



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 10:09 pm GMT

PDB ID : 5FSG  
Title : Structure of the hantavirus nucleoprotein provides insights into the mechanism of RNA encapsidation and a template for drug design  
Authors : Olal, D.; Daumke, O.  
Deposited on : 2016-01-05  
Resolution : 3.21 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](mailto: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  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
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) : recalc28949

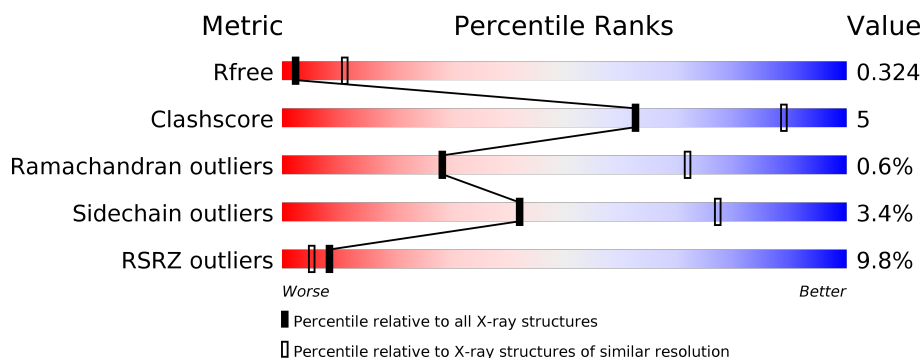
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.21 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	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

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  $\geq 3$ , 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	698	<div> <div>9%</div> <div>82%</div> <div>12%</div> <div>5%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10176 atoms, of which 5028 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.

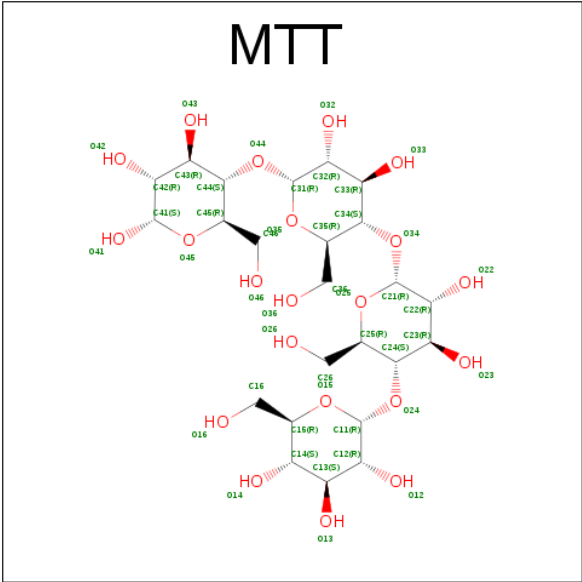
- Molecule 1 is a protein called MALTOSE-BINDING PERIPLASMIC PROTEIN, HANTAVIRUS NUCLEOPROTEIN.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
1	A	661	Total	C	H	N	O	S	Se		0	0	0
			10089	3269	4986	854	958	5	17				

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	ALA	-	LINKER	UNP P05133
A	-12	ALA	-	LINKER	UNP P05133
A	-11	LEU	-	LINKER	UNP P05133
A	-10	ALA	-	LINKER	UNP P05133
A	-9	ALA	-	LINKER	UNP P05133
A	-8	GLY	-	LINKER	UNP P05133
A	-7	SER	-	LINKER	UNP P05133
A	-6	ALA	-	LINKER	UNP P05133
A	-5	GLN	-	LINKER	UNP P05133
A	-4	THR	-	LINKER	UNP P05133
A	-3	ASN	-	LINKER	UNP P05133
A	-2	ALA	-	LINKER	UNP P05133
A	-1	ALA	-	LINKER	UNP P05133
A	0	ALA	-	LINKER	UNP P05133
A	430	LEU	-	EXPRESSION TAG	UNP P05133
A	431	GLU	-	EXPRESSION TAG	UNP P05133
A	432	HIS	-	EXPRESSION TAG	UNP P05133
A	433	HIS	-	EXPRESSION TAG	UNP P05133
A	434	HIS	-	EXPRESSION TAG	UNP P05133
A	435	HIS	-	EXPRESSION TAG	UNP P05133
A	436	HIS	-	EXPRESSION TAG	UNP P05133
A	437	HIS	-	EXPRESSION TAG	UNP P05133
A	185	ASN	GLN	CONFLICT	UNP P05133
A	377	ASP	GLY	CONFLICT	UNP P05133

- Molecule 2 is MALTOTETRAOSE (three-letter code: MTT) (formula:  $C_{24}H_{42}O_{21}$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			87	24	42	21		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	69.63Å 86.84Å 131.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.80 – 3.21 47.81 – 3.21	Depositor EDS
% Data completeness (in resolution range)	99.6 (47.80-3.21) 99.8 (47.81-3.21)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 3.19Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.290 , 0.299 0.306 , 0.324	Depositor DCC
$R_{free}$ test set	679 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	82.5	Xtriage
Anisotropy	0.642	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 55.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	10176	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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](#)

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

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	0.23	0/5197	0.43	0/7021

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	5103	4986	5082	51	0
2	A	45	42	42	1	0
All	All	5148	5028	5124	51	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 5.

All (51) 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:-37:GLN:N	1:A:-37:GLN:OE1	2.22	0.72
1:A:139:TYR:O	1:A:143:THR:HG23	1.90	0.71

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:-182:ALA:O	1:A:-179:THR:OG1	2.08	0.70
1:A:-368:GLU:HG2	1:A:-368:GLU:O	1.92	0.69
1:A:327:GLU:OE2	1:A:377:ASP:N	2.27	0.68
1:A:145:GLY:O	1:A:146:ARG:HB2	1.96	0.65
1:A:261:THR:OG1	1:A:264:ASP:OD2	2.12	0.64
1:A:263:ARG:NH2	1:A:298:ASP:OD2	2.31	0.62
1:A:298:ASP:OD1	1:A:299:ILE:N	2.34	0.60
1:A:-328:GLU:OE1	2:A:1431:MTT:O23	2.11	0.59
1:A:423:ASN:O	1:A:424:GLN:NE2	2.38	0.57
1:A:314:ARG:HG2	1:A:368:ARG:HB3	1.89	0.55
1:A:-3:ASN:HA	1:A:0:ALA:HB3	1.89	0.54
1:A:327:GLU:OE1	1:A:327:GLU:N	2.41	0.53
1:A:385:MSE:O	1:A:389:GLY:N	2.42	0.52
1:A:370:GLN:N	1:A:370:GLN:OE1	2.44	0.51
1:A:211:LYS:O	1:A:214:GLN:NE2	2.44	0.50
1:A:415:ASP:O	1:A:419:LYS:N	2.42	0.50
1:A:-7:SER:O	1:A:-3:ASN:ND2	2.41	0.48
1:A:-319:THR:O	1:A:-319:THR:HG22	2.14	0.47
1:A:-344:GLU:OE1	1:A:-339:ILE:N	2.48	0.47
1:A:231:ASP:OD2	1:A:235:ARG:NH2	2.48	0.47
1:A:-179:THR:O	1:A:-176:VAL:HG22	2.15	0.46
1:A:235:ARG:NH1	1:A:391:GLU:OE1	2.48	0.46
1:A:-345:PHE:O	1:A:-341:THR:HG22	2.14	0.46
1:A:-232:LYS:NZ	1:A:-170:LYS:O	2.47	0.46
1:A:-221:LEU:HD12	1:A:-164:THR:OG1	2.16	0.46
1:A:-290:ASP:N	1:A:-290:ASP:OD1	2.49	0.46
1:A:-160:ILE:HG13	1:A:-159:ALA:N	2.31	0.45
1:A:-366:LYS:HZ1	1:A:-336:THR:CB	2.30	0.45
1:A:-295:ALA:HB3	1:A:-106:ILE:HG23	1.98	0.45
1:A:-240:ILE:N	1:A:-239:PRO:HD2	2.31	0.45
1:A:258:GLY:N	1:A:259:PRO:CD	2.80	0.44
1:A:307:VAL:CG1	1:A:316:PRO:HG3	2.47	0.44
1:A:328:LEU:O	1:A:331:PHE:HB3	2.18	0.44
1:A:-369:GLU:O	1:A:-367:GLY:N	2.50	0.43
1:A:-122:PHE:O	1:A:-121:LYS:HG2	2.18	0.43
1:A:343:MSE:HE1	1:A:398:LEU:O	2.19	0.42
1:A:422:SER:O	1:A:424:GLN:N	2.52	0.42
1:A:-122:PHE:O	1:A:-121:LYS:CG	2.67	0.42
1:A:-214:TRP:N	1:A:-213:PRO:HD2	2.34	0.42
1:A:-327:GLU:O	1:A:-324:PRO:HD2	2.19	0.42
1:A:-164:THR:HA	1:A:-160:ILE:HD11	2.01	0.41

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:-307:ASP:OD1	1:A:-307:ASP:O	2.38	0.41
1:A:-307:ASP:OD1	1:A:-36:MSE:HB2	2.21	0.41
1:A:-307:ASP:OD1	1:A:-36:MSE:CB	2.68	0.41
1:A:137:ALA:O	1:A:141:LEU:HD13	2.21	0.41
1:A:-142:TRP:O	1:A:-75:LYS:NZ	2.48	0.41
1:A:-314:ASP:O	1:A:-106:ILE:HD12	2.21	0.41
1:A:295:MSE:HE3	1:A:318:THR:O	2.20	0.41
1:A:-127:THR:HG22	1:A:-126:VAL:H	1.86	0.40

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	651/698 (93%)	628 (96%)	19 (3%)	4 (1%)	28 72

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	-368	GLU
1	A	377	ASP
1	A	423	ASN
1	A	207	PRO

### 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	530/556 (95%)	512 (97%)	18 (3%)	42 77

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	-370	ILE
1	A	-363	ILE
1	A	-328	GLU
1	A	-307	ASP
1	A	-290	ASP
1	A	-114	PHE
1	A	-42	MSE
1	A	-14	ASP
1	A	140	MSE
1	A	226	LEU
1	A	239	TRP
1	A	263	ARG
1	A	265	TYR
1	A	280	GLU
1	A	327	GLU
1	A	328	LEU
1	A	385	MSE
1	A	429	LEU

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

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

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$
2	MTT	A	1431	-	48,48,48	2.15	16 (33%)	71,71,71	1.01	2 (2%)

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
2	MTT	A	1431	-	-	0/20/100/100	0/4/4/4

All (16) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1431	MTT	C26-C25	-3.42	1.40	1.51
2	A	1431	MTT	C36-C35	-3.11	1.41	1.51
2	A	1431	MTT	C16-C15	-2.91	1.41	1.51
2	A	1431	MTT	C46-C45	-2.86	1.42	1.51
2	A	1431	MTT	C32-C33	-2.08	1.47	1.52
2	A	1431	MTT	O32-C32	2.04	1.47	1.43
2	A	1431	MTT	O25-C21	2.14	1.47	1.41
2	A	1431	MTT	O35-C31	2.14	1.47	1.41
2	A	1431	MTT	O45-C41	2.44	1.47	1.43
2	A	1431	MTT	O33-C33	2.61	1.49	1.43
2	A	1431	MTT	O42-C42	2.77	1.49	1.43
2	A	1431	MTT	O15-C11	2.88	1.49	1.41
2	A	1431	MTT	O45-C45	4.59	1.55	1.44
2	A	1431	MTT	O35-C35	4.78	1.56	1.44
2	A	1431	MTT	O25-C25	4.80	1.56	1.44
2	A	1431	MTT	O15-C15	4.81	1.56	1.44

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	A	1431	MTT	C11-O24-C24	-2.51	111.88	118.00
2	A	1431	MTT	C31-O44-C44	-2.35	112.26	118.00

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
2	A	1431	MTT	1	0

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	643/698 (92%)	0.85	63 (9%) 8 5	50, 80, 107, 129	0

All (63) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	254	SER	5.6
1	A	-90	ASN	4.8
1	A	-371	LYS	4.5
1	A	-199	ASN	4.3
1	A	-366	LYS	4.3
1	A	246	LEU	4.3
1	A	398	LEU	4.0
1	A	-178	PHE	3.9
1	A	248	PRO	3.8
1	A	-88	LEU	3.8
1	A	-91	GLU	3.6
1	A	287	HIS	3.5
1	A	247	LEU	3.5
1	A	-272	ASN	3.4
1	A	-284	LYS	3.2
1	A	-201	TYR	3.2
1	A	-276	ALA	3.1
1	A	-106	ILE	3.1
1	A	177	LEU	3.1
1	A	422	SER	3.0
1	A	-295	ALA	3.0
1	A	172	ARG	2.9
1	A	136	LYS	2.9
1	A	293	CYS	2.8
1	A	-317	ASP	2.8
1	A	-94	GLU	2.6
1	A	-300	GLN	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	399	GLY	2.6
1	A	-323	GLN	2.6
1	A	159	PHE	2.6
1	A	369	THR	2.6
1	A	244	CYS	2.5
1	A	174	PRO	2.5
1	A	239	TRP	2.5
1	A	-208	ASP	2.5
1	A	183	ASN	2.5
1	A	185	ASN	2.5
1	A	324	GLY	2.4
1	A	-160	ILE	2.4
1	A	179	VAL	2.4
1	A	-165	ASP	2.4
1	A	330	ALA	2.4
1	A	-294	GLU	2.3
1	A	-239	PRO	2.3
1	A	-234	GLU	2.3
1	A	284	ILE	2.3
1	A	-305	PHE	2.3
1	A	-282	TYR	2.2
1	A	-179	THR	2.2
1	A	250	THR	2.2
1	A	275	ASN	2.2
1	A	-369	GLU	2.2
1	A	-278	TRP	2.1
1	A	-184	ALA	2.1
1	A	-343	LYS	2.1
1	A	-221	LEU	2.1
1	A	296	ILE	2.1
1	A	326	ALA	2.1
1	A	362	TYR	2.1
1	A	155	THR	2.1
1	A	-66	SER	2.1
1	A	-158	GLU	2.0
1	A	-283	LEU	2.0

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

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	MTT	A	1431	45/45	0.91	0.29	-0.04	19,31,63,65	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.