Full wwPDB NMR Structure Validation Report (i)

Mar 6, 2022 – 04:22 PM EST

PDB ID	:	2RQP
Title	:	The Solution Structure of Heterochromatin Protein 1-Binding Protein 74 Histone H1 like domain
Authors	:	Shimamoto, S.; Sugahara, H.; Ohkubo, T.
Deposited on	:	2009-09-04

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:

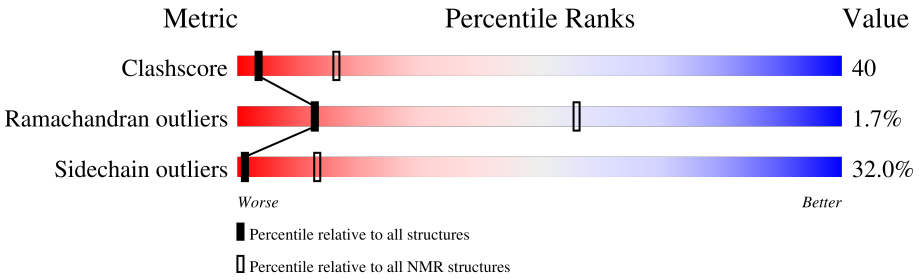
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	88	25% 25% 10% 40%

2 Ensemble composition and analysis (i)

This entry contains 12 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:161-A:213 (53)	0.30	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 3 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1422 atoms, of which 743 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Heterochromatin protein 1-binding protein 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	88	1422	431	743	128	117	3	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	150	GLY	-	expression tag	UNP Q5SSJ5
A	151	PRO	-	expression tag	UNP Q5SSJ5
A	152	GLY	-	expression tag	UNP Q5SSJ5

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: Heterochromatin protein 1-binding protein 3

Chain A: 25% 25% 10% 40%

G150 P151 G152 M153 A154 S155 S156 P157 R158 P159 K160 M161 I164 L165 A168 I169 K170 A171 C172 F173 S176 G177 A178 S179 V180 I183 R184 K185 Y186 I187 I188 H189 K190 Y191 L194 E197 R198 R199 G200 L203 K204 Q205 A206 L207 K208 R209 E210 R213 G214 V215 I216 K217

Q218 V219 K220 G221 K222 G223 A224 S225 G226 S227 F228 V229 V230 V231 Q232 K233 S234 R235 K236 T237

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: Heterochromatin protein 1-binding protein 3

Chain A: 23% 26% 11% 40%

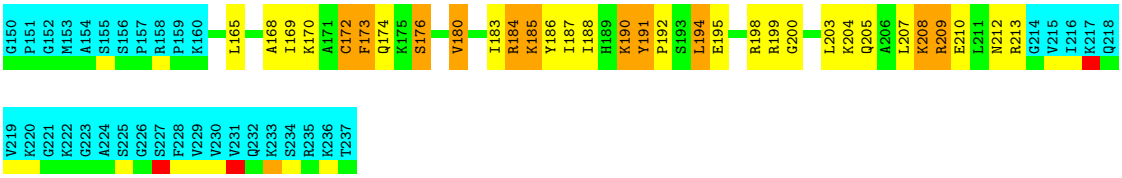
G150 P151 G152 M153 A154 S155 S156 P157 R158 P159 K160 M161 D162 A163 I164 L165 T166 E167 A168 I169 K170 A171 C172 F173 S176 G177 A178 S179 V180 I183 R184 K185 Y186 I187 I188 H189 K190 Y191 L194 E197 R198 R199 G200 L203 K204 Q205 A206 L207 K208 R209 E210 R213 G214 V215

I216 K217 Q218 V219 K220 G221 K222 G223 A224 S225 G226 S227 F228 V229 V230 V231 Q232 K233 S234 R235 K236 T237

4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: Heterochromatin protein 1-binding protein 3

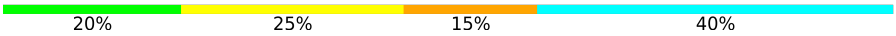
Chain A: 24% 24% 13% 40%

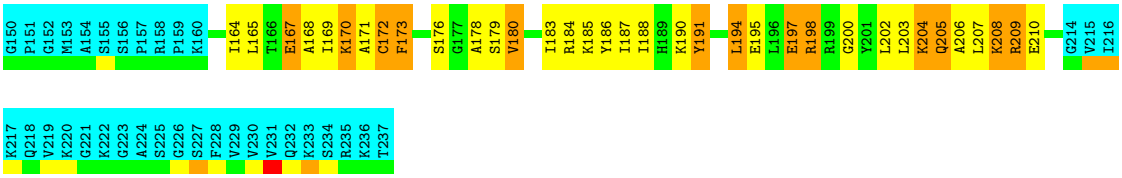


4.2.3 Score per residue for model 3

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A:

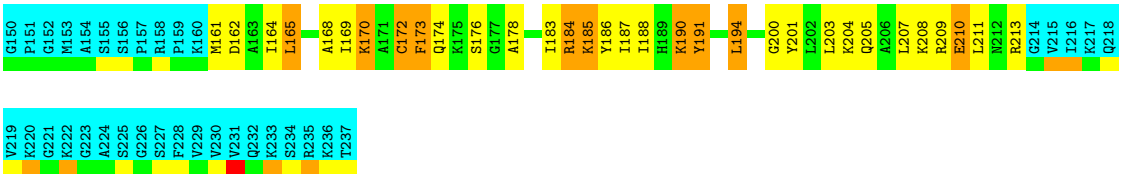




4.2.4 Score per residue for model 4

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A:

4.2.6 Score per residue for model 6

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A: 27% 22% 11% 40%

G150 P151 G152 M153 A154 S155 S156 P157 R158 P159 K160 M161 I164 L165 A168 I169 K170 A171 C172 F173 Q174 K175 S176 V180 I183 R184 K185 Y186 I187 I188 H189 K190 Y191 L194 R198 R199 G200 L203 K204 Q205 A206 L207 K208 R209 R213 G214 V215 I216 K217 Q218 V219 K220

G221 K222 G223 A224 S225 G226 S227 F228 V229 V230 V231 Q232 K233 S234 R235 K236 T237

4.2.7 Score per residue for model 7

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A: 20% 31% 9% 40%

G150 P151 G152 M153 A154 S155 S156 P157 R158 P159 K160 I164 L165 A168 I169 C172 F173 Q174 K175 A178 S179 I183 R184 K185 Y186 I187 I188 H189 K190 Y191 L194 E197 R198 R199 G200 Y201 L202 L203 K204 Q205 A206 L207 K208 R209 E210 L211 N212 R213 G214 V215 I216 K217 Q218

V219 K220 G221 K222 G223 A224 S225 G226 S227 F228 V229 V230 V231 Q232 K233 S234 R235 K236 T237

4.2.8 Score per residue for model 8

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A: 23% 25% 13% 40%

G150 P151 G152 M153 A154 S155 S156 P157 R158 P159 K160 L165 T166 E167 A168 I169 K170 A171 C172 F173 Q174 K175 A178 S179 V180 I183 R184 K185 Y186 I187 I188 H189 K190 Y191 L194 E197 R198 R199 G200 Y201 L202 L203 K204 Q205 A206 L207 K208 R209 E210 L211 N212 R213 G214 V215 I216

K217 Q218 V219 K220 G221 K222 G223 A224 S225 G226 S227 F228 V229 V230 V231 Q232 K233 S234 R235 K236 T237

4.2.9 Score per residue for model 9

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A: 22% 30% 9% 40%

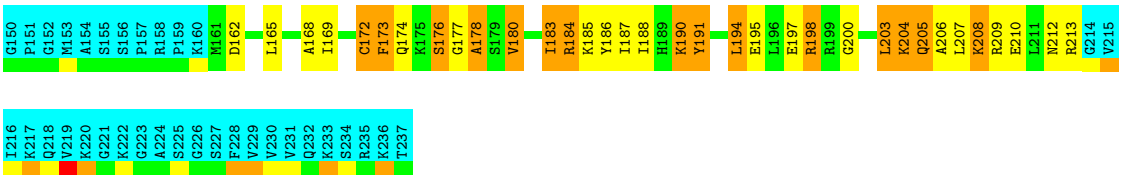
G150 P151 G152 M153 A154 S155 S156 P157 R158 P159 K160 M161 I164 L165 T166 E167 A168 I169 K170 A171 C172 F173 Q174 K175 S176 G177 A178 S179 V180 I183 R184 K185 Y186 I187 I188 H189 K190 Y191 L194 E197 R198 R199 G200 L203 K204 Q205 A206 L207 K208 R209 G214 V215 I216 K217

Q218 V219 K220 G221 K222 G223 A224 S225 G226 S227 F228 V229 V230 V231 Q232 K233 S234 R235 K236 T237

4.2.10 Score per residue for model 10

- Molecule 1: Heterochromatin protein 1-binding protein 3

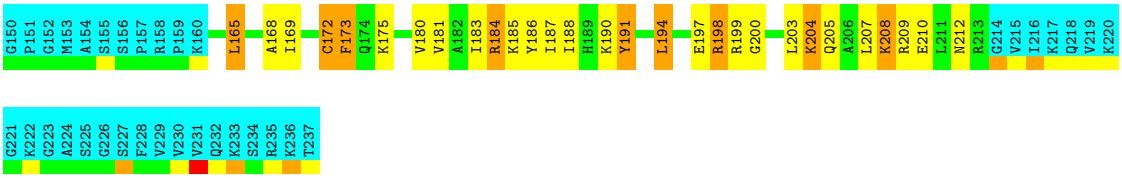
Chain A: 22% 22% 17% 40%



4.2.11 Score per residue for model 11

- Molecule 1: Heterochromatin protein 1-binding protein 3

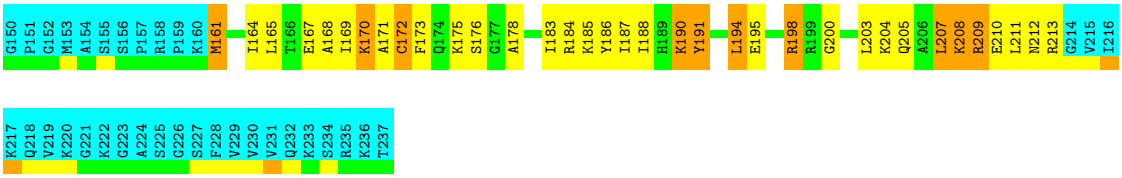
Chain A: 27% 23% 10% 40%



4.2.12 Score per residue for model 12

- Molecule 1: Heterochromatin protein 1-binding protein 3

Chain A: 20% 28% 11% 40%



5 Refinement protocol and experimental data overview (i)

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

Of the 1000 calculated structures, 12 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 1.1	refinement	

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	430	466	465	36±2
All	All	5160	5592	5580	429

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 40.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:187:ILE:HD11	1:A:203:LEU:HD13	0.97	1.36	10	1
1:A:165:LEU:HD21	1:A:183:ILE:HG22	0.91	1.42	2	6
1:A:183:ILE:HD13	1:A:207:LEU:CB	0.79	2.06	10	12
1:A:188:ILE:HG23	1:A:194:LEU:HD13	0.77	1.53	10	12
1:A:165:LEU:HD21	1:A:183:ILE:CG2	0.74	2.11	2	3
1:A:161:MET:HA	1:A:164:ILE:HD12	0.72	1.61	4	2
1:A:165:LEU:O	1:A:165:LEU:HD13	0.71	1.84	2	7
1:A:183:ILE:HG21	1:A:207:LEU:HD13	0.71	1.63	7	3
1:A:187:ILE:CD1	1:A:203:LEU:HD13	0.70	2.14	10	1
1:A:187:ILE:CG1	1:A:203:LEU:HD13	0.68	2.18	8	11
1:A:165:LEU:CD1	1:A:206:ALA:HB3	0.68	2.19	3	5
1:A:185:LYS:HA	1:A:188:ILE:HD12	0.68	1.64	10	9
1:A:172:CYS:HB3	1:A:178:ALA:HB2	0.68	1.64	9	5
1:A:183:ILE:CG2	1:A:207:LEU:HD13	0.65	2.21	7	4

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:183:ILE:C	1:A:203:LEU:HD21	0.65	2.11	10	1
1:A:169:ILE:HD12	1:A:170:LYS:N	0.65	2.06	3	1
1:A:165:LEU:HD23	1:A:165:LEU:O	0.64	1.93	9	5
1:A:161:MET:O	1:A:164:ILE:HG22	0.63	1.93	12	3
1:A:187:ILE:O	1:A:191:TYR:O	0.63	2.16	12	12
1:A:187:ILE:HG12	1:A:203:LEU:HD13	0.63	1.68	8	9
1:A:183:ILE:HD13	1:A:207:LEU:HB3	0.63	1.70	10	10
1:A:183:ILE:HD13	1:A:207:LEU:HD22	0.62	1.70	7	4
1:A:165:LEU:HD11	1:A:206:ALA:HB3	0.61	1.71	3	2
1:A:184:ARG:HA	1:A:203:LEU:HD11	0.61	1.72	8	11
1:A:166:THR:HA	1:A:169:ILE:HD12	0.60	1.73	8	1
1:A:187:ILE:HD11	1:A:203:LEU:CD1	0.59	2.22	10	1
1:A:164:ILE:HD11	1:A:186:TYR:OH	0.59	1.98	9	2
1:A:183:ILE:HD13	1:A:207:LEU:HB2	0.58	1.76	12	5
1:A:205:GLN:O	1:A:209:ARG:CB	0.58	2.52	6	12
1:A:188:ILE:HA	1:A:194:LEU:HD22	0.57	1.76	11	12
1:A:168:ALA:O	1:A:172:CYS:CB	0.56	2.53	12	12
1:A:167:GLU:O	1:A:171:ALA:HB3	0.56	2.00	9	4
1:A:200:GLY:O	1:A:204:LYS:HB3	0.55	2.01	6	12
1:A:165:LEU:HD11	1:A:183:ILE:CG2	0.54	2.33	4	2
1:A:183:ILE:HD13	1:A:207:LEU:CD2	0.53	2.34	7	3
1:A:165:LEU:HD21	1:A:207:LEU:CD1	0.53	2.33	5	1
1:A:169:ILE:O	1:A:173:PHE:CB	0.53	2.57	1	12
1:A:207:LEU:C	1:A:207:LEU:HD13	0.52	2.24	11	5
1:A:205:GLN:O	1:A:209:ARG:HB3	0.52	2.04	3	2
1:A:183:ILE:C	1:A:203:LEU:HD12	0.51	2.26	12	10
1:A:165:LEU:HD21	1:A:207:LEU:HD13	0.50	1.81	5	1
1:A:188:ILE:HG12	1:A:194:LEU:HD13	0.50	1.82	2	2
1:A:173:PHE:CD2	1:A:178:ALA:HB3	0.50	2.42	12	1
1:A:169:ILE:O	1:A:173:PHE:HB3	0.50	2.06	6	11
1:A:204:LYS:O	1:A:208:LYS:N	0.50	2.42	1	11
1:A:197:GLU:HG3	1:A:202:LEU:HD12	0.50	1.84	8	1
1:A:183:ILE:CD1	1:A:207:LEU:HD22	0.50	2.35	2	3
1:A:186:TYR:HD1	1:A:187:ILE:HD13	0.49	1.66	2	1
1:A:186:TYR:O	1:A:190:LYS:N	0.49	2.44	9	12
1:A:204:LYS:O	1:A:208:LYS:CG	0.49	2.61	5	8
1:A:205:GLN:O	1:A:209:ARG:HB2	0.49	2.07	10	10
1:A:164:ILE:HG23	1:A:186:TYR:CE1	0.49	2.43	3	1
1:A:169:ILE:HG21	1:A:210:GLU:OE1	0.49	2.07	8	1
1:A:183:ILE:HD13	1:A:207:LEU:CG	0.48	2.38	10	3
1:A:180:VAL:O	1:A:183:ILE:HG13	0.48	2.08	10	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:183:ILE:HG21	1:A:207:LEU:HG	0.48	1.86	6	1
1:A:184:ARG:N	1:A:203:LEU:HD21	0.48	2.24	10	1
1:A:173:PHE:HD2	1:A:178:ALA:HB3	0.48	1.69	12	4
1:A:164:ILE:CG2	1:A:186:TYR:CE1	0.47	2.98	5	3
1:A:203:LEU:HG	1:A:204:LYS:N	0.47	2.25	3	11
1:A:164:ILE:HG21	1:A:186:TYR:CE1	0.47	2.44	5	1
1:A:169:ILE:HD12	1:A:207:LEU:HD21	0.47	1.86	6	1
1:A:204:LYS:O	1:A:208:LYS:HG2	0.47	2.10	10	2
1:A:208:LYS:O	1:A:212:ASN:N	0.47	2.42	8	5
1:A:165:LEU:HD11	1:A:183:ILE:HG21	0.47	1.86	6	1
1:A:165:LEU:HD11	1:A:183:ILE:HG22	0.47	1.86	4	1
1:A:191:TYR:O	1:A:194:LEU:HD23	0.46	2.11	8	3
1:A:207:LEU:HG	1:A:211:LEU:HD12	0.46	1.85	4	1
1:A:165:LEU:CD2	1:A:183:ILE:HG22	0.46	2.30	8	2
1:A:180:VAL:HA	1:A:207:LEU:HD22	0.46	1.87	9	1
1:A:165:LEU:HD13	1:A:206:ALA:CB	0.45	2.41	9	1
1:A:177:GLY:O	1:A:178:ALA:C	0.45	2.55	10	1
1:A:207:LEU:CD1	1:A:211:LEU:HD13	0.45	2.41	12	1
1:A:183:ILE:HD13	1:A:207:LEU:HD13	0.45	1.86	9	1
1:A:165:LEU:HD11	1:A:207:LEU:HD23	0.45	1.88	12	2
1:A:183:ILE:HB	1:A:203:LEU:HD11	0.45	1.88	10	1
1:A:169:ILE:O	1:A:173:PHE:HB2	0.45	2.11	12	1
1:A:197:GLU:CD	1:A:202:LEU:HD12	0.45	2.32	3	1
1:A:165:LEU:HD13	1:A:206:ALA:HB3	0.45	1.87	9	1
1:A:183:ILE:HD13	1:A:207:LEU:HG	0.45	1.89	6	2
1:A:186:TYR:O	1:A:190:LYS:CB	0.45	2.65	12	7
1:A:204:LYS:O	1:A:208:LYS:CB	0.44	2.66	5	5
1:A:184:ARG:HA	1:A:203:LEU:CD1	0.44	2.43	11	8
1:A:198:ARG:C	1:A:200:GLY:H	0.44	2.16	5	11
1:A:169:ILE:CD1	1:A:207:LEU:HD21	0.44	2.43	6	1
1:A:186:TYR:O	1:A:190:LYS:HB2	0.43	2.12	11	1
1:A:172:CYS:O	1:A:176:SER:N	0.43	2.50	2	2
1:A:180:VAL:O	1:A:183:ILE:CG1	0.43	2.67	3	7
1:A:183:ILE:HB	1:A:203:LEU:HD21	0.43	1.90	10	1
1:A:172:CYS:O	1:A:174:GLN:N	0.43	2.52	2	3
1:A:194:LEU:HD12	1:A:198:ARG:NE	0.43	2.29	6	1
1:A:207:LEU:HD13	1:A:211:LEU:HD13	0.43	1.89	12	1
1:A:203:LEU:H	1:A:203:LEU:HD23	0.43	1.74	4	3
1:A:181:VAL:HG12	1:A:185:LYS:HG3	0.43	1.90	11	1
1:A:183:ILE:HB	1:A:203:LEU:HD12	0.42	1.91	2	2
1:A:187:ILE:HG22	1:A:194:LEU:CA	0.42	2.44	12	3

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:207:LEU:O	1:A:211:LEU:HD13	0.42	2.15	5	1
1:A:169:ILE:HG23	1:A:170:LYS:N	0.42	2.29	12	1
1:A:206:ALA:O	1:A:209:ARG:HG2	0.42	2.14	5	1
1:A:183:ILE:CG2	1:A:207:LEU:CD1	0.42	2.98	9	1
1:A:187:ILE:HG13	1:A:203:LEU:HD13	0.41	1.91	6	1
1:A:188:ILE:HA	1:A:194:LEU:CB	0.41	2.45	10	2
1:A:197:GLU:CG	1:A:202:LEU:HD12	0.41	2.46	7	1
1:A:201:TYR:O	1:A:205:GLN:CB	0.41	2.69	4	1
1:A:169:ILE:CG2	1:A:207:LEU:CD2	0.41	2.99	1	1
1:A:194:LEU:O	1:A:198:ARG:CB	0.41	2.69	3	1
1:A:169:ILE:CG2	1:A:207:LEU:CD1	0.41	2.99	4	1
1:A:183:ILE:HD12	1:A:203:LEU:CD2	0.40	2.45	10	1
1:A:183:ILE:HG13	1:A:184:ARG:N	0.40	2.32	1	2
1:A:169:ILE:HG13	1:A:170:LYS:N	0.40	2.31	4	1
1:A:194:LEU:O	1:A:198:ARG:HB2	0.40	2.16	5	3
1:A:191:TYR:CD2	1:A:192:PRO:HD2	0.40	2.51	2	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	53/88 (60%)	47±1 (89±1%)	5±1 (9±1%)	1±0 (2±1%)	13	56
All	All	636/1056 (60%)	565 (89%)	60 (9%)	11 (2%)	13	56

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	173	PHE	10
1	A	178	ALA	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	45/72 (62%)	31±2 (68±4%)	14±2 (32±4%)	1 13
All	All	540/864 (62%)	367 (68%)	173 (32%)	1 13

All 32 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	172	CYS	12
1	A	191	TYR	12
1	A	194	LEU	12
1	A	208	LYS	12
1	A	176	SER	9
1	A	190	LYS	9
1	A	210	GLU	9
1	A	184	ARG	9
1	A	185	LYS	8
1	A	213	ARG	8
1	A	198	ARG	7
1	A	204	LYS	7
1	A	170	LYS	7
1	A	175	LYS	6
1	A	197	GLU	5
1	A	180	VAL	5
1	A	199	ARG	5
1	A	162	ASP	4
1	A	195	GLU	4
1	A	209	ARG	3
1	A	161	MET	3
1	A	207	LEU	2
1	A	167	GLU	2
1	A	179	SER	2
1	A	205	GLN	2
1	A	165	LEU	2
1	A	174	GLN	2
1	A	189	HIS	1
1	A	211	LEU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	164	ILE	1
1	A	183	ILE	1
1	A	203	LEU	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