



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 19, 2023 – 05:46 AM EDT

PDB ID : 5CD4  
Title : The Type IE CRISPR Cascade complex from E. coli, with two assemblies in the asymmetric unit arranged back-to-back  
Authors : Jackson, R.N.; Golden, S.M.; Carter, J.; Wiedenheft, B.  
Deposited on : 2015-07-03  
Resolution : 3.20 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35.1  
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) : 2.35.1

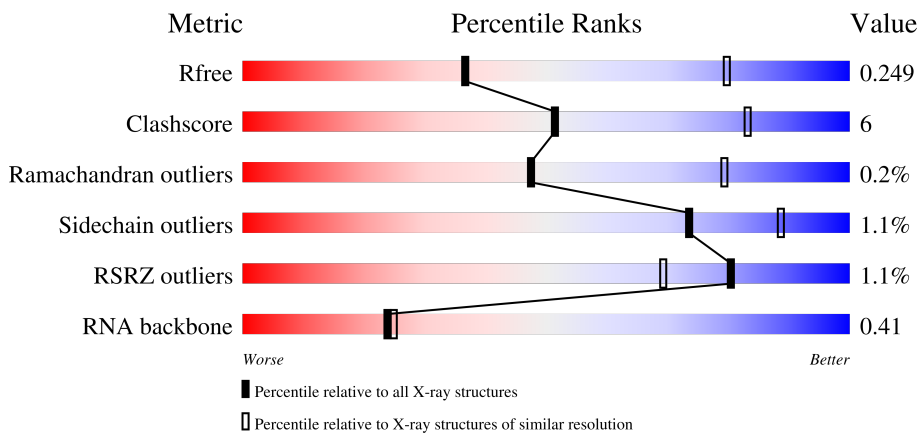
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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.



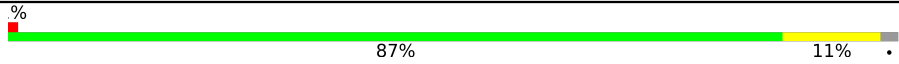

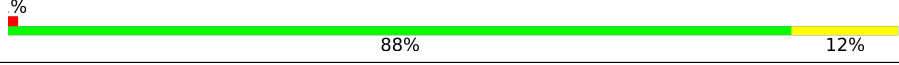



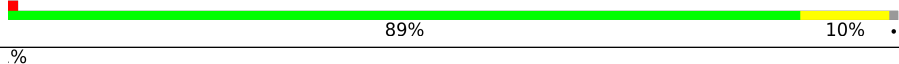

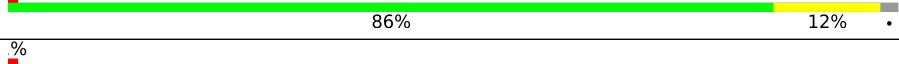

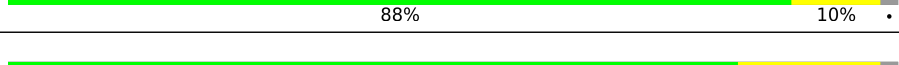
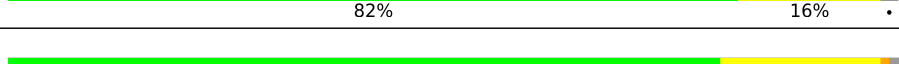

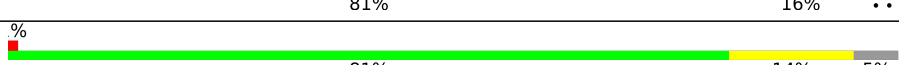

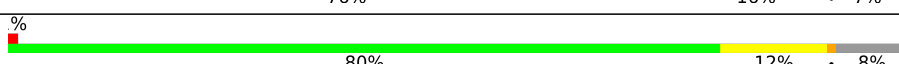
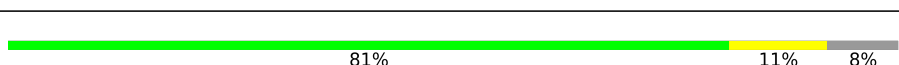
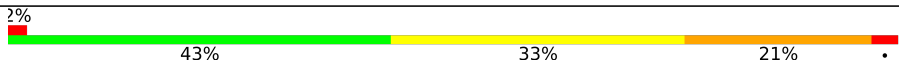


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)
RNA backbone	3102	1010 (3.50-2.90)

