Full wwPDB X-ray Structure Validation Report ⓘ

Feb 25, 2024 – 05:17 PM EST

PDB ID	:	5KSE
Title	:	Flap endonuclease 1 (FEN1) R100A with 5'-flap substrate DNA and Sm3+
Authors	:	Tsutakawa, S.E.; Arvai, A.S.; Tainer, J.A.
Deposited on	:	2016-07-08
Resolution	:	2.10 Å(reported)

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

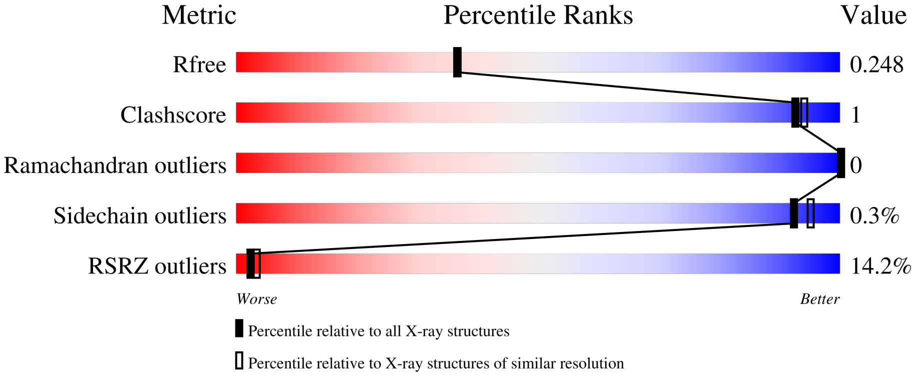
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

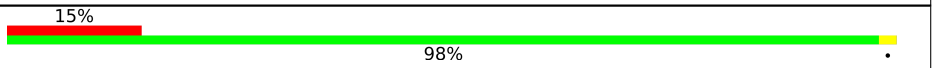
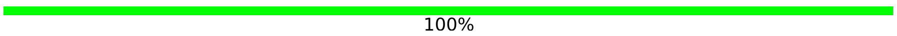
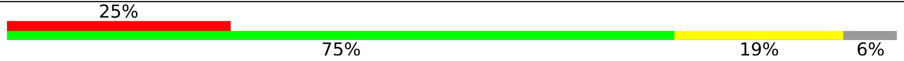
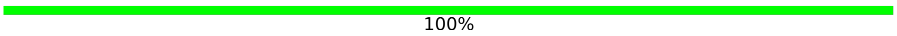
The reported resolution of this entry is 2.10 Å.

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}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	A	341	
2	D	18	
3	E	16	
4	F	7	

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 6885 atoms, of which 3187 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 Flap endonuclease 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	341	5439	1706	2733	469	515	16	0	3	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	100	ALA	ARG	engineered mutation	UNP P39748
A	337	LEU	-	expression tag	UNP P39748
A	338	GLU	-	expression tag	UNP P39748
A	339	VAL	-	expression tag	UNP P39748
A	340	LEU	-	expression tag	UNP P39748
A	341	PHE	-	expression tag	UNP P39748
A	342	GLN	-	expression tag	UNP P39748

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	D	18	567	174	204	66	106	17	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
3	E	15	484	149	170	61	89	15	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(*AP*CP*CP*GP*TP*CP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
4	F	7	216	66	80	24	40	6	0	0	0

- Molecule 5 is SAMARIUM (III) ION (three-letter code: SM) (formula: Sm).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Sm		
5	A	5	5	5	0	0

- Molecule 6 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	K		
6	A	1	1	1	0	0

- Molecule 7 is water.

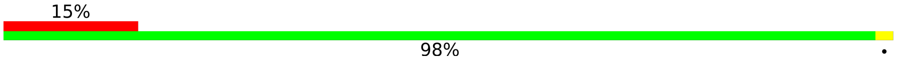
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	144	144	144	0	0
7	D	13	13	13	0	0
7	E	10	10	10	0	0
7	F	6	6	6	0	0

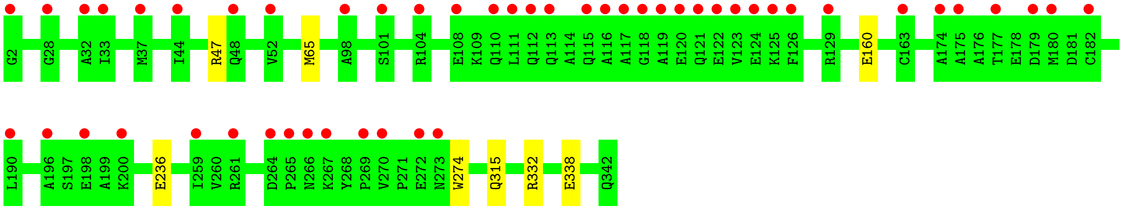
3 Residue-property plots

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: Flap endonuclease 1

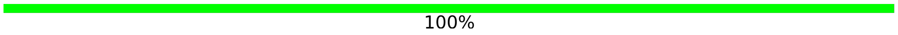
Chain A:





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

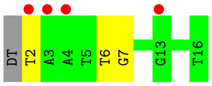
Chain D:



There are no outlier residues recorded for this chain.

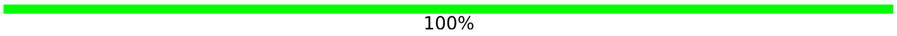
- Molecule 3: DNA (5’-D(P*TP*AP*AP*TP*TP*GP*AP*GP*GP*CP*AP*GP*AP*GP*T)-3’)

Chain E:



- Molecule 4: DNA (5’-D(*AP*CP*CP*GP*TP*CP*C)-3’)

Chain F:



There are no outlier residues recorded for this chain.

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	105.23Å 105.23Å 104.06Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.00 – 2.10 37.00 – 2.10	Depositor EDS
% Data completeness (in resolution range)	89.9 (37.00-2.10) 90.3 (37.00-2.10)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.20 (at 2.10Å)	Xtriage
Refinement program	PHENIX (dev_2383: ???)	Depositor
R, R_{free}	0.207 , 0.245 0.212 , 0.248	Depositor DCC
R_{free} test set	2004 reflections (5.68%)	wwPDB-VP
Wilson B-factor (Å^2)	63.4	Xtriage
Anisotropy	0.175	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 65.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.023 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6885	wwPDB-VP
Average B, all atoms (Å^2)	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.77% 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: K, SM

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	0.25	0/2762	0.40	0/3714
2	D	0.48	0/406	0.89	0/624
3	E	0.57	0/353	0.91	0/544
4	F	0.46	0/151	0.79	0/230
All	All	0.34	0/3672	0.57	0/5112

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	2706	2733	2723	6	0
2	D	363	204	204	0	0
3	E	314	170	170	3	0
4	F	136	80	80	0	0
5	A	5	0	0	0	0
6	A	1	0	0	0	0
7	A	144	0	0	3	1
7	D	13	0	0	1	0
7	E	10	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	F	6	0	0	0	0
All	All	3698	3187	3177	9	1

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 1.

All (9) 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:236:GLU:O	1:A:274:TRP:NE1	2.36	0.57
1:A:160:GLU:CD	7:A:512:HOH:O	2.48	0.52
1:A:315:GLN:NE2	7:A:504:HOH:O	2.42	0.52
3:E:6:DT:H4'	3:E:7:DG:OP1	2.10	0.51
1:A:332:ARG:NH2	1:A:338:GLU:OE2	2.43	0.51
1:A:160:GLU:CG	7:A:512:HOH:O	2.63	0.44
3:E:2:DT:O2	3:E:2:DT:O4'	2.36	0.44
3:E:6:DT:H2"	3:E:7:DG:O5'	2.18	0.42
1:A:65[B]:MET:HE3	7:D:106:HOH:O	2.20	0.42

All (1) 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)
7:A:590:HOH:O	7:A:614:HOH:O[5_554]	2.09	0.11

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	
1	A	342/341 (100%)	332 (97%)	10 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	
1	A	295/292 (101%)	294 (100%)	1 (0%)	92	95

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	47	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 ⓘ

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 (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	341/341 (100%)	0.83	50 (14%)	2	3	49, 72, 128, 160	0
2	D	18/18 (100%)	0.14	0	100	100	76, 96, 122, 126	0
3	E	15/16 (93%)	1.24	4 (26%)	0	0	95, 110, 159, 176	0
4	F	7/7 (100%)	-0.25	0	100	100	71, 81, 100, 103	0
All	All	381/382 (99%)	0.80	54 (14%)	2	3	49, 74, 136, 176	0

All (54) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	123	VAL	10.3
1	A	116	ALA	7.3
1	A	118	GLY	5.9
1	A	126	PHE	5.8
1	A	117	ALA	5.7
1	A	115	GLN	5.5
1	A	125	LYS	4.7
3	E	2	DT	4.6
1	A	112	GLN	4.4
1	A	122	GLU	4.2
3	E	3	DA	4.1
1	A	121	GLN	4.1
1	A	120	GLU	3.9
1	A	270	VAL	3.6
1	A	273	ASN	3.5
1	A	2	GLY	3.4
1	A	198	GLU	3.4
1	A	111	LEU	3.4
1	A	200	LYS	3.3
1	A	272	GLU	3.3
1	A	261	ARG	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	266	ASN	3.1
1	A	267	LYS	3.1
1	A	180	MET	3.0
1	A	119	ALA	3.0
1	A	269	PRO	3.0
1	A	175	ALA	2.8
1	A	52	VAL	2.7
1	A	124	GLU	2.7
1	A	44	ILE	2.7
1	A	113	GLN	2.6
1	A	37	MET	2.6
1	A	104	ARG	2.6
1	A	110	GLN	2.5
1	A	174	ALA	2.5
1	A	196	ALA	2.4
1	A	48	GLN	2.4
1	A	108	GLU	2.4
1	A	101	SER	2.4
1	A	264	ASP	2.4
1	A	177	THR	2.4
3	E	4	DA	2.4
1	A	190	LEU	2.2
1	A	265	PRO	2.2
1	A	28	GLY	2.2
1	A	33	ILE	2.2
1	A	179	ASP	2.2
1	A	259	ILE	2.1
3	E	13	DG	2.1
1	A	32	ALA	2.1
1	A	163	CYS	2.1
1	A	129	ARG	2.1
1	A	98	ALA	2.0
1	A	182	CYS	2.0

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 ⓘ

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, 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	B-factors(Å^2)	Q<0.9
6	K	A	404	1/1	0.91	0.08	122,122,122,122	0
5	SM	A	402	1/1	0.95	0.13	115,115,115,115	1
5	SM	A	403	1/1	0.96	0.14	106,106,106,106	1
5	SM	A	401	1/1	0.96	0.17	97,97,97,97	1
5	SM	A	406	1/1	1.00	0.14	63,63,63,63	0
5	SM	A	405	1/1	1.00	0.16	66,66,66,66	0

6.5 Other polymers ⓘ

There are no such residues in this entry.