



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2021 – 04:11 PM GMT

PDB ID : 6YQA
Title : Taka-amylase in complex with alpha-glucosyl epi-cyclophellitol aziridine inhibitor
Authors : Armstrong, Z.; Chen, Y.; Artola, M.; Overkleeft, H.; Davies, G.
Deposited on : 2020-04-16
Resolution : 1.67 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 ⓘ symbol.

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 : 2.17
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) : 2.17

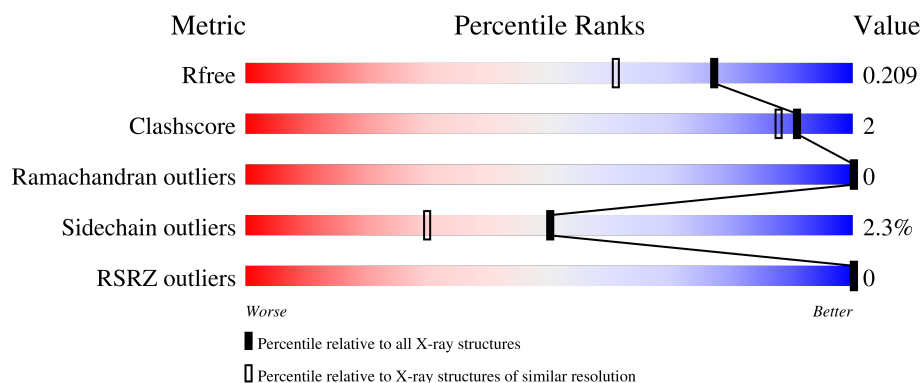
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION



The reported resolution of this entry is 1.67 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	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 $\leq 5\%$. 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	AAA	499	 90% 5%
1	BBB	499	 89% 7% 5%

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 15145 atoms, of which 7115 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 Alpha-amylase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	476	Total	C	H	N	O	S	215	3	0
			7193	2342	3493	599	741	18			
1	BBB	476	Total	C	H	N	O	S	215	3	0
			7200	2345	3496	598	743	18			

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



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	AAA	1	Total	C	H	N	O	3	0
			28	8	14	1	5		
2	BBB	1	Total	C	H	N	O	3	0
			28	8	14	1	5		

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
3	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
3	BBB	1	Total	C	H	O	1	0
			10	2	6	2		
3	BBB	1	Total	C	H	O	1	0
			10	2	6	2		
3	BBB	1	Total	C	H	O	1	0
			10	2	6	2		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

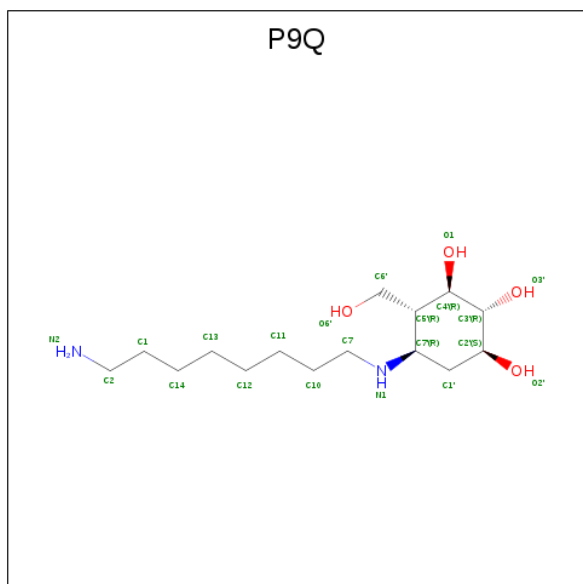
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AAA	1	Total	Ca	0	0
			1	1		
4	BBB	1	Total	Ca	0	0
			1	1		

- Molecule 5 is alpha-D-glucopyranose (three-letter code: GLC) (formula: C₆H₁₂O₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	AAA	1	Total	C	H	O	3	0
			22	6	11	5		
5	BBB	1	Total	C	H	O	3	0
			22	6	11	5		

- Molecule 6 is (1 {S},2 {R},3 {R},4 {R},5 {R})-5-(8-azanyloctylamino)-4-(hydroxymethyl)cyclohexane-1,2,3-triol (three-letter code: P9Q) (formula: C₁₅H₃₂N₂O₄).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	AAA	1	Total	C	H	N	O	4	0
			39	12	22	1	4		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	BBB	1	Total	C	H	N	O	4	0
			42	13	24	1	4		

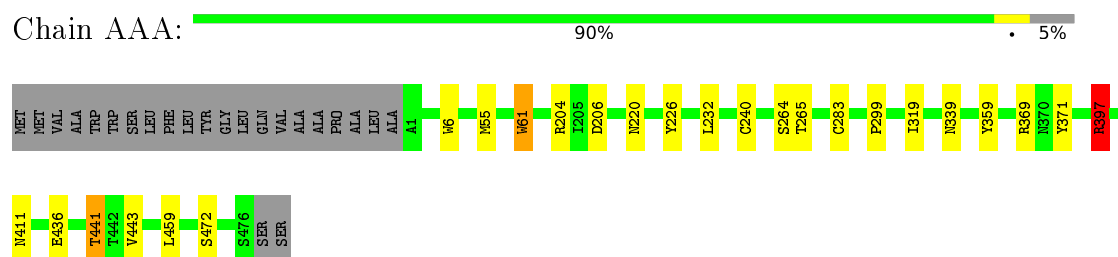
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	AAA	259	Total	O	0	0
			259	259		
7	BBB	260	Total	O	0	0
			260	260		

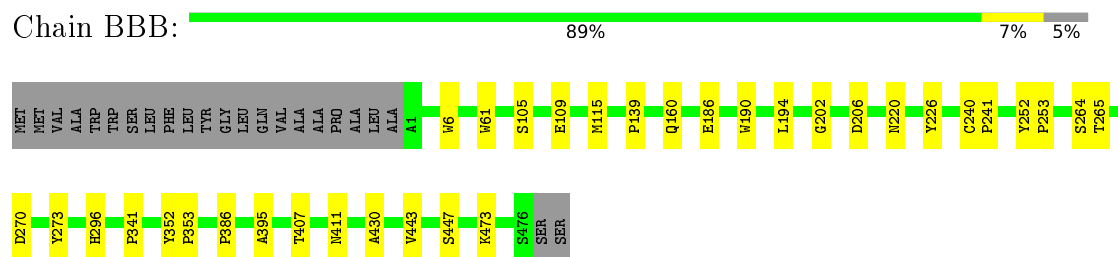
3 Residue-property plots [i](#)

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: Alpha-amylase



- Molecule 1: Alpha-amylase



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	65.68Å 103.32Å 75.51Å 90.00° 103.83° 90.00°	Depositor
Resolution (Å)	73.43 – 1.67 73.32 – 1.67	Depositor EDS
% Data completeness (in resolution range)	99.9 (73.43-1.67) 100.0 (73.32-1.67)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.86 (at 1.67Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.172 , 0.201 0.183 , 0.209	Depositor DCC
R_{free} test set	5597 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å ²)	22.0	Xtriage
Anisotropy	0.860	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.43 , 39.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	15145	wwPDB-VP
Average B, all atoms (Å ²)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.11% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: GLC, NAG, EDO, CA, P9Q

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	AAA	0.77	1/3800 (0.0%)	0.86	3/5190 (0.1%)
1	BBB	0.75	1/3804 (0.0%)	0.84	0/5195
All	All	0.76	2/7604 (0.0%)	0.85	3/10385 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	BBB	206	ASP	CG-OD1	6.85	1.41	1.25
1	AAA	206	ASP	CG-OD1	6.16	1.39	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	AAA	204	ARG	NE-CZ-NH2	-6.43	117.09	120.30
1	AAA	397[A]	ARG	NE-CZ-NH1	5.60	123.10	120.30
1	AAA	397[B]	ARG	NE-CZ-NH1	5.60	123.10	120.30

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	3700	3493	3484	10	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	BBB	3704	3496	3489	13	0
2	AAA	14	14	13	0	0
2	BBB	14	14	13	0	0
3	AAA	8	12	12	0	0
3	BBB	12	18	18	0	0
4	AAA	1	0	0	0	0
4	BBB	1	0	0	0	0
5	AAA	11	11	10	0	0
5	BBB	11	11	10	0	0
6	AAA	17	22	0	1	0
6	BBB	18	24	0	1	0
7	AAA	259	0	0	2	0
7	BBB	260	0	0	1	0
All	All	8030	7115	7049	25	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 2.

The worst 5 of 25 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:BBB:6:TRP:CD1	1:BBB:226:TYR:HB3	2.44	0.53
1:AAA:397[A]:ARG:HD3	1:AAA:397[A]:ARG:C	2.30	0.52
6:BBB:507:P9Q:C14	7:BBB:777:HOH:O	2.59	0.50
6:AAA:506:P9Q:N1	7:AAA:602:HOH:O	2.34	0.50
1:AAA:371:TYR:CE1	1:AAA:472:SER:HB3	2.47	0.49

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	AAA	477/499 (96%)	463 (97%)	14 (3%)	0	100	100
1	BBB	477/499 (96%)	462 (97%)	15 (3%)	0	100	100
All	All	954/998 (96%)	925 (97%)	29 (3%)	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	AAA	399/414 (96%)	389 (98%)	10 (2%)	47	26
1	BBB	400/414 (97%)	391 (98%)	9 (2%)	50	30
All	All	799/828 (96%)	780 (98%)	19 (2%)	50	28

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	270	ASP
1	BBB	443	VAL
1	BBB	473	LYS
1	BBB	341	PRO
1	AAA	443	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 13 ligands modelled in this entry, 2 are monoatomic - leaving 11 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	EDO	BBB	505	-	3,3,3	0.53	0	2,2,2	0.43	0
3	EDO	AAA	502	-	3,3,3	0.38	0	2,2,2	0.57	0
6	P9Q	AAA	506	5,1	17,17,21	1.47	3 (17%)	21,22,26	1.82	7 (33%)
2	NAG	AAA	501	1	14,14,15	0.83	0	17,19,21	2.08	6 (35%)
3	EDO	BBB	503	-	3,3,3	0.22	0	2,2,2	0.74	0
3	EDO	BBB	504	-	3,3,3	0.31	0	2,2,2	0.14	0
6	P9Q	BBB	507	5,1	18,18,21	1.12	2 (11%)	22,23,26	1.89	3 (13%)
2	NAG	BBB	502	1	14,14,15	0.75	0	17,19,21	1.33	3 (17%)
5	GLC	AAA	505	6	11,11,12	0.61	0	15,15,17	1.00	1 (6%)
3	EDO	AAA	503	-	3,3,3	0.18	0	2,2,2	0.15	0
5	GLC	BBB	506	6	11,11,12	0.51	0	15,15,17	1.51	2 (13%)

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	EDO	BBB	505	-	-	0/1/1/1	-
3	EDO	AAA	502	-	-	0/1/1/1	-
6	P9Q	AAA	506	5,1	-	3/8/28/32	0/1/1/1
2	NAG	AAA	501	1	-	2/6/23/26	0/1/1/1
3	EDO	BBB	503	-	-	0/1/1/1	-
3	EDO	BBB	504	-	-	0/1/1/1	-
6	P9Q	BBB	507	5,1	-	0/9/29/32	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	BBB	502	1	-	0/6/23/26	0/1/1/1
5	GLC	AAA	505	6	-	0/2/19/22	0/1/1/1
3	EDO	AAA	503	-	-	1/1/1/1	-
5	GLC	BBB	506	6	-	0/2/19/22	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	AAA	506	P9Q	C7'-N1	3.70	1.55	1.47
6	AAA	506	P9Q	C7'-N1	2.67	1.52	1.47
6	BBB	507	P9Q	C7'-N1	2.16	1.52	1.47
6	AAA	506	P9Q	C1'-C2'	2.12	1.56	1.52
6	BBB	507	P9Q	C1'-C7'	2.06	1.56	1.53

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	BBB	507	P9Q	C1'-C7'-C5'	5.53	117.84	111.49
2	AAA	501	NAG	O5-C5-C4	4.67	122.20	110.83
6	AAA	506	P9Q	C4'-C5'-C7'	4.40	118.91	109.26
6	BBB	507	P9Q	C4'-C5'-C7'	4.27	118.63	109.26
6	AAA	506	P9Q	C1'-C7'-C5'	3.90	115.96	111.49

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	501	NAG	C4-C5-C6-O6
2	AAA	501	NAG	O5-C5-C6-O6
6	AAA	506	P9Q	C7-C10-C11-C12
3	AAA	503	EDO	O1-C1-C2-O2
6	AAA	506	P9Q	C11-C10-C7-N1

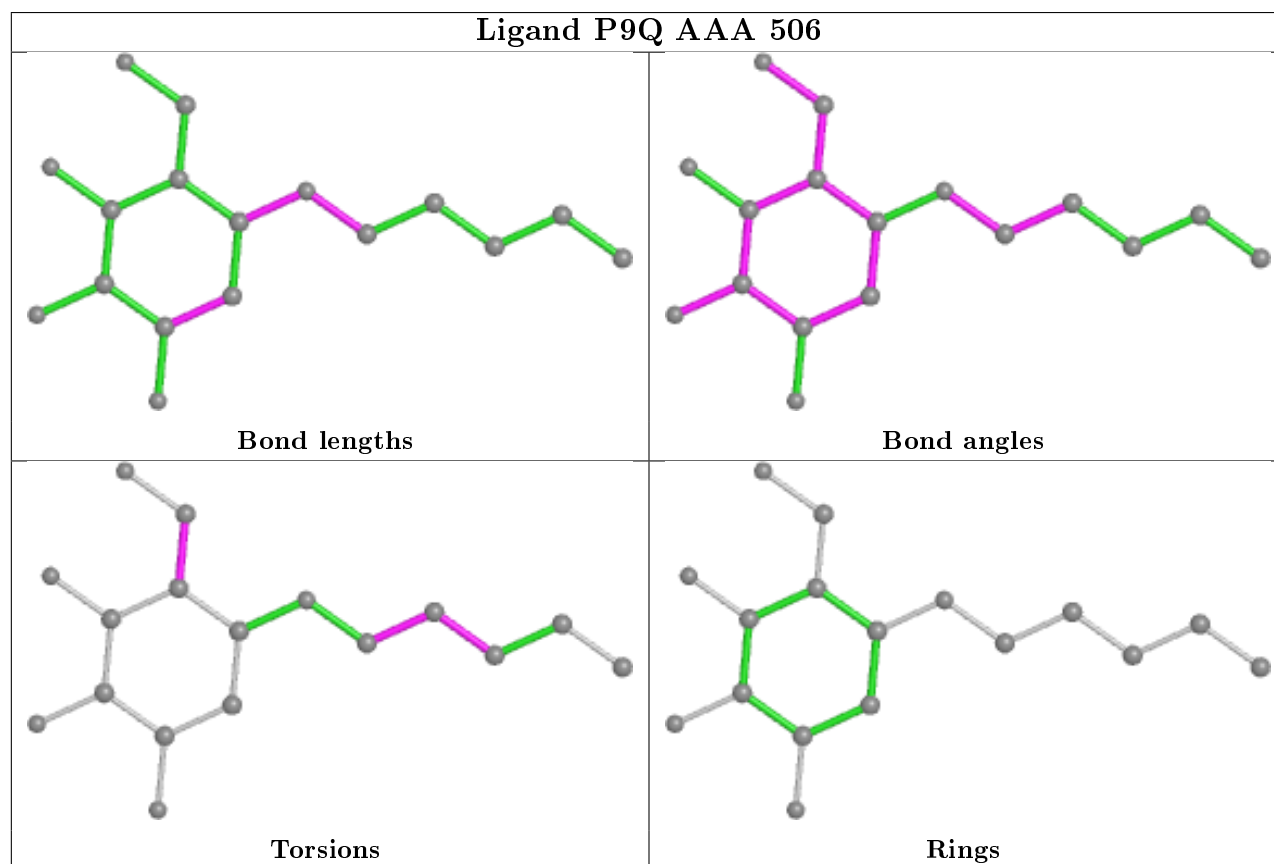
There are no ring outliers.

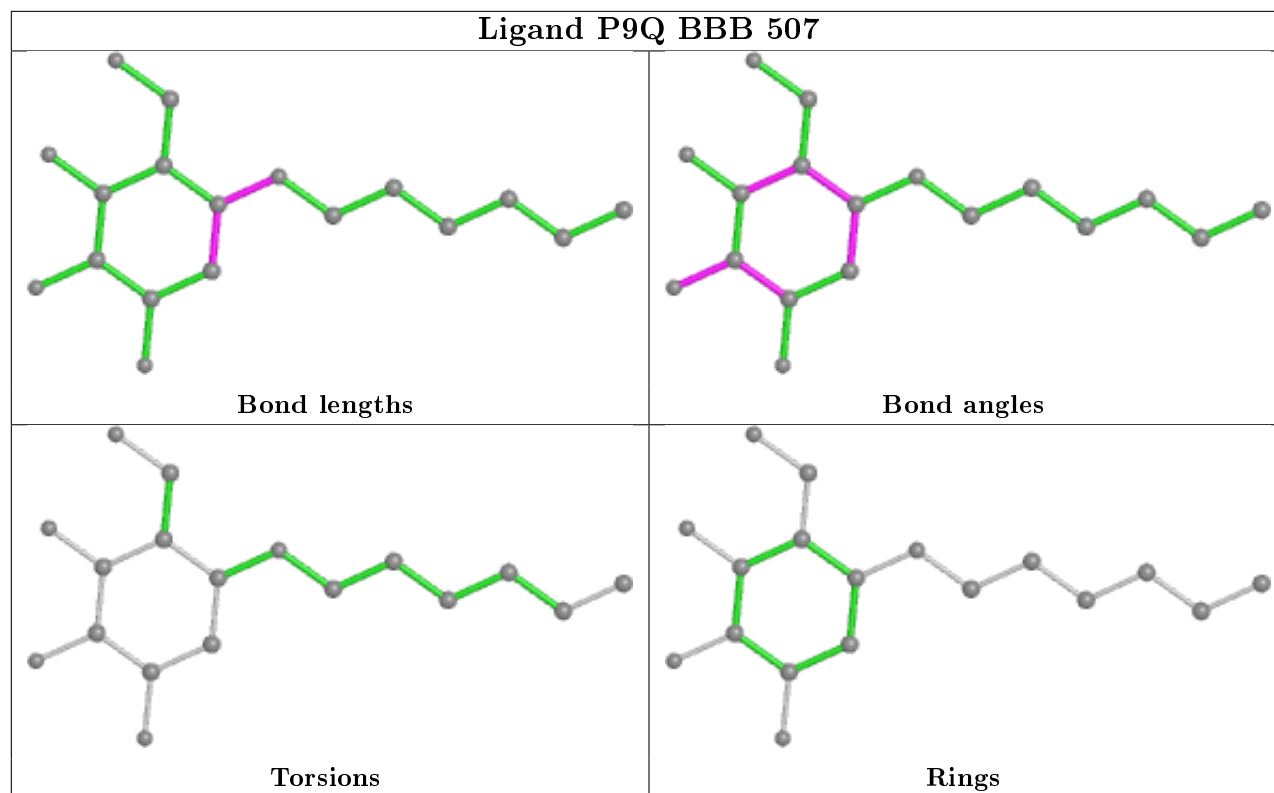
2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	AAA	506	P9Q	1	0
6	BBB	507	P9Q	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 than 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, 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	AAA	476/499 (95%)	-0.55	0 100 100	18, 25, 39, 60	3 (0%)
1	BBB	476/499 (95%)	-0.51	0 100 100	19, 28, 43, 69	2 (0%)
All	All	952/998 (95%)	-0.53	0 100 100	18, 26, 42, 69	5 (0%)

There are no RSRZ outliers to report.

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, 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.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	NAG	AAA	501	14/15	0.87	0.10	41,44,56,56	3
2	NAG	BBB	502	14/15	0.88	0.10	40,44,55,56	3
3	EDO	AAA	502	4/4	0.94	0.10	26,28,34,34	1
5	GLC	AAA	505	11/12	0.95	0.07	21,25,27,27	3
3	EDO	BBB	505	4/4	0.96	0.10	28,34,36,36	1
3	EDO	BBB	503	4/4	0.96	0.09	29,34,38,38	1
5	GLC	BBB	506	11/12	0.96	0.07	22,23,25,27	3

Continued on next page...

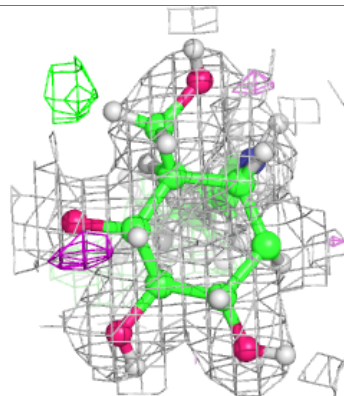
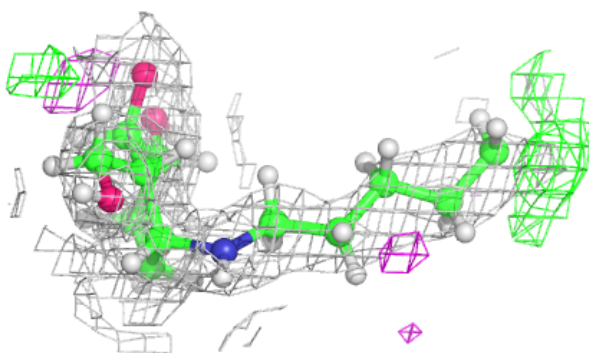
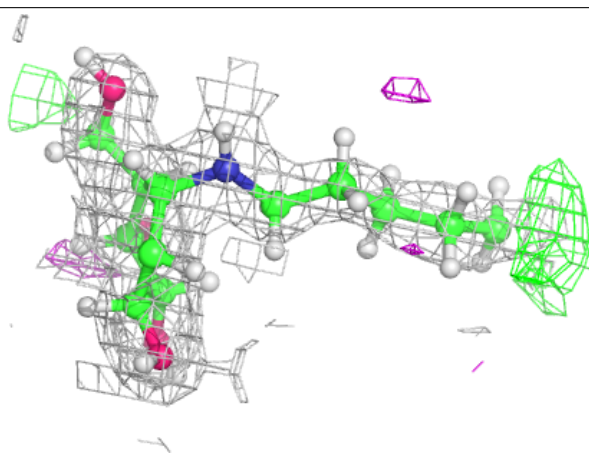
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
6	P9Q	AAA	506	17/21	0.96	0.08	22,29,47,47	4
3	EDO	AAA	503	4/4	0.97	0.07	30,36,39,39	1
6	P9Q	BBB	507	18/21	0.97	0.08	20,27,44,46	4
3	EDO	BBB	504	4/4	0.98	0.08	30,34,34,34	1
4	CA	AAA	504	1/1	0.98	0.09	22,22,22,22	0
4	CA	BBB	501	1/1	0.99	0.08	22,22,22,22	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.

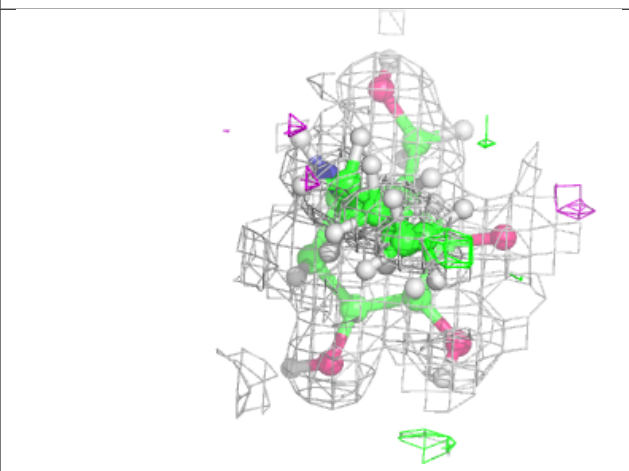
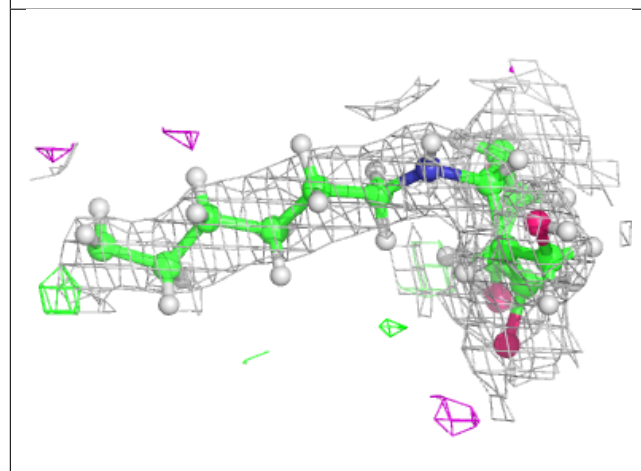
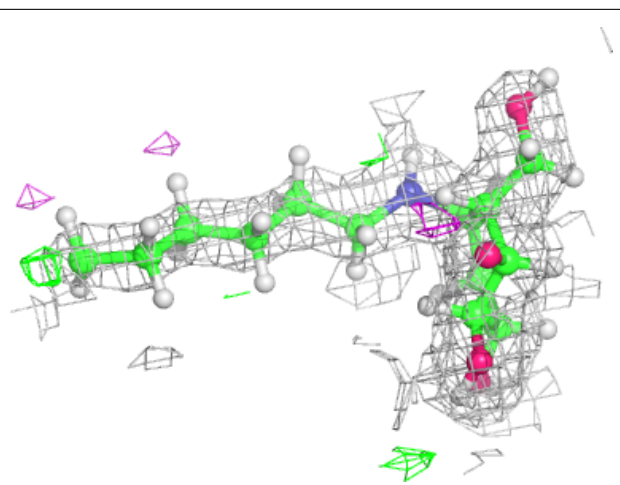
Electron density around P9Q AAA 506:

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



Electron density around P9Q BBB 507:

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



6.5 Other polymers [i](#)

There are no such residues in this entry.