wwPDB NMR Structure Validation Summary Report (i)

Dec 17, 2023 – 12:43 PM EST

PDB ID	:	4AYK
Title	:	CATALYTIC FRAGMENT OF HUMAN FIBROBLAST COLLAGENASE COMPLEXED WITH CGS-27023A, NMR, 30 STRUCTURES
Authors	:	Powers, R.; Moy, F.J.
Deposited on	:	1999-02-01

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:

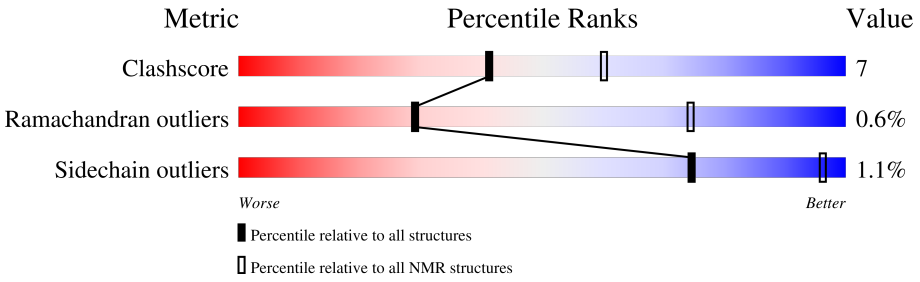
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 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	169	82% 7% 11%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 30 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:8-A:140, A:146-A:163 (151)	0.54	30

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	3, 4, 8, 9, 13, 14, 15, 19, 23, 30
2	5, 11, 16, 18, 21, 24, 26, 28
3	1, 2, 6, 12, 27
4	10, 25
5	17, 22
Single-model clusters	7; 20; 29

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2618 atoms, of which 1253 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN (COLLAGENASE).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	169	2565	834	1230	237	262	2	0

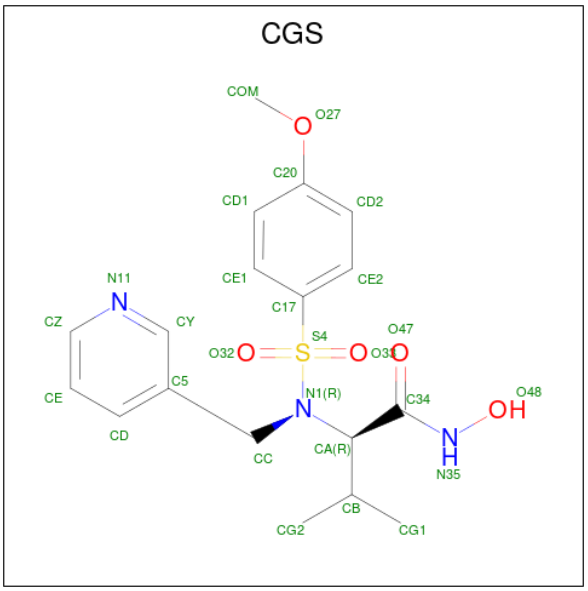
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
			Total	Ca
3	A	1	1	1

- Molecule 4 is N-HYDROXY-2(R)-[[(4-METHOXYPHENYL)SULFONYL](3-PICOLYL)AMINO]-3-METHYLBUTANAMIDE HYDROCHLORIDE (three-letter code: CGS) (formula: $C_{18}H_{23}N_3O_5S$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	S
4	A	1	50	18	23	3	5	1

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: PROTEIN (COLLAGENASE)

Chain A: 82% 7% 11%

V1 L2 T3 E4 G5 N6 P7 H13 T48 H68 R69 D70 N80 N106 F107 R108 E109 H118 E119 T141 F142 S143 G144 D145 Q157 G161 Q164 N165 P166 V167 Q168 P169

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

The representative model is number 30. Colouring as in section 4.1 above.

• Molecule 1: PROTEIN (COLLAGENASE)

Chain A: 78% 11% • 11%

V1 L2 T3 E4 G5 N6 P7 E19 N20 F38 F49 H68 R69 D70 N71 S72 T104 N105 N106 F107 R108 E109 Y110 N111 R114 V115 H118 E119 P138 T141 F142 S143 G144 D145 Q157 G161 Q164 N165 P166 V167 Q168 P169

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *HYBRID DISTANCE GEOMETRY-DYNAMICAL SIMULATED ANNEALING METHOD.*

Of the 100 calculated structures, 30 were deposited, based on the following criterion: *LOWEST ENERGY, LOWEST VIOLATIONS, CONSISTENT FOLD.*

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

Software name	Classification	Version
X-PLOR	refinement	3.84
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: ZN, CA, CGS

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	1201	1106	1106	17±5
4	A	27	23	22	3±2
All	All	36930	33870	33840	523

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 7.

5 of 186 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:45:THR:HG22	1:A:157:GLN:NE2	0.84	1.87	15	1
1:A:80:ASN:ND2	4:A:173:CGS:HG12	0.71	2.00	3	8
1:A:32:HIS:CE1	1:A:36:LYS:NZ	0.62	2.68	26	3
1:A:106:ASN:HD21	1:A:108:ARG:HE	0.62	1.36	15	2
1:A:45:THR:HG22	1:A:157:GLN:HE21	0.62	1.52	15	1

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	151/169 (89%)	139±1 (92±1%)	11±2 (7±1%)	1±1 (1±1%)	29 74
All	All	4530/5070 (89%)	4183 (92%)	322 (7%)	25 (1%)	29 74

5 of 8 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	138	PRO	8
1	A	109	GLU	7
1	A	26	PRO	3
1	A	89	PRO	2
1	A	107	PHE	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	125/141 (89%)	124±1 (99±1%)	1±1 (1±1%)	74 96
All	All	3750/4230 (89%)	3708 (99%)	42 (1%)	74 96

5 of 13 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	106	ASN	15
1	A	83	HIS	13
1	A	111	ASN	3
1	A	80	ASN	2
1	A	39	GLN	1

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)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

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
4	CGS	A	173	2	28,28,28	3.00±0.03	13±1 (45±2%)

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
4	CGS	A	173	2	37,39,39	1.46±0.04	3±1 (8±1%)

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
4	CGS	A	173	2	-	0±0,32,32,32	0±0,2,2,2

5 of 15 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
4	A	173	CGS	S4-N1	10.00	1.78	1.63	19	30
4	A	173	CGS	CE2-C17	4.57	1.46	1.38	4	30
4	A	173	CGS	CD2-C20	4.45	1.47	1.38	13	30
4	A	173	CGS	CD1-C20	4.39	1.47	1.38	10	30
4	A	173	CGS	CE1-C17	4.22	1.45	1.38	29	30

5 of 7 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
4	A	173	CGS	CB-CA-N1	5.43	122.27	112.99	25	30
4	A	173	CGS	C5-CC-N1	4.06	120.45	113.48	24	30
4	A	173	CGS	O33-S4-N1	2.54	102.33	106.97	4	23
4	A	173	CGS	O32-S4-C17	2.23	110.86	108.05	6	2
4	A	173	CGS	O32-S4-N1	2.07	103.19	106.97	17	2

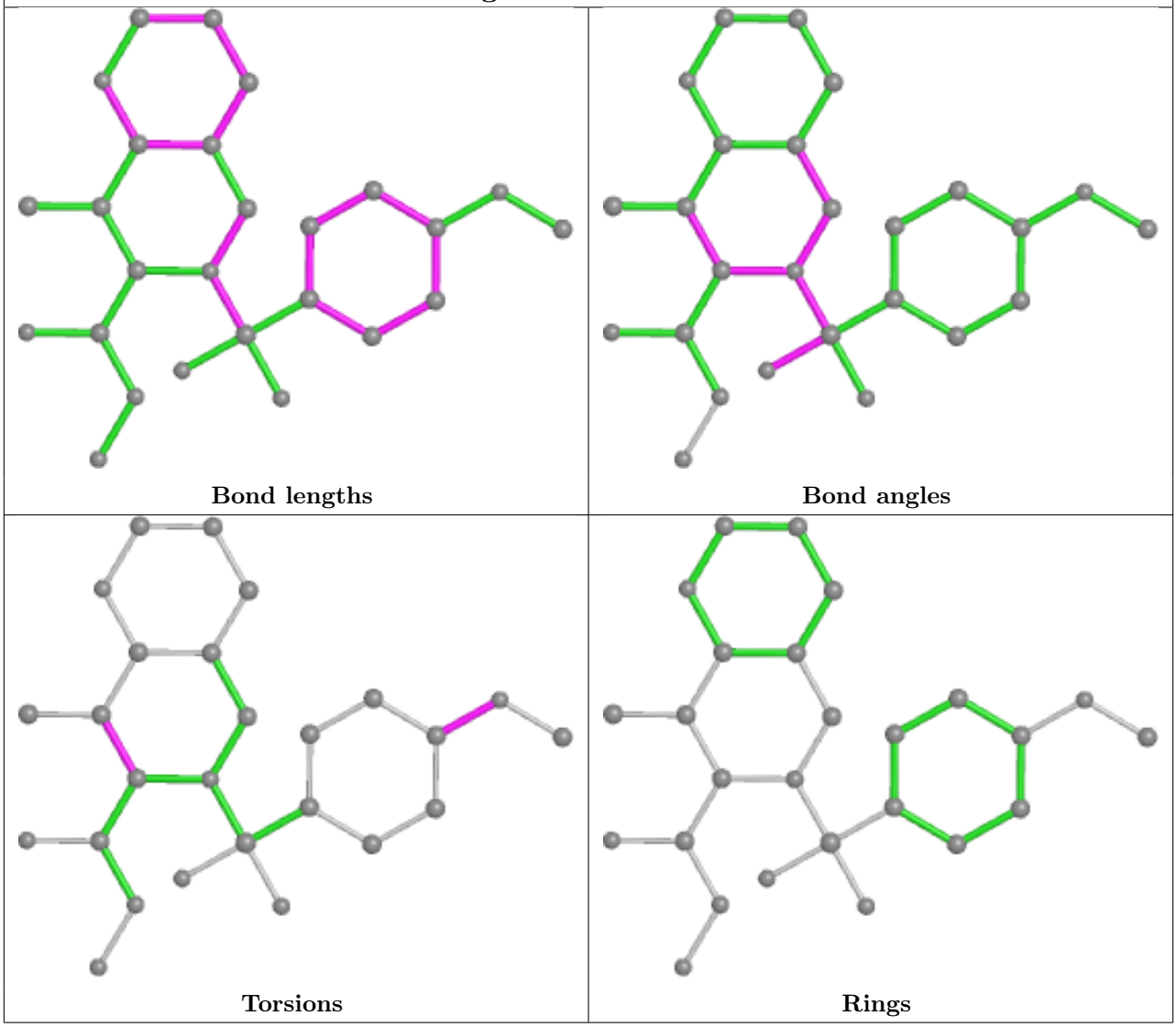
There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand CGS A 173



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