

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

HETEROLOGOUS EXPRESSION OF HUMAN MEMBRANE PROTEIN DRUG TARGETS AND THE X-RAY CRYSTALLOGRAPHIC DETERMINATION OF THE HUMAN AQUAPORIN 4 STRUCTURE

Permalink

<https://escholarship.org/uc/item/5fz218d9>

Author

Ho, Joseph

Publication Date

2009

Peer reviewed|Thesis/dissertation

**Heterologous Expression of Human Membrane Protein Drug Targets and the X-ray
Crystallographic Determination of the Human Aquaporin 4 Structure**

by

Joseph Daniel Ho

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

copyright © 2009

by

Joseph Daniel Ho

Dedicated to my family and

the brothers and sisters at 1st Home of Christ

for their encouragement, prayer and love

Acknowledgements

“I awoke this morning with devout thanksgiving for my friends, the old and the new.”

- Ralph Waldo Emerson

I would like to begin by acknowledging my thesis advisor, Professor Robert Stroud, for inspiring me to always seek out the most interesting problems in biology. I came into the lab with some work experience in the pharmaceutical industry. Desiring to work on membrane proteins, Bob encouraged me to study the G-protein coupled receptors, the hottest of all drug targets. This led to my pursuit to express these difficult human membrane proteins in *Pichia pastoris*. Throughout these 6 years, I thoroughly enjoyed Bob’s optimistic approach to science and his timely encouragement and guidance whenever I feel overwhelmed and dejected by negative results. Taking Bob’s crystallography course was also a delight. Bob, thank you for guiding me and supporting me these six years. Thank you for securing the funding needed for the lab so that I never had to worry about using the right reagent for the right experiment. Thank you for believing in me! You have worked to shape me into a great scientist and you deserve respect in all my current and future success. I also thank the rest of my thesis committee, Professor Robert Edwards and Professor David Julius. Thank you for always being available to critique my projects and helping me to think outside the box. You helped me to prioritize my projects so that I could finish my Ph.D. in a timely fashion.

I also thank all my fellow Stroudians (the Gang) for their encouragement and friendship.

Adrian K., David Savage, Zach Newby, Franz Gruswitz, Frank Hays, and Pascal Egea have

been my role models and have helped me to refine my experiments on a day-to-day basis.

Thanks to: Ian Harwood for organizing all the lab social activities, Tom Lee for helping me make pymol movies for my classes during the 1st year, Renee Robbins for being a wonderful bay mate, Ronald Yeh for his help with the GPCR and AQP4 projects, especially all the graveyard synchrotron trips, Rebecca Robbins & Larry Miercke for helping out in the TDA analyses on AQP4, Andrew Sandstrom for the assistance in purifying AQP4, Ilya Chorny for teaching me molecular dynamics simulations, James Holton, George Meigs, and Jane Tanamachi for making the ALS Beamline 8.3.1 such a wonderful place to collect data, Bill Harries for fixing just about anything that is broken in the lab as well as helping out with the growth and harvest of yeast cells and their membranes, Chris Waddling and Matt Harrington for being the best Macromolecular Structure Group support personnel. Thanks also to the rest of the “downstairs’ crew (Mimi Ho, Rachel Bond, Samantha Ngaw, and Zygy Roe-zurz, Mez, Linda Vuong, Melissa Del Rosario, and Min Li) for being wonderful and supportive friends. Thanks to Arceli Joves for making my yeast media and Rebecca Robbins for letting me share a wonderful office with her and helping me to be more organized.

I also have many good memories with my CCB classmates (Sarah, Abram, Janet, the Kristins, Jasmina, Jeremy, Dave). I treasured the time we hung out and studied together during our 1st year. Thanks to Janet and Abram for helping me through the statistical mechanics course. It was a hard class for me. Thank you for being so patient with my easy questions.

I appreciate Suzan Bethiel, Mom Keo, and Makiko Nakajima for their administrative assistance. They made it easy for Bob and me to just focus on doing great science. Thanks

also to Professor Charly Craik and the Chemistry and Chemical Biology graduate program. I believe my training in both chemistry and biology will be proven valuable in my future pursuits in the pharmaceutical industry. I give special honor to Chris Olson, the CCB mom. Thank you for adopting me and for always being there to listen to my stories inside and outside the lab. Thank you also for always pointing out my faults and deficiencies. I learned a lot from you.

During these 6 years, I was also able to engage in many meaningful extracurricular activities. I want to thank Richard Tjhen, Mark Borja and John Lee for the meaningful time of Bible study and for brotherly love, encouragement, and prayers. Many thanks also to the people at Living Stone Fellowship for praying for me and my Ph.D. projects. Special thanks to all the friends and family at Home of Christ for their love and support, to Pastor Joseph for Godly counsels, and to the praise band there—making music with you guys really helped me to decompress from the daily grind of life in the lab.

My family also contributed tremendously in shaping me as a scientist. Mom and Dad, thank you for your help throughout my life and my 21 years of education. I could not have walked this journey so successfully without your love and sacrifice. Richard, my brother, thank you for all your counsel and encouragement. My late grandmother, thank you for your prayers throughout my life and for believing in me that I would become a great scientist. Finally, I would like to express my deepest gratitude to my savior and Lord, Jesus. Thank you for giving me life and a passion for science. To You be the Glory!

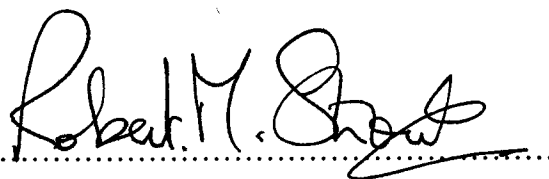
HETEROLOGOUS EXPRESSION OF HUMAN MEMBRANE PROTEIN DRUG TARGETS AND THE X-RAY CRYSTALLOGRAPHIC DETERMINATION OF THE HUMAN AQUAPORIN 4 STRUCTURE

Abstract

Membrane proteins comprise 60% of all known drug targets. They perform essential processes in the cell by functioning as receptors, transporters, or channels, controlling the flow of information and materials between the cell and its environment. In any given organism, about 1/3 of the genome encodes membrane proteins, but even given the clinical importance, the three-dimension structures of membrane proteins are scarce at best. As of June 2009, there are a total of ~58000 structures in the Protein Data Bank, but only 473 are from membrane proteins, and out of those, only 193 are unique structures, and out of those, only 6 structures are human integral membrane proteins. Factors that contributed to the paucity of membrane protein structures include difficulty in large quantity heterologous expression, purification and stabilization of the protein molecules in detergent, and forming crystals that diffract to high resolution relevant for biochemical studies and structure-based drug design.

We have tackled these problems on several fronts. First, the *Pichia pastoris* expression system was selected for heterologous expression of human membrane proteins. We have optimized the expression system for membrane protein expression, and from that system, we have obtained large quantity of human aquaporin 4 (hAQP4) as well as human GPCRs and acetylcholine receptors. In order to obtain the high resolution crystal structure of hAQP4, we

have characterized its protein-detergent complex using the Tetra Detector Array (Viscotek Corp.) and found ways to remove the flexible termini of the protein to improve diffraction resolution. We have also developed a way for academic laboratories to economically codon optimize and synthesize genes "in-house." Codon optimization is an important part of the expression optimization of eukaryotic membrane proteins due to the incompatibility of the codon usage in the gene of interest and the tRNA population of the expression host. By intelligently selecting the codons that are favored by the expression host, one can improve the expression significantly.

A handwritten signature in black ink, reading "Robert M. Stroud". The signature is written in a cursive style and is positioned above a horizontal dotted line.

Robert M. Stroud, Ph.D.

Advisor and Committee Chairman

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	2
CHAPTER 2 HETEROLOGOUS EXPRESSION OF HUMAN MEMBRANE PROTEIN DRUG TARGETS IN THE METHYLOTROPHIC YEAST, <i>PICHTIA PASTORIS</i>	
ABSTRACT	7
INTRODUCTION	8
RESULTS AND DISCUSSION	10
MATERIAL AND METHODS	19
REFERENCES	47
CHAPTER 3 CRYSTAL STRUCTURE OF HUMAN AQUAPORIN 4 AT 1.8Å AND ITS MECHANISM OF CONDUCTANCE	
ABSTRACT	51
INTRODUCTION	52
RESULTS AND DISCUSSION	55
MATERIAL AND METHODS	69
REFERENCES	101

**CHAPTER 4 | METHODS FOR IMPROVING X-RAY DIFFRACTION OF
MEMBRANE PROTEIN CRYSTALS: A CASE STUDY OF HUMAN**

AQUAPORIN 4

ABSTRACT	111
INTRODUCTION	112
RESULTS AND DISCUSSION	114
MATERIAL AND METHODS	118
REFERENCES	131

**CHAPTER 5 | AN “IN-HOUSE” APPROACH TO GENE OPTIMIZATION AND
SYNTHESIS**

ABSTRACT	134
INTRODUCTION	135
RESULTS AND DISCUSSION	137
MATERIAL AND METHODS	141
REFERENCES	154

LIST OF FIGURES

Figure 2-1	24
Construct designs for the protein expression in <i>Pichia pastoris</i>	
Figure 2-2	25
pPICZ and pPICZ α vectors	
Figure 2-3	26
<i>Pichia pastoris</i> expression studies of 12 eukaryotic membrane proteins from human and mouse	
Figure 2-4	28
Solubilization studies of the human AchR $\alpha 9$ from cells expressing the protein at the induction temperature of (a) 22°C and (b) 15°C	
Figure 2-5	30
Solubilization studies of the (a) human 5HT-1A and the (b) mouse δ opioid receptor	
Figure 2-6	32
Solubilization study of the human acetylcholine receptor subunit $\alpha 7$	

Figure 2-7	34
Size exclusion chromatography of the nickel affinity-purified 5HT-1A	
Figure 2-8	36
Size exclusion chromatography of the nickel affinity-purified 5HT-2A in FOS-CHOLINE [®] -14	
Figure 2-9	38
Size exclusion chromatography of the nickel affinity-purified mouse δ opioid receptor and the human acetylcholine receptor $\alpha 9$ in FOS- CHOLINE [®] -14	
Figure 2-10	40
Nickel affinity chromatography of human aquaporin 4 in n-decyl- β -D- maltopyranoside (DM)	
Figure 2-11	41
Size exclusion chromatography of the nickel affinity-purified human aquaporin 4 in n-decyl- β -D-maltopyranoside (DM)	
Figure 2-12	43
Size exclusion chromatography of the nickel affinity-purified human aquaporin 4 in n-octyl- β -D-glucoopyranoside (OG)	

Figure 3-1	77
General features of human aquaporin 4	
Figure 3-2	79
The conducting pore	
Figure 3-3	80
Electron density of the conducting pore	
Figure 3-4	82
Comparison of the hydrogen bond network of the selectivity filter arginine of hAQP4, bAQP1, and GlpF	
Figure 3-5	83
The NPA motifs	
Figure 3-6	85
Water and glycerol conduction assays and inhibitor assays	
Figure 3-7	87
Crystal lattice packing of (a) hAQP4 and (b) rAQP4	

Figure 3-8	88
Protein sequence alignment of all the aquaporin structures solved to date	
Figure 3-9	90
Comparison of the C loop between rat AQP4 and human AQP4	
Figure 3-10	91
Sphere representations of the van der Waals contact of the selectivity filter	
Figure 3-11	93
Comparison of the surfaces of the extracellular vestibule of human AQP4, PfAQP, <i>E. coli</i> GlpF, <i>E. coli</i> AqpZ , Archaeal AqpM, and bovine AQP0	
Figure 3-12	95
AQP4: post MD simulation from the third experiment	
Figure 4-1	119
Schematic representation of the protein-detergent interaction	
Figure 4-2	120

Tetra Detector Array Analysis (TDA) of the full-length AQP4

Figure 4-3 **121**

Crystals of the full-length AQP4

Figure 4-4 **122**

Crystals of full-length AQP4 after reducing the concentration of β -
octylglucoside in the crystallization sample

Figure 4-5 **123**

Initial diffraction patterns from the crystals in Figure 4

Figure 4-6 **124**

Schematic representation of the full-length human aquaporin 4
protein

Figure 4-7 **125**

Superdex 200 size exclusion chromatography traces of the full-
length and trypsinized AQP4

Figure 4-8 **126**

Matrix-assisted laser desorption/ionization mass spectrometry
(MALDI-MS) of the trypsinized AQP4

Figure 4-9	127
Tetra Detector Array Analysis (TDA) of the full-length AQP4	
Figure 4-10	128
Crystals of the trypsinized AQP4	
Figure 4-11	129
Crystals of the trypsinized AQP4 after using the Hampton Silver Bullet condition A1 as additive	
Figure 5-1	144
Gene design of (a) the 240 base-pair α -bungarotoxin and (b) the 717 base-pair acetylcholine binding protein (AchBP)	
Figure 5-2	145
Two methods for the PCR-based gene synthesis	
Figure 5-3	146
PCR products obtained from the “conventional” and the “inside-out” methods of PCR	
Figure 5-4	147

1% agarose DNA gel of the α -bungarotoxin used for the ligation reaction and the plasmid miniprep results

Figure 5-5 **148**

PCR products of the synthesis of the NcoI-PvuII segment and the PvuII-EcoRI segment of the 717 base-pair acetylcholine binding protein

Figure 5-6 **149**

1% agarose DNA gel of the AchBP fragments used for the ligation reaction and the plasmid miniprep results

Figure 5-7 **150**

Anti-flag Western result of the heterologous expression of human aquaporin 4 in *Pichia pastoris*

LIST OF TABLES

Table 2-1	45
A summary of the heterologous expression of human and mouse membrane protein in <i>Pichia pastoris</i>	
Table 2-2	46
A summary of constructs tested in <i>Pichia pastoris</i>	
Table 3-1	97
Data collection and refinement statistics	
Table 3-2	98
Summary of TLS refinement	
Table 4-1	130
A Summary of the synchrotron trips	
Table 5-1	151
Codon Adaptation Index of α -bungarotoxin and acetylcholine binding protein	
Table 5-2	152

CAI comparison for human AQP4

Table 5-3

153

A summary of the financial cost of the “in-house” approach to codon optimization and gene synthesis

“We ourselves feel that what we are doing is just a drop in the ocean. But the ocean would be less because of that missing drop.”

— Mother Teresa of Calcutta

Chapter 1

Introduction

When I was an undergraduate at University of California, Davis, I majored in Microbiology and was bound for medical school, eager to become a health care professional who can directly impact human health. It was not until the beginning of my 3rd year when I took a graduate course in microbial pathogenesis that I became interested in the molecular mechanism of how diseases work. Fueled by my interest in understanding how enzymes play a role in diseases, I asked to join the laboratory of Professor Irwin Segel in the Department of Biochemistry and Molecular Biology as an undergraduate research helper. I worked with graduate student Ian MacRae (who is now an assistant professor at The Scripps Research Institute) on the enzyme kinetics and the X-ray structure of ATP sulfurylase from *Penicillium chrysogenum*. It was there that I began to realize how X-ray crystallography can be used to probe the fine mechanistic details of enzymatic reactions.

Right after college, I started my first job with Professor Chris Garcia at Stanford University. I was trained in protein refolding, expression and the purification of a host of human proteins in the immune system. I also learned to relate the structural biology to the pathogenesis seen in auto-immune diseases. In the following 3 years, I joined the Structure Chemistry group at Celera Genomics where I was successful in the expression, purification and crystallization of human cathepsin 4 and human hepsin. After 3 years in industry, my background in molecular cloning, protein biochemistry, and the crystallization of several human proteins sparked my interest in pursuing a Ph.D. degree in Chemistry and Chemical Biology at UCSF to further my understanding in structural biology and structure-based drug design.

During the 1st year in graduate school, even though I knew I wanted to do structural biology, I did not have in mind to work on a specific type of protein or get involved in the process of a specific disease. It was not until I took Professor Charly Craik's Chemical Biology class that I realized the tremendous prospect of using membrane proteins as drug targets and the lack of membrane protein structures that are available for structure-based drug design. It was then I knew I wanted to study under Professor Robert Stroud. I wished to discover ways to more efficiently express eukaryotic membrane proteins at the levels that suit protein X-ray crystallography and solve a membrane protein structure of my own.

When I started in the lab, Bob said to me, "pick your own poisons, Joe." Indeed, fueled by a side interest in neuroscience, I immediately set out to select membrane proteins from the central nervous system that affect human health. In my six years in the Stroud Lab, I have drunk much poison by wrestling with these transmembrane macromolecules, and what didn't kill me indeed made me stronger. The highlight of this dissertation is the crystal structure of human aquaporin 4 at 1.8 Å.

In chapter 2, I described the pilot expression studies of several human membrane proteins in *Pichia pastoris*. In general, the heterologous expressions of these proteins in *Pichia* have not been difficult, but purifying them in their monodispersed and stable form has been the biggest challenge. Chapter 3 is the paper published in the *Proceedings of the National Academy of Science* that describes the crystal structure of human aquaporin 4 and its mechanism of conductance. The structure was solved to 1.8 Å and is currently the highest resolution human membrane protein structure. The process by which this high resolution

structure was obtained was, however, not straightforward. Therefore, I dedicated chapter 4 to describe how the diffraction resolution of the crystal was improved from 8 Å to 1.8 Å.

Finally, chapter 5 is a side project that I engaged in at the beginning of my stint in the Stroud Lab. Heterologous expression of eukaryotic proteins in *E. coli* bears an inherent difficulty—incompatibility of the codon usage of the gene of interest and the expression host. Most eukaryotic genes have synonymous codons for arginine, glycine, isoleucine, leucine, and proline that are rare in the tRNA pools of *E. coli*, and some eukaryotic genes may be too AT- or GC-rich so that the *E. coli* RNA polymerase cannot transcribe them efficiently. Therefore, a method to overcome this transcription/translation incompatibility is to reengineer the gene in such a way that it is friendly for *E. coli* expression. Chapter 5 describes an economic way by which gene synthesis can be accomplished “in-house”, without depending on a gene synthesis company. One could also use this method to synthesize genes that are *Pichia pastoris* friendly or *Spodoptera frugiperda* friendly, opening up endless possibilities of improving heterologous expression of human or eukaryotic membrane proteins in different hosts.

In hindsight, in the field of membrane protein expression, purification, and crystallization, this dissertation is only a drop in the ocean, but “ocean would be less without this drop!” By standing on the shoulders of giants, science will continue to ensure the success of its offspring.

Chapter 2

Heterologous Expression of Human Membrane Protein Drug Targets in the methylotrophic yeast, *Pichia pastoris*

Research completed by

Joseph D. Ho, Ronald Yeh and Robert M. Stroud

ABSTRACT

Pichia pastoris is a methylotrophic yeast that can utilize methanol as the sole carbon source for its growth. In this chapter, we examined the potential for using the *Pichia pastoris* expression system for heterologously expressing eukaryotic membrane proteins. Several G-protein coupled receptors, ion channels, and human aquaporin 4 were selected for pilot expression studies. Membrane solubilization and protein purification were carried through for some of these targets, and human aquaporin 4 is the only protein at this point that can be stabilized in detergent and purified into a pure, homogenous, and stable state.

INTRODUCTION

Since the report of the first membrane protein X-ray structure, the photosynthetic reaction center from *Rhodospseudomonas viridis*, in 1985 (1), membrane protein structure determination remains a challenging field, with the number of structures occupying less than 1% of all protein structures in the protein data bank (2, 3). During the 1990s, several eukaryotic membrane protein structures were determined, but the discovery remained limited to proteins purified from natural sources. It was not until 2005 that MacKinnon *et al.* reported the crystal structure of the *Rattus norvegicus* Kv1.2 potassium channel, the first eukaryotic membrane protein from heterologous expression (4). The protein was expressed in *Pichia pastoris* and since then, other successes of expressing eukaryotic membrane proteins in *Pichia pastoris* have been reported: the X-ray crystal structures of the spinach aquaporin (5), human aquaporin 4 and 5 (6, 7) and the *Mus musculus* P-Glycoprotein (8).

Pichia pastoris is a methylotrophic yeast that can metabolize methanol as the sole carbon source for growth. A homo-octomer enzyme, alcohol oxidase, converts methanol to formaldehyde utilizing molecular oxygen. Because this enzyme has poor affinity for O₂, the cell relies on a very strong promoter, the AOX1 promoter, to produce large amounts of the enzyme. The promoter is induced by methanol and is tightly regulated by methanol, so a gene of interest can be inserted after the AOX1 promoter, and its expression controlled by the presence of methanol (9, 10). Expression of eukaryotic protein in *Pichia pastoris* has at least two advantages: 1. *Pichia* can perform all the post-translational modifications of a eukaryotic protein to ensure its biological function whereas *E. coli* cannot, and 2. *Pichia* is more cost effective than baculovirus and mammalian cell expression systems.

Nevertheless, reports have also shown success with the *E. coli* and the baculovirus system. The expression of the malarial *Plasmodium falciparum* aquaporin in *E. coli* was reported and the X-ray structure was solved (11), but the gene needed to be codon optimized and synthesized to reduce the high AT content and to suit *E. coli* tRNA codon usage in order to obtain enough protein expression for crystallization trials. The X-ray structures of the three most recent G-protein coupled receptors have been solved from proteins expressed using the Baculovirus/insect cells: the turkey β_1 adrenergic receptor (12), the human β_2 adrenergic receptor (13, 14), and the human A_{2A} adenosine receptor (15). Indeed, most eukaryotic membrane proteins cannot be easily isolated from natural sources in sufficient quantity for X-ray structure determination; therefore, discovery of novel expression systems or improvement of existing expression systems will continue to be important to the field of membrane protein crystallography.

In this chapter, we present the prospect of using the *Pichia pastoris* expression system for the heterologous expression of eukaryotic membrane proteins. We selected 8 G-protein coupled receptors, 3 ion-channels, and 1 aquaporin from human and mouse. Pilot expression result, membrane solubilization studies, and initial purification results are presented.

RESULTS AND DISCUSSION

Heterologous Expression in *Pichia pastoris*

The X-33, SMD1168, and SMD1163 strains were evaluated for expression. SMD1168 and SMD1163 were the protease-deficient strains, but the X-33 strain was chosen for the expression experiments because the X-33 strain grew a lot faster than the other two, and there was no sign of proteolysis observed initially.

Both the flag tag (DYKDDDDK) and the 8xHis tag (HHHHHHHH) were included in the constructs (**Figure 2-1**). For the purpose of Western detection of protein expression in membrane fraction, the flag M2 antibody (Sigma) gave a cleaner Western than the His antibody (Santa Cruz Biotechnology). For the purpose of initial purification, Ni-NTA (Qiagen) was more economical than the M2 anti-flag antibody column (Sigma).

The human rhinovirus 3C protease was chosen as the protease for tag removal because of its high specificity, recognizing only an 8-amino acid sequence (LEVLFQ↓GP). Another protease of choice is the Tobacco Etch Virus (TEV) protease, a 7-amino acid recognition protease (ENLYFQ↓G), but inhibition by detergents such as glucopyranosides and fos-cholines make the TEV protease not ideal for tag removal of membrane proteins (16, 17). It was found later that the 3C protease is partially inhibited by the fos-cholines, and this can be overcome by using more protease. However, thrombin is not inhibited by any glucosides, maltosides, and fos-cholines (personal communications with Dr. Senyon Choe's lab at the Salk Institute and Dr. Michael Wiener at the University of Virginia). It is also known that since thrombin only

recognizes a 5-amino acid sequence (LVPR↓GS), there is a greater possibility of internal cleavage of the target protein.

The pPICZ and pPICZ α vectors were selected (**Figure 2-2a and 2-2b**). The pPICZ α vector has the α -Factor secretion signal upstream of the multiple cloning sites (**Figure 2-2b**). The function of the α -Factor secretion signal is to facilitate the extracellular secretion of expressed protein into the media. For membrane proteins that have a signal peptide or have an extracellular N-terminus, the α -Factor secretion signal can substitute for the native signal peptide. 5HT-3A and the nicotinic acetylcholine receptor $\alpha 7$ and $\alpha 9$ all have an extracellular N-terminus with a signal peptide. Experiments were performed to compare expression of these proteins using the native signal peptide versus substituting the native signal peptide with the α -Factor secretion signal using the pPICZ α vector. We were not able to get any expression with the native signal peptide, but expression was observed when the native signal peptide was swapped with the α -Factor secretion signal (**Figure 2-3, lanes 10-12**). The G-protein coupled receptors (GPCRs) also have extracellular N-termini, but they lack any signal peptides. GPCRs can be expressed in *Pichia* with and without the α -Factor secretion signal, but when expressing the protein with the α -Factor secretion signal, there was sometimes a portion of the expressed material still containing the α -Factor secretion signal. The α -Factor secretion signal should normally be removed by *Pichia*'s native proteases: KEX2 and STE13 (Invitrogen). This has been observed in literature (18) as well as in our initial expression trials. Therefore, all GPCRs were cloned into the pPICZ and not the pPICZ α vector.

Factors such as temperature, the usage of ligands, histidine, and DMSO have also been discussed in literature (19), and we found that by lowering the induction temperature to 15°C, the protein expressed could be solubilized by non-ionic detergents at greater efficiency. Human nicotinic acetylcholine receptor (AChR) $\alpha 9$ was initially expressed at 30°C, and the material could not be solubilized by any detergent except SDS, a harsh denaturing anionic detergent (data not shown). When expressed at 22°C, some solubilization was observed by FOS-CHOLINE[®] 14, a zwitterionic detergent (**Figure 2-4a**), but when the induction temperature was lowered further to 15°C, the expressed material could be solubilized by non-ionic detergents, such as the glucosides, maltosides, and the polyoxyethelenes (**Figure 2-4b**). This demonstrates that the materials expressed at 15°C is different from the materials expressed at 22°C, and lowering the temperature to 15°C most likely favors the production of the protein in the functional form. Hence, with the exception of hAQP4, all experiments were performed with the induction temperature of 15°C.

Out of the proteins tested, only the human dopamine D1 and D2 receptors did not express. Human aquaporin 4 (hAQP4) was purified to homogeneity (**Figure 2-11**) and was quantified to be about 5 mg of purified protein per liter of cell. By comparing the Western intensity of the band for hAQP4 with the other proteins, we can see that the rest all expressed at the level of at least 1 mg per liter (**Figure 2-3**). Some of the constructs ran as two bands on SDS-PAGE/Western. Membrane proteins tend to run a little bit smaller on SDS-PAGE because they are often not unfolded by SDS completely due to their highly hydrophobic nature. Based on the theoretical molecular weight of these constructs, the top band is most likely a dimer and the bottom band a monomer (**Figure 2-3**). Aquaporins can run as tetramers, dimers, and

monomers on SDS-PAGE (personal communication in the Stroud Lab), and GPCRs can form biological dimers, so it is likely that these higher order oligomers are biological species in the membrane that are SDS-resistant.

The 5HT-3A, AchR $\alpha 7$ and $\alpha 9$ are homopentameric ligand-gated ion channels (LGICs).

Besides the monomeric form, there was some SDS-resistant pentameric and dimeric species on the gel as well (**Figure 2-3, lanes 10-12**). However, the monomeric species ran a little larger than its theoretical molecular weight. There are two explanations. First, the larger molecular weight was attributed by the presence of glycosylations (5HT-3A has 3 sites; AchR $\alpha 7$ has 3 sites; AchR $\alpha 9$ has 2 sites). Second, the larger molecular weight was attributed by the presence of the α -Factor secretion signal (~9,300 Daltons), which was not removed by the yeast proteases. We have not been able to obtain enough pure protein of the LGICs to answer this question.

Solubilization of the human 5HT-1A receptor and the mouse δ opioid receptor

Both 5HT-1A and the δ opioid receptor could be solubilized by non-ionic detergents, but Fos-choline 14 still solubilized better than non-ionic detergents (**Figure 2-5**). This is attributed to the zwitterionic nature of the Fos-choline head group. For Sodium dodecyl sulfate (SDS), its anionic sulfate head group makes it the harshest detergent. It solubilized all the membrane proteins that we have tested, including those that could not be solubilized by Fos-choline 14. The solubilization result of 5HT-1A and the δ opioid receptor is representative of all the GPCRs that we have tested. β -octyl-glucopyranoside (OG) did not

solubilize any of the GPCRs, but β -dodecyl-maltopyranoside (DDM) and Fos-choline 14 solubilized all the GPCRs we have tested.

FOS-CHOLINE[®] 14 (FC14) solubilization is pH dependent

FC14 has a zwitterionic head group. By adjusting the pH, the zwitterionic head group can become anionic or cationic, potentially changing the property of the detergent. We tested the solubilization of AchR $\alpha 7$ from pH 4 to pH 9 and found that lowering the pH to 5 or 6 improved solubilization to nearly 100%, and raising the pH to 9 also slightly improved solubilization (**Figure 2-6a**). A similar experiment was conducted for β -dodecyl-maltopyranoside (DDM) and changes in pH did not have any effect on solubilization (**Figure 2-6b**). This showed that the improvement of solubilization in FC14 at higher and lower pHs was not attributed directly by the change in the protein conformation but by the change of the property of the FC14 zwitterionic head group. At lower pH, the FC14 head group becomes more cationic, and at higher pH, the FC14 head group becomes more anionic. It is likely that FC14 becomes a stronger detergent at pH 5 and pH 9. For both FC14 and DDM, solubilizing at pH 4 resulted in no solubilization, probably because the protein was denatured at that pH. It is not known whether the extra population of protein molecules that was solubilized by FC14 at pH 5 and pH 9 was functional as compared to the population that was solubilized by the non-ionic detergent DDM.

Size exclusion chromatography (SEC) of 5HT-1A in n-dodecyl- β -D-maltopyranoside (DDM) and n-octyl- β -D-maltopyranoside (OM)

In general, All the GPCRs that we have tested could not be solubilized by octyl- or nonyl-glucopyranosides, but they could be solubilized by maltopyranosides ranging from eight (octyl-) to twelve (dodecyl-) carbons long. However, the solubilized and nickel-affinity-purified material showed up as large aggregates on SEC. **Figure 2-7a** and **2-7b** are the SEC profiles of 5HT-1A in DDM and OM. These two profiles are representative of all the GPCRs we have tested in DDM, DM, and OM. Basically, the protein was mostly aggregated, showing up in the void volume of the column. There was always some protein in the included volume, but the peak was broad (**Figure 2-7**). We have conducted the same SEC runs in the presence of agonist, antagonist and inverse agonist, but they were similar to the traces in **Figure 2-7**. We have also tested using cholesterol and the brain lipid extract during solubilization, nickel purification, and the SEC runs, and we did not see a larger percentage of protein shifted to the included volume. It is likely that the maltopyranosides are not the right detergent for the GPCRs, or the GPCRs expressed in *Pichia* are somewhat misfolded or unstable.

Size exclusion chromatography (SEC) of 5HT-2A in FOS-CHOLINE[®]-14 (FC14)

Because of its zwitterionic property, FC14 is a stronger solubilizer than the non-ionic glucopyranosides and maltopyranosides. We think FC14 is not as strong and denaturing as sodium dodecyl sulfate (SDS) because we have encountered membrane proteins that could not be solubilized by FC14 but could be solubilized by SDS. Nevertheless, doubts have been raised concerning whether FC14 could destabilize or partially unfold membrane protein. Mirzabekov *et al.* have demonstrated that a conformation-dependent antibody directed against the GPCR CCR5 cannot recognize the epitope when CCR5 is solubilized in FC14,

but can recognize the epitope when CCR5 is solubilized in non-ionic detergents including DDM, DM and cymal-5 (20). It is also interesting that out of the ~180 unique membrane protein structures in the Protein Data Bank, only one is solubilized, purified, and crystallized in FC14, and this structure is only 3.8 Å in resolution (21).

We have solubilized and nickel-purified our GPCRs in FC14 and analyzed the materials on SEC. What we have found consistently in FC14-solubilized material was less protein in the void volume and more protein in the included volume. The 5HT-2A and the δ opioid receptor are examples (**Figure 2-8** and **Figure 2-9a**). However, the peak in the included volume is still very broad, not suitable for obtaining the monodisperse species (**Figure 2-8** and **Figure 2-9a**). We have also tested the homopentameric ligand-gated ion channels (5HT3A and the AchRs) in FC14, and they showed up mostly in the void volume. The AchR $\alpha 9$ is illustrated here (**Figure 2-9b**). We do not think the FC14-solubilized homopentamers (~300,000) should exceed the void volume molecular weight of the Superdex 200 column (~650,000), so it is likely that the FC14-solubilized material may be aggregated.

Given these reports, we cannot be certain that the FC14-solubilized materials still maintain their biological conformation. It is also likely that FC14 can solubilize materials that are partially misfolded or aggregated, contributing to the broadness of the peak in the included volume.

Human aquaporin 4 (hAQP4) can be solubilized and purified in n-decyl- β -D-maltopyranoside (DM)

Solubilization and nickel affinity purification have been performed in n-octyl- β -D-glucopyranoside (OG), n-nonyl- β -D-glucopyranoside (NG), n-decyl- β -D-maltopyranoside (DM), and n-dodecyl- β -D-maltopyranoside (DDM). Results in all four detergents were similar. As demonstrated by using DM (**Figure 2-10**), the detergent extracts nearly 100% of hAQP4 in the membrane, and the N-terminal his-tagged hAQP4 adsorbed to the nickel resin very well, and upon elution, the protein is already >95% pure. Human AQP4 ran as two bands on SDS-PAGE. The lower band corresponded to the monomer, and the higher band corresponded to the biological tetramer. Membrane proteins tend to run smaller than their theoretical molecular weight on SDS-PAGE because their hydrophobic nature prevents SDS from fully denaturing and extending the protein molecules. Upon removal of the tag by 3C protease, both bands became smaller, and the absence of the flag Western signal showed the 3C protease digestion was complete. The 3C protease was tagged by an N-terminal 6xHis and the maltose-binding protein (MBP). The combined molecular weight of the 6xHis-MBP-3C protein was about ~67,000. In **Figure 2-10**, lane 6, the residual protein band running at place of the tetrameric hAQP4 (pointed by the arrow) should be the 6His-MBP-3C protein because if that band is the remaining uncut tetrameric hAQP4, it would have given off anti-flag Western signal.

The DM-solubilized and nickel-purified material was also homogenous, running at the biological tetrameric size on SEC (**Figure 2-11a, left**). Upon storing the peak fraction at 4°C for 6 days and re-injecting it on SEC, hAQP4 was also stable over at least a week (**Figure 2-11a, right**). The SEC-purified material corresponded to the theoretical molecular weight of

the 3C-cleaved monomer (34,852) as analyzed by the Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) (**Figure 2-11b**).

Human AQP4 is not stable in n-octyl- β -D-glucopyranoside (OG) at pH 8.0

Protein purification and crystallization of hAQP4 have been attempted in the glucopyranosides and the maltopyranosides as well as FC14, but at the end, crystals were only obtained in OG. Interestingly, hAQP4 was also least stable in OG. Human AQP4 in OG oligomerized to a molecular weight of >650,000 (void volume of the Superdex 200 column) over a period of 1 week (**Figure 2-12a**) and the oligomerization was accelerated during protein concentration (**Figure 2-12b**). The oligomerization is prevented if hAQP4 was brought to pH 6.0 before concentration of the 3C-cleaved material. The material at pH 6.0 was stable over time and yielded crystals that also grew exclusively at pH 6.0 (**Figure 2-12c**). Please see chapter 4 for details on the crystallization of hAQP4 in OG.

Summary

Table 2-1 shows the summary of the progress of the expression, purification, and crystallization of the proteins heterologously expressed in *Pichia pastoris*. The human dopamine D1 and D2 did not express in *Pichia*, and human AQP4 was the only protein that we have been able to isolate as pure, homogenous and stable molecules. Apart from human AQP4, the spinach AQP and human AQP5 can also be expressed in *Pichia* (5, 7). We believe the *Pichia pastoris* expression system is great for expressing eukaryotic aquaporins. GPCRs can be expressed at milligrams level in *Pichia*, but it remains to be seen whether the material expressed is functional or can be purified into homogenous and stable molecules.

MATERIALS AND METHODS

Gene subcloning and design

Human 5HT-1A (Gene Bank accession number AF_498978), 5HT-2A (AF_498982) receptors, dopamine D1 (AF_498961), D2 (NM_000795), D5 (AY_136750) receptors were purchased through the Missouri S&T cDNA Resource Center (www.cdna.org). Human aquaporin 4 (NM_001650) and human acetylcholine receptor (AChR) $\alpha 9$ (NM_017581) were purchased through Origene (Rockville, MD). The human acetylcholine receptor (AChR) $\alpha 7$ gene was originally redesigned and synthesized to suit *E. coli* expression and was later applied to *Pichia* expression. Mouse opioid receptors δ , μ , and κ were gifts from professor Mark Von Zastrow (UCSF).

Except human aquaporin 4, the expression constructs were design with an N-terminal flag tag (DYKDDDDK) with a human rhinovirus 3C protease cleavage site (LEVLFQ↓GP) and a C-terminal 6xHis tag with the same cleavage site (**Figure 2-1**). The δ , μ , and κ opioid receptors were cloned into the EcoRI and SacII sites of the pPICZ expression vector (Invitrogen) (**Figure 2-2a**). Both AChR $\alpha 7$ and $\alpha 9$ were cloned into the EcoRI and NotI sites of the pPICZ α expression vector (Invitrogen) (**Figure 2-2b**). The rest were cloned into the EcoRI and NotI sites of the pPICZ expression vector (Invitrogen). For human aquaporin 4, the expression construct was designed with a N-terminal 8xHis followed by a flag tag (DYKDDDDK) and a human rhinovirus 3C protease cleavage site (LEVLFQ↓GP) and cloned into the EcoRI and NotI sites of the pPICZ expression vector (Invitrogen). For a complete summary, see **Table 2-2**.

The expression vector was then electroporated into *Pichia pastoris* X-33 cells (Invitrogen) using the BioRad Gene Pulser Xcell System following a standard yeast electroporation protocol. Transformation was then selected on YPD plates with 50 µg/mL Zeocin™ (Invitrogen). Four colonies were restreaked and tested for expression. For production, the yeast was cultured in BMY media (Invitrogen) in Fernbach flasks at 30°C for 24 hours, then the temperature was lowered to 15°C, and methanol was added directly to the cultures to a final concentration of 2.5%. The cultures were grown for another 48 hours before harvest. Cultures were harvested by centrifugation at 4°C at 6000 x g for 10 minutes. Pellets were washed once with TBS buffer with 1 mM β-ME, and 1 mM PMSF and pelleted again. Cells were then resuspended with the same buffer and lysed by bead beating with glass beads. Broken and unlysed cells were removed by centrifugation at 4°C at 6000 x g for 10 minutes while the membrane fragments remained in the supernatant. The membrane fragments were then pelleted at 160,000 x g at 4°C for 1 hour. Pellets were resuspended in **MR Buffer** (25 mM Tris-HCl, pH 7.4 at room temperature, 250 mM NaCl, 10% glycerol, 1 mM β-ME) and stored at -80°C until further processing.

Sample preparation for SDS-PAGE and Western transfer: the pilot expression studies

10 µL of resuspended membrane were mixed with 90 µL of SDS sample buffer, then 10 µL of that was loaded on a SDS-PAGE gel with no boiling. Western transfer was performed on a BioRad trans-plot SD semi-dry transfer cell for 30 minutes. Then it was block with 3% milk for 20 minutes and probed with the anti-flag M2-HRP peroxidase conjugate antibody (Sigma)

at 1:5000 dilution in 3% BSA for 1 hour. The western was then washed in TBS and developed with the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

Sample preparation for SDS-PAGE and Western transfer: the solubilization studies

500 μ L of resuspended membrane were mixed with 500 μ L of detergent at solubilization concentration in a 2 mL microcentrifuge tube and stirred for 1 hour at 4°C. The solubilization concentration for each detergent tested is described in **Figure 2-4** and **Figure 2-5**. Then, a 10- μ L pre-spin sample was taken and the rest was pelleted at 160,000 x g at 4°C for 30 minutes and then a 10- μ L post-spin sample was taken. The 10- μ L sample was mixed with 90 μ L of SDS sample buffer and loaded onto a SDS-PAGE gel without boiling and transferred to a Western blot as described above.

Sample preparation for SDS-PAGE and Western transfer: the size exclusion chromatography experiments

10 μ L were taken from the selected fractions and mixed with 3 μ L of 5X SDS sample buffer. The sample was then loaded onto a SDS-PAGE gel without boiling and transferred to a Western blot as described above.

When there was not enough signal for Western detection, groups of fractions are selected and concentrated together in a 50,000 molecular weight cut-off Amicon spin concentrator (Millipore) for about 100 fold. 10 μ L of the retente were then assayed on SDS-PAGE and Western transfer.

Nickel Affinity chromatography general protocol

To begin purification, resuspended membrane was solubilized by adding detergent to a final concentration at the **solubilization concentration** and stirred at 4°C for 1 hour.

Unsolubilized material was pelleted at 160,000 x g at 4°C for 30 minutes. 5 M imidazole, pH 7.4, was added to the supernatant to a final concentration of 50 mM. The supernatant was then batch bound with Ni-NTA resins (Qiagen) for 2 hours, loaded onto a BioRad Econo Column and washed with **MR Buffer** (25 mM Tris-HCl, pH 7.4 at room temperature, 250 mM NaCl, 10% glycerol, 1 mM β -ME) with detergent at the **chromatography concentration** and 50 mM imidazole, and then eluted with 300 mM imidazole. Imidazole was removed using Econo-Pac DG10 desalting column (BioRad) equilibrated with MR Buffer with detergent at the **chromatography concentration**.

See below for the concentrations of solubilization and chromatography:

OG (n-octyl- β -D-glucopyranoside): solubilization at 200 mM; chromatography at 40 mM.

DM (n-decyl- β -D-maltopyranoside): solubilization at 20 mM; chromatography at 5 mM.

DDM (n-dodecyl- β -D-maltopyranoside): solubilization at 20 mM; chromatography at 0.5 mM.

FC14 (FOS-CHOLINE[®]-14): solubilization at 20 mM; chromatography at 0.5 mM.

Size exclusion chromatography

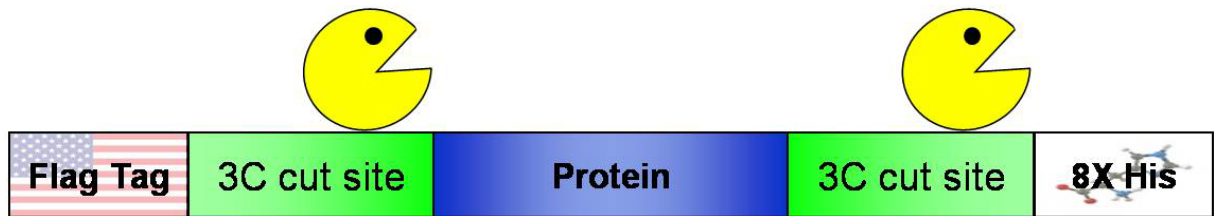
All size exclusion chromatography experiments were performed on a Superdex 200 10/300 GL column (GE Healthcare) in 25 mM Tris-HCl, pH 7.4 at room temperature, 250 mM NaCl, 10% glycerol, 2 mM DTT, and the detergent at the **chromatography concentration**. All

experiments were performed with the flow rate of 0.333 mL/min and a run length of either 50 or 60 minutes.

Purification of human aquaporin 4 in decyl-maltopyranoside (DM)

To begin purification, resuspended membrane was solubilized by adding 40 mM n-decyl- β -D-maltopyranoside (DM) (Anatrace) to a final concentration of 20 mM and stirred at 4°C for 1 hour. Unsolubilized material was pelleted at 160,000 x *g* at 4°C for 30 minutes. 5 M imidazole, pH 7.4, was added to the supernatant to a final concentration of 50 mM. The supernatant was then batch bound with Ni-NTA resins (Qiagen) for 2 hours, loaded onto a BioRad Econo Column and washed with MR Buffer with 5 mM DM and 50 mM imidazole, and then eluted with 300 mM imidazole. Imidazole was removed using Econo-Pac DG10 desalting column (BioRad) equilibrated with MR Buffer with 5 mM DM. The N-terminal tag was cleaved by His-tagged MBP fusion of human rhinovirus 3C protease (His6-MBP-3C) at 4°C overnight. Uncleaved hAQP4 (<5%) and 6xHis-MBP-3C were removed the next day with TALON® resin (Clontech). Then hAQP4 was concentrated in a 50,000 molecular weight cut-off Amicon spin concentrator (Millipore) and further purified by size exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in 25 mM Tris-HCl, pH 7.4 at room temperature, 250 mM NaCl, 10% glycerol, 5 mM DM, and 2 mM DTT.

(a)



(b)

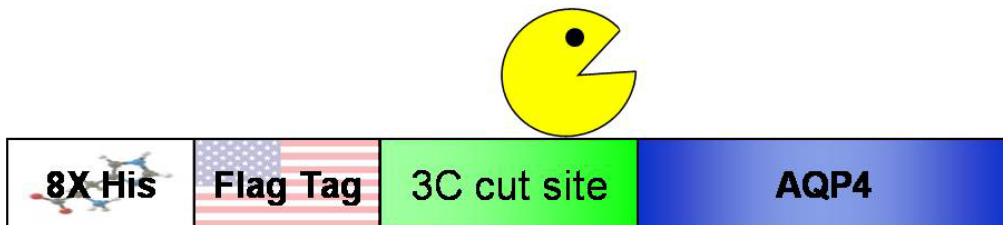
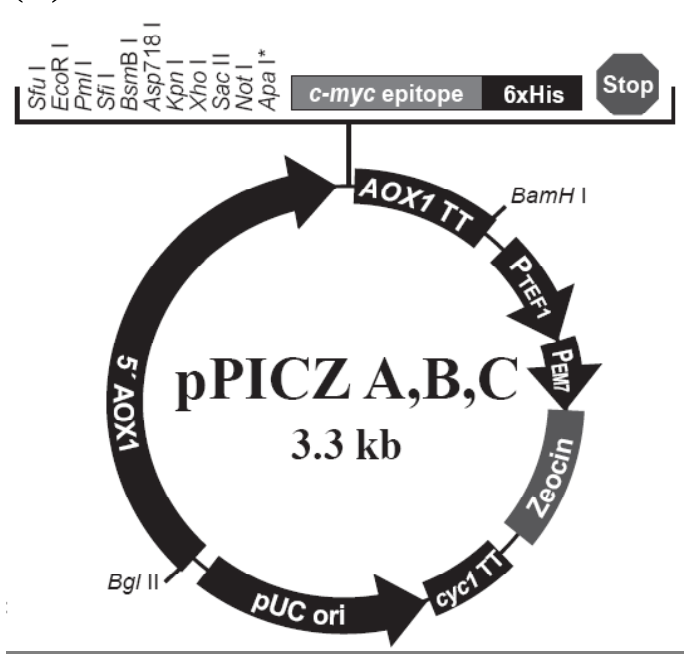


Figure 2-1

Construct designs for the protein expression in *Pichia pastoris*. Please see **Table 2-2** for more details.

- (a) Construct design for 5HT-1A, 5HT-2A, dopamine receptor D1, D2, D5, mouse δ , μ , and κ opioid receptors, and human acetylcholine receptor subunit $\alpha 7$ and $\alpha 9$.
- (b) Construct design for human aquaporin 4.

(a)



(b)

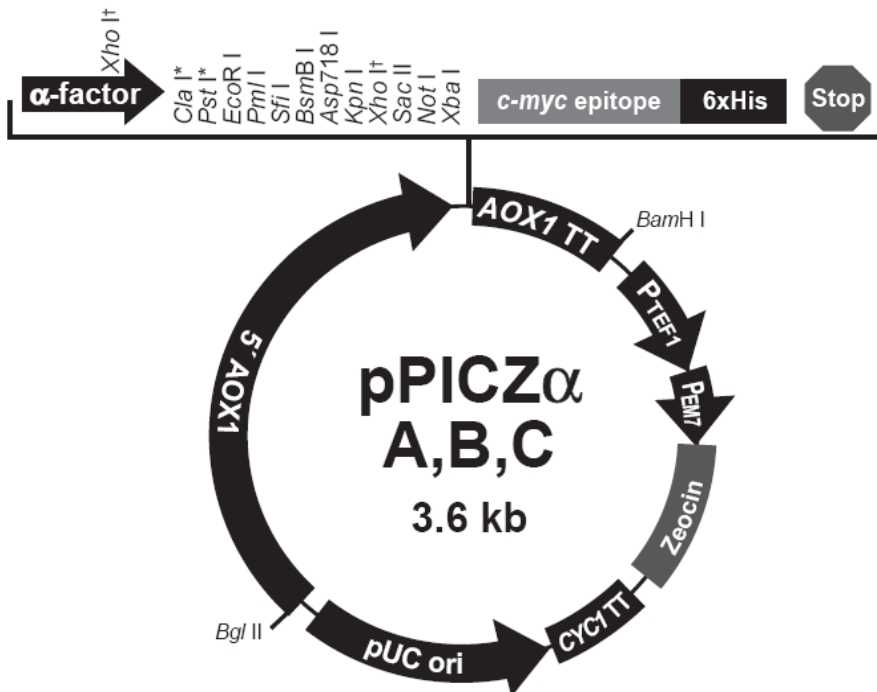


Figure 2-2

pPICZ and pPICZα vectors. (From www.invitrogen.com)

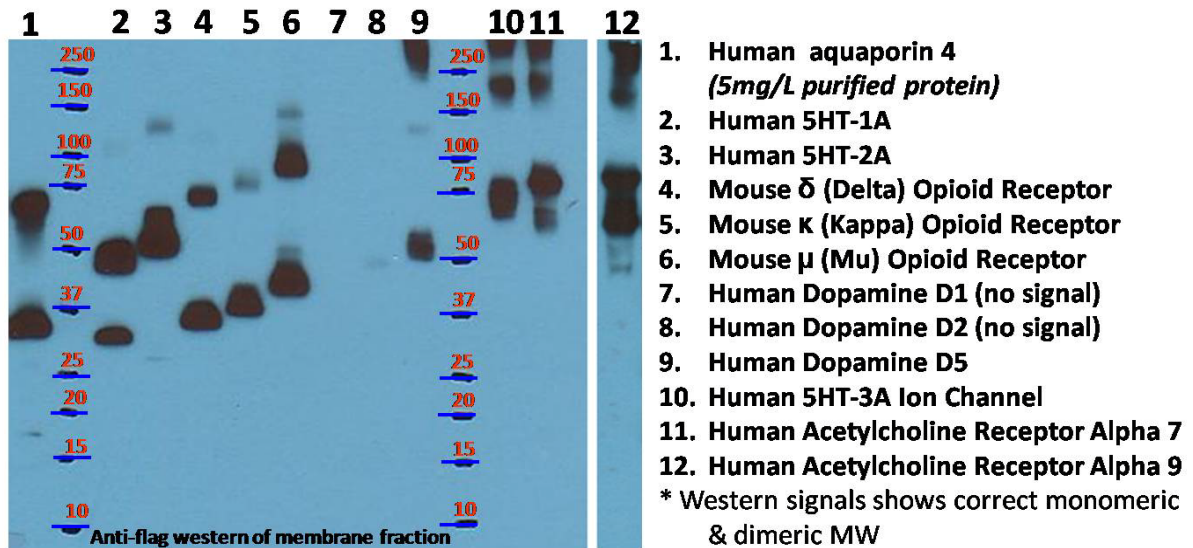


Figure 2-3

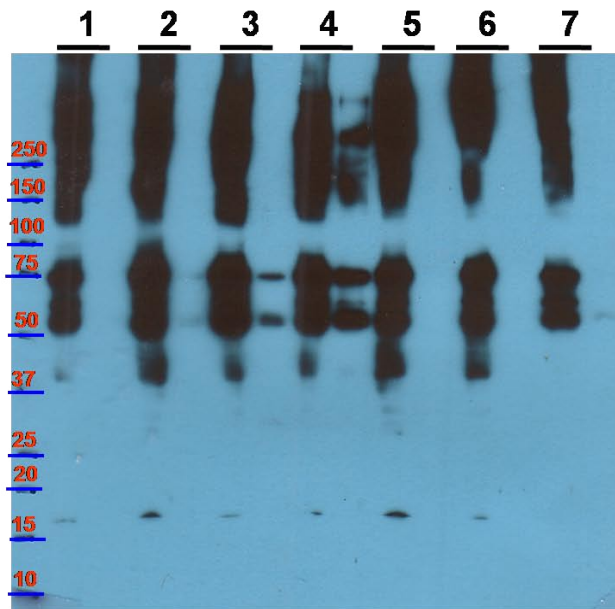
Pichia pastoris expression studies of 12 eukaryotic membrane proteins from human and mouse. Theoretical molecular weight for each protein with the tags and protease site:

1. Human AQP4: 37,862
2. Human 5HT-1A: 49,692
3. Human 5HT-2A: 56,189
4. Mouse δ opioid receptor: 44,100
5. Mouse κ opioid receptor: 46,238
6. Mouse μ opioid receptor: 47,646
7. Human DRD1: 52,879
8. Human DRD2: 54,205
9. Human DRD5: 56,522
10. Human 5HT-3A: 56,783
11. Human AchR α 7: 58,092

12. Human AchR α 9: 55,976

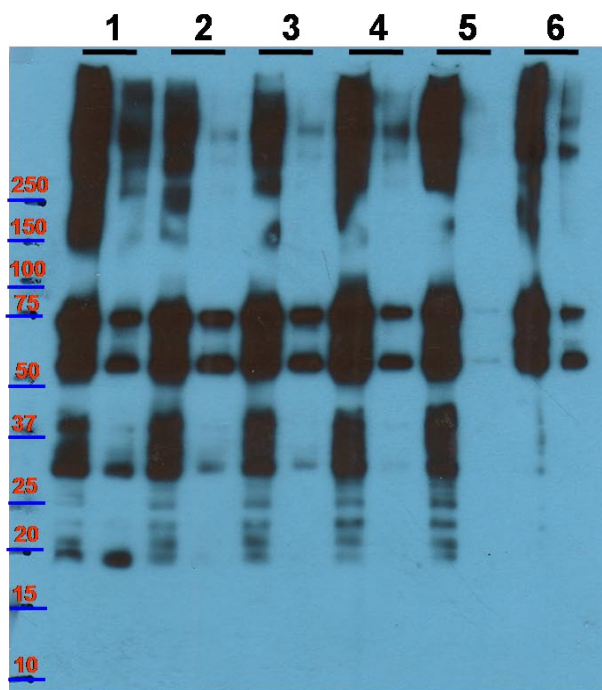
Details for running the SDS-PAGE and Western transfer are described under **Materials and Methods**.

(a)



1. Triton X-100 (1%)
2. Octyl- β -D-Glucoside (200 mM)
3. Nonyl- β -D-Glucoside (100 mM)
4. Fos-choline 14 (20 mM)
5. Dodecyl- β -D-Maltoside (20 mM)
6. Cymol-5 (20 mM)
7. $C_{12}E_8$ (1%)

(b)

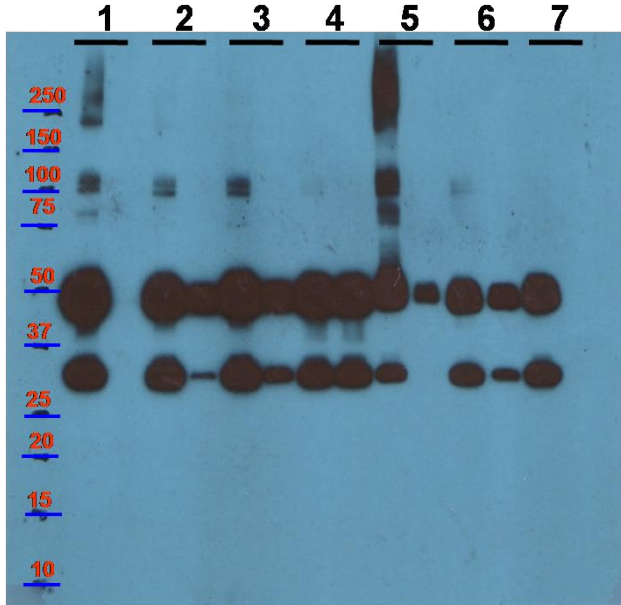


1. $C_{12}E_8$ (1%)
2. Cymol-5 (20 mM)
3. Dodecyl- β -D-Maltoside (20 mM)
4. Nonyl- β -D-Glucoside (100 mM)
5. Octyl- β -D-Glucoside (200 mM)
6. Triton X-100 (1%)

Figure 2-4

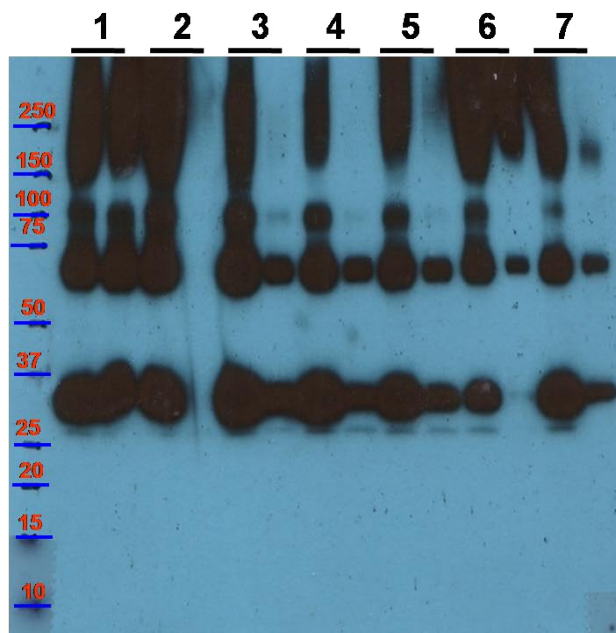
Solubilization studies of the human AchR $\alpha 9$ from cells expressing the protein at the induction temperature of (a) 22°C and (b) 15°C. Concentration of the detergents used during solubilization is shown in the parenthesis. Under each detergent, the *left* side is the sample taken before the ultracentrifugation spin, and the *right* side is the sample of the supernatant after the ultracentrifugation spin. Shown here is the Western signal from the anti-flag M2 antibody. Details for the solubilization are described under **Materials and Methods**.

(a)



1. β -Octylglucoside (200 mM)
2. Cymol-5 (20 mM)
3. Dodecyl- β -D-Maltoside (20 mM)
4. Fos-choline 14 (20 mM)
5. $C_{12}E_8$ (1%)
6. Triton X-100 (1%)
7. 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (20 mM)

(b)



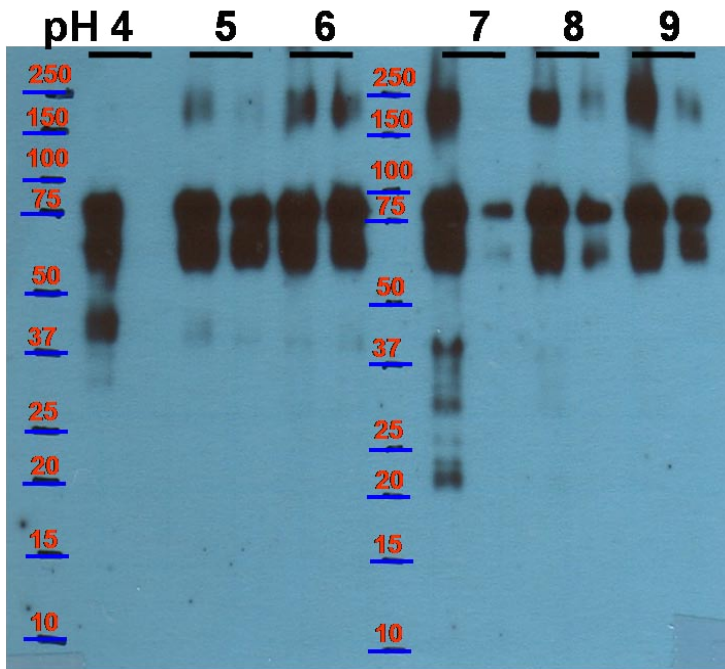
1. SDS (1%)
2. β -Octylglucoside (200 mM)
3. Cymol-5 (20 mM)
4. Dodecyl- β -D-Maltoside (20 mM)
5. Fos-choline 14 (20 mM)
6. $C_{12}E_8$ (1%)
7. Triton X-100 (1%)

Figure 2-5

Solubilization studies of the (a) human 5HT-1A and the (b) mouse δ opioid receptor.

Concentration of the detergents used during solubilization is shown in the parenthesis. Under each detergent, the *left* side is the sample taken before the ultracentrifugation spin, and the *right* side is the sample of the supernatant after the ultracentrifugation spin. Shown here is the Western signal from the anti-flag M2 antibody. Details for the solubilization are described under **Materials and Methods**.

(a)



(b)

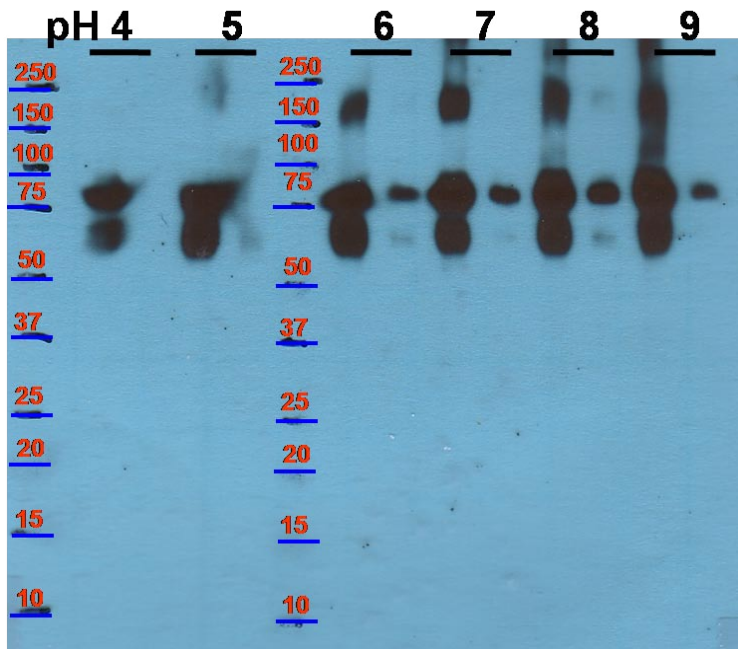
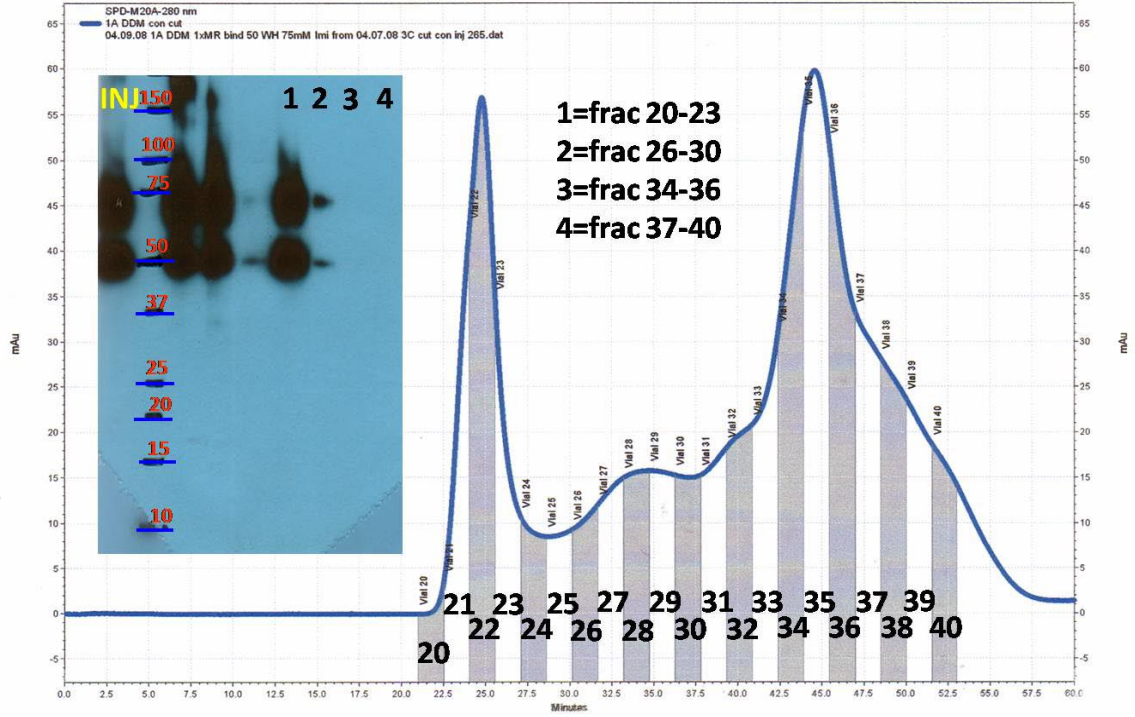


Figure 2-6

Solubilization study of the human acetylcholine receptor subunit $\alpha 7$ in (a) 20 mM FOS-CHOLINE[®]-14 and (b) 20 mM β -dodecyl-D-maltopyranoside (DDM) at different pHs. pH of the resuspended membrane were adjusted with 1 M buffer to a final concentration of 100 mM before mixed with detergents. Buffers used: 1 M Citrate, pH 4.0, 1 M Citrate, pH 5.0, 1 M MES, pH 6.0, 1 M HEPES, pH 7.0, 1 M Tris-HCl, pH 8.0, 1 M CAPS, pH 9.0. Under each detergent, the *left* side is the sample taken before the ultracentrifugation spin, and the *right* side is the sample of the supernatant after the ultracentrifugation spin. Shown here is the Western signal from the anti-flag M2 antibody. Details for the solubilization are described under **Materials and Methods**.

(a)



(b)

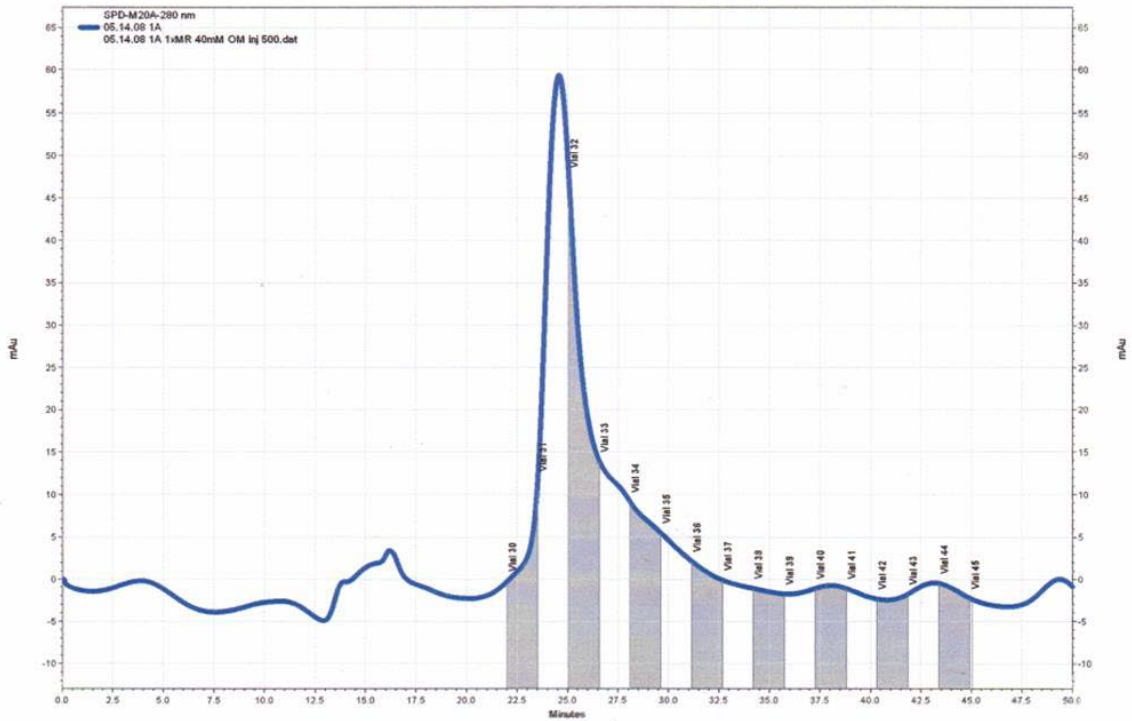
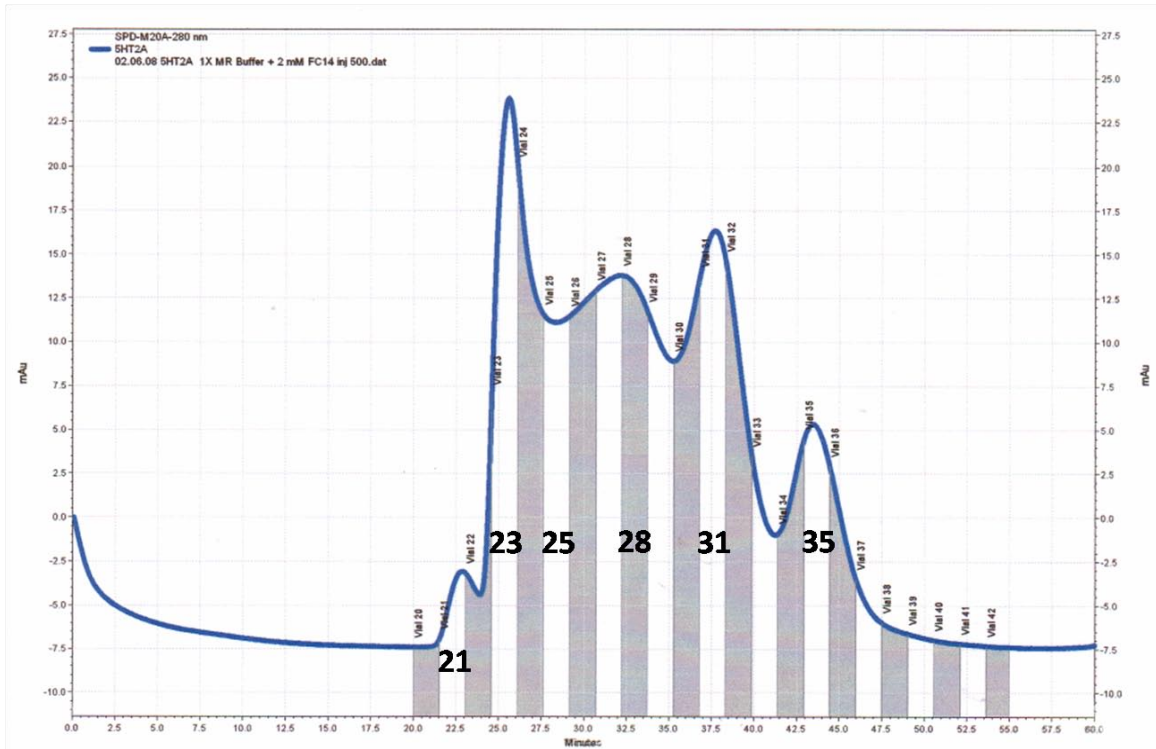


Figure 2-7

Size exclusion chromatography (SEC) of the nickel affinity-purified 5HT-1A. Shown here is the Western signal from the anti-flag M2 antibody. Details for the nickel affinity purification, size exclusion chromatography and the Western protocols are described under **Materials and Methods**.

- (a) SEC in n-dodecyl- β -D-maltopyranoside (DDM). INJ is the nickel affinity purified sample injected into the column. Due to the low protein concentration in each fraction, groups of fractions were selected and concentrated before the SDS-PAGE. Fraction numbers are labeled.
- (b) SEC in n-octyl- β -D-maltopyranoside (OM).

(a)



(b)

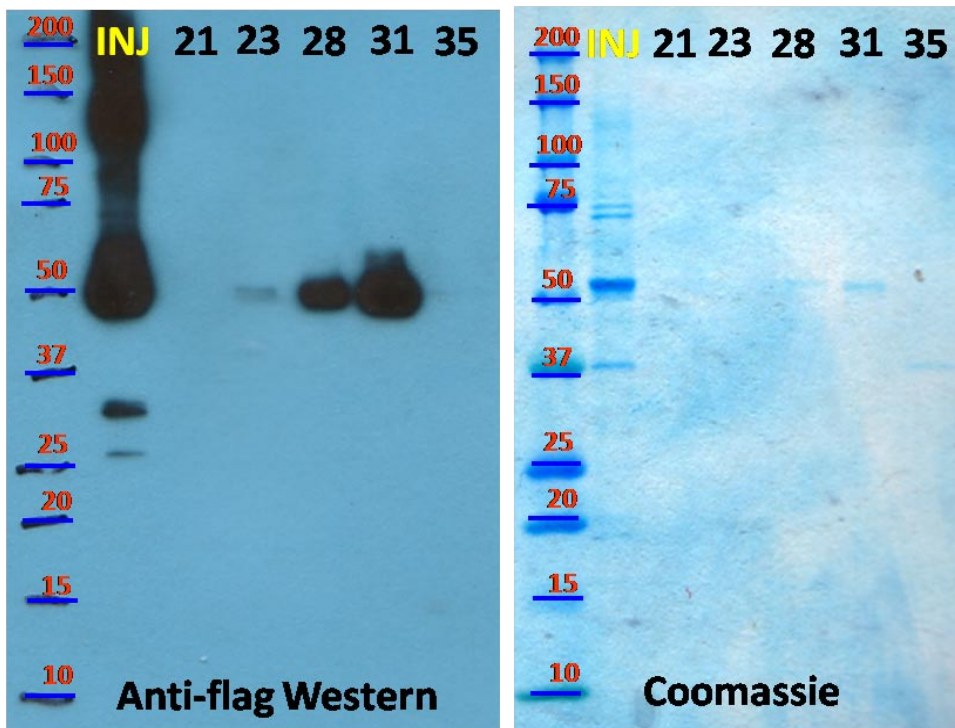
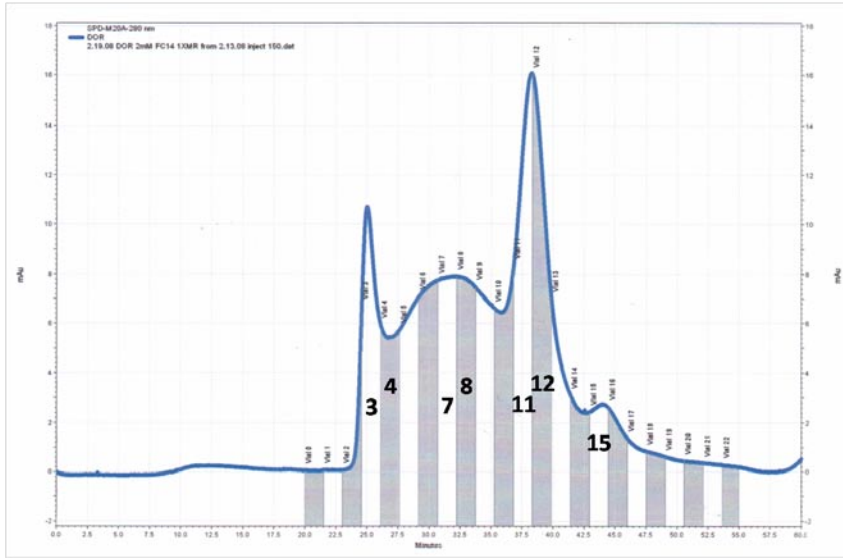


Figure 2-8

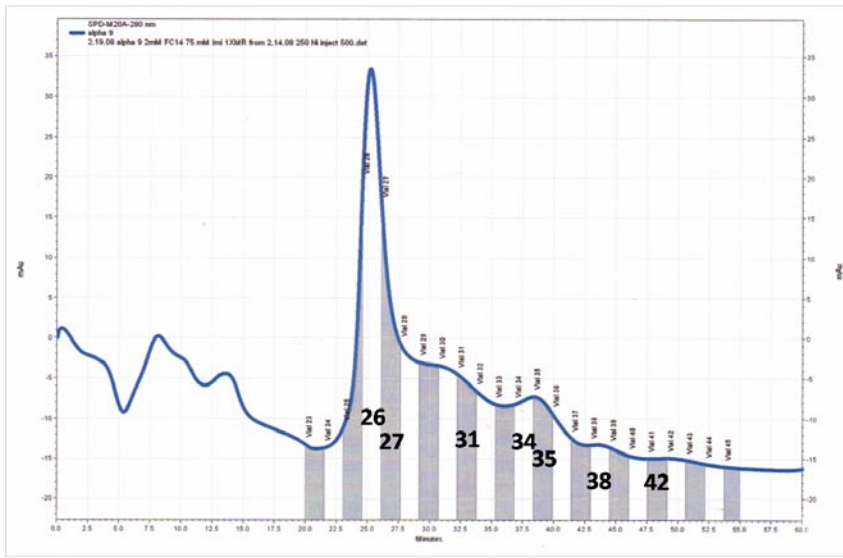
Size exclusion chromatography (SEC) of the nickel affinity-purified 5HT-2A in FOS-CHOLINE[®]-14. Details for the nickel affinity purification, size exclusion chromatography and the Western protocols are described under **Materials and Methods**.

- (a) The SEC profile. Fractions that were picked for SDS-PAGE are labeled.
- (b) The anti-flag Western and the Coomassie-stained gel of the fractions. INJ is the nickel affinity-purified sample injected into the column.

(a)



(b)



(c)

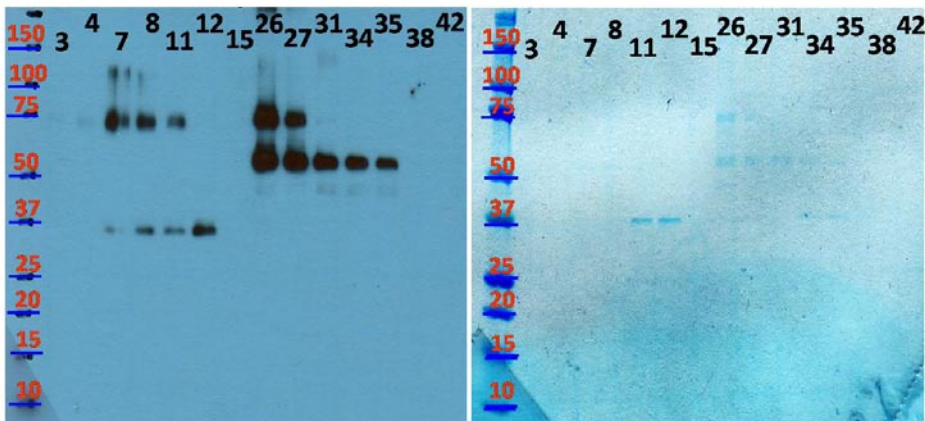


Figure 2-9

Size exclusion chromatography (SEC) of the nickel affinity-purified mouse δ opioid receptor and the human acetylcholine receptor subunit $\alpha 9$ in FOS-CHOLINE[®]-14. Details for the nickel affinity purification, size exclusion chromatography and the Western protocols are described under **Materials and Methods**.

- (a) The SEC profile for the mouse δ opioid receptor. Fractions that were picked for SDS-PAGE are labeled.
- (b) The SEC profile for the human acetylcholine receptor subunit $\alpha 9$. Fractions that were picked for SDS-PAGE are labeled.
- (c) The anti-flag Western and the Coomassie-stained gel of the selected fractions from both proteins.

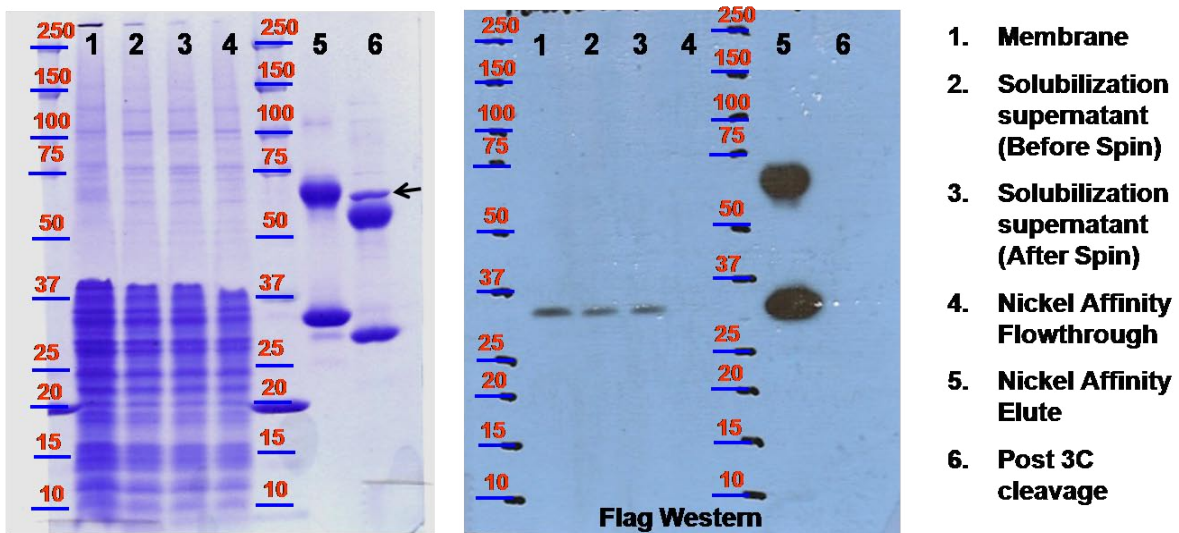


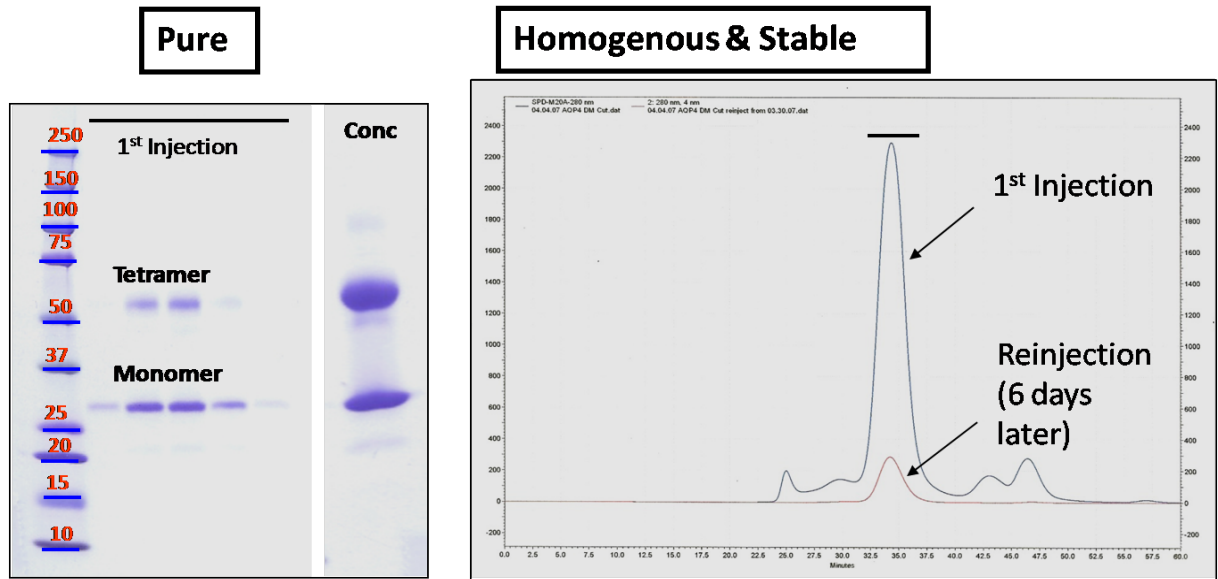
Figure 2-10

Nickel affinity chromatography of human aquaporin 4 in n-decyl- β -D-maltopyranoside (DM).

Left: Coomassie-stained SDS-PAGE. *Right:* Anti-flag Western blot. The **Arrow** points to the 6xHis-MBP-3C protease. Details for the nickel affinity purification, size exclusion

chromatography and the Western protocols are described under **Materials and Methods**.

(a)



(b)

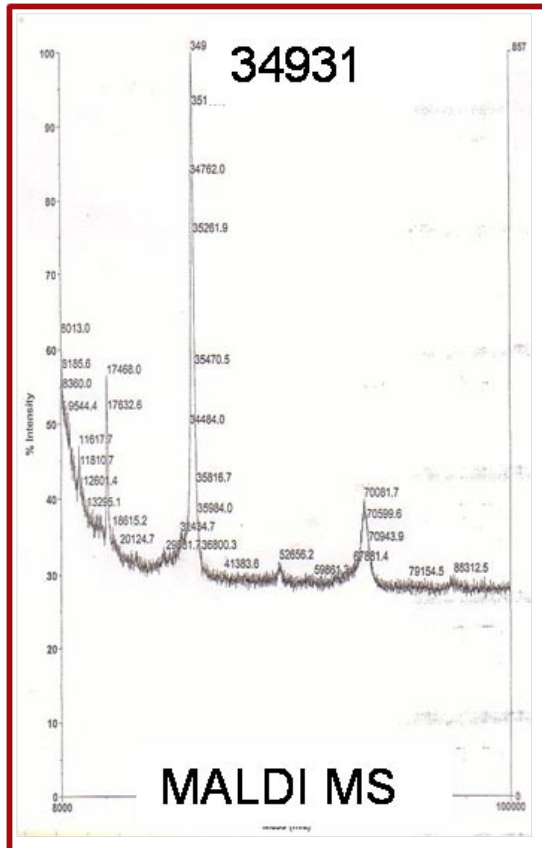


Figure 2-11

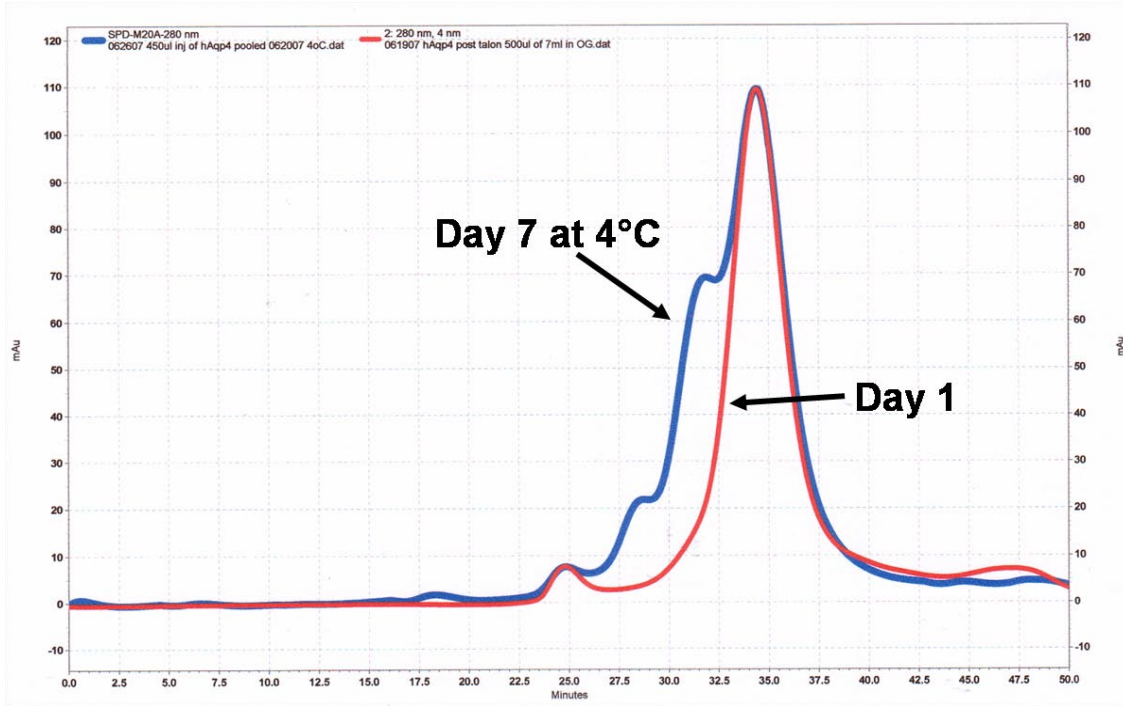
Size exclusion chromatography (SEC) of the nickel affinity-purified human aquaporin 4 in n-decyl- β -D-maltopyranoside (DM). Details of the run are described under **Materials and Methods**.

(a) *Left*: Coomassie-stained SDS-PAGE of SEC run.

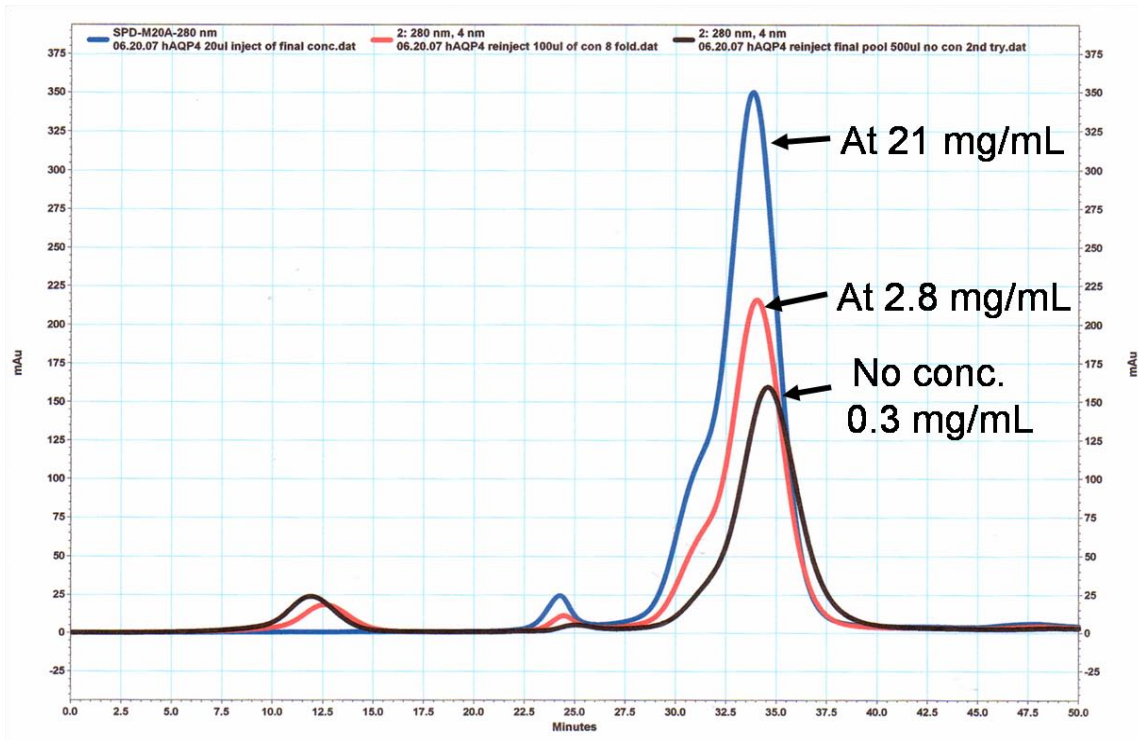
Right: The SEC run in decyl-maltopyranoside. The *red* trace indicates a reinjection of a sample from the peak that was stored at 4°C for 1 week.

(b) The MALDI-MS result of the peak fraction. The theoretical molecular weight of human aquaporin 4 is 34,852.

(a)



(b)



(c)

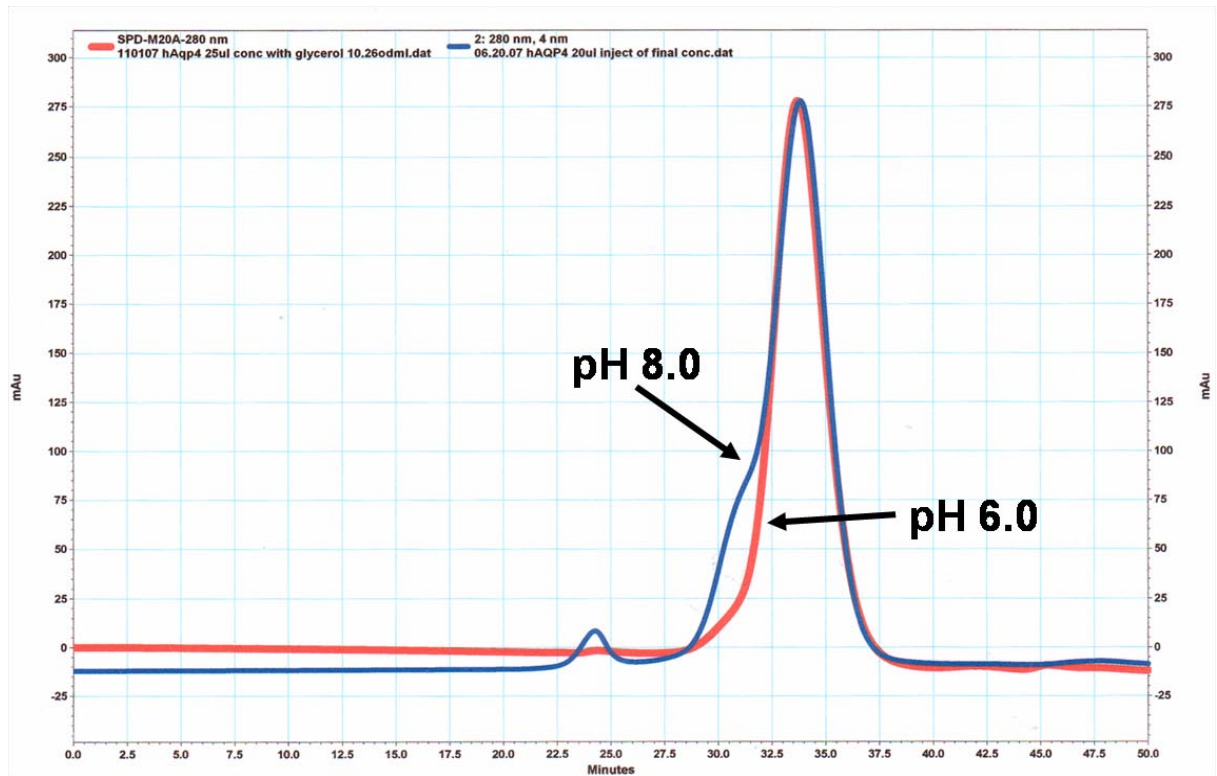


Figure 2-12

Size exclusion chromatography (SEC) of the nickel affinity-purified human aquaporin 4 (AQP4) in n-octyl- β -D-glucopyranoside (OG). Details of the run are described under

Materials and Methods.

- (a) Protein stability over a period of 7 days at 4°C when not concentrated.
- (b) Protein stability during protein concentration over a period of 2-3 hours at 4°C.
- (c) Protein stability of concentrating protein in pH 8.0 and pH 6.0. pH was first adjusted before protein concentration.

	<u>Expression</u>	<u>Solub.</u>	<u>Purif.</u>	<u>TagCleavage</u>	<u>PHS</u>	<u>Crystal/Structure</u>
hAQP4	√	√	√	√	√	√
h5HT-1A	√	√	√	√		
h5HT-2A	√	√	√	√		
mDOR	√	√				
mKOR	√	√				
mMOR	√	√	√			
hDRD1	No					
hDRD2	No					
hDRD5	√					
5HT-3A	√	√	√			
hAchRα7	√	√				
hAchRα9	√	√	√	√		

Table 2-1

A summary of the heterologous expression of human and mouse membrane protein in *Pichia pastoris*. The process is divided into 6 steps: 1. Expression (> 1mg/mL), 2. Solubilization in detergents, 3. Purification by Nickel affinity chromatography, 4. Tag cleavage/removal by 3C protease, 5. Protein purified in pure, homogenous, and stable state, 6. Crystal and structure obtained.

<u>Gene</u>	<u>Tag</u>	<u>Vector</u>	<u>Restriction Sites used</u>
5HT1A	N-flag, C-6xhis	pPICZ	EcoRI and NotI
5HT2A	N-flag, C-6xhis	pPICZ	EcoRI and NotI
DRD1	N-flag, C-6xhis	pPICZ	EcoRI and NotI
DRD2	N-flag, C-6xhis	pPICZ	EcoRI and NotI
DRD5	N-flag, C-6xhis	pPICZ	EcoRI and NotI
δ opioid	N-flag, C-6xhis	pPICZ	EcoRI and SacII
μ opioid	N-flag, C-6xhis	pPICZ	EcoRI and SacII
κ opioid	N-flag, C-6xhis	pPICZ	EcoRI and SacII
AQP4	N-8xhis-flag	pPICZ	EcoRI and NotI
AchR α 7	N-flag, C-6xhis	pPICZ α	EcoRI and NotI
AchR α 9	N-flag, C-6xhis	pPICZ α	EcoRI and NotI

Table 2-2

A summary of constructs tested in *Pichia pastoris*.

REFERENCES

1. Deisenhofer J, Epp O, Miki K, Huber R, Michel H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3[ångström] resolution. *Nature* 1985 318(6047): 618-624.
2. White SH. The progress of membrane protein structure determination. *Protein Sci* 2004 Jul; 13(7): 1948-1949.
3. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* 2008 Oct; 18(5): 581-586.
4. Long SB, Campbell EB, MacKinnon R. Crystal Structure of a Mammalian Voltage-Dependent Shaker Family K⁺ Channel. *Science* 2005 August 5, 2005; 309(5736): 897-903.
5. Tornroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P. Structural mechanism of plant aquaporin gating. *Nature* 2006 Feb 9; 439(7077): 688-694.
6. Ho JD, Yeh R, Sandstrom A, Chorny I, Harries WE, Robbins RA, Miercke LJ, Stroud RM. Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proc Natl Acad Sci U S A* 2009 May 5; 106(18): 7437-7442.
7. Horsefield R, Norden K, Fellert M, Backmark A, Tornroth-Horsefield S, Terwisscha van Scheltinga AC, Kvassman J, Kjellbom P, Johanson U, Neutze R. High-resolution x-ray structure of human aquaporin 5. *Proc Natl Acad Sci USA* 2008 September 9, 2008; 105(36): 13327-13332.
8. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G. Structure of P-Glycoprotein Reveals a Molecular

- Basis for Poly-Specific Drug Binding. *Science* 2009 March 27, 2009; 323(5922): 1718-1722.
9. Higgins DR. Overview of protein expression in *Pichia pastoris*. *Curr Protoc Protein Sci* 2001 May; Chapter 5(Unit5 7).
 10. Daly R, Hearn MT. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 2005 Mar-Apr; 18(2): 119-138.
 11. Newby ZE, O'Connell J, 3rd, Robles-Colmenares Y, Khademi S, Miercke LJ, Stroud RM. Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nat Struct Mol Biol* 2008 Jun; 15(6): 619-625.
 12. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX. Structure of a [bgr]1-adrenergic G-protein-coupled receptor. *Nature* 2008 454(7203): 486-491.
 13. Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI, Kobilka BK. Crystal structure of the human [bgr]2 adrenergic G-protein-coupled receptor. *Nature* 2007 450(7168): 383-387.
 14. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 2007 Nov 23; 318(5854): 1258-1265.
 15. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine

- Receptor Bound to an Antagonist. *Science* 2008 November 21, 2008; 322(5905): 1211-1217.
16. Mohanty AK, Simmons CR, Wiener MC. Inhibition of tobacco etch virus protease activity by detergents. *Protein Expression and Purification* 2003 27(1): 109-114.
 17. Lundbäck A-K, van den Berg S, Hebert H, Berglund H, Eshaghi S. Exploring the activity of tobacco etch virus protease in detergent solutions. *Analytical Biochemistry* 2008 382(1): 69-71.
 18. Zhang R, Kim T-K, Qiao Z-H, Cai J, Pierce Jr WM, Song Z-H. Biochemical and mass spectrometric characterization of the human CB2 cannabinoid receptor expressed in *Pichia pastoris*--Importance of correct processing of the N-terminus. *Protein Expression and Purification* 2007 55(2): 225-235.
 19. Andre N, Cherouati N, Prual C, Steffan T, Zeder-Lutz G, Magnin T, Pattus F, Michel H, Wagner R, Reinhart C. Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Sci* 2006 May; 15(5): 1115-1126.
 20. Mirzabekov T, Bannert N, Farzan M, Hofmann W, Kolchinsky P, Wu L, Wyatt R, Sodroski J. Enhanced expression, native purification, and characterization of CCR5, a principal HIV-1 coreceptor. *J Biol Chem* 1999 Oct 1; 274(40): 28745-28750.
 21. Bass RB, Strop P, Barclay M, Rees DC. Crystal Structure of *Escherichia coli* MscS, a Voltage-Modulated and Mechanosensitive Channel. *Science* 2002 November 22, 2002; 298(5598): 1582-1587.

Chapter 3

Crystal Structure of Human Aquaporin 4 at 1.8Å and its Mechanism of Conductance

Research completed in collaboration with

Ronald Yeh, Andrew Sandstrom, Ilya Chorny, William E.C. Harries, Rebecca A. Robbins,

Larry J.W. Miercke, and Robert M. Stroud

The following chapter was originally published in the *Proceedings of the National Academy
of Science U S A* 2009 May 5; 106(18): 7437-7442.

Data deposition: The atomic coordinates and structure factor data have been deposited in
the Protein Data Bank, www.pdb.org (PDB ID code 3GD8).

ABSTRACT

Aquaporin 4 is the predominant water channel in the mammalian brain, abundantly expressed in the blood-brain and brain-CSF interfaces of glial cells. Its function in cerebral water balance has implications in neuropathological disorders including brain edema, stroke, and head injuries. The 1.8-Å crystal structure reveals the molecular basis for the channel's water selectivity. Unlike the case in the structures of water-selective aquaporins AqpZ and AQP1, the asparagines of the two Asn-Pro-Ala (NPA) motifs do not hydrogen bond to the same water molecule; instead, they bond to two different water molecules in the center of the channel. Molecular dynamics simulations were performed to ask how this observation bears on the proposed mechanisms for how aquaporins remain totally insulating to any proton conductance while maintaining a single file of hydrogen bonded water molecules throughout the channel.

INTRODUCTION

The aquaporin (AQP) family includes both aquaporins that conduct water but not glycerol, and aquaglyceroporins that mediate diffusion of water, glycerol and certain other small molecules in their neutral form across biological membranes. They all exist as homotetramers. In humans, 13 different AQPs (AQP0-12) provide for transport in different tissues, each of which has broad clinical importance (1-3). Besides AQP4, AQP1 and AQP9 are also expressed in the brain (4); AQP1 is expressed in the epithelial cells of the choroid plexus and plays a role in cerebrospinal fluid production, whereas AQP4 is localized to the endfeet of astrocytes in contact with the blood vessels of the blood-brain barrier and in astrocytic processes in contact with synapses. From its tissue-specific concentrated localization in closely packed tetragonal arrays, and the improved response to water intoxication or stroke in AQP4^{-/-} knockout mouse, it is thought that AQP4 is primarily responsible for cerebral water homeostasis (5). AQP4 also may be involved in buffering altered potassium ion concentration after neuronal activity due to its co-distribution with KIR4.1 potassium channels in synapses (6). AQP9 is an aquaglyceroporin also found in astrocytes, and with suggested roles in glycerol and monocarboxylate diffusion and energy metabolism in catecholaminergic neurons.

The secretion and absorption of cerebrospinal fluid is precisely controlled because the brain is encased within the rigid cranium, thus any increase in intracranial pressure caused by edema can lead to compression of brain tissues resulting in neurological disorders and cell death. Because of its role, inhibitors of AQP4 are sought, though so far with conflicting results. Certain quaternary ammonium compounds (7), antiepileptic drugs (8, 9), and

serotonin receptor agonists (10) have been reported to inhibit AQP4 water transport in oocytes with IC_{50} values down to the low μM range. However Yang *et al.* report no inhibition up to 100 μM using different assay methods (11). Therefore, the atomic resolution structure of human AQP4 provides the means of validating, determining the binding interactions of drug lead compounds, discovering, and refining therapeutic agents that might diminish damage from stroke, tumor-associated edema, epilepsy, traumatic head injury and other CNS disorders associated with brain water imbalance.

AQP4 is the primary target in the autoimmune disease neuromyelitis optica (NMO), that primarily affects the optic nerves and spinal cord. In this rapidly progressing and disabling disease, AQP4-specific autoantibodies (NMO-IgG) activate the complement-mediated inflammatory demyelination and necrosis (12). Knowing the structure around the epitopes that are bound by NMO-IgG can facilitate discovery of agents that may compete for, or alter the site without triggering the complement cascade.

AQP4 distribution is concentrated in the perivascular membrane of astroglial cells. There AQP4 forms tetragonal arrays on the plasma membrane (13). On a structural level, AQP4 is unique among AQPs that it exists in two isoforms (M1 and M23) owing to the use of two different translation initiation sites at methionine M1, or at M23 (14). The M1 and M23 isoforms have very different effects on array formation with the shorter isoform favoring larger arrays mediated by two symmetric interactions between Arg108 of each molecule and Tyr250 of another molecule in the neighboring tetramer (15, 16). The M1 and M23 isoforms can form either homo- or heterotetramers, so providing opposing effects on array formation,

hence a mechanism for regulating the organization of the AQP4 arrays (15). The C-terminal three amino acids of AQP4, -SSV, serve as the ligand of a PDZ binding partner, α -syntrophin, which is a component of the dystrophin protein complex that links AQP4 to the actin cytoskeleton (17). Such bridged connection between AQP4 and the actin cytoskeleton allows AQP4 to be anchored at the endfeet of astrocytes, the blood-brain and brain-CSF interface, such that transgenic mice deficient in α -syntrophin completely lack such polarized expression in astrocytes (18, 19). The arrays of AQP4 may serve to tether many AQP4 molecules to each α -syntrophin.

AQP4 is not sensitive to inhibition by mercury (20) because it does not have the reactive cysteine residue in the lumen of the channel corresponding to Cys191 in AQP1 (21). AQP4 conductance is reduced >50% by phosphorylation mediated by protein kinase C at Ser180 (22, 23), and increased ~40% by protein kinase G activity at Ser111 (24). The gating mechanism by phosphorylation events may be similar to that of the spinach AQP SoPIP2;1 (25, 26).

To delineate the cellular organization, immune related properties, and the prospects for inhibitor development, we determined the 1.8 Å crystal structure of human AQP4 from heterologously expressed protein. To date there are only three crystal structures reported for mammalian AQPs, two purified from naturally rich sources, AQP1 in red blood cells, AQP0 from the eye lens (27, 28), and human AQP5, from protein heterologously expressed in *Pichia pastoris* (29).

RESULTS

Overall architecture

Crystals of the M1 isoform full length AQP4 diffracted to $\sim 8 \text{ \AA}$ in space group I4.

Trypsinization improved the resolution. Both full length and trypsinized protein (hAQP4)

have equal water conductance in reconstituted proteoliposomes (**Figure 3-6**). hAQP4

crystallized in space group P4₂1₂, though packing interactions between the tetramers are

completely different than in the electron diffraction structure of rat rAQP4 (16). The rAQP4

structure was determined to $3.2 \times 3.6 \text{ \AA}$ with no water or glycerol molecules observed. The

X-ray structure of hAQP4 at 1.8 \AA resolution shows discrete water molecules throughout the

channel, five glycerol, and one octylglucoside molecule. Each monomer, surrounded by six

and two half-length alpha-helices (M1 to M8) tetramerizes along the crystallographic four-

fold c axis (**Figure 3-1**) (30).

The central four-fold axis

The physiological four-fold axis also insulates against all solutes and water. On the

cytoplasmic side, a four-fold arrangement of water molecules is stabilized by the backbone

amides of Ser188 and Gly189 (**Figure 3-1c**). Throughout $\sim 21 \text{ \AA}$ of the mid-membrane

section, Phe195, Leu191, and Leu75, repeated four times, create a hydrophobic block. This

contrasts with the four-fold axis in *Plasmodium falciparum* aquaglyceroporin, PfAQP, where

the region around the four-fold symmetry axis is blocked by four aliphatic chains of

phospholipids or fatty acids (31), and the human AQP5 where a single lipid molecule is

found (29). There is little conservation around the four-fold symmetry axis among AQPs of

known structure.

Rat AQP4 and Human AQP4

Though hAQP4 and the electron diffraction structure of rAQP4 in lipid bilayers crystallize in the same space group (P4₂1₂), their crystal lattice contacts lie on different surfaces of the protein. The hAQP4 three-dimensional crystal contains head-to-head contacts only, as tetramers within the horizontal plane are too far apart ($a=82.1 \text{ \AA}$) to make contact with each other (**Figure 3-7a**). The rAQP4 two-dimensional crystal lattice has tetramers closer together ($a=69.0 \text{ \AA}$) and contains both in-plane contacts presumably representing the *in vivo* arrangement in a lipid bilayer, and between-plane contacts of the latticed tetramers. Based on the molecular contacts in the crystal, the interaction between the short 3₁₀ helices in the C-loop was proposed to be a possible mechanism for AQP4-mediated cell-cell adhesion (**Figure 3-7b**) (16). Though the sequences of the C-loop are the same in hAQP4 and rAQP4 (**Figure 3-8**), hAQP4 does not adopt the short 3₁₀ helix in this region (**Figure 3-9**) suggesting that it is induced on association with molecules of neighboring membranes.

The extracellular vestibule, the selectivity filter, and the conducting pore

AQP4 is a water-selective channel. Signature to the water-selective channels, His201 lies directly in the selectivity filter, reducing the channel diameter to $\sim 1.5 \text{ \AA}$, sterically excluding the passage of glycerol (**Figure 3-2**). This is unlike the aquaglyceroporins of known structures GlpF and PfAQP, where conserved aromatic rings of phenylalanine and tryptophan form walls of a hydrophobic corner that allows favorable interaction with the glycerol backbone (**Figure 3-10**). AQP4 was purified and crystallized in the presence of 5% (v/v) glycerol (0.7 Molar), and three glycerol molecules are found in the extracellular vestibule,

though not in the selectivity filter where the two glycerol-conducting AQPs, GlpF and PfAQP, bind glycerol identically to one another (30, 31) (**Figure 3-10**). Thus the basis of glycerol conductance is coupled to the presence of large aromatic groups at these two locations (Phe200, Trp48 in GlpF). In the water-selective rat AQP1, double mutant Phe56Ala and His180Ala (Phe77 and His201 in hAQP4) (**Figure 3-10**) allows for the passage of glycerol, showing that steric occlusion is one mechanism for exclusion of larger solutes (32).

The ~25 Å long conducting pore contains a line of water molecules and no solute molecule consistent with the highly water selective nature of the channel. However, the electron density of the water molecules are distributed along the pore with residual positive Fo-Fc density observed in between water positions indicating increased anisotropic distribution along the channel axis, implying low energy barriers between the water molecules along the direction of the channel (**Figure 3-3**). As in other aquaporins, the pathway through the channel is amphipathic. The hydrophobic sides are formed by the side chains of Phe77, Ile81, Val85, Leu170, Ile174, and Val197. The eight backbone carbonyls, of Gly93, Gly94, His95, Ile96, from the cytoplasmic side and Gly209, Ala210, Ser211 and Met212 form the hydrophilic hydrogen bond acceptors that traverse the entire length of the pore and function as hydrogen bond acceptors for 8 positions of water molecules in transit. This arrangement allows bi-directional conductance of water from either side of membrane.

The asparagines N213 and N97 of the two almost totally conserved Asn-Pro-Ala (NPA) motifs form the canonical 'fireman's grip-like' structure in the center of the pore (30), and provide the defining central force that orients water as it passes through the mid-point of the

channel. In AQP4 however, each asparagine donates its single, highly oriented hydrogen bond to a separate water molecule (**Figure 3-3 & Figure 3-5**). This is a key variant since in three other water-selective aquaporin structures (27, 28, 33), the two asparagines donate hydrogen bonds to a single water molecule, orienting the central water and so polarizing every other water molecule throughout the outgoing pathways of the pore. A comparison of the two variants in central coordination of water between hAQP4 and bovine AQP1 is shown in **Figure 3-5**.

Conductance of hAQP4 in proteoliposomes

Both full-length hAQP4 and trypsinized hAQP4 were reconstituted into proteoliposomes, and water and glycerol conduction were measured. Vesicle shrinkage in response to either water or glycerol efflux was monitored by light-scattering in a stopped-flow apparatus. The proteoliposomes containing the full-length protein conducts water at the rate of $120.3 \pm 9.5 \text{ s}^{-1}$, 12 times the rate for liposomes $9.80 \pm 0.70 \text{ s}^{-1}$, and do not conduct any glycerol, with measured rate of 0.071 ± 0.001 , the same as for protein-free liposomes 0.071 ± 0.0001 . Proteoliposomes containing the trypsinized protein conduct water at the rate (relative rates) of $107.4 \pm 1.5 \text{ s}^{-1}$ compared to $5.71 \pm 0.05 \text{ s}^{-1}$ for liposomes, and do not conduct glycerol with measured rates of 0.077 ± 0.002 , the same as 0.071 ± 0.0001 for protein-free liposomes (**Figure 3-6a**). Thus both forms of AQP4 conduct water but do not conduct glycerol. The data also demonstrate that the protein expressed in *Pichia pastoris* is functional, and that trypsinolysis of the protein does not have an apparent effect on water conductivity.

DISCUSSION

Implication of the C loop in cell-cell adhesion and neuromyelitis optica

X-ray structures of aquaporins derived from all three kingdoms: bacteria, archaea, and eukaryote have been determined. The C loop (residues 137 to 155 in hAQP4) that connects M4 and M5 and forms part of the extracellular vestibule shows remarkable similarities and differences across all kingdoms suggesting that they are conserved for functional reasons. GlpF is unique in that part of the C loop folds as two α -helices. Both AqpM and the spinach AQP SoPIP2;1 have a two-turn helix entering the C loop. PfAQP has a short helix that is very similar to its bacterial aquaglyceroporin homolog, GlpF (**Figure 3-1d**). Among others, rat AQP4 stands out in that it has a one-turn 3_{10} helix (**Figure 3-1e**), and is similar to the two-turn helix seen in AqpM and the spinach AQP SoPIP2;1 (**Figure 3-1f**). Hiroaki *et al.* proposed from the rAQP4 electron diffraction structure at $\sim 3.5\text{\AA}$ resolution that the 3_{10} helix in the C loop (residues 139 to 142, **Figure 3-1e** & **Figure 3-9**) is the main region that provides for AQP4-mediated cell-cell adhesion (16). Human AQP4 and rat AQP4 share the same protein sequences in this loop (**Figure 3-8**), but in the X-ray structure of hAQP4, even though this region has hydrogen bonds of an α -helix, it is a well-defined loop and does not adopt the dihedral angles (φ , ψ) of a 3_{10} helix (**Figure 3-9**). Therefore the 3_{10} helix in the C loop is a conformation that may be induced when two cells expressing AQP4 on the surface are in close proximity bringing two C loops into direct contact.

The affinity of individual tetramers of AQP4 for each other between cells may be extremely weak since they are not sufficient to favor the contacts in 3-dimensional crystals of either AQP4, or AQP0, while the preponderance of many, perhaps hundreds of such weak contacts

between lattices of equally spaced molecules, may provide the avidity necessary to support cell adhesion mediated by the transition to the 3_{10} helical conformation of the C-loop. It is also possible that the 3_{10} helix is non-physiological, induced by the packing of the rAQP4 double-layered structure in the electron microscopy samples (**Figure 3-7b**) (16). In two different cell types, each with or without expression of orthogonally arrayed AQP4, separate cell-cell adhesion and dynamic light scattering assays demonstrated that AQP4 does not have any cell-cell adhesion property (34). Since in our 1.8-Å hAQP4 structure, the C loop is not involved in the crystal packing, the lack of any helix in the C loop therefore supports the conclusion that AQP4 does not strongly drive adhesion and that the helices in the C loop observed in the rAQP4 structure may be induced by crystal contact. The forces that determine the in-plane contacts may also be weak by the same argument though detergent displaces the natural interface there.

Nevertheless, given the diversity and the flexibility of the C loop, it is possible that the C loop could play an important role in the pathogenesis of neuromyelitis optica (NMO), an auto-immune inflammatory disease. It has been proposed that the three extracellular loops (A, C, and E) could independently, or jointly form the antigenic epitopes for the NMO-IgG autoantibody (35). It is therefore possible that the variable conformations of the C loop could aid in the affinity of the autoantibody to AQP4 by providing a conformationally flexible epitope favorable to binding, so strengthening the downstream cascades of complement-mediated inflammatory response.

Recruitment of solutes in the extracellular vestibule

In the crystal structures of aquaglyceroporins GlpF (30), and PfAQP (31), and the archaeal aquaporin AqpM (36), glycerol molecules are located in the selectivity filter of the channel as well as in the entry vestibule (**Figure 3-11**). In aquaglyceroporins GlpF and PfAQP, the vestibule may serve as a place for glycerol recruitment and desolvation of solutes for transport through the channel. However, for water-specific aquaporins AQP1, the spinach AQP SoPIP2;1, bovine AQP0, and human AQP5, glycerol was not included in the crystallization conditions. But for water-specific aquaporins AqpZ where 5% glycerol (0.7 M, final concentration) was present during crystallization, glycerol molecules were also not found in the vestibule (**Figure 3-11**) (33). Human AQP4 is a water-selective channel. The 1.8-Å structure (crystallized in the presence of 5% glycerol), has no glycerol in the selectivity filter, but has three glycerol molecules non-specifically bound in the extracellular vestibule. The extracellular vestibule, a canyon defined primarily by the A loop and C loop, interacts with the glycerol molecules through hydrogen bonds of the amide backbone with the hydroxyl groups of glycerol. This supports the conclusion that the aquaglyceroporin-specific selectivity filter is the key to glycerol transport, and that the alternative water-specific arrangement prevents passage of glycerol. It is tantalizing to speculate that the locations of the glycerol molecules in the vestibule could represent ‘fragment binding sites’ and might suggest binding sites for defining inhibitors that would bind from the extracellular side.

Since there are tremendous prospects for drugs that inhibit AQP4, and indeed many inhibitors of AQP4 have been described in the literature to date, we co-crystallized hAQP4 with 5 mM of three such compounds, tetraethylammonium (TEA), acetazolamide, and

rizatriptan and determined their co-crystal structures. In oocytes swelling assays, these compounds were reported to be AQP4 inhibitors with IC_{50} in the μ M-range (7-10), however we have not been able to detect any compound bound in the structures (data not shown). We then reconstituted purified human AQP4 into liposomes and measured water conductance in the presence of these compounds. TEA up to 10 mM had no effect on AQP4 water conductance. Acetazolamide and rizatriptan do inhibit water conductance with an approximate IC_{50} in the low mM range (\sim 3 mM and \sim 1 mM respectively, **Figure 3-6b**). This is much higher than the μ M-range IC_{50} reported by Huber *et al.* (8, 10). We argue that water conductance measurement using purified proteoliposomes is more reliable than in oocytes. In a similar experiment, acetazolamide was also found to inhibit rat AQP4 in the mM range, but not human AQP1 (37). These results call into question the previous interpretation of acetazolamide and rizatriptan as μ M inhibitors of AQP4 inhibition.

Arginine 216 environment determines conductance of water in aquaporins

In hAQP4 and other water-selective aquaporins AQP1, AqpZ, AQP0, spinach AQP SoPIP2;1, and AQP5, the conserved arginine (216 in AQP4) at the heart of the selectivity filter is almost identically configured. Since the glycerol conducting PfAQP conducts water as well as a water channel while GlpF conducts water very poorly, we proposed that the efficiency of water conductance may be proportional to the effective neutralization of the formal positive charge on the arginine side chain, by five hydrogen bond acceptors, three from the protein, and two as donors to water in transit (31). Human AQP4 is a water-specific channel, and similar to all high water conductance AQPs, the $N\epsilon H$, $N\eta^1 H$, $N\eta^2 H$ of Arg216 in the selectivity filter are all hydrogen bonded to other acceptor oxygens of the protein, leaving

two, one each from $N\epsilon H$ and $N\eta H$ as donors to the waters in transit (**Figure 3-4**). For the bovine water-selective AQP1, the arginine (Arg197) environment is very similar to that in human AQP4 (**Figure 3-4**). However in the glycerol conducting GlpF, with low water conductance, the selectivity filter arginine (Arg206) has only one of the $N\eta H$ satisfied (**Figure 3-4**), so it is possible that the extra cost of desolvating the fully charged guanidinium leaves GlpF 'holding on' to water molecules in transit. Thus the unusual nature of arginine among all amino acid side chains, which uniquely is always positively charged, and its compensation that requires 5 in-plane hydrogen bond acceptors, may 'hold-on' to water more strongly when it is not fully compensated. This higher degree of the guanidinium cation buffering by the protein may be the basis for higher efficiency water transport among the AQPs.

The variant role of the dual NPA motifs in proton exclusion

The interlocking of two almost totally conserved NPA motifs in aquaporins provides two highly oriented donors from $N\delta H_2$ of the two asparagines to the center of the channel. We proposed they play a key role in insulation against any conduction of protons or ions while supporting a single file of hydrogen bonded water molecules throughout the entire length of the channel (38). In the crystal structures of the water-selective channels, *E. coli* AqpZ (33) and bovine AQP1 (27), the central water is positioned in between the two asparagines of the NPA motif and receives two hydrogen bonds from the two equidistant asparagine $N\delta H_2$ s (represented by bAQP1 in **Figure 3-5a & b**). Such orientation of the central water may be a factor in preventing proton conduction through the channel by the highly cooperative hop-and-turn Grotthuss relay mechanism (39, 40) because the central water cannot rotate to

accept or exchange its proton (30, 41). However, the hAQP4 crystal structure does not have a water molecule centrally located between the two NPA regions. Instead, each asparagine of the NPA motif (Asn213 and Asn97) donates a hydrogen bond to a different water molecule (**Figure 3-5**). This is also the case in the 2.0-Å structure of human AQP5 (29) and in spinach AQP SoPIP2;1 (26). Therefore, in this subclass of AQP structures, these two water molecules may each be somewhat freer to rotate in a concerted fashion leaving one additional non-hydrogen bonded water hydrogen. These AQPs equally insulate against any proton leakage, suggesting that the two models for the central waters may represent intermediates that are very close in free energy, and they also indicate that the barrier to water transport through this dual NPA region is small.

To see if this observation constitutes a difference in mechanistic terms, molecular dynamic (MD) simulations were performed in order to assess the probability of finding a single central water molecule bonding to both NPA asparagines simultaneously. The results of the MD simulations were used to calculate the number density of water molecules along the channel near the two NPA regions (**Figure 3-5c**). Four simulations were carried out. In the first simulation (*thin* solid line: —), the system commenced from the protein and crystallographic waters; the protein positions were kept frozen throughout the simulation. The resulting number density is consistent with the observed water positions in the hAQP4 crystal structure (solid red spheres) and suggests that this particular configuration of the protein supports two energy minima for water in the NPA region. In order to test whether this result was biased by the introduction of the crystallographic water molecules into the simulation, a second simulation (dotted line: - - -) was performed in which the protein

positions were kept frozen but the crystallographic waters were removed from the simulation and the pore evacuated. Following equilibration with the remaining simulation water molecules, the resulting number density is consistent with the experimentally observed water positions in the hAQP4 crystal structure and gives further support that the calculation is stable and reiterates the particular observed configuration of the human AQP4. In the third simulation (*thick solid line: ———*) the system contained both the protein and the crystallographic waters, but instead of freezing the atomic positions of the protein, the heavy atoms (N, C, O) were restrained to their crystallographic positions using a harmonic restraint. This slight release of the protein atoms from their crystallographic positions allows the number density to shift such that a water molecule positions itself in the middle of the NPA region, as is observed in the crystal structures of bAQP1 (**Figure 3-5a&b** & **Figure 3-12**). Finally, in the fourth simulation (dashed line: - - -) all the positional restraints were removed from the protein, allowing the protein to move freely. Again the resulting number density suggests the presence of a water molecule in the middle of the NPA region. Thus, we conclude that the observed position of the two waters near the NPA region, in the human AQP4 crystal structure, do not conform to a functionally distinct variant but rather, signal the lowered barriers to conductance between these sites, hence the low barriers to water molecule conductance by AQP4.

Analogous sets of simulations were performed on bAQP1, in which a water molecule is seen crystallographically to accept two donors in the center of the two NPA asparagines. In all four simulations, the number density supports the mechanism that as a water molecule travels through an aquaporin or aquaglyceroporin channel, it does transition into and through a

central position where it simultaneously accepts hydrogen bonds from both NPA asparagine donors, and that in this transition state, the line of waters throughout the channel are polarized such that all waters have their dipoles oriented away from the central water molecule (**Figure 3-12**).

Our first AQP structure, for GlpF (30), led to our suggestion that the highly ordered, polarized water orientation throughout the channel was a key factor in the exclusion of passage of protons in a fully concerted fashion without developing charge within the channel itself (38). Other considerations have also been raised. The dehydration cost for any ion is prohibitive, hence no small ions, or hydronium ions are conducted by a mechanism that would involve mass transport of the only partially solvated ion. The Warshel group calculated the electrostatic barrier of placing a proton/hydronium ion into the center of the channel and concluded that this electrostatic desolvation barrier alone excludes mass transfer of hydronium ions to the center of the channel, and that the contribution of the NPA asparagines and the helix-induced dipoles at the NPA region are an insignificant addition to a barrier for hydronium ions at this position (42-44). However, if a proton were transferred through the channel by a Grotthuss mechanism, no formal charge need ever appear within the channel because of the concerted hopping of protons, in at one side and out at the other. The Grotthuss mechanism has been assessed (45) by Poisson-Nernst-Planck theory that indicates that there are high barriers to the Grotthuss mechanism. Thus the cost of this degree of ordering may abrogate the Grotthuss mechanism through a line of water that is this long. Using free-energy simulations and continuum electrostatic calculations, the Roux group adds weight to the proposal that the dehydration of an ion as it approaches the channel presents a

major barrier to mass transfer of an ion through the channel, but concludes that other factors including the NPA regions within the channel are important as well (46). Using multistate empirical valence bond (MS-EVB) models for proton transfer, Ilan *et al.*(41) suggests that the bipolar line of water, the charge on the conserved arginine of the selectivity filter, and ion desolvation penalties are all relevant as we proposed initially (38).

Perspective

Because of the flexibility of the termini, the N-terminal 19 amino acids and the C-terminal 64 amino acids were removed from the M1 isoform by trypsinolysis, as determined by mass spectrometry and N-terminal sequencing. Thus the crystal structure we report at 1.8 Å resolution contained Glu20 to Lys259. The resulting structure has visible electron density from Gln32 to Pro254. In contrast, the electron diffraction structure of rat AQP4 was expressed as the M23 isoform, and revealed density for residues 31-254, consistent with a natively unstructured region at each of the termini. Thus while the structure fully elucidates all features of the water transport pathway it does not yet address the question of how the M1 isoforms works together with the M23 isoform to limit the extent of orthogonal array formation. It also remains to be determined how the C-terminus interacts with α -syntrophin in the context of polarized expression in astrocytes (6).

Questions also remain on the mechanism of how Ser111 and Ser180 phosphorylation affect AQP4 conductance. Studies have shown that the phosphorylation state of Ser111 on the B loop and Ser180 on the D loop affect water conductance of AQP4 (22-24) (**Figure 3-8**).

Phosphorylated Ser180 has been speculated to interact with Lys259, Arg260 and Arg261 as a

means of tethering the C-terminal domain to block the channel (47), and phosphorylated Ser111 has been proposed to mimic the gating mechanism of the spinach AQP SoPIP2;1 by disrupting the network of hydrogen bonds that anchor the D loop to the N-terminus and thereby opening the channel (24) (26). In our 1.8 Å crystal structure, the channel is open and while density of the B and D loops are clear, no phosphate is observed on either serine. Because the protein was trypsinized, neither the N and C terminal domains are present. While we cannot comment on the interaction of Ser180 with the C-terminal domain, we think it is unlikely that the gating of AQP4 at Ser111 is similar to SoPIP2;1 because the D loop of hAQP4 is 4 residues shorter than in SoPIP2;1 and residues in the N-terminal domain that are involved in binding to the D loop (Asp28 and Glu31) are different amino acids in hAQP4 (**Figure 3-8**).

The high resolution structure of the phosphorylated protein, the M1 isoform, and complexes with binding elements of syntrophin may address these issues of larger scale assemblies beyond that of the tetrameric AQP4.

MATERIALS AND METHODS

Protein expression and purification

The full-length human aquaporin 4 gene (NCBI accession number NM_001650) was purchased through Origene (Rockville, MD). The expression construct was designed with an N-terminal 8xHis followed by a flag tag (DYKDDDDK) and a human rhinovirus 3C protease cleavage site (LEVLFQ↓GP) and cloned into the EcoRI and NotI sites of pPICZ expression vector (Invitrogen). The expression vector was then electroporated into *Pichia pastoris* X-33 cells (Invitrogen) using the BioRad Gene Pulser Xcell System following a standard yeast electroporation protocol. Transformation was then selected on YPD plates with 50 µg/mL Zeocin™ (Invitrogen). Four colonies were restreaked and tested for expression. For production, the yeast was cultured in BMY media (Invitrogen) in Fernbach flasks at 30°C for 24 hrs, then the temperature was lowered to 26°C, and methanol was added directly to the cultures to a final concentration of 2.5%. The cultures were grown for another 48 hrs before harvest. Cultures harvested by centrifugation at 4°C at 6000 x g for 10 minutes. Pellets were washed once with TBS buffer with 1 mM β-ME, and 1 mM PMSF and pelleted again. Cells were then resuspended with the same buffer and lysed by bead beating with glass beads. Broken and unlysed cells were removed by centrifugation at 4°C at 6000 x g for 10 minutes while the membranes remained in the supernatant. The membranes were then pelleted at 160,000 x g at 4°C for 1 hr. Pellets were resuspended in MR Buffer (25 mM Tris-HCl, pH 7.4 at room temperature, 250 mM NaCl, 10% glycerol, 1 mM β-ME) and stored at -80°C until further processing.

To begin purification, resuspended membrane was solubilized by adding 400 mM n-octyl- β -D-glucopyranoside (OG) (Anatrace) to a final concentration of 200 mM and stirred at 4°C for 1 hr. Unsolubilized material was pelleted at 160,000 x g at 4°C for 30 min. 5 M imidazole, pH 7.4, was added to the supernatant to a final concentration of 50 mM. The supernatant was then batch bound with Ni-NTA resins (Qiagen) for 2 hrs, loaded onto a BioRad Econo Column and washed with MR Buffer with 40 mM OG and 50 mM imidazole, and then eluted with 300 mM imidazole. Imidazole was removed using Econo-Pac DG10 desalting column (BioRad) equilibrated with MR Buffer with 40 mM OG. The N-terminal tag was cleaved by His-tagged MBP fusion of human rhinovirus 3C protease (His6-MBP-3C) at 4°C overnight. Uncleaved hAQP4 (<5%) and 6xHis-MBP-3C were removed the next day with TALON® resin (Clontech). Then hAQP4 was concentrated in a 50,000 molecular weight cut-off Amicon spin concentrator (Millipore) and further purified by size exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in 25 mM Citrate, pH 6.0, 50 mM NaCl, 5% glycerol, 40 mM OG, and 2 mM DTT (Crystal Buffer). Purified hAQP4 was concentrated again to 30 mg/mL and stored as aliquots at -80°C. The final yield of hAQP4 was ~15 mg of purified and concentrated protein per liter of cells.

Trypsinolysis of hAQP4: Concentrated hAQP4 (post size exclusion chromatography) was diluted with MR Buffer to 1 mg/mL, and immobilized TPCK-trypsin (Cat # 20230, Thermo Scientific) was added. The mixture was incubated by gentle rocking at 25°C overnight. Immobilized TPCK-trypsin was removed by filtration, and trypsinized hAQP4 was further purified in the same way as the full length on size exclusion chromatography. MALDI-MS and N-terminal sequencing were performed on trypsinized hAQP4, and the stable trypsinized

product was identified as N-ENIMV...PDVEFK-C, which removes the N-terminal 19 residues and the C-terminal 64 residues. Trypsinized hAQP4 was concentrated to 30 mg/mL and stored at -80°C until crystallization.

Crystallization

Full length hAQP4 was crystallized in 25% PEG 2000-MME, 50 mM Citrate, pH 6.0, 5% Glycerol, 40 mM OG, and 2 mM DTT by hanging drop vapor diffusion at 25°C. Bi-pyramidal crystals grew to a maximum size of ~100 microns within 1 week and diffracted to a best resolution of 8 Å.

Trypsinized hAQP4 was crystallized in 25% PEG2000-MME, 50 mM MES, pH 6.5, 5% Glycerol by sitting drop vapor diffusion at 25°C. Both bi-pyramidal crystals and rod crystals were found within 1 week. The Silver Bullet Screen condition A1 (Hampton Research) was found helpful in growing the bi-pyramidal crystals. The rod crystals grew to a maximum size of 1.5 mm and diffracted to a best resolution of 6 Å. The bi-pyramidal crystals grew to a maximum size of 50 microns diffracted to a best resolution of 1.8Å. The best freezing condition for the crystals was a 3-second soak in a solution of 50% paraffin (v/v) and 50% paratone-N (v/v) (Hampton Research) before flash-freezing in liquid nitrogen.

Data collection and model building

Diffraction data were collected using a wavelength of 1.11Å at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, California, United States). Data were processed using HKL2000 (48). Molecular replacement was

performed with Phaser (49) using the rat AQP4 electron diffraction structure (PDB code: 2D57) (16) as a search model. Subsequent iterative cycles of manual building and retrained refinement were done using Coot (50) and Refmac5 in CCP4 (51). TLS refinement in Refmac5 was applied in the last stage using 19 TLS groups obtained from the TLS analysis server (52) (53) (**Table 3-2a & 3-2b**). Riding hydrogen atoms were generated during refinement, but not written to the output. Structure was assessed using PROCHECK (54, 55) and MolProbity (56). Data processing and refinement statistics are summarized in **Table 3-1**.

Molecular dynamic simulations of hAQP4 and bAQP1

MD simulations were carried out with the Gromacs (57) 3.3.1 simulation software. The simulations were performed on a 256 processor computer cluster running Microsoft Windows 2003 Compute Cluster Server operating system. We used the human aquaporin 4 (hAQP4) X-ray crystal structure as a starting conformation. The tetramer was centrally placed into a palmitoyloleoylphosphatidylethanolamine (POPE) lipid bilayer area containing 364 lipids, and was solvated on both sides with 25507 SPC water molecules (58). The OPLSAA force field (59, 60) was used to model the protein. Lipid parameters were an extension of the parameters of Berger *et al.* (61) that were optimized for the OPLSAA force field (62). 45 chloride ions and 45 sodium ions were added to the periodic simulation box to create a salt concentration of 50 mM. The total system size was 108,939 atoms.

Following energy minimization, two 500 ps simulations were carried out with the protein atoms frozen in their crystallographic positions. The first simulation contained the crystallographic waters, while the second simulation did not. A third 1 ns simulation was

then carried out with the starting positions taken from the end of the first simulation in which the heavy atom (N, C, O) positions of the protein were restrained to their crystallographic position with a harmonic spring-like restraint, having a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. A fourth 1 ns simulation was then performed, starting from the end of the third simulation, in which the restraints were removed and the protein allowed to move freely.

The number of water molecules along each of the four channels were collected every 1ps and used to calculate the number densities in **Figure 3-5c**. In the second simulation it took the water < 200 ps to fully populate the channel region. Number densities were calculated over 100 ps intervals and block averaged to test for convergence. In all cases, convergence of the water density in the channel region was observed.

Throughout these simulations, the temperature was kept constant by coupling the system to a temperature bath of 298 K (63). Likewise, the pressure was kept constant, in the third and fourth simulations, by coupling the system to a pressure bath at 1.01 bar (64). The xy (membrane plane) and z (membrane normal) directions, respectively, were independently coupled with a coupling constant τ of 1 ps. Electrostatic interactions were calculated using PME (65) with an electrostatic cut off of 1.2 nm. The Lennard–Jones potential, describing the van der Waals interaction, was cut off at 0.9 nm. The Settle algorithm (66) was used to constrain the bond lengths and angles of the water molecules, and Lincs was used to constrain all other bond lengths (67), allowing a time-step of 2 fs.

Proteoliposome assay

Lipid stocks were made from *E. coli* polar lipid extracts (Avanti Polar Lipids) at 50 mg/mL in water plus 4 mM β -mercaptoethanol and stored at -80°C . Before the assay, stocks were thawed and 250 μL lipid and 100 μL water were mixed in a 16 x 125 mm glass culture tubes (VWR) and overlaid with argon to reduce oxidation. Liposomes were formed by sonicating the culture tubes in a bath sonicator (Laboratory Supplies Company, Inc.). The reconstitution buffer mixture was formed by mixing (in order): 100 mM MOPS, pH7.5, 43 mM β -OG, full-length or trypsinized hAQP4 (final concentration 0.8 mg/mL) and sonicated lipids (final concentration 8 mg/mL) in a total volume of 1.5 mL. It was then dialyzed against the assay buffer (20mM HEPES, pH 7.4) in a 25,000 molecular weight cut off Spectra/Por Float-a-lyzer (Spectrum Laboratories, Inc.) for 48 hrs. Liposomes were then harvested by centrifugation at $75,000 \times g$ at 4°C for 1 hr. For water conduction, liposomes were resuspended in 1 mL of assay buffer. For glycerol conduction, liposomes were resuspended in 1 mL 20 mM HEPES, pH 7.3 plus ~ 550 mM glycerol (equal in osmolarity to sucrose buffer: 20 mM HEPES, pH 7.3 with 570 mM sucrose). Upon mixing 100 μL of liposome resuspension with 100 μL of sucrose buffer in a stopped-flow apparatus, vesicle shrinkage was monitored by light scattering at 440 nm over time at 12°C . For assaying water conduction in the presence of acetazolamide and rizatriptan, liposomes were incubated for 1 hr in assay buffer with 1 mM or 10 mM of compound and assayed against the sucrose buffer with compound. The resulting data points were fitted to a single exponential curve to extract the rate constant.

ACKNOWLEDGEMENTS

Research was supported by the US National Institutes of Health (NIH) grant GM24485 to R.M.S. and by the NIH Roadmap center grant P50 GM073210. I.C., the computer cluster and the simulations was supported from a scientific computing grant to R.M.S. from Microsoft Corporation. We thank Joseph Luft and Jennifer Wolfley at the Center for High-Throughput Structural Biology at the Hauptman-Woodward Medical Research Institute for setting up the drop-volume ratio crystallization trials of the full length AQP4 and Liang Li from the Ismagilov Laboratory in University of Chicago for his help in the capillary microbatch crystallization trials of the same; both produced crystals though neither yielded better than the 8Å resolution we obtained for the full length molecule. We thank John Lee and Zach Newby for helpful comments on the manuscript. We thank James Holton for his assistance at the Advanced Light Source (ALS) Beamline 8.3.1 supported by NIH grant GM074929 (R.M.S.).

AUTHOR CONTRIBUTIONS

J.D.H. performed all the molecular biology and optimized the protein expression, purification, crystallization, and the proteoliposome assays; J.D.H. and R.Y. collected the diffraction data and determined the structure; A.S., W.E.C.H., R.A.R. and L.J.W.M. assisted in the protein production and purification; I.C. performed the molecular dynamics simulation calculations; R.M.S. supervised the research; J.D.H., I.C., and R.M.S. wrote the manuscript.

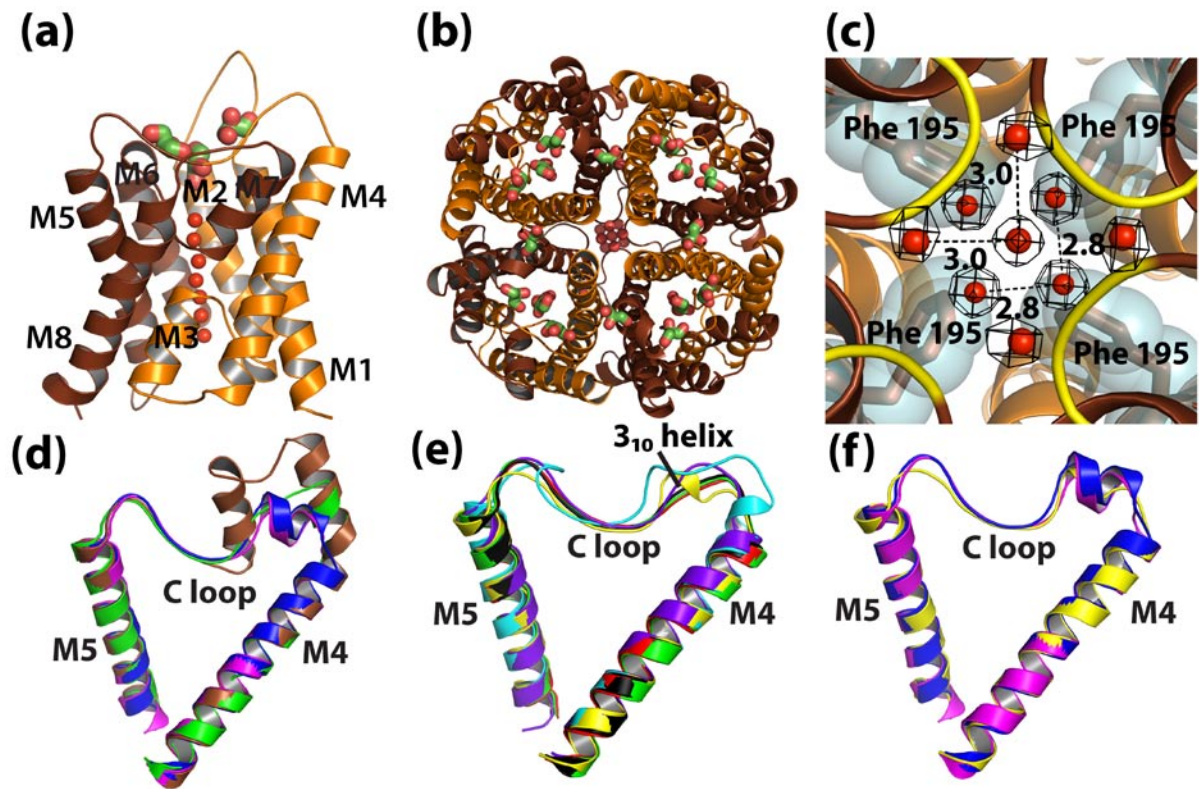


Figure 3-1

General features. (a-b) Monomer and tetramer views of human aquaporin 4 (hAQP4) in cartoon representation. Brown and orange colors represent the N- and C-terminal pseudo two-fold related portions. Water molecules are represented as red spheres, and glycerol molecules are shown as green sticks. (a) The side view of the monomer. Helices are labeled M1 to M8. (b) The tetramer viewed from the extracellular side down the crystallographic four-fold symmetry axis. (c) The network of water molecules found at the intracellular side of the central pore. The central pore is at the crystallographic four-fold symmetry axis and is formed by the tetramer. $2F_o - F_c$ density of the water molecules is shown in black, contoured at 1.2σ . The backbone amides of Ser188 and Gly189 are colored yellow in cartoon representation. Phe195 is shown as brown stick and cyan surface. (d-f) Cartoon

representation of the C loop of all the aquaporin X-ray structures solved to date. (d) *E. coli* GlpF (brown), archeal AqpM (magenta), spinach AQP SoPIP2;1 (blue), PfAQP (green). (e) rat AQP4 (yellow), human AQP4 (black), human AQP5 (red), *E. coli* AqpZ (cyan), bovine AQP0 (green), bovine AQP1 (purple). (f) Comparison of the 3_{10} helix of rat AQP4 (yellow) with the two-turn helix of AqpM (magenta) and spinach AQP (blue). All structural renderings were made with PyMOL (<http://www.pymol.org>).

