wwPDB NMR Structure Validation Summary Report

Jun 14, 2020 – 04:59 pm BST

PDB ID	:	2N8R
Title	:	Productive complex between MMP-12 and synthetic triple-helical collagen, revealed through paramagnetic NMR
Authors	:	Prior, S.H.; Van Doren, S.R.
Deposited on	:	2015-10-24

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

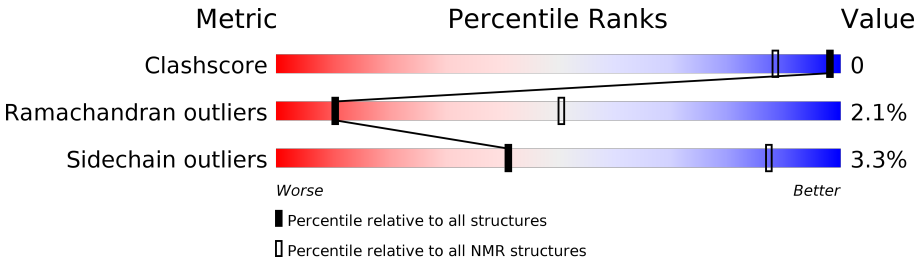
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 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	164	90% 9% .
2	B	36	100%
2	C	36	100%
2	D	36	89% 8% .

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3850 atoms, of which 1854 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Macrophage metalloelastase.

Mol	Chain	Residues	Atoms						Trace
1	A	164	Total	C	H	N	O	S	0
			2513	826	1222	225	236	4	

- Molecule 2 is a protein called Collagen triple helix repeat family protein.

Mol	Chain	Residues	Atoms					Trace
2	B	36	Total	C	H	N	O	0
			443	144	210	38	51	
2	C	36	Total	C	H	N	O	0
			443	144	210	38	51	
2	D	36	Total	C	H	N	O	0
			443	144	210	38	51	

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
3	A	2	Total	Zn
			2	2

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
4	A	3	Total	Ca
			3	3

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		
5	A	1	Total	H	O
			3	2	1

4 Residue-property plots (i)

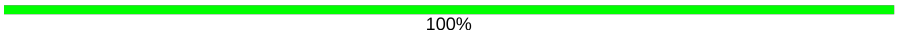
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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: Macrophage metalloelastase

Chain A:

- Molecule 2: Collagen triple helix repeat family protein

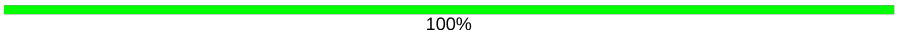
Chain B:



There are no outlier residues in this chain.

- Molecule 2: Collagen triple helix repeat family protein

Chain C:



There are no outlier residues in this chain.

- Molecule 2: Collagen triple helix repeat family protein

Chain D:

5 Refinement protocol and experimental data overview

The models were refined using the following method: *Rigid-body docking, Conformer selection.*

Of the 7500 calculated structures, 1 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum.*

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

Software name	Classification	Version
HADDOCK	structure solution	2.1
q test.py	structure solution	
GROMOS	refinement	
HADDOCK	refinement	2.1
q test.py	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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1753
Number of shifts mapped to atoms	1753
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%

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

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HYP, ZN, CA

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	0.75	0/1331 (0.0%)	1.11	10/1802 (0.6%)
2	B	0.72	0/170 (0.0%)	0.89	0/224 (0.0%)
2	C	0.67	0/170 (0.0%)	0.87	0/224 (0.0%)
2	D	0.72	0/170 (0.0%)	1.12	0/224 (0.0%)
All	All	0.73	0/1841 (0.0%)	1.07	10/2474 (0.4%)

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	2
All	All	0	2

There are no bond-length outliers.

5 of 10 angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	170	ASP	CB-CG-OD1	10.74	127.97	118.30
1	A	124	ASP	CB-CG-OD1	9.38	126.74	118.30
1	A	101	ARG	NE-CZ-NH1	7.74	124.17	120.30
1	A	170	ASP	OD1-CG-OD2	-6.92	110.15	123.30
1	A	256	ARG	NE-CZ-NH1	6.78	123.69	120.30

There are no chirality outliers.

