wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 08:13 PM GMT

PDB ID	:	4R79
Title	:	Mos1 transposase paired-end complex with left transposon end
Authors	:	Richardson, J.M.
Deposited on	:	2014-08-27
Resolution	:	3.10 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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:

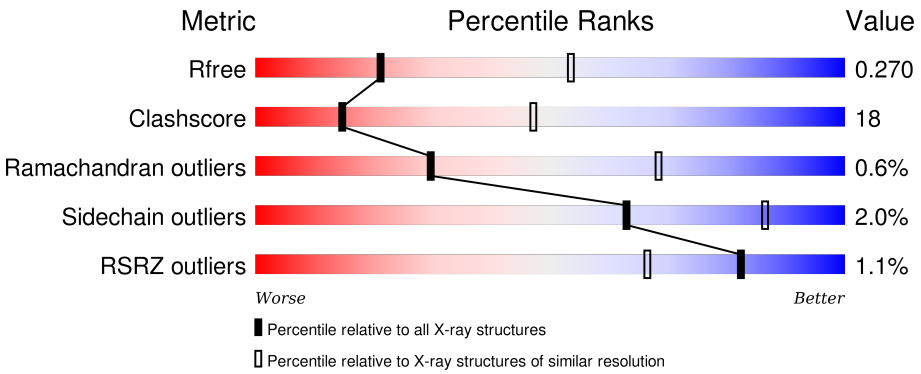
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 3.10 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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	C	25	36% 64%
1	E	25	48% 52%
1	G	25	12% 88%
2	D	28	36% 64%
2	F	28	50% 50%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	27	15% 85%
4	A	345	% 73% 23%
4	B	345	% 71% 26%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 8853 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 DNA chain called left Inverted repeat NTS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	25	521	248	100	149	24	0	0	0
1	E	25	521	248	100	149	24	0	0	0
1	G	25	524	248	100	151	25	0	0	0

- Molecule 2 is a DNA chain called left Inverted repeat TS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	28	562	270	99	166	27	0	0	0
2	F	28	562	270	99	166	27	0	0	0

- Molecule 3 is a DNA chain called left Inverted repeat NTS H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	H	27	540	260	94	160	26	0	0	0

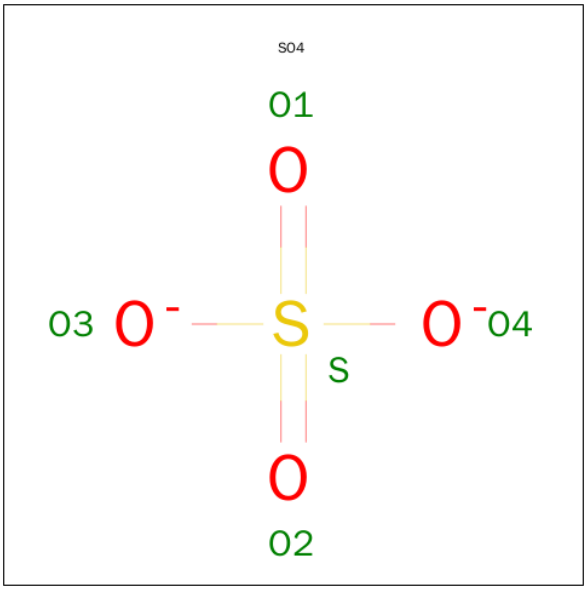
- Molecule 4 is a protein called Mariner Mos1 transposase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	336	2803	1778	507	508	10	0	0	0
4	B	336	2803	1778	507	508	10	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	45	THR	LYS	VARIANT	UNP Q7JQ07
A	164	ASN	SER	VARIANT	UNP Q7JQ07
A	210	PRO	ARG	VARIANT	UNP Q7JQ07
A	216	ALA	THR	ENGINEERED MUTATION	UNP Q7JQ07
A	344	PHE	LEU	VARIANT	UNP Q7JQ07
B	45	THR	LYS	VARIANT	UNP Q7JQ07
B	164	ASN	SER	VARIANT	UNP Q7JQ07
B	210	PRO	ARG	VARIANT	UNP Q7JQ07
B	216	ALA	THR	ENGINEERED MUTATION	UNP Q7JQ07
B	344	PHE	LEU	VARIANT	UNP Q7JQ07

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	1	Total O S 5 4 1	0	0
5	B	1	Total O S 5 4 1	0	0

- Molecule 6 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	H	1	Total Mn 1 1	0	0
6	A	1	Total Mn 1 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	1	Total O 1 1	0	0
7	D	1	Total O 1 1	0	0
7	F	1	Total O 1 1	0	0
7	A	1	Total O 1 1	0	0
7	B	1	Total O 1 1	0	0

3 Residue-property plots (i)

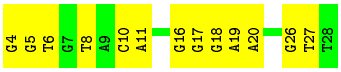
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: left Inverted repeat NTS

Chain C:

- Molecule 1: left Inverted repeat NTS

Chain E:

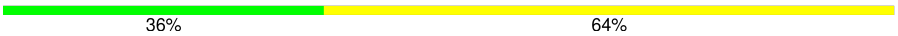


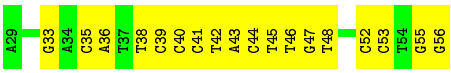
- Molecule 1: left Inverted repeat NTS

Chain G:

- Molecule 2: left Inverted repeat TS

Chain D:



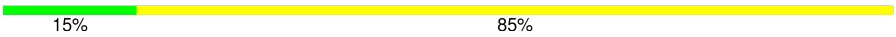


- Molecule 2: left Inverted repeat TS

Chain F:

- Molecule 3: left Inverted repeat NTS H

Chain H:



A29 A30 C31 C32 G33 A34 C35 A36 T37 T38 C39 C40 C41 T42 A43 C44 T45 T46 G47 T48 A49 C50 A51 C52 C53 T54 G55

- Molecule 4: Mariner Mos1 transposase

Chain A:

MET SER SER PHE V5 P6 T11 L15 T24 E27 R30 M31 L32 V33 V40 P41 T45 R52 V60 D61 K70 A75 Q78 E97 V98 S99 Q100 V103 R106 M110 V116 W119 V120 P121 L124 N125 E126 R127 Q128 R131

T135 L139 S147 F148 L149 H150 R151 I152 Y153 T154 G155 D156 E157 K158 N164 T191 M192 L193 C194 W197 L207 L208 K209 P210 G211 E212 T213 V214 N215 A216 Y219 Q220 L223 N227 L230 Q231 R234 Y237 GLN LYS ARG GLN HIS R243 V244 I245

H248 H254 A258 T262 L266 L271 S283 D284 Y285 H286 A289 E297 K307 E320 R324 G325 I326 L329 P330 Y343 F344 E345

- Molecule 4: Mariner Mos1 transposase

Chain B:

MET SER SER PHE V5 P6 N7 T11 R12 I16 K22 K23 T24 E27 L32 V33 G37 V40 P41 D57 F58 D59 Y72 L77 D84 Q87 T88 Q89 K90 Q91 S99 Q100 Q101 A102 V103 R106 M110 P121 N134 E137 I138 L139

L140 T154 G155 D156 E157 K158 W159 F162 R167 K168 K169 V172 D173 R183 P184 K190 T191 M192 L193 C194 D198 Q199 I203 L207 L208 K209 P210 G211 E212 N215 A216 Q220 N227 Q231 R234 P235 E236 Y237 GLN LYS ARG GLN HIS R243 V244

L247 R257 A258 L266 N267 W268 L271 S277 P278 D279 L280 S283 D284 Y285 H286 L287 F288 A289 S290 M291 L295 F300 K307 E320 R324 G325 L329 W333 E334 K335 C336 Y343 F344 E345

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	119.78Å 86.89Å 132.38Å 90.00° 99.03° 90.00°	Depositor
Resolution (Å)	29.87 – 3.10 29.74 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.4 (29.87-3.10) 99.5 (29.74-3.10)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.20 (at 3.11Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.227 , 0.270 0.227 , 0.270	Depositor DCC
R_{free} test set	2444 reflections (5.27%)	DCC
Wilson B-factor (Å^2)	42.9	Xtriage
Anisotropy	1.272	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 53.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Outliers	2 of 48780 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8853	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MN, SO4

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	C	0.31	0/586	0.71	0/906
1	E	0.36	0/586	0.73	0/906
1	G	0.25	0/589	0.68	0/910
2	D	0.38	0/628	0.75	0/965
2	F	0.35	0/628	0.71	0/965
3	H	0.26	0/603	0.72	0/926
4	A	0.30	0/2876	0.49	0/3881
4	B	0.29	0/2876	0.47	0/3881
All	All	0.30	0/9372	0.59	0/13340

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	C	521	0	284	29	0
1	E	521	0	284	23	0
1	G	524	0	283	39	0
2	D	562	0	317	16	0
2	F	562	0	317	14	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	H	540	0	306	39	0
4	A	2803	0	2739	77	0
4	B	2803	0	2739	75	0
5	B	5	0	0	0	0
5	D	5	0	0	0	0
6	A	1	0	0	0	0
6	H	1	0	0	0	0
7	A	1	0	0	0	0
7	B	1	0	0	0	0
7	C	1	0	0	0	0
7	D	1	0	0	0	0
7	F	1	0	0	1	0
All	All	8853	0	7269	283	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 18.

The worst 5 of 283 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:C:17:DG:H2"	1:C:18:DG:H5"	1.26	1.13
4:B:227:ASN:ND2	4:B:266:LEU:HD22	1.66	1.10
1:G:7:DG:H5'	4:B:216:ALA:HB2	1.31	1.08
1:E:4:DG:H2"	1:E:5:DG:C5'	1.83	1.07
4:A:262:THR:O	4:A:266:LEU:HD13	1.56	1.04

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	
4	A	332/345 (96%)	311 (94%)	19 (6%)	2 (1%)	30	68
4	B	332/345 (96%)	309 (93%)	21 (6%)	2 (1%)	30	68
All	All	664/690 (96%)	620 (93%)	40 (6%)	4 (1%)	30	68

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	B	210	PRO
4	A	210	PRO
4	B	199	GLN
4	A	41	PRO

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	
4	A	298/307 (97%)	290 (97%)	8 (3%)	52	82
4	B	298/307 (97%)	294 (99%)	4 (1%)	76	91
All	All	596/614 (97%)	584 (98%)	12 (2%)	63	86

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	135	THR
4	A	214	VAL
4	B	59	ASP
4	A	98	VAL
4	B	24	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	164	ASN
4	B	225	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	B	248	HIS
4	B	122	HIS
4	B	231	GLN

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)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
5	SO4	B	401	-	4,4,4	0.38	0	6,6,6	0.09	0
5	SO4	D	101	-	4,4,4	0.36	0	6,6,6	0.07	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
5	SO4	B	401	-	-	0/0/0/0	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	SO4	D	101	-	-	0/0/0/0	0/0/0/0

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	C	25/25 (100%)	-0.45	0	100	100	44, 65, 82, 88	0
1	E	25/25 (100%)	-0.43	0	100	100	35, 70, 87, 92	0
1	G	25/25 (100%)	0.17	0	100	100	52, 113, 147, 152	0
2	D	28/28 (100%)	-0.53	0	100	100	37, 62, 82, 89	0
2	F	28/28 (100%)	-0.54	0	100	100	42, 64, 90, 95	0
3	H	27/27 (100%)	0.12	0	100	100	53, 109, 145, 148	0
4	A	336/345 (97%)	-0.23	4 (1%)	81	64	33, 67, 109, 134	0
4	B	336/345 (97%)	-0.21	5 (1%)	76	58	27, 70, 111, 145	0
All	All	830/848 (97%)	-0.23	9 (1%)	82	66	27, 70, 117, 152	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	155	GLY	3.5
4	A	156	ASP	3.4
4	B	156	ASP	3.3
4	B	154	THR	3.2
4	A	154	THR	3.1

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 carbohydrates 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. 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, 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	LLDF	B-factors(Å^2)	Q<0.9
5	SO4	B	401	5/5	0.87	0.17	-	108,121,125,128	0
5	SO4	D	101	5/5	0.94	0.34	-	134,134,150,151	0
6	MN	A	401	1/1	0.98	0.28	-	69,69,69,69	0
6	MN	H	101	1/1	0.98	0.28	-	81,81,81,81	0

6.5 Other polymers (i)

There are no such residues in this entry.