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	199	
1	M	199	
2	B	363	
2	C	363	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	363	 % 87% 11%
2	E	363	 % 87% 13%
2	F	363	 % 88% 12%
2	G	363	 % 83% 16%
2	N	363	 2% 83% 15%
2	O	363	 % 82% 14%
2	P	363	 % 89% 10%
2	Q	363	 % 85% 13%
2	R	363	 % 86% 12%
2	S	363	 % 82% 13%
3	H	224	 % 88% 10%
3	T	224	 % 82% 16%
4	I	502	 % 80% 18%
4	U	502	 % 81% 16%
5	J	165	 % 81% 14% 5%
5	K	165	 % 76% 16% 7%
5	V	165	 % 80% 12% 8%
5	W	165	 % 81% 11% 8%
6	L	61	 2% 43% 33% 21%
6	X	61	 % 49% 30% 20%

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 53308 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 CasE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	199	1523	975	273	269	6	0	0	0
1	M	192	1354	864	245	240	5	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	357	2673	1677	474	507	15	0	0	0
2	C	356	2670	1671	477	507	15	0	0	0
2	D	355	2674	1672	478	508	16	0	0	0
2	E	362	2748	1722	490	520	16	0	0	0
2	F	363	2746	1719	490	521	16	0	0	0
2	G	362	2749	1716	490	528	15	0	0	0
2	N	356	2544	1595	464	471	14	0	0	0
2	O	351	2546	1593	464	475	14	0	0	0
2	P	359	2645	1659	478	493	15	0	0	0
2	Q	359	2706	1695	485	510	16	0	0	0
2	R	357	2688	1685	483	504	16	0	0	0
2	S	353	2634	1649	473	497	15	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	H	219	Total	C	N	O	S	0	0	0
			1715	1087	302	317	9			
3	T	219	Total	C	N	O	S	0	0	0
			1716	1087	303	317	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	I	494	Total	C	N	O	S	0	0	0
			3792	2424	671	678	19			
4	U	492	Total	C	N	O	S	0	0	0
			3709	2371	655	665	18			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	J	156	Total	C	N	O	S	0	0	0
			1215	767	231	210	7			
5	K	154	Total	C	N	O	S	0	0	0
			1263	791	243	222	7			
5	V	152	Total	C	N	O	S	0	0	0
			1161	741	216	197	7			
5	W	151	Total	C	N	O	S	0	0	0
			1235	775	239	214	7			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	-4	GLY	-	expression tag	UNP P76632
J	-3	PRO	-	expression tag	UNP P76632
J	-2	GLY	-	expression tag	UNP P76632
J	-1	TYR	-	expression tag	UNP P76632
J	0	GLN	-	expression tag	UNP P76632
K	-4	GLY	-	expression tag	UNP P76632
K	-3	PRO	-	expression tag	UNP P76632
K	-2	GLY	-	expression tag	UNP P76632
K	-1	TYR	-	expression tag	UNP P76632
K	0	GLN	-	expression tag	UNP P76632
V	-4	GLY	-	expression tag	UNP P76632
V	-3	PRO	-	expression tag	UNP P76632
V	-2	GLY	-	expression tag	UNP P76632

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
V	-1	TYR	-	expression tag	UNP P76632
V	0	GLN	-	expression tag	UNP P76632
W	-4	GLY	-	expression tag	UNP P76632
W	-3	PRO	-	expression tag	UNP P76632
W	-2	GLY	-	expression tag	UNP P76632
W	-1	TYR	-	expression tag	UNP P76632
W	0	GLN	-	expression tag	UNP P76632

- Molecule 6 is a RNA chain called crRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	L	61	1300	580	233	426	61	0	0	0
6	X	61	1300	580	233	426	61	0	0	0

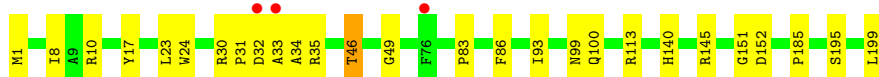
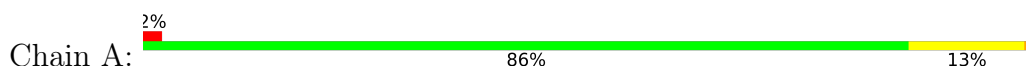
- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	I	1	Total	Zn	0	0
			1	1		
7	U	1	Total	Zn	0	0
			1	1		

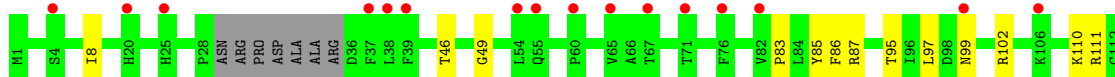
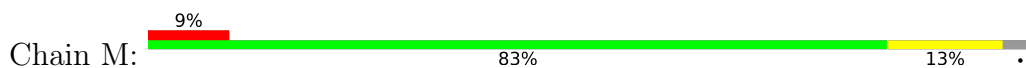
### 3 Residue-property plots [i](#)

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.

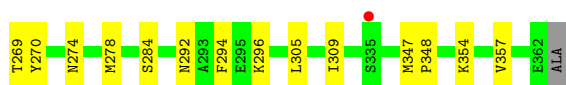
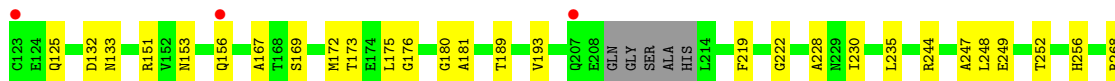
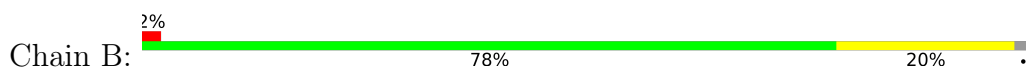
- Molecule 1: CRISPR system Cascade subunit CasE



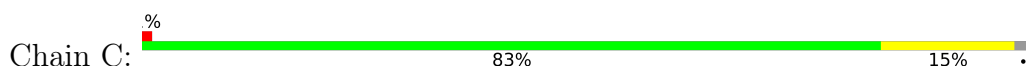
- Molecule 1: CRISPR system Cascade subunit CasE

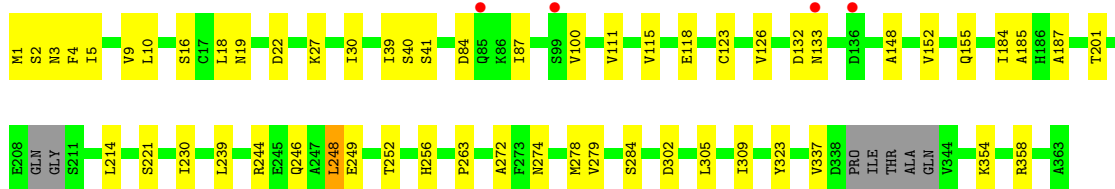


- Molecule 2: CRISPR system Cascade subunit CasC

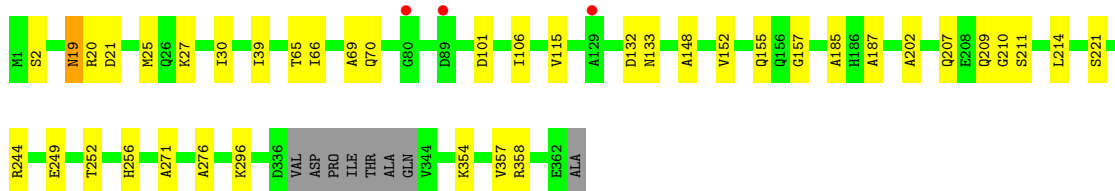
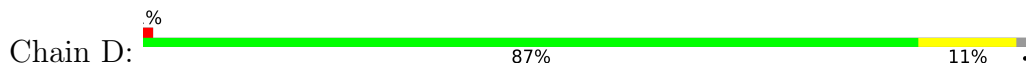


- Molecule 2: CRISPR system Cascade subunit CasC

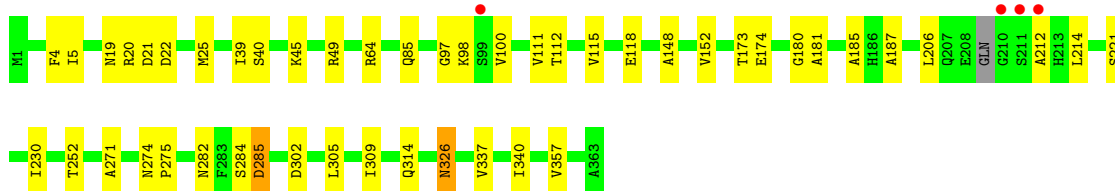
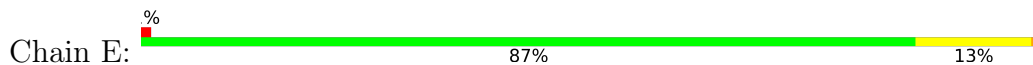




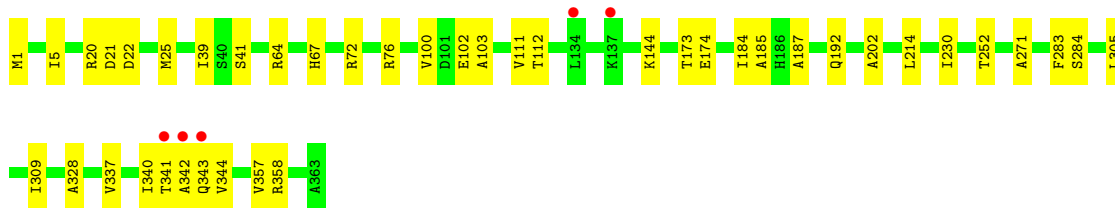
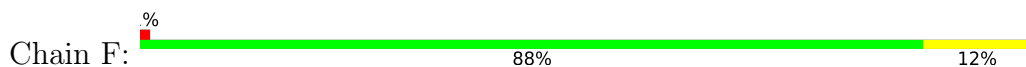
- Molecule 2: CRISPR system Cascade subunit CasC



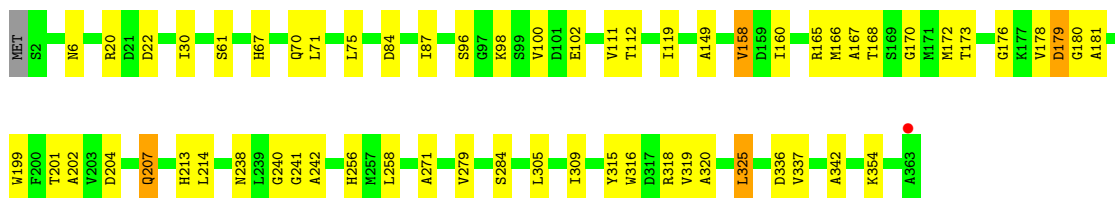
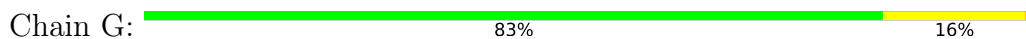
- Molecule 2: CRISPR system Cascade subunit CasC



- Molecule 2: CRISPR system Cascade subunit CasC

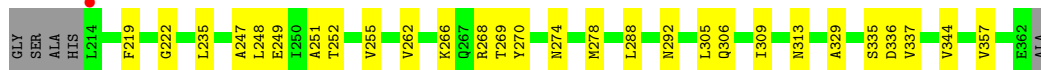
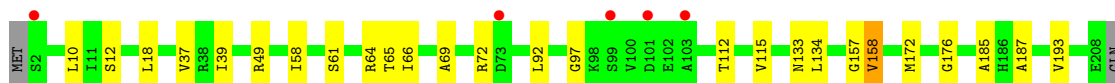
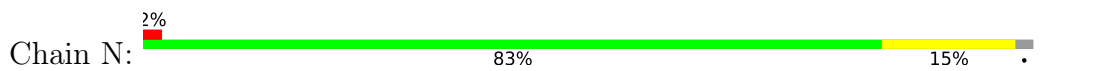


- Molecule 2: CRISPR system Cascade subunit CasC

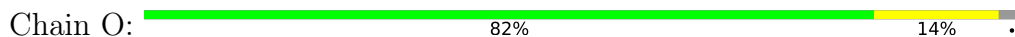


- Molecule 2: CRISPR system Cascade subunit CasC

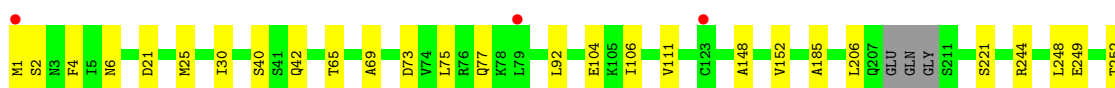
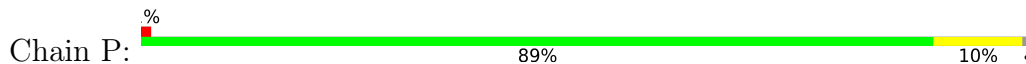




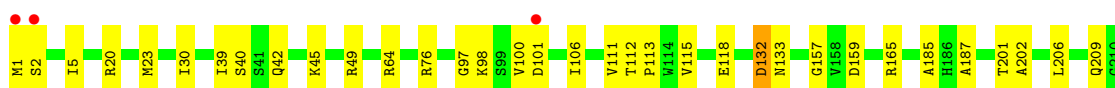
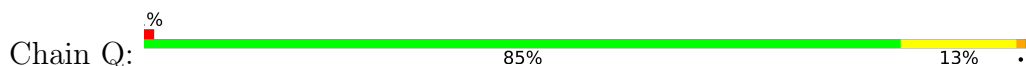
- Molecule 2: CRISPR system Cascade subunit CasC



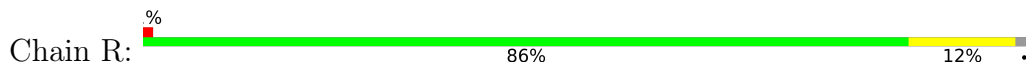
- Molecule 2: CRISPR system Cascade subunit CasC



- Molecule 2: CRISPR system Cascade subunit CasC

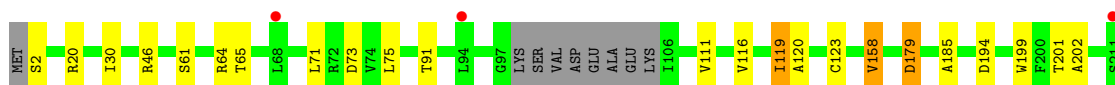
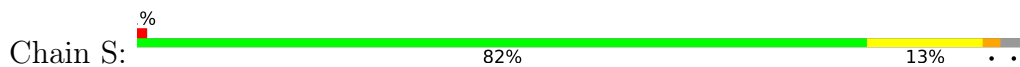


- Molecule 2: CRISPR system Cascade subunit CasC

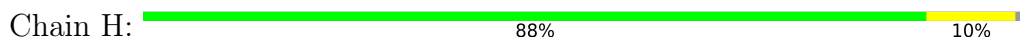




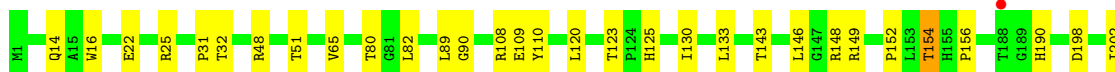
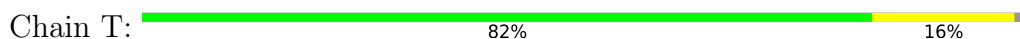
- Molecule 2: CRISPR system Cascade subunit CasC



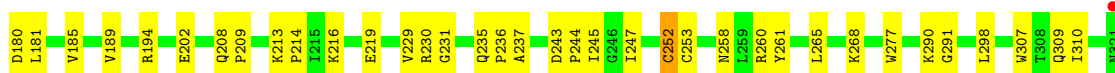
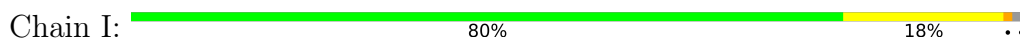
- Molecule 3: CRISPR system Cascade subunit CasD



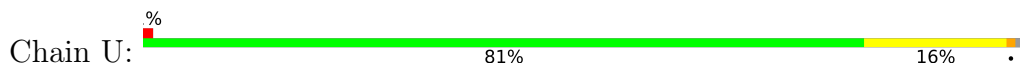
- Molecule 3: CRISPR system Cascade subunit CasD

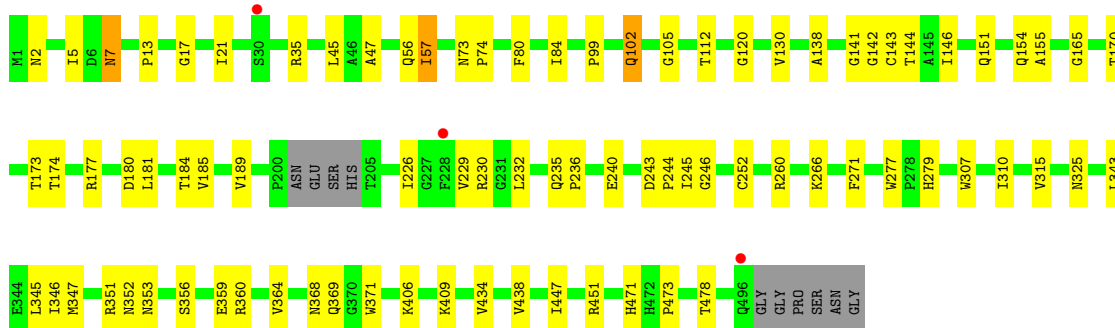


- Molecule 4: CRISPR system Cascade subunit CasA

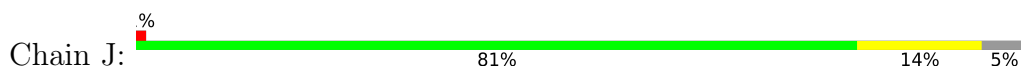


- Molecule 4: CRISPR system Cascade subunit CasA

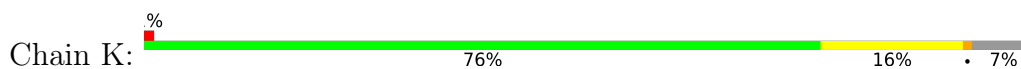




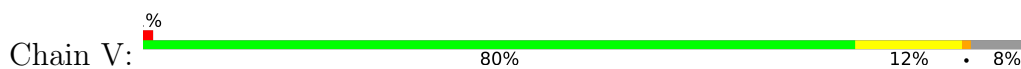
• Molecule 5: CRISPR system Cascade subunit CasB



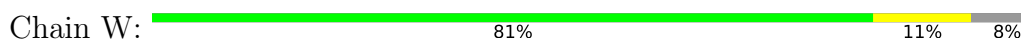
• Molecule 5: CRISPR system Cascade subunit CasB



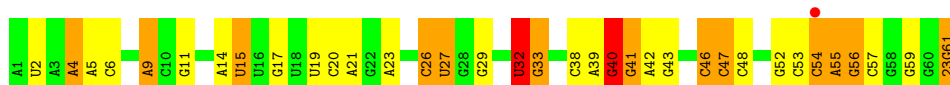
• Molecule 5: CRISPR system Cascade subunit CasB



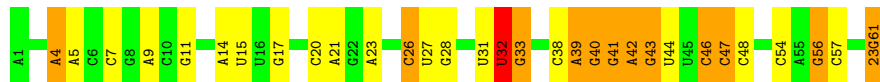
• Molecule 5: CRISPR system Cascade subunit CasB



• Molecule 6: crRNA



## ● Molecule 6: crRNA

Chain X:  49% 30% 20%

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.99Å 244.80Å 426.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.73 – 3.20 49.74 – 3.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.73-3.20) 97.1 (49.74-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.97 (at 3.19Å)	Xtrriage
Refinement program	PHENIX (1.10pre_2089: ???)	Depositor
R, $R_{free}$	0.212 , 0.250 0.212 , 0.249	Depositor DCC
$R_{free}$ test set	2000 reflections (1.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.6	Xtrriage
Anisotropy	0.741	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 47.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	53308	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.58% of the height of the origin peak. No significant pseudotranslation is detected.*

<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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, 23G

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	0.23	0/1554	0.47	0/2108
1	M	0.21	0/1379	0.43	0/1878
2	B	0.23	0/2716	0.43	0/3681
2	C	0.21	0/2712	0.37	0/3674
2	D	0.21	0/2717	0.36	0/3679
2	E	0.21	0/2791	0.35	0/3773
2	F	0.21	0/2790	0.38	0/3775
2	G	0.22	0/2794	0.39	0/3780
2	N	0.22	0/2586	0.43	1/3519 (0.0%)
2	O	0.21	0/2587	0.37	0/3512
2	P	0.21	0/2689	0.35	0/3648
2	Q	0.21	0/2748	0.36	0/3716
2	R	0.21	0/2730	0.35	0/3690
2	S	0.21	0/2678	0.37	0/3631
3	H	0.21	0/1756	0.43	0/2387
3	T	0.21	0/1758	0.43	0/2391
4	I	0.23	0/3883	0.44	0/5286
4	U	0.23	0/3799	0.42	0/5181
5	J	0.23	0/1240	0.42	0/1685
5	K	0.23	0/1287	0.41	0/1739
5	V	0.22	0/1185	0.42	1/1612 (0.1%)
5	W	0.22	0/1259	0.43	0/1702
6	L	0.24	0/1423	0.97	6/2216 (0.3%)
6	X	0.27	0/1423	0.98	7/2216 (0.3%)
All	All	0.22	0/54484	0.45	15/74479 (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
2	B	0	1

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	L	40	G	O4'-C1'-N9	9.32	115.66	108.20
6	X	48	C	C2-N1-C1'	7.72	127.29	118.80
2	N	97	GLY	N-CA-C	-7.52	94.30	113.10
6	L	26	C	C2-N1-C1'	6.30	125.73	118.80
6	X	26	C	C2-N1-C1'	6.16	125.57	118.80

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	110	ALA	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	1523	0	1551	17	0
1	M	1354	0	1291	15	0
2	B	2673	0	2583	48	0
2	C	2670	0	2561	34	0
2	D	2674	0	2583	25	0
2	E	2748	0	2701	30	0
2	F	2746	0	2693	28	0
2	G	2749	0	2670	43	0
2	N	2544	0	2348	33	0
2	O	2546	0	2371	31	0
2	P	2645	0	2510	25	0
2	Q	2706	0	2636	34	0
2	R	2688	0	2624	29	0
2	S	2634	0	2532	38	0
3	H	1715	0	1694	14	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	T	1716	0	1688	25	0
4	I	3792	0	3716	57	0
4	U	3709	0	3571	48	0
5	J	1215	0	1168	18	0
5	K	1263	0	1260	18	0
5	V	1161	0	1116	9	0
5	W	1235	0	1232	9	0
6	L	1300	0	659	20	0
6	X	1300	0	659	22	0
7	I	1	0	0	0	0
7	U	1	0	0	0	0
All	All	53308	0	50417	581	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 581 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)
6:X:61:23G:O4'	6:X:61:23G:C4'	1.67	1.24
6:L:61:23G:O4'	6:L:61:23G:C4'	1.66	1.21
1:A:8:ILE:O	1:A:49:GLY:HA3	1.72	0.89
1:M:151:GLY:HA3	1:M:154:LYS:H	1.41	0.85
4:I:290:LYS:H	4:I:291:GLY:HA3	1.46	0.80

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
1	A	197/199 (99%)	190 (96%)	7 (4%)	0	<b>100</b> <b>100</b>

Continued on next page...



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	M	188/199 (94%)	181 (96%)	7 (4%)	0	100	100
2	B	353/363 (97%)	327 (93%)	25 (7%)	1 (0%)	41	74
2	C	350/363 (96%)	332 (95%)	17 (5%)	1 (0%)	41	74
2	D	351/363 (97%)	339 (97%)	12 (3%)	0	100	100
2	E	358/363 (99%)	344 (96%)	13 (4%)	1 (0%)	41	74
2	F	361/363 (99%)	341 (94%)	19 (5%)	1 (0%)	41	74
2	G	360/363 (99%)	335 (93%)	24 (7%)	1 (0%)	41	74
2	N	352/363 (97%)	322 (92%)	30 (8%)	0	100	100
2	O	345/363 (95%)	328 (95%)	16 (5%)	1 (0%)	41	74
2	P	355/363 (98%)	330 (93%)	24 (7%)	1 (0%)	41	74
2	Q	353/363 (97%)	343 (97%)	10 (3%)	0	100	100
2	R	351/363 (97%)	336 (96%)	14 (4%)	1 (0%)	41	74
2	S	349/363 (96%)	328 (94%)	19 (5%)	2 (1%)	25	64
3	H	217/224 (97%)	210 (97%)	7 (3%)	0	100	100
3	T	217/224 (97%)	210 (97%)	7 (3%)	0	100	100
4	I	490/502 (98%)	468 (96%)	21 (4%)	1 (0%)	47	79
4	U	488/502 (97%)	467 (96%)	19 (4%)	2 (0%)	34	69
5	J	154/165 (93%)	151 (98%)	3 (2%)	0	100	100
5	K	150/165 (91%)	148 (99%)	2 (1%)	0	100	100
5	V	148/165 (90%)	147 (99%)	0	1 (1%)	22	61
5	W	147/165 (89%)	144 (98%)	2 (1%)	1 (1%)	22	61
All	All	6634/6866 (97%)	6321 (95%)	298 (4%)	15 (0%)	47	79

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	I	202	GLU
4	U	165	GLY
4	U	369	GLN
5	W	50	GLU
2	B	110	ALA

### 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	160/170 (94%)	158 (99%)	2 (1%)	69	87
1	M	125/170 (74%)	124 (99%)	1 (1%)	81	93
2	B	269/298 (90%)	269 (100%)	0	100	100
2	C	267/298 (90%)	265 (99%)	2 (1%)	84	94
2	D	271/298 (91%)	268 (99%)	3 (1%)	73	88
2	E	282/298 (95%)	278 (99%)	4 (1%)	67	86
2	F	282/298 (95%)	281 (100%)	1 (0%)	91	95
2	G	282/298 (95%)	275 (98%)	7 (2%)	47	77
2	N	231/298 (78%)	229 (99%)	2 (1%)	78	91
2	O	238/298 (80%)	235 (99%)	3 (1%)	69	87
2	P	254/298 (85%)	253 (100%)	1 (0%)	91	95
2	Q	274/298 (92%)	271 (99%)	3 (1%)	73	88
2	R	272/298 (91%)	272 (100%)	0	100	100
2	S	264/298 (89%)	257 (97%)	7 (3%)	44	75
3	H	183/192 (95%)	180 (98%)	3 (2%)	62	84
3	T	183/192 (95%)	181 (99%)	2 (1%)	73	88
4	I	396/426 (93%)	391 (99%)	5 (1%)	69	87
4	U	377/426 (88%)	373 (99%)	4 (1%)	73	88
5	J	118/141 (84%)	118 (100%)	0	100	100
5	K	132/141 (94%)	130 (98%)	2 (2%)	65	85
5	V	110/141 (78%)	108 (98%)	2 (2%)	59	82
5	W	128/141 (91%)	127 (99%)	1 (1%)	81	93
All	All	5098/5716 (89%)	5043 (99%)	55 (1%)	73	88

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	K	156	THR
2	P	282	ASN
5	W	43	LEU
4	U	57	ILE
1	M	46	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
2	N	331	GLN
2	Q	19	ASN
2	O	19	ASN
2	P	234	GLN
2	R	234	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	L	59/61 (96%)	22 (37%)	1 (1%)
6	X	59/61 (96%)	19 (32%)	1 (1%)
All	All	118/122 (96%)	41 (34%)	2 (1%)

