



Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 03:02 PM GMT

PDB ID : 4FGN
Title : Crystal structure of the SV40 large T-antigen origin binding domain bound to Site I DNA
Authors : Meinke, G.; Bohm, A.; Bullock, P.A.
Deposited on : 2012-06-04
Resolution : 3.20 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>

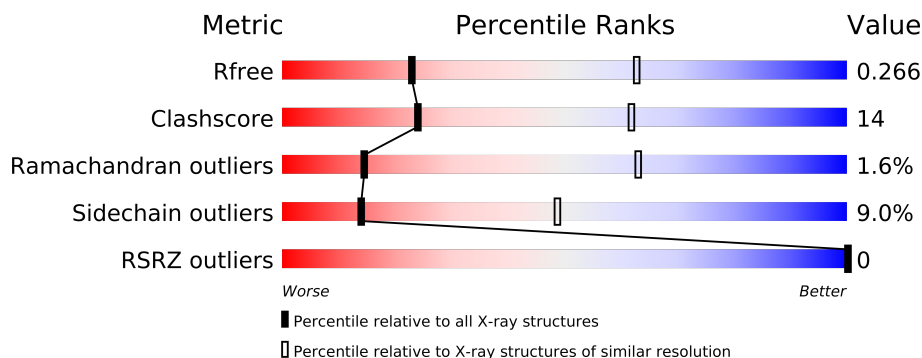
The following versions of software and data (see [references](#)) were used in the production of this report:

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 3.20 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	1824 (3.30-3.10)
Clashscore	79885	1078 (3.26-3.14)
Ramachandran outliers	78287	1059 (3.26-3.14)
Sidechain outliers	78261	1058 (3.26-3.14)
RSRZ outliers	66119	1825 (3.30-3.10)

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	132	<div><div></div><div><div></div><div></div><div></div><div></div></div></div>
1	B	132	<div><div></div><div><div></div><div></div><div></div><div></div></div></div>
2	Z	23	<div><div></div><div><div></div><div></div><div></div><div></div></div></div>
3	Y	23	<div><div></div><div><div></div><div></div><div></div><div></div></div></div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2967 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 Large T antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	126	Total	C	N	O	S	0	0	0
			1028	667	171	185	5			
1	B	124	Total	C	N	O	S	0	0	0
			999	649	165	180	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	129	GLY	-	EXPRESSION TAG	UNP P03070
A	130	SER	-	EXPRESSION TAG	UNP P03070
B	129	GLY	-	EXPRESSION TAG	UNP P03070
B	130	SER	-	EXPRESSION TAG	UNP P03070

- Molecule 2 is a DNA chain called Site I DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Z	23	Total	C	N	O	P	0	0	0
			462	221	88	131	22			

- Molecule 3 is a DNA chain called Site I DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Y	23	Total	C	N	O	P	0	0	0
			475	227	85	141	22			

- Molecule 4 is water.

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

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.

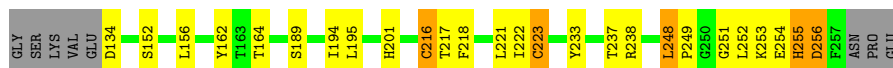
- Molecule 1: Large T antigen

Chain A: 



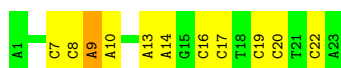
- Molecule 1: Large T antigen

Chain B: 



- Molecule 2: Site I DNA

Chain Z: 



- Molecule 3: Site I DNA

Chain Y: 



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	58.60Å 139.67Å 132.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.99 – 3.20 47.99 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.99-3.20) 99.9 (47.99-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.36 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.188 , 0.267 0.186 , 0.266	Depositor DCC
R_{free} test set	444 reflections (5.02%)	DCC
Wilson B-factor (Å ²)	83.3	Xtriage
Anisotropy	0.645	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 75.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	2 of 9285 reflections (0.022%)	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	2967	wwPDB-VP
Average B, all atoms (Å ²)	138.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.71% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.74	0/1056	0.84	1/1427 (0.1%)
1	B	0.64	0/1027	0.79	1/1392 (0.1%)
2	Z	0.48	0/518	1.10	5/795 (0.6%)
3	Y	0.63	0/532	1.16	3/822 (0.4%)
All	All	0.65	0/3133	0.94	10/4436 (0.2%)

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 outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	Z	14	DA	P-O3'-C3'	8.11	129.43	119.70
3	Y	9	DT	P-O3'-C3'	7.64	128.87	119.70
2	Z	16	DC	P-O3'-C3'	-6.99	111.32	119.70
3	Y	2	DA	P-O3'-C3'	-6.75	111.60	119.70
2	Z	9	DA	P-O3'-C3'	6.52	127.53	119.70
2	Z	13	DA	P-O3'-C3'	6.25	127.20	119.70
1	A	171	LEU	CA-CB-CG	-5.61	102.41	115.30
1	B	221	LEU	CA-CB-CG	5.51	127.97	115.30
3	Y	7	DG	P-O3'-C3'	5.43	126.22	119.70
2	Z	17	DC	P-O3'-C3'	5.34	126.11	119.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	257	PHE	Peptide
1	B	217	THR	Peptide

5.2 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 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	1028	0	0	11	0
1	B	999	0	0	11	0
2	Z	462	0	0	11	0
3	Y	475	0	0	11	0
4	A	2	0	0	0	0
4	B	1	0	0	0	0
All	All	2967	0	0	41	0

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 14.

All (41) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:216:CYS:SG	1:B:218:PHE:CZ	2.66	0.89
2:Z:8:DC:C2'	2:Z:9:DA:N7	2.35	0.88
3:Y:21:DC:C2'	3:Y:22:DT:OP2	2.32	0.78
1:B:216:CYS:SG	1:B:218:PHE:CE1	2.78	0.77
3:Y:22:DT:C2'	3:Y:23:DT:OP2	2.40	0.70
1:B:194:ILE:CG2	1:B:195:LEU:N	2.54	0.70
2:Z:19:DC:C2'	2:Z:20:DC:C6	2.76	0.69
1:A:250:GLY:CA	1:A:254:GLU:OE2	2.41	0.69
3:Y:20:DC:C2'	3:Y:21:DC:C5'	2.71	0.68
1:A:137:ASP:OD1	1:A:138:PHE:N	2.31	0.63
1:B:233:TYR:CE2	1:B:252:LEU:O	2.51	0.63
3:Y:2:DA:C5'	3:Y:2:DA:C8	2.83	0.61
2:Z:19:DC:C4	2:Z:20:DC:N4	2.70	0.59
2:Z:8:DC:C4'	2:Z:9:DA:OP1	2.50	0.58
2:Z:22:DC:C5	3:Y:1:DG:N2	2.71	0.58

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:254:GLU:O	1:B:256:ASP:N	2.38	0.56
2:Z:19:DC:C2'	2:Z:20:DC:C5	2.91	0.54
2:Z:22:DC:C5	3:Y:1:DG:N1	2.75	0.53
2:Z:8:DC:C2'	2:Z:9:DA:C8	2.92	0.53
2:Z:19:DC:C5	2:Z:20:DC:N4	2.81	0.49
1:A:171:LEU:O	1:A:172:TYR:C	2.51	0.47
1:A:202:ARG:O	1:A:203:HIS:CG	2.68	0.47
1:A:188:ASN:OD1	1:A:244:ILE:CD1	2.65	0.45
1:B:222:ILE:CG2	1:B:223:CYS:N	2.79	0.45
1:B:248:LEU:O	1:B:249:PRO:C	2.55	0.45
1:A:187:HIS:N	1:A:187:HIS:CD2	2.84	0.44
1:A:185:SER:OG	1:A:187:HIS:NE2	2.51	0.44
1:B:254:GLU:O	1:B:255:HIS:C	2.56	0.43
2:Z:9:DA:C2'	2:Z:10:DA:C8	3.01	0.43
3:Y:19:DG:C4	3:Y:20:DC:C5	3.06	0.43
3:Y:19:DG:C4	3:Y:20:DC:C6	3.06	0.43
1:A:172:TYR:CD2	1:A:172:TYR:C	2.92	0.43
1:A:141:GLU:OE1	1:A:141:GLU:N	2.51	0.43
1:B:248:LEU:O	1:B:251:GLY:N	2.51	0.42
1:B:152:SER:OG	3:Y:5:DA:N7	2.53	0.42
1:A:232:MET:O	1:A:233:TYR:C	2.57	0.41
3:Y:18:DG:C2	3:Y:19:DG:C5	3.08	0.41
3:Y:19:DG:N3	3:Y:20:DC:C6	2.88	0.41
1:A:239:ASP:CA	1:A:241:PHE:N	2.83	0.41
1:B:189:SER:O	1:B:189:SER:OG	2.37	0.41
2:Z:7:DC:C5	2:Z:8:DC:N4	2.89	0.40

There are no symmetry-related clashes.

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	124/132 (94%)	109 (88%)	13 (10%)	2 (2%)	14	63
1	B	122/132 (92%)	84 (69%)	36 (30%)	2 (2%)	14	63

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	246/264 (93%)	193 (78%)	49 (20%)	4 (2%)	14	63

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	140	SER
1	A	247	SER
1	B	255	HIS
1	B	253	LYS

5.3.2 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 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	114/120 (95%)	105 (92%)	9 (8%)	18	58
1	B	109/120 (91%)	98 (90%)	11 (10%)	11	41
All	All	223/240 (93%)	203 (91%)	20 (9%)	14	49

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	133	GLU
1	A	140	SER
1	A	144	SER
1	A	155	THR
1	A	162	TYR
1	A	170	LEU
1	A	181	VAL
1	A	223	CYS
1	A	239	ASP
1	B	134	ASP
1	B	156	LEU
1	B	162	TYR
1	B	164	THR
1	B	201	HIS
1	B	216	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	223	CYS
1	B	237	THR
1	B	238	ARG
1	B	248	LEU
1	B	256	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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	126/132 (95%)	-0.17	0 100 100	81, 115, 167, 216	0
1	B	124/132 (93%)	-0.22	0 100 100	96, 161, 217, 246	0
2	Z	23/23 (100%)	-0.67	0 100 100	95, 135, 154, 159	0
3	Y	23/23 (100%)	-0.59	0 100 100	98, 114, 168, 175	0
All	All	296/310 (95%)	-0.26	0 100 100	81, 132, 207, 246	0

There are no RSRZ outliers to report.

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.