wwPDB NMR Structure Validation Summary Report

May 28, 2020 – 10:04 pm BST

PDB ID	:	2HQO
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
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:

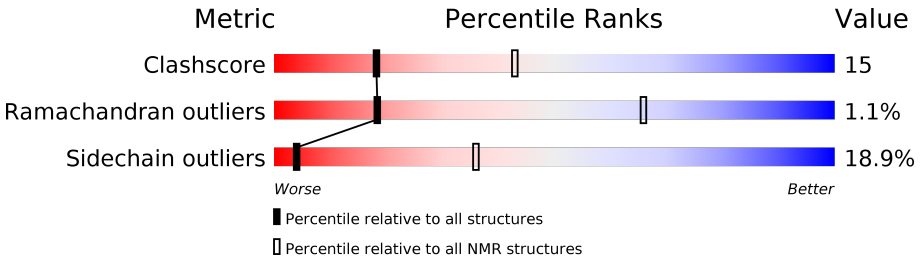
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)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

1 Overall quality at a glance

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	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	123	56% 32% · 6% ·
1	B	123	53% 37% · 6% ·

2 Ensemble composition and analysis

This entry contains 20 models. Model 2 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:2-A:113, B:302-B:413 (224)	0.21	2

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3762 atoms, of which 1882 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative TRANSCRIPTIONAL REGULATOR.

Mol	Chain	Residues	Atoms						Trace
1	A	119	Total	C	H	N	O	S	0
			1881	592	941	160	184	4	
1	B	119	Total	C	H	N	O	S	0
			1881	592	941	160	184	4	

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	120	LEU	-	CLONING ARTIFACT	UNP Q9ZM42
A	121	VAL	-	CLONING ARTIFACT	UNP Q9ZM42
A	122	PRO	-	CLONING ARTIFACT	UNP Q9ZM42
A	123	ARG	-	CLONING ARTIFACT	UNP Q9ZM42
B	420	LEU	-	CLONING ARTIFACT	UNP Q9ZM42
B	421	VAL	-	CLONING ARTIFACT	UNP Q9ZM42
B	422	PRO	-	CLONING ARTIFACT	UNP Q9ZM42
B	423	ARG	-	CLONING ARTIFACT	UNP Q9ZM42

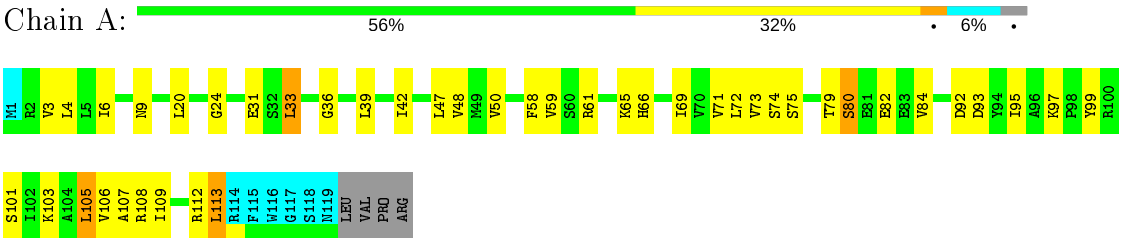
4 Residue-property plots (i)

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.

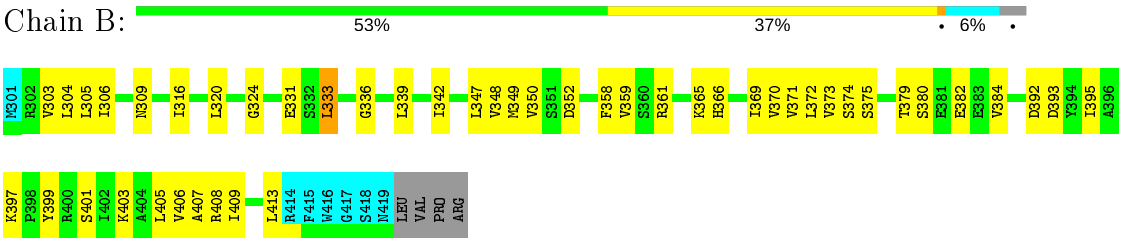
- Molecule 1: Putative TRANSCRIPTIONAL REGULATOR

Chain A:



- Molecule 1: Putative TRANSCRIPTIONAL REGULATOR

Chain B:

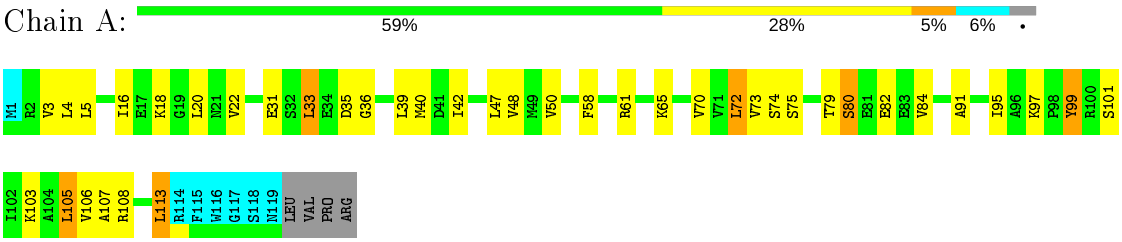


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

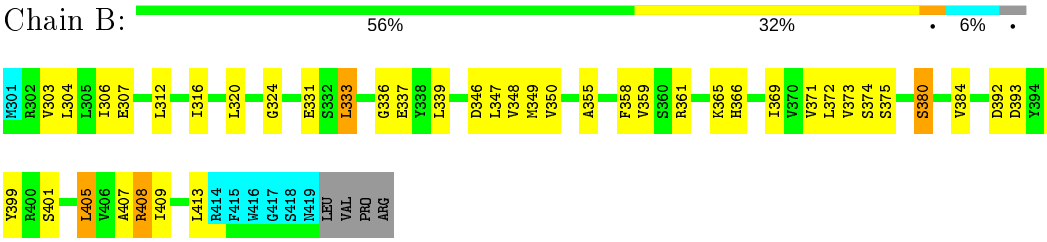
- Molecule 1: Putative TRANSCRIPTIONAL REGULATOR

Chain A:



- Molecule 1: Putative TRANSCRIPTIONAL REGULATOR

Chain B:


56%
32%
6%
M301 R302 V303 L304 L305 I306 E307 L312 I316 L320 G324 E331 S332 L333 G336 E337 Y338 L339 D346 L347 V348 M349 V350 A355 F358 V359 S360 R361 K365 H366 I369 V370 V371 L372 V373 S374 S375 S380 V384 D392 D393 Y394 I395 A396 K397 P398
Y399 R400 S401 L405 V406 A407 R408 I409 L413 R414 F415 W416 G417 S418 M419 LEU VAL PRO ARG

5 Refinement protocol and experimental data overview

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

6.1 Standard geometry

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

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	878	886	886	27±5
1	B	878	886	886	29±4
All	All	35120	35440	35440	1038

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 15.

5 of 252 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:347:LEU:HD23	1:B:413:LEU:HD11	0.90	1.39	15	2
1:B:303:VAL:HG13	1:B:347:LEU:HD12	0.89	1.45	8	14
1:A:3:VAL:HG13	1:A:47:LEU:HD12	0.87	1.43	5	11
1:B:347:LEU:HD11	1:B:372:LEU:HD12	0.83	1.48	19	12
1:A:72:LEU:HD22	1:A:95:ILE:HD11	0.79	1.54	13	12

6.3 Torsion angles

6.3.1 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	112/123 (91%)	102±2 (91±1%)	9±2 (8±1%)	1±1 (1±1%)	18 66
1	B	112/123 (91%)	102±1 (91±1%)	9±1 (8±1%)	1±1 (1±1%)	20 68
All	All	4480/4920 (91%)	4069 (91%)	363 (8%)	48 (1%)	18 66

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	B	324	GLY	14
1	A	24	GLY	12
1	A	9	ASN	9
1	B	309	ASN	6
1	B	390	GLY	3

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	98/108 (91%)	80±2 (81±2%)	18±2 (19±2%)	4 36
1	B	98/108 (91%)	79±2 (81±2%)	19±2 (19±2%)	4 36
All	All	3920/4320 (91%)	3178 (81%)	742 (19%)	4 36

5 of 106 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	401	SER	20
1	B	361	ARG	20
1	A	97	LYS	20
1	A	61	ARG	20
1	B	333	LEU	20

6.3.3 RNA

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates

There are no carbohydrates in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

No chemical shift data were provided