



Full wwPDB NMR Structure Validation Report ⓘ

May 28, 2020 – 11:15 pm BST

PDB ID : 2LP1
Title : The solution NMR structure of the transmembrane C-terminal domain of the amyloid precursor protein (C99)
Authors : Barrett, P.J.; Song, Y.; Van Horn, W.D.; Hustedt, E.J.; Schafer, J.M.; Hadzise-
limovic, A.; Beel, A.J.; Sanders, C.R.
Deposited on : 2012-01-30

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

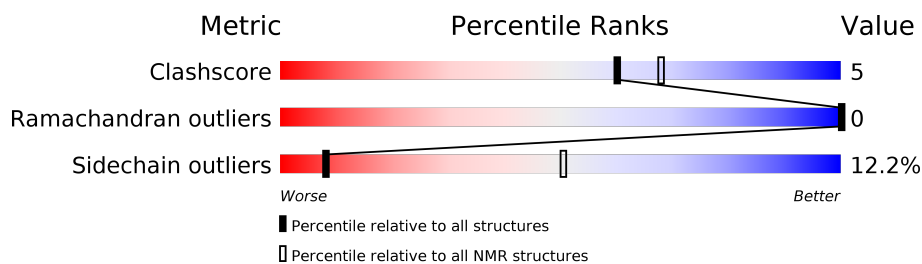
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 43%.

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	122	

2 Ensemble composition and analysis i

This entry contains 30 models. Model 24 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:701-A:723 (23)	0.32	24

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 3 clusters and 1 single-model cluster was found.

Cluster number	Models
1	2, 3, 4, 5, 9, 11, 12, 14, 19, 22, 24, 25, 26, 28, 29, 30
2	6, 8, 10, 13, 16, 20, 23
3	1, 7, 15, 17, 21, 27
Single-model clusters	18

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 732 atoms, of which 384 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called C99.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	46	732	231	384	58	57	2	0

There are 22 discrepancies between the modelled and reference sequences:

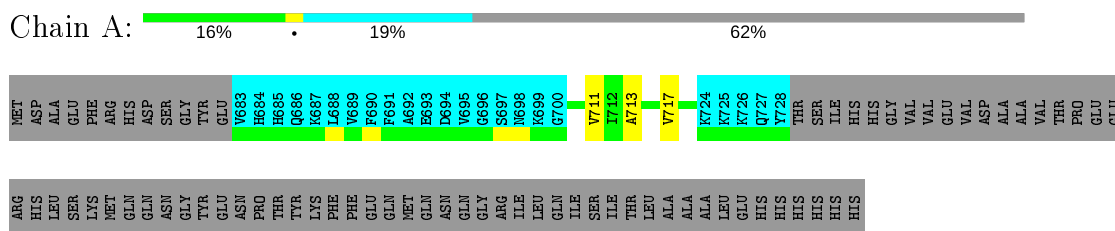
Chain	Residue	Modelled	Actual	Comment	Reference
A	771	GLN	-	EXPRESSION TAG	UNP P05067
A	772	GLY	-	EXPRESSION TAG	UNP P05067
A	773	ARG	-	EXPRESSION TAG	UNP P05067
A	774	ILE	-	EXPRESSION TAG	UNP P05067
A	775	LEU	-	EXPRESSION TAG	UNP P05067
A	776	GLN	-	EXPRESSION TAG	UNP P05067
A	777	ILE	-	EXPRESSION TAG	UNP P05067
A	778	SER	-	EXPRESSION TAG	UNP P05067
A	779	ILE	-	EXPRESSION TAG	UNP P05067
A	780	THR	-	EXPRESSION TAG	UNP P05067
A	781	LEU	-	EXPRESSION TAG	UNP P05067
A	782	ALA	-	EXPRESSION TAG	UNP P05067
A	783	ALA	-	EXPRESSION TAG	UNP P05067
A	784	ALA	-	EXPRESSION TAG	UNP P05067
A	785	LEU	-	EXPRESSION TAG	UNP P05067
A	786	GLU	-	EXPRESSION TAG	UNP P05067
A	787	HIS	-	EXPRESSION TAG	UNP P05067
A	788	HIS	-	EXPRESSION TAG	UNP P05067
A	789	HIS	-	EXPRESSION TAG	UNP P05067
A	790	HIS	-	EXPRESSION TAG	UNP P05067
A	791	HIS	-	EXPRESSION TAG	UNP P05067
A	792	HIS	-	EXPRESSION TAG	UNP P05067

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: C99

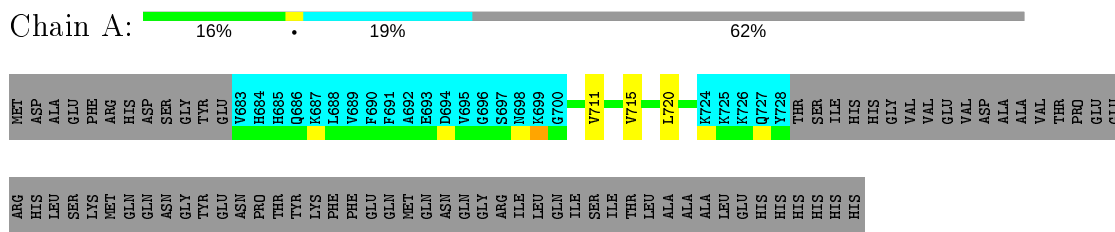


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

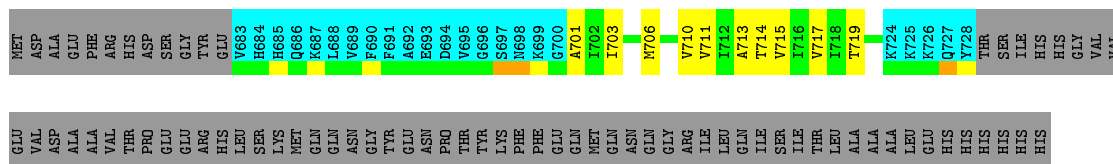
- Molecule 1: C99



4.2.2 Score per residue for model 2

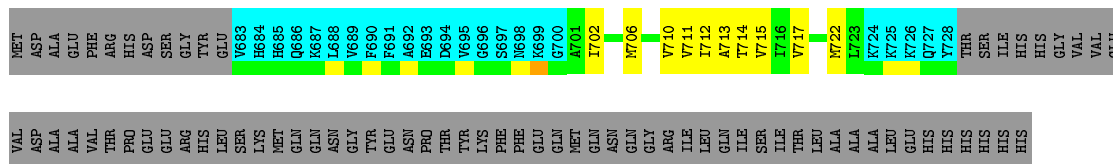
- Molecule 1: C99





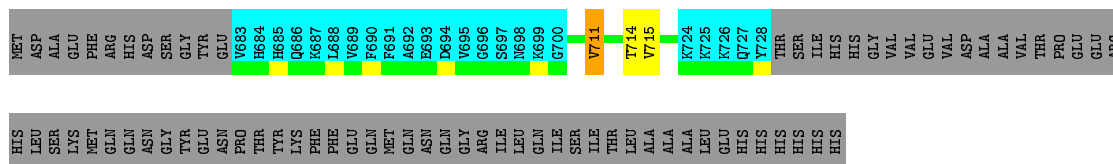
4.2.3 Score per residue for model 3

- Molecule 1: C99



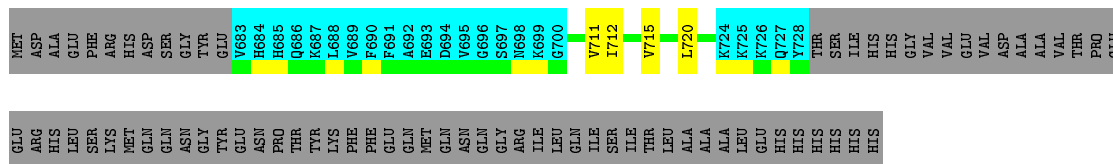
4.2.4 Score per residue for model 4

- Molecule 1: C99



4.2.5 Score per residue for model 5

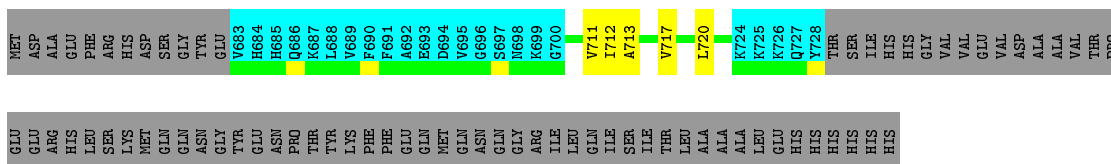
- Molecule 1: C99



4.2.6 Score per residue for model 6

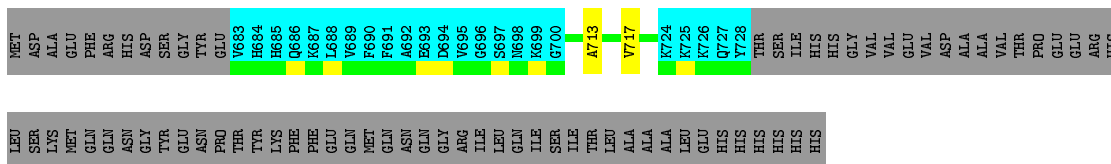
- Molecule 1: C99





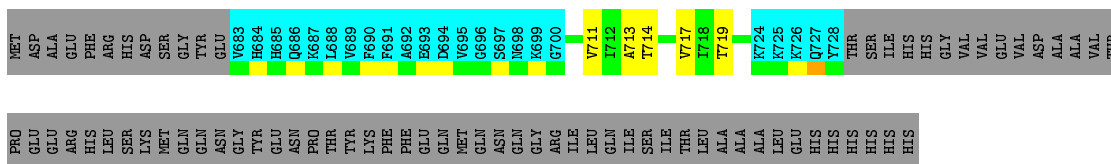
4.2.7 Score per residue for model 7

- Molecule 1: C99



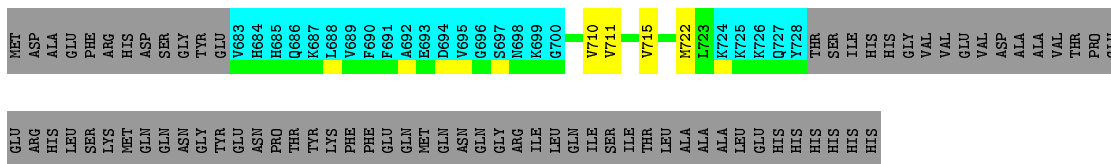
4.2.8 Score per residue for model 8

- Molecule 1: C99



4.2.9 Score per residue for model 9

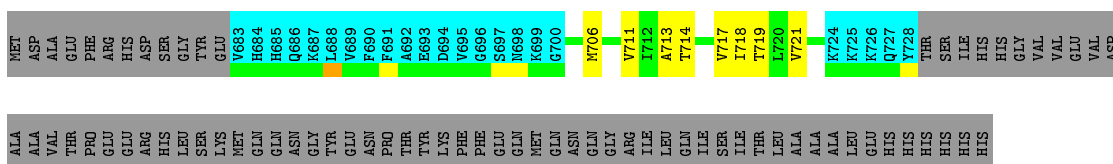
- Molecule 1: C99



4.2.10 Score per residue for model 10

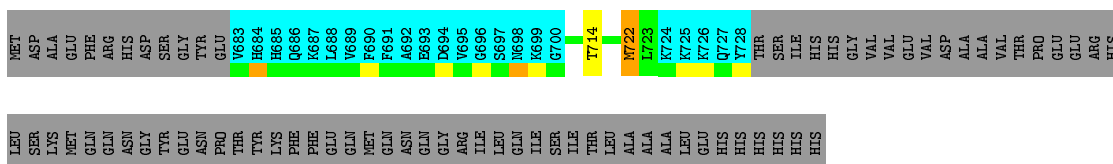
- Molecule 1: C99





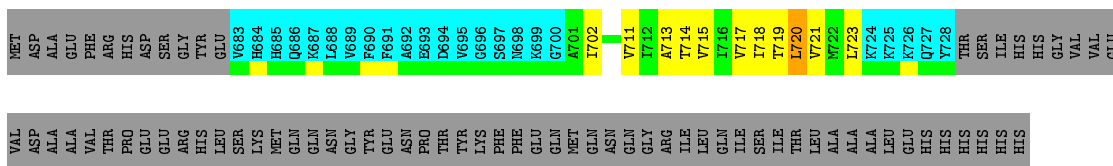
4.2.11 Score per residue for model 11

- Molecule 1: C99



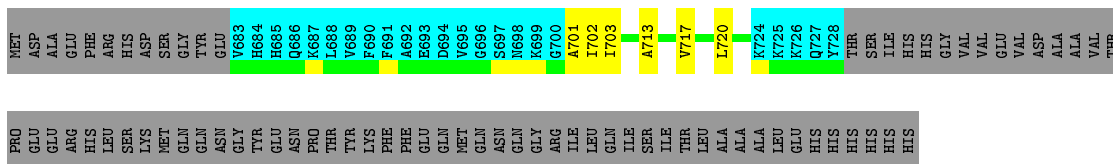
4.2.12 Score per residue for model 12

- Molecule 1: C99



4.2.13 Score per residue for model 13

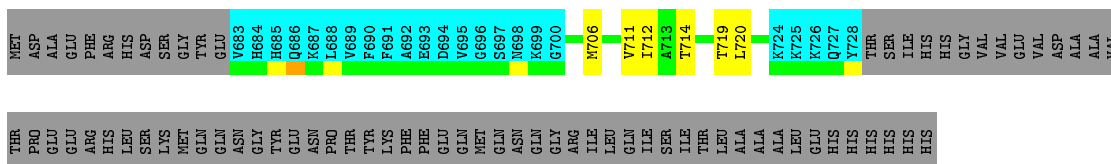
- Molecule 1: C99



4.2.14 Score per residue for model 14