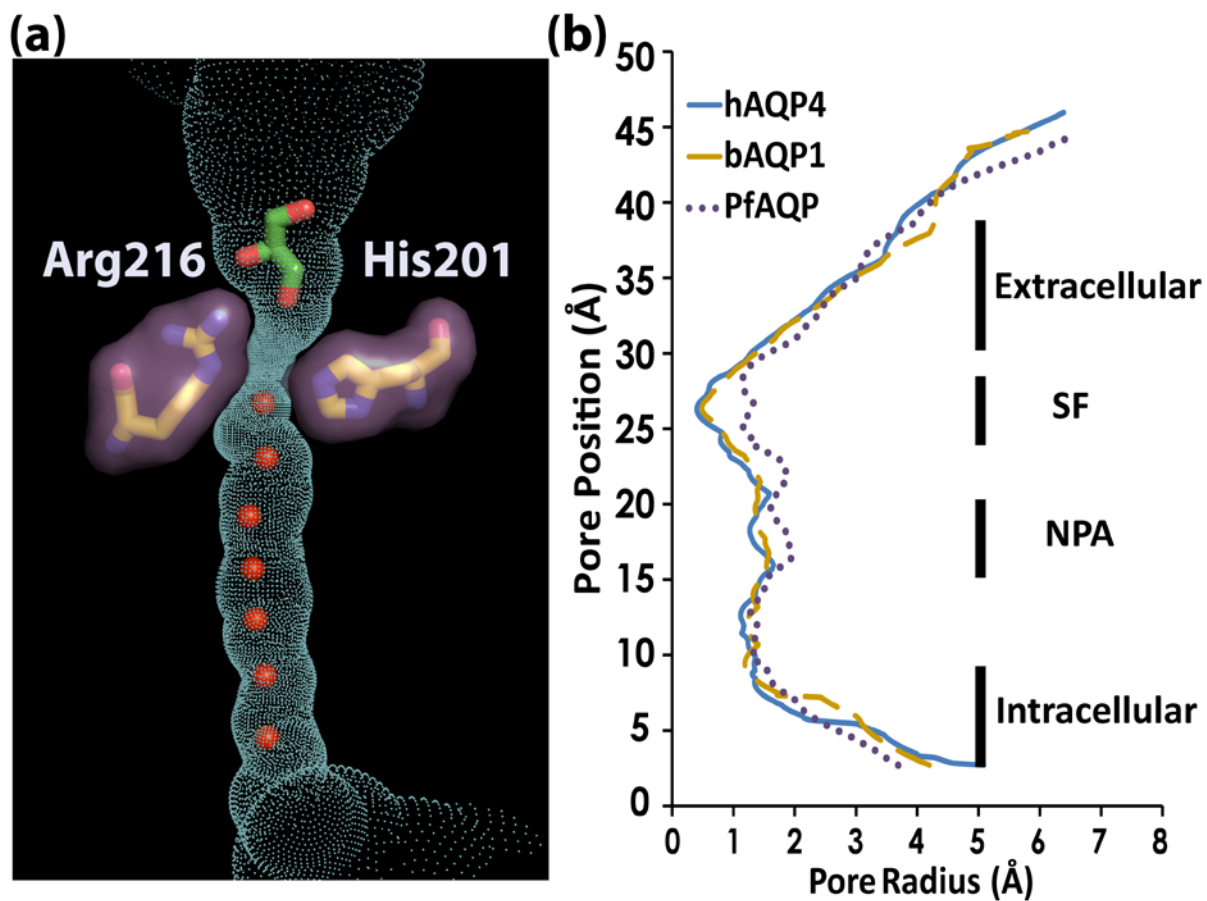


Figure 3-2

The conducting pore. (a) A schematic representation of the conducting pore. The trace of the pore inner surface is shown in cyan. The selectivity filter residues: Arg216 and His201 are shown as sticks with surfaces in purple. The glycerol molecule is shown as green stick, and the water molecules in the channel are shown as red spheres. (b) Plot of the channel radius versus position along the pore for human AQP4, bovine AQP1 (bAQP1), and the *Plasmodium falciparum* AQP (PfAQP). Regions of the channel are labeled as extracellular vestibule, the selectivity filter (SF), the NPA motif, and the intracellular vestibule. The pore inner surface and its dimension are calculated using Hole 2.0 (68).

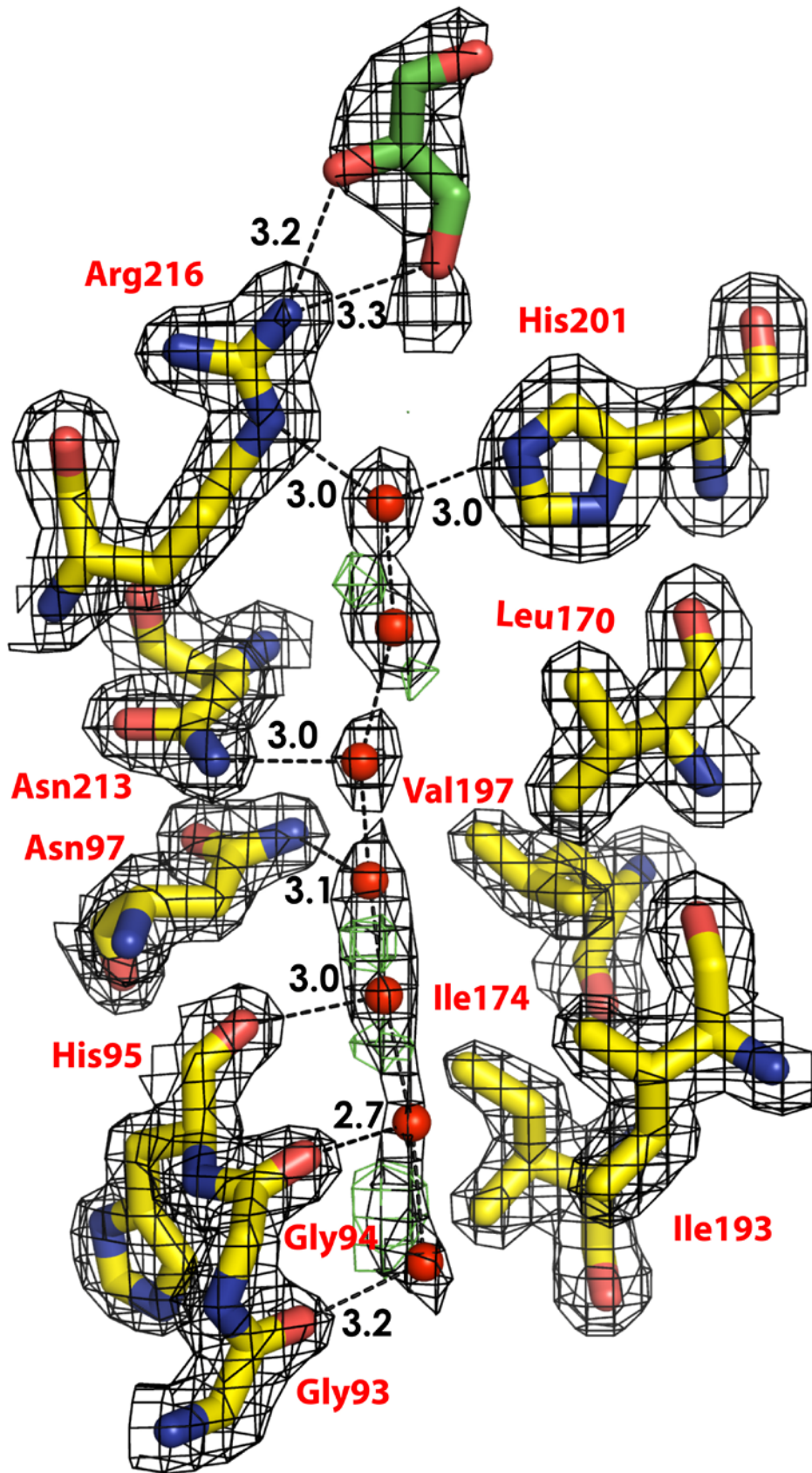


Figure 3-3

Electron density of the conducting pore. Residues that form the wall of the pore are shown in sticks. Water molecules are shown as red spheres. The glycerol molecule is shown in green stick. $2F_o-F_c$ density are shown in black, contoured at 1.2σ . Positive F_o-F_c density is shown in green, contoured at 3σ . There is no negative F_o-F_c density.

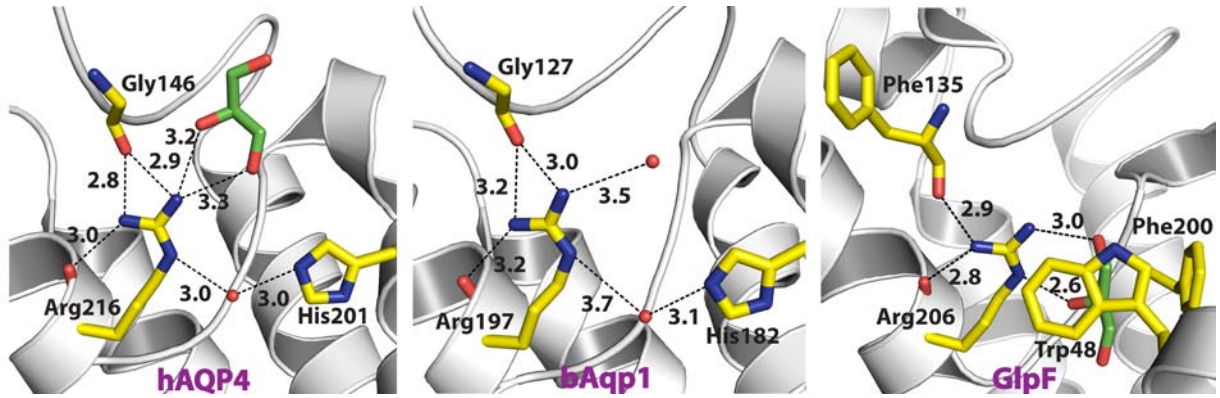


Figure 3-4

Comparison of the hydrogen bond network of the selectivity filter arginine of hAQP4, bAQP1, and GlpF. Protein C-alpha is shown in cartoon representation. Residues of the selectivity filter and glycerol molecules are shown as sticks. Water molecules are shown as red spheres.

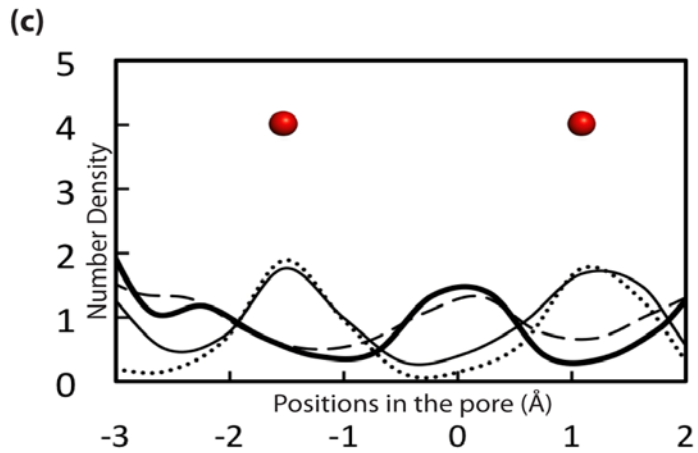
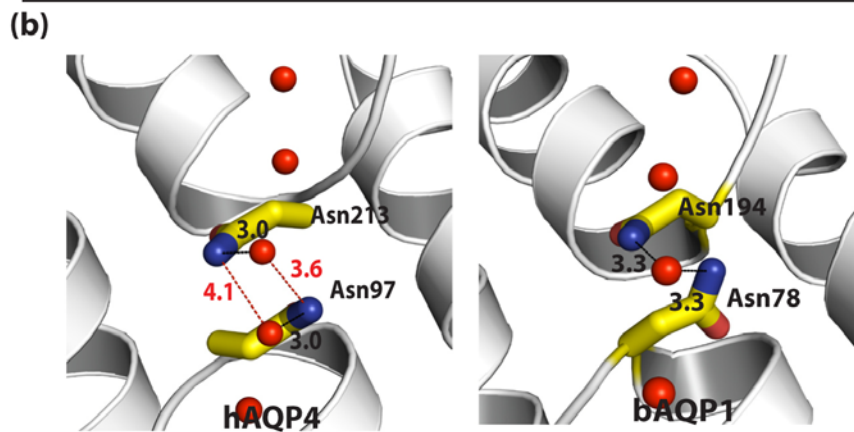
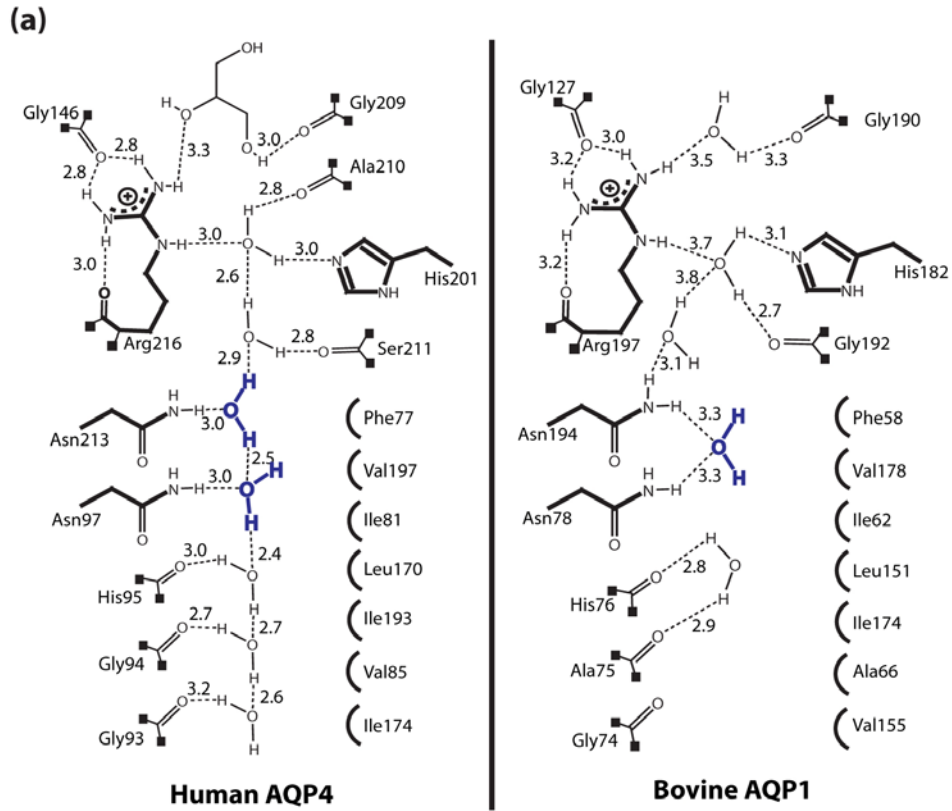
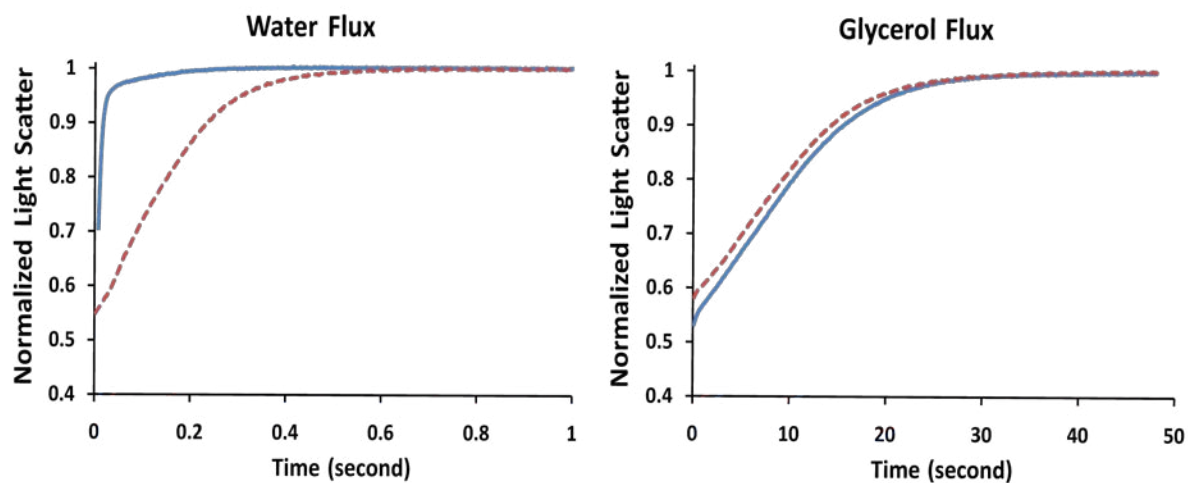


Figure 3-5

The NPA motifs. (a) Schematic representation of the hydrogen bonding network through the channels of hAQP4 and bAQP1. The distances are between heavy-atom to heavy-atom. (b) Stick representation of the NPA motif of hAQP4 and bAQP1. Distances that are too long to be a hydrogen bond are colored in red. (c) Plot of the MD simulations of hAQP4 from four different experiments. Legend for the plot is described in Discussion: The variant role of the dual NPA motifs in proton exclusion.

(I) Full Length Protein



(II) Trypsinized Protein

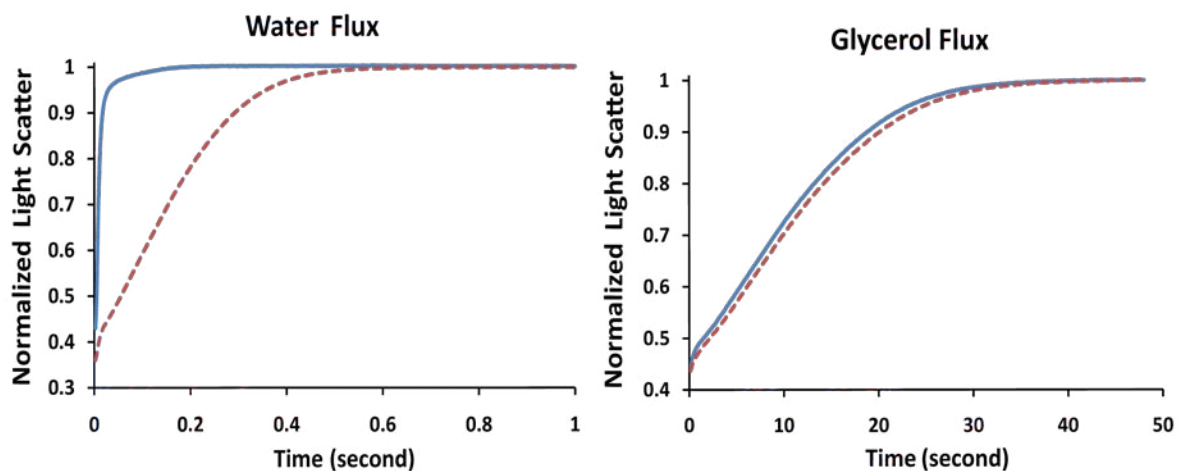


Figure 3-6a

Water and glycerol conduction assays for the (I) full-length and (II) trypsinized protein.

Proteoliposome with AQP4 is shown as blue line. Liposome without protein is shown as red dashes. Plots were generated from average values from six replicate measurements. Rates were determined from fitting a single exponential curve to the plot.

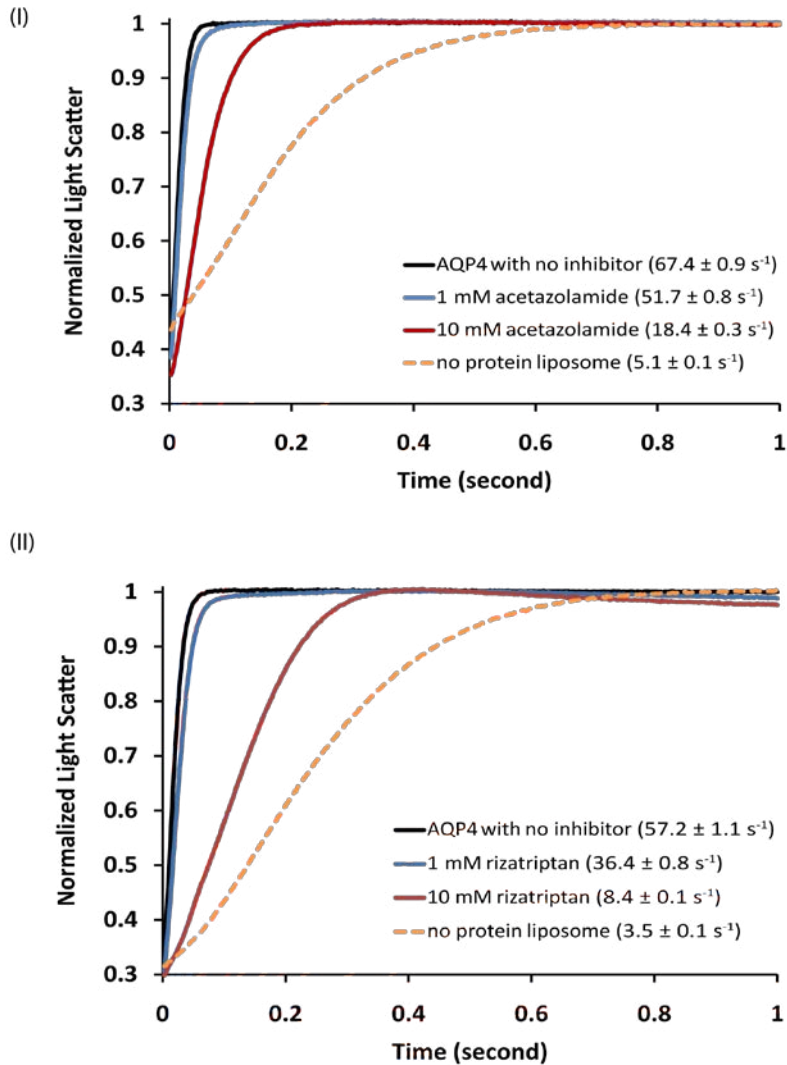


Figure 3-6b

Water conduction assays in the absence and presence of 1 mM or 10 mM of (I) acetazolamide and (II) rizatriptan. Results using the full-length human AQP4 are displayed, but the trypsinized protein yielded similar results. Proteoliposomes treated with 1 mM or 10 mM of TEA conducted as fast as without. Empty liposomes were assayed with and without inhibitors and no change of water conduction was observed. Plots were generated from average values from six replicate measurements. Rates were determined from fitting a single exponential curve to the plot.

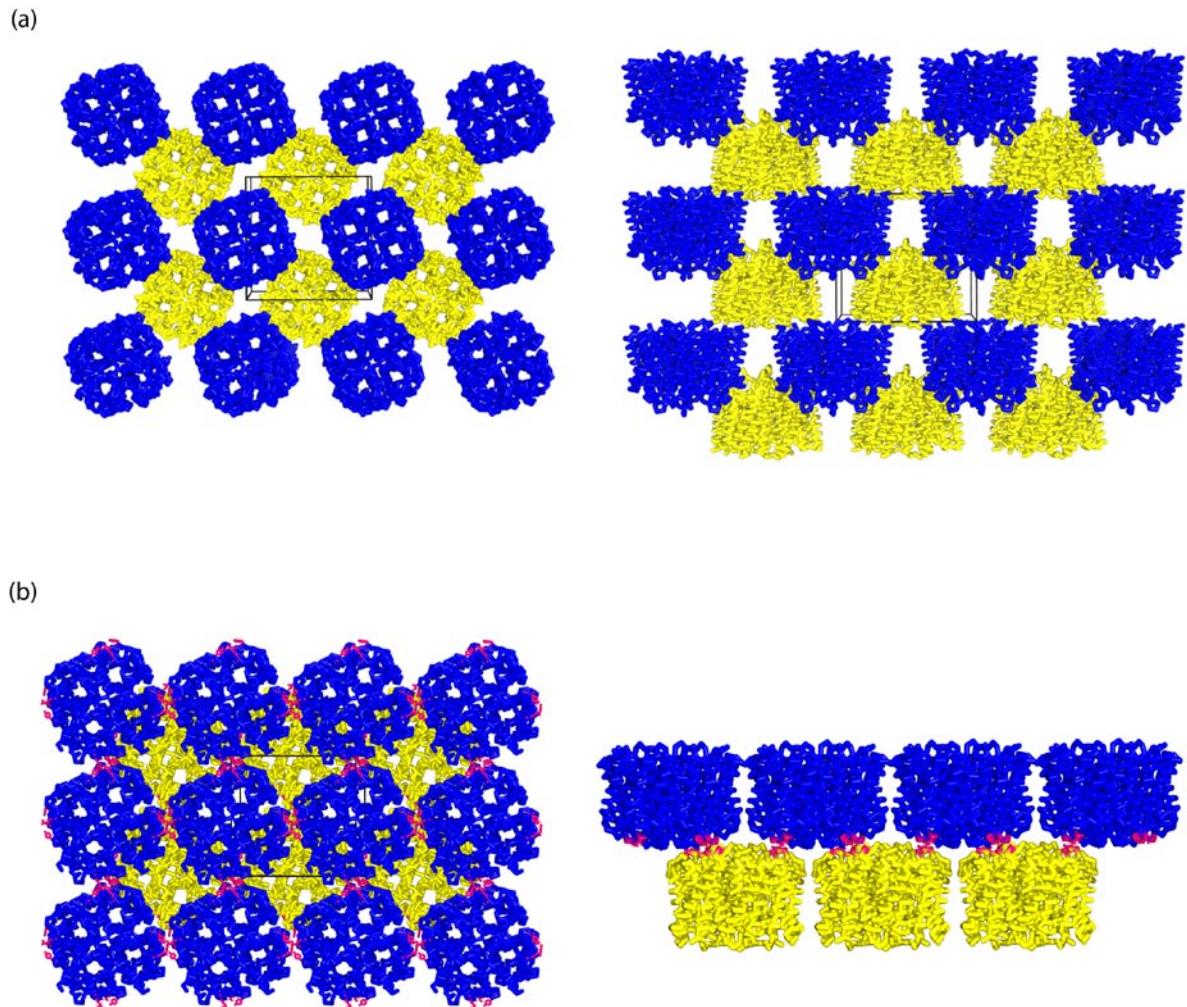


Figure 3-7

Crystal lattice packing of (a) hAQP4 (space group= $P4_21_2$, $a=82.1\text{ \AA}$, $c=76.4\text{ \AA}$) and (b) rAQP4 (space group= $P4_21_2$, $a=69\text{ \AA}$). Left: viewing the packing down the four-fold axis. Right: viewing the packing from the side. Tetramers that are not in the same horizontal plane are in different colors. Unit cell is shown in black. (b) Left: R108, G157, W231, I239 and Y250 are highlighted in pink, showing the interaction of the tetramers in the same plane. (b) Right: P139 and V142 from the rAQP4 3_{10} helix are highlighted in pink, showing that they are involved in the formation of the rAQP4 two-dimensional two-layered crystal (16).

M1 **A Loop**

hAQP4 **Met1** 10 **Met23** 30 40 50 60

```

hAQP4 . . . MSDRPTARRRWGKCGPLCTRENIMVAFKGVWTOAFWKAVTAEFLAMLI FVLLSLGSTIINWGGTEK . . .
rAQP4 . . . MSDGAAARRRWGKCGPPCSRESIMVAFKGVWTOAFWKAVTAEFLAMLI FVLLSVGSTIINWGGSEN . . .
bAQP0 . . . . . MWELRSASFWRAICAEFFASLFYVF FGLGASLRWAPG . . .
sAQP0 . . . . . MWELRSASFWRAIFAEFFATLFYVF FGLGASLRWAPG . . .
bAQP1 . . . . . MASEFKKKLFWRAVVAEFLAMILFIF ISISGALGFHYPIKSNQTT . . .
hAQP1 . . . . . MASEFKKKLFWRAVVAEFLATTLFVFI SSGALGFKYPVGNQT . . .
hAQP5 . . . . . MKKEVCSVAFLKAVFAEFLATLIFVF FGLGSAWKWPSA . . .
ScPIP MSKEVSEEAQAHQHGDYVDP PPAFFDLGELKLW SFWRAAIAEFLATL LFLYITVATVI GHSKETVVCG . . .
aqpM . . . . . MVSLTKRCIAEFGTFFLV FFGCAA IITLMIASGGTAPNPFN . . .
pFAQP . . . . . MHMLFYKSYVREFICEFLGT FVLMFLGEGATANFHTTG . . .
GlpF . . . . . MSQTS LKGCIAEFLGTGLLIF FFGVCVAALKVAG . . .
AqpZ . . . . . NDMFKRLAECFGTFWL VFGGCSAV LAAGFPPELG . . .

```

M2 **M3** **B Loop** **M4**

hAQP4 70 80 90 100 110 **Ser111** 120 130

```

hAQP4 . . . . . PLPVDMVLISL C FGLSIATM VQCFGH ISGCHINPAVTVM VCTRKISIAKSVFYTAAC LGAIGAGILY
rAQP4 . . . . . PLPVDMVLISL C FGLSIATM VQCFCH ISGCHINPAVTVM VCTRKISIAKSVFYITAAC LGAIGAGILY
bAQP0 . . . . . PLHV LQVALAF GLALATLVQAVGH ISGAHVNPAVTF AFV L VGSQMSLLRAICYVVAQL LGAVAGAVLY
sAQP0 . . . . . PLHV LQVALAF GLALATLVQAVGH ISGAHVNPAVTF AFV L VGSQMSLLRAICYVVAQL LGAVAGAVLY
bAQP1 . . . . . GAVQDNV KVS LAF GLSIAATLAQSVGH ISGAHLNPAVTL GLLSCQISVLRAIMYIIAOCVGAIVATAILS
hAQP1 . . . . . AVQDNV KVS LAF GLSIAATLAQSVGH ISGAHLNPAVTL GLLSCQISIFRALMYIIAOCVGAIVATAILS
hAQP5 . . . . . LPTILQIALAF GLAIGTLA QALGPVSGCHINPAITLAL LVGNQISLLRAFFYVAAQL VGAIVAGAVILY
ScPIP . . . . . SVGLLGIAWAF GGMIFVLVYCTAG ISGCHINPAVTFG LFLARKVSLLRALVYMAIAC LGAICGVGLVK
aqpM . . . . . IGIGLLGGLGDWVA IGLAF GF AIAAS IYALGN ISGCHINPAVTI GLWSVKKFPGRDVVPYIIAQL LGAAFSFI FL
pFAQP . . . . . LSGDWYKLC LGWGLAVFFG LILVSAKLSGAHLNPAVSI GLSSINKFDLKKIPVYFVACQL LGAFVGTSTVY
GlpF . . . . . ASFGQWEISVIW GLGVAMAIYLTAGVSGAHLNPAVTIAL WLFACFDKRVKIPVYFVACVAGFCAALVY
AqpZ . . . . . IGFAGV LAF GLTVLTMFAVGH ISGCHFNPAVTI GLWAGGRFPAK EVVGVYVIAOV VGGIVAAALLY

```

C Loop **M5** **D Loop**

hAQP4 140 150 160 170 **Ser180**

```

hAQP4 LVTP . . . . . PSVVG . . . . . GLGVTMVHGNLTAGHGLL VELITIFQLVFTIFAS CDSKR TDVTGSI . . .
rAQP4 LVTP . . . . . PSVVG . . . . . GLGVTMVHGNLTAGHGLL VELITIFQLVFTIFAS CDSKR TDVTGSI . . .
bAQP0 SVTP . . . . . PAVRG . . . . . NLA LNTLHPGVS VQGATI VEIFLTLQVFLC IFAIYDER RNRGLGVS . . .
sAQP0 SVTP . . . . . PAVRG . . . . . NLA LNTLHPGVS VQGATI VEIFLTLQVFLC IFAIYDER RNRGLGVS . . .
bAQP1 GITS . . . . . SLPDN . . . . . SLG LNALAPGVNSGQGLG IEIIGTLQVLVCLVAT TDRRRRDLGGSS . . .
hAQP1 GITS . . . . . SLTGN . . . . . SLG RNLADGVNSGQGLG IEIIGTLQVLVCLVAT TDRRRRDLGGSS . . .
hAQP5 GVAP . . . . . LNARG . . . . . NLA VNALNNTTQQQAMV VELITIFQLALC IFASTDSRRTSPVGS . . .
ScPIP AFMKG . . . . . PYNQF . . . . . GGGANSVALGYNKGTALG AEIIGTFVLVYTVFSA TDPKRSARDSHVPI L
aqpM QCAC . . . . . ITAATIG . . . . . GLGATAPFPY IGYWQAML AETVGTFLLMIT IMGIADVERAPKFPAG . . .
pFAQP GLYHGFISNSKIP . . . . . QFAWETSRNPS ISLTGAF NELILITGILLLV ILVVDEN ICGKPHIL . . . K
GlpF GLYLNKTFDFEQTHHIVRGSVESVDLAGT STYPNH IINFVQAF VEMVITAILMGLILALTD D . . . GNGVPRPAG . . . L
AqpZ LIASNGTGFDAAS . . . . . GFA SN GYGEHS PGGYSML SALVBLVLSAG FLV ITHGATDKFPAG . . . F

```

M6 **M7** **E Loop** **M8**

hAQP4 190 200 210 220 230 240 250

```

hAQP4 . . . LAIGFSVAIGH LFAINY TGAS MNPARSFG PAVI . . . . . MGNWE . . . . . NHW IYVVGPI IGA VLAGGLY EYV
rAQP4 . . . LAIGFSVAIGH LFAINY TGAS MNPARSFG PAVI . . . . . MGNWE . . . . . NHW IYVVGPI IGA VLAGALY EYV
bAQP0 . . . ALAVGFSLT LGHLFGMY YTGAG MNPARSFA PAI L . . . . . TRNF T . . . . . NHW VYVVGPI IGA GLSLLYDFL
sAQP0 . . . ALAVGFSLT LGHLFGMY YTGAG MNPARSFA PAI L . . . . . TRNF T . . . . . NHW VYVVGPI IGA GLSLLYDFL
bAQP1 . . . PLAIGFSVALGH LL AIDY TGCG INPARSFG SVI . . . . . THNF S . . . . . DHW IFWVGPF IGA ALAVLIYDFI
hAQP1 . . . PLAIGLSVALGH LL AIDY TGCG INPARSFG SVI . . . . . THNF S . . . . . NHW IFWVGPF IGA ALAVLIYDFI
hAQP5 . . . ALSIGLSVT LGHLVGIYFTGCS MNPARSFG PAVV . . . . . MNRFS P . . . . . AHW VVVGPI IGA VLAAILYFY L
ScPIP AP LPIGFVAV MVHLATIPITGTG INPARSFG PAVI . . . . . PNSN K V W . . . . . DDQW IFWVGPF IGA VAAAYHQYV
aqpM . . . I I IGLTVA GIIITIGNI TGSS LNPARTFG PYLN . . . . . DMVF AGTNLWNYFP IYV IGVVGA VLAALTYOYL
pFAQP LS SVVGLI IAVIG IITFGGN TGFA LNPSRD LGS RFLSLIA . YGKDTFT . . . . . KDNPFYFVPLVPL CVGVS VFCQFYDKV
GlpF AP LLIGLIIAVIG ASMGPL TGFA MNPARDFG KVF FAWLAGWGNVAF TGGRDIPYFLVPL FGP IGVAVI GFAYRKL
AqpZ AP IAIGLALTLIHLISIPVNTS VNPARS TA V A I F . . . . . OGGWAL . . . . . EQLW EFWVYV I VGG IIGGLIYRTL

```

hAQP4 260 270 280 290 300 310 320

```

hAQP4 FCPDVEFKRRFKEAFS KAAQQT KGSYMEVEDN . . . RSQVETD D LILKPGVVHVIDVDRGEEKKGDQSGEVLSSV
rAQP4 FCPDVELKRR LKEAFS KAAQQT KGSYMEVEDN . . . RSQVETD LILKPGVVHVIDIDRGDEKKKGDSSGEVLSSV
bAQP0 LFPR . . . . . LKSVSERLSILKGRSPSENG . . . QPEVTGEPVELKTQAL . . . . .
sAQP0 LFPR . . . . . LKSVSERLSILKGRSPSENG . . . QPEVTGEPVELKTQAL . . . . .
bAQP1 LAPR . . . . . SSDLTD R V K V W T S G Q V E E Y D L . . . DADDINSR VEMKPK . . . . .
hAQP1 LAPR . . . . . SSDLTD R V K V W T S G Q V E E Y D L . . . DADDINSR VEMKPK . . . . .
hAQP5 LFPN . . . . . SLSLSERVAI I K G T Y E P D E D W E E Q R E E R K K T M E L T T R . . . . .
ScPIP LRAA . . . . . AIKALGSRFNPTN . . . . .
aqpM TSE . . . . .
pFAQP I . . . . . CPLVDLANNEKDGVDL . . . . .
GlpF IGRH . . . . . LPCDICVVEEKETTT PSEQKASL . . . . .
AqpZ LEKRD . . . . .

