NA transmembrane domain
Amphiphilic drift to accommodate two functions

Johan Nordholm

Academic dissertation for the Degree of Doctor of Philosophy in Biochemistry at Stockholm University to be publicly defended on Monday 5 June 2017 at 10.00 in Magnelisalen, Kemiska övningslaboratoriet, Svante Arrhenius väg 16 B.

Abstract
Neuraminidase (NA) is one of two major antigens on the surface of influenza A viruses. It is comprised of a single N-terminal transmembrane domain (TMD), a stalk domain, and a C-terminal enzymatic head domain that cleaves sialic acid, most notably to release new particles from the host cell surface. NA is only enzymatically active as a homo-tetramer. However, it is not known which properties facilitate the oligomerization of NA during assembly. Our results show that, apart from anchoring the protein to the membrane, the NA TMD also contributes to the assembly process by keeping the stalk in a tetrameric conformation. The ability of the TMD to oligomerize is shown to be dependent on its amphiphilic characteristics that was largely conserved across the nine NA subtypes (N1-N9). Over time the NA TMDs in human H1N1 viruses were found to have become more amphipathic, which correlated with stronger oligomerization. An old H1N1 virus with a more recent N1 TMD had impaired growth, but readily acquired compensatory mutations in the TMD to restore growth, by reverting the TMD oligomerization strength back to that of the old TMD, demonstrating a biological role of the TMD in folding and assembly. NA and the other viral proteins are spatially and temporally coordinated to achieve optimal viral production. By using a co-transfection analysis, the high AU-content in the NA and HA ER-targeting sequence coding regions (for NA TMD as well as the HA signal sequence) were found to inhibit their expression. The inhibition was alleviated by the early expressed influenza RNA-binding protein NS1, which promoted translation and showed enriched foci at the endoplasmic reticulum (ER). NS1, which expresses early during infection, is therefore likely the regulator of NA and HA to prevent premature expression. These results show that the NA TMD is under substantial selection pressure at both the nucleotide and amino acid level to accommodate its roles in ER-targeting, protein folding, and post-transcriptional regulation.

Keywords: influenza, IAV, neuraminidase, NA, transmembrane domain, TMD, secretory protein, ER-targeting sequence, ER-targeting sequence coding region, protein regulation, NS1, GALLEX.

Stockholm 2017
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-142051


Department of Biochemistry and Biophysics
Stockholm University, 106 91 Stockholm
NA transmembrane domain
Amphiphilic drift to accommodate two functions

Johan Nordholm
NA transmembrane domain

Amphiphilic drift to accommodate two functions

Johan Nordholm
List of publications

I. Assembly of subtype 1 influenza neuraminidase is driven by both the transmembrane and head domains.
   D da Silva*, J Nordholm*, U Madjo, A Pfeiffer, R Daniels

II. Polar residues and their positional context dictate the transmembrane domain interactions of influenza A neuraminidases.
    J Nordholm*, D da Silva*, J Damjanovic, D Dou, R Daniels
    Journal of Biological Chemistry 288 (15), 10652-10660, 2013

III. The influenza virus neuraminidase protein transmembrane and head domains have coevolved.
     D da Silva, J Nordholm, D Dou, H Wang, JS Rossman, R Daniels
     Journal of Virology 89 (2), 1094-1104, 2015

IV. Translational regulation of viral secretory proteins by the 5’ coding regions and a viral RNA-binding protein.
    J Nordholm, J Petitou, H Östbye, D da Silva, D Dou, H Wang, R Daniels
    Journal of Cell Biology, 2017

* Both authors contributed equally to this work
Additional publications

I. Luminescent lanthanide complexes with analyte-triggered antenna formation.
   E Pershagen, J Nordholm, KE Borbas
   Journal of the American Chemical Society 134 (24), 9832-9835, 2012

II. Type II transmembrane domain hydrophobicity dictates the cotranslational dependence for inversion.
    D Dou, DV da Silva, J Nordholm, H Wang, R Daniels
    Molecular Biology of the Cell 25 (21), 3363-3374, 2014

III. Identification of residues that affect oligomerization and/or enzymatic activity of influenza virus H5N1 neuraminidase proteins.
     M Dai, H Guo, JCFM Dortmans, J Dekkers, J Nordholm, R Daniels, FJ van Kuppeveld, E de Vries, CA de Haan
     Journal of Virology 90 (20), 9457-9470, 2016

IV. Assembly co-cooperativity between the influenza 1 NA stalk and 2 transmembrane domain defines the insertion deletion boundary.
    R Revol, H Östbye, D da Silva, J Nordholm, R Daniels
    Manuscript
Abstract

Neuraminidase (NA) is one of two major antigens on the surface of influenza A viruses (IAV). It is comprised of a single N-terminal transmembrane domain (TMD), a stalk domain, and a C-terminal enzymatic head domain that cleaves sialic acid, most notably to release new particles from the host cell surface. NA is only enzymatically active as a homo-tetramer. However, it is not known which properties facilitate the oligomerization of NA during assembly. Our results show that, apart from anchoring the protein to the membrane, the NA TMD also contributes to the assembly process by keeping the stalk in a tetrameric conformation. The ability of the TMD to oligomerize is shown to be dependent on its amphiphilic characteristics that was largely conserved across the nine NA subtypes (N1-N9). Over time the NA TMDs in human H1N1 viruses were found to have become more amphiphilic, which correlated with stronger oligomerization. An old H1N1 virus with a more recent N1 TMD had impaired growth, but readily acquired compensatory mutations in the TMD to restore growth, by reverting the TMD oligomerization strength back to that of the old TMD, demonstrating a biological role of the TMD in folding and assembly. NA and the other viral proteins are spatially and temporally coordinated to achieve optimal viral production. By using a co-transfection analysis, the high AU-content in the NA and HA ER-targeting sequence coding regions (for NA TMD as well as the HA signal sequence) were found to inhibit their expression. The inhibition was alleviated by the early expressed influenza RNA-binding protein NS1, which promoted translation and showed enriched foci at the endoplasmic reticulum (ER). NS1, which expresses early during infection, is therefore likely the regulator of NA and HA to prevent premature expression. These results show that the NA TMD is under substantial selection pressure at both the nucleotide and amino acid level to accommodate its roles in ER-targeting, protein folding, and post-transcriptional regulation.
Contents

Introduction .................................................................12
Secretory protein and why we have them ..................................12
Birth, life and death of secretory proteins ..................................13
  Birth ..............................................................................13
  Life ..............................................................................15
  Oligomerization ...........................................................15
  Death .............................................................................17
Secretory protein regulation .....................................................17
The influenza virus ...............................................................20
Synthesis and regulation of influenza secretory proteins ...............21
  NS1, a major regulator ....................................................23
The influenza virus, a tool to study secretory proteins .................24
Results ...............................................................................25
Conclusions and future perspectives ........................................27
Sammanfattning på svenska ................................................29
Acknowledgments ...............................................................30
References .........................................................................31
Abbreviations

IAV – Influenza A virus
HA – Hemagglutinin
NA – Neuraminidase
TMD – Transmembrane domain
ER – Endoplasmic reticulum
NS1 – Non-structural protein 1
Introduction

Secretory protein and why we have them

Cells are surrounded by a lipid membrane that acts as a boundary between the cell and its environment. Eukaryotic cells have additional intracellular lipid membranes to separate the cell into smaller compartments. All proteins are synthesized in the cell’s interior, but many are targeted to the membrane (secretory proteins), after which they either travel across the membrane (secreted proteins) or reside within the membrane (membrane proteins) to exert their function. Approximately four out of ten of human genes are predicted to encode for secretory proteins and two-thirds of those to be membrane proteins\(^1^,\)\(^2^\), which comprise the majority of current pharmaceutical drug targets\(^3^\).

The plasma membrane in human cells is more than just a lipid bilayer, as it consists of approximately equal weight of lipids and proteins\(^4^\). The large amount of membrane proteins exert their functions in different ways, such as selective import and export of molecules, and signaling in and out of the cell. While most membrane proteins have one transmembrane domain (TMD)\(^5^\), membrane proteins can differ in how many TMDs they have, and which topology the proteins have in the membrane (Figure 1). With few mitochondrial outer membrane exceptions\(^5^\), TMDs in human membrane proteins form α-helices, which in multi-spanning membrane proteins often fold into separate domains that play functional roles (e.g. receptors, transporters and channels)\(^6^\). In contrast, the α-helices of single transmembrane proteins are more restricted in their repertoire, and are therefore often thought of as mere membrane anchors.

Viruses can be categorized based on the absence or presence of a lipid membrane surrounding the virus particle. The lipid membrane of enveloped viruses often originates from the host plasma membrane during viral budding\(^7^\), and is the site of contact for the virus with host cells. As such, enveloped viruses also encode their own secretory proteins to facilitate host cell entry and in some instances release. Because of their positioning on the viral surface, these viral proteins are also the main antigens presented to the immune system.
Birth, life and death of secretory proteins

Birth

Cellular genes encoding for secretory proteins are transcribed in the nucleus, and are co-transcriptionally modified to obtain a 5' cap²⁴ (7-methylguanosine linked with a 5'-5' triphosphate bridge) and a 3' polyadenylated tail²⁵ (Figure 2). While many viruses transcribe their genes in the cytoplasm and utilize internal ribosomal entry (IRES) sites instead of a 5' cap to initiate translation, some viral transcripts resembles host mRNA, such as those from the influenza virus. Following transcription, the mature viral and host mRNAs are actively exported through the nuclear pore to the cytoplasm by nuclear export complexes. Secretory proteins contain an N-terminal ER-targeting sequence¹⁰,¹¹ which is usually 15-30 residues and have a stretch of hydrophobic residues¹². When an mRNA in the cytosol is translated, the N-terminal targeting sequence is the first to exit the ribosomal tunnel and gains exposure to the cytosol. It is then bound by the signal recognition particle (SRP), a complex of one RNA and six proteins¹³. SRP has low affinity to the ribosome and scans emerging polypeptides for the presence of signal sequences. The binding of signal sequences to a large hydrophobic groove in SRP causes a stronger association...
to the ribosome, which results in GTP binding to SRP, and the arrest of elongation in the ribosome. Successive docking to the SRP receptor (SR), which is located close to the translocon in the ER membrane displaces the ribosome and signal sequence from SRP, thereby resuming elongation and the emerging polypeptide is channeled through the translocon. The arrest of elongation has been shown to be necessary for efficient targeting due to SRP receptors in limited number.\textsuperscript{14}

![Diagram](image)

Figure 2. Biosynthesis of a secretory protein. The host or viral gene is transcribed in the nucleus to an mRNA, which is actively exported to the cytoplasm. Translational initiation occurs on cytoplasmic ribosomes. Specific for secretory proteins is that they encode for a targeting sequence, which is recognized by SRP (blue sphere) to trigger transport to the ER, where translation resumes. Each segment of the protein can then either translocate through the translocon to become soluble, or insert laterally into the membrane to become a TMD.

The translocon is not a closed channel but has a lateral gate that allows for insertion of the polypeptide into the ER membrane. As the emerging polypeptide elongates, it can either continue to translocate through the channel into the ER-lumen, or partition laterally into the lipid bilayer. Exiting the translocon is kinetically determined and depends on the physical properties of the peptide. No specific sequence is required for partitioning into the bilayer, instead sufficient hydrophobicity and depletion of charged residues in a region that spans the bilayer is the major force to promote partitioning\textsuperscript{15}. A biological scale was developed by measuring the effect on insertion of each amino acid at each position within a hydrophobic peptide (ΔG\textsubscript{meas} < 0 will be inserted)\textsuperscript{10}, and resulted in a predictor of insertion (ΔG\textsubscript{apparent} < 0 are predicted to be inserted) for any peptide\textsuperscript{17}. TMDs with marginally hydrophobic TMDs where integration should be disfavored (ΔG\textsubscript{apparent} > 0) are mostly found in multi-spanning membrane proteins\textsuperscript{17}, where other TMDs or cofactors can bury hydrogen bond-capable residues internally instead of exposing them to the hydrophobic environment in the membrane\textsuperscript{18}.

A TMD may invert co-translationally within the membrane. Residues upstream of the transmembrane domain play a major role in which orientation the TMD will have in the bilayer, where positively charged residues are
strongly biased towards the cytoplasmic side (positive-inside rule\textsuperscript{19}). The orientation of each TMD is crucial since it is the deciding factor for which side of the membrane loops or globular domains downstream in the polypeptide will be located. The targeting sequence is not always inserted into the membrane. Downstream sequences in the polypeptide can also contain a TMD, and in these cases the ER-targeting sequence is often cleaved off from the mature protein. For proteins that do not contain a TMD, the entire protein translocates through the membrane, resulting in a soluble protein within the ER.

**Life**

Most secretory proteins move along the secretory pathway, to the Golgi and then to the plasma membrane, which for secreted soluble proteins results in their release from the cell. Some proteins stay in the ER or Golgi through protein-protein interactions, and specific retention signals such as a C-terminal KDEL/KKXX ER-retaining sequence\textsuperscript{20,21}. The affinity to the lipid environment also plays a role, explaining why the average TMD length and the membrane thickness both increase along the secretory pathway\textsuperscript{22}.

The oxidative environment of ER and outside the cell favors formation of disulfide bonds between pairs of cysteine residues. This can promote the correct folding pathway by covalently placing natively interacting residues close in space, and following folding they can stabilize the protein by reducing the entropy of the unfolded state, and thus increase the half-life of the protein. In contrast, free cysteines are often recognized by the ER quality control system and thus play a role in retaining misfolded proteins in the ER to make sure they are degraded.

Attachment of carbohydrates is perhaps the most common modification of secretory proteins\textsuperscript{23,24}, and they are most commonly attached to Asn (N-linked) and less commonly to Ser or Thr (O-linked)\textsuperscript{24}. N-linked glycosylation are co-translationally attached by the oligosaccharyltransferase complex in the ER-membrane. Glycans work in concert with ER-residing chaperones to contribute to correct folding, and to retain misfolded proteins in the ER\textsuperscript{25}. Many enveloped viruses insert glycans onto their main antigens both for folding, and to mask antibody epitopes ("glycan shield"\textsuperscript{26}), including HA and NA for influenza\textsuperscript{27}.

Secretory proteins, just as their cytoplasmic counterparts, can also acquire metal ions. Commonly the metal ion is Ca\textsuperscript{2+}, as the concentration in the ER and extracellularly is much higher than in the cytosol. Metallation of proteins have shown to play an active role in the folding process and for enzymes to be enzymatically active\textsuperscript{28-31}.

**Oligomerization**

For most proteins, the folding process is not finished after the full polypeptide has been synthesized. It is estimated that two-thirds of proteins are homo-
oligomers, with dimers and tetramers being the most and second most common states\(^{32}\). For many oligomeric proteins, the individual monomers are either unstable or incompletely folded, and oligomerization can be seen as an extension of the folding process. Uniquely to membrane proteins, the TMD(s) can drive the oligomerization. In contrast to the cytosolic aqueous environment where interactions are largely driven by hydrophobic forces, the hydrophobic environment in the lipid bilayer also favors polar interactions.

From the limited structural data for membrane proteins\(^{33}\), we know more about multi-spanning membrane proteins and their interactions than for single-spanning proteins where the soluble domain is often analyzed independently. Therefore, sequence motifs that interact in lipid bilayers have instead been extensively studied using reporter assays. In one such assay, GALLEX\(^{34}\), a TMD is fused to the DNA-binding domain of LexA (without the interaction domain), and expressed in cells where the β-galactosidase gene is put under the control of a LexA repressor. Following cell lysis, ortho-Nitrophenyl-β-galactoside (ONPG), which releases a chromophore upon β-galactosidase cleavage, can be added and the TMD interaction is measured by the repression of β-galactosidase activity. The GALLEX system has been used to study relative TMD interactions, and also to calculate apparent K\(_d\) and free energy values\(^{35}\).

The most studied protein in these systems is human glycoprotein A (GpA), which contains a strongly dimerizing TMD. It has been shown that the major driving force in GpA dimerization are glycines that form voids that valines can fill. This GxxxG sequence motif is highly overrepresented in TMDs\(^{36,37}\), but \textit{in silico} prediction of interacting TMDs has proven to be difficult. Even the GxxxG motif just provides a framework, and is highly dependent on the neighboring residues that can either stabilize or destabilize the interactions.

Randomized TMD libraries in different backbones have shown that many different amino acids can drive oligomerization\(^{38}\). However, known sequence motifs are not sufficient for an interaction, and many TMDs without known sequence motifs interact. Most α-helices have \(\approx 3.5\) residues per turn, meaning every seventh residue is on the same \(z\)-plane. Similar to the hydrophobic leucine heptad repeat in soluble proteins, TMDs can interact with polar residues on one side of the helix. These TMDs are usually amphipathic, with one side hydrophobic, facing the lipid bilayer, and the other side polar, facing the interacting helix. Amphipathicity can be determined by plotting the residues on a helical wheel, that illustrates the helix from the top to easy visualize if one or more neighboring heptad repeat (often termed seven faces, \(a\) to \(g\)) are more polar than the others\(^{39}\).

The lipid environment can play a role in modulating the TMD interactions. Mammalian plasma membranes consist largely of phospholipids asymmetrically distributed, and negative head groups mostly to the cytoplasmic side. Cholesterol, which is evenly distributed, can have a large effect on membrane
fluidity. Another discrepancy between the two membrane leaflets is the presence of glycoproteins and glycolipids in the outer leaflet. The *E. coli* inner membrane, which the vast majority of TMD interaction studies have used, share many of the characteristics of the mammalian plasma membrane, with asymmetrically distributed phospholipids and a large portion of membrane proteins. Due to the complex nature of biological membranes, their properties are difficult to reconstitute in vitro. Abnormalities have therefore been reported comparing in vitro interaction studies with those performed in vivo. In contrast, similarities have been reported comparing in vivo results from mammalian and bacterial membranes.

Death

Most human proteins live in the timescale of a few hours. Proteins are removed by two processes, cell division (affects all proteins) and degradation (protein specific). Most secretory proteins contain one or more N-linked glycans, and sequential remodeling of these structures by hydrolytic host proteins have been shown to occur to tag aging proteins as old. Turnover is not only a mean to control protein levels, but also to remove proteins that accumulate damage. With time, it is not uncommon that proteins get damaged, e.g. by deamidation or carbonylation on side groups, incorrect disulfide bonds, truncation, or aggregation. Another source of degradation is during the assembly process in the ER, where the ERAD (ER-associated protein degradation) pathway monitors the folding process of newly synthesized proteins. Folding intermediates or misfolded proteins are recognized by chaperones, by exposed hydrophobic patches, unpaired cysteines or immature glycan structures. These chaperones provide their substrates another opportunity to fold correctly, but if the proteins do not fold correctly, they are ultimately de-glycosylated and transported back to the cytosol and the ubiquitin-proteosome complex.

Secretory protein regulation

Regulation of secretory proteins, like their cytosolic counterparts, occur in all stages of their biosynthesis (Figure 3). Transcription of host proteins, including secretory proteins, requires accessible chromatin and the right transcriptional factors. The transcriptome is readily changed in the event of stimuli, such as sensing a viral infection, where secreted interferons are expressed which in turn induces a signaling cascade to initiate transcription of hundreds of antiviral proteins, many of which are secreted by the cell to signal for an infection. In the case of influenza, it has counter-evolved some of these processes using the NS1 protein to structurally mimic a histone, and induce repressive histone modifications that downregulate the antiviral cascade.
Once an mRNA is produced it also needs to be exported from the nucleus to the cytosol. It has been shown that the low adenine content in the ER-targeting sequence coding region can promote export efficiency and is less dependent on export factors. mRNA coding for secretory proteins has to ultimately become membrane bound, as the encoded protein will enter the ER. The dogma has been that protein synthesis occurs in the cytoplasm, and the nascent protein chain will be targeted together with the ribosome by SRP to the ER as soon as the N-terminal targeting sequence is exposed. More recently, reports have shown that (at least) two SRP-independent pathways exist. In one, the mRNA was shown to localize to the ER independent of being translated, and in the other, the ribosome was trafficked to the ER independent of SRP.

Most secretory proteins still contain an ER-targeting sequence. Due to the low sequence requirement of a targeting sequence (to be mainly hydrophobic), it has been questioned why targeting sequences differ so much. Targeting sequences can for example have different SRP binding strength. It was shown that mutations in an ER-targeting sequence reduced transport and expression, but also induced specific mRNA degradation, providing a link between the efficiency of SRP-targeting and mRNA turnover. However, these functions only provide fine-tuning to the expression, and do not provide any
temporal control. Yet, it has been shown that targeting sequences have unique structural features in their mRNA\textsuperscript{54}, and evolve faster than the mature protein\textsuperscript{55}, strongly indicating they perform additional functions.

The nucleotide content in an mRNA can have a large effect on expression and folding of proteins, by forming structures that modulate the rate of translation initiation and elongation by utilizing more or less rare tRNAs\textsuperscript{56}. The nucleotide composition of many viruses differs from that of human genes, and this is sometimes exploited by viruses which alters the tRNA pool and favor viral protein synthesis. For instance, HIV has been shown to alter the cellular tRNA pool during infection, while the cellular SLFN11 protein was shown to bind tRNA and selectively inhibit expression of HIV proteins with biased codon nucleotide composition\textsuperscript{57}. The nucleotides near the start codon are especially important. Initiating translation at the correct start codon can be mediated by a translation initiation site (Kozak sequence) near the start codon\textsuperscript{58}. Some proteins have a suboptimal Kozak site, which can be beneficial if more than one protein is encoded within same mRNA. Some viruses instead use internal ribosome entry sites in their mRNA that fold into tertiary structures to recruit ribosomal proteins\textsuperscript{59}. In addition, the untranslated regions (UTRs), in the 5’ and 3’ ends of each mRNA play important roles in regulating expression\textsuperscript{60}. These regions are often bound by RNA-binding proteins, which can have opposing regulatory functions, and be expressed in a cell and cell-cycle dependent manner\textsuperscript{61}. The ER-targeting sequence coding region is positioned just after the 5’ UTR, but it is not known whether this region can act as a 5’ UTR extension and play a role in regulation.

For many viruses, gene expression is temporally defined with genes that express either early or late during the infection. For papillomaviruses, temporal gene expression is regulated at the level of transcription and RNA processing, where expression of late genes is coordinated by a promoter switch as well alternative splicing and poly-adenylation signals\textsuperscript{62-64}. As mentioned, the main antigens of enveloped viruses are usually secretory proteins. These proteins often have potential to promote membrane curvature, and if they bud before all other viral components are ready, will cause inefficient viral production and extended exposure to the immune system. These antigens are therefore often expressed later (late genes) than some of their cytosolic counterparts (early genes). For HIV, the sole membrane protein encoding gene, Env, harbors a sequence that retains the mRNA in the nucleus, but is readily exported late in infection by the HIV RNA-binding protein Rev, which is expressed early in infection\textsuperscript{65}. 
The influenza virus

The influenza A virus (IAV) is a spherical or filamentous enveloped particle with its RNA genome divided into eight separate segments (Figure 4). Its primary natural reservoir is wild birds, but it is also found in numerous other species, including humans where it causes seasonal flu epidemics and at times devastating pandemics. The virus is one of the only RNA viruses that replicates in the nucleus. IAV encodes 10 major proteins, and a handful of strain-specific proteins. IAV has three membrane proteins, Hemagglutinin (HA), Neuraminidase (NA) and Matrix Protein 2 (M2) (Figure 4). HA and NA are involved in binding and cleavage of sialic acid, respectively. Sialic acids are a group of nine carbon sugars most often found at the terminal end of glycan chains of human secretory proteins and lipids. Due to their role in many crucial biological processes, numerous pathogens target this molecule as their cell receptor, just as influenza does.

Figure 4. Left: Schematic of the three IAV membrane proteins. The ER-targeting sequence (red) of HA is cleaved off and the TMD is located in the C-terminus. For NA and M2, the ER-targeting sequence also functions as a TMD. Right: An influenza particle with the eight gene segments. An NA homo-tetramer is shown with the a.a. numbering (WSN/H1N1) of the different domains.

The homo-trimeric HA is the most abundant membrane protein on the influenza surface, and is the main antigen. The HA0 precursor is proteolytically cleaved to form HA1 and HA2. The N-terminal HA1 domain makes up the globular head domain that contains the sialic binding site. HA2 makes up the stalk domain and contains both the fusion peptide and the C-terminal TMD. At the low pH of endosomes during entry, the protein is structurally altered to expose the fusion peptide, that can fuse the viral and endosomal membrane.

The homo-tetrameric NA is tethered to the membrane by an N-terminal TMD, and has a large C-terminal head domain (Figure 4, inset). No structural data is available for the TMD or for the stalk domain connecting the TMD and the head domain. The TMD of NA consists of 28-29 residues consistent with
many plasma membrane proteins and have a normal amino acid composition with increased levels of the polar residues Asn, Ser, Thr, Gln, and hydrophobic residue Ile. The TMD alone is enough for cell surface localization (although with different efficiencies between subtypes), apical targeting in polarized cells, and sorting into raft domains (Triton X-100 insoluble). The stalk domain connects the TMD to the enzymatic head domain. Given the apparent simple role of connecting two domains, the stalk region shows low sequence conservation and is rich in insertions/deletions. Mutations in both the TMD and stalk domain were reported to affect enzymatic activity, indicating that all domains work in concert for NA to function. An enzymatic pocket is placed in each monomer that cleaves off terminal sialic acids. The role of NA during budding is well established, it releases the progeny virus from the host, which HA otherwise bind to and cause cell retention. NA have also been shown to cleave off sialic acids from glycans on NA and HA themselves (which otherwise would promote HA-mediated influenza particle aggregation), from decoy receptors in the mucus, as well as from the host membrane before internalization (which results in HA-NA mediated movement on the cell surface, allowing for HA to find the correct receptor).

M2 is a homo-tetrameric protein that form a proton-selective ion channel. During cell entry in a new infection, the viral particle travels in the endocytic pathway, with a sequential drop in pH. Acidification will eventually activate M2 by protonation of a His residue, which will reposition its Trp gate and cause M2 to pump protons from the endosome to the inside of the viral particle. The low pH inside the virus will disrupt the interactions between M1 and the gene segments, which can then be released to the cell and enter the nucleus.

Synthesis and regulation of influenza secretory proteins

HA, NA and M2 mRNA are transcribed by the influenza polymerase complex (PB2, PB1, PA, together with the RNA-binding protein NP). The negative polarity gene segments serve as templates, and the polymerase binds to the promoter in the short UTR downstream of the stop codon (and conversely, the short UTR upstream of the start codon is used to generate new gene segments of negative polarity). To mimic host cell mRNA, a 5' cap is snatched from host pre-mRNA. First bound to PB2, the 5' RNA is cleaved by PA, and is then used to prime mRNA synthesis by PB1. A poly-adenylated tail is inserted by stuttering of the influenza polymerase during transcription, and so is independent of the host pre-mRNA processing machinery for poly-adenylation. Viral mRNAs utilizes both major export pathways to the cytosol.

The N-terminus of HA, NA and M2 contains an ER-targeting sequence (Figure 4), which for NA and M2 also functions as a TMD. The ER-targeting
sequence of HA is cleaved off, and the TMD is located in the very C-terminus of the protein. Specific for NA is inversion of the TMD for the protein to obtain an N_in-C_out orientation, driven by positive charge in the first six residues before the TMD\textsuperscript{19,99}.

Both HA, NA and M2 contain inter-disulfide bonds that stabilize their oligomeric structure. In contrast to M2, HA and NA have a large extracellular domain, which is glycosylated at multiple sites. For HA, it was shown that the co-translational addition of disulfide bonds and multiple glycans is strategically placed to play an active role in the folding process of HA\textsuperscript{100}. HA is subsequently proteolytically cleaved after maturation\textsuperscript{101} to form HA1 and HA2, which remain together through a disulfide bond. HA2, which contains the fusion peptide, is only active post-cleavage\textsuperscript{102}.

The large head domain of NA is enzymatically active only as a tetramer\textsuperscript{103}. Oligomerization of the protein occurs very early during NA biogenesis\textsuperscript{104}. Monoclonal antibodies have been generated that bind either dimeric and tetrameric, or just tetrameric NA\textsuperscript{104,105}. The epitope of these antibodies are not in the tetrameric interface, indicating that conformational change occurs during maturation which might explain the enzymatic requirement of an oligomer. Specific for NA is multiple Ca\textsuperscript{2+} ions bound to the head domain\textsuperscript{106}, and although their specific contribution is unknown, removal using EDTA renders the protein enzymatically inactive\textsuperscript{31,107,108}. Sialidase activity is routinely measured using 2-(4-methylumbelliferyl)-D-N-acetylneuraminic acid (MU-NANA), a sialic acid residue covalently attached to a fluorophore that becomes fluorescent upon NA cleavage.

NA from the WSN/H1N1 strain was one of two IAV proteins shown to have suboptimal translation initiation sites\textsuperscript{109}, which can lead to the first Met codon be missed by the ribosome. Whether this has a biological significance for NA is unknown, but IAV segment 2 encodes not only the polymerase PB1 protein but also PB1-F2\textsuperscript{110} and NV40\textsuperscript{111} through leaky scanning to alternative initiation sites. In addition, the NA segment of influenza B virus encodes roughly equal amounts of NB through an alternative start codon from an overlapping reading frame, a conserved 100 a.a. ion channel that similarly to NA is glycosylated and enter the membrane by a non-cleavable targeting sequence\textsuperscript{112}.

Influenza uses host proteases to post-translationally cleave HA to become active. A host interferon-stimulated gene, Plasminogen activator inhibitor (PAI-I), extracellularly binds covalently to host-proteases to inhibit their activity, and thus the maturation of HA\textsuperscript{113}. Interestingly, NA from the WSN/H1N1 strain has uniquely evolved the ability to recruit plasminogen to promote HA cleavage\textsuperscript{114}. As HA and NA perform reciprocal functions (binding and cleaving sialic acid), a balance of HA and NA activity has been reported in viruses and to have efficient transmission\textsuperscript{115-117}, where non-balanced chimeric viruses were reduced in transmission or acquired mutations to balance their activities.
HA must also make sure not to activate its low-pH conformation in the secretory pathway during its biosynthesis. The ion-channel M2 is therefore expressed earlier than HA, and as M2 travels the secretory pathway it will get activated by the low pH in the trans-Golgi, and pump protons from the trans-Golgi to the cytoplasm. This pH-equilibration of the trans-Golgi prevents a premature low-pH induced conformational change in HA\textsuperscript{118}. M2 is synthesized from alternative splicing of segment 7, which otherwise encodes for the soluble M1 protein. Regulation of M2 protein levels occurs mainly from altering the splicing efficiency. The ratio of M2 to M1 mRNA increases during the infection\textsuperscript{119}, M2 splicing has been reported to be promoted by NS1\textsuperscript{120,121}, and the amount of spliced mRNA differs between strains\textsuperscript{122}.

NA and HA have the capacity to form virus like particles (VLPs), when expressed on their own\textsuperscript{123,124}. Given that uncontrolled expression would likely result in depletion of cellular NA and HA as well as expose them to immune factors, it is not surprising that both NA\textsuperscript{125} and HA\textsuperscript{126,127} express later in infection. However, how NA and HA are temporally regulated by host or viral factors is not known.

**NS1, a major regulator**

The smallest influenza gene segment encodes both the multifunctional RNA-binding protein NS1, as well as NS2\textsuperscript{128}. NS1 is a two-domain protein consisting of an N-terminal RNA-binding domain and a C-terminal effector domain\textsuperscript{129}. Although the name (non-structural protein 1) suggests the protein is not present in the virus particle, which was demonstrated early\textsuperscript{130}, it has recently been shown to actually be present in the particle, but in low amounts\textsuperscript{75}. NS1 has been associated with numerous cellular processes that selectively increase influenza protein expression\textsuperscript{131}, and the RNA-binding domain has been shown to bind both host and viral RNA\textsuperscript{132-134}. NS1 binds and blocks proteins involved in host polyadenylation to inhibit host mRNA synthesis\textsuperscript{135}, binds factors involved in mRNA export to promote viral mRNA export\textsuperscript{136}, binds translational elongation factors\textsuperscript{137}, and multiple other proteins.

Importantly, NS1 is often also attributed to regulate members of the innate immune system\textsuperscript{138,139}. A warfare occurs during replication between the virus and the host's innate immunity system, with the chief commander being the interferon protein family\textsuperscript{140}. The main inducer of the interferon system in influenza infected cells is the RNA-binding protein RIG-I which detects the viral RNA\textsuperscript{141}. NS1 inhibits interferon activation through multiple mechanisms, including sequestering viral RNA\textsuperscript{142} and binding TRIM25 to inhibit RIG-I activation\textsuperscript{143}. Due to the crucial nature of NS1 in inhibiting the innate immunity, a ΔNS1 virus grows significantly better in cells (Vero cells\textsuperscript{144}) and mice (STAT\textsuperscript{-/-}\textsuperscript{145}) deficient in interferon signaling\textsuperscript{146}. Still, many cells remain unstimulated even after infection with a ΔNS1 virus, indicating other viral proteins also play an important role targeting the innate immune system\textsuperscript{147}. 

23
Together, a lot is known about virus-host interactions, many of which are performed by NS1. Although many aspects of viral infections are coordinated in a time-dependent manner, less is known whether interactions occur early or late. For M2, synthesis requires splicing of the M segment. It has been shown that NS1 promote splicing of the M segment120,121, and that the M1/M2 mRNA ratio is altered during infection119. Whether NS1, or any other viral protein, also is involved in the temporal regulation of the other membrane proteins HA and NA is not known.

The influenza virus, a tool to study secretory proteins

The influenza virus is constantly evolving by incremental changes through its error-prone RNA polymerase148, and less common significant changes by gene segment reassortment during co-infection due to its segmented genome149. Viable sequences can therefore be studied by introduction mutations and observe compensatory mutations the virus adds to regain normal growth. The biospace of secretory proteins HA and NA is so diverse that they are each separated into serotypes (H1-16, N1-9) based on antibody recognition150. This biological wealth together with collaborative sequencing efforts have resulted in a significant availability of genomic sequences unmatched by that of any other organism or pathogen151. This have turned into a tremendous resource for studies to be complemented with in silico analysis, where the viability of sequences can be studied on a large scale.

Viruses have since long been used to study various cellular processes, and also to discover cellular compartments152,153. Due to their small genomes, they most often use host pathways rather than inventing new pathways. However, in contrast to their hosts, they often use simple solutions, due to their genome size, and that can be easier to study. Influenza is not an exception to this, with just eight gene segments transcribed similar to host genes, which encode ten major proteins. Secretory proteins are known to be regulated by RNA-binding proteins, and IAV encode two RNA-binding proteins (NS1 and NP), and we show (Paper IV) that NS1 regulates NA and HA expression.

Many different aspects of secretory protein maturation can be studied using influenza, as the three secretory proteins have different oligomeric state (trimer and tetramer), ER-targeting sequence (non-cleaved and cleaved), TMD localization (N-terminus and C-terminus), protein orientation (N<sub>in</sub>-C<sub>out</sub> and N<sub>out</sub>-C<sub>in</sub>), TMD hydrophobicity (low and high99), post-translational modification (proteolytically cleaved and not), and function (substrate-binding and fusion peptide, enzymatic activity, and ion channel). In addition to all these properties, expression of these proteins is regulated, and we show (Paper IV) that NA and HA are potentially regulated by a different mechanism from M2.
Results

**Paper I:** Single-pass transmembrane helices are usually viewed as simply membrane anchoring domains. When the TMD was replaced with a cleavable signal peptide to create a secreted soluble NA, almost no enzymatic activity remained, and a large fraction was misfolded. The activity was sequentially rescued by truncating the stalk domain in this chimera, indicating that the NA TMD plays a major role in assembly when a stalk domain of biological length is present. The activity was also rescued by introducing a coiled-coil domain after the signal peptide, with the highest rescue when this domain had a tetrameric assembly thus matching the oligomeric assembly of the head domain. Together with showing the NA TMD can oligomerize on its own, these results indicate that the TMD plays a major role in assembly of the distal head domain, and that optimal assembly requires properties of the TMD to mimic those of the head domain.

**Paper II:** The NA TMD interaction was studied by a reporter assay (GAL-LEX) in a biological membrane. TMD oligomerization could be measured for consensus TMDs from all subtypes, except for N2. A conserved non-hydrophobic character was revealed in the TMDs, for some subtypes so low that partitioning into the ER bilayer should be disfavored according to predictions ($\Delta G_{\text{apparent}}$ for membrane insertion $<0$). Burying these non-hydrophobic residues by forming TMD interactions is a potential way to promote partitioning and increase the oligomerization strength. Indeed, TMD interaction was inversely correlated with hydrophobicity, and oligomerization was abolished when the polar residues were mutated. Analysis of human NAs from a sequence database showed substantial variation in N1 TMDs, but not for N2 TMDs. A temporal inspection of the human N1 dataset revealed a clear trend of more polar TMDs in recent strains, which correlated with higher interaction strength. Together, the key residues involved in NA TMD oligomerization have been identified, the oligomerization is mostly conserved between subtypes, and for human N1 the amphiphilicity and TMD interaction have gradually increased over time suggesting an evolutionary pressure on the TMD.

**Paper III:** Although membrane integration is the definition of a TMD, human N1 TMDs have evolved against this property by decreasing their hydrophobicity. To study the biological role of this change, we created a chimeric virus where the NA TMD from WSN/H1N1 that has medium interaction strength
was replaced with the TMD from Cal09/H1N1 that has high interaction strength. Although this NA should be more stable due to the new TMD, the resulting virus showed a severe growth defect at 37 °C, but not at 33 °C. Serial passaging at 37 °C, but not 33 °C, resulted in adaptive mutations in the NA TMD. Viruses with each TMD mutation were independently generated, and in all cases the growth profile returned to wild-type (WSN TMD) levels. Analyzing the TMD interaction revealed that all were decreased to levels of wild-type (WSN). In addition, although the original TMD chimera showed a folding defect, all mutations increased the folding efficiency of the NA protein. When a Cal09/H1N1 TMD chimera with altered amphiphilicity was generated, it showed even stronger attenuation at 37 °C, and again readily acquired adaptive mutations in the TMD that restored growth and increased folding efficiency. These results show that subtle property changes in the NA TMD can cause large defects in the distal head domain, and suggest that evolution of the NA TMD has been driven by changes in the head domain and to maintain compatibility between these domains.

**Paper IV:** Enveloped viruses often express their membrane proteins late during infection. Although the distinction between early and late genes is less defined for influenza, the viral RNA-binding proteins NS1 and NP have been shown to express earlier than NA and HA. To study viral factors that regulate gene expression, we employed a transient transfection approach where HA and NA were co-expressed with other genes and the resulting protein levels were monitored. The reduced plasmid-based expression mimics an infection by supporting nuclear transcription, mRNA export from nucleus to cytosol, and translation at the ER membrane. The RNA-binding domain of NS1 was shown to significantly increase HA and NA levels. The regulation was mapped to the 5’ coding region of HA and NA, as synonymous substitutions in the HA or NA 5’ ER-targeting sequence coding regions, or abolishing the RNA-binding activity of NS1, both reduced the regulation. As neither RNA levels or RNA export was affected, and NS1 showed enriched foci near the ER, the regulation likely occurs at the step of translational initiation or targeting to the ER. The regulation is likely before translational elongation, as regulation persisted when the 5’ regulatory region was moved to the 5’ UTR and thus not being translated. Sequence analysis of human mRNA revealed that 5’ ER-targeting sequence coding regions have a similar nucleotide content as their respective 5’ UTR. As 5’ UTRs are often involved in gene specific regulation, we speculate that the ER-targeting sequence coding region in human proteins not only encode an ER-targeting sequence, but may also act as an extension of the 5’ UTR and perform a regulatory role. This regulatory mechanism of an RNA-binding protein and ER-targeting sequence coding regions may therefore extend to other proteins except NA and HA.
Conclusions and future perspectives

Current vaccines are only standardized to their HA content, and the induced immunity is predominantly by a response to HA. Therefore, the amount of NA in the vaccine is unknown, and given the low stability of NA it is sometimes not functionally active in vaccines\textsuperscript{154-157}. However, NA antibodies have protective capacity and can significantly reduce titers\textsuperscript{158-161}, and NA was recently called “the forgotten antigen”\textsuperscript{160}. Although the head domain of NA was found to remain active after proteolytic cleavage from the rest of the protein\textsuperscript{77,162,163}, both the TMD and stalk domain play important roles in assembly and activity of the full-length protein. As structural data is still missing on the N-terminal part of the protein, biochemical experiments are greatly needed to elucidate the full functions of each domain.

Here, we have studied assembly and regulation of NA, and show that the N-terminal domains play an important role in these processes. I hope these findings can be useful in creating vaccines with more stable NA (the N-terminus is aiding in oligomerization (Paper I, II, and III)), and with higher levels of NA (the N-terminus is aiding in enhancing protein expression (Paper IV)). We also show (Paper IV) that host secretory genes may be regulated in a similar manner as NA and HA, and I look forward to see these additional RNA-substrates of NS1 and host RNA-binding proteins and whether any of them have therapeutic potential.

Specifically, we observed that human H1N1 NA TMDs have become more polar over time to the extent that TMDs from current strains are not predicted to partition into the membrane. This amphiphilic drift goes against the defining property of a TMD to be hydrophobic, and is therefore likely functional. We found that the amphiphilic drift correlated with stronger TMD interaction (Paper II) as well as stronger inhibition of expression in the absence of NS1, which resembles the early stage of infection (Paper IV and data not shown). Using NA TMDs with different hydrophobicity, co-translational insertion was recently suggested to lower the hydrophobicity threshold of insertion\textsuperscript{99}, possibly explaining how H1N1 NA can compromise TMD hydrophobicity to gain two additional TMD functions (oligomerization and regulation).

Apart from H1N1, H3N2 is also circulating in the human population. The NA TMD from H3N2 was found to have no temporal change in hydrophobicity (Paper II), little to no TMD interaction (Paper II), and very weak translational stimulation with NS1 co-expression (data not shown). The am-
philic drift to accommodate new functions seems therefore specific for hu-
man H1N1, and it remains to be seen how H3N2 NA assembles into a te-
tramer, how the protein is temporally regulated, and whether any constraints
exist that could explain the lack of additional roles in the TMD.
Sammanfattning på svenska

Trots att membranproteiner utgör en tredjedel av proteinerna i vår kropp, och att en majoritet av alla läkemedel attackerar membranproteiner, så har kunskapen om membranproteiner alltid varit lägre än den om proteiner som inte sitter fast i membran. Membranproteiner innehåller en eller flera ”transmembrana domäner”, som förankrar den funktionella delen av proteinet till membranet. Influenzaviruset har tre membranproteiner, varav ett av dem, neuraminidas (NA), utgör målet för de flesta influensaläkemedlen. NA är svampformat med en kort rot (transmembran domän), en stjälk, och ett huvud med enzymatisk aktivitet, som möjliggör att nya influensapartiklar kan frisläppas från en infekterad värdcell. Kritiskt för dess enzymatiska aktivitet är att den är en tetramer, d.v.s. fyra NA-proteiner tätt intill varandra. I denna avhandling påvisar jag att roten i NA har fler funktioner än bara förankrar NA i membranet.

Artikel I: Roten i NA stabiliserar stjälken i en tetramerisk struktur. Optimal stabilisering skedde när roten också tetrameriserade.

Artikel II: Tetrameriseringen i roten är mestadels konserverad i NA från olika influensastammar, och drivs av polära aminosyror i roten. Vi kunde också påvisa att roten har blivit mer polär genom åren, vilket antyder att det finns ett evolutionärt tryck på roten och att den därmed spelar en större roll än bara som en förankring.

Artikel III: När roten gjordes mindre polär i ett influensavirus växte detta viruset långsammare, men viruset muterade snabbt så att roten blev mer polär, vilket resulterade i att roten kunde tetramerisera och viruset återgick till sin normala tillväxtfart.

Acknowledgments

Rob I owe you a big thank you for all you have done for me. You took me under your wings and taught me so much. You have given me very interesting projects and have always been so helpful. It has been a great couple of years! I sincerely hope your scientific excellence will be appreciated by everyone.

Diogo, I miss the old times when we did experiments next to each other. You never hesitated to spend a lot of time helping me. You felt like my big brother, always taking care of me. I definitely think you can get whatever your dream job is! Dan & Hao, it has been great working with you both on a work and a personal level. Henrik, det har varit en väldigt kul tid ihop. Hoppas du inte kommer stagnera nu utan någon att se upp till, men du kan alltid slå en signal om du behöver hjälp med t.ex. grammatik eller organisering.

A big thanks to the French community. Jeanne & Antoine my NS1 project would not be where it is today without you. Merci pour votre aide. Stockholm or Paris, J’espère vous revoir bientôt! Rebecca, unfortunately we did not have many months overlap, but the short time has been great. I hope you enjoy your time at DBB as much as I did.

A big hug to everyone who made the years more enjoyable. Jörg, Claudio, Kievash, Stephen, Patrick, Bill, Thomas, Bea & Pedro, Catta, Rickard, Grant, Nina, Pontus, Jobst, Weihua & Biao amongst many. The list goes on for much longer, so even if I did not mention you by name, I will probably miss you!

Tack Mikael och Elzbieta för er roll som co-handledare och mentor, fixargänget (Matthew, Håkan, Peter) för att allt för det mesta fungerar, och hela sekretariatet för alla jobbiga frågor jag ställt.

Prof. Fodor, Schwarz and Smed Sörensen, thanks for being my opponent/committee. I am sure the dissertation day will be enjoyable.

Mest av allt vill jag tacka ♥ Chenge. Alldeles för många helger och sena kvällar medan maten blivit kall. Tack för att du är världens bästa flickvän.

Sist men inte minst, tack till alla i familjen & släkten. Ni har alltid stöttat mig i vått och torrt! Men nu är det slut med gulliga smeknamn, nu tilltalar ni mig som Doktor Johan!
References

28. Braakman, I. & Bulleid, N. J. Protein folding and modification in


105. Saito, T., Taylor, G., Laver, W. G., Kawaoka, Y. & Webster, R. G.


118. Ciampor, F. et al. Evidence that the amantadine-induced, M2-mediated conversion of influenza A virus hemagglutinin to the low pH


