



# Full wwPDB NMR Structure Validation Report ⓘ

Jun 22, 2022 – 01:11 pm BST

PDB ID : 7ZRF  
Title : Spatial structure of amyloidogenic SEM1(68-85) peptide  
Authors : Blokhin, D.S.; Sanchugova, D.A.  
Deposited on : 2022-05-04

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ 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
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	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

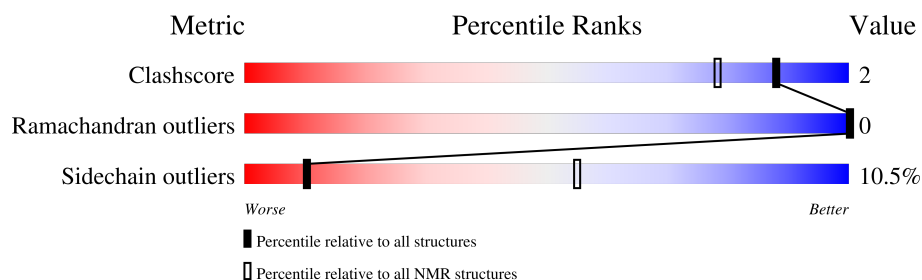
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment was not calculated.

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	18	

## 2 Ensemble composition and analysis

This entry contains 9 models. Model 6 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:68-A:85 (18)	0.89	6

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 4 single-model clusters were found.

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

### 3 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 290 atoms, of which 135 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Semenogelin-1.

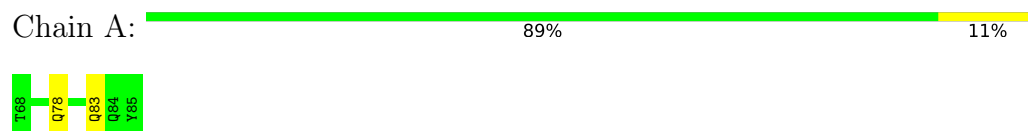
Mol	Chain	Residues	Atoms					Trace
1	A	18	Total	C	H	N	O	0
			290	91	135	30	34	

## 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: Semenogelin-1

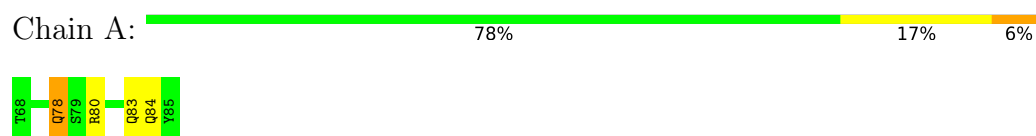


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

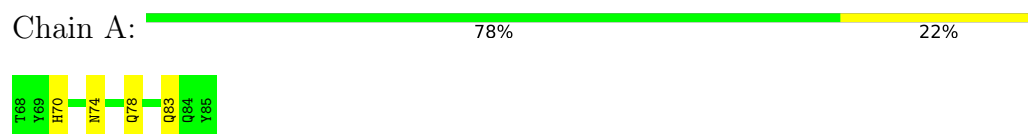
#### 4.2.1 Score per residue for model 1

- Molecule 1: Semenogelin-1




#### 4.2.2 Score per residue for model 2

- Molecule 1: Semenogelin-1



### 4.2.3 Score per residue for model 3

- Molecule 1: Semenogelin-1

Chain A:  89% 11%



### 4.2.4 Score per residue for model 4

- Molecule 1: Semenogelin-1

Chain A:  100%

There are no outlier residues in this chain.

### 4.2.5 Score per residue for model 5


- Molecule 1: Semenogelin-1

Chain A:  94% 6%



### 4.2.6 Score per residue for model 6 (medoid)


- Molecule 1: Semenogelin-1

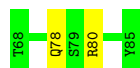
Chain A:  78% 22%



### 4.2.7 Score per residue for model 7

- Molecule 1: Semenogelin-1

Chain A:  89% 11%



#### 4.2.8 Score per residue for model 8


- Molecule 1: Semenogelin-1

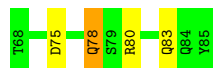
Chain A:  94% 6%



#### 4.2.9 Score per residue for model 9

- Molecule 1: Semenogelin-1

Chain A:  78% 17% 6%



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 50 calculated structures, 9 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
CNS	refinement	
X-PLOR NIH	structure calculation	3.2

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	173
Number of shifts mapped to atoms	0
Number of unparsed shifts	173
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	0%



## 6 Model quality [i](#)

### 6.1 Standard geometry [i](#)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	155	135	130	1±0
All	All	1395	1215	1170	5

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 2.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:70:HIS:HA	1:A:83:GLN:OE1	0.54	2.02	2	1
1:A:70:HIS:HB2	1:A:82:SER:CB	0.52	2.34	6	1
1:A:83:GLN:NE2	1:A:83:GLN:H	0.44	2.10	3	1
1:A:75:ASP:HA	1:A:78:GLN:OE1	0.41	2.15	9	1
1:A:78:GLN:CD	1:A:78:GLN:H	0.40	2.20	1	1

### 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	16/18 (89%)	14±1 (85±9%)	2±1 (15±9%)	0±0 (0±0%)	100	100
All	All	144/162 (89%)	123 (85%)	21 (15%)	0 (0%)	100	100

There are no Ramachandran outliers.

### 6.3.2 Protein sidechains ⓘ

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	17/17 (100%)	15±1 (90±7%)	2±1 (10±7%)	10	55
All	All	153/153 (100%)	137 (90%)	16 (10%)	10	55

All 6 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	78	GLN	6
1	A	80	ARG	3
1	A	83	GLN	3
1	A	84	GLN	2
1	A	74	ASN	1
1	A	79	SER	1

### 6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 6.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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 0% for the well-defined parts and 0% for the entire structure.

### 7.1 Chemical shift list 1

File name: working\_cs.cif

Chemical shift list name: *assigned\_chemical\_shifts\_1*

#### 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	173
Number of shifts mapped to atoms	0
Number of unparsed shifts	173
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following errors were found when reading this chemical shift list.

- Entity instance (chain) must be specified. All 173 occurrences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	?	1	THR	HA	3.736	0.003	1
2	?	1	THR	HB	3.987	0.008	1
3	?	1	THR	HG21	1.152	0.004	1
4	?	1	THR	HG22	1.152	0.004	1
5	?	1	THR	HG23	1.152	0.004	1
6	?	1	THR	C	170.315	0.01	1
7	?	1	THR	CA	61.336	0.043	1
8	?	1	THR	CB	69.078	0.042	1
9	?	1	THR	CG2	21.423	0.045	1
10	?	2	TYR	H	8.681	0.002	1
11	?	2	TYR	HA	4.483	0.003	1
12	?	2	TYR	HB2	2.904	0.004	2
13	?	2	TYR	HB3	2.827	0.003	2
14	?	2	TYR	C	174.733	?	1
15	?	2	TYR	CA	58.177	?	1

*Continued on next page...*

*Continued from previous page...*

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
16	?	2	TYR	CB	38.969	0.005	1
17	?	2	TYR	N	125.237	?	1
18	?	3	HIS	H	8.258	0.001	1
19	?	3	HIS	HA	4.513	0.001	1
20	?	3	HIS	HB2	3.043	0.004	2
21	?	3	HIS	HB3	2.950	0.004	2
22	?	3	HIS	C	173.206	0.013	1
23	?	3	HIS	CA	54.901	?	1
24	?	3	HIS	CB	29.567	0.001	1
25	?	3	HIS	N	122.896	?	1
26	?	4	VAL	H	8.113	0.003	1
27	?	4	VAL	HA	3.886	0.004	1
28	?	4	VAL	HB	1.896	0.004	1
29	?	4	VAL	HG11	0.829	0.011	1
30	?	4	VAL	HG12	0.829	0.011	1
31	?	4	VAL	HG13	0.829	0.011	1
32	?	4	VAL	C	175.462	0.005	1
33	?	4	VAL	CA	62.244	?	1
34	?	4	VAL	CB	32.846	0.042	1
35	?	4	VAL	CG1	20.601	0.068	1
36	?	4	VAL	N	122.655	?	1
37	?	5	ASP	H	8.488	0.003	1
38	?	5	ASP	HA	4.572	0.004	1
39	?	5	ASP	HB2	2.803	0.005	2
40	?	5	ASP	HB3	2.666	0.008	2
41	?	5	ASP	C	175.160	0.013	1
42	?	5	ASP	CA	53.125	0.059	1
43	?	5	ASP	CB	39.113	0.002	1
44	?	5	ASP	N	123.883	?	1
45	?	6	ALA	H	8.312	0.002	1
46	?	6	ALA	HA	4.150	0.007	1
47	?	6	ALA	HB1	1.278	0.004	1
48	?	6	ALA	HB2	1.278	0.004	1
49	?	6	ALA	HB3	1.278	0.004	1
50	?	6	ALA	C	177.445	0.015	1
51	?	6	ALA	CA	52.957	0.042	1
52	?	6	ALA	CB	19.136	0.052	1
53	?	6	ALA	N	125.399	?	1
54	?	7	ASN	H	8.246	0.002	1
55	?	7	ASN	HA	4.560	0.003	1
56	?	7	ASN	HB2	2.754	0.004	2

*Continued on next page...*

*Continued from previous page...*

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
57	?	7	ASN	HB3	2.646	0.007	2
58	?	7	ASN	HD21	7.534	0.002	1
59	?	7	ASN	HD22	6.848	0.001	1
60	?	7	ASN	C	174.964	0.016	1
61	?	7	ASN	CA	53.192	0.018	1
62	?	7	ASN	CB	38.643	0.018	1
63	?	7	ASN	N	116.719	?	1
64	?	7	ASN	ND2	112.865	0.0	1
65	?	8	ASP	H	8.077	0.001	1
66	?	8	ASP	HA	4.539	0.001	1
67	?	8	ASP	HB2	2.737	0.004	2
68	?	8	ASP	HB3	2.715	0.003	2
69	?	8	ASP	C	175.399	0.029	1
70	?	8	ASP	CA	53.401	0.001	1
71	?	8	ASP	CB	38.646	?	1
72	?	8	ASP	N	119.183	?	1
73	?	9	HIS	H	8.439	0.006	1
74	?	9	HIS	HA	4.602	0.001	1
75	?	9	HIS	HB2	3.232	0.006	2
76	?	9	HIS	HB3	3.109	0.005	2
77	?	9	HIS	C	174.311	0.04	1
78	?	9	HIS	CA	55.553	0.096	1
79	?	9	HIS	CB	28.461	0.007	1
80	?	9	HIS	N	118.728	?	1
81	?	10	ASP	H	8.282	0.002	1
82	?	10	ASP	HA	4.579	0.004	1
83	?	10	ASP	HB2	2.793	0.004	2
84	?	10	ASP	HB3	2.709	0.008	2
85	?	10	ASP	C	175.525	0.01	1
86	?	10	ASP	CA	53.671	0.016	1
87	?	10	ASP	CB	38.896	0.012	1
88	?	10	ASP	N	119.686	?	1
89	?	11	GLN	H	8.357	0.001	1
90	?	11	GLN	HA	4.255	0.008	1
91	?	11	GLN	HB2	2.062	0.004	2
92	?	11	GLN	HB3	1.903	0.007	2
93	?	11	GLN	HG2	2.291	0.006	1
94	?	11	GLN	C	176.110	0.009	1
95	?	11	GLN	CA	56.515	?	1
96	?	11	GLN	CB	29.262	0.003	1
97	?	11	GLN	CG	33.803	?	1

*Continued on next page...*

*Continued from previous page...*

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
98	?	11	GLN	N	121.044	?	1
99	?	12	SER	H	8.262	0.003	1
100	?	12	SER	HA	4.315	0.002	1
101	?	12	SER	HB2	3.795	0.004	1
102	?	12	SER	C	174.631	0.0	1
103	?	12	SER	CA	58.736	?	1
104	?	12	SER	CB	63.679	0.058	1
105	?	12	SER	N	116.666	?	1
106	?	13	ARG	H	8.198	0.002	1
107	?	13	ARG	HA	4.262	0.002	1
108	?	13	ARG	HB2	1.666	0.005	2
109	?	13	ARG	HB3	1.782	0.004	2
110	?	13	ARG	HG2	1.552	0.006	2
111	?	13	ARG	HG3	1.521	0.001	2
112	?	13	ARG	HD2	3.089	0.003	1
113	?	13	ARG	HE	7.071	0.002	1
114	?	13	ARG	C	174.738	?	1
115	?	13	ARG	CA	56.178	?	1
116	?	13	ARG	CB	30.680	0.04	1
117	?	13	ARG	CG	27.066	0.052	1
118	?	13	ARG	CD	43.367	?	1
119	?	13	ARG	N	122.797	?	1
120	?	14	LYS	H	8.251	0.003	1
121	?	14	LYS	HA	4.233	0.002	1
122	?	14	LYS	HB2	1.668	0.004	2
123	?	14	LYS	HB3	1.760	0.005	2
124	?	14	LYS	HG2	1.340	0.011	2
125	?	14	LYS	HG3	1.368	0.003	2
126	?	14	LYS	HD2	1.584	0.004	1
127	?	14	LYS	HE2	2.897	0.004	1
128	?	14	LYS	HZ1	7.422	0.002	1
129	?	14	LYS	HZ2	7.422	0.002	1
130	?	14	LYS	HZ3	7.422	0.002	1
131	?	14	LYS	C	176.612	0.009	1
132	?	14	LYS	CA	56.503	?	1
133	?	14	LYS	CB	33.007	0.001	1
134	?	14	LYS	CG	24.730	0.03	1
135	?	14	LYS	CD	29.087	?	1
136	?	14	LYS	CE	42.160	0.047	1
137	?	14	LYS	N	122.621	?	1
138	?	15	SER	H	8.222	0.002	1

*Continued on next page...*

*Continued from previous page...*

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
139	?	15	SER	HA	4.322	0.003	1
140	?	15	SER	HB2	3.794	0.001	2
141	?	15	SER	HB3	3.762	0.001	2
142	?	15	SER	C	174.261	0.013	1
143	?	15	SER	CA	58.464	?	1
144	?	15	SER	CB	63.758	0.026	1
145	?	15	SER	N	116.773	?	1
146	?	16	GLN	H	8.247	0.002	1
147	?	16	GLN	HA	4.216	0.013	1
148	?	16	GLN	HB2	1.947	0.001	2
149	?	16	GLN	HB3	1.850	0.004	2
150	?	16	GLN	HG2	2.187	0.003	1
151	?	16	GLN	C	175.234	0.005	1
152	?	16	GLN	CA	55.837	0.024	1
153	?	16	GLN	CB	29.510	0.01	1
154	?	16	GLN	CG	33.676	?	1
155	?	16	GLN	N	121.848	?	1
156	?	17	GLN	H	8.198	0.003	1
157	?	17	GLN	HA	4.171	0.003	1
158	?	17	GLN	HB2	1.889	0.004	2
159	?	17	GLN	HB3	1.798	0.005	2
160	?	17	GLN	HG2	2.146	0.021	1
161	?	17	GLN	C	174.922	0.024	1
162	?	17	GLN	CA	55.845	?	1
163	?	17	GLN	CB	29.708	0.003	1
164	?	17	GLN	CG	33.770	0.008	1
165	?	17	GLN	N	121.414	?	1
166	?	18	TYR	H	7.990	0.004	1
167	?	18	TYR	HA	4.464	0.003	1
168	?	18	TYR	HB2	3.055	0.006	2
169	?	18	TYR	HB3	2.828	0.003	2
170	?	18	TYR	C	178.414	0.034	1
171	?	18	TYR	CA	57.661	0.002	1
172	?	18	TYR	CB	38.851	0.001	1
173	?	18	TYR	N	123.258	?	1

### 7.1.2 Chemical shift referencing ⓘ

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



### 7.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 0%, i.e. 0 atoms were assigned a chemical shift out of a possible 232. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Backbone	0/90 (0%)	0/36 (0%)	0/36 (0%)	0/18 (0%)
Sidechain	0/110 (0%)	0/65 (0%)	0/37 (0%)	0/8 (0%)
Aromatic	0/32 (0%)	0/16 (0%)	0/12 (0%)	0/4 (0%)
Overall	0/232 (0%)	0/117 (0%)	0/85 (0%)	0/30 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 0%, i.e. 0 atoms were assigned a chemical shift out of a possible 232. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Backbone	0/90 (0%)	0/36 (0%)	0/36 (0%)	0/18 (0%)
Sidechain	0/110 (0%)	0/65 (0%)	0/37 (0%)	0/8 (0%)
Aromatic	0/32 (0%)	0/16 (0%)	0/12 (0%)	0/4 (0%)
Overall	0/232 (0%)	0/117 (0%)	0/85 (0%)	0/30 (0%)

### 7.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

### 7.1.5 Random Coil Index (RCI) plots [i](#)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned\_chemical\_shifts\_1). RCI is only applicable to proteins.