wwPDB NMR Structure Validation Summary Report (i)

Aug 8, 2016 – 05:20 PM EDT

PDB ID	:	5KIZ
Title	:	Solution Structure of a repacked version of HIF-2 alpha PAS-B
Authors	:	Correa, F.; Key, J.; Kuhlman, B.; Gardner, K.H.
Deposited on	:	2016-06-17

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:

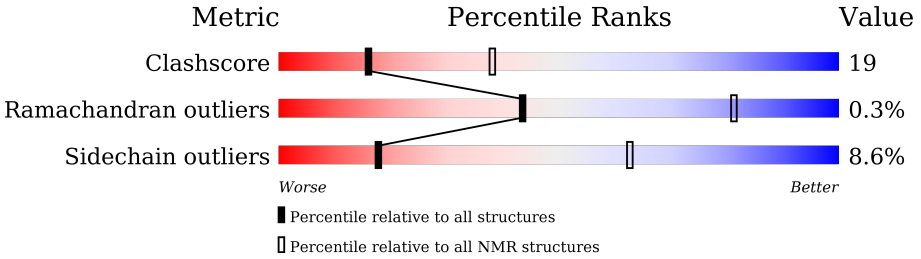
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027939
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 91%.

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	114	56% 27% 5% 11%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:242-A:294, A:302-A:349 (101)	0.24	8

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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1831 atoms, of which 903 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Endothelial PAS domain-containing protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	114	1831	587	903	156	176	9	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	236	GLY	-	expression tag	UNP Q99814
A	237	GLU	-	expression tag	UNP Q99814
A	238	PHE	-	expression tag	UNP Q99814
A	248	PHE	HIS	engineered mutation	UNP Q99814
A	278	SER	TYR	engineered mutation	UNP Q99814
A	281	TRP	TYR	engineered mutation	UNP Q99814
A	323	MET	GLY	engineered mutation	UNP Q99814
A	339	ALA	CYS	engineered mutation	UNP Q99814

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: Endothelial PAS domain-containing protein 1

Chain A: 56% 27% 5% 11%

G236 E237 F238 L239 D240 S241 K242 T243 F244 F254 C257 D258 D259 R260 I261 L264 I265 G266 Y267 H268 P269 E270 E271 L272 L273 G274 R275 E287 T290 K291 Q294 N295 L296 C297 T298 K299 G300 Q301 V302 R308 M309 L310 A311 K312 G315 Y316 V317 W318 L319 E320 M338

Y342 V343 I347 E348 K349

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

The representative model is number 8. Colouring as in section 4.1 above.

- Molecule 1: Endothelial PAS domain-containing protein 1

Chain A: 49% 34% 5% 11%

G236 E237 F238 L239 D240 S241 K242 T243 F244 L245 F254 C257 D258 D259 R260 I261 L264 I265 G266 Y267 H268 P269 E270 E271 L272 L273 G274 R275 D285 S286 E287 N288 M289 T290 K291 S292 H293 Q294 N295 L296 C297 T298 K299 G300 Q301 M309 L310 A311 K312 H313 Y316 V317 W318 L319

E320 T321 Q322 M323 T324 R330 M338 A339 Y342 V343 I347 E348 K349

5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *all calculated structures submitted*.

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

Software name	Classification	Version
ARIA	refinement	2.3

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	5kiz_cs.cif
Number of chemical shift lists	1
Total number of shifts	1448
Number of shifts mapped to atoms	1448
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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	A	0.48±0.03	0±0/851 (0.0±0.1%)	0.52±0.01	0±0/1147 (0.0±0.0%)
All	All	0.49	2/17020 (0.0%)	0.52	0/22940 (0.0%)

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	267	TYR	CE2-CZ	5.21	1.45	1.38	8	1
1	A	267	TYR	CE1-CZ	-5.12	1.31	1.38	8	1

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	831	810	806	31±3
All	All	16620	16200	16120	616

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 19.

5 of 110 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:272:LEU:O	1:A:275:ARG:HG2	0.70	1.86	1	20
1:A:267:TYR:CE1	1:A:312:LYS:HA	0.65	2.26	12	19
1:A:287:GLU:O	1:A:291:LYS:HD3	0.65	1.92	1	13
1:A:290:THR:O	1:A:294:GLN:HG3	0.64	1.92	11	10
1:A:260:ARG:HG3	1:A:264:LEU:HD11	0.63	1.68	5	3

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	100/114 (88%)	94±1 (94±1%)	6±1 (6±1%)	0±0 (0±0%)	43	81
All	All	2000/2280 (88%)	1876 (94%)	117 (6%)	7 (0%)	43	81

All 3 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	334	PRO	3
1	A	242	LYS	3
1	A	306	GLN	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	92/103 (89%)	84±1 (91±2%)	8±1 (9±2%)	18	63
All	All	1840/2060 (89%)	1682 (91%)	158 (9%)	18	63

5 of 22 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	261	ILE	20
1	A	310	LEU	20
1	A	319	LEU	20
1	A	338	MET	15
1	A	265	ILE	15

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)

The completeness of assignment taking into account all chemical shift lists is 91% for the well-defined parts and 90% for the entire structure.

7.1 Chemical shift list 1

File name: 5kiz_cs.cif

Chemical shift list name: *assigned_chem_shift_list_0*

7.1.1 Bookkeeping (i)

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	1448
Number of shifts mapped to atoms	1448
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	10

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	112	-0.46 ± 0.25	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	107	-0.10 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	99	-0.10 ± 0.16	None needed (< 0.5 ppm)
^{15}N	108	0.54 ± 0.35	None needed (imprecise)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 91%, i.e. 1171 atoms were assigned a chemical shift out of a possible 1291. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	485/499 (97%)	197/199 (99%)	190/202 (94%)	98/98 (100%)
Sidechain	564/652 (87%)	350/385 (91%)	204/238 (86%)	10/29 (34%)

Continued on next page...

Continued from previous page...

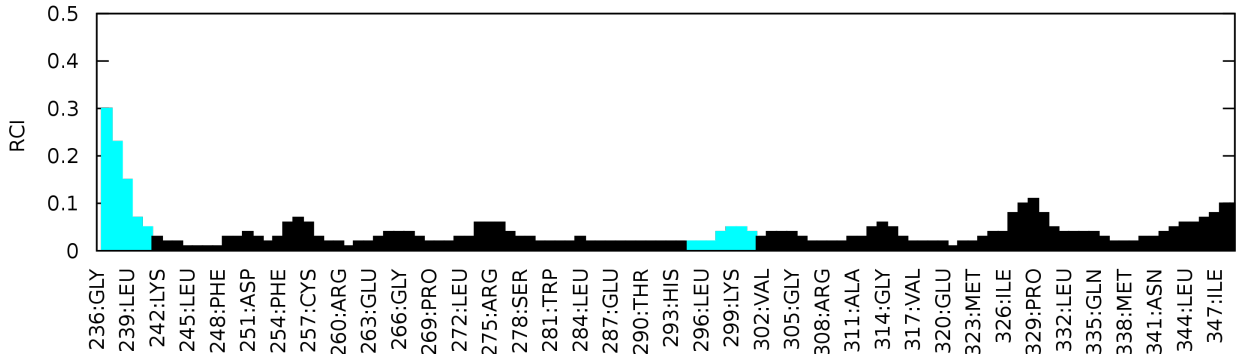
	Total	^{1}H	^{13}C	^{15}N
Aromatic	122/140 (87%)	63/72 (88%)	57/58 (98%)	2/10 (20%)
Overall	1171/1291 (91%)	610/656 (93%)	451/498 (91%)	110/137 (80%)

7.1.4 Statistically unusual chemical shifts (i)

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	308	ARG	HG2	-0.96	2.92 – 0.22	-9.4
1	A	308	ARG	HD3	1.17	4.36 – 1.86	-7.8
1	A	289	MET	HB3	-0.17	3.70 – 0.30	-6.4
1	A	312	LYS	HB3	0.03	3.10 – 0.40	-6.4
1	A	308	ARG	HG3	-0.25	3.00 – 0.10	-6.2
1	A	272	LEU	HD21	-0.93	2.14 – -0.66	-6.0
1	A	272	LEU	HD22	-0.93	2.14 – -0.66	-6.0
1	A	272	LEU	HD23	-0.93	2.14 – -0.66	-6.0
1	A	308	ARG	HD2	1.80	4.27 – 1.97	-5.8
1	A	289	MET	HG2	0.52	4.23 – 0.63	-5.3

7.1.5 Random Coil Index (RCI) plots (i)

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: