Full wwPDB X-ray Structure Validation Report (i)

Mar 16, 2022 – 12:06 am GMT

PDB ID	:	7QHY
Title	:	Structure of a Kluyveromyces lactis protein involved in RNA decay
Authors	:	Barbarin-Bocahu, I.; Graille, M.
Deposited on	:	2021-12-14
Resolution	:	2.45 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:

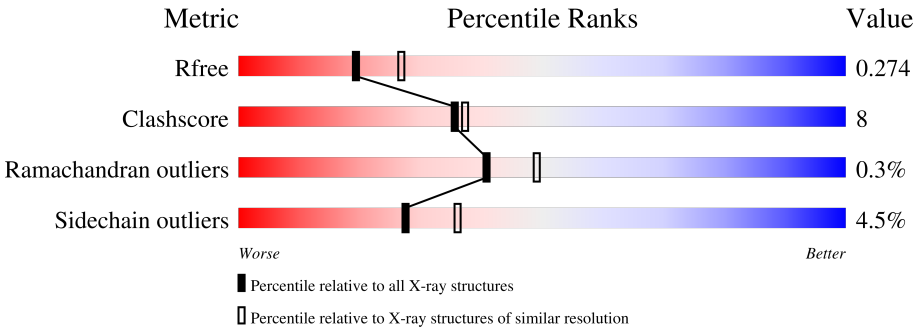
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.27
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.45 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)

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	252	58% 15% · 25%
1	B	252	56% 12% · 32%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3027 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.

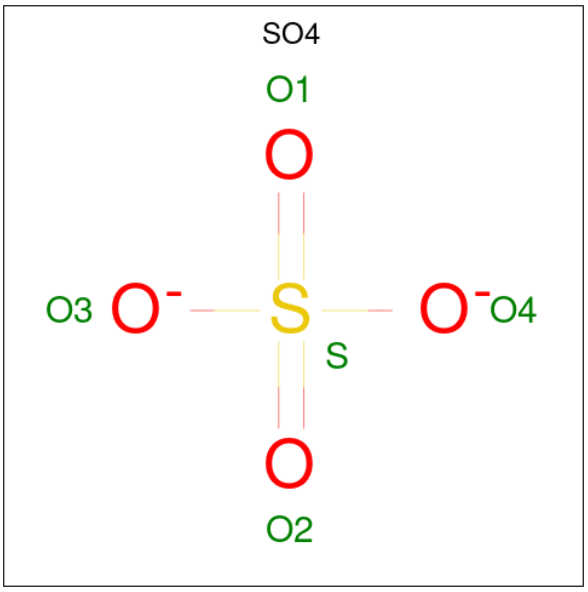
- Molecule 1 is a protein called Nonsense-mediated decay protein 4,Nonsense-mediated decay protein 4,Nonsense mediated mRNA decay protein 4 (Nmd4).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	190	Total	C	N	O	S	0	1	0
			1557	1010	259	284	4			
1	B	172	Total	C	N	O	S	0	0	0
			1398	904	235	255	4			

There are 8 discrepancies between the modelled and reference sequences:

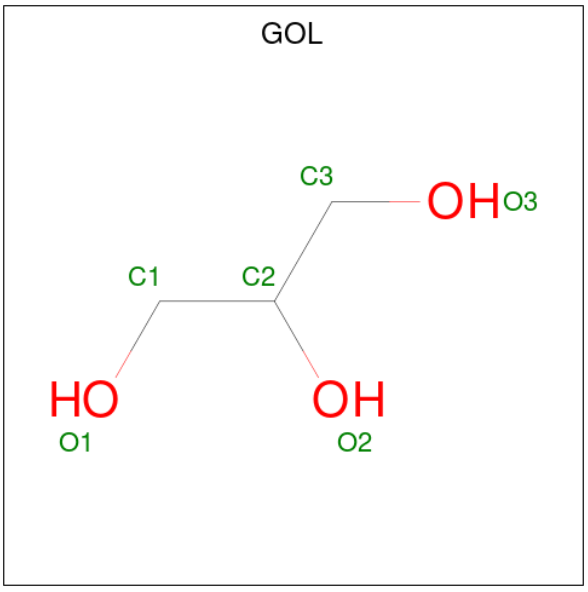
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q6CVZ8
A	-1	PRO	-	expression tag	UNP Q6CVZ8
A	0	GLY	-	expression tag	UNP Q6CVZ8
A	1	SER	-	expression tag	UNP Q6CVZ8
B	-2	GLY	-	expression tag	UNP Q6CVZ8
B	-1	PRO	-	expression tag	UNP Q6CVZ8
B	0	GLY	-	expression tag	UNP Q6CVZ8
B	1	SER	-	expression tag	UNP Q6CVZ8

- 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	B	1	Total O S 5 4 1	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

Continued on next page...

Continued from previous page...

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

- Molecule 4 is water.

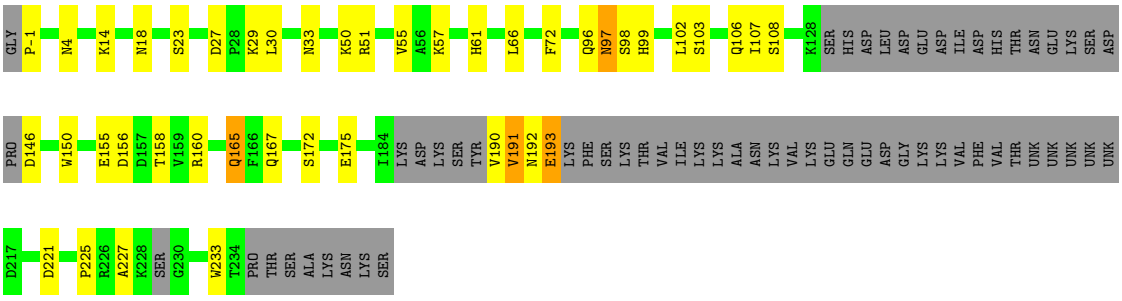
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	22	Total O 22 22	0	0
4	B	11	Total O 11 11	0	0

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.

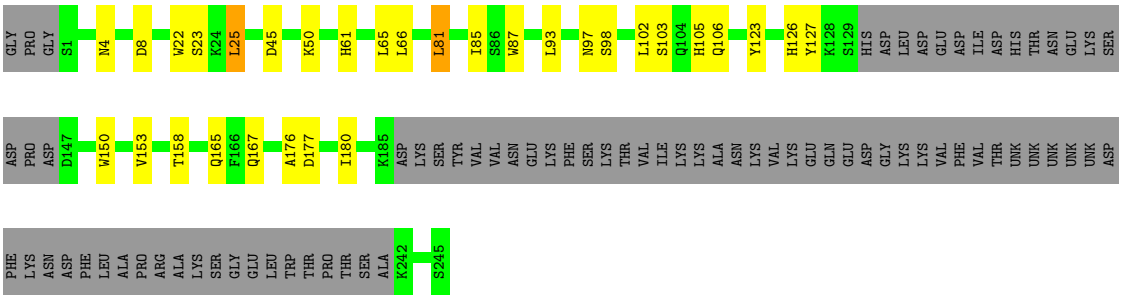
• Molecule 1: Nonsense-mediated decay protein 4,Nonsense-mediated decay protein 4,Nonsense mediated mRNA decay protein 4 (Nmd4)

Chain A:



• Molecule 1: Nonsense-mediated decay protein 4,Nonsense-mediated decay protein 4,Nonsense mediated mRNA decay protein 4 (Nmd4)

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	76.82Å 76.82Å 174.31Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.76 – 2.45 43.76 – 2.45	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.76-2.45) 99.9 (43.76-2.45)	Depositor EDS
R_{merge}	0.01	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.32 (at 2.45Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.230 , 0.269 0.233 , 0.274	Depositor DCC
R_{free} test set	1131 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	75.7	Xtriage
Anisotropy	0.117	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3027	wwPDB-VP
Average B, all atoms (Å^2)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.84% 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: GOL, 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	0.46	0/1591	0.60	0/2150
1	B	0.37	0/1426	0.56	0/1924
All	All	0.42	0/3017	0.58	0/4074

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	1557	0	1560	31	0
1	B	1398	0	1428	20	0
2	A	10	0	0	0	0
2	B	5	0	0	0	0
3	B	24	0	32	6	0
4	A	22	0	0	0	0
4	B	11	0	0	0	0
All	All	3027	0	3020	49	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 8.

All (49) 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:103:SER:H	1:A:106:GLN:HE21	1.35	0.74
1:A:-1:PRO:HD2	1:A:29:LYS:NZ	2.05	0.72
1:A:18:ASN:HD21	1:A:227:ALA:H	1.37	0.71
1:A:-1:PRO:HD2	1:A:29:LYS:HZ1	1.54	0.70
1:A:172:SER:OG	1:A:175:GLU:HG3	1.99	0.62
1:B:87:TRP:H	3:B:301:GOL:H2	1.63	0.62
1:A:55:VAL:HG21	1:A:225:PRO:HG2	1.80	0.61
1:A:191:VAL:HG23	1:A:193:GLU:HG3	1.82	0.60
1:A:221:ASP:OD1	1:B:105:HIS:HE1	1.86	0.58
1:A:103:SER:H	1:A:106:GLN:NE2	2.01	0.58
1:A:50:LYS:HD2	1:B:50:LYS:HD2	1.87	0.57
1:B:102:LEU:HG	1:B:165:GLN:HG2	1.88	0.55
1:A:4:ASN:ND2	1:A:33:ASN:HD22	2.06	0.54
1:A:51:ARG:HH11	3:B:303:GOL:H32	1.72	0.54
1:A:102:LEU:HG	1:A:165:GLN:HG2	1.90	0.54
1:A:156:ASP:O	1:A:160:ARG:HG3	2.09	0.53
1:B:81:LEU:O	1:B:85:ILE:HG13	2.09	0.53
1:A:175:GLU:OE2	1:A:192:ASN:ND2	2.43	0.52
1:B:103:SER:H	1:B:106:GLN:HE21	1.56	0.51
1:A:97:ASN:HB2	1:A:167:GLN:HE22	1.75	0.51
1:B:22:TRP:HA	1:B:25:LEU:HD22	1.94	0.50
1:B:123:TYR:HE2	3:B:302:GOL:H12	1.76	0.49
1:A:18:ASN:ND2	1:A:227:ALA:H	2.07	0.49
1:B:97:ASN:HB2	1:B:167:GLN:HE22	1.78	0.49
1:A:-1:PRO:CD	1:A:29:LYS:NZ	2.76	0.48
1:A:191:VAL:O	1:A:233[A]:TRP:O	2.31	0.48
1:A:51:ARG:NH1	3:B:303:GOL:H32	2.29	0.48
1:A:57:LYS:HE2	1:A:61:HIS:CE1	2.49	0.47
1:A:4:ASN:HB3	1:A:150:TRP:CD2	2.50	0.47
1:A:191:VAL:O	1:A:233[B]:TRP:O	2.32	0.47
1:A:23:SER:HA	1:A:72:PHE:CG	2.51	0.46
1:B:97:ASN:HB2	1:B:167:GLN:NE2	2.30	0.46
1:A:4:ASN:HD22	1:A:33:ASN:HB2	1.80	0.46
1:B:61:HIS:CE1	1:B:65:LEU:HD11	2.51	0.45
1:A:27:ASP:HB3	1:A:30:LEU:HB3	1.99	0.45
1:A:99:HIS:H	1:A:99:HIS:CD2	2.36	0.44
1:B:103:SER:H	1:B:106:GLN:NE2	2.15	0.44
1:A:103:SER:O	1:A:107:ILE:HG12	2.17	0.44
1:B:87:TRP:HB3	3:B:301:GOL:H2	2.00	0.43
1:A:23:SER:HB2	1:A:66:LEU:HD11	2.01	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:176:ALA:O	1:B:180:ILE:HG13	2.20	0.42
1:B:8:ASP:HB2	1:B:153:VAL:O	2.20	0.41
1:B:123:TYR:HA	1:B:127:TYR:HB2	2.01	0.41
1:A:14:LYS:HD3	1:A:155:GLU:HB3	2.02	0.41
1:B:93:LEU:HD13	1:B:126:HIS:CD2	2.55	0.41
1:B:123:TYR:CE2	3:B:302:GOL:H12	2.55	0.41
1:B:4:ASN:HB3	1:B:150:TRP:CD2	2.55	0.41
1:B:23:SER:HB2	1:B:66:LEU:HD11	2.03	0.40
1:A:97:ASN:HB2	1:A:167:GLN:NE2	2.37	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	181/252 (72%)	174 (96%)	6 (3%)	1 (1%)	25 29
1	B	166/252 (66%)	157 (95%)	9 (5%)	0	100 100
All	All	347/504 (69%)	331 (95%)	15 (4%)	1 (0%)	41 49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	96	GLN

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	175/227 (77%)	166 (95%)	9 (5%)	24	31
1	B	160/227 (70%)	154 (96%)	6 (4%)	33	43
All	All	335/454 (74%)	320 (96%)	15 (4%)	27	36

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	97	ASN
1	A	98	SER
1	A	108	SER
1	A	146	ASP
1	A	158	THR
1	A	165	GLN
1	A	190	VAL
1	A	191	VAL
1	A	193	GLU
1	B	25	LEU
1	B	45	ASP
1	B	81	LEU
1	B	98	SER
1	B	158	THR
1	B	177	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (15) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	4	ASN
1	A	18	ASN
1	A	61	HIS
1	A	95	GLN
1	A	99	HIS
1	A	106	GLN
1	A	192	ASN
1	B	4	ASN
1	B	47	GLN
1	B	61	HIS
1	B	75	HIS
1	B	105	HIS
1	B	106	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	126	HIS
1	B	167	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

7 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	$\#\|Z\| > 2$
3	GOL	B	304	-	5,5,5	0.16	0	5,5,5	0.45	0
3	GOL	B	303	-	5,5,5	0.17	0	5,5,5	0.29	0
3	GOL	B	302	-	5,5,5	0.07	0	5,5,5	0.26	0
2	SO4	B	305	-	4,4,4	0.18	0	6,6,6	0.05	0
3	GOL	B	301	-	5,5,5	0.19	0	5,5,5	0.24	0
2	SO4	A	302	-	4,4,4	0.19	0	6,6,6	0.13	0
2	SO4	A	301	-	4,4,4	0.14	0	6,6,6	0.14	0

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
3	GOL	B	304	-	-	0/4/4/4	-
3	GOL	B	303	-	-	0/4/4/4	-
3	GOL	B	302	-	-	0/4/4/4	-
3	GOL	B	301	-	-	0/4/4/4	-

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.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	303	GOL	2	0
3	B	302	GOL	2	0
3	B	301	GOL	2	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

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.