All planar outliers are listed below.

Mol	Chain	Res	Type	Group
1	A	174	PHE	Peptide
1	A	240	TYR	Sidechain

6.2 Too-close contacts

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	1291	1222	1219	1
All	All	1996	1854	1843	1

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)
1:A:237:PHE:CD1	1:A:238:PRO:HD2	0.46	2.46

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	162/164 (99%)	156 (96%)	4 (2%)	2 (1%)	17 64
2	B	27/36 (75%)	26 (96%)	1 (4%)	0 (0%)	100 100
2	C	27/36 (75%)	27 (100%)	0 (0%)	0 (0%)	100 100
2	D	27/36 (75%)	22 (81%)	2 (7%)	3 (11%)	1 8
All	All	243/272 (89%)	231 (95%)	7 (3%)	5 (2%)	10 50

All 5 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
2	D	92	GLU
1	A	171	PHE
1	A	238	PRO
2	D	88	GLY
2	D	89	VAL

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	132/132 (100%)	128 (97%)	4 (3%)	44 89
2	B	16/16 (100%)	16 (100%)	0 (0%)	100 100
2	C	16/16 (100%)	16 (100%)	0 (0%)	100 100
2	D	16/16 (100%)	14 (88%)	2 (12%)	8 50
All	All	180/180 (100%)	174 (97%)	6 (3%)	41 87

5 of 6 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	101	ARG
2	D	92	GLU
1	A	102	GLU
2	D	95	GLU
1	A	131	ASP

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

24 non-standard protein/DNA/RNA residues 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	HYP	D	99	2	6,8,9	0.53	0 (0%)
2	HYP	C	42	2	6,8,9	0.52	0 (0%)
2	HYP	D	84	2	6,8,9	0.54	0 (0%)
2	HYP	C	72	2	6,9,9	0.50	0 (0%)
2	HYP	B	12	2	6,8,9	0.52	0 (0%)
2	HYP	D	81	2	6,8,9	0.55	0 (0%)
2	HYP	C	66	2	6,8,9	0.49	0 (0%)
2	HYP	D	78	2	6,8,9	0.46	0 (0%)
2	HYP	C	63	2	6,8,9	0.59	0 (0%)
2	HYP	C	48	2	6,8,9	0.59	0 (0%)
2	HYP	B	3	2	6,8,9	0.53	0 (0%)
2	HYP	B	30	2	6,8,9	0.56	0 (0%)
2	HYP	D	102	2	6,8,9	0.51	0 (0%)
2	HYP	C	69	2	6,8,9	0.53	0 (0%)
2	HYP	C	39	2	6,8,9	0.48	0 (0%)
2	HYP	B	6	2	6,8,9	0.50	0 (0%)
2	HYP	B	33	2	6,8,9	0.58	0 (0%)
2	HYP	D	108	2	6,9,9	0.41	0 (0%)
2	HYP	D	105	2	6,8,9	0.56	0 (0%)
2	HYP	B	9	2	6,8,9	0.44	0 (0%)
2	HYP	B	27	2	6,8,9	0.54	0 (0%)
2	HYP	B	36	2	6,9,9	0.50	0 (0%)
2	HYP	C	45	2	6,8,9	0.55	0 (0%)
2	HYP	D	75	2	6,8,9	0.45	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	HYP	D	99	2	5,10,12	0.99	0 (0%)
2	HYP	C	42	2	5,10,12	0.74	0 (0%)
2	HYP	D	84	2	5,10,12	1.59	0 (0%)
2	HYP	C	72	2	5,12,12	1.26	0 (0%)

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	HYP	B	12	2	5,10,12	0.89	0 (0%)
2	HYP	D	81	2	5,10,12	0.70	0 (0%)
2	HYP	C	66	2	5,10,12	0.81	0 (0%)
2	HYP	D	78	2	5,10,12	1.08	0 (0%)
2	HYP	C	63	2	5,10,12	0.75	0 (0%)
2	HYP	C	48	2	5,10,12	0.69	0 (0%)
2	HYP	B	3	2	5,10,12	0.69	0 (0%)
2	HYP	B	30	2	5,10,12	0.76	0 (0%)
2	HYP	D	102	2	5,10,12	0.74	0 (0%)
2	HYP	C	69	2	5,10,12	0.73	0 (0%)
2	HYP	C	39	2	5,10,12	0.75	0 (0%)
2	HYP	B	6	2	5,10,12	0.84	0 (0%)
2	HYP	B	33	2	5,10,12	0.71	0 (0%)
2	HYP	D	108	2	5,12,12	0.80	0 (0%)
2	HYP	D	105	2	5,10,12	0.81	0 (0%)
2	HYP	B	9	2	5,10,12	1.41	0 (0%)
2	HYP	B	27	2	5,10,12	1.24	0 (0%)
2	HYP	B	36	2	5,12,12	0.44	0 (0%)
2	HYP	C	45	2	5,10,12	1.15	0 (0%)
2	HYP	D	75	2	5,10,12	1.36	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	HYP	D	99	2	-	0,0,11,13	0,1,1,1
2	HYP	C	42	2	-	0,0,11,13	0,1,1,1
2	HYP	D	84	2	-	0,0,11,13	0,1,1,1
2	HYP	C	72	2	-	0,0,13,13	0,1,1,1
2	HYP	B	12	2	-	0,0,11,13	0,1,1,1
2	HYP	D	81	2	-	0,0,11,13	0,1,1,1
2	HYP	C	66	2	-	0,0,11,13	0,1,1,1
2	HYP	D	78	2	-	0,0,11,13	0,1,1,1
2	HYP	C	63	2	-	0,0,11,13	0,1,1,1
2	HYP	C	48	2	-	0,0,11,13	0,1,1,1
2	HYP	B	3	2	-	0,0,11,13	0,1,1,1
2	HYP	B	30	2	-	0,0,11,13	0,1,1,1
2	HYP	D	102	2	-	0,0,11,13	0,1,1,1
2	HYP	C	69	2	-	0,0,11,13	0,1,1,1
2	HYP	C	39	2	-	0,0,11,13	0,1,1,1
2	HYP	B	6	2	-	0,0,11,13	0,1,1,1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	HYP	B	33	2	-	0,0,11,13	0,1,1,1
2	HYP	D	108	2	-	0,0,13,13	0,1,1,1
2	HYP	D	105	2	-	0,0,11,13	0,1,1,1
2	HYP	B	9	2	-	0,0,11,13	0,1,1,1
2	HYP	B	27	2	-	0,0,11,13	0,1,1,1
2	HYP	B	36	2	-	0,0,13,13	0,1,1,1
2	HYP	C	45	2	-	0,0,11,13	0,1,1,1
2	HYP	D	75	2	-	0,0,11,13	0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 5 are monoatomic - leaving 0 for Mogul analysis.

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

The completeness of assignment taking into account all chemical shift lists is 59% for the well-defined parts and 59% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *chem_shift_list_1*

7.1.1 Bookkeeping

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	1753
Number of shifts mapped to atoms	1753
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	159	-0.25 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	137	0.11 ± 0.18	None needed (< 0.5 ppm)
$^{13}C'$	128	-0.18 ± 0.08	None needed (< 0.5 ppm)
^{15}N	153	-0.75 ± 0.29	Should be applied

7.1.3 Completeness of resonance assignments

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. 1622 atoms were assigned a chemical shift out of a possible 2741. 15 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	752/1166 (64%)	312/459 (68%)	287/496 (58%)	153/211 (73%)
Sidechain	722/1312 (55%)	449/789 (57%)	264/477 (55%)	9/46 (20%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	148/263 (56%)	79/142 (56%)	66/109 (61%)	3/12 (25%)
Overall	1622/2741 (59%)	840/1390 (60%)	617/1082 (57%)	165/269 (61%)

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.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	226	LEU	CG	33.00	32.55 – 21.05	5.4
1	A	214	LEU	CG	33.00	32.55 – 21.05	5.4

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.

Random coil index (RCI) for chain A: