



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 15, 2017 – 08:07 PM EST

PDB ID : 5KPI  
Title : Mouse native PGP  
Authors : Xia, D.; Esser, L.; Zhou, F.  
Deposited on : 2016-07-04  
Resolution : 4.01 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-20028442  
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-20028442

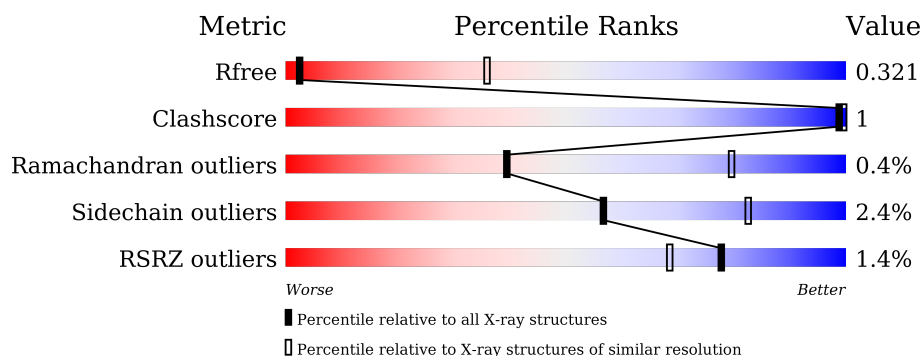
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 4.01 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	1015 (4.48-3.56)
Clashscore	102246	1054 (4.44-3.60)
Ramachandran outliers	100387	1006 (4.42-3.60)
Sidechain outliers	100360	1016 (4.46-3.58)
RSRZ outliers	91569	1018 (4.48-3.56)

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	1282	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 1%, green 88%, grey 8%);"></div> <div style="display: flex; justify-content: space-between; padding: 0 5px;"> <span>%</span> <span>88%</span> <span>8%</span> </div> </div>
1	B	1282	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 2%, orange 2%, yellow 1%, green 88%, grey 8%);"></div> <div style="display: flex; justify-content: space-between; padding: 0 5px;"> <span>2%</span> <span>88%</span> <span>5%</span> <span>8%</span> </div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 37077 atoms, of which 18703 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 Multidrug resistance protein 1A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1185	Total	C	H	N	O	S	0	0	0
			18542	5907	9353	1557	1687	38			
1	B	1184	Total	C	H	N	O	S	0	0	0
			18535	5905	9350	1556	1686	38			

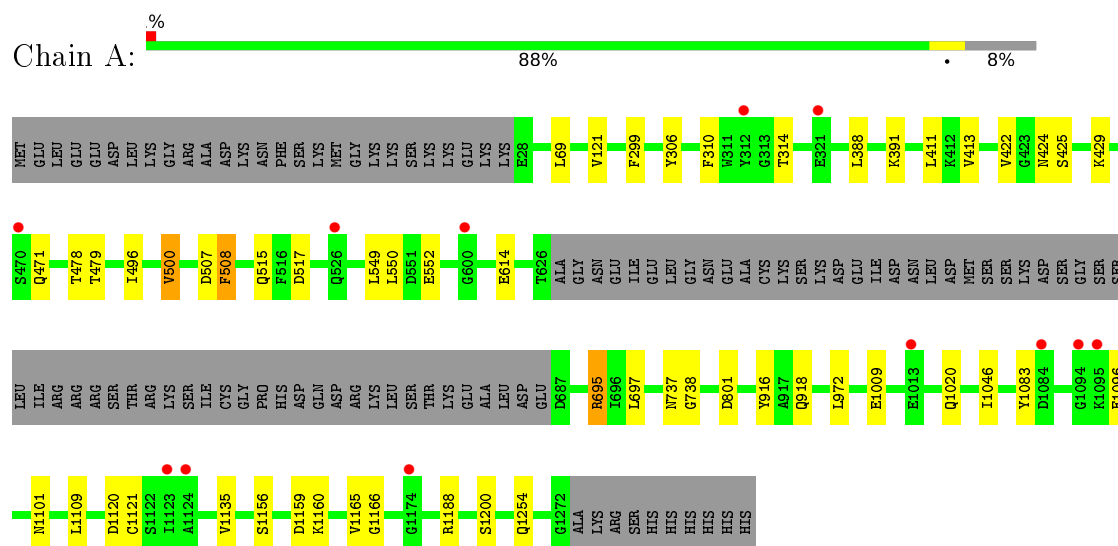
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1277	HIS	-	expression tag	UNP P21447
A	1278	HIS	-	expression tag	UNP P21447
A	1279	HIS	-	expression tag	UNP P21447
A	1280	HIS	-	expression tag	UNP P21447
A	1281	HIS	-	expression tag	UNP P21447
A	1282	HIS	-	expression tag	UNP P21447
B	1277	HIS	-	expression tag	UNP P21447
B	1278	HIS	-	expression tag	UNP P21447
B	1279	HIS	-	expression tag	UNP P21447
B	1280	HIS	-	expression tag	UNP P21447
B	1281	HIS	-	expression tag	UNP P21447
B	1282	HIS	-	expression tag	UNP P21447

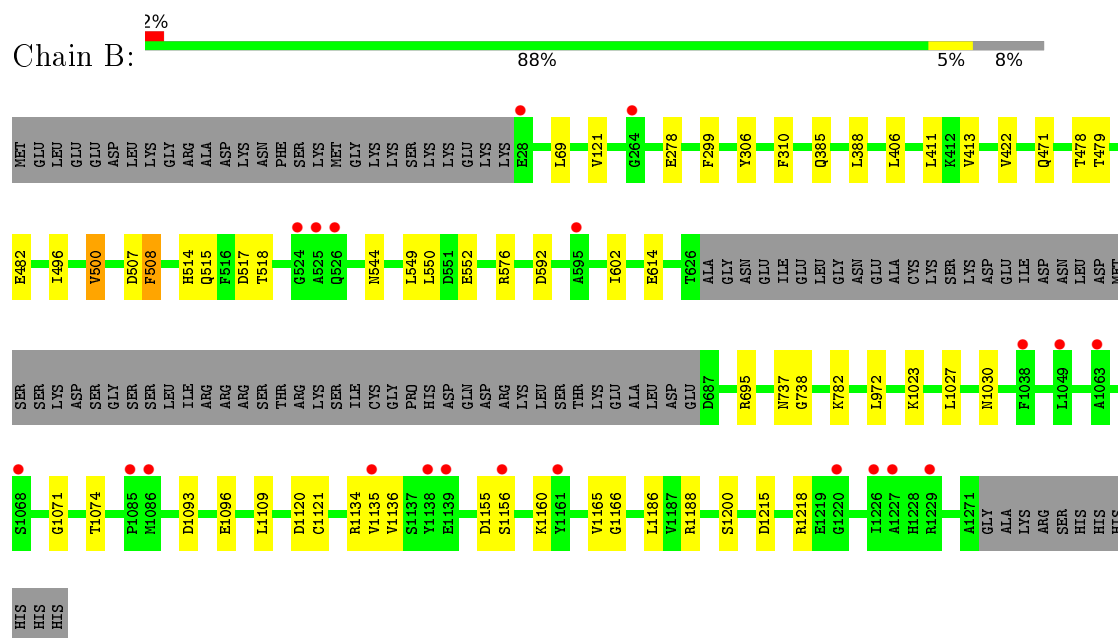
### 3 Residue-property plots [i](#)

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: Multidrug resistance protein 1A



#### • Molecule 1: Multidrug resistance protein 1A



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	100.61Å 119.53Å 383.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.14 – 4.01 35.37 – 4.01	Depositor EDS
% Data completeness (in resolution range)	94.8 (20.14-4.01) 94.7 (35.37-4.01)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 3.99Å)	Xtriage
Refinement program	PHENIX (dev_2443: ???)	Depositor
R, $R_{free}$	0.258 , 0.303 0.268 , 0.321	Depositor DCC
$R_{free}$ test set	1868 reflections (4.97%)	DCC
Wilson B-factor (Å <sup>2</sup> )	184.9	Xtriage
Anisotropy	0.427	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 142.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	37077	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	262.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.30% of the height of the origin peak. No significant pseudotranslation is detected.*

<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	0.34	0/9358	0.62	0/12650
1	B	0.34	0/9354	0.63	1/12645 (0.0%)
All	All	0.34	0/18712	0.63	1/25295 (0.0%)

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

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	406	LEU	CA-CB-CG	5.04	126.88	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1159	ASP	Peptide

### 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	9189	9353	9370	15	1
1	B	9185	9350	9367	17	1
All	All	18374	18703	18737	31	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 1.

All (31) 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:695:ARG:NH2	1:A:1009:GLU:OE2	2.26	0.68
1:A:471:GLN:NE2	1:A:552:GLU:OE1	2.36	0.59
1:B:278:GLU:OE1	1:B:782:LYS:NZ	2.38	0.56
1:A:801:ASP:OD2	1:A:1083:TYR:OH	2.24	0.56
1:A:1020:GLN:O	1:A:1101:ASN:ND2	2.40	0.55
1:B:514:HIS:O	1:B:518:THR:OG1	2.14	0.54
1:B:471:GLN:NE2	1:B:552:GLU:OE1	2.39	0.54
1:B:514:HIS:O	1:B:518:THR:N	2.40	0.54
1:B:544:ASN:OD1	1:B:576:ARG:NH2	2.42	0.52
1:B:1109:LEU:O	1:B:1188:ARG:NH1	2.44	0.51
1:A:424:ASN:OD1	1:A:425:SER:N	2.42	0.50
1:B:478:THR:HG22	1:B:479:THR:H	1.78	0.48
1:A:478:THR:HG22	1:A:479:THR:H	1.77	0.48
1:A:1165:VAL:HG12	1:A:1166:GLY:H	1.79	0.48
1:B:1165:VAL:HG12	1:B:1166:GLY:H	1.78	0.48
1:A:918:GLN:NE2	1:B:482:GLU:OE2	2.45	0.47
1:B:737:ASN:OD1	1:B:738:GLY:N	2.48	0.46
1:A:737:ASN:OD1	1:A:738:GLY:N	2.48	0.46
1:A:424:ASN:O	1:A:429:LYS:NZ	2.48	0.46
1:A:1254:GLN:N	1:A:1254:GLN:OE1	2.49	0.44
1:B:496:ILE:O	1:B:500:VAL:HG23	2.18	0.43
1:A:507:ASP:OD1	1:A:508:PHE:N	2.52	0.43
1:B:1215:ASP:OD1	1:B:1218:ARG:NH2	2.52	0.43
1:B:388:LEU:HB2	1:B:413:VAL:HB	2.00	0.43
1:B:507:ASP:OD1	1:B:508:PHE:N	2.52	0.42
1:B:1071:GLY:O	1:B:1074:THR:OG1	2.29	0.42
1:B:1030:ASN:N	1:B:1093:ASP:OD1	2.50	0.42
1:A:496:ILE:O	1:A:500:VAL:HG23	2.20	0.41
1:B:1134:ARG:O	1:B:1136:VAL:N	2.54	0.41
1:A:388:LEU:HB2	1:A:413:VAL:HB	2.03	0.41
1:A:1109:LEU:O	1:A:1188:ARG:NH1	2.54	0.41

