wwPDB NMR Structure Validation Summary Report

May 2, 2022 – 12:25 PM EDT

PDB ID	:	7RNO
Title	:	Model of the Ac-6-FP/hpMR1/bB2m/TAPBPR complex from integrated docking, NMR and restrained MD
Authors	:	McShan, A.C.; Sgourakis, N.G.
Deposited on	:	2021-07-29

This is a wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

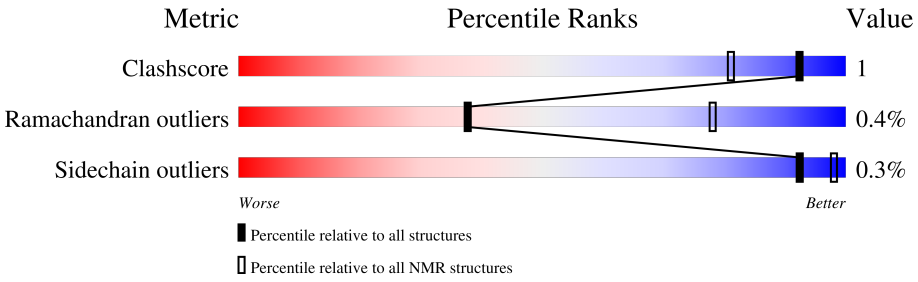
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.28.1
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 1%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	271	99% ·
2	B	99	95% · ·
3	C	396	80% · 14% ·

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
4	A	30W	301	-	6

2 Ensemble composition and analysis

This entry contains 10 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:4-A:271, B:3-B:99, C:10-C:21, C:37-C:47, C:70-C:144, C:150-C:290, C:295-C:384 (694)	1.25	10

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 6 single-model clusters were found.

Cluster number	Models
1	7, 10
2	6, 8
Single-model clusters	1; 2; 3; 4; 5; 9

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 11702 atoms, of which 5762 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Major histocompatibility complex class I-related gene protein.

Mol	Chain	Residues	Atoms						Trace
1	A	268	Total	C	H	N	O	S	0
			4303	1416	2088	379	411	9	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q95460

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms						Trace
2	B	97	Total	C	H	N	O	S	0
			1606	522	790	141	151	2	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP P01888

- Molecule 3 is a protein called TAP binding protein-like variant.

Mol	Chain	Residues	Atoms						Trace
3	C	383	Total	C	H	N	O	S	0
			5769	1818	2877	494	565	15	

There are 14 discrepancies between the modelled and reference sequences:

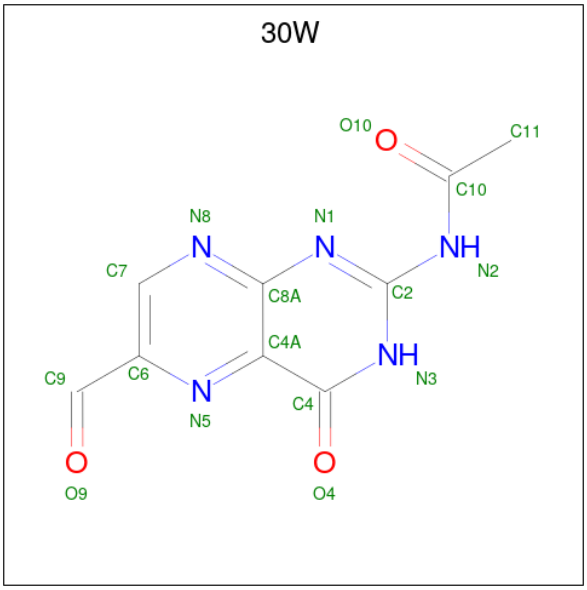
Chain	Residue	Modelled	Actual	Comment	Reference
C	30	3X9	LEU	conflict	UNP Q53GH5
C	120	TRP	ARG	conflict	UNP Q53GH5
C	385	LEU	-	expression tag	UNP Q53GH5
C	386	GLU	-	expression tag	UNP Q53GH5
C	387	GLY	-	expression tag	UNP Q53GH5
C	388	GLY	-	expression tag	UNP Q53GH5
C	389	LEU	-	expression tag	UNP Q53GH5
C	390	GLU	-	expression tag	UNP Q53GH5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	391	VAL	-	expression tag	UNP Q53GH5
C	392	LEU	-	expression tag	UNP Q53GH5
C	393	PHE	-	expression tag	UNP Q53GH5
C	394	GLN	-	expression tag	UNP Q53GH5
C	395	GLY	-	expression tag	UNP Q53GH5
C	396	PRO	-	expression tag	UNP Q53GH5

- Molecule 4 is N-(6-formyl-4-oxo-3,4-dihydropteridin-2-yl)acetamide (three-letter code: 30W) (formula: $C_9H_7N_5O_3$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
4	A	1	24	9	7	5	3

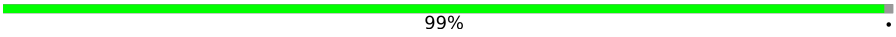
4 Residue-property plots

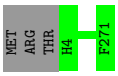
4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Major histocompatibility complex class I-related gene protein

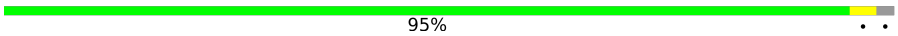
Chain A:

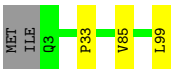




• Molecule 2: Beta-2-microglobulin

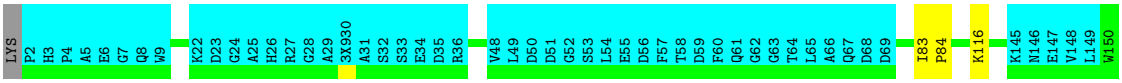
Chain B:





• Molecule 3: TAP binding protein-like variant

Chain C:

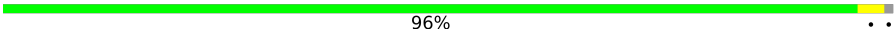


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 10. Colouring as in section 4.1 above.

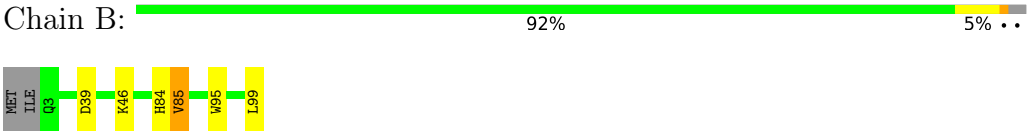
• Molecule 1: Major histocompatibility complex class I-related gene protein

Chain A:



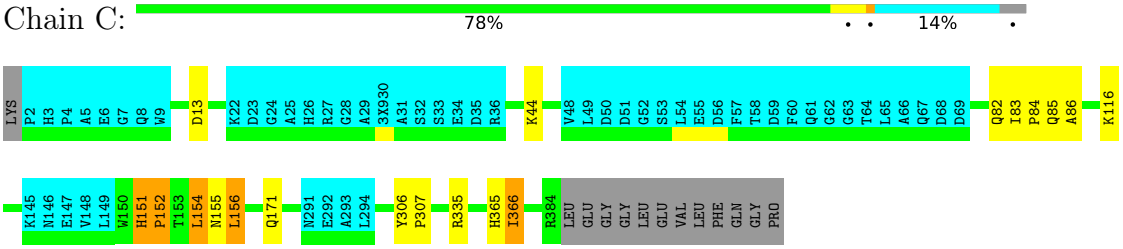
• Molecule 2: Beta-2-microglobulin

Chain B:



• Molecule 3: TAP binding protein-like variant

Chain C:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *molecular dynamics*.

Of the 10000 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
GROMACS	refinement	v2020.4
HADDOCK	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	204
Number of shifts mapped to atoms	204
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	1%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 3X9, 30W

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 (average) 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	0.68±0.01	0±0/2285 (0.0± 0.0%)	0.46±0.02	0±1/3106 (0.0± 0.0%)
2	B	0.65±0.00	0±0/842 (0.0± 0.0%)	0.43±0.00	0±0/1140 (0.0± 0.0%)
3	C	0.60±0.00	0±0/2521 (0.0± 0.0%)	0.48±0.15	0±1/3441 (0.0± 0.0%)
All	All	0.64	0/56480 (0.0%)	0.48	9/76870 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.3±0.5
3	C	0.0±0.0	1.5±0.7
All	All	0	18

There are no bond-length outliers.

5 of 8 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
3	C	70	PRO	C-N-CD	-48.43	14.05	120.60	2	1
1	A	80	ARG	NE-CZ-NH1	8.74	124.67	120.30	2	1
1	A	139	THR	CB-CA-C	8.56	134.71	111.60	9	1
1	A	139	THR	N-CA-CB	-7.69	95.69	110.30	9	1
1	A	144	TRP	CA-CB-CG	-6.74	100.89	113.70	6	1

There are no chirality outliers.

5 of 8 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
3	C	151	HIS	Peptide	9
3	C	307	PRO	Peptide	2
3	C	154	LEU	Peptide	2
1	A	80	ARG	Sidechain	1
3	C	117	THR	Peptide	1

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	2215	2088	2087	4±2
2	B	816	790	789	3±2
3	C	2472	2502	2502	9±4
All	All	55200	53870	53850	153

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.

5 of 99 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:190:LYS:NZ	2:B:99:LEU:O	1.00	1.95	8	1
1:A:202:ARG:NE	2:B:99:LEU:O	0.98	1.96	10	1
3:C:155:ASN:O	3:C:156:LEU:CB	0.74	2.35	2	2
3:C:155:ASN:O	3:C:156:LEU:HB2	0.69	1.88	2	3
1:A:128:LEU:HG	1:A:128:LEU:O	0.66	1.90	8	3

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	266/271 (98%)	260±2 (98±1%)	6±2 (2±1%)	0±0 (0±0%)	54 85
2	B	95/99 (96%)	91±1 (96±2%)	3±1 (3±1%)	1±1 (1±1%)	29 74
3	C	328/396 (83%)	316±2 (96±1%)	10±2 (3±1%)	2±1 (1±0%)	29 74
All	All	6890/7660 (90%)	6674 (97%)	186 (3%)	30 (0%)	38 78

5 of 14 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	C	156	LEU	5
2	B	85	VAL	5
3	C	116	LYS	4
3	C	155	ASN	3
3	C	366	ILE	3

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	237/240 (99%)	237±1 (100±0%)	0±1 (0±0%)	93 98
2	B	93/95 (98%)	93±0 (100±0%)	0±0 (0±0%)	92 98
3	C	274/325 (84%)	273±1 (99±0%)	1±1 (1±0%)	89 97
All	All	6040/6600 (92%)	6020 (100%)	20 (0%)	92 98

5 of 10 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	C	335	ARG	9
3	C	210	HIS	2
2	B	95	TRP	2
3	C	80	LEU	1
3	C	110	MET	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 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 for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
3	3X9	C	30	3	15,18,19	1.25±0.38	2±1 (12±8%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
3	3X9	C	30	3	12,27,29	2.91±2.06	3±2 (24±14%)

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	3X9	C	30	3	-	0±0,5,32,34	0±0,1,1,1

5 of 6 unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
3	C	30	3X9	CAJ-SAL	6.98	1.90	1.81	9	2
3	C	30	3X9	CAS-CAO	3.38	1.46	1.51	6	2
3	C	30	3X9	OAH-NAQ	3.00	1.27	1.43	6	10
3	C	30	3X9	CAI-CAO	2.62	1.30	1.32	6	2
3	C	30	3X9	CB-CA	2.33	1.59	1.53	6	1

5 of 10 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
3	C	30	3X9	CAJ-SAL-SG	18.94	124.78	103.67	6	7
3	C	30	3X9	CAA-CAR-CAI	14.66	127.75	112.79	6	2
3	C	30	3X9	CAB-CAR-CAI	9.70	102.90	112.79	6	3
3	C	30	3X9	CAI-CAR-NAQ	8.03	93.19	99.37	6	4
3	C	30	3X9	CA-CB-SG	4.61	134.24	114.55	6	1

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
4	30W	A	301	-	18,18,18	1.29±0.01	2±0 (11±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
4	30W	A	301	-	22,25,25	3.44±0.01	14±0 (63±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
4	30W	A	301	-	-	0±0,6,6,6	0±0,2,2,2

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
4	A	301	30W	C4A-C4	3.91	1.48	1.41	6	10
4	A	301	30W	C8A-N8	2.39	1.33	1.37	2	10

5 of 14 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
4	A	301	30W	C2-N2-C10	6.77	120.03	130.28	4	10
4	A	301	30W	C2-N1-C8A	5.81	121.88	115.28	2	10
4	A	301	30W	C9-C6-N5	5.47	120.02	114.66	4	10
4	A	301	30W	N8-C8A-N1	4.92	121.44	115.82	8	10
4	A	301	30W	O9-C9-C6	4.57	119.89	124.22	6	10

There are no chirality outliers.

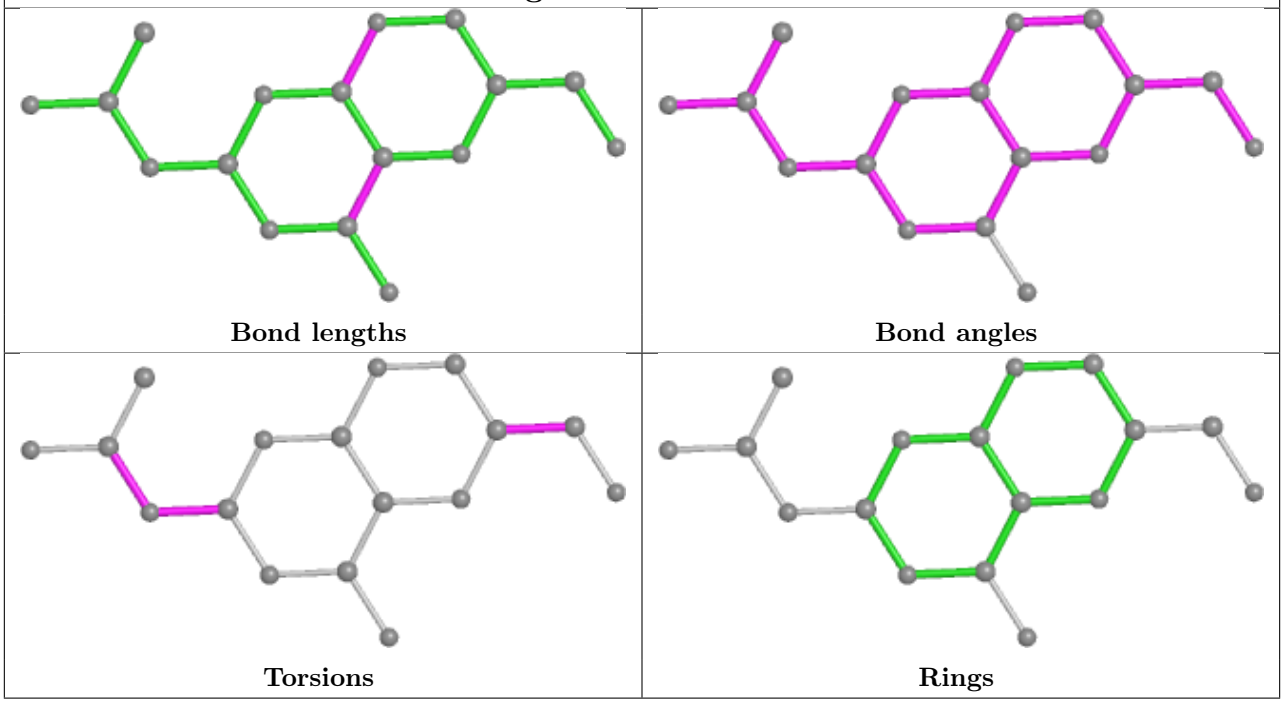
There are no torsion outliers.

There are no ring outliers.

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 30W A 301



6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 1% for the well-defined parts and 1% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_2*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	204
Number of shifts mapped to atoms	204
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 1%, i.e. 102 atoms were assigned a chemical shift out of a possible 8531. 0 out of 117 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/3386 (0%)	0/1346 (0%)	0/1388 (0%)	0/652 (0%)
Sidechain	102/4359 (2%)	51/2565 (2%)	51/1605 (3%)	0/189 (0%)
Aromatic	0/786 (0%)	0/415 (0%)	0/334 (0%)	0/37 (0%)
Overall	102/8531 (1%)	51/4326 (1%)	51/3327 (2%)	0/878 (0%)

7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chemical_shifts_2). RCI is only applicable to proteins.