wwPDB NMR Structure Validation Summary Report (i)

Jun 10, 2021 – 02:06 PM BST

PDB ID	:	7B2B
Title	:	Solution structure of a non-covalent extended docking domain complex of the Pax NRPS: PaxA T1-CDD/PaxB NDD
Authors	:	Watzel, J.; Sarawi, S.; Duchardt-Ferner, E.; Bode, H.B.; Woehnert, J.
Deposited on	:	2020-11-26

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:

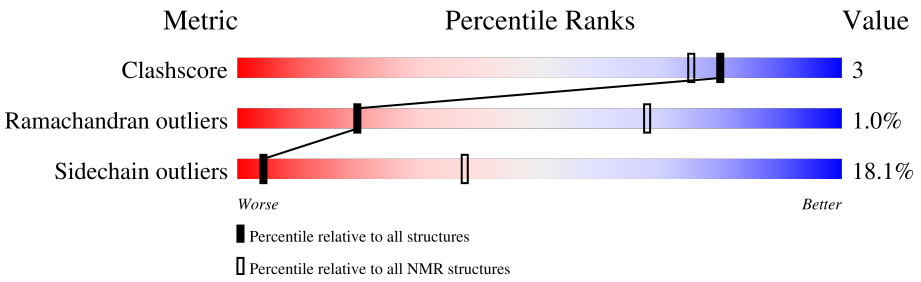
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.20
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.20

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 90%.

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	B	31	74% 10% 16%
2	A	104	76% 12% • 11%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 1 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	B:2-B:27, A:991-A:1083 (119)	0.28	1

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 and 6 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2176 atoms, of which 1082 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Amino acid adenylation domain-containing protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	B	31	520	157	268	49	45	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	23	ALA	THR	variant	UNP A0A3D9UGN9
B	31	TYR	THR	engineered mutation	UNP A0A3D9UGN9

- Molecule 2 is a protein called Peptide synthetase PaxA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	104	1656	534	814	137	168	3	0

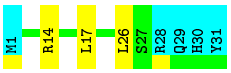
4 Residue-property plots (i)

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.

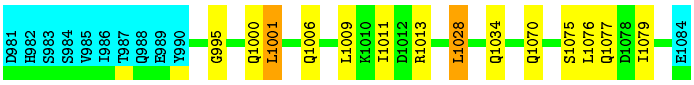
- Molecule 1: Amino acid adenylation domain-containing protein

Chain B:



- Molecule 2: Peptide synthetase PaxA

Chain A:



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

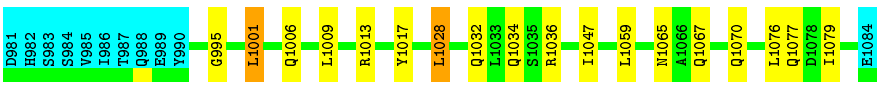
The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: Amino acid adenylation domain-containing protein

Chain B:

- Molecule 2: Peptide synthetase PaxA

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	**Classification**	**Version**
CYANA	structure calculation	3.98
OPALp	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	1718
Number of shifts mapped to atoms	1718
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

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	B	0.45±0.02	0±0/203 (0.0± 0.0%)	0.98±0.04	0±0/273 (0.1± 0.2%)
2	A	0.57±0.01	0±0/764 (0.0± 0.0%)	1.01±0.03	1±1/1035 (0.1± 0.1%)
All	All	0.55	0/19340 (0.0%)	1.01	24/26160 (0.1%)

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	B	0.0±0.0	0.6±0.6
2	A	0.0±0.0	0.7±0.8
All	All	0	24

There are no bond-length outliers.

5 of 14 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	1083	MET	CG-SD-CE	6.97	111.36	100.20	17	2
2	A	1013	ARG	CD-NE-CZ	6.88	133.23	123.60	12	2
2	A	1013	ARG	NE-CZ-NH1	6.79	123.69	120.30	19	1
2	A	1013	ARG	NE-CZ-NH2	-6.67	116.96	120.30	19	3
2	A	1018	ASP	CB-CG-OD1	-6.53	112.43	118.30	7	2

There are no chirality outliers.

5 of 8 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	B	14	ARG	Sidechain	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
2	A	1013	ARG	Sidechain	3
2	A	1016	ARG	Sidechain	3
2	A	1017	TYR	Sidechain	2
2	A	1080	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	B	202	222	222	2±1
2	A	751	737	737	5±2
All	All	19060	19180	19180	103

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 3.

5 of 32 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:26:LEU:HD13	2:A:1079:ILE:HD12	0.61	1.73	3	9
1:B:26:LEU:HD12	2:A:1075:SER:HB3	0.59	1.75	5	6
2:A:1045:ILE:HD11	2:A:1066:ALA:HB2	0.58	1.75	7	1
2:A:1014:ILE:CG2	2:A:1056:ILE:HD11	0.58	2.29	10	3
2:A:1008:ILE:CG2	2:A:1029:LEU:HD22	0.57	2.30	2	4

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	B	26/31 (84%)	25±1 (97±3%)	1±1 (2±3%)	0±0 (1±1%)	29 74

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	A	93/104 (89%)	88±1 (95±1%)	4±1 (4±1%)	1±0 (1±0%)	18 66
All	All	2380/2700 (88%)	2263 (95%)	94 (4%)	23 (1%)	20 68

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	A	995	GLY	20
1	B	27	SER	2
1	B	2	ASN	1

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	B	22/27 (81%)	18±1 (82±6%)	4±1 (18±6%)	4 38
2	A	83/94 (88%)	68±3 (82±3%)	15±3 (18±3%)	4 37
All	All	2100/2420 (87%)	1719 (82%)	381 (18%)	4 37

5 of 57 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	B	17	LEU	20
2	A	1001	LEU	20
2	A	1009	LEU	20
2	A	1028	LEU	20
2	A	1076	LEU	20

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 90% for the well-defined parts and 89% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *JW140_JW146new2.star*

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	1718
Number of shifts mapped to atoms	1718
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	135	-0.39 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	130	0.04 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	127	-0.19 ± 0.05	None needed (< 0.5 ppm)
^{15}N	127	0.14 ± 0.33	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 90%, i.e. 1362 atoms were assigned a chemical shift out of a possible 1508. 21 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	581/589 (99%)	233/235 (99%)	234/238 (98%)	114/116 (98%)
Sidechain	736/832 (88%)	451/484 (93%)	265/311 (85%)	20/37 (54%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	45/87 (52%)	23/47 (49%)	22/37 (59%)	0/3 (0%)
Overall	1362/1508 (90%)	707/766 (92%)	521/586 (89%)	134/156 (86%)

7.1.4 Statistically unusual chemical shifts (i)

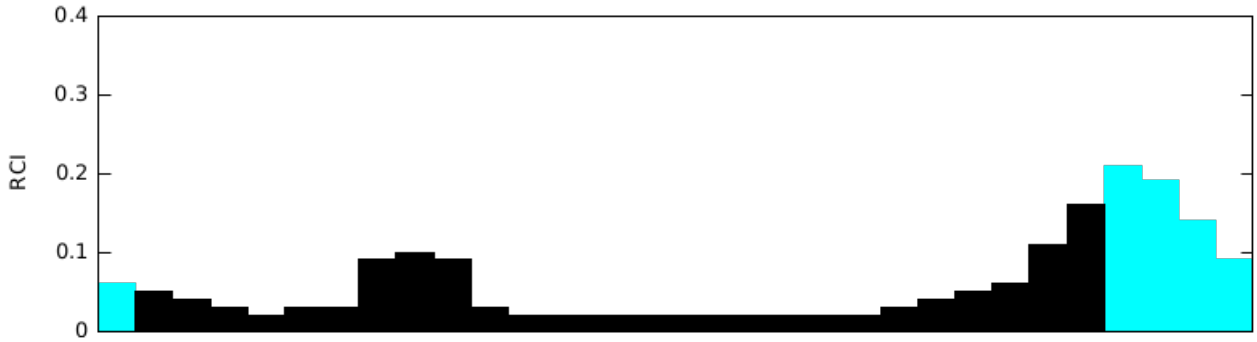
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	1013	ARG	NE	113.45	92.63 – 76.73	18.1
1	A	1036	ARG	NE	112.84	92.63 – 76.73	17.7
2	B	14	ARG	NE	112.77	92.63 – 76.73	17.7
1	A	1016	ARG	NE	112.10	92.63 – 76.73	17.2
2	B	16	ARG	NE	111.65	92.63 – 76.73	17.0
2	B	28	ARG	NE	110.78	92.63 – 76.73	16.4

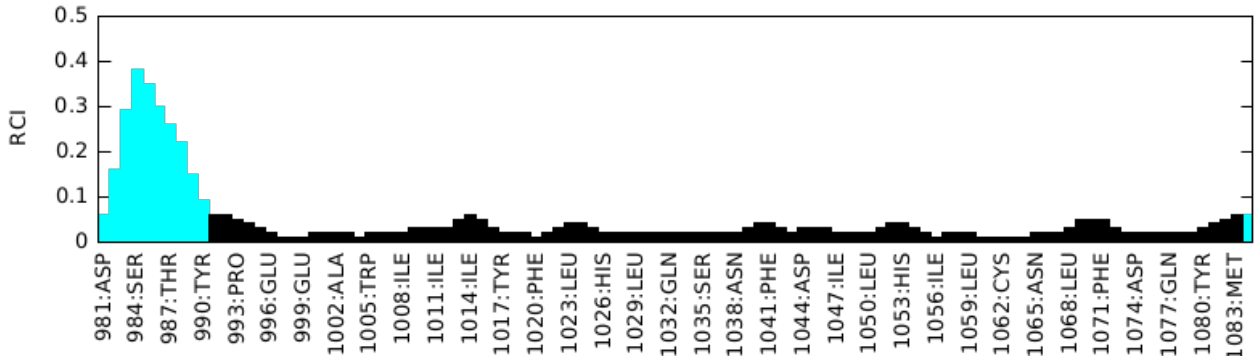
7.1.5 Random Coil Index (RCI) plots (i)

The images below report *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 B:



Random coil index (RCI) for chain A:


RCI
0.5
0.4
0.3
0.2
0.1
0
981:ASP
984:SER
987:THR
990:TYR
993:PRO
996:GLU
999:GLU
1002:ALA
1005:TRP
1008:ILE
1011:ILE
1014:ILE
1017:TYR
1020:PHE
1023:LEU
1026:HIS
1029:LEU
1032:GLN
1035:SER
1038:ASN
1041:PHE
1044:ASP
1047:ILE
1050:LEU
1053:HIS
1056:ILE
1059:LEU
1062:CYS
1065:ASN
1068:LEU
1071:PHE
1074:ASP
1077:GLN
1080:TYR
1083:MET