wwPDB X-ray Structure Validation Summary Report (i)

Apr 6, 2021 – 11:06 am BST

PDB ID	:	6YHJ
Title	:	Thrombin in complex with D-Phe-Pro-2-chlorothiophen derivative (16e)
Authors	:	Sandner, A.; Heine, A.; Klebe, G.; Collins, C.
Deposited on	:	2020-03-30
Resolution	:	1.44 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 following versions of software and data (see references (i)) were used in the production of this report:

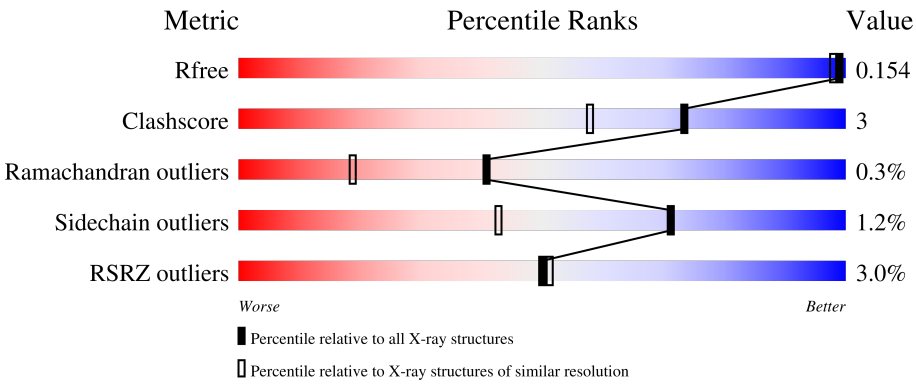
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.18
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.18

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 1.44 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2021 (1.46-1.42)
Clashscore	141614	2086 (1.46-1.42)
Ramachandran outliers	138981	2047 (1.46-1.42)
Sidechain outliers	138945	2047 (1.46-1.42)
RSRZ outliers	127900	1993 (1.46-1.42)

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	L	36	6%, 81%, 11%, 8%
2	H	259	2%, 91%, 6%, •
3	I	12	8%, 83%, 8%, 8%

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 5488 atoms, of which 2525 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	H	N	O	S			
1	L	33	596	189	291	50	65	1	0	11	0

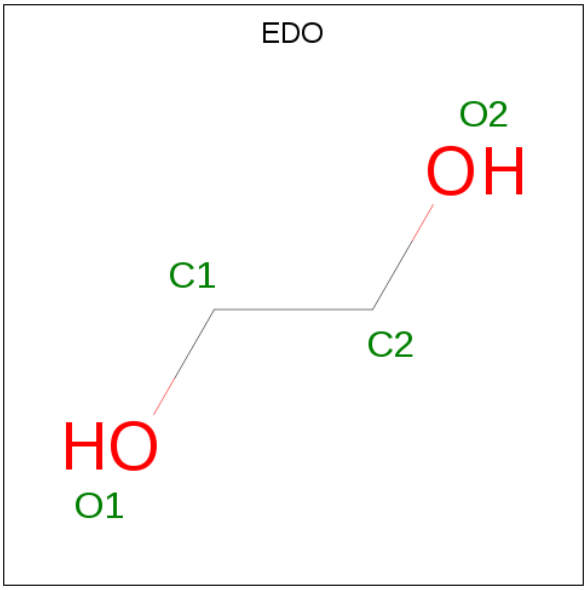
- Molecule 2 is a protein called Prothrombin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	H	253	4365	1400	2169	393	387	16	0	35	0

- Molecule 3 is a protein called Hirudin variant-2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	I	11	153	57	65	11	19	1	0	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

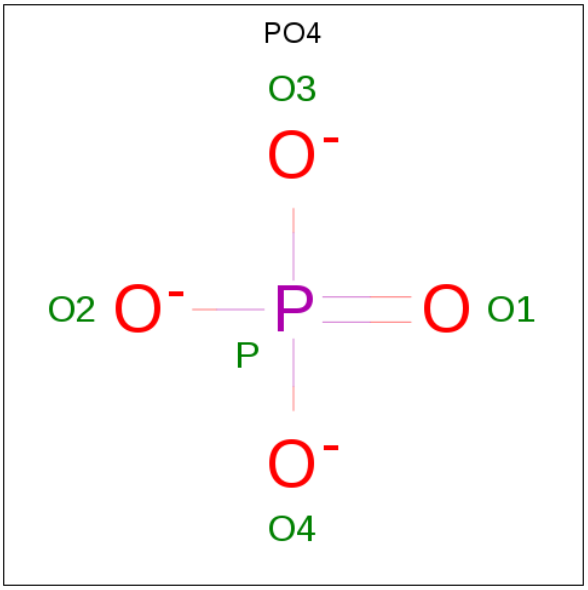


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	L	1	4	2	2	0	0

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

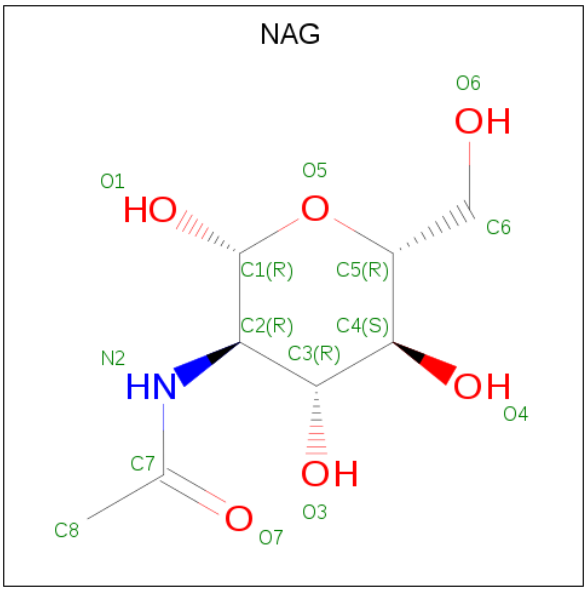
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
5	H	2	2	2	0	0

- Molecule 6 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



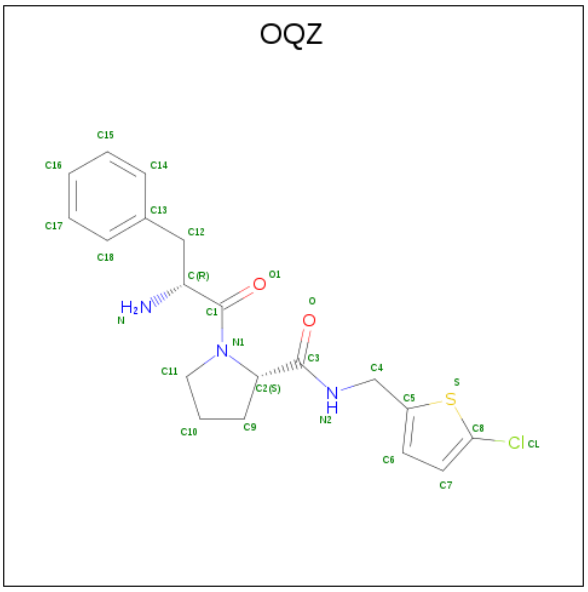
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
6	H	1	5	4	1	0	0

- Molecule 7 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



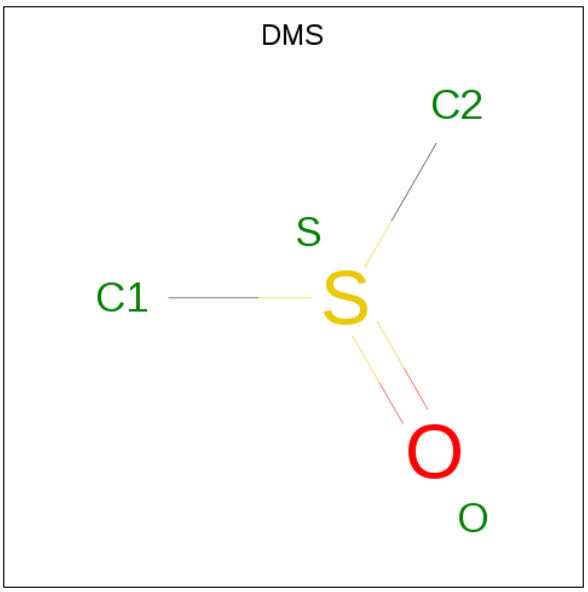
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
7	H	1	13	8	1	4	0	0

- Molecule 8 is (2 {S})-1-[(2 {R})-2-azanyl-3-phenyl-propanoyl]- {N}-[(5-chloranylthiophen-2 -yl)methyl]pyrrolidine-2-carboxamide (three-letter code: OQZ) (formula: $C_{19}H_{22}ClN_3O_2S$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	Cl	N	O	S		
8	H	1	26	19	1	3	2	1	0	0
			Total	C	Cl	N	O	S		
8	H	1	26	19	1	3	2	1	0	0

- Molecule 9 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	S		
9	H	1	4	2	1	1	0	0

- Molecule 10 is water.

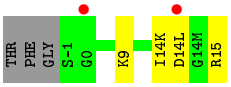
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
10	L	44	44	44	0	2
			Total	O		
10	H	245	247	247	0	4
			Total	O		
10	I	3	3	3	0	0

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.

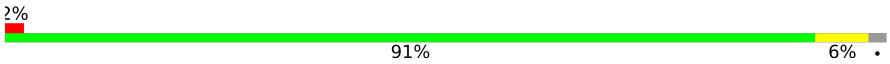
- Molecule 1: Prothrombin

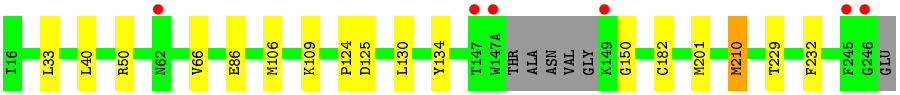
Chain L:



- Molecule 2: Prothrombin

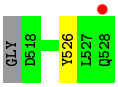
Chain H:





- Molecule 3: Hirudin variant-2

Chain I:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	70.28Å 71.40Å 72.56Å 90.00° 100.58° 90.00°	Depositor
Resolution (Å)	43.53 – 1.44 43.53 – 1.44	Depositor EDS
% Data completeness (in resolution range)	99.2 (43.53-1.44) 99.3 (43.53-1.44)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.49 (at 1.44Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.127 , 0.154 0.128 , 0.154	Depositor DCC
R_{free} test set	3170 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	14.5	Xtriage
Anisotropy	0.335	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 52.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	5488	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.66% 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: TYS, NA, PO4, OQZ, NAG, EDO, DMS

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	L	0.45	0/332	0.71	0/436
2	H	0.47	0/2402	0.70	0/3244
3	I	0.36	0/72	0.40	0/96
All	All	0.46	0/2806	0.70	0/3776

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	L	305	291	272	3	0
2	H	2196	2169	2017	15	0
3	I	88	65	66	0	0
4	L	4	0	6	0	0
5	H	2	0	0	0	0
6	H	5	0	0	0	0
7	H	13	0	10	0	0
8	H	52	0	0	2	0
9	H	4	0	6	0	0
10	H	247	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
10	I	3	0	0	0	0
10	L	44	0	0	1	0
All	All	2963	2525	2377	17	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.

The worst 5 of 17 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:L:9[B]:LYS:HG2	10:L:240:HOH:O	1.82	0.79
2:H:33[A]:LEU:CD2	2:H:66[A]:VAL:HG12	2.21	0.71
2:H:86[B]:GLU:HG2	2:H:109:LYS:HA	1.78	0.65
2:H:124:PRO:O	8:H:305:OQZ:CL	2.63	0.54
2:H:40[B]:LEU:C	2:H:40[B]:LEU:HD13	2.30	0.52

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	L	39/36 (108%)	38 (97%)	1 (3%)	0	100	100
2	H	284/259 (110%)	271 (95%)	12 (4%)	1 (0%)	34	13
3	I	8/12 (67%)	8 (100%)	0	0	100	100
All	All	331/307 (108%)	317 (96%)	13 (4%)	1 (0%)	41	19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	H	150	GLY

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	L	36/31 (116%)	36 (100%)	0	100	100
2	H	246/225 (109%)	241 (98%)	5 (2%)	55	21
3	I	6/10 (60%)	6 (100%)	0	100	100
All	All	288/266 (108%)	283 (98%)	5 (2%)	71	28

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	50[A]	ARG
2	H	50[B]	ARG
2	H	182	CYS
2	H	210[A]	MET
2	H	210[B]	MET

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)

1 non-standard protein/DNA/RNA residue is 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	TYS	I	526	3	15,16,17	1.12	2 (13%)	18,22,24	1.02	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
3	TYS	I	526	3	-	0/10/11/13	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	I	526	TYS	OH-CZ	-3.03	1.37	1.42
3	I	526	TYS	O3-S	2.00	1.62	1.50

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	I	526	TYS	CG-CB-CA	-2.01	110.03	114.10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	$\#\|Z\| > 2$
6	PO4	H	303	-	4,4,4	0.87	0	6,6,6	1.53	2 (33%)
8	OQZ	H	306	-	27,28,28	0.45	0	29,38,38	0.52	0
7	NAG	H	304	2	13,13,15	0.28	0	16,18,21	0.73	0
8	OQZ	H	305	-	27,28,28	0.61	1 (3%)	29,38,38	0.67	1 (3%)
9	DMS	H	307	-	3,3,3	0.63	0	3,3,3	0.78	0
4	EDO	L	101	-	3,3,3	0.56	0	2,2,2	0.37	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
8	OQZ	H	305	-	-	0/19/31/31	0/3/3/3
8	OQZ	H	306	-	-	0/19/31/31	0/3/3/3
7	NAG	H	304	2	-	1/4/21/26	0/1/1/1
4	EDO	L	101	-	-	1/1/1/1	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	H	305	OQZ	C4-C5	-2.09	1.49	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
8	H	305	OQZ	C5-C4-N2	-3.27	105.64	112.71
6	H	303	PO4	O3-P-O1	-2.30	102.48	110.89
6	H	303	PO4	O4-P-O2	2.05	114.56	107.97

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	H	304	NAG	C3-C2-N2-C7
4	L	101	EDO	O1-C1-C2-O2

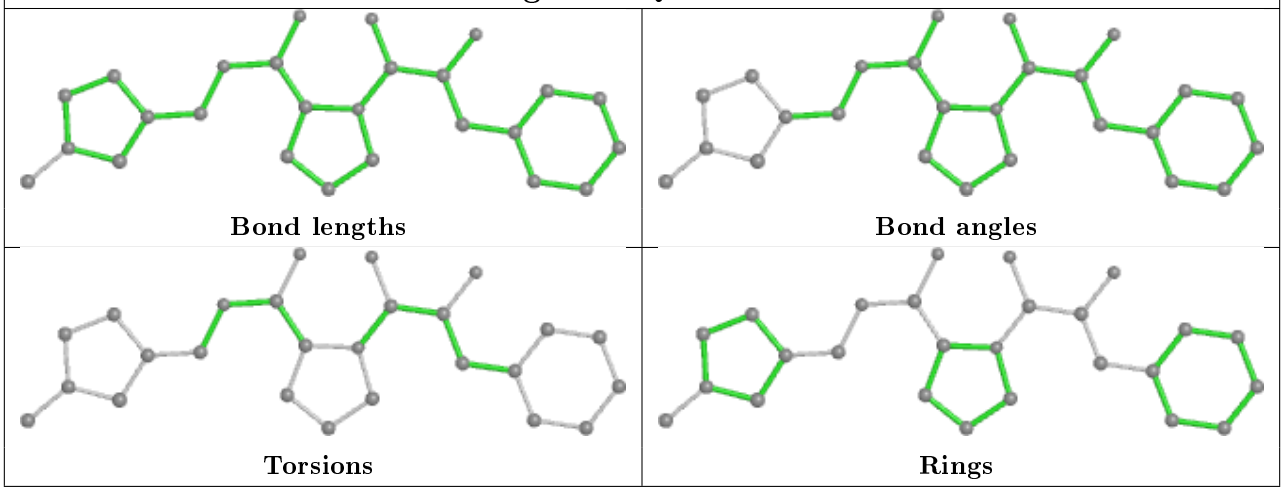
There are no ring outliers.

1 monomer is involved in 2 short contacts:

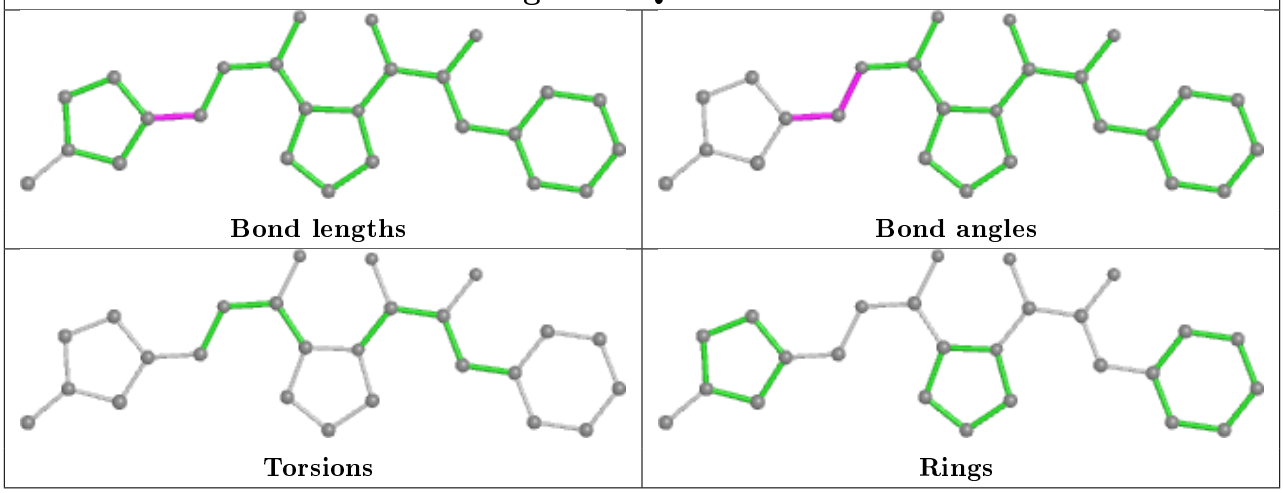
Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	H	305	OQZ	2	0

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 OQZ H 306



Ligand OQZ H 305



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	L	33/36 (91%)	0.08	2 (6%)	21	20	11, 15, 31, 37	4 (12%)
2	H	253/259 (97%)	-0.31	6 (2%)	59	60	10, 18, 37, 51	0
3	I	10/12 (83%)	0.49	1 (10%)	7	7	31, 39, 46, 51	0
All	All	296/307 (96%)	-0.24	9 (3%)	50	51	10, 18, 38, 51	4 (1%)

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	147(A)	TRP	7.9
2	H	149	LYS	4.3
2	H	245	PHE	3.8
2	H	147	THR	3.1
2	H	62	ASN	2.7

6.2 Non-standard residues in protein, DNA, RNA chains (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
3	TYS	I	526	16/17	0.97	0.09	31,37,44,46	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

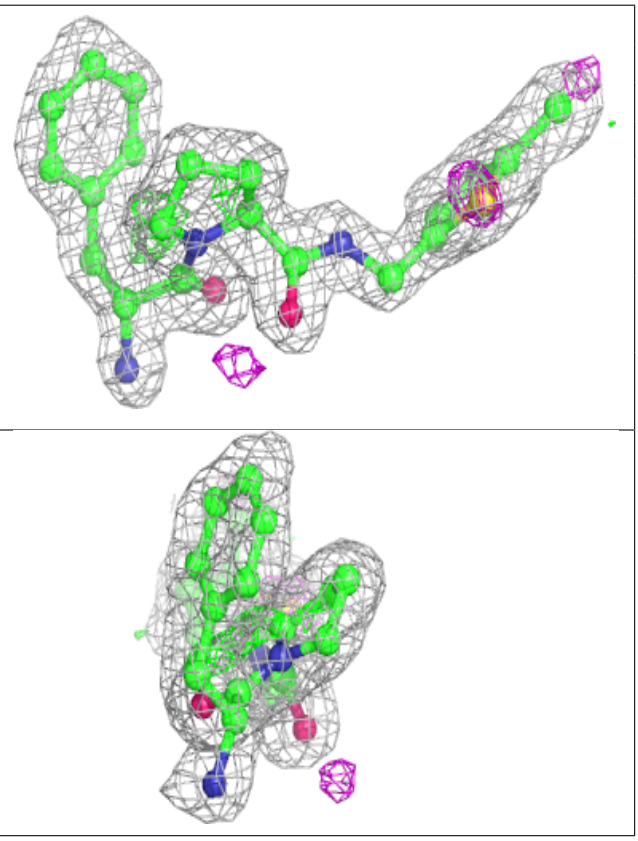
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

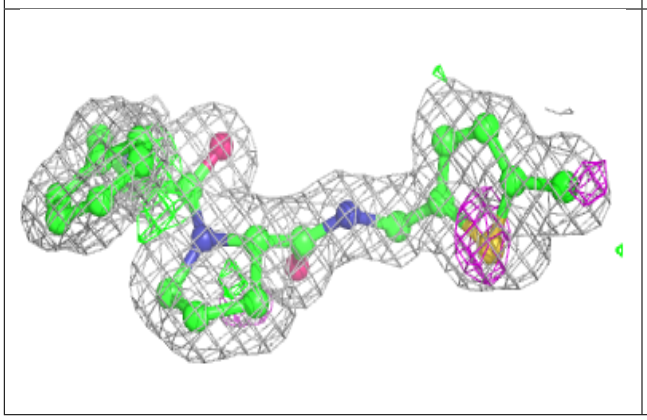
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	EDO	L	101	4/4	0.80	0.14	27,29,31,32	4
7	NAG	H	304	13/15	0.87	0.15	43,47,49,50	0
8	OQZ	H	305	26/26	0.90	0.10	21,26,29,32	26
8	OQZ	H	306	26/26	0.90	0.09	15,17,34,37	0
9	DMS	H	307	4/4	0.91	0.13	47,48,48,48	0
6	PO4	H	303	5/5	0.97	0.06	20,21,23,23	5
5	NA	H	302	1/1	0.99	0.06	13,13,13,13	0
5	NA	H	301	1/1	0.99	0.06	18,18,18,18	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 OQZ H 305:

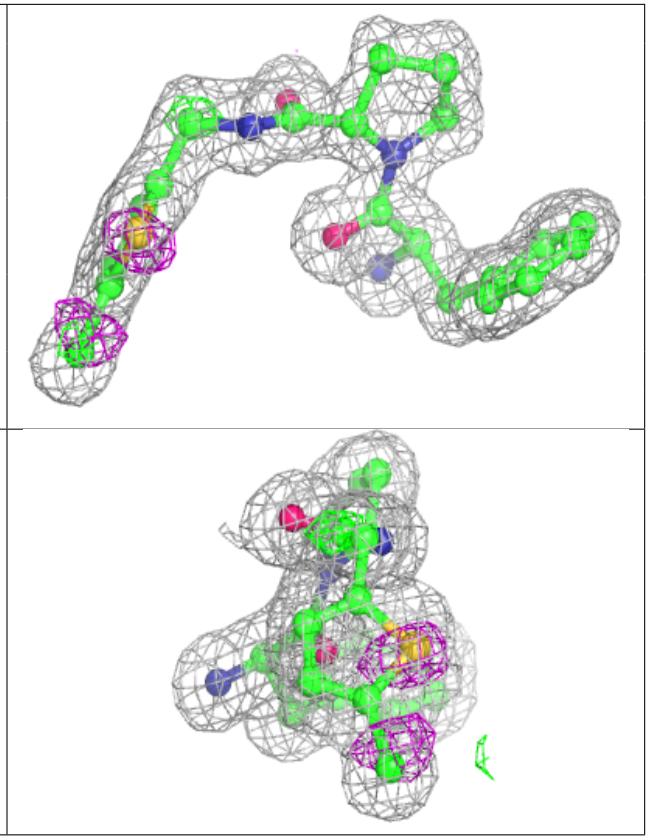
2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)





Electron density around OQZ H 306:

$2mF_o$-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.