wwPDB NMR Structure Validation Summary Report (i)

Mar 7, 2022 – 04:15 AM EST

PDB ID	:	3PHY
Title	:	PHOTOACTIVE YELLOW PROTEIN, DARK STATE (UNBLEACHED), SOLUTION STRUCTURE, NMR, 26 STRUCTURES
Authors	:	Dux, P.; Rubinstenn, G.; Vuister, G.W.; Boelens, R.; Mulder, F.A.A.; Hard, K.; Hoff, W.D.; Kroon, A.; Crielaard, W.; Hellingwerf, K.J.; Kaptein, R.
Deposited on	:	1998-02-06

This is a wwPDB NMR Structure Validation Summary 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 (1)) were used in the production of this report:

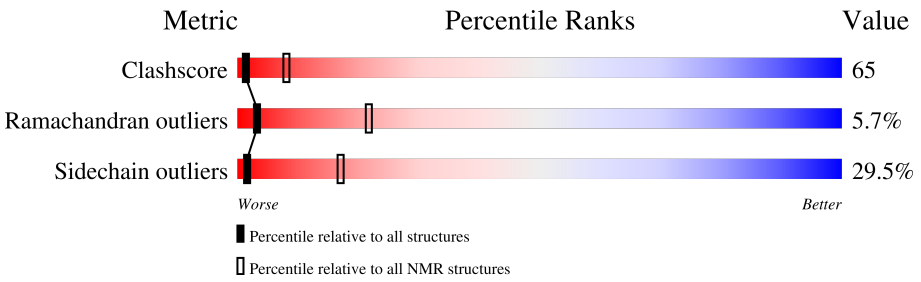
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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	125	24% 48% 12% • 15%

2 Ensemble composition and analysis (i)

This entry contains 26 models. Model 20 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:10-A:16, A:24-A:113, A:117-A:125 (106)	0.31	20

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 3 single-model clusters were found.

Cluster number	Models
1	10, 15, 17, 20, 21, 22, 23
2	1, 5, 6, 9, 12, 14, 26
3	3, 7, 13, 25
4	4, 18, 24
5	2, 16
Single-model clusters	8; 11; 19

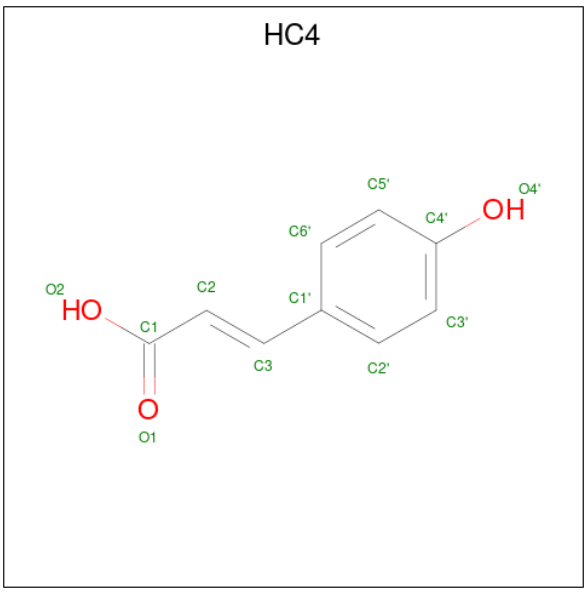
3 Entry composition (i)

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
			Total	C	H	N	O	S	
1	A	125	1912	620	936	158	192	6	0

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



Mol	Chain	Residues	Atoms			
			Total	C	H	O
2	A	1	17	9	6	2

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: PHOTOACTIVE YELLOW PROTEIN

Chain A: 24% 48% 12% • 15%

M1 E2 H3 V4 A5 F6 G7 S8 E9 D10 I11 E12 N13 T14 L15 A16 K17 M18 D19 D20 G21 Q22 L23 L26 A27 F28 G29 A30 I31 Q32 L33 D34 G35 D36 G37 N38 I39 L40 Q41 Y42 N43 E46 G47 D48 I49 T50 G51 R52 D53 P54 K55 Q56 V57 I58 N61 F62 F63

K64 D65 V66 A67 T70 D71 S72 P73 E74 F75 Y76 G77 K78 F79 K80 V83 L88 M89 T90 M91 F92 E93 Y94 T95 F96 D97 Y98 Q99 M100 T101 K104 V105 K106 V107 H108 M109 K110 K111 A112 L113 S114 G115 D116 S117 Y118 W119 V120 F121 V122 K123 R124 V125

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 20. Colouring as in section 4.1 above.

• Molecule 1: PHOTOACTIVE YELLOW PROTEIN

Chain A: 27% 43% 13% • 15%

M1 E2 H3 V4 A5 F6 G7 S8 E9 D10 I11 E12 N13 T14 L15 A16 K17 M18 D19 D20 G21 Q22 L23 D24 F28 G29 A30 I31 Q32 L33 D34 G35 D36 G37 L40 Q41 Y42 N43 A44 A45 E46 G47 D48 I49 T50 G51 R52 D53 P54 K55 Q56 V57 I58 N61 F62 F63 K64

D65 V66 T70 D71 S72 P73 E74 F75 Y76 G77 K78 F79 K80 V83 L88 M89 T90 M91 F92 E93 Y94 T95 F96 D97 Y98 Q99 M100 T101 P102 T103 K104 V105 K106 V107 H108 M109 K110 K111 A112 L113 S114 G115 D116 S117 Y118 W119 V120 F121 V122 K123 R124 V125

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *distance geometry*.

Of the 40 calculated structures, 26 were deposited, based on the following criterion: *LOW ENERGY*.

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

Software name	Classification	Version
X-PLOR	refinement	3.1
X-PLOR	structure solution	

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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

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	1.10±0.01	1±0/854 (0.1± 0.0%)	0.92±0.01	0±0/1151 (0.0± 0.0%)
All	All	1.10	26/22204 (0.1%)	0.92	0/29926 (0.0%)

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±0.4
All	All	0	46

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	46	GLU	CD-OE2	11.78	1.38	1.25	9	26

There are no bond-angle outliers.

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	124	ARG	Sidechain	26
1	A	52	ARG	Sidechain	20

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	835	811	810	109±6
2	A	11	6	6	6±4
All	All	21996	21242	21216	2829

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 65.

5 of 544 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:ILE:HG22	1:A:15:LEU:HD11	1.05	1.28	23	26
1:A:92:PHE:CE1	1:A:107:VAL:HG21	0.99	1.93	10	18
1:A:92:PHE:CE2	1:A:107:VAL:HG21	0.97	1.94	25	15
1:A:92:PHE:CZ	1:A:107:VAL:HG21	0.94	1.96	2	22
1:A:26:LEU:HD22	1:A:28:PHE:CZ	0.93	1.99	9	13

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	105/125 (84%)	85±2 (81±2%)	14±1 (14±1%)	6±2 (6±2%)	3 22
All	All	2730/3250 (84%)	2201 (81%)	373 (14%)	156 (6%)	3 22

5 of 19 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	43	ASN	26
1	A	99	GLN	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	35	GLY	18
1	A	54	PRO	13
1	A	10	ASP	11

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	88/103 (85%)	62±4 (70±4%)	26±4 (30±4%)	1 17
All	All	2288/2678 (85%)	1613 (70%)	675 (30%)	1 17

5 of 64 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	58	ILE	26
1	A	94	TYR	25
1	A	41	GLN	24
1	A	61	ASN	24
1	A	72	SER	23

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)

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	2.30±0.04	3±0 (23±4%)

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	1.38±0.04	2±0 (14±2%)

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

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
2	A	126	HC4	O4'-C4'	6.84	1.21	1.37	16	26
2	A	126	HC4	C2-C3	2.65	1.29	1.34	5	14
2	A	126	HC4	C2-C1	2.33	1.51	1.44	10	26

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
2	A	126	HC4	C1'-C3-C2	4.46	120.15	127.21	22	26
2	A	126	HC4	C6'-C1'-C2'	2.27	120.99	117.64	11	22

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	26

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 (i)

No chemical shift data were provided