



wwPDB NMR Structure Validation Summary Report ⓘ

Feb 12, 2017 – 10:26 pm GMT

PDB ID : 2JSD
Title : Solution structure of MMP20 complexed with NNGH
Authors : Arendt, Y.; Banci, L.; Bertini, I.; Cantini, F.; Cozzi, R.; Del Conte, R.;
Gonnelli, L.; Structural Proteomics in Europe (SPINE)
Deposited on : 2007-07-03

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 ⓘ symbol.

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

Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

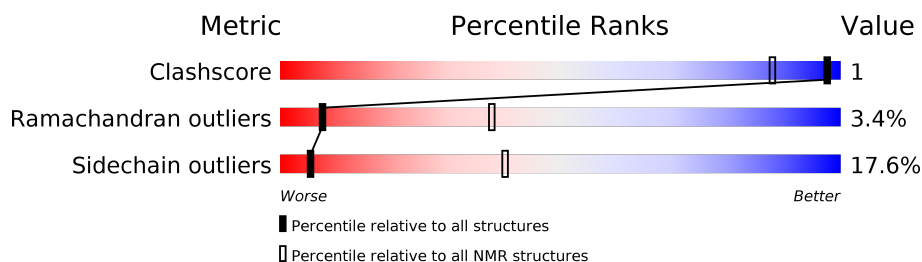
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR


The overall completeness of chemical shifts assignment is 87%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	160	

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
4	A	NGH	277	10	-

2 Ensemble composition and analysis

This entry contains 20 models. Model 18 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:115-A:271 (157)	0.54	18

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 3 clusters. No single-model clusters were found.

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

3 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2446 atoms, of which 1185 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Matrix metalloproteinase-20.

Mol	Chain	Residues	Atoms						Trace
1	A	160	Total	C	H	N	O	S	0
			2402	790	1166	205	236	5	

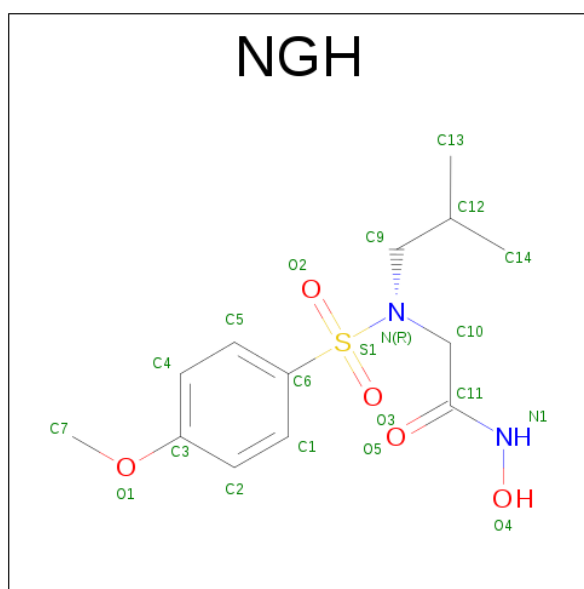
- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
2	A	2	Total	Ca
			2	2

- 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 N-ISOBUTYL-N-[4-METHOXYPHENYLSULFONYL]GLYCYL HYDROX-AMIC ACID (three-letter code: NGH) (formula: C₁₃H₂₀N₂O₅S).



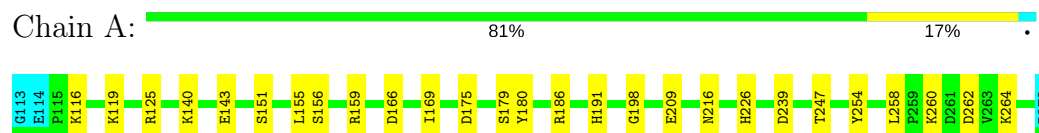
Mol	Chain	Residues	Atoms					
4	A	1	Total	C	H	N	O	S
			40	13	19	2	5	1

4 Residue-property plots [i](#)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA 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.

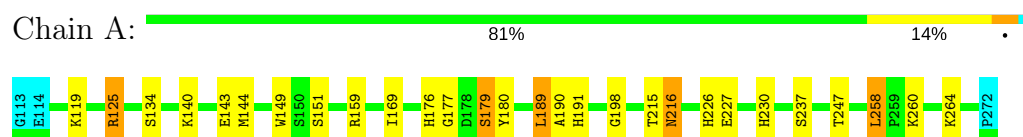
- Molecule 1: Matrix metalloproteinase-20



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

The representative model is number 18. Colouring as in section 4.1 above.

- Molecule 1: Matrix metalloproteinase-20



5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics*.

Of the 400 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
AMBER	refinement	8.0
QUEEN	refinement	
AQUA	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)	BMRB entry 15361
Number of chemical shift lists	1
Total number of shifts	1820
Number of shifts mapped to atoms	1802
Number of unparsed shifts	0
Number of shifts with mapping errors	18
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

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: ZN, NGH, 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	0.76±0.02	0±0/1255 (0.0±0.0%)	1.15±0.02	3±2/1706 (0.2±0.1%)
All	All	0.76	2/25100 (0.0%)	1.15	61/34120 (0.2%)

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	1.5±1.4
All	All	0	30

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	143	GLU	CB-CG	5.82	1.63	1.52	9	1
1	A	209	GLU	CD-OE1	-5.66	1.19	1.25	15	1

5 of 30 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	125	ARG	NE-CZ-NH1	11.39	125.99	120.30	10	4
1	A	125	ARG	NE-CZ-NH2	-10.08	115.26	120.30	11	2
1	A	186	ARG	NE-CZ-NH1	9.17	124.89	120.30	20	5
1	A	159	ARG	NE-CZ-NH1	7.97	124.28	120.30	13	8
1	A	186	ARG	NE-CZ-NH2	-7.05	116.78	120.30	8	1

There are no chirality outliers.

5 of 11 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	226	HIS	Sidechain	9
1	A	230	HIS	Sidechain	5
1	A	191	HIS	Sidechain	3
1	A	157	PHE	Sidechain	3
1	A	124	TYR	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	1215	1148	1148	1±1
4	A	21	19	19	0±1
All	All	24800	23340	23345	29

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 1.

5 of 14 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:189:LEU:HD22	1:A:189:LEU:H	0.61	1.56	18	1
1:A:149:TRP:CE3	1:A:258:LEU:HD21	0.58	2.33	6	3
1:A:149:TRP:CZ3	1:A:258:LEU:HD21	0.58	2.33	6	2
1:A:190:ALA:HB1	1:A:227:GLU:CD	0.57	2.19	14	3
1:A:226:HIS:CD2	4:A:277:NGH:O4	0.52	2.61	20	3

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	157/160 (98%)	134±3 (85±2%)	18±4 (11±3%)	5±2 (3±1%)	7	38
All	All	3140/3200 (98%)	2671 (85%)	361 (11%)	108 (3%)	7	38

5 of 25 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	260	LYS	20
1	A	198	GLY	15
1	A	186	ARG	8
1	A	176	HIS	7
1	A	179	SER	7

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	126/128 (98%)	104±3 (82±2%)	22±3 (18±2%)	5	40
All	All	2520/2560 (98%)	2076 (82%)	444 (18%)	5	40

5 of 79 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	169	ILE	20
1	A	116	LYS	19
1	A	180	TYR	19
1	A	119	LYS	18
1	A	264	LYS	18

6.3.3 RNA ⓘ

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 carbohydrates in this entry.

6.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 4 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	NGH	A	277	-	21,21,21	1.92±0.07	1±0 (4±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
4	NGH	A	277	-	26,29,29	1.57±0.41	0±1 (1±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
4	NGH	A	277	-	-	0±0,24,24,24	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
4	A	277	NGH	O4-N1	8.79	1.24	1.39	13	20

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
4	A	277	NGH	C6-S1-N	6.83	98.98	107.31	2	2
4	A	277	NGH	O3-S1-C6	6.52	116.49	108.00	2	1
4	A	277	NGH	O5-C11-N1	6.43	116.06	123.40	2	2
4	A	277	NGH	O2-S1-C6	5.61	100.68	108.00	13	1

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
4	A	277	NGH	N	10

All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
4	A	277	NGH	C10-C11-N1-O4	1
4	A	277	NGH	O5-C11-N1-O4	1

There are no ring outliers.

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 87% for the well-defined parts and 86% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15361

Chemical shift list name: *assigned_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	1820
Number of shifts mapped to atoms	1802
Number of unparsed shifts	0
Number of shifts with mapping errors	18
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	19

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Chain not found in structure. First 5 (of 18) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	1	NGH	H73	3.79	0.02	2
UNMAPPED	1	NGH	H131	0.928	0.02	1
UNMAPPED	1	NGH	H4	3.543	0.02	1
UNMAPPED	1	NGH	H142	0.928	0.02	1
UNMAPPED	1	NGH	H132	0.928	0.02	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	157	0.20 ± 0.19	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	141	0.44 ± 0.16	None needed (< 0.5 ppm)
$^{13}\text{C}'$	132	0.31 ± 0.18	None needed (< 0.5 ppm)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction \pm precision, ppm	Suggested action
^{15}N	144	0.15 ± 0.40	None needed (< 0.5 ppm)

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 87%, i.e. 1603 atoms were assigned a chemical shift out of a possible 1845. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	736/765 (96%)	303/304 (100%)	289/314 (92%)	144/147 (98%)
Sidechain	753/845 (89%)	469/500 (94%)	278/317 (88%)	6/28 (21%)
Aromatic	114/235 (49%)	102/123 (83%)	1/97 (1%)	11/15 (73%)
Overall	1603/1845 (87%)	874/927 (94%)	568/728 (78%)	161/190 (85%)

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	182	PHE	HE2	0.84	8.69 – 5.49	-19.5
1	A	182	PHE	HE1	0.84	8.69 – 5.49	-19.5
1	A	209	GLU	HB3	-0.04	3.10 – 0.90	-9.3
1	A	146	LEU	HD22	-1.31	2.14 – -0.66	-7.3
1	A	146	LEU	HD23	-1.31	2.14 – -0.66	-7.3
1	A	146	LEU	HD21	-1.31	2.14 – -0.66	-7.3
1	A	195	PRO	HB3	-0.25	3.81 – 0.21	-6.3
1	A	172	GLU	HB3	0.64	3.10 – 0.90	-6.2
1	A	240	PRO	HA	2.38	6.05 – 2.75	-6.1
1	A	206	ASP	HB3	0.99	4.07 – 1.27	-6.0
1	A	146	LEU	HD13	-0.90	2.16 – -0.64	-5.9
1	A	146	LEU	HD12	-0.90	2.16 – -0.64	-5.9
1	A	146	LEU	HD11	-0.90	2.16 – -0.64	-5.9
1	A	186	ARG	H	11.77	11.29 – 5.19	5.8
1	A	189	LEU	HD22	-0.86	2.14 – -0.66	-5.7
1	A	189	LEU	HD21	-0.86	2.14 – -0.66	-5.7
1	A	189	LEU	HD23	-0.86	2.14 – -0.66	-5.7
1	A	209	GLU	HB2	0.85	3.08 – 0.98	-5.6
1	A	172	GLU	HG3	1.16	3.31 – 1.21	-5.3

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:

