Full wwPDB NMR Structure Validation Report (i)

Feb 22, 2022 – 12:05 PM EST

PDB ID	:	7SFQ
Title	:	EmrE S64V Mutant Bound to tetra(4-fluorophenyl)phosphonium at pH 8.0
Authors	:	Shcherbakov, A.A.; Spreacker, P.J.; Dregni, A.J.; Henzler-Wildman, K.A.; Hong, M.
Deposited on	:	2021-10-04

This is a Full wwPDB NMR 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/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:

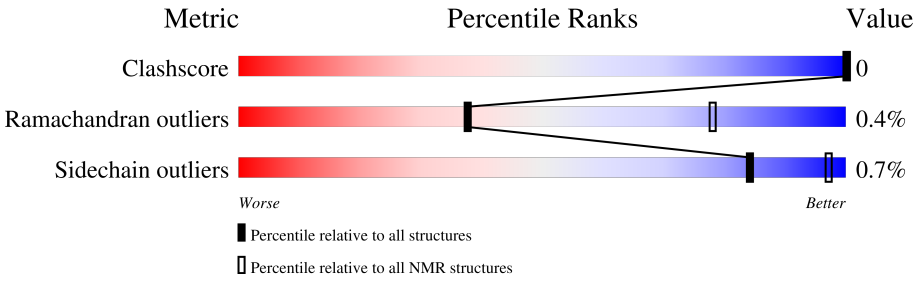
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 30%.

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	110	88% • • 7%
1	B	110	85% • 7% 7%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:28, A:33-A:105, B:4-B:24, B:33-B:105 (192)	1.27	7

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. No single-model clusters were found.

Cluster number	Models
1	5, 6, 7, 9, 10
2	1, 2, 3, 4, 8

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3237 atoms, of which 1652 are hydrogens and 0 are deuteriums.

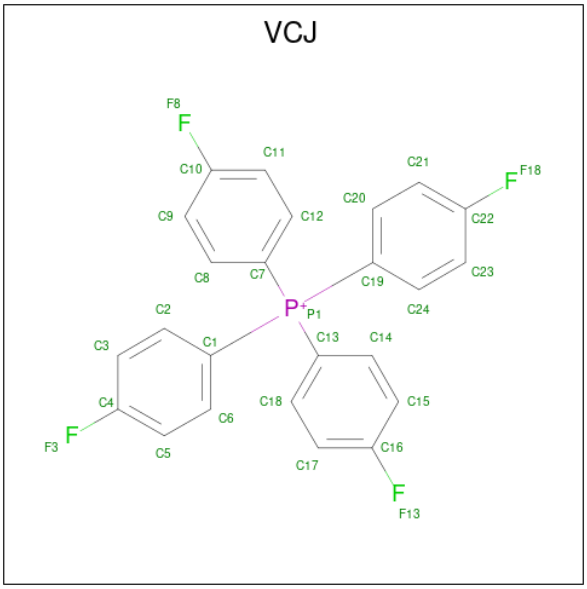
- Molecule 1 is a protein called Multidrug transporter EmrE.

Mol	Chain	Residues	Atoms						Trace
1	A	102	Total	C	H	N	O	S	0
			1596	528	818	116	128	6	
1	B	102	Total	C	H	N	O	S	0
			1596	528	818	116	128	6	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	64	VAL	SER	engineered mutation	UNP P23895
B	64	VAL	SER	engineered mutation	UNP P23895

- Molecule 2 is tetrakis(4-fluorophenyl)phosphanium (three-letter code: VCJ) (formula: $C_{24}H_{16}F_4P$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				
2	A	1	Total	C	F	H	P
			45	24	4	16	1

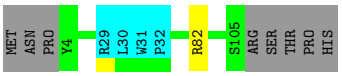
4 Residue-property plots

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.

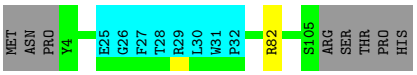
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

Chain B:



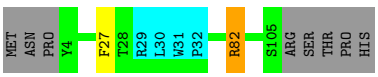
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

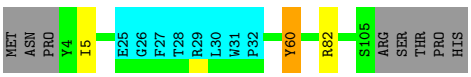
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

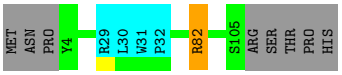
Chain B:



4.2.2 Score per residue for model 2

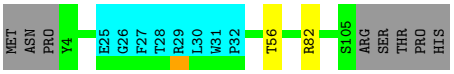
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

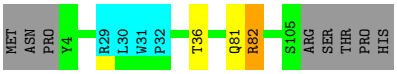
Chain B:



4.2.3 Score per residue for model 3

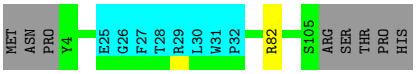
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

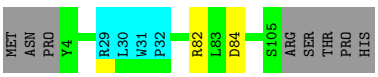
Chain B:



4.2.4 Score per residue for model 4

• Molecule 1: Multidrug transporter EmrE

Chain A:



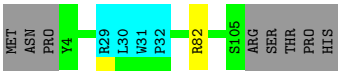
• Molecule 1: Multidrug transporter EmrE

Chain B:

4.2.5 Score per residue for model 5

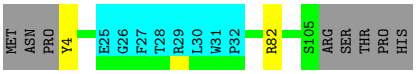
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

Chain B:



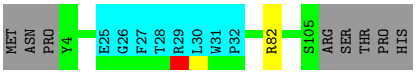
4.2.6 Score per residue for model 6

• Molecule 1: Multidrug transporter EmrE

Chain A:

• Molecule 1: Multidrug transporter EmrE

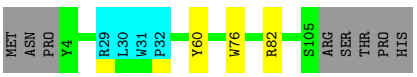
Chain B:



4.2.7 Score per residue for model 7 (medoid)

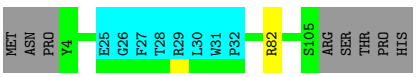
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

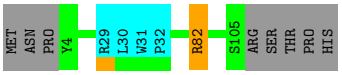
Chain B:



4.2.8 Score per residue for model 8

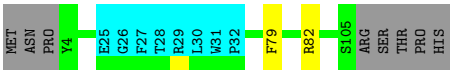
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

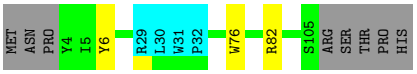
Chain B:



4.2.9 Score per residue for model 9

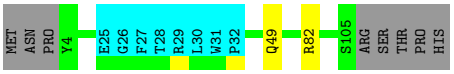
• Molecule 1: Multidrug transporter EmrE

Chain A:



• Molecule 1: Multidrug transporter EmrE

Chain B:



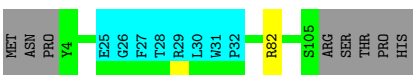
4.2.10 Score per residue for model 10

• Molecule 1: Multidrug transporter EmrE

Chain A:

• Molecule 1: Multidrug transporter EmrE

Chain B:



5 Refinement protocol and experimental data overview

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

Of the 82 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
HADDOCK	structure calculation	Webserver 2.4
GROMACS	refinement	2019.1

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	724
Number of shifts mapped to atoms	724
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	30%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: VCJ

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	0.71±0.01	0±0/755 (0.0± 0.0%)	0.95±0.04	2±1/1030 (0.2± 0.1%)
1	B	0.70±0.01	0±0/723 (0.0± 0.0%)	0.91±0.03	2±1/987 (0.2± 0.1%)
All	All	0.71	0/14780 (0.0%)	0.93	38/20170 (0.2%)

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	0.2±0.4
All	All	0	2

There are no bond-length outliers.

All 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
1	A	82	ARG	NE-CZ-NH1	11.55	126.08	120.30	9	10
1	A	82	ARG	NE-CZ-NH2	-10.62	114.99	120.30	9	7
1	B	82	ARG	NE-CZ-NH1	8.29	124.44	120.30	7	10
1	A	60	TYR	CB-CG-CD1	-6.72	116.97	121.00	7	3
1	A	60	TYR	CB-CG-CD2	6.63	124.98	121.00	7	1
1	B	60	TYR	CB-CG-CD1	-6.19	117.28	121.00	1	1
1	B	60	TYR	CB-CG-CD2	6.10	124.66	121.00	1	1
1	B	82	ARG	NE-CZ-NH2	-5.86	117.37	120.30	9	1
1	A	6	TYR	CB-CG-CD2	-5.61	117.64	121.00	9	1
1	A	6	TYR	CB-CG-CD1	5.45	124.27	121.00	9	1
1	B	4	TYR	CB-CG-CD2	-5.40	117.76	121.00	5	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	79	PHE	CB-CG-CD1	5.00	124.30	120.80	8	1

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	82	ARG	Sidechain	1
1	A	4	TYR	Sidechain	1

6.2 Too-close contacts

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
All	All	14740	15450	15270	-

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 -.

There are no clashes.

6.3 Torsion angles

6.3.1 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 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	96/110 (87%)	93±1 (97±1%)	2±1 (3±1%)	1±0 (1±1%)	29	74
1	B	92/110 (84%)	90±1 (97±1%)	2±1 (2±1%)	0±0 (0±0%)	54	85
All	All	1880/2200 (85%)	1825 (97%)	48 (3%)	7 (0%)	38	78

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	82	ARG	3
1	A	76	TRP	3
1	B	5	ILE	1

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	77/89 (87%)	76±1 (99±1%)	1±1 (1±1%)	82 97
1	B	74/89 (83%)	74±0 (99±1%)	0±0 (1±1%)	89 97
All	All	1510/1780 (85%)	1500 (99%)	10 (1%)	84 97

All 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)
1	A	27	PHE	1
1	B	60	TYR	1
1	B	56	THR	1
1	A	36	THR	1
1	A	81	GLN	1
1	A	84	ASP	1
1	B	5	ILE	1
1	A	49	GLN	1
1	B	49	GLN	1
1	A	44	PHE	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)

There are no non-standard protein/DNA/RNA residues in this entry.

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
2	VCJ	A	201	-	32,32,32	1.43±0.05	4±1 (11±3%)

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
2	VCJ	A	201	-	46,46,46	0.82±0.10	1±1 (1±1%)

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	VCJ	A	201	-	-	0±0,24,24,24	0±0,4,4,4

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
2	A	201	VCJ	P1-C19	3.08	1.85	1.79	10	9
2	A	201	VCJ	P1-C13	3.05	1.85	1.79	2	8
2	A	201	VCJ	P1-C1	2.99	1.85	1.79	5	8
2	A	201	VCJ	P1-C7	2.92	1.85	1.79	6	7
2	A	201	VCJ	C8-C7	2.07	1.43	1.39	6	1
2	A	201	VCJ	C24-C19	2.04	1.43	1.39	8	1
2	A	201	VCJ	C12-C7	2.03	1.43	1.39	3	1
2	A	201	VCJ	C20-C19	2.03	1.43	1.39	4	2

All 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
2	A	201	VCJ	P1-C19-C24	3.30	125.94	120.05	2	1
2	A	201	VCJ	P1-C1-C6	3.13	125.65	120.05	5	2
2	A	201	VCJ	P1-C1-C2	2.45	124.42	120.05	2	1
2	A	201	VCJ	P1-C7-C8	2.10	123.80	120.05	7	1
2	A	201	VCJ	P1-C13-C14	2.03	123.68	120.05	8	1

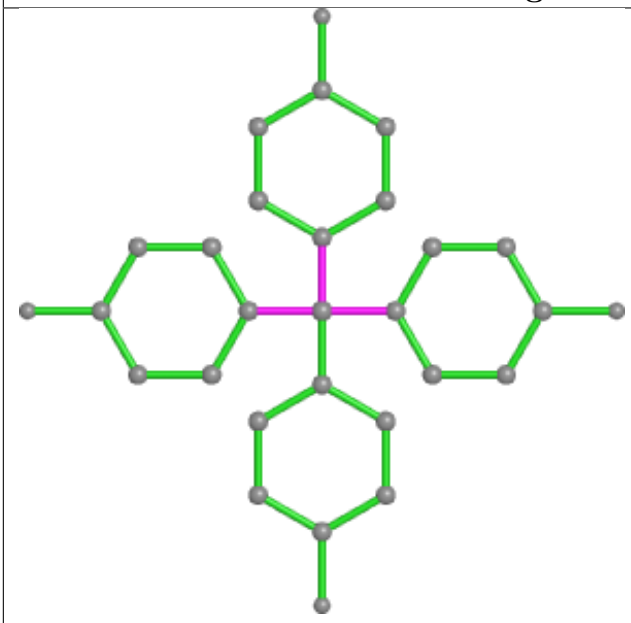
There are no chirality outliers.

There are no torsion outliers.

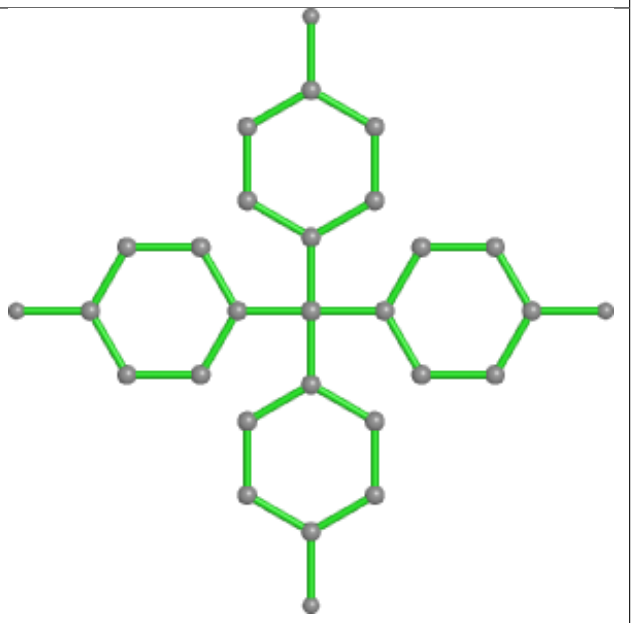
There are no ring outliers.

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.

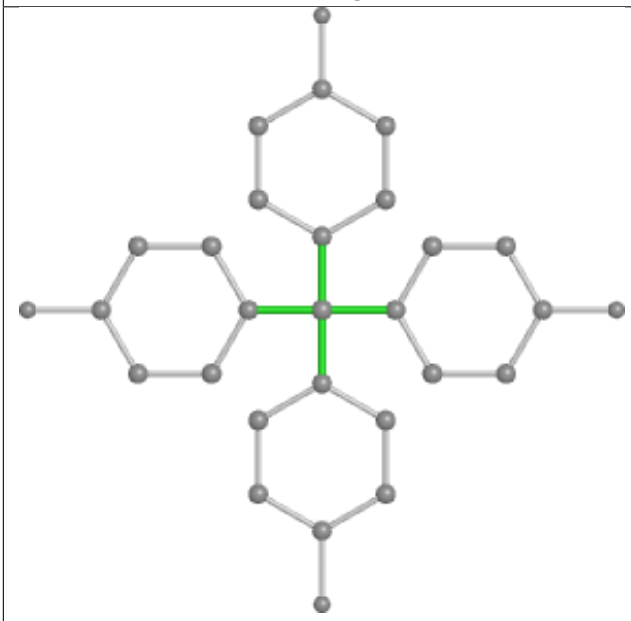
Ligand VCJ A 201



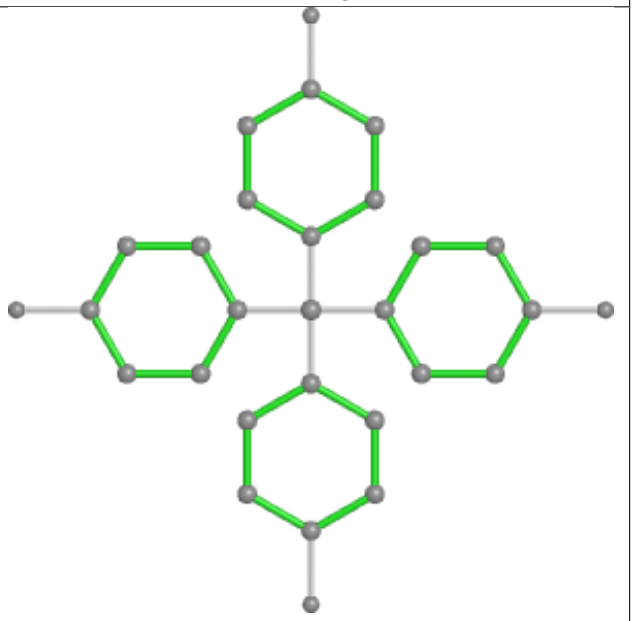
Bond lengths



Bond angles



Torsions



Rings

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 30% for the well-defined parts and 30% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *EmrE_AShifts_Star.str*

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	724
Number of shifts mapped to atoms	724
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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	162	-1.01 ± 0.17	Should be applied
$^{13}C_{\beta}$	116	1.39 ± 0.17	Should be applied
$^{13}C'$	160	-0.43 ± 0.18	None needed (< 0.5 ppm)
^{15}N	143	1.42 ± 0.34	Should be applied

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 30%, i.e. 675 atoms were assigned a chemical shift out of a possible 2235. 0 out of 42 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	567/952 (60%)	134/380 (35%)	299/384 (78%)	134/188 (71%)
Sidechain	108/1050 (10%)	0/602 (0%)	108/434 (25%)	0/14 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/233 (0%)	0/121 (0%)	0/106 (0%)	0/6 (0%)
Overall	675/2235 (30%)	134/1103 (12%)	407/924 (44%)	134/208 (64%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 30%, i.e. 724 atoms were assigned a chemical shift out of a possible 2416. 0 out of 44 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	608/1008 (60%)	143/402 (36%)	322/408 (79%)	143/198 (72%)
Sidechain	116/1142 (10%)	0/658 (0%)	116/464 (25%)	0/20 (0%)
Aromatic	0/266 (0%)	0/138 (0%)	0/120 (0%)	0/8 (0%)
Overall	724/2416 (30%)	143/1198 (12%)	438/992 (44%)	143/226 (63%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

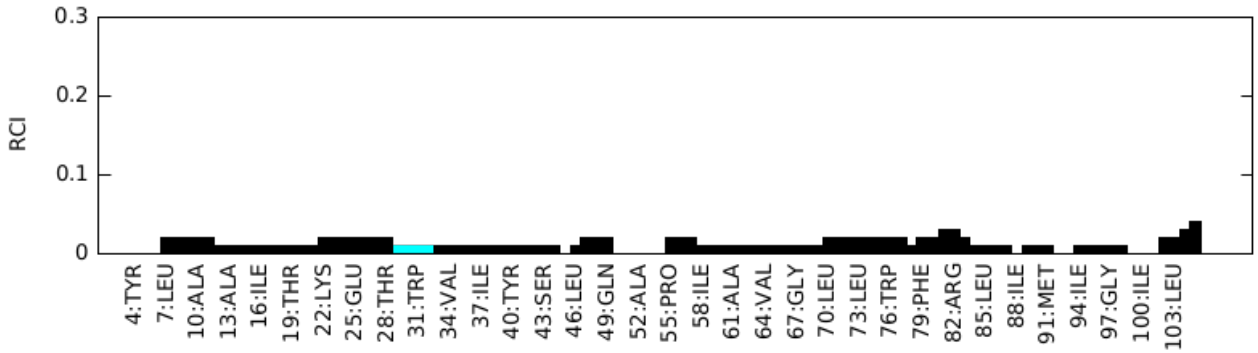
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B: