wwPDB X-ray Structure Validation Summary Report (i)

Aug 14, 2023 – 12:45 PM EDT

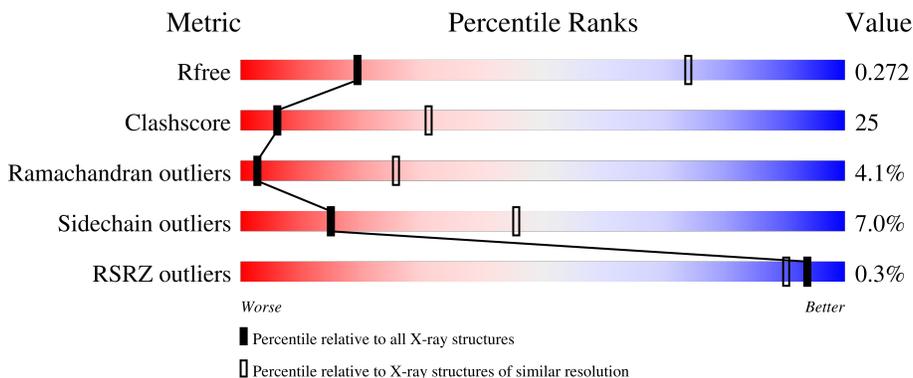|		
PDB ID	:	5W2M
Title	:	APOBEC3F Catalytic Domain Complex with a Single-Stranded DNA
Authors	:	Fang, Y.; Xiao, X.; Li, S.-X.; Wolfe, A.; Chen, X.S.
Deposited on	:	2017-06-06
Resolution	:	3.70 Å(reported)

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 3.70 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		0.272
Clashscore		25
Ramachandran outliers		4.1%
Sidechain outliers		7.0%
RSRZ outliers		0.3%

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	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)

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	184	54% / 39% / 7%
1	B	184	.% — 53% / 40% / 7%
1	C	184	.% — 52% / 40% / 8%
1	D	184	51% / 42% / 8%
1	J	184	.% — 54% / 39% / 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	K	184	54% 39% 7%
1	L	184	51% 42% 8%
1	M	184	% 50% 42% 8%
2	E	10	10% 80% 10%
2	N	10	10% 80% 10%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 12926 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 DNA dC->dU-editing enzyme APOBEC-3F.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	B	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	C	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	D	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	J	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	K	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	L	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			
1	M	184	Total	C	N	O	S	0	0	0
			1565	1017	255	282	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	10	Total	C	N	O	P	0	0	0
			201	100	20	71	10			
2	N	10	Total	C	N	O	P	0	0	0
			201	100	20	71	10			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		
3	B	1	Total	Zn	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	J	1	Total 1	Zn 1	0	0
3	K	1	Total 1	Zn 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: DNA dC->dU-editing enzyme APOBEC-3F

Chain A:

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain B:

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain C:

L362 F363 L364 L368 Q369 E370 I371 L372 E373

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain D:

51% 42% 8%

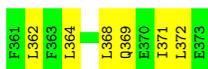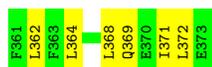N190 E193 A194 M195 Y196 P197 F200 Y201 F202 H203 F204 K205 N206 L207 R208 K209 A210 Y211 G212 R213 S216 C219 F220 T221 M222 E223 V224 V225 K226 H227 H228 S229 K234 N240 Q241 V242 D243 P244 H247 C248 H249 A250 E251 R252 C253 F254 W257 D261 I262 P265

N266 T267 N268 Y269 E270 V271 Y274 T275 S276 W277 S278 P279 C280 P281 E282 C283 A288 L299 T300 I301 F302 L306 Y307 Y308 F309 W310 D311 Y314 G317 L318 L321 S322 Q323 E324 K337 W340 E341 N342 F343 V344 Y345 N346 K352 P353 W354 L357 K358 N359 N360

F361 L362 F363 L364 L368 Q369 E370 I371 L372 E373

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain J:

.%

54% 39% 7%

N190 P191 M195 I199 F200 Y201 K205 N206 L207 R208 K209 A210 Y211 G212 R213 S216 C219 F220 T221 M222 E223 V224 V225 S229 P230 W233 K234 R235 F238 R239 N240 Q241 V242 D243 P244 E245 T246 H247 C248 E251 R252 C253 F254 L255 D261 I262 L263 N266

Y269 E270 Y274 W277 S278 P279 E282 C283 A284 G285 E286 F290 H294 L299 T300 R305 L306 Y307 Y308 F309 W310 D311 Y314 Q315 E316 G317 S320 L321 S322 Q323 E324 G325 A326 K337 Y338 C339 W340 E341 N342 F343 V344 F351 K352 P353 K358 Y359 N360

L364 L368 Q369 E370 I371 L372 E373

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain K:

54% 39% 7%

N190 P191 M195 I199 F200 Y201 K205 N206 L207 R208 K209 A210 Y211 G212 R213 S216 C219 F220 T221 M222 E223 V224 V225 W233 K234 R235 F238 R239 N240 Q241 V242 D243 P244 E245 T246 H247 C248 E251 R252 C253 F254 L255 D261 I262 L263 N266 Y269 E270

Y274 W277 S278 P279 C280 P281 E282 C283 A284 G285 E286 F290 H294 L299 T300 R305 L306 Y307 Y308 F309 W310 D311 Y314 Q315 E316 G317 S320 L321 S322 Q323 E324 G325 A326 Y333 K337 W340 E341 N342 F343 V344 F351 K352 P353 G356 L357 K358 Y359 N360

L364 L368 Q369 E370 I371 L372 E373

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain L:

51% 42% 8%

N190 E193 A194 M195 Y196 P197 F200 Y201 F202 H203 F204 K205 N206 L207 R208 K209 Y211 G212 R213 S216 C219 F220 T221 M222 E223 V224 V225 K226 H227 H228 S229 V231 K234 N240 Q241 V242 D243 P244 H247 C248 H249 A250 E251 R252 C253 F254 W257 D261 I262

P265 N266 T267 N268 Y269 E270 V271 Y274 T275 S276 W277 S278 P279 C280 P281 E282 C283 A288 L299 T300 I301 F302 L306 Y307 Y308 F309 W310 D311 Y314 G317 L318 L321 S322 Q323 E324 K337 W340 E341 N342 F343 V344 Y345 N346 P353 L357 K358 Y359 N360

F361 L362 F363 L364 L368 Q369 E370 I371 L372 E373

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3F

Chain M:

.%

50%   42%   8%

M190 E193 A194 M195 Y196 F200 Y201 F202 H203 F204 K205 N206 L207 R208 K209 A210 Y211 G212 R213 S216 W217 L218 C219 F220 T221 M222 E223 V224 V225 K226 H227 H228 S229 K234 N240 Q241 V242 D243 P244 H247 C248 H249 A250 E251 R252 C253 F254 W257 D261 I262 P265

N266 T267 N268 Y269 E270 V271 Y274 T275 S276 W277 S278 P279 C280 P281 E282 C283 A288 L299 T300 I301 F302 L306 Y307 Y308 F309 W310 D311 Y314 G317 L318 L321 S322 Q323 E324 K337 W340 E341 N342 F343 V344 Y345 N346 K352 P353 W354 L357 K358 Y359 N360

F361 L362 F363 L364 L368 Q369 E370 I371 L372 E373

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

Chain E: 10%   80%   10%

T901 T902 T903 T904 T905 T906 T907 T908 T909 T910

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

Chain N: 10%   80%   10%

T901 T902 T903 T904 T905 T906 T907 T908 T909 T910

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	68.19Å 108.49Å 108.55Å 79.60° 71.80° 71.75°	Depositor
Resolution (Å)	50.00 – 3.70 47.74 – 3.67	Depositor EDS
% Data completeness (in resolution range)	96.4 (50.00-3.70) 95.1 (47.74-3.67)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.35 (at 3.67Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.250 , 0.267 0.258 , 0.272	Depositor DCC
R_{free} test set	2866 reflections (9.84%)	wwPDB-VP
Wilson B-factor (Å^2)	87.7	Xtriage
Anisotropy	0.340	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 33.7	EDS
L-test for twinning[2]	$< |L| > = 0.44, < L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.327 for h,h-k,h-l 0.307 for -h,-l,-k 0.317 for -h,-h+l,-h+k	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	12926	wwPDB-VP
Average B, all atoms (Å^2)	107.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.56% 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: ZN

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	0.37	0/1623	0.52	0/2204
1	B	0.36	0/1623	0.52	0/2204
1	C	0.38	0/1623	0.55	0/2204
1	D	0.38	0/1623	0.55	0/2204
1	J	0.36	0/1623	0.52	0/2204
1	K	0.37	0/1623	0.52	0/2204
1	L	0.38	0/1623	0.55	0/2204
1	M	0.38	0/1623	0.55	0/2204
2	E	0.72	1/220 (0.5%)	0.87	0/336
2	N	0.72	1/220 (0.5%)	0.86	0/336
All	All	0.39	2/13424 (0.0%)	0.55	0/18304

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	N	901	DT	OP3-P	-7.27	1.52	1.61
2	E	901	DT	OP3-P	-7.10	1.52	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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1565	0	1443	70	0
1	B	1565	0	1443	73	0
1	C	1565	0	1443	76	0
1	D	1565	0	1443	82	0
1	J	1565	0	1443	66	0
1	K	1565	0	1443	74	0
1	L	1565	0	1443	81	0
1	M	1565	0	1443	84	0
2	E	201	0	121	21	0
2	N	201	0	121	23	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	J	1	0	0	0	0
3	K	1	0	0	0	0
All	All	12926	0	11786	611	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 25.

The worst 5 of 611 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:K:351:PHE:HB3	2:N:908:DT:H72	1.55	0.87
1:B:207:LEU:HB2	1:B:210:ALA:HB2	1.57	0.87
1:K:207:LEU:HB2	1:K:210:ALA:HB2	1.57	0.87
1:A:207:LEU:HB2	1:A:210:ALA:HB2	1.58	0.85
1:J:207:LEU:HB2	1:J:210:ALA:HB2	1.59	0.85

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	
1	A	182/184 (99%)	141 (78%)	33 (18%)	8 (4%)	2	24
1	B	182/184 (99%)	141 (78%)	33 (18%)	8 (4%)	2	24
1	C	182/184 (99%)	142 (78%)	33 (18%)	7 (4%)	3	27
1	D	182/184 (99%)	142 (78%)	33 (18%)	7 (4%)	3	27
1	J	182/184 (99%)	141 (78%)	33 (18%)	8 (4%)	2	24
1	K	182/184 (99%)	141 (78%)	33 (18%)	8 (4%)	2	24
1	L	182/184 (99%)	142 (78%)	33 (18%)	7 (4%)	3	27
1	M	182/184 (99%)	142 (78%)	33 (18%)	7 (4%)	3	27
All	All	1456/1472 (99%)	1132 (78%)	264 (18%)	60 (4%)	3	26

5 of 60 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	210	ALA
1	A	353	PRO
1	B	210	ALA
1	B	353	PRO
1	C	208	ARG

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	169/169 (100%)	157 (93%)	12 (7%)	14	45
1	B	169/169 (100%)	157 (93%)	12 (7%)	14	45
1	C	169/169 (100%)	158 (94%)	11 (6%)	17	48
1	D	169/169 (100%)	157 (93%)	12 (7%)	14	45
1	J	169/169 (100%)	157 (93%)	12 (7%)	14	45
1	K	169/169 (100%)	157 (93%)	12 (7%)	14	45
1	L	169/169 (100%)	157 (93%)	12 (7%)	14	45
1	M	169/169 (100%)	158 (94%)	11 (6%)	17	48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1352/1352 (100%)	1258 (93%)	94 (7%)	15 45

5 of 94 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	J	359	TYR
1	L	193	GLU
1	K	199	ILE
1	K	290	PHE
1	L	240	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 65 such sidechains are listed below:

Mol	Chain	Res	Type
1	M	247	HIS
1	M	266	ASN
1	C	369	GLN
1	C	360	ASN
1	M	294	HIS

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 4 ligands modelled in this entry, 4 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 ⓘ

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	A	184/184 (100%)	-0.27	0	100	100	80, 103, 141, 157	0
1	B	184/184 (100%)	-0.28	2 (1%)	80	71	79, 104, 142, 157	0
1	C	184/184 (100%)	-0.21	1 (0%)	91	85	81, 105, 139, 155	0
1	D	184/184 (100%)	-0.16	0	100	100	77, 105, 140, 154	0
1	J	184/184 (100%)	-0.25	1 (0%)	91	85	80, 103, 143, 157	0
1	K	184/184 (100%)	-0.35	0	100	100	78, 104, 143, 157	0
1	L	184/184 (100%)	-0.23	0	100	100	79, 105, 139, 154	0
1	M	184/184 (100%)	-0.20	1 (0%)	91	85	79, 105, 139, 155	0
2	E	10/10 (100%)	-0.34	0	100	100	82, 100, 156, 177	0
2	N	10/10 (100%)	-0.37	0	100	100	83, 102, 156, 177	0
All	All	1492/1492 (100%)	-0.25	5 (0%)	94	90	77, 105, 143, 177	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	242	VAL	4.1
1	C	279	PRO	3.8
1	B	241	GLN	2.5
1	B	242	VAL	2.4
1	M	279	PRO	2.4

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, 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
3	ZN	B	801	1/1	0.69	0.08	221,221,221,221	0
3	ZN	K	801	1/1	0.69	0.14	221,221,221,221	0
3	ZN	A	801	1/1	0.72	0.12	225,225,225,225	0
3	ZN	J	801	1/1	0.77	0.10	225,225,225,225	0

6.5 Other polymers ⓘ

There are no such residues in this entry.