Full wwPDB NMR Structure Validation Report (i)

Feb 19, 2022 – 10:03 PM EST

PDB ID	:	1UF0
Title	:	Solution structure of the N-terminal DCX domain of human doublecortin-like kinase
Authors	:	Saito, K.; Kigawa, T.; Koshiba, S.; Inoue, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2003-05-22

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:

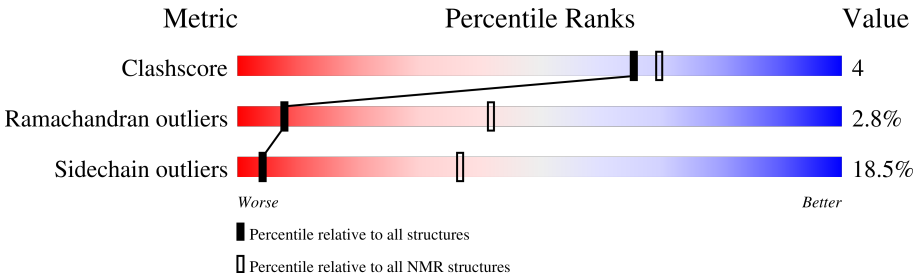
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	116	69% 14% 17%

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:10-A:105 (96)	0.37	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 5 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1818 atoms, of which 913 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Serine/threonine-protein kinase DCAMKL1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	116	1818	571	913	158	175	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP O15075
A	2	SER	-	cloning artifact	UNP O15075
A	3	SER	-	cloning artifact	UNP O15075
A	4	GLY	-	cloning artifact	UNP O15075
A	5	SER	-	cloning artifact	UNP O15075
A	6	SER	-	cloning artifact	UNP O15075
A	7	GLY	-	cloning artifact	UNP O15075
A	111	SER	-	cloning artifact	UNP O15075
A	112	GLY	-	cloning artifact	UNP O15075
A	113	PRO	-	cloning artifact	UNP O15075
A	114	SER	-	cloning artifact	UNP O15075
A	115	SER	-	cloning artifact	UNP O15075
A	116	GLY	-	cloning artifact	UNP O15075

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: Serine/threonine-protein kinase DCAMKL1

Chain A: 69% 14% 17%

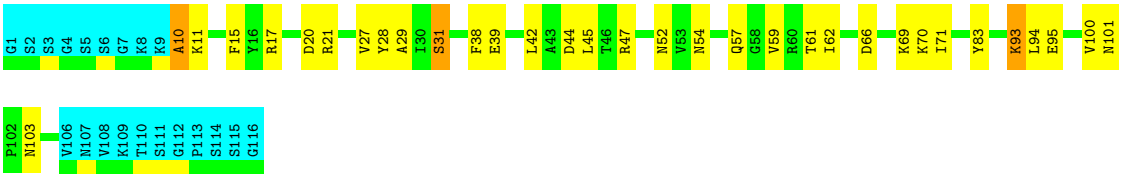
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Serine/threonine-protein kinase DCAMKL1

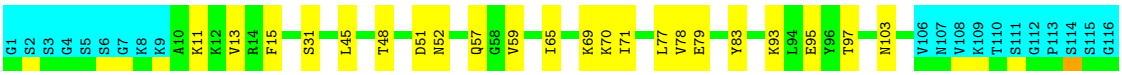
Chain A: 54% 26% • 17%



4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: Serine/threonine-protein kinase DCAMKL1

