



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 03:48 PM BST

PDB ID : 1O4X
Title : TERNARY COMPLEX OF THE DNA BINDING DOMAINS OF THE OCT1 AND SOX2 TRANSCRIPTION FACTORS WITH A 19MER OLIGONUCLEOTIDE FROM THE HOXB1 REGULATORY ELEMENT
Authors : Clore, G.M.; Williams, D.C.
Deposited on : 2003-07-17

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 ⓘ symbol.

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

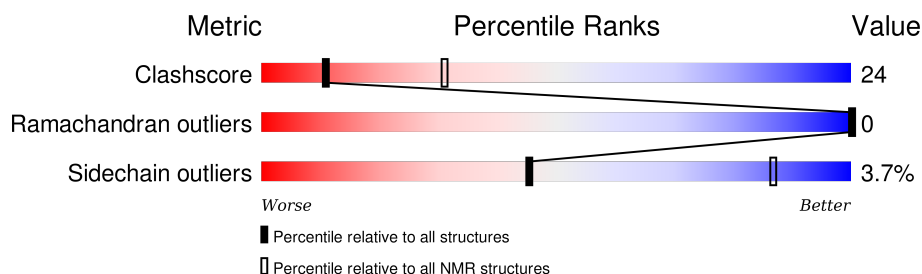
Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Mogul : unknown
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : rb-20027457
Ideal 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

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	C	19	
2	D	19	
3	A	167	
4	B	88	

2 Ensemble composition and analysis ⓘ

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4727 atoms, of which 2225 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called 5'-D(*TP*GP*TP*CP*TP*TP*TP*GP*TP*CP*AP*TP*GP*CP*TP*AP*AP*TP*G)-3'.

Mol	Chain	Residues	Atoms						Trace
1	C	19	Total	C	H	N	O	P	0
			605	187	220	62	118	18	

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

Mol	Chain	Residues	Atoms						Trace
2	D	19	Total	C	H	N	O	P	0
			602	186	214	78	106	18	

- Molecule 3 is a protein called transcription factor Oct-1.

Mol	Chain	Residues	Atoms						Trace
3	A	129	Total	C	H	N	O	S	0
			2107	662	1060	183	197	5	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P14859
A	2	SER	-	CLONING ARTIFACT	UNP P14859
A	3	HIS	-	CLONING ARTIFACT	UNP P14859
A	4	MET	-	CLONING ARTIFACT	UNP P14859
A	65	ALA	CYS	ENGINEERED	UNP P14859

- Molecule 4 is a protein called Transcription factor SOX-2.

Mol	Chain	Residues	Atoms						Trace
4	B	77	Total	C	H	N	O	S	0
			1413	426	731	142	109	5	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	201	GLY	-	CLONING ARTIFACT	UNP P48431

Continued on next page...

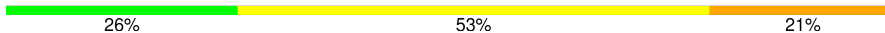
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	202	SER	-	CLONING ARTIFACT	UNP P48431
B	203	HIS	-	CLONING ARTIFACT	UNP P48431
B	204	MET	-	CLONING ARTIFACT	UNP P48431
B	205	PRO	-	CLONING ARTIFACT	UNP P48431

4 Residue-property plots [i](#)

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: 5'-D(*TP*GP*TP*CP*TP*TP*TP*GP*TP*CP*AP*TP*GP*CP*TP*AP*AP*TP*P*G)-3'

Chain C: 



- Molecule 2: 5'-D(*CP*AP*TP*TP*AP*GP*CP*AP*TP*GP*AP*CP*AP*AP*AP*GP*AP*C*P*A)-3'

Chain D: 



- Molecule 3: transcription factor Oct-1

Chain A: 



- Molecule 4: Transcription factor SOX-2

Chain B: 



5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *CONJOINED RIGID BODY/TORSION ANGLE DYNAMICS*.

Of the 100 calculated structures, 1 were deposited, based on the following criterion: *REGULARIZED MEAN STRUCTURE*.

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

Software name	Classification	Version
X-PLOR NIH	refinement	(HTTP://NMR.CIT.NIH.GOV/XPLOR_NIH)

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

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	C	1.31	6/429 (1.4%)	1.71	9/661 (1.4%)
2	D	1.17	3/437 (0.7%)	1.74	11/672 (1.6%)
3	A	0.85	0/1071 (0.0%)	0.94	1/1436 (0.1%)
4	B	1.01	0/717 (0.0%)	0.98	1/949 (0.1%)
All	All	1.03	9/2654 (0.3%)	1.29	22/3718 (0.6%)

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
3	A	0	1
All	All	0	1

5 of 9 bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	306	DT	C5-C7	5.82	1.53	1.50
1	C	318	DT	C5-C7	5.68	1.53	1.50
2	D	328	DT	C5-C7	5.54	1.53	1.50
2	D	323	DT	C5-C7	5.44	1.53	1.50
1	C	305	DT	C5-C7	5.42	1.53	1.50

5 of 22 angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	322	DT	C6-C5-C7	-6.36	119.08	122.90
2	D	336	DA	N1-C2-N3	-6.35	126.13	129.30
2	D	332	DA	N1-C2-N3	-6.31	126.14	129.30
2	D	321	DA	N1-C2-N3	-6.15	126.22	129.30
2	D	324	DA	N1-C2-N3	-6.14	126.23	129.30

There are no chirality outliers.

All planar outliers are listed below.

Mol	Chain	Res	Type	Group
3	A	155	ARG	Sidechain

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	C	385	220	220	15
2	D	388	214	214	3
3	A	1047	1060	1044	63
4	B	682	731	720	46
All	All	2502	2225	2198	113

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 24.

5 of 113 clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
3:A:31:GLN:NE2	3:A:48:GLN:O	1.29	1.62
4:B:233:MET:CE	4:B:241:ARG:CZ	1.04	2.35
4:B:224:ARG:HG2	4:B:224:ARG:HH11	0.98	1.17
4:B:233:MET:HE3	4:B:241:ARG:CZ	0.95	1.91
4:B:233:MET:CE	4:B:241:ARG:NH1	0.91	2.33

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	
3	A	127/167 (76%)	123 (97%)	4 (3%)	0 (0%)	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	B	79/88 (90%)	79 (100%)	0 (0%)	0 (0%)	100	100
All	All	206/255 (81%)	202 (98%)	4 (2%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 Protein sidechains ⓘ

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	
3	A	114/151 (75%)	110 (96%)	4 (4%)	47	88
4	B	74/80 (92%)	71 (96%)	3 (4%)	42	84
All	All	188/231 (81%)	181 (96%)	7 (4%)	45	87

5 of 7 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
4	B	281	ARG
3	A	149	ARG
4	B	224	ARG
3	A	79	GLU
3	A	140	GLN

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

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

No chemical shift data were provided