WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 11:45 pm GMT

PDB ID : 2L6U

Title : Solution NMR Structure of Med25(391-543) Comprising the Activator-Interacting Domain (ACID) of Human Mediator Subuniti 25. Northeast Structural Genomics Consortium Target HR6188A

Authors : Eletsky, A.; Ryuechan, W.T.; Sukumaran, D.K.; Shastry, R.; Ciccosanti, C.; Janjua, H.; Acton, T.B.; Xiao, R.; Everett, J.K.; Montelione, G.T.; Szyperski, T.; Northeast Structural Genomics Consortium (NESG)

Deposited on : 2010-11-24

The following versions of software and data (see references ⓘ) were used in the production of this report:

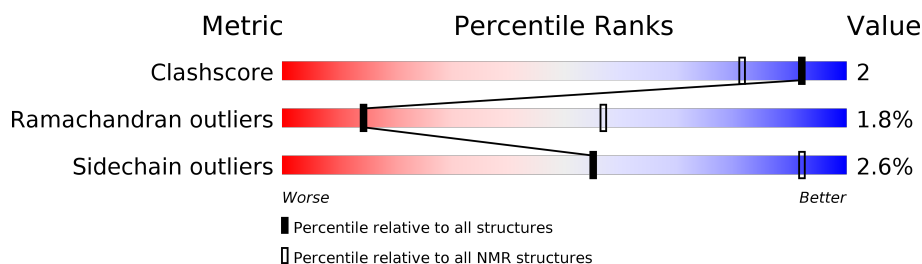|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 94%.

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		2
Ramachandran outliers		1.8%
Sidechain outliers		2.6%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	163	74% • 25%

WORLD WIDE
wwPDB
PROTEIN DATA BANK

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:410, A:423-A:434, A:438-A:501, A:508-A:540 (122)	0.44	20

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 5 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2626 atoms, of which 1316 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mediator complex subunit MED25.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	163	2626	837	1316	238	226	9	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	381	MET	-	EXPRESSION TAG	UNP Q71SY5
A	382	GLY	-	EXPRESSION TAG	UNP Q71SY5
A	383	HIS	-	EXPRESSION TAG	UNP Q71SY5
A	384	HIS	-	EXPRESSION TAG	UNP Q71SY5
A	385	HIS	-	EXPRESSION TAG	UNP Q71SY5
A	386	HIS	-	EXPRESSION TAG	UNP Q71SY5
A	387	HIS	-	EXPRESSION TAG	UNP Q71SY5
A	388	HIS	-	EXPRESSION TAG	UNP Q71SY5
A	389	SER	-	EXPRESSION TAG	UNP Q71SY5
A	390	HIS	-	EXPRESSION TAG	UNP Q71SY5

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.

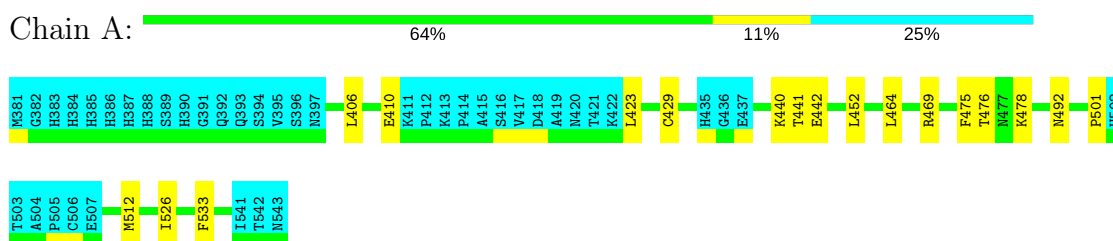• Molecule 1: Mediator complex subunit MED25

Chain A:
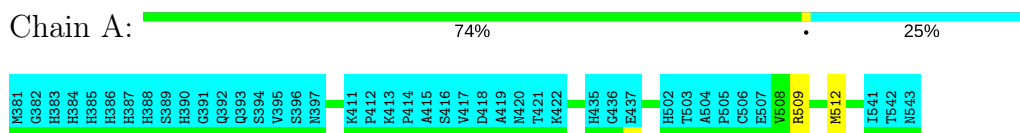
4.2 Scores per residue for each member of the ensemble
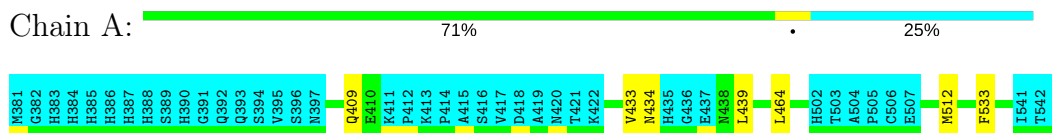
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
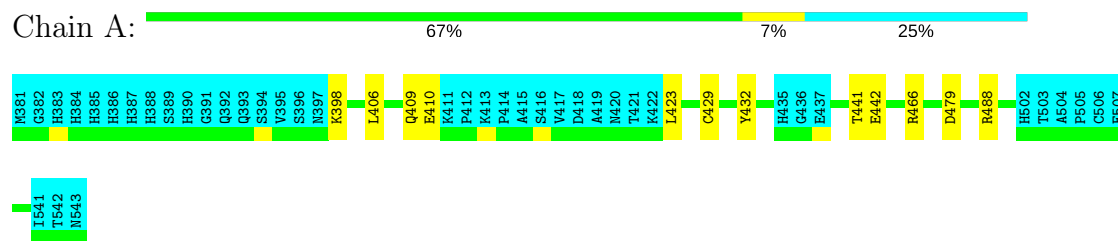
• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.2 Score per residue for model 2

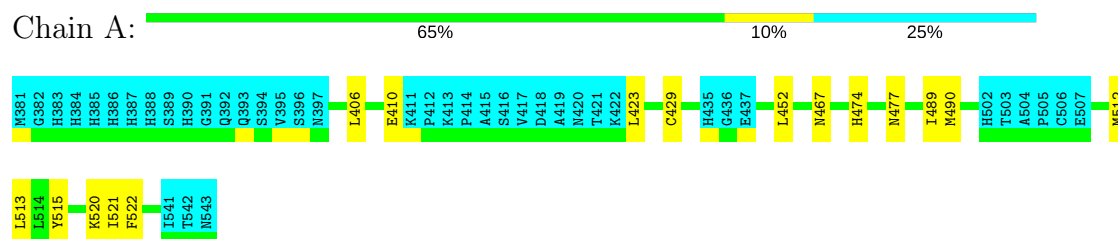• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.5 Score per residue for model 5

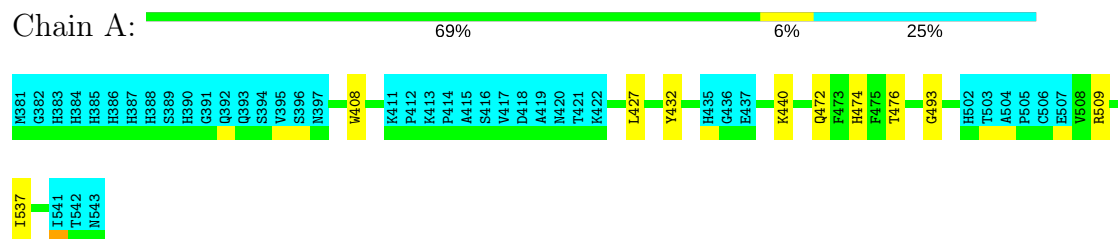• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.8 Score per residue for model 8

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.9 Score per residue for model 9

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Mediator complex subunit MED25

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Mediator complex subunit MED25

4.2.12 Score per residue for model 12

• Molecule 1: Mediator complex subunit MED25

4.2.13 Score per residue for model 13

• Molecule 1: Mediator complex subunit MED25

4.2.14 Score per residue for model 14
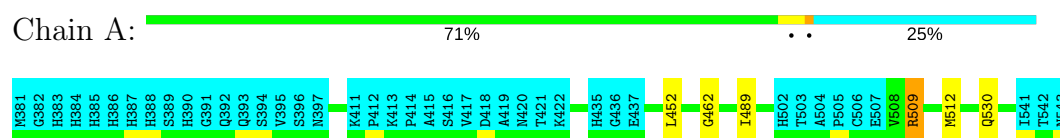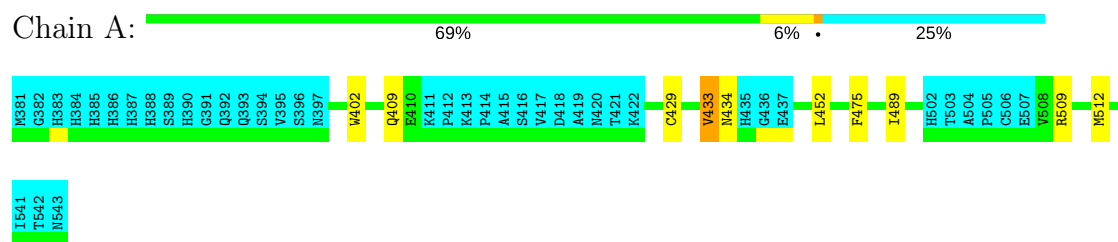
• Molecule 1: Mediator complex subunit MED25

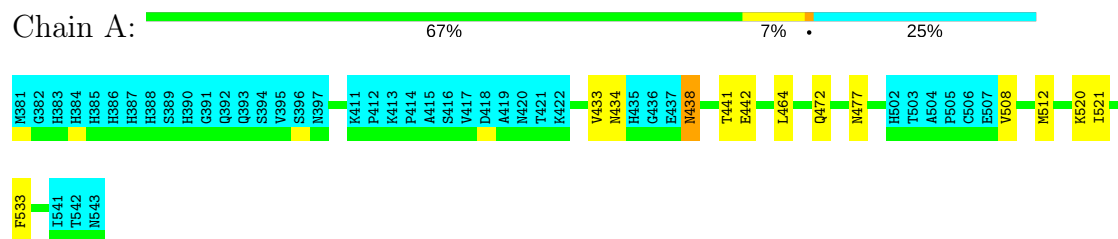#### 4.2.15 Score per residue for model 15

• Molecule 1: Mediator complex subunit MED25

4.2.16 Score per residue for model 16

• Molecule 1: Mediator complex subunit MED25

4.2.17 Score per residue for model 17

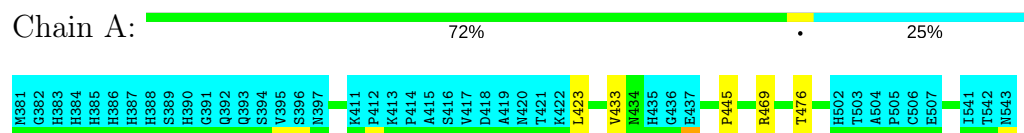• Molecule 1: Mediator complex subunit MED25
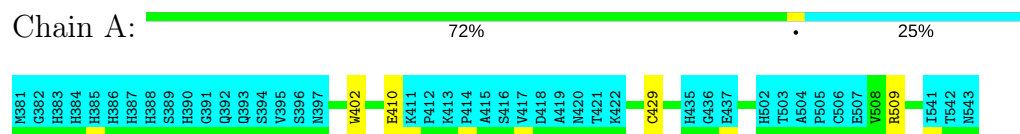
4.2.18 Score per residue for model 18

• Molecule 1: Mediator complex subunit MED25
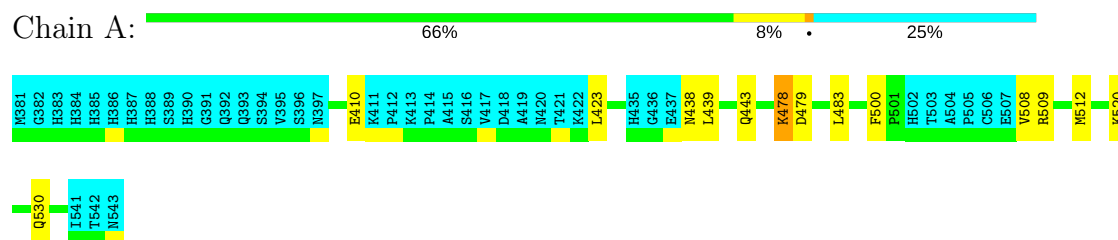
4.2.19 Score per residue for model 19
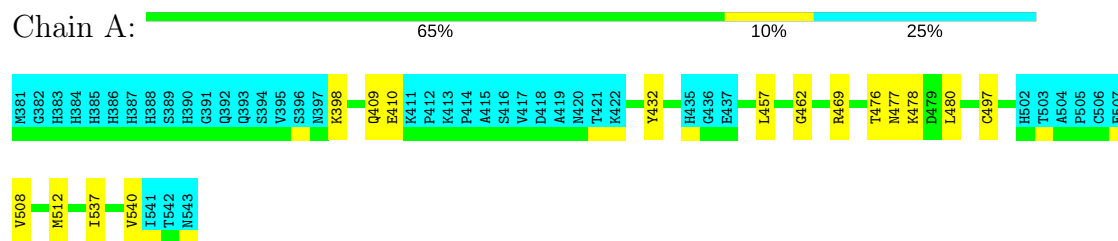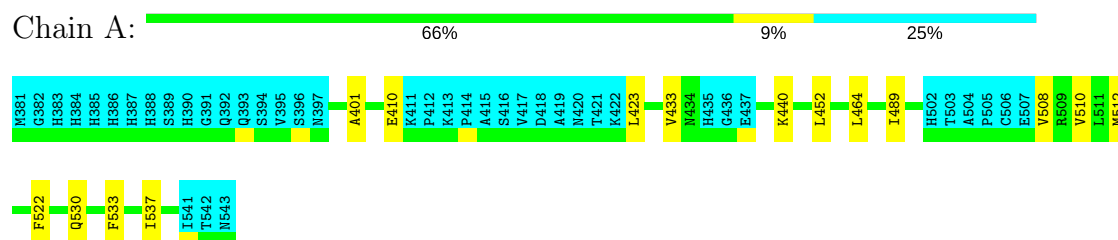
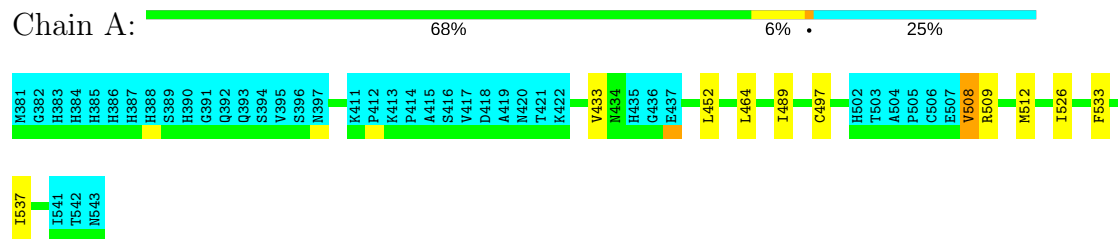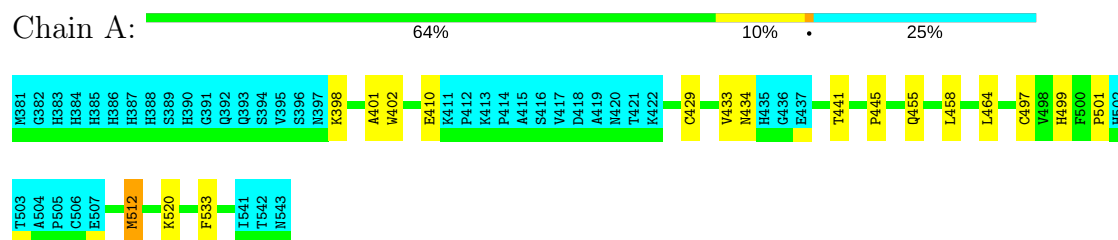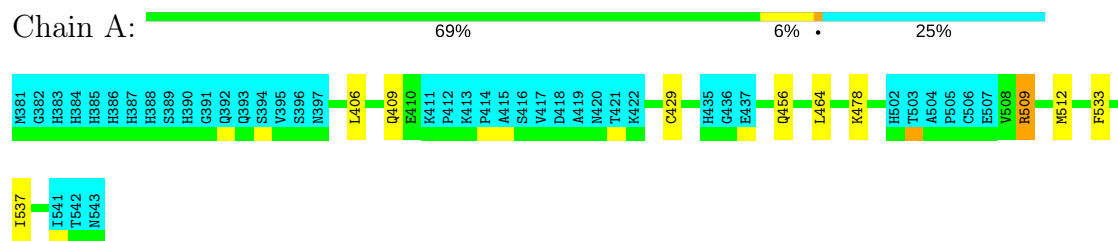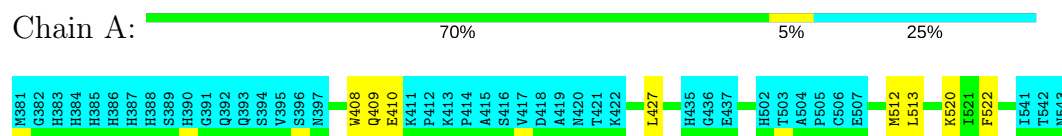• Molecule 1: Mediator complex subunit MED25

Chain A:
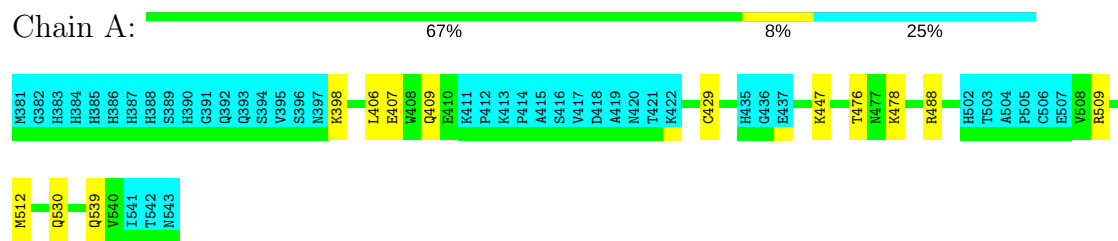
4.2.20 Score per residue for model 20 (medoid)

• Molecule 1: Mediator complex subunit MED25

Chain A:
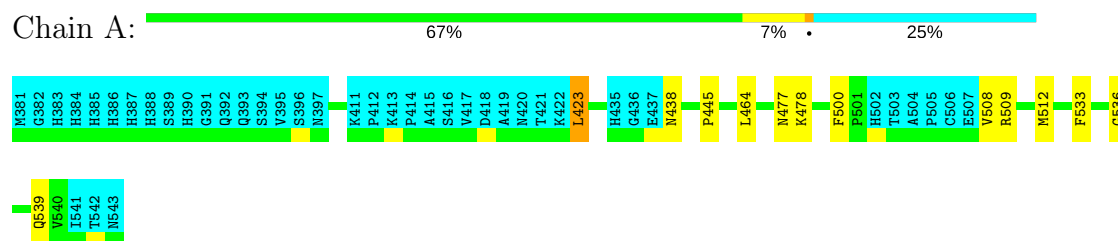
5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *SIMULATED ANNEALING*.

Of the 100 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
CNS	refinement	1.2
CNS	structure solution	1.2
CNS	geometry optimization	1.2
CYANA	refinement	3.0
CYANA	geometry optimization	3.0
CYANA	structure solution	3.0
AutoStructure	refinement	2.2.1
TALOS+	geometry optimization	1.2009.0721.18

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)	BMRB entry 17323
Number of chemical shift lists	1
Total number of shifts	2053
Number of shifts mapped to atoms	2053
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	94%

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	995	1033	1033	3±2
All	All	19900	20660	20660	67

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:402:TRP:HZ3	1:A:429:CYS:HG	0.66	1.34	17	3
1:A:410:GLU:HB2	1:A:423:LEU:HB2	0.59	1.73	12	5
1:A:406:LEU:HB2	1:A:429:CYS:SG	0.51	2.46	1	5
1:A:464:LEU:HD13	1:A:533:PHE:HA	0.51	1.83	15	8
1:A:457:LEU:HD13	1:A:540:VAL:HB	0.50	1.81	11	1
1:A:410:GLU:HB2	1:A:423:LEU:HB3	0.50	1.82	4	1
1:A:452:LEU:HD13	1:A:489:ILE:HG21	0.50	1.83	8	5
1:A:433:VAL:HG23	1:A:434:ASN:H	0.49	1.66	14	4
1:A:469:ARG:HB3	1:A:526:ILE:HB	0.49	1.85	1	1
1:A:508:VAL:HG22	1:A:526:ILE:HG21	0.48	1.85	13	1
1:A:423:LEU:H	1:A:423:LEU:HD23	0.48	1.68	20	1
1:A:510:VAL:HG12	1:A:530:GLN:HG3	0.47	1.86	12	1
1:A:429:CYS:HA	1:A:475:PHE:HA	0.47	1.86	1	2
1:A:445:PRO:HG3	1:A:501:PRO:HD2	0.47	1.86	14	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:398:LYS:HB3	1:A:432:TYR:HB3	0.44	1.88	3	2
1:A:490:MET:HB3	1:A:515:TYR:HB3	0.44	1.88	4	1
1:A:432:TYR:CE1	1:A:474:HIS:HB3	0.44	2.48	5	1
1:A:401:ALA:HB2	1:A:433:VAL:HG22	0.44	1.89	14	2
1:A:440:LYS:HE2	1:A:440:LYS:HA	0.44	1.90	5	1
1:A:408:TRP:HH2	1:A:427:LEU:HD12	0.43	1.72	5	2
1:A:410:GLU:HA	1:A:452:LEU:HD12	0.43	1.90	1	1
1:A:479:ASP:HA	1:A:483:LEU:HD23	0.43	1.89	18	1
1:A:461:LEU:HD22	1:A:464:LEU:HD12	0.43	1.91	7	1
1:A:513:LEU:HD11	1:A:522:PHE:HB3	0.43	1.91	4	2
1:A:497:CYS:SG	1:A:537:ILE:HD12	0.43	2.54	13	2
1:A:455:GLN:HA	1:A:458:LEU:HD12	0.42	1.91	14	1
1:A:500:PHE:HB2	1:A:509:ARG:HA	0.42	1.90	20	2
1:A:497:CYS:HG	1:A:499:HIS:CE1	0.41	2.33	14	1
1:A:512:MET:HG2	1:A:533:PHE:HZ	0.41	1.76	14	1
1:A:407:GLU:HB3	1:A:447:LYS:HE3	0.41	1.90	19	1
1:A:536:GLY:HA2	1:A:539:GLN:HE21	0.41	1.75	20	1
1:A:433:VAL:HG11	1:A:439:LEU:HD23	0.41	1.91	2	1
1:A:401:ALA:HB1	1:A:441:THR:HG21	0.41	1.93	14	1
1:A:402:TRP:HZ3	1:A:429:CYS:SG	0.40	2.38	17	1
1:A:477:ASN:HB2	1:A:480:LEU:HD21	0.40	1.92	11	1
1:A:472:GLN:HE21	1:A:521:ILE:HD12	0.40	1.77	9	1

6.3　Torsion angles ⓘ

6.3.1　Protein backbone ⓘ

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	122/163 (75%)	110±3 (90±2%)	10±3 (8±2%)	2±2 (2±1%)	14	57
All	All	2440/3260 (75%)	2193 (90%)	203 (8%)	44 (2%)	14	57

All 18 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	508	VAL	6
1	A	476	THR	5
1	A	442	GLU	4
1	A	441	THR	4
1	A	438	ASN	4
1	A	530	GLN	3
1	A	477	ASN	3
1	A	462	GLY	2
1	A	509	ARG	2
1	A	398	LYS	2
1	A	445	PRO	2
1	A	522	PHE	1
1	A	478	LYS	1
1	A	433	VAL	1
1	A	439	LEU	1
1	A	501	PRO	1
1	A	493	GLY	1
1	A	479	ASP	1

6.3.2 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	111/146 (76%)	108±1 (97±1%)	3±1 (3±1%)	55	92
All	All	2220/2920 (76%)	2163 (97%)	57 (3%)	55	92

All 17 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	512	MET	15
1	A	409	GLN	8
1	A	509	ARG	7
1	A	478	LYS	6
1	A	520	LYS	5
1	A	410	GLU	4
1	A	440	LYS	2
1	A	423	LEU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	438	ASN	1
1	A	539	GLN	1
1	A	488	ARG	1
1	A	492	ASN	1
1	A	456	GLN	1
1	A	466	ARG	1
1	A	469	ARG	1
1	A	467	ASN	1
1	A	443	GLN	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 carbohydrates 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 94% for the well-defined parts and 88% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 17323

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	2053
Number of shifts mapped to atoms	2053
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	154	-0.00 ± 0.06	None needed (< 0.5 ppm)
$^{13}C_\beta$	143	0.14 ± 0.06	None needed (< 0.5 ppm)
$^{13}C'$	153	0.06 ± 0.10	None needed (< 0.5 ppm)
^{15}N	142	0.18 ± 0.25	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 94%, i.e. 1505 atoms were assigned a chemical shift out of a possible 1605. 28 out of 28 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	595/598 (99%)	237/238 (100%)	243/244 (100%)	115/116 (99%)
Sidechain	773/862 (90%)	477/509 (94%)	278/311 (89%)	18/42 (43%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	137/145 (94%)	72/77 (94%)	62/63 (98%)	3/5 (60%)
Overall	1505/1605 (94%)	786/824 (95%)	583/618 (94%)	136/163 (83%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 88%, i.e. 1834 atoms were assigned a chemical shift out of a possible 2083. 30 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	745/797 (93%)	296/317 (93%)	307/326 (94%)	142/154 (92%)
Sidechain	941/1076 (87%)	584/638 (92%)	334/388 (86%)	23/50 (46%)
Aromatic	148/210 (70%)	76/113 (67%)	65/81 (80%)	7/16 (44%)
Overall	1834/2083 (88%)	956/1068 (90%)	706/795 (89%)	172/220 (78%)

7.1.4 Statistically unusual chemical shifts ⓘ

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	425	ARG	HB2	-0.22	3.15 − 0.45	-7.5
1	A	441	THR	HG21	-0.28	2.29 − -0.01	-6.2
1	A	441	THR	HG23	-0.28	2.29 − -0.01	-6.2
1	A	441	THR	HG22	-0.28	2.29 − -0.01	-6.2

7.1.5 Random Coil Index (RCI) plots ⓘ

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: