Full wwPDB NMR Structure Validation Report (i)

Aug 17, 2022 – 04:31 PM EDT

PDB ID	:	2MY7
Title	:	NMR Structure of unfolding intermediate state of RRM-3 domain of ETR-3
Authors	:	Bhatt, H.P.; Ganguly, A.K.; Bhavesh, N.S.
Deposited on	:	2015-01-21

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:

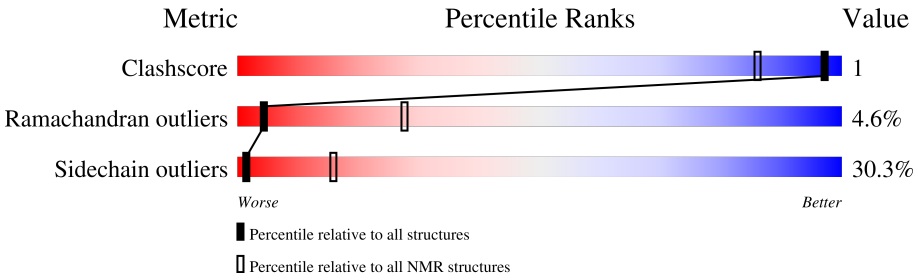
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 83%.

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	97	57% 22% 22%

2 Ensemble composition and analysis (i)

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: *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:398-A:437, A:448-A:483 (76)	0.39	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 3 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1531 atoms, of which 764 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CUGBP Elav-like family member 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	97	1531	490	764	131	140	6	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP O95319
A	-2	SER	-	expression tag	UNP O95319
A	-1	HIS	-	expression tag	UNP O95319
A	0	MET	-	expression tag	UNP O95319

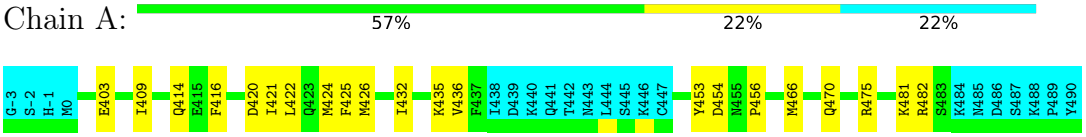
4 Residue-property plots

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: CUGBP Elav-like family member 2

Chain A:



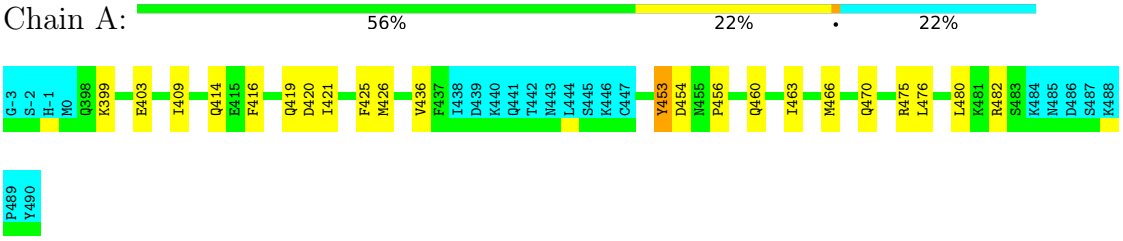
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: CUGBP Elav-like family member 2

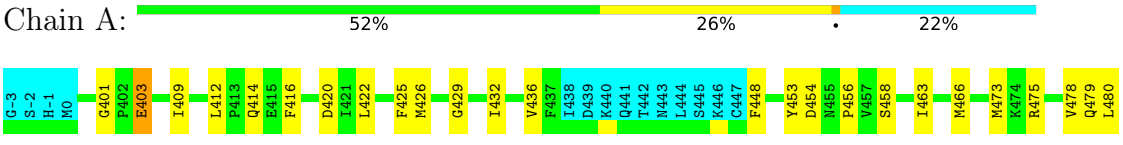
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: CUGBP Elav-like family member 2

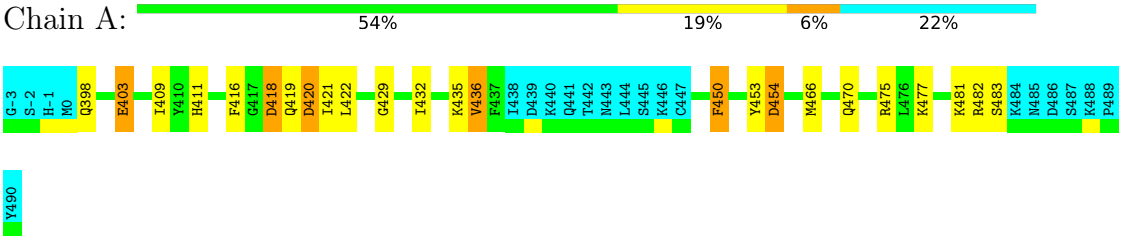
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: CUGBP Elav-like family member 2

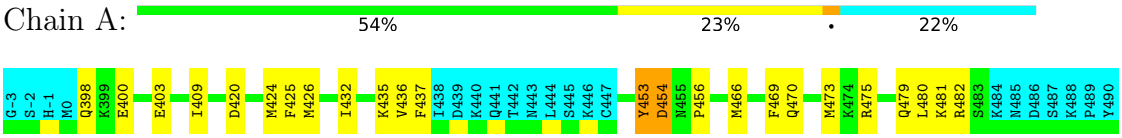
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: CUGBP Elav-like family member 2

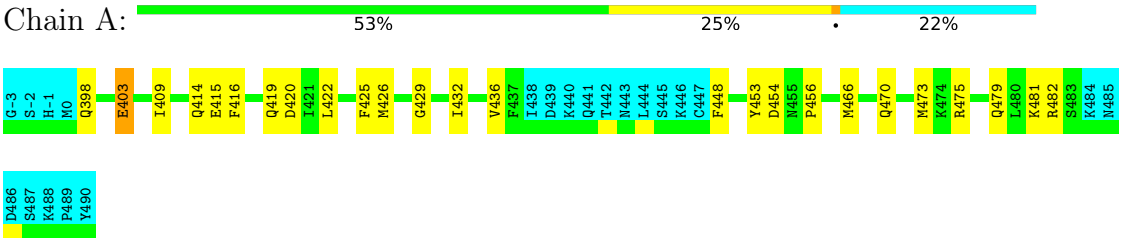
Chain A:



4.2.5 Score per residue for model 5 (medoid)

• Molecule 1: CUGBP Elav-like family member 2

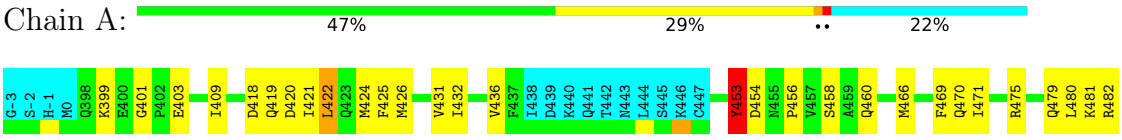
Chain A:

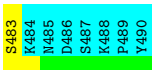


4.2.6 Score per residue for model 6

• Molecule 1: CUGBP Elav-like family member 2

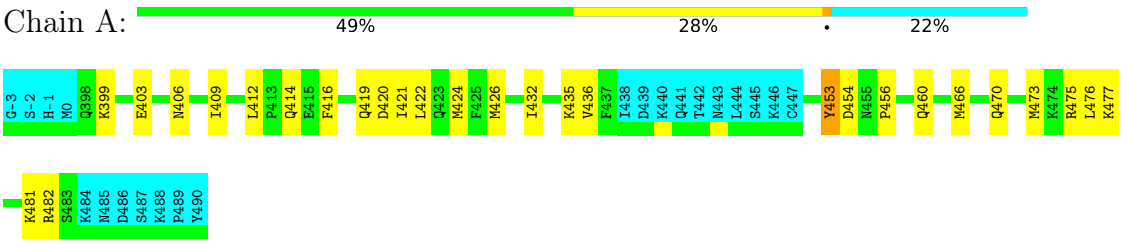
Chain A:





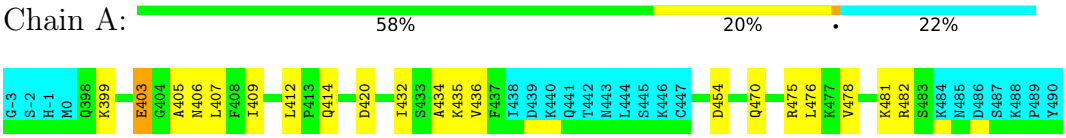
4.2.7 Score per residue for model 7

• Molecule 1: CUGBP Elav-like family member 2



4.2.8 Score per residue for model 8

• Molecule 1: CUGBP Elav-like family member 2

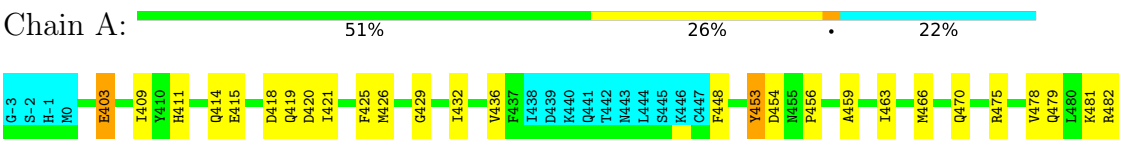


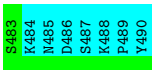
4.2.9 Score per residue for model 9

• Molecule 1: CUGBP Elav-like family member 2

4.2.10 Score per residue for model 10

• Molecule 1: CUGBP Elav-like family member 2

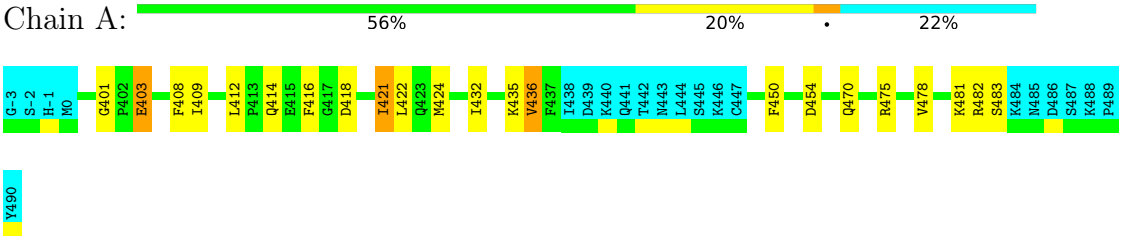




4.2.11 Score per residue for model 11

• Molecule 1: CUGBP Elav-like family member 2

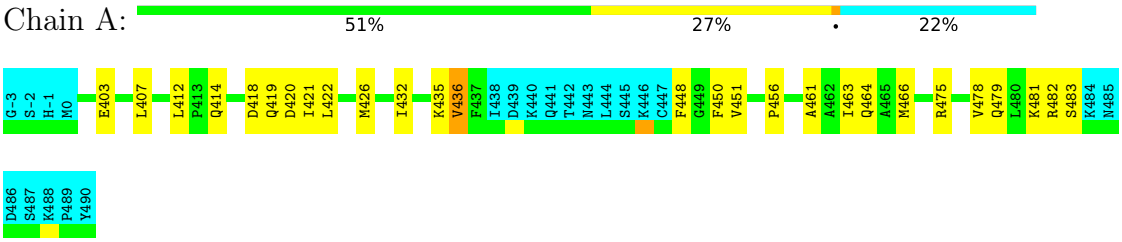
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: CUGBP Elav-like family member 2

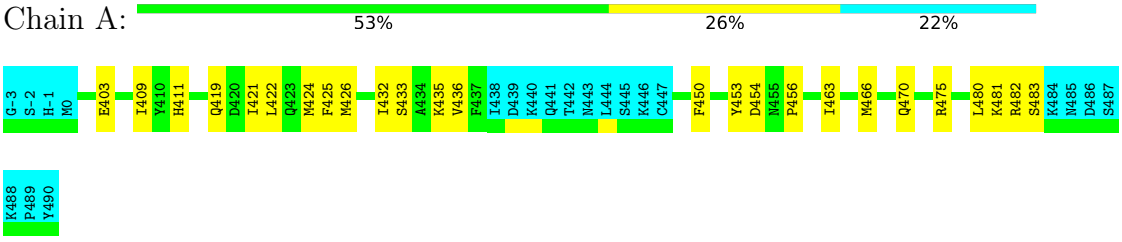
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: CUGBP Elav-like family member 2

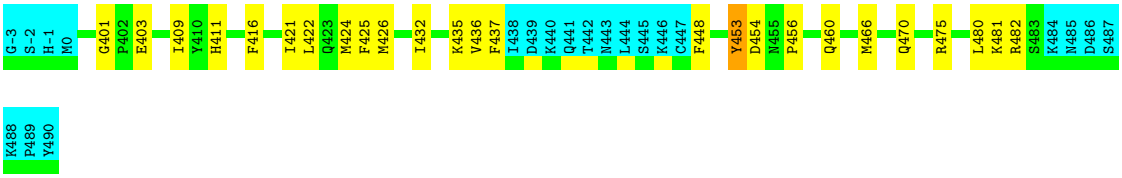
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: CUGBP Elav-like family member 2

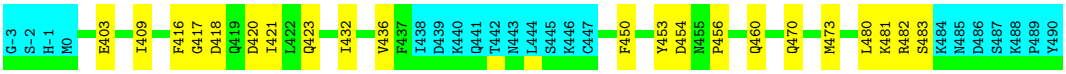
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: CUGBP Elav-like family member 2

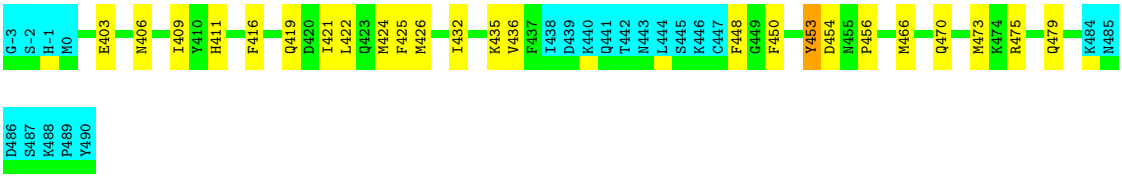
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: CUGBP Elav-like family member 2

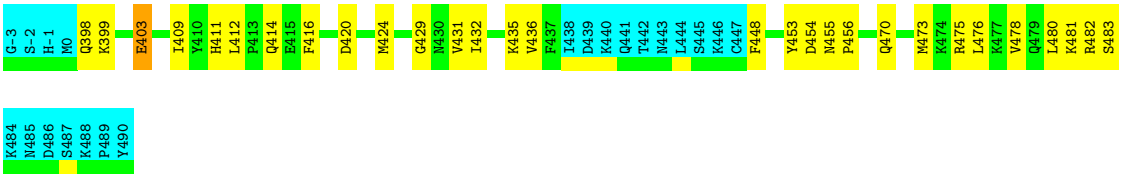
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: CUGBP Elav-like family member 2

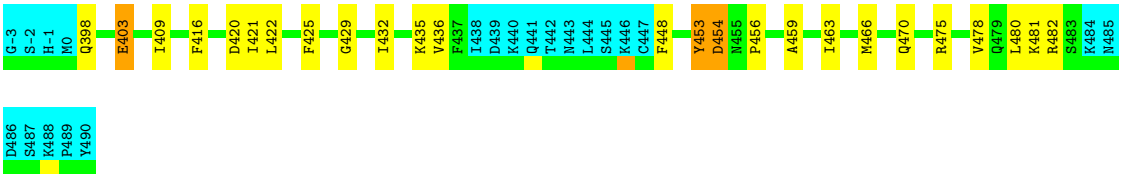
Chain A:



4.2.18 Score per residue for model 18

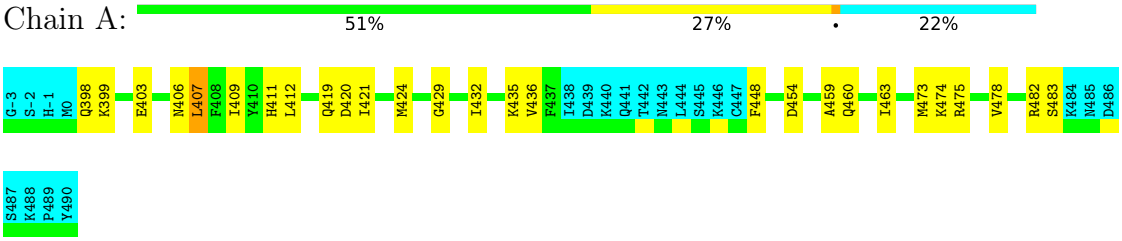
• Molecule 1: CUGBP Elav-like family member 2

Chain A:



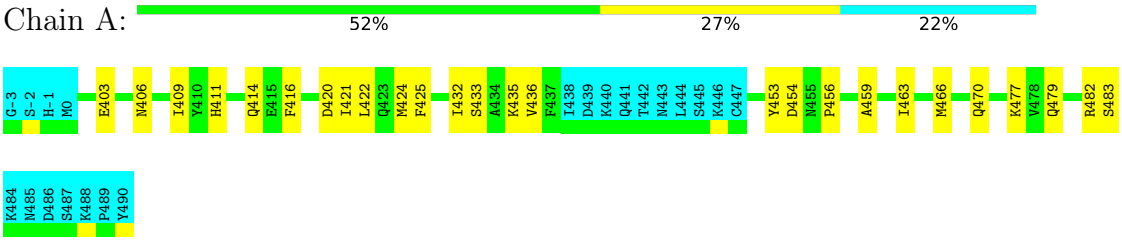
4.2.19 Score per residue for model 19

• Molecule 1: CUGBP Elav-like family member 2



4.2.20 Score per residue for model 20

• Molecule 1: CUGBP Elav-like family member 2



5 Refinement protocol and experimental data overview

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

Of the 100 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 solution	
CYANA	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	1063
Number of shifts mapped to atoms	1063
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

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	A	1.39±0.02	0±1/616 (0.1± 0.1%)	1.20±0.03	2±1/828 (0.2± 0.1%)
All	All	1.39	10/12320 (0.1%)	1.20	35/16560 (0.2%)

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	A	0.0±0.0	0.3±0.5
All	All	0	7

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	407	LEU	CA-CB	5.56	1.66	1.53	19	1
1	A	429	GLY	CA-C	5.46	1.60	1.51	3	8
1	A	453	TYR	C-N	5.11	1.45	1.34	5	1

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	A	416	PHE	CB-CG-CD1	7.19	125.83	120.80	14	8
1	A	403	GLU	N-CA-CB	-6.33	99.21	110.60	11	9
1	A	453	TYR	CB-CG-CD1	6.06	124.64	121.00	9	8
1	A	416	PHE	CB-CG-CD2	-5.91	116.66	120.80	7	3
1	A	453	TYR	CA-CB-CG	5.57	123.99	113.40	6	2
1	A	453	TYR	N-CA-CB	-5.54	100.63	110.60	5	3
1	A	461	ALA	CB-CA-C	5.28	118.02	110.10	12	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	434	ALA	CB-CA-C	-5.19	102.32	110.10	8	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	A	416	PHE	Peptide	3
1	A	453	TYR	Sidechain	3
1	A	399	LYS	Peptide	1

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	601	597	596	1±1
All	All	12020	11940	11920	15

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:418:ASP:HB3	1:A:436:VAL:HG22	0.57	1.77	12	3
1:A:405:ALA:HB1	1:A:407:LEU:HD21	0.50	1.84	8	1
1:A:459:ALA:O	1:A:463:ILE:HG12	0.49	2.07	18	3
1:A:398:GLN:HB2	1:A:450:PHE:CZ	0.46	2.44	3	1
1:A:469:PHE:CE2	1:A:471:ILE:HD11	0.45	2.46	6	1
1:A:417:GLY:O	1:A:421:ILE:HB	0.45	2.12	15	1
1:A:459:ALA:O	1:A:463:ILE:HG13	0.44	2.13	19	1
1:A:422:LEU:HB3	1:A:431:VAL:HG11	0.42	1.91	6	1
1:A:416:PHE:CE2	1:A:421:ILE:HA	0.42	2.50	11	1
1:A:416:PHE:HA	1:A:420:ASP:OD1	0.41	2.16	3	1
1:A:453:TYR:CE2	1:A:458:SER:HB2	0.40	2.51	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	76/97 (78%)	67±2 (88±2%)	6±2 (8±2%)	4±1 (5±1%)	4 27
All	All	1520/1940 (78%)	1335 (88%)	115 (8%)	70 (5%)	4 27

All 6 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	403	GLU	20
1	A	454	ASP	19
1	A	456	PRO	16
1	A	483	SER	9
1	A	401	GLY	4
1	A	481	LYS	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	A	63/83 (76%)	44±3 (70±4%)	19±3 (30±4%)	1 16
All	All	1260/1660 (76%)	878 (70%)	382 (30%)	1 16

All 51 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	436	VAL	20
1	A	409	ILE	19
1	A	482	ARG	19
1	A	432	ILE	19

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	475	ARG	18
1	A	470	GLN	17
1	A	420	ASP	15
1	A	466	MET	14
1	A	421	ILE	13
1	A	422	LEU	13
1	A	435	LYS	13
1	A	481	LYS	12
1	A	414	GLN	11
1	A	425	PHE	11
1	A	426	MET	11
1	A	424	MET	11
1	A	419	GLN	10
1	A	448	PHE	10
1	A	480	LEU	9
1	A	478	VAL	9
1	A	453	TYR	8
1	A	412	LEU	8
1	A	473	MET	8
1	A	479	GLN	8
1	A	411	HIS	8
1	A	399	LYS	6
1	A	460	GLN	6
1	A	450	PHE	6
1	A	406	ASN	6
1	A	476	LEU	5
1	A	398	GLN	5
1	A	463	ILE	4
1	A	418	ASP	4
1	A	454	ASP	3
1	A	477	LYS	3
1	A	437	PHE	2
1	A	415	GLU	2
1	A	407	LEU	2
1	A	433	SER	2
1	A	400	GLU	1
1	A	469	PHE	1
1	A	458	SER	1
1	A	452	SER	1
1	A	408	PHE	1
1	A	451	VAL	1
1	A	464	GLN	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	423	GLN	1
1	A	431	VAL	1
1	A	455	ASN	1
1	A	474	LYS	1
1	A	483	SER	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 83% for the well-defined parts and 81% 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	1063
Number of shifts mapped to atoms	1063
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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	97	0.41 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	89	0.15 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	97	0.65 ± 0.11	Should be applied
^{15}N	92	-0.04 ± 0.45	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 83%, i.e. 788 atoms were assigned a chemical shift out of a possible 955. 4 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	372/372 (100%)	148/148 (100%)	152/152 (100%)	72/72 (100%)
Sidechain	374/487 (77%)	218/289 (75%)	144/175 (82%)	12/23 (52%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	42/96 (44%)	21/52 (40%)	21/42 (50%)	0/2 (0%)
Overall	788/955 (83%)	387/489 (79%)	317/369 (86%)	84/97 (87%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 81%, i.e. 983 atoms were assigned a chemical shift out of a possible 1215. 4 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	473/475 (100%)	187/189 (99%)	194/194 (100%)	92/92 (100%)
Sidechain	464/628 (74%)	271/374 (72%)	178/224 (79%)	15/30 (50%)
Aromatic	46/112 (41%)	23/60 (38%)	23/48 (48%)	0/4 (0%)
Overall	983/1215 (81%)	481/623 (77%)	395/466 (85%)	107/126 (85%)

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: