



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 6, 2023 – 07:09 PM EDT

PDB ID : 4GSK
Title : Crystal structure of an Atg7-Atg10 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 : 2.90 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.35
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) : 2.35

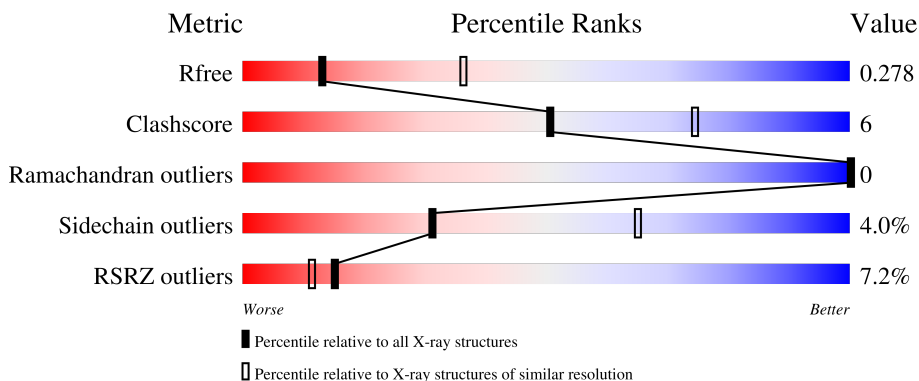
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 $\leq 5\%$. 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	
1	B	615	
2	Y	173	
2	Z	173	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 11408 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
			Total	C	N	O	S			
1	A	570	Total	C	N	O	S	0	0	0
			4548	2930	768	829	21			
1	B	572	Total	C	N	O	S	0	0	0
			4573	2946	774	832	21			

There are 10 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
A	39	SER	CYS	engineered mutation	UNP P38862
A	195	SER	CYS	engineered mutation	UNP P38862
A	375	ALA	CYS	engineered mutation	UNP P38862
B	-1	GLY	-	expression tag	UNP P38862
B	0	SER	-	expression tag	UNP P38862
B	39	SER	CYS	engineered mutation	UNP P38862
B	195	SER	CYS	engineered mutation	UNP P38862
B	375	ALA	CYS	engineered mutation	UNP P38862

- Molecule 2 is a protein called Ubiquitin-like-conjugating enzyme ATG10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	Y	137	Total	C	N	O	S	0	0	0
			1161	758	190	210	3			
2	Z	131	Total	C	N	O	S	0	0	0
			1124	737	184	200	3			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Y	-5	GLY	-	expression tag	UNP Q07879
Y	-4	SER	-	expression tag	UNP Q07879

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
Y	-3	GLY	-	expression tag	UNP Q07879
Y	-2	GLY	-	expression tag	UNP Q07879
Y	-1	SER	-	expression tag	UNP Q07879
Y	0	GLY	-	expression tag	UNP Q07879
Y	26	SER	CYS	engineered mutation	UNP Q07879
Y	137	SER	CYS	engineered mutation	UNP Q07879
Z	-5	GLY	-	expression tag	UNP Q07879
Z	-4	SER	-	expression tag	UNP Q07879
Z	-3	GLY	-	expression tag	UNP Q07879
Z	-2	GLY	-	expression tag	UNP Q07879
Z	-1	SER	-	expression tag	UNP Q07879
Z	0	GLY	-	expression tag	UNP Q07879
Z	26	SER	CYS	engineered mutation	UNP Q07879
Z	137	SER	CYS	engineered mutation	UNP Q07879

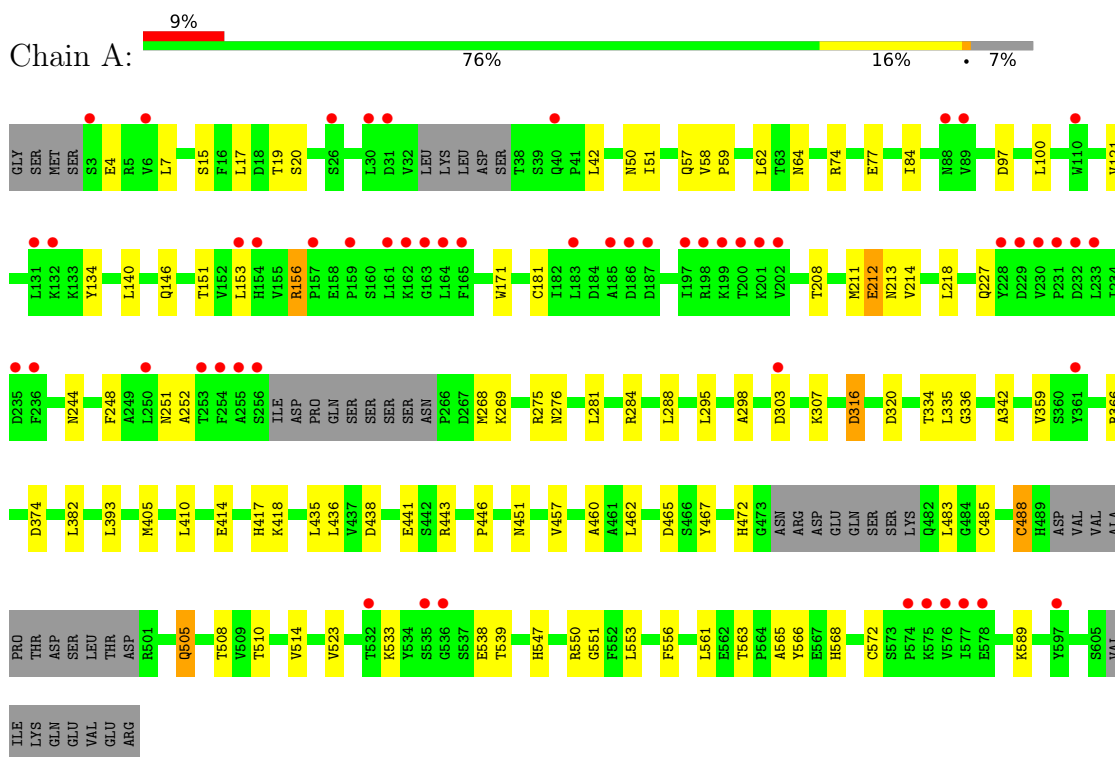
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

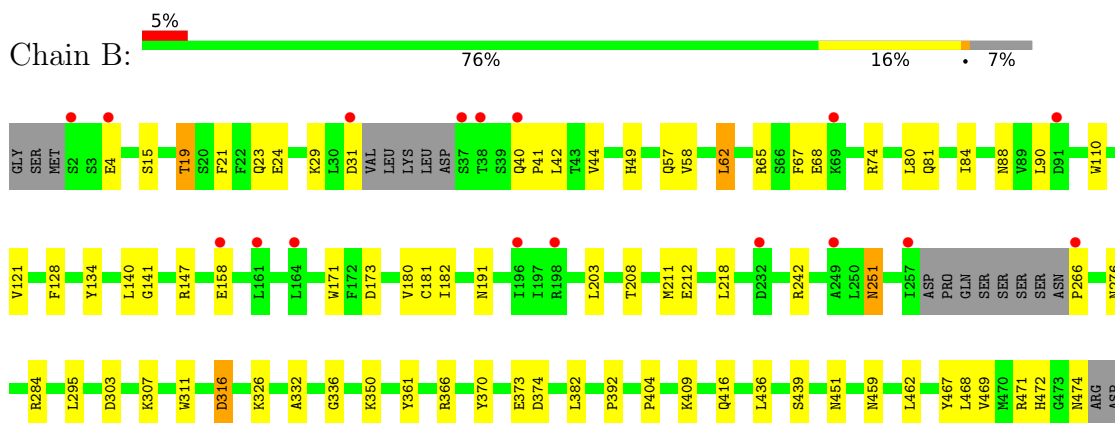
3 Residue-property plots

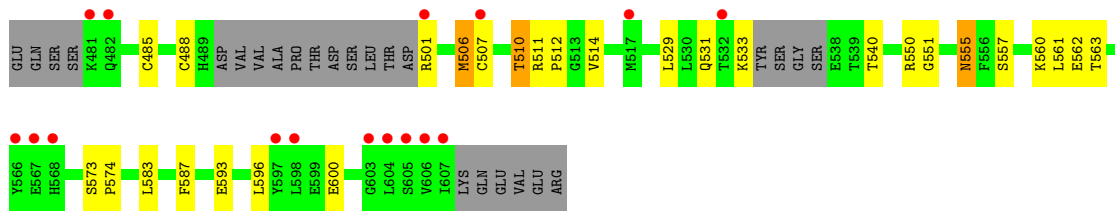
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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 ATG7

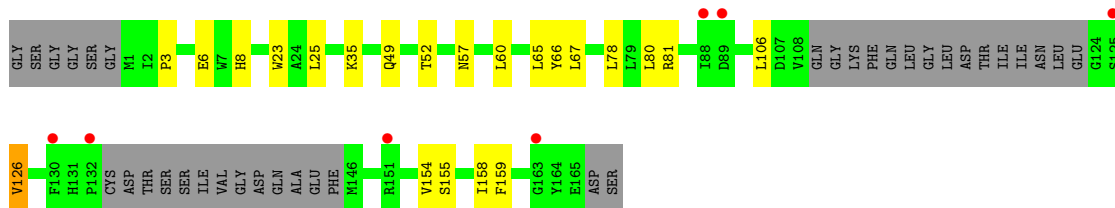


- Molecule 1: Ubiquitin-like modifier-activating enzyme ATG7

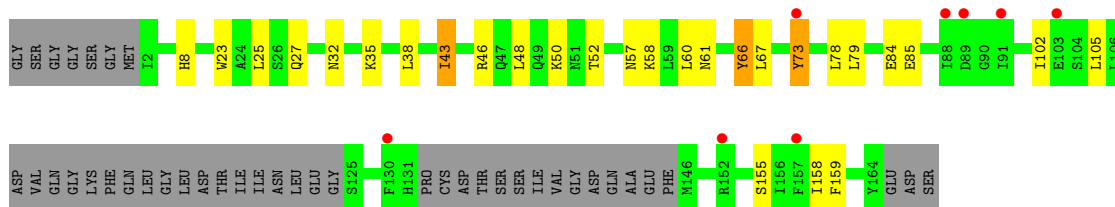




● Molecule 2: Ubiquitin-like-conjugating enzyme ATG10



● Molecule 2: Ubiquitin-like-conjugating enzyme ATG10



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	120.56Å 146.19Å 108.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.56 – 2.90 49.57 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.9 (49.56-2.90) 92.4 (49.57-2.90)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.92 (at 2.91Å)	Xtrriage
Refinement program	PHENIX dev_1000	Depositor
R, R_{free}	0.236 , 0.285 0.230 , 0.278	Depositor DCC
R_{free} test set	2154 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	51.5	Xtrriage
Anisotropy	0.398	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 53.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	11408	wwPDB-VP
Average B, all atoms (Å ²)	79.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.86% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.26	0/4646	0.45	0/6291
1	B	0.26	0/4669	0.46	0/6318
2	Y	0.25	0/1193	0.45	0/1621
2	Z	0.26	0/1155	0.44	0/1568
All	All	0.26	0/11663	0.45	0/15798

There are no bond length outliers.

There are no bond angle outliers.

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	A	4548	0	4604	52	0
1	B	4573	0	4648	49	0
2	Y	1161	0	1139	14	0
2	Z	1124	0	1114	18	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
All	All	11408	0	11505	130	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 6.

The worst 5 of 130 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:Y:49:GLN:HE22	2:Y:60:LEU:H	1.19	0.89
1:A:15:SER:HB3	1:A:140:LEU:HD22	1.59	0.84
2:Y:81:ARG:HH21	2:Y:126:VAL:HB	1.56	0.70
1:B:15:SER:HB3	1:B:140:LEU:HD22	1.77	0.66
2:Z:61:ASN:HD22	2:Z:85:GLU:HG2	1.62	0.64

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	560/615 (91%)	549 (98%)	11 (2%)	0	100	100
1	B	560/615 (91%)	550 (98%)	10 (2%)	0	100	100
2	Y	131/173 (76%)	125 (95%)	6 (5%)	0	100	100
2	Z	125/173 (72%)	121 (97%)	4 (3%)	0	100	100
All	All	1376/1576 (87%)	1345 (98%)	31 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	510/555 (92%)	492 (96%)	18 (4%)	36	70
1	B	515/555 (93%)	491 (95%)	24 (5%)	26	59
2	Y	130/160 (81%)	126 (97%)	4 (3%)	40	74
2	Z	127/160 (79%)	122 (96%)	5 (4%)	32	66
All	All	1282/1430 (90%)	1231 (96%)	51 (4%)	31	65

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	251	ASN
1	B	506	MET
2	Z	73	TYR
1	B	316	ASP
1	B	474	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	474	ASN
1	B	555	ASN
2	Z	61	ASN
1	B	548	GLN
2	Y	5	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 monosaccharides 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, 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	570/615 (92%)	0.66	54 (9%) 8 6	61, 80, 97, 105	0
1	B	572/615 (93%)	0.53	33 (5%) 23 19	60, 76, 94, 104	0
2	Y	137/173 (79%)	0.49	7 (5%) 28 24	66, 83, 96, 103	0
2	Z	131/173 (75%)	0.52	8 (6%) 21 17	68, 79, 93, 100	0
All	All	1410/1576 (89%)	0.58	102 (7%) 15 11	60, 78, 96, 105	0

The worst 5 of 102 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	37	SER	6.0
1	B	481	LYS	5.2
1	B	38	THR	5.1
1	A	255	ALA	4.6
1	A	597	TYR	4.5

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 monosaccharides 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, 95th 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(\AA^2)	Q<0.9
3	ZN	B	701	1/1	0.91	0.21	92,92,92,92	0
3	ZN	A	701	1/1	0.98	0.28	88,88,88,88	0

6.5 Other polymers [i](#)

There are no such residues in this entry.