W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2023 – 06:59 am GMT

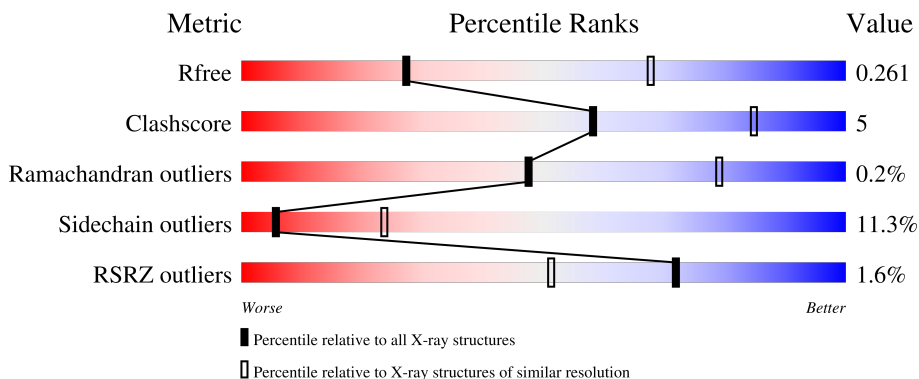|		
PDB ID	:	4U7B
Title	:	Crystal structure of a pre-cleavage Mos1 transpososome
Authors	:	Richardson, J.M.
Deposited on	:	2014-07-30
Resolution	:	3.09 Å(reported)

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 3.09 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	Percentile Ranks	Value
Rfree		0.261
Clashscore		5
Ramachandran outliers		0.2%
Sidechain outliers		11.3%
RSRZ outliers		1.6%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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 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	C	25	48% 48% •
1	E	25	56% 44%
1	H	25	56% 44%
2	D	31	6% 71% 29%
2	F	31	65% 35%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	31	71% ... 29%
3	A	342	3% ... 82% ... 17% ...
3	B	342	2% ... 81% ... 17% ...
3	G	342	82% ... 15% ...

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 12039 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 DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	25	Total 520	C 248	N 94	O 153	P 25	0	0	0
1	E	25	Total 520	C 248	N 94	O 153	P 25	0	0	0
1	H	25	Total 520	C 248	N 94	O 153	P 25	0	0	0

- Molecule 2 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	31	Total 629	C 303	N 114	O 182	P 30	0	0	0
2	F	31	Total 629	C 303	N 114	O 182	P 30	0	0	0
2	I	31	Total 629	C 303	N 114	O 182	P 30	0	0	0

- Molecule 3 is a protein called Mariner Mos1 transposase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	342	Total 2858	C 1814	N 521	O 513	S 10	0	0	0
3	B	342	Total 2859	C 1814	N 521	O 514	S 10	0	0	0
3	G	342	Total 2858	C 1814	N 521	O 513	S 10	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	45	THR	LYS	conflict	UNP Q7JQ07
A	164	ASN	SER	conflict	UNP Q7JQ07

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	210	PRO	ARG	conflict	UNP Q7JQ07
A	216	ALA	THR	conflict	UNP Q7JQ07
A	249	ALA	ASP	conflict	UNP Q7JQ07
A	344	PHE	LEU	conflict	UNP Q7JQ07
B	45	THR	LYS	conflict	UNP Q7JQ07
B	164	ASN	SER	conflict	UNP Q7JQ07
B	210	PRO	ARG	conflict	UNP Q7JQ07
B	216	ALA	THR	conflict	UNP Q7JQ07
B	249	ALA	ASP	conflict	UNP Q7JQ07
B	344	PHE	LEU	conflict	UNP Q7JQ07
G	45	THR	LYS	conflict	UNP Q7JQ07
G	164	ASN	SER	conflict	UNP Q7JQ07
G	210	PRO	ARG	conflict	UNP Q7JQ07
G	216	ALA	THR	conflict	UNP Q7JQ07
G	249	ALA	ASP	conflict	UNP Q7JQ07
G	344	PHE	LEU	conflict	UNP Q7JQ07

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
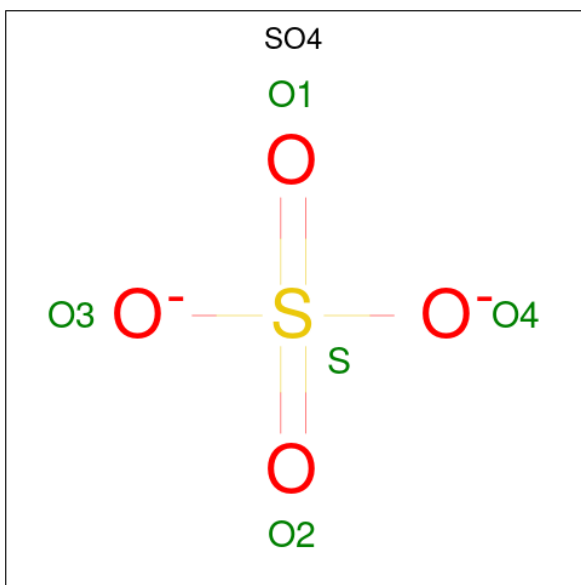
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total 5	O 4	S 1	0	0
4	D	1	Total 5	O 4	S 1	0	0
4	E	1	Total 5	O 4	S 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total O 1 1	0	0
5	D	1	Total O 1 1	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.
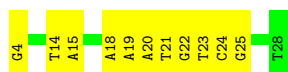
- Molecule 1: DNA (25-MER)

Chain C:

- Molecule 1: DNA (25-MER)

Chain E:

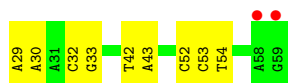- Molecule 1: DNA (25-MER)

Chain H:

- Molecule 2: DNA (31-MER)

Chain D:
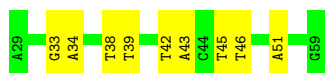
- Molecule 2: DNA (31-MER)

Chain F:

- Molecule 2: DNA (31-MER)

Chain I:
71%   29%

A29 G33 A34 T38 T39 T42 A43 C44 T45 T46 A51 G59

• Molecule 3: Mariner Mos1 transposase

3%

Chain A:
82%   17%   •

F4 N7 K8 R30 T45 V60 D61 D62 K63 Y72 Q78 L96 Q100 R106 L107 R108 V120 P121 N125 E126 R127 Q128 M129 E130 E137 L140 S141 R142 L149 H150 R151 I152 G155 D156 E157 V163 N164 R167 K168 K169 S170 T181

N185 R186 F187 T191 M192 W196 G201 T213 V214 N215 L226 L230 P235 E236 Y237 Q238 K239 R240 R243 H248 A249 N250 A251 P252 S253 N267 L271 P272 H273 S283 D284 Y285 Y322 K328 L329 R332 W333 E334 K335 C336 E345

• Molecule 3: Mariner Mos1 transposase

2%

Chain B:
81%   17%   •

F4 K8 Q39 V40 P41 T45 S55 Y72 Q78 A86 Q87 T88 V98 Q101 R106 R118 E123 L124 N125 K133 N134 T135 K146 L149 H150 R151 I152 V153 T154 G155 D156 E157 K158 W159 N164 Q176 T181 N185 K190 T191

M192 W196 L207 L208 K209 T213 Q221 N225 R228 E236 Y237 Q238 K239 H242 H248 A251 P252 S253 H254 R260 L266 L271 P272 S277 P278 D279 L280 A281 P282 H286 L310 F314 E320 F321 Y322 W323 R324 H327 K328 L329

D340 E345

• Molecule 3: Mariner Mos1 transposase

Chain G:
82%   15%   •

F4 T24 L32 V40 K44 T45 R48 D57 D62 K67 Y72 Q78 D85 K90 L96 E97 V98 S99 Q100 Q101 R106 V120 E123 L124 N125 E126 R127 E130 R131 L140 K144 R145 V153 E157 K158 W159 I160 N164

T181 N185 K190 T191 M192 K209 T213 V214 R218 E236 R240 L247 A251 P252 L271 S277 L280 Y285 H286 M291 K307 W313 D318 G325 K328 L329 R332 C336 E345

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	96.47Å 340.09Å 160.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.75 – 3.09 29.73 – 3.09	Depositor EDS
% Data completeness (in resolution range)	98.6 (29.75-3.09) 98.8 (29.73-3.09)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.09 (at 3.11Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.228 , 0.261 0.224 , 0.261	Depositor DCC
R_{free} test set	2447 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	79.6	Xtriage
Anisotropy	0.838	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 60.6	EDS
L-test for twinning [2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	12039	wwPDB-VP
Average B, all atoms (Å^2)	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.16% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.50	1/583 (0.2%)	0.71	0/898
1	E	0.52	1/583 (0.2%)	0.77	0/898
1	H	0.51	1/583 (0.2%)	0.76	0/898
2	D	0.36	0/705	0.73	0/1085
2	F	0.36	0/705	0.73	0/1085
2	I	0.31	0/705	0.71	0/1085
3	A	0.31	0/2934	0.52	0/3960
3	B	0.32	0/2935	0.55	0/3960
3	G	0.32	0/2934	0.53	0/3960
All	All	0.36	3/12667 (0.0%)	0.61	0/17829

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	4	DG	OP3-P	-10.08	1.49	1.61
1	C	4	DG	OP3-P	-9.91	1.49	1.61
1	H	4	DG	OP3-P	-9.88	1.49	1.61

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.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	520	0	285	13	0
1	E	520	0	285	9	0
1	H	520	0	285	9	0
2	D	629	0	352	11	0
2	F	629	0	352	8	0
2	I	629	0	352	7	0
3	A	2858	0	2799	26	0
3	B	2859	0	2799	22	0
3	G	2858	0	2799	24	0
4	C	5	0	0	0	0
4	D	5	0	0	0	0
4	E	5	0	0	0	0
5	C	1	0	0	0	0
5	D	1	0	0	0	0
All	All	12039	0	10308	107	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 5.

The worst 5 of 107 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:4:DG:C6	2:D:54:DT:H72	2.05	0.92
3:G:72:TYR:O	3:G:106:ARG:NH1	2.13	0.82
3:A:130:GLU:HB2	3:G:127:ARG:HD2	1.63	0.80
3:A:130:GLU:OE1	3:G:127:ARG:NH2	2.24	0.71
3:A:267:ASN:OD1	3:G:145:ARG:NH1	2.24	0.70

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	
3	A	340/342 (99%)	312 (92%)	27 (8%)	1 (0%)	41	73
3	B	340/342 (99%)	314 (92%)	25 (7%)	1 (0%)	41	73
3	G	340/342 (99%)	325 (96%)	15 (4%)	0	100	100
All	All	1020/1026 (99%)	951 (93%)	67 (7%)	2 (0%)	47	79

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	7	ASN
3	B	236	GLU

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	
3	A	303/303 (100%)	271 (89%)	32 (11%)	6	26
3	B	303/303 (100%)	270 (89%)	33 (11%)	6	25
3	G	303/303 (100%)	265 (88%)	38 (12%)	4	18
All	All	909/909 (100%)	806 (89%)	103 (11%)	6	23

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	248	HIS
3	G	62	ASP
3	G	318	ASP
3	B	266	LEU
3	G	24	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	G	125	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	G	267	ASN
3	G	128	GLN
3	G	286	HIS
3	G	185	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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

3 ligands 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	$\#|Z| > 2$
4	SO4	E	101	-	4,4,4	0.32	0	6,6,6	0.11	0
4	SO4	D	101	-	4,4,4	0.33	0	6,6,6	0.05	0
4	SO4	C	101	-	4,4,4	0.34	0	6,6,6	0.10	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 ⓘ

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	C	25/25 (100%)	-0.56	0	100	100	73, 98, 108, 118	0
1	E	25/25 (100%)	-0.26	0	100	100	69, 93, 152, 169	0
1	H	25/25 (100%)	-0.50	0	100	100	80, 101, 124, 130	0
2	D	31/31 (100%)	-0.25	2 (6%)	18	8	82, 96, 145, 211	0
2	F	31/31 (100%)	-0.22	0	100	100	74, 88, 136, 193	0
2	I	31/31 (100%)	-0.34	0	100	100	84, 99, 123, 179	0
3	A	342/342 (100%)	-0.06	9 (2%)	56	33	66, 108, 158, 218	0
3	B	342/342 (100%)	-0.10	7 (2%)	65	44	70, 108, 150, 218	0
3	G	342/342 (100%)	-0.14	1 (0%)	94	88	61, 104, 149, 175	0
All	All	1194/1194 (100%)	-0.13	19 (1%)	72	51	61, 105, 154, 218	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	345	GLU	3.6
3	A	155	GLY	3.4
3	B	345	GLU	3.0
3	B	155	GLY	3.0
3	B	242	HIS	2.9

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 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	SO4	D	101	5/5	0.84	0.41	191,191,193,193	0
4	SO4	C	101	5/5	0.90	0.41	150,154,157,160	0
4	SO4	E	101	5/5	0.92	0.27	155,156,161,164	0

6.5 Other polymers (i)

There are no such residues in this entry.