

wwPDB X-ray Structure Validation Summary Report (i)

Jan 3, 2024 – 01:34 pm GMT

PDB ID : 5FUV

Title : catalytic domain of Thymidine kinase from Trypanosoma brucei with dThd Authors : Timm, J.; Valente, M.; Castillo-Acosta, V.; Balzarini, T.; Nettleship, J.E.;

Rada, H.; Wilson, K.S.; Gonzalez-Pacanowska, D.

Deposited on : 2016-01-31

Resolution : 2.30 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 (1)) 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)

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

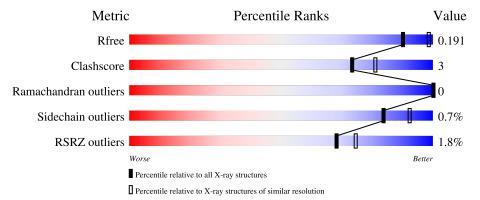
Validation Pipeline (wwPDB-VP) : 2.36

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 2.30 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	Similar resolution
Metric	$(\# ext{Entries})$	$(\# ext{Entries}, ext{ resolution range}(ext{Å}))$
R_{free}	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	182	86%	8% • 5%
1	В	182	80%	9% • 10%

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:



\mathbf{Mol}	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GOL	A	1386	-	X	-	-



2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2733 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.

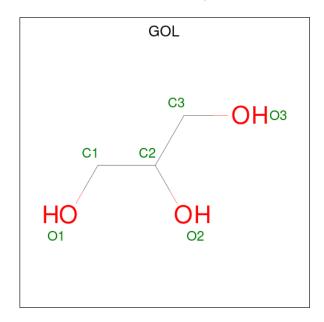
• Molecule 1 is a protein called THYMDINE KINASE.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	173	Total 1333	_		O 256	S 12	0	0	0
1	В	164	Total 1255	_		_	S 12	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chai	in	Residue	Modelled	Actual	Comment	Reference
A		292	ASP	GLY	engineered mutation	UNP Q38CF2
В		292	ASP	GLY	engineered mutation	UNP Q38CF2

• Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 6 3 3	0	0

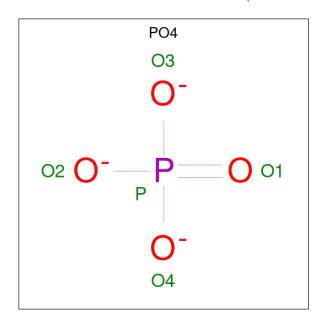
Continued on next page...



Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 6 3 3	0	0
2	В	1	Total C O 6 3 3	0	0

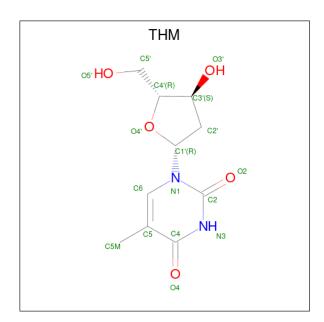
 \bullet Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: $\mathrm{O_4P}).$



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

 \bullet Molecule 4 is THYMIDINE (three-letter code: THM) (formula: $C_{10}H_{14}N_2O_5).$





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 17				0	0
4	В	1	Total 17	_		_	0	0

• Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Zn 1 1	0	0
5	В	1	Total Zn 1 1	0	0

• Molecule 6 is water.

\mathbf{Mol}	Chain	Residues	${f Atoms}$	ZeroOcc	AltConf
6	A	62	Total O 62 62	0	0
6	В	24	Total O 24 24	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: THYMDINE KINASE

Chain A:

86%
8%
• Molecule 1: THYMDINE KINASE

Chain B:

80%
9%
• 10%



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants	114.68Å 114.68Å 105.60Å	Donogiton
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	114.68 - 2.30	Depositor
Resolution (A)	46.13 - 2.12	EDS
% Data completeness	99.8 (114.68-2.30)	Depositor
(in resolution range)	99.1 (46.13-2.12)	EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.29 (at 2.12Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
D D.	0.179 , 0.207	Depositor
R, R_{free}	0.187 , 0.191	DCC
R_{free} test set	2016 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	50.0	Xtriage
Anisotropy	0.168	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.33, 48.3	EDS
L-test for twinning ²	$ < L > = 0.49, < L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	2733	wwPDB-VP
Average B, all atoms (Å ²)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.59% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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: GOL, PO4, ZN, THM

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		nd lengths	Bond angles		
MIOI	Chain	RMSZ	# Z > 5	RMSZ	# Z > 5	
1	A	1.21	5/1353 (0.4%)	1.09	$6/1823 \; (0.3\%)$	
1	В	1.03	0/1275	1.06	7/1720 (0.4%)	
All	All	1.13	5/2628~(0.2%)	1.07	13/3543 (0.4%)	

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\text{\AA})$	Ideal(Å)
1	A	292	ASP	CB-CG	9.29	1.71	1.51
1	A	274	ASP	CB-CG	5.70	1.63	1.51
1	A	372	SER	CB-OG	5.50	1.49	1.42
1	A	322	GLU	CG-CD	5.38	1.60	1.51
1	A	322	GLU	CD-OE1	5.24	1.31	1.25

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
1	A	228	ARG	NE-CZ-NH2	-7.75	116.42	120.30
1	A	228	ARG	NE-CZ-NH1	6.74	123.67	120.30
1	A	315	ASP	CB-CG-OD2	-6.21	112.71	118.30
1	В	369	MET	CG-SD-CE	6.11	109.97	100.20
1	A	244	ASP	CB-CG-OD1	6.09	123.78	118.30

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	1333	0	1284	6	0
1	В	1255	0	1209	8	0
2	A	12	0	15	3	0
2	В	6	0	8	1	0
3	A	5	0	0	0	0
4	A	17	0	14	0	0
4	В	17	0	14	0	0
5	A	1	0	0	0	0
5	В	1	0	0	0	0
6	A	62	0	0	2	0
6	В	24	0	0	1	0
All	All	2733	0	2544	17	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 3.

The worst 5 of 17 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	$egin{aligned} ext{Clash} \ ext{overlap } (ext{Å}) \end{aligned}$
2:A:1386:GOL:O2	6:A:2062:HOH:O	2.12	0.68
1:B:343:GLU:OE1	1:B:381:LYS:HE2	2.04	0.58
1:B:264:VAL:HG12	1:B:266:VAL:HG23	1.86	0.56
2:A:1386:GOL:H31	6:A:2004:HOH:O	2.08	0.53
2:B:1384:GOL:H32	6:B:2003:HOH:O	2.09	0.53

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	Analysed Favoured Allowed		Outliers	Percentiles	
1	A	$169/182 \ (93\%)$	167 (99%)	2 (1%)	0	100	100
1	В	160/182 (88%)	154 (96%)	6 (4%)	0	100	100
All	All	329/364 (90%)	321 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	142/156 (91%)	141 (99%)	1 (1%)	84 92		
1	В	133/156 (85%)	132 (99%)	1 (1%)	81 91		
All	All	275/312 (88%)	273 (99%)	2 (1%)	84 92		

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	376	SER
1	В	376	SER

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

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)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 for Mogul analysis.

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		Chain	Res	Link	Во	ond leng	$\overline{ ext{gths}}$	В	ond ang	gles
IVIOI	туре	Chain	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	PO4	A	1387	-	4,4,4	0.87	0	6,6,6	2.43	2 (33%)
2	GOL	A	1386	-	5,5,5	1.75	1 (20%)	5,5,5	3.40	4 (80%)
4	THM	A	1388	-	18,18,18	1.87	6 (33%)	26,26,26	2.11	9 (34%)
4	THM	В	1385	-	18,18,18	1.57	5 (27%)	26,26,26	1.99	10 (38%)
2	GOL	В	1384	-	5,5,5	1.33	1 (20%)	5,5,5	1.03	0
2	GOL	A	1385	-	5,5,5	0.66	0	5,5,5	0.86	0

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	GOL	A	1386	-	-	3/4/4/4	-
4	THM	A	1388	-	-	0/6/18/18	0/2/2/2
4	THM	В	1385	-	-	2/6/18/18	0/2/2/2
2	GOL	В	1384	-	-	3/4/4/4	-
2	GOL	A	1385	-	-	4/4/4/4	-

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	Observed(A)	$Ideal(\AA)$
4	A	1388	THM	C4-N3	-3.61	1.32	1.38
4	A	1388	THM	C4-C5	3.46	1.50	1.44

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	Observed(A)	$\operatorname{Ideal}(\text{\AA})$
4	В	1385	THM	C6-C5	3.41	1.40	1.34
4	A	1388	THM	C6-N1	-2.94	1.33	1.38
4	A	1388	THM	C2-N3	-2.74	1.33	1.38

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
2	A	1386	GOL	O3-C3-C2	5.41	136.16	110.20
3	A	1387	PO4	O4-P-O2	4.04	120.93	107.97
4	A	1388	THM	N3-C2-N1	4.04	120.25	114.89
4	A	1388	THM	O4'-C4'-C5'	-3.95	100.67	109.21
3	A	1387	PO4	O2-P-O1	-3.90	96.62	110.89

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1385	GOL	O1-C1-C2-C3
2	В	1384	GOL	O1-C1-C2-C3
2	В	1384	GOL	C1-C2-C3-O3
4	В	1385	THM	O4'-C4'-C5'-O5'
4	В	1385	THM	C3'-C4'-C5'-O5'

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1386	GOL	2	0
2	В	1384	GOL	1	0
2	A	1385	GOL	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^{th} 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	$\langle { m RSRZ} \rangle$	$\# \mathrm{RSRZ}{>}2$	$OWAB(Å^2)$	Q<0.9
1	A	173/182 (95%)	-0.25	3 (1%) 70 76	33, 47, 90, 124	1 (0%)
1	В	164/182 (90%)	-0.22	3 (1%) 68 74	42, 64, 98, 129	0
All	All	337/364 (92%)	-0.23	6 (1%) 68 74	33, 56, 93, 129	1 (0%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	245	THR	3.6
1	A	246	ARG	3.0
1	В	262	ALA	2.7
1	A	359	ASP	2.5
1	В	272	VAL	2.5

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, 95^{th} 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
2	GOL	В	1384	6/6	0.81	0.23	58,85,93,94	0
2	GOL	A	1386	6/6	0.84	0.19	43,50,73,75	0
2	GOL	A	1385	6/6	0.91	0.28	58,77,91,92	0
4	THM	A	1388	17/17	0.97	0.07	40,44,48,51	0
4	THM	В	1385	17/17	0.97	0.08	52,57,67,69	0
3	PO4	A	1387	5/5	0.99	0.11	34,47,55,56	5
5	ZN	В	1386	1/1	0.99	0.08	60,60,60,60	0
5	ZN	A	1389	1/1	1.00	0.08	51,51,51,51	0

6.5 Other polymers (i)

There are no such residues in this entry.

