Full wwPDB NMR Structure Validation Report (i)

Jan 25, 2023 – 01:30 PM EST

PDB ID	:	2RNK
Title	:	NMR structure of the domain 513-651 of the SARS-CoV nonstructural protein nsp3
Authors	:	Chatterjee, A.; Johnson, M.A.; Serrano, P.; Pedrini, B.; Joseph, J.; Saikatendu, K.; Neuman, B.W.; Wilson, I.A.; Stevens, R.C.; Buchmeier, M.J.; Kuhn, P.; Wuthrich, K.; Joint Center for Structural Genomics (JCSG)
Deposited on	:	2008-01-11

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 (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:

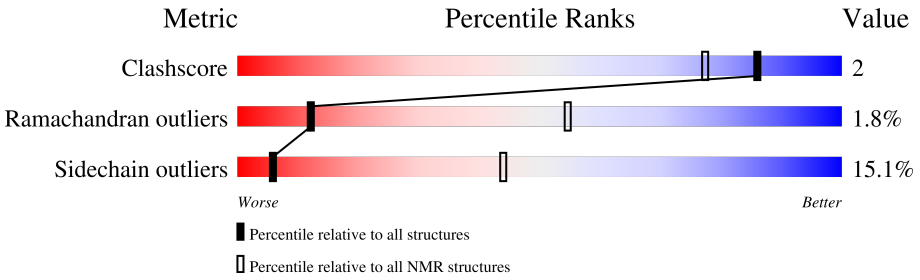
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.31.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 was not calculated.

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	143	76% 13% 12%

2 Ensemble composition and analysis

This entry contains 20 models. Model 12 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:524-A:649 (126)	0.35	12

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2239 atoms, of which 1131 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Replicase polyprotein 1ab.

Mol	Chain	Residues	Atoms						Trace
1	A	143	Total	C	H	N	O	S	0
			2239	705	1131	187	207	9	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP P59641
A	-3	SER	-	expression tag	UNP P59641
A	-2	HIS	-	expression tag	UNP P59641
A	-1	MET	-	expression tag	UNP P59641

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: Replicase polyprotein 1ab

Chain A:

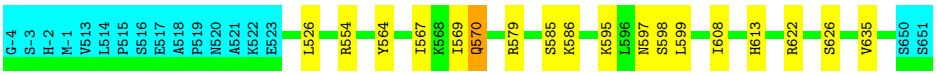
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Replicase polyprotein 1ab

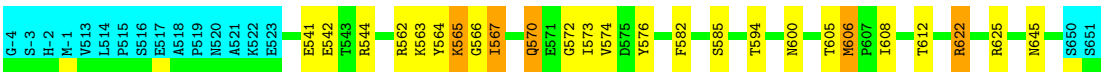
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Replicase polyprotein 1ab

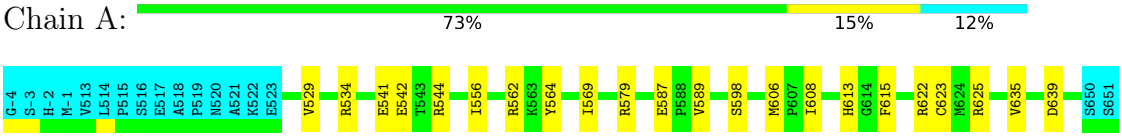
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Replicase polyprotein 1ab

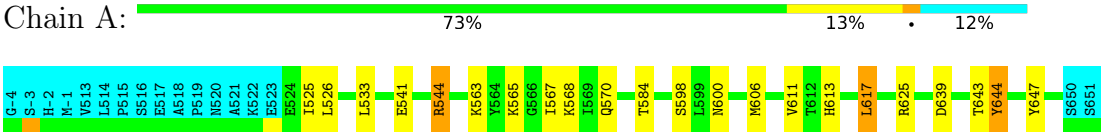
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Replicase polyprotein 1ab

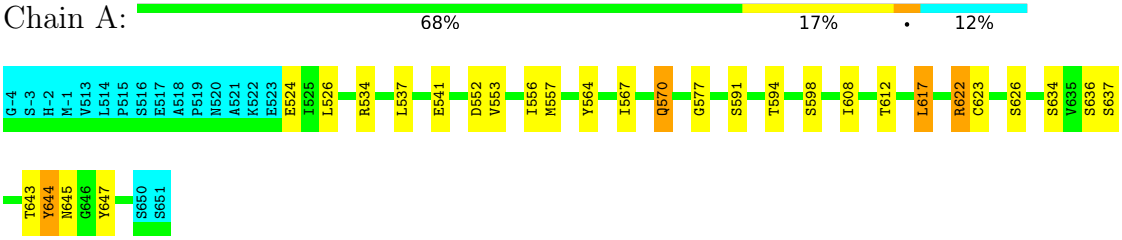
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Replicase polyprotein 1ab

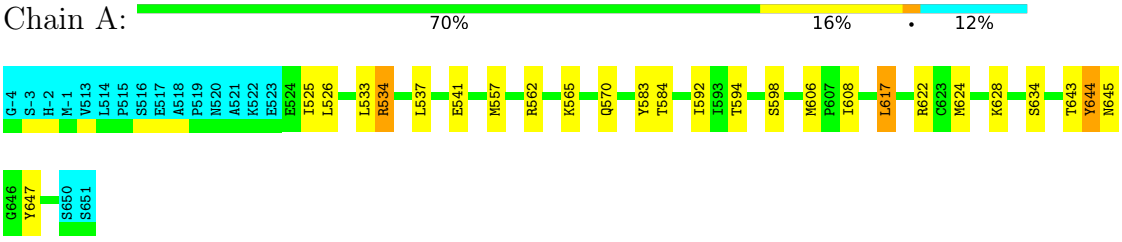
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Replicase polyprotein 1ab

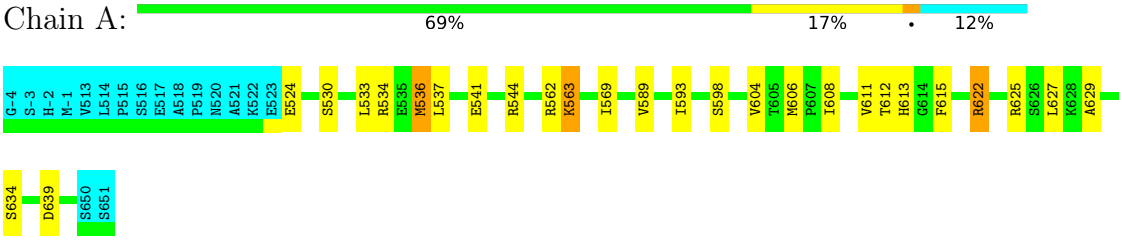
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Replicase polyprotein 1ab

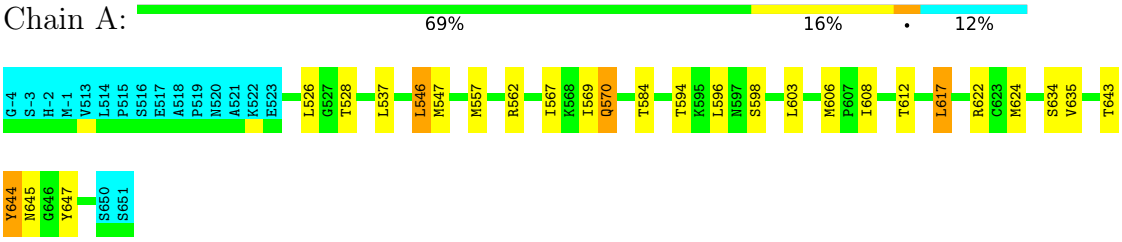
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Replicase polyprotein 1ab

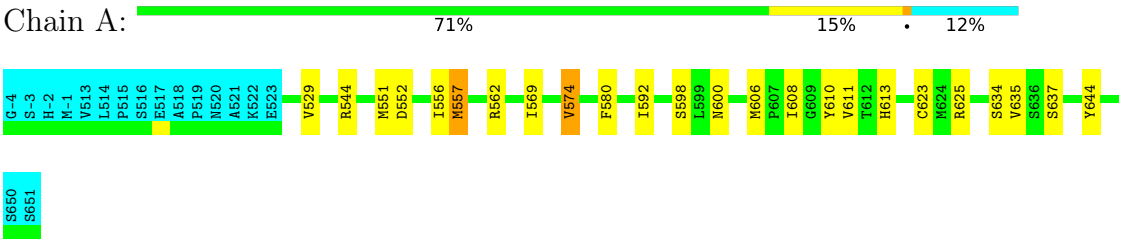
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Replicase polyprotein 1ab

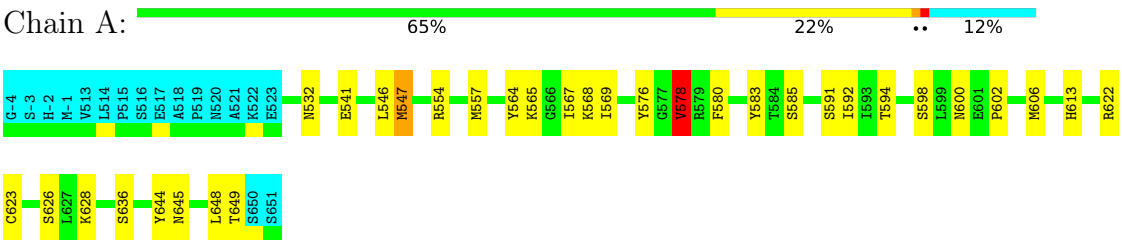
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Replicase polyprotein 1ab

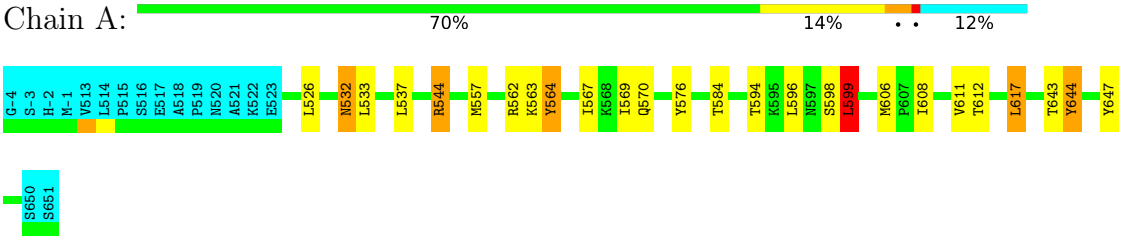
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Replicase polyprotein 1ab

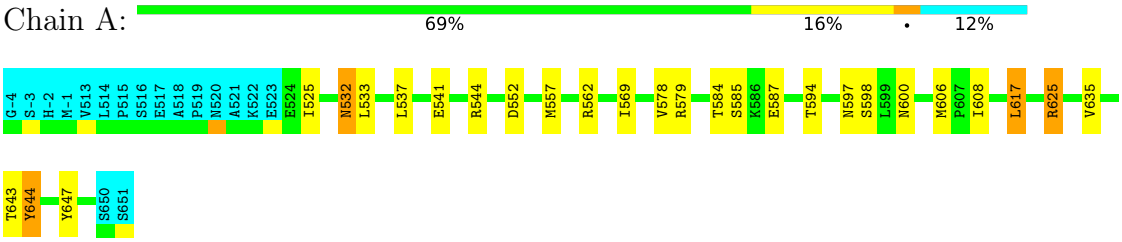
Chain A:



4.2.12 Score per residue for model 12 (medoid)

• Molecule 1: Replicase polyprotein 1ab

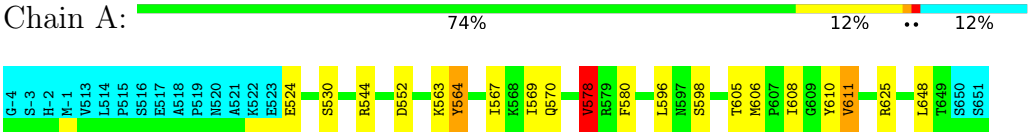
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Replicase polyprotein 1ab

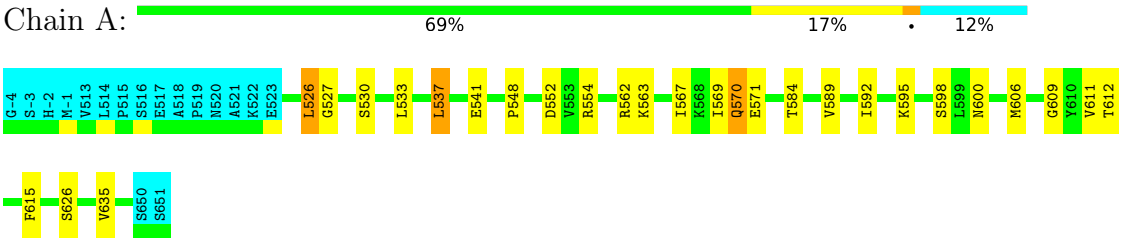
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Replicase polyprotein 1ab

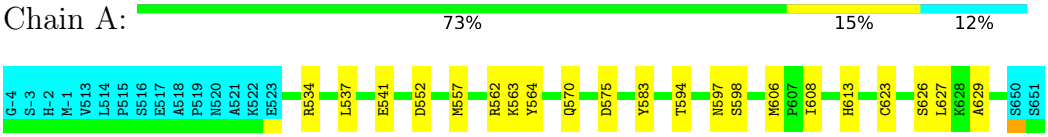
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Replicase polyprotein 1ab

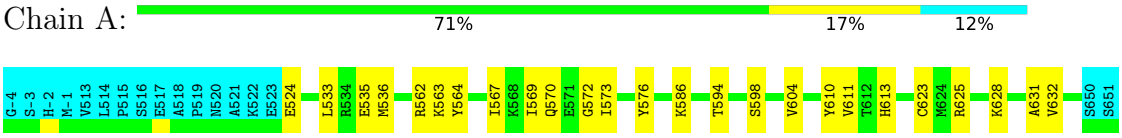
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Replicase polyprotein 1ab

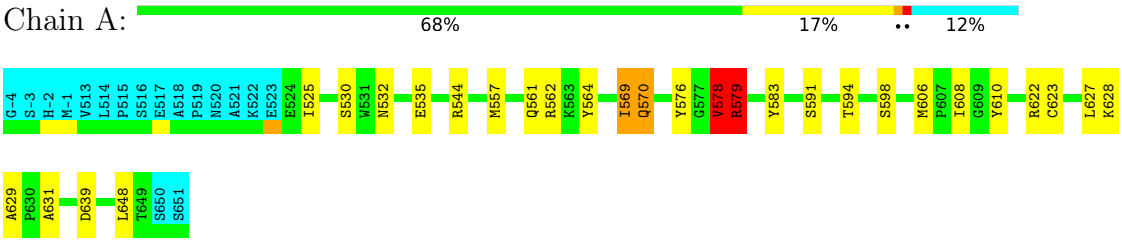
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Replicase polyprotein 1ab

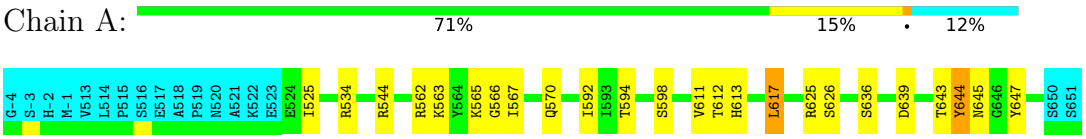
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Replicase polyprotein 1ab

Chain A:



4.2.19 Score per residue for model 19

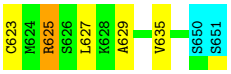
• Molecule 1: Replicase polyprotein 1ab

Chain A:

4.2.20 Score per residue for model 20

• Molecule 1: Replicase polyprotein 1ab

Chain A:



5 Refinement protocol and experimental data overview

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

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
CYANA	refinement	1.2

No chemical shift data was provided.

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	0.61±0.01	0±0/1008 (0.0± 0.0%)	1.10±0.03	2±1/1367 (0.2± 0.1%)
All	All	0.61	0/20160 (0.0%)	1.10	50/27340 (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	2.0±1.8
All	All	0	41

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	625	ARG	CD-NE-CZ	9.35	136.69	123.60	12	2
1	A	625	ARG	NE-CZ-NH1	8.36	124.48	120.30	20	6
1	A	546	LEU	CB-CG-CD2	7.05	122.99	111.00	8	2
1	A	625	ARG	NE-CZ-NH2	-7.02	116.79	120.30	12	3
1	A	644	TYR	CB-CG-CD2	-6.81	116.92	121.00	8	8
1	A	537	LEU	CB-CG-CD1	6.62	122.25	111.00	5	8
1	A	578	VAL	CA-CB-CG1	6.55	120.73	110.90	13	2
1	A	544	ARG	NE-CZ-NH2	-6.43	117.08	120.30	4	4
1	A	579	ARG	NE-CZ-NH2	-6.06	117.27	120.30	17	1
1	A	589	VAL	CA-CB-CG2	6.05	119.97	110.90	20	1
1	A	578	VAL	CB-CA-C	6.04	122.88	111.40	10	2
1	A	544	ARG	NE-CZ-NH1	5.92	123.26	120.30	2	2
1	A	622	ARG	CD-NE-CZ	5.72	131.61	123.60	2	1
1	A	579	ARG	NE-CZ-NH1	5.68	123.14	120.30	17	1
1	A	575	ASP	CB-CG-OD1	5.60	123.34	118.30	15	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	622	ARG	N-CA-CB	-5.58	100.55	110.60	5	1
1	A	622	ARG	NE-CZ-NH1	5.34	122.97	120.30	1	1
1	A	599	LEU	CB-CG-CD2	5.17	119.79	111.00	11	1
1	A	554	ARG	NE-CZ-NH2	-5.11	117.75	120.30	10	1
1	A	576	TYR	CB-CG-CD2	-5.05	117.97	121.00	17	1
1	A	574	VAL	CA-CB-CG1	5.01	118.42	110.90	9	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	564	TYR	Sidechain,Peptide	8
1	A	625	ARG	Sidechain	6
1	A	583	TYR	Sidechain	5
1	A	610	TYR	Sidechain	4
1	A	622	ARG	Sidechain	3
1	A	534	ARG	Sidechain	2
1	A	544	ARG	Sidechain	2
1	A	579	ARG	Sidechain	2
1	A	631	ALA	Peptide	2
1	A	565	LYS	Peptide	1
1	A	576	TYR	Sidechain	1
1	A	634	SER	Peptide	1
1	A	562	ARG	Sidechain	1
1	A	644	TYR	Sidechain	1
1	A	554	ARG	Sidechain	1
1	A	578	VAL	Mainchain	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	988	1016	1014	4±2
All	All	19760	20320	20280	86

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:617:LEU:HD11	1:A:644:TYR:CD2	0.74	2.17	12	7
1:A:578:VAL:O	1:A:578:VAL:HG12	0.65	1.90	17	1
1:A:536:MET:SD	1:A:604:VAL:HG22	0.65	2.30	7	1
1:A:546:LEU:HB3	1:A:603:LEU:HD23	0.64	1.69	8	1
1:A:604:VAL:HG13	1:A:632:VAL:HB	0.64	1.70	16	1
1:A:589:VAL:HG11	1:A:615:PHE:CE2	0.61	2.31	3	3
1:A:627:LEU:HD23	1:A:629:ALA:H	0.58	1.57	7	4
1:A:541:GLU:CB	1:A:578:VAL:HG22	0.57	2.29	12	1
1:A:605:THR:HG23	1:A:606:MET:O	0.55	2.01	2	1
1:A:578:VAL:HG13	1:A:580:PHE:CE1	0.52	2.39	13	2
1:A:617:LEU:HD12	1:A:647:TYR:CD2	0.52	2.39	8	7
1:A:526:LEU:HD13	1:A:527:GLY:H	0.52	1.63	14	1
1:A:617:LEU:O	1:A:617:LEU:HD13	0.52	2.04	18	6
1:A:541:GLU:HB3	1:A:578:VAL:CG2	0.52	2.35	12	1
1:A:567:ILE:HD11	1:A:570:GLN:HE22	0.52	1.63	1	5
1:A:617:LEU:HD12	1:A:647:TYR:CE2	0.51	2.41	12	5
1:A:578:VAL:HG22	1:A:580:PHE:CE1	0.51	2.41	10	2
1:A:574:VAL:HG13	1:A:580:PHE:HB2	0.49	1.84	9	1
1:A:564:TYR:O	1:A:567:ILE:HG22	0.46	2.11	16	3
1:A:607:PRO:HB2	1:A:620:ALA:HB1	0.46	1.87	20	1
1:A:596:LEU:HA	1:A:599:LEU:HD23	0.45	1.88	11	1
1:A:574:VAL:HG21	1:A:582:PHE:CD2	0.45	2.47	2	1
1:A:541:GLU:HB3	1:A:578:VAL:HG22	0.45	1.89	12	1
1:A:537:LEU:CD2	1:A:563:LYS:HE3	0.44	2.42	14	2
1:A:627:LEU:CD2	1:A:629:ALA:H	0.44	2.25	7	1
1:A:541:GLU:CB	1:A:578:VAL:CG2	0.44	2.96	12	1
1:A:599:LEU:HD12	1:A:599:LEU:N	0.44	2.27	1	1
1:A:604:VAL:HG13	1:A:632:VAL:CB	0.44	2.41	16	1
1:A:610:TYR:CD2	1:A:611:VAL:HG23	0.43	2.49	20	1
1:A:546:LEU:HD22	1:A:547:MET:H	0.43	1.74	10	1
1:A:578:VAL:O	1:A:578:VAL:CG1	0.43	2.65	17	1
1:A:551:MET:HA	1:A:557:MET:SD	0.42	2.54	20	2
1:A:594:THR:HA	1:A:597:ASN:ND2	0.42	2.29	12	2
1:A:593:ILE:HG23	1:A:627:LEU:HG	0.42	1.91	7	1
1:A:627:LEU:HD23	1:A:628:LYS:N	0.42	2.30	17	1
1:A:578:VAL:HG13	1:A:580:PHE:HE1	0.42	1.71	13	1
1:A:564:TYR:CD1	1:A:577:GLY:HA3	0.42	2.50	5	1
1:A:537:LEU:CD2	1:A:563:LYS:HE2	0.42	2.44	7	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:606:MET:HA	1:A:634:SER:H	0.42	1.73	19	1
1:A:578:VAL:O	1:A:579:ARG:C	0.42	2.58	17	1
1:A:589:VAL:HG21	1:A:609:GLY:HA3	0.42	1.92	14	1
1:A:553:VAL:HG22	1:A:556:ILE:H	0.41	1.73	19	2
1:A:546:LEU:HD23	1:A:547:MET:N	0.41	2.30	8	1
1:A:546:LEU:HD22	1:A:603:LEU:HD22	0.41	1.91	8	1
1:A:617:LEU:HD11	1:A:644:TYR:HD2	0.41	1.66	12	1
1:A:617:LEU:HD13	1:A:617:LEU:C	0.40	2.36	6	1
1:A:617:LEU:HD13	1:A:617:LEU:O	0.40	2.16	6	1
1:A:569:ILE:CA	1:A:570:GLN:HE21	0.40	2.30	17	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	126/143 (88%)	108±2 (86±2%)	16±3 (13±2%)	2±1 (2±1%)	12	54
All	All	2520/2860 (88%)	2157 (86%)	318 (13%)	45 (2%)	12	54

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	569	ILE	14
1	A	608	ILE	14
1	A	565	LYS	3
1	A	532	ASN	3
1	A	566	GLY	2
1	A	572	GLY	2
1	A	573	ILE	2
1	A	612	THR	1
1	A	544	ARG	1
1	A	611	VAL	1
1	A	578	VAL	1
1	A	579	ARG	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	109/123 (89%)	92±3 (85±3%)	16±3 (15±3%)	6 44
All	All	2180/2460 (89%)	1850 (85%)	330 (15%)	6 44

All 67 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	598	SER	19
1	A	606	MET	15
1	A	570	GLN	14
1	A	562	ARG	13
1	A	594	THR	11
1	A	613	HIS	10
1	A	541	GLU	10
1	A	563	LYS	10
1	A	557	MET	10
1	A	526	LEU	8
1	A	612	THR	8
1	A	623	CYS	8
1	A	533	LEU	8
1	A	611	VAL	8
1	A	635	VAL	7
1	A	600	ASN	7
1	A	525	ILE	7
1	A	617	LEU	7
1	A	643	THR	7
1	A	626	SER	6
1	A	622	ARG	6
1	A	645	ASN	6
1	A	544	ARG	6
1	A	639	ASP	6
1	A	584	THR	6
1	A	552	ASP	6
1	A	585	SER	5
1	A	534	ARG	5
1	A	524	GLU	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	634	SER	5
1	A	592	ILE	5
1	A	530	SER	5
1	A	567	ILE	4
1	A	648	LEU	4
1	A	579	ARG	3
1	A	591	SER	3
1	A	636	SER	3
1	A	628	LYS	3
1	A	596	LEU	3
1	A	532	ASN	3
1	A	576	TYR	3
1	A	554	ARG	2
1	A	586	LYS	2
1	A	595	LYS	2
1	A	542	GLU	2
1	A	529	VAL	2
1	A	556	ILE	2
1	A	587	GLU	2
1	A	568	LYS	2
1	A	637	SER	2
1	A	565	LYS	2
1	A	624	MET	2
1	A	536	MET	2
1	A	578	VAL	2
1	A	605	THR	2
1	A	535	GLU	2
1	A	561	GLN	2
1	A	597	ASN	1
1	A	528	THR	1
1	A	547	MET	1
1	A	602	PRO	1
1	A	649	THR	1
1	A	599	LEU	1
1	A	625	ARG	1
1	A	548	PRO	1
1	A	571	GLU	1
1	A	589	VAL	1

6.3.3 RNA

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 (i)

No chemical shift data were provided