Full wwPDB NMR Structure Validation Report (i)

Apr 7, 2022 – 01:12 pm BST

PDB ID	:	7OJ9
Title	:	NMR solution structure of SNX9 SH3 - EEEV nsP3 peptide complex
Authors	:	Tossavainen, H.; Permi, P.
Deposited on	:	2021-05-14

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 (i)) were used in the production of this report:

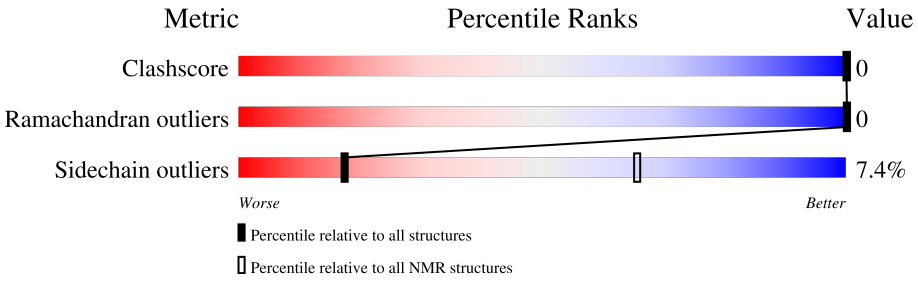
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 79%.

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	B	23	35% 9% 57%
2	A	67	75% . . 18%

2 Ensemble composition and analysis

This entry contains 20 models. Model 5 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	B:1667-B:1676, A:3-A:13, A:17-A:60 (65)	0.15	5

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 5 clusters and 3 single-model clusters were found.

Cluster number	Models
1	2, 6, 7, 8, 13
2	11, 14, 16, 17, 20
3	1, 5, 18
4	4, 9
5	3, 10
Single-model clusters	12; 15; 19

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1368 atoms, of which 684 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called EEEV nsP3 peptide.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	B	23	380	115	200	38	27	0

- Molecule 2 is a protein called Sorting nexin-9.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	67	988	312	484	86	104	2	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q9Y5X1
A	-1	SER	-	expression tag	UNP Q9Y5X1
A	0	HIS	-	expression tag	UNP Q9Y5X1

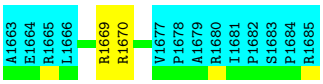
4 Residue-property plots

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.

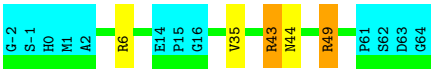
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

Chain A:



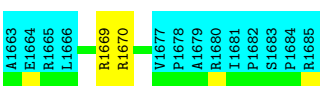
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

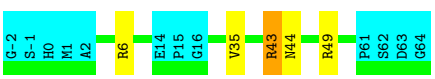
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

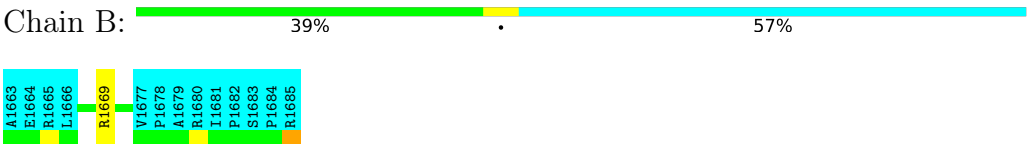
Chain A:



4.2.2 Score per residue for model 2

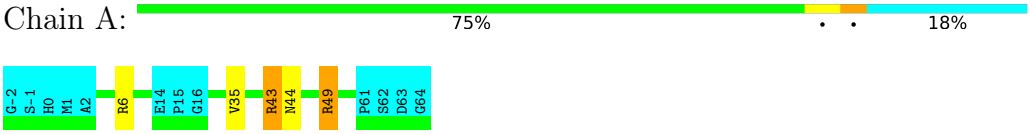
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

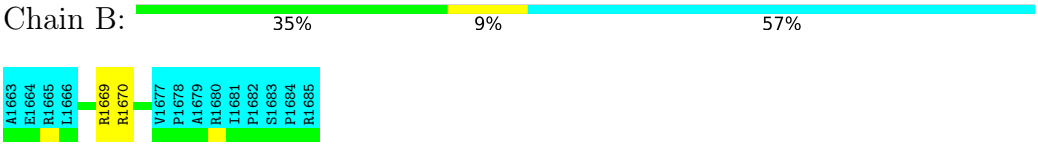
Chain A:



4.2.3 Score per residue for model 3

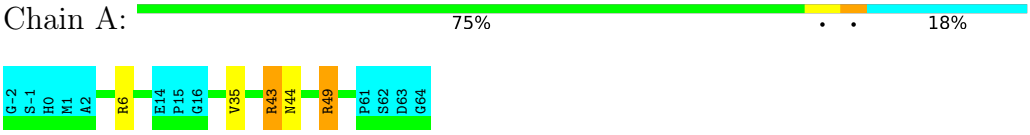
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

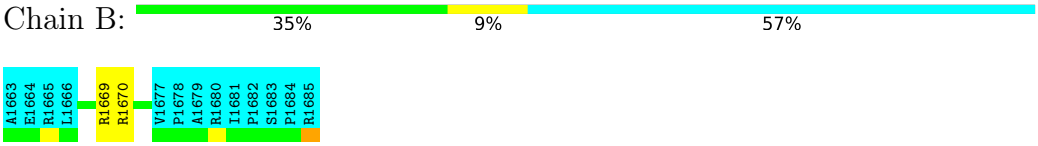
Chain A:



4.2.4 Score per residue for model 4

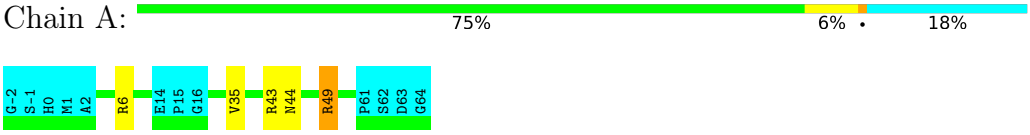
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

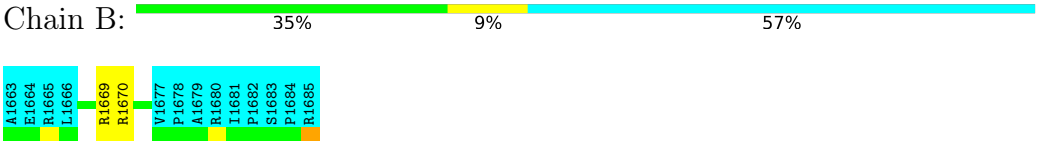
Chain A:



4.2.5 Score per residue for model 5 (medoid)

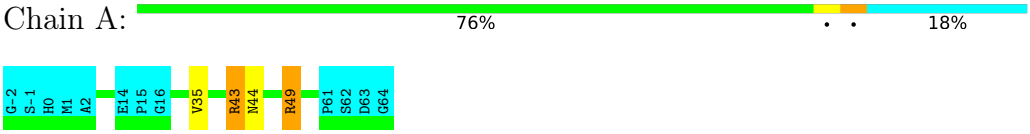
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

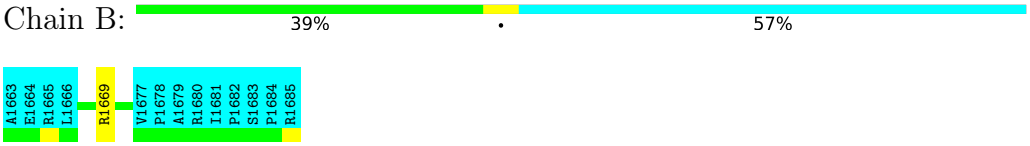
Chain A:



4.2.6 Score per residue for model 6

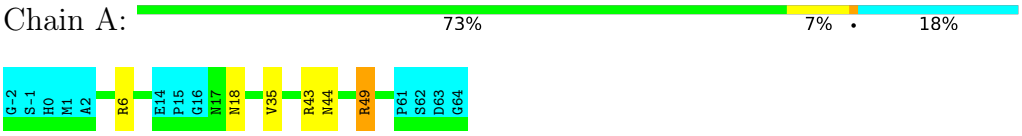
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

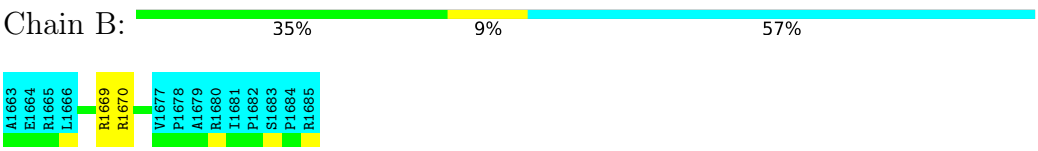
Chain A:



