Development and application of a high-throughput cell-free expression system

Vom Fachbereich Biologie der Technischen Universität Darmstadt

zur

Erlangung des akademischen Grades

eines Doctor rerum naturalium

genehmigte Dissertation von

BSc (Biological Engineering) Lei Kai

aus Xuzhou (P.R. China)

Berichterstatter: Prof. Dr. Ralf Kaldenhoff

Mitberichterstatter: Prof. Dr. H. Ulrich Göringer

Mitberichterstatter: Prof. Dr. Volker Dötsch

Tag der Einreichung: 25.07.2012

Tag der mündlichen Prüfung: 16.10.2012

Darmstadt 2012

D 17

Table of content

Summary

At the post genomics age, the study of proteomics attracts more and more attention and help to decipher the hidden information inside genes. However, the study of proteomics required large amount or large number of proteins to be synthesized. During the past several decades, heterologous overexpression of recombinant protein from bacteria, yeast, insect cells and mammalian cells was developed. Especially the *E. coli* system, which is the most studied expression system, was used to obtain different recombinant proteins from both prokaryotic and eukaryotic organisms. Still, there are a large number of eukaryotic proteins that cannot be functionally overexpressed in the *E. coli* system. Techniques were developed to overexpress recombinant protein in eukaryotic host cells including yeast, insect cells and mammalian cells. However, these eukaryotic overexpression systems normally require more complicated processes and take more time to finally get the protein. Cell-free (CF) expression systems, using cell lysates, which contain the required enzymes, extra energy resources and substrates needed for translation, were developed. The CF system was first used to discover the genetic code by *in vitro* translation of poly-U RNA sequences. Later on, it was applied to protein synthesis, attracted more attention and became an alternative method for protein expression besides traditional cell based expression systems.

Eliminating the need of viable cells opens the CF expression system for various modifications. The CF expression system was widely used to express difficult target proteins like toxins, membrane proteins (MPs), insoluble or aggregation-prone proteins that failed to be expressed in *in vivo*. The open nature and emendation for throughput automation made the CF system an ideal system for proteome research. In this thesis, a high throughput CF expression platform, in particular, addressing difficult targets including MPs and aggregation prone proteins is developed.

In order to establish the high-throughput CF expression platform, a batch configuration which is less time consuming, was selected. The amount of protein expressed in low-volume batch configurations can be addressed by more sensitive detection methods. With the TECAN Freedom EVO 200 liquid handling robot, the full automation of the CF expression system was achieved. This platform provides parallel expressions in 96-well plates with a minimum volume of 25 µl. The whole expression takes 2-4 hours with an expression yield above 0.5 mg per 1 ml reaction mixture. A custom designed software E.Y.E.S which provides a more versatile and friendly user interface to design the screening of CF expression conditions, was created. With the help of E.Y.E.S program the time for robot programming and pipetting was largely reduced. The whole pipetting time was reduced to 30-45 minutes for 96 reactions. The high-throughput expression optimization can be combined with fluorescence detection by fusing shifted green fluorescent protein (sGFP) to the target protein, allowing a fast read out.

With the established high-throughput CF expression system, we systematically screened the effect of common chemical stabilizers. The sGFP was chosen as a target protein. In the first step, the compatibility of the selected chemical stabilizers was evaluated by measuring the fluorescence of CF expressed sGFP. With these experiments, the effect of selected stabilizers can be classified into three different types: positive, negative and tolerant. In case of a positive effect, the selected stabilizer increased the expression or folding of sGFP, which resulted in an increase of the measured fluorescence. A negative effect indicated that selected chemical compounds were toxic to the CF expression system and inhibited the expression at low compound concentrations. A tolerant effect represented that the expression of sGFP in CF system had a linear decrease with the increase concentration of this type of chemical stabilizers. Those stabilizers which had a positive or tolerant effect on the CF expression system can be used and applied to those unstable protein or aggregation prone proteins to improve the solubility and stability of the sample as soon as they were expressed. In order to check the application of these chemical stabilizers, we choose two targets. One is human GNA1 (EC 2.3.1.4) which catalyzes the transfer of the Acetyl group from Acetyl Coenzyme A (AcCoA) to the primary amine of D-glucosamine-6-phosphate to form N-acetyl-D-glucosamine-6-phosphate (GlcNAc-6P) and Coenzyme A (CoA). Another target is CurA-halogenase, which is the first enzyme within the biosynthesis of curacin A (a bioactive compound with potent anti-proliferative activity) pathway. The biosynthesis of CurA is mediated by a 2.2 MDa hybrid polyketide synthase (PKSs) and non-ribosomal peptide synthetases (NRPSs). Both of the two targets were expressed as partially soluble *in vivo* and *in vitro*. We selected several chemical stabilizers to check the ability of these chemical stabilizers to improve the soluble expression of the selected two targets in CF expression system. For the case of halogenase, a maximum of around 25% increased soluble expression was observed with choline. For GNA1, a beneficial effect of choline was also observed. To further check if the fluorescence signals from the protein fused with sGFP correlated with the function of the fusion partners, we created a GNA1-sGFP construct which sGFP was fused at the C-terminal of GNA1. Results of the experiment showed that the function of GNA1 was not always correlated with the fluorescence signal from sGFP. In the case of L-arginine, the fluorescence of the fusion protein was increased without the increase of the functionality of GNA1, which might explained by that L-arginine had a beneficial effect on the sGFP folding. While for the case of choline, the increase of fluorescence signal correlated with the functional assay. Finally, in order to provide an example of correlated screening of two chemical compounds, we select PEG8000 and choline to perform this experiment. When using PEG8000 alone, an increase of soluble expression was observed around 17% was observed. While with choline alone around 23% more soluble expression was achieved. While when using both PEG8000 and choline at the optima concentrations, an increase of around 60% was obtained. This 60 % increase was much more than simply sum up of the 17% and 23%. Which might indicate an additional beneficial effect came from

the interaction of the two compounds. The successful application of the systematically screening of chemical stabilizers offered an example to use the automatic throughput CF platform.

Preliminary studies on the application of the established throughput CF platform on MPs were performed. We choose the D-CF mode for directly soluble expression of MPs in this platform. However, one challenge was to find a monitor to characterize the expression and even the folding of expressed MPs. Taken the idea from *in vivo* expression system for MPs, which use a C-terminal GFP fusion for the fast detection of soluble expressed MPs in either *E. coli* or yeast expression system. The same approach was used in this CF system. By following the fluorescence of the fused sGFP, the soluble expression of MPs can be detected and quantified. Since detergents was used to provide hydrophobic environment for MPs, the folding of GFP in presence of detergents needs to be evaluated. However, results showed that the folding of normal sGFP was highly influenced by detergents. sGFP only gave relative high fluorescence accounts in limited detergents like Brij derivatives. With the best detergent, less than 60% sGFP was folded compared to the control without detergent. With other commonly used detergents like Digitonin, TritonX-100 and DDM very low fluorescence signals were detected, indicating that sGFP was poorly folded in these detergents. The limitation of choosing detergents reduced the application this method for many MPs because many MPs cannot be soluble expressed with the Brij detergents. In order to solve this problem, a mutant of GFP was selected, which was called superfolderGFP. This superfolderGFP was super stable, less affected by the out environment and fast folding compared to normal sGFP. A comparison experiment of detergent screening was performed with both sGFP and superfolderGFP showed that in general the superfolderGFP had a better detergent tolerance. Those detergents like TritonX-100, Digitonin and DDM can allow the folding of superfolderGFP to more than 70%. With Brij-78 the fluorescence was almost the same compared to the control. Furthermore, a MP SugE (small multidrug transporter) was selected as an example for detergent screening. Result showed that SugE fused with superfolderGFP gave relative good fluorescence signal in Brij-78, Brij-58, Digitonin and Lauryl-MNG. With Brij-58, the fluorescence was even higher than Brij-78. These two experiments showed that with superfolderGFP more detergents can be select for D-CF expression of MPs fusion partners. Introducing the superfolderGFP as fusion monitor increases the application of the throughput CF platform in detergents screening for MPs. Further experiment with detergents and lipids mixture provide another promising tool for obtain soluble MPs with functions. However, more research still need to be done to provide more information about the ratio of lipid/detergents in practical applications.

Two different MPs, SugE (small multidrug transporter from *E. coli*), AQP4 M23 (aquaporin 4 from mouse), were selected and studied. The sample preparation of SugE, offered a new approach for obtain pure, homogeneous and stable MP protein sample. SugE was expressed in CF system first as precipitate and later on re-solubilized with detergent. A two-step re-solubilization strategy was applied. First treat the protein pellet from the reaction with relative mild detergent like $NDSB_{256}$.

Second, re-solubilize the remaining pellet with DDM. The first step, the mild detergent will solubilize quite a lot of impurities not the target protein SugE. In the second step, DDM will solubilize most of the SugE protein and small amount of impurities. With this two-step re-solubilization approach, without any chromatography approach, the sample was already relative pure. Later on with another SEC, the sample was purified as one pure band on SDS-PAGE and showed homogeneous elution profile on the gel filtration column. Later on, in order to obtain more detergent conditions which can keep SugE homogeneous and stable, a detergent exchange was performed directly on the SEC column. Detergents with different micelle size were selected and some of them also showed comparable elution profile of DDM.

Another MP we selected was AQP4 M23. The mouse aquaporin 4 was selected as a representative of mammalian aquaporins. The protein was synthesized in an *E. coli* extract based CF system with two different expression modes, and the efficiencies of two modes were compared. In both, the P-CF (CF MP expression as precipitate) mode generating initial aquaporin precipitates as well as in the D-CF (CF MP expression in presence of detergent) mode, generating directly detergent solubilized samples, we were able to obtain mg amounts of protein per ml of CF reaction. Purified aquaporin samples solubilized in different detergents were reconstituted into liposomes, and analyzed for the water channel activity. The calculated Pf value of proteoliposome samples isolated from the D-CF mode was 133 µm/s at 10^{o}C, which was 5 times higher as that of the control. A reversible inhibitory effect of mercury chloride was observed, which is consistent with previous observations of *in vitro* reconstituted aquaporin 4. In this study, a fast and convenient protocol was established for functional expression of aquaporins, which could serve as basis for further applications such as water filtration.

Zusammenfassung

In der postgenomischen Ära werden Proteomstudien attraktiver und helfen dabei die in den Genen verborgene Information zu entschlüsseln. Jedoch erfordern Proteomstudien die Synthese größerer Proteinmenge oder Proteinanzahl. In den letzten Jahrzehnten wurde die heterologe Überexpression von rekombinantem Protein in Bakterien, Hefe, Insektenzellen und Säugetierzellen entwickelt. Mit dem *E. coli* System, welches am besten untersucht ist, ist man in der Lage, gänzlich verschiedene rekombinante Proteine sowohl prokaryotischen als auch eukaryotischen Ursprungs zu erhalten. Nichtsdestotrotz gibt es eukaryotische Proteine die im bakteriellen System nicht funktionell aktiv exprimiert werden können. Diese können oft mit Hilfe von eukaryotischen Expressionssystemen wie Hefe, Insektenzellen oder Säugetierzellen hergestellt werden. Eukaryotische Überexpressionssysteme erfordern jedoch komplexere Verfahren und es erfordert einen zeitaufwendigen Prozess um letztendlich das rekombinante Protein zu erhalten. Das zellfreie (CF) Expressionssystem nutzt Zelllysate welche die Maschinerie für Transkription, Translation, zusätzliche Energieressourcen und Substrate für die Proteinsynthese enthält. Das CF System wurde als erstes dazu genutzt um den genetischen 'Code' durch *in-vitro* Translation von Poly-U RNA Sequenzen zu entdecken. Erst später wurde die Methode weiterentwickelt um Proteine zu synthetisieren, zog mehr Aufmerksamkeit auf sich und etablierte sich als alternative Methode neben den konventionellen zellulären Expressionssystemen.

Der Verzicht auf die Nutzung lebender Zellen macht das CF Expressionssystem offen für eine Vielzahl an Modifikationen. Es kann eine breitere Anwendung finden, insbesondere in der Expression von schwierigen Untersuchungsobjekten wie Toxinen, Membranproteinen (MPs), unlöslichen oder aggregationsgefährdeten Proteinen, welche in konventionellen *in-vivo* Expressionssystemen nicht hergestellt werden können. Die offene Natur und die Verbesserung der Hochdurchsatz-Automatisierung machte das CF System zu einem idealen System für die Proteomforschung. In dieser Arbeit wird versucht eine Hochdurchsatz-CF Expressionsplattform zu entwickeln, die schwierigen Untersuchungsobjekten einschließlich Membranproteinen und aggregationsgefährdeten Proteinen gerecht wird.

Um eine Hochdurchsatzplattform zu etablieren wurde ein CF 'Batch' Expressionssystem ausgewählt, welches relativ einfach gehandhabt werden kann und zeitsparend ist. Die relativ niedrige Expressionsausbeute kann durch sensitivere Detektionsmethoden ausgeglichen werden. Mit dem TECAN Freedom EVO 200 'liquid handling' Roboter wurde eine volle Automatisierung des CF Expressionssystems erreicht. Die Plattform ermöglicht die schnelle, parallele CF Expression von bis zu 96 Reaktionen in einem Minimalvolumen von 25 µl. Die komplette Expression dauert 2-4 Stunden und erreichte eine Ausbeute von 0.5 bis 1 mg pro mL Reaktionsmixtur. Die speziell entworfene Software E.Y.E.S stellt eine vielfältige und nutzerfreundliche Umgebung für den

Hochdurchsatzscreen von CF-Expressionen zur Verfügung. Das E.Y.E.S Programm reduzierte die Zeit für das Programmieren und Pipettieren des Roboters in hohem Maße. Die gesamte Pipettierzeit für 96 Reaktionen wurde auf 30-45 Minuten reduziert. Mit Hilfe dieser robotergestützten Plattform kann die Optimierung von Expressionsbedingungen mit Fluoreszenz-Detektion durch Fusion von grün fluoreszierendem Protein (GFP) an das zu untersuchende Protein kombiniert werden.

Mittels dieses etablierten Hochdurchsatz CF Expressionssystems kann der Effekt von gängigen chemischen Stabilisatoren auf das CF System untersucht werden. 'shifted GFP' (sGFP) wurde als das zu untersuchende Protein gewählt. Zunächst wurde die Kompatibilität der selektierten chemischen Stabilisatoren mittels Fluoreszenz von CF exprimierten sGFP evaluiert. Mit diesen Experimenten können die Effekte der selektierten chemischen Stabilisatoren in drei Klassen eingeteilt werden: positiv, negativ und tolerant. Ein positiver Effekt eines selektierten Stabilisators hatte eine Erhöhung der Fluoreszenz von sGFP zur Folge. Ein negativer Effekt impliziert dass die selektierte chemische Verbindung toxisch auf die CF Expression wirkt und die Expression inhibiert. Ein toleranter Effekt bedeutet dass die Expression von sGFP im CF System linear mit zunehmender Konzentration der chemischen Verbindung abnimmt. Die Stabilisatoren welche einen positiven oder toleranten Effekt auf das CF Expressionssystem haben, können genutzt werden um die Löslichkeit und Stabilität von instabilen oder aggregationsgefährdeten Proteinen zu erhöhen, sobald sie exprimiert worden sind. Um die Anwendung von chemischen Stabilisatoren beurteilen zu können, haben wir zwei Proteine ausgewählt: Das GNA1 (EC 2.3.1.4) katalysiert den Transfer einer Acetylgruppe vom AcetylCoenzymA (AcCoA) auf das primäre Amin von D-Glucosamin-6-phosphat mit den Nebenprodukten N-acetyl-D-Glucosamin-6-phosphat (GlcNAc-6P) und Coenzym A (CoA). Die Halogenase ist das erste Enzym in der Biosynthese von Curacin A (CurA, eine bioaktive Verbindung mit potenter anit-proliferativer Aktivität). Die Biosynthese von CurA wird durch die 2.2 MDa große Hybridpolyketidsynthase (PKS) und nicht-ribosomale Peptidsynthetasen (NRPS) durchgeführt. Beide werden teilweise löslich *in-vivo* bzw. *in-vitro* exprimiert. Verschiedene chemische Stabilisatoren wurden selektiert um die Möglichkeit einer verbesserten löslichen Expression im CF System zu testen. Im Fall der Halogenase wurde eine um 25% erhöhte Expression mit Cholin beobachtet. Auch im Fall von GNA1 wurde ein zuträglicher Effekt von Cholin beobachtet. sGFP wurde an den C-Terminus von GNA1 fusioniert, um zu testen ob die Amplitude des Fluoreszenzsignals mit der Funktion des Fusionspartners korreliert ist. Es zeigte sich dass die Funktion nicht vollständig mit dem Fluoreszenzsignal von sGFP korreliert. Im Fall von L-Arginin erhöhte sich die Fluoreszenz des Fusionsproteins ohne einen Anstieg der Funktionalität von GNA1. Dies lässt sich durch einen positiven Effekt auf die Faltung von sGFP erklären. Schlussendlich wurden PEG 8000 und Cholin ausgewählt, um ein Beispiel eines korrelierten Effekts von zwei chemischen Verbindungen zu zeigen. PEG 8000 alleine erhöhte die lösliche Expression um 17% während Cholin eine 23% höhere Expression bewirkte. Im Kontrast dazu erhöhte die Zugabe von beiden Verbindungen bei optimierter Konzentration die lösliche Expression um 60 %. Diese

60 %ige Zunahme war mehr als einfach die Summe von 17 % und 23 % was dafür spricht, dass der positive Effekt aus einer Interaktion der beiden Verbindungen resultiert. Die erfolgreiche Anwendung des 'Screenings' nach chemischen Stabilisatoren ermöglichte ein Anwendungsbeispiel für die vollautomatische CF-Hochdurchsatzplattform.

Vorläufige Studien bezügliche der Anwendung der etablierten Hochdurchsatzplattform auf Membranproteine (MP) wurden durchgeführt. Der D-CF Modus wurde für die direkte lösliche Expression von MPs gewählt. Eine Herausforderung war es einen Weg zu finden, die Expression und die Faltung von MPs verfolgen zu können. Das Verfahren welches bei der *in-vivo* Expression von MPs genutzt wird, nämlich die C-terminale Fusion an sGFP zur schnellen Detektion von MPs aus *E. coli* oder Hefe, wurde für das CF System übernommen. Über die Fluoreszenz des fusionierten sGFP wird die lösliche Expression von MPs detektiert und quantifiziert. Da Detergenzien genutzt wurden um den MPs eine hydrophobe Umgebung zu geben, musste die Faltung von sGFP in Anwesenheit von Detergenzien evaluiert werden. Die Faltung von sGFP zeigt sich sehr stark beeinflusst durch die Detergenzien. Ein relativ hohes Fluoreszenzsignal wurde lediglich für Brij-Derivate beobachtet. Im besten Fall waren weniger als 60% GFP im Vergleich zur Kontrolle ohne Detergenz gefaltet. Häufig genutzte Detergenzien wie Digitonin, TritonX-100 und DDM führten zu einem geringeren Fluroeszenzsignal, implizierend dass sGFP in diesen Detergenzien in geringfügigen Ausmaß faltet. Die Einschränkung der Detergenzwahl schränkte die Anwendung der Methode für viele MPs ein, da diese in Brij-Detergenzien oft nicht löslich gehalten werden können. Um dieses Problem zu lösen wurde eine GFP-Mutante ausgesucht, das sogenannte ''superfolderGFP''. Dieses ist extrem stabil, weniger durch seine Umgebung beeinflusst und faltet schneller als normales sGFP. Ein Vergleichsexperiment zeigte dass 'superfolderGFP' generell Detergenz-toleranter als sGFP ist. In TritonX-100, Digitonin und DDM ist 'superfolderGFP' zu mehr als 70% gefaltet. In Brij-78 zeigt 'superfolderGFP' ein fast identisches Fluoreszenzsignal im Vergleich zur Kontrolle. Das Membranprotein SugE wurde als Beispiel für einen 'Detergenzscreen' gewählt. SugE fusioniertes 'superfolderGFP' zeigt ein relativ gutes Fluoreszenzsignal in Brij-78, Brij-58, Digitonin und Lauryl-MNG. Brij-58 zeigt sogar ein höheres Fluoreszenzsignal als Brij-78. Diese beiden Experimente zeigten, dass mit 'superfolderGFP' mehr Detergenzien für die D-CF Expression von MP-Fusionspartnern ausgewählt werden können. Dies erhöht die Anwendung der Hochdurchsatzplattform für das 'Screening' von Membranproteinen in Detergenzien enorm. Auch die Zugabe von Detergenz- und Lipidmixturen ermöglicht die lösliche und funktionelle Expression von MPs. Jedoch ist mehr Wissen über das optimale Lipid/Detergenz-Verhältnis für die praktische Anwendung nötig.

Zwei Membranproteine wurden für diese Arbeit ausgewählt und studiert: SugE ('small multidrug transporter' aus *E. coli*) und AQP4 M23 (Aquaporin 4 aus der Maus). Die Präparation von SugE eröffnete einen neuen Weg um reines, homogenes und stabiles MP zu erhalten. SugE wurde im CF System als Präzipitat exprimiert und später mit einem Detergenz resolubilisiert. Diese

Resolubiliserung wurde in zwei Schritten durchgeführt. Als erstes wurde das Pellet mit einem relativ milden Detergenz wie $NDSB_{256}$ behandelt, welches somit Verunreinigungen durch Resolubilisierung entfernt. Als nächstes wurde SugE sowie eine geringe Menge an Verunreinigungen mit DDM resolubilisiert. Mittels dieser zweistufigen Resolubilisierung wird ohne Anwendung eines Chromatographie-Schritts bereits eine relativ reine Probe erhalten. Eine folgende Größenaufschlusschromatographie zeigte ein homogenes Elutionsprofil sowie eine einzelne Bande auf einem SDS-PAGE (Natriumdodecylsulfat-Polyakrylamid Gelelektrophorese) Gel. Um die finale Ausbeute zu maximieren wurden verschiedene Detergenzien getestet, welche das SugE homogen und stabil halten. Detergenzien mit unterschiedlicher Mizellengröße wurden ausgewählt und zeigten ähnliche Elutionsprofile im Vergleich zu DDM.

Das Aquaporin 4 (AQP4 M23) aus der Maus wurde als Repräsentant für Aquaporine aus Säugetieren zu Studien selektiert. Das Protein wurde in einem *E. coli* Extrakt-basierten CF System in zwei Expressionsmodi synthetisiert und deren Effizienz wurde verglichen. Im P-CF (zellfreie Membranproteinexpression als Präzipitat) sowie im D-CF (zellfreie Membranproteinexpression in der Anwesenheit von Detergenz) Modus in welchen Aquaporin als Präzipitat bzw. direkt in Detergenzien gelöst vorliegt, wurde das Protein im Milligrammmaßstab pro mL Zell-frei Reaktion erhalten. Gereinigte Aquaporinproben wurden in verschiedenen Detergenzien gelöst, in Liposomen rekonstituiert und auf ihre Wasserkanalfunktionalität hin analysiert. Der berechnete Pf-Wert von Proteoliposomproben aus dem D-CF Modus betrug 133 µm/s bei 10 °C, und liegt damit um den Faktor 5 höher als die Kontrolle. Ein reversibler inhibitorischer Effekt von Quecksilberchlorid wurde beobachtet, welcher konsistent mit früheren Untersuchungen von *in-vitro* rekonstituiertem Aquaporin 4 ist. In dieser Studie wurde ein schnelles und geeignetes Protokoll zur funktionellen Expression von Aquaporinen erstellt, welches als Basis für weitere Anwendungen wie z.B. der Wasserfiltration dienen kann.

1. Introduction

MPs act as the link between the cytoplasm and the cellular environment. Essential cell processes like signal transmission, transport of substances through the membrane, generation of energy are associated with MPs. Therefore, many drugs are designed against MPs, which become important targets for medical and pharmaceutical studies. However, a major bottle neck for study MPs is the limitation in sample preparation. Traditional *in-vivo* expression system was well established, especially the bacterial Escherichia coli expression system, for producing recombinant proteins. However, *in-vivo* expression system based either on prokaryotic or eukaryotic cells do not work for the MPs production. The same difficulty also appears when try to express other problematic proteins like, toxins, instable enzymes, aggregation prone proteins. The toxic effect on the host cells, when producing MPs and other problematic proteins, always results in low expression yield and forming aggregation or unfolded state.

CF expression systems have been developed in recent times as alternative and promising tools to address the challenge of producing MPs and other problematic targets. The open nature and elimination of living host cells during protein expression offers a variety of advantages. It is proved that the toxic or inhibitory effects via recombinant MPs, toxins and other problematic proteins are minimized or even completely eliminated. Remove of the cell wall and membrane barrier gave the open nature of the CF expression systems. Additional compounds can be directly introduced into the system without the concern of transport and metabolic conversion problems. Additional protease inhibitors, ligands, cofactors or chemical stabilizers can be used as additives that might be helpful to stabilize and folding of the freshly expressed proteins. This characteristic of CF expression system provides a flexible and changeable environment for different targets. In the case of MPs, MPs can be expressed in the soluble form by offering an artificial hydrophobic environment via using detergent micelles, liposomes or nanolipoprotein particles. In case of instable and aggregation prone soluble proteins, additional chemical stabilizer can be introduced into the CF system as stabilizer to hold the correct folding of expressed proteins. Furthermore, CF reactions are often carried out in a small volume from a few milliliters to microliters and require only short incubation time of a few hours. High expression yield of several milligrams of proteins can be obtained per milliliter of reaction. With all the advantages above, CF expression system draws great attention for its application in preparative scale protein production, throughput screening and proteome studies.

1.1 Membrane proteins

MPs are attached to, or associated with, biological membranes. They consist 20-40% of all open reading frames in fully sequenced genomes and are targets of around 60% of all modern drugs. Human genome analysis in 2002 led to estimation of 6000-8000 targets of pharmacological

interest. Only a small part of these targets relate to approved drugs. A consensus number of 324 drug targets for all classes of approved therapeutic drugs was proposed [2]. Of these, 266 are human genome-derived proteins, and 58 are bacterial, viral, fungal or other pathogenic organism targets. The analysis by Overington *et al.* identifies in excess of 21 000 drug products marketed in US corresponding to 1357 unique drugs, of which 1204 are 'small-molecule drugs' and 166 are 'biological drugs'. 27% of these drugs bind to G-protein-coupled receptors (GPCRs; ~20 families, each family including up to five members), 13% to nuclear receptors, 7.9% to ligand-gated ion channels and 5.5% to voltage-gated ion channels. A selected target may have a unique approved drug, or a large number of me-too molecules. This draws great attention for researchers to study MPs in view of their fundamental role in all biological processes and for finding new targets for therapies.

1.1.1 Water channel proteins

Water is the most abundant and basic molecule in cell. The cell separates itself by the plasma membrane, which has a basic structure of phospholipid bilayer. The phospholipid bilayer is permeable to small molecule, and water can diffuse slowly through the membrane. However, there is a type of integral MP, which forms as a channel to facilitate the passing of water across the biomebranes. Such water channel proteins have been found in organisms from bacteria to plant and animals, named aquaporins (AQPs)[3]. AQPs are integral MPs that serve as channels in the transfer of water, and in some cases, small solutes across the membrane. They are conserved in bacteria, plants, and animals. Structural analyzes of the molecules have revealed the presence of a pore in the center of each aquaporin molecule. Since multiple isoforms of aquaporins are differentially expressed in cells and tissues, understanding of different roles of each AQP isoform therefore become important.

In human, 13 different and tissue specific AQPs are responsible for transport mechanisms and showed considerable clinical relevance[4]. Among them, AQP4 is known as a water-transporting aquaporin with two isoforms differing from their N-termini, which result from variable translation starting sites either from methionine M1 (323 aa) or methionine M23 (301 aa) [5]. AQP4 is found in many organs and tissues like kidney, skeletal muscle, stomach, lung, air epithelium and brain. Phenotype analysis of AQP4-knockout mice indicated its involvement in the brain water balance [6]. AQP4 expression was furthermore shown to be induced during spongiform encephalopathy [7], neuromyelitis optica, multiple sclerosis [8], and Alzheimer's disease [9]. In addition, AQP4 provides a molecular pathway for water permeability and homeostasis in the brain, and its astrocytic end-feet localization makes AQP4 a partner to blood-brain-barrier function [10].

Sufficient amounts of protein have been obtained for 2D crystallization and functional characterization of rat AQP4 [11]. Nevertheless, the purification processes are time-consuming and the yield with approx. 3 mg/l produced in insect cells [11] is still not satisfying if compared with

that of soluble proteins. Further biochemical and biophysical characterizations as well as high-throughput research towards the relationship between aquaporin and certain diseases remain therefore to be difficult. Recently, an increasing interest is to prepare some functional biomembranes integrated with bioactive AQPs to filtrate water [12]. However, the success of this amazing strategy largely depends on the availability of sufficient and functional AQPs, which still remains challenging [13, 14].

1.1.2 Transporters- multidrug resistance efflux pumps

The number of antibiotics to which bacteria have developed resistance has increased greatly, during the last decade. Therefore, the problem of inefficient antibiotics for treatment of infections became a big challenge. Another concern is about the increase number of bacteria, which is becoming resistant to more than one antibiotic. Efflux-pump genes and proteins are present in nearly all organisms. In bacteria, the genes that encode efflux pumps are located on the chromosome or on transmissible genetic elements, such as plasmids. It is well known that, in bacteria, efflux pumps can confer decreased susceptibility to antibiotics; however, such decreases do not always result in clinical levels of antibiotic resistance. Efflux pumps can be specific for one substrate or can transport a range of structurally dissimilar compounds (including antibiotics of different chemical classes). Those pumps that transport several compounds can be associated with multidrug (antibiotic) resistance (MDR). Many of these efflux pumps are therefore of clinical relevance because they can render a bacterial infection untreatable by the agent(s) of choice [15]. MDR mediate by specialized multidrug efflux transporters, which contain several different protein super families. They are the ATP binding cassette (ABC) superfamily, the major facilitator superfamily (MFS), the multidrug and toxic-compound extrusion (MATE) family, the small multidrug resistance (SMR) family and the resistance nodulation division (RND) family as shown in Fig. 1. MDR efflux pumps are classified as belonging to these families on the basis of the number of components that the pump has (single or multiple), the number of transmembrane-spanning regions that the transporter protein has, the energy source that the pump uses and the types of substrate that the pump exports. A single organism can express MDR efflux pumps from more than one family and/or more than one type of efflux pump belonging to the same family.

For instance, *Pseudomonas aeruginosa* can express more than one type of Mex efflux pump, and *E. coli* can express more than one type of Acr efflux pump (both of which are pumps belonging to the RND family). Efflux pumps can consist of either a single component or multiple components. Examples of efflux pumps from each family are shown in Fig.1

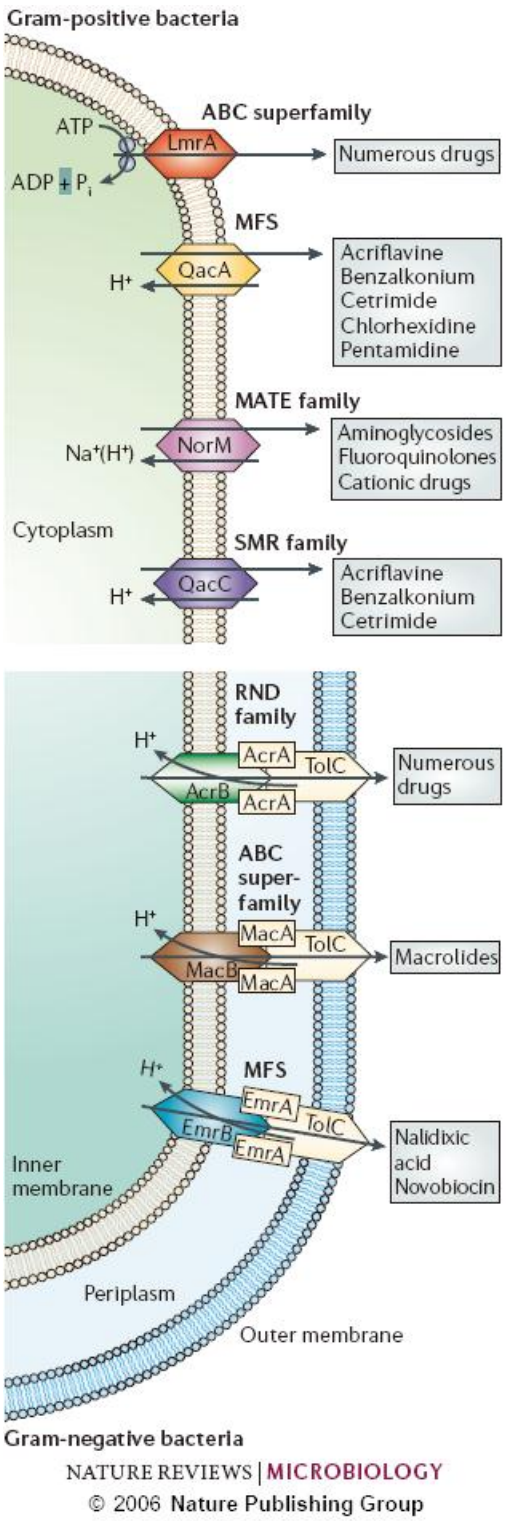


Figure 1. Multidrug-resistance efflux pumps. There are five families of multidrug-resistance efflux pumps: the ATP-binding cassette (ABC) superfamily, the major facilitator superfamily (MFS), the multidrug and toxic-compound extrusion (MATE) family, the small multidrug resistance (SMR) family and the resistance nodulation division (RND) family. A diagrammatic representation of the structure and membrane location of efflux pumps from each of these families is shown. Common examples of the individual proteins that form each class of efflux pump are indicated. Antibiotic substrates and examples of other substrates are also listed for each class of efflux pump. Multidrug-resistance efflux pumps expressed by Gram-negative bacteria usually have several components, and the outermembrane protein is typically TolC. Pi, inorganic phosphate [15].

Efflux pumps, which belong to RNA family, often expressed by Gram-negative bacteria, are associated with clinically important MDR. They are organized as tripartite systems (Fig. 1). These efflux pumps comprise of a transporter (efflux) protein (i.e. AcrB), which is located in the inner (cytoplasmic) membrane of bacteria; an accessory protein (known as a membrane-fusion protein, i.e. AcrA); and an outer-MP (also known as an outer-MP channel) (i.e. TolC) [16], which is located in the outer membrane of the bacterium. Using the AcrAB-TolC system (that is, the efflux pump comprising AcrA, AcrB and TolC) as an example, it is thought that the transporter protein AcrB, similar to AcrD [17], captures its substrates either from the phospholipids bilayer of the inner membrane or from the cytoplasm, and then transports them to the extracellular medium through TolC, which forms a channel in the outer membrane[18]. The cooperation between AcrB and TolC is mediated by the periplasmic accessory protein AcrA. Efflux through RND-family pumps is driven by the proton motive force [18], an electrochemical gradient in which the movement of hydrogen ions drives transport of the substrate. In *Enterobacteriaceae*, TolC function as the protein channel for different RND-family efflux pumps, and it can also interact with MFS transporters (for instance, EmrAB of *E. coli*) and ABC-superfamily transporters (for instance, MacAB of *E. coli*). Similarly, the RND-family outer-MP OprM of *P. aeruginosa* can interact with several RND-family proteins.

1.2 Aggregation and inclusion body - aggregation prone proteins

Protein aggregation is an inevitable result of cellular existence. Protein aggregates are oligomeric complexes of non-native conformers that arise from non-native interactions among structured, kinetically trapped intermediates in protein folding or assembly [19-21] (Fig. 2). Protein aggregation is facilitated by partial unfolding during thermal or oxidative stress and by alterations in primary structure caused by mutation, RNA modification or translational misincorporation [20, 22]. Protein aggregates can be either structured (e.g. amyloid; Fig. 2a) or amorphous (Fig. 2b). In both case, they are prone to behave insoluble and metabolically stable under physiological conditions. The accumulation of aggregation is tightly linked to neuronal degeneration or organ failure in many 'protein deposition' diseases. The fact that protein aggregates do not accumulate in unstressed cells, despite their continued production, is due in part to the existence of cellular 'quality control' machinery. This suppresses the formation of aggregates by ensuring the fidelity of transcription and translation, by chaperoning nascent or unfolded proteins, and by selectively degrading improperly folded polypeptides before they can aggregate [23, 24]. The prominence of protein aggregates in intracellular and extracellular lesions associated with cell death in many degenerative diseases (e.g. amyloid diseases, Alzheimer's disease, Parkinson's disease, Huntington's disease and alcoholic liver disease [25]) underscores the physiological importance of these quality control pathways and suggests that the failure of these systems to control aggregation might have catastrophic consequences for the organism. However, protein aggregation need not be pathogenic. For example, in yeast, cytoplasmic inheritance of aggregated prion proteins underlies the propagation of stable epigenetic traits that are not associated with any known pathology [26]. Although considerable research has addressed the mechanisms by which cells avoid forming protein aggregates, the way that cells deal with proteins once they have aggregated is relatively unexplored.

With the development of heterogeneous expression system, scientists are able to over express heterogeneous proteins in other host cells as recombinant proteins. The enteric bacterium *E. coli* is one of the most extensively used prokaryotic organisms for genetic manipulations and for the industrial production of proteins of therapeutic or commercial interest. Compared with other established and emerging expression systems[27], *E. coli* offers several advantages, including growth on inexpensive carbon sources, rapid biomass accumulation, amenability to high cell-density fermentations and simple process scale-up. It is, however, not uncommon that overexpressed recombinant proteins fail to reach a correct conformation and undergo proteolytic degradation or associate with each other to form insoluble aggregates of nonnative proteins known as inclusion bodies.

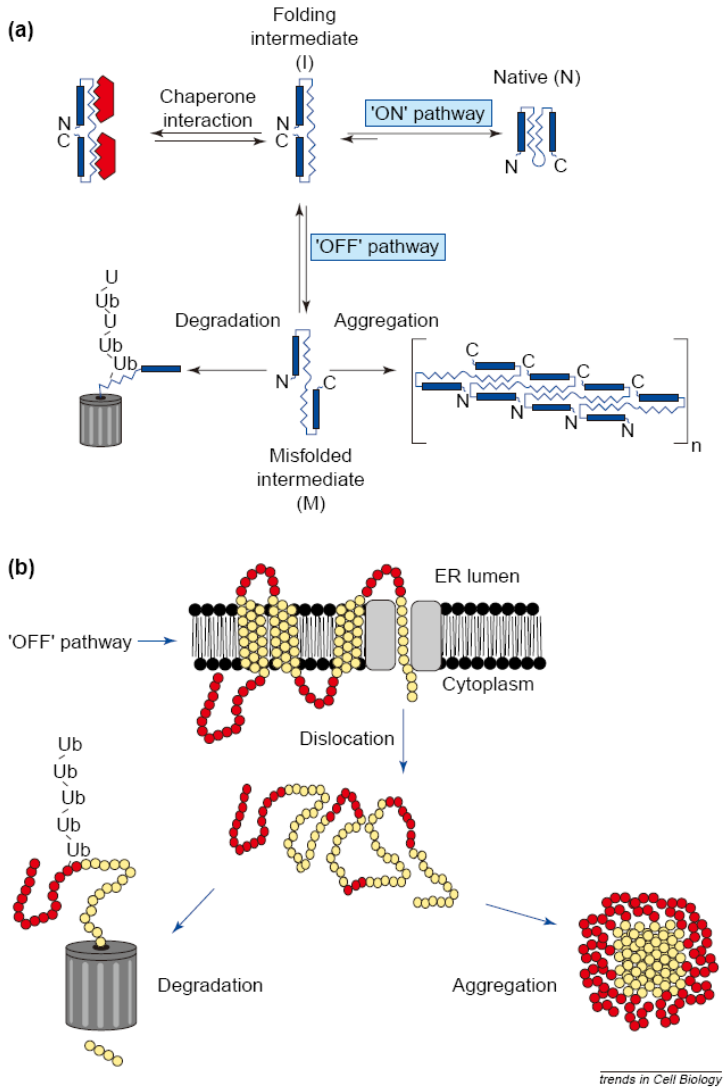


Figure 2. Protein aggregates arise from kinetically trapped products of protein folding or degradation. (a) Formation of structured aggregates. The overall efficiency of protein folding is the proportion of intermediate that assumes a native conformation. Intermediates can also assume misfolded, off-pathway conformations, which are prone to oligomerize into non-native aggregates. The ubiquitin–proteasome pathway competes for and degrades off-pathway intermediates before they can aggregate. (b) Misfolded integral membrane and secretory proteins are dislocated across the endoplasmic reticulum (ER) membrane to the cytoplasm for degradation by the ubiquitin-proteasome pathway. Such dislocated intermediates might be highly aggregation prone in the aqueous, reducing, environment of the cytoplasm. The proportions that are degraded and that aggregate are determined by the relative rates of proteasomal degradation and aggregation, and by the degree of coupling between dislocation and degradation[28].

For a heterologous protein, failure to rapidly reach a native conformation or to interact with folding modulators in a timely fashion has two possible consequences: partial or complete deposition into insoluble aggregates known as inclusion bodies or degradation. The likelihood of misfolding is increased by the routine use of strong promoters and high inducer concentrations that can lead to product yields exceeding 50% of the total cellular protein. Under such conditions, the rate of protein aggregation is often much greater than that of proper folding and folding modulators are rapidly titrated. A second factor contributing to inclusion body formation is the inability of bacteria to support all post-translational modifications that a protein requires to fold. For instance, the formation of intra- or intermolecular disulfides is not possible in the reducing cytoplasm of wild-type *E. coli*, which results in the aggregation of certain disulfide bond-rich proteins (e.g., Fab antibody fragments).

Inclusion bodies can accumulate in the cytoplasm or periplasm depending on whether or not a recombinant protein has been engineered for secretion. The target typically accounts for 80-95% of the inclusion body material and is contaminated by outer MPs, ribosomal components and a small amount of phospholipids and nucleic acids that likely adsorb upon cell lysis [29]. Folding modulators (e.g., DnaK, GroEL and IbpA/B) are sometimes-but not always-associated with inclusion bodies [30, 31]. Cytoplasmic inclusion bodies are porous ovoids or cylinders with

maximum characteristic length and volume of 1 μm and 0.6 μm [32], respectively [31, 33, 34]. However, hemispherical inclusion bodies of 0.5-μm diameter have been observed in the periplasm [33]. In the cytoplasm, inclusion bodies grow from structured folding intermediates9 at nearly constant rates and around nucleation cores that are mutually exclusive. Thus, multiple inclusions of different sizes may be present within a single cell [31]. Because inclusion bodies are resistant to proteolysis and contain large amounts of relatively pure material, their formation is often exploited for the production of proteins that are toxic, unstable or easy to refold. Finding optimal conditions for efficient refolding requires considerable optimization, but acceptable yields can usually be achieved using established strategies [35].

1.3 Systems for expressing challenging protein targets

Lot of researches has been done to optimize the expression yield and quality of challenging targets like aggregation prone proteins, toxins and especially MPs. In each step of expression, there can be potential improvement. For instance, the design of suitable vector (e.g. include fusion partners for target protein to improve the solubility and folding), cultivation conditions (usage of special chemical compound in the media, control the cultivate temperature and so on), expression strategies. It is especially hard for expression MPs in a soluble and correctly folded state. The hydrophobic characteristic of MPs results in non-soluble and aggregated protein during overexpression.

1.3.1 Conventional *in-vivo* expression systems

Selection of suitable expression system is crucial for the successful expression of proteins. It is always the first choice to look for the organism, in which certain target is naturally abundant. Then, this organism can be used to over produce certain abundant protein or related proteins. However, due to the fact that not all the organisms have well established expression strategies, it is often the case that one has to choose another easy handle host for heterogeneous expression.

Few elaborated, well-documented and studied system like *E. coli*, yeast, insect cells and mammalian cells are therefore prevalent for challenging protein production. There are various modifications of those expression systems, which give better production for difficult targets (e.g. MPs). Those well established and reliable protocols often give the first idea to overcome the difficulty of expression. For instance, *E. coli* is frequently the first and only choice for prokaryotic proteins, while eukaryotic proteins are often attempt to be produced in prokaryotic as well as in eukaryotic backgrounds.

Take MPs as example to handle challenging targets. Central factors that could determine the yield, integrity, activity and stability of synthesized MPs are the availability of highly processive transcription and translation machineries, suitable folding environments, the lipid composition of

cellular membranes, the presence of efficient targeting systems and appropriate pathways for posttranslational modifications (PTMs). Strong biases between the codon usage of the target gene and the host`s translation machinery could result in premature termination of the polypeptide chain or in mis-incorporation of amino acids, both having severe effects on sample quality and functionality of the final MPs [36, 37]. Most PTMs of recombinant eukaryotic MPs are unlikely to occur in prokaryotic expression hosts such as *E. coli* and *Lactococcus lactis*, which sometimes may hamper functional folding but in other cases could be beneficial for crystallization studies. In addition, bacterial MPs typically have a cytoplasmic location for both the N as well as the C-terminus [38]. The prevalent requirement of an N-terminus-out topology for eukaryotic MPs like G-protein-coupled receptors (GPCRs) could therefore cause problems in correct targeting and membrane insertion.

Specific lipids or hydrophobic environments can play important roles in the functional folding or stabilization of MPs [39, 40]. Many crystallized MP complexes show defined protein-lipid contacts. Especially sterol derivatives such as cholesterol appear to influence the function of distinct mammalian MPs. Also, folding and function of bacterial MPs can depend on the presence of specific lipids, as reported for the lactose permease in *E. coli* [41]. Most observed lipid effects on MPs are still difficult to interpret. They could result from altered membrane properties like changes in fluidity, bilayer thickness and lateral pressure, from positioning MPs in correct orientations within the membrane or from structural stabilization by direct lipid-MP interactions. While specificity could be shown in some cases, it is not clear for many identified interactions whether they are promiscuous and whether the lipids could be replaced by other structurally similar compounds. The lipid composition of host cells from the different expression systems varies considerably, and an overview is given by [40]. Bacterial membranes are generally devoid of sterols, sphingolipids and polyunsaturated side chains, which are prevalent in eukaryotic membranes. However, insect cells also have much lower cholesterol levels compared with mammalian cells, and they completely lack phosphatidyl-serine. The selection of an expression system or cellular background should therefore be of major concern for the lipid composition of the host cell membranes in addition to specific codon usages and prevalent glycosylation patterns. Expression of MPs in cellular backgrounds as closely related to their origin as possible thus appears to be the most reasonable strategy. However, the efficiency and workload of the individual expression systems are quite different, and preparative amounts of MPs are often only obtained with bacterial, yeast systems. Therefore, compromises must be made, if no specific requirements for the synthesis, folding or transport of an MP is evident, heterologous expression even in unrelated backgrounds might still provide the greatest success, especially in view of obtaining higher yields.

1.3.2 Cell-free systems

CF system is simply activating biological processes without a living cell. It is based on crude extracts, which provide low-cost catalyst and greater system compatibilities. CF biology has been used and developed for decades, mostly to study biological phenomena, and it was used first to decipher the genetic code 50 years ago [42]. For long time, the researchers were intimidated by the complexity of crude cell extracts containing hundreds of active biological catalysts. Recently, the 'black box' has been illuminated by number of examples indicating that the reaction networks can be understood, changed, and controlled. CF technologies simplify and harness cellular biology and also utilize the entire reactor volume instead of the 5 to 20% volume fraction (the intracellular volume) that is used in a typical fermentation process. Importantly, CF approaches provide direct access to the complex set of chemical reactions that take place in a living cell. The process then takes on the form and function of chemical processes employing homogeneous catalysis; processes that have been practiced and improved by chemical engineers for decades. Two most important applications of CF system draw great attention of biologist to explore and use this technique. They are CF protein synthesis (CFPS) and CF metabolic engineering (CFME).

The initial and well developed CF technology is CFPS. Extract from different source was used as the source of protein synthesis catalysts. In all cases, the cells are lyzed and the released content diluted. The dilution has effect on reducing the crowding effect in the cells. As a result, few or no background protein synthesis exists directed by the chromosomal DNA. The loosely associated sigma factors fall of the native RNA polymerase so that it is no longer functional to produce messenger RNA from the DNA remaining in the cell extract. For CFPS, however, monomolecular RNA polymerases such as those from the T7 and SP6 bacteriophages are added and function effectively with their respective promoters to produce only the desired mRNAs (and proteins).

The dilution also appears to be beneficial for protein folding. Translational protein elongation is slowed to about 10% of its normal rate because of the lower concentrations of the required elongation factors. This allows more time for newly synthesized protein domains to fold before the following domains emerge from the ribosome. Also, since only one type of protein is being produced into the diluted solution, there is less chance for nascent, unfolded proteins to become tangled with each other. While the dilution may also reduce the effectiveness of chaperones, these can be overexpressed in the source cell before lysis, if required. Notably, the catalytic system will also be stabilized as the dilution also reduces the concentrations of proteases and nucleases.

Amino acids are typically added to the lysate to serve as a major class of substrates, but protein synthesis is also one of the most energy demanding processes in typical cells. The entropic cost of assembling such precise macromolecules is high. This is another reason that expressing a single protein during CFPS is beneficial as chemical energy is not dissipated toward the synthesis of

undesired proteins. Most CFPS systems use a high-energy phosphate bond donor such as phosphoenolpyruvate (PEP), acetyl phosphate, or creatine phosphate, but newer systems activate normal aerobic catabolism to use glucose or glutamate as the energy source. The discovery that such complex metabolism could be activated and controlled in cell lysates has now sparked interest in using CF biology to produce important smaller biomolecules as well as proteins.

CFME is an emerging area that is trying to take full advantage of the reaction accessibility offered by the CF approach. CF metabolic engineering methods provide unprecedented opportunities to monitor and control the entire metabolic system for maximal production rates and conversion efficiencies.

1.3.2.1 Extract source for CF system

As mentioned above, extract act as the pool of catalysts for CF systems for protein synthesis. There are extract from prokaryotic and eukaryotic organisms. The most well developed and characterized were the *E. coli* S30 extract and wheat germ extract. However, other sources like insect cells [43], rabbit reticulocyte [44], yeast [45], parasitic protozoan *Leishmania tarentolae*[46], human HeLa cells[47], *Neurospora crassa* [48], plant cells (liquid callus cultures of Arabidopsis)[49], are also developed to produce different recombinant protein from different organisms. However, *E. coli* and wheat germ extracts are still the most popular systems due to the easy availability and standardized protocols. The well-studied *E. coli* S30 extract is obtained by centrifugation of lyzed cells at 30,000*g* and it includes most enzymes necessary for protein synthesis [50, 51]. During the preparation, endogenous mRNA as well as low molecular weight compounds is eliminated from the extract. The normal procedure for S30 extract preparation consists of the following steps: fermentation of *E. coli* cells; harvesting and washing the cells; cell disruption via high pressure; clarification of the lysate by centrifugation at 30,000g; run off step; final dialysis against buffers. However, there are modifications of several steps, like using low centrifugation speed to get the so call S12 extract [52]; different treatment of the run off steps or totally run an empty run off step[53]; using different genetically modified *E. coli* strans [54]; and modifying the fermentation conditions [55, 56]. In a practical view, CF expression should be performed according to these published protocols [57, 58]. In addition, commercial kits are available for CF expressions from several companies.

Besides the *E. coli* S30 extract, the wheat germ extract is also well studies and the preparation protocols are well established. Although the preparation and material availability are not as easy as *E. coli* extract, the advantage of eukaryotic background may offer better folding environment for certain eukaryotic targets. Generally, the wheat germ CF system is frequently operated as translation system with supplied mRNA as template, while the *E. coli* CF system is transcription and translation coupled system. The wheat germ CF reactions could be extended for even weeks with the final protein yields up to 10 mg/ml at optimal conditions [59]. Wheat germ

extracts are more difficult to prepare as the embryo is surrounded by the endosperm containing high levels of protein and nucleic acids degrading enzymes as well as translation inhibitors. In addition, high batch variations in extract quality of individually prepared extracts exist that depend on the origin of the wheat germs. However, commercial suppliers of wheat germ extracts or expression systems are available such as Cell Free Sciences Co. Ltd (Japan) or Roche Diagnostics.

Further extract sources or alternative CF expression systems exist but only few have been used so far for the production of MPs. The PURE (Protein synthesis using recombinant elements) system represents a nearly complete reconstituted translation machinery of all aminoacyl-tRNA-synthetases and translation factors that have been separately produced and purified from conventional *in vivo E. coli* expression [60, 61]. The ribosomes are still isolated from E. coli extract, but with a significantly higher purity after centrifugation at 100,000*g*. The PURE system allows CF synthesis under better defined conditions which can be advantageous for the study of folding pathways or translation kinetics. However, expression yields are lower if compared with other CF expression systems and it is only used for the analytical scale production of proteins.

1.3.2.2 Template design

In CF expression system, the template DNA has to be provided under control of specific and strong promoter sequence, most commonly recognized by phage polymerase like T7, SP6 or T3 RNA polymerase. Several suitable vectors like pET (Novagen) or pIVEX (Roche Diagnostics) derivatives which contain the T7 promoter region are frequently used. The design of template in CF expression system to solve the expression of difficult targets seems very important. Frequent rare codons as well as unfavorable initiation of translation has been identified as major limiting steps in the protein synthesis [62, 63]. Secondary structures formation involving critical areas at the 5-prime end of the mRNA such as the ribosomal binding site could slow down or even prevent the initiation process [64-66]. Although algorithms have been established to optimize the DNA sequences to optimize the secondary structure of mRNA, it is still difficult to make reliable and efficient prediction of the mRNA structures due to its complexity. A high frequency of rare codons could cause the pause of translation, mis-incorporation of amino acids or even premature termination, but such problems could be solved by using codon optimized synthetic genes.

Besides the traditional vectors which can be used as template DNA in CF expression systems, a PCR linear template, which contains the basic sequence region like T7-promotor and terminator, can also be used as efficient template. Recently, several researches were done to develop a practical method for preparing the linear PCR template for CF expression. A two-step PCR method developed by Yabuki, *et. al*, can provide robust, high-throughput production of linear template for CF protein synthesis. The procedure consist two step PCR: firstly, a target region

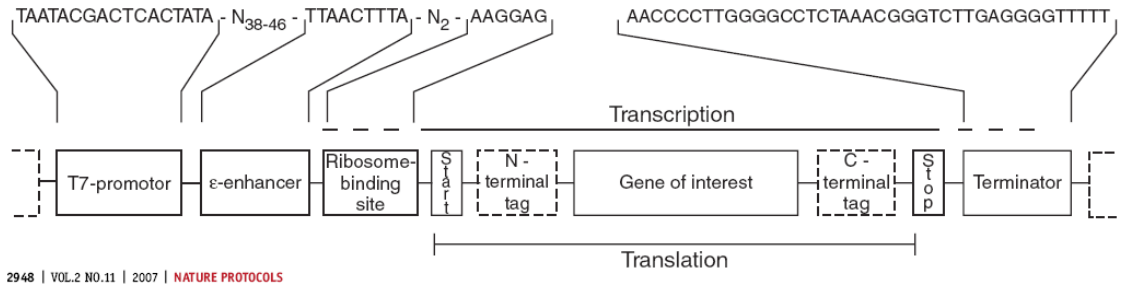


Figure 3. DNA template for CF expression. Essential T7 regulatory sequences are shaded and the corresponding sequences are indicated. Optional translated elements are dashed [58].

within the coding sequence is amplified using two gene-specific forward and reverse primers, which contain the linker sequences and the 5-prime sequences of the target region; secondly, PCR concatenates the first PCR product with the 5-prime and 3-prime double-stranded fragments, which contain the linker sequences as well as the sequences for the tag(s) and the initiation and termination, respectively, for T7 transcription and ribosomal translation, and amplifies it with the universal primer. Proteins can be fused with a variety of tags, such as natural poly-histidine, glutathione-S-transferase, maltose-binding protein, or streptavidin-binding peptide [67].

A strategy, so called 'tag variation', which focusing on the potential problems with translation initiation [68, 69], has been developed. This fast tag variation strategy was designed by introducing non-functional tags of minimal size which located right at the 5-prime end of the target gene. This extra short sequence functions as translation enhancer. With combination of using linear PCR product, a fast and efficient protocol from the sequence design to the final protein expression yield determination was established. This strategy can be classified into several sequential steps: (I) Preparing DNA templates by overlap PCR for each tag – target combination, (II) DNA template screening in analytical scale continuous exchange CF (CECF) reactions, (III) the potential further optimization of the most efficient DNA template and (IV) final CECF protocol optimization with the optimized DNA template for preparative scale production. With this strategy difficult MPs targets including GPCRs are used to exemplify the application. Results showed great improvement of expression yield was observed for most of the targets [70].

1.3.2.3 Compounds in CF system

Besides the extracts, CF expression reactions consist of various essential compounds with defined and sometime critical concentration optima (Fig. 4). Divalent Mg^{2+} ions are essential for many biological reactions and the free Mg^{2+} concentration in CF extracts is a very hard to determine value. In order to obtain the maximum expression yield, it is therefore advisable to optimize Mg^{2+} concentration for each new batch of prepared extract or even freshly prepared stock solution. With different CF systems the optimum concentration range of basic compounds like ions, can be quite different. For *E. coli* systems, the total Mg^{2+} ion concentrations typically range from 8-20 mM,

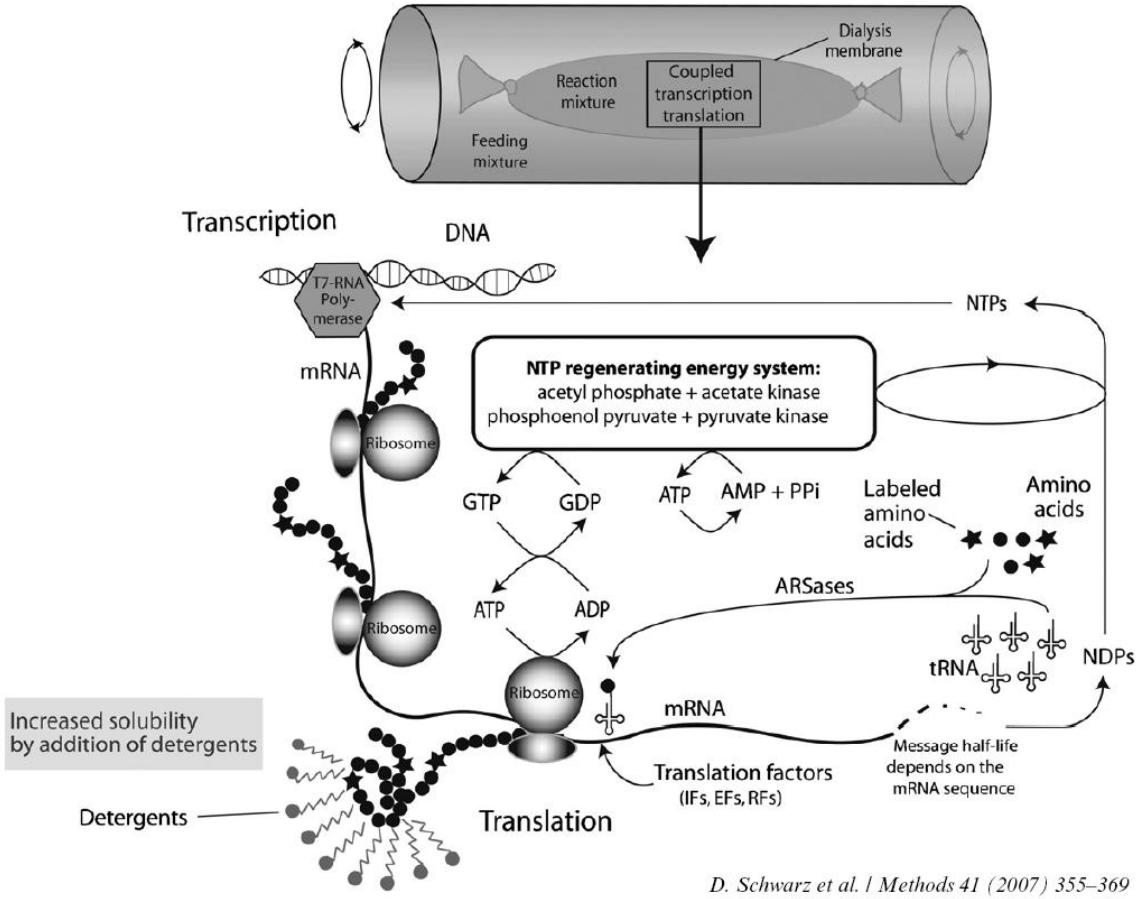


Figure 4. Schematic configuration of a coupled transcription/translation reaction in a CECF system. The CF reaction can be carried out in simple dialysis tubes placed into suitable plastic vials that hold the FM. The complete set-up is incubated e.g., on a turning device that ensures continuous agitation of the reaction and substance exchange between the two compartments.

while eukaryotic translation systems include only 2-4 mM Mg^{2+} [71]. Potassium and ammonium acetate or the corresponding glutamates can be added into *E. coli* systems at high concentrations exceeding 200 mM. Wheat germ systems often contain 100 mM potassium acetate [71, 72]. ATP and GTP are sufficient to provide the required energy for mRNA translation in wheat germ systems, while four nucleoside triphosphates ATP/CTP/GTP/UTP are necessary in the coupled transcription/translation reaction type of *E. coli* CF expression systems. Amino acids are supplied in concentrations between 0.3 mM and 2 mM while even higher amounts of unstable amino acids (e.g. R, C, W, M, D, E) could significantly increase expression. MPs often have a strong bias in amino acid composition and adjusting the amino acid pool of the reaction accordingly could further increase the expression yields. Problems with the codon usage upon expression of heterologous genes could be addressed by modulating the total tRNA concentration or by addition of rare codon enriched tRNA. An alternative option would be to use synthetic genes designed according to the preferred codon usage of the chosen CF expression system.

NTPs are essential energy sources, which are rapidly consumed in CF expression reactions. In addition, bacterial cell extracts have high endogenous ATPase and phosphatase activities which

will cause further uncoupled NTP hydrolysis. Sufficient energy supply is therefore the main bottleneck for CF protein synthesis and a number of modified protocols for *E. coli* systems have been developed as a major variation use different energy sources and pathways for the efficient energy regeneration. In conventional protocols, high-energy phosphate donors like phosphoenol pyruvate, acetyl phosphate or creatine phosphate are supplied as secondary energy substrate together with the corresponding kinases [73, 74]. However, the excess of phosphate generated as byproduct during the breakdown of energy sources and nucleotides is identified as an inhibitor of translation. In particular in batch configurations, the strategy of energy regeneration is of importance as accumulation of inhibitory phosphate should be minimized. However, those secondary energy substrate with high-energy phosphate bond are normaly quite expensive. Researches which designed to use other less expressive energy sources have been done to improve the application of this CF technique. Proposed modified energy regeneration pathways for batch systems use the oxidation of substrates from the glycolytic pathway (e.g. pyruvate, glucose-6-phosphate, glucose) for the generation of ATP concomitant with the consumption of reaction by-products [75]. The "PANOx" system combines phosphoenol pyruvate as the conventional energy source with the addition of nicotinamide adenine dinucleotide and coenzyme A in order to use in addition pyruvate for ATP regeneration [76]. Protocols using alternative nonphosphorylated energy sources like pyruvate in the "Cytomim" system have further been developed [77]. Oxidative phosphorylation, the most efficient natural source of ATP, is used and potassium and magnesium glutamate as well as the polycations spermidine and putrescine should also more closely mimic a cytoplasmic composition of the reaction. Having efficient energy regenerating systems, the relatively expensive NTPs could also be replaced by NMPs in order to generate a more cost-effective system [75].

1.3.2.3 Configurations of CF system

When the CF technique was first developed, the least complicated configuration was used. There was only one compartment contain all the necessary components of CF reaction. This kind of CF system configurations could allow continuous protein synthesis for a few hours or less depending on the speed of precursor consumption and the accumulation of inhibitory by-products. The final expression yield was relatively low. Research was done to improve and prolong the reaction time by introducing different metabolic path way of certain compounds. The results show that with optimization it was possible to obtain close to 1 mg of recombinant protein per 1 ml of reaction [78].

A remarkable strategy to optimize the efficiency of CF reactions was the design of continuous exchange CF (CECF) systems [79]. The principle is to separate the reaction into two compartments that are separated by a semipermeable membrane (Fig. 4). One compartment holds the reaction mixture (RM) with all high molecular weight compounds like the CF extract, enzymes

and nucleic acids. The second compartment provides a reservoir of low molecular weight precursors like NTPs, energy sources and amino acids and represents the feeding mixture (FM). The molecular weight cut-off of the separating membrane is usually relatively high and in between 10-50 kDa. It has to be emphasized that even at higher pore sizes no significant leakage of essential proteins from the translation machinery of the extract are leaking from the RM into the FM as they form a large complex through multiple interactions. Synthesized MPs also stay in the RM as the insertion into detergent micelles usually increases the molecular weight of even small proteins above 50 kDa. Efficient substance exchange between the two compartments through the membrane is ensured by vigorous stirring or shaking during incubation. The reaction time is significantly extended by the continuous supply of fresh precursors from the FM into the RM concomitant with the continuous dilution of inhibitory by-products from the RM into the FM. The ratio of the volumes of RM : FM is usually in between 1 : 10 and 1 : 50 and it is an important parameter for the final yield of recombinant protein. Higher FM volumes will certainly increase product yields, but not in a linear correlation. The exchange rate of compounds between the two compartments as well as the capacity of the CF extract will become more and more limiting and reduce the gain of yield. Considering the relatively high costs of the precursors in the FM, a ratio of approx. 1: 20 may be a reasonable compromise for maintaining CF expression for 6-8 hrs. with final yields of up to several mg recombinant protein per one ml of RM.

Since CECF reactions have to be equipped with a membrane, their set up and handling is relatively complicated and limits their use in throughput applications. A bilayer system has recently been proposed as an interesting alternative configuration in particular for throughput purposes [80]. The bilayer system resembles a simplified version of the CECF system, in which the RM is carefully overlaid by the FM without separation by a membrane. Substances between the two layers exchange simply by diffusion and translation can be prolonged for several hrs yielding sub-milligram amounts of protein which is sufficient for functional analysis. The reaction can be set up in micro-plates and a ratio of RM: FM of approx. 1: 5 is recommended for best performance.

1.3.2.4 Expression mode of membrane proteins

MPs can be synthesized in CF system in different ways. Without any additional hydrophobic reagent, the produced MP will result in the precipitant; with the addition of detergent in the CF reaction mixtures, the produced MPs will stay in soluble form. Recently, via adding lipids instead of detergent, MP can also be produced in their functional form or even soluble form. Both supplying with detergents and lipids are not possible with any other expression system, only with CF system as an open expression system that can create such artificial hydrophobic environment for MPs. Different expression modes are shown in Fig. 5.

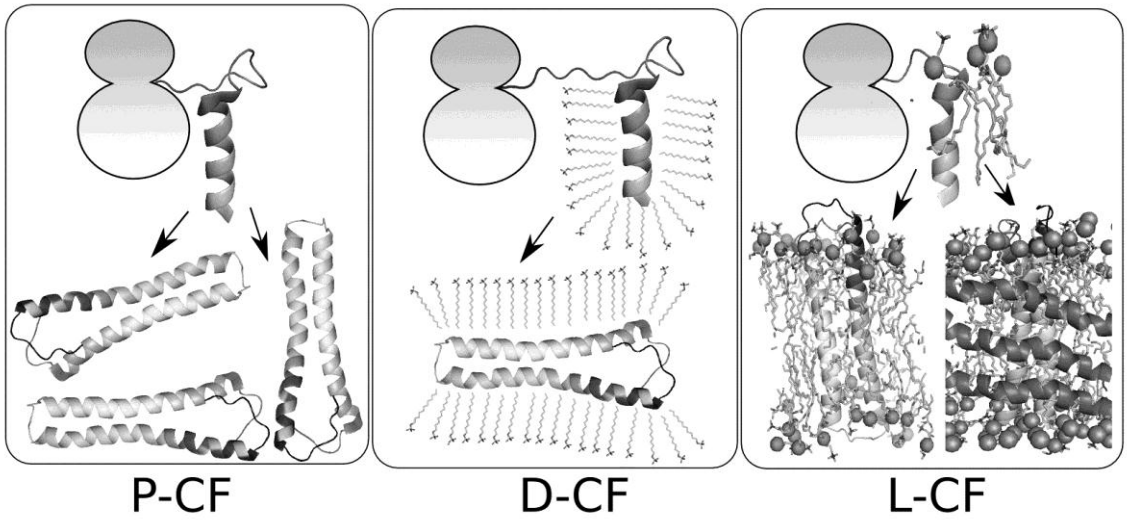


Figure 5. Expression mode of CF production of MPs. In the precipitate mode (P-CF), no hydrophobic environments are provided, and the MPs precipitate after translation. In the detergent-based mode (D-CF), the synthesized MPs can stay soluble by insertion into the provided detergent micelles. In the lipid-based mode (L-CF), the synthesized MPs have the possibility to integrate into supplemented liposomes.

1.3.2.4.1 P-CF mode

In this standard and most simple CF expression mode, so called P-CF mode, MPs will be produced without any artificial hydrophobic environment. Therefore, all the MPs produced will result in the pellet form. This pellet is mainly the precipitated MPs, however, it is quite different from the precipitant when expressed *in vivo*. There are several differences: (I) Precipitates are always formed during the successful P-CF expression of MPs, while inclusion body formation is a much rarer event mostly observed with outer MPs. (II) P-CF generated precipitates can efficiently be solubilized with detergents by gentle shaking for few hours [81]. In contrast, inclusion bodies usually have to be solubilized with strong denaturants like high concentrations of urea or guanidinium hydrochloride. The subsequent refolding procedures require excessive dialysis steps and buffer exchanges in high dilutions. (III) The solubilization of P-CF generated MP precipitates can already result in functionally folded MPs as shown for example for the multidrug transporter EmrE [81, 82] or for the human histamine-1 receptor [83]. The successful refolding of MPs from inclusion bodies is so far mostly confined to outer MPs, whereas the functional refolding of α-helical MPs remains the very exception.

Precipitates obtained from P-CF reactions appear therefore to consist of less denatured or aggregated MPs if compared with inclusion bodies. This assumption is further supported by the recent observation of folded proteorhodopsin in P-CF generated precipitates. Structural features might therefore persist that could support the folding process of distinct MPs upon solubilization. Detergents that are highly efficient in the solubilization of P-CF generated precipitates and that still might allow a functional folding include DPC or the lyso-phosphoglycerols LMPG and LPPG. In addition, these detergents provide suitable environments for structural analysis of MPs by nuclear magnetic resonance spectroscopy (NMR). However, the functional solubilization of MPs from

precipitates is not a general process. Activities of P-CF produced and solublized samples of the nucleoside transporter Tsx and of the human endothelin-B receptor could not be observed [84, 85].

1.3.2.4.2 D-CF mode

When the P-CF expressed MPs did not function, there is another way, which directly introduce reagents that can create a hydrophobic environment in the CF expression system. Therefore, the freshly produce MPs can get access to the hydrophobic environment, which can keep the MPs in a correctly folded soluble form. Detergents, one of the reagents which usually used as hydrophobic reagent, form micelles above certain concentration in solution. These micelles will provide the hydrophobic environment for MPs. The formation of MPs and detergents micelles complex, proteomicelles, is mainly based on non-specific hydrophobic interactions and do not require a translocation process, which *in vivo* will help the MPs insert into the membrane. Therefore, D-CF expression mode does not have the problems which originate from the transport and translocation process that often restrict the efficient production of MPs in the cell based conventional systems.

Not all the detergents can be used in CF systems some with relatively high critical micellar concentration (CMC) like n-octyl-ß-D-glucopyranosid (β-OG) or 3-[(3-cholamidopropyl)dimethylammonio]-1-propansulfonat (CHAPS) start to inhibit the expression machinery of cell extracts already at concentrations hardly exceeding their CMCs, while the most of commonly used detergents are tolerated by CF systems in concentrations of at least several times CMC [84]. Especially, the relative mild detergents from the group of long chain polyoxyethylene-ethers (Brij-derivatives) can be supplied into CF systems at final concentrations of more than 100 x CMC. Other detergents like Triton X-100 or DDM are also effective for the D-CF expression and solubilization of MPs, but the presence of higher concentrations starts to reduce the performance of the expression reaction. Further options of detergents include alkyl-glucosides like n-dodecyl-β-D-maltoside (DDM) or steroid derivatives like digitonin. Digitonin and Brij-derivatives appear to be exceptionally efficient for the solubilization of a diverse variety of MPs in the D-CF expression mode. With both detergent groups, preparative scale amounts of solubilized samples of the multidrug transporter EmrE, the nucleoside transporter Tsx and the rat vasopressin-2 receptor could be produced in D-CF reactions [84]. Type and final concentration of the supplied detergent can have a strong impact on the quality and the yield of solubilized MP and these parameters should be subject of optimization in each D-CF expression approach [40, 84, 86]. The addition of detergent mixtures for the formation of mixed micelles could be a further interesting alternative. Lipids are discussed as important structural elements for the folding or stabilization of certain MPs [40] and providing detergent/lipid combinations in D-CF reactions could further be considered in order to optimize the quality of the produced MP.

Emphasize should be given to the extremely mild detergents, like the mentioned Brij-derivatives, which could increasingly play a role in MP analysis. Those detergents are not able to disintegrate membranes and therefore have not been used so far for the solubilization of MPs by the conventional membrane extraction approach after MP production in cellular expression systems. However, their mildness is now a valuable property for their combination with CF synthesis as they are highly efficient in solubilizing MPs but do not interfere with protein synthesis. A similar effect might have fluorinated or hemifluorinated surfactants that are strongly hydrophobic but not lipophilic [87]. These amphipols probably do not compete effectively with protein-protein or protein-lipid interactions and could therefore stabilize the native fold of MPs under particularly mild conditions. The mechanosensitive channel MscL had similar electrophysiological activity after CF synthesis in presence of perfluorocarbons as after expression *in vivo* [87].

1.3.2.4.3 L-CF mode

Lipids are the natural environment of MPs and in many cases essential for folding, stability and function. Defined lipid environment are often beneficial for functional and structural studies. In cells, it is known that MPs needs different mechanisms to be transported or targeted into the lipid bilayer. These processes require specific enzyme machineries, for instance, the Sec and SRP pathways for MP insertion [88, 89]. Although the *in vitro* reconstitution protocol was well established and most case efficient [81-83, 90], there are also some cases that lipids are necessary co-translational to obtain functional MPs. CF can provide such an system which lipids can be directly introduced. Recent reports have shown an effect of the charge of the employed lipids towards transcription and translation efficiency in the CF system. While transcription is usually not affected by the presence of any liposomes, translation may be inhibited by cationic lipids. Lipid concentration may furthermore influence the expression level [91, 92].

Lipids can be supplied into CF reactions in different formulations such as (i) preformed liposomes of defined compositions, (ii) isolated fractions of cell membranes, (iii) detergent solubilized lipomicelles providing a mixed environment for MPs, (iv) bicelles consisting of planar bilayers surrounded by detergents, and (v) nano-lipid particles or nanodiscs providing highly soluble and confined bilayer areas. Mechanisms of co-translational MP translocation could thus be analyzed in dependency of a large variety of lipid mixtures. Furthermore, MP integration into the provided bilayers may be more directed during L-CF mode expression, compared with the random orientation in classical posttranslational reconstitution approaches. In addition, the synthesized targets will be the only MP inserted into the bilayers and measured activities could thus clearly be attributed. These characteristics can have considerable benefits for subsequent functional studies.

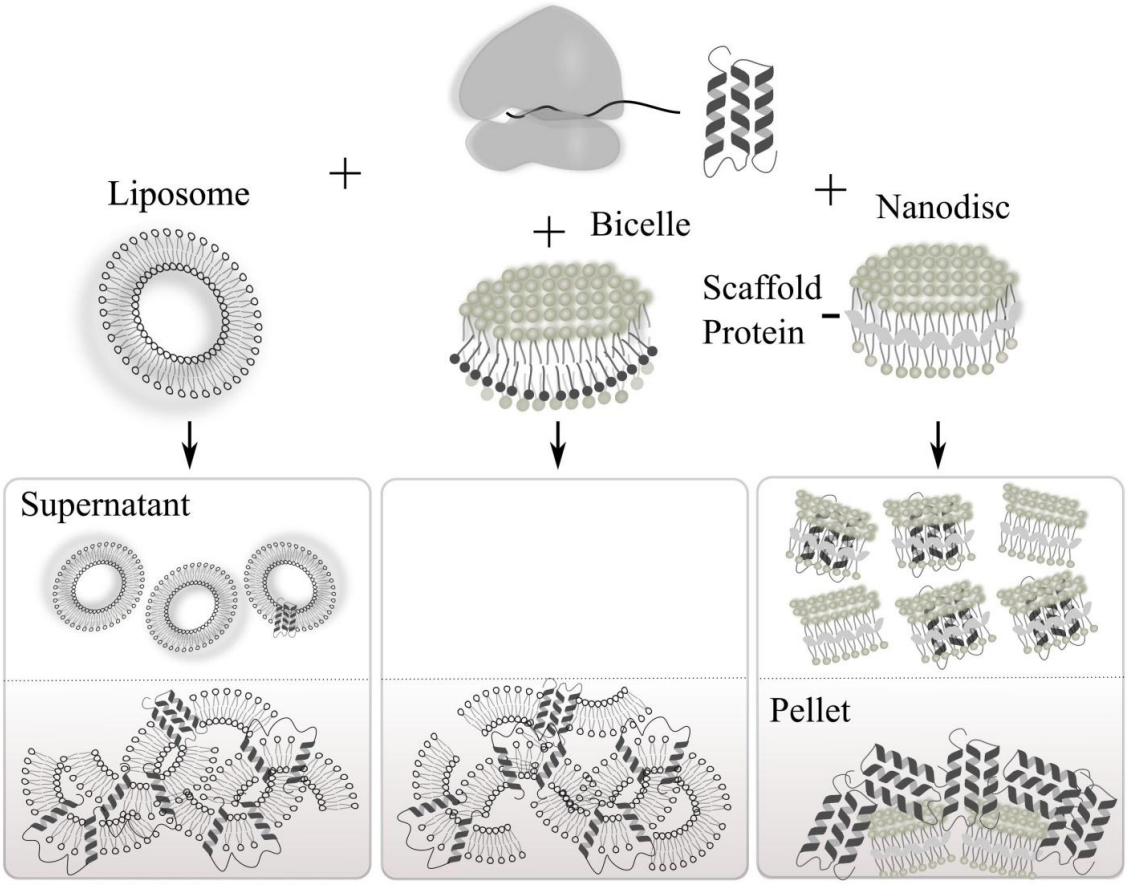


Figure 6. L-CF modes for MP production. Lipids are provided in three basic forms, either as liposomes, lipid/detergent mixtures or nanodiscs. Illustrated is the recovery of the lipids together with the synthesized MPs after incubation in either supernatant or precipitated fraction of the reaction.

A number of recent researches showed that functional MPs can be produced functionally. For example, bacteriorrhodopsin could be produced as membrane integrated protein after addition of artificial liposomes of defined composition in *E. coli* extracts [93]. A plant transporter and the membrane associated protein apo cytochrome b5 could be synthesized into preformed liposomes by using wheat germ systems [94, 95]. However, in all cases again only relatively low amounts of MP could be synthesized. In a very recent report, two *E. coli* transporters could be CF synthesized into purified *E. coli* inner membrane vesicles in amounts of several 100 µg per ml of RM [96]. A batch system was used and the activity of the L-CF produced tetracycline pump TetA was verified by a specific transport assay. The calculated surface coverage of the vesicles with up to 40% of recombinant protein was significantly higher if compared with data obtained from cellular expression systems. This work demonstrated that CF expression of MPs in the L-CF mode can even approach preparative scales and this technique might play a significant role in the next future for the functional analysis of MPs as well as for studying their specific translocation mechanism.

1.4 Cell-free in proteomic research

With the completion of the genome sequences of many organisms, there is an increasing need for a high throughput expression of genome encoded proteins. However, the traditional *in vivo* cell based heterologous protein expression systems are not able to address this challenge because of the time consuming cloning procedures and later on sample preparations. However, CF technique, emerging as an alternative approach to cell-based methods, offers a simple, open and flexible system for fast synthesis of functional proteins. The time required from DNA sequence to the functional protein is dramatically shortened in CF expression system.

1.4.1 Cell-free system for throughput application

1.4.1.1 Large-scale analysis of proteins

CF expression system has been widely used to determine the presence of a gene or open reading frame (ORF), especially when multiple samples or large genes are to be examined [97]. It can also be used to confirm an ORF predicted by DNA sequencing. In each case, the synthesized proteins are usually analyzed for size, biophysical properties and function [98]. CF protein synthesis has also been used as a routine screen for translation-terminating mutations in the diagnosis of genetic disease [99], a rapid method referred to as a protein truncation test (PTT). The truncated mutant alleles can be easily distinguished from the normal full-length protein product by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

The successful rates in throughput expression screens are higher than *in vivo* expression systems, especially for difficult proteins like MPs. In addition, many critical and highly variable parameters like efficient gene delivery, template stability in cells, and suppression of background expression and induction of expression as well as considerations about optimal cell physiologies are eliminated. Furthermore, samples can be taken at any time point without disrupting the reaction. The possibility to use linear DNA fragments as template for CF expression is a further benefit for throughput applications. Products from appropriately designed PCR reactions including suitable promoter sequences could be used directly for expression analysis and time consuming sub-cloning procedures are not necessary [100, 101]. With the help of throughput CF expression system, not only screening large number of different targets at similar conditions is possible, but also large number of targets specific expression conditions screening is becoming feasible and efficient. The open nature of the CF system allowed target specific expression environment become possible. For instance, different chaperons and chemical compounds, which had effect on the protein folding, stability, can be added into the CF reaction system.

Large number of expression screenings was performed with CF technique, including *E. coli* inner MPs, human proteins and so on [102, 103]. In the study of *E. coli* inner MPs, bioinformatics

analysis showed that about 1000 of the 4288 predicted genes encode integral inner MPs [38]. For a number of 737 prevalent α-helical topologies contain at least two trans-membrane segments (TMS). 134 representative targets were selected and studied by CF expression system, covering the diverse MP families, protein sizes up to 1000 amino acids and up to 15 predicted TMSs. 112 targets, consisting of 84% of the total 134 targets, could be expressed either in P-CF or D-CF mode. This successful rate is particular high if compared with the cellular systems where usually the expression of not more than 30% of the targets is monitored [104-106]. Not only difficult prokaryotic proteins, but also eukaryotic proteins can be successfully expressed and studied in CF systems. Both *E. coli* CF system and wheat germ CF system were selected for large scale of eukaryotic proteome research [107, 108].

1.4.2 Cell-free system for generation of protein microarrays

In the post-genomic era, high-throughput protein expression platforms are becoming increasingly important. Protein microarrays, which have gained popularity following the success of DNA microarrays, are miniaturized arrays containing small amounts of immobilized proteins. In microarrays [109], proteins are typically printed on glass slides, polyacrylamide gel pad slides, or microwells, each having its own advantages and drawbacks [110, 111]. This protein array approach allows studying hundreds or even thousands of proteins simultaneously in a single experiment [111]. There is great potential of protein microarrays in being flexible tools for large-scale analysis of functional proteins [112], such as antibody profiling [113], biomarker identification[114], protein-protein interactions[115], and enzyme assays [116].

Traditional cellular approaches for the generation of protein microarrays involve expression of proteins in suitable heterologous system (e.g. *E. coli*, yeast, or mammalian cells), purification of expressed proteins, and immobilization of the purified proteins on to a solid support. With this cell-based methods, laborious and time-consuming procedure that poses other technical challenges such as protein purity, maintenance of protein folding, and functionality during the purification and immobilization steps are often the bottleneck.

CF expression system, as an alternative expression platform, has big advantage in developing a high-throughput protein express method. Using linear PCR template or circular plasmid for gene expression; short expression time and low reaction volume, all make the CF system an ideal tool for protein microarray. To circumvent the obstacles posed by traditional cell-based methods, CF expression systems are now largely being adopted for the generation of protein microarrays. An optimized technology for this purpose must meet certain criteria: (i) It should be able to utilize a wide variety of DNA templates, in the form of PCR products or plasmids. (ii) The process should be simple, quick, and cost effective. (iii) Protein arrays should be produced on demand to avoid storage effects. (iv) It should allow for simultaneous production of thousands of

proteins in a single reaction. (v) Methods for the detection and analysis of the bound protein must be simple.

1.4.2.1 Protein *in situ* array (PISA)

PISA (protein in situ array, shown in Fig. 7) was the first well known cell free based in situ protein microarray technology that provided a rapid, single-step approach for the generation of a protein array from DNA fragments coupled with CF expression system followed by immobilization onto a solid support [117]. In this technique, the DNA constructs encoding the protein of interest contain a T7 promoter and sequences for translation initiation (Shine-Dalgarno or Kozak), an N- or C-terminal tag sequence for immobilization along with suitable termination sequences. The DNA constructs are produced by PCR amplification using a high fidelity TAQ polymerase. The surface is precoated with a tag capturing agent and the protein expression is carried out, commonly with *E. coli* S30 or rabbit reticulocyte lysates. After transcription/ translation, the expressed proteins bind specifically onto the surface through the tag sequence and the unbound material could be washed away. The authors of this technology made use of 6x histidine tag and a microtiter plate having 24 wells coated with nickel nitrilo-triacetic acid (Ni-NTA). They successfully achieved the expression and functional immobilization of a fragment of the human antiprogesterone antibody in microtiter wells and that of luciferase enzyme on Ni-NTA-coated magnetic beads. Quantification by Western blot analysis unveiled that 120 ng of protein was produced from the standard reaction volume of 25 µl, having 0.5-1 µg of PCR DNA, of which 50% was successfully immobilized in the microtiter wells [118].

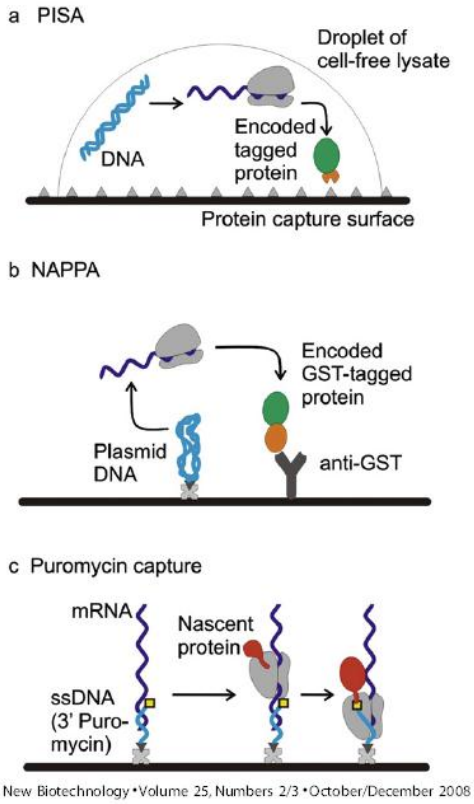


Figure 7. Methods for making protein arrays by in situ protein synthesis. (a) Protein *in situ* array (PISA) [118]. (b) Nucleic Acid Programmable Protein Array (NAPPA) [119]. (c) Puromycin capture [120]. ssDNA: single-stranded DNA[1].

PISA was the first technique and its success opened the door for other similar technologies in cell free based protein *in situ* arrays. However, the short comings of this technique included the need to utilize the PCR-generated DNA immediately after synthesis, relative high volume of CF lysate required, as well as a possible loss of protein function if the tag sequence interfered with the folding.

1.4.2.2 Nucleic acid programmable protein array (NAPPA)

NAPPA was known as the first method to proof the possibility of direct conversion of DNA arrays into protein arrays by in situ CF protein synthesis. Unlike the method of PISA, in this method, DNA was first spotted onto a glass slide together with a protein-capturing reagent (an antibody in this case). The resultant DNA array was then covered with a CF transcription/ translation system on the array surface to express the proteins, which became trapped by the antibody colocalised in each spot. LaBaer and colleagues developed a method that used immobilized DNA templates, encoding the desired protein as a GST fusion, for *in vitro* transcription and translation (IVTT). In NAPPA, the biotinylated plasmid DNA was bound through avidin onto an aminopropyltriethoxysilane (APTES)-coated slide, which also contained anti-GST antibodies bound to it for protein capture. Protein expression was then carried out by adding RRL to the arrays. The newly synthesized proteins, ranging in signal from 4 to 29 fmols, were captured by the antibody through the GST tag, and generated an array in which the protein was co-localized with its DNA[119].

This NAPPA method was further modified to generate high-density protein microarrays [121]. High yields of high quality DNA were obtained for immobilization by using a diamine-derivatized resin. They also found that BSA improved the binding efficiency of DNA. A test array of cDNAs for 96 genes along with a negative control of non-expressing plasmid DNA and a concentration series of recombinant, purified DNA were printed. Around 97% of the printed samples were found to be detected by using the PicoGreen stain for ds-DNA, whereas anti-GST antibody detected protein signals for 99% of these 96 genes. This technique was effectively used to array up to 1000 unique human cDNAs, with an average protein yield of 9 fmol per feature.

1.4.2.3 DNA array to protein array (DAPA)

This innovative technique, developed by He *et al*., makes it possible to repeatedly use the same DNA template slide for printing multiple protein arrays. Different from NAPPA, the DAPA procedure produces pure protein arrays on a separated surface as shown in Fig.8. The idea of DAPA is that by carrying out CF protein synthesis within a membrane filter sandwiched between two glass slides, on one of which is an array of immobilized PCR DNA molecules while the other is coated with a reagent to capture translated proteins [1]. Individual proteins are produced in parallel within the membrane. Then the synthesized proteins diffuse through the membrane filter and are immobilized on the opposite surface via interaction with the protein-capturing reagent, which form a protein array with the same pattern as the DNA array [1].

Data showed that more than 20 copies of a protein array can be produced by use of a single DNA template, repeatedly [122], which made it useful for repeated printing of protein arrays in laboratories without access to microarray spotters. But, there is also some limitation for this

technique. That is the protein diffusion during membrane penetration, especially with regard to larger multi-meric proteins.

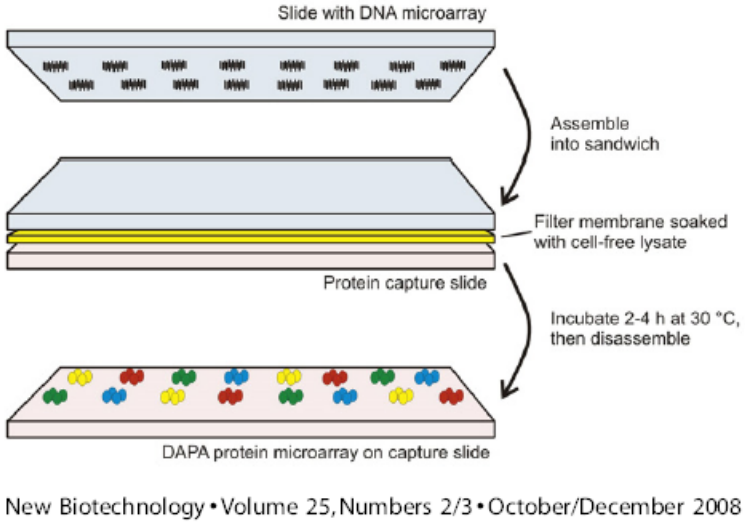


Figure 8. Principle of DAPA technology[1].

1.4.2.4 Arraying of nascent proteins by in situ puromycin capture

New research was done to make an adjustment of mRNA display [123]. This adaption using in situ puromycin capture of nascent polypeptides to facilitate protein arrays [120]. This approach requires additional manipulation procedures to make mRNA–ssDNA hybrid molecules before printing and protein yields are limited by the amount of mRNA spotted. Similar method, exemplified by solid phase immobilisation on beads, has also been reported [1, 124].

2. Materials

2.1 Laboratory equipment

0.22-µm polysulfone filters (Roth)
10-liter fermenter (B.Braun)
96F Nunclon Delta Black Microwell SI (Nunc, ON 137101)
Äkta purifier (GE Healthcare Amersham)
Autoclave (Getinge)
Balance (Sartorius)
Centrifuge
Centriprep devices YM-10 (Amicon)
Cooled table top centrifuge
Dialysis tubes type 27/32 MWCO 14 kDa (Spectrum, Rancho Dominguez)
DispoDialyser 25 kDa MWCO, regenerated cellulose (Spectrum, Rancho Dominguez)
French press / Cell disrupter
Gel imager (Biometra)
Glass vials (e.g., Rotilabo vials; Roth)
Heating block (VWR)
ITHACO electrometer (ITHACO, NY, USA)
Lumi Imager P1 (Roche Diagnostics)
MicroDispoDialyser 25 kDa MWCO, regenerated cellulose (e.g., Roth)
Microwave (Alaska)
Midi DNA preparation kit (e.g., Qiagen)
Mini-extruder (Avanti lipids)
Nanodrop 100 (peqlab)
Ni-NTA Superflow resin (Qiagen)
pH-meter PHM 210 (Radiometer)
Pipettes (Abimed, Gilson)
Plate Centrifuge Rotana 46RSC-Robotic (Hettich)
PS-microplate 96well V-shape ON 651101 (Greiner bio-one)
QIAquick gel extraction kit (Qiagen)
QIAquick PCR purification kit (Qiagen)
Rolling device (e.g., Fröbel Labortechnik)
SDS-polyacrylamide gel electrophoresis system
SEC columns: Superdex200 3.2/30
Shaker (Heidolph)

Shaking incubator (New Brunswick)

Slide-A-lyzer dialysis cassette 10 kDa MWCO (Pierce)

Sonifier Labsonic U (B.Braun)

TECAN cooled carrier

TECAN FreedomEVO 200/8

TECAN GENiosPro plate reader

TECAN shaked plate incubator / Thermoshaker

TECAN Te-mags

Temperature-controlled shaking incubator

Thermocycler (Bio-Rad, Eppendorf)

UV/Vis spectrometer (Cary)

Vacuum pump (Abm)

Vortexer Reax 2000 (Heidolph)

Waterbath

Western blotting system (BioRad)

2.2 Reagents

1,4-Dihydrobutane dihydrochloride (Putrescine) (Sigma)

1,4-Dithiothreitol (Roth)

2-Mercaptoethanol (Sigma Aldrich)

2x yeast tryptophan peptone glucose (YTPG) medium (see Buffers)

Acetyl phosphate lithium potassium salt (Sigma Aldrich)

Adenosine 5-triphosphate disodium salt trihydrate (Roche Diagnostics)

Amino acids (Sigma Aldrich)

Antibodies: anti-biotin peroxidase conjugate (Sigma); anti-gree fluorescent protein from rabbit (Calbiochem); anti-mouse IgG HRP conjugate from goat (Sigma Aldrich); anti-penta His IgG from mouse (Qiagen); goat anti-rabbit IgG (Calbiochem); T7 tag antibody HRP conjugate (Novagen)

Antifoam Y-30 emulsion (Sigma Aldrich)

Bactotryptone (Roth)

Benchmark protein ladder (Invitrogen)

Coenzyme A sodium salt hydrate (Sigma)

Complete protease inhibitor mix (Roche Diagnostics)

Cytidine 5`triphosphate disodium salt hydrate (Fluka Sigma Aldrich)

Detergents (Sigma Aldrich, Anatrace, Avanti Polar Lipids, Glycon): Brij-35, polyoxyethylene-(23)-lauryl-ether; Brij-56, polyoxyethylene-(10)-cetyl-ether; Brij-58, polyoxyethylene-(20)-cetyl-ether; Brij-72, polyoxyethylene-(2)-stearylether; Brij-78, polyoxyethylene-(20)-stearyl-ether; Brij-97,

polyoxyethylene-(10)-oleyl-ether; Brij-98, polyoxyethylene-(20)-oleyl-ether; C10E5, Pentaethylene Glycol Monodecyl Ether; C8E5, Pentaethylene Glycol Monooctyl Ether; C12E8, Octaethylene glycol monododecyl ether; β-OG, n-octyl-β-glucopyranoside; FG, n-Heptyl-β-D-Glucopyranoside; NG, n-Nonyl-β-D-Glucopyranoside; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propansulfonat; DHPC, 1,2-diheptanoyl-sn-glycero-3-phosphocholine; diC6PC, 1,2-dihexanoyl-sn-glycero-3-phosphocholine; diC8PC, 1,2-dioctanoyl-sn-glycero-3-phosphocholine; DM, n-decyl-β-maltoside; DDM, n-dodecyl-β-D-maltoside; NM, n-Nonyl-β-maltoside; DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine; DPC, dodecyl-phosphocholine; LMPG, 1-myristoyl-2-hydroxy-snglycero-3-[phospho-rac-(1-glycerol)]; LPPG, 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]; Lauryl-MNG, 2,2-didecylpropane-1,3-bis-β-D-maltopyranoside; Decyl- MNG, 2,2-dioctylpropane-1,3-bis-b-D-maltopyranoside; C6F-TAC, C6F13C2H4-S-poly[tris(hydroxymethyl)aminomethane; C8F-TAC, C8F17C2H4-S-poly[tris(hydroxymethyl); SDS, sodium dodecylsulfate; Triton X-100,; Tween 20, polyoxy-ethylene sorbitan monolaurate 20; Tyloxapol.

E. coli A19 (E. coli Genetic Stock Center)

Ethidiumbromide (Roth)

Folinic acid calcium salt (Sigma Aldrich)

Gene ruler 1kb DNA ladder (Fermentas)

Glucose monohydrate (Roth)

Guanosine 5`triphosphate disodium salt hydrate (Fluka Sigma Aldrich)

Hemin (Sigma)

HEPES (Roth) Imidazole (Roth)

2.3 Bacterial strains

Table 1. Bacterial strains

Strain	Genotype	Reference
***E. coli* DH5α**	[*F-,endA1, hsdR17 (rk-mk-), supE44, thi-1, recA1, gyrA (Nalr), relA1, Δ(lacZYA-argF)U169, Φ80lacZΔM15*]	[125]
***E. coli* BL21(DE3)**	*hsdS, gal* [*λcI, ts857, cnd1, hsdR17, recA1, endA1, gyrA96, thi-1, relA1*]	[126]
***E. coli* K-12 A19**	[*rna19 gdh A2 his95 relA1 spoT1 metB1*]	*E. coli* Genetic stock center, New Heaven, CT

2.4 Software

Gemini 4.2.17.304

Tecan Magellan 5.03

UNICORN 5.11

2.5 Common buffers/media/reagents

Buffers and Media for S30 extract and T7 polymerase preparation:

40 x S30-A/B buffer: 400 mM Tris-acetate, pH 8.2, 560 mM $Mg(OAc)_2$, 2.4 M KCl.

Supplement 1 x S30-A buffer with 6 mM ß-mercaptoethanol.

Supplement 1 x S30-B buffer with 1 mM DTT and 1 mM PMSF.

40 x S30-C buffer: 400 mM Tris-acetate, pH 8.2, 560 mM $Mg(OAc)_2$, 2.4 M KOAc.

Supplement 1 x S30-C buffer with 0.5 mM DTT.

2x YTPG medium: 22 mM KH_2PO_4, 40 mM K2HPO$_4$, 100 mM glucose, tryptone 16 g/l, yeast extract 10 g/l, NaCl 5 g/l.

LB medium: Peptone 10 g/l, yeast extract 5 g/l, NaCl 5 g/l.

Buffer T7RNAP-A: 30 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM EDTA, 10mM ß-mercaptoethanol, 5% glycerol.

Buffer T7RNAP-B: 30 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM EDTA, 10 mM ß-mercaptoethanol, 5% glycerol.

Buffer T7RNAP-C: 30 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.

Buffer T7RNAP-D: 10 mM K2HPO4/KH_2PO_4, pH 8.0, 10 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% glycerol.

20% streptomycin sulphate.

Reagents for CF reaction:

50 x Complete® protease inhibitor cocktail (Roche Diagnostics) 1 tablet per 1 ml of MilliQ water.

Amino acid mixtures containing 4 mM or 8 mM of each of the 20 natural amino acids.

RCWMDE mixture containing 16.7 mM of each amino acid.

1 M acetyl phosphate lithium potassium salt (AcP) (Sigma-Aldrich), adjusted to pH 7.0 with KOH.

1 M phospho(enol)pyruvic acid K^+ salt (PEP) (Sigma-Aldrich), adjusted to pH 7.0 with KOH.

NTP mixture containing 90 mM ATP, 60 mM CTP, 60 mM GTP and 60 mM UTP, adjusted to pH 7.0 with NaOH.

Pyruvat kinase (Roche Diagnostics), 10 mg/ml.

RiboLock® RNAse inhibitor (Fermentas), 40 U/µl.

Total *E. coli* tRNA (Roche Diagnostics), 40 mg/ml.

Folinic acid, Ca^{2+} salt, 10 mg/ml (Sigma-Aldrich).

Polyethylene glycol 8000 (PEG 8000), 40% (w/v).

4 M potassium acetate (KOAc).

2.4 M Hepes/ 20 mM EDTA, pH 8.0 adjusted with KOH.

500 mM 1,4-dithiothreitol (DTT).

E. coli S30 extract, store frozen at -80°C

T7-RNA polymerase (T7RNAP), store frozen at -80°C

Template DNA (plasmid DNA or linear PCR products) 200-500 ng/µl

Reaction container: Analytical and preparative scale reaction container; D-tube containers, 12-14 kDa MWCO (Merck Biosciences); Slide-A-Lyzer, 10 kDa MWCO (Pierce); dialysis tubes, 14 kDa MWCO.

10 x Premix: 15 mM putrescine, 15 mM spermidine, 2.5 M K^+-glutamate, 100 mM NH4+-glutamate, 100 mM Mg^{2+}-glutamate, 40 mM Na^+-oxalate, 330 mM Na^+-pyruvate, 340 µg/ml folinic acid, 10 mM DTT, 5.3 mM NAD^+.

30 mM CoA-Na^+.

GFP assay buffer: 20 mM Tris, 150 mM NaCl, pH 7.8.

1000×Ampicillin stock: Dissolve 100 mg/ml Na^+-Ampicillin salt in 50% H_2O and 50% EtOH. Store at -20 ℃.

1000×Kanamycin stock: Dissolve 30 mg/ml kanamycin sulfate in H2O. Store at -20 ℃.

SDS-gel buffers:

Stacking gel buffer: 0.4% (w/v) SDS, 0.5 M Tris-HCl, pH 6.8.

Separating gel buffer: 0.4% (w/v) SDS, 1.5 M Tris-HCl, pH 8.9.

5×SDS-PAGE sample buffer: 25% (w/v) glycerol, 25% (v/v) β-mercaptoethanol, 7.5% (v/v) SDS, 0.05% (w/v) bromphenol blue, 300 M Tris-HCl, pH 6.8.

Running buffer: 0.025 M Tris-HCl, pH 8, 0.1% (w/v) SDS and 0.2 M glycine.

Tricine-SDS- PAGE buffers:

10×Anode buffer: 1.0 M Tris-HCl, pH 8.9

10×Cathode buffer: 1.0 M Tris, 1.0 M Tricine, 1.0% (w/v) SDS, pH ~8.25

3×Gel buffer: 3.0 M Tris-HCl, 0.3% (w/v) SDS, pH 8.45

Coomassie brilliant blue-staining solution for SDS gels:

50% (v/v) ethanol (96%), 10% (v/v) acetic acid (100%) and 0.1% (w/v) Coomassie Brilliant Blue G250 + R250. Dissolve in H2O and store at RT in a dark bottle.

Western-blotting buffer (Towbin): Dissolve 25 mM Tris, 192 mM Glycin, 3.5 mM (1%) SDS, 15% MeOH in H_2O and adjust to pH 8.3 by HCl. Store at 4 ℃.

ECL1: 100mM Tris (pH 8.5), 2.5 mM Luminol, 0.4 mM p-cumaric acid. Store at 4 ℃.

ECL2: 100mM Tris (pH 8.5), 0.0183% H_2O_2. Store at 4 ℃

3. Methods

3.1 DNA template

sGFP (shifted green fluorescence protein, red mutant) was cloned into the pIVEX 2.3d (Roche Diagnostics, Penzberg, Germany) vector with C-terminal poly(His)$_{10}$ tag.

The 552 bp GNA1 cloned from plasmid pRGNA1 (conserved by Wang's lab in South China University of Technology), contains BamHI or XhoI restriction site indicated by underlines. 756 bp sGFP which consists of an intermediate TEV cleavage site obtained from plasmid sGFP (red mutant) respectively. Used primers are described in Tab. 2. The PCR cloning products were inserted into a modified vector pET21a (Merck Biosciences, Darmstadt, Germany) carrying T7 promoter and a poly(His)10-tag using RF cloning as described by [127]. The sequences of the primers used for cloning are given in Tab. 2

Table 2. construct for GNA1 and GNA1-sGFP and sGFP

Construct	Vector	Type	Sequence(5'- 3')
GNA1	pET21a	*BamH*I	CGCGGATCCATGAAACCTGATGAAACTCCT
		*Xho*I	CCGCTCGAGCTTTAGAAACCTCCGACACA
GNA1-sGFP	pET21a	RF	CTACATGTGTCGGAGGTTTCTAAAGGGCGAAAACCTGTACTTCCAGGGCG
		RF	GGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTTGTAGAGC

The CurA-halogenase is a dimeric protein which was crystalized in 2010 [128]. We received the construct as a gift from D. Sherman. CurA-halogenase was cloned into the vector pET28b with an N-terminal His$_6$-tag. The native protein sequence covers the amino acids 1599 to 1930 of CurA according to the sequence accessible at NCBI (GenBank accession code: AAT70096.1)

DNA templates used for CF expression were transformed into DH5α competent cells and isolated by plasmid purification kit (Macherey-Nagel, Germany).

The 903 bp mAQP4 M23 gene was amplified from the plasmid T-vector-AQP4 M23 (obtained from Dongbei Normal University, China). The following two primers were used: 5'-

GGAATTCCATATGGTGGCTTTCAAAGG-3' and 5'-GCCCGCTCGAG TACGGAAGACAAT-3'. The *Nde*I and *Xho*I restriction sites used for cloning are shown underlined. The restricted PCR products were ligated into a modified vector pIVEX 2.3d (Roche Diagnostics, Penzberg, Germany) containing a poly-$(His)_{10}$-tag. DNA templates used for CF expression were isolated from the resulting plasmid pIVEX2.3d-AQP4 M23 using commercial kits (Qiagen, Hilden, Germany).

3.2 Cell-free expression

3.2.1 Continuous exchange cell-free configuration

The preparation of T7-RNA polymerase, S30 extract and the final set-up of analytical as well as preparative scale reactions were performed as described in detail [57].

3.2.2 Batch configuration

As reaction container, standard 96-well v-shaped micro-plates or reaction tubes can be used. Total volumes of ≥ 25 µl are recommended for batch screening reactions in 96-well micro-plates.

The standard reaction mixture (RM) contained the following components: 2.5 mM ATP, 1.7 mM each of GTP, UTP and CTP, 34 mg/mL folinic acid, 170 mg/mL E. coli tRNA mixture (Roche, Indianapolis, IN, USA), 4-15 ng/µL of plasmid vector, 10 µg/mL T7 RNA polymerase, 2mM each of 20 unlabeled amino acids, 0.53 mM NAD, 0.26 mM CoA, 280 mM potassium glutamate, 10mM ammonium glutamate, 10 mM magnesium glutamate, 1.5 mM spermidine, 1.5 mM putrescine, 4mM sodium oxalate, 2 mM DTT and 0.24 volume of S30 extract [57].

TECAN Freedom EVO 200/8 equipped with one liquid handling and two transport arms was used to perform the pipetting the batch reactions. In order to perform fast and reliable pipetting, 4 needles each combined with a 1000µL syringe and 4 needles each combined with a 50 µL syringe were combined with liquid handling. All substances and destination vessels (e.g. 2 mL reaction-tubes) are kept during pipetting at 4 ℃ on a water bath cooled carrier (e.g. Cooling carrier for 2x16 reaction tubes; Cooling carrier for 3x96/394well plates in half-height or deep-well format). As the penultimate component the NTP-mix is added (before diluent), which functions as a kind of 'reaction starter'. Finally the plate was manually sealed with Parafilm and incubated at 32 ℃ and 9Hz shaking speed on a TECAN shaking incubator for 3h.

3.3 Electrophoresis

3.3.1 SDS-PAGE and western blotting

For SDS-gel analysis, protein samples supplemented with SDS sample buffer were loaded on 12% or 16.5% (w/v) Tris/glycine/SDS or Tricine/SDS gels and stained with Coomassie blue (shown in Tab. 3, Tab. 4). Protein sample was mixed with the 5 x SDS-PAGE Sample buffer and incubated at 98 ^{o}C for soluble protein, 37 ^{o}C for MP for 10 min. Then the sample was spin down at 18000g for 30 seconds. The supernatant was loading onto the gel. The running of the gel started with 100 V for 15 min and followed by 200 V for 45 min with constant current (time might change with different percentage of separating gels).

Table 3. Pipetting scheme for SDS-PAGE gels

Compounds	Separating gel 12%	Separating gel 16%	compounds	Staking gel 4%
1.5 M Tris/HCl pH 8.8	2.5 ml	2.5 ml	**0.5 M Tris/HCl pH 6.8**	2.5 ml
30% Acrylamide	4 ml	5.3 ml	**30% Acrylamide**	1.3 ml
10% SDS	0.1 ml	0.1 ml	**10% SDS**	0.1 ml
10% APS	50 µl	50 µl	**10% APS**	25 µl
TEMED	5 µl	5 µl	**TEMED**	5
H_2O	3.4 ml	2.1 ml	**H_2O**	6.1
Total	~10 ml	~10 ml	**Total**	~10 ml

Table 4. Pipetting scheme for Tricine-SDS-PAGE

Compounds	Separating gel (10%)	Separating gel (12%)	Stacking gel (4%)
30% (w/v) Polyacrylamid (37.5:1)	10 ml	12 ml	1.65 ml
Gel buffers 3×	10 ml	10 ml	3 ml
Glycerol	3 ml	3 ml	0
APS (10%)	150 µl	100 µl	30 µl
TEMED	15 µl	10 µl	3 µl
H_2O	6.83 ml	4.9 ml	7.32 ml
Total	~30 ml	~ 30 ml	~ 12 ml

For western blot analysis, the gels were transferred on a 0.45 µm Immobilon-P poly(vinylidene difluoride) membrane (Millipore) in a BioRad wet western blot apparatus for 50min at 350 mA. For Dot blot analysis samples were directly applied on a nitrocellulose membrane (Sartorius, Göttingen, Germany) and stored at 37 ℃ until the sample was totally dry. Membranes were blocked for 1 h in blocking-buffer containing 1 x Tris buffered saline, 4% skim milk powder (Fluka) and 0.05% (w/v) Tween20.

Horseradish peroxidase-conjugated T7-tag antibody (Merck Biosciences) was used for detection at a dilution of 1:5.000 and incubated for 1 h with the membrane. His-tag was detected by

anti-penta His IgG from mouse (Qiagen) in a 1:2.000 dilution, the membrane was washed 3times with 1 x Tris buffered saline and 0.05% Tween20. As a second antibody an anti-mouse IgG HRP conjugate from goat (Sigma Aldrich) was used in a 1:5.000 dilution.

3.4 Protein purification

3.4.1 mAQP4 M23 purification

The CF expressed mAQP4 M23 was purified in one step by imobilized metal-chelated affinity chromatography (IMAC). 1 ml of either D-CF soluble expressed mAQP4 M23 or P-CF resolubilized mAQP4 M23 was mixed with 300 µl of Co^{2+} loaded NTA resin slurry (Qiagen, Hilden, Germany). The mixture was diluted 10-fold with column buffer (20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 20 mM imidazole, 0.05% DDM or 0.05% Fos-12) and incubated at 4oC for overnight with gently shaking. The mixture was poured into an empty column and washed at gravity flow rate of ten column volumes of column buffer supplemented with 80 mM imidazole. The bound mAQP4 M23 was then eluted with column buffer supplemented with 300 mM imidazole.

3.4.2 Size exclusion chromatography

SEC with prepack columns (GE Healthcare, Munich, Germay) Superdex 200 PC 10/300 or Superdex 200 PC3.2/300.

3.5 Protein quantification

3.5.1 Fluorescence

After expression, keep GFP/sGFP samples at 4^{o}C for 12 hours to allow complete GFP folding. Add 3 µl of sample into 297 µl GFP assay buffer in a 96-well dark micro-plate. Incubate the plate with shaking for 5 min at 22^{o}C. Fluorescence measurement for sGFP is performed at an excitation wavelength of 484 nm and an emission wavelength of 510 nm. For wild type GFP the excitation wavelength is 395 nm with an emission wavelength of 510 nm. Use a calibration curve with purified GFP/sGFP to quantify the measured fluorescence.

3.5.2 Western blotting

Positope™ was used as standard protein positive control for western blotting quantification. The procedure for western blotting was performed as above in 3.3.1

3.5.3 ^{35}S-Met Incorporation

S35 labeled methionine was used for radioactive quantification. S35 labeled methionine was mixed with non-label amino acids mix with a ratio of 1: 40,000 and then the mixture were added into the reaction mixture for expression. After incubation, the reaction mix was collected and transferred into eppendorf tubes. Then the total protein was quantified by a TCA precipitation. Add ice-cold TCA (15%, w/v) into the sample to reach final concentration of 10%, centrifuge at 22,000xg for 10 min at 4 ^{o}C. The pellet was washed two times by 10% ice-cold TCA and one time of 95% ethanol. Carefully remove the supernatant and dry the pellet then the eppendorf tube containing the dried pellet was dropped into a scintillation vial, 20 ml (Carl Roth, Karlsruhe, Germany) filled with 5 ml scintillation cocktail Rotiszint®eco (Carl Roth, Karlsruhe, Germany). Turn the vial upside down to allow efficient contact between the scintillation liquid and the pellet. After 1-2 hours, the scintillation is counted for 1 min using a liquid scintillation counter (LS6500, Beckmann).

3.6 Functional analysis

3.6.1 Liposome preparation

E. coli polar lipids or other artificial lipids were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL). Lipid mixtures solubilized in chloroform were first transferred to a round bottom flask. A thin lipid film was formed by evaporating the chloroform under a nitrogen stream and placing the flask in a vacuum chamber for overnight. The lipids were then reconstituted in 1 ml of assay buffer (100 mM MOPS-KOH, pH 7.5) or S30 C buffer to a final concentration of 20 mg/ml by vortexing for 15 min to form multilamellar vesicles. The multilamellar vesicle solution was passed at least 21 times through an Avanti Polar Lipids mini extruder holding a 200 nm Whatman polycarbonate membrane filter (Florham Park, NJ) sandwiched with two filter supports on each side. The resulting unilamellar liposome solution was used for MP reconstitution.

3.6.2 Reconstitution

IMAC purified mAQP4 M23 protein was reconstituted into liposomes composed of E. coli polar lipids (Avanti Polar Lipids, Alabaster, AL, U.S.A.) by modification of a previously published protocol [129]. Briefly, a reconstitution mixture was prepared in a microtube at room temperature by sequentially adding reconstitution buffer (100 mM Mops, pH 7.5), 10% (v/w) Triton X-100

(final concentration 4 mM), 20 mg/ml preformed liposomes (final concentration 4 mg/ml), and 100 µg/ml purified mAQP4 M23. The reconstitution mixture was incubated at room temperature with gently shaking for 30 min. The detergent was removed by SM-2 beads (Bio-Rad, München, Germany) according to the manual. Finally, the liquid reconstitution mixture was sent to ultracentrifugation at 500,000 g for 45 min. Then, the pellet was washed again with reconstitution buffer. After wash step, the proteoliposome solution was ultracentrifugated again and finally resuspended in 1.6 ml reconstitution buffer.

3.6.3 Water channel activity assay

Water permeability was measured by 90 degree light scattering at 436 nm in a stopped-flow spectrophotometer (SFM 300, BioLogic). Before measurement, the proteoliposomes were extruded once through a 200 nm membrane filter for homogenization. Proteoliposomes suspended in reconstitution buffer were quickly mixed with equal volumes of a hyper-osmotic solution (reconstitution buffer with 400 mM sucrose). Because sucrose does not penetrate into the proteoliposomes, the applied osmotic gradient initiates a water efflux. The resulting shrinking of the liposomes can be recorded by light scattering analysis. Data were fitted to an exponential rise equation, and the initial shrinkage rate (k) was determined by the average of at least three fitted equations. The water permeability factor P_f of the proteoliposome samples was calculated as described previously [130] using the equation:

$$P_f = \frac{k}{(S/V_0) \times V_w \times \Delta_{osm}}$$

where S/V_0 is the vesicle's initial surface to volume ratio, V_w the partial molar volume of water (18 cm^3 mol^{-1}), and Δ_{osm} the osmotic driving force. The S/V_0 was calculated by determining the diameter of the proteoliposomes using dynamic light scattering (ZetaPlus particle sizing software 2.27). The diameter of the proteoliposomes and empty liposomes were determined to be 113 nm. The Δ_{osm} was 200 mM in this case.

Here one specific example was given as follows:

We take the D-CF reconstituted proteoliposome at 10°C as an example. k were obtained by three curve fitting to be 24.42, 25.33, 26.57. Then S/V_0 was 5.31×10^7 m^{-1}. So the final P_f were 127.75 µm/s, 133.0 µm/s, 139 µm/s, respectively. Then the mean value and SD was given as 133.1 $\pm$ 5.6 µm/s

3.6.4 Activity Assay of the Cell-free synthesized GNA1-sGFP

After expression, reaction mixture were collected and transferred into D-tubes. In order to get a better exchange of the dialysis, reaction mixture was diluted into 100 µl with buffer (50 mM

Tris-HCl, pH 8.0, 5.0). Then the samples were dialysis against 500 ml buffer with stirring at 4 ^{o}C for 2 hours. Then the sample was collected and centrifuge at 22,000xg for 10 min. Supernatant was then used for enzyme activity assay. Activity was measured as previously described [2] with modifications. The assay was performed in 50 µl of mixture buffer containing 500mM GlcN6P, 500mM AcCoA, 50 mM Tris-HCl, pH 8.0, 5.0 mM MgCl2 and 10% glycerol in 96-well flat bottom plates, and approximately 0.2 µg of purified CECF expressed GNA1 and 0.4 µg unpurified GNA1-sGFP were added to start the reaction, respectively. After incubation at 30 ^{o}C for 5 mins, the reaction was terminated by adding 50 µl of the stop solution (50 mM Tris–HCl, pH 8.0, and 6.4M guanidine hydrochloride) and then 50 µl of CR buffer (50 mM Tris–HCl, pH 8.0, 1 mM EDTA, and 200 uM DTNB). The amount of CoA produced by the GNA1 was estimated by 4-nitrothiophenolate and measured at the optimal density of 412 nm on full wave length spectrophotometer microplate reader (Thermo Scientific, USA). A blank reaction using CF reaction mixture without GNA1 or sGPF protein was set up as background control. The amount of CoA produced was calculated using the extinction coefficient of DTNB at 30 ^{o}C (13,800 M^{-1} cm^{-1}).

4. Results

4.1 Establishing high-throughput cell-free expression platform

The open nature and the high success rate of CF systems enable throughput expression condition screening as a new rational strategy for protein expression. Quality and production of proteins can be optimized by extensive modification of their direct expression environment. However, even with CF system, to find suitable stabilizers or folding enhancers still require large amount of screening experiments for each specific protein target. In order to find a solution to this time consuming procedure, we introduce a robotic pipetting system, based on Tecan Freedom EVO 200 workstation, to perform all the pipetting work for compounds screening via CF expression system[57]. Studies were already done to introduce the robotic pipetting system to perform the CF expression. Aoki *et al* introduced robotic system to pipetting the CF reactions using the continuous exchange mode, which suitable for preparative scale sample production [131]. However, we still chose the batch mode because of its easy handling and high-throughput. For screening purpose, the expression level of batch mode was already sufficient for detection[132]. With the help of custom-designed operation software E.Y.E.S program, time required for programing and pipetting was greatly reduced, around 30-45 min per 96 reactions. The use of robotic liquid handling systems and calculation programs will largely facilitate such throughput applications.

4.1.1 Setting up batch cell-free system via robotic pipetting

With the goal of setting up an automated CF expression platform, we introduced the robotic pipetting platform of TECAN Freedom EVO 200, which contains one liquid handling and two transport arm as shown in Fig. 9. In considering to minimizing the CF reaction volume, and covering the most frequent used pipetting volume, we select a set of tips and syringes (4 tips each combined with 1000 µl syringe; 4 tips each combined with 50 µl syringe) for the robot to allow us perform accurate and fast liquid pipetting work ranging from nanoscale volume to milliscale. Cooling carriers for 50 ml falcon tubes; 15 ml falcon tubes; 2×16, 1.5-2 ml eppendorf tubes; 3 position for 96/394 well plates are introduced to keep all the temperature sensitive compounds and CF reactions. Thermo-shaking devices are also introduced for CF reaction incubation and shaking.

4.1.1.1 Establishing an efficient batch CF expression protocol

The simple batch format CF expression protocols are becoming more and more productive and the yield even approaching 1 mg protein per 1 ml of reaction volume[78, 133]. Batch format CF reactions are ideal for throughput systems with small volumes as low as a few microliters. The set up and further processing of the reaction can then be handled via robots. The ‘Cytomim’ system was

used as a basic CF protocol in batch format as described [77]. Modifications are made in order to

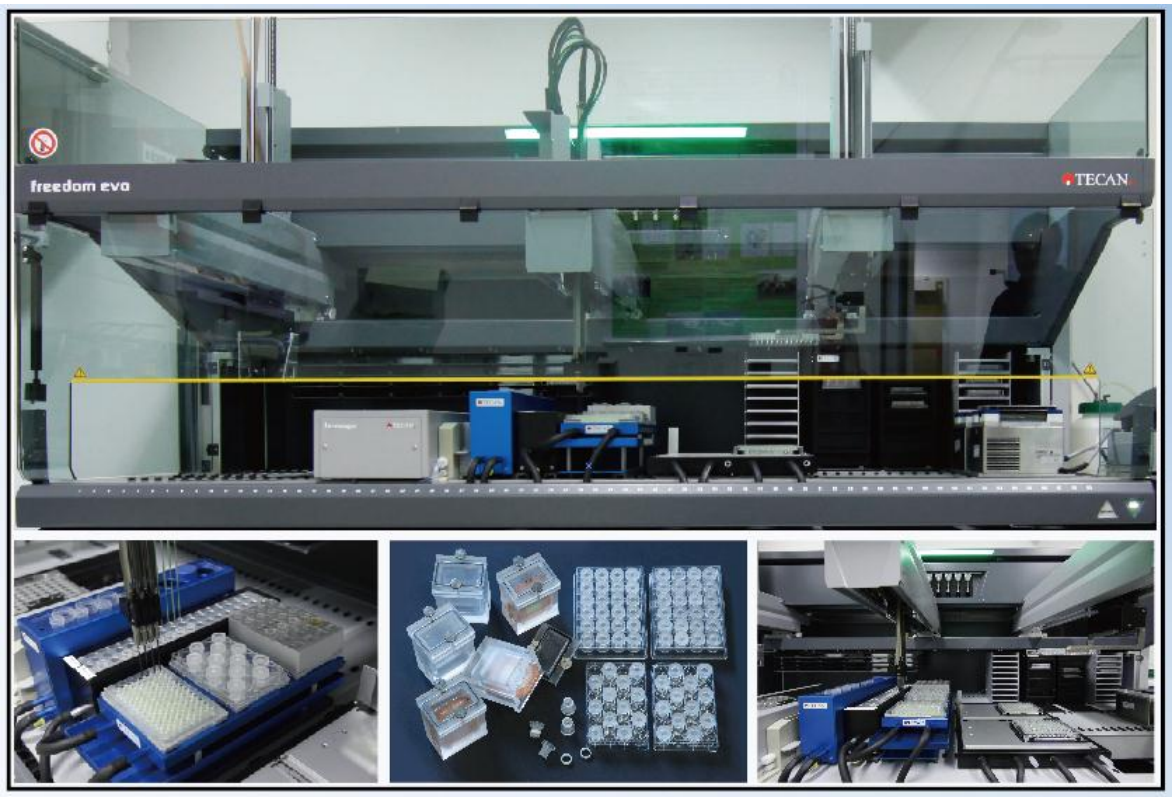


Figure 9. Platform for high-throughput optimization of CF MP production. An integrated process for pipetting, incubation, separation, purification, and quantification.

obtain the highest yield, less cost, and suitable for robot pipetting [57]. Basic compounds and the recommend optima concentrations are listed in Tab. 5.

Table 5. list of all the compounds for batch CF reaction and with the recommended concentration range.

Compound	Stock	Final
Premix	10 ×	1 ×
Amino acid mix	8 mM each	2 mM each
PEP	1 M	30-50 mM
CoA-Na$^+$	30 mM	0.26 mM
***E. coli* tRNA**	40 mg/ml	0.17 mg/ml
T7RNAP	1.4 mg/ml	10 µg/ml
36 x NTP-Mix	ATP: 90 mM C/G/UTP: 60 mM each	ATP: 2.5 mM C/G/UTP: 1.7 mM each
Plasmid template	0.2 - 0.5 mg/ml	0.002-0.015 mg/ml
***E. coli* S30 extract**	100%	24%-30%
MilliQ water		fill up to final volume

In the compound list, we introduced a premix (Tab. 6), which contains several standard compounds, in order to reduce the time for pipetting. Premix could be stored at -20^{o}C and refrozen without detectable loss of efficiency. Extract, enzymes, unstable compounds and potential screening components were not included into the premix in order to keep options for further protocol

optimizations. Reaction volumes were miniaturized down to 25 µl in order to reduce cost but still allowing sufficient volume for screening compounds. The pipetting time of a complete 96-well micro-plate could be reduced to 30-45 min depending on the number of screening compounds and additives.

Table 6. Compounds that contained in the premix.

Compound	Stock	Final
Putrescine	15 mM	1.5 mM
Spermidine	15 mM	1.5 mM
K^+-glutamate	2500 mM	250 mM
NH_4^+-glutamate	100 mM	10 mM
Mg^{2+}-glutamate	100 mM	10 mM
Na^+-oxalate	40 mM	4 mM
Na^+-pyruvate	330 mM	33 mM
Folinic acid	340 µg/ml	34 µg/ml
DTT	10 mM	1 mM
NAD^+	5.3 mM	0.53 mM

Folding of sGFP is oxygen dependent and the plates were incubated for 2 hours at 4 ^oC, after the reaction in order to allow sufficient time for folding of the monitor.

4.1.1.2 Perform the CF protocol via robotic pipetting

With established CF protocol, CF reactions can be then pipetted by the robot. However, several important parameters with regard to the liquid classes need to be considered and optimized for the robotic pipetting. Different liquid classes were defined as: diluent, ProteinExp and Protein ExpVisc. As each liquid requires due to its properties special handling features, we focused specially on the viscosity, which gives a major impact on velocities of aspiration and dispensing, introduction of air-gaps and excess volumes or the positioning of the tip relative to the liquid surface. So water was finally defined as diluent, medium viscous liquids as e.g. co-factors as ProtExp and very viscous or crucial liquids like premix, extracts, enzymes as ProteinExpVisc.

In order to further reduce the time required for pipetting and also obtain reliable results of compounds screening. A mastermix (MM) has to be created, which contains all the components that are kept constant in each reactions of a screening. Then, the MM can be aliquoted into each well. Finally, the screening compounds were then pipetted. During all the pipetting process, the plate was kept at 4 ^oC on the cooled carrier. Normally, the NTP-mix is added in the final step to act as a reaction starter. After finishing pipetting, the microplate was sealed with Parafilm and incubated at 30 ^oC, 9 Hz shaking speed on the TECAN incubator.

In order to handle the complicated pipetting process, we introduced custom designed software to create the pipetting lists, which can be carried out by the TECAN software Gemini. All

the reagents with stock concentrations are stored in the reagent data base. When starting the program, all the reagent which are needed in the following experiment, are selected. Secondly, set up the final concentration of all the compounds. For the reagents that you want to perform a concentration screening, in this step, choose the highest concentration in the screening. Thirdly, assign the usage of all the compounds. In the current version, two screening compounds can be defined. The final concentration of two screening compounds can be programed to vary per row or per columns. At the same time, up to 4 additives can be selected. The additives are chosen to be added per rows. Additive 1 is reserved for the starter reagent. Normally, the NTP mix was used as starter. However, other reagent like DNA template can also be the starter reagent.

4.1.2 Basic compounds screening

With the established CF expression platform, two basic screening modes can be performed. One is linear screening, where the concentration of one compounds can be screening within certain range. Another is correlated screening, where the concentrations of two compounds can be screened at the same time. One is screened per row; the other is screened per column as shown in Fig. 10.

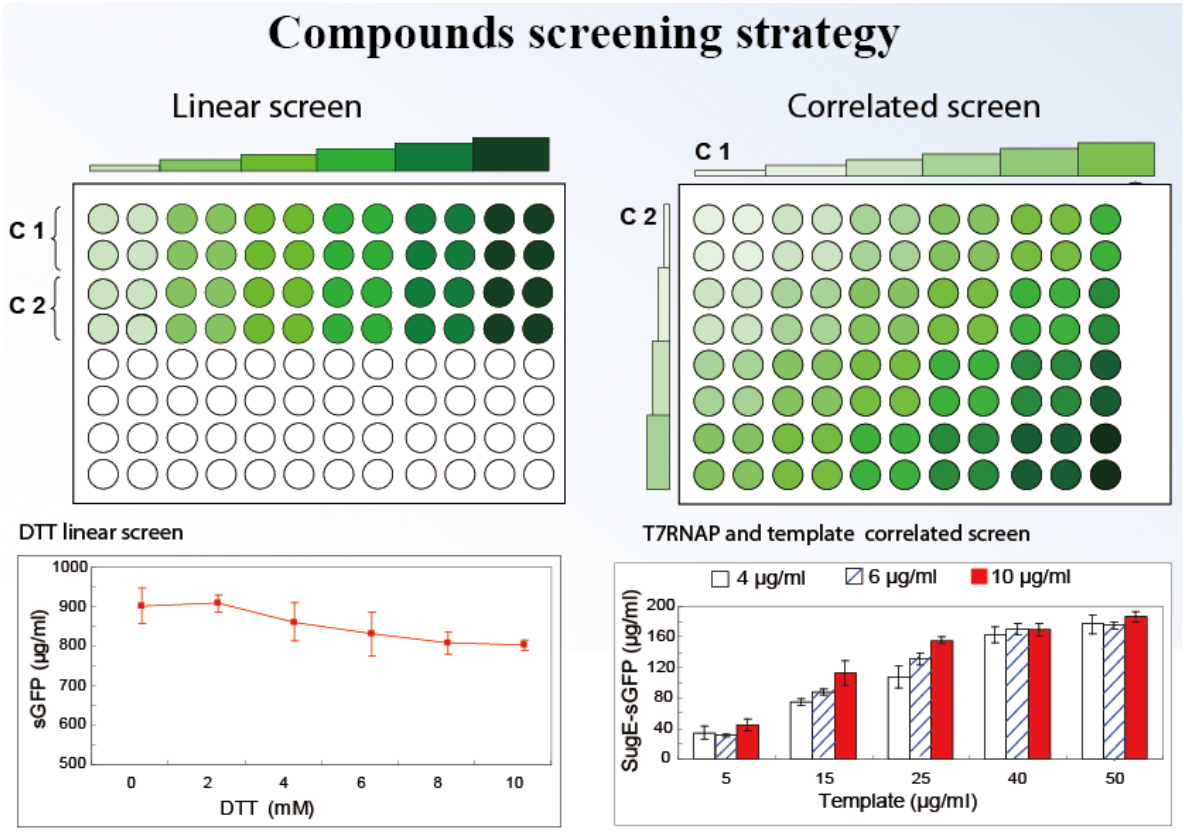


Figure 10. CF expression protocol optimization by throughput linear and correlated compound screens. The expression of sGFP and small multidrug transporter SugE-sGFP has been analyzed. The screening compounds are indicated in the figure.

4.1.2.1 Linear screening of basic compounds using sGFP as a monitor

In order to perform a fast determination of CF expression yields, a shifted green fluorescent protein (sGFP) was selected. The fluorescent measurement can be easily and accurately quantified and correlated with the protein concentration. Several critical basic compounds were selected and evaluated to determine their optima concentrations.

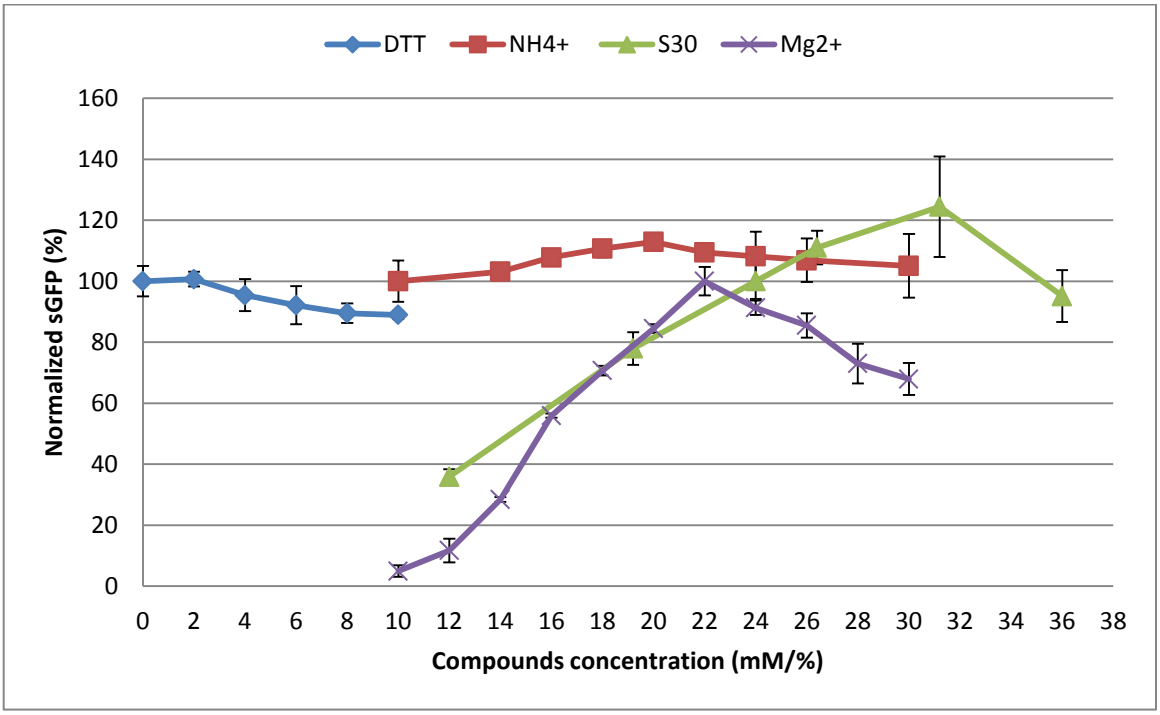


Figure 11. Linear screening of the basic compounds of CF system. Each compound was screened according to the concentrations mentioned above while other compounds were kept with the optima concentration.

To establish an expression protocol, several important compounds have to be optimized like, Mg^{2+}, K^{+} and S30 extract. Here we select several compounds, which are not included in the premix, and further optimize them by using the linear screening function of the E.Y.E.S. program (Fig. 11). A major compound is the S30 extract and the optimal final concentration in the reaction was found at approx. 30% (Fig. 11). The Mg^{2+} optima of different S30 extract preparations can be different and the therefore a rescreening of the ion concentration was performed with each new S30 extract batch (Fig. 11).

Reducing conditions could become critical depending on the structure of the synthesized target proteins. DTT as reducing agent is tolerated in the reaction from zero to 10 mM final concentration without significant effects and could thus be supplemented if necessary (Fig. 11). NH_4^{+} titration from 1 to 30 mM had no effect on GFP expression.

The template concentration always plays an important role in the final expression yield. Not only the quality of the template was important, but also the concentration of the template is crucial to the final expression yield. sGFP and soluble enzyme GNA1-sGFP were selected as example for the template concentration screening as shown in Fig. 12. The results showed that below 1 ng/µl the

expression yield is linear correlate with the template concentration. When the template concentration is above 2 ng/ µl, the yield of the CF reaction reached a plateau.

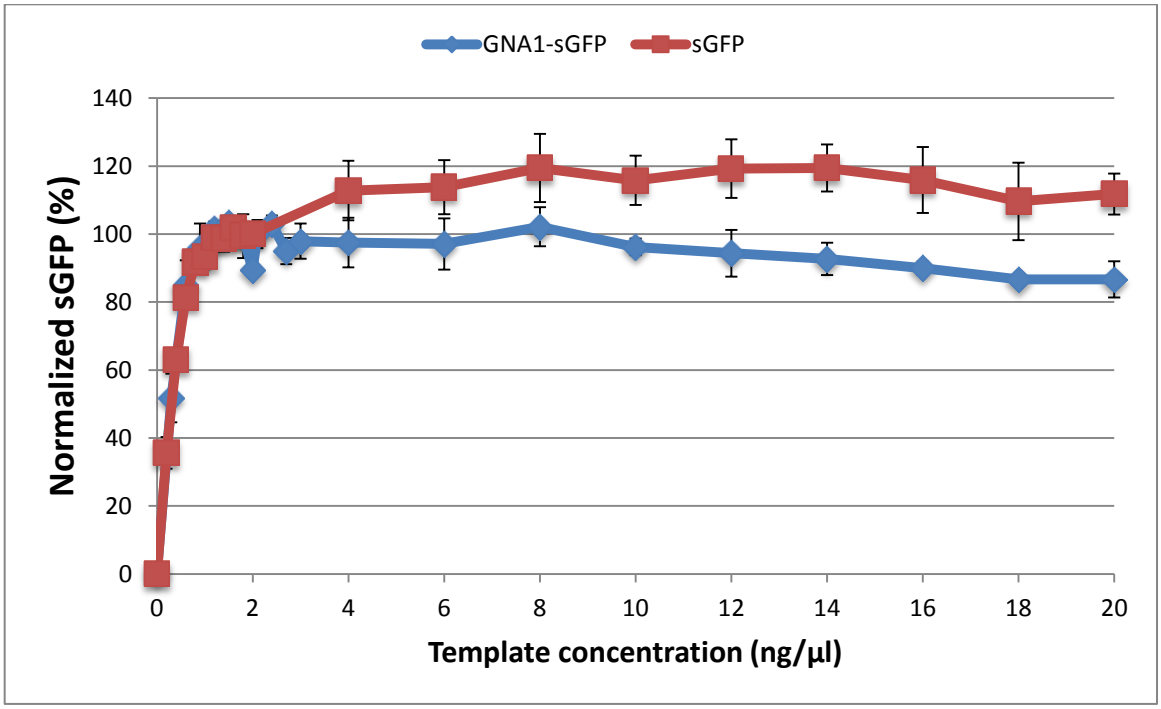


Figure 12. Template concentration screening in batch CF system. Two targets sGFP and GNA1-sGFP were selected as model proteins.

The usage of linear PCR product as DNA template great accelerates the whole preparation of CF reactions. In order to evaluate the difference between linear template and plasmid template, experiment with different concentration of linear template was performed. As shown in Fig. 13

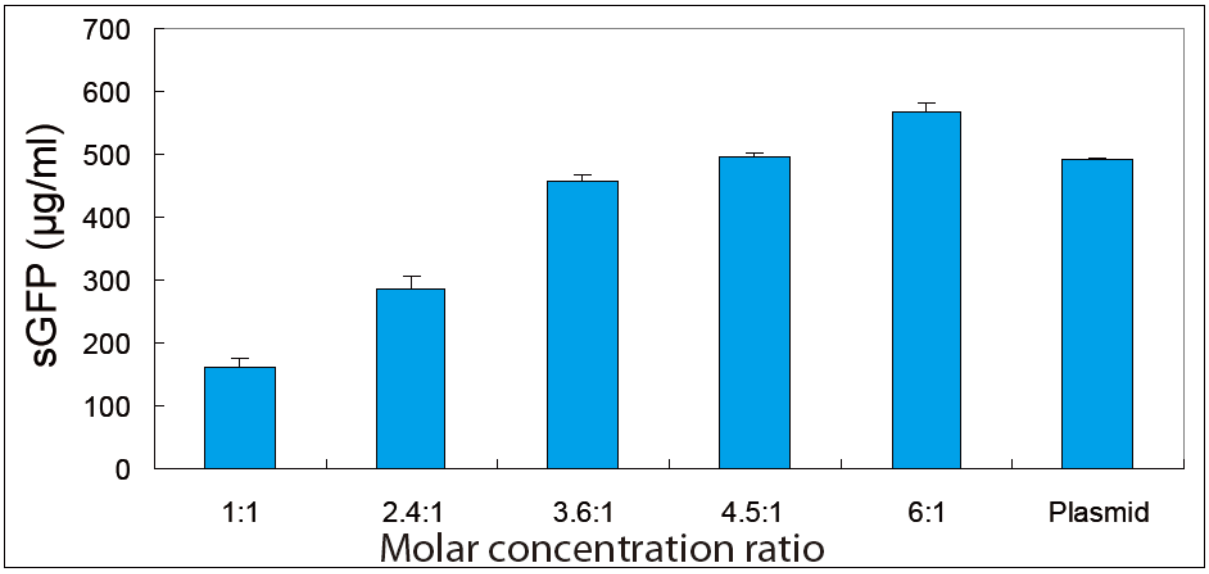


Figure 13. Comparison of linear template and plasmid template in CF reaction systems. Different concentrations of linear template was used in CF system and compared with the plasmid concentration. Different linear template to plasmid template molar ratios are indicated in the figure.

Several very sensitive parameters like Mg^{2+} concentration are recommended to be screening whenever changing reagents like extract and buffers. In the following experiment, two different batch of S30 extract was used and the Mg^{2+} concentration was screened. As shown in Fig.14, there

is a clear optimum concentration shift between the two batches of extract. A small change in the Mg^{2+} concentration may cause great reduction in the final protein yield.

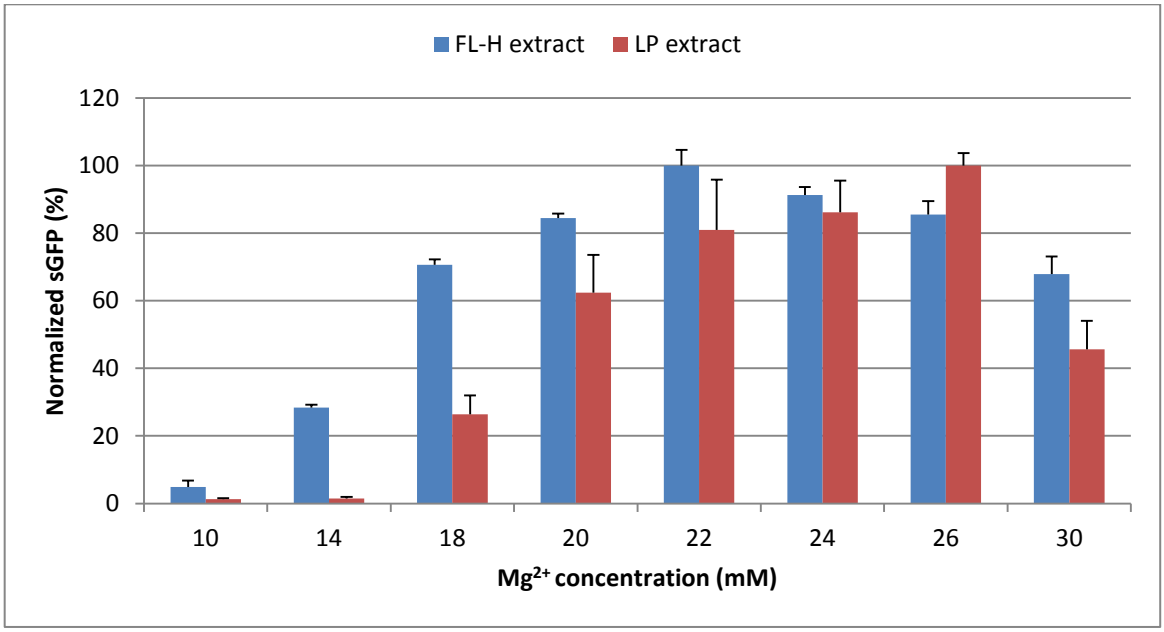


Figure 14. Comparison the optima Mg^{2+} concentrations of two different batches of S30 extract.

Kinetic experiment of the batch CF expression system was performed and shown in Fig.15. Results shown that after 2 hours the expression is nearly finished.

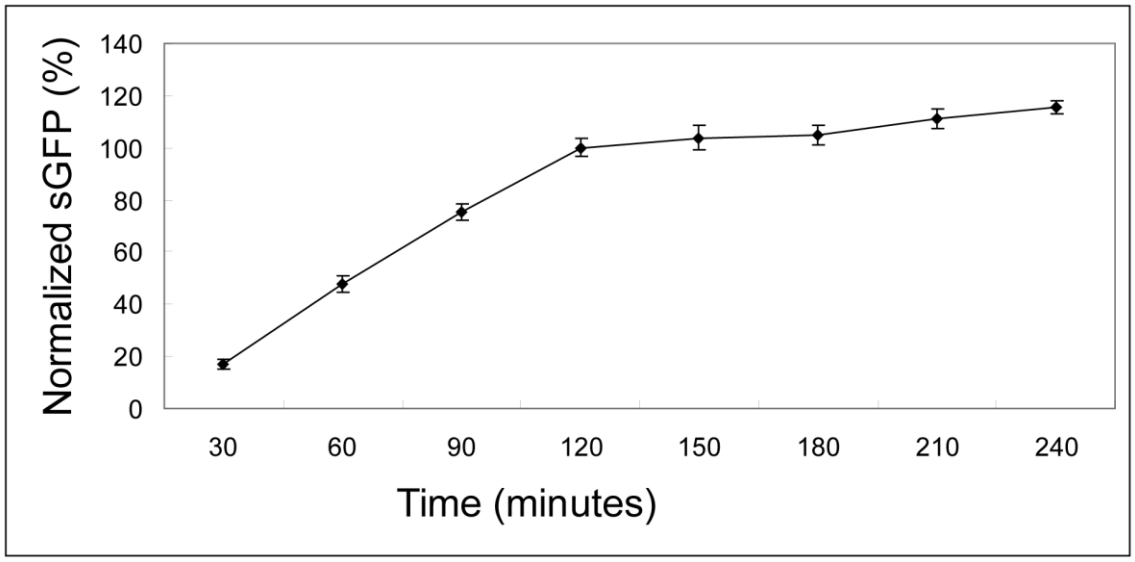


Figure 15. Kinetics of batch CF expression system. At different time point, reactions was stopped by lowing the temperature to 4°C for 1 hour, then the fluorescence of the reaction was measured and sGFP concentrations were calculated using standard curve.

In order to test, if this optimized batch CF system can be easily scaled up without loss in the yield. We performed the following experiment with different volume. Volume range from 25 µl to 1 ml was selected. The experiment was performed in 96-well plate and 24-well plate. 96-well v-shaped plate was used when the volume is below 200 µl, while 24-well plate was used when the volume is above 200 µl. The result of this experiment indicated that simply scale up to 1 ml reaction volume will not decrease the efficiency of the expression system (Fig. 16).

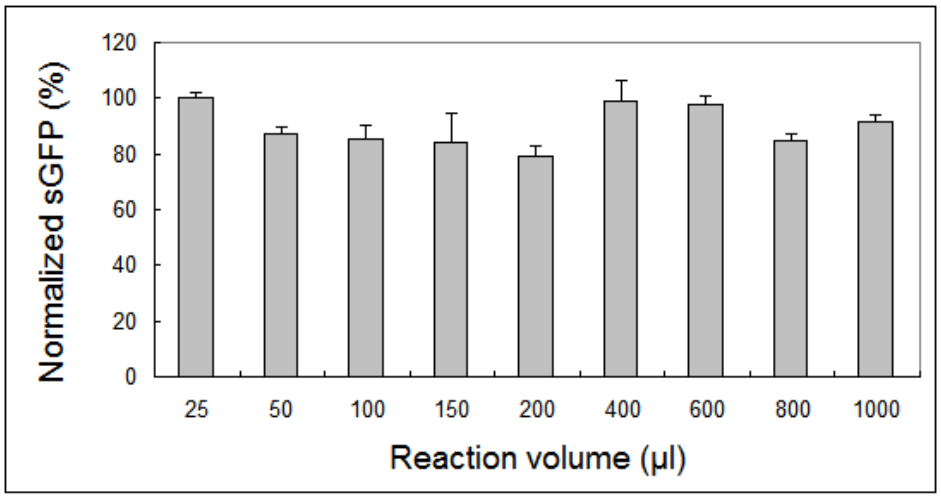


Figure 16. Scale up of CF batch reactions into preparative scales. Reaction conditions for sGFP expression optimized in 25 µl volumes were stepwise scaled up to one ml reaction volumes. For reaction volumes ≤ 200 µl, 96-well microplates were used. For reaction volumes ≥ 400 µl, 24-well microplates were used.

4.1.2.2 Correlated screening of basic compounds using sGFP as a monitor

In linear screening, the effect of individual compounds were evaluated, however, there are cases, where the concentration optima of several compounds could depend on each other. In this correlated screening, two compounds are screened at the same time. With the E.Y.E.S program, this

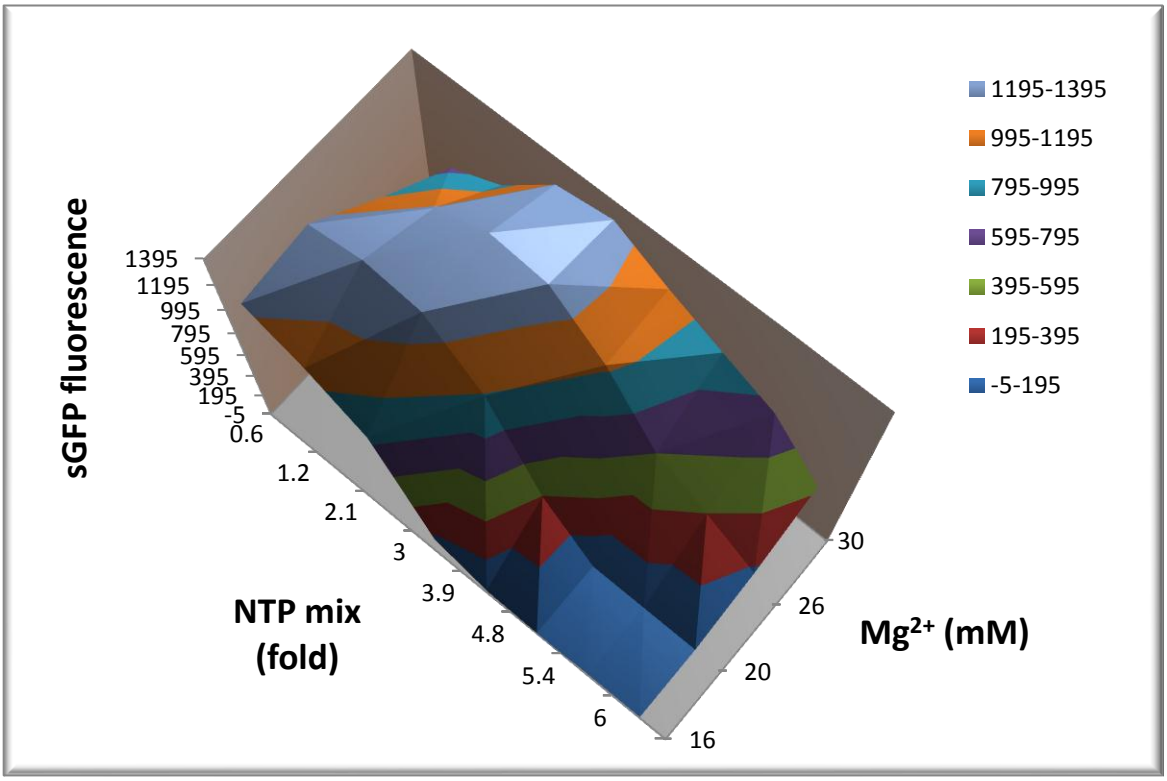


Figure 17. Correlated screening of Mg^{2+} and NTP mix. 1 fold of NTP mix corresponding

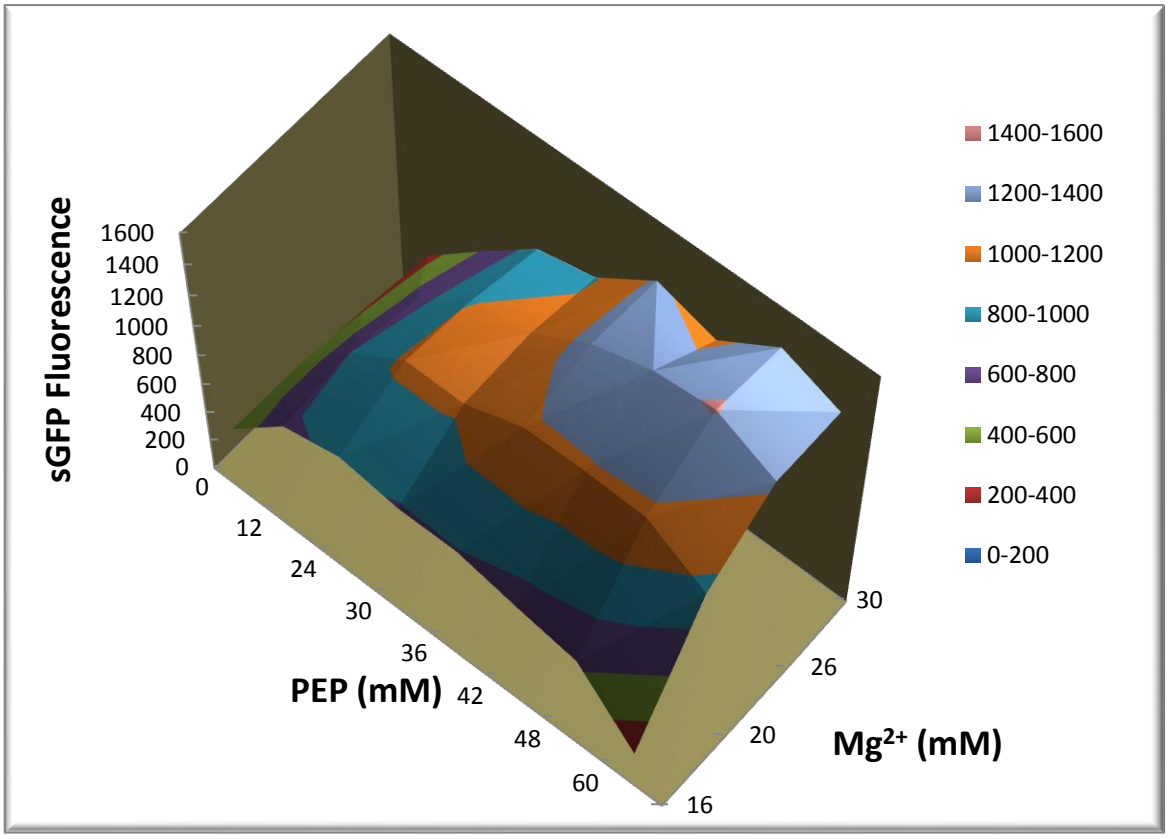


Figure 18. Correlated screening of PEP and Mg^{2+}. Different concentration of the two selected compounds were sleceted

correlated screening can be easily achieved as shown in Fig. 8. Mg^{2+} and NTP mix are selected, as shown in Fig. 17. With certain concentration range, the increase of NTP mix concentration could be implemented by increasing Mg^{2+} concentrations. The optimum concentration for NTP mix in this screening is between 1 and 2 folds with Mg^{2+} between 20 and 26 mM. Another example was PEP and Mg^{2+} (shown in Fig. 18); however, here the effect of PEP cannot fully be implemented by Mg^{2+}. In this case, the highest expression yield was achieved with the fixed optimum concentration (PEP 48 mM, Mg^{2+} 26 mM) for these two compounds.

4.2 Systematic chemical compounds screening via throughput cell-free system

There quite a large number of proteins which tend to form precipitate, soluble aggregation and mis-folded inclusion bodies when expressed *in vivo*. Well known class of protein was enzymes, which were quite sensitive to temperature, pH, and buffers. It was quite difficult for traditional *in vivo* expression system to produce such kind of difficult proteins and often results in low yield or non-functional mis-folded proteins. Therefore quite lots of studies were done to improve the stability of these unstable proteins. In natural environment, both prokaryotic and eukaryotic cells develop a strategy to protect their proteins by producing low molecular weight organic substance called osmolytes when subjected into harsh environmental conditions such high salts, cold and hot stress1. In addition, to apply such kind of low molecular compounds to the protein as soon as they were produced was quite difficult because of the cell wall and membrane barrier. As a result, those proteins often were expressed as inclusion bodies or aggregations *in vivo*. Later on, the mis-folded protein was refolded after a complete denaturation via harsh chemical reagent like urea or guanidine hydrochloride. The final ratio of successful refolded protein was often in a relative low amount.

According to previous reports, different osmolytes (i.e., amino acids and their derivatives, polyols and sugars), PEG serious, PEI, betaine, ectoine and choline, behaving as protein stabilizers [134, 135], inducing refolding of misfolded proteins [136, 137] and removing protein aggregation[138-140], were selected for testing. However, those experiments were performed post-translational; the effects of those chemical compounds on the CF expression system still need to be systematically evaluated. The effects were shown as co-translational level. Red shifted green fluorescence protein (sGFP) was chosen as model protein for fast quantification by fluorescence. Amount the selected osmolytes, three basic effects were observed, which we classified as positive, negative and tolerant. For positive effect, an increase of sGFP fluorescence was observed. For negative effect, the CF system was killed or terminated with very low concentration of certain compounds. With tolerant effect, the fluorescence of sGFP decreased with the increase concentration of added compounds.

In order to exemplify the application of those compatible chemical protein stabilizers, we choose two targets from our lab, which are difficult to expression and tend to form precipitate and aggregation. One is Halogenase, which is the first enzyme within the biosynthesis of curacin A (a bioactive compound with potent anti-proliferative activity) pathway. The biosynthesis of CurA is mediated by a 2.2 MDa hybrid polyketide synthase (PKSs) and non-ribosomal peptide synthetases (NRPSs) as shown in Fig. 19. Another target we select was the human GNA1 (EC 2.3.1.4) catalyzes the transfer of the Acetyl group from Acetyl Coenzyme A (AcCoA) to the primary amine of D-glucosamine-6-phosphate to form N-acetyl-D-glucosamine-6-phosphate (GlcNAc-6P) and Coenzyme A (CoA)[141, 142].

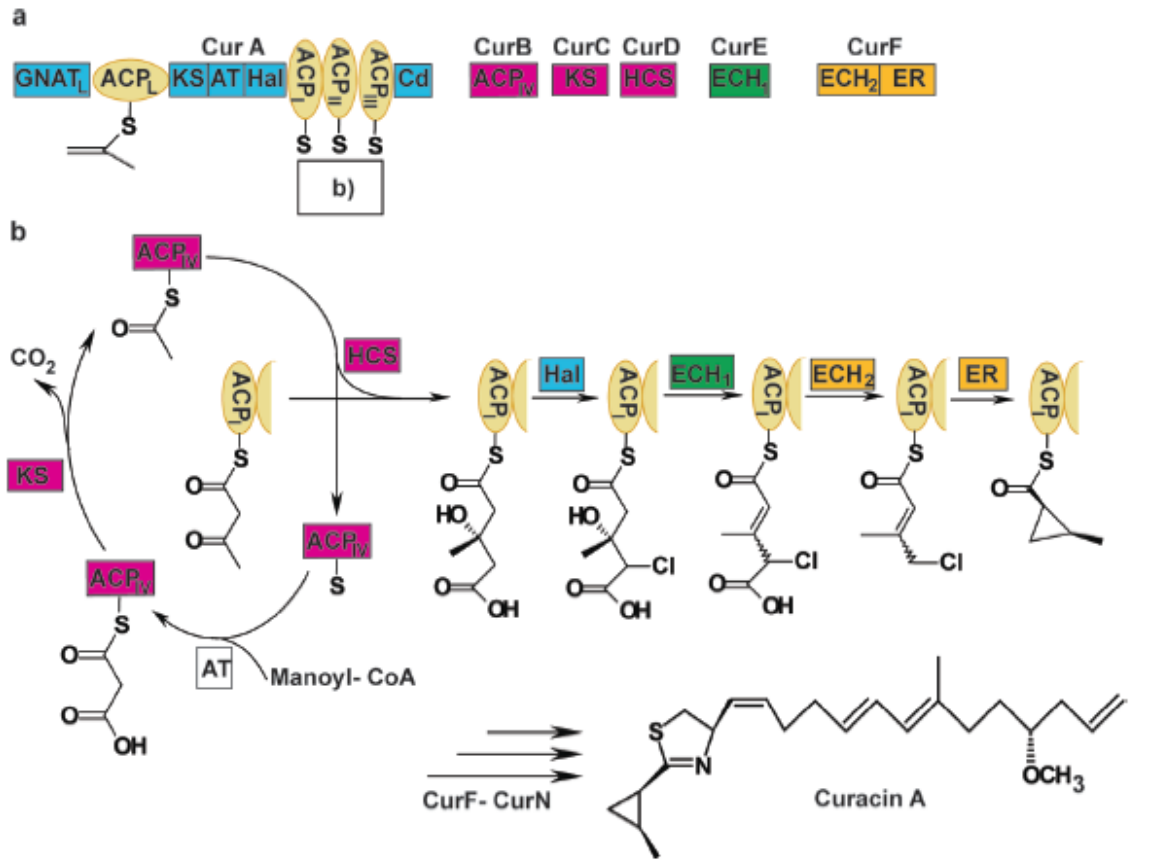


Figure 19. Ten-enzyme assembly catalyzing the cyclopropane ring formation. (a) The 10 enzymes that are involved in the cyclopropane ring formation are encoded on different proteins. (b) Representation of the biosynthesis of the cyclopropane ring formation. GNATL = loading module, KS = ketosynthase, AT = acyltransferase, Hal = halogenase, ACP = acetyl-carrier-protein, HCS = HMG-CoA synthase-like enzyme, ECH1 = dehydratase, ECH2 = decarboxylase, ER = enoyl-reductase

4.2.1 Linear screening of chemical compounds using sGFP as monitor

sGFP was used as monitor to see if the selected chemical compounds can be tolerant or beneficial effect on the CF system. The fluorescence of sGFP was used as a quantification method for protein quantification.

4.2.1.1 Effect of alcohols

Alcohols, due to their hydrophobicity, help for stabilizing MP and are often used in crystallization studies. In this experiment, different alcohols were select from methanol to hexanol, the effect on the expression of sGFP was shown in Fig. 20. In addition, we reduced the Mg^{2+} concentration and performed the same experiment to see if the effects of the alcohols on sGFP are dependent on the Mg^{2+}. As shown in Fig. 20, different effects were observed for different alcohols. Methanol, isopropanol and butanol are tolerant by the system to certain concentrations. The expression yield decreased linearly with the increase those three alcohols. Ethanol was found to have even a beneficial effect on the expression with up to 60% more if compared to the control. In contrast, pentanol and hexanol immediately killed the CF system with very low concentration. A

similar experiment was performed with lower Mg^{2+} concentration. The effects were the same as compared to the optima Mg^{2+}, except for the case of ethanol. The expression yield was increase much more up to 300% compared to the control.

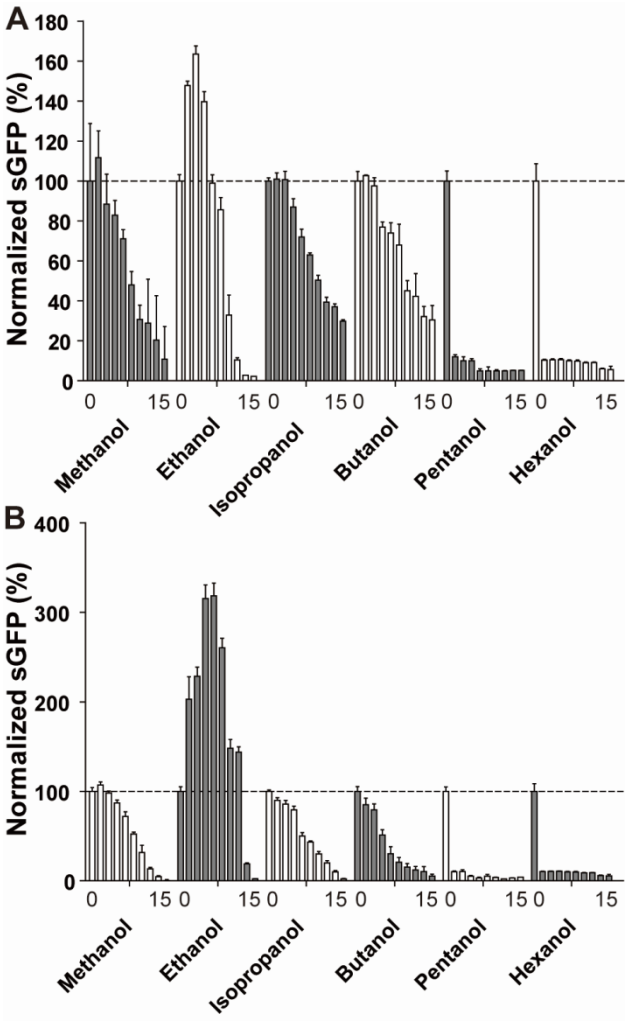


Figure 20. The effects of different alcohols on the expression of sGFP. The control without any alcohol is taken as 100%. The concentration range of each alcohol is from 0-15% (w/v).

A: Within this screening all the compounds are set up with their optima concentrations. Alcohols are screened with changes in their final concentrations in the CF expression system.

B: In this screening, the Mg^{2+} concentration was reduced compared to the optimum concentration. In the control reaction without any alcohol, the expression of sGFP was lower than that control in (a).

4.2.1.2 Effect of sugars and polyols

Several commonly used sugars and polyols, including sucrose, glycerol and D-trehalose and so on are selected for this experiment. The effects of these compounds are listed in Fig. 21. As shown in the figure, this type of chemicals behaves in the CF expression system quite similar. The expression yield of the sGFP in CF system linearly decreased with the increasing concentration of the tested sugars and polyols. However, CF system showed better tolerance to sucrose than all the others. With up to 10% sucrose the CF system still reserved 80% of the expression yield.

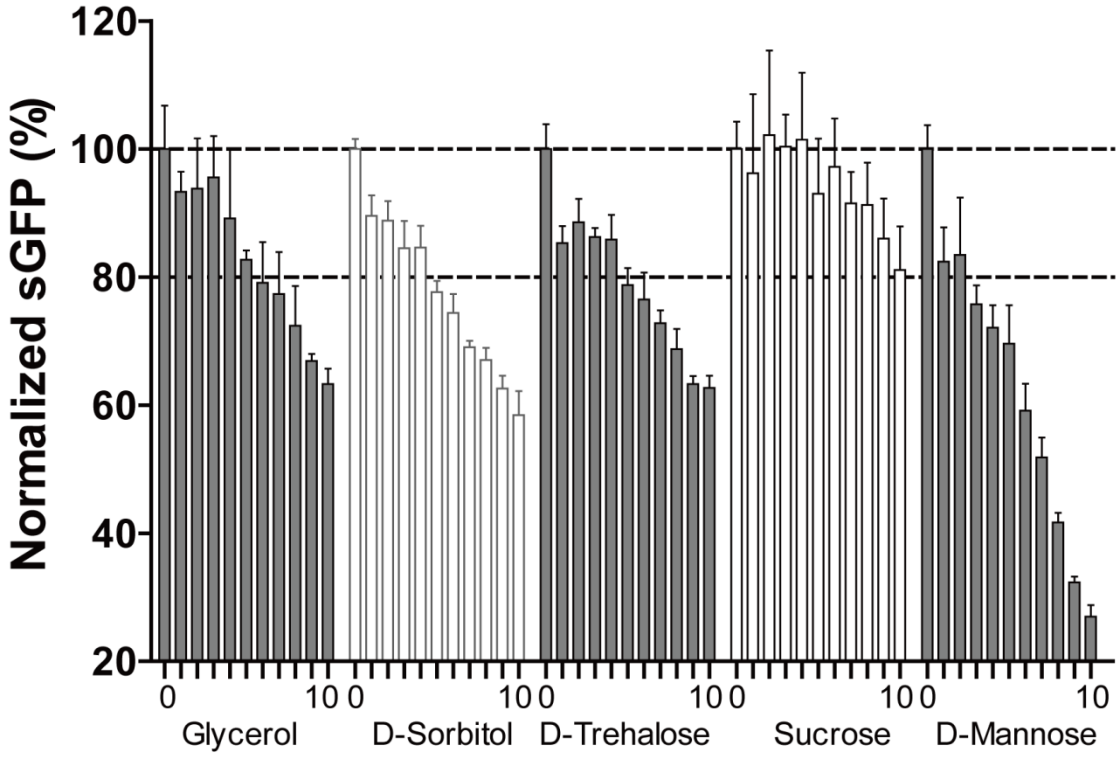


Figure 21. Effect of different sugars and polyols on CF expression of sGFP. The screening range of the corresponding chemicals is from 0-10% (w/v).

4.2.1.3 Effect of betaine and other natural osmolytes

Natural osmolytes are known to be effectively protected the cells and proteins under extreme conditions like: high salt, high temperature, high pressure and so on. We selected several of them except those sugars and polyols to be tested in the CF system. As shown in Fig. 22, betaine and ectoine both showed a very high tolerant concentration range. Up to 200 mM for betaine, 150 mM ectoine, more than 80% of the expression yields were obtained. With choline, there is even a beneficial effect on the expression, around 30% more expression yield was achieve with around 10 mM choline.

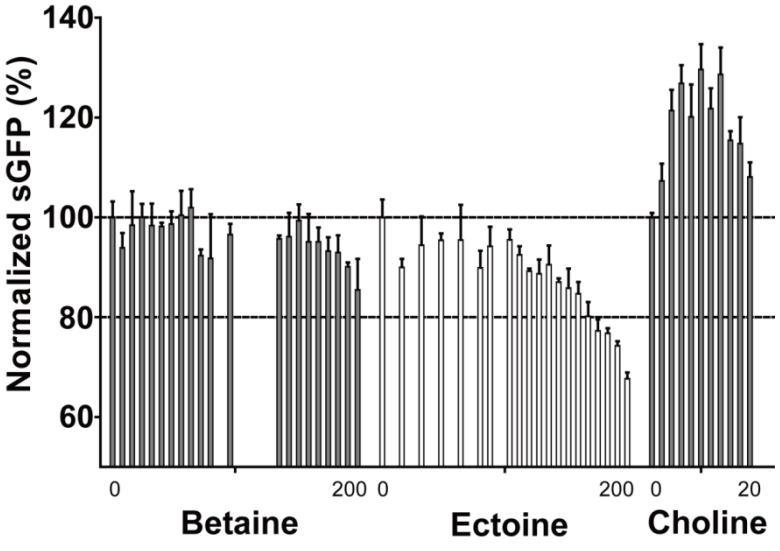


Figure 22. Effect of several natural osmolytes on CF expression of sGFP. The screening range of the corresponding chemicals is indicated in the figure with unit of mM.

4.2.1.4 Effect of amino acids

Amino acids and their derivatives are also known to be one class of osmolytes. They

also greatly stabilized certain kind of protein. Well known is Arginine, which is frequently used in buffers to stabilize protein after expression and isolation. Here we selected several of them to be tested in the CF expression system, including L-agrinine, N-acetyl-L-Lysine, L-Carnitine, Sarcosine, *trans*-OH-L-Proline. A buffer system of 200 mM HEPES, pH 8.0 was used when the pH of certain compounds is not neutral. As shown in the following figure, nearly all the amino acid showed a slightly beneficial effect on the sGFP expression.

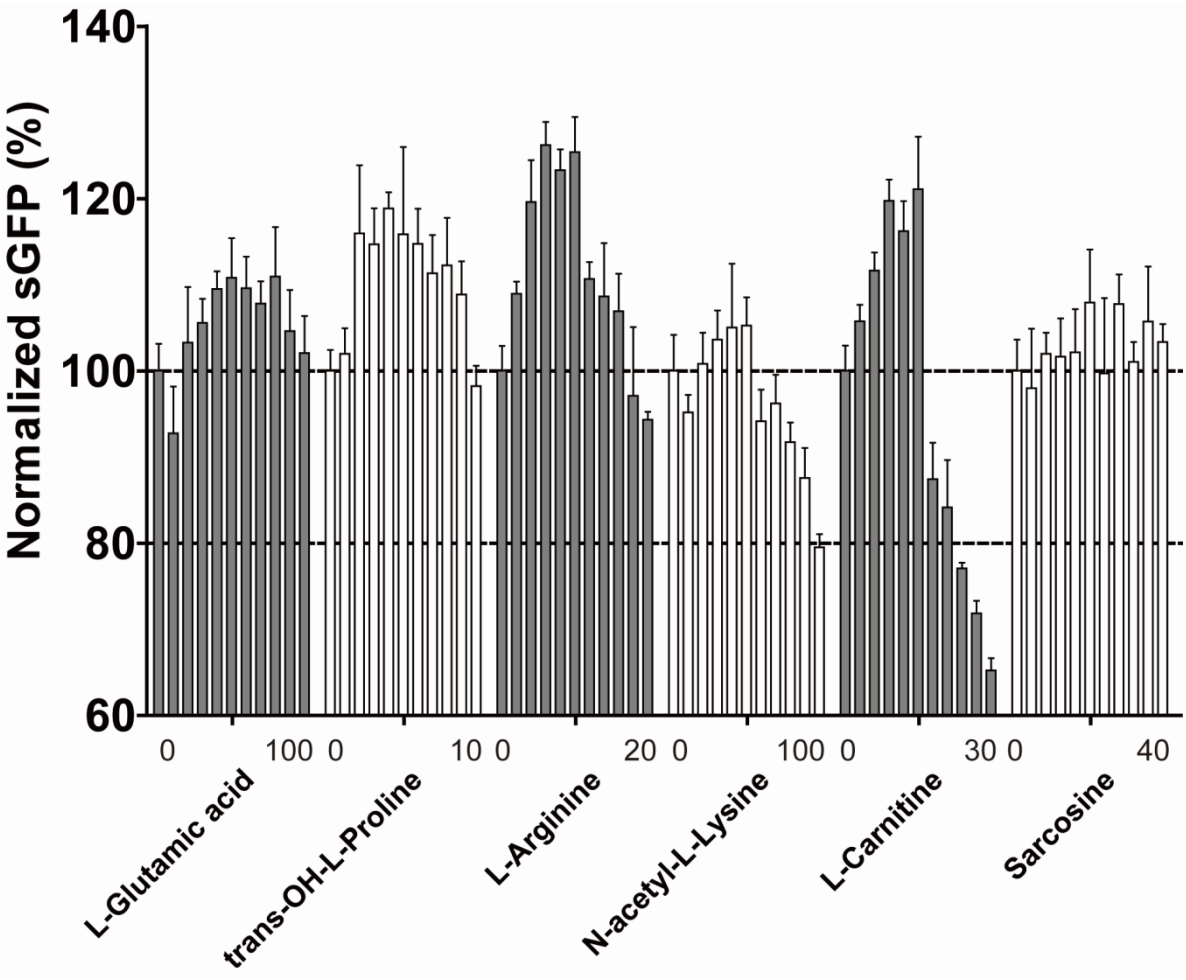


Figure 23. Effect of several amino acids and their derivatives on CF expression of sGFP. The screening range of the corresponding chemicals is indicated in the figure with concentration unit of mM.

4.2.2 Correlated screening of two compounds using sGFP

PEG was frequently used molecular crowding reagent and PEG8000 was used in several CF expression systems to mimic the *in vivo* environment. We select all the PEGs and test their effect on this batch CF expression system. As mentioned in the basic screening of batch CF system above, two compounds may interact with each other when used together. In this experiment, different Mg^{2+} concentration (low and high) and one other compound were selected to perform the screening.

At 20 mM Mg^{2+} concentrations, PEG slowly reduces the expression efficiency starting from 2% final concentration after an initial slightly increase. At the lower 14 mM Mg^{2+} concentration the

effect is reversed and the expression efficiency is rapidly increased with a peak at concentrations of approx. 4% PEG. This observation can be explained by the general effect of PEG to attract water thus making other reaction compounds more readily accessible. 14 mM Mg^{2+} was suboptimal conditions and the increase in available Mg^{2+} ions rapidly increases expression efficiencies until the optimum in between 20-22 mM is reached. Vice versa, at the already almost optimal Mg^{2+} concentration of 20 mM, the additionally available Mg^{2+} ions result into an excess that again decreases the expression efficiency. PEG therefore substitutes for part of the Mg^{2+} ions and vice versa, increased Mg^{2+} concentrations make PEG addition dispensable.

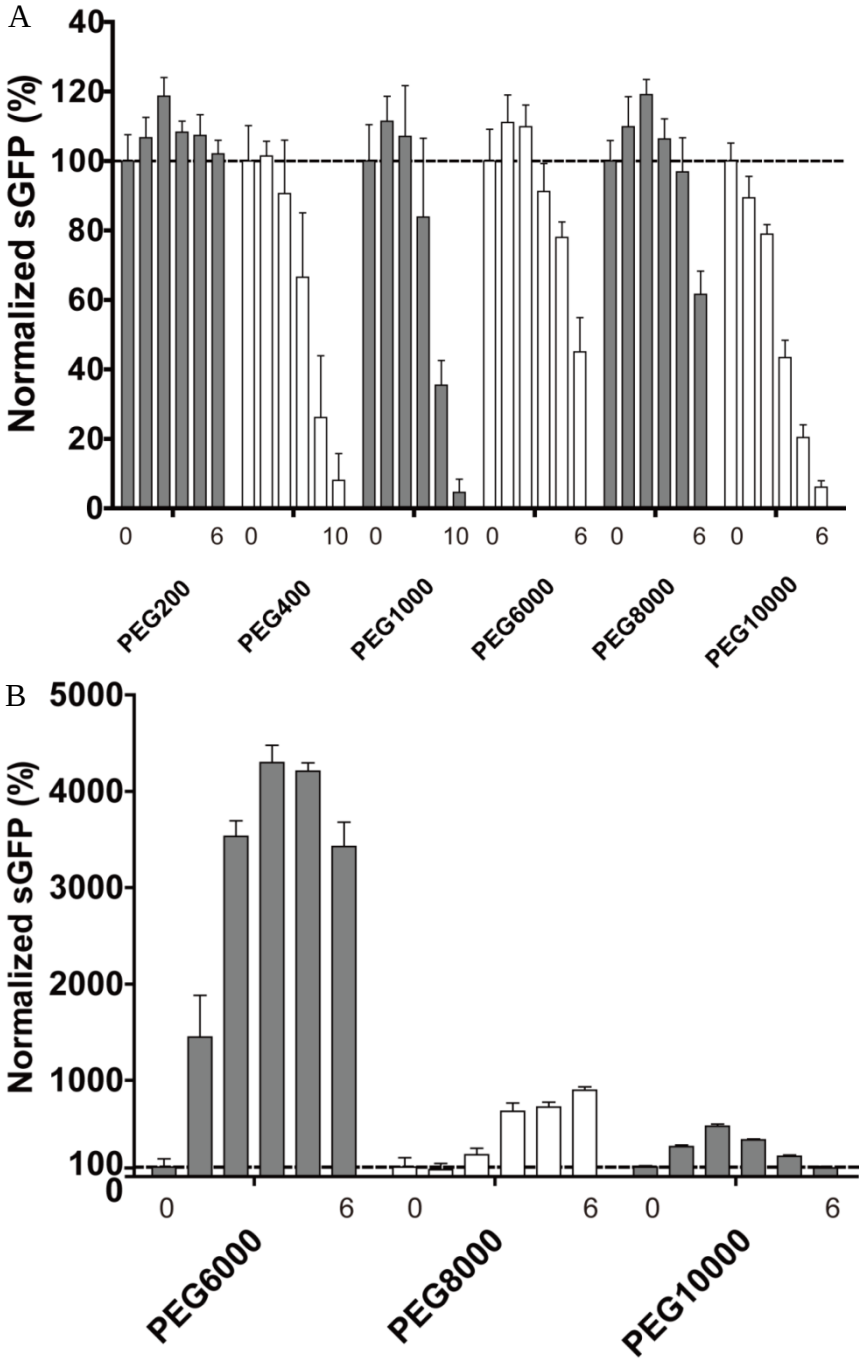


Figure 24. Effect of PEGs in the expression of sGFP. Different Mg^{2+} concentrations (16 mM, 26 mM)was used in A, B.

In the following table we summarized all the results of the chemical stabilizer screening for sGFP in batch CF expression system, screening concentration range of each compound was given and the effect of the compounds also given as positive, negative and tolerant (Tab. 7).

Table 7. Effects of different chemical compounds on the CF expression of sGFP

Compounds	Concentration range	Effect	Optimum/ Tolerant concentration
Methanol	1-15%	T	5.6%
Ethanol	1-15%	P	3.75%
Isopropanol	1-15%	T	3.75%
Butanol	1-15%	T	n.a
Pentanol	1-15%	N	n.a
Hexanol	1-15%	N	n.a
Glycerol	1-10%	T	4%
D-sorbitol	1-10%	T	4%
D-trehalose	1-10%	T	4%
Sucrose	1-10%	T	10%
D-mannose	1-10%	T	2%
Betaine	0-200 mM	T	250 mM
Ectoine	0-200 mM	T	150 mM
Choline	0-20 mM	P	10 mM
L-Glutamic acid	0-100 mM	P	50 mM
Trans-OH-L-Proline	0-10 mM	P	5 mM
L-Arginine	0-20 mM	P	10 mM
N-acetyl-L-Lysine	0-100 mM	P	100 mM
L-Carnitine Cl	0-30 mM	P	5 mM
Sarcosine	0-40 mM	P	40 mM

4.2.3 Stabilizer screening for better soluble expression of GNA1 and Halogenase

4.2.3.1 Linear screening of potential chemical stabilizer for GNA1 and Halogenase

Two soluble proteins GNA1 and halogenase, which form precipitate either *in vivo* or *in vitro*, were selected as models for testing the effect of the above mentioned compounds. sGFP fusion of GNA1 was made for fast soluble expression quantification. We select several compounds for the first screening as shown in Fig 26. In order to confirm if the fluorescence signal is correlated with the enzymatic activity of GNA1-sGFP, a functional assay was performed in parallel as shown in Fig 25. The results showed that the fluorescence signal was not always correlated with the enzymatic activity. Like the case of Trehalose, especially L-Arginine, the increase of fluorescence does not finally results in the same increase of enzymatic activity. This might due to the fact that these reagents are beneficial for the folding of sGFP. However, in the case of choline, both the fluorescence and enzymatic activity were increased by 20%.

Further experiment was performed to compare the total protein concentration, fluorescence signal (which we consider as soluble folded sGFP) and the enzymatic activity of the GNA1-sGFP as indicated in Fig.27. The reaction mixture of the CF expressed GNA1-sGFP was dialysis against enzymatic buffer, and then used for enzymatic assay. Positive control without any additives and negative control without any DNA template were made for each screening experiment. As shown in

Fig. 27, addition of choline gave an increase of around 20% percent soluble expression, 7% increase of total protein expression, and 20% increase of enzymatic activity. However, in the case of Arginine, the increase of fluorescence did not result in the increase of enzyme activity. The total protein expression was reduced by 8%, which was consistent with around 8% decrease in the enzymatic activity. This experiment confirmed that Arginine might have a beneficial effect on the folding of sGFP but not the GNA1.

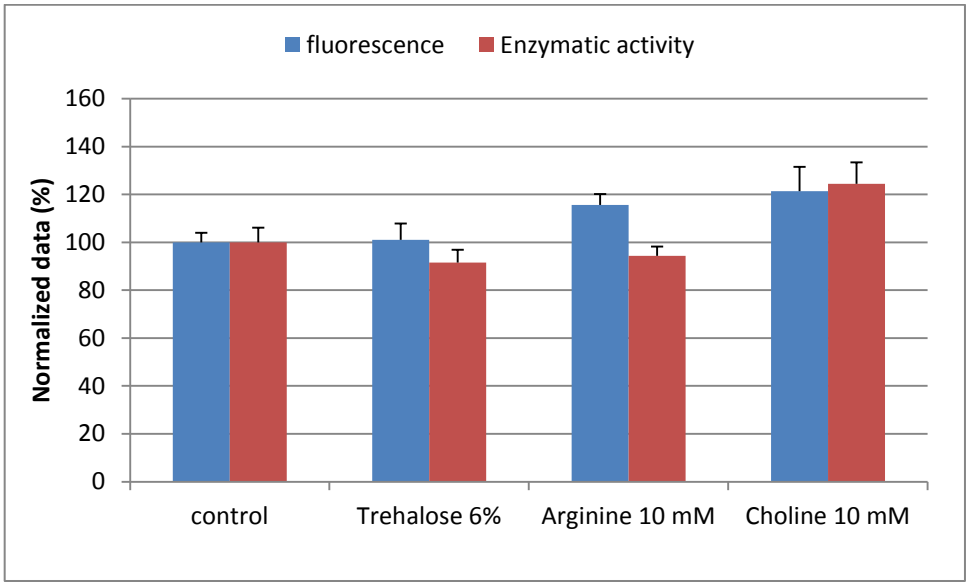
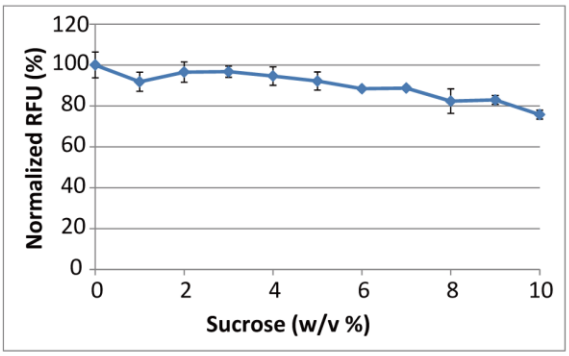
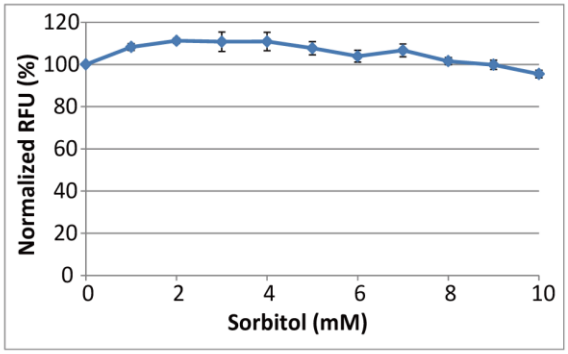
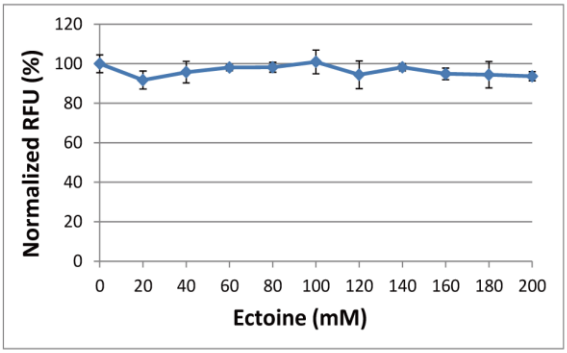
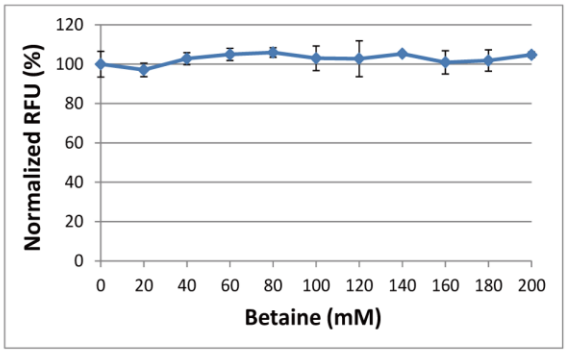


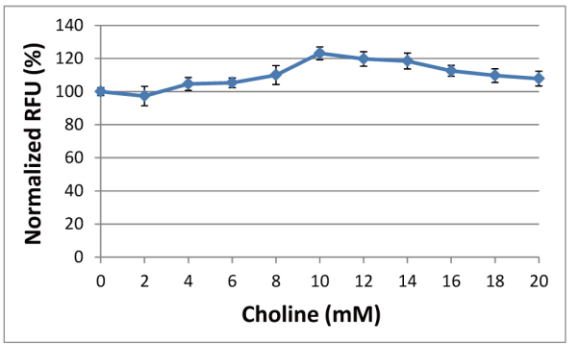
Figure 25. Effect of selected chemical stabilizers on the soluble expression and enzymatic activity of GNA1-sGFP.

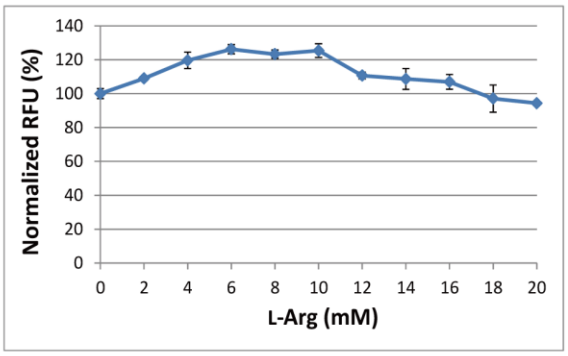












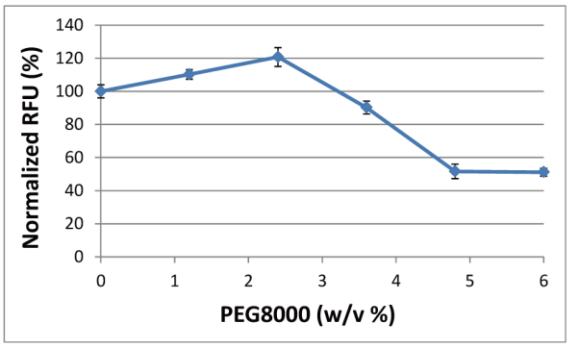


Figure 26. Effect of selected chemical stabilizers on the soluble expression of GNA1-sGFP. All the chemicals was added with the final concentration indicated in the figure. In case of charged compounds like choline and Arginine which affect the pH, a HEPES buffer system was used with additional glutamic acid.

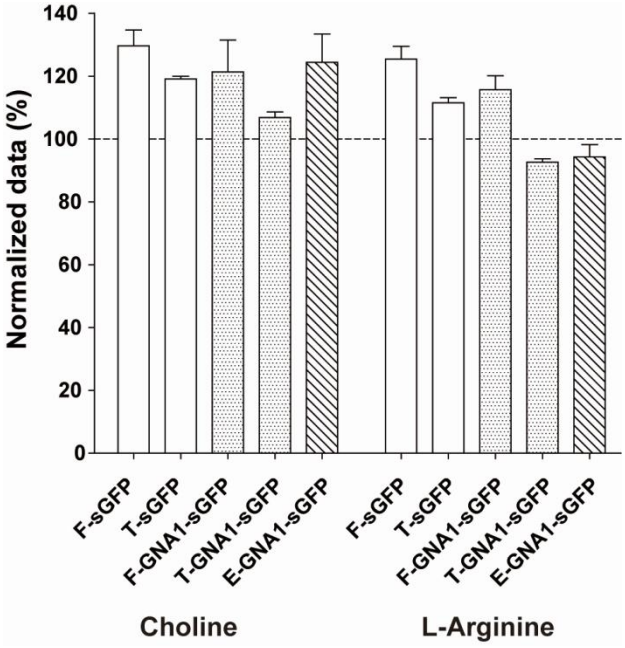


Figure 27. Effect of Choline and L-Arginine in the total protein expression, fluorescence of sGFP and GNA1-sGFP and enzymatic activity of GNA1-sGFP.F: Fluorescence; T: Total protein concentration measured by radioactive labeling; E, enzymatic activity.

Another experiment with Halogenase was performed. Result for halogenase was shown in Fig. 28, soluble fractions was collected after expression and applied on to western blot for quantification. Results showed that choline and arginine gave increase of around 20% and 10% in soluble expression.

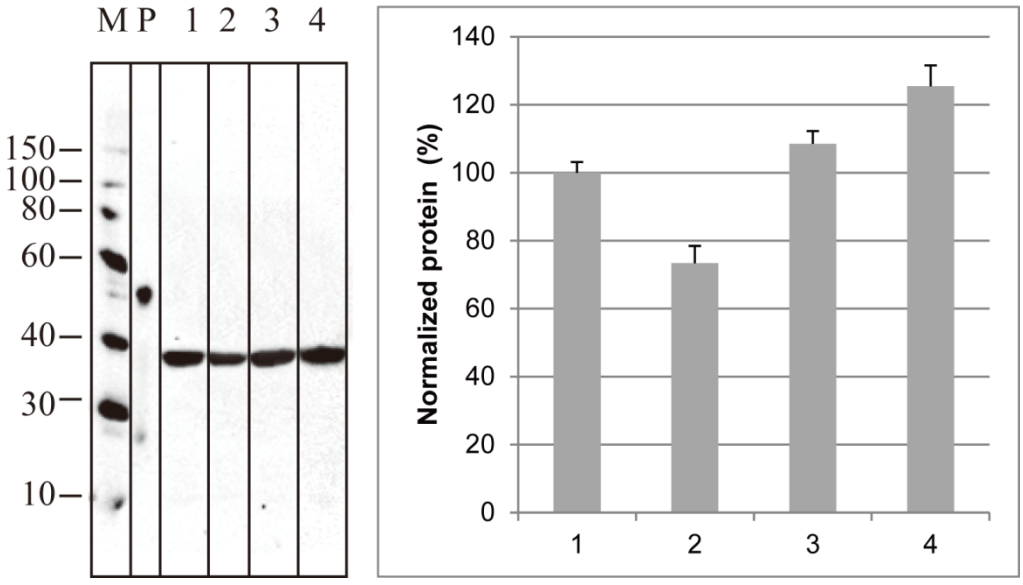


Figure 28. Screening of several compounds on the soluble expression of Halogenase. M, molecular marker; P, 1, 2, 3, 4 represent: positive control for quantification (Positope™, invitrogene), control, 6% Trehalose, 10 mM Arginine, 10 mM Choline. The band intensity was used to quantify the protein concentration by quantity one (Bio-Rad).Sample was collected after 2 hours expression and transferred into eppendorf tubes. After centrifuge at 22,000xg for 10 min, 2 µl of supernatant was used for western blotting. 5 µl of positope was loaded.The results was normalized taken control as 100%, which corresponding to protein concentration of 80 ng/µl (calculated by comparison with Positope). The normalized data were: 100%, 73.3%, 108%, 125% for lane 2-5.

4.2.3.2 Correlated screening of PEG and choline for GNA1-sGFP

In order to check if the beneficial of two stabilizers or expression enhancer had a accumulative effect on the specific target, GNA1-sGFP was chosen and subject into the correlated screening. Two compounds, PEG8000 and choline, were selected for this experiment. As shown in Fig. 29, the beneficial effect of PEG8000 and choline is accumulative, which is better than using one of them alone. There was more than 100% increase in the fluorescence than the control without any chemical compounds. By using PEG8000 and choline alone, there was around 30% increase in the fluorescence. However, when using together, the increase was more than 60%, but more than 100%. This result showed that not only accumulative beneficial effect was observed, but also a beneficial increased beneficial effect was achieved.

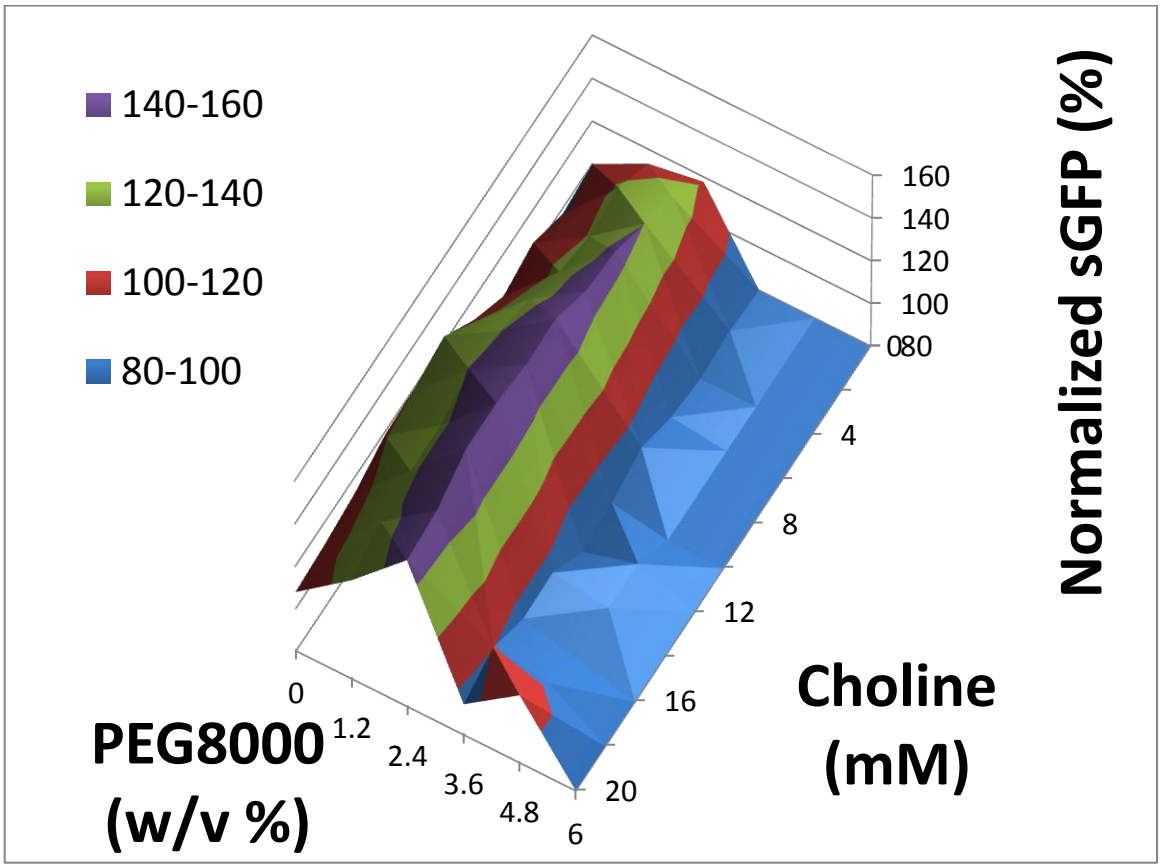


Figure 29. Correlated screening of PEG8000 and choline for the soluble expression of GNA1-sGFP.

4.3 Application of throughput cell-free expression system on membrane proteins

Sufficient yield in combination with high quality of the target MP are the two key benchmarks which have to be approached by a multiple screening strategy. Fast and efficient monitoring systems would be helpful in order to accelerate protocol development. Derivatives of green fluorescent protein (GFP) have been proposed as expression and folding monitor for MPs synthesized in cellular systems [143]. In MP-GFP fusions, the C-terminal GFP moiety appears to have even sometimes a correlated folding behavior with the attached MP. GFP fusions could thus be used (I) for monitoring of MP expression rates, (II) as preliminary folding indication of the attached MP and (III) for fast tracking and establishing MP purification protocols.

For the implementation of GFP as monitor in CF systems different aspects have to be considered. In the P-CF mode, the synthesized MP should not be able to fold into functional conformations as no hydrophobic environments for its stabilization are present. The MPs therefore instantly precipitate and consequently direct the attached GFP into the pellet. If the folding of the two partners in the MP-GFP fusion is correlated, the precipitated GFP should be non-fluorescent, which indeed is usually the case [144].

4.3.1 D-CF expression of membrane proteins using different sGFP derivatives fusion as expression and folding monitor

In the D-CF mode, detergents are supplied which instantly solubilize the freshly translated MP-GFP fusions into micelles. The functional folding of a number of MPs after expression in the D-CF mode has already been shown [145, 146]. However, unfortunately the folding of GFP is specifically affected by the presence of detergent [102]. In many commonly used detergents such as DDM, Digitonin or Triton X-100, the folding of GFP is almost completely abolished resulting in only low fluorescence signals (Fig. 30). Only in the mild detergents Brij-58, Brij-78, Brij-98, in the amphipol Nvoy10 and in decyl-MNG 30% - 60% of the fluorescence signal can be recovered while the rest of the GFP fraction remains in a non-fluorescent conformation. The detergent induced GFP misfolding is non-reversible and the protein stays non-fluorescent after detergent removal. However, once folded, GFP is highly resistant against unfolding in presence of a large variety of detergents [57, 102]. Nevertheless, the residual activity of GFP can still be sufficient for throughput expression screening in the D-CF mode as shown with the CF expression profiling of a library of E. coli inner MPs [102]. However, future applications might rather implement the recently developed derivative superfolderGFP having improved stability [147]. The overall detergent resistance of superfolderGFP is significantly increased and recovery of fluorescence signals exceeded 70% with most analyzed detergents.

In order to test the detergents tolerance of these two different GFP derivatives, experiment was performed using sGFP and superfolderGFP as template in the CF expression system in presence of different detergents. After expression, the fluorescence of the CF reactions was measured. As indicated in Fig. 30, the comparison of sGFP and superfolderGFP with most commonly used detergent for MPs in CF expression systems showed that the detergents compatibility of superfolderGFP was much better than that of sGFP, which only showed good folding in Brij detergents and Decyl-MNG. In contrast, superfoderGFP showed much better tolerance to different kind of detergents in CF systems. Even those detergents like TritonX-100 and DDM which are not suitable for D-CF expression with MP-sGFP fusions are now possible to use with superfolderGFP fusion.

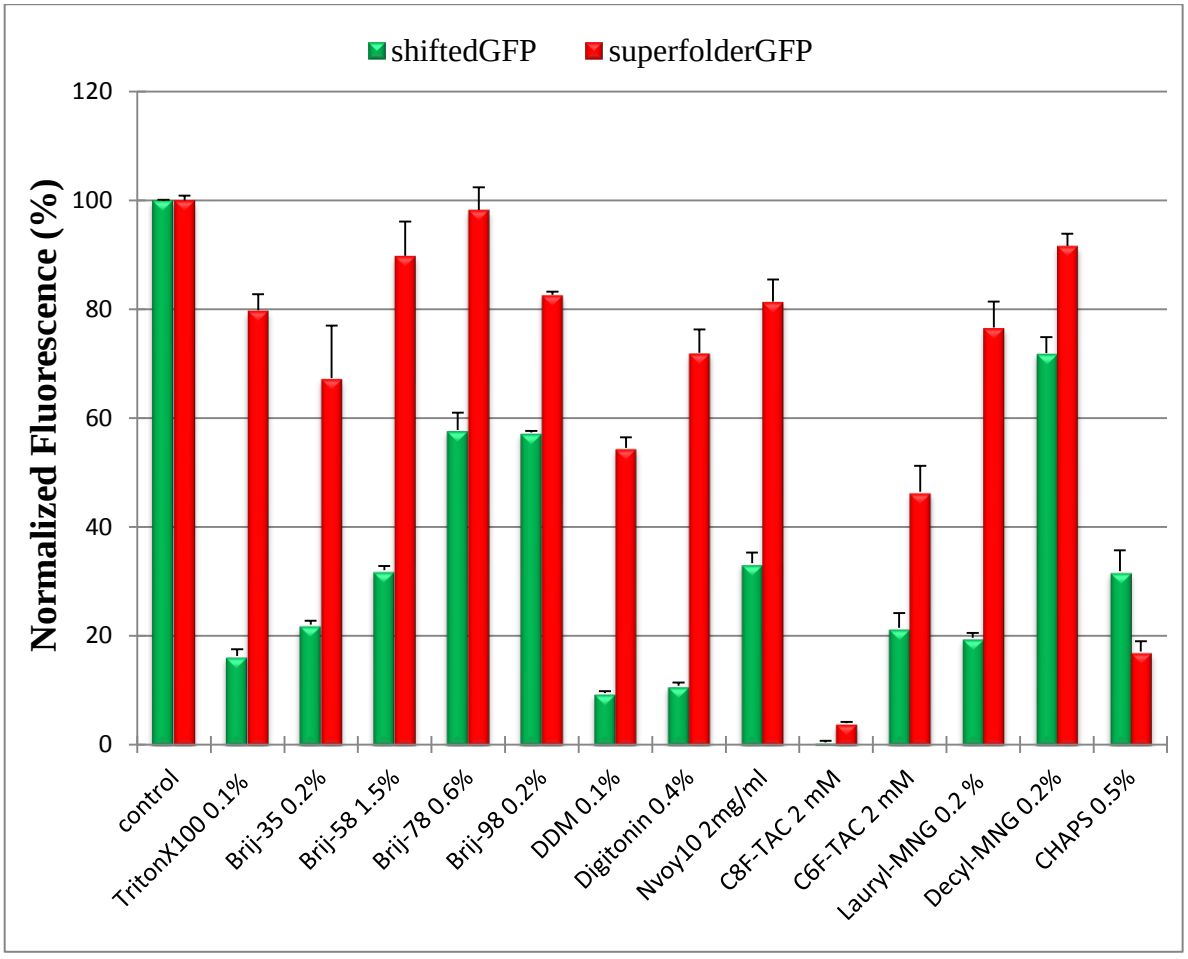


Figure 30. Comparison the folding of sGFP and superfolderGFP in presence of detergents via CF expression systems. Concentrations shown in percentage were w/v. Control was performed without any detergent and was taken as 100% for the normalization.

Further experiment with a MP target was performed. SugE, which is a small multidrug transporter, was used as a model protein. Different constructs with sGFP or superfolderGFP fusion in the C-terminal of the SugE gene were made. Then, these two construct were used in the CF expression with different detergents. After expression, the supernatants of the CF reaction mixtures were used for fluorescence measurement. As shown in Fig.31, SugE-superfolderGFP showed quite

good fluorescence signal in many detergents, including Brij-35, Brij-78, Brij-58, Digitonin, Laury-MNG, and Decyl-MNG. In contrast, SugE-sGFP only showed good fluorescence signal in Brij-78 and Brij-58. This experiment gave the idea that D-CF expression of MPs using the GFP as monitor is largely depending on the folding of GFP itself. In quite some cases, the MPs can be expressed as soluble form in certain detergents, which were not tolerant by the fusion GFP.

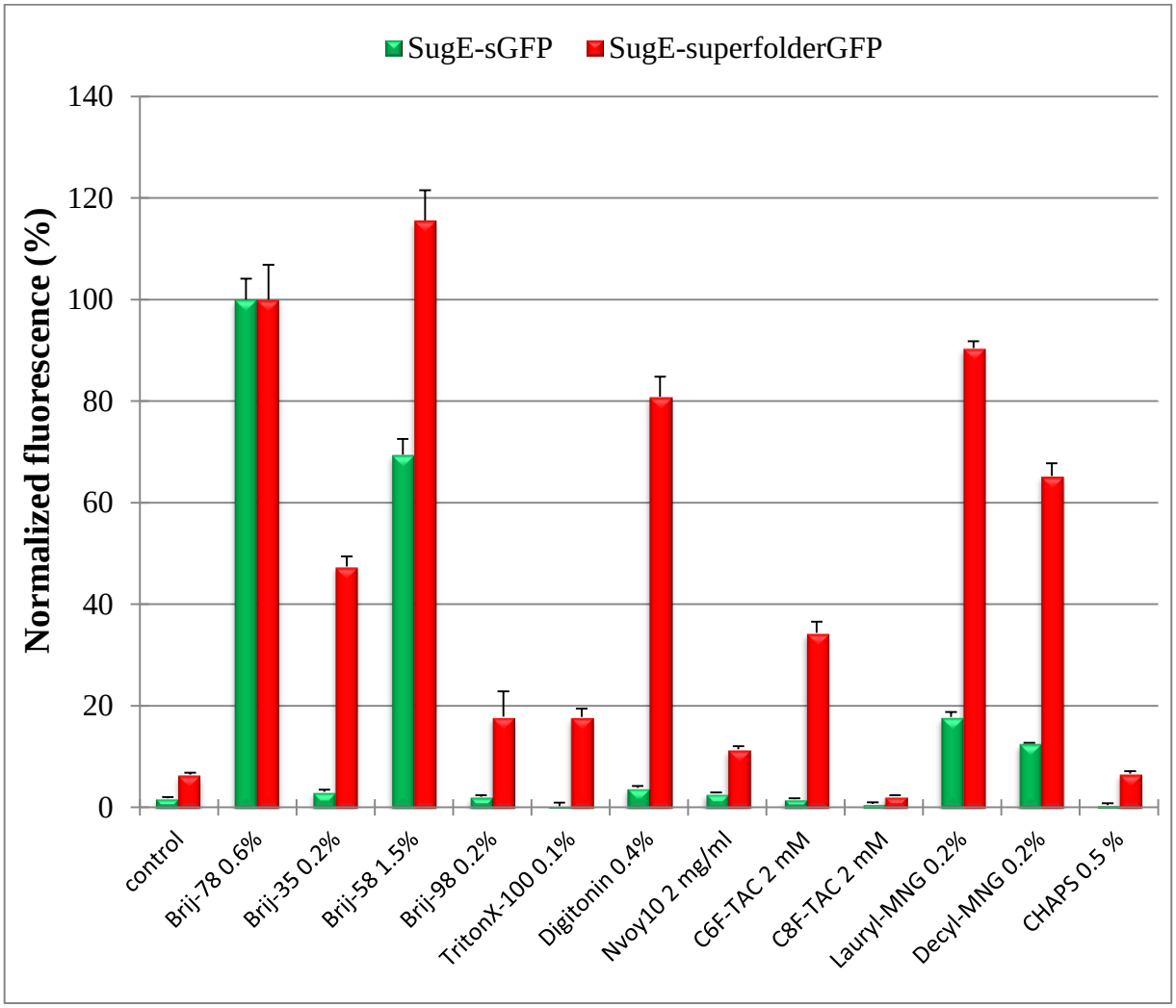


Figure 31. Comparison of the cell free produced SugE-sGFP and SugE-superfolder GFP in presence of different detergents. The fluorescence signal was normalized and the value of SugE-sGFP and SugE-superfolderGFP in Brij-78 was set up as 100%. Control sample was performed without any detergent. Detergent concentrations shown in percent were all w/v.

4.3.2 Improving MPs translocation with lipid/detergent mixtures

In L-CF mode approaches, detergents are only present if initially bicelles or lipid/detergent mixtures are supplied into the RM. In contrast to detergents, lipids generally do not seem to have negative effects on the folding of GFP. However, some lipids might have negative effects on the protein expression efficiency depending on the supplied concentration. While transcription appears

to mostly benefit from the presence of different types of liposomes, the translation was enhanced by anionic and neutral liposomes and inhibited by cationic liposomes [92]. GFP fluorescence of MP-GFP fusions can certainly only taken as first indication of MP folding which requires further subsequent verification by specific assays in each individual case. A relatively fast approach is the analysis of L-CF synthesized proteoliposome with confocal microscopes which could indicate the co-localization of the MP-GFP fusion with the liposomes [148].

Complete or partial solubilization of lipid bilayers by detergents could improve their availability and interaction with synthesized MPs. Corresponding Lipid/detergent mixtures are prepared *in vitro* in fixed ratios (Tab. 8) and they can be implemented for the soluble CF expression of MPs in two different approaches. One strategy is to add a certain amount of detergent only into the RM in order to initially solubilize the supplied lipids. During incubation, the detergent will than constantly be diluted out from the RM by diffusion into the FM. The kinetics of dilution will depend on the specific CMC and micelle size of the detergent. Mixtures of lipids and detergents form water-soluble assemblies most likely containing solubilized individual lipids as well as variable patterns and stretches of solubilized bilayers. The first six hrs of CECF reactions are most productive and some 80% of the final MP will be synthesized during that time [149]. The detergent will support the solubilization of the synthesized MPs and could promote their association with the solubilized lipids. During progress of the reaction, the detergent concentration will constantly be reduced and the forming liposomes may include the associated MPs. With this approach, a better folding of bacteriorhodopsin was obtained if initialyy steroid detergents such as cholate, CHAPS or digitonin have been provided [149].

Alternatively, the detergent concentration can be kept constant by its addition in RM and FM. Relatively well analyzed are bicelles formed by mixtures of DHPC and DMPC and consisting of central lipid discs surrounded by the detergent [150]. For preparing bicelles, the lipids ratio, q (the molar ratio of long- to short chain lipids or detergent) is crucial for the size of bicelles and need to be optimized for specific MP targets [151]. One of the most important parameter for structural studies is the stability of the samples. MPs in bicelles tend to precipitate with halftimes of some 3-4 days for q = 0.33 bicelles. One reason for this instability could be ester hydrolysis of phospholipids. This can be solved by using the more stable ether-linked analogues of DMPC and DHPC [152]. Another alternative solution is introducing a fraction of negative charged phospholipids, which will increase the repulsion between bicelles and minimize protein aggregation [153]. Several MPs like ATP synthase, bacteriorhodopsin were functionally L-CF produced in presence of bicelles [149, 154]. In addition, recent studies showed that NMR analysis of MPs in bicelles is feasible [153, 155, 156].

Table 8. Preparation of lipid/detergent mixtures for L (D)-CF expression approaches

Lipid	Detergent	Detergent conc. (%)[1]	Lipid:detergent molar ratio
DMPC (4 mg/ml)	CHAPS	0.15	2.42
	Digitonin	0.5	1.45
	Triton-X100	0.25	1.56
	Brij-35	0.04	17.70
	Brij-78	0.2	3.40
	DPC	0.2	1.03
	DDM	0.5	0.59
DMPC (10 mg/ml)[2]	DHPC	2	0.36
Egg PC (6.7 mg/ml)[3]	β-OG	1	0.25
	Triton X100	0.03	18.77
	DDM	0.02	22.23
	Brij-58	0.02	48.90
	Digitonin	0.2	5.35
	Cholate	1	0.37
	CHAPS	1	0.54
	Deoxycholate	0.5	0.72
	PFO	2	0.18
	Lauryl sarcosine	1	0.26
	SDS	0.5	0.50

1, concentrations necessary for complete lipid solubilization

2, concentrations for bicelle formation [154].

3, [149]

4.4 Expression, purification and optimization of membrane proteins for functional and structural analysis

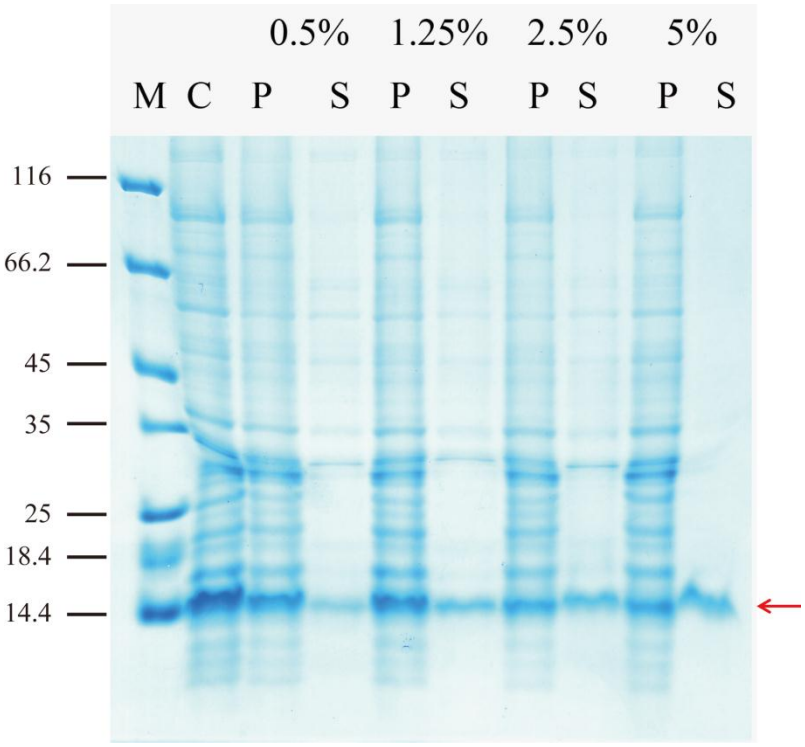
4.4.1 Investigation on the small multidrug resistance protein-SugE

Small multidrug resistance (SMR) proteins are integral MPs, located within the plasma membrane of most Bacteria and Euryarchaea [157], that confer resistance to quaternary cation compounds (QCC) via proton motive force [158]. As indicated in the name, these small (~12 kDa) integral inner MPs range from 100 to 140 amino acids in length, 4 α-helix transmembrane strands (TMS) [159]. Different from other multidrug transporter protein, the SMR family has only been proved transport of lipophilic compounds, primary QAC and a variety of antibiotics [158, 160, 161].

The SMR protein family is composed of three major protein subclasses: small multidrug proteins (SMP), suppressor of *groEL* mutations (SUG), and paired SMR proteins (PSMR). Among the three SMR subclasses, only SMP and SUG subclass members are capable of conferring isogenic QCC resistance to its host and members from either subclass are frequently encoded in conserved regions of integrons [162, 163] and on multidrug resistant plasmids [164, 165]. The narrow QCC spectrum of SugE is limited to antiseptics, specifically those with one or more elongated acyl chain R groups that range from C_{12} to C_{18} covalently bound to a single N cation [166]. More specifically, SugE can confer host resistance to the antispectics, cetylpyridinium ($C_{21}H_{38}N$), cetyldimethylethyl ammonium ($C_{20}H_{42}N$), and hexadecyltrimethyl ammonium ($C_{19}H_{42}N$) also known as cetrimide [166].

4.4.1.1Sample preparation of SugE produced in cell-free system

With the goal of functional and structural investigation of CF produced SugE, different CF expression strategies were applied to obtain large amount of pure protein sample in a soluble form. After first optimization of the basic conditions like Mg^{2+} and K^{+}, an expression yield around 2 mg per 1 ml reaction mixture of native SugE from *E. coli* without any tags was obtained in the P-CF mode. In order to resolubilize the pellet of SugE, different detergent was selected. According to the previous results, a two-step resolubilization strategy was used. Pellet sample from the reaction mixture was firstly washed with 20% $NDSB_{256}$, which found to solubilize the co-precipitate protein not SugE [81]. Finally, the SugE could be efficiently resolubilized in 2% DDM, when incubated for 2 hours at 37°C with shaking. Further experiment was performed with the commonly used detergents like β-OG and TritonX-100. These two detergents showed good selectivity, however, there were still relative large amount of proteins still remained in the pellet form after detergents treatment. As shown in Fig. 32, with 5% TritonX-100, SugE was selective resolubilized with very few impurities. However, still quite a lot of SugE protein remains in the pellet form. The same effect was observed for β-OG resolubilization.



Different resolubilized samples were checked by SEC, the results were shown in Fig. 33. From the results, the NDSB256 washed and DDM resolubilized sample gave the best peak shape, indicating the best sample homogeneity. However, β-OG directly resolubilized sample showed a very broad peak, while TritonX-100 showed a medium quality in the peak shape.

Figure 32. Tricine-SDS-PAGE of TritonX-100 resolubilized SugE. SguE were separated by Tricine-SDS-PAGE (12%). M: Protein marker; C, Control pellet without any detergent treatment. P, pellet after detergent treatment; S, supernatant after detergent treatment.Percentage shown on top was w/v.

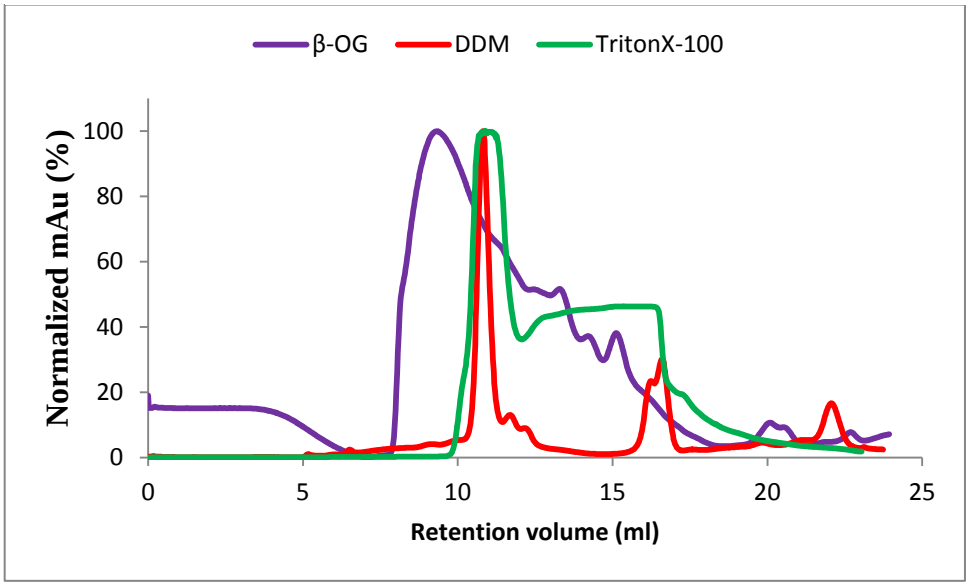


Figure 33. Elution profiles of detergent resolubilized SugE sample on a Superdex 200 10/300 gel filtration colume.

4.4.1.2 Sample preparation of SugE for crystallization trials

In order to prepare the sample for crystallization, a stable and homogenous sample with high concentration was required. The protein sample was concentrated via a centricon centrifugal filter device. Concentration was performed step by step and sample was taken for each step and the protein concentration was measured. The concentrated sample was checked by Tricine-SDS- PAGE as shown in Fig. 30. The concentrated sample was store at 4^{o}C to check the stability. Results

showed that the concentrated sample was stable at 4°C with concentration around 10 mg/ml. The sample of SugE already started to precipitate after stored at 4°C for 2 days at the concentration of 15 mg/ml.

Since there were still quite some impurities in the resolubilized sample and there are no affinity tags in this protein construct, a further semi-preparative scale SEC was performed in order to purify the SugE protein sample. Resolubilized SugE protein sample in 2% DDM was loaded on to the Superdex 200 10/300 column. As indicated in Fig. 35 after collecting the SugE protein peak fraction, the SDS-PAGE showed a better purity compared to the loaded sample in Fig. 34. The collected sample from the semi-preparative gel filtration column was pooled and concentrated again to the concentration around 10 mg/ml.

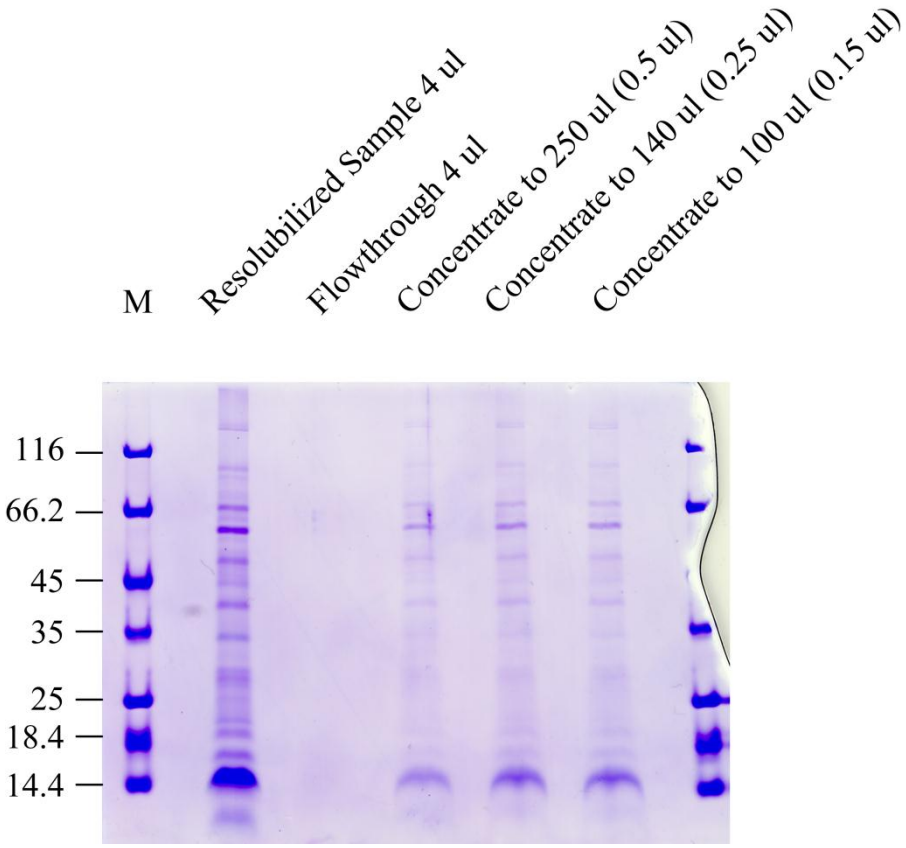


Figure 34. Tricine-SDS-PAGE of concentrating the resolubilized SugE sample in 2% DDM using a Centricon centrifugal filter device. The starting volume was 1.2 ml resolubilized SugE with a protein concentration around 1.6 mg/ml. The flowthrough from the filter device was also loaded on the gel to check if there are SugE samples in it. The volumes indicated in the labeling were the final volume remained after concentrating. And the volume in the brace was the volume of concentrated sample that was loaded on to the gel. M, Protein marker.

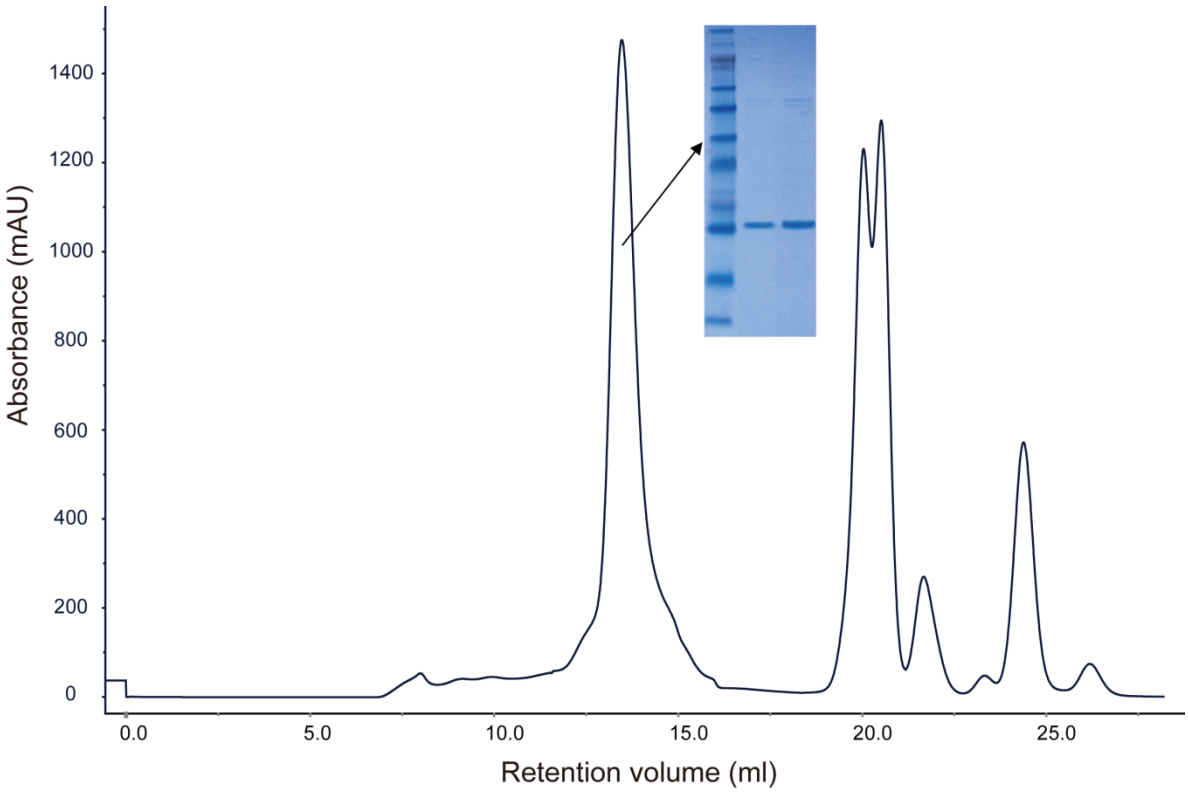


Figure 35. Profile of concentrated SugE samples in DDM on a Superdex 200 10/300 semi-preparative colume. Arrow indicated peak fraction was collected, pooled, and analyzed by gradient SDS-PAGE gel. Running buffer contained 0.05% DDM

Detergents were known to be important for MPs crystallization. Therefore, the experiment was performed to check if SugE protein was stable and homogenous in other commonly used detergents. Since the construct currently using was without any affinity tags, the detergent exchange was performed on the gel filtration column, which was pre-equilibrated with the exchange detergent. Maltosides and glucosides with different Alkyl chains were chosen because previous result showed that SugE stay stable in DDM and β-OG. SEC elution profile of SugE detergent exchange was shown in Fig. 36, results showed that NG, DDM, C8E5 and C10E5 gave the best SugE peak shapes, which indicated homogenous samples.

Stability test was performed in 4^{o}C, SugE samples only showed good stability in DM and DDM. In other detergent, SugE protein started to precipitate within 1-3 days. Therefore, we chose the sample in DDM and DM for the first crystallization trials.

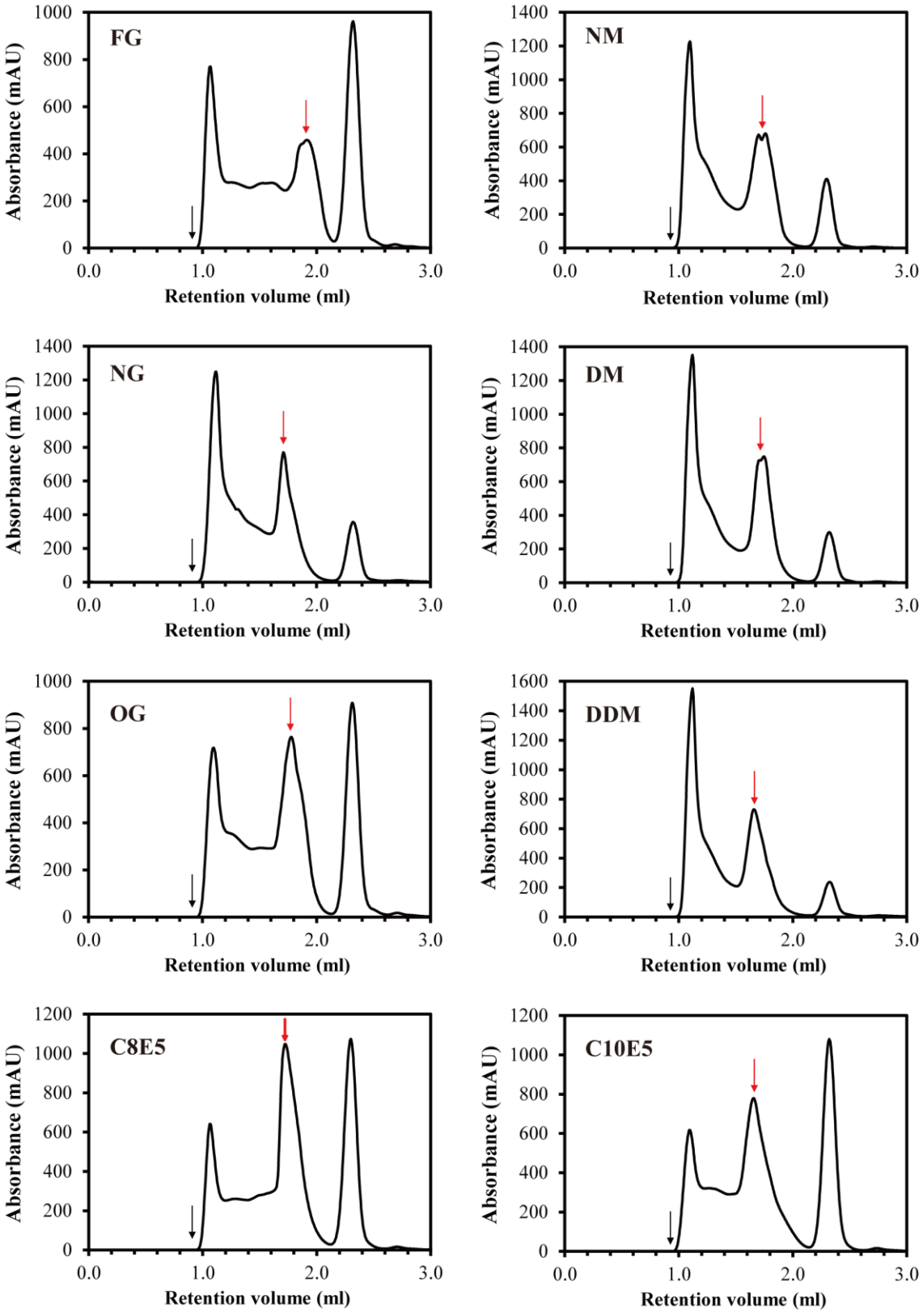
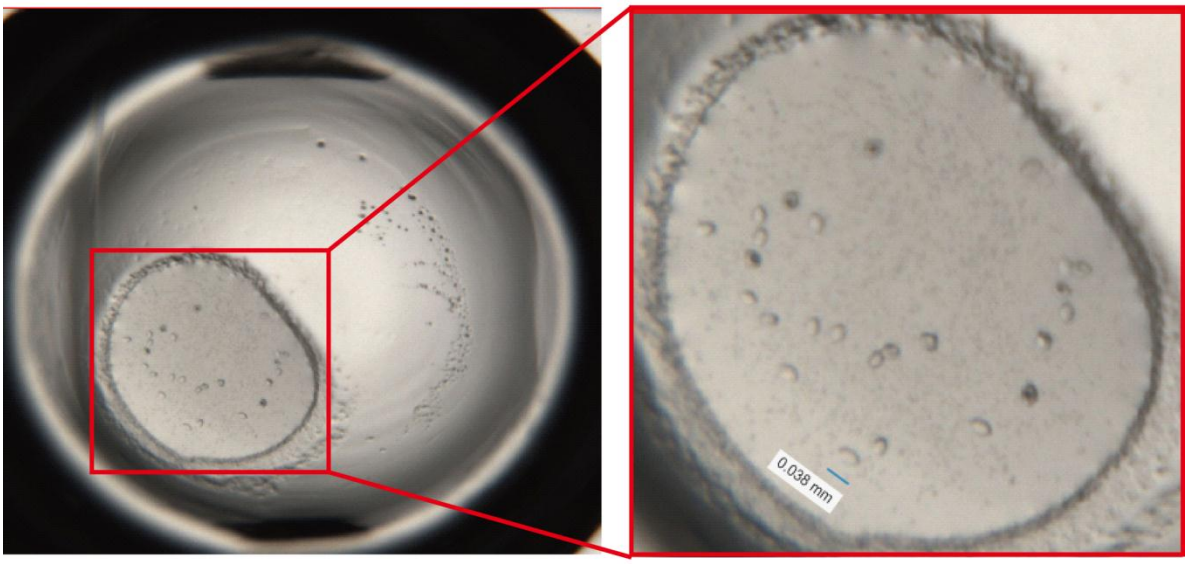


Figure 36. Detergent exchange on a SEC using Superdex 200 3.2/30 gel filtration column. Resolubilized SugE samples in 2% DDM was loaded on to the gel filtration column, pre-equilibrated with the indicated secondary detergent. The running buffer only differed in the corresponding detergent: 2% FG, 0.3% NG, 0.9% β-OG, 0.4% NM, 0.25% DM, 0.05% DDM, 0.3% C8E5, and 0.05% C10E5. The black arrow indicated the void volume and the red arrow indicate the SugE protein peak.

Several commercial MP crystallization kits were used for the first condition screening and a custom designed pH screening was performed. Results showed that SugE was quite stable from pH 5.1 to 9.1, no significant precipitation was observed. There are several conditions, which showed microcrystals, as shown in Fig. 37. However, when check by X-ray they are all detergent crystals not protein.

A



B

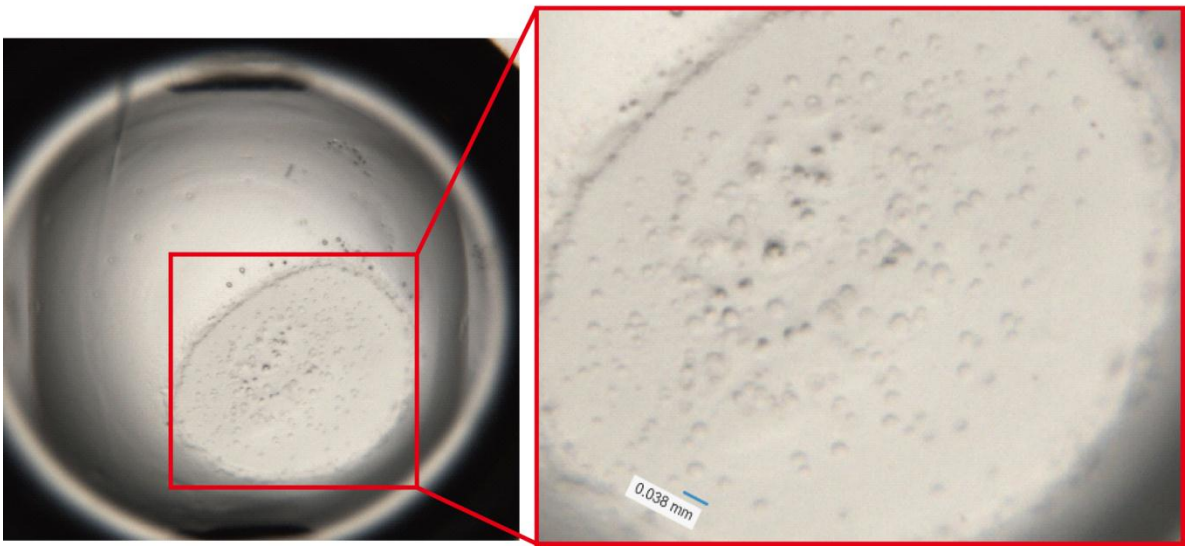


Figure 37. Crystallization conditions where microcrystals appeared.

4.4.2 Preparative scale production of functional mouse aquaporin 4 using different cell-free expression modes

The continuous progress in the structural and functional characterization of aquaporins increasingly attracts attention to study their roles in certain mammalian diseases. Although several structures of aquaporins have already been solved by crystallization, the challenge of producing sufficient amounts of functional proteins still remains. CF expression has emerged in recent times as a promising alternative option in order to synthesize large quantities of MPs, and the focus of this report was to evaluate the potential of this technique for the production of eukaryotic aquaporins. We have selected the mouse aquaporin 4 as a representative of mammalian aquaporins. The protein was synthesized in an E. coli extract based CF system with two different expression modes, and the efficiencies of two modes were compared. In both, the P-CF (CF MP expression as precipitate) mode generating initial aquaporin precipitates as well as in the D-CF (CF MP expression in presence of detergent) mode, generating directly detergent solubilized samples, we were able to obtain mg amounts of protein per ml of CF reaction. Purified aquaporin samples solubilized in different detergents were reconstituted into liposomes, and analyzed for the water channel activity. The calculated Pf value of proteoliposome samples isolated from the D-CF mode was 133 µm/s at 10^{o}C, which was 5 times higher as that of the control. A reversible inhibitory effect of mercury chloride was observed, which is consistent with previous observations of *in vitro* reconstituted aquaporin 4. In this study, a fast and convenient protocol was established for functional expression of aquaporins, which could serve as basis for further applications such as water filtration.

4.4.2.1 Development of cell-free expression protocols for the preparative scale production of mAQP4 M23.

Previously described CECF reaction conditions were used as criteria for the expression protocol development [145]. The first 22 codons are not essential for the mAQP4 water transport activity and thus were deleted in our construct [167, 168]. The mAQP4 M23 coding sequence was cloned into the vector pIVEX2.3MCS and the template was designed for the production of a modified mAQP4 M23 containing an additional 12 amino acid N-terminal T7-tag as well as a poly(His)$_{10}$ purification-tag at the C-terminal end. The calculated molecular mass of this mAQP4 M23 derivative is 30 kDa. Critical parameters for CF protein production are optimal ion concentrations in particular of Mg^{2+} and K^{+}. After screening of Mg^{2+} and K^{+} ion concentrations in the P-CF expression mode in a range between 7 – 25 mM and 200 – 400 mM, respectively, optima were determined at 17 - 22 mM Mg^{2+} and 250 - 340 mM K^{+}. With these conditions, the yield of CF produced mAQP4 M23 in 50 µl analytical scale micro-reactor reactions as well as in 3 ml preparative scale maxi-reactor reactions were routinely in the range of 1.5 - 2 mg protein per one ml RM.

4.4.2.2 Production of mAQP4 M23 proteomicelles in the P-CF and D-CF expression modes.

In the P-CF mode, the mAQP4 M23 precipitated directly after translation within the RM. The precipitates were harvested by centrifugation, washed once with resolubilization buffer (20 mM Tris, pH 7.3, 150 mM NaCl) and instantly resuspended in a variety of detergents for resolubilization, with the same volume as the initial RM volume. For this P-CF resolubilization screen, the detergents Fos-12 (1%), DHPC (2%), Fos-16 (2%), LMPG (2%) and LPPG (1%) were evaluated. The initial suspensions were incubated at 37°C for 1 hour with gentle shaking to allow efficient mAQP4 M23 solubilization. Residual mAQP4 M23 precipitate was then removed by centrifugation at 18,800 x g for 10 min. The supernatant and residual pellets were subsequently analyzed by SDS-PAGE, western blotting and immunodetection of the C-terminal poly-$(His)_{10}$-tag. From tested detergents, 1% Fos-12, 2% Fos-16 and 1% LPPG showed optimal solubilization of the target protein. Nearly all precipitates were redissolved (Fig. 38), whereas 2% LMPG and 2% DHPC could only partially solubilize the mAQP4 M23 precipitates and approx. 50% of the protein still remained non-soluble (Fig. 38).

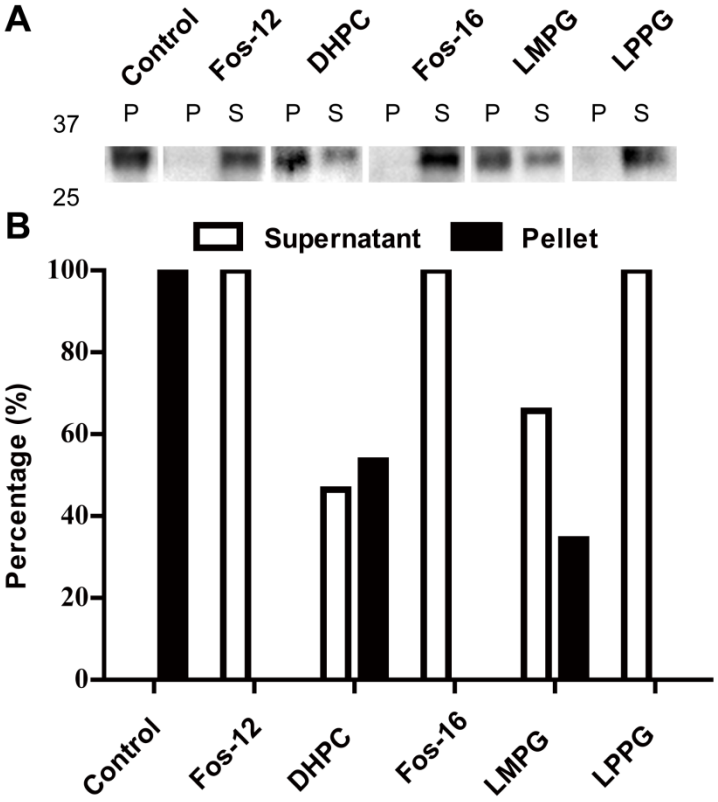


Figure 38. Resolubilization screening of P-CF produced mAQP4 M23.

The pellet from the P-CF reaction mix was resuspended with 1% (w/v) Fos-12, 2% (w/v) DHPC, 2% (w/v) Fos-16, 2% (w/v) LMPG, or 1% (w/v) LPPG. Sample volumes of 4 µl were analyzed by 16% SDS-PAGE. The solubilization efficiency was determined by densitometry after immunoblotting using anti-His antibodies. Control is P-CF expressed mAQP4 M23.

A: immunoblotting using anti-His antibodies. S, supernatant; P, pellet

B: The solubilization efficiency determined by densitometry of the immunoblotting

Alternatively, soluble mAQP4 M23 was directly produced in the D-CF expression mode. The detergents supplied into the RM provided hydrophobic environments for the co-translational solubilization of CF expressed mAQP4 M23. A number of D-CF suitable detergents including Brij-35 (polyoxyethylene-(23)-lauryl-ether) (0.1%, w/v), Digitonin (0.4%, w/v), Triton X-100 (0.1%, w/v) and Tyloxapol (0.05%, w/v) were screened for their efficiency (Fig. 39). In the ideal case, all produced mAQP4 M23

should become soluble while the expression efficiency should not be reduced by the detergent. The detergents Brij-35 and Digitonin showed high efficiency to solubilize mAQP4 M23 without significant effect on the protein expression. With 0.2% Brij-35 in the RM, almost 90% of the expressed mAQP4 M23 was solubilized in proteomicelles. In the presence of 0.4% Digitonin, approximately 75% of synthesized mAQP4 M23 was solubilized. However, Triton X-100 (0.1%) and Tyloxapol (0.05%) were much less effective on the co-translational solubilization of mAQP4 M23, with efficiencies of not more than 10%.

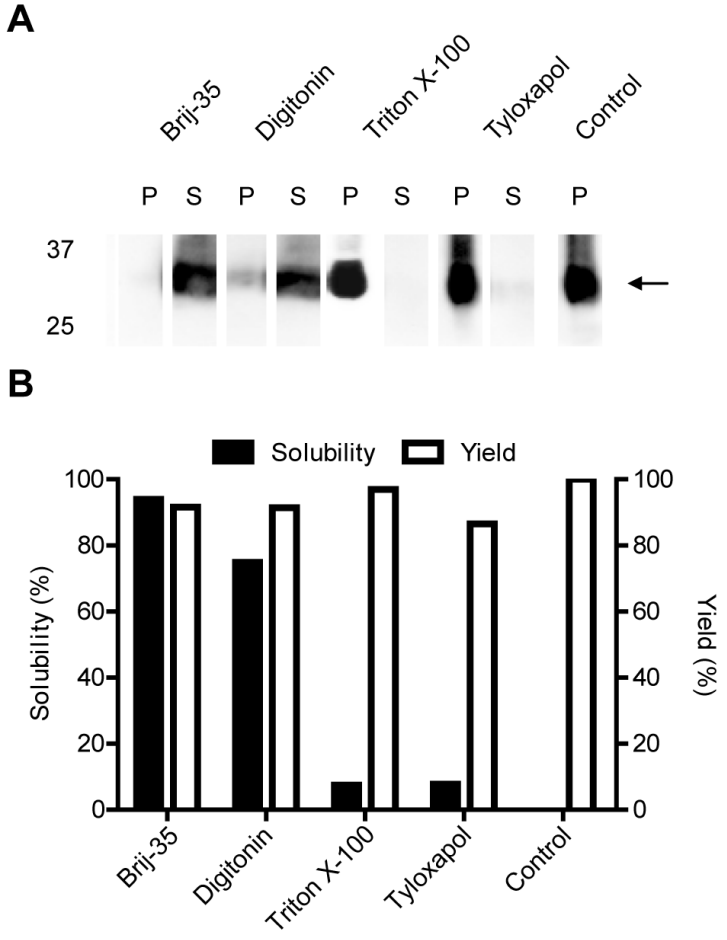


Figure 39. Detergent screening of mAQP4 M23 expressed in the D-CF mode.

A: RM samples of 2 µl were analyzed by 16% SDS-PAGE and immunoblotted using anti-His antibodies.

B: Solubility of D-CF expressed mAQP4 M23 in presence of 0.2% Brij-35, 0.4% Digitonin, 0.1% Triton X-100, and 0.05% Tyloxapol. Control is P-CF expressed mAQP4 M23. S, supernatant; P, pellet.

4.4.2.3 Purification of mAQP4 M23 and *in vitro* liposome reconstitution

Depending on the C-terminal poly-$(His)_{10}$-tag, a one-step IMAC purification was applied to get relative pure proteins from both P-CF and D-CF mode samples (Fig. 40). Binding of mAQP4 M23 to the IMAC column was relatively strong in the analyzed detergents Fos-12 (1%), Fos-16 (2%), LPPG (1%) and Brij-35 (0.2%) and no elution was detected below 100 mM imidazol. Most impurities were washed off with 80 mM imidazol and the bound protein was eluted with 300 mM imidazol.

The SDS-PAGE indicates a prominent 30 kDa signal as the mAQP4 M23 monomer. An additional protein band at 66 kDa detected by immunoblotting could correspond to mAQP4 M23

dimeric complexes (Fig. 40). After immobilization of mAQP4 M23 to the Co^{2+} loaded IMAC column, detergents used for the initial solubilization of mAQP4 M23 could be exchanged by secondary detergents. Only a limited number of detergents are suitable for D-CF expression or for resolubilization after P-CF expression. While the D-CF mode requires very mild detergents, for the resolubilization of P-CF produced precipitates only relatively harsh detergents are useful. Those detergents used for the initial MP solubilization might not be optimal for subsequent assays and exchange against a second probably better suitable detergent could be beneficial. To test the possibility of primary detergent substitution of mAQP4 M23, Brij-35 in the proteomicelles of D-CF expressed mAQP4 M23 was exchanged to DDM (N-dodecyl-β-D-maltoside) (0.05%) or Fos-12 (0.05%), respectively. Fos-12 in the proteomicelles of P-CF expressed mAQP4 M23 was furthermore exchanged to DDM (0.05%). The detergent exchange from Brij-35 to Fos-12 resulted in apparent aggregation and precipitation of mAQP4 M23, while the protein remained soluble after the substitution with DDM from either Brij-35 or Fos-12. After elution of the resulting proteomicelles from the IMAC column, the concentrations of mAQP4 M23 in the peak fractions were determined in the range of 0.7 - 1 mg/ml.

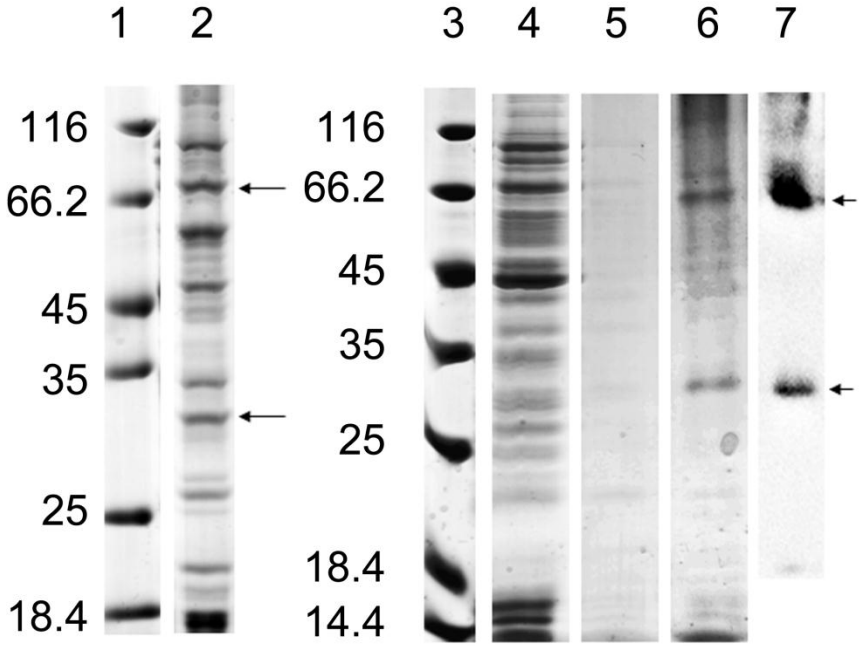


Figure 40. Purification of D-CF produced mAQP4 M23 in 0.2% Brij-35 by Co^{2+}-NTA chromatography. Samples were separated by 12 % (lanes 1-2) or 16 % (lanes 3-7) SDS-PAGE and analyzed by Coomassie staining. Lanes 1 and 3, protein marker; Lane 2, precipitate after P-CF expression; Lane 4, flow through; Lane 5, washing fraction; Lane 6, elution fraction; Lane 7, immunoblot of lane 6 using anti-His antibodies. Samples of 2 µl were applied to each lane.

Reconstitution of mAQP4 M23 into lipid bilayers is a prerequisite for the functional characterization of its water channel activity. The purified mAQP4 M23 was reconstituted into liposomes by following a previously published protocol [129]. The final concentration of mAQP4 M23 in reconstitution mixtures was approx. 100 µg/ml. Initial attempts to destabilize preformed liposomes (4 mg/ml) with DDM (0.04%) were not successful, which led to very low reconstitution rates or in even empty liposomes. When lipid concentration was increased to 6 mg/ml and DDM was replaced by Triton X-100 (0.36%), the mAQP4 M23 reconstitution rate was significantly

increased. The prepared proteoliposomes could be stored at 4°C for up to one week. Extended storage as well as freezing in liquid nitrogen resulted in completely inactive proteoliposomes.

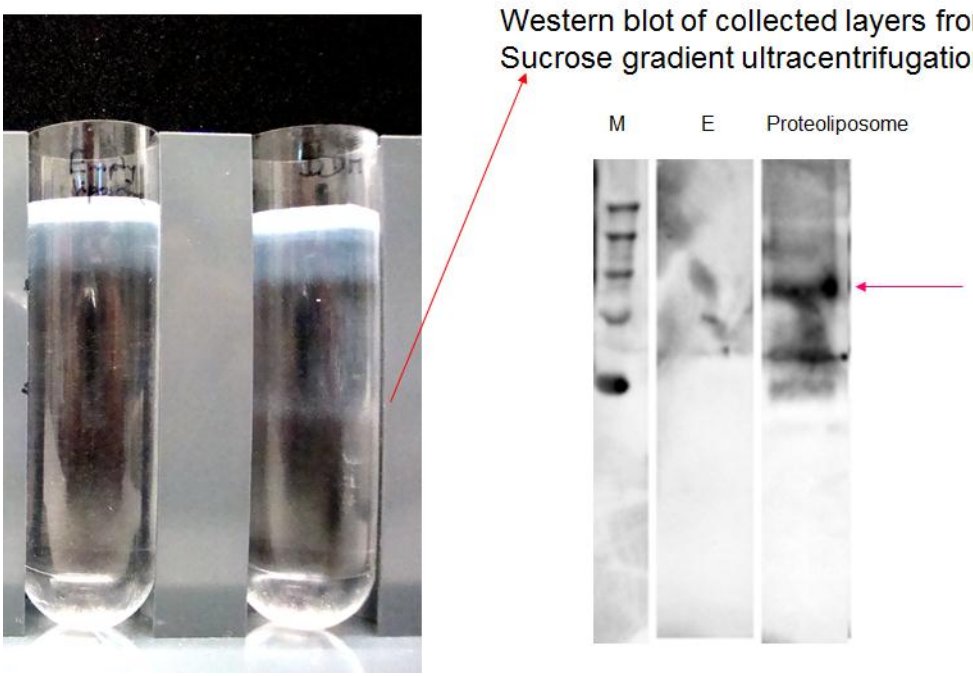


Figure 41. Ultracentrifugation of reconstituted AQP4 M23 liposomes in a sucrose gradient buffer. Sucrose gradient was made by pooling 40%, 30%, 20% and 10% from bottom to top. The preteoliposome was loaded on to the top of the sucrose buffer. Centrifuge at 101,000 ×g at 4°C overnight.

4.4.2.4 Water channel activity of cell-free expressed mAQP4 M23

Water channel activity of reconstituted mAQP4 M23 samples obtained from either P-CF or D-CF expression mode was analyzed in a 100 µl reaction mixture, which was composed of 50 µl proteoliposomes and 50 µl reconstitution buffers with a final sucrose concentration of 200 mM. Reconstituted mAQP4 M23 proteoliposome and control empty liposomes were quickly mixed with the high osmotic reconstitution buffer by stopped-flow equipment at 10°C. The change of liposome volume as the result of water channel activity of inserted mAQP4 M23 was measured by light scattering at λ=436 nm (Fig. 42). All mAQP4 M23 samples obtained from two different expression modes and primarily solubilized in four different detergents were analyzed at 10°C. In all cases, the proteoliposomes had presented a higher water channel activity comparing with the control liposomes, which clearly suggested the functional reconstitution of mAQP4 M23 (Fig. 42). The P_f values of P-CF produced mAQP4 M23 samples resolubilized in Fos-12, Fos-16 and LPPG were determined as 54.4 ± 2.5 µm/s, 49.4 ± 2.0 µm/s and 50.7 ± 2.7 µm/s, respectively. The P_f value obtained from D-CF produced mAQP4 M23 in Brij-35 was 133.1 ± 5.6 µm/s. All control liposomes had an average P_f value of 25.5 ± 1.6 µm/s.

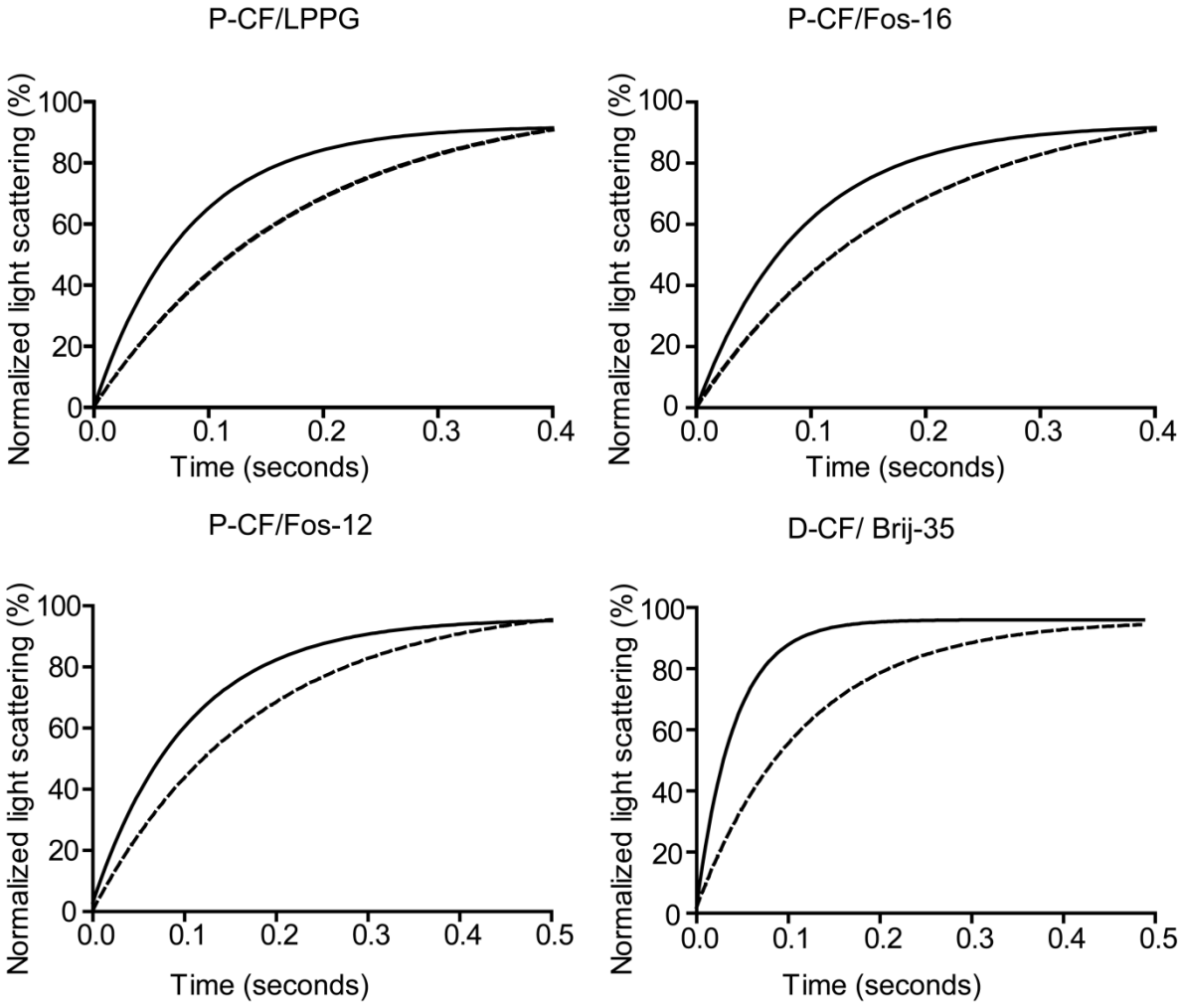


Figure 42. Water transport activity of P-CF and D-CF mode produced mAQP4 M23.

Precipitates of P-CF produced mAQP4 M23 were resolubilized in the indicated detergents. The solubilized proteins were purified by Co^{2+}-NTA chromatography, the initial detergent exchanged to 0.05 % DDM and the samples were reconstituted into E. coli polar lipids. Water transport activity was determined by stopped-flow light scattering measurements of mAQP4 M23 proteoliposomes at 10°C. A 200 mM osmotic gradient was established by rapidly mixing vesicles suspended in reconstitution buffer with an equal volume of reconstitution buffer + 400 mM sucrose. Data represent the average of three independent measurements. Fitted curves of mAQP4 M23 proteoliposome light scattering are shown. Solid line, mAQP4 M23 proteoliposomes; Dashed line, *E. coli* polar lipid empty liposomes.

The water channel activity of aquaporins can be inhibited by the binding of $HgCl_2$ to essential cysteine residues in the protein [169]. Pre-incubation of proteoliposomes containing D-CF produced mAQP4 M23 with 300 µM $HgCl_2$ at 25°C for 5 min resulted in a clear reduction of the water channel activity from 158.3 ± 3.0 µm/s to 103.8 ± 2.7 µm/s (Fig. 43). The inhibition was partially reversible by treatment with ß-mercaptoethanol, which was presumably due to the regeneration of the essential thiol-residues, and the related P_f value was recovered to 126.0 ± 2.8 µm/s. The P_f value of empty liposomes was 55.9 ± 3.7 µm/s, with or without the treatment of $HgCl_2$. The established complete process for the preparative scale CF production of mAQP4 M23 from expression to functionally active protein therefore takes no more than 2 days (Fig. 44).

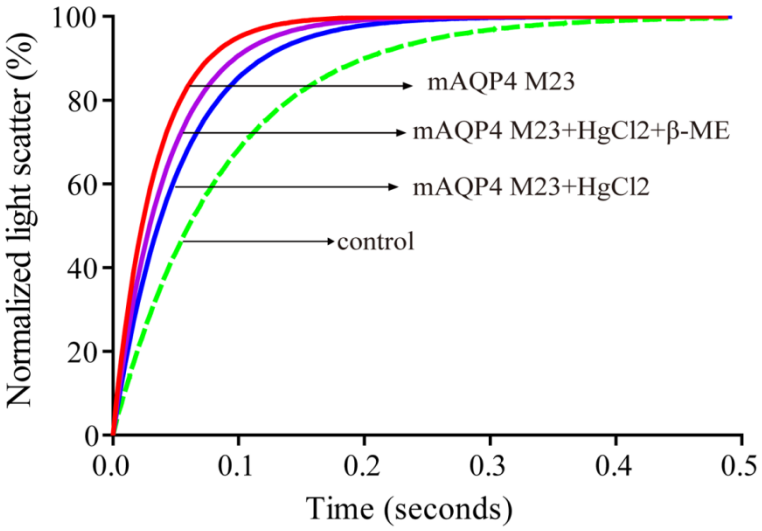


Figure 43. Specific inhibition of mAQP4 M23 water transport.

Proteoliposomes containing D-CF produced mAQP4 M23 were treated with 300 µM HgCl2 for 5 min. at 23°C. For the recovery of the function of mAQP4 M23 function, 2 mM β-mercaptoethanol (β-ME) was added and incubated 10 min. at 23°C after incubation with HgCl2. Empty liposomes with and without treatment by $HgCl_2$ were used as control and showed identical curves (dashed line).

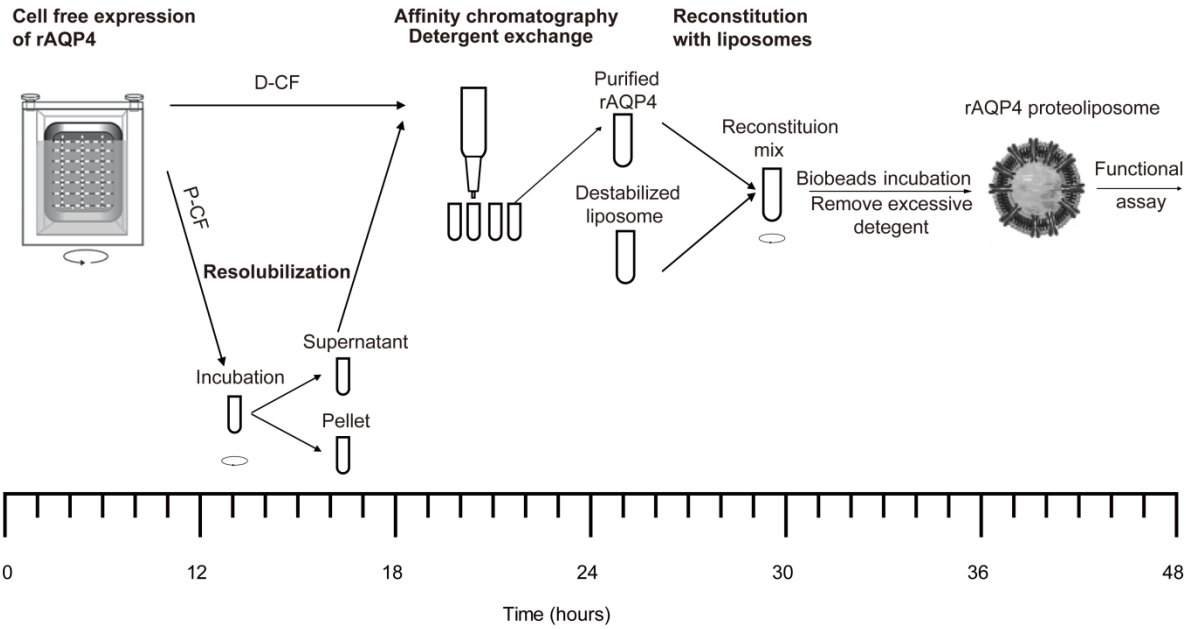


Figure 44. Flow-chart of mAQP4 M23 production by CF expression.

The complete process from expression to functional analysis is finished within 2 days.

5. Discussion

5.1 Establishing high-throughput cell-free expression platform

5.1.1 Setting up an automated cell-free expression platform for screening parameters

In the traditional throughput protein expression approaches, very limited expression parameters can be screened mainly focused on changing the vector or host environments. The sensitive physiology of living cells and membrane barriers do not allow extensive variations of the expression protocols. Therefore, expression conditions of quite a large number of target proteins are quite identical. The low successful rate does not improved further even with the normal optimization of expression conditions. However, those targets, which seemed hard to express with traditional approach are quite often the targets with great interest, i.e., toxins, MPs, aggregation-prone proteins, and unstable sensitive proteins like enzymes.

With the development of CF expression technique, the strategy and expression parameters can be completely different from traditional *in vivo* approach and more versatile. Without the membrane barrier and other metabolic process of living cells, CF expression system provided efficient machinery for protein expression. In addition, the open nature of CF expression system allowed a large number of different expression parameters to be introduced in the system directly to address specific expression problems. For instance, the hydrophobic characterization of MPs always made them difficult to be expressed and analyzed. In addition, some MPs directly lead to the death of the host cells when overexpressed. In contrast, the CF expression system can introduce large number of natural or artificial compounds to the system directly to address the expression and folding problem for those difficult targets. For instance, introducing detergents and lipids will MPs soluble and functional folded when expressed (see the part 1.3.2.4).

With the increase of expression parameters that can be optimized and screened, a more efficient and fast expression CF platform is therefore required. This platform should not only able to provide a platform for large number of targets expression screening, but more importantly provide the platform for individual expression protocol development. Therefore, a more complicated compounds concentrations screening is desired. Since a larger number of screening experiments are required for this screening purpose, a robotic device which can provide a precise pipetting work is therefore necessary. In addition, the selection of suitable CF expression protocols is also critical for establishing such an automatic throughput system.

Quite a lot of CF expression protocols are published and modified for various application, among them the batch CF expression system with the easy setting up in one compartment and around 1 mg protein per 1 ml reaction volume [78, 133] is the most suitable CF system for throughput platform. Besides the easy set ups of the batch system, the possibility to minimize the

reaction volume is also a beneficial parameter for successful throughput application, considering the cost of screening. Low volume like a few microliters can be directly transferred to micro-plate where the reaction can be hold and incubated for protein expression.

With the combination of batch CF expression system and the robotic pipetting system TECAN Freedom EVO 200/8, a prototype of the throughput CF expression system with the purpose of condition screening for specific target can be successfully set up. In order to perform fast and precise pipetting, liquid handling arms with 8 channels was introduced, which can precisely pipetting the volume between 300 nl to 750 µl at once. Besides, in order to handle different liquid, three different liquid classes are therefore defined. As indicated in **4.1.2,** a custom designed software E.Y.E.S program which can create worklists for the robot operation system Germini to carry out was introduced. With this software, the design of the screening experiment is quite straightforward. The whole screening experiment then can be carried out in the following steps: creating a common master mix and aliquot in to each well; pipetting all the screening compounds and additives; supplement with H_2O; adding start reagent to start the expression; incubating and quantification.

Based on the "Cytomim" system, which was established for soluble protein expression and reported to have a relative high yield, we systematically screening the optima concentrations for most of the important compounds like those energy precursors, NTP mix, and T7 polymerase. After optimization, a stable soluble expression yield of sGFP around 500 µg was obtained with the minimum reaction volume of 25 µl in a V-shaped 96-well plate.

5.1.2 Developing individual expression protocols

There are several different CF expression protocols available. They can be classified with the different source of extract: like the *E. coli* [170, 171], wheat germ [172], yeast [45], insect cell [43, 173], and rabbit reticulocyte [174]. Even with the most well developed *E. coli* systems, they differed in energy source and other precursors like NTPs. The concentration of each compound needed to be optimized. Although recommended concentrations are given in those protocols, in practice, those concentrations are highly depend on individual setting ups. That is why in the stage of CF expression protocol development for specific target required rescreening several basic compounds. Small changes in the concentration of specific compounds are often not avoidable, therefore this always cause the decrease of the efficiency of the whole system. As indicated in Fig.14, the difference of optima Mg^{2+} concentration is around 4 mM, which can lead to more than 20% decrease in the expression yield if use the optimum Mg^{2+} concentration from another. Therefore it is recommended to re-optimize several key compounds like Mg^{2+}, T7 polymerase, S30 extract and so on to obtain the basic expression protocol.

A systematic screening of several basic compounds was shown in Fig. 11. Different compounds had different effluence on the expression yield. Compounds like S30 extract and Mg^{2+} showed quite sharp curve with fixed optima concentration. While NH_4^+ showed less influence on

the expression with quite wide range of tolerance. DTT act as a reducing reagent showed a compatible concentration around 2 mM. Further increase will cause the decrease in the expression yield. Linear screening of the compounds concentration would give the information on the independent effect of the compounds on the efficiency of CF expression system. However, the influence of the interaction between different compounds on the efficiency was not analyzed and most likely ignored.

DNA template played a crucial role in the CF expression. Different studies were done in order to improve the design of the template. For instance, with different N-terminal sequence, the stability of the transcribed mRNA [175] can be improved therefore a stable high concentration of mRNA can be achieved. High level of mRNA will greatly improve the expression yield. In the procedure of over production of recombinant protein, the translation initiation [176] play a crucial role. With the tag variation strategy, which adding a short non-functional tag sequence on the 5' coding region of the protein, we successfully optimized the expression of large eukaryotic MPs to a high expression level (above 1 mg per 1 ml reaction mixture) [177].

In this study, we screened the template concentration in our batch system to find the optimum. As shown in Fig. 12, the expression yield reached a plateau with the template concentration around 2-4 ng/µl. Linear template was very convenient for perform different target screening and especially for throughput applications, an overlap PCR strategy was developed to create suitable linear template for CF expression [70]. In order to check the efficiency of the linear template, the comparison experiment with the traditional plasmid template was performed. As shown in Fig.13 with equal weight of linear template and plasmid template, the expression yields were the same. However, with equal molar concentration ratio the expression yield was only 30% compared to plasmid template. This decrease was probably due to the instability of the linear template.

5.2 Introducing chemical stabilizers into the cell-free expression system

5.2.1 Evaluation of chemical stabilizers in cell-free system via sGFP

The development of CF technology allowed high-throughput application in proteomics. Especially, for large number of targets or mutations, protein samples can be obtained with hours. However, to process large number of reactions was still quite challenging and time-consuming if handled manually. Several previous study already showed the possibility to automate the pipetting process of CF reactions [103, 131]. We introduced the Tecan Evo 200® robotic pipetting system, together with custom designed software E.Y.E.S, to perform the pipetting of CF reactions. With this platform, large number of CF free reactions can be performed with a short time periods. In addition, with the usage of throughput monitor methods like fluorescent measurement, the evaluation of the

screening also can be performed in the throughput form. With this a throughput strategy can be achieved for both large proteomic studies.

Quite a lot of researches were done for the throughput application of CF system. However, most of them were based on improving the successful expression of targets [103] or specific application of one type of proteins [102, 178]. A few researches were done to combine CF system and specific protein functional assay to obtain a high-throughput platform, which can be used for protein evolution screening [179] or drug screening[180]. However, our aim was to address those proteins which are difficult to be expressed or to be functional expressed via the CF expression systems. Use the open nature of the system itself to develop a strategy to handle those difficult targets. It was often found that quite large number of difficulty proteins, i.e., aggregation prone, toxic, and MPs could not be addressed by traditional *in vivo* expression system and also not easily handled *in vitro* CF system. Therefore we took the advantage of CF system as an open system and amendable for throughput automation to develop an application of systematic screening of chemical compounds co-translational, which in previous reports beneficial for protein production, stability and folding.

Taking the advantage of fast evaluation of fluorescence, sGFP was taken as the monitor of the CF expression system. Molecular crowding reagents (PEG, Ficoll) were used to mimic the environment inside the cells and the effects on CF expression system were studied recently. Results showed that these macromolecular crowding reagents showed a beneficial effect on the transcription level, while an inhibitory effect on the translation level. In this study, an increase of 2.2 fold proteins yield can be achieved with applying macromolecular reagent in the transcription process rather not in the translation process [181]. Considering the overall effect of transcription and translation, we observed an overall beneficial effect of PEG on the expression of sGFP, up to 20% increase. In addition, via correlated screening of PEG and Mg^{2+}, we observed an interaction of these two compounds. As a result, the effect of Mg^{2+} can be supplemented by certain amount of PEG, without losing the final protein yield.

Furthermore, chemical compounds which acted like protein stabilizer were also evaluated. From the evaluation we found most of the tested protein stabilizer had no or less inhibitory effect on the CF system itself. In order to give a general guide for choosing proper stabilizers, we classified all the compounds into three levels. For those compounds like ethanol, choline and the amino acids gave an increase in the protein yield can behave like expression enhancer and as well as protein stabilizer. While for the tolerant compounds like methanol, isopropanol, betaine, ectoine, sugars, and polyols which did not result in the increase of the protein yield, still could be used within certain concentration (indicated in Tab. 7) to stabilize specific targets during expression to obtain more functional form of proteins. Compounds like butanol, petanol, haxanol, PEI should be avoided because they almost completely killed the system with very low concentration. Important

information for this kind of screening was that all these compounds were available to protein immediately when they were produced. They act as co-translational, not post-translational partners. With this one can expected better protection and more functional protein recovery rate.

As shown in the screening of basic compounds, there might be an interaction between two compounds which influenced the final expression of the CF system. In the screening of PEGs and alcohols, different Mg^{2+} concentrations were used as shown in Fig. 20 and Fig. 24. For the screening of alcohols, the effect was basically the same with different Mg^{2+}. Ethanol was the only alcohol, showing an increasing effect on the expression of sGFP. The increase of yield reached nearly 3 fold, compared to the control, with low Mg^{2+} concentration. While, with high Mg^{2+} concentration, the expression yield increased 60%. The same situation was found in the case of PEG. With low Mg^{2+} concentration, the PEG6000 will really increase dramatically the expression yields, up to 4000%. With high Mg^{2+}, most of the PEG had a 20% increase. In the case of PEG, we assume that the effect of reducing Mg^{2+} on the expression yield can be implemented by adding certain amount of PEG.

5.2.2 Evaluation of chemical stabilizers on improving soluble expression of proteins in cell-free system

We select two targets: Halogenase [182] and GNA1, which tend to form insoluble form during expression *in vivo* or *in vitro*. Halogenase, when expressed *in vivo* required low temperature conditions [183], and still formed quite a lot of precipitated form (data not shown). GNA1 also tent to form precipitate when expressed in CF system. For further proof of the principle to using these stabilizers to generate a beneficial effect on the protein stability and folding, we pick up several stabilizers to test their effect on the solubility and function of the two targets. Among the selected stabilizers, L-arginine and choline showed beneficial effect on the soluble expression of Halogenase. The soluble expression improved up to 8% or 25%, when using 10 mM L-arginine or 10 mM choline as additives.

Results of GNA1 also showed that with both choline and L-arginine, the soluble expression was improved. However, the fluorescence of fusion protein might be influenced by sGFP itself. Comparison of sGFP indicated that both two compounds had beneficial effect on sGFP folding, since the ratio of supernatant and pellet of sGFP increased. Result of enzymatic assay in presence of choline was consistent with the fluorescence data. But, for L-arginine the enzymatic activity was slightly reduced while the fluorescence was improved with addition of 10 mM L-arginine. This effect probably due to the folding of GFP was improved by Arginine [184], but not the GNA1. In this case, the fluorescence did not correlate with the function of the target protein. However, we select two compounds, PEG8000 and choline, to present an example of correlated screening of two compounds which act either as expression enhancer or protein stabilizer. As indicated in Fig. 29, when combine the two compounds, more soluble expression was observed than the sum up of using

each single compound. When using PEG8000 alone, an increase of soluble expression was observed around 17% was observed. While with choline alone around 23% more soluble expression was achieved. While when using both PEG8000 and choline at the optima concentrations, an increase of around 60% was obtained. This 60 % increase was much more than simply sum up of the 17% and 23%. Which might indicate an additional beneficial effect came from the interaction of the two compounds.

5.3 Application of throughput cell-free expression system on membrane proteins

5.3.1 Fast evaluation of membrane protein expression

The usage of green fluorescence protein (GFP) as a folding indicator *in vivo* provided an efficient tool for MPs expression. The folding of GFP was used as a monitor if the overexpressed MPs fusion partner was end up with inclusion bodies or inserted to membranes. Successful method was developed with both prokaryotic *E. coli* expression system [144] and eukaryotic yeast system [185]. Taken the same idea for the CF expression system, we introduced the construct of MPs fused with GFP at the C-terminal. However, the successful usage of the GFP as a monitor in the CF expression system was largely depended on the folding of GFP in presence of detergent, since the hydrophobicity of MPs. However, with previous results, the fluorescence was dramatically decrease in most of the detergents, like TritonX-100, maltoside, glucoside, Digitonin and so on [102]. Finally, the best folding of GFP was observed in polyoxyethylenes or phosphocholines derivatives, which still retained around 30-55% of the fluorescence (shown in Fig. 30 and [102]). With this reduced signal it might be sufficient for the first analysis of soluble MPs expression in D-CF mode. However, detergent screening of MPs fused with GFP only showed high fluorescence in Brij-78 and Brij-58. Therefore, most of the D-CF mode expression of MPs fused with GFP was performed with the Brij-78 and Brij-58 [102].

However, with more target tested, the effect of C-terminal GFP folding really limit the application of this method, due to the fact that there were many MPs which cannot be expressed soluble in Brij-78. In addition, there was still quite some part of the soluble expressed MPs which did not give the fluorescence signals. In order to address this problem for the application of high throughput CF platform in MPs, we select a new mutant of GFP which should an improved stability and faster folding compared to the normal GFP [147]. Due to the fast folding property of the superfolder GFP, the tolerance to different detergents might be much better. Experiment which compared the folding of normal shifted GFP (sGFP) and superfolder GFP was performed. As indicated in Fig. 30, the sGFP did not show fluorescence in quite some detergent, like DDM, TritoX100, Digitonin, fluorinated surfactance and Nvoy. However, the superfolder GFP exhibit a broader detergent resistance, in Brij-78 the superfolder GFP fold up to around 100%. With detergent

like TritonX-100 and Brij-35, Digitonin, DDM the folding of the superfolder GFP were still above 70%. For the case of CHAPS both sGFP and superfolder GFP did not give high fluorescence signals, this might due to the inhibitory effect of the detergent on the CF expression system itself.

With the promising results from the detergent tolerance experiment of this superfolder GFP (Fig. 30), further analysis of MP expression using the superfolderGFP was performed. This example was given for the detergent screening of soluble expression for SugE (a small multidrug transporter) with two different GFP fusions. As shown in Fig. 31, compared to the most well folded Brij-78 for sGFP fusion construct, the superfolder GFP fusion construct also showed good fluorescence (above 80% compared to the control in Brij-78) in detergents like Digitonin and Lauryl-MNG. Overall, with all the detergents selected, the superfolder GFP fusion construct showed much higher fluorescence than the normal sGFP, especially with Brij-58 showed higher fluorescence than the positive control with Brij-78.

5.3.2 Improving the membrane proteins translocation with lipids/detergent mixtures

Since MPs are targeting at the bio-membrane, therefore in several cases, lipids are required for the functionally or folding of certain MPs. Although detergent micelles can also provide hydrophobic environment, lipids sometime are necessary. Taking this effect into consideration, experiment which try to introduce additional lipids together with the detergent into the CF expression system. As summarized in Tab. 8, different MPs including large target like bacteriorhodopsin and ATP synthase were functional produced in lipid/detergent environment, with increased functionality than using detergent alone (see 4.3.3).

5.4 Characterization of cell-free produced membrane proteins

5.4.1 Strategy for preparing homogeneous membrane protein sample

One important parameter for sample quality is its homogeneity. Different methods like SEC, EM and other spectroscopy methods to characterize the sample status. The select SugE (small multidrug transporter) protein was chosen for further sample quality optimization. The SugE protein was highly expressed in P-CF mode as precipitates and later on resolubized in detergents to form soluble sample. P-CF produce SugE with its native N-terminal and C-terminal amino acids sequence was evaluated via SEC. As shown in Fig. 33, re-solubilized sample showed different elution profile on the SEC column. For example, the β-OG resolubilized SugE showed a very broad peak over the whole elution time, indicating a multimeric oligomer status or even aggregations. While the DDM sample showed a single sharp peak, which indicated a more homogeneous sample than in TritonX-100 and β-OG.

Since the construct was without any affinity tags for purification, special strategy was developed for a purification purpose. With a mild detergent like $NDSB_{256}$ was first used for the

resolubilization. However, NDSB256 only solubilize the impurities and very few of the SugE target. A second step of resolubilization with other detergent like DDM was used to mainly solubilize the SugE protein. With this two-step resolubilization, a relative pure solubilized protein can be obtained (shown in Fig.34). However, after the resolubilization, the sample was still now pure and quite some other impurities remaining. SEC was then used to further purifying the protein sample. The peak fractions that were collected and pooled showed a relative pure band on a SDS-PAGE (see Fig. 35). With this particular purification protocol, this MP SugE was purified to homogeneous and further structural studies can then be performed with this sample.

The hydrophobic environment for MPs structure characterization is crucial. The interactions of MPs and the hydrophobic environment are always needed to be considered when trying to determine the structure of a MP. Especially, for MPs crystallization, the role of detergent for getting well diffracted MP crystals is important. Therefore, a lot of studies have been done to characterize the effect of different detergent on the crystallization process [186]. Several detergents have been used frequently for MPs crystallization like DDM, β-OG, LDAO, and so on[186]. Some derivatives of these detergents with modification of the alkyl chain were created in order to provide different sized micelles. The experiment of detergent exchange on a SEC column provided more information for potential detergents that can be used for crystallization conditions screening (Fig. 36). We observed in this experiment, with small micelle, the peak shape of the SugE on the SEC was less homogeneous. This can be explained by that small micelles had problem to hold the protein stable. However, small micelles can provide a better protein-protein contact which might facilitate the crystallization. Therefore, exploring more detergents which can keep the MPs stable with small micelle was necessary to improve the success rate of crystallization. As shown in Fig. 36, the SugE sample behaved the best in DDM. However, the NM and DM sample showed a very close double peak, this might due to the change of detergent micelle from DDM to NM or DM. For the sample in FG, NG and OG, the peak looked broad with a small shoulder, which might indicate inhomogeneity. Finally, the C8E5 and C10E5 also showed relative good peaks. With this experiment, more detergents which showed comparable protein quality can be used for the further crystallization. In the early elution volume, there is a huge peak appeared in nearly all the elution profile. By checking the absorbance of 260 nm, we find this peak might be some nuclear acids. And after peak collection of the protein sample, this peak was disappeared.

Although the first crystallization screening only showed small crystals of detergent, the results of detergents screening and purification procedure provide an example of quality optimization of CF produced MPs.

5.4.2 Functional characterization of cell-free produced mAPQ4 M23

Aquaporins are a ubiquitous class of MPs present in prokaryotes and eukaryotes that provide the cellular gatekeepers for water as well as for other small molecules such as glycerol. In

human, 13 different and tissue specific AQPs are responsible for transport mechanisms and showed considerable clinical relevance [4]. AQP4 is expressed in the brain and thought to be primarily responsible for cerebral water homeostasis [187]. This protein is an important central regulator of cerebrospinal fluid which has to be very tightly controlled in order to prevent intracranial pressure resulting in compression of brain tissue, neurological disorders and even cell death. The two isoforms of AQP4 resulting from two optional translation initiation sites at methionine M1 or M23 are very unique. The shorter M23 isoform favors larger array formation in distinct tissues based on improved intermolecular contacts [188, 189]. Several structural information of AQPs has already been obtained by X-ray crystallization. The abundant AQP0 and AQP1 could be isolated in sufficient amounts from natural tissues whereas AQP4 and AQP5 were heterologously expressed in yeast cells [190-193]. However, obtaining sufficient amounts of functional AQP samples still remains challenging and hampers their detailed molecular study. In addition, developing and screening for therapeutic drugs targeting on AQP4 would be valuable in addressing damages caused by stroke, edema, epilepsy and other CNS disorders.

This report demonstrated the first example for the preparative scale CF production of a functional eukaryotic AQP. Expression of rAQP4 that is approx. 92% identical to the mouse AQP4 has been approached in conventional *in vivo* systems before. But the expression in *E. coli* was not successful, and only a few µg of rAQP4 per liter of culture could be obtained from *Pichia pastoris* [194]. Best expression rates of up to 3 mg/l were obtained in Sf9 cell cultures after 72 hours infection [11, 194]. Summarizing from recent reports associated with AQP production, up to 15 mg/l hAQP4 could be produced in *P. pastoris* [193]; 3–5 mg/l of Methanothermobacter marburgensis AQPM was produced in *E. coli* [195]; 0.5 mg/l of human AQP2 was expressed in insect cells [196]; 9–13 mg/l of *E. coli* AQPZ was produced in *E. coli* [197] and 25 mg/L of spinach AQP-PM28A was produced in *P. pastoris* [198]. It was further suggested that construction of large fusion proteins may further improve the expression of AQPs [199]. However, most of AQPs are still difficult to produce, and solublization of overproduced AQPs from cellular membrane remains a challenging task. With the established CF expression protocol in this report, we were able to generate soluble and functional mAQP4 M23 within 24 hours. In addition to the considerably shortened expression time, the downstream purification process is significantly faster with the application of affinity purification column. The handling volumes by using CF expression technology are much smaller and expression rates of mg per ml can be achieved if compared with mg per litres with *in vivo* expression systems. This strategy will therefore dramatically accelerate the preparation of bioactive AQPs for both functional and structural assays.

The *E. coli* AQPZ was previously synthesized in a similar CF system using the batch configuration [200]. The protein was directly synthesized into preformed artificial liposomes that have been added into the reaction, resulting into proteoliposomes. In this study, we have used two different CF expression modes that both provided solubilized mAQP4 M23 samples, representing

the preferred samples for structural analysis. In the P-CF mode, the protein was first precipitated and then post-translationally solubilized by detergents. In the D-CF mode, the mAQP4 M23 was co-translationally solubilized by detergents supplemented into the CF reaction. The results of water channel activity assays have indicated that both CF expression modes could produce functionally folded mAQP4 M23, while higher Pf values were obtained from D-CF samples. Osmotic water permeability values (Pf) in reconstituted mAQP4 M23 proteoliposomes were calculated according to the results derived from light scattering assays and can be compared with corresponding values in the literature. The measured Pf values of ~133 µm/s for mAQP4 M23 produced by D-CF at 10°C are comparable with reported Pf values of 28 µm/s for AQPZ at 6.5°C [130], Pf values of 57 µm/s for AQPM at 4°C [195], and Pf values of 112 µm/s for rAQP4 expressed in yeast at 10°C [169]. Considering high variations in reconstitution rates and the assay temperature bias, our data obtained from the CF expressed mAQP4 M23 are in good agreement with the above mentioned Pf values. The lower Pf values of the P-CF produced samples at approx. 50 µm/s might also be a result of different reconstitution rates as different detergents were used for mAQP4 M23 solubilization.

Several higher Pf values of approx 200 µm/s of mAQP4 M23 are reported in the literature [201]. However, one should notice that most high Pf values were obtained by *in vivo* functional assays [201]. Pf values calculated by *in vitro* assays after reconstitution into artificial liposomes were usually below or around 100 µm/s, e.g. 89 µm/s [202] or 117 µm/s [169]. In addition, lower Pf values (around 75 µm/s) were further observed in other *in vivo* measurements [203]. While there still potential exists for quality optimization of CF produced AQP in future, the functional parameter Pf clearly appears to be influenced by a number of factors which have to be considered if different samples were compared.

In the present work, two different CF expression approaches were developed that obtained sufficient amounts of functional eukaryotic AQPs within a short time. The presented protocols could be useful for structural characterization as well as for industrial applications of water filtration such as proposed by Swartz 2006 [12]. Lipid bilayers with functional AQP could be immobilized on top of porous supports e.g. regenerated cellulose. With the specificity and high permeability of the selected AQP, only water can pass through the membrane while contaminants like ions and other molecules were not transported. This prototype could therefore provide an alternative way for water filtration. However, the requirement of large amounts of functional AQPs is the key bottleneck of this application.

6. Perspectives

With the established high-throughput CF platform, more applications in addressing challenge targets are possible and more efficient, especially MPs. Preliminary work of using a superfolderGFP as monitor for MPs expression and even folding enable both large number of MPs target expression screening and target specific condition screening. More researches needs to be done in the future to further improve this platform and set up a general standard optimization strategy for MPs expression, purification suitable for functional and structural characterization.

New nanoparticle technology with different set ups like lipid/detergent mix (bicelles or detergent solubilized liposomes) and nanodiscs (lipid/protein) are merging to optimize the hydrophobic environment. However, still quite lot of research need to be done to proof the advantage of these new approaches. In combination with the throughput CF platform, it will be more efficient for applying these new nanoparticles to selected MPs.

Successful examples of functional produce AQP4 M23 in different CF expression mode, both P-CF and D-CF, proved that eukaryotic MPs can be successfully produced with the *E. coli* based CF expression system with very high yield. It is always a good choice to select the most simple and efficient system to start with. However, in the future, a combination of prokaryotic and eukaryotic CF system would be more attracting, which retain the high efficiency of protein production and the post-translational modification for the functional folding.

The advantage of CF expression system for providing protein sample for structural study was already recognized. Especially for NMR study, selective labeling and very less scrambling with different amino acids compared to traditional *in vivo* expression system are the biggest advantages. High expression yield and less time-consuming purification process is also attractive for obtain large amount of homogeneous sample for structural approach.

7. Reference

1. He, M., Cell-free protein synthesis: applications in proteomics and biotechnology. N Biotechnol, 2008. 25(2-3): p. 126-32.

2. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, How many drug targets are there? Nat Rev Drug Discov, 2006. 5(12): p. 993-6.

3. Agre, P., S. Sasaki, and M.J. Chrispeels, Aquaporins: a family of water channel proteins. Am J Physiol, 1993. 265(3 Pt 2): p. F461.

4. Rojek, A., et al., A current view of the mammalian aquaglyceroporins. Annu Rev Physiol, 2008. 70: p. 301-27.

5. Lu, M.Q., et al., The human AQP4 gene: Definition of the locus encoding two water channel polypeptides in brain. Proceedings of the National Academy of Sciences of the United States of America, 1996. 93(20): p. 10908-10912.

6. Papadopoulos, M.C. and A.S. Verkman, Potential utility of aquaporin modulators for therapy of brain disorders. Prog Brain Res, 2008. 170: p. 589-601.

7. Costa, C., et al., Aquaporin 1 and aquaporin 4 overexpression in bovine spongiform encephalopathy in a transgenic murine model and in cattle field cases. Brain Res, 2007. 1175: p. 96-106.

8. Misu, T., et al., Loss of aquaporin 4 in lesions of neuromyelitis optica: distinction from multiple sclerosis. Brain, 2007. 130(Pt 5): p. 1224-34.

9. Perez, E., et al., Aquaporin expression in the cerebral cortex is increased at early stages of Alzheimer disease. Brain Res, 2007. 1128(1): p. 164-74.

10. Badaut, J., et al., Temporal and regional evolution of aquaporin-4 expression and magnetic resonance imaging in a rat pup model of neonatal stroke. Pediatr Res, 2007. 62(3): p. 248-54.

11. Hiroaki, Y., et al., Implications of the aquaporin-4 structure on array formation and cell adhesion. J Mol Biol, 2006. 355(4): p. 628-39.

12. Swartz, J., Developing cell-free biology for industrial applications. J Ind Microbiol Biotechnol, 2006. 33(7): p. 476-85.

13. Xu, Z., J. Lian, and J. Cai, Efficient Expression of Aquaporin Z in Escherichia coli Cell-free System Using Different Fusion Vectors. Protein Pept Lett, 2009.

14. Lian, J., et al., Efficient expression of membrane-bound water channel protein (Aquaporin Z) in Escherichia coli. Protein Pept Lett, 2008. 15(7): p. 687-91.

15. Piddock, L.J., Multidrug-resistance efflux pumps - not just for resistance. Nat Rev Microbiol, 2006. 4(8): p. 629-36.

16. Koronakis, V., J. Eswaran, and C. Hughes, Structure and function of TolC: the bacterial exit duct for proteins and drugs. Annu Rev Biochem, 2004. 73: p. 467-89.

17. Aires, J.R. and H. Nikaido, Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of Escherichia coli. J Bacteriol, 2005. 187(6): p. 1923-9.

18. Eswaran, J., et al., Three's company: component structures bring a closer view of tripartite drug efflux pumps. Curr Opin Struct Biol, 2004. 14(6): p. 741-7.

19. London, J., C. Skrzynia, and M.E. Goldberg, Renaturation of Escherichia coli tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers. Eur J Biochem, 1974. 47(2): p. 409-15.

20. Wetzel, R., Mutations and off-pathway aggregation of proteins. Trends Biotechnol, 1994. 12(5): p. 193-8.

21. Haase-Pettingell, C.A. and J. King, Formation of aggregates from a thermolabile *in vivo* folding intermediate in P22 tailspike maturation. A model for inclusion body formation. J Biol Chem, 1988. 263(10): p. 4977-83.

22. Fink, A.L., Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des, 1998. 3(1): p. R9-23.

23. Ellgaard, L., M. Molinari, and A. Helenius, Setting the standards: quality control in the secretory pathway. Science, 1999. 286(5446): p. 1882-8.

24. Wickner, S., M.R. Maurizi, and S. Gottesman, Posttranslational quality control: folding, refolding, and degrading proteins. Science, 1999. 286(5446): p. 1888-93.

25. Tran, P.B. and R.J. Miller, Aggregates in neurodegenerative disease: crowds and power? Trends Neurosci, 1999. 22(5): p. 194-7.

26. Serio, T.R. and S.L. Lindquist, Protein-only inheritance in yeast: something to get [PSI+]-ched about. Trends Cell Biol, 2000. 10(3): p. 98-105.

27. Baneyx, F. and M. Mujacic, Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol, 2004. 22(11): p. 1399-408.

28. Kopito, R.R., Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol, 2000. 10(12): p. 524-30.

29. Valax, P. and G. Georgiou, Molecular characterization of beta-lactamase inclusion bodies produced in Escherichia coli. 1. Composition. Biotechnol Prog, 1993. 9(5): p. 539-47.

30. Allen, S.P., et al., Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli. J Bacteriol, 1992. 174(21): p. 6938-47.

31. Carrio, M.M., J.L. Corchero, and A. Villaverde, Dynamics of *in vivo* protein aggregation: building inclusion bodies in recombinant bacteria. FEMS Microbiol Lett, 1998. 169(1): p. 9-15.

32. Ellis, R.J. and A.P. Minton, Cell biology: join the crowd. Nature, 2003. 425(6953): p. 27-8.

33. Bowden, G.A., A.M. Paredes, and G. Georgiou, Structure and morphology of protein inclusion bodies in Escherichia coli. Biotechnology (N Y), 1991. 9(8): p. 725-30.

34. Taylor, G., et al., Size and Density of Protein Inclusion-Bodies. Bio-Technology, 1986. 4(6): p. 553-557.

35. Tsumoto, K., et al., Practical considerations in refolding proteins from inclusion bodies. Protein Expr Purif, 2003. 28(1): p. 1-8.

36. Kurland, C. and J. Gallant, Errors of heterologous protein expression. Curr Opin Biotechnol, 1996. 7(5): p. 489-93.

37. Sorensen, H.P., H.U. Sperling-Petersen, and K.K. Mortensen, Production of recombinant thermostable proteins expressed in Escherichia coli: completion of protein synthesis is the bottleneck. J Chromatogr B Analyt Technol Biomed Life Sci, 2003. 786(1-2): p. 207-14.

38. Daley, D.O., et al., Global topology analysis of the Escherichia coli inner membrane proteome. Science, 2005. 308(5726): p. 1321-3.

39. Hunte, C., Specific protein-lipid interactions in membrane proteins. Biochem Soc Trans, 2005. 33: p. 938-942.

40. Opekarova, M. and W. Tanner, Specific lipid requirements of membrane proteins--a putative bottleneck in heterologous expression. Biochim Biophys Acta, 2003. 1610(1): p. 11-22.

41. Xie, J., et al., Phosphatidylethanolamine and monoglucosyldiacylglycerol are interchangeable in supporting topogenesis and function of the polytopic membrane protein lactose permease. J Biol Chem, 2006. 281(28): p. 19172-8.

42. Matthaei, J.H. and M.W. Nirenberg, Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. Proc Natl Acad Sci U S A, 1961. 47: p. 1580-8.

43. Ezure, T., et al., A cell-free protein synthesis system from insect cells. Methods Mol Biol, 2010. 607: p. 31-42.

44. Lockard, R.E. and J.B. Lingrel, Synthesis of Mouse Hemoglobin Beta-Chains in a Rabbit Reticulocyte Cell-Free System Programmed with Mouse Reticulocyte 9s Rna. Biochem Biophys Res Commun, 1969. 37(2): p. 204-&.

45. Wang, X., et al., An optimized yeast cell-free system: sufficient for translation of human papillomavirus 58 L1 mRNA and assembly of virus-like particles. J Biosci Bioeng, 2008. 106(1): p. 8-15.

46. Kovtun, O., et al., Towards the construction of expressed proteomes using a Leishmania tarentolae based cell-free expression system. PLoS One, 2010. 5(12): p. e14388.

47. Witherell, G., *In vitro* translation using HeLa extract. Curr Protoc Cell Biol, 2001. Chapter 11: p. Unit 11 8.

48. Wang, Z. and M.S. Sachs, Arginine-specific regulation mediated by the Neurospora crassa arg-2 upstream open reading frame in a homologous, cell-free *in vitro* translation system. J Biol Chem, 1997. 272(1): p. 255-61.

49. Murota, K., et al., Arabidopsis cell-free extract, ACE, a new *in vitro* translation system derived from Arabidopsis callus cultures. Plant Cell Physiol, 2011. 52(8): p. 1443-53.

50. Kigawa, T., et al., Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. J Struct Funct Genomics, 2004. 5(1-2): p. 63-8.

51. Zubay, G., In-Vitro Synthesis of Protein in Microbial Systems. Annual Review of Genetics, 1973. 7: p. 267-287.

52. Kim, T.W., et al., Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. J Biotechnol, 2006. 126(4): p. 554-61.

53. Liu, D.V., J.F. Zawada, and J.R. Swartz, Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis. Biotechnol Prog, 2005. 21(2): p. 460-5.

54. Jiang, X., et al., Reduction of protein degradation by use of protease-deficient mutants in cell-free protein synthesis system of Escherichia coli. J Biosci Bioeng, 2002. 93(2): p. 151-6.

55. Yamane, T., et al., Enhanced cell-free protein synthesis using a S30 extract from Escherichia coli grown rapidly at 42 degrees C in an amino acid enriched medium. Biotechnol Prog, 2005. 21(2): p. 608-13.

56. Zawada, J. and J. Swartz, Maintaining rapid growth in moderate-density Escherichia coli fermentations. Biotechnol Bioeng, 2005. 89(4): p. 407-15.

57. Kai, L., et al., Systems for the cell-free synthesis of proteins. Methods Mol Biol, 2012. 800: p. 201-25.

58. Schwarz, D., et al., Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. Nat Protoc, 2007. 2(11): p. 2945-57.

59. Sawasaki, T., et al., [High-throughput expression of proteins from cDNAs catalogue from Arabidopsis in wheat germ cell-free protein synthesis system]. Tanpakushitsu Kakusan Koso, 2002. 47(8 Suppl): p. 1003-8.

60. Ohashi, H., et al., Efficient protein selection based on ribosome display system with purified components. Biochem Biophys Res Commun, 2007. 352(1): p. 270-6.

61. Shimizu, Y., et al., Cell-free translation reconstituted with purified components. Nat Biotechnol, 2001. 19(8): p. 751-5.

62. Griswold, K.E., et al., Effects of codon usage versus putative 5'-mRNA structure on the expression of Fusarium solani cutinase in the Escherichia coli cytoplasm. Protein Expr Purif, 2003. 27(1): p. 134-42.

63. Kudla, G., et al., Coding-sequence determinants of gene expression in Escherichia coli. Science, 2009. 324(5924): p. 255-8.

64. Hall, M.N., et al., A role for mRNA secondary structure in the control of translation initiation. Nature, 1982. 295(5850): p. 616-8.

65. Laursen, B.S., et al., Initiation of protein synthesis in bacteria. Microbiol Mol Biol Rev, 2005. 69(1): p. 101-23.

66. de Smit, M.H. and J. van Duin, Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci U S A, 1990. 87(19): p. 7668-72.

67. Yabuki, T., et al., A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis. J Struct Funct Genomics, 2007. 8(4): p. 173-91.

68. Qing, G., B. Xia, and M. Inouye, Enhancement of translation initiation by A/T-rich sequences downstream of the initiation codon in Escherichia coli. J Mol Microbiol Biotechnol, 2003. 6(3-4): p. 133-44.

69. Ahn, J.H., et al., Use of signal sequences as an in situ removable sequence element to stimulate protein synthesis in cell-free extracts. Nucleic Acids Res, 2007. 35(4): p. e21.

70. Haberstock, S., et al., A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. Protein Expr Purif, 2012. 82(2): p. 308-316.

71. Madin, K., et al., A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc Natl Acad Sci U S A, 2000. 97(2): p. 559-64.

72. Sawasaki, T., et al., A cell-free protein synthesis system for high-throughput proteomics. Proc Natl Acad Sci U S A, 2002. 99(23): p. 14652-7.

73. Kigawa, T., et al., Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Lett, 1999. 442(1): p. 15-9.

74. Ryabova, L.A., et al., Acetyl phosphate as an energy source for bacterial cell-free translation systems. Anal Biochem, 1995. 226(1): p. 184-6.

75. Calhoun, K.A. and J.R. Swartz, An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates. Biotechnol Prog, 2005. 21(4): p. 1146-53.

76. Kim, D.M. and J.R. Swartz, Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. Biotechnol Bioeng, 2001. 74(4): p. 309-16.

77. Jewett, M.C. and J.R. Swartz, Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng, 2004. 86(1): p. 19-26.

78. Jewett, M.C. and J.R. Swartz, Rapid expression and purification of 100 nmol quantities of active protein using cell-free protein synthesis. Biotechnol Prog, 2004. 20(1): p. 102-9.

79. Spirin, A.S., et al., A continuous cell-free translation system capable of producing polypeptides in high yield. Science, 1988. 242(4882): p. 1162-4.

80. Sawasaki, T., et al., A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett, 2002. 514(1): p. 102-5.

81. Klammt, C., et al., High level cell-free expression and specific labeling of integral membrane proteins. European Journal of Biochemistry, 2004. 271(3): p. 568-580.

82. Elbaz, Y., et al., *In vitro* synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. Proc Natl Acad Sci U S A, 2004. 101(6): p. 1519-24.

83. Sansuk, K., et al., GPCR proteomics: Mass spectrometric and functional analysis of histamine H-1 receptor after baculovirus-driven and *in vitro* cell free expression. J Proteome Res, 2008. 7(2): p. 621-629.

84. Klammt, C., et al., Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. FEBS J, 2005. 272(23): p. 6024-38.

85. Klammt, C., et al., Functional analysis of cell-free-produced human endothelin B receptor reveals transmembrane segment 1 as an essential area for ET-1 binding and homodimer formation. FEBS J, 2007. 274(13): p. 3257-69.

86. Schwarz, D., et al., Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies. Methods, 2007. 41(4): p. 355-69.

87. Park, K.H., et al., Fluorinated and hemifluorinated surfactants as alternatives to detergents for membrane protein cell-free synthesis. Biochem J, 2007. 403(1): p. 183-7.

88. Neumann-Haefelin, C., et al., SRP-dependent co-translational targeting and SecA-dependent translocation analyzed as individual steps in the export of a bacterial protein. EMBO J, 2000. 19(23): p. 6419-26.

89. Wickner, W., A.J. Driessen, and F.U. Hartl, The enzymology of protein translocation across the Escherichia coli plasma membrane. Annu Rev Biochem, 1991. 60: p. 101-24.

90. Berrier, C., et al., Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent. Biochemistry, 2004. 43(39): p. 12585-91.

91. Marques, B., et al., Liposome-mediated cellular delivery of active gp91(phox). PLoS One, 2007. 2(9): p. e856.

92. Umakoshi, H., et al., Charged liposome affects the translation and folding steps of *in vitro* expression of green fluorescent protein. J Biosci Bioeng, 2009. 108(5): p. 450-4.

93. Kalmbach, R., et al., Functional cell-free synthesis of a seven helix membrane protein: in situ insertion of bacteriorhodopsin into liposomes. J Mol Biol, 2007. 371(3): p. 639-48.

94. Nomura, S.M., et al., Direct preparation of giant proteo-liposomes by *in vitro* membrane protein synthesis. J Biotechnol, 2008. 133(2): p. 190-5.

95. Nozawa, A., et al., A cell-free translation and proteoliposome reconstitution system for functional analysis of plant solute transporters. Plant Cell Physiol, 2007. 48(12): p. 1815-20.

96. Wuu, J.J. and J.R. Swartz, High yield cell-free production of integral membrane proteins without refolding or detergents. Biochim Biophys Acta, 2008. 1778(5): p. 1237-50.

97. Switzer, W.M. and W. Heneine, Rapid screening of open reading frames by protein synthesis with an *in vitro* transcription and translation assay. Biotechniques, 1995. 18(2): p. 244-8.

98. Roberts, B.E. and B.M. Paterson, Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc Natl Acad Sci U S A, 1973. 70(8): p. 2330-4.

99. Roest, P.A., et al., Protein truncation test (PTT) for rapid detection of translation-terminating mutations. Hum Mol Genet, 1993. 2(10): p. 1719-21.

100. Ahn, J.H., et al., Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. Biochem Biophys Res Commun, 2005. 338(3): p. 1346-52.

101. Woodrow, K.A. and J.R. Swartz, A sequential expression system for high-throughput functional genomic analysis. Proteomics, 2007. 7(21): p. 3870-9.

102. Schwarz, D., et al., Cell-free expression profiling of E. coli inner membrane proteins. Proteomics, 2010. 10(9): p. 1762-79.

103. Endo, Y. and T. Sawasaki, High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. J Struct Funct Genomics, 2004. 5(1-2): p. 45-57.

104. Savage, D.F., et al., Cell-free complements *in vivo* expression of the E. coli membrane proteome. Protein Sci, 2007. 16(5): p. 966-76.

105. Hammon, J., et al., A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. Protein Sci, 2009. 18(1): p. 121-33.

106. Osterberg, M., et al., Phenotypic effects of membrane protein overexpression in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A, 2006. 103(30): p. 11148-53.

107. Michel, E. and K. Wuthrich, High-yield Escherichia coli-based cell-free expression of human proteins. J Biomol NMR, 2012.

108. Endo, Y. and T. Sawasaki, Cell-free expression systems for eukaryotic protein production. Curr Opin Biotechnol, 2006. 17(4): p. 373-80.

109. Brown, P.O. and D. Botstein, Exploring the new world of the genome with DNA microarrays. Nat Genet, 1999. 21(1 Suppl): p. 33-7.

110. Arenkov, P., et al., Protein microchips: use for immunoassay and enzymatic reactions. Anal Biochem, 2000. 278(2): p. 123-31.

111. Zhu, H. and M. Snyder, Protein arrays and microarrays. Curr Opin Chem Biol, 2001. 5(1): p. 40-5.

112. Predki, P.F., Functional protein microarrays: ripe for discovery. Curr Opin Chem Biol, 2004. 8(1): p. 8-13.

113. Ramachandran, N., et al., Tracking humoral responses using self assembling protein microarrays. Proteomics Clin Appl, 2008. 2(10-11): p. 1518-27.

114. Ramachandran, N., S. Srivastava, and J. Labaer, Applications of protein microarrays for biomarker discovery. Proteomics Clin Appl, 2008. 2(10-11): p. 1444-59.

115. Kawahashi, Y., et al., *In vitro* protein microarrays for detecting protein-protein interactions: application of a new method for fluorescence labeling of proteins. Proteomics, 2003. 3(7): p. 1236-43.

116. Feilner, T., et al., High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. Mol Cell Proteomics, 2005. 4(10): p. 1558-68.

117. He, M. and M.J. Taussig, DiscernArray technology: a cell-free method for the generation of protein arrays from PCR DNA. J Immunol Methods, 2003. 274(1-2): p. 265-70.

118. He, M. and M.J. Taussig, Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). Nucleic Acids Res, 2001. 29(15): p. E73-3.

119. Ramachandran, N., et al., Self-assembling protein microarrays. Science, 2004. 305(5680): p. 86-90.

120. Tao, S.C. and H. Zhu, Protein chip fabrication by capture of nascent polypeptides. Nat Biotechnol, 2006. 24(10): p. 1253-4.

121. Ramachandran, N., et al., Next-generation high-density self-assembling functional protein arrays. Nature Methods, 2008. 5(6): p. 535-8.

122. He, M., et al., Printing protein arrays from DNA arrays. Nature Methods, 2008. 5(2): p. 175-7.

123. Roberts, R.W. and J.W. Szostak, RNA-peptide fusions for the *in vitro* selection of peptides and proteins. Proc Natl Acad Sci U S A, 1997. 94(23): p. 12297-302.

124. Biyani, M., Y. Husimi, and N. Nemoto, Solid-phase translation and RNA-protein fusion: a novel approach for folding quality control and direct immobilization of proteins using anchored mRNA. Nucleic Acids Res, 2006. 34(20): p. e140.

125. Kai, L. and Q. Kemao, Fast frequency-guided sequential demodulation of a single fringe pattern. Opt Lett, 2010. 35(22): p. 3718-20.

126. Studier, F.W. and B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol, 1986. 189(1): p. 113-30.

127. van den Ent, F. and J. Lowe, RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods, 2006. 67(1): p. 67-74.

128. Khare, D., et al., Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis. Proc Natl Acad Sci U S A, 2010. 107(32): p. 14099-104.

129. Zeidel, M.L., et al., Ultrastructure, pharmacologic inhibition, and transport selectivity of aquaporin channel-forming integral protein in proteoliposomes. Biochemistry, 1994. 33(6): p. 1606-15.

130. Borgnia, M.J., et al., Functional reconstitution and characterization of AqpZ, the E. coli water channel protein. J Mol Biol, 1999. 291(5): p. 1169-79.

131. Aoki, M., et al., Automated system for high-throughput protein production using the dialysis cell-free method. Protein Expr Purif, 2009. 68(2): p. 128-36.

132. Kim, D.M. and J.R. Swartz, Prolonging cell-free protein synthesis with a novel ATP regeneration system. Biotechnol Bioeng, 1999. 66(3): p. 180-8.

133. Chen, Y.J., et al., X-ray structure of EmrE supports dual topology model. Proc Natl Acad Sci U S A, 2007. 104(48): p. 18999-9004.

134. Bolen, D.W. and I.V. Baskakov, The osmophobic effect: natural selection of a thermodynamic force in protein folding. J Mol Biol, 2001. 310(5): p. 955-63.

135. Timasheff, S.N., Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. Proc Natl Acad Sci U S A, 2002. 99(15): p. 9721-6.

136. Uversky, V.N., J. Li, and A.L. Fink, Trimethylamine-N-oxide-induced folding of alpha-synuclein. FEBS Lett, 2001. 509(1): p. 31-35.

137. Meng, F., Y. Park, and H. Zhou, Role of proline, glycerol, and heparin as protein folding aids during refolding of rabbit muscle creatine kinase. Int J Biochem Cell Biol, 2001. 33(7): p. 701-9.

138. Tanaka, M., Y. Machida, and N. Nukina, A novel therapeutic strategy for polyglutamine diseases by stabilizing aggregation-prone proteins with small molecules. J Mol Med (Berl), 2005. 83(5): p. 343-52.

139. Tanaka, M., et al., Trehalose alleviates polyglutamine-mediated pathology in a mouse model of Huntington disease. Nat Med, 2004. 10(2): p. 148-54.

140. Yang, D.S., et al., Manipulating the amyloid-beta aggregation pathway with chemical chaperones. J Biol Chem, 1999. 274(46): p. 32970-4.

141. Selitrennikoff, C.P. and D.R. Sonneborn, The last two pathway-specific enzyme activities of hexosamine biosynthesis are present in Blastocladiella emersonii zoospores prior to germination. Biochim Biophys Acta, 1976. 451(2): p. 408-16.

142. Mio, T., et al., Saccharomyces cerevisiae GNA1, an essential gene encoding a novel acetyltransferase involved in UDP-N-acetylglucosamine synthesis. J Biol Chem, 1999. 274(1): p. 424-9.

143. Drew, D., et al., Optimization of membrane protein overexpression and purification using GFP fusions. Nature Methods, 2006. 3(4): p. 303-313.

144. Drew, D.E., et al., Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. FEBS Lett, 2001. 507(2): p. 220-4.

145. Klammt, C., et al., Cell-free production of G protein-coupled receptors for functional and structural studies. J Struct Biol, 2007. 158(3): p. 482-93.

146. Kai, L., et al., Preparative scale production of functional mouse aquaporin 4 using different cell-free expression modes. PLoS One, 2010. 5(9): p. e12972.

147. Pedelacq, J.D., et al., Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol, 2006. 24(1): p. 79-88.

148. Moritani, Y., et al., Direct integration of cell-free-synthesized connexin-43 into liposomes and hemichannel formation. FEBS J, 2010. 277(16): p. 3343-52.

149. Shimono, K., et al., Production of functional bacteriorhodopsin by an Escherichia coli cell-free protein synthesis system supplemented with steroid detergent and lipid. Protein Sci, 2009. 18(10): p. 2160-71.

150. Sanders, C.R. and R.S. Prosser, Bicelles: a model membrane system for all seasons? Structure, 1998. 6(10): p. 1227-34.

151. De Angelis, A.A. and S.J. Opella, Bicelle samples for solid-state NMR of membrane proteins. Nat Protoc, 2007. 2(10): p. 2332-8.

152. Ottiger, M. and A. Bax, Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. J Biomol NMR, 1999. 13(2): p. 187-91.

153. Poget, S.F. and M.E. Girvin, Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. Biochim Biophys Acta, 2007. 1768(12): p. 3098-106.

154. Uhlemann, E.M., et al., Cell-free synthesis of membrane subunits of ATP synthase in phospholipid bicelles: NMR shows subunit a fold similar to the protein in the cell membrane. Protein Sci, 2012. 21(2): p. 279-88.

155. Dittmer, J., et al., Incorporation of antimicrobial peptides into membranes: a combined liquid-state NMR and molecular dynamics study of alamethicin in DMPC/DHPC bicelles. J Phys Chem B, 2009. 113(19): p. 6928-37.

156. Lyukmanova, E.N., et al., Lipid-protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: Comparison with detergent micelles, bicelles and liposomes. Biochim Biophys Acta, 2012. 1818(3): p. 349-58.

157. Bay, D.C. and R.J. Turner, Diversity and evolution of the small multidrug resistance protein family. BMC Evol Biol, 2009. 9: p. 140.

158. Littlejohn, T.G., et al., Substrate specificity and energetics of antiseptic and disinfectant resistance in Staphylococcus aureus. FEMS Microbiol Lett, 1992. 74(2-3): p. 259-65.

159. Bay, D.C., K.L. Rommens, and R.J. Turner, Small multidrug resistance proteins: a multidrug transporter family that continues to grow. Biochim Biophys Acta, 2008. 1778(9): p. 1814-38.

160. Heir, E., G. Sundheim, and A.L. Holck, Identification and characterization of quaternary ammonium compound resistant staphylococci from the food industry. Int J Food Microbiol, 1999. 48(3): p. 211-9.

161. Jack, D.L., et al., A broad-specificity multidrug efflux pump requiring a pair of homologous SMR-type proteins. J Bacteriol, 2000. 182(8): p. 2311-3.

162. Paulsen, I.T., et al., The 3' conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. Antimicrob Agents Chemother, 1993. 37(4): p. 761-8.

163. Su, L.H., et al., Distribution of a transposon-like element carrying bla(CMY-2) among Salmonella and other Enterobacteriaceae. J Antimicrob Chemother, 2006. 57(3): p. 424-9.

164. Schluter, A., et al., Plasmid pB8 is closely related to the prototype IncP-1beta plasmid R751 but transfers poorly to Escherichia coli and carries a new transposon encoding a small multidrug resistance efflux protein. Plasmid, 2005. 54(2): p. 135-48.

165. Wu, S.W., et al., Characterization and nucleotide sequence of a Klebsiella oxytoca cryptic plasmid encoding a CMY-type beta-lactamase: confirmation that the plasmid-mediated cephamycinase originated from the Citrobacter freundii AmpC beta-lactamase. Antimicrob Agents Chemother, 1999. 43(6): p. 1350-7.

166. Chung, Y.J. and M.H. Saier, Jr., Overexpression of the Escherichia coli sugE gene confers resistance to a narrow range of quaternary ammonium compounds. J Bacteriol, 2002. 184(9): p. 2543-5.

167. Jung, J.S., et al., Molecular characterization of an aquaporin cDNA from brain: candidate osmoreceptor and regulator of water balance. Proc Natl Acad Sci U S A, 1994. 91(26): p. 13052-6.

168. Neely, J.D., et al., Heterotetrameric composition of aquaporin-4 water channels. Biochemistry, 1999. 38(34): p. 11156-63.

169. Yukutake, Y., et al., Mercury chloride decreases the water permeability of aquaporin-4-reconstituted proteoliposomes. Biol Cell, 2008. 100(6): p. 355-63.

170. Spirin, A.S., Cell-free systems of polypeptide biosynthesis and approaches to the evolution of translation apparatus. Orig Life, 1976. 7(2): p. 109-18.

171. Klammt, C., et al., Cell-free expression as an emerging technique for the large scale production of integral membrane protein. FEBS J, 2006. 273(18): p. 4141-53.

172. Sawasaki, T., et al., The wheat germ cell-free expression system: methods for high-throughput materialization of genetic information. Methods Mol Biol, 2005. 310: p. 131-44.

173. Tarui, H., S. Imanishi, and T. Hara, A novel cell-free translation/glycosylation system prepared from insect cells. J Biosci Bioeng, 2000. 90(5): p. 508-14.

174. Oliver, C.L. and C.D. Boyd, *In vitro* translation of messenger RNA in a rabbit reticulocyte lysate cell-free system. Methods Mol Biol, 1985. 2: p. 145-55.

175. de Zoysa, P.A., et al., Contrasting expression of KAL in cell-free systems: 5' UTR and coding region structural effects on translation. Protein Expr Purif, 1998. 13(2): p. 235-42.

176. de Smit, M.H. and J. van Duin, Control of translation by mRNA secondary structure in Escherichia coli. A quantitative analysis of literature data. J Mol Biol, 1994. 244(2): p. 144-50.

177. Haberstock, S., et al., A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. Protein Expr Purif, 2012. 82(2): p. 308-16.

178. Schwarz, D., V. Dotsch, and F. Bernhard, Production of membrane proteins using cell-free expression systems. Proteomics, 2008. 8(19): p. 3933-46.

179. Chandrasekaran, A., et al., A microscale platform for integrated cell-free expression and activity screening of cellulases. J Proteome Res, 2010. 9(11): p. 5677-83.

180. Khnouf, R., et al., Cell-free expression of soluble and membrane proteins in an array device for drug screening. Anal Chem, 2010. 82(16): p. 7021-6.

181. Ge, X., D. Luo, and J. Xu, Cell-free protein expression under macromolecular crowding conditions. PLoS One, 2011. 6(12): p. e28707.

182. Gu, L., et al., Metamorphic enzyme assembly in polyketide diversification. Nature, 2009. 459(7247): p. 731-5.

183. Busche, A., et al., Characterization of Molecular Interactions between ACP and Halogenase Domains in the Curacin A Polyketide Synthase. Acs Chemical Biology, 2012. 7(2): p. 377-385.

184. Tsumoto, K., et al., Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine. Biochem Biophys Res Commun, 2003. 312(4): p. 1383-6.

185. Drew, D., et al., GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. Nat Protoc, 2008. 3(5): p. 784-98.

186. Prive, G.G., Detergents for the stabilization and crystallization of membrane proteins. Methods, 2007. 41(4): p. 388-97.

187. Manley, G.T., et al., Aquaporin-4 deletion in mice reduces brain edema after acute water intoxication and ischemic stroke. Nat Med, 2000. 6(2): p. 159-63.

188. Furman, C.S., et al., Aquaporin-4 square array assembly: opposing actions of M1 and M23 isoforms. Proc Natl Acad Sci U S A, 2003. 100(23): p. 13609-14.

189. Strand, L., et al., Roles of aquaporin-4 isoforms and amino acids in square array assembly. Biochemistry, 2009. 48(25): p. 5785-93.

190. Harries, W.E., et al., The channel architecture of aquaporin 0 at a 2.2-A resolution. Proc Natl Acad Sci U S A, 2004. 101(39): p. 14045-50.

191. Horsefield, R., et al., High-resolution x-ray structure of human aquaporin 5. Proc Natl Acad Sci U S A, 2008. 105(36): p. 13327-32.

192. Fu, D., et al., Structure of a glycerol-conducting channel and the basis for its selectivity. Science, 2000. 290(5491): p. 481-6.

193. Ho, J.D., et al., Crystal structure of human aquaporin 4 at 1.8 A and its mechanism of conductance. Proc Natl Acad Sci U S A, 2009. 106(18): p. 7437-42.

194. Yang, B., A.N. van Hoek, and A.S. Verkman, Very high single channel water permeability of aquaporin-4 in baculovirus-infected insect cells and liposomes reconstituted with purified aquaporin-4. Biochemistry, 1997. 36(24): p. 7625-32.

195. Kozono, D., et al., Functional expression and characterization of an archaeal aquaporin. AqpM from methanothermobacter marburgensis. J Biol Chem, 2003. 278(12): p. 10649-56.

196. Werten, P.J., et al., Large-scale purification of functional recombinant human aquaporin-2. FEBS Lett, 2001. 504(3): p. 200-5.

197. Daniels, B.V., J.S. Jiang, and D. Fu, Crystallization and preliminary crystallographic analysis of the Escherichia coli water channel AqpZ. Acta Crystallogr D Biol Crystallogr, 2004. 60(Pt 3): p. 561-3.

198. Karlsson, M., et al., Reconstitution of water channel function of an aquaporin overexpressed and purified from Pichia pastoris. FEBS Lett, 2003. 537(1-3): p. 68-72.

199. Lian, J., et al., Improving aquaporin Z expression in Escherichia coli by fusion partners and subsequent condition optimization. Appl Microbiol Biotechnol, 2009. 82(3): p. 463-70.

200. Hovijitra, N.T., et al., Cell-free synthesis of functional aquaporin Z in synthetic liposomes. Biotechnol Bioeng, 2009. 104(1): p. 40-9.

201. Silberstein, C., et al., Membrane organization and function of M1 and M23 isoforms of aquaporin-4 in epithelial cells. Am J Physiol Renal Physiol, 2004. 287(3): p. F501-11.

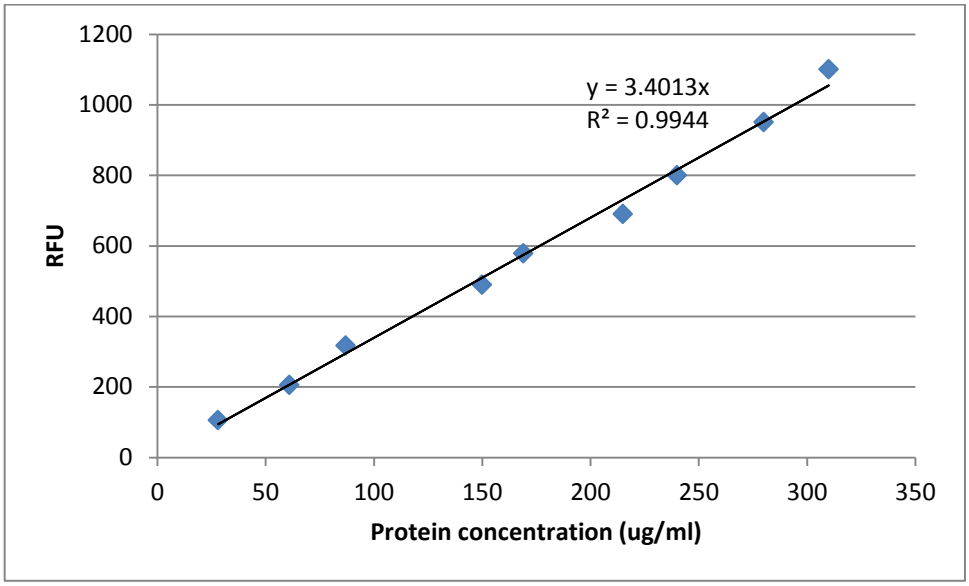
202. Yukutake, Y., et al., Rapid and reversible inhibition of aquaporin-4 by zinc. Biochemistry, 2009. 48(51): p. 12059-61.

203. Splinter, P.L., et al., AQP4 transfected into mouse cholangiocytes promotes water transport in biliary epithelia. Hepatology, 2004. 39(1): p. 109-16.

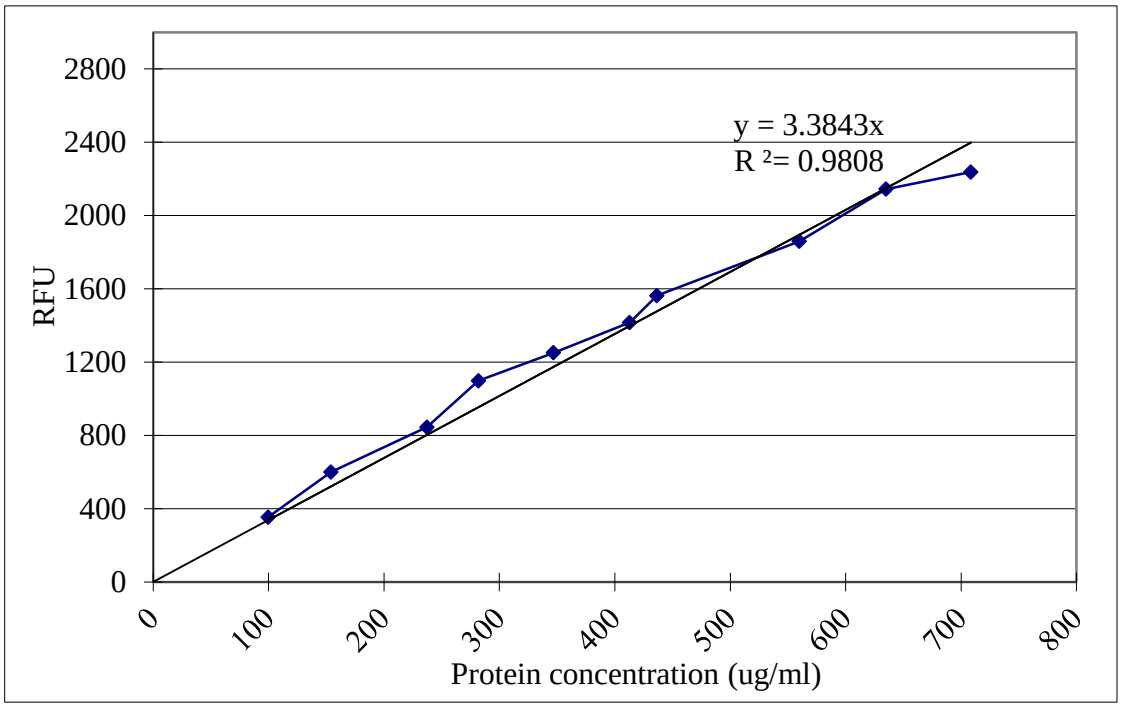
8. Appendix

8.1 Standard curves

8.1.1 Standard curve of sGFP

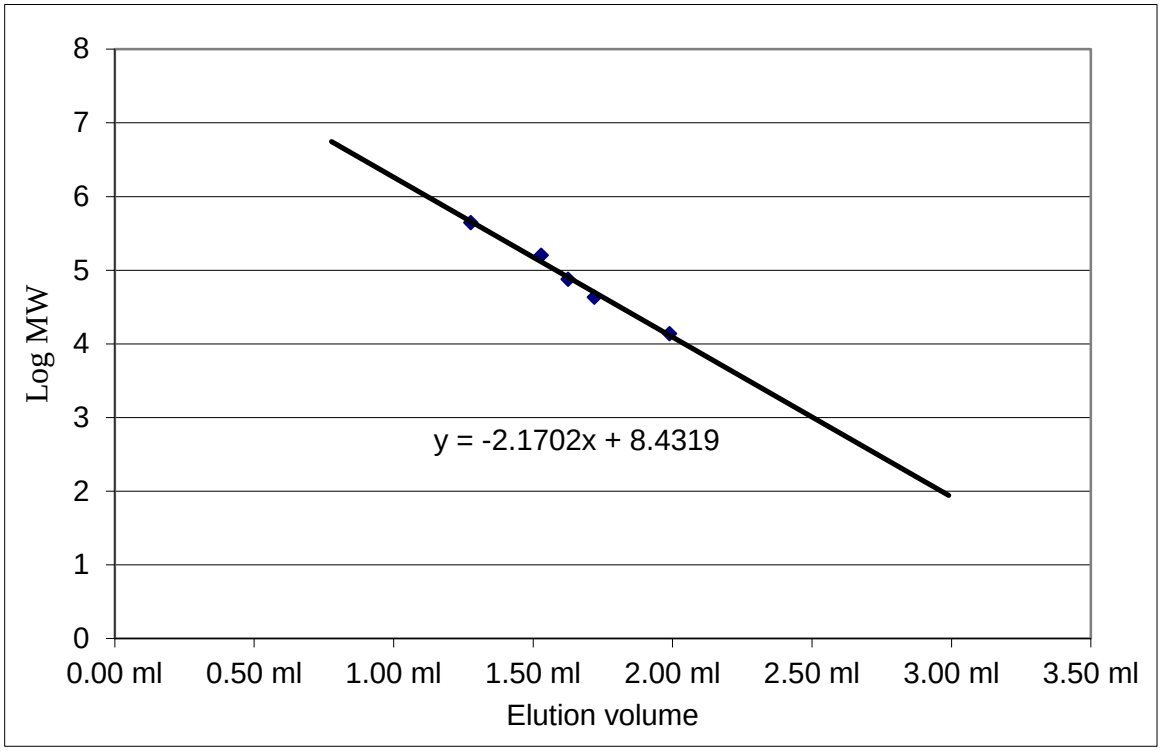


8.1.2 Standard curve of superfolderGFP

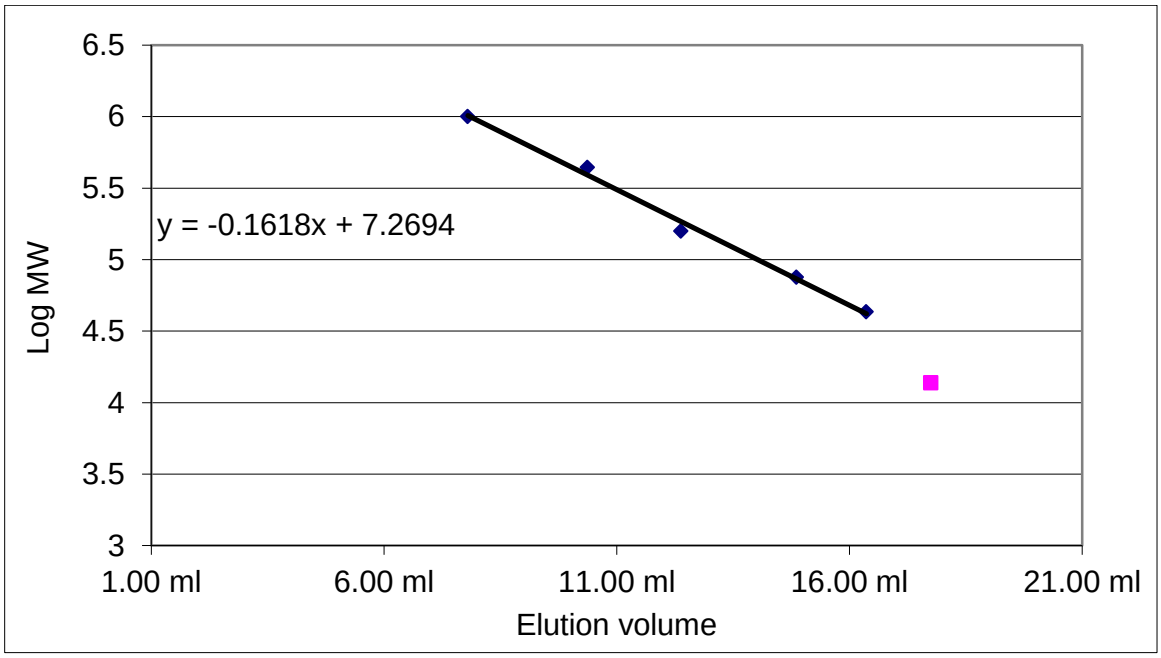


8.1.3 Calibration curve of size exclusion chromatography columns

Superdex 200 3.2/30



Superdex 200 10/300



Acknowledgement

First of all I would like to thank my supervisors Dr. Frank Bernhard, Prof. Volker Dötsch and Prof. Ralf Kaldenhoff for the opportunity to perform my Ph.D thesis research in their laboratories. Their enthusiasm for science and inspiration were great encouragement for me to solve all the challenges during my whole Ph.D work. Their open-minded characters and expertise in science create great atmosphere for do scientific research. With all this, they set up good examples of pursuing a scientific career.

Special thank goes to Dr. Frank Bernhard, who not only support my research work, but also help me really a lot in my personal life, which allowed me easily to get used to live in Frankfurt.

I would also like to thank Prof. Zhinan Xu in Zhejiang University, Hangzhou, China, who introduced me into the cell-free field, which became later on my main research topic of my Ph.D.

Great thanks to all the members of Dötsch group. Particularly, I want to thank Dr. Daniel Schwarz for introducing me to my Ph.D project and a lot of help for handling everything in the lab. I would like to thank Daniel Coutandin and Dr. Solmaz Sobhanifar for their nice discussion about my research and great help in the lab as well as the life in Frankfurt. I would also like to thank the member of cell-free group: Dr. Sina Reckel, Dr. Friederike Junge, Dr. Sebastian Richers, Davide Proverbio, Christian Roos, Stefan Haberstock, Yi Ma, Dr. Aisha Laguerre, Dr. Edith Buchinger and Christopher Hein. Their contribution and cooperation during daily research life provide a really good working atmosphere and made all the experiments more efficient. I want to thank Birgit Schäfer for her great help in organizing all the routine things in the lab and really a lot of help for my project. Particular thanks to Sigrid Oğuzer-Fachinger, who can always help me to find a solution to solve the problems of preparing all kinds of documentations and gave me the best suggestions. Again, thanks to all the people in the group.

I would like to thank for all the cooperation partners Dr. Yvonne Thielmann for the nice work in crystallization, Sven Siegfart for the AQP project.

Finally, I would like to thank my wife and my family, who always stand by my side and provide support whenever and wherever I need!

Curriculum Vitae

Lei, Kai

Institute of Botany, Applied Plant Science

Darmstadt University of Technology, Darmstadt, Germany

Institute of biophysical chemistry

Goethe-Universität Frankfurt am Main, Frankfurt am Main Germany

kaileism@yahoo.com.cn Tel : 49-69-79829626

Webpage: http://www.bpc.uni-frankfurt.de/doetsch/wiki/index.php/Lei_Kai

Personal details:

Date of birth: 14th, October, 1982
Place of birth: Xuzhou, P.R.China
Nationality: Chinese

Education:

2011.9- present	**TECHNICAL UNIVERSITY DARMSTADT** Ph.D student	**DARMSTADT GERMANY**
2008.9- present	**GOETHE-UNIVERSITÄT FRANKFURT AM MAIN** Guest PhD student	**FRANKFURT AM MAIN GERMANY**
2007.9-2008.9	**ZHEJIANG UNIVERSITY** Candidate for Doctoral degree in Biochemical Engineering	**HANGZHOU P.R.CHINA**
2005.9-2007.6	**ZHEJIANG UNIVERSITY** Candidate for Master degree in applied microbiology. Teaching assistant in the course of Microbiology Experiment.	**HANGZHOU P.R.CHINA**
2001.9-2005.7	**CHINA UNIVERSITY OF MINING AND TECHNOLOGY** Bachelor degree in Biological Engineering	**XUZHOU, P.R.CHINA**

Research:

2004-2005	Scientific and Research Training Program: The study of clarification agents of vinegar
2005-present	Throughput application of cell-free expression system Functional and structural analysis of membrane proteins based on cell-free expression system Cell-free based functional and structure analysis of Multidrug transporters and related proteins Isolation and characterization of a new bacterium capable of biotransforming cis-epoxysuccinic acid to D(-)-tartaric acid Purificaiton and characterization of a thermostable uricase from strain *Microbacterium* sp. Expression and characterization of a thermostable sarcosine oxidase (SOX) from *Bacillus sp*. in *Escherichia coli*.

Publications:

- **Lei Kai**, Volker Dötsch, Ralf Kaldenhoff, Frank Bernhard. Screening method for the co-translational stabilization of cell-free expressed proteins. Analytical Chemistry, submitted
- Christian Roos, **Lei Kai**, Davide Proverbio, Umesh Ghoshdastider, Slawomir Filipek, Volker Dötsch and Frank Bernhard. Co-translational insertion of cell-free expressed membrane proteins into lipid bilayers. Molecular Membrane Biology, 2012 Early Online.
- Christian Roos, **Lei Kai**, Stefan Haberstock, Davide Proverbio, Umesh Ghoshdastider, Yi Ma, Slawomir Filipek, Xiaoning Wang, Volker Dötsch1 and Frank Bernhard. High Level Cell-Free Production of Membrane Proteins into Nanodiscs. Methods in Molecular Biology, In Press
- Julia von Buelow, Annika Mueller-Lucks, **Lei Kai**, Frank Bernhard, and Eric Beitz. Functional characterization of a novel aquaporin from Dictyostelium discoideum amoebae implies a unique gating mechanism Journal of Biological Chemistry. (2012) Mar; 287: 7487–7494
- **Lei Kai**, Christian Roos, Stefan Haberstock, Davide Proverbio and Yi Ma, et al. Systems for the Cell-Free Synthesis of Proteins. Methods in Molecular Biology (2012), 1, Volume 800, Chemical Genomics and Proteomics, Part 2, Pages 201-225
- **Kai L**, Kaldenhoff R, Lian J, Zhu X, Dötsch V, Bernhard F, Cen P, Xu Z. Preparative Scale Production of Functional Mouse Aquaporin 4 Using Different Cell-Free Expression Modes. PLoS ONE (2010) 5(9): e12972
- Sobhanifar S, Reckel S, Junge F, Schwarz D, **Kai L**, Karbyshev M, Löhr F, Bernhard F, Dötsch V. Cell-free expression and stable isotope labelling strategies for membrane proteins. Journal of Biomolecular NMR. (2010) Jan; 46 (1): 33-43
- Xia Li, Tongcheng Xu, Hongbo Lu, Xiaohang Ma, **Lei Kai**, Kangping Guo, Yuhua Zhao. Purification and characterization of a cis-epoxysuccinic acid hydrolase from *Bordetella* sp. strain 1–3. Protein Expression and Purification. (2010) Jan;69 (1):16-20
- Rongrong Ma, Zhong Yang, Lei Huang, Xiangcheng Zhu, **Lei Kai**, Jin Cai, Xiaoning Wang, Zhinan Xu Construction of an efficient *Escherichia coli* cell-free system for *in vitro* expression of several kinds of proteins. Engineering in life Science (2010) Aug; 10 (4): 333-338
- **Lei Kai**, Xiao-hang Ma, Xiao-ming Jia, Xue-lai Zhou, Xia Li, Kang-ping Guo. Purification and characterization of a thermostable uricase from *Microbacterium* sp. strain ZZJ4-1. World journal of Microbiology and Biotechnology. (2008) 24:401–406
- Jiazhang Lian, Xiangming Fang, Jin Cai, Qixing Chen, Qiang Zheng, **Lei Kai**, Zhinan Xu. Efficient expression of membrane-bound water channel protein (Aquaporin Z) in *Escherichia coli*. Protein & Peptide Letters (2008) 15: 687-691
- Xia Li, Tongcheng Xu, Xiaohang Ma, Kangping Guo, **Lei Kai**, Yuhua Zhao, Xiaoming Jia, Yin Ma. Optimization of culture conditions for production of cis-epoxysuccinic acid hydrolase using response surface methodology. Bioresource Technology (2008) Sep; 99 (13): 5391–5396
- **Lei Kai**, Ke Yue, Xiao-hang Ma. Research and Application of Uricase. Bulletin of Science and Technology. (2008) 2: 183-187. (Chinese review paper)
- Yuanyuan Wang, Xiaohang Ma, Weifeng Zhao, Xiaoming Jia, **Lei Kai**. Study on the creatinase from *Paracoccus* sp. strain WB1, Process Biochemistry (2006) 41: 2072 – 2077
- Xia Li, Xiaohang Ma, Yuhua Zhao, Xiaoming Jia, **Lei Kai**, Kangping Guo, Weifeng Zhao. Isolation and characterization of a new bacterium capable of biotransforming *cis*-epoxysuccinic acid to D(-)-tartaric acid. FEMS Microbiology letters (2007) 267: 214-220
- Kangping Guo, Xiaohang Ma, Guiqin Sun, Yuhua Zhao, Xia Li, Weifeng Zhao, **Lei Kai**. Expression and characterization of a thermostable sarcosine oxidase (SOX) from *Bacillus* sp. in Escherichia coli, Appl Microbiol Biotechnol, (2006) 73: 559-566

Intership:

- 2004.7, summer Intern in Lvjian Dairy Company.
- 2004.7, Summer Intern in Nuobei Citric acid Company.
- 2004.8, Summer Intern in Hasen Pharmacy Company (producing Erythromycin and so on).

Workshops:

- SFB 807/TRAM Retreat – Naurod, 22nd-24th, March, 2010
- SMBP (Structural Biology of Membrane Proteins, Marie Curie Initial Training Networks FP-7-PEOPLE-2007-1-1-ITN) workshop of Protein crystallization in Grenoble, 15th-18th, March 2010
- SMBP workshop of cell-free expression of membrane proteins in Frankfurt as one of the organizers
- Workshop of Liquid Handling basic organized by TECAN, Männedorf, Switzerland, December 2010

Conferences:

- Invited talk: "Application of Detergents in a cell-free approach", Ringberg Symposium on Crystallization and Cell-Free Protein Production of Membrane Proteins, May 25th-28th, 2010
- Talk and poster: "High-throughput platform for cell-free expression and condition optimization: Analysis of cell-free expressed water channel proteins" The 2nd Annual Meeting of SBMPs, 8th-11th, June 2010, Basel, Switzerland.
- Talk: "Throughput applications of cell-free expression systems", poster: "Throughput strategies for cell-free expression protocol development" The 3rd Annual Meeting of SBMPs, 23rd-25th, May 2011, Utrecht, Netherland.
- Poster: "Throughput strategies for the optimization of membrane protein samples by cell-free expression", EUROMAR 2011, conjunction with the 33rd Discussion Meeting of the Magnetic Resonance Spectroscopy Division of the Gesellschaft Deutscher Chemiker, 21st-25th, August, 2011, Frankfurt/Main, Germany
- Poster: "Throughput applications of cell-free expression systems" was presented in GBM 2011 (International Symposium of the German Society for Biochemistry and Molecular Biology), 25th-28th, September, 2011, Frankfurt/Main, Germany
- Poster: "Throughput expression screening for optimized membrane protein sample" The 4th Annual Meeting of SBMPs, 30th May- 2nd June, 2012, Acquafredda di Maratea, Italy

“Word of honor”
I assure herewith on my word of honor, that I wrote this thesis by myself. All quotes, whether word by word, or in my own words, have been put in quotation marks or otherwise identified as such. The thesis has not been published anywhere else or has been presented to any other examination board.

Darmstadt, (date) ……

_____
(signature)