Full wwPDB X-ray Structure Validation Report (i)

May 16, 2020 – 12:46 pm BST

PDB ID	:	5NR8
Title	:	Crystal structure of human chitotriosidase-1 (hCHIT) catalytic domain in complex with compound 7a
Authors	:	Podjarny, A.; Fadel, F.; Golebiowski, A.
Deposited on	:	2017-04-22
Resolution	:	1.35 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

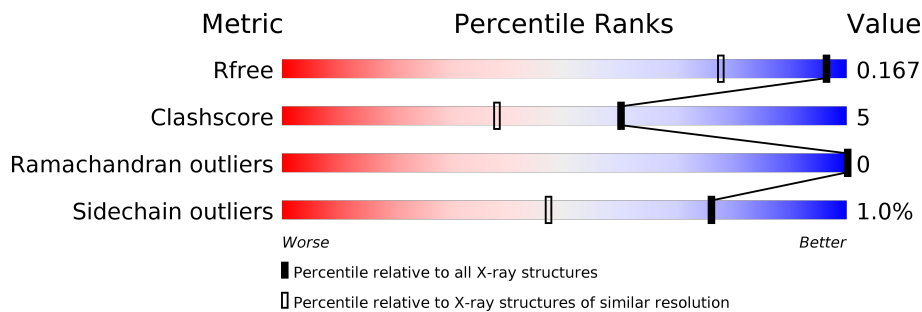|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1　Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
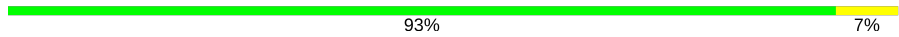*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.35 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		0.167
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		1.0%

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	1385 (1.36-1.32)
Clashscore	141614	1417 (1.36-1.32)
Ramachandran outliers	138981	1397 (1.36-1.32)
Sidechain outliers	138945	1397 (1.36-1.32)

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 on 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%

Mol	Chain	Length	Quality of chain
1	A	377	93%　　　　　　　　　　　　　　　　7%

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	GOL	A	402	-	-	X	-
3	GOL	A	404	-	-	X	-
3	GOL	A	406	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3440 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 Chitotriosidase-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	377	Total 3146	C 2002	N 539	O 592	S 13	0	20	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	388	VAL	-	expression tag	UNP Q13231
A	389	PRO	-	expression tag	UNP Q13231
A	390	ARG	-	expression tag	UNP Q13231
A	391	GLY	-	expression tag	UNP Q13231
A	392	SER	-	expression tag	UNP Q13231
A	393	HIS	-	expression tag	UNP Q13231
A	394	HIS	-	expression tag	UNP Q13231
A	395	HIS	-	expression tag	UNP Q13231
A	396	HIS	-	expression tag	UNP Q13231
A	397	HIS	-	expression tag	UNP Q13231
A	398	HIS	-	expression tag	UNP Q13231

- Molecule 2 is 1-(3-azanyl-1 {H}-1,2,4-triazol-5-yl)- {N}-[2-(4-bromophenyl)ethyl]- {N}-methyl-piperidin-4-amine (three-letter code: 95N) (formula: $C_{16}H_{23}BrN_6$).

95N
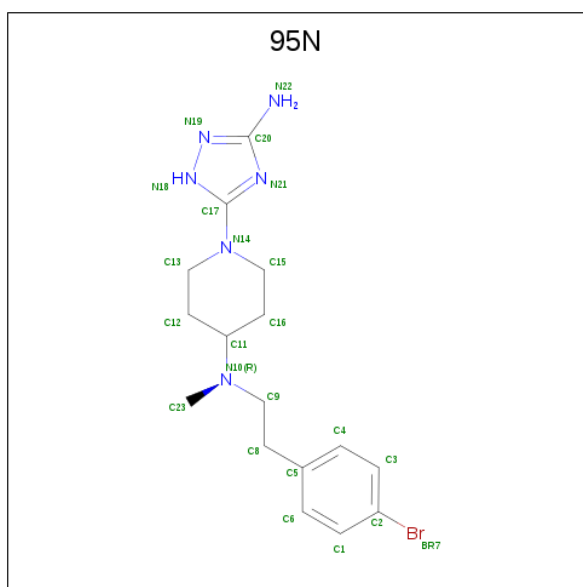
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	Br	C	N	0	0
			23	1	16	6		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 12	C 6	O 6	0	1
3	A	1	Total 6	C 3	O 3	0	0
3	A	1	Total 6	C 3	O 3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	222	Total 229	O 229	0	7

3 Residue-property plots (i)

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: Chitotriosidase-1

Chain A:

93%   7%

A22 W31 A32 Q33 Y34 R35 P43 K44 D45 L46 S49 L50 C51 N72 L83 W218 E219 K230 W309 K314 I325 F326 R327 Q349 L360 A366 G367 F368 T380 Q383 P389 R390 G391 H396 H397 H398
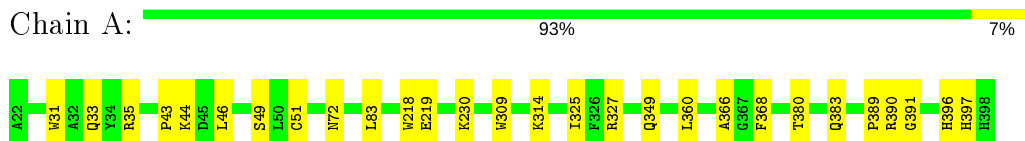
4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	85.90Å 106.01Å 42.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 1.35 30.09 – 1.35	Depositor EDS
% Data completeness (in resolution range)	97.4 (30.00-1.35) 97.4 (30.09-1.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.93 (at 1.35Å)	Xtriage
Refinement program	REFMAC, PHENIX	Depositor
R, R_{free}	(Not available) , (Not available) 0.143 , 0.167	Depositor DCC
R_{free} test set	4172 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	14.2	Xtriage
Anisotropy	0.026	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 44.8	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3440	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.32% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, 95N

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	0.41	0/3229	0.63	0/4378

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 (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	3146	0	3014	29	0
2	A	23	0	0	0	0
3	A	42	0	56	14	0
4	A	229	0	0	3	0
All	All	3440	0	3070	29	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 5.

All (29) 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:391:GLY:HA2	3:A:402:GOL:H11	1.54	0.89
1:A:391:GLY:HA2	3:A:402:GOL:C1	2.17	0.73
1:A:391:GLY:HA2	3:A:402:GOL:H31	1.74	0.70
1:A:35:ARG:NH2	3:A:406:GOL:H32	2.11	0.66
1:A:72:ASN:ND2	4:A:501:HOH:O	2.32	0.62
1:A:35:ARG:HH21	3:A:406:GOL:H32	1.66	0.59
1:A:46:LEU:HD21	1:A:360:LEU:HD13	1.85	0.58
1:A:31:TRP:CD1	3:A:406:GOL:H11	2.40	0.57
1:A:391:GLY:CA	3:A:402:GOL:H11	2.33	0.56
1:A:314:LYS:HA	1:A:325[B]:ILE:HG12	1.87	0.56
1:A:46:LEU:CD2	1:A:360:LEU:HD13	2.37	0.54
1:A:33:GLN:OE1	4:A:501:HOH:O	2.19	0.50
1:A:383[A]:GLN:OE1	1:A:397:HIS:CE1	2.65	0.50
1:A:43[B]:PRO:HA	1:A:46:LEU:HD12	1.94	0.49
1:A:383[A]:GLN:OE1	1:A:397:HIS:ND1	2.46	0.49
1:A:349:GLN:NE2	4:A:506:HOH:O	2.48	0.47
1:A:309[A]:TRP:HE1	1:A:327:ARG:NH2	2.13	0.46
1:A:46:LEU:HD13	1:A:83:LEU:HD11	1.98	0.46
1:A:43[B]:PRO:O	1:A:46:LEU:HB2	2.15	0.46
1:A:380:THR:O	1:A:383[B]:GLN:HG2	2.17	0.44
1:A:391:GLY:HA2	3:A:402:GOL:C3	2.45	0.43
1:A:51:CYS:O	3:A:404:GOL:H11	2.19	0.42
1:A:366:ALA:HB3	1:A:368[A]:PHE:HD2	1.84	0.42
1:A:49[B]:SER:HA	3:A:404:GOL:H12	2.02	0.41
1:A:218:TRP:CD1	1:A:219:GLU:HG3	2.56	0.41
1:A:35:ARG:HH21	3:A:406:GOL:C1	2.34	0.41
1:A:389:PRO:HG3	3:A:404:GOL:O2	2.21	0.41
1:A:390:ARG:O	3:A:402:GOL:O3	2.22	0.41
1:A:49[A]:SER:HA	3:A:404:GOL:H12	2.03	0.41

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	395/377 (105%)	391 (99%)	4 (1%)	0	100 100

There are no Ramachandran outliers to report.

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	332/312 (106%)	328 (99%)	4 (1%)	71 40

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	44[A]	LYS
1	A	44[B]	LYS
1	A	230	LYS
1	A	396	HIS

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)

8 ligands 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	GOL	A	404	-	5,5,5	0.62	0	5,5,5	0.68	0
3	GOL	A	406	-	5,5,5	0.62	0	5,5,5	0.76	0
3	GOL	A	407	-	5,5,5	0.50	0	5,5,5	1.10	0
3	GOL	A	405[B]	-	5,5,5	0.36	0	5,5,5	0.28	0
3	GOL	A	403	-	5,5,5	0.42	0	5,5,5	0.40	0
3	GOL	A	405[A]	-	5,5,5	0.37	0	5,5,5	0.15	0
3	GOL	A	402	-	5,5,5	0.41	0	5,5,5	0.42	0
2	95N	A	401	-	21,25,25	0.84	0	25,34,34	1.70	2 (8%)

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	GOL	A	404	-	-	4/4/4/4	-
3	GOL	A	406	-	-	4/4/4/4	-
3	GOL	A	407	-	-	2/4/4/4	-
3	GOL	A	405[B]	-	-	3/4/4/4	-
3	GOL	A	403	-	-	4/4/4/4	-
3	GOL	A	405[A]	-	-	2/4/4/4	-
3	GOL	A	402	-	-	4/4/4/4	-
2	95N	A	401	-	-	1/9/23/23	0/3/3/3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	401	95N	C15-N14-C17	-5.64	112.34	121.69

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	401	95N	C13-N14-C17	-5.08	113.27	121.69

There are no chirality outliers.

All (24) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	406	GOL	O1-C1-C2-C3
3	A	407	GOL	C1-C2-C3-O3
3	A	405[B]	GOL	C1-C2-C3-O3
3	A	403	GOL	C1-C2-C3-O3
3	A	405[A]	GOL	O1-C1-C2-C3
3	A	402	GOL	O1-C1-C2-C3
3	A	402	GOL	C1-C2-C3-O3
3	A	402	GOL	O2-C2-C3-O3
3	A	404	GOL	O1-C1-C2-O2
3	A	405[A]	GOL	O1-C1-C2-O2
3	A	404	GOL	O1-C1-C2-C3
3	A	406	GOL	C1-C2-C3-O3
3	A	403	GOL	O1-C1-C2-C3
3	A	406	GOL	O1-C1-C2-O2
3	A	406	GOL	O2-C2-C3-O3
3	A	407	GOL	O2-C2-C3-O3
3	A	405[B]	GOL	O2-C2-C3-O3
3	A	403	GOL	O2-C2-C3-O3
3	A	402	GOL	O1-C1-C2-O2
3	A	404	GOL	O2-C2-C3-O3
3	A	405[B]	GOL	O1-C1-C2-O2
3	A	403	GOL	O1-C1-C2-O2
2	A	401	95N	C12-C11-N10-C9
3	A	404	GOL	C1-C2-C3-O3

There are no ring outliers.

3 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	404	GOL	4	0
3	A	406	GOL	4	0
3	A	402	GOL	6	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In

addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand 95N A 401
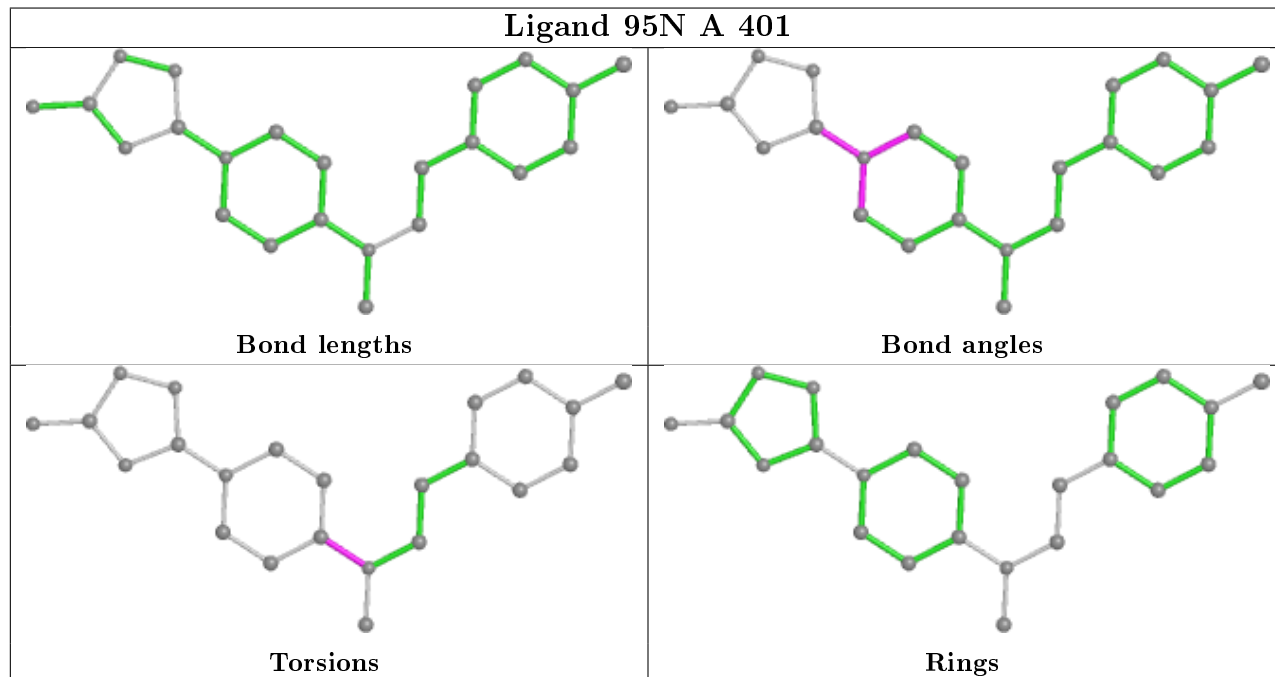
5.7 Other polymers (i)

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 (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

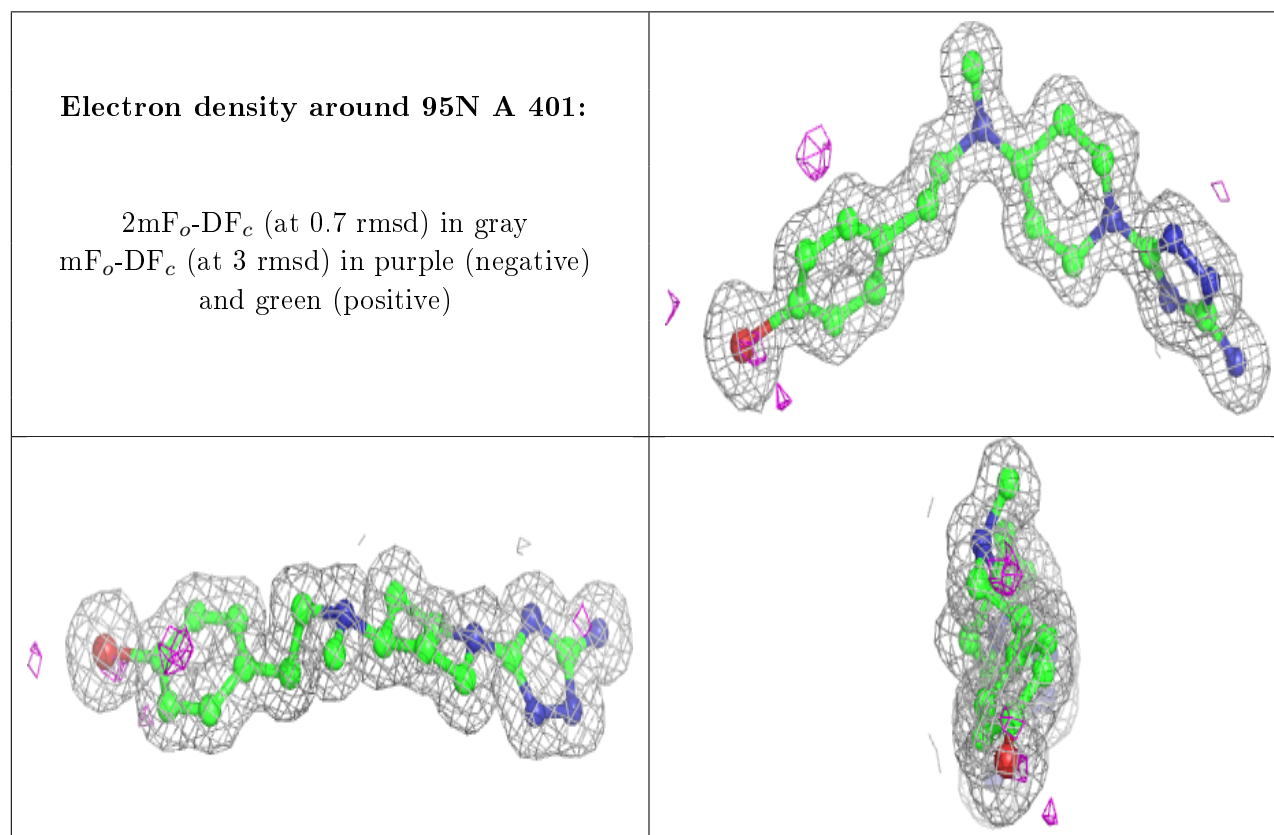Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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 95N A 401:

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

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.