



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 6, 2023 – 10:19 am BST

PDB ID : 8Q0P
Title : Crystal Structure of an N-terminal Domain of Variant Surface Glycoprotein 21 (VSG21) of Trypanosome brucei brucei Lister 427
Authors : Zeelen, J.P.; Stebbins, C.E.
Deposited on : 2023-07-28
Resolution : 2.01 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 ⓘ 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
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

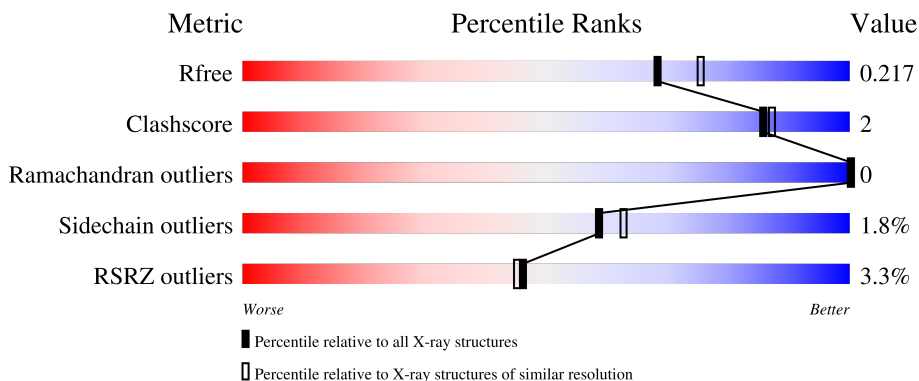
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.01 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 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 $\leq 5\%$. 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	476	 2% 47% 51%
1	B	476	 2% 46% 52%
1	C	476	 46% 51%
1	D	476	 2% 45% 51%
1	E	476	 2% 45% 51%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	476	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a small red segment at the beginning labeled '3%', a large green segment labeled '44%', a small yellow segment, and a large grey segment at the end labeled '51%'. A small black dot is located on the boundary between the yellow and grey segments.</p>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 10970 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 Variant surface glycoprotein MITat 1.21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	235	1677	1037	290	345	5	0	0	0
1	B	230	1650	1022	286	337	5	0	0	0
1	C	234	1701	1051	294	351	5	0	3	0
1	D	232	1678	1039	289	345	5	0	3	0
1	E	235	1698	1050	293	350	5	0	2	0
1	F	232	1675	1036	290	344	5	0	2	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	167	Total 167	O 167	0	0
2	B	156	Total 156	O 156	0	0
2	C	166	Total 166	O 166	0	0
2	D	154	Total 154	O 154	0	0
2	E	126	Total 126	O 126	0	0
2	F	122	Total 122	O 122	0	0

CYS
LYS
LYS
HIS
THR
THR
SER
ASP
ASP
CYS
LYS
LYS
ASP
GLU
LYS
GLY
CYS
CYS
GLU
PHE
GLU
GLU
LYS
LYS
ASP
PRO
LYS
CYS
PHE
PRO
PHE
LYS
VAL
VAL
ASP
THR
GLU
LYS
LYS
ASP
LYS
LYS
SER
PHE
SER
SER
ASN
VAL
VAL
ARG
VAL
VAL
SER
VAL
VAL
GLN
VAL
PHE
ALA
ALA
VAL
LEU
VAL
LEU
ALA
PHE

● Molecule 1: Variant surface glycoprotein MITat 1.21

Chain F: 

MET
ALA
ASN
TYR
TRP
LEU
ILE
SER
ALA
ALA
THR
GLY
GLY
ALA
LEU
LEU
LEU
LEU
ALA
VAL
VAL
THR
THR
GLN
GLY
LYS
T23
T24
L24
L25
A28
K29
K29
P35
P35
E50
E50
L57
L57
L64
L64
GLU
THR
THR
SER
ASN
LYS
LYS
LEU
LEU
ARG
VAL
VAL
VAL
ALA
SER
SER
SER
LYS
LYS
THR
THR
GLY
GLY
ALA
ALA
THR
THR
ALA
ALA
SER
SER
ILE
LEU

ALA
ALA
ASN
LEU
SER
S94
E95
A96
I100
E101
I102
I103
I111
L111
A112
G113
G113
A118
A118
S128
E129
I130
I130
I153
I153
D170
D171
D171
V175
SER
ASP
ASP
GLU
ALA
ALA
GLN
GLN
PRO
PRO
ASN
GLN
G186
G186
H213
H213
T218
T218
K240
K240
S251
S252
A253
G254
L282
L282
V286
V286
L289
L289
D290
A291
A291

I292
S293
ASP
VAL
VAL
SER
SER
PRO
GLU
GLU
ASP
ILE
LYS
LYS
LEU
ALA
ALA
ARG
PRO
GLU
ILE
GLY
GLY
ASP
LYS
LYS
ALA
ALA
VAL
VAL
VAL
GLN
GLU
THR
THR
GLN
GLN
GLY
LYS
LYS
GLY
LYS
GLU
SER
ASN
GLN
ALA
ILE
VAL
VAL
LYS
LYS
THR
THR
ILE
VAL
VAL
LYS
LYS
ASP
LEU
LEU
LYS
LYS
ASP
LEU
LEU
LEU
LEU
PRO
PRO
PRO

LYS
SER
ALA
VAL
VAL
GLY
SER
GLY
GLY
GLU
GLU
LYS
LYS
LEU
LEU
THR
THR
ILE
ASN
ASN
ASP
GLU
GLU
LYS
LYS
GLY
GLY
LEU
ALA
LYS
ASP
ALA
ALA
GLN
ILE
TYR
TYR
THR
THR
VAL
VAL
LYS
LYS
PHE
PHE
SER
PHE
VAL
VAL
ASP
GLU
GLN
GLU
GLU
GLN
GLY
LYS
LYS
ASP
LYS
ASN
GLN
ALA
ILE
VAL
VAL
SER
SER
SER
ASN
VAL
VAL
ARG
VAL
THR
THR
ASN
LYS
ASP
LEU
LEU
LYS
LYS
THR
THR
GLU
GLU
PRO
PRO
ALA

LYS
SER
ALA
ASP
GLU
CYS
LYS
LYS
HIS
THR
THR
THR
SER
SER
ASP
ASP
CYS
LYS
ASP
GLU
GLU
LYS
GLY
CYS
GLU
PHE
PHE
ASP
GLU
LYS
LYS
ASP
PRO
PRO
CYS
PHE
PHE
THR
THR
GLU
LYS
LYS
ASP
PRO
PRO
CYS
PHE
PHE
SER
SER
PHE
VAL
VAL
VAL
VAL
GLN
GLN
VAL
PHE
PHE
ALA
ALA
LEU

VAL
LEU
ALA
ALA
PHE

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	174.33Å 84.42Å 120.09Å 90.00° 100.89° 90.00°	Depositor
Resolution (Å)	48.53 – 2.01 48.53 – 2.01	Depositor EDS
% Data completeness (in resolution range)	97.7 (48.53-2.01) 97.7 (48.53-2.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.45 (at 2.01Å)	Xtrriage
Refinement program	PHENIX (1.19.2_4158: ???)	Depositor
R, R_{free}	0.179 , 0.219 0.178 , 0.217	Depositor DCC
R_{free} test set	5579 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	38.0	Xtrriage
Anisotropy	0.285	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 45.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	10970	wwPDB-VP
Average B, all atoms (Å ²)	51.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.00% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.34	0/1696	0.51	0/2312
1	B	0.35	0/1669	0.53	0/2273
1	C	0.33	0/1720	0.52	0/2343
1	D	0.34	0/1697	0.51	0/2316
1	E	0.29	0/1717	0.51	0/2339
1	F	0.33	0/1694	0.51	0/2308
All	All	0.33	0/10193	0.52	0/13891

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	1677	0	1687	8	0
1	B	1650	0	1670	6	0
1	C	1701	0	1714	7	0
1	D	1678	0	1687	8	0
1	E	1698	0	1712	14	0
1	F	1675	0	1686	14	0
2	A	167	0	0	0	0
2	B	156	0	0	0	0
2	C	166	0	0	0	0
2	D	154	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	126	0	0	0	0
2	F	122	0	0	0	0
All	All	10970	0	10156	45	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.

The worst 5 of 45 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:C:111:LEU:HD12	1:D:50:GLU:HG3	1.74	0.70
1:E:100:ILE:HD11	1:F:100:ILE:HD11	1.75	0.67
1:F:113:GLY:HA3	1:F:282:LEU:HB2	1.78	0.66
1:A:50:GLU:HG3	1:B:111:LEU:HD12	1.79	0.64
1:B:25:LEU:HD21	1:B:286:VAL:HG21	1.81	0.62

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	229/476 (48%)	224 (98%)	5 (2%)	0	100	100
1	B	224/476 (47%)	215 (96%)	9 (4%)	0	100	100
1	C	231/476 (48%)	228 (99%)	3 (1%)	0	100	100
1	D	229/476 (48%)	226 (99%)	3 (1%)	0	100	100
1	E	231/476 (48%)	227 (98%)	4 (2%)	0	100	100
1	F	228/476 (48%)	226 (99%)	2 (1%)	0	100	100
All	All	1372/2856 (48%)	1346 (98%)	26 (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	179/376 (48%)	176 (98%)	3 (2%)	60	65
1	B	176/376 (47%)	172 (98%)	4 (2%)	50	53
1	C	183/376 (49%)	182 (100%)	1 (0%)	88	92
1	D	179/376 (48%)	174 (97%)	5 (3%)	43	44
1	E	182/376 (48%)	181 (100%)	1 (0%)	88	92
1	F	179/376 (48%)	174 (97%)	5 (3%)	43	44
All	All	1078/2256 (48%)	1059 (98%)	19 (2%)	59	63

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	29	LYS
1	F	240	LYS
1	F	292	ILE
1	F	171	ASP
1	D	58	LYS

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](#)

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, 95th 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	A	235/476 (49%)	-0.19	6 (2%) 56 54	29, 42, 90, 114	0
1	B	230/476 (48%)	-0.22	4 (1%) 70 68	30, 43, 85, 115	0
1	C	234/476 (49%)	-0.17	2 (0%) 84 83	32, 43, 79, 107	0
1	D	232/476 (48%)	-0.11	10 (4%) 35 34	31, 44, 93, 116	0
1	E	235/476 (49%)	0.09	11 (4%) 31 30	32, 47, 109, 144	0
1	F	232/476 (48%)	0.13	13 (5%) 24 23	32, 47, 101, 116	0
All	All	1398/2856 (48%)	-0.08	46 (3%) 46 45	29, 44, 95, 144	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	253	ALA	5.6
1	A	294	ASP	4.5
1	D	253	ALA	4.3
1	A	176	SER	4.1
1	F	28	ALA	4.0

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.