



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2017 – 03:20 PM EDT

PDB ID : 3JSU  
Title : Quadruple mutant(N51I+C59R+S108N+I164L) plasmodium falciparum dihydrofolate reductase-thymidylate synthase(PFDHFR-TS) complexed with QN254, NADPH, and dUMP  
Authors : Chitnumsub, P.; Maneeruttanarungroj, C.; Kamchonwongpaisan, S.; Yuthavong, Y.; Diagana, T.T.  
Deposited on : unknown  
Resolution : 2.70 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](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	:	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

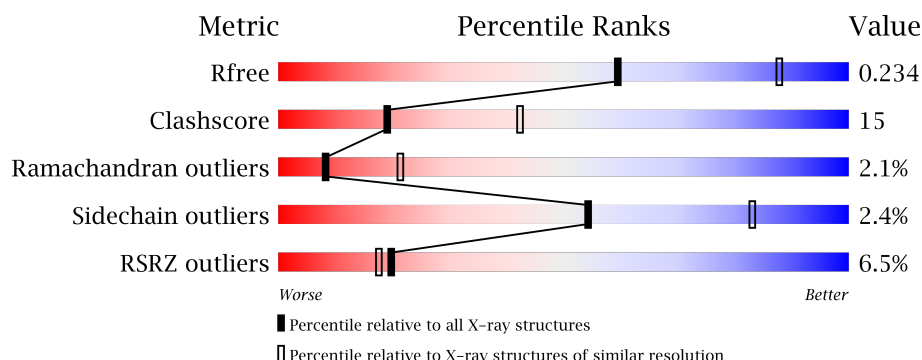
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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  $\geq 3$ , 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	608	<div> <div>5%</div> <div>62%</div> <div>26%</div> <div>10%</div> </div>
1	B	608	<div> <div>7%</div> <div>63%</div> <div>24%</div> <div>11%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	UMP	A	611	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 9487 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.

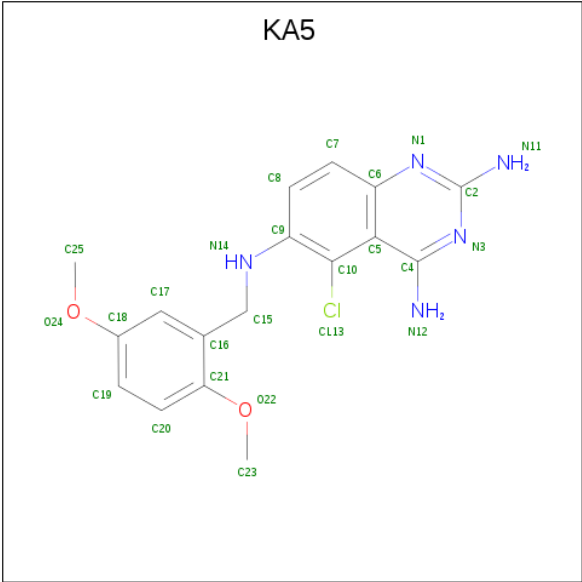
- Molecule 1 is a protein called Dihydrofolate reductase-thymidylate synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	545	Total	C	N	O	S	0	0	0
			4539	2931	751	830	27			
1	B	542	Total	C	N	O	S	0	0	0
			4511	2914	747	825	25			

There are 8 discrepancies between the modelled and reference sequences:

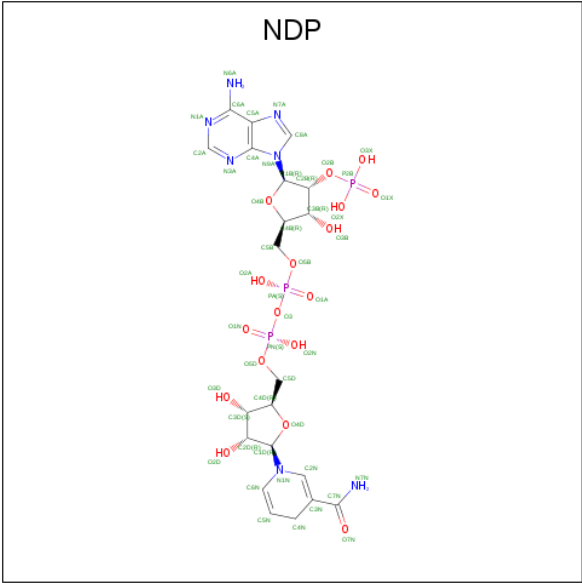
Chain	Residue	Modelled	Actual	Comment	Reference
A	51	ILE	ASN	ENGINEERED MUTATION	UNP A7UD79
A	59	ARG	TYR	ENGINEERED MUTATION	UNP A7UD79
A	108	ASN	SER	ENGINEERED MUTATION	UNP A7UD79
A	164	LEU	ILE	ENGINEERED MUTATION	UNP A7UD79
B	51	ILE	ASN	ENGINEERED MUTATION	UNP A7UD79
B	59	ARG	TYR	ENGINEERED MUTATION	UNP A7UD79
B	108	ASN	SER	ENGINEERED MUTATION	UNP A7UD79
B	164	LEU	ILE	ENGINEERED MUTATION	UNP A7UD79

- Molecule 2 is 5-chloro-N 6 -(2,5-dimethoxybenzyl)quinazoline-2,4,6-triamine (three-letter code: KA5) (formula: C<sub>17</sub>H<sub>18</sub>ClN<sub>5</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 25	C 17	Cl 1	N 5	O 2	0	0
2	B	1	Total 25	C 17	Cl 1	N 5	O 2	0	0

- Molecule 3 is NADPH DIHYDRO-NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NDP) (formula:  $C_{21}H_{30}N_7O_{17}P_3$ ).



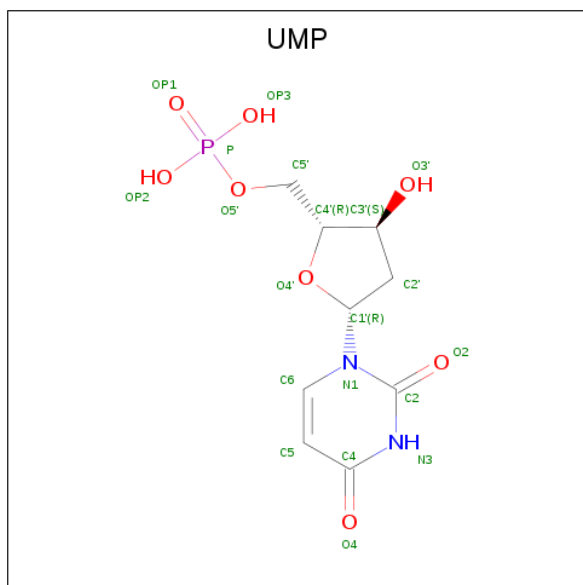
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			48	21	7	17	3		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	B	1	Total	C	N	O	P	0	0
			48	21	7	17	3		

- Molecule 4 is 2'-DEOXYURIDINE 5'-MONOPHOSPHATE (three-letter code: UMP) (formula:  $C_9H_{13}N_2O_8P$ ).

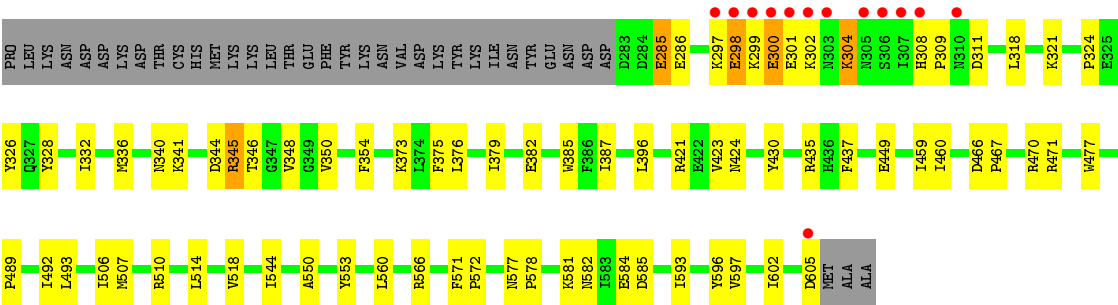


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	B	1	Total	C	N	O	P	0	0
			20	9	2	8	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	124	Total	O	0	0
			124	124		
5	B	127	Total	O	0	0
			127	127		







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	58.99Å 157.38Å 165.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.50 – 2.70 45.39 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.7 (29.50-2.70) 99.6 (45.39-2.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.92 (at 2.69Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.206 , 0.233 0.206 , 0.234	Depositor DCC
$R_{free}$ test set	2163 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	42.9	Xtriage
Anisotropy	0.407	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 43.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.006 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	9487	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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: NDP, UMP, KA5

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	0.38	0/4644	0.62	2/6267 (0.0%)
1	B	0.38	0/4617	0.66	5/6233 (0.1%)
All	All	0.38	0/9261	0.64	7/12500 (0.1%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	117	LYS	C-N-CD	-9.96	98.69	120.60
1	A	166	GLY	N-CA-C	-8.01	93.07	113.10
1	B	118	PRO	CA-N-CD	-7.31	101.27	111.50
1	B	117	LYS	CA-CB-CG	-7.16	97.64	113.40
1	B	166	GLY	N-CA-C	-5.52	99.31	113.10

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	4539	0	4501	140	0
1	B	4511	0	4467	134	0
2	A	25	0	18	1	0
2	B	25	0	18	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	48	0	26	2	0
3	B	48	0	26	4	0
4	A	20	0	11	0	0
4	B	20	0	11	1	0
5	A	124	0	0	6	0
5	B	127	0	0	4	0
All	All	9487	0	9078	267	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 15.

The worst 5 of 267 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:117:LYS:HB3	1:B:118:PRO:HD2	1.21	1.16
1:B:376:LEU:HD22	1:B:379:ILE:HD11	1.36	1.08
1:A:1:MET:HG3	1:A:2:MET:H	1.18	1.06
1:B:346:THR:HG22	1:B:348:VAL:HG23	1.39	1.03
1:A:376:LEU:HD22	1:A:379:ILE:HD11	1.45	0.99

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	539/608 (89%)	486 (90%)	45 (8%)	8 (2%)	12	30
1	B	536/608 (88%)	475 (89%)	46 (9%)	15 (3%)	6	14
All	All	1075/1216 (88%)	961 (89%)	91 (8%)	23 (2%)	8	21

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	24	ASN
1	B	118	PRO
1	B	135	ASP
1	B	298	GLU
1	B	345	ARG

### 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	510/570 (90%)	497 (98%)	13 (2%)	53	82
1	B	507/570 (89%)	496 (98%)	11 (2%)	57	84
All	All	1017/1140 (89%)	993 (98%)	24 (2%)	54	83

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	299	LYS
1	B	25	GLU
1	B	285	GLU
1	A	303	ASN
1	A	304	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	415	ASN
1	A	491	HIS
1	B	171	GLN
1	A	407	ASN
1	B	394	ASN

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

6 ligands are 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	KA5	A	609	-	27,27,27	1.35	3 (11%)	34,38,38	1.93	7 (20%)
3	NDP	A	610	-	43,52,52	1.63	6 (13%)	49,80,80	1.77	10 (20%)
4	UMP	A	611	-	17,21,21	2.13	5 (29%)	23,31,31	3.60	8 (34%)
2	KA5	B	609	-	27,27,27	1.42	3 (11%)	34,38,38	1.93	6 (17%)
3	NDP	B	610	-	43,52,52	1.68	7 (16%)	49,80,80	1.75	8 (16%)
4	UMP	B	611	-	17,21,21	2.20	4 (23%)	23,31,31	3.62	7 (30%)

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	KA5	A	609	-	-	0/9/9/9	0/3/3/3
3	NDP	A	610	-	-	0/30/77/77	0/5/5/5
4	UMP	A	611	-	-	0/6/22/22	0/2/2/2
2	KA5	B	609	-	-	0/9/9/9	0/3/3/3
3	NDP	B	610	-	-	0/30/77/77	0/5/5/5
4	UMP	B	611	-	-	0/6/22/22	0/2/2/2

The worst 5 of 28 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	610	NDP	C4N-C5N	-4.95	1.38	1.49
3	A	610	NDP	C4N-C5N	-4.84	1.38	1.49
2	B	609	KA5	C6-N1	-4.71	1.29	1.37
2	A	609	KA5	C6-N1	-4.42	1.30	1.37
2	A	609	KA5	C2-N3	-3.38	1.29	1.35

The worst 5 of 46 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	610	NDP	C4B-O4B-C1B	-4.84	104.62	109.77
3	B	610	NDP	C4B-O4B-C1B	-4.76	104.70	109.77
2	A	609	KA5	N1-C2-N3	-4.51	120.87	127.46
3	B	610	NDP	N3A-C2A-N1A	-4.48	124.96	128.86
2	B	609	KA5	N1-C2-N3	-4.44	120.97	127.46

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	609	KA5	1	0
3	A	610	NDP	2	0
3	B	610	NDP	4	0
4	B	611	UMP	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 [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, 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	545/608 (89%)	0.00	30 (5%) 26 24	22, 40, 89, 91	0
1	B	542/608 (89%)	0.22	41 (7%) 15 12	22, 44, 90, 91	0
All	All	1087/1216 (89%)	0.11	71 (6%) 20 17	22, 41, 90, 91	0

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	24	ASN	7.2
1	A	605	ASP	6.9
1	B	303	ASN	6.4
1	A	2	MET	6.3
1	B	301	GLU	6.3

### 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](#)

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
4	UMP	A	611	20/20	0.95	0.26	2.10	49,72,80,82	0
4	UMP	B	611	20/20	0.95	0.23	1.15	44,63,71,72	0
2	KA5	A	609	25/25	0.95	0.20	0.74	39,44,49,55	0
2	KA5	B	609	25/25	0.88	0.24	0.06	67,74,77,79	0
3	NDP	B	610	48/48	0.85	0.21	-0.04	88,90,90,90	0
3	NDP	A	610	48/48	0.97	0.16	-0.37	51,60,63,63	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.