



Full wwPDB NMR Structure Validation Report ⓘ

Mar 5, 2026 – 07:27 PM UTC

PDB ID : 2KHI / pdb_00002khi
Title : NMR structure of the domain 4 of the E. coli ribosomal protein S1
Authors : Salah, P.; Bisaglia, M.; Aliprandi, P.; Uzan, M.; Sizun, C.; Bontems, F.
Deposited on : 2009-04-07

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

MolProbity : 4-5-2 with Phenix2.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
wwPDB-RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
wwPDB-ShiftChecker : v1.2
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

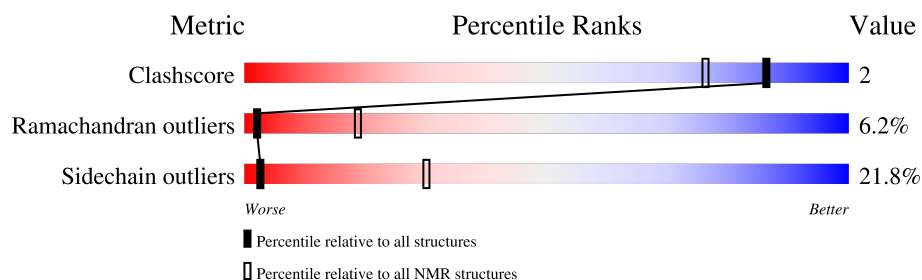
1 Overall quality at a glance

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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	115	33% 21% . 26% 17%

2 Ensemble composition and analysis

This entry contains 12 models. Model 6 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:274-A:306, A:322-A:353 (65)	1.12	6

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 and 2 single-model clusters were found.

Cluster number	Models
1	1, 4, 8, 10, 11, 12
2	2, 6, 7, 9
Single-model clusters	3; 5

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1510 atoms, of which 749 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called 30S ribosomal protein S1.

Mol	Chain	Residues	Atoms						Trace
1	A	95	Total	C	H	N	O	S	0
			1510	479	749	133	145	4	

There are 20 discrepancies between the modelled and reference sequences:

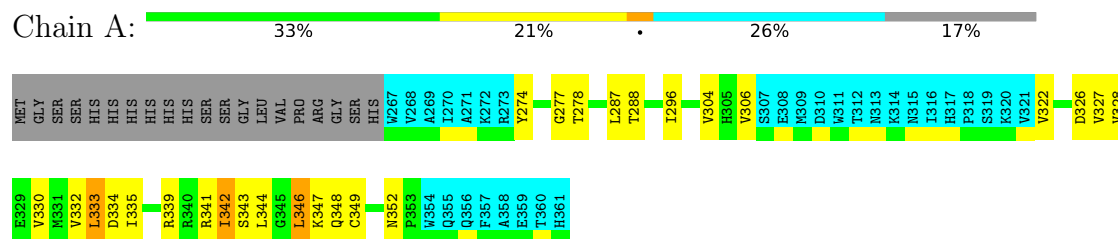
Chain	Residue	Modelled	Actual	Comment	Reference
A	247	MET	-	expression tag	UNP P0AG67
A	248	GLY	-	expression tag	UNP P0AG67
A	249	SER	-	expression tag	UNP P0AG67
A	250	SER	-	expression tag	UNP P0AG67
A	251	HIS	-	expression tag	UNP P0AG67
A	252	HIS	-	expression tag	UNP P0AG67
A	253	HIS	-	expression tag	UNP P0AG67
A	254	HIS	-	expression tag	UNP P0AG67
A	255	HIS	-	expression tag	UNP P0AG67
A	256	HIS	-	expression tag	UNP P0AG67
A	257	SER	-	expression tag	UNP P0AG67
A	258	SER	-	expression tag	UNP P0AG67
A	259	GLY	-	expression tag	UNP P0AG67
A	260	LEU	-	expression tag	UNP P0AG67
A	261	VAL	-	expression tag	UNP P0AG67
A	262	PRO	-	expression tag	UNP P0AG67
A	263	ARG	-	expression tag	UNP P0AG67
A	264	GLY	-	expression tag	UNP P0AG67
A	265	SER	-	expression tag	UNP P0AG67
A	266	HIS	-	expression tag	UNP P0AG67

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: 30S ribosomal protein S1

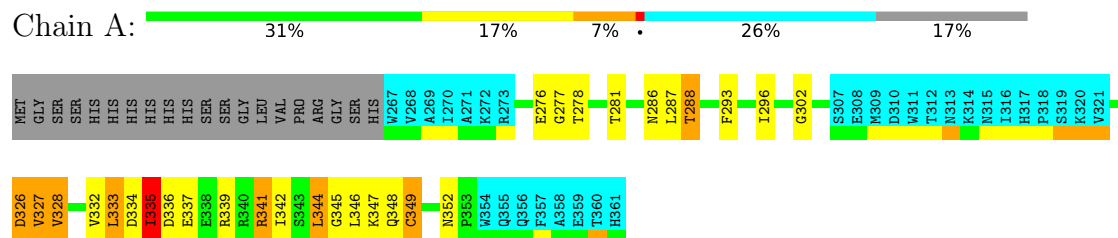


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

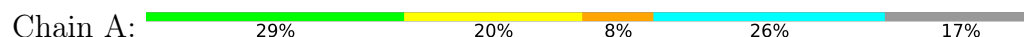
4.2.1 Score per residue for model 1

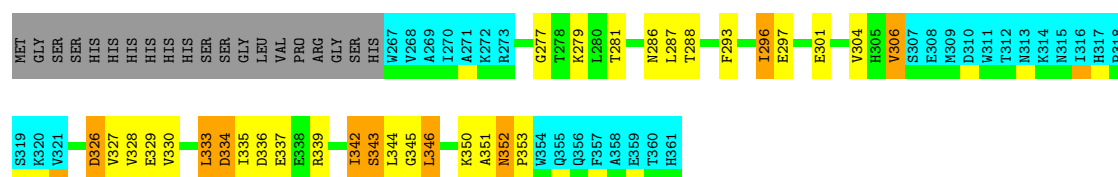
- Molecule 1: 30S ribosomal protein S1



4.2.2 Score per residue for model 2

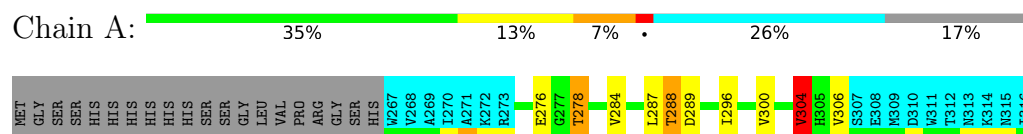
- Molecule 1: 30S ribosomal protein S1





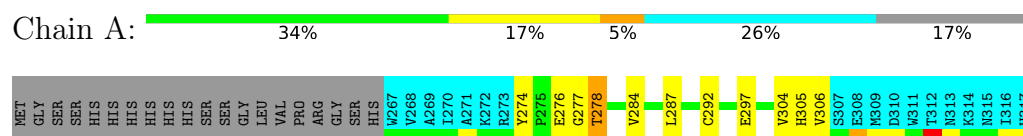
4.2.3 Score per residue for model 3

- Molecule 1: 30S ribosomal protein S1



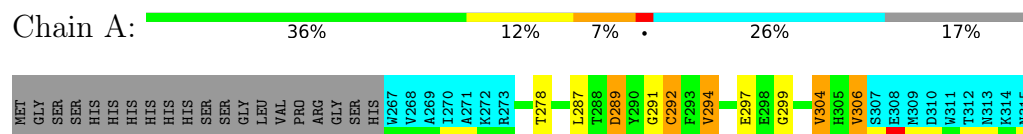
4.2.4 Score per residue for model 4

- Molecule 1: 30S ribosomal protein S1



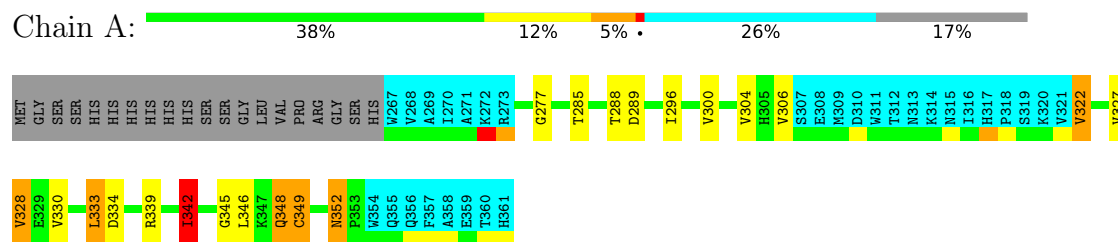
4.2.5 Score per residue for model 5

- Molecule 1: 30S ribosomal protein S1



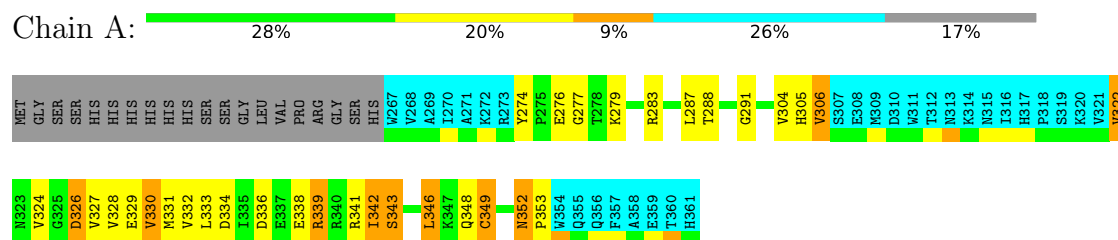
4.2.6 Score per residue for model 6 (medoid)

- Molecule 1: 30S ribosomal protein S1



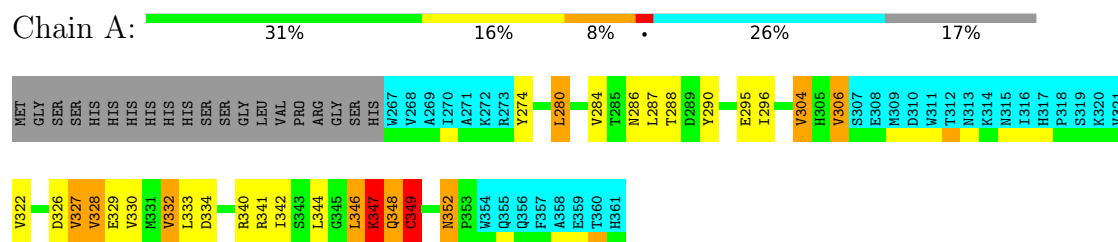
4.2.7 Score per residue for model 7

- Molecule 1: 30S ribosomal protein S1



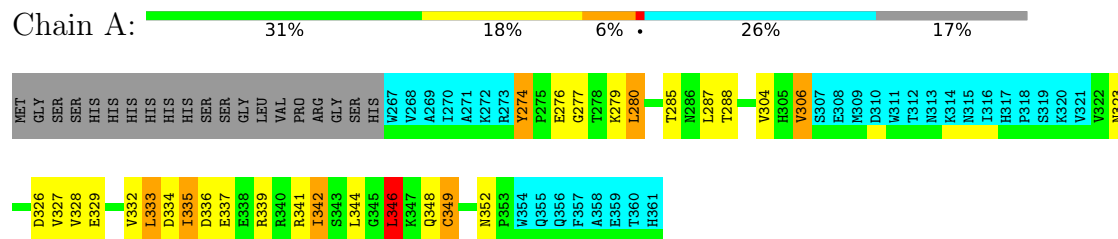
4.2.8 Score per residue for model 8

- Molecule 1: 30S ribosomal protein S1



4.2.9 Score per residue for model 9

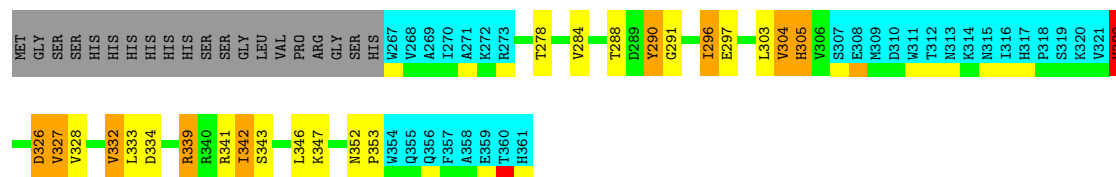
- Molecule 1: 30S ribosomal protein S1



4.2.10 Score per residue for model 10

- Molecule 1: 30S ribosomal protein S1

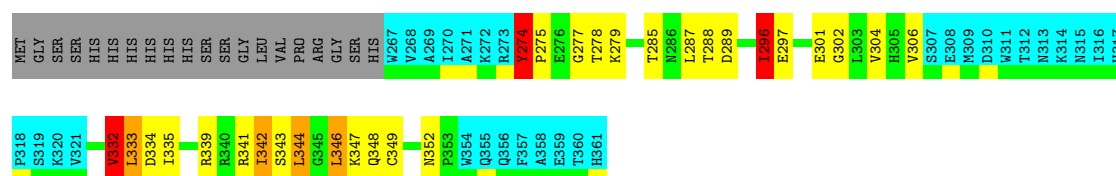
Chain A: 35% 13% 8% 26% 17%



4.2.11 Score per residue for model 11

- Molecule 1: 30S ribosomal protein S1

Chain A: 31% 19% 26% 17%



4.2.12 Score per residue for model 12

- Molecule 1: 30S ribosomal protein S1

Chain A: 31% 19% 5% 26% 17%



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 12 were deposited, based on the following criterion: *12 structures for lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
INCA	structure solution	
INCA	refinement	

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	1.40±0.03	0±0/509 (0.1± 0.1%)	2.15±0.05	25±4/690 (3.7± 0.6%)
All	All	1.40	4/6108 (0.1%)	2.15	305/8280 (3.7%)

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	0.2±0.4
All	All	0	3

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	343	SER	CA-C	-6.15	1.45	1.53	12	3
1	A	305	HIS	N-CA	-5.09	1.40	1.46	10	1

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	297	GLU	CA-C-N	14.10	143.07	120.94	5	1
1	A	297	GLU	C-N-CA	14.10	143.07	120.94	5	1
1	A	346	LEU	N-CA-C	13.82	128.78	110.53	3	7
1	A	348	GLN	CA-C-N	12.79	139.63	120.82	6	5
1	A	348	GLN	C-N-CA	12.79	139.63	120.82	6	5
1	A	342	ILE	N-CA-C	-9.95	94.18	108.11	12	10
1	A	334	ASP	N-CA-C	9.49	122.82	109.15	3	12
1	A	346	LEU	N-CA-CB	9.09	125.85	110.49	8	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	304	VAL	N-CA-CB	8.99	121.25	111.00	10	8
1	A	301	GLU	CA-C-N	8.95	130.44	122.47	11	1
1	A	301	GLU	C-N-CA	8.95	130.44	122.47	11	1
1	A	326	ASP	CA-CB-CG	8.92	121.52	112.60	9	6
1	A	338	GLU	CA-C-N	8.92	132.96	120.29	7	2
1	A	338	GLU	C-N-CA	8.92	132.96	120.29	7	2
1	A	342	ILE	CA-C-N	-8.61	109.22	123.33	5	12
1	A	342	ILE	C-N-CA	-8.61	109.22	123.33	5	12
1	A	341	ARG	CA-C-N	8.38	133.58	123.19	5	6
1	A	341	ARG	C-N-CA	8.38	133.58	123.19	5	6
1	A	284	VAL	N-CA-C	7.87	119.13	109.30	3	5
1	A	337	GLU	CA-C-N	7.86	134.02	120.58	2	5
1	A	337	GLU	C-N-CA	7.86	134.02	120.58	2	5
1	A	339	ARG	N-CA-CB	-7.67	98.81	110.16	7	1
1	A	277	GLY	CA-C-N	7.64	133.27	122.72	7	8
1	A	277	GLY	C-N-CA	7.64	133.27	122.72	7	8
1	A	306	VAL	CA-C-N	7.56	131.17	120.28	5	3
1	A	306	VAL	C-N-CA	7.56	131.17	120.28	5	3
1	A	327	VAL	CA-C-N	7.50	134.30	123.27	1	4
1	A	327	VAL	C-N-CA	7.50	134.30	123.27	1	4
1	A	335	ILE	N-CA-C	7.46	118.61	107.80	12	2
1	A	291	GLY	N-CA-C	-7.33	100.47	110.80	10	3
1	A	332	VAL	N-CA-C	7.27	118.48	108.89	11	8
1	A	345	GLY	CA-C-N	7.24	135.37	121.54	5	5
1	A	345	GLY	C-N-CA	7.24	135.37	121.54	5	5
1	A	352	ASN	CA-CB-CG	7.01	119.61	112.60	2	2
1	A	339	ARG	CA-C-N	6.97	134.86	121.54	9	8
1	A	339	ARG	C-N-CA	6.97	134.86	121.54	9	8
1	A	297	GLU	N-CA-C	-6.97	98.72	109.52	11	4
1	A	348	GLN	N-CA-C	-6.96	97.77	108.76	3	3
1	A	293	PHE	N-CA-CB	6.93	121.85	110.69	1	2
1	A	288	THR	CA-C-N	6.84	130.13	120.28	1	2
1	A	288	THR	C-N-CA	6.84	130.13	120.28	1	2
1	A	349	CYS	CA-C-N	6.80	134.53	121.54	12	1
1	A	349	CYS	C-N-CA	6.80	134.53	121.54	12	1
1	A	352	ASN	CA-C-N	6.71	126.39	119.82	8	5
1	A	352	ASN	C-N-CA	6.71	126.39	119.82	8	5
1	A	305	HIS	CA-CB-CG	6.70	120.50	113.80	4	1
1	A	349	CYS	N-CA-CB	6.57	121.59	110.49	9	3
1	A	329	GLU	N-CA-C	-6.49	98.82	109.40	2	4
1	A	297	GLU	CA-C-O	6.46	127.31	119.31	5	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	353	PRO	CA-C-N	6.39	133.74	121.54	7	1
1	A	353	PRO	C-N-CA	6.39	133.74	121.54	7	1
1	A	349	CYS	CA-C-O	6.38	127.78	120.00	3	1
1	A	274	TYR	CA-C-N	6.28	125.51	118.97	9	2
1	A	274	TYR	C-N-CA	6.28	125.51	118.97	9	2
1	A	294	VAL	CA-CB-CG1	6.22	120.98	110.40	5	1
1	A	350	LYS	CB-CG-CD	6.20	125.55	111.30	12	1
1	A	278	THR	N-CA-CB	6.18	120.24	110.23	5	1
1	A	347	LYS	N-CA-CB	6.14	120.86	110.49	8	1
1	A	280	LEU	N-CA-CB	6.13	121.55	111.62	9	1
1	A	335	ILE	CB-CA-C	6.11	119.86	111.31	9	3
1	A	306	VAL	N-CA-C	6.09	122.00	109.34	2	7
1	A	333	LEU	N-CA-C	-6.05	97.92	110.80	1	8
1	A	278	THR	N-CA-C	-5.99	100.80	110.32	11	4
1	A	323	ASN	CA-CB-CG	-5.99	106.61	112.60	9	1
1	A	351	ALA	CA-C-N	5.90	136.20	121.80	2	1
1	A	351	ALA	C-N-CA	5.90	136.20	121.80	2	1
1	A	296	ILE	CA-C-N	5.86	131.80	122.59	11	3
1	A	296	ILE	C-N-CA	5.86	131.80	122.59	11	3
1	A	335	ILE	CA-CB-CG1	5.84	120.33	110.40	3	2
1	A	286	ASN	CA-CB-CG	5.78	118.38	112.60	12	2
1	A	290	TYR	N-CA-CB	5.77	120.25	110.49	10	1
1	A	340	ARG	CB-CA-C	5.75	119.19	111.82	8	1
1	A	275	PRO	CA-C-N	5.63	128.87	120.87	11	1
1	A	275	PRO	C-N-CA	5.63	128.87	120.87	11	1
1	A	349	CYS	N-CA-C	5.57	117.48	110.24	6	3
1	A	327	VAL	CB-CA-C	5.49	118.76	110.63	8	1
1	A	276	GLU	N-CA-C	5.40	118.12	110.50	4	1
1	A	274	TYR	N-CA-C	5.40	112.99	108.13	11	2
1	A	304	VAL	CB-CA-C	-5.38	103.14	110.77	2	1
1	A	284	VAL	CA-CB-CG1	5.37	119.54	110.40	3	1
1	A	277	GLY	CA-C-O	5.36	124.75	118.96	4	4
1	A	352	ASN	N-CA-CB	5.33	119.86	110.37	2	2
1	A	336	ASP	CA-C-N	5.33	129.12	120.60	1	1
1	A	336	ASP	C-N-CA	5.33	129.12	120.60	1	1
1	A	289	ASP	N-CA-C	5.31	117.81	111.71	5	1
1	A	330	VAL	N-CA-C	-5.26	100.67	108.45	2	1
1	A	292	CYS	CB-CA-C	5.22	120.50	109.79	5	1
1	A	285	THR	CA-CB-OG1	5.19	117.38	109.60	9	1
1	A	333	LEU	N-CA-CB	5.14	119.18	110.49	4	2
1	A	328	VAL	CB-CA-C	5.11	118.38	110.82	9	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	284	VAL	CA-CB-CG2	-5.10	101.72	110.40	4	1
1	A	299	GLY	CA-C-N	5.10	129.68	123.10	5	1
1	A	299	GLY	C-N-CA	5.10	129.68	123.10	5	1
1	A	306	VAL	CA-CB-CG1	5.07	119.02	110.40	7	1
1	A	322	VAL	CB-CA-C	5.03	119.55	111.29	10	1
1	A	328	VAL	N-CA-CB	-5.01	102.96	111.23	6	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	305	HIS	Sidechain	1
1	A	290	TYR	Sidechain	1
1	A	274	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	503	505	505	2±1
All	All	6036	6060	6060	26

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:328:VAL:HG23	1:A:349:CYS:H	0.64	1.53	5	1
1:A:332:VAL:HG12	1:A:344:LEU:HD12	0.55	1.79	11	1
1:A:304:VAL:HG21	1:A:347:LYS:HG3	0.54	1.78	8	1
1:A:335:ILE:HD13	1:A:342:ILE:HG23	0.52	1.82	5	1
1:A:328:VAL:HG21	1:A:346:LEU:HG	0.51	1.82	5	1
1:A:304:VAL:CG2	1:A:346:LEU:H	0.51	2.18	3	1
1:A:328:VAL:HA	1:A:349:CYS:HA	0.50	1.81	8	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:332:VAL:HB	1:A:344:LEU:HD12	0.47	1.87	3	1
1:A:304:VAL:HG13	1:A:346:LEU:H	0.44	1.72	9	1
1:A:300:VAL:HG13	1:A:342:ILE:HG13	0.44	1.88	6	1
1:A:283:ARG:HG2	1:A:327:VAL:HG22	0.43	1.90	7	1
1:A:296:ILE:HG12	1:A:344:LEU:HD21	0.43	1.88	11	1
1:A:304:VAL:HG21	1:A:347:LYS:HB2	0.43	1.89	4	1
1:A:302:GLY:HA3	1:A:344:LEU:HD22	0.42	1.90	11	2
1:A:280:LEU:HD21	1:A:332:VAL:HG22	0.42	1.89	8	1
1:A:334:ASP:N	1:A:343:SER:HB2	0.42	2.29	12	1
1:A:304:VAL:HA	1:A:345:GLY:HA2	0.42	1.90	3	1
1:A:278:THR:O	1:A:332:VAL:HG12	0.42	2.14	10	2
1:A:296:ILE:HB	1:A:300:VAL:HG12	0.42	1.91	6	1
1:A:330:VAL:HA	1:A:346:LEU:HA	0.41	1.91	7	1
1:A:279:LYS:HA	1:A:331:MET:HA	0.41	1.91	7	1
1:A:336:ASP:HB3	1:A:339:ARG:HB2	0.40	1.91	7	1
1:A:328:VAL:HG13	1:A:349:CYS:HA	0.40	1.94	1	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	65/115 (57%)	52±2 (81±3%)	9±2 (13±2%)	4±1 (6±2%)	2	19
All	All	780/1380 (57%)	629 (81%)	103 (13%)	48 (6%)	2	19

All 13 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	352	ASN	10
1	A	296	ILE	6
1	A	349	CYS	6
1	A	276	GLU	5
1	A	347	LYS	5
1	A	322	VAL	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	306	VAL	4
1	A	353	PRO	2
1	A	274	TYR	1
1	A	346	LEU	1
1	A	348	GLN	1
1	A	350	LYS	1
1	A	351	ALA	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	57/101 (56%)	45±3 (78±4%)	12±3 (22±4%)	2	30
All	All	684/1212 (56%)	535 (78%)	149 (22%)	2	30

All 38 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	287	LEU	10
1	A	288	THR	10
1	A	333	LEU	10
1	A	342	ILE	10
1	A	327	VAL	9
1	A	344	LEU	8
1	A	328	VAL	7
1	A	346	LEU	7
1	A	330	VAL	6
1	A	335	ILE	5
1	A	341	ARG	5
1	A	343	SER	5
1	A	289	ASP	5
1	A	326	ASP	4
1	A	279	LYS	4
1	A	336	ASP	3
1	A	350	LYS	3
1	A	304	VAL	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	332	VAL	3
1	A	322	VAL	3
1	A	348	GLN	3
1	A	274	TYR	3
1	A	278	THR	2
1	A	281	THR	2
1	A	286	ASN	2
1	A	292	CYS	2
1	A	285	THR	2
1	A	306	VAL	2
1	A	280	LEU	2
1	A	301	GLU	1
1	A	334	ASP	1
1	A	300	VAL	1
1	A	294	VAL	1
1	A	324	VAL	1
1	A	290	TYR	1
1	A	295	GLU	1
1	A	339	ARG	1
1	A	296	ILE	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 oligosaccharides in this entry.

6.6 Ligand geometry ⓘ

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