



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:30 AM GMT

PDB ID : 3M99
Title : Structure of the Ubp8-Sgf11-Sgf73-Sus1 SAGA DUB module
Authors : Kohler, A.; Zimmerman, E.; Schneider, M.; Hurt, E.; Zheng, N.
Deposited on : 2010-03-21
Resolution : 2.70 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
<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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

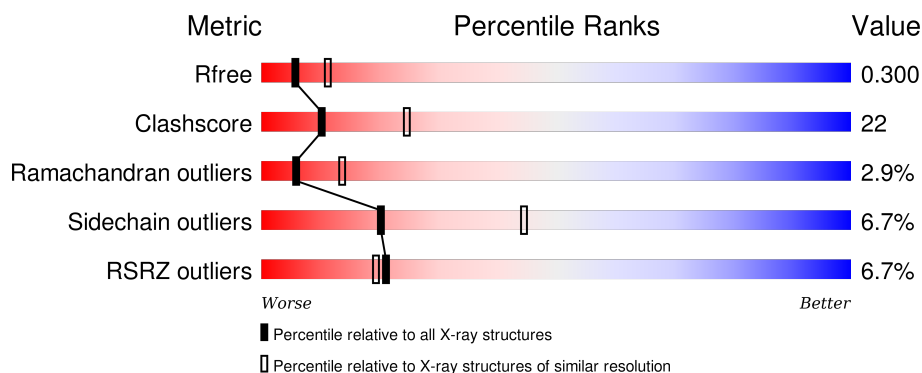
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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. 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	471	<div> <div>7%</div> <div> <div></div> <div>53%</div> <div>31%</div> <div>5%</div> <div>10%</div> </div> </div>
2	B	99	<div> <div>3%</div> <div> <div></div> <div>55%</div> <div>32%</div> <div>•</div> <div>10%</div> </div> </div>
3	C	96	<div> <div>4%</div> <div> <div></div> <div>55%</div> <div>35%</div> <div>•</div> <div>5%</div> </div> </div>
4	D	104	<div> <div>5%</div> <div> <div></div> <div>42%</div> <div>35%</div> <div>•</div> <div>19%</div> </div> </div>

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 5557 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 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	422	Total	C	N	O	S	0	0	0
			3382	2154	575	620	33			

- Molecule 2 is a protein called SAGA-associated factor 11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	89	Total	C	N	O	S	0	0	0
			710	434	132	141	3			

- Molecule 3 is a protein called Protein SUS1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	91	Total	C	N	O	S	0	0	0
			737	460	121	154	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	84	Total	C	N	O	S	0	0	0
			680	432	114	129	5			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Zn	0	0
			1	1		
5	A	5	Total	Zn	0	0
			5	5		
5	D	1	Total	Zn	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	32	Total 32	O 32	0	0
6	B	2	Total 2	O 2	0	0
6	C	2	Total 2	O 2	0	0
6	D	5	Total 5	O 5	0	0

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

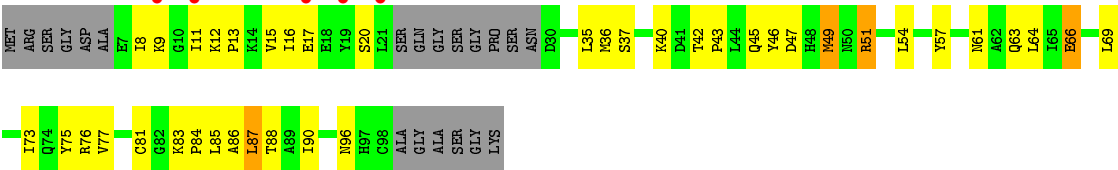
- Chain A:
-
- The visualization for Chain A includes a sequence of amino acids arranged in a grid, a horizontal bar chart at the top, and a network graph on the right. The bar chart shows five segments: 7% (red), 53% (green), 31% (yellow), 5% (orange), and 10% (grey). The network graph consists of nodes (amino acids) connected by lines, with some nodes highlighted in red.
- Sequence of amino acids (from top to bottom):
- Met, I3, C4, P5, H6, I7, Q8, Q9, V10, F11, Q12, M13, E14, K15, L21, C24, R28, L31, K37, E38, L41, M42, T43, M44, T48, C49, H50, E51, I52, N53, G55, A56, F58, M59, C60, L61, C62, C63, C66, C68, H71, H73, S78, T83.
- Horizontal bar chart segments:
- 7% (Red)
 - 53% (Green)
 - 31% (Yellow)
 - 5% (Orange)
 - 10% (Grey)
- Network graph nodes (from top to bottom):
- ARG, SER, PRO, D180, I387, C182, F183, A186, L187, D188, K189, I190, H191, H192, E193, L194, A197, L198, ASN, THR, LYS, GLN, A115, K116, D119, V121, T125, M126, S129, M130, E131, R132, S137, I140, M148, I151, I156, H157, M158, I162, R163, H164, S165, M166, N173, C174, K175, V176, S209, T210, R211, R212, Q213, L219, L226, N227, GLN, ASN, LEU, ALA, GLY, THR, VLY, Q235, Q236, D237, E240, F241, F244, L245, L246, N247, C248, P250, G251, C252, Q253, L254, R255, S256, T257, L258, K259, C260, L261, G262, D263, E264, A265, S266, H267, L268, K269, I270, H271, C272, C273, L274, V275, H276, T277, S282, S285, S286, L287, V288, C289, P290, G291, C292, Q293, L294, R295, S296, C297, L298, T299, L300, R301, P302, F303, S307, I310, K311, D312, K313, R314, S322, K325, L329, R330, D331, P332, I333, TYR, HIS, CYS, GLY, GLU, CYS, ASN, SER, THR, GLN, ASP, A345, I346, V357, L362, R363, R364, G371, F365, E366, H367, L368, N370, G371, S372, N373, R374, K375, I380, E381, F382, P383, T384, V385, L386, N387, M388, Y391, C392, S393, THR, LYS, GLU, LYS, ASP, LYS, HIS, SER, GLU, ASN, GLY, V406, I409.

- Chain B:
-
- 3% 55% 32% 10%
- MET THR GLU THR THR ILE I7 I8 D9 S10 I11 I15 L16 N17 N18 L19 L20 T21 T22 L23 I24 Q25 D26 I27 R30 Q34 Q35 Q36 R46 Y49 F50 D51 P52 N53 G54 S55 L56 D57 I58 L61 E66 S67 S68 Q69 Y70 C73 E74 N75 C76 G77 R78 D79
- R84 H88 R91 C92 L93 S94 R95 GLY ALA ARG ARG

- Chain C:  4% 55% 35% 5%



● 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, α , β , γ	64.42Å 103.52Å 70.35Å 90.00° 108.09° 90.00°	Depositor
Resolution (Å)	48.12 – 2.70 48.12 – 2.49	Depositor EDS
% Data completeness (in resolution range)	88.8 (48.12-2.70) 75.4 (48.12-2.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.48 (at 2.48Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, R_{free}	0.234 , 0.300 0.234 , 0.300	Depositor DCC
R_{free} test set	1102 reflections (5.35%)	DCC
Wilson B-factor (Å ²)	45.5	Xtriage
Anisotropy	0.818	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 34.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 24441 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5557	wwPDB-VP
Average B, all atoms (Å ²)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.69% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.33	0/3453	0.51	1/4647 (0.0%)
2	B	0.28	0/718	0.78	3/971 (0.3%)
3	C	0.30	0/743	0.45	0/1000
4	D	0.31	0/694	0.54	0/935
All	All	0.31	0/5608	0.55	4/7553 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	66	GLU	CB-CA-C	-15.39	79.61	110.40
2	B	67	SER	N-CA-CB	-7.34	99.50	110.50
2	B	67	SER	N-CA-C	-7.03	92.01	111.00
1	A	406	VAL	CB-CA-C	-5.10	101.70	111.40

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	3382	0	3312	170	0
2	B	710	0	698	30	0
3	C	737	0	742	30	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	680	0	678	48	0
5	A	5	0	0	1	0
5	B	1	0	0	1	0
5	D	1	0	0	0	0
6	A	32	0	0	4	0
6	B	2	0	0	0	0
6	C	2	0	0	3	0
6	D	5	0	0	4	0
All	All	5557	0	5430	238	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 238 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)
2:B:66:GLU:O	2:B:66:GLU:CG	1.80	1.19
4:D:37:SER:HA	6:D:110:HOH:O	1.51	1.09
1:A:364:ARG:HH22	1:A:459:GLN:HG3	1.23	1.03
2:B:66:GLU:O	2:B:66:GLU:HG3	1.21	1.00
1:A:60:CYS:CB	1:A:63:CYS:SG	2.51	0.99

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	408/471 (87%)	337 (83%)	57 (14%)	14 (3%)	5	10
2	B	87/99 (88%)	75 (86%)	11 (13%)	1 (1%)	17	42
3	C	89/96 (93%)	83 (93%)	4 (4%)	2 (2%)	8	22

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	D	80/104 (77%)	68 (85%)	10 (12%)	2 (2%)	7	18
All	All	664/770 (86%)	563 (85%)	82 (12%)	19 (3%)	6	14

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	54	SER
1	A	105	ASN
1	A	130	MET
1	A	212	ARG
1	A	443	ASN

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	386/429 (90%)	363 (94%)	23 (6%)	24	50
2	B	81/89 (91%)	78 (96%)	3 (4%)	41	72
3	C	86/91 (94%)	77 (90%)	9 (10%)	8	19
4	D	78/90 (87%)	71 (91%)	7 (9%)	12	27
All	All	631/699 (90%)	589 (93%)	42 (7%)	20	44

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	364	ARG
2	B	9	ASP
4	D	49	MET
1	A	375	LYS
1	A	438	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	239	HIS
1	A	251	GLN
3	C	13	GLN
1	A	91	ASN
1	A	227	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 7 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 [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	422/471 (89%)	0.31	34 (8%) 15 12	38, 67, 126, 194	0
2	B	89/99 (89%)	0.24	3 (3%) 49 49	51, 75, 111, 123	0
3	C	91/96 (94%)	0.22	4 (4%) 38 37	54, 76, 123, 129	0
4	D	84/104 (80%)	0.38	5 (5%) 25 24	46, 60, 133, 140	0
All	All	686/770 (89%)	0.30	46 (6%) 21 19	38, 69, 128, 194	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	21	LEU	6.5
1	A	368	LEU	5.2
4	D	19	TYR	5.2
1	A	293	GLN	4.6
1	A	267	ASN	4.6

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 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, 95th 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(\AA^2)	Q<0.9
5	ZN	D	105	1/1	0.88	0.09	-0.91	104,104,104,104	0
5	ZN	A	476	1/1	0.90	0.10	-1.34	81,81,81,81	0
5	ZN	A	473	1/1	0.88	0.09	-1.69	89,89,89,89	0
5	ZN	A	475	1/1	0.73	0.09	-1.88	107,107,107,107	0
5	ZN	B	100	1/1	0.78	0.10	-2.09	146,146,146,146	0
5	ZN	A	472	1/1	0.82	0.07	-2.17	126,126,126,126	0
5	ZN	A	474	1/1	0.72	0.10	-2.82	93,93,93,93	0

6.5 Other polymers [i](#)

There are no such residues in this entry.