STRUCTURAL BIOLOGY
COMMUNICATIONS

Volume 73 (2017)

Supporting information for article:

Identification, biochemical characterization and crystallization of the central region of human ATG16L1

Archna Archna and Andrea Scrima

Table S1 Data collection and processing

Values for the outer shell are given in parentheses.

PDB code	5NPV	5NPW
Data collection		
Diffraction source	RIGAKU MICROMAX-007HF	SLS BEAMLINE X06DA
Wavelength (Å)	1.5418	1.0000
Space group	I4	C2
Temperature (K)	100	100
a, *b*, *c* (Å)	143.58, 143.58, 62.32	200.86, 75.99, 142.44
α, β, γ (°)	90, 90, 90	90, 131.415, 90
Resolution range (Å)	45.4 - 3.1 (3.2 - 3.1)	48.3 - 3.1 (3.2 - 3.1)
Total No. of reflections	64354 (5930)	108826 (10152)
No. of unique reflections	11680 (1069)	29428 (2643)
Completeness (%)	99.6 (100.0)	99.1 (99.3)
Multiplicity	5.51 (5.55)	3.70 (3.84)
$R_{r.i.m.}$	0.185 (0.799)	0.147 (0.865)
$\langle I/\sigma(I) \rangle$	11.45 (2.22)	9.48 (2.08)
R_{merge}	0.167 (0.723)	0.126 (0.745)
Refinement		
Resolution (Å)	45.4 - 3.1	48.3 - 3.1
R_{work}/R_{free}	25.29/27.81	21.75/26.39
No. of non-hydrogen atoms (total/protein)	4634/4623	9952/9945
R.m.s. deviation		
Bond angles (°)	0.003	0.003
Bond lengths (Å)	0.489	0.504
Average B factor protein (Å^2)	45.7	63.3
Ramachandran plot		
Most favored	94.99	97.23
Additionally allowed	5.01	2.77

$R_{merge} = \sum_{hkl}\sum_j |I_{hkl,j} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_j I_{hkl,j}$ and $R_{r.i.m} = \sum_{hkl}\sqrt{(n/(n-1)} \sum^{n}_{j=1}|I_{hkl,j} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_j I_{hkl,j}$. With I being the measured intensity and $\langle I \rangle$ being the averaged intensity of each unique reflection with indices *hkl*. $I/\sigma(I)$ corresponds to the average of the intensity divided by its average standard deviation. $R_{work/free} = \sum_{hkl} \big| |F_{obs}| - |F_{calc}| \big| / \sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} is the same as R_{work}, calculated for the 5% of the data that was randomly omitted from refinement.

Table S2 List of r.m.s.d. of all Cα atoms between ATG5/ATG16L1 structures from this work and previously solved structures of ATG12-ATG5/ATG16L1 or ATG5/ATG16L1.

Protein complexes (PBD code, space groups)	r.m.s.d. (all Cα) between dimers
ATG5/ATG16L1 structures from this work (5NPW, C2 and 5NPV, I4)	1.458 Å
ATG5/ATG16L1 (5NPW, C2)§ and ATG12-ATG5/ATG16L1* (4GDL, C2)	0.484 Å
ATG5/ATG16L1 (5NPW, C2)§ and ATG12-ATG5/ATG16L1* (4GDK, $P2_12_12_1$)	0.953 Å
ATG5/ATG16L1 (5NPW, C2)§ and ATG5/ATG16L1# (5D7G, C2)	1.321 Å
ATG5/ATG16L1 (5NPW, C2)§ and ATG5/ATG16L1& (4TQ0, $P4_12_12$)	0.944 Å
ATG5/ATG16L1 (5NPV, I4)§ and ATG12-ATG5/ATG16L1* (4GDL, C2)	1.258 Å
ATG5/ATG16L1 (5NPV, I4)§ and ATG12-ATG5/ATG16L1* (4GDK, $P2_12_12_1$)	2.053 Å
ATG5/ATG16L1 (5NPV, I4)§ and ATG5/ATG16L1# (5D7G, C2)	2.217 Å
ATG5/ATG16L1 (5NPV, I4)§ and ATG5/ATG16L1& (4TQ0, $P4_12_12$)	1.399 Å

§ structures solved in this work

*structures from (Otomo *et al*., 2013)

#structures from (Kim *et al*., 2016)

&structure from (Kim *et al*., 2015)

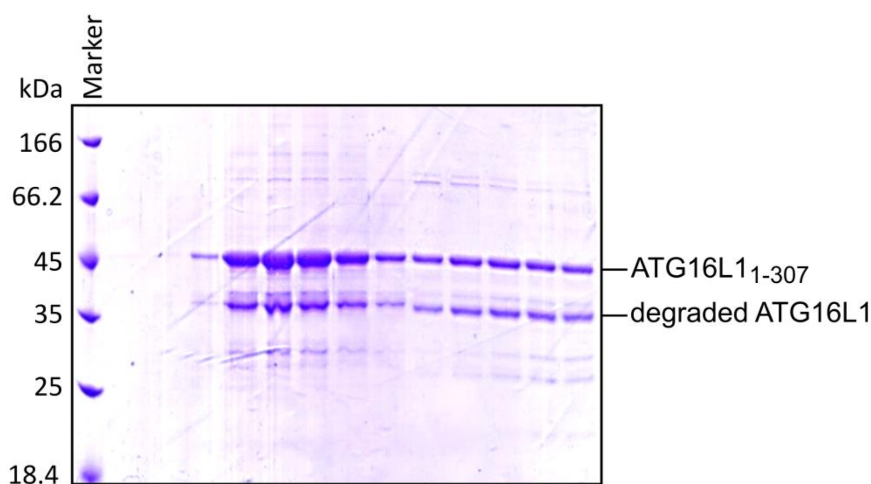


Figure S1 Coomassie stained SDS-PAGE gel showing the protein fractions from size exclusion chromatography (SEC) of ATG16L1$_{1-307}$. ATG16L1$_{1-307}$ was cloned into a modified pCDF Duet vector similar to ATG16L1$_{11-307}$ and was expressed in fusion with an N-terminal His$_6$ tag. Protein was expressed in TB media by induction with 0.5 mM IPTG at OD$_{600}$ 0.8 followed by incubation at 37°C for 3 h. Cell lysate was prepared as described for ATG16L1$_{72-307}$. Ni-sepharose resin (4 ml) was incubated with the clear lysate followed by washing the beads with wash buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM DTT and 30 mM Imidazole) and eluting the protein with elution buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM DTT and 250 mM Imidazole). Finally, His$_6$-ATG16L$_{11-307}$ was purified by size exclusion chromatography on Superdex75 16/600 column. Protein fractions corresponding to the elution peak of SEC were analyzed on 12 % SDS-PAGE. All the fractions were consisting of two protein bands; ATG16L1$_{1-307}$ and a lower MW band for degradation product.

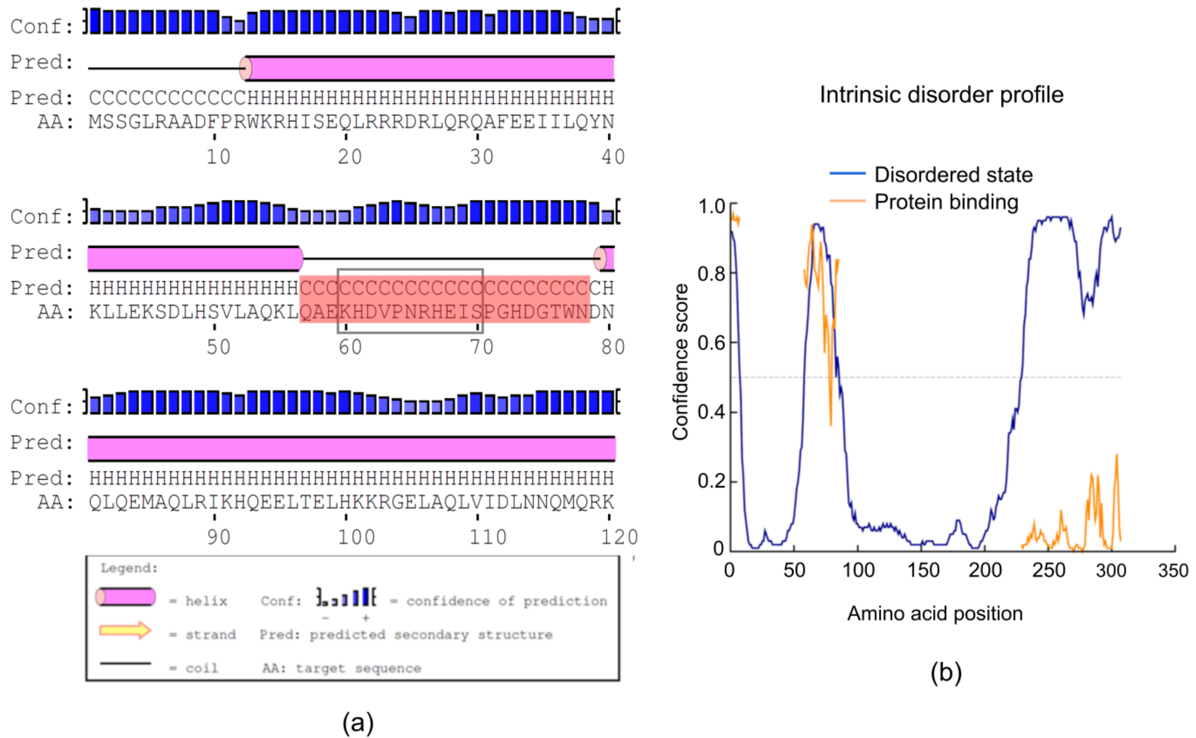


Figure S2 (a) Secondary structure prediction of N-terminal ATG16L1 region (aa 1-120) including the region close to aa 60-70 predicted to lack any secondary structure. Red shaded part of the sequence is showing the flexible part based on secondary structure prediction and the sequence in black box is representing the unstable region identified in this study. (b) Intrinsic disorder prediction of N-terminal ATG16L1. Only N-terminal sections of the ATG16L1 full-length sequence are shown.

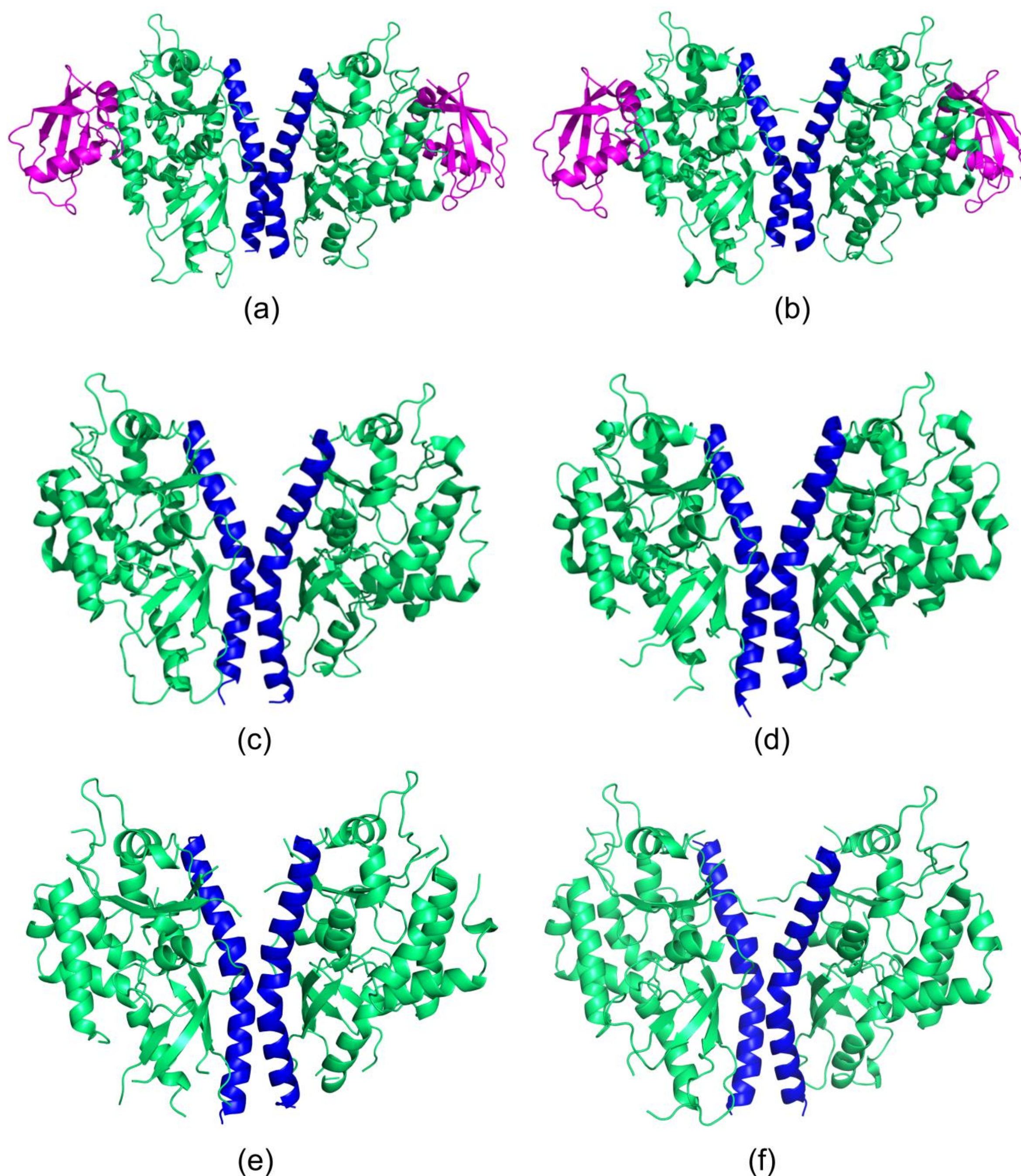
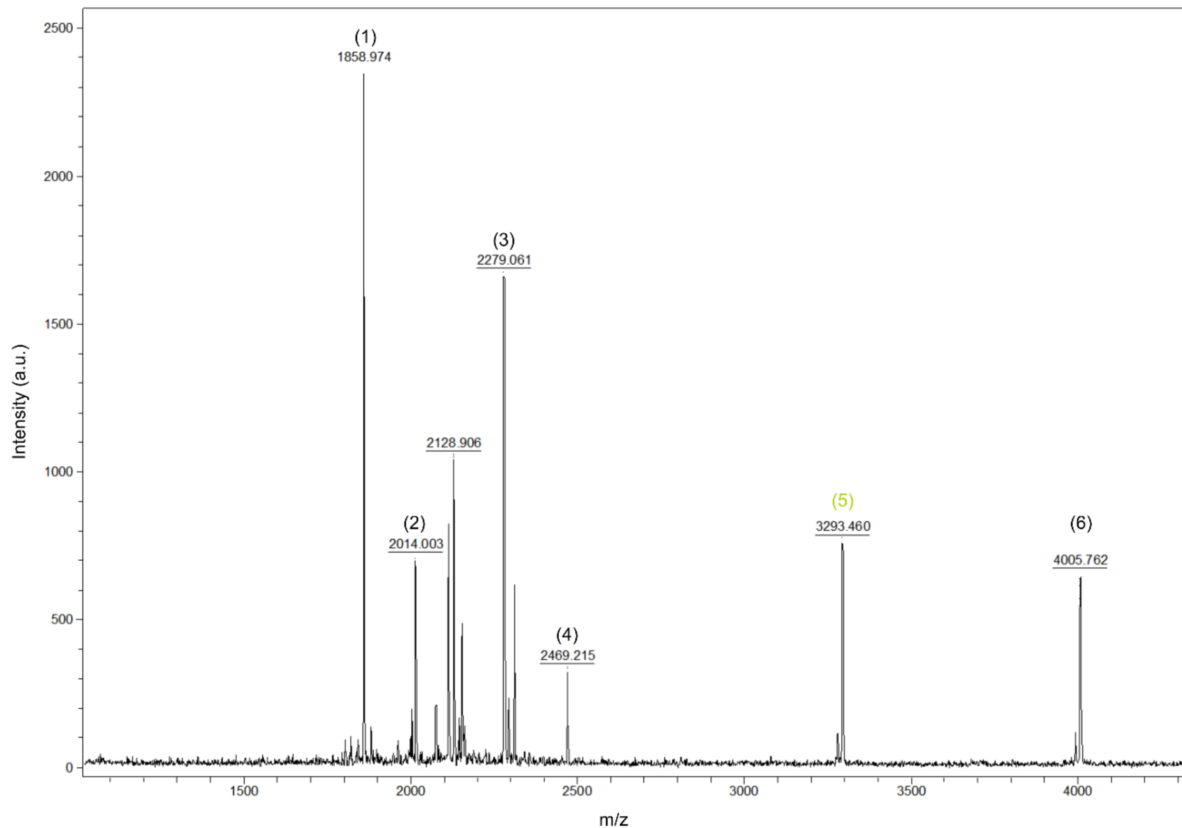


Figure S3 Dimerization of ATG5/ATG16L1 via the ATG5BD. Overview of structures of human ATG5/ATG16L1 complexes and human ATG12~ATG5/ATG16L complexes. (a) PDB code 4GDK, (b) PDB code 4GDL, (c) PDB code 4GD7 and (d) PDB code 4TQ0 are previously solved structures (e) and (f) are PDB code 5NPV and 5NPW, the structures solved in this work. ATG5 is shown in green, ATG16L1 in blue and ATG12 in magenta.



1 MSSGLRAADF PRWKRHISEQ LRRRDRLQRQ AFEEIILQYN KLLEKSDLHS

51 VLAQKLQAEK HDVPNRHEIS PGHDGTWNDN QLQEMAQLRI KHQEELTELH

101 KK**RGELAQLV IDLNNQMQR**K DREMQMNEAK **IAECLQTISD LETECLDLR**T

151 KLCDLER**ANQ TLKDEYDALQ ITFTALEGK**L RKTTEENQEL VTRWMAEKAQ

201 EANRLNAENE KDSRRRQARL QK**ELAEAAKE PLPVEQDDDI EVIVDETSDH**

251 **TEETSPVR**AI SRAATK**RLSQ PAGGLLDSIT NIFG**RRSVSS FPVPQDNVDT

301 HPGSGKEVRV PATALCVFDA HDGEVNAVQF SPGSRLLATG GMDRRVKLWE

(2) (3) (4) (6) (5) (1)

Figure S4 MALDI-MS analysis of dissolved crystals. In-gel tryptic digestion of the protein was performed and resulting peptide sequences were determined by MS. Mascot protein database was used to identify the protein. Bold red peptide sequences are indicating the matching peptide sequences from MS spectra. Each MS-peak and corresponding peptide is represented with respective number. Boundaries of the ATG16L1$_{72\text{-}307}$ construct used for crystallization are highlighted with a grey background.