



Full wwPDB NMR Structure Validation Report ⓘ

May 31, 2020 – 11:35 am BST

PDB ID : 6O22
Title : Structure of Asf1-H3:H4-Rtt109-Vps75 histone chaperone-lysine acetyltransferase complex with the histone substrate.
Authors : Danilenko, N.; Carlomagno, T.; Kirkpatrick, J.P.
Deposited on : 2019-02-22

This is a Full wwPDB NMR 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/NMRValidationReportHelp>

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:

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : 2.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.11

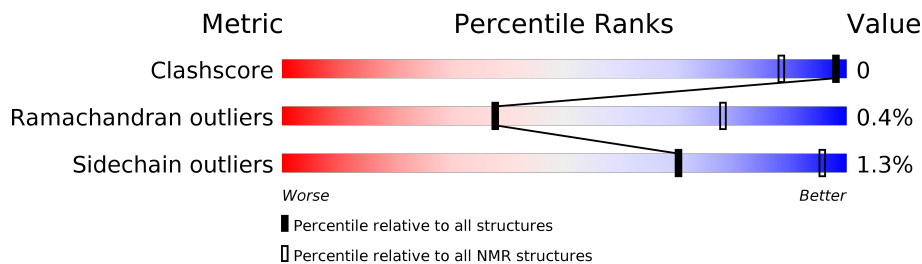
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR, SOLUTION SCATTERING

The overall completeness of chemical shifts assignment is 1%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	264	68% 17% . 13%
1	B	264	63% 17% . 18%
2	C	442	74% 20% . .
3	D	279	46% 11% . 41%
4	E	136	44% 9% . 45%
5	F	103	58% 17% . . 20%

2 Ensemble composition and analysis

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 19411 atoms, of which 9680 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Vacuolar protein sorting-associated protein 75.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	230	3733	1221	1825	309	372	6	0
1	B	217	3562	1168	1752	294	343	5	0

- Molecule 2 is a protein called Histone acetyltransferase RTT109.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	C	424	6917	2205	3480	585	637	10	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-5	GLY	-	expression tag	UNP Q07794
C	-4	MET	-	expression tag	UNP Q07794
C	-3	ASP	-	expression tag	UNP Q07794
C	-2	PRO	-	expression tag	UNP Q07794
C	-1	ASN	-	expression tag	UNP Q07794
C	0	SER	-	expression tag	UNP Q07794

- Molecule 3 is a protein called Histone chaperone ASF1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	D	164	2592	840	1280	212	258	2	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	SER	-	expression tag	UNP P32447

- Molecule 4 is a protein called Histone H3.2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
4	E	75	1243	384	636	114	106	3	0

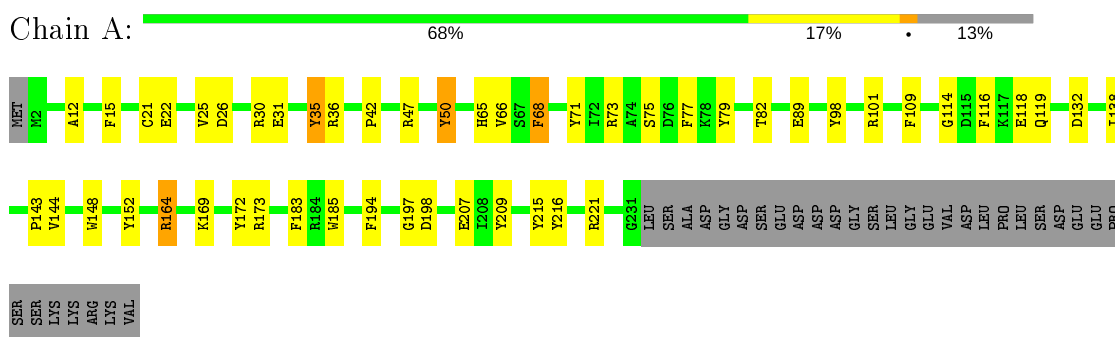
- Molecule 5 is a protein called Histone H4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
5	F	82	1364	416	707	128	112	1	0

