



Full wwPDB NMR Structure Validation Report ⓘ

Mar 5, 2026 – 04:02 PM UTC

PDB ID : 1SOY / pdb_00001soy
Title : Solution structure of the bacterial frataxin orthologue, CyaY
Authors : Nair, M.; Adinolfi, S.; Pastore, C.; Kelly, G.; Temussi, P.; Pastore, A.
Deposited on : 2004-03-16

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

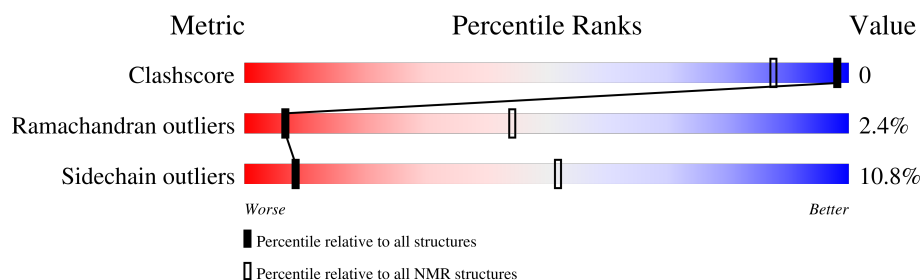
1 Overall quality at a glance

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	108	 88% 8% ...

2 Ensemble composition and analysis

This entry contains 20 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 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:2-A:106 (105)	1.12	3

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 16 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1656 atoms, of which 793 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CyaY protein.

Mol	Chain	Residues	Atoms						Trace
1	A	106	Total	C	H	N	O	S	0
			1656	539	793	146	175	3	

There are 2 discrepancies between the modelled and reference sequences:

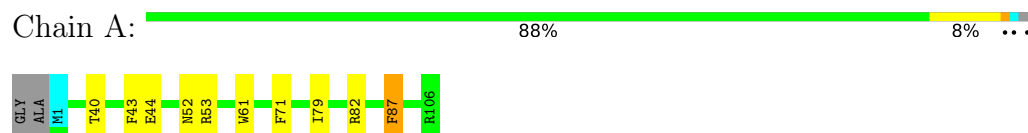
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	cloning artifact	UNP P27838
A	0	ALA	-	cloning artifact	UNP P27838

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: CyaY protein

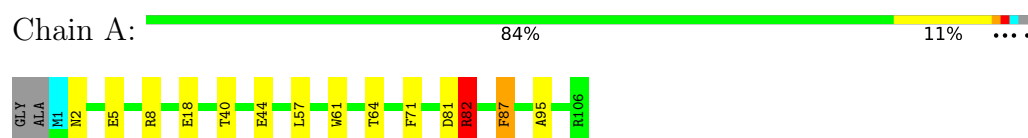


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

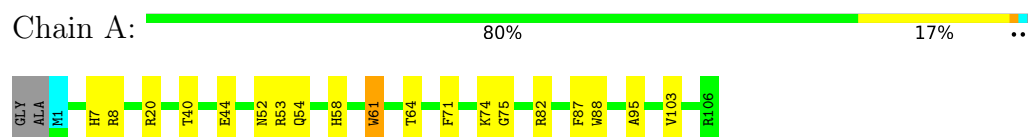
4.2.1 Score per residue for model 1

- Molecule 1: CyaY protein



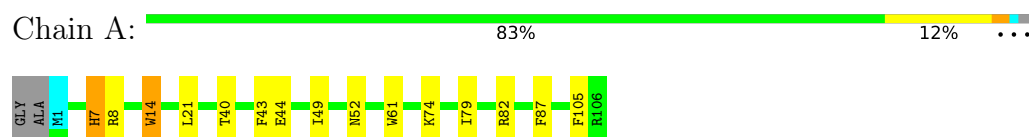
4.2.2 Score per residue for model 2

- Molecule 1: CyaY protein



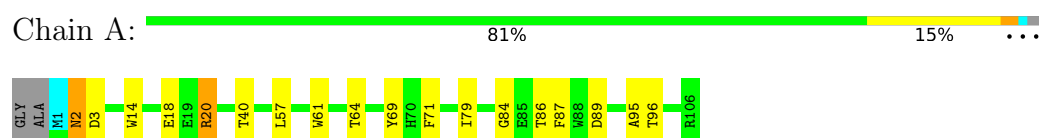
4.2.3 Score per residue for model 3 (medoid)

- Molecule 1: CyaY protein



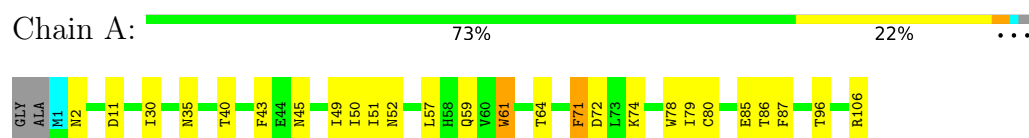
4.2.4 Score per residue for model 4

- Molecule 1: CyaY protein



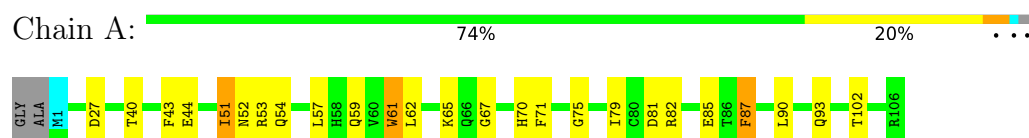
4.2.5 Score per residue for model 5

- Molecule 1: CyaY protein



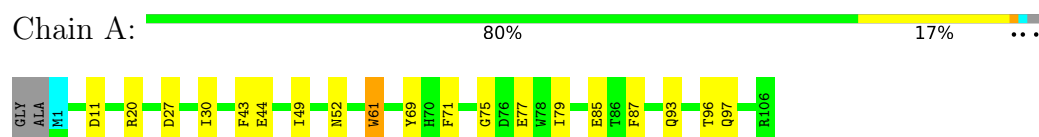
4.2.6 Score per residue for model 6

- Molecule 1: CyaY protein



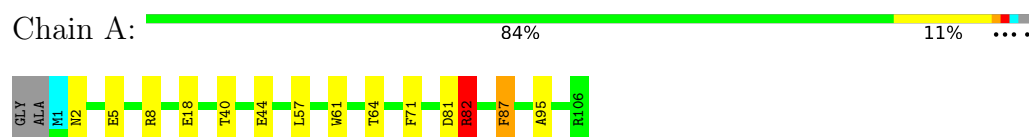
4.2.7 Score per residue for model 7

- Molecule 1: CyaY protein



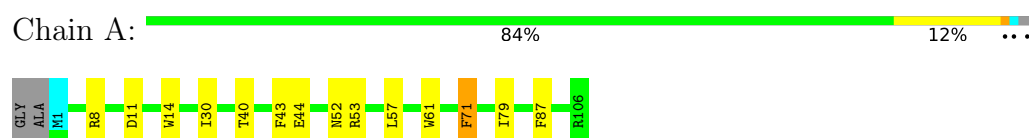
4.2.8 Score per residue for model 8

- Molecule 1: CyaY protein



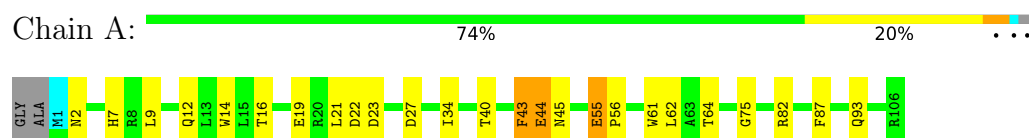
4.2.9 Score per residue for model 9

- Molecule 1: CyaY protein



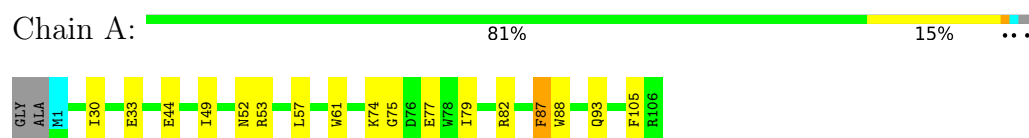
4.2.10 Score per residue for model 10

- Molecule 1: CyaY protein



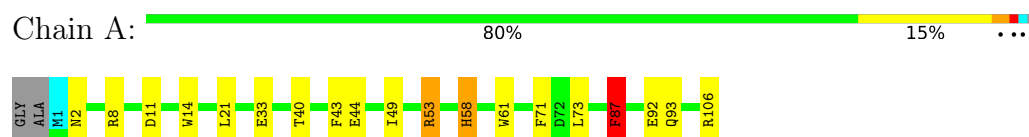
4.2.11 Score per residue for model 11

- Molecule 1: CyaY protein



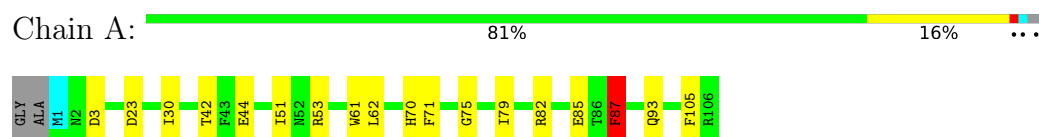
4.2.12 Score per residue for model 12

- Molecule 1: CyaY protein



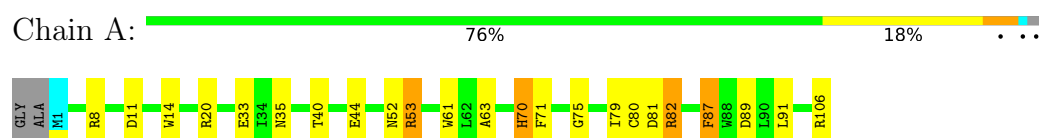
4.2.13 Score per residue for model 13

- Molecule 1: CyaY protein



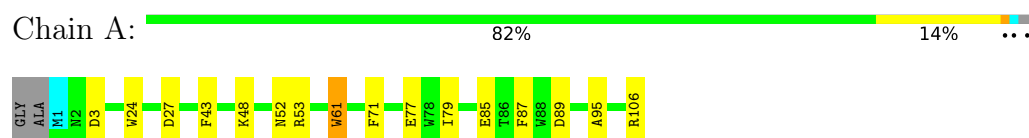
4.2.14 Score per residue for model 14

- Molecule 1: CyaY protein



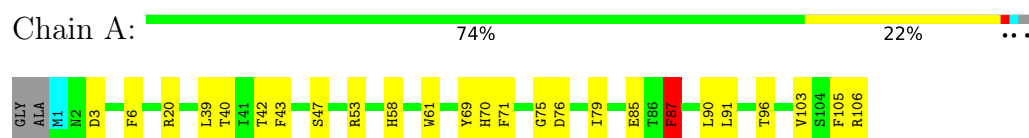
4.2.15 Score per residue for model 15

- Molecule 1: CyaY protein



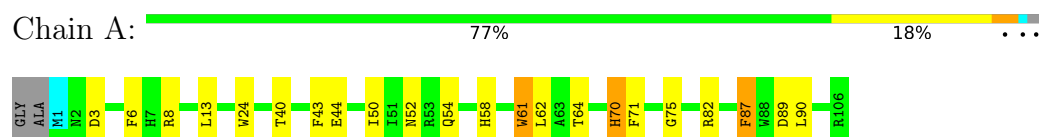
4.2.16 Score per residue for model 16

- Molecule 1: CyaY protein



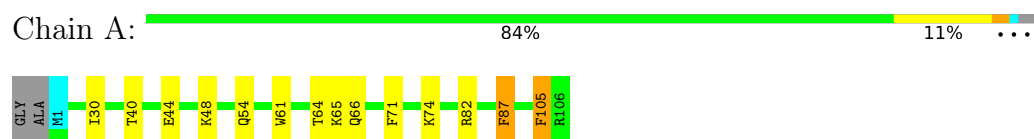
4.2.17 Score per residue for model 17

- Molecule 1: CyaY protein



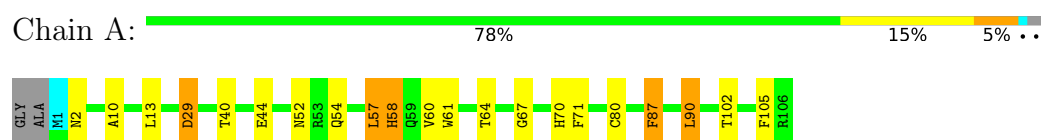
4.2.18 Score per residue for model 18

- Molecule 1: CyaY protein



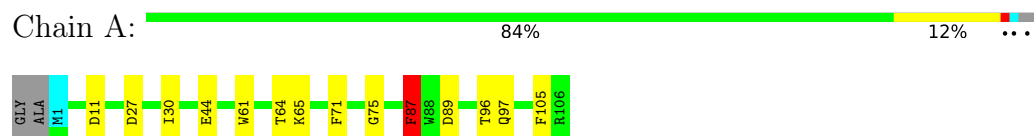
4.2.19 Score per residue for model 19

- Molecule 1: CyaY protein



4.2.20 Score per residue for model 20

- Molecule 1: CyaY protein



5 Refinement protocol and experimental data overview

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

Of the 250 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy*.

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

Software name	Classification	Version
CYANA	structure solution	1.0.6
X-PLOR	refinement	2.8.1

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	1.42±0.01	0±0/874 (0.0± 0.0%)	1.61±0.03	9±4/1184 (0.7± 0.3%)
All	All	1.42	5/17480 (0.0%)	1.61	171/23680 (0.7%)

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	1.6±1.3
All	All	0	32

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	64	THR	CA-C	-6.03	1.49	1.52	19	1
1	A	57	LEU	CA-C	-5.88	1.49	1.52	1	3
1	A	74	LYS	CA-C	-5.42	1.49	1.52	18	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	71	PHE	CA-CB-CG	9.61	123.41	113.80	5	12
1	A	87	PHE	CA-CB-CG	9.03	122.83	113.80	12	12
1	A	64	THR	CA-C-N	7.44	130.12	120.44	1	3
1	A	64	THR	C-N-CA	7.44	130.12	120.44	1	3
1	A	85	GLU	N-CA-C	6.84	119.42	108.41	7	4
1	A	71	PHE	N-CA-CB	-6.79	99.70	110.77	14	9
1	A	52	ASN	CA-CB-CG	6.45	119.05	112.60	19	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	12	GLN	N-CA-C	6.33	117.84	111.07	10	1
1	A	90	LEU	CA-C-N	6.22	128.94	120.54	6	2
1	A	90	LEU	C-N-CA	6.22	128.94	120.54	6	2
1	A	70	HIS	CA-CB-CG	6.21	120.01	113.80	16	1
1	A	7	HIS	CA-CB-CG	6.14	119.94	113.80	10	3
1	A	58	HIS	CB-CG-CD2	-6.08	123.29	131.20	12	1
1	A	44	GLU	N-CA-C	6.03	118.63	111.33	14	7
1	A	2	ASN	CA-C-N	6.00	128.24	120.44	5	5
1	A	2	ASN	C-N-CA	6.00	128.24	120.44	5	5
1	A	43	PHE	CA-C-N	5.88	128.44	120.38	6	7
1	A	43	PHE	C-N-CA	5.88	128.44	120.38	6	7
1	A	105	PHE	CA-CB-CG	5.76	119.56	113.80	18	2
1	A	80	CYS	CA-C-N	5.73	128.27	120.54	5	2
1	A	80	CYS	C-N-CA	5.73	128.27	120.54	5	2
1	A	3	ASP	CA-CB-CG	5.72	118.32	112.60	16	2
1	A	6	PHE	CA-CB-CG	-5.72	108.08	113.80	16	1
1	A	57	LEU	CA-C-N	5.68	132.38	121.54	19	1
1	A	57	LEU	C-N-CA	5.68	132.38	121.54	19	1
1	A	23	ASP	CA-C-N	5.67	128.20	120.54	13	1
1	A	23	ASP	C-N-CA	5.67	128.20	120.54	13	1
1	A	51	ILE	CB-CA-C	5.65	116.94	111.23	6	1
1	A	13	LEU	CA-C-N	5.62	127.75	120.44	17	2
1	A	13	LEU	C-N-CA	5.62	127.75	120.44	17	2
1	A	29	ASP	N-CA-CB	-5.60	103.74	112.41	19	1
1	A	22	ASP	CA-C-N	5.59	128.08	120.54	10	1
1	A	22	ASP	C-N-CA	5.59	128.08	120.54	10	1
