Full wwPDB NMR Structure Validation Report (i)

Aug 20, 2022 – 08:12 AM EDT

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 Full wwPDB NMR Structure Validation 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)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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.

There are no overall percentile quality scores available for this entry.

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	C	19	100%
2	D	19	100%
3	A	167	77% 23%
4	B	88	88% 13%

2 Ensemble composition and analysis (i)

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
			Total	C	H	N	O	P	
1	C	19	605	187	220	62	118	18	0

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

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

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

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

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 mutation	UNP P14859

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

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

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...

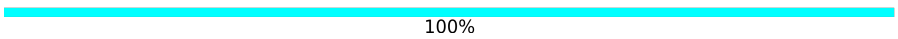
Chain	Residue	Modelled	Actual	Comment	Reference
B	202	SER	-	cloning artifact	UNP P48431
B	203	HIS	-	cloning artifact	UNP P48431
B	204	MET	-	cloning artifact	UNP P48431
B	205	PRO	-	cloning artifact	UNP P48431

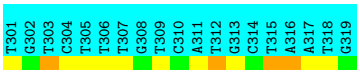
4 Residue-property plots (i)

These plots are provided for all protein, RNA, 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: 5'-D(*TP*GP*TP*CP*TP*TP*TP*GP*TP*CP*AP*TP*GP*CP*TP*AP*AP*TP*G)-3'

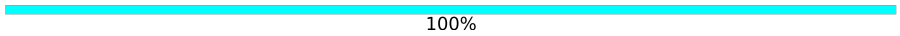
Chain C:

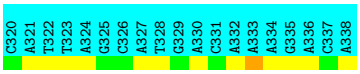




• Molecule 2: 5'-D(*CP*AP*TP*TP*AP*GP*CP*AP*TP*GP*AP*CP*AP*AP*AP*GP*AP*CP*A)-3'

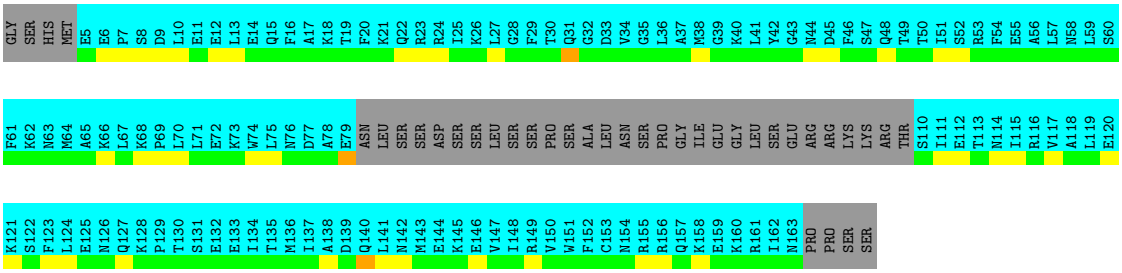
Chain D:





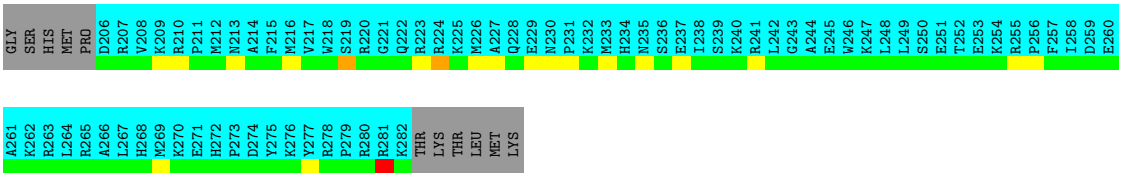
• Molecule 3: transcription factor Oct-1

Chain A:



• Molecule 4: Transcription factor SOX-2

Chain B:



5 Refinement protocol and experimental data overview (i)

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.

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	C	0	0	0	0
2	D	0	0	0	0
3	A	-26	-38	0	0
4	B	-64	-118	0	0
All	All	-90	-156	0	-

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 -.

There are no clashes.

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	0	-	-	-	-
4	B	0	-	-	-	-
All	All	0	-	-	-	-

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
3	A	0	-	-	-
4	B	0	-	-	-
All	All	0	-	-	-

There are no protein residues with a non-rotameric sidechain to report.

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)

No chemical shift data were provided