Full wwPDB X-ray Structure Validation Report (i)

Mar 21, 2016 – 04:46 PM EDT

PDB ID	:	5AZ9
Title	:	Crystal structure of (5-residue deleted)MBP-Tom20 fusion protein tethered with ALDH presequence via a disulfide bond
Authors	:	Matsuoka, R.; Kohda, D.
Deposited on	:	2015-09-27
Resolution	:	1.82 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

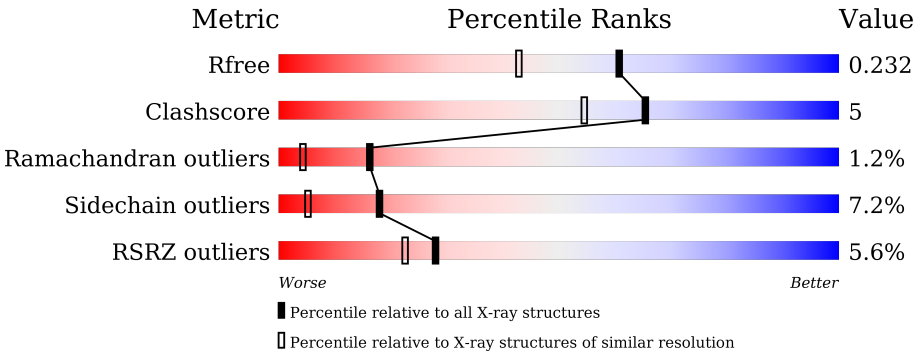
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027107
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0122
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027107

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 1.82 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}	91344	5422 (1.84-1.80)
Clashscore	102246	6347 (1.84-1.80)
Ramachandran outliers	100387	6276 (1.84-1.80)
Sidechain outliers	100360	6276 (1.84-1.80)
RSRZ outliers	91569	5439 (1.84-1.80)

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	430	6% 83% 13% ...

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3498 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 Maltose-binding periplasmic protein,Mitochondrial import receptor subunit TOM20 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	425	3326	2151	539	628	8	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	THR	deletion	UNP P0AEX9
A	?	-	GLY	deletion	UNP P0AEX9
A	?	-	ASP	deletion	UNP P0AEX9
A	308	VAL	ALA	engineered mutation	UNP P0AEX9
A	365	LYS	-	linker	UNP P0AEX9
A	366	GLU	-	linker	UNP P0AEX9
A	367	ALA	-	linker	UNP P0AEX9
A	368	LEU	-	linker	UNP P0AEX9

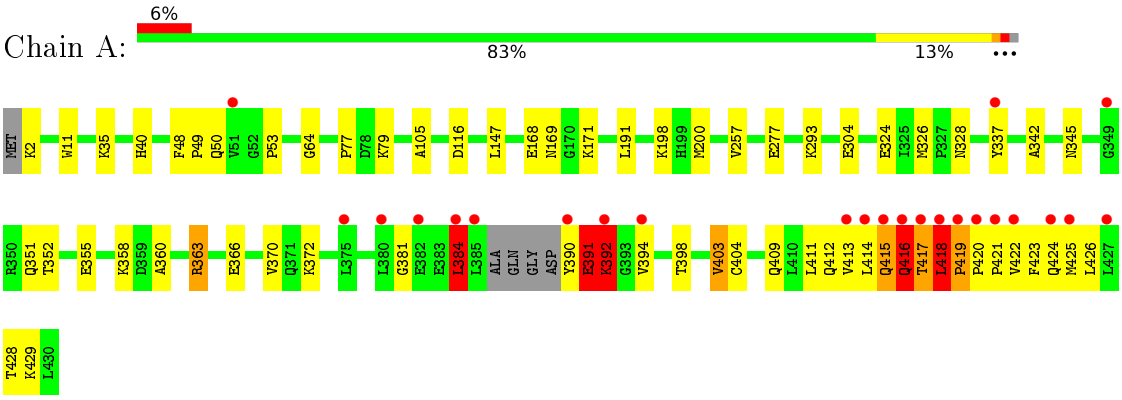
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	172	172	172	0	0

3 Residue-property plots (i)

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: Maltose-binding periplasmic protein,Mitochondrial import receptor subunit TOM20 homolog



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	160.48Å 71.40Å 47.54Å 90.00° 104.91° 90.00°	Depositor
Resolution (Å)	39.55 – 1.82 39.55 – 1.82	Depositor EDS
% Data completeness (in resolution range)	99.8 (39.55-1.82) 99.8 (39.55-1.82)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 1.82Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.192 , 0.227 0.201 , 0.232	Depositor DCC
R_{free} test set	2351 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	28.8	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 41.9	EDS
Estimated twinning fraction	0.020 for -h-2*l,-k,l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 46512 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3498	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.55% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.94	2/3403 (0.1%)	0.92	5/4618 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	324	GLU	CD-OE1	5.59	1.31	1.25
1	A	304	GLU	CD-OE2	5.06	1.31	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	363	ARG	CG-CD-NE	6.39	125.23	111.80
1	A	418	LEU	CA-CB-CG	5.29	127.48	115.30
1	A	392	LYS	N-CA-C	5.29	125.28	111.00
1	A	116	ASP	CB-CG-OD1	5.10	122.89	118.30
1	A	384	LEU	CA-CB-CG	5.05	126.93	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	391	GLU	Peptide

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	3326	0	3330	33	1
2	A	172	0	0	2	0
All	All	3498	0	3330	33	1

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 5.

All (33) 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:363:ARG:NH1	1:A:366:GLU:OE2	1.99	0.94
1:A:352:THR:HG23	1:A:355:GLU:H	1.55	0.71
1:A:394:VAL:O	1:A:398:THR:HG23	1.92	0.70
1:A:422:VAL:HG22	1:A:425:MET:CE	2.24	0.68
1:A:417:THR:HG22	1:A:418:LEU:HD13	1.86	0.58
1:A:422:VAL:HG22	1:A:425:MET:HE2	1.87	0.57
1:A:390:TYR:CE1	1:A:391:GLU:HB3	2.40	0.56
1:A:11:TRP:CG	1:A:53:PRO:HG3	2.42	0.55
1:A:40:HIS:HD2	2:A:650:HOH:O	1.90	0.53
1:A:415:GLN:HG2	1:A:416:GLN:HE22	1.72	0.53
1:A:415:GLN:HG2	1:A:416:GLN:NE2	2.25	0.52
1:A:64:GLY:HA3	1:A:328:ASN:O	2.12	0.50
1:A:392:LYS:N	1:A:392:LYS:HD2	2.29	0.48
1:A:168:GLU:HG2	1:A:169:ASN:ND2	2.29	0.47
1:A:398:THR:HG22	1:A:426:LEU:HD12	1.96	0.47
1:A:412:GLN:O	1:A:416:GLN:NE2	2.48	0.46
1:A:352:THR:CG2	1:A:355:GLU:HB2	2.45	0.46
1:A:370:VAL:CG1	1:A:403:VAL:HG13	2.47	0.45
1:A:390:TYR:OH	1:A:422:VAL:HG21	2.16	0.45
1:A:345:ASN:HD22	1:A:351:GLN:NE2	2.14	0.45
1:A:35:LYS:HG3	2:A:551:HOH:O	2.16	0.45
1:A:398:THR:HG21	1:A:429:LYS:HB2	1.99	0.44
1:A:342:ALA:HB2	1:A:360:ALA:HB2	2.00	0.43
1:A:423:PHE:O	1:A:426:LEU:HB3	2.18	0.42
1:A:352:THR:HG22	1:A:355:GLU:HB2	2.00	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:381:GLY:O	1:A:384:LEU:HG	2.20	0.42
1:A:77:PRO:HA	1:A:277:GLU:OE2	2.20	0.42
1:A:105:ALA:O	1:A:257:VAL:HA	2.20	0.42
1:A:191:LEU:HD12	1:A:200:MET:HE1	2.01	0.41
1:A:420:PRO:N	1:A:421:PRO:CD	2.83	0.41
1:A:48:PHE:HB3	1:A:49:PRO:HD3	2.02	0.41
1:A:419:PRO:HG2	1:A:422:VAL:HB	2.02	0.40
1:A:363:ARG:CZ	1:A:366:GLU:OE2	2.66	0.40

All (1) 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:A:363:ARG:NH1	1:A:363:ARG:NH1[2_755]	2.14	0.06

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	421/430 (98%)	407 (97%)	9 (2%)	5 (1%)	16 4

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	391	GLU
1	A	392	LYS
1	A	418	LEU
1	A	416	GLN
1	A	419	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	349/352 (99%)	324 (93%)	25 (7%)	18 5

All (25) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	LYS
1	A	50	GLN
1	A	79	LYS
1	A	147	LEU
1	A	171	LYS
1	A	198	LYS
1	A	293	LYS
1	A	326	MET
1	A	337	TYR
1	A	358	LYS
1	A	372	LYS
1	A	384	LEU
1	A	391	GLU
1	A	403	VAL
1	A	404	CYS
1	A	409	GLN
1	A	411	LEU
1	A	413	VAL
1	A	414	LEU
1	A	415	GLN
1	A	416	GLN
1	A	417	THR
1	A	418	LEU
1	A	424	GLN
1	A	428	THR

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	351	GLN
1	A	412	GLN
1	A	416	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 carbohydrates 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	425/430 (98%)	-0.08	24 (5%)	28	22	20, 31, 68, 99	0

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	418	LEU	10.4
1	A	390	TYR	9.0
1	A	421	PRO	6.3
1	A	419	PRO	5.2
1	A	392	LYS	4.8
1	A	417	THR	4.8
1	A	422	VAL	4.7
1	A	385	LEU	4.1
1	A	427	LEU	3.8
1	A	425	MET	3.6
1	A	424	GLN	3.5
1	A	420	PRO	3.4
1	A	51	VAL	3.0
1	A	394	VAL	2.9
1	A	384	LEU	2.7
1	A	337	TYR	2.5
1	A	414	LEU	2.4
1	A	416	GLN	2.4
1	A	413	VAL	2.3
1	A	382	GLU	2.1
1	A	380	LEU	2.1
1	A	349	GLY	2.1
1	A	375	LEU	2.0
1	A	415	GLN	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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.