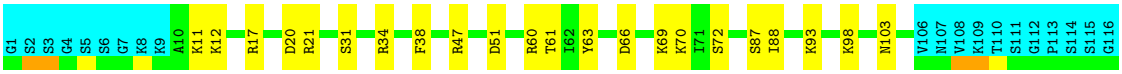
Chain A: 64% 19% 17%



4.2.3 Score per residue for model 3

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

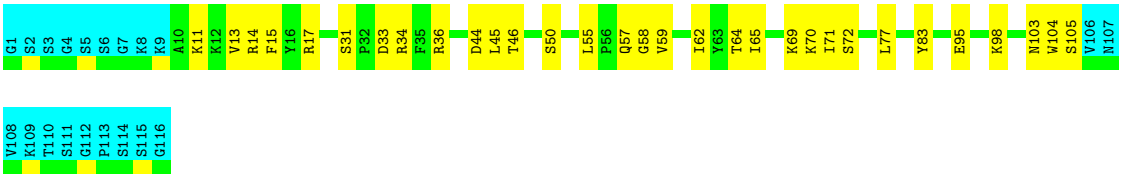
Chain A: 64% 19% 17%



4.2.4 Score per residue for model 4

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

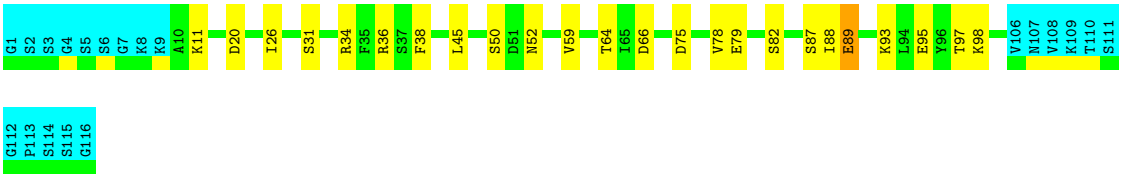
Chain A: 56% 27% 17%



4.2.5 Score per residue for model 5

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

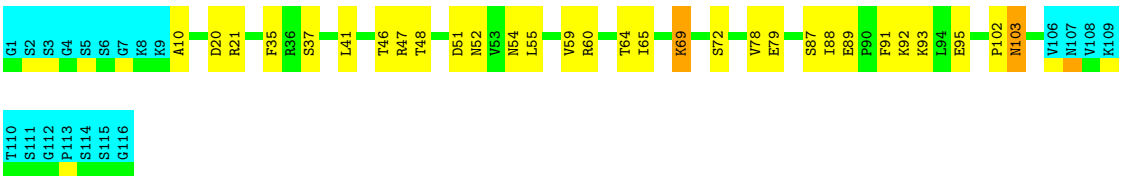
Chain A: 62% 20% • 17%



4.2.6 Score per residue for model 6

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

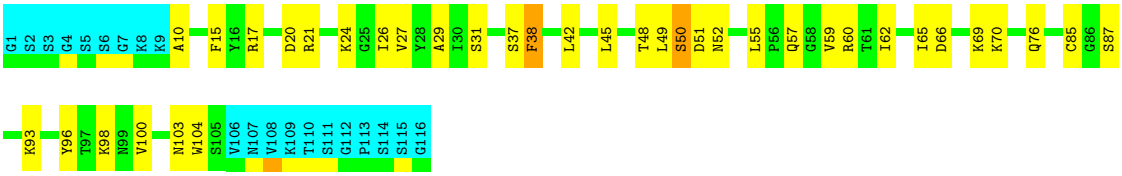
Chain A: 57% 24% • 17%



4.2.7 Score per residue for model 7

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

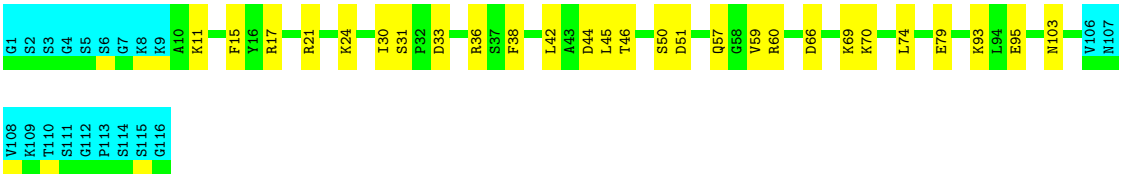
Chain A: 51% 30% • 17%



4.2.8 Score per residue for model 8

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

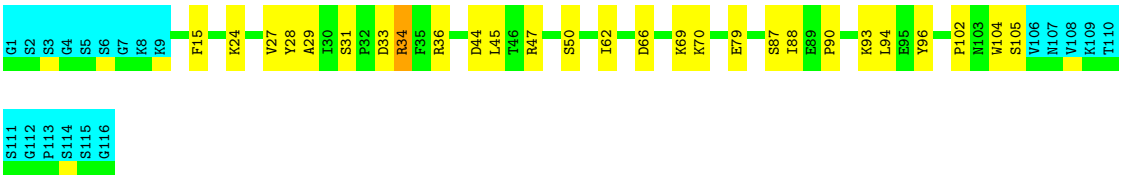
Chain A: 59% 23% 17%



4.2.9 Score per residue for model 9

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

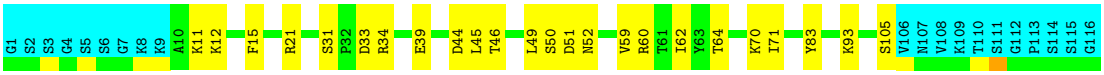
Chain A: 59% 22% • 17%



4.2.10 Score per residue for model 10

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 62% 21% 17%



4.2.11 Score per residue for model 11

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 59% 21% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 K9 A10 K11 I26 A29 I30 S31 R34 E39 T46 L49 S50 D51 N52 V59 R60 T61 T64 K69 K70 I71 V78 E79 Y83 E89 E95 Y96 T97 V100 N101 P102 N103 V106 N107 V108 K109 T110

S111 G112 P113 S114 S115 G116

4.2.12 Score per residue for model 12

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 54% 27% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 K9 A10 K11 R14 F15 Y16 R17 K24 G25 I26 V27 S31 R34 F35 R36 S37 F38 E39 L42 L45 T46 R47 T48 L49 N52 L55 P56 Q57 R60 T64 I65 D66 K69 K70 I71 S72 S87 I88 K93

T97 V100 N103 V106 N107 V108 K109 T110 S111 G112 P113 S114 S115 G116

4.2.13 Score per residue for model 13

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 59% 22% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 K9 A10 K11 F15 G25 I26 S31 S37 D44 L45 T46 R47 N52 Q57 G58 V59 R60 T61 T64 I71 S72 S73 L74 V78 E79 Y83 S87 I88 E89 P90 K93 V100 V106 N107 V108 K109 T110

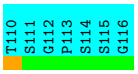
S111 G112 P113 S114 S115 G116

4.2.14 Score per residue for model 14

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 60% 22% • 17%

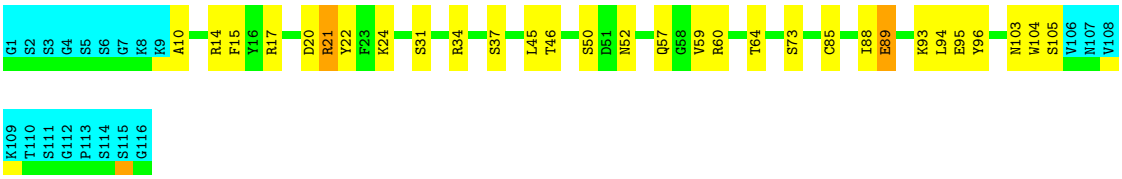
G1 S2 S3 G4 S5 S6 G7 K8 K9 A10 K11 F15 Y16 R17 V27 Y28 S31 P32 D33 S37 L41 R47 T48 D51 N52 R60 T61 I62 I65 K70 L74 D75 E89 P90 K93 K98 N99 V100 N101 P102 N103 V106 N107 V108 K109



4.2.15 Score per residue for model 15

• Molecule 1: Serine/threonine-protein kinase DCAMKL1

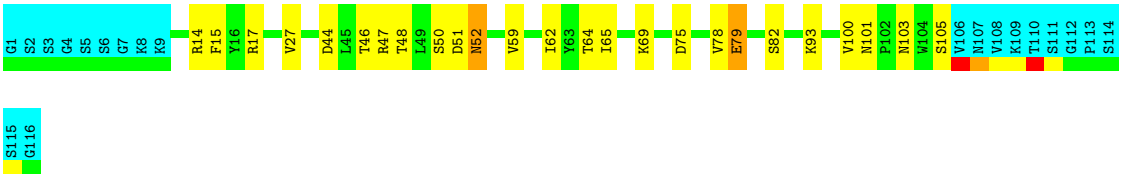
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Serine/threonine-protein kinase DCAMKL1

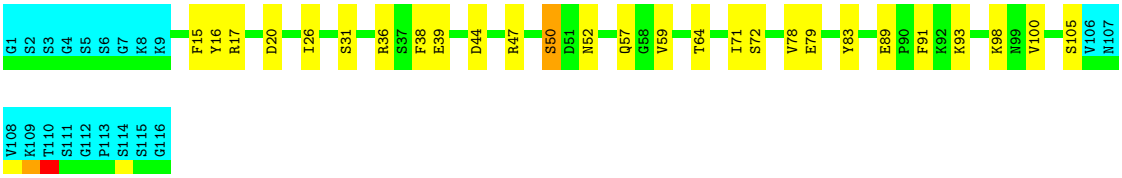
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Serine/threonine-protein kinase DCAMKL1

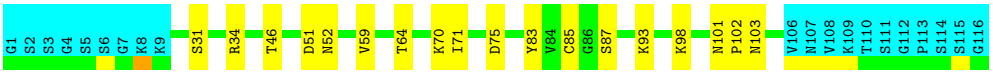
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A:



4.2.19 Score per residue for model 19

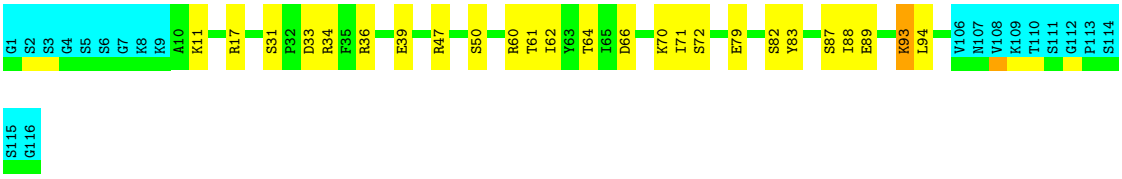
- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 57% 24% • 17%

4.2.20 Score per residue for model 20

- Molecule 1: Serine/threonine-protein kinase DCAMKL1

Chain A: 61% 21% • 17%



5 Refinement protocol and experimental data overview (i)

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CNS	structure solution	1.1
CNS	refinement	1.1

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	779	784	784	6±2
All	All	15580	15680	15680	128

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 4.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:71:ILE:HD11	1:A:83:TYR:CE2	0.81	2.11	13	9
1:A:71:ILE:HD11	1:A:83:TYR:CE1	0.70	2.21	2	1
1:A:45:LEU:CB	1:A:59:VAL:HG11	0.66	2.19	19	5
1:A:87:SER:O	1:A:88:ILE:HD13	0.66	1.90	3	1
1:A:88:ILE:HD12	1:A:89:GLU:N	0.65	2.07	20	4
1:A:15:PHE:CD2	1:A:45:LEU:HD21	0.65	2.27	2	9
1:A:45:LEU:HD22	1:A:49:LEU:HD12	0.64	1.68	12	3
1:A:10:ALA:HB2	1:A:31:SER:OG	0.62	1.95	1	3
1:A:88:ILE:HD13	1:A:88:ILE:N	0.61	2.11	13	1
1:A:45:LEU:HD22	1:A:85:CYS:SG	0.61	2.36	15	1
1:A:46:THR:OG1	1:A:59:VAL:HG12	0.60	1.97	15	8
1:A:45:LEU:HB2	1:A:59:VAL:HG11	0.59	1.73	19	4
1:A:64:THR:HG23	1:A:69:LYS:O	0.59	1.98	6	2
1:A:96:TYR:O	1:A:100:VAL:HG23	0.59	1.96	7	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:13:VAL:HG11	1:A:77:LEU:HD13	0.57	1.76	4	1
1:A:15:PHE:CE1	1:A:62:ILE:HG21	0.57	2.34	7	7
1:A:93:LYS:O	1:A:94:LEU:HD23	0.55	2.02	20	2
1:A:38:PHE:CD2	1:A:42:LEU:HD11	0.54	2.38	12	3
1:A:14:ARG:HG3	1:A:27:VAL:HG22	0.53	1.80	16	1
1:A:87:SER:HB3	1:A:88:ILE:HD13	0.53	1.79	13	1
1:A:10:ALA:HB3	1:A:29:ALA:HB1	0.53	1.81	11	1
1:A:45:LEU:CD2	1:A:49:LEU:HD12	0.52	2.34	10	1
1:A:15:PHE:O	1:A:26:ILE:HG22	0.52	2.05	13	2
1:A:88:ILE:O	1:A:88:ILE:HG22	0.51	2.05	12	2
1:A:62:ILE:HG22	1:A:83:TYR:HB3	0.50	1.82	20	2
1:A:38:PHE:CE2	1:A:42:LEU:HD21	0.50	2.41	1	1
1:A:88:ILE:HD13	1:A:88:ILE:H	0.49	1.66	13	1
1:A:59:VAL:HG21	1:A:85:CYS:SG	0.49	2.47	18	2
1:A:94:LEU:HD12	1:A:96:TYR:OH	0.49	2.08	9	1
1:A:46:THR:HG23	1:A:58:GLY:HA3	0.49	1.84	4	1
1:A:26:ILE:CG2	1:A:49:LEU:HD11	0.49	2.37	11	1
1:A:10:ALA:HB1	1:A:29:ALA:HB1	0.48	1.83	1	1
1:A:35:PHE:CE1	1:A:41:LEU:HD12	0.48	2.44	6	1
1:A:26:ILE:O	1:A:26:ILE:HG23	0.48	2.08	13	3
1:A:10:ALA:CB	1:A:29:ALA:HB1	0.47	2.39	7	1
1:A:59:VAL:HG13	1:A:59:VAL:O	0.47	2.09	17	3
1:A:16:TYR:OH	1:A:100:VAL:HG21	0.46	2.10	17	1
1:A:14:ARG:NH1	1:A:104:TRP:CD1	0.45	2.85	15	1
1:A:91:PHE:CD1	1:A:92:LYS:N	0.45	2.85	6	1
1:A:45:LEU:HB3	1:A:59:VAL:HG11	0.45	1.87	19	1
1:A:10:ALA:HB1	1:A:30:ILE:C	0.44	2.33	11	1
1:A:88:ILE:N	1:A:88:ILE:CD1	0.44	2.81	13	1
1:A:41:LEU:HD22	1:A:74:LEU:HD23	0.44	1.89	14	1
1:A:13:VAL:HG11	1:A:77:LEU:HB3	0.44	1.90	2	1
1:A:21:ARG:NH1	1:A:22:TYR:CE1	0.43	2.86	15	1
1:A:71:ILE:CD1	1:A:83:TYR:CE1	0.43	3.00	2	1
1:A:83:TYR:CD1	1:A:83:TYR:N	0.43	2.86	4	1
1:A:15:PHE:HE1	1:A:62:ILE:HG21	0.43	1.70	7	1
1:A:55:LEU:HD22	1:A:87:SER:CB	0.43	2.44	6	1
1:A:71:ILE:HD11	1:A:83:TYR:HE2	0.43	1.74	4	2
1:A:29:ALA:HB3	1:A:34:ARG:HE	0.42	1.74	9	1
1:A:78:VAL:HG12	1:A:79:GLU:N	0.42	2.29	11	7
1:A:30:ILE:HG23	1:A:74:LEU:HD22	0.42	1.91	8	1
1:A:61:THR:HG21	1:A:63:TYR:CZ	0.42	2.50	3	1
1:A:55:LEU:HD23	1:A:87:SER:HB2	0.41	1.92	7	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:27:VAL:HG12	1:A:28:TYR:N	0.41	2.30	14	3
1:A:64:THR:HG22	1:A:83:TYR:CE2	0.41	2.51	19	1
1:A:14:ARG:NH1	1:A:104:TRP:CG	0.41	2.89	4	1
1:A:61:THR:CG2	1:A:63:TYR:CZ	0.40	3.04	3	1
1:A:64:THR:O	1:A:91:PHE:CE1	0.40	2.75	17	1
1:A:27:VAL:HG13	1:A:104:TRP:CE3	0.40	2.51	7	1
1:A:14:ARG:NH1	1:A:27:VAL:HG22	0.40	2.30	12	1
1:A:37:SER:C	1:A:74:LEU:HD11	0.40	2.37	13	1
1:A:94:LEU:O	1:A:96:TYR:CD2	0.40	2.74	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	96/116 (83%)	79±4 (83±4%)	14±3 (15±3%)	3±1 (3±2%)	8	42
All	All	1920/2320 (83%)	1588 (83%)	279 (15%)	53 (3%)	8	42

All 12 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	52	ASN	12
1	A	103	ASN	7
1	A	100	VAL	6
1	A	50	SER	5
1	A	102	PRO	5
1	A	10	ALA	4
1	A	101	ASN	4
1	A	90	PRO	3
1	A	25	GLY	3
1	A	87	SER	2
1	A	104	TRP	1
1	A	51	ASP	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	86/101 (85%)	70±3 (81±4%)	16±3 (19±4%)	4 37
All	All	1720/2020 (85%)	1401 (81%)	319 (19%)	4 37

All 45 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	31	SER	18
1	A	93	LYS	18
1	A	11	LYS	12
1	A	70	LYS	12
1	A	17	ARG	11
1	A	69	LYS	11
1	A	47	ARG	10
1	A	57	GLN	10
1	A	103	ASN	10
1	A	34	ARG	10
1	A	60	ARG	10
1	A	64	THR	10
1	A	44	ASP	9
1	A	66	ASP	9
1	A	95	GLU	9
1	A	51	ASP	9
1	A	65	ILE	8
1	A	98	LYS	8
1	A	36	ARG	8
1	A	50	SER	8
1	A	20	ASP	7
1	A	21	ARG	7
1	A	72	SER	7
1	A	33	ASP	7
1	A	89	GLU	7
1	A	39	GLU	6
1	A	52	ASN	6
1	A	48	THR	6
1	A	105	SER	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	97	THR	5
1	A	37	SER	5
1	A	24	LYS	5
1	A	61	THR	4
1	A	38	PHE	4
1	A	75	ASP	4
1	A	79	GLU	4
1	A	101	ASN	4
1	A	82	SER	3
1	A	87	SER	3
1	A	54	ASN	2
1	A	12	LYS	2
1	A	55	LEU	2
1	A	76	GLN	1
1	A	88	ILE	1
1	A	73	SER	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