wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2023 – 08:04 PM EDT

PDB ID	:	3UAI
Title	:	Structure of the Shq1-Cbf5-Nop10-Gar1 complex from Saccharomyces cerevisiae
Authors	:	Ye, K.
Deposited on	:	2011-10-21
Resolution	:	3.06 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

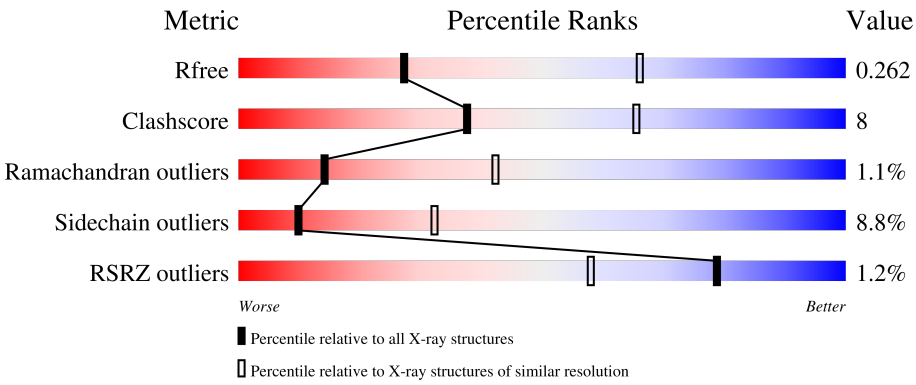
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.06 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	1754 (3.10-3.02)
Clashscore	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)
RSRZ outliers	127900	1713 (3.10-3.02)

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	400	.% 64% 20% 5% 11%
2	B	58	72% 9% • 17%
3	C	114	2% 61% 14% 5% 19%
4	D	366	.% 73% 15% • 11%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6627 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 H/ACA ribonucleoprotein complex subunit 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	357	2804	1765	504	517	18	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP P33322
A	2	ALA	-	expression tag	UNP P33322
A	395	HIS	-	expression tag	UNP P33322
A	396	HIS	-	expression tag	UNP P33322
A	397	HIS	-	expression tag	UNP P33322
A	398	HIS	-	expression tag	UNP P33322
A	399	HIS	-	expression tag	UNP P33322
A	400	HIS	-	expression tag	UNP P33322

- Molecule 2 is a protein called H/ACA ribonucleoprotein complex subunit 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	48	387	244	70	71	2	0	0	0

- Molecule 3 is a protein called H/ACA ribonucleoprotein complex subunit 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	92	721	470	113	134	4	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	11	MET	-	expression tag	UNP P28007
C	12	GLY	-	expression tag	UNP P28007

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	13	SER	-	expression tag	UNP P28007
C	14	SER	-	expression tag	UNP P28007
C	15	HIS	-	expression tag	UNP P28007
C	16	HIS	-	expression tag	UNP P28007
C	17	HIS	-	expression tag	UNP P28007
C	18	HIS	-	expression tag	UNP P28007
C	19	HIS	-	expression tag	UNP P28007
C	20	HIS	-	expression tag	UNP P28007
C	21	SER	-	expression tag	UNP P28007
C	22	SER	-	expression tag	UNP P28007
C	23	GLY	-	expression tag	UNP P28007
C	24	LEU	-	expression tag	UNP P28007
C	25	VAL	-	expression tag	UNP P28007
C	26	PRO	-	expression tag	UNP P28007
C	27	ARG	-	expression tag	UNP P28007
C	28	GLY	-	expression tag	UNP P28007
C	29	SER	-	expression tag	UNP P28007
C	30	HIS	-	expression tag	UNP P28007
C	31	MET	-	expression tag	UNP P28007

- Molecule 4 is a protein called Protein SHQ1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	327	2715	1734	440	532	9	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

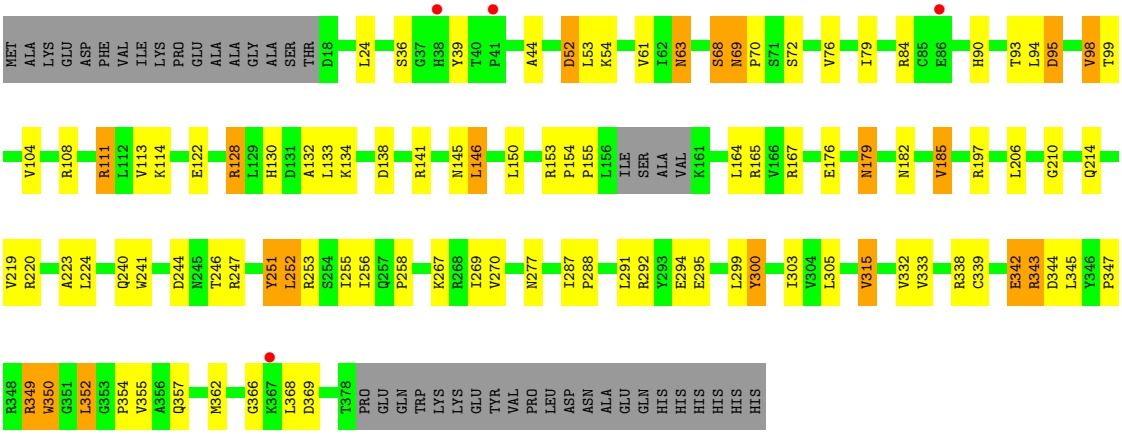
Chain	Residue	Modelled	Actual	Comment	Reference
D	142	GLY	-	expression tag	UNP P40486
D	143	PRO	-	expression tag	UNP P40486
D	144	GLU	-	expression tag	UNP P40486
D	145	ALA	-	expression tag	UNP P40486
D	146	SER	-	expression tag	UNP P40486

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.

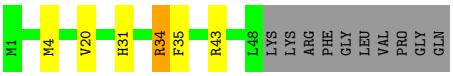
• Molecule 1: H/ACA ribonucleoprotein complex subunit 4

Chain A:



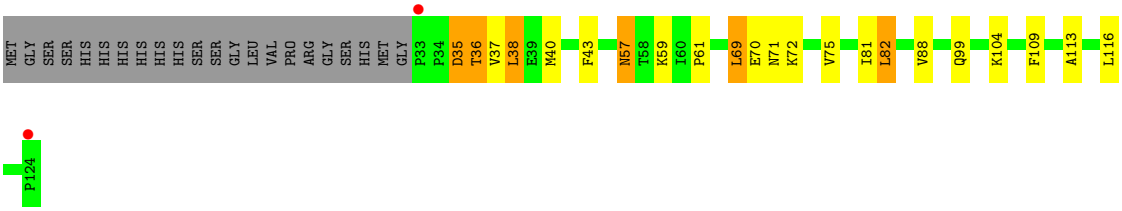
• Molecule 2: H/ACA ribonucleoprotein complex subunit 3

Chain B:



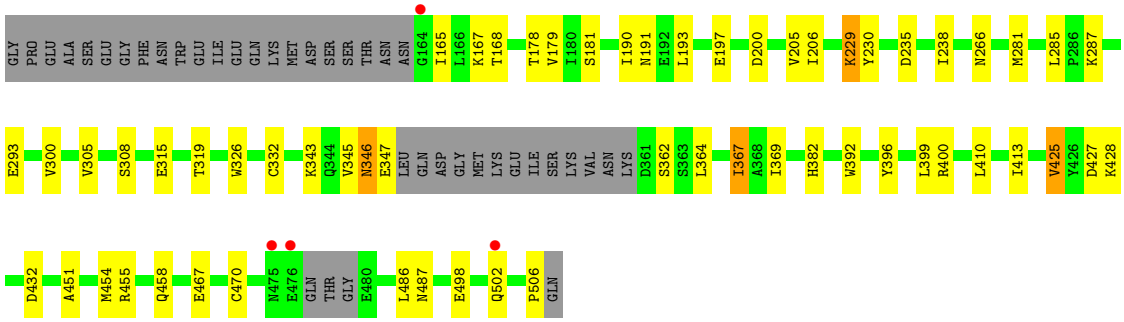
• Molecule 3: H/ACA ribonucleoprotein complex subunit 1

Chain C:



• Molecule 4: Protein SHQ1

Chain D:


GLY PRO GLU ALA SER GLU GLY PHE ASN TRP GLU ILE GLU GLN LYS MET ASP SER SER THR ASN ASN G164 I165 L166 K167 T168 T178 V179 I180 S181 I190 N191 E192 L193 E197 D200 V205 I206 K229 Y230 D235 I238 N266 M281 L285 P286 K287
E293 V300 V305 S308 E315 T319 W326 C332 K343 Q344 V345 N346 E347 LEU GLN ASP GLY MET LYS GLU ILE SER LYS VAL ASN LYS D361 S362 S363 L364 I367 A368 I369 H382 W392 Y396 L399 R400 L410 I413 V425 Y426 D427 K428
D432 A451 M454 R455 Q458 E467 C470 N475 E476 GLN THR GLY E480 L486 N487 E498 Q502 P506 GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	101.00Å 101.00Å 272.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.06 19.98 – 3.06	Depositor EDS
% Data completeness (in resolution range)	99.9 (20.00-3.06) 99.9 (19.98-3.06)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.35 (at 3.04Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.219 , 0.275 0.209 , 0.262	Depositor DCC
R_{free} test set	1370 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	92.5	Xtriage
Anisotropy	0.039	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 48.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6627	wwPDB-VP
Average B, all atoms (Å^2)	100.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.53% 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	0.46	0/2850	0.64	0/3847
2	B	0.45	0/395	0.60	0/530
3	C	0.48	0/739	0.58	1/1002 (0.1%)
4	D	0.46	0/2771	0.56	0/3749
All	All	0.46	0/6755	0.60	1/9128 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	82	LEU	CA-CB-CG	5.26	127.41	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	2804	0	2875	61	0
2	B	387	0	398	4	0
3	C	721	0	721	8	0
4	D	2715	0	2644	31	0
All	All	6627	0	6638	102	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.

The worst 5 of 102 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:111:ARG:HH21	1:A:111:ARG:HG3	1.49	0.78
1:A:111:ARG:HH21	1:A:111:ARG:CG	1.96	0.77
1:A:277:ASN:ND2	1:A:343:ARG:O	2.19	0.75
1:A:315:VAL:HG13	1:A:338:ARG:HB3	1.67	0.75
1:A:362:MET:O	1:A:366:GLY:O	2.10	0.70

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	353/400 (88%)	317 (90%)	28 (8%)	8 (2%)	6 24
2	B	46/58 (79%)	44 (96%)	2 (4%)	0	100 100
3	C	90/114 (79%)	74 (82%)	16 (18%)	0	100 100
4	D	321/366 (88%)	305 (95%)	15 (5%)	1 (0%)	41 70
All	All	810/938 (86%)	740 (91%)	61 (8%)	9 (1%)	14 42

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	69	ASN
1	A	342	GLU
1	A	252	LEU
1	A	352	LEU
1	A	84	ARG

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	309/345 (90%)	273 (88%)	36 (12%)	5	19
2	B	43/51 (84%)	41 (95%)	2 (5%)	26	56
3	C	80/98 (82%)	68 (85%)	12 (15%)	3	11
4	D	304/339 (90%)	289 (95%)	15 (5%)	25	55
All	All	736/833 (88%)	671 (91%)	65 (9%)	10	32

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	346	ASN
4	D	367	ILE
1	A	247	ARG
1	A	246	THR
4	D	425	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	377	ASN
4	D	458	GLN
3	C	57	ASN
4	D	504	GLN
4	D	316	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)

There are no ligands in this entry.

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)

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(Å^2)	Q<0.9
1	A	357/400 (89%)	-0.22	4 (1%) 80 60	68, 96, 146, 158	0
2	B	48/58 (82%)	-0.39	0 100 100	71, 91, 147, 152	0
3	C	92/114 (80%)	-0.13	2 (2%) 62 38	90, 114, 137, 147	0
4	D	327/366 (89%)	-0.35	4 (1%) 79 58	64, 91, 131, 173	0
All	All	824/938 (87%)	-0.27	10 (1%) 79 58	64, 96, 141, 173	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	164	GLY	5.7
4	D	475	ASN	3.8
1	A	38	HIS	3.6
3	C	124	PRO	3.2
1	A	367	LYS	2.9

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

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.