Full wwPDB NMR Structure Validation Report (i)

Mar 15, 2026 – 10:03 AM UTC

PDB ID	:	7Y0I / pdb_00007y0i
BMRB ID	:	36492
Title	:	Solution structures of ASH1L PHD domain in complex with H3K4me2 peptide
Authors	:	Yu, M.; Zeng, L.
Deposited on	:	2022-06-05

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 (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 (i)) were used in the production of this report:

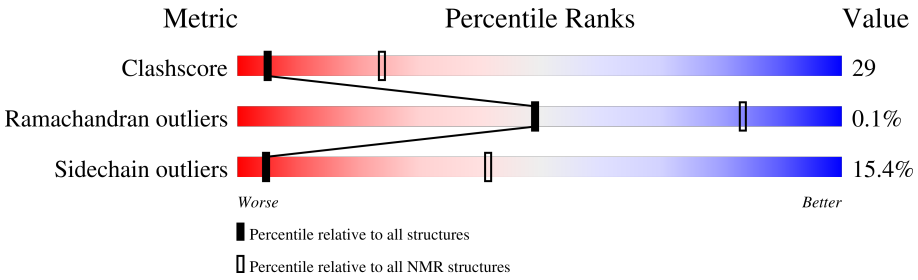
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 82%.

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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	56	39% 38% 7% 16%
2	B	15	13% 13% 73%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:6-A:52, B:301-B:303, B:305-B:305 (51)	0.30	3

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 6, 7, 8, 9, 11, 15, 16, 18, 19, 20
2	5, 10, 12, 17
3	13, 14

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1090 atoms, of which 536 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone-lysine N-methyltransferase ASH1L.

Mol	Chain	Residues	Atoms						Trace
1	A	56	Total	C	H	N	O	S	0
			852	266	411	79	85	11	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP Q9NR48
A	2	GLY	-	expression tag	UNP Q9NR48
A	3	GLY	-	expression tag	UNP Q9NR48
A	4	ARG	-	expression tag	UNP Q9NR48
A	39	THR	SER	conflict	UNP Q9NR48

- Molecule 2 is a protein called ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-ALA.

Mol	Chain	Residues	Atoms					Trace
2	B	15	Total	C	H	N	O	0
			236	65	125	25	21	

- Molecule 3 is ZINC ION (CCD ID: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	
3	A	2	Total	Zn
			2	2

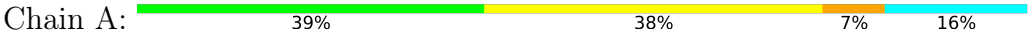
4 Residue-property plots

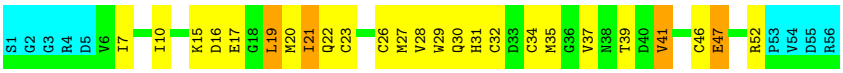
4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Histone-lysine N-methyltransferase ASH1L

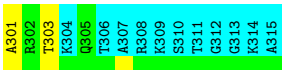
Chain A:





• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-ALA

Chain B:



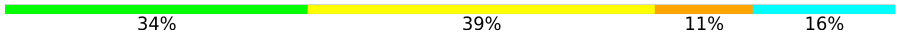
4.2 Scores per residue for each member of the ensemble

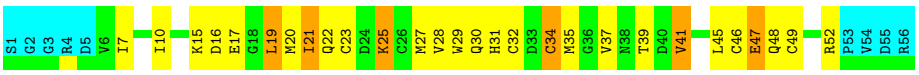
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

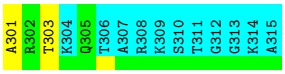
Chain A:





• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-ALA

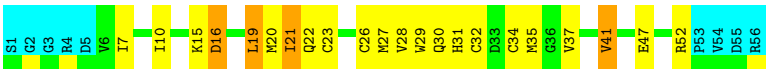
Chain B:



4.2.2 Score per residue for model 2

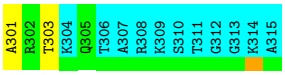
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

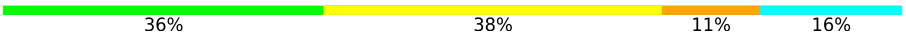
Chain B:



4.2.3 Score per residue for model 3 (medoid)

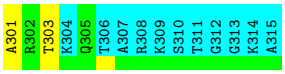
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

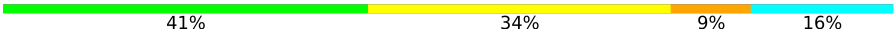
Chain B:

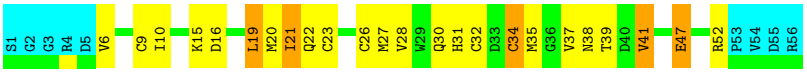


4.2.4 Score per residue for model 4

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:





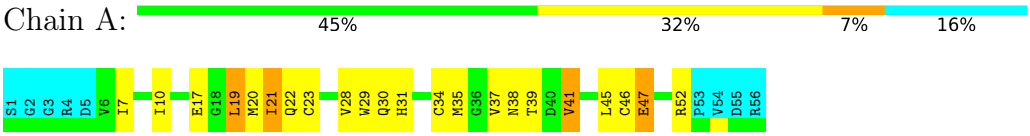
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

Chain B:

4.2.5 Score per residue for model 5

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



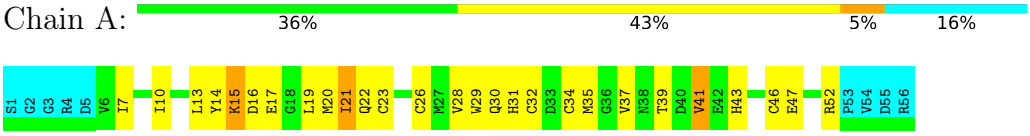
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

Chain B:

4.2.6 Score per residue for model 6

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



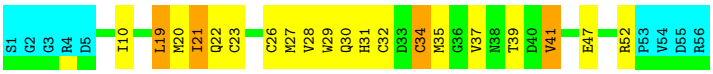
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

Chain B:

4.2.7 Score per residue for model 7

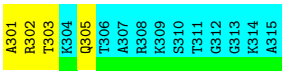
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



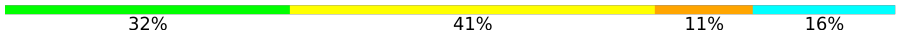
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

Chain B:



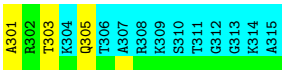
4.2.8 Score per residue for model 8

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A: 

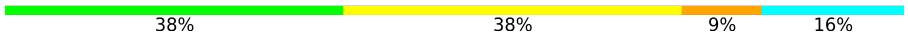
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

Chain B:



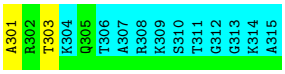
4.2.9 Score per residue for model 9

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A: 

• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

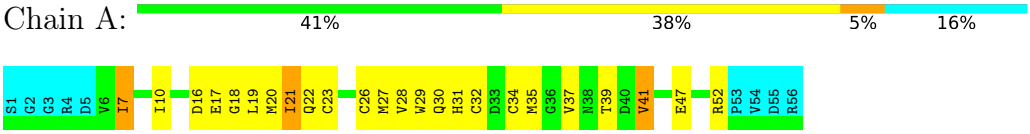
Chain B:



4.2.10 Score per residue for model 10

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



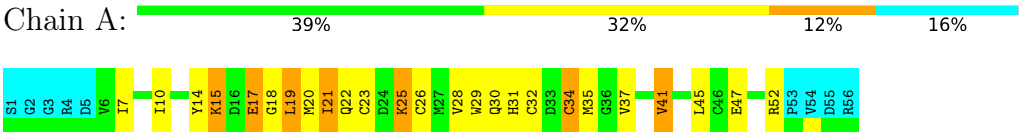
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-ALA

Chain B:

4.2.11 Score per residue for model 11

• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



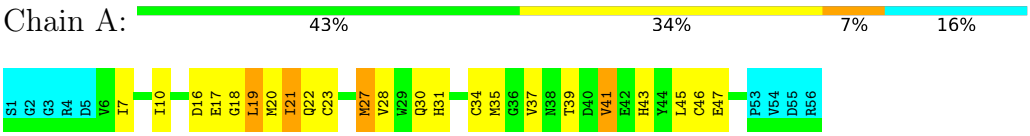
• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-ALA

Chain B:

4.2.12 Score per residue for model 12

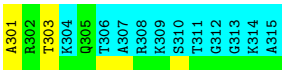
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-ALA

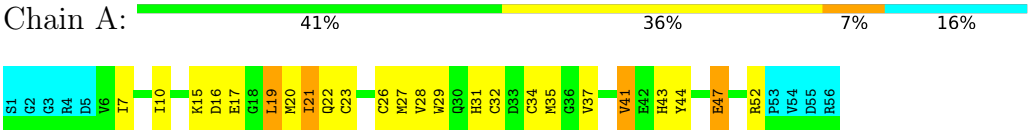
Chain B:



4.2.13 Score per residue for model 13

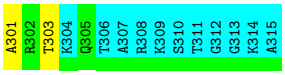
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

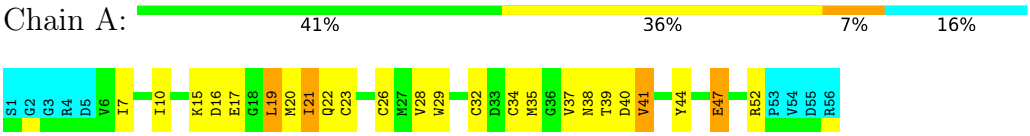
Chain B:



4.2.14 Score per residue for model 14

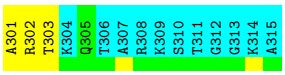
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

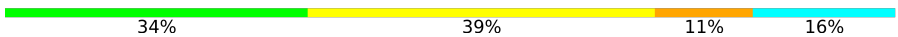
Chain B:



4.2.15 Score per residue for model 15

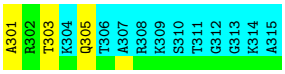
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

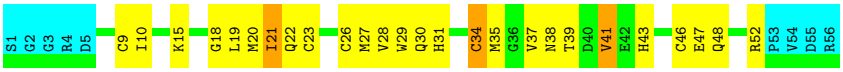
Chain B: 7% 20% 73%



4.2.16 Score per residue for model 16

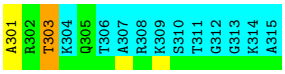
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A: 38% 41% 5% 16%



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

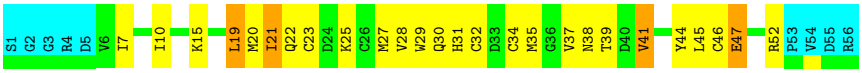
Chain B: 13% 7% 7% 73%



4.2.17 Score per residue for model 17

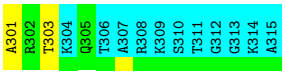
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A: 38% 39% 7% 16%



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

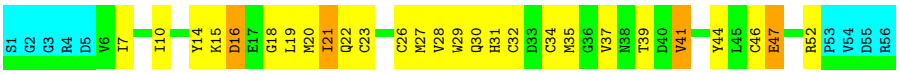
Chain B: 13% 13% 73%



4.2.18 Score per residue for model 18

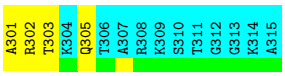
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A: 36% 41% 7% 16%



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

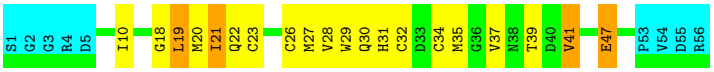
Chain B:



4.2.19 Score per residue for model 19

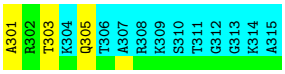
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

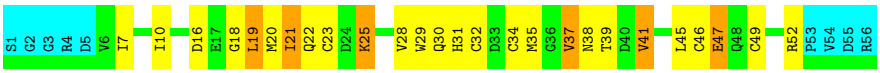
Chain B:



4.2.20 Score per residue for model 20

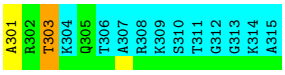
• Molecule 1: Histone-lysine N-methyltransferase ASH1L

Chain A:



• Molecule 2: ALA-ARG-THR-MLY-GLN-THR-ALA-ARG-LYS-SER-THR-GLY-GLY-LYS-AL
A

Chain B:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
ARIA	refinement	
CNS	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	685
Number of shifts mapped to atoms	685
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MLY, 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 (average) 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	0.51±0.02	0±0/381 (0.0± 0.0%)	0.85±0.02	0±0/515 (0.0± 0.0%)
2	B	0.36±0.02	0±0/31 (0.0± 0.0%)	0.40±0.06	0±0/40 (0.0± 0.0%)
All	All	0.50	0/8240 (0.0%)	0.83	1/11100 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	7	ILE	CB-CA-C	-5.65	106.24	112.68	10	1

There are no chirality outliers.

There are no planarity outliers.

6.2 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 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	374	348	347	23±3
2	B	32	35	32	4±1
All	All	8160	7660	7580	463

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 29.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:41:VAL:HG22	2:B:301:ALA:HB2	1.01	1.33	19	20
1:A:35:MET:HE1	1:A:47:GLU:HG3	0.85	1.48	13	3
1:A:7:ILE:HD12	1:A:20:MET:HE1	0.82	1.51	18	15
1:A:22:GLN:HE21	1:A:27:MET:HG3	0.80	1.35	15	8
1:A:31:HIS:HB2	1:A:34:CYS:SG	0.79	2.18	16	19
1:A:26:CYS:HB3	1:A:28:VAL:HG13	0.78	1.54	4	13
1:A:22:GLN:HE21	1:A:27:MET:HG2	0.77	1.38	8	5
1:A:20:MET:HE2	1:A:29:TRP:HB3	0.76	1.58	18	17
1:A:10:ILE:HG23	1:A:35:MET:HE1	0.74	1.59	11	17
1:A:25:LYS:HD3	1:A:45:LEU:HD21	0.72	1.61	20	4
1:A:39:THR:HB	2:B:303:THR:HG21	0.72	1.62	19	17
1:A:35:MET:HG2	1:A:44:TYR:OH	0.70	1.86	13	3
1:A:16:ASP:HA	1:A:20:MET:SD	0.69	2.28	2	11
1:A:8:ARG:HB2	1:A:28:VAL:HG21	0.68	1.65	15	1
1:A:19:LEU:HD21	2:B:303:THR:HB	0.66	1.65	5	15
1:A:21:ILE:HA	2:B:302:ARG:O	0.65	1.92	18	5
1:A:21:ILE:HD13	1:A:37:VAL:HG21	0.63	1.70	2	20
1:A:22:GLN:HA	1:A:28:VAL:O	0.63	1.93	2	19
1:A:21:ILE:H	1:A:21:ILE:HD12	0.59	1.58	14	16
1:A:23:CYS:HB3	1:A:26:CYS:HB2	0.57	1.74	19	11
1:A:31:HIS:O	1:A:35:MET:HG2	0.57	1.99	3	14
1:A:35:MET:HA	1:A:52:ARG:HH21	0.57	1.60	11	1
1:A:19:LEU:HD22	2:B:303:THR:HB	0.56	1.78	18	1
1:A:21:ILE:HD11	1:A:32:CYS:SG	0.56	2.40	19	16
1:A:30:GLN:NE2	1:A:35:MET:HE2	0.55	2.16	19	17
1:A:41:VAL:HG22	2:B:301:ALA:CB	0.55	2.26	3	5
1:A:38:ASN:HB3	1:A:41:VAL:HG12	0.54	1.79	17	6
1:A:21:ILE:HD11	1:A:32:CYS:HA	0.53	1.80	9	5
1:A:23:CYS:HB3	1:A:28:VAL:HG22	0.53	1.80	2	19
1:A:39:THR:CB	2:B:303:THR:HG21	0.53	2.34	18	1
1:A:22:GLN:NE2	1:A:27:MET:HG3	0.53	2.15	4	2
1:A:19:LEU:HG	2:B:305:GLN:HB3	0.53	1.81	15	2
1:A:30:GLN:OE1	1:A:46:CYS:HB3	0.51	2.05	5	10
1:A:25:LYS:HB2	1:A:49:CYS:SG	0.51	2.45	20	3
1:A:21:ILE:HD12	1:A:21:ILE:N	0.51	2.19	14	18
1:A:19:LEU:HG	2:B:305:GLN:HA	0.50	1.84	19	1
1:A:27:MET:HG2	1:A:27:MET:O	0.50	2.06	2	4
1:A:20:MET:HE2	1:A:29:TRP:CB	0.50	2.37	1	8
1:A:26:CYS:O	1:A:27:MET:HB3	0.50	2.06	9	3
1:A:39:THR:HB	2:B:303:THR:CG2	0.49	2.36	19	2
1:A:46:CYS:SG	1:A:48:GLN:HB3	0.49	2.47	16	1
1:A:21:ILE:HG12	1:A:37:VAL:HG11	0.49	1.82	14	5

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:23:CYS:SG	1:A:45:LEU:HA	0.49	2.47	12	3
1:A:26:CYS:HB2	1:A:28:VAL:HG22	0.48	1.84	18	6
1:A:20:MET:CE	1:A:29:TRP:HB3	0.48	2.38	6	5
1:A:35:MET:HA	1:A:52:ARG:NH2	0.48	2.24	13	1
1:A:19:LEU:HD23	2:B:305:GLN:HB2	0.48	1.86	18	1
1:A:25:LYS:HG3	1:A:45:LEU:HD21	0.48	1.85	9	2
1:A:21:ILE:HD12	1:A:21:ILE:H	0.47	1.69	16	4
1:A:10:ILE:HD11	1:A:34:CYS:O	0.47	2.09	12	2
1:A:19:LEU:CD2	2:B:303:THR:HB	0.47	2.39	9	2
1:A:19:LEU:HB2	1:A:32:CYS:SG	0.47	2.49	18	1
1:A:38:ASN:HB2	1:A:41:VAL:HG12	0.46	1.87	20	1
1:A:46:CYS:SG	1:A:48:GLN:HB2	0.46	2.51	3	3
1:A:47:GLU:H	1:A:47:GLU:CD	0.46	2.19	15	2
1:A:47:GLU:CD	1:A:47:GLU:H	0.45	2.19	4	5
1:A:21:ILE:CD1	1:A:37:VAL:HG21	0.45	2.40	14	3
1:A:13:LEU:HB3	1:A:15:LYS:HG3	0.45	1.89	6	1
1:A:35:MET:HG2	1:A:44:TYR:HH	0.45	1.70	13	1
1:A:21:ILE:HD11	1:A:32:CYS:CA	0.45	2.42	2	4
1:A:35:MET:HG2	1:A:44:TYR:CZ	0.44	2.47	13	1
1:A:8:ARG:CB	1:A:28:VAL:HG21	0.44	2.39	15	1
1:A:23:CYS:SG	1:A:45:LEU:HD23	0.44	2.52	8	2
1:A:15:LYS:HD3	1:A:17:GLU:OE1	0.44	2.13	11	1
1:A:19:LEU:HA	2:B:305:GLN:HA	0.43	1.90	18	1
1:A:9:CYS:SG	1:A:35:MET:SD	0.43	3.16	4	3
1:A:8:ARG:HB2	1:A:28:VAL:CG2	0.43	2.39	15	1
1:A:10:ILE:HD11	1:A:34:CYS:C	0.43	2.38	5	5
1:A:26:CYS:SG	1:A:28:VAL:HB	0.43	2.54	15	1
1:A:35:MET:CE	1:A:47:GLU:HG3	0.43	2.38	17	1
1:A:35:MET:HB2	1:A:37:VAL:CG2	0.42	2.44	10	4
1:A:26:CYS:HB3	1:A:28:VAL:CG1	0.42	2.39	6	1
1:A:27:MET:SD	1:A:27:MET:O	0.42	2.77	16	1
1:A:19:LEU:HG	2:B:305:GLN:HB2	0.42	1.92	8	1
1:A:7:ILE:HB	1:A:14:TYR:CD2	0.41	2.50	11	2
1:A:38:ASN:C	1:A:40:ASP:H	0.41	2.24	14	1
1:A:35:MET:HG3	1:A:44:TYR:OH	0.41	2.16	18	1
1:A:22:GLN:NE2	1:A:27:MET:HG2	0.41	2.24	19	1
1:A:22:GLN:OE1	2:B:302:ARG:HD3	0.41	2.15	11	1
1:A:9:CYS:SG	1:A:10:ILE:N	0.40	2.94	9	1
1:A:21:ILE:HD12	1:A:30:GLN:O	0.40	2.16	10	1
1:A:35:MET:HE1	1:A:47:GLU:CG	0.40	2.37	14	1

6.3 Torsion angles (i)

6.3.1 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 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	47/56 (84%)	43±1 (91±2%)	4±1 (8±1%)	0±0 (0±0%)	49 83
2	B	3/15 (20%)	3±0 (93±13%)	0±0 (7±13%)	0±0 (0±0%)	100 100
All	All	1000/1420 (70%)	916 (92%)	83 (8%)	1 (0%)	49 83

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	14	TYR	1

6.3.2 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 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	44/51 (86%)	37±1 (84±3%)	7±1 (16±3%)	4 40
2	B	3/9 (33%)	3±0 (95±12%)	0±0 (5±12%)	23 74
All	All	940/1200 (78%)	795 (85%)	145 (15%)	5 41

All 18 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	21	ILE	20
1	A	41	VAL	20
1	A	47	GLU	20
1	A	19	LEU	19
1	A	52	ARG	15
1	A	15	LYS	11
1	A	34	CYS	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	17	GLU	6
1	A	16	ASP	5
1	A	43	HIS	5
1	A	25	LYS	4
2	B	303	THR	3
1	A	27	MET	3
1	A	14	TYR	2
1	A	37	VAL	2
1	A	24	ASP	1
1	A	10	ILE	1
1	A	28	VAL	1

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	MLY	B	304	2	9,10,11	0.84±0.03	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	MLY	B	304	2	6,11,13	0.45±0.09	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	MLY	B	304	2	-	0±0,8,9,11	-

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.

6.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

6.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

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	685
Number of shifts mapped to atoms	685
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	63	-0.43 ± 0.31	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	59	0.22 ± 0.20	None needed (< 0.5 ppm)
$^{13}C'$	2	—	None (insufficient data)
^{15}N	50	-0.72 ± 0.80	None needed (imprecise)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 82%, i.e. 556 atoms were assigned a chemical shift out of a possible 681. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	197/256 (77%)	100/104 (96%)	52/102 (51%)	45/50 (90%)
Sidechain	325/380 (86%)	217/245 (89%)	102/119 (86%)	6/16 (38%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	34/45 (76%)	17/22 (77%)	16/19 (84%)	1/4 (25%)
Overall	556/681 (82%)	334/371 (90%)	170/240 (71%)	52/70 (74%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 74%, i.e. 670 atoms were assigned a chemical shift out of a possible 911. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	240/353 (68%)	125/145 (86%)	65/140 (46%)	50/68 (74%)
Sidechain	396/513 (77%)	263/329 (80%)	126/157 (80%)	7/27 (26%)
Aromatic	34/45 (76%)	17/22 (77%)	16/19 (84%)	1/4 (25%)
Overall	670/911 (74%)	405/496 (82%)	207/316 (66%)	58/99 (59%)

7.1.4 Statistically unusual chemical shifts (i)

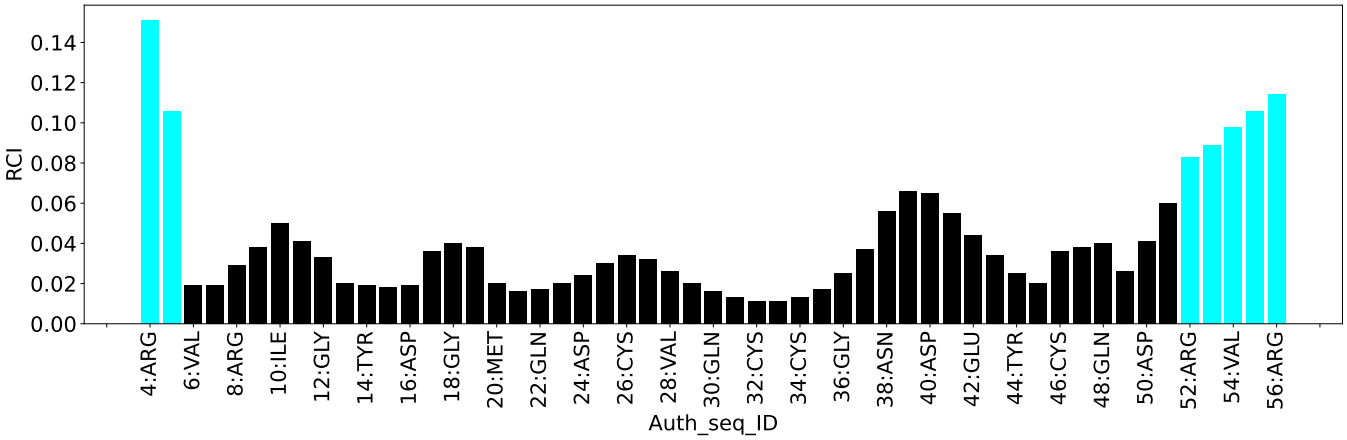
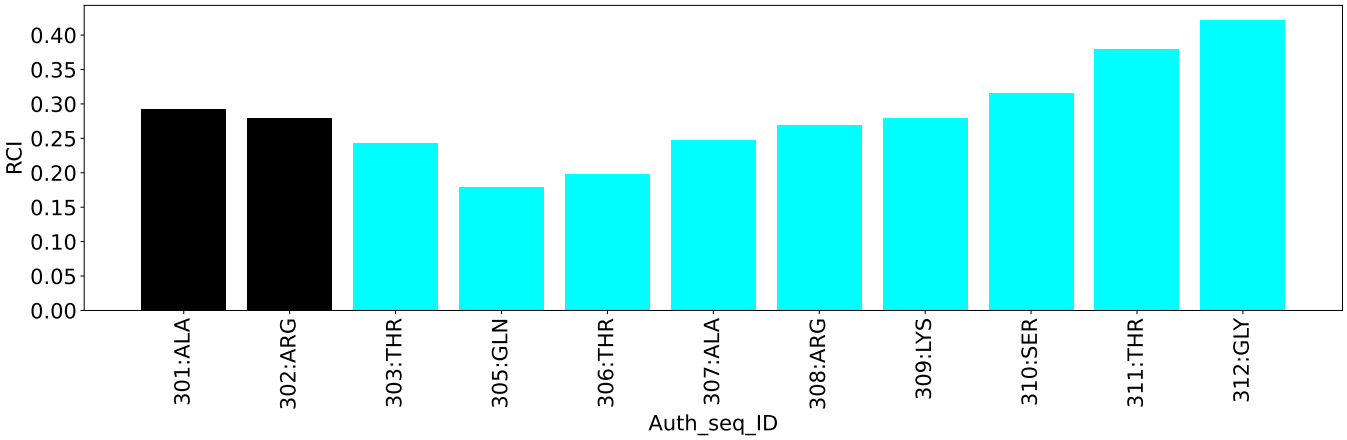
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	22	GLN	HB3	-1.14	0.71 – 3.33	-12.0
1	A	22	GLN	HG3	0.66	0.91 – 3.68	-5.9

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B: