wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:54 PM BST

PDB ID	:	2HQN
Title	:	Structure of a Atypical Orphan Response Regulator Protein Revealed a New Phosphorylation-Independent Regulatory Mechanism
Authors	:	Hong, E.; Lee, W.
Deposited on	:	2006-07-19

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:

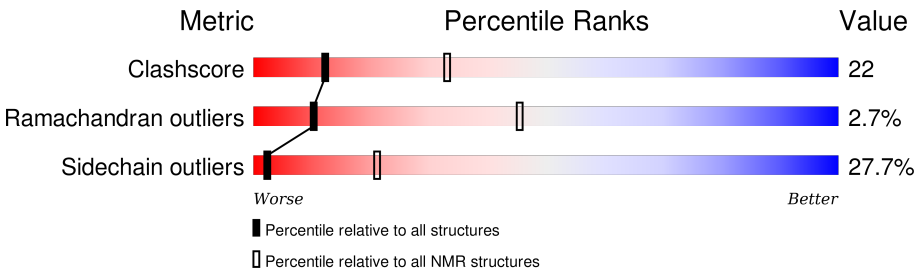
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	109	42% 39% 8% 10%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 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:120-A:217 (98)	0.17	3

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 and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1760 atoms, of which 886 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative TRANSCRIPTIONAL REGULATOR.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	109	1760	550	886	153	167	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	115	GLY	-	CLONING ARTIFACT	UNP Q9ZM42
A	116	SER	-	CLONING ARTIFACT	UNP Q9ZM42
A	117	GLY	-	CLONING ARTIFACT	UNP Q9ZM42
A	118	SER	-	CLONING ARTIFACT	UNP Q9ZM42

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: Putative TRANSCRIPTIONAL REGULATOR

Chain A: 42% 39% 8% 10%

G115 S116 G117 S118 N119 V120 I121 E122 I123 L126 T127 I128 S129 E132 E133 K134 Y137 R140 K145 G146 K147 P148 F149 E150 V151 L152 T153 H154 L155 A156 R157 H158 R159 D160 Q161 I162 V163 S164 L168 L169 I172 W173 M178 V179 T180 P181 N182 V183 I184 E185 V186

I191 M195 D196 K197 P198 L199 G200 I201 S202 T203 V204 E205 R209 F214 C215 Y216 P217 K218 P219 A220 C221 E222 E223

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: Putative TRANSCRIPTIONAL REGULATOR

Chain A: 45% 38% 7% 10%

G115 S116 G117 S118 N119 I123 G124 D125 L126 T127 I128 S129 E132 E133 K134 Y137 V142 K147 P148 F149 E150 V151 L152 T153 H154 L155 A156 R157 H158 R159 D160 Q161 I162 V163 S164 L168 I172 W173 P176 E177 M178 V179 T180 P181 N182 V183 I184 E185 V186 M195 D196

K197 P198 L199 G200 I201 S202 T203 V204 E205 R209 Y212 R213 F214 C215 Y216 P217 K218 P219 A220 C221 E222 E223

5 Refinement protocol and experimental data overview (i)

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

Of the 100 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
Cyana	refinement	2.1

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	801	822	822	36±4
All	All	16020	16440	16440	725

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 22.

5 of 129 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:151:VAL:HG12	1:A:172:ILE:HG21	1.06	1.26	9	20
1:A:142:VAL:HG21	1:A:199:LEU:HD21	0.95	1.35	15	7
1:A:163:VAL:HG11	1:A:168:LEU:HD12	0.88	1.46	8	4
1:A:152:LEU:HD22	1:A:191:ILE:HG21	0.86	1.45	16	2
1:A:155:LEU:HD11	1:A:168:LEU:HD13	0.84	1.48	20	1

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	98/109 (90%)	83±1 (85±1%)	12±2 (12±2%)	3±1 (3±1%)	10 45
All	All	1960/2180 (90%)	1667 (85%)	241 (12%)	52 (3%)	10 45

5 of 7 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	161	GLN	20
1	A	217	PRO	18
1	A	160	ASP	6
1	A	124	GLY	5
1	A	176	PRO	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	90/98 (92%)	65±3 (72±3%)	25±3 (28±3%)	2 21
All	All	1800/1960 (92%)	1301 (72%)	499 (28%)	2 21

5 of 56 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	162	ILE	20
1	A	127	THR	20
1	A	129	SER	20
1	A	128	ILE	20
1	A	149	PHE	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 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