wwPDB NMR Structure Validation Summary Report (i)

Aug 9, 2021 – 06:04 PM EDT

PDB ID	:	2KTV
Title	:	Human eRF1 C-domain, "open" conformer
Authors	:	Mantsyzov, A.B.; Polshakov, V.I.; Birdsall, B.
Deposited on	:	2010-02-09

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 (i)) were used in the production of this report:

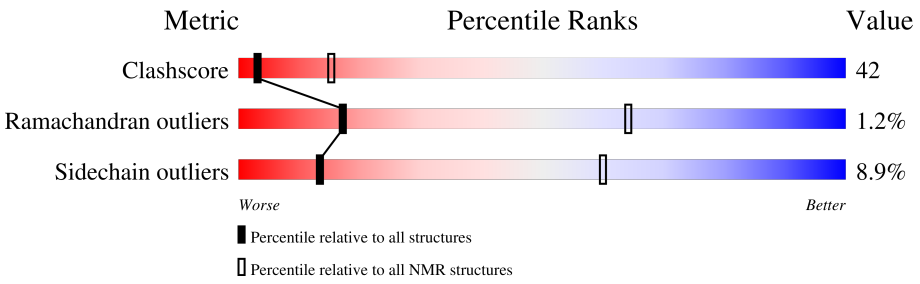
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.1
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

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	170	46% 40% • 7% •

2 Ensemble composition and analysis (i)

This entry contains 24 models. Model 22 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:278-A:334, A:342-A:416 (132)	0.37	22
2	A:421-A:440 (20)	0.94	15

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 4 clusters and 3 single-model clusters were found.

Cluster number	Models
1	4, 8, 11, 12, 15, 17, 18, 19, 20
2	2, 7, 9, 10, 14, 16, 21, 22
3	6, 23
4	1, 13
Single-model clusters	3; 5; 24

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2626 atoms, of which 1284 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Eukaryotic peptide chain release factor subunit 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	164	2626	858	1284	211	267	6	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	438	LEU	-	expression tag	UNP P62495
A	439	GLU	-	expression tag	UNP P62495
A	440	HIS	-	expression tag	UNP P62495
A	441	HIS	-	expression tag	UNP P62495
A	442	HIS	-	expression tag	UNP P62495
A	443	HIS	-	expression tag	UNP P62495
A	444	HIS	-	expression tag	UNP P62495
A	445	HIS	-	expression tag	UNP P62495

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: Eukaryotic peptide chain release factor subunit 1

Chain A: 46% 40% • 7% •

SER N277 V278 I281 Q282 E283 K284 K285 L286 I287 Y290 E293 I294 S295 Q296 D297 T298 G299 K300 Y301 C302 F303 G304 V305 T308 L309 K310 A311 L312 V317 E318 I319 L320 I321 V322 Y323 E324 N325 L326 D327 I328 Y331 V332 L333 H334 C335 Q336 G337 T338 E339 E340 E341 K342 I343

L344 Y345 L346 T347 P348 E351 F357 T358 D359 T362 G363 Q364 E365 H366 E367 L368 I369 E370 S371 M372 P373 L374 L375 F378 Y382 A387 T388 L389 E390 I391 V392 F402 F406 G407 G408 I409 G410 G411 I412 L413 R414 Y415 R416 V417 D418 F419 Q420 H440 HIS HIS HIS

HIS HIS

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

The representative model is number 22. Colouring as in section 4.1 above.

• Molecule 1: Eukaryotic peptide chain release factor subunit 1

Chain A: 39% 44% 5% • 7% •

SER N277 V278 K279 F280 I281 Q282 E283 K284 K285 L286 I287 G288 R289 Y290 F291 D292 E293 I294 S295 T298 G299 K300 Y301 C302 F303 G304 V305 E306 D307 T308 L309 K310 A311 L312 A316 V317 E318 I319 L320 I321 V322 Y323 E324 N325 L326 D327 I328 Y331 V332 L333 H334 C335 Q336 G337 T338 E339

E340 E341 K342 I343 L344 Y345 L346 T347 P348 E349 Q350 E351 D359 T362 G363 Q364 I369 E370 S371 L374 L375 F378 Y382 A387 T388 L389 E390 I391 V392 E398 G399 S400 Q401 F402 V403 K404 G405 F406 G407 G408 I409 G410 G411 I412 L413 R414 Y415 R416 V417 D418 F419 Q420

G421 M422 Q425 H440 HIS HIS HIS HIS HIS

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 100 calculated structures, 24 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CNS	refinement	2.1
Sparky	structure solution	
CNS	structure solution	2.1
ARIA	structure solution	
CNSSOLVE	structure solution	

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	1246	1207	1203	104±9
All	All	29904	28968	28872	2497

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 42.

5 of 558 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:294:ILE:HD12	1:A:301:TYR:CZ	1.14	1.75	20	1
1:A:344:LEU:HD23	1:A:346:LEU:HD11	0.94	1.39	21	7
1:A:305:VAL:HG11	1:A:374:LEU:HD13	0.91	1.43	4	17
1:A:311:ALA:HB1	1:A:413:LEU:HD21	0.91	1.42	21	15
1:A:318:GLU:O	1:A:319:ILE:HD13	0.88	1.67	20	22

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	151/170 (89%)	140±2 (93±1%)	9±2 (6±1%)	2±1 (1±1%)	17	64
All	All	3624/4080 (89%)	3364 (93%)	217 (6%)	43 (1%)	17	64

5 of 9 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	407	GLY	24
1	A	421	GLY	7
1	A	426	GLY	3
1	A	299	GLY	2
1	A	300	LYS	2

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	133/150 (89%)	121±2 (91±1%)	12±2 (9±1%)	13	60
All	All	3192/3600 (89%)	2909 (91%)	283 (9%)	13	60

5 of 40 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	305	VAL	23
1	A	345	TYR	22
1	A	332	VAL	22
1	A	282	GLN	21
1	A	312	LEU	21

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