



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 23, 2023 – 07:09 AM EDT

PDB ID : 3C5X  
Title : Crystal structure of the precursor membrane protein- envelope protein heterodimer from the dengue 2 virus at low pH  
Authors : Li, L.  
Deposited on : 2008-02-01  
Resolution : 2.20 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  
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

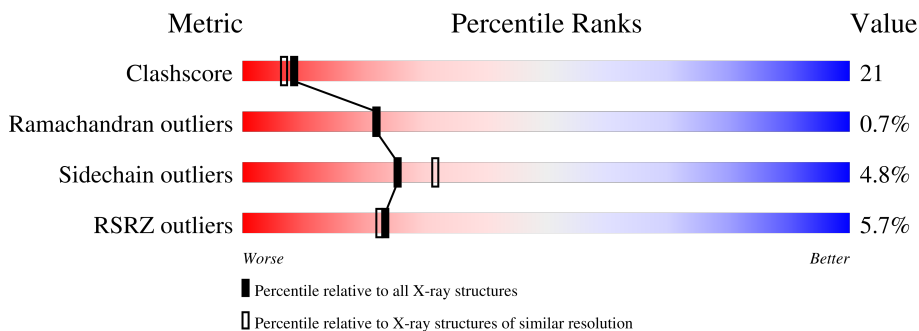
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	402	 7% 64% 29%
2	C	130	 47% 13% 38%
3	B	2	 50% 50%
4	D	5	 40% 60%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NDG	B	2	-	-	-	X
4	NAG	D	2	-	-	X	-
4	BMA	D	4	X	-	-	X
4	BMA	D	5	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3860 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 Envelope protein E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	390	3047	1926	525	569	27	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	GLY	-	expression tag	UNP O09234
A	-6	GLU	-	expression tag	UNP O09234
A	-5	ASN	-	expression tag	UNP O09234
A	-4	LEU	-	expression tag	UNP O09234
A	-3	TYR	-	expression tag	UNP O09234
A	-2	PHE	-	expression tag	UNP O09234
A	-1	GLN	-	expression tag	UNP O09234
A	0	GLY	-	expression tag	UNP O09234

- Molecule 2 is a protein called prM.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	81	640	396	106	128	10	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

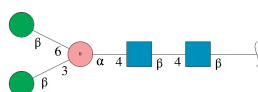
Chain	Residue	Modelled	Actual	Comment	Reference
C	87	SER	ARG	engineered mutation	UNP O09234
C	88	THR	ARG	engineered mutation	UNP O09234
C	91	SER	ARG	engineered mutation	UNP O09234

- Molecule 3 is an oligosaccharide called 2-acetamido-2-deoxy-alpha-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	B	2	28	16	2	10	0	0	0

- Molecule 4 is an oligosaccharide called beta-D-mannopyranose-(1-3)-[beta-D-mannopyranose-(1-6)]alpha-D-altropyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	D	5	61	34	2	25	0	0	0

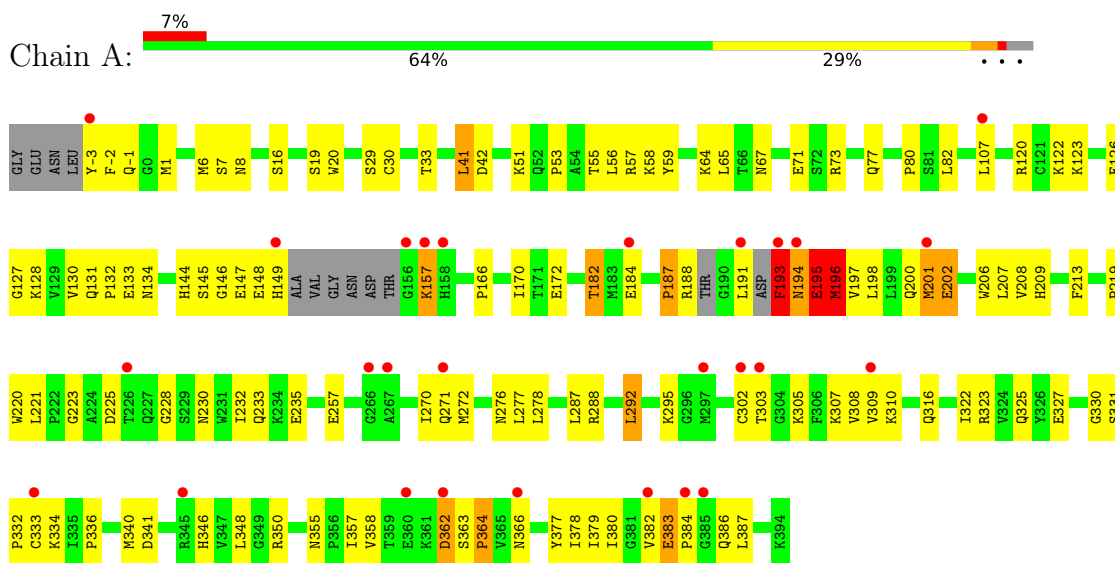
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	61	Total	O	0	0
			61	61		
5	C	23	Total	O	0	0
			23	23		

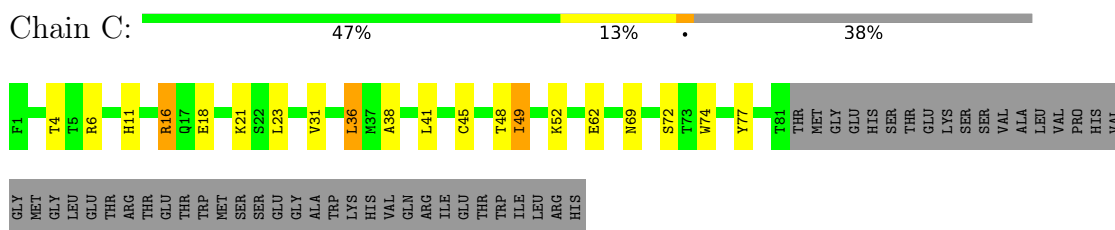
### 3 Residue-property plots [i](#)

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: Envelope protein E



- Molecule 2: prM



- Molecule 3: 2-acetamido-2-deoxy-alpha-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose



- Molecule 4: beta-D-mannopyranose-(1-3)-[beta-D-mannopyranose-(1-6)]alpha-D-altropyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose



MAG1  
MAG2  
SHD3  
EM44  
EM45

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.21Å 108.59Å 108.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 43.21 – 2.21	Depositor EDS
% Data completeness (in resolution range)	100.0 (50.00-2.20) 98.2 (43.21-2.21)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.74 (at 2.20Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.253 , 0.275 0.247 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.1	Xtrriage
Anisotropy	0.561	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 44.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.016 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3860	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	55.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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: SHD, BMA, NDG, NAG

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	0.34	1/3108 (0.0%)	0.72	7/4189 (0.2%)
2	C	0.53	1/652 (0.2%)	0.67	0/883
All	All	0.38	2/3760 (0.1%)	0.71	7/5072 (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	3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	49	ILE	C-N	-8.66	1.14	1.34
1	A	77	GLN	C-N	5.20	1.42	1.33

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	195	GLU	N-CA-C	13.72	148.06	111.00
1	A	196	MET	N-CA-CB	-10.48	91.74	110.60
1	A	195	GLU	CB-CA-C	-7.94	94.51	110.40
1	A	384	PRO	N-CA-C	5.92	127.50	112.10
1	A	196	MET	N-CA-C	-5.47	96.22	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	193	PHE	Peptide
1	A	194	ASN	Peptide
1	A	362	ASP	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	3047	0	3047	130	0
2	C	640	0	609	20	0
3	B	28	0	24	5	0
4	D	61	0	46	11	0
5	A	61	0	0	6	0
5	C	23	0	0	0	0
All	All	3860	0	3726	156	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 21.

The worst 5 of 156 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:A:195:GLU:CG	1:A:195:GLU:O	1.87	1.20
2:C:31:VAL:HG21	4:D:2:NAG:C8	1.72	1.19
2:C:31:VAL:CG2	4:D:2:NAG:H81	1.74	1.18
1:A:195:GLU:O	1:A:195:GLU:HG2	1.57	0.98
1:A:195:GLU:O	1:A:195:GLU:HG3	1.62	0.97

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	382/402 (95%)	365 (96%)	14 (4%)	3 (1%)	19	19
2	C	79/130 (61%)	74 (94%)	5 (6%)	0	100	100
All	All	461/532 (87%)	439 (95%)	19 (4%)	3 (1%)	22	22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	195	GLU
1	A	187	PRO
1	A	364	PRO

### 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	339/348 (97%)	324 (96%)	15 (4%)	28	35
2	C	75/118 (64%)	70 (93%)	5 (7%)	16	18
All	All	414/466 (89%)	394 (95%)	20 (5%)	25	32

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	383	GLU
2	C	36	LEU
2	C	62	GLU
2	C	41	LEU
1	A	196	MET

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	A	316	GLN
1	A	325	GLN
1	A	366	ASN
1	A	233	GLN
1	A	242	ASN

### 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](#)

7 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	$\# Z  > 2$
3	NAG	B	1	3	14,14,15	0.61	0	17,19,21	0.73	0
3	NDG	B	2	3	14,14,15	0.58	0	17,19,21	0.96	1 (5%)
4	NAG	D	1	4,2	14,14,15	0.83	1 (7%)	17,19,21	1.43	1 (5%)
4	NAG	D	2	4	14,14,15	0.55	0	17,19,21	1.73	4 (23%)
4	SHD	D	3	4	11,11,12	1.75	3 (27%)	15,15,17	6.06	5 (33%)
4	BMA	D	4	4	11,11,12	0.87	0	15,15,17	1.19	2 (13%)
4	BMA	D	5	4	11,11,12	0.50	0	15,15,17	1.71	3 (20%)

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
3	NAG	B	1	3	-	3/6/23/26	0/1/1/1
3	NDG	B	2	3	-	5/6/23/26	0/1/1/1
4	NAG	D	1	4,2	-	2/6/23/26	0/1/1/1
4	NAG	D	2	4	-	2/6/23/26	0/1/1/1
4	SHD	D	3	4	-	2/2/19/22	0/1/1/1
4	BMA	D	4	4	1/1/4/5	0/2/19/22	1/1/1/1
4	BMA	D	5	4	-	2/2/19/22	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	3	SHD	C2-C3	-4.03	1.46	1.52
4	D	3	SHD	O3-C3	-2.68	1.36	1.43
4	D	3	SHD	C4-C3	2.53	1.58	1.52
4	D	1	NAG	O5-C1	2.03	1.47	1.43

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	3	SHD	O2-C2-C3	-19.04	71.99	110.14
4	D	3	SHD	O4-C4-C3	9.79	132.98	110.35
4	D	3	SHD	C1-C2-C3	7.38	118.74	109.67
4	D	3	SHD	O3-C3-C4	-5.31	98.08	110.35
4	D	2	NAG	C3-C4-C5	-4.29	102.59	110.24

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	D	4	BMA	C1

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	1	NAG	C8-C7-N2-C2
3	B	1	NAG	O7-C7-N2-C2
3	B	2	NDG	C8-C7-N2-C2
3	B	2	NDG	O7-C7-N2-C2
4	D	3	SHD	C4-C5-C6-O6

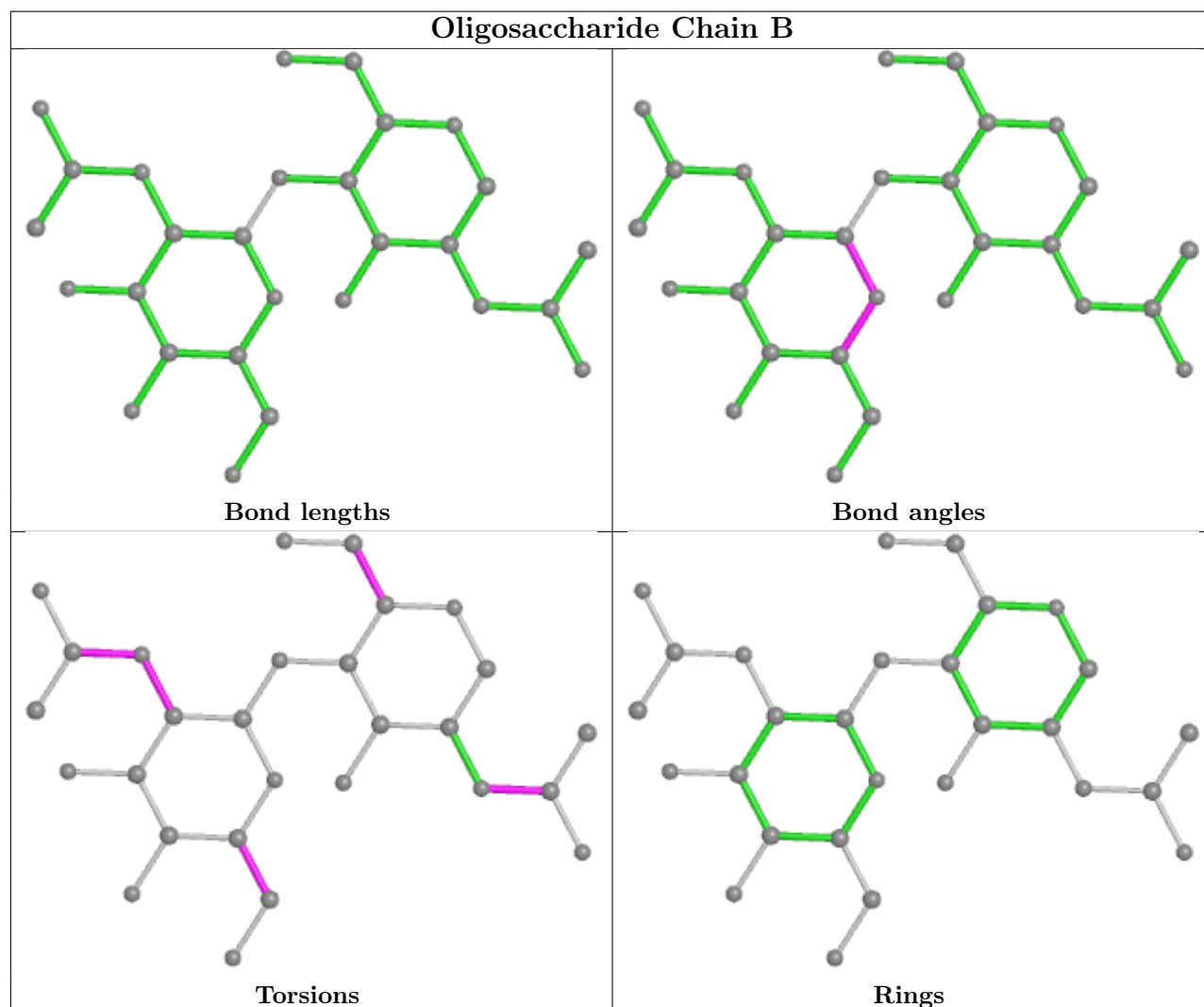
All (1) ring outliers are listed below:

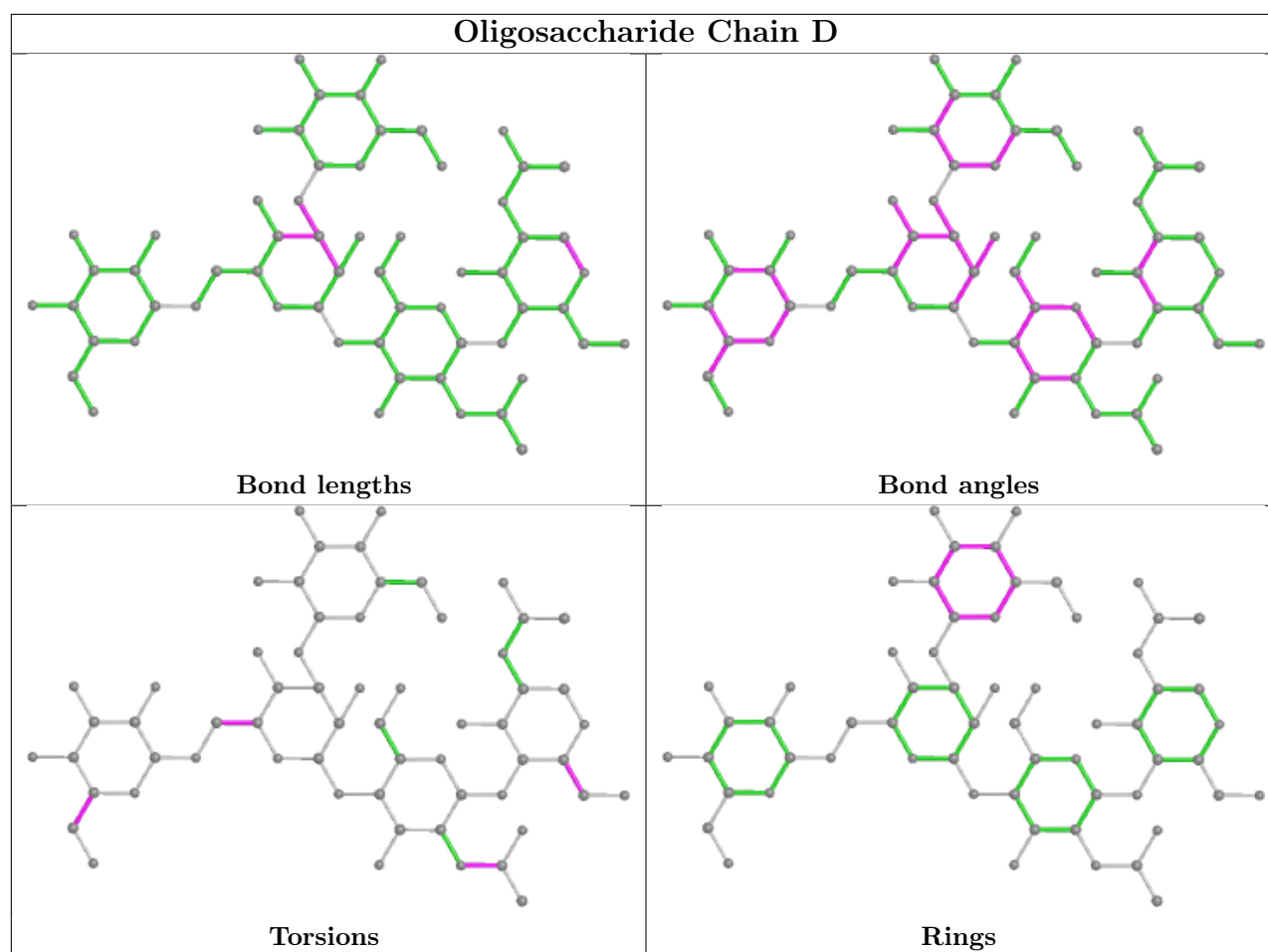
Mol	Chain	Res	Type	Atoms
4	D	4	BMA	C1-C2-C3-C4-C5-O5

5 monomers are involved in 16 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	1	NAG	2	0
4	D	2	NAG	10	0
4	D	3	SHD	1	0
3	B	2	NDG	4	0
3	B	1	NAG	5	0

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





## 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 [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	49:ILE	C	50:THR	N	1.14



## 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	390/402 (97%)	0.48	27 (6%) 16 15	30, 52, 87, 113	0
2	C	81/130 (62%)	0.34	0 100 100	22, 41, 73, 97	0
All	All	471/532 (88%)	0.46	27 (5%) 23 22	22, 50, 85, 113	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	149	HIS	6.3
1	A	157	LYS	4.0
1	A	-3	TYR	3.9
1	A	303	THR	3.9
1	A	309	VAL	3.3

### 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](#)

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.

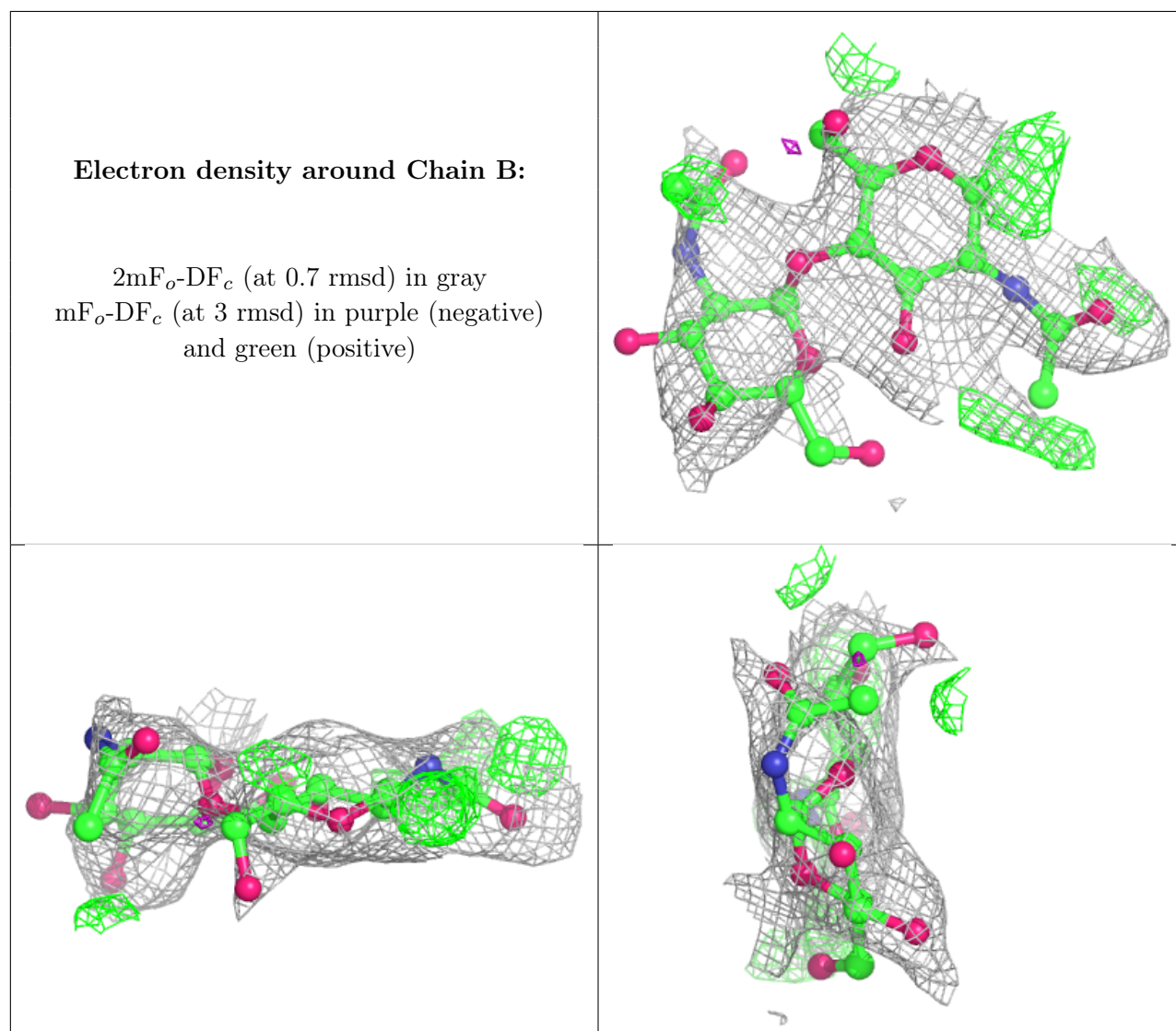
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
4	BMA	D	4	11/12	0.47	0.49	119,120,121,122	0
4	SHD	D	3	11/12	0.58	0.40	124,127,130,135	0
4	BMA	D	5	11/12	0.63	0.47	138,141,142,142	0
3	NAG	B	1	14/15	0.66	0.21	79,84,91,108	0

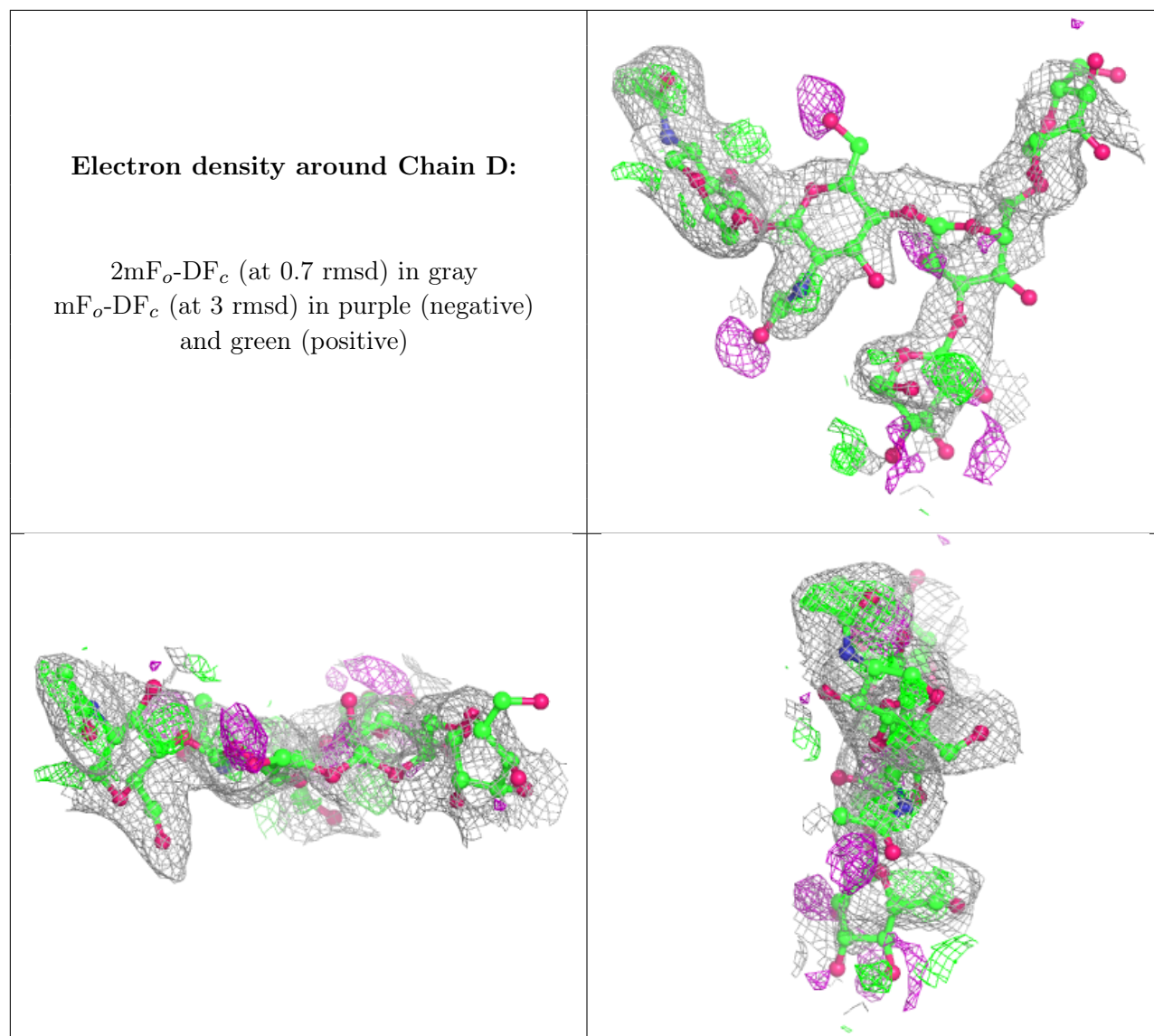
*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NDG	B	2	14/15	0.67	0.41	118,132,133,133	0
4	NAG	D	2	14/15	0.71	0.40	104,119,121,124	0
4	NAG	D	1	14/15	0.75	0.18	64,69,76,93	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.





## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.