WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 07:49 PM GMT

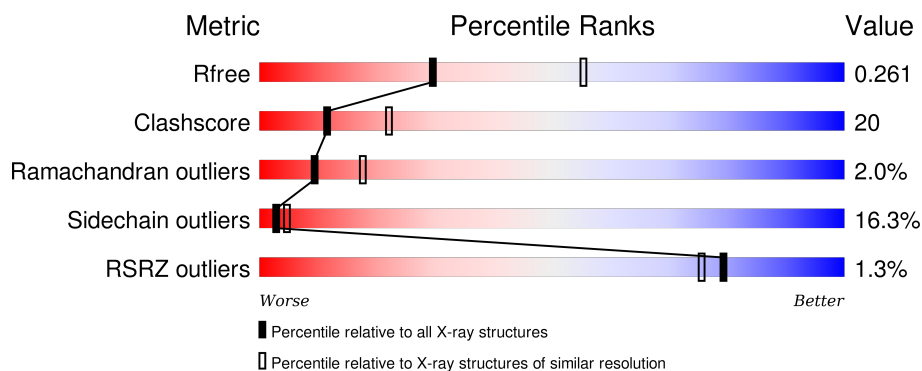|		
PDB ID	:	4Q2J
Title	:	A novel structure-based mechanism for DNA-binding of SATB1
Authors	:	Wang, Z.; Long, J.; Yang, X.; Shen, Y.
Deposited on	:	2014-04-08
Resolution	:	2.60 Å(reported)

1   Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	Percentile Ranks	Value
Rfree		0.261
Clashscore		20
Ramachandran outliers		2.0%
Sidechain outliers		16.3%
RSRZ outliers		1.3%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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. 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	180	49% 36% 9% 6%
1	B	180	35% 14% · 48%
1	C	180	32% 15% · · 49%
1	D	180	53% 33% 7% 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4034 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 DNA-binding protein SATB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	169	Total	C	N	O	S	0	0	0
			1316	839	222	245	10			
1	B	94	Total	C	N	O	S	0	0	0
			720	464	117	134	5			
1	C	92	Total	C	N	O	S	0	0	0
			702	454	115	128	5			
1	D	169	Total	C	N	O	S	0	0	0
			1283	818	215	240	10			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	67	GLY	-	EXPRESSION TAG	UNP Q60611
A	68	PRO	-	EXPRESSION TAG	UNP Q60611
A	69	GLY	-	EXPRESSION TAG	UNP Q60611
A	70	SER	-	EXPRESSION TAG	UNP Q60611
B	67	GLY	-	EXPRESSION TAG	UNP Q60611
B	68	PRO	-	EXPRESSION TAG	UNP Q60611
B	69	GLY	-	EXPRESSION TAG	UNP Q60611
B	70	SER	-	EXPRESSION TAG	UNP Q60611
C	67	GLY	-	EXPRESSION TAG	UNP Q60611
C	68	PRO	-	EXPRESSION TAG	UNP Q60611
C	69	GLY	-	EXPRESSION TAG	UNP Q60611
C	70	SER	-	EXPRESSION TAG	UNP Q60611
D	67	GLY	-	EXPRESSION TAG	UNP Q60611
D	68	PRO	-	EXPRESSION TAG	UNP Q60611
D	69	GLY	-	EXPRESSION TAG	UNP Q60611
D	70	SER	-	EXPRESSION TAG	UNP Q60611

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	4	Total 4	O 4	0	0
2	B	4	Total 4	O 4	0	0
2	C	1	Total 1	O 1	0	0
2	D	4	Total 4	O 4	0	0

3 Residue-property plots ⓘ

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: DNA-binding protein SATB1

- Molecule 1: DNA-binding protein SATB1

- Molecule 1: DNA-binding protein SATB1
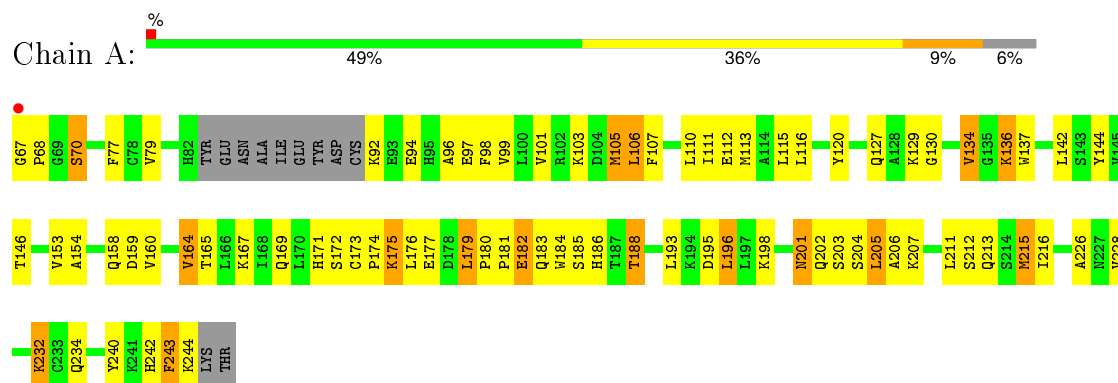
• Molecule 1: DNA-binding protein SATB1
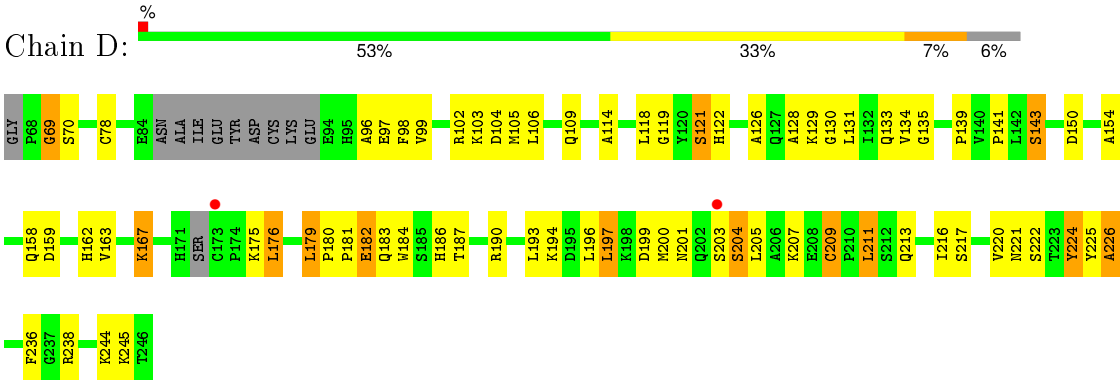
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	128.98Å 91.97Å 100.24Å 90.00° 128.97° 90.00°	Depositor
Resolution (Å)	42.47 – 2.60 42.47 – 2.60	Depositor EDS
% Data completeness (in resolution range)	90.8 (42.47-2.60) 90.6 (42.47-2.60)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	0.08	Depositor
$<I/\sigma(I)>$ [1]	2.37 (at 2.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, R_{free}	0.228 , 0.271 0.220 , 0.261	Depositor DCC
R_{free} test set	1265 reflections (4.97%)	DCC
Wilson B-factor (Å^2)	49.7	Xtriage
Anisotropy	0.830	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 46.6	EDS
Estimated twinning fraction	0.015 for h,-k,-h-l	Xtriage
L-test for twinning [2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Outliers	0 of 27707 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	4034	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.72% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.
[2] Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 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 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	0.48	0/1346	0.63	0/1826
1	B	0.55	0/734	0.71	1/998 (0.1%)
1	C	0.49	0/716	0.65	0/974
1	D	0.45	0/1308	0.60	0/1775
All	All	0.49	0/4104	0.64	1/5573 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	68	PRO	N-CA-CB	6.20	110.75	103.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

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 hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1316	0	1306	61	0
1	B	720	0	716	26	0
1	C	702	0	704	33	0
1	D	1283	0	1264	49	0
2	A	4	0	0	0	0
2	B	4	0	0	1	0
2	C	1	0	0	0	0
2	D	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4034	0	3990	163	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 20.

The worst 5 of 163 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:203:SER:HA	1:A:213:GLN:HE21	1.22	1.04
1:A:243:PHE:HA	1:A:244:LYS:C	1.86	0.94
1:A:212:SER:O	1:A:216:ILE:HG13	1.74	0.87
1:C:105:MET:SD	1:C:113:MET:HG3	2.19	0.82
1:A:154:ALA:O	1:A:158:GLN:HB2	1.78	0.82

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 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 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	165/180 (92%)	150 (91%)	12 (7%)	3 (2%)	11	21
1	B	90/180 (50%)	82 (91%)	7 (8%)	1 (1%)	17	36
1	C	88/180 (49%)	79 (90%)	8 (9%)	1 (1%)	17	36
1	D	163/180 (91%)	146 (90%)	12 (7%)	5 (3%)	5	8
All	All	506/720 (70%)	457 (90%)	39 (8%)	10 (2%)	9	18

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	95	HIS
1	A	176	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	242	HIS
1	B	142	LEU
1	D	69	GLY

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	147/159 (92%)	117 (80%)	30 (20%)	1	2
1	B	79/159 (50%)	72 (91%)	7 (9%)	12	23
1	C	77/159 (48%)	66 (86%)	11 (14%)	4	7
1	D	140/159 (88%)	116 (83%)	24 (17%)	2	4
All	All	443/636 (70%)	371 (84%)	72 (16%)	3	5

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	105	MET
1	C	99	VAL
1	D	203	SER
1	B	106	LEU
1	B	167	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	133	GLN
1	B	169	GLN
1	D	169	GLN
1	B	127	GLN
1	D	186	HIS

5.3.3 RNA (i)

There are no RNA molecules 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)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

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(Å²)	Q<0.9
1	A	169/180 (93%)	-0.11	1 (0%)	90	88	39, 67, 89, 103	0
1	B	94/180 (52%)	-0.09	2 (2%)	67	61	40, 58, 94, 109	0
1	C	92/180 (51%)	-0.15	2 (2%)	65	59	53, 67, 87, 95	0
1	D	169/180 (93%)	-0.08	2 (1%)	81	77	47, 72, 95, 106	0
All	All	524/720 (72%)	-0.10	7 (1%)	79	75	39, 67, 92, 109	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	173	CYS	2.9
1	A	67	GLY	2.8
1	B	93	GLU	2.7
1	D	203	SER	2.6
1	B	94	GLU	2.5

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.