Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:37 pm BST

PDB ID	:	2KQW
Title	:	SARS coronavirus-unique domain (SUD): Three-domain molecular architecture in solution and RNA binding. II: Structure of the SUD-C domain of SUD-MC
Authors	:	Johnson, M.A.; Chatterjee, A.; Wuthrich, K.; Joint Center for Structural Genomics (JCSG)
Deposited on	:	2009-11-19

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:

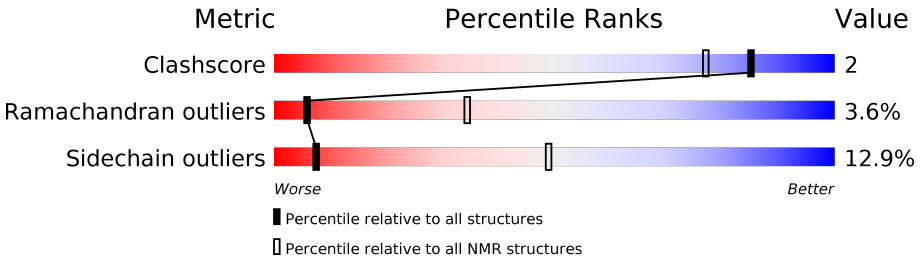
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 >=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	198	30% · · 66%

2 Ensemble composition and analysis

This entry contains 20 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	A:132-A:198 (67)	0.41	10

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

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

The models can be grouped into 4 clusters. No single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1062 atoms, of which 523 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Non-structural protein 3.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	67	1062	343	523	88	108	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP P0C6U8
A	2	SER	-	EXPRESSION TAG	UNP P0C6U8
A	3	HIS	-	EXPRESSION TAG	UNP P0C6U8
A	4	MET	-	EXPRESSION TAG	UNP P0C6U8

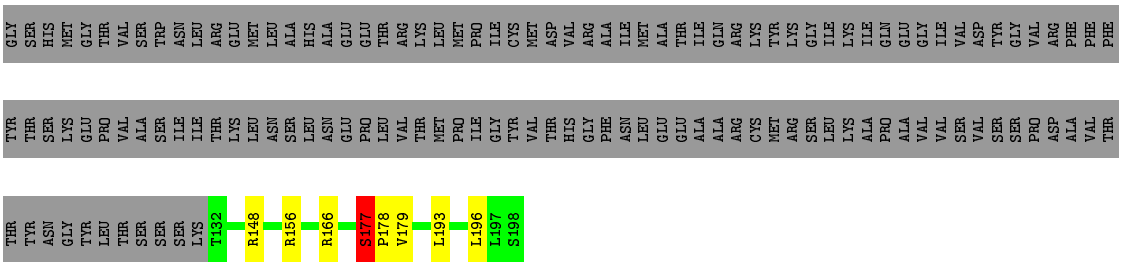
4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Non-structural protein 3

Chain A:



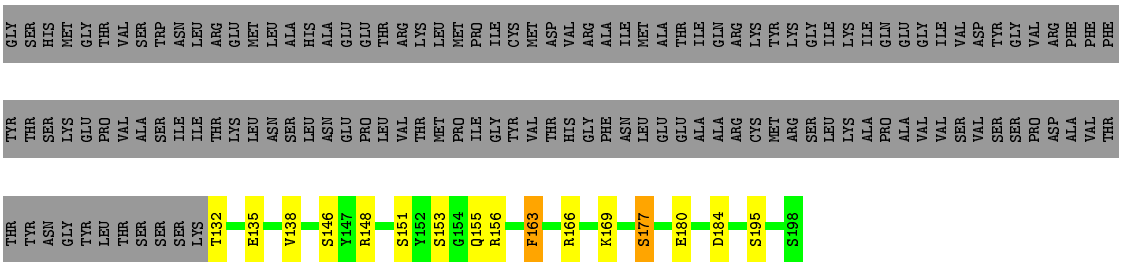
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Non-structural protein 3

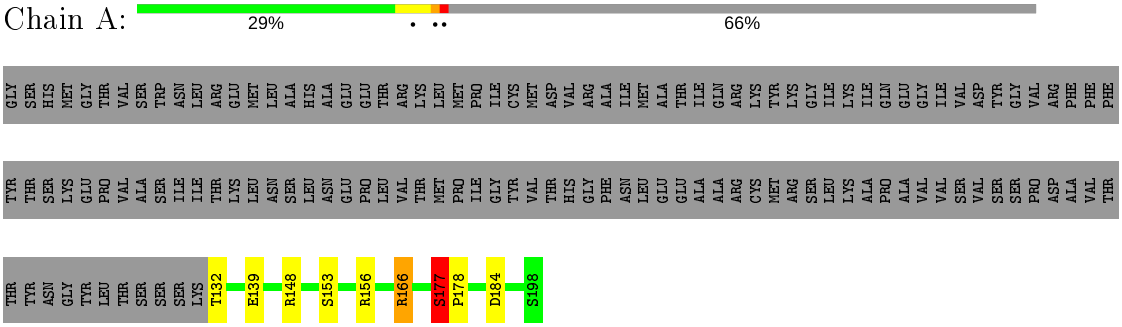
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Non-structural protein 3

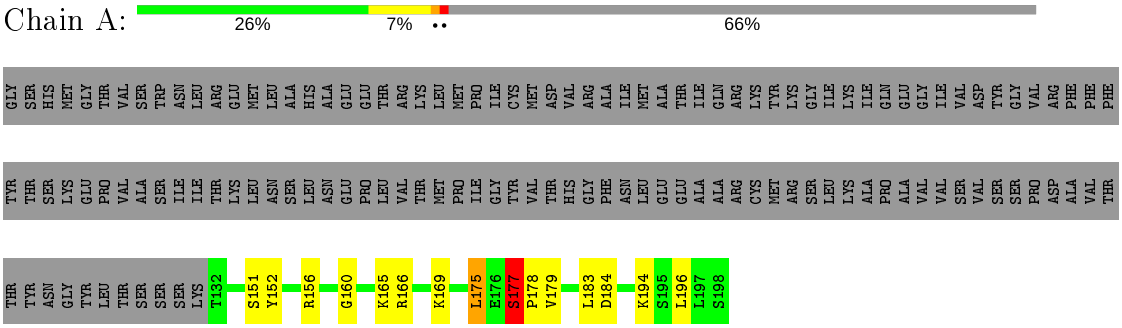
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Non-structural protein 3

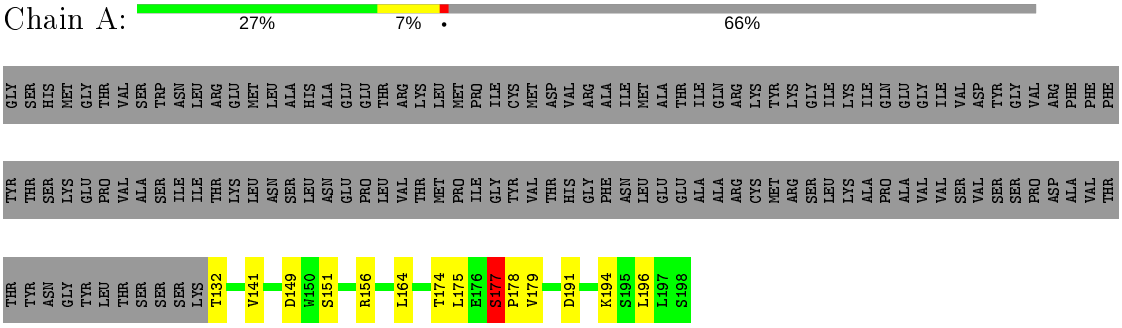
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Non-structural protein 3

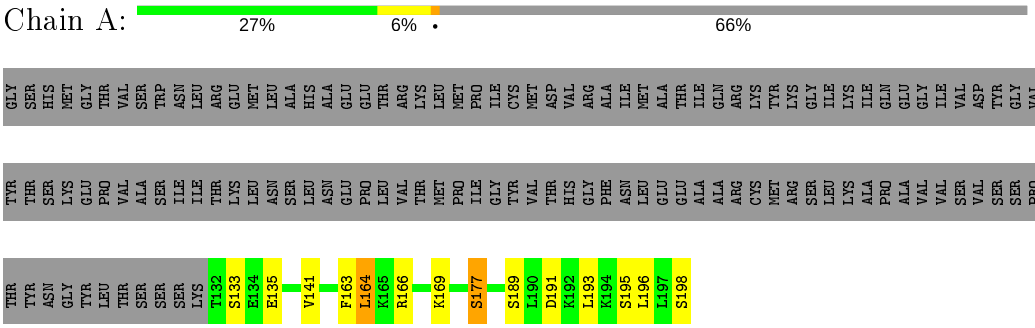
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Non-structural protein 3

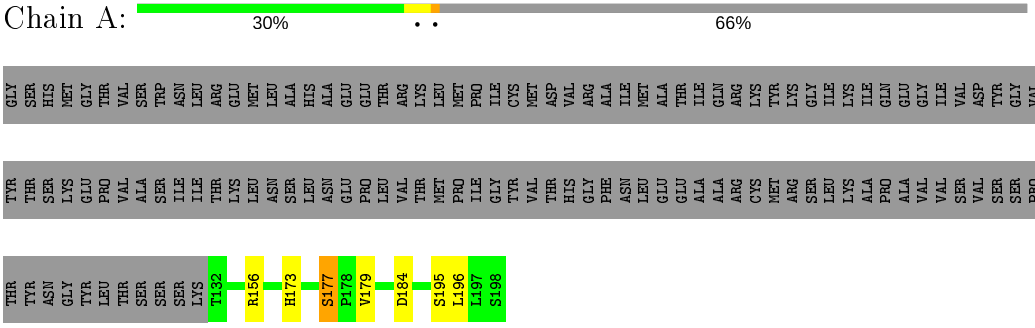
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Non-structural protein 3

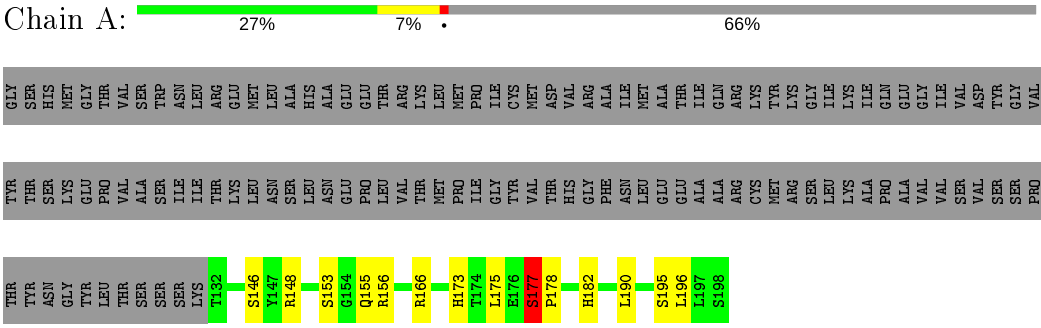
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Non-structural protein 3

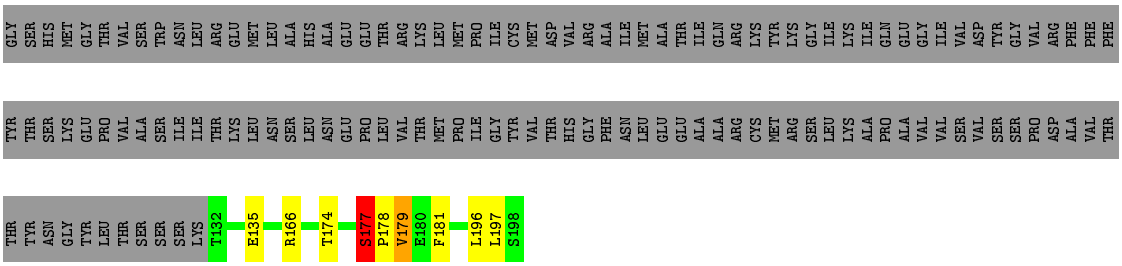
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Non-structural protein 3

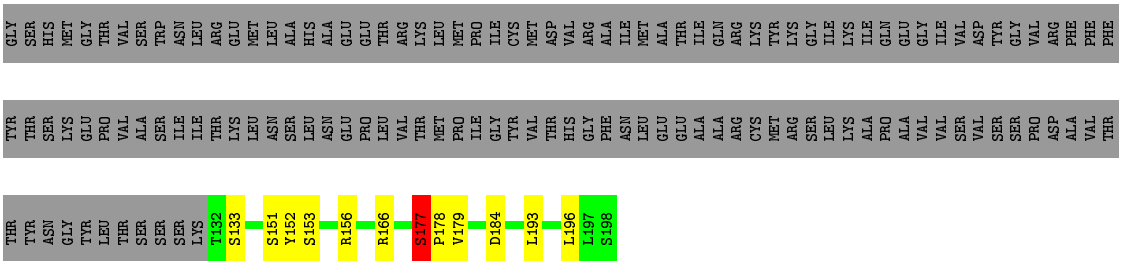
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Non-structural protein 3

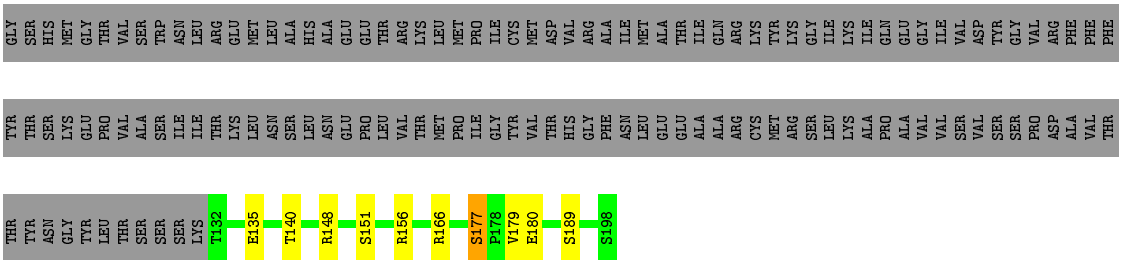
Chain A:



4.2.10 Score per residue for model 10 (medoid)

- Molecule 1: Non-structural protein 3

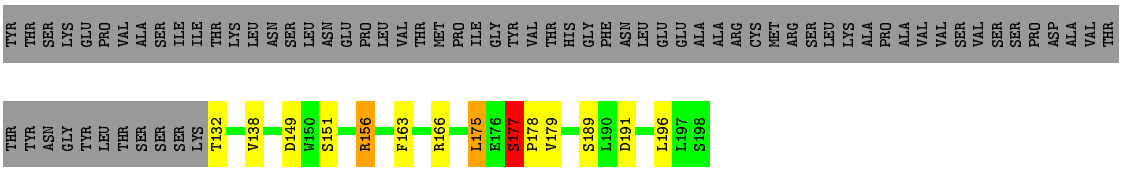
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Non-structural protein 3

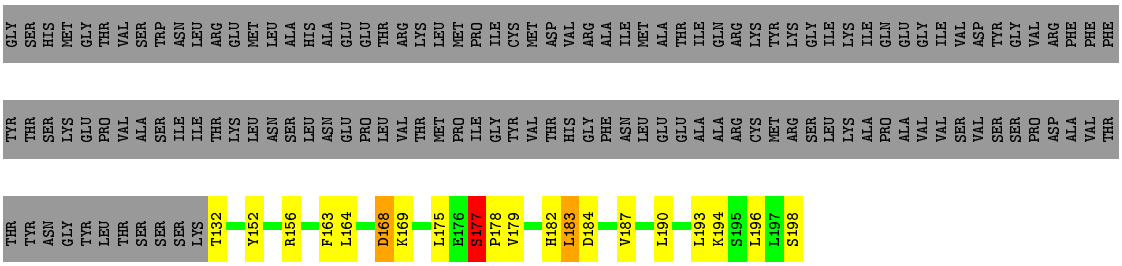
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Non-structural protein 3

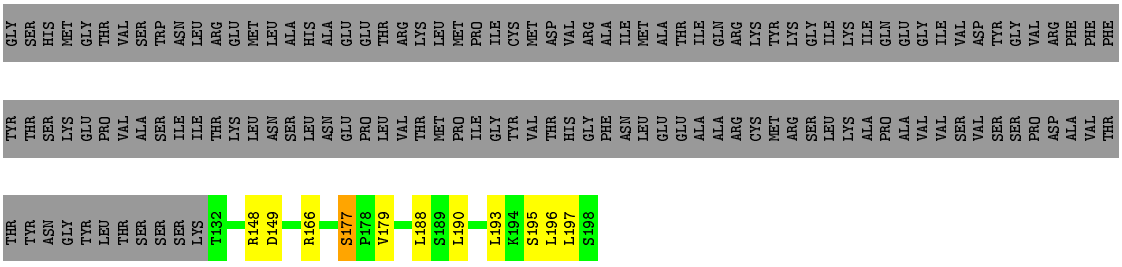
Chain A: 24% 9% •• 66%



4.2.13 Score per residue for model 13

- Molecule 1: Non-structural protein 3

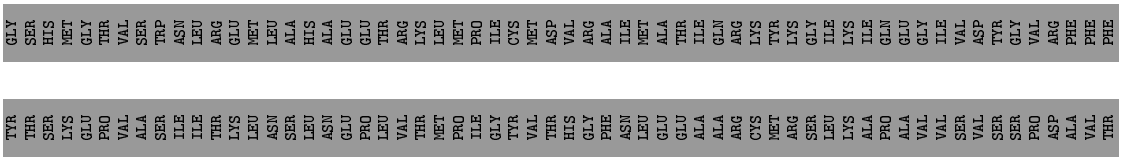
Chain A: 28% 5% • 66%



4.2.14 Score per residue for model 14

- Molecule 1: Non-structural protein 3

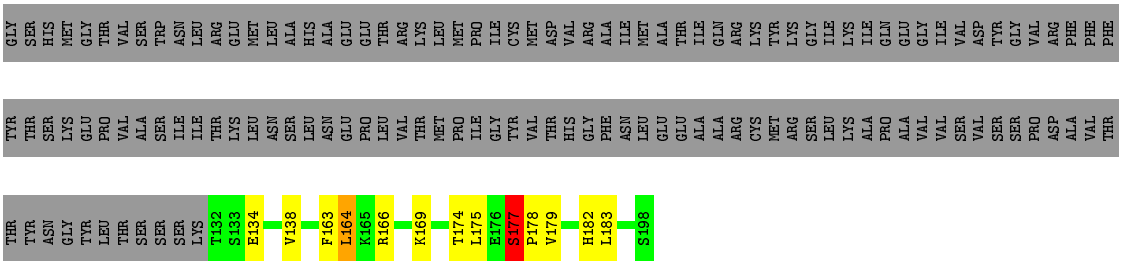
Chain A: 26% 7% • 66%



4.2.15 Score per residue for model 15

- Molecule 1: Non-structural protein 3

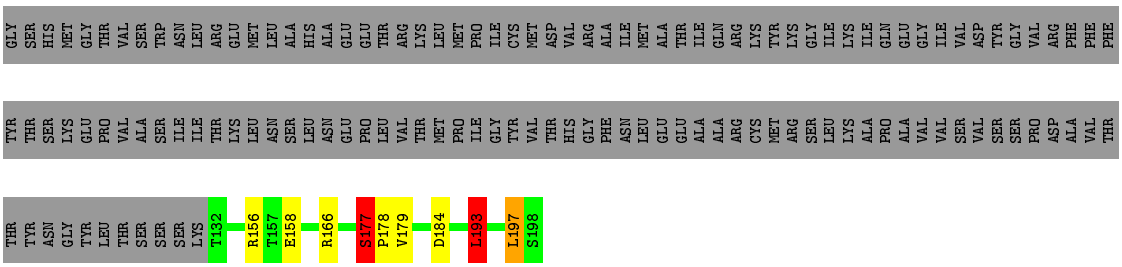
Chain A: 27% 6% 66%



4.2.16 Score per residue for model 16

- Molecule 1: Non-structural protein 3

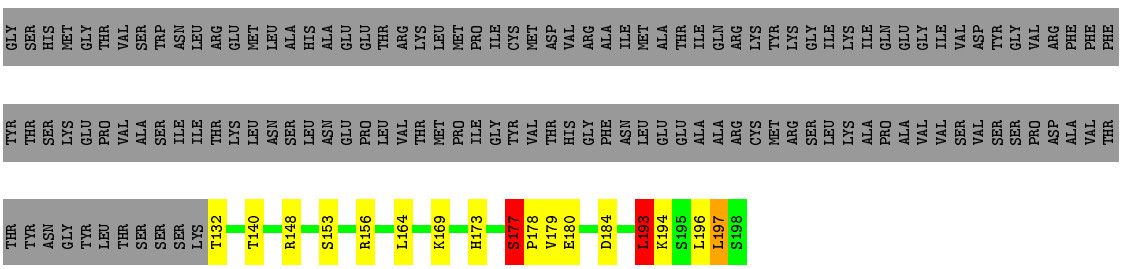
Chain A: 29% 66%



4.2.17 Score per residue for model 17

- Molecule 1: Non-structural protein 3

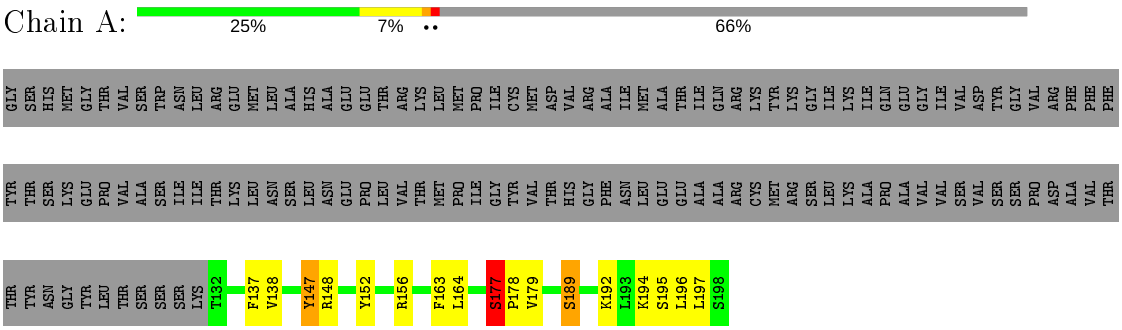
Chain A: 25% 7% 66%



4.2.18 Score per residue for model 18

• Molecule 1: Non-structural protein 3

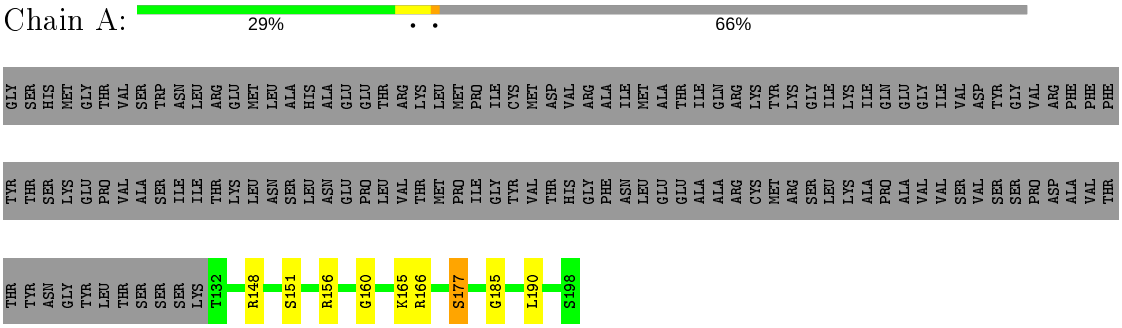
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Non-structural protein 3

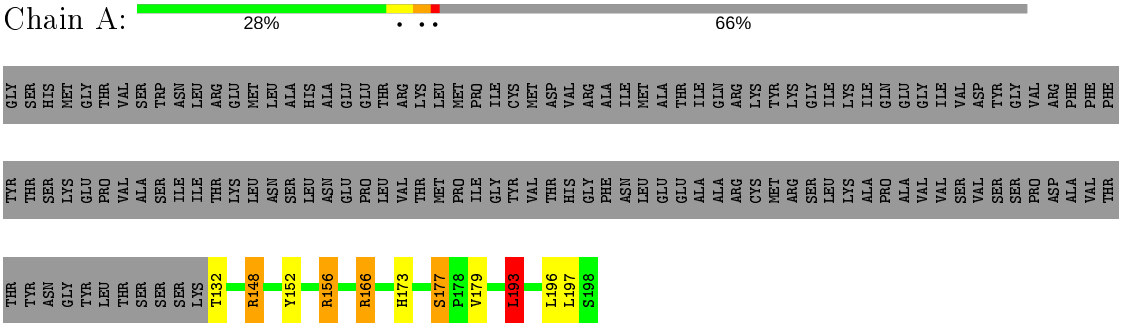
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Non-structural protein 3

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 80 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
UNIO	refinement	
CANDID	refinement	
OpalP	refinement	

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	A	539	523	522	2±2
All	All	10780	10460	10440	46

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:183:LEU:HD13	1:A:184:ASP:HB3	0.64	1.67	12	1
1:A:193:LEU:CD1	1:A:197:LEU:HD21	0.57	2.29	20	3
1:A:182:HIS:CD2	1:A:187:VAL:HG22	0.57	2.35	12	2
1:A:141:VAL:HG11	1:A:164:LEU:HD11	0.56	1.77	4	1
1:A:174:THR:HG23	1:A:180:GLU:OE1	0.55	2.02	14	1
1:A:174:THR:HG23	1:A:180:GLU:CD	0.54	2.23	14	1
1:A:163:PHE:CD2	1:A:164:LEU:HD23	0.52	2.38	5	3
1:A:188:LEU:HD12	1:A:196:LEU:HD11	0.48	1.85	13	1
1:A:175:LEU:H	1:A:175:LEU:HD22	0.48	1.69	11	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:197:LEU:HD22	1:A:197:LEU:H	0.47	1.70	17	1
1:A:141:VAL:HG11	1:A:164:LEU:HD21	0.47	1.86	5	1
1:A:163:PHE:CG	1:A:164:LEU:HD23	0.47	2.45	5	1
1:A:138:VAL:HA	1:A:163:PHE:CZ	0.46	2.45	14	5
1:A:188:LEU:CD1	1:A:196:LEU:HD11	0.46	2.40	13	1
1:A:193:LEU:O	1:A:197:LEU:HD22	0.46	2.10	13	2
1:A:194:LYS:HA	1:A:197:LEU:CD2	0.46	2.41	17	1
1:A:169:LYS:HB3	1:A:183:LEU:HD11	0.44	1.90	15	2
1:A:177:SER:N	1:A:178:PRO:HA	0.43	2.28	12	12
1:A:197:LEU:H	1:A:197:LEU:HD22	0.42	1.75	16	1
1:A:163:PHE:CE2	1:A:164:LEU:CD2	0.41	3.02	15	1
1:A:137:PHE:CD1	1:A:194:LYS:HE2	0.41	2.49	18	1
1:A:190:LEU:HD11	1:A:194:LYS:HE3	0.41	1.92	12	1
1:A:160:GLY:HA2	1:A:175:LEU:HD21	0.41	1.92	3	1
1:A:179:VAL:HG12	1:A:181:PHE:CZ	0.40	2.51	8	1

5.2 Torsion angles (i)

5.2.1 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	65/198 (33%)	55±2 (85±3%)	8±2 (12±3%)	2±1 (4±1%)	6	34
All	All	1300/3960 (33%)	1099 (85%)	154 (12%)	47 (4%)	6	34

All 10 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	177	SER	20
1	A	179	VAL	15
1	A	174	THR	4
1	A	148	ARG	2
1	A	134	GLU	1
1	A	168	ASP	1
1	A	160	GLY	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	153	SER	1
1	A	189	SER	1
1	A	185	GLY	1

5.2.2 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	61/174 (35%)	53±3 (87±4%)	8±3 (13±4%)	7 49
All	All	1220/3480 (35%)	1063 (87%)	157 (13%)	7 49

All 37 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	177	SER	20
1	A	156	ARG	14
1	A	196	LEU	13
1	A	166	ARG	8
1	A	132	THR	8
1	A	151	SER	7
1	A	184	ASP	7
1	A	195	SER	7
1	A	193	LEU	6
1	A	175	LEU	6
1	A	164	LEU	4
1	A	189	SER	4
1	A	169	LYS	4
1	A	135	GLU	4
1	A	180	GLU	4
1	A	153	SER	4
1	A	190	LEU	3
1	A	149	ASP	3
1	A	191	ASP	3
1	A	182	HIS	2
1	A	173	HIS	2
1	A	155	GLN	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	168	ASP	2
1	A	198	SER	2
1	A	133	SER	2
1	A	165	LYS	2
1	A	140	THR	2
1	A	146	SER	2
1	A	194	LYS	2
1	A	148	ARG	1
1	A	147	TYR	1
1	A	139	GLU	1
1	A	152	TYR	1
1	A	192	LYS	1
1	A	183	LEU	1
1	A	197	LEU	1
1	A	158	GLU	1

5.2.3 RNA

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains

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

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

There are no ligands in this entry.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

No chemical shift data were provided