Full wwPDB NMR Structure Validation Report (i)

May 29, 2020 – 07:50 am BST

PDB ID	:	5NPA
Title	:	Solution structure of Drosophila melanogaster Loquacious dsRBD2
Authors	:	Tants, J.-N.; Fesser, S.; Kern, T.; Stehle, R.; Geerlof, A.; Wunderlich, C.; Boettcher, R.; Kunzelmann, S.; Lange, O.; Kreutz, C.; Foerstemann, K.; Sattler, M.
Deposited on	:	2017-04-16

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 following versions of software and data (see references (1)) were used in the production of this report:

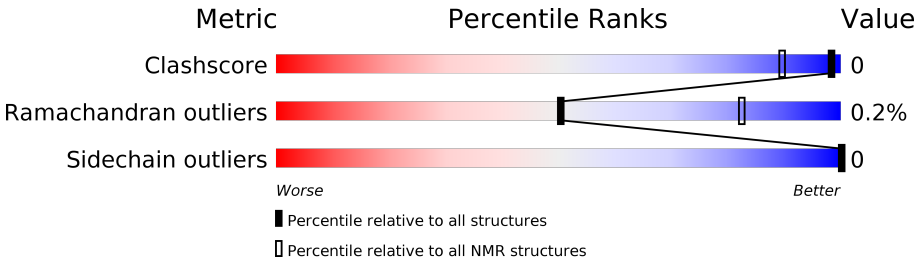
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 52%.

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	72	90% 7%

2 Ensemble composition and analysis

This entry contains 10 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: *none*.

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:9-A:74 (66)	0.77	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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	3, 4, 5, 7, 9, 10
2	1, 2, 6, 8

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1127 atoms, of which 567 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Loquacious.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	67	1127	355	567	104	94	7	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	ASP	-	expression tag	UNP Q4TZM6
A	4	LYS	-	expression tag	UNP Q4TZM6
A	5	THR	-	expression tag	UNP Q4TZM6
A	6	VAL	-	expression tag	UNP Q4TZM6
A	7	GLY	-	expression tag	UNP Q4TZM6

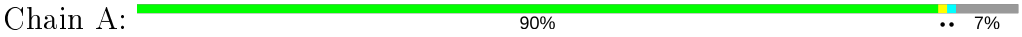
4 Residue-property plots

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: Loquacious

Chain A:



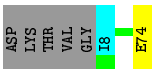
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Loquacious

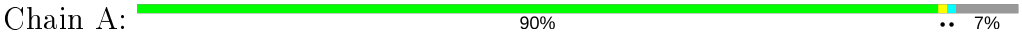
Chain A:

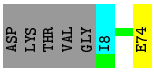


4.2.2 Score per residue for model 2

- Molecule 1: Loquacious

Chain A:

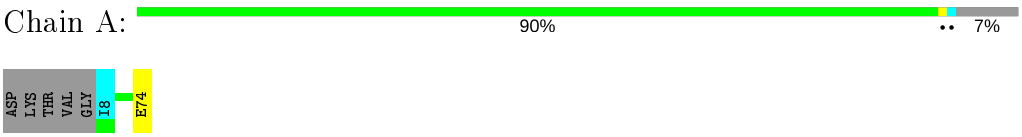




4.2.3 Score per residue for model 3

• Molecule 1: Loquacious

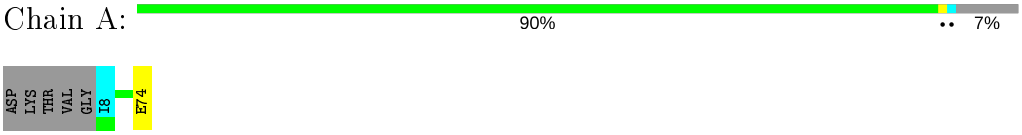
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Loquacious

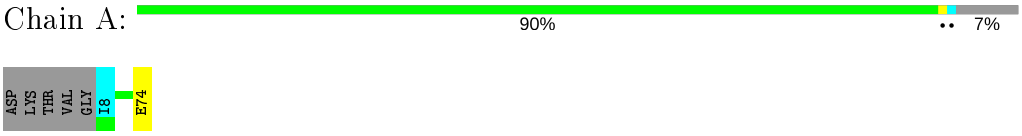
Chain A:



4.2.5 Score per residue for model 5 (medoid)

• Molecule 1: Loquacious

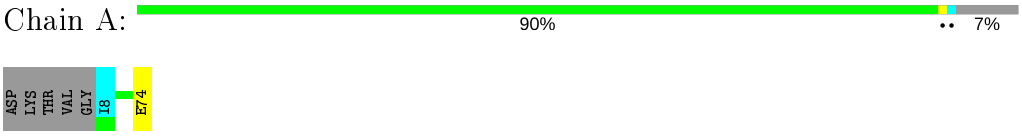
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Loquacious

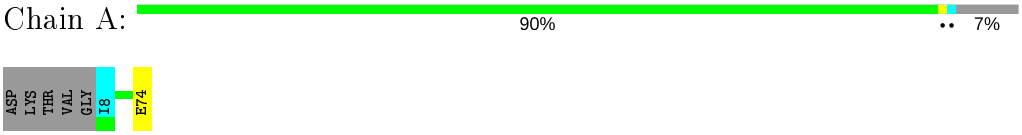
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Loquacious

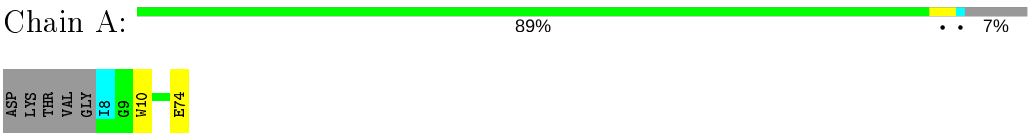
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Loquacious

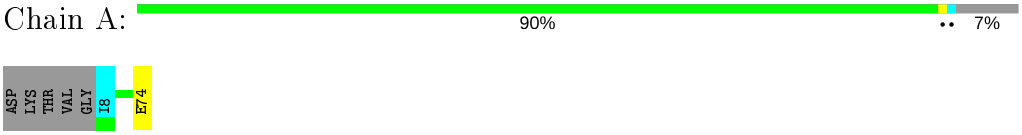
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Loquacious

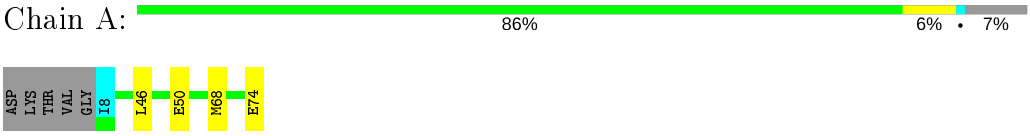
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: Loquacious

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *Rasrec algorithm.*

Of the 5000 calculated structures, 10 were deposited, based on the following criterion: *Rosetta score.*

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

Software name	Classification	Version
CS-ROSETTA	structure calculation	autoNOE Rosetta

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	560
Number of shifts mapped to atoms	533
Number of unparsed shifts	0
Number of shifts with mapping errors	27
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	52%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	552	556	558	0±0
All	All	5520	5560	5580	2

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:10:TRP:C	1:A:10:TRP:CD1	0.46	2.88	8	1
1:A:50:GLU:HB3	1:A:68:MET:SD	0.41	2.56	10	1

5.2 Torsion angles

5.2.1 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	65/72 (90%)	62±1 (96±1%)	3±1 (4±1%)	0±0 (0±0%)	50	82
All	All	650/720 (90%)	622 (96%)	27 (4%)	1 (0%)	50	82

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	46	LEU	1

5.2.2 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	58/63 (92%)	58±0 (100±0%)	0±0 (0±0%)	100	100
All	All	580/630 (92%)	580 (100%)	0 (0%)	100	100

There are no protein residues with a non-rotameric sidechain to report.

5.2.3 RNA

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

There are no ligands in this entry.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 52% for the well-defined parts and 52% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *dsRBD2_shifts.tab*

6.1.1 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	560
Number of shifts mapped to atoms	533
Number of unparsed shifts	0
Number of shifts with mapping errors	27
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

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

- Residue not found in structure. All 27 occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	3	THR	HG22	0.767	-1.0	1
A	3	THR	CB	69.711	-1.0	1
A	3	THR	HA	4.18	-1.0	1
A	4	VAL	N	121.721	-1.0	1
A	4	VAL	CB	32.527	-1.0	1
A	3	THR	HB	4.072	-1.0	1
A	1	ASP	N	121.126	-1.0	1
A	1	ASP	HA	4.453	-1.0	1
A	3	THR	HG23	0.767	-1.0	1
A	4	VAL	HA	4.068	-1.0	1
A	4	VAL	CG2	20.667	-1.0	1
A	4	VAL	H	7.883	-1.0	1
A	4	VAL	HG22	0.758	-1.0	1
A	1	ASP	H	8.048	-1.0	1

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	4	VAL	HG23	0.758	-1.0	1
A	2	LYS	N	121.358	-1.0	1
A	5	GLY	CA	44.564	-1.0	1
A	3	THR	HG21	0.767	-1.0	1
A	5	GLY	HA3	3.806	-1.0	1
A	2	LYS	H	8.112	-1.0	1
A	4	VAL	HB	1.923	-1.0	1
A	5	GLY	HA2	4.043	-1.0	1
A	1	ASP	HB3	2.583	-1.0	1
A	4	VAL	HG21	0.758	-1.0	1
A	3	THR	CG2	21.37	-1.0	1
A	3	THR	CA	62.264	-1.0	1
A	4	VAL	CA	62.11	-1.0	1

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	44	-0.07 ± 0.20	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	46	0.13 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	59	0.37 ± 0.51	None needed (< 0.5 ppm)

6.1.3 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 52%, i.e. 463 atoms were assigned a chemical shift out of a possible 894. 5 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	207/322 (64%)	112/128 (88%)	40/132 (30%)	55/62 (89%)
Sidechain	256/499 (51%)	169/300 (56%)	87/169 (51%)	0/30 (0%)
Aromatic	0/73 (0%)	0/39 (0%)	0/31 (0%)	0/3 (0%)
Overall	463/894 (52%)	281/467 (60%)	127/332 (38%)	55/95 (58%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 52%, i.e. 475 atoms were assigned a chemical shift out of a possible 908. 5 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	210/327 (64%)	113/130 (87%)	41/134 (31%)	56/63 (89%)
Sidechain	265/508 (52%)	174/305 (57%)	91/173 (53%)	0/30 (0%)
Aromatic	0/73 (0%)	0/39 (0%)	0/31 (0%)	0/3 (0%)
Overall	475/908 (52%)	287/474 (61%)	132/338 (39%)	56/96 (58%)

6.1.4 Statistically unusual chemical shifts

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	63	LEU	HD12	4.53	2.16 – -0.64	13.5
1	A	63	LEU	HD13	4.53	2.16 – -0.64	13.5
1	A	63	LEU	HD11	4.53	2.16 – -0.64	13.5
1	A	18	ARG	HB3	0.26	3.17 – 0.37	-5.4
1	A	18	ARG	HB2	0.44	3.15 – 0.45	-5.1

6.1.5 Random Coil Index (RCI) plots

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: