



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 04:54 am BST

PDB ID : 4GSL  
Title : Crystal structure of an Atg7-Atg3 crosslinked complex  
Authors : Kaiser, S.E.; Mao, K.; Taherbhoy, A.M.; Yu, S.; Olszewski, J.L.; Duda, D.M.;  
Kurinov, I.; Deng, A.; Fenn, T.D.; Klionsky, D.J.; Schulman, B.A.  
Deposited on : 2012-08-27  
Resolution : 2.70 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

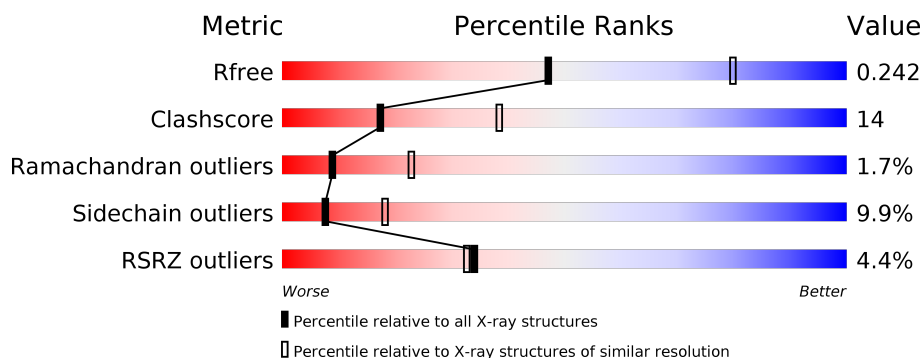
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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))
$R_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 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  $\leq 5\%$ . 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	615	<div> <div>2%</div> <div> <div></div> <div>69%</div> <div>26%</div> <div>• •</div> </div> </div>
1	B	615	<div> <div>3%</div> <div> <div></div> <div>67%</div> <div>25%</div> <div>• •</div> </div> </div>
2	C	312	<div> <div>6%</div> <div> <div></div> <div>34%</div> <div>21%</div> <div>• •</div> <div>40%</div> </div> </div>
2	D	312	<div> <div>6%</div> <div> <div></div> <div>34%</div> <div>21%</div> <div>•</div> <div>42%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 12543 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	A	598	Total	C	N	O	S	0	0	0
			4767	3064	804	875	24			
1	B	592	Total	C	N	O	S	0	0	0
			4717	3031	795	867	24			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP P38862
A	0	SER	-	EXPRESSION TAG	UNP P38862
B	-1	GLY	-	EXPRESSION TAG	UNP P38862
B	0	SER	-	EXPRESSION TAG	UNP P38862

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	186	Total	C	N	O	S	0	0	0
			1534	993	246	288	7			
2	D	180	Total	C	N	O	S	0	0	0
			1480	958	239	277	6			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	GLY	-	EXPRESSION TAG	UNP P40344
C	0	SER	-	EXPRESSION TAG	UNP P40344
C	41	ALA	CYS	ENGINEERED MUTATION	UNP P40344
C	76	ALA	CYS	ENGINEERED MUTATION	UNP P40344
C	83	ALA	CYS	ENGINEERED MUTATION	UNP P40344
D	-1	GLY	-	EXPRESSION TAG	UNP P40344
D	0	SER	-	EXPRESSION TAG	UNP P40344
D	41	ALA	CYS	ENGINEERED MUTATION	UNP P40344

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	76	ALA	CYS	ENGINEERED MUTATION	UNP P40344
D	83	ALA	CYS	ENGINEERED MUTATION	UNP P40344

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

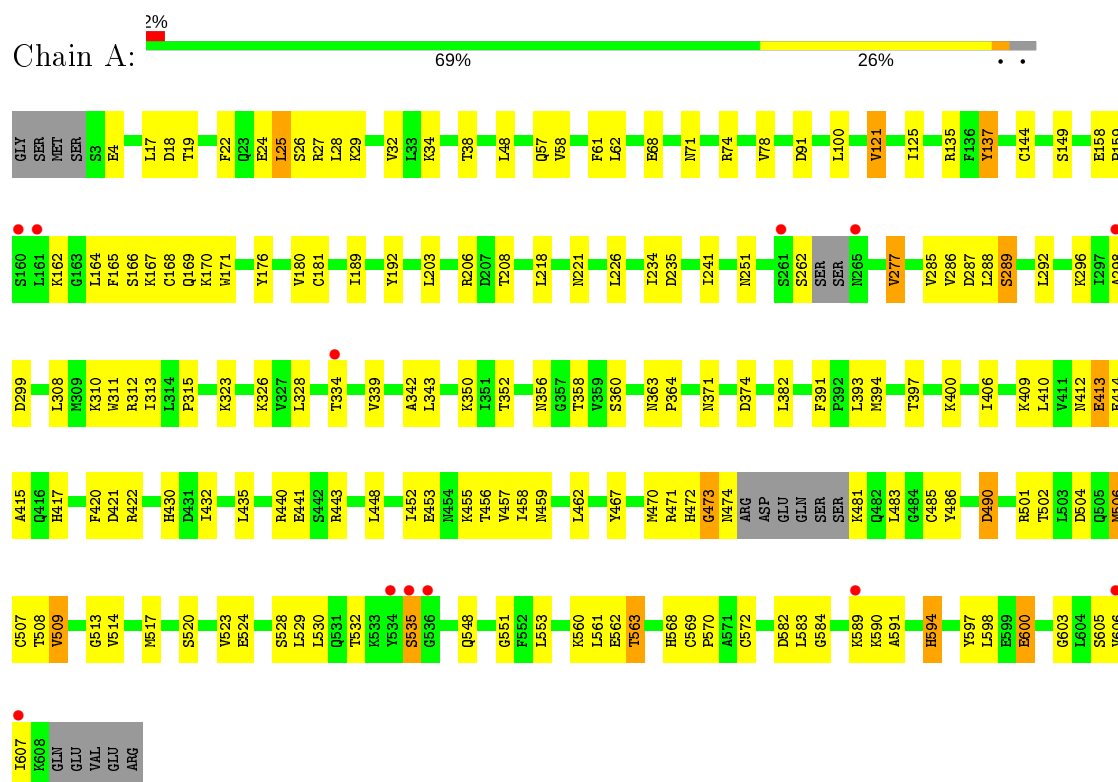
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	19	Total 19	O 19	0	0
4	B	19	Total 19	O 19	0	0
4	C	5	Total 5	O 5	0	0

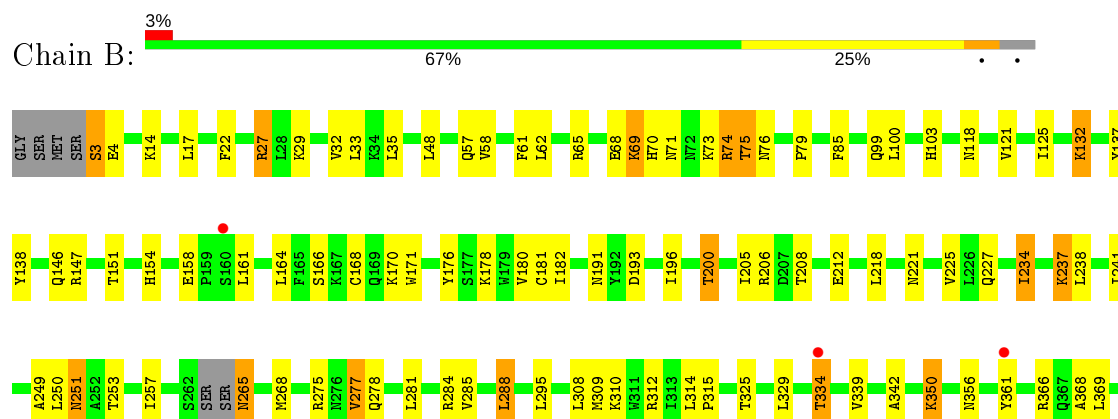
### 3 Residue-property plots [i](#)

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: Ubiquitin-like modifier-activating enzyme ATG7



#### • Molecule 1: Ubiquitin-like modifier-activating enzyme ATG7





H3O3
D3O4
TYR
THR
MET
GLU
GLY
TRP

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	281.56Å 125.08Å 71.08Å 90.00° 103.32° 90.00°	Depositor
Resolution (Å)	37.07 – 2.70 37.35 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.7 (37.07-2.70) 94.7 (37.35-2.70)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.69 (at 2.68Å)	Xtriage
Refinement program	PHENIX 1.7.1_743	Depositor
R, $R_{free}$	0.193 , 0.246 0.184 , 0.242	Depositor DCC
$R_{free}$ test set	1904 reflections (2.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.1	Xtriage
Anisotropy	0.254	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 55.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.046 for -h-2*k,l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12543	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.52% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.46	0/4869	0.61	0/6598
1	B	0.47	0/4819	0.60	0/6532
2	C	0.38	0/1570	0.63	0/2123
2	D	0.34	0/1515	0.55	0/2050
All	All	0.44	0/12773	0.60	0/17303

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
2	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	165	GLN	Peptide

### 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	A	4767	0	4833	116	0
1	B	4717	0	4766	126	0
2	C	1534	0	1514	64	0
2	D	1480	0	1462	54	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	19	0	0	0	0
4	B	19	0	0	1	0
4	C	5	0	0	1	0
All	All	12543	0	12575	345	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 14.

The worst 5 of 345 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:237:ALA:HB2	2:C:240:MET:HB3	1.40	1.02
2:D:218:LYS:HG3	2:D:225:SER:HB3	1.43	1.01
1:B:502:THR:HB	1:B:508:THR:HG21	1.49	0.94
1:B:427:ILE:HG23	1:B:433:ILE:HD11	1.51	0.91
1:A:287:ASP:OD1	1:A:289:SER:OG	1.88	0.90

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	592/615 (96%)	551 (93%)	36 (6%)	5 (1%)	19 43
1	B	586/615 (95%)	559 (95%)	23 (4%)	4 (1%)	22 46
2	C	178/312 (57%)	157 (88%)	13 (7%)	8 (4%)	2 5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	172/312 (55%)	156 (91%)	7 (4%)	9 (5%)	2	3
All	All	1528/1854 (82%)	1423 (93%)	79 (5%)	26 (2%)	9	23

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	232	HIS
2	D	180	ARG
2	D	245	ASP
1	A	27	ARG
1	A	473	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	540/556 (97%)	490 (91%)	50 (9%)	9	21
1	B	533/556 (96%)	482 (90%)	51 (10%)	8	19
2	C	172/282 (61%)	147 (86%)	25 (14%)	3	8
2	D	166/282 (59%)	152 (92%)	14 (8%)	11	25
All	All	1411/1676 (84%)	1271 (90%)	140 (10%)	8	18

5 of 140 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	164	LEU
1	B	393	LEU
2	D	70	ILE
1	B	170	LYS
1	B	251	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	57	GLN
1	B	71	ASN
2	D	67	GLN
1	A	594	HIS
1	B	548	GLN

### 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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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	598/615 (97%)	-0.18	12 (2%) 65 67	27, 60, 102, 132	0
1	B	592/615 (96%)	-0.16	17 (2%) 51 52	25, 59, 110, 157	0
2	C	186/312 (59%)	0.32	20 (10%) 5 4	38, 91, 138, 174	0
2	D	180/312 (57%)	0.34	19 (10%) 6 4	43, 93, 145, 160	0
All	All	1556/1854 (83%)	-0.05	68 (4%) 34 33	25, 65, 126, 174	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	164	ALA	9.9
2	C	245	ASP	7.4
2	C	281	LEU	6.4
1	B	498	LEU	5.1
1	B	494	PRO	5.0

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZN	A	700	1/1	0.98	0.15	74,74,74,74	0
3	ZN	B	700	1/1	0.99	0.12	86,86,86,86	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.