wwPDB NMR Structure Validation Summary Report

Apr 1, 2024 – 10:07 AM EDT

PDB ID	:	8EPY
BMRB ID	:	31046
Title	:	The solution structure of abxF in complex with its product (-)-ABX, an enzyme catalyzing the formation of the chiral spiroketal of an anthrabenzoxocinone antibiotic, (-)-ABX
Authors	:	Jia, X.; Yan, X.; Qu, X.; Mobli, M.
Deposited on	:	2022-10-06

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 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:

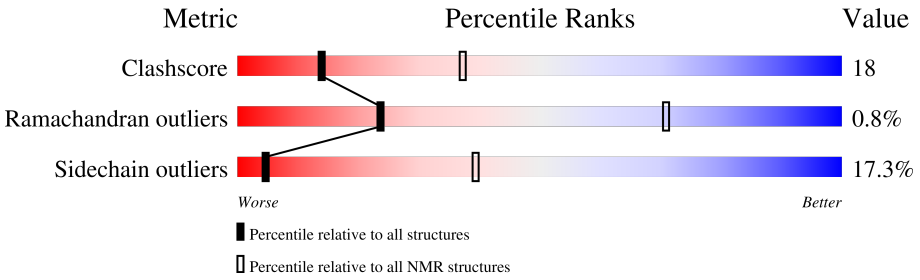
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.1

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 is 91%.

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	245	56% 22% . 20%

2 Ensemble composition and analysis

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: *target function*.

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:38, A:48-A:129, A:138-A:163, A:180-A:242 (197)	0.43	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 3 clusters and 4 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3518 atoms, of which 1712 are hydrogens and 0 are deuteriums.

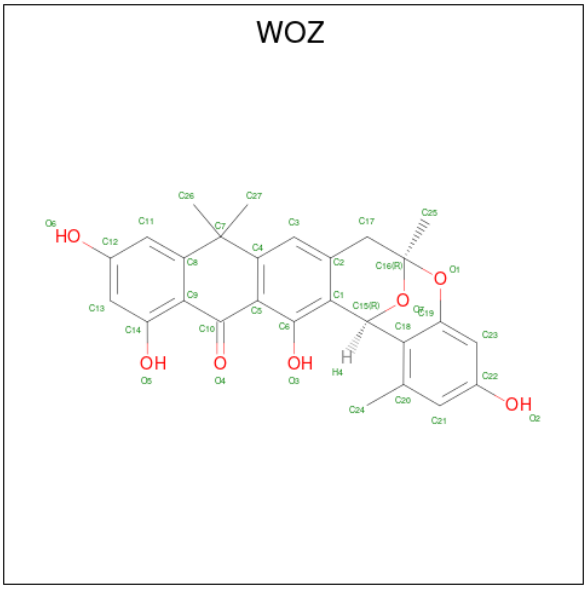
- Molecule 1 is a protein called Glyoxalase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	245	3460	1110	1688	307	347	8	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP A0A2I6B3F9
A	-1	SER	-	expression tag	UNP A0A2I6B3F9
A	0	HIS	-	expression tag	UNP A0A2I6B3F9

- Molecule 2 is (6R,16R)-3,11,13,15-tetrahydroxy-1,6,9,9-tetramethyl-6,7,9,16-tetrahydro-14H-6,16-epoxyanthra[2,3-e]benzo[b]oxocin-14-one (three-letter code: WOZ) (formula: $C_{27}H_{24}O_7$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			
			Total	C	H	O
2	A	1	58	27	24	7

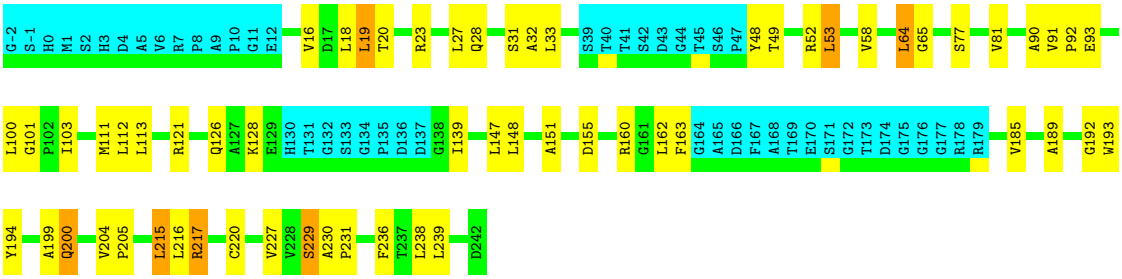
4 Residue-property plots (i)

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: Glyoxalase

Chain A:

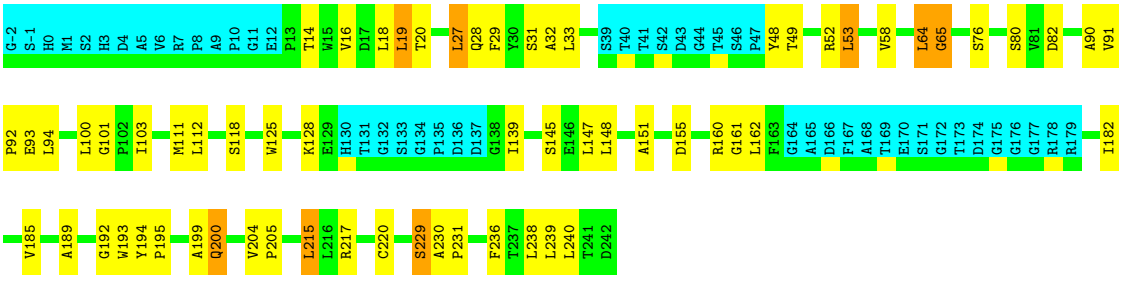


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: Glyoxalase

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 200 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
CYANA	refinement	
CYANA	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	2684
Number of shifts mapped to atoms	2681
Number of unparsed shifts	0
Number of shifts with mapping errors	3
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

6 Model quality

6.1 Standard geometry

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

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	1445	1410	1410	51±3
All	All	29580	28680	28200	1015

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 18.

5 of 152 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:19:LEU:HD23	1:A:64:LEU:HD12	0.85	1.48	17	15
1:A:64:LEU:HD13	1:A:64:LEU:N	0.81	1.91	13	15
1:A:19:LEU:HD23	1:A:64:LEU:HD22	0.81	1.52	15	1
1:A:19:LEU:O	1:A:19:LEU:HD13	0.78	1.79	14	20
1:A:19:LEU:HD22	1:A:19:LEU:C	0.76	2.01	9	20

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	196/245 (80%)	171±2 (87±1%)	23±3 (12±1%)	2±1 (1±0%)	24 71
All	All	3920/4900 (80%)	3426 (87%)	463 (12%)	31 (1%)	24 71

All 4 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	151	ALA	17
1	A	65	GLY	10
1	A	186	GLY	2
1	A	138	GLY	2

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	142/176 (81%)	118±2 (83±1%)	24±2 (17±1%)	5 39
All	All	2840/3520 (81%)	2350 (83%)	490 (17%)	5 39

5 of 53 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	19	LEU	20
1	A	53	LEU	20
1	A	64	LEU	20
1	A	128	LYS	20
1	A	215	LEU	20

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 monosaccharides 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	WOZ	A	301	-	38,39,39	0.82±0.00	2±0 (5±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	WOZ	A	301	-	52,65,65	0.95±0.00	3±0 (5±0%)

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	301	WOZ	C1-C15	2.22	1.53	1.51	19	20
2	A	301	WOZ	C18-C15	2.17	1.53	1.51	17	20

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	301	WOZ	C18-C15-C1	3.28	115.63	110.19	20	20
2	A	301	WOZ	C6-C1-C15	2.86	125.63	119.71	16	20
2	A	301	WOZ	C8-C7-C4	2.03	113.88	109.07	19	17

There are no chirality outliers.

There are no torsion outliers.

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

The completeness of assignment taking into account all chemical shift lists is 91% for the well-defined parts and 89% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_ output*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	2684
Number of shifts mapped to atoms	2681
Number of unparsed shifts	0
Number of shifts with mapping errors	3
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 3 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	301	WOZ	Q24	2.435	0.000	.
1	A	301	WOZ	Q26	0.737	0.000	.
1	A	301	WOZ	Q27	0.573	0.000	.

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	241	-0.02 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	205	0.14 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	220	0.17 ± 0.10	None needed (< 0.5 ppm)
^{15}N	217	-0.29 ± 0.21	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 91%, i.e. 2268 atoms were assigned a chemical shift out of a possible 2500. 0 out of 38 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	952/983 (97%)	399/406 (98%)	376/394 (95%)	177/183 (97%)
Sidechain	1172/1317 (89%)	803/874 (92%)	356/403 (88%)	13/40 (32%)
Aromatic	144/200 (72%)	76/97 (78%)	64/97 (66%)	4/6 (67%)
Overall	2268/2500 (91%)	1278/1377 (93%)	796/894 (89%)	194/229 (85%)

7.1.4 Statistically unusual chemical shifts

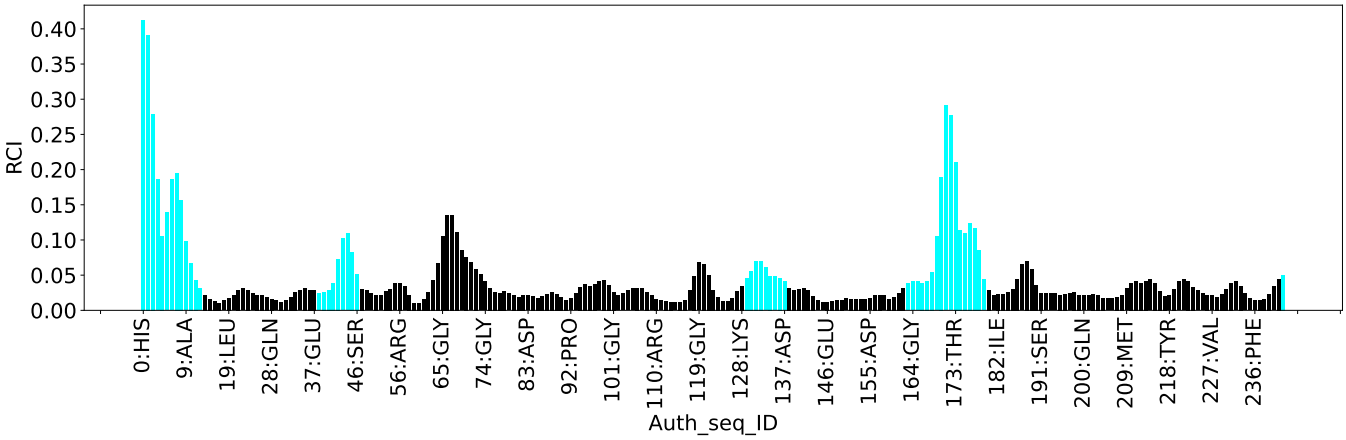
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	147	LEU	HB2	-1.79	-0.07 – 3.30	-10.1
1	A	18	LEU	HB2	-1.65	-0.07 – 3.30	-9.7
1	A	147	LEU	HB3	-1.79	-0.26 – 3.31	-9.3
1	A	18	LEU	HB3	-0.96	-0.26 – 3.31	-7.0
1	A	34	PHE	HB2	1.15	1.20 – 4.80	-5.1

7.1.5 Random Coil Index (RCI) plots

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:


RCI
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0:HIS
9:ALA
19:LEU
28:GLN
37:GLU
46:SER
56:ARG
65:GLY
74:GLY
83:ASP
92:PRO
101:GLY
110:ARG
119:GLY
128:LYS
137:ASP
146:GLU
155:ASP
164:GLY
173:THR
182:ILE
191:SER
200:GLN
209:MET
218:TYR
227:VAL
236:PHE
Auth_seq_ID