

wwPDB X-ray Structure Validation Summary Report (i)

Sep 26, 2022 – 04:56 pm BST

PDB ID : 7PRI

Title: Carbonic Anhydrase from Schistosoma Mansoni in complex with clorsulon

Authors : Angeli, A.; Ferraroni, M.

Deposited on : 2021-09-21

Resolution : 1.68 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
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

MolProbity: 4.02b-467

Mogul : 1.8.4, CSD as541be (2020)

Xtriage (Phenix) : 1.13

EDS : 2.31.2buster-report : 1.1.7 (2018)

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

 $Refmac \quad : \quad 5.8.0267$

CCP4 : 7.1.010 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

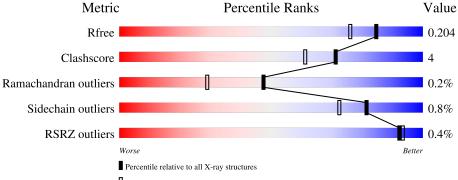
Validation Pipeline (wwPDB-VP) : 2.31.2

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 1.68 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.



Percentile relative to X-ray structures of similar resolution

Metric	Whole archive	Similar resolution
Metric	$(\# { m Entries})$	$(\# ext{Entries}, ext{ resolution range}(\mathring{A}))$
R_{free}	130704	6780 (1.70-1.66)
Clashscore	141614	7310 (1.70-1.66)
Ramachandran outliers	138981	7173 (1.70-1.66)
Sidechain outliers	138945	7172 (1.70-1.66)
RSRZ outliers	127900	6661 (1.70-1.66)

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	AAA	280	91%	7%	•
1	BBB	280	91%	8%	



2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4998 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 Carbonic anhydrase.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	AAA	277	Total 2259	C 1441	11	O 425	S 7	0	2	0
1	BBB	277	Total 2247	C 1433		O 422	S 7	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

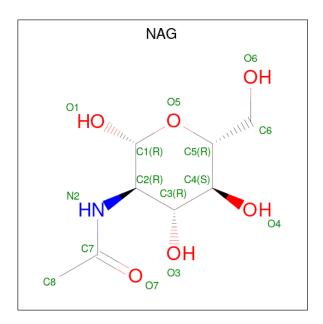
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	299	ARG	-	expression tag	UNP A0A3Q0KSG2
AAA	300	GLY	-	expression tag	UNP A0A3Q0KSG2
BBB	299	ARG	-	expression tag	UNP A0A3Q0KSG2
BBB	300	GLY	-	expression tag	UNP A0A3Q0KSG2

• Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	1	Total Zn 1 1	0	0
2	BBB	1	Total Zn 1 1	0	0

• Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$) (labeled as "Ligand of Interest" by depositor).

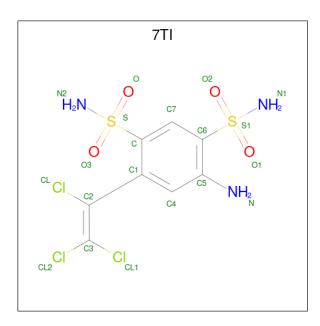




Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total C N O	0	0
	717171	1	14 8 1 5	0	Ü
3	AAA	1	Total C N O	O	0
3	AAA	1	14 8 1 5	0	0
3	BBB	1	Total C N O	0	0
3	ррр	1	14 8 1 5	0	0
3	BBB	1	Total C N O	0	0
3	ррр	1	14 8 1 5	0	0
3	BBB	1	Total C N O	0	0
3	DDD	1	14 8 1 5	U	U

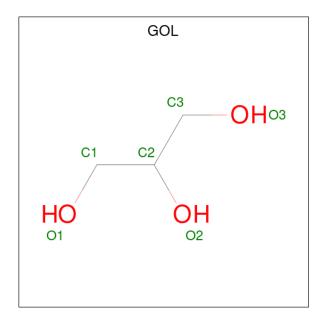
• Molecule 4 is 4-azanyl-6-[1,2,2-tris(chloranyl)ethenyl]benzene-1,3-disulfonamide (three-letter code: 7TI) (formula: $C_8H_8Cl_3N_3O_4S_2$) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf			
4	A A A	1	Total	С	Cl	N	О	S	0	0	
4	4 AAA	1	20	8	3	3	4	2	0	U	
4	DDD	1	Total	С	Cl	N	О	S	0	0	
4	4 BBB	BB 1		8	3	3	4	2		U	

• Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

• Molecule 6 is water.



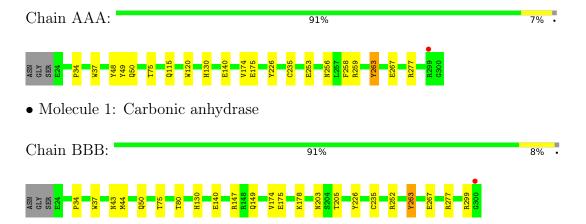
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AAA	190	Total O 190 190	0	0
6	BBB	184	Total O 184 184	0	0



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: Carbonic anhydrase





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants	103.80Å 103.80Å 132.90Å	Denogitor
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	89.89 - 1.68	Depositor
Resolution (A)	89.89 - 1.68	EDS
% Data completeness	99.7 (89.89-1.68)	Depositor
(in resolution range)	99.7 (89.89-1.68)	EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.28 (at 1.68Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.167 , 0.195	Depositor
it, it free	0.178 , 0.204	DCC
R_{free} test set	4543 reflections $(4.81%)$	wwPDB-VP
Wilson B-factor (Å ²)	20.4	Xtriage
Anisotropy	0.014	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	(Not available), (Not available)	EDS
L-test for twinning ²	$< L > = 0.50, < L^2> = 0.33$	Xtriage
Estimated twinning fraction	0.020 for -h,-k,l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4998	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.50% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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: 7TI, ZN, NAG, GOL

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		
IVIOI	Chain	77 1		RMSZ	# Z > 5	
1	AAA	0.76	0/2325	0.92	1/3163 (0.0%)	
1	BBB	0.76	0/2307	0.89	1/3139 (0.0%)	
All	All	0.76	0/4632	0.90	2/6302 (0.0%)	

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
1	AAA	75	THR	CA-CB-OG1	-6.88	94.55	109.00
1	BBB	75	THR	CA-CB-OG1	-6.04	96.32	109.00

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	AAA	2259	0	2194	14	0
1	BBB	2247	0	2180	16	0
2	AAA	1	0	0	0	0
2	BBB	1	0	0	0	0
3	AAA	28	0	26	0	0
3	BBB	42	0	39	3	0
4	AAA	20	0	0	1	0

Continued on next page...



Continued	trom	mmoninonic	maaa
COHABABACA		DIEUIUU	DUIUE
0 0 1000100000			

Mol	Chain	Non-H	H(model)	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
4	BBB	20	0	0	0	0
5	BBB	6	0	8	0	0
6	AAA	190	0	0	1	0
6	BBB	184	0	0	2	0
All	All	4998	0	4447	31	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 4.

The worst 5 of 31 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	$\begin{array}{c} \text{Clash} \\ \text{overlap } (\text{\AA}) \end{array}$
1:BBB:203:ASN:HD21	3:BBB:405:NAG:C1	1.38	1.35
1:AAA:50:GLN:HE22	1:AAA:235:CYS:HB3	1.49	0.77
1:BBB:130:HIS:HE1	1:BBB:226:TYR:OH	1.72	0.71
1:AAA:130:HIS:HE1	1:AAA:226:TYR:OH	1.75	0.70
1:BBB:50:GLN:HE22	1:BBB:235:CYS:HB3	1.56	0.69

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	Perce	entiles
1	AAA	277/280 (99%)	267 (96%)	10 (4%)	0	100	100
1	BBB	275/280 (98%)	269 (98%)	5 (2%)	1 (0%)	34	17
All	All	552/560 (99%)	536 (97%)	15 (3%)	1 (0%)	47	29

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	BBB	299	ARG

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	AAA	253/253 (100%)	251 (99%)	2 (1%)	81 72
1	BBB	251/253 (99%)	249 (99%)	2 (1%)	81 72
All	All	504/506 (100%)	500 (99%)	4 (1%)	81 72

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

Mol	Chain	Res	Type
1	AAA	115	GLN
1	AAA	263	TYR
1	BBB	178	LYS
1	BBB	263	TYR

Sometimes 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 monosaccharides in this entry.



5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 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	Во	ond leng	ths	В	ond ang	les
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	7TI	BBB	406	2	20,20,20	1.11	2 (10%)	27,32,32	1.07	2 (7%)
3	NAG	AAA	403	1	14,14,15	1.11	1 (7%)	17,19,21	3.27	9 (52%)
3	NAG	AAA	402	1	14,14,15	1.12	1 (7%)	17,19,21	1.35	3 (17%)
5	GOL	BBB	401	-	5,5,5	0.53	0	5,5,5	1.29	1 (20%)
4	7TI	AAA	404	2	20,20,20	1.03	2 (10%)	27,32,32	0.83	2 (7%)
3	NAG	BBB	405	1	14,14,15	0.92	0	17,19,21	1.81	4 (23%)
3	NAG	BBB	403	1	14,14,15	1.08	1 (7%)	17,19,21	1.90	5 (29%)
3	NAG	BBB	404	1	14,14,15	0.72	0	17,19,21	1.09	1 (5%)

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
4	7TI	BBB	406	2	-	3/20/20/20	0/1/1/1
3	NAG	AAA	403	1	-	2/6/23/26	0/1/1/1
3	NAG	AAA	402	1	-	0/6/23/26	0/1/1/1
5	GOL	BBB	401	-	-	0/4/4/4	-
4	7TI	AAA	404	2	-	2/20/20/20	0/1/1/1
3	NAG	BBB	405	1	-	0/6/23/26	0/1/1/1
3	NAG	BBB	403	1	-	2/6/23/26	0/1/1/1
3	NAG	BBB	404	1	-	0/6/23/26	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\operatorname{\AA})$	$[Ideal(\AA)]$
4	BBB	406	7TI	C-S	2.80	1.81	1.77

Continued on next page...



$\alpha \cdots$, r	•	
Continued	trom	mromonie	maaa
-	110116	DICULUUS	Duuc
	J	1	1

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	Observed(A)	$[Ideal(\AA)]$
4	AAA	404	7TI	C-S	2.78	1.81	1.77
4	AAA	404	7TI	C2-CL	2.53	1.79	1.74
3	AAA	402	NAG	O5-C5	2.24	1.48	1.43
3	AAA	403	NAG	C6-C5	2.16	1.59	1.51

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$Ideal(^{o})$
3	AAA	403	NAG	C1-O5-C5	10.59	126.53	112.19
3	BBB	405	NAG	C1-O5-C5	4.67	118.52	112.19
3	BBB	403	NAG	O7-C7-C8	-4.00	114.63	122.06
3	BBB	403	NAG	C8-C7-N2	3.97	122.82	116.10
3	BBB	403	NAG	C2-N2-C7	3.80	128.31	122.90

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	AAA	403	NAG	C4-C5-C6-O6
3	BBB	403	NAG	C8-C7-N2-C2
3	BBB	403	NAG	O7-C7-N2-C2
3	AAA	403	NAG	O5-C5-C6-O6
4	BBB	406	7TI	C5-C6-S1-O1

There are no ring outliers.

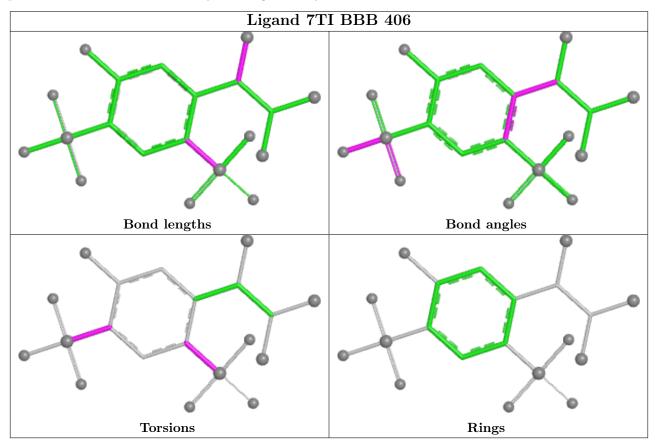
3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	AAA	404	7TI	1	0
3	BBB	405	NAG	2	0
3	BBB	403	NAG	1	0

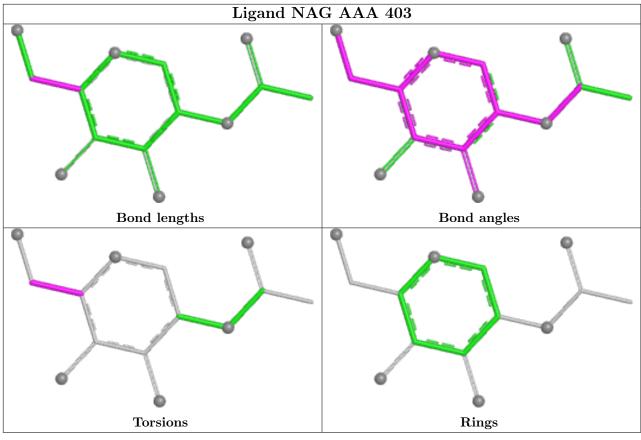
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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.

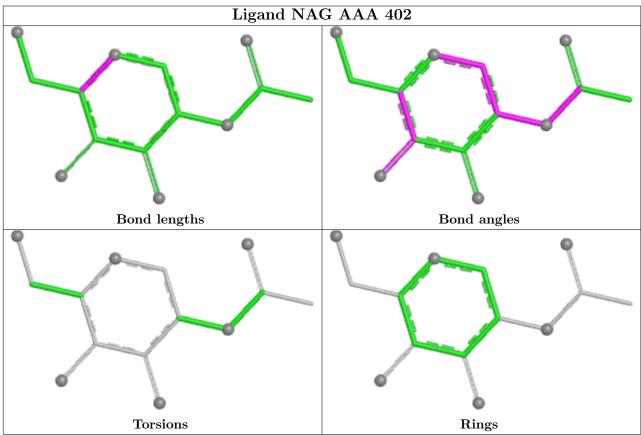


The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

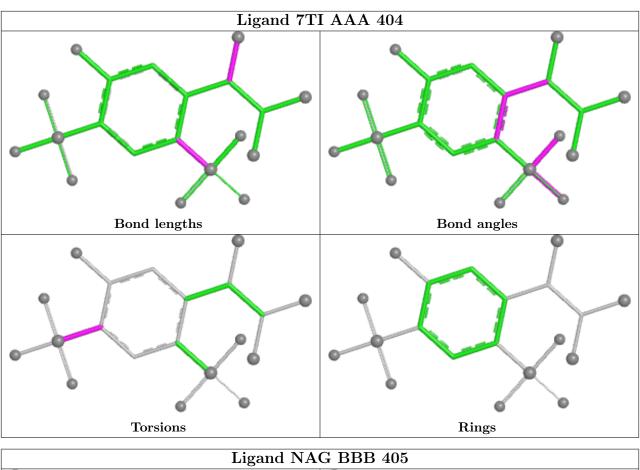


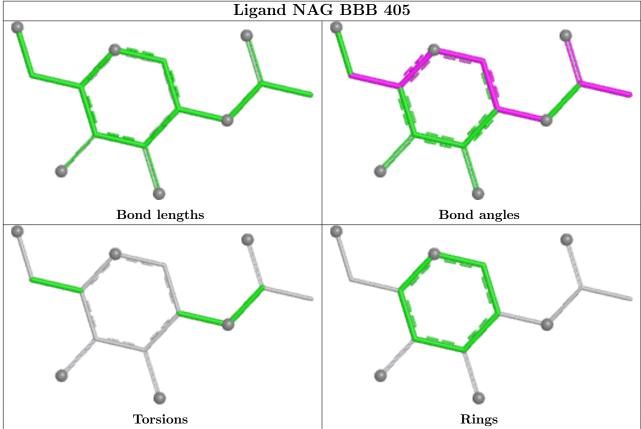




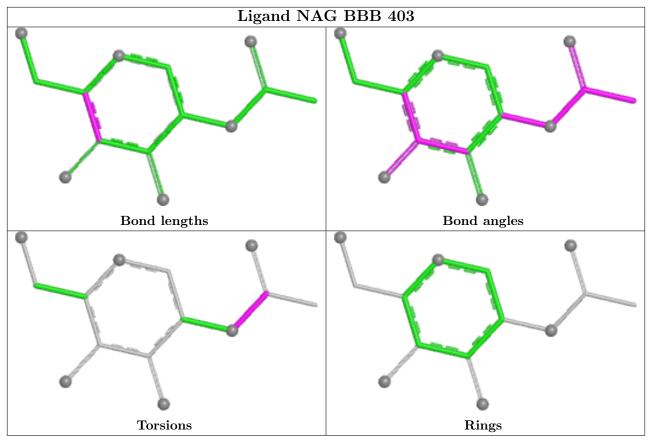


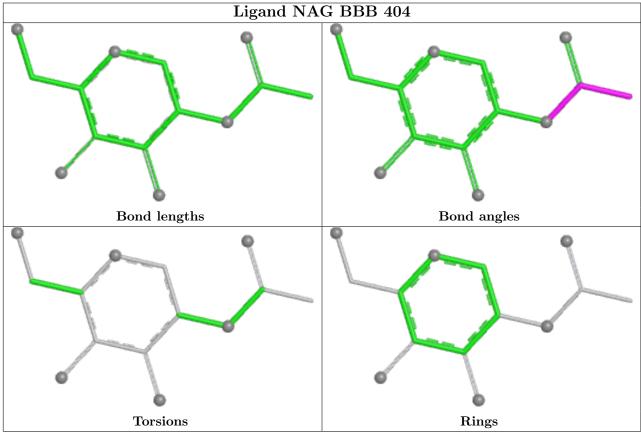














5.7 Other polymers (i)

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)

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^{th} 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		$OWAB(A^2)$	Q < 0.9	
1	AAA	277/280 (98%)	-0.22	1 (0%)	92	93	13, 20, 38, 61	0
1	BBB	277/280 (98%)	-0.25	1 (0%)	92	93	13, 22, 40, 65	0
All	All	554/560 (98%)	-0.23	2 (0%)	92	93	13, 21, 39, 65	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	300	GLY	4.1
1	AAA	299	ARG	3.4

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)

There are no monosaccharides in this entry.

6.4 Ligands (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^{th} 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{ ilde{A}}^2)$	Q<0.9
3	NAG	BBB	405	14/15	0.66	0.23	60,68,75,77	0
3	NAG	BBB	403	14/15	0.85	0.11	35,42,47,48	0
3	NAG	AAA	402	14/15	0.86	0.13	31,35,43,43	0

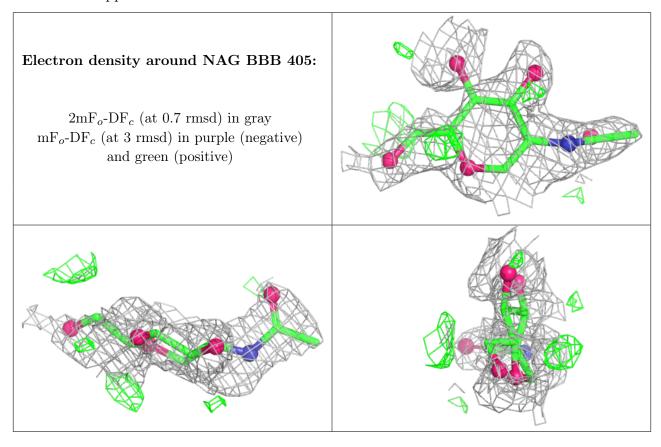
Continued on next page...



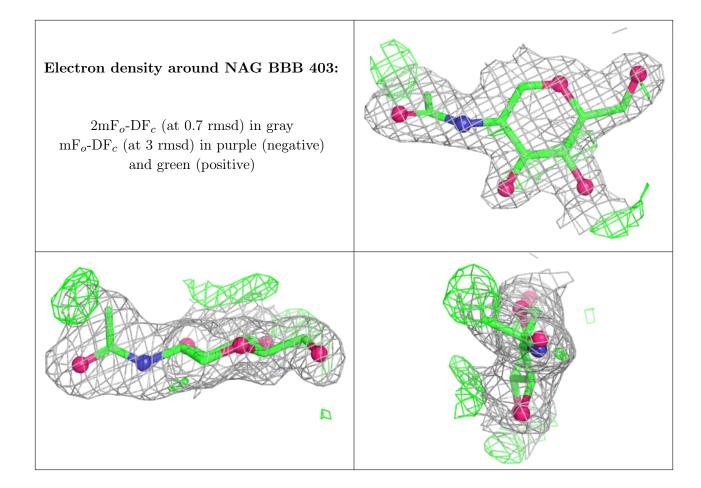
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
3	NAG	AAA	403	14/15	0.87	0.11	36,46,56,57	0
3	NAG	BBB	404	14/15	0.91	0.06	34,37,43,43	0
5	GOL	BBB	401	6/6	0.94	0.11	18,25,26,31	0
4	7TI	BBB	406	20/20	0.95	0.14	18,32,77,100	0
4	7TI	AAA	404	20/20	0.96	0.14	17,28,69,83	0
2	ZN	AAA	401	1/1	1.00	0.09	15,15,15,15	0
2	ZN	BBB	402	1/1	1.00	0.08	17,17,17,17	0

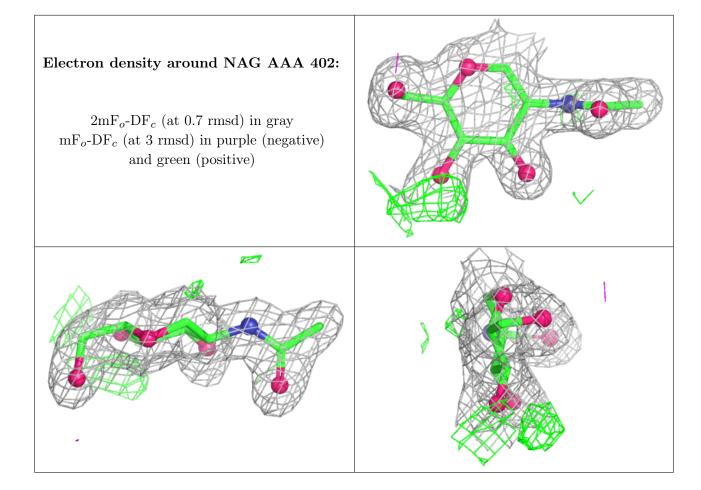
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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





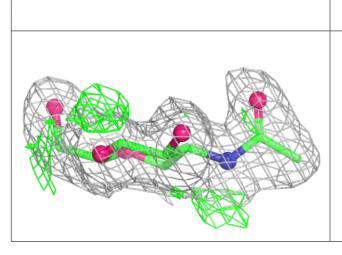


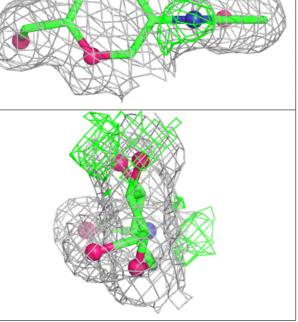




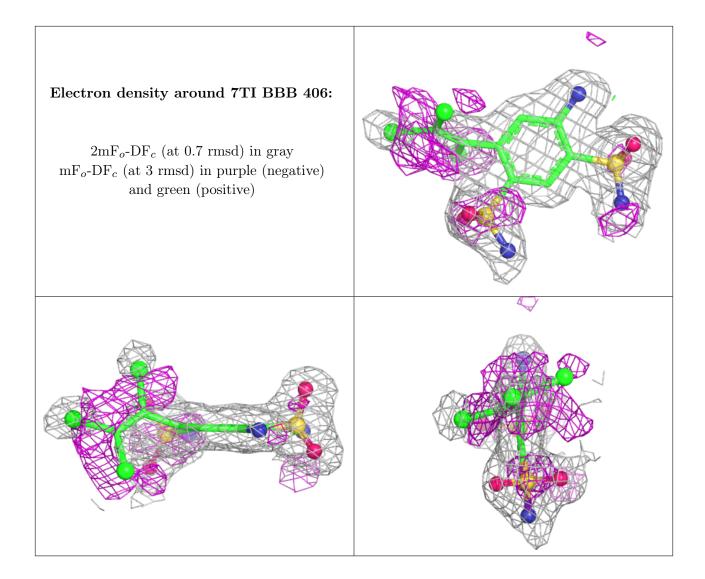


Electron density around NAG AAA 403: $2 \mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 0.7 rmsd) in gray ${ m mF}_o{ m -DF}_c$ (at 3 rmsd) in purple (negative) and green (positive) Electron density around NAG BBB 404: $2 \mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 0.7 rmsd) in gray mF_o -DF_c (at 3 rmsd) in purple (negative) and green (positive)

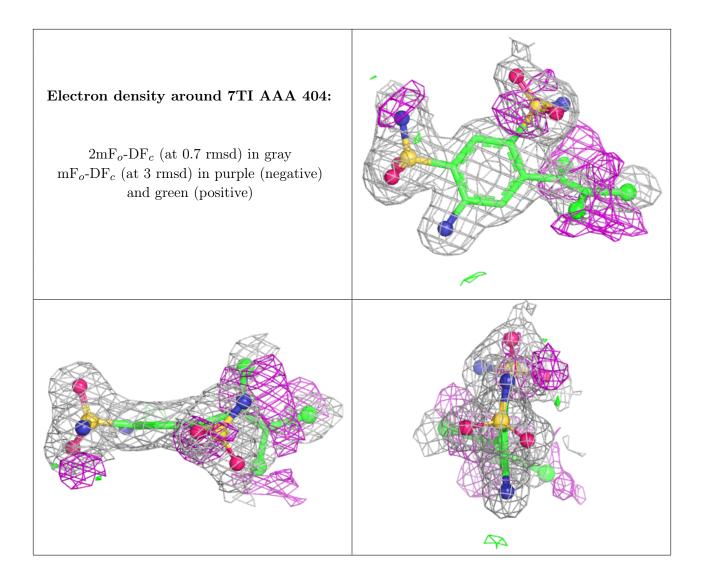












6.5 Other polymers (i)

There are no such residues in this entry.

