



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Jul 24, 2017 – 04:56 AM EDT

PDB ID : 5MG3  
EMDB ID: : EMD-3506  
Title : EM fitted model of bacterial holo-translocon  
Authors : Schaffitzel, C.; Botte, M.  
Deposited on : unknown  
Resolution : 14.00 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029824

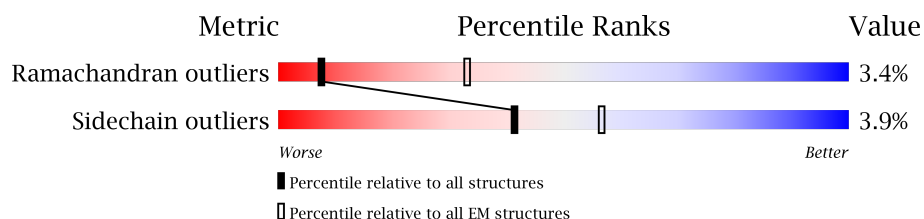
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 14.00 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)	EM structures (#Entries)
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

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  $\geq 3$ , 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	Y	458	
2	E	140	
3	G	136	
4	D	622	
5	F	323	
6	C	559	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 26650 atoms, of which 13553 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 Protein translocase subunit SecY.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
1	Y	443	7004	2259	3581	566	580	18	0	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Y	-14	VAL	-	expression tag	UNP P0AGA2
Y	-13	TRP	-	expression tag	UNP P0AGA2
Y	-12	ASN	-	expression tag	UNP P0AGA2
Y	-11	CYS	-	expression tag	UNP P0AGA2
Y	-10	GLU	-	expression tag	UNP P0AGA2
Y	-9	ARG	-	expression tag	UNP P0AGA2
Y	-8	ILE	-	expression tag	UNP P0AGA2
Y	-7	THR	-	expression tag	UNP P0AGA2
Y	-6	ILE	-	expression tag	UNP P0AGA2
Y	-5	SER	-	expression tag	UNP P0AGA2
Y	-4	HIS	-	expression tag	UNP P0AGA2
Y	-3	ARG	-	expression tag	UNP P0AGA2
Y	-2	LYS	-	expression tag	UNP P0AGA2
Y	-1	GLN	-	expression tag	UNP P0AGA2
Y	0	THR	-	expression tag	UNP P0AGA2

- Molecule 2 is a protein called Protein translocase subunit SecE.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
2	E	65	1062	332	552	91	86	1	0	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	369	MET	-	initiating methionine	UNP P0AG96
E	370	HIS	-	expression tag	UNP P0AG96

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	371	HIS	-	expression tag	UNP P0AG96
E	372	HIS	-	expression tag	UNP P0AG96
E	373	HIS	-	expression tag	UNP P0AG96
E	374	HIS	-	expression tag	UNP P0AG96
E	375	HIS	-	expression tag	UNP P0AG96
E	376	ASP	-	expression tag	UNP P0AG96
E	377	ASP	-	expression tag	UNP P0AG96
E	378	ASP	-	expression tag	UNP P0AG96
E	379	ASP	-	expression tag	UNP P0AG96
E	380	LYS	-	expression tag	UNP P0AG96
E	381	ALA	-	expression tag	UNP P0AG96
E	382	MET	-	expression tag	UNP P0AG96
E	383	GLY	-	expression tag	UNP P0AG96

- Molecule 3 is a protein called Protein-export membrane protein SecG.

Mol	Chain	Residues	Atoms						AltConf	Trace
3	G	32	Total	C	H	N	O	S	0	0
			480	151	244	39	44	2		

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	439	VAL	-	expression tag	UNP P0AG99
G	440	GLY	-	expression tag	UNP P0AG99
G	441	THR	-	expression tag	UNP P0AG99
G	442	GLY	-	expression tag	UNP P0AG99
G	443	TRP	-	expression tag	UNP P0AG99
G	444	TYR	-	expression tag	UNP P0AG99
G	445	SER	-	expression tag	UNP P0AG99
G	446	GLY	-	expression tag	UNP P0AG99
G	447	SER	-	expression tag	UNP P0AG99
G	448	PRO	-	expression tag	UNP P0AG99
G	449	GLY	-	expression tag	UNP P0AG99
G	450	ILE	-	expression tag	UNP P0AG99
G	451	LEU	-	expression tag	UNP P0AG99
G	452	TYR	-	expression tag	UNP P0AG99
G	453	HIS	-	expression tag	UNP P0AG99
G	454	TRP	-	expression tag	UNP P0AG99
G	455	PRO	-	expression tag	UNP P0AG99
G	456	GLU	-	expression tag	UNP P0AG99
G	457	VAL	-	expression tag	UNP P0AG99

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	458	LEU	-	expression tag	UNP P0AG99
G	459	ARG	-	expression tag	UNP P0AG99
G	460	ILE	-	expression tag	UNP P0AG99
G	461	GLN	-	expression tag	UNP P0AG99
G	462	GLU	-	expression tag	UNP P0AG99
G	463	LEU	-	expression tag	UNP P0AG99
G	464	ILE	-	expression tag	UNP P0AG99

- Molecule 4 is a protein called Protein translocase subunit SecD.

Mol	Chain	Residues	Atoms						AltConf	Trace
4	D	414	Total	C	H	N	O	S	0	0
			6419	1991	3291	545	582	10		

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-6	MET	-	initiating methionine	UNP P0AG90
D	-5	HIS	-	expression tag	UNP P0AG90
D	-4	HIS	-	expression tag	UNP P0AG90
D	-3	HIS	-	expression tag	UNP P0AG90
D	-2	HIS	-	expression tag	UNP P0AG90
D	-1	HIS	-	expression tag	UNP P0AG90
D	0	HIS	-	expression tag	UNP P0AG90
D	1	MET	-	expression tag	UNP P0AG90
D	142	VAL	ALA	conflict	UNP P0AG90

- Molecule 5 is a protein called Protein translocase subunit SecF.

Mol	Chain	Residues	Atoms						AltConf	Trace
5	F	289	Total	C	H	N	O	S	0	0
			4502	1434	2291	366	398	13		

- Molecule 6 is a protein called Membrane protein insertase YidC.

Mol	Chain	Residues	Atoms						AltConf	Trace
6	C	455	Total	C	H	N	O	S	0	0
			7183	2336	3594	582	650	21		

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-4	MET	-	initiating methionine	UNP P25714
C	-3	ASP	-	expression tag	UNP P25714
C	-2	PRO	-	expression tag	UNP P25714
C	-1	SER	-	expression tag	UNP P25714
C	0	SER	-	expression tag	UNP P25714
C	1	ARG	-	expression tag	UNP P25714
C	228	ALA	GLU	conflict	UNP P25714
C	229	ALA	LYS	conflict	UNP P25714
C	231	ALA	GLU	conflict	UNP P25714
C	232	ALA	LYS	conflict	UNP P25714
C	234	ALA	LYS	conflict	UNP P25714
C	549	HIS	-	expression tag	UNP P25714
C	550	HIS	-	expression tag	UNP P25714
C	551	HIS	-	expression tag	UNP P25714
C	552	HIS	-	expression tag	UNP P25714
C	553	HIS	-	expression tag	UNP P25714
C	554	HIS	-	expression tag	UNP P25714



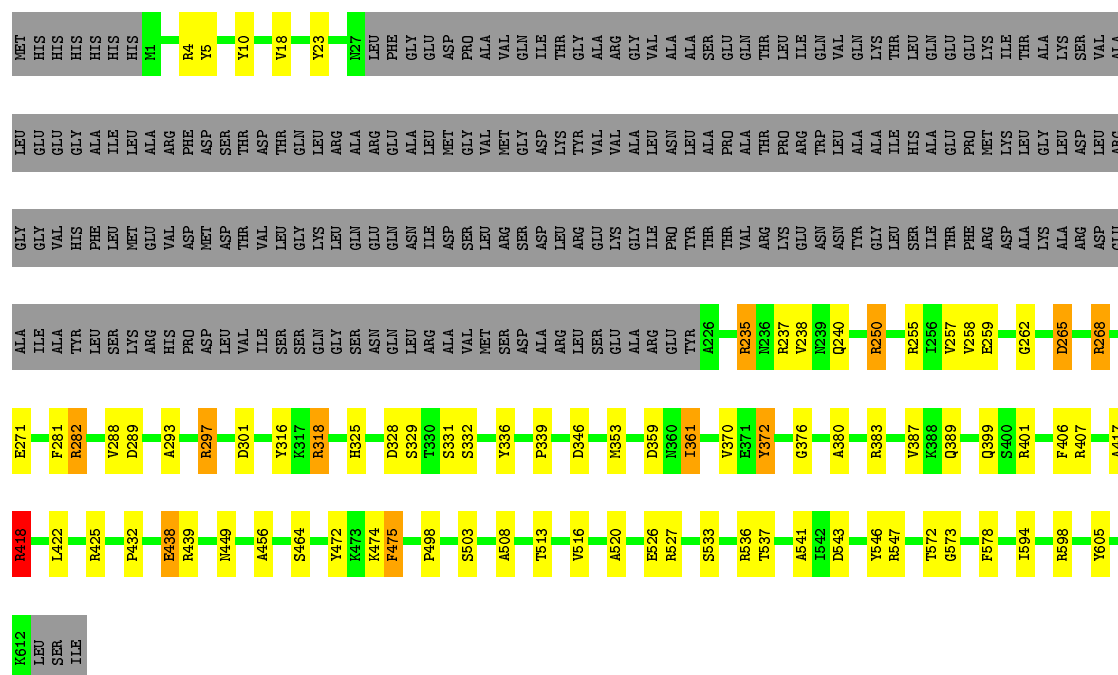
- Molecule 1: Protein translocase subunit SecY

[illegible]

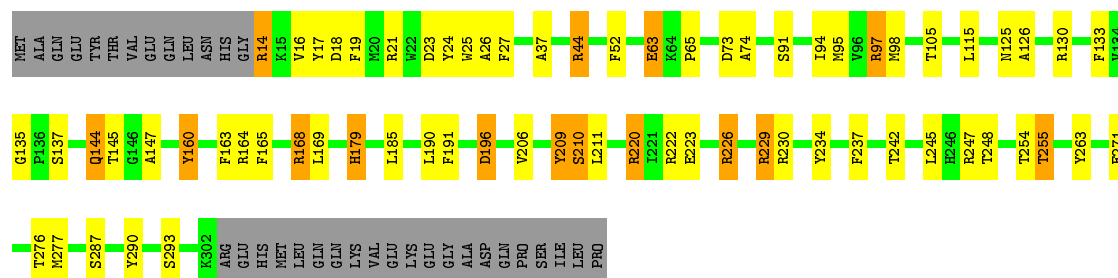
VAL	VAL	ILE	LEU	ILE	ALA	ALA	GLY	VAL	ALA	LEU	THR	T444	T449	R454	E455	A456	R457	T458	R461	W465	R468	T471	T475	W485	W490	R498	S501	F502	G505	L506	R507	F508	MET	HIS	HIS	HIS	HIS	HIS	ASP	ASP	ASP	LYS	ALA	MET	GLY	ALA	ASN	THR	GLU	ALA	GLN	GLY	GLY	SER	GLY	ARG	LEU	GLU	LEU	GLU	ALA	MET	LYS	TRP	VAL	VAL	VAL	VAL	VAL	ALA	LEU	LEU	VAL	ALA	ILE	VAL	GLY	ASN	TYR	LEU	TYR	ARG	ASP	MET	LEU	PRO	LEU	ARG	ALA	LEU	ALA
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

GLY	ALA	GLY	GLY	ALA	ALA	SER	THR	LEU	PHE	GLY	PRO	GLY	ILE	TYR	HIS	TRP	GLU	VAL	LEU	ARG	ILE	GLN	GLU	LEU	ILE	GLY	NET	TYR	GLU	ALA	LEU	VAL	VAL	PHE	LEU	ILE	VAL	ALA	ILE	GLY	LEU	VAL	GLN	LEU	ILE	MET	ASP	ALA	GLY	ALA	SER
VAL	THR	GLY	TRP	TYR	SER	GLY	LEU	GLY	SER	PRO	ILE	GLY	TYR	HIS	TRP	GLU	VAL	LEU	ARG	ILE	GLN	GLU	LEU	ILE	GLY	NET	TYR	GLU	ALA	LEU	VAL	VAL	PHE	LEU	ILE	VAL	ALA	ILE	GLY	LEU	VAL	GLN	LEU	ILE	MET	ASP	ALA	GLY	ALA	SER	

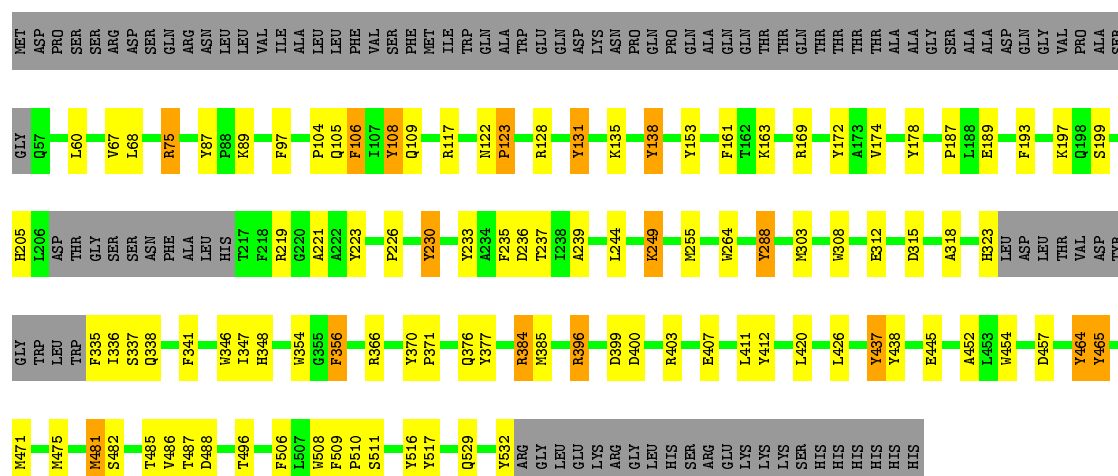
Chain D:  53% 11% 3% 33%



• Molecule 5: Protein translocase subunit SecF



• Molecule 6: Membrane protein insertase YidC





## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	53648	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	100	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	10	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	FEI FALCON I (4k x 4k)	Depositor

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

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	Y	2.09	27/3501 (0.8%)	2.13	116/4744 (2.4%)
2	E	1.78	6/518 (1.2%)	2.24	19/702 (2.7%)
3	G	1.72	3/238 (1.3%)	1.91	5/320 (1.6%)
4	D	1.62	20/3163 (0.6%)	2.06	82/4288 (1.9%)
5	F	1.59	4/2250 (0.2%)	2.18	77/3049 (2.5%)
6	C	1.62	16/3683 (0.4%)	1.99	97/5013 (1.9%)
All	All	1.76	76/13353 (0.6%)	2.09	396/18116 (2.2%)

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	Y	0	11
2	E	0	3
4	D	0	9
5	F	0	13
6	C	0	21
All	All	0	57

The worst 5 of 76 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	Y	317	TYR	CG-CD2	35.73	1.85	1.39
1	Y	317	TYR	CG-CD1	34.43	1.83	1.39
1	Y	317	TYR	CE2-CZ	33.58	1.82	1.38
1	Y	317	TYR	CE1-CZ	33.16	1.81	1.38
1	Y	317	TYR	CD1-CE1	23.16	1.74	1.39

The worst 5 of 396 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	F	226	ARG	NE-CZ-NH1	28.99	134.80	120.30
4	D	297	ARG	NE-CZ-NH1	22.63	131.61	120.30
1	Y	392	ARG	NE-CZ-NH1	18.04	129.32	120.30
4	D	268	ARG	NE-CZ-NH1	17.26	128.93	120.30
5	F	168	ARG	NE-CZ-NH1	17.14	128.87	120.30

There are no chirality outliers.

5 of 57 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	Y	113	ARG	Sidechain
1	Y	122	TYR	Sidechain
1	Y	22	ARG	Sidechain
1	Y	38	PHE	Sidechain
1	Y	7	LEU	Peptide

## 5.2 Too-close contacts ⓘ

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	Y	3423	3581	3583	0	0
2	E	510	552	551	0	0
3	G	236	244	243	0	0
4	D	3128	3291	3292	0	0
5	F	2211	2291	2290	0	0
6	C	3589	3594	3591	0	0
All	All	13097	13553	13550	0	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). Clashscore could not be calculated for this entry.

There are no clashes within the asymmetric unit.

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	Y	441/458 (96%)	367 (83%)	50 (11%)	24 (5%)	2	25
2	E	63/140 (45%)	57 (90%)	5 (8%)	1 (2%)	11	51
3	G	30/136 (22%)	29 (97%)	1 (3%)	0	100	100
4	D	410/622 (66%)	373 (91%)	25 (6%)	12 (3%)	5	38
5	F	287/323 (89%)	257 (90%)	24 (8%)	6 (2%)	8	45
6	C	449/559 (80%)	395 (88%)	40 (9%)	14 (3%)	5	37
All	All	1680/2238 (75%)	1478 (88%)	145 (9%)	57 (3%)	7	35

5 of 57 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	Y	48	VAL
1	Y	71	ALA
1	Y	113	ARG
1	Y	199	LEU
1	Y	305	THR

### 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 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	Y	359/374 (96%)	338 (94%)	21 (6%)	23	56
2	E	54/110 (49%)	52 (96%)	2 (4%)	39	68
3	G	27/106 (26%)	27 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	D	337/509 (66%)	328 (97%)	9 (3%)	50	74
5	F	237/267 (89%)	227 (96%)	10 (4%)	34	64
6	C	387/475 (82%)	375 (97%)	12 (3%)	45	71
All	All	1401/1841 (76%)	1347 (96%)	54 (4%)	41	66

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	297	ARG
4	D	474	LYS
6	C	376	GLN
4	D	339	PRO
4	D	418	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 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](#)

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 ⓘ

There are no chain breaks in this entry.