



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 6, 2017 – 01:35 PM EST

PDB ID : 4JEP  
Title : Crystal structure of Toxoplasma gondii nucleoside triphosphate diphosphohydrolase 1 (NTPDase1)  
Authors : Krug, U.; Totzauer, R.; Strater, N.  
Deposited on : unknown  
Resolution : 3.10 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  
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

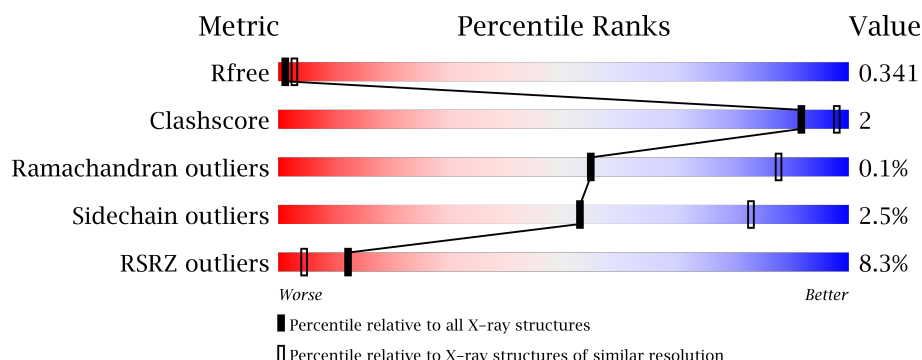
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1001 (3.12-3.08)
Clashscore	112137	1099 (3.12-3.08)
Ramachandran outliers	110173	1057 (3.12-3.08)
Sidechain outliers	110143	1057 (3.12-3.08)
RSRZ outliers	101464	1006 (3.12-3.08)

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	611	<div> <div>5%</div> <div>86%</div> <div>8%</div> <div>6%</div> </div>
1	B	611	<div> <div>10%</div> <div>83%</div> <div>7%</div> <div>10%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 8772 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 Nucleoside-triphosphatase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	574	Total	C	N	O	S	0	0	0
			4471	2805	789	853	24			
1	B	552	Total	C	N	O	S	0	0	0
			4301	2703	756	816	26			

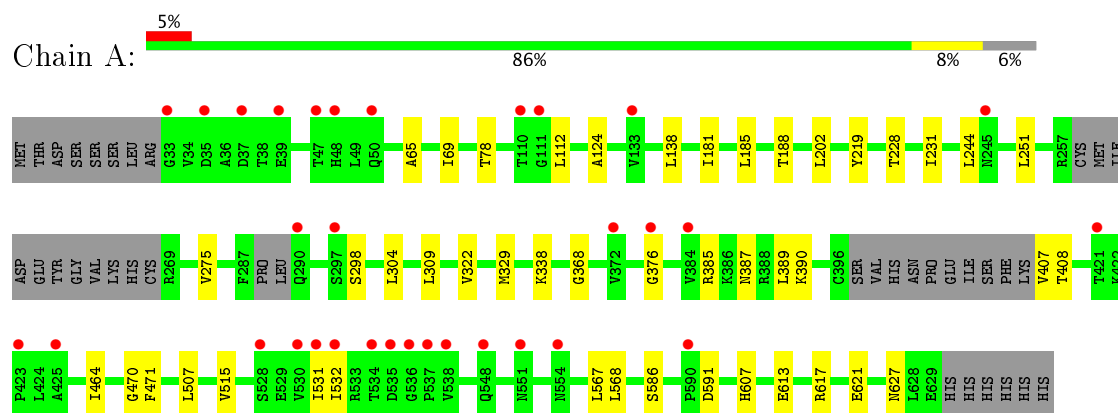
There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	25	MET	-	EXPRESSION TAG	UNP Q27895
A	629	GLU	-	EXPRESSION TAG	UNP Q27895
A	630	HIS	-	EXPRESSION TAG	UNP Q27895
A	631	HIS	-	EXPRESSION TAG	UNP Q27895
A	632	HIS	-	EXPRESSION TAG	UNP Q27895
A	633	HIS	-	EXPRESSION TAG	UNP Q27895
A	634	HIS	-	EXPRESSION TAG	UNP Q27895
A	635	HIS	-	EXPRESSION TAG	UNP Q27895
B	25	MET	-	EXPRESSION TAG	UNP Q27895
B	629	GLU	-	EXPRESSION TAG	UNP Q27895
B	630	HIS	-	EXPRESSION TAG	UNP Q27895
B	631	HIS	-	EXPRESSION TAG	UNP Q27895
B	632	HIS	-	EXPRESSION TAG	UNP Q27895
B	633	HIS	-	EXPRESSION TAG	UNP Q27895
B	634	HIS	-	EXPRESSION TAG	UNP Q27895
B	635	HIS	-	EXPRESSION TAG	UNP Q27895

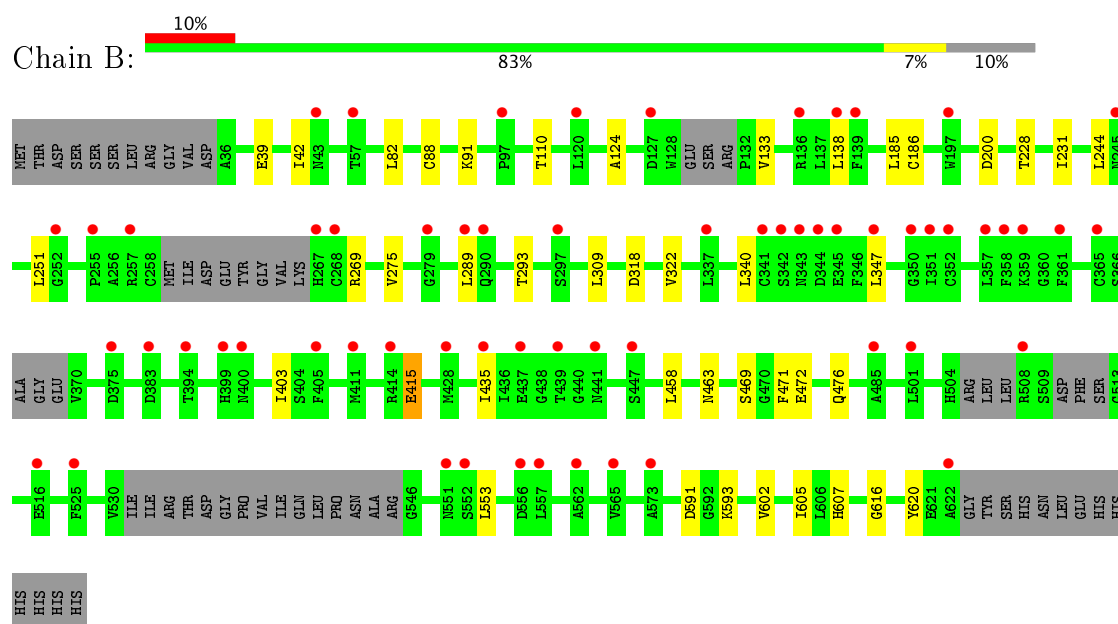
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Nucleoside-triphosphatase 2



#### • Molecule 1: Nucleoside-triphosphatase 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.74Å 161.52Å 236.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.31 – 3.10 39.49 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.4 (36.31-3.10) 99.7 (39.49-3.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.20 (at 3.12Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, $R_{free}$	0.270 , 0.305 0.299 , 0.341	Depositor DCC
$R_{free}$ test set	1305 reflections (5.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	65.1	Xtriage
Anisotropy	0.228	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 39.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.81	EDS
Total number of atoms	8772	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% 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

### 5.1 Standard geometry

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	0.36	0/4554	0.55	0/6161
1	B	0.37	0/4381	0.54	0/5922
All	All	0.37	0/8935	0.55	0/12083

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts

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	4471	0	4420	19	0
1	B	4301	0	4249	16	0
All	All	8772	0	8669	33	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 2.

All (33) 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:338:LYS:HB2	1:A:389:LEU:HD11	1.86	0.56
1:B:616:GLY:HA2	1:B:620:TYR:HD1	1.73	0.54

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:185:LEU:HB3	1:B:228:THR:HG23	1.89	0.54
1:A:69:ILE:HG12	1:A:78:THR:HG22	1.89	0.53
1:B:289:LEU:HD11	1:B:293:THR:HG21	1.90	0.53
1:A:185:LEU:HB3	1:A:228:THR:HG23	1.91	0.52
1:A:368:GLY:HA2	1:A:387:ASN:HA	1.92	0.52
1:B:322:VAL:HG11	1:B:471:PHE:HB2	1.91	0.51
1:A:124:ALA:HB2	1:A:138:LEU:HD21	1.93	0.50
1:A:407:VAL:HG22	1:B:476:GLN:HE21	1.76	0.49
1:B:244:LEU:HD22	1:B:275:VAL:HB	1.96	0.47
1:B:415:GLU:HG2	1:B:458:LEU:HD13	1.96	0.47
1:A:188:THR:HG22	1:A:231:ILE:HD11	1.95	0.46
1:B:124:ALA:HB2	1:B:138:LEU:HD21	1.97	0.46
1:B:269:ARG:HH21	1:B:318:ASP:HB2	1.81	0.46
1:B:309:LEU:HA	1:B:607:HIS:HD2	1.81	0.45
1:B:39:GLU:HA	1:B:42:ILE:HD12	1.98	0.45
1:A:244:LEU:HD22	1:A:275:VAL:HB	1.98	0.44
1:A:112:LEU:HD22	1:A:202:LEU:HD21	1.98	0.44
1:A:464:ILE:HG12	1:A:470:GLY:HA2	2.00	0.44
1:A:376:GLY:HA3	1:A:532:ILE:HG12	2.00	0.43
1:B:602:VAL:HA	1:B:605:ILE:HD12	1.99	0.43
1:A:181:ILE:HD12	1:A:219:TYR:HB3	2.01	0.43
1:B:322:VAL:HG13	1:B:469:SER:HB2	2.00	0.43
1:A:322:VAL:HG11	1:A:471:PHE:HB2	2.00	0.43
1:A:385:ARG:HA	1:A:390:LYS:HD2	2.00	0.43
1:B:88:CYS:HB2	1:B:91:LYS:HB2	2.00	0.43
1:A:617:ARG:HA	1:A:621:GLU:HB2	2.00	0.42
1:A:298:SER:HB3	1:B:463:ASN:HD22	1.85	0.41
1:A:304:LEU:HD23	1:A:309:LEU:HD13	2.02	0.41
1:A:515:VAL:HG22	1:A:568:LEU:HB3	2.03	0.41
1:B:186:CYS:HB3	1:B:231:ILE:HG12	2.03	0.40
1:A:65:ALA:HB3	1:A:181:ILE:HG12	2.04	0.40

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	566/611 (93%)	545 (96%)	20 (4%)	1 (0%)	51	84
1	B	538/611 (88%)	516 (96%)	22 (4%)	0	100	100
All	All	1104/1222 (90%)	1061 (96%)	42 (4%)	1 (0%)	55	88

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	329	MET

### 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	491/527 (93%)	481 (98%)	10 (2%)	60	86
1	B	475/527 (90%)	461 (97%)	14 (3%)	48	80
All	All	966/1054 (92%)	942 (98%)	24 (2%)	53	83

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	251	LEU
1	A	408	THR
1	A	507	LEU
1	A	531	ILE
1	A	567	LEU
1	A	586	SER
1	A	591	ASP
1	A	607	HIS
1	A	613	GLU
1	A	627	ASN
1	B	82	LEU
1	B	110	THR

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	133	VAL
1	B	200	ASP
1	B	251	LEU
1	B	340	LEU
1	B	347	LEU
1	B	403	ILE
1	B	415	GLU
1	B	435	ILE
1	B	472	GLU
1	B	553	LEU
1	B	591	ASP
1	B	593	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	101	GLN
1	B	463	ASN
1	B	476	GLN
1	B	607	HIS
1	B	608	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 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	574/611 (93%)	0.47	32 (5%) 25 11	32, 61, 99, 128	0
1	B	552/611 (90%)	0.63	61 (11%) 6 2	38, 78, 112, 173	0
All	All	1126/1222 (92%)	0.54	93 (8%) 12 4	32, 69, 107, 173	0

All (93) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	39	GLU	5.2
1	A	50	GLN	4.7
1	B	341	CYS	4.4
1	B	357	LEU	4.3
1	A	37	ASP	4.0
1	B	344	ASP	4.0
1	A	376	GLY	3.9
1	A	530	VAL	3.9
1	A	48	HIS	3.7
1	B	525	PHE	3.5
1	B	252	GLY	3.5
1	B	297	SER	3.4
1	A	421	THR	3.3
1	B	289	LEU	3.2
1	B	268	CYS	3.2
1	B	437	GLU	3.2
1	B	557	LEU	3.1
1	A	532	ILE	3.1
1	B	138	LEU	3.1
1	A	110	THR	3.1
1	A	536	GLY	3.0
1	B	255	PRO	3.0
1	A	528	SER	3.0
1	B	139	PHE	3.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	535	ASP	2.9
1	A	33	GLY	2.9
1	B	290	GLN	2.9
1	B	337	LEU	2.8
1	B	375	ASP	2.7
1	B	352	CYS	2.7
1	A	372	VAL	2.7
1	A	290	GLN	2.7
1	A	423	PRO	2.7
1	B	411	MET	2.7
1	B	359	LYS	2.7
1	B	267	HIS	2.7
1	B	358	PHE	2.6
1	B	485	ALA	2.6
1	B	361	PHE	2.6
1	B	556	ASP	2.6
1	B	197	TRP	2.6
1	B	399	HIS	2.6
1	A	133	VAL	2.5
1	B	343	ASN	2.5
1	B	136	ARG	2.5
1	A	35	ASP	2.5
1	A	551	ASN	2.5
1	B	562	ALA	2.5
1	B	552	SER	2.5
1	B	405	PHE	2.4
1	B	414	ARG	2.4
1	B	573	ALA	2.4
1	B	245	ASN	2.4
1	B	439	THR	2.4
1	A	425	ALA	2.4
1	A	537	PRO	2.4
1	B	501	LEU	2.4
1	A	245	ASN	2.4
1	B	120	LEU	2.4
1	A	538	VAL	2.4
1	B	347	LEU	2.4
1	B	394	THR	2.4
1	A	111	GLY	2.3
1	B	516	GLU	2.3
1	A	47	THR	2.3
1	B	127	ASP	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	622	ALA	2.3
1	B	365	CYS	2.3
1	B	350	GLY	2.2
1	B	342	SER	2.2
1	B	383	ASP	2.2
1	B	400	ASN	2.2
1	A	531	ILE	2.2
1	A	590	PRO	2.2
1	B	57	THR	2.2
1	B	565	VAL	2.1
1	B	257	ARG	2.1
1	B	345	GLU	2.1
1	B	43	ASN	2.1
1	A	554	ASN	2.1
1	B	508	ARG	2.1
1	A	548	GLN	2.1
1	B	441	ASN	2.1
1	A	384	VAL	2.1
1	B	279	GLY	2.1
1	B	551	ASN	2.1
1	B	351	ILE	2.0
1	B	435	ILE	2.0
1	B	447	SER	2.0
1	B	97	PRO	2.0
1	B	428	MET	2.0
1	A	297	SER	2.0
1	A	534	THR	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.