wwPDB NMR Structure Validation Summary Report (i)

Apr 20, 2024 – 07:30 PM EDT

PDB ID	:	2KRE
Title	:	Solution structure of E4B/UFD2A U-Box domain
Authors	:	Nomine, Y.; Wasielewski, E.; Botuyan, M.; Mer, G.
Deposited on	:	2009-12-16

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

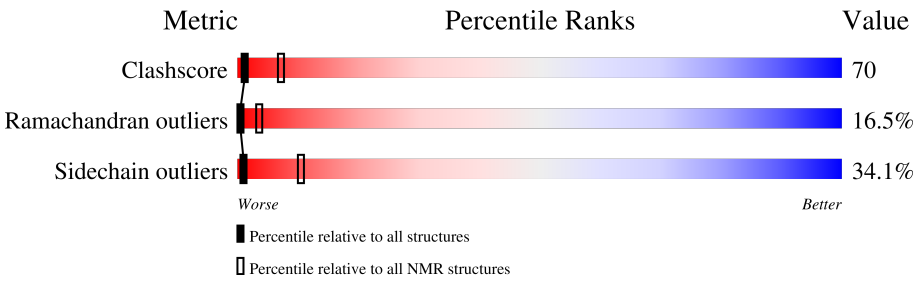
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	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	100	9% 42% 14% • 33%

2 Ensemble composition and analysis (i)

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:1228-A:1294 (67)	0.84	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 and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1600 atoms, of which 792 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ubiquitin conjugation factor E4 B.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	100	1600	501	792	141	161	5	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1203	GLY	-	expression tag	UNP O95155
A	1204	SER	-	expression tag	UNP O95155
A	1205	HIS	-	expression tag	UNP O95155
A	1206	LYS	-	expression tag	UNP O95155
A	1207	PHE	-	expression tag	UNP O95155

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Ubiquitin conjugation factor E4 B

Chain A: 9% 42% 14% • 33%

G1203 S1204 H1205 K1206 F1207 A1208 E1209 K1210 V1211 E1212 E1213 I1214 V1215 A1216 K1217 N1218 A1219 R1220 A1221 E1222 I1223 D1224 Y1225 S1226 D1227 A1228 P1229 D1230 E1231 F1232 R1233 D1234 P1235 L1236 M1237 D1238 T1239 L1240 M1241 T1242 D1243 R1246 L1247 P1248 S1249 G1250 T1251 I1252 M1253 D1254 R1255 S1256 I1257 I1258 L1259 R1260 H1261 L1262 L1263

N1264 S1265 P1266 T1267 D1268 P1269 F1270 N1271 R1272 Q1273 T1274 L1275 T1276 E1277 S1278 M1279 L1280 P1284 E1285 L1286 K1287 E1288 Q1289 I1290 Q1291 A1292 W1293 M1294 R1295 E1296 K1297 Q1298 N1299 S1300 D1301 H1302

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: Ubiquitin conjugation factor E4 B

Chain A: 10% 37% 13% 7% 33%

G1203 S1204 H1205 K1206 F1207 A1208 E1209 K1210 V1211 E1212 E1213 I1214 V1215 A1216 K1217 N1218 A1219 R1220 A1221 E1222 I1223 D1224 Y1225 S1226 D1227 A1228 P1229 F1232 R1233 D1234 P1235 L1236 M1237 D1238 T1239 L1240 M1241 T1242 D1243 R1246 L1247 P1248 S1249 G1250 T1251 I1252 M1253 D1254 R1255 S1256 I1257 I1258 L1259 R1260 H1261 L1262 L1263 N1264

S1265 P1266 T1267 D1268 P1269 F1270 N1271 R1272 Q1273 T1274 L1275 T1276 E1277 S1278 M1279 L1280 E1281 P1282 V1283 P1284 E1285 L1286 K1287 E1288 Q1289 I1290 Q1291 A1292 W1293 M1294 R1295 E1296 K1297 Q1298 N1299 S1300 D1301 H1302

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DISTANCE GEOMETRY, SIMULATED ANNEALING*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
XPLOR-NIH	refinement	2.11.1
CYANA	structure solution	2.1

No chemical shift data was provided.

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	540	540	540	76±11
All	All	10800	10800	10800	1519

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 70.

5 of 626 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1286:LEU:HD13	1:A:1287:LYS:N	1.03	1.66	11	1
1:A:1247:LEU:HD13	1:A:1247:LEU:N	1.00	1.71	20	7
1:A:1247:LEU:HD23	1:A:1247:LEU:O	0.94	1.63	12	3
1:A:1247:LEU:HD22	1:A:1247:LEU:H	0.93	1.17	18	8
1:A:1286:LEU:HD22	1:A:1286:LEU:O	0.92	1.64	11	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	67/100 (67%)	41±4 (61±5%)	15±3 (23±5%)	11±3 (16±5%)	0 3
All	All	1340/2000 (67%)	813 (61%)	306 (23%)	221 (16%)	0 3

5 of 38 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	1255	ARG	17
1	A	1265	SER	17
1	A	1268	ASP	16
1	A	1271	ASN	16
1	A	1274	THR	16

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	64/92 (70%)	42±3 (66±5%)	22±3 (34±5%)	1 10
All	All	1280/1840 (70%)	843 (66%)	437 (34%)	1 10

5 of 51 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	1255	ARG	19
1	A	1276	THR	19
1	A	1268	ASP	18
1	A	1289	GLN	16
1	A	1243	ASP	14

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 monosaccharides 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