



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 5, 2016 – 10:41 AM EDT

PDB ID : 5DTH  
Title : Crystal structure of MUPP1 PDZ8 domain from rattus norvegicus  
Authors : Li, J.; Lv, Y.; Zhu, H.; Liu, W.  
Deposited on : 2015-09-18  
Resolution : 1.95 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](mailto: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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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)	:	rb-20027939

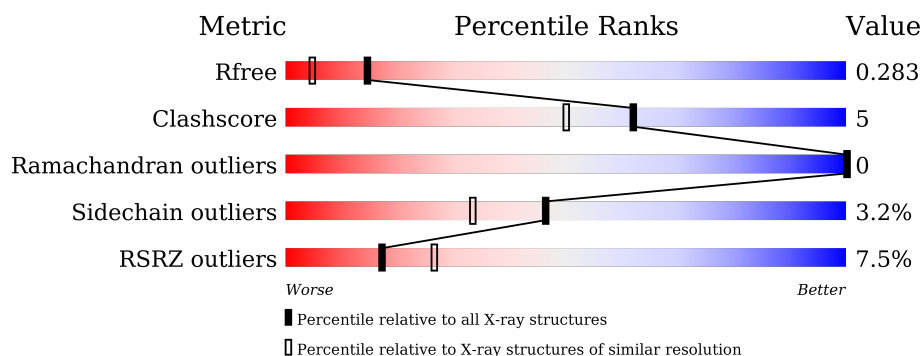
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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	1833 (1.96-1.96)
Clashscore	102246	1953 (1.96-1.96)
Ramachandran outliers	100387	1936 (1.96-1.96)
Sidechain outliers	100360	1936 (1.96-1.96)
RSRZ outliers	91569	1835 (1.96-1.96)

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  $\geq 3$ , 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	111	<div> <div>8%</div> <div> <div></div> <div>82%</div> <div>12%</div> <div>.. 5%</div> </div> </div>
1	B	111	<div> <div>4%</div> <div> <div></div> <div>86%</div> <div>9%</div> <div>..</div> </div> </div>
1	C	111	<div> <div>12%</div> <div> <div></div> <div>81%</div> <div>14%</div> <div>..</div> </div> </div>
1	D	111	<div> <div>5%</div> <div> <div></div> <div>85%</div> <div>10%</div> <div>• 5%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3420 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 Multiple PDZ domain protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	106	Total	C	N	O	S	0	0	0
			788	493	140	152	3			
1	B	109	Total	C	N	O	S	0	2	0
			829	518	149	159	3			
1	C	107	Total	C	N	O	S	0	4	0
			814	511	140	160	3			
1	D	106	Total	C	N	O	S	0	0	0
			779	491	140	145	3			

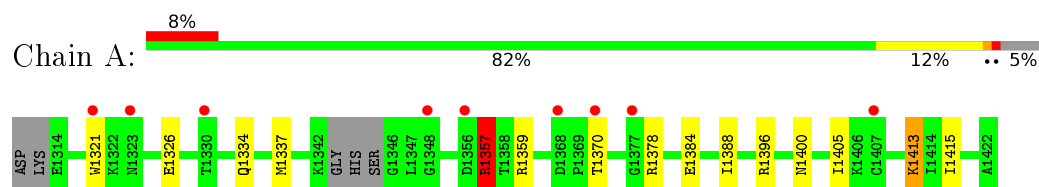
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	40	Total	O	0	0
			40	40		
2	B	53	Total	O	0	0
			53	53		
2	C	57	Total	O	0	0
			57	57		
2	D	60	Total	O	0	0
			60	60		

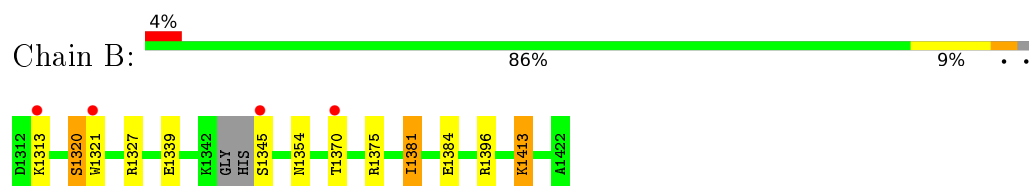
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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.

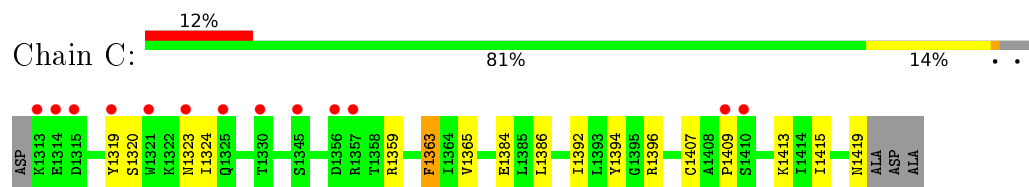
- Molecule 1: Multiple PDZ domain protein



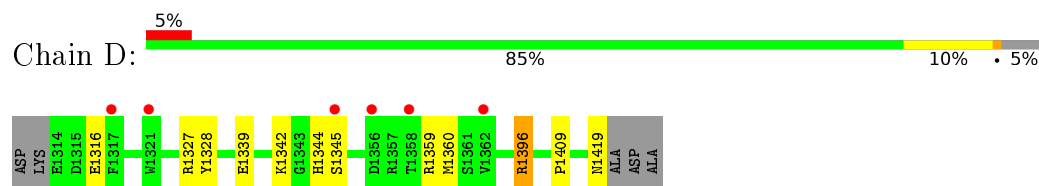
- Molecule 1: Multiple PDZ domain protein



- Molecule 1: Multiple PDZ domain protein



- Molecule 1: Multiple PDZ domain protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.10Å 39.24Å 134.45Å 90.00° 92.73° 90.00°	Depositor
Resolution (Å)	40.64 – 1.95 40.64 – 1.86	Depositor EDS
% Data completeness (in resolution range)	85.7 (40.64-1.95) 89.4 (40.64-1.86)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.00 (at 1.86Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.225 , 0.265 0.240 , 0.283	Depositor DCC
$R_{free}$ test set	1625 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	20.6	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 48.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.36$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l	Xtriage
Reported twinning fraction	0.896 for H, K, L 0.104 for -H, -K, L	Depositor
Outliers	0 of 33468 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	3420	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 50.65 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 6.2737e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical 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 ⓘ

### 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	0.90	1/798 (0.1%)	1.06	7/1076 (0.7%)
1	B	0.95	2/845 (0.2%)	1.12	7/1136 (0.6%)
1	C	1.03	1/837 (0.1%)	1.05	3/1130 (0.3%)
1	D	0.99	1/791 (0.1%)	1.08	5/1069 (0.5%)
All	All	0.97	5/3271 (0.2%)	1.08	22/4411 (0.5%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1326	GLU	CD-OE1	6.85	1.33	1.25
1	D	1339	GLU	CG-CD	6.08	1.61	1.51
1	B	1320[A]	SER	CB-OG	-5.42	1.35	1.42
1	B	1320[B]	SER	CB-OG	-5.42	1.35	1.42
1	C	1419	ASN	N-CA	5.26	1.56	1.46

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1396	ARG	NE-CZ-NH1	8.06	124.33	120.30
1	C	1396	ARG	NE-CZ-NH1	7.71	124.16	120.30
1	D	1396	ARG	NE-CZ-NH2	-7.16	116.72	120.30
1	D	1342	LYS	CB-CA-C	-6.80	96.80	110.40
1	D	1396	ARG	NE-CZ-NH1	6.67	123.64	120.30
1	A	1357	ARG	NE-CZ-NH2	6.53	123.56	120.30
1	A	1396	ARG	NE-CZ-NH2	-6.49	117.06	120.30
1	B	1381	ILE	CA-CB-CG1	-6.35	98.93	111.00
1	B	1396	ARG	NE-CZ-NH2	5.97	123.29	120.30
1	B	1384	GLU	OE1-CD-OE2	-5.92	116.19	123.30
1	A	1378	ARG	NE-CZ-NH1	-5.71	117.44	120.30
1	D	1360	MET	CA-CB-CG	5.48	122.61	113.30
1	D	1339	GLU	OE1-CD-OE2	-5.44	116.77	123.30
1	B	1396	ARG	NE-CZ-NH1	-5.38	117.61	120.30
1	A	1378	ARG	NE-CZ-NH2	5.23	122.92	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1384	GLU	CA-CB-CG	5.17	124.77	113.40
1	B	1327	ARG	NE-CZ-NH1	5.11	122.86	120.30
1	A	1384	GLU	OE1-CD-OE2	-5.09	117.19	123.30
1	B	1320[A]	SER	N-CA-CB	5.08	118.12	110.50
1	B	1320[B]	SER	N-CA-CB	5.08	118.12	110.50
1	C	1396	ARG	NE-CZ-NH2	-5.01	117.79	120.30
1	C	1363	PHE	CB-CA-C	-5.00	100.39	110.40

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	788	0	757	10	0
1	B	829	0	811	7	0
1	C	814	0	784	13	0
1	D	779	0	750	7	0
2	A	40	0	0	1	0
2	B	53	0	0	0	0
2	C	57	0	0	6	0
2	D	60	0	0	4	0
All	All	3420	0	3102	34	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 5.

All (34) 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:1337:MET:HE3	1:A:1415:ILE:HG12	1.42	0.99
1:C:1407:CYS:SG	2:C:1541:HOH:O	2.42	0.77
1:A:1388:ILE:HB	2:A:1535:HOH:O	1.85	0.75
1:C:1320:SER:N	1:C:1323[B]:ASN:OD1	2.19	0.75
1:D:1316:GLU:CB	2:D:1558:HOH:O	2.39	0.71