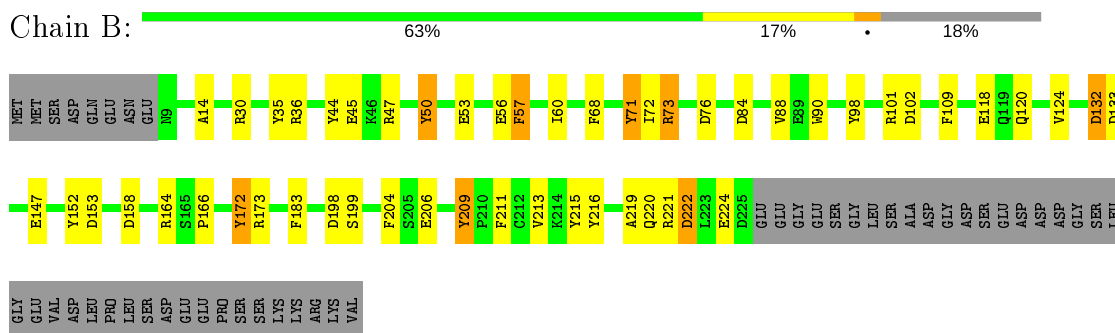
4 Residue-property plots [i](#)

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

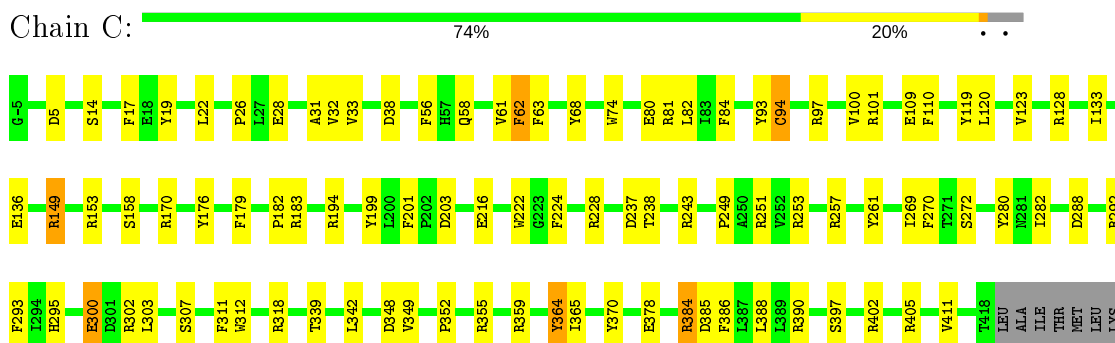
- Molecule 1: Vacuolar protein sorting-associated protein 75



- Molecule 1: Vacuolar protein sorting-associated protein 75



- Molecule 2: Histone acetyltransferase RTT109



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 150 calculated structures, 1 were deposited, based on the following criterion: *structures with the least restraint violations*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
HADDOCK	structure calculation	
Amber	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	368
Number of shifts mapped to atoms	368
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	1%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	1908	1825	1824	2
2	C	3437	3480	3480	2
4	E	607	636	634	1
All	All	9731	9680	9675	5

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 0.

All clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
2:C:365:ILE:N	2:C:365:ILE:HD12	0.52	2.19
1:A:164:ARG:O	1:A:169:LYS:HE3	0.44	2.12
4:E:71:VAL:HG12	4:E:84:PHE:CZ	0.43	2.49
1:A:116:PHE:CE1	1:A:119:GLN:NE2	0.40	2.88
2:C:293:PHE:CD1	2:C:311:PHE:CE1	0.40	3.10

5.2 Torsion angles [i](#)

5.2.1 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 PDB entries followed by that with respect to all NMR entries. 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	228/264 (86%)	209 (92%)	19 (8%)	0 (0%)	100	100
1	B	215/264 (81%)	203 (94%)	11 (5%)	1 (0%)	32	76
2	C	422/442 (95%)	389 (92%)	31 (7%)	2 (0%)	32	76
3	D	162/279 (58%)	155 (96%)	6 (4%)	1 (1%)	29	74
4	E	73/136 (54%)	67 (92%)	5 (7%)	1 (1%)	15	61
5	F	80/103 (78%)	78 (98%)	2 (2%)	0 (0%)	100	100
All	All	1180/1488 (79%)	1101 (93%)	74 (6%)	5 (0%)	38	78

All 5 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
4	E	133	GLU
2	C	14	SER
3	D	90	VAL
1	B	132	ASP
2	C	203	ASP

5.2.2 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 PDB entries followed by that with respect to all NMR

entries. 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	205/235 (87%)	204 (100%)	1 (0%)	89	97
1	B	194/235 (83%)	191 (98%)	3 (2%)	66	95
2	C	388/403 (96%)	383 (99%)	5 (1%)	70	96
3	D	151/252 (60%)	148 (98%)	3 (2%)	57	93
4	E	64/111 (58%)	64 (100%)	0 (0%)	100	100
5	F	68/79 (86%)	66 (97%)	2 (3%)	45	89
All	All	1070/1315 (81%)	1056 (99%)	14 (1%)	70	96

All 14 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	B	222	ASP
2	C	300	GLU
1	B	220	GLN
2	C	62	PHE
2	C	237	ASP
3	D	61	LEU
3	D	49	ARG
2	C	22	LEU
1	A	68	PHE
5	F	63	GLU
1	B	56	GLU
2	C	94	CYS
5	F	45	ARG
3	D	51	LEU

5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.5 Ligand geometry [i](#)

There are no ligands in this entry.

5.6 Other polymers [i](#)

There are no such molecules in this entry.

5.7 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Chemical shift validation [i](#)

The completeness of assignment taking into account all chemical shift lists is 1% for the well-defined parts and 1% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list*

6.1.1 Bookkeeping [i](#)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	368
Number of shifts mapped to atoms	368
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

6.1.2 Chemical shift referencing [i](#)

No chemical shift referencing corrections were calculated (not enough data).

6.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 1%, i.e. 184 atoms were assigned a chemical shift out of a possible 15339. 20 out of 181 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	0/5846 (0%)	0/2327 (0%)	0/2384 (0%)	0/1135 (0%)
Sidechain	184/8165 (2%)	92/4804 (2%)	92/2970 (3%)	0/391 (0%)
Aromatic	0/1328 (0%)	0/707 (0%)	0/583 (0%)	0/38 (0%)
Overall	184/15339 (1%)	92/7838 (1%)	92/5937 (2%)	0/1564 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 1%, i.e. 184 atoms were assigned a chemical shift out of a possible 15339. 20 out of 181 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	0/5846 (0%)	0/2327 (0%)	0/2384 (0%)	0/1135 (0%)
Sidechain	184/8165 (2%)	92/4804 (2%)	92/2970 (3%)	0/391 (0%)
Aromatic	0/1328 (0%)	0/707 (0%)	0/583 (0%)	0/38 (0%)
Overall	184/15339 (1%)	92/7838 (1%)	92/5937 (2%)	0/1564 (0%)

6.1.4 Statistically unusual chemical shifts [i](#)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	200	LEU	HD23	-0.69	2.14 – -0.66	-5.1
1	A	200	LEU	HD22	-0.69	2.14 – -0.66	-5.1
1	A	200	LEU	HD21	-0.69	2.14 – -0.66	-5.1
1	B	200	LEU	HD23	-0.68	2.14 – -0.66	-5.1
1	B	200	LEU	HD21	-0.68	2.14 – -0.66	-5.1
1	B	200	LEU	HD22	-0.68	2.14 – -0.66	-5.1

6.1.5 Random Coil Index (RCI) plots [i](#)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list). RCI is only applicable to proteins.