WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:53 pm BST

PDB ID	:	2L6Y
Title	:	haddock model of GATA1NF:Lmo2LIM2-Ldb1LID
Authors	:	Wilkinson-White, L.; Gamsjaeger, R.; Dastmalchi, S.; Wienert, B.; Stokes, P.H.; Crossley, M.; Mackay, J.P.; Matthews, J.M.
Deposited on	:	2010-12-01

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:

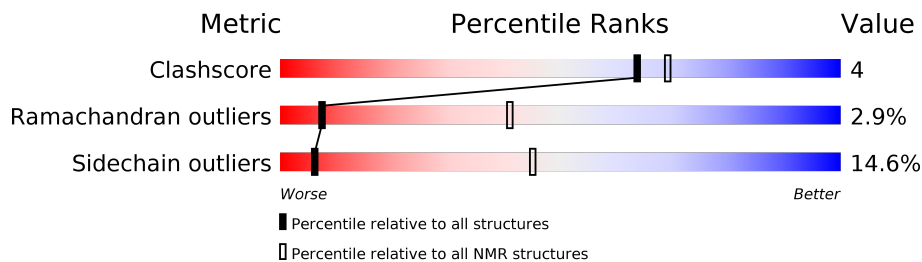|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	Percentile Ranks	Value
Clashscore		4
Ramachandran outliers		2.9%
Sidechain outliers		14.6%

Worse   *Better*

■ Percentile relative to all structures
❙ Percentile relative to all NMR structures

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	39	77% 15% 5% ·
2	B	96	72% 16% · 11%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 4 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:201-A:238, B:91-B:175 (123)	0.42	4

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 2 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2040 atoms, of which 997 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Erythroid transcription factor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	39	592	183	286	62	56	5	0

- Molecule 2 is a protein called LIM domain only 2, linker, LIM domain-binding protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	96	1445	457	711	127	140	10	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	130	SER	CYS	ENGINEERED MUTATION	UNP Q544Z2

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

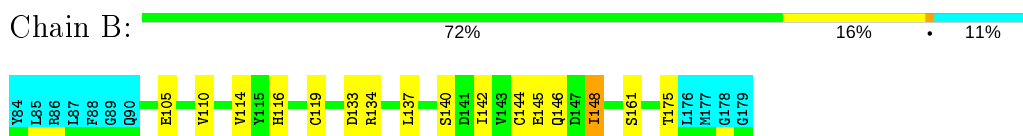| Mol	Chain	Residues	Atoms	
3	B	2	Total	Zn
			2	2
3	A	1	Total	Zn
			1	1

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Erythroid transcription factor

Chain A:

77%   15%   5%   •

E200 A201 R202 E203 N206 C207 G208 H222 C225 N226 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

72%   16%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 E105 V110 V114 Y115 H116 C119 D133 R134 L137 S140 D141 I142 V143 C144 E145 Q146 D147 I148 S161 T175 L176 M177 G178 G179
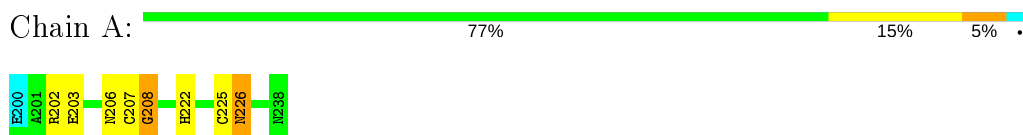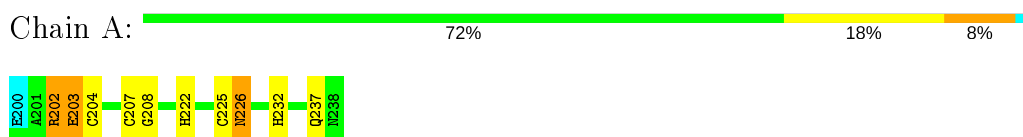
4.2 Scores per residue for each member of the ensemble
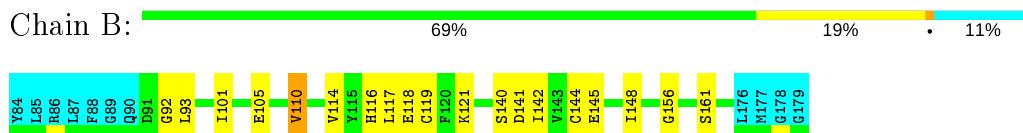
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
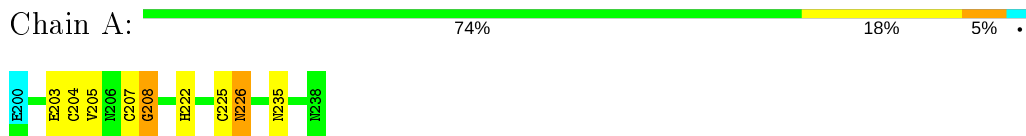
• Molecule 1: Erythroid transcription factor

Chain A:

72%   18%   8%   •

E200 A201 R202 E203 C204 C207 G208 H222 C225 N226 H232 Q237 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

69%   19%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 D91 G92 L93 I101 E105 V110 V114 Y115 H116 L117 E118 C119 F120 K121 S140 D141 I142 V143 C144 E145 I148 G156 S161 L176 M177 G178 G179

4.2.2 Score per residue for model 2

• Molecule 1: Erythroid transcription factor

Chain A: 74%   18%   5%   •

E200 E203 C204 V205 N206 C207 G208 H222 C225 N226 N235 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B: 66%   22%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 D98 E105 M106 T107 V110 V114 Y115 H116 C119 F120 K121 S130 D133 R134 S140 D141 I142 V143 C144 E145 Q146 D147 I148 G157 S161 S164 P174 T175 L176 M177 G178 G179
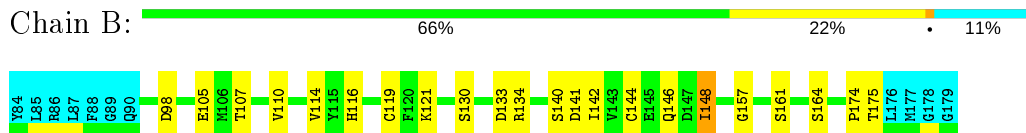
4.2.3 Score per residue for model 3

• Molecule 1: Erythroid transcription factor

Chain A: 72%   18%   8%   •

E200 E203 C207 G208 A209 T210 W215 H222 C225 N226 H232 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B: 70%   17%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 S96 R102 E105 V110 V114 Y115 H116 C119 R134 L137 I138 I142 V143 C144 E145 Q146 D147 I148 G157 S164 T175 L176 M177 G178 G179
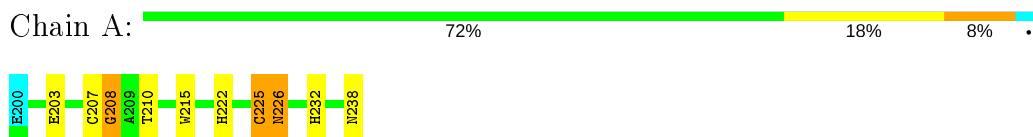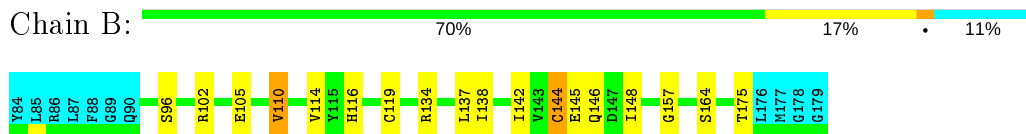
4.2.4 Score per residue for model 4 (medoid)

• Molecule 1: Erythroid transcription factor

Chain A: 69%   23%   5%   •

E200 A201 R202 E203 N206 C207 G208 A209 T210 D218 H222 C225 N226 A227 C228 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B: 70%   18%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 C94 A95 R102 E105 H116 L117 E118 C119 F120 K121 Q126 S140 D141 I142 V143 C144 E145 Q146 D147 I148 S158 S161 D167 T175 L176 M177 G178 G179
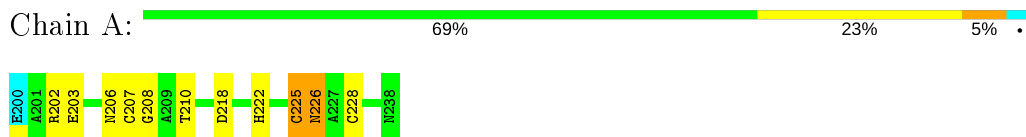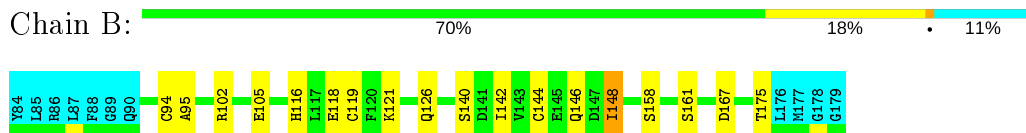
4.2.5 Score per residue for model 5

• Molecule 1: Erythroid transcription factor

Chain A:

77%   15%   5%   •

E200 E203 N206 C207 G208 H222 C225 N226 A227 C228 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

69%   18%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 D98 E105 V110 V114 Y115 H116 C119 A123 Q126 D133 R134 L137 D141 I142 V143 C144 E145 Q146 D147 I148 D167 V171 L176 M177 G178 G179
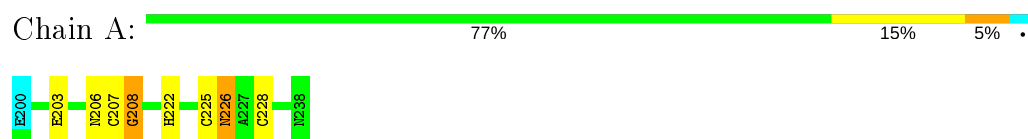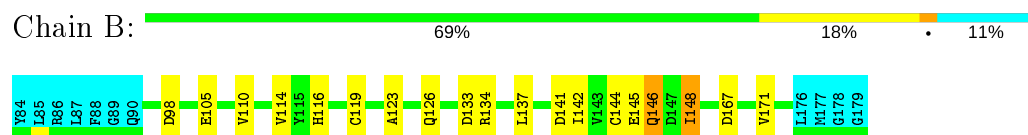
4.2.6 Score per residue for model 6

• Molecule 1: Erythroid transcription factor

Chain A:

74%   18%   5%   •

E200 A201 R202 E203 C207 G208 W215 G221 H222 C225 N226 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

68%   21%   11%

Y84 L85 R86 L87 F88 G89 Q90 D91 G92 L93 C94 A95 I101 E105 R109 V110 H116 C119 S130 D133 I138 N139 S140 D141 I142 V143 C144 I148 S161 V171 T175 L176 M177 G178 G179
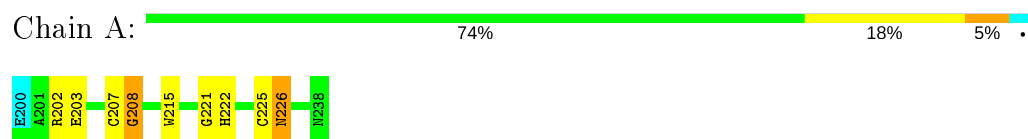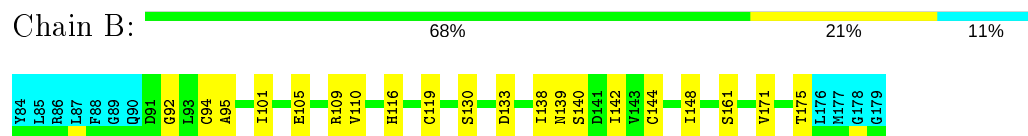
4.2.7 Score per residue for model 7

• Molecule 1: Erythroid transcription factor

Chain A:

79%   10%   8%   •

E200 N206 C207 G208 C225 N226 A227 C228 N235 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

69%   17%   •   11%

Y84 L85 R86 L87 F88 G89 Q90 V114 Y115 H116 L117 E118 C119 C125 Q126 K127 R134 L137 S140 D141 I142 V143 C144 E145 Q146 D147 I148 T152 N155 G156 G157 S164 V171 L176 M177 G178 G179
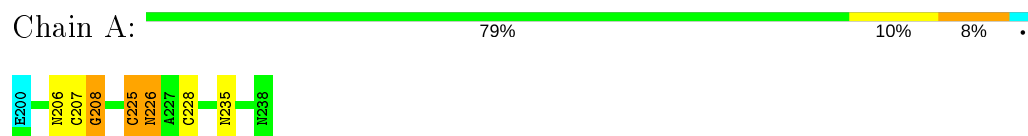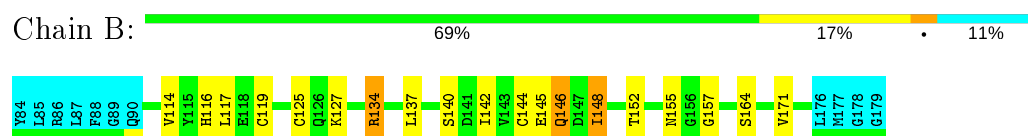
4.2.8 Score per residue for model 8

• Molecule 1: Erythroid transcription factor

Chain A:

77% 13% 8% •

E200 A201 R202 E203 C207 G208 D218 H222 C225 N226 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

67% 20% • 11%

Y84 L85 R86 L87 F88 G89 Q90 S96 C97 D98 E105 V110 H116 C119 F120 K121 R134 L137 I138 M139 S140 D141 I142 V143 C144 E145 Q146 D147 I148 Y149 M155 S158 D167 L176 M177 G178 G179
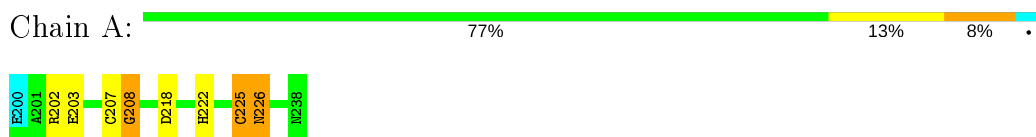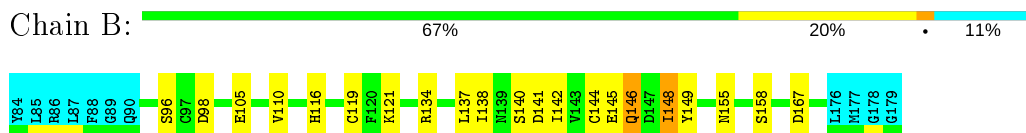
4.2.9 Score per residue for model 9

• Molecule 1: Erythroid transcription factor

Chain A:

67% 28% • •

E200 C204 V205 N206 C207 G208 H222 C225 N226 Y231 H232 N235 G236 Q237 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

71% 17% • 11%

Y84 L85 R86 L87 F88 G89 Q90 R102 E105 H116 L117 E118 C119 S130 D133 L137 I142 V143 C144 I148 G157 S158 S161 V171 T175 L176 M177 G178 G179
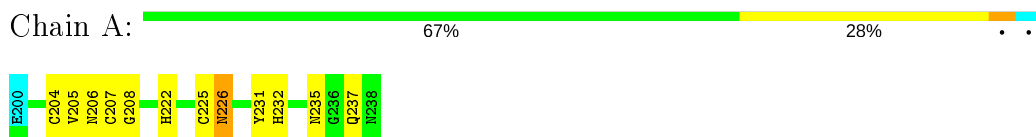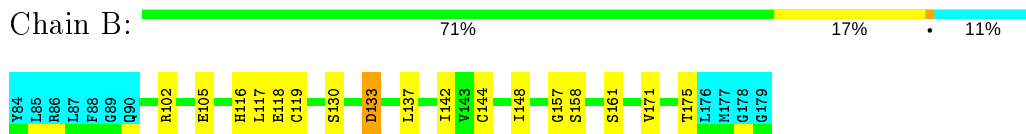
4.2.10 Score per residue for model 10

• Molecule 1: Erythroid transcription factor

Chain A:

67% 28% • •

E200 A201 R202 E203 C204 V205 N206 C207 G208 R217 H222 C225 N226 A227 C228 G229 L230 N238

• Molecule 2: LIM domain only 2, linker, LIM domain-binding protein 1

Chain B:

68% 20% • 11%

Y84 L85 R86 L87 F88 G89 Q90 E105 R109 V110 K113 V114 Y115 H116 C119 R134 I138 M139 S140 D141 I142 V143 C144 E145 Q146 D147 I148 T152 S158 S164 G165 T175 L176 M177 G178 G179
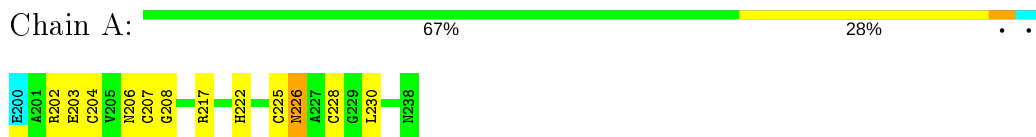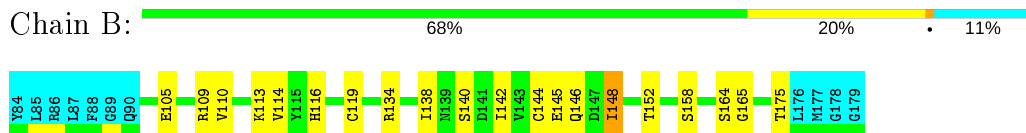
5 Refinement protocol and experimental data overview (i)

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

Of the 10 calculated structures, 10 were deposited, based on the following criterion: *all calculated structures submitted*.

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

Software name	Classification	Version
CNS	refinement	

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	297	278	276	4±1
2	B	646	619	618	4±1
All	All	9460	8970	8940	80

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	Models Total
2:B:144:CYS:O	2:B:148:ILE:HB	0.65	1.91	2	10
2:B:130:SER:HB3	2:B:133:ASP:OD2	0.61	1.95	2	3
2:B:138:ILE:HG22	2:B:139:ASN:H	0.60	1.56	6	1
2:B:116:HIS:HB2	2:B:119:CYS:SG	0.60	2.37	6	10
1:A:207:CYS:SG	1:A:208:GLY:N	0.59	2.75	6	10
1:A:225:CYS:SG	1:A:226:ASN:N	0.54	2.79	8	10
1:A:232:HIS:CD2	1:A:238:ASN:HA	0.50	2.41	3	1
1:A:206:ASN:HB3	1:A:228:CYS:SG	0.50	2.46	7	4
2:B:92:GLY:H	2:B:101:ILE:HD12	0.48	1.69	1	2
1:A:204:CYS:SG	1:A:205:VAL:N	0.48	2.87	9	2
2:B:134:ARG:HG2	2:B:145:GLU:OE1	0.48	2.09	8	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
2:B:102:ARG:NE	2:B:102:ARG:HA	0.47	2.23	3	2
1:A:202:ARG:NE	1:A:202:ARG:HA	0.47	2.24	10	2
1:A:204:CYS:HB3	1:A:208:GLY:N	0.47	2.25	10	2
1:A:210:THR:HA	1:A:215:TRP:CZ2	0.46	2.45	3	1
2:B:94:CYS:SG	2:B:95:ALA:N	0.46	2.88	6	2
1:A:226:ASN:O	1:A:230:LEU:HG	0.46	2.11	10	1
1:A:232:HIS:HA	1:A:237:GLN:O	0.46	2.11	9	2
2:B:116:HIS:HB3	2:B:118:GLU:OE1	0.45	2.11	9	1
1:A:202:ARG:HD3	1:A:203:GLU:OE1	0.45	2.11	1	1
2:B:125:CYS:SG	2:B:127:LYS:HG2	0.45	2.52	7	1
2:B:144:CYS:SG	2:B:146:GLN:HB3	0.45	2.52	5	3
1:A:202:ARG:HB3	1:A:215:TRP:CE2	0.44	2.47	6	1
2:B:134:ARG:HD2	2:B:145:GLU:OE1	0.42	2.14	7	1
2:B:123:ALA:HB2	2:B:141:ASP:HB2	0.42	1.91	5	1
2:B:133:ASP:O	2:B:171:VAL:HG22	0.41	2.15	5	1
2:B:107:THR:O	2:B:174:PRO:HA	0.41	2.15	2	1
1:A:231:TYR:O	1:A:235:ASN:HB2	0.41	2.15	9	1
2:B:116:HIS:H	2:B:119:CYS:HB2	0.40	1.77	4	2

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	37/39 (95%)	30±2 (82±5%)	5±2 (13±6%)	2±1 (6±2%)	3	22
2	B	85/96 (89%)	72±1 (84±2%)	12±2 (14±3%)	1±1 (2±1%)	13	57
All	All	1220/1350 (90%)	1018 (83%)	167 (14%)	35 (3%)	7	41

All 10 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	226	ASN	10
1	A	208	GLY	6
2	B	110	VAL	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	157	GLY	4
1	A	225	CYS	4
2	B	138	ILE	2
2	B	144	CYS	1
2	B	156	GLY	1
1	A	221	GLY	1
2	B	165	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	30/31 (97%)	27±1 (91±4%)	3±1 (9±4%)	13	60
2	B	70/78 (90%)	58±2 (83±3%)	12±2 (17±3%)	5	40
All	All	1000/1090 (92%)	854 (85%)	146 (15%)	6	45

All 40 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)
2	B	142	ILE	10
2	B	105	GLU	9
1	A	222	HIS	9
1	A	203	GLU	8
2	B	146	GLN	7
2	B	140	SER	7
2	B	114	VAL	6
2	B	148	ILE	6
2	B	175	THR	6
2	B	137	LEU	5
2	B	161	SER	5
2	B	134	ARG	5
2	B	121	LYS	4
2	B	110	VAL	4
2	B	145	GLU	4
2	B	164	SER	4
2	B	158	SER	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	117	LEU	3
2	B	141	ASP	3
1	A	202	ARG	3
2	B	171	VAL	3
2	B	167	ASP	3
2	B	98	ASP	3
1	A	218	ASP	2
1	A	235	ASN	2
2	B	118	GLU	2
2	B	152	THR	2
2	B	96	SER	2
2	B	155	ASN	2
2	B	109	ARG	2
2	B	126	GLN	2
2	B	133	ASP	1
2	B	138	ILE	1
1	A	217	ARG	1
2	B	102	ARG	1
2	B	93	LEU	1
1	A	210	THR	1
1	A	206	ASN	1
2	B	149	TYR	1
2	B	113	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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