wwPDB NMR Structure Validation Summary Report (i)

May 7, 2024 – 12:31 pm BST

PDB ID	:	2J2S
Title	:	Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase
Authors	:	Allen, M.D.; Grummitt, C.G.; Hilcenko, C.; Young-Min, S.; Tonkin, L.M.; Johnson, C.M.; Bycroft, M.; Warren, A.J.
Deposited on	:	2006-08-17

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 (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:

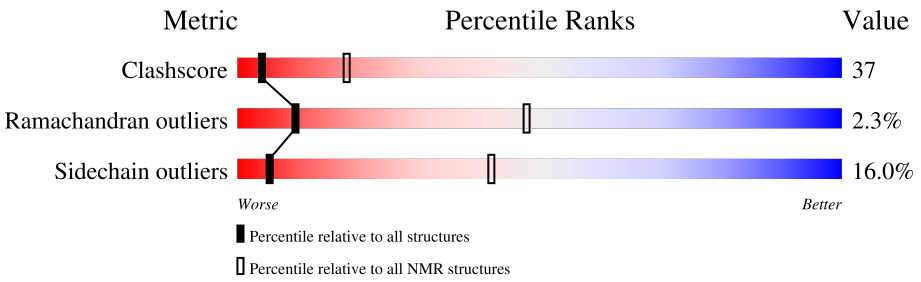
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 was not calculated.

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	72	28% 26% 6% • 39%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:1154-A:1197 (44)	0.29	20

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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1136 atoms, of which 580 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ZINC FINGER PROTEIN HRX.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	72	1134	336	580	113	94	11	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1143	GLY	GLU	engineered mutation	UNP Q03164
A	1144	GLY	PRO	engineered mutation	UNP Q03164
A	1145	SER	PRO	engineered mutation	UNP Q03164

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

4 Residue-property plots (i)

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: ZINC FINGER PROTEIN HRX

Chain A: 28% 26% 6% • 39%

G1143 G1144 S1145 V1146 K1147 K1148 G1149 R1150 R1151 S1152 R1153 R1154 C1158 P1159 G1160 V1163 P1164 E1165 G1168 V1169 C1170 T1171 N1172 C1173 L1174 D1175 K1176 F1179 G1180 G1181 R1182 N1183 I1184 K1185 K1186 C1189 K1190 M1191 R1192 L1197 Q1198 W1199 M1200 P1201 S1202 K1203 A1204 Y1205 L1206 Q1207 K1208 Q1209 A1210 K1211

A1212 V1213 K1214

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

The representative model is number 20. Colouring as in section 4.1 above.

• Molecule 1: ZINC FINGER PROTEIN HRX

Chain A: 31% 25% • • 39%

G1143 G1144 S1145 V1146 K1147 K1148 G1149 R1150 R1151 S1152 R1153 R1154 C1158 P1159 G1160 V1163 P1164 E1165 D1166 C1167 G1168 V1169 C1170 T1171 N1172 C1173 L1174 D1175 K1176 F1179 G1180 G1181 R1182 N1183 I1184 K1185 K1186 M1191 C1194 L1197 Q1198 W1199 M1200 P1201 S1202 K1203 A1204 Y1205 L1206 Q1207 K1208 Q1209 A1210 K1211

A1212 V1213 K1214

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *CNS*.

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *NO VIOLATIONS*.

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

Software name	Classification	Version
CNS	refinement	
ANSIG	structure solution	

No chemical shift data was provided.

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

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	331	334	330	24±2
All	All	6660	6680	6600	487

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 37.

5 of 51 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1168:GLY:HA3	1:A:1179:PHE:CE1	0.82	2.09	20	20
1:A:1169:VAL:HA	1:A:1174:LEU:HD21	0.77	1.56	8	20
1:A:1154:ARG:HA	1:A:1197:LEU:HD23	0.76	1.56	10	20
1:A:1168:GLY:CA	1:A:1179:PHE:CE1	0.74	2.71	10	20
1:A:1159:PRO:O	1:A:1163:VAL:HG23	0.71	1.85	8	20

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	44/72 (61%)	40±1 (90±3%)	3±1 (8±3%)	1±0 (2±0%)	9 48
All	All	880/1440 (61%)	792 (90%)	68 (8%)	20 (2%)	9 48

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	1186	LYS	20

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	39/61 (64%)	33±2 (84±4%)	6±2 (16±4%)	5 42
All	All	780/1220 (64%)	655 (84%)	125 (16%)	5 42

5 of 15 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	1171	THR	20
1	A	1197	LEU	20
1	A	1154	ARG	16
1	A	1186	LYS	15
1	A	1165	GLU	13

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 monosaccharides 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 (i)

No chemical shift data were provided