wwPDB EM Map/Model Validation Report (i)

Oct 3, 2016 – 08:54 PM EDT

PDB ID	:	5GAM
EMDB ID:	:	EMD-8011
Title	:	Foot region of the yeast spliceosomal U4/U6.U5 tri-snRNP
Authors	:	Nguyen, T.H.D.; Galej, W.P.; Bai, X.C.; Oubridge, C.; Scheres, S.H.W.; Newman, A.J.; Nagai, K.
Deposited on	:	2015-12-15
Resolution	:	3.70 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.

For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.

We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
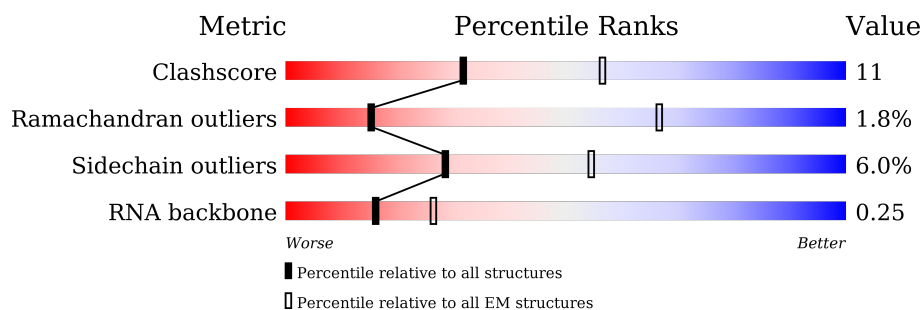http://wwpdb.org/validation/2016/EMValidationReportHelp

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.70 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	Percentile Ranks	Value
Clashscore		11
Ramachandran outliers		1.8%
Sidechain outliers		6.0%
RNA backbone		0.25

Worse ... *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all EM structures

Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686
RNA backbone	3027	244

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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 <=5%

Mol	Chain	Length	Quality of chain
1	U	178	24% 30% 24% • 21%
2	A	735	62% 17% • 19%
3	C	1008	57% 25% • 15%
4	W	31	35% 32% 13% 19%
5	b	196	41% 59%
6	e	94	74% 5% 20%
7	f	86	80% • 16%
8	g	77	87% •• 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	x	18	100%
10	d	101	78% · 19%
11	h	146	52% · 44%
12	j	110	81% 5% 15%

2 Entry composition (i)

There are 13 unique types of molecules in this entry. The entry contains 19123 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 RNA chain called U5 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	U	141	Total	C	N	O	P	0	0
			2999	1342	530	986	141		

- Molecule 2 is a protein called Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	A	593	Total	C	N	O	S	0	0
			4702	3052	817	817	16		

- Molecule 3 is a protein called Pre-mRNA-splicing factor SNU114.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	855	Total	C	N	O	S	0	0
			6450	4195	1089	1143	23		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	530	SER	GLU	conflict	UNP P36048
C	531	LYS	ASP	conflict	UNP P36048
C	532	THR	ASP	conflict	UNP P36048

- Molecule 4 is a RNA chain called U6 snRNA, 5' end.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	W	25	Total	C	N	O	P	0	0
			532	237	91	179	25		

- Molecule 5 is a protein called Small nuclear ribonucleoprotein-associated protein B.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	b	80	Total	C	N	O	S	0	0
			631	403	114	111	3		

- Molecule 6 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	e	75	575	379	92	101	3	0	0

- Molecule 7 is a protein called Small nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	f	72	573	368	101	103	1	0	0

- Molecule 8 is a protein called Small nuclear ribonucleoprotein G.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	g	69	529	337	93	97	2	0	0

- Molecule 9 is a protein called Unknown polypeptide.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
9	x	18	90	54	18	18	0	0

- Molecule 10 is a protein called Small nuclear ribonucleoprotein Sm D3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	d	82	625	399	109	115	2	0	0

- Molecule 11 is a protein called Small nuclear ribonucleoprotein Sm D1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	h	82	644	409	110	123	2	0	0

- Molecule 12 is a protein called Small nuclear ribonucleoprotein Sm D2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
12	j	94	741	477	141	119	4	0	0

- Molecule 13 is GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GTP) (formula:

$C_{10}H_{16}N_5O_{14}P_3$).

Mol	Chain	Residues	Atoms					AltConf
			Total	C	N	O	P	
13	C	1	32	10	5	14	3	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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: U5 snRNA

Chain U:

- Molecule 2: Pre-mRNA-splicing factor 8

Chain A:

- Molecule 3: Pre-mRNA-splicing factor SNU114

Chain C:

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

GLU LEU GLY ILE SER LEU GLU GLU HIS PRO TYR GLY LYS GLU VAL VAL GLU VAL LEU MET GLU GLU LYS ASN THR GLN SER PRO GLN THR PRO LEU VAL GLU PRO VAL THR GLU ARG THR LYS LEU GLN **E102** **Q108** **K111** **N112** **I113** **L124** **I129** **P130** **I133** **V138** **L142** **H143**

K146 **L149** **L152** **L153** **V154** **S157** **R160** **K173** **R176** **Y177** **I178** **D179** **N180** **Q183** **R187** **G188** **L189** **S190** **I191** **K192** **L193** **N194** **G195** **S196** **T197** **L198** **L199** **C200** **S207** **R208** **M209** **N211** **I210** **F212** **L213** **D214** **V219** **N220** **F221** **M222** **D223** **L234** **V235** **L236** **I239** **D240** **V241**

V242 **V245** **Q251** **Q255** **N260** **V261** **K269** **L270** **D271** **R272** **L273** **L277** **P281** **Y285** **L288** **I295** **P306** **N309** **F313** **A314** **S315** **L318** **G319** **F320** **T321** **F322** **T323** **I324** **F327** **I336** **D342** **T345** **T346** **R347** **L348** **W349** **K356** **F359** **R360** **T361**

V367 **E368** **K369** **Y370** **P371** **V374** **E375** **F376** **I377** **L378** **I379** **P380** **D381** **Y382** **F385** **S386** **Y387** **L396** **L399** **L400** **N403** **F404** **R405** **V406** **N407** **L408** **A412** **Y415** **D416** **P417** **L421** **K422** **H423** **V424** **L425** **Q426** **L427** **I428** **Q431** **Q432** **L435** **Y436** **D437** **A438** **I439** **T440** **Q444** **P445**

F446 **S457** **I458** **P464** **L468** **W469** **A470** **H471** **V472** **T475** **V476** **D477** **Y478** **S484** **L485** **V486** **L492** **L493** **T498** **V499** **R500** **I501** **L502** **I518** SER LYS THR GLU THR SER ASN GLU ASP GLU ASP **S530** **T534** **P535** **S536** **C537** **E538** **V539** **I542** **L545** **G546** **V553** **Q560**

I564 **I567** **S568** **Y571** **I572** **K573** **S574** **A575** **T576** **L577** **Y578** **S582** **K583** **K590** **F591** **F592** **D596** **Y597** **I598** **T599** **V602** **I605** **V606** **L607** **L615** **P616** **K617** **I625** **S626** **K627** **Y628** **Y629** **P630** **G631** **V632** **I633** **I634** **Y651** **M652** **D653** **C654** **L655** **L656** **Y657** **D658** **I659** **R660** **I666**

E667 **I668** **K669** **I670** **L674** **T675** **S678** **E679** **S680** **S688** **N693** SER ILE SER ARG LEU GLY GLU GLU ASN LEU PRO GLY LEU **S707** **V710** **A711** **A712** **L723** **S724** **R725** ASN THR LEU GLY LYS GLY GLN ASN CYS LEU ILE ASP GLY ILE MET ASP ASN **P744** **L760** **A761** **S762**

V765 **W766** **S767** **F768** **Y769** **N770** **G771** **N772** **V773** **L774** **I775** **P780** **I783** **S784** **Y791** **Q797** **Y800** **V803** **E811** **P812** **I813** **V816** **Q817** **Y818** **K819** **L820** **L821** **S836** **P840** **L851** **I856** **L857** **I861** **Y862** **E863** **V864** **D865** **I866** **T867** **V868** **H869** **L872** **L873** **P874** **I875**

V876 **L879** **R883** **R884** **G885** **S886** **R887** **I888** **L900** **E901** **V902** **V906** **P907** **V908** **I909** **E910** **S911** **A912** **G913** **F914** **L918** **S921** **T922** **L925** **G926** **M927** **C928** **Q929** **L930** **W933** **H934** **K935** **K939** **V944** **I951** **A957** **P958** **S961** **L962** **F966** **V967** **R971** **T988** **L989**

Y998 ALA GLN LEU ARG GLU ASN GLY LEU VAL PRO

● Molecule 4: U6 snRNA, 5' end

Chain W: [35%]   [32%]   [13%]   [19%]

G1 **U2** **G5** **C6** **G7** **A8** **A9** G U A A C C **C16** **U17** **C25** **A26** **U27** **U28** **U29** **G30** **G31**

● Molecule 5: Small nuclear ribonucleoprotein-associated protein B

Chain b: [41%]   [59%]

MET SER LYS **I4** **K55** THR GLN LEU ASP LYS LEU ARG PRO ARG LYS ASP SER LYS ASP GLY THR THR LEU THR LEU ASN **I75** **L102** LEU SER LYS LYS GLU LEU ARG LEU VAL ARG ASP LYS LYS GLU LYS GLN GLN ALA GLN LYS LYS THR LYS LEU ARG LYS GLU GLU LYS LYS PRO GLY

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

ARG LYS

• Molecule 6: Small nuclear ribonucleoprotein E

Chain e:
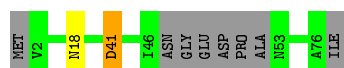

• Molecule 7: Small nuclear ribonucleoprotein F

Chain f:

• Molecule 8: Small nuclear ribonucleoprotein G

Chain g:
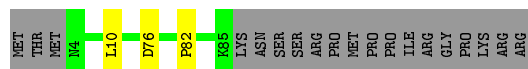

• Molecule 9: Unknown polypeptide

Chain x:
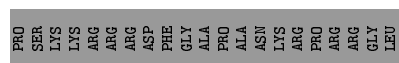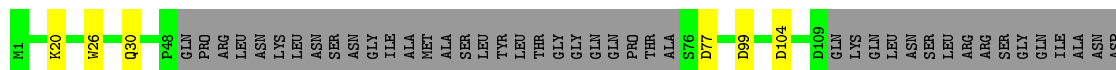

There are no outlier residues recorded for this chain.

• Molecule 10: Small nuclear ribonucleoprotein Sm D3

Chain d:

• Molecule 11: Small nuclear ribonucleoprotein Sm D1

Chain h:

• Molecule 12: Small nuclear ribonucleoprotein Sm D2

Chain j:

4 Experimental information ⓘ

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	140155	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/\text{Å}^2)$	Not provided	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	81000	Depositor
Image detector	Not provided	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GTP

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	#$|Z|$ >2	RMSZ	#$|Z|$ >2
1	U	0.38	0/3351	0.87	3/5213 (0.1%)
10	d	0.36	0/634	0.66	0/859
11	h	0.36	0/649	0.59	0/880
12	j	0.41	0/753	0.67	0/1013
2	A	0.42	0/4840	0.74	2/6596 (0.0%)
3	C	0.44	0/6590	0.78	1/8975 (0.0%)
4	W	0.26	0/592	0.73	0/918
5	b	0.36	0/636	0.64	0/856
6	e	0.39	0/585	0.60	0/795
7	f	0.43	0/585	0.62	0/791
8	g	0.35	0/532	0.62	0/715
All	All	0.41	0/19747	0.76	6/27611 (0.0%)

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 outliers	#Planarity outliers
2	A	0	5
3	C	0	4
8	g	0	1
All	All	0	10

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	U	82	A	C2'-C3'-O3'	7.10	125.12	109.50
1	U	128	A	C2'-C3'-O3'	6.70	124.43	113.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	U	62	G	C4'-C3'-O3'	5.50	124.01	113.00
2	A	425	ASP	CB-CG-OD2	5.26	123.03	118.30
3	C	659	LEU	CA-CB-CG	5.17	127.19	115.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	460	PRO	Peptide
2	A	464	GLU	Peptide
2	A	466	GLU	Peptide
2	A	553	ASN	Peptide
2	A	556	TYR	Peptide

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	U	2999	0	1515	103	0
2	A	4702	0	4527	109	0
3	C	6450	0	6420	186	0
4	W	532	0	269	4	0
5	b	631	0	670	0	0
6	e	575	0	597	0	0
7	f	573	0	572	0	0
8	g	529	0	557	0	0
9	x	90	0	20	0	0
10	d	625	0	647	0	0
11	h	644	0	686	0	0
12	j	741	0	778	0	0
13	C	32	0	12	1	0
All	All	19123	0	17270	377	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 11.

The worst 5 of 377 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)
2:A:139:LEU:HD13	2:A:193:TYR:CD2	1.31	1.63
2:A:139:LEU:CD1	2:A:193:TYR:CD2	2.15	1.29
1:U:42:A:H3'	1:U:43:G:H5"	1.24	1.15
2:A:139:LEU:HD13	2:A:193:TYR:CE2	1.87	1.08
2:A:139:LEU:CD1	2:A:193:TYR:HD2	1.60	1.07

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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 EM 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	
2	A	587/735 (80%)	511 (87%)	67 (11%)	9 (2%)	13	59
3	C	847/1008 (84%)	730 (86%)	93 (11%)	24 (3%)	6	47
5	b	76/196 (39%)	69 (91%)	7 (9%)	0	100	100
6	e	71/94 (76%)	66 (93%)	5 (7%)	0	100	100
7	f	70/86 (81%)	62 (89%)	6 (9%)	2 (3%)	6	46
8	g	65/77 (84%)	64 (98%)	1 (2%)	0	100	100
10	d	80/101 (79%)	72 (90%)	7 (9%)	1 (1%)	15	62
11	h	78/146 (53%)	74 (95%)	4 (5%)	0	100	100
12	j	92/110 (84%)	86 (94%)	6 (6%)	0	100	100
All	All	1966/2553 (77%)	1734 (88%)	196 (10%)	36 (2%)	15	56

5 of 36 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	247	PRO
2	A	615	LEU
3	C	269	LYS
3	C	356	LYS
3	C	367	VAL

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 PDB entries followed by that with respect to all EM 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	
2	A	482/677 (71%)	460 (95%)	22 (5%)	33	73
3	C	673/910 (74%)	618 (92%)	55 (8%)	14	53
5	b	70/176 (40%)	70 (100%)	0	100	100
6	e	65/83 (78%)	60 (92%)	5 (8%)	16	56
7	f	63/77 (82%)	62 (98%)	1 (2%)	70	90
8	g	58/66 (88%)	56 (97%)	2 (3%)	44	79
10	d	69/89 (78%)	67 (97%)	2 (3%)	50	81
11	h	77/129 (60%)	71 (92%)	6 (8%)	16	56
12	j	79/103 (77%)	74 (94%)	5 (6%)	22	64
All	All	1636/2310 (71%)	1538 (94%)	98 (6%)	28	66

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	468	LEU
3	C	658	ASP
11	h	77	ASP
3	C	469	TRP
3	C	534	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
2	A	576	HIS
3	C	112	ASN
8	g	66	ASN
2	A	617	ASN
2	A	658	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	U	139/178 (78%)	81 (58%)	23 (16%)
4	W	24/31 (77%)	11 (45%)	4 (16%)
All	All	163/209 (77%)	92 (56%)	27 (16%)

5 of 92 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	U	5	A
1	U	7	C
1	U	8	U
1	U	9	U
1	U	10	U

5 of 27 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	U	78	A
1	U	95	C
4	W	25	C
1	U	82	A
1	U	26	A

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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
13	GTP	C	1101	-	26,34,34	0.99	2 (7%)	29,54,54	1.79	7 (24%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
13	GTP	C	1101	-	-	0/18/38/38	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
13	C	1101	GTP	C5-C4	2.70	1.46	1.40
13	C	1101	GTP	C6-C5	2.93	1.47	1.41

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
13	C	1101	GTP	N3-C2-N1	-3.85	122.31	127.56
13	C	1101	GTP	C5-C6-N1	-3.36	119.14	123.52
13	C	1101	GTP	C6-C5-C4	-3.18	117.22	120.86
13	C	1101	GTP	C1'-N9-C4	-2.42	124.11	126.81
13	C	1101	GTP	O3G-PG-O2G	2.23	115.63	107.44

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
13	C	1101	GTP	1	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.