



Full wwPDB X-ray Structure Validation Report i

Mar 1, 2014 – 04:05 AM GMT

PDB ID : 4A3N
Title : Crystal Structure of HMG-BOX of Human SOX17
Authors : Gao, N.; Gao, H.; Qian, H.; Si, S.; Xie, Y.
Deposited on : 2011-10-01
Resolution : 2.40 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

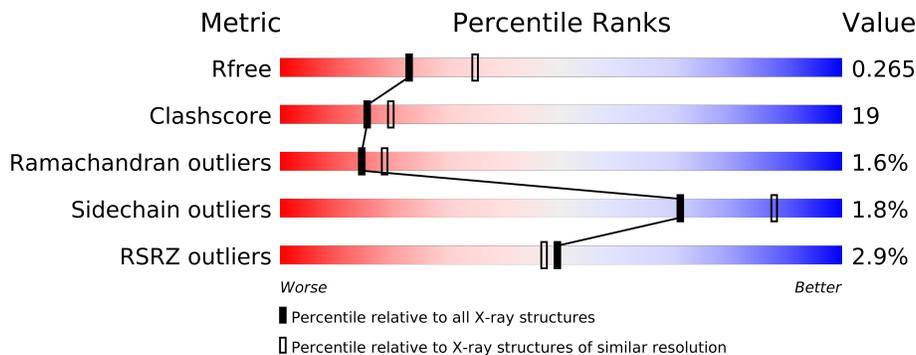
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.40 Å.

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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	2207 (2.40-2.40)
Clashscore	79885	2789 (2.40-2.40)
Ramachandran outliers	78287	2736 (2.40-2.40)
Sidechain outliers	78261	2737 (2.40-2.40)
RSRZ outliers	66119	2210 (2.40-2.40)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	71	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 609 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called TRANSCRIPTION FACTOR SOX-17.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	65	534	334	104	92	4	0	0	1

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	66	GLY	-	EXPRESSION TAG	UNP Q9H6I2
A	67	SER	-	EXPRESSION TAG	UNP Q9H6I2

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
2	A	3	3	3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	72	72	72	0	0

3 Residue-property plots i

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: TRANSCRIPTION FACTOR SOX-17

Chain A: 



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	94.28Å 22.52Å 66.67Å 90.00° 129.09° 90.00°	Depositor
Resolution (Å)	23.66 – 2.40 23.66 – 2.30	Depositor EDS
% Data completeness (in resolution range)	96.1 (23.66-2.40) 98.2 (23.66-2.30)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	8.40 (at 2.31Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.202 , 0.267 0.202 , 0.265	Depositor DCC
R_{free} test set	453 reflections (11.72%)	DCC
Wilson B-factor (Å ²)	27.7	Xtrriage
Anisotropy	0.700	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 56.1	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Outliers	0 of 4992 reflections	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	609	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.14% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section:
ZN

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 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	0.42	0/546	0.53	0/733

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	534	0	535	20	1
2	A	3	0	0	0	0
3	A	72	0	0	3	0
All	All	609	0	535	20	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 19.

All (20) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:89:GLN:C	1:A:90:ASN:HD22	1.72	0.93

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:106:TRP:CZ2	1:A:114:LYS:HD2	2.33	0.63
1:A:106:TRP:CH2	1:A:114:LYS:HD2	2.38	0.58
1:A:78:TRP:CZ2	1:A:82:GLU:HG3	2.39	0.58
1:A:89:GLN:C	1:A:90:ASN:ND2	2.52	0.57
1:A:72:MET:HB3	1:A:76:MET:HB2	1.90	0.54
1:A:70:ARG:N	1:A:71:PRO:CD	2.71	0.54
1:A:92:ASP:OD1	1:A:92:ASP:N	2.35	0.52
1:A:120:GLU:CD	3:A:2061:HOH:O	2.48	0.52
1:A:106:TRP:CE2	1:A:114:LYS:HD2	2.45	0.51
1:A:93:LEU:HD23	1:A:94:HIS:O	2.11	0.50
1:A:90:ASN:N	1:A:90:ASN:HD22	2.01	0.50
1:A:127:GLN:HG3	3:A:2047:HOH:O	2.12	0.49
1:A:90:ASN:N	1:A:90:ASN:ND2	2.61	0.48
1:A:93:LEU:HD23	1:A:93:LEU:C	2.35	0.48
1:A:97:GLU:O	1:A:101:MET:HG3	2.15	0.47
1:A:70:ARG:N	1:A:71:PRO:HD3	2.33	0.44
1:A:97:GLU:HB2	3:A:2022:HOH:O	2.20	0.41
1:A:132:HIS:N	1:A:133:PRO:HD3	2.37	0.40
1:A:118:VAL:O	1:A:122:GLU:HG3	2.21	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:90:ASN:OD1	1:A:92:ASP:OD1[2_555]	2.18	0.02

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. 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	63/71 (89%)	60 (95%)	2 (3%)	1 (2%)	14 18

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	133	PRO

5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution. 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	56/62 (90%)	55 (98%)	1 (2%)	71 88

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	81	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	90	ASN

5.3.3 RNA [i](#)

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	65/71 (91%)	-0.15	2 (3%) 47 44	19, 41, 69, 111	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	70	ARG	5.1
1	A	88	GLN	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å ²)	Q<0.9
2	ZN	A	1136	1/1	0.12	0.20	50,50,50,50	0
2	ZN	A	1135	1/1	0.07	-0.94	39,39,39,39	0
2	ZN	A	1137	1/1	0.09	-2.24	88,88,88,88	0

6.5 Other polymers

There are no such residues in this entry.