



# Full wwPDB NMR Structure Validation Report ⓘ

Oct 5, 2017 – 08:20 AM EDT

PDB ID : 5O1T  
Title : Solution structure of the RNA binding domain of Nrd1  
Authors : Martinez-Lumbreras, S.; Perez-Canadillas, J.M.  
Deposited on : unknown

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 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 : rb-20030345  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

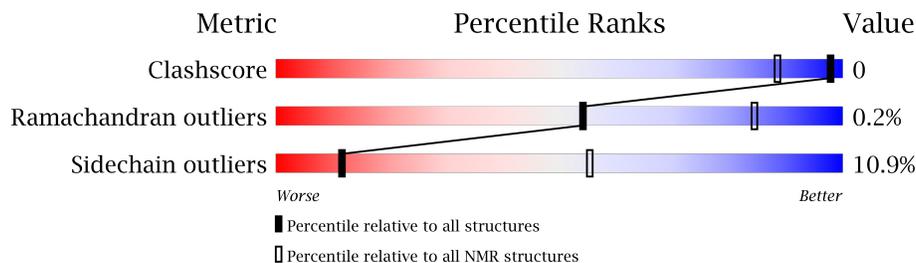
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 76%.

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	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  $\geq 3$ , 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	179	

## 2 Ensemble composition and analysis i

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

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:303-A:306, A:310-A:311, A:317-A:463 (153)	0.32	1

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 2 clusters. No single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2788 atoms, of which 1371 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein NRD1.

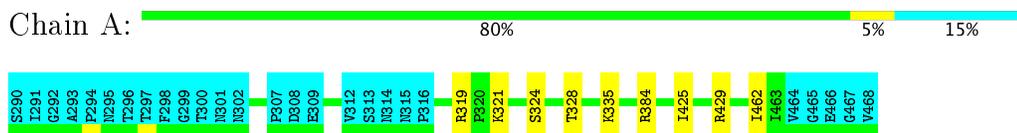
Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	179	2788	901	1371	250	262	4	0

## 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: Protein NRD1

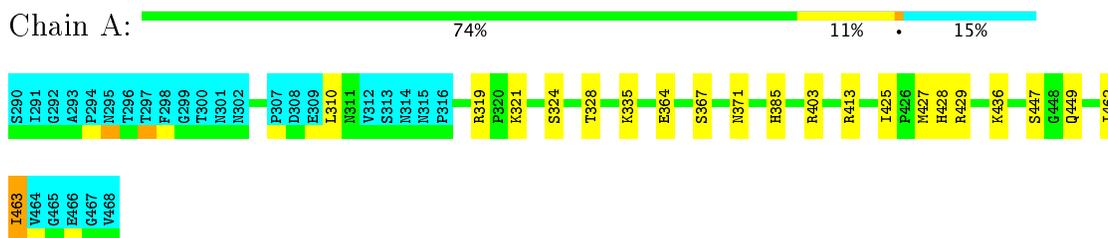


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

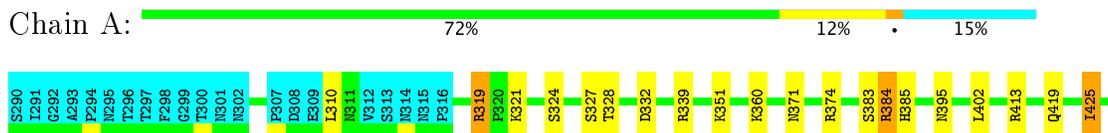
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Protein NRD1



#### 4.2.2 Score per residue for model 2

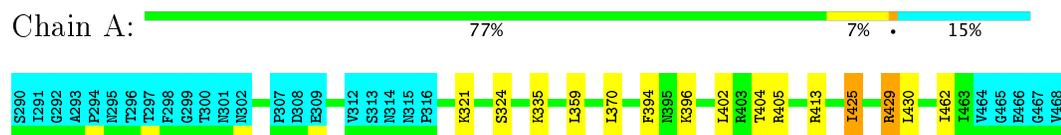
- Molecule 1: Protein NRD1





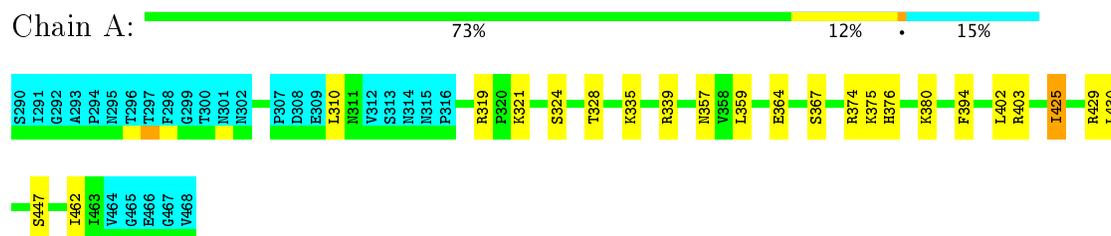
### 4.2.3 Score per residue for model 3

- Molecule 1: Protein NRD1



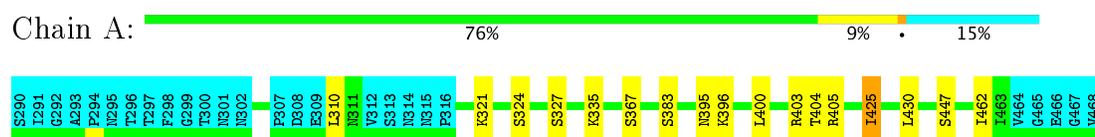
### 4.2.4 Score per residue for model 4

- Molecule 1: Protein NRD1



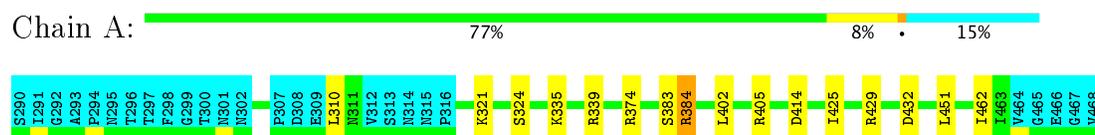
### 4.2.5 Score per residue for model 5

- Molecule 1: Protein NRD1



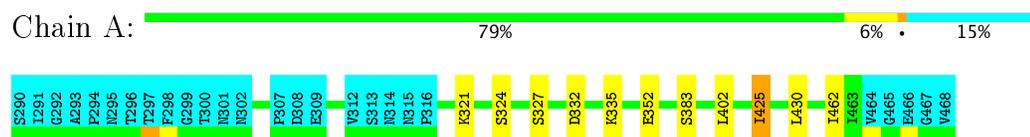
### 4.2.6 Score per residue for model 6

- Molecule 1: Protein NRD1



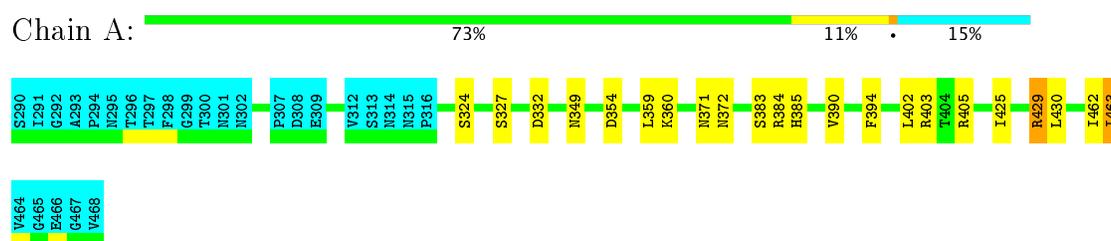
### 4.2.7 Score per residue for model 7

- Molecule 1: Protein NRD1



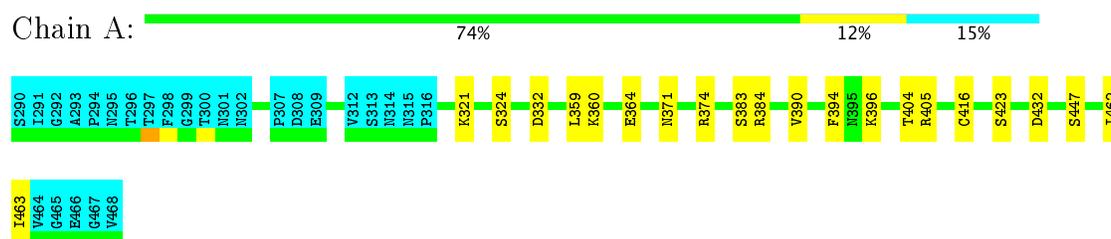
### 4.2.8 Score per residue for model 8

- Molecule 1: Protein NRD1



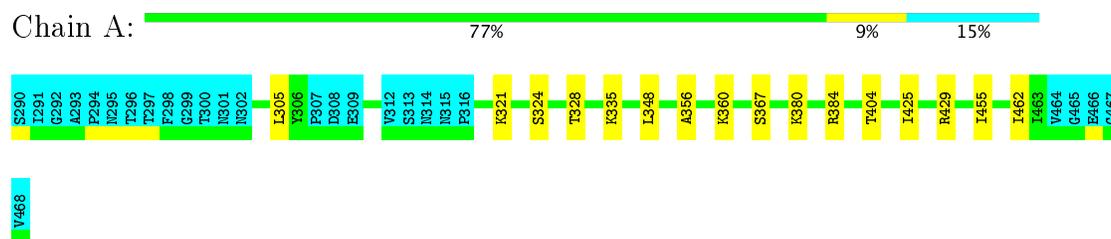
### 4.2.9 Score per residue for model 9

- Molecule 1: Protein NRD1



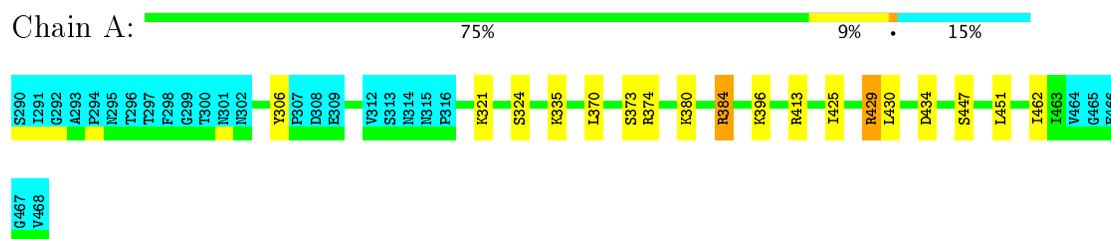
### 4.2.10 Score per residue for model 10

- Molecule 1: Protein NRD1



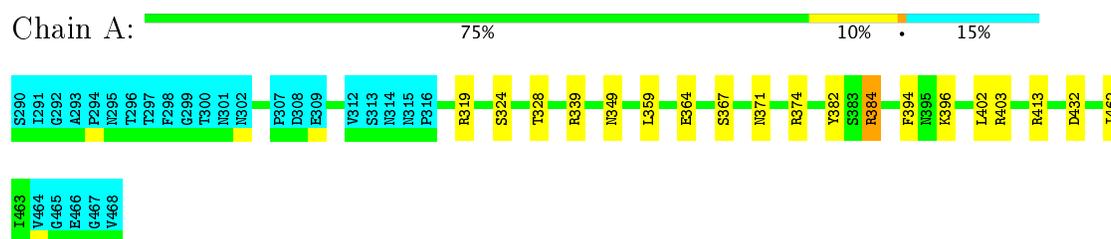
### 4.2.11 Score per residue for model 11

- Molecule 1: Protein NRD1



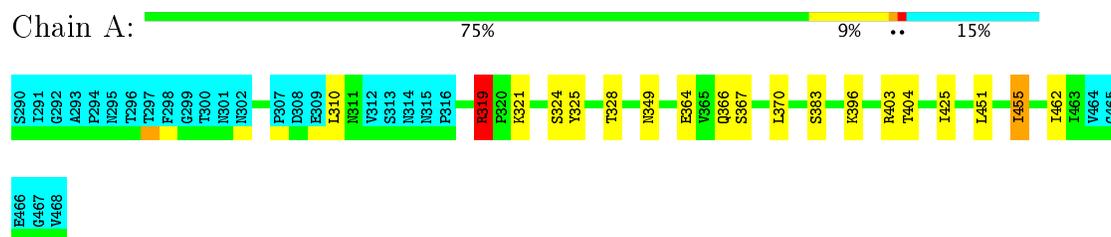
### 4.2.12 Score per residue for model 12

- Molecule 1: Protein NRD1



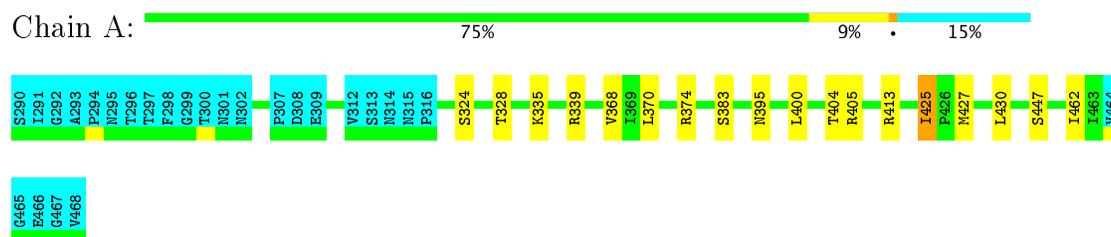
### 4.2.13 Score per residue for model 13

- Molecule 1: Protein NRD1



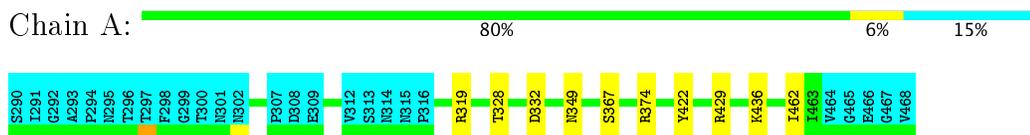
### 4.2.14 Score per residue for model 14

- Molecule 1: Protein NRD1



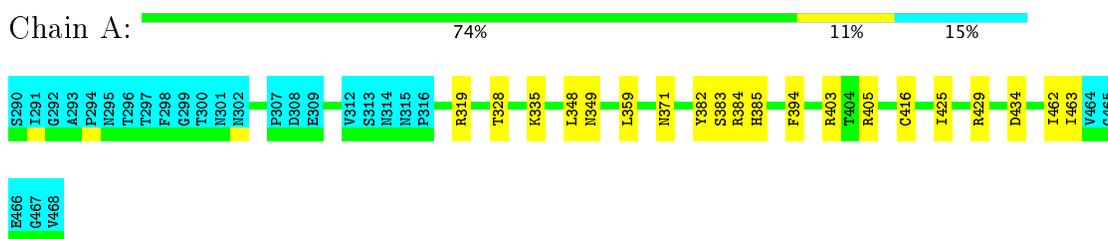
#### 4.2.15 Score per residue for model 15

- Molecule 1: Protein NRD1



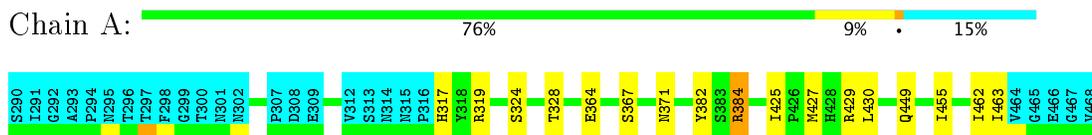
#### 4.2.16 Score per residue for model 16

- Molecule 1: Protein NRD1



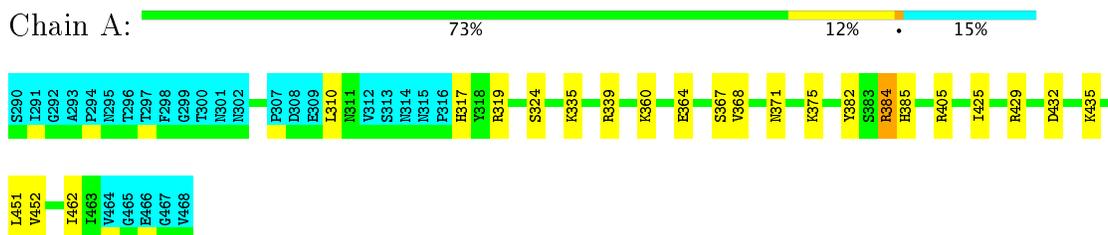
#### 4.2.17 Score per residue for model 17

- Molecule 1: Protein NRD1



#### 4.2.18 Score per residue for model 18

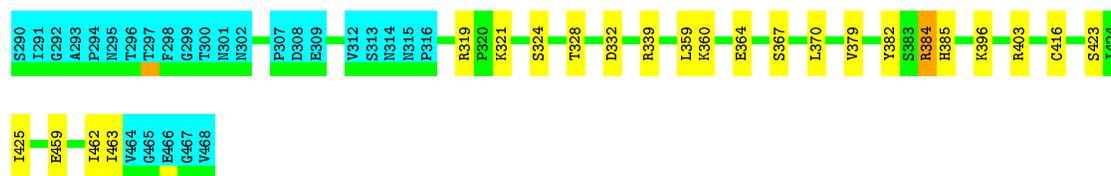
- Molecule 1: Protein NRD1



#### 4.2.19 Score per residue for model 19

- Molecule 1: Protein NRD1

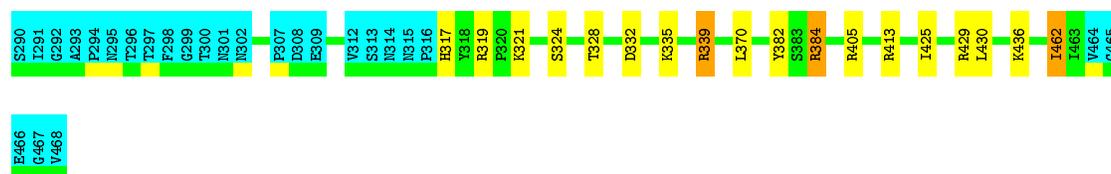
Chain A:  73% 12% 15%



#### 4.2.20 Score per residue for model 20

- Molecule 1: Protein NRD1

Chain A:  75% 8% 15%



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *Minimization*.

Of the 50 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
CYANA	structure calculation	
AMBER	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)	5o1t_cs.cif
Number of chemical shift lists	1
Total number of shifts	1838
Number of shifts mapped to atoms	1838
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

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	0.69±0.02	0±0/1275 (0.0±0.0%)	0.97±0.03	2±1/1738 (0.1±0.1%)
All	All	0.70	0/25500 (0.0%)	0.97	37/34760 (0.1%)

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	1.2±0.4
All	All	0	24

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	319	ARG	NE-CZ-NH2	-7.57	116.52	120.30	19	7
1	A	384	ARG	NE-CZ-NH2	-7.37	116.61	120.30	12	8
1	A	339	ARG	NE-CZ-NH1	6.77	123.68	120.30	2	5
1	A	429	ARG	NE-CZ-NH1	6.63	123.62	120.30	15	6
1	A	384	ARG	NE-CZ-NH1	6.20	123.40	120.30	12	4
1	A	319	ARG	NE-CZ-NH1	5.77	123.19	120.30	4	7

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	462	ILE	Peptide	20
1	A	339	ARG	Sidechain	2

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	319	ARG	Sidechain	1
1	A	422	TYR	Sidechain	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	1235	1207	1207	1±1
All	All	24700	24140	24140	22

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:425:ILE:HG21	1:A:430:LEU:HD21	0.65	1.67	3	10
1:A:359:LEU:HD23	1:A:394:PHE:CE1	0.55	2.36	9	6
1:A:359:LEU:HD22	1:A:390:VAL:HG11	0.51	1.82	8	2
1:A:325:TYR:CE2	1:A:455:ILE:HD12	0.45	2.46	13	1
1:A:305:LEU:CD2	1:A:356:ALA:HB2	0.43	2.43	10	1
1:A:359:LEU:CD1	1:A:379:VAL:HG11	0.40	2.47	19	1
1:A:352:GLU:CD	1:A:352:GLU:H	0.40	2.20	7	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	153/179 (85%)	145±2 (95±1%)	7±2 (5±1%)	0±0 (0±0%)	54	84
All	All	3060/3580 (85%)	2906 (95%)	148 (5%)	6 (0%)	54	84

All 3 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	463	ILE	4
1	A	462	ILE	1
1	A	447	SER	1

### 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	135/156 (87%)	120±4 (89±3%)	15±4 (11±3%)	12 55
All	All	2700/3120 (87%)	2405 (89%)	295 (11%)	12 55

All 60 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	324	SER	18
1	A	321	LYS	13
1	A	425	ILE	13
1	A	384	ARG	12
1	A	328	THR	12
1	A	335	LYS	12
1	A	367	SER	10
1	A	405	ARG	9
1	A	429	ARG	9
1	A	383	SER	9
1	A	371	ASN	8
1	A	374	ARG	8
1	A	364	GLU	8
1	A	403	ARG	8
1	A	402	LEU	7
1	A	413	ARG	7
1	A	332	ASP	7
1	A	310	LEU	7
1	A	396	LYS	7
1	A	404	THR	6
1	A	382	TYR	6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	360	LYS	6
1	A	385	HIS	6
1	A	370	LEU	6
1	A	447	SER	5
1	A	349	ASN	5
1	A	463	ILE	5
1	A	436	LYS	4
1	A	327	SER	4
1	A	432	ASP	4
1	A	451	LEU	4
1	A	416	CYS	3
1	A	317	HIS	3
1	A	427	MET	3
1	A	395	ASN	3
1	A	380	LYS	3
1	A	455	ILE	3
1	A	339	ARG	2
1	A	423	SER	2
1	A	449	GLN	2
1	A	400	LEU	2
1	A	375	LYS	2
1	A	348	LEU	2
1	A	368	VAL	2
1	A	319	ARG	2
1	A	434	ASP	2
1	A	452	VAL	1
1	A	419	GLN	1
1	A	373	SER	1
1	A	354	ASP	1
1	A	435	LYS	1
1	A	372	ASN	1
1	A	351	LYS	1
1	A	459	GLU	1
1	A	414	ASP	1
1	A	428	HIS	1
1	A	306	TYR	1
1	A	366	GLN	1
1	A	376	HIS	1
1	A	357	ASN	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 76% for the well-defined parts and 75% for the entire structure.

### 7.1 Chemical shift list 1

File name: 5o1t\_cs.cif

Chemical shift list name: *Nrd1\_290-468\_curated.tbl*

#### 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	1838
Number of shifts mapped to atoms	1838
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	8

#### 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
$^{13}\text{C}_\alpha$	127	$0.51 \pm 0.11$	Should be applied
$^{13}\text{C}_\beta$	113	$0.31 \pm 0.19$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	160	$0.72 \pm 0.58$	None needed (imprecise)

#### 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 76%, i.e. 1477 atoms were assigned a chemical shift out of a possible 1933. 0 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	550/743 (74%)	293/295 (99%)	116/306 (38%)	141/142 (99%)
Sidechain	763/976 (78%)	521/577 (90%)	229/353 (65%)	13/46 (28%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	164/214 (77%)	93/114 (82%)	67/88 (76%)	4/12 (33%)
Overall	1477/1933 (76%)	907/986 (92%)	412/747 (55%)	158/200 (79%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 75%, i.e. 1642 atoms were assigned a chemical shift out of a possible 2196. 0 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	621/867 (72%)	334/344 (97%)	127/358 (35%)	160/165 (97%)
Sidechain	850/1106 (77%)	582/652 (89%)	250/403 (62%)	18/51 (35%)
Aromatic	171/223 (77%)	98/119 (82%)	69/92 (75%)	4/12 (33%)
Overall	1642/2196 (75%)	1014/1115 (91%)	446/853 (52%)	182/228 (80%)

#### 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	366	GLN	HB3	-0.99	3.37 – 0.67	-11.2
1	A	311	ASN	HB2	-0.50	4.36 – 1.26	-10.7
1	A	335	LYS	HE3	0.92	3.86 – 1.96	-10.5
1	A	395	ASN	HB3	0.07	4.41 – 1.11	-8.2
1	A	335	LYS	HE2	1.48	3.87 – 1.97	-7.6
1	A	421	GLY	HA3	1.72	5.80 – 2.00	-5.7
1	A	352	GLU	HG2	1.12	3.33 – 1.23	-5.5
1	A	366	GLN	HB2	0.76	3.30 – 0.80	-5.2

#### 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:

