Full wwPDB X-ray Structure Validation Report (i)

Nov 14, 2023 – 07:40 PM JST

PDB ID	:	6A47
Title	:	Structure of TREX2 in complex with a Y structured dsDNA
Authors	:	Hsiao, Y.Y.
Deposited on	:	2018-06-19
Resolution	:	1.90 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 (i) 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 (1)) were used in the production of this report:

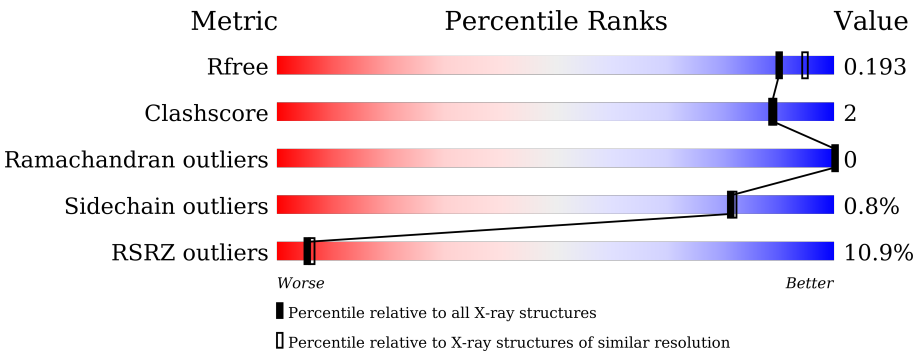
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 <=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	256	10%, 83%, 5%, 12%
1	B	256	6%, 84%, •, 15%
2	C	23	26%, 70%, 13%, 17%
2	D	23	22%, 65%, 17%, 17%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8290 atoms, of which 3772 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 Three prime repair exonuclease 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	224	3433	1101	1692	304	326	10	0	0	0
1	B	217	3340	1070	1651	293	316	10	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP Q9R1A9
A	-18	GLY	-	expression tag	UNP Q9R1A9
A	-17	SER	-	expression tag	UNP Q9R1A9
A	-16	SER	-	expression tag	UNP Q9R1A9
A	-15	HIS	-	expression tag	UNP Q9R1A9
A	-14	HIS	-	expression tag	UNP Q9R1A9
A	-13	HIS	-	expression tag	UNP Q9R1A9
A	-12	HIS	-	expression tag	UNP Q9R1A9
A	-11	HIS	-	expression tag	UNP Q9R1A9
A	-10	HIS	-	expression tag	UNP Q9R1A9
A	-9	SER	-	expression tag	UNP Q9R1A9
A	-8	SER	-	expression tag	UNP Q9R1A9
A	-7	GLY	-	expression tag	UNP Q9R1A9
A	-6	LEU	-	expression tag	UNP Q9R1A9
A	-5	VAL	-	expression tag	UNP Q9R1A9
A	-4	PRO	-	expression tag	UNP Q9R1A9
A	-3	ARG	-	expression tag	UNP Q9R1A9
A	-2	GLY	-	expression tag	UNP Q9R1A9
A	-1	SER	-	expression tag	UNP Q9R1A9
A	0	HIS	-	expression tag	UNP Q9R1A9
B	-19	MET	-	initiating methionine	UNP Q9R1A9
B	-18	GLY	-	expression tag	UNP Q9R1A9
B	-17	SER	-	expression tag	UNP Q9R1A9
B	-16	SER	-	expression tag	UNP Q9R1A9
B	-15	HIS	-	expression tag	UNP Q9R1A9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-14	HIS	-	expression tag	UNP Q9R1A9
B	-13	HIS	-	expression tag	UNP Q9R1A9
B	-12	HIS	-	expression tag	UNP Q9R1A9
B	-11	HIS	-	expression tag	UNP Q9R1A9
B	-10	HIS	-	expression tag	UNP Q9R1A9
B	-9	SER	-	expression tag	UNP Q9R1A9
B	-8	SER	-	expression tag	UNP Q9R1A9
B	-7	GLY	-	expression tag	UNP Q9R1A9
B	-6	LEU	-	expression tag	UNP Q9R1A9
B	-5	VAL	-	expression tag	UNP Q9R1A9
B	-4	PRO	-	expression tag	UNP Q9R1A9
B	-3	ARG	-	expression tag	UNP Q9R1A9
B	-2	GLY	-	expression tag	UNP Q9R1A9
B	-1	SER	-	expression tag	UNP Q9R1A9
B	0	HIS	-	expression tag	UNP Q9R1A9

- Molecule 2 is a DNA chain called DNA (5'-D(P*CP*CP*AP*GP*GP*CP*CP*CP*TP*CP*TP*AP*GP*GP*GP*CP*CP*TP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	C	19	598	182	214	67	116	19	0	0	0
			Total	C	H	N	O	P			
2	D	19	600	183	215	66	117	19	0	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
3	A	1	1	1	0	0
			Total	Mg		
3	B	1	1	1	0	0
			Total	Mg		
3	C	1	1	1	0	0
			Total	Mg		
3	D	1	1	1	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
4	A	1	1	1	0	0

- Molecule 5 is water.

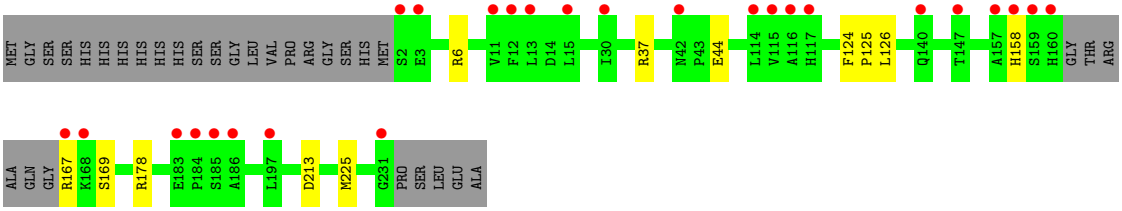
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	129	129	129	0	0
			Total	O		
5	B	158	158	158	0	0
			Total	O		
5	C	15	15	15	0	0
			Total	O		
5	D	12	12	12	0	0

3 Residue-property plots (i)

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.

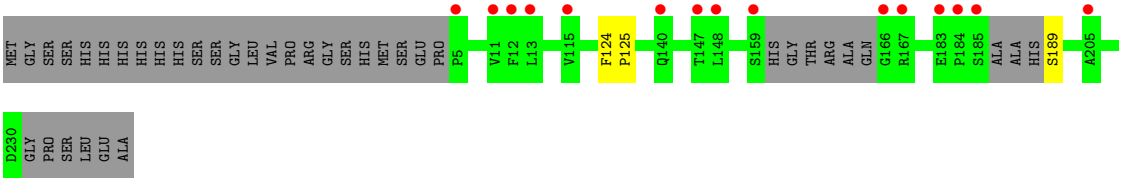
• Molecule 1: Three prime repair exonuclease 2

Chain A:



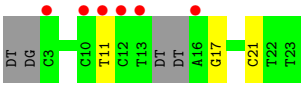
• Molecule 1: Three prime repair exonuclease 2

Chain B:



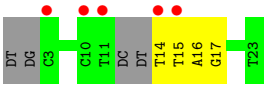
• Molecule 2: DNA (5'-D(P*CP*CP*AP*GP*GP*CP*CP*CP*TP*CP*TP*AP*GP*GP*GP*CP*CP*TP*T)-3')

Chain C:



• Molecule 2: DNA (5'-D(P*CP*CP*AP*GP*GP*CP*CP*CP*TP*CP*TP*AP*GP*GP*GP*CP*CP*TP*T)-3')

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	145.67Å 48.40Å 90.59Å 90.00° 97.36° 90.00°	Depositor
Resolution (Å)	28.45 – 1.90 28.45 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.8 (28.45-1.90) 99.8 (28.45-1.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	2.65 (at 1.91Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.163 , 0.191 0.167 , 0.193	Depositor DCC
R_{free} test set	3751 reflections (7.55%)	wwPDB-VP
Wilson B-factor (Å^2)	31.0	Xtriage
Anisotropy	0.375	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 54.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	8290	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.31% 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, 0.333 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: MG, NA

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	0.30	0/1786	0.48	0/2429
1	B	0.32	0/1730	0.49	0/2349
2	C	0.56	0/427	0.90	0/653
2	D	0.68	0/428	1.00	0/655
All	All	0.39	0/4371	0.62	0/6086

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	1741	1692	1691	7	0
1	B	1689	1651	1650	2	0
2	C	384	214	215	2	0
2	D	385	215	216	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	129	0	0	2	0
5	B	158	0	0	1	0
5	C	15	0	0	1	0
5	D	12	0	0	0	0
All	All	4518	3772	3772	15	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 2.

