



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2022 – 04:55 pm BST

PDB ID : 7Z4D
Title : Crystal structure of SpCas9 bound to a 10 nucleotide complementary DNA substrate
Authors : Pacesa, M.; Jinek, M.
Deposited on : 2022-03-03
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

<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
Xtrriage (Phenix) : 1.13
EDS : 2.29
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.29

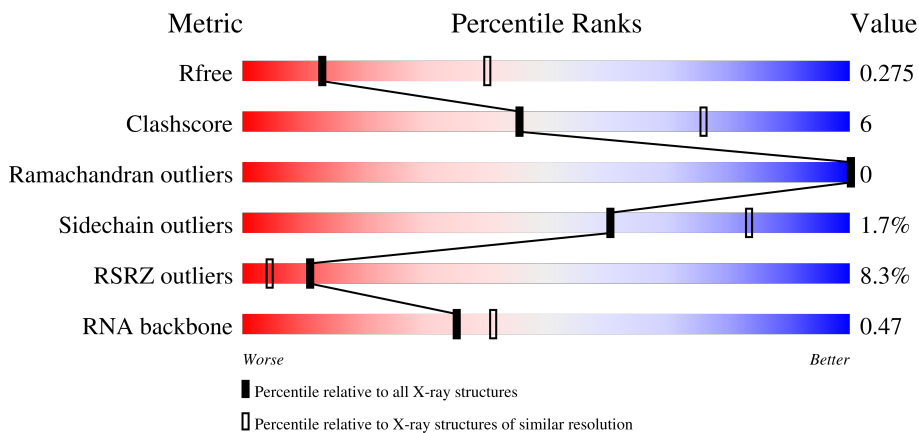
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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.



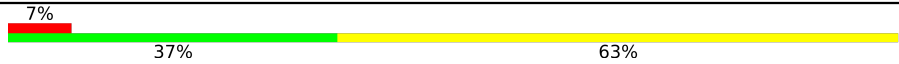


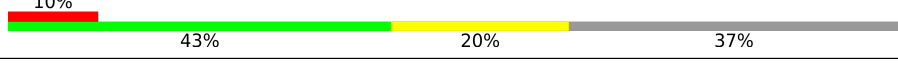
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)
RNA backbone	3102	1116 (3.40-2.80)

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 $\leq 5\%$. 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	84	 39% 43% 7% 11%
1	F	84	 6% 35% 44% 13% 8%
2	B	1368	 8% 79% 13% 8%
2	E	1368	 7% 71% 14% 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	30	
3	G	30	
4	D	30	
4	H	30	

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 25082 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 sgRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	75	Total 1610	C 723	N 303	O 510	P 74	0	0	0
1	F	77	Total 1652	C 742	N 310	O 524	P 76	0	0	0

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1259	Total 10259	C 6549	N 1769	O 1921	S 20	0	0	0
2	E	1166	Total 9537	C 6102	N 1640	O 1776	S 19	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
E	10	ALA	ASP	engineered mutation	UNP Q99ZW2
E	840	ALA	HIS	engineered mutation	UNP Q99ZW2

- Molecule 3 is a DNA chain called Target strand of 10 nucleotide complementary DNA substrate.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	30	Total 606	C 291	N 105	O 181	P 29	0	0	0
3	G	27	Total 526	C 252	N 90	O 159	P 25	0	0	1

- Molecule 4 is a DNA chain called Non-target strand of 10 nucleotide complementary DNA substrate.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	22	Total	C	N	O	P	0	0	1
			429	207	75	127	20			
4	H	19	Total	C	N	O	P	0	0	1
			359	172	64	106	17			

- Molecule 5 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	6	Total	K	0	0
			6	6		
5	B	3	Total	K	0	0
			3	3		
5	C	1	Total	K	0	0
			1	1		
5	E	2	Total	K	0	0
			2	2		
5	F	3	Total	K	0	0
			3	3		
5	G	1	Total	K	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	15	Total	O	0	0
			15	15		
6	B	22	Total	O	0	0
			22	22		
6	C	1	Total	O	0	0
			1	1		
6	D	2	Total	O	0	0
			2	2		
6	E	29	Total	O	0	0
			29	29		
6	F	14	Total	O	0	0
			14	14		
6	G	1	Total	O	0	0
			1	1		
6	H	4	Total	O	0	0
			4	4		

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.

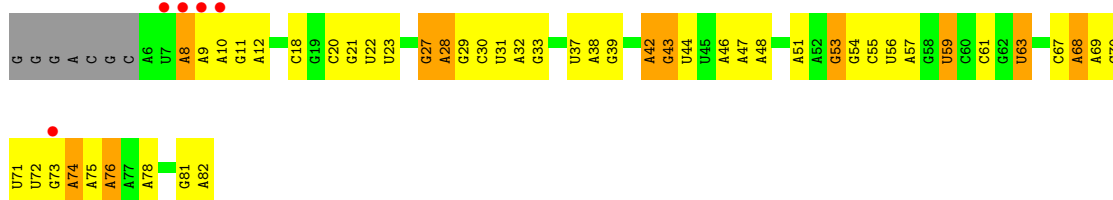
- Molecule 1: sgRNA

Chain A: 




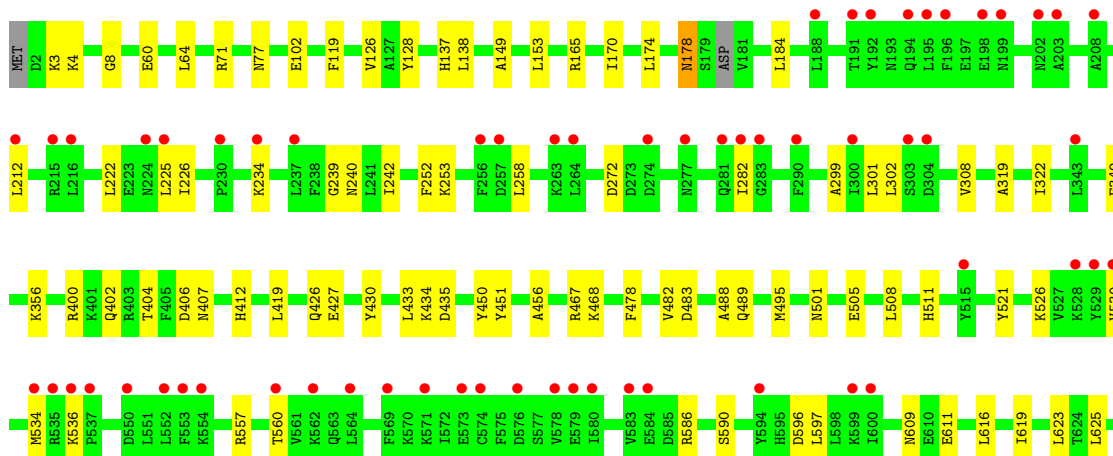
- Molecule 1: sgRNA

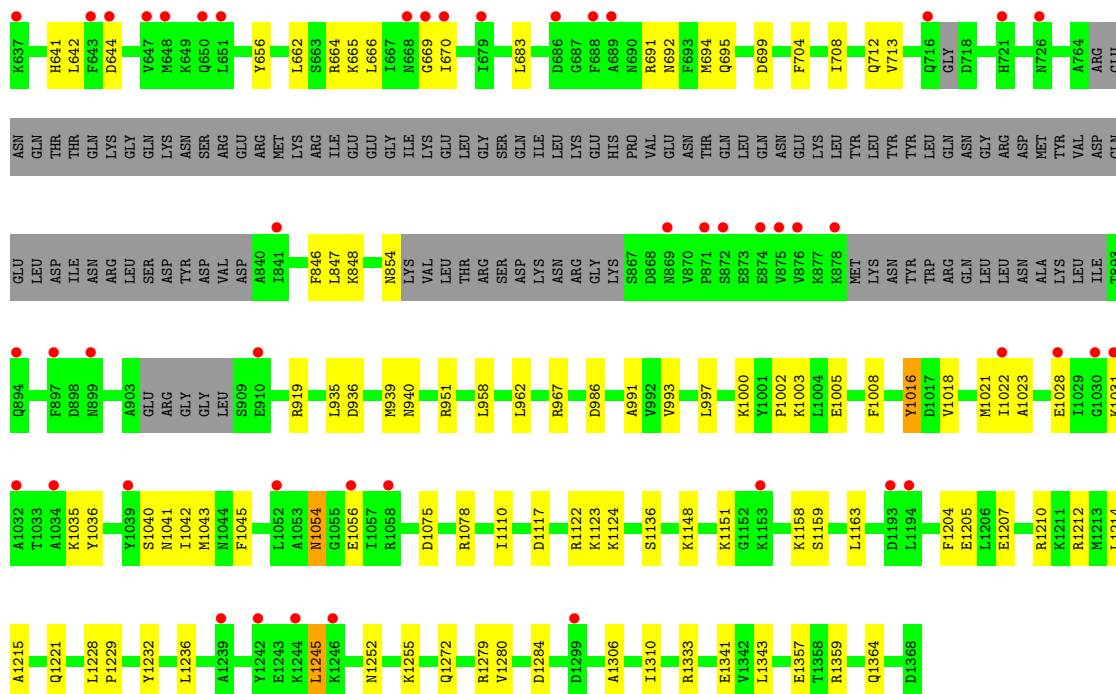
Chain F: 



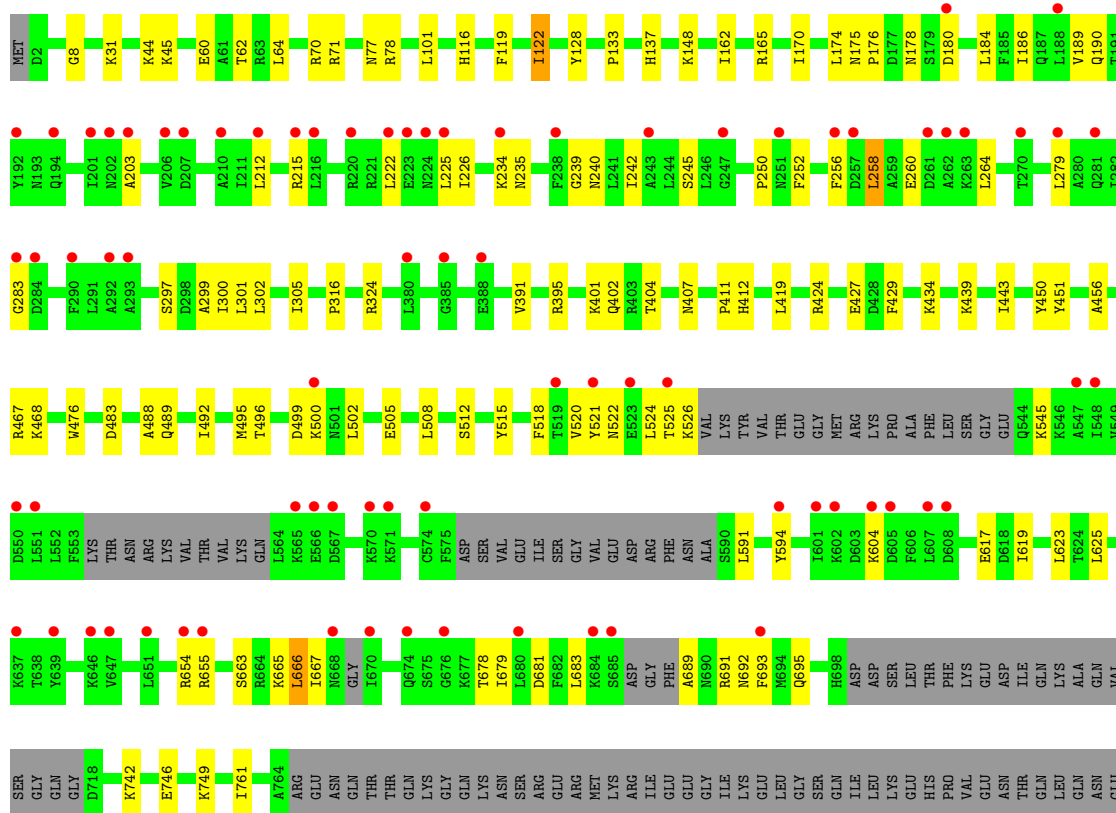
- Molecule 2: CRISPR-associated endonuclease Cas9/Csn1

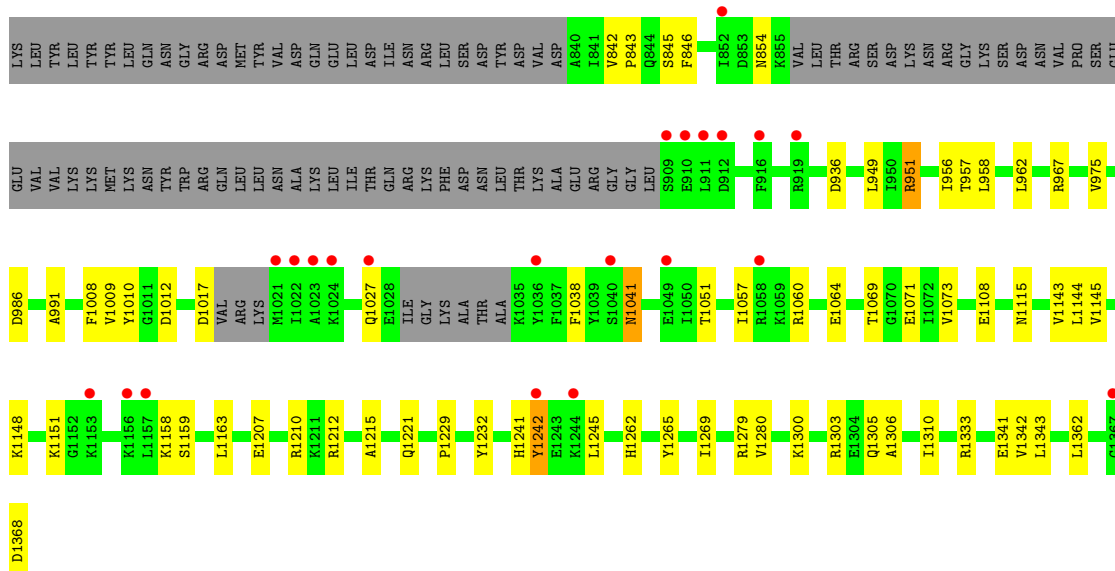
Chain B: 





- Molecule 2: CRISPR-associated endonuclease Cas9/Csn1

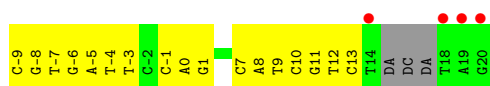
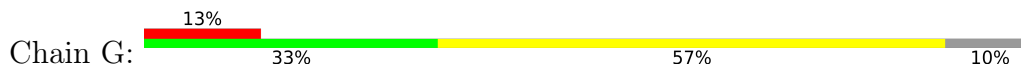




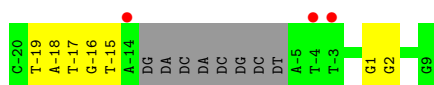
- Molecule 3: Target strand of 10 nucleotide complementary DNA substrate



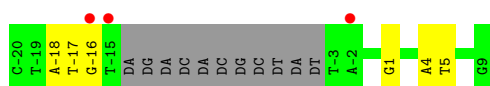
- Molecule 3: Target strand of 10 nucleotide complementary DNA substrate



- Molecule 4: Non-target strand of 10 nucleotide complementary DNA substrate



- Molecule 4: Non-target strand of 10 nucleotide complementary DNA substrate



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	88.64Å 95.51Å 169.80Å 80.83° 78.04° 62.34°	Depositor
Resolution (Å)	66.49 – 3.10 66.48 – 3.10	Depositor EDS
% Data completeness (in resolution range)	89.2 (66.49-3.10) 89.2 (66.48-3.10)	Depositor EDS
R_{merge}	0.34	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.86 (at 3.13Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.246 , 0.276 0.247 , 0.275	Depositor DCC
R_{free} test set	3882 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å ²)	52.8	Xtriage
Anisotropy	0.009	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	25082	wwPDB-VP
Average B, all atoms (Å ²)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.13% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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

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	0.26	0/1807	0.84	0/2816
1	F	0.26	0/1854	0.82	0/2889
2	B	0.25	1/10440 (0.0%)	0.48	0/14026
2	E	0.25	0/9706	0.47	1/13036 (0.0%)
3	C	0.59	0/677	0.94	0/1042
3	G	0.55	0/586	0.94	0/901
4	D	0.56	0/479	0.95	0/737
4	H	0.56	0/401	0.95	0/616
All	All	0.29	1/25950 (0.0%)	0.60	1/36063 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1021	MET	CG-SD	5.24	1.94	1.81

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	666	LEU	CA-CB-CG	5.29	127.47	115.30

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	A	1610	0	809	27	0
1	F	1652	0	830	37	0
2	B	10259	0	10425	102	0
2	E	9537	0	9689	118	0
3	C	606	0	341	18	0
3	G	526	0	297	18	0
4	D	429	0	242	5	0
4	H	359	0	201	7	0
5	A	6	0	0	0	0
5	B	3	0	0	0	0
5	C	1	0	0	0	0
5	E	2	0	0	0	0
5	F	3	0	0	0	0
5	G	1	0	0	0	0
6	A	15	0	0	0	0
6	B	22	0	0	1	0
6	C	1	0	0	0	0
6	D	2	0	0	0	0
6	E	29	0	0	1	0
6	F	14	0	0	1	0
6	G	1	0	0	0	0
6	H	4	0	0	0	0
All	All	25082	0	22834	307	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 6.

The worst 5 of 307 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:10:A:H62	3:C:11:DG:H1	1.14	0.92
2:E:1051:THR:HG22	2:E:1057:ILE:HG22	1.65	0.78
2:B:847:LEU:HD21	2:B:919:ARG:HH11	1.49	0.76
3:G:-8:DG:H2''	3:G:-7:DT:H5'	1.68	0.75
2:E:1148:LYS:HG2	2:E:1159:SER:HA	1.66	0.75

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	
2	B	1245/1368 (91%)	1189 (96%)	56 (4%)	0	100	100
2	E	1144/1368 (84%)	1102 (96%)	42 (4%)	0	100	100
All	All	2389/2736 (87%)	2291 (96%)	98 (4%)	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	
2	B	1125/1225 (92%)	1105 (98%)	20 (2%)	59	82
2	E	1045/1225 (85%)	1028 (98%)	17 (2%)	62	84
All	All	2170/2450 (89%)	2133 (98%)	37 (2%)	60	83

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	951	ARG
2	E	1242	TYR
2	E	1012	ASP
2	E	1115	ASN
2	B	1016	TYR

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

Mol	Chain	Res	Type
2	E	1027	GLN
2	E	1044	ASN
2	B	854	ASN
2	B	1044	ASN
2	B	1054	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	74/84 (88%)	18 (24%)	2 (2%)
1	F	76/84 (90%)	18 (23%)	3 (3%)
All	All	150/168 (89%)	36 (24%)	5 (3%)

5 of 36 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	20	C
1	A	28	A
1	A	29	G
1	A	37	U
1	A	51	A

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	27	G
1	A	28	A
1	F	27	G
1	F	28	A
1	F	42	A

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](#)

Of 16 ligands modelled in this entry, 16 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 [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	A	75/84 (89%)	0.39	0 100 100	22, 43, 142, 151	0
1	F	77/84 (91%)	0.63	5 (6%) 18 8	25, 46, 153, 210	0
2	B	1259/1368 (92%)	0.40	108 (8%) 10 4	18, 67, 118, 149	0
2	E	1166/1368 (85%)	0.44	98 (8%) 11 4	21, 65, 117, 142	0
3	C	30/30 (100%)	0.38	2 (6%) 17 7	36, 61, 129, 140	0
3	G	27/30 (90%)	0.54	4 (14%) 2 1	32, 55, 126, 130	0
4	D	22/30 (73%)	0.71	3 (13%) 3 1	32, 89, 137, 141	0
4	H	19/30 (63%)	0.64	3 (15%) 2 1	33, 83, 122, 141	0
All	All	2675/3024 (88%)	0.43	223 (8%) 11 4	18, 64, 119, 210	0

The worst 5 of 223 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	207	ASP	6.2
2	E	224	ASN	6.1
2	E	281	GLN	5.9
2	E	216	LEU	5.5
2	E	912	ASP	5.5

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 [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
5	K	F	103	1/1	0.73	0.19	86,86,86,86	0
5	K	F	101	1/1	0.89	0.19	93,93,93,93	0
5	K	A	106	1/1	0.89	0.24	80,80,80,80	0
5	K	B	1402	1/1	0.90	0.21	60,60,60,60	0
5	K	A	105	1/1	0.92	0.20	73,73,73,73	0
5	K	A	104	1/1	0.93	0.21	68,68,68,68	0
5	K	A	101	1/1	0.94	0.15	98,98,98,98	0
5	K	E	1401	1/1	0.94	0.16	61,61,61,61	0
5	K	A	103	1/1	0.95	0.18	52,52,52,52	0
5	K	B	1401	1/1	0.95	0.31	53,53,53,53	0
5	K	A	102	1/1	0.96	0.15	69,69,69,69	0
5	K	B	1403	1/1	0.96	0.17	56,56,56,56	0
5	K	C	101	1/1	0.96	0.24	43,43,43,43	0
5	K	E	1402	1/1	0.97	0.24	47,47,47,47	0
5	K	G	101	1/1	0.97	0.21	48,48,48,48	0
5	K	F	102	1/1	0.99	0.18	58,58,58,58	0

6.5 Other polymers [i](#)

There are no such residues in this entry.