Full wwPDB X-ray Structure Validation Report (i)

Oct 15, 2019 – 08:30 am BST

PDB ID	:	6T4R
Title	:	Crystal structure of Trypanosoma brucei Morn1
Deposited on	:	2019-10-14
Resolution	:	2.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

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 following versions of software and data (see references (1)) were used in the production of this report:

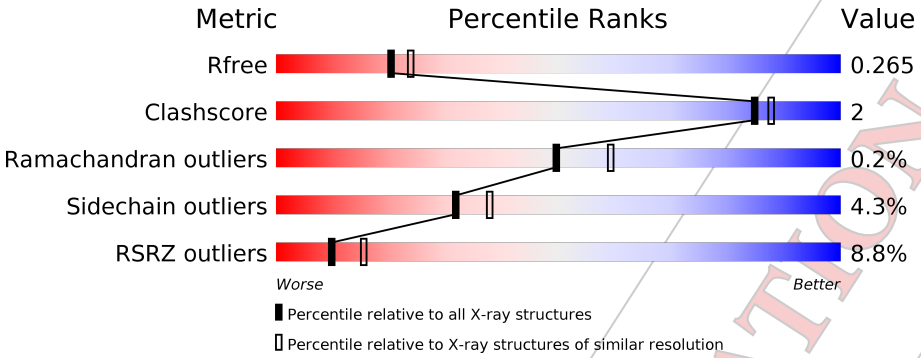
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.5
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.5

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 2.35 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}	111664	1015 (2.36-2.36)
Clashscore	122126	1081 (2.36-2.36)
Ramachandran outliers	120053	1066 (2.36-2.36)
Sidechain outliers	120020	1067 (2.36-2.36)
RSRZ outliers	108989	1002 (2.36-2.36)

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.

Mol	Chain	Length	Quality of chain
1	AAA	216	13% 87% 11%
1	CCC	216	4% 93% 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6556 atoms, of which 3078 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 MORN repeat-containing protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	AAA	212	3256	1079	1547	296	329	5	74	0	0
			Total	C	H	N	O	S			
1	CCC	210	3226	1070	1531	294	326	5	73	0	0

- Molecule 2 is water.

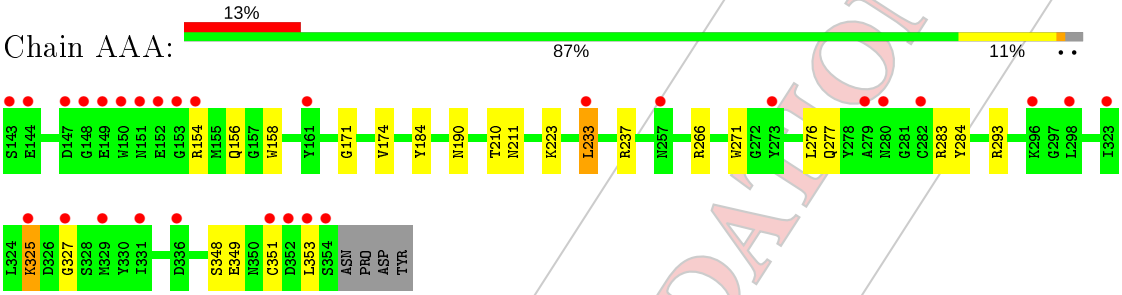
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	AAA	25	25	25	0	0
			Total	O		
2	CCC	49	49	49	0	0

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 the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometryand 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.

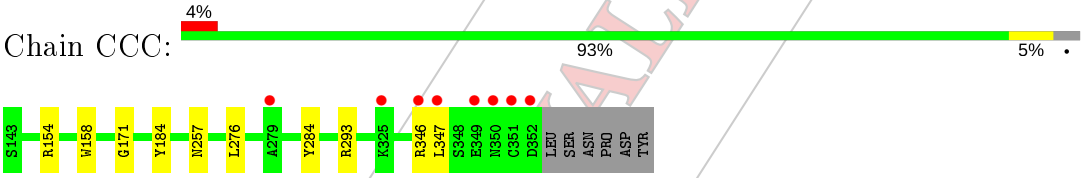
- Molecule 1: MORN repeat-containing protein 1

Chain AAA:



- Molecule 1: MORN repeat-containing protein 1

Chain CCC:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	69.04Å 27.63Å 114.54Å 90.00° 101.83° 90.00°	Depositor
Resolution (Å)	48.28 – 2.35 48.28 – 2.35	Depositor EDS
% Data completeness (in resolution range)	95.2 (48.28-2.35) 95.3 (48.28-2.35)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.10 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.221 , 0.259 0.225 , 0.265	Depositor DCC
R_{free} test set	818 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	31.4	Xtriage
Anisotropy	0.463	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 45.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6556	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.59% 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, 0.333 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	AAA	0.62	0/1763	0.76	0/2375
1	CCC	0.61	0/1749	0.75	0/2356
All	All	0.62	0/3512	0.75	0/4731

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	AAA	1709	1547	1536	14	0
1	CCC	1695	1531	1520	2	0
2	AAA	25	0	0	0	0
2	CCC	49	0	0	0	0
All	All	3478	3078	3056	16	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 (16) 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:AAA:211:ASN:O	1:AAA:233:LEU:CD1	2.51	0.58

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:AAA:233:LEU:N	1:AAA:233:LEU:HD22	2.20	0.56
1:CCC:284:TYR:CD1	1:CCC:293:ARG:HD2	2.44	0.53
1:AAA:284:TYR:CD1	1:AAA:293:ARG:HD2	2.44	0.52
1:AAA:351:CYS:SG	1:AAA:351:CYS:O	2.69	0.50
1:AAA:233:LEU:H	1:AAA:233:LEU:HD22	1.78	0.49
1:AAA:277:GLN:HE22	1:AAA:283:ARG:HD3	1.79	0.47
1:AAA:327:GLY:C	1:AAA:348:SER:HB2	2.36	0.46
1:AAA:190:ASN:ND2	1:AAA:210:THR:H	2.12	0.46
1:AAA:353:LEU:HD12	1:AAA:353:LEU:C	2.36	0.46
1:AAA:349:GLU:HA	1:AAA:353:LEU:HD11	2.00	0.43
1:AAA:325:LYS:C	1:AAA:327:GLY:N	2.73	0.41
1:CCC:158:TRP:CH2	1:CCC:171:GLY:HA2	2.55	0.41
1:AAA:156:GLN:HG3	1:AAA:174:VAL:HA	2.03	0.41
1:AAA:158:TRP:CH2	1:AAA:171:GLY:HA2	2.56	0.41
1:AAA:211:ASN:O	1:AAA:233:LEU:HD12	2.18	0.41

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	AAA	210/216 (97%)	199 (95%)	10 (5%)	1 (0%)	31 34
1	CCC	208/216 (96%)	201 (97%)	7 (3%)	0	100 100
All	All	418/432 (97%)	400 (96%)	17 (4%)	1 (0%)	49 59

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	325	LYS

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	AAA	165/169 (98%)	157 (95%)	8 (5%)	28 33
1	CCC	163/169 (96%)	157 (96%)	6 (4%)	37 45
All	All	328/338 (97%)	314 (96%)	14 (4%)	32 38

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	154	ARG
1	AAA	184	TYR
1	AAA	223	LYS
1	AAA	233	LEU
1	AAA	237	ARG
1	AAA	266	ARG
1	AAA	271	TRP
1	AAA	276	LEU
1	CCC	154	ARG
1	CCC	184	TYR
1	CCC	257	ASN
1	CCC	276	LEU
1	CCC	346	ARG
1	CCC	347	LEU

Some 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 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

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^{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	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	AAA	212/216 (98%)	0.98	29 (13%) 3 4	19, 40, 83, 99	0
1	CCC	210/216 (97%)	0.61	8 (3%) 40 54	15, 28, 65, 103	0
All	All	422/432 (97%)	0.80	37 (8%) 10 15	15, 34, 75, 103	0

All (37) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	153	GLY	8.3
1	AAA	150	TRP	4.1
1	AAA	148	GLY	4.1
1	CCC	351	CYS	4.0
1	CCC	347	LEU	3.9
1	AAA	149	GLU	3.8
1	AAA	352	ASP	3.7
1	AAA	353	LEU	3.6
1	CCC	346	ARG	3.4
1	AAA	152	GLU	3.4
1	AAA	143	SER	3.4
1	AAA	257	ASN	3.4
1	AAA	144	GLU	3.3
1	AAA	329	MET	3.2
1	AAA	325	LYS	3.0
1	AAA	151	ASN	2.9
1	CCC	352	ASP	2.8
1	AAA	298	LEU	2.8
1	AAA	147	ASP	2.7
1	AAA	233	LEU	2.6
1	AAA	280	ASN	2.6
1	AAA	273	TYR	2.5
1	AAA	351	CYS	2.5
1	AAA	336	ASP	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	AAA	327	GLY	2.4
1	AAA	282	CYS	2.4
1	AAA	296	LYS	2.4
1	AAA	279	ALA	2.3
1	AAA	154	ARG	2.3
1	AAA	354	SER	2.3
1	CCC	279	ALA	2.2
1	AAA	161	TYR	2.2
1	AAA	331	ILE	2.2
1	CCC	325	LYS	2.1
1	CCC	349	GLU	2.0
1	AAA	323	ILE	2.0
1	CCC	350	ASN	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 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.