Full wwPDB NMR Structure Validation Report (i)

Mar 3, 2022 – 07:15 AM EST

PDB ID	:	2HH3
Title	:	Solution structure of the third KH domain of KSRP
Authors	:	Garcia-Mayoral, M.F.
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:

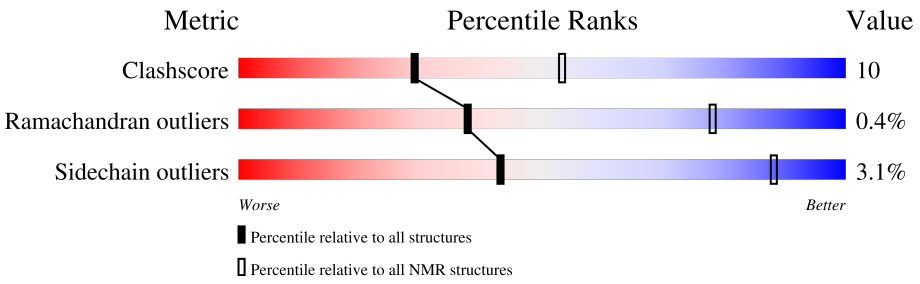
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	106	41% 14% • 11% 33%

2 Ensemble composition and analysis (i)

This entry contains 20 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:13-A:47, A:55-A:78 (59)	0.25	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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1101 atoms, of which 559 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called KH-type splicing regulatory protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	71	1101	335	559	105	99	3	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP Q59EZ9
A	2	ALA	-	cloning artifact	UNP Q59EZ9
A	3	MET	-	cloning artifact	UNP Q59EZ9
A	4	ALA	-	cloning artifact	UNP Q59EZ9
A	5	THR	-	cloning artifact	UNP Q59EZ9

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: KH-type splicing regulatory protein

Chain A: 41% 14% • 11% 33%

GLY ALA MET ALA THR GLY SER ARG ILE GLY GLY G12 I13 D14 V15 I26 G30 I33 K34 A40 I44 Q45 F46 K47 Q48 D49 D50 G51 T52 G53 P54 E55 K56 I57 A58 H59 I60 P63 R66 I74 L78 Q79 S80 L81 R82 SER GLY PRO PRO GLY PRO

PRO GLY GLY PRO GLY MET PRO PRO GLY GLY ARG GLY ARG GLY ARG GLY GLN GLY

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: KH-type splicing regulatory protein

Chain A: 41% 14% • 11% 33%

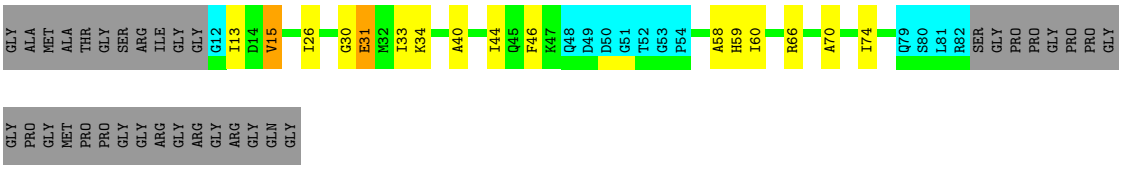
GLY ALA MET ALA THR GLY SER ARG ILE GLY GLY G12 I13 D14 V15 P18 R19 H20 S21 I26 G30 K34 D39 A40 Q48 D49 D50 G51 T52 G53 P54 E55 K56 I57 A58 H59 I60 P63 R66 Q79 S80 L81 R82 SER GLY PRO PRO GLY PRO PRO GLY

GLY PRO GLY MET PRO PRO GLY GLY ARG GLY ARG GLY ARG GLY GLN GLY

4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: KH-type splicing regulatory protein

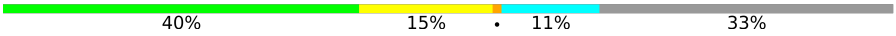
Chain A: 41% 13% • 11% 33%

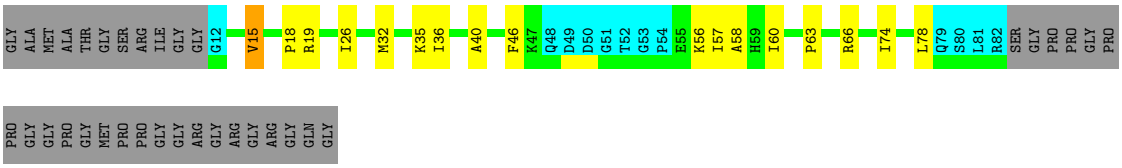


4.2.3 Score per residue for model 3

- Molecule 1: KH-type splicing regulatory protein

Chain A:

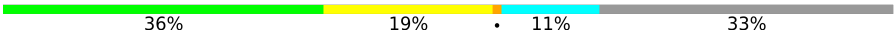


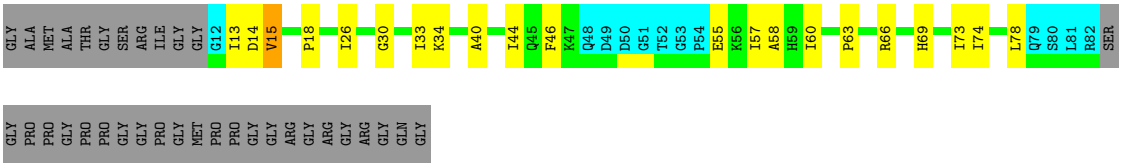


4.2.4 Score per residue for model 4

- Molecule 1: KH-type splicing regulatory protein

Chain A:

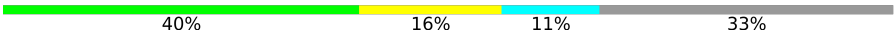


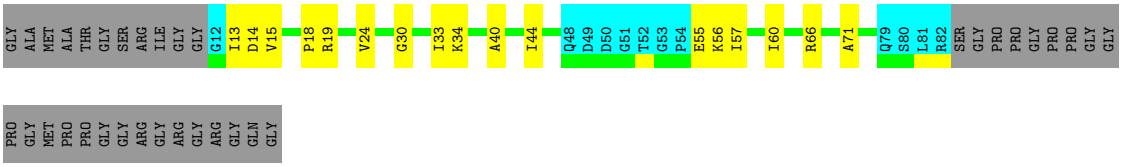


4.2.5 Score per residue for model 5

- Molecule 1: KH-type splicing regulatory protein

Chain A:





4.2.6 Score per residue for model 6

- Molecule 1: KH-type splicing regulatory protein

Chain A: 35% 20% • 11% 33%

GLY ALA MET ALA THR GLY SER ARG ILE GLY GLY G12 V15 P16 V17 P18 R19 V22 G23 V24 V25 I26 I33 A40 G41 V42 R43 I44 Q45 F46 K47 Q48 D49 D50 G51 T52 G53 P54 E55 K56 I57 A58 H59 I60 P63 R66 A70 I74 L78 Q79 S80

L81 R82 SER GLY PRO PRO GLY PRO PRO GLY GLY PRO GLY MET PRO PRO GLY GLY ARG GLY ARG GLY ARG GLY GLN GLY

4.2.7 Score per residue for model 7

- Molecule 1: KH-type splicing regulatory protein

Chain A: 42% 12% • 11% 33%

GLY ALA MET ALA THR GLY SER ARG ILE GLY GLY G12 I13 D14 V15 G30 E31 M32 I33 K34 K35 I36 A40 R43 I44 Q48 D49 D50 G51 T52 G53 P54 I60 P63 R66 I74 L78 Q79 S80 L81 R82 SER GLY PRO PRO GLY PRO PRO GLY GLY PRO

GLY MET PRO PRO GLY GLY ARG GLY ARG GLY ARG GLY GLN GLY

4.2.8 Score per residue for model 8

- Molecule 1: KH-type splicing regulatory protein

Chain A: 37% 17% • 11% 33%

GLY ALA MET ALA THR GLY SER ARG ILE GLY GLY G12 I13 D14 V15 R19 R28 S29 G30 E31 M32 I33 K34 K35 I36 I44 Q48 D49 D50 G51 T52 G53 P54 E55 K56 I57 A58 H59 I60 P63 R66 I73 I74 N75 D76 L77 L78 Q79 S80 L81 R82 SER GLY

PRO PRO GLY PRO PRO GLY GLY PRO GLY MET PRO PRO GLY GLY ARG GLY ARG GLY ARG GLY GLN GLY

4.2.9 Score per residue for model 9

- Molecule 1: KH-type splicing regulatory protein

Chain A: 41% 14% • 11% 33%

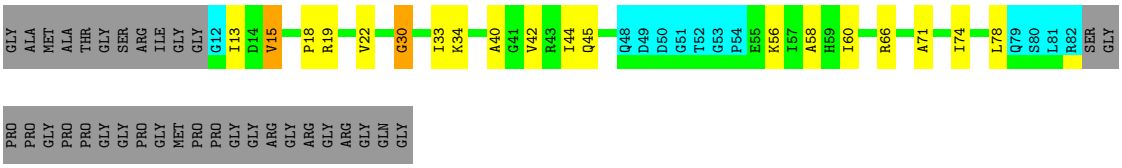
GLY ALA MET ALA THR GLY SER ARG ILE GLY GLY G12 I13 D14 V15 P16 V24 V25 I26 G30 K34 I44 Q45 F46 K47 Q48 D49 D50 G51 T52 G53 P54 E55 K56 I57 A58 H59 I60 A70 I74 Q79 S80 L81 R82 SER GLY PRO PRO GLY PRO PRO GLY GLY

PRO GLY MET PRO PRO GLY GLY ARG GLY ARG GLY GLN GLY

4.2.10 Score per residue for model 10

- Molecule 1: KH-type splicing regulatory protein

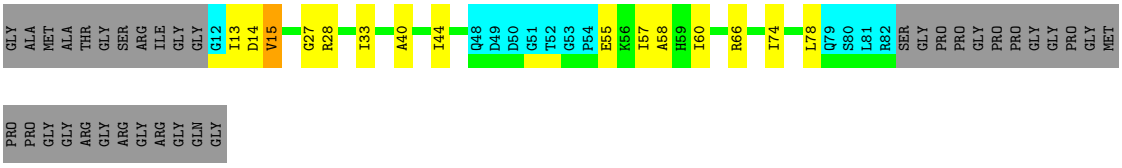
Chain A: 38% 16% • 11% 33%



4.2.11 Score per residue for model 11

- Molecule 1: KH-type splicing regulatory protein

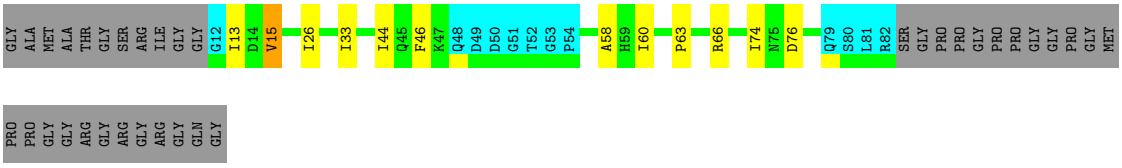
Chain A: 42% 13% • 11% 33%



4.2.12 Score per residue for model 12

- Molecule 1: KH-type splicing regulatory protein

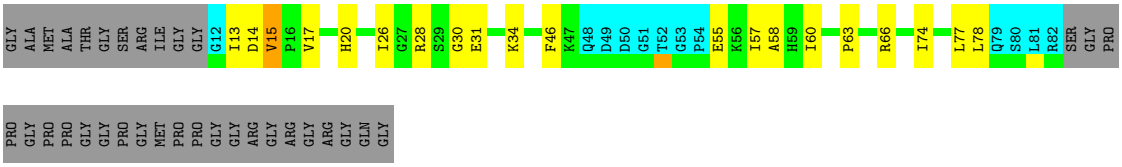
Chain A: 44% 10% • 11% 33%



4.2.13 Score per residue for model 13

- Molecule 1: KH-type splicing regulatory protein

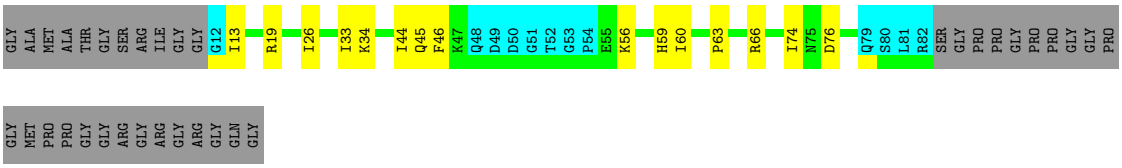
Chain A: 37% 18% • 11% 33%



4.2.14 Score per residue for model 14

- Molecule 1: KH-type splicing regulatory protein

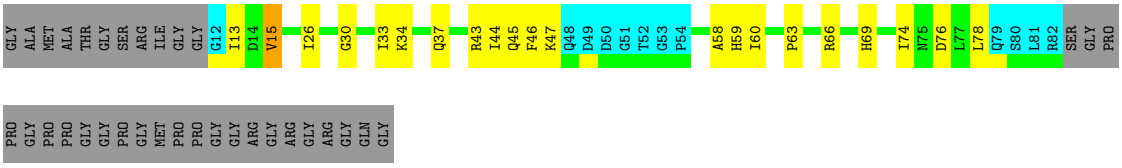
Chain A: 42% 14% 11% 33%



4.2.15 Score per residue for model 15

- Molecule 1: KH-type splicing regulatory protein

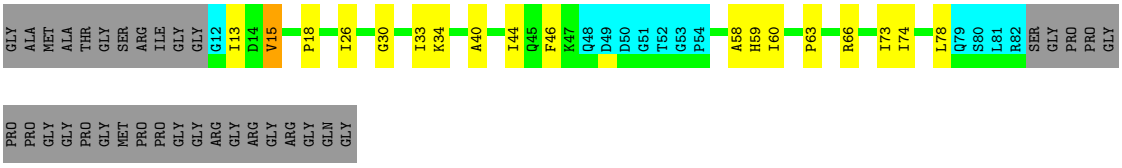
Chain A: 36% 19% • 11% 33%



4.2.16 Score per residue for model 16

- Molecule 1: KH-type splicing regulatory protein

Chain A: 39% 16% • 11% 33%



4.2.17 Score per residue for model 17

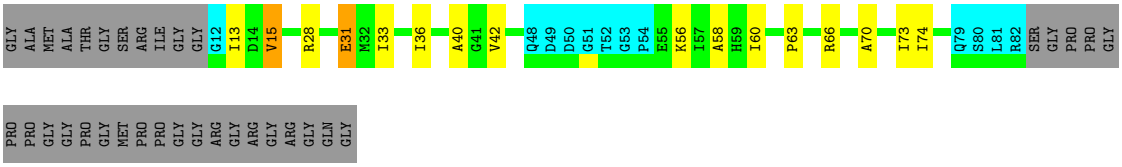
- Molecule 1: KH-type splicing regulatory protein

Chain A: 42% 12% • 11% 33%

4.2.18 Score per residue for model 18

- Molecule 1: KH-type splicing regulatory protein

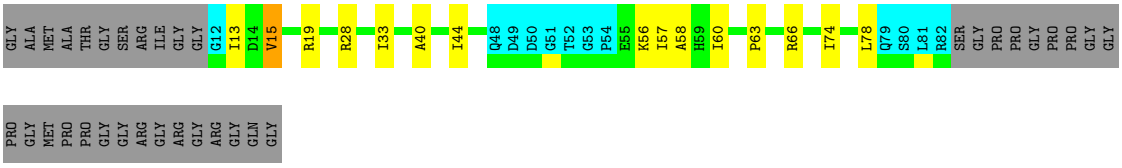
Chain A: 41% 13% • 11% 33%



4.2.19 Score per residue for model 19

- Molecule 1: KH-type splicing regulatory protein

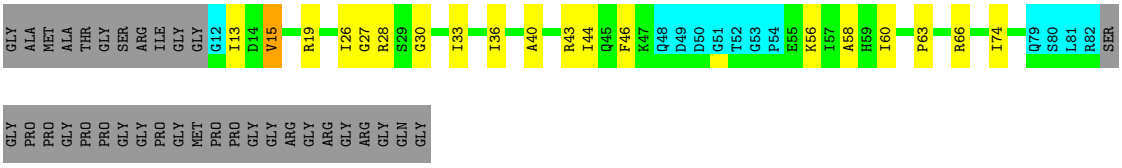
Chain A: 42% 13% • 11% 33%



4.2.20 Score per residue for model 20

- Molecule 1: KH-type splicing regulatory protein

Chain A: 38% 17% • 11% 33%



5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 20 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
ARIA	structure solution	1.2
ARIA	refinement	1.2

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	457	483	480	10±3
All	All	9140	9660	9600	193

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 10.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:13:ILE:HD11	1:A:60:ILE:HD12	0.77	1.55	8	18
1:A:19:ARG:HG2	1:A:56:LYS:HG3	0.66	1.68	1	9
1:A:15:VAL:HG23	1:A:58:ALA:HB3	0.64	1.70	18	11
1:A:40:ALA:O	1:A:66:ARG:HD2	0.64	1.91	1	14
1:A:18:PRO:HA	1:A:55:GLU:CD	0.60	2.17	5	1
1:A:63:PRO:HG2	1:A:66:ARG:CB	0.59	2.27	16	1
1:A:63:PRO:HG2	1:A:66:ARG:HB3	0.58	1.75	16	1
1:A:45:GLN:HB2	1:A:59:HIS:HB3	0.58	1.73	15	1
1:A:30:GLY:O	1:A:34:LYS:HG2	0.56	2.00	8	8
1:A:18:PRO:O	1:A:22:VAL:HG23	0.56	2.00	6	2
1:A:74:ILE:O	1:A:78:LEU:HG	0.55	2.01	13	11
1:A:30:GLY:O	1:A:34:LYS:HG3	0.54	2.01	17	4
1:A:19:ARG:CG	1:A:56:LYS:HG3	0.54	2.32	5	1
1:A:33:ILE:HG13	1:A:36:ILE:HD12	0.54	1.80	8	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:44:ILE:HD13	1:A:74:ILE:HD11	0.53	1.80	10	5
1:A:15:VAL:HG23	1:A:58:ALA:O	0.52	2.04	12	6
1:A:18:PRO:HA	1:A:55:GLU:HA	0.52	1.80	4	1
1:A:33:ILE:HA	1:A:36:ILE:HD12	0.52	1.82	8	4
1:A:43:ARG:O	1:A:60:ILE:HA	0.52	2.05	20	2
1:A:33:ILE:HD11	1:A:44:ILE:CG2	0.51	2.35	19	10
1:A:26:ILE:HD11	1:A:46:PHE:CE1	0.50	2.42	2	7
1:A:63:PRO:HD2	1:A:66:ARG:HB3	0.49	1.85	16	11
1:A:17:VAL:HG22	1:A:22:VAL:HG22	0.48	1.84	6	1
1:A:63:PRO:HB2	1:A:66:ARG:HB2	0.48	1.85	14	1
1:A:60:ILE:HD11	1:A:74:ILE:HD12	0.48	1.84	16	8
1:A:26:ILE:HD11	1:A:46:PHE:CE2	0.48	2.42	15	4
1:A:18:PRO:HD2	1:A:78:LEU:HD13	0.48	1.86	3	2
1:A:14:ASP:HB2	1:A:57:ILE:CG2	0.48	2.39	4	3
1:A:13:ILE:HD11	1:A:71:ALA:HB2	0.47	1.87	5	1
1:A:27:GLY:O	1:A:28:ARG:HD2	0.46	2.10	11	1
1:A:36:ILE:HG23	1:A:73:ILE:HG21	0.46	1.86	8	1
1:A:40:ALA:HB2	1:A:73:ILE:HD12	0.46	1.88	18	2
1:A:18:PRO:HA	1:A:55:GLU:HG2	0.46	1.88	1	1
1:A:18:PRO:HA	1:A:55:GLU:HG3	0.46	1.88	6	1
1:A:29:SER:HB2	1:A:31:GLU:OE1	0.46	2.11	8	1
1:A:37:GLN:OE1	1:A:43:ARG:HD2	0.45	2.10	15	1
1:A:60:ILE:HG21	1:A:70:ALA:HB3	0.45	1.88	9	3
1:A:63:PRO:CG	1:A:66:ARG:HB3	0.45	2.41	16	2
1:A:33:ILE:HD11	1:A:44:ILE:HG21	0.45	1.89	15	4
1:A:17:VAL:HG23	1:A:78:LEU:HD21	0.45	1.86	13	2
1:A:45:GLN:HB3	1:A:59:HIS:CE1	0.45	2.47	14	1
1:A:32:MET:O	1:A:36:ILE:HG13	0.45	2.12	7	2
1:A:42:VAL:HG11	1:A:70:ALA:HB2	0.44	1.90	18	1
1:A:28:ARG:O	1:A:31:GLU:HB2	0.44	2.13	18	1
1:A:45:GLN:HA	1:A:45:GLN:OE1	0.44	2.13	6	1
1:A:13:ILE:HG12	1:A:60:ILE:HB	0.44	1.89	8	2
1:A:26:ILE:HG23	1:A:30:GLY:HA2	0.43	1.90	1	1
1:A:14:ASP:HB3	1:A:57:ILE:CG2	0.43	2.43	13	1
1:A:42:VAL:CG1	1:A:66:ARG:HG3	0.43	2.44	6	2
1:A:42:VAL:HG12	1:A:66:ARG:HG3	0.43	1.90	10	1
1:A:31:GLU:H	1:A:31:GLU:CD	0.42	2.18	2	1
1:A:30:GLY:O	1:A:33:ILE:HG22	0.42	2.14	2	1
1:A:46:PHE:HB3	1:A:56:LYS:HE3	0.42	1.92	9	1
1:A:44:ILE:HG12	1:A:60:ILE:HG12	0.42	1.91	12	1
1:A:28:ARG:C	1:A:30:GLY:H	0.42	2.19	20	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:13:ILE:CD1	1:A:71:ALA:HB2	0.42	2.45	10	1
1:A:69:HIS:O	1:A:73:ILE:HG13	0.41	2.15	4	1
1:A:34:LYS:HE2	1:A:34:LYS:HA	0.41	1.92	14	1
1:A:63:PRO:HB2	1:A:66:ARG:HB3	0.40	1.92	20	1
1:A:35:LYS:C	1:A:35:LYS:HD3	0.40	2.36	3	1
1:A:33:ILE:HD11	1:A:44:ILE:HB	0.40	1.93	14	1
1:A:13:ILE:O	1:A:59:HIS:HA	0.40	2.16	16	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	59/106 (56%)	57±1 (96±2%)	2±1 (3±2%)	0±0 (0±1%)	38	78
All	All	1180/2120 (56%)	1134 (96%)	41 (3%)	5 (0%)	38	78

All 3 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	30	GLY	2
1	A	28	ARG	2
1	A	27	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	50/77 (65%)	48±1 (97±2%)	2±1 (3±2%)	43	88
All	All	1000/1540 (65%)	969 (97%)	31 (3%)	43	88

All 10 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	15	VAL	19
1	A	31	GLU	4
1	A	39	ASP	1
1	A	59	HIS	1
1	A	45	GLN	1
1	A	14	ASP	1
1	A	20	HIS	1
1	A	28	ARG	1
1	A	69	HIS	1
1	A	72	ARG	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