



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:04 PM GMT

PDB ID : 3VX8  
Title : Crystal structure of Arabidopsis thaliana Atg7NTD-Atg3 complex  
Authors : Matoba, K.; Fujioka, Y.; Noda, N.N.  
Deposited on : 2012-09-11  
Resolution : 3.11 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

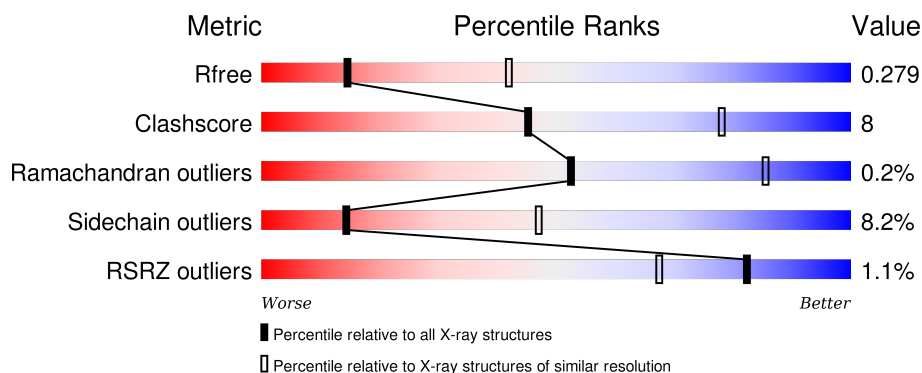
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.11 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	1112 (3.16-3.08)
Clashscore	102246	1218 (3.16-3.08)
Ramachandran outliers	100387	1175 (3.16-3.08)
Sidechain outliers	100360	1175 (3.16-3.08)
RSRZ outliers	91569	1114 (3.16-3.08)

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	323	<div> <div> <div></div> <div>73%</div> <div>18%</div> <div>• 6%</div> </div> </div>
1	D	323	<div> <div> <div></div> <div>72%</div> <div>19%</div> <div>• 7%</div> </div> </div>
2	B	292	<div> <div> <div></div> <div>48%</div> <div>11%</div> <div>• 40%</div> </div> </div>
2	C	292	<div> <div> <div></div> <div>47%</div> <div>10%</div> <div>• 41%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7608 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 Ubiquitin-like modifier-activating enzyme atg7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	300	Total	C	N	O	S	0	0	0
			2380	1533	396	444	7			
1	A	305	Total	C	N	O	S	0	0	0
			2418	1557	403	451	7			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	3	GLY	-	EXPRESSION TAG	UNP Q94CD5
D	4	PRO	-	EXPRESSION TAG	UNP Q94CD5
D	5	HIS	-	EXPRESSION TAG	UNP Q94CD5
D	6	MET	-	EXPRESSION TAG	UNP Q94CD5
A	3	GLY	-	EXPRESSION TAG	UNP Q94CD5
A	4	PRO	-	EXPRESSION TAG	UNP Q94CD5
A	5	HIS	-	EXPRESSION TAG	UNP Q94CD5
A	6	MET	-	EXPRESSION TAG	UNP Q94CD5

- Molecule 2 is a protein called Autophagy-related protein 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	175	Total	C	N	O	S	0	0	0
			1416	909	238	259	10			
2	C	172	Total	C	N	O	S	0	0	0
			1391	892	233	256	10			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	22	GLY	-	EXPRESSION TAG	UNP Q0WWQ1
B	23	PRO	-	EXPRESSION TAG	UNP Q0WWQ1
B	24	HIS	-	EXPRESSION TAG	UNP Q0WWQ1
B	25	MET	-	EXPRESSION TAG	UNP Q0WWQ1

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	22	GLY	-	EXPRESSION TAG	UNP Q0WWQ1
C	23	PRO	-	EXPRESSION TAG	UNP Q0WWQ1
C	24	HIS	-	EXPRESSION TAG	UNP Q0WWQ1
C	25	MET	-	EXPRESSION TAG	UNP Q0WWQ1

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O 1 1	0	0
3	B	1	Total O 1 1	0	0
3	C	1	Total O 1 1	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.52Å 132.68Å 102.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.85 – 3.11 34.85 – 3.11	Depositor EDS
% Data completeness (in resolution range)	99.1 (34.85-3.11) 99.2 (34.85-3.11)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.69 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.226 , 0.278 0.226 , 0.279	Depositor DCC
$R_{free}$ test set	1295 reflections (5.39%)	DCC
Wilson B-factor (Å <sup>2</sup> )	44.6	Xtriage
Anisotropy	0.076	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 20.7	EDS
Estimated twinning fraction	0.018 for l,-k,h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 25339 reflections	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	7608	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 3.51% 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, 0.375 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	0.50	0/2487	0.60	2/3383 (0.1%)
1	D	0.39	0/2447	0.54	0/3328
2	B	0.28	0/1452	0.51	0/1966
2	C	0.28	0/1425	0.49	0/1929
All	All	0.40	0/7811	0.55	2/10606 (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 (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	227	GLY	N-CA-C	6.10	128.35	113.10
1	A	309	LEU	CA-CB-CG	5.55	128.07	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	305	VAL	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	2418	0	2358	47	0
1	D	2380	0	2323	45	0
2	B	1416	0	1412	15	0
2	C	1391	0	1386	17	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
All	All	7608	0	7479	118	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 8.

The worst 5 of 118 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:D:210:SER:HB2	1:D:229:HIS:O	1.55	1.06
1:A:228:ASP:OD1	1:A:230:GLN:HB3	1.59	1.03
1:D:304:SER:O	1:D:305:VAL:HG23	1.65	0.97
2:B:193:ARG:HD2	2:B:222:LEU:HD11	1.51	0.92
1:A:230:GLN:HG3	1:A:230:GLN:O	1.72	0.89

There are no symmetry-related clashes.

## 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	299/323 (93%)	284 (95%)	14 (5%)	1 (0%)	46	81
1	D	294/323 (91%)	275 (94%)	18 (6%)	1 (0%)	46	81
2	B	169/292 (58%)	164 (97%)	5 (3%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	166/292 (57%)	161 (97%)	5 (3%)	0	100	100
All	All	928/1230 (75%)	884 (95%)	42 (4%)	2 (0%)	52	85

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	ARG
1	D	306	GLY

### 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	271/286 (95%)	247 (91%)	24 (9%)	12	42
1	D	267/286 (93%)	250 (94%)	17 (6%)	22	57
2	B	160/259 (62%)	144 (90%)	16 (10%)	9	34
2	C	157/259 (61%)	144 (92%)	13 (8%)	14	47
All	All	855/1090 (78%)	785 (92%)	70 (8%)	14	47

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	260	ARG
1	A	310	ASN
2	C	217	GLU
1	A	282	LEU
1	A	303	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	54	HIS
1	A	60	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	287	GLN
1	A	17	ASN
1	A	310	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	305/323 (94%)	-0.42	3 (0%) 84 70	18, 32, 56, 87	0
1	D	300/323 (92%)	0.03	3 (1%) 84 70	33, 56, 85, 108	0
2	B	175/292 (59%)	-0.38	1 (0%) 90 81	18, 33, 58, 89	0
2	C	172/292 (58%)	-0.18	3 (1%) 73 53	33, 49, 74, 102	0
All	All	952/1230 (77%)	-0.23	10 (1%) 82 67	18, 43, 79, 108	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	228	ASP	4.5
1	A	70	LEU	3.6
1	D	174	SER	3.4
1	D	162	ASP	3.3
2	C	277	GLU	2.9

### 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.