



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 31, 2017 – 01:42 AM EDT

PDB ID : 3Q2T
Title : Crystal Structure of CFIm68 RRM/CFIm25/RNA complex
Authors : Yang, Q.; Coseno, M.; Gilmartin, G.M.; Doublié, S.
Deposited on : unknown
Resolution : 3.06 Å(reported)

This is a wwPDB X-ray 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/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

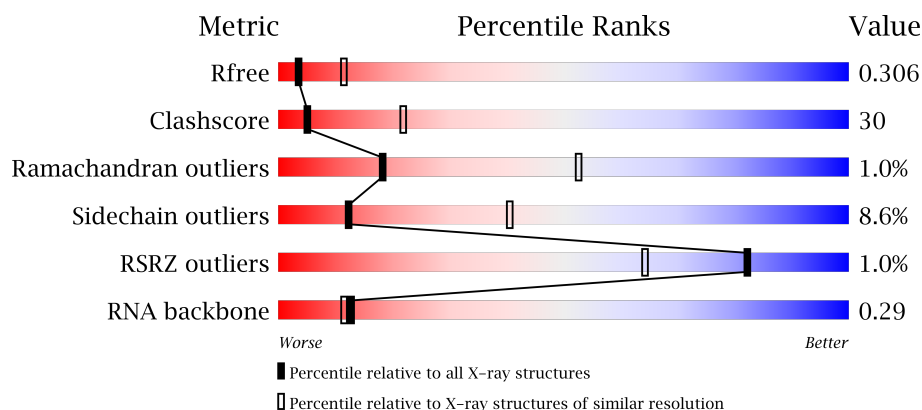
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.06 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}	100719	1348 (3.10-3.02)
Clashscore	112137	1462 (3.10-3.02)
Ramachandran outliers	110173	1410 (3.10-3.02)
Sidechain outliers	110143	1410 (3.10-3.02)
RSRZ outliers	101464	1355 (3.10-3.02)
RNA backbone	2435	1037 (3.42-2.70)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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 $\leq 5\%$. 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	207	<div> <div>53%</div> <div>43%</div> <div>..</div> </div>
1	B	207	<div> <div>50%</div> <div>44%</div> <div>..</div> </div>
2	C	229	<div> <div>21%</div> <div>17%</div> <div>.</div> <div>59%</div> </div>
2	D	229	<div> <div>17%</div> <div>20%</div> <div>.</div> <div>60%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	5	<p>20% 40% 40% 20%</p>
3	F	5	<p>60% 60% 40%</p>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 5035 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 Cleavage and polyadenylation specificity factor subunit 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	205	Total	C	N	O	S	4	2	0
			1699	1106	286	303	4			
1	B	203	Total	C	N	O	S	8	0	0
			1672	1089	282	297	4			

- Molecule 2 is a protein called Cleavage and polyadenylation specificity factor subunit 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	93	Total	C	N	O	S	0	0	0
			731	465	125	139	2			
2	D	92	Total	C	N	O	S	0	0	0
			726	462	124	138	2			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	159	VAL	CYS	ENGINEERED MUTATION	UNP Q16630
C	236	HIS	-	EXPRESSION TAG	UNP Q16630
C	237	HIS	-	EXPRESSION TAG	UNP Q16630
C	238	HIS	-	EXPRESSION TAG	UNP Q16630
C	239	HIS	-	EXPRESSION TAG	UNP Q16630
C	240	HIS	-	EXPRESSION TAG	UNP Q16630
C	241	HIS	-	EXPRESSION TAG	UNP Q16630
D	159	VAL	CYS	ENGINEERED MUTATION	UNP Q16630
D	236	HIS	-	EXPRESSION TAG	UNP Q16630
D	237	HIS	-	EXPRESSION TAG	UNP Q16630
D	238	HIS	-	EXPRESSION TAG	UNP Q16630
D	239	HIS	-	EXPRESSION TAG	UNP Q16630
D	240	HIS	-	EXPRESSION TAG	UNP Q16630
D	241	HIS	-	EXPRESSION TAG	UNP Q16630

- Molecule 3 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	5	Total	C	N	O	P	10	0	0
			102	47	16	35	4			
3	F	5	Total	C	N	O	P	8	0	0
			102	47	16	35	4			

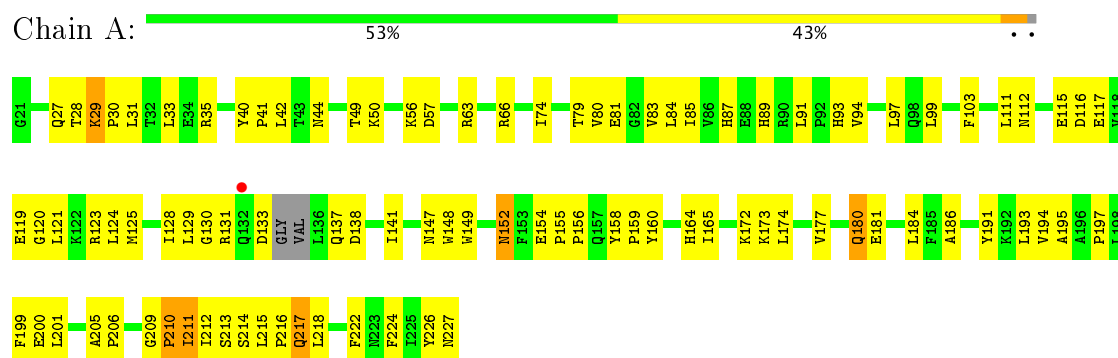
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	O	0	0
			2	2		
4	D	1	Total	O	0	0
			1	1		

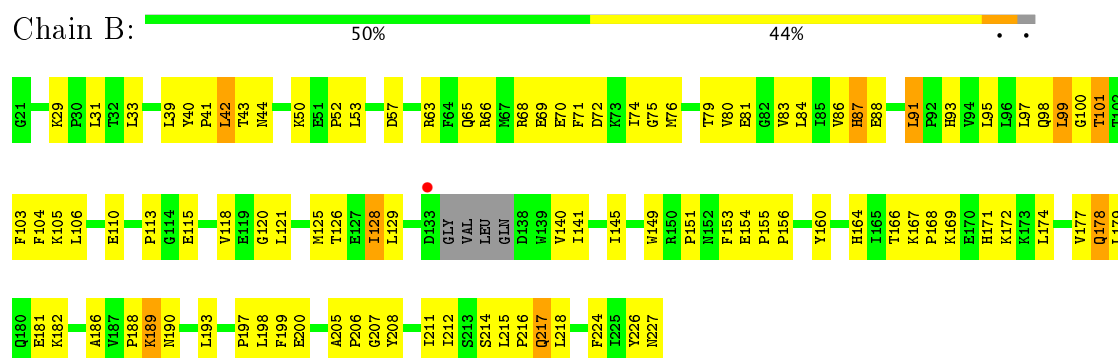
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

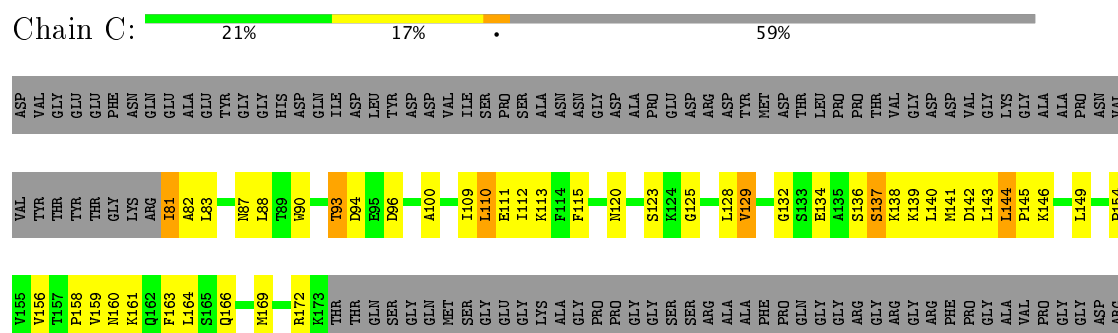
- Molecule 1: Cleavage and polyadenylation specificity factor subunit 5



- Molecule 1: Cleavage and polyadenylation specificity factor subunit 5



- Molecule 2: Cleavage and polyadenylation specificity factor subunit 6



PHE
PRO
GLY
PRO
ALA
GLY
PRO
GLY
GLY
PRO
PRO
ALA
GLY
PHE
PRO
ALA
ALA
GLY
GLN
THR
HIS
HIS
HIS
HIS
HIS
HIS

- Molecule 2: Cleavage and polyadenylation specificity factor subunit 6



ASP
VAL
GLY
GLU
GLU
PHE
ASN
GLN
ALA
GLU
TYR
GLY
HIS
ASP
GLN
ILE
ASP
LEU
TYR
ASP
ASP
VAL
ILE
SER
PRO
SER
ALA
ASN
ASN
GLY
ASP
ALA
PRO
GLU
ASP
ARG
ASP
TYR
MET
ASP
THR
LEU
PRO
PRO
THR
VAL
GLY
ASP
ASP
VAL
GLY
LYS
GLY
ALA
ALA
PRO
ASN
VAL

VAL
TYR
THR
THR
THR
GLY
LYS
ARG
I81
I82
L83
Y84
I85
G86
N87
L88
T92
T93
D94
E95
D96
L97
T98
E99
A100
V101
H102
S103
L104
G105
V106
M107
D108
I109
L110
E111
I112
K113
F114
F115
E116
N117
R118
S123
K124
G125
F126
A127
L128
V129
S133
E134
A135
S136
S137
K138

M141
D142
L143
L144
P145
K146
L149
Q152
N153
P154
V155
V156
T157
P158
V159
N160
F163
L164
F167
Q170
S171
R172
LYS
THR
THR
GLN
SER
SER
GLY
GLN
MET
SER
GLY
GLY
GLY
LYS
ALA
GLY
PRO
PRO
GLY
GLY
SER
SER
SER
ARG
ALA
ALA
PHE
PRO
GLN
GLY
GLY
ARG
GLY
ARG
GLY

ARG
PHE
PRO
GLY
ALA
VAL
PRO
GLY
GLY
ASP
ARG
PHE
PRO
GLY
PRO
ALA
GLY
PRO
GLY
GLY
GLY
PRO
PHE
PRO
ALA
GLN
THR
HIS
HIS
HIS
HIS
HIS

- Molecule 3: RNA



U1
U2
G3
U4
A5

- Molecule 3: RNA



U1
U2
G3
U4
A5

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α , β , γ	138.41Å 138.41Å 138.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 3.06 19.98 – 3.06	Depositor EDS
% Data completeness (in resolution range)	84.7 (19.98-3.06) 84.2 (19.98-3.06)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.72 (at 3.04Å)	Xtriage
Refinement program	PHENIX 1.6 _289	Depositor
R, R_{free}	0.211 , 0.286 0.222 , 0.306	Depositor DCC
R_{free} test set	736 reflections (5.04%)	DCC
Wilson B-factor (Å ²)	69.6	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.24 , 50.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.034 for l,-k,h	Xtriage
Reported twinning fraction	0.033 for l,-k,h	Depositor
Outliers	0 of 16899 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5035	wwPDB-VP
Average B, all atoms (Å ²)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.36% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.35	0/1752	0.51	0/2376
1	B	0.35	0/1719	0.53	0/2331
2	C	0.31	0/745	0.49	0/1007
2	D	0.30	0/740	0.47	0/1000
3	E	0.30	0/113	0.80	0/174
3	F	0.33	0/113	0.84	1/174 (0.6%)
All	All	0.34	0/5182	0.53	1/7062 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	2	U	O4'-C1'-N1	5.06	112.25	108.20

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	1699	0	1706	89	0
1	B	1672	0	1677	119	0
2	C	731	0	723	48	0
2	D	726	0	721	47	0
3	E	102	0	54	5	0
3	F	102	0	54	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	2	0	0	1	0
4	D	1	0	0	0	0
All	All	5035	0	4935	294	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 30.

The worst 5 of 294 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:B:217:GLN:HE21	1:B:217:GLN:N	1.51	1.07
1:B:217:GLN:NE2	1:B:217:GLN:H	1.50	1.07
1:A:83:VAL:HG12	1:A:85:ILE:HD11	1.49	0.94
1:A:217:GLN:HE21	1:A:217:GLN:H	0.94	0.92
1:B:217:GLN:HE21	1:B:217:GLN:H	0.93	0.92

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	203/207 (98%)	185 (91%)	16 (8%)	2 (1%)	18	52
1	B	199/207 (96%)	169 (85%)	28 (14%)	2 (1%)	18	52
2	C	91/229 (40%)	75 (82%)	16 (18%)	0	100	100
2	D	90/229 (39%)	78 (87%)	10 (11%)	2 (2%)	8	31
All	All	583/872 (67%)	507 (87%)	70 (12%)	6 (1%)	18	52

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	211	ILE
2	D	154	PRO
1	B	101	THR
1	A	210	PRO
1	B	211	ILE

5.3.2 Protein sidechains ⓘ

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	186/185 (100%)	174 (94%)	12 (6%)	20	52
1	B	182/185 (98%)	171 (94%)	11 (6%)	22	55
2	C	80/180 (44%)	69 (86%)	11 (14%)	4	17
2	D	80/180 (44%)	68 (85%)	12 (15%)	3	13
All	All	528/730 (72%)	482 (91%)	46 (9%)	12	40

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	178	GLN
2	C	93	THR
2	D	113	LYS
1	B	189	LYS
2	C	81	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	217	GLN
1	A	223	ASN
2	C	166	GLN
1	A	180	GLN
2	C	87	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	E	4/5 (80%)	2 (50%)	0
3	F	4/5 (80%)	0	0
All	All	8/10 (80%)	2 (25%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	E	3	G
3	E	5	A

There are no RNA pucker outliers to report.

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 carbohydrates 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	205/207 (99%)	-0.69	1 (0%) 90 78	35, 54, 85, 100	13 (6%)
1	B	203/207 (98%)	-0.64	1 (0%) 90 78	33, 56, 92, 115	11 (5%)
2	C	93/229 (40%)	-0.55	0 100 100	44, 68, 92, 104	6 (6%)
2	D	92/229 (40%)	-0.49	0 100 100	54, 75, 92, 98	9 (9%)
3	E	5/5 (100%)	1.20	1 (20%) 1 0	83, 100, 113, 114	2 (40%)
3	F	5/5 (100%)	2.20	3 (60%) 0 0	91, 111, 127, 142	2 (40%)
All	All	603/882 (68%)	-0.58	6 (0%) 82 64	33, 61, 92, 142	43 (7%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	3	G	3.0
3	F	5	A	2.8
3	E	5	A	2.4
1	B	133	ASP	2.3
1	A	132	GLN	2.2

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 carbohydrates 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.