Full wwPDB NMR Structure Validation Report (i)

Mar 1, 2022 – 03:42 PM EST

PDB ID	:	2GQB
Title	:	Solution Structure of a conserved unknown protein RPA2825 from Rhodopseudomonas palustris; (Northeast Structural Genomics Consortium Target RpT4; Ontario Centre for Structural Proteomics Target rp2812)
Authors	:	Srisailam, S.; Lukin, J.A.; Yee, A.; Lemak, A.; Arrowsmith, C.H.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2006-04-20

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:

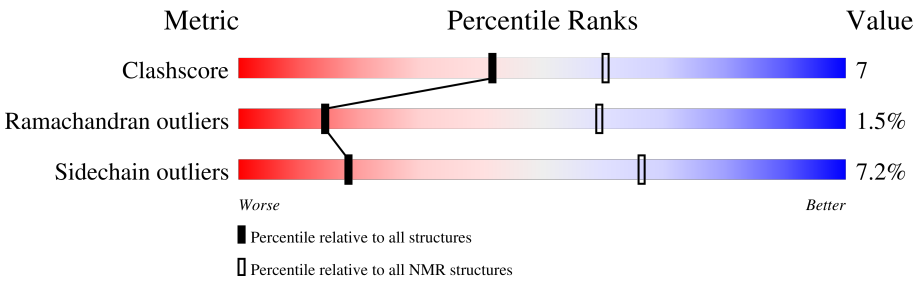
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	130	55% 10% • 34%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 14 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:2-A:13, A:46-A:95, A:103-A:126 (86)	0.36	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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1907 atoms, of which 967 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called conserved hypothetical protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	130	1907	591	967	160	183	6	0

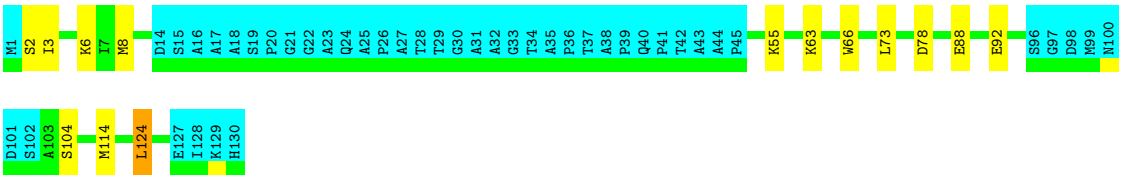
4 Residue-property plots (i)

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: conserved hypothetical protein

Chain A: 55% 10% • 34%



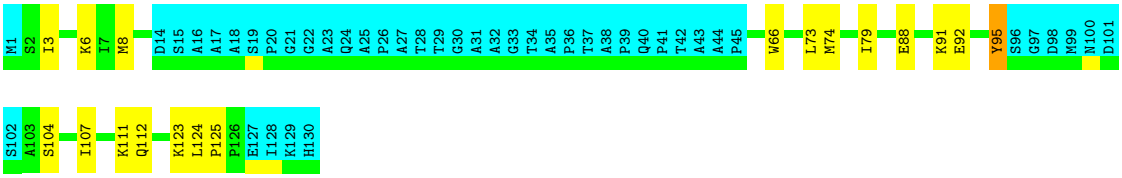
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: conserved hypothetical protein

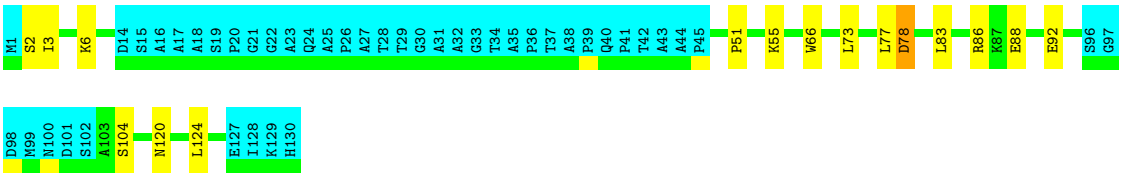
Chain A: 52% 13% • 34%



4.2.2 Score per residue for model 2

• Molecule 1: conserved hypothetical protein

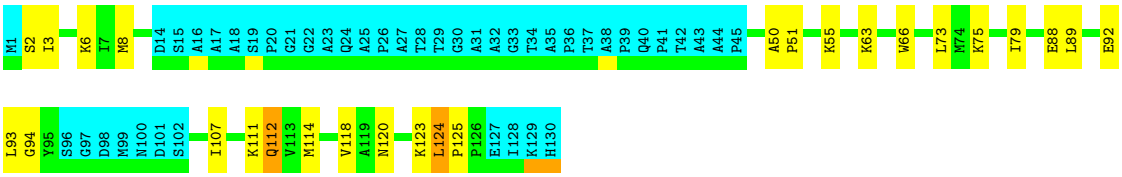
Chain A: 54% 12% • 34%



4.2.3 Score per residue for model 3

- Molecule 1: conserved hypothetical protein

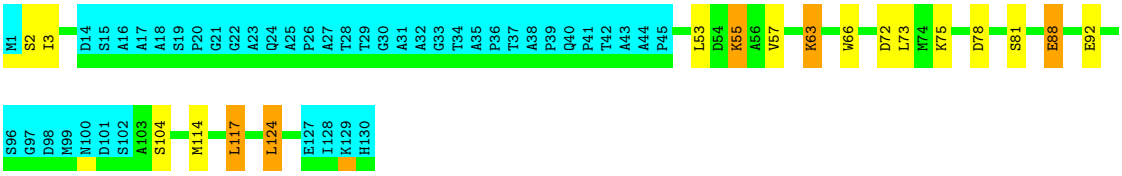
Chain A: 46% 18% • 34%



4.2.4 Score per residue for model 4

- Molecule 1: conserved hypothetical protein

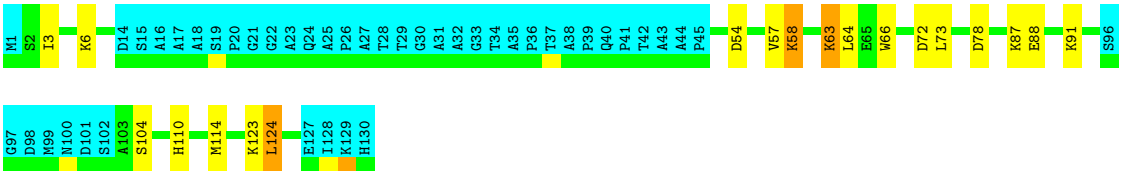
Chain A: 52% 10% • 34%



4.2.5 Score per residue for model 5

- Molecule 1: conserved hypothetical protein

Chain A: 52% 12% • 34%



4.2.6 Score per residue for model 6

- Molecule 1: conserved hypothetical protein

Chain A: 48% 16% • 34%

M1 S2 I3 K6 T7 M8 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 K55 K60 K63 W66 L73 M74 K75 A76 L77 D78 L83 E88 L89

A90 K91 E92 S96 G97 D98 M99 N100 D101 S102 A103 S104 M105 N106 H110 K111 M114 L124 E127 I128 K129 H130

4.2.7 Score per residue for model 7

- Molecule 1: conserved hypothetical protein

Chain A: 48% 15% • 34%

M1 M8 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 P51 I52 L53 D54 K55 A56 V57 K58 A59 K60 K63 W66 R67 L73 D78 S82 R86

K87 E88 E92 S96 G97 D98 M99 N100 D101 S102 A103 S104 M105 N106 I107 K111 Q112 V113 M114 K123 L124 E127 I128 K129 H130

4.2.8 Score per residue for model 8

- Molecule 1: conserved hypothetical protein

Chain A: 50% 13% • 34%

M1 M8 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 P51 I52 L53 D54 K55 A56 V57 K60 K63 W66 L73 L77 D78 S84 E88

K91 E92 L93 G94 Y95 S96 G97 D98 M99 N100 D101 S102 A103 S104 Q112 L124 E127 I128 K129 H130

4.2.9 Score per residue for model 9

- Molecule 1: conserved hypothetical protein

Chain A: 52% 13% • 34%

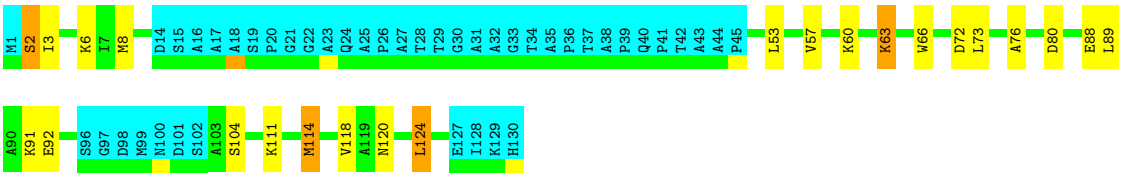
M1 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 P51 K55 K60 L64 E65 W66 I70 V71 D72 L73 M74 K75 D78 E88 E92 S96 G97

D98 M99 N100 D101 S102 A103 S104 M105 N106 I107 W108 K111 Q112 L117 L124 E127 I128 K129 H130

4.2.10 Score per residue for model 10

- Molecule 1: conserved hypothetical protein

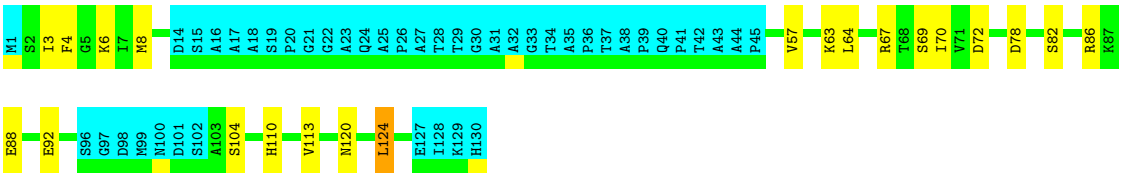
Chain A: 48% 15% • 34%



4.2.11 Score per residue for model 11

- Molecule 1: conserved hypothetical protein

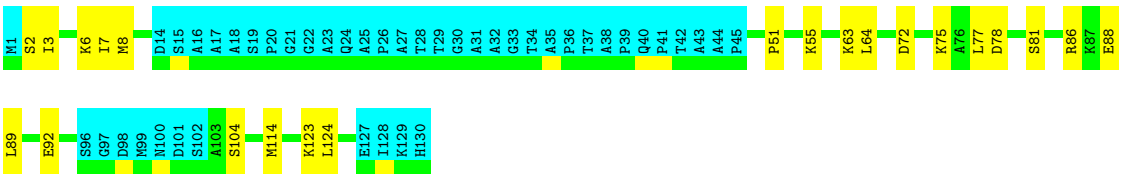
Chain A: 50% 15% • 34%



4.2.12 Score per residue for model 12

- Molecule 1: conserved hypothetical protein

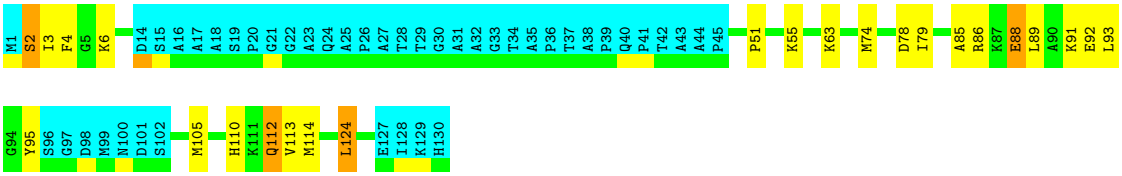
Chain A: 49% 17% 34%



4.2.13 Score per residue for model 13

- Molecule 1: conserved hypothetical protein

Chain A: 48% 15% • 34%



4.2.14 Score per residue for model 14 (medoid)

- Molecule 1: conserved hypothetical protein

Chain A: 55% 11% • 34%

M1 S2 I3 K6 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 W66 D72 L73 M74 K75 D78 E88 E92 S96 G97 D98 M99 N100 D101 S102 A103

S104 H110 K111 M114 N120 L124 E127 I128 K129 H130

4.2.15 Score per residue for model 15

- Molecule 1: conserved hypothetical protein

Chain A: 54% 9% • 34%

M1 S2 I3 K6 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 K55 A56 V57 K63 W66 R67 L73 M74 K75 D78 S81 Y95 S96 G97 D98

M99 N100 D101 S102 W108 L109 N120 L124 E127 I128 K129 H130

4.2.16 Score per residue for model 16

- Molecule 1: conserved hypothetical protein

Chain A: 52% 12% • 34%

M1 S2 I3 M8 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 K63 L64 E65 K75 A76 L77 D78 I79 D80 S81 R86 K87 E88 L89 A90 K91 Y95 S96

G97 D98 M99 N100 D101 S102 Q112 L117 N120 L124 E127 I128 K129 H130

4.2.17 Score per residue for model 17

- Molecule 1: conserved hypothetical protein

Chain A: 52% 12% • 34%

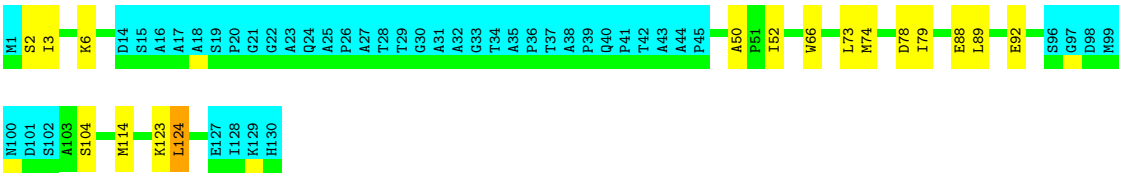
M1 S2 I3 K6 D14 S15 A16 A17 A18 S19 P20 G21 G22 A23 Q24 A25 P26 A27 T28 T29 G30 A31 A32 G33 T34 A35 P36 T37 A38 P39 Q40 P41 T42 A43 A44 P45 P51 I52 L53 D54 K55 A56 V57 K63 L64 A76 L77 D78 E88 E92 S96 G97

D98 M99 N100 D101 S102 A103 S104 H110 K111 M114 N120 L124 E127 I128 K129 H130

4.2.18 Score per residue for model 18

- Molecule 1: conserved hypothetical protein

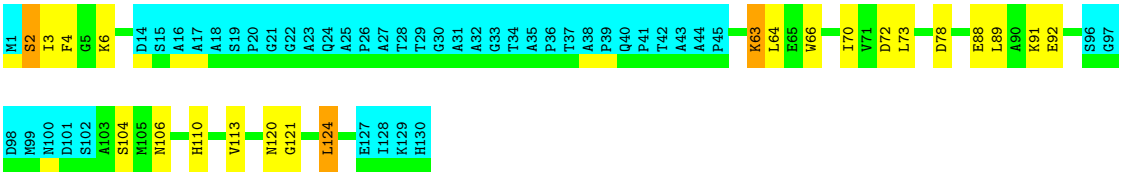
Chain A: 53% 12% • 34%



4.2.19 Score per residue for model 19

- Molecule 1: conserved hypothetical protein

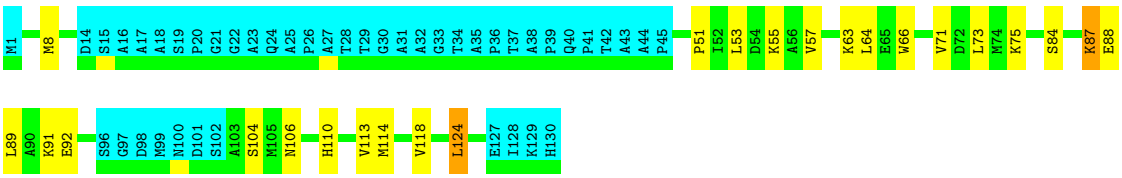
Chain A: 49% 15% • 34%



4.2.20 Score per residue for model 20

- Molecule 1: conserved hypothetical protein

Chain A: 48% 17% • 34%



5 Refinement protocol and experimental data overview (i)

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

Of the 100 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
CYANA	structure solution	2.1
CNS	refinement	1.1

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	656	703	702	10±2
All	All	13120	14060	14040	197

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 7.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:3:ILE:HA	1:A:6:LYS:HD3	0.86	1.46	5	11
1:A:66:TRP:HB2	1:A:73:LEU:HB3	0.73	1.60	8	10
1:A:88:GLU:O	1:A:92:GLU:HG2	0.70	1.87	12	17
1:A:69:SER:HB2	1:A:72:ASP:HB2	0.67	1.66	11	1
1:A:75:LYS:HB3	1:A:81:SER:HB3	0.67	1.65	4	4
1:A:111:LYS:HA	1:A:114:MET:HE2	0.64	1.68	17	2
1:A:84:SER:HA	1:A:87:LYS:HE3	0.64	1.68	20	1
1:A:51:PRO:O	1:A:55:LYS:HG2	0.60	1.96	13	5
1:A:64:LEU:HB3	1:A:72:ASP:OD1	0.60	1.95	19	3
1:A:110:HIS:O	1:A:114:MET:HG3	0.58	1.98	17	4
1:A:2:SER:HA	1:A:92:GLU:HG3	0.57	1.77	6	4
1:A:108:TRP:HE3	1:A:109:LEU:HD12	0.57	1.58	15	1
1:A:55:LYS:N	1:A:55:LYS:HE2	0.56	2.16	17	4
1:A:64:LEU:HD13	1:A:76:ALA:HB2	0.56	1.76	17	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:66:TRP:CB	1:A:73:LEU:HB3	0.54	2.33	14	10
1:A:72:ASP:O	1:A:75:LYS:HG2	0.54	2.03	12	1
1:A:70:ILE:HB	1:A:110:HIS:HA	0.53	1.80	11	2
1:A:75:LYS:HA	1:A:79:ILE:O	0.53	2.03	16	2
1:A:7:ILE:HB	1:A:77:LEU:HG	0.52	1.80	12	1
1:A:88:GLU:HG3	1:A:89:LEU:N	0.52	2.19	13	7
1:A:107:ILE:O	1:A:111:LYS:HG2	0.52	2.05	1	3
1:A:63:LYS:C	1:A:63:LYS:HD3	0.52	2.25	5	3
1:A:4:PHE:HB3	1:A:113:VAL:HG13	0.51	1.83	11	3
1:A:114:MET:O	1:A:118:VAL:HG22	0.50	2.06	10	3
1:A:123:LYS:HD3	1:A:124:LEU:N	0.50	2.22	3	1
1:A:87:LYS:O	1:A:91:LYS:HG3	0.50	2.05	5	1
1:A:69:SER:CB	1:A:72:ASP:HB2	0.50	2.36	11	1
1:A:106:ASN:O	1:A:110:HIS:HB2	0.49	2.07	6	3
1:A:3:ILE:HA	1:A:6:LYS:CD	0.49	2.37	17	1
1:A:63:LYS:HE2	1:A:64:LEU:O	0.49	2.08	19	1
1:A:74:MET:HG2	1:A:79:ILE:HB	0.48	1.84	1	3
1:A:70:ILE:HB	1:A:110:HIS:CA	0.48	2.39	11	1
1:A:91:LYS:HA	1:A:95:TYR:HB2	0.48	1.86	8	4
1:A:88:GLU:HA	1:A:91:LYS:HG2	0.48	1.83	6	2
1:A:60:LYS:HE2	1:A:75:LYS:HD3	0.48	1.84	9	1
1:A:77:LEU:O	1:A:78:ASP:HB2	0.47	2.09	2	2
1:A:3:ILE:HA	1:A:6:LYS:HD2	0.47	1.86	6	2
1:A:60:LYS:HD3	1:A:76:ALA:O	0.47	2.09	10	2
1:A:3:ILE:HG12	1:A:6:LYS:HD2	0.47	1.87	19	1
1:A:124:LEU:HD13	1:A:124:LEU:N	0.47	2.25	14	11
1:A:60:LYS:HA	1:A:60:LYS:HE3	0.47	1.85	7	1
1:A:57:VAL:HG13	1:A:64:LEU:HB2	0.47	1.86	11	1
1:A:53:LEU:O	1:A:57:VAL:HG23	0.46	2.11	4	6
1:A:93:LEU:HD13	1:A:112:GLN:HB3	0.46	1.86	3	2
1:A:114:MET:HA	1:A:117:LEU:HB2	0.46	1.87	4	1
1:A:60:LYS:HE2	1:A:75:LYS:O	0.46	2.11	6	1
1:A:75:LYS:HB3	1:A:81:SER:CB	0.45	2.41	12	1
1:A:3:ILE:HG13	1:A:88:GLU:OE2	0.45	2.12	4	1
1:A:64:LEU:HB2	1:A:72:ASP:CG	0.44	2.33	9	1
1:A:111:LYS:HG3	1:A:112:GLN:HE21	0.44	1.72	9	1
1:A:51:PRO:O	1:A:55:LYS:HB2	0.44	2.13	7	1
1:A:108:TRP:CE3	1:A:109:LEU:HD12	0.43	2.44	15	1
1:A:88:GLU:O	1:A:91:LYS:HG2	0.43	2.13	20	2
1:A:2:SER:O	1:A:6:LYS:HG3	0.43	2.12	18	2
1:A:51:PRO:O	1:A:55:LYS:HD2	0.43	2.13	8	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:57:VAL:HG22	1:A:64:LEU:HB2	0.43	1.90	20	1
1:A:54:ASP:O	1:A:57:VAL:HB	0.43	2.13	5	1
1:A:88:GLU:HA	1:A:91:LYS:HE2	0.43	1.90	5	2
1:A:70:ILE:HG22	1:A:106:ASN:OD1	0.43	2.14	9	1
1:A:124:LEU:HD13	1:A:124:LEU:H	0.43	1.74	10	1
1:A:108:TRP:O	1:A:112:GLN:HG2	0.42	2.14	9	1
1:A:64:LEU:HD13	1:A:72:ASP:OD1	0.42	2.14	12	1
1:A:120:ASN:HD22	1:A:120:ASN:N	0.42	2.12	15	1
1:A:50:ALA:HA	1:A:66:TRP:CH2	0.42	2.50	18	1
1:A:54:ASP:O	1:A:58:LYS:HD2	0.42	2.13	5	1
1:A:2:SER:CA	1:A:92:GLU:HG3	0.42	2.44	6	1
1:A:55:LYS:HE2	1:A:55:LYS:HA	0.42	1.90	15	1
1:A:3:ILE:HB	1:A:89:LEU:HD13	0.42	1.91	16	1
1:A:3:ILE:HG23	1:A:79:ILE:HG13	0.42	1.90	16	1
1:A:93:LEU:HD13	1:A:112:GLN:CB	0.42	2.44	8	1
1:A:82:SER:O	1:A:86:ARG:HG2	0.41	2.15	11	1
1:A:71:VAL:O	1:A:75:LYS:HG2	0.41	2.15	20	1
1:A:124:LEU:O	1:A:124:LEU:HD22	0.41	2.15	15	2
1:A:74:MET:SD	1:A:85:ALA:HB1	0.41	2.56	13	1
1:A:83:LEU:HD22	1:A:86:ARG:HD2	0.41	1.93	2	1
1:A:124:LEU:N	1:A:124:LEU:HD13	0.41	2.31	3	1
1:A:89:LEU:HD21	1:A:113:VAL:HG21	0.41	1.93	20	1
1:A:50:ALA:HB2	1:A:125:PRO:HD3	0.41	1.93	3	1
1:A:82:SER:O	1:A:86:ARG:HD3	0.41	2.16	7	1
1:A:3:ILE:HA	1:A:6:LYS:CG	0.41	2.46	17	1
1:A:63:LYS:HD3	1:A:63:LYS:O	0.40	2.16	4	1
1:A:93:LEU:HD13	1:A:112:GLN:HB2	0.40	1.93	8	1
1:A:3:ILE:CA	1:A:6:LYS:HD3	0.40	2.36	14	1
1:A:72:ASP:HA	1:A:75:LYS:HE2	0.40	1.94	14	1
1:A:8:MET:HB3	1:A:117:LEU:CD1	0.40	2.46	16	1
1:A:66:TRP:HB3	1:A:73:LEU:HB3	0.40	1.92	7	1
1:A:63:LYS:HA	1:A:63:LYS:HE3	0.40	1.93	10	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	86/130 (66%)	77±1 (89±2%)	8±1 (9±1%)	1±1 (2±1%)	14 59
All	All	1720/2600 (66%)	1539 (89%)	155 (9%)	26 (2%)	14 59

All 5 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	78	ASP	15
1	A	2	SER	6
1	A	95	TYR	3
1	A	94	GLY	1
1	A	121	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	71/98 (72%)	66±2 (93±3%)	5±2 (7±3%)	18 66
All	All	1420/1960 (72%)	1318 (93%)	102 (7%)	18 66

All 27 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	124	LEU	20
1	A	63	LYS	15
1	A	8	MET	10
1	A	120	ASN	9
1	A	55	LYS	6
1	A	112	GLN	5
1	A	123	LYS	5
1	A	86	ARG	4
1	A	88	GLU	3
1	A	114	MET	3
1	A	72	ASP	2
1	A	117	LEU	2
1	A	58	LYS	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	77	LEU	2
1	A	67	ARG	2
1	A	83	LEU	1
1	A	105	MET	1
1	A	54	ASP	1
1	A	60	LYS	1
1	A	84	SER	1
1	A	64	LEU	1
1	A	111	LYS	1
1	A	80	ASP	1
1	A	57	VAL	1
1	A	2	SER	1
1	A	65	GLU	1
1	A	87	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