wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 03:03 PM BST

PDB ID	:	1FNX
Title	:	SOLUTION STRUCTURE OF THE HUC RBD1-RBD2 COMPLEXED WITH THE AU-RICH ELEMENT
Authors	:	Inoue, M.; Hirao, M.; Kasashima, K.; Kim, I.-S.; Kawai, G.; Kigawa, T.; Sakamoto, H.; Muto, Y.; Yokoyama, S.
Deposited on	:	2000-08-24

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:

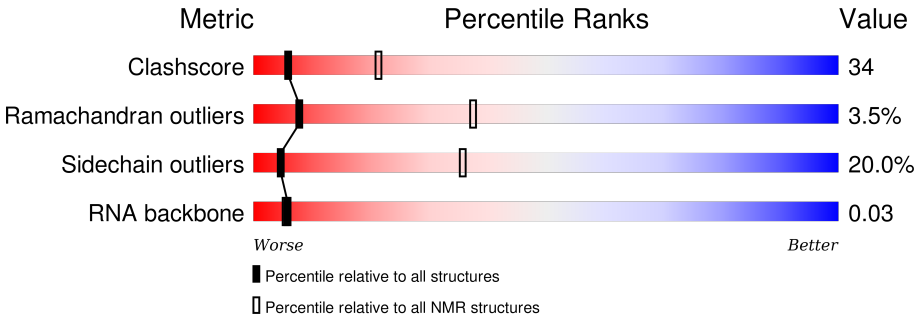
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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
RNA backbone	3027	600

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	R	10	10% 30% 60%
2	H	174	37% 46% • 16%

2 Ensemble composition and analysis (i)

This entry contains 21 models. Model 21 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	H:38-H:71, H:79-H:119, H:124-H:188, H:198-H:204 (147)	0.96	21

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, 4, 6, 8, 9, 12, 15, 17, 19, 20, 21
2	3, 5
3	11, 18
Single-model clusters	2; 7; 10; 13; 14; 16

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3022 atoms, of which 1471 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called AU-RICH RNA ELEMENT.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	R	10	309	92	104	26	77	10	0

- Molecule 2 is a protein called HU ANTIGEN C.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	H	174	2713	844	1367	233	264	5	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	35	MET	ASP	CLONING ARTIFACT	UNP Q60900

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.

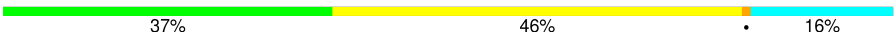
- Molecule 1: AU-RICH RNA ELEMENT

Chain R:

U0 A1 U2 U3 U4 A5 U6 U7 U8 U9

- Molecule 2: HU ANTIGEN C

Chain H:



M35 D36 S37 K38 T39 N40 L41 I42 V43 N44 Y45 L46 P47 F59 G60 D64 I65 E66 K69 L70 V71 R72 D73 K74 I75 T76 G77 Q78 S79 L80 G81 Y82 G83 F84 V85 N86 Y87 S88 D89 D94 K95 A96 I97 N98 T99 L100 N101 G102 L103 K104 L105 K108 T109 I110 K111

V112 A115 R116 P117 S118 S119 A120 S121 I122 R123 D124 A125 N126 L127 Y128 V129 K134 T135 M136 S137 Q138 K139 E140 M141 E142 Q143 L144 F145 S146 Q147 Y148 G149 R150 I151 R155 I156 L157 L158 D159 Q160 S165 R166 G167 V168 G169 F170 I171 R172 F173 D174 K175 R176 A179 I183

K184 G185 L186 N187 G188 Q189 K190 P191 L192 G193 A194 A195 E196 P197 I198 T199 V200 K201 M205 P206 S207 Q208

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

The representative model is number 21. Colouring as in section 4.1 above.

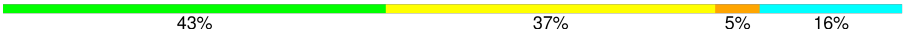
- Molecule 1: AU-RICH RNA ELEMENT

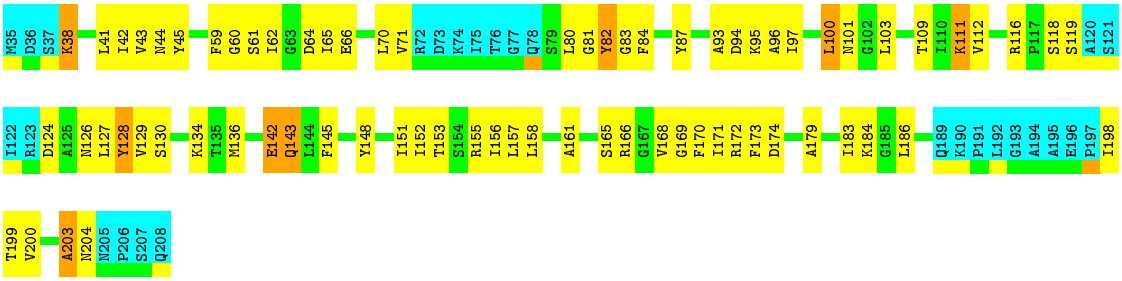
Chain R:

U0 A1 U2 U3 U4 A5 U6 U7 U8 U9

- Molecule 2: HU ANTIGEN C

Chain H:




M35 D36 S37 K38
L41 I42 V43 N44 Y45
F59 G60 S61 I62
G63 D64 I65 E66
L70 V71
R72 D73 K74 I75 T76 G77 Q78
S79 L80 G81 Y82 G83 F84
Y87
A93 D94 K95 A96 I97
L100 N101 G102 L103
T109 I110 K111 V112
R116 P117 S118 S119 A120 S121
I122 R123 D124 A125 N126 L127 Y128 V129 S130
K134 T135 M136
E142 Q143 L144 F145
Y148
I151 I152 T153 S154 R155 I156 L157 L158
A161
S165 R166 G167 V168 G169 F170 I171 R172 F173 D174
A179
I183 K184 G185 L186
Q189 K190 P191 L192 G193 A194 A195 E196 P197 I198
T199 V200
A203 N204 N205 P206 S207 Q208

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 21 were deposited, based on the following criterion: *structures with the lowest energy. CONFORMER 21 IS THE MINIMIZED AVERAGE STRUCTURE OF CONFORMERS 1-20.*

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

Software name	Classification	Version
CNS	structure solution	0.9
CNS	refinement	0.9

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)

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	R	0.30±0.02	0±0/226 (0.0±0.0%)	0.72±0.02	0±0/346 (0.0±0.0%)
2	H	0.30±0.00	0±0/1161 (0.0±0.0%)	0.44±0.01	1±1/1564 (0.1±0.1%)
All	All	0.30	0/29127 (0.0%)	0.50	20/40110 (0.0%)

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	R	0.0±0.0	0.0±0.2
All	All	0	1

There are no bond-length outliers.

All 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	H	128	TYR	CB-CG-CD1	6.21	124.72	121.00	12	10
2	H	128	TYR	CB-CG-CD2	-6.05	117.37	121.00	12	10

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	R	2	U	Sidechain	1

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	R	205	104	103	32±7
2	H	1144	1163	1163	86±11
All	All	28329	26607	26586	1879

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 34.

5 of 651 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:R:8:U:O4'	2:H:71:VAL:HG22	1.05	1.50	19	13
2:H:157:LEU:HD21	2:H:168:VAL:HB	0.97	1.37	13	10
1:R:2:U:C6	2:H:170:PHE:CZ	0.93	2.57	2	21
1:R:5:A:C2	2:H:80:LEU:HD21	0.89	2.02	5	1
1:R:5:A:N3	2:H:80:LEU:HD21	0.89	1.81	5	2

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	
2	H	147/174 (84%)	119±3 (81±2%)	23±3 (16±2%)	5±2 (4±1%)	8	37
All	All	3087/3654 (84%)	2493 (81%)	485 (16%)	109 (4%)	8	37

5 of 28 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	H	102	GLY	16
2	H	198	ILE	11

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	H	150	ARG	10
2	H	161	ALA	9
2	H	81	GLY	9

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
2	H	127/149 (85%)	102±4 (80±3%)	25±4 (20±3%)	5 36
All	All	2667/3129 (85%)	2134 (80%)	533 (20%)	5 36

5 of 70 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)
2	H	95	LYS	20
2	H	143	GLN	19
2	H	100	LEU	17
2	H	82	TYR	16
2	H	116	ARG	14

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	R	9/10 (90%)	8±0 (94±6%)	2±1 (19±10%)	0.03±0.03
All	All	193/210 (92%)	178 (92%)	36 (19%)	0.03

The overall RNA backbone suiteness is 0.03.

5 of 9 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	R	6	U	21
1	R	8	U	21
1	R	9	U	21
1	R	4	U	21

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	R	2	U	20

5 of 7 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	R	5	A	17
1	R	6	U	5
1	R	8	U	4
1	R	0	U	4
1	R	2	U	2

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