WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 8, 2024 – 05:16 am GMT

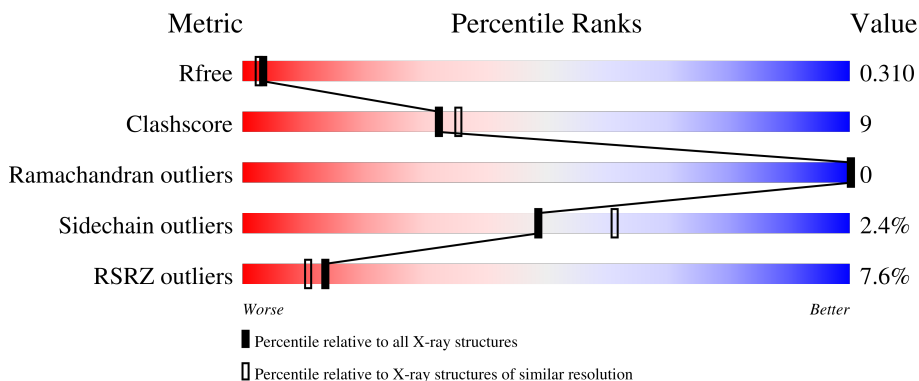|		
PDB ID	:	6GH8
Title	:	Crystal structure of GP1 domain of Lujo virus in complex with the first CUB domain of neuropilin-2
Authors	:	Cohen-Dvashi, H.; Kilimnik, I.; Diskin, R.
Deposited on	:	2018-05-06
Resolution	:	2.44 Å(reported)

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 2.44 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		0.310
Clashscore		9
Ramachandran outliers		0
Sidechain outliers		2.4%
RSRZ outliers		7.6%

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	1564 (2.46-2.42)
Clashscore	141614	1631 (2.46-2.42)
Ramachandran outliers	138981	1617 (2.46-2.42)
Sidechain outliers	138945	1617 (2.46-2.42)
RSRZ outliers	127900	1547 (2.46-2.42)

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	B	139	14% / 63% 15% • 21%
1	D	139	4% / 60% 18% • 21%
2	A	133	5% / 74% 16% • 10%
2	C	133	2% / 76% 14% 10%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 3801 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 Glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	110	Total	C	N	O	S	0	0	0
			908	568	163	173	4			
1	B	110	Total	C	N	O	S	0	0	0
			908	568	163	173	4			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	61	ALA	-	expression tag	UNP C5ILC1
D	62	ASP	-	expression tag	UNP C5ILC1
D	63	LEU	-	expression tag	UNP C5ILC1
D	64	GLY	-	expression tag	UNP C5ILC1
D	65	SER	-	expression tag	UNP C5ILC1
D	66	HIS	-	expression tag	UNP C5ILC1
D	67	HIS	-	expression tag	UNP C5ILC1
D	68	HIS	-	expression tag	UNP C5ILC1
D	69	HIS	-	expression tag	UNP C5ILC1
D	70	HIS	-	expression tag	UNP C5ILC1
D	71	HIS	-	expression tag	UNP C5ILC1
D	72	GLY	-	expression tag	UNP C5ILC1
D	73	GLY	-	expression tag	UNP C5ILC1
B	61	ALA	-	expression tag	UNP C5ILC1
B	62	ASP	-	expression tag	UNP C5ILC1
B	63	LEU	-	expression tag	UNP C5ILC1
B	64	GLY	-	expression tag	UNP C5ILC1
B	65	SER	-	expression tag	UNP C5ILC1
B	66	HIS	-	expression tag	UNP C5ILC1
B	67	HIS	-	expression tag	UNP C5ILC1
B	68	HIS	-	expression tag	UNP C5ILC1
B	69	HIS	-	expression tag	UNP C5ILC1
B	70	HIS	-	expression tag	UNP C5ILC1
B	71	HIS	-	expression tag	UNP C5ILC1
B	72	GLY	-	expression tag	UNP C5ILC1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	73	GLY	-	expression tag	UNP C5ILC1

• Molecule 2 is a protein called Neuropilin-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	120	Total	C	N	O	S	0	0	0
			950	603	160	182	5			
2	C	120	Total	C	N	O	S	0	0	0
			950	603	160	182	5			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	22	ALA	-	expression tag	UNP O60462
A	23	ASP	-	expression tag	UNP O60462
A	24	LEU	-	expression tag	UNP O60462
A	25	GLY	-	expression tag	UNP O60462
A	26	SER	-	expression tag	UNP O60462
A	123	ARG	LYS	conflict	UNP O60462
A	147	GLY	-	expression tag	UNP O60462
A	148	THR	-	expression tag	UNP O60462
A	149	HIS	-	expression tag	UNP O60462
A	150	HIS	-	expression tag	UNP O60462
A	151	HIS	-	expression tag	UNP O60462
A	152	HIS	-	expression tag	UNP O60462
A	153	HIS	-	expression tag	UNP O60462
A	154	HIS	-	expression tag	UNP O60462
C	22	ALA	-	expression tag	UNP O60462
C	23	ASP	-	expression tag	UNP O60462
C	24	LEU	-	expression tag	UNP O60462
C	25	GLY	-	expression tag	UNP O60462
C	26	SER	-	expression tag	UNP O60462
C	123	ARG	LYS	conflict	UNP O60462
C	147	GLY	-	expression tag	UNP O60462
C	148	THR	-	expression tag	UNP O60462
C	149	HIS	-	expression tag	UNP O60462
C	150	HIS	-	expression tag	UNP O60462
C	151	HIS	-	expression tag	UNP O60462
C	152	HIS	-	expression tag	UNP O60462
C	153	HIS	-	expression tag	UNP O60462
C	154	HIS	-	expression tag	UNP O60462

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	D	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Ca	0	0
			1	1		
4	C	1	Total	Ca	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	11	Total	O	0	0
			11	11		
5	A	9	Total	O	0	0
			9	9		
5	B	4	Total	O	0	0
			4	4		

Continued on next page...

Continued from previous page...

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	17	Total 17	O 17	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.
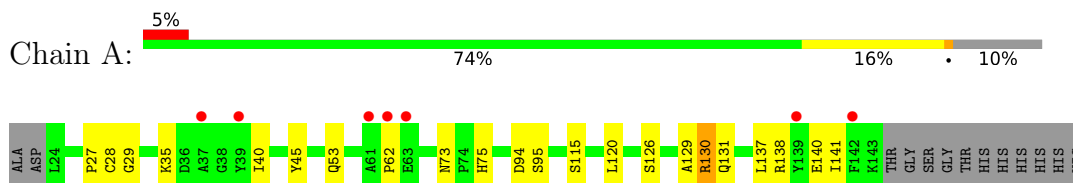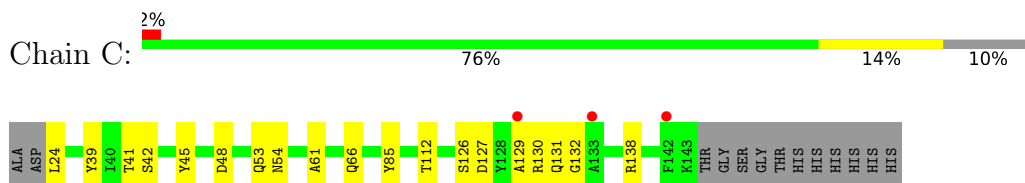
• Molecule 1: Glycoprotein
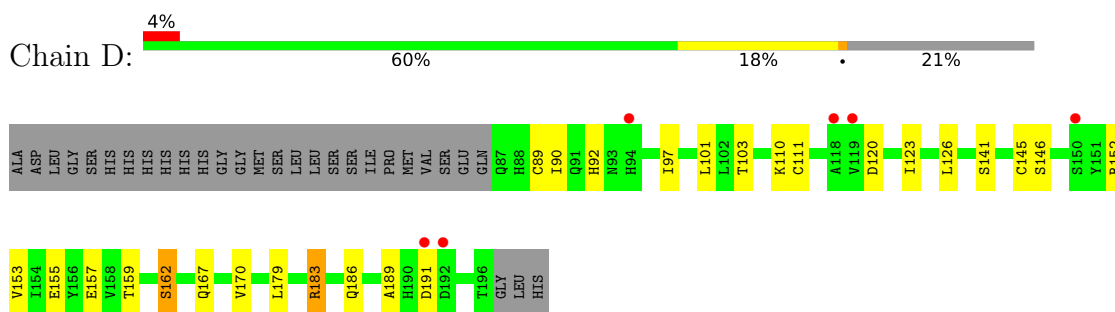
• Molecule 1: Glycoprotein
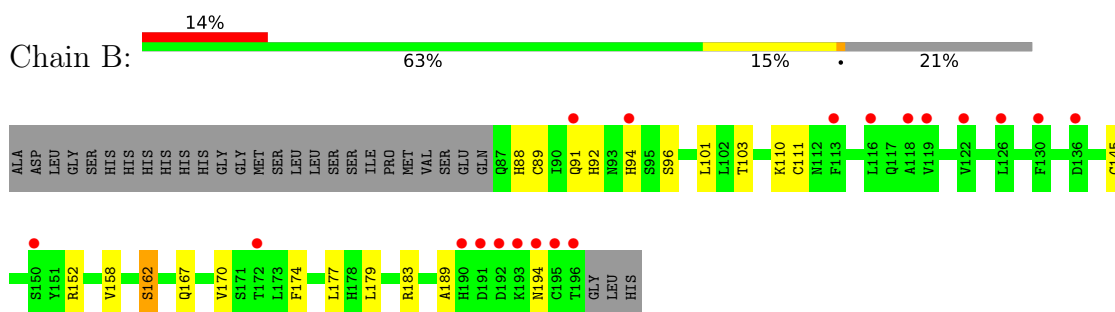
• Molecule 2: Neuropilin-2

• Molecule 2: Neuropilin-2

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	141.81Å 59.29Å 83.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	72.00 – 2.44 72.00 – 2.43	Depositor EDS
% Data completeness (in resolution range)	99.1 (72.00-2.44) 99.4 (72.00-2.43)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.45 (at 2.42Å)	Xtriage
Refinement program	PHENIX (dev_2871: ???)	Depositor
R, R_{free}	0.266 , 0.311 0.265 , 0.310	Depositor DCC
R_{free} test set	1341 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	48.3	Xtriage
Anisotropy	0.615	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 50.4	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	0.92	EDS
Total number of atoms	3801	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 21.56 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.8365e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: CA, 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	B	0.28	0/928	0.58	0/1255
1	D	0.33	0/928	0.54	0/1255
2	A	0.28	0/978	0.49	0/1325
2	C	0.27	0/978	0.50	0/1325
All	All	0.29	0/3812	0.53	0/5160

There are no bond length outliers.

There are no bond angle outliers.

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	B	908	0	866	16	0
1	D	908	0	867	23	0
2	A	950	0	891	18	0
2	C	950	0	891	13	0
3	B	28	0	26	1	0
3	D	14	0	13	0	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
5	A	9	0	0	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	4	0	0	1	0
5	C	17	0	0	2	0
5	D	11	0	0	2	0
All	All	3801	0	3554	64	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 9.

The worst 5 of 64 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)
2:C:112:THR:O	5:C:301:HOH:O	1.87	0.92
1:B:91:GLN:O	1:B:92:HIS:ND1	2.14	0.81
2:A:35:LYS:O	2:A:35:LYS:NZ	2.16	0.76
1:B:91:GLN:OE1	1:B:194:ASN:HB3	1.89	0.73
1:D:103:THR:HG22	1:D:183:ARG:HE	1.54	0.71

There are no symmetry-related clashes.

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	B	108/139 (78%)	106 (98%)	2 (2%)	0	100	100
1	D	108/139 (78%)	105 (97%)	3 (3%)	0	100	100
2	A	118/133 (89%)	111 (94%)	7 (6%)	0	100	100
2	C	118/133 (89%)	111 (94%)	7 (6%)	0	100	100
All	All	452/544 (83%)	433 (96%)	19 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	B	106/130 (82%)	103 (97%)	3 (3%)	43	56
1	D	106/130 (82%)	104 (98%)	2 (2%)	57	69
2	A	103/113 (91%)	100 (97%)	3 (3%)	42	54
2	C	103/113 (91%)	101 (98%)	2 (2%)	57	69
All	All	418/486 (86%)	408 (98%)	10 (2%)	49	61

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	183	ARG
2	C	54	ASN
2	C	127	ASP
2	A	115	SER
2	A	130	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	188	GLN
2	A	131	GLN
2	C	53	GLN
2	C	54	ASN

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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 5 ligands modelled in this entry, 2 are monoatomic - leaving 3 for Mogul analysis.

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	D	201	1	14,14,15	0.31	0	17,19,21	0.70	1 (5%)
3	NAG	B	202	1	14,14,15	0.52	0	17,19,21	1.07	1 (5%)
3	NAG	B	201	1	14,14,15	0.42	0	17,19,21	0.52	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
3	NAG	D	201	1	-	0/6/23/26	0/1/1/1
3	NAG	B	202	1	-	2/6/23/26	0/1/1/1
3	NAG	B	201	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	202	NAG	C1-O5-C5	3.57	117.02	112.19
3	D	201	NAG	C1-O5-C5	2.56	115.67	112.19

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	202	NAG	C1-C2-N2-C7
3	B	202	NAG	C3-C2-N2-C7

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	202	NAG	1	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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	B	110/139 (79%)	1.15	19 (17%)	1	1	60, 91, 136, 179	0
1	D	110/139 (79%)	0.57	6 (5%)	25	21	39, 72, 121, 146	0
2	A	120/133 (90%)	0.67	7 (5%)	23	19	39, 60, 100, 122	0
2	C	120/133 (90%)	0.48	3 (2%)	57	53	41, 58, 85, 104	0
All	All	460/544 (84%)	0.71	35 (7%)	13	11	39, 69, 120, 179	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	192	ASP	11.2
1	B	191	ASP	6.8
1	D	191	ASP	5.4
1	B	94	HIS	5.3
1	D	192	ASP	5.1

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 ⓘ
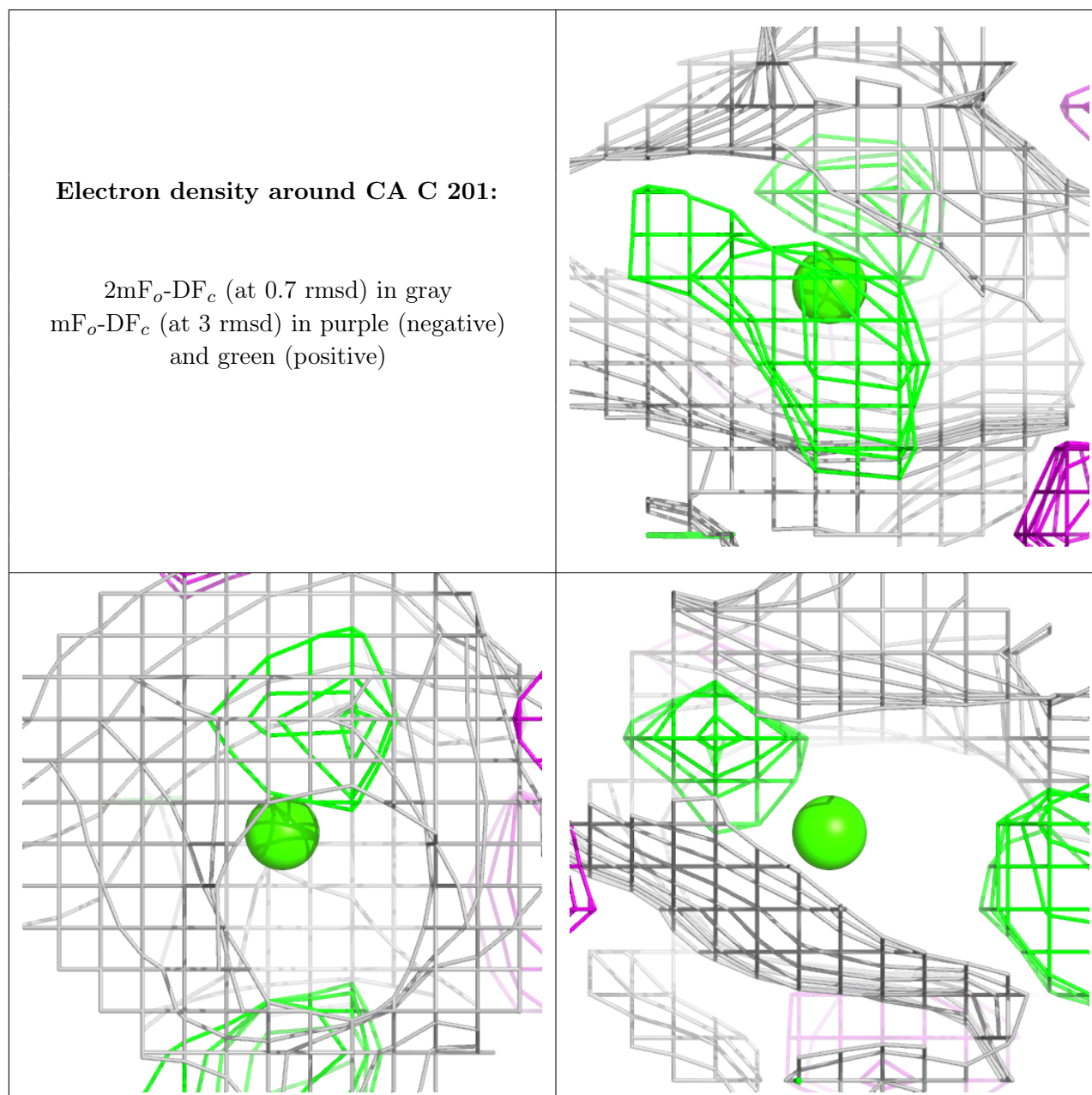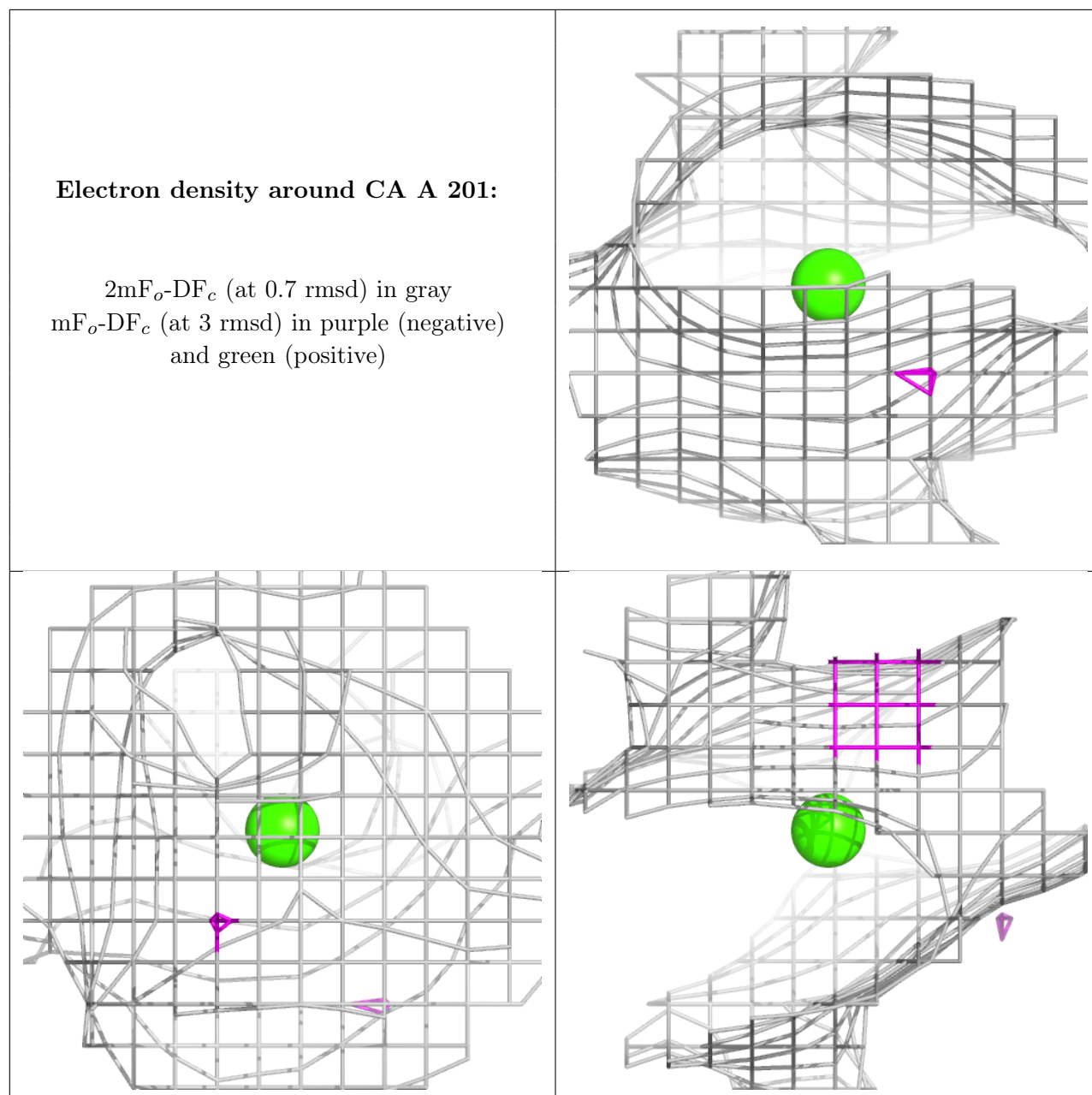
There are no monosaccharides in this entry.

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
3	NAG	B	202	14/15	0.68	0.39	97,104,111,112	0
3	NAG	B	201	14/15	0.83	0.31	73,86,91,92	0
3	NAG	D	201	14/15	0.89	0.22	59,65,82,85	0
4	CA	C	201	1/1	0.89	0.19	51,51,51,51	0
4	CA	A	201	1/1	0.95	0.17	47,47,47,47	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around CA C 201:

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

Electron density around CA A 201:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.