

Full wwPDB NMR Structure Validation Report (i)

Jun 12, 2024 – 02:53 AM EDT

PDB ID	:	2KG1
BMRB ID	:	16194
Title	:	Structure of the third qRRM domain of hnRNP F in complex with a AGGGAU
		G-tract RNA
Authors	:	Allain, F.H.T.; Dominguez, C.
Deposited on	:	2009-03-02

This is a Full wwPDB NMR Structure Validation 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 (1)) 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 is 76%.

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.



Motric	Whole archive	NMR archive
Meth	$(\# { m Entries})$	$(\# { m Entries})$
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
RNA backbone	4643	676

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	А	139	34%	11% •	29%	24%	
2	В	6	17%	50%		33%	



2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 10 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) p	protein residues	
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:289-A:296, A:300-A:320,	0.63	10
	A:331-A:365 (64)		

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 5 clusters and 3 single-model clusters were found.

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



3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1755 atoms, of which 819 are hydrogens and 0 are deuteriums.

• Molecule 1 is a protein called Heterogeneous nuclear ribonucleoprotein F.

Mol	Chain	Residues		Atoms					Trace
1	٨	105	Total	С	Η	Ν	0	S	0
	A	105	1558	496	752	142	162	6	

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	243	MET	-	expression tag	UNP P52597
А	244	GLY	-	expression tag	UNP P52597
А	245	SER	-	expression tag	UNP P52597
А	246	SER	-	expression tag	UNP P52597
А	247	HIS	-	expression tag	UNP P52597
А	248	HIS	-	expression tag	UNP P52597
А	249	HIS	-	expression tag	UNP P52597
А	250	HIS	-	expression tag	UNP P52597
А	251	HIS	-	expression tag	UNP P52597
А	252	HIS	-	expression tag	UNP P52597
А	253	SER	-	expression tag	UNP P52597
А	254	SER	-	expression tag	UNP P52597
А	255	GLY	-	expression tag	UNP P52597
А	256	LEU	-	expression tag	UNP P52597
А	257	VAL	-	expression tag	UNP P52597
А	258	PRO	-	expression tag	UNP P52597
А	259	ARG	-	expression tag	UNP P52597
А	260	GLY	-	expression tag	UNP P52597
А	261	SER	-	expression tag	UNP P52597
А	262	HIS	-	expression tag	UNP P52597
А	263	MET	-	expression tag	UNP P52597
А	264	ALA	-	expression tag	UNP P52597
А	265	SER	-	expression tag	UNP P52597
А	266	MET	-	expression tag	UNP P52597
А	267	THR	-	expression tag	UNP P52597
А	268	GLY	-	expression tag	UNP P52597
А	269	GLY	-	expression tag	UNP P52597
А	270	GLN	-	expression tag	UNP P52597
А	271	GLN	-	expression tag	UNP P52597
А	272	MET	-	expression tag	UNP P52597
А	273	GLY	-	expression tag	UNP P52597



Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
А	274	ARG	-	expression tag	UNP P52597
А	275	GLY	-	expression tag	UNP P52597
А	276	SER	-	expression tag	UNP P52597

• Molecule 2 is a RNA chain called 5'-R(*AP*GP*GP*GP*AP*U)-3'.

Mol	Chain	Residues	Atoms				Trace		
0	D	6	Total	С	Η	Ν	Ο	Р	0
Δ	D	0	197	59	67	27	39	5	0



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: Heterogeneous nuclear ribonucleoprotein F

Chain A:	34%	11% •	29%	24%
MET GLY SER SER HIS HIS HIS HIS	SER SER GLY LEU VAL LEU PRO ARC GLY SER HIS SER ALA SER MET	THR GLY GLN GLN MET MET ARG GLY SER SER	228 5279 5279 5281 7281 7283 7285 7285 7285 7286 7286 7288 7288 7288 7288 7288 7288	L296 L296 7297 7299 8300 1301 1301 1301 1301 1300 1303 1303
L312 V315 V315 V315 C315 C317 C322 C322 C322 C324 C324 C324 C324 C324	R326 R326 C329 C329 C329 C329 C329 C329 C323 R326 R356 R355 R355 R355 R355 R355 R355 R35	E358 L359 F360 L359 A367 A367 A367 C370 C370 A371	Y372 S373 S374 Q375 V376 Q376 Q378 C379 G379 C381	
• Molecule 2:	5'-R(*AP*GP*GP	*GP*AP*U)-3	5	
Chain B:	17%	50%	33	%
A 1 6 2 0 6 1				

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

 \bullet Molecule 1: Heterogeneous nuclear ribonucleoprotein F



• Molecule 2: 5'-R(*AP*GP*GP*GP*AP*U)-3'



Chain B:	50%			50%	•
A1 62 64 85 06					
4.2.2 Score p	er residue f	or model 2			
• Molecule 1: He	eterogeneous	nuclear ribon	ucleoprotein F		
Chain A:	35%	12%	29%	24%	
MET MET SER SER SER HIS HIS HIS HIS SER SER SER	GLY LEU VAL PRO ARG GLY SER HIS MET ALA	SER MET THR THR GLY GLY GLN MET GLN GLN	Алы GLY SER G277 G277 C277 S279 E280 F281 T282 V283 V283 Q284	S285 1286 1286 1288 6288 1288 1288 1288 1288	F308 F309 V315
R316 V317 1321 1321 0325 P323 0326 N326 N326 N326 N326 N328 C328 C328	E330 T337 M345 M352 Q353 H354 H355 H355	Y356 F360 L361 L361 L361 S368 S368	03009 4371 8373 8374 8375 8374 0375 0376 0375 0379 0379	11380 0381	
• Molecule 2: 5 [*]	-R(*AP*GP*	GP*GP*AP*	U)-3'	_	
Chain B: 17%	6 17%	33	%	33%	
1 2 2 3 3 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9					
4.2.3 Score p	er residue f	or model 3			
• Molecule 1: He	eterogeneous	nuclear ribon	ucleoprotein F		
Chain A:	37%	9%	29%	24%	
计不能推迟迟迟迟迟迟振振	7 8 1 9 9 7 8 S L 7	現在おかえるでない	277 277 2778 281 281 282 283 283 284	285 287 287 288 296 298 298 298 201 201	309 309 321
GI CI CI CI CI CI CI CI CI CI CI CI CI CI	GI PPR AR AR AR AR AR AR AR AR AR AR AR AR AR	SH CILL CILL CILL CILL CILL CILL CILL CIL			
G322 P323 D324 C325 C326 C329 C329 C329 C329 C329 C329 C329 C329	q353 Y356 Y356 E358 E358 E358 F360 F360 L361 L361 C366	A367 S368 N369 G370 A371 Y372 S373 S373 S374 Q375	v377 M377 Q378 G379 M380 G381		
• Molecule 2: 5 [*]	-R(*AP*GP*	GP*GP*AP*	U)-3'		
Chain B:		67%		33%	•
A1 G2 G4 U6					

 \bullet Molecule 1: Heterogeneous nuclear ribonucleoprotein F

N I N N H H H H N N N	AF A	A G G G G G H M S S S S S S S S S S S S S S S S S S		622 622 622 722 622 722 722 722 732 732 732 732 732 732 7	F3 F3 L3
V315 V316 V317 1321 3221 3222 7325 7326 7326 7326 7326 7326 7326	T1328 0329 E330 M345 Q353 Q353	7356 1357 1357 1359 1359 1356 1356 13366 13366 13368 13368	NJ69 G370 S373 S373 S374 S374 M375 M377 Q376 C3778	M380 G381	
• Molecule 2: 5'-	R(*AP*GP*C	GP*GP*AP*U	J)-3'	_	
Chain B: 17%	17%		50%	17%	•
A 1 G 2 A 5 U 6					
4.2.5 Score p	er residue fo	or model 5			
• Molecule 1: He	eterogeneous n	uclear ribonu	cleoprotein F		
Chain A:	36%	9% •	29%	24%	
MET GLY SER SER HIS HIS HIS HIS HIS RIS SER SER	GLY LEU VAL ARG GLY GLY HIS MET ALA	SER MET THR GLY GLY GLN MET GLN MET GLY	GLY SER G277 D277 5279 E280 F281 T282 T282 T283 Q284 Q284	T286 T287 T286 T287 T287 T287 T287 T287 T287 T280 T299 T301 T301 T301 T301	F308 F309 V315
1321 1322 1322 13224 13226 13226 13228 1328 1328 1328 1328 1328 1328 132	D348 R349 A350 R355 Y356 F369	L361 T364 T365 G366 A367 S368 N369 G370 G370	X372 X373 S374 Q375 V376 M377 Q378 Q379 M380 M380		
• Molecule 2: 5'	$\mathbf{R}(*\Delta \mathbf{P}^* \mathbf{C} \mathbf{P}^* \mathbf{C})$		I) ay		
• Molecule 2. 5 -		JP*GP*AP*U)-3		
Chain B: 17%		50%))-3'	33%	
Chain B: 17%		50%))-3	33%	
• Molecule 2. 5 - Chain B: 17% ≈ 8 8 5 2 5 4.2.6 Score p	er residue fo	50% 50% or model 6))-3	33%	
• Molecule 2. 3 - Chain B: 17% ₹ 8 8 5 2 5 4.2.6 Score p • Molecule 1: He	er residue fo	50% 50% 50% Dr model 6 nuclear ribonu))-3 [°] cleoprotein F	33%	
 Molecule 2. 3 - Chain B: 17% ■ 8 8 8 9 2 8 4.2.6 Score p Molecule 1: H€ Chain A: 	er residue fo eterogeneous n 36%	50% 50% or model 6 nuclear ribonu 7% •	cleoprotein F	33%	
Molecule 2. 3 - Chain B: 17% ■ 8 8 8 2 8 4.2.6 Score p Molecule 1: He Chain A: ■ 5 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	er residue fo eterogeneous n 36%	50% 50% or model 6 nuclear ribonu 7% 。	cleoprotein F 29%	24%	1306 1307 1308 1309 1309
Molecule 2. 3 - Chain B: 17% 17% 2 8 8 8 8 8 8 4.2.6 Score p • Molecule 1: He Chain A: 5 6 6 6 8 8 8 8 8 8 8 8 8 8 8 8 8 8	er residue fo eterogeneous n 36%	50% 50% or model 6 nuclear ribonu 7% 8 9 1 3 3 3 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	cleoprotein F 29% 55 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	33% 24% 24% 24% 24%	Y 306 N 307 F 308 F 309
Molecule 2: 5 ⁻ Chain B: 17% 17%	er residue fo eterogeneous n 36% E E E E E E E E E E E E E E E E E E E	50% 50% or model 6 nuclear ribonu 7% ۰ قوق قوق قوق قوق GP*GP*AP*U	cleoprotein F 29% 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	33% 24% 24% 8820 1001 1001 1001 1001 1001 1001 1001	Y306 N307 F308 F309
Molecule 2: 5 ⁻ Chain B: 17% 17% 17% 17% 17% 17% 17% 17% 17%	er residue fo eterogeneous n 36%	50% 50% or model 6 nuclear ribonu 7% 7% 8 8 8 8 8 8 8 9))-3 [°] cleoprotein F 29% 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	33% 24% 28% 18% 1001 1001 1001 1001 1001 1001 100	2008 1300 1300 1300 1300



A1	G2	G3	G4	A5	U6

4.2.7 Score per residue for model 7

• Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A:	32%	13% •	29%	24%	-
MET GLY SER SER HIS HIS HIS HIS HIS SER SER SER	GLY GLY LEU VAL PRO ARG GLY SER HIS MET ALA	SER MET THR GLY GLN GLN MET GLN MET ARG	GLY SER G277 D278 5279 5279 F281 T282 V283	2285 2285 1286 1288 1288 1288 1288 1289 1289 1296 1296 1296 1296 1296 1296 1296 129	E302 N303 F308 P311
L312 V315 V315 V316 V317 V317 V317 V326 C322 D324 C322 D324 C325 C325 C325 C325 C325 C325 C325 C325	V327 V328 C329 C329 C329 C329 C335 A336 A344 A344	M345 A350 1367 1367 1367 1358 1358 1358 1356	N362 G366 A367 S368 N369 G370 A371 Y372	S374 Q375 M377 Q378 Q378 G379 M380 G381 C381	
• Molecule 2: 5	'-R(*AP*GP*0	GP*GP*AP*	U)-3'		
Chain B:	50%			50%	-
A1 62 63 64 85 U6					
4.2.8 Score j	per residue fo	or model 8			
• Molecule 1: H	leterogeneous r	uclear ribonu	icleoprotein F		
Chain A:	37%	7% •	29%	24%	I
MET GLY SER SER HIS HIS HIS HIS SER SER SER	GLY CLEU VAL PRO PRO GLY SER HIS MET ALA	SER NET GLY GLY GLN GLN MET ARG	GLY SER 6277 6277 5279 5279 5281 7282 V283	8285 8285 1286 1287 6288 8298 17298 17298 17299 17299 18303 18303 18303 18303	V315 R316 V317 I321
6322 P323 D324 6326 6326 7326 7326 7328 6329 6329 6330 6330	A350 A350 H354 R355 Y356 L357 F360 F360 L361	G366 A367 A367 S368 N369 G370 A371 Y372 S374 S374	Q375 V376 Q379 G379 M380 G381		
• Molecule 2: 5	'-R(*AP*GP*0	GP*GP*AP*	U)-3'		
Chain B: 17	%	50%		33%	-
A1 G2 G6 A5 U6					
4.2.9 Score J	per residue fo	or model 9			
• Molecule 1: H	leterogeneous r	uclear ribonu	icleoprotein F		
Chain A:	35%	10% •	29%	24%	-



MET GLY SER SER HIS HIS HIS HIS HIS SER SER SER	GLY LEU VAL PRO ARG GLY SER HIS ALA	SEA MET THR GLY GLY GLN GLN GLN ARG	GLY SER G277 G277 S279 E280 F281 T282 V283 V283	4204 5285 7286 7287 6288 4289 1296 1296	P297 Y298 K299 A300 T301 E302 N303	F308 F309
V315 V315 V317 C325 C325 C325 C325 C325 C325 C325 C325	<mark>6329</mark> E330 M345 M345 Y356	L359 F360 L361 C365 G366 G366 A367 S368 S368 S368 N369	G370 A371 S372 S373 S374 0375 M377 M377 C378	M380 G381		
• Molecule 2: 5'-	-R(*AP*GP*C	GP*GP*AP*U	U)-3'			
Chain B:	50%			50%		
A1 G2 G4 A5 U6						
4.2.10 Score	per residue f	for model 1	0 (medoid)			
• Molecule 1: He	eterogeneous n	uclear ribonu	icleoprotein F	م		
Chain A:	37%	7% •	29%	24%	%	
MET GLY SER SER HIS HIS HIS HIS HIS SER SER	GLY LEU VAL PRO ARG GLY SER HIS MET ALA	AFR MET GLY GLY GLN MET GLN ARG	GLY SER G277 G277 S279 F280 F281 T282 V283	8285 1286 1286 6288 1296 1296 1296	Y 298 K 299 A 300 E 301 E 302 N 303	F308 F309 V315
R316 V317 1321 6322 6325 6325 6325 7326 7326 7328 7328 7329 7329	E330 H354 H355 Y356 I357 E358 F360 F360	G366 A367 S368 N369 N369 G370 A371 Y372 S373 S374	Q375 V376 M377 Q378 G379 M380 G381			
• Molecule 2: 5'-	-R(*AP*GP*C	GP*GP*AP*U	U)-3'			
Chain B: 17%	, ,	50%		17%	17%	
41 62 64 45 UG						
4.2.11 Score	per residue f	for model 1	1			
• Molecule 1: He	eterogeneous n	uclear ribonu	icleoprotein F	r		
Chain A:	37%	8% •	29%	24%	%	
MET GLY SER SER HIS HIS HIS HIS SER SER SER	GLY LEU VAL PRO ARG GLY SER HIS MET ALA	AER MET THR GLY GLY GLN GLN GLN MET ARG	GLY SER G277 G277 S279 S279 E280 F281 T282 V283 V283	2285 7286 7287 6288 6288 7297 7298	K299 A300 E302 N303 N303 N307	F308 F309 V315
1321 1321 1322 1325 1325 1328 1328 1328 1328 1328 1328 1328 1328	D348 R349 A350 L355 L359 L359 L350 L361 L361	G3 <mark>6</mark> 6 A367 S368 N369 G370 A371 Y372 S373 S374	Q375 V376 M377 Q378 G379 G381 G381			
• Molecule 2: 5'-	-R(*AP*GP*C	GP*GP*AP*U	IJ)-3'			
Chain B:	50%			50%		



A1 G2 G3 G4 A5 U6

4.2.12Score per residue for model 12

• Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A:	37%	8% •	29%	24%
MET CLY SER SER SER HIS HIS HIS HIS HIS SER SER SER CLEU	VAL PRO ARG GLY SER HIS ALA ALA SER MET THR	GLY GLY GLN GLN GLN MET GLY GLY SER	627 (1627 (1228 1 1281 1282 1283 1288 1288 1288 1288 1288 1288 1288 1288	P297 Y299 K299 R300 F301 F301 F302 F308 F308 F308 F308 F308 F308
V317 1321 1321 1322 1322 1324 1326 1326 1328 1328 1328 1337	q353 4354 4355 7356 7356 1356 1366 1366	A367 S368 N369 A371 Y372 S373 S374 Q375	M377 Q377 Q379 M380 M380 C381	
• Molecule 2: 5'-R((*AP*GP*GP*	GP*AP*U)-3	3,	
Chain B: 17%			83%	
A1 62 A5 A5 U6				
4.2.13 Score pe	r residue for	model 13		
• Molecule 1: Hete	rogeneous nucl	ear ribonucle	oprotein F	
Chain A:	35%	7% •	29%	24%
MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER SER SER CLEU	VAL PRO ARG GLY SER HIS MET ALA SER MET THR	GLY GLY GLN GLN GLN MET GLY GLY SER SER	62/ 62/ 5278 5280 5281 5281 7282 7283 7285 7285 7286 7286 7286 7287 7286 7287 7286	L296 P297 Y298 K299 K299 K299 K299 E301 E301 E301 E301 E301 F311
V315 N316 V317 1321 1321 1321 1322 1322 1328 1328 1328	T337 T337 R349 H354 R355 Y3556 T357 E358	F360 L361 C366 A367 S368 M369 C370 C370	5373 5375 5375 4375 4375 4376 6379 6379 6379 6381	
• Molecule 2: 5'-R((*AP*GP*GP*	GP*AP*U)-3	3,	
Chain B:	67%		33	%
A1 62 64 85 U6				
4.2.14 Score pe	r residue for	model 14		
• Molecule 1: Hete	rogeneous nucl	ear ribonucle	oprotein F	
Chain A: 29	9% 14	% •	29%	24%



	00017440042	A C C C C C C C C C C C C C C C C C C C		(128 (129 (129 (129 (129 (129 (129 (129 (129	N30 F30 F30
• Molecule 2:	5'-R(*AP*GF	8 10 - 8 80 - 80 80 - 80 80 80 80 80 80 80 80 80 80 80 80 80 	99 - <mark>88 99 19 - 98 99 88 99 19 99 99 99 99 99 99 99 99 99 99 99 </mark>	8372 8374 8374 8376 8376 8377 8377 8379 8378 8379 8378 8378 8378	
Chain B:	17%	50%		33%	
A1 62 63 86 86 06					
4.2.15 Scor	re per residu	e for model 13	5		
• Molecule 1:	Heterogeneou	s nuclear ribonu	cleoprotein F		
Chain A:	34%	11% •	29%	24%	
MET GLY SER HIS HIS HIS HIS HIS HIS	SER SER GLY LEU VAL PRO ARG GLY SER HIS	ALLA SER SER SER SER GLY GLY GLY GLN GLY GLY ARG	GLIY SER 6277 6277 6277 6277 7281 7280 7283 7285 7285 7285 7286	1267 (1288 (1289 (1299 (1299 (1299 (1299 (1299) (1290) (1290) (1290) (1290) (1290) (1290) (1290) (1290) (1290) (1290) (1290) (1290) (1200) (12	N307 F308 F309
V315 1321 0322 P323 P324 C325 C325 C325 C325 C325 C325 C325 C325	G329 E330 T337 T337 A350 A350 A353 A350 A353 A350	1355 1355 1359 1359 1356 1366 1366 1366 1366 1368 1368 1368 136	A371 Y372 S372 S374 S375 Q375 Q376 Q378 Q378 M380 M380 C381		
• Molecule 2:	5'-R(*AP*GF	P*GP*GP*AP*U	J)-3'		
Chain B:		67%		33%	
Chain B:		67%		33%	
Chain B:	re per residu	67% le for model 16	3	33%	
Chain B:	re per residu Heterogeneou	^{67%} l e for model 16 s nuclear ribonu	3 cleoprotein F	33%	
Chain B: 4.2.16 Sco • Molecule 1: Chain A:	re per residu Heterogeneou ^{34%}	67% le for model 16 s nuclear ribonu 9% •	3 cleoprotein F 29%	33%	
Chain B: 4.2.16 Sco • Molecule 1: Chain A: E S S S E E E E E E E	re per residu Heterogeneou 34%	67% le for model 16 s nuclear ribonu 9% 。	5 cleoprotein F 29%	24%	F300 F310 F311 L312
Chain B: 4.2.16 Scor • Molecule 1: Chain A: E Chain A: E Chain A: Chain B: E Chain B: E C Chain B: E C C C C C C C C C C C C C C C C C C C	re per residu Heterogeneou 34%	67% e for model 16 s nuclear ribonu 9% 。	3 cleoprotein F 29% 22%	24%	F300 F300 P311 L312
Chain B: • 8 8 8 8 9 4.2.16 Scor • Molecule 1: Chain A: • 5 8 8 9 9 9 • 5 8 8 9 9 • 6 8 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	re per residu Heterogeneou ^{34%} S S S S S S S S S S S S S S S S S S S	67% e for model 16 s nuclear ribonu 9% • e \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ 8 \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	5 cleoprotein F 29% 5 5 5 5 5 5 5 5 5 5	24%	F300 F309 P311 L312
Chain B: • 8 8 8 8 9 4.2.16 Scor • Molecule 1: Chain A: • Se 8 8 8 8 8 8 8 8 8 8 • Se 8 8 8 8 8 8 8 8 8 8 8 • Molecule 2: Chain B:	re per residu Heterogeneou ^{34%} S S S S S S S S S S S S S S S S S S S	67% e for model 16 s nuclear ribonu 9% e 1 % e # 5 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	5 cleoprotein F 29% 5 5 5 5 5 5 5 5 5 5	33%	F300 F300 P311 L312

R L D W I D E PDB TEIN DATA BANK

A1 G2 G3 G4 A5 U6

4.2.17 Score per residue for model 17

• Molecule 1: Heterogeneous nuclear ribonucleoprotein F

Chain A:	37%	8% •	29%	24%	-
MET CLY SER SER HIS HIS HIS HIS HIS	SER SER GLY VAL PRO PRO GLY SER HIS MAT	SER MET THR GLY GLY GLN GLN MET ARG	GLY SER G277 G277 B278 E280 F281 T282 V283 Q284 Q284	S285 T286 T286 G288 R294 F297 F298 K299 K299	R316 V317 I321 G322
P323 D324 C325 C325 C325 C325 C325 C325 C329 E332 E334 E334	D348 R349 A350 A353 H354 R355 F356 L359 L359	L361 L361 A367 A367 A367 A371 Y372 S372	S374 Q375 M377 Q377 Q378 G379 M380 G381		
• Molecule 2:	5'-R(*AP*GP*	GP*GP*AP*U	U)-3'		
Chain B:	17%	33%		50%	-
A1 G2 G3 G4 V6 U6					
4.2.18 Scor	e per residue	for model 1	8		
• Molecule 1:	Heterogeneous	nuclear ribonu	cleoprotein F		
Chain A:	33%	13%	29%	24%	-
MET GLY SER SER HIS HIS HIS HIS	SER SER GLY LEU VAL PRO PRO GLY SER HIS MET	SER MET GLY GLY GLN GLN MET GLN ARG	GLY SER 5277 5279 5279 5279 5281 7282 7283 7283 7284	S285 1286 1286 1288 1286 1288 1286 1288 1286 1288 1286 1286	E302 N303 F308 F309
L312 V315 V315 V317 V317 C322 P323 P324	6325 R326 V327 C329 E330 E330 A341 A341	1353 1355 1356 1359 1356 1356 1356 1356 1356	G366 A367 S368 N369 G370 A371 Y372 S373 S374 S374 C375	V376 M377 G379 M380 M380 G381	
• Molecule 2:	5'-R(*AP*GP*	GP*GP*AP*U	U)-3'		
Chain B:		83%		17%	-
A1 G2 G3 G4 U6					
4.2.19 Scor	e per residue	for model 1	9		
• Molecule 1:	Heterogeneous	nuclear ribonu	cleoprotein F		
Chain A:	36%	7% •	29%	24%	-



MET GLY SER SER HIS HIS HIS HIS	SER SER GLY CLU CLU LEU PRO ARC CLY FRO ARC CLY SER HIS	ALA ALA MET THR THR GLY GLN GLN GLN GLN GLN GLY SER SER	G277 G277 S279 F281 F281 F281 V283 V283 V283 7285 F285 F286 F286 F286 F286 F286 F286 F286 F286	R294 6296 12966 12968 8299 8299 8299 8299 8303 8303 1312 1312
V315 1321 6322 6322 1323 1324 6325 8325 8325 8325 8325 8325 8325 8325 8	T328 C329 E320 T337 M345 M345 R349 Y356	L359 F360 L361 L361 A367 A367 N369 A371 A371 Y377 Y377 S373	8374 0375 N377 0378 0379 M380 0381	
• Molecule 2:	5'-R(*AP*GP*	GP*GP*AP*U)-	3'	
Chain B:	17%	50%	3	3%
A1 62 63 64 86 U6				
4.2.20 Sco	re per residue	e for model 20		
• Molecule 1:	Heterogeneous	nuclear ribonucle	oprotein F	
Chain A:	26%	17% •	29%	24%
MET GLY SER SER HIS HIS HIS HIS HIS	SER SER GLY LEU VAL PRO ARG GLY SER HIS	ALA SER MET THR CLY CLY GLN GLN GLN GLN GLN GLY SER SER	C277 D278 D278 C280 E280 E281 T282 T282 C283 C288 C288 C288 C288 C288 C288 C	R294 G295 I_296 R294 G295 R299 R301 F301 F301 F301 F302 B304
F308 F309 S310 P311 L312 L312 N313 P314 V315 R315 R315 R315 V317	1321 1321 1322 1322 1323 1324 1326 1326 1326 1328 1329 1329 1330	1337 1337 1334 1334 1334 1334 1335 1335 1355 1355	1357 E358 E358 E356 F356 F366 M367 A367 A367 A367 A367 A367 A367 A367 A	Y372 8373 8374 0376 10376 0378 0378 0379 0381 0381
• Molecule 2:	5'-R(*AP*GP*	GP*GP*AP*U)-	3'	
Chain B:				
	17%	33%	50%	



5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 20 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
CYANA	structure solution	2.1
Amber	refinement	

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1097
Number of shifts mapped to atoms	1097
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%



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		E	Bond lengths	Bond angles	
WIOI	Ullalli	RMSZ	$\#Z{>}5$	RMSZ	$\#Z{>}5$
1	А	$0.49 {\pm} 0.00$	$0{\pm}0/531~(~0.0{\pm}~0.0\%)$	0.87 ± 0.02	$0{\pm}0/722~(~0.1{\pm}~0.1\%)$
2	В	1.12 ± 0.01	$0{\pm}0/146~(~0.0{\pm}~0.0\%)$	1.81 ± 0.11	$5{\pm}1/227~(~2.3{\pm}~0.6\%)$
All	All	0.68	0/13540 ($0.0%$)	1.16	111/18980~(~0.6%)

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	А	$0.0{\pm}0.0$	0.1 ± 0.3
2	В	$0.0{\pm}0.0$	$0.5 {\pm} 0.7$
All	All	0	11

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.

Mal	Chain	Dec	Turne	Atoma	7	Obcomrod ⁽⁰⁾	Ideal(0)	Mod	dels
	Unam	nes	туре	Atoms	2	Observed()	Ideal()	Worst	Total
2	В	4	G	O4'-C1'-N9	12.31	118.05	108.20	20	15
2	В	4	G	C1'-O4'-C4'	-11.09	101.03	109.90	18	20
2	В	1	А	O4'-C1'-N9	8.88	115.30	108.20	11	12
2	В	2	G	O4'-C1'-N9	8.03	114.62	108.20	19	13
2	В	5	А	O4'-C1'-N9	7.55	114.24	108.20	16	3
2	В	2	G	C1'-O4'-C4'	-7.52	103.88	109.90	19	4
2	В	1	А	C1'-O4'-C4'	-7.46	103.93	109.90	7	11
2	В	5	А	C1'-O4'-C4'	-7.46	103.93	109.90	16	12
1	А	356	TYR	CB-CG-CD2	-6.42	117.15	121.00	8	8
2	В	2	G	C5'-C4'-C3'	-6.17	106.12	116.00	6	2
2	В	5	А	P-O3'-C3'	5.55	126.36	119.70	18	8
2	В	3	G	O4'-C1'-N9	5.29	112.43	108.20	5	1
2	В	4	G	C5'-C4'-O4'	5.03	115.14	109.10	16	2



There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	В	2	G	Sidechain	3
2	В	5	А	Sidechain	3
1	А	296	LEU	Peptide	2
2	В	4	G	Sidechain	2
2	В	3	G	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	А	518	489	489	4 ± 3
2	В	130	67	67	4 ± 2
All	All	12960	11120	11120	120

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 5.

Atom 1	Atom 2	$Cleah(\lambda)$	Distance(Å)	Models	
Atom-1	Atom-2	Clash(A)	Distance(A)	Worst	Total
2:B:3:G:H1'	2:B:4:G:C8	0.60	2.32	13	12
1:A:356:TYR:CD2	2:B:3:G:C6	0.60	2.90	14	6
1:A:289:HIS:CE1	1:A:341:ALA:HB2	0.55	2.35	18	1
1:A:356:TYR:CD2	2:B:3:G:C2	0.55	2.94	19	1
1:A:360:PHE:CD2	2:B:5:A:C6	0.54	2.96	20	5
1:A:358:GLU:CB	2:B:3:G:H22	0.54	2.16	20	1
1:A:356:TYR:CE2	2:B:3:G:C5	0.54	2.96	10	3
1:A:360:PHE:CE2	2:B:6:U:C6	0.54	2.96	6	4
1:A:289:HIS:CD2	1:A:337:THR:HA	0.52	2.40	13	11
1:A:356:TYR:CE1	2:B:3:G:C5	0.52	2.97	20	2
1:A:359:LEU:C	2:B:5:A:H61	0.51	2.08	7	8
2:B:5:A:H1'	2:B:6:U:C6	0.50	2.41	13	8
1:A:356:TYR:CD1	2:B:3:G:C6	0.50	2.99	20	3
2:B:1:A:H1'	2:B:2:G:C8	0.49	2.42	5	2

All unique clashes are listed below, sorted by their clash magnitude.



	A trave 2	$C = \left(\begin{pmatrix} \lambda \\ \lambda \end{pmatrix} \right)$	\mathbf{D}	Models	
Atom-1	Atom-2	Clash(A)	Distance(A)	Worst	Total
1:A:306:TYR:CD1	1:A:317:VAL:HG11	0.49	2.42	6	1
1:A:354:HIS:CG	1:A:355:ARG:N	0.49	2.81	15	1
1:A:294:ARG:HD3	2:B:4:G:C5	0.48	2.42	17	1
1:A:356:TYR:O	1:A:356:TYR:CD1	0.47	2.67	1	2
2:B:3:G:H1'	2:B:4:G:C5	0.47	2.45	8	1
1:A:356:TYR:CE2	2:B:3:G:C4	0.46	3.03	18	4
1:A:359:LEU:H	2:B:5:A:N6	0.46	2.09	6	2
2:B:6:U:C6	2:B:6:U:H5'	0.46	2.46	1	9
1:A:358:GLU:HG3	2:B:3:G:H21	0.45	1.71	13	1
1:A:294:ARG:HD3	1:A:294:ARG:C	0.45	2.32	19	1
1:A:356:TYR:CE1	2:B:3:G:C4	0.45	3.05	15	1
1:A:356:TYR:O	1:A:356:TYR:CG	0.44	2.70	4	4
1:A:335:PHE:CZ	1:A:344:ALA:CB	0.44	3.01	7	2
1:A:310:SER:CB	1:A:311:PRO:CD	0.44	2.96	20	1
1:A:312:LEU:CD1	1:A:312:LEU:N	0.43	2.81	19	1
1:A:358:GLU:CD	2:B:4:G:C8	0.43	2.92	13	1
2:B:3:G:C8	2:B:3:G:H5'	0.43	2.48	5	1
1:A:294:ARG:CZ	2:B:3:G:C5	0.43	3.02	20	1
1:A:352:MET:CG	1:A:353:GLN:H	0.42	2.28	2	1
1:A:359:LEU:CD1	1:A:359:LEU:N	0.42	2.82	10	2
1:A:349:ARG:H	2:B:5:A:N6	0.42	2.12	14	2
2:B:3:G:O4'	2:B:4:G:C8	0.42	2.73	19	1
1:A:303:ASN:HD22	1:A:303:ASN:C	0.42	2.18	20	1
1:A:358:GLU:CG	2:B:3:G:H21	0.41	2.29	6	1
1:A:312:LEU:HD12	1:A:316:ARG:CZ	0.41	2.45	20	1
1:A:310:SER:HB3	1:A:311:PRO:HD2	0.41	1.92	13	1
2:B:6:U:H5'	2:B:6:U:C6	0.41	2.51	8	1
1:A:310:SER:C	1:A:312:LEU:H	0.41	2.19	16	2
1:A:353:GLN:O	1:A:355:ARG:N	0.40	2.54	20	1
1:A:316:ARG:HB2	1:A:334:GLU:H	0.40	1.77	17	1
1:A:312:LEU:HD11	1:A:344:ALA:HB2	0.40	1.92	20	1
2:B:5:A:H4'	2:B:6:U:OP1	0.40	2.16	17	1
1:A:303:ASN:HD22	1:A:304:ASP:N	0.40	2.14	20	1

Continued from previous page...

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



Mol	Chain	Analysed	Favoured	'avoured Allowed		Percentiles		
1	А	64/139~(46%)	48 ± 2 (74±3%)	$12\pm2~(19\pm3\%)$	$4\pm2~(6\pm2\%)$	3 19		
All	All	1280/2780~(46%)	953 (74%)	248 (19%)	79~(6%)	3 19		

was analysed and the total number of residues.

All 15 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	\mathbf{Res}	Type	Models (Total)
1	А	315	VAL	18
1	А	350	ALA	10
1	А	353	GLN	10
1	А	348	ASP	7
1	А	354	HIS	7
1	А	312	LEU	6
1	А	317	VAL	4
1	А	296	LEU	4
1	А	365	THR	3
1	А	311	PRO	3
1	А	349	ARG	3
1	А	364	THR	1
1	А	310	SER	1
1	А	355	ARG	1
1	А	337	THR	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	Perc	entile	es
1	А	56/112~(50%)	47 ± 2 (83±3%)	$9\pm2~(17\pm3\%)$	5	41	
All	All	1120/2240 (50%)	934 (83%)	186 (17%)	5	41	

All 20 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	А	303	ASN	20
1	А	359	LEU	20



OVC	1
$2 \Lambda G$	Т

Mol	Chain	Res	Type	Models (Total)
1	А	360	PHE	20
1	А	361	LEU	20
1	А	308	PHE	18
1	А	301	THR	17
1	А	309	PHE	14
1	А	356	TYR	10
1	А	317	VAL	10
1	А	358	GLU	9
1	А	345	MET	8
1	А	354	HIS	7
1	А	307	ASN	3
1	А	310	SER	3
1	А	362	ASN	2
1	А	355	ARG	1
1	А	294	ARG	1
1	A	296	LEU	1
1	А	312	LEU	1
1	А	313	ASN	1

Continued from previous page...

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	В	5/6~(83%)	5 ± 0 (96 $\pm8\%$)	$4\pm1~(74\pm19\%)$	$0.02{\pm}0.04$
All	All	113/120~(94%)	96~(85%)	74~(65%)	0.02

The overall RNA backbone suiteness is 0.02.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	В	4	G	20
2	В	5	А	20
2	В	6	U	20
2	В	2	G	18
2	В	3	G	18

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	В	5	A	20
2	В	4	G	17



Continued from previous page	Continued f	rom previous	page
------------------------------	-------------	--------------	------

Mol	Chain	Res	Type	Models (Total)
2	В	2	G	15
2	В	1	А	13
2	В	3	G	9

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)

The completeness of assignment taking into account all chemical shift lists is 76% for the well-defined parts and 75% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	1097
Number of shifts mapped to atoms	1097
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	${\rm Correction}\pm{\rm precision},ppm$	Suggested action
$^{13}C_{\alpha}$	89	-0.15 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	88	-0.01 ± 0.25	None needed (< 0.5 ppm)
$^{13}C'$	0		None (insufficient data)
^{15}N	101	0.00 ± 0.57	None needed (< 0.5 ppm)

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 76%, i.e. 760 atoms were assigned a chemical shift out of a possible 995. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^{1}\mathrm{H}$	$^{13}\mathrm{C}$	$^{15}\mathbf{N}$
Backbone	246/317 (78%)	127/127~(100%)	57/128~(45%)	62/62~(100%)
Sidechain	412/470 (88%)	282/306~(92%)	120/145~(83%)	10/19~(53%)



	J I I.)		
	Total	$^{1}\mathrm{H}$	$^{13}\mathrm{C}$	$^{15}\mathbf{N}$
Aromatic	55/93~(59%)	35/48~(73%)	20/40~(50%)	0/5~(0%)
Sugar	36/66~(55%)	36/36~(100%)	0/30~(0%)	$0/0 \ (\%)$
Base	11/49~(22%)	11/31~(35%)	0/9~(0%)	0/9~(0%)
Overall	760/995 (76%)	491/548~(90%)	197/352~(56%)	72/95~(76%)

Continued from previous page...

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 75%, i.e. 1096 atoms were assigned a chemical shift out of a possible 1460. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^{1}\mathrm{H}$	$^{13}\mathrm{C}$	$^{15}\mathbf{N}$
Backbone	401/527~(76%)	211/216 (98%)	89/210~(42%)	101/101~(100%)
Sidechain	575/697~(82%)	394/453~(87%)	168/217~(77%)	13/27~(48%)
Aromatic	73/121~(60%)	47/61 (77%)	26/55~(47%)	0/5~(0%)
Sugar	36/66~(55%)	36/36~(100%)	0/30~(0%)	0/0~(-%)
Base	11/49~(22%)	11/31~(35%)	0/9~(0%)	0/9~(0%)
Overall	1096/1460~(75%)	699/797~(88%)	283/521 (54%)	114/142~(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.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	В	3	G	H5"	2.95	2.98 - 5.38	-5.1

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. If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:





