Full wwPDB NMR Structure Validation Report (i)

Jan 25, 2023 – 04:38 AM EST

PDB ID	:	2JZD
Title	:	NMR structure of the domain 527-651 of the SARS-CoV nonstructural protein nsp3
Authors	:	Chatterjee, A.; Johnson, M.A.; Serrano, P.; Pedrini, B.; Joseph, J.; Saikatendu, K.; Neuman, B.; Stevens, R.C.; Wilson, I.A.; Buchmeier, M.J.; Kuhn, P.; Wuthrich, K.; Joint Center for Structural Genomics (JCSG)
Deposited on	:	2008-01-04

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:

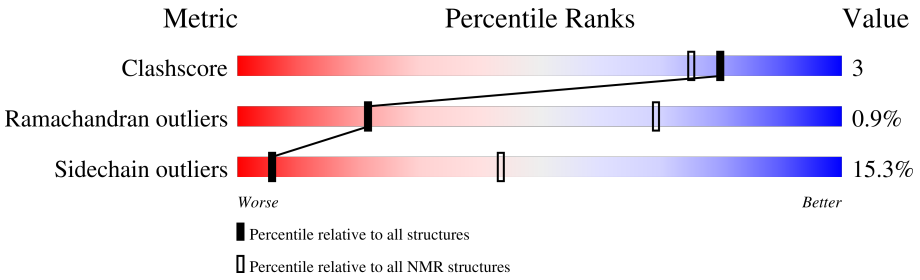
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.31.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al.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

1 Overall quality at a glance

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	129	84% 10% 5%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 6 as representative, based on the following criterion: *closest to the average*.

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:527-A:648 (122)	0.36	7

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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2026 atoms, of which 1023 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	129	2026	638	1023	171	185	9	0

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:

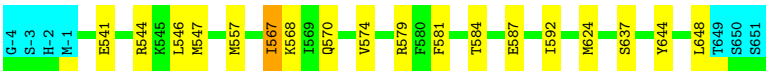
4.2 Scores per residue for each member of the ensemble

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

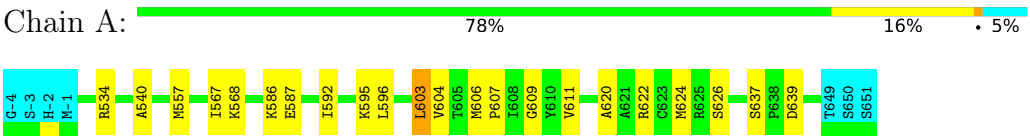
• Molecule 1: Replicase polyprotein 1ab

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Replicase polyprotein 1ab

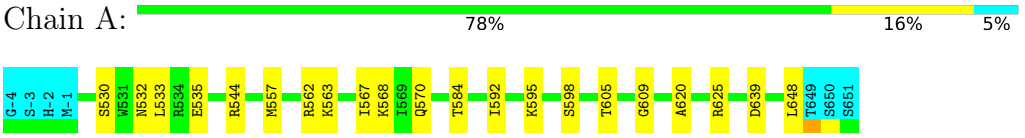
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Replicase polyprotein 1ab

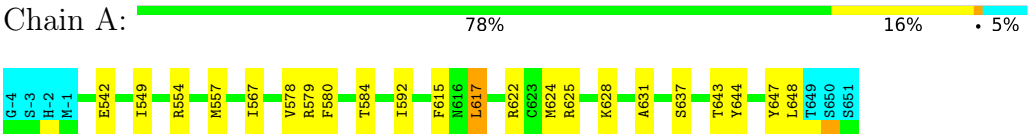
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Replicase polyprotein 1ab

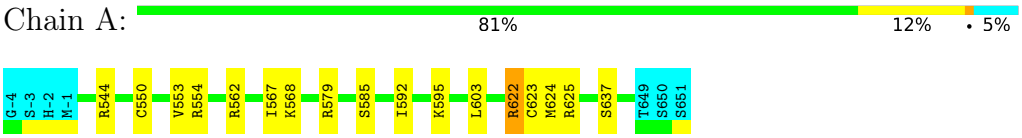
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Replicase polyprotein 1ab

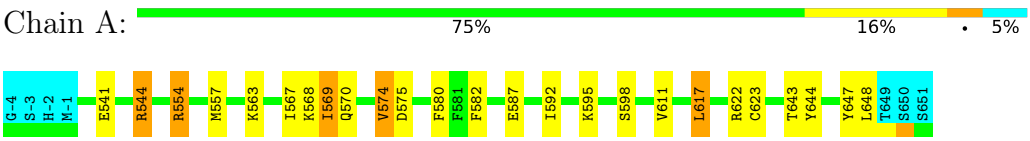
Chain A:



4.2.7 Score per residue for model 7 (medoid)

• Molecule 1: Replicase polyprotein 1ab

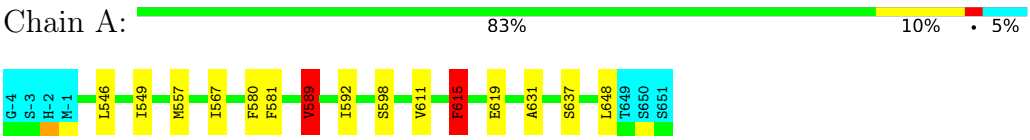
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Replicase polyprotein 1ab

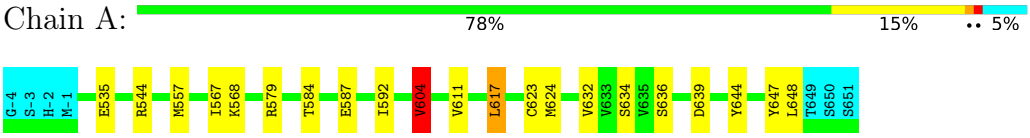
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Replicase polyprotein 1ab

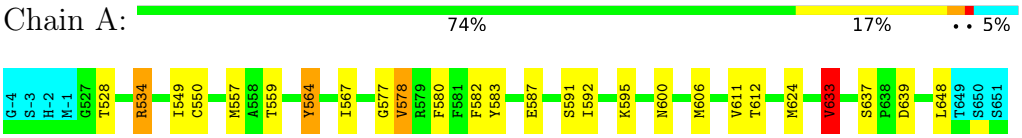
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Replicase polyprotein 1ab

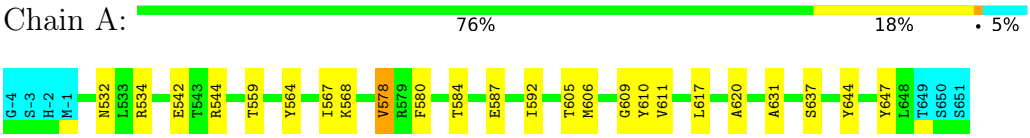
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Replicase polyprotein 1ab

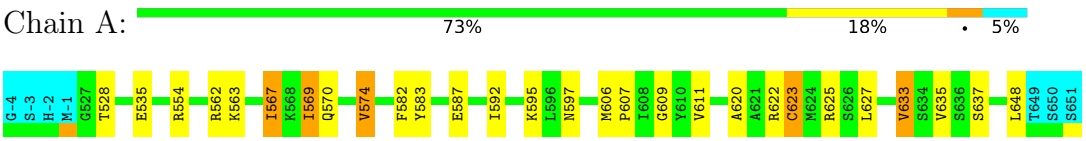
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Replicase polyprotein 1ab

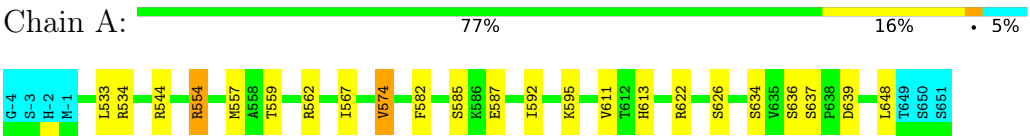
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Replicase polyprotein 1ab

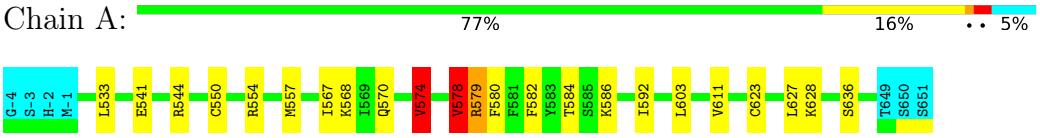
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Replicase polyprotein 1ab

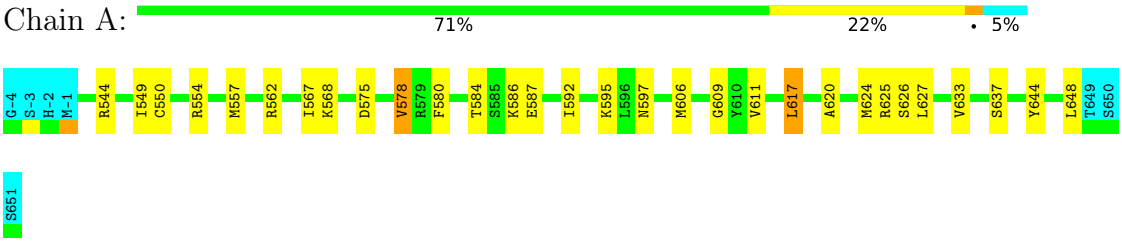
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Replicase polyprotein 1ab

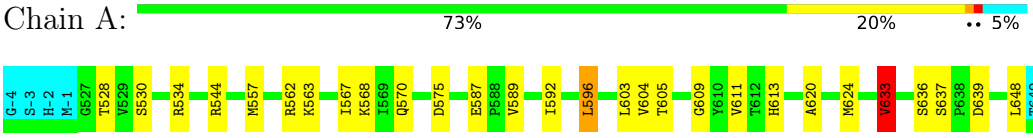
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Replicase polyprotein 1ab

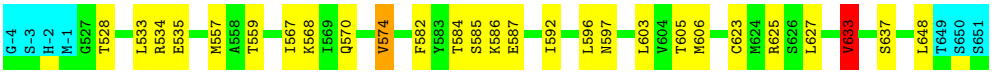
Chain A:



4.2.17 Score per residue for model 17

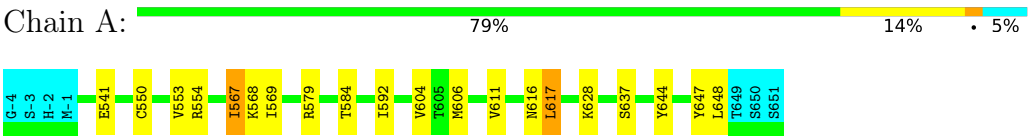
• Molecule 1: Replicase polyprotein 1ab

Chain A:



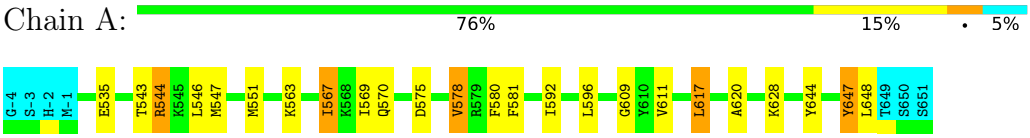
4.2.18 Score per residue for model 18

• Molecule 1: Replicase polyprotein 1ab



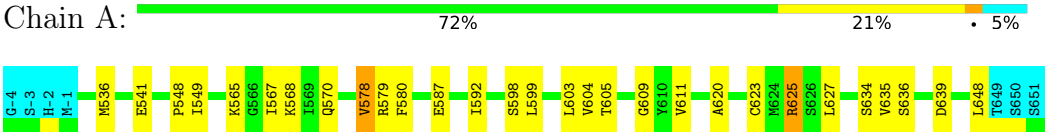
4.2.19 Score per residue for model 19

• Molecule 1: Replicase polyprotein 1ab



4.2.20 Score per residue for model 20

• Molecule 1: Replicase polyprotein 1ab



5 Refinement protocol and experimental data overview

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

Of the 80 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 solution	1.2
CYANA	refinement	1.2

No chemical shift data was provided.

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	0.60±0.01	0±0/976 (0.0± 0.0%)	1.07±0.04	1±1/1323 (0.1± 0.1%)
All	All	0.60	0/19520 (0.0%)	1.07	27/26460 (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	1.4±1.2
All	All	0	29

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	641	VAL	CA-CB-CG1	9.29	124.83	110.90	2	1
1	A	578	VAL	CA-CB-CG1	8.43	123.54	110.90	14	1
1	A	633	VAL	CA-CB-CG1	8.36	123.44	110.90	10	4
1	A	625	ARG	NE-CZ-NH2	-6.31	117.14	120.30	12	3
1	A	564	TYR	CB-CG-CD2	-6.28	117.23	121.00	10	1
1	A	641	VAL	CG1-CB-CG2	-6.18	101.01	110.90	2	1
1	A	577	GLY	C-N-CA	6.09	136.94	121.70	10	1
1	A	615	PHE	CB-CG-CD2	-6.06	116.56	120.80	5	2
1	A	579	ARG	NE-CZ-NH2	-6.03	117.28	120.30	14	1
1	A	603	LEU	CB-CG-CD2	5.99	121.18	111.00	3	1
1	A	554	ARG	NE-CZ-NH1	5.94	123.27	120.30	15	1
1	A	562	ARG	NE-CZ-NH2	-5.76	117.42	120.30	15	2
1	A	574	VAL	CA-CB-CG1	5.54	119.20	110.90	14	1
1	A	554	ARG	NE-CZ-NH2	-5.51	117.55	120.30	18	1
1	A	562	ARG	NE-CZ-NH1	5.47	123.03	120.30	4	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	644	TYR	CB-CG-CD1	-5.27	117.84	121.00	1	1
1	A	625	ARG	NE-CZ-NH1	5.18	122.89	120.30	20	1
1	A	604	VAL	CA-CB-CG1	5.11	118.57	110.90	9	1
1	A	534	ARG	NE-CZ-NH2	-5.03	117.78	120.30	17	1
1	A	579	ARG	NE-CZ-NH1	5.01	122.81	120.30	1	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	622	ARG	Sidechain	5
1	A	631	ALA	Peptide	4
1	A	625	ARG	Sidechain	4
1	A	579	ARG	Sidechain	2
1	A	554	ARG	Sidechain	2
1	A	534	ARG	Sidechain	2
1	A	564	TYR	Sidechain	2
1	A	583	TYR	Sidechain	2
1	A	544	ARG	Sidechain	1
1	A	580	PHE	Sidechain	1
1	A	615	PHE	Sidechain	1
1	A	610	TYR	Sidechain	1
1	A	562	ARG	Sidechain	1
1	A	647	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	956	981	979	5±4
All	All	19120	19620	19580	103

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:574:VAL:HG11	1:A:582:PHE:CE2	0.79	2.12	17	5
1:A:617:LEU:HD11	1:A:644:TYR:CD2	0.77	2.15	19	7
1:A:549:ILE:HD11	1:A:580:PHE:CD1	0.68	2.24	10	5
1:A:528:THR:HA	1:A:633:VAL:HG13	0.63	1.70	17	4
1:A:570:GLN:NE2	1:A:574:VAL:HG12	0.60	2.12	17	1
1:A:578:VAL:HG13	1:A:580:PHE:CZ	0.55	2.35	11	4
1:A:578:VAL:CG1	1:A:580:PHE:CZ	0.54	2.90	15	4
1:A:528:THR:CA	1:A:633:VAL:HG13	0.54	2.32	17	3
1:A:609:GLY:O	1:A:620:ALA:HB2	0.53	2.03	16	8
1:A:589:VAL:HG11	1:A:615:PHE:CZ	0.52	2.39	8	1
1:A:540:ALA:HB2	1:A:604:VAL:CG1	0.52	2.34	3	1
1:A:617:LEU:HD11	1:A:644:TYR:HD2	0.52	1.64	19	3
1:A:569:ILE:HD12	1:A:582:PHE:CZ	0.52	2.40	7	1
1:A:578:VAL:HG13	1:A:580:PHE:CE2	0.52	2.40	14	3
1:A:574:VAL:HG11	1:A:582:PHE:CD2	0.51	2.40	17	1
1:A:617:LEU:HD12	1:A:647:TYR:CE2	0.50	2.41	19	5
1:A:617:LEU:C	1:A:617:LEU:HD13	0.50	2.27	15	6
1:A:567:ILE:CD1	1:A:569:ILE:HD13	0.50	2.37	18	2
1:A:546:LEU:HD23	1:A:547:MET:N	0.50	2.22	1	2
1:A:578:VAL:CG1	1:A:580:PHE:CE2	0.49	2.95	10	1
1:A:567:ILE:HD13	1:A:569:ILE:H	0.47	1.68	12	1
1:A:546:LEU:HD21	1:A:581:PHE:HB2	0.46	1.86	19	2
1:A:597:ASN:OD1	1:A:627:LEU:HD12	0.46	2.10	17	3
1:A:596:LEU:HD11	1:A:603:LEU:HD21	0.46	1.87	16	1
1:A:617:LEU:HD13	1:A:617:LEU:O	0.46	2.10	9	4
1:A:617:LEU:HD12	1:A:647:TYR:CD2	0.46	2.46	11	6
1:A:548:PRO:HG2	1:A:605:THR:HG21	0.45	1.89	20	1
1:A:617:LEU:CD1	1:A:644:TYR:CD2	0.45	2.98	15	5
1:A:549:ILE:HD11	1:A:580:PHE:HD1	0.44	1.69	10	1
1:A:599:LEU:HD12	1:A:599:LEU:H	0.44	1.72	20	1
1:A:604:VAL:HG13	1:A:632:VAL:HB	0.41	1.92	9	1
1:A:567:ILE:HG13	1:A:574:VAL:CG1	0.41	2.45	1	1
1:A:546:LEU:HD11	1:A:581:PHE:HB2	0.41	1.92	8	1
1:A:606:MET:HB3	1:A:607:PRO:HD2	0.41	1.92	12	1
1:A:596:LEU:HD12	1:A:603:LEU:HD21	0.41	1.93	17	1
1:A:536:MET:CE	1:A:634:SER:HB3	0.41	2.45	20	1
1:A:620:ALA:HA	1:A:623:CYS:SG	0.41	2.56	12	1
1:A:549:ILE:HD13	1:A:582:PHE:CE1	0.41	2.50	10	1
1:A:607:PRO:HA	1:A:624:MET:SD	0.40	2.56	3	1
1:A:589:VAL:CG2	1:A:613:HIS:CD2	0.40	3.03	16	1
1:A:617:LEU:C	1:A:617:LEU:CD1	0.40	2.89	15	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	122/129 (95%)	110±3 (90±2%)	11±3 (9±2%)	1±1 (1±1%)	21 69
All	All	2440/2580 (95%)	2193 (90%)	224 (9%)	23 (1%)	21 69

All 7 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	611	VAL	15
1	A	569	ILE	2
1	A	604	VAL	2
1	A	579	ARG	1
1	A	589	VAL	1
1	A	543	THR	1
1	A	544	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	105/111 (95%)	89±3 (85±3%)	16±3 (15±3%)	6 43
All	All	2100/2220 (95%)	1779 (85%)	321 (15%)	6 43

All 65 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	567	ILE	20
1	A	592	ILE	20
1	A	648	LEU	16

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	568	LYS	14
1	A	637	SER	14
1	A	557	MET	13
1	A	587	GLU	13
1	A	544	ARG	12
1	A	570	GLN	9
1	A	584	THR	9
1	A	623	CYS	8
1	A	595	LYS	8
1	A	624	MET	7
1	A	639	ASP	7
1	A	578	VAL	7
1	A	535	GLU	6
1	A	563	LYS	6
1	A	636	SER	6
1	A	606	MET	6
1	A	554	ARG	6
1	A	617	LEU	6
1	A	541	GLU	5
1	A	559	THR	5
1	A	603	LEU	5
1	A	550	CYS	5
1	A	574	VAL	5
1	A	534	ARG	4
1	A	586	LYS	4
1	A	533	LEU	4
1	A	598	SER	4
1	A	605	THR	4
1	A	628	LYS	4
1	A	575	ASP	4
1	A	633	VAL	4
1	A	562	ARG	3
1	A	596	LEU	3
1	A	626	SER	3
1	A	585	SER	3
1	A	579	ARG	3
1	A	530	SER	2
1	A	532	ASN	2
1	A	542	GLU	2
1	A	643	THR	2
1	A	553	VAL	2
1	A	622	ARG	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	604	VAL	2
1	A	634	SER	2
1	A	635	VAL	2
1	A	627	LEU	2
1	A	528	THR	1
1	A	552	ASP	1
1	A	597	ASN	1
1	A	618	GLU	1
1	A	641	VAL	1
1	A	642	THR	1
1	A	589	VAL	1
1	A	619	GLU	1
1	A	591	SER	1
1	A	600	ASN	1
1	A	612	THR	1
1	A	613	HIS	1
1	A	625	ARG	1
1	A	616	ASN	1
1	A	551	MET	1
1	A	565	LYS	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 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

No chemical shift data were provided