Full wwPDB X-ray Structure Validation Report

Jan 20, 2024 – 03:42 pm GMT

PDB ID	:	7A0D
Title	:	The Crystal Structure of Bovine Thrombin in complex with Hirudin (C16U/C28U) at 1.6 Angstroms Resolution
Authors	:	Hidmi, T.; Mousa, R.; Pomyalov, S.; Lansky, S.; Khouri, L.; Metanis, N.; Shoham, G.
Deposited on	:	2020-08-07
Resolution	:	1.60 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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:

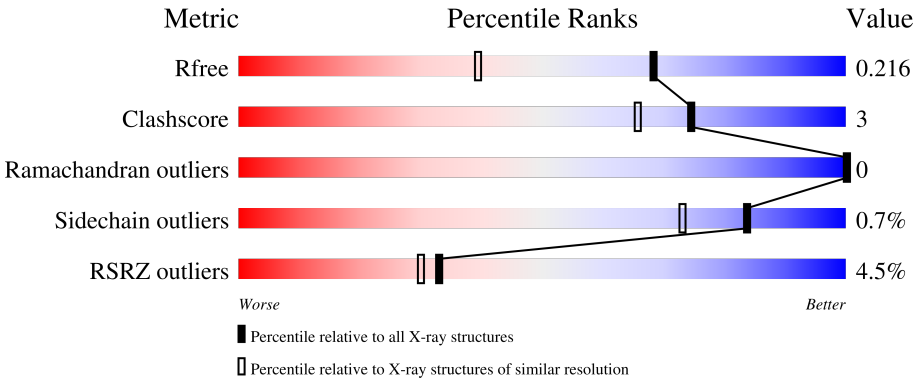
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.60 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))
R_{free}	130704	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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	HHH	259	2% 92% 7%
2	III	65	11% 91% 8%
3	LLL	49	8% 65% 6% 29%
4	HaH	2	100%

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
4	NAG	HaH	2	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3207 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 Prothrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	HHH	257	2076	1327	371	366	12	0	4	0

- Molecule 2 is a protein called Hirudin variant-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	III	64	462	275	78	103	4	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
III	16	SEC	CYS	modified residue	UNP P01050
III	28	SEC	CYS	modified residue	UNP P01050

- Molecule 3 is a protein called Prothrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	LLL	35	283	179	45	58	1	0	0	0

- Molecule 4 is an oligosaccharide called 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	HaH	2	28	16	2	10	0	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	HHH	1	Total Na 1 1	0	0

- Molecule 6 is water.

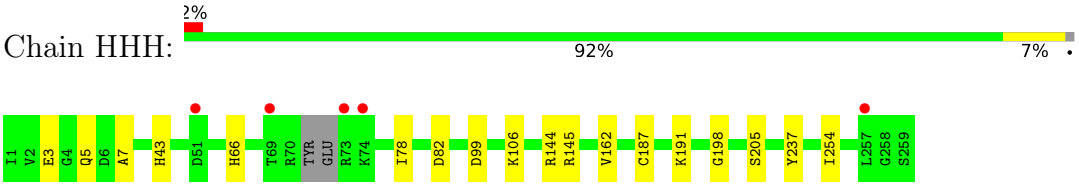
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	HHH	254	Total O 254 254	0	0
6	III	69	Total O 69 69	0	0
6	LLL	34	Total O 34 34	0	0

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.

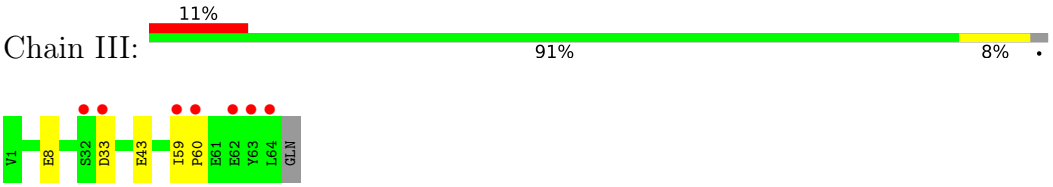
• Molecule 1: Prothrombin

Chain HHH:



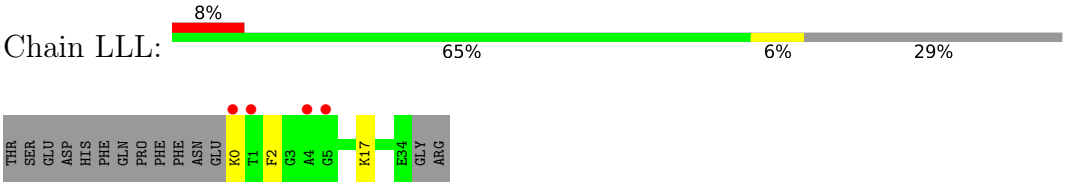
• Molecule 2: Hirudin variant-1

Chain III:



• Molecule 3: Prothrombin

Chain LLL:



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

Chain HaH:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	58.67Å 102.37Å 142.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.18 – 1.60 48.18 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (48.18-1.60) 99.9 (48.18-1.60)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.21 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.183 , 0.215 0.184 , 0.216	Depositor DCC
R_{free} test set	2797 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	25.5	Xtriage
Anisotropy	0.008	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 42.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.018 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.031 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3207	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.75% 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: NA, SEC, 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	HHH	0.50	0/2140	0.89	0/2896
2	III	0.44	0/454	0.93	2/611 (0.3%)
3	LLL	0.51	0/287	0.93	0/382
All	All	0.49	0/2881	0.90	2/3889 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	III	33	ASP	CB-CA-C	5.24	120.88	110.40
2	III	8	GLU	CB-CA-C	-5.12	100.16	110.40

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	HHH	2076	0	2081	15	0
2	III	462	0	400	4	0
3	LLL	283	0	273	3	0
4	HaH	28	0	25	0	0
5	HHH	1	0	0	0	0
6	HHH	254	0	0	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	III	69	0	0	1	0
6	LLL	34	0	0	1	0
All	All	3207	0	2779	18	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 3.

All (18) 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:HHH:144[A]:ARG:NH2	6:HHH:401:HOH:O	2.22	0.72
2:III:43:GLU:HG3	6:III:166:HOH:O	1.89	0.71
1:HHH:43:HIS:HD2	1:HHH:99:ASP:OD2	1.77	0.68
1:HHH:5[B]:GLN:HG2	1:HHH:162:VAL:HG13	1.80	0.63
1:HHH:254:ILE:HG21	3:LLL:2:PHE:HD1	1.66	0.60
2:III:59:ILE:HG23	2:III:60:PRO:HD2	1.87	0.57
1:HHH:7:ALA:O	1:HHH:66:HIS:HE1	1.95	0.49
1:HHH:78:ILE:HD13	2:III:59:ILE:HG23	1.93	0.49
1:HHH:5[A]:GLN:HB3	6:HHH:442:HOH:O	2.11	0.49
1:HHH:191:LYS:HG2	1:HHH:237:TYR:OH	2.12	0.49
1:HHH:5[B]:GLN:HG2	1:HHH:162:VAL:CG1	2.42	0.48
1:HHH:145:ARG:NH1	6:HHH:404:HOH:O	2.46	0.48
1:HHH:43:HIS:CE1	1:HHH:205:SER:HB3	2.49	0.47
1:HHH:82:ASP:HB2	1:HHH:106:LYS:HA	1.98	0.45
1:HHH:78:ILE:HD13	2:III:59:ILE:CG2	2.48	0.43
1:HHH:254:ILE:HD13	3:LLL:2:PHE:CD1	2.53	0.43
1:HHH:3:GLU:HB2	1:HHH:198:GLY:HA2	2.01	0.43
3:LLL:17:LYS:HE3	6:LLL:126:HOH:O	2.20	0.42

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	HHH	257/259 (99%)	251 (98%)	6 (2%)	0	100	100
2	III	60/65 (92%)	56 (93%)	4 (7%)	0	100	100
3	LLL	33/49 (67%)	32 (97%)	1 (3%)	0	100	100
All	All	350/373 (94%)	339 (97%)	11 (3%)	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	HHH	225/226 (100%)	224 (100%)	1 (0%)	91	84
2	III	50/54 (93%)	50 (100%)	0	100	100
3	LLL	30/43 (70%)	29 (97%)	1 (3%)	38	14
All	All	305/323 (94%)	303 (99%)	2 (1%)	84	73

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	HHH	187	CYS
3	LLL	0	LYS

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)

2 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	$\#\|Z\| > 2$
4	NAG	HaH	1	1,4	14,14,15	0.65	0	17,19,21	1.21	3 (17%)
4	NAG	HaH	2	4	14,14,15	0.83	1 (7%)	17,19,21	1.65	2 (11%)

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	NAG	HaH	1	1,4	-	0/6/23/26	0/1/1/1
4	NAG	HaH	2	4	-	4/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	HaH	2	NAG	C1-C2	2.03	1.55	1.52

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	HaH	2	NAG	C1-O5-C5	4.84	118.76	112.19
4	HaH	2	NAG	C2-N2-C7	3.74	128.23	122.90
4	HaH	1	NAG	C2-N2-C7	2.19	126.03	122.90
4	HaH	1	NAG	O4-C4-C5	2.07	114.44	109.30
4	HaH	1	NAG	C1-C2-N2	2.04	113.97	110.49

There are no chirality outliers.

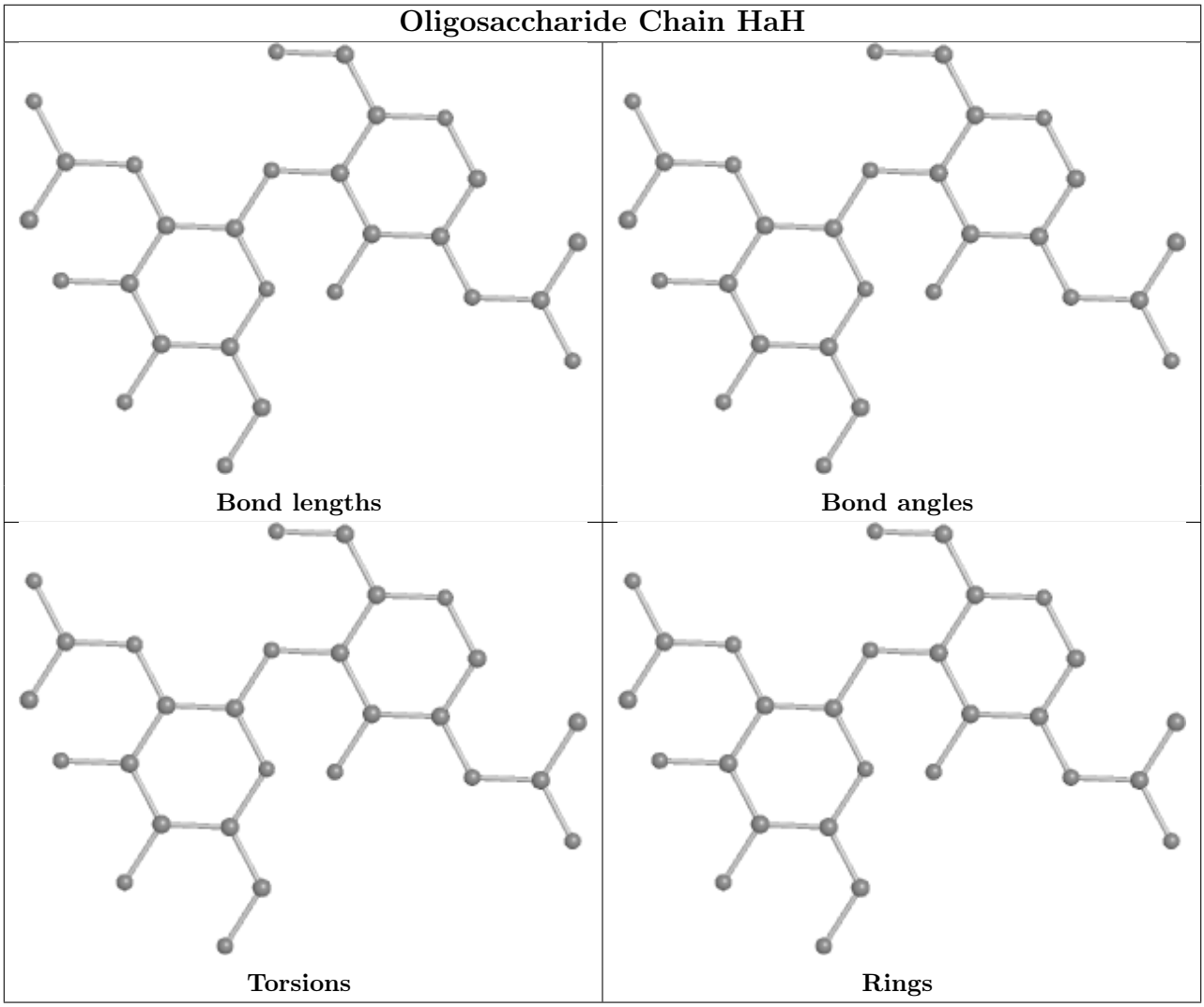
All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	HaH	2	NAG	C8-C7-N2-C2
4	HaH	2	NAG	O7-C7-N2-C2
4	HaH	2	NAG	C1-C2-N2-C7
4	HaH	2	NAG	C3-C2-N2-C7

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.



5.6 Ligand geometry

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

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, 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(Å^2)	Q<0.9
1	HHH	257/259 (99%)	-0.09	5 (1%)	66	65	16, 26, 49, 101	0
2	III	62/65 (95%)	0.36	7 (11%)	5	4	22, 32, 65, 89	0
3	LLL	35/49 (71%)	0.27	4 (11%)	5	4	21, 28, 53, 84	0
All	All	354/373 (94%)	0.02	16 (4%)	33	30	16, 27, 54, 101	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	III	64	LEU	10.9
1	HHH	69	THR	4.9
2	III	63	TYR	4.5
1	HHH	74	LYS	4.2
1	HHH	257	LEU	3.8
2	III	60	PRO	3.7
1	HHH	73	ARG	3.5
3	LLL	5	GLY	3.3
2	III	62	GLU	3.1
3	LLL	1	THR	3.0
2	III	33	ASP	3.0
2	III	59	ILE	3.0
3	LLL	0	LYS	2.8
2	III	32	SER	2.7
1	HHH	51	ASP	2.3
3	LLL	4	ALA	2.3

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

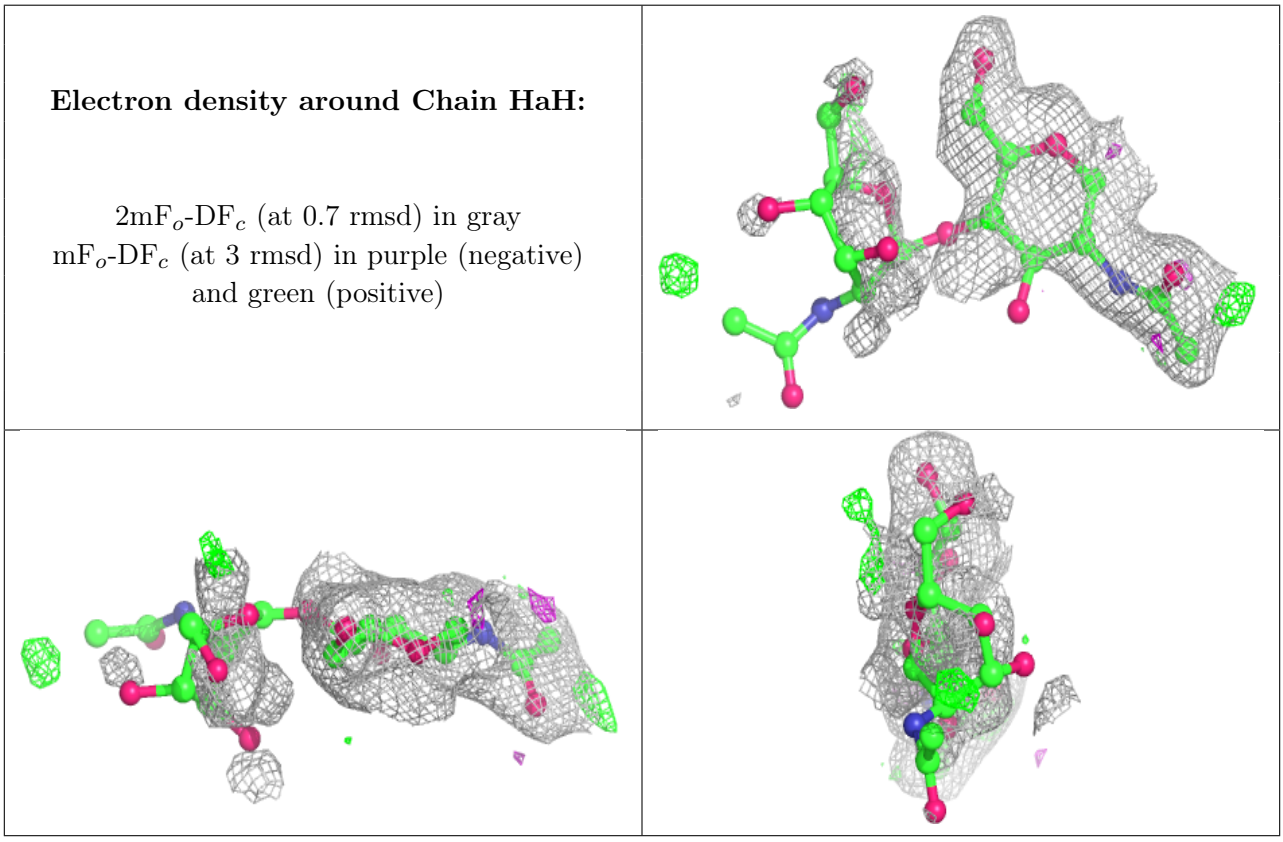
There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (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	B-factors(Å^2)	Q<0.9
4	NAG	HaH	2	14/15	0.62	0.41	102,124,142,143	0
4	NAG	HaH	1	14/15	0.74	0.14	48,70,84,100	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)

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	B-factors(Å^2)	Q<0.9
5	NA	HHH	301	1/1	0.99	0.07	24,24,24,24	0

6.5 Other polymers (i)

There are no such residues in this entry.