



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 06:10 am GMT

PDB ID : 2YOG  
Title : Plasmodium falciparum thymidylate kinase in complex with a (thio)urea-alpha-deoxythymidine inhibitor  
Authors : Huaqing, C.; Carrero-Lerida, J.; Silva, A.P.G.; Whittingham, J.L.; Brannigan, J.A.; Ruiz-Perez, L.M.; Read, K.D.; Wilson, K.S.; Gonzalez-Pacanowska, D.; Gilbert, I.H.  
Deposited on : 2012-10-24  
Resolution : 1.50 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

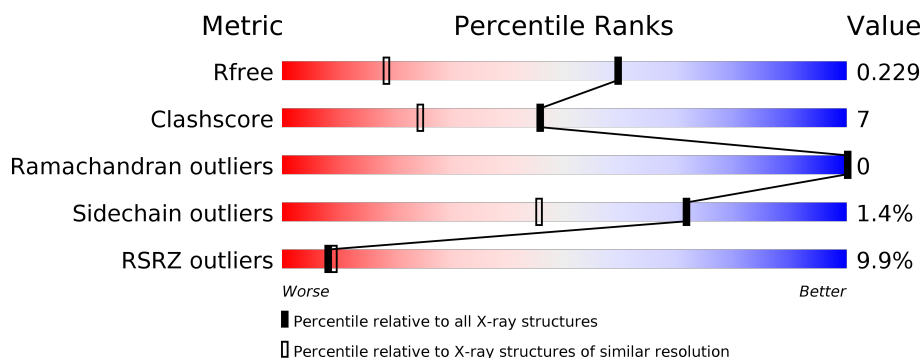
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.50 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	2279 (1.50-1.50)
Clashscore	112137	2503 (1.50-1.50)
Ramachandran outliers	110173	2445 (1.50-1.50)
Sidechain outliers	110143	2443 (1.50-1.50)
RSRZ outliers	101464	2305 (1.50-1.50)

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	210	<div> <div>11%</div> <div> <div></div> <div>79%</div> <div>14%</div> <div>• 6%</div> </div> </div>
1	B	210	<div> <div>8%</div> <div> <div></div> <div>81%</div> <div>12%</div> <div>7%</div> </div> </div>

## 2 Entry composition [i](#)

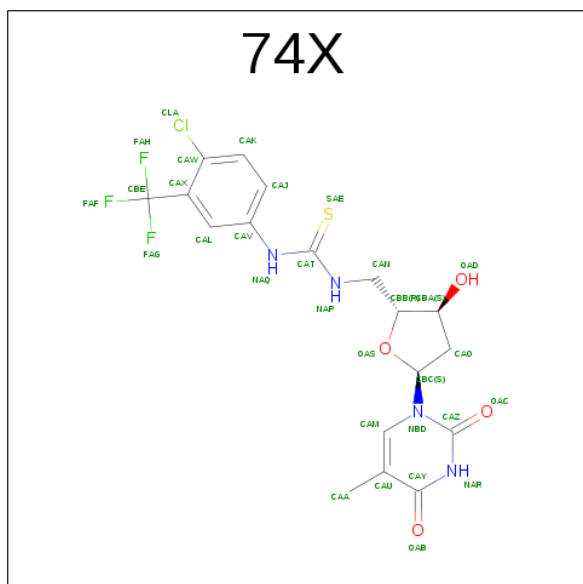
There are 3 unique types of molecules in this entry. The entry contains 3969 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 THYMIDYLATE KINASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	197	Total	C	N	O	S	6	23	0
			1752	1144	279	323	6			
1	B	196	Total	C	N	O	S	0	14	0
			1721	1123	276	316	6			

- Molecule 2 is 1-[4-CHLORANYL-3-(TRIFLUOROMETHYL)PHENYL]-3-[[[(2R,3S)-5-[5-METHYL-2,4-BIS(OXIDANYLIDENE)PYRIMIDIN-1-YL]-3-OXIDANYL-OXOLAN-2-YL]METHYL]THIOUREA (three-letter code: 74X) (formula: C<sub>18</sub>H<sub>18</sub>ClF<sub>3</sub>N<sub>4</sub>O<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Cl	F	N	O	S
			31	18	1	3	4	4	1
2	B	1	Total	C	Cl	F	N	O	S
			31	18	1	3	4	4	1

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	220	Total 220	O 220	0	0
3	B	214	Total 214	O 214	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.06Å 58.79Å 71.71Å 90.00° 95.78° 90.00°	Depositor
Resolution (Å)	23.37 – 1.50 23.37 – 1.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (23.37-1.50) 99.9 (23.37-1.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.79 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.5.0088	Depositor
R, $R_{free}$	0.182 , 0.233 0.178 , 0.229	Depositor DCC
$R_{free}$ test set	3308 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.8	Xtriage
Anisotropy	0.592	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 45.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3969	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.32% 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: 74X

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	0.71	0/1862	0.73	1/2514 (0.0%)
1	B	0.72	0/1799	0.71	1/2424 (0.0%)
All	All	0.72	0/3661	0.72	2/4938 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	78	ARG	NE-CZ-NH2	-5.49	117.56	120.30
1	A	78	ARG	NE-CZ-NH2	-5.41	117.60	120.30

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	1752	0	1795	26	1
1	B	1721	0	1730	26	0
2	A	31	0	18	0	0
2	B	31	0	18	0	0
3	A	220	0	0	4	1
3	B	214	0	0	3	0
All	All	3969	0	3561	52	1

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 7.

All (52) 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:41[B]:HIS:CD2	1:B:43[B]:TYR:CE1	2.12	1.36
1:B:41[B]:HIS:NE2	1:B:43[B]:TYR:CE1	1.98	1.30
1:B:41[B]:HIS:NE2	1:B:43[B]:TYR:CZ	2.08	1.17
1:B:21:LYS:HA	1:B:135:LEU:HD12	1.33	1.10
1:B:41[B]:HIS:CD2	1:B:43[B]:TYR:HE1	1.57	1.02
1:B:21:LYS:HA	1:B:135:LEU:CD1	1.91	0.99
1:A:131[A]:VAL:HG21	1:A:176[A]:ILE:HD12	1.45	0.98
1:B:41[B]:HIS:HD2	1:B:43[B]:TYR:HE1	1.11	0.95
1:B:46:ASN:H	1:B:77:ASN:HD21	1.31	0.77
1:A:131[A]:VAL:CG2	1:A:176[A]:ILE:HD12	2.14	0.76
1:B:21:LYS:CA	1:B:135:LEU:HD12	2.17	0.73
1:A:21:LYS:HA	1:A:135:LEU:HD12	1.70	0.73
1:A:126:LEU:HB2	1:A:208[B]:LEU:HD13	1.73	0.70
1:A:140:ASN:C	1:A:140:ASN:HD22	1.96	0.69
1:A:139:PRO:HG2	1:A:160:LYS:HG2	1.77	0.67
1:A:63:ASN:C	1:A:63:ASN:HD22	1.98	0.66
1:A:21:LYS:HA	1:A:135:LEU:CD1	2.26	0.66
1:B:115:ASN:HD22	1:B:118:TRP:H	1.42	0.65
1:B:131:VAL:HG21	1:B:176:ILE:HD12	1.79	0.65
1:A:115:ASN:HD22	1:A:118:TRP:H	1.44	0.63
1:B:41[B]:HIS:NE2	1:B:43[B]:TYR:OH	2.31	0.63
1:B:157[A]:GLU:H	1:B:157[A]:GLU:CD	2.02	0.62
1:B:41[B]:HIS:HD2	1:B:43[B]:TYR:CE1	1.86	0.61
1:B:21:LYS:HE2	1:B:99:ARG:CZ	2.31	0.60
1:A:83[B]:ASN:OD1	3:A:2113:HOH:O	2.16	0.60
1:A:116:LYS:HE3	1:A:165[B]:THR:HG23	1.85	0.58
1:B:131:VAL:CG2	1:B:176:ILE:HD12	2.34	0.57
1:A:137[A]:VAL:HG13	3:A:2163:HOH:O	2.04	0.56
1:A:131[A]:VAL:HG21	1:A:176[A]:ILE:CD1	2.29	0.56
1:A:21:LYS:HE2	1:A:99:ARG:CZ	2.36	0.55
1:A:46:ASN:H	1:A:77:ASN:HD21	1.55	0.54
1:B:46:ASN:N	1:B:77:ASN:HD21	2.04	0.54
1:A:152:ILE:N	3:A:2166:HOH:O	2.46	0.49
1:B:17:ASP:HB2	1:B:21:LYS:HG3	1.94	0.48
1:B:152[B]:ILE:O	3:B:2155:HOH:O	2.20	0.48
1:A:18:ARG:HD3	1:A:18:ARG:HA	1.49	0.47
1:B:170:ALA:HA	1:B:175:TRP:CD1	2.50	0.47

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:91:LYS:HE2	3:B:2109:HOH:O	2.13	0.47
1:B:40:LYS:HE3	1:B:42:LEU:HD21	1.97	0.46
1:B:178:ILE:HG21	1:B:187:ILE:HG23	1.99	0.44
1:A:179:ASP:HB3	1:A:182:ARG:HE	1.82	0.44
1:A:183:LYS:HG2	3:A:2194:HOH:O	2.18	0.44
1:A:140:ASN:C	1:A:140:ASN:ND2	2.67	0.44
1:A:115:ASN:ND2	1:A:118:TRP:H	2.12	0.43
1:A:108:SER:HB3	1:A:114:LEU:HD12	2.00	0.43
1:A:19:SER:HB3	1:A:137[A]:VAL:HG12	2.01	0.41
1:A:21:LYS:NZ	1:A:21:LYS:HB2	2.35	0.41
1:B:160:LYS:NZ	3:B:2149:HOH:O	2.54	0.41
1:A:42:LEU:O	1:A:97:CYS:HA	2.20	0.40
1:A:63:ASN:C	1:A:63:ASN:ND2	2.72	0.40
1:B:21:LYS:HB2	1:B:21:LYS:NZ	2.36	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:173[B]:ASP:OD2	3:A:2042:HOH:O[2_545]	2.19	0.01

## 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	216/210 (103%)	209 (97%)	7 (3%)	0	100	100
1	B	203/210 (97%)	198 (98%)	5 (2%)	0	100	100
All	All	419/420 (100%)	407 (97%)	12 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	205/193 (106%)	201 (98%)	4 (2%)	60	28
1	B	195/193 (101%)	193 (99%)	2 (1%)	80	60
All	All	400/386 (104%)	394 (98%)	6 (2%)	71	42

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	17[A]	ASP
1	A	17[B]	ASP
1	A	63	ASN
1	A	140	ASN
1	B	166	TYR
1	B	200	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	34	ASN
1	A	63	ASN
1	A	77	ASN
1	A	115	ASN
1	A	140	ASN
1	A	159	GLN
1	B	77	ASN
1	B	115	ASN
1	B	159	GLN
1	B	168	HIS

### 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](#)

2 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	74X	A	211	-	28,33,33	1.42	4 (14%)	39,49,49	1.81	6 (15%)
2	74X	B	211	-	28,33,33	1.31	5 (17%)	39,49,49	1.87	5 (12%)

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	74X	A	211	-	-	0/15/31/31	0/3/3/3
2	74X	B	211	-	-	0/15/31/31	0/3/3/3

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	211	74X	CAV-NAQ	-3.98	1.33	1.41
2	B	211	74X	CAV-NAQ	-3.98	1.33	1.41
2	B	211	74X	CAY-NAR	2.36	1.37	1.33
2	A	211	74X	CAT-SAE	2.41	1.74	1.68
2	B	211	74X	CAT-SAE	2.42	1.74	1.68
2	B	211	74X	CBE-CAX	2.51	1.56	1.50

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	211	74X	CAN-CBB	2.51	1.57	1.52
2	B	211	74X	CAN-CBB	2.60	1.58	1.52
2	A	211	74X	CAY-NAR	4.04	1.40	1.33

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	211	74X	CAU-CAY-NAR	-5.77	118.88	125.24
2	A	211	74X	CAU-CAY-NAR	-5.19	119.51	125.24
2	B	211	74X	CBB-CAN-NAP	-4.76	103.21	112.68
2	A	211	74X	CBB-CAN-NAP	-4.04	104.64	112.68
2	A	211	74X	CAV-NAQ-CAT	-3.05	119.98	128.53
2	B	211	74X	CAU-CAM-NBD	-2.43	119.52	122.15
2	A	211	74X	SAE-CAT-NAP	-2.22	119.33	123.19
2	A	211	74X	FAH-CBE-CAX	-2.13	108.85	112.69
2	B	211	74X	OAS-CBC-CAO	2.60	111.24	106.25
2	B	211	74X	CAY-NAR-CAZ	6.04	120.44	115.16
2	A	211	74X	CAY-NAR-CAZ	6.56	120.90	115.16

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	197/210 (93%)	0.67	23 (11%) <b>5</b> <b>5</b>	8, 20, 34, 44	0
1	B	196/210 (93%)	0.56	16 (8%) <b>12</b> <b>14</b>	9, 19, 35, 43	0
All	All	393/420 (93%)	0.62	39 (9%) <b>8</b> <b>9</b>	8, 20, 35, 44	0

All (39) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	141	TYR	9.9
1	B	142	ALA	8.5
1	A	4	ASP	7.9
1	B	141	TYR	7.4
1	B	199	LYS	6.6
1	A	18	ARG	5.3
1	B	200	VAL	5.1
1	A	22	SER	4.6
1	A	152	ILE	4.6
1	A	153	TYR	4.3
1	B	36	ASN	4.1
1	A	17[A]	ASP	3.9
1	A	64[A]	SER	3.8
1	B	154	GLU	3.1
1	B	64	SER	3.1
1	A	155[A]	LYS	3.0
1	A	21	LYS	3.0
1	A	20	GLY	2.9
1	A	140	ASN	2.9
1	B	34	ASN	2.8
1	A	61	MET	2.7
1	B	12	VAL	2.7
1	A	154	GLU	2.6
1	B	43[A]	TYR	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	200	VAL	2.5
1	A	19	SER	2.5
1	B	201	GLU	2.4
1	B	21	LYS	2.4
1	B	184	ILE	2.4
1	A	138	PRO	2.4
1	B	22	SER	2.3
1	A	12	VAL	2.3
1	A	63	ASN	2.3
1	A	105	VAL	2.2
1	B	198	ILE	2.2
1	A	34	ASN	2.1
1	B	204	GLU	2.1
1	A	65	MET	2.0
1	A	111	ALA	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 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(Å <sup>2</sup> )	Q<0.9
2	74X	A	211	31/31	0.94	0.13	-0.47	12,16,42,44	10
2	74X	B	211	31/31	0.93	0.10	-0.75	10,13,28,34	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.