wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 01:33 PM EDT

PDB ID	:	2LAA
BMRB ID	:	17518
Title	:	Solution Strucuture of the CBM25-1 of beta/alpha-amylase from Paenibacillus polymyxa
Authors	:	Horibe, I.; Nishimura, S.; Takahashi, R.; Ohkubo, T.; Yoshida, T.
Deposited on	:	2011-03-09

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:

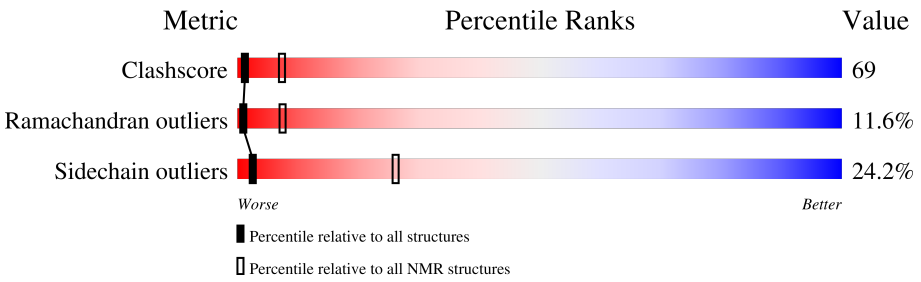
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	104	15% 50% 12% 22%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:424-A:483, A:495-A:507, A:514-A:521 (81)	0.32	11

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1458 atoms, of which 700 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Beta/alpha-amylase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	104	1458	469	700	130	158	1	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: Beta/alpha-amylase

Chain A: 15% 50% 12% 22%

G420 G421 T422 G423 N424 K425 V426 T427 I428 Y429 Y430 K431 K432 G433 F434 N435 S436 P437 Y438 I439 H440 Y441 R442 P443 A444 S447 W448 T449 A450 A451 P452 G453 V454 K455 M456 Q457 D458 A459 E460 I461 S462 G463 Y464 A465 K466 I467 T468 V469 D470 I471 G472 S473 A474 S475 Q476 L477 E478 A479 A480

F481 N482 D483 G484 N485 N486 N487 W488 D489 S490 N491 N492 T493 K494 N495 Y496 S497 F498 S499 T500 G501 T502 T506 P507 G508 N509 S510 G511 N512 A513 G514 T515 I516 T517 A520 P521 A522 G523

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

The representative model is number 11. Colouring as in section 4.1 above.

- Molecule 1: Beta/alpha-amylase

Chain A: 18% 46% 12% • 22%

G420 G421 T422 G423 N424 K425 V426 T427 I428 Y429 Y430 K431 K432 G433 F434 N435 S436 P437 Y438 I439 H440 Y441 R442 P443 A444 S447 W448 A451 P452 G453 V454 K455 M456 Q457 D458 A459 E460 I461 S462 G463 Y464 A465 K466 I467 T468 V469 D470 I471 G472 S473 A474 S475 Q476 L477 E478 A479 A480 F481

N482 D483 G484 N485 N486 N487 W488 D489 S490 N491 N492 T493 K494 N495 Y496 S497 F498 S499 T500 G501 T502 T506 P507 G508 N509 S510 G511 N512 A513 G514 T515 I516 T517 A522 G523

5 Refinement protocol and experimental data overview (i)

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

Of the 2000 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
DYANA	refinement	

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	1132
Number of shifts mapped to atoms	1132
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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	607	576	577	82±7
All	All	12140	11520	11540	1642

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 69.

5 of 333 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:498:PHE:CE1	1:A:516:ILE:HG21	1.05	1.86	3	8
1:A:441:TYR:CE1	1:A:469:VAL:HG11	0.97	1.93	11	14
1:A:496:TYR:CD1	1:A:516:ILE:HD11	0.96	1.93	5	17
1:A:479:ALA:HB3	1:A:498:PHE:CE1	0.94	1.96	18	3
1:A:477:LEU:C	1:A:477:LEU:HD13	0.92	1.85	12	3

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	81/104 (78%)	68±1 (84±1%)	4±1 (5±1%)	9±1 (12±1%)	1 7
All	All	1620/2080 (78%)	1354 (84%)	78 (5%)	188 (12%)	1 7

5 of 12 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	435	ASN	20
1	A	453	GLY	20
1	A	461	ILE	20
1	A	462	SER	20
1	A	495	ASN	20

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	63/77 (82%)	48±2 (76±3%)	15±2 (24±3%)	2 26
All	All	1260/1540 (82%)	955 (76%)	305 (24%)	2 26

5 of 39 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	458	ASP	20
1	A	477	LEU	20
1	A	499	SER	20
1	A	515	THR	20
1	A	517	THR	19

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)

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

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

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	1132
Number of shifts mapped to atoms	1132
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	10

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	104	0.08 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	88	-0.45 ± 0.19	None needed (< 0.5 ppm)
$^{13}C'$	102	0.41 ± 0.15	None needed (< 0.5 ppm)
^{15}N	96	-0.69 ± 0.57	None needed (imprecise)

7.1.3 Completeness of resonance assignments (i)

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. 908 atoms were assigned a chemical shift out of a possible 1002. 0 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	400/404 (99%)	165/166 (99%)	160/162 (99%)	75/76 (99%)
Sidechain	453/486 (93%)	298/319 (93%)	148/153 (97%)	7/14 (50%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	55/112 (49%)	28/53 (53%)	26/57 (46%)	1/2 (50%)
Overall	908/1002 (91%)	491/538 (91%)	334/372 (90%)	83/92 (90%)

7.1.4 Statistically unusual chemical shifts (i)

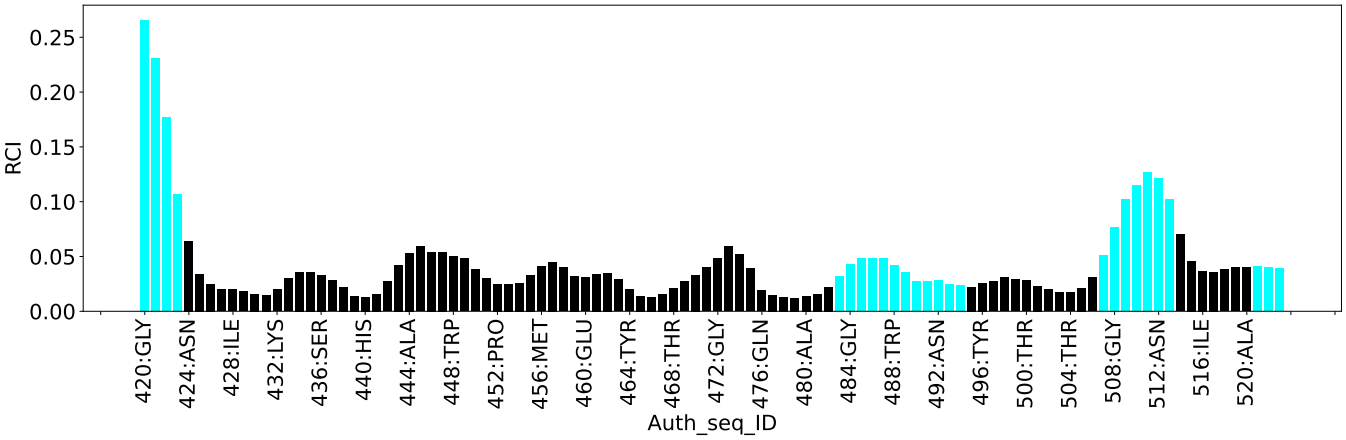
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	496	TYR	CZ	105.25	143.41 – 170.10	-19.3
1	A	442	ARG	HD3	1.35	1.81 – 4.39	-6.8
1	A	482	ASN	HD22	3.96	4.69 – 9.61	-6.5
1	A	442	ARG	HB3	0.04	0.43 – 3.11	-6.4
1	A	442	ARG	HB2	0.21	0.52 – 3.08	-6.2
1	A	442	ARG	HG3	-0.08	0.15 – 2.94	-5.8
1	A	483	ASP	HB3	1.11	1.32 – 4.00	-5.8
1	A	482	ASN	HB2	1.08	1.27 – 4.34	-5.6
1	A	521	PRO	HD3	1.60	1.76 – 5.48	-5.4
1	A	482	ASN	HB3	1.03	1.12 – 4.38	-5.3

7.1.5 Random Coil Index (RCI) plots (i)

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
0.25
0.20
0.15
0.10
0.05
0.00
420:GLY
424:ASN
428:ILE
432:LYS
436:SER
440:HIS
444:ALA
448:TRP
452:PRO
456:MET
460:GLU
464:TYR
468:THR
472:GLY
476:GLN
480:ALA
484:GLY
488:TRP
492:ASN
496:TYR
500:THR
504:THR
508:GLY
512:ASN
516:ILE
520:ALA
Auth_seq_ID