wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 01:39 PM BST

PDB ID	:	1ACZ
Title	:	GLUCOAMYLASE, GRANULAR STARCH-BINDING DOMAIN COMPLEX WITH CYCLODEXTRIN, NMR, 5 STRUCTURES
Authors	:	Sorimachi, K.; Le Gal-Coeffet, M.-F.; Williamson, G.; Archer, D.B.; Williamson, M.P.
Deposited on	:	1997-02-10

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
http://wwpdb.org/validation/2016/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:

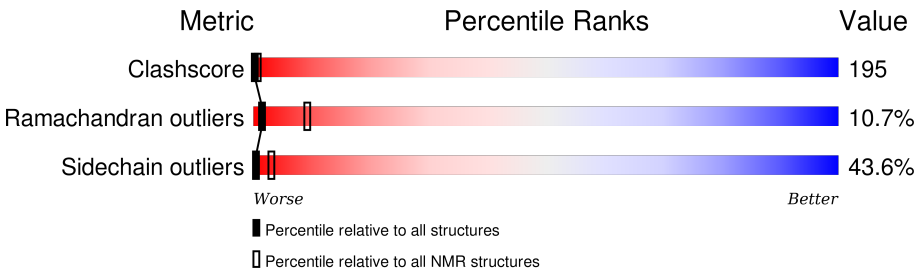
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	108	10% 47% 31% • 11%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	A	GLC	1	2	-
2	A	GLC	2	2	-
2	A	GLC	3	1	-
2	A	GLC	4	2	-
2	A	GLC	6	1	-
2	A	GLC	617(B)	1	-
2	A	GLC	619(B)	1	-

Continued on next page...

Continued from previous page...

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	A	GLC	620(B)	2	-
2	A	GLC	621(B)	1	-
2	A	GLC	622(B)	2	-
2	A	GLC	623(B)	1	-

2 Ensemble composition and analysis (i)

This entry contains 5 models. Model 3 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:513-A:525, A:530-A:555, A:560-A:616 (96)	0.65	3

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 1 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 3, 4
Single-model clusters	2; 5

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1767 atoms, of which 775 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called GLUCOAMYLASE.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	108	1613	527	775	127	182	2	0

- Molecule 2 is a polymer of unknown type called SUGAR (7-MER).

Mol	Chain	Residues	Atoms		
			Total	C	O
2	A	7	77	42	35
			Total	C	O
2	A	7	77	42	35

4 Residue-property plots (i)

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: GLUCOAMYLASE

Chain A: 10% 47% 31% • 11%

C509 T510 T511 P512 V515 A516 V517 T518 F519 D520 L521 T522 A523 T524 T525 T526 Y527 G528 E529 N530 I531 Y532 L533 V534 G535 S536 I537 S538 Q539 L540 G541 D542 W543 E544 T545 S546 D547 G548 I549 A550 L551 S552 A553 D554 K555 Y556 T557 S558 S559 D560 P561 L562 W563 Y564 V565 T566 V567 T568 L569

P570 A571 G572 E573 S574 F575 E576 Y577 K578 F579 I580 R581 I582 E583 S584 D585 D586 S587 V588 E589 W590 E591 S592 D593 P594 N595 R596 E597 Y598 T599 V600 P601 Q602 T606 S607 T608 A609 T610 V611 T612 D613 T614 W615 R616

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: GLUCOAMYLASE

Chain A: 15% 42% 29% • 11%

C509 T510 T511 P512 V515 A516 V517 T518 F519 D520 L521 T522 A523 T524 T525 T526 Y527 G528 E529 N530 I531 Y532 L533 V534 G535 S536 I537 S538 Q539 L540 G541 D542 W543 E544 T545 I549 A550 L551 S552 A553 D554 K555 Y556 T557 S558 S559 D560 P561 L562 W563 Y564 V565 T566 V567 T568 L569 P570 A571

G572 E573 S574 F575 E576 Y577 K578 F579 I580 R581 I582 E583 S584 D585 D586 S587 V588 E589 W590 E591 S592 D593 P594 N595 R596 E597 Y598 T599 V600 P601 Q602 A603 T606 S607 T608 A609 T610 V611 T612 D613 T614 W615 R616

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING*.

Of the 100 calculated structures, 5 were deposited, based on the following criterion: *RANDOM FROM 81 GOOD STRUCTURES*.

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. No validations of the models with respect to experimental NMR restraints is performed at this time.

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: GLC

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	1.14±0.01	0±0/766 (0.0±0.0%)	1.12±0.02	0±0/1052 (0.0±0.0%)
2	A	0.00±0.00	-	0.00±0.00	-
All	All	1.14	0/3830 (0.0%)	1.12	2/5260 (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	2.8±0.4
2	A	3.2±1.6	0.0±0.0
All	All	16	14

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	532	TYR	CB-CG-CD2	-6.22	117.27	121.00	5	1
1	A	534	VAL	CG1-CB-CG2	-5.04	102.84	110.90	4	1

5 of 11 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	2	GLC	C1	2
2	A	4	GLC	C1	2
2	A	1	GLC	C1	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	622(B)	GLC	C1	2
2	A	620(B)	GLC	C1	2

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	596	ARG	Sidechain	5
1	A	581	ARG	Sidechain	5
1	A	616	ARG	Sidechain	4

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	748	697	697	323±35
2	A	154	0	126	32±15
All	All	4510	3485	4115	1679

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 195.

5 of 1070 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:537:ILE:HG21	1:A:569:LEU:HD21	1.13	1.18	2	4
1:A:532:TYR:CZ	1:A:550:ALA:HB2	1.13	1.79	1	4
1:A:515:VAL:HG21	1:A:600:VAL:HG13	1.11	1.22	2	1
1:A:545:THR:HG22	1:A:580:ILE:HD12	1.09	1.10	3	5
1:A:534:VAL:HG22	1:A:542:ASP:CA	1.06	1.81	3	4

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	95/108 (88%)	55±5 (58±5%)	29±6 (31±6%)	10±2 (11±2%)	1 9
All	All	475/540 (88%)	277 (58%)	147 (31%)	51 (11%)	1 9

5 of 23 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	611	VAL	5
1	A	563	TRP	5
1	A	590	TRP	4
1	A	565	VAL	4
1	A	560	ASP	4

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	83/94 (88%)	47±5 (56±5%)	36±5 (44±5%)	0 3
All	All	415/470 (88%)	234 (56%)	181 (44%)	0 3

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	533	LEU	5
1	A	552	SER	5
1	A	614	THR	5
1	A	517	VAL	5
1	A	611	VAL	5

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)

14 carbohydrates are 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	GLC	A	1	2	11,11,12	0.46±0.04	0±0 (0±0%)
2	GLC	A	2	2	11,11,12	0.45±0.02	0±0 (0±0%)
2	GLC	A	3	2	11,11,12	0.55±0.14	0±0 (0±0%)
2	GLC	A	4	2	11,11,12	0.45±0.04	0±0 (0±0%)
2	GLC	A	5	2	11,11,12	0.43±0.02	0±0 (0±0%)
2	GLC	A	6	2	11,11,12	0.42±0.03	0±0 (0±0%)
2	GLC	A	617(B)	2	11,11,12	0.70±0.14	0±0 (0±0%)
2	GLC	A	618(B)	2	11,11,12	0.46±0.03	0±0 (0±0%)
2	GLC	A	619(B)	2	11,11,12	0.49±0.07	0±0 (0±0%)
2	GLC	A	620(B)	2	11,11,12	0.48±0.11	0±0 (0±0%)
2	GLC	A	621(B)	2	11,11,12	0.43±0.02	0±0 (0±0%)
2	GLC	A	622(B)	2	11,11,12	0.45±0.05	0±0 (0±0%)
2	GLC	A	623(B)	2	11,11,12	0.56±0.17	0±0 (0±0%)
2	GLC	A	7	2	11,11,12	0.45±0.05	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	GLC	A	1	2	15,15,17	0.60±0.12	0±0 (0±0%)
2	GLC	A	2	2	15,15,17	0.55±0.08	0±0 (0±0%)

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	GLC	A	3	2	15,15,17	1.06±0.59	0±0 (1±2%)
2	GLC	A	4	2	15,15,17	0.56±0.09	0±0 (0±0%)
2	GLC	A	5	2	15,15,17	0.47±0.01	0±0 (0±0%)
2	GLC	A	6	2	15,15,17	0.55±0.07	0±0 (0±0%)
2	GLC	A	617(B)	2	15,15,17	1.27±0.42	0±0 (1±2%)
2	GLC	A	618(B)	2	15,15,17	0.74±0.23	0±0 (0±0%)
2	GLC	A	619(B)	2	15,15,17	0.86±0.27	0±0 (0±0%)
2	GLC	A	620(B)	2	15,15,17	0.69±0.35	0±0 (0±0%)
2	GLC	A	621(B)	2	15,15,17	0.56±0.09	0±0 (0±0%)
2	GLC	A	622(B)	2	15,15,17	0.66±0.27	0±0 (0±0%)
2	GLC	A	623(B)	2	15,15,17	0.77±0.36	0±0 (0±0%)
2	GLC	A	7	2	15,15,17	0.59±0.17	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	GLC	A	1	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	2	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	3	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	4	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	5	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	6	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	617(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	618(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	619(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	620(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	621(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	622(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	623(B)	2	-	0±0,2,19,22	0±0,1,1,1
2	GLC	A	7	2	-	0±0,2,19,22	0±0,1,1,1

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
2	A	3	GLC	C1-C2-C3	6.88	117.89	109.55	2	1
2	A	617(B)	GLC	C1-C2-C3	6.08	116.92	109.55	5	1

5 of 11 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	1	GLC	C1	2
2	A	4	GLC	C1	2
2	A	620(B)	GLC	C1	2
2	A	2	GLC	C1	2
2	A	622(B)	GLC	C1	2

There are no torsion outliers.

There are no ring outliers.

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)

No chemical shift data were provided