



Topology of membrane proteins in Stockholm

# An Energy Code for Helices

One of the greatest challenges for protein science is the question of how the sequence of amino acids determines the topology of a protein. Gunnar von Heijne and his team approach the answer in the case of transmembrane helices.

The insertion of pore proteins or receptors into the cell membrane takes place at the ER. The translating ribosomes interact with a protein complex called the translocon which somehow anchors parts of the newly synthesized peptides in the lipid bi-layer while it smuggles others through it. In most cases the transmembrane segments are very tightly packed bundles of  $\alpha$ -helices. The recognition of such helical structures in the sequence of amino acids takes place inside the translocon but how does the translocon “know”, which part of the protein will be a transmembrane helix?

For a long time the answer has been very general, the transmembrane segments have to be hydrophobic in order to interact optimally with the lipid bi-layer. Therefore the primary sequence has to contain mainly lysines, alanines or other non-polar amino acids. All in all this is an unspecific hint because the membrane of the ER is not homogenous and the sequential distribution of amino acids in the polypeptide chain may play a role. Gunnar von Heijne and his colleagues from the Bioinformatics Centre of Stockholm University now offer a more exact answer. They present an amino acid code for transmembrane-helix recognition. In future this code might even allow the prediction of transmembrane helices from the amino acid sequence of unknown proteins.

## With the help of dogs

“I have been trying to understand how the amino acid sequence determines the topology of proteins for a very long time,” says von Heijne, professor at the Centre for Bioinformatics. In 1975 he graduated in chemistry. During his PhD he turned towards theoretical biophysics and studied protein chemistry. Since then he has been working on the modelling of peptide se-

quences and processes of protein assembly. Four years ago he and his group eventually started to focus on transmembrane helices. Von Heijne says, “Our aim is complicated. How must an amino acid sequence look like in order to be inserted into the membrane of the ER?”

The Bioinformatics Centre offers optimal conditions for his work. His own group is never bigger than 12 people. “A perfect size. We don’t want to be bigger,” he says. In fact the campus offers a very good local environment for interaction with other labs if necessary. Inter-disciplinarity is one of the most crucial conditions for von Heijne’s work. “We have a lot of good people here. There is much knowhow around us which we can use for our ideas,” he says

To investigate the insertion of transmembrane helices into the ER membrane



Annie von Heijne

Gunnar von Heijne

von Heijne has come a long way. In the beginning he and his group used artificial peptides with random sequences of amino acids and tested to see if they could be inserted into a lipid bi-layer. “From these initial experiments we gained a very general idea about the degree of hydrophobicity required for insertion of transmembrane segments,” says von Heijne. Then he met Stephen White from the University of California at Irvine who experimented with very well defined sequences. Von Heijne says, “After I had talked to Stephen White I thought why not combine our knowledge?”

## It's like a cell but it is not a cell

From this moment on von Heijne and his colleagues started to test peptides with exactly determined sequences. “We attached one component after the other, until we had constructs of about 20 amino acids in length with defined positions for some of them,” he says. The rest was routine work; in order to simulate the processes at the ER

the researchers cloned their constructs into an *E. coli* protein which is normally inserted into the membrane. In a second step they translated them *in vitro* in the presence of dog microsomes. Von Heijne says, “The dog microsome system is very useful. You take the whole extract of a dog’s pancreatic cell with one difference: there is no cell membrane. It’s like a cell, but it is not a cell.”

In fact the system contains ribosomes, the ER-membrane and the translocon complexes. The constructs are transcribed and translated like normal proteins: The ribosomes interact with the translocons which mediate the nascent polypeptide chain through the membrane of the ER. The interesting question for von Heijne was: To what extent will each of the artificial segments in the *E. coli* protein be inserted into the lipid bi-layer?

## An elegant trick

To answer this von Heijne and his colleagues invented an intelligent trick. In order to control the insertion they flanked their peptides with two sequences which are normally glycosylated by enzymes in the lumen of the ER. In the case of insertion, only one of them would show in the lumen of the microsome and would be glycosylated. In the other case the complete segment would be pushed through the membrane and both of the glycosylation sites would be processed. With this paradigm von Heijne and his group were able to quantify the probability of insertion for every single sequence they wanted because the protein with two glycosylated sites is heavier than the other one; it runs slower on a SDH-Gel.

From the ratio of inserted and non inserted peptides von Heijne and his group computed free energy values describing the chemical interaction between the bi-layer and their peptide segments. Negative values mean a high rate of insertion and therefore a gain in the free energy of the protein-lipid system. In contrast positive values mean a low rate of insertion. “In our initial experiments with random sequences we discovered very general principles. Now we can go even further. We can say something

about the required sequence characteristics and their impact on hydrophobic interactions between the peptide and the membrane," says von Heijne.

The researchers tested each amino acid separately for its optimal position in the sequence of a transmembrane segment. Therefore they built standard constructs from the hydrophobic amino acids Leucine and Alanine and only at one position they placed one of the 20 amino acids in question. "We moved for example an aspartate from the middle position to the position next to the middle and measured how the ratio of glycosylated proteins changed," says von Heijne. They had to test over 400 different constructs but the effort was worth it.

The outcome was a complete profile of free energy values for each amino acid in each position of a transmembrane helix (*Nature*, 2007, vol. 450, p. 1026). This profile predicted the frequencies of amino acids in transmembrane segments of natural proteins quite well. When von Heijne and his group compared their profiles to the X-Ray data of known proteins, they gained a good fit. The probability that they would find a particular amino acid in a particular position of the natural transmembrane segment corresponded to the free energy value for this amino acid in this position gained from the artificial constructs.

"Of course this finding was great. It led us to a mathematical model describing the probability for membrane insertion," says von Heijne. This model defines a compound free energy value for an amino acid sequence of a particular length. In other words: it predicts whether a segment will be inserted into the membrane of the ER or not, depending on its amino acid sequence. It does so by summing up the free energy values of every single amino acid in every single position to an integral. This integral is then corrected for the length of the segment, because the length turned out to affect the free energy values in some other experiments and that's it.

### Knowing while reading

Again the fit between the predicted behaviour of sequences and the knowledge gained from natural transmembrane proteins matched quite well. This was shown by another experiment which von Heijne and Co. performed with their new model. In this experiment they compared the amino acid sequences of known transmembrane-helix proteins, multi-spanning transmembrane-helix proteins and secretory proteins which are not integrated into the membrane of the

ER. They scanned these sequences for segments with the lowest free energy values as predicted by their model.

"The outcome was really surprising," says von Heijne. Only transmembrane-helix proteins had segments with negative free energy values and this fits with the fact that these proteins consist of segments that are really inserted into the membrane. In contrast even the lowest values for secretory proteins were above zero. This might be the reason why the translocon pushes them into the lumen of the ER.

### Sequence in, topology out

"This shows that our model not only describes the recognition of simplified artificial segments by the translocon but also of natural ones," says von Heijne. It seems that the translocon "reads" the sequence of the nascent protein and somehow "knows" which segment fits into the lipid bi-layer. Therefore the profiles of free energy values for single amino acids represent a code for hydrophobicity which the translocon uses.

Of course nobody yet knows what exactly happens inside a translocon. "Other data, however, implies that the translocon handles the helical transmembrane segments not step by step but in one piece," says von Heijne. Furthermore the helices of multi-spanning transmembrane proteins seem to interact with each other inside the protein complex. In the experiment not all of them show negative free energy values. Von Heijne says, "Obviously the translocon recognizes only some of the helices in multi-spanning transmembrane proteins. These then influence their neighbours in order to adjust them properly so that they can better be inserted into the lipid bi-layer."

To test this prediction future work is needed. Another aim of von Heijne's group is to use their free energy profiles in order to predict transmembrane-helices in unknown proteins. "The next step is to feed our model into a computer. If I may dream: I'd like to have something with the amino acid sequence as input and the topology of a protein as output," says von Heijne.

Together with other groups at the Centre of Biomembrane Science he is entering his knowledge into a computer programme; the first trials are already promising. The prototypes deliver a prediction success of 75-80%. This is similar to already existing programmes which are based on statistical approaches and machine learning algorithms. "So far our programme is as good as others but of course we would like to make it better," says von Heijne. MATTHIAS NAWRAT