Full wwPDB NMR Structure Validation Report (i)

Dec 13, 2023 – 06:55 PM EST

PDB ID	:	2JQF
BMRB ID	:	15277
Title	:	Full Length Leader Protease of Foot and Mouth Disease Virus C51A Mutant
Authors	:	Cencic, R.; Mayer, C.; Juliano, M.A.; Juliano, L.; Konrat, R.; Kontaxis, G.; Skern, T.
Deposited on	:	2007-06-01

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

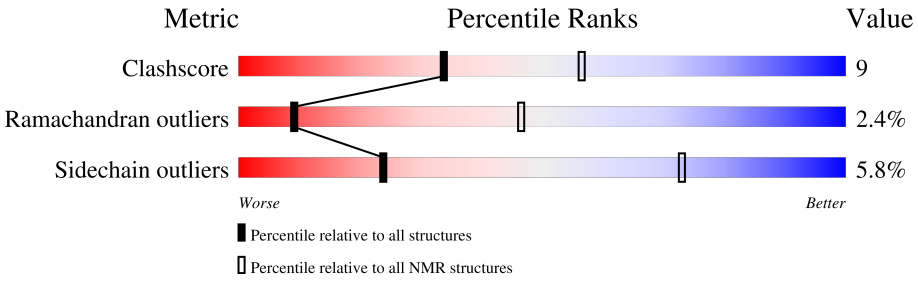
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 25%.

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	R	173	64% 24% • 5% •
1	S	173	64% 25% • 10%

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: *closest to the average*.

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	R:30-R:185, R:193-R:194 (158)	0.10	10
2	S:229-S:384 (156)	0.08	4

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 4 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5352 atoms, of which 2617 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms						Trace
1	R	166	Total	C	H	N	O	S	0
			2622	867	1284	217	249	5	
1	S	173	Total	C	H	N	O	S	0
			2730	905	1333	226	261	5	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
R	51	ALA	CYS	engineered mutation	UNP P03305
R	126	VAL	MET	engineered mutation	UNP P03305
S	251	ALA	CYS	engineered mutation	UNP P03305
S	326	VAL	MET	engineered mutation	UNP P03305

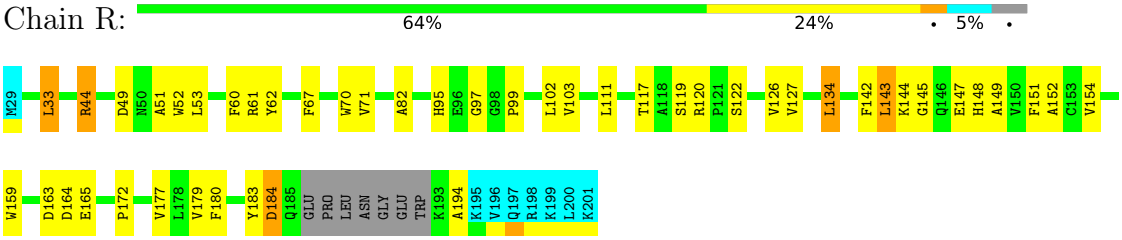
4 Residue-property plots

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.

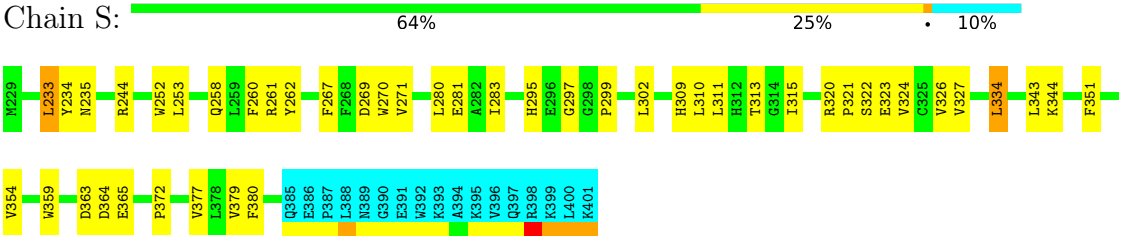
• Molecule 1: Genome polyprotein

Chain R:



• Molecule 1: Genome polyprotein

Chain S:



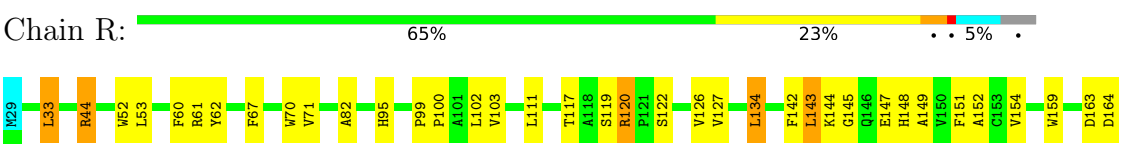
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

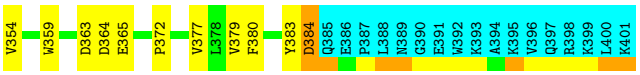
4.2.1 Score per residue for model 1

• Molecule 1: Genome polyprotein

Chain R:

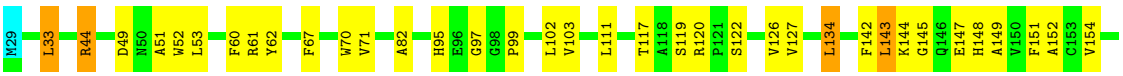


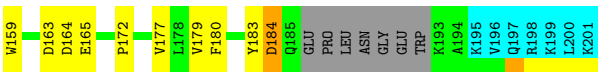
• Molecule 1: Genome polyprotein



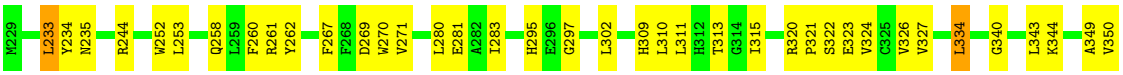
4.2.2 Score per residue for model 2

• Molecule 1: Genome polyprotein



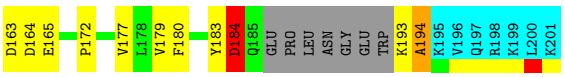


• Molecule 1: Genome polyprotein

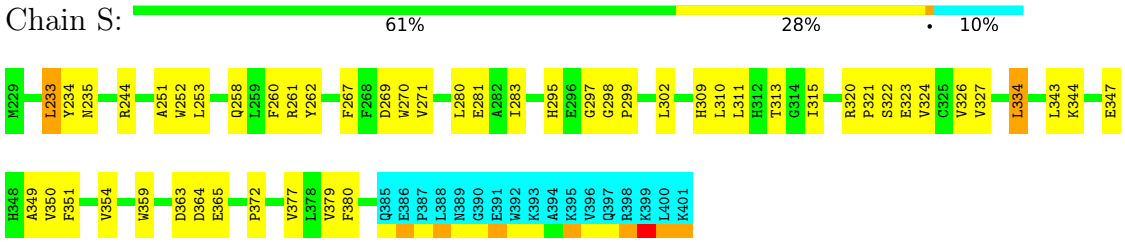


4.2.3 Score per residue for model 3

• Molecule 1: Genome polyprotein

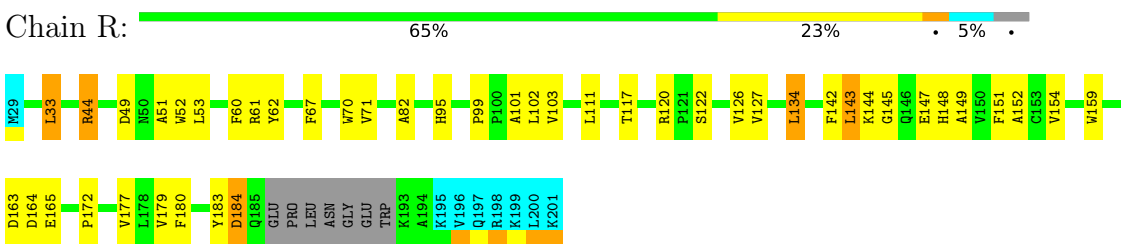


• Molecule 1: Genome polyprotein

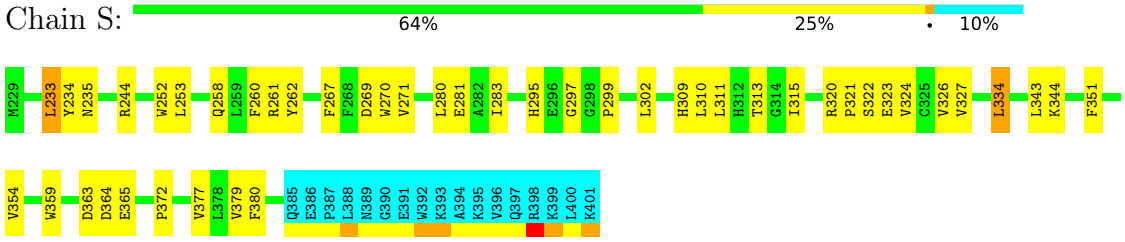


4.2.4 Score per residue for model 4

• Molecule 1: Genome polyprotein

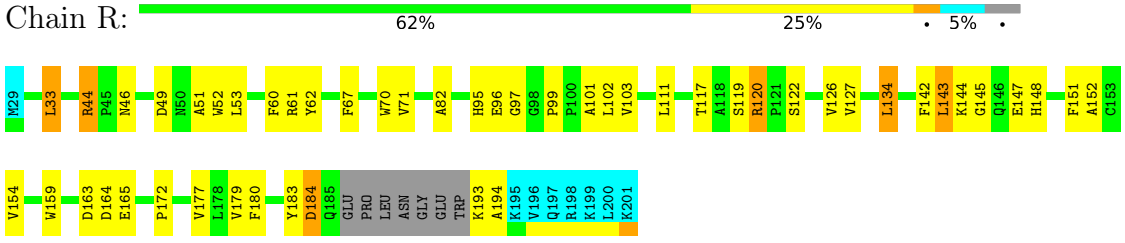


• Molecule 1: Genome polyprotein

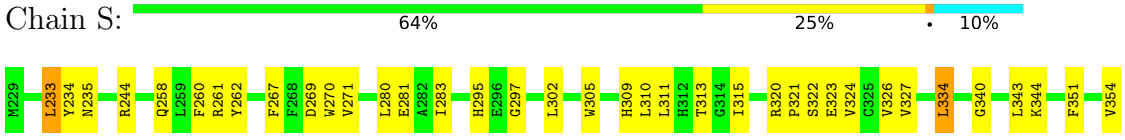


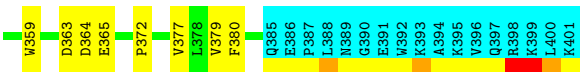
4.2.5 Score per residue for model 5

• Molecule 1: Genome polyprotein



• Molecule 1: Genome polyprotein

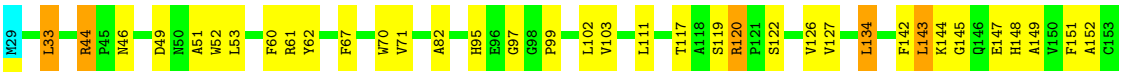


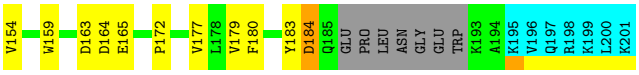


4.2.6 Score per residue for model 6

• Molecule 1: Genome polyprotein

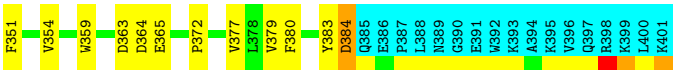
Chain R:





• Molecule 1: Genome polyprotein

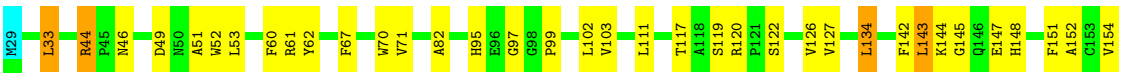
Chain S:



4.2.7 Score per residue for model 7

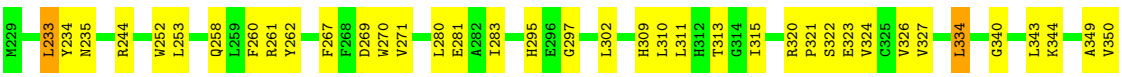
• Molecule 1: Genome polyprotein

Chain R:



• Molecule 1: Genome polyprotein

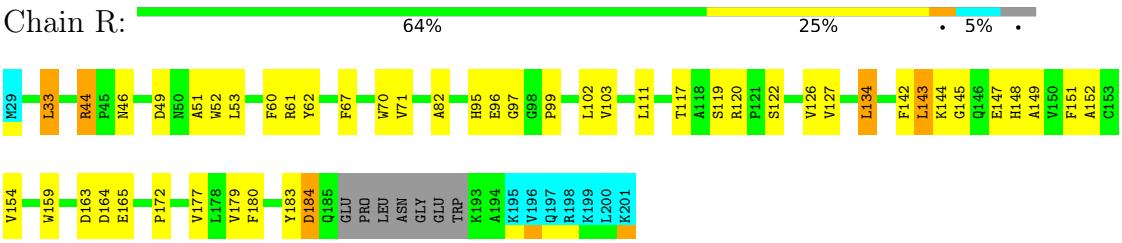
Chain S:



4.2.8 Score per residue for model 8

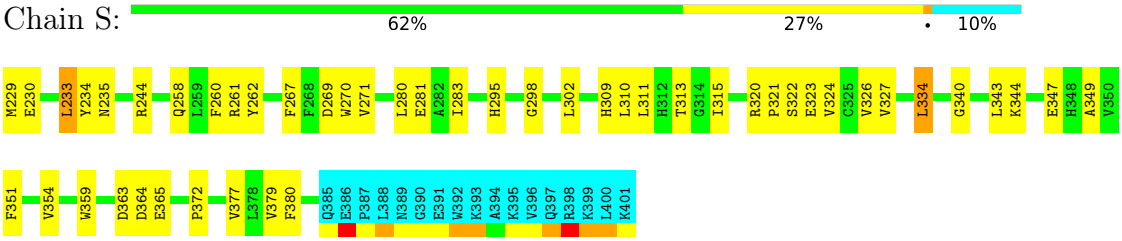
• Molecule 1: Genome polyprotein

Chain R:



• Molecule 1: Genome polyprotein

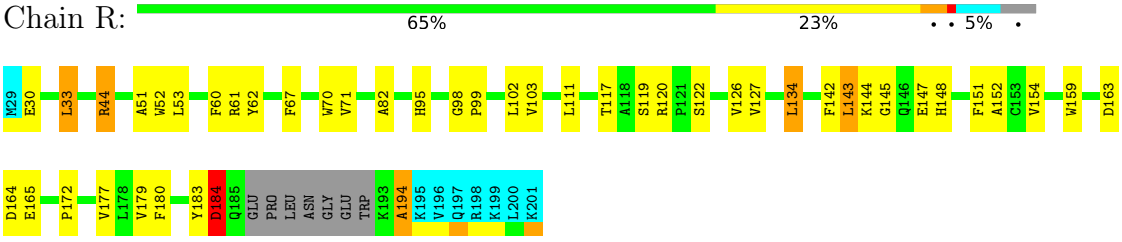
Chain S:



4.2.9 Score per residue for model 9

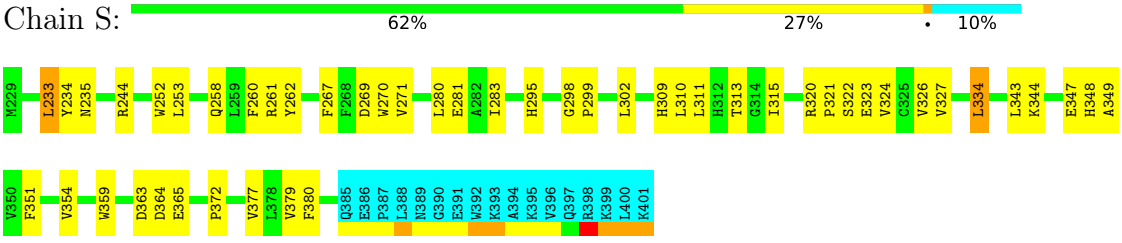
• Molecule 1: Genome polyprotein

Chain R:



• Molecule 1: Genome polyprotein

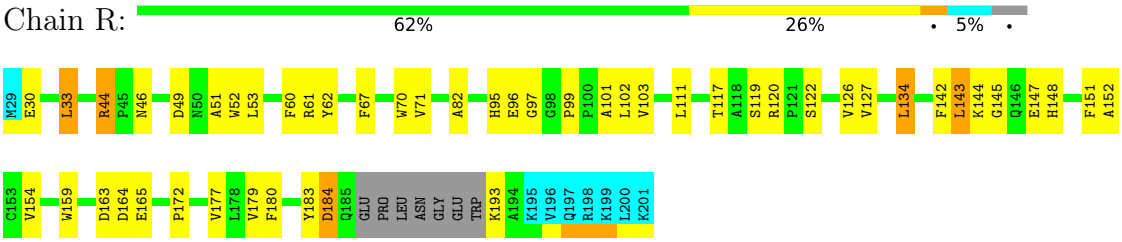
Chain S:



4.2.10 Score per residue for model 10 (medoid)

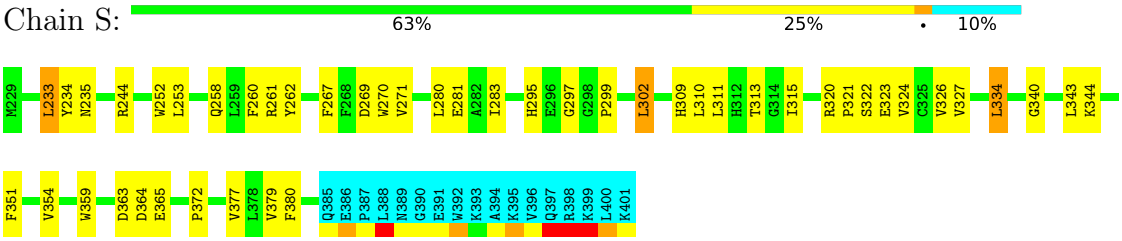
• Molecule 1: Genome polyprotein

Chain R:



• Molecule 1: Genome polyprotein

Chain S:



5 Refinement protocol and experimental data overview

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

Of the 10 calculated structures, 10 were deposited, based on the following criterion: *all calculated structures submitted*.

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

Software name	Classification	Version
X-PLOR NIH	structure solution	
X-PLOR NIH	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	1096
Number of shifts mapped to atoms	1081
Number of unparsed shifts	0
Number of shifts with mapping errors	15
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	25%

6 Model quality

6.1 Standard geometry

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

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	R	0.0±0.0	3.0±0.0
1	S	0.0±0.0	3.0±0.0
All	All	0	60

There are no bond-length outliers.

There are no bond-angle outliers.

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	R	44	ARG	Sidechain	10
1	R	61	ARG	Sidechain	10
1	R	120	ARG	Sidechain	10
1	S	244	ARG	Sidechain	10
1	S	261	ARG	Sidechain	10
1	S	320	ARG	Sidechain	10

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
1	R	1269	1193	1192	24±1
1	S	1254	1178	1175	21±1
All	All	25230	23710	23670	442

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 9.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:S:383:TYR:O	1:S:384:ASP:CB	0.98	2.10	1	2
1:S:383:TYR:O	1:S:384:ASP:HB2	0.90	1.66	6	2
1:S:383:TYR:O	1:S:384:ASP:HB3	0.79	1.76	1	2
1:R:183:TYR:O	1:R:184:ASP:C	0.74	2.26	1	10
1:R:99:PRO:O	1:R:103:VAL:HG23	0.70	1.85	7	10
1:R:143:LEU:HD23	1:R:148:HIS:O	0.65	1.92	5	10
1:S:313:THR:HG23	1:S:323:GLU:HG2	0.65	1.69	6	10
1:R:127:VAL:HG11	1:R:177:VAL:O	0.64	1.92	1	10
1:R:119:SER:HA	1:R:194:ALA:HB3	0.63	1.70	7	2
1:S:233:LEU:HD22	1:S:262:TYR:CE2	0.60	2.31	10	10
1:R:119:SER:HA	1:R:194:ALA:CB	0.57	2.29	7	3
1:R:151:PHE:CD1	1:R:152:ALA:N	0.54	2.75	9	10
1:R:154:VAL:HG22	1:R:159:TRP:CE2	0.53	2.39	8	10
1:R:193:LYS:O	1:R:194:ALA:HB2	0.53	2.04	1	1
1:S:280:LEU:HA	1:S:283:ILE:HD12	0.52	1.81	1	10
1:S:327:VAL:HG11	1:S:377:VAL:O	0.52	2.05	1	10
1:R:33:LEU:HD22	1:R:62:TYR:CE2	0.51	2.41	7	10
1:R:126:VAL:HG12	1:R:179:VAL:HG22	0.50	1.82	1	10
1:S:270:TRP:CE2	1:S:271:VAL:CG2	0.49	2.96	8	10
1:R:142:PHE:CD2	1:R:172:PRO:HB2	0.49	2.41	3	10
1:R:151:PHE:CD1	1:R:151:PHE:C	0.49	2.85	4	10
1:R:52:TRP:CD2	1:R:53:LEU:N	0.48	2.80	5	10
1:S:354:VAL:HG22	1:S:359:TRP:CE2	0.48	2.44	6	10
1:S:315:ILE:HD13	1:S:324:VAL:HB	0.48	1.85	10	10
1:R:70:TRP:CE2	1:R:71:VAL:CG2	0.48	2.97	1	10
1:S:270:TRP:CE2	1:S:271:VAL:HG23	0.47	2.45	8	10
1:S:233:LEU:HD22	1:S:262:TYR:CZ	0.46	2.46	1	10
1:S:334:LEU:HD11	1:S:372:PRO:O	0.45	2.12	6	10
1:S:234:TYR:CG	1:S:235:ASN:N	0.45	2.85	5	10
1:R:134:LEU:HD11	1:R:172:PRO:O	0.45	2.12	2	10
1:R:154:VAL:CG2	1:R:159:TRP:CZ2	0.44	2.99	1	10
1:S:327:VAL:HG23	1:S:327:VAL:O	0.43	2.13	1	9
1:S:327:VAL:O	1:S:327:VAL:HG23	0.43	2.13	6	1
1:R:70:TRP:CE2	1:R:71:VAL:HG23	0.43	2.48	3	10
1:R:117:THR:HG22	1:R:126:VAL:CG2	0.43	2.44	2	10
1:S:326:VAL:HG12	1:S:379:VAL:HG22	0.42	1.90	7	10
1:R:177:VAL:O	1:R:177:VAL:HG13	0.42	2.14	1	10
1:S:260:PHE:CE2	1:S:311:LEU:HD21	0.42	2.49	3	10
1:S:258:GLN:HB3	1:S:351:PHE:CD2	0.42	2.49	2	10

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:S:363:ASP:O	1:S:365:GLU:N	0.42	2.53	4	10
1:R:120:ARG:NH1	1:S:305:TRP:CE2	0.42	2.88	6	1
1:R:120:ARG:NH1	1:S:305:TRP:NE1	0.42	2.68	5	2
1:R:163:ASP:O	1:R:165:GLU:N	0.41	2.52	6	10
1:S:309:HIS:CD2	1:S:310:LEU:HG	0.41	2.50	4	10
1:R:143:LEU:HD22	1:R:145:GLY:O	0.41	2.15	4	10
1:R:60:PHE:CE2	1:R:111:LEU:HD21	0.41	2.50	3	10
1:R:180:PHE:CD1	1:R:180:PHE:N	0.41	2.89	4	10
1:S:380:PHE:C	1:S:380:PHE:CD1	0.41	2.94	5	9
1:S:380:PHE:CD1	1:S:380:PHE:C	0.41	2.94	1	1
1:S:252:TRP:CD2	1:S:253:LEU:N	0.41	2.89	1	8
1:R:71:VAL:HG11	1:R:82:ALA:HB2	0.41	1.92	3	10
1:S:229:MET:HG3	1:S:230:GLU:N	0.41	2.31	8	1
1:R:142:PHE:CD1	1:R:142:PHE:N	0.41	2.88	2	10
1:S:340:GLY:CA	1:S:379:VAL:O	0.41	2.69	8	5
1:S:349:ALA:C	1:S:350:VAL:HG13	0.40	2.37	3	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	R	156/173 (90%)	134±0 (86±0%)	17±1 (11±1%)	4±0 (3±0%)	8	42
1	S	155/173 (90%)	136±0 (88±0%)	16±0 (10±0%)	3±0 (2±0%)	10	50
All	All	3110/3460 (90%)	2703 (87%)	331 (11%)	76 (2%)	9	46

All 9 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	R	33	LEU	10
1	R	122	SER	10
1	R	164	ASP	10
1	R	184	ASP	10
1	S	233	LEU	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	S	322	SER	10
1	S	364	ASP	10
1	R	194	ALA	4
1	S	384	ASP	2

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	R	137/151 (91%)	129±1 (94±0%)	8±1 (6±0%)	25	74
1	S	136/151 (90%)	128±0 (94±0%)	8±0 (6±0%)	23	72
All	All	2730/3020 (90%)	2572 (94%)	158 (6%)	24	73

All 19 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	R	44	ARG	10
1	R	67	PHE	10
1	R	95	HIS	10
1	R	102	LEU	10
1	R	134	LEU	10
1	R	143	LEU	10
1	R	144	LYS	10
1	S	267	PHE	10
1	S	269	ASP	10
1	S	281	GLU	10
1	S	295	HIS	10
1	S	302	LEU	10
1	S	334	LEU	10
1	S	343	LEU	10
1	S	344	LYS	10
1	R	184	ASP	3
1	R	193	LYS	3
1	S	229	MET	1
1	R	185	GLN	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

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 25% for the well-defined parts and 23% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_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	1096
Number of shifts mapped to atoms	1081
Number of unparsed shifts	0
Number of shifts with mapping errors	15
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 15 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	R	186	GLU	H	8.45	0.030	1
1	R	186	GLU	CA	54.65	0.100	1
1	R	186	GLU	CB	29.76	0.100	1
1	R	186	GLU	N	123.6	0.050	1
1	R	187	PRO	C	177.06	0.100	1
1	R	187	PRO	CA	62.99	0.100	1
1	R	188	LEU	H	8.52	0.030	1
1	R	188	LEU	CA	55.08	0.100	1
1	R	188	LEU	N	123.5	0.050	1
1	R	189	ASN	H	8.58	0.030	1
1	R	189	ASN	CA	52.98	0.100	1
1	R	189	ASN	N	120.1	0.050	1
1	R	190	GLY	H	8.1	0.030	1
1	R	190	GLY	CA	45.22	0.100	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	R	190	GLY	N	108.0	0.050	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	290	-0.23 ± 0.18	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	108	0.08 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	174	0.10 ± 0.20	None needed (< 0.5 ppm)
^{15}N	262	0.72 ± 0.32	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 25%, i.e. 1062 atoms were assigned a chemical shift out of a possible 4225. 0 out of 56 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	957/1550 (62%)	253/628 (40%)	451/628 (72%)	253/294 (86%)
Sidechain	105/2165 (5%)	0/1414 (0%)	105/699 (15%)	0/52 (0%)
Aromatic	0/510 (0%)	0/254 (0%)	0/234 (0%)	0/22 (0%)
Overall	1062/4225 (25%)	253/2296 (11%)	556/1561 (36%)	253/368 (69%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 23%, i.e. 1081 atoms were assigned a chemical shift out of a possible 4624. 0 out of 61 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	974/1674 (58%)	258/678 (38%)	458/678 (68%)	258/318 (81%)
Sidechain	107/2428 (4%)	0/1581 (0%)	107/778 (14%)	0/69 (0%)
Aromatic	0/522 (0%)	0/260 (0%)	0/239 (0%)	0/23 (0%)
Overall	1081/4624 (23%)	258/2519 (10%)	565/1695 (33%)	258/410 (63%)

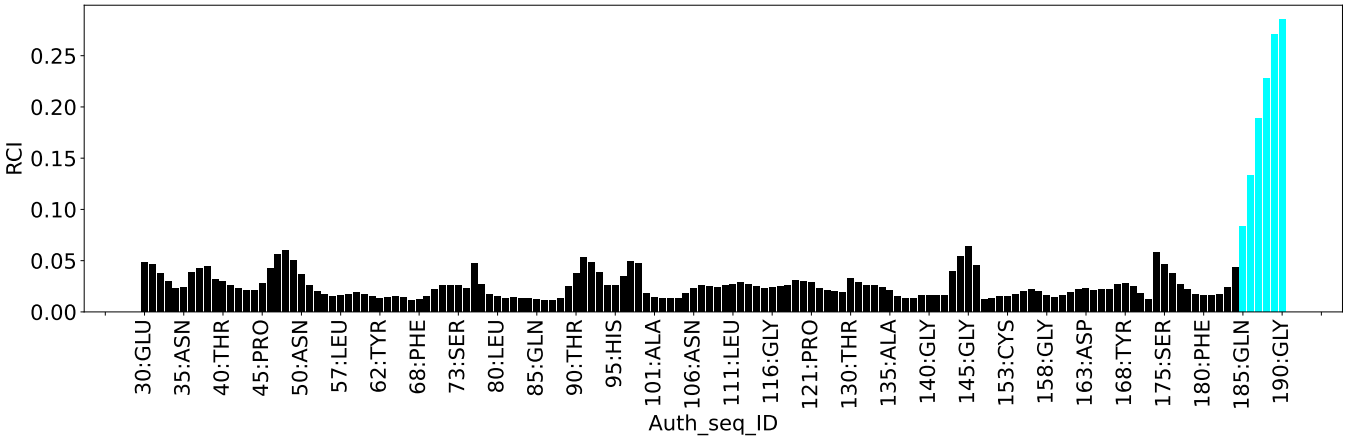
7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots

The image below reports *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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain R:



Random coil index (RCI) for chain S: