

# **Charge pair interactions in transmembrane helices and turn propensity of the connecting sequence promote helical hairpin insertion**

Manuel Bañó-Polo <sup>a</sup>, Luis Martínez-Gil <sup>a^</sup>, Björn Wallner <sup>b</sup>, José L. Nieva <sup>c</sup>,  
Arne Elofsson <sup>d</sup> and Ismael Mingarro <sup>a\*</sup>

<sup>a</sup> Departament de Bioquímica i Biologia Molecular, Universitat de València.  
E-46100 Burjassot, Spain

<sup>b</sup> Department of Physics, Chemistry and Biology, Swedish e-Science Research  
Center, Linköping University, SE-581 83 Linköping, Sweden.

<sup>c</sup> Unidad de Biofísica (CSIC-UPV/ EHU) and Departamento de Bioquímica,  
Universidad del País Vasco (UPV/ EHU) E-48080 Bilbao, Spain.

<sup>d</sup> Department of Biochemistry and Biophysics, Science for Life Laboratory,  
Stockholm Bioinformatics Center, Center for Biomembrane Research, Swedish  
e-Science Research Center, Stockholm University.  
SE-106 91 Stockholm, Sweden.

Running title: Membrane integration of an  $\alpha$ -helical hairpin

\* Corresponding author: I. Mingarro.

Phone: Int+34-96-354 3796.

Fax: Int+34-96-354 4635.

E-mail: Ismael.Mingarro@uv.es

^ Present address: Department of Microbiology, Mount Sinai School of  
Medicine, New York, USA

## **Abstract**

$\alpha$ -helical hairpins, consisting of a pair of closely spaced transmembrane (TM) helices that are connected by a short interfacial turn, are the simplest structural motifs found in multi-spanning membrane proteins. In naturally occurring hairpins, the presence of polar residues is common and predicted to complicate membrane insertion. We postulate that the pre-packing process offsets any energetic cost of allocating polar and charged residues within the hydrophobic environment of biological membranes. Consistent with this idea, we provide here experimental evidence demonstrating that helical hairpin insertion into biological membranes can be driven by electrostatic interactions between closely separated, poorly hydrophobic sequences. Additionally, we observe that the integral hairpin can be stabilized by a short loop heavily populated by turn-promoting residues. We conclude that the combined effect of TM-TM electrostatic interactions and tight turns plays an important role in generating the functional architecture of membrane proteins, and propose that helical hairpin motifs can be acquired within the context of the Sec61 translocon at the early stages of membrane protein biosynthesis. Taken together, these data further underline the potential complexities involved in accurately predicting TM domains from primary structures.

## **KEYWORDS**

Electrostatic interactions; helical hairpin; membrane integration; salt bridge; translocon.

**Abbreviations**

Ct: C-terminus; ER: endoplasmic reticulum; Lep: leader peptidase; RMs: rough microsomal membranes; Nt: N-terminus; PV: poliovirus; SDS-PAGE: sodium dodecylsulfate polyacryamide-gel electrophoresis; TM: transmembrane.

## Introduction

Multi-spanning membrane proteins (those including two or more membrane spanning segments) are important for many biological functions. The basic structural unit of such membrane proteins is a hydrophobic  $\alpha$ -helix. In folded proteins these individual helix-forming sequences are engaged in a rich network of interactions with other helices. Whereas individual helices are formed in response to main-chain hydrogen bonding and the hydrophobic effect, other interactions must be responsible of side-to-side assembly. Such interactions might include hydrophobic packing, electrostatic effects, turns between helices, and binding to components placed in the aqueous environments that surround the membrane.

$\alpha$ -helical hairpins, consisting of a pair of closely spaced transmembrane (TM) helices that are connected by a short extramembrane or interfacial turn, are the simplest structural motifs found in multi-spanning membrane proteins <sup>1</sup>. This motif is thought to occur relatively frequently in integral membrane proteins and may serve as an important structural and/or functional element <sup>2</sup>.

The insertion of most helical plasma membrane proteins occurs co-translationally, whereby protein synthesis and integration into the membrane are coupled. For the integration of individual TM sequences into the membrane, it is expected that TM segments will preadopt a helical state <sup>3; 4</sup>, due to the significant free-energy penalty of embedding an exposed polypeptide backbone into the hydrophobic membrane core <sup>5</sup>. Similarly, the

formation of inter-helical hydrogen bonds facilitates the integration of polar residues present in adjacent TM regions. Then, there is no doubt that hydrogen bond interactions can play key roles in helical hairpin stabilization.

While many studies addressing the formation of helical hairpins in membranes have been carried out on model hydrophobic TM segments<sup>6; 7; 8; 9</sup>, naturally occurring helical hairpins are not always highly hydrophobic<sup>10</sup> and the role of helix-helix interactions and turn-propensities of the residues interconnecting the two helices in their folding and stability are poorly understood. We previously showed that poliovirus (PV) 2B, which is a small protein involved in PV virulence, is a double-spanning integral membrane protein, in which the two TM segments are interconnected by a short turn forming a putative helical hairpin<sup>11</sup>. As a first step towards understanding its biogenesis, we demonstrated that *in vitro* PV 2B integrates into the endoplasmic reticulum (ER) membrane through the translocon<sup>11</sup>.

Here, we present a detailed investigation on structural determinants underlying helical hairpin formation in the viral membrane protein 2B. Using an *in vivo*-like translation-glycosylation system of the naturally occurring helical hairpin from PV 2B we have determined the importance of helix-helix interactions for hairpin formation. In addition, we show that the hairpin structure is stabilized by the turn propensity of the amino acid residues in the short loop between the two TM helices. Our results suggest that integral helical hairpins may form in biological membranes driven by electrostatic interactions between marginally hydrophobic sequences and be additionally stabilized by short, tight connecting turns.



## Results

### Insertion of the viroporin 2B hairpin region into biological membranes

We have recently shown that *in vitro* PV 2B product inserts into the ER membrane as a double-spanning integral membrane protein with an N-/ C-terminal cytoplasmic orientation <sup>11</sup>. Such topology is attained upon insertion of a helical hairpin whose constituent TM helices are marginally hydrophobic (Fig. 1A). *In vitro* synthesis of several truncated protein versions indeed put forward that the two hydrophobic regions cooperate to insert into the ER-derived microsomal membranes. Here, we explore the structural grounds for such effect.

As in our previous study of membrane insertion of the PV 2B <sup>11</sup>, we used a well-characterized *in vitro* experimental system based on glycosylation <sup>12</sup> that accurately reports the integration of TM regions into microsomal membranes. Upon insertion, the oligosaccharyl transferase (OST) enzyme modifies the protein of interest. OST adds sugar molecules to a NX(S/ T) consensus sequence <sup>13</sup>, with X being any amino acid except proline <sup>14</sup>, after the protein emerges from the translocon channel. Glycosylation of a protein region synthesized *in vitro* in the presence of microsomal membranes therefore indicates the exposure of this region to the OST active site on the luminal side of the ER membrane. When assayed independently, the two hydrophobic regions of the PV 2B did not span the ER-derived membranes <sup>11</sup>, as expected according to the predicted apparent free energy of insertion (Fig. 1A). It has been shown previously that in some cases, a neighboring TM helix can promote membrane insertion of a marginally hydrophobic TM region <sup>15; 16;</sup>



<sup>17, 18</sup> and that there is a correlation between the polarity of a TM helix and its interaction area with the rest of the protein <sup>19</sup>. Therefore, we investigated the insertion of the full  $\alpha$ -helical hairpin region (residues 35-81) in this *in vitro* translation system.

In our first experimental set-up, the helical hairpin region was introduced into the “host” protein leader peptidase (Lep) (Fig. 1B), which contains two TM helices (H1 and H2) and a large lumenally exposed C-terminal domain (P2). In this first Lep construct, the 2B hairpin sequence (residues 35-81, panel A) was placed near the middle of the P2 domain (panel B) and was flanked by two engineered NXT acceptor sites for *N*-linked glycosylation (G1 and G2), with the G2 site located immediately downstream the hairpin region (see Materials and Methods). It has previously been demonstrated that efficient glycosylation occurs when the acceptor Asn is ~12-14 residues away from the membrane <sup>20, 21</sup>. If the hairpin is translocated across the membrane, both G1 and G2 sites will be modified by the lumenally oriented OST; if the helical hairpin is inserted into the membrane, only G1 will receive a glycan (Fig 1B). If one of the two hydrophobic regions is inserted, only G1 will be modified but in that case the large P2 domain will be non-translocated across the microsomal membrane. In this way, a single glycosylation suggest either correct hairpin integration (Fig. 1B, right) or the integration of only one hydrophobic region, whereas double glycosylation reports non-integration capability of the hairpin region (Fig. 1B, left). Single glycosylation of the molecule results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of Lep expressed in the

absence of microsomes, and around 5 kDa in the case of double glycosylation.

As shown in Fig. 1D, *in vitro* synthesis of this construct yielded single and double glycosylated forms of the protein (lanes 1 and 2). Proteinase K treatment of this translation reaction mixtures produced two main bands corresponding to P1-digested molecules (Fig. 1D, lane 3), which contained the second TM segment from Lep (H2) plus the hairpin of 2B and the P2 domain either singly-glycosylated (lower band) or doubly-glycosylated (upper band). It should be noted that the G2 acceptor site is located too close to the C-terminal region of the  $\alpha$ -helical hairpin (see Materials and Methods) to be efficiently glycosylated (Fig. 1B), since as previously demonstrated, for efficient glycosylation the acceptor site should be located away from the membrane interface<sup>20; 21</sup>. To facilitate the interpretation of these experimental results it is important to mention that short interconnecting turns like the one present in the 2B  $\alpha$ -helical hairpin are poor substrates for proteinase K digestion<sup>22</sup>. Then, the presence of a significant proportion of singly-glycosylated molecules is indicative of a hairpin insertion.

Next, we tested the 2B hairpin arranged according to its predicted topology<sup>11</sup> using a different set-up (Fig. 1C). In the Lep-derived construct (Lep') used to measure hairpin insertion propensity in this case, the 2B hairpin sequence replaced the Lep H2 domain (Fig. 1C). The glycosylation site (G2) located in the beginning of the P2 domain will be modified only if this Ct domain is translocated across the membrane, while an engineered G1' site, embedded in an extended Nt sequence of 24 amino acids is always glycosylated. As previously observed, when assayed in the Lep' construct the

first hydrophobic region inserted efficiently into the membrane (up to 60% of the molecules were doubly-glycosylated, see figure 4 in reference <sup>11</sup>). However, synthesizing the helical hairpin in this Lep' construct in the presence of microsomal membranes the doubly-glycosylated band diminished to 34% of the molecules (Fig. 1D, lane 5), indicating a significant achievement of hairpin insertion reported as singly-glycosylated molecules (see Fig. 1C for a scheme). Furthermore, proteinase K digestions of these samples (Fig. 1D, lanes 4-6) render protected forms derived from the doubly-glycosylated molecules. Overall, these data suggest that, in the context of Lep-derived model proteins, the two hydrophobic regions of the viroporin 2B can insert into the membrane when expressed as an  $\alpha$ -helical hairpin. They further emphasize that the preferred orientation is with the N-/ C-termini facing towards the cytosolic side (Fig. 1C).

### **Effects on helical hairpin formation of a single Asp residue in the second hydrophobic region**

A phylogenetic analysis of picornavirus 2B proteins has highlighted the presence of cationic residues in the first hydrophobic  $\alpha$ -helix, and in the case of several genus the existence of an aspartic acid residue within the second hydrophobic segment <sup>23</sup>.

In our attempts to identify possible helix-helix interactions that stabilize hairpin formation we focused first on the aspartic residue mentioned above since this appeared to be a likely candidate to promote electrostatic interactions between the TM helices. Because *N*-glycosylation acceptor sites

are absent from the PV 2B sequence, we added a C-terminal glycosylation tag that has been proven to be efficiently modified <sup>24</sup>, and synthesized the first 76 residues of the PV 2B (76-mer), which contained the native aspartic residue (Asp74) (Fig. 2A, see also Fig. 6 in <sup>11</sup>), to emphasize the role of this polar residue. In these truncated polypeptides glycosylation scores translocation of the Ct region across the ER membrane, while non-glycosylated protein bands correspond to insertion of this Ct region (see schemes in Fig. 2B).

To prove the role of Asp74 within these truncated protein forms, we replaced this residue by a lysine residue (76-mer(K) and 76-mer(K2)). As shown in Fig. 2B the glycosylation level for these mutants increased significantly compared to the wild-type truncated sequence (lanes 1 and 2), suggesting that the presence of the lysine residues complicates membrane insertion of the second hydrophobic region (see equilibrium schemes at the bottom of panel B). A similar effect was found when a second aspartic acid residue was included (76-mer (D2)) (Fig. 2C). The most likely explanation for the disruptive effect of this second aspartic acid residue (position 76) is that it lies on the side of the helix that faces the lipid (2 positions away from Asp74 or  $\sim 200^\circ$ ), imposing an unsurpassable penalty for membrane insertion of the second helix. Interestingly, replacement of Asp74 by glutamic residue rendered a glycosylation level similar to the native aspartic residue (Fig. 2C).

### **Position-specific effects on helical hairpin formation by Lys-Asp pairs**

We next focused on the positively charged residues present in the first hydrophobic region of the PV 2B sequence (residues 35-55, Fig. 1A). The

protein sequence in this region contains three lysine residues Lys39, Lys42 and Lys46. To find out which of these residues would be involved in electrostatic interactions with Asp74 we generated possible structural models using the Rosetta membrane protocol <sup>25</sup>. These models clearly predicted a preference for Lys46 and Lys42 to interact with Asp74. Among the five largest structural clusters obtained, as compared to Lys39, interactions of Lys42 and Lys46 with Asp74 are 53 and 75 times more frequent respectively (Figure 3).

In a series of experiments performed in hairpin truncations (76 residues long polypeptides, 76-mers) we separately replaced each lysine residues by an aspartic acid residue. As expected, K39D mutation exhibited a glycosylation level indistinguishable from the wild-type sequence (data not shown), indicating that it does not perturb hairpin insertion and confirming our computational predictions. On the contrary, as shown in Fig. 4A, replacement of Lys42 by a negatively charged aspartic acid residue (K42D mutant) substantially increased glycosylation (compare lanes 1 and 2 with lanes 5 and 6), indicating that Lys42 is involved in some electrostatic interactions between the two adjacent helices. In fact, these interactions can be partially restored as reported with the double mutant K42D/ D74K, which is only weakly glycosylated (Fig. 4A, lanes 3 and 4). Thus, hairpin capacity for inserting into the ER membrane was almost unaffected by swapping charges between helices, which underscores the involvement of these residues in inter-helical electrostatic interactions.

The same set of constructs was used to test Lys46 involvement in these electrostatic interactions (Fig. 4B). In this set of experiments a similar effect

was observed: the single mutant K46D was efficiently glycosylated (C-terminal tag translocated across the membrane, lanes 1 and 2), whereas double mutant K46D/ D74K was poorly glycosylated (lanes 3 and 4), indicating again that the interactions between the two hydrophobic regions were restored in this later construct.

To confirm that these interactions are relevant in the context of the 2B full-length protein, both Lys42 and Lys46 were separately mutated to aspartic acid and their membrane disposition experimentally determined. The synthesis of Ct-tagged 2B full-length protein (FL) in the presence of RMs resulted in glycosylation in <5% of the molecules (Fig. 5, lanes 5 and 6). Both K42D and K46D mutants consistently showed a significantly increased glycosylation level, suggesting that the more efficient hairpin integration observed in the native 2B protein might be promoted by electrostatic interactions between Lys-Asp pairs placed in the adjacent TM segments.

### **Influence of turn-residues on hairpin stabilization**

Next we focused on the turn between the two TM segments as an additional source of stability of the helical hairpin motif. To investigate the role of the residues interconnecting both 2B TM segments we performed an alanine-scanning mutagenesis approach with the full-length Ct-tagged 2B protein. Alanine was used because it carries a non-bulky and chemically inert side chain, and more importantly, because it has a very low turn-promoting effect<sup>26</sup>. Thus, we individually replaced residues 56 to 60 with alanine. The results of these experiments are shown in Fig. 6A. Interestingly, a gradual effect was

observed. Hence, Y58A and D60A mutants displayed glycosylation levels similar to native full-length PV 2B (compare lanes 4 and 6 in Fig. 6A with lane 6 in Fig. 5), suggesting that individual replacement of Tyr58 or Asp60 by the poor turn-promoting alanine residue does not preclude hairpin integration. However, Glu59 and specially Arg56 and Asn57 single amino acid replacements rendered a clear increase in the glycosylation efficiency of these constructs.

Further, we collected all turns in helical hairpins from predicted integral membrane proteins of viral origin (see Materials and Methods). A total of 3,654 such turns were found. Turn length and amino acid frequencies were calculated for these. In Fig. 7, we show that 5-residue turn helical hairpins, like the one observed in PV 2B sequence, are among the most frequent in viral membrane, only surpassed by turns of length two.

Finally, we sought to test if there is any combined effect on the helical hairpin integration between the charge pairs found in the hydrophobic regions and the turn residues. To this end we designed double mutations including the more sensitive residue Asn57 (N57A), which appears to play an important role in hairpin formation, and the lysines and aspartic acid residues involved in the electrostatic interactions found above. As shown in Fig. 6B, the glycosylation efficiencies for these double mutants were indeed higher compared to the single mutants analysed above (Fig. 5 and 6A), and demonstrated an additive effect of the different protein domains in terms of helical hairpin stabilization.

## Discussion

Helical hairpins appear to be extremely common in multi-spanning integral membrane proteins, and is thus an essential structural motif to our understanding of membrane protein folding and topology. In these structural motifs, the insertion of a polypeptide segment containing polar amino acids can be facilitated by the insertion of a closely spaced more hydrophobic region<sup>27</sup>. Then, the membrane-buried polar groups are saturated with internal hydrogen bonds and salt bridges, and non-polar side chains are preferentially exposed on the bilayer-facing surfaces, balancing the physical constraints imposed by the non-polar core of the membrane bilayer and resulting in thermodynamic stability of the multi-spanning membrane protein.

In the hydrophobic environment of the membrane core, van der Waals helix-helix packing, hydrogen bonding and ionic interactions are the governing contributors to multi-spanning membrane protein assembly. These interactions can be modulated by the sequence context and the lipid bilayer properties<sup>28</sup>. Although hydrophobic interactions between TM segments are the more abundant, hydrogen bonding or salt-bridge formation between membrane-spanning charged residues are essential to drive membrane protein folding, while at the same time, reduces the unfavourable energetics of inserting charged residues into the membrane<sup>29</sup>. Hence, interactions between polar residues in adjacent TM segments have been shown to favour membrane insertion<sup>30; 31; 32</sup>. PV 2B comprises two TM segments harbouring conserved charged residues<sup>23</sup>. Individual integration of these sequences in a model protein construct (Lep) proved to be inefficient<sup>11</sup>. However, the



glycosylation pattern found in this system when both regions were expressed in-block (including their native turn region) suggests that the helical hairpin components cooperate to facilitate their insertion into the membrane (Fig. 1). We show here that a central component for the stabilization of the helical hairpin is the charged residues found in the TM regions. Structure predictions placed lysine residues 42 and 46 in the first TM segment and Asp74 in the second TM segment at optimum positions to stabilize the helical hairpin conformation by inter-helical hydrogen bonding (Fig. 3). Experimentally, we found that the negatively charged residue (Asp74) appears to be relevant for hairpin integration (Fig. 2). Furthermore, Lys42 and Lys46 have an impact on hairpin integration of both truncated (Fig. 4) and full-length proteins (Fig. 5). In all these cases, substitutions of the lysine residues by aspartic acid residues have a significant impact on the glycosylation level of wild-type downstream sequences, whereas similar mutations on a background where the Asp74 was replaced by a lysine residue restored the glycosylation levels of native sequences. It is important to mention that all these mutations have a higher impact on the glycosylation level of the truncated molecules than on the full-length constructs. Interestingly, it has been previously demonstrated that extramembraneous positively charged residues placed near the cytosolic end of TM segments in membrane proteins promote membrane insertion of precedent hydrophobic helices<sup>18; 33</sup>. Inspection of PV 2B sequence reveals the presence of three positively charged flanking residues adjacent to the helical hairpin C-terminal end (see Fig. 5, top), which could then contribute to the

higher integration efficiency for the full-length protein compared to truncated molecules where these flanking positively charged residues are absent.

One of the aims of our study was to search for patterns to aid the prediction of helical hairpins from viral membrane protein sequences. The two-residue turns are most common in our viral membrane protein collection, but five-residue turns (like the one found in PV 2B) are second most abundant, Figure 7.

The first two residues in the turn (Arg56 and Asn57) showed to be the more sensitive to alanine replacement (Fig. 6A). Interestingly, asparagine and arginine residues together with prolines have been found to behave as the strongest turn-promoting residues when placed in the middle of a 40-residue polyleucine segment <sup>26, 34</sup>. In addition, multiple sequence analysis of picornavirus 2B protein sequences highlighted Arg56 and Asn57 as fully conserved residues <sup>35</sup>, suggesting a relevant role in protein function. We also find that the contribution of turn residues can be combined with the interaction between charged residues appropriately located to stabilize the helical hairpin conformation by inter-helical hydrogen bonding, since the higher levels of glycosylation have been obtained when perturbed both turn and charged residues at the same time (Fig. 6B). Similarly, structural perturbations introduced by extracellular loop mutations in a cystic fibrosis TM conductance regulator (CFTR) appeared to affect hairpin conformation to a lesser extent in the context of more stable TM helix-helix interactions <sup>36</sup>, highlighting the interplay between helix-helix interactions and turn contribution in modulating hairpin formation.

Our observation that Lys-Asp pairs together with turn-promoting residues stabilize a TM helical hairpin conformation supports helical hairpin formation at the initial stages of membrane protein biogenesis, i.e. within the translocon. In this regard, two TM segments have been found to accumulate within or adjacent to the translocon before final integration into the lipid bilayer<sup>37; 38; 39; 40; 41</sup>. Furthermore, molecular dynamics simulations of the translocon in a fully solvated lipid bilayer showed that two helices can coexist within the translocon (J. Gumbart, personal communication). An interesting possibility arising from these observations is that the translocon may facilitate membrane protein integration by allowing efficient polar interactions between consecutive TM segments harbouring charged residues, and thereby stabilizing them in the nascent polypeptide. In this particular case, the lysine residues located in the first TM segment may participate in keeping this domain in the translocon until the second TM segment reaches this location. If this were the case, helical hairpin formation inside the translocon might facilitate partitioning into the lipid bilayer by shielding the charged amino acids that would otherwise constrain membrane integration efficiency. Although further studies are needed to unravel the details of this hypothesis, our results suggest that very specific helix-helix interactions can be formed within the context of the translocon and that such interactions together with turn constraints can have a dramatic effect into the poorly understood mechanism of multi-spanning membrane protein folding.

## Materials and methods

**Enzymes and chemicals.** All enzymes as well as plasmid pGEM1 and rabbit reticulocyte lysate were from Promega (Madison, WI). SP6 RNA polymerase and ER rough microsomes from dog pancreas were from tRNA Probes (College Station, TX). [<sup>35</sup>S]Met and <sup>14</sup>C-methylated markers were from GE Healthcare. Restriction enzymes and Endoglycosidase H were from Roche Molecular Biochemicals. The DNA plasmid, RNA clean up and PCR purification kits were from Qiagen (Hilden, Germany). The PCR mutagenesis kit QuikChange was from Stratagen (La Jolla, CA). All the oligonucleotides were from Thermo (Ulm, Germany).

**DNA manipulations.** The helical hairpin from PV 2B was introduced into the modified Lep sequence from the pGEM1 plasmid<sup>12; 42</sup> between the *Spe I* and *Kpn I* sites by PCR amplified PV 2B sequence containing residues 32-81 with primers containing appropriate restriction sites. After PCR amplification, the PCR products were purified, digested, and ligated to the corresponding Lep vector digested with the same enzymes. In this construct, the acceptor asparagine is located seven residues downstream the end of the 2B hairpin sequence (...WL<sup>81</sup>QVPGQQNAT..., where L<sup>81</sup> is the Ct residue of the hairpin and the glycosylation site is underlined). The Lep' construct carried one glycosylation acceptor site in positions 3-5 of an extended sequence of 24 residues previously described<sup>43</sup>. The same strategy was used to insert PV 2B hairpin sequence by PCR into the Lep' construct.

Full-length 2B sequence (without a stop codon) fused to the P2 domain of the *Escherichia coli* leader peptidase in a pGEM1 plasmid was performed as described previously<sup>11, 44</sup>. The K42D, K46D and D74K site-directed mutagenesis were done using the QuikChange mutagenesis kit from Stratagene (La Jolla, CA) following the manufacturer's protocol. All DNA manipulations were confirmed by sequencing of plasmid DNAs. For the 76-residue long truncated molecules (76-mers) replacement of Asp74 by lysine, alanine, glycine and glutamic acid as well as mutants 76-mer(K2) and 76-mer(D2) by using reverse primers with the appropriated sequences in the PCR amplifications (see below).

**Expression *in vitro*.** Lep construct with 2B helical hairpin was transcribed and translated as previously reported<sup>42, 45</sup>.

Full-length 2B DNA was amplified from 2B/ P2 plasmid using a reverse primer with an optimized Ct glycosylation tag<sup>24</sup> and a stop codon at the end of the 2B sequence (2B-derived expressions). Truncated 2B DNA was amplified by PCR from 2B/ P2 plasmid using reverse primers with an optimized Ct glycosylation tag and a stop codon at the end annealing at specific position to obtain the desired polypeptide length (76 residues). The transcription of the DNA derived from pGEM1 plasmid was done as previously<sup>46</sup>. Briefly, the transcription mixture was incubated at 37°C for 2 h. The mRNAs were purified using a Qiagen RNeasy clean up kit and verified on a 1% agarose gel.

*In vitro* translation of *in vitro* transcribed mRNA was done in presence of reticulocyte lysate, [<sup>35</sup>S]Met and dog pancreas microsomes as described

previously<sup>46; 47</sup>. After polypeptide synthesis membranes were collected by ultra-centrifugation and analyzed by SDS-PAGE, finally gels were visualised on a Fuji FLA3000 phosphorimager using the ImageGauge software.

For the proteinase K protection assay, the translation mixture was supplemented with 1  $\mu$ L of 50 mM CaCl<sub>2</sub> and 1  $\mu$ L of proteinase K (4 mg/ mL), then digested for 20 min on ice. Adding 1 mM PMSF before SDS-PAGE analysis stopped the reaction.

**Computational modelling.** Rosetta membrane<sup>25</sup> was used to generate 10,181 conformations of the helix hairpin from structural fragments and the topology prediction by Topcons (topcons ref: PMID: 19429891). All conformations were clustered on the helix-helix region using 2Å cut-off. The models belonging to the five largest clusters (20% of all clustered models) were subjected to all-atom refinement using the *relax* application in the Rosetta package. The all-atom conformational space around each model was explored by performing 250 independent all-atom refinement runs for each model before the interaction between Asp74, and Lys39, Lys42 and Lys46 - were analysed. Asp and Lys were considered to interact if the polar groups of their side-chains were within 3.5Å.

### **Sequence analysis of Viral TM proteins.**

Potential viral proteins containing membrane hairpins were detected in the following way. All Viral sequences were downloaded from the refseq<sup>48</sup>. The topcons-single algorithm (ref:PMID 21493661) was then used to predict the topology of all 81932 viral proteins. 13848 of these proteins were predicted to

be TM and 3654 of these had two predicted TM-helices. These proteins were used to analyse the length and composition of the loops.

## Figure legends

**Figure 1.** In-block insertion of PV 2B  $\alpha$ -helical hairpin (residues 35-81) into microsomal membranes using a model protein. (A) Sequences for the predicted TM helices (boxes) and connecting turn. The predicted free energy of insertion ( $\Delta G$ ) apparent values (estimated using the  $\Delta G$  prediction server at <http://dgpred.cbr.su.se><sup>12, 49</sup>) is indicated at the bottom of each TM segment. In this algorithm, positive values are indicative of translocation across the membrane (i.e., absence of stable insertion). Lysine and aspartic acid residues are shown in blue and red, respectively. Schematic representations of the Lep (B) and Lep' (C) constructs used to report insertion into ER membrane of 2B helical hairpin. In the Lep construct (B), the helical hairpin under study is inserted into the P2 domain and flanked by two artificial glycosylation acceptor sites (G1 and G2). Recognition of the hairpin by the translocon machinery as an integrating domain locates G1 and G2 in the luminal side of the ER membrane, but the short distance to the membrane prevents G2 glycosylation. The Lep chimera will be doubly-glycosylated when the hairpin is translocated into the lumen of the microsomes, and insertion of only one hydrophobic region would render singly-glycosylated proteinase K-digestible molecules. (C) In this Lep' construct, PV 2B hairpin replaces the H2 domain from Lep. The glycosylation acceptor site (G2) will be modified only if translocated across the membrane, while the G1' site, embedded in an extended N-terminal sequence is always glycosylated. (D) *In vitro* translation in the presence (+) or absence (-) of RMs and proteinase K (PK). Bands of non-glycosylated protein are indicated by a white dot; singly



and doubly glycosylated proteins are indicated by one and two black dots, respectively. The protected singly- and doubly-glycosylated H2/ helical hairpin/ P2 fragments are indicated by one and two black triangles, respectively. Lane Mw contains radioactive molecular mass markers as indicate.

**Figure 2.** Insertion of PV 2B-derived truncated proteins. (A) Structural organization of the 76-mer truncated constructs. The *N*-glycosylation site added as a C-terminal reporter tag (rectangles) is highlighted by a Y-shaped symbol. Mutant residues are shown in bold and acidic and basic residues are highlighted in red and blue, respectively. (B) *In vitro* translations were performed in the presence (+) and in the absence (–) of RMs as indicated. Non-glycosylated and singly-glycosylated proteins are indicated by an empty and black dot, respectively. A scheme for the equilibrium between non-glycosylated and glycosylated forms for the wild type (left) and the lysine-containing constructs (right) is shown at the bottom, where the presence of lysine residues at the Ct is highlighted in blue and the thickness of the arrowhead lines refers to the prevalence of each form. (C) *In vitro* glycosylation of 76-mer truncated proteins. The level of glycosylation is quantified from SDS-PAGE gels by measuring the fraction of glycosylated ( $f_g$ ) versus glycosylated plus non-glycosylated ( $f_{ng}$ ) molecules,  $p = f_g / (f_g + f_{ng})$ . Data correspond to averages of at least three independent experiments and error bars show standard deviations. Lane Mw contains radioactive molecular mass markers as indicate.

**Figure 3.** Observed interactions of Asp74 among low energy models generated by the Rosetta-Membrane protocol (see Methods). Clearly, it is energetically favorable for the hairpin to hide the polar groups of the Asp and Lys residues. In particular the polar group of Asp74 can be shielded from the hydrophobic lipid environment by interactions with one of the three Lys in the first helix. The data analysis shows that interactions between Asp74 and either Lys42 or Lys46 are found much more frequently than interactions between Asp74 and Lys39, which are almost never observed. The inset shows one of the models where interactions are seen between Asp74 and both Lys42 and Lys46.

**Figure 4.** Effects of Lys42-Asp74 (A) and Lys46-Asp74 (B) charge pairs on the integration of 2B truncated molecules into the ER membrane. Schematic representations of the Ct-tagged 2B 76-mer truncated constructs are shown on top of each panel. Bottom: *in vitro* translation in the presence (+) or absence (–) of RMs of 76-residue truncated molecules carrying wild-type (Asp74, lanes 1, 2, 5 and 6) or mutant D74K (lanes 3 and 4) sequences. Glycosylation levels are quantified as in in Fig. 2, and non-glycosylated and singly-glycosylated proteins are indicated by an empty and black dot, respectively. Data correspond to averages of at least three independent experiments. Lane Mw contains radioactive molecular mass markers as indicate.

**Figure 5.** Effects of mutations in the first hydrophobic segment on the integration of full-length 2B into the ER membrane. Structural organizations of full-length PV 2B constructs are shown on top, with the flanking basic residues Arg82, Lys83 and Lys84 highlighted in blue. Bottom: *in vitro*

translation in the presence (+) or absence (–) of RMs of full-length molecules carrying wild-type (lanes 5, and 6) or mutant K42D (lanes 1 and 2), K46D (lanes 3 and 4) sequences. The level of glycosylation is quantified as in Fig. 2, and non-glycosylated and singly-glycosylated proteins are indicated by an empty and black dot, respectively. Data correspond to averages of at least three independent experiments. Lane Mw contains radioactive molecular mass markers as indicate.

**Fig. 6.** Effects of mutations in the turn region. (A) Ala-scanning mutagenesis of the residues located in the turn region. *In vitro* translations were performed in the presence (+) or in the absence of RMs as indicated. Turn region amino acid sequence is shown on top. (B) Additive effect of mutations in the turn and in the charged residues involved in pairwise interactions. *In vitro* translations were performed in the presence (+) or in the absence of RMs as indicated. Data correspond to averages of at least three independent experiments.

**Fig. 7.** Distribution of the turn-lengths among viral membrane proteins. Inset: schematic diagram of a helical hairpin. Turn residues are defined by using the topcons-single prediction algorithm<sup>50</sup>.

**Acknowledgments**

This work was supported by grants BFU2009-08401 (to I.M.) and BIO2011-29792 (to J.L.N.) from the Spanish Ministry of Science and Innovation (MICINN, ERDF supported from the EU), ACOMP/ 2012/ 226 from the Generalitat Valenciana (to I.M.) and GIU06/ 42 from the Basque Government (to J.L.N.). A.E. was funded by the Swedish National Research center, Vinnova and EU FP7 through the EDICT grant (FP7-HEALTH-F4-2007-201924). M.B.-P. was recipient of an FPU fellowship from the Spanish Ministry of Education. B.W. was funded by the Swedish e-Science Research Center.

## References

1. Wehbi, H., Gasmi-Seabrook, G., Choi, M. Y. & Deber, C. M. (2008). Positional dependence of non-native polar mutations on folding of CFTR helical hairpins. *Biochimica et biophysica acta* **1778**, 79-87.
2. Janovick, J. A. & Conn, P. M. (2010). Salt bridge integrates GPCR activation with protein trafficking. *Proc Natl Acad Sci U S A* **107**, 4454-8.
3. Mingarro, I., Nilsson, I., Whitley, P. & von Heijne, G. (2000). Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biol* **1**, 3.
4. Gumbart, J., Chipot, C. & Schulten, K. (2011). Free energy of nascent-chain folding in the translocon. *Journal of the American Chemical Society* **133**, 7602-7.
5. White, S. H. & von Heijne, G. (2008). How translocons select transmembrane helices. *Annu Rev Biophys* **37**, 23-42.
6. Hermansson, M., Monne, M. & von Heijne, G. (2001). Formation of helical hairpins during membrane protein integration into the endoplasmic reticulum membrane. Role of the N and C-terminal flanking regions. *Journal of molecular biology* **313**, 1171-9.
7. Hermansson, M. & von Heijne, G. (2003). Inter-helical hydrogen bond formation during membrane protein integration into the ER membrane. *Journal of molecular biology* **334**, 803-9.
8. Saaf, A., Hermansson, M. & von Heijne, G. (2000). Formation of cytoplasmic turns between two closely spaced transmembrane helices during membrane protein integration into the ER membrane. *Journal of molecular biology* **301**, 191-7.
9. Johnson, R. M., Heslop, C. L. & Deber, C. M. (2004). Hydrophobic helical hairpins: design and packing interactions in membrane environments. *Biochemistry* **43**, 14361-9.
10. Baeza-Delgado, C., Marti-Renom, M. A. & Mingarro, I. (2012). Structure-based statistical analysis of transmembrane helices. *European biophysics journal : EBJ*.
11. Martinez-Gil, L., Bano-Polo, M., Redondo, N., Sanchez-Martinez, S., Nieva, J. L., Carrasco, L. & Mingarro, I. (2011). Membrane integration of poliovirus 2B viroporin. *Journal of virology* **85**, 11315-24.
12. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. & von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-81.
13. Silberstein, S. & Gilmore, R. (1996). Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *Faseb J* **10**, 849-58.
14. Shakin-Eshleman, S. H., Spitalnik, S. L. & Kasturi, L. (1996). The amino acid at the X position of an Asn-X-ser sequon is an important

- determinant of N-linked core-glycosylation efficiency. *J Biol Chem* **271**, 6363-6366.
15. Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L., Pettersson, A., Johansson, S., von Heijne, G. & Nilsson, I. (2009). Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J Mol Biol* **387**, 1153-64.
  16. Hedin, L. E., Ojemalm, K., Bernsel, A., Hennerdal, A., Illergard, K., Enquist, K., Kauko, A., Cristobal, S., von Heijne, G., Lerch-Bader, M., Nilsson, I. & Elofsson, A. (2010). Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* **396**, 221-9.
  17. Tamborero, S., Vilar, M., Martinez-Gil, L., Johnson, A. E. & Mingarro, I. (2011). Membrane Insertion and Topology of the Translocating Chain-Associating Membrane Protein (TRAM). *J Mol Biol* **406**, 571-82.
  18. Ojemalm, K., Halling, K. K., Nilsson, I. & von Heijne, G. (2012). Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix. *Molecular cell* **45**, 529-40.
  19. Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. & Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci U S A* **105**, 7177-81.
  20. Nilsson, I. & von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J Biol Chem* **268**, 5798-5801.
  21. Orzaez, M., Salgado, J., Gimenez-Giner, A., Perez-Paya, E. & Mingarro, I. (2004). Influence of proline residues in transmembrane helix packing. *J Mol Biol* **335**, 631-40.
  22. Sauri, A., Tamborero, S., Martinez-Gil, L., Johnson, A. E. & Mingarro, I. (2009). Viral membrane protein topology is dictated by multiple determinants in its sequence. *J Mol Biol* **387**, 113-28.
  23. de Jong, A. S., de Mattia, F., Van Dommelen, M. M., Lanke, K., Melchers, W. J., Willems, P. H. & van Kuppeveld, F. J. (2008). Functional analysis of picornavirus 2B proteins: effects on calcium homeostasis and intracellular protein trafficking. *J Virol* **82**, 3782-90.
  24. Bano-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M. A. & Mingarro, I. (2011). N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci* **20**, 179-86.
  25. Barth, P., Wallner, B. & Baker, D. (2009). Prediction of membrane protein structures with complex topologies using limited constraints. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1409-14.
  26. Monne, M., Nilsson, I., Elofsson, A. & von Heijne, G. (1999). Turns in transmembrane helices: determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. *J Mol Biol* **293**, 807-14.

27. Engelman, D. M. & Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: The helical hairpin hypothesis. *Cell* **23**, 411-422.
28. Cymer, F., Veerappan, A. & Schneider, D. (2012). Transmembrane helix-helix interactions are modulated by the sequence context and by lipid bilayer properties. *Biochimica et biophysica acta* **1818**, 963-73.
29. Bano-Polo, M., Baeza-Delgado, C., Orzaez, M., Marti-Renom, M. A., Abad, C. & Mingarro, I. (2012). Polar/ Ionizable residues in transmembrane segments: effects on helix-helix packing. *PloS one* **7**, e44263.
30. Buck, T. M., Wagner, J., Grund, S. & Skach, W. R. (2007). A novel tripartite motif involved in aquaporin topogenesis, monomer folding and tetramerization. *Nature structural & molecular biology* **14**, 762-9.
31. Meindl-Beinker, N. M., Lundin, C., Nilsson, I., White, S. H. & von Heijne, G. (2006). Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep* **7**, 1111-6.
32. Zhang, L., Sato, Y., Hessa, T., von Heijne, G., Lee, J. K., Kodama, I., Sakaguchi, M. & Uozumi, N. (2007). Contribution of hydrophobic and electrostatic interactions to the membrane integration of the Shaker K<sup>+</sup> channel voltage sensor domain. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8263-8.
33. Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I. & von Heijne, G. (2008). Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **105**, 4127-32.
34. Monne, M., Hermansson, M. & von Heijne, G. (1999). A turn propensity scale for transmembrane helices. *J Mol Biol* **288**, 141-5.
35. Nieva, J. L., Agirre, A., Nir, S. & Carrasco, L. (2003). Mechanisms of membrane permeabilization by picornavirus 2B viroporin. *FEBS Lett* **552**, 68-73.
36. Wehbi, H., Rath, A., Glibowicka, M. & Deber, C. M. (2007). Role of the extracellular loop in the folding of a CFTR transmembrane helical hairpin. *Biochemistry* **46**, 7099-106.
37. Pitonzo, D., Yang, Z., Matsumura, Y., Johnson, A. E. & Skach, W. R. (2009). Sequence-Specific Retention and Regulated Integration of a Nascent Membrane Protein by the ER Sec61 Translocon. *Mol Biol Cell* **20**, 685-698.
38. Sauri, A., McCormick, P. J., Johnson, A. E. & Mingarro, I. (2007). Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. *J Mol Biol* **366**, 366-74.
39. Sauri, A., Saksena, S., Salgado, J., Johnson, A. E. & Mingarro, I. (2005). Double-spanning Plant Viral Movement Protein Integration into the Endoplasmic Reticulum Membrane Is Signal Recognition Particle-dependent, Translocon-mediated, and Concerted. *J Biol Chem* **280**, 25907-25912.

40. Cross, B. C. & High, S. (2009). Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *J Cell Sci* **122**, 1768-77.
41. Meacock, S. L., Lecomte, F. J., Crawshaw, S. G. & High, S. (2002). Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol Biol Cell* **13**, 4114-29.
42. Martinez-Gil, L., Perez-Gil, J. & Mingarro, I. (2008). The surfactant peptide KL4 sequence is inserted with a transmembrane orientation into the endoplasmic reticulum membrane. *Biophys J* **95**, L36-8.
43. Lundin, C., Kim, H., Nilsson, I., White, S. H. & von Heijne, G. (2008). Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for Nin/ Cout and Nout/ Cin transmembrane helices. *Proceedings of the National Academy of Sciences* **105**, 15702-15707.
44. Navarro, J. A., Genoves, A., Climent, J., Sauri, A., Martinez-Gil, L., Mingarro, I. & Pallas, V. (2006). RNA-binding properties and membrane insertion of Melon necrotic spot virus (MNSV) double gene block movement proteins. *Virology* **356**, 57-67.
45. Martinez-Gil, L., Johnson, A. E. & Mingarro, I. (2010). Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. *J Virol* **84**, 5520-7.
46. Vilar, M., Sauri, A., Monne, M., Marcos, J. F., von Heijne, G., Perez-Paya, E. & Mingarro, I. (2002). Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. *J. Biol. Chem.* **277**, 23447-23452.
47. Garcia-Saez, A. J., Mingarro, I., Perez-Paya, E. & Salgado, J. (2004). Membrane-insertion fragments of Bcl-xL, Bax, and Bid. *Biochemistry* **43**, 10930-43.
48. Pruitt, K. D., Tatusova, T., Klimke, W. & Maglott, D. R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic acids research* **37**, D32-6.
49. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. & von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-30.
50. Hennerdal, A. & Elofsson, A. (2011). Rapid membrane protein topology prediction. *Bioinformatics* **27**, 1322-3.



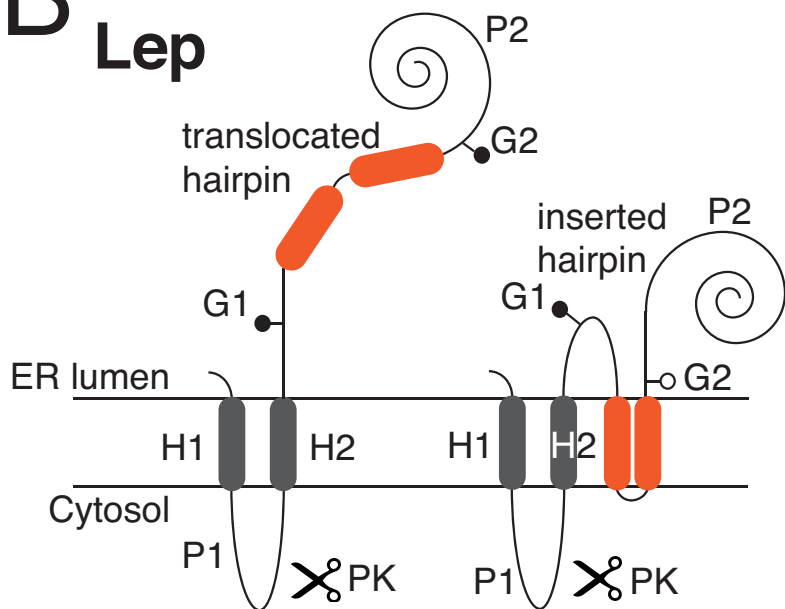
Figure 1

**A**



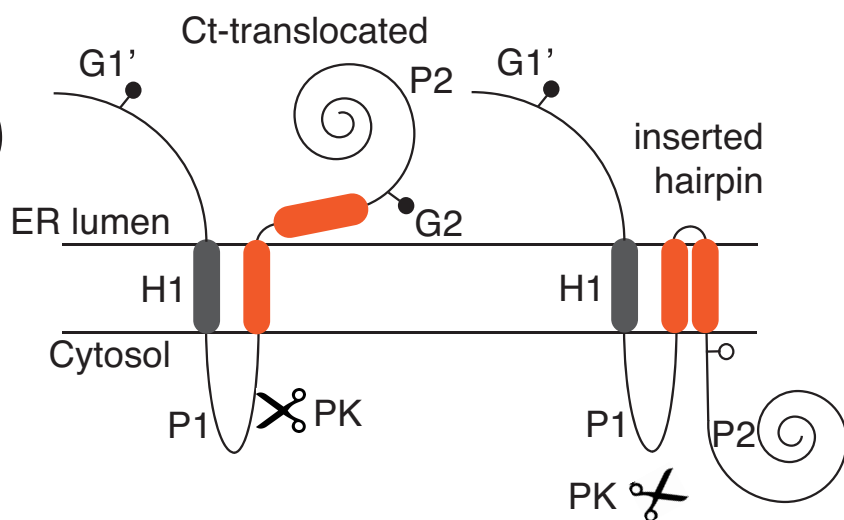
**B**

**Lep**



**C**

**Lep'**



**D**

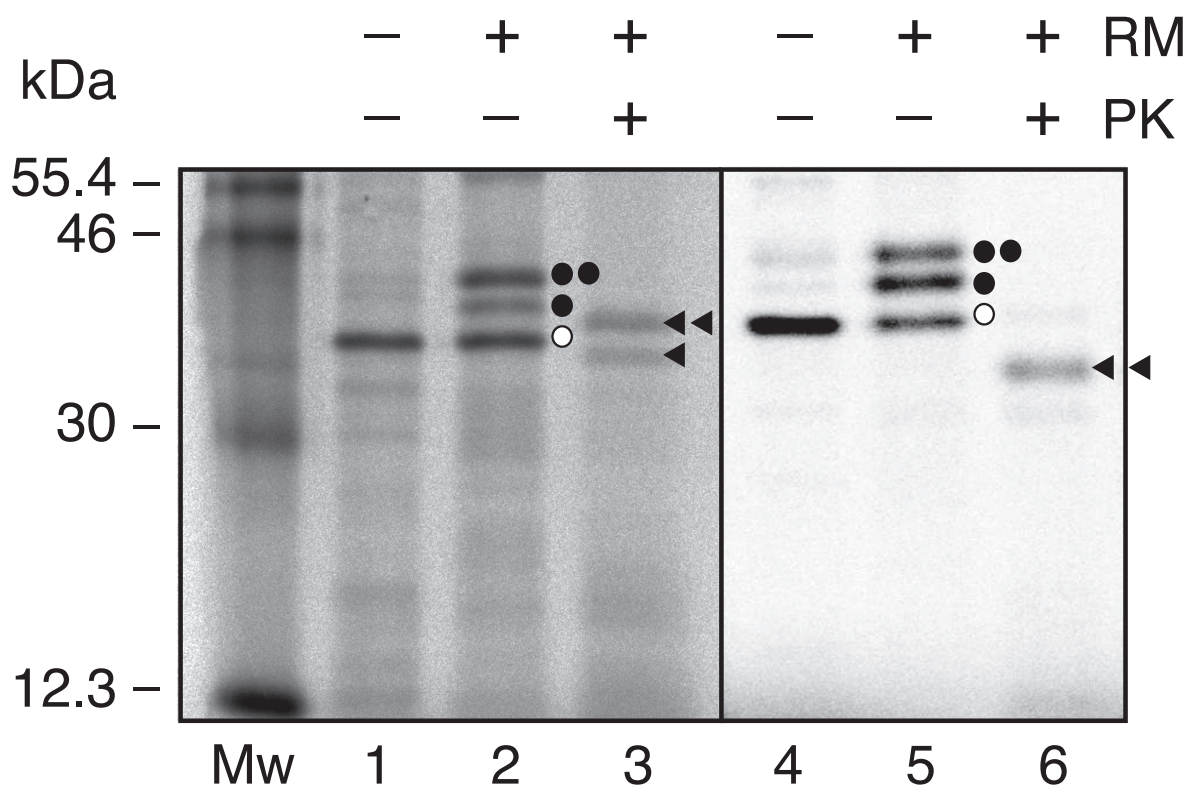


Figure 2

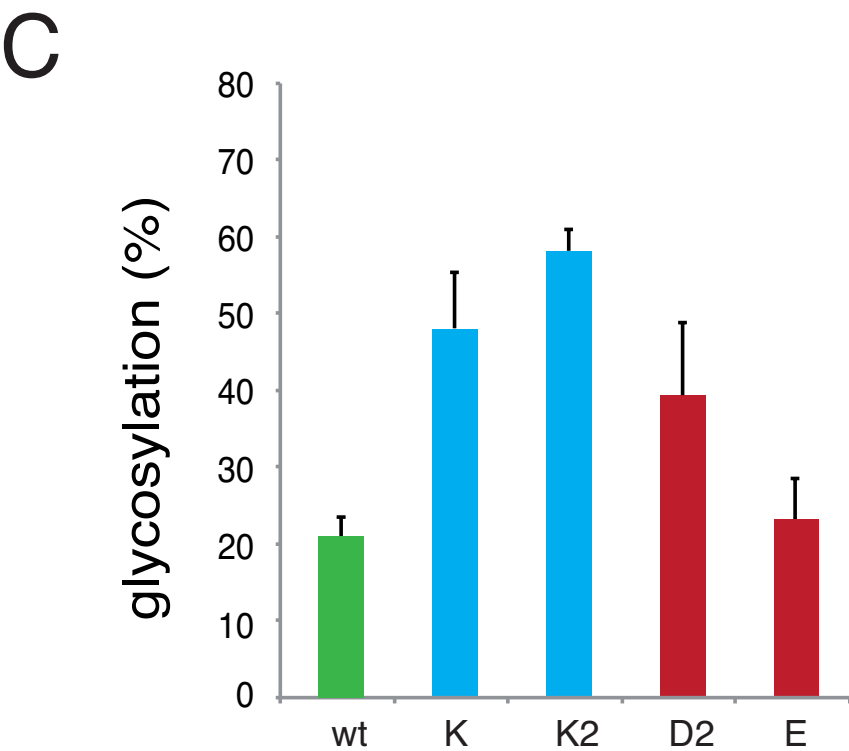
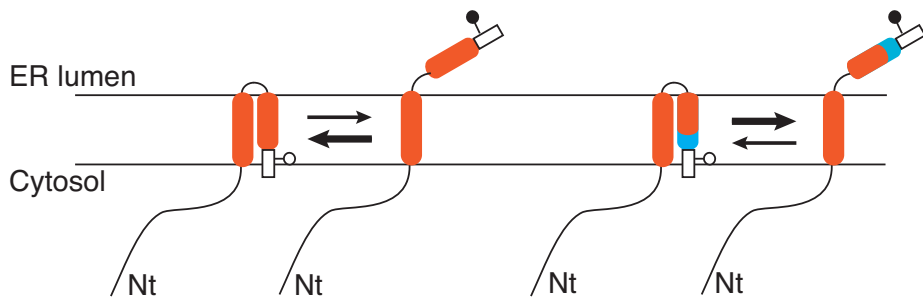
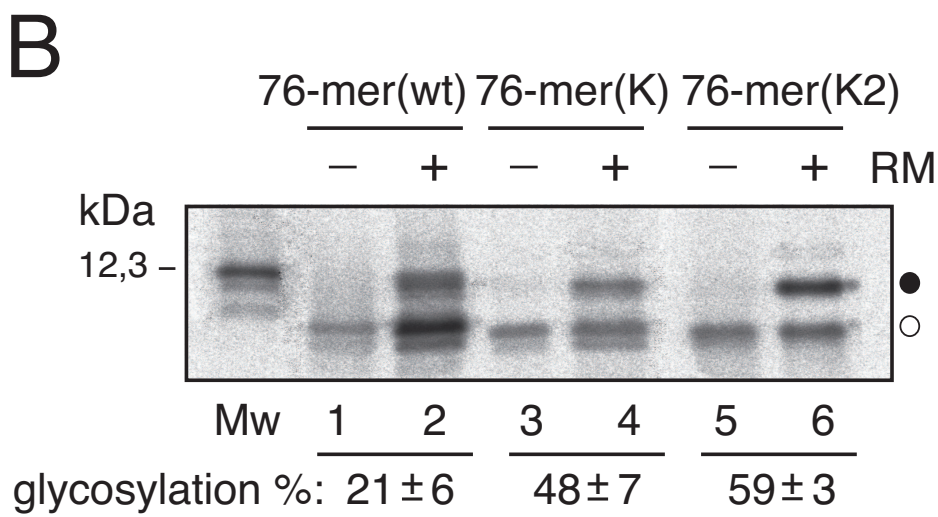
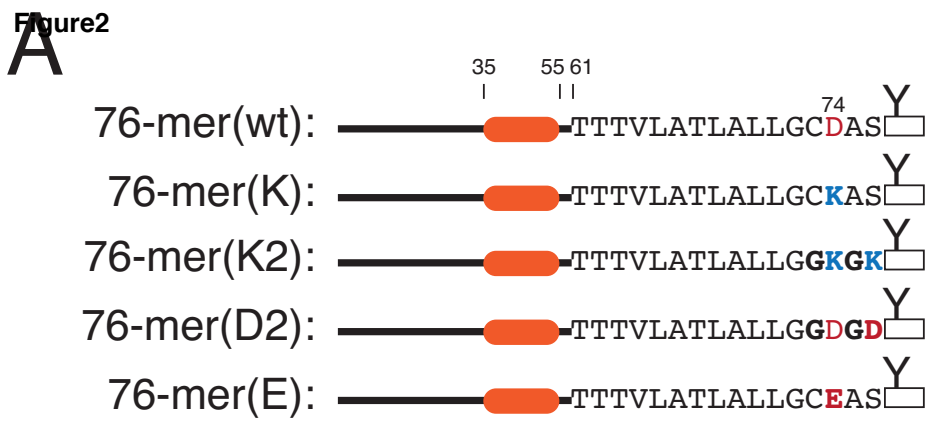
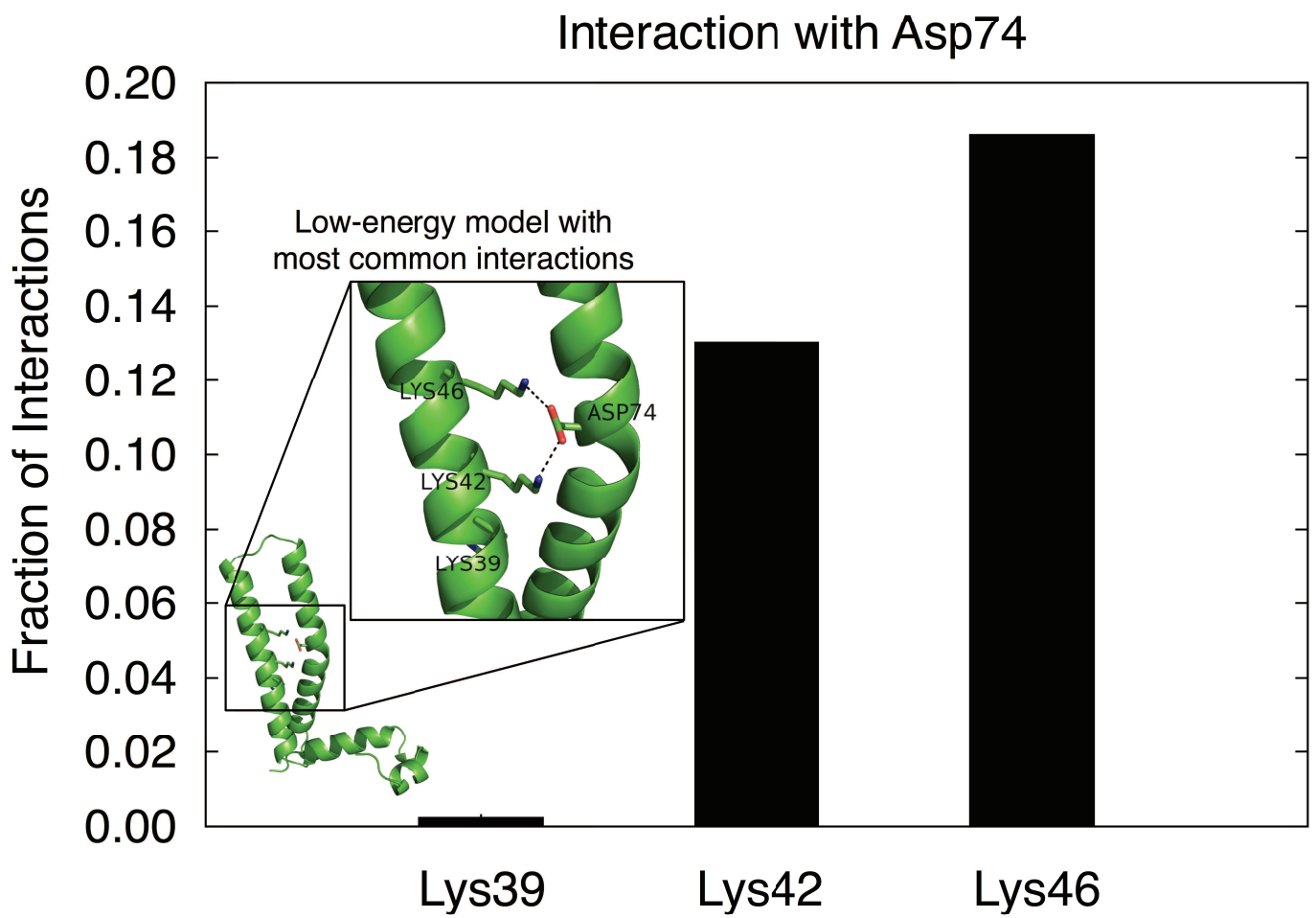
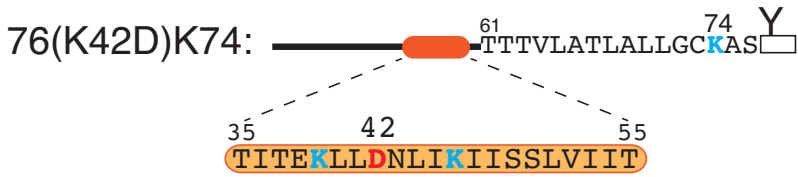
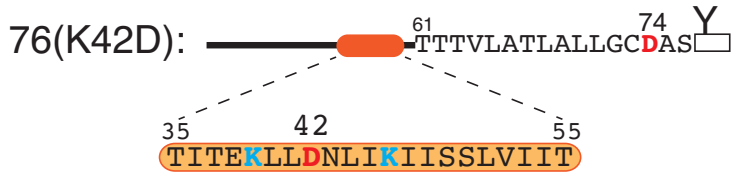
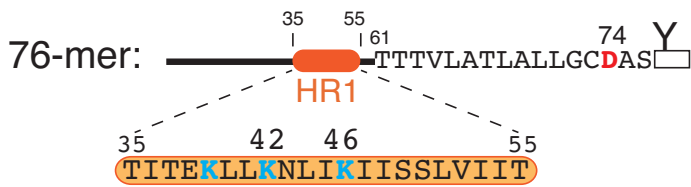


Figure3



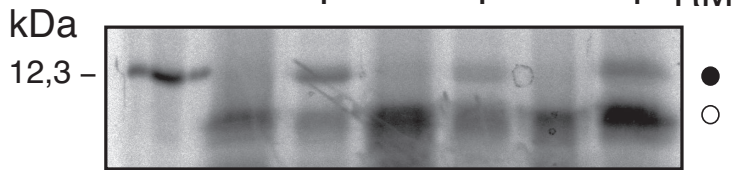
**A** Figure 4



76-mer(K42D)

D74      D74K      76-mer

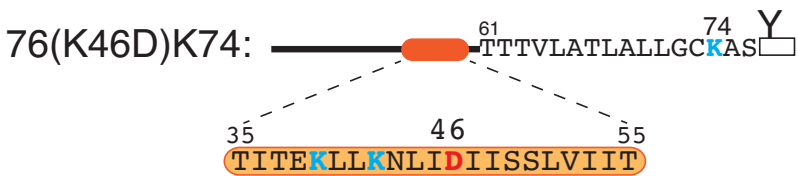
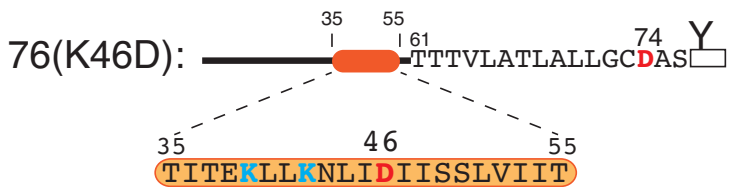
-    +    -    +    -    +    RM



Mw    1    2    3    4    5    6

glycosylation %:    47 ± 4    20 ± 3    16 ± 5

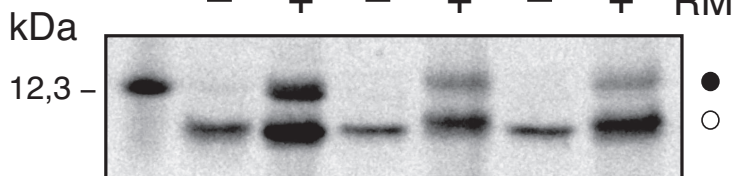
**B**



76-mer(K46D)

D74      D74K      76-mer

-    +    -    +    -    +    RM



Mw    1    2    3    4    5    6

glycosylation %:    42 ± 3    19 ± 6    16 ± 5

Figure 5

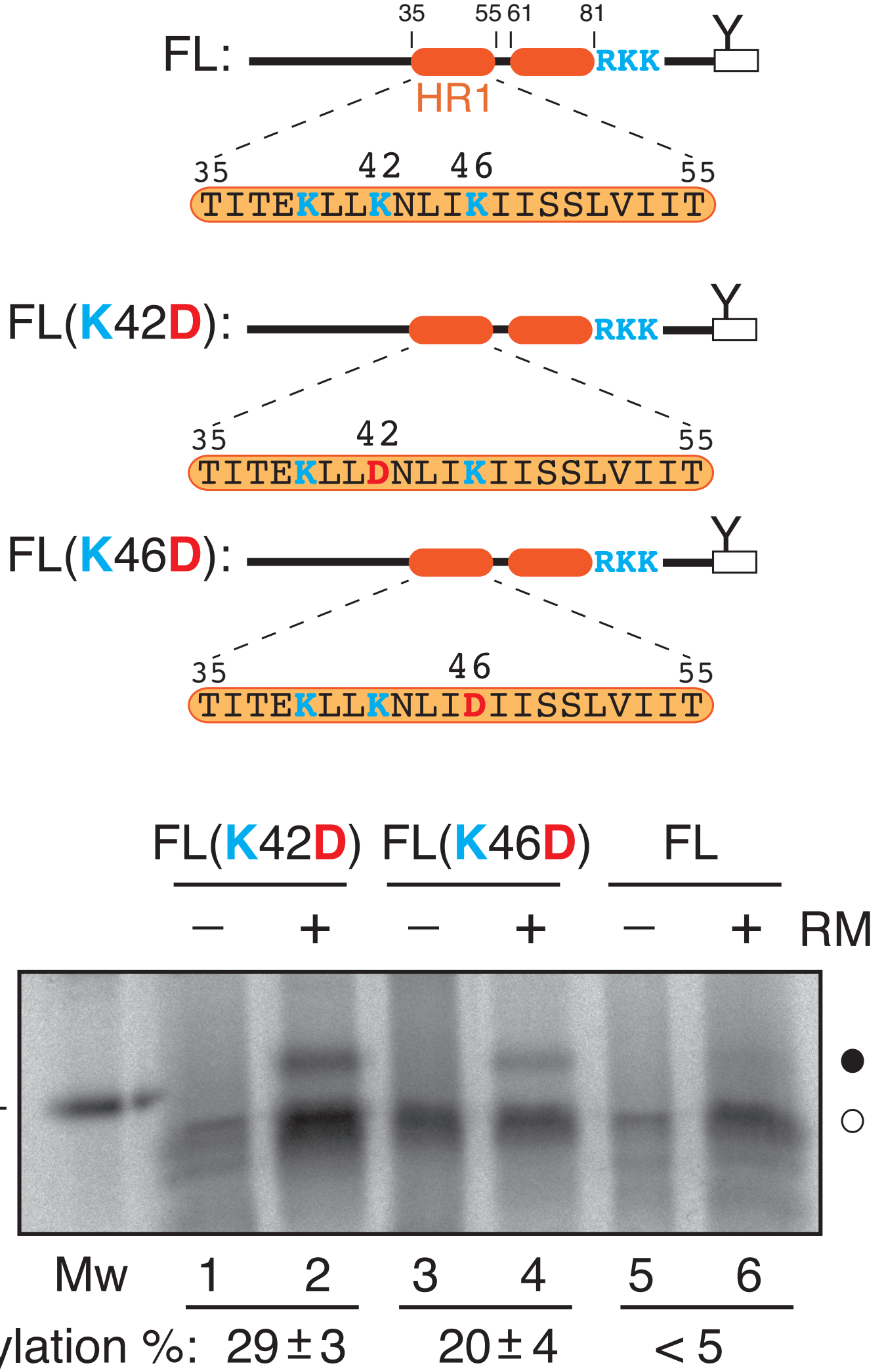


Figure 6

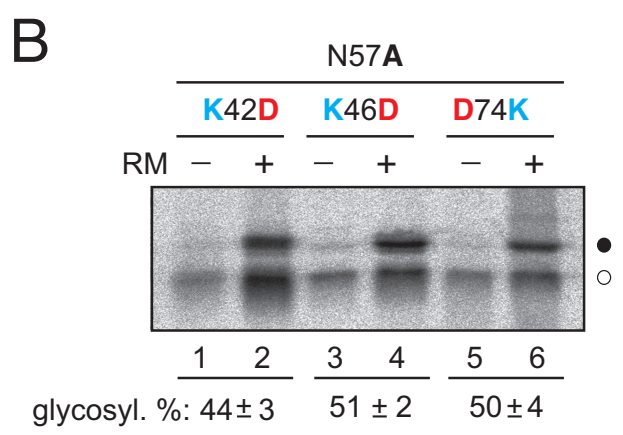
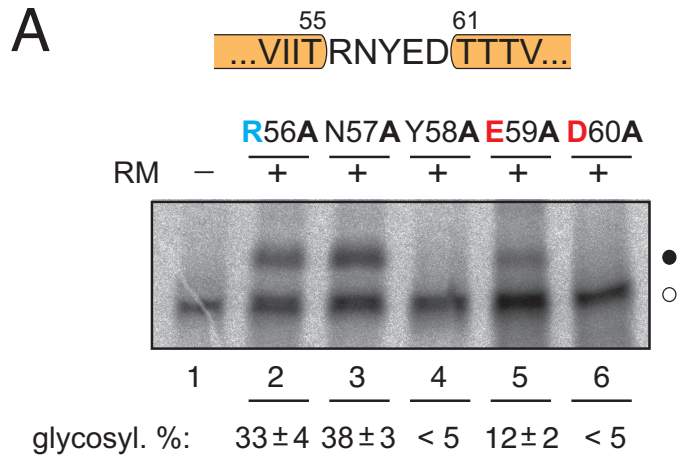


Figure 7

