W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 23, 2023 – 12:44 AM JST

PDB ID	:	7Y18
Title	:	Crystal structure of ribosomal ITS2 pre-rRNA processing complex from Saccharomyces cerevisiae
Authors	:	Chen, J.; Liu, L.
Deposited on	:	2022-06-07
Resolution	:	3.69 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

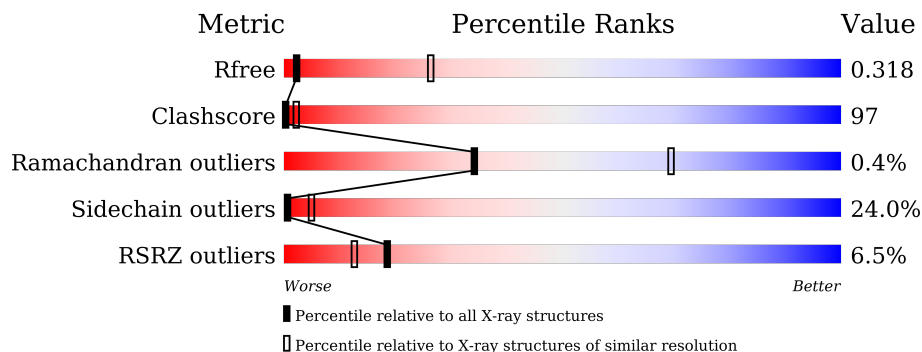|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 3.69 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	Percentile Ranks	Value
Rfree		0.318
Clashscore		97
Ramachandran outliers		0.4%
Sidechain outliers		24.0%
RSRZ outliers		6.5%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

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)
RSRZ outliers	127900	1578 (3.90-3.50)

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 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	632	2% 18% 43% 22% 16%
1	B	632	4% 23% 42% 17% • 16%
1	E	632	2% 22% 43% 17% • 17%
2	C	502	8% 31% 40% 14% 14%
2	D	502	7% 34% 39% 11% • 15%
2	F	502	12% 33% 34% 11% • 21%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 23039 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.

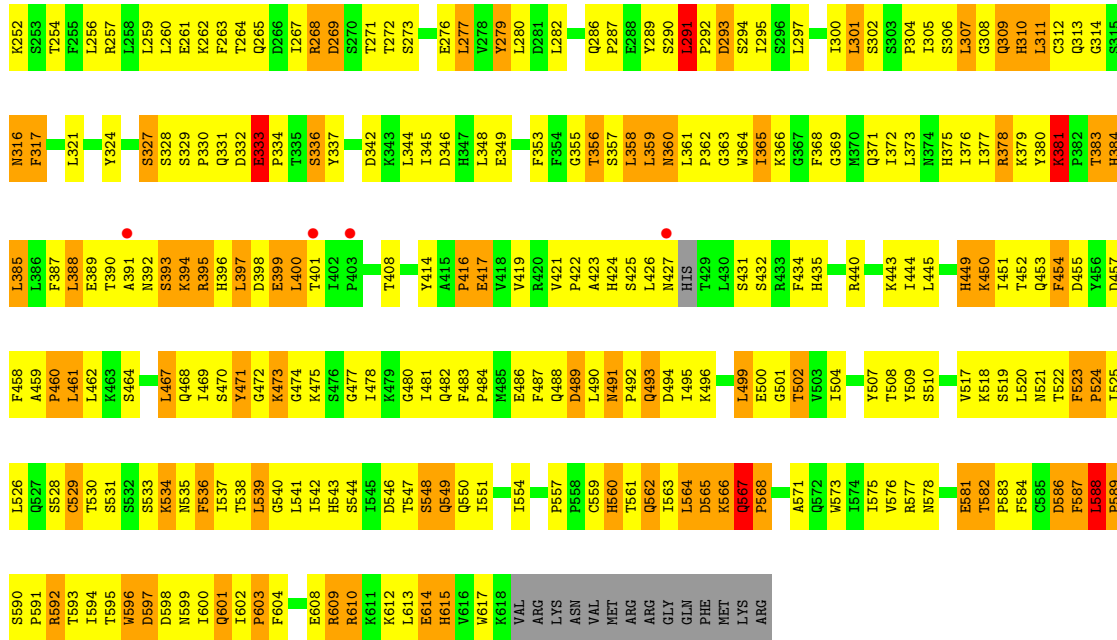- Molecule 1 is a protein called Polynucleotide 5'-hydroxyl-kinase GRC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	529	Total	C	N	O	S	0	0	0
			4122	2649	693	771	9			
1	B	528	Total	C	N	O	S	0	0	0
			4069	2607	693	759	10			
1	E	527	Total	C	N	O	S	0	0	0
			4076	2614	691	761	10			

- Molecule 2 is a protein called Protein LAS1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	430	Total	C	N	O	S	0	0	0
			3614	2324	629	652	9			
2	D	429	Total	C	N	O	S	0	0	0
			3614	2325	628	652	9			
2	F	397	Total	C	N	O	S	0	0	0
			3364	2176	584	595	9			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	36	Total	O	0	0
			36	36		
3	C	31	Total	O	0	0
			31	31		
3	D	21	Total	O	0	0
			21	21		
3	B	30	Total	O	0	0
			30	30		
3	E	43	Total	O	0	0
			43	43		
3	F	19	Total	O	0	0
			19	19		

3   Residue-property plots ⓘ

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: Polynucleotide 5'-hydroxyl-kinase GRC3

Chain A:

• Molecule 1: Polynucleotide 5'-hydroxyl-kinase GRC3

Chain B:

• Molecule 1: Polynucleotide 5'-hydroxyl-kinase GRC3

• Molecule 2: Protein LAS1

Chain C:

• Molecule 2: Protein LAS1

Chain D:

• Molecule 2: Protein LAS1

Chain F:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	233.56Å 116.14Å 159.31Å 90.00° 96.41° 90.00°	Depositor
Resolution (Å)	48.63 – 3.69 48.63 – 3.69	Depositor EDS
% Data completeness (in resolution range)	80.1 (48.63-3.69) 80.2 (48.63-3.69)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.03 (at 3.67Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.278 , 0.318 0.278 , 0.318	Depositor DCC
R_{free} test set	1824 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	43.4	Xtriage
Anisotropy	0.586	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 63.0	EDS
L-test for twinning[2]	$< |L| > = 0.39, < L^2 > = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.77	EDS
Total number of atoms	23039	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 3.42% 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)

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	0.56	15/4230 (0.4%)	0.72	19/5764 (0.3%)
1	B	0.46	10/4169 (0.2%)	0.73	22/5676 (0.4%)
1	E	0.49	14/4179 (0.3%)	0.67	15/5690 (0.3%)
2	C	0.42	5/3703 (0.1%)	0.57	7/5016 (0.1%)
2	D	0.42	4/3704 (0.1%)	0.56	7/5017 (0.1%)
2	F	0.43	4/3445 (0.1%)	0.59	8/4658 (0.2%)
All	All	0.47	52/23430 (0.2%)	0.65	78/31821 (0.2%)

The worst 5 of 52 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	304	PRO	N-CD	5.72	1.55	1.47
1	A	304	PRO	N-CD	5.55	1.55	1.47
1	B	191	PRO	N-CD	5.38	1.55	1.47
1	B	194	PRO	N-CD	5.38	1.55	1.47
1	B	123	PRO	N-CD	5.38	1.55	1.47

The worst 5 of 78 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	197	PRO	N-CA-C	-13.72	76.42	112.10
1	B	618	LYS	N-CA-C	-10.92	81.53	111.00
1	B	198	LEU	N-CA-CB	-9.77	90.87	110.40
1	B	617	TRP	CB-CA-C	-8.25	93.90	110.40
1	B	617	TRP	N-CA-C	-7.62	90.42	111.00

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	4122	0	3931	1015	0
1	B	4069	0	3873	963	0
1	E	4076	0	3880	878	3
2	C	3614	0	3553	694	0
2	D	3614	0	3551	665	0
2	F	3364	0	3350	537	6
3	A	36	0	0	4	0
3	B	30	0	0	2	0
3	C	31	0	0	3	0
3	D	21	0	0	1	0
3	E	43	0	0	8	0
3	F	19	0	0	1	0
All	All	23039	0	22138	4373	6

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 97.

The worst 5 of 4373 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:84:LYS:CB	1:A:139:TRP:CD1	1.77	1.65
2:C:202:TRP:CZ3	2:C:209:PHE:CE1	1.77	1.62
2:D:4:PRO:HB3	1:B:607:PHE:CE1	1.10	1.62
1:E:331:GLN:HG3	1:E:364:TRP:CZ2	1.21	1.62
1:A:607:PHE:CE1	2:C:4:PRO:CB	1.82	1.62

The worst 5 of 6 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:E:331:GLN:O	2:F:143:MET:CE[2_555]	1.76	0.44
2:F:49:SER:C	2:F:136:ARG:NH2[2_555]	1.94	0.26
2:F:49:SER:O	2:F:136:ARG:NH2[2_555]	1.99	0.21
1:E:486:GLU:O	2:F:81:SER:OG[2_555]	2.08	0.12

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:94:ARG:NH2	2:F:136:ARG:O[2_555]	2.08	0.12

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	525/632 (83%)	494 (94%)	29 (6%)	2 (0%)	34	69
1	B	522/632 (83%)	488 (94%)	31 (6%)	3 (1%)	25	62
1	E	521/632 (82%)	492 (94%)	28 (5%)	1 (0%)	47	78
2	C	422/502 (84%)	404 (96%)	18 (4%)	0	100	100
2	D	421/502 (84%)	399 (95%)	20 (5%)	2 (0%)	29	66
2	F	385/502 (77%)	372 (97%)	11 (3%)	2 (0%)	29	66
All	All	2796/3402 (82%)	2649 (95%)	137 (5%)	10 (0%)	34	69

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	416	PRO
2	D	131	TRP
1	B	416	PRO
1	E	416	PRO
2	F	501	VAL

5.3.2 Protein sidechains ⓘ

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	441/579 (76%)	303 (69%)	138 (31%)	0	1
1	B	430/579 (74%)	319 (74%)	111 (26%)	0	4
1	E	433/579 (75%)	313 (72%)	120 (28%)	0	2
2	C	392/461 (85%)	311 (79%)	81 (21%)	1	7
2	D	391/461 (85%)	320 (82%)	71 (18%)	1	11
2	F	364/461 (79%)	297 (82%)	67 (18%)	1	10
All	All	2451/3120 (79%)	1863 (76%)	588 (24%)	0	5

5 of 588 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	365	ILE
2	F	352	LYS
1	E	417	GLU
1	E	360	ASN
1	E	615	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 77 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	186	ASN
2	F	42	GLN
1	E	235	ASN
1	E	491	ASN
2	F	345	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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(Å²)	Q<0.9
1	A	529/632 (83%)	-0.20	13 (2%)	57	45	30, 74, 134, 150	0
1	B	528/632 (83%)	0.17	27 (5%)	28	21	28, 109, 157, 177	0
1	E	527/632 (83%)	-0.01	10 (1%)	66	55	49, 89, 133, 163	0
2	C	430/502 (85%)	0.11	39 (9%)	9	7	29, 72, 158, 200	0
2	D	429/502 (85%)	0.16	36 (8%)	11	8	45, 93, 152, 198	0
2	F	397/502 (79%)	0.61	59 (14%)	2	2	45, 133, 176, 198	0
All	All	2840/3402 (83%)	0.12	184 (6%)	18	12	28, 91, 158, 200	0

The worst 5 of 184 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	387	ASN	13.3
1	E	137	ALA	11.3
2	C	433	GLU	8.3
2	C	233	ASN	7.3
2	F	386	GLY	6.6

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.