wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:21 PM BST

PDB ID	:	2EQS
Title	:	Solution structure of the S1 RNA binding domain of human ATP-dependent RNA helicase DHX8
Authors	:	Suzuki, S.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2007-03-30

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 (i) symbol.

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

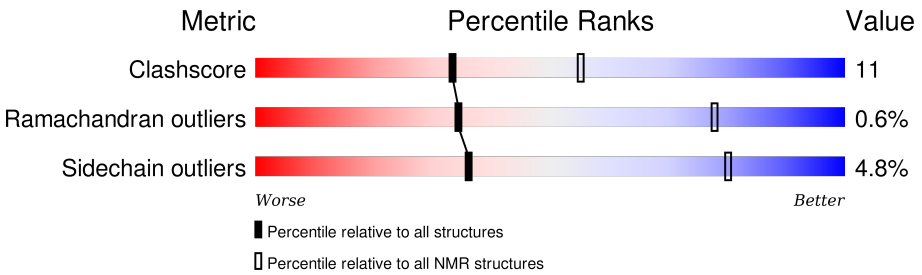
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 was not calculated.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	103	62% 14% 24%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 17 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:260-A:300, A:309-A:345 (78)	0.26	17

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1601 atoms, of which 806 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ATP-dependent RNA helicase DHX8.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	103	1601	488	806	149	155	3	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	253	GLY	-	EXPRESSION TAG	UNP Q14562
A	254	SER	-	EXPRESSION TAG	UNP Q14562
A	255	SER	-	EXPRESSION TAG	UNP Q14562
A	256	GLY	-	EXPRESSION TAG	UNP Q14562
A	257	SER	-	EXPRESSION TAG	UNP Q14562
A	258	SER	-	EXPRESSION TAG	UNP Q14562
A	259	GLY	-	EXPRESSION TAG	UNP Q14562

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: ATP-dependent RNA helicase DHX8

Chain A: 62% 14% 24%

G253 S254 S255 G256 S257 S258 G259 I275 M276 Q277 F278 F281 R290 W291 L294 I297 S298 E299 L300 R301 R302 E303 G304 R305 V306 A307 N308 V309 V312 V313 T327 L346 N347 P348 N349 R350 R351 R352 N353 L354 V355

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

The representative model is number 17. Colouring as in section 4.1 above.

- Molecule 1: ATP-dependent RNA helicase DHX8

Chain A: 54% 20% • 24%

G253 S254 S255 G256 S257 S258 G259 I264 G265 D266 R271 I275 M276 Q277 F278 F281 L287 R290 L294 L300 R301 R302 E303 G304 R305 V306 A307 N308 V309 A310 D311 V312 L324 T327 S334 D337 V338 D339 G343 E344 D345 L346 N347 P348 N349 R350 R351

R352 N353 L354 V355

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dyanamics, simulated annealing.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy, target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	606	614	614	13±3
All	All	12120	12280	12280	266

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 11.

5 of 132 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:300:LEU:HD13	1:A:312:VAL:HG11	0.93	1.37	20	7
1:A:297:ILE:HD13	1:A:297:ILE:O	0.71	1.85	3	1
1:A:281:PHE:CE1	1:A:294:LEU:HD13	0.71	2.20	1	5
1:A:309:VAL:HG12	1:A:313:VAL:HG22	0.68	1.64	19	12
1:A:284:LEU:HD12	1:A:291:TRP:HB2	0.65	1.66	15	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	A	78/103 (76%)	69±3 (88±3%)	9±2 (11±3%)	1±1 (1±1%)	34 78
All	All	1560/2060 (76%)	1372 (88%)	178 (11%)	10 (1%)	34 78

5 of 6 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	287	LEU	3
1	A	290	ARG	2
1	A	296	HIS	2
1	A	338	VAL	1
1	A	337	ASP	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	A	69/89 (78%)	66±1 (95±2%)	3±1 (5±2%)	36 80
All	All	1380/1780 (78%)	1314 (95%)	66 (5%)	36 80

5 of 25 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	266	ASP	8
1	A	290	ARG	7
1	A	278	PHE	6
1	A	271	LYS	6
1	A	299	GLU	6

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 carbohydrates 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)

No chemical shift data were provided