



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 10:50 am GMT

PDB ID : 2EPG
Title : Crystal structure of TTHA1785
Authors : Sekine, S.; Bessho, Y.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on : 2007-03-30
Resolution : 2.10 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

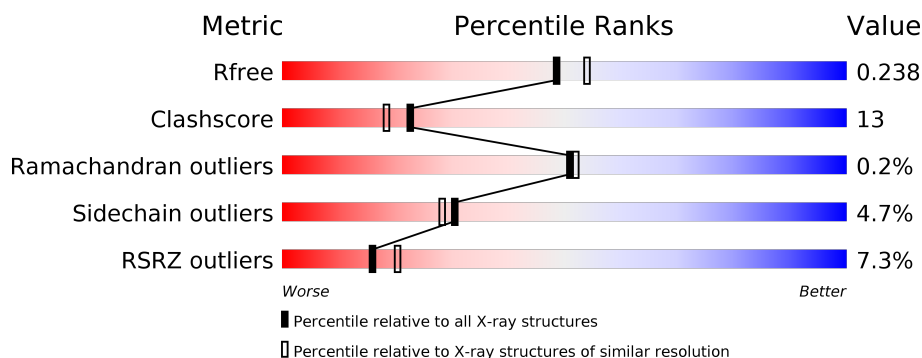
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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}	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (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. 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	487	<div> <div>7%</div> <div> <div></div> <div>67%</div> <div>21%</div> <div>9%</div> </div> </div>
1	B	487	<div> <div>6%</div> <div> <div></div> <div>68%</div> <div>21%</div> <div>9%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7188 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 Hypothetical protein TTHA1785.

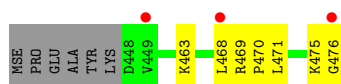
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	441	Total	C	N	O	S	Se	0	0	0
			3440	2197	617	616	3	7			
1	B	444	Total	C	N	O	S	Se	0	0	0
			3459	2207	620	622	3	7			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	MSE	-	EXPRESSION TAG	UNP Q5SHE5
A	-9	ALA	-	EXPRESSION TAG	UNP Q5SHE5
A	-8	ARG	-	EXPRESSION TAG	UNP Q5SHE5
A	-7	THR	-	EXPRESSION TAG	UNP Q5SHE5
A	-6	LEU	-	EXPRESSION TAG	UNP Q5SHE5
A	-5	PHE	-	EXPRESSION TAG	UNP Q5SHE5
A	-4	LEU	-	EXPRESSION TAG	UNP Q5SHE5
A	-3	GLU	-	EXPRESSION TAG	UNP Q5SHE5
A	-2	GLU	-	EXPRESSION TAG	UNP Q5SHE5
A	-1	GLY	-	EXPRESSION TAG	UNP Q5SHE5
A	0	PRO	-	EXPRESSION TAG	UNP Q5SHE5
B	-10	MSE	-	EXPRESSION TAG	UNP Q5SHE5
B	-9	ALA	-	EXPRESSION TAG	UNP Q5SHE5
B	-8	ARG	-	EXPRESSION TAG	UNP Q5SHE5
B	-7	THR	-	EXPRESSION TAG	UNP Q5SHE5
B	-6	LEU	-	EXPRESSION TAG	UNP Q5SHE5
B	-5	PHE	-	EXPRESSION TAG	UNP Q5SHE5
B	-4	LEU	-	EXPRESSION TAG	UNP Q5SHE5
B	-3	GLU	-	EXPRESSION TAG	UNP Q5SHE5
B	-2	GLU	-	EXPRESSION TAG	UNP Q5SHE5
B	-1	GLY	-	EXPRESSION TAG	UNP Q5SHE5
B	0	PRO	-	EXPRESSION TAG	UNP Q5SHE5

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	148	Total 148	O 148	0	0
2	B	141	Total 141	O 141	0	0



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	64.69Å 66.69Å 116.99Å 90.00° 101.21° 90.00°	Depositor
Resolution (Å)	38.63 – 2.10 38.63 – 2.11	Depositor EDS
% Data completeness (in resolution range)	99.0 (38.63-2.10) 99.1 (38.63-2.11)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.27 (at 2.10Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.199 , 0.238 0.200 , 0.238	Depositor DCC
R_{free} test set	2856 reflections (5.07%)	DCC
Wilson B-factor (Å ²)	32.5	Xtriage
Anisotropy	0.371	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 46.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7188	wwPDB-VP
Average B, all atoms (Å ²)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.79% 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	0.41	0/3510	0.66	1/4728 (0.0%)
1	B	0.42	0/3529	0.67	1/4754 (0.0%)
All	All	0.42	0/7039	0.66	2/9482 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	98	CYS	N-CA-C	-6.39	93.73	111.00
1	B	98	CYS	N-CA-C	-5.70	95.62	111.00

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	3440	0	3444	99	0
1	B	3459	0	3458	81	0
2	A	148	0	0	8	0
2	B	141	0	0	3	0
All	All	7188	0	6902	179	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 13.

The worst 5 of 179 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:376:MSE:HE1	1:A:475:LYS:HG2	1.23	1.10
1:A:176:LEU:H	1:A:291:ASN:HD21	1.00	0.95
1:A:205:PHE:HB3	1:A:376:MSE:HE3	1.50	0.92
1:B:176:LEU:H	1:B:291:ASN:HD21	0.98	0.91
1:A:286:MSE:HE1	1:A:331:ALA:HB3	1.54	0.89

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	433/487 (89%)	415 (96%)	16 (4%)	2 (0%)	32	28
1	B	436/487 (90%)	415 (95%)	21 (5%)	0	100	100
All	All	869/974 (89%)	830 (96%)	37 (4%)	2 (0%)	51	52

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	256	VAL
1	A	128	PRO

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	349/377 (93%)	333 (95%)	16 (5%)	31	29
1	B	351/377 (93%)	334 (95%)	17 (5%)	30	27
All	All	700/754 (93%)	667 (95%)	33 (5%)	30	28

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	418	ASN
1	B	58	PRO
1	B	419	LEU
1	A	468	LEU
1	B	15	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	291	ASN
1	A	330	ASN
1	B	291	ASN
1	A	268	GLN
1	B	66	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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	434/487 (89%)	0.34	35 (8%) 13 16	20, 34, 66, 116	0
1	B	437/487 (89%)	0.28	29 (6%) 19 24	22, 35, 62, 103	0
All	All	871/974 (89%)	0.31	64 (7%) 16 20	20, 35, 64, 116	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	130	GLY	8.1
1	A	420	VAL	7.6
1	A	419	LEU	7.1
1	B	476	GLY	6.5
1	A	425	GLU	6.1

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.