wwPDB X-ray Structure Validation Summary Report (i)

Jun 22, 2024 – 12:01 PM EDT

PDB ID	:	4QYZ
Title	:	Crystal structure of a CRISPR RNA-guided surveillance complex, Cascade, bound to a ssDNA target
Authors	:	Mulepati, S.; Bailey, S.
Deposited on	:	2014-07-26
Resolution	:	3.03 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 (i) 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 (1)) were used in the production of this report:

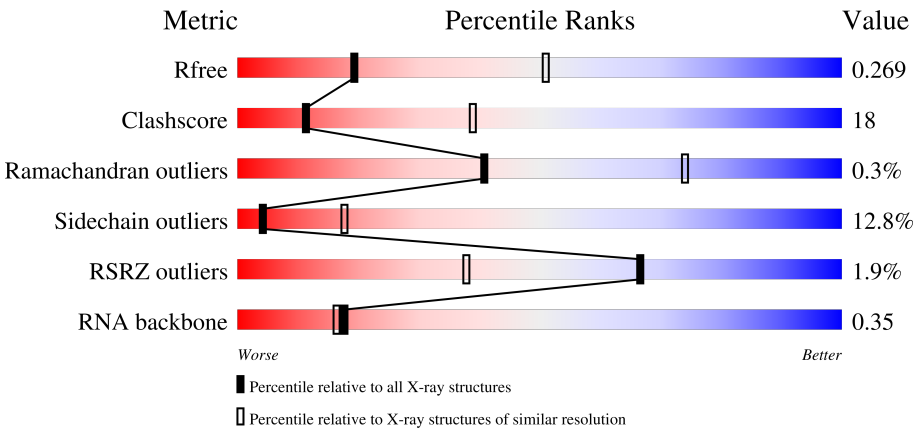
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.03 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	2752 (3.08-3.00)
Clashscore	141614	3096 (3.08-3.00)
Ramachandran outliers	138981	2986 (3.08-3.00)
Sidechain outliers	138945	2988 (3.08-3.00)
RSRZ outliers	127900	2636 (3.08-3.00)
RNA backbone	3102	1034 (3.30-2.78)

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	502	.% 52% 34% 7% 7%
2	B	160	58% 32% • 6%
2	C	160	52% 36% 5% • 6%
3	D	363	2% 52% 33% 9% • 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	363	%, 62%, 33%
3	F	363	%, 63%, 34%
3	G	363	61%, 34%
3	H	363	64%, 31%
3	I	363	%, 43%, 25%, 29%
4	J	224	%, 37%, 41%, 9%, 13%
5	K	199	18%, 43%, 30%, 24%
6	L	61	3%, 13%, 48%, 23%, 7%, 10%
7	M	40	32%, 48%, 18%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 26543 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 CRISPR system Cascade subunit CasA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	468	3702	2368	653	662	19	0	0	0

- Molecule 2 is a protein called CRISPR system Cascade subunit CasB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	150	1238	777	239	215	7	0	0	0
2	C	150	1238	777	239	215	7	0	0	0

- Molecule 3 is a protein called CRISPR system Cascade subunit CasC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	340	2630	1643	469	502	16	0	0	0
3	E	361	2794	1747	496	535	16	0	0	0
3	F	362	2803	1752	497	538	16	0	0	0
3	G	360	2788	1742	495	535	16	0	0	0
3	H	359	2781	1738	494	533	16	0	0	0
3	I	256	1978	1231	354	380	13	0	0	0

- Molecule 4 is a protein called CRISPR system Cascade subunit CasD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	J	195	1545	987	273	276	9	0	0	0

- Molecule 5 is a protein called CRISPR system Cascade subunit CasE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	K	152	1197	777	209	206	5	0	0	0

- Molecule 6 is a RNA chain called RNA (55-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	L	55	1178	526	216	382	54	0	0	0

- Molecule 7 is a DNA chain called DNA (33-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	M	33	670	319	122	196	33	0	0	0

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

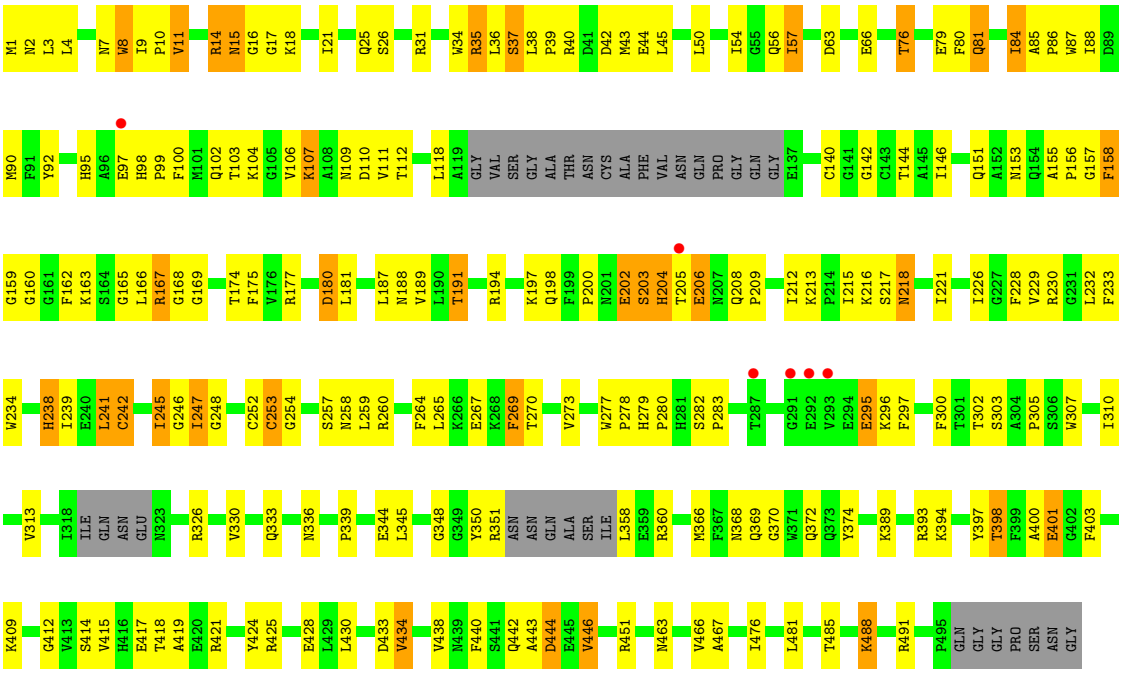
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
8	A	1	1	1	0	0

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.

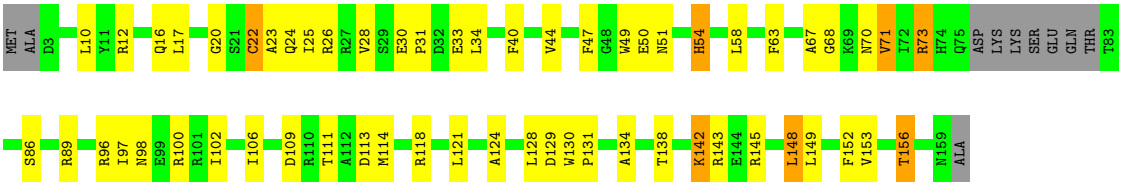
• Molecule 1: CRISPR system Cascade subunit CasA

Chain A:



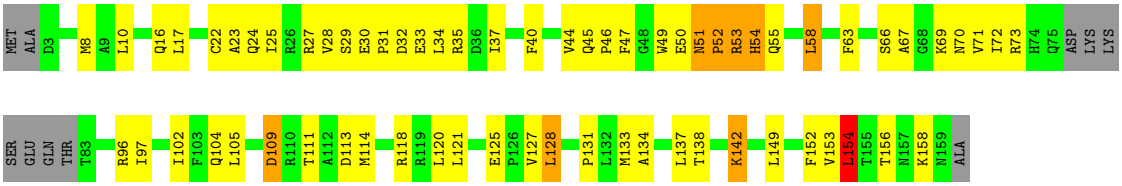
• Molecule 2: CRISPR system Cascade subunit CasB

Chain B:



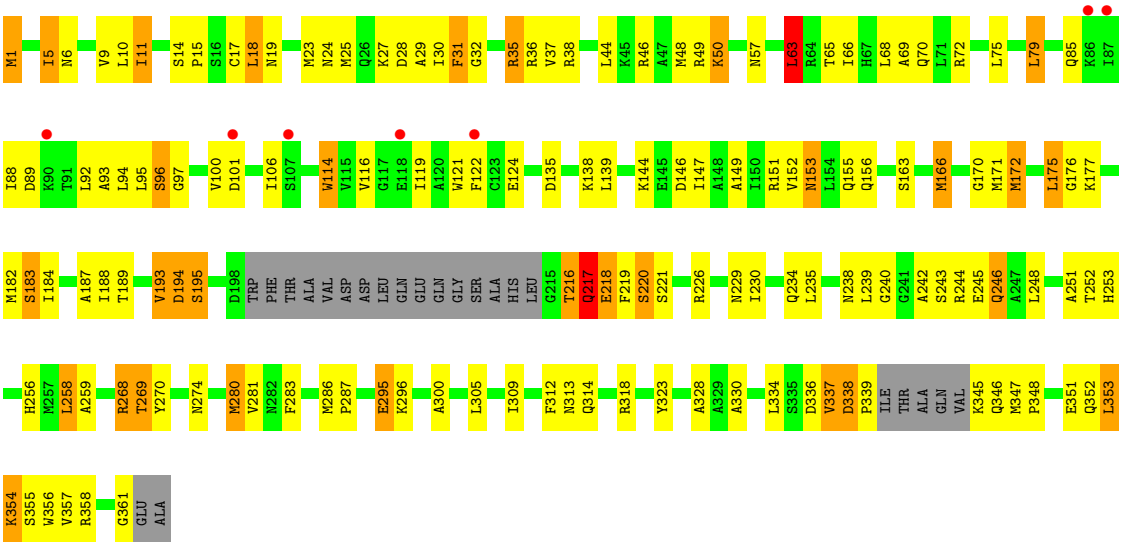
• Molecule 2: CRISPR system Cascade subunit CasB

Chain C:



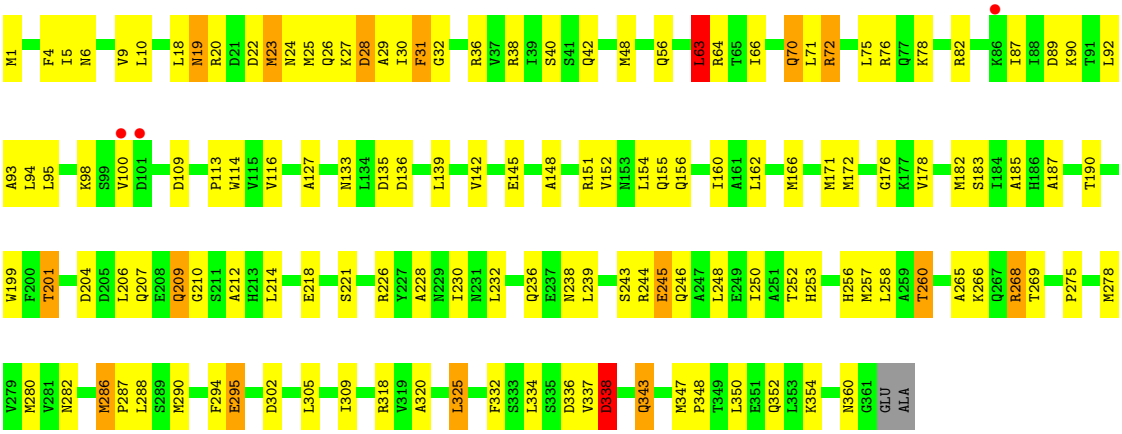
• Molecule 3: CRISPR system Cascade subunit CasC

Chain D:



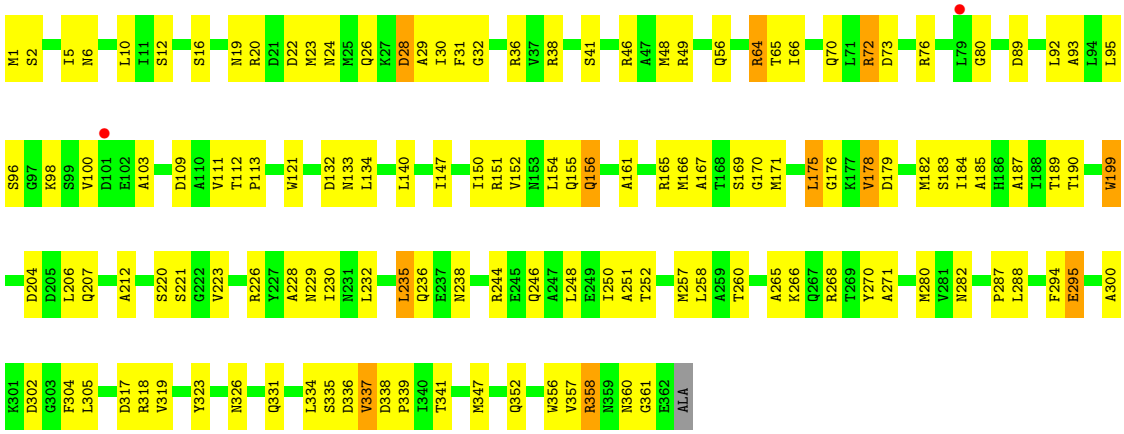
• Molecule 3: CRISPR system Cascade subunit CasC

Chain E:



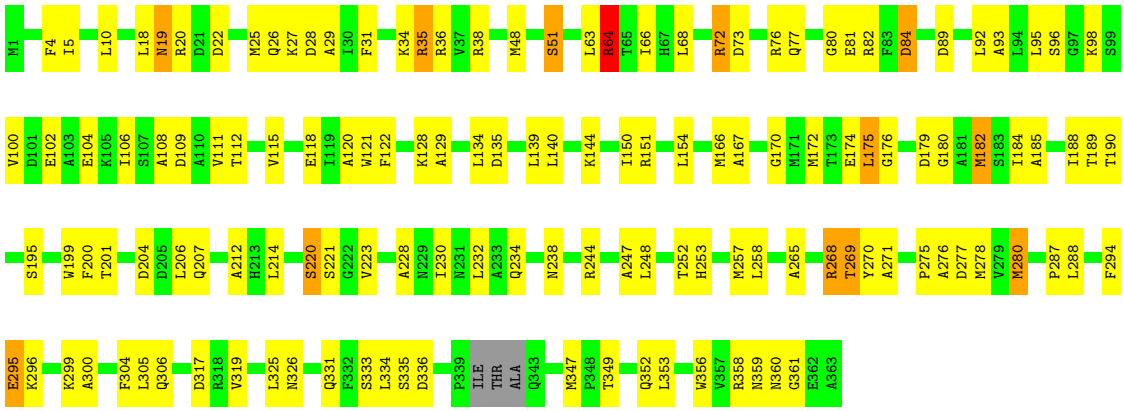
• Molecule 3: CRISPR system Cascade subunit CasC

Chain F:



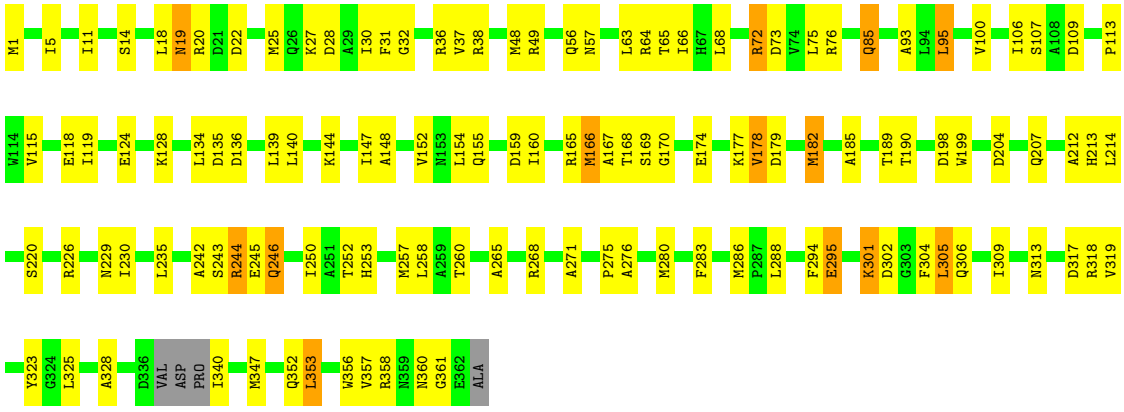
• Molecule 3: CRISPR system Cascade subunit CasC

Chain G:



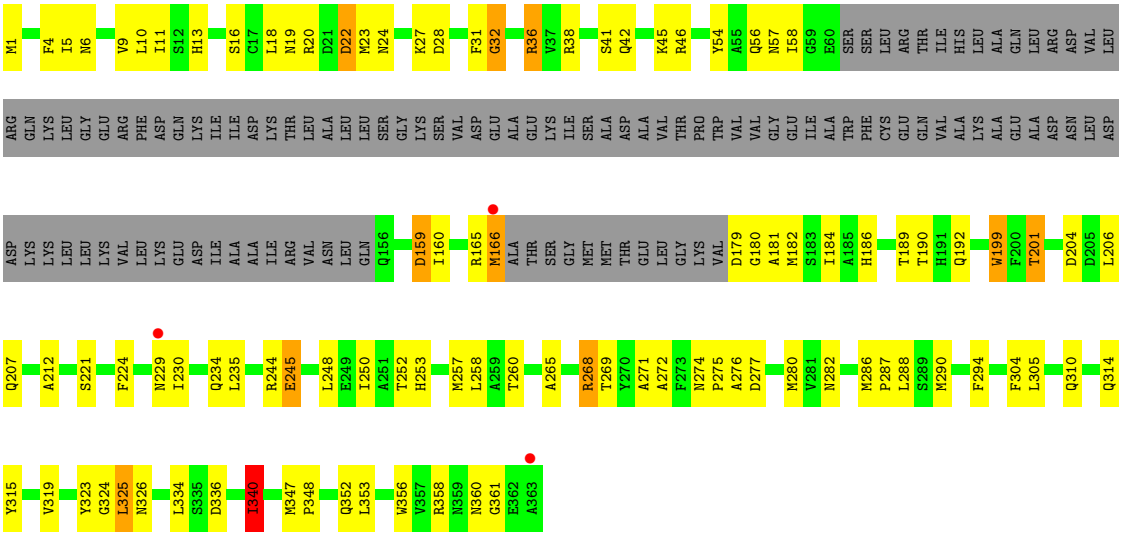
• Molecule 3: CRISPR system Cascade subunit CasC

Chain H:



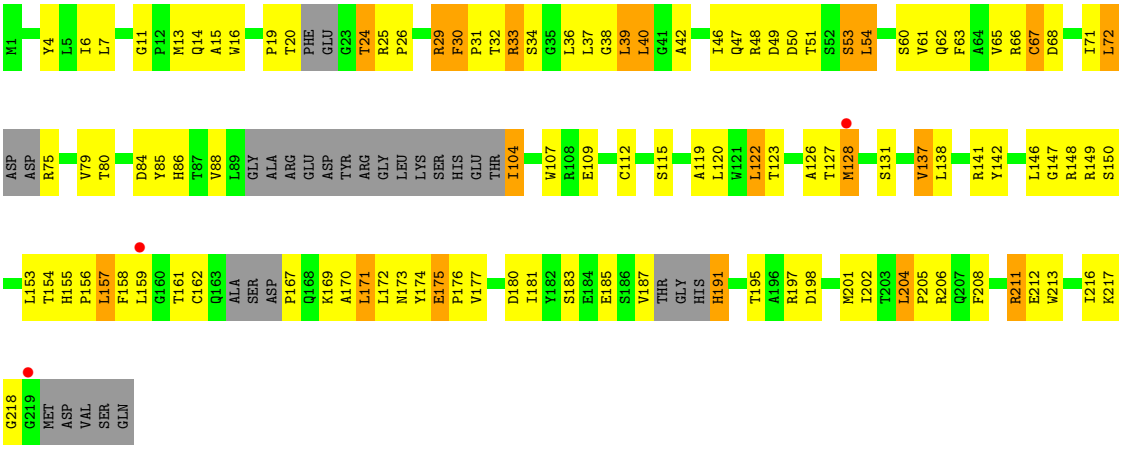
• Molecule 3: CRISPR system Cascade subunit CasC

Chain I:



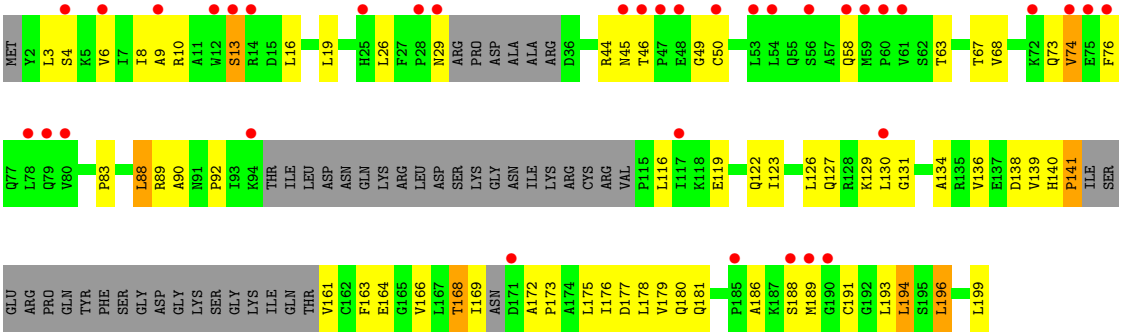
- Molecule 4: CRISPR system Cascade subunit CasD

Chain J:



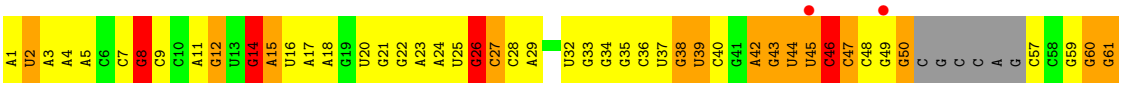
- Molecule 5: CRISPR system Cascade subunit CasE

Chain K:



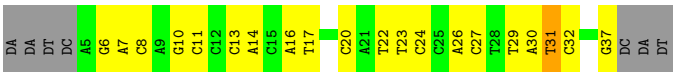
- Molecule 6: RNA (55-MER)

Chain L:



- Molecule 7: DNA (33-MER)

Chain M:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	223.87Å 223.87Å 290.65Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.47 – 3.03 39.47 – 3.03	Depositor EDS
% Data completeness (in resolution range)	96.2 (39.47-3.03) 96.3 (39.47-3.03)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.224 , 0.267 0.228 , 0.269	Depositor DCC
R_{free} test set	7932 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	86.4	Xtriage
Anisotropy	0.268	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 66.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.016 for -h,-k,l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	26543	wwPDB-VP
Average B, all atoms (Å^2)	106.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.29% 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 (i)

5.1 Standard geometry (i)

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	0.54	2/3789 (0.1%)	0.76	2/5138 (0.0%)
2	B	0.52	0/1262	0.74	1/1704 (0.1%)
2	C	0.56	1/1262 (0.1%)	0.76	1/1704 (0.1%)
3	D	0.43	0/2669	0.65	2/3595 (0.1%)
3	E	0.54	0/2839	0.76	3/3832 (0.1%)
3	F	0.55	0/2848	0.76	1/3844 (0.0%)
3	G	0.58	0/2832	0.76	1/3820 (0.0%)
3	H	0.53	0/2824	0.75	1/3808 (0.0%)
3	I	0.58	0/2013	0.81	3/2716 (0.1%)
4	J	0.55	1/1578 (0.1%)	0.81	0/2136
5	K	0.37	0/1222	0.56	0/1654
6	L	2.04	20/1317 (1.5%)	1.60	25/2051 (1.2%)
7	M	0.97	2/750 (0.3%)	1.05	1/1153 (0.1%)
All	All	0.70	26/27205 (0.1%)	0.83	41/37155 (0.1%)

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
1	A	0	2
3	D	0	1
3	E	0	1
All	All	0	4

The worst 5 of 26 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	L	57	C	N1-C2	20.95	1.61	1.40
6	L	57	C	N1-C6	19.63	1.49	1.37

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	L	50	G	C8-N7	18.89	1.42	1.30
6	L	50	G	N9-C4	17.88	1.52	1.38
6	L	50	G	C2-N3	17.86	1.47	1.32

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	L	50	G	C5-C6-O6	-13.47	120.52	128.60
6	L	50	G	N1-C6-O6	12.30	127.28	119.90
6	L	57	C	C2-N1-C1'	12.27	132.30	118.80
6	L	57	C	N3-C2-O2	-10.69	114.42	121.90
6	L	50	G	N3-C4-C5	-10.02	123.59	128.60

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	157	GLY	Peptide
1	A	442	GLN	Peptide
3	D	217	GLN	Peptide
3	E	31	PHE	Peptide

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	3702	0	3706	137	0
2	B	1238	0	1242	35	0
2	C	1238	0	1242	48	0
3	D	2630	0	2626	110	1
3	E	2794	0	2780	108	0
3	F	2803	0	2786	108	0
3	G	2788	0	2767	114	1
3	H	2781	0	2765	92	0
3	I	1978	0	1918	81	0
4	J	1545	0	1562	89	0
5	K	1197	0	1232	44	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	L	1178	0	597	68	0
7	M	670	0	371	18	0
8	A	1	0	0	0	0
All	All	26543	0	25594	925	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 18.

The worst 5 of 925 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)
3:E:337:VAL:N	3:E:338:ASP:HB2	1.40	1.35
3:E:337:VAL:H	3:E:338:ASP:CB	1.61	1.13
1:A:4:LEU:HB3	1:A:84:ILE:HD11	1.37	1.06
3:E:337:VAL:CA	3:E:338:ASP:HB2	1.85	1.06
3:H:185:ALA:HB2	3:I:271:ALA:HB3	1.41	1.02

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)
3:D:318:ARG:NH1	3:G:80:GLY:O[2_655]	2.10	0.10

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
1	A	460/502 (92%)	414 (90%)	42 (9%)	4 (1%)	17 52
2	B	146/160 (91%)	139 (95%)	7 (5%)	0	100 100
2	C	146/160 (91%)	136 (93%)	8 (6%)	2 (1%)	11 40

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	D	334/363 (92%)	303 (91%)	29 (9%)	2 (1%)	25	60
3	E	359/363 (99%)	346 (96%)	12 (3%)	1 (0%)	41	74
3	F	360/363 (99%)	349 (97%)	11 (3%)	0	100	100
3	G	356/363 (98%)	349 (98%)	7 (2%)	0	100	100
3	H	355/363 (98%)	342 (96%)	13 (4%)	0	100	100
3	I	250/363 (69%)	245 (98%)	5 (2%)	0	100	100
4	J	183/224 (82%)	174 (95%)	8 (4%)	1 (0%)	29	65
5	K	142/199 (71%)	136 (96%)	6 (4%)	0	100	100
All	All	3091/3423 (90%)	2933 (95%)	148 (5%)	10 (0%)	41	74

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	52	PRO
3	E	338	ASP
1	A	369	GLN
2	C	53	ARG
1	A	253	CYS

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	
1	A	401/426 (94%)	335 (84%)	66 (16%)	2	10
2	B	130/138 (94%)	114 (88%)	16 (12%)	4	19
2	C	130/138 (94%)	115 (88%)	15 (12%)	5	22
3	D	280/298 (94%)	231 (82%)	49 (18%)	2	9
3	E	297/298 (100%)	266 (90%)	31 (10%)	7	25
3	F	298/298 (100%)	274 (92%)	24 (8%)	11	37
3	G	296/298 (99%)	265 (90%)	31 (10%)	7	25
3	H	295/298 (99%)	265 (90%)	30 (10%)	7	26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	I	208/298 (70%)	185 (89%)	23 (11%)	6 23
4	J	168/192 (88%)	128 (76%)	40 (24%)	0 3
5	K	129/170 (76%)	116 (90%)	13 (10%)	7 27
All	All	2632/2852 (92%)	2294 (87%)	338 (13%)	4 18

5 of 338 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	359	ASN
4	J	29	ARG
3	H	95	LEU
3	H	317	ASP
4	J	115	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
3	H	306	GLN
4	J	207	GLN
5	K	127	GLN
4	J	62	GLN
3	F	42	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	L	53/61 (86%)	23 (43%)	6 (11%)

5 of 23 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	L	2	U
6	L	8	G
6	L	9	C
6	L	14	G
6	L	15	A

5 of 6 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
6	L	42	A
6	L	44	U
6	L	45	U
6	L	21	G
6	L	9	C

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 1 ligands modelled in this entry, 1 is 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	468/502 (93%)	-0.36	6 (1%) 77 51	63, 99, 129, 161	0
2	B	150/160 (93%)	-0.08	0 100 100	90, 117, 141, 158	0
2	C	150/160 (93%)	-0.43	0 100 100	78, 110, 139, 166	0
3	D	340/363 (93%)	-0.16	7 (2%) 63 34	82, 124, 176, 195	0
3	E	361/363 (99%)	-0.24	3 (0%) 86 65	68, 101, 155, 175	0
3	F	362/363 (99%)	-0.41	2 (0%) 89 72	60, 90, 125, 146	0
3	G	360/363 (99%)	-0.39	0 100 100	56, 86, 121, 140	0
3	H	359/363 (98%)	-0.46	0 100 100	56, 91, 131, 147	0
3	I	256/363 (70%)	-0.22	3 (1%) 79 53	59, 92, 129, 149	0
4	J	195/224 (87%)	-0.12	3 (1%) 73 46	72, 102, 137, 158	0
5	K	152/199 (76%)	1.22	36 (23%) 0 0	115, 174, 210, 234	0
6	L	55/61 (90%)	0.16	2 (3%) 42 18	72, 96, 195, 199	0
7	M	33/40 (82%)	-0.72	0 100 100	79, 102, 148, 168	0
All	All	3241/3524 (91%)	-0.23	62 (1%) 66 38	56, 100, 162, 234	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	K	60	PRO	4.7
3	D	101	ASP	4.4
1	A	292	GLU	3.8
5	K	12	TRP	3.6
5	K	74	VAL	3.6

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
8	ZN	A	601	1/1	0.99	0.07	170,170,170,170	0

6.5 Other polymers (i)

There are no such residues in this entry.