



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 7, 2023 – 10:23 AM EDT

PDB ID : 4H1I  
Title : Structure of human thymidylate synthase at low salt conditions  
Authors : Brunn, N.; Dibrov, S.; Hermann, T.  
Deposited on : 2012-09-10  
Resolution : 3.10 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](mailto: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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

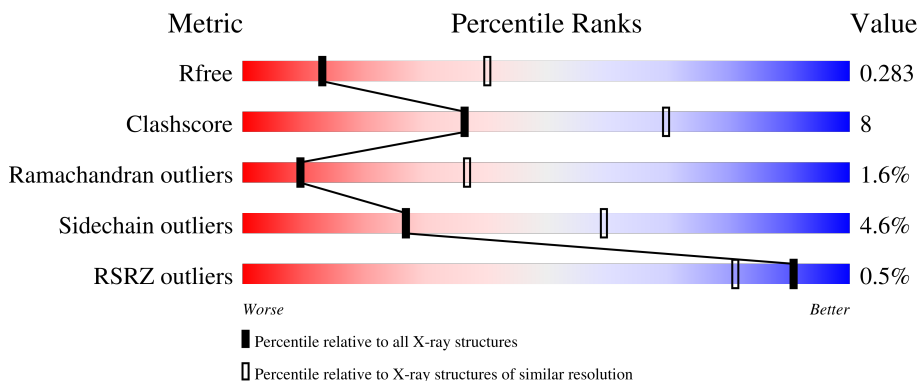
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
$R_{free}$	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	318	 70% 17% 12%
1	B	318	 63% 24% 12%
1	C	318	 71% 15% 12%
1	D	318	 68% 19% 12%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9121 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 synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	281	2270	1451	397	411	11	0	0	0
1	B	281	2270	1451	397	411	11	0	0	0
1	C	281	2270	1451	397	411	11	0	0	0
1	D	281	2270	1451	397	411	11	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	314	HIS	-	expression tag	UNP P04818
A	315	HIS	-	expression tag	UNP P04818
A	316	HIS	-	expression tag	UNP P04818
A	317	HIS	-	expression tag	UNP P04818
A	318	HIS	-	expression tag	UNP P04818
B	314	HIS	-	expression tag	UNP P04818
B	315	HIS	-	expression tag	UNP P04818
B	316	HIS	-	expression tag	UNP P04818
B	317	HIS	-	expression tag	UNP P04818
B	318	HIS	-	expression tag	UNP P04818
C	314	HIS	-	expression tag	UNP P04818
C	315	HIS	-	expression tag	UNP P04818
C	316	HIS	-	expression tag	UNP P04818
C	317	HIS	-	expression tag	UNP P04818
C	318	HIS	-	expression tag	UNP P04818
D	314	HIS	-	expression tag	UNP P04818
D	315	HIS	-	expression tag	UNP P04818
D	316	HIS	-	expression tag	UNP P04818
D	317	HIS	-	expression tag	UNP P04818
D	318	HIS	-	expression tag	UNP P04818

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).

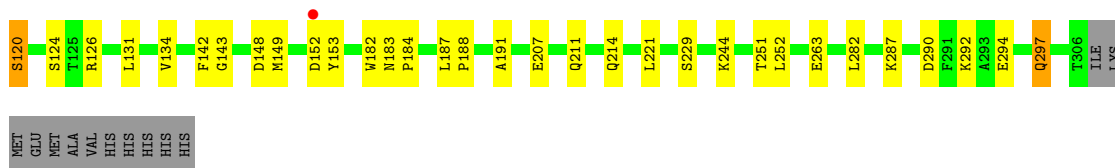


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	C	1	Total O S 5 4 1	0	0
2	D	1	Total O S 5 4 1	0	0
2	D	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

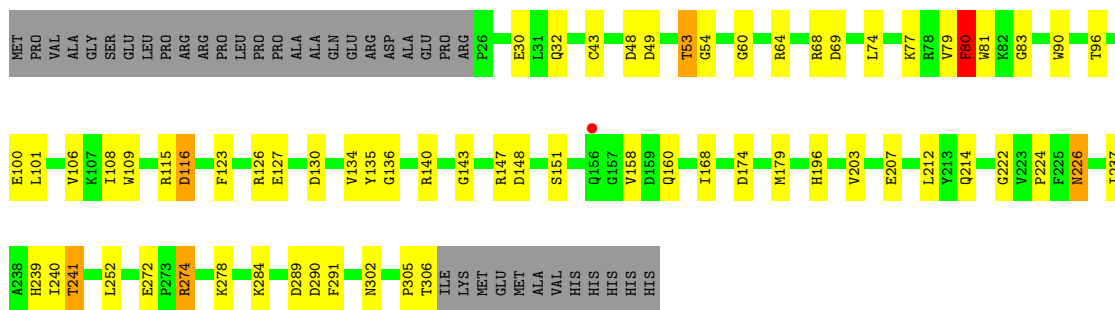
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total O 2 2	0	0
3	B	5	Total O 5 5	0	0
3	C	7	Total O 7 7	0	0
3	D	7	Total O 7 7	0	0





• Molecule 1: Thymidylate synthase

Chain D: 68% 19% 12%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	157.85Å 94.93Å 131.44Å 90.00° 122.75° 90.00°	Depositor
Resolution (Å)	19.89 – 3.10 19.95 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.6 (19.89-3.10) 94.3 (19.95-3.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.89 (at 3.09Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.232 , 0.285 0.229 , 0.283	Depositor DCC
$R_{free}$ test set	1509 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	50.0	Xtrriage
Anisotropy	0.289	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 46.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	9121	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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: SO4

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	0.22	0/2330	0.42	0/3153
1	B	0.23	0/2330	0.44	0/3153
1	C	0.22	0/2330	0.43	0/3153
1	D	0.23	0/2330	0.43	0/3153
All	All	0.23	0/9320	0.43	0/12612

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 [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	2270	0	2235	30	0
1	B	2270	0	2235	54	0
1	C	2270	0	2235	28	0
1	D	2270	0	2235	42	0
2	A	5	0	0	0	0
2	C	5	0	0	0	0
2	D	10	0	0	0	0
3	A	2	0	0	0	0
3	B	5	0	0	1	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	7	0	0	0	0
3	D	7	0	0	2	0
All	All	9121	0	8940	143	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 8.

The worst 5 of 143 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:D:90:TRP:HB2	1:D:101:LEU:HD13	1.66	0.77
1:A:162:GLN:OE1	1:A:287:LYS:NZ	2.18	0.77
1:C:115:ARG:HH22	1:C:124:SER:HA	1.50	0.76
1:D:116:ASP:OD1	1:D:116:ASP:N	2.15	0.70
1:A:41:LEU:O	1:B:46:ARG:NH2	2.25	0.69

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	279/318 (88%)	264 (95%)	13 (5%)	2 (1%)	22 57
1	B	279/318 (88%)	257 (92%)	15 (5%)	7 (2%)	5 27
1	C	279/318 (88%)	256 (92%)	19 (7%)	4 (1%)	11 40
1	D	279/318 (88%)	249 (89%)	25 (9%)	5 (2%)	8 34
All	All	1116/1272 (88%)	1026 (92%)	72 (6%)	18 (2%)	9 37

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	108	ILE
1	B	128	GLU
1	D	135	TYR
1	D	305	PRO
1	C	191	ALA

### 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	245/276 (89%)	236 (96%)	9 (4%)	34	66
1	B	245/276 (89%)	233 (95%)	12 (5%)	25	57
1	C	245/276 (89%)	233 (95%)	12 (5%)	25	57
1	D	245/276 (89%)	233 (95%)	12 (5%)	25	57
All	All	980/1104 (89%)	935 (95%)	45 (5%)	27	59

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

Mol	Chain	Res	Type
1	C	221	LEU
1	D	96	THR
1	C	229	SER
1	C	297	GLN
1	D	116	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	268	GLN
1	C	211	GLN
1	D	211	GLN

### 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](#)

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	$\# Z  > 2$
2	SO4	D	401	-	4,4,4	0.14	0	6,6,6	0.06	0
2	SO4	D	402	-	4,4,4	0.15	0	6,6,6	0.04	0
2	SO4	A	401	-	4,4,4	0.14	0	6,6,6	0.05	0
2	SO4	C	401	-	4,4,4	0.14	0	6,6,6	0.05	0

There are no bond length outliers.

There are no bond angle outliers.

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 [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<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	281/318 (88%)	-0.53	2 (0%) 87 75	17, 39, 74, 99	0
1	B	281/318 (88%)	-0.40	2 (0%) 87 75	16, 47, 93, 115	0
1	C	281/318 (88%)	-0.50	1 (0%) 92 84	23, 47, 75, 98	0
1	D	281/318 (88%)	-0.40	1 (0%) 92 84	22, 49, 90, 114	0
All	All	1124/1272 (88%)	-0.46	6 (0%) 91 81	16, 45, 82, 115	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	124	SER	2.9
1	D	156	GLN	2.2
1	A	306	THR	2.1
1	C	152	ASP	2.1
1	B	136	GLY	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 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<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	B-factors( $\text{\AA}^2$ )	Q<0.9
2	SO4	A	401	5/5	0.98	0.09	55,57,58,58	0
2	SO4	C	401	5/5	0.98	0.14	52,53,54,54	0
2	SO4	D	401	5/5	0.98	0.08	50,51,52,54	0
2	SO4	D	402	5/5	0.98	0.09	59,60,60,61	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.