wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 07:58 AM EDT

PDB ID	:	2LKX
BMRB ID	:	18015
Title	:	NMR structure of the homeodomain of Pitx2 in complex with a TAATCC DNA binding site
Authors	:	Baird-Titus, J.M.; Doerdelmann, T.; Chaney, B.A.; Clark-Baldwin, K.; Dave, V.; Ma, J.
Deposited on	:	2011-10-21

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:

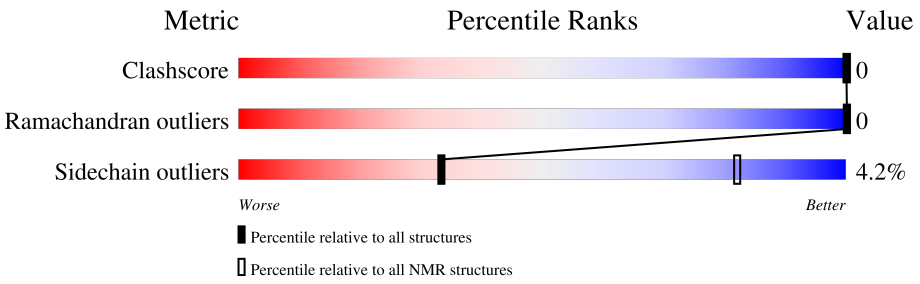
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 46%.

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	68	57% 6% • 35%
2	B	13	8% 92%
3	C	13	8% 92%

2 Ensemble composition and analysis (i)

This entry contains 17 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:9-A:52 (44)	0.46	8

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 6 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2010 atoms, of which 884 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Pituitary homeobox 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	68	1188	370	589	121	107	1	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP O75364
A	-1	SER	-	expression tag	UNP O75364
A	61	GLU	-	expression tag	UNP O75364
A	62	PHE	-	expression tag	UNP O75364
A	63	ILE	-	expression tag	UNP O75364
A	64	VAL	-	expression tag	UNP O75364
A	65	THR	-	expression tag	UNP O75364
A	66	ASP	-	expression tag	UNP O75364

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	13	405	124	148	44	77	12	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
3	C	13	417	128	147	55	75	12	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: Pituitary homeobox 3

Chain A: 57% 6% • 35%

G-2 S-1 Q1 R2 R3 Q4 R5 T6 H7 F8 R22 N23 R24 R31 R44 R52 R53 A54 K55 W56 R57 K58 R59 E60 E61 F62 I63 V64 T65 D66

- Molecule 2: DNA (5'-D(*GP*CP*TP*CP*TP*AP*AP*TP*CP*CP*CP*CP*G)-3')

Chain B: 8% 92%

G67 C68 T69 C70 T71 A72 A73 T74 C75 C76 C77 C78 G79

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

Chain C: 8% 92%

C80 G81 G82 G83 G84 A85 T86 T87 A88 G89 A90 G91 C92

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

The representative model is number 8. Colouring as in section 4.1 above.

- Molecule 1: Pituitary homeobox 3

Chain A: 57% 6% • 35%

G-2 S-1 Q1 R2 R3 Q4 R5 T6 H7 F8 R22 N23 R24 R31 R44 R52 R53 A54 K55 W56 R57 K58 R59 E60 E61 F62 I63 V64 T65 D66

- Molecule 2: DNA (5'-D(*GP*CP*TP*CP*TP*AP*AP*TP*CP*CP*CP*CP*G)-3')

Chain B: 8% 69% 23%

G67 C68 T69 C70 T71 A72 A73 T74 C75 C76 C77 C78 G79

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

Chain C: 31% 62% 8%

C80 G81 G82 G83 G84 A85 T86 T87 A88 G89 A90 G91 C92

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 30 calculated structures, 17 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CYANA	structure solution	
Amber	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	833
Number of shifts mapped to atoms	832
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	46%

6 Model quality (i)

6.1 Standard geometry (i)

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	0.78±0.01	0±0/388 (0.0± 0.0%)	1.28±0.06	5±1/526 (1.0± 0.2%)
2	B	1.54±0.02	0±0/286 (0.0± 0.0%)	2.36±0.04	23±2/438 (5.2± 0.4%)
3	C	1.57±0.02	0±0/304 (0.0± 0.0%)	2.32±0.07	21±3/469 (4.4± 0.6%)
All	All	1.30	0/16626 (0.0%)	2.02	830/24361 (3.4%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.6±0.6
2	B	0.0±0.0	2.9±1.3
3	C	0.0±0.0	3.0±1.4
All	All	0	112

There are no bond-length outliers.

5 of 111 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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	44	ARG	NE-CZ-NH1	10.70	125.65	120.30	15	15
3	C	90	DA	N1-C6-N6	-10.59	112.25	118.60	2	17
1	A	24	ARG	NE-CZ-NH1	10.55	125.57	120.30	5	15
1	A	31	ARG	NE-CZ-NH1	10.46	125.53	120.30	11	17
3	C	85	DA	N1-C6-N6	-10.44	112.33	118.60	7	17

There are no chirality outliers.

5 of 25 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	69	DT	Sidechain	8
3	C	81	DG	Sidechain	8
3	C	82	DG	Sidechain	8
3	C	83	DG	Sidechain	8
2	B	72	DA	Sidechain	6

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
3	C	270	147	147	0±0
All	All	15419	11288	11288	1

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:C:88:DA:H1'	3:C:89:DG:C8	0.41	2.50	17	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	44/68 (65%)	43±1 (98±2%)	1±1 (2±2%)	0±0 (0±0%)	100	100
All	All	748/1156 (65%)	734 (98%)	14 (2%)	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	41/63 (65%)	39±1 (96±2%)	2±1 (4±2%)	33	82
All	All	697/1071 (65%)	668 (96%)	29 (4%)	33	82

5 of 10 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	52	ARG	12
1	A	21	GLN	4
1	A	50	LYS	3
1	A	41	THR	3
1	A	27	ASP	2

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 (i)

The completeness of assignment taking into account all chemical shift lists is 46% for the well-defined parts and 53% 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 (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	833
Number of shifts mapped to atoms	832
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	14

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

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	7	HIS	HD1	8.744	0.020	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
$^{13}C_{\alpha}$	61	-0.55 ± 0.21	Should be checked
$^{13}C_{\beta}$	59	0.12 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	65	-0.42 ± 0.43	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 46%, i.e. 540 atoms were assigned a chemical shift out of a possible 1174. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	172/218 (79%)	87/87 (100%)	42/88 (48%)	43/43 (100%)
Sidechain	322/385 (84%)	220/245 (90%)	89/114 (78%)	13/26 (50%)
Aromatic	46/53 (87%)	26/26 (100%)	18/25 (72%)	2/2 (100%)
Sugar	0/312 (0%)	0/182 (0%)	0/130 (0%)	0/0 (—%)
Base	0/206 (0%)	0/128 (0%)	0/44 (0%)	0/34 (0%)
Overall	540/1174 (46%)	333/668 (50%)	149/401 (37%)	58/105 (55%)

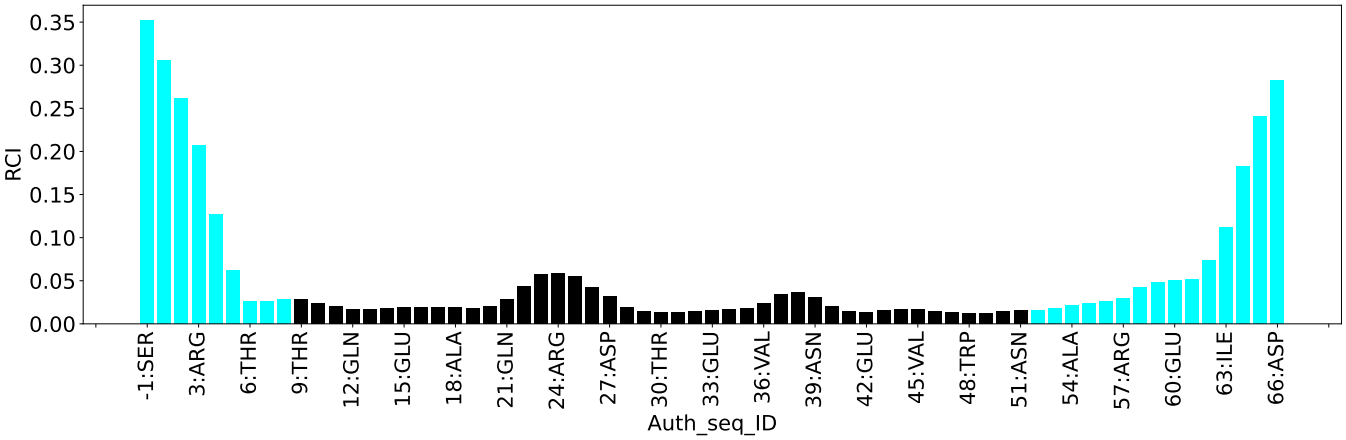
7.1.4 Statistically unusual chemical shifts (i)

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	9	THR	HG1	5.83	0.08 – 2.19	22.2
1	A	19	THR	HG1	5.36	0.08 – 2.19	20.0
1	A	38	THR	HG1	4.08	0.08 – 2.19	14.0
1	A	52	ARG	HB2	-0.51	0.52 – 3.08	-9.0
1	A	52	ARG	HG3	-0.60	0.15 – 2.94	-7.7
1	A	16	LEU	HB3	-1.20	-0.26 – 3.31	-7.7
1	A	52	ARG	HG2	-0.21	0.26 – 2.87	-6.8
1	A	24	ARG	HG3	-0.29	0.15 – 2.94	-6.6
1	A	26	PRO	HG3	-0.12	0.33 – 3.48	-6.4
1	A	24	ARG	HD2	1.71	1.97 – 4.26	-6.2
1	A	24	ARG	HD3	1.63	1.81 – 4.39	-5.7
1	A	16	LEU	HD11	-0.80	-0.61 – 2.12	-5.7
1	A	16	LEU	HD12	-0.80	-0.61 – 2.12	-5.7
1	A	16	LEU	HD13	-0.80	-0.61 – 2.12	-5.7

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: