



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 2, 2023 – 12:03 PM EDT

PDB ID : 8DS8
Title : Crystal structure of human TNRC18 BAH domain in complex with H3K9me3 peptide
Authors : Song, J.K.; Lu, J.W.
Deposited on : 2022-07-21
Resolution : 1.84 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.34
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.34

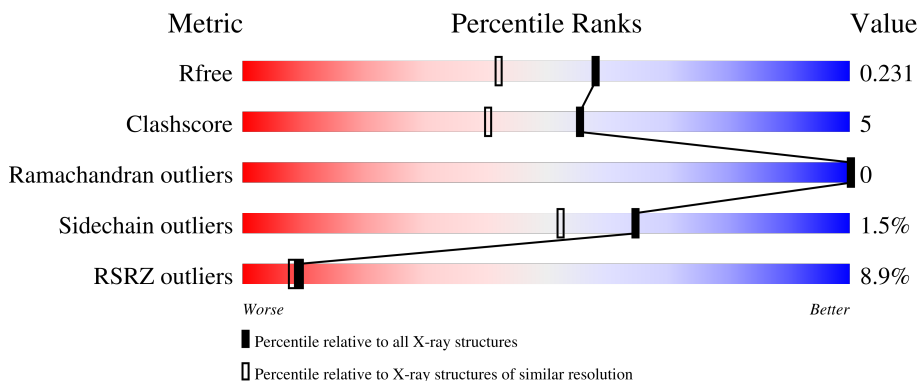
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.84 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}	130704	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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 $\leq 5\%$. 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	184	
1	B	184	
2	C	23	
2	D	23	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2970 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 Trinucleotide repeat-containing gene 18 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	166	1349	867	238	234	10	0	4	0
1	B	157	1256	808	218	222	8	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2784	GLY	-	expression tag	UNP O15417
B	2784	GLY	-	expression tag	UNP O15417

- Molecule 2 is a protein called Histone H3.1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	8	52	32	9	11	0	0	0
2	D	8	62	37	13	12	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	23	TRP	LYS	conflict	UNP V9H1G0
D	23	TRP	LYS	conflict	UNP V9H1G0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	132	Total 132	O 132	0	0
3	C	3	Total 3	O 3	0	0

Continued on next page...

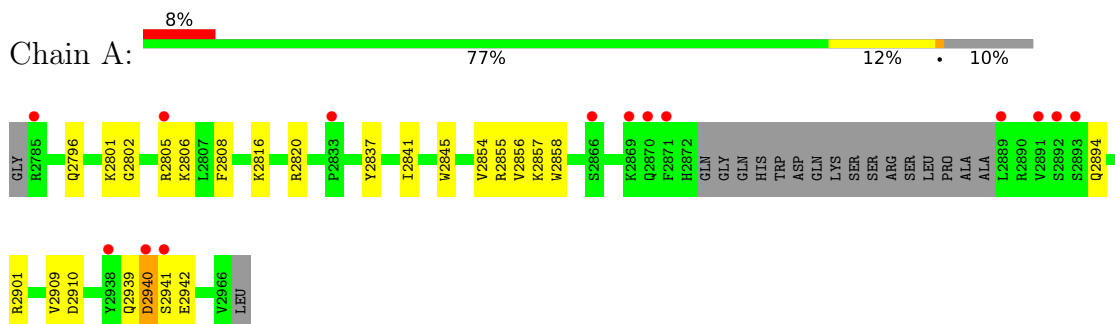
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	115	Total 115	O 115	0	0
3	D	1	Total 1	O 1	0	0

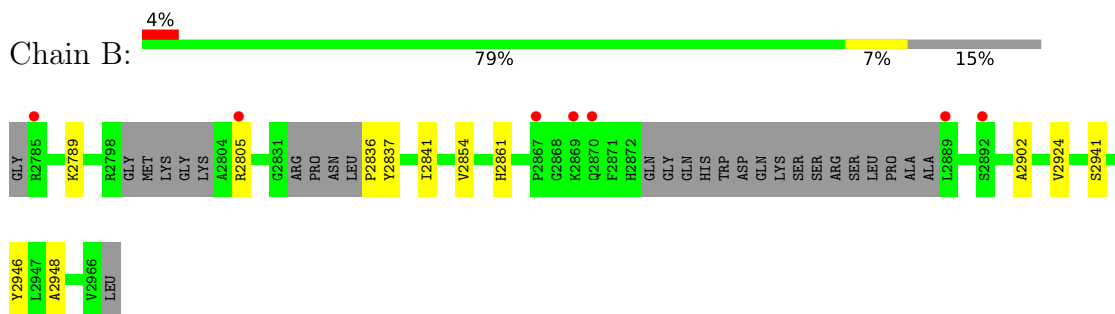
3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

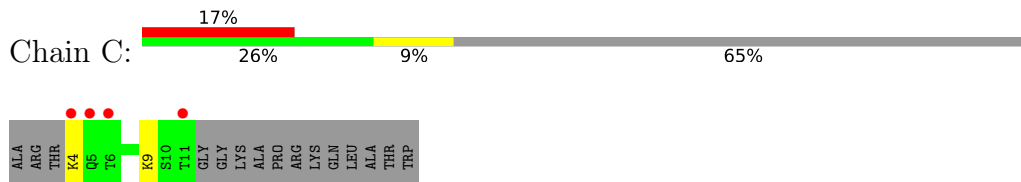
- Molecule 1: Trinucleotide repeat-containing gene 18 protein



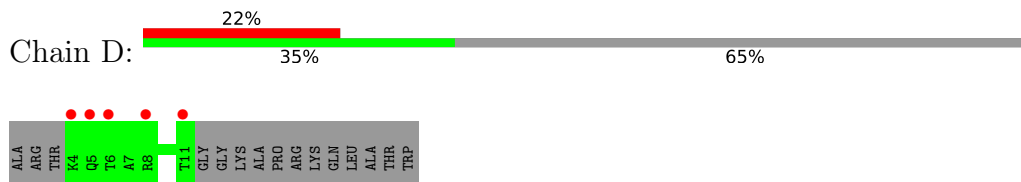
- Molecule 1: Trinucleotide repeat-containing gene 18 protein



- Molecule 2: Histone H3.1



- Molecule 2: Histone H3.1



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	73.80Å 120.83Å 46.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.48 – 1.84 37.48 – 1.84	Depositor EDS
% Data completeness (in resolution range)	98.6 (37.48-1.84) 98.6 (37.48-1.84)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.49 (at 1.84Å)	Xtrriage
Refinement program	PHENIX 1.20.1_4487, PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.203 , 0.235 0.201 , 0.231	Depositor DCC
R_{free} test set	1984 reflections (5.44%)	wwPDB-VP
Wilson B-factor (Å ²)	26.5	Xtrriage
Anisotropy	0.678	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 54.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2970	wwPDB-VP
Average B, all atoms (Å ²)	38.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.20 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.5792e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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: M3L

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	0.46	0/1399	0.66	0/1884
1	B	0.47	0/1291	0.64	0/1740
2	C	0.31	0/39	0.68	0/53
2	D	0.34	0/49	0.74	0/65
All	All	0.46	0/2778	0.65	0/3742

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 [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 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	1349	0	1317	19	0
1	B	1256	0	1198	7	0
2	C	52	0	47	2	0
2	D	62	0	64	0	0
3	A	132	0	0	6	2
3	B	115	0	0	1	2
3	C	3	0	0	1	0
3	D	1	0	0	0	0
All	All	2970	0	2626	27	2

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 5.

All (27) 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:2901:ARG:NH1	3:A:3002:HOH:O	2.04	0.90
1:A:2857:LYS:NZ	3:A:3003:HOH:O	2.06	0.88
1:B:2805:ARG:NH1	3:B:3001:HOH:O	2.19	0.75
1:A:2816:LYS:O	3:A:3001:HOH:O	2.03	0.75
1:A:2894:GLN:OE1	3:A:3004:HOH:O	2.10	0.69
1:A:2941:SER:O	3:A:3002:HOH:O	2.10	0.68
1:A:2801:LYS:HA	1:A:2805[C]:ARG:HD3	1.78	0.65
1:A:2802:GLY:H	1:A:2805[B]:ARG:HH21	1.48	0.62
1:A:2806:LYS:HE3	1:A:2845:TRP:CD1	2.39	0.57
1:A:2855:ARG:NH2	1:A:2909:VAL:HG21	2.27	0.49
1:A:2801:LYS:HE2	3:A:3071:HOH:O	2.13	0.47
1:A:2841:ILE:HG23	1:A:2854:VAL:HG23	1.96	0.47
1:A:2939:GLN:O	1:A:2940:ASP:HB2	2.15	0.46
1:A:2858:TRP:HE1	1:A:2910:ASP:HB2	1.82	0.44
1:A:2856:VAL:O	1:A:2909:VAL:HA	2.18	0.43
1:B:2902:ALA:HB1	1:B:2948:ALA:HB2	2.00	0.43
1:A:2837:TYR:CD1	2:C:9:M3L:HM12	2.54	0.42
1:B:2836:PRO:HD3	1:B:2861:HIS:NE2	2.35	0.42
1:A:2796[B]:GLN:HG2	1:A:2808:PHE:HB2	2.01	0.42
1:A:2806:LYS:HE2	1:A:2808:PHE:CZ	2.54	0.42
1:B:2941:SER:O	1:B:2941:SER:OG	2.30	0.41
1:A:2806:LYS:HE3	1:A:2845:TRP:CG	2.55	0.41
1:B:2836:PRO:HD3	1:B:2861:HIS:CE1	2.56	0.41
1:B:2841:ILE:HG23	1:B:2854:VAL:HG23	2.03	0.41
1:A:2942:GLU:H	1:A:2942:GLU:CD	2.23	0.40
2:C:4:LYS:N	3:C:102:HOH:O	2.54	0.40
1:B:2924:VAL:O	1:B:2946:TYR:HA	2.22	0.40

All (2) 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	Interatomic distance (Å)	Clash overlap (Å)
3:A:3075:HOH:O	3:B:3002:HOH:O[1_556]	2.15	0.05
3:A:3122:HOH:O	3:B:3092:HOH:O[3_453]	2.18	0.02

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	167/184 (91%)	166 (99%)	1 (1%)	0	100	100
1	B	150/184 (82%)	149 (99%)	1 (1%)	0	100	100
2	C	5/23 (22%)	3 (60%)	2 (40%)	0	100	100
2	D	5/23 (22%)	5 (100%)	0	0	100	100
All	All	327/414 (79%)	323 (99%)	4 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	139/160 (87%)	137 (99%)	2 (1%)	67	55
1	B	128/160 (80%)	126 (98%)	2 (2%)	62	49
2	C	3/16 (19%)	3 (100%)	0	100	100
2	D	5/16 (31%)	5 (100%)	0	100	100
All	All	275/352 (78%)	271 (98%)	4 (2%)	65	52

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2820	ARG
1	A	2940	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	2789	LYS
1	B	2837	TYR

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

Mol	Chain	Res	Type
1	A	2912	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	M3L	D	9	2	10,11,12	0.64	0	9,14,16	0.58	0
2	M3L	C	9	2	10,11,12	0.59	0	9,14,16	0.54	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	M3L	D	9	2	-	4/9/10/12	-
2	M3L	C	9	2	-	2/9/10/12	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	9	M3L	CG-CD-CE-NZ
2	C	9	M3L	CA-CB-CG-CD
2	D	9	M3L	CA-CB-CG-CD
2	D	9	M3L	CD-CE-NZ-CM2
2	D	9	M3L	CD-CE-NZ-CM3
2	D	9	M3L	CD-CE-NZ-CM1

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	9	M3L	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides 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(Å ²)	Q<0.9
1	A	166/184 (90%)	0.30	14 (8%) 11 10	19, 31, 76, 87	0
1	B	157/184 (85%)	0.15	7 (4%) 33 30	19, 32, 69, 81	0
2	C	7/23 (30%)	3.74	4 (57%) 0 0	53, 59, 72, 79	0
2	D	7/23 (30%)	4.41	5 (71%) 0 0	53, 62, 80, 82	0
All	All	337/414 (81%)	0.39	30 (8%) 9 8	19, 32, 75, 87	0

All (30) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	11	THR	8.4
2	C	5	GLN	7.6
2	D	4	LYS	5.9
2	C	11	THR	5.7
1	B	2889	LEU	5.1
1	A	2889	LEU	5.0
2	C	4	LYS	4.9
1	A	2892	SER	4.8
2	D	6	THR	4.6
1	B	2892	SER	4.4
2	D	5	GLN	4.3
2	C	6	THR	4.2
2	D	8	ARG	4.1
1	A	2833	PRO	4.1
1	A	2870	GLN	3.9
1	A	2805[A]	ARG	3.3
1	B	2785	ARG	3.0
1	A	2871	PHE	2.9
1	A	2940	ASP	2.9
1	A	2866	SER	2.8
1	A	2891	VAL	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	2938	TYR	2.8
1	A	2941	SER	2.8
1	A	2893	SER	2.5
1	A	2785	ARG	2.4
1	B	2805	ARG	2.2
1	B	2870	GLN	2.2
1	B	2869	LYS	2.2
1	B	2867	PRO	2.1
1	A	2869	LYS	2.0

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

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. 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	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	M3L	D	9	12/13	0.74	0.27	44,64,69,71	0
2	M3L	C	9	12/13	0.77	0.21	39,59,66,69	0

6.3 Carbohydrates [i](#)

There are no monosaccharides 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.