Full wwPDB NMR Structure Validation Report (i)

Jun 12, 2024 – 11:40 PM EDT

PDB ID	:	1KYJ
Title	:	Tumor Associated Mucin Motif from CD43 protein
Authors	:	Coltart, D.M.; Williams, L.J.; Glunz, P.W.; Sames, D.; Kuduk, S.D.; Schwarz, J.B.; Chen, X.-T.; Royyuru, A.K.; Danishefsky, S.D.; Live, D.H.
Deposited on	:	2002-02-04

The following versions of software and data (see references (i)) were used in the production of this report:

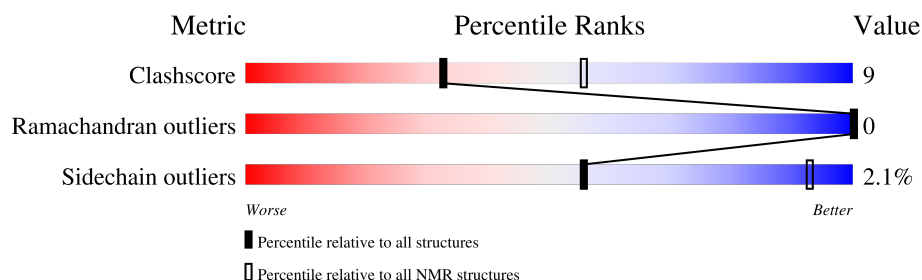|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1   Overall quality at a glance ⓘ

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	Percentile Ranks	Value
Clashscore		9
Ramachandran outliers		0
Sidechain outliers		2.1%

Worse *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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 seg-
ment 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	6	83% 17%

2 Ensemble composition and analysis (i)

This entry contains 59 models.

Cyrange was unable to find well-defined residues.

Error message: The number of core atoms (5) was below the domain threshold value (8).

NmrClust was unable to cluster the ensemble.

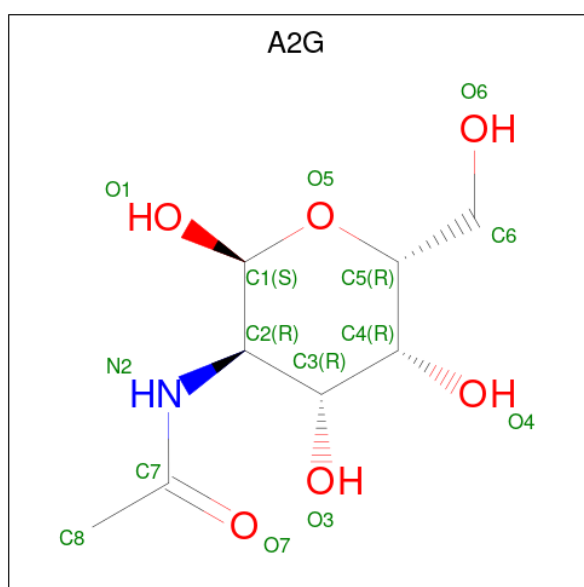Error message: Wrapper check: not enough residues in core to run NmrClust

3 Entity composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 144 atoms, of which 69 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Leukosialin (CD43) fragment.

Mol	Chain	Residues	Atoms					Trace
1	A	6	Total	C	H	N	O	0
			69	21	33	5	10	

- Molecule 2 is 2-acetamido-2-deoxy-alpha-D-galactopyranose (three-letter code: A2G) (formula: $C_8H_{15}NO_6$).
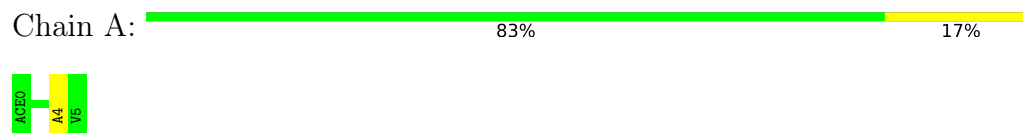
Mol	Chain	Residues	Atoms				
2	A	1	Total	C	H	N	O
			25	8	12	1	4
2	A	1	Total	C	H	N	O
			25	8	12	1	4
2	A	1	Total	C	H	N	O
			25	8	12	1	4

4 Residue-property plots ⓘ

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.
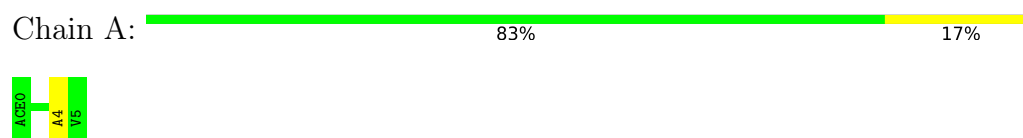
• Molecule 1: Leukosialin (CD43) fragment

Chain A:

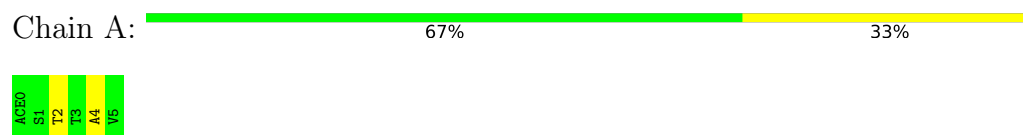## 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 83% 17%

4.2.4 Score per residue for model 4

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 100%

There are no outlier residues in this chain.

4.2.5 Score per residue for model 5

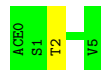• Molecule 1: Leukosialin (CD43) fragment

Chain A: 83% 17%

4.2.6 Score per residue for model 6

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 83% 17%

4.2.7 Score per residue for model 7

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 100%

There are no outlier residues in this chain.

4.2.8 Score per residue for model 8
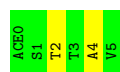
• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.9 Score per residue for model 9

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.12 Score per residue for model 12

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.13 Score per residue for model 13

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

67%   33%

4.2.14 Score per residue for model 14

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.15 Score per residue for model 15

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.16 Score per residue for model 16

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.17 Score per residue for model 17

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.18　Score per residue for model 18

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.19　Score per residue for model 19

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

67%　　　　　33%

ACE0 A4 V5

4.2.20　Score per residue for model 20

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%　　　　　17%

ACE0 A4 V5

4.2.21　Score per residue for model 21

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

67%　　　　　33%

ACE0 S1 T2 T3 A4 V5

4.2.22　Score per residue for model 22

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%　　　　　17%

ACE0 T3 A4 V5

4.2.23 Score per residue for model 23

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 83%   17%

4.2.24 Score per residue for model 24

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 50%   50%

4.2.25 Score per residue for model 25

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 100%

There are no outlier residues in this chain.

4.2.26 Score per residue for model 26

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 50%   50%

4.2.27 Score per residue for model 27

• Molecule 1: Leukosialin (CD43) fragment

Chain A: 83%   17%

4.2.28 Score per residue for model 28
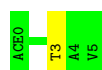
• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.29 Score per residue for model 29

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

There are no outlier residues in this chain.

4.2.30 Score per residue for model 30

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.31 Score per residue for model 31

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

There are no outlier residues in this chain.

4.2.32 Score per residue for model 32

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.33 Score per residue for model 33

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.34 Score per residue for model 34

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.35 Score per residue for model 35

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

There are no outlier residues in this chain.

4.2.36 Score per residue for model 36
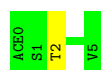
• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.37 Score per residue for model 37

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.38 Score per residue for model 38

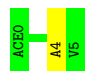• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.39 Score per residue for model 39

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.40 Score per residue for model 40

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.41 Score per residue for model 41

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

67%   33%

4.2.42 Score per residue for model 42

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.43 Score per residue for model 43

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.44 Score per residue for model 44
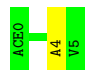
• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.45 Score per residue for model 45

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

67%   33%

4.2.46 Score per residue for model 46

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

67%   33%

4.2.47 Score per residue for model 47

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.48 Score per residue for model 48

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

ACE0   A4   V5

4.2.49 Score per residue for model 49

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

ACE0   T3 A4 V5

4.2.50 Score per residue for model 50

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

100%

There are no outlier residues in this chain.

4.2.51 Score per residue for model 51

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

ACE0   A4 V5

4.2.52 Score per residue for model 52

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

ACE0 S1 T2   V5

4.2.53   Score per residue for model 53

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.54   Score per residue for model 54

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.55   Score per residue for model 55

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.56   Score per residue for model 56

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

4.2.57   Score per residue for model 57

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

There are no outlier residues in this chain.

4.2.58 Score per residue for model 58

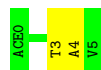• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

4.2.59 Score per residue for model 59

• Molecule 1: Leukosialin (CD43) fragment

Chain A:

83%   17%

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *VNMR, NMRPIPE, X-PLOR, Torsion Space Simulated Annealing.*

Of the 200 calculated structures, 59 were deposited, based on the following criterion: *No NOE violations greter than 0.15A, no coupling violation greater than 0.5 Hz, no angle violation greater than 5 deg..*

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

Software name	Classification	Version
X-PLOR	refinement	98

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: ACE, A2G

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	36	33	33	1±1
2	A	39	36	27	1±1
All	All	4425	4071	3540	74

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 9.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:4:ALA:HB2	2:A:10:A2G:O7	0.79	1.77	33	26
1:A:4:ALA:HB2	2:A:10:A2G:H8	0.66	1.68	54	9
1:A:4:ALA:HB2	2:A:10:A2G:C8	0.63	2.22	41	3
1:A:3:THR:O	2:A:13:A2G:H8A	0.59	1.98	55	12
1:A:2:THR:O	2:A:7:A2G:H8A	0.56	2.00	42	9
1:A:4:ALA:HB2	2:A:10:A2G:C7	0.51	2.35	16	12
1:A:2:THR:C	2:A:7:A2G:H8A	0.49	2.28	33	1
1:A:2:THR:O	2:A:10:A2G:H8A	0.45	2.11	2	2

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	4/6 (67%)	3±1 (67±19%)	1±1 (33±19%)	0±0 (0±0%)	100	100
All	All	236/354 (67%)	159 (67%)	77 (33%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 Protein sidechains ⓘ

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	4/4 (100%)	4±0 (98±7%)	0±0 (2±7%)	56	93
All	All	236/236 (100%)	231 (98%)	5 (2%)	56	93

All 2 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	5	VAL	4
1	A	2	THR	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 monosaccharides in this entry.

6.6 Ligand geometry ⓘ

3 ligands 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	A2G	A	7	1	13,13,15	1.07±0.05	0±0 (0±0%)
2	A2G	A	13	1	13,13,15	1.08±0.04	0±0 (0±0%)
2	A2G	A	10	1	13,13,15	1.05±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	A2G	A	7	1	16,18,21	1.29±0.04	2±0 (13±2%)
2	A2G	A	13	1	16,18,21	1.33±0.04	2±1 (14±3%)
2	A2G	A	10	1	16,18,21	1.31±0.03	1±1 (9±3%)

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	A2G	A	10	1	-	0±0,4,21,26	0±0,1,1,1
2	A2G	A	13	1	-	0±0,4,21,26	0±0,1,1,1

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	A2G	A	7	1	-	0±0,4,21,26	0±0,1,1,1

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models Worst	Total
2	A	13	A2G	C4-C3	2.09	1.57	1.52	5	1

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	13	A2G	C2-N2-C7	3.13	118.44	122.90	14	59
2	A	7	A2G	C2-N2-C7	3.03	118.59	122.90	33	59
2	A	10	A2G	C2-N2-C7	2.87	118.82	122.90	16	59
2	A	7	A2G	C1-C2-N2	2.81	105.68	110.49	9	11
2	A	13	A2G	C4-C3-C2	2.60	114.83	111.02	59	59
2	A	7	A2G	C4-C3-C2	2.51	114.69	111.02	57	52
2	A	13	A2G	C3-C4-C5	2.50	113.67	109.77	58	16
2	A	7	A2G	C3-C4-C5	2.49	113.65	109.77	57	2
2	A	10	A2G	C3-C4-C5	2.43	113.56	109.77	34	5
2	A	13	A2G	C1-C2-N2	2.42	106.36	110.49	55	3
2	A	10	A2G	C4-C3-C2	2.38	114.50	111.02	13	16
2	A	10	A2G	C1-C2-N2	2.27	106.61	110.49	37	6
2	A	13	A2G	C8-C7-N2	2.01	112.70	116.10	4	1
2	A	10	A2G	C8-C7-N2	2.00	112.71	116.10	54	1

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided