



Full wwPDB NMR Structure Validation Report ⓘ

Nov 9, 2020 – 06:08 PM GMT

PDB ID : 6T3I
Title : Solution structure of the HRP2 IBD
Authors : Veverka, V.
Deposited on : 2019-10-11

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 : 2.14.6
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.14.6

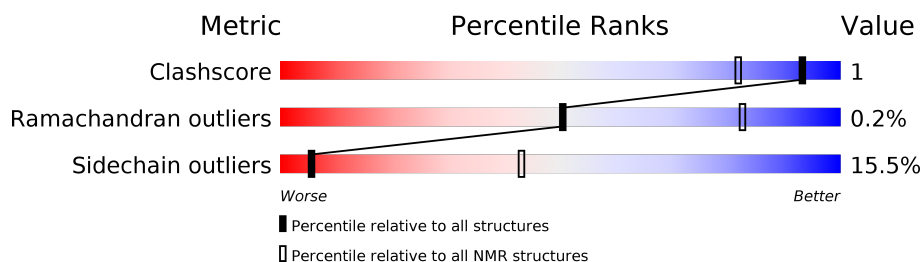
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 89%.

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	86	

2 Ensemble composition and analysis ⓘ

This entry contains 30 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: *lowest energy*.

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:470-A:548 (79)	0.09	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 4 clusters and 6 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1401 atoms, of which 724 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hepatoma-derived growth factor-related protein 2.

Mol	Chain	Residues	Atoms						Trace
1	A	86	Total	C	H	N	O	S	0
			1401	423	724	122	130	2	

There are 5 discrepancies between the modelled and reference sequences:

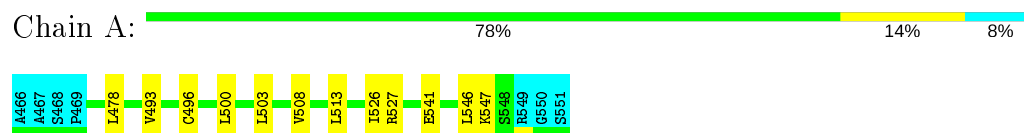
Chain	Residue	Modelled	Actual	Comment	Reference
A	466	ALA	-	expression tag	UNP Q7Z4V5
A	467	ALA	-	expression tag	UNP Q7Z4V5
A	468	SER	-	expression tag	UNP Q7Z4V5
A	550	GLY	-	expression tag	UNP Q7Z4V5
A	551	SER	-	expression tag	UNP Q7Z4V5

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: Hepatoma-derived growth factor-related protein 2

