wwPDB X-ray Structure Validation Summary Report (i)

Dec 7, 2023 – 09:19 pm GMT

PDB ID	:	2WK6
Title	:	Structural features of native human thymidine phosphorylase and in complex with 5-iodouracil
Authors	:	Mitsiki, E.; Papageorgiou, A.C.; Iyer, S.; Thiyagarajan, N.; Prior, S.H.; Sleep, D.; Finnis, C.; Acharya, K.R.
Deposited on	:	2009-06-05
Resolution	:	2.50 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

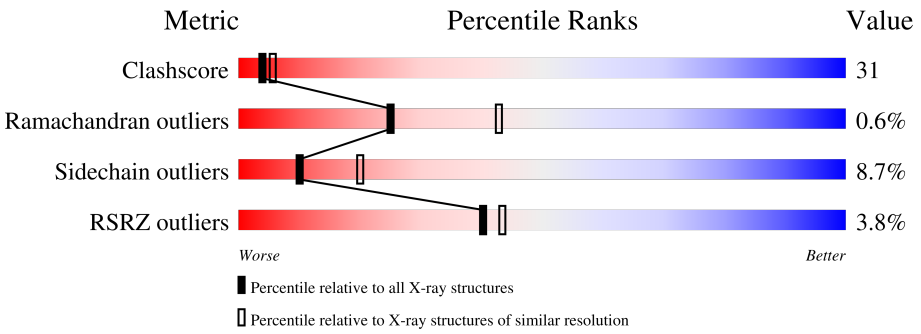
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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))
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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% 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	482	% 51% 37% • 7%
1	B	482	6% 47% 41% • 7%

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
2	IUR	B	1482	-	-	X	-

2 Entry composition (i)

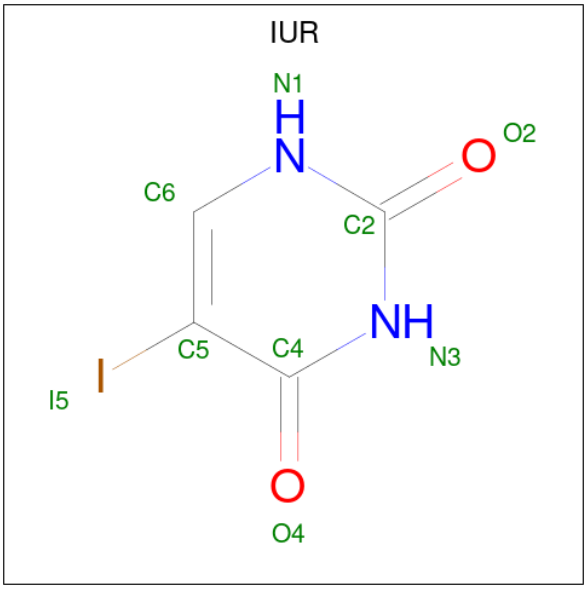
There are 3 unique types of molecules in this entry. The entry contains 6576 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 THYMIDINE PHOSPHORYLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	447	3257	2034	597	610	16	0	0	1
1	B	447	3257	2034	597	610	16	0	0	1

- Molecule 2 is 5-IODOURACIL (three-letter code: IUR) (formula: $C_4H_3IN_2O_2$).



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

- Molecule 3 is water.

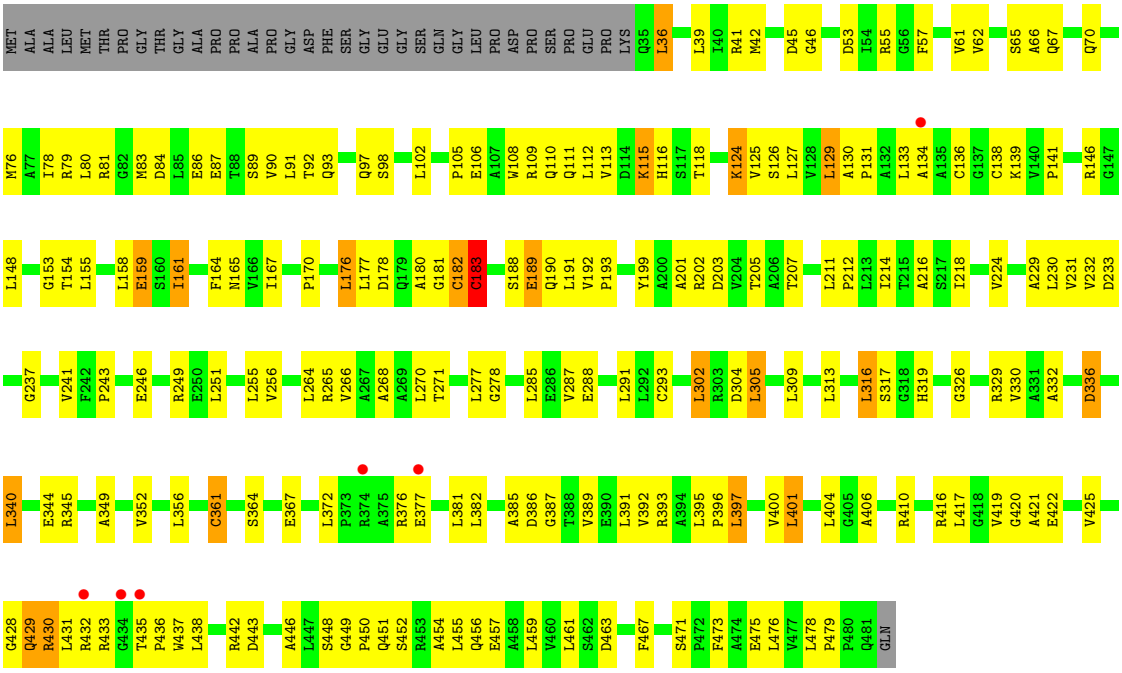
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	33	33	33	0	0
			Total	O		
3	B	11	11	11	0	0

3 Residue-property plots (i)

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

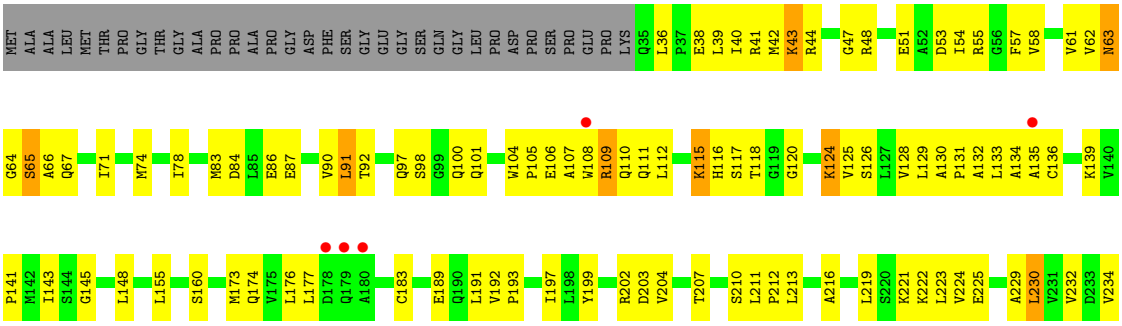
- Molecule 1: THYMIDINE PHOSPHORYLASE

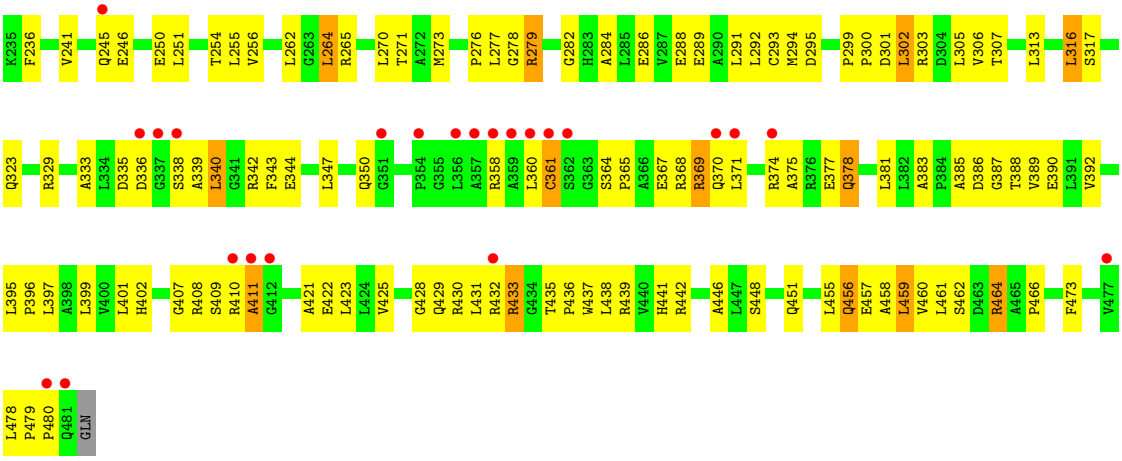
Chain A:



- Molecule 1: THYMIDINE PHOSPHORYLASE

Chain B:




K235 F236 V241 Q245 E246 E250 L251 T254 L255 V256 L262 G263 L264 R265 L270 T271 A272 M273 P276 L277 G278 R279 G282 H283 A284 L285 E286 V287 E288 E289 A290 L291 L292 C293 M294 D295 P299 P300 D301 L302 R303 D304 L305 V306 T307 L313 L316 S317
Q323 R329 A333 L334 D335 D336 G337 S338 A339 L340 G341 R342 F343 E344 L347 Q350 G351 P354 G355 L356 A357 R358 A359 L360 C361 S362 G363 S364 P365 A366 E367 R368 R369 Q370 L371 R374 A375 R376 E377 Q378 L381 L382 A383 P384 A385 D386 G387 T388 V389 E390 L391 V392
L395 P396 L397 A398 L399 V400 L401 H402 G407 R408 S409 R410 A411 G412 A421 E422 L423 L424 V425 G428 Q429 R430 L431 R432 R433 G434 T435 P436 W437 L438 R439 V440 H441 R442 A446 L447 S448 Q451 L455 Q456 E457 A458 L459 V460 L461 S462 D463 R464 A465 P466 F473 V477
L478 P479 P480 Q481 GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	61.91Å 67.27Å 212.27Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.59 – 2.50 46.59 – 2.46	Depositor EDS
% Data completeness (in resolution range)	95.2 (46.59-2.50) 93.2 (46.59-2.46)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.29 (at 2.45Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.227 , 0.283 0.223 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	41.6	Xtriage
Anisotropy	0.465	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 40.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6576	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.90% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: IUR

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	0.50	3/3304 (0.1%)	0.67	1/4484 (0.0%)
1	B	0.36	0/3304	0.63	0/4484
All	All	0.44	3/6608 (0.0%)	0.65	1/8968 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	159	GLU	CD-OE1	-12.20	1.12	1.25
1	A	159	GLU	CD-OE2	-11.94	1.12	1.25
1	A	159	GLU	C-O	-5.87	1.12	1.23

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	159	GLU	OE1-CD-OE2	-9.83	111.51	123.30

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	3257	0	3349	197	0
1	B	3257	0	3349	230	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	9	0	3	2	0
2	B	9	0	3	4	0
3	A	33	0	0	2	0
3	B	11	0	0	0	0
All	All	6576	0	6704	404	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 31.

The worst 5 of 404 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:65:SER:CB	1:B:410:ARG:NH2	1.73	1.49
1:A:65:SER:HB2	1:B:410:ARG:CZ	1.48	1.42
1:A:65:SER:HB2	1:B:410:ARG:NH2	0.85	1.17
1:A:65:SER:CB	1:B:410:ARG:HH22	1.41	1.15
1:A:417:LEU:O	3:A:2028:HOH:O	1.79	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	445/482 (92%)	426 (96%)	17 (4%)	2 (0%)	34 54
1	B	445/482 (92%)	418 (94%)	24 (5%)	3 (1%)	22 39
All	All	890/964 (92%)	844 (95%)	41 (5%)	5 (1%)	25 43

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	109	ARG
1	A	183	CYS
1	B	411	ALA
1	A	109	ARG
1	B	464	ARG

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	333/359 (93%)	304 (91%)	29 (9%)	10 20
1	B	333/359 (93%)	304 (91%)	29 (9%)	10 20
All	All	666/718 (93%)	608 (91%)	58 (9%)	10 20

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	471	SER
1	B	456	GLN
1	B	106	GLU
1	B	433	ARG
1	B	340	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	370	GLN
1	B	378	GLN
1	B	456	GLN
1	A	378	GLN
1	B	63	ASN

5.3.3 RNA (i)

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 monosaccharides 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	IUR	A	1482	-	9,9,9	2.67	4 (44%)	11,12,12	1.35	1 (9%)
2	IUR	B	1482	-	9,9,9	2.66	4 (44%)	11,12,12	1.32	1 (9%)

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	IUR	A	1482	-	-	-	0/1/1/1
2	IUR	B	1482	-	-	-	0/1/1/1

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1482	IUR	C6-C5	4.64	1.41	1.35
2	A	1482	IUR	C2-N1	4.49	1.42	1.36
2	A	1482	IUR	C6-C5	4.44	1.41	1.35
2	B	1482	IUR	C2-N1	4.29	1.42	1.36
2	A	1482	IUR	C2-N3	2.96	1.42	1.37

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1482	IUR	O2-C2-N1	3.42	126.56	122.79
2	B	1482	IUR	O2-C2-N1	3.40	126.54	122.79

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1482	IUR	2	0
2	B	1482	IUR	4	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, 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	447/482 (92%)	0.08	6 (1%)	77	79	21, 36, 56, 78	0
1	B	447/482 (92%)	0.41	28 (6%)	20	21	22, 43, 69, 92	0
All	All	894/964 (92%)	0.25	34 (3%)	40	43	21, 39, 65, 92	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	337	GLY	8.2
1	B	411	ALA	4.9
1	B	336	ASP	4.9
1	B	374	ARG	3.8
1	B	480	PRO	3.8

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 monosaccharides 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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
2	IUR	A	1482	9/9	0.94	0.12	26,30,40,72	0
2	IUR	B	1482	9/9	0.95	0.13	29,33,42,69	0

6.5 Other polymers (i)

There are no such residues in this entry.