



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 27, 2023 – 01:11 pm BST

PDB ID : 7ZB3  
Title : Crystal structure of beta-xylosidase from *Thermotoga maritima* in complex with xylohexaose hydrolysed to xylobiose  
Authors : Gloster, T.M.; Foltanyi, F.  
Deposited on : 2022-03-23  
Resolution : 1.51 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](mailto: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)  
Xtrriage (Phenix) : 1.13  
EDS : 2.32.2  
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) : 2.32.2

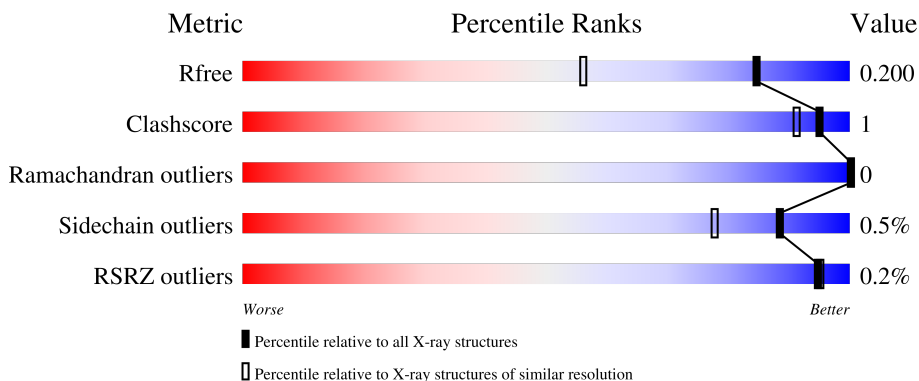
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.51 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	4009 (1.54-1.50)
Clashscore	141614	4249 (1.54-1.50)
Ramachandran outliers	138981	4148 (1.54-1.50)
Sidechain outliers	138945	4146 (1.54-1.50)
RSRZ outliers	127900	3943 (1.54-1.50)

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  $\geq 3$ , 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	A	801	
1	B	801	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 25150 atoms, of which 12020 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 Beta-xylosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	765	12019	3843	5987	1033	1132	24	0	12	0
1	B	761	12004	3835	5983	1034	1129	23	0	12	0

There are 46 discrepancies between the modelled and reference sequences:

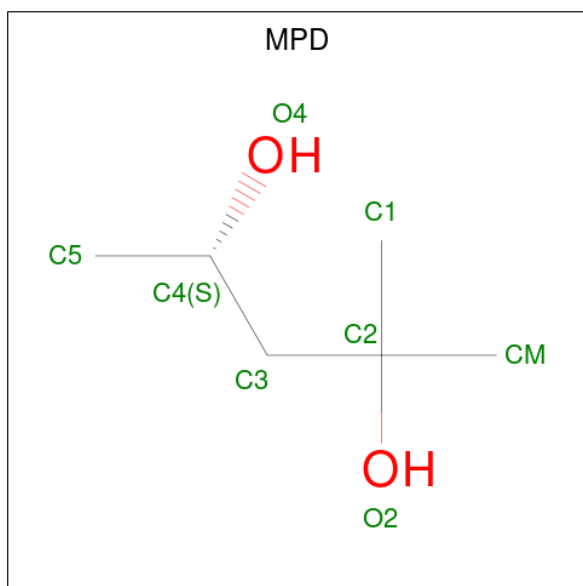
Chain	Residue	Modelled	Actual	Comment	Reference
A	779	ALA	-	expression tag	UNP R4NX63
A	780	ALA	-	expression tag	UNP R4NX63
A	781	ALA	-	expression tag	UNP R4NX63
A	782	LEU	-	expression tag	UNP R4NX63
A	783	GLU	-	expression tag	UNP R4NX63
A	784	GLU	-	expression tag	UNP R4NX63
A	785	ASN	-	expression tag	UNP R4NX63
A	786	LEU	-	expression tag	UNP R4NX63
A	787	TYR	-	expression tag	UNP R4NX63
A	788	PHE	-	expression tag	UNP R4NX63
A	789	GLN	-	expression tag	UNP R4NX63
A	790	GLY	-	expression tag	UNP R4NX63
A	791	ALA	-	expression tag	UNP R4NX63
A	792	HIS	-	expression tag	UNP R4NX63
A	793	HIS	-	expression tag	UNP R4NX63
A	794	HIS	-	expression tag	UNP R4NX63
A	795	HIS	-	expression tag	UNP R4NX63
A	796	HIS	-	expression tag	UNP R4NX63
A	797	HIS	-	expression tag	UNP R4NX63
A	798	HIS	-	expression tag	UNP R4NX63
A	799	HIS	-	expression tag	UNP R4NX63
A	800	HIS	-	expression tag	UNP R4NX63
A	801	HIS	-	expression tag	UNP R4NX63
B	779	ALA	-	expression tag	UNP R4NX63
B	780	ALA	-	expression tag	UNP R4NX63

*Continued on next page...*

Continued from previous page...

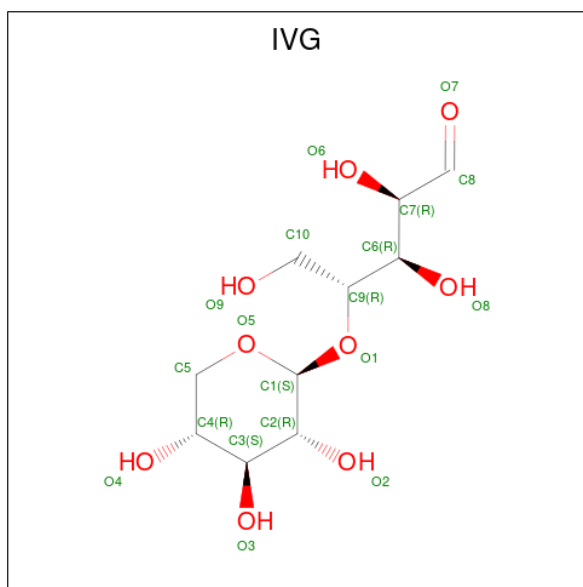
Chain	Residue	Modelled	Actual	Comment	Reference
B	781	ALA	-	expression tag	UNP R4NX63
B	782	LEU	-	expression tag	UNP R4NX63
B	783	GLU	-	expression tag	UNP R4NX63
B	784	GLU	-	expression tag	UNP R4NX63
B	785	ASN	-	expression tag	UNP R4NX63
B	786	LEU	-	expression tag	UNP R4NX63
B	787	TYR	-	expression tag	UNP R4NX63
B	788	PHE	-	expression tag	UNP R4NX63
B	789	GLN	-	expression tag	UNP R4NX63
B	790	GLY	-	expression tag	UNP R4NX63
B	791	ALA	-	expression tag	UNP R4NX63
B	792	HIS	-	expression tag	UNP R4NX63
B	793	HIS	-	expression tag	UNP R4NX63
B	794	HIS	-	expression tag	UNP R4NX63
B	795	HIS	-	expression tag	UNP R4NX63
B	796	HIS	-	expression tag	UNP R4NX63
B	797	HIS	-	expression tag	UNP R4NX63
B	798	HIS	-	expression tag	UNP R4NX63
B	799	HIS	-	expression tag	UNP R4NX63
B	800	HIS	-	expression tag	UNP R4NX63
B	801	HIS	-	expression tag	UNP R4NX63

- Molecule 2 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			21	6	13	2		
2	B	1	Total	C	H	O	0	0
			21	6	13	2		

- Molecule 3 is D-Xylose, 4-O-beta-D-xylopyranosyl- (three-letter code: IVG) (formula: C<sub>10</sub>H<sub>18</sub>O<sub>9</sub>) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			31	10	12	9		
3	B	1	Total	C	H	O	0	0
			31	10	12	9		

- Molecule 4 is water.

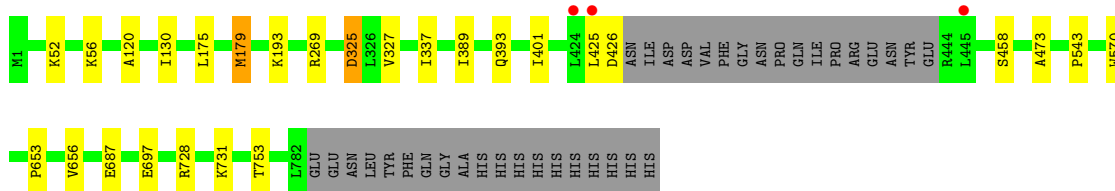
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	513	Total	O	0	0
			513	513		
4	B	510	Total	O	0	0
			510	510		

### 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.

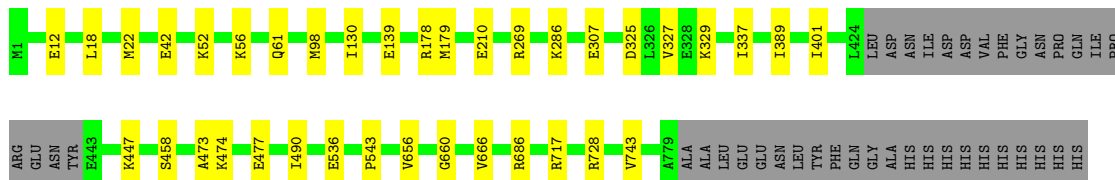
- Molecule 1: Beta-xylosidase

Chain A:  92%



- Molecule 1: Beta-xylosidase

Chain B:  90% 5% 5%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	178.53Å 98.72Å 100.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	87.71 – 1.51 87.71 – 1.51	Depositor EDS
% Data completeness (in resolution range)	93.5 (87.71-1.51) 93.5 (87.71-1.51)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.22 (at 1.51Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.167 , 0.190 0.179 , 0.200	Depositor DCC
$R_{free}$ test set	12762 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.4	Xtrriage
Anisotropy	0.617	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 41.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.000 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	25150	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.83% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: MPD, IVG

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	0.76	3/6167 (0.0%)	0.85	4/8329 (0.0%)
1	B	0.77	4/6147 (0.1%)	0.88	10/8300 (0.1%)
All	All	0.77	7/12314 (0.1%)	0.86	14/16629 (0.1%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	210	GLU	CD-OE1	7.59	1.34	1.25
1	A	687	GLU	CD-OE2	-7.25	1.17	1.25
1	B	307	GLU	CD-OE2	-5.89	1.19	1.25
1	B	42	GLU	CD-OE2	5.85	1.32	1.25
1	B	12	GLU	CD-OE2	5.30	1.31	1.25

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	269[A]	ARG	NE-CZ-NH2	-13.17	113.71	120.30
1	B	269[B]	ARG	NE-CZ-NH2	-13.17	113.71	120.30
1	A	269[A]	ARG	NE-CZ-NH2	-9.72	115.44	120.30
1	A	269[B]	ARG	NE-CZ-NH2	-9.72	115.44	120.30
1	A	179	MET	CG-SD-CE	-9.65	84.76	100.20

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	6032	5987	6090	15	0
1	B	6021	5983	6082	14	0
2	A	8	13	14	0	0
2	B	8	13	14	1	0
3	A	19	12	0	0	0
3	B	19	12	0	0	0
4	A	513	0	0	7	0
4	B	510	0	0	3	0
All	All	13130	12020	12200	30	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 1.

The worst 5 of 30 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:B:139[A]:GLU:HG2	4:B:1096:HOH:O	1.77	0.84
1:B:286[B]:LYS:NZ	4:B:1001:HOH:O	2.32	0.63
1:B:327:VAL:HG21	1:B:337:ILE:HD12	1.79	0.62
1:B:474:LYS:NZ	1:B:477:GLU:OE2	2.37	0.57
1:A:753[A]:THR:HG23	4:A:1153:HOH:O	2.04	0.57

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	773/801 (96%)	757 (98%)	16 (2%)	0	100 100
1	B	769/801 (96%)	751 (98%)	18 (2%)	0	100 100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	1542/1602 (96%)	1508 (98%)	34 (2%)	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	644/678 (95%)	641 (100%)	3 (0%)	88	78
1	B	644/678 (95%)	641 (100%)	3 (0%)	88	78
All	All	1288/1356 (95%)	1282 (100%)	6 (0%)	88	78

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	130	ILE
1	B	447	LYS
1	B	717	ARG
1	A	325	ASP
1	A	130	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	241	ASN
1	B	57	ASN
1	B	241	ASN

### 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](#)

4 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	IVG	B	902	-	18,19,19	0.94	2 (11%)	23,26,26	1.58	3 (13%)
2	MPD	B	901	-	7,7,7	0.71	0	9,10,10	0.74	0
3	IVG	A	902	-	18,19,19	1.25	3 (16%)	23,26,26	2.66	8 (34%)
2	MPD	A	901	-	7,7,7	0.71	0	9,10,10	0.74	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
3	IVG	B	902	-	-	6/14/33/33	0/1/1/1
2	MPD	B	901	-	-	0/5/5/5	-
3	IVG	A	902	-	-	9/14/33/33	0/1/1/1
2	MPD	A	901	-	-	0/5/5/5	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	902	IVG	O1-C9	3.23	1.50	1.44
3	A	902	IVG	O5-C1	2.44	1.46	1.41

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	902	IVG	C6-C7	2.38	1.57	1.53
3	B	902	IVG	O5-C1	2.37	1.46	1.41
3	B	902	IVG	O1-C9	2.22	1.48	1.44

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	902	IVG	O1-C9-C6	8.28	123.22	107.47
3	A	902	IVG	O8-C6-C9	-5.73	95.99	109.16
3	B	902	IVG	O1-C9-C6	4.88	116.75	107.47
3	A	902	IVG	C1-O1-C9	3.74	121.11	115.33
3	A	902	IVG	O6-C7-C6	3.59	117.98	109.46

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

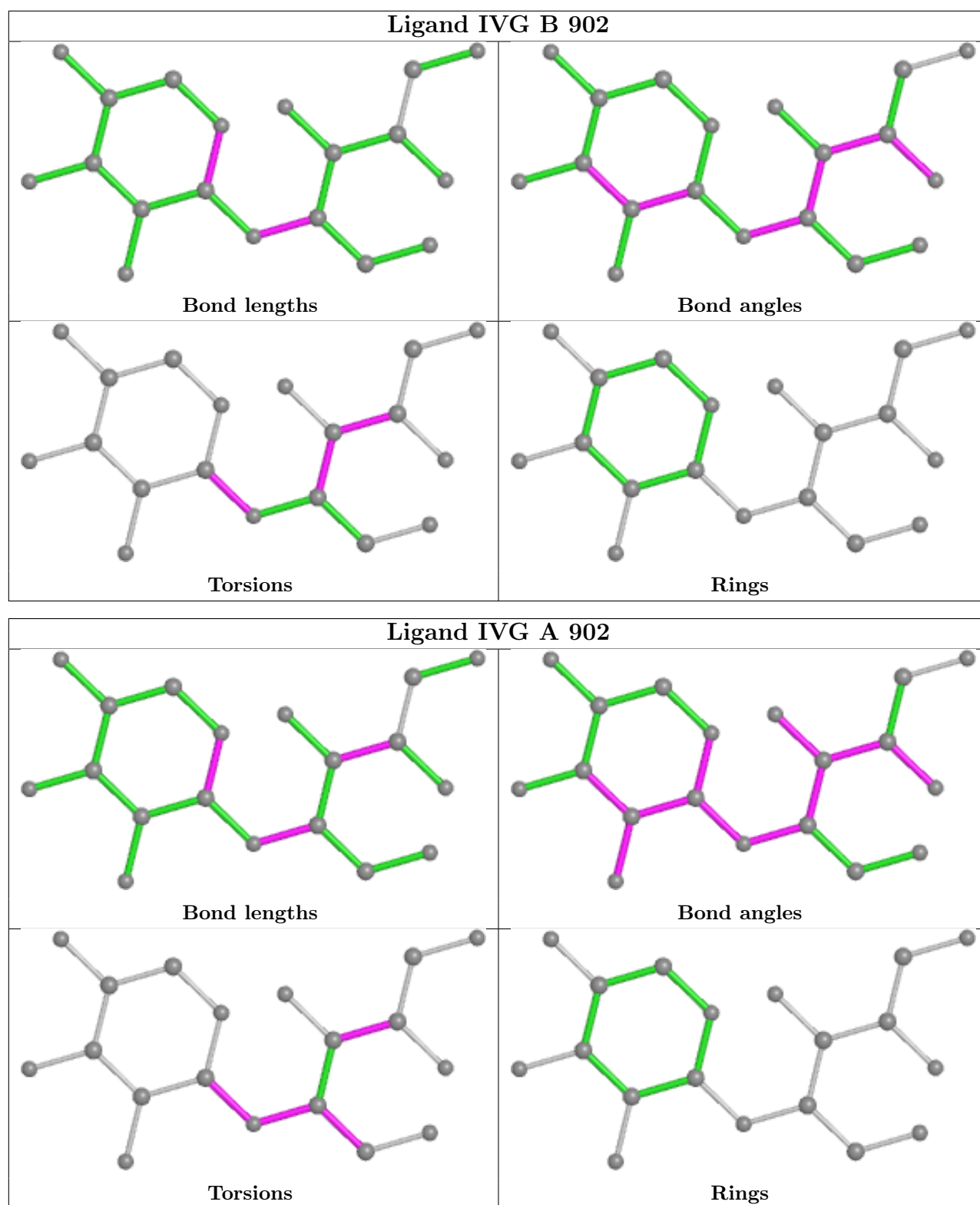
Mol	Chain	Res	Type	Atoms
3	A	902	IVG	C10-C9-O1-C1
3	A	902	IVG	C6-C9-O1-C1
3	A	902	IVG	O8-C6-C7-O6
3	A	902	IVG	C9-C6-C7-O6
3	A	902	IVG	C9-C6-C7-C8

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	901	MPD	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

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	765/801 (95%)	-0.56	3 (0%) 92   94	16, 23, 38, 79	0
1	B	761/801 (95%)	-0.63	0 100   100	16, 22, 37, 56	0
All	All	1526/1602 (95%)	-0.59	3 (0%) 95   95	16, 22, 38, 79	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	425	LEU	2.7
1	A	445	LEU	2.3
1	A	424	LEU	2.3

### 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<sup>th</sup> 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.

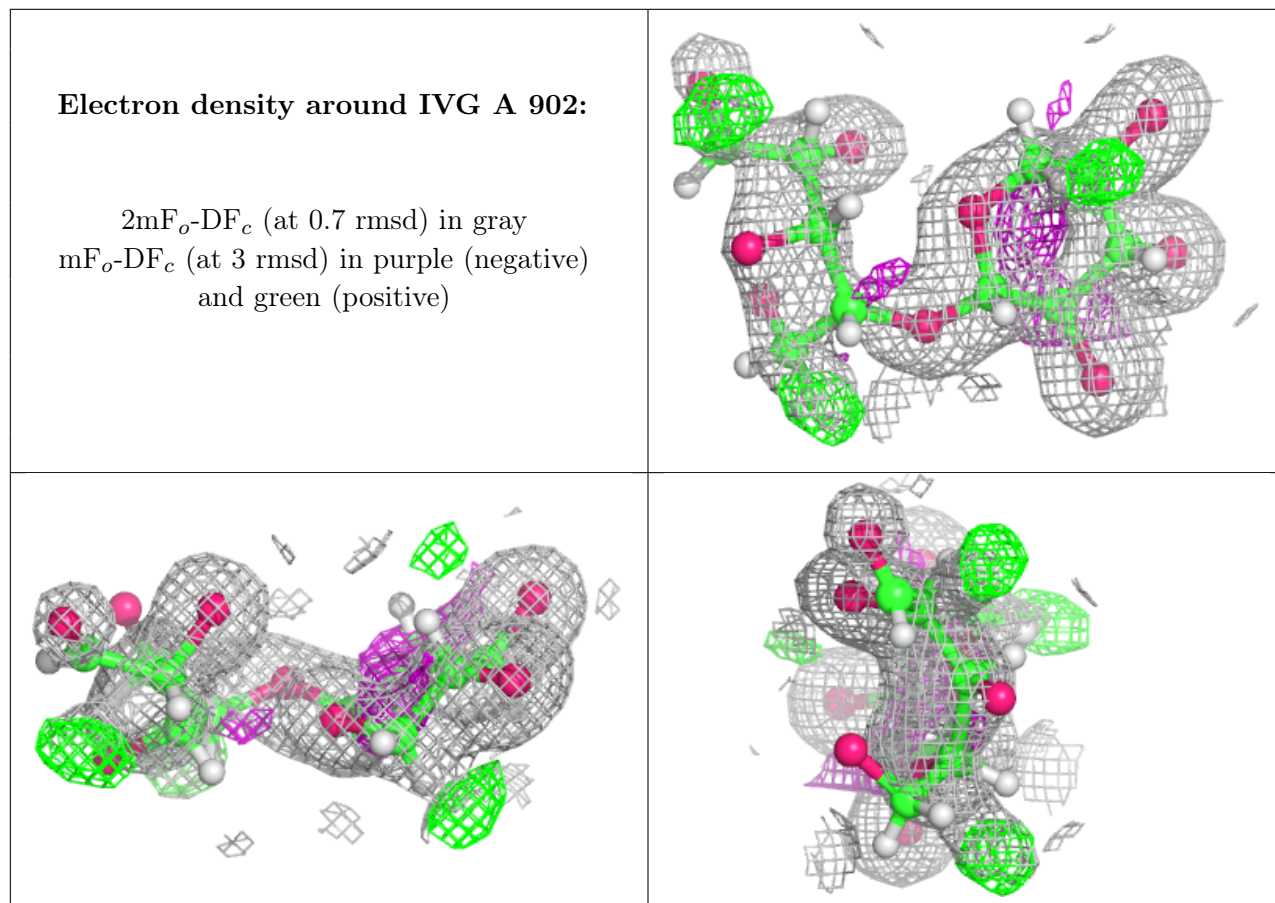
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	MPD	A	901	8/8	0.88	0.08	24,25,37,38	0
3	IVG	A	902	19/19	0.89	0.18	23,36,64,71	0

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	IVG	B	902	19/19	0.91	0.22	21,36,59,73	0
2	MPD	B	901	8/8	0.93	0.11	23,28,38,39	0

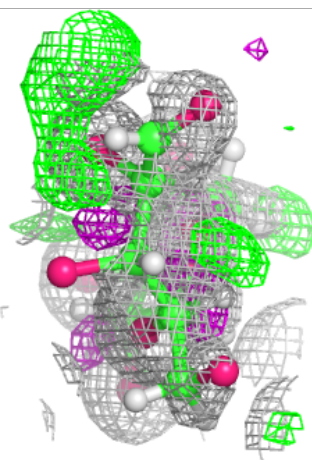
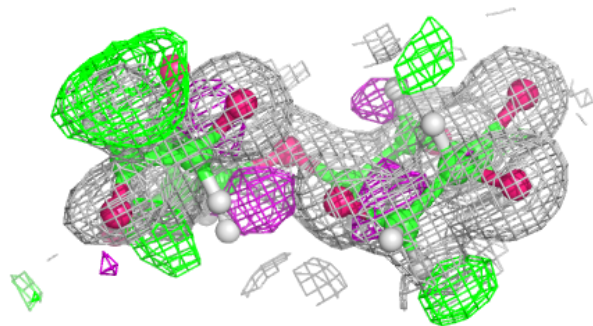
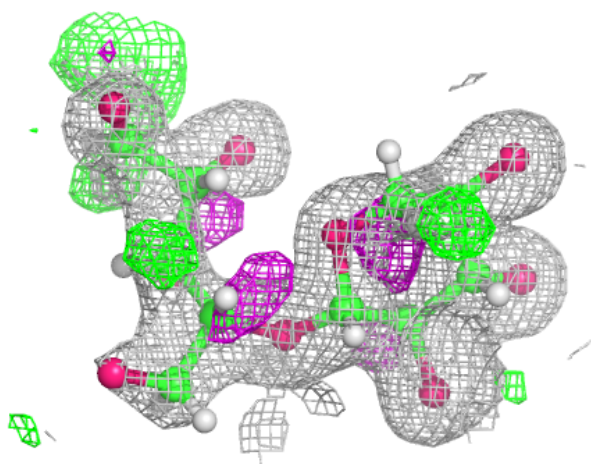
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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





**Electron density around IVG B 902:**

$2mF_o-DF_c$  (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.