Full wwPDB NMR Structure Validation Report (i)

Feb 28, 2022 – 02:07 PM EST

PDB ID	:	2EE1
Title	:	Solution structures of the Chromo domain of human chromodomain helicase-DNA-binding protein 4
Authors	:	Sato, M.; Tochio, N.; Koshiba, S.; Watanabe, S.; Harada, T.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2007-02-15

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:

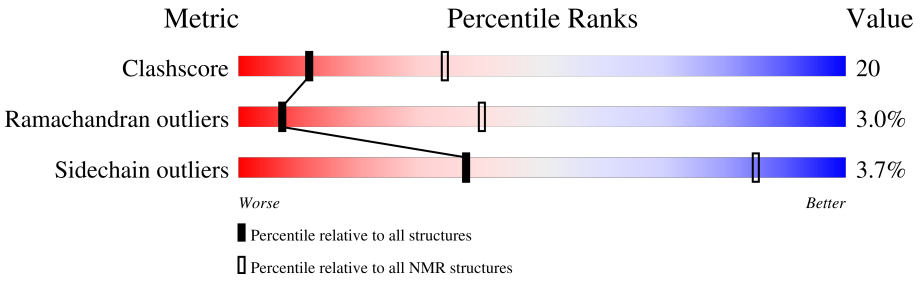
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	64	55% 27% • 17%

2 Ensemble composition and analysis (i)

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: *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:12-A:64 (53)	0.17	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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1047 atoms, of which 504 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Chromodomain helicase-DNA-binding protein 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	64	1047	346	504	94	101	2	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q14839
A	2	SER	-	expression tag	UNP Q14839
A	3	SER	-	expression tag	UNP Q14839
A	4	GLY	-	expression tag	UNP Q14839
A	5	SER	-	expression tag	UNP Q14839
A	6	SER	-	expression tag	UNP Q14839
A	7	GLY	-	expression tag	UNP Q14839

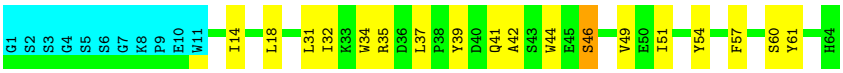
4 Residue-property plots (i)

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: Chromodomain helicase-DNA-binding protein 4

Chain A: 55% 27% • 17%



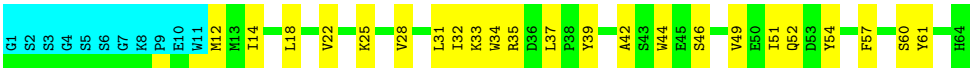
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

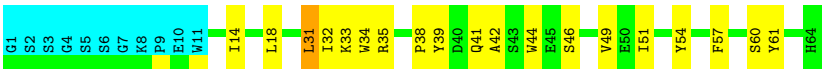
Chain A: 47% 36% 17%



4.2.2 Score per residue for model 2

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

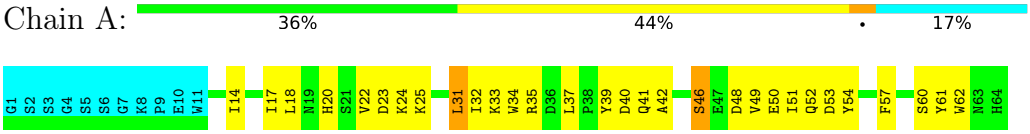
Chain A: 53% 28% • 17%



4.2.3 Score per residue for model 3 (medoid)

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

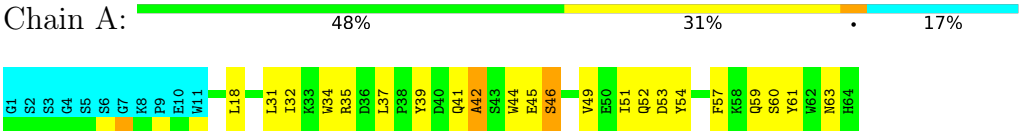
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

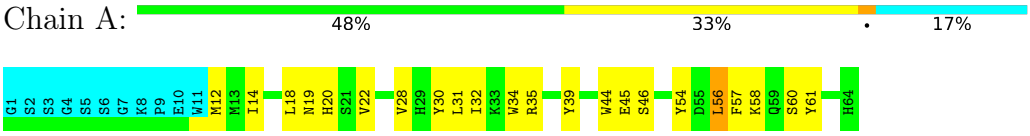
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

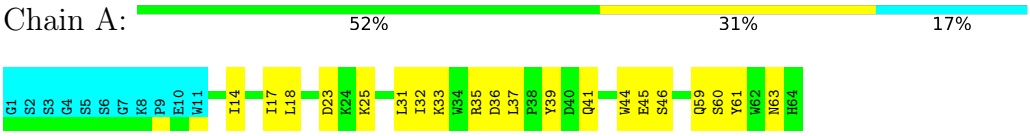
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

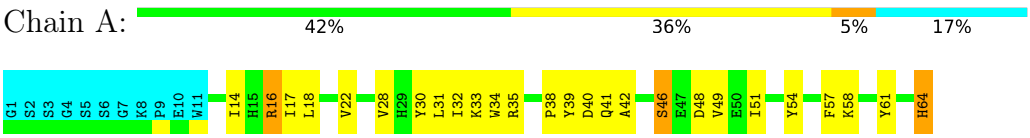
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

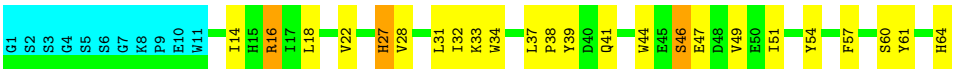
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 45% 33% 5% 17%



4.2.9 Score per residue for model 9

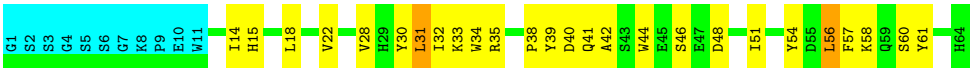
- Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 56% 25% • 17%

4.2.10 Score per residue for model 10

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

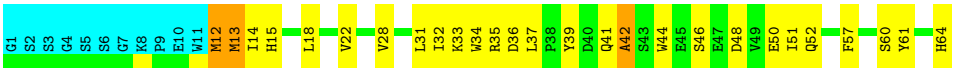
Chain A: 42% 38% • 17%



4.2.11 Score per residue for model 11

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

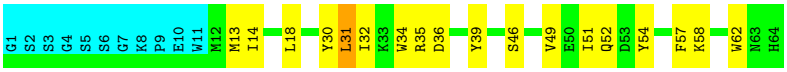
Chain A: 41% 38% 5% 17%



4.2.12 Score per residue for model 12

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 55% 27% • 17%



4.2.13 Score per residue for model 13

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 45% 34% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 I14 L18 V22 D23 V28 H29 Y30 L31 I32 K33 W34 R35 D36 L37 P38 Y39 D40 Q41 W44 E45 S46 E47 Y54 D55 L56 F57 K58 D59 S60 Y61 H64

4.2.14 Score per residue for model 14

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 50% 31% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 I14 L18 V22 K25 V28 L31 I32 K33 W34 R35 D36 L37 P38 Y39 D40 Q41 E45 S46 V49 E50 I51 Y54 D55 L56 Q59 N63 H64

4.2.15 Score per residue for model 15

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 61% 19% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 I14 L18 L31 I32 K33 W34 R35 P38 Y39 D40 Q41 S46 I51 Y54 F57 H64

4.2.16 Score per residue for model 16

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 53% 27% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 I14 L18 D23 K24 H27 L31 I32 K33 W34 R35 P38 Y39 D40 Q41 E45 S46 I51 Q52 D53 Y54 F57 H64

4.2.17 Score per residue for model 17

• Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 53% 27% • 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 I14 L18 D23 L31 I32 K33 W34 R35 D36 L37 P38 Y39 D40 Q41 W44 E45 S46 V49 E50 I51 Y54 F57 H64

4.2.18 Score per residue for model 18

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 44% 31% 8% 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 M12 M13 I14 H15 L18 V22 V28 L31 I32 K33 W34 R35 D36 Y39 W44 E45 S46 V49 E50 I51 Q52 D53 L56 F57 S60 Y61 H64

4.2.19 Score per residue for model 19

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 56% 23% • • 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 I14 L18 L31 I32 K33 W34 R35 D36 L37 P38 Y39 D40 Q41 A42 E45 S46 V49 Y54 F57 S60 Y61 H64

4.2.20 Score per residue for model 20

- Molecule 1: Chromodomain helicase-DNA-binding protein 4

Chain A: 45% 33% 5% 17%

G1 S2 S3 G4 S5 S6 G7 K8 P9 E10 W11 M12 M13 I14 H15 L18 Y30 L31 I32 K33 W34 R35 Y39 W44 E45 S46 V49 E50 I51 Y54 F57 K58 Q59 S60 Y61 W62 N63 H64

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations, target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

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	468	439	434	18±4
All	All	9360	8780	8680	358

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 20.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:18:LEU:HD22	1:A:39:TYR:CE2	0.97	1.95	10	4
1:A:18:LEU:HD13	1:A:39:TYR:CE1	0.90	2.01	14	19
1:A:22:VAL:HG22	1:A:28:VAL:HG22	0.90	1.42	10	7
1:A:37:LEU:HD22	1:A:41:GLN:OE1	0.85	1.70	13	4
1:A:37:LEU:HD22	1:A:41:GLN:NE2	0.79	1.93	14	3
1:A:46:SER:O	1:A:49:VAL:HG22	0.74	1.82	1	12
1:A:16:ARG:HD2	1:A:18:LEU:HD23	0.70	1.64	7	2
1:A:18:LEU:HD12	1:A:31:LEU:HG	0.70	1.63	20	19
1:A:14:ILE:HG23	1:A:32:ILE:HG23	0.69	1.64	13	12
1:A:18:LEU:HD22	1:A:39:TYR:CZ	0.65	2.27	10	5
1:A:31:LEU:N	1:A:44:TRP:CZ3	0.60	2.70	2	10
1:A:49:VAL:HG23	1:A:54:TYR:CE2	0.60	2.31	4	11
1:A:14:ILE:O	1:A:57:PHE:CZ	0.60	2.55	19	9
1:A:14:ILE:O	1:A:57:PHE:CE2	0.60	2.55	7	17

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:53:ASP:OD2	1:A:57:PHE:CE1	0.60	2.55	4	2
1:A:22:VAL:CG2	1:A:28:VAL:HG22	0.59	2.23	10	2
1:A:12:MET:SD	1:A:34:TRP:CZ3	0.59	2.95	1	1
1:A:32:ILE:HD11	1:A:54:TYR:CE1	0.59	2.32	17	8
1:A:32:ILE:CD1	1:A:51:ILE:CD1	0.59	2.79	18	8
1:A:53:ASP:OD1	1:A:57:PHE:CE1	0.58	2.57	3	1
1:A:18:LEU:HD13	1:A:39:TYR:HE1	0.56	1.59	9	3
1:A:34:TRP:CD1	1:A:37:LEU:HD12	0.56	2.35	17	6
1:A:19:ASN:ND2	1:A:20:HIS:H	0.56	1.98	5	1
1:A:18:LEU:CD1	1:A:42:ALA:CB	0.53	2.86	10	1
1:A:32:ILE:HD12	1:A:45:GLU:HB2	0.53	1.79	5	7
1:A:56:LEU:CD2	1:A:56:LEU:N	0.52	2.72	14	3
1:A:64:HIS:CD2	1:A:64:HIS:C	0.52	2.83	8	4
1:A:18:LEU:HD21	1:A:33:LYS:HB2	0.51	1.82	13	7
1:A:61:TYR:C	1:A:61:TYR:CD1	0.50	2.84	3	1
1:A:17:ILE:HG21	1:A:61:TYR:CD2	0.50	2.42	7	2
1:A:18:LEU:HD11	1:A:33:LYS:HB2	0.50	1.83	6	3
1:A:64:HIS:C	1:A:64:HIS:CD2	0.50	2.85	15	2
1:A:34:TRP:O	1:A:37:LEU:N	0.49	2.45	13	2
1:A:32:ILE:HD13	1:A:51:ILE:CD1	0.49	2.38	18	14
1:A:51:ILE:HG22	1:A:52:GLN:N	0.49	2.22	3	7
1:A:56:LEU:N	1:A:56:LEU:CD2	0.48	2.76	18	2
1:A:18:LEU:HD22	1:A:39:TYR:HE1	0.48	1.67	7	4
1:A:15:HIS:HB3	1:A:33:LYS:CG	0.48	2.39	10	1
1:A:38:PRO:CD	1:A:41:GLN:OE1	0.48	2.61	8	1
1:A:30:TYR:CE2	1:A:58:LYS:HE2	0.48	2.44	13	4
1:A:15:HIS:CB	1:A:33:LYS:O	0.48	2.62	20	1
1:A:32:ILE:HD13	1:A:51:ILE:HD13	0.47	1.86	18	3
1:A:59:GLN:O	1:A:63:ASN:ND2	0.47	2.47	4	3
1:A:36:ASP:CG	1:A:37:LEU:N	0.47	2.68	6	1
1:A:50:GLU:O	1:A:50:GLU:CG	0.47	2.62	18	4
1:A:18:LEU:HD21	1:A:33:LYS:CB	0.47	2.40	7	4
1:A:31:LEU:HB2	1:A:44:TRP:CZ2	0.47	2.45	2	4
1:A:22:VAL:HB	1:A:28:VAL:HG22	0.46	1.87	18	2
1:A:31:LEU:N	1:A:44:TRP:CE3	0.46	2.84	2	1
1:A:18:LEU:HD12	1:A:31:LEU:CG	0.46	2.39	18	7
1:A:34:TRP:O	1:A:35:ARG:C	0.46	2.54	13	18
1:A:54:TYR:CD1	1:A:58:LYS:HD2	0.46	2.46	5	4
1:A:19:ASN:ND2	1:A:20:HIS:N	0.46	2.64	5	1
1:A:49:VAL:O	1:A:54:TYR:CD2	0.46	2.68	19	5
1:A:12:MET:SD	1:A:34:TRP:CE3	0.45	3.09	1	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:60:SER:O	1:A:61:TYR:C	0.45	2.55	18	13
1:A:17:ILE:CG2	1:A:61:TYR:CD2	0.45	2.99	3	1
1:A:52:GLN:O	1:A:53:ASP:CG	0.45	2.55	18	1
1:A:30:TYR:CD2	1:A:58:LYS:HE2	0.45	2.45	10	1
1:A:18:LEU:HD11	1:A:42:ALA:CB	0.45	2.41	10	1
1:A:18:LEU:HD13	1:A:39:TYR:CZ	0.45	2.47	19	2
1:A:41:GLN:O	1:A:42:ALA:C	0.45	2.55	19	3
1:A:35:ARG:O	1:A:36:ASP:C	0.45	2.55	6	3
1:A:27:HIS:CD2	1:A:27:HIS:N	0.45	2.83	8	1
1:A:46:SER:OG	1:A:47:GLU:N	0.44	2.50	8	2
1:A:30:TYR:OH	1:A:62:TRP:NE1	0.44	2.50	12	2
1:A:54:TYR:CE1	1:A:58:LYS:HD2	0.44	2.48	10	4
1:A:14:ILE:O	1:A:15:HIS:C	0.44	2.56	11	2
1:A:18:LEU:HD22	1:A:39:TYR:CE1	0.43	2.47	7	1
1:A:14:ILE:O	1:A:14:ILE:HG22	0.43	2.14	11	1
1:A:31:LEU:HB2	1:A:44:TRP:CH2	0.43	2.49	13	2
1:A:38:PRO:CG	1:A:41:GLN:OE1	0.43	2.67	8	1
1:A:45:GLU:HB3	1:A:49:VAL:HG21	0.43	1.90	14	2
1:A:17:ILE:HG22	1:A:61:TYR:CD2	0.42	2.49	3	1
1:A:38:PRO:HD2	1:A:41:GLN:NE2	0.42	2.29	10	1
1:A:52:GLN:O	1:A:53:ASP:C	0.42	2.58	9	2
1:A:18:LEU:HD21	1:A:33:LYS:HD2	0.42	1.92	11	2
1:A:58:LYS:O	1:A:62:TRP:CD1	0.42	2.73	12	1
1:A:32:ILE:HD12	1:A:45:GLU:CB	0.42	2.44	16	1
1:A:38:PRO:CG	1:A:41:GLN:HG3	0.42	2.45	2	1
1:A:64:HIS:CD2	1:A:64:HIS:N	0.41	2.87	7	1
1:A:38:PRO:HG2	1:A:41:GLN:CG	0.41	2.45	7	4
1:A:18:LEU:HD11	1:A:42:ALA:HB2	0.41	1.91	10	1
1:A:20:HIS:CE1	1:A:62:TRP:CH2	0.41	3.08	3	1
1:A:31:LEU:HB2	1:A:44:TRP:CE2	0.41	2.51	10	1
1:A:31:LEU:HD13	1:A:44:TRP:CD1	0.41	2.51	10	1
1:A:27:HIS:CD2	1:A:27:HIS:H	0.40	2.34	8	1
1:A:64:HIS:CD2	1:A:64:HIS:H	0.40	2.34	7	1
1:A:23:ASP:OD1	1:A:27:HIS:N	0.40	2.55	16	1
1:A:36:ASP:OD1	1:A:36:ASP:C	0.40	2.59	17	1
1:A:23:ASP:CG	1:A:24:LYS:N	0.40	2.74	3	1
1:A:45:GLU:O	1:A:46:SER:O	0.40	2.40	19	1
1:A:15:HIS:N	1:A:35:ARG:HG3	0.40	2.32	11	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	52/64 (81%)	39±2 (76±3%)	11±2 (21±3%)	2±1 (3±1%)	7 40
All	All	1040/1280 (81%)	788 (76%)	221 (21%)	31 (3%)	7 40

All 4 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	46	SER	14
1	A	42	ALA	7
1	A	31	LEU	7
1	A	23	ASP	3

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	51/59 (86%)	49±1 (96±2%)	2±1 (4±2%)	37 85
All	All	1020/1180 (86%)	982 (96%)	38 (4%)	37 85

All 16 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	56	LEU	5
1	A	25	LYS	4
1	A	48	ASP	4
1	A	13	MET	4
1	A	46	SER	4
1	A	40	ASP	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	12	MET	3
1	A	16	ARG	2
1	A	36	ASP	2
1	A	22	VAL	1
1	A	64	HIS	1
1	A	27	HIS	1
1	A	55	ASP	1
1	A	52	GLN	1
1	A	24	LYS	1
1	A	23	ASP	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