THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN NATURAL SCIENCE

AQUAPORINS

PRODUCTION OPTIMIZATION AND CHARACTERIZATION

FREDRIK ÖBERG

UNIVERSITY OF GOTHENBURG

Department of Chemistry – Biochemistry

Göteborg, Sweden

2011

Thesis for the Degree of Doctor of Philosophy in Natural Science

AQUAPORINS: Production Optimization and Characterization

Fredrik Öberg

Cover: Cells of the yeast *Pichia pastoris*, producing hAQP5 fused to green fluorescence protein. Visualized using confocal microscopy.

ISBN 978-91-628-8290-7
Available online at http://hdl.handle.net/2077/25277

Department of Chemistry
Biochemistry and Biophysics
SE-413 90 Göteborg, Sweden

Printed by Chalmers Reproservice
Göteborg, Sweden 2011

Till min familj

ABSTRACT

Aquaporins are water facilitating proteins embedded in the cellular membranes. Such channels have been identified in almost every living organism – including humans. They are vital molecules and their malfunction can lead to several severe disorders. An increased understanding of their structure, function and regulation is of utmost importance for developing current and future drugs.

The first problem to overcome is to acquire the proteins in sufficient amounts to enable characterization. To achieve this, proteins are often produced in a host organism. One of the most successful hosts for recombinant overproduction is the yeast *Pichia pastoris*. Using this yeast we could obtain exceptional yield of aquaporin 1, whereas some others were below the threshold needed for successful subsequent characterization. In this process, we have established methods allowing fast and accurate determination of the initial production yield. Furthermore, we optimized the yield for low producing targets, enabling studies of proteins previously out of reach, exemplified with human aquaporin 4.

Characterization has been performed on aquaporins obtained in sufficient quantities, and the functionality of aquaporin 1, 5 and 10 has been assessed. Furthermore, a glycosylation was found to stabilize the aquaporin 10 tetramer although only a minority of the monomers where modified. Moreover, we used protein crystallography to determine the three dimensional structure of a hAQP5 mutant, providing insight into regulation of the protein by trafficking.

Taken together, these results provide insight into factors directing high production of eukaryotic membrane proteins. The subsequent characterization, including functional and structural determination, reveals new knowledge about aquaporin activity and regulation.

LIST OF PUBLICATIONS

This thesis is based on the following papers listed in reverse chronological order. They are appended at the end of the thesis and will be referred to in the text by their roman numerals.

Paper I. Nyblom, M., **Öberg, F.**, Lindkvist-Petersson, K., Hallgren, K., Findlay, H., Wikström, J., Karlsson, A., Hansson, Ö., Booth, P. J., Bill, R. M., Neutze, R. and Hedfalk, K. (2007) Exceptional overproduction of a functional human membrane protein. *Protein Expr Purif*, 56, 110-20.

Paper II. Hedfalk, K., Pettersson, N., **Öberg, F.**, Hohmann, S. and Gordon, E. (2008) Production, characterization and crystallization of the *Plasmodium falciparum* aquaporin. *Protein Expr Purif*, 59, 69-78.

Paper III. **Öberg, F.**, Ekvall, M., Nyblom, M., Backmark, A., Neutze, R. and Hedfalk, K. (2009) Insight into factors directing high production of eukaryotic membrane proteins; production of 13 human AQPs in *Pichia pastoris*. *Mol Membr Biol*, 1-13.

Paper IV. **Öberg, F.**, Sjöhamn, J., Conner, M.T., Bill, R.M., Hedfalk, K. (2011) Improving recombinant eukaryotic membrane protein yields in *Pichia pastoris*: the importance of codon optimisation and clone selection. *Submitted*

Paper V. **Öberg, F.**, Sjöhamn, J., Fischer, G., Moberg, A., Pedersen, A., Neutze, R., Hedfalk, K. (2011) Glycosylation increases the thermostability of human aquaporin 10. *Submitted*

Paper VI. **Öberg, F.**, Sjöhamn, J., Hedfalk, K., Neutze, R., Törnroth-Horsefield, S. (2011) Crystal structure of the S156E-mutant of human aquaporin 5. *Manuscript.*

Related publication

Paper VII. Wöhri, A. B., Johansson, L. C., Wadsten-Hindrichsen, P., Wahlgren, W. Y., Fischer, G., Horsefield, R., Katona, G., Nyblom, M., **Öberg, F.**, Young, G., Cogdell, R. J., Fraser, N. J., Engström, S. and Neutze, R. (2008) A lipidic-sponge phase screen for membrane protein crystallization. *Structure*, 16, 1003-9.

CONTRIBUTION REPORT

There are multiple authors on the papers presented here and my contribution to each of them is listed below. The focus of my thesis is on areas where I have made major contributions.

Paper I. I was involved in planning the project and cloned the tagged construct, screened for protein production, and performed optimization and production experiments. I was involved in data processing and analysis as well as manuscript preparation.

Paper II. I was involved in planning the project and cloned the constructs for production in *Pichia pastoris*. I was involved in data analysis, figure preparations, and writing of the manuscript.

Paper III. I planned the project and was responsible for cloning the constructs, protein production, quantitation, and localization studies. I took a major part in interpretation of the results, preparing figures, and writing of the manuscript.

Paper IV. I was involved in planning the project and was responsible for designing the cloning of the constructs, transformation, protein production, quantitation, Zeocin screens, and preparing figures. I took part in the localization studies with GFP, interpretation of the results, and writing of the manuscript.

Paper V. I planned the project and was responsible for cloning the constructs, producing and purifying the protein, glycosylation studies, circular dichroism, crystallization, and functional studies using stopped-flow. I took a major part in interpretation of the results, preparing figures, and writing of the manuscript.

Paper VI. I planned the project and was responsible for cloning the constructs, producing and purifying the protein, crystallization, collecting diffraction data, structure determination, structure refinement, and figure preparation. I took part in the interpretation of the structure and writing of the manuscript.

CONTENTS

ABBREVIATIONS AND SYMBOLS

AOX	Alcohol Oxidase
AQP	Aquaporin
AQPap	AQP adipose; early name for AQP7
ar/R	Aromatic/Arginine constriction region
AVP	Arginine-Vasopressin
Å	Ångström (10^{-10}m or 0.1nm)
CAI	Codon Adaptation Index
CD	Circular Dichroism
CHIP28	Channel-Like Integral Protein of 28kDa; initial name for AQP1
DDM	n-dodecyl-β-D-maltopyranoside
DLS	Dynamic Light Scattering
DM	n-decyl-β-D-maltopyranoside
E. coli	*Escherichia coli*
ER	Endoplasmic Reticulum
Glc/GlcNAc	Glucose/N-acetylglucosamine
GLIP	Glycerol Intrinsic Protein; early name for AQP3
H. Sapiens	*Homo Sapiens*
hAQPX	Human Aquaporin X
hKID	Human Kidney Aquaporin; early name for AQP6
Man	Mannose
MIP	Major Intrinsic Protein; early name for AQP0 and the aquaporins
MIWC	Mercurial Insensitive Water Channel; initial name for AQP4
NG	n-nonyl-β-D-glucopyranoside
NPA	Asparagine-Proline-Alanine signature motif of aquaporins
OG	n-octyl-β-D-glucopyranoside
P. falciparum	*Plasmodium falciparum*
P. pastoris	*Pichia pastoris*
P_f	Osmotic water permeability
PfAQP	Aquaporin from the parasite *Plasmodium falciparum*
PKA/PKC	Protein Kinase A/Protein Kinase C
PM28A	Plasma Membrane Protein of 28kDa; initial name for SoPIP2;1
S. cerevisiae	*Saccharomyces cerevisiae*
SCP	Single Cell Protein
SoPIP2;1	Plasma Intrinsic Protein of subgroup 2 from *Spinacia oleracea*
TMD	Transmembrane domain
WCH3	Water Channel 3; early name for AQP6
WCH-CD	Water Channel of Collecting Duct; initial name for AQP2

PREFACE

In 1992 Peter Agre and co-workers discovered the long sought molecular channel responsible for facilitated flux of water across cellular membranes. This family of proteins was later named aquaporins and they will be of central importance in this thesis. Today more than 5000 research articles have been published on the relevance of aquaporins and numerous groups throughout the world have turned their focus towards them. A milestone event was in 2003 when Peter Agre was awarded the Nobel Prize in chemistry, which confirms the importance of aquaporins.

The following year I started my research career by a master project aiming to produce these aquaporins in yeast. Eventually, I ended up as a PhD student with the aim to take a doctorate degree. Inevitably, the Day is coming closer and I here present a summary of what I have done during the almost seven years that have passed since I first started in the lab. It has required substantial work, with many failed and unclear experiments on the way. Nevertheless, at least in retrospect, my five years as a PhD student have been thrilling and loads of fun. And as Albert Einstein supposedly said *"Anyone who has never made a mistake has never tried anything new"*. That captures the very essence of why I like research; to investigate questions that have never been answered before, thus providing new solutions – and problems.

My thesis is an attempt to summarize all my work and put it together as one story. It is not only a story about aquaporins, but also about how to obtain the purified protein in sufficient amounts and how you can use it when you have achieved this initial goal. The journey will start with a background on the very basics of proteins and quickly move on to the essentials of membrane proteins and why we want to study them. Continuing with a description of how to study them and what we might find in doing so, we splash into the intriguing aquaporins and their whereabouts in our bodies. With the text in the introduction as support, I formulate the aims of my thesis followed by two major chapters including the actual results and discussion. The chapters are entitled 'Production optimization' and 'Characterization', respectively, and their names represent what this thesis is all about; describing ways to obtain enough protein material of interesting targets and what analyses we can do once we have sufficient amount.

To conclude, I have very much liked putting this book together, and I hope you will find it interesting to read.

Enjoy!

1 INTRODUCTION

The human body consists of as many cells as there are stars in our galaxy. Each of these cells has its own protective lipid bilayer membrane. It is hydrophobic and generally not permeable to solutes, thus allowing the cell to maintain necessary substances at the precise concentrations for the biochemical reactions to occur. Small and neutral molecules, such as water, can passively diffuse through the cellular membrane. However, for larger and more rapid changes, channels and pumps are vital. Transport along the concentration gradient, through a channel, requires no energy and is referred to as passive transport or facilitated diffusion. On the other hand, when molecules are pumped against the gradient, active transporters must utilize the cellular energy to mediate the transport.

1.1 PROTEINS – UNIQUE STRUCTURAL ELEMENTS

These essential membrane channels and pumps are examples of proteins, the most versatile macromolecules of the cell. Fundamentally, proteins are constructed from amino acids and the chemical properties of their side chains give them functional diversity and versatility. Amino acids are bound together by peptide bonds to form long polypeptide chains. These polypeptides first fold into a secondary structure of α-helices and β-sheets which secondly create more advanced tertiary and quaternary structures. The exact structure is unique for each protein, giving exclusive properties and functions. Hence, determining the protein structure will aid the process of understanding its function and detailed molecular mechanisms.

Proteins were first described in 1838 by the Dutch chemist Gerhardus Johannes Mulder and named by the Swedish chemist Jöns Jakob Berzelius (Hartley, 1951). The first protein to have its complete amino acid sequence determined was insulin; the work was lead by Frederick Sanger, who received the Nobel Prize for this achievement in 1958 (Nobelprize.org, 2011a). The same year, the first three dimensional structure of a protein was published by John Kendrew showing a low resolution structure of myoglobin (Kendrew *et al.*, 1958), for which he was awarded the Nobel Prize in 1962 (Nobelprize.org, 2011b). This breakthrough was a key step towards a deeper knowledge in proteins and their function.

1.2 MEMBRANE PROTEINS – FUNDAMENTAL MOLECULES OF LIFE

Membrane proteins (Figure 1.1), such as the channels and pumps described above, play many vital roles in all cellular life, including the human body. For example, they are involved in transport across the membrane, signal transduction, cell-cell interactions, and controlling the shape of the organelles within the cell (von Heijne, 2007). The number of proteins in a plasma membrane varies between cell types, but typically more than half of the membrane mass constitutes membrane proteins (Bretscher *et al.*, 1975). Approximately 26% of all proteins in the human genome have been predicted to be membrane proteins (Fagerberg *et al.*, 2010).

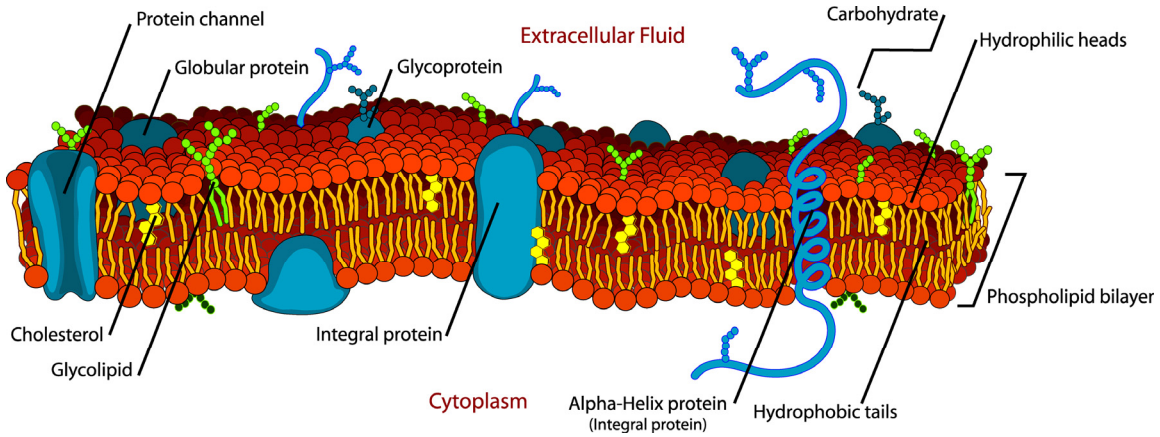


Figure 1.1. Schematic representation of the cell membrane with lipids and proteins, including integral proteins functioning as protein channels.

Membrane proteins constitute a majority of all drug targets (Lundstrom, 2006). Thus, the importance of this class of molecules is of great relevance for both academia and pharmaceutical industry. Despite their crucial function in the cell, they are generally poorly characterized. Hence, a better understanding of their properties and functions is essential to gain deeper insight into the action of existing drugs as well as aiding the development of new pharmaceuticals. Such understanding could be achieved by unifying biochemical and functional data with protein structure determination. However, membrane proteins are still dramatically underrepresented in structural databases (White, 2011). This could partly be attributed to the more complex and difficult procedure of producing and purifying them in sufficient amounts allowing detailed characterization. Furthermore, a majority of medically important membrane protein targets are present at very low concentrations in their native membranes (Mus-Veteau, 2002), requesting several strategies for recombinant overproduction. Today, the main bottleneck for structural determination and characterization of a membrane protein is the task of overproducing a stable and functional protein in sufficient amounts (Forstner *et al.*, 2007, Grisshammer *et al.*, 1995). Developing effective strategies for producing recombinant eukaryotic membrane proteins remains a particular challenge.

1.3 Recombinant Production of Membrane Proteins

Membrane protein overproduction is often a matter of a trial-and-error exercise (Grisshammer, 2006). Interestingly, it has historically even been considered an art instead of science (Bonander *et al.*, 2009). This clearly indicates the lack of knowledge and methods to overcome the problems associated with producing the protein of interest. Moreover, eukaryotic membrane proteins are known to be even more difficult to produce relative to their prokaryotic counterparts (Tate, 2001, Grisshammer, 2006).

Unravelling the collection of unknown factors in the process from gene transcription to a fully functional protein would be beneficial for controlling future overproduction experiments. Although the picture is still incomplete, important findings have been

presented including how the structure and sequence of mRNA affects translation initiation (Kozak, 1991, Kozak, 1992) and determinants for protein segments insertion into the membrane (Hessa *et al.*, 2005, Hessa *et al.*, 2007).

Especially for membrane protein overproduction, the host's folding machinery could be overloaded due to the strong promoters and multicopy vectors commonly used, resulting in an increased number of proteins passing through the endoplasmic reticulum (ER). This could become a potential problem as it may initiate an unfolded protein response in the host. Tuning the protein production level could thereby increase the final yield, as exemplified by the P2 adenosine transporter (Griffith *et al.*, 2003).

As a consequence of the difficulties related to the production of membrane protein targets, several ways of circumventing overproduction exists. Extracting large quantities of protein from naturally abundant sources has so far been quite successful (Figure 1.2). However, it limits the selection of targets, especially of human origin. Consequently, for future studies, this method has to be replaced by recombinant overproduction.

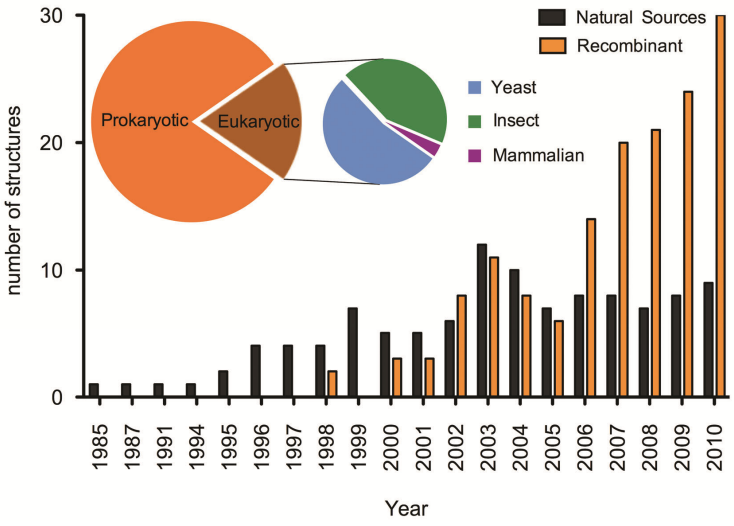


Figure 1.2. Bar chart showing the number of membrane protein structures determined over the last 25 years. The protein source has been indicated with black bars for natural sources and orange bars for recombinant protein production. Recombinant production is performed in prokaryotic or eukaryotic cell systems, where for the latter yeast-, insect- and mammalian cells are the most common. Reprinted with permission from Simon Newstead and Nature Publishing Group (Bill *et al.*, 2011).

Another common strategy is to try multiple initial constructs including different truncated forms and/or several homologous genes. By testing a vast number of genes, you are likely to find at least one which results in sufficient protein yield. This 'solution' is utilized at the Structural Genomics Consortium. However, judging from data available at their homepage (Structural Genomics Consortium, 2009), the success rates

are still fairly low. Of all targets, only 1/3 passed the cloning, overproduction and purification phase and about half of these targets had their structure determined. For organisations focusing only on membrane proteins, such as the European Membrane Protein Consortium, the success rates are even lower; less than 1/5 were produced in quantities enabling successful purification (European Membrane Protein Consortium, 2009). Even though the numbers will probably increase as these projects progress, achieving high yield for eukaryotic membrane protein overproduction is clearly the major bottleneck.

The recombinant overproduction host of choice depends on the protein target. During the last one and a half decades, 80% of all recombinantly produced proteins where produced in the bacterium *Escherichia coli* or the yeast *Pichia pastoris* (Sorensen, 2010). For bacterial targets *E. coli* is the obvious system of choice, being a very simple and cheap system. However, if the protein requires posttranslational modifications such as glycosylation (Chapter 1.5), a eukaryotic system is necessary. Popular eukaryotic hosts are yeast, including *Saccharomyces cerevisiae* or *P. pastoris*. They are almost as easy to handle as the bacterial systems but have higher success rates in producing eukaryotic proteins. Nevertheless, for production of human membrane proteins there is sometimes a need for systems having a more human-like cells, such as insect cells or mammalian cells. These are, however, often slower and more expensive to use. To date, the most successful host for overproduction of eukaryotic membrane proteins for structural determination is yeast, followed by insect cells and mammalian cells (Figure 1.2). In particular, overproduction from the yeast *P. pastoris* has yielded several high resolution structures of aquaporins (Fischer *et al.*, 2009, Horsefield *et al.*, 2008, Tornroth-Horsefield *et al.*, 2006).

In conclusion, a shift towards more targeted approaches for recombinant membrane protein production is desirable. This includes finding the optimal host for the protein target and utilizing all available knowledge to optimize both the host and the protein. Several possible host organisms are today available for overproduction of eukaryotic membrane proteins – ranging from the simplest bacterial systems to human cell lines. However, based on the previous successes, *P. pastoris* is one of the most suitable hosts for membrane protein production.

1.4 *PICHIA PASTORIS* AS PRODUCTION HOST

Apart from the posttranslational modifications, there are other reasons to avoid a prokaryotic host for eukaryotic membrane proteins; the translation rate and the translocon and lipid composition differs, which in combination could have impact on the final yield (Tate, 2001, Tate *et al.*, 2003). *P. pastoris* is a commonly used host due to the strong and tightly regulated promoter used to drive recombinant protein production, the similarity in manipulation techniques to those used for *S. cerevisiae*, the preferred respiratory growth allowing high cell density cultures to grow, and its record as a successful host for overproduction for structural characterization. Moreover,

stable transformants from linearized vector DNA can be made through homologous recombination between the vector and the host genome making stable strains which can grow without selection pressure.

1.4.1 FROM DISCOVERY TO CURRENT USE

Zygosaccharomyces Pastori was first isolated by Alexandre Guilliermond from the secretion of a chestnut tree (Guilliermond *et al.*, 1920). This initial classification was subsequently revised and the yeast got the new name *P. pastoris* (Phaff, 1956, Phaff *et al.*, 1956).

The interest for this particular yeast started in the petroleum industry during the sixties, with the intention to use waste products as nutrition for growing microorganisms in the production of single cell protein (SCP) (Cregg *et al.*, 2000). The usage of a single cell to produce protein was a potential solution to the anticipated food shortage in the world (Israelidis, 2004). A convenient carbon source was methanol readily available from natural gas. As *P. pastoris* was found in a small subset of yeast strains capable to assimilate methanol, it was attractive to use this specie in the SCP process (Hazeu *et al.*, 1972).

Phillips Petroleum Company performed extensive work on methanol assimilating yeast SCP and a high cell density process was developed (Wegner, 1983). However, as the price of methane increased during the second oil crisis in the U.S., the methods of producing SCP by the utilization of methanol was never profitable. Hence, another application was emerging; to use *P. pastoris* as a system for heterologous protein production (Cregg *et al.*, 1987, Wegner, 1990). Vectors, strains and methods were eventually sold to Research Corporation Technologies who made it available to academic users through Invitrogen.

1.4.2 STRENGTH IN PROTEIN PRODUCTION

The tightly regulated alcohol oxidase promoter is exploited in the *P. pastoris* system to drive expression of the gene of interest when induced by methanol. The different methylotrophic yeast genera (*Hansenula*, *Pichia*, *Candida*, and *Torulopsis*) all have the same pathway for methanol metabolism involving several unique enzymes (Veenhuis *et al.*, 1983). The metabolic process starts with alcohol oxidase (AOX) which catalyzes the oxidation of methanol (CH_3OH) to formaldehyde (CH_2O), also giving hydrogen peroxide (H_2O_2) from oxygen (O_2). The enzyme AOX has poor affinity for oxygen which is compensated by the production of excessive quantities of the protein. In wild type *P. pastoris* it can constitute over 30% of the total protein content in a cell (Couderc *et al.*, 1980). Thus, by inserting the gene of interest under the control of the AOX promoter, high yields of recombinant protein can be obtained.

A common strategy for overproduction of soluble proteins in *P. pastoris* has been to increase the number of gene copies inserted into the genome (Scorer *et al.*, 1994). However, this has far from always resulted in a higher yield and it has been observed that using high copy number vectors could actually result in less protein (Cregg, 2007).

A too strong methanol induction of the promoter might instead cause an overload of the endoplasmic reticulum (ER) machinery. This could lead to large amount of incorrectly folded protein or protein lacking posttranslational modifications needed for proper function. Instead, slowing the rate of translation could be beneficial for production of functional proteins (Griffith *et al.*, 2003) including aquaporins (Bonander *et al.*, 2005).

The host *P. pastoris* has been a vital part of the pipeline leading up to structure determination and lately its popularity has increased. Thus, since more researchers will use the system in the future, understanding how to achieve high yields is of vital importance. In the work presented in this thesis, the use of *P. pastoris* as a host for high eukaryotic membrane protein production yields has been further analysed and developed.

1.5 PROTEIN GLYCOSYLATION - MODIFICATION OF PROTEINS

For selection of overproduction system, the host's ability to perform certain protein modifications must match the needs of the protein of interest. The most common of all such modifications, protein glycosylation, could have a huge impact on the newly synthesized protein, with respect to function and stability (Lis *et al.*, 1993).

Mammalian membrane proteins are synthesized in the rough endoplasmic reticulum membrane where they are cotranslationally modified with the addition of carbohydrates, also named glycans (Kornfeld *et al.*, 1985). From analysis of the protein sequence data bank, glycosylation occurs on the majority of all proteins (Apweiler *et al.*, 1999). For proteins passing through the eukaryotic secretory pathway, including membrane proteins, almost all will be glycosylated (Alberts, 2002, Lodish, 2000).

Glycosylations do not occur in *E. coli* (Lis *et al.*, 1993) but have been found in other bacteria. *Campylobacter jejuni* was the first bacteria in which the glycosylation pathway was described (Nothaft *et al.*, 2010, Szymanski *et al.*, 2005, Szymanski *et al.*, 1999). It share functions with eukaryotic systems, but the structures of the glycans are strikingly different (Messner, 2004, Weerapana *et al.*, 2006).

There are two major types of protein-carbohydrate linkages found in eukaryotic glycoproteins. The glycan is attached covalently to a nitrogen or oxygen, hence they are referred to as N-linked and O-linked glycosylation, respectively (Lodish, 2000). For O-linked glycans attached to a serine or threonine residue, no consensus sequences have been found. Instead, the secondary and tertiary structure of the protein specifies the glycosylation sites, thus making them difficult to predict (Voet *et al.*, 2004). Furthermore, these glycans are less easily classified and fewer generalizations can be drawn as compared to the N-linked glycans (Bill *et al.*, 1998). O-linked oligosaccharides are generally not found in proteins produced in *P. pastoris*, and very few of the secreted proteins in this yeast are found to be O-linked (Grinna *et al.*, 1989, Invitrogen, 2010)

1.5.1 N-LINKED GLYCOSYLATION

N-linked glycoproteins are formed in the ER and are processed to the final structure in the Golgi apparatus. The oligosaccharide to be attached is initially synthesized as a lipid-linked precursor. The lipid, which is dolichol, anchors the growing oligosaccharide to the ER membrane until the complete core structure has been synthesized. It has the structure $Glc_3Man_9GlcNAc_2$ as shown in Figure 1.3 (Voet *et al.*, 2004). The whole precursor is subsequently attached, cotranslationally, to the asparagine residue at the recognition sequence, the "sequon", Asn-X-Ser/Thr where X can be any amino acid with the exception of proline (Marshall, 1974).

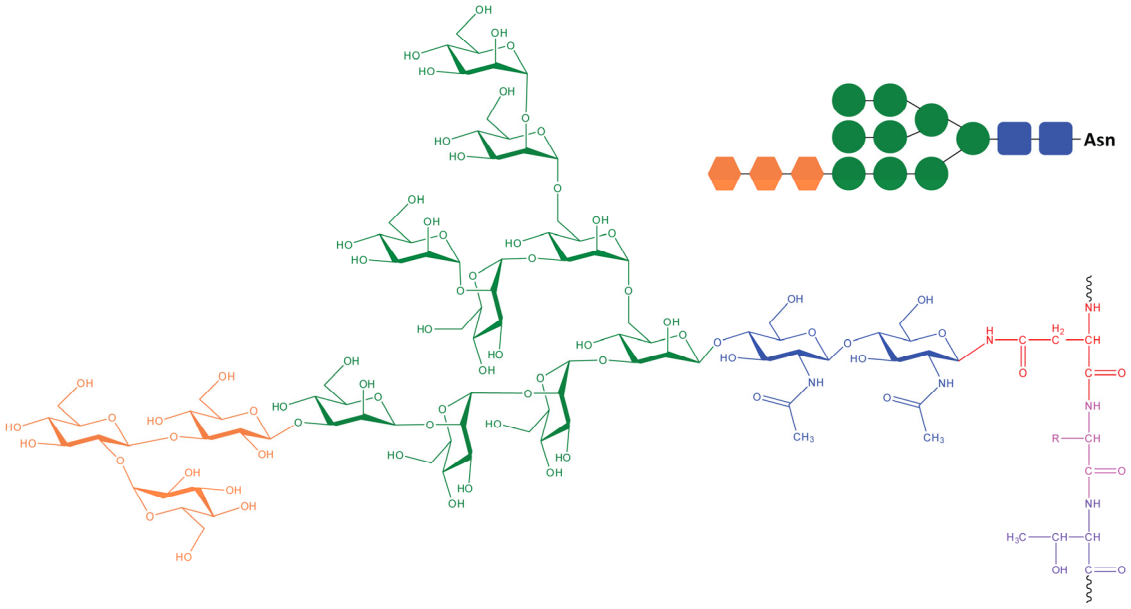


Figure 1.3. Chemical structure of an N-linked glycan attached to the asparagine in the sequon. A schematic image of the same structure is inserted. Shown in blue: N-acetyl-glucosamine (GlcNAc), in green: Mannose (Man), in orange: Glucose (Glc).

Processing of the precursor glycan starts already in the ER lumen, before the protein has been fully synthesized and folded. First, the three glucose residues are trimmed from the glycan (Figure 1.4A) (Atkinson *et al.*, 1984). Subsequently the protein is transported to the Golgi apparatus where the Golgi stacks contain different sets of processing enzymes (Jamieson *et al.*, 1968). Mannose residues are trimmed and acetylglucosamine (GlcNAc), galactose, fucose and/or sialic acid residues are added or removed from the complex. The inner core of the glycan, $Man_3GlcNAc_2$, remains intact and is the common structure for all glycoproteins (Kornfeld *et al.*, 1985). The protein is eventually leaving the Golgi network for transport to the cellular destination it is destined for (Alberts, 2002).

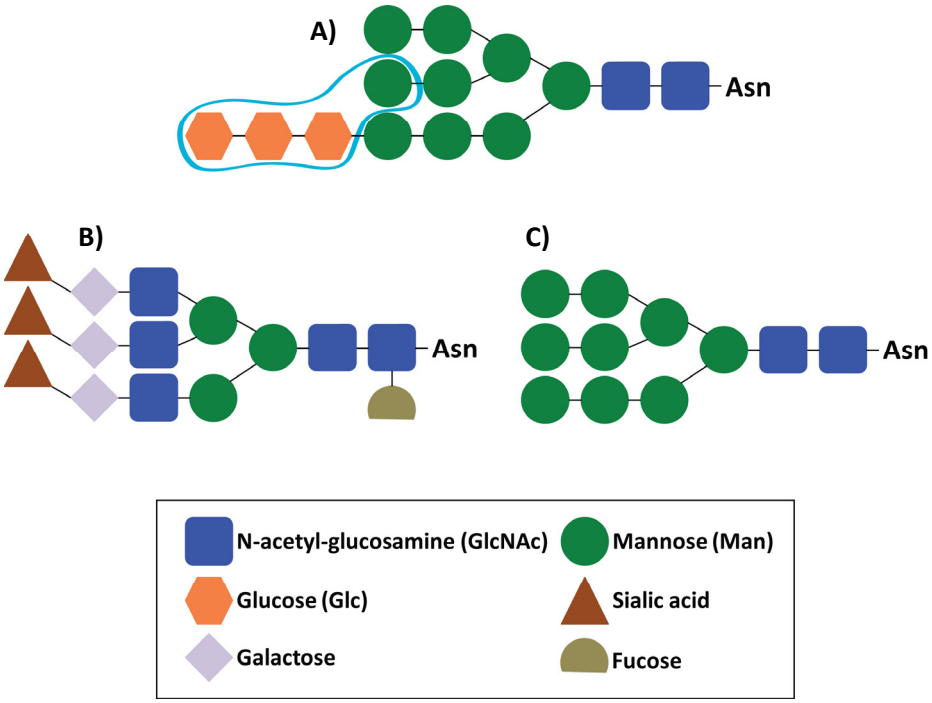


Figure 1.4. Schematic illustrations of the processing of N-linked glycans. A) Showing the precursor attached to an asparagine. The initial trimming in the ER results in removal of glucose and mannose residues, which have been circumscribed in blue. After the final trimming in the Golgi apparatus, two different classes of N-linked glycans are generally identified; B) complex oligosaccharides, and C) high-mannose oligosaccharides.

Two general classes of N-linked oligosaccharides can be identified after Golgi apparatus processing: complex (Figure 1.4B) and high-mannose oligosaccharides (Figure 1.4C) (Alberts, 2002). Complex oligosaccharides contain a number of acetylglucosamine-galactose-sialic acid units attached to the common core structure, even though they are frequently processed and truncated. In addition, fucose residues could be attached. In the case of high-mannose oligosaccharides, not all of the initial mannoses are truncated and additional mannoses could be attached. Hybrid oligosaccharides, with different structure on the different branches of the sugar tree, can also be found. Which type of processing each oligosaccharide receive depends largely on its position within the protein; if the glycan is inaccessible it is likely to remain as a high-mannose oligosaccharide whereas a glycan accessible to the Golgi apparatus is more likely to be processed into a more complex form (Alberts, 2002).

The vast number of possible ways to build up a mammalian glycan makes glycosylation the most diverse of all post-translational modifications. Two different protein molecules produced in the same cell and exposed to the same enzymes and glycan machinery can differ in glycan structure (Daly *et al.*, 2005). In addition to this difference, which is referred to as glycoform microheterogeneity, the occupancy of potential glycosylation sites can vary between proteins, called glycoform macroheterogeneity (Harmon *et al.*, 1996). However, the microheterogeneity in glycan structure does not necessary lead to an altered effect of a protein.

1.5.2 N-LINKED GLYCOSYLATION IN *PICHIA PASTORIS*

The initial glycosylation processes in the ER are similar in both yeast and mammals, but the processing in the Golgi apparatus is vastly different. Although there are some differences between yeast species, they generally do not trim the glycan down to three mannoses. Instead they have a core composed of eight or nine mannoses. Further mannose and galactose residues can be attached (Gemmill *et al.*, 1999). Moreover, yeast does not add sialic acids to their glycans, but they instead add other negative charges with unknown function to their N-glycans (Gemmill *et al.*, 1999).

In *S. cerevisiae*, hyperglycosylation is common. During the passage through the Golgi apparatus a large number of mannose residues are attached, often giving a final glycan with more than 100 mannose residues (Hamilton *et al.*, 2003). In *P. pastoris*, on the other hand, the elongated N-glycans are shorter and the final length of the oligosaccharide chains normally range between 8 and 14 mannose residues. (Grinna *et al.*, 1989, Trimble *et al.*, 1991).

1.5.3 BIOLOGICAL RELEVANCE OF N-LINKED GLYCOSYLATION

Although glycosylations have been thoroughly studies for many decades, no unambiguous purpose of the modification has yet been presented and the specific role of glycans remains uncertain. Considering the frequency and complexity of glycan attachments, it is easy to assume that they conduct important functions in the cell. Blocking glycosylation or glycan processing using inhibitors or mutants is commonly done to study the effect of a glycan (Gong *et al.*, 2002, Nakagawa *et al.*, 2009, Zhou *et al.*, 2005). In several cases, the protein is not affected whereas other studies have shown a decrease in protein function, amount, and/or stability (Elbein, 1991).

One suggested glycan function is the sorting of membrane proteins in polarized cells. It occurs in the *trans* Golgi network where the proteins incorporated into vesicles are destined for different domains of the cell membrane. A typical epithelial cell has an apical domain facing the lumen, with features such as cilia or microvilli, and a basolateral domain, covering the rest of the cell. For the two domains to remain different, tight junctions separate the membrane preventing proteins and lipids to diffuse between them. Membrane proteins targeted to the basolateral domain have signal sequences, for example a critical tyrosine residue near a large hydrophobic amino acid or a leucine based motif (Keller *et al.*, 1997). Targeting to the apical membrane is fundamentally different and no signal sequence exists. Instead GPI anchor (Lisanti *et al.*, 1990), specific membrane spanning regions (Kundu *et al.*, 1996), and N-glycans (Scheiffele *et al.*, 1995, Keller *et al.*, 1997) target the protein to the apical domain. Basolateral targeting signal is likely dominant since glycoproteins containing a basolateral signal sequence will end up in the basolateral domain (Simons *et al.*, 1997). As an example, the N-linked glycans have been shown to be essential for proper trafficking of human organic anion transporter to the plasma membrane (Zhou *et al.*, 2005).

Glycosylation has also been implicated in protein folding. The chaperon calnexin is aiding proteins to fold, *via* untrimmed glucose residues in the glycan. Upon reaching the mature fold, the glycan is trimmed and the protein can leave ER (Bergeron *et al.*, 1994).

Less specific functions based on the general chemical properties of the glycan have also been suggested, such as stabilizing the protein to avoid denaturation or proteolysis (West, 1986, Nakagawa *et al.*, 2009, Waetzig *et al.*, 2010). Thermodynamic stabilization calculations have been made and, in principle, the covalent binding of a glycan to a protein surface may enhance the thermal stability of the protein. This can be measured as an increase in melting temperature. Interestingly, the length of each glycan had only a minor effect on the degree of stabilization (Shental-Bechor *et al.*, 2008). Moreover, chains of sugar have limited flexibility and even small N-linked glycans can protrude from a protein surface, thereby sterically hinder a protease or another protein from reaching the glycoprotein (Pletcher *et al.*, 1980).

Clearly, glycosylation cannot be neglected when overproducing proteins as they might alter the proteins function or stability. Keeping this in mind, the heterogeneity often occurring from glycosylation can influence the measurements on the protein, since all results will be dependent on multiple populations with potentially different properties. Hence, for some purposes, like crystallization, the removal of a glycan can be beneficial for protein behaviour.

1.6 WATER AND AQUAPORINS

Life, as we know it, occurs in water. Water is a unique substance and its properties make it the most central molecule in biology. These extraordinary solvent properties of water arise from the polarity and the many possibilities for hydrogen bonds, making it an ideal biological solvent. The earliest forms of life appeared in an aqueous solution some 4 billion years ago, and water is a major component of all organisms living today. About 2/3 of the mass in the human body is water. Regulation of the fluid content within a cell is fundamental and the water flux across biological membranes is mediated by intrinsic membrane proteins acting as water channels. They have been named aquaporins (AQPs) and they are widespread throughout nature: from bacteria and yeast to plants and animals.

1.6.1 DISCOVERY OF AQUAPORINS

The discovery of the plasma membrane in the 1920s started the discussion on how water can be transported across this membrane. Initially it was believed to be by passive diffusion, but several studies found that cellular membranes had a higher permeability than could be explained by diffusion alone (Agre, 2006). The presence of water-filled channels in the membrane of red blood cells was first indicated in 1957 (Paganelli *et al.*, 1957) and later evidence for the existence of such channels arose when the water permeability could be inhibited by $HgCl_2$ (Macey, 1984). Since this compound could not inhibit diffusion of water across the membrane itself, a channel

protein sensitive to mercury was the logical conclusion. In the same period of time, several other groups were also predicting water channel proteins in red blood cells (Benga *et al.*, 1986, Brown *et al.*, 1975).

The first water channel was identified during an isolation of a Rhesus blood group antigen from red blood cells. The protein, expected to be around 32kDa, was poorly stained with coomassie and when silver stain was used, an unexpected protein of 28kDa was also detected (Agre *et al.*, 1987). This was initially thought to be a degradation product of the larger protein, but rabbits immunized with the 28kDa protein developed an immune response specific for the smaller fragment (Denker *et al.*, 1988). In addition, the antibodies could also detect a broader high molecular weight signal, an N-linked glycoprotein. Further studies (Smith *et al.*, 1991) characterized the new protein as a membrane channel with several similarities to MIP, the major intrinsic protein of eye lens which at the time was a putative membrane channel protein with undefined function (Gorin *et al.*, 1984). The evidence that channel-forming integral protein of 28kDa (CHIP28) was transporting water came in 1992 when Peter Agre and co-workers injected *Xenopus laevis* oocytes with CHIP28 RNA. The oocytes showed increased osmotic water permeability (P_f) as compared to control oocytes (Preston *et al.*, 1992). This effect was reversibly inhibited by mercury, and later cysteine 189 was found to be the mercury binding and inhibition site (Preston *et al.*, 1993). As more proteins transporting water were discovered, the common name of 'aquaporins' was introduced (Agre *et al.*, 1993) and CHIP28 was renamed aquaporin 1 (AQP1) (Agre, 1997).

1.6.2 STRUCTURAL FEATURES

Several high resolution structures have revealed common features of the aquaporins. As predicted in the very first studies of aquaporins, they have six α-helical transmembrane domains (TMDs), denoted 1-6, and five connecting loops, denoted A-E. The termini are located on the intracellular side. Loop B and E both contain the signature motif for aquaporins: a highly conserved asparagine-proline-alanine (NPA) sequence. These two motifs meet in the middle of the bilayer (Jung *et al.*, 1994b) (Figure 1.5A). The protein thus consists of two similar halves, arising from gene duplication of three TMD segments (Pao *et al.*, 1991). In the native membrane, aquaporins are assembled into homotetramers (Smith *et al.*, 1991), where each subunit is a functional unit and contains a single water channel. This creates a central pore in the middle of the four monomers. Whether or not this channel has any biological relevance is still debated.

While facilitating water transport, aquaporins must be able to exclude protons as a proton leakage would destroy the proton motive force that is essential for the cell and used to drive energy production. In bulk water, protons are conducted by the Grotthuss mechanism. They are transferred between oriented water molecules *via* the hydrogen bonds. In the NPA motif in aquaporins, two positively charged asparagines

have been suggested to be the key for proton exclusion (de Groot *et al.*, 2001) (Figure 1.5B). The electrostatic field around the NPA motif cases a strict orientation of water molecules resulting in an interruption of the proton wire created by the Grotthuss mechanism. However, it has also been suggested that protons are directly excluded by the positive electrostatic field and its repulsion on small positively charged particles (de Groot *et al.*, 2005).

The narrowest part of the pore is the selectivity filter, an aromatic/arginine (ar/R) constriction region formed by four amino acids (Figure 1.5B). It is located in the proximity of the extracellular entrance with a typical width of 2.8Å in aquaporins and 3.4Å in aquaglyceroporins (Walz *et al.*, 2009). In AQP1 the four amino acids are Phe58, His182, Cys191, and Arg197 with histidine being typical for water transporters.

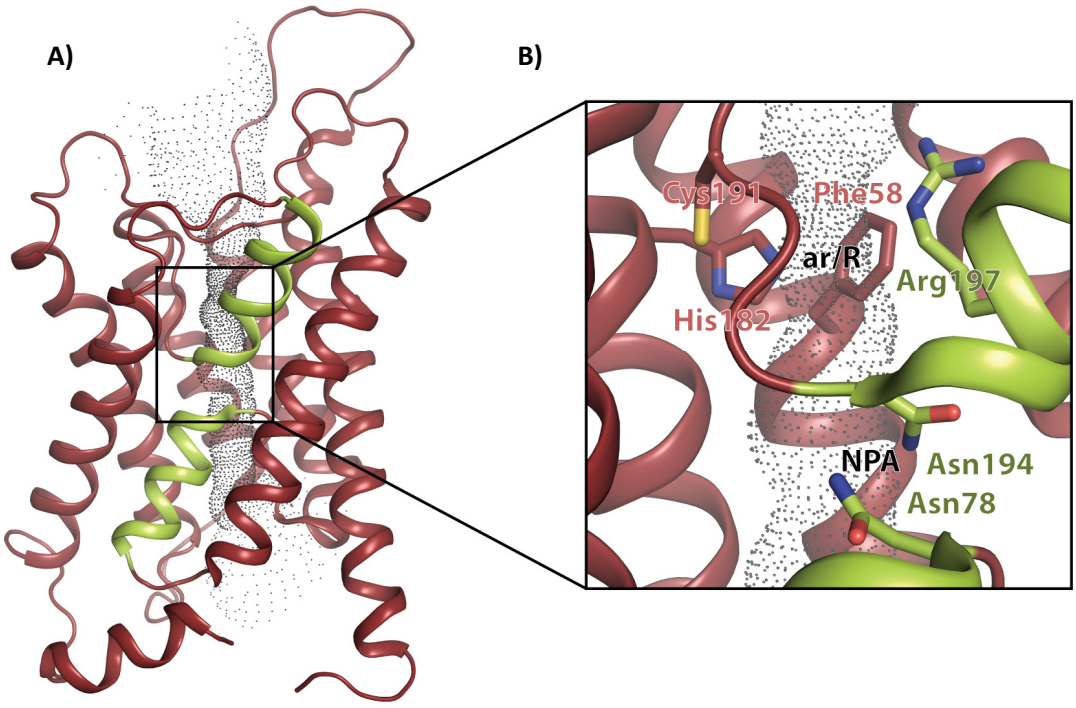


Figure 1.5. The general structure of aquaporins exemplified with AQP1 (Sui *et al.*, 2001). A) The structure shows six transmembrane domains (red) and two half helices (green). The water pore is contoured in black. B) Detailed view of the aromatic/arginine region and the four amino acids constricting the region. The two asparagines in the NPA motif are also shown.

Moreover, structural changes within some aquaporins can cause the water channel to be opened or closed. This mechanism is referred to as gating and is well established for a plant aquaporin (Chapter 1.10.1). However, the gating of mammalian aquaporins has been much more controversial and no structural features have been able to verify its existence (Tornroth-Horsefield *et al.*, 2010). Instead, the human aquaporins are regulated by trafficking from intracellular vesicles to the plasma membrane. This provides a rapid first response to changes in the biological environment. Trafficking

was first indentified for AQP2 but has also been implied for other aquaporins: AQP1, AQP5, and AQP8. The trafficking of AQP5 is further discussed in Chapter 1.8.2.

1.7 HUMAN AQUAPORINS

In humans, 13 aquaporin homologues have been identified. They are commonly divided into two subgroups: the orthodox aquaporins (AQP0, AQP1, AQP2, AQP4, AQP5, AQP6, AQP8), mainly transporting water; and the aquaglyceroporins (AQP3, AQP7, AQP9, AQP10), transporting water and glycerol. The two remaining aquaporins (AQP11, AQP12) still have undetermined transport specificity and are usually placed in their own group, sometimes referred to as superaquaporins. Table 1.1 provides a short summary followed by a brief introduction of all the human aquaporins. Two of them, the orthodox aquaporin 5 and aquaglyceroporin 10, are described in more detail since they are directly relevant to the studies described in this thesis.

Protein	Found	Substrate	Major tissue distribution
AQP0	1984	water	eye lens
AQP1	1987	water	red blood cells, kidney, lung, brain, eye, vascular endothelium
AQP2	1993	water	kidney
AQP3	1994	water, glycerol, urea	skin, kidney, lung, eye, small intestine, colon
AQP4	1994	water	brain, kidney, lung, muscle, stomach
AQP5	1995	water	Salivary-, lacrimal- and sweat gland, lung, eye.
AQP6	1993	water, anions	kidney
AQP7	1997	water, glycerol, urea, arsenite	adipose tissue, kidney, testis
AQP8	1997	water	Kidney, liver, pancreas, small intestine, testis, colon
AQP9	1998	water, glycerol, urea, small solutes, arsenite	liver, white blood cells, brain, testis
AQP10	2001	water, glycerol, urea	small intestine
AQP11	2000, 2005	water(?)	brain, liver, kidney
AQP12	2000, 2005	?	pancreas

Table 1.1 Table presenting substrate specificity and the major tissue distribution of all human aquaporins. For AQP11 and AQP12, two years of discovery are given due to some ambiguousness as further discussed in Chapter 1.7.10. The information in this table is collected from reviews: (Carbrey *et al.*, 2009, Castle, 2005, King *et al.*, 2004).

1.7.1 AQUAPORIN 0

The mRNA encoding AQP0 was first identified in 1984 (Gorin *et al.*, 1984) and it was believed to be an aqueous channel and/or a gap junctional protein. At this time it was referred to as MIP – major intrinsic protein of the lens. However, after the discovery of AQP1 and the development of the functional assays for water transporters, it was renamed to AQP0 (Agre, 1997). This channel transports water at a slower rate than that of AQP1 (Mulders *et al.*, 1995) and in addition to facilitating water, AQP0 plays a role in cell-to-cell adhesion of the lens fibre. Whether the junction or the water flux

properties plays the major role in maintaining the lens transparent is not unequivocally determined. However, the importance of AQP0 is clear; human individuals with mutations in AQP0 suffer from cataracts, a symptom ranging from cloudy vision to blindness (Berry *et al.*, 2000). High resolution structures have also been determined for AQP0: visualizing protein-lipid interactions, possible regulatory mechanisms, and the cell-cell adhesion role (Gonen *et al.*, 2005, Gonen *et al.*, 2004, Harries *et al.*, 2004).

1.7.2 AQUAPORIN 1

AQP1 was the first protein for which water transport was measured and today it is the most studied aquaporin. It was also the first aquaporin for which a high resolution structure was determined (Murata *et al.*, 2000), which together with the many biochemical studies has made AQP1 much of a model protein for future research. No clear gating mechanism has been identified, but alteration of osmotic conditions can induce a reversible protein kinase C (PKC) dependent change in the membrane localization of AQP1 (Conner *et al.*, 2010), suggesting a regulatory mechanism by trafficking. The protein has been found in many different tissues in the body, including red blood cells, kidneys, and lungs. Mice and humans lacking AQP1 have shown to have urinary concentration deficiency during water deprivation (Ma *et al.*, 1998, King *et al.*, 2001). AQP1 null mice are protected against lung edema in certain situations (Bai *et al.*, 1999) and the protein has been found in the aqueous producing epithelium in the eye and in the choroid plexus of the brain (Nielsen *et al.*, 1993). In both eye and brain, AQP1 is involved in water accumulation, but not in water efflux (Carbrey *et al.*, 2009) causing null mice to have a higher survival rate after targeted injury to the brain due to a lower intracranial pressure (Oshio *et al.*, 2005).

1.7.3 AQUAPORIN 2

AQP2 was identified shortly after the discovery of AQP1. It was found in the renal collecting duct and hence called water channel of collecting duct (WCH-CD) (Fushimi *et al.*, 1993). The trafficking of AQP2 is one of the most studied aquaporin regulation mechanisms. Vasopressin triggers cAMP signalling, leading to activation of protein kinase A which phosphorylates AQP2 resulting in translocation to the apical plasma membrane (Nedvetsky *et al.*, 2009). The result is a major increase in the water permeability of the collecting duct and hence absorption of water from the primary urine. When the levels of vasopressin drop, AQP2 will be endocytosed; a fraction of the protein is ubiquitinated and degraded, and the water permeability of the membrane is reduced back to the initial situation (Kamsteeg *et al.*, 2006). A mutation in AQP2 causes nephrogenic diabetes insipidus (Deen *et al.*, 1994) and mice with mutations in this gene show severe urine concentration defects (Yang *et al.*, 2001).

1.7.4 AQUAPORIN 3

AQP3 was first identified in the basolateral membrane of the collecting duct in the kidney. (Ma *et al.*, 1994, Ishibashi *et al.*, 1994, Echevarria *et al.*, 1994). It was named glycerol intrinsic protein (GLIP) or AQP3 and could in addition to water also transport

glycerol and urea. AQP3 is also abundant in keratinocytes in the basal layer of the epidermis in human skin (Sougrat *et al.*, 2002). Inhibition of AQP3 can be achieved by low pH and nickel (Zeuthen *et al.*, 1999, Zelenina *et al.*, 2003). However, the physiological relevance for this inhibition is not clear (Carbrey *et al.*, 2009). For unknown reason, AQP2 is down regulated in AQP3 null mice causing deficiency in urine concentration and nephrogenic diabetes insipidus (Ma *et al.*, 2000). The deletion of the gene in mice caused a reduction in skin elasticity, slower wound healing, reduced glycerol contents and dry skin (Hara *et al.*, 2002). As a consequence of finding a connection to dry skin, companies producing skin care productions have tried to make products stimulating AQP3 levels in the skin cell layers (Dumas *et al.*, 2007).

1.7.5 AQUAPORIN 4

AQP4 was first cloned from rat lung (Hasegawa *et al.*, 1994) and rat brain (Jung *et al.*, 1994a). It was given the name mercurial insensitive water channel (MIWC) due to the lack of mercury inhibition. The isoform identified in the brain was several amino acids longer and it transported water at higher rates. Subsequent publications found more possible versions of the gene (Yang *et al.*, 1995, Lu *et al.*, 1996). Today there are two recognized human isoforms; AQP4-M1 is the full length protein while the shorter, lacking the first 22 amino acids, is referred to as hAQP4-M23. From the first structure determination of AQP4, it was proposed that solely M23 can form orthogonal arrays, giving a plausible physiological role for the existence of two isoforms (Hiroaki *et al.*, 2006). Another role of AQP4 is to control the water balance in the brain and AQP4 null mice have a higher chance of survival after brain edema which could be a result from hyponatremia or external damage to the brain (Carbrey *et al.*, 2009). Moreover, the water flux through AQP4 is helping the rapid clearance of potassium ions thereby aiding the recovery after neuronal activation (Amiry-Moghaddam *et al.*, 2003). A high resolution structure of truncated hAQP4 has also been presented with some differences in the interaction with waters along the channel, as compared to other water-selective AQPs (Ho *et al.*, 2009).

1.7.6 AQUAPORIN 6

AQP6 was first cloned from rat kidney and was initially referred to as WCH3 (Ma *et al.*, 1993). The water permeability of the protein was relatively low, but could be increased by mercury – in contrast to the more commonly observed inhibitory function of this compound on aquaporins (Yasui *et al.*, 1999a). In addition, AQP6 was found to transport anions. A human AQP6 variant with slightly different sequence was also identified and referred to as hKID (Ma *et al.*, 1996). All forms were exclusively found in the kidney and the rat isoform was found to be modified by an N-linked glycosylation (Yasui *et al.*, 1999b). In contrast to other aquaporins located in the kidney, AQP6 was found to be located into intracellular vesicles, making it less likely to be involved in reabsorption water. Instead it could function as an acid-base regulator with pH being the activating mechanism (Yasui *et al.*, 1999a).

1.7.7 AQUAPORIN 7

AQP7 was first cloned from rat testis (Ishibashi *et al.*, 1997a) and was found to transport glycerol. However, in humans it was first detected in adipose tissue (Kuriyama *et al.*, 1997), giving it the initial name AQP adipose (AQPap). The role in this tissue is to provide the glycerol needed for gluconeogenesis (Maeda *et al.*, 2004). In addition, adult null mice have a severe increase in body fat mass indicating a role in obesity (Hara-Chikuma *et al.*, 2005). AQP7 has also been found to reabsorb glycerol in the kidney (Skowronski *et al.*, 2007) and have been detected in testis (Suzuki-Toyota *et al.*, 1999). Although controversial, AQP7 was not produced in the sperm of some infertile male patients suggesting this as a plausible reason for infertility (Saito *et al.*, 2004).

1.7.8 AQUAPORIN 8

AQP8 was found in different tissues: colon, placenta, liver, heart (Ma *et al.*, 1997), testis (Ishibashi *et al.*, 1997b), and pancreas (Koyama *et al.*, 1997). In rat liver cells, AQP8 was observed to be trafficked from intracellular vesicles to the plasma membrane in response to cAMP (Garcia *et al.*, 2001). The mechanism of trafficking is not as well studied as for AQP2, and the influence of possible phosphorylations is still unknown.

1.7.9 AQUAPORIN 9

AQP9 was first identified in human white blood cells, where it was found to transport water and urea but not glycerol (Ishibashi *et al.*, 1998). However, later studies have detected glycerol transport for this aquaglyceroporin, as well as a broad range of other solutes (Tsukaguchi *et al.*, 1998). Yeast cells producing mammalian AQP7 and AQP9 have been observed to conduct arsenic transport (Liu *et al.*, 2002). Roles of AQP9 includes facilitating the uptake of glycerol in the liver (Maeda *et al.*, 2009) and acting as a glucose metabolite channel in the brain (Badaut *et al.*, 2004).

1.7.10 AQUAPORIN 11

The molecular cloning of AQP11 and AQP12 is often attributed to a book chapter from 2000 where they are named AQPX1 and AQPX2 (Hohmann *et al.*, 2000). Nevertheless, they were further characterized and published about five years later. AQP11 null mice develop polycystic kidneys, a fatal symptom, but not clearly linked to aquaporin activity (Ishibashi, 2009). In immunohistochemical studies, AQP11 has been found in intracellular compartments of kidney proximal tubes (Morishita *et al.*, 2005). The ability of AQP11 to facilitate water has been assayed in both oocytes and reconstituted into liposomes. The results have been contradicting; the former assay showed no transport at all (Gorelick *et al.*, 2006), whereas full water activity could be observed in the artificial liposomes (Yakata *et al.*, 2007).

1.7.11 AQUAPORIN 12

AQP12 was found by searching for homologues to AQP11. The protein was localized intracellularly in pancreas and no functional assays were performed (Itoh *et al.*, 2005). From studies with null mice, no obvious differences were observed under normal

conditions, but differences could be seen upon inducing an inflammation in the pancreas (Ohta *et al.*, 2009). Nevertheless, the physiological relevance remains unclear.

1.8 THE ORTHODOX HUMAN AQUAPORIN 5

Aquaporin 5 is one of three human aquaporins with known structure (Horsefield *et al.*, 2008), the others being aquaporin 1 (Murata *et al.*, 2000) and aquaporin 4 (Ho *et al.*, 2009). Based on the high sequence identity with hAQP2 and findings of several studies described below, it is believed to be regulated by trafficking. However, the complete mechanism of this regulation has yet not been revealed.

1.8.1 DISCOVERY AND PHYSIOLOGICAL ROLE

AQP5 was first identified from a rat salivary gland and in the same study mRNA was detected in salivary glands, lacrimal glands, sweat glands, eyes and lungs (Raina *et al.*, 1995). In the lungs, AQP5 have been found in the secretory cells of the submucosal glands (Kreda *et al.*, 2001). These glands significantly contribute to the liquid film found in the airways (Ballard *et al.*, 1999) and AQP5-null mice have been found to have a reduced secretion from these glands (Song *et al.*, 2001). In the eye, AQP5-null mice have significantly thicker corneas giving them a reduced response rate to osmotic gradients (Thiagarajah *et al.*, 2002). From the sweat glands, a reduced secretion has been observed for AQP5-null mice (Nejsum *et al.*, 2002) but it had been contradicted by other work (Song *et al.*, 2002).

Human AQP5 was found in the apical membrane of salivary glands, but for patients with Sjögren's syndrome it was primarily located in the basal membranes (Tsubota *et al.*, 2001). This would imply defective hAQP5 trafficking, causing the dry mouth and dry eyes which is typical symptoms of patients suffering from Sjögren's syndrome. Moreover, AQP5 null mice have a major reduction in saliva production (Ma *et al.*, 1999). In contrast, there are reports indicating that the tear secretion is independent of any aquaporin (Moore *et al.*, 2000).

1.8.2 PROTEIN TRAFFICKING

AQP2 was the first aquaporin for which the subcellular localization was observed to change in response to stimuli. The process starts with arginine-vasopressin (AVP) binding to vasopressin V2-receptors in the basolateral membranes of the collecting duct (Sabolic *et al.*, 1995). This activates adenylate cyclase which results in an increase in the intracellular cAMP and subsequent activation of protein kinase A (PKA) (van Balkom *et al.*, 2002). Activated PKA is targeting the intracellular vesicles containing AQP2 and phosphorylates Ser256 in the carboxyl terminus of AQP2 (Fushimi *et al.*, 1997). The modified AQP2 is thereafter targeted to the apical plasma membrane where it increases the permeability of the collecting duct.

The main PKA consensus sequence in hAQP5 is Ser156 located in the D-loop. Additionally, Thr259 has been identified as another possible PKA site (Diegelmann *et al.*, 2006). As seen in the sequence alignment (Figure 1.6), Thr259 in hAQP5

corresponds to Ser256 in hAQP2 and could possibly be of importance for the regulatory trafficking of hAQP5.

```
CLUSTAL 2.1 multiple sequence alignment

hAQP2           -MWELRSIAFSRAVFAEFLATLLFVFFGLGSALNWPQALPSVLQIAMAFGLGIGTLVQAL 59
hAQP5           MKKEVCSVAFLKAVFAEFLATLIFVFFGLGSALKWPSALPTILQIALAFGLAIGTLAQAL 60
                   *: *:** :**********:**********:**.***::****:****.****.***

hAQP2           GHISGAHINPAVTVACLVGCHVSVLRAAFYVAAQLLGAVAGAALLHEITPADIRGDLAVN 119
hAQP5           GPVSGGHINPAITLALLVGNQISLLRAFFYVAAQLVGAIAGAGILYGVAPLNARGNLAVN 120
                * :**.*****:*:* *** ::*:*** *******:**:***.:*: ::* : **:****

hAQP2           ALSNSTTAGQAVTVELFLTLQLVLCIFASTDERRGENPGTPALSIGFSVALGHLLGIHYT 179
hAQP5           ALNNNTTQGQAMVVELILTFQLALCIFASTDSRRTSPVGSPALSIGLSVTLGHLVGIYFT 180
                **.*.** ***:.***:**:**.********.** .  *:******:**:****:**::*

hAQP2           GCSMNPARSLAPAVVTGKFD-DHWVFWIGPLVGAILGSLLYNYVLFPPAKSLSERLAVLK 238
hAQP5           GCSMNPARSFGPAVVMNRFSPAHWVFWVGPIVGAVLAAILYFYLLFPNSLSLSERVAIIK 240
                *********:.**** .:*.  *****:**:***:*.::** *:*** : *****:*::*

hAQP2           G-LEPDTDWEEREVRRRQSVELHSPQSLPRGTKA 271
hAQP5           GTYEPDEDWEEQREERKKTMELTTR--------- 265
                *  *** ****:. .*::::** :
```

Figure 1.6. Sequence alignment of hAQP2 and hAQP5, uniprot accession number P41181 and P55064, respectively. The sequences share a 63% sequence identity. The amino acids predicted to be phosphorylated by PKA have been marked with boxes. Predicted using pkaPS (Neuberger *et al.*, 2007).

For localization studies, green fluorescent protein (GFP) is commonly fused to one of the termini of the protein of interest. For a GFP-AQP5 construct, the protein was localized to intracellular vesicles which were trafficked to the plasma membrane upon stimulation with cAMP (Kosugi-Tanaka *et al.*, 2006). The movement was inhibited by H89, a known PKA inhibitor. The same result was obtained when antibodies targeting the carboxyl terminus where used instead of GFP (Yang *et al.*, 2003). Together, these results suggest similar regulation mechanisms for AQP5 as for AQP2. In contrast, another group's GFP-AQP5 protein was localized to the apical membrane, and by using H89 to inhibit PKA, the membrane localization was increased even further (Karabasil *et al.*, 2009). Since they observed the same increase when the PKA sequence was mutated, they concluded that AQP5 can be targeted to the membrane irrespective of phosphorylation of the PKA-motif.

Constructs with C-terminal tags have also been used. AQP5-GFP was localized constitutively to the plasma membrane and was not affected by any stimuli (Kosugi-Tanaka *et al.*, 2006). The same behaviour was observed for an AQP5-3xFLAG construct, also mutated at the phosphorylation site, but the protein remained localized to the membrane (Woo *et al.*, 2008). This suggests an important role for the carboxyl terminus since swapping the GFP tag between termini had a major impact. In addition, the AQP5-T259A-GFP mutation did not change the proteins localization, indicating that phosphorylation of this residue is not sufficient for membrane trafficking (Kosugi-Tanaka *et al.*, 2006). Consequently, phosphorylation does not explain the trafficking of constructs with tags fused to the carboxyl terminus. In addition, the results from tags attached to the amino terminus are inconclusive.

The termini of AQP5 have been further studied. The N-terminal region does not play a major role in trafficking since its removal did not move the protein from the apical membranes (Wellner *et al.*, 2005). In contrast, a construct with a deleted C-terminal was unstably expressed and found in intracellular sites in the cells. Moreover, from healthy mice the prolactin-inducible protein (PIP) binding to the C-terminus of AQP5 has been identified, possibly affecting the trafficking (Ohashi *et al.*, 2008). The gene expression of PIP was reduced in a mouse model for Sjögren's syndrome suggesting the involvement of AQP5 and this interaction partner in the disorder (Ohashi *et al.*, 2008).

From the 2Å resolution structure of hAQP5 (Horsefield *et al.*, 2008), features of potential importance regarding the trafficking were identified. The C-terminus of the four monomers had two different conformations, which could imply a disorder important for trafficking. Moreover, the C-terminus was anchored to loop D where the PKA consensus serine 156 is located. A phosphorylation at this site could possibly trigger conformational changes in the loop and consequently affect the interaction with the carboxyl terminus leading to trafficking of the protein. Taken together, these studies suggest an important role for the carboxyl terminus in hAQP5 trafficking.

1.9 THE HUMAN AQUAGLYCEROPORIN 10

Aquaporin 10, 11 and 12 are classified as the newest members of the aquaporin family. Although aquaporins have been extensively studied for two decades, surprisingly little is known about these members.

1.9.1 DISCOVERY AND CELLULAR LOCALIZATION

The first identification of AQP10 was in 2001 (Hatakeyama *et al.*, 2001). Even though AQP10 is suggested to belong to the aquaglyceroporin subfamily, the protein showed no glycerol transport. Nevertheless, a water flux could be measured and the protein contained two NPA-boxes, verifying it as a member of the aquaporin family. The protein was exclusively observed in duodenum and jejunum, the proximal parts of the small intestine, and not in ileum, the distal part of the small intestine.

Shortly thereafter, an independent study found a different isoform of AQP10 which was functionally characterized as an aquaglyceroporin; both water and glycerol permeability could be observed (Ishibashi *et al.*, 2002). Conversely to the first study, this isoform was longer and had a conserved C-terminus as compared to the other human aquaglyceroporins (Figure 1.7). Thus, the former isoform is an incompletely spliced version, causing a frame shift, a different termination, and thus a shorter protein (Morinaga *et al.*, 2002, Ishibashi *et al.*, 2002) (Figure 1.8). To distinguish between the isoforms the former will be referred to as AQP10sv (splicing variant) while AQP10 will be reserved to describe the fully functional full length protein.

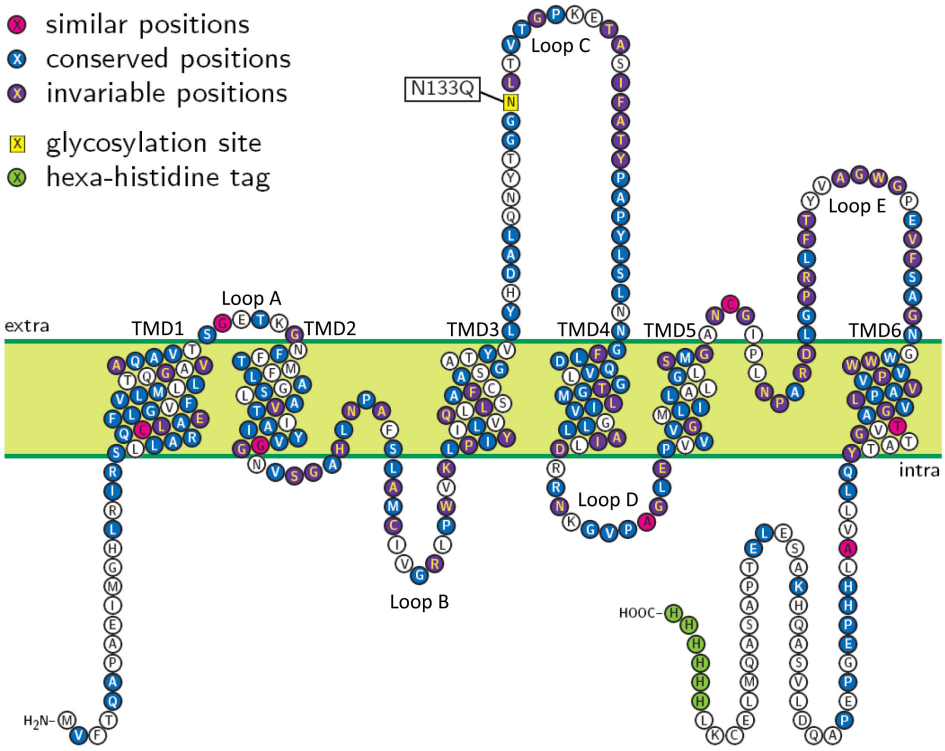


Figure 1.7. Snake plot of human aquaporin 10 compared with the other human aquaglyceroporins (hAQP3, hAQP7, hAQP9). Similar residues are shown in red, conserved in blue and invariable in purple. The glycosylation site mutated in **Paper V** has been marked, as well as the histidine tag.

```
CLUSTAL 2.0.12 multiple sequence alignment

                              ==============TMD1==============       =======
hAQP10          MVFTQAPAEIMGHLRIRSLLARQCLAEFLGVFVLMLLTQGAVAQAVTSGETKGNFFTMFL 60
hAQP10sv        MVFTQAPAEIMGHLRIRSLLARQCLAEFLGVFVLMLLTQGAVAQAVTSGETKGNFFTMFL 60
                ************************************************************

                ==TMD2========                                 ========TMD3=========
hAQP10          AGSLAVTIAIYVGGNVSGAHLNPAFSLAMCIVGRLPWVKLPIYILVQLLSAFCASGATYV 120
hAQP10sv        AGSLAVTIAIYVGGNVSGAHLNPAFSLAMCIVGRLPWVKLPIYILVQLLSAFCASGATYV 120
                **********************************************************

                                                        ========TMD4=========
hAQP10          LYHDALQNYTGGNLTVTGPKETASIFATYPAPYLSLNNGFLDQVLGTGMLIVGLLAILDR 180
hAQP10sv        LYHDALQNYTGGNLTVTGPKETASIFATYPAPYLSLNNGFLDQVLGTGMLIVGLLAILDR 180
                ************************************************************

                       ======TMD5======                                      =
hAQP10          RNKGVPAGLEPVVVGMLILALGLSMGANCGIPLNPARDLGPRLFTYVAGWGPEVFSAGNG 240
hAQP10sv        RNKGVPAGLEPVVVGMLILALGLSMGANCGIPLNPARDLGPRLFTYVAGWGPEVFRWETD 240
                ********************************************************   ..

                ========TMD6========
hAQP10          WWWVPVVAPLVGATVGTATYQLLVALHHPEGPEPAQDLVSAQHKASELETPASAQMLECK 300
hAQP10sv        SPGAGLHSPSSAKGSVPGSTALCL---------------------------------- 264
                   . : :*  .    ..:  * :

hAQP10          L 301
hAQP10sv        -
```

Figure 1.8. Sequence alignment of the two hAQP10 isoforms; hAQP10 denotes the full length protein (Ishibashi *et al.*, 2002) and hAQP10sv the splicing variant (Hatakeyama *et al.*, 2001). The locations of the transmembrane domains have been indicated as well as the NPA signature motifs (boxed).

More recent research has analysed the subcellular distribution and membrane polarization of AQP10 and AQP10sv. In one study, AQP10 was found in the apical membrane of absorptive intestinal epithelial cells of the ileum and AQP10sv was not detected at all (Mobasheri *et al.*, 2004). In another study, both versions were found in human duodenum and jejunum cells of the small intestine: AQP10sv in capillary endothelial cells of the small intestinal villi and hAQP10 in the epithelium (Li *et al.*, 2005). More insight into the localization of AQP10 was obtained from an *in silico* study made on AQP10 and AQP3 to compare the difference in protein targeting in epithelial cells (Cohly *et al.*, 2008); for AQP3 a signal sequence targeting it to the basolateral domain has been found (Rai *et al.*, 2006) whereas no such signal sequence have been identified for AQP10. On the other hand, AQP10 has glycosylation sites which provide the necessary signal to sort this aquaporin in the apical region (Cohly *et al.*, 2008).

1.9.2 TRANSPORT SPECIFICITY

The initial studies on AQP10 showed only water transport in the oocyte system (Hatakeyama *et al.*, 2001) but subsequent studies using the same method measured flux of water, glycerol and urea (Ishibashi *et al.*, 2002). From these data and from sequence homology, hAQP10 has been classified as an aquaglyceroporin with only limited water transport. Other aquaglyceroporins, hAQP7 and hAQP9, transport arsenite (Liu *et al.*, 2002), something not yet assayed for hAQP10. However, the homologue in zebrafish (*Danio rerio*), *aqp10,* has been shown to be permeable to water, glycerol and arsenite (Hamdi *et al.*, 2009). The corresponding homologue in European eel (*Anguilla anguilla*), AQP10e, is permeable to water, glycerol and urea (MacIver *et al.*, 2009). However, related studies have not been made on hAQP10 to determine the specificity for similar compounds.

1.9.3 PHYSIOLOGICAL ROLE

The physiological importance of aquaporin 10 has not yet been fully unravelled. Even thought the protein localization was determined in the initial studies, no obvious disease or health condition could be attributed to the protein. Initial knock out studies in mice gave no clear phenotype and only limited knowledge about the proteins actual function *in vivo* exists. However, the mouse gene contains many defects, causing a frame shift, and was later denoted as a pseudogene (Morinaga *et al.*, 2002).

AQP10 has more recently been shown to be regulated in connection to diseases, such as cholera. *Vibrio cholerae* is the causative agent of cholera and still causes a large number of deaths in the world. The bacteria colonises the epithelial cells in the small intestine and secretes cholera toxin which leads to a secretion response causing a massive loss of body fluid by vomiting and diarrhoea. Several membrane transporters are found to be down regulated during the acute phase in cholera patient, including hAQP10 and hAQP10sv (Flach *et al.*, 2007). A reduced amount of protein in the

epithelial cells reduces the permeability of the membrane thus limiting the secretary response.

The small intestine is the part of the gastrointestinal tract following the stomach, (Figure 1.9A) and is the primary site for nutrient absorption (Lodish, 2000). To facilitate an efficient absorption, the wall of the small intestine is covered with villi (Figure 1.9B) which increase its surface area to facilitate absorption of nutrients. The villi are formed of the epithelial cells (Figure 1.9C), called enterocytes or intestinal absorptive cells, and they are brush border cells covered with microvilli. Microvilli greatly enhance the surface area of the apical membrane and thus, they increase the number of integral membrane proteins it can contain (Figure 1.9D-E). Consequently the absorptive capacity of the small intestine will be enhanced. Covering the brush border is a layer called glycocalyx. The glycocalyx is a highly concentrated glycoprotein layer composed of N- and O-linked oligosaccharides, for example from integral membrane glycoproteins (Frey *et al.*, 1996, Ishimura *et al.*, 1984).

Several aquaporins have been detected in the small intestine but only AQP3, AQP8 and AQP10 have been shown to reside in the enterocytes. In contrast to AQP10, AQP8 has been found intracellularly (Calamita *et al.*, 2001) and AQP3 has been found to target the basolateral domain due to a signal motif (Rai *et al.*, 2006). Together, AQP10 and AQP3 could function as an entry pathway for small solutes such as water and glycerol from the lumen across the membrane (Figure 1.9E) (Mobasheri et al., 2004).

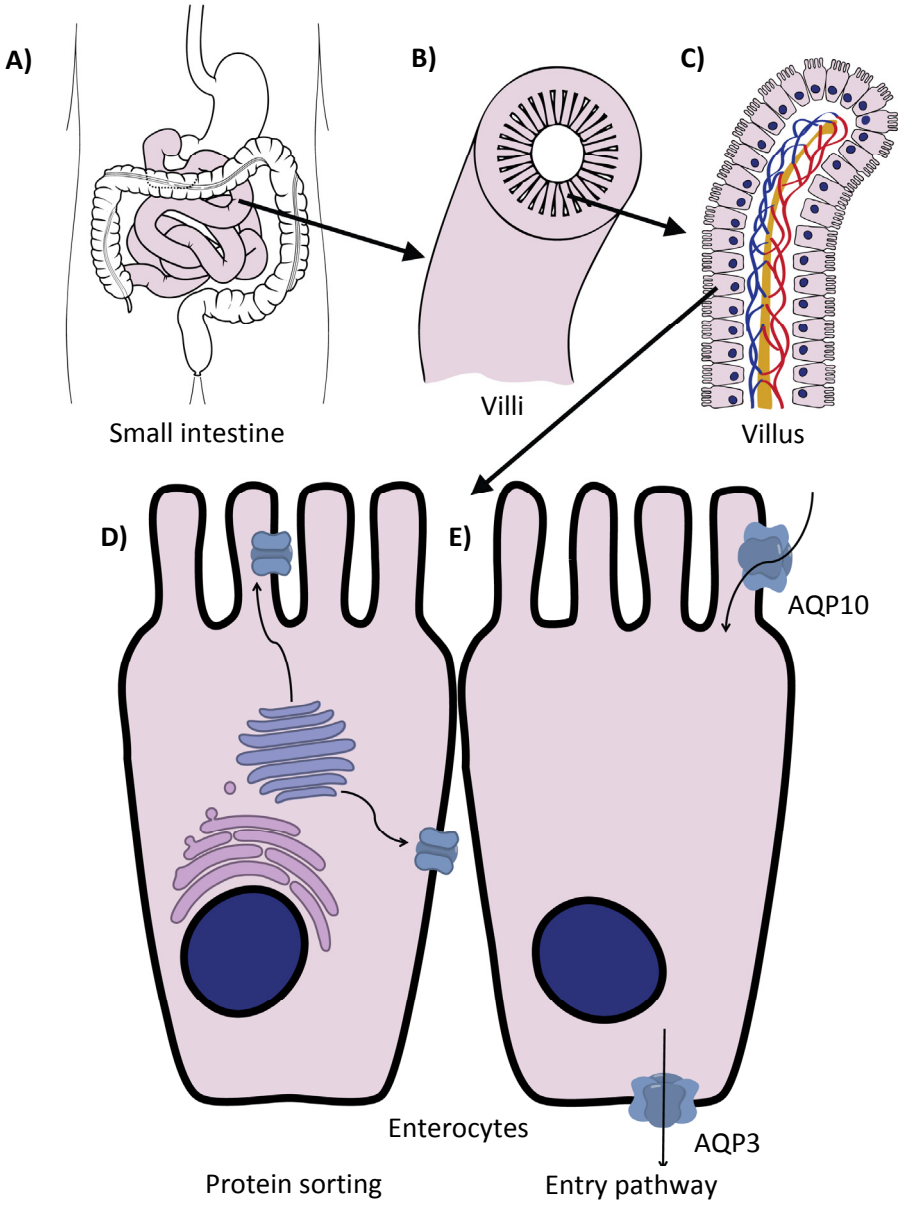


Figure 1.9. Summary of the localization and function of hAQP10 in the small intestine. A) Human small intestine. B) Cross section of the small intestine reveals the villi. C) A villus formed from enterocytes. D) Targeting of newly synthesized proteins to the basolateral membrane is mediated by specific amino acid motifs while for apical surface targeting sphingolipids and oligosaccharides have been shown to be involved (Keller *et al.*, 1997). E) Suggested entry pathway for small solutes such as water and glycerol from the lumen across the membrane via AQP10 and AQP3.

1.10 OTHER AQUAPORINS RELEVANT FOR THE PRESENT STUDY

Apart from the human aquaporins previously described, there are two other aquaporins of relevance for this study: namely the spinach aquaporin SoPIP2;1 and the *Plasmodium falciparum* aquaporin PfAQP.

1.10.1 SOPIP2;1

SoPIP2;1, at the time referred to as PM28A, was first isolated from *Spinacia oleracea*. It was found to be a major intrinsic protein of the leaf plasma membrane and migrated on an SDS-PAGE with an apparent size of 28kDa (Fotiadis *et al.*, 2001). Plant aquaporins can be divided into several groups, where the plasma intrinsic proteins (PIPs) can be divided into subgroup 1 and 2 with the latter (PIP2) being efficient water channels (Kaldenhoff *et al.*, 2006). In plants, the turgor pressure is crucial for cellular structure and rigidity and they have evolved rapid responses to water availability by regulation of aquaporins in the plasma membranes. During drought, dephosphorylation causes the channel to close (Johansson *et al.*, 1998), whereas flooding triggers channel closure by protonation of a conserved histidine (Tournaire-Roux *et al.*, 2003). The regulatory details of the protein has been revealed by the high resolution structure (Tornroth-Horsefield *et al.*, 2006), a study in which the protein was recombinantly produced to high yields in *P. pastoris*.

1.10.2 PFAQP

Malaria is one of the major diseases in the world. It causes millions of deaths every year, especially among children in Africa (Snow *et al.*, 2005). The cause of malaria is the four parasites of the genus plasmodium. The mosquito *Anopheles gambiae* can transfer them to humans, with *P. falciparum* being the most lethal. This parasite has only one single aquaporin (PfAQP) and it has been suggested that this channel might be of importance to protect the parasite during rapid proliferation, relive the osmotic stress during kidney passage, and take up glycerol during biosynthesis of new lipids (Hansen *et al.*, 2002). Thus, PfAQP is a potential drug target and interference with the aquaporin function may eliminate the parasite (Beitz, 2005). Today, *P. falciparum* has circumvented many of the available drugs. Targeting other molecular machineries within the parasite is therefore a potential part of a future vaccine (Richie *et al.*, 2002), highlighting PfAQP as an interesting target for malaria treatment.

2 SCOPE OF THE THESIS

Membrane proteins, including aquaporins, are vital for all cellular life and today they constitute the vast majority of all drug targets. Further studies on those highly important molecules are absolutely essential since many questions concerning their function and regulation still remain to be answered.

When the initial production level is high for a given protein target, this enables characterization. This is exemplified in **Paper I** were the exceptional yield of human aquaporin 1 allowed extensive functional assessments as well as early studies related to structure determination. In **Paper II**, the yield for PfAQP was significantly enhanced by gene optimization and structure characterization was initiated.

Noteworthy, not all membrane protein targets are easily produced and the factors directing a high production level is not understood. In the majority of cases, the first hurdle is the lack of strategies for acquisition of enough quantities of protein. To address these issues, overproduction of all the human aquaporins along with analyses of factors which could influence the final protein yield are presented in **Paper III**. Aquaporins were selected as the targets of interest since they constitute a family of highly homologous proteins, although with significant variation in terms of production yield, enabling a comparative study. Based upon this work, further production optimization was undertaken and evaluated in **Paper IV**.

In **Paper V** we investigated the influence and biological relevance of the glycan attached to hAQP10 during protein synthesis in *P. pastoris*. This is the only aquaporin which acquires such a modification within this production system, hence functional assays including stopped-flow and stability assays with circular dichroism were undertaken to explore the function and thermostability of the protein with and without a glycan.

Finally, a high resolution structure of a hAQP5 mutant was determined in **Paper VI**. We investigated the relevance for a conserved phosphorylation site as well as the importance of the carboxyl terminus for protein regulation by membrane trafficking.

Taken together, the scope of this thesis includes the optimization of production and the characterization possible after a high production yield has been achieved.

3 RESULTS AND DISCUSSION

Whilst over 5000 articles have been published mentioning aquaporins, many questions regarding their function and regulation remain. Several members of the aquaporin family are relatively uncharacterized (Figure 3.1). In contrast, AQP1 has been extensively studied, not the least since it was the first homologue with a determined structure. Noteworthy, the number of AQP2 articles is not related to structure, since no high resolution structure is currently available, but rather it reflects the involvement in protein trafficking. The merge of structural insights for AQP5 with the details of regulation for AQP2 might prove to be useful to further understand the molecular details involved in membrane protein regulation.

There are several possible explanations as to the scarce information available for certain aquaporins: they were discovered more recently, their transport specificity and physiological importance are debated, they lack homologues in common model systems, *etc.* A distressing reason is the lack of success in attaining sufficient yields of functional and pure protein to perform the desired research. Hence, systematic studies focusing on solving the overproduction are of utmost importance, where the production of several homologous in the same system could be very informative. For example, the human aquaporins constitute a family of highly homologous proteins, making them suitable candidates for systematic studies on factors directing high membrane protein yields from recombinant production.

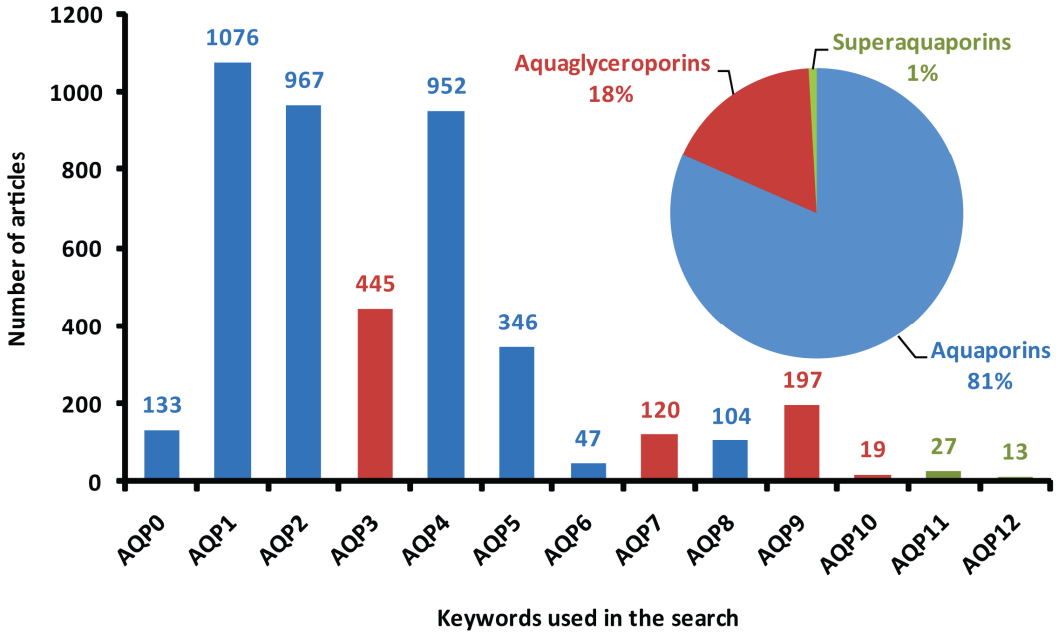


Figure 3.1. Number of published articles with a given keyword. Data acquired March 1st by searching PubMed http://www.ncbi.nlm.nih.gov/pubmed. In total 4446 articles were found divided among aquaporins (blue), aquaglyceroporins (red) and superaquaporins (green) as shown in the insert. Searching for the keyword 'aquaporins' resulted in 5383 hits.

3.1 PRODUCTION OPTIMIZATION

This chapter is mostly related to the work presented in **Paper I**, **Paper II**, **Paper III** and **Paper IV**. It will highlight several important factors connected to protein production and optimization.

The gene coding for the protein of interest has been cloned into the plasmid pPICZB (Invitrogen, 2010) and multiplied in *E. coli*. Subsequently, the plasmid was linearized and transformed to *P. pastoris* where homologous recombination integrates the whole plasmid into the genome. The pPICZ vectors contain a Zeocin resistance gene which makes it possible to select for positive transformants, both in bacteria and yeast. Thereafter, clones were selected based on their yield.

3.1.1 IDENTIFYING HIGH YIELDING CLONES BY PRODUCTION SCREENING

To have a fast and reliable way to determine the initial production yield is of key importance. From this, it is possible to conclude if the yields are sufficient for further characterization or if optimization is needed. Thus, we used an initial production screen for *P. pastoris* previously set up in the laboratory (Fantoni *et al.*, 2007). This screening is a robust and reproducible method to estimate the yield from cultivation of 24 colonies in a deep well block in a total volume of 2ml. Controls for high and low production yields, respectively, were included. Moreover, in order to verify proper membrane localization, a quick fractionation was made where whole cell lysates and membrane fractions were analysed with Western blot. We further developed the assay by more accurate densitometry analysis of the signals from the Western blot to select high producing clones (**Paper I**). Our results show a successful identification of a clone producing functional hAQP1 to exceptional levels in the *P. pastoris* membrane.

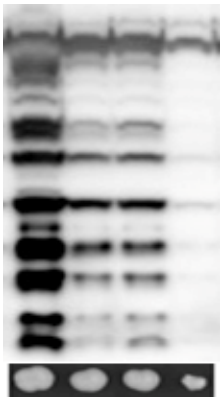


Figure 3.2. Immunoblot showing the yields after the initial production screen from a selection of colonies, shown underneath. There is a clear correlation between large colonies, i.e. the ability to grown on high concentration of Zeocin, and the production yield.

An alternative initial production screen for integral membrane proteins was assessed and developed (**Paper IV**). We examined the relationship between the yield and the ability of *P. pastoris* cells to grow on high concentration of Zeocin. Such cells are likely

to have multiple versions of the Zeocin resistance gene. Thus, they will also contain multiple copies of the desired gene, since they are both located on the expression plasmid integrated in the genome. We investigated the impact of having multiple copies of the gene by selecting several *P. pastoris* clones and comparing the yield with that from the previous initial production screen (Figure 3.2). As shown, there is a clear correlation between the ability to grow on Zeocin and the yield, where a fast-growing colony gives more protein. Thus, increasing the gene dosage can be beneficial, even for membrane proteins. Consequently, the latter method is to be preferred, since it is faster and more convenient.

For generation of *P. pastoris* strains, the linearized expression plasmids have routinely been transformed by chemical transformation using the Lithium Chloride Method (Cregg *et al.*, 1990). However, multiple insertion events occur at higher frequency when electroporation is used (Invitrogen, 2010). Consequently, as we had established the benefits for such a multi-copy insertion, also for integral membrane proteins, the use of electroporation for transformation was evaluated. Results showed a significant improvement in yield for the low producing targets, suggesting major benefits from an increased number of transcripts for these aquaporins. Hence, in contrast to previous findings (discussed in Chapter 1.4.2), increasing the gene dosage was found to be beneficial for aquaporin production in the *P. pastoris* production system (**Paper IV**).

Independent of choice of initial production screen and transformation method, a subsequent medium scale quantitative production screening was performed (**Paper I** and **Paper III**). To accurately estimate the relative production level, triplicate cultures of each promising candidate were cultivated in Erlenmeyer flasks. They were incubated at 30°C for 6h with a cultivation volume of 25ml. Growth in a shaker flask allows better aeration and is less sensitive to measuring errors due to the larger volume as compared to deep well blocks used in the initial screen. After harvest the obtained cells were broken with glass beads and cell debris was removed by centrifugation. The resulting cell lysates, referred to as the 'total protein production', are subsequently centrifuged to collect the membranes. The obtained fraction was referred to as the 'membrane fraction'. Furthermore, the total protein concentration of both these samples was determined and equal amount of total protein was analysed on Western blot.

Quantitative Western blot is a particular challenge, and variations occur between different Western blot kits, SDS-PAGE gels, and the transfer to the membrane. In particular, the detection system used to visualize the immunoblot is of importance; some detection kits indicate when too much protein/antibody is present for the signal to be reliable whereas others enhance all signals as much as possible on the cost of the linearity (Öberg, unpublished observation). Moreover, an even protein transfer between gel and membrane is crucial and a continuous agitation during antibody incubation is important to avoid a membrane blot with systematic variations. To

properly compare different Western blots, an internal scaling standard was used. The standard is a protein sample which was prepared and aliquoted at one single occasion. By loading this sample onto each gel the signal from the Western blot was assumed to correspond to the same amount of protein each time. Thus it was possible to remove variation arising from development of the Western blot membrane.

Moreover, as a part of the quality assurance, a growth control has always been included in the medium scale quantitative production screen. Its main purpose was to make sure the growth and production experiment progressed as expected. However, it also serves as an extra quality measurement of the Western blot; since it is the same protein (SoPIP2;1), it should always give the same result after scaling with the internal scaling standard. Indeed, this was observed for all successful Western blots as illustrated in Figure 3.3.

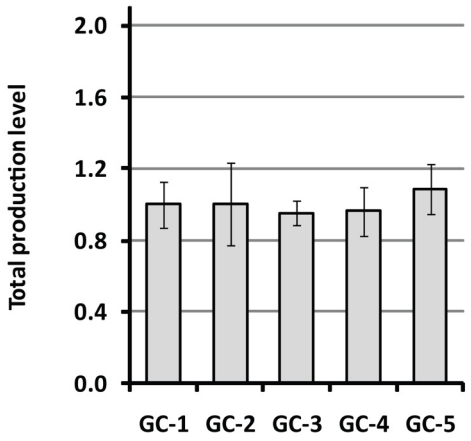


Figure 3.3. Variation between the growth controls (GC). The same control was cultured at different time points, over several years, and the production yield has been estimated from five independent Western blots.

To summarize, we have developed fast and robust screening methods, allowing us to estimate the yield of integral membrane protein in *P. pastoris* in a reproducible way. We have validated protocols to determine the quality of the quantitation, thus the final result will provide an accurate and reliable estimation of the production level. Our screens allow us to find the most promising clones with the highest yield and thereby the best candidates for further studies.

3.1.2 Significant Variation in the Production Yield between Homologous Aquaporins

Using our established screening method we could successfully quantitate the membrane protein yield of all aquaporins produced in the host *P. pastoris*. The development of the initial and the medium scale screen described in the previous chapter was an absolute prerequisite to assess the differences in yield. Thus, using these screens, we can identify the clone with the highest production for each homologue, as well as evaluate any optimization experiments with great accuracy.

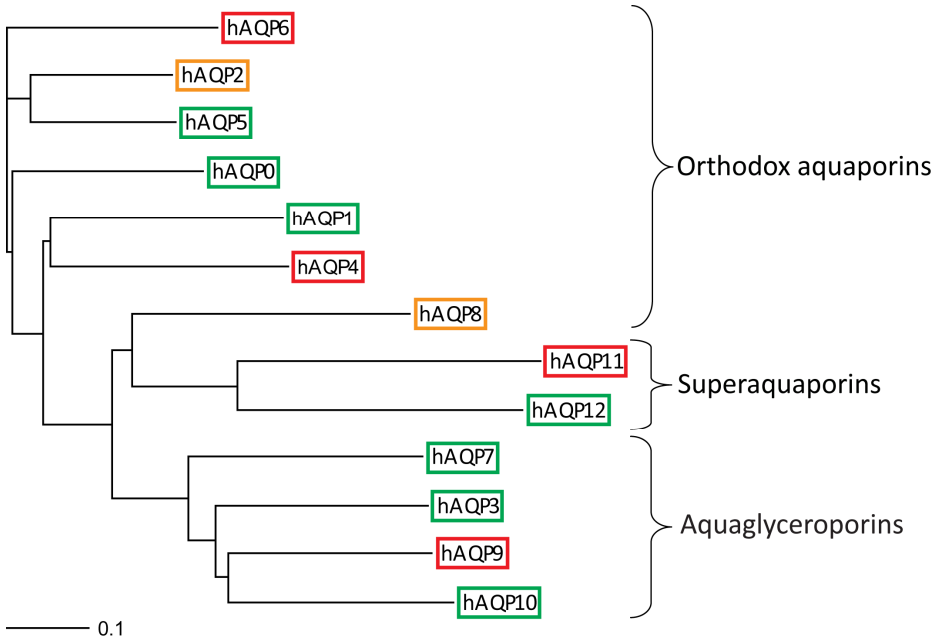


Figure 3.4. Phylogenetic tree of all 13 human aquaporins. Aquaporins giving a high protein yield is shown in green boxes, poor in orange boxes, and proteins with a yield below the detection limit in the medium scale production screen are shown in red boxes.

All human aquaporins are closely related homologues (Figure 3.4) with a high sequence similarity. For example 63% of the protein sequence is identical between hAQP2 and hAQP5. Nevertheless, there was a substantial variation in the yield when produced recombinantly in *P. pastoris* (Figure 3.5). Another striking example of the significant variation is the close relationship of hAQP1 and hAQP4, where the former belongs to the high producers and the latter is below the detection limit in the medium scale quantitative production screen.

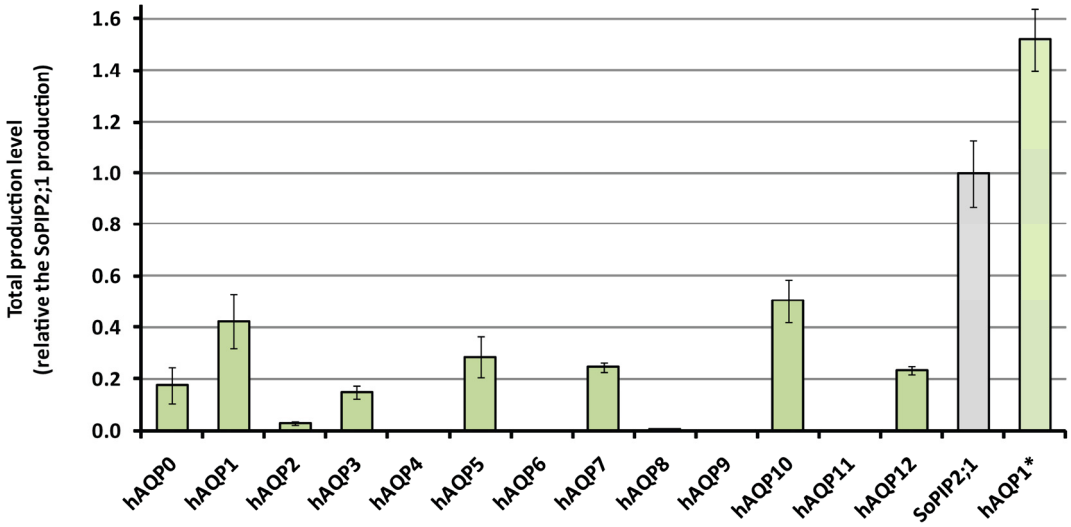


Figure 3.5. Bar chart showing the total production level of the different human aquaporins produced in the host *P. pastoris* relative the SoPIP2;1 production, for which the production is set to one (shown in gray). The y-axis represents average from triplicate cultures, error bars show the standard deviation (n=3). All results are from **Paper III**, with the rightmost bar (hAQP1*) also presented in **Paper I**.

Interestingly, the two aquaporins produced to the highest yields in *P. pastoris*, hAQP1 (**Paper I**) and SoPIP2;1 from spinach leaves, also have a high natural abundance in their native plasma membranes. An interpretation would be that their intrinsic properties results in high protein yields as the overproduced protein can be densely packed within the membrane of *P. pastoris* without forming aggregates or denaturing.

Aquaporins have been found in most organs in the human body. Some of them have only been detected in intracellular vesicles whereas others are found in the plasma membrane. At least two of them are known to be trafficked to the plasma membrane (AQP2 and AQP8), thus they reside in intracellular vesicles when trafficking has not been triggered. During protein synthesis, signals within the protein sequence itself or external recognition systems can determine the subcellular localization for that specific protein. To see if the localization could affect the protein yield in the *P. pastoris* production system, we analysed the yield versus the localization (Table 3.1). Although, not statistically ensured, there is a clear tendency for proteins targeted to the plasma membrane to have a high yield as compared to the aquaporins found in intracellular vesicles. The relevance of this finding is not fully clear, but it suggests that unknown signals or sequences within the protein itself are of importance for the final yield.

Protein	Recombinant yield	Native localization
AQP0	High	Plasma membrane (Chepelinsky, 2009)
AQP1	High	Plasma membrane (Nielsen *et al.*, 1993)
AQP2	Poor	Intracellular vesicles (untrafficked) (Nedvetsky *et al.*, 2009)
AQP3	High	Basolateral plasma membrane (Rai *et al.*, 2006)
AQP4	No	Basolateral plasma membrane (Nielsen *et al.*, 1997, Neely *et al.*, 2001)
AQP5	High	Apical plasma membrane (Karabasil *et al.*, 2009)
AQP6	No	Intracellular vesicles (Yasui *et al.*, 1999a)
AQP7	High	Apical plasma membrane (Skowronski *et al.*, 2007)
AQP8	Poor	Intracellular vesicles (untrafficked) (Garcia *et al.*, 2001)
AQP9	No	Plasma membrane (Elkjaer *et al.*, 2000)
AQP10	High	Plasma membrane (Mobasheri et al., 2004)
AQP11	No	Intracellular (Morishita *et al.*, 2005)
AQP12	High	Intracellular (Itoh *et al.*, 2005)

Table 3.1. Table showing the recombinant yield and subcellular localization in the native membrane for the different aquaporins found in mammalians. The localizations have been extracted from the shown references, but they are also stated, with only minor differences, in a review (King *et al.*, 2004).

In conclusion, we observed a significant variation in the production yield between the homologous human aquaporins. The factors behind this variation are not fully unravelled, but specific properties of the proteins are a plausible explanation. Nevertheless, our results establish the possibility to take overproduction yields of recombinant eukaryotic membrane proteins to a quantitative level, which is an essential step towards revealing the complex factors influencing the final yield of a eukaryotic membrane protein.

3.1.3 Significant Variation in Membrane Insertion between Homologous Aquaporins

It is noteworthy that no apparent connection between the yields (Figure 3.5) and the location in the phylogenetic three (Figure 3.4) could be found. However, this was changed when analyzing the fraction of the protein localized to the membrane. Consequently, for all the human aquaporins two set of signals were obtained: one corresponds to the total protein production (cell lysates, scaled and presented in Figure 3.5) and one corresponds to the membrane fraction (**Paper III**). To get the fraction of membrane localized material, the signal strength for the membrane fraction was divided by the total protein production for each specific target (Figure 3.6). Due to error propagation arising from this exercise, the error bars are relatively wide. However, our results showed a higher membrane insertion for the orthodox aquaporins as compared to the other aquaporins. A two tailed Fisher's Exact Test comparing the orthodox aquaporins to the non-orthodox gives a statistically significant association ($P<0.05$) where the orthodox aquaporins are more prone to be integrated. Thus, this test suggests that the substrate specificity also give rise to protein properties of relevance for the overproduction experiment aiming for correctly folded and membrane localised protein.

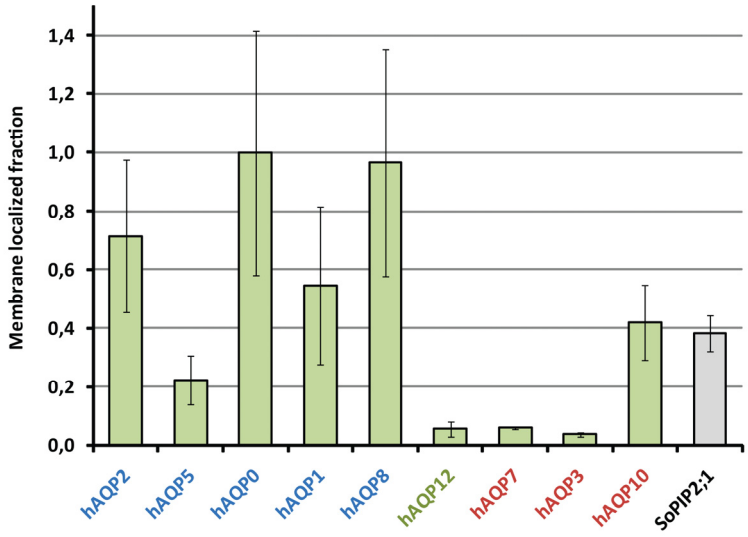


Figure 3.6. Bar chart showing the membrane localized fraction for the hAQPs detected in the medium scale quantitative production screen. The typical membrane insertion is shown for the reference protein SoPIP2;1. The aquaporins are grouped by their position in the phylogenetic three (Figure 3.4) and coloured as in Figure 3.1.

To evaluate the quality of an overproduction experiment, it would be appealing to directly measure the amount of correctly folded aquaporin in a membrane. We have conducted initial tests with GFP tagged AQPs and produced them in the *P. pastoris* cell (**Paper IV**). Our results support the production of correctly folded and properly inserted aquaporin for both a high (hAQP1) and low (hAQP8) yielding target in this particular host. However, their degree of membrane insertion appears to be different.

Without staining, it is hard to distinguish between the different cellular compartments, but nevertheless, there seems to be a higher degree of insertion of the low producing hAQP8 (Figure 3.5), indicating saturation and overload of the cellular machinery for proteins being produced to exceptionally high yields, as exemplified by hAQP1.

To summarize, we have produced eukaryotic membrane proteins to high levels in the yeast *P. pastoris* and found them to be properly inserted into the membrane. However, variations correlating with the proteins substrate specificity were observed as well as a possible saturation of the yeast's cellular machinery, for highly produced targets.

3.1.4 FERMENTOR GROWTH IS ESSENTIAL TO ACHIEVE HIGH YIELDS

When high producing clones have been selected, the large scale production using controlled growth can boost the yield even further. In a fermentor, the growth conditions are monitored and controllable: for example, the amount of dissolved oxygen can be closely measured and addition of pure oxygen gas can be used to improve the production even further. This is especially important for *P. pastoris*, as the host requires large amount of oxygen during methanol growth. Since methanol is used to induce the protein production, the yield is negatively influences by oxygen limitation (Cregg *et al.*, 2000). Consequently, we were able to achieve exceptional high yields of hAQP1; 90mg of pure protein was extracted per litre of culture when cultivated in a fermentor (**Paper I**).

Moreover, by using the appropriate sensors, the amount of viable cells in the reactor can easily be monitored. In contrast, in the classical optical density measurements, often used to analyse the growth in shaker flask, all cells are observed. With this feedback, we can optimize the induction more carefully, since initial addition of methanol can cause too high concentrations thereby killing the cells instead of only inducing the promoter. Interestingly, not all *P. pastoris* clones producing aquaporin continued to grow upon switching from glycerol to methanol to trigger induction; in general, the clones overproducing aquaglyceroporins had a much slower growth rate and in some cases a decay of living cells were observed. A plausible explanation could be that the slightly wider aquaglyceroporins can transport the small methanol molecule into the cell where it would be toxic. To evaluate this possibility, three amino acids lining the pore entrance at the ar/R constriction region in hAQP5 were mutated to create a larger pore and thereby changing the pore specificity, as have been made for AQP1 (Beitz *et al.*, 2006). Indeed, the growth on methanol was hampered for this mutant (Öberg, unpublished observation).

A mixed feed containing 60% sorbitol and 40% methanol was evaluated for hAQP10 with the intention to lower the concentrations of the toxic methanol for aquaglyceroporins (related to **Paper V**). This provides an additional carbon source not giving rise to the gene repression associated with glycerol and it has been shown to increase the protein yield by increasing biomass (Jungo *et al.*, 2007a, Jungo *et al.*, 2007b) and by weakening induction to better match the requirements of the cells metabolism

(Holmes *et al.*, 2009). For hAQP10, we observed no significant changes in protein yield from the mixed feed. However, a protein degradation product commonly seen in the pure methanol feed, disappeared. Hence, the mixed feed resulted in a more pure and stable protein, suggesting a reduced cellular stress response under these conditions. Therefore, this could provide a solution for stable production of all aquaglyceroporins in order to achieve homogenous samples suitable for further characterization.

Finally, the fermentor cultures made it possible for us to verify the quality and reliability of the medium scale quantitative production screen. In one study, we produced 90mg hAQP1* per litre of fermentor culture (**Paper I**) and in another we produced 30mg hAQP10 per litre of fermentor culture (**Paper V**), a third of the amount obtained for hAQP1. Comparing hAQP1* and hAQP10 from the medium scale screen (Figure 3.5) gives exactly three times higher yield for hAQP1*. Hence, the data from the medium scale quantitative production screen can be extrapolated to the large scale production in the fermentor. In comparison, targets resulting in low yields in Figure 3.5, such as hAQP2, hAQP6, and hAQP8, have been problematic to overproduce, even in fermentor, to sufficient yields for subsequent analysis.

3.1.5 CONSTRUCT DESIGN TO INCREASE YIELD

Several different approaches to identify factors directing high production of eukaryotic membrane proteins were evaluated, including alterations around the initiation codon of the mRNA, exploiting Mistic-fusions, chimeric/fusion AQP constructs, as well as directing the topological maturation of the aquaporin monomer by direct mutations.

Translation is controlled by the rate of initiation, thus being affected by the 5' sequence of the mRNA (Romanos *et al.*, 1992). For eukaryotic mRNA the sequence flanking the initiator codon (underlined) was found to be (A/G)NN<u>AUG</u>G[1] by Marilyn Kozak, henceforth called the Kozak sequence (Kozak, 1981). In yeast, the consensus sequence is different (A/Y)A(A/U)A<u>AUG</u>UCU[1] (Cigan *et al.*, 1987). Interestingly, alteration of the yeast consensus sequence have shown a 2-3 fold effect in translation efficiency, in contrast, modification of the Kozak sequence shows a much greater effect in mammalian cells (Baim *et al.*, 1988).

In 2007, when this project was underway, the recommendation in the EasySelect Pichia Expression Kit manual was to use the yeast consensus sequence A/YAA/TA<u>ATG</u>TCT (Invitrogen, 2005). Hence, we aimed to mimic this sequence by just allowing silent mutants in the second codon (position +4-6, where adenine in <u>ATG</u> is defined as +1) (**Paper III**). We also introduced a Y in the -4 position and triple A in the codon before ATG for the overproduction of all human aquaporins in *P. pastoris* (Table 3.2). Interestingly, the results showed a clear preference for G in the +4 position. However, this is likely not related to efficiency of translation initiation but is rather related to the small non-polar amino acids Ala and Gly codons (always starting

[1] N = Any base, Y = Pyrimidine: thymine/uracil or cytosine (Cornish-Bowden, 1985)

with G) and their importance for a successful cleavage of the initiator Met from the nascent polypeptide (Xia, 2007). The importance of a G in the +4 position was confirmed for hAQP8 for which the production yield was increased upon mutation from the wild type TCT sequence to GCT (Figure 3.7A). Moreover, a significant decrease in the production yield for hAQP1 was seen upon imitation of the yeast consensus sequence (**Paper III**) as compared to a construct with unaltered sequence for the second codon (**Paper I**) (Figure 3.7B). This could partly be explained by the intrinsic cytosine in the +6 position, which upon change to a thymine reduced the yield. Taken together, mimicking TCT in the second codon did not show a positive influence on the aquaporin yield. In 2009, after the submission of **Paper III**, the recommendation in the Invitrogen manual was changed to the Kozak sequence ANNATGG with the motivation "from other studies we have seen that the easier Kozak can be used" [2].

Gene	Yield	Initiation sequence	2nd residue
hAQP1*	+++	C AAA ATG GCC	Ala
hAQP10	++	C AAA ATG GT**T**(C)	Val
hAQP1	++	C AAA ATG GC**T**(C)	Ala
hAQP5	++	C AAA ATG AAG	Lys
hAQP7	++	C AAA ATG GTT	Val
hAQP12	++	C AAA ATG GC**T**(A)	Ala
hAQP0	++	C AAA ATG TGG	Trp
hAQP3	++	C AAA ATG GGT	Gly
hAQP2	+	C AAA ATG TGG	Trp
hAQP8	+	T AAA ATG TCT	Ser
hAQP4 m1	-	C AAA ATG **TC**T(AG)	Ser
hAQP4 m23	-	C AAA ATG GT**T**(G)	Val
hAQP6	-	T AAA ATG GAT	Asp
hAQP9	-	C AAA ATG CAG	Gln
hAQP11	-	C AAA ATG TC**T**(G)	Ser

Table 3.2. Table showing protein production yields for all human AQPs overproduced in *P. pastoris*. AQPs are listed based on their production level, starting with the hAQP giving the highest observed yield. The silent mutations made in the second codon are underlined. The original sequence is given in brackets. The two smallest non-polar residues (Ala and Gly) in the second position are underlined. All results are from **Paper III**, with the top entry (hAQP1*) also presented in **Paper I**.

[2] Personal communication with Dr. Marco Polidori, Invitrogen, 16 March 2011.

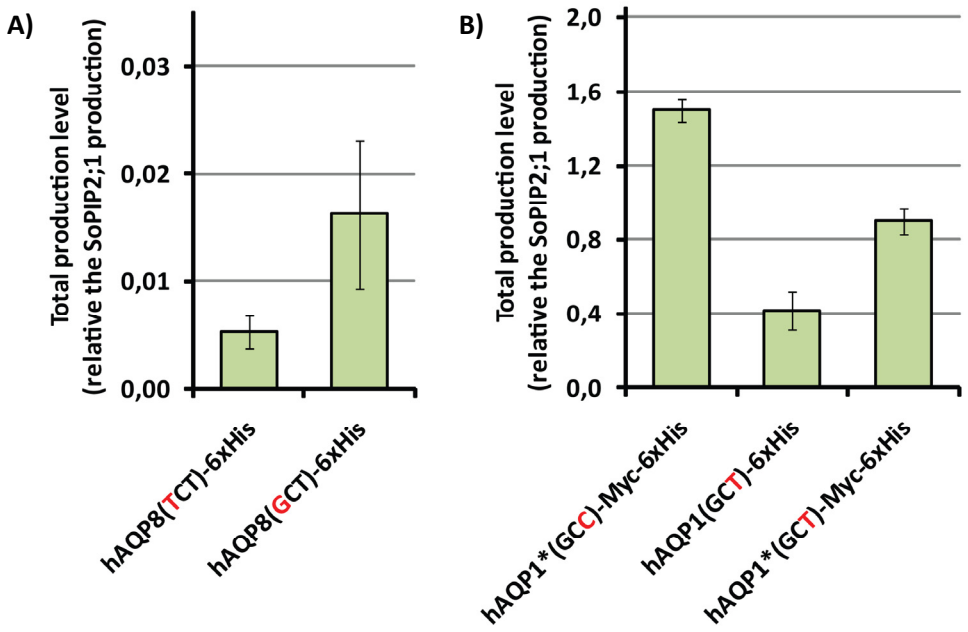


Figure 3.7. Bar chart showing the total production yield for A) two hAQP8 constructs and B) three hAQP1 constructs. Variations in the nucleotide sequence for the second codon, shown in brackets, are marked in red. C-terminal affinity tags are also shown. All results are from **Paper III**, with hAQP1* also presented in **Paper I**. Production is relative the SoPIP2;1 production, for which the production is set to one. The y-axis represents average from triplicate cultures, error bars show the standard deviation (n=3).

Moreover, the genome of *P. pastoris* is now available for the public (De Schutter *et al.*, 2009). Thus, the sequence flanking the initiator codon of AOX1 in the wild type *P. pastoris* cell can be investigated. Since the cell produces extensive amount of AOX1, this sequence is likely to be beneficial for efficient translation initiation. From the database the sequence was found to be ACGATGG, perfectly matching the Kozak sequence. This confirms our observation that the eukaryotic mRNA sequence is superior to the yeast consensus sequence when overproducing eukaryotic membrane proteins in the yeast *P. pastoris*.

Apart from the flanking sequence of the initiator codon, other aspects of the construct design were also evaluated. In particular, the common strategy to improve production by fusion partners was evaluated for production of eukaryotic membrane protein in a simple eukaryotic host. The intention is that the fusion to a stable and highly-produced protein or peptide sequence would enhance the production of the protein of interest. This approach has been successful for soluble proteins in both *E. coli* and *S. cerevisiae* using maltose binding protein as well as other fusion partners (Wang *et al.*, 2003, Hennig *et al.*, 1998, Perez-Martin *et al.*, 1997, Lian *et al.*, 2008). Mistic, acronym for 'membrane-integrating sequence for translation of integral membrane protein constructs', is a membrane anchored protein found in the bacteria *Bacillus subtilis*. It has been evaluated as a fusion partner in a bacterial system (Roosild *et al.*, 2005) where it has been able to increase the production of membrane proteins (Petrovskaya *et al.*,

2010). The Mistic sequence was codon optimized (described in more detailed in Chapter 3.1.6) to evaluate the usability of this approach in a eukaryotic host (**Paper IV**). Our results showed a remarkably stable protein production in the same range as the hAQP1 homologue. We further examined whether the production of a low yielding AQP could be enhanced by making a Mistic-fusion in analogy to the design used in bacteria. Unfortunately, our attempts with Mistic-AQP fusions resulted in a lower yield than for the AQP alone. Hence, these data imply that the concept of Mistic fusions to increase eukaryotic membrane protein yields cannot be directly transferred from *E. coli* to *P. pastoris*.

Due to the unsuccessful trials to increase the yield in *P. pastoris* using the Mistic approach, an aquaporin with naturally high yield in this particular host was tested as fusion partner. However, both the full length hAQP1 and parts thereof failed to increase the production of full length hAQP8. Additional chimeric constructs were tested where the amino-terminus, transmembrane domain 1 (TMD1), TMD1-2 or TMD1-3 of hAQP8 was substituted for the corresponding protein sequence of hAQP1. The absence of positive results from these constructs indicates the importance of the carboxyl terminal half of hAQP1 for the high production. Thus, both termini of hAQP1*, respectively, was engineered upon SoPIP2;1 replacing the existing terminus. Ultimately we found a chimera with increased production yield; the C-terminus of hAQP1* could increase the production of SoPIP2;1 (**Paper IV**) further supporting the importance of the hAQP1 C-terminus for high production yields.

Yet another method of construct design was evaluated, this time based on detailed information of the topology maturation for certain aquaporin targets: AQP1 and AQP4. During folding of AQP1 in the ER, a four spanning intermediate is first observed (Skach *et al.*, 1994), which subsequently matures into the final six transmembrane fold (Lu *et al.*, 2000) as schematically shown in Figure 3.8. In contrast, AQP4 folds sequentially into the six spanning topology (Shi *et al.*, 1995, Sadlish *et al.*, 2005). More recent studies identified two amino acids near TMD2 to be responsible for the difference in maturation (Foster *et al.*, 2000): Asn49, Lys51 for AQP1 corresponding to Met48, Leu50 for AQP4. Interestingly, transferring these specific amino acids between AQP1 and AQP4 lead to a swap in folding pathway (Buck *et al.*, 2007). This information provided a unique opportunity to evaluate the effect of topology maturation on the membrane protein yield for these aquaporin targets.

Noteworthy, by testing this hypothesis we could significantly increase the production of hAQP4 by mutating two amino acids to resemble those in hAQP1; the anticipated effect is likely related to a shift in the folding pathway (**Paper III**). Figure 3.9 shows the relevant amino acids in hAQP1 and hAQP4, respectively, as well as the corresponding residues for the other human aquaporins.

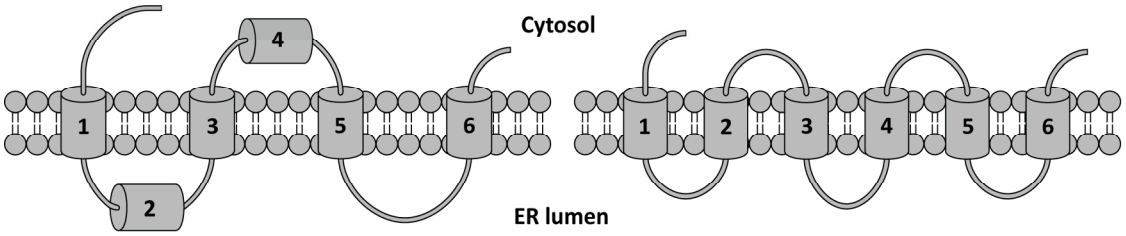


Figure 3.8. Topology maturation for AQP1 with a four spanning intermediate (left) and the mature fold with six TMD (right) (Lu *et al.*, 2000). Asn49 and Lys51 in the N-terminal part of TMD2 are determinants for this fold (Foster *et al.*, 2000).

Interestingly, a unique presence of positively charged amino acids is observed in hAQP1 which is completely lacking in hAQP4. In correlation to this, the hydrophobicity of the amino acids in and in close proximity of a given amino acid segment is known determinants for the insertion of that TMD into the membrane (Hessa *et al.*, 2005, Hessa *et al.*, 2007). Hence, a plausible explanation for the observed difference in folding pathway could be the variation in hydrophobicity in the TMD2 segment, where AQP1 is less hydrophobic due to two positively charged residues before TMD2. The concept to increase the production yield by altering the folding pathway for hAQP4 is unlikely to be transferable to other human aquaporins, since no other family members possess two hydrophobic amino acids in the current positions (Figure 3.9). Furthermore, the folding pathway is yet to be established for these proteins. Nonetheless, this presents an excellent example of how detailed knowledge acquired from other biochemical experiments can provide the details needed for a successful overproduction experiment.

```
CLUSTAL 2.0.12 multiple sequence alignment

hAQP1       (49) NVKVSLAFGLSIATLAQSVGHISGAHLNPA (78)
hAQP4 M23   (48) MVLISLCFGLSIATMVQCFGHISGGHINPA (77)

hAQP0       (41) VLQVAMAFGLALATLVQSVGHISGAHVNPA (70)
hAQP2       (41) VLQIAMAFGLGIGTLVQALGHISGAHINPA (70)
hAQP3       (56) FLTINLAFGFAVTLGILIAGQVSGAHLNPA (85)
hAQP5       (42) ILQIALAFGLAIGTLAQALGPVSGGHINPA (71)
hAQP6       (55) VLQIAITFNLVTAMAVQVTWKASGAHANPA (84)
hAQP7       (67) YLGVNLGFGFGVTMGVHVAGRISGAHMNAA (96)
hAQP8       (65) LLQPALAHGLALGLVIATLGNISGGHFNPA (94)
hAQP9       (57) VITINVGFSMAVAMAIYVAGGVSGGHINPA (86)
hAQP10      (55) FFTMFLAGSLAVTIAIYVGGNVSGAHLNPA (84)
hAQP11      (72) PAHPTWTLTLVYFFSLVHGLTLVGTSSNPC (101)
hAQP12      (66) DFGPDLLLTLLFLLFLAHGVTLDGASANPT (95)
```

Figure 3.9. Sequence alignment of all hAQPs around TMD2, with the amino acids corresponding to Asn49 and Lys51 in AQP1. For AQP4 M23 the corresponding amino acids are Met48 and Leu50. Nonpolar amino acids (hydrophobic) are marked in bold and polar and electrically charged (hydrophilic) are marked in bold and underlined.

Taken together we have evaluated the importance of the construct design for recovery of high membrane protein yields in *P. pastoris*. In conclusion, the sequence flanking the initiator methionine should mimic the Kozak sequence and the use of fusion partners is a non-trivial approach for eukaryotic membrane proteins. For the few cases where the target of interest is known to great detail, alteration of the topology maturation in the translation and translocation process might prove to be useful.

3.1.6 GENE OPTIMIZATION TO INCREASE YIELD

A different way of improving the production is to optimize the gene sequence. As the ease of creating synthetic genes has improved, more and more services now offer gene optimization. The goal is to increase the yields for the protein target of interest in a certain host system. During gene optimization several parameters are considered including the codon usage, adaptation of the GC content, inhibition of splicing within the gene, prevention of stable mRNA secondary structures, avoidance of DNA repeats, and where applicable, finding the most efficient signal peptide for your gene (GeneArt, 2011, GenScript, 2011). Nevertheless, the main focus has been on the codon optimization where the codon usage and adaptation to the host was first presented almost 25 years ago (Sharp *et al.*, 1987).

The genetic code has 64 possible ways of combining a triplet of nucleotides, codons, to code for one of the 20 amino acids or the three stop codons. The tRNA, facilitated by the ribosome, will translate the codon sequence into a polypeptide. Most amino acids can be encoded by several codons, Met and Trp being the only exceptions, and the frequency of a given combination of a nucleotide codon and tRNA anticodon can vary substantially between different organisms (Table 3.3). Several explanations to the origin of this phenomenon exists, one being that the bias have evolved to increase the resource efficiency within the cell. One such improvement would be to enhance the efficiency of protein synthesis and thereby the ability to divide and grow more rapidly (Akashi, 2001). Thus, organisms which undergo rapid growth can reduce the diversity of tRNAs to reduce the metabolic load resulting in a biased tRNA pool.

As a consequence of recombinant gene expression, the difference in the preferred codon usage in the native host and the tRNA pools of the expression system may cause inefficient translation hampering the protein production. One solution is to increase the intracellular tRNA pool by overexpressing genes coding for the rare tRNAs. This common strategy for *E. coli* was evaluated for PfAQP produced in *E. coli* (**Paper II**) but without any successful outcome. Moreover, moving this concept to higher organisms has been found to be impractical (Gustafsson *et al.*, 2004). Hence, the alternative approach of changing the gene is left as a possible solution.

Codon	AA	Pp	Hs	Pf	Codon	AA	Pp	Hs	Pf
UAA	*	0.51	0.30	0.69	UUA	Leu	0.16	0.08	0.63
UGA	*	0.20	0.47	0.21	UUG	Leu	0.33	0.13	0.14
UAG	*	0.29	0.24	0.10	CUU	Leu	0.16	0.13	0.11
GCU	Ala	0.45	0.27	0.42	CUC	Leu	0.08	0.20	0.02
GCC	Ala	0.26	0.40	0.11	CUA	Leu	0.11	0.07	0.08
GCA	Ala	0.23	0.23	0.43	CUG	Leu	0.16	0.40	0.02
GCG	Ala	0.06	0.11	0.05	AAA	Lys	0.47	0.43	0.82
CGU	Arg	0.17	0.08	0.11	AAG	Lys	0.53	0.57	0.18
CGC	Arg	0.05	0.18	0.02	AUG	Met	1.00	1.00	1.00
CGA	Arg	0.10	0.11	0.09	UUU	Phe	0.54	0.46	0.84
CGG	Arg	0.05	0.20	0.01	UUC	Phe	0.46	0.54	0.16
AGA	Arg	0.48	0.21	0.60	CCU	Pro	0.35	0.29	0.40
AGG	Arg	0.16	0.21	0.16	CCC	Pro	0.15	0.32	0.10
AAU	Asp	0.48	0.47	0.86	CCA	Pro	0.42	0.28	0.46
AAC	Asp	0.52	0.53	0.14	CCG	Pro	0.09	0.11	0.05
GAU	Asp	0.58	0.46	0.87	UCU	Ser	0.29	0.19	0.23
GAC	Asp	0.42	0.54	0.13	UCC	Ser	0.20	0.22	0.08
UGU	Cys	0.64	0.46	0.87	UCA	Ser	0.18	0.15	0.26
UGC	Cys	0.36	0.54	0.13	UCG	Ser	0.09	0.05	0.05
CAA	Gln	0.61	0.27	0.87	AGU	Ser	0.15	0.15	0.32
CAG	Gln	0.39	0.73	0.13	AGC	Ser	0.09	0.24	0.06
GAA	Glu	0.56	0.42	0.86	ACU	Thr	0.40	0.25	0.26
GAG	Glu	0.44	0.58	0.14	ACC	Thr	0.26	0.36	0.12
GGU	Gly	0.44	0.16	0.42	ACA	Thr	0.24	0.28	0.53
GGC	Gly	0.14	0.34	0.05	ACG	Thr	0.11	0.11	0.09
GGA	Gly	0.33	0.25	0.44	UGG	Trp	1.00	1.00	1.00
GGG	Gly	0.10	0.25	0.10	UAU	Tyr	0.47	0.44	0.89
CAU	His	0.57	0.42	0.86	UAC	Tyr	0.53	0.56	0.11
CAC	His	0.43	0.58	0.14	GUU	Val	0.42	0.18	0.40
AUU	Ile	0.50	0.36	0.39	GUC	Val	0.23	0.24	0.06
AUC	Ile	0.31	0.47	0.07	GUA	Val	0.15	0.12	0.41
AUA	Ile	0.18	0.17	0.54	GUG	Val	0.19	0.46	0.12
					Coding GC (%):		42.7	52.3	23.8

Table 3.3. Table showing the codon usage for three different species. The genetic code includes 64 possible ways of making a triplet, of these, three are stop codons in most organisms and are here marked with a star (*). The usage is given as fractions where 1 equals 'always used'. The GC content of the coded DNA is also given. Gene frequencies adapted from Codon Usage Database (Kazusa, 2007) for *Pichia pastoris* (Pp), *Homo sapiens* (Hs) and *Plasmodium falciparum* (Pf) where the species had the numbers 4922, 9606 and 36329 respectively.

By codon optimization, it is possible to introduce silent mutations in the foreign gene to change the codons for the encoded amino acids, thereby making use of the more favourable codons of the host. Codon Adaptation Index (CAI) is a measurement of how well the codon usage in a protein coding DNA sequence matches the bias of a certain host. Table 3.4 shows CAI values for all hAQPs and PfAQP in their native host and in *P. pastoris*. Two of the calculators, CAIcal (Puigbo *et al.*, 2008) and EMBOSS:cai (Bleasby, 2001), are using the same codon table, thus giving very similar values. Contrary, JCat (Grote *et al.*, 2005) uses its own table giving more conservative values but with the same trends between the high and low values. Moreover, CAI can also be used as a numerical estimator of gene expressivity where a high value would indicate a highly expressed gene (Wu *et al.*, 2005). Two constructs optimized for production in

P. pastoris have been included in Table 3.4 and are named "Opt-". As seen, these have a higher CAI value in *P. pastoris* than in their native hosts and would thus be expected to give a higher protein yield.

			CAIcal			EMBOSS:cai			JCat
Gene	**N_c**	**%GC**	**Pp**	**Hs**	**Pf**	**Pp**	**Hs**	**Pf**	**Hs**
hAQP0	52.4	57.7	0.670	0.798		0.598	0.794		0.343
hAQP1	42.0	61.2	0.613	0.826		0.540	0.826		0.516
hAQP2	40.2	64.0	0.616	0.815		0.535	0.811		0.476
hAQP3	38.6	60.3	0.635	0.843		0.579	0.838		0.470
hAQP4	56.6	49.2	0.707	0.760		0.664	0.750		0.228
Opt-hAQP4	25.6	40.8	0.907	0.750		0.909	0.693		0.135
hAQP5	36.5	63.1	0.602	0.832		0.534	0.825		0.546
hAQP6	44.0	62.9	0.591	0.803		0.540	0.805		0.457
hAQP7	45.7	56.6	0.638	0.791		0.593	0.791		0.382
hAQP8	44.7	61.2	0.610	0.801		0.556	0.798		0.415
hAQP9	52.7	48.9	0.727	0.757		0.680	0.736		0.215
hAQP10	49.7	58.5	0.681	0.810		0.617	0.797		0.357
hAQP11	55.3	55.6	0.664	0.766		0.583	0.744		0.341
hAQP12	38.0	66.6	0.605	0.824		0.527	0.816		0.528
PfAQP	36.5	30.7	0.750		0.787	0.724		0.736	
Opt-PfAQP	32.6	44.0	0.895		0.376	0.886		0.373	

Table 3.4. Table showing the codon adaptation index (CAI) for aquaporins in their native host and *P. pastoris*. CAI calculations using CAIcal (Puigbo *et al.*, 2008) and EMBOSS:cai (Bleasby, 2001) were based on the Codon Usage Database (Kazusa, 2007) for *Pichia pastoris* (Pp), *Homo sapiens* (Hs) and *Plasmodium falciparum* (Pf) where the species had the numbers 4922, 9606 and 36329 respectively. CAI calculations from JCat (Grote *et al.*, 2005) use its own codon tables and since it does not contain data for Pp or Pf, these columns have been omitted. For clarity, the values obtained from the calculators are separated by a dotted line. The GC content and the N_c value is also given. N_c represents the extent of the gene's codon bias; 20: strong bias, 61: no bias.

It is also worth mentioning, that although an efficient translation is necessary to produce high yields of functional proteins, it is not sufficient. The nascent polypeptide must be correctly folded, translocated, and undergo any necessary posttranslational modifications.

We set out to produce the malaria parasite aquaporin (PfAQP) introduced in Chapter 1.10.2 (**Paper II**). However, initial production attempts only gave low production in *E. coli* and the protein was below the detection limit in our *P. pastoris* screens. We anticipated that the explanation to this was to be found in the unusual low GC content: only 31% in the PfAQP gene and 24% in the coding region of the whole genome (Table 3.4 and Table 3.3). Thus, a gene with a substantial higher CAI and GC content was synthesised (Opt-PfAQP in Table 3.4). As predicted, the final protein yield was significantly increased, giving a functional and membrane localized protein product.

Whether the reason for an increased production of PfAQP in *P. pastoris* is found in the codon adaptation, in the alterations of the GC content or in the combination thereof is

not obvious. It has been reported that the change in GC content is the major contributor to the increased translational efficiency in *P. pastoris* (Sinclair *et al.*, 2002) and considering the large deviation in GC content for *P. falciparum* it seems to be a probable explanation. Interestingly, the wild type PfAQP gene has a higher CAI in *P. pastoris* than all the human aquaporins. This suggests it could possibly be produced to high yield. This however, was not observed in our study. Consequently, the usage of CAI for prediction of membrane protein production is limited.

The difference in GC of the coding genome is not always as distinct as for *P. falciparum* and *P. pastoris*. In the case of human AQP4, the difference between *H. Sapiens* and *P. pastoris* is significantly smaller and no obvious production problems can be predicted based on the GC content. The CAI values were also similar for hAQP4 in the native host and in *P. pastoris* giving no indication of production problems related to codon usage in the two hosts. Nevertheless, the wild type sequence was submitted to two large companies specialized on engineering genes: GenScript and GeneArt. As comparison, a manual codon optimized gene was made using JCat (Grote *et al.*, 2005). The result is presented in Table 3.5. The 'effective number of codons used', N_c, is a measurement on the extent of codon bias in a gene; its value range from 20, corresponding to one codon being exclusively used for an amino acid, to 61 when the probability is equal for all possible codons (Wright, 1990). The strategy for the JCat program is simply to optimize each single codon without considering what has already been used. In contrast, the algorithms used by GenScript and GeneArt appears to be more sophisticated, but surprisingly different, although the percentage of changed nucleotides are similar. The major differences are seen in nucleotide no3 of each codon, as expected, since substitutions at this site often do not change the amino acid. Nevertheless, the variation between the optimizations performed by the companies is substantial and obviously different strategies have been used. From Table 3.3 it appears that human genes have a relatively high GC content, 52%, as compared to *P. pastoris*, 43%. Generally a decreased GC content was observed for all the adapted genes (Table 3.5), due to the more AT biased codon choice for the lower eukaryote.

Name	%GC	%GC (1)	%GC (2)	%GC (3)	N_c	Changed
hAQP4 wt	49.2	53.3	43.3	51.1	55.9	---
JCat	43.6	45.5	43.3	41.8	20.0	25.3%
GenScript	40.8	45.5	43.3	33.4	32.7	24.0%
GeneArt	45.7	45.8	43.3	48.0	23.7	24.5%

Table 3.5. Table summarizing the different results obtained when optimizing hAQP4 using JCat, GenScript or GeneArt. Total GC content is shown as well as the GC content of the first, second and third position in the codon. The number of bases changed is also presented as a percentage of the 969 nucleotides in the gene. N_c is a value representing the extend of the codon bias; 20: strong bias, 61: no bias.

To further analyse the importance of the GC content, a plot was made for all the optimized genes in Table 3.5 and compared to the wild type hAQP4 (Figure 3.10). There are large variations in the native protein, with some regions spiking over 67% or dipping to 25%. In contrast, the optimized genes stay within a much more narrow range.

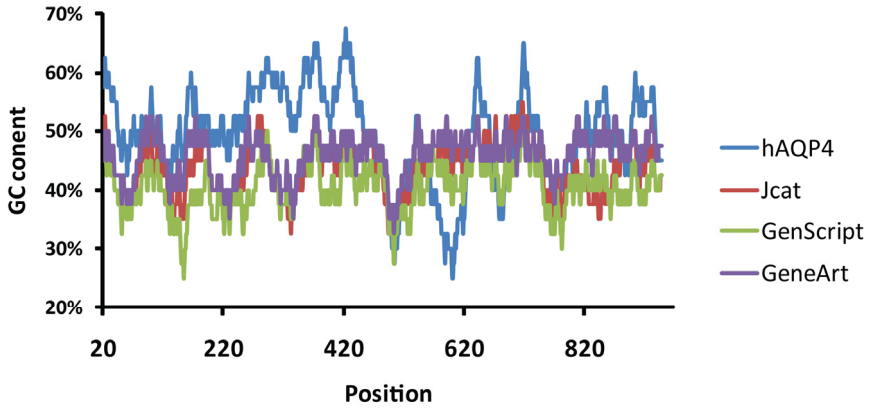


Figure 3.10. Plot showing the GC content in a 40bp window centered at the indicated nucleotide position. Data where acquired using the EMBOSS:isochore (Rice, 2011).

To analyse the effect of the codon adaptation in practice, we used the GenScript optimized version of hAQP4 for production in *P. pastoris* (**Paper IV**). The yield of the optimized gene (Opt-hAQP4) could be significantly increased as shown in Figure 3.11. If this could be attributed to the codon usage or the GC content is not known. Nevertheless, since other processes are known to affect the translocation and topology maturation, we combined the folding pathway mutants of hAQP4* (Chapter 3.1.5) with codon optimized hAQP4 to investigate the combinatorial effect of enhancing both the translation and translocation of a protein. Indeed, the yield was increased even further giving a final yield in the same range as hAQP1 (shown as Opt-hAQP4* in Figure 3.11).

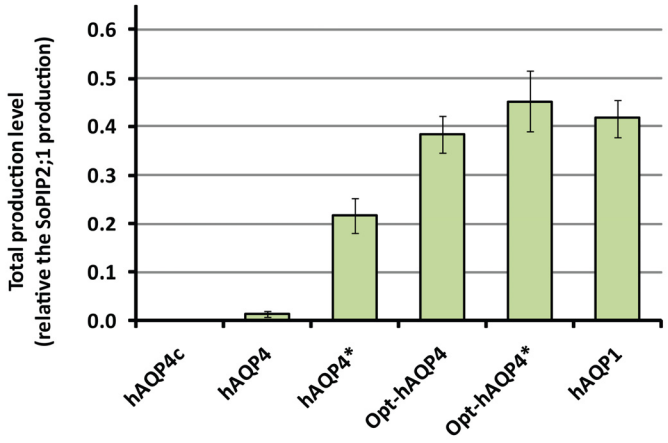


Figure 3.11. Bar chart showing the yield of four different hAQP4 versions. Electroporation has been the transformation method of choice except for hAQP4c where chemical transformation has been used. Constructs with mutations casing an AQP1 like topology maturation have been indicated with a star (*). The prefix Opt symbolizes constructs where the gene has been codon optimized.

The strategies used for hAQP4 illustrates how detailed knowledge of the protein target can aid the overproduction experiments. However, such knowledge is rarely available which emphasizes the importance of further studies of unknown proteins. Nevertheless, when the required information exists, combining strategies such as those for hAQP4 are likely to improve the success rate.

It is clear that gene optimization can be used to increase the production yield of membrane proteins. In the process, codons will be adapted to use the host tRNA pool in an efficient way, the GC content will be changed to a value more similar to the host, and changes can be made to influence the mRNA stability. However, the precise reasons behind the actual effect of the codon adaptation remain elusive. In the codon adaptation experiments results were significantly improved when the GC content of the gene was close to the GC content of the host's coding DNA (**Paper IV** and **Paper II**). However, several of the unmodified human aquaporins giving high yield have a very high average GC content (hAQP1 and hAQP5) while some of the low producers (hAQP4 and hAQP9) have GC content close to *P. pastoris*. Consequently, the reasons behind the improved results are likely to be a combination of different, possibly unknown, factors. When applicable, combining multiple strategies, such as alteration of the topology maturation with codon optimization, is likely to have a high success rate.

3.2 CHARACTERIZATION

The characterization described here will include functional studies using proteoliposomes and dynamic light scattering, analysis of aquaporin stability by circular dichroism, and structural studies using x-ray crystallography. These experiments can only take place after sufficient amount of protein has been achieved and thus, these steps are completely dependent on a highly pure and stable protein product. This chapter is mostly related to the work presented in **Paper I**, **Paper V** and **Paper VI.**

To purify the overproduced protein, the *P. pastoris* cells are disrupted and membranes collected by centrifugation. Subsequently, detergent is used to extract the protein from the membrane. However, different detergents will solubilise different proteins, making the choice of detergent crucial to receive micelles in which the protein of interest is stable. Finally, the protein is purified with ion exchange- or affinity chromatography followed by size exclusion chromatography to obtain highly pure protein.

3.2.1 FUNCTIONAL CHARACTERIZATION OF AQUAPORINS

During extraction from the native membrane into an artificial lipid/detergent environment, the protein is subjected to relatively harsh conditions. In addition, the recombinant overproduction itself could influence the protein activity since it not fully mimics the wild type membrane environment. Thus, it is of key importance to verify the protein functionality of the purified product. Reconstitution of the pure protein into liposomes, now referred to as proteoliposomes, is therefore a common procedure to measure protein activity.

For measurements of hAQP1, *E. coli* lipids where sonicated to form liposomes (**Paper I**). By treatment with detergent (OG) the liposomes where destabilized and a fraction of the pure protein was inserted into the vesicles. Subsequently, the detergent was removed by dilution and the protein was fully incorporated, forming proteoliposomes and possibly a minor population of empty liposomes.

These proteoliposomes were rapidly mixed with a hypertonic solution containing sucrose in a stopped-flow apparatus. Hence, the surrounding buffer had a higher osmolality causing water to flow out from the vesicles (Figure 3.12A). Subsequent shrinkage is measured as an increased light scattering due to the increased index of refraction (Figure 3.12B). The increase is based on the condensing lipids of the bilayer and the concentration increase of osmolytes inside the shrinking vesicles (Lawaczeck, 1984). The increased light scattering is measured as an exponentially increasing curve and a single exponential curve can be fitted for liposomes without protein and a double exponential function for the proteoliposomes (Figure 3.12B). Thus, it indicates that the curve is composed of two phases: one slow corresponding to the passive diffusion of water across the vesicle and one fast representing the facilitated water flux.

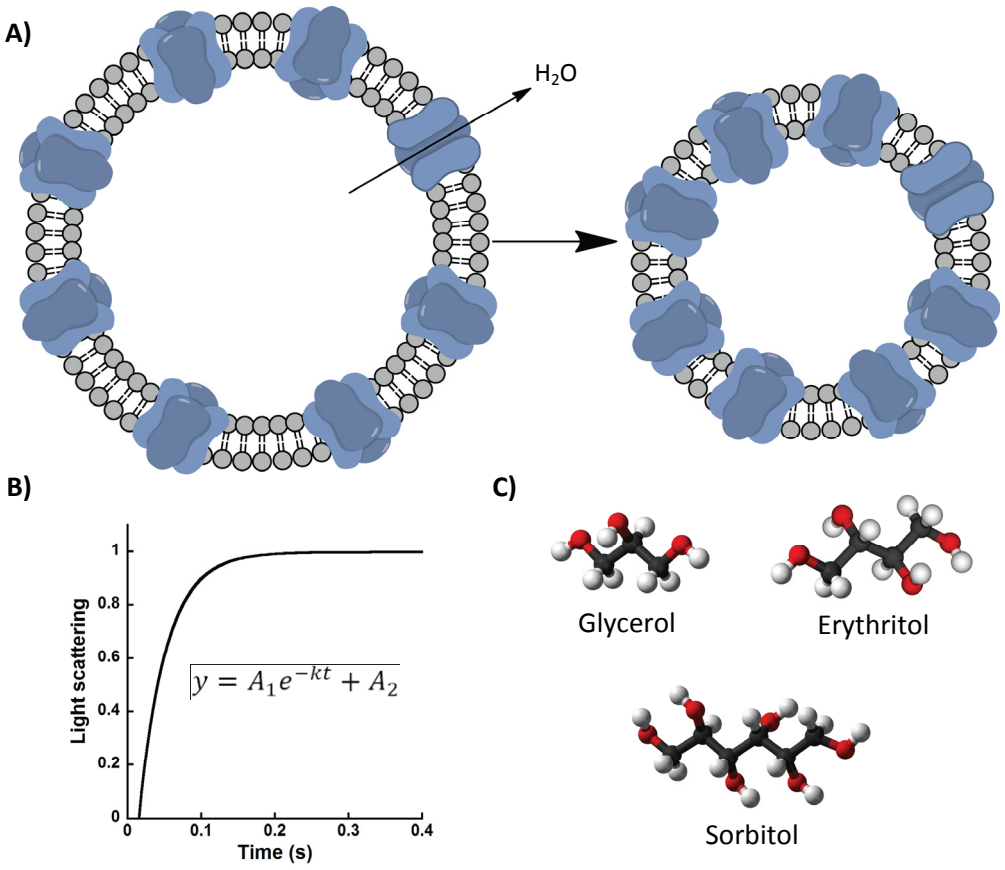


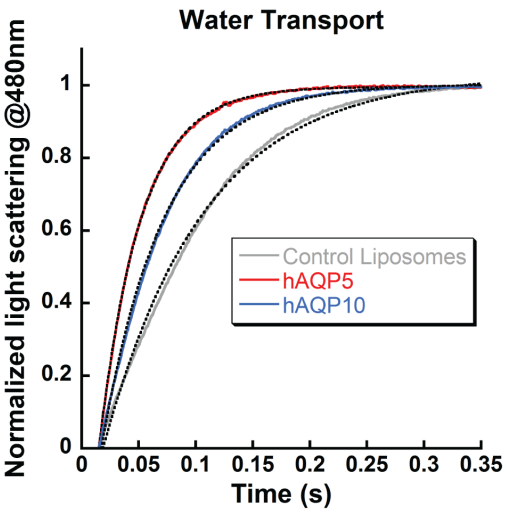
Figure 3.12. Functional assay of aquaporins using stopped-flow experiment on proteoliposomes. A) The osmolality gradient drives water out from the vesicle causing it to shrink. B) Shrinkage is observed as an increased light scattering. Curve fit using an exponential function gives the rate constant (k) of transport. C) Sorbitol will not cross the membrane at any relevant rate when measuring rapid water transport. For hAQP10, the transport of glycerol and erythritol was also assessed.

The diameter of the vesicles was determined with dynamic light scattering (DLS) and the value for the osmotic water permeability, P_f, was determined. This value depends on the rate constant (acquired from the curve fits, Figure 3.12B), the size of the vesicle and the external osmolality. However, since the rate constant will increase as the osmotic gradient increases, the P_f should remain constant upon changes in the osmolality gradient. The system is also sensitive to the amount of functional protein in the vesicle. This can be attributed by analysis on Western blot, where the signals of all samples being compared should be equally strong (as in Figure 3.14D). The amounts received with such an assay assume that all detected protein is active.

For hAQP1 we could verify proper localization of the overproduced protein to the *P. pastoris* membrane indicating a correctly folded product. Moreover, we successfully determined a threefold increase in water permeability of liposomes having purified hAQP1 inserted (**Paper I**). The exceptional yield of hAQP1 has been commented on earlier, but to verify the protein as fully functional is an essential aspect, since this allows us to characterize the properties and structure of the biologically relevant form of the protein.

Another high yielding target in *P. pastoris* is hAQP10 (**Paper V**). In this study, we had several forms of hAQP10: the wild type hAQP10, the engineered mutant lacking the possibility of acquiring a glycan (hAQP10-N133Q), and a sample subjected to an extra purification step where only the non-glycosylated fraction were collected (hAQP10-ΔGlyc). As described in Chapter 1.9.2, only a limited water conductance through this aquaporin was reported when assayed in oocytes. In contrast, we measured a significant water transport for all versions of hAQP10 in proteoliposomes (Figure 3.14A). A plausible explanation is the different membrane curvature and lipid composition between the systems. Such a difference caused the mechanosensitive AQY1 to activate its water transport when assayed in artificial liposomes as compared to spheroplasts of its native host (Fischer *et al.*, 2009).

Moreover, the measured water transport rate of hAQP10 was comparable with the transport of the orthodox aquaporin 5 when assayed in proteoliposomes. Interestingly, the resulting proteoliposomes showed a major difference in vesicle size with a radius of 80nm for hAQP5 and 138nm for hAQP10, possibly due to the different detergents used. The large difference in rate constants was thus compensated for by the variation in vesicle size, resulting in similar value for P_f (Figure 3.13).



Sample	k	r	C_{out}	P_f
Control	11±0.06	70	610	2.4±0.01
hAQP5	27±0.03	80	610	6.6±0.01
hAQP10	18±0.05	132	610	7.3±0.02
hAQP1				9.1±0.4

$$P_f = \frac{k}{\frac{S}{V_0} \cdot V_w \cdot C_{out}}$$

Figure 3.13. Water transport for hAQP5 and hAQP10. The rate constant (k [s^{-1}]), vesicle size (r [nm]), external osmolality (C_{out} [mOsm/kg]), and the osmotic water permeability (P_f) are shown in the table. S/V_0 is the vesicle surface area to initial volume ratio, V_w is the partial molar volume of water. For reference, the P_f value for hAQP1 is also presented (**Paper I**). The P_f for control liposomes in that experiment was 3.1±0.01.

For measurement of rapid transport, such as the facilitation of water *via* aquaporins, sorbitol will not cross the membrane at any relevant rate. However, for slower transport, movement of sorbitol could interfere with the measured transport and thus, the use of sucrose could be beneficial. This larger compound does not pass through the membrane of the proteoliposomes and have been used during measurements of larger sugar alcohols, such as glycerol and erythritol.

To measure these transports, proteoliposomes are resuspended in buffer containing glycerol or erythritol, respectively. The vesicles, now loaded with the sugar alcohol to study, are mixed with a buffer containing sucrose giving identical osmolality on both sides of the vesicle. During this isoosmolal condition, the difference in glycerol concentration will cause a flux of glycerol out from the cell resulting in a hyperosmolal condition where water is also transported out from the proteoliposomes. The simultaneous transport, which is caused by the sugar alcohol, will result in shrinkage of the vesicle. This was measured as an increased light scattering. If the initial condition is not isoosmolal, a rapid water flux will cause the cell to burst or shrink without being based on the transport of the sugar alcohol.

When the protein is not stable in OG, and detergents with much lower CMC are used, such as DDM, the incorporation of protein often causes leaky liposomes. To analyse the transport for hAQP10, which needs DDM or DM to remain stable (**Paper V**), a different reconstitution method was used to minimize this effect. Lipids, protein and detergents where mixed and by the addition of Bio-beads, detergent was slowly removed to create proteoliposomes. The main advantage with this approach is the more reproducible reconstitution of protein. The drawbacks include the possible

formation of multilaminar vesicles (which will not shrink as single laminar vesicles) and a larger spread in vesicle size, even if the average size can be accurately controlled by the amount of salt in the buffer.

Consequently, we used this new method for measuring transport through hAQP10. In agreement with earlier studies we obtained a glycerol transport (Figure 3.14B), clearly putting this homologue among the aquaglyceroporins. In addition, we show transport of longer sugar alcohols such as erythritol (Figure 3.14C). Erythritol is used as an artificial sweetener and is known to be absorbed in the small intestine and, hence, it does not induce the same laxative effect as can occur for other sugar alcohols (Munro *et al.*, 1998). This suggests a potential absorbing role for hAQP10 within the human body.

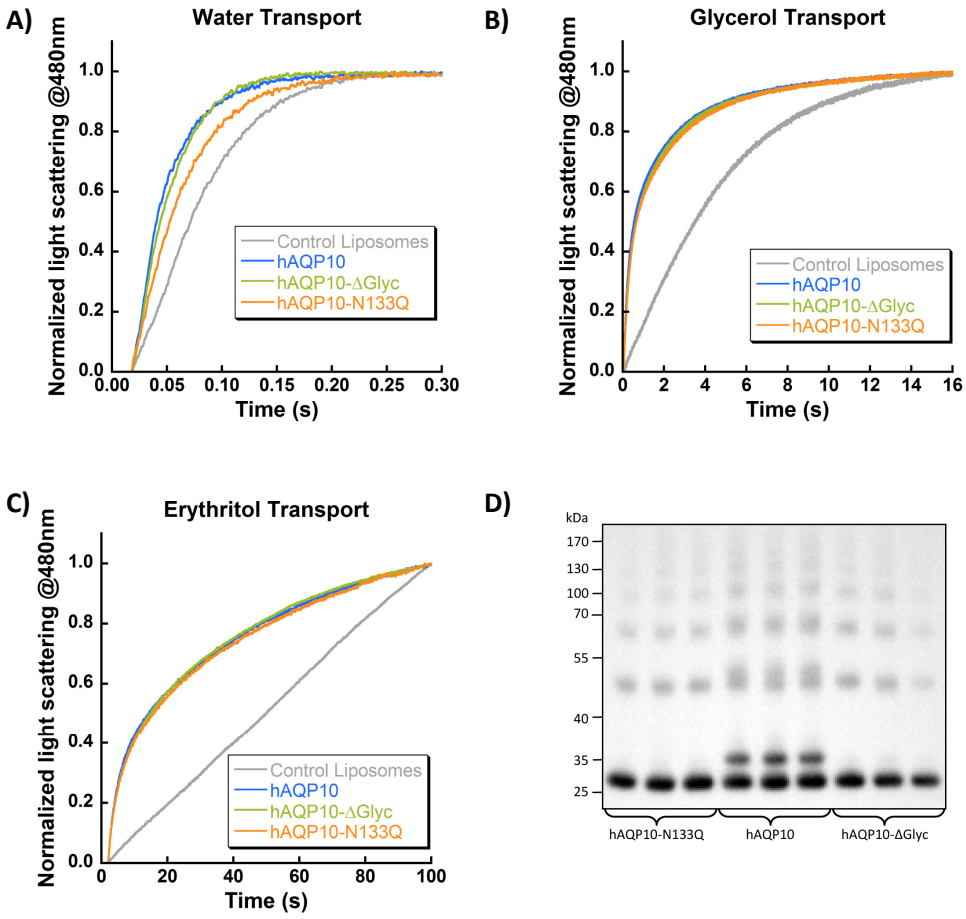


Figure 3.14. Results acquired from the stopped-flow experiment showing A) water, B) glycerol, and C) erythritol transport for hAQP10. D) The proteoliposomes for each construct and for each assay are shown on the Western blot.

In summary, we have used protein reconstituted into liposomes to measure the flux of water and solutes using stopped-flow methods. Our results confirm high water permeability of hAQP1 and hAQP5, but somewhat surprisingly also for hAQP10. In addition, we observed the passage of glycerol and longer sugar alcohols through the aquaglyceroporin 10.

3.2.2 CHARACTERIZATION OF AQUAPORIN STABILITY

As described in Chapter 3.1.5, two amino acids near TMD2 are important for the topology maturation of hAQP1 and hAQP4 (Figure 3.9). In addition, a residue near TMD5 has been shown to be important for the tetramer stability of AQP1; in AQP1 Lys51 forms a polar interaction with Asp185 in an adjacent monomer (Buck *et al.*, 2007). Thus, we examined the possibilities of increasing the stability of the aquaporin 4 tetramer by introducing an additional mutation, casing the whole tripartite motif (Asn49, Lys51 and Asp185) to be transferred from hAQP1 (**Paper III**). We observed an increased production yield for the triple mutant, suggesting that we obtained an increased stability of the protein tetramer. The mutated hAQP4 protein was also found to be inserted into the membrane. Analysis of all hAQPs reveals four other family members where a polar interaction is possible in a similar manner: hAQP0 (Gln43-Tyr177), hAQP2 (Gln43-His177), hAQP5 (Gln44-Tyr178), and hAQP11 (His74-Ser209). In the hAQP5-S156E structure (**Paper VI**), the side chains of Gln44 and Tyr178 are separated by around 3Å, thus enabling a hydrogen bond between them (Figure 3.15). How this is affecting the tetramer stability is yet to be evaluated. However, it is noteworthy that for all of the above mentioned hAQPs, for which the production and membrane insertion was quantitated (hAQP0, hAQP2 and hAQP5), a high membrane insertion was observed (Figure 3.6). Thus, higher tetramer stability could be important for overproduction yield as well as for correct membrane insertion.

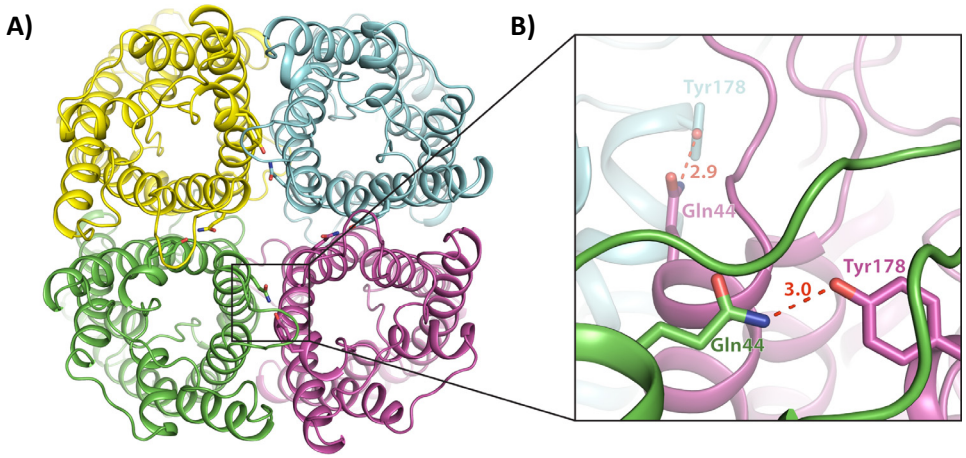


Figure 3.15. A) The hAQP5 tetramer seen from the extracellular side with the four monomers in different colours (**Paper VI**). B) Enlarged and rotated image of the interaction between Gln44 from chain A (green) with Tyr178 in chain C (purple).

AQP10 is the only human aquaporin where an N-linked glycan is attached during the recombinant protein synthesis in *P. pastoris*. Hence, we conducted a series of experiments to conclude the exact location and nature of this modification (**Paper V**). Of the three theoretical sequons, only Asn133 was found to be glycosylated and in total 35% of the hAQP10 sample was modified. From an additional purification step (hAQP10-ΔGlyc), as well as from the construction of an engineered mutant (hAQP10-N133Q), we could obtain protein completely lacking the N-linked glycan.

Hence, we could examine the effect of glycosylation on protein stability. By using circular dichroism (CD) the secondary structure content of a helical protein can be observed by monitoring the signal at 208nm and 222nm which are the two largest features associated with α-helices. By gradually increasing the temperature of the sample, the thermostability of a protein can be monitored and compared to other versions of the same protein or other proteins.

From the CD value at 222nm, the approximate helical content could be estimated to 45% at ambient temperature, but dropped to 9% near 100°C. From the CD measurements of secondary structure content as a function of the temperature we found an increased thermostability of up to 6°C for the glycosylated protein as shown in Figure 3.16. Interestingly, the shape of the melting curve does not indicate the presence of two distinct populations, even though only 35% is glycosylated, suggesting that the presence of one glycan per tetramer is sufficient for an enhanced structural stability for the whole tetramer complex. Possibly, such a cooperative mechanism may be a generic mechanism of membrane protein stabilization by glycosylation. In this context it is noteworthy that only one of the monomers in hAQP1 is glycosylated in the human body (Heymann *et al.*, 1998).

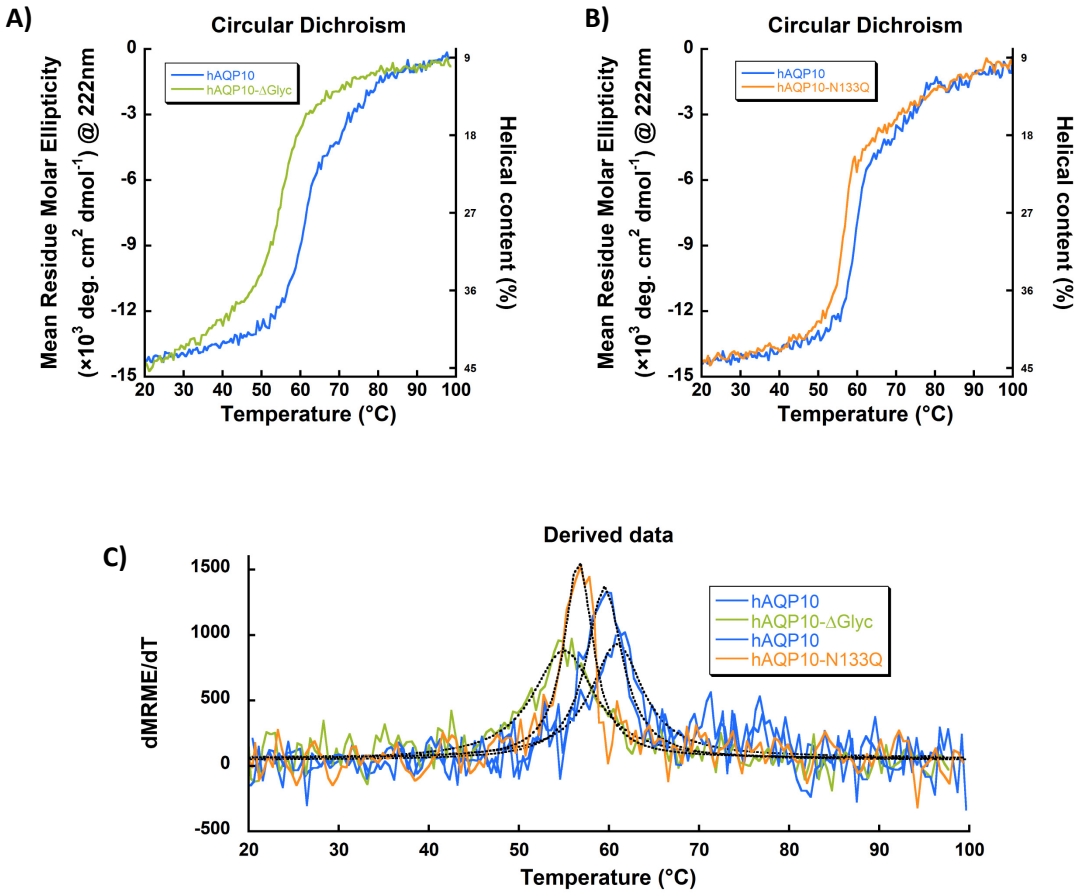


Figure 3.16. Results from circular dichroism. A-B) Melting curves and C) derived melting curves representing the thermostability of hAQP10. Traces for wild type hAQP10 are marked in blue, hAQP10-ΔGlyc in green, hAQP10-N133Q in orange.

Moreover, for the protein sample were the glycosylated population has been removed (hAQP10-ΔGlyc), the melting temperature is lower than for the mutant hAQP10-N133Q, where no glycan can be attached during protein synthesis (Figure 3.16). The purification includes binding to the Concanavalin A lectin which binds the carbohydrates α-D-mannopyranose, α-D-glucopyranose and sterically related residues with available C-3, C-4, or C-5 hydroxyl groups. It is used for isolation of cell surface glycoproteins from detergent-solubilized membranes (GE Healthcare, 2011). However, several of the most common detergents for solubilisation of membrane proteins will interact with the lectin due to their glucose or maltose parts. Nevertheless, we managed to separate a fraction of the non-glycosylated hAQP10 (hAQP10-ΔGlyc) from the total population. Noteworthy, this extra purification step caused a reduced thermostability for the protein as compared to the engineered mutant (hAQP10-N133Q) (Figure 3.16). A possible explanation could be that during the lectin binding, the protein lost a stabilizing lipid. Such a lipid could for example be the lipid present in the central pore of the aquaporin tetramer, as seen in the hAQP5 crystal structure (Horsefield *et al.*, 2008).

3.2.3 Structural Characterization of Aquaporins

The three dimensional structure of a protein can provide detailed insight into the protein's mechanism and regulation. Here, we have utilized crystallography and x-ray diffraction to determine the arrangement of atoms within a protein crystal. Thus, the initial task of growing such crystals was attempted for hAQP1 (**Paper I**), for PfAQP (**Paper II**), for hAQP10 (**Paper V**), and for a hAQP5 mutant (**Paper VI**).

For hAQP1, we were initially unable to obtain protein crystals. A new construct, lacking the purification tags, was cloned to remove the possibilities for the affinity tag to affect the solubility, aggregation or crystallization in a negative way. We obtained overproduction yields in the same range as for the previous construct but, in contrast, diffracting crystals were recovered (**Paper I**).

The situation for PfAQP was different; the main obstacle was to overcome the production phase and once that had been accomplished, crystals could be obtained from crystallization screens (**Paper II**). The construct had the same purification tags as the hAQP1 protein which did not crystallize (**Paper I**), hence variation in result from similar constructs has been observed. Thus, the construct is not only important for the overproduction yield, but also for crystallization. Affinity tags do not necessarily impact the crystallization or the three dimensional structure in a negative or non-biological way. Indeed, they have been shown to enhance the resolution (Carson *et al.*, 2007) or reduce the twinning of aquaporin crystals (Backmark *et al.*, 2008). Nevertheless, no straightforward method of predicting the constructs impact on the formation of protein crystals has yet been presented.

For hAQP10 the wild type protein, which was partly glycosylated, failed to form crystals. As ordered protein crystals are more likely to be formed from a homogenous

sample, the engineered mutant of hAQP10, lacking the possibility to acquire a glycan, was used in subsequent trials. Indeed, we reproducibly obtained diffracting crystals, both in small volumes from robot setups and in larger volumes from hanging drop experiments. Thus, this study emphasizes the importance of knowing the heterogeneity of your protein sample, and the presence of glycans, as it can affect your experiments in a negative way (**Paper V**).

In terms of obtaining diffracting crystals, our attempts on the hAQP5-S156E mutant were the most successful. From one crystal, we could collect high resolution diffraction data which we processed to 2.6Å resolution (Figure 3.17A) (**Paper VI**). The reason for such a mutation is that phosphorylation of serine 156 has been suggested to be involved in protein trafficking (Chapter 1.8.2). Since glutamate supposedly mimics a phosphorylated serine, we aimed to structurally evaluate the importance of a phosphorylation at this site for the trafficking of AQP5.

The overall structure of the hAQP5-S156E mutant was strikingly similar to the wild type hAQP5 (Horsefield et al., 2008) (Figure 3.17B). It has been speculated that the phosphorylation of Ser156 in loop D would affect the interaction between this loop and the C-terminus. Such a change could relieve the forces keeping the carboxyl terminus locked in a fixed position. Hence, allowing it to take a different conformation and flag the protein for trafficking. However, loop D retained its native conformation and no structural changes were observed. Thus, phosphorylation of Ser156 is unlikely to be the triggering factor for protein trafficking.

Moreover, the loop region between TMD6 and the C-terminal helix showed a different conformation (Figure 3.17C). In addition, a variation in the location of the side chains of Asn80, Gln81 in loop B was observed. Consequently, new interactions between the C-terminus and loop B were seen in the hAQP5-S156E structure. The distance between Asn80 and Arg235 decreased from 4.3Å till 2.7Å, thus enabling the creation of a hydrogen bond. Moreover, Gln81 can interact with Asn288 *via* a water molecule not present in the wild type structure (Figure 3.17C). In conclusion, the C-terminus is stabilized by the new conformation of the loop region and the amino acids in the close proximity. The observed structural changes in the C-terminus is likely not due to the mutation, as no structural connection has been established.

In the wild type hAQP5 structure, the C-terminus of the four monomers was found to adopt two different conformations, whereas in the hAQP5-S156E structure, all eight monomers had the same. The difference is a backbone flip of two amino acids: Glu244 and Pro245 (Figure 3.17D). Such a flip could have a large impact on the location of the carboxyl terminus. Unfortunately, this could not be assessed since the terminus is disordered and no structure is visible after Pro245. In addition, the protein sequence has been truncated at this location in the high resolution hAQP5-S156E structure, as the full length mutant construct did not diffract better than about 3.6Å. However, the

low resolution structure supports the observation that all monomers have the same conformation.

Taken together, we were not able to find any structural support for involvement of Ser156 in protein trafficking. Instead, the multiple conformations possible for the C-terminus could be a possible signal resulting in trafficking of the protein to the apical membrane.

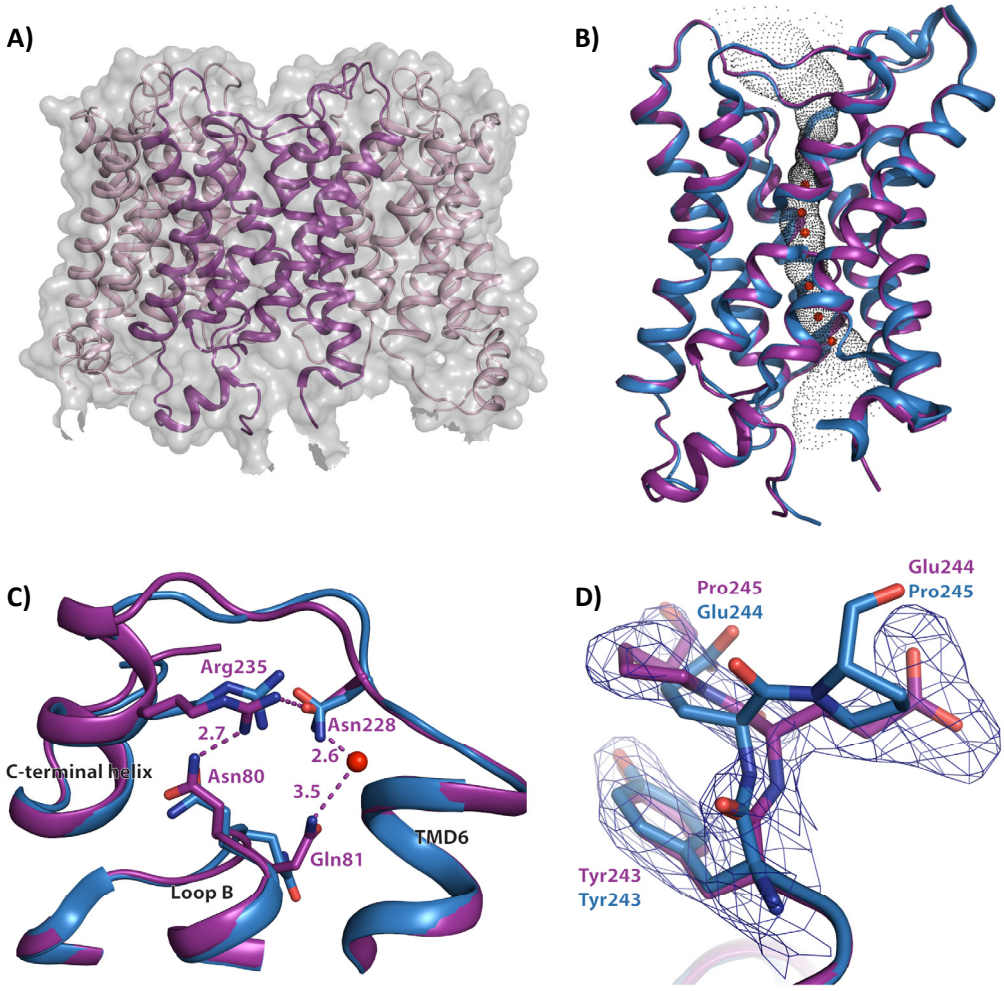


Figure 3.17. Structural comparison of wild type hAQP5 (blue) and hAQP5-S156E (purple, **Paper VI**). A) The crystal structure of hAQP5-S156E showing the tetramer. The surface accessible to solvent or lipids is contoured in grey. B) Overlay of one monomer from each structure with the water pore contoured in black. Water molecules in the pore are from the mutant structure. C) Differences in the loop region after TMD6 causing new interactions to form in the mutant structure (labeled with the distance). D) View of the C-terminal flip. Electron density for the mutant ($2F_{obs}$-F_{calc} map contoured at 0.8 σ) is also shown.

In conclusion, we obtained crystals for hAQP1, PfAQP and hAQP10 providing promising leads for structural characterization. Furthermore, we obtained diffracting crystals and were able to determine the structure of a hAQP5 mutant providing information to how protein trafficking occurs within a cell.

4 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Membrane proteins are key players in all living cells and they constitute the majority of drug targets. Although they are a highly relevant class of proteins, they are generally poorly characterized and underrepresented in structure databases. Thus, a better understanding of the properties and function of membrane proteins is essential to gain deeper insight into the action of existing drugs as well as to aid the development of new treatment strategies. For this purpose, a significant amount of purified protein is required to study the molecular mechanisms and physiological functions of key players in the cellular membranes. Eukaryotic membrane protein production is commonly treated as a matter of trial-and-error, limiting the number of available targets to be studied. This approach does not match the demands on detailed protein characterization and structure determination of highly interesting targets, as they require large amount of purified protein. Hence, many research groups would benefit from improved and controlled production of this group of molecules.

The scope of this thesis was to use a family of highly homologous aquaporins to understand the bottleneck to high recombinant production and use that information to optimize the production of any target. When a high production yield had been achieved, characterization, including functional and structural analysis, was performed.

4.1 PRODUCTION OF AQUAPORINS

To determine the initial production yield, we have developed screening methods for fast and accurate estimation. By using these screens we can estimate our final yield and evaluate whether it will be sufficient for downstream characterization or if optimizations needs to be undertaken.

The human aquaporins are highly homologous proteins, but significant variations in production yield were nevertheless observed for them. From these initial results, we have investigated and analysed several factors of importance for overproduction: the transformation method, amino acid mutation effects, protein stability, sequence flanking the initiator methionine, fusion partners, GC content, and codon adaptation. Moreover, the specificity and function of the protein and fraction of membrane insertion have been studied. To some extent, the observed variations in production yields among homologous proteins still remain elusive. However, when we know the factors directing the high production of a specific eukaryotic membrane protein, they could be used to drive high yielding protein production in future experiments.

Our results establish the possibility to quantitate the overproduction yields of recombinant eukaryotic membrane proteins in a suitable host. This is an essential tool of great use in identifying complex factors which can influence the final protein yield.

Besides engineering of the DNA, another aspect to consider is the optimization of growth conditions to maximize the production yield. The optimal growth conditions

for the host do not necessarily correlate with the optimal conditions for protein production, something that is more pronounced for integral membrane proteins as compared to their soluble counterparts. Although several growth parameters have been analysed in this thesis, a systematic approach with multiple identical fermentors would be beneficial to elucidate the importance of different growth conditions, both for specific targets as well as for membrane proteins in general.

Yet another way of improving the final yield would be to adapt the host for high protein production. For example, stress responses can be triggered within the host cell causing degradation of protein and lack of chaperones and other important proteins for the folding machinery could cause an overload also leading to degradation of your protein. Quite some efforts have been made in the fields of finding optimal growth conditions and strain development. However, as in our study, they are developed for a specific protein family or to a subset of different proteins. Thus, to identify general factors valid for all kinds of proteins is a great and challenging task. We have gained insight into factors directing high production of human aquaporins in *P. pastoris*, and future work can hopefully be based upon this to find even more, possibly unknown, aspects of protein production.

4.2 CHARACTERIZATION OF AQUAPORINS

The characterization presented in this thesis have aimed to investigate the functionality, substrate specificity, secondary structure content, thermostability, ability to form crystals, and the three dimensional structure of the overproduced aquaporins. Verifying the functionality of the aquaporin is important, since this establishes that exceptional high protein yields are not necessarily incompatible with the biological function, folding, and proper localisation of the protein. Furthermore, analyzing the substrate specificity of an aquaporin can give valuable clues to its function and relevance *in vivo*. There are several aquaporins with unknown or controversial substrate specificity which will require substantially more attention to unravel the different possibilities of conductance.

Aquaporins are assembled into tetramers, thereby enabling stabilizing interactions to take place between the monomers. Furthermore, some of the aquaporins have been found to be glycosylated, which could be yet another way to increase their stability. This could have implications for the half life of the protein as well as providing a protection in a harsh milieu, as found within the intestine. Possibly, assemblies of proteins can convoy the positive properties of a glycan to the whole complex, something that needs further assessments.

From the crystal structure of hAQP5-S156E, we have shown that phosphorylation of serine 156 is not likely the cause of protein trafficking. Instead, several conformation changes were observed for the C-terminus, suggesting a flexible domain with multiple possible locations. In particular, a flip of two amino acids in the C-terminus (Glu244 and Pro245) would cause a major rearrangement of the remaining amino acids of the

protein. Further studies would be needed to find exactly what residues, posttranslational events and/or interactions that are necessary for aquaporin regulation by trafficking. As the literature is inconclusive about the role of the consensus sites within the protein, future studies on interaction partners (targeting the c-terminus) would be of particular interest.

The ultimate goal would be to determine the three dimensional structure of all the members of the aquaporin family and combine that information with the biochemical data to learn more about the function, regulation, and the physiological role in a wider context. This information would be fundamental for understanding the processes happening in the human body, and why they sometimes fail. Hence, this could provide new insight into diseases and disorders affecting millions of people, and hopefully provide cures for the people of today and for coming generations.

ACKNOWLEDGEMENTS

To complete a PhD requires considerable effort, but it is a journey worth doing! During my years here at the Lundberg laboratory I have met many new people and acquired many new experiences.

First and foremost I would like to thank my main supervisor **Kristina**. I started as a diploma worker in 2004 and apparently I did not scare you off completely. After that, I got the possibility to continue, first on short term contracts and finally as a PhD student. As a researcher, you have introduced me to the ways of thinking and conducting experiments. However, just mentioning the science would not be sufficient; you have always considered my education as a scientist as an important part of my years in the lab. Looking at other PhD students, this is something unique and a great ability to have as a supervisor. Thank you for guiding me through this journey!

I would also like to thank my boss and co-supervisor **Richard** for letting me join his group when I started my PhD studies. I very much enjoyed discussing science with you, even though many of the discussions took place over a bear in a sauna or abroad on some conference. I have very much enjoyed being a part of your group!

Furthermore, I would also like to thank **Roslyn** at Aston University in Birmingham, UK. I have had the pleasure to work in your lab for a few months which I greatly appreciate. Thank you for always taking your time to discuss – both the 'trivial' problems found in everyday lab work and to formulate comprehensive scientific research questions! In Roslyn's lab I also met **Mohammed**, **Richard** and more recently **Matt**. It was great to work with you all! And Mohammed, I think it is time for some pool or bowling again soon!

Apart from the above mentioned seniors, two persons have helped me a lot with the most recent work: **Susanna** and **Anders**. Thank you for the all the help on those manuscripts!

Big thanks to all the former and current PhD students in the group! Some of you have already left, **Maria**, **Pontus**, **Annemarie**, **Magnus**, **Anna**, but it was fun while you were around! I think it is about time to reset the Beer Club points and start over, don't you agree? :) I miss all friendly insults from you Maria, those were some good late evening in the lab!

A few of you started around the same time as I did: **Gerhard**, **Anna** and **Erik**. It has been great to work with you! Gerhard, you learn Swedish amazingly fast and I think you will never forget the classic Swedish phrase "Mot Sherwoodskogen!" :) Gerhard, Anna and **Sylwia**, thanks for the nice trip to Japan, and especially to Sylwia for following me up on mount Fuji! And a big cheers to the future Beer Club king and

queen: **David** and **Linda**. Good luck with collecting all the debts and in taking the first place on the BC list!

I also had the privilege to be supervisor for a few master students: **Mikael**, **Jennie** and **Eskil**. Thanks for letting me be bossy :) As both Jennie and Mikael continued as PhD students, I can't have been all that bad (I hope!). And Jennie, sorry for all those "Jaha, du gör så?", I never realized until you told me :) Mikael, we should try and find more conferences in Greece :). Anyway, thanks for your hard work and good luck with the rest of your PhD studies!

Thanks to everyone else in the corridor, past as well as present members. You are great to have around to ask questions or to discuss things with: **Rob**, **Euan**, **Gergely**, **Sebastian**, **Rosie**, **Urszula**, **Karin**, **Karin**, **Nicklas**, **Elena**, **Etienne**, **Annette**, **Weixiao**, **Ida**, **Mike**, **Johan**, **Linnéa**, **Maria**, **Karin**, **Madde**.

Also thanks to the people I supervised for shorter projects: **David**, **Leif** and **Anna**. You will probably soon be stuck in PhDs all of you :)

Avslutningsvis, ett stort tack min familj: **Mamma**, **Pappa**, **Marie**, **Farfar**, **Mormor**, **Anna-Karin**, **Ullrika**, **Henrik**. Jag tror inte någon av er blev speciellt förvånad när jag inte kunde slita mig från skolan och fortsatte med en forskarutbildning :) Tack för allt stöd!

Till sist, ett jättetack till **Hanna**, min snygging och älskling!

REFERENCES

Agre, P. (1997) Molecular physiology of water transport: aquaporin nomenclature workshop. Mammalian aquaporins. *Biol Cell,* 89**,** 255-7.

Agre, P. (2006) The aquaporin water channels. *Proc Am Thorac Soc,* 3**,** 5-13.

Agre, P., Saboori, A. M., Asimos, A. and Smith, B. L. (1987) Purification and partial characterization of the Mr 30,000 integral membrane protein associated with the erythrocyte Rh(D) antigen. *J Biol Chem,* 262**,** 17497-503.

Agre, P., Sasaki, S. and Chrispeels, M. J. (1993) Aquaporins: a family of water channel proteins. *Am J Physiol,* 265**,** F461.

Akashi, H. (2001) Gene expression and molecular evolution. *Curr Opin Genet Dev,* 11**,** 660-6.

Alberts, B. (2002) Molecular biology of the cell*,* New York, *Garland Science.*

Amiry-Moghaddam, M., Williamson, A., Palomba, M., Eid, T., de Lanerolle, N. C., Nagelhus, E. A., Adams, M. E., Froehner, S. C., Agre, P. and Ottersen, O. P. (2003) Delayed K+ clearance associated with aquaporin-4 mislocalization: phenotypic defects in brains of alpha-syntrophin-null mice. *Proc Natl Acad Sci U S A,* 100**,** 13615-20.

Apweiler, R., Hermjakob, H. and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta,* 1473**,** 4-8.

Atkinson, P. H. and Lee, J. T. (1984) Co-translational excision of alpha-glucose and alpha-mannose in nascent vesicular stomatitis virus G protein. *J Cell Biol,* 98**,** 2245-9.

Backmark, A., Nyblom, M., Tornroth-Horsefield, S., Kosinska-Eriksson, U., Norden, K., Fellert, M., Kjellbom, P., Johanson, U., Hedfalk, K., Lindkvist-Petersson, K., Neutze, R. and Horsefield, R. (2008) Affinity tags can reduce merohedral twinning of membrane protein crystals. *Acta Crystallogr D Biol Crystallogr,* 64**,** 1183-1186.

Badaut, J., Petit, J. M., Brunet, J. F., Magistretti, P. J., Charriaut-Marlangue, C. and Regli, L. (2004) Distribution of Aquaporin 9 in the adult rat brain: preferential expression in catecholaminergic neurons and in glial cells. *Neuroscience,* 128**,** 27-38.

Bai, C., Fukuda, N., Song, Y., Ma, T., Matthay, M. A. and Verkman, A. S. (1999) Lung fluid transport in aquaporin-1 and aquaporin-4 knockout mice. *J Clin Invest,* 103**,** 555-61.

Baim, S. B. and Sherman, F. (1988) mRNA structures influencing translation in the yeast Saccharomyces cerevisiae. *Mol Cell Biol,* 8**,** 1591-601.

Ballard, S. T., Trout, L., Bebok, Z., Sorscher, E. J. and Crews, A. (1999) CFTR involvement in chloride, bicarbonate, and liquid secretion by airway submucosal glands. *Am J Physiol,* 277**,** L694-9.

Beitz, E. (2005) Aquaporins from pathogenic protozoan parasites: structure, function and potential for chemotherapy. *Biol Cell,* 97**,** 373-83.

Beitz, E., Wu, B., Holm, L. M., Schultz, J. E. and Zeuthen, T. (2006) Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons. *Proc Natl Acad Sci U S A,* 103**,** 269-74.

Benga, G., Popescu, O., Pop, V. I. and Holmes, R. P. (1986) p-(Chloromercuri)benzenesulfonate binding by membrane proteins and the inhibition of water transport in human erythrocytes. *Biochemistry,* 25**,** 1535-8.

Bergeron, J. J., Brenner, M. B., Thomas, D. Y. and Williams, D. B. (1994) Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. *Trends Biochem Sci,* 19**,** 124-8.

Berry, V., Francis, P., Kaushal, S., Moore, A. and Bhattacharya, S. (2000) Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nat Genet,* 25**,** 15-7.

Bill, R. M., Henderson, P. J., Iwata, S., Kunji, E. R., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G. and Vogel, H. (2011) Overcoming barriers to membrane protein structure determination. *Nat Biotechnol,* 29**,** 335-340.

Bill, R. M., Revers, L. and Wilson, I. B. H. (1998) Protein glycosylation*,* Boston ; London, *Kluwer Academic Publishers.*

Bleasby, A. (2001). EMBOSS: cai [Online]. *Available: http://emboss.bioinformatics.nl/cgi-bin/emboss/cai [Accessed 25 Mar 2011].*

Bonander, N. and Bill, R. M. (2009) Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalize membrane protein production. *Expert Rev Proteomics,* 6**,** 501-5.

Bonander, N., Hedfalk, K., Larsson, C., Mostad, P., Chang, C., Gustafsson, L. and Bill, R. M. (2005) Design of improved membrane protein production experiments: quantitation of the host response. *Protein Sci,* 14**,** 1729-40.

Bretscher, M. S. and Raff, M. C. (1975) Mammalian plasma membranes. *Nature,* 258**,** 43-9.

Brown, P. A., Feinstein, M. B. and Sha'afi, R. I. (1975) Membrane proteins related to water transport in human erythrocytes. *Nature,* 254**,** 523-5.

Buck, T. M., Wagner, J., Grund, S. and Skach, W. R. (2007) A novel tripartite motif involved in aquaporin topogenesis, monomer folding and tetramerization. *Nat Struct Mol Biol,* 14**,** 762-9.

Calamita, G., Mazzone, A., Bizzoca, A., Cavalier, A., Cassano, G., Thomas, D. and Svelto, M. (2001) Expression and immunolocalization of the aquaporin-8 water channel in rat gastrointestinal tract. *Eur J Cell Biol,* 80**,** 711-9.

Carbrey, J. M. and Agre, P. (2009) Discovery of the aquaporins and development of the field. *Handb Exp Pharmacol***,** 3-28.

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C. and Delucas, L. J. (2007) His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr,* 63**,** 295-301.

Castle, N. A. (2005) Aquaporins as targets for drug discovery. *Drug Discov Today,* 10**,** 485-93.

Chepelinsky, A. B. (2009) Structural function of MIP/aquaporin 0 in the eye lens; genetic defects lead to congenital inherited cataracts. *Handb Exp Pharmacol***,** 265-97.

Cigan, A. M. and Donahue, T. F. (1987) Sequence and structural features associated with translational initiator regions in yeast--a review. *Gene,* 59**,** 1-18.

Cohly, H. H., Isokpehi, R. and Rajnarayanan, R. V. (2008) Compartmentalization of aquaporins in the human intestine. *International journal of environmental research and public health,* 5**,** 115-9.

Conner, M. T., Conner, A. C., Brown, J. E. and Bill, R. M. (2010) Membrane Trafficking of Aquaporin 1 Is Mediated by Protein Kinase C via Microtubules and Regulated by Tonicity. *Biochemistry.*

Cornish-Bowden, A. (1985) Nomenclature Committee of the International Union of Biochemistry (NC-IUB). Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *Biochem J,* 229**,** 281-6.

Couderc, R. and Baratti, J. (1980) Oxidation of Methanol by the Yeast, Pichia pastoris. Purification and Properties of the Alcohol Oxidase. *Agric biol chem,* 44**,** 2279-2289.

Cregg, J. M. (2007) Pichia protocols*,* Totowa, N.J., *Humana ; [Paisley : Quantum, distributor].*

Cregg, J. M. and Barringer, K. J. (1990) Pichia transformation. *United States patent application 07/110148*

Cregg, J. M., Cereghino, J. L., Shi, J. and Higgins, D. R. (2000) Recombinant protein expression in Pichia pastoris. *Mol Biotechnol,* 16**,** 23-52.

Cregg, J. M., Tschopp, J. F., Stillman, C., Siegel, R., Akong, M., Craig, W. S., Buckholz, R. G., Madden1, K. R., Kellaris, P. A., Davis, G. R., Smiley, B. L., Cruze, J., Torregrossa, R., Velicecedillebi, G. and Thill, G. P. (1987) High–Level Expression and Efficient Assembly of Hepatitis B Surface Antigen in the Methylotrophic Yeast, Pichia Pastoris. *Nat Biotechnol,* 5**,** 479-485.

Daly, R. and Hearn, M. T. (2005) Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *J Mol Recognit,* 18**,** 119-38.

de Groot, B. L. and Grubmuller, H. (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science,* 294**,** 2353-7.

de Groot, B. L. and Grubmuller, H. (2005) The dynamics and energetics of water permeation and proton exclusion in aquaporins. *Curr Opin Struct Biol,* 15**,** 176-83.

De Schutter, K., Lin, Y. C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouze, P., Van de Peer, Y. and Callewaert, N. (2009) Genome sequence of the recombinant protein production host Pichia pastoris. *Nat Biotechnol,* 27**,** 561-6.

Deen, P. M., Verdijk, M. A., Knoers, N. V., Wieringa, B., Monnens, L. A., van Os, C. H. and van Oost, B. A. (1994) Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science,* 264**,** 92-5.

Denker, B. M., Smith, B. L., Kuhajda, F. P. and Agre, P. (1988) Identification, purification, and partial characterization of a novel Mr 28,000 integral membrane protein from erythrocytes and renal tubules. *J Biol Chem,* 263**,** 15634-42.

Diegelmann, S., Nieratschker, V., Werner, U., Hoppe, J., Zars, T. and Buchner, E. (2006) The conserved protein kinase-A target motif in synapsin of Drosophila is effectively modified by pre-mRNA editing. *BMC Neurosci,* 7**,** 76.

Dumas, M., Sadick, N. S., Noblesse, E., Juan, M., Lachmann-Weber, N., Boury-Jamot, M., Sougrat, R., Verbavatz, J. M., Schnebert, S. and Bonte, F. (2007) Hydrating skin by stimulating biosynthesis of aquaporins. *J Drugs Dermatol,* 6**,** s20-4.

Echevarria, M., Windhager, E. E., Tate, S. S. and Frindt, G. (1994) Cloning and expression of AQP3, a water channel from the medullary collecting duct of rat kidney. *Proc Natl Acad Sci U S A,* 91**,** 10997-1001.

Elbein, A. D. (1991) The role of N-linked oligosaccharides in glycoprotein function. *Trends in biotechnology,* 9**,** 346-52.

Elkjaer, M., Vajda, Z., Nejsum, L. N., Kwon, T., Jensen, U. B., Amiry-Moghaddam, M., Frokiaer, J. and Nielsen, S. (2000) Immunolocalization of AQP9 in liver, epididymis, testis, spleen, and brain. *Biochem Biophys Res Commun,* 276**,** 1118-28.

European Membrane Protein Consortium. (2009). E-MeP Progress [Online]. *Available: http://www.e-mep.org/results/index.htm [Accessed 18 Feb 2011].*

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M. and Berglund, L. (2010) Prediction of the human membrane proteome. *Proteomics,* 10**,** 1141-9.

Fantoni, A., Bill, R. M., Gustafsson, L. and Hedfalk, K. (2007) Improved yields of full-length functional human FGF1 can be achieved using the methylotrophic yeast Pichia pastoris. *Protein Expr Purif,* 52**,** 31-9.

Fischer, G., Kosinska-Eriksson, U., Aponte-Santamaria, C., Palmgren, M., Geijer, C., Hedfalk, K., Hohmann, S., de Groot, B. L., Neutze, R. and Lindkvist-Petersson, K. (2009) Crystal structure of a yeast aquaporin at 1.15 angstrom reveals a novel gating mechanism. *PLoS Biol,* 7**,** e1000130.

Flach, C. F., Qadri, F., Bhuiyan, T. R., Alam, N. H., Jennische, E., Holmgren, J. and Lonnroth, I. (2007) Differential expression of intestinal membrane transporters in cholera patients. *FEBS letters,* 581**,** 3183-8.

Forstner, M., Leder, L. and Mayr, L. M. (2007) Optimization of protein expression systems for modern drug discovery. *Expert Rev Proteomics,* 4**,** 67-78.

Foster, W., Helm, A., Turnbull, I., Gulati, H., Yang, B., Verkman, A. S. and Skach, W. R. (2000) Identification of sequence determinants that direct different intracellular folding pathways for aquaporin-1 and aquaporin-4. *J Biol Chem,* 275**,** 34157-65.

Fotiadis, D., Jeno, P., Mini, T., Wirtz, S., Muller, S. A., Fraysse, L., Kjellbom, P. and Engel, A. (2001) Structural characterization of two aquaporins isolated from native spinach leaf plasma membranes. *J Biol Chem,* 276**,** 1707-14.

Frey, A., Giannasca, K. T., Weltzin, R., Giannasca, P. J., Reggio, H., Lencer, W. I. and Neutra, M. R. (1996) Role of the glycocalyx in regulating access of microparticles to apical plasma membranes of intestinal epithelial cells: implications for microbial attachment and oral vaccine targeting. *J Exp Med,* 184**,** 1045-59.

Fushimi, K., Sasaki, S. and Marumo, F. (1997) Phosphorylation of serine 256 is required for cAMP-dependent regulatory exocytosis of the aquaporin-2 water channel. *J Biol Chem,* 272**,** 14800-4.

Fushimi, K., Uchida, S., Hara, Y., Hirata, Y., Marumo, F. and Sasaki, S. (1993) Cloning and expression of apical membrane water channel of rat kidney collecting tubule. *Nature,* 361**,** 549-52.

Garcia, F., Kierbel, A., Larocca, M. C., Gradilone, S. A., Splinter, P., LaRusso, N. F. and Marinelli, R. A. (2001) The water channel aquaporin-8 is mainly intracellular in rat hepatocytes, and its plasma membrane insertion is stimulated by cyclic AMP. *J Biol Chem,* 276**,** 12147-52.

GE Healthcare. (2011). Products for Protein Purification [Online]. *Available: http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&parentid=572&moduleid=13277&zone=Labsep [Accessed 23 Mar 2011].*

Gemmill, T. R. and Trimble, R. B. (1999) Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica et biophysica acta,* 1426**,** 227-37.

GeneArt. (2011). Gene Synthesis [Online]. *Available: http://www.geneart.com/english/products-services/gene-synthesis/index.html [Accessed 25 Mar 2011].*

GenScript. (2011). Gene Synthesis [Online]. *Available: http://www.genscript.com/gene_synthesis.html [Accessed 25 Mar 2011].*

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C. and Walz, T. (2005) Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature,* 438**,** 633-8.

Gonen, T., Sliz, P., Kistler, J., Cheng, Y. and Walz, T. (2004) Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature,* 429**,** 193-7.

Gong, Q., Anderson, C. L., January, C. T. and Zhou, Z. (2002) Role of glycosylation in cell surface expression and stability of HERG potassium channels. *American journal of physiology. Heart and circulatory physiology,* 283**,** H77-84.

Gorelick, D. A., Praetorius, J., Tsunenari, T., Nielsen, S. and Agre, P. (2006) Aquaporin-11: a channel protein lacking apparent transport function expressed in brain. *BMC Biochem,* 7**,** 14.

Gorin, M. B., Yancey, S. B., Cline, J., Revel, J. P. and Horwitz, J. (1984) The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning. *Cell,* 39**,** 49-59.

Griffith, D. A., Delipala, C., Leadsham, J., Jarvis, S. M. and Oesterhelt, D. (2003) A novel yeast expression system for the overproduction of quality-controlled membrane proteins. *FEBS Lett,* 553**,** 45-50.

Grinna, L. S. and Tschopp, J. F. (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, Pichia pastoris. *Yeast,* 5**,** 107-15.

Grisshammer, R. (2006) Understanding recombinant expression of membrane proteins. *Curr Opin Biotechnol,* 17**,** 337-40.

Grisshammer, R. and Tate, C. G. (1995) Overexpression of integral membrane proteins for structural studies. *Q Rev Biophys,* 28**,** 315-422.

Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C. and Jahn, D. (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res,* 33**,** W526-31.

Guilliermond, A. and Tanner, F. W. (1920) The yeasts*,* New York, *John Wiley and Sons ; London : Chapman and Hall.*

Gustafsson, C., Govindarajan, S. and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol,* 22**,** 346-53.

Hamdi, M., Sanchez, M. A., Beene, L. C., Liu, Q., Landfear, S. M., Rosen, B. P. and Liu, Z. (2009) Arsenic transport by zebrafish aquaglyceroporins. *BMC molecular biology,* 10**,** 104.

Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., Nett, J. H., Rausch, S., Stadheim, T. A., Wischnewski, H., Wildt, S. and Gerngross, T. U. (2003) Production of complex human glycoproteins in yeast. *Science,* 301**,** 1244-6.

Hansen, M., Kun, J. F., Schultz, J. E. and Beitz, E. (2002) A single, bi-functional aquaglyceroporin in blood-stage Plasmodium falciparum malaria parasites. *J Biol Chem,* 277**,** 4874-82.

Hara-Chikuma, M., Sohara, E., Rai, T., Ikawa, M., Okabe, M., Sasaki, S., Uchida, S. and Verkman, A. S. (2005) Progressive adipocyte hypertrophy in aquaporin-7-deficient mice: adipocyte glycerol permeability as a novel regulator of fat accumulation. *J Biol Chem,* 280**,** 15493-6.

Hara, M., Ma, T. and Verkman, A. S. (2002) Selectively reduced glycerol in skin of aquaporin-3-deficient mice may account for impaired skin hydration, elasticity, and barrier recovery. *J Biol Chem,* 277**,** 46616-21.

Harmon, B. J., Gu, X. and Wang, D. I. (1996) Rapid monitoring of site-specific glycosylation microheterogeneity of recombinant human interferon-gamma. *Anal Chem,* 68**,** 1465-73.

Harries, W. E., Akhavan, D., Miercke, L. J., Khademi, S. and Stroud, R. M. (2004) The channel architecture of aquaporin 0 at a 2.2-A resolution. *Proc Natl Acad Sci U S A,* 101**,** 14045-50.

Hartley, H. (1951) Origin of the word 'protein'. *Nature,* 168**,** 244.

Hasegawa, H., Ma, T., Skach, W., Matthay, M. A. and Verkman, A. S. (1994) Molecular cloning of a mercurial-insensitive water channel expressed in selected water-transporting tissues. *J Biol Chem,* 269**,** 5497-500.

Hatakeyama, S., Yoshida, Y., Tani, T., Koyama, Y., Nihei, K., Ohshiro, K., Kamiie, J. I., Yaoita, E., Suda, T., Hatakeyama, K. and Yamamoto, T. (2001) Cloning of a new aquaporin (AQP10) abundantly expressed in duodenum and jejunum. *Biochem Biophys Res Commun,* 287**,** 814-9.

Hazeu, W., de Bruyn, J. C. and Bos, P. (1972) Methanol assimilation by yeasts. *Arch Mikrobiol,* 87**,** 185-8.

Hennig, L. and Schafer, E. (1998) Protein purification with C-terminal fusion of maltose binding protein. *Protein Expr Purif,* 14**,** 367-70.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. and von Heijne, G. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature,* 433**,** 377-81.

Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. and von Heijne, G. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature,* 450**,** 1026-30.

Heymann, J. B., Agre, P. and Engel, A. (1998) Progress on the structure and function of aquaporin 1. *Journal of structural biology,* 121**,** 191-206.

Hiroaki, Y., Tani, K., Kamegawa, A., Gyobu, N., Nishikawa, K., Suzuki, H., Walz, T., Sasaki, S., Mitsuoka, K., Kimura, K., Mizoguchi, A. and Fujiyoshi, Y. (2006) Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol,* 355**,** 628-39.

Ho, J. D., Yeh, R., Sandstrom, A., Chorny, I., Harries, W. E., Robbins, R. A., Miercke, L. J. and Stroud, R. M. (2009) Crystal structure of human aquaporin 4 at 1.8 A and its mechanism of conductance. *Proc Natl Acad Sci U S A,* 106**,** 7437-42.

Hohmann, S. and Nielsen, S. N. (2000) Molecular biology and physiology of water and solute transport*,* New York ; London, *Kluwer Academic/Plenum*.

Holmes, W. J., Darby, R. A., Wilks, M. D., Smith, R. and Bill, R. M. (2009) Developing a scalable model of recombinant protein yield from Pichia pastoris: the influence of culture conditions, biomass and induction regime. *Microb Cell Fact,* 8**,** 35.

Horsefield, R., Norden, K., Fellert, M., Backmark, A., Tornroth-Horsefield, S., Terwisscha van Scheltinga, A. C., Kvassman, J., Kjellbom, P., Johanson, U. and Neutze, R. (2008) High-resolution x-ray structure of human aquaporin 5. *Proc Natl Acad Sci U S A,* 105**,** 13327-32.

Invitrogen. (2005). EasySelect Pichia Expression Kit, version G [Online]. *Available: http://www.invitrogen.com/content/sfs/manuals/easyselect_man.pdf [Accessed 14 Feb 2005]*.

Invitrogen. (2010). EasySelect Pichia Expression Kit, Version MAN0000042 [Online]. *Available: http://www.invitrogen.com/content/sfs/manuals/easyselect_man.pdf [Accessed 18 Jun 2010]*.

Ishibashi, K. (2009) New members of mammalian aquaporins: AQP10-AQP12. *Handb Exp Pharmacol***,** 251-62.

Ishibashi, K., Kuwahara, M., Gu, Y., Kageyama, Y., Tohsaka, A., Suzuki, F., Marumo, F. and Sasaki, S. (1997a) Cloning and functional expression of a new water channel abundantly expressed in the testis permeable to water, glycerol, and urea. *J Biol Chem,* 272**,** 20782-6.

Ishibashi, K., Kuwahara, M., Gu, Y., Tanaka, Y., Marumo, F. and Sasaki, S. (1998) Cloning and functional expression of a new aquaporin (AQP9) abundantly expressed in the peripheral leukocytes permeable to water and urea, but not to glycerol. *Biochem Biophys Res Commun,* 244**,** 268-74.

Ishibashi, K., Kuwahara, M., Kageyama, Y., Tohsaka, A., Marumo, F. and Sasaki, S. (1997b) Cloning and functional expression of a second new aquaporin abundantly expressed in testis. *Biochem Biophys Res Commun,* 237**,** 714-8.

Ishibashi, K., Morinaga, T., Kuwahara, M., Sasaki, S. and Imai, M. (2002) Cloning and identification of a new member of water channel (AQP10) as an aquaglyceroporin. *Biochim Biophys Acta,* 1576**,** 335-40.

Ishibashi, K., Sasaki, S., Fushimi, K., Uchida, S., Kuwahara, M., Saito, H., Furukawa, T., Nakajima, K., Yamaguchi, Y., Gojobori, T. and et al. (1994) Molecular cloning and expression of a member of the

aquaporin family with permeability to glycerol and urea in addition to water expressed at the basolateral membrane of kidney collecting duct cells. *Proc Natl Acad Sci U S A,* 91**,** 6269-73.

Ishimura, K., Kurihara, H. and Fujita, H. (1984) Effects of tunicamycin on thin-section and freeze-fracture images of microvilli of the duodenal epithelial cells of the mouse. *Cell and tissue research,* 238**,** 653-6.

Israelidis, C. J. (2004). Nutrition - Single cell protein, twenty years later [Online]. *Available: http://www.biopolitics.gr/HTML/PUBS/VOL1/isreali.htm [Accessed 21 Feb 2011].*

Itoh, T., Rai, T., Kuwahara, M., Ko, S. B., Uchida, S., Sasaki, S. and Ishibashi, K. (2005) Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. *Biochem Biophys Res Commun,* 330**,** 832-8.

Jamieson, J. D. and Palade, G. E. (1968) Intracellular transport of secretory proteins in the pancreatic exocrine cell. IV. Metabolic requirements. *J Cell Biol,* 39**,** 589-603.

Johansson, I., Karlsson, M., Shukla, V. K., Chrispeels, M. J., Larsson, C. and Kjellbom, P. (1998) Water transport activity of the plasma membrane aquaporin PM28A is regulated by phosphorylation. *Plant Cell,* 10**,** 451-9.

Jung, J. S., Bhat, R. V., Preston, G. M., Guggino, W. B., Baraban, J. M. and Agre, P. (1994a) Molecular characterization of an aquaporin cDNA from brain: candidate osmoreceptor and regulator of water balance. *Proc Natl Acad Sci U S A,* 91**,** 13052-6.

Jung, J. S., Preston, G. M., Smith, B. L., Guggino, W. B. and Agre, P. (1994b) Molecular structure of the water channel through aquaporin CHIP. The hourglass model. *J Biol Chem,* 269**,** 14648-54.

Jungo, C., Marison, I. and von Stockar, U. (2007a) Mixed feeds of glycerol and methanol can improve the performance of Pichia pastoris cultures: A quantitative study based on concentration gradients in transient continuous cultures. *J Biotechnol,* 128**,** 824-37.

Jungo, C., Schenk, J., Pasquier, M., Marison, I. W. and von Stockar, U. (2007b) A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with Pichia pastoris. *J Biotechnol,* 131**,** 57-66.

Kaldenhoff, R. and Fischer, M. (2006) Functional aquaporin diversity in plants. *Biochim Biophys Acta,* 1758**,** 1134-41.

Kamsteeg, E. J., Hendriks, G., Boone, M., Konings, I. B., Oorschot, V., van der Sluijs, P., Klumperman, J. and Deen, P. M. (2006) Short-chain ubiquitination mediates the regulated endocytosis of the aquaporin-2 water channel. *Proc Natl Acad Sci U S A,* 103**,** 18344-9.

Karabasil, M. R., Hasegawa, T., Azlina, A., Purwanti, N., Purevjav, J., Yao, C., Akamatsu, T. and Hosoi, K. (2009) Trafficking of GFP-AQP5 chimeric proteins conferred with unphosphorylated amino acids at their PKA-target motif ((152)SRRTS) in MDCK-II cells. *J Med Invest,* 56**,** 55-63.

Kazusa. (2007). Codon Usage Database [Online]. *Available: http://www.kazusa.or.jp/codon/ [Accessed 22 Feb 2011].*

Keller, P. and Simons, K. (1997) Post-Golgi biosynthetic trafficking. *J Cell Sci,* 110 (Pt 24)**,** 3001-9.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature,* 181**,** 662-6.

King, L. S., Choi, M., Fernandez, P. C., Cartron, J. P. and Agre, P. (2001) Defective urinary-concentrating ability due to a complete deficiency of aquaporin-1. *N Engl J Med,* 345**,** 175-9.

King, L. S., Kozono, D. and Agre, P. (2004) From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell. Biol.,* 5**,** 687-698.

Kornfeld, R. and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem,* 54**,** 631-64.

Kosugi-Tanaka, C., Li, X., Yao, C., Akamatsu, T., Kanamori, N. and Hosoi, K. (2006) Protein kinase A-regulated membrane trafficking of a green fluorescent protein-aquaporin 5 chimera in MDCK cells. *Biochim Biophys Acta,* 1763**,** 337-44.

Koyama, Y., Yamamoto, T., Kondo, D., Funaki, H., Yaoita, E., Kawasaki, K., Sato, N., Hatakeyama, K. and Kihara, I. (1997) Molecular cloning of a new aquaporin from rat pancreas and liver. *J Biol Chem,* 272**,** 30329-33.

Kozak, M. (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res,* 9**,** 5233-52.

Kozak, M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem,* 266**,** 19867-70.

Kozak, M. (1992) Regulation of translation in eukaryotic systems. *Annu Rev Cell Biol,* 8**,** 197-225.

Kreda, S. M., Gynn, M. C., Fenstermacher, D. A., Boucher, R. C. and Gabriel, S. E. (2001) Expression and localization of epithelial aquaporins in the adult human lung. *Am J Respir Cell Mol Biol,* 24**,** 224-34.

Kundu, A., Avalos, R. T., Sanderson, C. M. and Nayak, D. P. (1996) Transmembrane domain of influenza virus neuraminidase, a type II protein, possesses an apical sorting signal in polarized MDCK cells. *J Virol,* 70**,** 6508-15.

Kuriyama, H., Kawamoto, S., Ishida, N., Ohno, I., Mita, S., Matsuzawa, Y., Matsubara, K. and Okubo, K. (1997) Molecular cloning and expression of a novel human aquaporin from adipose tissue with glycerol permeability. *Biochem Biophys Res Commun,* 241**,** 53-8.

Lawaczeck, R. (1984) Water permeability through biological membranes by isotopic effects of fluorescence and light scattering. *Biophys J,* 45**,** 491-4.

Li, H., Kamiie, J., Morishita, Y., Yoshida, Y., Yaoita, E., Ishibashi, K. and Yamamoto, T. (2005) Expression and localization of two isoforms of AQP10 in human small intestine. *Biol Cell,* 97**,** 823-9.

Lian, J., Ding, S., Cai, J., Zhang, D., Xu, Z. and Wang, X. (2008) Improving aquaporin Z expression in Escherichia coli by fusion partners and subsequent condition optimization. *Appl Microbiol Biotechnol.*

Lis, H. and Sharon, N. (1993) Protein glycosylation. Structural and functional aspects. *European journal of biochemistry / FEBS,* 218**,** 1-27.

Lisanti, M. P., Le Bivic, A., Saltiel, A. R. and Rodriguez-Boulan, E. (1990) Preferred apical distribution of glycosyl-phosphatidylinositol (GPI) anchored proteins: a highly conserved feature of the polarized epithelial cell phenotype. *J Membr Biol,* 113**,** 155-67.

Liu, Z., Shen, J., Carbrey, J. M., Mukhopadhyay, R., Agre, P. and Rosen, B. P. (2002) Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. *Proc Natl Acad Sci U S A,* 99**,** 6053-8.

Lodish, H. F. (2000) Molecular cell biology*,* New York, *W. H. Freeman and Co.*

Lu, M., Lee, M. D., Smith, B. L., Jung, J. S., Agre, P., Verdijk, M. A., Merkx, G., Rijss, J. P. and Deen, P. M. (1996) The human AQP4 gene: definition of the locus encoding two water channel polypeptides in brain. *Proc Natl Acad Sci U S A,* 93**,** 10908-12.

Lu, Y., Turnbull, I. R., Bragin, A., Carveth, K., Verkman, A. S. and Skach, W. R. (2000) Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol Biol Cell,* 11**,** 2973-85.

Lundstrom, K. (2006) Structural genomics: the ultimate approach for rational drug design. *Mol Biotechnol,* 34**,** 205-12.

Ma, T., Frigeri, A., Hasegawa, H. and Verkman, A. S. (1994) Cloning of a water channel homolog expressed in brain meningeal cells and kidney collecting duct that functions as a stilbene-sensitive glycerol transporter. *J Biol Chem,* 269**,** 21845-9.

Ma, T., Frigeri, A., Skach, W. and Verkman, A. S. (1993) Cloning of a novel rat kidney cDNA homologous to CHIP28 and WCH-CD water channels. *Biochem Biophys Res Commun,* 197**,** 654-9.

Ma, T., Song, Y., Gillespie, A., Carlson, E. J., Epstein, C. J. and Verkman, A. S. (1999) Defective secretion of saliva in transgenic mice lacking aquaporin-5 water channels. *J Biol Chem,* 274**,** 20071-4.

Ma, T., Song, Y., Yang, B., Gillespie, A., Carlson, E. J., Epstein, C. J. and Verkman, A. S. (2000) Nephrogenic diabetes insipidus in mice lacking aquaporin-3 water channels. *Proc Natl Acad Sci U S A,* 97**,** 4386-91.

Ma, T., Yang, B., Gillespie, A., Carlson, E. J., Epstein, C. J. and Verkman, A. S. (1998) Severely impaired urinary concentrating ability in transgenic mice lacking aquaporin-1 water channels. *J Biol Chem,* 273**,** 4296-9.

Ma, T., Yang, B., Kuo, W. L. and Verkman, A. S. (1996) cDNA cloning and gene structure of a novel water channel expressed exclusively in human kidney: evidence for a gene cluster of aquaporins at chromosome locus 12q13. *Genomics,* 35**,** 543-50.

Ma, T., Yang, B. and Verkman, A. S. (1997) Cloning of a novel water and urea-permeable aquaporin from mouse expressed strongly in colon, placenta, liver, and heart. *Biochem Biophys Res Commun,* 240**,** 324-8.

Macey, R. I. (1984) Transport of water and urea in red blood cells. *Am J Physiol,* 246**,** C195-203.

MacIver, B., Cutler, C. P., Yin, J., Hill, M. G., Zeidel, M. L. and Hill, W. G. (2009) Expression and functional characterization of four aquaporin water channels from the European eel (Anguilla anguilla). *J Exp Biol,* 212**,** 2856-63.

Maeda, N., Funahashi, T., Hibuse, T., Nagasawa, A., Kishida, K., Kuriyama, H., Nakamura, T., Kihara, S., Shimomura, I. and Matsuzawa, Y. (2004) Adaptation to fasting by glycerol transport through aquaporin 7 in adipose tissue. *Proc Natl Acad Sci U S A,* 101**,** 17801-6.

Maeda, N., Hibuse, T. and Funahashi, T. (2009) Role of aquaporin-7 and aquaporin-9 in glycerol metabolism; involvement in obesity. *Handb Exp Pharmacol***,** 233-49.

Marshall, R. D. (1974) The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp***,** 17-26.

Messner, P. (2004) Prokaryotic glycoproteins: unexplored but important. *J Bacteriol,* 186**,** 2517-9.

Mobasheri, A., Shakibaei, M. and Marples, D. (2004) Immunohistochemical localization of aquaporin 10 in the apical membranes of the human ileum: a potential pathway for luminal water and small solute absorption. *Histochemistry and cell biology,* 121**,** 463-71.

Moore, M., Ma, T., Yang, B. and Verkman, A. S. (2000) Tear secretion by lacrimal glands in transgenic mice lacking water channels AQP1, AQP3, AQP4 and AQP5. *Exp Eye Res,* 70**,** 557-62.

Morinaga, T., Nakakoshi, M., Hirao, A., Imai, M. and Ishibashi, K. (2002) Mouse aquaporin 10 gene (AQP10) is a pseudogene. *Biochemical and biophysical research communications,* 294**,** 630-4.

Morishita, Y., Matsuzaki, T., Hara-chikuma, M., Andoo, A., Shimono, M., Matsuki, A., Kobayashi, K., Ikeda, M., Yamamoto, T., Verkman, A., Kusano, E., Ookawara, S., Takata, K., Sasaki, S. and Ishibashi, K. (2005) Disruption of aquaporin-11 produces polycystic kidneys following vacuolization of the proximal tubule. *Mol Cell Biol,* 25**,** 7770-9.

Mulders, S. M., Preston, G. M., Deen, P. M., Guggino, W. B., van Os, C. H. and Agre, P. (1995) Water channel properties of major intrinsic protein of lens. *J Biol Chem,* 270**,** 9010-16.

Munro, I. C., Berndt, W. O., Borzelleca, J. F., Flamm, G., Lynch, B. S., Kennepohl, E., Bar, E. A. and Modderman, J. (1998) Erythritol: an interpretive summary of biochemical, metabolic, toxicological and

clinical data. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association,* 36**,** 1139-74.

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A. and Fujiyoshi, Y. (2000) Structural determinants of water permeation through aquaporin-1. *Nature,* 407**,** 599-605.

Mus-Veteau, I. (2002) Heterologous expression and purification systems for structural proteomics of mammalian membrane proteins. *Comp Funct Genomics,* 3**,** 511-7.

Nakagawa, H., Wakabayashi-Nakao, K., Tamura, A., Toyoda, Y., Koshiba, S. and Ishikawa, T. (2009) Disruption of N-linked glycosylation enhances ubiquitin-mediated proteasomal degradation of the human ATP-binding cassette transporter ABCG2. *The FEBS journal,* 276**,** 7237-52.

Nedvetsky, P. I., Tamma, G., Beulshausen, S., Valenti, G., Rosenthal, W. and Klussmann, E. (2009) Regulation of aquaporin-2 trafficking. *Handb Exp Pharmacol***,** 133-57.

Neely, J. D., Amiry-Moghaddam, M., Ottersen, O. P., Froehner, S. C., Agre, P. and Adams, M. E. (2001) Syntrophin-dependent expression and localization of Aquaporin-4 water channel protein. *Proc Natl Acad Sci U S A,* 98**,** 14108-13.

Nejsum, L. N., Kwon, T. H., Jensen, U. B., Fumagalli, O., Frokiaer, J., Krane, C. M., Menon, A. G., King, L. S., Agre, P. C. and Nielsen, S. (2002) Functional requirement of aquaporin-5 in plasma membranes of sweat glands. *Proc Natl Acad Sci U S A,* 99**,** 511-6.

Neuberger, G., Schneider, G. and Eisenhaber, F. (2007) pkaPS: prediction of protein kinase A phosphorylation sites with the simplified kinase-substrate binding model. *Biol Direct,* 2**,** 1.

Nielsen, S., Nagelhus, E. A., Amiry-Moghaddam, M., Bourque, C., Agre, P. and Ottersen, O. P. (1997) Specialized membrane domains for water transport in glial cells: high-resolution immunogold cytochemistry of aquaporin-4 in rat brain. *J Neurosci,* 17**,** 171-80.

Nielsen, S., Smith, B. L., Christensen, E. I. and Agre, P. (1993) Distribution of the aquaporin CHIP in secretory and resorptive epithelia and capillary endothelia. *Proc Natl Acad Sci U S A,* 90**,** 7275-9.

Nobelprize.org. (2011a). The Nobel Prize in Chemistry 1958 [Online]. *Available: http://nobelprize.org/nobel_prizes/chemistry/laureates/1958/ [Accessed 17 Feb 2011].*

Nobelprize.org. (2011b). The Nobel Prize in Chemistry 1962 [Online]. *Available: http://nobelprize.org/nobel_prizes/chemistry/laureates/1962/ [Accessed 17 Feb 2011].*

Nothaft, H. and Szymanski, C. M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol,* 8**,** 765-78.

Ogata, K., Nishikawa, H. and Ohsugi, M. (1969) A Yeast Capable of Utilizing Methanol. *Agric biol chem,* Vol.33 pp.1519-1520.

Ohashi, Y., Tsuzaka, K., Takeuchi, T., Sasaki, Y. and Tsubota, K. (2008) Altered distribution of aquaporin 5 and its C-terminal binding protein in the lacrimal glands of a mouse model for Sjogren's syndrome. *Current eye research,* 33**,** 621-9.

Ohta, E., Itoh, T., Nemoto, T., Kumagai, J., Ko, S. B., Ishibashi, K., Ohno, M., Uchida, K., Ohta, A., Sohara, E., Uchida, S., Sasaki, S. and Rai, T. (2009) Pancreas-specific aquaporin 12 null mice showed increased susceptibility to caerulein-induced acute pancreatitis. *Am J Physiol Cell Physiol,* 297**,** C1368-78.

Oshio, K., Watanabe, H., Song, Y., Verkman, A. S. and Manley, G. T. (2005) Reduced cerebrospinal fluid production and intracranial pressure in mice lacking choroid plexus water channel Aquaporin-1. *FASEB J,* 19**,** 76-8.

Paganelli, C. V. and Solomon, A. K. (1957) The rate of exchange of tritiated water across the human red cell membrane. *J Gen Physiol,* 41**,** 259-77.

Pao, G. M., Wu, L. F., Johnson, K. D., Hofte, H., Chrispeels, M. J., Sweet, G., Sandal, N. N. and Saier, M. H., Jr. (1991) Evolution of the MIP family of integral membrane transport proteins. *Mol Microbiol,* 5**,** 33-7.

Perez-Martin, J., Cases, I. and de Lorenzo, V. (1997) Design of a solubilization pathway for recombinant polypeptides in vivo through processing of a bi-protein with a viral protease. *Protein Eng,* 10**,** 725-30.

Petrovskaya, L. E., Shulga, A. A., Bocharova, O. V., Ermolyuk, Y. S., Kryukova, E. A., Chupin, V. V., Blommers, M. J., Arseniev, A. S. and Kirpichnikov, M. P. (2010) Expression of G-protein coupled receptors in Escherichia coli for structural studies. *Biochemistry (Mosc),* 75**,** 881-91.

Phaff, H. J. (1956) A proposal for amendment of the diagnosis of the genus Pichia hansen. *Antonie Van Leeuwenhoek,* 22**,** 113-6.

Phaff, H. J. and Knapp, E. P. (1956) The taxonomy of yeasts found in exudates of certain trees and other natural breeding sites of some species of Drosophila. *Antonie Van Leeuwenhoek,* 22**,** 117-30.

Pletcher, C. H., Resnick, R. M., Wei, G. J., Bloomfield, V. A. and Nelsestuen, G. L. (1980) Deglycosylated prothrombin fragment 1. Calcium binding, phospholipid interaction, and self-assocation. *J Biol Chem,* 255**,** 7433-8.

Preston, G. M., Carroll, T. P., Guggino, W. B. and Agre, P. (1992) Appearance of water channels in Xenopus oocytes expressing red cell CHIP28 protein. *Science,* 256**,** 385-7.

Preston, G. M., Jung, J. S., Guggino, W. B. and Agre, P. (1993) The mercury-sensitive residue at cysteine 189 in the CHIP28 water channel. *J Biol Chem,* 268**,** 17-20.

Puigbo, P., Bravo, I. G. and Garcia-Vallve, S. (2008) CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct,* 3**,** 38.

Rai, T., Sasaki, S. and Uchida, S. (2006) Polarized trafficking of the aquaporin-3 water channel is mediated by an NH2-terminal sorting signal. *American journal of physiology. Cell physiology,* 290**,** C298-304.

Raina, S., Preston, G. M., Guggino, W. B. and Agre, P. (1995) Molecular cloning and characterization of an aquaporin cDNA from salivary, lacrimal, and respiratory tissues. *J Biol Chem,* 270**,** 1908-12.

Rice, P. (2011). EMBOSS:isochore [Online]. *Available: http://emboss.bioinformatics.nl/cgi-bin/emboss/isochore [Accessed 28 Mar 2011]*.

Richie, T. L. and Saul, A. (2002) Progress and challenges for malaria vaccines. *Nature,* 415**,** 694-701.

Romanos, M. A., Scorer, C. A. and Clare, J. J. (1992) Foreign gene expression in yeast: a review. *Yeast,* 8**,** 423-88.

Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R. and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science,* 307**,** 1317-21.

Sabolic, I., Katsura, T., Verbavatz, J. M. and Brown, D. (1995) The AQP2 water channel: effect of vasopressin treatment, microtubule disruption, and distribution in neonatal rats. *J Membr Biol,* 143**,** 165-75.

Sadlish, H., Pitonzo, D., Johnson, A. E. and Skach, W. R. (2005) Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. *Nat Struct Mol Biol,* 12**,** 870-8.

Saito, K., Kageyama, Y., Okada, Y., Kawakami, S., Kihara, K., Ishibashi, K. and Sasaki, S. (2004) Localization of aquaporin-7 in human testis and ejaculated sperm: possible involvement in maintenance of sperm quality. *J Urol,* 172**,** 2073-6.

Scheiffele, P., Peranen, J. and Simons, K. (1995) N-glycans as apical sorting signals in epithelial cells. *Nature,* 378**,** 96-8.

Scorer, C. A., Clare, J. J., McCombie, W. R., Romanos, M. A. and Sreekrishna, K. (1994) Rapid selection using G418 of high copy number transformants of Pichia pastoris for high-level foreign gene expression. *Biotechnology (N Y),* 12**,** 181-4.

Sharp, P. M. and Li, W. H. (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res,* 15**,** 1281-95.

Shental-Bechor, D. and Levy, Y. (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 8256-61.

Shi, L. B., Skach, W. R., Ma, T. and Verkman, A. S. (1995) Distinct biogenesis mechanisms for the water channels MIWC and CHIP28 at the endoplasmic reticulum. *Biochemistry,* 34**,** 8250-6.

Simons, K. and Ikonen, E. (1997) Functional rafts in cell membranes. *Nature,* 387**,** 569-72.

Sinclair, G. and Choy, F. Y. (2002) Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, Pichia pastoris. *Protein Expr Purif,* 26**,** 96-105.

Skach, W. R., Shi, L. B., Calayag, M. C., Frigeri, A., Lingappa, V. R. and Verkman, A. S. (1994) Biogenesis and transmembrane topology of the CHIP28 water channel at the endoplasmic reticulum. *J Cell Biol,* 125**,** 803-15.

Skowronski, M. T., Lebeck, J., Rojek, A., Praetorius, J., Fuchtbauer, E. M., Frokiaer, J. and Nielsen, S. (2007) AQP7 is localized in capillaries of adipose tissue, cardiac and striated muscle: implications in glycerol metabolism. *Am J Physiol Renal Physiol,* 292**,** F956-65.

Smith, B. L. and Agre, P. (1991) Erythrocyte Mr 28,000 transmembrane protein exists as a multisubunit oligomer similar to channel proteins. *J Biol Chem,* 266**,** 6407-15.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. and Hay, S. I. (2005) The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature,* 434**,** 214-7.

Song, Y., Sonawane, N. and Verkman, A. S. (2002) Localization of aquaporin-5 in sweat glands and functional analysis using knockout mice. *J Physiol,* 541**,** 561-8.

Song, Y. and Verkman, A. S. (2001) Aquaporin-5 dependent fluid secretion in airway submucosal glands. *J Biol Chem,* 276**,** 41288-92.

Sorensen, H. P. (2010) Towards universal systems for recombinant gene expression. *Microb Cell Fact,* 9**,** 27.

Sougrat, R., Morand, M., Gondran, C., Barre, P., Gobin, R., Bonte, F., Dumas, M. and Verbavatz, J. M. (2002) Functional expression of AQP3 in human skin epidermis and reconstructed epidermis. *J Invest Dermatol,* 118**,** 678-85.

Structural Genomics Consortium. (2009). SGC Target Progress [Online]. *Available: http://www.thesgc.org/structures/target_progress.php [Accessed 18 Feb 2011]*.

Sui, H., Han, B. G., Lee, J. K., Walian, P. and Jap, B. K. (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature,* 414**,** 872-8.

Suzuki-Toyota, F., Ishibashi, K. and Yuasa, S. (1999) Immunohistochemical localization of a water channel, aquaporin 7 (AQP7), in the rat testis. *Cell and tissue research,* 295**,** 279-85.

Szymanski, C. M. and Wren, B. W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol,* 3**,** 225-37.

Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. and Guerry, P. (1999) Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol,* 32**,** 1022-30.

Tate, C. G. (2001) Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett,* 504**,** 94-8.

Tate, C. G., Haase, J., Baker, C., Boorsma, M., Magnani, F., Vallis, Y. and Williams, D. C. (2003) Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. *Biochim Biophys Acta,* 1610**,** 141-53.

Thiagarajah, J. R. and Verkman, A. S. (2002) Aquaporin deletion in mice reduces corneal water permeability and delays restoration of transparency after swelling. *J Biol Chem,* 277**,** 19139-44.

Tornroth-Horsefield, S., Hedfalk, K., Fischer, G., Lindkvist-Petersson, K. and Neutze, R. (2010) Structural insights into eukaryotic aquaporin regulation. *FEBS Lett,* 584**,** 2580-8.

Tornroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R. and Kjellbom, P. (2006) Structural mechanism of plant aquaporin gating. *Nature,* 439**,** 688-94.

Tournaire-Roux, C., Sutka, M., Javot, H., Gout, E., Gerbeau, P., Luu, D. T., Bligny, R. and Maurel, C. (2003) Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. *Nature,* 425**,** 393-7.

Trimble, R. B., Atkinson, P. H., Tschopp, J. F., Townsend, R. R. and Maley, F. (1991) Structure of oligosaccharides on Saccharomyces SUC2 invertase secreted by the methylotrophic yeast Pichia pastoris. *The Journal of biological chemistry,* 266**,** 22807-17.

Tsubota, K., Hirai, S., King, L. S., Agre, P. and Ishida, N. (2001) Defective cellular trafficking of lacrimal gland aquaporin-5 in Sjogren's syndrome. *Lancet,* 357**,** 688-9.

Tsukaguchi, H., Shayakul, C., Berger, U. V., Mackenzie, B., Devidas, S., Guggino, W. B., van Hoek, A. N. and Hediger, M. A. (1998) Molecular characterization of a broad selectivity neutral solute channel. *J Biol Chem,* 273**,** 24737-43.

Waetzig, G. H., Chalaris, A., Rosenstiel, P., Suthaus, J., Holland, C., Karl, N., Valles Uriarte, L., Till, A., Scheller, J., Grotzinger, J., Schreiber, S., Rose-John, S. and Seegert, D. (2010) N-linked glycosylation is essential for the stability but not the signaling function of the interleukin-6 signal transducer glycoprotein 130. *The Journal of biological chemistry,* 285**,** 1781-9.

Walz, T., Fujiyoshi, Y. and Engel, A. (2009) The AQP structure and functional implications. *Handb Exp Pharmacol***,** 31-56.

van Balkom, B. W., Savelkoul, P. J., Markovich, D., Hofman, E., Nielsen, S., van der Sluijs, P. and Deen, P. M. (2002) The role of putative phosphorylation sites in the targeting and shuttling of the aquaporin-2 water channel. *J Biol Chem,* 277**,** 41473-9.

Wang, A., Clapper, J., Guderian, J. A., Foy, T. M., Fanger, G. R., Retter, M. W. and Skeiky, Y. A. (2003) A novel method for increasing the expression level of recombinant proteins. *Protein Expr Purif,* 30**,** 124-33.

Veenhuis, M., Van Dijken, J. P. and Harder, W. (1983) The significance of peroxisomes in the metabolism of one-carbon compounds in yeasts. *Adv Microb Physiol,* 24**,** 1-82.

Weerapana, E. and Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology,* 16**,** 91R-101R.

Wegner, E. H. (1983) Biochemical conversions by yeast fermentation at high cell densities. *United States patent application 06/534848*

Wegner, G. H. (1990) Emerging applications of the methylotrophic yeasts. *FEMS Microbiol Rev,* 7**,** 279-83.

Wellner, R. B., Hong, S., Cotrim, A. P., Swaim, W. D. and Baum, B. J. (2005) Modifying the NH2 and COOH termini of aquaporin-5: effects on localization in polarized epithelial cells. *Tissue Eng,* 11**,** 1449-58.

West, C. M. (1986) Current ideas on the significance of protein glycosylation. *Molecular and cellular biochemistry,* 72**,** 3-20.

White, S. (2011). Database of Membrane Proteins of Known Structure [Online]. *Available: http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html [Accessed 8 Apr 2011].*

Voet, D. and Voet, J. G. (2004) Biochemistry*,* New York, *J. Wiley & Sons.*

von Heijne, G. (2007) The membrane protein universe: what's out there and why bother? *J Intern Med,* 261**,** 543-57.

Woo, J., Chae, Y. K., Jang, S. J., Kim, M. S., Baek, J. H., Park, J. C., Trink, B., Ratovitski, E., Lee, T., Park, B., Park, M., Kang, J. H., Soria, J. C., Lee, J., Califano, J., Sidransky, D. and Moon, C. (2008) Membrane trafficking of AQP5 and cAMP dependent phosphorylation in bronchial epithelium. *Biochem Biophys Res Commun,* 366**,** 321-7.

Wright, F. (1990) The 'effective number of codons' used in a gene. *Gene,* 87**,** 23-9.

Wu, G., Culley, D. E. and Zhang, W. (2005) Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. *Microbiology,* 151**,** 2175-87.

Xia, X. (2007) The +4G site in Kozak consensus is not related to the efficiency of translation initiation. *PLoS ONE,* 2**,** e188.

Yakata, K., Hiroaki, Y., Ishibashi, K., Sohara, E., Sasaki, S., Mitsuoka, K. and Fujiyoshi, Y. (2007) Aquaporin-11 containing a divergent NPA motif has normal water channel activity. *Biochim Biophys Acta,* 1768**,** 688-93.

Yang, B., Gillespie, A., Carlson, E. J., Epstein, C. J. and Verkman, A. S. (2001) Neonatal mortality in an aquaporin-2 knock-in mouse model of recessive nephrogenic diabetes insipidus. *J Biol Chem,* 276**,** 2775-9.

Yang, B., Ma, T. and Verkman, A. S. (1995) cDNA cloning, gene organization, and chromosomal localization of a human mercurial insensitive water channel. Evidence for distinct transcriptional units. *J Biol Chem,* 270**,** 22907-13.

Yang, F., Kawedia, J. D. and Menon, A. G. (2003) Cyclic AMP regulates aquaporin 5 expression at both transcriptional and post-transcriptional levels through a protein kinase A pathway. *J Biol Chem,* 278**,** 32173-80.

Yasui, M., Hazama, A., Kwon, T. H., Nielsen, S., Guggino, W. B. and Agre, P. (1999a) Rapid gating and anion permeability of an intracellular aquaporin. *Nature,* 402**,** 184-7.

Yasui, M., Kwon, T. H., Knepper, M. A., Nielsen, S. and Agre, P. (1999b) Aquaporin-6: An intracellular vesicle water channel protein in renal epithelia. *Proc Natl Acad Sci U S A,* 96**,** 5808-13.

Zelenina, M., Bondar, A. A., Zelenin, S. and Aperia, A. (2003) Nickel and extracellular acidification inhibit the water permeability of human aquaporin-3 in lung epithelial cells. *J Biol Chem,* 278**,** 30037-43.

Zeuthen, T. and Klaerke, D. A. (1999) Transport of water and glycerol in aquaporin 3 is gated by H(+). *J Biol Chem,* 274**,** 21631-6.

Zhou, F., Xu, W., Hong, M., Pan, Z., Sinko, P. J., Ma, J. and You, G. (2005) The role of N-linked glycosylation in protein folding, membrane targeting, and substrate binding of human organic anion transporter hOAT4. *Molecular pharmacology,* 67**,** 868-76.