Full wwPDB NMR Structure Validation Report (i)

Oct 22, 2024 – 07:09 PM EDT

PDB ID	:	2JPO
Title	:	NMR structure of Antheraea polyphemus pheromone-binding protein 1 at pH 4.5
Authors	:	Damberger, F.F.; Wuthrich, K.; Leal, W.S.; Ishida, Y.
Deposited on	:	2007-05-20

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 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:

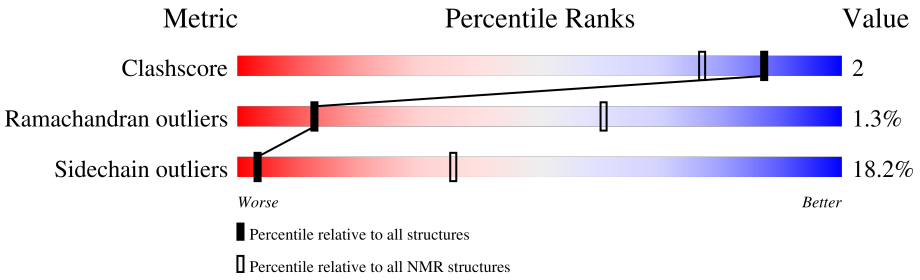
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	142	81% 13% • 5%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:8-A:142 (135)	0.60	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 5 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2167 atoms, of which 1070 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Pheromone-binding protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	142	2167	686	1070	183	214	14	0

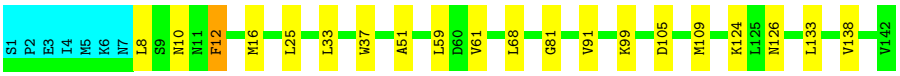
4 Residue-property plots

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: Pheromone-binding protein

Chain A:



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

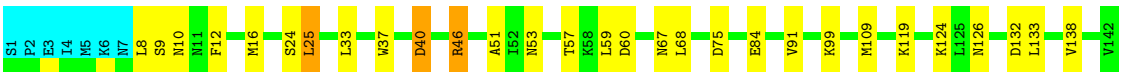
• Molecule 1: Pheromone-binding protein

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Pheromone-binding protein

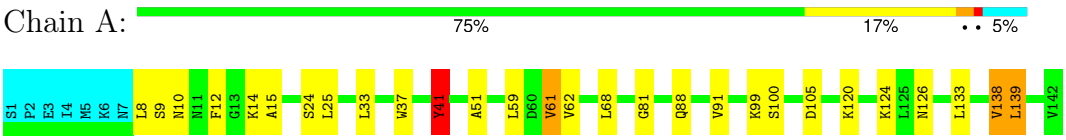
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Pheromone-binding protein

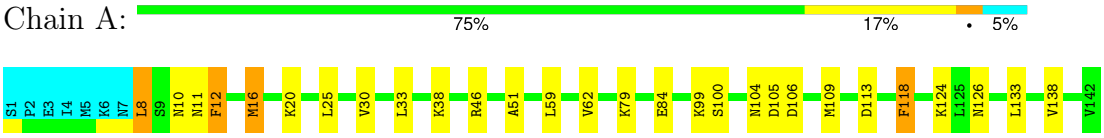
Chain A:



4.2.4 Score per residue for model 4 (medoid)

• Molecule 1: Pheromone-binding protein

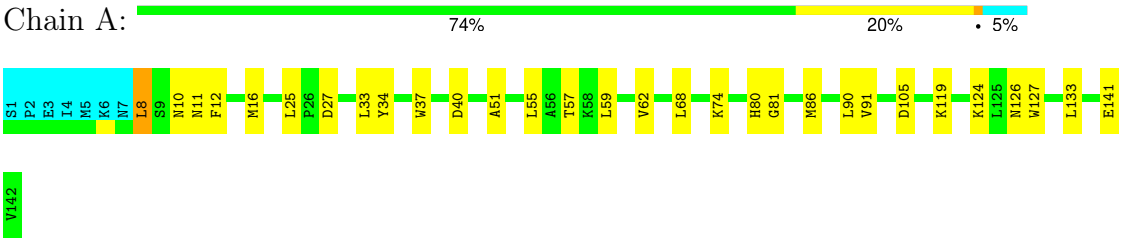
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Pheromone-binding protein

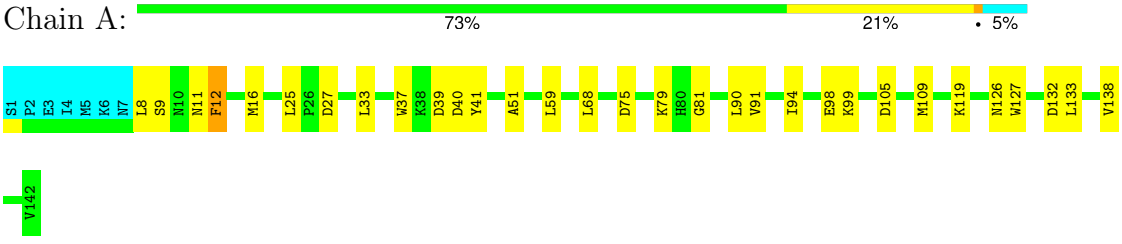
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Pheromone-binding protein

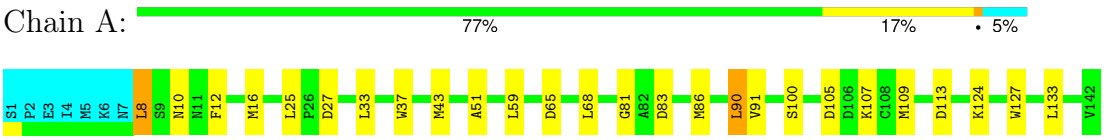
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Pheromone-binding protein

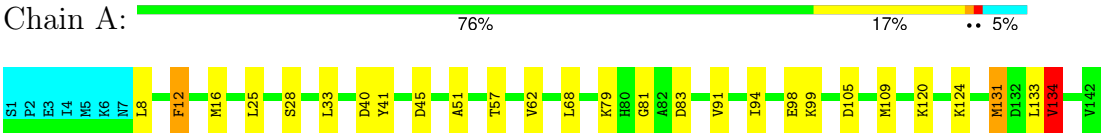
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Pheromone-binding protein

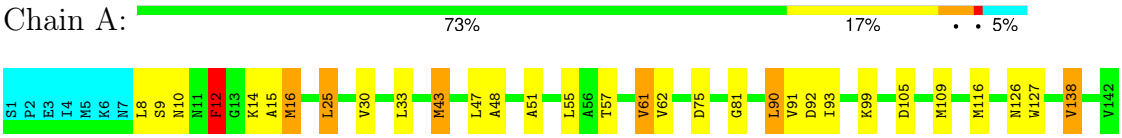
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Pheromone-binding protein

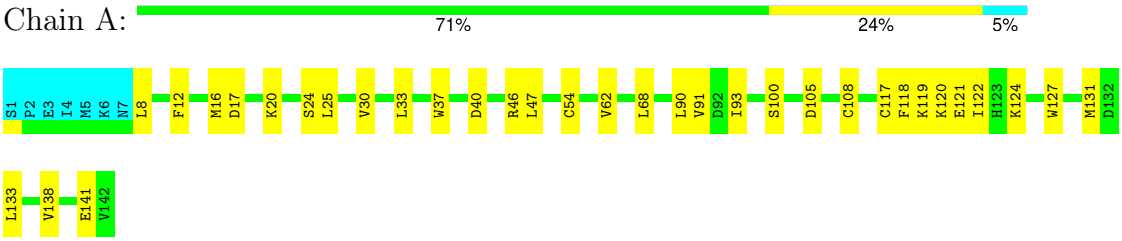
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Pheromone-binding protein

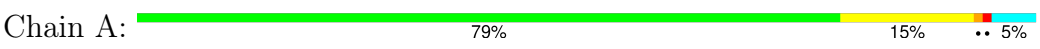
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Pheromone-binding protein

Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Pheromone-binding protein

Chain A:

4.2.13 Score per residue for model 13

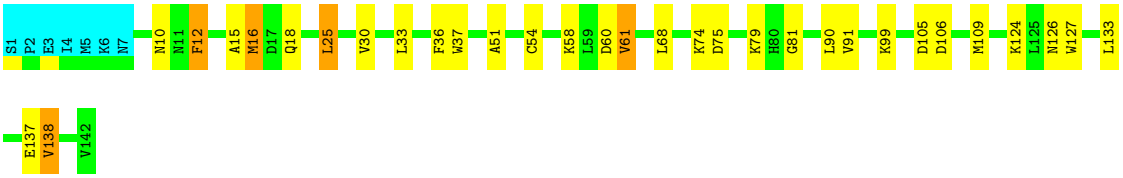
• Molecule 1: Pheromone-binding protein

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: Pheromone-binding protein

Chain A:



4.2.15 Score per residue for model 15

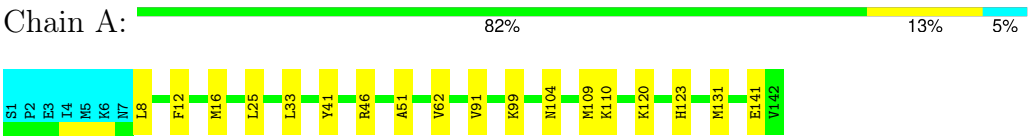
• Molecule 1: Pheromone-binding protein

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: Pheromone-binding protein

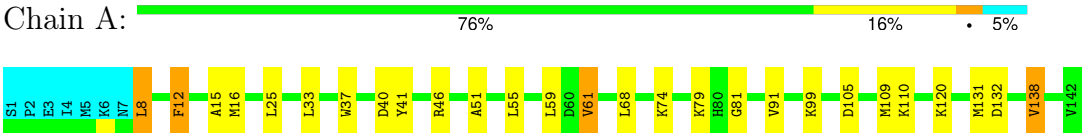
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Pheromone-binding protein

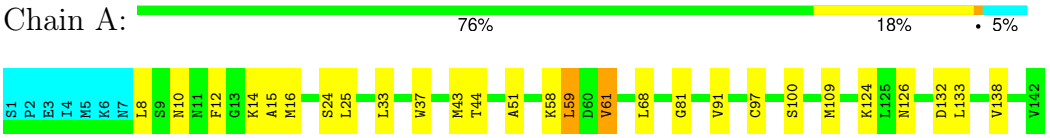
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Pheromone-binding protein

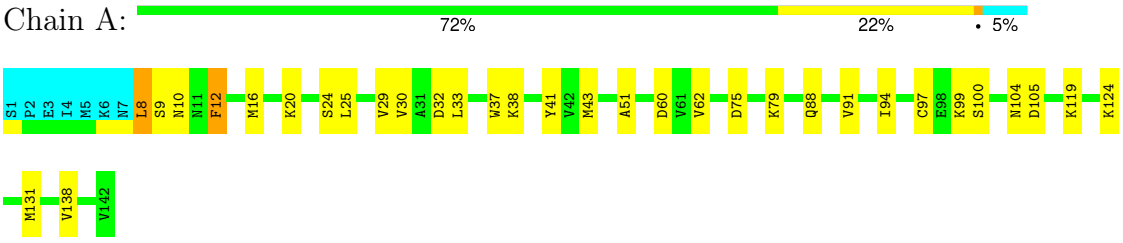
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Pheromone-binding protein

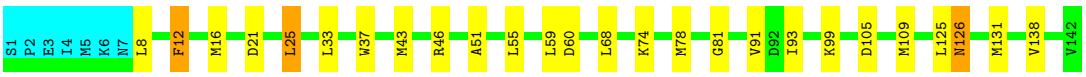
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Pheromone-binding protein

Chain A:


S1
P2
E3
I4
M5
K6
N7
L8
F12
M16
D21
L26
L33
W37
M43
R46
A51
L55
L59
D60
L68
K74
M78
G81
V91
D92
I93
K99
D105
M109
L125
N126
M131
V138
V142

5 Refinement protocol and experimental data overview

The models were refined using the following method: *TORSION ANGLE DYNAMICS, SIMULATED ANNEALING.*

Of the 80 calculated structures, 20 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
OPALp	refinement	1.2
CARA	structure solution	1.8
ATNOS/CANDID	structure solution	1.2
DYANA	structure solution	1.0.3
MOLMOL	structure solution	2.2K

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

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	0.54±0.01	0±0/1061 (0.0± 0.0%)	1.05±0.03	1±1/1435 (0.1± 0.1%)
All	All	0.54	0/21220 (0.0%)	1.05	24/28700 (0.1%)

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	0.8±0.9
All	All	0	17

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	47	LEU	CB-CG-CD1	7.02	122.93	111.00	15	2
1	A	108	CYS	CA-CB-SG	6.43	125.58	114.00	10	1
1	A	138	VAL	CA-CB-CG2	6.08	120.02	110.90	14	2
1	A	138	VAL	CA-CB-CG1	6.06	120.00	110.90	12	9
1	A	90	LEU	CB-CG-CD1	6.01	121.22	111.00	7	2
1	A	46	ARG	NE-CZ-NH2	-5.60	117.50	120.30	2	1
1	A	8	LEU	CB-CA-C	5.52	120.69	110.20	3	1
1	A	68	LEU	CB-CG-CD1	5.48	120.32	111.00	17	2
1	A	41	TYR	CB-CG-CD1	-5.43	117.74	121.00	3	1
1	A	61	VAL	CG1-CB-CG2	-5.14	102.68	110.90	11	1
1	A	62	VAL	CA-CB-CG1	5.05	118.48	110.90	3	1
1	A	134	VAL	CA-CB-CG2	5.04	118.47	110.90	8	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	12	PHE	Sidechain	7
1	A	41	TYR	Sidechain	4
1	A	118	PHE	Sidechain	1
1	A	34	TYR	Sidechain	1
1	A	46	ARG	Sidechain	1
1	A	36	PHE	Sidechain	1
1	A	106	ASP	Peptide	1
1	A	39	ASP	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	1042	1013	1008	4±2
All	All	20840	20260	20160	79

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:25:LEU:CD1	1:A:51:ALA:HB2	0.72	2.14	7	17
1:A:25:LEU:HD12	1:A:51:ALA:HB2	0.68	1.65	1	16
1:A:15:ALA:HB2	1:A:61:VAL:CG1	0.66	2.21	18	6
1:A:43:MET:SD	1:A:48:ALA:HB1	0.59	2.37	13	1
1:A:15:ALA:HB2	1:A:61:VAL:HG13	0.53	1.81	18	1
1:A:25:LEU:HD11	1:A:51:ALA:HB2	0.53	1.79	9	3
1:A:93:ILE:HD11	1:A:125:LEU:HD11	0.53	1.81	20	1
1:A:90:LEU:HD11	1:A:127:TRP:CH2	0.52	2.39	6	7
1:A:93:ILE:HG23	1:A:121:GLU:HB2	0.49	1.84	10	1
1:A:8:LEU:HD21	1:A:12:PHE:CD2	0.48	2.44	17	2
1:A:15:ALA:HB2	1:A:61:VAL:HG12	0.48	1.86	3	2
1:A:88:GLN:HA	1:A:91:VAL:HG23	0.47	1.86	11	1
1:A:131:MET:HA	1:A:134:VAL:CG2	0.46	2.40	8	1
1:A:59:LEU:HD22	1:A:68:LEU:HG	0.46	1.87	20	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:47:LEU:HD22	1:A:47:LEU:H	0.45	1.72	9	1
1:A:41:TYR:C	1:A:41:TYR:CD1	0.45	2.89	3	1
1:A:16:MET:CE	1:A:30:VAL:HG22	0.45	2.42	9	3
1:A:20:LYS:CD	1:A:30:VAL:HG21	0.45	2.42	4	3
1:A:125:LEU:C	1:A:126:ASN:HD22	0.43	2.16	20	1
1:A:90:LEU:HD11	1:A:127:TRP:CZ2	0.43	2.49	6	1
1:A:59:LEU:HD21	1:A:139:LEU:HB3	0.43	1.91	3	1
1:A:12:PHE:HA	1:A:62:VAL:HG23	0.42	1.90	9	1
1:A:8:LEU:HD13	1:A:80:HIS:CE1	0.42	2.50	5	1
1:A:43:MET:SD	1:A:48:ALA:HB3	0.41	2.55	9	1
1:A:36:PHE:CD1	1:A:43:MET:HG3	0.41	2.51	13	1
1:A:93:ILE:HG22	1:A:118:PHE:HA	0.41	1.91	10	1
1:A:59:LEU:H	1:A:59:LEU:HD22	0.41	1.76	18	1
1:A:14:LYS:HE3	1:A:61:VAL:HG22	0.40	1.90	9	1
1:A:118:PHE:CD1	1:A:122:ILE:HD11	0.40	2.51	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	134/142 (94%)	120±2 (89±1%)	13±2 (9±2%)	2±1 (1±1%)	13	60
All	All	2680/2840 (94%)	2392 (89%)	254 (9%)	34 (1%)	13	60

All 9 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	81	GLY	13
1	A	105	ASP	6
1	A	40	ASP	5
1	A	8	LEU	5
1	A	38	LYS	1
1	A	43	MET	1
1	A	39	ASP	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	123	HIS	1
1	A	104	ASN	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	116/123 (94%)	95±3 (82±3%)	21±3 (18±3%)	3	36
All	All	2320/2460 (94%)	1898 (82%)	422 (18%)	3	36

All 80 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	12	PHE	20
1	A	33	LEU	20
1	A	8	LEU	18
1	A	16	MET	18
1	A	91	VAL	17
1	A	37	TRP	15
1	A	99	LYS	14
1	A	124	LYS	14
1	A	68	LEU	13
1	A	109	MET	13
1	A	133	LEU	12
1	A	10	ASN	11
1	A	126	ASN	11
1	A	59	LEU	9
1	A	100	SER	9
1	A	105	ASP	9
1	A	57	THR	8
1	A	138	VAL	8
1	A	62	VAL	8
1	A	79	LYS	7
1	A	24	SER	7
1	A	9	SER	6
1	A	25	LEU	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	46	ARG	6
1	A	60	ASP	6
1	A	61	VAL	6
1	A	55	LEU	6
1	A	131	MET	6
1	A	43	MET	5
1	A	75	ASP	5
1	A	119	LYS	5
1	A	132	ASP	5
1	A	120	LYS	5
1	A	40	ASP	4
1	A	74	LYS	4
1	A	110	LYS	4
1	A	65	ASP	3
1	A	83	ASP	3
1	A	84	GLU	3
1	A	11	ASN	3
1	A	104	ASN	3
1	A	113	ASP	3
1	A	27	ASP	3
1	A	141	GLU	3
1	A	94	ILE	3
1	A	98	GLU	3
1	A	47	LEU	3
1	A	97	CYS	3
1	A	137	GLU	3
1	A	28	SER	2
1	A	92	ASP	2
1	A	14	LYS	2
1	A	41	TYR	2
1	A	88	GLN	2
1	A	106	ASP	2
1	A	86	MET	2
1	A	39	ASP	2
1	A	107	LYS	2
1	A	116	MET	2
1	A	54	CYS	2
1	A	58	LYS	2
1	A	53	ASN	1
1	A	67	ASN	1
1	A	139	LEU	1
1	A	118	PHE	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	45	ASP	1
1	A	134	VAL	1
1	A	93	ILE	1
1	A	17	ASP	1
1	A	117	CYS	1
1	A	89	GLN	1
1	A	142	VAL	1
1	A	72	ASN	1
1	A	18	GLN	1
1	A	44	THR	1
1	A	29	VAL	1
1	A	32	ASP	1
1	A	38	LYS	1
1	A	21	ASP	1
1	A	78	MET	1

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no oligosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

No chemical shift data were provided