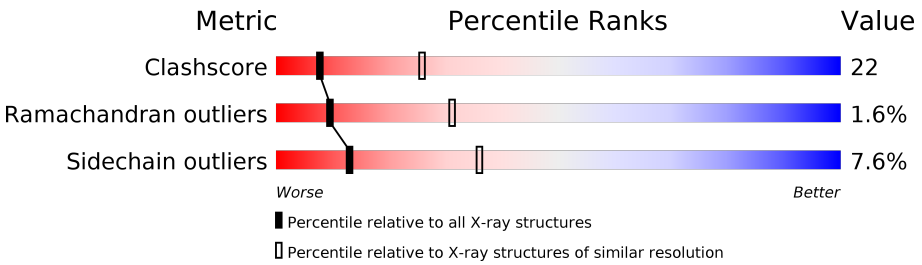
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (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 2.80 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))
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	203	60% 28% 5% • 7%
1	B	203	50% 39% 6% 5%
1	C	203	54% 33% 5% 7%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	GOL	C	1901	-	X	-	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4771 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.

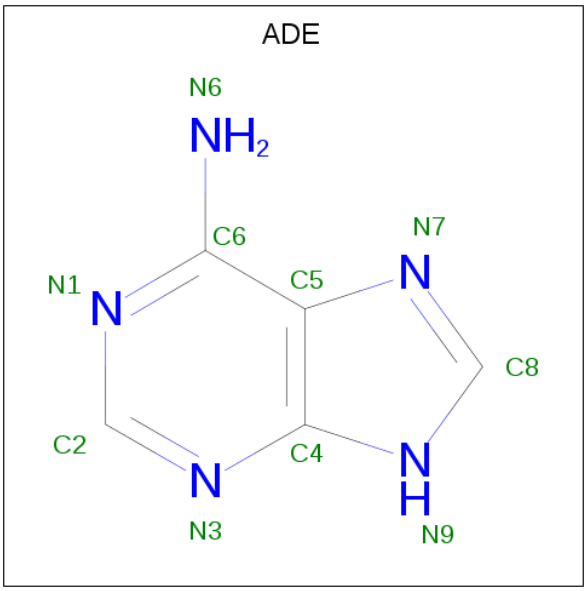
- Molecule 1 is a protein called Hypothetical 22.5 kDa protein in TUB1-CPR3 intergenic region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	189	1515	983	249	280	3	0	0	0
1	B	193	1542	998	254	287	3	0	0	0
1	C	188	1505	977	247	278	3	0	0	0

There are 15 discrepancies between the modelled and reference sequences:

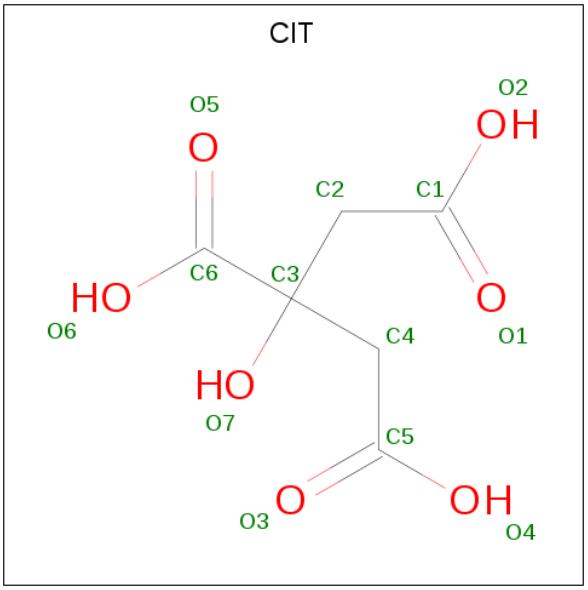
Chain	Residue	Modelled	Actual	Comment	Reference
A	23	MSE	MET	MODIFIED RESIDUE	UNP Q03629
A	58	MSE	MET	MODIFIED RESIDUE	UNP Q03629
A	73	MSE	MET	MODIFIED RESIDUE	UNP Q03629
A	202	HIS	-	EXPRESSION TAG	UNP Q03629
A	203	HIS	-	EXPRESSION TAG	UNP Q03629
B	23	MSE	MET	MODIFIED RESIDUE	UNP Q03629
B	58	MSE	MET	MODIFIED RESIDUE	UNP Q03629
B	73	MSE	MET	MODIFIED RESIDUE	UNP Q03629
B	202	HIS	-	EXPRESSION TAG	UNP Q03629
B	203	HIS	-	EXPRESSION TAG	UNP Q03629
C	23	MSE	MET	MODIFIED RESIDUE	UNP Q03629
C	58	MSE	MET	MODIFIED RESIDUE	UNP Q03629
C	73	MSE	MET	MODIFIED RESIDUE	UNP Q03629
C	202	HIS	-	EXPRESSION TAG	UNP Q03629
C	203	HIS	-	EXPRESSION TAG	UNP Q03629

- Molecule 2 is ADENINE (three-letter code: ADE) (formula: $C_5H_5N_5$).



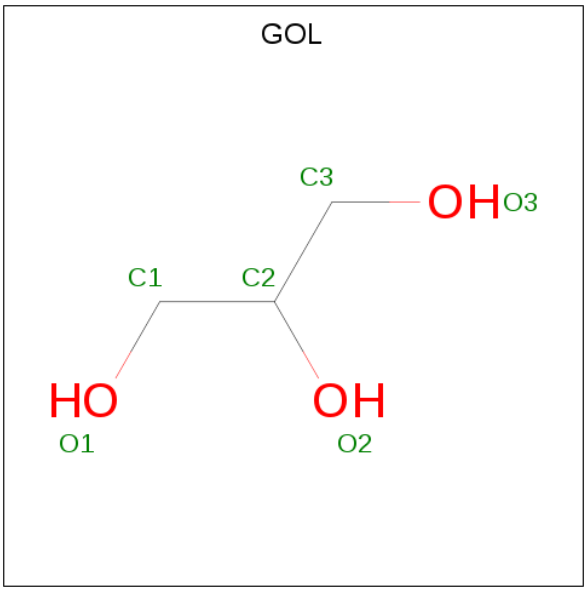
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C N 10 5 5	0	0

- Molecule 3 is CITRIC ACID (three-letter code: CIT) (formula: $C_6H_8O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 13 6 7	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	C	1	6	3	3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	78	78	78	0	0
			Total	O		
5	B	40	40	40	0	0
			Total	O		
5	C	62	62	62	0	0

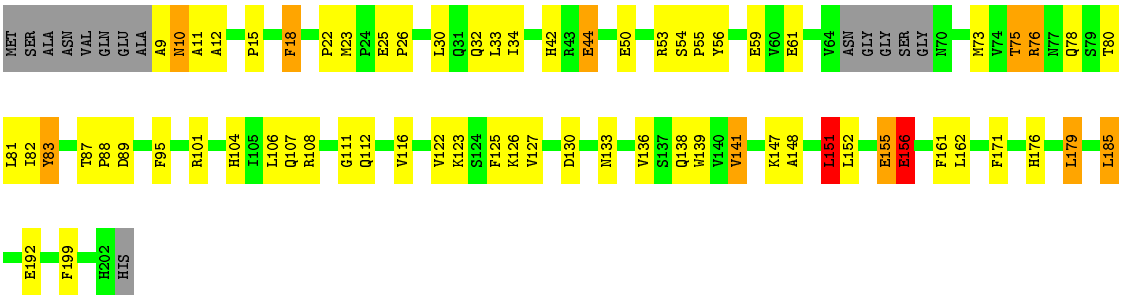
3 Residue-property plots (i)

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.

Note EDS was not executed.

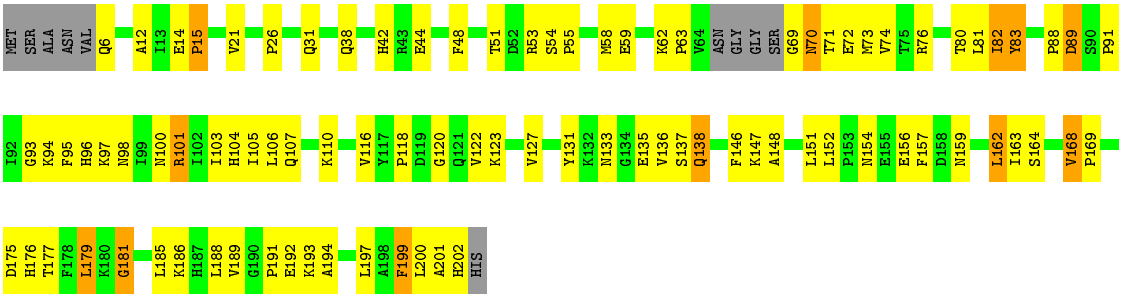
• Molecule 1: Hypothetical 22.5 kDa protein in TUB1-CPR3 intergenic region

Chain A: 60% 28% 5% • 7%



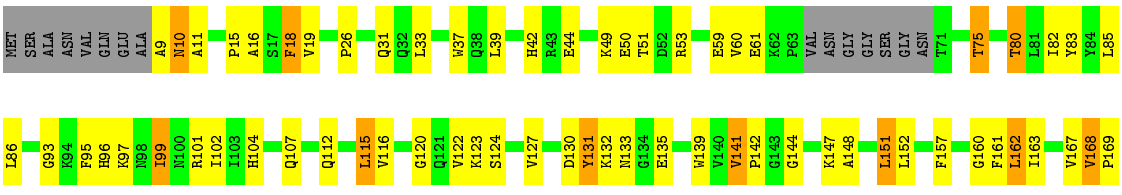
• Molecule 1: Hypothetical 22.5 kDa protein in TUB1-CPR3 intergenic region

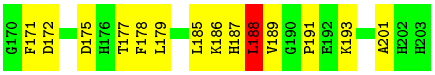
Chain B: 50% 39% 6% 5%



• Molecule 1: Hypothetical 22.5 kDa protein in TUB1-CPR3 intergenic region

Chain C: 54% 33% 5% 7%




G170
F171
D172
D175
H176
T177
F178
L179
L185
K186
H187
L188
V189
G190
P191
E192
K193
A201
H202
H203

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	I 4 3 2	Depositor
Cell constants a, b, c, α, β, γ	207.97Å 207.97Å 207.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.80	Depositor
% Data completeness (in resolution range)	98.8 (20.00-2.80)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.20	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.223 , 0.284	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4771	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

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: GOL, ADE, CIT

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	0.43	0/1554	0.70	3/2102 (0.1%)
1	B	0.38	0/1581	0.64	1/2138 (0.0%)
1	C	0.42	0/1544	0.68	1/2088 (0.0%)
All	All	0.41	0/4679	0.67	5/6328 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	83	TYR	N-CA-C	-5.53	96.06	111.00
1	A	155	GLU	N-CA-C	-5.50	96.15	111.00
1	B	93	GLY	N-CA-C	-5.23	100.01	113.10
1	C	93	GLY	N-CA-C	-5.20	100.11	113.10
1	A	151	LEU	CA-CB-CG	5.03	126.88	115.30

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	1515	0	1487	59	0
1	B	1542	0	1509	77	0
1	C	1505	0	1474	68	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	10	0	4	1	0
3	A	13	0	5	1	0
4	C	6	0	4	0	0
5	A	78	0	0	4	0
5	B	40	0	0	1	0
5	C	62	0	0	1	0
All	All	4771	0	4483	198	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 22.

The worst 5 of 198 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:138:GLN:HE21	1:B:51:THR:HG21	1.25	0.99
1:B:58:MSE:HE2	1:B:169:PRO:HG3	1.51	0.93
1:A:138:GLN:NE2	1:B:51:THR:HG21	1.87	0.89
1:C:188:LEU:HD13	1:C:189:VAL:HG23	1.52	0.89
1:C:102:ILE:HG12	1:C:167:VAL:HG22	1.54	0.89

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	185/203 (91%)	173 (94%)	10 (5%)	2 (1%)	14 41
1	B	189/203 (93%)	170 (90%)	15 (8%)	4 (2%)	7 23
1	C	184/203 (91%)	173 (94%)	8 (4%)	3 (2%)	9 31
All	All	558/609 (92%)	516 (92%)	33 (6%)	9 (2%)	9 31

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	10	ASN
1	B	15	PRO
1	B	181	GLY
1	C	10	ASN
1	C	131	TYR

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	166/172 (96%)	153 (92%)	13 (8%)	12	35
1	B	168/172 (98%)	157 (94%)	11 (6%)	17	44
1	C	164/172 (95%)	150 (92%)	14 (8%)	10	31
All	All	498/516 (96%)	460 (92%)	38 (8%)	13	36

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	91	PRO
1	B	168	VAL
1	C	162	LEU
1	B	138	GLN
1	B	179	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	38	GLN
1	B	96	HIS
1	C	133	ASN
1	B	42	HIS
1	B	78	GLN

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)

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	ADE	A	500	-	9,11,11	0.92	0	7,15,15	1.16	1 (14%)
4	GOL	C	1901	-	5,5,5	4.20	4 (80%)	5,5,5	5.77	3 (60%)
3	CIT	A	1102	-	3,12,12	2.63	1 (33%)	3,17,17	4.01	2 (66%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	ADE	A	500	-	-	-	0/2/2/2
4	GOL	C	1901	-	-	2/4/4/4	-
3	CIT	A	1102	-	-	3/6/16/16	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	1901	GOL	C3-C2	-7.21	1.22	1.51
4	C	1901	GOL	O1-C1	4.68	1.62	1.42
3	A	1102	CIT	O7-C3	4.49	1.50	1.43
4	C	1901	GOL	C1-C2	-2.60	1.41	1.51
4	C	1901	GOL	O2-C2	-2.28	1.36	1.43

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	1901	GOL	O3-C3-C2	10.59	160.98	110.20
4	C	1901	GOL	O2-C2-C3	6.60	138.19	109.12
3	A	1102	CIT	C3-C4-C5	5.06	123.09	114.98
3	A	1102	CIT	C3-C2-C1	4.70	122.51	114.98
4	C	1901	GOL	O1-C1-C2	3.18	125.46	110.20

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	1901	GOL	C1-C2-C3-O3
3	A	1102	CIT	C1-C2-C3-O7
3	A	1102	CIT	C1-C2-C3-C4
3	A	1102	CIT	C1-C2-C3-C6
4	C	1901	GOL	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	500	ADE	1	0
3	A	1102	CIT	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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.