W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 2, 2017 – 02:59 AM EDT

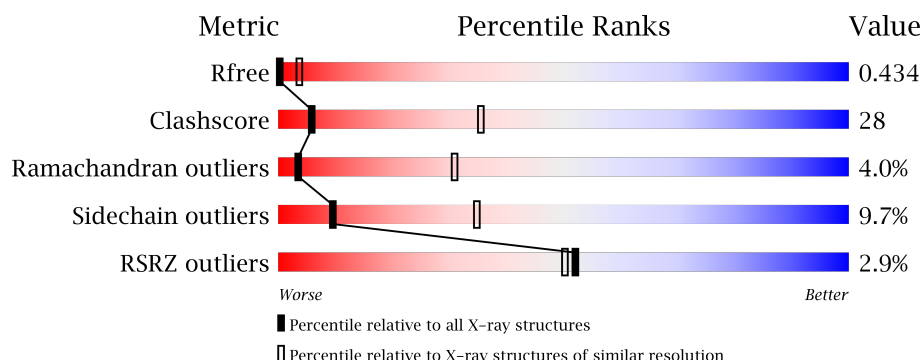|		
PDB ID	:	4HTT
Title	:	Crystal Structure of Twin Arginine Translocase Receptor- TatC in DDM
Authors	:	Ramasamy, S.; Suloway, C.J.M.; Clemons Jr., W.M.
Deposited on	:	unknown
Resolution	:	6.80 Å(reported)

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 6.80 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	Percentile Ranks	Value
Rfree		0.434
Clashscore		28
Ramachandran outliers		4.0%
Sidechain outliers		9.7%
RSRZ outliers		2.9%

Worse　　　　　　　　　　　　　　　　　　Better

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1099 (9.50-3.70)
Clashscore	112137	1033 (9.50-3.80)
Ramachandran outliers	110173	1002 (9.50-3.76)
Sidechain outliers	110143	1097 (9.50-3.70)
RSRZ outliers	101464	1002 (9.50-3.72)

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	A	418	2% / 24% 26% • 46%
1	B	418	.% / 23% 26% • 46%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3608 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 Sec-independent protein translocase protein TatC, Lysozyme.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	225	Total	C	N	O	S	16	0	0
			1804	1248	263	289	4			
1	B	225	Total	C	N	O	S	16	0	0
			1804	1248	263	289	4			

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	EXPRESSION TAG	UNP O67305
A	0	GLY	-	EXPRESSION TAG	UNP O67305
A	236	GLU	-	EXPRESSION TAG	UNP P00720
A	237	ILE	-	EXPRESSION TAG	UNP P00720
A	238	GLY	-	EXPRESSION TAG	UNP P00720
A	239	SER	-	EXPRESSION TAG	UNP P00720
A	240	GLY	-	EXPRESSION TAG	UNP P00720
A	241	ALA	-	EXPRESSION TAG	UNP P00720
A	242	SER	-	EXPRESSION TAG	UNP P00720
A	296	THR	CYS	CONFLICT	UNP P00720
A	339	ALA	CYS	CONFLICT	UNP P00720
A	407	GLU	-	EXPRESSION TAG	UNP P00720
A	408	LEU	-	EXPRESSION TAG	UNP P00720
A	409	TYR	-	EXPRESSION TAG	UNP P00720
A	410	LYS	-	EXPRESSION TAG	UNP P00720
A	411	HIS	-	EXPRESSION TAG	UNP P00720
A	412	HIS	-	EXPRESSION TAG	UNP P00720
A	413	HIS	-	EXPRESSION TAG	UNP P00720
A	414	HIS	-	EXPRESSION TAG	UNP P00720
B	-1	MET	-	EXPRESSION TAG	UNP O67305
B	0	GLY	-	EXPRESSION TAG	UNP O67305
B	236	GLU	-	EXPRESSION TAG	UNP P00720
B	237	ILE	-	EXPRESSION TAG	UNP P00720
B	238	GLY	-	EXPRESSION TAG	UNP P00720
B	239	SER	-	EXPRESSION TAG	UNP P00720

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	240	GLY	-	EXPRESSION TAG	UNP P00720
B	241	ALA	-	EXPRESSION TAG	UNP P00720
B	242	SER	-	EXPRESSION TAG	UNP P00720
B	296	THR	CYS	CONFLICT	UNP P00720
B	339	ALA	CYS	CONFLICT	UNP P00720
B	407	GLU	-	EXPRESSION TAG	UNP P00720
B	408	LEU	-	EXPRESSION TAG	UNP P00720
B	409	TYR	-	EXPRESSION TAG	UNP P00720
B	410	LYS	-	EXPRESSION TAG	UNP P00720
B	411	HIS	-	EXPRESSION TAG	UNP P00720
B	412	HIS	-	EXPRESSION TAG	UNP P00720
B	413	HIS	-	EXPRESSION TAG	UNP P00720
B	414	HIS	-	EXPRESSION TAG	UNP P00720

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 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: Sec-independent protein translocase protein TatC, Lysozyme

- Molecule 1: Sec-independent protein translocase protein TatC, Lysozyme

K228 L229 ALA THR ARG LYS LYS LYS GLU ILE GLY SER GLY ALA SER MET ASN ILE PHE GLU MET LEU ARG ILE ASP GLU GLY LEU ARG LEU LYS ILE TYR LYS ASP THR GLU GLY TYR TYR THR ILE GLY ILE GLY HIS LEU LEU THR LYS SER PRO SER LEU ASN ALA ALA LYS SER GLU

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

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

LEU TYR LYS HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	I 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	142.01Å 142.01Å 251.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 6.80 39.25 – 6.80	Depositor EDS
% Data completeness (in resolution range)	95.2 (30.00-6.80) 95.0 (39.25-6.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$<I/\sigma(I)>$ [1]	2.81 (at 6.64Å)	Xtriage
Refinement program	REFMAC, CNS	Depositor
R, R_{free}	0.344 , 0.418 0.356 , 0.434	Depositor DCC
R_{free} test set	98 reflections (4.53%)	DCC
Wilson B-factor (Å^2)	409.8	Xtriage
Anisotropy	0.065	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 398.7	EDS
L-test for twinning[2]	$<|L|> = 0.46$, $<L^2> = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.76	EDS
Total number of atoms	3608	wwPDB-VP
Average B, all atoms (Å^2)	474.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.24% 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 (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	0.28	0/1853	0.49	0/2522
1	B	0.28	0/1853	0.49	0/2522
All	All	0.28	0/3706	0.49	0/5044

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	1804	0	1946	104	0
1	B	1804	0	1946	108	2
All	All	3608	0	3892	207	2

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 28.

The worst 5 of 207 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:189:LYS:HB3	1:B:206:VAL:HG22	1.48	0.96
1:A:214:ILE:HA	1:A:217:LEU:HD12	1.54	0.89

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:141:THR:HG23	1:A:142:PRO:HD2	1.56	0.87
1:B:141:THR:HG23	1:B:142:PRO:HD2	1.59	0.85
1:A:164:PHE:O	1:A:167:PRO:HD2	1.80	0.81

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:95:HIS:NE2	1:B:95:HIS:NE2[8_775]	1.93	0.27
1:B:22:LEU:CD1	1:B:22:LEU:CD1[6_565]	2.13	0.07

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	223/418 (53%)	192 (86%)	22 (10%)	9 (4%)	3	31
1	B	223/418 (53%)	193 (86%)	21 (9%)	9 (4%)	3	31
All	All	446/836 (53%)	385 (86%)	43 (10%)	18 (4%)	3	31

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	58	THR
1	A	141	THR
1	A	211	LEU
1	B	94	SER
1	B	141	THR

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	195/356 (55%)	176 (90%)	19 (10%)	9	35
1	B	195/356 (55%)	176 (90%)	19 (10%)	9	35
All	All	390/712 (55%)	352 (90%)	38 (10%)	9	35

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	193	VAL
1	B	39	LEU
1	B	165	GLU
1	B	7	LEU
1	B	45	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	137	GLN
1	B	137	GLN
1	B	209	GLN

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 (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	225/418 (53%)	-0.16	8 (3%)	43	43	212, 472, 626, 717	4 (1%)
1	B	225/418 (53%)	-0.23	5 (2%)	62	60	211, 472, 626, 717	4 (1%)
All	All	450/836 (53%)	-0.19	13 (2%)	52	50	211, 472, 638, 717	8 (1%)

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	137	GLN	3.8
1	B	145	SER	3.6
1	A	134	GLY	3.4
1	B	52	LEU	3.0
1	A	135	PHE	2.4

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.