



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 21, 2025 – 11:16 AM EDT

PDB ID : 9CVD / pdb_00009cvd
BMRB ID : 31192
Title : SET Domain of Histone-Lysine N-Methyltransferase NSD2 in Complex with Selective Inhibitor
Authors : Szczepski, K.; Jaremko, L.
Deposited on : 2024-07-29

This is a wwPDB NMR 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/NMRValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 2022.3.0, CSD as543be (2022)
buster-report : 1.1.7 (2018)
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
wwPDB-ShiftChecker : v1.2
BMRB Restraints Analysis : v1.2
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.42

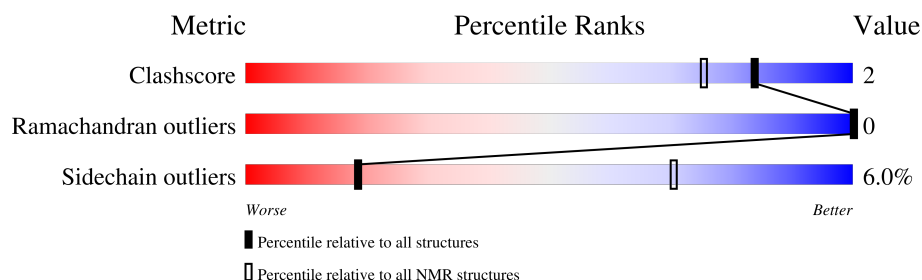
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 27%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	220	<div> <div style="width: 74%; background-color: green;"></div> <div style="width: 9%; background-color: yellow;"></div> <div style="width: 17%; background-color: cyan;"></div> </div> <div>74% 9% 17%</div>

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 4 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:990-A:1069, A:1075-A:1112, A:1118-A:1179, A:1195-A:1197 (183)	0.13	4

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 7, 9, 10
2	6, 8

3 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3451 atoms, of which 1678 are hydrogens and 0 are deuteriums.

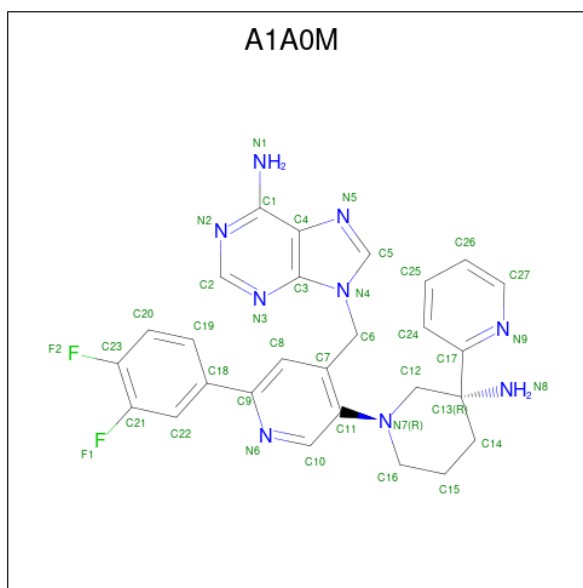
- Molecule 1 is a protein called Histone-lysine N-methyltransferase NSD2.

Mol	Chain	Residues	Atoms						Trace
1	A	220	Total	C	H	N	O	S	0
			3385	1085	1653	297	329	21	

- Molecule 2 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
2	A	3	Total	Zn
			3	3

- Molecule 3 is 9-{[(2M)-5-[(3R)-3-amino-3-(pyridin-2-yl)piperidin-1-yl]-2-(3,4-difluorophenyl)pyridin-4-yl]methyl}-9H-purin-6-amine (CCD ID: A1A0M) (formula: C₂₇H₂₅F₂N₉).



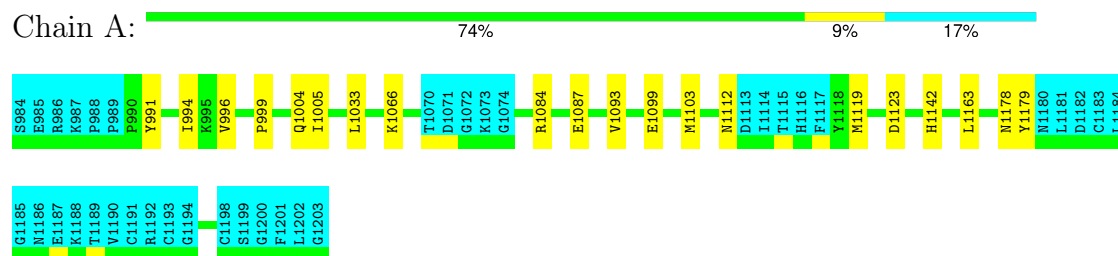
Mol	Chain	Residues	Atoms				
3	A	1	Total	C	F	H	N
			63	27	2	25	9

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.

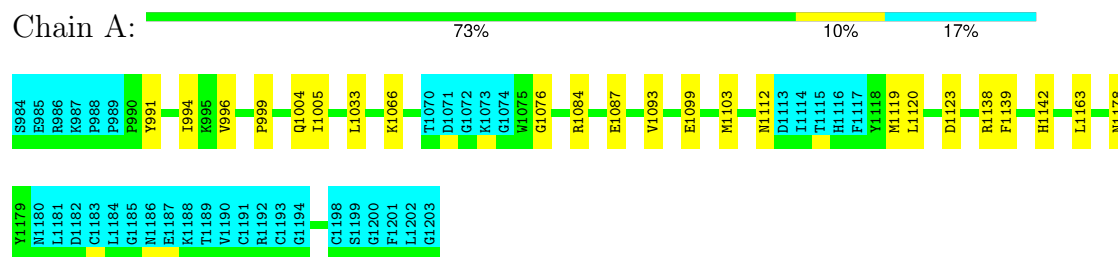
- Molecule 1: Histone-lysine N-methyltransferase NSD2



4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Histone-lysine N-methyltransferase NSD2



5 Refinement protocol and experimental data overview

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

Of the 500 calculated structures, 10 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
CNS	structure calculation	
HADDOCK	refinement	

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	794
Number of shifts mapped to atoms	794
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	27%

6 Model quality [i](#)

6.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, A1A0M

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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 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	1463	1403	1406	6±2
3	A	38	25	0	0±1
All	All	15040	14280	14061	65

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 2.

5 of 15 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1142:HIS:HB3	3:A:1304:A1A0M:N5	0.61	2.10	6	2
1:A:999:PRO:HB3	1:A:1005:ILE:HD11	0.58	1.75	6	10
1:A:1142:HIS:CE1	1:A:1178:ASN:HA	0.52	2.40	6	7
1:A:994:ILE:HD12	1:A:996:VAL:O	0.52	2.05	2	7
1:A:1068:ILE:HG12	1:A:1078:VAL:HG13	0.51	1.83	6	3

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	183/220 (83%)	175±1 (96±0%)	8±1 (4±0%)	0±0 (0±0%)	100	100
All	All	1830/2200 (83%)	1754 (96%)	76 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	162/194 (84%)	152±1 (94±0%)	10±1 (6±0%)	18	70
All	All	1620/1940 (84%)	1522 (94%)	98 (6%)	18	70

5 of 12 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	991	TYR	10
1	A	1004	GLN	10
1	A	1033	LEU	10
1	A	1066	LYS	10
1	A	1093	VAL	10

6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

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
3	A1A0M	A	1304	-	37,43,43	2.15±0.02	9±0 (24±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
3	A1A0M	A	1304	-	44,63,63	1.58±0.00	4±0 (9±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
3	A1A0M	A	1304	-	-	0±0,16,30,30	0±0,6,6,6

5 of 9 unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
3	A	1304	A1A0M	C18-C9	6.33	1.39	1.49	3	10
3	A	1304	A1A0M	C6-C7	6.04	1.39	1.51	6	10
3	A	1304	A1A0M	C2-N3	5.65	1.40	1.32	3	10
3	A	1304	A1A0M	C2-N2	3.98	1.41	1.33	1	10
3	A	1304	A1A0M	C11-N7	3.23	1.34	1.41	1	10

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

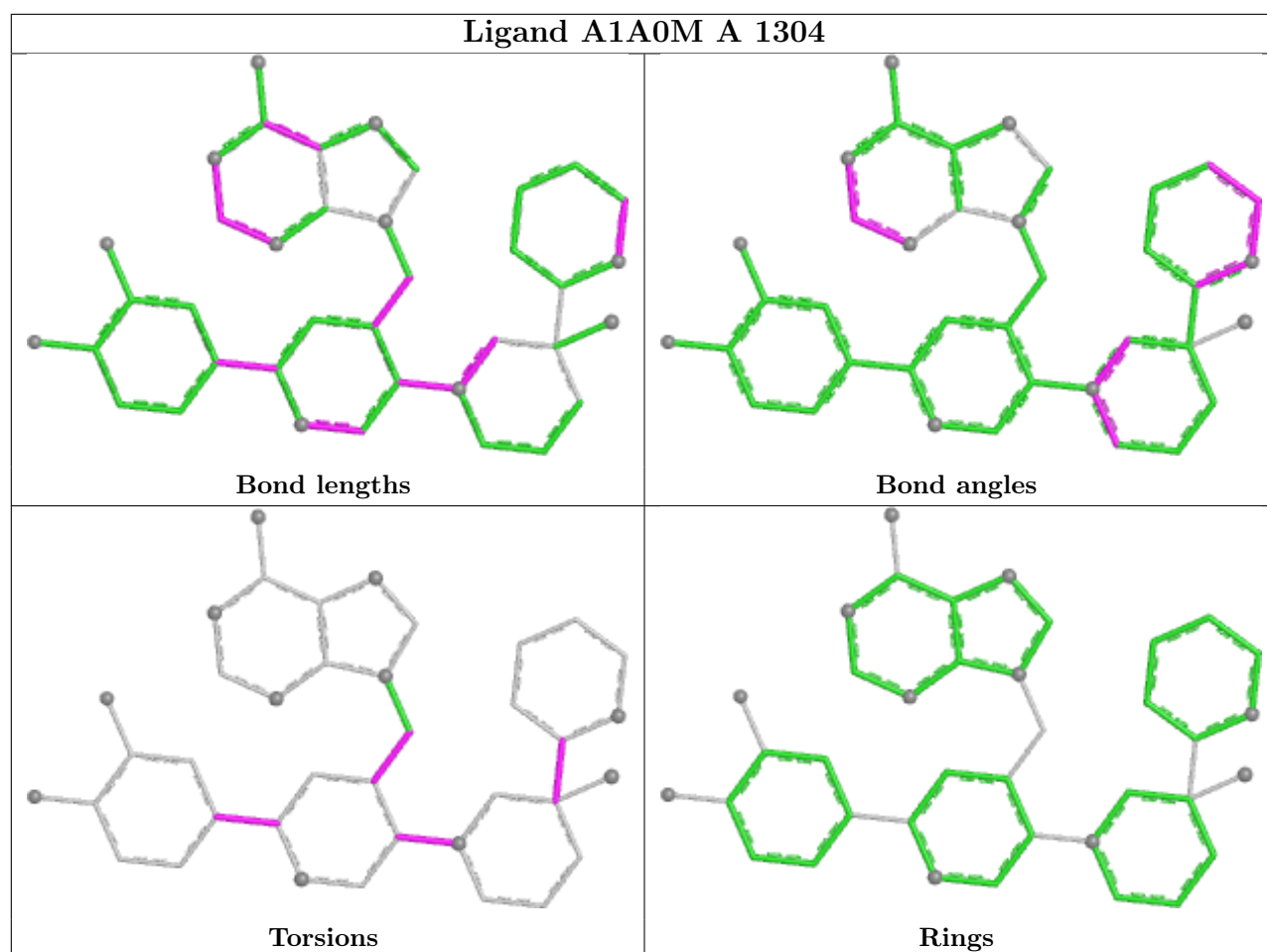
Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
3	A	1304	A1A0M	N3-C2-N2	7.35	118.70	128.67	6	10
3	A	1304	A1A0M	C27-N9-C17	2.54	120.77	117.48	5	10
3	A	1304	A1A0M	C26-C27-N9	2.45	119.55	123.42	2	10
3	A	1304	A1A0M	C16-N7-C12	2.17	118.37	113.86	4	10
3	A	1304	A1A0M	C10-N6-C9	2.04	120.58	117.93	2	2

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 27% for the well-defined parts and 27% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *D_1000285482_cs_P1.str.V5*

7.1.1 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	794
Number of shifts mapped to atoms	794
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

7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	205	2.73 ± 0.13	Should be checked
$^{13}\text{C}_\beta$	0	—	None (insufficient data)
$^{13}\text{C}'$	211	2.59 ± 0.14	Should be applied
^{15}N	189	-0.50 ± 0.33	None needed (< 0.5 ppm)

7.1.3 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 27%, i.e. 676 atoms were assigned a chemical shift out of a possible 2475. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	676/904 (75%)	160/366 (44%)	356/366 (97%)	160/172 (93%)
Sidechain	0/1363 (0%)	0/876 (0%)	0/429 (0%)	0/58 (0%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	0/208 (0%)	0/104 (0%)	0/100 (0%)	0/4 (0%)
Overall	676/2475 (27%)	160/1346 (12%)	356/895 (40%)	160/234 (68%)

7.1.4 Statistically unusual chemical shifts [i](#)

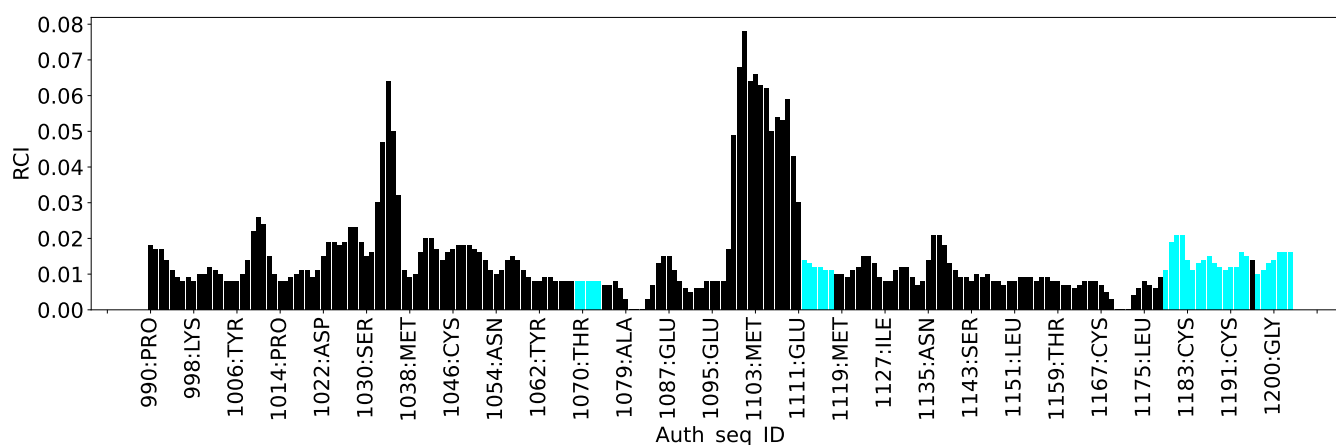
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	1170	PRO	C	999.90	169.47 – 184.06	564.2
1	A	1082	ASP	C	999.90	168.07 – 184.82	491.6
1	A	1172	GLY	C	999.90	164.92 – 182.89	459.6
1	A	1080	LYS	C	999.90	167.28 – 186.22	434.4
1	A	1081	ARG	C	999.90	166.60 – 186.40	415.9
1	A	1171	ALA	C	999.90	167.61 – 188.05	402.2

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:



8 NMR restraints analysis [i](#)

8.1 Conformationally restricting restraints [i](#)

The following table provides the summary of experimentally observed NMR restraints in different categories. Restraints are classified into different categories based on the sequence separation of the atoms involved.

Description	Value
Total distance restraints	50
Intra-residue ($ i-j =0$)	2
Sequential ($ i-j =1$)	0
Medium range ($ i-j >1$ and $ i-j <5$)	0
Long range ($ i-j \geq 5$)	8
Inter-chain	40
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	42
Number of restraints per residue	0.2
Number of long range restraints per residue ¹	0.0

¹Long range hydrogen bonds and disulfide bonds are counted as long range restraints while calculating the number of long range restraints per residue

8.2 Residual restraint violations [i](#)

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

8.2.1 Average number of distance violations per model [i](#)

Distance violations less than 0.1 Å are not included in the calculation. There are no distance violations

8.2.2 Average number of dihedral-angle violations per model [i](#)

Dihedral-angle violations less than 1° are not included in the calculation. There are no dihedral-angle violations

9 Distance violation analysis [i](#)

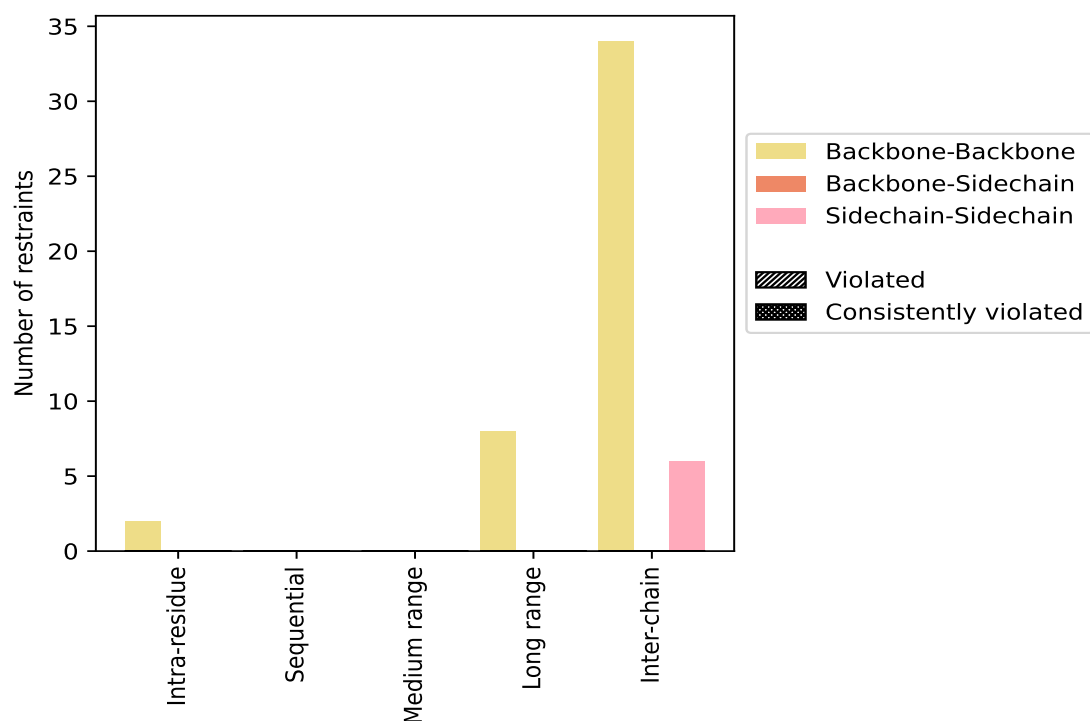
9.1 Summary of distance violations [i](#)

The following table shows the summary of distance violations in different restraint categories based on the sequence separation of the atoms involved. Each category is further sub-divided into three sub-categories based on the atoms involved. Violations less than 0.1 Å are not included in the statistics.

Restrains type	Count	% ¹	Violated ³			Consistently Violated ⁴		
			Count	% ²	% ¹	Count	% ²	% ¹
Intra-residue (i-j =0)	2	4.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	2	4.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sequential (i-j =1)	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Medium range (i-j >1 & i-j <5)	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Long range (i-j ≥5)	8	16.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	8	16.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Inter-chain	40	80.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	34	68.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	6	12.0	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	50	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	44	88.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	6	12.0	0	0.0	0.0	0	0.0	0.0

¹ percentage calculated with respect to the total number of distance restraints, ² percentage calculated with respect to the number of restraints in a particular restraint category, ³ violated in at least one model, ⁴ violated in all the models

9.1.1 Bar chart : Distribution of distance restraints and violations [i](#)



Violated and consistently violated restraints are shown using different hatch patterns in their respective categories. The hydrogen bonds and disulfide bonds are counted in their appropriate category on the x-axis

9.2 Distance violation statistics for each model [i](#)

No violations found

9.3 Distance violation statistics for the ensemble [i](#)

No violations found

9.4 Most violated distance restraints in the ensemble [i](#)

No violations found

9.5 All violated distance restraints [i](#)

No violations found

10 Dihedral-angle violation analysis ⓘ

No dihedral-angle restraints found