Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 08:08 AM EDT

PDB ID	:	6COO
BMRB ID	:	30427
Title	:	CSP1-E1A
Authors	:	Yang, Y.
Deposited on	:	2018-03-12

This is a Full wwPDB NMR Structure Validation 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:

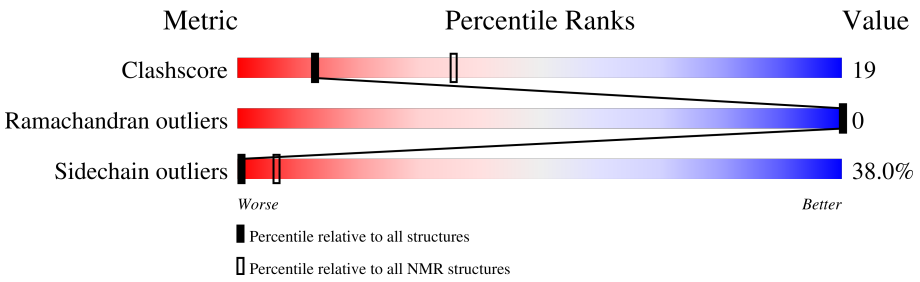
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 59%.

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	17	18% 41% 18% 12% 12%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 5 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:3-A:17 (15)	0.12	5

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 4 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 325 atoms, of which 171 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Competence-stimulating peptide type 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	17	325	101	171	30	22	1	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	ALA	GLU	engineered mutation	UNP P60242

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: Competence-stimulating peptide type 1

Chain A: 18% 41% 18% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

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: Competence-stimulating peptide type 1

Chain A: 18% 29% 24% 18% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.2 Score per residue for model 2

• Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 29% 18% 18% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.3 Score per residue for model 3

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 29% 35% 12% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 K17

4.2.4 Score per residue for model 4

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 18% 35% 18% 18% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.5 Score per residue for model 5 (medoid)

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 35% 12% 18% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.6 Score per residue for model 6

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 35% 18% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.7 Score per residue for model 7

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 18% 35% 18% 18% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.8 Score per residue for model 8

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 41% 12% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.9 Score per residue for model 9

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 18% 41% 18% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.10 Score per residue for model 10

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 29% 29% 18% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 K16 K17

4.2.11 Score per residue for model 11

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 35% 18% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 K16 K17

4.2.12 Score per residue for model 12

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 18% 41% 24% 6% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.13 Score per residue for model 13

• Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 35% 18% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.14 Score per residue for model 14

• Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 29% 24% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.15 Score per residue for model 15

• Molecule 1: Competence-stimulating peptide type 1

Chain A: 24% 35% 24% 6% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.16 Score per residue for model 16

• Molecule 1: Competence-stimulating peptide type 1

Chain A: 29% 35% 12% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 K16 K17

4.2.17 Score per residue for model 17

• Molecule 1: Competence-stimulating peptide type 1

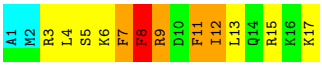
Chain A: 18% 47% 12% 12% 12%

A1 M2 R3 L4 S5 K6 F7 F8 R9 D10 F11 I12 L13 Q14 R15 K16 K17

4.2.18 Score per residue for model 18

- Molecule 1: Competence-stimulating peptide type 1

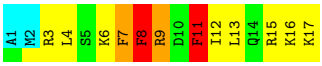
Chain A: 18% 41% 24% 6% 12%



4.2.19 Score per residue for model 19

- Molecule 1: Competence-stimulating peptide type 1

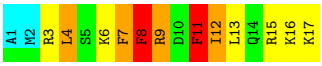
Chain A: 18% 47% 12% 12% 12%



4.2.20 Score per residue for model 20

- Molecule 1: Competence-stimulating peptide type 1

Chain A: 18% 35% 24% 12% 12%



5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	**Classification**	**Version**
X-PLOR NIH	structure calculation	
X-PLOR NIH	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	160
Number of shifts mapped to atoms	160
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	59%

6 Model quality (i)

6.1 Standard geometry (i)

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	2.37±0.02	8±1/143 (5.8± 0.7%)	3.29±0.03	14±1/184 (7.4± 0.7%)
All	All	2.37	167/2860 (5.8%)	3.29	272/3680 (7.4%)

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	1.1±0.3
All	All	0	22

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
1	A	8	PHE	CD1-CE1	-9.48	1.20	1.39	10	20
1	A	9	ARG	CA-C	-6.49	1.36	1.52	7	20
1	A	8	PHE	CB-CG	-6.03	1.41	1.51	10	18
1	A	11	PHE	CD1-CE1	-5.49	1.28	1.39	12	20
1	A	8	PHE	CG-CD1	-5.40	1.30	1.38	9	17
1	A	8	PHE	CG-CD2	-5.38	1.30	1.38	4	18
1	A	9	ARG	NE-CZ	5.31	1.40	1.33	17	20
1	A	3	ARG	NE-CZ	5.29	1.40	1.33	4	18
1	A	8	PHE	C-O	-5.18	1.13	1.23	12	1
1	A	15	ARG	NE-CZ	5.17	1.39	1.33	19	15

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	7	PHE	CB-CG-CD2	-24.85	103.40	120.80	12	20
1	A	8	PHE	CB-CG-CD1	-20.10	106.73	120.80	16	20
1	A	7	PHE	CB-CG-CD1	-14.88	110.38	120.80	14	20
1	A	11	PHE	CB-CG-CD1	-13.83	111.12	120.80	5	20
1	A	7	PHE	N-CA-CB	11.61	131.50	110.60	6	18
1	A	11	PHE	CB-CG-CD2	11.33	128.73	120.80	14	20
1	A	7	PHE	CB-CA-C	-10.79	88.82	110.40	5	18
1	A	8	PHE	CB-CG-CD2	9.19	127.23	120.80	16	20
1	A	8	PHE	N-CA-C	-7.46	90.86	111.00	18	20
1	A	8	PHE	N-CA-CB	7.38	123.88	110.60	1	20
1	A	7	PHE	CA-CB-CG	6.44	129.37	113.90	8	18
1	A	11	PHE	CA-CB-CG	5.78	127.78	113.90	17	20
1	A	11	PHE	CB-CA-C	-5.60	99.21	110.40	17	20
1	A	8	PHE	CG-CD2-CE2	-5.52	114.72	120.80	17	18

There are no chirality outliers.