All (15) 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:225:MET:SD	5:A:484:HOH:O	2.22	0.98
1:A:169:SER:O	1:A:178:ARG:NH2	2.19	0.75
1:B:189:SER:N	5:B:401:HOH:O	2.18	0.75
2:D:15:DT:H3'	2:D:16:DA:H5"	1.86	0.57
2:D:15:DT:C3'	2:D:16:DA:H5"	2.39	0.53
1:A:44:GLU:OE1	5:A:401:HOH:O	2.19	0.51
1:A:126:LEU:HD13	1:A:225:MET:CE	2.45	0.47
2:D:16:DA:H1'	2:D:17:DG:H5"	1.97	0.46
1:A:126:LEU:HD13	1:A:225:MET:HE1	1.98	0.45
1:B:124:PHE:HB2	1:B:125:PRO:HD3	1.99	0.44
1:A:124:PHE:HB2	1:A:125:PRO:HD3	2.01	0.43
1:A:37:ARG:NH2	1:A:213:ASP:OD2	2.44	0.42
2:D:14:DT:H2"	2:D:15:DT:H2'	2.03	0.41
2:C:21:DC:OP2	5:C:201:HOH:O	2.22	0.40
2:C:11:DT:O2	2:C:17:DG:N2	2.55	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	220/256 (86%)	216 (98%)	4 (2%)	0	100	100
1	B	211/256 (82%)	208 (99%)	3 (1%)	0	100	100
All	All	431/512 (84%)	424 (98%)	7 (2%)	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	189/214 (88%)	186 (98%)	3 (2%)	62	60
1	B	184/214 (86%)	184 (100%)	0	100	100
All	All	373/428 (87%)	370 (99%)	3 (1%)	81	82

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	6	ARG
1	A	158	HIS
1	A	167	ARG

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	188	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 5 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 (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	224/256 (87%)	0.46	26 (11%)	4	5	22, 35, 68, 115	0
1	B	217/256 (84%)	0.25	15 (6%)	16	19	23, 34, 64, 104	0
2	C	19/23 (82%)	1.34	6 (31%)	0	0	32, 66, 134, 142	0
2	D	19/23 (82%)	1.09	5 (26%)	0	0	35, 64, 134, 169	0
All	All	479/558 (85%)	0.42	52 (10%)	5	6	22, 35, 86, 169	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	160	HIS	7.4
1	A	159	SER	6.8
1	B	5	PRO	5.6
1	B	167	ARG	5.6
1	B	166	GLY	5.5
1	A	167	ARG	5.3
2	D	14	DT	5.0
1	A	2	SER	4.9
2	D	15	DT	4.9
1	A	185	SER	4.9
2	C	12	DC	4.6
1	A	186	ALA	4.5
1	A	115	VAL	4.5
1	B	185	SER	4.1
1	A	116	ALA	4.1
1	A	158	HIS	4.0
2	D	11	DT	4.0
1	A	3	GLU	4.0
2	C	13	DT	3.9
1	B	159	SER	3.8
1	A	12	PHE	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	C	11	DT	3.7
1	A	184	PRO	3.7
1	A	183	GLU	3.4
2	D	3	DC	3.4
1	B	115	VAL	3.3
2	C	3	DC	3.2
1	A	13	LEU	3.2
1	B	184	PRO	3.2
2	C	10	DC	3.1
1	A	140	GLN	2.9
1	A	197	LEU	2.9
2	C	16	DA	2.8
1	A	11	VAL	2.8
1	A	117	HIS	2.7
1	A	15	LEU	2.7
1	A	114	LEU	2.6
1	A	30	ILE	2.6
1	B	183	GLU	2.5
1	B	11	VAL	2.5
1	A	157	ALA	2.4
1	A	147	THR	2.4
1	B	140	GLN	2.4
1	A	42	ASN	2.4
1	B	148	LEU	2.3
1	B	147	THR	2.3
1	B	12	PHE	2.2
1	A	231	GLY	2.2
2	D	10	DC	2.2
1	A	168	LYS	2.1
1	B	205	ALA	2.1
1	B	13	LEU	2.1

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 monosaccharides in this entry.

6.4 Ligands (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
4	NA	A	302	1/1	0.94	0.08	44,44,44,44	0
3	MG	D	101	1/1	0.97	0.13	32,32,32,32	0
3	MG	A	301	1/1	0.98	0.07	40,40,40,40	0
3	MG	B	301	1/1	0.99	0.12	50,50,50,50	0
3	MG	C	101	1/1	0.99	0.22	30,30,30,30	0

6.5 Other polymers (i)

There are no such residues in this entry.