Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 11:43 AM EDT

PDB ID	:	2HGL
Title	:	NMR structure of the first 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

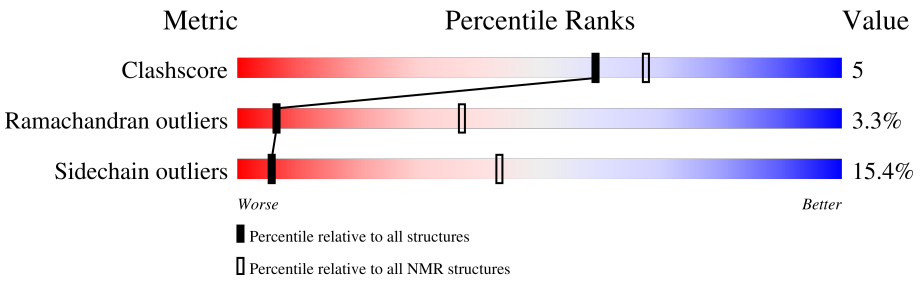
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	136	41% 15% • 18% 25%

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 2 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:11-A:47, A:54-A:77, A:81-A:97 (78)	0.49	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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1571 atoms, of which 768 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	102	1571	502	768	141	153	7	0

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-33	MET	-	cloning artifact	UNP Q5T0N2
A	-32	GLY	-	cloning artifact	UNP Q5T0N2
A	-31	SER	-	cloning artifact	UNP Q5T0N2
A	-30	SER	-	cloning artifact	UNP Q5T0N2
A	-29	HIS	-	expression tag	UNP Q5T0N2
A	-28	HIS	-	expression tag	UNP Q5T0N2
A	-27	HIS	-	expression tag	UNP Q5T0N2
A	-26	HIS	-	expression tag	UNP Q5T0N2
A	-25	HIS	-	expression tag	UNP Q5T0N2
A	-24	HIS	-	expression tag	UNP Q5T0N2
A	-23	SER	-	cloning artifact	UNP Q5T0N2
A	-22	SER	-	cloning artifact	UNP Q5T0N2
A	-21	GLY	-	cloning artifact	UNP Q5T0N2
A	-20	LEU	-	cloning artifact	UNP Q5T0N2
A	-19	VAL	-	cloning artifact	UNP Q5T0N2
A	-18	PRO	-	cloning artifact	UNP Q5T0N2
A	-17	ARG	-	cloning artifact	UNP Q5T0N2
A	-16	GLY	-	cloning artifact	UNP Q5T0N2
A	-15	SER	-	cloning artifact	UNP Q5T0N2
A	-14	HIS	-	cloning artifact	UNP Q5T0N2
A	-13	MET	-	cloning artifact	UNP Q5T0N2
A	-12	ALA	-	cloning artifact	UNP Q5T0N2
A	-11	SER	-	cloning artifact	UNP Q5T0N2
A	-10	MET	-	cloning artifact	UNP Q5T0N2
A	-9	THR	-	cloning artifact	UNP Q5T0N2
A	-8	GLY	-	cloning artifact	UNP Q5T0N2
A	-7	GLY	-	cloning artifact	UNP Q5T0N2
A	-6	GLN	-	cloning artifact	UNP Q5T0N2
A	-5	GLN	-	cloning artifact	UNP Q5T0N2
A	-4	MET	-	cloning artifact	UNP Q5T0N2
A	-3	GLY	-	cloning artifact	UNP Q5T0N2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	ARG	-	cloning artifact	UNP Q5T0N2
A	-1	GLY	-	cloning artifact	UNP Q5T0N2
A	0	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: 41% 15% • 18% 25%

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 M1 M2 L3 G4 P5 E6 G7 G8 E9 G10 K14 L15 R16 G17 L18 V24 V27 D33 C34 T35 I36 H37

D38 V43 I46 Y47 T48 R49 E50 G51 R52 Q53 F58 E64 V67 D74 R75 E76 S77 M78 G79 H80 R81 R90 W95 V96 L97 K98 H99 S100 G101 P102

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: 39% 14% • • 18% 25%

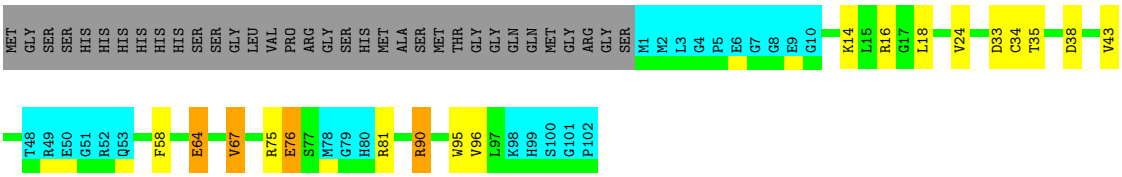
MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER M1 M2 L3 G4 P5 E6 G7 G8 E9 G10 K14 L15 R16 G17 L18 P19 W20 V24 V27 D33 C34 T35

I36 H37 D38 V43 T48 R49 E50 G51 R52 Q53 F58 V59 E60 L61 E64 V67 D74 R75 E76 S77 M78 G79 H80 R81 K87 R90 T91 W95 V96 L97 K98 H99 S100 G101 P102

4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: Heterogeneous nuclear ribonucleoprotein F

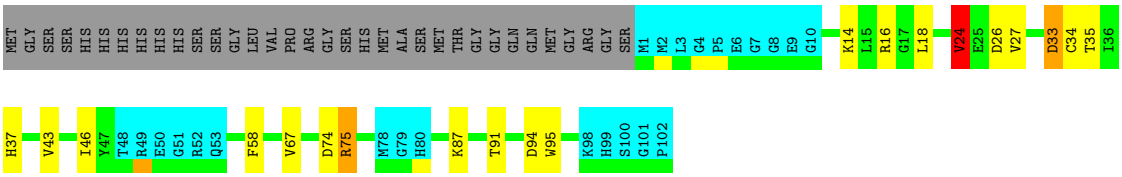
Chain A: 44% 10% • 18% 25%



4.2.3 Score per residue for model 3

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

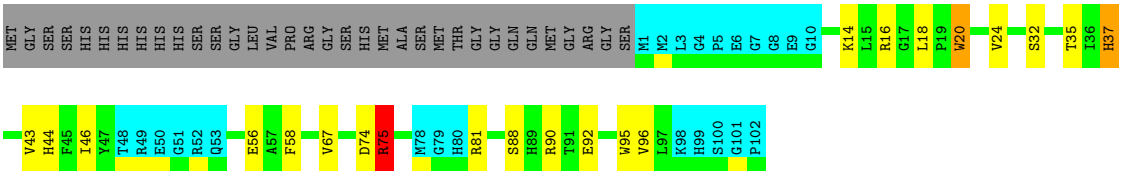
Chain A: 43% 12% •• 18% 25%



4.2.4 Score per residue for model 4

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

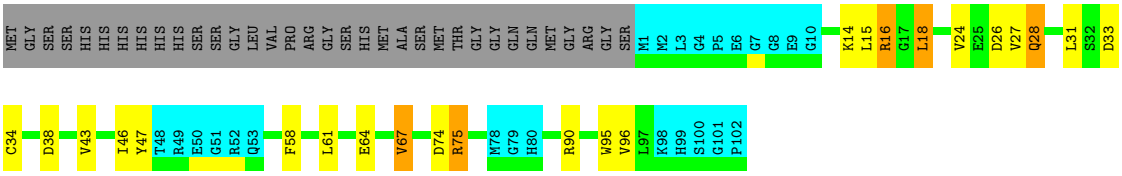
Chain A: 41% 14% •• 18% 25%



4.2.5 Score per residue for model 5

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 40% 14% • 18% 25%



4.2.6 Score per residue for model 6

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 39% 16% • 18% 25%

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 M1 M2 L3 G4 P5 E6 G7 G8 E9 G10 K14 L15 R16 G17 L18 S21 C22 S23 V24 E25 D26 V27 Q28

D33 C34 T35 D38 V43 I46 Y47 T48 R49 E50 G51 R52 Q53 F58 V59 E60 V67 R75 M78 G79 H80 R81 R90 T91 D94 W95 K98 H99 S100 G101 P102

4.2.7 Score per residue for model 7

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 39% 15% • • 18% 25%

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 M1 M2 L3 G4 P5 E6 G7 G8 E9 G10 K14 L15 R16 G17 L18 V24 V27 Q28 N29 F30 D33 C34

T35 D38 V43 H44 F45 I46 Y47 T48 R49 E50 G51 R52 Q53 F58 V59 E60 E64 V67 R75 E76 S77 M78 G79 H80 R81 Y82 I83 S88 H89 R90 W95 K98 H99 S100 G101 P102

4.2.8 Score per residue for model 8

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 39% 15% • 18% 25%

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 M1 M2 L3 G4 P5 E6 G7 G8 E9 G10 F11 K14 L15 R16 G17 L18 P19 W20 D33 C34 T35 I36 H37

V43 Y47 T48 R49 E50 G51 R52 Q53 E56 A57 F58 L61 E64 V67 K68 D74 R75 M78 G79 H80 R81 F86 K87 R90 W95 V96 L97 K98 H99 S100 G101 P102

4.2.9 Score per residue for model 9

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 40% 13% • 18% 25%

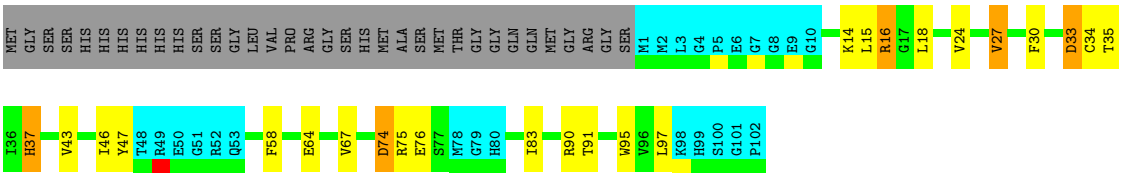
MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET ALA SER MET THR GLY GLY GLN GLN MET GLY ARG GLY SER M1 M2 L3 G4 P5 E6 G7 G8 E9 G10 K14 L15 R16 W20 V24 V27 F30 L31 S32 D33

I36 H37 D38 V43 I46 Y47 T48 R49 E50 G51 R52 Q53 F58 L61 E64 V67 R75 E76 S77 M78 G79 H80 I83 R90 W95 V96 L97 K98 H99 S100 G101 P102

4.2.10 Score per residue for model 10

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

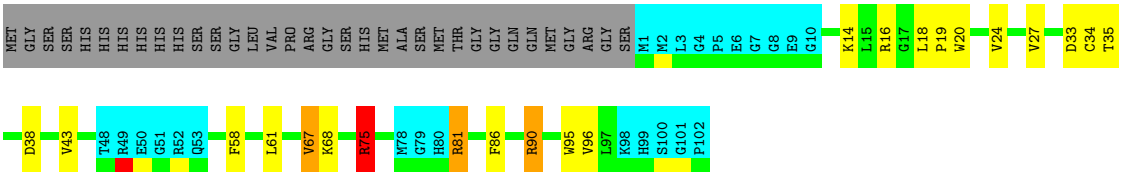
Chain A: 39% 15% • 18% 25%



4.2.11 Score per residue for model 11

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

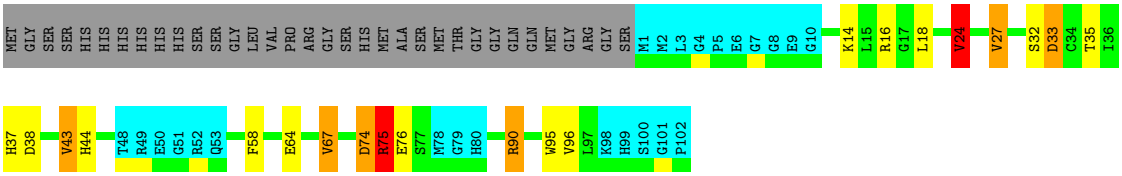
Chain A: 41% 13% • • 18% 25%



4.2.12 Score per residue for model 12

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

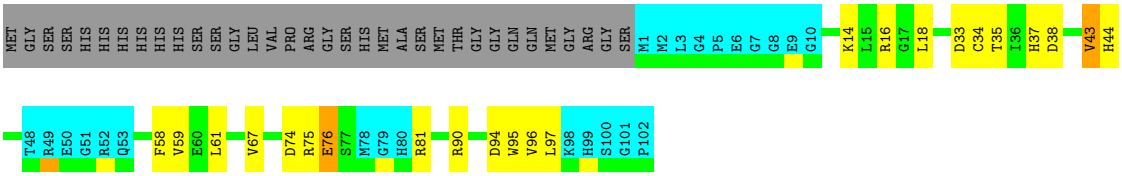
Chain A: 42% 10% • • 18% 25%



4.2.13 Score per residue for model 13

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

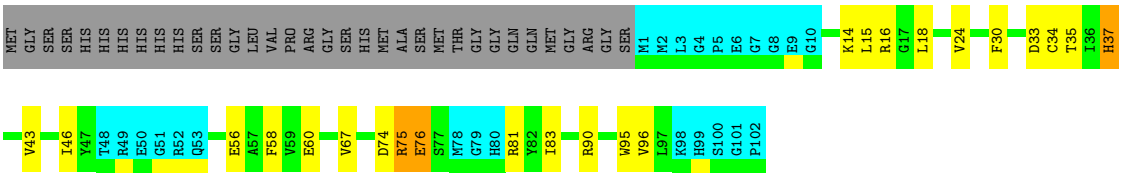
Chain A: 40% 15% • 18% 25%



4.2.14 Score per residue for model 14

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

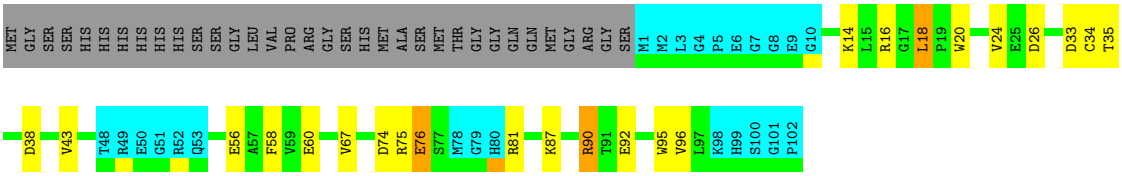
Chain A: 40% 15% • 18% 25%



4.2.15 Score per residue for model 15

- Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A: 40% 15% • 18% 25%



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±0/643 (0.0± 0.0%)	1.37±0.03	7±1/868 (0.8± 0.1%)
All	All	0.73	0/9645 (0.0%)	1.37	108/13020 (0.8%)

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	1.8±0.7
All	All	0	27

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	90	ARG	NE-CZ-NH1	10.09	125.34	120.30	1	14
1	A	75	ARG	NE-CZ-NH1	8.51	124.55	120.30	13	13
1	A	67	VAL	CA-CB-CG2	7.64	122.37	110.90	5	14
1	A	27	VAL	CA-CB-CG2	7.49	122.13	110.90	1	3
1	A	24	VAL	CA-CB-CG1	7.05	121.48	110.90	1	4
1	A	37	HIS	CA-CB-CG	7.03	125.55	113.60	8	4
1	A	81	ARG	NE-CZ-NH1	6.93	123.77	120.30	13	8
1	A	16	ARG	NE-CZ-NH2	-6.83	116.89	120.30	10	4
1	A	67	VAL	CG1-CB-CG2	6.80	121.78	110.90	9	13
1	A	90	ARG	NE-CZ-NH2	-6.72	116.94	120.30	4	1
1	A	16	ARG	NE-CZ-NH1	6.70	123.65	120.30	10	15
1	A	43	VAL	CG1-CB-CG2	-6.42	100.62	110.90	8	3
1	A	76	GLU	CB-CA-C	5.97	122.34	110.40	13	4
1	A	27	VAL	CA-CB-CG1	5.67	119.40	110.90	10	1
1	A	11	PHE	CB-CG-CD1	-5.61	116.87	120.80	8	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	28	GLN	CA-CB-CG	5.51	125.52	113.40	5	2
1	A	20	TRP	CB-CG-CD2	5.41	133.63	126.60	4	1
1	A	16	ARG	CD-NE-CZ	5.22	130.91	123.60	10	1
1	A	43	VAL	CA-CB-CG1	5.20	118.70	110.90	9	1
1	A	31	LEU	CA-CB-CG	-5.04	103.70	115.30	5	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	33	ASP	Peptide	13
1	A	75	ARG	Sidechain	4
1	A	46	ILE	Peptide	2
1	A	28	GLN	Peptide	2
1	A	90	ARG	Sidechain,Peptide	2
1	A	21	SER	Peptide	1
1	A	47	TYR	Peptide	1
1	A	76	GLU	Peptide	1
1	A	97	LEU	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	628	604	604	6±2
All	All	9420	9060	9060	89

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:24:VAL:HG22	1:A:43:VAL:CG2	0.64	2.22	9	8
1:A:24:VAL:HG22	1:A:43:VAL:HG21	0.62	1.70	6	8

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:58:PHE:CZ	1:A:96:VAL:HG21	0.60	2.31	12	9
1:A:58:PHE:CE1	1:A:96:VAL:HG21	0.60	2.32	8	8
1:A:61:LEU:HD13	1:A:67:VAL:HA	0.58	1.75	9	6
1:A:43:VAL:HG12	1:A:59:VAL:HG22	0.57	1.75	13	1
1:A:24:VAL:HG22	1:A:43:VAL:HG11	0.57	1.75	5	1
1:A:24:VAL:HB	1:A:43:VAL:HG21	0.55	1.77	1	4
1:A:14:LYS:HB2	1:A:58:PHE:CE2	0.53	2.39	15	15
1:A:24:VAL:HG22	1:A:43:VAL:CG1	0.52	2.34	5	1
1:A:24:VAL:HG13	1:A:43:VAL:HG21	0.49	1.83	14	6
1:A:24:VAL:HG13	1:A:43:VAL:HG11	0.49	1.84	5	1
1:A:24:VAL:HA	1:A:27:VAL:HG13	0.48	1.86	12	3
1:A:30:PHE:CE2	1:A:83:ILE:HG21	0.48	2.43	9	3
1:A:64:GLU:HA	1:A:67:VAL:HG13	0.47	1.86	5	3
1:A:24:VAL:CG2	1:A:43:VAL:HG11	0.45	2.42	12	3
1:A:24:VAL:HA	1:A:27:VAL:HG12	0.43	1.89	11	2
1:A:46:ILE:HG22	1:A:47:TYR:H	0.43	1.74	9	1
1:A:30:PHE:CD2	1:A:83:ILE:HG21	0.42	2.48	14	1
1:A:24:VAL:CB	1:A:43:VAL:HG21	0.41	2.45	3	1
1:A:14:LYS:HB3	1:A:86:PHE:HB2	0.41	1.93	8	2
1:A:24:VAL:HB	1:A:43:VAL:HG11	0.41	1.93	1	1
1:A:44:HIS:HB3	1:A:97:LEU:HD21	0.40	1.93	13	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	78/136 (57%)	63±2 (81±3%)	13±2 (16±3%)	3±1 (3±2%)	6 37
All	All	1170/2040 (57%)	943 (81%)	188 (16%)	39 (3%)	6 37

All 11 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	38	ASP	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	74	ASP	9
1	A	60	GLU	5
1	A	76	GLU	3
1	A	32	SER	3
1	A	18	LEU	3
1	A	36	ILE	2
1	A	37	HIS	2
1	A	20	TRP	1
1	A	19	PRO	1
1	A	90	ARG	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	69/112 (62%)	58±2 (85±3%)	11±2 (15±3%)	6 43
All	All	1035/1680 (62%)	876 (85%)	159 (15%)	6 43

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	95	TRP	15
1	A	18	LEU	14
1	A	34	CYS	12
1	A	35	THR	12
1	A	75	ARG	11
1	A	33	ASP	6
1	A	64	GLU	6
1	A	76	GLU	6
1	A	37	HIS	6
1	A	46	ILE	6
1	A	20	TRP	5
1	A	24	VAL	4
1	A	81	ARG	4
1	A	87	LYS	4
1	A	91	THR	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	26	ASP	4
1	A	27	VAL	4
1	A	56	GLU	4
1	A	15	LEU	4
1	A	43	VAL	3
1	A	74	ASP	3
1	A	90	ARG	3
1	A	94	ASP	3
1	A	44	HIS	3
1	A	88	SER	2
1	A	92	GLU	2
1	A	16	ARG	2
1	A	47	TYR	2
1	A	68	LYS	2
1	A	22	CYS	1
1	A	29	ASN	1
1	A	38	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