Full wwPDB X-ray Structure Validation Report (i)

Mar 4, 2024 – 09:31 AM EST

PDB ID	:	1FH0
Title	:	CRYSTAL STRUCTURE OF HUMAN CATHEPSIN V COMPLEXED WITH AN IRREVERSIBLE VINYL SULFONE INHIBITOR
Authors	:	Somoza, J.R.
Deposited on	:	2000-07-30
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 (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:

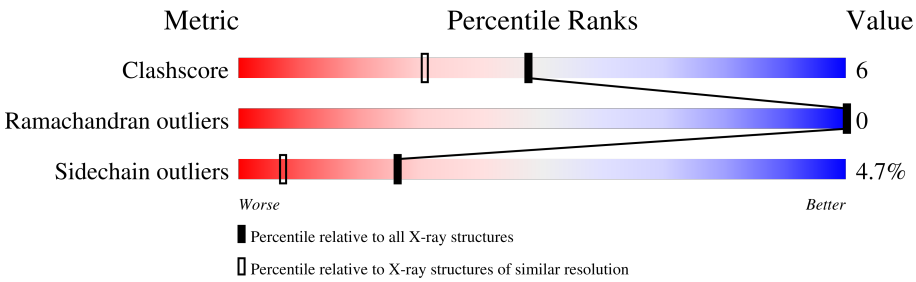
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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))
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	221	85% 14% •
1	B	221	88% 12%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3757 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.

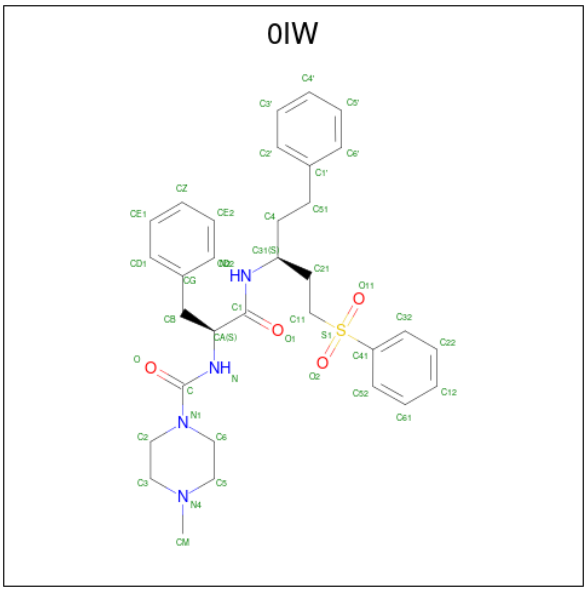
- Molecule 1 is a protein called CATHEPSIN V.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	221	1689	1064	289	325	11	0	0	0
1	B	221	1689	1064	289	325	11	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

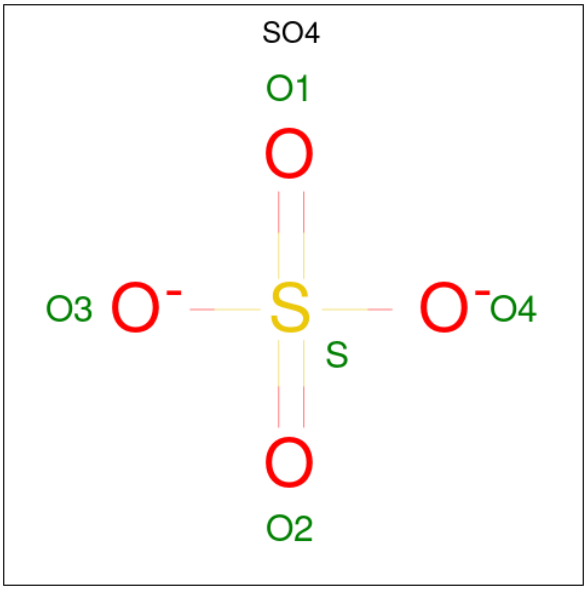
Chain	Residue	Modelled	Actual	Comment	Reference
A	108	GLN	ASN	engineered mutation	UNP O60911
A	178	ASP	ASN	engineered mutation	UNP O60911
B	2108	GLN	ASN	engineered mutation	UNP O60911
B	2178	ASP	ASN	engineered mutation	UNP O60911