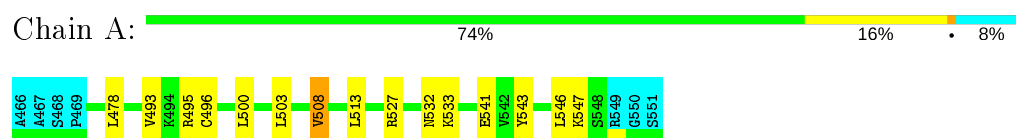


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

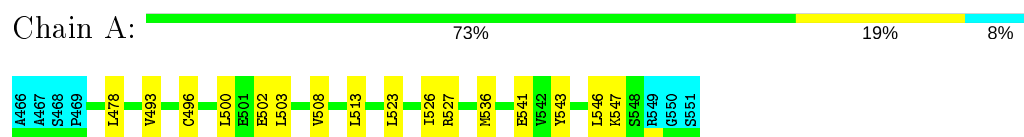
4.2.1 Score per residue for model 1

- Molecule 1: Hepatoma-derived growth factor-related protein 2



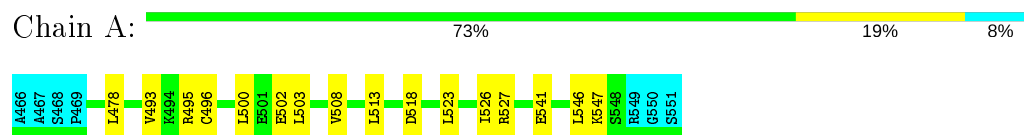
4.2.2 Score per residue for model 2

- Molecule 1: Hepatoma-derived growth factor-related protein 2



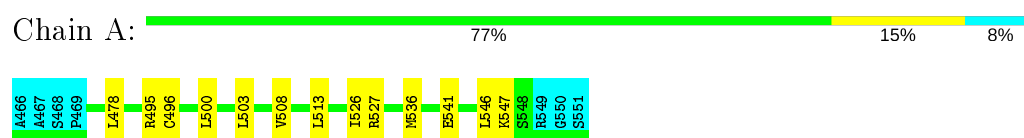
4.2.3 Score per residue for model 3

- Molecule 1: Hepatoma-derived growth factor-related protein 2



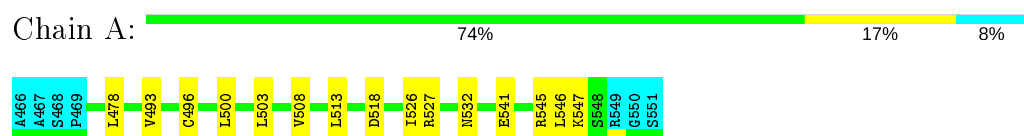
4.2.4 Score per residue for model 4

- Molecule 1: Hepatoma-derived growth factor-related protein 2



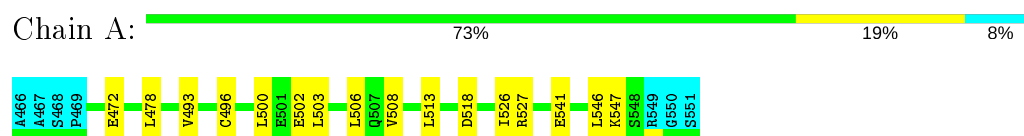
4.2.5 Score per residue for model 5 (medoid)

- Molecule 1: Hepatoma-derived growth factor-related protein 2



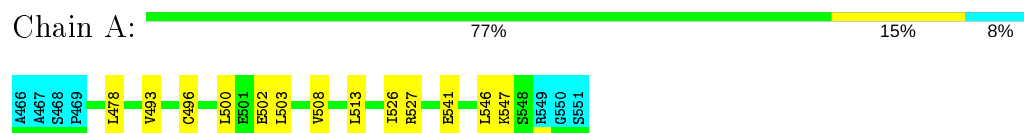
4.2.6 Score per residue for model 6

- Molecule 1: Hepatoma-derived growth factor-related protein 2



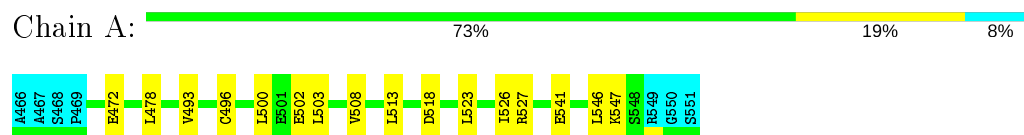
4.2.7 Score per residue for model 7

- Molecule 1: Hepatoma-derived growth factor-related protein 2



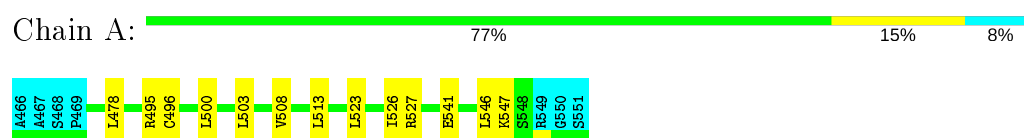
4.2.8 Score per residue for model 8

- Molecule 1: Hepatoma-derived growth factor-related protein 2



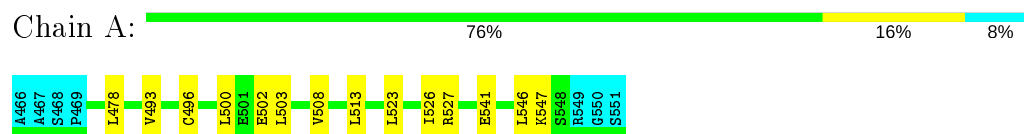
4.2.9 Score per residue for model 9

- Molecule 1: Hepatoma-derived growth factor-related protein 2



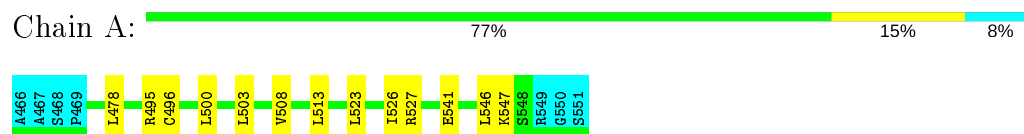
4.2.10 Score per residue for model 10

- Molecule 1: Hepatoma-derived growth factor-related protein 2



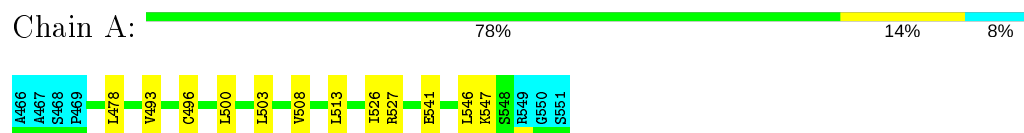
4.2.11 Score per residue for model 11

- Molecule 1: Hepatoma-derived growth factor-related protein 2



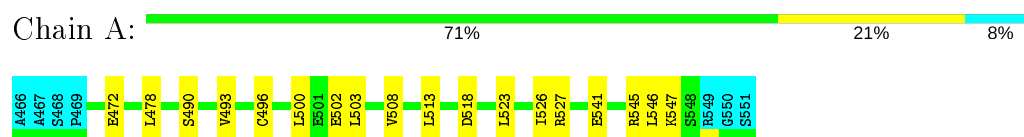
4.2.12 Score per residue for model 12

- Molecule 1: Hepatoma-derived growth factor-related protein 2



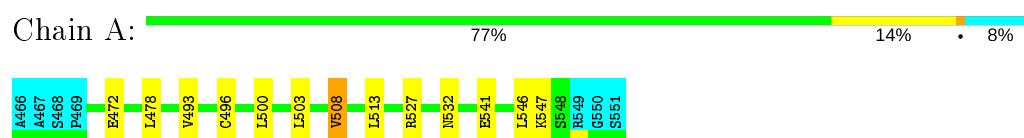
4.2.13 Score per residue for model 13

- Molecule 1: Hepatoma-derived growth factor-related protein 2



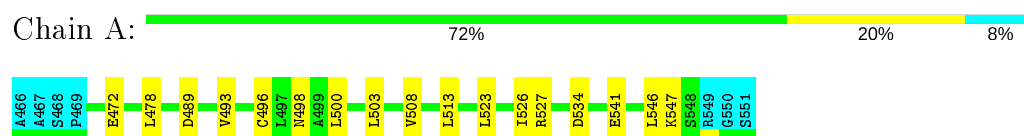
4.2.14 Score per residue for model 14

- Molecule 1: Hepatoma-derived growth factor-related protein 2



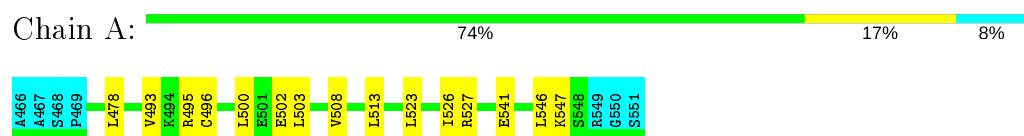
4.2.15 Score per residue for model 15

- Molecule 1: Hepatoma-derived growth factor-related protein 2



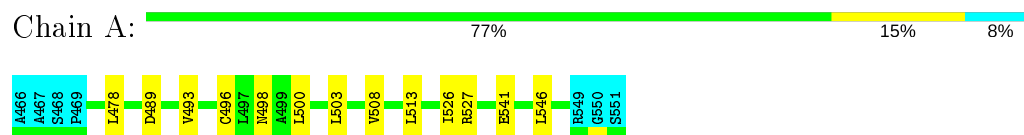
4.2.16 Score per residue for model 16

- Molecule 1: Hepatoma-derived growth factor-related protein 2



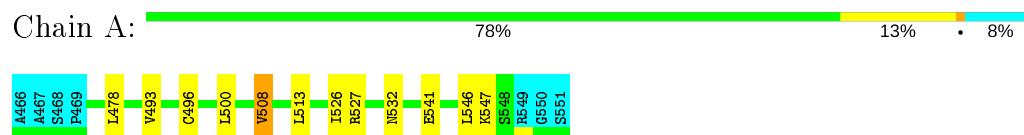
4.2.17 Score per residue for model 17

- Molecule 1: Hepatoma-derived growth factor-related protein 2



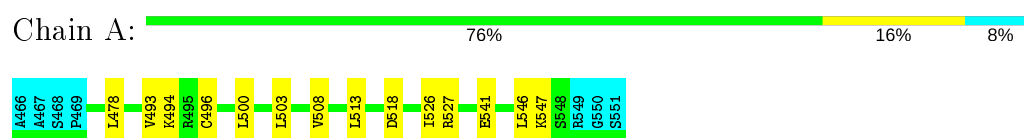
4.2.18 Score per residue for model 18

- Molecule 1: Hepatoma-derived growth factor-related protein 2



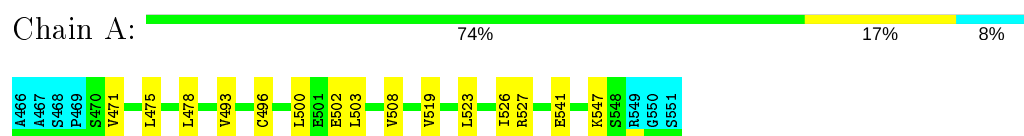
4.2.19 Score per residue for model 19

- Molecule 1: Hepatoma-derived growth factor-related protein 2



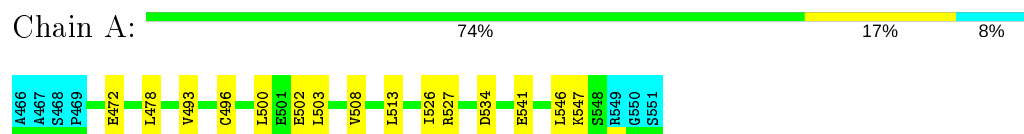
4.2.20 Score per residue for model 20

- Molecule 1: Hepatoma-derived growth factor-related protein 2



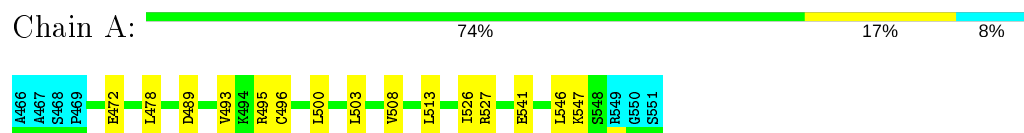
4.2.21 Score per residue for model 21

- Molecule 1: Hepatoma-derived growth factor-related protein 2



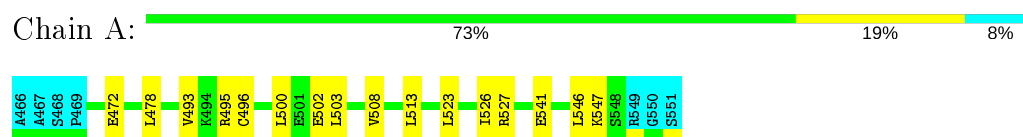
4.2.22 Score per residue for model 22

- Molecule 1: Hepatoma-derived growth factor-related protein 2



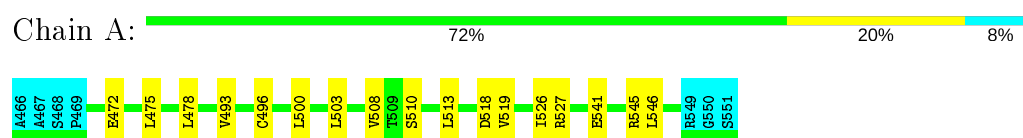
4.2.23 Score per residue for model 23

- Molecule 1: Hepatoma-derived growth factor-related protein 2



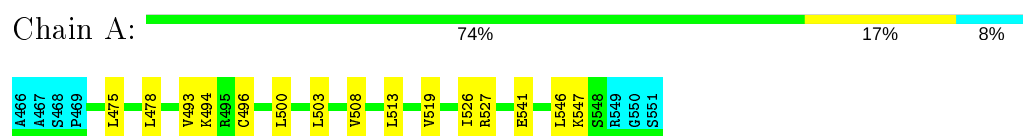
4.2.24 Score per residue for model 24

- Molecule 1: Hepatoma-derived growth factor-related protein 2



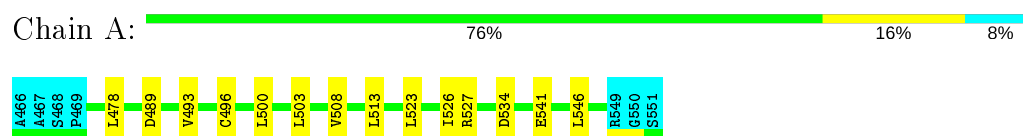
4.2.25 Score per residue for model 25

- Molecule 1: Hepatoma-derived growth factor-related protein 2



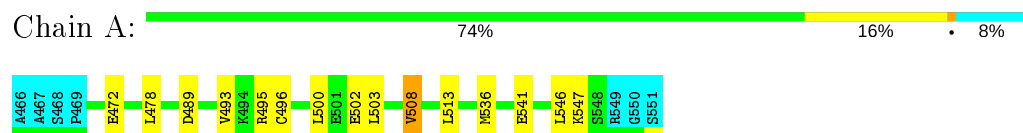
4.2.26 Score per residue for model 26

- Molecule 1: Hepatoma-derived growth factor-related protein 2



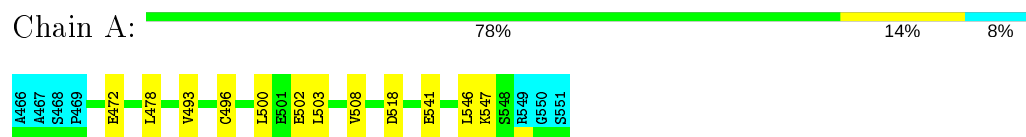
4.2.27 Score per residue for model 27

- Molecule 1: Hepatoma-derived growth factor-related protein 2



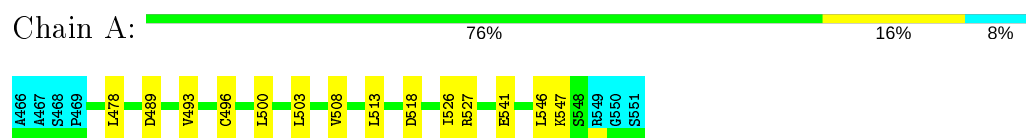
4.2.28 Score per residue for model 28

- Molecule 1: Hepatoma-derived growth factor-related protein 2



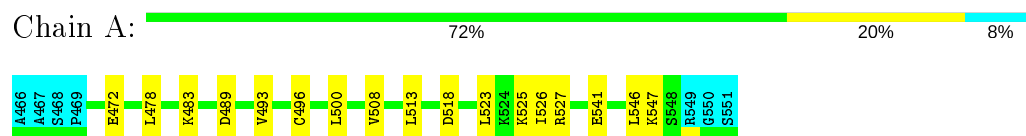
4.2.29 Score per residue for model 29

- Molecule 1: Hepatoma-derived growth factor-related protein 2



4.2.30 Score per residue for model 30

- Molecule 1: Hepatoma-derived growth factor-related protein 2



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 30 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	
YASARA	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	1098
Number of shifts mapped to atoms	1098
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

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

6 Model quality [i](#)

6.1 Standard geometry [i](#)

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	0.62±0.01	0±0/637 (0.0± 0.0%)	0.77±0.01	0±0/854 (0.0± 0.1%)
All	All	0.62	0/19110 (0.0%)	0.77	10/25620 (0.0%)

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	495	ARG	NE-CZ-NH1	5.79	123.20	120.30	4	9
1	A	545	ARG	NE-CZ-NH1	5.04	122.82	120.30	5	1

There are no chirality outliers.

There are no planarity outliers.

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	632	679	679	2±1
All	All	18960	20370	20370	50

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:508:VAL:HG21	1:A:546:LEU:HD22	0.76	1.58	15	28
1:A:475:LEU:HD21	1:A:519:VAL:HG21	0.53	1.80	20	3
1:A:508:VAL:CG2	1:A:546:LEU:HD22	0.50	2.37	23	2
1:A:523:LEU:HD23	1:A:526:ILE:HD11	0.49	1.84	26	13
1:A:508:VAL:O	1:A:508:VAL:HG23	0.43	2.13	20	2
1:A:543:TYR:CE1	1:A:547:LYS:HD3	0.41	2.50	2	2

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	79/86 (92%)	78±1 (98±1%)	1±1 (2±1%)	0±0 (0±0%)	50	82
All	All	2370/2580 (92%)	2329 (98%)	37 (2%)	4 (0%)	50	82

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	508	VAL	4

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	72/76 (95%)	61±1 (85±2%)	11±1 (15±2%)	5	43
All	All	2160/2280 (95%)	1826 (85%)	334 (15%)	5	43

All 27 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	496	CYS	30
1	A	541	GLU	30
1	A	478	LEU	30
1	A	500	LEU	30
1	A	513	LEU	28
1	A	503	LEU	28
1	A	527	ARG	28
1	A	493	VAL	27
1	A	547	LYS	25
1	A	502	GLU	13
1	A	526	ILE	13
1	A	472	GLU	12
1	A	518	ASP	10
1	A	489	ASP	7
1	A	532	ASN	4
1	A	536	MET	3
1	A	534	ASP	3
1	A	545	ARG	2
1	A	494	LYS	2
1	A	498	ASN	2
1	A	533	LYS	1
1	A	525	LYS	1
1	A	506	LEU	1
1	A	510	SER	1
1	A	483	LYS	1
1	A	490	SER	1
1	A	471	VAL	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 89% for the well-defined parts and 89% for the entire structure.

7.1 Chemical shift list 1

File name: `working_cs.cif`

Chemical shift list name: *starch_output*

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	1098
Number of shifts mapped to atoms	1098
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	86	-0.48 ± 0.07	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	83	0.23 ± 0.10	None needed (< 0.5 ppm)
$^{13}\text{C}'$	86	-0.34 ± 0.05	None needed (< 0.5 ppm)
^{15}N	84	0.34 ± 0.31	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 89%, i.e. 915 atoms were assigned a chemical shift out of a possible 1023. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	393/393 (100%)	157/157 (100%)	158/158 (100%)	78/78 (100%)
Sidechain	501/598 (84%)	311/348 (89%)	189/219 (86%)	1/31 (3%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	21/32 (66%)	14/17 (82%)	7/14 (50%)	0/1 (0%)
Overall	915/1023 (89%)	482/522 (92%)	354/391 (91%)	79/110 (72%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 89%, i.e. 972 atoms were assigned a chemical shift out of a possible 1093. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	424/426 (100%)	168/170 (99%)	172/172 (100%)	84/84 (100%)
Sidechain	527/635 (83%)	327/371 (88%)	199/230 (87%)	1/34 (3%)
Aromatic	21/32 (66%)	14/17 (82%)	7/14 (50%)	0/1 (0%)
Overall	972/1093 (89%)	509/558 (91%)	378/416 (91%)	85/119 (71%)

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	483	LYS	HE3	1.89	3.86 – 1.96	-5.4

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:

