wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 06:18 PM BST

PDB ID	:	1X2N
Title	:	Solution structure of the homeobox domain of human homeobox protein PKNOX1
Authors	:	Kamatari, Y.O.; Tomizawa, T.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-04-26

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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

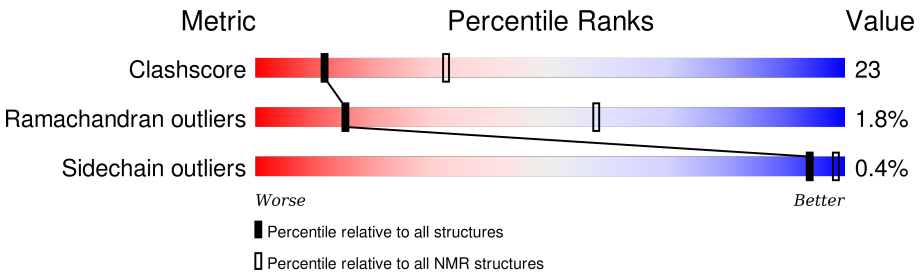
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	73	29% 25% 47%

2 Ensemble composition and analysis (i)

This entry contains 20 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:19-A:57 (39)	0.07	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 6 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1148 atoms, of which 576 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Homeobox protein PKNOX1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	73	1148	356	576	112	103	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P55347
A	2	SER	-	CLONING ARTIFACT	UNP P55347
A	3	SER	-	CLONING ARTIFACT	UNP P55347
A	4	GLY	-	CLONING ARTIFACT	UNP P55347
A	5	SER	-	CLONING ARTIFACT	UNP P55347
A	6	SER	-	CLONING ARTIFACT	UNP P55347
A	7	GLY	-	CLONING ARTIFACT	UNP P55347
A	68	SER	-	CLONING ARTIFACT	UNP P55347
A	69	GLY	-	CLONING ARTIFACT	UNP P55347
A	70	PRO	-	CLONING ARTIFACT	UNP P55347
A	71	SER	-	CLONING ARTIFACT	UNP P55347
A	72	SER	-	CLONING ARTIFACT	UNP P55347
A	73	GLY	-	CLONING ARTIFACT	UNP P55347

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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 PKNOX1

Chain A: 29% 25% 47%

G1 S2 S3 G4 S5 S6 G7 K8 N9 K10 R11 G12 V13 L14 P15 K16 H17 A18 M22 W25 H29 I30 G31 H32 P33 Y34 P35 T36 E37 D38 E39 A44 N48 L49 T50 L51 L52 Q53 W57 F58 I59 N60 A61 R62 R63 R64 I65 L66 Q67 S68 G69 P70 S71 S72

G73

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: Homeobox protein PKNOX1

Chain A: 30% 23% 47%

G1 S2 S3 G4 S5 S6 G7 K8 N9 K10 R11 G12 V13 L14 P15 K16 H17 A18 W25 H29 I30 G31 H32 P33 Y34 P35 T36 E37 D38 E39 K40 A44 N48 L49 T50 L51 L52 F58 I59 N60 A61 R62 R63 R64 I65 L66 Q67 S68 G69 P70 S71 S72 G73

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function, 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	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

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	326	318	316	15±3
All	All	6520	6360	6320	294

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 23.

5 of 48 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:32:HIS:CD2	1:A:34:TYR:CD1	0.81	2.69	20	14
1:A:44:ALA:HB1	1:A:49:LEU:O	0.72	1.85	11	20
1:A:32:HIS:CE1	1:A:34:TYR:CE1	0.66	2.83	11	6
1:A:32:HIS:NE2	1:A:34:TYR:CG	0.60	2.70	17	10
1:A:32:HIS:CE1	1:A:34:TYR:CD1	0.59	2.90	3	4

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	39/73 (53%)	33±1 (84±3%)	6±1 (14±3%)	1±1 (2±1%)	15 58
All	All	780/1460 (53%)	655 (84%)	111 (14%)	14 (2%)	15 58

All 2 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	37	GLU	10
1	A	51	LEU	4

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	36/62 (58%)	36±0 (100±1%)	0±0 (0±1%)	94 98
All	All	720/1240 (58%)	717 (100%)	3 (0%)	94 98

All 2 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	38	ASP	2
1	A	28	GLN	1

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 carbohydrates 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)

No chemical shift data were provided