Full wwPDB X-ray Structure Validation Report (i)

Jun 18, 2024 – 09:05 PM EDT

PDB ID	:	4CEM
Title	:	Crystal structure of the first MIF4G domain of human nonsense mediated decay factor UPF2
Authors	:	Clerici, M.; Deniaud, A.; Boehm, V.; Gehring, N.H.; Schaffitzel, C.; Cusack, S.
Deposited on	:	2013-11-11
Resolution	:	2.60 Å(reported)

This is a Full wwPDB X-ray 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/XrayValidationReportHelp
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:

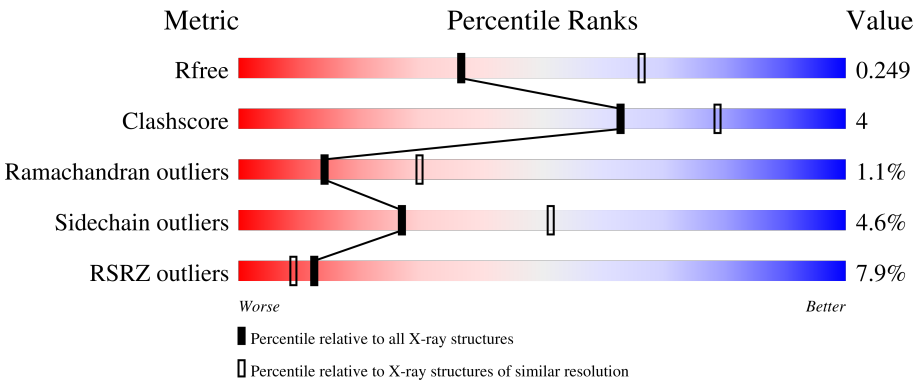
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 Å.

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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	370	7% 71% 11% • 17%
1	B	370	6% 73% 9% • 16%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5115 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called REGULATOR OF NONSENSE TRANSCRIPTS 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	308	2526	1588	456	475	4	3	0	0	0
1	B	309	2535	1593	458	477	4	3	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	117	GLY	-	expression tag	UNP Q9HAU5
A	118	ALA	-	expression tag	UNP Q9HAU5
A	119	MSE	-	expression tag	UNP Q9HAU5
A	120	GLY	-	expression tag	UNP Q9HAU5
A	168	ILE	LEU	conflict	UNP Q9HAU5
B	117	GLY	-	expression tag	UNP Q9HAU5
B	118	ALA	-	expression tag	UNP Q9HAU5
B	119	MSE	-	expression tag	UNP Q9HAU5
B	120	GLY	-	expression tag	UNP Q9HAU5
B	168	ILE	LEU	conflict	UNP Q9HAU5

- Molecule 2 is water.

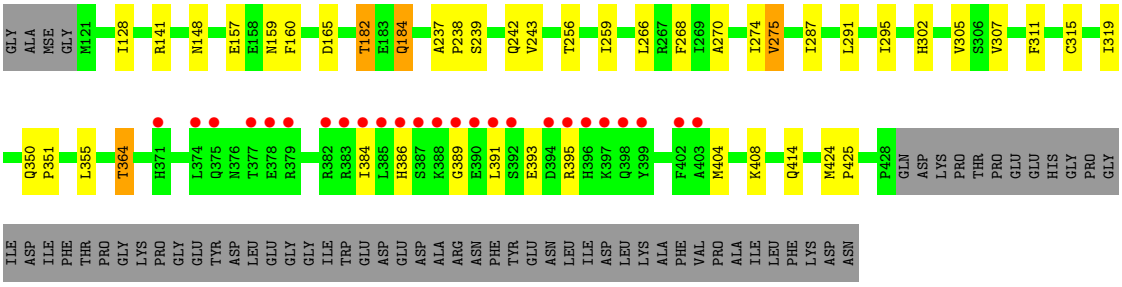
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	20	20	20	0	0
2	B	34	34	34	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

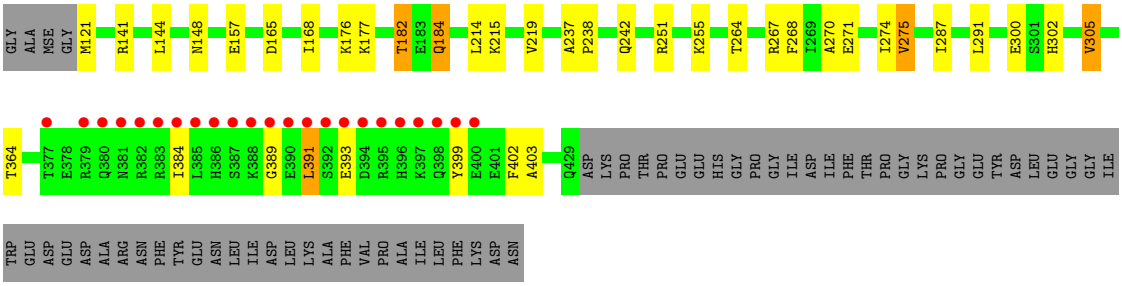
• Molecule 1: REGULATOR OF NONSENSE TRANSCRIPTS 2

Chain A:



• Molecule 1: REGULATOR OF NONSENSE TRANSCRIPTS 2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	101.08Å 101.08Å 158.88Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	87.54 – 2.60 48.16 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (87.54-2.60) 99.9 (48.16-2.60)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.6.0116	Depositor
R, R_{free}	0.205 , 0.251 0.211 , 0.249	Depositor DCC
R_{free} test set	1428 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	46.1	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 32.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.058 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5115	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 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 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	0/2573	0.55	0/3457
1	B	0.49	0/2582	0.59	0/3469
All	All	0.48	0/5155	0.57	0/6926

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2526	0	2528	23	0
1	B	2535	0	2536	20	0
2	A	20	0	0	1	0
2	B	34	0	0	3	0
All	All	5115	0	5064	43	0

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 4.

All (43) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:414:GLN:HA	1:A:424:MSE:HE1	1.53	0.89

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:384:ILE:HG22	1:B:384:ILE:O	1.94	0.67
1:A:148:ASN:HD22	1:A:275:VAL:HG13	1.59	0.67
1:A:148:ASN:ND2	1:A:275:VAL:HG13	2.11	0.65
1:A:141:ARG:HD3	1:A:274:ILE:HG21	1.78	0.65
1:A:404:MSE:HE2	1:A:408:LYS:HE3	1.80	0.64
1:B:148:ASN:HD22	1:B:275:VAL:HG13	1.63	0.63
1:A:414:GLN:HG2	1:A:424:MSE:HE3	1.82	0.61
1:A:364:THR:HG23	1:A:425:PRO:HG3	1.83	0.61
1:B:148:ASN:ND2	1:B:275:VAL:HG13	2.18	0.58
1:A:182:THR:HG23	1:A:184:GLN:OE1	2.05	0.56
1:B:182:THR:HG21	2:B:2006:HOH:O	2.05	0.56
1:B:182:THR:HG23	1:B:184:GLN:OE1	2.06	0.55
1:B:237:ALA:HB3	1:B:238:PRO:HD3	1.87	0.55
1:B:214:LEU:HB2	1:B:264:THR:HG21	1.90	0.52
1:B:168:ILE:H	1:B:168:ILE:HD12	1.74	0.51
1:A:159:ASN:OD1	1:A:160:PHE:N	2.44	0.51
1:A:414:GLN:HA	1:A:424:MSE:CE	2.34	0.50
1:A:237:ALA:HB3	1:A:238:PRO:HD3	1.92	0.50
1:A:239:SER:O	1:A:243:VAL:HG23	2.11	0.50
1:A:128:ILE:HG23	2:A:2002:HOH:O	2.12	0.49
1:B:219:VAL:HG21	1:B:267:ARG:HD3	1.93	0.49
1:B:275:VAL:HG13	1:B:275:VAL:O	2.13	0.48
1:A:295:ILE:HD11	1:A:355:LEU:O	2.13	0.48
1:B:270:ALA:HA	1:B:287:ILE:HD11	1.95	0.47
1:A:315:CYS:HB3	1:A:319:ILE:HD12	1.97	0.46
1:B:271:GLU:O	1:B:275:VAL:HB	2.16	0.46
1:B:141:ARG:HD3	1:B:274:ILE:HG21	1.98	0.45
1:B:242:GLN:HG3	2:B:2019:HOH:O	2.17	0.44
1:B:251:ARG:HD2	2:B:2021:HOH:O	2.18	0.44
1:B:384:ILE:HG21	1:B:391:LEU:HD23	1.99	0.44
1:A:424:MSE:HE2	1:A:424:MSE:HB2	1.74	0.44
1:B:302:HIS:O	1:B:305:VAL:HG22	2.18	0.43
1:A:384:ILE:HD12	1:A:395:ARG:HH12	1.84	0.43
1:B:399:TYR:O	1:B:403:ALA:N	2.50	0.42
1:B:144:LEU:HB3	1:B:275:VAL:HG22	2.01	0.42
1:A:259:ILE:HG23	1:A:307:VAL:CG2	2.49	0.42
1:A:266:LEU:HD23	1:A:311:PHE:HB2	2.02	0.42
1:A:350:GLN:HB3	1:A:351:PRO:HD3	2.02	0.42
1:A:384:ILE:HG22	1:A:384:ILE:O	2.19	0.42
1:B:182:THR:CG2	1:B:184:GLN:HG2	2.50	0.41
1:A:270:ALA:HA	1:A:287:ILE:HD11	2.02	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:302:HIS:O	1:A:305:VAL:HG23	2.21	0.41

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.

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	306/370 (83%)	290 (95%)	12 (4%)	4 (1%)	12	24
1	B	307/370 (83%)	294 (96%)	10 (3%)	3 (1%)	15	32
All	All	613/740 (83%)	584 (95%)	22 (4%)	7 (1%)	14	30

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	393	GLU
1	A	389	GLY
1	A	391	LEU
1	A	393	GLU
1	B	389	GLY
1	B	391	LEU
1	A	256	THR

5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	283/329 (86%)	273 (96%)	10 (4%)	36	62
1	B	284/329 (86%)	268 (94%)	16 (6%)	21	42
All	All	567/658 (86%)	541 (95%)	26 (5%)	27	51

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	157	GLU
1	A	165	ASP
1	A	182	THR
1	A	184	GLN
1	A	242	GLN
1	A	268	PHE
1	A	275	VAL
1	A	291	LEU
1	A	364	THR
1	A	386	HIS
1	B	121	MSE
1	B	157	GLU
1	B	165	ASP
1	B	176	LYS
1	B	177	LYS
1	B	182	THR
1	B	184	GLN
1	B	215	LYS
1	B	255	LYS
1	B	268	PHE
1	B	275	VAL
1	B	291	LEU
1	B	300	GLU
1	B	305	VAL
1	B	364	THR
1	B	402	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	305/370 (82%)	0.28	25 (8%)	11	8	31, 49, 116, 148	0
1	B	306/370 (82%)	0.17	23 (7%)	14	10	27, 40, 109, 141	0
All	All	611/740 (82%)	0.23	48 (7%)	12	9	27, 44, 113, 148	0

All (48) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	390	GLU	8.0
1	B	391	LEU	6.9
1	A	391	LEU	5.9
1	B	392	SER	5.7
1	B	388	LYS	5.5
1	A	382	ARG	5.5
1	B	385	LEU	5.2
1	B	389	GLY	5.2
1	B	381	ASN	5.2
1	A	388	LYS	5.1
1	A	386	HIS	5.0
1	B	383	ARG	4.9
1	B	382	ARG	4.7
1	B	386	HIS	4.6
1	A	389	GLY	4.6
1	A	379	ARG	4.6
1	B	395	ARG	4.4
1	A	392	SER	4.4
1	A	387	SER	4.2
1	A	383	ARG	4.1
1	A	399	TYR	4.1
1	A	396	HIS	4.0
1	B	380	GLN	4.0
1	A	384	ILE	3.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	384	ILE	3.7
1	B	390	GLU	3.5
1	B	393	GLU	3.5
1	A	377	THR	3.5
1	B	396	HIS	3.4
1	A	375	GLN	3.4
1	A	394	ASP	3.3
1	A	397	LYS	3.1
1	B	394	ASP	3.0
1	B	397	LYS	2.9
1	A	374	LEU	2.9
1	A	385	LEU	2.9
1	B	379	ARG	2.8
1	B	398	GLN	2.6
1	B	387	SER	2.6
1	A	402	PHE	2.5
1	B	377	THR	2.4
1	A	371	HIS	2.3
1	A	395	ARG	2.3
1	B	400	GLU	2.3
1	A	403	ALA	2.2
1	A	398	GLN	2.2
1	A	378	GLU	2.1
1	B	399	TYR	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.