



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Sep 7, 2019 – 02:10 PM EDT

PDB ID : 6O7I
EMDB ID: : EMD-0642
Title : Cryo-EM structure of Csm-crRNA-target RNA ternary bigger complex in complex with cA4 in type III-A CRISPR-Cas system
Authors : Jia, N.; Patel, D.J.
Deposited on : 2019-03-07
Resolution : 3.20 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : 2.4

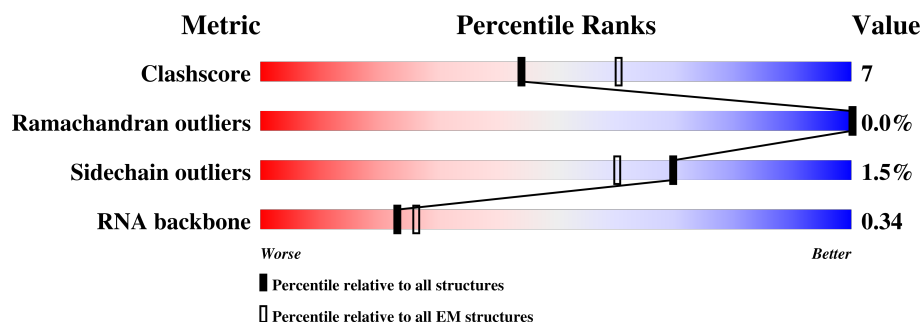
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531
RNA backbone	3747	458

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	791	79% 15% • 6%
2	B	187	66% 12% 22%
2	J	187	59% 6% 35%
3	C	291	76% 18% • 5%
3	D	291	76% 18% 7%
3	K	291	76% 17% • 5%
4	E	289	84% 12% •
5	G	38	13% 26% 29% 32%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	H	40	<div><div></div><div>15%</div><div>23%</div><div>10%</div><div>53%</div></div>
7	I	4	<div><div></div><div>50%</div><div>50%</div></div>
8	F	378	<div><div></div><div>82%</div><div>13%</div><div>..</div></div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 20276 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 single-strand-specific deoxyribonuclease Cas10/Csm1 (subtype III-A).

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	741	Total	C	N	O	S	0	0
			5935	3808	1031	1077	19		

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP B6YWB8
A	-12	GLY	-	expression tag	UNP B6YWB8
A	-11	SER	-	expression tag	UNP B6YWB8
A	-10	SER	-	expression tag	UNP B6YWB8
A	-9	HIS	-	expression tag	UNP B6YWB8
A	-8	HIS	-	expression tag	UNP B6YWB8
A	-7	HIS	-	expression tag	UNP B6YWB8
A	-6	HIS	-	expression tag	UNP B6YWB8
A	-5	HIS	-	expression tag	UNP B6YWB8
A	-4	HIS	-	expression tag	UNP B6YWB8
A	-3	SER	-	expression tag	UNP B6YWB8
A	-2	GLN	-	expression tag	UNP B6YWB8
A	-1	ASP	-	expression tag	UNP B6YWB8
A	0	PRO	-	expression tag	UNP B6YWB8
A	589	ALA	ASP	engineered mutation	UNP B6YWB8

- Molecule 2 is a protein called Csm2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	146	Total	C	N	O	S	0	0
			1065	677	189	195	4		
2	J	121	Total	C	N	O	S	0	0
			907	582	159	162	4		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	SER	-	expression tag	UNP B6YWB9
J	0	SER	-	expression tag	UNP B6YWB9

- Molecule 3 is a protein called Csm3.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	275	Total	C	N	O	S	0	0
			2191	1392	383	410	6		
3	D	272	Total	C	N	O	S	0	0
			2178	1384	381	406	7		
3	K	275	Total	C	N	O	S	0	0
			2198	1396	385	410	7		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	SER	-	expression tag	UNP B6YWC0
D	0	SER	-	expression tag	UNP B6YWC0
K	0	SER	-	expression tag	UNP B6YWC0

- Molecule 4 is a protein called Csm4.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	279	Total	C	N	O	S	0	0
			2207	1436	373	394	4		

- Molecule 5 is a RNA chain called RNA (38-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	G	26	Total	C	N	O	P	0	0
			569	253	112	179	25		

- Molecule 6 is a RNA chain called RNA (40-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	H	19	Total	C	N	O	P	0	0
			398	178	69	132	19		

- Molecule 7 is a RNA chain called RNA (5'-R(P*AP*AP*AP*A)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
7	I	4	Total	C	N	O	P	0	0
			88	40	20	24	4		

- Molecule 8 is a protein called Csm5.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	F	361	Total	C	N	O	S	0	0
			2536	1603	453	471	9		

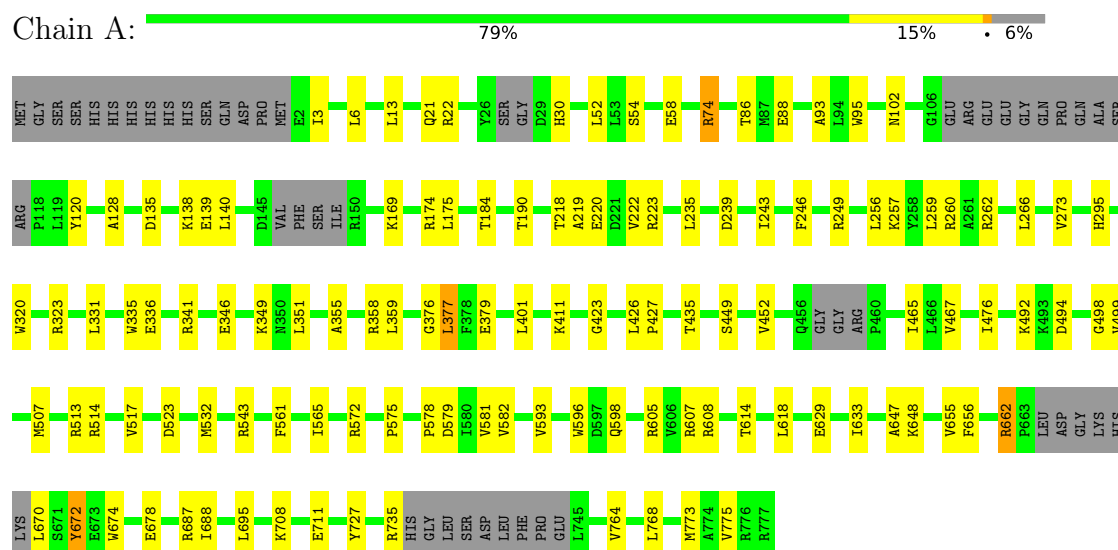
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
9	A	1	Total	Zn	0
			1	1	
9	D	1	Total	Zn	0
			1	1	
9	C	1	Total	Zn	0
			1	1	
9	K	1	Total	Zn	0
			1	1	

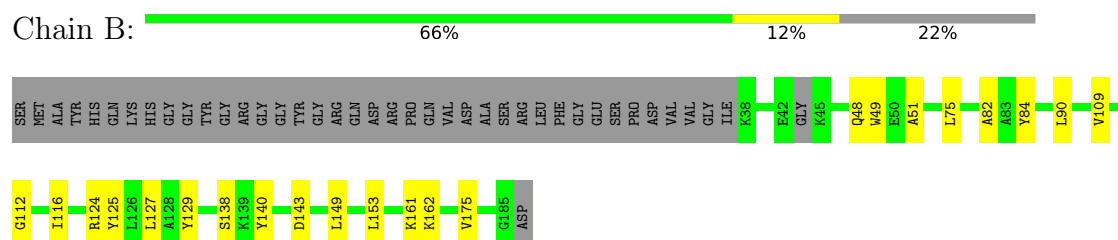
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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. 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. 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.

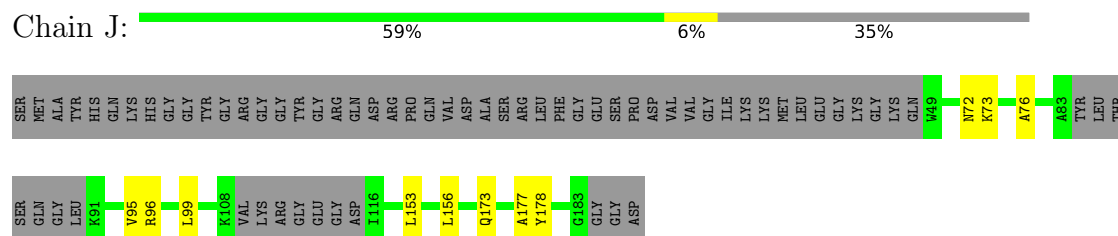
- Molecule 1: CRISPR system single-strand-specific deoxyribonuclease Cas10/Csm1 (subtype III-A)




- Molecule 2: Csm2

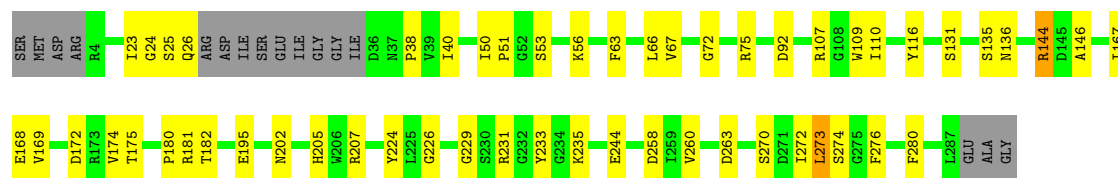


- Molecule 2: Csm2




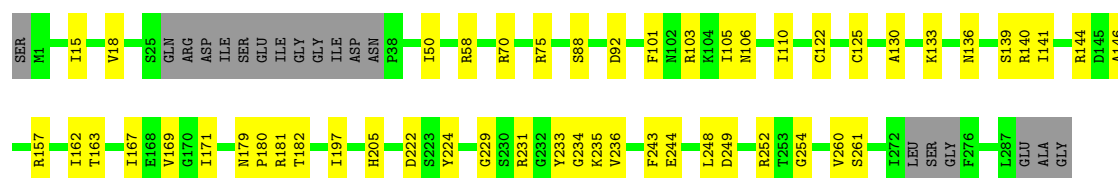
- Molecule 3: Csm3

Chain C:  76% 18% 5%




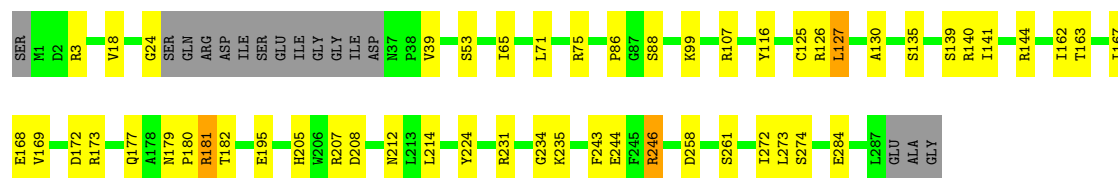
• Molecule 3: Csm3

Chain D:  76% 18% 7%




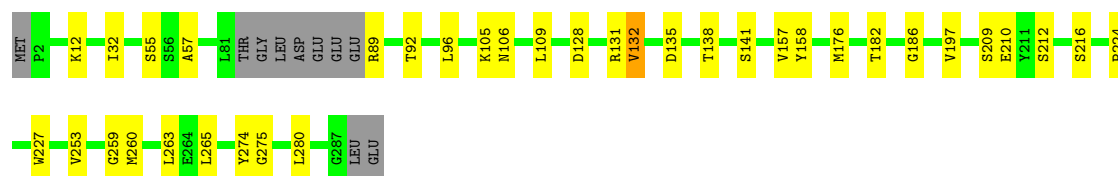
• Molecule 3: Csm3

Chain K:  76% 17% 5%



• Molecule 4: Csm4

Chain E:  84% 12% 4%




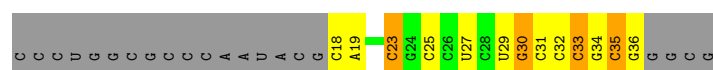
• Molecule 5: RNA (38-MER)

Chain G:  13% 26% 29% 32%



• Molecule 6: RNA (40-MER)

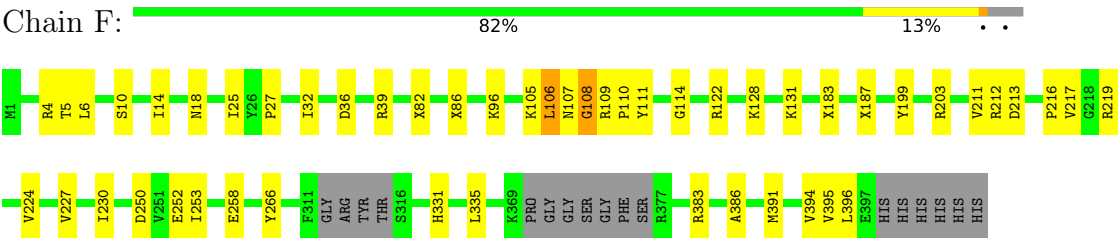
Chain H:  15% 23% 10% 53%



● Molecule 7: RNA (5'-R(P*AP*AP*AP*A)-3')



● Molecule 8: Csm5



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	41472	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	1.35	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

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 > 2$	RMSZ	# $ Z > 2$
1	A	0.54	0/6066	0.68	0/8179
2	B	0.30	0/1079	0.56	0/1453
2	J	0.26	0/919	0.50	0/1236
3	C	0.75	0/2240	0.73	0/3021
3	D	0.69	0/2226	0.72	0/2998
3	K	0.47	0/2247	0.61	0/3029
4	E	0.74	0/2267	0.70	0/3070
5	G	1.16	0/639	1.14	0/999
6	H	0.53	0/442	1.23	3/685 (0.4%)
7	I	1.01	0/99	1.29	0/152
8	F	0.30	0/2122	0.55	0/2875
All	All	0.60	0/20346	0.71	3/27697 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	H	33	C	C5-C6-N1	7.53	124.77	121.00
6	H	33	C	P-O3'-C3'	6.03	126.93	119.70
6	H	33	C	C6-N1-C2	-5.39	118.14	120.30

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	5935	0	5937	64	0
2	B	1065	0	1003	22	0
2	J	907	0	872	8	0
3	C	2191	0	2154	52	0
3	D	2178	0	2148	49	0
3	K	2198	0	2168	39	0
4	E	2207	0	2221	25	0
5	G	569	0	284	39	0
6	H	398	0	208	9	0
7	I	88	0	45	0	0
8	F	2536	0	2139	45	0
9	A	1	0	0	0	0
9	C	1	0	0	0	0
9	D	1	0	0	0	0
9	K	1	0	0	0	0
All	All	20276	0	19179	294	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 7.

The worst 5 of 294 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)
8:F:106:LEU:HD22	8:F:111:TYR:CD1	1.63	1.33
8:F:106:LEU:CD2	8:F:111:TYR:HD1	1.71	1.03
8:F:106:LEU:HD23	8:F:106:LEU:H	1.17	1.03
5:G:9:G:O2'	5:G:10:C:C5	2.13	1.00
3:K:273:LEU:HD23	3:K:274:SER:N	1.81	0.95

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 PDB entries followed by that with respect to all EM entries.

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	727/791 (92%)	670 (92%)	57 (8%)	0	100	100
2	B	142/187 (76%)	125 (88%)	17 (12%)	0	100	100
2	J	115/187 (62%)	106 (92%)	9 (8%)	0	100	100
3	C	271/291 (93%)	246 (91%)	25 (9%)	0	100	100
3	D	266/291 (91%)	241 (91%)	25 (9%)	0	100	100
3	K	271/291 (93%)	245 (90%)	26 (10%)	0	100	100
4	E	275/289 (95%)	261 (95%)	14 (5%)	0	100	100
8	F	259/378 (68%)	241 (93%)	17 (7%)	1 (0%)	36	74
All	All	2326/2705 (86%)	2135 (92%)	190 (8%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
8	F	108	GLY

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

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	622/664 (94%)	611 (98%)	11 (2%)	62	85
2	B	95/154 (62%)	95 (100%)	0	100	100
2	J	84/154 (54%)	84 (100%)	0	100	100
3	C	235/248 (95%)	230 (98%)	5 (2%)	56	83
3	D	234/248 (94%)	231 (99%)	3 (1%)	71	89
3	K	236/248 (95%)	231 (98%)	5 (2%)	56	83
4	E	231/240 (96%)	228 (99%)	3 (1%)	71	89
8	F	212/245 (86%)	210 (99%)	2 (1%)	81	93
All	All	1949/2201 (89%)	1920 (98%)	29 (2%)	70	88

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	181	ARG
3	D	75	ARG
3	K	246	ARG
3	C	207	ARG
3	D	179	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	179	ASN
3	C	205	HIS
3	K	22	HIS
3	C	136	ASN
4	E	139	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	G	25/38 (65%)	11 (44%)	8 (32%)
6	H	18/40 (45%)	9 (50%)	1 (5%)
7	I	3/4 (75%)	2 (66%)	0
All	All	46/82 (56%)	22 (47%)	9 (19%)

5 of 22 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
5	G	2	U
5	G	8	G
5	G	9	G
5	G	10	C
5	G	14	C

5 of 9 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
5	G	13	G
6	H	33	C
5	G	20	C
5	G	10	C

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	G	14	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
8	F	8

The worst 5 of 8 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	169:UNK	C	175:UNK	N	18.34
1	F	91:UNK	C	95:ARG	N	9.58

Continued on next page...

Continued from previous page...

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	191:UNK	C	197:ILE	N	8.20
1	F	143:UNK	C	150:UNK	N	7.29
1	F	156:UNK	C	159:UNK	N	7.06