- Molecule 1: C99

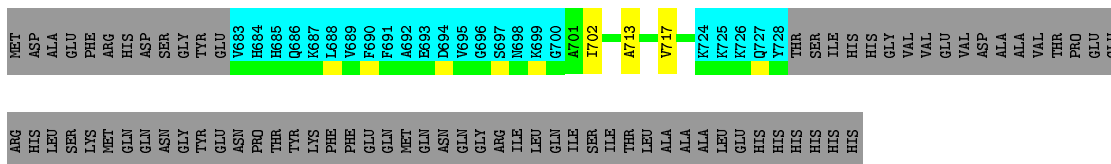




4.2.15 Score per residue for model 15

- Molecule 1: C99

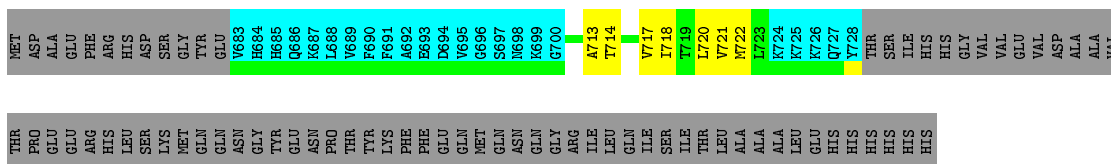
Chain A: 16% 19% 62%



4.2.16 Score per residue for model 16

- Molecule 1: C99

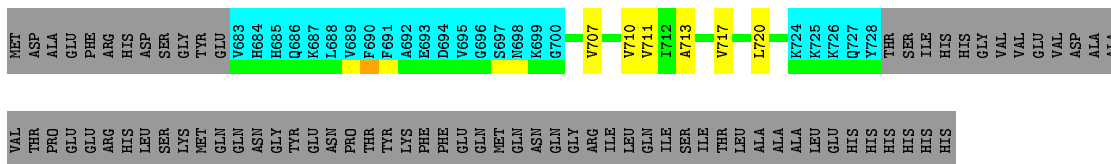
Chain A: 13% 6% 19% 62%



4.2.17 Score per residue for model 17

- Molecule 1: C99

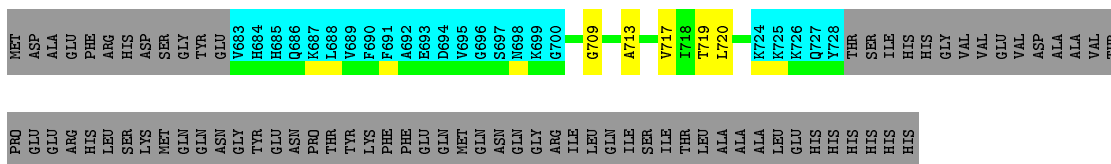
Chain A: 14% 5% 19% 62%



4.2.18 Score per residue for model 18

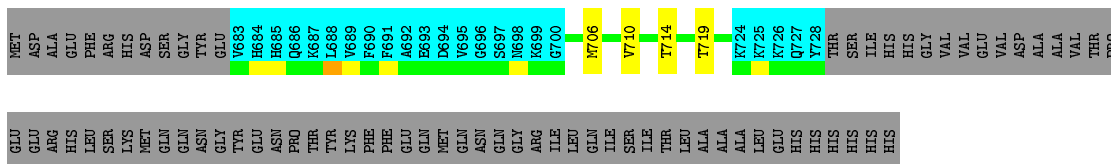
- Molecule 1: C99

Chain A: 15% 19% 62%



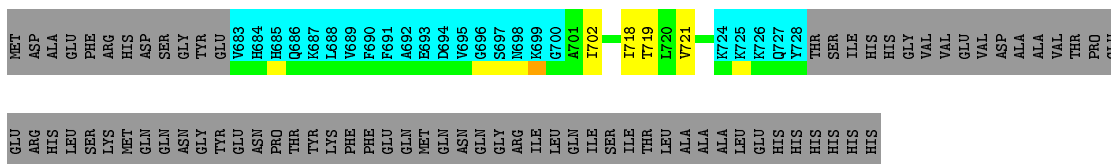
4.2.19 Score per residue for model 19

- Molecule 1: C99



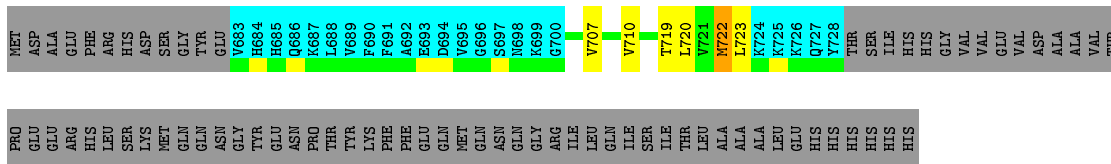
4.2.20 Score per residue for model 20

- Molecule 1: C99



4.2.21 Score per residue for model 21

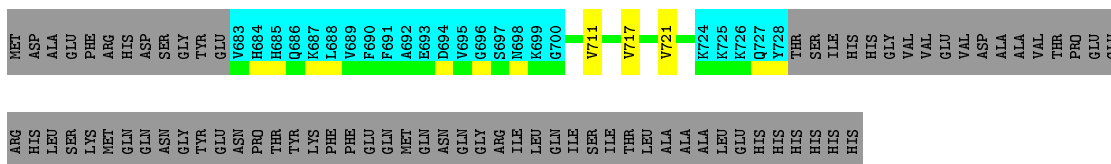
- Molecule 1: C99



4.2.22 Score per residue for model 22

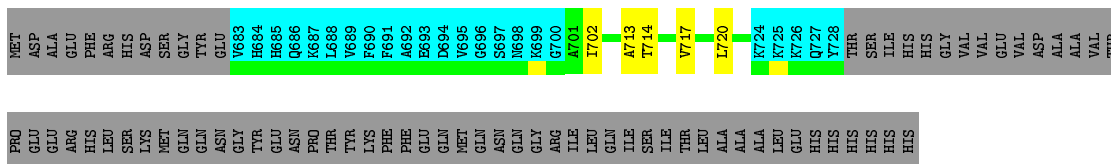
- Molecule 1: C99





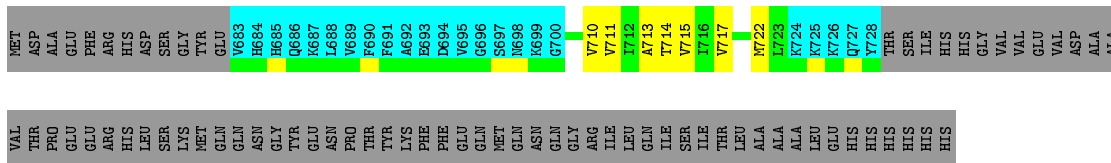
4.2.23 Score per residue for model 23

- Molecule 1: C99



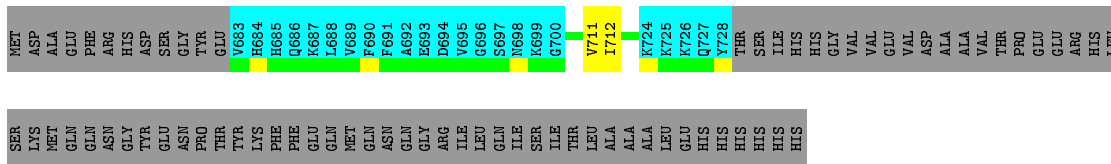
4.2.24 Score per residue for model 24 (medoid)

- Molecule 1: C99



4.2.25 Score per residue for model 25

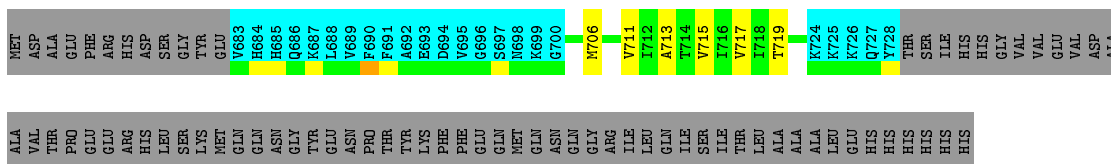
- Molecule 1: C99



4.2.26 Score per residue for model 26

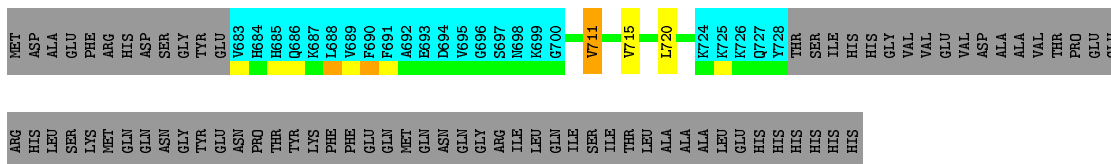
- Molecule 1: C99





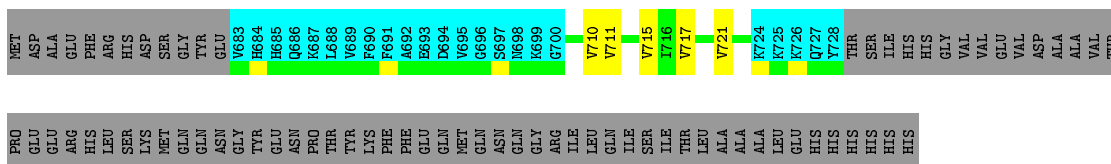
4.2.27 Score per residue for model 27

- Molecule 1: C99



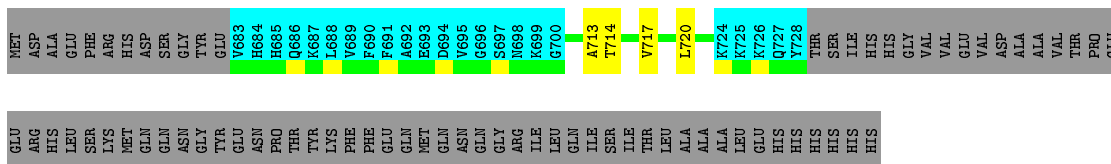
4.2.28 Score per residue for model 28

- Molecule 1: C99



4.2.29 Score per residue for model 29

- Molecule 1: C99



4.2.30 Score per residue for model 30

- Molecule 1: C99



MET	ALA
ASP	VAL
ALA	THR
GLU	PRO
PHE	GLU
ARG	ARG
HIS	HIS
ASP	LEU
SER	SER
GLY	LYS
TYR	TYR
GLU	MET
1683	GLN
1684	GLN
1685	ASN
1686	GLY
1687	TYR
1688	GLU
1689	ASN
1690	PRO
1691	THR
1692	TYR
1693	LYS
1694	PHE
1695	PHE
1696	GLU
1697	GLN
1698	MET
1699	GLN
1700	ASN
1706	GLN
1711	GLY
1712	ANG
1713	ILE
1714	LEU
1715	GLN
1716	ILE
1717	SER
1720	ILE
1724	THR
1725	LEU
1727	ALA
1728	ALA
	ALA
	LEU
	GLU
	HIS
	HIS
	HIS
	HIS
	HIS
	THR
	SER
	ILE
	ILE
	HIS
	HIS
	GLY
	VAL
	VAL
	VAL
	GLU
	VAL
	ASP
	ALA

5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 30 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
X-PLOR NIH	structure solution	2.24
X-PLOR NIH	refinement	2.24
TALOS	geometry optimization	Linux9
TALOS	structure solution	Linux9
ProcheckNMR	structure solution	3.5.4

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	179
Number of shifts mapped to atoms	179
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	43%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

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

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	158	193	193	2±1
All	All	4740	5790	5790	49

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 5.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:720:LEU:HD13	1:A:723:LEU:HD12	0.79	1.52	12	1
1:A:707:VAL:O	1:A:710:VAL:HG23	0.61	1.96	21	2
1:A:711:VAL:O	1:A:715:VAL:HG23	0.55	2.01	2	12
1:A:717:VAL:O	1:A:721:VAL:HG23	0.55	2.01	22	2
1:A:720:LEU:CD1	1:A:723:LEU:HD12	0.54	2.29	12	1
1:A:713:ALA:O	1:A:717:VAL:HG23	0.51	2.05	23	17
1:A:718:ILE:O	1:A:721:VAL:HG22	0.49	2.07	12	4
1:A:722:MET:N	1:A:722:MET:SD	0.48	2.86	11	1
1:A:709:GLY:O	1:A:713:ALA:HB2	0.45	2.12	18	1
1:A:703:ILE:HD12	1:A:703:ILE:N	0.44	2.26	2	1
1:A:711:VAL:HG13	1:A:712:ILE:N	0.44	2.28	14	3
1:A:719:THR:HG22	1:A:723:LEU:HD12	0.44	1.87	21	1
1:A:710:VAL:HG13	1:A:711:VAL:N	0.42	2.29	2	1
1:A:722:MET:SD	1:A:722:MET:C	0.41	2.98	21	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:710:VAL:O	1:A:714:THR:HG23	0.41	2.16	2	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	23/122 (19%)	23±0 (99±2%)	0±0 (1±2%)	0±0 (0±0%)	100	100
All	All	690/3660 (19%)	685 (99%)	5 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	18/104 (17%)	16±1 (88±6%)	2±1 (12±6%)	8	50
All	All	540/3120 (17%)	474 (88%)	66 (12%)	8	50

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	720	LEU	14
1	A	714	THR	12
1	A	719	THR	9
1	A	711	VAL	8
1	A	706	MET	7
1	A	722	MET	6
1	A	710	VAL	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	702	ILE	2
1	A	712	ILE	2
1	A	703	ILE	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 carbohydrates 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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 43% for the well-defined parts and 32% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping [i](#)

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	179
Number of shifts mapped to atoms	179
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 [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
¹³ C _α	37	-0.71 \pm 0.18	Should be applied
¹³ C _β	30	1.27 \pm 0.07	Should be applied
¹³ C'	38	-0.36 \pm 0.13	None needed (< 0.5 ppm)
¹⁵ N	37	0.57 \pm 0.28	Should be applied

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 43%, i.e. 109 atoms were assigned a chemical shift out of a possible 251. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	91/115 (79%)	23/46 (50%)	45/46 (98%)	23/23 (100%)
Sidechain	18/136 (13%)	0/74 (0%)	18/62 (29%)	0/0 (—%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Overall	109/251 (43%)	23/120 (19%)	63/108 (58%)	23/23 (100%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 32%, i.e. 179 atoms were assigned a chemical shift out of a possible 556. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	149/230 (65%)	37/92 (40%)	75/92 (82%)	37/46 (80%)
Sidechain	30/286 (10%)	0/163 (0%)	30/115 (26%)	0/8 (0%)
Aromatic	0/40 (0%)	0/22 (0%)	0/16 (0%)	0/2 (0%)
Overall	179/556 (32%)	37/277 (13%)	105/223 (47%)	37/56 (66%)

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 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.

Random coil index (RCI) for chain A:

