Full wwPDB X-ray Structure Validation Report ⓘ

Jun 24, 2024 – 02:49 PM EDT

PDB ID	:	5OW4
Title	:	Crystal structure of a protease-resistant fragment of the Trypanosoma cruzi gamete fusion protein HAP2 ectodomain
Authors	:	Fedry, J.; Rey, F.A.; Krey, T.
Deposited on	:	2017-08-30
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*
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:

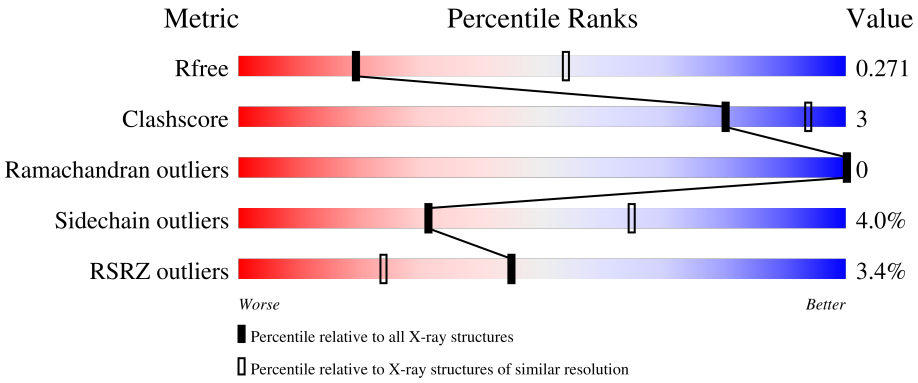
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 >=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 <=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	591	% 35% 5% 60%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1857 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 Uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	237	1855	1174	328	336	17	0	0	0

There are 41 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	ARG	-	expression tag	UNP Q4DKJ2
A	25	SER	-	expression tag	UNP Q4DKJ2
A	576	GLY	-	expression tag	UNP Q4DKJ2
A	577	PRO	-	expression tag	UNP Q4DKJ2
A	578	PHE	-	expression tag	UNP Q4DKJ2
A	579	GLU	-	expression tag	UNP Q4DKJ2
A	580	ASP	-	expression tag	UNP Q4DKJ2
A	581	ASP	-	expression tag	UNP Q4DKJ2
A	582	ASP	-	expression tag	UNP Q4DKJ2
A	583	ASP	-	expression tag	UNP Q4DKJ2
A	584	LYS	-	expression tag	UNP Q4DKJ2
A	585	ALA	-	expression tag	UNP Q4DKJ2
A	586	GLY	-	expression tag	UNP Q4DKJ2
A	587	TRP	-	expression tag	UNP Q4DKJ2
A	588	SER	-	expression tag	UNP Q4DKJ2
A	589	HIS	-	expression tag	UNP Q4DKJ2
A	590	PRO	-	expression tag	UNP Q4DKJ2
A	591	GLN	-	expression tag	UNP Q4DKJ2
A	592	PHE	-	expression tag	UNP Q4DKJ2
A	593	GLU	-	expression tag	UNP Q4DKJ2
A	594	LYS	-	expression tag	UNP Q4DKJ2
A	595	GLY	-	expression tag	UNP Q4DKJ2
A	596	GLY	-	expression tag	UNP Q4DKJ2
A	597	GLY	-	expression tag	UNP Q4DKJ2
A	598	SER	-	expression tag	UNP Q4DKJ2
A	599	GLY	-	expression tag	UNP Q4DKJ2
A	600	GLY	-	expression tag	UNP Q4DKJ2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	601	GLY	-	expression tag	UNP Q4DKJ2
A	602	SER	-	expression tag	UNP Q4DKJ2
A	603	GLY	-	expression tag	UNP Q4DKJ2
A	604	GLY	-	expression tag	UNP Q4DKJ2
A	605	GLY	-	expression tag	UNP Q4DKJ2
A	606	SER	-	expression tag	UNP Q4DKJ2
A	607	TRP	-	expression tag	UNP Q4DKJ2
A	608	SER	-	expression tag	UNP Q4DKJ2
A	609	HIS	-	expression tag	UNP Q4DKJ2
A	610	PRO	-	expression tag	UNP Q4DKJ2
A	611	GLN	-	expression tag	UNP Q4DKJ2
A	612	PHE	-	expression tag	UNP Q4DKJ2
A	613	GLU	-	expression tag	UNP Q4DKJ2
A	614	LYS	-	expression tag	UNP Q4DKJ2

- Molecule 2 is water.

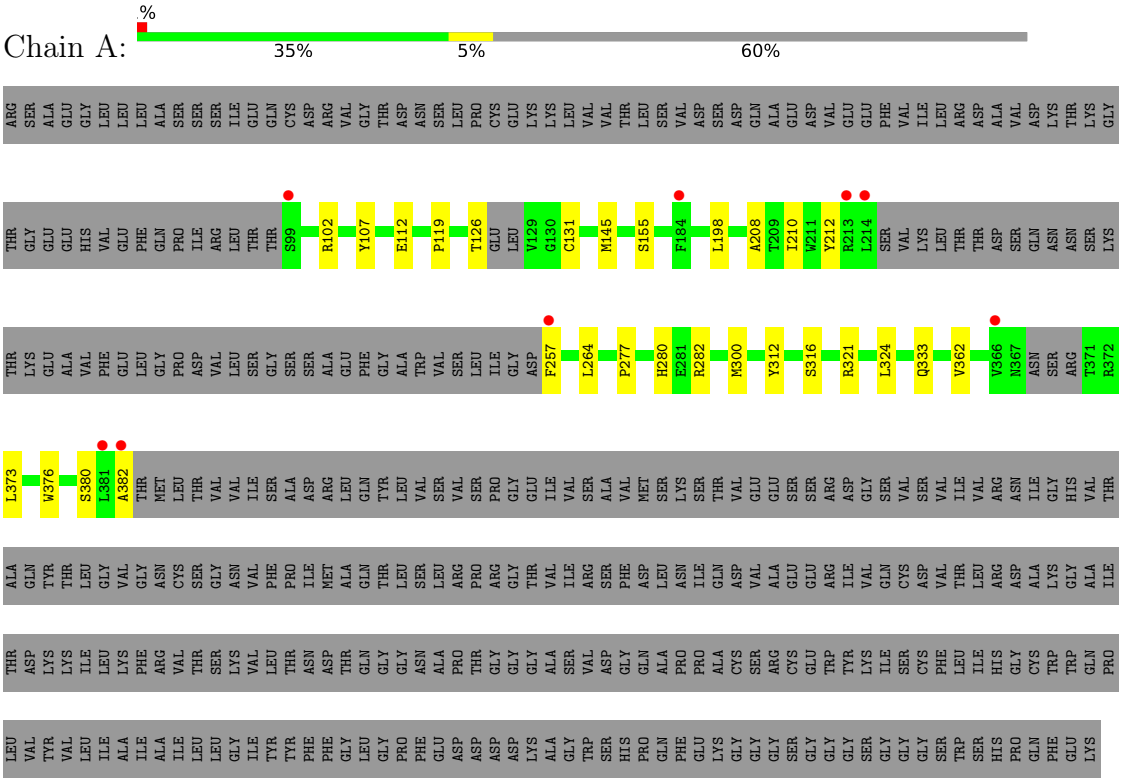
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total O 2 2	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Uncharacterized protein

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	150.09Å 150.09Å 123.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.13 – 3.10 49.13 – 3.10	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.13-3.10) 99.8 (49.13-3.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 3.12Å)	Xtriage
Refinement program	BUSTER 2.10.1	Depositor
R, R_{free}	0.233 , 0.260 0.248 , 0.271	Depositor DCC
R_{free} test set	781 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	88.5	Xtriage
Anisotropy	0.098	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 58.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	1857	wwPDB-VP
Average B, all atoms (Å^2)	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% 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	A	0.42	0/1897	0.64	0/2557

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	1855	0	1800	10	0
2	A	2	0	0	0	0
All	All	1857	0	1800	10	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 3.

All (10) 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:119:PRO:HD2	1:A:333:GLN:HG3	1.83	0.59
1:A:198:LEU:HD21	1:A:282:ARG:HD2	1.84	0.57
1:A:277:PRO:HG2	1:A:280:HIS:HB2	1.96	0.47
1:A:321:ARG:HA	1:A:324:LEU:HD12	1.97	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:126:THR:HG21	1:A:131:CYS:HA	1.98	0.46
1:A:257:PHE:HE1	1:A:382:ALA:H	1.66	0.44
1:A:107:TYR:CE2	1:A:208:ALA:HB2	2.53	0.44
1:A:155:SER:HA	1:A:312:TYR:CD1	2.55	0.41
1:A:102:ARG:HB2	1:A:376:TRP:CE3	2.55	0.41
1:A:112:GLU:HG3	1:A:300:MET:HE1	2.02	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	A	229/591 (39%)	221 (96%)	8 (4%)	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	198/495 (40%)	190 (96%)	8 (4%)	31 65

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	145	MET
1	A	210	ILE
1	A	212	TYR
1	A	264	LEU
1	A	316	SER
1	A	362	VAL
1	A	373	LEU
1	A	380	SER

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

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	A	237/591 (40%)	0.03	8 (3%) 45 24	58, 79, 127, 153	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	382	ALA	3.0
1	A	184	PHE	2.7
1	A	214	LEU	2.6
1	A	257	PHE	2.4
1	A	99	SER	2.3
1	A	213	ARG	2.3
1	A	381	LEU	2.2
1	A	366	VAL	2.1

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.