wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 02:26 AM EDT

PDB ID	:	2MH2
BMRB ID	:	19613
Title	:	Structural insights into the DNA recognition and protein interaction domains reveal fundamental homologous DNA pairing properties of HOP2
Authors	:	Moktan, H.; Guiraldelli, M.F.; Eyter, C.A.; Zhao, W.; Camerini-Otero, R.D.; Sung, P.; Zhou, D.H.; Pezza, R.J.
Deposited on	:	2013-11-13

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:

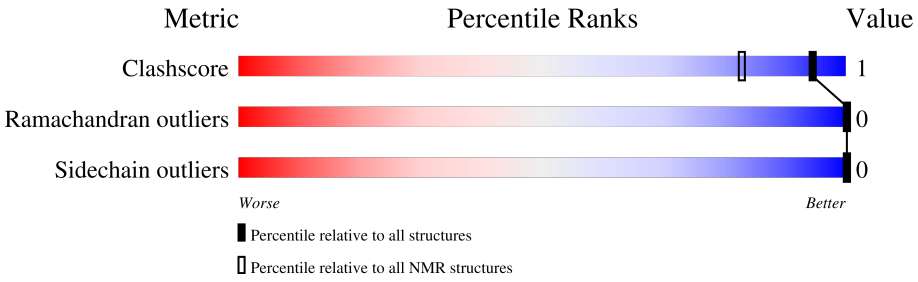
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

1 Overall quality at a glance

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

The overall completeness of chemical shifts assignment is 81%.

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	84	64% · 10% 24%

2 Ensemble composition and analysis

This entry contains 20 models. Model 15 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:12-A:62, A:69-A:73 (56)	0.21	15

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 and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1018 atoms, of which 514 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Homologous-pairing protein 2 homolog.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	64	1018	322	514	89	93	0

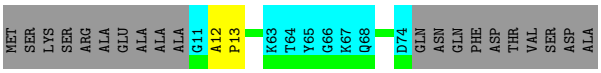
4 Residue-property plots (i)

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: Homologous-pairing protein 2 homolog

Chain A:

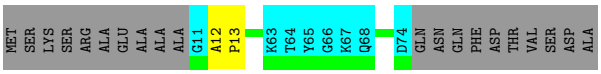


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: Homologous-pairing protein 2 homolog

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *distance geometry*.

Of the 40000 calculated structures, 20 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum*.

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

Software name	Classification	Version
CS-ROSETTA	structure solution	
RPF	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	877
Number of shifts mapped to atoms	718
Number of unparsed shifts	0
Number of shifts with mapping errors	159
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

6 Model quality

6.1 Standard geometry

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

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	442	454	454	1±0
All	All	8840	9080	9080	20

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:12:ALA:N	1:A:13:PRO:CD	0.56	2.69	3	20

6.3 Torsion angles

6.3.1 Protein backbone

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	56/84 (67%)	56±0 (100±1%)	0±0 (0±1%)	0±0 (0±0%)	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1120/1680 (67%)	1118 (100%)	2 (0%)	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	44/65 (68%)	44±0 (100±0%)	0±0 (0±0%)	100 100
All	All	880/1300 (68%)	880 (100%)	0 (0%)	100 100

There are no protein residues with a non-rotameric sidechain to report.

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)

There are no ligands in this entry.

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 81% for the well-defined parts and 81% 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	877
Number of shifts mapped to atoms	718
Number of unparsed shifts	0
Number of shifts with mapping errors	159
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 159) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	5	ARG	C	176.95	0.04	1
1	A	5	ARG	CA	56.95	0.03	1
1	A	5	ARG	CB	30.45	0.05	1
1	A	5	ARG	CD	43.16	0.00	1
1	A	5	ARG	CG	26.97	0.00	1
1	A	5	ARG	HA	4.44	0.00	1
1	A	5	ARG	HB2	1.79	0.00	1
1	A	5	ARG	HB3	1.79	0.00	1
1	A	5	ARG	HD2	3.17	0.00	1
1	A	5	ARG	HD3	3.17	0.00	1
1	A	6	ALA	C	178.65	0.00	1
1	A	6	ALA	CA	53.37	0.04	1
1	A	6	ALA	CB	18.72	0.03	1
1	A	6	ALA	H	8.31	0.01	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	6	ALA	HA	4.22	0.00	1
1	A	6	ALA	HB1	1.37	0.02	1
1	A	6	ALA	HB2	1.37	0.02	1
1	A	6	ALA	HB3	1.37	0.02	1
1	A	6	ALA	N	124.41	0.05	1
1	A	7	GLU	C	177.25	0.01	1
1	A	7	GLU	CA	57.18	0.06	1
1	A	7	GLU	CB	29.99	0.05	1
1	A	7	GLU	CG	36.26	0.01	1
1	A	7	GLU	H	8.32	0.01	1
1	A	7	GLU	HA	4.18	0.01	1
1	A	7	GLU	HB2	2.0	0.01	2
1	A	7	GLU	HB3	1.99	0.01	2
1	A	7	GLU	HG2	2.26	0.01	1
1	A	7	GLU	HG3	2.26	0.00	1
1	A	7	GLU	N	119.8	0.03	1
1	A	8	ALA	C	178.17	0.00	1
1	A	8	ALA	CA	53.14	0.05	1
1	A	8	ALA	CB	18.8	0.09	1
1	A	8	ALA	H	8.22	0.01	1
1	A	8	ALA	HA	4.21	0.00	1
1	A	8	ALA	HB1	1.37	0.01	1
1	A	8	ALA	HB2	1.37	0.01	1
1	A	8	ALA	HB3	1.37	0.01	1
1	A	8	ALA	N	124.25	0.05	1
1	A	9	ALA	C	177.92	0.03	1
1	A	9	ALA	CA	53.06	0.12	1
1	A	9	ALA	CB	18.75	0.12	1
1	A	9	ALA	H	8.11	0.00	1
1	A	9	ALA	HA	4.22	0.00	1
1	A	9	ALA	HB1	1.37	0.01	1
1	A	9	ALA	HB2	1.37	0.01	1
1	A	9	ALA	HB3	1.37	0.01	1
1	A	9	ALA	N	121.75	0.04	1
1	A	10	ALA	C	178.5	0.01	1
1	A	10	ALA	CA	52.8	0.11	1
1	A	10	ALA	CB	19.12	0.03	1
1	A	10	ALA	H	8.02	0.00	1
1	A	10	ALA	HA	4.2	0.00	1
1	A	10	ALA	HB1	1.39	0.01	1
1	A	10	ALA	HB2	1.39	0.01	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	10	ALA	HB3	1.39	0.01	1
1	A	10	ALA	N	121.75	0.03	1
1	A	75	GLN	C	176.02	0.01	1
1	A	75	GLN	CA	56.24	0.05	1
1	A	75	GLN	CB	29.23	0.08	1
1	A	75	GLN	CG	33.45	0.12	1
1	A	75	GLN	H	8.59	0.00	1
1	A	75	GLN	HA	4.25	0.01	1
1	A	75	GLN	HB2	2.07	0.00	2
1	A	75	GLN	HB3	1.96	0.00	2
1	A	75	GLN	HE21	7.5	0.04	2
1	A	75	GLN	HE22	6.79	0.00	2
1	A	75	GLN	HG2	2.33	0.00	1
1	A	75	GLN	HG3	2.33	0.01	1
1	A	75	GLN	N	122.15	0.01	1
1	A	75	GLN	NE2	111.87	0.08	1
1	A	76	ASN	C	175.28	0.00	1
1	A	76	ASN	CA	53.65	0.05	1
1	A	76	ASN	CB	38.78	0.05	1
1	A	76	ASN	H	8.56	0.01	1
1	A	76	ASN	HA	4.61	0.02	1
1	A	76	ASN	HB2	2.72	0.05	1
1	A	76	ASN	HB3	2.72	0.05	1
1	A	76	ASN	HD21	7.67	0.00	2
1	A	76	ASN	HD22	6.93	0.00	2
1	A	76	ASN	N	119.27	0.03	1
1	A	76	ASN	ND2	112.78	0.34	1
1	A	77	GLN	C	175.64	0.01	1
1	A	77	GLN	CA	56.13	0.03	1
1	A	77	GLN	CB	29.18	0.06	1
1	A	77	GLN	CG	33.42	0.00	1
1	A	77	GLN	H	8.23	0.11	1
1	A	77	GLN	HA	3.78	0.60	1
1	A	77	GLN	HB2	2.05	0.07	2
1	A	77	GLN	HB3	1.83	0.02	2
1	A	77	GLN	N	119.99	0.02	1
1	A	78	PHE	C	175.42	0.01	1
1	A	78	PHE	CA	57.53	0.08	1
1	A	78	PHE	CB	39.24	0.15	1
1	A	78	PHE	H	8.08	0.00	1
1	A	78	PHE	HA	4.57	0.01	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	78	PHE	HB2	3.15	0.01	2
1	A	78	PHE	HB3	2.9	0.01	2
1	A	78	PHE	N	120.01	0.01	1
1	A	79	ASP	C	176.2	0.01	1
1	A	79	ASP	CA	54.45	0.04	1
1	A	79	ASP	CB	41.2	0.01	1
1	A	79	ASP	H	8.27	0.01	1
1	A	79	ASP	HA	4.62	0.01	1
1	A	79	ASP	HB2	2.62	0.00	1
1	A	79	ASP	HB3	2.62	0.00	1
1	A	79	ASP	N	121.2	0.01	1
1	A	80	THR	C	174.51	0.00	1
1	A	80	THR	CA	61.91	0.11	1
1	A	80	THR	CB	69.93	0.15	1
1	A	80	THR	CG2	21.48	0.00	1
1	A	80	THR	H	8.05	0.01	1
1	A	80	THR	HA	4.32	0.01	1
1	A	80	THR	HB	4.18	0.01	1
1	A	80	THR	HG21	1.16	0.00	0
1	A	80	THR	HG22	1.16	0.00	0
1	A	80	THR	HG23	1.16	0.00	0
1	A	80	THR	N	114.33	0.01	1
1	A	81	VAL	C	176.09	0.00	1
1	A	81	VAL	CA	62.2	0.17	1
1	A	81	VAL	CB	32.6	0.03	1
1	A	81	VAL	CG1	20.8	0.00	2
1	A	81	VAL	CG2	20.24	0.00	2
1	A	81	VAL	H	8.17	0.00	1
1	A	81	VAL	HA	4.15	0.01	1
1	A	81	VAL	HB	2.05	0.00	1
1	A	81	VAL	HG11	0.89	0.00	2
1	A	81	VAL	HG12	0.89	0.00	2
1	A	81	VAL	HG13	0.89	0.00	2
1	A	81	VAL	HG21	0.88	0.00	2
1	A	81	VAL	HG22	0.88	0.00	2
1	A	81	VAL	HG23	0.88	0.00	2
1	A	81	VAL	N	122.63	0.02	1
1	A	82	SER	C	174.13	0.01	1
1	A	82	SER	CA	58.16	0.03	1
1	A	82	SER	CB	63.8	0.24	1
1	A	82	SER	H	8.4	0.01	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	82	SER	HA	4.45	0.01	1
1	A	82	SER	HB2	3.82	0.01	2
1	A	82	SER	HB3	3.83	0.00	2
1	A	82	SER	N	119.64	0.02	1
1	A	83	ASP	C	174.85	0.01	1
1	A	83	ASP	CA	54.34	0.05	1
1	A	83	ASP	CB	41.05	0.04	1
1	A	83	ASP	CG	177.04	0.01	1
1	A	83	ASP	H	8.36	0.00	1
1	A	83	ASP	HA	4.58	0.01	1
1	A	83	ASP	HB2	2.81	0.01	2
1	A	83	ASP	HB3	3.21	0.01	2
1	A	83	ASP	N	123.22	0.02	1
1	A	84	ALA	C	182.52	0.00	1
1	A	84	ALA	CA	53.85	0.06	1
1	A	84	ALA	CB	20.07	0.07	1
1	A	84	ALA	H	7.82	0.01	1
1	A	84	ALA	HA	4.06	0.01	1
1	A	84	ALA	HB1	1.27	0.01	1
1	A	84	ALA	HB2	1.27	0.01	1
1	A	84	ALA	HB3	1.27	0.01	1
1	A	84	ALA	N	129.06	0.03	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	78	-0.30 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	72	0.11 ± 0.18	None needed (< 0.5 ppm)
$^{13}C'$	76	-0.21 ± 0.20	None needed (< 0.5 ppm)
^{15}N	75	1.24 ± 0.34	Should be applied

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 81%, i.e. 640 atoms were assigned a chemical shift out of a possible 786. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	275/281 (98%)	113/115 (98%)	109/112 (97%)	53/54 (98%)
Sidechain	365/451 (81%)	244/292 (84%)	117/139 (84%)	4/20 (20%)
Aromatic	0/54 (0%)	0/26 (0%)	0/27 (0%)	0/1 (0%)
Overall	640/786 (81%)	357/433 (82%)	226/278 (81%)	57/75 (76%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: