wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 08:29 PM GMT

PDB ID	:	1KI6
Title	:	CRYSTAL STRUCTURE OF THYMIDINE KINASE FROM HERPES SIMPLEX VIRUS TYPE I COMPLEXED WITH A 5-IODOURACIL ANHYDROHEXITOL NUCLEOSIDE
Authors	:	Champness, J.N.; Bennett, M.S.; Wien, F.; Herdewijn, P.; Ostrowski, T.; Summers, W.C.; Sanderson, M.R.
Deposited on	:	1998-05-18
Resolution	:	2.37 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

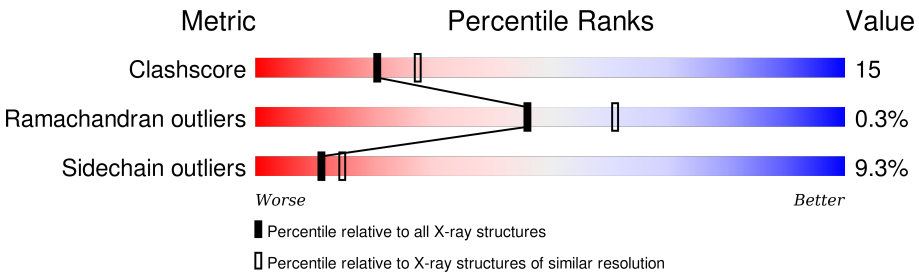
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	FAILED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.37 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	102246	4595 (2.40-2.36)
Ramachandran outliers	100387	4520 (2.40-2.36)
Sidechain outliers	100360	4522 (2.40-2.36)

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 >=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% 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.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	331	63% 24% • 9%
2	B	331	60% 27% 5% 8%

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
3	SO4	B	4	-	X	-	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4946 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 KINASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	300	2300	1467	402	416	15	0	0	0

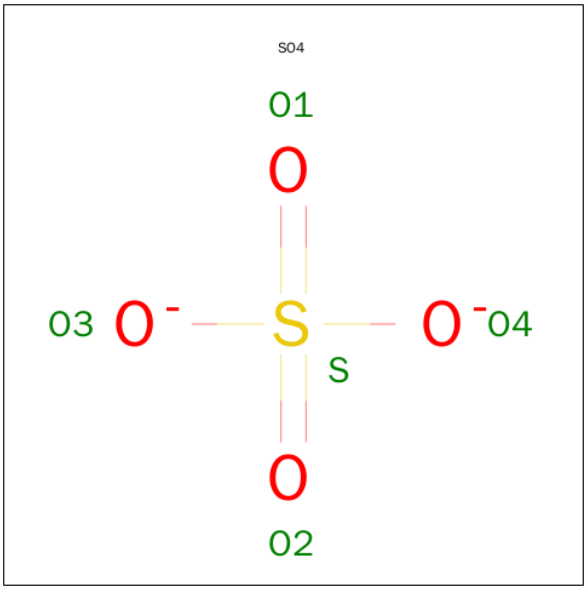
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	ALA	PRO	CONFLICT	UNP P03176

- Molecule 2 is a protein called THYMIDINE KINASE.

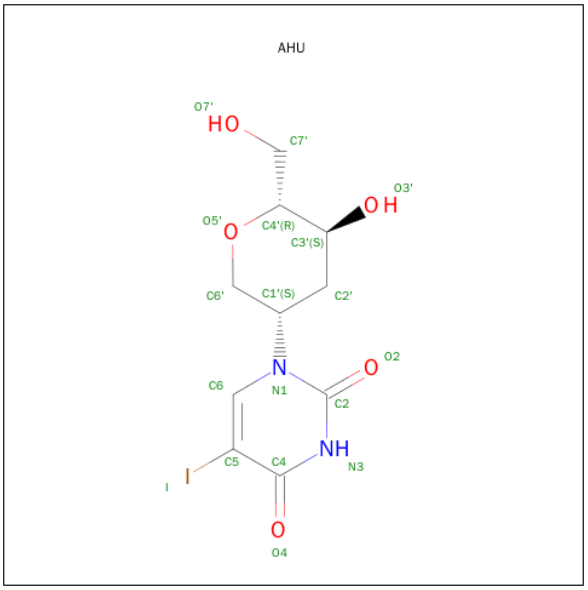
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	304	2325	1485	406	419	15	0	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
3	A	1	5	4	1	0	0
			Total	O	S		
3	B	1	5	4	1	0	0

- Molecule 4 is 1',5'-ANHYDRO-2',3'-DIDEOXY-2'-(5-IODOURACIL-1-YL)-D-ABABINO-HEXITOL (three-letter code: AHU) (formula: $C_{10}H_{13}IN_2O_5$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	I	N	O		
4	A	1	18	10	1	2	5	0	0
			Total	C	I	N	O		
4	B	1	18	10	1	2	5	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	135	135	135	0	0
			Total	O		
5	B	140	140	140	0	0

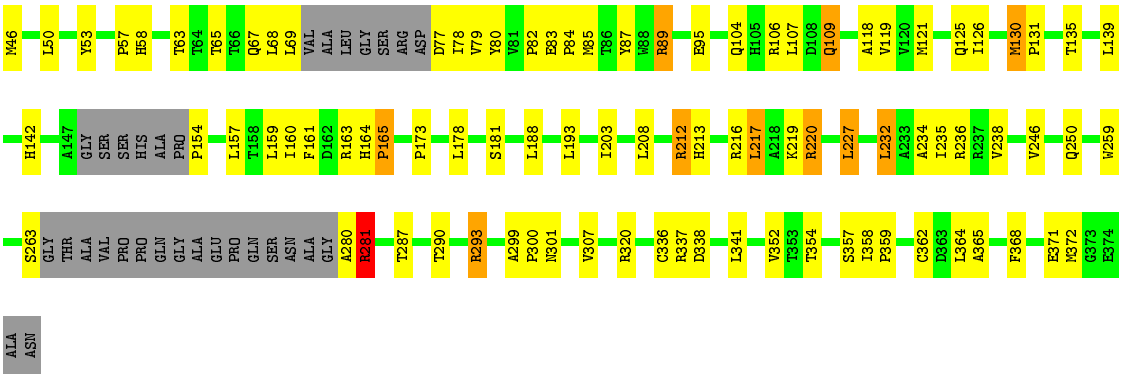
3 Residue-property plots (i)

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 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.

Note EDS failed to run properly.

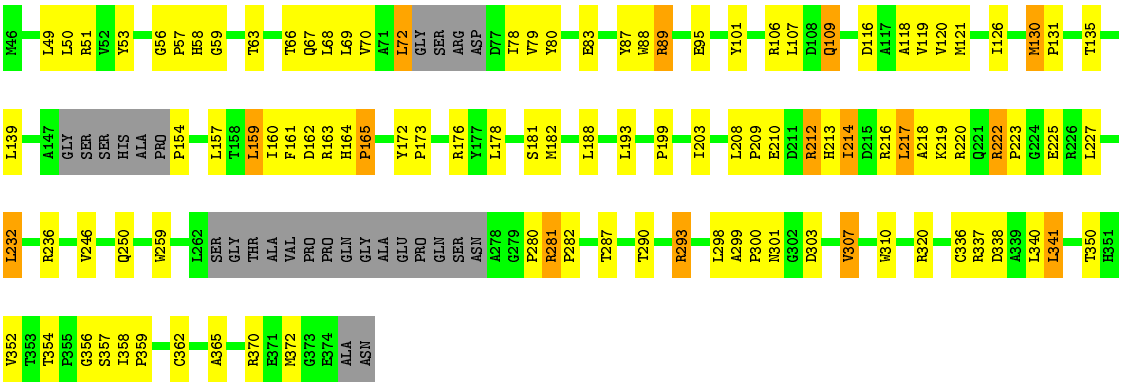
- Molecule 1: THYMIDINE KINASE

Chain A: 63% 24% • 9%



- Molecule 2: THYMIDINE KINASE

Chain B: 60% 27% 5% 8%



4 Data and refinement statistics (i)

EDS failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	113.60Å 116.40Å 108.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	12.00 – 2.37	Depositor
% Data completeness (in resolution range)	88.0 (12.00-2.37)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	2.65 (at 2.38Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.223 , 0.317	Depositor
Wilson B-factor (Å^2)	24.5	Xtriage
Anisotropy	0.009	Xtriage
Estimated twinning fraction	0.036 for -k,-h,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 26498 reflections	Xtriage
Total number of atoms	4946	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.82% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: SO4, AHU

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	1/2351 (0.0%)	0.78	1/3207 (0.0%)
2	B	0.72	1/2377 (0.0%)	0.78	1/3244 (0.0%)
All	All	0.72	2/4728 (0.0%)	0.78	2/6451 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	46	MET	C-N	8.48	1.50	1.34
2	B	336	CYS	CB-SG	-5.26	1.73	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	154	PRO	N-CA-CB	6.18	110.72	103.30
1	A	154	PRO	N-CA-CB	5.47	109.86	103.30

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	2300	0	2313	65	0
2	B	2325	0	2343	76	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	5	0	0	0	0
3	B	5	0	0	0	0
4	A	18	0	12	3	0
4	B	18	0	12	1	0
5	A	135	0	0	5	0
5	B	140	0	0	6	0
All	All	4946	0	4680	136	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 136 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:126:ILE:HD11	2:B:126:ILE:HD11	1.56	0.88
2:B:232:LEU:HD13	2:B:236:ARG:HD2	1.66	0.77
2:B:59:GLY:HA3	2:B:217:LEU:HD22	1.67	0.76
1:A:69:LEU:HD23	1:A:337:ARG:HG2	1.68	0.75
4:B:2:AHU:O7'	4:B:2:AHU:H6'1	1.86	0.75

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	292/331 (88%)	277 (95%)	13 (4%)	2 (1%)	26 36
2	B	296/331 (89%)	284 (96%)	12 (4%)	0	100 100
All	All	588/662 (89%)	561 (95%)	25 (4%)	2 (0%)	46 61

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	281	ARG
1	A	300	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	241/263 (92%)	220 (91%)	21 (9%)	13 17
2	B	243/264 (92%)	219 (90%)	24 (10%)	10 13
All	All	484/527 (92%)	439 (91%)	45 (9%)	11 15

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	362	CYS
2	B	109	GLN
2	B	307	VAL
2	B	72	LEU
2	B	130	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	109	GLN
2	B	342	GLN
2	B	221	GLN
1	A	342	GLN
2	B	142	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

4 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	$\#\|Z\| > 2$
4	AHU	A	1	-	15,19,19	2.11	4 (26%)	12,27,27	4.91	8 (66%)
3	SO4	A	3	-	4,4,4	1.75	1 (25%)	6,6,6	0.64	0
4	AHU	B	2	-	15,19,19	2.58	7 (46%)	12,27,27	4.79	6 (50%)
3	SO4	B	4	-	4,4,4	2.62	3 (75%)	6,6,6	0.97	1 (16%)

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
4	AHU	A	1	-	-	0/4/19/19	0/2/2/2
3	SO4	A	3	-	-	0/0/0/0	0/0/0/0
4	AHU	B	2	-	-	0/4/19/19	0/2/2/2
3	SO4	B	4	-	-	0/0/0/0	0/0/0/0

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	2	AHU	C2'-C1'	-4.55	1.44	1.53
4	B	2	AHU	C3'-C4'	-4.40	1.44	1.52

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1	AHU	C6-C5	-3.91	1.29	1.38
3	B	4	SO4	O4-S	-3.58	1.34	1.47
4	A	1	AHU	O3'-C3'	-3.30	1.35	1.43

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	2	AHU	O5'-C4'-C3'	-7.65	96.49	109.63
4	A	1	AHU	C7'-C4'-C3'	-4.16	107.98	113.29
4	A	1	AHU	O3'-C3'-C4'	-3.61	102.62	109.88
4	B	2	AHU	O3'-C3'-C4'	-3.26	103.32	109.88
4	A	1	AHU	O7'-C7'-C4'	-2.91	101.72	111.33

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1	AHU	3	0
4	B	2	AHU	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)

EDS failed to run properly - this section will therefore be empty.

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

EDS failed to run properly - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS failed to run properly - this section will therefore be empty.

6.4 Ligands (i)

EDS failed to run properly - this section will therefore be empty.

6.5 Other polymers (i)

EDS failed to run properly - this section will therefore be empty.