- Molecule 2 is Nalpha-[(4-methylpiperazin-1-yl)carbonyl]-N-[(3S)-1-phenyl-5-(phenylsulfonyl)pentan-3-yl]-L-phenylalaninamide (three-letter code: 0IW) (formula: $C_{32}H_{40}N_4O_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	41	32	4	4	1	0	0
			Total	C	N	O	S		
2	B	1	41	32	4	4	1	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
3	A	1	10	8	2	0	1

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	123	123	123	0	0
			Total	O		
4	B	164	164	164	0	0

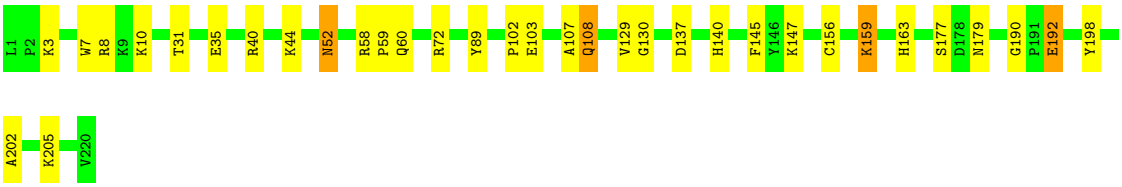
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

• Molecule 1: CATHEPSIN V

Chain A:



• Molecule 1: CATHEPSIN V

Chain B:

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 64 2 2	Depositor
Cell constants a, b, c, α, β, γ	104.70Å 104.70Å 179.20Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 1.60	Depositor
% Data completeness (in resolution range)	95.0 (50.00-1.60)	Depositor
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNX	Depositor
R, R_{free}	0.197 , 0.211	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3757	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

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: 0IW, SO4

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	0.29	0/1731	0.56	0/2339
1	B	0.28	0/1731	0.54	0/2339
All	All	0.29	0/3462	0.55	0/4678

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	A	1689	0	1617	23	0
1	B	1689	0	1614	18	0
2	A	41	0	39	1	0
2	B	41	0	39	0	0
3	A	10	0	0	0	0
4	A	123	0	0	2	0
4	B	164	0	0	1	0
All	All	3757	0	3309	41	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 6.

All (41) 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:A:108:GLN:N	1:A:108:GLN:HE21	1.68	0.90
1:A:108:GLN:HB3	4:A:527:HOH:O	1.70	0.89
1:B:2159:LYS:HA	1:B:2159:LYS:HE2	1.62	0.80
1:B:2077:VAL:HG12	1:B:2108:GLN:OE1	1.97	0.63
1:A:192:GLU:CD	1:A:192:GLU:H	2.03	0.61
1:A:159:LYS:HE2	1:A:159:LYS:HA	1.83	0.60
1:B:2007:TRP:CE2	1:B:2130:GLY:HA2	2.36	0.60
1:A:145:PHE:O	1:A:147:LYS:HE3	2.01	0.60
1:B:2178:ASP:OD2	4:B:309:HOH:O	2.18	0.56
1:A:140:HIS:HD2	4:A:577:HOH:O	1.89	0.55
1:A:163:HIS:ND1	2:A:1280:0IW:H11	2.20	0.55
1:A:190:GLY:HA3	1:A:192:GLU:OE2	2.10	0.52
1:B:2159:LYS:HE2	1:B:2159:LYS:CA	2.39	0.50
1:A:58:ARG:N	1:A:59:PRO:CD	2.76	0.49
1:A:108:GLN:N	1:A:108:GLN:NE2	2.51	0.48
1:A:137:ASP:OD1	1:A:140:HIS:HE1	1.97	0.48
1:A:7:TRP:CE2	1:A:130:GLY:HA2	2.48	0.48
1:B:2031:THR:O	1:B:2035:GLU:HG3	2.13	0.48
1:A:31:THR:O	1:A:35:GLU:HG3	2.14	0.48
1:B:2008:ARG:HD3	1:B:2198:TYR:CZ	2.50	0.47
1:B:2172:PHE:CG	1:B:2177:SER:HA	2.50	0.47
1:B:2058:ARG:N	1:B:2059:PRO:CD	2.77	0.47
1:A:40:ARG:HD2	1:A:129:VAL:O	2.14	0.46
1:A:107:ALA:C	1:A:108:GLN:HE21	2.19	0.46
1:B:2159:LYS:HA	1:B:2159:LYS:CE	2.38	0.46
1:B:2020:LYS:NZ	1:B:2020:LYS:HB3	2.31	0.46
1:B:2108:GLN:OE1	1:B:2108:GLN:HA	2.15	0.46
1:A:202:ALA:HB1	1:A:205:LYS:HG3	1.97	0.45
1:B:2181:LYS:HE2	1:B:2204:ASP:OD2	2.16	0.45
1:B:2176:ASN:ND2	1:B:2178:ASP:H	2.15	0.45
1:B:2077:VAL:O	1:B:2108:GLN:NE2	2.50	0.44
1:B:2108:GLN:HG3	1:B:2109:ASP:N	2.33	0.44
1:A:102:PRO:HG2	1:A:103:GLU:OE2	2.19	0.43
1:B:2073:ALA:O	1:B:2077:VAL:HG23	2.19	0.42
1:A:44:LYS:HB3	1:A:44:LYS:HE2	1.83	0.42
1:A:52:ASN:HD22	1:A:52:ASN:C	2.23	0.42
1:A:8:ARG:HD3	1:A:198:TYR:CZ	2.55	0.42
1:A:3:LYS:HD3	1:A:177:SER:HB3	2.01	0.41
1:A:103:GLU:H	1:A:103:GLU:CD	2.24	0.41
1:A:72:ARG:HA	1:A:72:ARG:HD2	1.79	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:2178:ASP:C	1:B:2181:LYS:HD2	2.42	0.40

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	219/221 (99%)	210 (96%)	9 (4%)	0	100	100
1	B	219/221 (99%)	213 (97%)	6 (3%)	0	100	100
All	All	438/442 (99%)	423 (97%)	15 (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	A	179/179 (100%)	170 (95%)	9 (5%)	24	6
1	B	179/179 (100%)	171 (96%)	8 (4%)	27	8
All	All	358/358 (100%)	341 (95%)	17 (5%)	26	7

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	10	LYS
1	A	52	ASN
1	A	60	GLN
1	A	89	TYR
1	A	108	GLN
1	A	156	CYS
1	A	159	LYS
1	A	179	ASN
1	A	192	GLU
1	B	2003	LYS
1	B	2052	ASN
1	B	2060	GLN
1	B	2089	TYR
1	B	2118	LYS
1	B	2156	CYS
1	B	2176	ASN
1	B	2184	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	60	GLN
1	A	63	GLN
1	A	66	ASN
1	A	104	ASN
1	A	108	GLN
1	A	140	HIS
1	A	144	GLN
1	A	176	ASN
1	B	2060	GLN
1	B	2063	GLN
1	B	2104	ASN
1	B	2176	ASN
1	B	2219	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	SO4	A	467[A]	-	4,4,4	2.39	2 (50%)	6,6,6	0.10	0
2	0IW	B	2280	1	44,44,44	2.25	5 (11%)	56,59,59	2.01	6 (10%)
3	SO4	A	467[B]	-	4,4,4	2.37	2 (50%)	6,6,6	0.13	0
2	0IW	A	1280	1	44,44,44	2.46	6 (13%)	56,59,59	1.96	10 (17%)

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
2	0IW	B	2280	1	-	6/37/47/47	0/4/4/4
2	0IW	A	1280	1	-	9/37/47/47	0/4/4/4

All (15) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1280	0IW	O11-S1	10.93	1.60	1.44
2	B	2280	0IW	O11-S1	10.68	1.59	1.44
2	A	1280	0IW	O2-S1	9.81	1.58	1.44

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2280	0IW	O2-S1	8.65	1.56	1.44
2	A	1280	0IW	C6-N1	3.30	1.52	1.47
3	A	467[B]	SO4	O1-S	3.09	1.62	1.46
2	A	1280	0IW	C11-S1	3.08	1.84	1.77
3	A	467[A]	SO4	O2-S	2.93	1.61	1.46
3	A	467[A]	SO4	O1-S	2.89	1.61	1.46
3	A	467[B]	SO4	O2-S	2.79	1.61	1.46
2	B	2280	0IW	C11-S1	2.60	1.83	1.77
2	B	2280	0IW	C6-N1	2.60	1.51	1.47
2	B	2280	0IW	C2-N1	2.47	1.51	1.47
2	A	1280	0IW	C21-C11	2.27	1.54	1.52
2	A	1280	0IW	C32-C41	2.11	1.42	1.38

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	2280	0IW	O2-S1-O11	-10.31	106.39	118.44
2	A	1280	0IW	O2-S1-O11	-8.88	108.06	118.44
2	A	1280	0IW	O11-S1-C11	7.24	121.20	108.20
2	B	2280	0IW	O11-S1-C11	7.04	120.85	108.20
2	A	1280	0IW	O2-S1-C41	-3.50	104.81	108.36
2	B	2280	0IW	O2-S1-C41	-3.00	105.31	108.36
2	B	2280	0IW	O2-S1-C11	2.98	113.56	108.20
2	B	2280	0IW	O11-S1-C41	-2.81	105.51	108.36
2	A	1280	0IW	C21-C11-S1	-2.57	107.02	114.29
2	A	1280	0IW	C11-S1-C41	-2.38	101.77	105.03
2	A	1280	0IW	CM-N4-C3	-2.37	107.12	110.66
2	A	1280	0IW	C52-C41-C32	-2.24	117.32	120.44
2	A	1280	0IW	C61-C52-C41	2.18	121.22	118.95
2	A	1280	0IW	C22-C32-C41	2.09	121.13	118.95
2	B	2280	0IW	CM-N4-C5	-2.09	107.54	110.66
2	A	1280	0IW	CM-N4-C5	-2.02	107.64	110.66

There are no chirality outliers.

All (15) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	2280	0IW	C32-C41-S1-O11
2	A	1280	0IW	C32-C41-S1-C11
2	A	1280	0IW	C52-C41-S1-C11
2	A	1280	0IW	S1-C11-C21-C31
2	B	2280	0IW	S1-C11-C21-C31

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	B	2280	0IW	C52-C41-S1-O11
2	A	1280	0IW	C52-C41-S1-O11
2	B	2280	0IW	C32-C41-S1-C11
2	A	1280	0IW	C32-C41-S1-O2
2	A	1280	0IW	CA-CB-CG-CD1
2	A	1280	0IW	CA-CB-CG-CD2
2	B	2280	0IW	CA-CB-CG-CD1
2	B	2280	0IW	CA-CB-CG-CD2
2	A	1280	0IW	C32-C41-S1-O11
2	A	1280	0IW	C21-C11-S1-O11

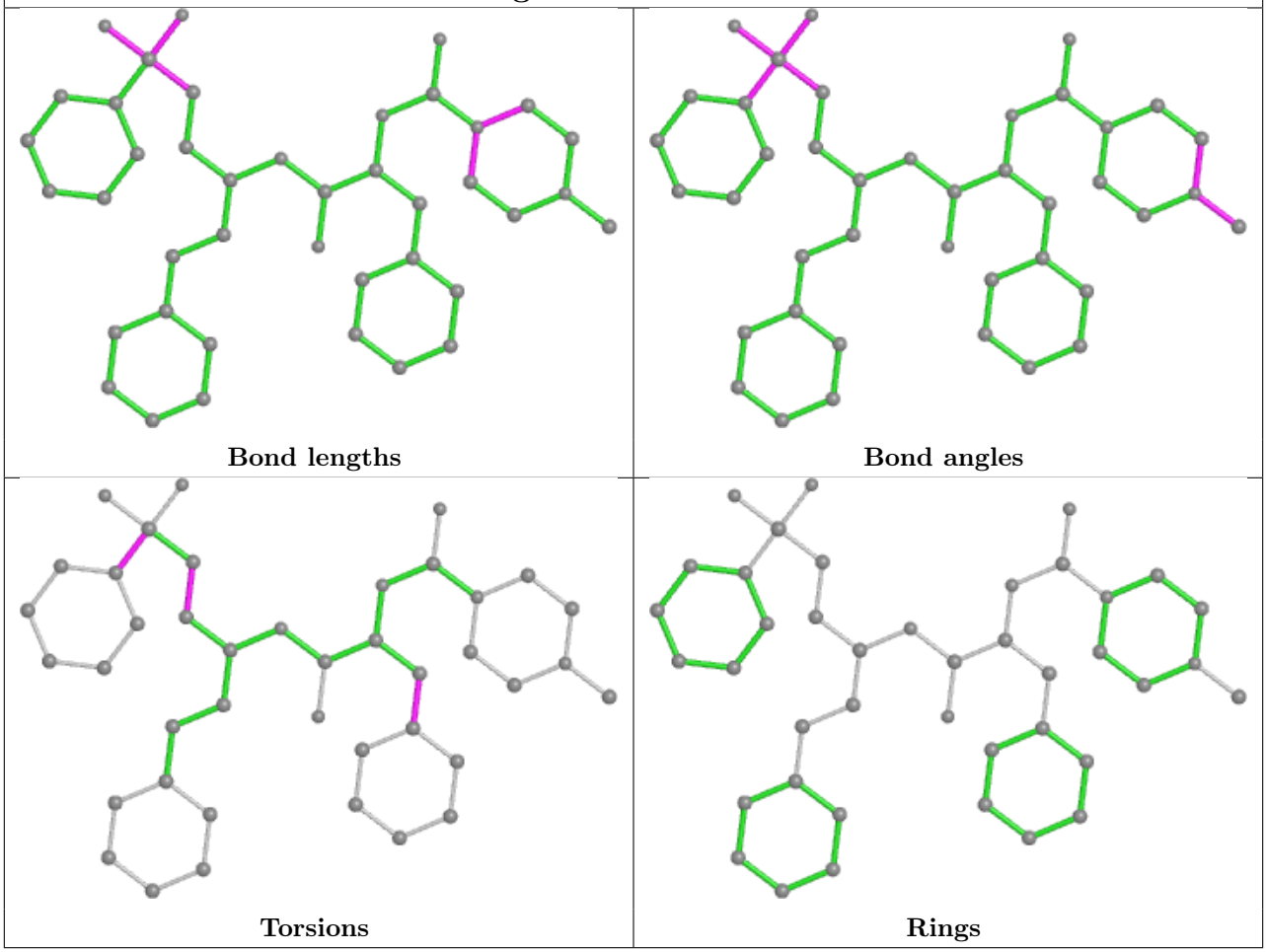
There are no ring outliers.

1 monomer is involved in 1 short contact:

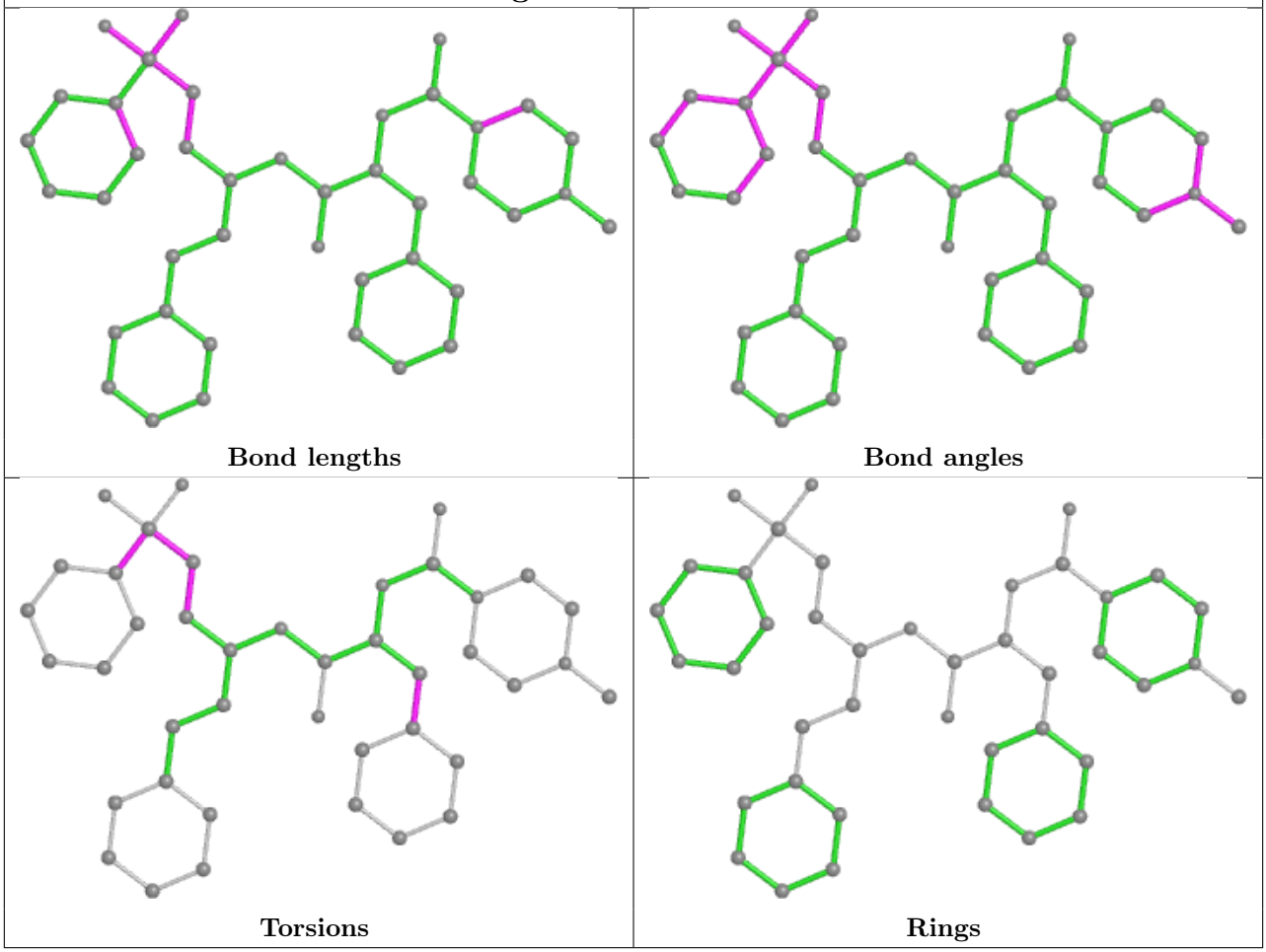
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1280	0IW	1	0

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.

Ligand 0IW B 2280



Ligand 0IW A 1280



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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.