4.2.7 Score per residue for model 7

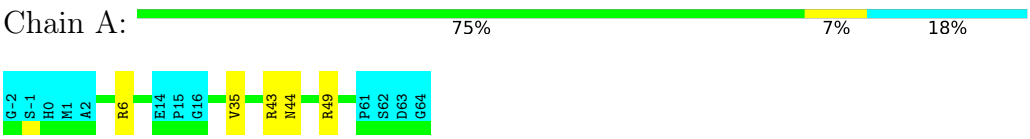
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

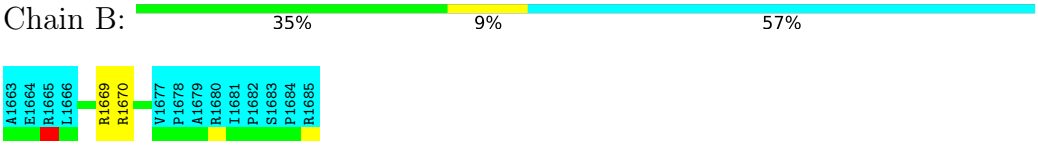
Chain A:



4.2.8 Score per residue for model 8

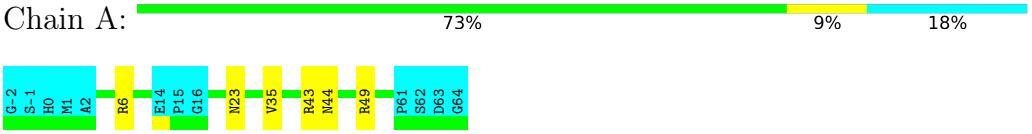
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

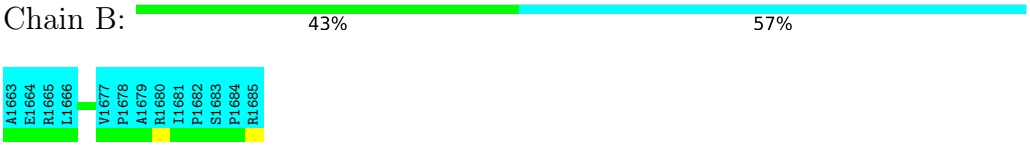
Chain A:



4.2.9 Score per residue for model 9

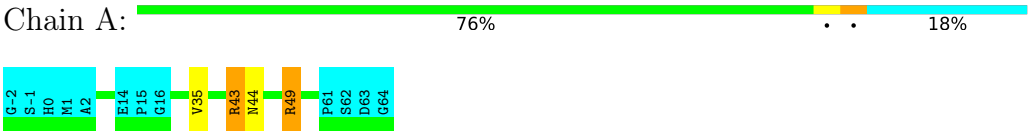
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

Chain A:



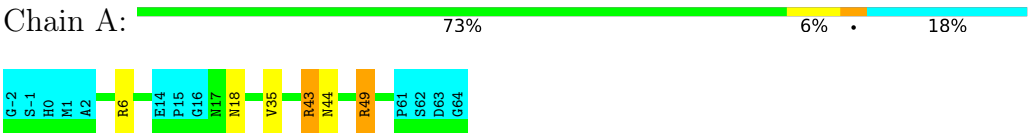
4.2.10 Score per residue for model 10

• Molecule 1: EEEV nsP3 peptide

Chain B:

• Molecule 2: Sorting nexin-9

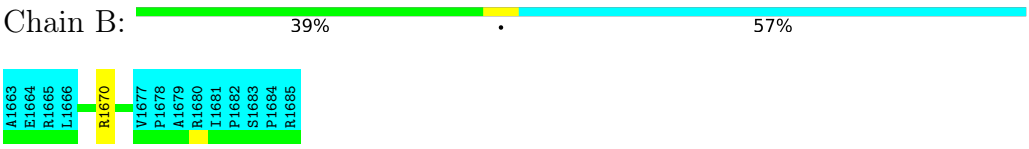
Chain A:



4.2.11 Score per residue for model 11

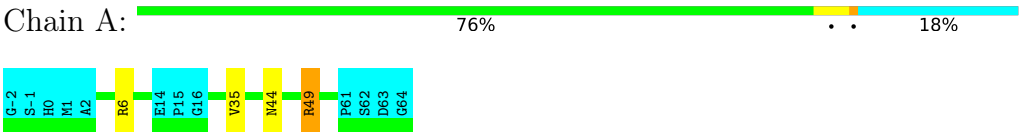
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

