



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 12, 2017 – 09:55 pm GMT

PDB ID : 4X1Q  
Title : The crystal structure of mupain-1 in complex with murinised human uPA at pH7.4  
Authors : Jiang, L.; Zhao, B.; Xu, P.; Andreasen, P.; Huang, M.  
Deposited on : 2014-11-25  
Resolution : 2.28 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](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  
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

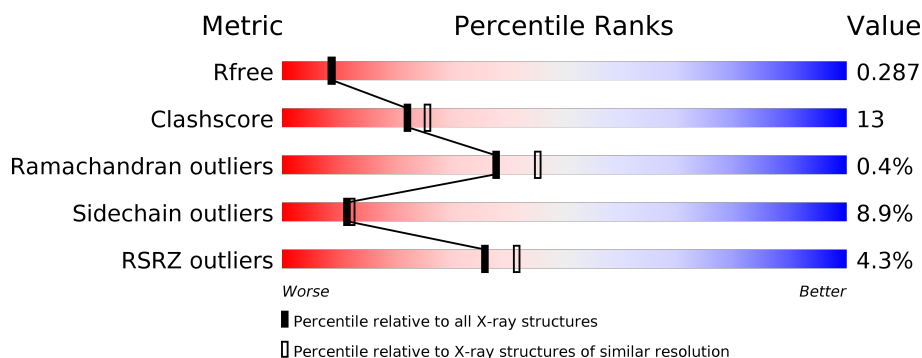
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.28 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	5609 (2.30-2.26)
Clashscore	112137	6364 (2.30-2.26)
Ramachandran outliers	110173	6281 (2.30-2.26)
Sidechain outliers	110143	6281 (2.30-2.26)
RSRZ outliers	101464	5639 (2.30-2.26)

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	P	10	<div> <div>30%</div> <div> <div></div> <div>50%</div> <div>40%</div> <div>10%</div> </div> </div>
2	U	247	<div> <div>3%</div> <div> <div></div> <div>71%</div> <div>26%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2060 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 mupain-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	P	10	Total	C	N	O	S	0	0	0
			81	51	13	15	2			

- Molecule 2 is a protein called Urokinase-type plasminogen activator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	U	247	Total	C	N	O	S	0	1	0
			1959	1237	340	366	16			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
U	99	TYR	HIS	engineered mutation	UNP P00749
U	122	ALA	CYS	engineered mutation	UNP P00749
U	145	GLN	ASN	engineered mutation	UNP P00749

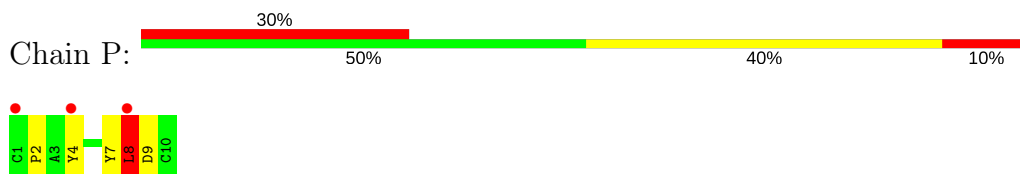
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	U	20	Total	O	0	0
			20	20		

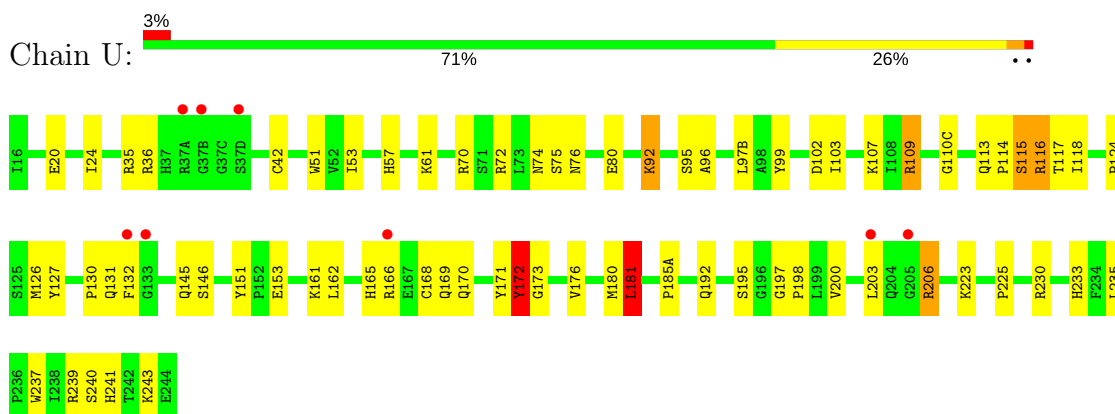
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: mupain-1



#### • Molecule 2: Urokinase-type plasminogen activator



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	122.11Å 122.11Å 42.55Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	60.00 – 2.28 35.25 – 2.28	Depositor EDS
% Data completeness (in resolution range)	99.4 (60.00-2.28) 99.4 (35.25-2.28)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.08 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.5.0110	Depositor
R, $R_{free}$	0.205 , 0.287 0.203 , 0.287	Depositor DCC
$R_{free}$ test set	511 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	38.5	Xtriage
Anisotropy	0.132	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 51.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.026 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2060	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.84% 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 [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	P	0.69	0/83	0.85	1/112 (0.9%)
2	U	0.83	0/2008	0.85	4/2720 (0.1%)
All	All	0.82	0/2091	0.85	5/2832 (0.2%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	U	181	LEU	CA-CB-CG	6.63	130.55	115.30
2	U	168	CYS	CA-CB-SG	-5.68	103.78	114.00
1	P	8	LEU	CA-CB-CG	5.66	128.32	115.30
2	U	172	TYR	N-CA-C	5.49	125.82	111.00
2	U	109	ARG	NE-CZ-NH1	5.07	122.84	120.30

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	P	81	0	73	11	0
2	U	1959	0	1906	51	1
3	U	20	0	0	1	0
All	All	2060	0	1979	53	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 13.

The worst 5 of 53 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)
2:U:116:ARG:HH21	2:U:116:ARG:HG2	1.28	0.98
2:U:116:ARG:HH21	2:U:116:ARG:CG	1.86	0.87
1:P:8:LEU:HB2	2:U:151:TYR:OH	1.73	0.87
2:U:53:ILE:HD11	2:U:103:ILE:HD11	1.62	0.80
2:U:72:ARG:NH2	2:U:153:GLU:HB3	1.97	0.79

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)
2:U:76:ASN:OD1	2:U:206:ARG:NH1[8_444]	2.07	0.13

## 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	P	8/10 (80%)	8 (100%)	0	0	100	100
2	U	246/247 (100%)	239 (97%)	6 (2%)	1 (0%)	38	45
All	All	254/257 (99%)	247 (97%)	6 (2%)	1 (0%)	38	45

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	U	172	TYR

### 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	P	9/9 (100%)	8 (89%)	1 (11%)	7	7
2	U	216/215 (100%)	197 (91%)	19 (9%)	12	13
All	All	225/224 (100%)	205 (91%)	20 (9%)	11	12

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	U	116	ARG
2	U	131	GLN
2	U	181	LEU
2	U	113	GLN
2	U	115	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	U	165	HIS
2	U	233	HIS
2	U	170	GLN
2	U	128	ASN
2	U	169	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 [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	P	10/10 (100%)	1.39	3 (30%) <b>1</b> <b>0</b>	59, 70, 73, 76	0
2	U	247/247 (100%)	0.03	8 (3%) 48 54	20, 42, 63, 75	0
All	All	257/257 (100%)	0.08	11 (4%) 36 41	20, 43, 66, 76	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	P	1	CYS	5.9
2	U	37(A)	ARG	5.0
2	U	132	PHE	3.5
2	U	166	ARG	3.4
2	U	37(B)	GLY	3.2

### 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.