



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 07:52 PM GMT

PDB ID : 4W4U  
Title : Structure of yeast SAGA DUBm with Sgf73 Y57A mutant at 2.8 angstroms resolution  
Authors : Wolberger, C.; Yan, M.  
Deposited on : 2014-08-15  
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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20026982

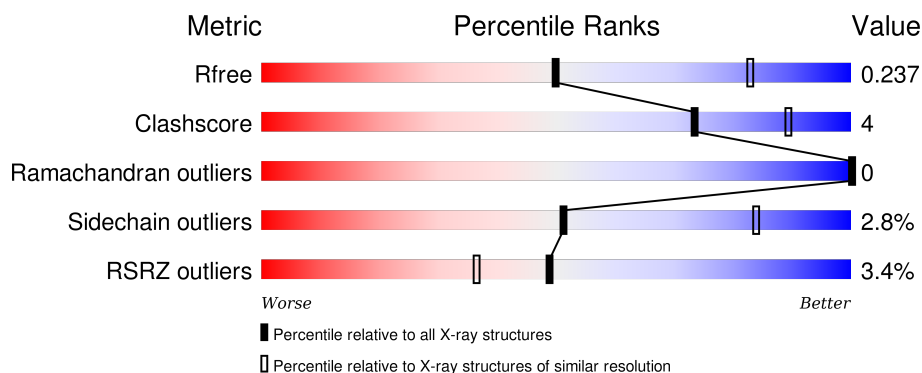
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . 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	476	<div> <div>4%</div> <div>80%</div> <div>11%</div> <div>8%</div> </div>
1	D	476	<div> <div>4%</div> <div>76%</div> <div>10%</div> <div>14%</div> </div>
2	B	96	<div> <div>%</div> <div>93%</div> <div>7%</div> </div>
2	F	96	<div> <div>2%</div> <div>84%</div> <div>14%</div> <div>.</div> </div>
3	C	99	<div> <div>%</div> <div>51%</div> <div>8%</div> <div>.</div> <div>40%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	99	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>43%</div><div>53%</div></div></div>
4	E	96	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>79%</div><div>8%</div><div>13%</div></div></div>
4	H	96	<div><div><div>2%</div><div><div></div><div></div><div></div></div><div>77%</div><div>8%</div><div>15%</div></div></div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10594 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 Ubiquitin carboxyl-terminal hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	438	Total	C	N	O	S	0	4	0
			3526	2245	598	647	36			
1	D	409	Total	C	N	O	S	0	2	0
			3294	2107	557	599	31			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP N1P0J5
A	-3	ALA	-	expression tag	UNP N1P0J5
A	-2	ALA	-	expression tag	UNP N1P0J5
A	-1	ALA	-	expression tag	UNP N1P0J5
A	0	ALA	-	expression tag	UNP N1P0J5
D	-4	GLY	-	expression tag	UNP N1P0J5
D	-3	ALA	-	expression tag	UNP N1P0J5
D	-2	ALA	-	expression tag	UNP N1P0J5
D	-1	ALA	-	expression tag	UNP N1P0J5
D	0	ALA	-	expression tag	UNP N1P0J5

- Molecule 2 is a protein called Transcription and mRNA export factor SUS1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	96	Total	C	N	O	S	0	1	0
			780	487	126	163	4			
2	F	94	Total	C	N	O	S	0	1	0
			765	478	124	160	3			

- Molecule 3 is a protein called SAGA-associated factor 11.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	59	Total	C	N	O	0	0	0
			471	296	80	95			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	47	Total	C	N	O	S	0	0	0
			374	231	64	78	1			

- Molecule 4 is a protein called SAGA-associated factor 73.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	84	Total	C	N	O	S	0	0	0
			669	422	114	129	4			
4	H	82	Total	C	N	O	S	0	0	0
			649	408	110	127	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	57	ALA	TYR	engineered mutation	UNP P53165
H	57	ALA	TYR	engineered mutation	UNP P53165

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	1	Total	Zn	0	0
			1	1		
5	A	6	Total	Zn	0	0
			6	6		
5	D	5	Total	Zn	0	0
			5	5		
5	E	1	Total	Zn	0	0
			1	1		

- Molecule 6 is water.

