Universidade de Aveiro Departamento de Biologia
2012

Ana Marisa Henriques Duarte Simões

Caraterização estrutural de proteínas associadas a autofagia

Structural characterization of proteins associated to autophagy

Universidade de Aveiro Departamento de Biologia
2012

Ana Marisa Henriques Duarte Simões

Caraterização estrutural de proteínas associadas a autofagia

Structural characterization of proteins associated to autophagy

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Sandra de Macedo Ribeiro, Professora auxiliar convidada do Departamento de Biologia da Universidade de Aveiro

Apoio financeiro do Programa de Cooperação Territorial SUDOE

o júri

presidente

Prof. Doutora Maria do Céu Gomes dos Santos
Professora auxiliar convidada do Departamento de Biologia da Universidade de Aveiro

Prof. Doutora Sandra de Macedo Ribeiro
Professora auxiliar convidada do Departamento de Biologia da Universidade de Aveiro
Investigadora principal no Instituto de Biologia Molecular e Celular, Porto

Doutor Frederico Silva
Técnico Superior de 1ª classe no Instituto de Biologia Molecular e Celular, Porto

palavras-chave

DOR, GATE16, LC3, domínio LIR, autofagia, interação, ressonância plasmónica de superfície, circular dicroísmo, estabilidade térmica, cristalização, estrutura

resumo

DOR (ou Tumor protein p53 inducible nuclear protein 2 - tp53inp2) é uma proteína bifuncional que atua no núcleo e no citosol. No núcleo DOR atua como um co-fator nuclear, liga-se e co-ativa no recetor da hormona da tiróide. No último, DOR desloca-se do núcleo para o citoplasma em situações de ativação da autofagia ou stress celular, localiza-se no autofagossoma e interage diretamente com as proteínas associadas à membrana do autofagossoma, LC3 e GATE16. A caraterização da interação entre a DOR e os seus parceiros e a relevância da DOR na autofagia é muito importante. A autofagia tem um papel importante no envelhecimento, morte celular, defesa contra agentes intracelulares patogénicos, doenças neurodegenerativas e tumorogenesis, o que demonstra a importância biológica e médica de estudar as proteínas envolvidas neste processo.

A proteína de fusão NusA-DOR e os seus interatores, LC3 e GATE16, foram expressos em *E.coli*. Todas as proteínas foram purificadas por cromatografia de afinindade, seguida por cromatografia de exclusão molecular (DOR) ou por cromatografia de troca iónica (LC3 e GATE16). A estabilidade da DOR e a interação com os seus parceiros intracelulares foi analisada estruturalmente, através de ressonância plasmónica de superfície, circular dicroísmo e estabilidade térmica. Um péptido da DOR contendo o local de interação (região LIR) foi produzido para os ensaios de co-cristalização por difusão vapor. O péptido da DOR liga num sulco da LC3 numa conformação em gancho, dois importantes aminoácidos medeiam a interação com LC3, Trp^{35} e a Leu^{38}. A conformação desta estrutura é diferente das outas estruturas conhecidas da LC3 com domínios LIR.

keywords

DOR, GATE16, LC3, LIR motif, autophagy, interaction, surface plasmon resonance, circular dichroism, thermal stabilty, crystalliztion, structure

abstract

DOR (or Tumor protein p53 inducible nuclear protein 2 - tp53inp2) is a bifunctional protein that operates both in the nucleus and in the cytosol. In the nucleus, DOR acts as a nuclear co-factor, and binds to and co-activates the thyroid hormone receptor. In the later, DOR moves from the nucleus to the cytoplasm under conditions characterized by the activation of autophagy or cellular stress and can be localized to early autophagosome and interact directly with the autophagosome membrane associated protein LC3 and GATE16. Characterization of the interaction between DOR and its interacting partners is very important to understand the relevance of DOR in autophagy. Since autophagy plays a protective role in aging, cell death, defense against intracellular pathogens, neurodegenerative diseases and tumorogenesis, studying DOR might have a large biological and medical relevance.

The fusion protein NusA-DOR and its interactors, LC3 and GATE16, were expressed in *E.coli*. All the proteins were purified by affinity chromatography, followed by size exclusion chromatography (DOR) or ion exchange chromatography, (LC3 and GATE16). The stability of DOR and the interaction with intracellular partners has been structurally analyzed, by surface plasmon resonance, circular dichroism and differential scanning fluorimetry. A DOR peptide containing the interaction site (LIR motif) has been produced for co-crystallization experimentss. The DOR peptide binds within LC3 groove in a hairpin conformation, two important amino acids, Trp^{35} and Leu^{38} mediated the insertion into pockets of LC3. This peptide displays a new conformation, when compared with the known three-dimensional strutures of LC3:LIR complexes.

Table of Contents

Abbreviations

Arg	Arginine
Asp	Aspartic acid
AU	Absorbance units
CD	Circular dichroism
cDNA	Complementary deoxyribonucleic acid
C-terminus	Carboxyl-terminus
DLS	Dynamic Light Scattering
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxy Nucleotides
DOR	Diabetes- and obesity-regulated gene
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GATE16	Golgi-associated ATPase enhancer of 16 kDa
Glut	Glutamic acid
His	Histidine
Ile	Isoleucine
IMAC	Ion Immobilized Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilodaltons
LB	Luria-Bertani
LC3	Autophagy-related protein LC3 B
Leu	Leucine
LIR	LC3 interacting region
Lys	Lysine
MES	2-Morpholinoethanesulfonic acid monohydrate
Min	Minutes
N-terminus	Amino-terminus
OD	Optical density
PCR	Polymerase chain reaction

PEG	Poly(ethanol) glycol
Phe	Phenylalanine
pI	Isoelectric point
Pro	Proline
PVDF	Polyvinylidene Fluoride
rpm	Rotation per minute
RT-PCR	Real time Polymerase Chain Reaction (machine)
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate –polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SPR	Surface plasmon resonance
TAE	Tris, acetic acid; EDTA
TEV	Tobaco Etch Virus
TFE	Trifluoroethanol
T_m	Melting temperature
TMAO	Trimethylamine *N*-oxide
Trp	Tryptophan
Trx	Thioredoxin
Val	Valine

1. Introduction

1.1. The Human DOR protein

The diabetes- and obesity-regulated gene (DOR), also known as “tumor protein p53-inducible nuclear protein 2” (TP53INP2) is a novel protein abundantly expressed in tissues with high metabolic rates such as skeletal muscle and heart, which was orginally identified in Zucker diabetic fatty (ZDF) rats, and located in promyelocytic leukemia nuclear bodies under basal conditions [1].

DOR was also identified in the locus for autosomal recessive infantile cerebellar ataxia (CLA3 or SCAR6) in chromosome 20q11-q13. The expression of Tp53inp2 in some regions of the developing nervous system suggests that this gene may be involved in certain aspects of neuronal development and in mammalian neurogenesis [2].

Human DOR it is composed of five exons spanning 9 kb of genomic DNA. The predicted open reading frame of *TP53INP2* cDNA codes for 220 amino acids with a predicted molecular weight of 24kDa and located at human chromosome 20q11.2. *TP53INP2* shows a conserved exon-intron structure with the *TP53INP1* [3].

The only homologous protein described to date is a human p53-dependent apoptosis regulator *TP53INP1* with which it shares 36% amino acid identity [1]. There are two *TP53INP1* isoforms, *TP53INP1*α and *TP53INP1*β, like DOR this protein is located in the nucleus. *TP53INP1* must be involved in 53-mediated cell death, since their synthesis is increased during cellular stress by p53-mediated activation of transcription [4]. DOR and *TP53INP1* sequences are phylogenetically related, there are two clusters corresponding to both sequences in vertebrates spanning from cartilaginous and bony fish to mammals (Figure 1) [5].

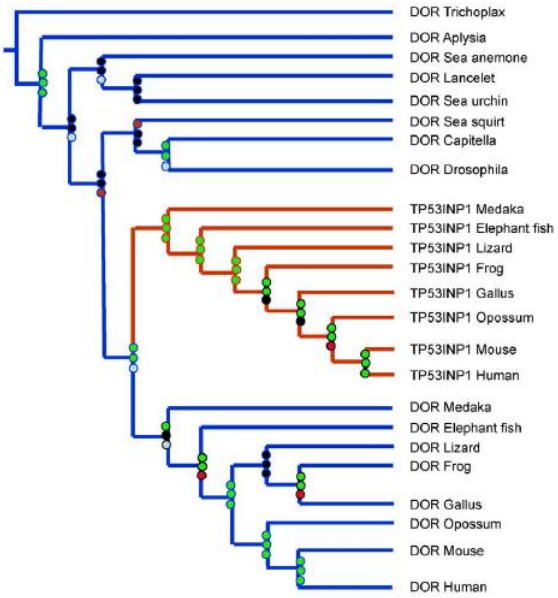


Figure 1- Phylogenetic tree showing two clusters corresponding to DOR and *Tp53inp1* vertebrate sequences and indicating a duplication of the ancestral DOR gene in the common ancestor of vertebrates (adapted from Sancho, A. et al., 2012) [5].

Human DOR is well conserved on other homologous (48.5 ±11.6%) But, the only homologous protein described to date *TP53INP1* only have 32.9 ±11.6% identify with human DOR. In terms of structural organization, this protein contains two functionally distinct regions and unique to the DOR protein family. These regions localize in residues 28-42 (region 1) and residues 66-112 (region 2) of human DOR [5]. A nuclear export sequence (NES) was identified in region 1, in residues 32 to 40, responsible for cytoplasmic dislocation. The N-terminal half is predicted to have random structure (Figure 2), however, the C-terminal region of DOR is predicted to form a positively charged alpha-helix structure (Figure 2) [1], [5].

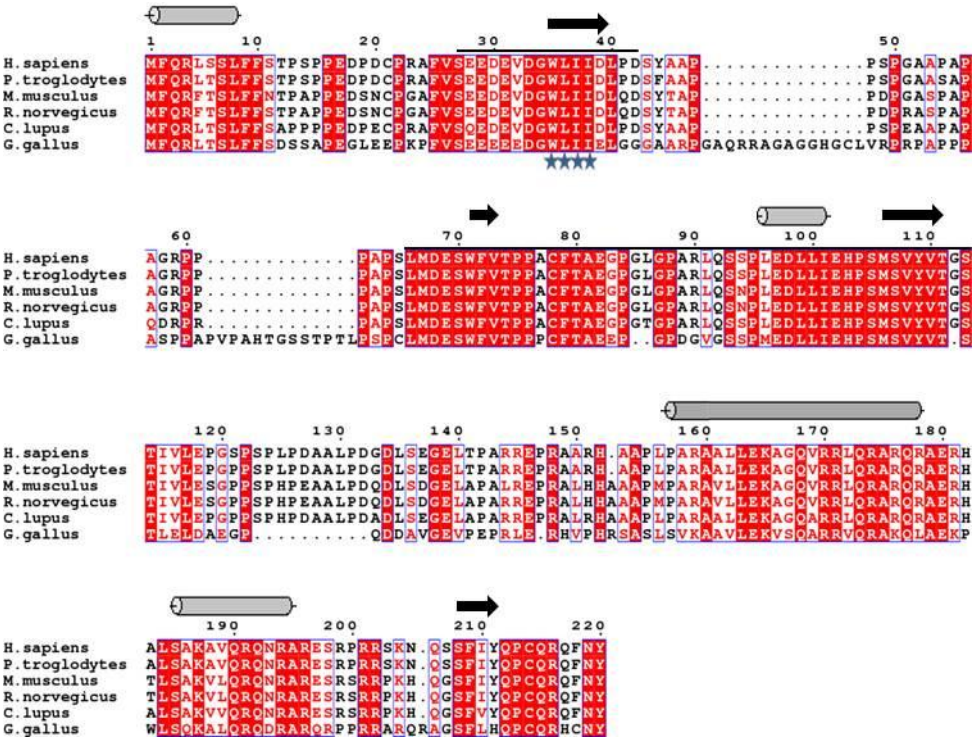


Figure 2 - Sequence alignment of DOR homologues. Sequence alignment of human (*H.sapiens*), chimpanzee (*P. troglodytes*), mouse (*M.musculus*), rat (*R. norvegicus*), dog (*C.lupus*) and chicken (*G.gallus*) DOR protein (accession numbers: NP_067025.1, XP_001160044.1, NP_835212.1, XP_001074013.1, XP_852549.1, XP_001232258.1, respectively) . Conserved residues are show in white against a red background, identical residues are show in red. The conserved LIR domain is indicated by blue stars. Conserved regions, 1(residues 28-42) and 2 (residues 66-112) are indicated by a black line above the alignment. The secondary structural elements of DOR are show above the alignment (PsiPred server). The figure was prepared with ESPript (http://espript.ibcp.fr).

1.2. DOR shows transcriptional activity

The high expression of DOR in tissues characterized by high metabolic requirements led to speculate about the regulatory role of this protein in thyroid hormone action. To test this hypothesis, Baumgartner and coworkers, transfected HeLa cells with DNA encoding thyroid hormone nuclear receptor, $TR_{\alpha 1}$, and luciferase reporters gene fused to thyroid hormone receptor (TR) elements, in the presence or absence of DOR. Co-transfection of DOR and $TR_{\alpha 1}$ enhanced the transcriptional activity of luciferase reporter gene, whereas a small effect was showed with DOR alone. These observations indicate that DOR acted as an activator, binding to TRs and enhances the transcriptional activity of thyroid

receptors (TRα1) in a ligand dependent manner [1]. The transcriptional activity of DOR must be from the N-terminal, the C-terminal has no transcriptional activation capacity and the activity of the N-terminal decrease when the C-terminal half is present [1], [5]. Mutant forms of DOR in region 1 and 2, L36A/L40A and E97K/D98K, show less transactivation activity than the wild-type DOR, confirming the relevance in transcriptional activity of these regions [5].

The *in vivo* function of DOR was analyzed by studying flies mutant for Drosophila DOR (dDOR), the Drosophila homologs of DOR, and the coactivator of ecdysone receptor (EcR) function was confirmed. The binding of dDOR with EcR, needed during metamorphosis, is required to transcriptional activity of EcR. A number of ecdysone loss-of-function phenotypes were observed in the absence of dDOR, such as spiracle eversion, salivary gland degradation and pupal lethality [6].

Recent reports, shown that DOR is a bifunctional protein that operates both in the nucleus and in the cytosol where it plays an essential role in autophagy.

1.3. The role of DOR in Autophagy

Autophagy is a major cellular pathway for the degradation of long-lived proteins and organelles, in mammalian cells [7]. The autophagosome, a double membrane vesicle, collects intact organelles and portions of the cytosol. After that, form an autolysosome, by a fusion with an lysosome [8]. The cargo are exposes to lysosomal hydrolases, resulting macromolecules that are transported back into the cytosol through membrane permeases for reuse [9] . Autophagy plays an important role in cellular homeostasis, eliminating damage/old organelles. Defective autophagy has been implicated in the pathogenesis of certain types of neuronal degeneration and cancer [9]. Questions related to the identification of the components of autophagy, signals that regulate and the mechanism of degradation were opened due medicine interest [5]. Nuclear regulators of gene transcription and autophagy have a bidirectional functional link, as example, p53 and E2F1, nuclear regulators, stimulate autophagy after activation by DNA

damage. The autophagic proteins, LC3 and p62, have other important characteristic to the regulation of autophagy, they are nucleocytoplasmic shuttling in cells [5].

Autophagy must be induced in the presence of stress by the nutrient sensor mammalian target of rapamycin (mTOR) inhibition, whereas in favorable conditions, the autophagy is negatively regulated (Figure 3). The nucleation of the autophagosome membrane is initiated by a complex formation between beclin 1 with vps34, whereas phagophore formation is promoting by the autophagic-related (Atg) proteins. In autophagy condition, DOR exits the nucleus, localizes to early autophagosomes and interacts directly with the autophagosome- membrane-associated proteins (LC3 and GATE16) [10]. LC3 and GATE16 remains bound to the autophagosomal membrane during autophagosome formation until autolysosome formation, where autophagic substrates are degraded [10].

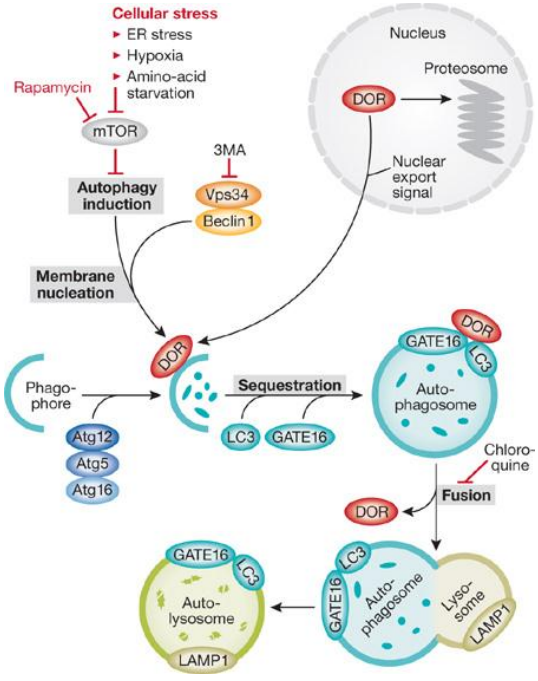


Figure 3- Incorporation of DOR into the autophagy. DOR translocates from the nucleus to autophagosomes upon induction of autophagy, where it forms a complex with LC3 and GATE16 (adapted from J. Spowart and J. J. Lum, 2010).

Because DOR doesn't interact with LAMP, a autophagosome- associated protein, could be a cofactor to target material to the autophagosome in a retrieval recycling manner [10].

In the absence of cellular stress, neither chloroquine nor 3-methyladenine (drugs are known to inhibit autophagy acting as inhibitors of and inhibitor of class III PI3K (Phosphatidylinositol 3-kinases) perturbed the nuclear localization of DOR, indicating that a specific activation of autophagy by cellular stress is crucial to release DOR from the nucleus [10].

The interaction of DOR with components of autophagy machinery such as LC3 or Golgi-associated ATPase enhancer of 16 kDa (GATE16) was supported for pull down, immunoprecipitation and two-hybrid assays [9]. By immunofluorescence microscopy, DOR with LC3, and DOR with GATE16 were detected in nucleus, on basal conditions, however in starvation they were detected in cytosolic punctate structures [9]. HeLa cells transfected with DOR show an increase in protein degradation, an increment in the number of GFP-LC3 contaning puncta per cell and lead to an increase in the number of autophagosomes under basal and starvation conditions. In parallel, DOR reduction caused a decrease in LC3 – containing puncta per cell, and decreased the number of autophagosomes, accompanied by a concomitant a inhibition of protein degradation [9].

The interaction of GATE16 and LC3 with DOR requires solely the N-terminal part of DOR [11]. Regions 1 and 2 in DOR sequence in mammalian cells and in fruit fly lines contribute to the autophagic activity. It was shown that under basal conditions, wild-type DOR was detected mainly in the nucleus, while GFP-LC3 and p62 were found mainly in cytosol. During stimulated autophagy, wild-type DOR and GFP-LC3 co-localized in cytosolic punctuate structures. In contrast, DOR mutants (L36A/L40A and E96K/D98K) caused only a minimal exit of the protein from the nucleus upon amino acid starvation or rapamycin treatment. Overexpression of these DOR mutants also prevented proper autophagosome formation in response to these autophagy-inducing stimuli [12]

These results indicate that mutations in regions 1 and 2 alter the normal function of DOR in autophagy, and retain DOR in the nucleus in response to rapamycin treatment [12].

1.4. The Autophagy-related proteins LC3 and GATE16

LC3 and GATE16 proteins were shown to interact with DOR and involved in several processes related to intracellular vesicle formation and transport, including autophagy [11].

Microtubule-associated protein light chain 3 (LC3) and Atg8p, yeast homolog of LC3, requires several post-translational modifications to associated with autophagosome membranes. The proprotein undergoes cleavage of its C-terminus of the conserved Gly120, to form a soluble cytosolic 18 kDa LC3-I, and is ultimately modified by the attachment of phosphatidylethanolamine to form LC3-II, a 16 kDa protein that localizes to autophagosomal membrane, a conversion that is commonly used to monitor autophagic activity [13]. Intra-autophagosomal LC3-II is subsequently degraded, whereas cytosolically-localized LC3-II can be released from the autophagosome membrane following delipidation [13], [14].

Golgi-associated ATPase enhancer of 16 kDa (GATE16) and $GABA_A$ receptor-associated protein (GABARAP) are two additional proteins identified as human homologues of Atg8. As in the case of LC3, GATE16 and GABARAP are also converted into form II and then located in autophagosomes [13].

Structural analysis of GABARAP, GATE16 and LC3 using X-ray crystallography and nuclear magnetic resonance (NMR) shows subtle structural differences. The structure of LC3 (PDB code 1UGM) contains a five-stranded central β-sheet as a core, which is flanked by two pairs of α-helices (α1 and α2, α3 and α4). The major portion of the structure comprising five β-stands and two α-helices (α3 and α4) closely resembles a ubiquitin fold, and the two additional α-helices (α1 and α2) associate with its N-terminus (Figure 4) [15].

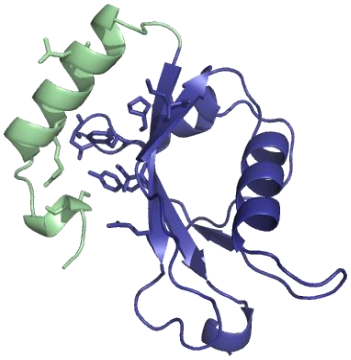


Figure 4 - Ribbon representation of LC3 three-dimensional structure (PDB ID 1UGM, [15]). The core ubiquitin-like fold is shown in blue and the additional N-terminal helices are shown in green. The side chains of residues involved in intra-domain interaction is represented as sticks.

Analysis of the LC3 structure in comparison with that of GATE16 (PDB code 1EO6), revealed strong structural similarity, with small in two regions: the loop between β1 and β2, and the other is the segment between β3 and β4. Both loops in LC3 have one amino acid residue insertion, Gln43 inserted in the β1-β2 loop, and Gly85 inserted in the β3-β4 loop. The latter insertion makes it possible to form a hydrogen bond with the main chain carbonyl group of Val83, resulting in the formation of an extra strand, β4. The presence of α1 and α2 is a structural feature of LC3 family proteins. The conservation of hydrophilic and hydrophobic interactions are strictly conserved in LC3 family proteins including Lys8, Arg16, Ile23, Pro32, Ile34, Glu36, Leu53, Asp106, Phe108 and Tyr110. The conservation of these residues implies that α1 and α2 are indispensable for biological function [15]. Comparing the electrostatic surface potential of LC3, GATE16 and GABARAP, the surface of α1 of LC3 is basic in contrast to the acidic nature of GATE16 and GABARAP (Figure 5). The surface of α2 moiety is acid, neutral and basic in LC3, GATE16 and GABARAP, respectively. The difference in electrostatic surface potential of α1 and α2 in LC3 family proteins may rather confer specificity toward their respective target proteins since LC3 family proteins are reported to interact with different target proteins [15].

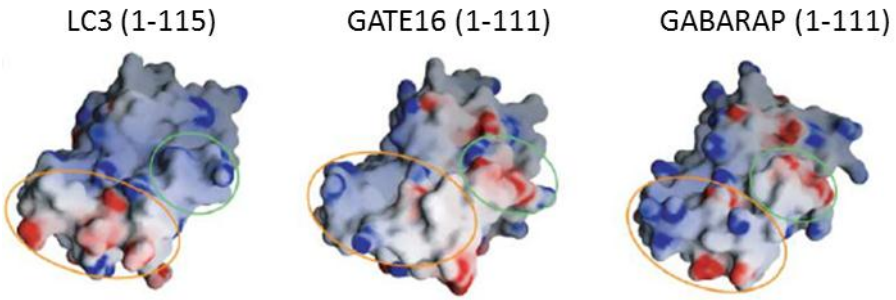


Figure 5- Electrostatic surface potential of LC3 family proteins. Surface representation of LC3, GATE16 and GABARAP colored according to electrostatic potential, red color (negative potential) and blue color (positive potential). α1 and α2 are indicated by circles in green and orange, respectively. In each protein, the C-terminal flexible tail is deleted to reveal the hydrophobic patch (adapted from Sugawara, K. et al., 2004).

1.5. The LC3 interacting region of DOR

Another factor that supports the idea of the interaction between these proteins is the existence of a typical LIR "LC3-interacting-region" motif found in DOR sequence with similarity to the LIR motif of p62 (Figure 6) and other LC3 interacting proteins as neighbor of BRCA1 gene 1 protein (NBR1) and Nip-like protein x (Nix). LIR motif is a LC3 interacting region that mediates the interaction between autophagic adaptors and the autophagosomal marker protein LC3 [16], [17], [18].

LIR sequences were defined as a WXXL motif, originally, but neither the tryptophan nor the leucine residues are strictly conserved, so a more general definition of the LIR sequence should be ΘXXΓ where Θ and Γ are respectively an aromatic and hydrophobic residues [17].

The tertiary structure of LC3-p62 revealed that both acidic cluster (DDD or DEE) and hydrophobic motif (WXXL or WXXV) in LIR motif of p62 are involved in the interaction with LC3 (Figure 6). In the structure of p62 in complex with LC3, the aromatic Trp340 residue interacts with the first hydrophobic pocket of LC3 and is situated between the core of the protein, which adopts an ubiquitin-like fold, and an N-terminal extension consisting of the two helices. The hydrophobic Leu343 residue binds to the second hydrophobic pocket located on the surface of the ubiquitin-like core domain [16], [17].

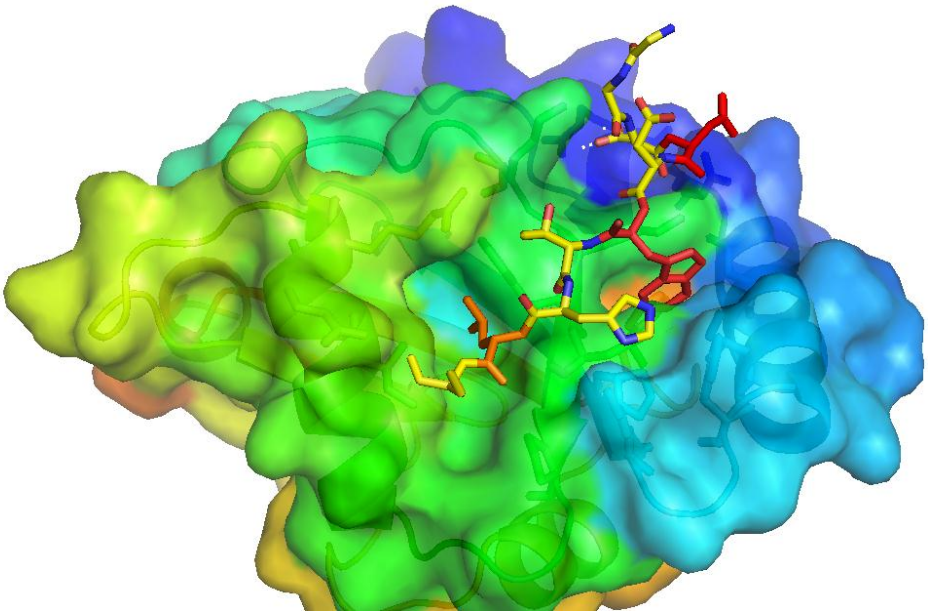


Figure 6- LIR motif of p62 in complex with LC3. Surface representing of LC3 in complex with a LIR peptide belonging to the LC3-interacting region of p62 (PDB code 2ZJD, [16]). Strictly conserved residues in DOR are shown in red and conserved residues in orange.

Sequence of Atg19, like p62, have tryptophan at position 1 and leucine at position 4 of LIR motif and bind to the conserved hydrophobic pockets on LC3 and Atg8 in a similar manner. Intriguingly, these binding pockets are located at the opposite surface of the hydrophobic patch that is conserved among ubiquitin-like modifiers and define the binding site for various target molecules. In both complexes, the WXXL motif adopts as extended β-conformation, which is considered to be essential for its interaction with Atg8/LC3-family proteins [19]. Contrary to the most LIR sequences described, LIR domain of NBR1 contains tyrosine at position 1 and isoleucine at position 2. The NMR structure of the GABARAPL-1 – NBR1-LIR complex is the first example of the structure of a complex involving a non-tryptophan autophagy receptor. The NBR1-LIR domain (YIIL) has a more hydrophobic nature that the p62-LIR domain (WTHL), which can be explained by the strong influence on the binding affinity mediated by the presence or the absence of a tryptophan residue in position 1 [17].

Region 1 of DOR sequence ($_{27}$SEEDEVDGWLIIDLPDSYAA$_{46}$) reveals the most common LIR sequence, the aromatic (W) and the hydrophobic (L) residues required for binding to the two hydrophobic pockets on the LC3 and GATE16 and additionally acidic residues on the N-terminal of this sequence, aspartic acids and glutamic acids [5]. As shown for p62, mutation of the core LIR residues W35 and Ile38 to alanine in DOR markedly reduced binding to and blocked the nuclear exit in response to autophagy activation [5].

Summarizing, region 1 of DOR protein contains a NES and an LIR motif, which participates in nucleocytoplasmic shuttling as well as in binding to LC3/GABARAP proteins and the induction of autophagosome formation. Another function of region 1 of DOR is transcriptional activity. The region 2 of DOR shows an α-helical structure and E97K/D98K DOR mutant showed the formation of aberrant autophagosomes under basal and autophagy activation conditions [5].

1.6. Objectives

Characterization of the interaction between DOR and its interacting partner, LC3 and GATE16, and the relevance of DOR in autophagy is very important. Since autophagy plays a protective role in aging, cell death, and defense against intracellular pathogens, neurodegenerative diseases and tumorogenesis, studying this protein and its interactions with their molecular partners might have biomedical implications.

To better understand and the role of this protein in autophagy, the knowledge of the three-dimensional structure of selected targets is required. X-ray crystallography is the most common method for determining the three-dimensional structure of a protein, which allows the visualization of the complex at atomic or near atomic level. Crystallization is the primary obstacle for solving a protein structure by X-ray crystallography, ant it is highly dependent upon the quality of the protein.

The crystallization workflow requires expression and purification of high-quality protein. Pure, homogeneous protein is the most critical prerequisite for successful protein crystallization, this can be assessed by many methods, like SDS-PAGE, size exclusion chromatography (nonaggregated protein), dynamic light scattering (polydispersity) and circular dichroism (protein secondary structure) [20]. Often, interaction with macromolecular partners improves the chances of obtaining stable and conformationally homogeneous protein for crystallization trials. The interaction between DOR and LC3 or GATE16 were studied and for that purpose recombinant proteins were expressed in *E.coli* and purified. To determine and characterize the optimal interaction partners for structural studies, a set of biophysical assays were performed, such as differential scanning fluorimetry (DSF), circular dichroism (CD) and surface plasmon resonance (SPR). The main research effort was invested in the characterization of the interaction between DOR and its interacting partners, crucial for the growth of crystals for three-dimensional structure determination.

2. Materials and Methods

2.1. Molecular Biology

2.1.1. Transformation

E.coli DH5α (Invitrogen) competent cells were thawed on ice and 1 µl of DNA was added to 50 µl of cells. The transformation reaction was incubated on ice for 30 minutes (min) and heat-shocked at 42 ºC for 45 seconds (sec). After 2 min of incubation on ice 900 µl of LB medium (ForMedium) were added, cells were recovered at 37 ºC for 1 hour (h) with shaking. The mixture was centrifuged for 1 min at 17 000 g, the supernatant discarded and the pellet re-suspended in the remaining medium (~150 µl). The culture was spread on LB agar plates (Appendix I) containing 50 µg/ml of kanamycin (kan). The plates were incubated at 37 ºC overnight.

2.1.2. Isolation of Nucleic Acids

A single colony from the relevant culture of *E. coli* DH5α transformed was grown overnight at 37 ºC with shaking in 10 ml of LB medium containing 50 µg/ml of kan. The culture was harvested by centrifugation for 10 min at 12 000 g at room temperature and the supernatant was discarded. To isolate and prepare plasmid DNA, the High Pure Plasmid Isolation Kit, from Roche Applied Science was followed. The plasmid DNA was quantified by measuring the absorbance at 260 nm on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific). The concentration value for each plasmid was determined based in Beer-Lambert equation. This equation allowed the use of an extinction coefficient with units expressed in ng/cm/ml. The equation used was $C= (A \times \varepsilon) / b$, where C is the nucleic acid concentration in ng/microliter, A is the absorbance in Absorbance unit (AU), ε is the wavelength dependent extinction coefficient in ng-cm/µl and b is the path length in cm.

2.1.3. Cloning of DOR

The synthetic cDNA encoding for DOR (in pUC57, GenScript, USA), with codons optimized for *E.coli* expression, were digested with NcoI and Acc65I (New England Biolabs) restriction enzymes and inserted on expression vectors (EMBL vectors).

To improve the chances of success, a set of expression vectors based on pET28 (containing a hexahistidine tag) that have been modified to express fusion proteins with additional N-terminal tags were used. The different fusion proteins were Maltose Binding Protein (MBP), Nus A, GB1, Thioredoxin (Trx) and Z2 domain. All the vectors code for a Tobaco Etch Virus (TEV) Protease recognition sequence before the cloning site of the protein of interest to allow for fusion tag/protein removal after protein purification. The vectors were digested with Acc65I and NcoI and treated with alkaline phosphatase (New England Biolabs) before ligation with DOR cDNA.

The fragments obtained after the restriction digestion of DOR gene and EMBL vectors were separated by agarose gel electrophoresis. The band of interest was excised from the gel with a blade under UV light and the cDNA purified using the High Pure PCR Product Purification Kit, for purification of PCR reaction products from Roche Applied Science.

Ligation was done using a molar ratio 3:1 vector to insert, 1 µl of T4 DNA ligase (New England Biolabs) in a total reaction volume of 20 µl, overnight at 16 ºC in 1x T4 DNA ligase buffer (50 mM Tris-HCl, pH 7.5; 10 mM $MgCl_2$, 1 mM ATP, 10 mM Dithiothreitol).

After ligation, 10 µl were transformed as described in 2.1.1. and positive clones selected by restriction analysis, using Acc65I and NcoI, as described in section 2.1.2.. The selected clone was sent to sequencing

2.1.4. Agarose gel electrophoresis of DNA

DNA was analyzed on a 1% agarose gel containing 10 µg of ethidium bromide. Samples were prepared by adding 1/10 sample volume of DNA Loading Dye (0.5 M Ficoll 400; 100 mM EDTA; 1% SDS; 3.6 mM Bromophenol blue; 4.6 mM de

xylene cyanol).The samples were loaded on the gel and ran at 100 V in TAE buffer (40 mM Tris; 1 mM EDTA; 0.11% glacial acetic acid).

2.1.5. LC3 and GATE 16 Cloning

Gate 16 and LC3 cDNAs (kindly provided by Dr. Zorzano, IRB, Barcelona, Spain) were amplified by Polymerase chain reaction (PCR) to insert NcoI and Acc65I restriction sites, before the ligation with EMBL vectors.

The sequences coding for the GATE16 and LC3 were amplified by PCR, to obtain the mature form, GATE16 1-116 and LC3 1-120. The reaction mixture contained 50 ng of DNA; 0.2 mM of each dNTP; 0.5 µM of forward primer (for Gate 16: 5'-CCATGGGCAAGTGGATGTTCAAGGAGGACC- 3' and for LC3: 5'-CCATGGGCCCGTCCGAGAAGACCTTC- 3'); 0.5 µM of reverse primer (for Gate 16: 5'-GGTACCTCAGCCGAAAGTGTTCTCGCCAC-3' and for LC3: 5'-GGTACCTCACC CGAACGTCTCCTGGGAG - 3'); *Taq* polymerase buffer (500 mM KCl, 100 mM Tris-HCl pH 8.3 (at 25°C), 15 mM Mg^{2+})(5Prime) and 1.25 U *Taq* polymerase (5Prime) to a final volume of 50 µl. The cDNA was amplified by an initial denaturation step at 94 ºC for 5 min, followed by 35 cycles of denaturation at 94 ºC for 45 sec, for 35 cycles, anneling at 58 ºC for 45 sec and extension at 65 ºC for 1 min and 30 sec, with a final extension step. The PCR product was purified using the High Pure PCR Product Purification Kit, for purification of PCR reaction products from Roche Applied Science. Thereafter, ligation to EMBL vectors and selection of positive clones were made as described for DOR.

2.1.6. DOR Mutagenesis

The DOR W35AI38A mutant was obtained by site-directed mutagenesis of the wild- type DOR expression construct.

For this mutagenesis 50 ng of DNA was used; with 0.2 mM of each dNTP; *Pfu* buffer (20 mM Tris-HCl pH8.8; 10 mM $(NH_4)_2SO_4$; 10 mM KCl; 0.1 mg/ml BSA; 0.1% (v/v) Triton X-100; 2 mM $MgSO_4$) (Fermentas); 0.5 µM of forward primer (GAAGATGAAGTTGACGGCGCGCTGATTGCAGATCTGCCGGACTCGTATG),

0.5 µM of reverse primer (CATACGAGTCCGGCAGATCTGCAATCAGCGCGCC GTCAACTTCATCTTC); 2.5% DMSO and 1 µl of *Pfu* Turbo DNA polimerase. The cDNA was amplified by 18 cycles of denaturation steps at 95 ºC for 30 sec, annealing at 55 ºC for 1min and extension at 68 ºC for 16 min preceded by a initial denaturation step at 95 ºC for 5 min. After PCR reaction, digestion with 1 µl of DpnI was done at 37 ºC, for 1 h. The mutagenic product was transformed in *E. coli* DH5α competent cells and plated on LB/kan plates.

2.1.7. DNA sequence analysis

The cDNA for DOR, LC3 and GATE16 on cloning vectors were analyzed by DNA sequencing using T7 reverse primer (5'-GCTAGTTATTGCTCAGCGG-3') to StabVida (Portugal). The results were analyzed with VectorNTI sequence alignment software (Invitrogen).

2.2. Protein Expression

2.2.1. Screening for protein expression

Small-scale expression tests were performed in order to find the best set of expression conditions yielding soluble protein. The variables tested were: Fusion partner (MBP, NusA, GB1, Trx and Z2); *E.coli* strain (BL21 Star (DE3) (Invitrogen), JM109 (DE3) (Promega)), expression medium (LB, 2YT, ZMY5052 (Appendix I)); IPTG concentration (0.1 mM; 0.5 mM) and temperature at 25 ºC. Since the number of experimental points of the full combination of variables increases exponentially, an incomplete factorial approach was used to decrease the number of combinations to test while retaining the statistical significance of the full factorial (Table 1) [21]. To analyze the results of protein expression, 1 ml of each culture at T0 (before induction with IPTG) and T3 (after 3 h of expression) were collected. Samples were harvested at 17 000 g, 3 min, at 4 ºC, the supernatant was discarded and the pellets were resuspended in BugBuster Protein Extraction Reagent (Novagen) (100 µl for T0 and 200 µl for T3 collected sample), next the

sample was incubated for 20 min at room temperature on a shaking platform. After that the samples were centrifuged at 17 000 g for 5 min at 4ºC to remove cell debris. The soluble fractions were removed to a fresh tube and the pellet resuspended in 200 µl of buffer, then SDS-PAGE Loading Buffer Dye was added to both fractions. Expression levels were analyzed by SDS-PAGE and western-blot using an anti-histidine tag antibody (Mouse Anti-His mAb, GenScript, USA). The condition yielding the higher levels of soluble recombinant protein was selected for experimental scale-up.

Table 1 – Incomplete factorial of small-scale expression tests of DOR

Fusion partner	*E.coli* Strain	Medium	[IPTG] mM
GB1	JM109 (DE3)	LB	0.5
GB1	JM109 (D3)	ZMY5052	0.5
Trx	BL21 Star (DE3)	LB	0.5
NusA	JM109 (DE3)	ZMY5052	
GB1	BL21 Star (DE3)	ZMY5052	
Trx	JM109 (DE3)	ZMY5052	
NusA	BL21 Star (DE3)	LB	0.1
Trx	JM109 (DE3)	2YT	0.1
MBP	JM109 (DE3)	LB	0.5
Z2	JM109 (DE3)	ZMY5052	
Z2	BL21 Star (DE3)	2YT	0.1
NusA	BL21 Star (DE3)	2YT	0.5
NusA	BL21 Star (DE3)	ZMY5052	
MBP	BL21 Star (DE3)	ZMY5052	
Trx	BL21 Star (DE3)	2YT	0.5
Z2	BL21 Star (DE3)	LB	0.1
Z2	JM109 (DE3)	2YT	0.5
GB1	BL21 Star (DE3)	2YT	0.1
MBP	JM109 (DE3)	2YT	0.1
MBP	BL21 Star (DE3)	LB	0.1

2.2.2. Overexpression of proteins

Large scale production of all proteins, NusA-DOR; NusA-DOR W35AI38A; Trx-LC3 and Trx-GATE16, were done in BL21 Star (DE3) *E.coli*. Expression cultures (LB medium with 50 mg ml^{-1} kan) were incubated at 37 ºC until the OD_{600} reached ~0.5 AU. At this point, the incubation temperature was lowered to 25 ºC and the bacterial cells were grown to 0.7 AU<OD_{600}<0.9 AU. The expression was induced by addition of 0.1 mM IPTG and continued for 3 h. Cells were harvested by centrifugation (4000 g, 20 min, 4 ºC), resuspended in lysis buffer (25 ml per liter of culture) supplemented with 50 ug ml^{-1} lysozyme and stored at -20 ºC. The lysis buffer for NusA-DOR and NusA-DOR W35AI38A was 20 mM sodium phosphate pH 7.5; 500mM NaCl; 10 mM imidazole; 1% (v/v) glycerol; 5 mM 2-mercaptoethanol and Complete EDTA-free protease inhibitor cocktail (Roche), for Trx-LC3 and Trx-GATE16 was 20 mM sodium phosphate pH 7.5; 500 mM NaCl; 10 mM imidazole and Complete EDTA-free protease inhibitor cocktail (Roche).

2.3. Protein Purification

Column chromatography is a method that can be applied to obtain a pure protein preparation from microbial cell, this method separates a mixture of proteins that is passed through a column with a matrix of beads. The nature of these beads determines whether the separation of proteins is based on differences in shape (gel filtration chromatography), net-charge (ion-exchange chromatography), hydrophobicity (hydrophobic chromatography) or binding affinity (affinity chromatography).

The ability of proteins to bind specifically to other molecules is the basis of affinity chromatography. Immobilized Metal Affinity Chromatography (IMAC) is a special form of affinity chromatography, based on the specific interaction between immobilized metal ions (e.g. Ni2+) and recombinant proteins bearing a 6xHis affinity tag on its C- or N-terminal. The most frequently used elution procedure for histidine-tagged proteins is based on a competitive displacement by imidazole [22].

Size exclusion chromatography separates a mixture of proteins based on differences in their molecular size and shape. Although proteins flow around the spherical beads in gel filtration chromatography, they spend some time within the large depressions that cover a beads surface. Because smaller proteins can penetrate into these depressions more easily than can larger proteins, they travel through a gel filtration column more slowly than the larger proteins [23].

Ion exchange chromatography separates molecules on the basis of differences in their net surface charge. Elution is usually performed by increasing salt concentration (usually NaCl) or by changing pH. The separation is based on the reversible interaction between a charged protein and an oppositely charged chromatography medium. The net surface charge of proteins varies according to the surrounding pH. Typically, when above its isoeletric point (pI) a protein will bind to an anion exchanger, when below its pI a protein will bind to a cation exchanger [24].

2.3.1. Purification of DOR

The cells were lysed by cycles of freeze-thawing and the suspension was then supplemented with DNase I (12 µg ml^{-1} final concentration) and $MgCl_2$ (20 mM final concentration). After a centrifugation step for clarification (20 000 g, 20 min, 4 ºC), the NusA-DOR-containing extract was loaded onto a 5 ml HisTrap HP column (GE-Healthcare) pre-equilibrated with lysis buffer. The protein was eluted stepwise with increasing concentrations of imidazole in lysis buffer (10-500 mM). Fractions were analyzed by SDS-PAGE and the ones containing pure protein were pooled. The final purification step was as size exclusion chromatography Sephacryl-200 (26/60, Amerscham Biosciences) using 20 mM HEPES pH7,5; 100 mM NaCl; 1 mM EDTA; 10% (v/v) glycerol and 5 mM 2-mercaptoethanol as elution buffer. NusA-DOR W35AI38A purification was performed similarly.

2.3.2. Purification of LC3 and GATE16

Each cell lysate was thawed in water and treated with DNase I (12 µg/ml) and $MgCl_2$ (20 mM). After centrifugation (as described in 2.4.1), the protein-containing extract was loaded onto a 5 ml HisTrap HP column (GE-Healthcare) pre-equilibrated with lysis buffer and the protein was eluted stepwise with increasing concentrations of imidazole in lysis buffer (10-500 mM). The protein was dialyzed against 20 mM HEPES pH7.5; 100 mM NaCl; 1 mM EDTA. Subsequently, the Trx and His tag were digested in a reaction mixture containing TEV protease. Purified Trx-LC3 and Trx-GATE16 were cleaved by recombinant TEV protease in order to remove the Trx and the histidine tag. The cleavage was performed at 4 ºC; for 3 h; in 20 mM HEPES pH7.5; 100 mM NaCl; 1 mM EDTA reaction buffer, using a 1:25 enzyme: substrate molar ratio. The reaction was applied again onto a His Trap HP column (GE-Healthcare). The flow-through of this column containing LC3 or GATE16 without the Trx and His tag was dialyzed against buffer A (20 mM HEPES pH7.5; 1 mM EDTA) before loading into a cation-exchange column. LC3 or GATE16 containing extracts were injected into a Mono-S column (5/50 GL, 5 cm x 2 cm, GE-Healthcare) equilibrated with buffer A and eluted with 20 ml in a linear NaCl gradient (0-500 mM for LC3 and 0-1000 mM for GATE16).

2.3.3. TEV protease expression and purification

BL21 CodonPlus cells transformed with pRK793 were grown at 37 ºC in LB medium containing 100 µg/ml ampicillin and 30 µg/ml chloramphenicol. Expression culture was incubated at 37 ºC until the OD_{600} reached ~0.5 AU, and induction was performed by addition of 1mM IPTG at 30 ºC. After 4h of induction, the cells were collected by centrifugation (4000 g, 20 min, 4 ºC) and resuspended in 10 ml of lysis buffer (50 mM sodium phosphate pH7.5; 100 mM NaCl; 10% glycerol; 25 mM imidazole) per liter of expression. The lysis buffer was supplemented with 0.2 mg/ml lysozyme; 10 µg/ml DNase; 1 mM $MgCl_2$; 1 mM PMSF and stored at -20 ºC.

The cell suspension was disrupted by freeze-thawing and supplemented and centrifuged at 20 000 g, 20 min, 4 ºC. The extract was loaded onto a 5 ml HisTrap HP column (GE-Healthcare) pre-equilibrated with lysis buffer. TEV protease was

elute with a 10 column volumes (CVs) linear gradient of lysis buffer to 50 mM sodium phosphate pH7.5; 100 mM NaCl; 10% glycerol; 300 mM imidazole.
Fractions containing protein were pooled and loaded on a HiTrap Desalting column (GE-Healthcare) pre-equilibrated with 25 mM sodium phosphate pH7.5; 200 mM NaCl; 10% glycerol; 2 mM EDTA; 10 mM DTT. After elution protein was stored at -80 ºC [25].

2.3.4. DOR peptide

A DOR peptide, representing LIR domain, $_{27}$SEEDEVDGWLIIDLPDSYAA$_{46}$ was commercially synthetized (GenScript, USA) and was diluted in water to a final concentration of 14 mg ml^{-1}. Complexes of DOR peptide with LC3 or GATE16, with different peptide to protein ratios, were prepared at 4 ºC and incubated for 3h before use.

2.3.5. Protein Quantification

Purified proteins were quantified by measuring the absorbance at 280 nm on a Nanodrop ND-100 Spectrophotometer (Thermo Scientific). This quantification is based on the Lambert-Beer law, which considers a linear relationship between absorbance and protein concentration, according to the equation $A=c \varepsilon l$, in which A corresponds to the absorbance represented in AU, c to the protein concentration in molarity (M), ε to the extinction coefficient of the protein in units of $M^{-1}cm^{-1}$ and l to the path length of the sample in centimeter (cm). Before quantification, the molar extinction coefficient for each protein was estimated from the respective amino acid sequence using the ProtParam Tools of the Expasy Proteomics Server (http://expasy.org/).

2.3.6. SDS-PAGE

SDS-PAGE was made using 12.5% polyacrylamide gels. The samples were loaded with 5x loading buffer (10% w/v SDS; 20% v/v glycerol; 0.2 M Tris-HCl pH 6.8; 0.05% Bromophenol Blue; 1.8% 2-mercaptoethanol) followed by incubation at 95 ºC for 8 min. The running buffer used was 24.8 mM Tris; 191.8 mM glycine and 0.1% SDS. Gels were run at 150 V, at room temperature, for approximately 45 min, using a Mini Protean 3 system (Bio-Rad). For staining the gels were heated for 1 min in deionized water and this step was repeated three times with fresh water. Washed gels were boiled with Page Blue (Fermentas) solution for 15 sec, and transferred to a shaking platform for 20 min, after which, the staining solution was removed, and destaining was achieved by washing with distilled water. The stained gels were analyzed with a scanning densitometer (Molecular Image GS800 Calibrated Densitometer, BioRad).

2.3.7. Western blot

For western blot analysis, SDS-PAGE gels were electrotransferred onto Polyvinylidene fluoride (PVDF) membranes (Roche Applied Science), previously activated with methanol. The electroctransfer was made using a Trans-blot Electrophoretic Transfer Cell (Bio-Rad), overnight, at 40 V, 11 ºC, in 25 mM Tris, 192 mM glycine, 20% methanol buffer. Membranes were blocked for 1 h with TBS-T buffer (20 mM Tris, pH7.6; 137 mM NaCl and 1% Tween 20) containing 5% milk (Molico, Nestlé). After that membranes were incubated with primary antibody solution (monoclonal mouse anti-His, GenScript, USA) 1:10 000 in TBS-T buffer with 0.5% milk, for 1 h. After incubation with primary antibody, membranes were washed with 0.5% milk solution, 5 times 5 min, and then incubated with secondary antibody. The secondary antibody used was Anti-Mouse IgG + IgM alkaline phosphatase linked whole antibody (from goat, Amersham Biosciences). Membranes were developed with ECF substrate (GE Healthcare) for 5 min and then visualized in a Molecular Imager FX (Bio-Rad).

2.4. Protein structural studies

2.4.1. Dynamic Light Scattering (DLS) Analysis

Dynamic Light Scattering (DLS) is a powerful technique to characterize dilute and transparent dispersions of particles. DLS measures scattered light fluctuations caused by the Brownian motion of particles, across very short time intervals to produce a correlation curve. The correlation calculation derived from the scattered light intensity fluctuations allows, then, to determine the diffusion coefficient and subsequently the particle size. Knowing the medium viscosity, the hydrodynamic radius of objects can be calculated using the Stokes-Einstein law [26].

$$R_H = kT/6\pi\eta D$$

Where *D* is the diffusion coefficient of particles, *K* the Boltzmann constant, *T* the temperature, η the dynamic viscosity and R_H the hydrodynamic radius. This law only applies to homogeneous and dense spheres without interactions. The hydrodynamic radius, that is being reported, is the radius of the hard sphere that diffuses at the same speed as the particle or molecule being measured. The size distribution obtained by DLS is a plot of the relative intensity of light scattered by particles in various size classes and is therefore known as an intensity size distribution. If the plot shows one peak with a substantial tail or more than one peak, the intensity size distribution must be converted to a volume size distribution for a more realistic view of the data, considering the importance of the tail or a second peak. Polydispersity (PdI) in the area of light scattering is used to describe the width of the particle size distribution [27], [26].

NusA-DOR (1 mg/ml in 20 mM HEPES pH7.5, 100 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol and 5 mM 2-mercaptoethanol) was centrifuged at 100 000 rpm for 45 min at 4ºC in an Airfuge™ air driven ultracentrifuge (Beckman Coulter). Three independent measurements were obtained at 20 ºC, in a 45 µl DTS 2112 cuvette, in a Zeta size Nano Zs DLS system (Malvern Instruments).The result was analyzed using the software for the instrument DTS (nano) 6.30 [26].

2.4.2. Circular Dichroism

The secondary structure content of all proteins were determined its far UV circular dichroism (CD) spectrum. CD spectroscopy is a powerful method in structural biology that has been used to examine proteins, polypeptides and peptide structures. CD is defined as the difference in absorption of left hand and right hand circular polarized light with optically active compounds [28]. Protein secondary structure can be determined by CD spectroscopy in the “far-UV” spectral region (190-250 nm). At these wavelengths the chromophore is the peptide bond, and the signal arises when it is located in a regular, folded environment. The α-helix, β-sheet and random coil structures each give rise to a characteristic shape and magnitude of CD spectrum [29].

Measurements were performed at 20 ºC on a Jasco J-815 spectrometer fitted with a Peltier temperature controller. Spectra were acquired between 190 and 260 nm, set up to 1 nm band width, 1s response, 500 nm/min scanning speed and 3 accumulations. NusA-DOR (2 µM) and NusA (1 µM) samples were in 20 mM HEPES pH7.5, 100 mM NaCl, 1 mM EDTA, 10% glycerol and 5 mM 2-mercaptoethanol, LC3 (7 µM) and GATE16 (6 µM) in 20 mM HEPES pH7.5, 100 mM NaCl, 1 mM EDTA, these samples were diluted in water to desired concentration, to a final volume of 250 µl. DOR peptide (6 µM and 12 µM) was prepared in water, to a final volume of 250 µl. In the case of experiments where trifluoroethanol (TFE) was included, NusA-DOR and NusA were diluted in 50% of TFE.

Each spectrum was the average of two scans with the average buffer control spectrum subtracted. All measurements were taken in ellipticities in mdeg and converted to molar ellipticities ([θ], deg cm2 dmol-1) by normalizing for the concentration of peptide bonds.

The spectra were analyzed with the CONTIN program of the software Dichroweb [30]. Dichroweb server includes a number of the most popular analysis programs. It provides access to all reference databases for the various algorithms enabling simple comparisons to be made between various combinations of algorithms and databases. The normalized root mean square (NRMSD), a standard goodness-of-fit parameter, provides an indication as to how closely the back-calculated spectra

produced from the predicted secondary structures reproduce the experimental spectrum [28]

2.4.3. Thermal Stability Assay

To assess protein thermal stability, a differential scanning fluorimetry (DSF) or thermal shift assay was performed. This technique allows the determination of the unfolding/melting temperature (T_m), the midpoint of the protein unfolding transition. An environmentally sensitive fluorescent dye, such as Sypro orange, that interacts specifically with nonnative protein is used. During DSF assay as the protein is subjected to increasing temperature it undergoes (thermal) unfolding exposing its hydrophobic regions. Sypro orange as the temperature rises, the protein undergoes thermal unfolding and exposes its hydrophobic core region, Sypro orange then binds to the hydrophobic regions and becomes unquenched, resulting in an increase on the intensity of the fluorescence. As a result, by plotting the fluorescence signal against temperature a thermogram is obtained, allowing the determination of the melting temperature as the maximum of the derivative of the plotted curve [31].

Protein melting temperature (T_m) determination was performed by monitoring protein unfolding using the fluoroprobe Sypro orange dye (Invitrogen). The assay was performed in a 96-well white plate (Thermo-Fast 96, Non-Skirted, Thermo Scientific) and in a Real Time – PCR machine ($IQ_{TM}5$ Multicolor Real-Time PCR Detection) with a Cy3 filter (~550 nm excitation and ~570 nm emission). To each well a final volume of 50 µl was pipetted and for each protein four replicates were prepared. Pre-dilutions from a stock solution of each protein to be studied (Table 2) and Sypro orange concentrated stock of 5000x to 10x, were prepared before the final master mix preparation. Solutions of 25 µl of protein in buffer condition and 25 µl of Sypro orange were added to each well. The plate was covered with sealing tape (Thermo Scientific); centrifuged at 4 ºC, for 1 min, at 1000 g; then the plate was placed in the RT-PCR machine, and the fluorescence signal was measured every 0.15 sec with a variation in temperature of 0.5 ºC, over the 35-85 ºC range. The melting curves were analyzed using the CFX Manager software

(Bio-Rad) which calculates the T_m from the maximum value of the first derivate curve of the melting curve.

Table 2- Concentrations of proteins used in Thermal shift assay

	pre-dilution solution (µM)	Concentration well (µM)
NusA-DOR	8	4
NusA	8	4
NusA-DOR W35AI38A	8	4
DOR peptide	48	24
LC3	24	12
GATE16	24	12
random peptide	48	24

2.4.4. Surface Plasmon Resonance

Biacore systems exploit the phenomenon of surface plasmon resonance (SPR) to monitor the interaction between molecules in real time. The approach involves attaching one interacting partners (ligand) to the surface of a sensor chip, and then passing sample containing other interaction partners over the surface (analyte). The sensor chip consists of a glass surface, coated with a thin layer of gold. In the most widely used sensor chip, the gold surface is modified with a carboxymethylated dextran layer, which forms a hydrophilic environment for the attached biomolecules, preserving them in a non-denatured stated [32]. The CM5 sensor chip is the most versatile and widely used for directed immobilization of a ligand. It is coated with a layer of carboxymethylated dextran and through covalent derivation, it is possible to immobilize a peptide/protein ligand via amine coupling ($-NH_2$), coupling by thyol-disulphide exchange (-SH) or aldehyde coupling (-CHO) [33].

Essentially, SPR detects changes in mass in the aqueous layer close to the sensor chip surface by measuring changes in refractive index. When the sample

containing the analyte is injected, the upward slope of the curve indicates the association of the analyte with the immobilized ligand (complex formation). After injection, the bound analyte dissociates while buffer is flowing (dissociation). Before performing a new injection, a regeneration solution is injected to dissociate the remaining analyte until the base line is reached (regeneration). This simple principle forms the basis of the sensorgram, a continuous, real time monitoring of the association and dissociation of the interacting molecules. The sensorgram provides quantitative information in real time on specificity of binding, active concentration of molecules in a sample, kinetics and affinity [34]·[32].

A Biacore T-100 Instrument (GE Healthcare Europe) was used for analyzing the interaction between NusA-DOR and LC3 or GATE16.The running buffer used was HBSEP (10 mM HEPES, pH 7.4; 150 mM NaCl; 3 mM EDTA; 0.05% Tween 20) supplemented with 0.84% of glycerol.

LC3 was initially diluted in NaAc pH 4.5 to a final concentration of 1 µg/ml and immobilized on a sensor chip CM5 (Biacore, GE Healthcare Europe) at flow rate of 5 µl/min, the chip was activated with NHS/EDC at 10 µl/min for 7 minutes followed by coupling of approximately 50 resonance units (RU) to 300 RUs of LC3 and deactivation with Ethanolamine at 10 µl/min for 7 minutes. One channel was left blank and was used as reference.

GATE16 was diluted to 1 µg/ml in NaAc pH 6.0 buffer and immobilized like LC3 in a range of 30-170 RUs.

Kinetic interaction experiments using Nus-DOR, NusA (negative control), NusA-DOR W35AI38A and DOR peptide as analytes were performed at 10 ºC, at flow rate of 30 µl/min during 30 seconds. NusA-DOR, NusA and NusA-DOR W35AI38A were diluted in HBSEP buffer and injected at concentrations between 31.25 nM and 5000 nM. The DOR peptide was diluted in HBSEP buffer supplemented with 0.84% of glycerol, because DOR peptide was prepared in water and the others samples have glycerol in the elution buffer, and injected at concentrations between 58.63 nM and 30 µM. Regeneration was obtained by injecting a solution of 25 mM NaOH at a flow rate of 60 µl/min for 60 sec at the end of each cycle of interaction. Data was evaluated with BIAevaluation software (Biacore, GE Healthcare Europe), applying a simple 1:1 binding mass transfer model.

2.4.5. Crystallization

The LC3: DOR peptide complexes were prepared by incubating LC3 (1600 µM) with DOR peptide (in 2 or 3- fold molar excess) on ice for 2-3 hours.

Initial crystallization screenings were performed using commercially available kits, at 20 ºC, using a Cartesian PixSys 4200 crystallization robot (Genomic solutions, U.K.) and Greiner Crystal Quick plates, at High Throughput Crystallization Laboratory at EMBL Grenoble.

Crystals of the 2 (peptide):1 (LC3) complex were obtained using 20% PEG 8000, 0.05 M Potassium phosphate as precipitant (The Classics, Qiagen/Nextal). Crystals of the 3 (peptide) : 1 (LC3) complex were obtained using 1 M lithium chloride, 0.1 M MES pH 6.0 and 30% PEG 6000 (PEG-LiCl Kit, Hampton Research).The first condition was further optimized to obtain better quality crystals by changing the precipitant concentration or using additives (Additive Screen, Hampton Research). The drops contained 1 µl protein solution mixed, 0.7 µl precipitant solution and 0.3 µl of additive solution and were equilibrated against 300 µl of mother solution in the reservoir. The best crystals (when the peptide was added in a 2-fold excess) were obtained using 0.1 M L-cysteine and 0.05 M Potassium phosphate with 20% PEG 8000 as precipitant. New optimization was performed using seed beads (Hampton Research) in an attempt to obtain larger and single crystals. These crystals were sequentially transferred to solutions of precipitant containing 5, 10 and 15% of glycerol as cryoprotectant and then cryo-cooled in liquid nitrogen.

For optimization of the crystallization condition identified when the peptide was added in a 3-fold molar excess, new solutions were prepared by varying the pH (5.5, 6.0 and 6.5) and the percentage of PEG 6000 (25%, 30% and 35%). The complex solution was incubated with these new solutions at 20 ºC in sitting-drop plates. The best crystals were obtained using 30% PEG 6000 and 0.1 M MES pH 6.0 as precipitant. A new optimization attempt was done using TMAO as an additive at 0.1 M, 0.2 M, 0.3 M and 0.4 M final concentration. The best crystals were directly plunged into liquid nitrogen for cryo-cooling before data collection.

2.4.6. Data collection and processing

Diffraction data was collected on a CCD Detector in an Oxford Diffraction Gemini PX Ultra diffractometer using CuKα radiation. A cryo-cooled crystal of DOR peptide with LC3 (2:1ratio) was used and was measured in 1º oscillation steps with an 70 mm sample to detector distance and 6 minutes exposure per frame (164 frames).

Diffraction images were integrated with CrysAlis ^{Pro} Software (Oxford Diffraction) and scaled with SCALA (Collaborative Computational Project No. 4, 1994).

2.4.7. Structure Solution

The three dimensional structure of DOR peptide with LC3 was solved by molecular replacement with PHASER [35] from the CCP4 suite, using the coordinates for LC3-p62 complex as the search model (PDB entry 2ZJD [16]). The model was built by alternated cycles of manual model using Coot [36] and refinement with PHENIX [37]. 3D molecular structure images were made with pymol software (http://www.pymol.org/).

3. Results and Discussion

3.1. Cloning of Human DOR into prokaryotic expression vectors

E.coli was selected as the host system to express the recombinant protein DOR. A synthetic cDNA coding for DOR with codon usage optimized for expression in *E.coli* was synthesized with the NcoI and Acc65I restriction sites included and inserted in pUC57 cloning plasmid by GenScript, USA (pUC57-DOR). The optimization takes into account codon adaptability, mRNA structure and various cis-elements in transcription and translation which leads to increase protein expression level [38].

For cloning into the *E. coli* expression vectors, pUC57-DOR was first digested using NcoI and Acc65I restriction enzymes), and the DOR cDNA fragment (670 bp) was purified after separation on agarose gel (Figure 7).

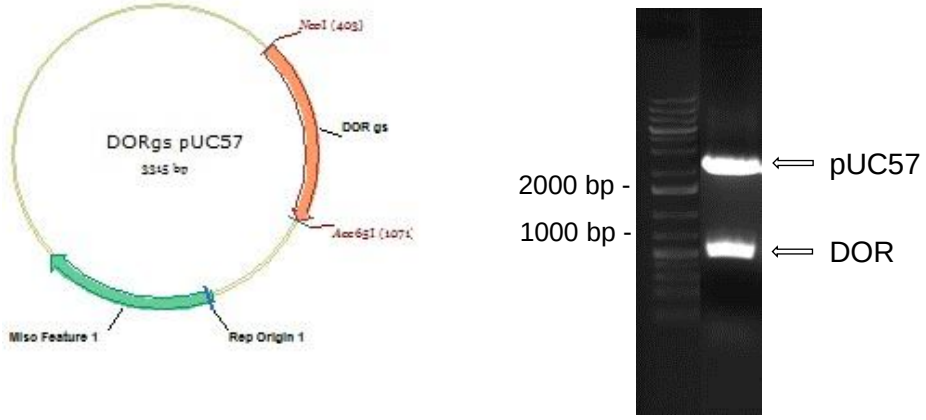


Figure 7 - pUC57-DOR digestion with NcoI and Acc65I restriction enzymes. Agarose gel electrophoresis separates two fragments corresponding to pUC57 with 2647 bp and DOR cDNA with 670 bp.

The DNA of the insert was cloned into EMBL vectors (Appendix III) for bacterial expression, of the following fusion proteins: Z2-DOR; NusA-DOR; MBP-DOR; GB1-DOR and Trx-DOR, all containing a histidine tag in N-terminal and a recognition site for TEV protease. After ligation the constructs were verified by restriction analysis with NcoI/Acc65I. The fragments thus obtained were separated by agarose gel electrophoresis. Two fragments were obtained, one with 670 bp corresponding to DOR insert and the other corresponding to the size of the vector used (e.g. Z2-containig plasmid with 5500 bp and NusA with 6800 bp (Figure 8)).

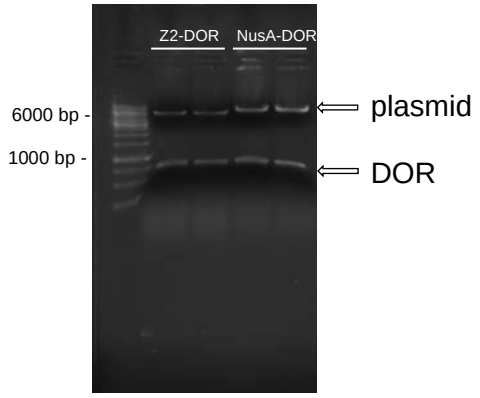


Figure 8- Restriction analysis of DOR expression plasmids. Example of restriction analysis of two clones of each construct of Z2-DOR and NusA-DOR fusion digested with NcoI and Acc65I.

The positive clones, containing an insert with the predicted size, were sent for sequencing and the results were analyzed with VectorNTI sequence alignment software (Invitrogen). The DOR cDNA was successfully inserted into EMBL vectors and was found to be in the correct orientation and in frame (Figure 9). Appendix IV shows the result of the NusA-DOR sequencing analysis.

```
atgggctttcagcgtctgagctctctgttctttagtaccccgtccccgccggaagatccg
 M  G  F  Q  R  L  S  S  L  F  F  S  T  P  S  P  P  E  D  P
gactgcccgcgcgcgttcgtgtcagaagaagatgaagttgacggctggctgattatcgat
 D  C  P  R  A  F  V  S  E  E  D  E  V  D  G  W  L  I  I  D
ctgccggactcgtatgccgcaccgccgagcccgggtgctgccccggcaccggcaggtcgt
 L  P  D  S  Y  A  A  P  P  S  P  G  A  A  P  A  P  A  G  R
cctccgccagctccgtctctgatggatgaaagttggttcgtcacgccgccggcctgtttc
 P  P  P  A  P  S  L  M  D  E  S  W  F  V  T  P  P  A  C  F
accgcggaaggtccgggtctgggtccggcacgtctgcaatcctcaccgctggaagacctg
 T  A  E  G  P  G  L  G  P  A  R  L  Q  S  S  P  L  E  D  L
ctgattgaacatccgtcgatgagcgtgtacgttacgggctctaccattgttctggaaccg
 L  I  E  H  P  S  M  S  V  Y  V  T  G  S  T  I  V  L  E  P
ggtagtccgtccccgctgccggatgctgccctgccggatggtgatctgtcagaaggtgaa
 G  S  P  S  P  L  P  D  A  A  L  P  D  G  D  L  S  E  G  E
ctgaccccggcccgtcgcgaaccgcgtgcagctcgtcatgccgccccgctgccggcacgt
 L  T  P  A  R  R  E  P  R  A  A  R  H  A  A  P  L  P  A  R
gctgccctgctggaaaaagcgggtcaggtgcgccgtctgcaacgtgcacgtcagcgtgct
 A  A  L  L  E  K  A  G  Q  V  R  R  L  Q  R  A  R  Q  R  A
gaacgtcatgccctgagcgccaaagcagttcaacgccagaaccgtgcccgcgaaagccgt
 E  R  H  A  L  S  A  K  A  V  Q  R  Q  N  R  A  R  E  S  R
ccgcgccgttctaaaaatcaaagcagctttatttatcagccgtgccagcgccagttcaac
 P  R  R  S  K  N  Q  S  S  F  I  Y  Q  P  C  Q  R  Q  F  N
tac
 Y
```

Figure 9- Human DOR cDNA translation

3.2. Expression trials

To find the best set of expression conditions yielding soluble protein DOR small-scale tests were performed. Different fusion partners, *E.coli* strain, expression media, IPTG concentration and temperature were tested, using the incomplete factorial approach, setup by the SAmBA software (http://igs-server.cnrs-mrs.fr/samba/).

Expression levels were analyzed by SDS-PAGE (12.5% resolving gel) and by western blot using an anti-hexahistidine antibody (Figure 10).

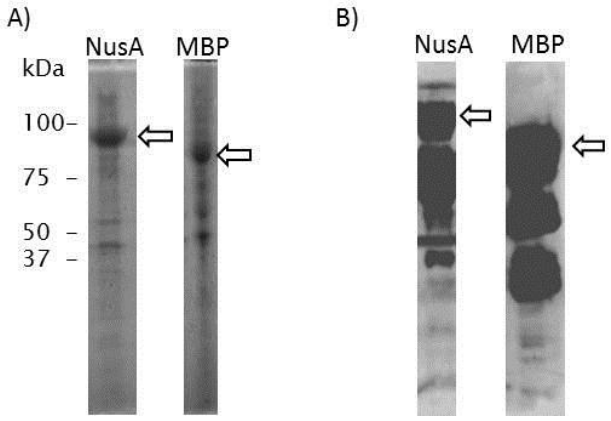


Figure 10- Analysis of expression conditions using two different constructs. A) Cell lysates were separated by SDS-PAGE and proteins were visualized by Coomassie Blue Staining B) by immunoblot analysis with anti-hexahistidine. The best conditions found are with NusA and MBP fusion partner, BL21Star (DE3) strain, 0.1 mM IPTG and 25ºC.

Higher expression levels were obtained for MBP-DOR (~67 kDa) and NusA-DOR (~82 kDa) recombinant proteins, in *E.coli* BL21 Star (DE3), grown in LB medium at 25 ºC, and induced with 0.1 mM IPTG. However significant levels of protein degradation were observed as indicated by the observed lower weight molecular bands (Figure 10, b).

Originally, the MBP-DOR fusion construct was selected for expression scale up and purification. However, size exclusion chromatography (SEC) analysis of the recombinant purified protein showed that the protein was predominantly oligomerized (Figure 11, limited by vertical lines). Therefore, the NusA-DOR fusion protein was selected for further purification and biochemical/structural analysis.

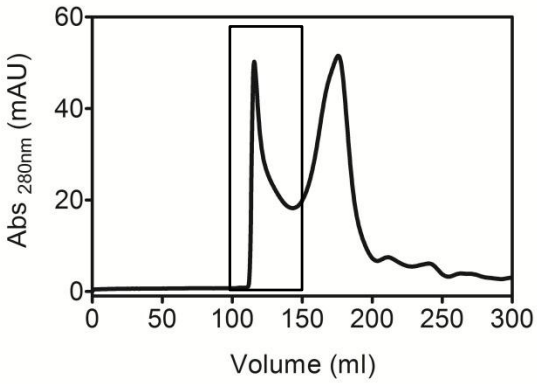


Figure 11- Size exclusion chromatography of MBP-DOR after IMAC purification. A 5 mL Fraction from HisTrap eluted with 300 mM imidazole was applied to a Superdex 200 (26/60) size exclusion column equilibrated in 20 mM HEPES pH7.5; 100 mM NaCl; 1 mM EDTA; 10% (v/v) glycerol and 5 mM 2-mercaptoethanol.

3.3. NusA-DOR purification

NusA-DOR fusion protein was expressed from six liters of cell culture. After cell lysis, the soluble fraction was purified in two steps, a first purification step using an ion immobilized affinity chromatography (IMAC), and a preparative SEC as a second step.

In the first purification step, the protein was eluted from the HisTrap HP column by a stepwise increase in the imidazole concentration. An example of the elution profile obtained is shown in Figure 12.

As it can be seen in the chromatogram, a peak containing DOR with a small amount of contaminants obtained with 300 mM imidazole. After SDS-PAGE separation and protein staining it is clear that the protein eluted corresponds to NusA-DOR with the predicted molecular weight of 82 kDa.

A preparative size exclusion chromatography was the second step of purification, where the most abundant specie was separated from the species with a different weight. The proteins eluted from the HisTrap were applied to a Sephacryl-200 (26/60) and a typical elution profile is shown in Figure 13.

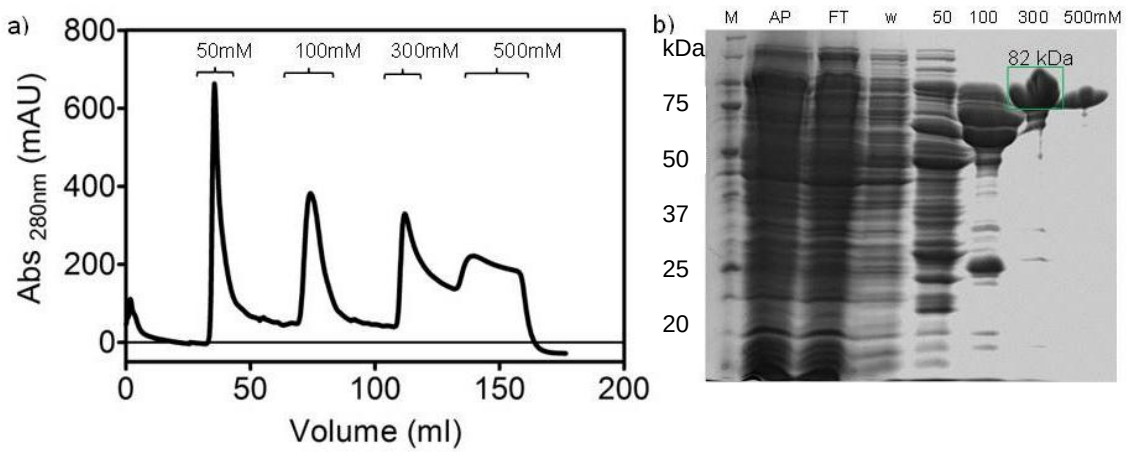


Figure 12 - Purification of NusA-DOR by affinity chromatography (IMAC). a) Chromatogram of NusA-DOR purified using a HisTrap HP 5 ml column equilibrated in 20 mM sodium phosphate, pH 7.5; 500 mM NaCl; 10 mM imidazole; 1% glycerol and 5 mM 2-mercaptoethanol. Protein was eluted with increasing concentrations of imidazole (50 mM, 100 mM, 300 mM, 500 mM). b) SDS-PAGE (12.5%) analysis of the IMAC fractions obtained after elution with each imidazole concentration, visualized by Coomassie Blue Staining. The highest yield of NusA-DOR was obtained after elution with 300 mM imidazole.

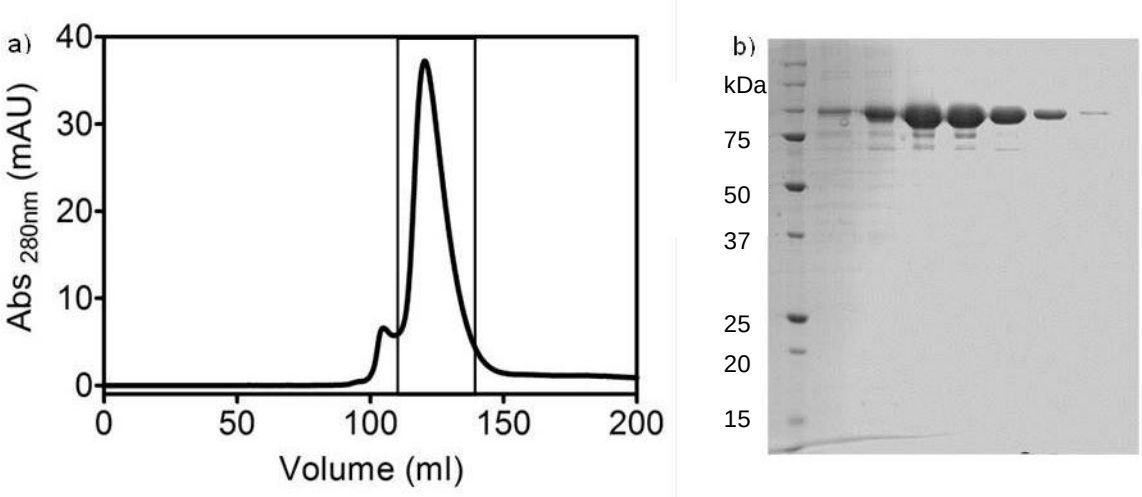
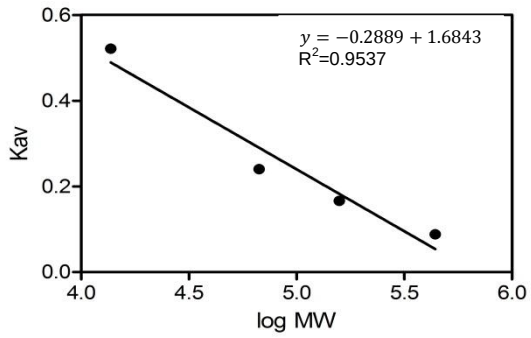


Figure 13 – Size exclusion chromatography of NusA-DOR after IMAC purification. a) A 5 mL Fraction from HisTrap eluted with 300 mM imidazole was applied to a Sephacryl-200 (26/60) size exclusion column equilibrated in 20 mM HEPES pH7.5; 100 mM NaCl; 1 mM EDTA; 10% (v/v) glycerol and 5 mM 2-mercaptoethanol. b) Coomassie blue stained SDS-PAGE (12.5%) analysis of a 5 ml fractions collected in peak delimited by vertical lines.

Although some of the protein was eluted in the void volume (first peak just before the main peak), significant amount eluted as a peak with elution of ~120 ml.

Table 3- Sephacryl-200 (26/60) size exclusion column calibration in 20 mM HEPES pH7.5; 100 mM NaCl; 1 mM EDTA; 10% (v/v) glycerol and 5 mM 2-mercaptoethanol



Protein	Volume elution (ml)	MW
Ribonuclease	207.5	13700
Albumina	141.4	67000
Aldolase	124.1	158000
Ferritin	105.7	440000
NusA-DOR	119.5	130000
NusA-DOR W35AI38A	121.1	124000

Comparison of the NusA-DOR chromatogram obtained with size exclusion standards chromatograms indicated that the purified NusA-DOR had a molecular weight higher than the expected 82 kDa compatible with a monomeric form (Table 3). This may be due to NusA-DOR having a more elongated form, non-globular.

The fractions corresponding to monomeric NusA-DOR (rectangle in Figure 13a) were pooled and concentrated by ultrafiltration on a centrifugal concentration device (30 kDa cutoff). Protein was quantified with by measuring the absorbance at 280 nm. The value obtained was used to calculate protein concentration using the theoretical extinction coefficient obtained with ProtParam tool (http://expasy.org/) (ε = 49390 M^{-1} cm^{-1}). The protein was frozen in liquid nitrogen and stored at -80 ºC.

Because NusA-DOR is not stable, with a lot of degradation and precipitation after TEV digestion, further biochemical and biophysical studies were pursued with the NusA-DOR fusion protein.

Previous analysis of the interaction between LC3/GATE16 and DOR showed that it was dependent on the integrity of LIR of DOR [12]. To test the relevance of these residues for the interaction, the conserved W35 and I38 of DOR were mutated to alanines and the recombinant protein was expressed and purified.

The mutant protein NusA-DOR W35AI38A was expressed and purified as described for WT NusA-DOR (Figure 14).

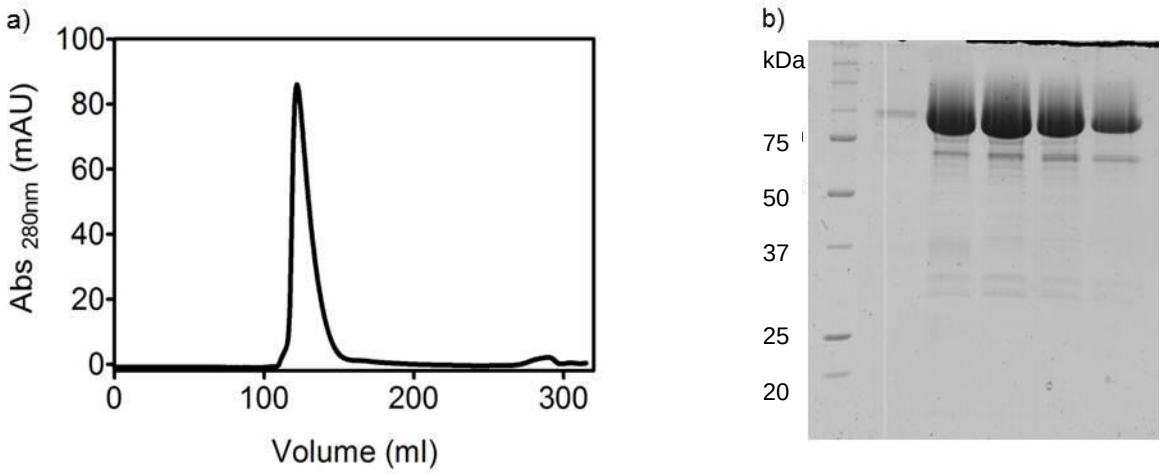


Figure 14 - Purification of NusA-DOR W35AI38A. a) NusA-DOR W35AI38A was purified using an immobilized metal-affinity chromatography, followed by a preparative SEC. a) The NusA-DOR W35AI38A behaves as monomer in solution (20 mM HEPES, pH7.5; 100 mM NaCl; 1mM EDTA; 10% glycerol and 5 mM 2-mercaptoethanol), as seen by gel filtration analysis, with an apparent molecular weight of 82 kDa. b) Coomassie blue stained SDS-PAGE (12.5%) of a 5 ml fractions collected in peak delimited by vertical lines of recombinant protein.

The mutated form eluted at 121.1 ml, indicated that the purified NusA-DOR mutant had a molecular weight higher than the expected 82 kDa compatible with a monomeric form (Table 3), as happened with wild type.

3.4. LC3 and GATE16 cloning

Recent reports show that DOR interacts with the autophagy related proteins LC3 and GATE16 [9], [11], [12] and therefore these interaction partners were expressed and purified for quantitative interaction analysis. The cDNA of $LC3_{120}$ and $GATE16_{116}$ was amplified by PCR using specific primers and plasmids containing the cDNA of LC3 and GATE16 templates, obtained from Dr. Antonio Zorzano lab (IRB, Barcelona, Spain). After DNA purification from agarose gel

(Figure 15), the cDNAs were subcloned into the NcoI/Acc65I restriction sites in MBP and Trx containing vectors (Figure 16).

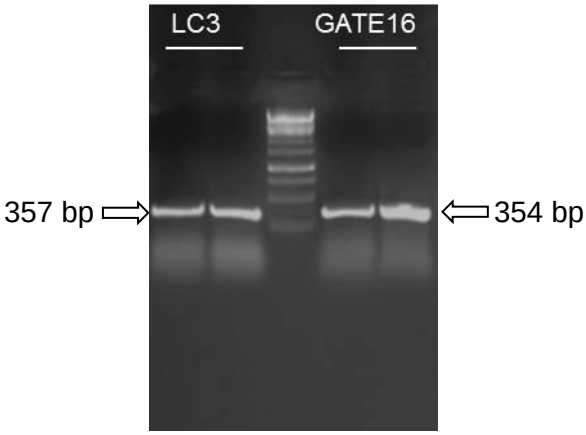


Figure 15- PCR amplification of LC3 and GATE16 cDNA. PCR products were analyzed by agarose gel electrophoresis and purified before insertion into the expression vectors.

The insertion into the selected expression vectors was confirmed by restriction and sequencing analysis.

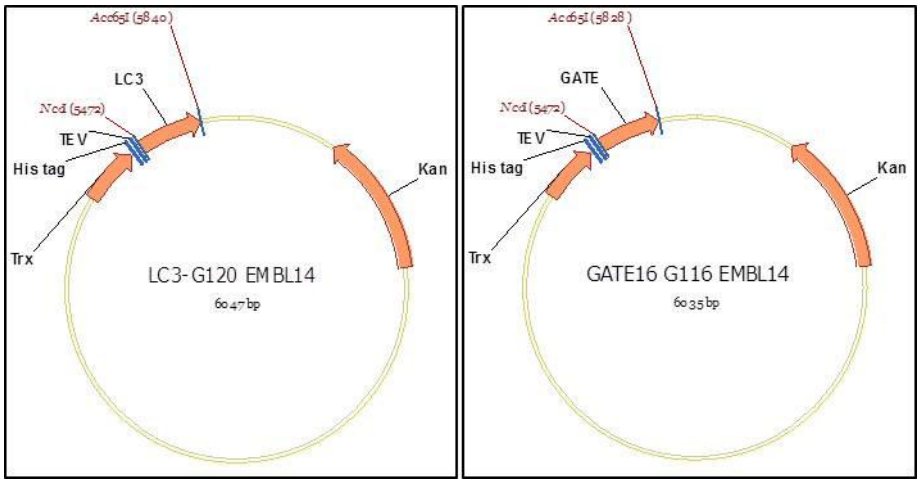


Figure 16- Schematic representation of LC3 and GATE16 cloned onto expression vectors. The proteins expressed from this vector contain an N-terminal Trx, followed by the hexahistidine-tag and a TEV protease recognition site.

The sequencing result was analyzed with VectorNTI sequence alignment software (Invitrogen) (Appendix IV).

3.5. LC3 and GATE16 Purification

The Trx containing vector was selected for experiment scale-up. A protocol for the expression and high-yield purification of Trx-LC3 and Trx-GATE16 was developed. Fusion proteins were expressed in *E.coli* BL21 Star (DE3) grown in LB medium at 25 ºC, and induced with 0.1 mM IPTG. Figure 17 shows an example of protein expression obtained under those conditions for LC3 and GATE16, showing that these proteins were expressed mainly in the soluble faction. A clear band with molecular weight of about 28 kDa was visible.

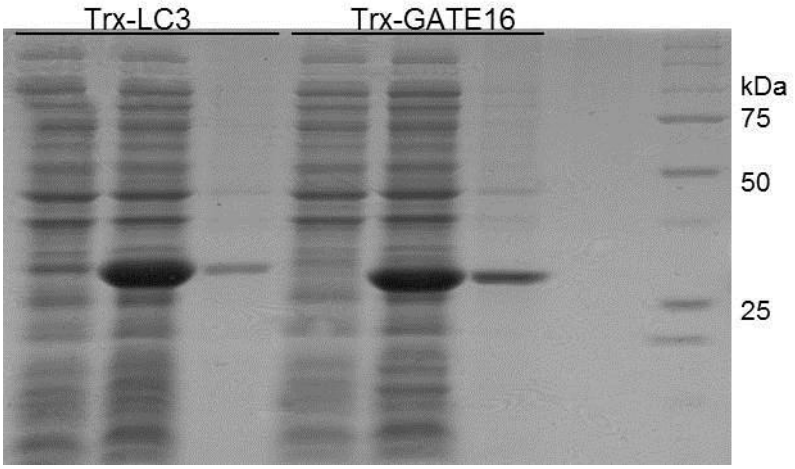


Figure 17 - Trx-LC3 and Trx-GATE expression. Trx-LC3 (28 kDa) and Trx-GATE (27 kDa) expression in *E.coli* Bl21 Star (DE3) strain at 25 ºC, mainly in soluble fraction (Sol). TF corresponds to the total protein extract before induction, Ins corresponds to insoluble fractions after induction.

These proteins were expressed with an N-terminal hexa-histidine tag, which enables purification by IMAC. Trx-LC3 was eluted using 50 mM imidazole, while Trx-GATE16 was eluted using 100 mM imidazole (data not shown). After changing the buffer by dialysis against 20 mM HEPES pH7.5; 100 mM NaCl; 1 mM EDTA, a TEV protease cleavage assay was performed by incubating for 3 h at 4 ºC with a protein to TEV protease molar ratio of 1:25.

The fusion tags were cleaved by a hexahistidine-tagged TEV protease and purified in a HisTrap column after the digestion. Both the tags and the TEV protease were retained in the HisTrap column, while LC3 and GATE16 were eluted in unbound fraction (Figure 18).

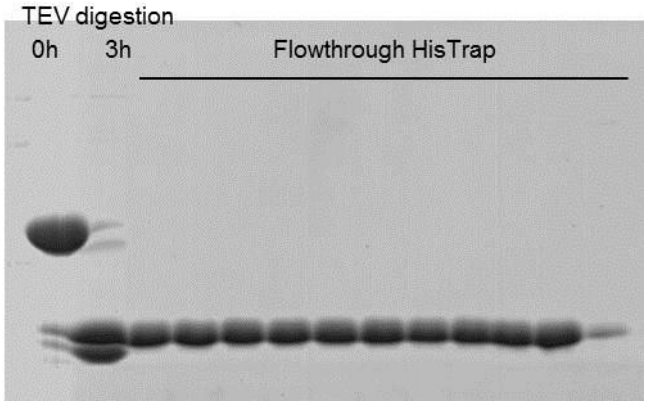
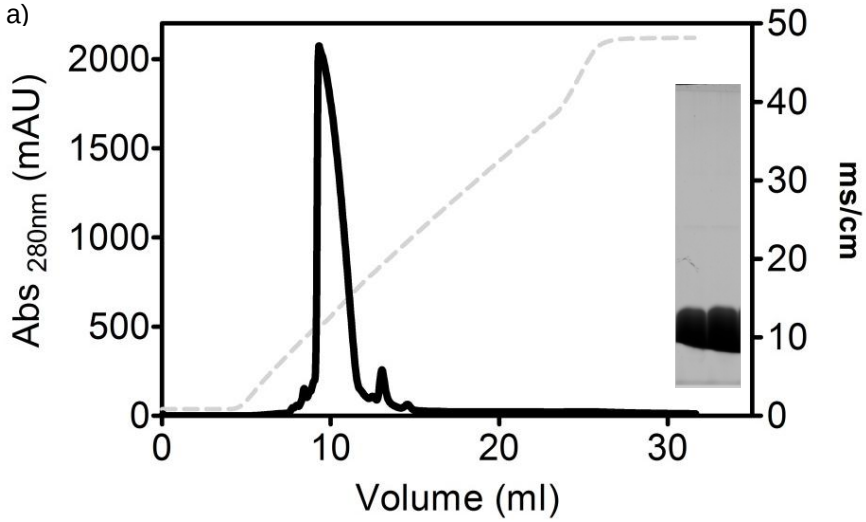


Figure 18 - Analysis of LC3 TEV protease cleavage. After TEV cleavage, the protein was applied to HisTrap, the flow through collected and analyzed in SDS-PAGE. Molecular weight expected: Trx-LC3 fusion protein (28 kDa), TEV protease (27 kDa) and LC3 (14kDa).

After buffer exchange by dialysis in 20 mM HEPES pH7.5; 1 mM EDTA, to remove the salt and the imidazole, the samples were loaded into a cationic exchange column (MonoS, 5'50GL, GE-Healthcare). LC3 and GATE16 were eluted by a linear gradient of 0-1M NaCl in 20 CVs (Figure 19).



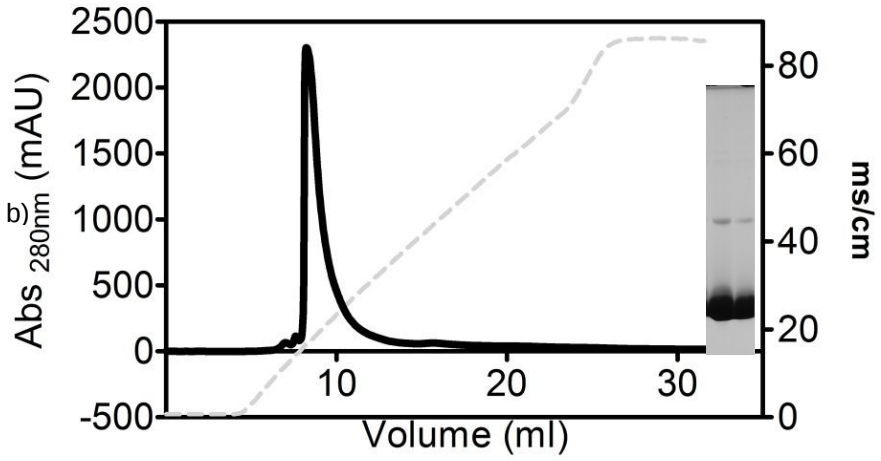


Figure 19 - Purification of LC3 (a) and GATE16 (b) in anionic exchange chromatography after IMAC. Fractions from IMAC, after TEV digestion, were applied to a MonoS 5'50GL column, equilibrated in 20 mM HEPES, pH 7.5 and 1 mM EDTA. The elution was carried out in a linear gradient of NaCl (0-1 M) in the same buffer, at a flow rate of 0.75 ml/min. The dotted line gray refers to conductivity. The fractions corresponding to the first peak were analyzed by SDS-PAGE.

Protein was quantified with by measuring the absorbance at 280 nm. The value obtained was used to calculate protein concentration using the theoretical extinction coefficient obtained with ProtParam tool (http://expasy.org/) (LC3 ε= 5960 M^{-1} cm^{-1}; GATE16 ε=18450 M^{-1} cm^{-1}). The protein was frozen in liquid nitrogen and stored at -80 ºC.

3.6. Structural analysis of the NusA-DOR fusion protein

The homogeneity, stability and secondary structure content of recombinant NusA-DOR were analyzed by DLS, DSF and CD spectroscopy.

3.6.1. Dynamic light scattering (DLS)

The homogeneity of the purified protein was assessed by DLS. The plots of the size distribution by intensity were converted to volume distribution for a more

accurate interpretation of the results, as usually done when size distribution by intensity shows a substantial tail [26].

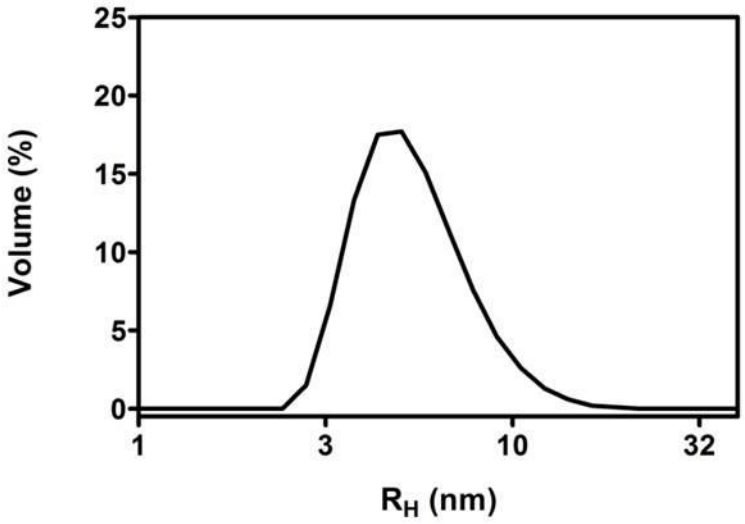


Figure 20 - Dynamic light scattering (DLS) analysis. Purified NusA-DOR, at 1 mg/ml, was analyzed by DLS in order to determine the homogeneity of the sample. The Hydrodynamic radius (R_H) value was obtained from plots of size distribution as a function of volume.

The value of RH, determined from the volume distribution, was used to estimate the molecular weight using an empirical calibration graph developed by Malvern Instruments available in the Nano Software [26], assuming a globular protein. The DLS parameters of NusA-DOR indicated it is a protein monomer, with a RH value of 4.360 nm, corresponds to a predicted molecular weight of 95 kDa and a polydispersity of 19.3%. As a rule of thumb, samples with % Pd <~20%, are considered to be monodisperse [26], [27].

3.6.2. Circular Dichroism analysis

Circular dichroism spectroscopy is a widely used technique to analyze the secondary structure of proteins in solution. Each type of polypeptide secondary structure gives rise to a characteristic CD spectrum in the Far UV region [29].

The secondary structure of NusA-DOR and NusA were determined by Circular dichroism spectroscopy using concentrations of 0.15 and 0.05 mg/ml respectively. The CD spectra show that both the secondary structure of proteins NusA and NusA-DOR is mainly composed of α-helices. However, the secondary structure of DOR, obtained for subtraction of the NusA spectrum from that of NusA-DOR, is mostly unstructured (Figure 21). The spectrum shows a single negative peak at around 202 nm and a shoulder around 222 nm, suggesting that the protein does not have a regular secondary structure [28].

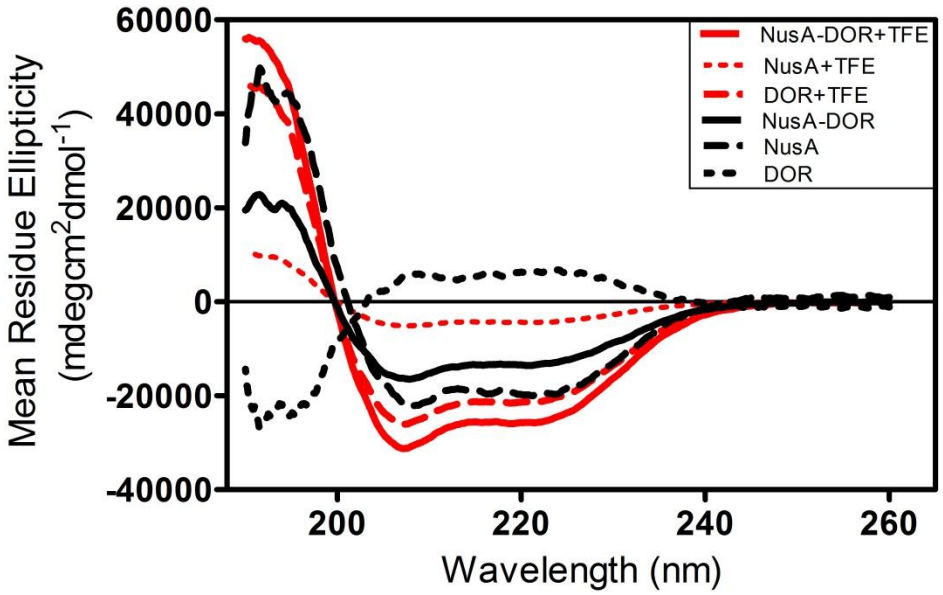


Figure 21 - Far UV CD spectra of NusA-DOR, NusA and DOR. DOR spectrum was obtained by subtracting the spectra of the isolated tag from that of the fusion protein. The spectra resulting from incubation of the proteins with TFE are shown in red.

The effect of trifluoroethanol (TFE) on the secondary structure of NusA and NusA-DOR was also tested. This reagent mimics hydrophobic environments and intracellular molecular crowding effects, inducing an increase in α-helical conformation if the proteins have a tendency to form this structure [39]. Analyzing the CD spectrum of DOR with 50 % of TFE (Figure 21), two negative peaks at 208 nm and 222 nm with a large positive peak at shorter wavelengths were observed, suggesting that DOR acquired an α-helical conformation upon incubation with TFE.

The CD spectrum of a protein can be deconvoluted to estimate its secondary structure content [29]. Analysis methods, typically, utilize as reference a database comprised of spectra of proteins whose three dimensional structures are known. The deconvolution was done with DICHROWEB web server using CONTIN analysis algorithms [30].

Table 4 – Deconvolution of the far UV CD spectra of the analysed proteins. Secondary structure of NusA-DOR, NusA and DOR (by subtraction) in the absence and presence of TFE. NMRS values: NusA-DOR, 0.065; NusA-DOR with TFE, 0.035; NusA, 0.081; NusA with TFE, 0.045; DOR, 0.397 and DOR with TFE, 0.048.

		% Struture		
		NusA-DOR	NusA	DOR
0% TFE	α-helix	40%	63%	0%
	β-sheets	6%	2%	28%
	β-turns	13%	12.5%	8%
	Coil	41%	22.5%	64%
50% TFE	α-helix	70%	100%	62%
	β-sheets	1%	0%	2%
	β-turns	10%	0%	11%
	Coil	19%	0%	25%

Table 5- Secondary structure of NusA from PDB (code 2KWP)

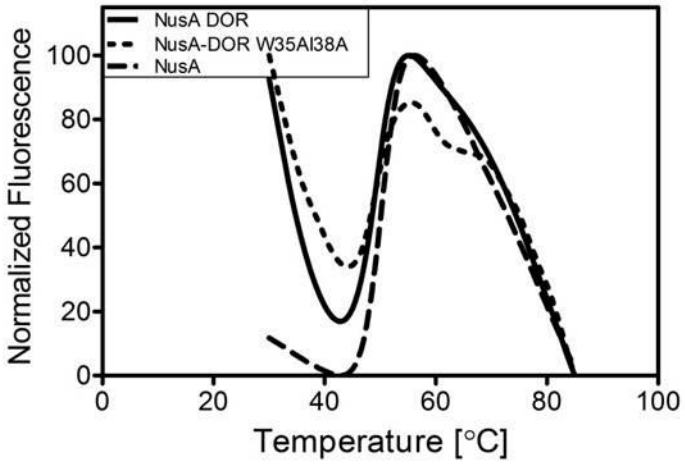
	% structure
α-helix	48%
β-sheets	19%

Upon deconvolution (Table 4) the secondary structure of DOR was 64% random coil, while with 50% TFE the total content of α-helix increased to 62% with 13% of β-structures and 25% random coil. Comparing the results obtained for NusA with the described in PDB (Table 5), these are similar.

3.6.3. DOR thermal stability assay

The thermal stability of all proteins was evaluated using DSF, a miniaturized biophysical binding assay. This technique monitors changes in the fluorescent intensity of dyes that are quenched in aqueous environments but increase in fluorescence on binding to hydrophobic core of denatured proteins [40]. As the temperature increases, the fraction of nonnative protein increases, producing a cooperative unfolding transition. The midpoint temperature, T_m, of such a transition is defined as the temperature at which the concentration of native and nonnative protein is equivalent [41].

Each protein was submitted to a temperature gradient (30-85 ºC) in the presence of Sypro orange and, as a result, a melting temperature curve was obtained. The Sypro orange dye was dilute into the buffer solution (1:500) before adding the protein, to prevent damage to the protein by contact with high concentrations of DMSO, since the Sypro orange dye is provided as a 5000 X solution in 100% (v/v) DMSO. Four assays for each protein were performed and the results obtained were analyzed and normalized with GraphPad (GraphPad Prism), and the average of the four replicates for each protein was calculated. The thermal stability of NusA-DOR fusion protein, NusA and NusA-DOR W35AI38A were analyzed and compared (Figure 22).



	T_m (ºC)
NusA	50.5±0.2
NusA-DOR	48.7±0.5
NusA-DOR W35AI38A	49.0±0.5

Figure 22 - Normalized differential scanning fluorimetry data. Fluorescence-based thermal shift curves and melting temperature of DOR constructs, NusA–DOR and NusA-DOR W35AI38A. These assays were done in 20 mM HEPES, pH7.5; 100 mM NaCl; 1 mM EDTA; 10% glycerol and 5 mM 2-mercaptoethanol.

NusA-DOR fusion proteins display a thermal unfolding profile commonly observed in proteins that do not crystallize [42]. The results obtained for NusA-DOR and NusA-DOR W35AI38A were very similar. The melting temperature of each protein differs only in 0.3 ºC, which indicates that mutant does not differ in stability when compared to the wild type. Altogether, the results show that DOR is a monomeric protein with a predominantly random coil structure, a feature that might explain the fact that all trials to obtain crystals of the fusion protein were unsuccessful. Table 6 summarized all the crystallization conditions assayed in High Throughput Crystallization Laboratory at EMBL Grenoble with NusA-DOR at 3.5 mg/ml.

Table 6 - Crystallization conditions assayed in High Throughput Crystallization Laboratory at EMBL Grenoble with NusA-DOR

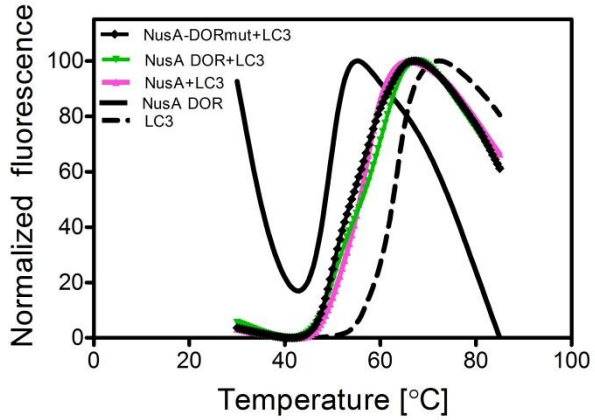
Screen	Supplier
The Classics	Qiagen/Nextal
Crystal Screen Lite & PEG/Ion	Hampton Research
MembFac & Natrix	Hampton Research
QuickScreen & Grid screens	Hampton Research - Home Made
Grid screens PEG 6K, PEG/LiCl,	Hampton Research – Home Made
Index Screen	Hampton Research

The analysis of the NusA-DOR sequence showed the presence of a region which similarity to the LIR motif common in LC3 interacting proteins [16]. Therefore, the interacting partners LC3 and GATE16 were purified for further biochemical and structural analysis. Also, for further mapping the role of different DOR regions on the interaction with those proteins, the peptide $_{27}$SEEDEVDGWLIIDLPDSYAA$_{46}$, representing the LIR motif of DOR was synthesized (GenScript, USA).

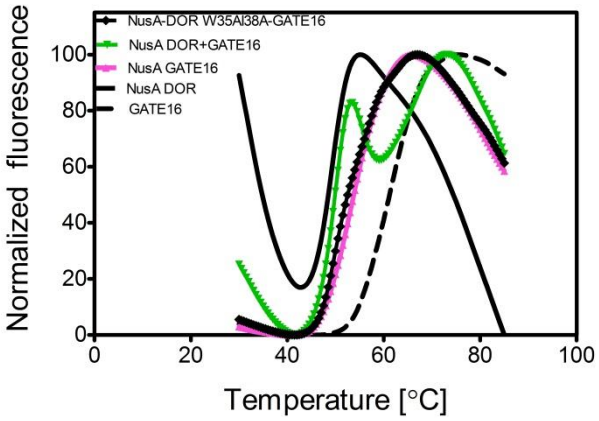
3.7. Analysis of the interaction between DOR and the autophagy-related proteins LC3 and GATE16

3.7.1. Thermal stability analysis of LC3 and GATE16 complexes

The thermal stability of NusA-DOR in complex with LC3 and GATE16 was analyzed and compared with that of NusA-DOR alone (Figure 23).



	T_m (ºC)
NusA-DOR + LC3	54.6±2.0
NusA-DOR W35AI38A + LC3	53.1±2.0
NusA + LC3	54.7±1.0
LC3	62.1±1.0
NusA	50.5±0.5
NusA-DOR	48.7±0.2
NusA-DOR W35AI38A	49.0±0.5



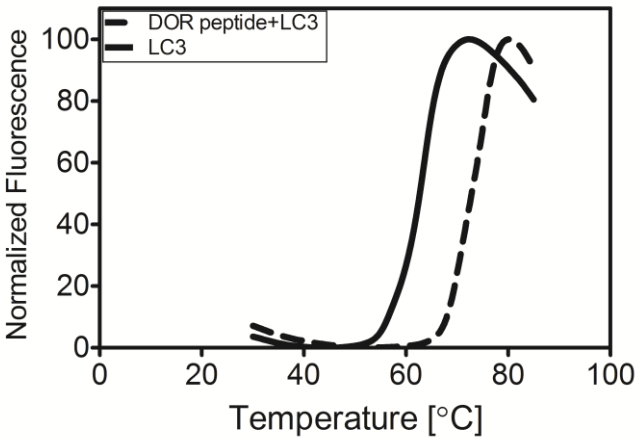
	T_m (ºC)
NusA-DOR + GATE16	n.d
NusA-DOR W35AI38A + GATE16	52.0±1.5
NusA + GATE16	52.9±1.5
GATE16	61.2±1.0
NusA	50.5±0.5
NusA-DOR	48.7±0.2
NusA-DOR W35AI38A	49.0±0.5

Figure 23- Normalized DSF data. Fluorescence-based thermal shift curves and melting temperature of NusA-DOR constructs in complex with LC3 and GATE16 1:3. These assays were done in 20 mM HEPES, pH7.5; 100 mM NaCl; 1 mM EDTA; 10% glycerol and 5 mM 2-mercaptoethanol

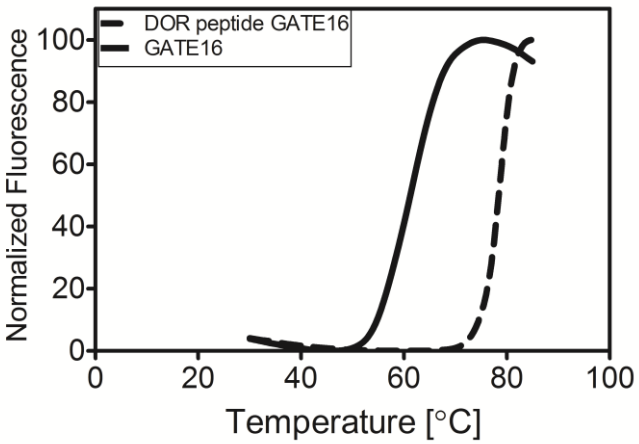
For NusA-DOR, LC3 induces an increase in stability, ΔT_m = +6 ºC, whereas for NusA-DOR W35AI38A the ΔT_m = +4 ºC. In the thermal stability assay with

GATE16, for NusA-DOR two melting curves were obtained and W35AI38A mutant increased 3 ºC, indicating that the unfolding of NusA-DOR and GATE16, likely occur independently. The complex does not seem to be as strong with GATE16 as it is with LC3. As the melting temperature is higher with DOR wild type than with the mutant, appears to be another interaction site with LC3.

To understand the effect of DOR–LIR peptide on the thermal stability of LC3 and GATE16 the DSF assay was done after incubation of the proteins with different protein:peptide ratios (Figure 24).



	T_m (ºC)
DOR peptide + LC3	72.8±2.0
LC3	62.1±1.0



	T_m (ºC)
DOR peptide + GATE16	74.5±2.0
GATE16	61.2±1.0

Figure 24 - Normalized DSF data. Thermal stability analysis was done for DOR peptide:LC3 or DOR peptide:GATE16 2:1 complexes. All assays were done in 20 mM HEPES, pH7.5; 100 mM NaCl; 1 mM EDTA.

Analyzing the results of melting temperature and melting curves it is possible verify that upon addition of the peptide an increase in melting temperature of 10 ºC and 13 ºC for LC3 and GATE16 complexes, respectively. In the case of a random peptide (MTQTSGTNLTSEELRKRREAYFEKQQQKQQ), no differences in stability were observed. These results support that the DOR peptide specifically interacts with LC3 and GATE16, forming a stable complex.

It has been shown that binding constants for interactions can be obtained thorough analysis of dose-response curves [40]. Therefore, a ligand-binding affinity assay using DSF was done using different concentrations of DOR peptide in complex with LC3 or GATE16. Protein thermal stability increases in this experiment by an amount that is proportional to the concentration and affinity of DOR peptide (Figure 25).

A typical dose-response curve was obtained for interaction between DOR peptide and LC3 or GATE16, as illustrated in the semi-log plot represented in Figure 25. The dose-response curves show an increase in T_m and reach a plateau at the highest DOR peptide concentrations. The constant for LC3 and GATE16 was determined by GraphPad analysis, k_d for LC3 was 2.2 µM and for GATE16 2.3 µM.

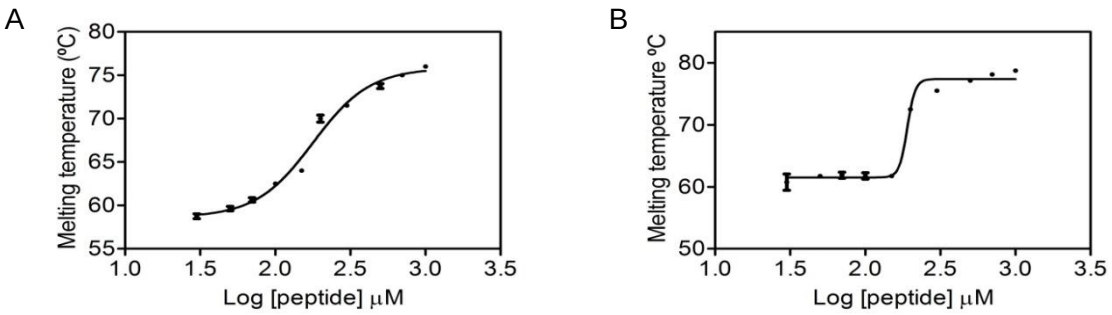


Figure 25 – Thermal stability of LC3 and GATE16 increases in the presence of DOR peptide. a) Analysis of LC3 in complex with increasing concentration of DOR peptide (30 µM; 50 µM; 70 µM; 100 µM; 150 µM; 200 µM; 300 µM; 500 µM; 700 µM; 1000 µM), with an affinity constant of 2.2 µM and a Robust Standard Deviation of the Residuals (RSDR) of 0.5. b) Thermal stability of GATE16 was measured in the presence of different DOR peptide concentrations (30 µM; 50 µM; 70 µM; 100 µM; 150 µM; 200 µM; 300 µM; 500 µM; 700 µM; 1000 µM), with an affinity constant of 2.3 µM and a RSDR of 1.1.

3.7.2. Circular Dichroism of DOR-LIR peptide/complexes

The conformational changes of the DOR peptide, when interacting with LC3 and GATE16 were studied by CD spectroscopy. The complexes were incubated three hours before the experiment at 4 ºC and 1:1 ratio peptide to protein were tested. The secondary structure of complexes where obtained by circular dichroism (Figure 26) using 7 µM of LC3, and 6 µM GATE.

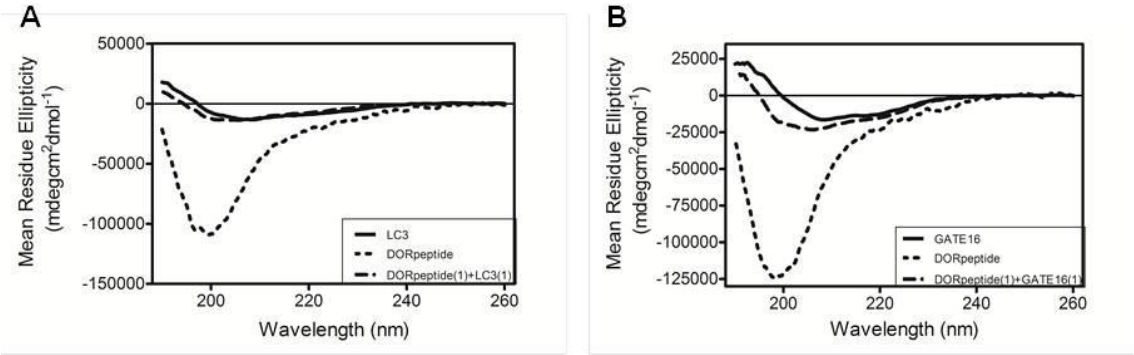


Figure 26 – CD spectra of DOR peptide/complexes. A) Far UV CD spectra of LC3, DOR peptide and DOR peptide:LC3 complex. B) Far UV CD spectra of GATE16, DOR peptide and DOR peptide:GATE16 complex.

The CD spectra show that DOR peptide is predominantly unstructured. The spectrum shows a single negative peak at around 202 nm, characteristic of random coil structures. The deconvolution was done with DICHROWEB web server using CONTIN analysis algorithms (Table 7 and 8) [30].

Table 7- Deconvolution of the far UV CD spectra of DOR peptide:LC3 complex using CONTIN analysis algorithms. NRMSD results for LC3, 0.11; DOR peptide, 0.08 and complex, 0.12.

	% Structure		
	LC3	DOR peptide	LC3+DOR peptide (1:1)
α-helix	22%	8%	14%
β-sheets	23%	0%	21%
B-turns	17%	0%	17%
Coil	38%	92%	48%

Table 8 Deconvolution of the far UV CD spectra of DOR peptide: GATE16 complex. NRMSD results for GATE16, 0.11; DOR peptide and complex, 0.16.

	% Structure		
	GATE16	DOR peptide	GATE16+DOR peptide (1:1)
α-helix	27%	11%	23%
β-sheets	25%	0%	9%
B-turns	23%	0%	21%
Coil	25%	89%	47%

Structural characteristics of the DOR peptide within the complex, can be observed from the difference spectra where the CD spectra of LC3 or GATE16 alone is subtracted from those of the DOR peptide complexes (Figure 27). This is valid if we assumes that LC3 or GATE16 do not suffer any significant conformational changes upon complex formation, assumption supported by evidences from experiment crystal structures of LC3-p62 complex (2ZJD) [16] and GABARAPL-1 with NBR1-LIR (2L8J) [17]. The peptide clearly shows an increase in secondary structure content upon complexion with LC3 or GATE16 (Table 9 and 10).

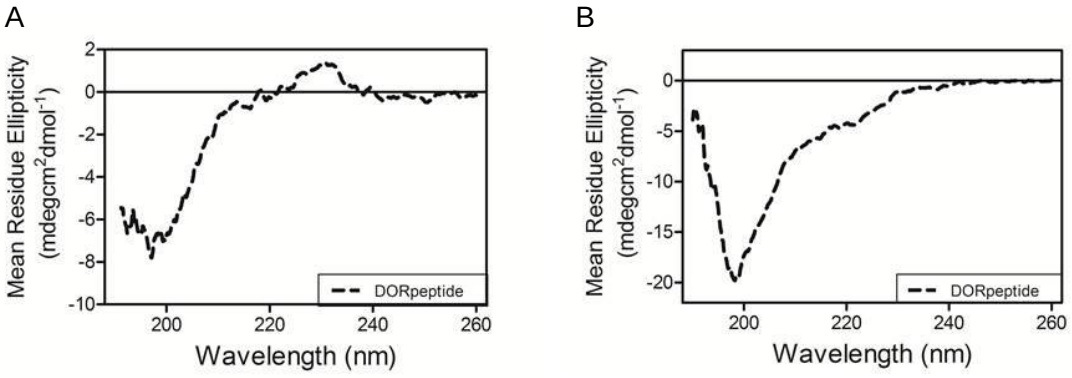


Figure 27 - Far UV CD spectra of DOR peptide, when interacting with LC3 (A) or GATE16 (B). The spectra were obtained by subtracting the LC3 or GATTE16 spectra from the LC3:DOR peptide or GATE16:DOR peptide spectra, respectively.

Table 9 - Deconvolution of the far UV CD spectra of DOR peptide subtracting LC3 from spectra obtained in complexes. NRMSD value 0.3.

	% Structure	
	DOR peptide	DOR peptide + LC3 (1:1) – LC3
α-helix	8%	14%
β-sheets	0%	21%
B-turns	0%	17%
Coil	92%	48%

Table 10 - Deconvololution of the far UV CD spectra of DOR peptide subtracting GATE16 from spectra obtained in complexes. NRMSD value 0.1.

	% Structure	
	DOR peptide	DOR peptide + GATE16 (1:1) – GATE16
α-helix	11%	5%
β-sheets	0%	17%
B-turns	0%	14%
Coil	89%	64%

The CD spectrum of DOR peptide measured alone is different from the CD spectrum of DOR peptide obtained by subtraction, indicating that interaction with LC3 or GATE16 induces conformational change on the interacting peptide.

3.7.3. Surface plasmon resonance (BIAcore)

To confirm the interactions of DOR with the autophagy-related proteins, LC3 and GATE16, an *in vitro* approach, surface plasmon resonance (SPR) was chosen. Using the BIAcore T-100 system (IRB, Barcelona, Spain), LC3 and GATE16 were immobilized into the surface of the sensor chip (CM5) and the potential interaction partners were injected in solution by flowing in HBSEP supplemented with 0.84 % of glycerol.

We performed SPR analysis to determine if: 1) DOR interacts with LC3 and GATE16; 2) DOR mutation in the LIR region abolishes LC3 and GATE16 interaction and 3) LIR peptide is the only binding region in DOR.

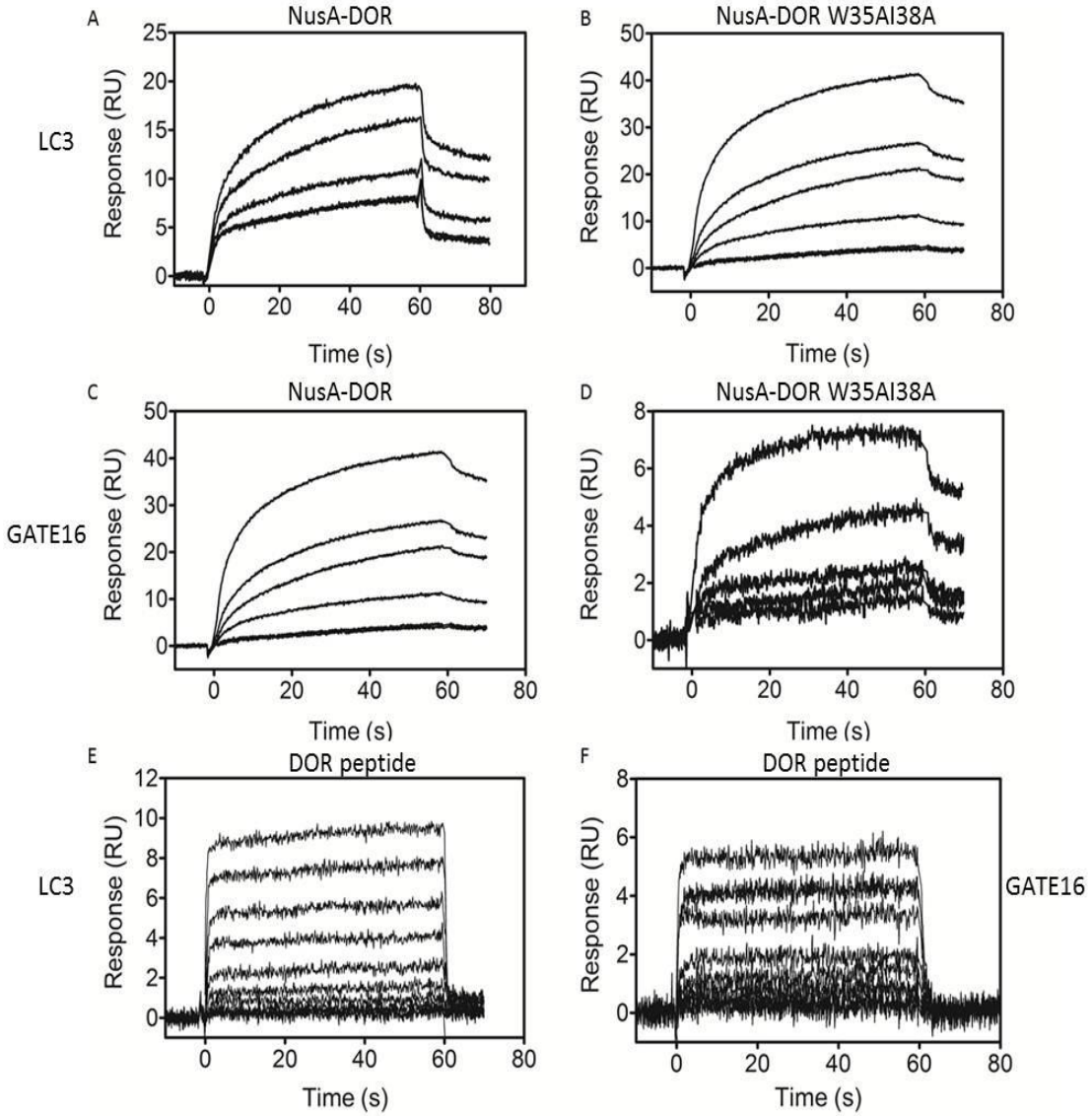


Figure 28 – Biacore sensorgrams. Sensorgram showing responses to the NusA-DOR at 31.25 nM; 62.5 nM; 125 nM and 250 nM (A), Nus-DOR W35AI38A using 31.25 nM; 62.5 nM; 125 nM; 250 nM and 500 nM concentrations (B), in multiple cycle kinetics analysis to immobilized LC3. Multiple cycle kinetics analysis to immobilized GATE16 with NusA-DOR using 31.25 nM; 62.5 nM; 125 nM and 250 nM (C), Nus-DOR W35AI38A at 31.25 nM; 62.5 nM; 125 nM; 250 nM and 500 nM concentrations (D). Sensorgram showing response to DOR peptide to immobilized LC3 (E) and to immobilized GATE16 (F) using 58.63 nM; to 30 000 nM using doubling concentrations.

Table 11 – Binding kinetics. Data sets were analyzed and fitted with a 1:1 binding model, generating global rate constants (association rate k_a, dissociation rate k_d, and equilibrium dissociation constant k_D).

		k_a (1/Ms)	k_d (1/s)	k_D (M)	Chi^2
LC3	NusA-DOR	2.2E+5	0.01	5.5E-8	0.2
	NusA-DOR W35AI38A	1.1E+5	0.02	1.6E-7	1.3
	DOR peptide	8.7E+4	0.67	7.7E-6	0.09
GATE16	NusA-DOR	6.2E+5	0.07	1.2E-7	0.09
	NusA-DOR W35AI38A	2.0E+5	0.02	1.1E-7	0.09
	DOR peptide	1.3E+5	0.63	4.9E-6	0.09

To obtain the kinetic parameters, sensorgrams were modeled with BIAevaluation software 3.0 using 1:1 binding model. For all reactions, individual sensorgrams that did not fit well were excluded from the global fit for kinetic analysis. For the NusA-DOR, analyte concentrations of 500 nM, 1000 nM and 5000 nM yielded a poor fit. Individual curves use for the global analysis of NusADOR interactions were generated by NusA-DOR concentrations of 31.25 nM; 62.5 nM; 125 nM and 250 nM.

In the case of NusA-DOR W35AI38A, it was predicted that these amino acid changes would reduce LC3 binding affinity. For the DOR mutant, binding at 1000 nM and 5000 nM resulted in sensorgrams with a poor fit to the binding model. Curves derived from 31.25 nM; 62.5 nM; 125 nM; 250 nM and 500 nM concentrations exhibited good individual fits and were subjected to global analysis.

For the DOR-LIR peptide all the analyte concentrations of 58.63 nM; 117.3 nM; 234.5 nM; 469 nM; 938 nM; 1875 nM; 3750 nM; 7500 nM; 15 000 nM and 30 000 nM were used for the global analysis.

Analysis of the data from SPR revealed a high-affinity interaction between LC3 and NusA-DOR (k_D = 54.9 nM). This interaction was higher than that with GATE16 (k_D = 117 nM). K_D values of NusA-DOR mutant were 155 nM for LC3 interaction and 105 nM for GATE16. The binding analysis showed that DOR mutation in LIR domain does not abolish LC3 and GATE16 interaction. The mutation only reduces

the affinity for LC3 at about three times, however the affinity for GATE16 is only marginally affected.

When we analyzed the results for the interaction with DOR-LIR peptide, k_D values are higher than NusA-DOR, in other words the interaction is weaker. This result can be explained by the existence of another binding region in DOR or by the requirement of a particular conformation that the peptide only acquires in the full-length protein. The k_d and k_a were much faster for LC3-DOR peptide and GATE16-DOR peptide than for LC3-NusADOR and GATE16-NusADOR. Comparing the k_d obtained from thermal shift analysis, by semi-log plot, the results obtained for DOR peptide in interaction with LC3 and GATE16 are in the same magnitude.

These results indicated that NusA-DOR is able to form a complex with LC3 and GATE16 but there must be another site of interaction besides the LIR domain, since the mutation in DOR did not abolish the interaction and the affinity for DOR peptide was lower than that of full length DOR.

3.8. Crystallization of DOR-peptide in complex with LC3

Crystallization assays with NusA-DOR with LC3 (2:1 ratio, at 4.1 mg/ml) was performed at High Throughput Crystallization Laboratory at EMBL Grenoble, using the conditions summarized in Table 12.

Table 12 - Crystallization conditions assayed in High Throughput Crystallization Laboratory at EMBL Grenoble with NusA-DOR in complex with LC3

Screen	Supplier
The Classics	Qiagen/Nextal
Crystal Screen Lite & PEG/Ion	Hampton Research
MembFac & Natrix	Hampton Research
QuickScreen & Grid screens	Hampton Research - Home Made
Grid screens PEG 6K, PEG/LiCl,	Hampton Research – Home Made
Index Screen	Hampton Research

Crystallization assays of DOR peptide in complex with LC3 and GATE16 were performed in House (IBMC, Porto), at 2:1 and 3:1 ratios, using Morpheus MD 1-46; PACT premier MD 1-29; Proplex MD 1-38 screening kits from Molecular Dimensions and PEG/ION screen from Hampton Research. DOR complexes were tested too at EMBL Grenoble using conditions referred in table 6 using the sitting drop method. From High Throughput Crystallization Laboratory crystals of LC3 in complex with DOR peptide were obtained at ratio 2:1 (14.8 mg/ml) using 0.05 M Potassium Phosphate and 20% (v/v) PEG 8000 as precipitant, with The Classics from Qiagen kit. Crystals of the 3:1 LC3:DOR peptide complex (12.4 mg/ml) were also obtained with Peg/LiCl kit from Hampton Research in condition 1.0 M lithium chloride, 0.1 M PEG 6000 and 30% PEG 6000 as precipitant. Both the conditions were reproduced and further optimized in the lab.

Crystals of the 3 (peptide):1 (LC3) complex were grown using 1.0 M lithium chloride, 30% PEG 6000 and 0.1 M MES pH6.0 as precipitant. Crystals appeared after one week (Figure 29, A) and were directly plunged into liquid nitrogen for cryo-cooling before data collection. Optimization with 0.2 M of TMAO originated larger crystals after two weeks (Figure 29, B), and were used for X-ray diffraction analysis.

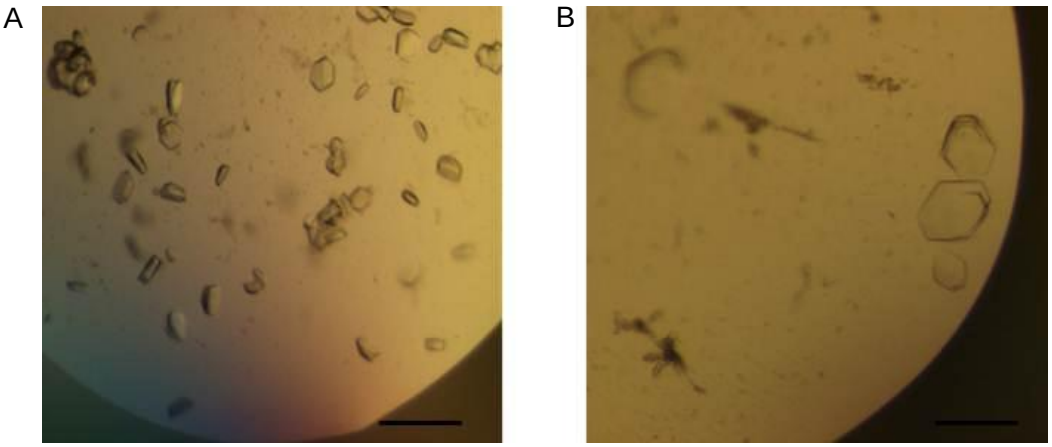


Figure 29 – Crystals of DOR peptide: LC3 complex (3:1) obtained using 30% PEG 6000 and 0.1 M MES pH6.0 as precipitant. A) The drop contained equal volumes of DOR complex solution and precipitant. The crystals were first observed on the 8th day. B) Optimization crystals upon addition of 0.2 M of TMAO (scale bars, 0.1 mm).

Crystals for DOR peptide in complex with LC3, 2:1 ratio, appeared after one month (Figure 30, A). These crystals were used for optimization with seed beads (Hampton Research). Using small crystals obtained before, a seed stock was created to perform subsequent experiments. A solution with seeds was pipetted in a drop and used to grow larger crystals. New crystals were observed two weeks after the incubation of complex with the precipitant (Figure 30, B).

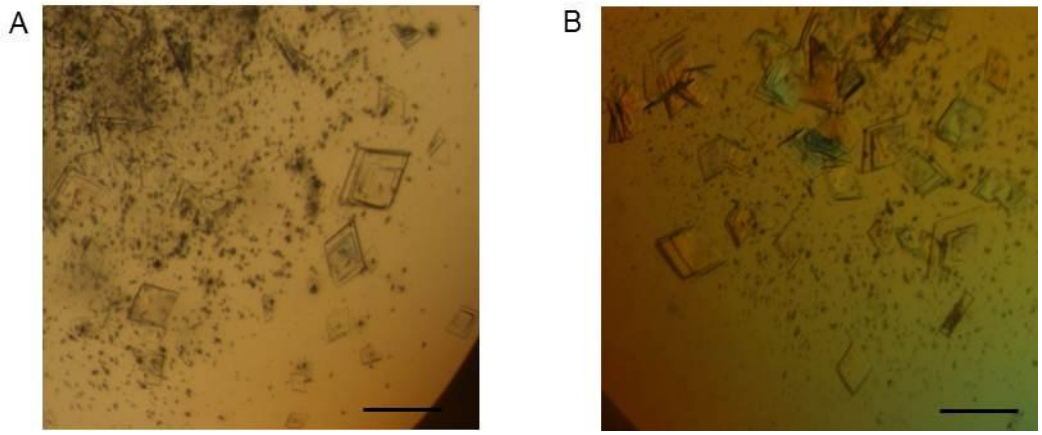


Figure 30 – Crystals of DOR peptide and LC3 complex (2:1) obtained using 0.1 M L-cysteine and 0.05 M Potassium phosphate with 20% PEG 8000 as precipitant. A) The drop contained 1 µl of complex solution; 0.7 µl precipitant and 0.3 µl of cysteine. B) The drop was an optimization of A) with seed beads (scale bars, 0.1 mm).

The crystals this obtained were cryo-cooled in liquid nitrogen and used for X-ray diffraction. The collected data was used to determine the structure of complex.

3.9. Data collection, processing and refinement

Diffraction data was collected from a single cryo-cooled crystal to 2.3 Å. The crystal belonged to the orthorhombic space group $C2_1$, with unit cell constant of a = 56.33 Å, b = 42.53 Å, c = 53.51 Å, and 44.6 % of solvent content. The coordinates of LC3-p62 complex from PDB entry 2ZJD [16] were used as a search model to solve the structure of LC3-DOR peptide. Alternate cycles of manual building with Coot [36] and refinement with PHENIX [35] allowed partial building of

the experimental three-dimensional structure of the DOR-peptide/LC3 complex. Current statistics on data collection, processing and partial refinement are summarized in Table 13.

Table 13 – Data collection and refinement

Crystallographic analysis		
Resolution range Å	2.3 – 7.3	
Space group	$C2_1$	
Unit cell dimensions	a=56.33 b=42.53 c=53.51	β=103.5
Multiplicity (overall/outer shell)	3.0	
R_{merge} [a] (overall/outer shell)	8.3%	
Completeness (%) (overall/outer shell)	99.6	
I/σ (I) (overall/outer shell)	6.6	
Mathews coefficient (Å Da^{-1})	2.2	
Solvent content	44.6	
Structure Refinement		
R_{factor} [b] / Free R_{factor} [c] (%)	21.2/ 26.6	
Nº of unique reflections (working/test set)	5379	
Water molecules	12	
Total number of non-hydrogen atoms	1074	
Number of protein non-hydrogen atoms	1074	
r.m.s.d. bond lengths (Å)	0.08	
r.m.s.d. bond angles (º)	1.17	
Ramachandran plot statistics		
Residues in allowed regions	96	
Residues in disallowed regions	1.6	

3.10. Structure of DOR-peptide in complex with LC3

The final model of LC3 – DOR peptide complex consists of residues 5-120 of LC3 and residues 29-43 of DOR peptide. The structure of LC3, which consists of a five-stranded β-sheet and four α-helices, is similar to the previously reported structure for the LC3-p62 LIR peptide complex [15].

LIR binds to one side of the LC3 surface apposite to C-terminal region, in a narrow channel, acquiring a β- hairpin conformation (Figure 31).

Figure 31- Structure of LC3-DOR peptide complex. The structure of DOR-LIR peptide (red) shows that complex formation extends the central β-sheet of LC3. The DOR peptide acquires an anti-parallel β-hairpin structure when in complex with LC3.

The two conserved amino acids, Trp^{35} and Ile^{38}, insert into well delineated pockets in LC3 a structure (Figure 32). There are two striking features of this interaction; one is the indole ring of the strictly conserved DOR Trp^{35} inserted into a site surrounded by LC3 residues, Asp^{19}, Ile^{23}, Pro^{32}, Lys^{51}, Leu^{53} and Phe^{108}. The other is between Ile^{38} of DOR-LIR which makes van der Waals contacts with Ile^{35}, Phe^{52}, Val^{54}, Ile^{66} and Arg^{70} of LC3. Other important site of interaction is between the N-terminal of DOR-LIR, Glu^{29}, and a basic conserved group on the surface of LC3, composed of Arg^{10}, Arg^{11} and Lys^{49}.

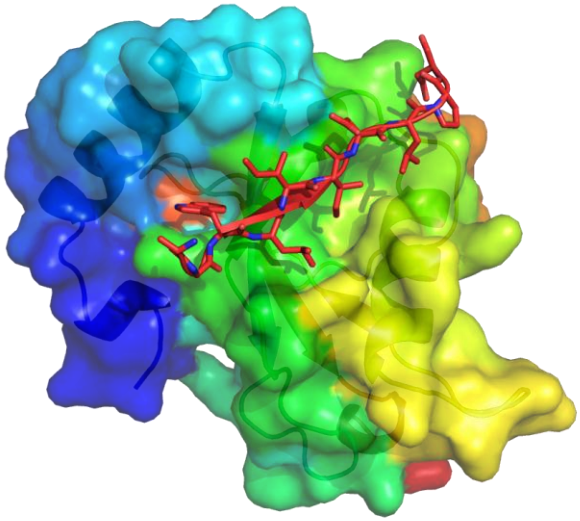


Figure 32 – Surface representation of LC3-DOR peptide complex. DOR-LIR is represented as sticks binding to the groove of LC3 surface.

Comparing this structure with other structures described for proteins that have a LIR domain in complex with LC3 or GABARAP-L, the hydrophobic amino acids are always essential for the interaction. The structure of the LC3-p62 peptide revealed interaction with Trp^{340} and Leu^{343} of p62 (PDB code 2ZJD) with the same hydrophobic pockets on LC3 (Figure 33, A) [16]. The same is observed with NBR1(PDB code 2L8J), where it was shows that the YIII motif is crucial for its interaction with all members of the MAP1LC3 protein family (Figure 33, B) [17].
However, this structure suggests that the LIR domain of DOR associates with amino acids of LC3 in a hairpin conformation, different from the other known structures known of LC3 in complex with LIR motifs [16], [17].

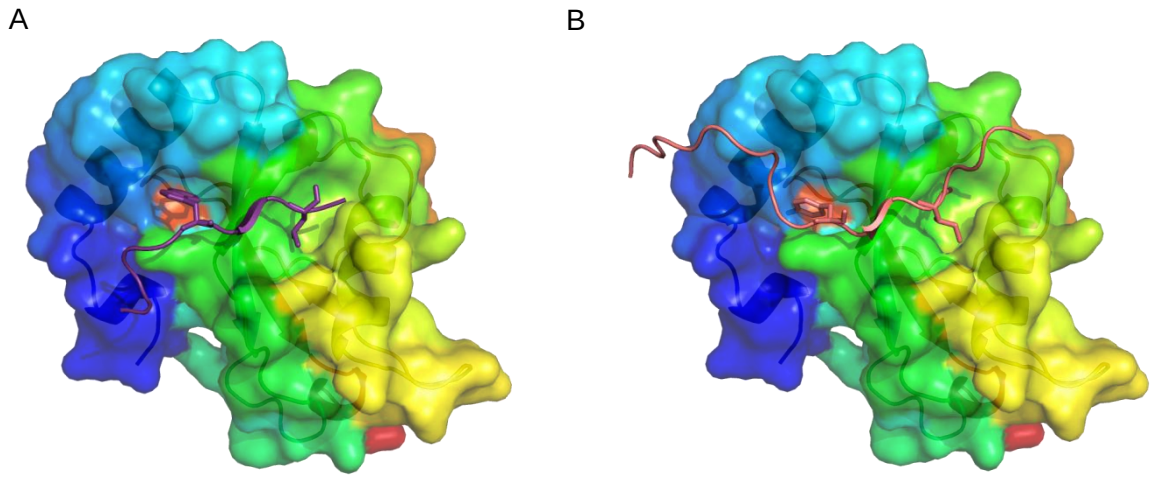


Figure 33 – Surface representation of A) LC3-p62 peptide interaction with Trp^{340} and Leu^{343} of p62 (PDB code 2ZJD) and B) GABARAPL-1 NRBR-LIR complex (PDB code 2L8J) interaction with Tyr^{732} and Ile^{735} of NRBR.

This structure is still not completely refined, and a more complete dataset is required to further analyze the molecular details of this novel interaction. This will require growth of better crystals and data collection in a synchrotron radiation source.

4. Conclusions

In this section are summarize the main conclusions of this work, regarding the expression and purification of NusA-DOR, and the expression and purification of its interactors and the interaction between them. Our interest in this work comes essentially from the recent discovery that DOR is implicated in autophagy, interacting with LC3 and GATE16 in the autophagosome formation.

We have succeeded in establishing a protocol for expression and purification of pure and homogeneous recombinant NusA-DOR fusion protein. Biochemical and biophysical studies, DLS, DSF and CD spectroscopy, were pursued with the NusA-DOR fusion protein.
The CD spectra of DOR, obtained for subtraction of the NusA spectrum from that of NusA-DOR shows a single negative peak, around 202 nm and a shoulder around 222 nm, a characteristic of random coil structures, suggesting that the protein is mostly unstructured, with 64% random coil. The detailed structural characterization of the NusA-DOR fusion protein seems to raise the hypothesis that DOR might be a predominantly random coil structure, a feature that might explain the fact that all trials to obtain crystals of the fusion protein were unsuccessful. However, incubation with TFE (50%) induces change conformation in DOR, the total content of α-helix increased to 62% with 13% of β-structures and 25% random coil, as expected, since TFE induces an increase in α-helical conformation [39]. Crystallization trials can be made in presence of TFE to improve the formation of NusA-DOR crystals.

Previous analysis of the interaction between LC3/GATE16 and DOR showed that it was dependent on the integrity of LIR identified in DOR [12] and therefore these interaction partners were expressed and purified for quantitative interaction analysis. Biochemical and biophysical characterization of the interaction was performed using different techniques such as DSF, CD and SPR. The melting temperatures of each protein and complex were compared with DSF. The results obtained indicate that DOR peptide strongly interacts with LC3 and GATE16, stabilizing the proteins by 10-13 ºC. CD analysis, show that, upon interaction with LC3 or GATE16, the peptide clearly shows an increase in secondary structure

content, mainly β-structure. As shown by SPR the protein interacts with LC3 and GATE16 and this interaction is in part mediated by the LIR motif identified in DOR, but there must be another site of interaction, since (1) the mutation in DOR, W35A I38A, did not completely abolish the interaction, and (2) the affinity for DOR peptide was lower than that of full-length DOR.

All the information achieved in these biophysical analyses were essential for preparing complexes to obtain crystals of LC3 in complex with a DOR peptide containing the LIR motif, and to determine the structure of the complex at 2.3 Å resolution. This structure shows that the LIR domain of DOR associates with LC3, extending the central LC3 β-sheet and adopting a hairpin conformation. The crystal structure also showed the atomic details of the interaction of the hydrophobic (WXXI) motif in LIR with narrow pockets on the LC3, similarly to what is found in other structures of LC3/LC3 orthologs in complex with LIR motifs (e.g. LC3-p62 complex and NBR1-LIR, [16],[17]). However, this hairpin found for the DOR peptide conformation has not been observed before in LIR peptides.

A lot of aspects about this interaction were clarified but much more remains to be addressed. Since the affinity of LC3/GATE16 for the LIR motif was weaker than the one determined for full-length DOR, another construct with the N-terminal of DOR (1-46) will be prepared, to further define the boundaries of the interaction site(s), since recent studies show that the interaction with LC3 occurs in the N-terminal of DOR [5]. The structure is still not completely refined, and a more complete dataset is required to further analyze the molecular details of this novel interaction. This will require growth of better crystals and data collection in a synchrotron radiation source.

5. Bibliography

[1] B. G. Baumgartner et al., “Identification of a novel modulator of thyroid hormone receptor-mediated action.,” *PloS one*, vol. 2, no. 11, p. e1183, Jan. 2007.

[2] J. S. Bennetts, N. D. Rendtorff, F. Simpson, L. Tranebjaerg, and C. Wicking, “The coding region of TP53INP2, a gene expressed in the developing nervous system, is not altered in a family with autosomal recessive non-progressive infantile ataxia on chromosome 20q11-q13.,” *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 236, no. 3, pp. 843-52, Mar. 2007.

[3] J. Nowak, D. Depetris, J.-L. Iovanna, M.-G. Mattei, and M.-J. Pébusque, “Assignment of the tumor protein p53 induced nuclear protein 2 (TP53INP2) gene to human chromosome band 20q11.2 by in situ hybridization,” *Cytogenetic and Genome Research*, vol. 108, no. 4, p. 362B-362B, 2005.

[4] R. Tomasini et al., “TP53INP1s and homeodomain-interacting protein kinase-2 (HIPK2) are partners in regulating p53 activity.,” *The Journal of biological chemistry*, vol. 278, no. 39, pp. 37722-9, Sep. 2003.

[5] A. Sancho et al., “DOR/Tp53inp2 and Tp53inp1 constitute a metazoan gene family encoding dual regulators of autophagy and transcription.,” *PloS one*, vol. 7, no. 3, p. e34034, Jan. 2012.

[6] V. a Francis, A. Zorzano, and A. a Teleman, “dDOR is an EcR coactivator that forms a feed-forward loop connecting insulin and ecdysone signaling.,” *Current biology : CB*, vol. 20, no. 20, pp. 1799-808, Oct. 2010.

[7] a Longatti and S. a Tooze, “Vesicular trafficking and autophagosome formation.,” *Cell death and differentiation*, vol. 16, no. 7, pp. 956-65, Jul. 2009.

[8] Z. Yang and D. J. Klionsky, “Mammalian autophagy: core molecular machinery and signaling regulation.,” *Current opinion in cell biology*, vol. 22, no. 2, pp. 124-31, Apr. 2010.

[9] C. Mauvezin et al., “The nuclear cofactor DOR regulates autophagy in mammalian and Drosophila cells.,” *EMBO reports*, vol. 11, no. 1, pp. 37-44, Jan. 2010.

[10] J. Spowart and J. J. Lum, “Opening a new DOR to autophagy.,” *EMBO reports*, vol. 11, no. 1, pp. 4-5, Jan. 2010.

[11] J. Nowak et al., “The TP53INP2 Protein Is Required for Autophagy in Mammalian Cells,” *Molecular Biology of the Cell*, vol. 20, pp. 870 - 881, 2009.

[12] A. Sancho et al., “DOR and Tp53inp1 constitute a metazoan gene family , and encode for dual regulators of autophagy and transcription Submission intended as a Research Article,” *Molecular Cancer Research*.

[13] Y. Kabeya, N. Mizushima, A. Yamamoto, S. Oshitani-Okamoto, Y. Ohsumi, and T. Yoshimori, “LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation.,” *Journal of cell science*, vol. 117, no. Pt 13, pp. 2805-12, Jun. 2004.

[14] K. R. Drake, M. Kang, and A. K. Kenworthy, “Nucleocytoplasmic distribution and dynamics of the autophagosome marker EGFP-LC3.,” *PloS one*, vol. 5, no. 3, p. e9806, Jan. 2010.

[15] K. Sugawara, N. N. Suzuki, Y. Fujioka, N. Mizushima, Y. Ohsumi, and F. Inagaki, “The crystal structure of microtubule-associated protein light chain 3 , a mammalian homologue of Saccharomyces cerevisiae Atg8,” *Genes to Cells*, pp. 611-618, 2004.

[16] Y. Ichimura et al., “Structural basis for sorting mechanism of p62 in selective autophagy.,” *The Journal of biological chemistry*, vol. 283, no. 33, pp. 22847-57, Aug. 2008.

[17] A. Rozenknop et al., “Characterization of the interaction of GABARAPL-1 with the LIR motif of NBR1.,” *Journal of molecular biology*, vol. 410, no. 3, pp. 477-87, Jul. 2011.

[18] N. N. Noda, Y. Ohsumi, and F. Inagaki, “Atg8-family interacting motif crucial for selective autophagy.,” *FEBS letters*, vol. 584, no. 7, pp. 1379-85, Apr. 2010.

[19] N. N. Noda et al., “Structural basis of target recognition by Atg8/LC3 during selective autophagy.,” *Genes to cells : devoted to molecular & cellular mechanisms*, vol. 13, no. 12, pp. 1211-8, Dec. 2008.

[20] “TechnicalInformation Critical factors for successful protein crystallization Sample & Assay Technologies.”

[21] I. Benoit, B. Coutard, R. Oubelaid, M. Asther, and C. Bignon, “Expression in Escherichia coli, refolding and crystallization of Aspergillus niger feruloyl esterase A using a serial factorial approach.,” *Protein expression and purification*, vol. 55, no. 1, pp. 166-74, Sep. 2007.

[22] G. E. Healthcare, “Recombinant Protein Purification Handbook,” *Methods*.

[23] G. E. Healthcare, “Gel Filtration,” *Methods*.

[24] G. E. Healthcare, “Purifying Challenging Proteins,” *Methods*.

[25] S. Nallamsetty et al., “Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro.,” *Protein expression and purification*, vol. 38, no. 1, pp. 108-15, Nov. 2004.

[26] D. Arzen, R. P. Co-advisor, and D. K. Ljubljana, “Dynamic light scattering and application to proteins in solutions,” 2010.

[27] Z. Sárkány, T. P. Ikonen, F. Ferreira-da-Silva, M. J. Saraiva, D. Svergun, and A. M. Damas, "Solution structure of the soluble receptor for advanced glycation end products (sRAGE).," *The Journal of biological chemistry*, vol. 286, no. 43, pp. 37525-34, Oct. 2011.

[28] L. Whitmore and B. a Wallace, "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases.," *Biopolymers*, vol. 89, no. 5, pp. 392-400, May 2008.

[29] S. M. Kelly, T. J. Jess, and N. C. Price, "How to study proteins by circular dichroism.," *Biochimica et biophysica acta*, vol. 1751, no. 2, pp. 119-39, Aug. 2005.

[30] L. Whitmore and B. a Wallace, "DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data.," *Nucleic acids research*, vol. 32, no. Web Server issue, pp. W668-73, Jul. 2004.

[31] M. D. Cs, M. A. Farnum, and M. I. Nelen, "Journal of Biomolecular Screening," *Journal of Biomolecular Screening*, 2006.

[32] S. S. Handbook, "Biacore," *Strategy*.

[33] P. Torreri, M. Ceccarini, P. Macioce, T. C. Petrucci, and B. Cellulare, "Biomolecular interactions by Surface Plasmon Resonance technology," *differences*, vol. 41, no. 4, pp. 437-441, 2005.

[34] "An introduction to Biacore ' s SPR technology Biacore ' s SPR technology – at the core of scientific research," *Molecules*.

[35] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software.," *Journal of applied crystallography*, vol. 40, no. Pt 4, pp. 658-674, Aug. 2007.

[36] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot.," *Acta crystallographica. Section D, Biological crystallography*, vol. 66, no. Pt 4, pp. 486-501, Apr. 2010.

[37] "PHENIX: a comprehensive Python-based system for macromolecular structure solution." [Online]. Available: http://journals.iucr.org/d/issues/2010/02/00/dz5186/dz5186.pdf. [Accessed: 01-May-2012].

[38] N. a Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, and O. Gileadi, "Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study.," *Protein expression and purification*, vol. 59, no. 1, pp. 94-102, May 2008.

[39] M. Matsubara, N. Hayashi, K. Titani, and H. Taniguchi, "Circular Dichroism and 1 H NMR Studies on the Structures of Peptides Derived from the Calmodulin-binding

Domains of Inducible and Endothelial Nitric-oxide Synthase in Solution and in Complex with Calmodulin," *Biochemistry*, vol. 272, no. 37, pp. 23050-23056, 1997.

[40] J. Kervinen et al., "Effect of construct design on MAPKAP kinase-2 activity, thermodynamic stability and ligand-binding affinity.," *Archives of biochemistry and biophysics*, vol. 449, no. 1–2, pp. 47-56, May 2006.

[41] T. M. Mezzasalma et al., "Enhancing recombinant protein quality and yield by protein stability profiling.," *Journal of biomolecular screening*, vol. 12, no. 3, pp. 418-28, Apr. 2007.

[42] U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund, "Thermofluor-based high-throughput stability optimization of proteins for structural studies.," *Analytical biochemistry*, vol. 357, no. 2, pp. 289-98, Oct. 2006.

6. Appendix

I- Bacterial culture media

i) Luria Bertani (LB medium)

1% tryptone
1% NaCl
0.5% Yeast extract

Dissolve tryptone, NaCl and yeast extract in deionized water. Sterilize by autoclaving.

ii) 2YT

1.6% Bacto Trypton
1% NaCl
0.5% Yeast extract

Dissolve Bacto Trypton, NaCl and yeast extract in deionized water. Adjust pH to 7.0 with 1 M NaOH. Sterilize by autoclaving.

iii) ZYM-5052 auto-inducing complex medium

1% N-Z-amine AS
0.5% yeast extract
25 mM Na_2HPO_4
25 mM KH_2PO_4
50 mM NH_4Cl
5 mM $NaSO_4$
2 mM $MgSO_4$
0.2% trace metals
0.05% glycerol
0.05% glucose
0.2% α-lactose

Trace metals only add after sterilize the medium by autoclaving.

II- DOR protein sequence

```
E.colioptmized      MGFQRLSSLFFSTPSPPEDPDCPRAFVSEEDEVDGWLIIDLPDSYAAPPSPGAAPAPAGR
HumanDOR            -MFQRLSSLFFSTPSPPEDPDCPRAFVSEEDEVDGWLIIDLPDSYAAPPSPGAAPAPAGR
                      ************************************************************

E.colioptmized      PPPAPSLMDESWFVTPPACFTAEGPGLGPARLQSSPLEDLLIEHPSMSVYVTGSTIVLEP
HumanDOR            PPPAPSLMDESWFVTPPACFTAEGPGLGPARLQSSPLEDLLIEHPSMSVYVTGSTIVLEP
                    ************************************************************

E.colioptmized      GSPSPLPDAALPDGDLSEGELTPARREPRAARHAAPLPARAALLEKAGQVRRLQRARQRA
HumanDOR            GSPSPLPDAALPDGDLSEGELTPARREPRAARHAAPLPARAALLEKAGQVRRLQRARQRA
                    ************************************************************

E.colioptmized      ERHALSAKAVQRQNRARESRPRRSKNQSSFIYQPCQRQFNY 221
HumanDOR            ERHALSAKAVQRQNRARESRPRRSKNQSSFIYQPCQRQFNY 220
                    ******************************************
```

Figure 34 – Alignment sequence of Human DOR and DOR optimized to *E.coli*, used in this project.

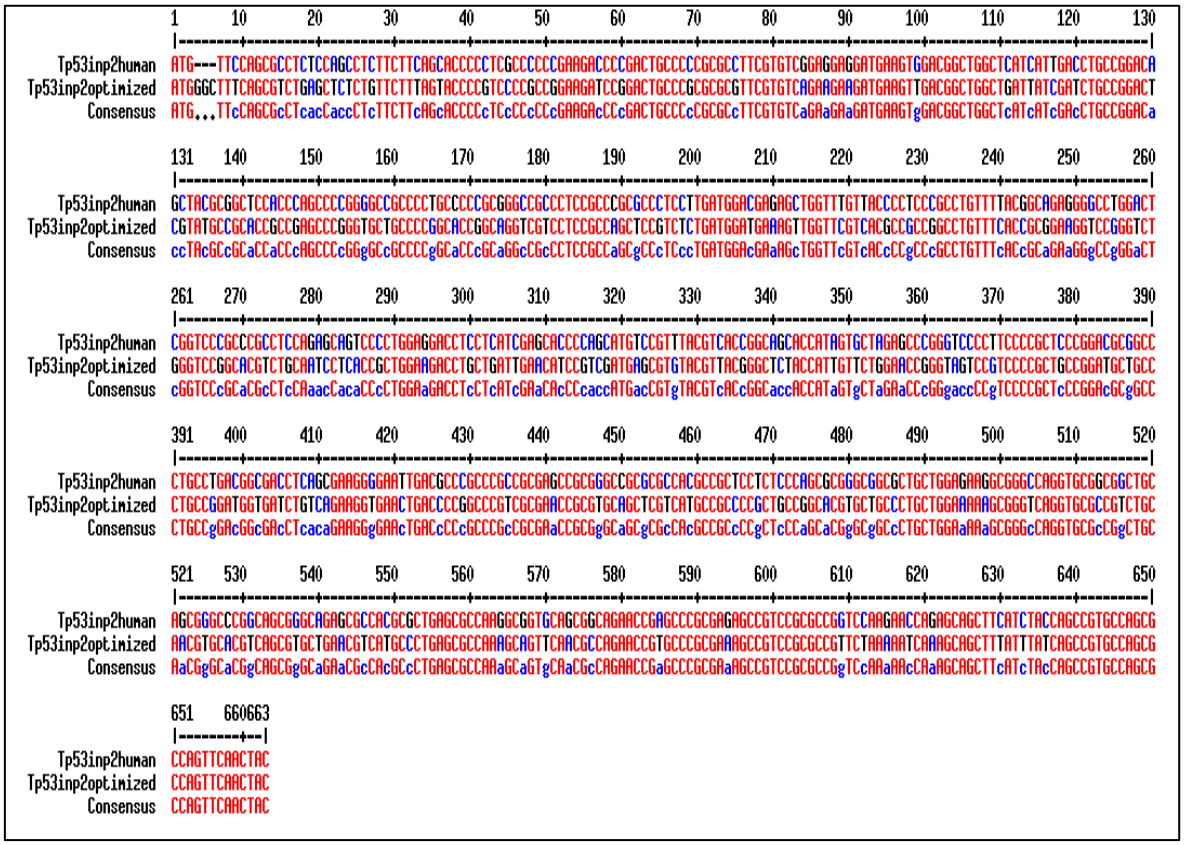


Figure 35 – Alignment of DNA sequence of Human DOR and DOR optimized to *E.coli*, used in this project performed with MultiAlin software.

III- Expression Vectors

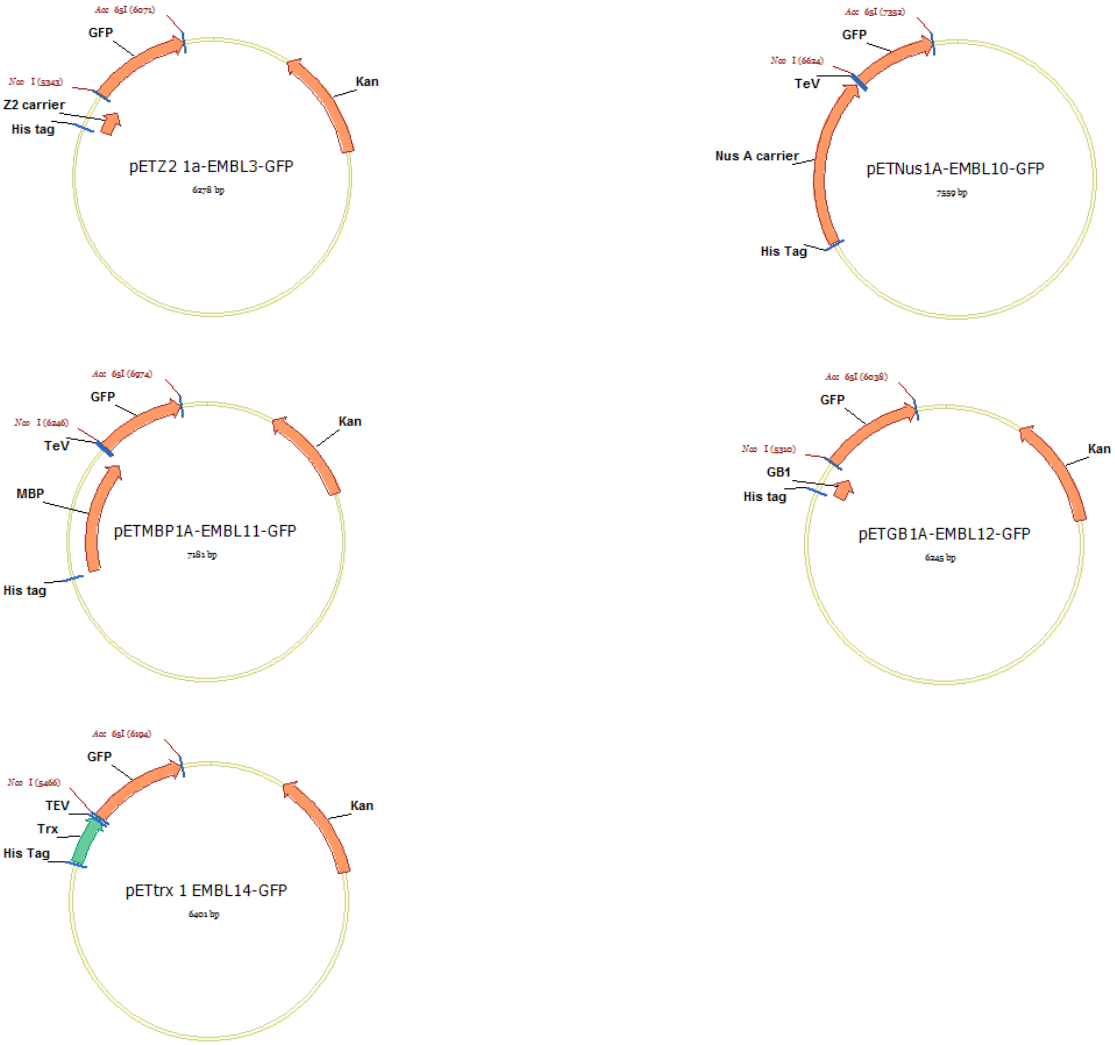


Figure 35- Map of expression vectors. All the vectors has the kanamycin resistance gene (Kan) for the selection of transformants in *E.coli*; T7 promotor, that allows the IPTG inducible; f1 origin; an N-terminal hexahistidine tag, that allows the purification of the recombinant fusion protein on a metal-chelating resin; a fusion partner, to improve the solubility of the protein and also the detection of the recombinant fusion protein with an ant-his antibody and a TEVp recognition site before the cloning site of the protein of interest to allow for fusion tag/protein removal after protein purification.

IV- Sequencing analysis

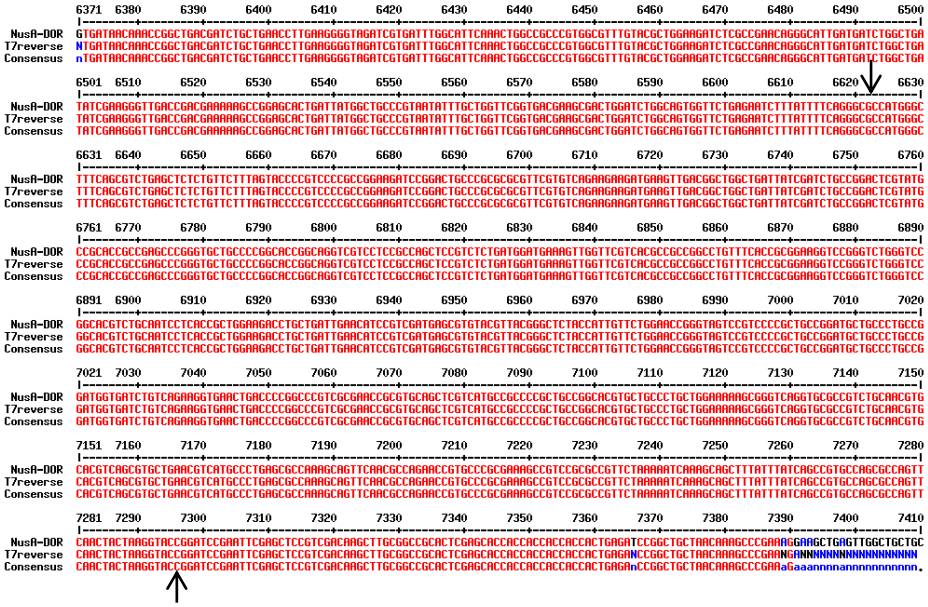


Figure 36- Alignment of DNA sequence of NusA-DOR. The positive clone was sequenced with T7 reverse primer, comparison with NusA-DOR expected was performed with MultiAlin software. The consensus sequence is represented in red. DOR sequence is indicated between the black arrow.

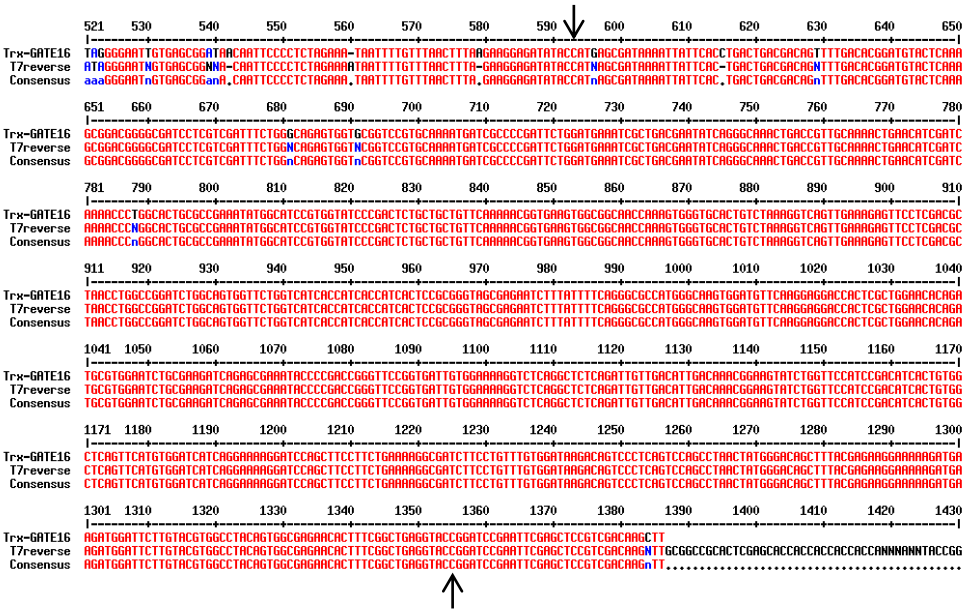


Figure 38- Alignment of DNA sequence of Trx-GATE16. One positive clone was sequenced with T7 reverse primer, comparison with the expected was performed with MultiAlin software. The consensus sequence is represented in red. Trx-GATE16 sequence is indicated between the black arrow.

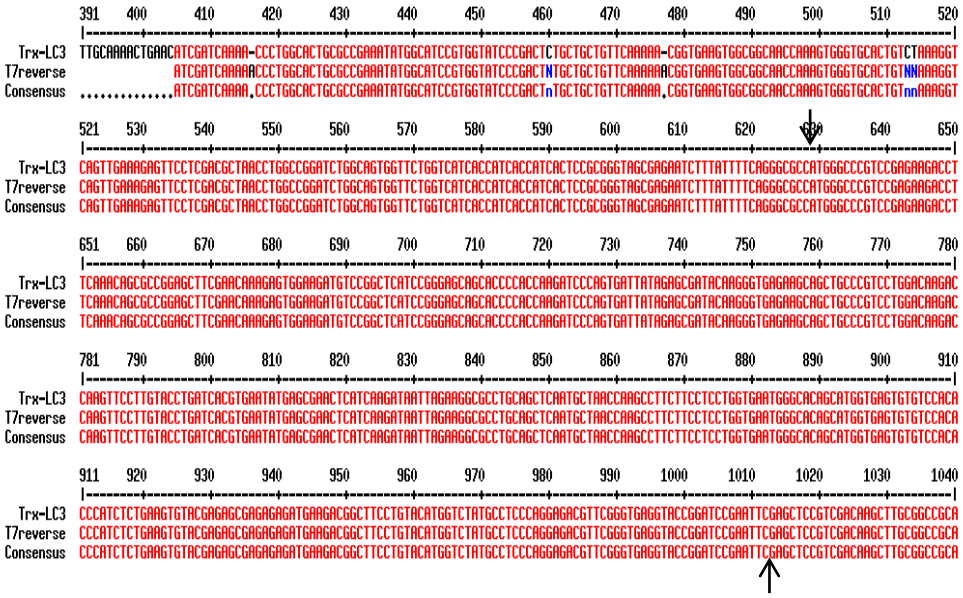


Figure 39- Alignment of DNA sequence of Trx-LC3. One positive clone was sequenced with T7 reverse primer, comparison with the expected was performed with MultiAlin software. The consensus sequence is represented in red. LC3 sequence is indicated between the black arrow.

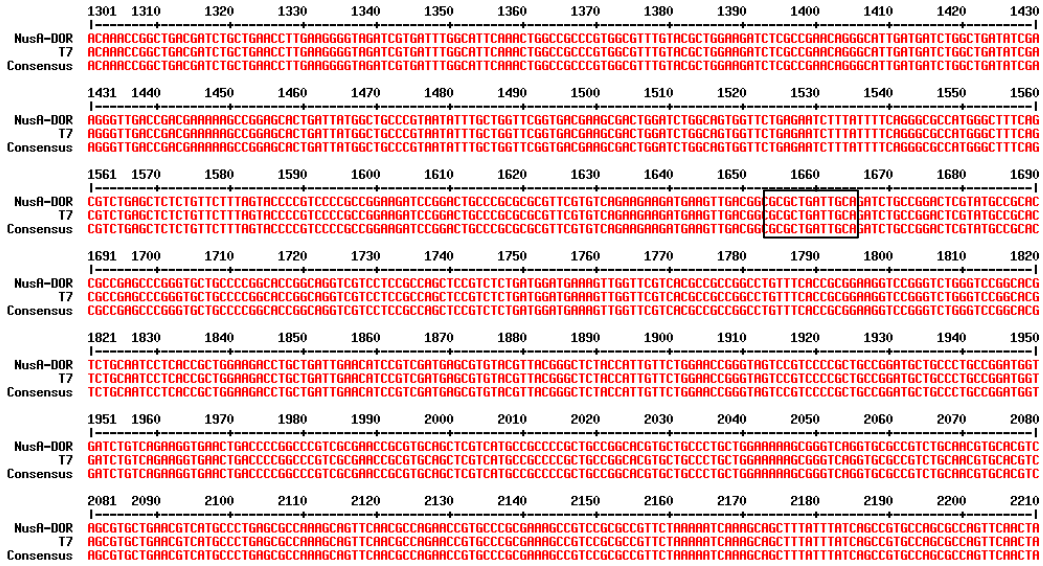


Figure 40- Alignment of DNA sequence of NusA-DOR W35AI38A. One positive clone was sequenced with T7 reverse primer, comparison with the expected was performed with MultiAlin software. The consensus sequence is represented in red. W35AI38A is delimited by a black rectangle.