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:1359:ARG:CZ	2:C:1503:HOH:O	2.39	0.71
1:A:1337:MET:HE3	1:A:1415:ILE:CG1	2.21	0.69
1:B:1381:ILE:HD13	2:D:1510:HOH:O	1.96	0.64
1:C:1359:ARG:NH2	2:C:1503:HOH:O	2.30	0.63
1:B:1370:THR:O	1:B:1375:ARG:NE	2.33	0.61
1:B:1339:GLU:HB2	1:B:1413:LYS:HD3	1.88	0.56
1:C:1319:TYR:HD2	1:C:1413:LYS:HG2	1.70	0.55
1:B:1354:ASN:HD21	1:D:1327:ARG:HE	1.55	0.55
1:C:1365:VAL:O	1:D:1396:ARG:NH2	2.41	0.54
1:C:1386:LEU:HD21	2:C:1503:HOH:O	2.11	0.50
1:B:1370:THR:O	1:B:1375:ARG:CZ	2.60	0.50
1:C:1319:TYR:CD2	1:C:1413:LYS:HG2	2.47	0.49
1:D:1419:ASN:ND2	2:D:1503:HOH:O	2.44	0.49
1:A:1357:ARG:CD	1:A:1359:ARG:O	2.61	0.49
1:B:1313:LYS:O	1:B:1320[A]:SER:OG	2.30	0.49
1:A:1357:ARG:HD3	1:A:1359:ARG:O	2.14	0.47
1:B:1381:ILE:CD1	2:D:1510:HOH:O	2.60	0.45
2:C:1523:HOH:O	1:D:1409:PRO:HB3	2.17	0.45
1:A:1388:ILE:HD13	1:A:1405:ILE:HG12	1.98	0.45
1:D:1327:ARG:HD2	1:D:1328:TYR:CE2	2.50	0.45
1:A:1334:GLN:OE1	1:D:1344:HIS:HD2	2.00	0.44
1:C:1363:PHE:CE1	1:C:1384:GLU:HB2	2.53	0.43
1:C:1324:ILE:HD12	1:C:1415:ILE:HD13	2.01	0.42
1:C:1392:ILE:HG23	1:C:1394:TYR:CE1	2.55	0.42
1:A:1337:MET:CE	1:A:1415:ILE:CG1	2.95	0.41
1:C:1409:PRO:HD2	2:C:1516:HOH:O	2.20	0.41
1:A:1357:ARG:HD2	1:A:1359:ARG:O	2.21	0.41
1:A:1337:MET:SD	1:A:1413:LYS:HE2	2.61	0.41
1:C:1320:SER:O	1:C:1324:ILE:HG12	2.22	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	102/111 (92%)	99 (97%)	3 (3%)	0	100	100
1	B	107/111 (96%)	104 (97%)	3 (3%)	0	100	100
1	C	109/111 (98%)	106 (97%)	3 (3%)	0	100	100
1	D	104/111 (94%)	101 (97%)	3 (3%)	0	100	100
All	All	422/444 (95%)	410 (97%)	12 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	78/90 (87%)	73 (94%)	5 (6%)	22	8
1	B	84/90 (93%)	81 (96%)	3 (4%)	42	28
1	C	83/90 (92%)	83 (100%)	0	100	100
1	D	76/90 (84%)	74 (97%)	2 (3%)	54	43
All	All	321/360 (89%)	311 (97%)	10 (3%)	46	34

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1321	TRP
1	A	1357	ARG
1	A	1370	THR
1	A	1400	ASN
1	A	1413	LYS
1	B	1321	TRP
1	B	1345	SER
1	B	1413	LYS
1	D	1345	SER
1	D	1359	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1354	ASN
1	C	1398	HIS
1	D	1344	HIS
1	D	1419	ASN

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	106/111 (95%)	0.82	9 (8%) 13 21	12, 29, 45, 50	0
1	B	109/111 (98%)	0.50	4 (3%) 45 56	11, 23, 38, 47	0
1	C	107/111 (96%)	0.77	13 (12%) 6 9	12, 21, 37, 50	0
1	D	106/111 (95%)	0.60	6 (5%) 27 37	13, 22, 37, 47	0
All	All	428/444 (96%)	0.67	32 (7%) 17 26	11, 24, 41, 50	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1357	ARG	4.9
1	D	1356	ASP	4.3
1	B	1321	TRP	4.0
1	C	1330	THR	3.9
1	B	1313	LYS	3.8
1	C	1345	SER	3.4
1	A	1323	ASN	3.1
1	D	1321	TRP	3.1
1	C	1314	GLU	3.1
1	C	1356	ASP	3.0
1	A	1321	TRP	2.9
1	C	1321	TRP	2.9
1	C	1315	ASP	2.9
1	B	1345	SER	2.9
1	D	1358	THR	2.8
1	A	1370	THR	2.7
1	C	1313	LYS	2.7
1	A	1407	CYS	2.6
1	C	1319	TYR	2.4
1	B	1370	THR	2.3
1	C	1323[A]	ASN	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	1345	SER	2.3
1	A	1368	ASP	2.3
1	A	1377	GLY	2.2
1	C	1409	PRO	2.2
1	C	1325	GLN	2.2
1	A	1348	GLY	2.1
1	D	1362	VAL	2.1
1	D	1317	PHE	2.1
1	A	1356	ASP	2.1
1	C	1410	SER	2.1
1	A	1330	THR	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 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.