Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 11:11 pm BST

PDB ID	:	2KX6
Title	:	Signaling state of Photoactive Yellow Protein
Authors	:	Ramachandran, P.L.; Lovett, J.E.; Carl, P.J.; Cammarata, M.; Lee, J.H.; Yang, J.O.; Ihee, H.; Timmel, C.R.; van Thor, J.
Deposited on	:	2010-04-27

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:

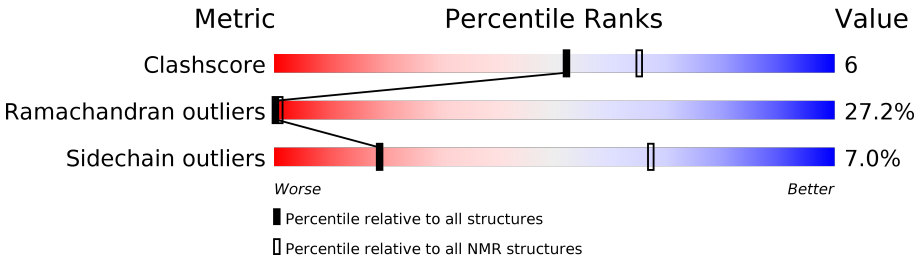
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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

The following experimental techniques were used to determine the structure:
SOLUTION NMR, SOLUTION SCATTERING

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	50% 26% • 18% •

2 Ensemble composition and analysis

This entry contains 14 models. Model 13 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:14-A:15, A:27-A:125 (101)	0.16	13

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 3 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1929 atoms, of which 942 are hydrogens and 0 are deuteriums.

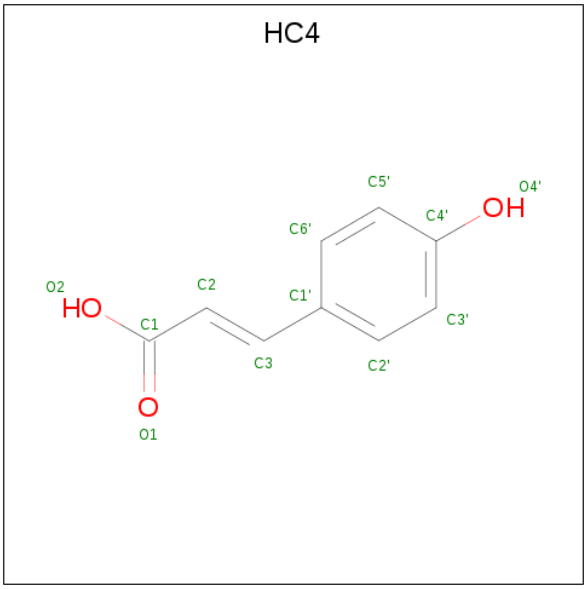
- Molecule 1 is a protein called Photoactive yellow protein.

Mol	Chain	Residues	Atoms						Trace
1	A	125	Total	C	H	N	O	S	0
			1911	620	935	158	192	6	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	PHE	-	EXPRESSION TAG	UNP P16113
A	-3	GLY	-	EXPRESSION TAG	UNP P16113
A	-2	LEU	-	EXPRESSION TAG	UNP P16113
A	-1	GLY	-	EXPRESSION TAG	UNP P16113
A	0	SER	-	EXPRESSION TAG	UNP P16113

- Molecule 2 is 4'-HYDROXYCINNAMIC ACID (three-letter code: HC4) (formula: $C_9H_8O_3$).



Mol	Chain	Residues	Atoms			
2	A	1	Total	C	H	O
			18	9	7	2

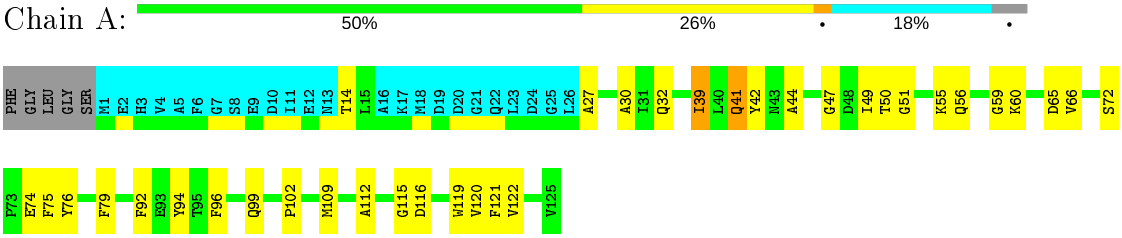
4 Residue-property plots

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: Photoactive yellow protein

Chain A:



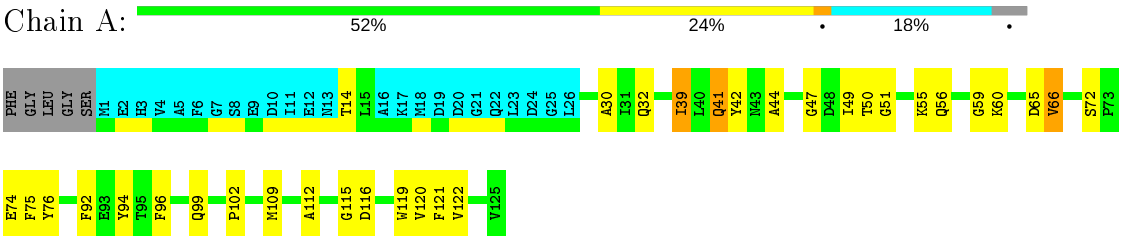
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Photoactive yellow protein

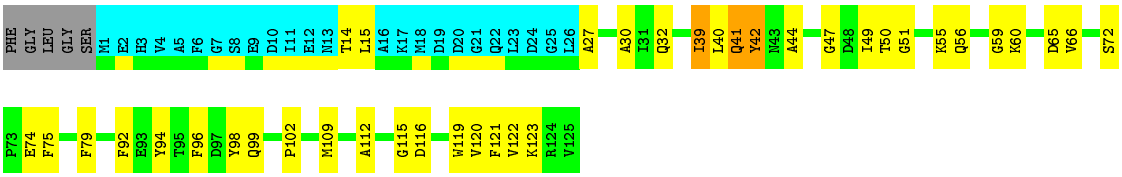
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Photoactive yellow protein

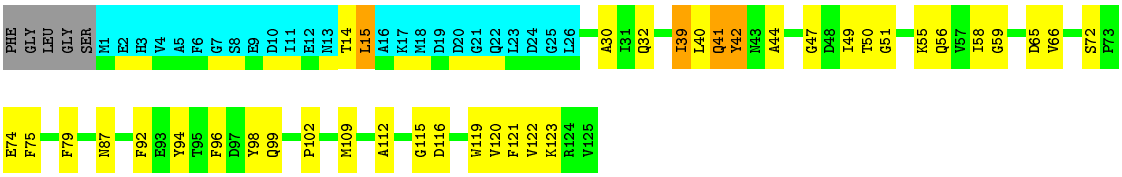
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Photoactive yellow protein

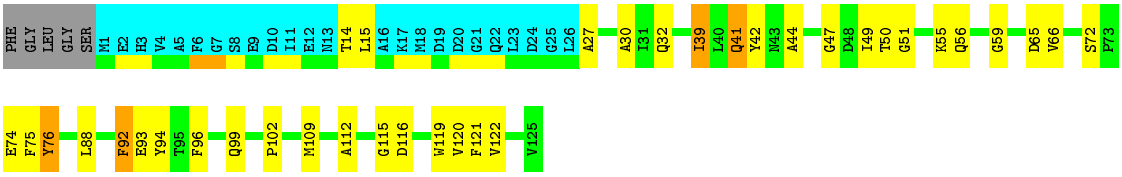
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Photoactive yellow protein

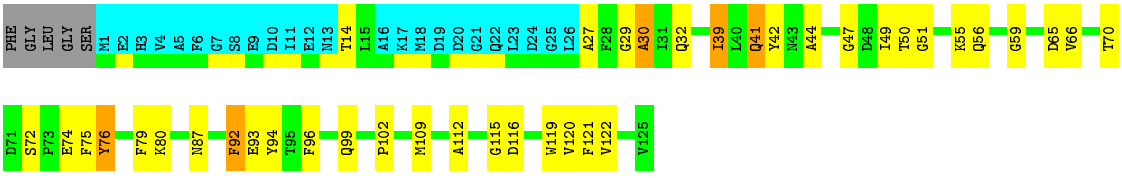
Chain A:



4.2.5 Score per residue for model 5

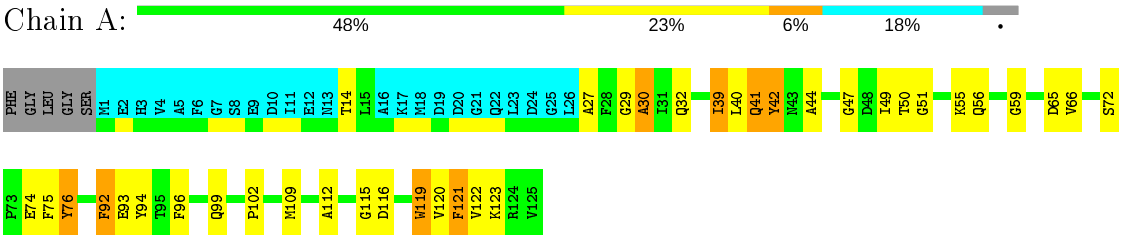
- Molecule 1: Photoactive yellow protein

Chain A:



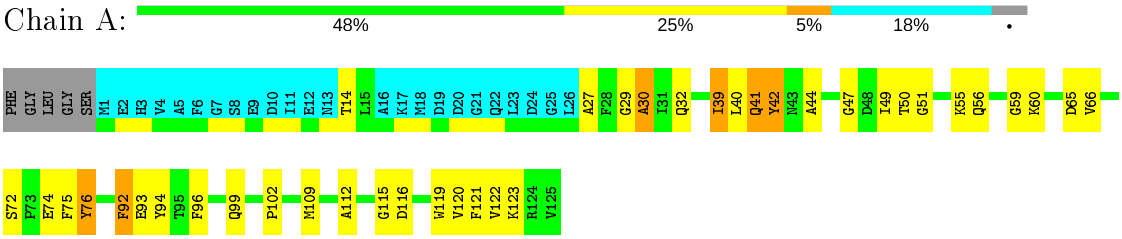
4.2.6 Score per residue for model 6

- Molecule 1: Photoactive yellow protein



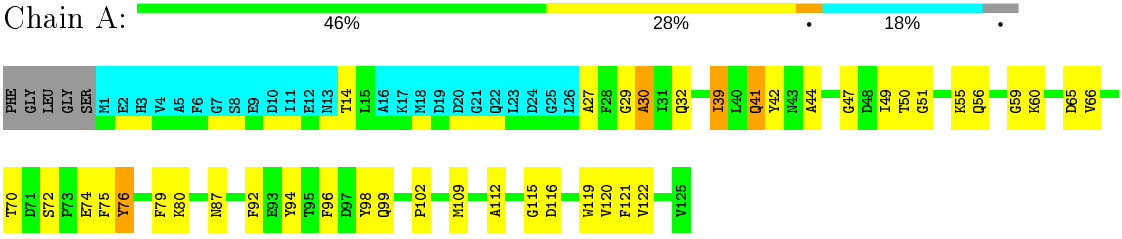
4.2.7 Score per residue for model 7

- Molecule 1: Photoactive yellow protein



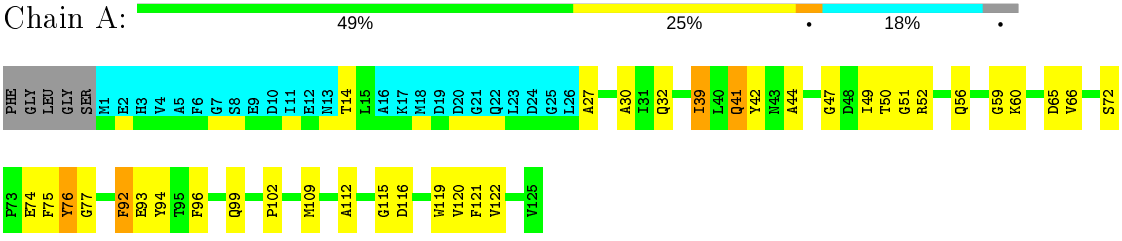
4.2.8 Score per residue for model 8

- Molecule 1: Photoactive yellow protein



4.2.9 Score per residue for model 9

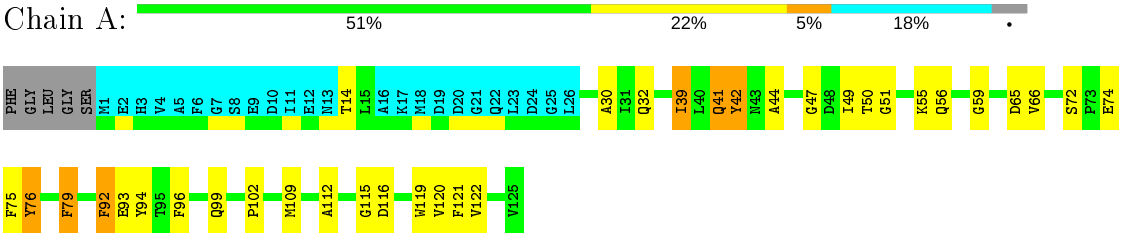
- Molecule 1: Photoactive yellow protein



4.2.10 Score per residue for model 10

• Molecule 1: Photoactive yellow protein

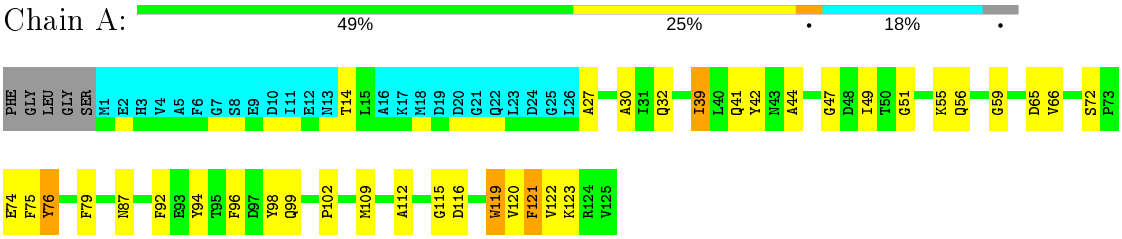
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Photoactive yellow protein

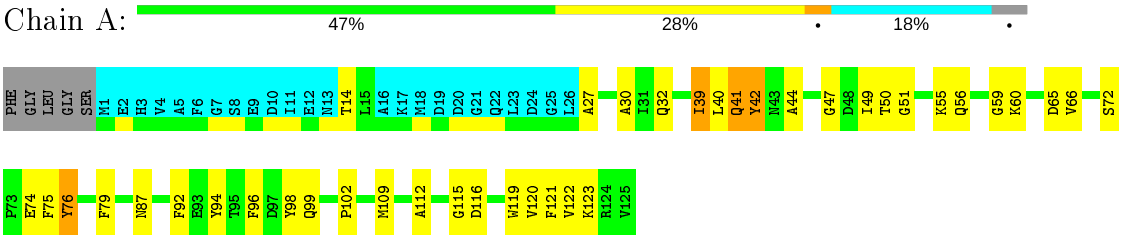
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Photoactive yellow protein

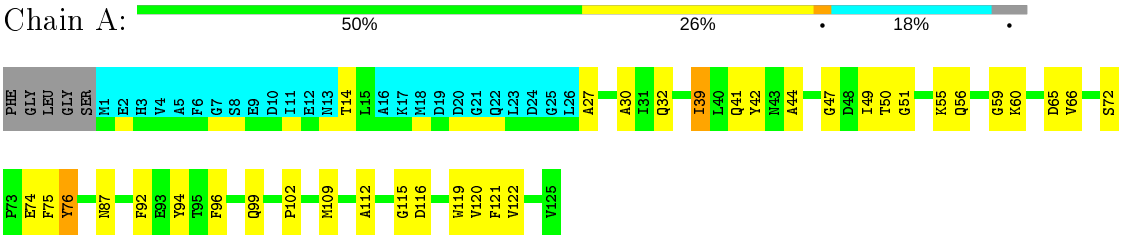
Chain A:



4.2.13 Score per residue for model 13 (medoid)

• Molecule 1: Photoactive yellow protein

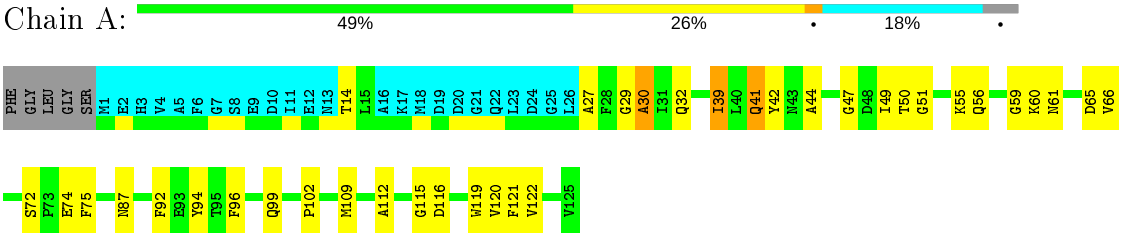
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Photoactive yellow protein

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 15 calculated structures, 14 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CNS	structure solution	
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

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HC4

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

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	795	773	772	8±1
2	A	11	7	7	1±0
All	All	11284	10920	10906	127

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 6.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:126:HC4:O1	2:A:126:HC4:H2'	0.92	1.65	2	6
2:A:126:HC4:H2'	2:A:126:HC4:O1	0.92	1.65	13	5
2:A:126:HC4:O1	2:A:126:HC4:H6'	0.91	1.65	3	2
2:A:126:HC4:H6'	2:A:126:HC4:O1	0.89	1.65	8	1
1:A:39:ILE:HD11	1:A:59:GLY:N	0.82	1.89	7	14
1:A:58:ILE:HD12	1:A:58:ILE:O	0.54	2.02	3	1
1:A:39:ILE:HD11	1:A:59:GLY:H	0.51	1.63	7	14
1:A:39:ILE:N	1:A:39:ILE:HD13	0.50	2.22	5	3
1:A:39:ILE:HD13	1:A:39:ILE:N	0.46	2.25	4	6
1:A:79:PHE:CE1	1:A:80:LYS:NZ	0.46	2.83	5	2
2:A:126:HC4:C1	2:A:126:HC4:H6'	0.46	2.34	8	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:75:PHE:CD2	1:A:109:MET:SD	0.46	3.09	6	14
1:A:39:ILE:N	1:A:39:ILE:CD1	0.45	2.80	5	2
1:A:40:LEU:O	1:A:42:TYR:N	0.45	2.50	3	5
1:A:79:PHE:CZ	1:A:80:LYS:NZ	0.45	2.85	5	1
1:A:79:PHE:O	1:A:79:PHE:CG	0.43	2.72	10	1
1:A:39:ILE:HD12	1:A:60:LYS:O	0.42	2.14	12	8
1:A:120:VAL:O	1:A:120:VAL:HG23	0.42	2.13	12	7
1:A:119:TRP:CE2	1:A:121:PHE:CE1	0.42	3.08	6	2
1:A:120:VAL:HG23	1:A:120:VAL:O	0.42	2.13	11	7
1:A:39:ILE:CD1	1:A:39:ILE:N	0.42	2.83	4	3
1:A:92:PHE:CG	1:A:93:GLU:N	0.41	2.87	9	6
1:A:79:PHE:O	1:A:79:PHE:CD2	0.41	2.73	2	3
1:A:29:GLY:O	1:A:30:ALA:HB2	0.41	2.15	8	5
1:A:75:PHE:O	1:A:76:TYR:CD2	0.41	2.74	8	1
1:A:75:PHE:CE2	1:A:109:MET:SD	0.41	3.14	14	3
1:A:75:PHE:CG	1:A:109:MET:SD	0.40	3.14	10	2
1:A:42:TYR:O	1:A:42:TYR:CD1	0.40	2.74	10	1
1:A:79:PHE:CD2	1:A:79:PHE:O	0.40	2.74	3	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	100/130 (77%)	45±2 (45±2%)	28±2 (28±2%)	27±1 (27±1%)	0	1
All	All	1400/1820 (77%)	625 (45%)	394 (28%)	381 (27%)	0	1

All 34 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	92	PHE	14
1	A	14	THR	14
1	A	30	ALA	14
1	A	44	ALA	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	74	GLU	14
1	A	51	GLY	14
1	A	122	VAL	14
1	A	56	GLN	14
1	A	42	TYR	14
1	A	116	ASP	14
1	A	115	GLY	14
1	A	102	PRO	14
1	A	32	GLN	14
1	A	112	ALA	14
1	A	65	ASP	14
1	A	94	TYR	14
1	A	121	PHE	14
1	A	49	ILE	14
1	A	96	PHE	14
1	A	47	GLY	14
1	A	41	GLN	14
1	A	99	GLN	14
1	A	55	LYS	13
1	A	50	THR	13
1	A	76	TYR	11
1	A	27	ALA	11
1	A	87	ASN	7
1	A	123	LYS	6
1	A	98	TYR	4
1	A	15	LEU	3
1	A	70	THR	2
1	A	52	ARG	1
1	A	66	VAL	1
1	A	77	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	84/106 (79%)	78±1 (93±1%)	6±1 (7±1%)	19	67
All	All	1176/1484 (79%)	1094 (93%)	82 (7%)	19	67

All 11 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	39	ILE	14
1	A	72	SER	14
1	A	119	TRP	14
1	A	66	VAL	14
1	A	41	GLN	12
1	A	76	TYR	9
1	A	88	LEU	1
1	A	61	ASN	1
1	A	79	PHE	1
1	A	15	LEU	1
1	A	98	TYR	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 carbohydrates in this entry.

6.6 Ligand geometry

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	HC4	A	126	1	11,11,12	1.18±0.12	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	HC4	A	126	1	13,13,15	0.93±0.02	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	HC4	A	126	1	-	1±0,4,4,5	0±0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique torsion outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	126	HC4	O1-C1-C2-C3	14

There are no ring outliers.

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

No chemical shift data were provided