wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 07:08 AM EDT

PDB ID	:	2LD5
BMRB ID	:	16577
Title	:	Solution NMR-derived complex structure of Hoxa13 DNA binding domain bound to DNA
Authors	:	Zhang, Y.
Deposited on	:	2011-05-14

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:

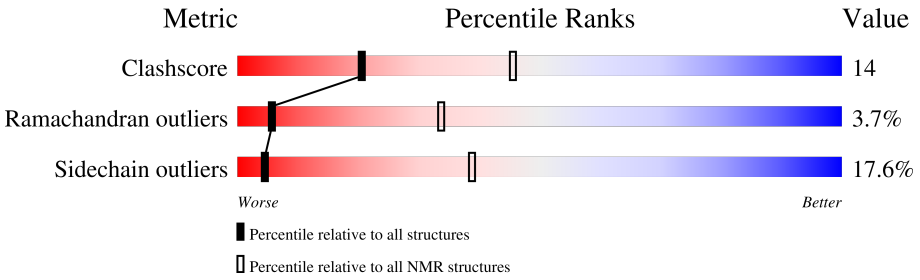
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 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	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	71	75% 15% · 6%
2	B	11	18% 82%
3	C	11	9% 91%

2 Ensemble composition and analysis

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:10-A:73 (64)	0.64	1

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, 5, 8, 9, 10
2	4, 6, 7

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1450 atoms, of which 428 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Homeobox protein Hox-A13.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	67	749	360	175	114	100	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	HIS	-	expression tag	UNP Q62424
A	4	MET	-	expression tag	UNP Q62424
A	5	LEU	-	expression tag	UNP Q62424
A	6	GLU	-	expression tag	UNP Q62424

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*AP*AP*AP*TP*AP*AP*AP*AP*TP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	11	346	108	124	45	59	10	0

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*AP*TP*TP*TP*TP*AP*TP*TP*TP*G)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
3	C	11	355	110	129	34	71	11	0

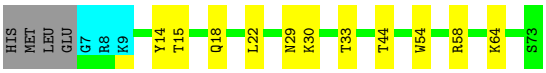
4 Residue-property plots

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.

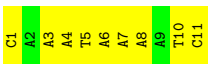
• Molecule 1: Homeobox protein Hox-A13

Chain A:



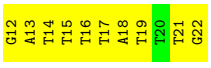
• Molecule 2: DNA (5'-D(*CP*AP*AP*AP*TP*AP*AP*AP*AP*TP*C)-3')

Chain B:



• Molecule 3: DNA (5'-D(P*GP*AP*TP*TP*TP*TP*AP*TP*TP*TP*G)-3')

Chain C:

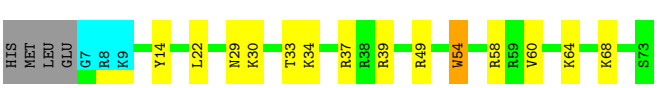


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: Homeobox protein Hox-A13

Chain A:



• Molecule 2: DNA (5'-D(*CP*AP*AP*AP*TP*AP*AP*AP*AP*TP*C)-3')

Chain B:

• Molecule 3: DNA (5'-D(P*GP*AP*TP*TP*TP*TP*AP*TP*TP*TP*G)-3')

Chain C:

5 Refinement protocol and experimental data overview

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

Of the 200 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	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	423
Number of shifts mapped to atoms	333
Number of unparsed shifts	0
Number of shifts with mapping errors	90
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	27%

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	550	164	596	8±2
2	B	222	124	125	10±2
3	C	226	129	129	15±4
All	All	9980	4170	8500	250

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 14.

5 of 104 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:53:ILE:HG23	3:C:19:DT:H71	0.87	1.44	3	2
2:B:3:DA:H1'	2:B:4:DA:H5'	0.79	1.54	10	3
2:B:4:DA:H2"	2:B:5:DT:H71	0.78	1.56	3	3
2:B:11:DC:O2	3:C:12:DG:N2	0.77	2.17	4	10
2:B:1:DC:O2	3:C:22:DG:N2	0.77	2.18	4	10

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	63/71 (89%)	56±1 (88±1%)	5±1 (8±1%)	2±1 (4±1%)	6 34
All	All	630/710 (89%)	555 (88%)	52 (8%)	23 (4%)	6 34

5 of 6 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	30	LYS	10
1	A	29	ASN	6
1	A	14	TYR	4
1	A	10	LYS	1
1	A	71	THR	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	62/68 (91%)	51±3 (82±4%)	11±3 (18±4%)	4 39
All	All	620/680 (91%)	511 (82%)	109 (18%)	4 39

5 of 36 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	22	LEU	9
1	A	54	TRP	9
1	A	58	ARG	8
1	A	15	THR	7
1	A	33	THR	6

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 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: *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	423
Number of shifts mapped to atoms	333
Number of unparsed shifts	0
Number of shifts with mapping errors	90
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

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 90) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	3	HIS	HA	4.68	0.000	1
1	A	3	HIS	HB2	3.156	0.000	2
1	A	3	HIS	C	174.771	0.000	1
1	A	3	HIS	CA	55.866	0.037	1
1	A	3	HIS	CB	29.79	0.000	1
1	A	4	MET	H	8.341	0.003	1
1	A	4	MET	C	175.951	0.000	1
1	A	4	MET	CA	55.919	0.131	1
1	A	4	MET	CB	32.67	0.231	1
1	A	4	MET	N	121.938	0.033	1
1	A	5	LEU	H	8.3	0.003	1
1	A	5	LEU	HA	4.329	0.000	1
1	A	5	LEU	HB2	1.595	0.000	2
1	A	5	LEU	C	177.221	0.000	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	5	LEU	CA	55.453	0.159	1
1	A	5	LEU	CB	42.083	0.256	1
1	A	5	LEU	N	123.735	0.031	1
1	A	6	GLU	H	8.306	0.004	1
1	A	6	GLU	HA	4.274	0.000	1
1	A	6	GLU	HB2	2.275	0.000	2
1	A	6	GLU	HB3	1.997	0.000	2
1	A	6	GLU	C	176.841	0.000	1
1	A	6	GLU	CA	56.832	0.110	1
1	A	6	GLU	CB	30.142	0.279	1
1	A	6	GLU	N	121.823	0.161	1
1	A	7	GLY	HA2	3.934	0.014	2
1	A	7	GLY	HA3	3.934	0.014	2
1	A	8	ARG	HA	4.308	0.000	1
1	A	8	ARG	HB2	1.737	0.000	2
1	A	14	TYR	HD1	6.961	0.001	3
1	A	16	LYS	HA	4.039	0.000	1
1	A	17	VAL	HA	3.778	0.007	1
1	A	17	VAL	HB	1.963	0.000	1
1	A	21	GLU	HA	4.15	0.000	1
1	A	22	LEU	HA	3.804	0.000	1
1	A	22	LEU	HB2	-1.015	0.001	2
1	A	22	LEU	HD21	-0.294	0.006	2
1	A	22	LEU	HD22	-0.294	0.006	2
1	A	22	LEU	HD23	-0.294	0.006	2
1	A	27	ALA	HA	4.046	0.018	1
1	A	27	ALA	HB1	1.531	0.007	1
1	A	27	ALA	HB2	1.531	0.007	1
1	A	27	ALA	HB3	1.531	0.007	1
1	A	28	THR	HA	4.209	0.006	1
1	A	28	THR	HB	4.427	0.000	1
1	A	28	THR	HG21	1.324	0.000	1
1	A	28	THR	HG22	1.324	0.000	1
1	A	28	THR	HG23	1.324	0.000	1
1	A	31	PHE	HA	4.656	0.000	1
1	A	33	THR	HA	4.446	0.000	1
1	A	33	THR	HB	4.663	0.008	1
1	A	33	THR	HG21	1.33	0.000	1
1	A	33	THR	HG22	1.33	0.000	1
1	A	33	THR	HG23	1.33	0.000	1
1	A	35	ASP	HA	4.379	0.000	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	39	ARG	HA	4.157	0.000	1
1	A	42	ALA	HA	4.217	0.006	1
1	A	42	ALA	HB1	1.573	0.021	1
1	A	42	ALA	HB2	1.573	0.021	1
1	A	42	ALA	HB3	1.573	0.021	1
1	A	43	THR	HA	4.145	0.022	1
1	A	44	THR	HG21	1.238	0.000	1
1	A	44	THR	HG22	1.238	0.000	1
1	A	44	THR	HG23	1.238	0.000	1
1	A	45	ASN	HA	4.399	0.000	1
1	A	46	LEU	HA	4.507	0.000	1
1	A	53	ILE	HA	3.676	0.000	1
1	A	54	TRP	HD1	7.377	0.000	1
1	A	58	ARG	HB2	-0.541	0.000	2
1	A	60	VAL	HA	3.726	0.000	1
1	A	63	LYS	HA	4.004	0.000	1
1	A	63	LYS	HB2	2.038	0.000	2
1	A	64	LYS	HA	4.079	0.010	1
1	A	64	LYS	HB2	1.974	0.002	2
1	A	65	VAL	HA	3.8	0.003	1
1	A	65	VAL	HB	2.261	0.030	1
1	A	67	ASN	HA	4.589	0.000	1
1	A	67	ASN	HB2	2.873	0.000	2
1	A	68	LYS	HA	4.344	0.000	1
1	A	68	LYS	HB2	1.839	0.000	2
1	A	70	LYS	HA	4.359	0.000	1
1	A	70	LYS	HB2	1.916	0.000	2
1	A	71	THR	HA	4.417	0.045	1
1	A	71	THR	HG21	1.255	0.000	1
1	A	71	THR	HG22	1.255	0.000	1
1	A	71	THR	HG23	1.255	0.000	1
1	A	72	THR	HA	4.465	0.000	1
1	A	72	THR	HG21	1.244	0.000	1
1	A	72	THR	HG22	1.244	0.000	1
1	A	72	THR	HG23	1.244	0.000	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	71	-0.72 ± 0.21	Should be checked
$^{13}C_{\beta}$	70	0.37 ± 0.05	None needed (< 0.5 ppm)
$^{13}C'$	69	-0.58 ± 0.24	Should be applied
^{15}N	69	0.13 ± 0.42	None needed (< 0.5 ppm)

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 27%, i.e. 380 atoms were assigned a chemical shift out of a possible 1419. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	276/318 (87%)	87/127 (69%)	126/128 (98%)	63/63 (100%)
Sidechain	100/627 (16%)	35/399 (9%)	65/186 (35%)	0/42 (0%)
Aromatic	4/50 (8%)	3/24 (12%)	0/25 (0%)	1/1 (100%)
Sugar	0/264 (0%)	0/154 (0%)	0/110 (0%)	0/0 (—%)
Base	0/160 (0%)	0/94 (0%)	0/42 (0%)	0/24 (0%)
Overall	380/1419 (27%)	125/798 (16%)	191/491 (39%)	64/130 (49%)

7.1.4 Statistically unusual chemical shifts

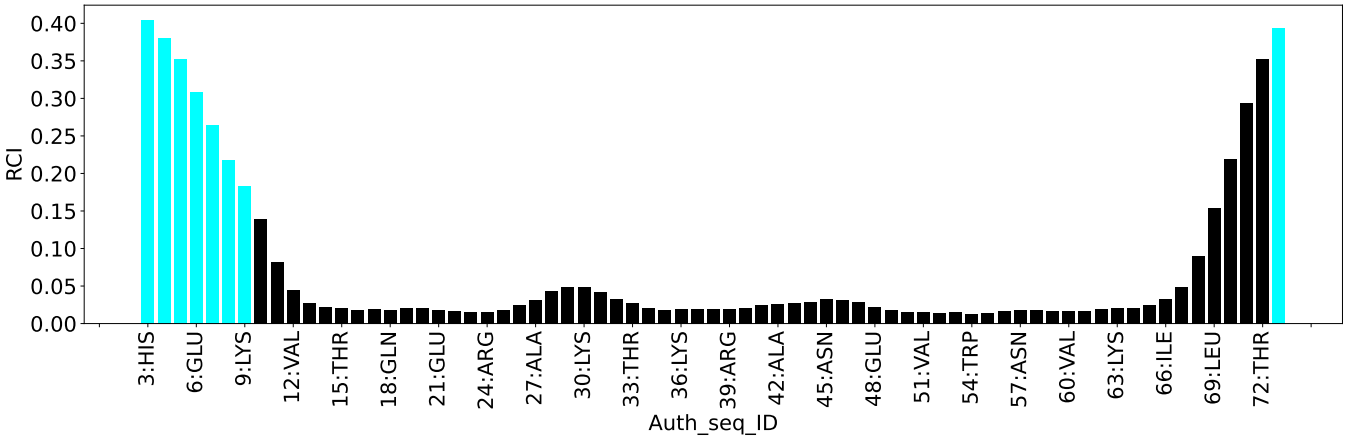
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	58	ARG	HB2	-0.54	0.52 – 3.08	-9.1
1	A	22	LEU	HB2	-1.01	-0.07 – 3.30	-7.8

7.1.5 Random Coil Index (RCI) plots

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:


RCI
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
3:HIS
6:GLU
9:LYS
12:VAL
15:THR
18:GLN
21:GLU
24:ARG
27:ALA
30:LYS
33:THR
36:LYS
39:ARG
42:ALA
45:ASN
48:GLU
51:VAL
54:TRP
57:ASN
60:VAL
63:LYS
66:ILE
69:LEU
72:THR
Auth_seq_ID