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	8	PHE	Sidechain	20
1	A	7	PHE	Sidechain	2

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	141	155	155	6±1
All	All	2820	3100	3100	110

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 19.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:8:PHE:O	1:A:11:PHE:HB2	0.93	1.63	9	20

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:9:ARG:O	1:A:13:LEU:HG	0.80	1.77	1	20
1:A:8:PHE:C	1:A:8:PHE:CD1	0.58	2.76	1	20
1:A:8:PHE:O	1:A:12:ILE:HD13	0.48	2.08	18	11
1:A:9:ARG:O	1:A:13:LEU:CG	0.46	2.60	1	4
1:A:7:PHE:CG	1:A:8:PHE:N	0.45	2.84	5	18
1:A:11:PHE:HB2	1:A:12:ILE:HD13	0.44	1.90	18	1
1:A:5:SER:O	1:A:9:ARG:HB2	0.43	2.12	12	2
1:A:9:ARG:HE	1:A:13:LEU:HD21	0.43	1.73	2	4
1:A:11:PHE:O	1:A:15:ARG:HG3	0.43	2.14	12	1
1:A:4:LEU:HD13	1:A:7:PHE:CD2	0.42	2.49	9	9

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	14/17 (82%)	14±0 (100±2%)	0±0 (0±2%)	0±0 (0±0%)	100	100
All	All	280/340 (82%)	279 (100%)	1 (0%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	15/16 (94%)	9±1 (62±7%)	6±1 (38±7%)	1	6
All	All	300/320 (94%)	186 (62%)	114 (38%)	1	6

All 9 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	6	LYS	20
1	A	12	ILE	20
1	A	4	LEU	19
1	A	17	LYS	19
1	A	11	PHE	17
1	A	16	LYS	12
1	A	9	ARG	5
1	A	13	LEU	1
1	A	3	ARG	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)

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 59% for the well-defined parts and 56% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *shiftFile_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	160
Number of shifts mapped to atoms	160
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

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 59%, i.e. 154 atoms were assigned a chemical shift out of a possible 263. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}**H**	^{13}**C**	^{15}**N**
Backbone	60/75 (80%)	30/30 (100%)	15/30 (50%)	15/15 (100%)
Sidechain	88/158 (56%)	73/100 (73%)	15/45 (33%)	0/13 (0%)
Aromatic	6/30 (20%)	6/15 (40%)	0/15 (0%)	0/0 (—%)
Overall	154/263 (59%)	109/145 (75%)	30/90 (33%)	15/28 (54%)

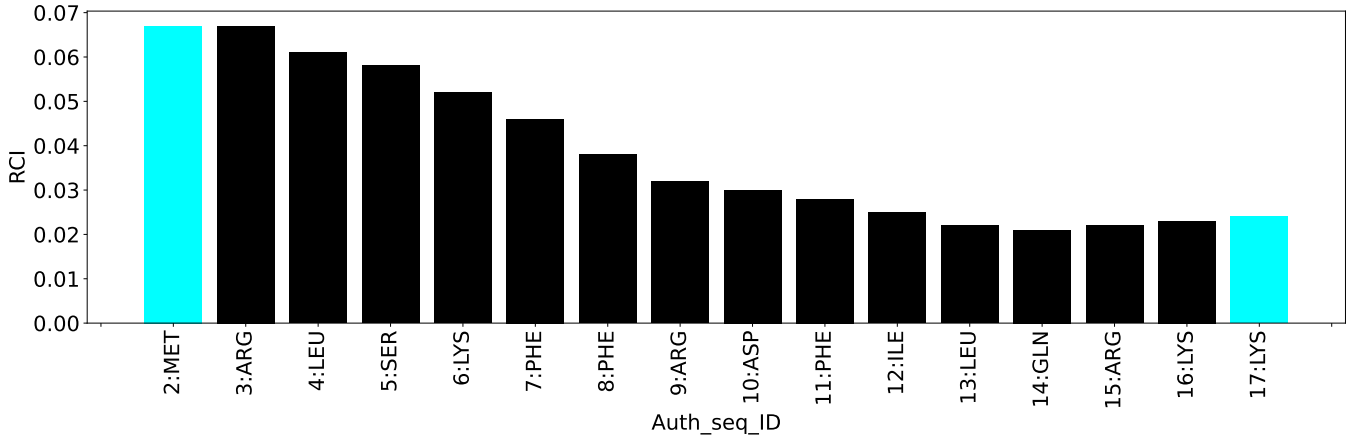
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 56%, i.e. 160 atoms were assigned a chemical shift out of a possible 287. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	62/85 (73%)	31/34 (91%)	16/34 (47%)	15/17 (88%)
Sidechain	92/172 (53%)	76/110 (69%)	16/49 (33%)	0/13 (0%)
Aromatic	6/30 (20%)	6/15 (40%)	0/15 (0%)	0/0 (—%)
Overall	160/287 (56%)	113/159 (71%)	32/98 (33%)	15/30 (50%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: