



wwPDB NMR Structure Validation Summary Report ⓘ

Mar 8, 2026 – 04:01 PM UTC

PDB ID : 1I11 / pdb_00001i11
Title : SOLUTION STRUCTURE OF THE DNA BINDING DOMAIN, SOX-5 HMG
BOX FROM MOUSE
Authors : Cary, P.D.; Read, C.M.; Davis, B.; Driscoll, P.C.; Crane-Robinson, C.
Deposited on : 2001-01-30

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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
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)	:	2.49

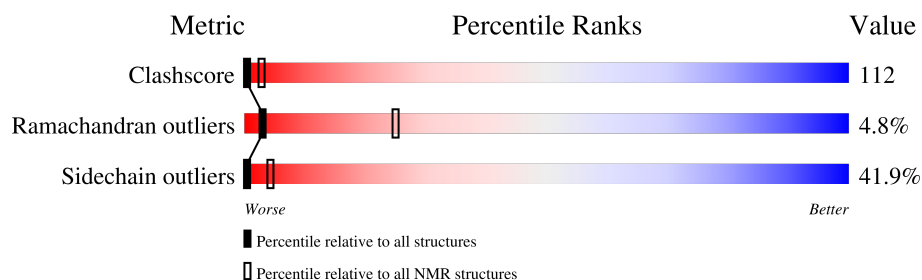
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment 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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	81	

2 Ensemble composition and analysis

This entry contains 30 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 9 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:8-A:56 (49)	0.55	9

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 3 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 20, 22, 24, 25, 27, 28
2	11, 23, 30
3	21, 26, 29
Single-model clusters	19

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1194 atoms, of which 596 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TRANSCRIPTION FACTOR SOX-5.

Mol	Chain	Residues	Atoms						Trace
1	A	70	Total	C	H	N	O	S	0
			1194	382	596	109	103	4	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	SEE REMARK 999	UNP P35710
A	0	SER	-	SEE REMARK 999	UNP P35710

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *SIMULATING ANNEALING*.

Of the 50 calculated structures, 30 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy*.

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

Software name	Classification	Version
X-PLOR	refinement	3.851

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

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	1.39±0.01	0±0/426 (0.0± 0.1%)	1.47±0.03	3±1/573 (0.5± 0.2%)
All	All	1.39	3/12780 (0.0%)	1.47	87/17190 (0.5%)

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	2.8±0.5
All	All	0	84

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	28	MET	C-N	-5.50	1.26	1.33	30	3

5 of 9 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(°)	Ideal(°)	Models	
								Worst	Total
1	A	29	HIS	CA-CB-CG	-6.37	107.43	113.80	27	4
1	A	27	ASP	N-CA-C	-5.74	106.82	113.88	14	19
1	A	40	ARG	N-CA-C	-5.57	105.60	112.90	17	29
1	A	39	SER	N-CA-C	-5.48	106.76	113.50	17	25
1	A	16	ASP	N-CA-C	-5.32	106.92	113.41	2	5

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	40	ARG	Sidechain	29
1	A	19	ARG	Sidechain	28
1	A	18	ARG	Sidechain	27

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	415	408	408	92±11
All	All	12450	12240	12240	2775

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 112.

5 of 538 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:14:ALA:HB1	1:A:37:LEU:HD13	1.12	1.14	1	30
1:A:18:ARG:HG3	1:A:37:LEU:HD11	1.01	1.31	9	8
1:A:14:ALA:HB1	1:A:37:LEU:CD1	0.98	1.88	7	30
1:A:14:ALA:CB	1:A:37:LEU:HD13	0.96	1.90	3	27
1:A:18:ARG:CG	1:A:37:LEU:HD11	0.94	1.92	17	8

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	49/81 (60%)	41±1 (83±3%)	6±2 (12±4%)	2±1 (5±2%)	3	25
All	All	1470/2430 (60%)	1216 (83%)	183 (12%)	71 (5%)	3	25

5 of 7 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	31	SER	23
1	A	28	MET	14
1	A	46	ASN	12
1	A	44	MET	10
1	A	51	PRO	6

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	44/74 (59%)	26±2 (58±5%)	18±2 (42±5%)	0	4
All	All	1320/2220 (59%)	767 (58%)	553 (42%)	0	4

5 of 35 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	13	TRP	30
1	A	17	GLU	30
1	A	19	ARG	30
1	A	21	ILE	30
1	A	34	SER	28

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 oligosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

No chemical shift data were provided