Full wwPDB X-ray Structure Validation Report ⓘ

Sep 25, 2024 – 01:18 pm BST

PDB ID	:	9F8W
Title	:	Crystal structure of the apo Pex5 peroxisomal cargo receptor from Trypanosoma brucei
Authors	:	Banasik, M.K.; Dubin, G.
Deposited on	:	2024-05-07
Resolution	:	2.35 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:

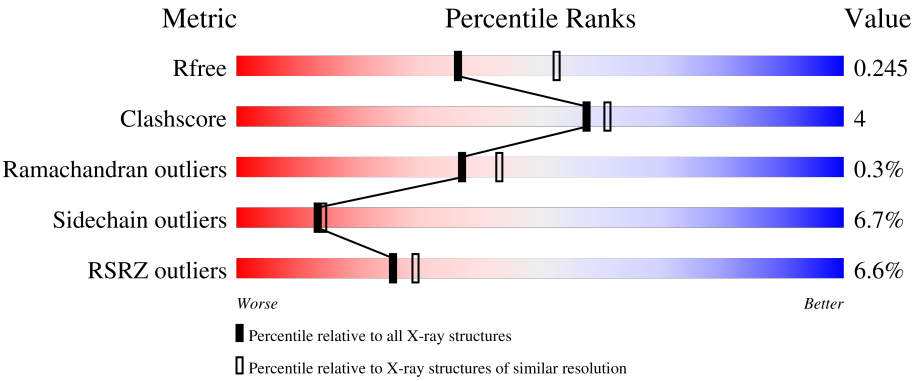
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

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}	164625	1460 (2.36-2.36)
Clashscore	180529	1571 (2.36-2.36)
Ramachandran outliers	177936	1559 (2.36-2.36)
Sidechain outliers	177891	1559 (2.36-2.36)
RSRZ outliers	164620	1460 (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 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	347	8% 79% 10% • 10%
1	B	347	4% 79% 13% 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4949 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 Peroxisome targeting signal 1 receptor PEX5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	313	Total	C	N	O	S	0	1	0
			2377	1492	415	455	15			
1	B	319	Total	C	N	O	S	0	1	0
			2460	1538	429	475	18			

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	315	MET	-	initiating methionine	UNP Q9U7C3
A	316	ALA	-	expression tag	UNP Q9U7C3
A	317	SER	-	expression tag	UNP Q9U7C3
A	318	MET	-	expression tag	UNP Q9U7C3
A	319	THR	-	expression tag	UNP Q9U7C3
A	320	GLY	-	expression tag	UNP Q9U7C3
A	321	GLY	-	expression tag	UNP Q9U7C3
A	322	GLN	-	expression tag	UNP Q9U7C3
A	323	GLN	-	expression tag	UNP Q9U7C3
A	324	MET	-	expression tag	UNP Q9U7C3
A	325	GLY	-	expression tag	UNP Q9U7C3
A	326	ARG	-	expression tag	UNP Q9U7C3
A	327	GLY	-	expression tag	UNP Q9U7C3
A	328	SER	-	expression tag	UNP Q9U7C3
A	329	GLU	-	expression tag	UNP Q9U7C3
A	330	PHE	-	expression tag	UNP Q9U7C3
A	331	MET	-	expression tag	UNP Q9U7C3
A	656	GLU	-	expression tag	UNP Q9U7C3
A	657	ASN	-	expression tag	UNP Q9U7C3
A	658	LEU	-	expression tag	UNP Q9U7C3
A	659	TYR	-	expression tag	UNP Q9U7C3
A	660	PHE	-	expression tag	UNP Q9U7C3
A	661	GLN	-	expression tag	UNP Q9U7C3
B	315	MET	-	initiating methionine	UNP Q9U7C3
B	316	ALA	-	expression tag	UNP Q9U7C3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	317	SER	-	expression tag	UNP Q9U7C3
B	318	MET	-	expression tag	UNP Q9U7C3
B	319	THR	-	expression tag	UNP Q9U7C3
B	320	GLY	-	expression tag	UNP Q9U7C3
B	321	GLY	-	expression tag	UNP Q9U7C3
B	322	GLN	-	expression tag	UNP Q9U7C3
B	323	GLN	-	expression tag	UNP Q9U7C3
B	324	MET	-	expression tag	UNP Q9U7C3
B	325	GLY	-	expression tag	UNP Q9U7C3
B	326	ARG	-	expression tag	UNP Q9U7C3
B	327	GLY	-	expression tag	UNP Q9U7C3
B	328	SER	-	expression tag	UNP Q9U7C3
B	329	GLU	-	expression tag	UNP Q9U7C3
B	330	PHE	-	expression tag	UNP Q9U7C3
B	331	MET	-	expression tag	UNP Q9U7C3
B	656	GLU	-	expression tag	UNP Q9U7C3
B	657	ASN	-	expression tag	UNP Q9U7C3
B	658	LEU	-	expression tag	UNP Q9U7C3
B	659	TYR	-	expression tag	UNP Q9U7C3
B	660	PHE	-	expression tag	UNP Q9U7C3
B	661	GLN	-	expression tag	UNP Q9U7C3

- Molecule 2 is water.

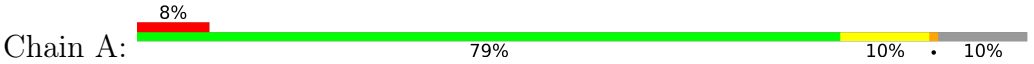
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	35	Total O 35 35	0	0
2	B	77	Total O 77 77	0	0

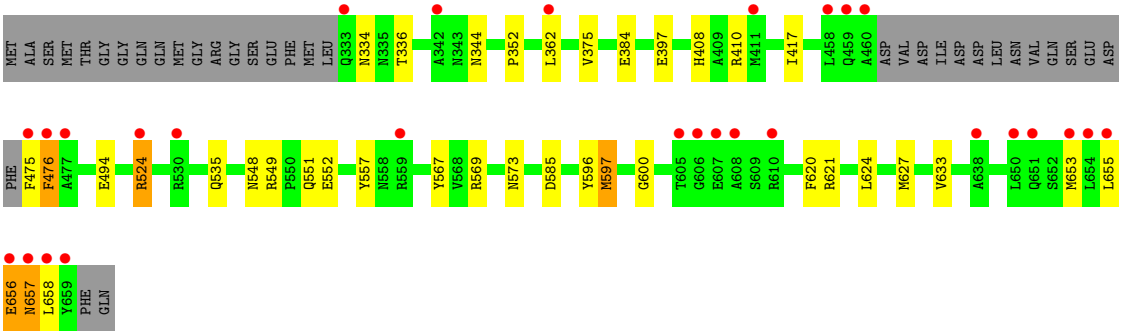
3 Residue-property plots (i)

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: Peroxisome targeting signal 1 receptor PEX5

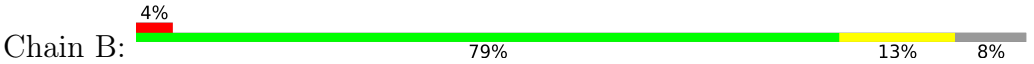
Chain A:





• Molecule 1: Peroxisome targeting signal 1 receptor PEX5

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	206.54Å 206.54Å 67.87Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.89 – 2.35 47.89 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.89-2.35) 99.9 (47.89-2.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.58 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.198 , 0.246 0.204 , 0.245	Depositor DCC
R_{free} test set	2309 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å^2)	49.1	Xtriage
Anisotropy	0.028	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 44.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.013 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4949	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% 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	0.75	0/2428	0.82	0/3310
1	B	0.74	0/2511	0.84	0/3409
All	All	0.74	0/4939	0.83	0/6719

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	2377	0	2228	18	0
1	B	2460	0	2343	23	0
2	A	35	0	0	2	0
2	B	77	0	0	4	0
All	All	4949	0	4571	41	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 4.

All (41) 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:596:TYR:HD2	1:A:655:LEU:HD12	1.34	0.92

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:624:LEU:HD23	1:A:627:MET:HE3	1.63	0.81
1:B:349:HIS:NE2	1:B:355:GLU:HG3	1.97	0.79
1:A:596:TYR:CD2	1:A:655:LEU:HD12	2.23	0.71
1:B:349:HIS:CD2	1:B:355:GLU:HG3	2.28	0.67
1:B:408:HIS:HE1	2:B:708:HOH:O	1.78	0.66
1:A:600:GLY:HA2	1:A:658:LEU:HD22	1.76	0.66
1:B:349:HIS:NE2	1:B:355:GLU:CG	2.60	0.63
1:A:336:THR:O	1:A:408:HIS:HD2	1.84	0.61
1:B:501:GLN:HG2	1:B:530:ARG:NH2	2.15	0.61
1:B:443:TRP:CZ2	1:B:501:GLN:HB3	2.36	0.60
1:A:597:MET:HG2	1:A:655:LEU:HD13	1.84	0.58
1:B:349:HIS:HE2	1:B:355:GLU:CG	2.18	0.57
1:A:656:GLU:C	1:A:658:LEU:H	2.09	0.56
1:B:352:PRO:HG2	1:B:375:VAL:HG22	1.89	0.55
1:A:475:PHE:O	1:A:476:PHE:HB2	2.06	0.55
1:A:352:PRO:HG2	1:A:375:VAL:HG22	1.90	0.54
1:B:410:ARG:HD2	2:B:764:HOH:O	2.08	0.54
1:B:524:ARG:HD2	2:B:757:HOH:O	2.07	0.53
1:B:530:ARG:NH1	1:B:533:ASP:OD2	2.42	0.53
1:B:524:ARG:CD	2:B:757:HOH:O	2.57	0.51
1:A:557:TYR:CZ	1:A:573:ASN:HB3	2.49	0.47
1:B:554:LEU:HD13	1:B:578:TYR:CE1	2.50	0.46
1:A:597:MET:CG	1:A:655:LEU:HD13	2.46	0.46
1:B:357:LEU:O	1:B:360:LEU:HB2	2.15	0.46
1:B:397:GLU:HG2	1:B:569:ARG:NH2	2.32	0.45
1:B:317:SER:HB2	1:B:323:GLN:NE2	2.33	0.44
1:A:535:GLN:HG3	2:A:708:HOH:O	2.19	0.42
1:B:501:GLN:HG2	1:B:530:ARG:HH22	1.82	0.42
1:A:621:ARG:HG2	1:A:633:VAL:HG13	2.02	0.42
1:B:579:SER:OG	1:B:623:LEU:HD21	2.20	0.42
1:B:325:GLY:O	1:B:331:MET:HG3	2.20	0.41
1:B:355:GLU:O	1:B:359:MET:HG2	2.19	0.41
1:B:557:TYR:CZ	1:B:573:ASN:HB3	2.55	0.41
1:A:334:ASN:HD22	1:A:410:ARG:HH12	1.69	0.41
1:A:524[A]:ARG:NH2	2:A:704:HOH:O	2.54	0.41
1:B:652:SER:OG	1:B:653:MET:N	2.52	0.41
1:B:428:THR:HG23	1:B:512:LEU:HD13	2.03	0.41
1:A:569:ARG:HD2	1:A:569:ARG:HA	1.89	0.40
1:A:549:ARG:HD2	1:A:552:GLU:OE2	2.21	0.40
1:A:656:GLU:C	1:A:658:LEU:N	2.75	0.40

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	310/347 (89%)	300 (97%)	8 (3%)	2 (1%)	22 24
1	B	316/347 (91%)	301 (95%)	15 (5%)	0	100 100
All	All	626/694 (90%)	601 (96%)	23 (4%)	2 (0%)	37 43

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	657	ASN
1	A	476	PHE

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	231/284 (81%)	214 (93%)	17 (7%)	11 11
1	B	248/284 (87%)	232 (94%)	16 (6%)	14 15
All	All	479/568 (84%)	446 (93%)	33 (7%)	13 13

All (33) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	344	ASN
1	A	362	LEU
1	A	384	GLU
1	A	397	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	417	ILE
1	A	494	GLU
1	A	524[A]	ARG
1	A	524[B]	ARG
1	A	548	ASN
1	A	551	GLN
1	A	567	TYR
1	A	585	ASP
1	A	597	MET
1	A	620	PHE
1	A	653	MET
1	A	656	GLU
1	A	657	ASN
1	B	322	GLN
1	B	332	LEU
1	B	336	THR
1	B	365	LEU
1	B	395	GLU
1	B	449	GLN
1	B	452	GLN
1	B	518	SER
1	B	530	ARG
1	B	567	TYR
1	B	585	ASP
1	B	597	MET
1	B	610	ARG
1	B	619	PHE
1	B	620	PHE
1	B	635	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (24) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	334	ASN
1	A	335	ASN
1	A	343	ASN
1	A	344	ASN
1	A	408	HIS
1	A	447	GLN
1	A	449	GLN
1	A	501	GLN
1	A	522	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	628	ASN
1	A	639	GLN
1	A	640	ASN
1	B	323	GLN
1	B	377	GLN
1	B	408	HIS
1	B	447	GLN
1	B	449	GLN
1	B	452	GLN
1	B	514	ASN
1	B	522	ASN
1	B	548	ASN
1	B	551	GLN
1	B	558	ASN
1	B	639	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	313/347 (90%)	0.46	28 (8%)	17	19	25, 58, 92, 118	2 (0%)
1	B	319/347 (91%)	0.17	14 (4%)	39	46	29, 53, 88, 120	1 (0%)
All	All	632/694 (91%)	0.32	42 (6%)	26	29	25, 56, 90, 120	3 (0%)

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	475	PHE	7.3
1	B	316	ALA	7.2
1	A	460	ALA	6.0
1	A	559	ARG	5.8
1	B	607	GLU	5.3
1	A	459	GLN	5.3
1	B	477	ALA	5.2
1	A	653	MET	4.8
1	A	655	LEU	4.7
1	A	476	PHE	4.3
1	A	651	GLN	4.3
1	B	654	LEU	4.3
1	A	659	TYR	3.9
1	A	605	THR	3.9
1	B	478	ALA	3.7
1	B	456	VAL	3.7
1	A	654	LEU	3.5
1	A	362	LEU	3.1
1	B	479	PRO	3.1
1	A	606	GLY	3.1
1	A	411	MET	3.0
1	B	653	MET	3.0
1	A	608	ALA	3.0
1	A	657	ASN	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	656	GLU	2.7
1	B	610	ARG	2.7
1	B	650	LEU	2.6
1	B	634	GLU	2.5
1	A	658	LEU	2.5
1	B	481	GLU	2.4
1	A	477	ALA	2.3
1	A	610	ARG	2.3
1	B	652	SER	2.3
1	A	342	ALA	2.3
1	A	458	LEU	2.3
1	B	606	GLY	2.2
1	A	607	GLU	2.2
1	A	638	ALA	2.1
1	A	530	ARG	2.1
1	A	650	LEU	2.0
1	A	524[A]	ARG	2.0
1	A	333	GLN	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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.