W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 21, 2024 – 04:56 PM EST

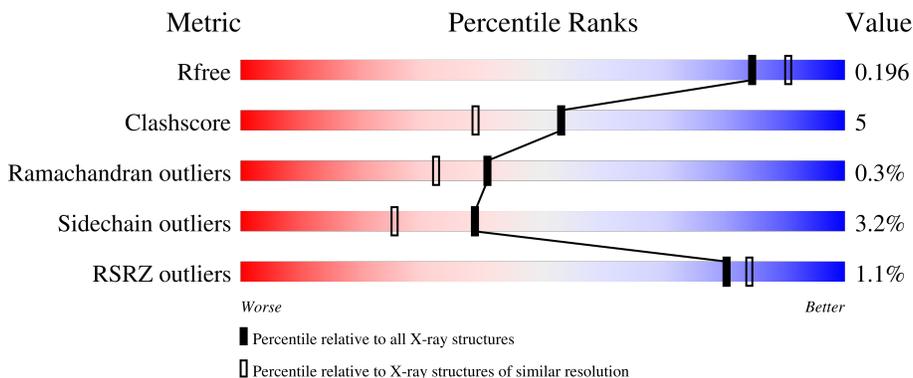|		
PDB ID	:	4OJP
Title	:	Crystal Structure of Putative Tailspike Protein (TSP1, orf210) from Escherichia coli O157:H7 Bacteriohage CBA120 in Complex with Maltose
Authors	:	Chen, C.; Herzberg, O.
Deposited on	:	2014-01-21
Resolution	:	1.95 Å(reported)

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 1.95 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	Percentile Ranks	Value
Rfree		0.196
Clashscore		5
Ramachandran outliers		0.3%
Sidechain outliers		3.2%
RSRZ outliers		1.1%

Worse　　　　　　　　　　　　　　　　　*Better*

▊ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	776	% 88% ... 9% ..
1	B	776	2% 84% ... 12% ..
1	C	776	% 87% ... 9% ..
2	D	2	100%
2	E	2	50% ... 50%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	2	100%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 19572 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 Tailspike protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	758	Total 5670	C 3576	N 954	O 1121	S 19	1	0	0
1	B	756	Total 5653	C 3568	N 947	O 1119	S 19	0	0	0
1	C	758	Total 5667	C 3573	N 955	O 1120	S 19	1	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	771	HIS	-	expression tag	UNP G3M189
A	772	HIS	-	expression tag	UNP G3M189
A	773	HIS	-	expression tag	UNP G3M189
A	774	HIS	-	expression tag	UNP G3M189
A	775	HIS	-	expression tag	UNP G3M189
A	776	HIS	-	expression tag	UNP G3M189
B	771	HIS	-	expression tag	UNP G3M189
B	772	HIS	-	expression tag	UNP G3M189
B	773	HIS	-	expression tag	UNP G3M189
B	774	HIS	-	expression tag	UNP G3M189
B	775	HIS	-	expression tag	UNP G3M189
B	776	HIS	-	expression tag	UNP G3M189
C	771	HIS	-	expression tag	UNP G3M189
C	772	HIS	-	expression tag	UNP G3M189
C	773	HIS	-	expression tag	UNP G3M189
C	774	HIS	-	expression tag	UNP G3M189
C	775	HIS	-	expression tag	UNP G3M189
C	776	HIS	-	expression tag	UNP G3M189

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	D	2	Total 23	C 12	O 11	0	0	0
2	E	2	Total 23	C 12	O 11	0	0	0
2	F	2	Total 23	C 12	O 11	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 4 is water.

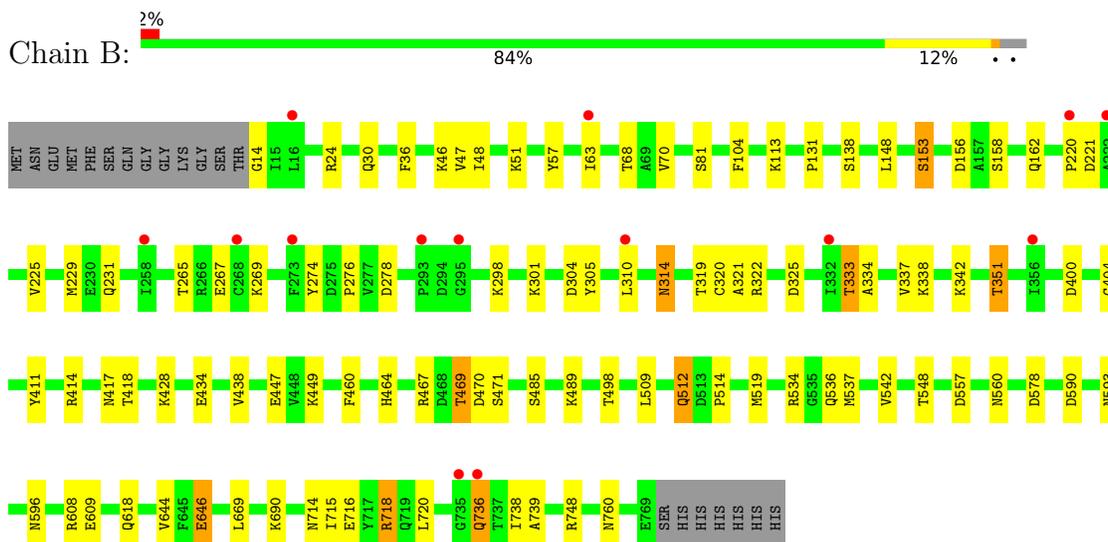| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	774	Total 774	O 774	0	0
4	B	819	Total 819	O 819	0	0
4	C	919	Total 919	O 919	0	0

3 Residue-property plots ⓘ

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.
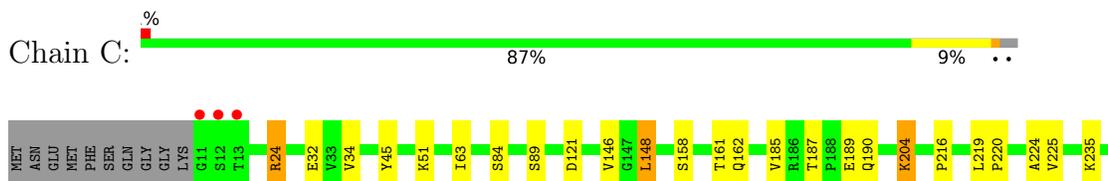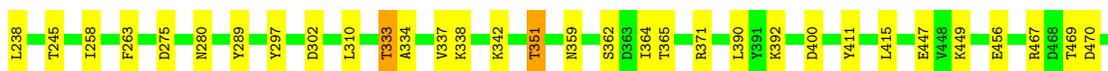
• Molecule 1: Tailspike protein
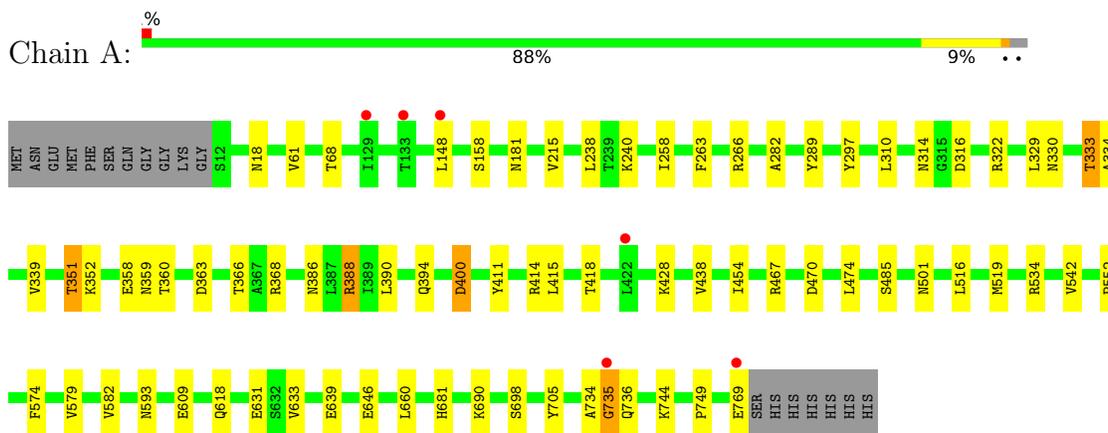
• Molecule 1: Tailspike protein

• Molecule 1: Tailspike protein

L238 T245 I258 F263 D275 N280 Y289 Y297 D302 L310 T333 A334 V337 K338 K342 T351 N359 S362 D363 I364 T365 R371 L390 Y391 K392 D400 Y411 L415 E447 V448 K449 E456 R467 D468 T469 D470

F475 S500 L509 M519 R534 V542 P552 F574 D578 N596 T604 E609 Q618 E631 S632 V633 E689 K690 I715 E716 G735 Q736 A739 T750 N760 I768 GLU SER HIS HIS HIS HIS HIS HIS

• Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain D:

100%

GLC1 GLC2

• Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain E:

50%   50%

GLC1 GLC2

• Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain F:

100%

GLC1 GLC2

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	123.44Å 147.98Å 171.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	61.70 – 1.95 67.91 – 1.95	Depositor EDS
% Data completeness (in resolution range)	98.5 (61.70-1.95) 98.5 (67.91-1.95)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 1.95Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.192 , 0.220 0.194 , 0.196	Depositor DCC
R_{free} test set	11323 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	22.6	Xtriage
Anisotropy	0.425	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 46.5	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	19572	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.80% 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 ⓘ

5.1 Standard geometry ⓘ

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

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	0.38	0/5774	0.65	0/7850
1	B	0.39	0/5757	0.68	2/7828 (0.0%)
1	C	0.41	0/5771	0.68	0/7846
All	All	0.39	0/17302	0.67	2/23524 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	718	ARG	NE-CZ-NH2	-7.37	116.62	120.30
1	B	718	ARG	NE-CZ-NH1	6.81	123.70	120.30

There are no chirality outliers.

There are no planarity outliers.

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	A	5670	0	5629	50	0
1	B	5653	0	5600	90	0
1	C	5667	0	5624	49	0
2	D	23	0	21	0	0
2	E	23	0	21	0	0
2	F	23	0	21	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
4	A	774	0	0	17	0
4	B	819	0	0	53	1
4	C	919	0	0	24	0
All	All	19572	0	16916	183	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 5.

The worst 5 of 183 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:C:371:ARG:HB2	4:C:1746:HOH:O	1.28	1.27
1:C:235:LYS:HE3	4:C:1540:HOH:O	1.38	1.20
1:B:47:VAL:HG13	4:B:1701:HOH:O	1.38	1.19
1:C:338:LYS:HA	4:C:1746:HOH:O	1.46	1.16
1:A:593:ASN:ND2	4:A:1358:HOH:O	1.81	1.12

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)
4:B:1282:HOH:O	4:B:1674:HOH:O[2_585]	1.57	0.63

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	756/776 (97%)	727 (96%)	26 (3%)	3 (0%)	34	24
1	B	754/776 (97%)	724 (96%)	28 (4%)	2 (0%)	41	32

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	756/776 (97%)	731 (97%)	23 (3%)	2 (0%)	41	32
All	All	2266/2328 (97%)	2182 (96%)	77 (3%)	7 (0%)	41	32

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	736	GLN
1	B	736	GLN
1	C	735	GLY
1	C	736	GLN
1	A	734	ALA

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 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	619/637 (97%)	603 (97%)	16 (3%)	46	32
1	B	615/637 (96%)	596 (97%)	19 (3%)	40	26
1	C	618/637 (97%)	594 (96%)	24 (4%)	32	17
All	All	1852/1911 (97%)	1793 (97%)	59 (3%)	39	25

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	469	THR
1	C	509	LEU
1	C	24	ARG
1	C	469	THR
1	C	390	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	262	ASN
1	C	280	ASN
1	C	327	ASN
1	A	501	ASN
1	A	681	HIS

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

5.4   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5   Carbohydrates ⓘ

6 monosaccharides 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	$\#\|Z\| > 2$
2	GLC	D	1	2	12,12,12	0.54	0	17,17,17	0.70	0
2	GLC	D	2	2	11,11,12	0.50	0	15,15,17	0.93	0
2	GLC	E	1	2	12,12,12	0.57	0	17,17,17	0.94	0
2	GLC	E	2	2	11,11,12	0.56	0	15,15,17	0.92	1 (6%)
2	GLC	F	1	2	12,12,12	0.54	0	17,17,17	0.85	0
2	GLC	F	2	2	11,11,12	0.57	0	15,15,17	0.66	0

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
2	GLC	D	1	2	-	1/2/22/22	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GLC	D	2	2	-	0/2/19/22	0/1/1/1
2	GLC	E	1	2	-	0/2/22/22	0/1/1/1
2	GLC	E	2	2	-	0/2/19/22	0/1/1/1
2	GLC	F	1	2	-	2/2/22/22	0/1/1/1
2	GLC	F	2	2	-	2/2/19/22	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	2	GLC	O5-C5-C6	2.60	111.28	107.20

There are no chirality outliers.

All (5) torsion outliers are listed below:

| Mol	Chain	Res	Type	Atoms
2	F	1	GLC	O5-C5-C6-O6
2	F	1	GLC	C4-C5-C6-O6
2	F	2	GLC	C4-C5-C6-O6
2	F	2	GLC	O5-C5-C6-O6
2	D	1	GLC	C4-C5-C6-O6

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

Oligosaccharide Chain D

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain E

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain F

Bond lengths | Bond angles

Torsions | Rings
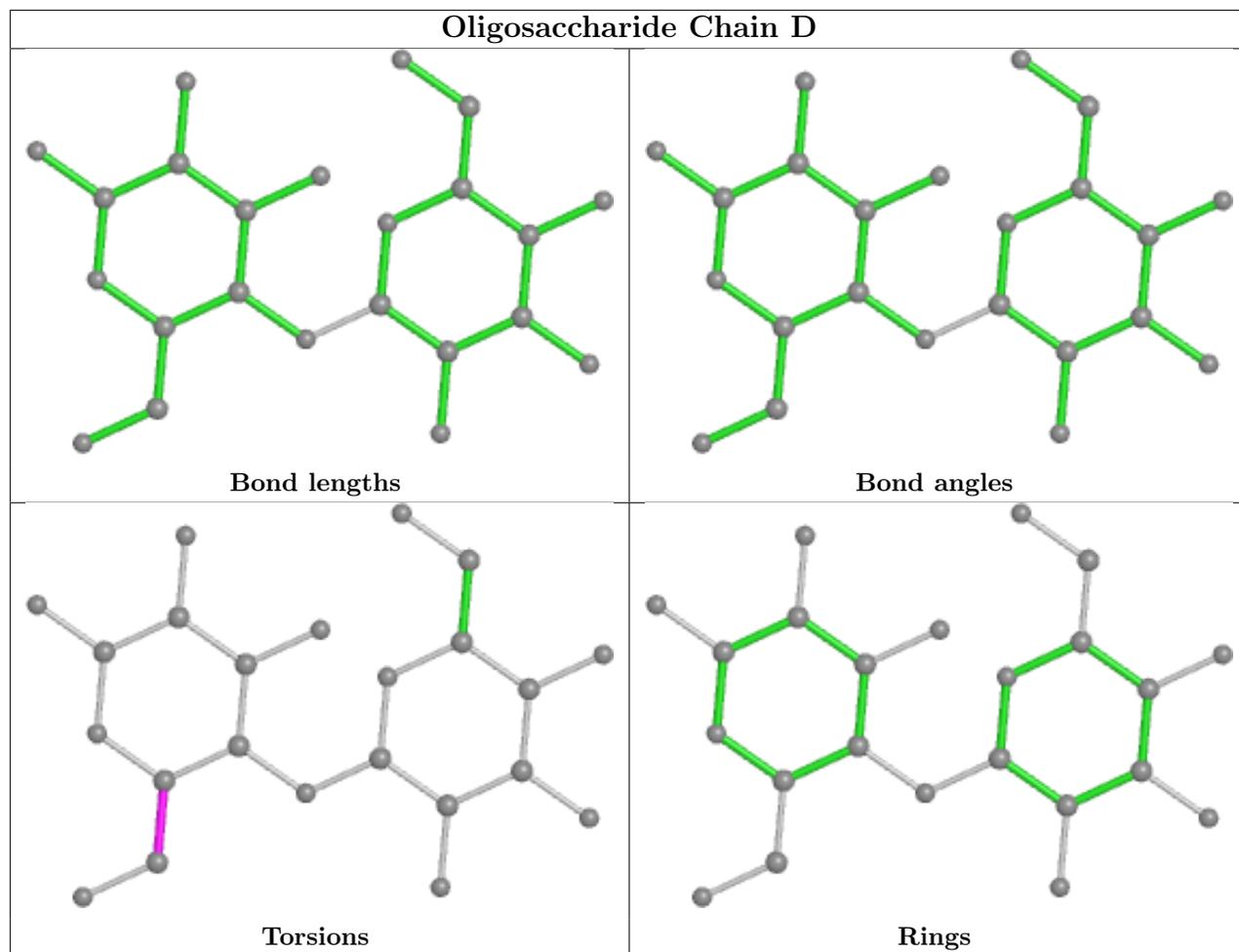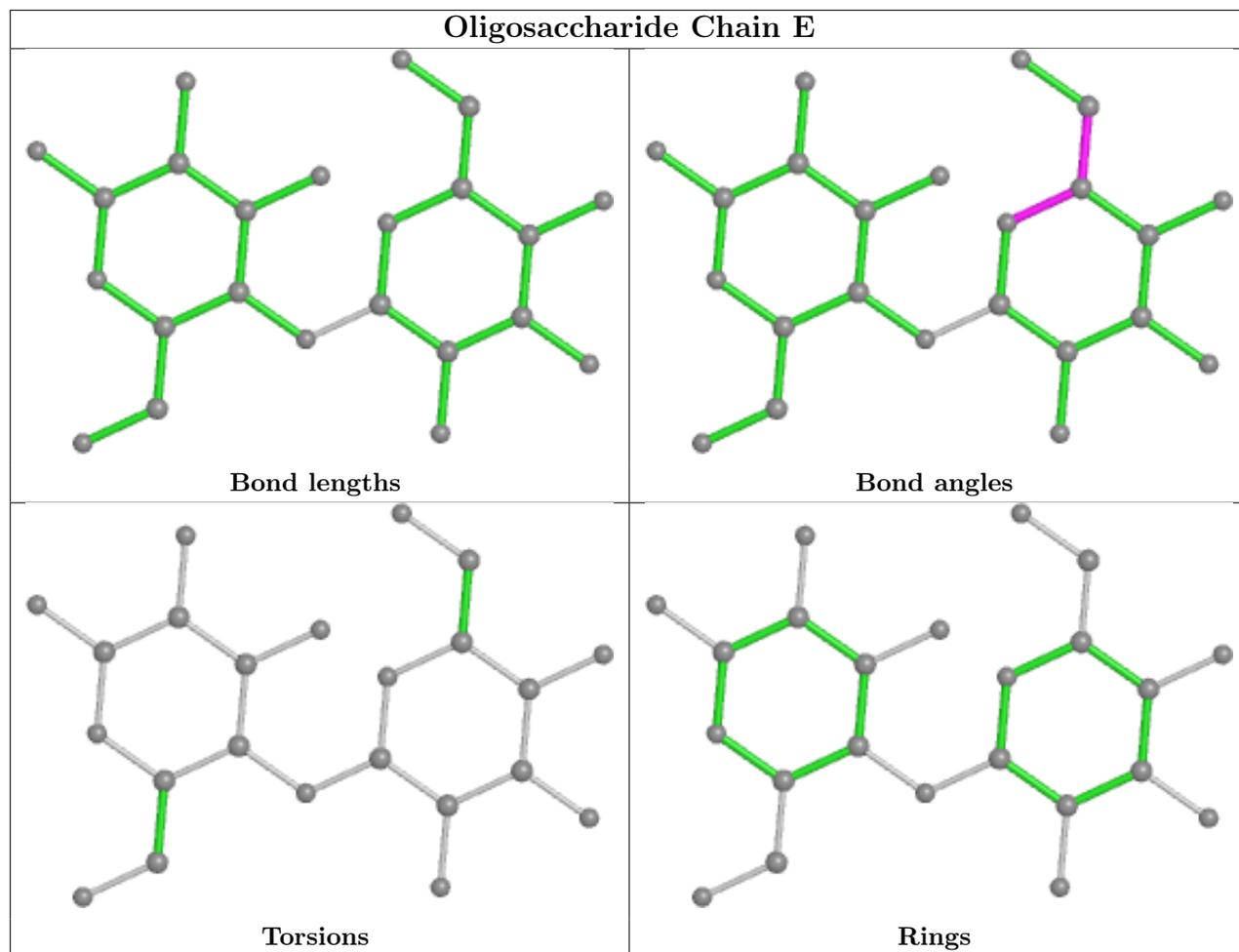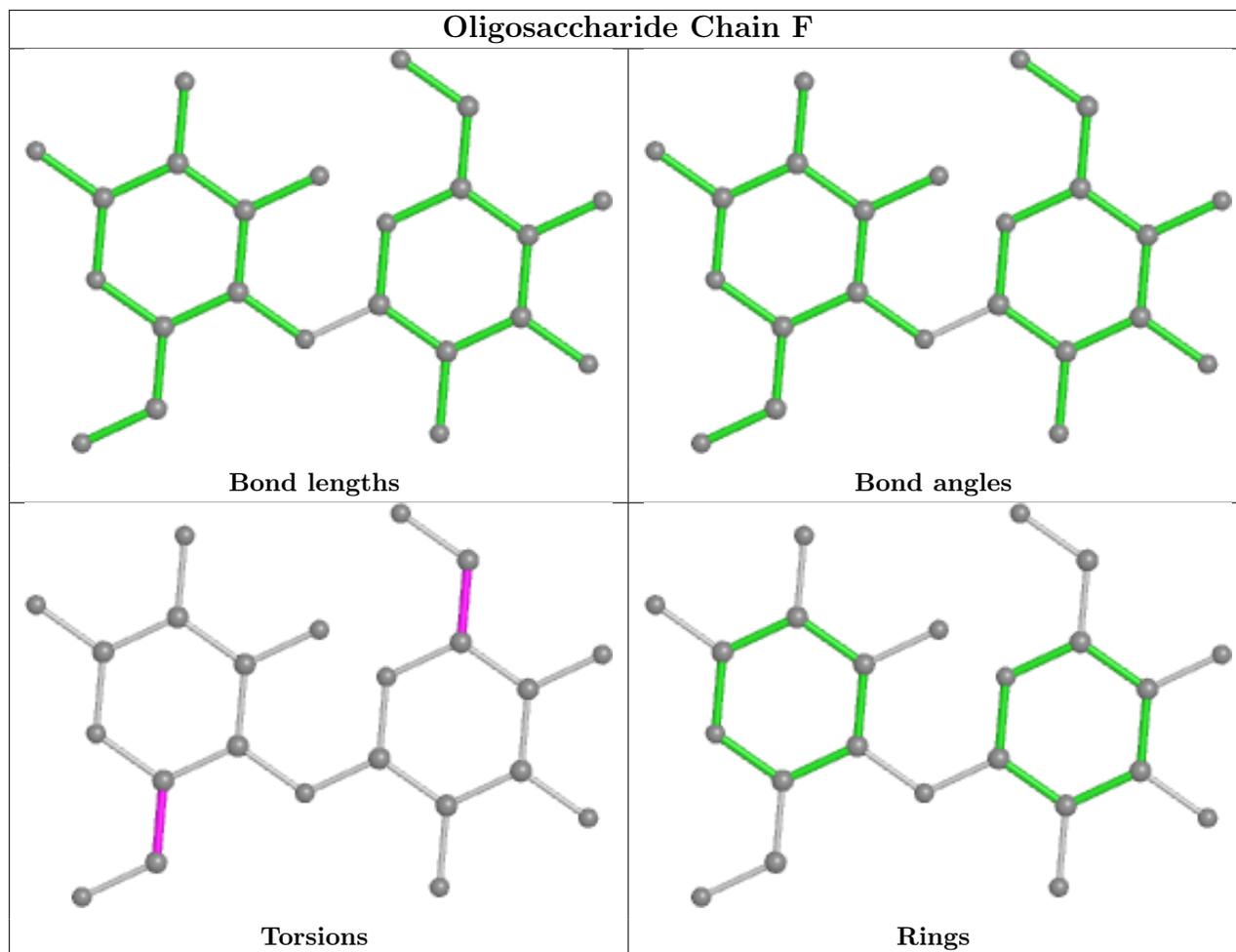
5.6 Ligand geometry (i)

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 ⓘ

6.1   Protein, DNA and RNA chains ⓘ

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	758/776 (97%)	-0.04	6 (0%)	86	89	8, 18, 31, 58	1 (0%)
1	B	756/776 (97%)	0.03	14 (1%)	66	72	5, 15, 32, 49	0
1	C	758/776 (97%)	-0.21	4 (0%)	91	93	3, 12, 29, 67	0
All	All	2272/2328 (97%)	-0.07	24 (1%)	80	84	3, 15, 31, 67	1 (0%)

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	735	GLY	5.2
1	C	11	GLY	5.2
1	A	129	ILE	3.6
1	B	16	LEU	3.4
1	B	63	ILE	3.4

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3   Carbohydrates ⓘ

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.
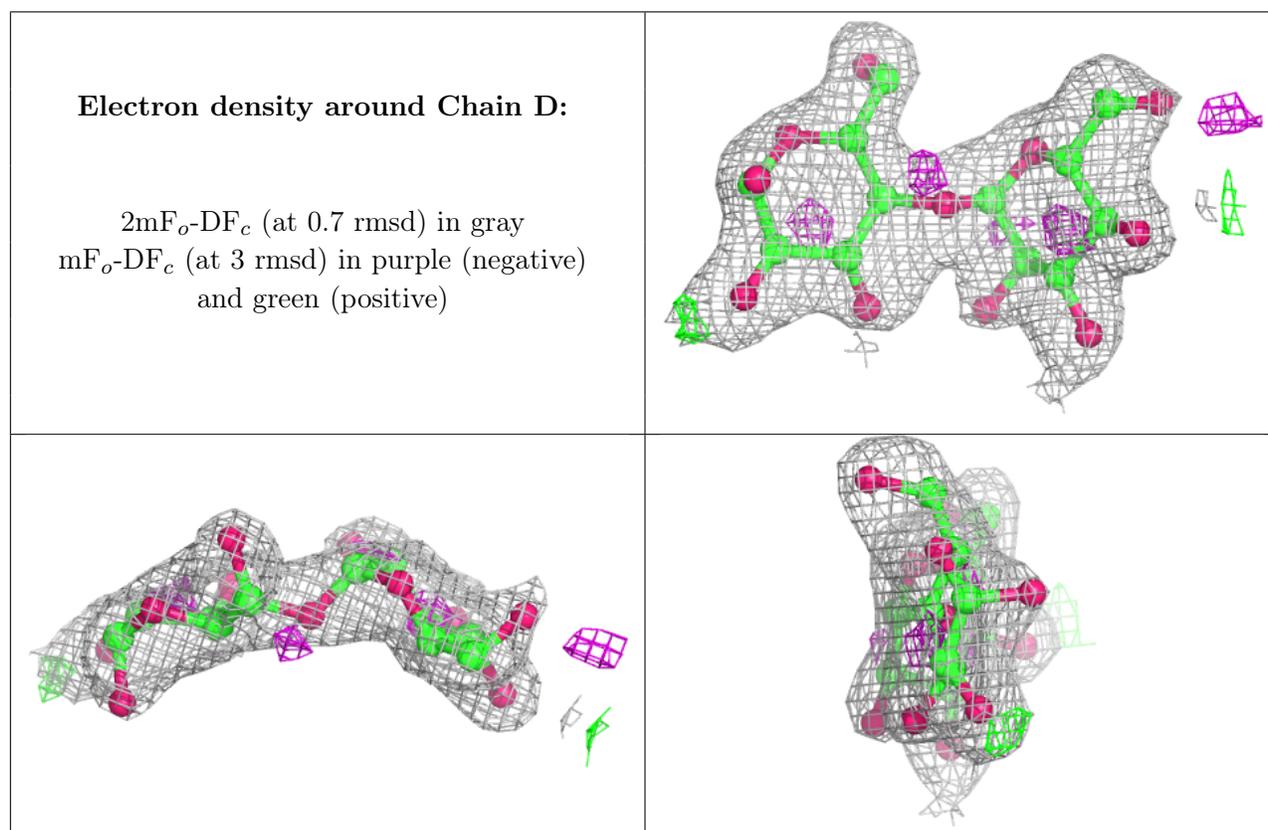
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	GLC	D	1	12/12	0.89	0.28	26,38,42,42	0
2	GLC	E	1	12/12	0.91	0.26	25,38,41,44	0
2	GLC	D	2	11/12	0.92	0.28	25,35,40,41	0

Continued on next page...

Continued from previous page...

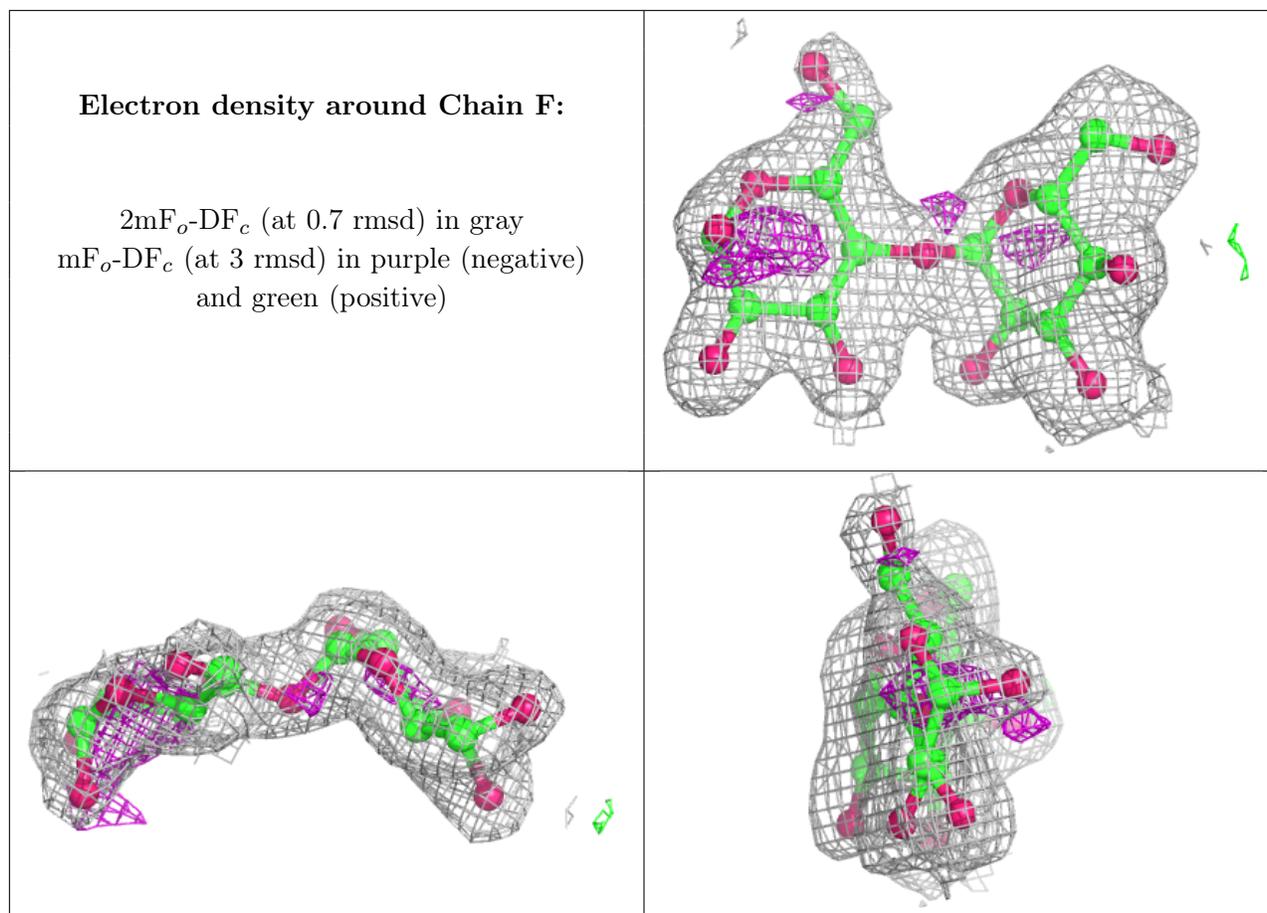| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	GLC	F	1	12/12	0.92	0.27	15,27,36,38	0
2	GLC	F	2	11/12	0.93	0.18	19,25,29,30	0
2	GLC	E	2	11/12	0.94	0.17	24,31,37,40	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around Chain D:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around Chain E:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
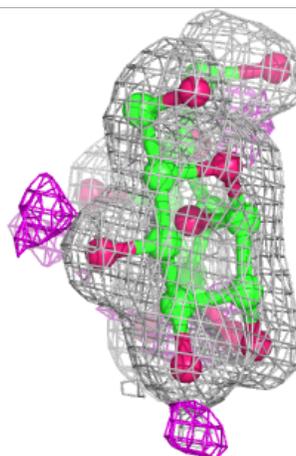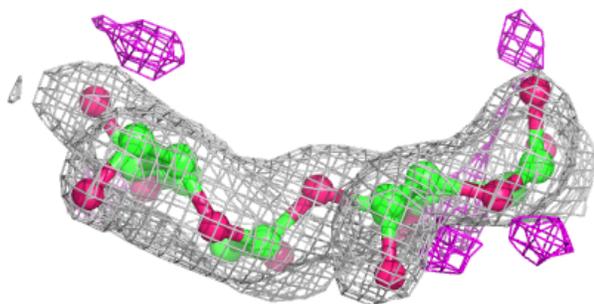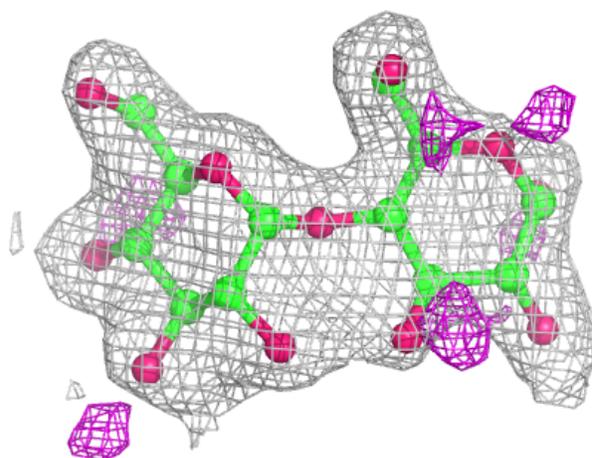
Electron density around Chain F:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative) and green (positive)

6.4 Ligands ⓘ

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(Å²)	Q<0.9
3	ZN	A	802[A]	1/1	0.99	0.06	32,32,32,32	1

6.5 Other polymers ⓘ

There are no such residues in this entry.