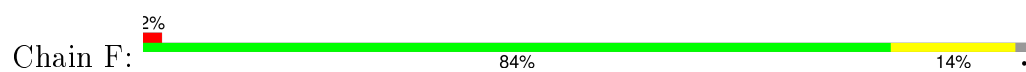
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	8	Total	O	0	0
			8	8		
6	B	5	Total	O	0	0
			5	5		
6	E	3	Total	O	0	0
			3	3		
6	D	25	Total	O	0	0
			25	25		
6	F	4	Total	O	0	0
			4	4		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	G	2	Total	O	0	0
			2	2		
6	H	6	Total	O	0	0
			6	6		

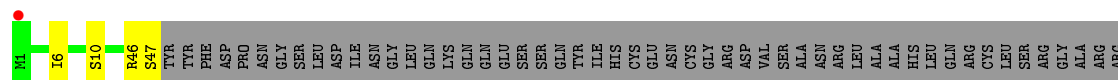
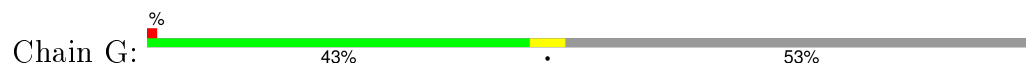




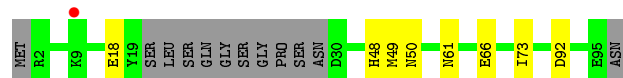
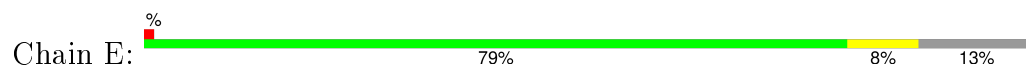
- Molecule 3: SAGA-associated factor 11



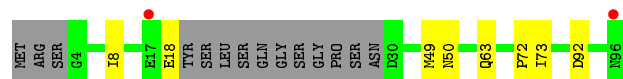
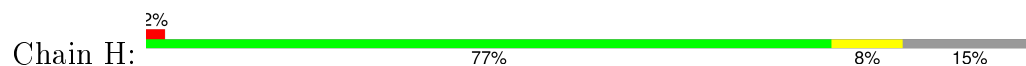
- Molecule 3: SAGA-associated factor 11



- Molecule 4: SAGA-associated factor 73



- Molecule 4: SAGA-associated factor 73





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	80.74Å 67.27Å 137.15Å 90.00° 106.84° 90.00°	Depositor
Resolution (Å)	47.02 – 2.80 46.98 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.02-2.80) 99.8 (46.98-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.39 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.182 , 0.241 0.184 , 0.237	Depositor DCC
$R_{free}$ test set	1755 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	64.8	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 55.2	EDS
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 34989 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	10594	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.76% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 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: ZN

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	0.52	0/3612	0.66	0/4868
1	D	0.57	0/3370	0.67	0/4536
2	B	0.56	0/789	0.71	0/1062
2	F	0.56	0/774	0.69	0/1042
3	C	0.49	0/476	0.64	0/647
3	G	0.53	0/375	0.68	0/509
4	E	0.49	0/681	0.68	0/916
4	H	0.56	0/660	0.68	0/887
All	All	0.54	0/10737	0.67	0/14467

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 [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	3526	0	3459	30	1
1	D	3294	0	3251	33	0
2	B	780	0	792	5	0
2	F	765	0	773	12	1
3	C	471	0	478	7	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	374	0	390	5	0
4	E	669	0	671	7	0
4	H	649	0	650	9	0
5	A	6	0	0	0	0
5	D	5	0	0	0	0
5	E	1	0	0	0	0
5	H	1	0	0	0	0
6	A	8	0	0	1	0
6	B	5	0	0	1	0
6	D	25	0	0	0	0
6	E	3	0	0	1	0
6	F	4	0	0	1	0
6	G	2	0	0	1	0
6	H	6	0	0	0	0
All	All	10594	0	10464	86	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 4.

The worst 5 of 86 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:104:GLY:HA3	3:C:33:THR:HG23	1.73	0.71
4:E:66:GLU:HA	6:E:203:HOH:O	1.89	0.70
1:D:7[B]:ILE:HG23	1:D:85:PHE:CE1	2.30	0.67
1:D:193:GLU:OE2	1:D:215:GLY:N	2.28	0.67
2:F:5:THR:HG22	3:G:6:ILE:HG12	1.76	0.66

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:76:SER:OG	2:F:4:ASP:OD2[2_556]	2.13	0.07

## 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	434/476 (91%)	406 (94%)	28 (6%)	0	100	100
1	D	397/476 (83%)	380 (96%)	17 (4%)	0	100	100
2	B	95/96 (99%)	95 (100%)	0	0	100	100
2	F	93/96 (97%)	93 (100%)	0	0	100	100
3	C	57/99 (58%)	55 (96%)	2 (4%)	0	100	100
3	G	45/99 (46%)	44 (98%)	1 (2%)	0	100	100
4	E	80/96 (83%)	76 (95%)	4 (5%)	0	100	100
4	H	78/96 (81%)	76 (97%)	2 (3%)	0	100	100
All	All	1279/1534 (83%)	1225 (96%)	54 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	403/429 (94%)	390 (97%)	13 (3%)	46	80
1	D	376/429 (88%)	368 (98%)	8 (2%)	61	90
2	B	92/91 (101%)	90 (98%)	2 (2%)	60	89
2	F	90/91 (99%)	87 (97%)	3 (3%)	45	79
3	C	55/89 (62%)	51 (93%)	4 (7%)	17	44
3	G	45/89 (51%)	44 (98%)	1 (2%)	60	89

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	E	75/85 (88%)	73 (97%)	2 (3%)	52	85
4	H	73/85 (86%)	71 (97%)	2 (3%)	52	85
All	All	1209/1388 (87%)	1174 (97%)	35 (3%)	51	83

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	10	SER
4	E	73	ILE
3	G	10	SER
3	C	47	SER
3	C	55	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	427	HIS
3	G	34	GLN
1	D	452	GLN
1	D	268	ASN
1	D	438	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](#)

Of 13 ligands modelled in this entry, 13 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

## 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	438/476 (92%)	0.20	20 (4%) 36 25	47, 80, 121, 149	0
1	D	409/476 (85%)	0.03	17 (4%) 40 28	44, 69, 115, 151	1 (0%)
2	B	96/96 (100%)	-0.04	1 (1%) 84 77	54, 70, 90, 113	0
2	F	94/96 (97%)	-0.11	2 (2%) 67 56	52, 67, 91, 110	0
3	C	59/99 (59%)	0.13	1 (1%) 73 63	49, 75, 126, 132	0
3	G	47/99 (47%)	-0.25	1 (2%) 67 56	46, 63, 91, 92	0
4	E	84/96 (87%)	0.18	1 (1%) 81 73	64, 83, 111, 121	0
4	H	82/96 (85%)	-0.08	2 (2%) 62 50	51, 70, 94, 112	0
All	All	1309/1534 (85%)	0.07	45 (3%) 49 36	44, 74, 115, 151	1 (0%)

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	224	TRP	6.6
1	A	216	PHE	5.0
1	D	198	LEU	4.6
1	A	218	TYR	4.6
1	D	216	PHE	4.5

### 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 carbohydrates in this entry.

## 6.4 Ligands [i](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	ZN	A	504	1/1	1.00	0.18	0.88	65,65,65,65	0
5	ZN	A	501	1/1	0.99	0.15	0.35	73,73,73,73	0
5	ZN	D	505	1/1	1.00	0.16	0.25	70,70,70,70	0
5	ZN	A	502	1/1	1.00	0.18	0.22	55,55,55,55	0
5	ZN	D	504	1/1	1.00	0.15	0.21	64,64,64,64	0
5	ZN	A	503	1/1	0.99	0.15	-0.07	76,76,76,76	0
5	ZN	D	503	1/1	1.00	0.14	-0.45	58,58,58,58	0
5	ZN	A	505	1/1	0.99	0.13	-0.85	90,90,90,90	0
5	ZN	D	501	1/1	0.99	0.10	-0.95	75,75,75,75	0
5	ZN	D	502	1/1	0.99	0.11	-1.01	80,80,80,80	0
5	ZN	A	506	1/1	0.98	0.05	-1.88	115,115,115,115	0
5	ZN	H	101	1/1	0.99	0.12	-	73,73,73,73	0
5	ZN	E	101	1/1	0.98	0.13	-	102,102,102,102	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.