Full wwPDB NMR Structure Validation Report (i)

Mar 3, 2022 – 07:08 AM EST

PDB ID	:	2HGM
Title	:	NMR structure of the second qRRM domain of human hnRNP F
Authors	:	Dominguez, C.; Allain, F.H.-T.
Deposited on	:	2006-06-27

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:

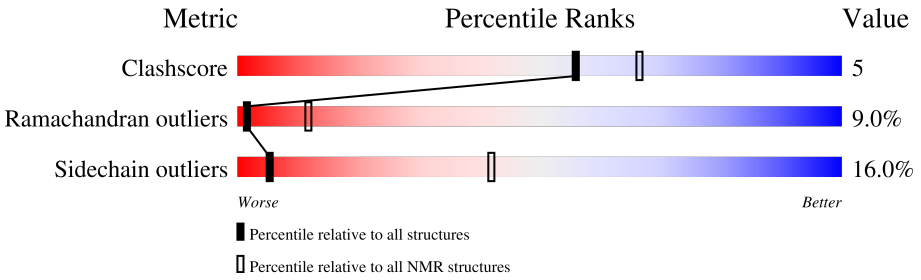
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	126	46% 16% • 7% 27%

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 14 is the overall representative, medoid model (most similar to other models).

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:109-A:191 (83)	0.28	14

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1429 atoms, of which 707 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Heterogeneous nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	92	1429	457	707	124	140	1	0

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	69	MET	-	cloning artifact	UNP Q5T0N2
A	70	GLY	-	cloning artifact	UNP Q5T0N2
A	71	SER	-	cloning artifact	UNP Q5T0N2
A	72	SER	-	cloning artifact	UNP Q5T0N2
A	73	HIS	-	expression tag	UNP Q5T0N2
A	74	HIS	-	expression tag	UNP Q5T0N2
A	75	HIS	-	expression tag	UNP Q5T0N2
A	76	HIS	-	expression tag	UNP Q5T0N2
A	77	HIS	-	expression tag	UNP Q5T0N2
A	78	HIS	-	expression tag	UNP Q5T0N2
A	79	SER	-	cloning artifact	UNP Q5T0N2
A	80	SER	-	cloning artifact	UNP Q5T0N2
A	81	GLY	-	cloning artifact	UNP Q5T0N2
A	82	LEU	-	cloning artifact	UNP Q5T0N2
A	83	VAL	-	cloning artifact	UNP Q5T0N2
A	84	PRO	-	cloning artifact	UNP Q5T0N2
A	85	ARG	-	cloning artifact	UNP Q5T0N2
A	86	GLY	-	cloning artifact	UNP Q5T0N2
A	87	SER	-	cloning artifact	UNP Q5T0N2
A	88	HIS	-	cloning artifact	UNP Q5T0N2
A	89	MET	-	cloning artifact	UNP Q5T0N2
A	90	ALA	-	cloning artifact	UNP Q5T0N2
A	91	SER	-	cloning artifact	UNP Q5T0N2
A	92	MET	-	cloning artifact	UNP Q5T0N2
A	93	THR	-	cloning artifact	UNP Q5T0N2
A	94	GLY	-	cloning artifact	UNP Q5T0N2
A	95	GLY	-	cloning artifact	UNP Q5T0N2
A	96	GLN	-	cloning artifact	UNP Q5T0N2
A	97	GLN	-	cloning artifact	UNP Q5T0N2
A	98	MET	-	cloning artifact	UNP Q5T0N2
A	99	GLY	-	cloning artifact	UNP Q5T0N2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	100	ARG	-	cloning artifact	UNP Q5T0N2
A	101	GLY	-	cloning artifact	UNP Q5T0N2
A	102	SER	-	cloning artifact	UNP Q5T0N2

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: Heterogeneous nuclear ribonucleoprotein F

Chain A: 46% 16% • 7% 27%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER N103 S104 A105 D106 S107 A108 N109 F112 R116 G117 L118 V128 P144 V145 I151 T152 G153 F156 V157

H172 R175 I176 G177 H178 R179 Y180 I181 E182 S187 Q188 V191 R192 S193 Y194

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: Heterogeneous nuclear ribonucleoprotein F

Chain A: 40% 21% 5% • 7% 27%

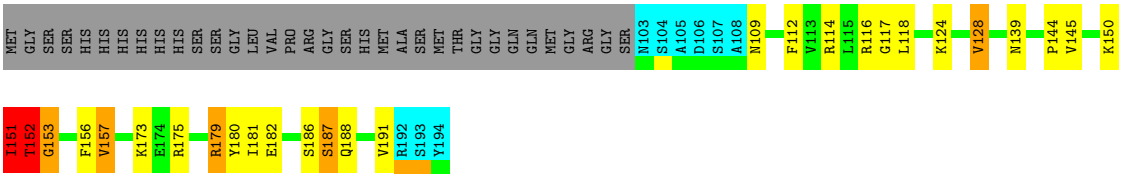
MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER N103 S104 A105 D106 S107 A108 N109 D110 G111 F112 V113 R114 L115 R116 G117 L118 P119 K124 V128 I136 N139

P144 V145 K150 I151 T152 G153 F156 V157 Q158 F159 E174 R175 I176 G177 H178 R179 Y180 I181 E182 S187 Q188 E189 E190 V191 R192 S193 Y194

4.2.2 Score per residue for model 2

• Molecule 1: Heterogeneous nuclear ribonucleoprotein F

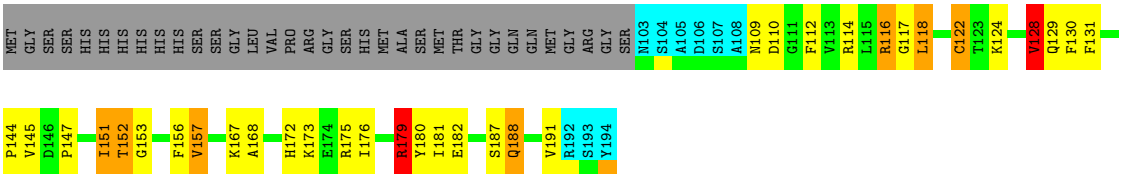
Chain A: 44% 16% • • 7% 27%



4.2.3 Score per residue for model 3

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

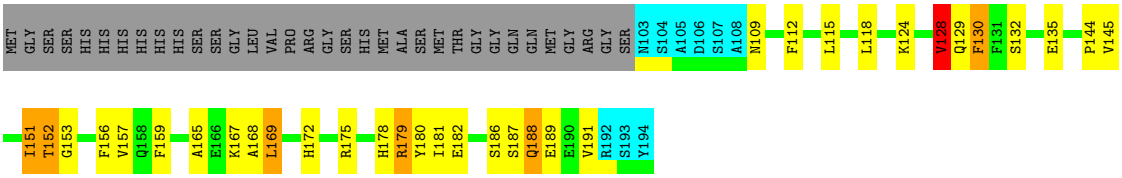
Chain A: 39% 20% 6% • 7% 27%



4.2.4 Score per residue for model 4

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

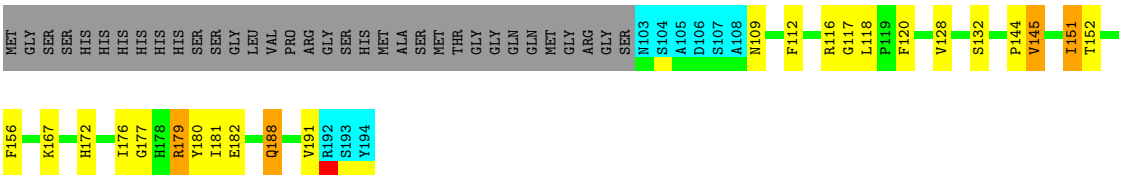
Chain A: 39% 21% 5% • 7% 27%



4.2.5 Score per residue for model 5

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 48% 15% • 7% 27%



4.2.6 Score per residue for model 6

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 42% 19% 5% 7% 27%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER N103 S104 A105 D106 S107 A108 N109 F112 V113 R114 L115 R116 G117 L118 C122 V128 I136 N139 P144 V145

I151 T152 G153 F156 F159 E163 L164 A165 A168 L169 H172 R175 I176 G177 H178 R179 Y180 I181 E182 Q188 V191 R192 S193 Y194

4.2.7 Score per residue for model 7

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 47% 13% 6% 7% 27%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER N103 S104 A105 D106 S107 A108 N109 D110 G111 G117 V128 I136 N139 P144 V145 I151 T152 V157 Q158

F159 Q162 E163 H172 R175 I176 G177 H178 R179 Y180 I181 E182 Q188 V191 R192 S193 Y194

4.2.8 Score per residue for model 8

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 41% 17% 7% 7% 27%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER N103 S104 A105 D106 S107 A108 G111 F112 V113 R114 L115 R116 G117 L118 K124 V128 F131 S132 G133 P144 V145

I151 T152 G153 F156 V157 K167 A168 H172 R175 I176 R179 Y180 I181 E182 V183 S187 Q188 V191 R192 S193 Y194

4.2.9 Score per residue for model 9

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 44% 16% • • 7% 27%

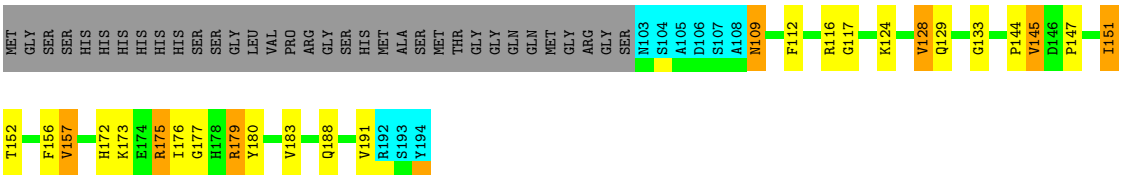
MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER N103 S104 A105 D106 S107 A108 F112 L115 R116 G117 L118 C122 T123 K124 V128 P144 V145 I151 T152 G153

F156 V157 Q158 E163 H172 K173 I176 G177 H178 R179 Y180 S187 Q188 E189 E190 V191 R192 S193 Y194

4.2.10 Score per residue for model 10

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

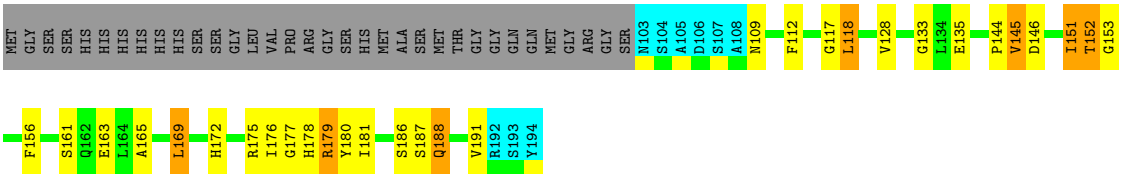
Chain A: 46% 14% 6% 7% 27%



4.2.11 Score per residue for model 11

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

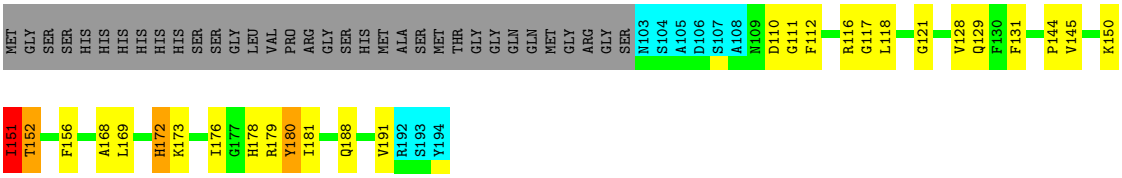
Chain A: 42% 18% 6% 7% 27%



4.2.12 Score per residue for model 12

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

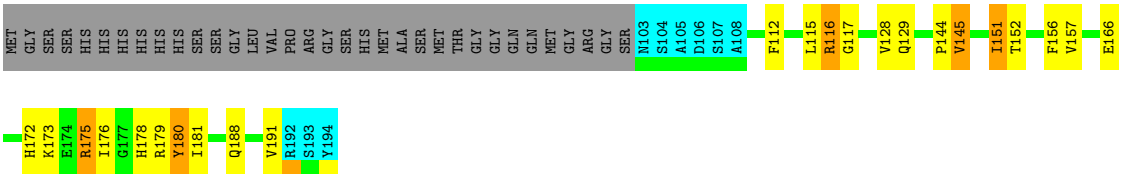
Chain A: 44% 18% • • 7% 27%



4.2.13 Score per residue for model 13

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

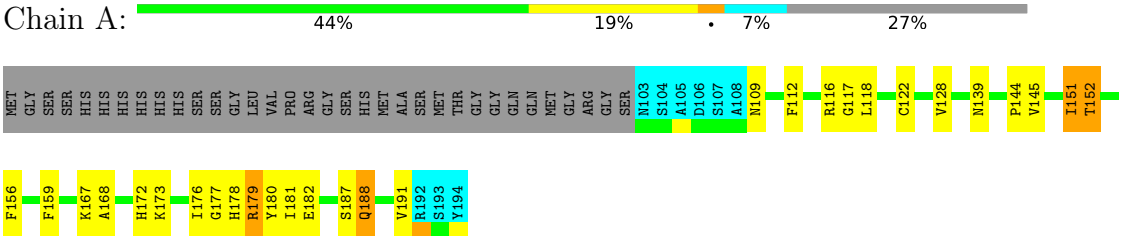
Chain A: 48% 14% • 7% 27%



4.2.14 Score per residue for model 14 (medoid)

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

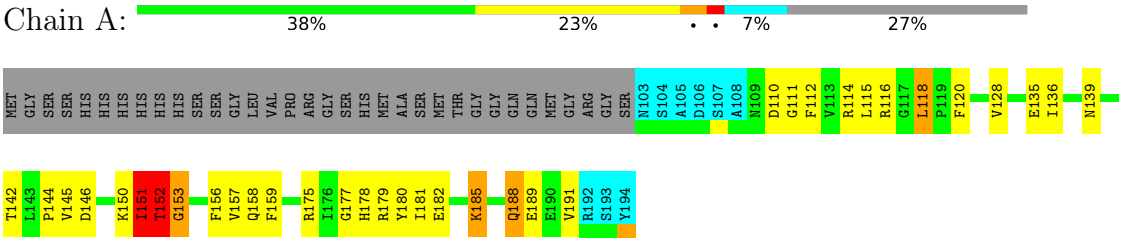
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 15 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
CYANA	structure solution	2.1
Amber	refinement	7.0

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	0.73±0.01	0±0/668 (0.0± 0.0%)	1.29±0.03	5±1/899 (0.6± 0.2%)
All	All	0.73	0/10020 (0.0%)	1.29	77/13485 (0.6%)

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	3.3±0.9
All	All	0	49

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	191	VAL	CA-CB-CG1	8.57	123.75	110.90	7	2
1	A	175	ARG	NE-CZ-NH1	8.05	124.33	120.30	7	9
1	A	114	ARG	NE-CZ-NH1	6.94	123.77	120.30	8	6
1	A	128	VAL	CA-CB-CG1	6.92	121.27	110.90	12	15
1	A	116	ARG	NE-CZ-NH1	6.81	123.71	120.30	1	10
1	A	179	ARG	NE-CZ-NH1	6.62	123.61	120.30	3	9
1	A	128	VAL	CA-CB-CG2	6.17	120.15	110.90	14	11
1	A	157	VAL	CA-CB-CG1	6.05	119.98	110.90	10	6
1	A	191	VAL	CA-CB-CG2	5.99	119.89	110.90	8	3
1	A	187	SER	C-N-CA	5.58	135.64	121.70	2	4
1	A	130	PHE	CB-CG-CD2	-5.20	117.16	120.80	4	1
1	A	182	GLU	N-CA-CB	-5.15	101.33	110.60	4	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	179	ARG	Peptide,Sidechain	15
1	A	180	TYR	Peptide	15
1	A	153	GLY	Peptide	6
1	A	115	LEU	Peptide	4
1	A	175	ARG	Sidechain	3
1	A	189	GLU	Peptide	1
1	A	191	VAL	Peptide	1
1	A	117	GLY	Peptide	1
1	A	190	GLU	Peptide	1
1	A	133	GLY	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	654	650	650	7±3
All	All	9810	9750	9750	101

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 5.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:118:LEU:HD23	1:A:181:ILE:CD1	0.72	2.14	15	3
1:A:156:PHE:CD1	1:A:191:VAL:HG12	0.67	2.25	6	8
1:A:118:LEU:HD23	1:A:181:ILE:HD13	0.60	1.72	15	2
1:A:112:PHE:CD2	1:A:191:VAL:HG11	0.57	2.34	10	11
1:A:118:LEU:HD13	1:A:122:CYS:SG	0.56	2.41	6	3
1:A:156:PHE:CD2	1:A:191:VAL:HG12	0.54	2.38	4	5
1:A:124:LYS:O	1:A:128:VAL:HG13	0.53	2.04	10	7
1:A:187:SER:O	1:A:191:VAL:HG13	0.52	2.05	8	3
1:A:112:PHE:CG	1:A:191:VAL:HG11	0.51	2.41	15	5
1:A:112:PHE:CD1	1:A:191:VAL:HG11	0.51	2.40	15	2
1:A:156:PHE:CE2	1:A:191:VAL:HA	0.50	2.41	4	3
1:A:118:LEU:HD22	1:A:176:ILE:HD12	0.50	1.82	3	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:118:LEU:HB2	1:A:153:GLY:HA2	0.49	1.84	4	5
1:A:118:LEU:HD12	1:A:152:THR:O	0.47	2.08	9	1
1:A:131:PHE:CE1	1:A:168:ALA:HB1	0.47	2.45	8	3
1:A:176:ILE:HD12	1:A:181:ILE:CD1	0.46	2.40	14	3
1:A:188:GLN:HA	1:A:191:VAL:CG2	0.45	2.41	4	6
1:A:165:ALA:O	1:A:169:LEU:HD12	0.45	2.11	6	3
1:A:118:LEU:HD22	1:A:181:ILE:HD13	0.45	1.88	11	1
1:A:159:PHE:CE2	1:A:168:ALA:HB2	0.44	2.48	6	3
1:A:118:LEU:HD23	1:A:181:ILE:HG13	0.43	1.90	3	1
1:A:146:ASP:CA	1:A:153:GLY:H	0.43	2.27	11	2
1:A:119:PRO:CD	1:A:176:ILE:HG22	0.43	2.44	1	1
1:A:121:GLY:H	1:A:151:ILE:HD11	0.42	1.74	12	1
1:A:151:ILE:HG23	1:A:152:THR:H	0.42	1.74	2	3
1:A:136:ILE:HA	1:A:159:PHE:HA	0.41	1.93	7	4
1:A:180:TYR:N	1:A:181:ILE:HD12	0.41	2.30	12	1
1:A:136:ILE:HG23	1:A:158:GLN:O	0.41	2.16	15	2
1:A:118:LEU:HD23	1:A:181:ILE:HG12	0.41	1.92	6	1
1:A:176:ILE:HB	1:A:181:ILE:CD1	0.41	2.46	8	1
1:A:151:ILE:HD13	1:A:152:THR:H	0.41	1.76	9	1
1:A:112:PHE:CD1	1:A:158:GLN:OE1	0.41	2.74	9	1
1:A:112:PHE:O	1:A:185:LYS:HE2	0.41	2.16	15	1
1:A:180:TYR:C	1:A:181:ILE:HD12	0.40	2.36	12	2
1:A:122:CYS:SG	1:A:176:ILE:HG21	0.40	2.57	14	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	83/126 (66%)	65±3 (78±3%)	10±2 (13±3%)	7±2 (9±2%)	1 12
All	All	1245/1890 (66%)	977 (78%)	156 (13%)	112 (9%)	1 12

All 17 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	144	PRO	15
1	A	151	ILE	15
1	A	188	GLN	15
1	A	152	THR	14
1	A	117	GLY	12
1	A	172	HIS	11
1	A	177	GLY	8
1	A	145	VAL	5
1	A	111	GLY	4
1	A	110	ASP	3
1	A	109	ASN	2
1	A	147	PRO	2
1	A	133	GLY	2
1	A	174	GLU	1
1	A	181	ILE	1
1	A	132	SER	1
1	A	153	GLY	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	70/103 (68%)	59±3 (84±4%)	11±3 (16±4%)	5 42
All	All	1050/1545 (68%)	882 (84%)	168 (16%)	5 42

All 38 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	145	VAL	15
1	A	151	ILE	15
1	A	157	VAL	10
1	A	152	THR	10
1	A	182	GLU	9
1	A	178	HIS	8
1	A	173	LYS	7
1	A	109	ASN	7
1	A	118	LEU	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	176	ILE	7
1	A	139	ASN	6
1	A	129	GLN	5
1	A	167	LYS	5
1	A	150	LYS	4
1	A	116	ARG	4
1	A	163	GLU	4
1	A	175	ARG	4
1	A	189	GLU	3
1	A	186	SER	3
1	A	179	ARG	3
1	A	135	GLU	3
1	A	169	LEU	3
1	A	172	HIS	3
1	A	110	ASP	2
1	A	128	VAL	2
1	A	130	PHE	2
1	A	132	SER	2
1	A	120	PHE	2
1	A	180	TYR	2
1	A	183	VAL	2
1	A	188	GLN	2
1	A	190	GLU	1
1	A	122	CYS	1
1	A	187	SER	1
1	A	162	GLN	1
1	A	161	SER	1
1	A	166	GLU	1
1	A	185	LYS	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