1	A	93	GLN	CA-C-N	5.58	128.02	120.38	10	6
1	A	93	GLN	C-N-CA	5.58	128.02	120.38	10	6
1	A	95	ALA	CA-C-N	5.42	127.86	120.54	1	5
1	A	95	ALA	C-N-CA	5.42	127.86	120.54	1	5
1	A	105	PHE	CB-CA-C	5.40	118.14	109.07	18	1
1	A	102	THR	N-CA-CB	5.32	118.01	110.46	19	2
1	A	81	ASP	CA-C-N	5.28	131.63	121.54	1	2
1	A	81	ASP	C-N-CA	5.28	131.63	121.54	1	2
1	A	103	VAL	CA-C-N	5.28	128.84	121.24	2	1
1	A	103	VAL	C-N-CA	5.28	128.84	121.24	2	1
1	A	55	GLU	N-CA-C	5.26	117.24	109.50	10	1
1	A	96	THR	N-CA-CB	5.26	117.85	110.12	5	1
1	A	10	ALA	CA-C-N	5.24	127.61	120.54	19	1
1	A	10	ALA	C-N-CA	5.24	127.61	120.54	19	1
1	A	72	ASP	CA-CB-CG	5.21	117.81	112.60	5	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	14	TRP	CA-CB-CG	5.21	123.50	113.60	3	1
1	A	96	THR	CA-C-N	5.20	127.71	120.63	7	2
1	A	96	THR	C-N-CA	5.20	127.71	120.63	7	2
1	A	50	ILE	CB-CA-C	5.19	117.43	110.42	5	1
1	A	91	LEU	CA-C-N	5.13	127.11	120.44	16	2
1	A	91	LEU	C-N-CA	5.13	127.11	120.44	16	2
1	A	9	LEU	CA-C-N	5.11	127.09	120.44	10	1
1	A	9	LEU	C-N-CA	5.11	127.09	120.44	10	1
1	A	5	GLU	CA-C-N	5.10	127.42	120.54	1	2
1	A	5	GLU	C-N-CA	5.10	127.42	120.54	1	2
1	A	23	ASP	N-CA-C	5.08	117.21	111.11	10	1
1	A	90	LEU	N-CA-CB	5.06	117.66	110.16	19	1
1	A	29	ASP	CB-CA-C	5.04	119.58	111.36	19	1
1	A	92	GLU	CA-C-N	5.04	127.34	120.54	12	1
1	A	92	GLU	C-N-CA	5.04	127.34	120.54	12	1
1	A	16	THR	CA-C-N	5.01	126.88	120.56	10	1
1	A	16	THR	C-N-CA	5.01	126.88	120.56	10	1
1	A	78	TRP	CA-C-N	5.00	129.39	123.19	5	1
1	A	78	TRP	C-N-CA	5.00	129.39	123.19	5	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	82	ARG	Sidechain	7
1	A	53	ARG	Sidechain	6
1	A	8	ARG	Sidechain	5
1	A	106	ARG	Sidechain	4
1	A	20	ARG	Sidechain	3
1	A	70	HIS	Peptide	3
1	A	105	PHE	Sidechain	2
1	A	84	GLY	Peptide	1
1	A	63	ALA	Peptide	1

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	855	785	785	1±1
All	All	17100	15700	15700	14

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:52:ASN:H	1:A:61:TRP:CD1	0.60	2.13	17	6
1:A:57:LEU:O	1:A:58:HIS:CD2	0.51	2.64	19	1
1:A:58:HIS:C	1:A:58:HIS:CD2	0.49	2.90	12	1
1:A:96:THR:HG22	1:A:103:VAL:H	0.45	1.72	16	1
1:A:49:ILE:HD12	1:A:64:THR:HG23	0.44	1.88	5	1
1:A:70:HIS:CE1	1:A:81:ASP:OD2	0.43	2.71	14	2
1:A:29:ASP:HA	1:A:44:GLU:H	0.42	1.74	19	1
1:A:61:TRP:CD1	1:A:61:TRP:C	0.41	2.98	2	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	104/108 (96%)	89±3 (86±3%)	12±3 (12±3%)	2±1 (2±1%)	7	44
All	All	2080/2160 (96%)	1781 (86%)	249 (12%)	50 (2%)	7	44

All 11 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	87	PHE	20
1	A	75	GLY	10
1	A	57	LEU	4
1	A	82	ARG	3
1	A	58	HIS	3
1	A	2	ASN	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	67	GLY	2
1	A	64	THR	2
1	A	56	PRO	1
1	A	30	ILE	1
1	A	74	LYS	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	91/92 (99%)	81±2 (89±3%)	10±2 (11±3%)	8	52
All	All	1820/1840 (99%)	1623 (89%)	197 (11%)	8	52

All 58 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	61	TRP	20
1	A	40	THR	15
1	A	79	ILE	11
1	A	44	GLU	9
1	A	82	ARG	6
1	A	14	TRP	6
1	A	11	ASP	6
1	A	30	ILE	6
1	A	54	GLN	5
1	A	89	ASP	5
1	A	27	ASP	5
1	A	53	ARG	5
1	A	43	PHE	4
1	A	49	ILE	4
1	A	105	PHE	4
1	A	62	LEU	4
1	A	87	PHE	4
1	A	8	ARG	3
1	A	18	GLU	3
1	A	64	THR	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	74	LYS	3
1	A	21	LEU	3
1	A	3	ASP	3
1	A	20	ARG	3
1	A	69	TYR	3
1	A	51	ILE	3
1	A	65	LYS	3
1	A	77	GLU	3
1	A	33	GLU	3
1	A	88	TRP	2
1	A	86	THR	2
1	A	35	ASN	2
1	A	45	ASN	2
1	A	59	GLN	2
1	A	71	PHE	2
1	A	85	GLU	2
1	A	97	GLN	2
1	A	42	THR	2
1	A	70	HIS	2
1	A	24	TRP	2
1	A	48	LYS	2
1	A	90	LEU	2
1	A	7	HIS	1
1	A	96	THR	1
1	A	19	GLU	1
1	A	34	ILE	1
1	A	55	GLU	1
1	A	73	LEU	1
1	A	106	ARG	1
1	A	39	LEU	1
1	A	47	SER	1
1	A	76	ASP	1
1	A	6	PHE	1
1	A	50	ILE	1
1	A	58	HIS	1
1	A	66	GLN	1
1	A	60	VAL	1
1	A	80	CYS	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 oligosaccharides 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