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:391:LYS:NZ	1:B:1023:LYS:O[1_565]	2.18	0.02

## 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	1181/1282 (92%)	1121 (95%)	55 (5%)	5 (0%)	39 79
1	B	1180/1282 (92%)	1122 (95%)	53 (4%)	5 (0%)	39 79
All	All	2361/2564 (92%)	2243 (95%)	108 (5%)	10 (0%)	39 79

All (10) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	515	GLN
1	A	1160	LYS
1	B	515	GLN
1	B	1160	LYS
1	A	1120	ASP
1	B	1120	ASP
1	B	411	LEU
1	A	411	LEU
1	B	1135	VAL
1	A	1135	VAL

### 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	977/1063 (92%)	955 (98%)	22 (2%)	58	83
1	B	977/1063 (92%)	953 (98%)	24 (2%)	55	82
All	All	1954/2126 (92%)	1908 (98%)	46 (2%)	57	82

All (46) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	69	LEU
1	A	121	VAL
1	A	299	PHE
1	A	306	TYR
1	A	310	PHE
1	A	314	THR
1	A	422	VAL
1	A	500	VAL
1	A	508	PHE
1	A	517	ASP
1	A	549	LEU
1	A	550	LEU
1	A	614	GLU
1	A	695	ARG
1	A	697	LEU
1	A	916	TYR
1	A	972	LEU
1	A	1046	ILE
1	A	1096	GLU
1	A	1121	CYS
1	A	1156	SER
1	A	1200	SER
1	B	69	LEU
1	B	121	VAL
1	B	299	PHE
1	B	306	TYR
1	B	310	PHE
1	B	385	GLN
1	B	422	VAL
1	B	500	VAL
1	B	508	PHE
1	B	517	ASP
1	B	549	LEU
1	B	550	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	592	ASP
1	B	602	ILE
1	B	614	GLU
1	B	695	ARG
1	B	972	LEU
1	B	1027	LEU
1	B	1096	GLU
1	B	1121	CYS
1	B	1155	ASP
1	B	1156	SER
1	B	1186	LEU
1	B	1200	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	87	ASN
1	A	458	ASN
1	A	471	GLN
1	A	608	HIS
1	A	878	GLN
1	A	1020	GLN
1	A	1077	GLN
1	B	458	ASN
1	B	471	GLN
1	B	608	HIS
1	B	962	GLN
1	B	1020	GLN
1	B	1050	GLN
1	B	1077	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 [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

### 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<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	1185/1282 (92%)	-0.19	12 (1%) 84 77	147, 228, 323, 498	0
1	B	1184/1282 (92%)	-0.17	21 (1%) 71 61	140, 234, 354, 592	0
All	All	2369/2564 (92%)	-0.18	33 (1%) 78 68	140, 231, 340, 592	0

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1227	ALA	6.6
1	B	28	GLU	5.3
1	A	1124	ALA	3.9
1	B	1139	GLU	3.3
1	B	526	GLN	3.2
1	B	1068	SER	2.8
1	A	1095	LYS	2.7
1	B	525	ALA	2.7
1	B	1138	TYR	2.6
1	A	1123	ILE	2.6
1	B	1135	VAL	2.5
1	B	1063	ALA	2.5
1	B	1161	TYR	2.5
1	B	1049	LEU	2.5
1	B	1229	ARG	2.5
1	A	321	GLU	2.4
1	B	1226	ILE	2.4
1	A	470	SER	2.4
1	A	312	TYR	2.3
1	B	264	GLY	2.3
1	B	1156	SER	2.3
1	B	1220	GLY	2.2
1	B	1038	PHE	2.2
1	A	1174	GLY	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	1094	GLY	2.2
1	A	1013	GLU	2.2
1	A	600	GLY	2.1
1	B	1086	MET	2.1
1	B	524	GLY	2.1
1	A	526	GLN	2.1
1	B	1085	PRO	2.1
1	A	1084	ASP	2.0
1	B	595	ALA	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.