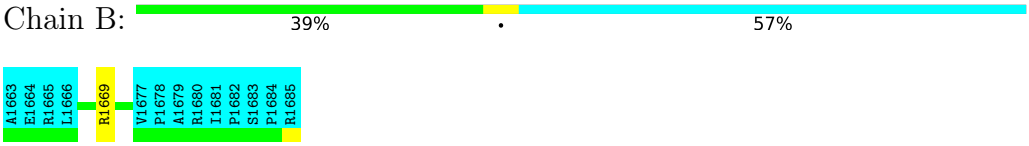
Chain A:



4.2.12 Score per residue for model 12

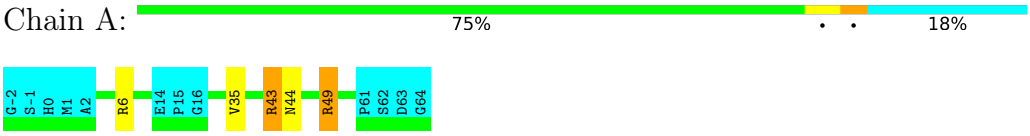
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

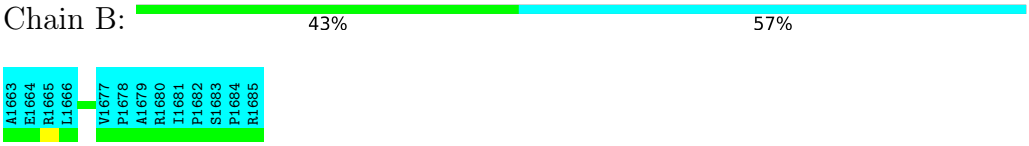
Chain A:



4.2.13 Score per residue for model 13

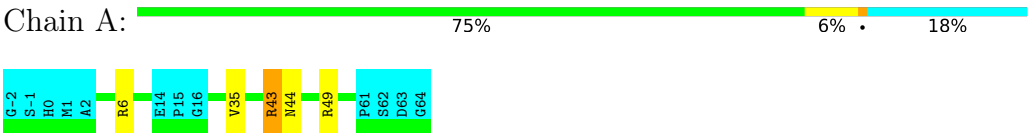
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

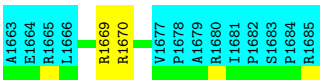
Chain A:



4.2.14 Score per residue for model 14

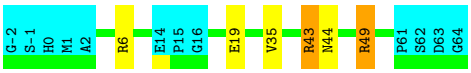
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

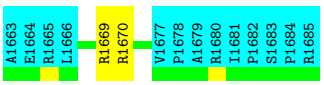
Chain A:



4.2.15 Score per residue for model 15

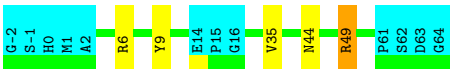
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

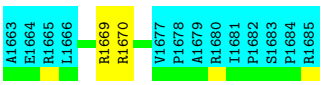
Chain A:



4.2.16 Score per residue for model 16

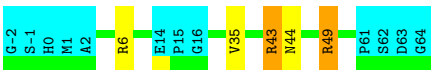
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

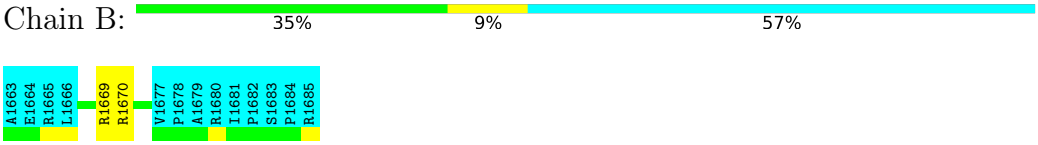
Chain A:



4.2.17 Score per residue for model 17

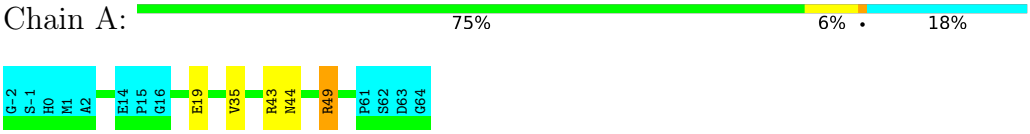
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

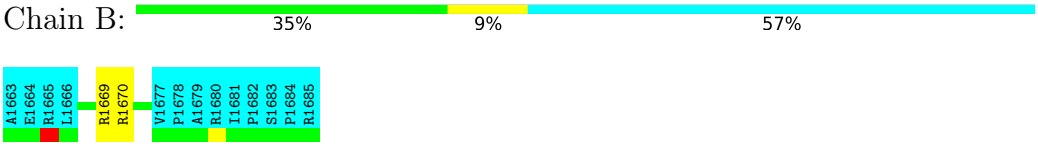
Chain A:



4.2.18 Score per residue for model 18

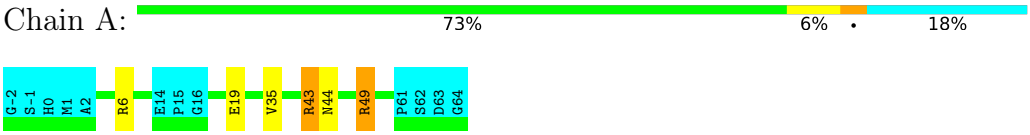
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

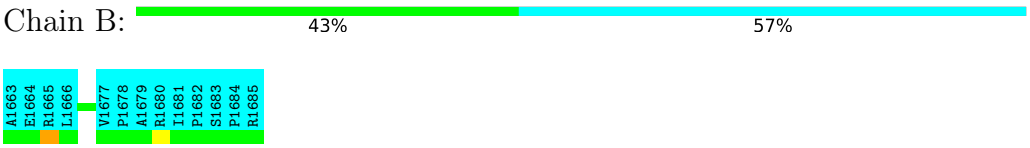
Chain A:



4.2.19 Score per residue for model 19

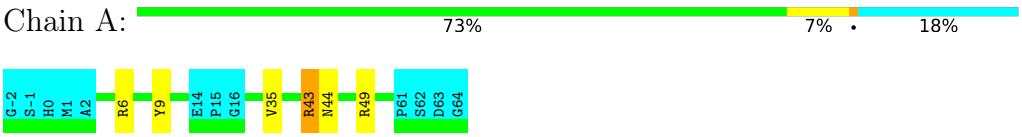
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

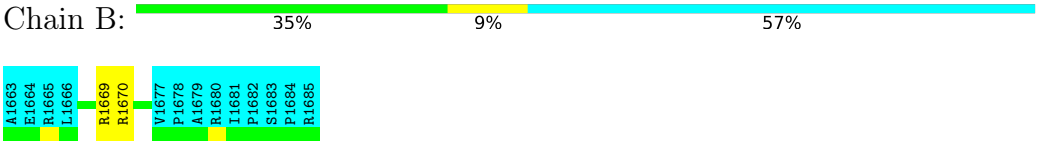
Chain A:



4.2.20 Score per residue for model 20

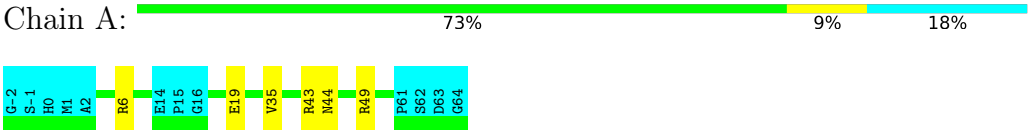
• Molecule 1: EEEV nsP3 peptide

Chain B:



• Molecule 2: Sorting nexin-9

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 30 calculated structures, 20 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
CYANA	structure calculation	
Amber	refinement	

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	913
Number of shifts mapped to atoms	913
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	79%

6 Model quality

6.1 Standard geometry

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	B	0.75±0.02	0±0/82 (0.0± 0.0%)	1.38±0.12	2±1/116 (1.3± 0.7%)
2	A	0.68±0.01	0±0/432 (0.0± 0.0%)	1.14±0.03	3±1/589 (0.4± 0.1%)
All	All	0.69	0/10280 (0.0%)	1.19	83/14100 (0.6%)

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	B	0.0±0.0	0.1±0.3
2	A	0.0±0.0	0.1±0.3
All	All	0	4

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	B	1670	ARG	NE-CZ-NH1	8.81	124.70	120.30	11	13
2	A	43	ARG	NE-CZ-NH1	8.66	124.63	120.30	13	16
2	A	6	ARG	NE-CZ-NH1	8.61	124.60	120.30	2	17
2	A	49	ARG	NE-CZ-NH1	7.89	124.25	120.30	6	14
1	B	1669	ARG	NE-CZ-NH1	7.86	124.23	120.30	1	16
2	A	43	ARG	NE-CZ-NH2	-6.36	117.12	120.30	6	4
1	B	1670	ARG	NE-CZ-NH2	-5.28	117.66	120.30	18	1
1	B	1670	ARG	NH1-CZ-NH2	-5.02	113.88	119.40	7	1
2	A	6	ARG	NH1-CZ-NH2	-5.00	113.90	119.40	4	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	B	1670	ARG	Sidechain	2
2	A	9	TYR	Sidechain	2

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	10040	10080	10080	-

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	B	10/23 (43%)	9±0 (94±5%)	1±0 (6±5%)	0±0 (0±0%)	100	100
2	A	55/67 (82%)	53±1 (97±2%)	2±1 (3±2%)	0±0 (0±0%)	100	100
All	All	1300/1800 (72%)	1255 (97%)	45 (3%)	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	B	9/20 (45%)	9±0 (100±0%)	0±0 (0±0%)	100	100
2	A	45/53 (85%)	41±1 (91±1%)	4±1 (9±1%)	13	60
All	All	1080/1460 (74%)	1000 (93%)	80 (7%)	17	65

All 7 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)
2	A	35	VAL	20
2	A	44	ASN	20
2	A	49	ARG	20
2	A	43	ARG	13
2	A	19	GLU	4
2	A	18	ASN	2
2	A	23	ASN	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)

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 79% for the well-defined parts and 73% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *eeev.bmrb*

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	913
Number of shifts mapped to atoms	913
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	65	0.13 ± 0.20	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	57	-0.10 ± 0.28	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	59	0.06 ± 0.60	None needed (< 0.5 ppm)

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 79%, i.e. 622 atoms were assigned a chemical shift out of a possible 790. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	228/311 (73%)	122/123 (99%)	53/130 (41%)	53/58 (91%)
Sidechain	357/442 (81%)	234/258 (91%)	117/162 (72%)	6/22 (27%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	37/37 (100%)	19/19 (100%)	17/17 (100%)	1/1 (100%)
Overall	622/790 (79%)	375/400 (94%)	187/309 (61%)	60/81 (74%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 73%, i.e. 787 atoms were assigned a chemical shift out of a possible 1079. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	285/426 (67%)	161/168 (96%)	65/180 (36%)	59/78 (76%)
Sidechain	461/609 (76%)	321/360 (89%)	134/218 (61%)	6/31 (19%)
Aromatic	41/44 (93%)	21/23 (91%)	19/19 (100%)	1/2 (50%)
Overall	787/1079 (73%)	503/551 (91%)	218/417 (52%)	66/111 (59%)

7.1.4 Statistically unusual chemical shifts (i)

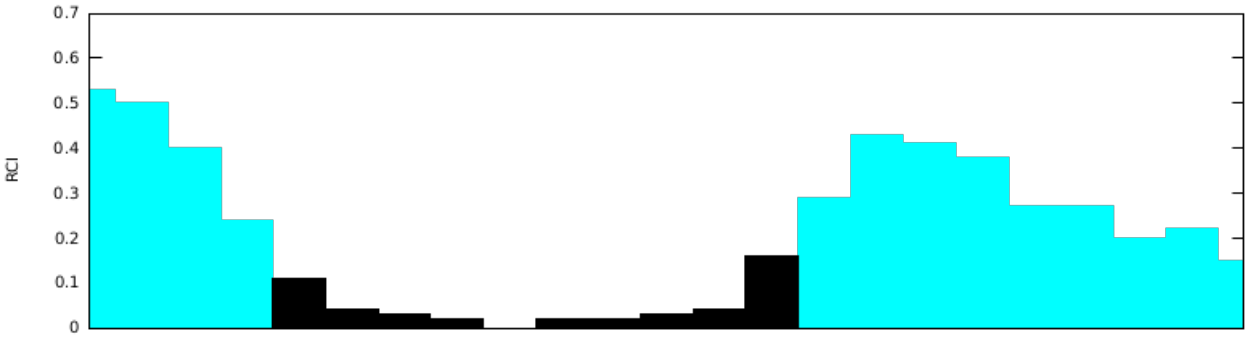
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	53	PRO	HG3	0.03	3.56 – 0.26	-5.7
1	A	53	PRO	HG2	0.18	3.48 – 0.38	-5.6

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 B:



Random coil index (RCI) for chain A: