Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 07:54 PM GMT

PDB ID	:	4QBQ
Title	:	Crystal structure of DNMT3a ADD domain bound to H3 peptide
Authors	:	Li, H.; Patel, D.J.
Deposited on	:	2014-05-08
Resolution	:	2.41 Å(reported)

This is a Full wwPDB X-ray Structure 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/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

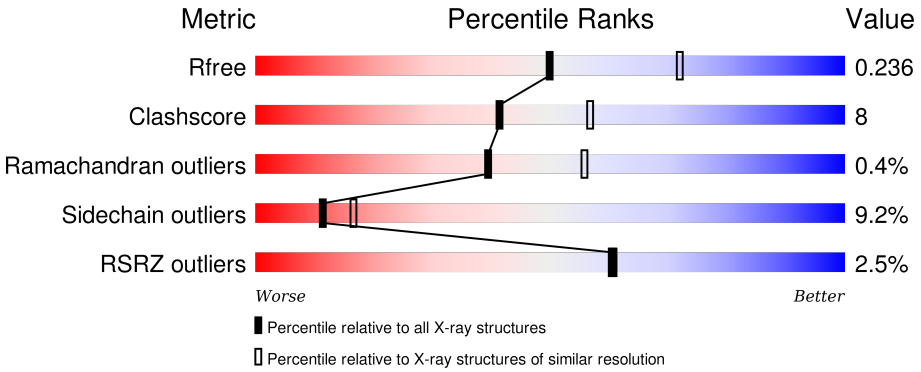
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.41 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}	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (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. 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	137	77% 21%
1	C	137	2% 74% 21%
2	P	8	25% 63% 38%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2299 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 (cytosine-5)-methyltransferase 3A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	137	Total	C	N	O	S	0	0	0
			1062	658	185	197	22			
1	C	137	Total	C	N	O	S	0	0	0
			1062	658	185	197	22			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	474	GLY	-	EXPRESSION TAG	UNP Q9Y6K1
A	475	PRO	-	EXPRESSION TAG	UNP Q9Y6K1
A	476	LEU	-	EXPRESSION TAG	UNP Q9Y6K1
A	477	GLY	-	EXPRESSION TAG	UNP Q9Y6K1
A	478	SER	-	EXPRESSION TAG	UNP Q9Y6K1
C	474	GLY	-	EXPRESSION TAG	UNP Q9Y6K1
C	475	PRO	-	EXPRESSION TAG	UNP Q9Y6K1
C	476	LEU	-	EXPRESSION TAG	UNP Q9Y6K1
C	477	GLY	-	EXPRESSION TAG	UNP Q9Y6K1
C	478	SER	-	EXPRESSION TAG	UNP Q9Y6K1

- Molecule 2 is a protein called Histone H3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	P	8	Total	C	N	O	0	0	0
			65	37	16	12			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 4 is water.

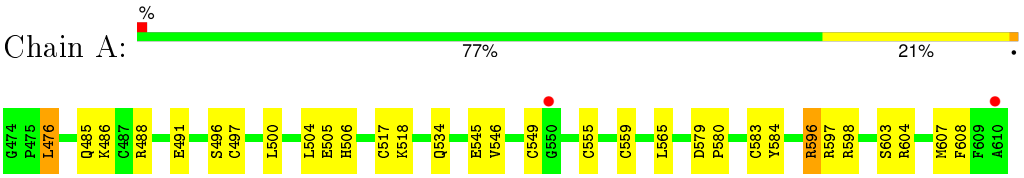
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	53	Total O 53 53	0	0
4	C	49	Total O 49 49	0	0
4	P	2	Total O 2 2	0	0

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.

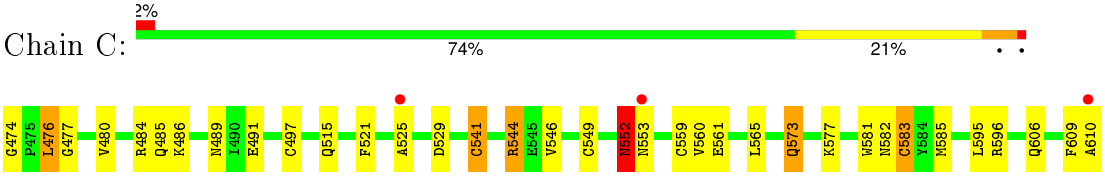
- Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain A:



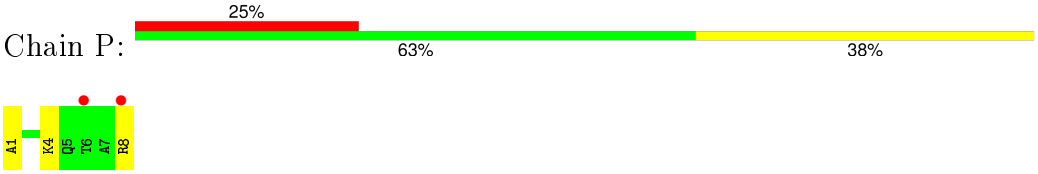
- Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain C:



- Molecule 2: Histone H3

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	41.56Å 56.42Å 57.30Å 90.00° 90.27° 90.00°	Depositor
Resolution (Å)	41.56 – 2.41 41.56 – 2.41	Depositor EDS
% Data completeness (in resolution range)	99.1 (41.56-2.41) 99.2 (41.56-2.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	2.09 (at 2.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.174 , 0.230 0.180 , 0.236	Depositor DCC
R_{free} test set	498 reflections (5.06%)	DCC
Wilson B-factor (Å^2)	28.1	Xtriage
Anisotropy	0.465	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 22.9	EDS
Estimated twinning fraction	0.004 for -h,l,k 0.037 for -h,-l,-k 0.157 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 10337 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2299	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.68% 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

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	0.51	0/1084	0.61	0/1461
1	C	0.50	0/1084	0.63	0/1461
2	P	0.39	0/64	0.61	0/82
All	All	0.50	0/2232	0.62	0/3004

There are no bond length outliers.

There are no bond angle outliers.

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	1062	0	995	13	0
1	C	1062	0	993	23	0
2	P	65	0	73	2	0
3	A	3	0	0	0	0
3	C	3	0	0	0	0
4	A	53	0	0	2	0
4	C	49	0	0	3	0
4	P	2	0	0	0	0
All	All	2299	0	2061	36	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 8.

All (36) 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:596:ARG:NH2	4:A:838:HOH:O	2.14	0.78
1:C:484:ARG:NH1	4:C:843:HOH:O	2.20	0.75
1:A:505:GLU:OE1	1:A:604:ARG:NH2	2.19	0.74
1:C:525:ALA:HA	4:C:838:HOH:O	1.88	0.73
1:C:549:CYS:HB2	1:C:583:CYS:HB3	1.74	0.69
1:C:474:GLY:N	1:C:497:CYS:O	2.34	0.60
1:C:544:ARG:NH2	1:C:560:VAL:HG22	2.16	0.59
1:C:546:VAL:HA	1:C:559:CYS:HA	1.86	0.58
1:C:544:ARG:HH21	1:C:560:VAL:HG22	1.69	0.57
1:A:579:ASP:OD1	1:A:580:PRO:HA	2.06	0.55
1:C:544:ARG:NH1	1:C:561:GLU:HB2	2.24	0.53
1:A:534:GLN:HG3	1:A:555:CYS:O	2.09	0.52
1:A:518:LYS:HG3	1:A:608:PHE:CZ	2.44	0.52
1:C:529:ASP:OD2	2:P:4:LYS:NZ	2.39	0.51
1:C:477:GLY:HA2	1:C:565:LEU:HD11	1.96	0.48
1:A:546:VAL:HA	1:A:559:CYS:HA	1.96	0.47
1:A:485:GLN:HA	1:A:485:GLN:HE21	1.81	0.46
1:A:549:CYS:HB2	1:A:583:CYS:HB3	1.98	0.45
1:C:596:ARG:HD2	4:C:846:HOH:O	2.16	0.45
1:C:581:TRP:CB	2:P:1:ALA:HB2	2.47	0.45
1:C:476:LEU:HD13	1:C:565:LEU:HD23	1.99	0.44
1:A:584:TYR:O	1:A:597:ARG:NH1	2.47	0.43
1:A:476:LEU:HD13	1:A:565:LEU:CD2	2.49	0.43
1:C:489:ASN:HB3	1:C:491:GLU:OE2	2.17	0.43
1:A:506:HIS:O	1:A:598:ARG:NE	2.49	0.43
1:C:544:ARG:H	1:C:544:ARG:CZ	2.32	0.42
4:A:829:HOH:O	1:C:515:GLN:HG2	2.20	0.42
1:C:552:ASN:OD1	1:C:552:ASN:N	2.53	0.42
1:A:496:SER:HB2	1:A:517:CYS:SG	2.60	0.41
1:C:485:GLN:O	1:C:486:LYS:HB2	2.21	0.41
1:C:476:LEU:HD22	1:C:480:VAL:HG23	2.03	0.41
1:C:573:GLN:OE1	1:C:577:LYS:NZ	2.52	0.41
1:C:521:PHE:CE1	1:C:541:CYS:HB2	2.55	0.41
1:A:579:ASP:HA	1:A:580:PRO:HA	1.91	0.41
1:C:552:ASN:HB2	1:C:553:ASN:H	1.73	0.41
1:C:609:PHE:O	1:C:610:ALA:HB2	2.21	0.40

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	135/137 (98%)	128 (95%)	7 (5%)	0	100 100
1	C	135/137 (98%)	125 (93%)	9 (7%)	1 (1%)	26 38
2	P	6/8 (75%)	5 (83%)	1 (17%)	0	100 100
All	All	276/282 (98%)	258 (94%)	17 (6%)	1 (0%)	39 56

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	552	ASN

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	117/117 (100%)	106 (91%)	11 (9%)	11 16
1	C	117/117 (100%)	107 (92%)	10 (8%)	13 20
2	P	6/6 (100%)	5 (83%)	1 (17%)	3 3
All	All	240/240 (100%)	218 (91%)	22 (9%)	11 16

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	476	LEU
1	A	486	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	488	ARG
1	A	491	GLU
1	A	497	CYS
1	A	500	LEU
1	A	504	LEU
1	A	545	GLU
1	A	596	ARG
1	A	603	SER
1	A	607	MET
1	C	476	LEU
1	C	541	CYS
1	C	544	ARG
1	C	552	ASN
1	C	573	GLN
1	C	582	ASN
1	C	583	CYS
1	C	585	MET
1	C	595	LEU
1	C	606	GLN
2	P	8	ARG

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	551	ASN

5.3.3 RNA

There are no RNA molecules 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

Of 6 ligands modelled in this entry, 6 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.

No monomer is involved in short contacts.

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, 95^{th} 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	137/137 (100%)	-0.53	2 (1%) 76 75	18, 26, 43, 57	0
1	C	137/137 (100%)	-0.46	3 (2%) 65 64	18, 27, 49, 57	0
2	P	8/8 (100%)	0.88	2 (25%) 1 1	37, 41, 50, 54	0
All	All	282/282 (100%)	-0.46	7 (2%) 61 60	18, 27, 46, 57	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	610	ALA	3.2
1	A	550	GLY	3.2
2	P	6	THR	3.2
1	C	610	ALA	3.0
2	P	8	ARG	2.6
1	C	553	ASN	2.4
1	C	525	ALA	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, 95^{th} 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	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
3	ZN	A	702	1/1	1.00	0.11	1.76	20,20,20,20	0
3	ZN	A	701	1/1	1.00	0.12	1.48	24,24,24,24	0
3	ZN	A	703	1/1	1.00	0.12	0.43	27,27,27,27	0
3	ZN	C	701	1/1	0.99	0.12	0.23	27,27,27,27	0
3	ZN	C	703	1/1	1.00	0.09	-0.31	20,20,20,20	0
3	ZN	C	702	1/1	0.99	0.09	-1.75	21,21,21,21	0

6.5 Other polymers (i)

There are no such residues in this entry.