5 of 41 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	L	4	A
6	L	9	A
6	L	14	A
6	L	15	U
6	L	21	A

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
6	L	54	C
6	X	42	A

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	23G	L	61	6	19,29,30	4.76	15 (78%)	20,45,48	1.82	7 (35%)
6	23G	X	61	6	19,29,30	4.96	15 (78%)	20,45,48	1.88	6 (30%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	23G	L	61	6	-	2/3/35/36	0/4/4/4
6	23G	X	61	6	-	2/3/35/36	0/4/4/4

The worst 5 of 30 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	X	61	23G	O4'-C4'	10.13	1.67	1.45
6	L	61	23G	O4'-C4'	9.72	1.66	1.45
6	X	61	23G	C3'-C4'	-8.55	1.29	1.52
6	L	61	23G	C3'-C4'	-8.51	1.30	1.52
6	X	61	23G	O2'-C2'	-8.33	1.27	1.45

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	X	61	23G	C5-C6-N1	3.86	120.77	113.95
6	L	61	23G	C5-C6-N1	3.69	120.46	113.95
6	X	61	23G	C2-N1-C6	-3.56	118.55	125.10
6	L	61	23G	C2-N1-C6	-3.23	119.15	125.10
6	X	61	23G	C8-N7-C5	3.12	108.94	102.99

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	X	61	23G	C3'-C4'-C5'-O5'
6	X	61	23G	O4'-C4'-C5'-O5'
6	L	61	23G	O4'-C4'-C5'-O5'
6	L	61	23G	C3'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	L	61	23G	1	0
6	X	61	23G	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 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, 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	199/199 (100%)	0.10	3 (1%) 73 61	40, 79, 126, 149	0
1	M	192/199 (96%)	0.61	17 (8%) 9 5	73, 117, 146, 164	0
2	B	357/363 (98%)	-0.00	8 (2%) 62 48	34, 71, 131, 174	0
2	C	356/363 (98%)	-0.17	4 (1%) 80 69	29, 59, 113, 158	0
2	D	355/363 (97%)	-0.26	3 (0%) 86 78	21, 54, 120, 151	0
2	E	362/363 (99%)	-0.32	4 (1%) 80 69	19, 42, 87, 122	0
2	F	363/363 (100%)	-0.22	5 (1%) 75 63	16, 39, 100, 135	0
2	G	362/363 (99%)	-0.33	1 (0%) 94 92	15, 42, 103, 132	0
2	N	356/363 (98%)	0.16	6 (1%) 70 57	63, 100, 148, 183	0
2	O	351/363 (96%)	-0.12	1 (0%) 94 92	52, 82, 122, 166	0
2	P	359/363 (98%)	-0.08	4 (1%) 80 69	47, 78, 138, 164	0
2	Q	359/363 (98%)	-0.25	3 (0%) 86 78	36, 60, 107, 134	0
2	R	357/363 (98%)	-0.30	2 (0%) 89 83	33, 61, 102, 143	0
2	S	353/363 (97%)	-0.13	3 (0%) 86 78	34, 63, 116, 135	0
3	H	219/224 (97%)	-0.34	1 (0%) 91 86	22, 46, 82, 107	0
3	T	219/224 (97%)	-0.23	1 (0%) 91 86	37, 68, 109, 136	0
4	I	494/502 (98%)	-0.30	1 (0%) 95 94	24, 54, 96, 131	0
4	U	492/502 (98%)	-0.14	3 (0%) 89 83	33, 82, 123, 154	0
5	J	156/165 (94%)	-0.05	2 (1%) 77 65	39, 74, 122, 152	0
5	K	154/165 (93%)	-0.20	1 (0%) 89 83	23, 51, 94, 148	0
5	V	152/165 (92%)	0.15	2 (1%) 77 65	61, 94, 129, 141	0
5	W	151/165 (91%)	-0.13	0 100 100	34, 66, 102, 119	0
6	L	60/61 (98%)	-0.14	1 (1%) 70 57	29, 52, 95, 138	0
6	X	60/61 (98%)	-0.07	0 100 100	41, 71, 137, 157	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
All	All	6838/6988 (97%)	-0.14	76 (1%) 80 69	15, 67, 125, 183	0

The worst 5 of 76 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	K	3	ASP	4.8
2	N	73	ASP	4.2
2	R	363	ALA	4.2
2	C	133	ASN	3.7
2	G	363	ALA	3.6

## 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
6	23G	X	61	26/27	0.89	0.26	59,94,132,137	0
6	23G	L	61	26/27	0.90	0.18	42,77,110,118	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
7	ZN	I	601	1/1	0.93	0.15	113,113,113,113	0
7	ZN	U	601	1/1	0.97	0.15	134,134,134,134	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.