```

Figure 3-8

Protein sequence alignment of all the aquaporin structures solved to date: human and rat AQP4, bovine and sheep AQP0, bovine and human AQP1, human AQP5, spinach AQP SoPIP2;1, archeal AQP AqpM, *Plasmodium falciparum* PfAQP, *E. coli* GlpF and AqpZ. Transmembrane helices and loop regions are defined. Initiating methionine for M1 and M23 isoforms of AQP4 are labeled. Ser111 and Ser180 are potential phosphorylation sites that affect gating for AQP4. The C-terminal SSV of AQP4 is the ligand that recruits α -syntrophin.

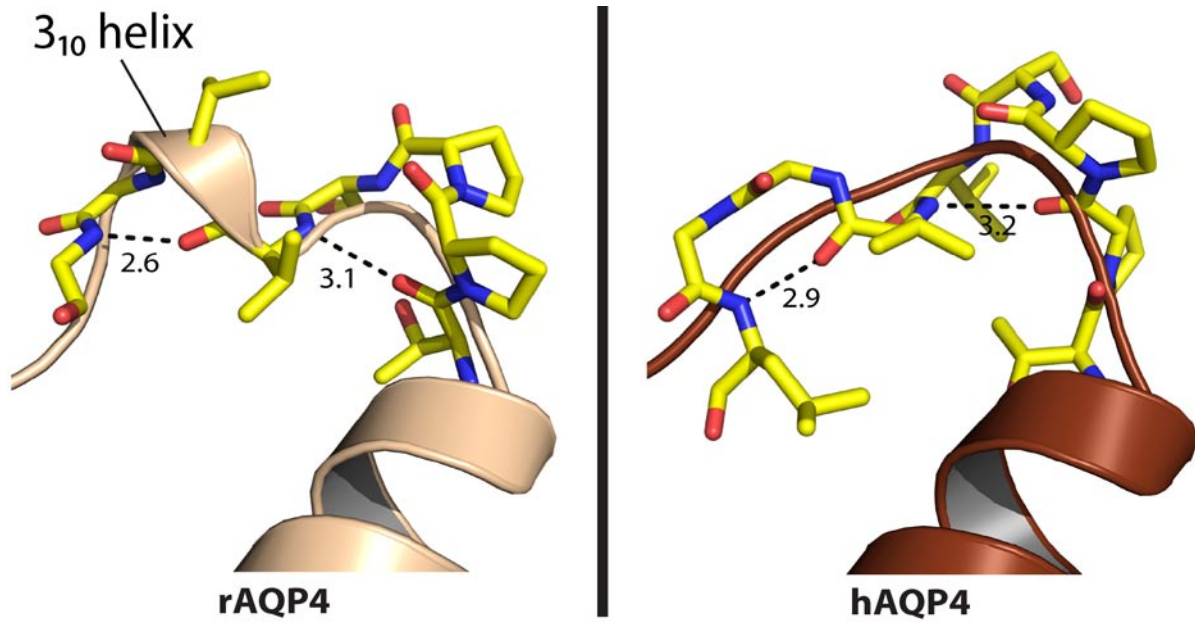


Figure 3-9

Comparison of the C loop between rat AQP4 (light brown) and human AQP4 (dark brown).

Notice the 3₁₀ helix is missing in human AQP4.

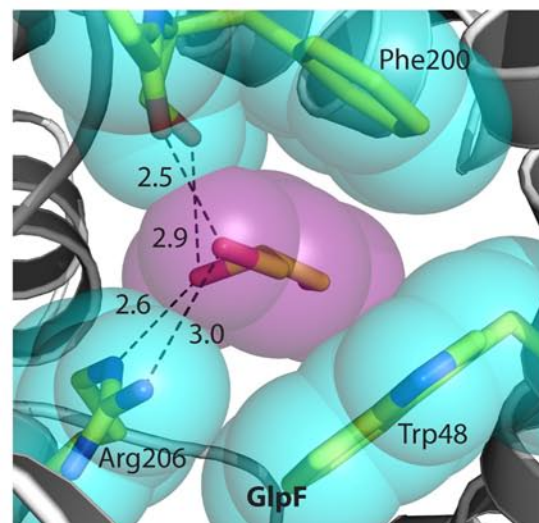
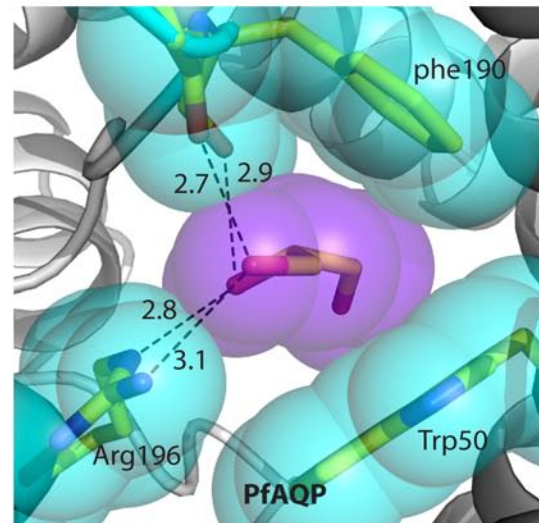
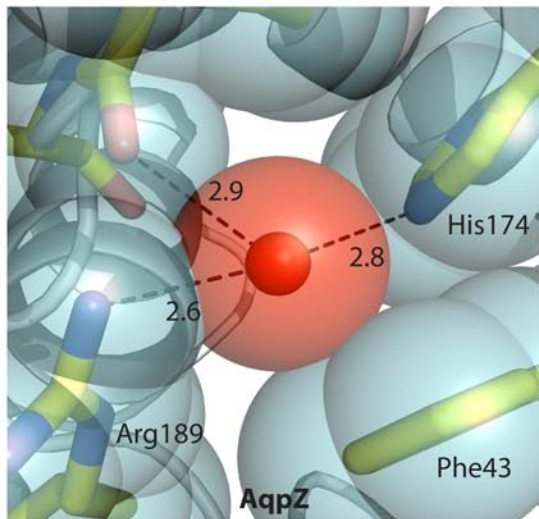
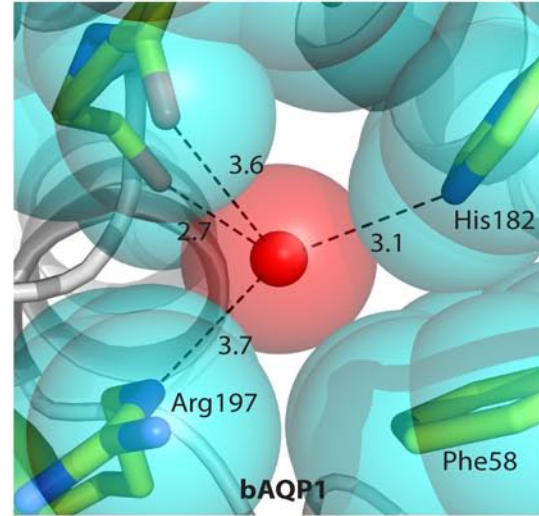
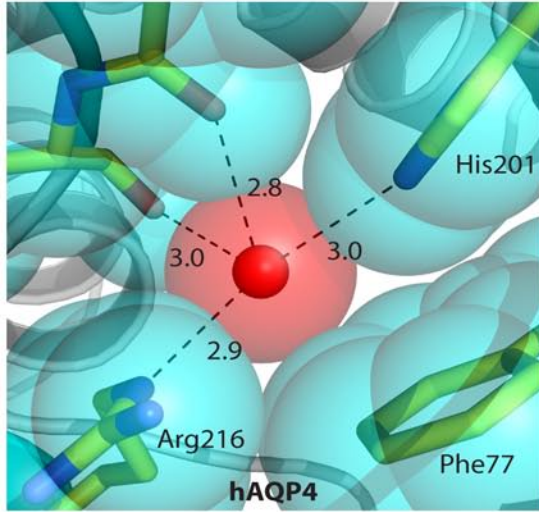


Figure 3-10

Sphere representations of the van der Waals contact of the selectivity filter (SF) residues of hAQP4, bAQP1, AqpZ, PfAQP and GlpF, viewing from the extracellular side down the conducting pore.

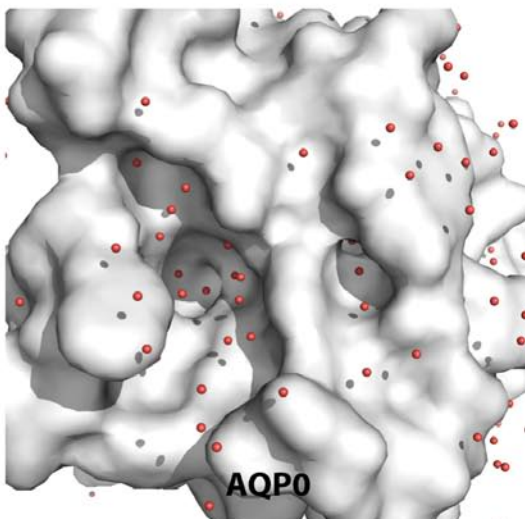
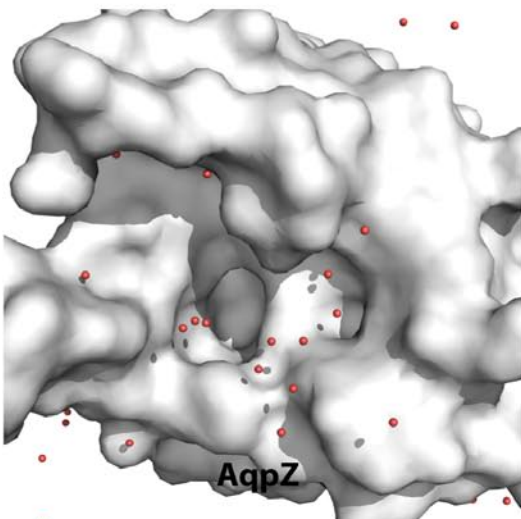
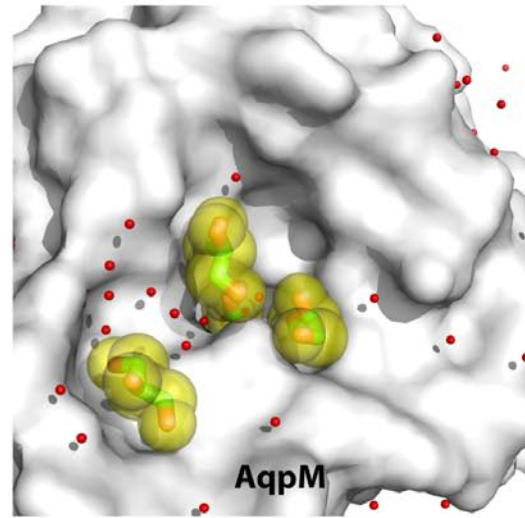
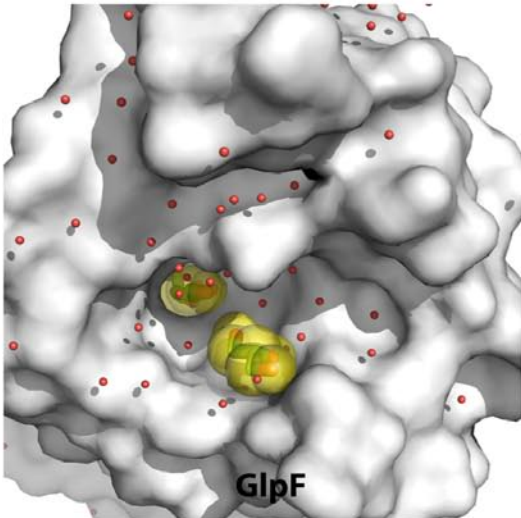
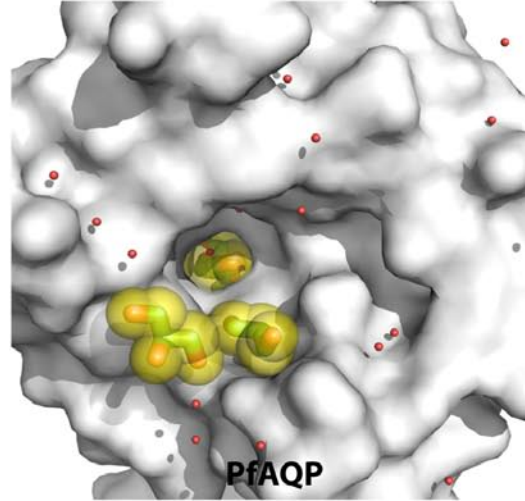
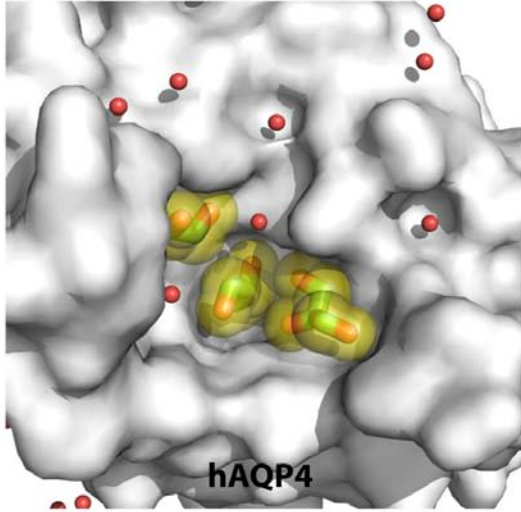


Figure 3-11

Comparison of the surfaces of the extracellular vestibule of human AQP4, PfAQP, *E. coli* GlpF, *E. coli* AqpZ, Archaeal AqpM, and bovine AQP0. All are in the same orientation. Glycerol molecules are shown as green sticks with yellow surfaces. Water molecules are shown as red spheres with gray shadows.

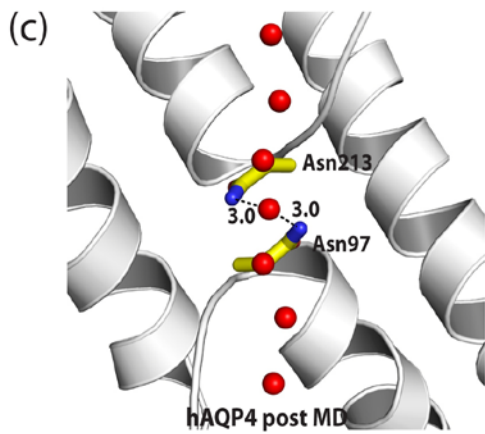
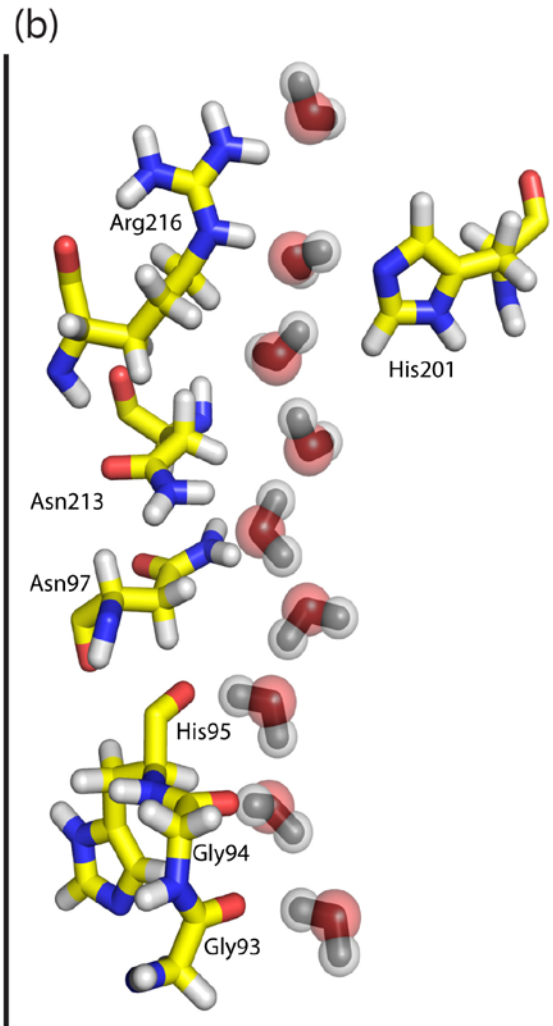
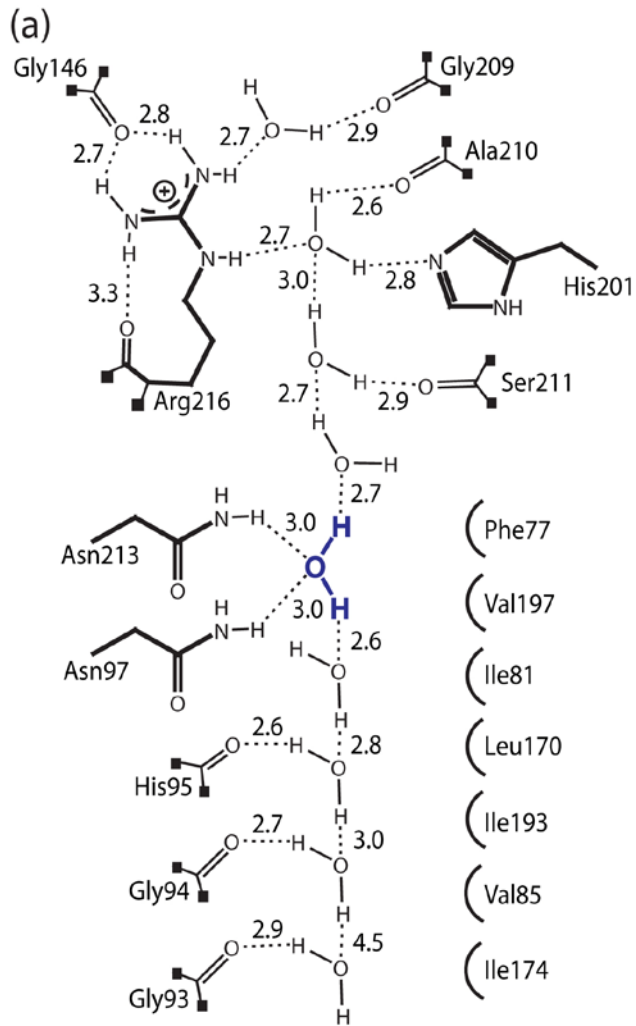


Figure 3-12

AQP4: post MD simulation from the third experiment (see text under Discussion: The variant role of the dual NPA motifs in proton exclusion.) (a) Schematic representation of the hydrogen bonding network through the channels. The distances are between heavy-atom to heavy-atom. (b) Stick representation of the conducting channel with hydrogen atoms from the simulation. (c) Another view of the NPA motifs and the central water.

Table 3-1. Data collection and refinement statistics

Human AQP4	
Data collection	
Space group	P4 ₂ 1 ₂
Cell dimensions	
<i>a</i> , <i>c</i> (Å)	82.1, 76.4
Resolution (Å)	1.8 (1.86-1.8)
R _{sym}	9.0 (79.6) ^a
<i>I</i> / σI	20 (3.3)
Completeness (%)	99.9 (100)
Redundancy	11.9 (12.0)
Refinement	
Resolution (Å)	36-1.8
No. unique reflections	23583
R _{work} / R _{free} ^b	16.4/17.4
No. atoms	
Protein	1667
Glycerol	30
β -OG	20
Water	61
Average <i>B</i> -factors, Å ²	
Protein	25
Glycerol	44
β -OG	59
Water	47
R.m.s. deviations	
Bond lengths (Å)	0.08
Bond Angles (°)	1.141

^aValue in parentheses are for the highest resolution shell.

^bR_{free} is calculated from 5% of reflections chosen randomly.

Table 3-2a: Summary of TLS refinement trials

<u># TLS Groups</u>	<u>Rfact</u>	<u>Rfree</u>	<u>FOM</u>	<u>LLG</u>	<u>rmsBOND</u>	<u>rmsANGLE</u>	<u>rmsCHIRAL</u>
1	0.181	0.194	0.898	118283.2	0.01	1.217	0.081
2	0.18	0.191	0.9	117942.1	0.01	1.216	0.082
3	0.178	0.189	0.902	117673.9	0.01	1.209	0.082
4	0.178	0.189	0.904	117552.8	0.01	1.201	0.081
5	0.178	0.19	0.903	117593	0.01	1.206	0.081
6	0.177	0.189	0.905	117338	0.01	1.189	0.08
7	0.177	0.188	0.905	117267.7	0.01	1.183	0.079
8	0.176	0.187	0.906	117085	0.01	1.171	0.079
9	0.176	0.187	0.906	117078	0.01	1.166	0.078
10	0.175	0.186	0.906	116935.9	0.01	1.17	0.078
11	0.175	0.188	0.905	116933	0.01	1.173	0.078
12	0.174	0.188	0.906	116838.4	0.01	1.174	0.079
13	0.173	0.186	0.909	116610.8	0.01	1.172	0.078
14	0.173	0.185	0.909	116547	0.01	1.174	0.078
15	0.173	0.186	0.91	116453.9	0.011	1.182	0.079
16	0.173	0.186	0.91	116440.4	0.01	1.178	0.079
17	0.173	0.186	0.91	116440.4	0.01	1.178	0.079
18	0.172	0.184	0.911	116299.9	0.011	1.177	0.079
19	0.172	0.184	0.911	116269.7	0.011	1.176	0.078
20	0.172	0.184	0.91	116285.3	0.011	1.17	0.078
19 + H ride	0.164	0.174	0.922	114509	0.008	1.141	0.068
20 + H ride	0.166	0.174	0.92	114859.7	0.008	1.149	0.068

Different numbers of TLS groups were tried, and the best result was with 19 TLS groups.

Statistics were taken from the Refmac5 refinement output log.

The TLS server (<http://skuld.bmsc.washington.edu/~tlsmd/>) was used to generate the different numbers of TLS groups.

For comparison purpose, riding hydrogen was applied to refinement with 19 and 20 TLS groups.

Table 3-2b: The 19 TLS groups used for refinement

<u>Group #</u>	<u>Residue Range</u>
1	Q32 to K36
2	A37 to I47
3	F48 to W59
4	G60 to P65
5	L66 to V71
6	L72 to S92
7	G93 to V102
8	A103 to S115
9	V116 to I132
10	L133 to T148
11	M149 to L154
12	T155 to V171
13	F172 to S177
14	C178 to V185
15	T186 to V197
16	A198 to A210
17	S211 to I232
18	Y233 to Y248
19	E249 to P254

REFERENCES

1. Rojek A, Praetorius J, Frokiaer J, Nielsen S, Fenton RA. A Current View of the Mammalian Aquaglyceroporins. *Annual Review of Physiology* 2008 70(1): 301-327.
2. King LS, Kozono D, Agre P. From structure to disease: the evolving tale of aquaporin biology. *Nat Rev Mol Cell Biol* 2004 5(9): 687-698.
3. Itoh T, Rai T, Kuwahara M, Ko SBH, Uchida S, Sasaki S, Ishibashi K. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. *Biochemical and Biophysical Research Communications* 2005 330(3): 832-838.
4. Badaut J, Brunet J-F, Regli L. Aquaporins in the brain: from aqueduct to “multi-duct”. *Metabolic Brain Disease* 2007 22(3): 251-263.
5. Manley GT, Fujimura M, Ma T, Noshita N, Filiz F, Bollen AW, Chan P, Verkman AS. Aquaporin-4 deletion in mice reduces brain edema after acute water intoxication and ischemic stroke. *Nat Med* 2000 6(2): 159-163.
6. Badaut J, Lasbennes F, Magistretti PJ, Regli L. Aquaporins in Brain[colon] Distribution, Physiology, and Pathophysiology. *J Cereb Blood Flow Metab* 2002 22(4): 367-378.
7. Detmers FJM, de Groot BL, Muller EM, Hinton A, Konings IBM, Sze M, Flitsch SL, Grubmuller H, Deen PMT. Quaternary Ammonium Compounds as Water Channel Blockers: SPECIFICITY, POTENCY, AND SITE OF ACTION. *J Biol Chem* 2006 May 19, 2006; 281(20): 14207-14214.
8. Huber VJ, Tsujita M, Yamazaki M, Sakimura K, Nakada T. Identification of arylsulfonamides as Aquaporin 4 inhibitors. *Bioorg Med Chem Lett* 2007 Mar 1; 17(5): 1270-1273.

9. Huber VJ, Tsujita M, Kwee IL, Nakada T. Inhibition of Aquaporin 4 by antiepileptic drugs. *Bioorg Med Chem* 2009 17(1): 418-424.
10. Huber VJ, Tsujita M, Nakada T. Identification of aquaporin 4 inhibitors using in vitro and in silico methods. *Bioorg Med Chem* 2009 Jan 1; 17(1): 411-417.
11. Yang B, Zhang H, Verkman AS. Lack of aquaporin-4 water transport inhibition by antiepileptics and arylsulfonamides. *Bioorganic & Medicinal Chemistry* 2008 16(15): 7489-7493.
12. Wingerchuk DM, Lennon VA, Lucchinetti CF, Pittock SJ, Weinshenker BG. The spectrum of neuromyelitis optica. *The Lancet Neurology* 2007 6(9): 805-815.
13. Rash JE, Yasumura T, Hudson CS, Agre P, Nielsen S. Direct immunogold labeling of aquaporin-4 in square arrays of astrocyte and ependymocyte plasma membranes in rat brain and spinal cord. *Proc Natl Acad Sci U S A* 1998 Sep 29; 95(20): 11981-11986.
14. Lu M, Lee MD, Smith BL, Jung JS, Agre P, Verdijk MA, Merckx G, Rijss JP, Deen PM. The human AQP4 gene: definition of the locus encoding two water channel polypeptides in brain. *Proc Natl Acad Sci U S A* 1996 Oct 1; 93(20): 10908-10912.
15. Furman CS, Gorelick-Feldman DA, Davidson KG, Yasumura T, Neely JD, Agre P, Rash JE. Aquaporin-4 square array assembly: opposing actions of M1 and M23 isoforms. *Proc Natl Acad Sci USA* 2003 100(13609 - 13614).
16. Hiroaki Y, Tani K, Kamegawa A, Gyobu N, Nishikawa K, Suzuki H, Walz T, Sasaki S, Mitsuoka K, Kimura K, Mizoguchi A, Fujiyoshi Y. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol* 2006 Jan 27; 355(4): 628-639.

17. Michele DE, Campbell KP. Dystrophin-Glycoprotein Complex: Post-translational Processing and Dystroglycan Function. *J Biol Chem* 2003 April 25, 2003; 278(18): 15457-15460.
18. Amiry-Moghaddam M, Frydenlund DS, Ottersen OP. Anchoring of aquaporin-4 in brain: Molecular mechanisms and implications for the physiology and pathophysiology of water transport. *Neuroscience* 2004 129(4): 997-1008.
19. Neely JD, Amiry-Moghaddam M, Ottersen OP, Froehner SC, Agre P, Adams ME. Syntrophin-dependent expression and localization of Aquaporin-4 water channel protein. *Proceedings of the National Academy of Sciences of the United States of America* 2001 November 20, 2001; 98(24): 14108-14113.
20. Hasegawa H, Ma T, Skach W, Matthay MA, Verkman AS. Molecular cloning of a mercurial-insensitive water channel expressed in selected water-transporting tissues. *J Biol Chem* 1994 February 25, 1994; 269(8): 5497-5500.
21. Savage DF, Stroud RM. Structural basis of aquaporin inhibition by mercury. *J Mol Biol* 2007 May 4; 368(3): 607-617.
22. Han Z, Wax MB, Patil RV. Regulation of aquaporin-4 water channels by phorbol ester-dependent protein phosphorylation. *J Biol Chem* 1998 Mar 13; 273(11): 6001-6004.
23. Zelenina M, Zelenin S, Bondar AA, Brismar H, Aperia A. Water permeability of aquaporin-4 is decreased by protein kinase C and dopamine. *Am J Physiol Renal Physiol* 2002 August 1, 2002; 283(2): F309-318.

24. Gunnarson E, Zelenina M, Axehult G, Song Y, Bondar A, Krieger P, Brismar H, Zelenin S, Aperia A. Identification of a molecular target for glutamate regulation of astrocyte water permeability. *Glia* 2008 56(6): 587-596.
25. Hedfalk K, Tornroth-Horsefield S, Nyblom M, Johanson U, Kjellbom P, Neutze R. Aquaporin gating. *Curr Opin Struct Biol* 2006 Aug; 16(4): 447-456.
26. Tornroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P. Structural mechanism of plant aquaporin gating. *Nature* 2006 Feb 9; 439(7077): 688-694.
27. Sui H, Han B-G, Lee JK, Walian P, Jap BK. Structural basis of water-specific transport through the AQP1 water channel. *Nature* 2001 414(6866): 872-878.
28. Harries WE, Akhavan D, Miercke LJ, Khademi S, Stroud RM. The channel architecture of aquaporin 0 at a 2.2-Å resolution. *Proc Natl Acad Sci U S A* 2004 Sep 28; 101(39): 14045-14050.
29. Horsefield R, Norden K, Fellert M, Backmark A, Tornroth-Horsefield S, Terwisscha van Scheltinga AC, Kvassman J, Kjellbom P, Johanson U, Neutze R. High-resolution x-ray structure of human aquaporin 5. *Proc Natl Acad Sci U S A* 2008 September 9, 2008; 105(36): 13327-13332.
30. Fu D, Libson A, Miercke LJW, Weitzman C, Nollert P, Krucinski J, Stroud RM. Structure of a Glycerol-Conducting Channel and the Basis for Its Selectivity. *Science* 2000 October 20, 2000; 290(5491): 481-486.
31. Newby ZER, O'Connell Iii J, Robles-Colmenares Y, Khademi S, Miercke LJ, Stroud RM. Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nat Struct Mol Biol* 2008 15(6): 619-625.

32. Beitz E, Wu B, Holm LM, Schultz JE, Zeuthen T. 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* 2006 Jan 10; 103(2): 269-274.
33. Savage DF, Egea PF, Robles-Colmenares Y, O'Connell JD, 3rd, Stroud RM. Architecture and selectivity in aquaporins: 2.5 a X-ray structure of aquaporin Z. *PLoS Biol* 2003 Dec; 1(3): E72.
34. Zhang H, Verkman AS. Evidence against Involvement of Aquaporin-4 in Cell-Cell Adhesion. *Journal of Molecular Biology* 2008 382(5): 1136-1143.
35. Graber D, Levy M, Kerr D, Wade W. Neuromyelitis optica pathogenesis and aquaporin 4. *Journal of Neuroinflammation* 2008 5(1): 22.
36. Lee JK, Kozono D, Remis J, Kitagawa Y, Agre P, Stroud RM. Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å. *Proc Natl Acad Sci U S A* 2005 Dec 27; 102(52): 18932-18937.
37. Tanimura Y, Hiroaki Y, Fujiyoshi Y. Acetazolamide reversibly inhibits water conduction by aquaporin-4. *J Struct Biol* 2008 Dec 10.
38. Tajkhorshid E, Nollert P, Jensen MO, Miercke LJW, O'Connell J, Stroud RM, Schulten K. Control of the Selectivity of the Aquaporin Water Channel Family by Global Orientational Tuning. *Science* 2002 April 19, 2002; 296(5567): 525-530.
39. Grotthuss CJTd. Mémoire sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann Chim (Paris)* 1806 58(54-74).
40. Grotthuss CJTd. Memoir on the decomposition of water and of the bodies that it holds in solution by means of galvanic electricity. 1805. *Biochim Biophys Acta* 2006 Aug; 1757(8): 871-875.

41. Ilan B, Tajkhorshid E, Schulten K, Voth GA. The mechanism of proton exclusion in aquaporin channels. *Proteins* 2004 May 1; 55(2): 223-228.
42. Kato M, Pislakov AV, Warshel A. The barrier for proton transport in aquaporins as a challenge for electrostatic models: the role of protein relaxation in mutational calculations. *Proteins* 2006 Sep 1; 64(4): 829-844.
43. Burykin A, Warshel A. On the origin of the electrostatic barrier for proton transport in aquaporin. *FEBS Letters* 2004 570(1-3): 41-46.
44. Burykin A, Warshel A. What really prevents proton transport through aquaporin? Charge self-energy versus proton wire proposals. *Biophys J* 2003 Dec; 85(6): 3696-3706.
45. Chen H, Ilan B, Wu Y, Zhu F, Schulten K, Voth GA. Charge delocalization in proton channels, I: the aquaporin channels and proton blockage. *Biophys J* 2007 Jan 1; 92(1): 46-60.
46. Chakrabarti N, Tajkhorshid E, Roux B, Pomès R. Molecular basis of proton blockage in aquaporins. *Structure* 2004 12(1): 65 - 74.
47. Hiroaki Y, Tani K, Kamegawa A, Gyobu N, Nishikawa K, Suzuki H, Walz T, Sasaki S, Mitsuoka K, Kimura K, Mizoguchi A, Fujiyoshi Y. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol* 2006 355(628 - 639).
48. Otwinowski Z, Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode *Methods in Enzymology* 1997 276(307-326).
49. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 2005 Apr; 61(Pt 4): 458-464.

50. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004 Dec; 60(Pt 12 Pt 1): 2126-2132.
51. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994 Sep 1; 50(Pt 5): 760-763.
52. Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. *J Appl Cryst* 2006 39(1): 109-111.
53. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D* 2006 62(4): 439-450.
54. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993 26(283-291).
55. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* 1992 Apr; 12(4): 345-364.
56. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 2003 Feb 15; 50(3): 437-450.
57. Lindahl E, Hess B, Van Der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* 2001 7(8): 306-317.
58. Pullman B (1981) *Intermolecular forces : proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry held in Jerusalem, Israel, April 13-16, 1981* (D. Reidel ;, Dordrecht, Holland ; Boston, U.S.A.).

59. Xu Z, Luo HH, Tieleman DP. Modifying the OPLS-AA force field to improve hydration free energies for several amino acid side chains using new atomic charges and an off-plane charge model for aromatic residues. *J Comput Chem* 2007 Feb; 28(3): 689-697.
60. Jorgensen WL, Maxwell DS, TiradoRives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* 1996 Nov 13; 118(45): 11225-11236.
61. Berger O, Edholm O, Jahnig F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* 1997 May; 72(5): 2002-2013.
62. Monticelli L, Tieleman P (2007).
63. Nose S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* 1984 52(2): 255-268.
64. Parrinello M, Rahman A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Journal of Applied Physics* 1981 52(12): 7182-7190.
65. Darden T, York D, Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* 1993 Jun 15; 98(12): 10089-10092.
66. Miyamoto S, Kollman PA. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of Computational Chemistry* 1992 Oct; 13(8): 952-962.

67. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 1997 Sep; 18(12): 1463-1472.
68. Smart OS, Goodfellow JM, Wallace BA. The pore dimensions of gramicidin A. *Biophys J* 1993 Dec; 65(6): 2455-2460.

Chapter 4

Methods for Improving X-ray Diffraction of Membrane Protein Crystals:

A case study of human aquaporin 4

Research conducted in collaboration with Ronald Yeh, Andrew Sandstrom, Rebecca A.

Robbins, Larry J.W. Miercke, and Robert M. Stroud

ABSTRACT

Full-length human Aquaporin 4 was heterologously expressed in *Pichia pastoris* and purified.

Initial crystals of the protein took 6 months to grow and diffracted only to 8 Å. The final

improved crystals took 1 week to grow and diffracted to a best resolution of 1.8 Å. The

detergent concentration of the concentrated sample, the flexible termini of AQP4, and the

cryoprotectant conditions were found to be critical for improving crystal growth and

diffraction.

INTRODUCTION

Out of the current total of ~58,000 structures in the protein data bank, only ~473 are of membrane protein structures (www.pdb.org). Besides the difficulty in heterologous expression of membrane proteins when natural sources are not available, another major obstacle is obtaining crystals that diffract to 3 Å or better—a limit of resolution required for structural biologists to answer pertinent biochemical and mechanistic questions of the macromolecule.

During crystallization, membrane proteins can typically form one of three types of crystals: 2D crystals, type I 3D crystals, and type II 3D crystals (1). 2D crystals are excellent candidates for electron diffraction, and some examples are the structures of bacteriorhodopsin (2), the *Torpedo* acetylcholine receptor (3), human aquaporin 0 (4), and rat aquaporin 4 (5). Type I 3D crystals are formed by reconstituting detergent-solubilized membrane protein into lipidic cubic phase (LCP) (6). The most recent and prominent example for the use of the LCP is the crystal structure of human β 2-adrenergic G protein-coupled receptor (7). Finally, type II 3D crystals are formed by detergent-solubilized membrane proteins. The type II 3D crystal is the most common type of membrane protein crystals, and they are grown using the same techniques as applied to soluble proteins. However, the presence of detergent molecules makes the polar contacts between protein molecules in the crystal lattice more difficult, often resulting in poor-diffracting crystals.

In this chapter, we describe human aquaporin 4 (AQP4) as a case study for growing high-diffracting type II membrane protein crystals. The AQP4 crystals started with a resolution of

8 Å, but upon optimization, a best resolution of 1.8 Å was obtained. We investigated three factors that are critical for getting good X-ray diffraction from these AQP4 crystals: the detergent concentration in the final sample, the protein topology, and the cryoprotectant condition for freezing the crystals for synchrotron diffraction experiments. We believe these factors can also be important for forming high quality type II 3D crystals of other proteins.

RESULTS AND DISCUSSION

Increased β -octylglucoside concentration of the final sample impeded crystallization

Size exclusion chromatography (SEC)-purified Full-length AQP4 was pooled (2.4 OD/mL) and concentrated 32 fold to 75 OD/mL on an Amicon 50,000 molecular weight cut off spin filter without any resuspension of the retente during concentration. 40mM of β -octylglucoside (OG) was maintained throughout the purification to ensure sufficient micelle species were present to contribute to the protein-detergent complex (**Figure 4-1**). OG has a measured micelle size of ~22,000, so in theory, no extra OG micelles should accumulate in the retente when concentrating the protein in a 50,000 molecular weight cut off spin filter. However, the Tetra Detector Analysis (TDA) of the final concentrated sample showed the OG micelle peak with a concentration of 280 mM (**Figure 4-2**). A 32 fold increase of protein concentration resulted in a 7 fold increase of detergent concentration. Upon setting up the protein sample in crystallization, the crystals took ~2 months to form and grew to a maximum size of 50 microns (**Figure 4-3**). Interestingly, another crystallization experiment was set up where the sample was diluted 4 fold with water to reduce the concentration of OG to 70 mM, and these crystals grew in 2 days to a maximum size of 100 microns (**Figure 4-4**). The drastic difference in the speed of the crystal growth shows that excess OG molecules above CMC cause phase separation (indicated by arrows in **Figure 4-3**) which then impedes crystal growth.

Crystals that took 2 months to grow did not diffract at all, and the crystals that grew in the OG-reduced condition diffracted to a best resolution of 8.6 Å (**Figure 4-5**).

Trypsin removed the flexible regions of the protein

By looking at the topology of human AQP4, we postulate that the N-terminal end before the first transmembrane helix and the C-terminal end after the last transmembrane helix may be flexible (**Figure 4-6**). This flexibility could contribute adversely to the packing of the crystal, resulting in poor diffraction. To test this hypothesis, we mixed the full-length protein with trypsin (protocol described in detail in chapter 3: Material and Methods) and let the reaction to go to completion (overnight at 4°C). The trypsinized product still maintained the biological tetrameric assembly on size exclusion chromatography (**Figure 4-7**), and MALDI-MS of the peak showed a discrete molecular weight, indicating no secondary cleavage product (**Figure 4-8**). Combining the information of the MALDI-MS, N-terminal sequencing, and the knowledge that trypsin cleaves after lysine or arginine, we can deduce the trypsinized product (**Figure 4-6**). The fact that the N and C termini were susceptible to trypsin cleavage shows that they were indeed flexible.

The trypsinized AQP4 was concentrated similarly as the full-length protein, but less OG was retained by concentration. A 12 fold concentration of protein resulted in a 1.5 fold increase in OG concentration (40 mM to 63 mM, **Figure 4-9**). This shows that the OG retention on the 50,000 Dalton cut off concentrator is protein-dependent. The full-length molecule probably contains domains that bind OG molecules and may other detergents, causing the increase of detergent concentration during protein concentration. Two crystal morphologies were observed in the same condition for the trypsinized AQP4 (**Figure 4-10**), but the pyramidal-shaped crystals diffracted much better than the rod-shaped (**Table 4-1**). A 96-condition additive screen using the Silver Bullet (Hampton Research) was then used, and condition A1

was found helpful to selectively promote the growth of the pyramidal-shaped crystals (**Figure 4-11**). The best condition was found mixing 1 uL of protein with 1 uL of condition A1 and then with 2.5 uL of well solution.

Cryoprotectant condition optimization further improved diffraction resolution

For both the full-length and trypsinized AQP4 crystals, the initial cryoprotectant condition used was the well solution plus 20% glycerol. This cryoprotectant condition worked well when the crystal was only briefly soaked in it (~5 seconds). When the crystals were soaked in this condition for more than 1 minute, the crystals started to crack and disintegrate. Doing a step-wise increase of glycerol or ethylene glycol concentration from 5% to 20% also did not help.

Cryoprotectants can generally be divided into two classes: oil-based or water-based (8).

Three well-known oil-based cryoprotectants are paratone-N, paraffin, and NVH oil. During the cryoprotectant optimization for the trypsinized AQP4 crystals, all three oils were tried with brief soaking (~5 sec), but the crystals were all dissolved immediately. Since it is known that paraffin can be mixed with paratone-N to reduce viscosity and increase visibility of the crystals in oil, we tried a solution of 50:50 paraffin:paratone-N oil mixture. Interestingly, this cryoprotectant keeps both the full-length and the trypsinized AQP4 crystals intact over a period of 30 minutes. Indeed, trypsinized AQP4 crystals diffracted even better when cryoprotected with a 3-second soak in 50:50 paraffin:paratone-N (**Table 4-1**). This cryoprotectant, however, did not improve diffraction of the full-length AQP4 crystals. There are also a few other crystal structures solved using 50:50 paraffin:paratone-N as

cryoprotectant: the *Bacillus subtilis* levansucrase (1.5 Å) (9), DHNA synthetase from *Geobacillus kaustophilus* (2.2 Å), and the carboxymuconolactone decarboxylase family member protein from *Thermus thermophilus* (1.9 Å) (10).

MATERIALS AND METHODS

Crystal Imaging

Crystal pictures were taken on a Leica MZ 16 Stereomicroscope with a Leica KL 1500 LCD fiber optic illuminator (Spectronic Analytical Instruments, United Kingdom) using the Canon LA-DC58E as the camera. UV fluorescence images of the crystals were taken on a Korima PRS-1000 microscope (Korima, Inc.).

Tetra Detector Array (TDA) Analyses

The Viscotek Model 302-050 Tetra Detector Array instrument (Viscotek Corporation, Texas) was connected in-line with a Superdex 200 10/300 GL column (GE Healthcare). For the analysis of the full-length AQP4, protein was concentrated 35 fold in an Amicon 50 KD MWCO spin concentrator, and 10 μ L at 53.6 mg/mL (75 OD/mL) of protein was injected. For the analysis of the trypsinized AQP4, protein was concentrated 12 fold in an Amicon 50 KD MWCO spin concentrator, and 30 μ L at 21.4 mg/mL (30 OD/mL) of protein was injected. A flow rate of 0.35 mL/min was used for both experiments. The molecular weight and concentration of the protein and detergent under the peaks were calculated from the viscosity, right-angle light scattering, and refractive index measurement (11-13).

X-ray data collection

Diffraction data were collected using a wavelength of 1.11 \AA at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, California, United States).

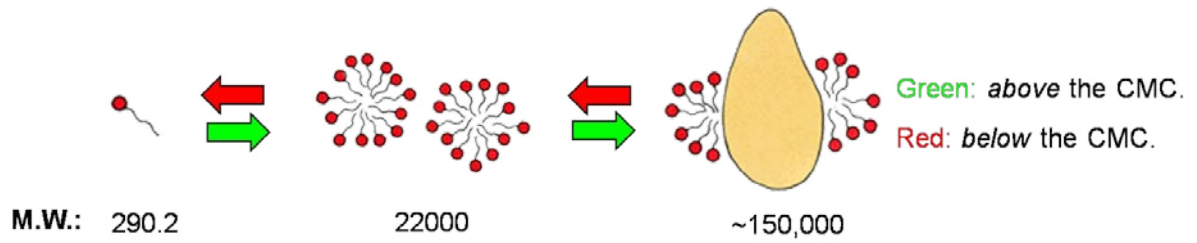


Figure 4-1

Schematic representation of the protein-detergent interaction. CMC stands for Critical Micelle Concentration. The detergent molecule represented is octyl- β -D-glucopyranoside (OG) with a monomer molecular weight of 290.2 g/mol. OG has a CMC of 20 mM, and at its CMC, it forms micelles of 22,000 Dalton in size. At the concentration of 40 mM OG, a complex of full-length AQP4 with OG micelles has a combined molecule weight of ~150,000 Dalton.

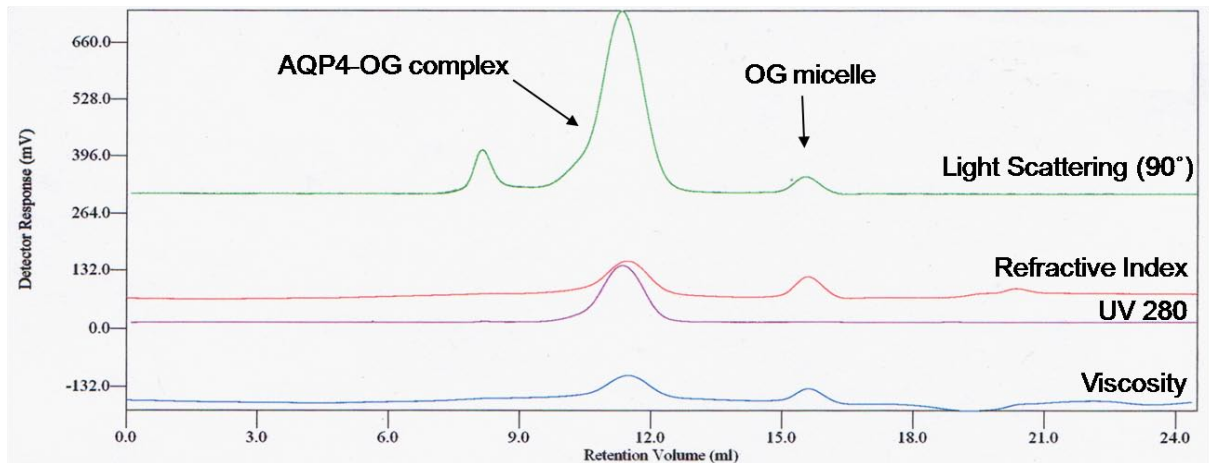
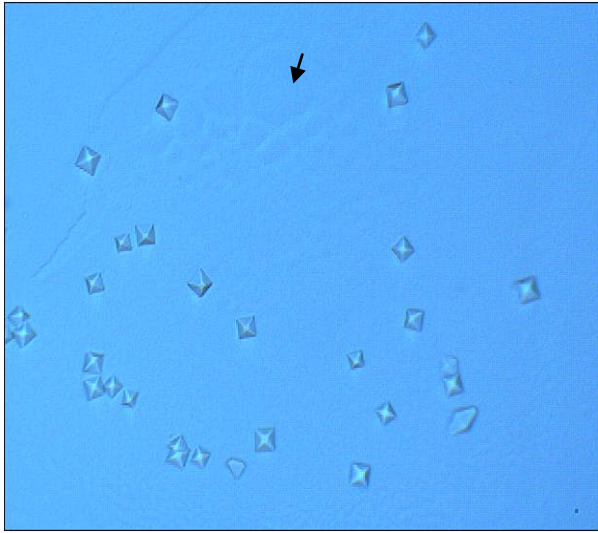


Figure 4-2

Tetra Detector Array Analysis (TDA) of the full-length AQP4 on size exclusion chromatography. OG is octyl- β -D-glucopyranoside. Peaks represent molecular species of different sizes. **Green** trace measures the light scattering of the sample at 90°. **Pink** trace measures the refractive index. **Purple** trace measures absorbance at 280 nm. **Blue** trace measures the viscosity of the sample.

With white light



With UV fluorescence

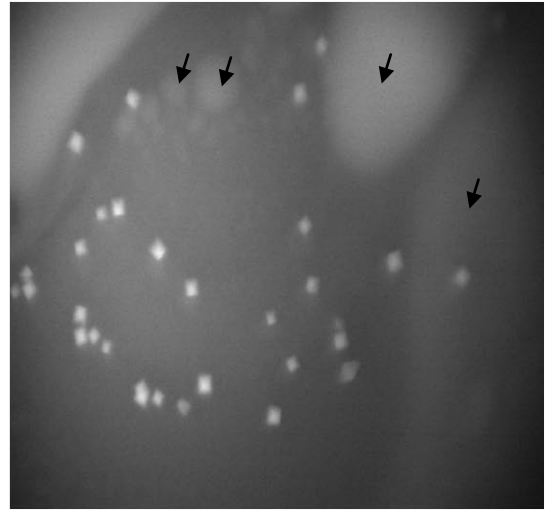


Figure 4-3

Crystals of the full-length AQP4 protein. The crystals are about 20 microns in size. The left and right images are from the same view of the tray. UV fluorescence microscope identifies the crystals as protein crystals. Regions of white 'blobs' are phase separation with concentrated protein in that area (indicated by arrows).

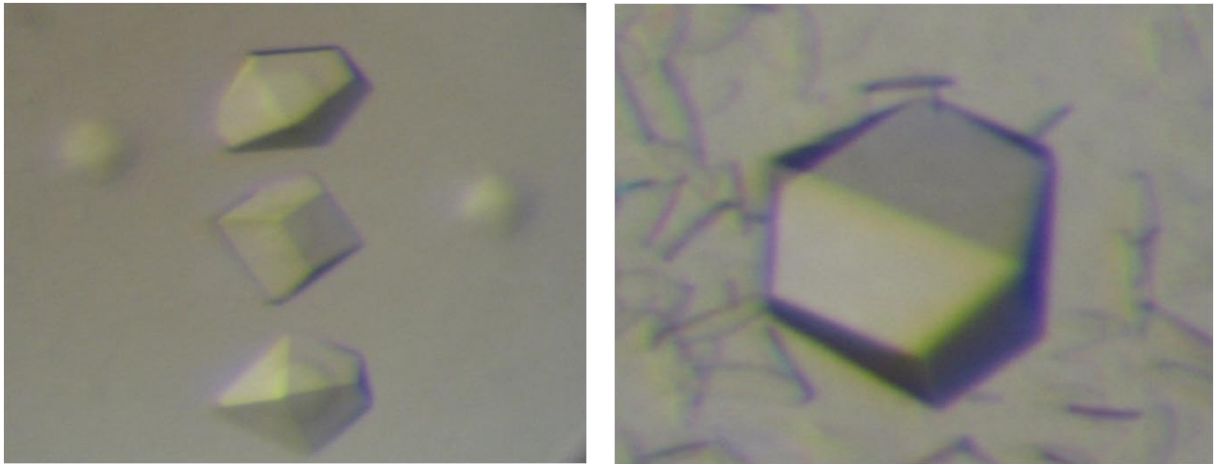


Figure 4-4

Crystals of full-length AQP4 after reducing the concentration of octyl- β -D-glucopyranoside for about 4 fold in the crystallization sample. Crystals are about 50-100 microns in size.

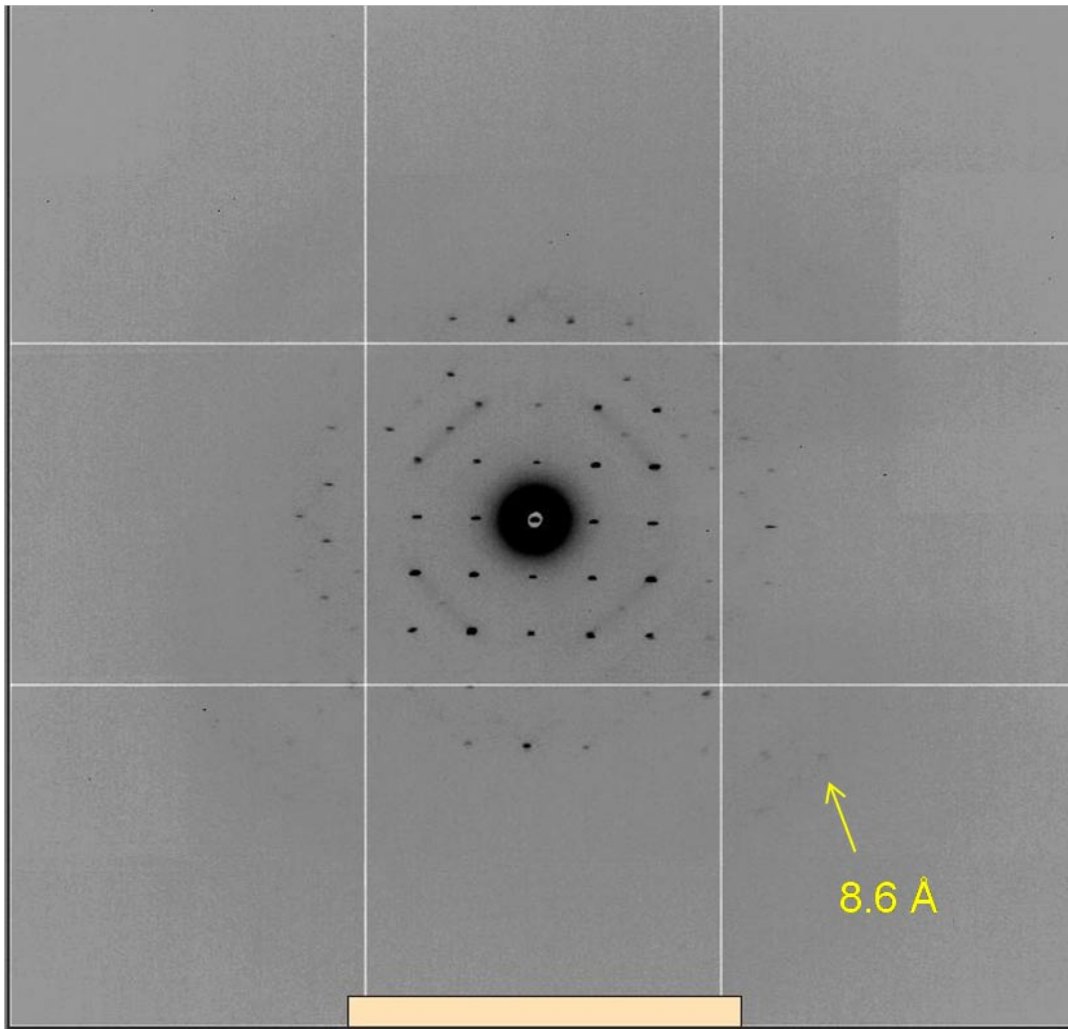


Figure 4-5

Initial diffraction patterns from the crystals in **Figure 4-4**. Best resolution spot is at 8.6 Å.

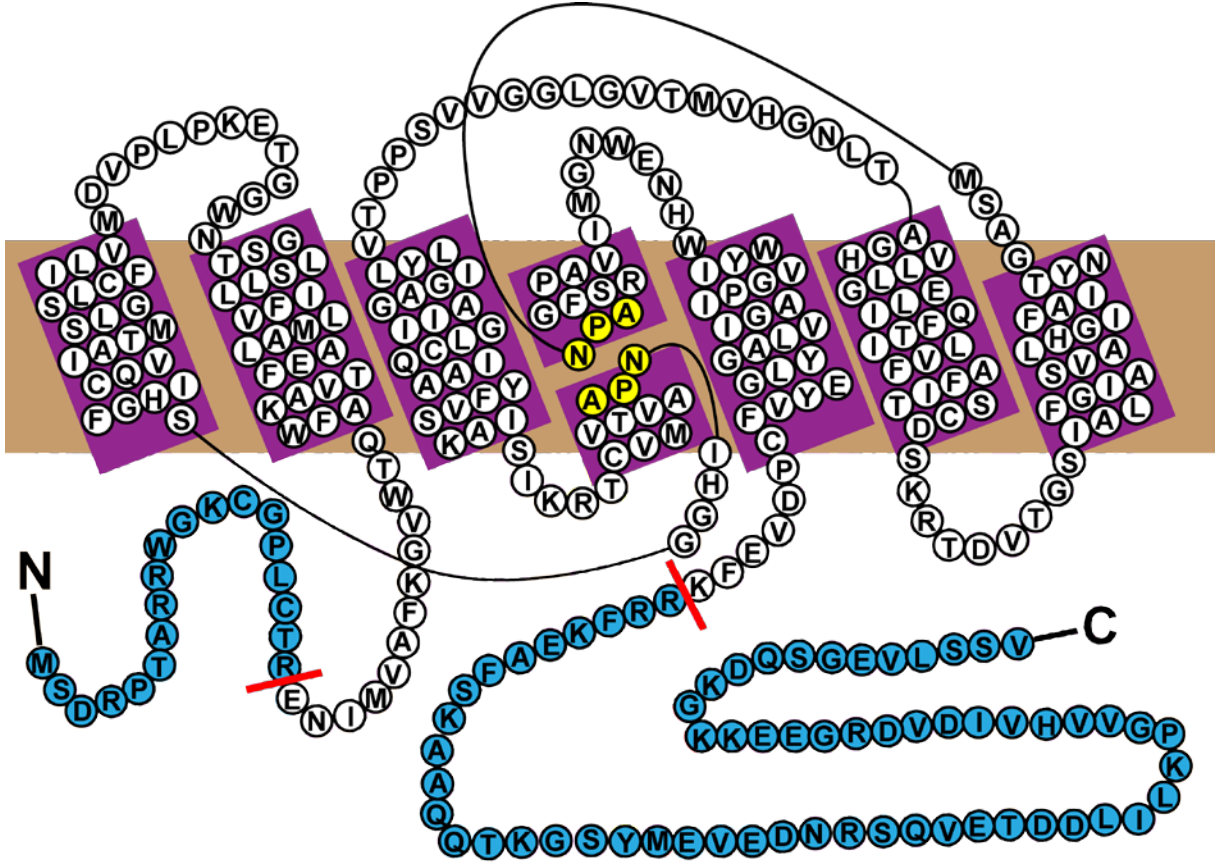


Figure 4-6

Schematic representation of the full-length human aquaporin 4 protein. Brown region represents cell membrane. Purple region represents transmembrane helices of the protein. The NPA motifs are conserved throughout most aquaporins and aquaglyceroporins. Red bars represent locations that trypsin cleaves in AQP4, removing the amino acids colored in blue.

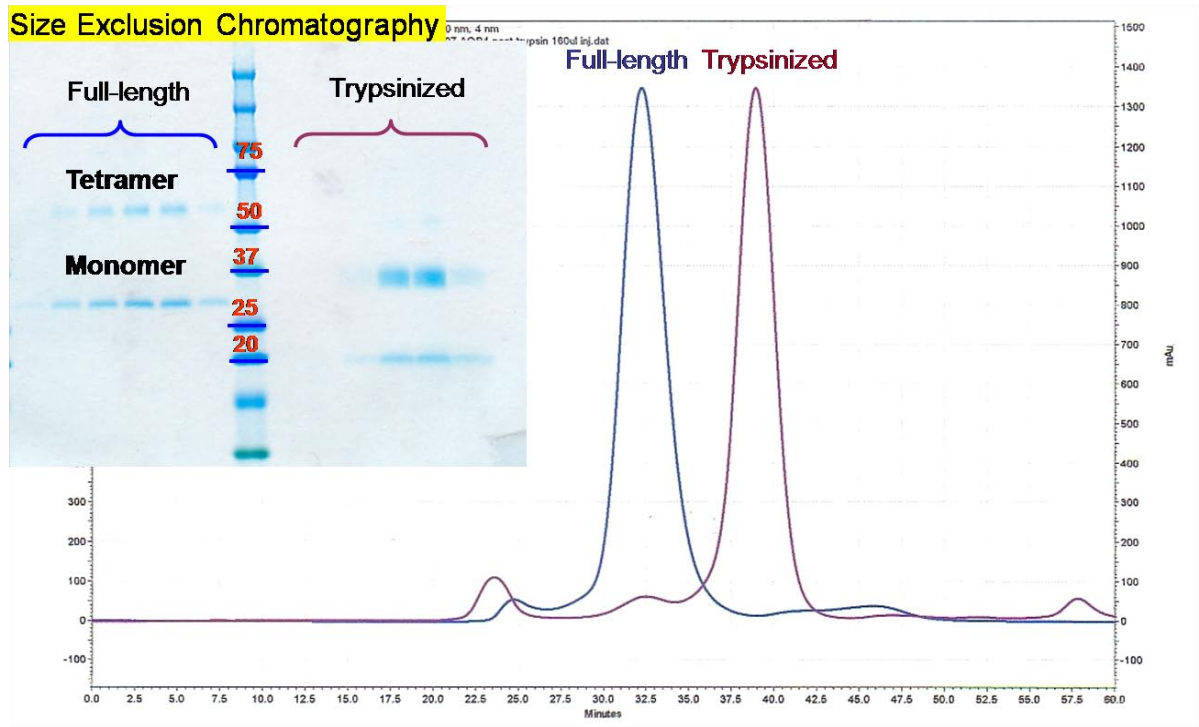


Figure 4-7

Superdex 200 size exclusion chromatography traces of the full-length and trypsinized AQP4.

Top left: Coomassie-stained SDS-PAGE of the full-length and trypsinized AQP4 across the peaks. The two bands on SDS-PAGE represent the SDS-denatured AQP4 monomers and the SDS-resistant AQP4 tetramers.

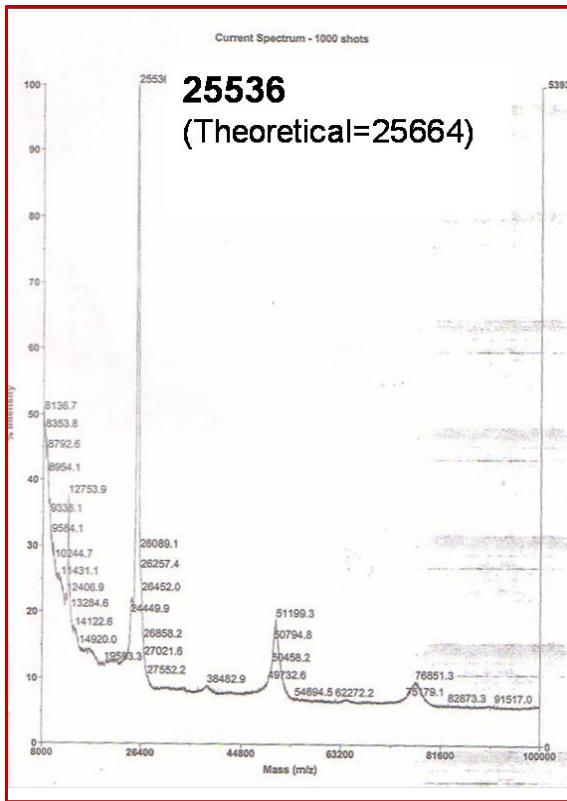


Figure 4-8

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) of the trypsinized AQP4. The theoretical molecular weight of the trypsinized product is calculated based on the N-terminal sequencing result and the consensus sequence for trypsin digest (cleavage after lysine or arginine) (**Figure 4-6**).

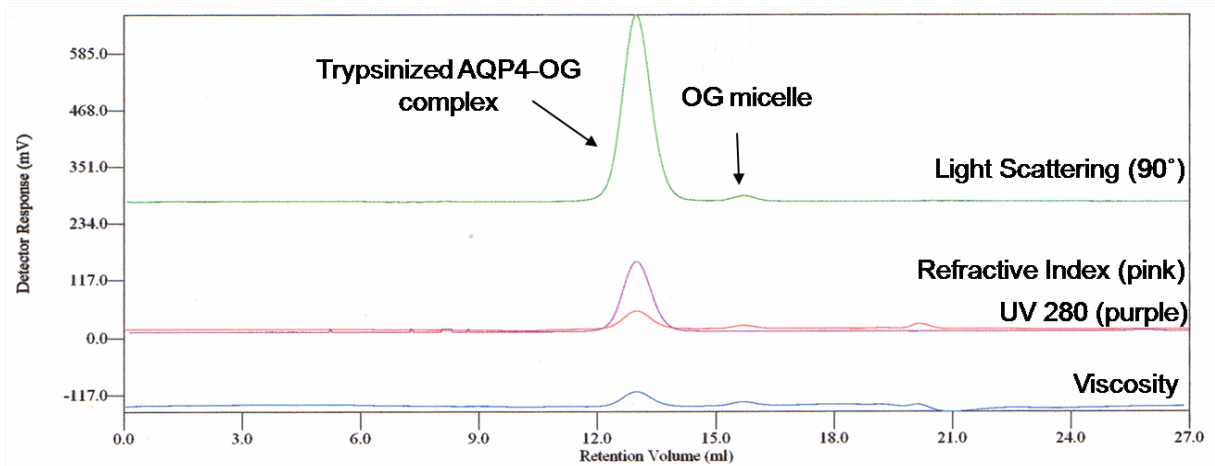
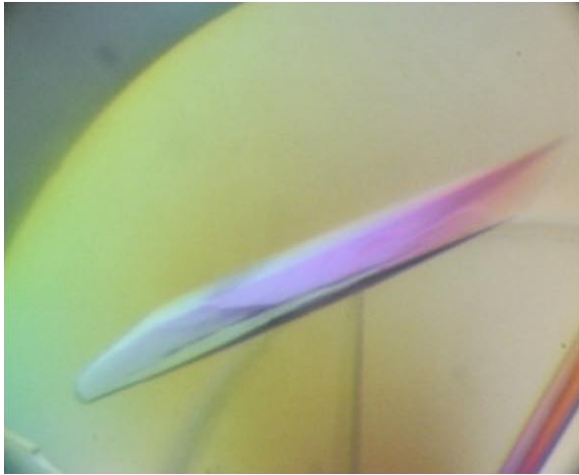


Figure 4-9

Tetra Detector Array Analysis (TDA) of the trypsinized AQP4 on size exclusion chromatography. OG is octyl- β -D-glucopyranoside. Peaks represent molecular species of different sizes. **Green** trace measures the light scattering of the sample at 90°. **Pink** trace measures the refractive index. **Purple** trace measures absorbance at 280 nm. **Blue** trace measures the viscosity of the sample.

Rods



Pyramids

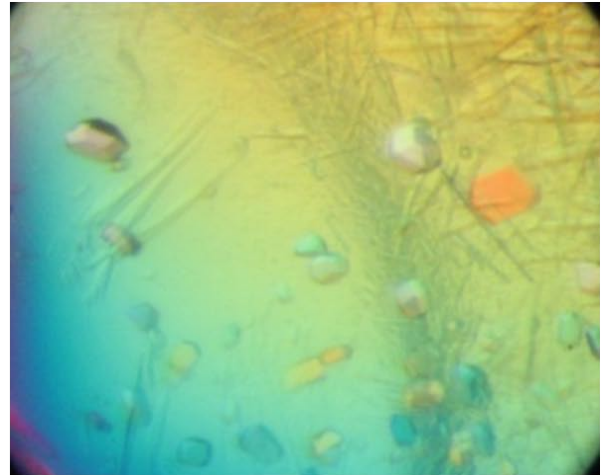


Figure 4-10

Crystals of the trypsinized AQP4. The *rods* are about 1.5 mm in length and 0.2 mm in width.

The *pyramids* are about 30-50 microns in size. Both are found in the same crystallization condition.

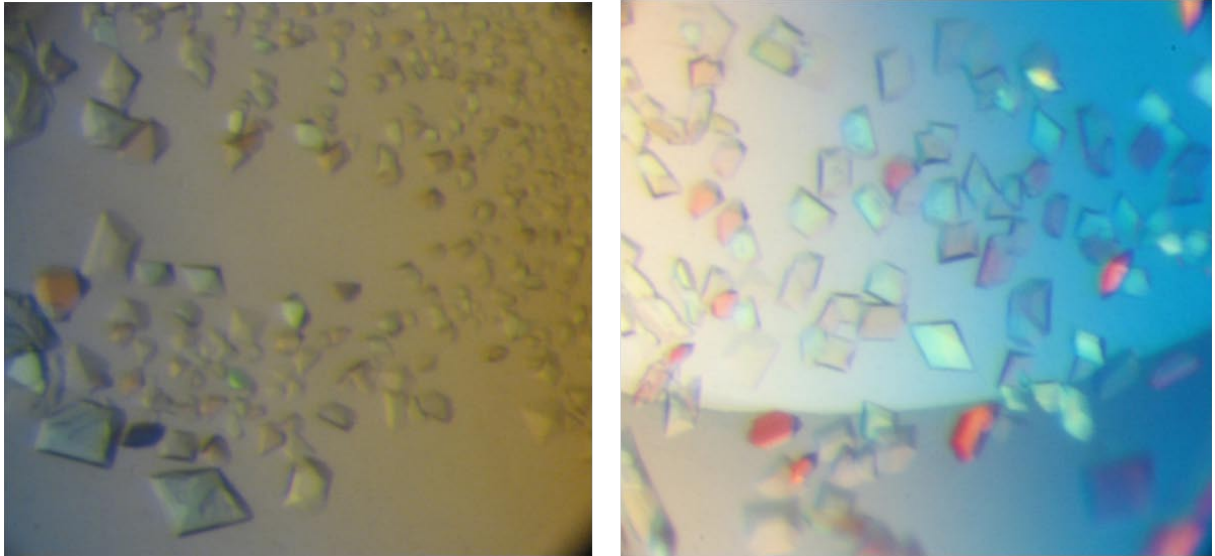


Figure 4-11

Crystals of the trypsinized AQP4 after using the Hampton Silver Bullet condition A1 as additive. Only pyramids are grown in the drops. The *pyramids* are about 50-100 microns in size.

X-ray Trips (ALS)	Crystals	Best Resolution
10/05/07	21 (full length)	8.5
10/24/07	34	8.2
11/09/07	49	8.5
11/28/07	52	7.6
12/07/07	24	11
1/10/08	56	8.5
1/16/08	28	14
2/04/08	28 (trypsinized, rod shaped)	6
2/22/08	69 (trypsinized, pyramids)	3.2
3/05/08	44	3.3
3/28/08	44	2.5
4/10/08	40 (cryo: 50% paraffin, 50% paratone-N)	1.8
Total: 489 crystals, 114 24-well trays, screen >2000 conditions		

Table 4-1

A Summary of the synchrotron trips to the Adance Light Source at Lawrence Berkeley National Laboratory. Crystals of the full-length protein were screened on trips until 1/16/2008, and then the pyramidal-shaped crystals from the trypsinized protein were screened after 2/22/2008.

REFERENCES

1. Michel H. General and practical aspects of membrane protein crystallization. *In Crystallization of Membrane Proteins* 1991: 73-88.
2. Kimura Y, Vassilyev DG, Miyazawa A, Kidera A, Matsushima M, Mitsuoka K, Murata K, Hirai T, Fujiyoshi Y. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* 1997 Sep 11; 389(6647): 206-211.
3. Miyazawa A, Fujiyoshi Y, Unwin N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* 2003 Jun 26; 423(6943): 949-955.
4. Gonen T, Cheng Y, Sliz P, Hiroaki Y, Fujiyoshi Y, Harrison SC, Walz T. Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* 2005 Dec 1; 438(7068): 633-638.
5. Hiroaki Y, Tani K, Kamegawa A, Gyobu N, Nishikawa K, Suzuki H, Walz T, Sasaki S, Mitsuoka K, Kimura K, Mizoguchi A, Fujiyoshi Y. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol* 2006 Jan 27; 355(4): 628-639.
6. Landau EM, Rosenbusch JP. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci U S A* 1996 Dec 10; 93(25): 14532-14535.
7. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 2007 Nov 23; 318(5854): 1258-1265.

8. Sugahara M, Kunishima N. Novel versatile cryoprotectants for heavy-atom derivatization of protein crystals. *Acta Crystallogr D Biol Crystallogr* 2006 May; 62(Pt 5): 520-526.
9. Meng G, Futterer K. Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. *Nat Struct Biol* 2003 Nov; 10(11): 935-941.
10. Ito K, Arai R, Fusatomi E, Kamo-Uchikubo T, Kawaguchi S, Akasaka R, Terada T, Kuramitsu S, Shirouzu M, Yokoyama S. Crystal structure of the conserved protein TTHA0727 from *Thermus thermophilus* HB8 at 1.9 Å resolution: A CMD family member distinct from carboxymuconolactone decarboxylase (CMD) and AhpD. *Protein Sci* 2006 May; 15(5): 1187-1192.
11. Hayashi Y, Matsui H, Takagi T. Membrane protein molecular weight determined by low-angle laser light-scattering photometry coupled with high-performance gel chromatography. *Methods Enzymol* 1989 172(514-528).
12. Maezawa S, Hayashi Y, Nakae T, Ishii J, Kameyama K, Takagi T. Determination of molecular weight of membrane proteins by the use of low-angle laser light scattering combined with high-performance gel chromatography in the presence of a non-ionic surfactant. *Biochim Biophys Acta* 1983 Sep 28; 747(3): 291-297.
13. Wen J, Arakawa T, Philo JS. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal Biochem* 1996 Sep 5; 240(2): 155-166.

Chapter 5

An “in-house” Approach to Gene Optimization and Synthesis

Research completed by

Joseph D. Ho and Robert M. Stroud

ABSTRACT

With the increasing number of available genomes of eukaryotic organisms, many more unknown proteins are now available for characterization. Although the use of natural gene sequences for protein expression may appear to be the quickest approach, such sequences may not be optimal for expression in *E. coli* or other host organisms due to their high G/C or A/T content, codon bias, mRNA secondary structures, or the presence of introns. With gene synthesis, the protein coding sequence can be optimized for the expression system of choice. While gene synthesis service is commercially available, the cost is relatively high, ranging from \$0.49 per base pair to \$1.60 per base pair plus minimum setup charge and shipping. An efficient “in-house” approach is developed here where academic laboratories can perform gene optimization and synthesis at less than 45 cents per base pair covering primers, PCR reactions, and sequence verification.

INTRODUCTION

It was Dr. Francis Crick in 1958 who first coined the term “the central dogma of molecular biology” stating that as information is transferred from DNA to RNA to protein, “once information has passed into protein, it cannot get out again” (1). Crick later reiterated this point in 1970 that “information cannot be transferred back from protein to either protein or nucleic acid” (2). While machineries that can synthesize RNA or DNA from protein molecules are yet to be found (if there are any!), one realizes that there are indeed many different combinations by which a piece of protein molecule can be encoded in DNA. Due to the redundancy in the genetic code (20 amino acids encoded by 61 tRNAs), each amino acid, with the exception of methionine and tryptophan, can be encoded by 2 to 6 different synonymous codons. By using different synonymous codons, there can be many way to encode the same amino acid in a protein molecule.

The frequencies by which the cell uses these synonymous codons are different across diverse organisms (3). In general, mRNA expressivity correlates with tRNA abundance so that highly expressed genes tend to contain synonymous codons that are recognized by the more abundant tRNAs (4, 5). Another factor that can affect the bias in tRNA population is the A+T/G+C content of the genome. For example, the genome of *P. falciparum* is 80% A+T rich, where the human genome is 59% A+T. The extent to which the expressivity of the gene correlates with codon bias can be quantitated by calculating the codon adaptation index (CAI) (6-8), namely that gene with high CAI tends to have high expression level.

Codon bias across organisms has implications in recombinant gene expression. A recombinant gene cloned from one organism may have poor heterologous expression in another host because the gene may have a low CAI, namely that the synonymous codon usage in the gene does not match well to the host's tRNA populations; and hence, there is a lower expression level in the host. This problem can be circumvented by codon optimization, in which all rarely used codons are substituted with the synonymous codons that are more frequently found in the expression host. Codon optimization has been applied successfully in the area of DNA vaccination by which parasitic and viral proteins can be expressed in mammals at high levels to elicit a strong immune response (9-12). Codon optimization is also important in generating high amounts of recombinant protein for X-ray structure determination. The gene for the *Plasmodium falciparum* aquaporin (PfAQP) was resynthesized to suit heterologous expression in *E. coli*, and the increased expression of the protein allowed for the determination of the crystal structure (13).

Currently, many commercial services are available for codon optimization and gene synthesis (DNA2.0, GENEART, GenScript, Biomatik, Mr.Gene to name a few), but the cost is still relatively high, ranging from \$0.49 per base pair to \$1.60 per base pair plus setup charge and shipping. In this chapter, an "in-house" approach is presented where two eukaryotic genes, the 240 base-pair α -bungarotoxin from *Bungarus multicinctus* and the 717 base-pair acetylcholine binding protein from *Lymnaea stagnalis*, were optimized for heterologous expression in *E.coli* and synthesized. The cost averaged to less than 45 cents per base pair covering primers, PCR reactions, and sequence verification.

RESULTS AND DISCUSSION

Codon optimization and gene design of α -bungarotoxin and the acetylcholine binding protein (AchBP)

Protein sequences of α -bungarotoxin and the AchBP were entered into the Gene Morphing System (GeMS) (<http://software.kosan.com/GeMS>) (14) and the genes were designed with flanking NcoI and EcoRI sites (**Figure 5-1a & 1b**). An extra PvuII site was design to divide the AchBP into 2 parts (**Figure 5-1b**). The purpose of dividing the gene into two segments was to synthesize the two halves of the gene in parallel and ligate them together at the ligation step. Because the NcoI, EcoRI, and PvuII sites do not share the same overhang, during ligation, there is only one way for the gene to be ligated into the vector.

The codon adaptation index (CAI) for expression in *E. coli* was also calculated for both genes before and after the codon optimization by GeMS (**Table 5-1a and 5-1b**). In both cases, the GeMS algorithms improved the CAIs of the genes. Please see “Differential expression of human aquaporin 4 (AQP4) in *Pichia pastoris*” for a discussion on the effect of CAIs on protein expression in *Pichia pastoris*.

Because of the recent acquisition of Kosan Biosciences by the Bristol-Myers Squibb Company, the website that hosts GeMS is no longer available. Another academic source is the OPTIMIZER (<http://genomes.urv.es/OPTIMIZER/>) (15), which can also perform the same computer algorithm to generates a DNA sequence that is optimized for expression in a particular host along with user input for the inclusion and exclusion of restriction enzyme

sites in the gene. One may also find some commercial websites that allow users to generate the optimized gene sequence without purchasing the gene synthesis package.

Two Methods for the PCR-based gene synthesis and the gene synthesis of the 240 base-pair α -bungarotoxin from *Bungarus multicinctus*

A piece of double-stranded DNA can be synthesized one of two ways. First, pairs of overlapping primers can be designed to cover the gene from the 5' end to the 3' end. This is the “conventional” method (**Figure 5-2, left**). Because the “conventional” method involves all the primers at equal concentration, there can be many smaller segments generated that are not completely incorporated into the final product. In the case of the 240 base-pair α -bungarotoxin, the “conventional” method was attempted, and non-specific products were observed (**Figure 5-3, left**). Or, pairs of overlapping primers can be designed “inside-out”, and the primers are mixed together with the innermost set at the lowest concentration (16) (**Figure 5-2, right**). The “inside-out” method produces a more specific product because during the course of a 30-cycle PCR, all the intermediate sizes are incorporated into the final product because the outermost set of primers are at the highest concentration and can therefore turn the intermediate fragments into the final product. Comparing with the “conventional” method, the synthesis of the 240 base-pair α -bungarotoxin using the “inside-out” method generated a cleaner product than the “conventional” method (**Figure 5-3, right**).

The product from the “inside-out” PCR was cloned into pET27b and transformed. After minipreping the DNA of the transformants, 3 out of 3 positive clones with the α -bungarotoxin gene were found (**Figure 5-4**). Clones were verified by DNA sequencing, and

all 3 clones of the α -bungarotoxin gene have no mutation, and the process from primers to final construct in pET27b took no more than 3 days (**Table 5-3**).

Gene synthesis of the 717 base-pair acetylcholine binding protein (AchBP) from *Lymnaea stagnalis*

The AchBP was split into two segments for the purpose of doing PCR of both segments in parallel (**Figure 5-1b**). Each segment is spanned by 10 primers (**Figure 5-5**). Putting all 10 primers together using the “inside-out” method resulted in non-specific products, so only 6 primers were included in the first round (**Figure 5-5, left**). Then the 2 templates of the first PCR (~250 base pairs) were used with the outer 4 primers to generate the two final products (~350 base pairs) (**Figure 5-5, right**). The extension of the template size (from ~250 to ~350 base pairs) showed that the outer 4 primers were successfully incorporated into the original template. Then, the two segments were ligated together into pET27b and transformed into *E. coli*. After minipreping the DNA of the transformants, 2 out of 6 positive clones were found for the AchBP gene (**Figure 5-6**). Clones were verified by DNA sequencing. Each of the clones of the acetylcholine binding protein gene contained two mutations (different in each clone). The mutations were easily fixed by QuikChange site-directed mutagenesis (Stratagene). There were less positive clones for AchBP gene because the ligation reaction was a tri-molecular event at NcoI, PvuII, and EcoRI (**Figure 5-1b** and **Figure 5-6, left**). We think the strategy of engineering a restriction site for ligating the two segments together is not as efficient as using another set of oligoes to PCR the two segments together.

Differential expression of human aquaporin 4 (AQP4) in *Pichia pastoris*

The human AQP4 gene was codon optimized and synthesized for *E. coli* expression. However, no protein expression was observed in *E. coli*, so the gene was subcloned into pPICZ for expression in *Pichia pastoris*. About 1 mg/mL of protein expression was observed. Another expression study was conducted where the original human AQP4 gene was also cloned into pPICZ and expressed. The codon adaptation index (CAI) was calculated for both sequences using the codon usage table of *Pichia pastoris* (<http://genomes.urv.es/CAIcal/>) (8). In this case, the human AQP4 gene has a higher CAI than the *E. coli*-optimized human AQP4 gene (**Table 5-2**), meaning that the original human sequence has a better fit for the codon usage of *Pichia pastoris* than the *E. coli*-optimized sequence. Indeed, the original AQP4 sequence had a better expression than the *E. coli*-optimized sequence in *Pichia pastoris* as measured by Western analysis (**Figure 5-7**). Expressing the original human AQP4 gene in *Pichia* resulted in about 5 mg/mL of purified protein (see chapter 2 and 3). This shows that codon usage is an important parameter in heterologous expression.

Cost effectiveness of “in-house” gene optimization and synthesis

The approximate cost of dollar per base pair for gene optimization and synthesis was calculated for both genes (**Table 5-3**). The “in-house” approach averages ~ \$0.43 per base pair, which is only slightly lower than the minimum rate found on the internet of \$0.49 per base pair offered by Mr. Gene (www.mrgene.com). However, the “in-house” approach is without shipping and minimum setup charge per gene and is faster and more versatile in terms of making mutations and small changes in the sequence during synthesis. Overall, some academic laboratories will find this approach to be faster and more cost-effective when synthesizing a gene along with a set of mutants together.

MATERIALS AND METHODS

Codon optimization and primer design

Protein sequence of α -bungarotoxin from *Bungarus multicinctus* (Accession number: AF056400) and the acetylcholine binding protein from *Lymnaea stagnalis* (Accession number: AF364899) were taken from the NCBI database. The signal peptides from both proteins were removed based on the analysis on SignalP 3.0 server (<http://www.cbs.dtu.dk/services/SignalP/>). The two protein sequences were entered into the Gene Morphing System (GeMS) (<http://software.kosan.com/GeMS>) (14) and codon-optimized DNA sequences suitable for *E. coli* expression were generated by the website. The two genes were also designed with flanking NcoI and EcoRI ends for subcloning into the *E. coli* expression vector pET27b. For the acetylcholine binding protein, the gene was divided into two segments with one segment flanking by NcoI and PvuII, and the other, PvuII and EcoRI. The PvuII site was used to ligate the two segments together. The final designed gene was 240 base-pair long for α -bungarotoxin and 717 base-pair long for acetylcholine binding protein. Then, the codon-optimized DNA sequences were entered into DNABWorks website (<http://helixweb.nih.gov/dnaworks/>) (17) using these options: annealing temperature at 62°C, oligo length at 60 nucleotides, codon frequency threshold at 10%, oligonucleotide concentration at 1×10^{-7} M, Na⁺/K⁺ concentration at 0.05 M, Mg²⁺ concentration at 0.002 M, number of solutions at 1, and the TBIO (thermodynamically Balanced Inside-Out Mode Output) option checked (16). The oligonucleotides generated from the output from DNABWorks were ordered through Elim Biopharmaceuticals (<http://www.elimbio.com/>) using the 50 nmol synthesis scale.

Gene synthesis of the 240 base-pair α -bungarotoxin from *Bungarus multicinctus*

A total of 1 40-mer and 5 60-mer oligonucleotides were generated by DNAWorks for the 240 base-pair gene. For the “inside-out” PCR, all six primers were mixed in 50 μ L PCR reactions with primer 3 & 4 at 40 nM, primer 2 & 5 at 100 nM, and primer 1 & 6 at 200 nM (**Figure 5-3, right**). For the “conventional” PCR, all six primers were mixed at equal concentration of 200 nM (**Figure 5-3, left**). Phusion[®] DNA polymerase was used along with the HF buffer (Finnzymes). A total of 30 PCR cycles was performed. The PCR product was gel purified using the Qiagen QIAquick Gel Extraction Kit. The PCR product was further PCR’ed again with primer 1 and 6 and gel purified to ensure the completeness of the insert, and gel purified again.

Ligation of the 240 base-pair α -bungarotoxin into pET 27b

The Insert was then digested with NcoI and EcoRI and ligated into pET27b with T4 DNA ligase (NEB) at NcoI and EcoRI sites for 45 minutes at room temperature. Sub-cloning efficiency DH5 α chemical competent cells (Invitrogen) were used for transformation of the ligation reaction. Transformants were plated on fresh LB plates with 50 μ g/mL kanamycin.

Gene synthesis of the 717 base-pair acetylcholine binding protein from *Lymnaea stagnalis*

A total of 9 55-mer, 9 60-mer, 1 43-mer, and 1 27-mer oligonucleotides were generated by DNAWorks for the 717 base-pair gene. Only the “inside-out” PCR reactions were performed. Two separate PCR experiments were conducted. Primer 1 to 10 covered the first 384 nucleotides of the gene and primer 11 to 20 covered the second 339 nucleotides of the gene

(**Figure 5-1b**). First set of PCR reactions involved only 6 primers (primer 5 & 6 and primer 15 & 16 at 40 nM, primer 4 & 7 and primer 14 & 17 at 100 nM, and primer 3 & 8 and primers 13 & 18 at 200 nM (**Figure 5-5**). PCR products were gel purified and used as template for the second set of PCR reactions with primers 2 & 9 and primers 12 & 19 at 40 nM and primers 1 and 10 and primers 11 and 20 at 200 nM. Phusion[®] DNA polymerase was used along with the HF buffer (Finnzymes). A total of 30 PCR cycles was performed. The PCR product was gel purified using the Qiagen QIAquick Gel Extraction Kit.

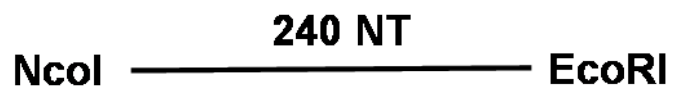
Ligation of the 717 base-pair acetylcholine binding protein into pET 27b

The two PCR products: 384 nucleotides and 339 nucleotides (**Figure 5-1b** and **Figure 5-5**) were digested with NcoI, EcoRI, and PvuII and ligated into pET27b with T4 DNA ligase (NEB) at NcoI and EcoRI sites, for 2 hours at room temperature. Sub-cloning efficiency DH5 α chemical competent cells (Invitrogen) were used for transformation of the ligation reaction. Transformants were plated on fresh LB plates with 50 μ g/mL kanamycin.

DNA minipreps

Transformants were picked from fresh plate with antibiotics and inoculated into 15 mL LB cultures with antibiotics and grown overnight (~16 hours) at 37°C. Cultures were then centrifuged at 5,000 g for 10 minutes, and the pellet resuspended in 250 μ L P1 buffer (Qiagen). All subsequent steps were performed according to the protocol in the QIAprep Spin Miniprep Kit (Qiagen).

(a)



(b)



Figure 5-1

Gene design of (a) the 240 base-pair α -bungarotoxin and (b) the 717 base-pair acetylcholine binding protein (AchBP). NcoI, PvuII, and EcoRI are restriction sites. NT stands for nucleotides, or base pairs.

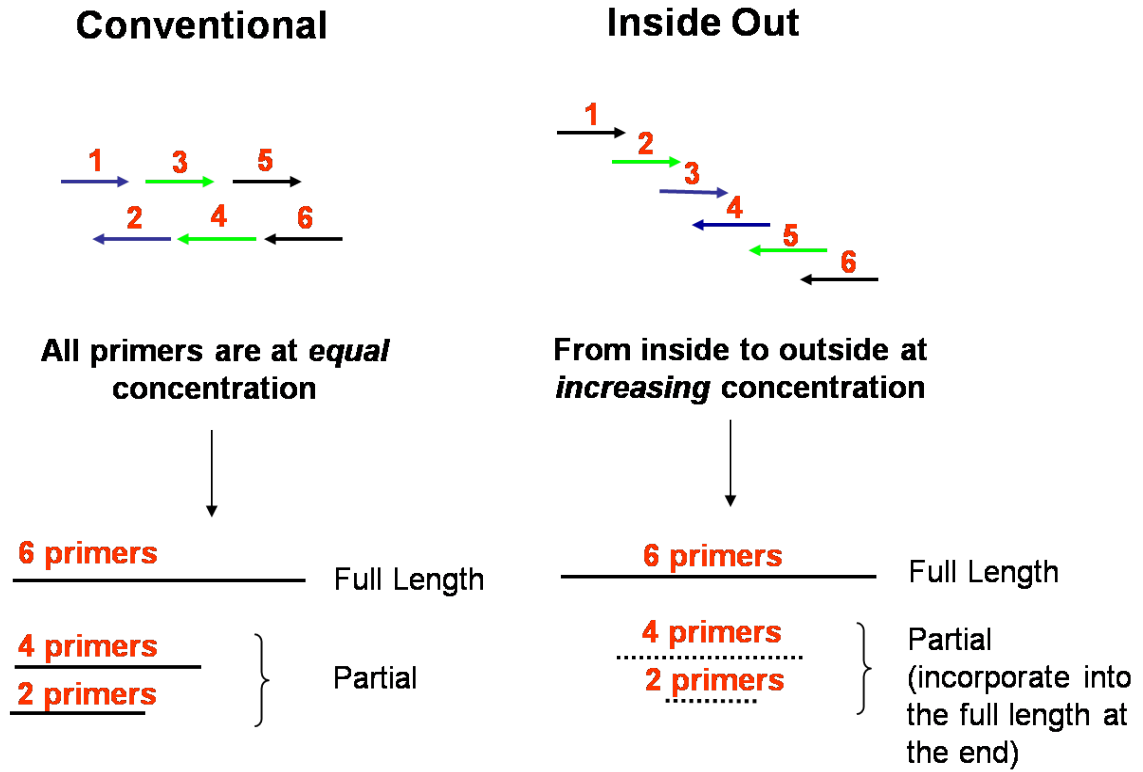
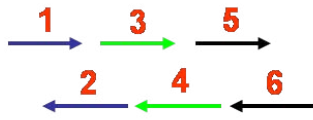


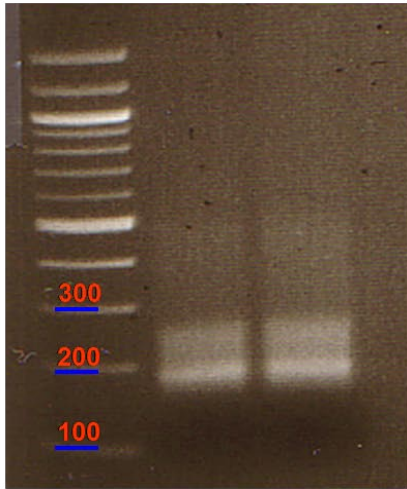
Figure 5-2

Two methods for the PCR-based gene synthesis. Arrows are individual oligoes/primers for the PCR reactions. The “conventional” method requires all primers at equal concentration during PCR. The “inside out” method uses sequentially more primers as the DNA lengthens, and by the time the PCR reaction is completed, all initial products should be incorporated into the final product because they are at lower concentration to start with.

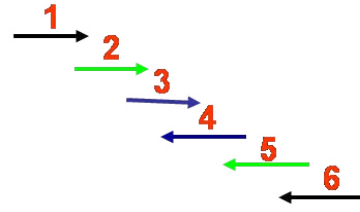
Conventional



All primers are at *equal* concentration



Inside Out



From inside to outside at *increasing* concentration

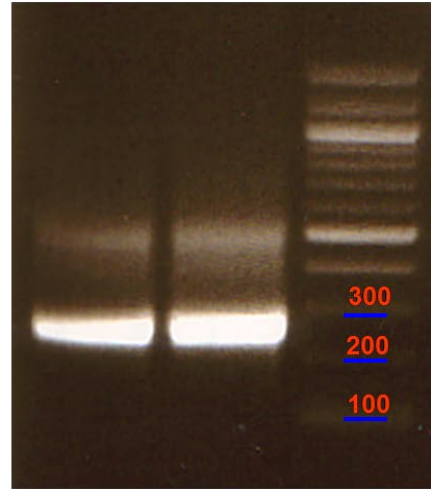


Figure 5-3

PCR products obtained from the “conventional” and the “inside-out” methods of PCR.

Arrows represent the oligoes/primers used for the PCR reactions. PCR products were loaded onto a 2% agarose gel. Showing here are duplicates of the PCR reaction. The DNA ladder is the New England Biolab 100 bp ladder (N3231, NEB).

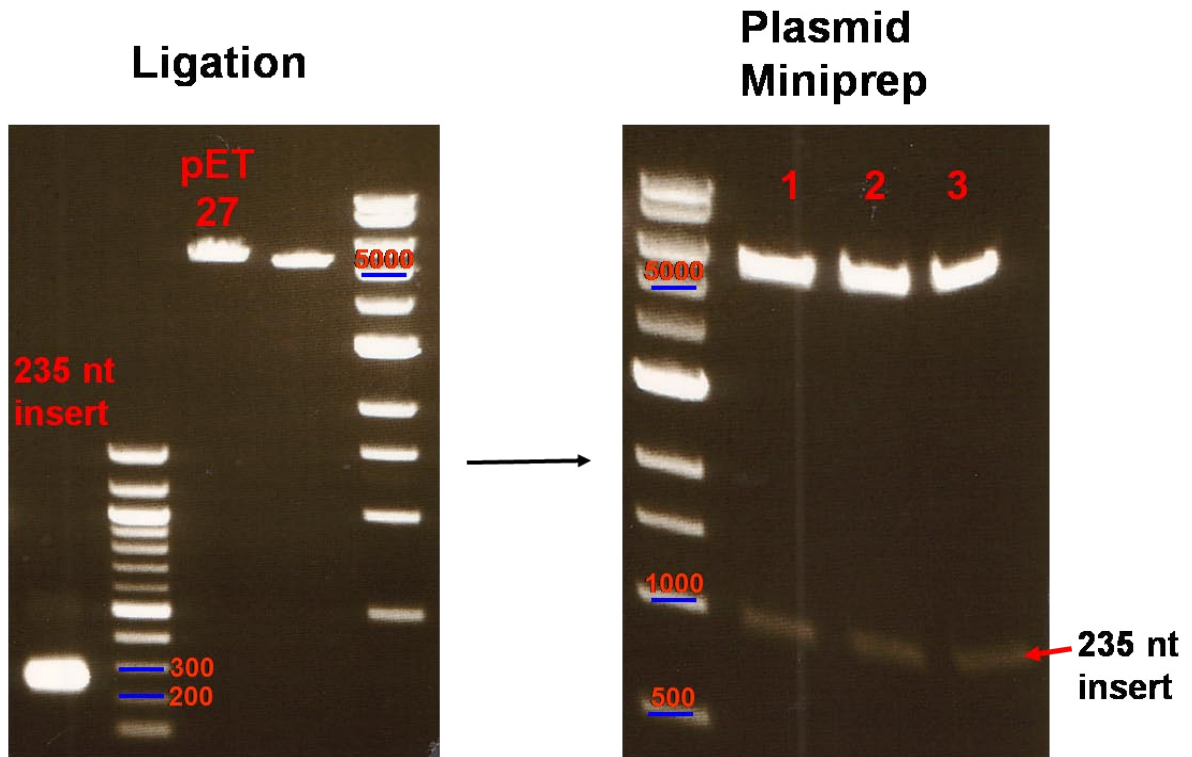


Figure 5-4

1% agarose DNA gel of the α -bungarotoxin used for the ligation reaction and the plasmid miniprep results. The plasmid minipreps were digested with NcoI and EcoRI prior to running the gel.

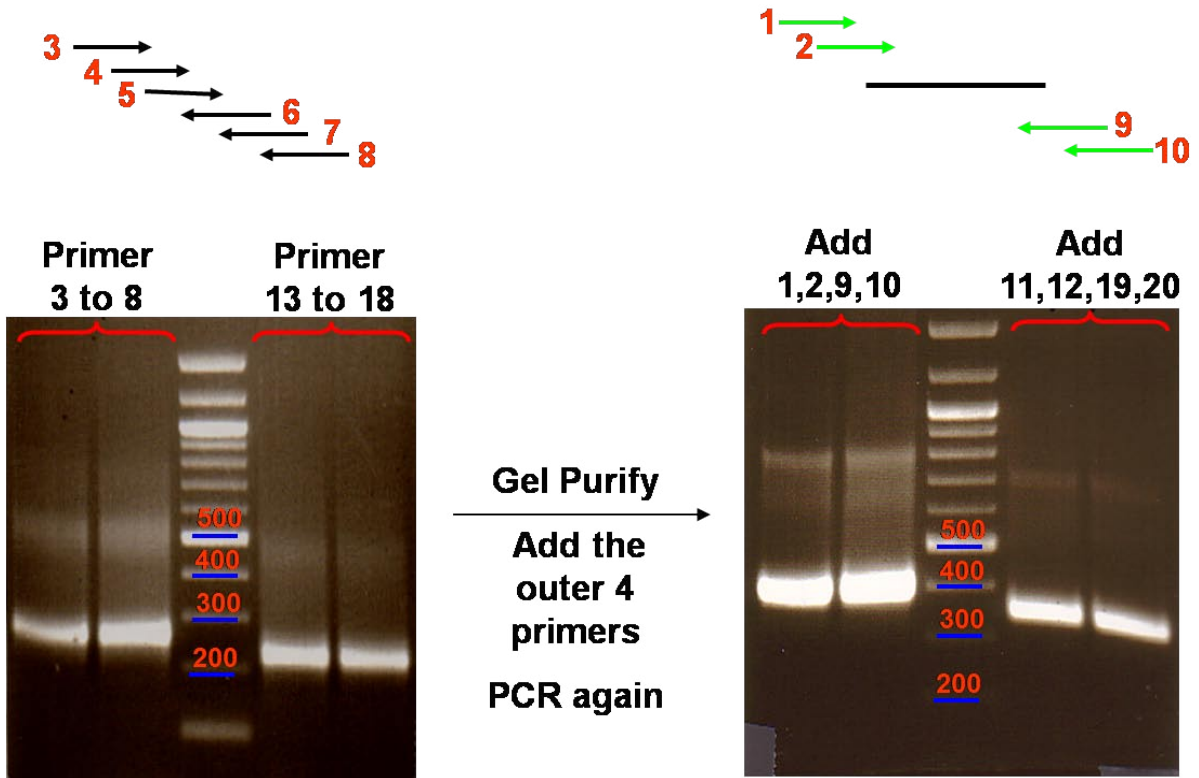


Figure 5-5

PCR products of the synthesis of the NcoI-PvuII segment and the PvuII-EcoRI segment of the 717 base-pair acetylcholine binding protein. Arrows represent the oligoes/primers used for the PCR reactions. The NcoI-PvuII segment was generated by primer 1 through 10, and the PvuII-EcoRI segment was generated by primer 11 through 20. PCR products were loaded onto a 2% agarose gel. Showing here are duplicates of the PCR reaction. The DNA ladder is the New England Biolab 100 bp ladder (N3231, NEB).

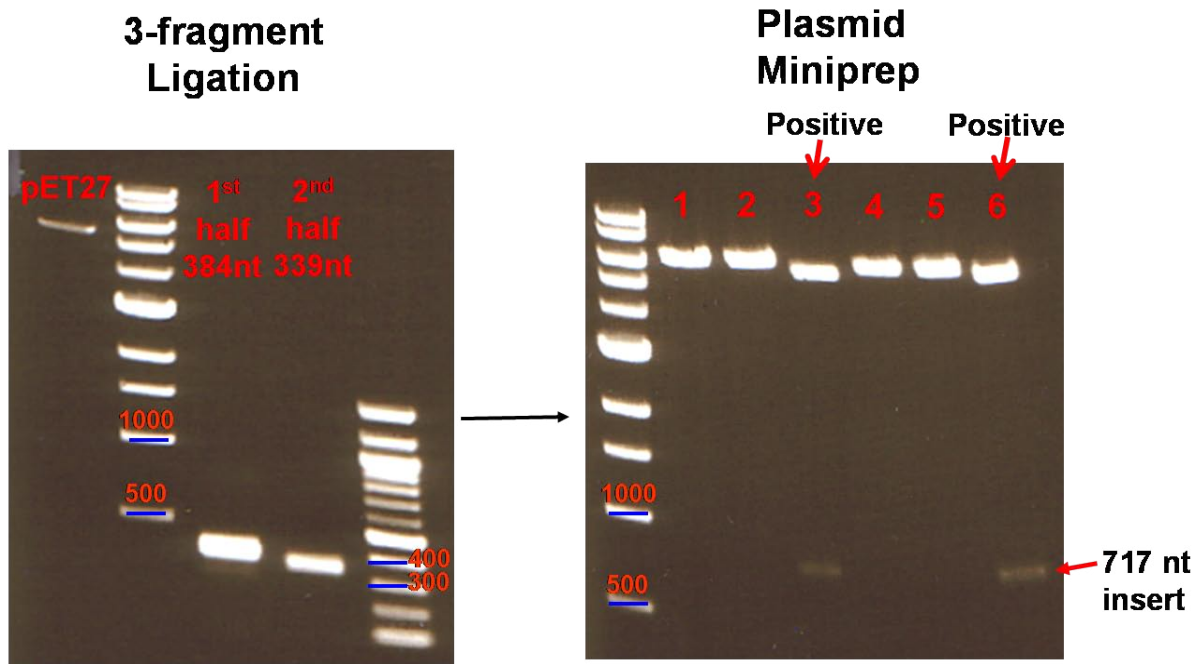


Figure 5-6

1% agarose DNA gel of the AchBP fragments used for the ligation reaction and the plasmid miniprep results. The plasmid minipreps were digested with NcoI and EcoRI prior to running the gel.

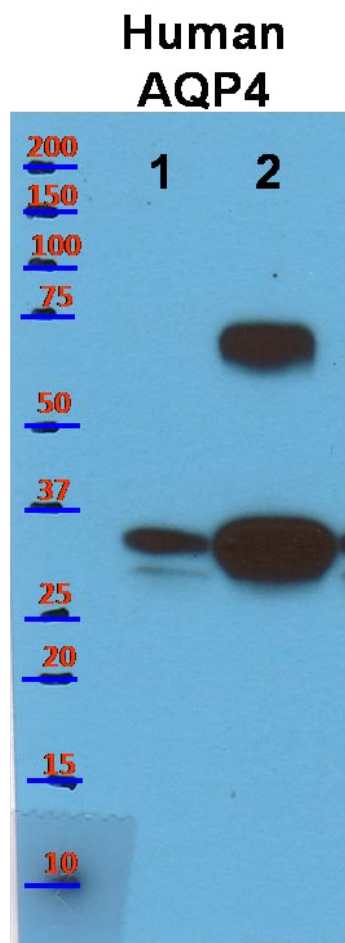


Figure 5-7

Anti-flag Western result of the heterologous expression of human aquaporin 4 (AQP4) in *Pichia pastoris*. Lane 1: Expression using the *E. coli*-optimized AQP4 gene. Lane 2: Expression using the original AQP4 gene from human.

Table 5-1a

CAI comparison of α -bungarotoxin

<u>α-bungarotoxin</u>	<u>CAI</u>	<u>%G+C</u>
Original	0.657	48.6
Optimized for <i>E. coli</i>	0.78	54.4

Using the codon usage table for *Escherichia coli* CFT073

<http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=199310>

Table 5-1b

CAI comparison of AchBP

<u>acetylcholine binding protein</u>	<u>CAI</u>	<u>%G+C</u>
Original	0.693	50.6
Optimized for <i>E. coli</i>	0.775	50.1

Using the codon usage table for *Escherichia coli* CFT073

<http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=199310>

Table 5-2

CAI comparison for human AQP4

<u>Human AQP4</u>	<u>CAI</u>	<u>%GC</u>
Optimized for <i>E. coli</i>	0.664	50.9
Original human sequence	0.711	49.2

Using the codon usage table for *Pichia pastoris*

<http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4922>

235 bp α -bungarotoxin

Primer cost: \$91 at \$0.28 per base (Operon)

Sequencing Cost: \$9 per reaction X 1 reaction

PCR Cost: \$5 (1 reaction)

Total: ~\$105

Average: \$0.45 per base **Time: 3 days**

717 bp acetylcholine binding protein

Primer Cost: \$248 at \$0.28 per base (Operon)

Sequencing Cost: \$9 per reaction X 1 reaction

PCR Cost: \$20 (4 reactions)

Total: ~\$277

Average: \$0.39 per base **Time: 5 days**

Table 5-3

A summary of the financial cost of the “in-house” approach to codon optimization and gene synthesis. The primer/oligos cost is calculated based on the pricing of the 10 nmol synthesis scale by Operon (www.operon.com).

REFERENCES

1. Crick FH. On protein synthesis. *Symp Soc Exp Biol* 1958 12(138-163).
2. Crick F. Central dogma of molecular biology. *Nature* 1970 Aug 8; 227(5258): 561-563.
3. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985 January 1, 1985; 2(1): 13-34.
4. Bennetzen JL, Hall BD. Codon selection in yeast. *J Biol Chem* 1982 March 25, 1982; 257(6): 3026-3031.
5. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucl Acids Res* 1982 November 25, 1982; 10(22): 7055-7074.
6. Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987 Feb 11; 15(3): 1281-1295.
7. Puigbo P, Bravo IG, Garcia-Vallve S. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 2008 9(65).
8. Puigbo P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* 2008 3(38).
9. Ivory C, Chadee K. DNA vaccines: designing strategies against parasitic infections. *Genet Vaccines Ther* 2004 Dec 3; 2(1): 17.
10. Narum DL, Kumar S, Rogers WO, Fuhrmann SR, Liang H, Oakley M, Taye A, Sim BK, Hoffman SL. Codon optimization of gene fragments encoding Plasmodium falciparum merzoite proteins enhances DNA vaccine protein expression and immunogenicity in mice. *Infect Immun* 2001 Dec; 69(12): 7250-7253.

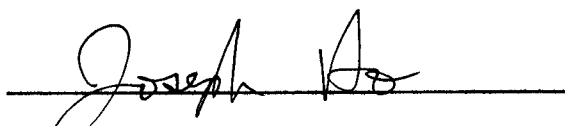
11. Mossadegh N, Gissmann L, Muller M, Zentgraf H, Alonso A, Tomakidi P. Codon optimization of the human papillomavirus 11 (HPV 11) L1 gene leads to increased gene expression and formation of virus-like particles in mammalian epithelial cells. *Virology* 2004 Aug 15; 326(1): 57-66.
12. Ramakrishna L, Anand KK, Mohankumar KM, Ranga U. Codon optimization of the tat antigen of human immunodeficiency virus type 1 generates strong immune responses in mice following genetic immunization. *J Virol* 2004 Sep; 78(17): 9174-9189.
13. Newby ZE, O'Connell J, 3rd, Robles-Colmenares Y, Khademi S, Miercke LJ, Stroud RM. Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nat Struct Mol Biol* 2008 Jun; 15(6): 619-625.
14. Jayaraj S, Reid R, Santi DV. GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res* 2005 33(9): 3011-3016.
15. Puigbo P, Guzman E, Romeu A, Garcia-Vallve S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucl Acids Res* 2007 April 16, 2007: gkm219.
16. Gao X, Yo P, Keith A, Ragan TJ, Harris TK. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res* 2003 Nov 15; 31(22): e143.
17. Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* 2002 May 15; 30(10): e43.

Publishing Agreement

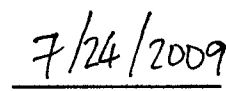
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date