



Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 11:36 PM GMT

PDB ID : 3PJQ
Title : Trypanosoma cruzi trans-sialidase-like inactive isoform (including the natural mutation Tyr342His) in complex with lactose
Authors : Oppezzo, P.; Baraibar, M.; Obal, G.; Pritsch, O.; Alzari, P.M.; Buschiazzi, A.
Deposited on : 2010-11-10
Resolution : 2.10 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

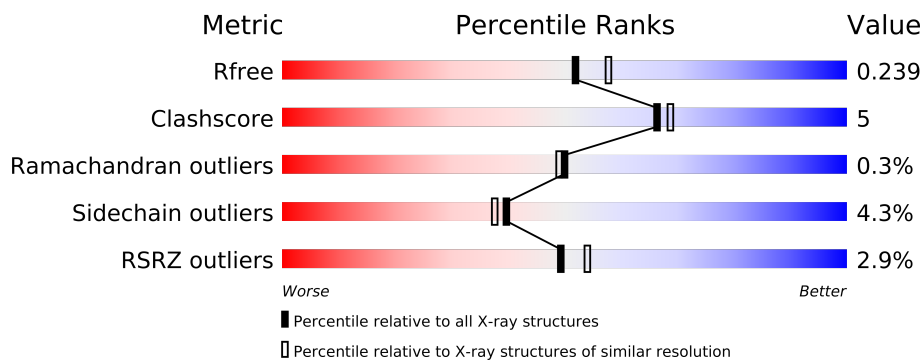
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.10 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}	66092	3012 (2.10-2.10)
Clashscore	79885	3649 (2.10-2.10)
Ramachandran outliers	78287	3610 (2.10-2.10)
Sidechain outliers	78261	3611 (2.10-2.10)
RSRZ outliers	66119	3013 (2.10-2.10)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	648	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5115 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

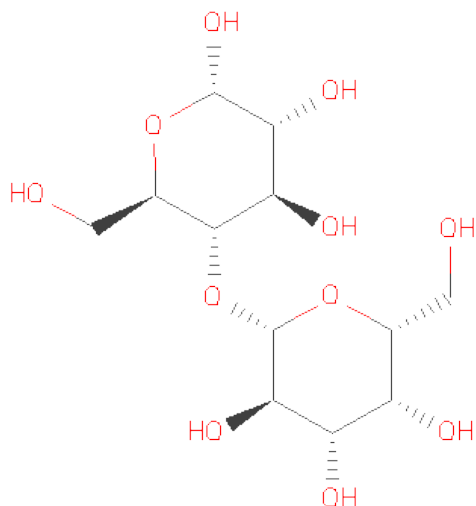
- Molecule 1 is a protein called Trans-sialidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	626	4879	3088	853	923	15	0	3	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	INITIATING METHIONINE	UNP Q26966
A	-12	GLY	-	EXPRESSION TAG	UNP Q26966
A	-11	GLY	-	EXPRESSION TAG	UNP Q26966
A	-10	SER	-	EXPRESSION TAG	UNP Q26966
A	-9	HIS	-	EXPRESSION TAG	UNP Q26966
A	-8	HIS	-	EXPRESSION TAG	UNP Q26966
A	-7	HIS	-	EXPRESSION TAG	UNP Q26966
A	-6	HIS	-	EXPRESSION TAG	UNP Q26966
A	-5	HIS	-	EXPRESSION TAG	UNP Q26966
A	-4	HIS	-	EXPRESSION TAG	UNP Q26966
A	-3	GLY	-	EXPRESSION TAG	UNP Q26966
A	-2	MET	-	EXPRESSION TAG	UNP Q26966
A	-1	ALA	-	EXPRESSION TAG	UNP Q26966
A	0	SER	-	EXPRESSION TAG	UNP Q26966
A	58	PHE	ASN	ENGINEERED MUTATION	UNP Q26966
A	262	THR	SER	SEE REMARK 999	UNP Q26966
A	342	HIS	TYR	SEE REMARK 999	UNP Q26966
A	476	HIS	ARG	SEE REMARK 999	UNP Q26966
A	484	LEU	VAL	SEE REMARK 999	UNP Q26966
A	495	LYS	SER	ENGINEERED MUTATION	UNP Q26966
A	496	GLY	VAL	ENGINEERED MUTATION	UNP Q26966
A	520	LYS	GLU	ENGINEERED MUTATION	UNP Q26966
A	558	VAL	GLU	SEE REMARK 999	UNP Q26966
A	593	GLY	ASP	ENGINEERED MUTATION	UNP Q26966
A	597	ASP	ILE	ENGINEERED MUTATION	UNP Q26966
A	599	ARG	HIS	ENGINEERED MUTATION	UNP Q26966

- Molecule 2 is SUGAR (ALPHA-LACTOSE) (three-letter code: LBT) (formula: $C_{12}H_{22}O_{11}$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			23	12	11		

- Molecule 3 is water.

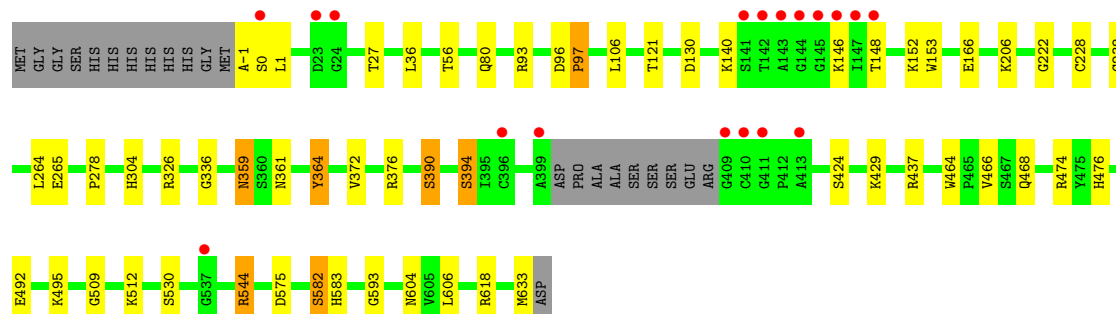
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	213	Total	O	0	0
			213	213		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Trans-sialidase

Chain A: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	54.39Å 130.22Å 54.53Å 90.00° 107.90° 90.00°	Depositor
Resolution (Å)	30.00 – 2.10 40.58 – 2.10	Depositor EDS
% Data completeness (in resolution range)	70.7 (30.00-2.10) 71.2 (40.58-2.10)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	0.05	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.49 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.6.0081	Depositor
R, R_{free}	0.186 , 0.247 0.194 , 0.239	Depositor DCC
R_{free} test set	1503 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	27.2	Xtriage
Anisotropy	0.500	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 31.8	EDS
Estimated twinning fraction	0.965 for H, K, L 0.035 for L, -K, H 0.037 for l,-k,h	Xtriage
Reported twinning fraction	0.965 for H, K, L 0.035 for L, -K, H	Depositor
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 29896 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5115	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 5.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: LBT

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	1.04	2/5001 (0.0%)	0.97	8/6787 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	228	CYS	CB-SG	-6.28	1.71	1.82
1	A	372	VAL	CB-CG1	5.34	1.64	1.52

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	544	ARG	CG-CD-NE	-6.68	97.77	111.80
1	A	376	ARG	NE-CZ-NH2	6.41	123.50	120.30
1	A	130	ASP	CB-CG-OD1	6.12	123.81	118.30
1	A	376	ARG	NE-CZ-NH1	-5.80	117.40	120.30
1	A	93	ARG	NE-CZ-NH2	-5.60	117.50	120.30
1	A	106	LEU	CB-CG-CD1	-5.28	102.03	111.00
1	A	618	ARG	NE-CZ-NH1	-5.13	117.73	120.30
1	A	474	ARG	NE-CZ-NH2	-5.10	117.75	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the

chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4879	0	0	24	0
2	A	23	0	22	1	0
3	A	213	0	0	14	0
All	All	5115	0	22	25	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 5.

All (25) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:476:HIS:CD2	3:A:665:HOH:O	2.21	0.92
1:A:361:ASN:CB	3:A:739:HOH:O	2.38	0.71
1:A:575:ASP:OD2	3:A:843:HOH:O	2.08	0.70
1:A:-1:ALA:CB	1:A:336:GLY:N	2.56	0.68
1:A:530:SER:OG	3:A:671:HOH:O	2.14	0.65
1:A:96:ASP:O	3:A:686:HOH:O	2.16	0.58
1:A:361:ASN:CG	3:A:739:HOH:O	2.42	0.57
1:A:437:ARG:NE	3:A:814:HOH:O	2.40	0.55
1:A:424:SER:OG	1:A:604:ASN:ND2	2.41	0.54
2:A:901:LBT:H6'2	3:A:776:HOH:O	2.14	0.47
1:A:0:SER:O	1:A:1:LEU:C	2.53	0.47
1:A:140:LYS:C	3:A:828:HOH:O	2.53	0.46
1:A:238:GLY:N	3:A:800:HOH:O	2.51	0.45
1:A:265:GLU:N	3:A:714:HOH:O	2.50	0.44
1:A:495:LYS:N	3:A:773:HOH:O	2.50	0.44
1:A:582:SER:OG	1:A:583:HIS:ND1	2.50	0.44
1:A:206:LYS:NZ	1:A:222:GLY:O	2.51	0.43
1:A:80:GLN:NE2	1:A:153:TRP:NE1	2.67	0.42
1:A:593:GLY:C	3:A:784:HOH:O	2.58	0.42
1:A:509:GLY:O	1:A:512:LYS:NZ	2.53	0.42
1:A:304:HIS:CE1	3:A:657:HOH:O	2.73	0.41
1:A:359:ASN:ND2	1:A:359:ASN:C	2.75	0.41
1:A:96:ASP:N	1:A:97:PRO:CD	2.84	0.41
1:A:390:SER:O	1:A:394:SER:OG	2.39	0.41
1:A:364:TYR:N	1:A:364:TYR:CD1	2.90	0.40

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	625/648 (96%)	584 (93%)	39 (6%)	2 (0%)	50	49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	468	GLN
1	A	466	VAL

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	533/547 (97%)	509 (96%)	24 (4%)	38	35

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	THR
1	A	36[A]	LEU
1	A	36[B]	LEU
1	A	56	THR
1	A	97	PRO
1	A	121	THR
1	A	146	LYS
1	A	148	THR
1	A	152	LYS
1	A	166	GLU
1	A	264	LEU
1	A	278	PRO

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	326	ARG
1	A	359	ASN
1	A	364	TYR
1	A	390	SER
1	A	394	SER
1	A	429	LYS
1	A	464	TRP
1	A	492	GLU
1	A	544	ARG
1	A	582	SER
1	A	606	LEU
1	A	633	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	LBT	A	901	-	24,24,24	0.74	0	35,35,35	1.31	3 (8%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	LBT	A	901	-	-	0/8/48/48	0/2/2/2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	901	LBT	O2'-C2'-C1'	3.21	116.72	109.89
2	A	901	LBT	C1'-O5'-C5'	2.52	117.91	113.40
2	A	901	LBT	O3-C3-C4	2.17	115.22	110.35

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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, 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	626/648 (96%)	0.01	18 (2%)	49 54	18, 36, 63, 121	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	147	ILE	5.7
1	A	146	LYS	5.7
1	A	409	GLY	5.3
1	A	410	CYS	5.0
1	A	142	THR	5.0
1	A	141	SER	4.5
1	A	148	THR	3.9
1	A	145	GLY	3.8
1	A	143	ALA	3.5
1	A	396	CYS	2.7
1	A	24	GLY	2.7
1	A	399	ALA	2.7
1	A	144	GLY	2.7
1	A	0	SER	2.7
1	A	537	GLY	2.4
1	A	23	ASP	2.1
1	A	411	GLY	2.1
1	A	413	ALA	2.0

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 carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
2	LBT	A	901	23/23	0.12	0.87	29,39,47,60	0

6.5 Other polymers

There are no such residues in this entry.