WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:05 pm BST

PDB ID	:	2ITH
Title	:	NMR Structure of Haloferax volcanii DHFR
Authors	:	Binbuga, B.
Deposited on	:	2006-10-19

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 (i)) were used in the production of this report:

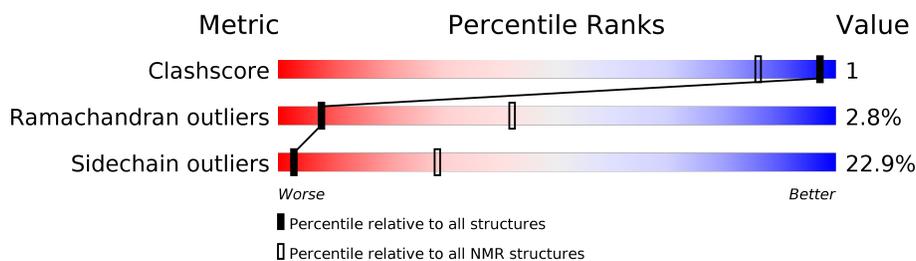|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)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

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 80%.

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	Percentile Ranks	Value
Clashscore		1
Ramachandran outliers		2.8%
Sidechain outliers		22.9%

Worse   *Better*

■ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	162	55% 21% 5% 19%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 17 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:2-A:15, A:28-A:65, A:73-A:121, A:129-A:158 (131)	0.17	17

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2464 atoms, of which 1199 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Dihydrofolate reductase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	162	2464	788	1199	218	255	4	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	37	VAL	ILE	SEE REMARK 999	UNP P15093

4 Residue-property plots (i)

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.

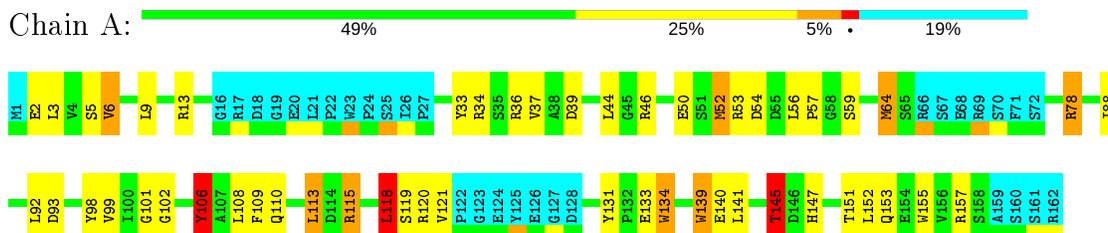• Molecule 1: Dihydrofolate reductase
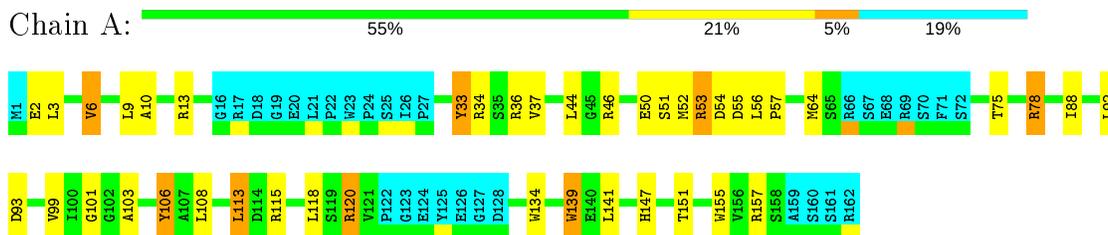
4.2 Scores per residue for each member of the ensemble

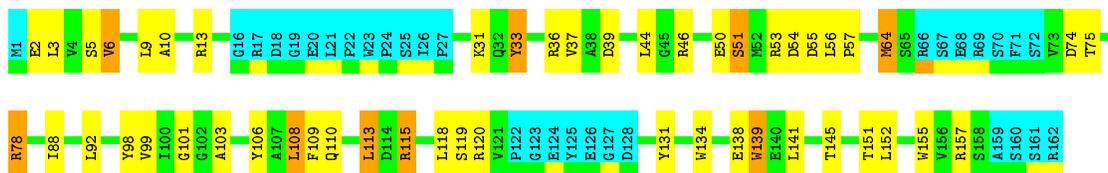Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Dihydrofolate reductase

4.2.2 Score per residue for model 2

• Molecule 1: Dihydrofolate reductase

4.2.3 Score per residue for model 3

• Molecule 1: Dihydrofolate reductase

4.2.4 Score per residue for model 4

• Molecule 1: Dihydrofolate reductase

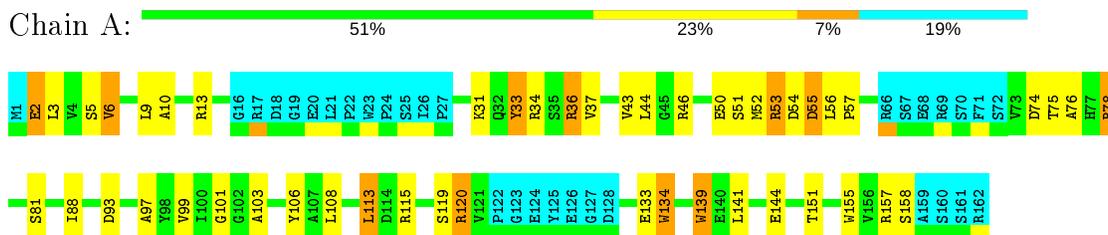### 4.2.5 Score per residue for model 5

• Molecule 1: Dihydrofolate reductase

4.2.6 Score per residue for model 6

• Molecule 1: Dihydrofolate reductase

Chain A: 52% 22% 6% 19%

4.2.7 Score per residue for model 7

- Molecule 1: Dihydrofolate reductase

Chain A: 56% 22% • 19%

4.2.8 Score per residue for model 8

- Molecule 1: Dihydrofolate reductase

Chain A: 56% 20% 5% • 19%

4.2.9 Score per residue for model 9

- Molecule 1: Dihydrofolate reductase

Chain A: 53% 22% 5% • 19%

4.2.10　Score per residue for model 10

• Molecule 1: Dihydrofolate reductase

Chain A:

4.2.11　Score per residue for model 11

• Molecule 1: Dihydrofolate reductase

Chain A:

4.2.12　Score per residue for model 12

• Molecule 1: Dihydrofolate reductase

Chain A:

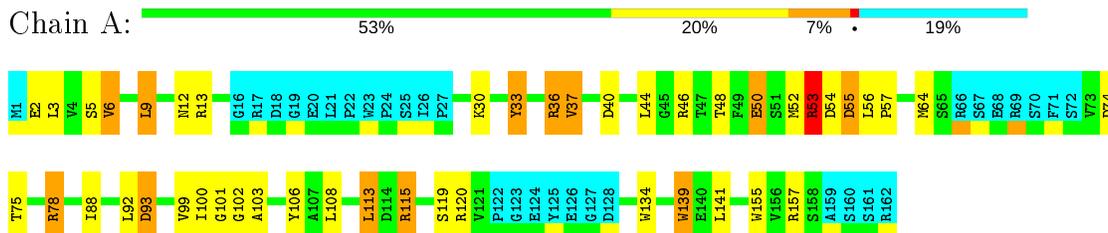### 4.2.13　Score per residue for model 13

• Molecule 1: Dihydrofolate reductase

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: Dihydrofolate reductase

Chain A: 52% 24% 19%

M1 E2 L3 V4 S5 V6 L9 A10 E11 N12 G16 R17 D18 G19 E20 L21 P22 W23 P24 S25 I26 P27 Y33 R36 V37 A38 D39 L44 G45 R46 E50 S51 M52 R53 D54 D55 L56 P57 R66 S67 E68 R69 S70 F71 S72 T75 R78 S81 E84

I88 D93 V99 I100 G101 G102 A103 A104 I105 Y106 A107 L108 F109 Q110 L113 D114 R115 L118 S119 R120 V121 P122 G123 E124 Y125 E126 G127 D128 T129 Y130 Y131 W134 W139 E140 L141 H147 T151 W155 V156 R157 S158 A159 S160 S161 R162

4.2.15 Score per residue for model 15

• Molecule 1: Dihydrofolate reductase

Chain A: 51% 23% 6% 19%

M1 E2 L3 V6 L9 A10 R13 G16 R17 D18 G19 E20 L21 P22 W23 P24 S25 I26 P27 Y33 R34 S35 R36 V37 L44 G45 R46 E50 S51 M52 R53 D54 D55 L56 P57 M64 S65 R66 S67 E68 R69 S70 F71 S72 V73 D74 T75 A76 H77 R78 I88

D93 V99 I100 G101 G102 A103 Y106 A107 L108 F109 Q110 L113 D114 R115 L118 S119 R120 V121 P122 G123 E124 Y125 E126 G127 D128 Y131 W134 D135 W139 E140 L141 H147 T151 W155 V156 R157 S158 A159 S160 S161 R162

4.2.16 Score per residue for model 16

• Molecule 1: Dihydrofolate reductase

Chain A: 48% 27% 6% 19%

M1 E2 L3 V6 L9 A10 E11 N12 R13 G16 R17 D18 G19 E20 L21 P22 W23 P24 S25 I26 P27 Y33 R36 V37 L44 G45 R46 T47 E50 S51 M52 R53 D54 D55 L56 P57 Q61 M64 S65 R66 S67 E68 R69 S70 F71 S72 V73 D74 T75 A76 H77

R78 S81 I88 D93 V99 I100 G101 G102 A103 Y106 A107 L108 F109 Q110 P111 H112 D114 L113 R115 L118 S119 R120 V121 P122 G123 E124 Y125 E126 G127 D128 T129 Y130 Y131 P132 E133 W134 W139 E140 L141 H147 Q153 E154 W155 V156 R157 S158 A159 S160 S161 R162

4.2.17 Score per residue for model 17 (medoid)

• Molecule 1: Dihydrofolate reductase

Chain A: 49% 25% 19%

M1 E2 L3 V4 S5 V6 L9 A10 E11 N12 R13 G16 R17 D18 G19 E20 L21 P22 W23 S25 I26 P27 K30 K31 Q32 Y33 R36 V37 A38 D39 L44 G45 R46 E50 S51 M52 R53 D54 D55 L56 P57 M64 R66 S67 E68 R69 S70 F71 S72 T75

A76 H77 R78 I88 L92 D93 V99 I100 G101 G102 A103 Y106 A107 L108 F109 Q110 L113 D114 R115 L118 S119 R120 V121 P122 G123 E124 Y125 E126 G127 D128 W134 W139 E140 L141 T145 T151 L152 Q153 E154 W155 V156 R157 S158 A159 S160 S161 R162
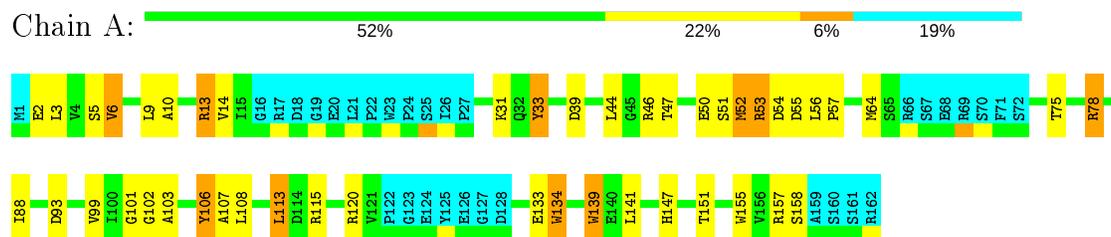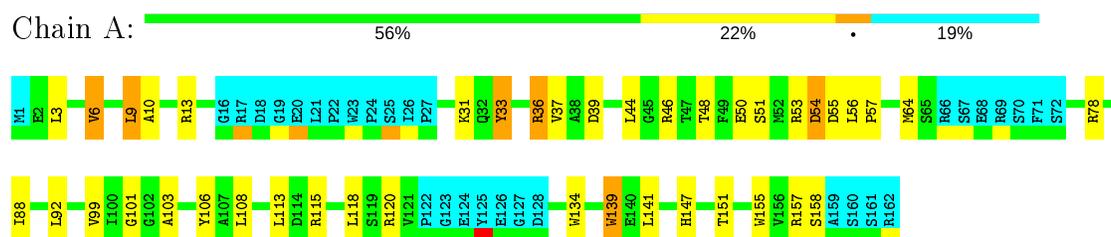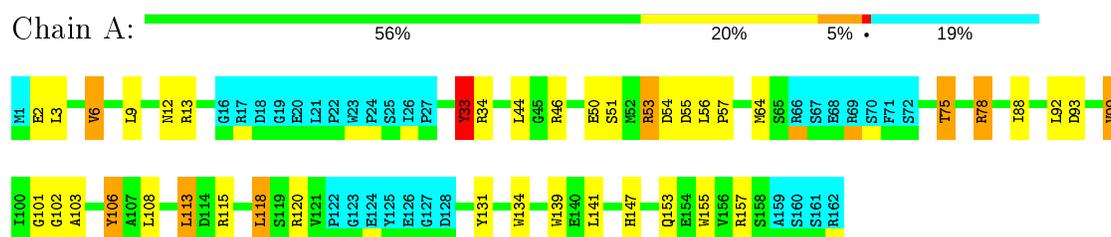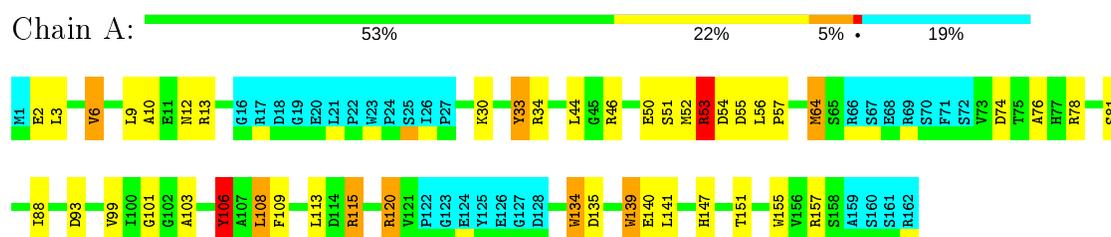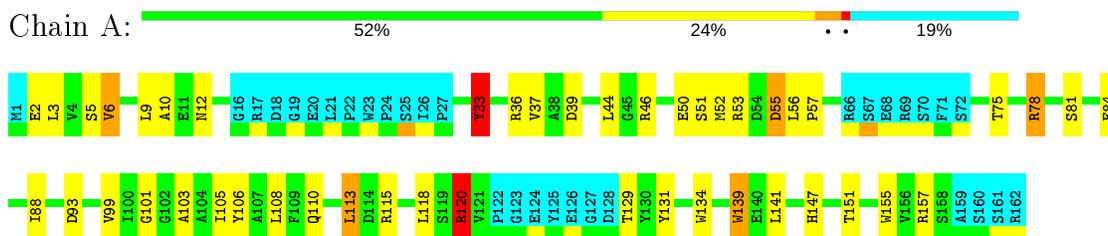
4.2.18 Score per residue for model 18

• Molecule 1: Dihydrofolate reductase

4.2.19 Score per residue for model 19
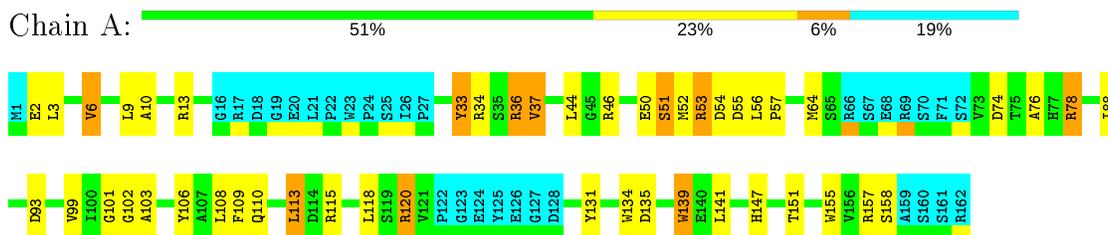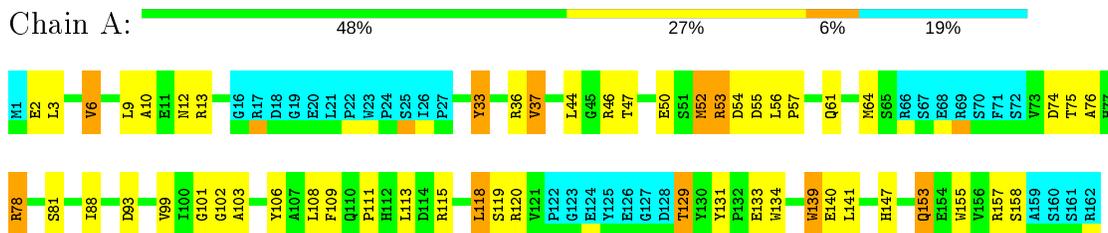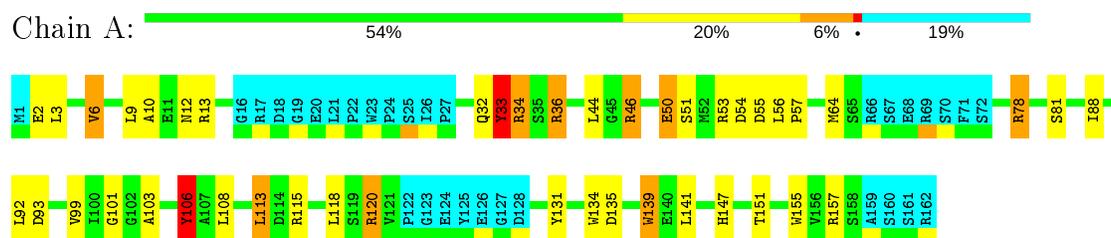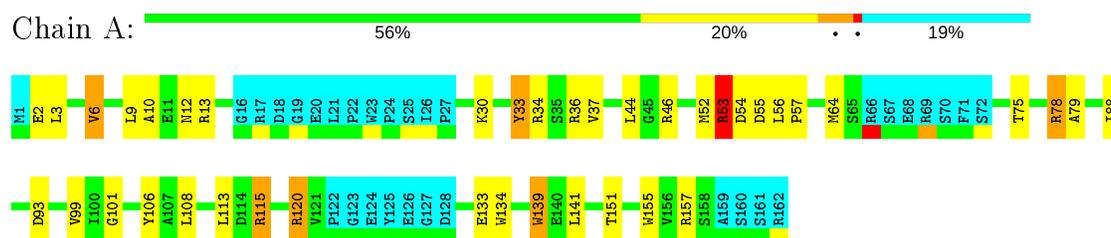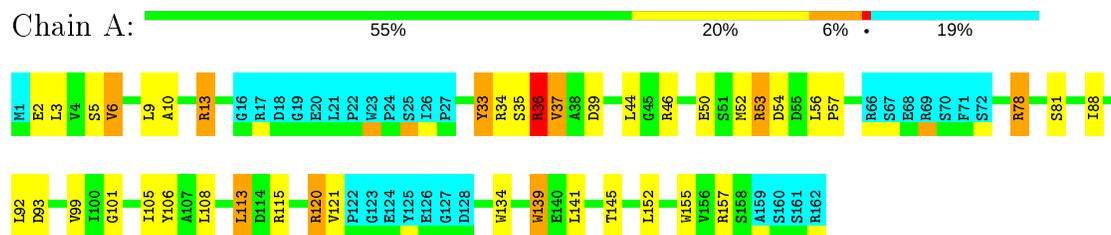
• Molecule 1: Dihydrofolate reductase

4.2.20 Score per residue for model 20

• Molecule 1: Dihydrofolate reductase

5 Refinement protocol and experimental data overview (i)

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

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
CNS	refinement	1.1

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1694
Number of shifts mapped to atoms	1694
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	1023	979	979	1±1
All	All	20460	19580	19580	25

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	Models Total
1:A:145:THR:HG22	1:A:152:LEU:HD12	0.56	1.74	1	6
1:A:47:THR:HG21	1:A:129:THR:OG1	0.54	2.03	16	1
1:A:52:MET:SD	1:A:55:ASP:O	0.47	2.73	4	5
1:A:64:MET:SD	1:A:108:LEU:HD11	0.47	2.50	9	2
1:A:118:LEU:HD12	1:A:131:TYR:CZ	0.46	2.45	8	7
1:A:118:LEU:HD23	1:A:153:GLN:HB3	0.45	1.86	16	1
1:A:64:MET:SD	1:A:109:PHE:CE2	0.41	3.13	1	3

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	131/162 (81%)	107±2 (82±2%)	20±3 (15±2%)	4±1 (3±1%)	8	42
All	All	2620/3240 (81%)	2148 (82%)	398 (15%)	74 (3%)	8	42

All 12 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	57	PRO	20
1	A	93	ASP	14
1	A	53	ARG	11
1	A	75	THR	8
1	A	158	SER	6
1	A	51	SER	6
1	A	110	GLN	3
1	A	105	ILE	2
1	A	52	MET	1
1	A	47	THR	1
1	A	92	LEU	1
1	A	100	ILE	1

5.2.2 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	106/132 (80%)	82±3 (77±3%)	24±3 (23±3%)	3	28
All	All	2120/2640 (80%)	1634 (77%)	486 (23%)	3	28

All 63 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	44	LEU	20
1	A	3	LEU	20
1	A	88	ILE	20
1	A	139	TRP	20
1	A	108	LEU	20
1	A	6	VAL	20
1	A	56	LEU	20
1	A	141	LEU	20
1	A	33	TYR	19
1	A	9	LEU	19
1	A	113	LEU	19
1	A	78	ARG	17
1	A	55	ASP	17
1	A	151	THR	15
1	A	54	ASP	14
1	A	64	MET	13
1	A	37	VAL	12
1	A	147	HIS	11
1	A	12	ASN	9
1	A	39	ASP	9
1	A	36	ARG	9
1	A	53	ARG	9
1	A	120	ARG	8
1	A	51	SER	8
1	A	81	SER	8
1	A	106	TYR	7
1	A	74	ASP	7
1	A	134	TRP	6
1	A	110	GLN	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	30	LYS	6
1	A	31	LYS	5
1	A	118	LEU	5
1	A	133	GLU	5
1	A	119	SER	5
1	A	52	MET	4
1	A	140	GLU	4
1	A	32	GLN	4
1	A	109	PHE	4
1	A	13	ARG	4
1	A	153	GLN	4
1	A	135	ASP	3
1	A	75	THR	3
1	A	93	ASP	3
1	A	92	LEU	3
1	A	129	THR	3
1	A	111	PRO	2
1	A	14	VAL	1
1	A	34	ARG	1
1	A	145	THR	1
1	A	99	VAL	1
1	A	59	SER	1
1	A	2	GLU	1
1	A	158	SER	1
1	A	11	GLU	1
1	A	115	ARG	1
1	A	84	GLU	1
1	A	96	THR	1
1	A	116	MET	1
1	A	121	VAL	1
1	A	144	GLU	1
1	A	138	GLU	1
1	A	83	GLU	1
1	A	61	GLN	1

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 80% for the well-defined parts and 77% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 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	1694
Number of shifts mapped to atoms	1694
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

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}\mathrm{C}_\alpha$	159	-0.26 ± 0.12	None needed (< 0.5 ppm)
$^{13}\mathrm{C}_\beta$	147	0.54 ± 0.09	Should be applied
$^{13}\mathrm{C}'$	148	-0.15 ± 0.06	None needed (< 0.5 ppm)
$^{15}\mathrm{N}$	147	0.26 ± 0.21	None needed (< 0.5 ppm)

6.1.3 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 80%, i.e. 1241 atoms were assigned a chemical shift out of a possible 1554. 2 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	629/647 (97%)	252/258 (98%)	253/262 (97%)	124/127 (98%)
Sidechain	612/783 (78%)	371/451 (82%)	241/298 (81%)	0/34 (0%)

Continued on next page...

Continued from previous page...

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Aromatic	0/124 (0%)	0/65 (0%)	0/53 (0%)	0/6 (0%)
Overall	1241/1554 (80%)	623/774 (80%)	494/613 (81%)	124/167 (74%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 77%, i.e. 1494 atoms were assigned a chemical shift out of a possible 1929. 2 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	758/794 (95%)	304/316 (96%)	307/324 (95%)	147/154 (95%)
Sidechain	736/982 (75%)	451/573 (79%)	285/363 (79%)	0/46 (0%)
Aromatic	0/153 (0%)	0/80 (0%)	0/66 (0%)	0/7 (0%)
Overall	1494/1929 (77%)	755/969 (78%)	592/753 (79%)	147/207 (71%)

6.1.4   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	22	PRO	CB	25.44	37.79 − 25.89	-5.4

6.1.5   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: