



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 01:30 pm BST

PDB ID : 3OSG  
Title : The structure of protozoan parasite *Trichomonas vaginalis* Myb2 in complex with MRE-1-12 DNA  
Authors : Jiang, I.; Tsai, C.K.; Chen, S.C.; Wang, S.H.; Amiraslanov, I.; Chang, C.F.; Wu, W.J.; Tai, J.H.; Liaw, Y.C.; Huang, T.H.  
Deposited on : 2010-09-09  
Resolution : 2.00 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

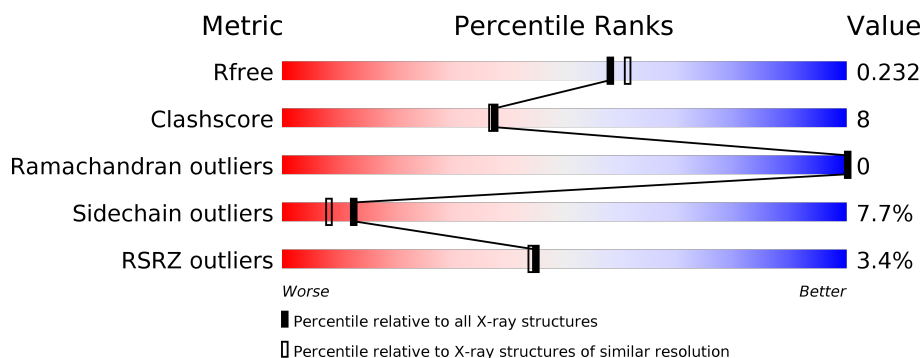
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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	126	<div> <div>3%</div> <div> <div></div> <div>70%</div> <div>13%</div> <div>• •</div> <div>13%</div> </div> </div>
1	D	126	<div> <div>4%</div> <div> <div></div> <div>63%</div> <div>16%</div> <div>•</div> <div>18%</div> </div> </div>
2	B	12	<div> <div></div> <div> <div>42%</div> <div>50%</div> <div>8%</div> </div> </div>
2	E	12	<div> <div></div> <div> <div>58%</div> <div>25%</div> <div>17%</div> </div> </div>
3	C	12	<div> <div></div> <div> <div>50%</div> <div>42%</div> <div>8%</div> </div> </div>
3	F	12	<div> <div></div> <div> <div>58%</div> <div>25%</div> <div>17%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3034 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 MYB21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	110	Total	C	N	O	S	0	0	0
			906	577	169	157	3			
1	D	103	Total	C	N	O	S	0	0	0
			856	546	159	148	3			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	39	MET	-	EXPRESSION TAG	UNP Q58HP3
A	157	LEU	-	EXPRESSION TAG	UNP Q58HP3
A	158	GLU	-	EXPRESSION TAG	UNP Q58HP3
A	159	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	160	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	161	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	162	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	163	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	164	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	39	MET	-	EXPRESSION TAG	UNP Q58HP3
D	157	LEU	-	EXPRESSION TAG	UNP Q58HP3
D	158	GLU	-	EXPRESSION TAG	UNP Q58HP3
D	159	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	160	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	161	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	162	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	163	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	164	HIS	-	EXPRESSION TAG	UNP Q58HP3

- Molecule 2 is a DNA chain called 5'-D(\*AP\*AP\*AP\*TP\*AP\*TP\*CP\*GP\*TP\*TP\*AP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	12	Total	C	N	O	P	0	0	0
			243	119	43	70	11			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	12	Total	C	N	O	P	0	0	0
			243	119	43	70	11			

- Molecule 3 is a DNA chain called 5'-D(\*AP\*TP\*AP\*AP\*CP\*GP\*AP\*TP\*AP\*TP\*TP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total	C	N	O	P	0	0	0
			243	119	43	70	11			
3	F	12	Total	C	N	O	P	0	0	0
			243	119	43	70	11			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	83	Total	O	0	0
			83	83		
4	B	44	Total	O	0	0
			44	44		
4	C	42	Total	O	0	0
			42	42		
4	D	82	Total	O	0	0
			82	82		
4	E	24	Total	O	0	0
			24	24		
4	F	25	Total	O	0	0
			25	25		

**i**

- Molecule 1: MYB21



- [illegible]

- 
- A1  
A2  
A3
- T6  
C7  
G8
- T9  
T10  
A11  
T12

- 

-

- Molecule 3: 5'-D(\*AP\*TP\*AP\*AP\*CP\*GP\*AP\*TP\*AP\*TP\*TP\*T)-3'

Chain F:  58% 25% 17%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	40.12Å 127.53Å 41.47Å 90.00° 100.23° 90.00°	Depositor
Resolution (Å)	24.19 – 2.00 24.19 – 2.00	Depositor EDS
% Data completeness (in resolution range)	93.8 (24.19-2.00) 93.9 (24.19-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.59 (at 1.99Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6 _289)	Depositor
R, $R_{free}$	0.191 , 0.233 0.186 , 0.232	Depositor DCC
$R_{free}$ test set	1339 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.5	Xtriage
Anisotropy	0.531	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 50.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.041 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3034	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.08% 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](#)

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	0.41	0/928	0.58	0/1252
1	D	0.41	0/879	0.56	0/1187
2	B	0.83	0/272	1.61	6/418 (1.4%)
2	E	0.69	0/272	1.48	7/418 (1.7%)
3	C	0.84	0/272	1.60	7/418 (1.7%)
3	F	0.74	0/272	1.61	7/418 (1.7%)
All	All	0.58	0/2895	1.10	27/4111 (0.7%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	10	DT	C1'-O4'-C4'	-7.38	102.72	110.10
2	B	10	DT	O4'-C1'-N1	-7.09	103.03	108.00
2	B	3	DA	O4'-C1'-N9	-6.95	103.14	108.00
3	C	9	DA	O4'-C1'-N9	-6.78	103.26	108.00
3	F	2	DT	O4'-C1'-C2'	-6.16	100.97	105.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	40	VAL	Peptide



## 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	906	0	909	18	0
1	D	856	0	855	18	0
2	B	243	0	139	2	0
2	E	243	0	139	2	0
3	C	243	0	139	1	0
3	F	243	0	139	1	0
4	A	83	0	0	4	0
4	B	44	0	0	0	0
4	C	42	0	0	0	0
4	D	82	0	0	0	0
4	E	24	0	0	1	0
4	F	25	0	0	0	0
All	All	3034	0	2320	39	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 39 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:129:PHE:HB3	1:A:132:ARG:HG3	1.55	0.87
1:D:48:LYS:HB3	1:D:48:LYS:NZ	1.95	0.81
1:A:42:VAL:O	1:A:43:ASN:HB2	1.93	0.68
3:F:8:DT:H2"	3:F:9:DA:C8	2.30	0.67
1:D:70:ASP:OD2	1:D:73:MET:HE1	1.96	0.66

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	106/126 (84%)	102 (96%)	4 (4%)	0	100	100
1	D	101/126 (80%)	97 (96%)	4 (4%)	0	100	100
All	All	207/252 (82%)	199 (96%)	8 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	94/110 (86%)	86 (92%)	8 (8%)	10	6
1	D	89/110 (81%)	83 (93%)	6 (7%)	16	11
All	All	183/220 (83%)	169 (92%)	14 (8%)	13	8

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	84	ARG
1	A	132	ARG
1	D	73	MET
1	A	60	LEU
1	D	60	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	136	HIS
1	D	136	HIS
1	D	67	HIS
1	A	113	GLN
1	D	50	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 carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

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 [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	110/126 (87%)	0.37	4 (3%) 42 42	21, 34, 47, 54	0
1	D	103/126 (81%)	0.38	5 (4%) 29 28	19, 34, 48, 61	0
2	B	12/12 (100%)	-0.45	0 100 100	20, 26, 32, 33	0
2	E	12/12 (100%)	-0.20	0 100 100	25, 35, 40, 41	0
3	C	12/12 (100%)	-0.64	0 100 100	19, 24, 29, 30	0
3	F	12/12 (100%)	-0.42	0 100 100	31, 34, 38, 39	0
All	All	261/300 (87%)	0.22	9 (3%) 45 44	19, 34, 47, 61	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	150	ILE	5.2
1	A	135	ILE	4.3
1	A	150	ILE	4.0
1	D	149	GLY	3.6
1	D	135	ILE	3.2

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.