W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Aug 15, 2022 – 10:04 am BST

PDB ID	:	7PYV
Title	:	Crystal structure of human UBA6 in complex with the ubiquitin-like modifier FAT10
Authors	:	Li, S.; Truongvan, N.; Schindelin, H.
Deposited on	:	2021-10-11
Resolution	:	3.27 Å(reported)

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

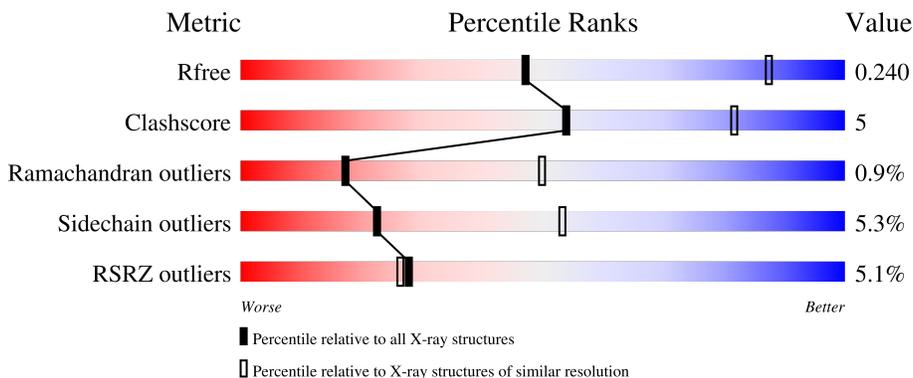|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 3.27 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	Percentile Ranks	Value
Rfree		0.240
Clashscore		5
Ramachandran outliers		0.9%
Sidechain outliers		5.3%
RSRZ outliers		5.1%

Worse ··· *Better*

▮ Percentile relative to all X-ray structures
◻ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1177 (3.32-3.24)
Clashscore	141614	1044 (3.30-3.26)
Ramachandran outliers	138981	1026 (3.30-3.26)
Sidechain outliers	138945	1025 (3.30-3.26)
RSRZ outliers	127900	1141 (3.32-3.24)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	1045	5% / 80% / 11% · 7%
1	B	1045	4% / 80% / 11% · 7%
2	C	159	6% / 89% / 7% ·· ·

W O R L D W I D E
PDB
PROTEIN DATA BANK

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 32955 atoms, of which 16452 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 Ubiquitin-like modifier-activating enzyme 6,Ubiquitin-like modifier-activating enzyme 1,Ubiquitin-like modifier-activating enzyme 6.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	967	Total 15195	C 4872	H 7567	N 1282	O 1437	S 37	0	0	0
1	B	971	Total 15258	C 4892	H 7600	N 1286	O 1443	S 37	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	625	ALA	CYS	conflict	UNP P22314
B	625	ALA	CYS	conflict	UNP P22314

- Molecule 2 is a protein called UBD.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	154	Total 2502	C 773	H 1285	N 214	O 226	S 4	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

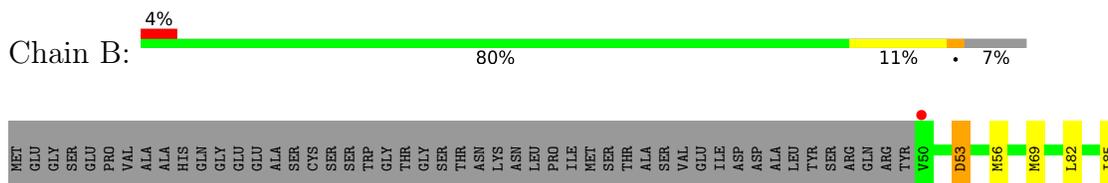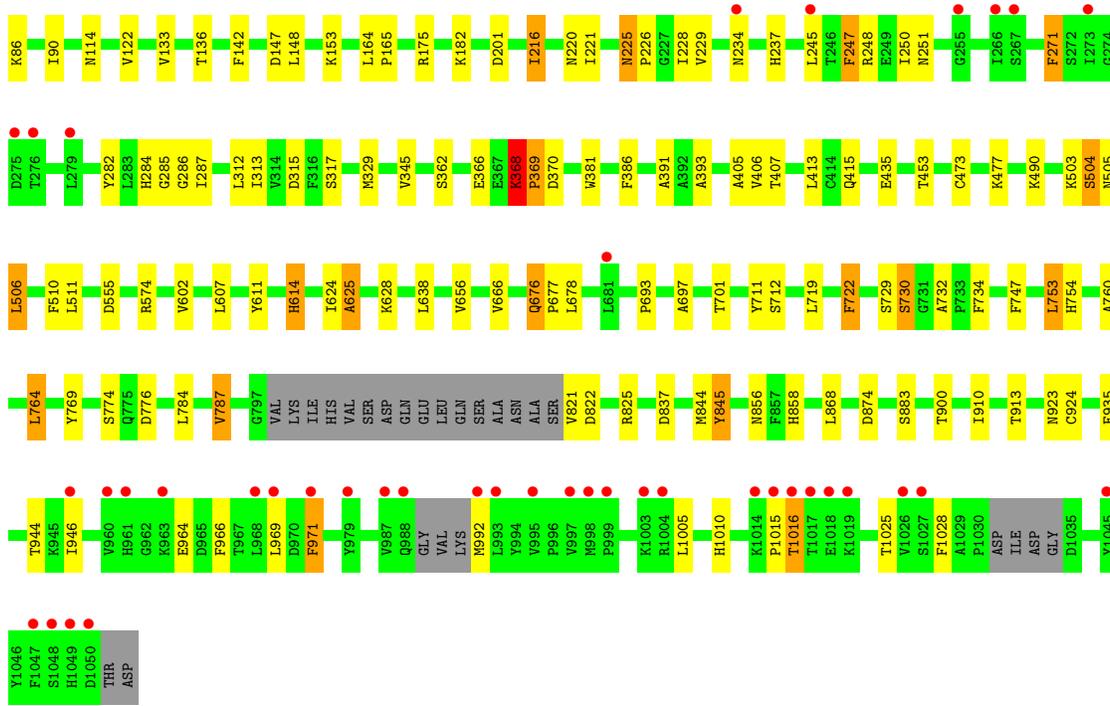| Chain	Residue	Modelled	Actual	Comment	Reference
C	7	THR	CYS	conflict	UNP A0A1U9X8S9
C	9	THR	CYS	conflict	UNP A0A1U9X8S9
C	134	LEU	CYS	conflict	UNP A0A1U9X8S9
C	162	SER	CYS	conflict	UNP A0A1U9X8S9

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.
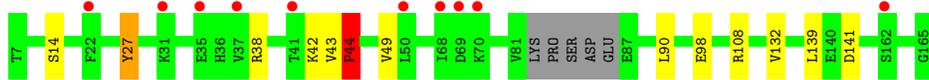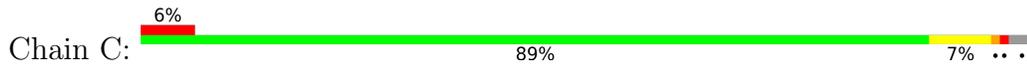
• Molecule 1: Ubiquitin-like modifier-activating enzyme 6,Ubiquitin-like modifier-activating enzyme 1,Ubiquitin-like modifier-activating enzyme 6

• Molecule 1: Ubiquitin-like modifier-activating enzyme 6,Ubiquitin-like modifier-activating enzyme 1,Ubiquitin-like modifier-activating enzyme 6

K86 I90 N114 V122 V133 T136 F142 D147 L148 K153 L164 P165 R175 K182 D201 I216 N220 I221 N225 P226 G227 I228 V229 N234 H237 L245 T246 F247 R248 E249 I250 N251 G255 I266 S267 F271 S272 I273 G274

D275 T276 L279 Y282 L283 H284 G285 G286 I287 L312 I313 V314 D315 F316 S317 M329 V345 S362 E366 E367 K368 P369 D370 W381 F386 A391 A392 A393 A405 V406 T407 L413 C414 Q415 E435 T453 C473 K477 K490 K503 S504 N505

L506 F510 L511 D555 R574 V602 L607 Y611 H614 I624 A625 K628 L638 V656 V666 Q676 P677 L678 L681 P693 A697 T701 Y711 S712 L719 F722 S729 S730 G731 A732 P733 F734 F747 L753 H754 A760

L764 L769 S774 Q775 D776 L784 V787 G797 VAL LYS ILE HIS VAL SER ASP GLN GLU LEU GLN SER ALA ASN ALA SER V821 D822 R825 D837 M844 Y845 N856 F857 H858 L868 D874 S883 T900 I910 T913 N923 C924 F935

T944 K945 I946 V960 H961 G962 K963 E964 D965 F966 T967 L968 L969 D970 F971 Y979 V987 Q988 GLY VAL LYS M992 L993 Y994 V995 P996 V997 M998 P999 K1003 R1004 L1005 H1010 K1014 P1015 T1016 T1017 E1018 K1019 T1025 V1026 S1027 F1028 A1029 P1030 ASP ILE ASP GLY D1035 Y1045

Y1046 F1047 S1048 H1049 D1050 THR ASP

• Molecule 2: UBD

Chain C: 6% 89% 7%

T7 S14 F22 Y27 K31 E35 H36 V37 R38 T41 K42 V43 P44 V49 L50 I68 D69 K70 V81 LYS PRO SER ASP GLU E87 L90 E98 R108 V132 L139 E140 D141 S162 G165

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	93.46Å 93.59Å 109.54Å 70.52° 88.54° 74.59°	Depositor
Resolution (Å)	45.14 – 3.27 46.09 – 3.27	Depositor EDS
% Data completeness (in resolution range)	59.0 (45.14-3.27) 59.1 (46.09-3.27)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.28 (at 3.25Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.216 , 0.239 0.216 , 0.240	Depositor DCC
R_{free} test set	875 reflections (2.84%)	wwPDB-VP
Wilson B-factor (Å^2)	91.8	Xtriage
Anisotropy	0.007	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$<|L|>$ = 0.47, $<L^2>$ = 0.29	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	32955	wwPDB-VP
Average B, all atoms (Å^2)	125.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.46% 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, 0.333 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	0.29	0/7810	0.47	0/10615
1	B	0.29	0/7840	0.47	0/10657
2	C	0.34	0/1233	0.55	1/1655 (0.1%)
All	All	0.30	0/16883	0.48	1/22927 (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	2
1	B	0	3
All	All	0	5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	44	PRO	CA-N-CD	-6.07	103.00	111.50

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	225	ASN	Peptide
1	A	368	LYS	Peptide
1	B	225	ASN	Peptide
1	B	368	LYS	Peptide
1	B	614	HIS	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	7628	7567	7567	86	0
1	B	7658	7600	7600	85	0
2	C	1217	1285	1285	5	0
All	All	16503	16452	16452	175	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.

The worst 5 of 175 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:C:43:VAL:HG12	2:C:44:PRO:HD2	1.38	1.04
1:B:284:HIS:CG	1:B:285:GLY:H	2.08	0.72
2:C:43:VAL:CG1	2:C:44:PRO:HD2	2.19	0.70
1:A:251:ASN:OD1	1:A:284:HIS:CD2	2.46	0.68
1:B:747:PHE:CE1	1:B:754:HIS:HB3	2.32	0.65

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	959/1045 (92%)	925 (96%)	25 (3%)	9 (1%)	17	50
1	B	963/1045 (92%)	929 (96%)	26 (3%)	8 (1%)	19	52
2	C	150/159 (94%)	143 (95%)	6 (4%)	1 (1%)	22	56

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	2072/2249 (92%)	1997 (96%)	57 (3%)	18 (1%)	17	50

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	368	LYS
1	A	369	PRO
1	A	615	ARG
2	C	44	PRO
1	B	368	LYS

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	848/913 (93%)	805 (95%)	43 (5%)	24	54
1	B	852/913 (93%)	807 (95%)	45 (5%)	22	53
2	C	138/143 (96%)	128 (93%)	10 (7%)	14	41
All	All	1838/1969 (93%)	1740 (95%)	98 (5%)	22	53

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	153	LYS
1	B	504	SER
1	B	228	ILE
1	B	317	SER
1	B	614	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	284	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 monosaccharides 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, 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	967/1045 (92%)	0.18	54 (5%)	24	23	29, 90, 223, 298	0
1	B	971/1045 (92%)	0.04	42 (4%)	35	33	29, 90, 223, 298	0
2	C	154/159 (96%)	0.36	10 (6%)	18	19	90, 150, 219, 238	0
All	All	2092/2249 (93%)	0.13	106 (5%)	28	26	29, 94, 223, 298	0

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	266	ILE	9.7
1	A	1016	THR	9.1
1	B	1016	THR	8.5
1	A	999	PRO	8.3
2	C	70	LYS	7.7

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 monosaccharides 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.