



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 14, 2023 – 10:00 PM JST

PDB ID : 6AEB  
Title : Crystal structure of xCas9 in complex with sgRNA and target DNA (AAG PAM)  
Authors : Guo, M.; Ren, K.; Zhu, Y.; Huang, Z.  
Deposited on : 2018-08-04  
Resolution : 3.00 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](mailto: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 ⓘ 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 ⓘ](#)) 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

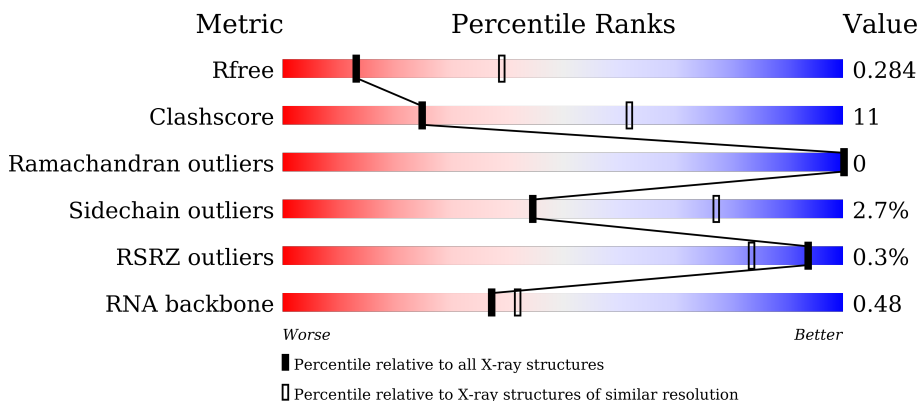
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	100	
1	E	100	
2	C	28	
2	G	28	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	D	11	 55% 45%
3	H	11	 45% 55%
4	B	1368	 72% 23% ..
4	F	1368	 72% 24% ..

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 26351 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 RNA chain called RNA (95-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	95	2032	911	372	654	95	0	0	0
1	E	95	2032	911	372	654	95	0	0	0

- Molecule 2 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	25	506	247	83	152	24	0	0	0
2	G	25	506	247	83	152	24	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*TP\*AP\*TP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	11	227	110	46	61	10	0	0	0
3	H	11	227	110	46	61	10	0	0	0

- Molecule 4 is a protein called DNA Nuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	1322	10411	6656	1811	1923	21	0	0	0
4	F	1322	10409	6655	1811	1922	21	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	262	THR	ALA	engineered mutation	UNP Q99ZW2
B	324	LEU	ARG	engineered mutation	UNP Q99ZW2
B	409	ILE	SER	engineered mutation	UNP Q99ZW2
B	480	LYS	GLU	engineered mutation	UNP Q99ZW2
B	543	ASP	GLU	engineered mutation	UNP Q99ZW2
B	694	ILE	MET	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
B	1219	VAL	GLU	engineered mutation	UNP Q99ZW2
F	10	ALA	ASP	engineered mutation	UNP Q99ZW2
F	262	THR	ALA	engineered mutation	UNP Q99ZW2
F	324	LEU	ARG	engineered mutation	UNP Q99ZW2
F	409	ILE	SER	engineered mutation	UNP Q99ZW2
F	480	LYS	GLU	engineered mutation	UNP Q99ZW2
F	543	ASP	GLU	engineered mutation	UNP Q99ZW2
F	694	ILE	MET	engineered mutation	UNP Q99ZW2
F	840	ALA	HIS	engineered mutation	UNP Q99ZW2
F	1219	VAL	GLU	engineered mutation	UNP Q99ZW2

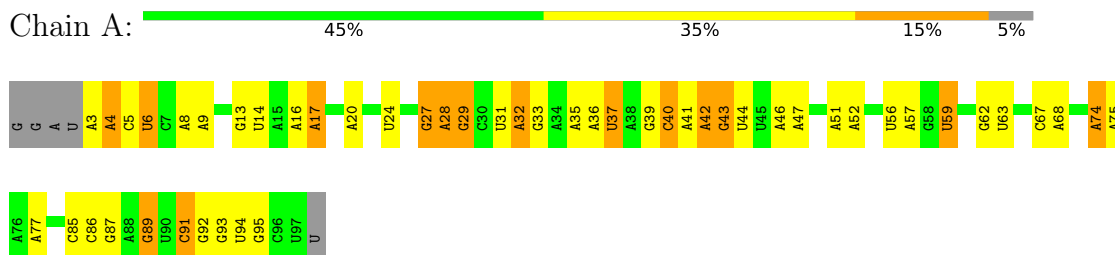
- Molecule 5 is water.

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

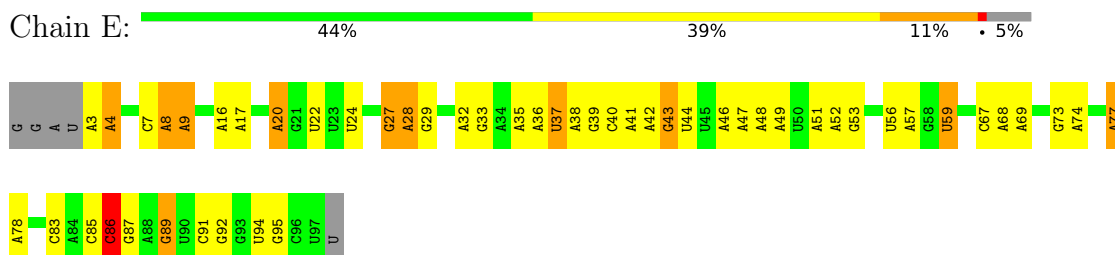
### 3 Residue-property plots [i](#)

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.

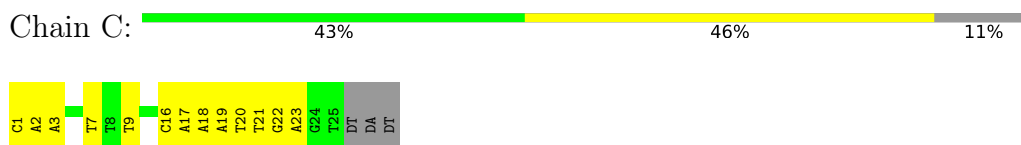
- Molecule 1: RNA (95-MER)



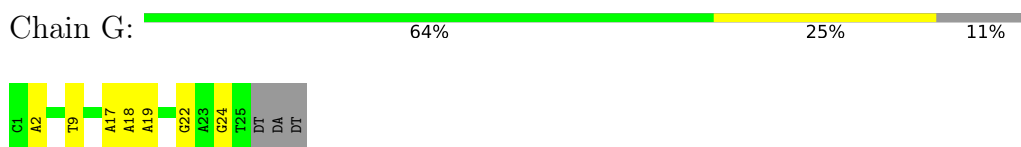
- Molecule 1: RNA (95-MER)



- Molecule 2: DNA (25-MER)

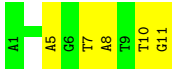


- Molecule 2: DNA (25-MER)

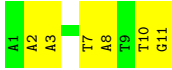


- Molecule 3: DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*TP\*AP\*TP\*TP\*G)-3')

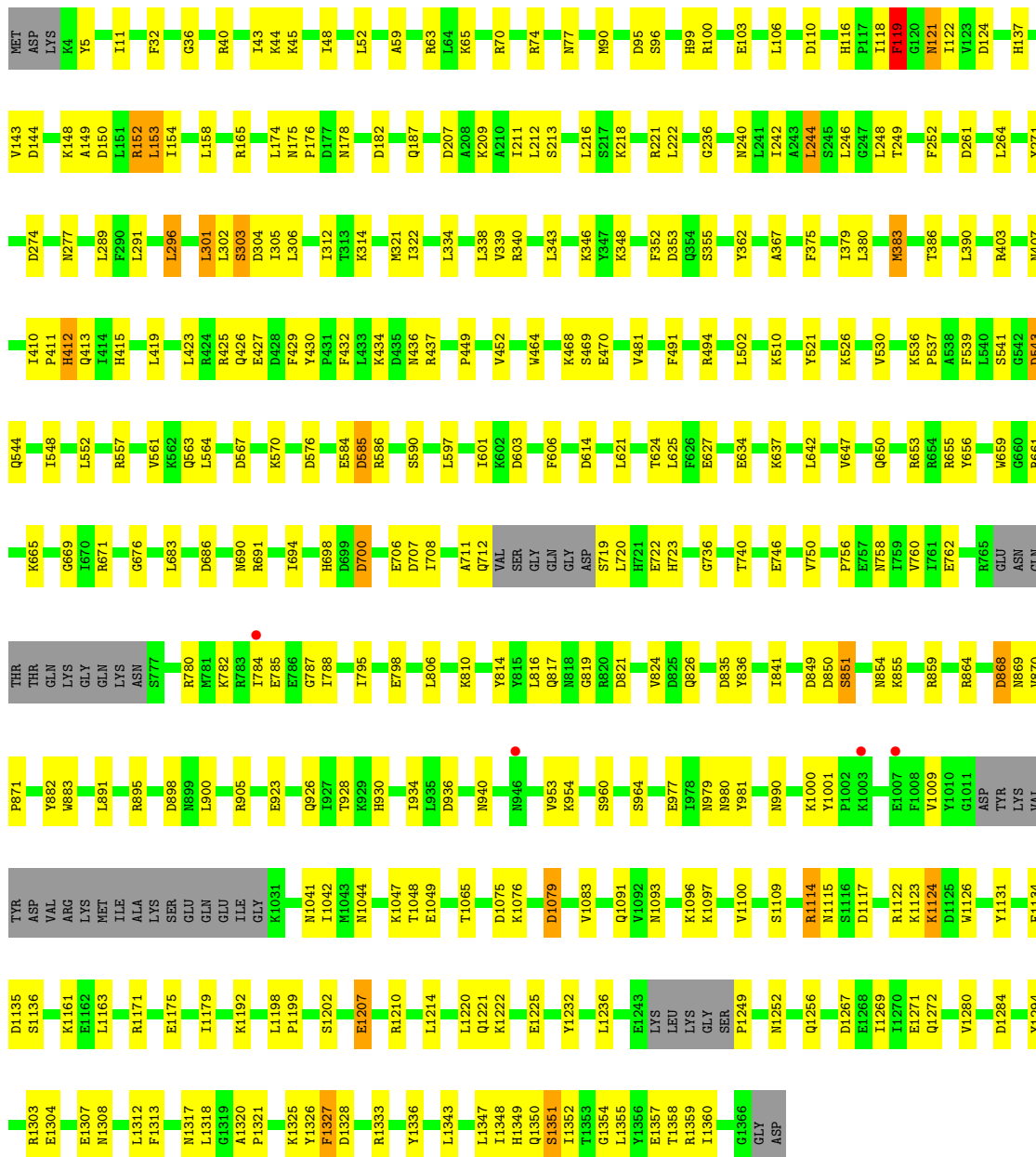




● Molecule 3: DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*TP\*AP\*TP\*TP\*G)-3')



● Molecule 4: DNA Nuclease



● Molecule 4: DNA Nuclease

Chain F:  72% 24%

MET	ASP	LYS	K4	G8	I11	T22	Y25	K31	N37	R40	K44	K45	M46	L47	I48	K65	R70	R74	L82	F86	M90	D85	S96	F97	R100	E103	V107	K111	K112	I118	F119	G120	N121	D124	I135	Y136	H137																							
D144	K148	R152	H160	M161	I162	R165	I170	L174	N175	P176	S179	F185	I186	Q187	L188	V189	A203	K209	L212	S213	S217	K218	R221	G236	N240	L244	S245	L246	T249	L264	Q265	Y271	D274	N277	L278	I282																								
A287	L291	I300	L301	L302	I305	K434	I312	F313	K314	I322	L334	L338	V339	R340	L343	P344	I350	F351	F352	D353	G354	S355	S368	E371	I379	L380	M383	L389	L390	E396	R400	R403	T404	F405	H412	Q413																								
R425	Q426	E427	D428	F429	F432	L433	K434	D435	M436	R437	E441	T445	P449	V452	S469	E470	E471	T472	I473	T474	P475	W476	M477	S490	T496	M497	L514	F518	L524	L651	T525	K526	R553	R554	R555	K536	D543	Q544	I548	L552	R557																			
V561	Q563	D567	E573	E584	D585	R586	T593	V594	H595	D596	L601	D605	F606	L607	D608	M612	T615	T624	L625	F626	E627	D628	R629	E630	R631	V639	D644	D645	K646	V647	Q650	L651	K652	R653	R654	R655	Y656	W659	K665	L666	I667	L670																		
R671	D672	S675	G676	K677	F682	N690	Q695	I697	H698	D699	D700	S701	A711	Q712	VAL	SER	GLY	GLN	GLY	ASP	S719	L720	G729	I733	K734	R737	T740	K755	V760	R765	GLU	ASN	GLN	THR	THR	GLN	LYS	LYS	GLN	LYS	ASN	S777	R778	E779	R780															
L784	E785	I788	H799	N803	L806	E809	Y812	Y815	H899	Q817	Y824	R825	Q826	E827	L828	D829	I830	R831	R832	L833	D835	I841	F846	L847	D853	N854	K855	R859	R864	N869	V870	P871	Y882	W883	I892	T893	R894	R895	K896	L900	T901																			
K902	L908	K918	R919	Q920	L921	V922	E923	T928	M939	N940	V953	K961	R967	Q971	F972	R976	E977	I978	D986	Y988	L989	N990	A991	V992	V993	L997	K1000	Y1001	S1006	E1007	G1011	ASP	TYR	LYS	VAL	TYR	ASP	VAL	ARG	LYS	MET	ILE	ALA	LYS																
SER	GLU	GLN	ILE	GLY	K1031	Y1039	I1042	M1043	N1044	F1045	F1046	K1047	A1053	N1054	R1060	P1061	L1062	I1063	E1064	T1065	N1066	T1069	D1075	K1076	V1083	R1084	K1085	V1086	L1087	S1088	N1093	K1097	K1107	I1110	R1114	N1115	S1116	D1117	I1120	F1134	D1135	S1136	K1148																	
V1149	L1157	K1158	S1159	V1160	K1161	E1162	L1163	M1169	L1194	K1200	Y1201	E1205	L1206	E1207	R1210	K1211	R1212	M1213	L1214	A1215	S1216	L1220	G1221	K1222	V1233	N1234	F1235	L1236	S1240	E1243	LYS	LEU	LYS	SER	P1249	M1252	H1264	D1267	I1270	F1271	Q1272	D1284	A1285																	
M1286	Y1294	M1295	K1296	H1297	R1298	D1299	K1300	L1312	P1321	F1324	K1325	D1328	D1332	R1333	Y1336	E1341	V1342	L1343	S1351	G1354	L1355	Y1356	E1357	R1359	I1360	D1361	L1365	G1366	GLY	ASP																														



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	361.85Å 71.13Å 198.32Å 90.00° 101.75° 90.00°	Depositor
Resolution (Å)	49.45 – 3.00 49.45 – 3.01	Depositor EDS
% Data completeness (in resolution range)	97.7 (49.45-3.00) 86.1 (49.45-3.01)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.74 (at 3.01Å)	Xtrriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, $R_{free}$	0.231 , 0.283 0.232 , 0.284	Depositor DCC
$R_{free}$ test set	1981 reflections (2.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	62.2	Xtrriage
Anisotropy	0.704	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 41.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	26351	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 41.30 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.4091e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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

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	0.35	0/2277	1.00	0/3546
1	E	0.36	0/2277	0.99	1/3546 (0.0%)
2	C	0.77	0/565	1.19	0/870
2	G	0.77	0/565	1.17	0/870
3	D	0.73	0/256	0.98	0/394
3	H	0.74	0/256	0.99	0/394
4	B	0.36	1/10600 (0.0%)	0.52	11/14316 (0.1%)
4	F	0.34	0/10598	0.50	4/14313 (0.0%)
All	All	0.39	1/27394 (0.0%)	0.68	16/38249 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
4	B	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	119	PHE	C-N	-6.62	1.21	1.33

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	F	175	ASN	C-N-CD	-10.49	97.52	120.60
4	B	119	PHE	N-CA-C	-8.90	86.98	111.00
4	B	119	PHE	O-C-N	-8.07	109.48	123.20
4	B	118	ILE	CB-CA-C	-6.44	98.71	111.60
4	B	586	ARG	N-CA-CB	-6.35	99.17	110.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	119	PHE	Mainchain

## 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	2032	0	1021	42	0
1	E	2032	0	1021	27	0
2	C	506	0	289	18	0
2	G	506	0	289	9	0
3	D	227	0	126	7	0
3	H	227	0	126	5	0
4	B	10411	0	10286	238	0
4	F	10409	0	10281	224	0
5	B	1	0	0	0	0
All	All	26351	0	23439	517	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 11.

The worst 5 of 517 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:63:U:N3	4:B:65:LYS:NZ	1.95	1.13
4:F:1047:LYS:O	4:F:1076:LYS:NZ	1.89	1.05
4:B:150:ASP:OD1	4:B:152:ARG:HG2	1.69	0.93
4:F:1062:LEU:O	4:F:1076:LYS:HG3	1.69	0.92
4:B:561:VAL:HG23	4:B:585:ASP:O	1.70	0.91

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	B	1312/1368 (96%)	1188 (90%)	124 (10%)	0	100	100
4	F	1312/1368 (96%)	1181 (90%)	131 (10%)	0	100	100
All	All	2624/2736 (96%)	2369 (90%)	255 (10%)	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	
4	B	1086/1226 (89%)	1052 (97%)	34 (3%)	40	75
4	F	1085/1226 (88%)	1061 (98%)	24 (2%)	52	81
All	All	2171/2452 (88%)	2113 (97%)	58 (3%)	44	77

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	1124	LYS
4	F	1234	ASN
4	F	65	LYS
4	F	1222	LYS
4	F	847	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
4	F	1264	HIS
4	F	1308	ASN
4	F	394	ASN
4	F	497	ASN
4	F	650	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	94/100 (94%)	28 (29%)	4 (4%)
1	E	94/100 (94%)	28 (29%)	5 (5%)
All	All	188/200 (94%)	56 (29%)	9 (4%)

5 of 56 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	4	A
1	A	6	U
1	A	9	A
1	A	17	A
1	A	20	A

5 of 9 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	E	42	A
1	E	68	A
1	A	42	A
1	E	8	A
1	E	27	G

## 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](#)

There are no ligands in this entry.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9	
1	A	95/100 (95%)	-0.11	0	100 100	51, 79, 116, 149	0
1	E	95/100 (95%)	-0.12	0	100 100	49, 79, 118, 148	0
2	C	25/28 (89%)	-0.23	0	100 100	58, 67, 105, 120	0
2	G	25/28 (89%)	-0.14	0	100 100	56, 68, 112, 126	0
3	D	11/11 (100%)	0.10	0	100 100	53, 72, 117, 118	0
3	H	11/11 (100%)	0.05	0	100 100	55, 69, 115, 121	0
4	B	1322/1368 (96%)	-0.07	4 (0%)	94 84	51, 79, 103, 138	0
4	F	1322/1368 (96%)	-0.05	6 (0%)	91 75	46, 79, 104, 134	0
All	All	2906/3014 (96%)	-0.07	10 (0%)	94 84	46, 79, 105, 149	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	F	1053	ALA	3.3
4	B	1003	LYS	3.0
4	F	1312	LEU	2.8
4	F	1054	ASN	2.6
4	B	946	ASN	2.4

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.