Full wwPDB NMR Structure Validation Report (i)

Dec 13, 2023 – 09:36 PM EST

PDB ID	:	2JQA
Title	:	Solution structure of apo-DR1885 from Deinococcus radiodurans
Authors	:	Banci, L.; Bertini, I.; Ciofi Baffoni, S.; Katsari, E.; Katsaros, N.; Kubicek, K.
Deposited on	:	2007-05-30

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:

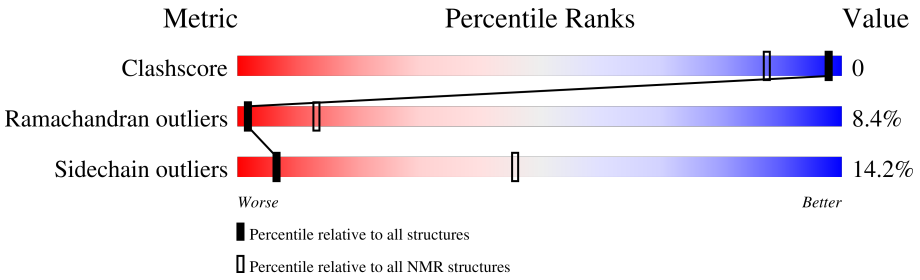
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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	149	70% 11% 19%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:28-A:147 (120)	1.78	15

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 4 clusters and 6 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2215 atoms, of which 1137 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hypothetical protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	149	2215	673	1137	195	200	10	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	cloning artifact	UNP Q9RT80
A	146	ILE	-	cloning artifact	UNP Q9RT80
A	147	GLU	-	cloning artifact	UNP Q9RT80
A	148	GLY	-	cloning artifact	UNP Q9RT80
A	149	ARG	-	cloning artifact	UNP Q9RT80

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: Hypothetical protein

Chain A:

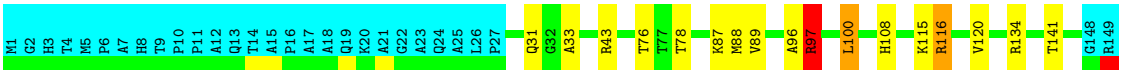
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Hypothetical protein

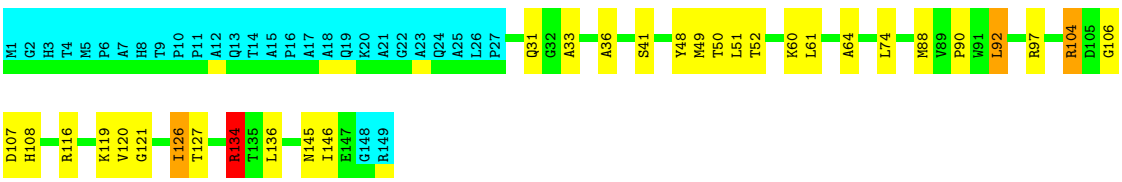
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Hypothetical protein

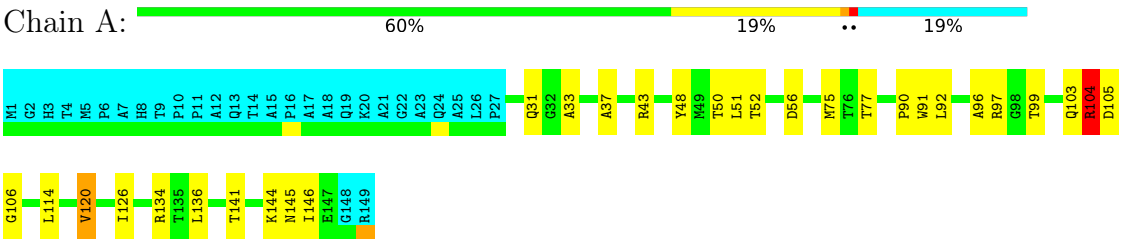
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Hypothetical protein

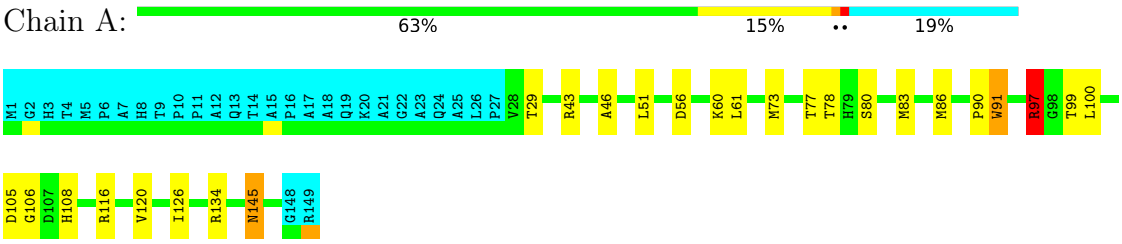
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Hypothetical protein

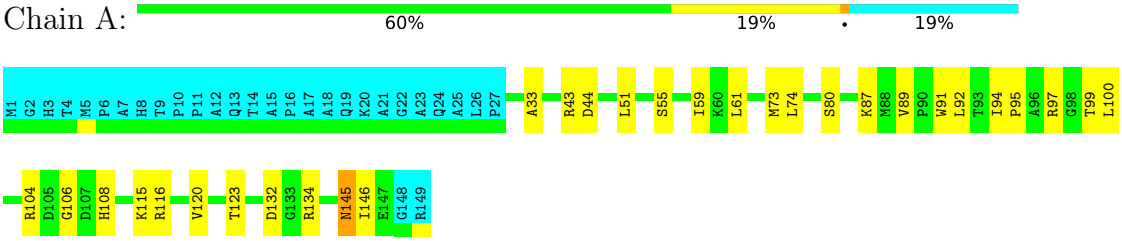
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Hypothetical protein

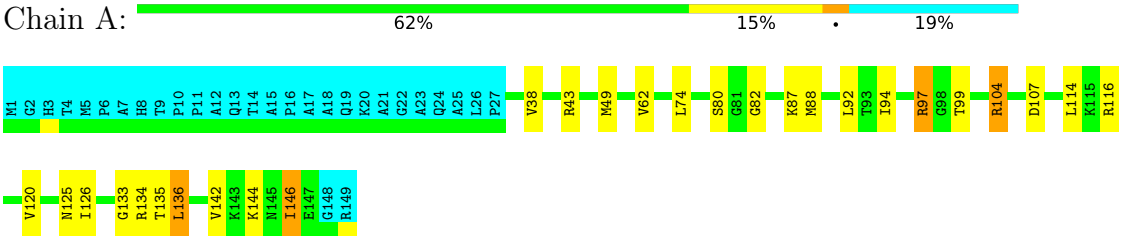
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Hypothetical protein

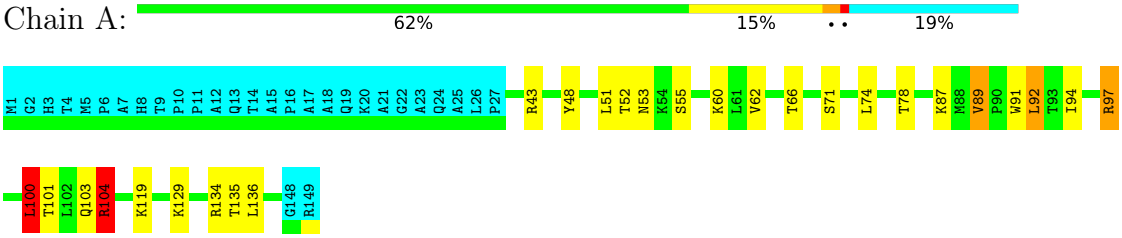
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Hypothetical protein

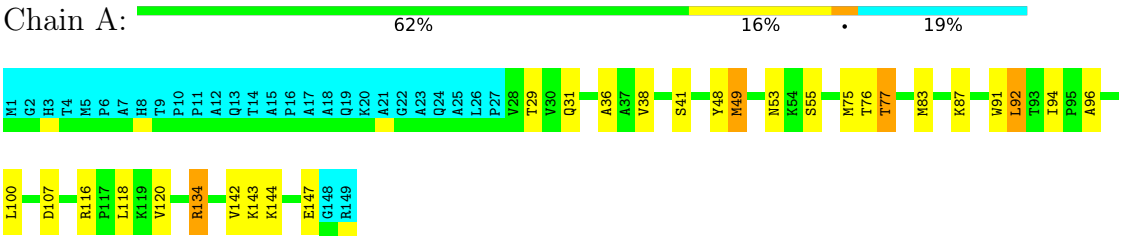
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Hypothetical protein

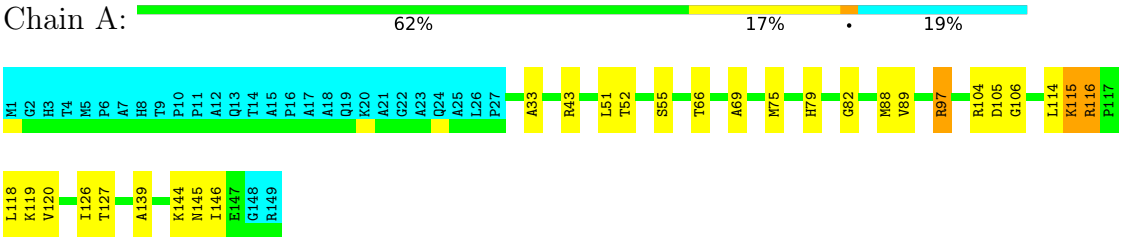
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Hypothetical protein

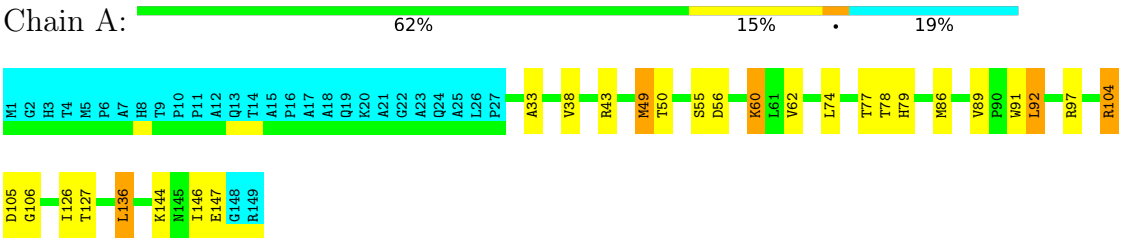
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Hypothetical protein

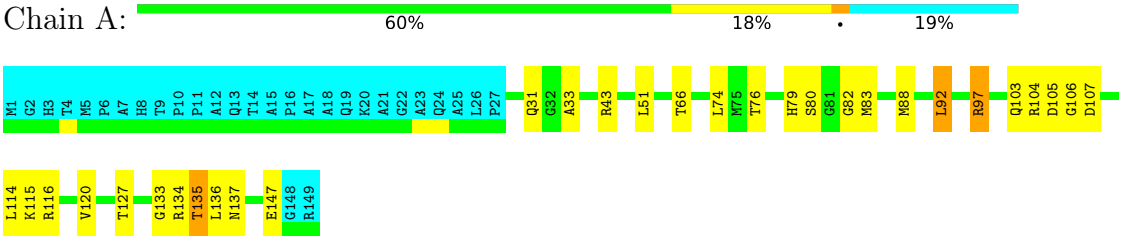
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Hypothetical protein

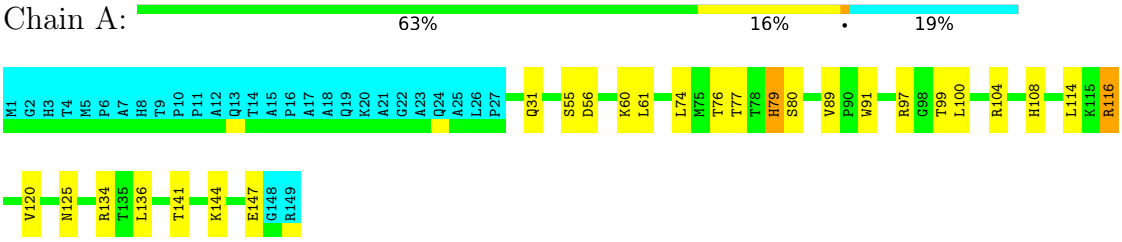
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Hypothetical protein

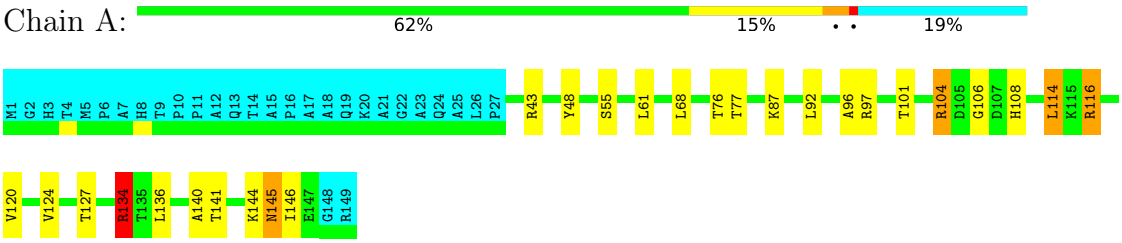
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Hypothetical protein

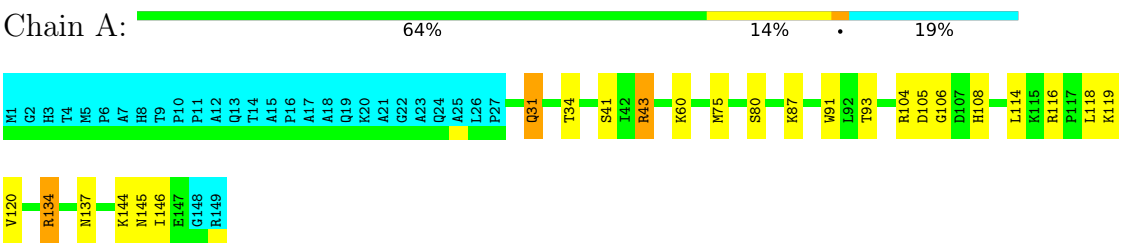
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Hypothetical protein

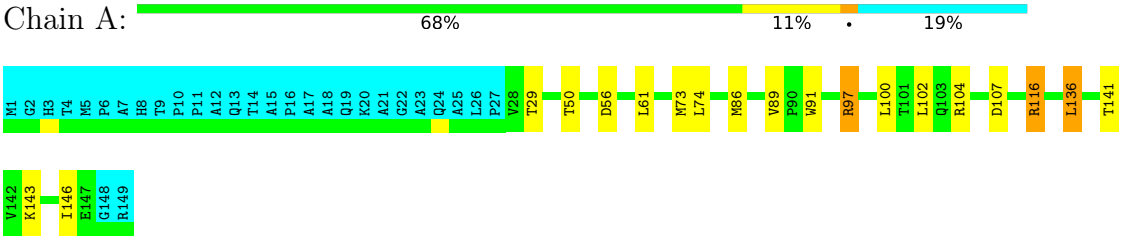
Chain A:



4.2.15 Score per residue for model 15 (medoid)

• Molecule 1: Hypothetical protein

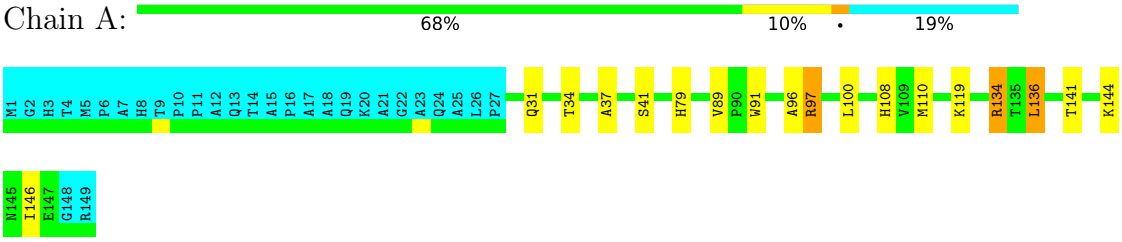
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Hypothetical protein

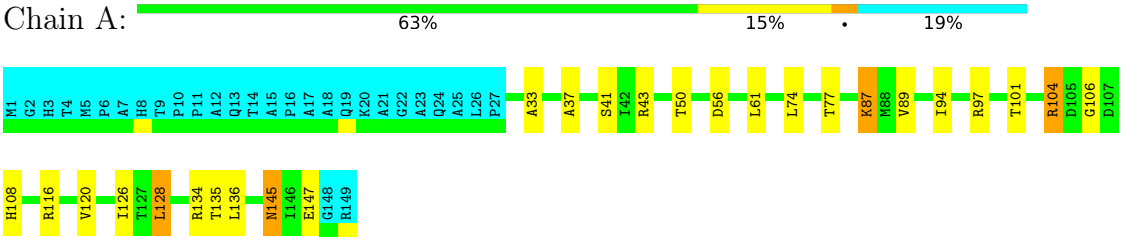
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Hypothetical protein

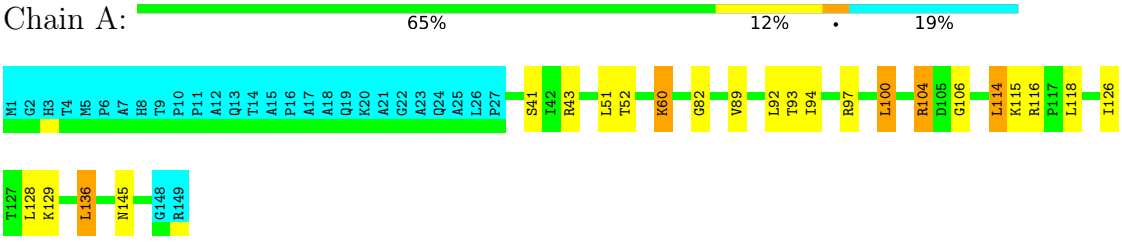
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Hypothetical protein

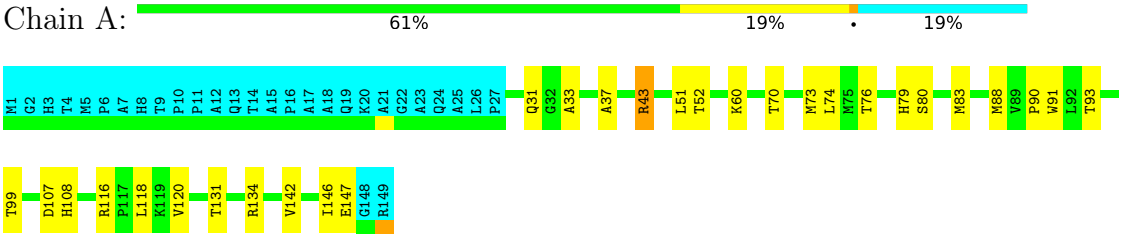
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Hypothetical protein

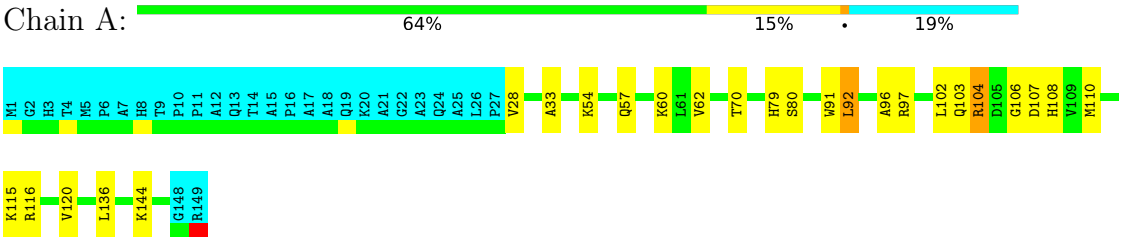
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Hypothetical protein

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 300 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
Amber	refinement	8
DYANA	structure solution	

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	0.62±0.01	0±0/891 (0.0± 0.0%)	1.16±0.04	4±1/1214 (0.3± 0.1%)
All	All	0.62	0/17820 (0.0%)	1.16	79/24280 (0.3%)

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	1.0±1.3
All	All	0	20

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	97	ARG	NE-CZ-NH1	8.60	124.60	120.30	10	13
1	A	116	ARG	NE-CZ-NH1	8.43	124.52	120.30	14	15
1	A	134	ARG	NE-CZ-NH1	7.82	124.21	120.30	17	10
1	A	43	ARG	NE-CZ-NH1	7.76	124.18	120.30	9	12
1	A	134	ARG	NE-CZ-NH2	-7.42	116.59	120.30	8	4
1	A	104	ARG	NE-CZ-NH1	7.40	124.00	120.30	2	11
1	A	43	ARG	NE-CZ-NH2	-6.88	116.86	120.30	11	2
1	A	48	TYR	CB-CG-CD2	-6.61	117.03	121.00	8	1
1	A	116	ARG	NE-CZ-NH2	-6.37	117.12	120.30	6	4
1	A	97	ARG	NE-CZ-NH2	-6.33	117.13	120.30	15	3
1	A	104	ARG	NE-CZ-NH2	-6.22	117.19	120.30	2	2
1	A	99	THR	CA-CB-CG2	5.66	120.32	112.40	4	1
1	A	106	GLY	C-N-CA	5.33	135.03	121.70	20	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	100	LEU	Peptide	2
1	A	104	ARG	Sidechain,Peptide	2
1	A	50	THR	Peptide	1
1	A	61	LEU	Peptide	1
1	A	142	VAL	Peptide	1
1	A	71	SER	Peptide	1
1	A	91	TRP	Peptide	1
1	A	119	LYS	Peptide	1
1	A	129	LYS	Peptide	1
1	A	49	MET	Peptide	1
1	A	143	LYS	Peptide	1
1	A	147	GLU	Peptide	1
1	A	82	GLY	Peptide	1
1	A	115	LYS	Peptide	1
1	A	79	HIS	Peptide	1
1	A	68	LEU	Peptide	1
1	A	116	ARG	Sidechain	1
1	A	140	ALA	Peptide	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	878	937	937	1±1
All	All	17560	18740	18740	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:100:LEU:HD12	1:A:100:LEU:H	0.70	1.46	18	1
1:A:118:LEU:HB3	1:A:142:VAL:HG21	0.55	1.78	8	1
1:A:64:ALA:HB1	1:A:126:ILE:HD11	0.52	1.82	2	1
1:A:94:ILE:HG22	1:A:100:LEU:H	0.49	1.68	7	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:114:LEU:HD23	1:A:115:LYS:H	0.48	1.68	18	1
1:A:92:LEU:HA	1:A:100:LEU:HD21	0.47	1.85	8	1
1:A:91:TRP:CD1	1:A:100:LEU:HD12	0.46	2.46	4	1
1:A:37:ALA:HB2	1:A:142:VAL:CG1	0.46	2.40	19	1
1:A:69:ALA:HB2	1:A:115:LYS:HE3	0.45	1.88	9	1
1:A:61:LEU:HD21	1:A:128:LEU:HB2	0.45	1.87	17	1
1:A:126:ILE:HD13	1:A:127:THR:N	0.44	2.28	2	1
1:A:100:LEU:CD2	1:A:102:LEU:HD11	0.43	2.44	15	1
1:A:36:ALA:HB2	1:A:48:TYR:CE1	0.42	2.50	2	1
1:A:46:ALA:HB1	1:A:108:HIS:CE1	0.40	2.52	4	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	120/149 (81%)	89±4 (74±4%)	21±5 (17±4%)	10±3 (8±2%)	2	13
All	All	2400/2980 (81%)	1781 (74%)	417 (17%)	202 (8%)	2	13

All 50 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	120	VAL	15
1	A	106	GLY	11
1	A	146	ILE	11
1	A	33	ALA	10
1	A	89	VAL	10
1	A	136	LEU	9
1	A	97	ARG	8
1	A	145	ASN	8
1	A	96	ALA	6
1	A	80	SER	6
1	A	55	SER	6
1	A	88	MET	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	31	GLN	5
1	A	92	LEU	5
1	A	104	ARG	5
1	A	115	LYS	4
1	A	90	PRO	4
1	A	107	ASP	4
1	A	56	ASP	4
1	A	91	TRP	4
1	A	99	THR	4
1	A	62	VAL	4
1	A	41	SER	4
1	A	144	LYS	4
1	A	147	GLU	4
1	A	78	THR	3
1	A	37	ALA	3
1	A	60	LYS	3
1	A	82	GLY	3
1	A	79	HIS	3
1	A	134	ARG	2
1	A	114	LEU	2
1	A	133	GLY	2
1	A	87	LYS	2
1	A	77	THR	2
1	A	135	THR	2
1	A	93	THR	2
1	A	76	THR	1
1	A	121	GLY	1
1	A	94	ILE	1
1	A	95	PRO	1
1	A	36	ALA	1
1	A	116	ARG	1
1	A	139	ALA	1
1	A	49	MET	1
1	A	137	ASN	1
1	A	101	THR	1
1	A	124	VAL	1
1	A	119	LYS	1
1	A	28	VAL	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	97/115 (84%)	83±3 (86±3%)	14±3 (14±3%)	6 46
All	All	1940/2300 (84%)	1664 (86%)	276 (14%)	6 46

All 73 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	92	LEU	11
1	A	108	HIS	10
1	A	134	ARG	10
1	A	74	LEU	10
1	A	51	LEU	9
1	A	136	LEU	9
1	A	60	LYS	8
1	A	126	ILE	8
1	A	87	LYS	7
1	A	104	ARG	7
1	A	114	LEU	7
1	A	91	TRP	7
1	A	97	ARG	6
1	A	100	LEU	6
1	A	141	THR	6
1	A	52	THR	6
1	A	77	THR	6
1	A	105	ASP	6
1	A	144	LYS	6
1	A	31	GLN	5
1	A	61	LEU	5
1	A	145	ASN	5
1	A	76	THR	5
1	A	49	MET	4
1	A	50	THR	4
1	A	75	MET	4
1	A	103	GLN	4
1	A	73	MET	4
1	A	83	MET	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	94	ILE	4
1	A	118	LEU	4
1	A	127	THR	4
1	A	79	HIS	4
1	A	43	ARG	4
1	A	116	ARG	3
1	A	119	LYS	3
1	A	48	TYR	3
1	A	29	THR	3
1	A	86	MET	3
1	A	38	VAL	3
1	A	135	THR	3
1	A	66	THR	3
1	A	107	ASP	3
1	A	41	SER	2
1	A	125	ASN	2
1	A	53	ASN	2
1	A	101	THR	2
1	A	56	ASP	2
1	A	80	SER	2
1	A	34	THR	2
1	A	110	MET	2
1	A	128	LEU	2
1	A	70	THR	2
1	A	120	VAL	1
1	A	44	ASP	1
1	A	59	ILE	1
1	A	123	THR	1
1	A	132	ASP	1
1	A	146	ILE	1
1	A	89	VAL	1
1	A	55	SER	1
1	A	78	THR	1
1	A	147	GLU	1
1	A	93	THR	1
1	A	137	ASN	1
1	A	143	LYS	1
1	A	129	LYS	1
1	A	88	MET	1
1	A	99	THR	1
1	A	131	THR	1
1	A	54	LYS	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	57	GLN	1
1	A	102	LEU	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 monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

No chemical shift data were provided