Full wwPDB NMR Structure Validation Report (i)

Mar 6, 2022 – 02:52 PM EST

PDB ID	:	2KKP
Title	:	Solution NMR structure of the phage integrase SAM-like Domain from Moth 1796 from Moorella thermoacetica. Northeast Structural Genomics Consortium Target MtR39K (residues 64-171).
Authors	:	Ramelot, T.A.; Wang, H.; Ciccosanti, C.; Jang, M.; Nair, R.; Rost, B.; Swapna, G.; Acton, T.B.; Xiao, R.; Everett, J.K.; Montelione, G.T.; Kennedy, M.A.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2009-06-26

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 following versions of software and data (see references (i)) were used in the production of this report:

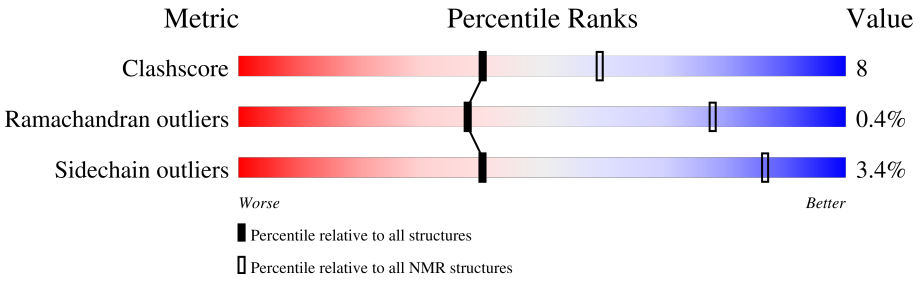
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	117	62% 9% 30%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:8-A:16, A:32-A:104 (82)	0.53	2

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1946 atoms, of which 982 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Phage integrase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	117	1946	614	982	180	168	2	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q2RHJ2
A	110	LEU	-	expression tag	UNP Q2RHJ2
A	111	GLU	-	expression tag	UNP Q2RHJ2
A	112	HIS	-	expression tag	UNP Q2RHJ2
A	113	HIS	-	expression tag	UNP Q2RHJ2
A	114	HIS	-	expression tag	UNP Q2RHJ2
A	115	HIS	-	expression tag	UNP Q2RHJ2
A	116	HIS	-	expression tag	UNP Q2RHJ2
A	117	HIS	-	expression tag	UNP Q2RHJ2

4 Residue-property plots (i)

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: Phage integrase

Chain A: 62% 9% 30%

M1 I2 E3 P4 S5 K6 I7 T8 V9 L17 T18 D19 Y20 A21 K22 P23 H24 L25 R26 Q27 S28 T29 W30 E31 E34 L37 R38 V41 I42 G46 L70 L96 P105 P106 R107 H108 P109 L110 E111 H112 H113 H114 H115 H116 H117

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: Phage integrase

Chain A: 55% 15% 30%

M1 I2 E3 P4 S5 K6 I7 T8 V9 L17 T18 D19 Y20 A21 K22 P23 H24 L25 R26 Q27 S28 T29 W30 E31 E34 L37 R38 V41 G46 S47 I48 P49 K52 K65 L70 V80 M86 A89 L94 L95 L96 P99 A103 K104 P105 P106

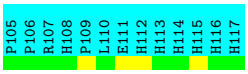
R107 H108 P109 L110 E111 H112 H113 H114 H115 H116 H117

4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: Phage integrase

Chain A: 52% 18% 30%

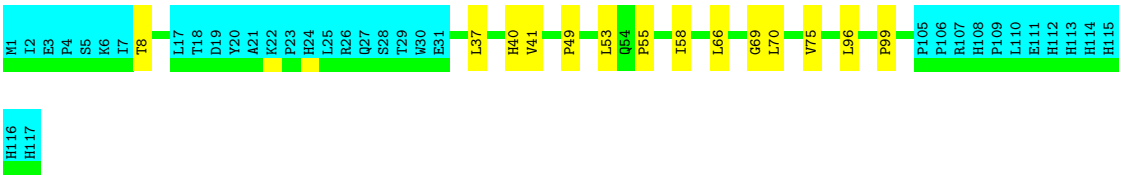
M1 I2 E3 P4 S5 K6 I7 T8 V9 L17 T18 D19 Y20 A21 K22 P23 H24 L25 R26 Q27 S28 T29 W30 E31 E34 L37 R38 V41 I42 L45 I48 K52 L53 Q54 I58 Y62 K65 L70 I78 H79 V80 V81 L96 T100 E101



4.2.3 Score per residue for model 3

• Molecule 1: Phage integrase

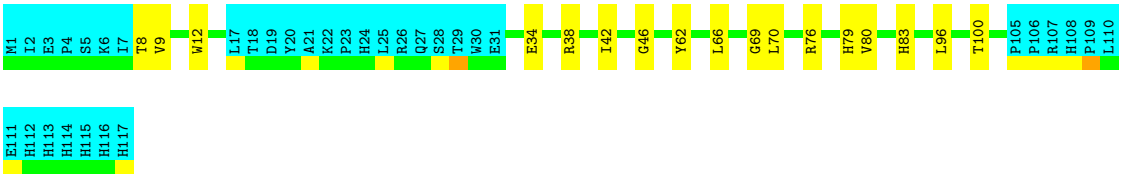
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Phage integrase

Chain A:



4.2.5 Score per residue for model 5

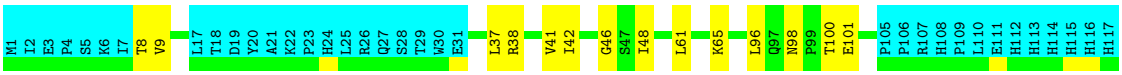
• Molecule 1: Phage integrase

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Phage integrase

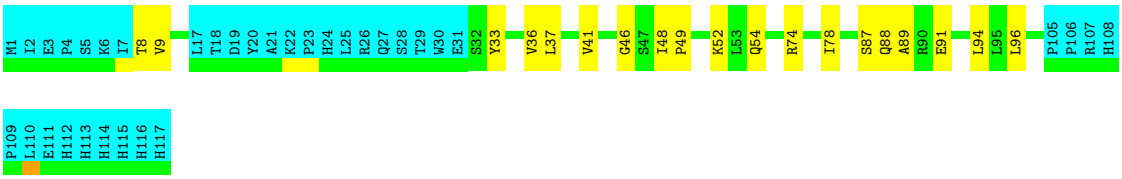
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Phage integrase

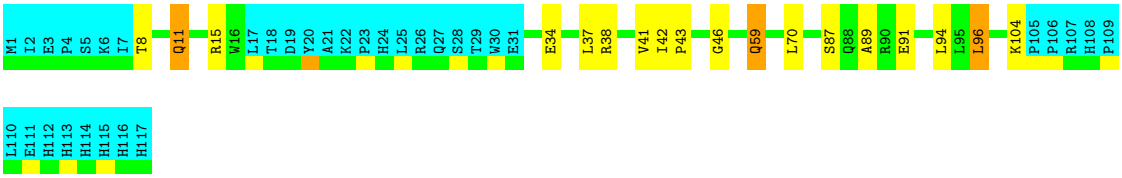
Chain A: 54% 16% 30%



4.2.8 Score per residue for model 8

- Molecule 1: Phage integrase

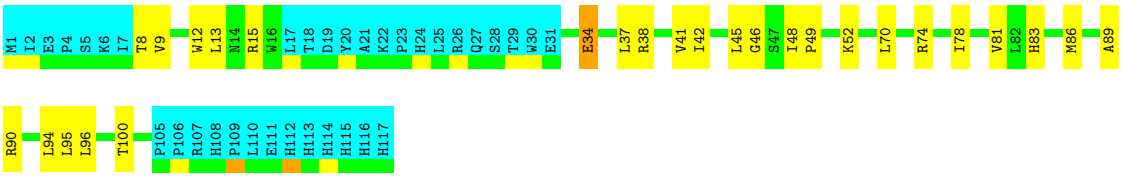
Chain A: 55% 13% • 30%



4.2.9 Score per residue for model 9

- Molecule 1: Phage integrase

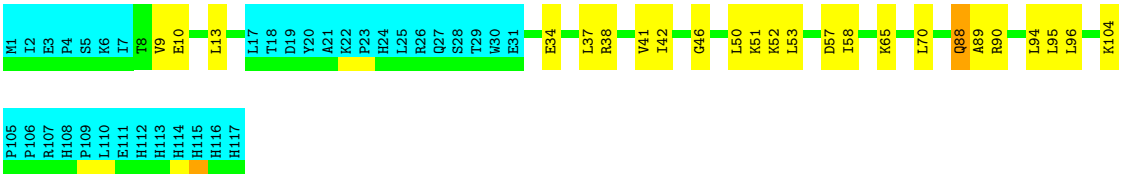
Chain A: 47% 22% • 30%



4.2.10 Score per residue for model 10

- Molecule 1: Phage integrase

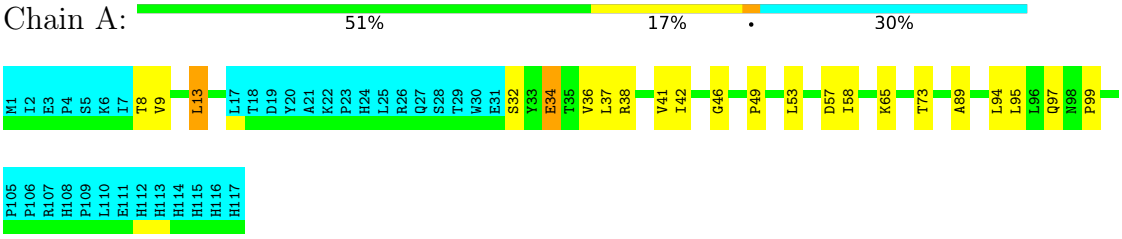
Chain A: 50% 20% • 30%



4.2.11 Score per residue for model 11

- Molecule 1: Phage integrase

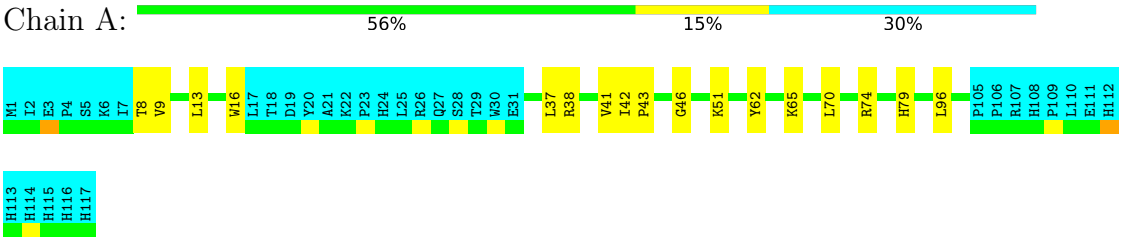
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Phage integrase

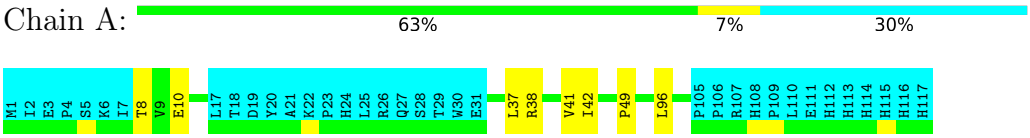
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Phage integrase

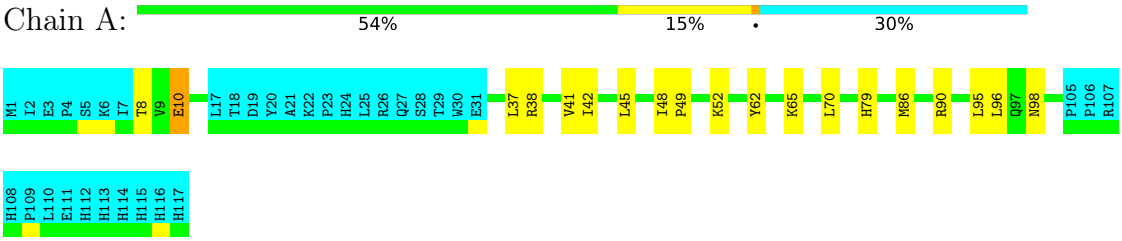
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Phage integrase

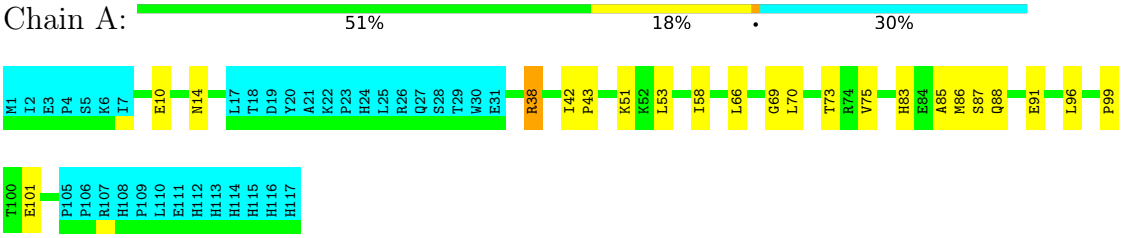
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Phage integrase

Chain A:



4.2.16 Score per residue for model 16

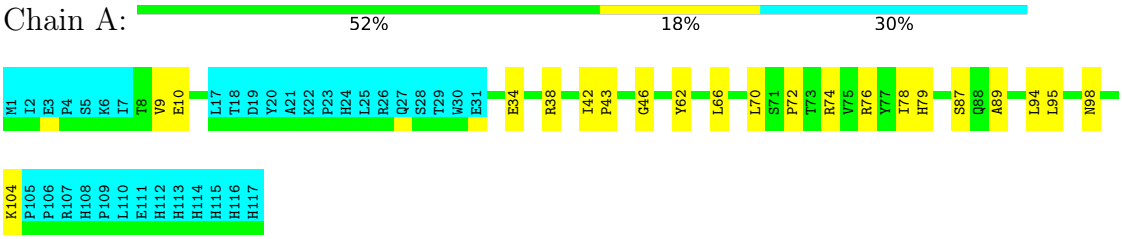
- Molecule 1: Phage integrase

Chain A:

4.2.17 Score per residue for model 17

- Molecule 1: Phage integrase

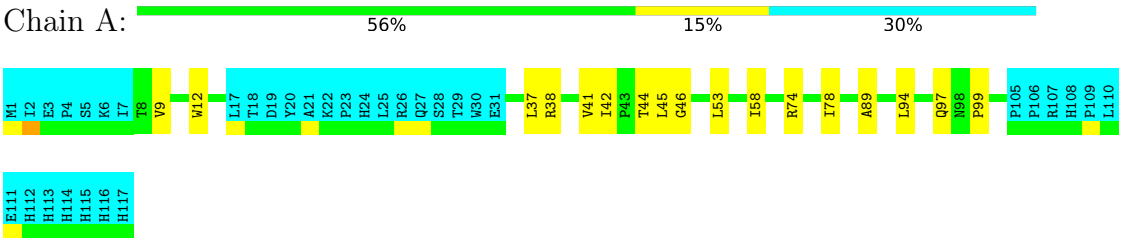
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Phage integrase

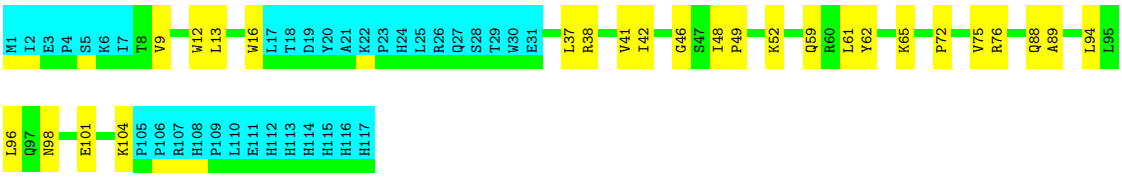
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Phage integrase

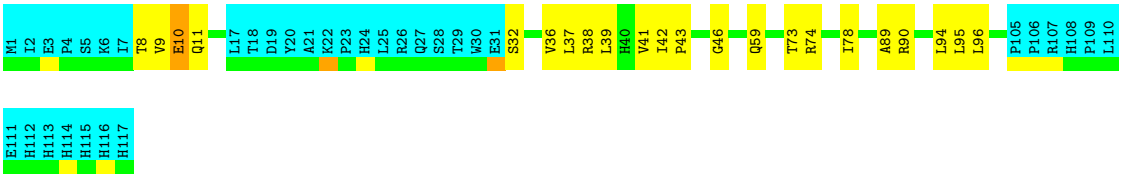
Chain A: 48% 22% 30%



4.2.20 Score per residue for model 20

- Molecule 1: Phage integrase

Chain A: 51% 18% • 30%



5 Refinement protocol and experimental data overview (i)

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

Of the 150 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
AutoStructure	structure solution	2.2.1
X-PLOR NIH	structure solution	
CNS	refinement	1.2
PSVS	structure solution	
PdbStat	structure solution	5.1
CYANA	structure solution	

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	659	687	684	10±3
All	All	13180	13740	13680	203

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 8.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:9:VAL:HB	1:A:46:GLY:HA2	0.80	1.49	11	10
1:A:89:ALA:HB1	1:A:94:LEU:HD11	0.76	1.57	9	9
1:A:88:GLN:HE21	1:A:88:GLN:HA	0.75	1.41	10	1
1:A:65:LYS:HB3	1:A:70:LEU:HD12	0.69	1.62	1	4
1:A:11:GLN:HE21	1:A:11:GLN:HA	0.69	1.47	16	1
1:A:53:LEU:HD11	1:A:58:ILE:HD11	0.69	1.65	16	7
1:A:9:VAL:HA	1:A:12:TRP:NE1	0.65	2.06	18	3
1:A:8:THR:HB	1:A:10:GLU:OE2	0.63	1.94	14	2
1:A:96:LEU:H	1:A:96:LEU:HD23	0.61	1.56	1	14
1:A:8:THR:HB	1:A:46:GLY:O	0.59	1.97	7	5
1:A:90:ARG:HA	1:A:95:LEU:HB3	0.59	1.74	9	2
1:A:38:ARG:O	1:A:42:ILE:HB	0.58	1.98	9	10
1:A:13:LEU:HD12	1:A:42:ILE:HD11	0.58	1.75	10	2
1:A:34:GLU:O	1:A:38:ARG:HG2	0.57	1.99	2	5

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:45:LEU:HG	1:A:48:ILE:HD12	0.57	1.76	9	3
1:A:44:THR:HG23	1:A:45:LEU:HG	0.57	1.76	18	1
1:A:66:LEU:HA	1:A:70:LEU:HB2	0.56	1.76	3	4
1:A:38:ARG:O	1:A:43:PRO:HD3	0.56	2.00	17	4
1:A:8:THR:HA	1:A:49:PRO:HA	0.55	1.78	11	3
1:A:13:LEU:HD13	1:A:37:LEU:HD21	0.55	1.79	12	1
1:A:33:TYR:HA	1:A:36:VAL:HG22	0.55	1.77	7	2
1:A:49:PRO:HD2	1:A:52:LYS:HD3	0.55	1.78	14	2
1:A:49:PRO:HG2	1:A:52:LYS:HB3	0.54	1.79	9	2
1:A:72:PRO:O	1:A:76:ARG:HG3	0.54	2.02	19	2
1:A:87:SER:O	1:A:91:GLU:HG2	0.54	2.02	8	3
1:A:38:ARG:HB3	1:A:42:ILE:HD12	0.54	1.78	17	1
1:A:61:LEU:O	1:A:65:LYS:HG2	0.53	2.03	6	1
1:A:9:VAL:HG23	1:A:48:ILE:O	0.52	2.03	1	5
1:A:86:MET:SD	1:A:99:PRO:HD2	0.52	2.45	1	2
1:A:76:ARG:O	1:A:80:VAL:HG23	0.51	2.05	4	1
1:A:62:TYR:HA	1:A:65:LYS:HE2	0.51	1.83	19	1
1:A:38:ARG:HB2	1:A:42:ILE:HD12	0.51	1.81	6	3
1:A:37:LEU:O	1:A:41:VAL:HG12	0.51	2.06	11	13
1:A:42:ILE:N	1:A:43:PRO:HD2	0.51	2.21	20	3
1:A:90:ARG:HA	1:A:95:LEU:O	0.50	2.06	14	2
1:A:49:PRO:HG2	1:A:52:LYS:CB	0.50	2.36	19	1
1:A:38:ARG:HG3	1:A:39:LEU:N	0.50	2.20	20	1
1:A:8:THR:HB	1:A:11:GLN:HB2	0.50	1.82	8	1
1:A:97:GLN:O	1:A:99:PRO:HD3	0.50	2.06	11	2
1:A:11:GLN:O	1:A:15:ARG:HG2	0.50	2.07	8	1
1:A:62:TYR:OH	1:A:79:HIS:HA	0.49	2.08	14	6
1:A:38:ARG:CB	1:A:42:ILE:HD12	0.49	2.37	17	1
1:A:12:TRP:HE3	1:A:13:LEU:HD22	0.48	1.68	19	1
1:A:37:LEU:O	1:A:41:VAL:HG22	0.48	2.07	10	4
1:A:83:HIS:CD2	1:A:101:GLU:HB2	0.48	2.43	15	1
1:A:74:ARG:O	1:A:78:ILE:HG13	0.48	2.09	17	3
1:A:16:TRP:HE1	1:A:88:GLN:HG2	0.47	1.69	19	1
1:A:55:PRO:HG3	1:A:99:PRO:HA	0.47	1.85	3	1
1:A:82:LEU:HG	1:A:100:THR:HG21	0.47	1.86	16	1
1:A:8:THR:HG22	1:A:9:VAL:H	0.47	1.70	20	1
1:A:66:LEU:HD13	1:A:75:VAL:HG21	0.47	1.86	3	2
1:A:83:HIS:HA	1:A:100:THR:HB	0.46	1.86	4	1
1:A:38:ARG:HA	1:A:42:ILE:HB	0.46	1.88	8	3
1:A:58:ILE:HG22	1:A:62:TYR:CE2	0.46	2.45	2	1
1:A:87:SER:HA	1:A:98:ASN:OD1	0.46	2.11	17	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:38:ARG:HA	1:A:42:ILE:HD13	0.45	1.86	9	4
1:A:78:ILE:HA	1:A:81:VAL:HG12	0.45	1.88	2	2
1:A:50:LEU:HB3	1:A:95:LEU:HD13	0.45	1.87	10	1
1:A:42:ILE:HA	1:A:46:GLY:HA3	0.45	1.89	8	1
1:A:53:LEU:CD1	1:A:58:ILE:HD11	0.45	2.42	2	2
1:A:32:SER:O	1:A:36:VAL:HG22	0.45	2.11	20	2
1:A:8:THR:HB	1:A:11:GLN:CG	0.45	2.42	16	1
1:A:65:LYS:HG2	1:A:70:LEU:HD12	0.44	1.88	2	1
1:A:8:THR:HB	1:A:11:GLN:HG2	0.44	1.89	16	1
1:A:34:GLU:H	1:A:34:GLU:CD	0.44	2.15	11	1
1:A:52:LYS:HG2	1:A:52:LYS:O	0.44	2.13	10	1
1:A:38:ARG:HA	1:A:42:ILE:HG12	0.44	1.90	10	1
1:A:74:ARG:O	1:A:78:ILE:HG12	0.43	2.13	9	2
1:A:88:GLN:HE21	1:A:88:GLN:CA	0.43	2.21	10	1
1:A:98:ASN:OD1	1:A:101:GLU:HG2	0.43	2.13	19	1
1:A:40:HIS:CE1	1:A:41:VAL:HB	0.42	2.49	3	1
1:A:98:ASN:ND2	1:A:101:GLU:HB2	0.42	2.29	6	1
1:A:61:LEU:O	1:A:65:LYS:HG3	0.42	2.15	19	1
1:A:13:LEU:HG	1:A:42:ILE:HD11	0.42	1.90	11	1
1:A:15:ARG:HA	1:A:15:ARG:NE	0.42	2.30	9	1
1:A:52:LYS:O	1:A:54:GLN:HG3	0.41	2.14	2	2
1:A:89:ALA:O	1:A:94:LEU:HG	0.41	2.16	11	2
1:A:85:ALA:O	1:A:88:GLN:HG2	0.41	2.16	15	1
1:A:59:GLN:HG2	1:A:104:LYS:HB3	0.41	1.91	8	1
1:A:38:ARG:HA	1:A:42:ILE:CG1	0.40	2.46	10	1
1:A:34:GLU:OE2	1:A:38:ARG:HD2	0.40	2.17	8	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	82/117 (70%)	79±1 (97±1%)	2±1 (3±1%)	0±0 (0±1%)	38	78
All	All	1640/2340 (70%)	1584 (97%)	50 (3%)	6 (0%)	38	78

All 2 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	69	GLY	4
1	A	96	LEU	2

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	73/107 (68%)	71±2 (97±3%)	2±2 (3±3%)	40 87
All	All	1460/2140 (68%)	1411 (97%)	49 (3%)	40 87

All 22 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	10	GLU	6
1	A	34	GLU	5
1	A	100	THR	3
1	A	11	GLN	3
1	A	59	GLN	3
1	A	51	LYS	3
1	A	104	LYS	3
1	A	73	THR	3
1	A	88	GLN	2
1	A	70	LEU	2
1	A	86	MET	2
1	A	57	ASP	2
1	A	13	LEU	2
1	A	95	LEU	2
1	A	101	GLU	1
1	A	96	LEU	1
1	A	83	HIS	1
1	A	65	LYS	1
1	A	74	ARG	1
1	A	98	ASN	1
1	A	14	ASN	1
1	A	38	ARG	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 (i)

No chemical shift data were provided