UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO DI BIOTECNOLOGIE

FACOLTÀ DI SCIENZE MM. FF. NN.

DOTTORATO DI RICERCA IN
BIOTECNOLOGIE MOLECOLARI, INDUSTRIALI ED AMBIENTALI
XXII CICLO

Heterologous expression of three enzymes for structural studies

S.S.D. Bio/11 Biologia Molecolare

Coordinatore: Prof. Roberto Bassi

Tutor: Prof. Hugo Luis Monaco

Dottoranda: Dott. Laura Civiero

Riassunto

In questo lavoro di tesi ci si è occupati dell'espressione, della purificazione e della cristallizzazione di tre enzimi (bile acid CoA:amino acid N-acyltransferase, BAAT; farnesyl cysteine-carboxyl methyltransferase, Ste14; e stearoyl-CoA desaturase, SCD) con lo scopo finale di determinarne la struttura tridimensionale mediante analisi di diffrazione di raggi X.
I cDNA delle proteine in questione sono stati clonati in vettori per l'espressione in sistemi eterologhi. Le scelta del sistema di espressione opportuno per la produzione su larga scala è stata condotta dopo una valutazione della resa, della stabilità ed integrità del prodotto proteico.

La proteina BAAT umana è stata espressa utilizzando il sistema procariotico *E.coli*. La proteina ricombinante è stata purificata tramite due cromatografie di affinità al nichel e gel filtrazione. L'integrità del campione ottenuto è stata controllata attraverso spettrometria di massa e l'omogeneità mediante DLS. E' stata condotta anche una valutazione dell'attività enzimatica verificando la presenza dei prodotti mediante spettrometria di massa. Oltre alla proteina *wild type* è stato prodotto il mutante privo di attività enzimatica (C235A) per procedere con le prove di cristallizzazione in presenza del substrato. I campioni ottenuti si presentano sufficientemente puri, stabili ed omogenei per allestire le prove di cristallizzazione. Le prove di cristallizzazione hanno dato esito positivo con la proteina mutata, anche se i cristalli fino ad ora ottenuti non sono idonei per gli esperimenti di diffrazione di raggi X.
La proteina di membrana Ste14 di *C.glabrata* è stata individuata come buon candidato per la produzione su larga scala al fine di condurre studi strutturali attraverso l'utilizzo di un sistema *high-throughput* pubblicato da Drew *et al*., 2008. Il sistema prevede l'espressione eterologa in *S. cerevisiae*; la proteina di interesse viene prodotta in fusione alla GFP permettendo la valutazione del livello di espressione e la stabilità nei diversi detergenti attraverso misure di fluorescenza sull'estratto cellulare. L'enzima è stato purificato attraverso due cromatografie di affinità al nichel e gel filtrazione in presenza sia del detergente DDM sia del detergente LDAO. Il profilo della gel filtrazione suggerisce che il campione è

‘monodisperso’ e quindi adeguato per l’allestimento di prove di cristallizzazione. Ad oggi non sono però stati ottenuti cristalli adatti ad esperimenti di diffrazione.
E’ stato infine messo appunto un protocollo di espressione e purificazione per l’enzima transmembrana umano SCD. La proteina ricombinante è stata espressa in cellule di insetto e purificata in presenza del detergente DDM attraverso IMAC. La resa e la purezza del prodotto proteico ad oggi non è ancora ottimale ma la quantità di enzima è sufficiente per allestire prove di cristallizzazione e per condurre studi biochimici e strutturali.

Abstract

This thesis work was aimed at the expression, purification and crystallization of three human proteins (bile acid CoA:amino acid N-acyltransferase, BAAT; farnesyl cysteine-carboxyl methyltransferase, Ste14; and stearoyl-CoA desaturase, SCD) in order to determine their three-dimensional structure using X-ray crystallography.

The cDNA sequences were cloned into specific vectors for overexpression in heterologous systems. The choice of the best expression system was made considering the protein yield, stability and integrity before scaling up.

Human BAAT was expressed in *E.coli*. The recombinant protein was purified by IMAC, reverse IMAC and gel filtration. The integrity of the sample was assessed by mass spectrometry and its homogeneity using DLS. The enzymatic activity was controlled verifying the presence of the products by mass spectrometry. In addition to the wild type protein, the catalytic mutant (C235A) was cloned and expressed to carry out crystallization trials in the presence of the substrate. The samples obtained are sufficiently pure, stable and homogeneous to set up crystallization trials. Crystals were obtained only with the mutant but they are not sufficiently ordered for X-ray diffraction experiments.

Using the high-throughput system published by Drew *et al*., 2008, the membrane protein Ste14 from *C.glabrata* was selected as a good candidate for a large scale production in order to conduct structural studies. The system uses *S.cerevisiae* to overexpress membrane proteins; the protein of interest is cloned into a GFP-fusion vector allowing to estimate the expression level and the stability in several detergents by measuring fluorescence directly in cell extracts. The enzyme was purified by IMAC, reverse IMAC and gel filtration in the presence of the detergents DDM and LDAO. The gel filtration profile indicates that the sample is ‘monodisperse’ and hence adequate to set crystallization trials. Up to now no suitable crystals were obtained.

Finally, a protocol for the expression and purification of the human membrane enzyme SCD was developed. The recombinant protein was expressed in insect cells and purified in the presence of the detergent DDM by IMAC. Until now the

yield and the purity obtained are not optimal but a sufficient quantity of the enzyme can be extracted to carry out crystallization trials aimed at structural and biochemical studies.

Table of contents

Abbreviations

BAAT: bile acid CoA:amino acid N-acytransferase
$C_{12}E_9$: polyethylene(9)dodecyl ether
CRB: cell resuspention buffer
DB: dyalisis buffer
DDM: n-dodecyl-b-D-maltopyranoside
DLS: dynamic light scattering
DM: n-decyl-b-D-maltopyranoside
DTT: dithiothreitol
EB: elution buffer
EDTA: ethylenediaminetetraacetic acid
FSEC: fluorescence size exclusion chromatography
GFP: green fluorescence protein
Icmt: isoprenylcysteine carboxyl methyltransferase
IMAC: immobilized metal affinity chromatography
IPTG: isopropil β-D-1-tiogalattopiranoside
LB: Luria-Bertani
LDAO: N,N-dimethyldodecylamine N-oxide
MRB: membrane resuspention buffer
PBS: phosphate buffered saline
PCR: polymerase chain reaction
PEG: poli-ethylene glycol
SCD: stearoyl-CoA desaturase
SDS: sodium dodecyl sulfate
SEC: size exclusion chromatography
Ste14: farnesyl cysteine carboxyl-methyltransferase
TBE: tris-borate-EDTA
TBS: tris buffered saline
TEV: tobacco etch virus
URA: uracil
YPD: yeast peptone dextrose

Chapter 1

Introduction

I. Protein X-ray crystallography

Protein X-ray crystallography is a powerful technique for obtaining atomic structural information. In the last decade this technique has found renewed interest since the genomes of humans and other organism were sequenced. This enabled the birth of high throughput crystallographic platforms aimed at solving the structures of all the encoded proteins, even in the absence of functional characterisations.

Determining the 3D structure of biological macromolecules by X-ray crystallography involves several steps including cloning, expression, purification, and crystallisation of the target molecule, acquiring diffraction data, phasing the data, model building and refinement. However, the preparation of protein crystals remains the main bottleneck for protein crystallography, especially in the case of membrane proteins.

I.I. Protein crystallization

A battery of different experimental setups and strategies exist for attempting to crystallize proteins. The starting point for crystallization is to obtain pure protein that is highly concentrated in a water-based buffer solution. The procedure for crystallising proteins involves disolving the protein in a mixture of precipitant agents, such as organic solvents (e.g., PEG, MPD), highly concentrated salt solutions (e.g., ammonium sulphate or sodium malonate), or combinations of both. In these conditions the solubility of the proteins is reduced resulting in the formation of an amorphous protein precipitate. However if the appropriate conditions are selected so that specific attractive interactions can occur between the protein molecules, this can lead to protein crystallization. The crystallization process can be differentiated into two steps: (1) the nucleation process and (2) crystal growth. Under the correctly selected conditions, crystal nucleation and growth can be achieved within the supersaturated regions of the phase diagram. Nucleation is the process by which molecules or non crystalline aggregates which are free in solution come together in such a way as to produce a

thermodynamically stable aggregate with a repeating lattice. The nucleation range is also called the labile zone, while the growth range is known as the metastable zone. For crystal nucleation to occur, the experimenter must push the protein solution into the labile zone, which is also the region where fast growth of crystal nuclei occurs. This means that for a successful experiment the protein/precipitant mixture must approach the nucleation zone very slowly so that the developing nuclei have enough time to growth. The transition from a stable solution to a supersaturated solution can be achieved by adjusting the position of the protein/precipitant mixture in the example phase diagram (Fig.1.1).

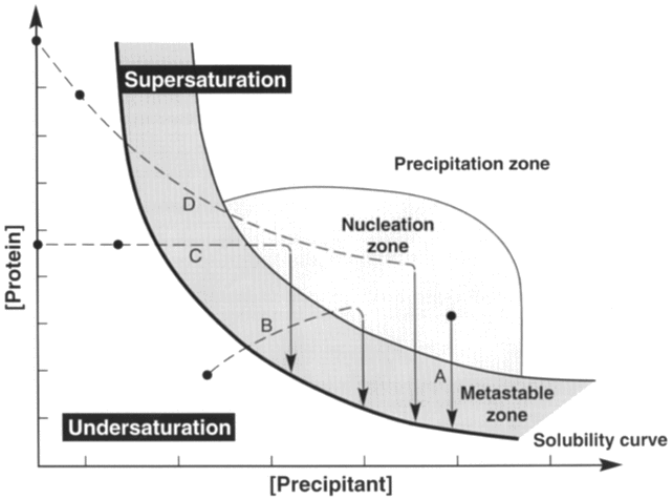


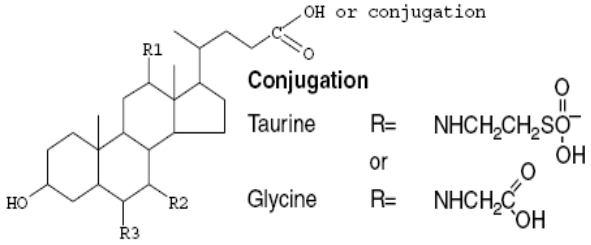
Fig.1.1. A schematic phase diagram showing the solubility of a protein in solution as a function of the concentration of the precipitant present.

These experiments are based on trial-and-error, and the selected precipitants, their concentration, and the pH range are selected based on previous crystallisation successes. In one of such screening experiments, hundreds of conditions are tested. The crystallisation conditions are screened for promising results, such as very small crystals or crystalline precipitates, that can guide further crystallization rounds. Null results also provide useful information since they suggest that the protein concentration should be increased. A screening of nearly random conditions is required due to uncertainty in the effects of agents like ions, temperature, and pH on protein crystal growth (McPherson, 1990).

A. Bile Acid Coenzyme A:Amino Acid N-Acyltransferase

A.1.1. Bile Acids

Cells utilize several enzymatic mechanisms to clear the excess of cholesterol. The major pathway, accounting for approximately 90% of cholesterol breakdown, is the enzymatic conversion of cholesterol into bile acids by enzymes in the liver.
The conversion is known to involve multiple hydroxylations of the ring structures and the oxidation and shortening of the side chain of cholesterol to produce a C_{24} bile acid; but the precise number of steps, the identities of the enzymes involved, and the mechanisms by which the pathway is regulated are currently largely unknown (Russell, 2009).
Bile acids are amphipathic molecules, which contain a sterol nucleus with hydroxyl groups and a side chain that terminates in a carboxyl group. Their amphipathic nature is essential to solubilize dietary lipids, which subsequently promotes their absorption in the digestive tract.



R1	R2	R3	
OH	OH	H	Cholic acid
H	OH	H	Chenodeoxycholic acid
OH	H	H	Deoxycholic acid
H	H	H	Lithocholic acid
H	OH	OH	Muricholic acid

Fig.1.2. Structure of selected bile acids (Houten *et al.*, 2006).

Bile acids that are synthesized from cholesterol in the hepatocyte are termed primary bile acids. Bile acids that are formed by bacterial modification of primary bile acids are termed secondary bile acids. The principal bile acids include in humans the primary bile acids cholic acid (CA) and chenodeoxycholic acid (CDCA), their glycine and taurine conjugates, and the secondary bile acids deoxycholic acid and lithocholic acid. In mice, CDCA is efficiently converted into muricholic acid (Fig.1.2).

Most of the bile acids are present within the enterohepatic organs. Usually bile acids are stored in the gallbladder; however, when a meal is ingested, they flow into the duodenum and intestine. Afterward they are efficiently (95%) reabsorbed by passive diffusion and active transport in the terminal ileum, and transported back to the liver via the portal vein. In the liver, bile acids are taken up at the basolateral (sinusoidal) membrane and exported again at the apical (canalicular) membrane of the hepatocytes into the bile canaliculus (transhepatic bile acid flux). This completes their enterohepatic circulation. Each bile acid molecule may complete 4–12 cycles between the liver and intestine per day. Owing to this efficient recirculation, only a small amount of the bile acid pool size is derived from *de novo* biosynthesis (Fig.1.3).

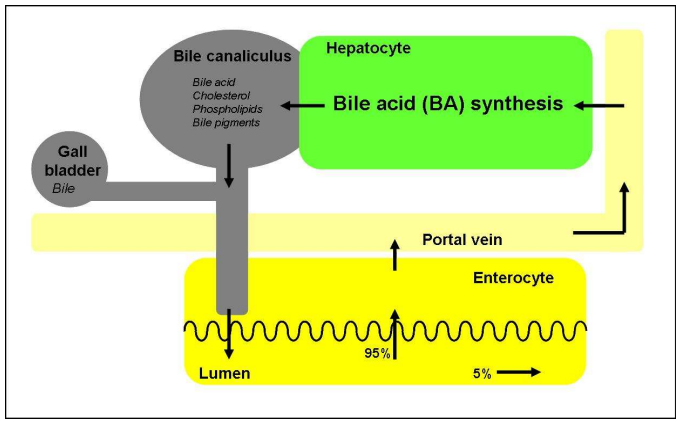


Fig.1.3. Schematic depiction of the enterohepatic circulation of bile acids.

Before the excretion of bile by the liver bile acids have to be conjugated to an amino acid. In normal bile, bile acids conjugated with glycine and taurine are the major solutes and unconjugated bile acids are almost nondetectable. Conjugated bile acids are less toxic and are more efficient promoters of intestinal absorption of dietary lipids than the unconjugated molecules (Solaas *et al*., 2000).

It has also become clear that bile acids besides their roles in dietary lipid absorption and cholesterol homeostasis, are also signaling molecules. Three major signaling mechanisms have been identified. Bile acids activate mitogen-activated protein kinase (MAPK) pathways, are ligands for the G-protein-coupled receptor (GPCR) TGR5, and activate nuclear hormone receptors such as farnesoid X receptor a (FXRa; NR1H4)(Houten *et al*., 2006; Chiang, 2002).

A.1.1.1. Bile acid biosynthesis

Cholesterol is converted into bile acids via multiple pathways which involve many different enzymes (Fig.1.4). Many of these enzymes are predominantly expressed in the liver and are localized in several different subcellular compartments.

In the classic pathway, bile acid biosynthesis begins with the conversion of cholesterol into 7α-hydroxycholesterol by cholesterol 7α-hydroxylase (CYP7A1). This is a microsomal cytochrome P450 enzyme localized exclusively in the liver. Its expression is highly regulated and shows a strong diurnal rhythm. The classic pathway involving CYP7A1 is the major pathway in bile acid biosynthesis because its contribution to total bile acid synthesis is 90% in humans and 75% in mice.

In the acidic pathway, oxysterols rather than cholesterol serve as substrates for 7α-hydroxylation. Sterol 27-hydroxylase, a mitochondrial cytochrome P450 (CYP27A1), forms predominantly 27-hydroxycholesterol but can also hydroxylate cholesterol at carbons 24 and 25. Alternatively, 24- and 25-hydroxycholesterol are also the product of separate cholesterol 24- and 25-hydroxylases. To be converted into bile acids, these oxysterols must undergo 7α-hydroxylation. Oxysterol 7α-hydroxylase (CYP7B1) catalyzes the conversion of 27-hydroxycholesterol and 25 hydroxycholesterol, and is highly expressed in liver. The CYP39A1 oxysterol 7α-hydroxylas acts on 24-hydroxycholesterol. The acidic pathway is responsible for the remaining bile acid synthesis (10% in humans and 25% in mice).

The next step in the ring structure modification involves isomerization of the double bond from the 5 to the 4 position and the oxidation of the 3β-hydroxyl-group to a 3-oxo-group. This step is catalyzed by the microsomal 3β-hydroxy-Δ5-C27-steroid oxidoreductase (HSD3B7), which can only handle 7α-hydroxylated intermediates derived from both cholesterol and oxysterols. At this point in the biosynthesis pathway, the intermediates can be modified by sterol 12α-hydroxylase (CYP8B1), and this will decide the fate of the produced bile acid. If the reaction is catalyzed by CYP8B1, another microsomal cytochrome P450, the final product will be cholic acid, if not either chenodeoxycholic acid (in humans) or muricholic acid (in mice) will be formed.

Subsequently, the products of HSD3B7, whether or not 12α-hydroxylated, are subject to reduction of the double bond by the enzyme Δ4-3-oxosteroid 5β-reductase (AKR1D1), which is a cytosolic enzyme. The final step of the ring modifications involves reduction of the 3-oxo-group to an alcohol-group and is catalyzed by 3α-hydroxysteroid dehydrogenase (AKR1C4).

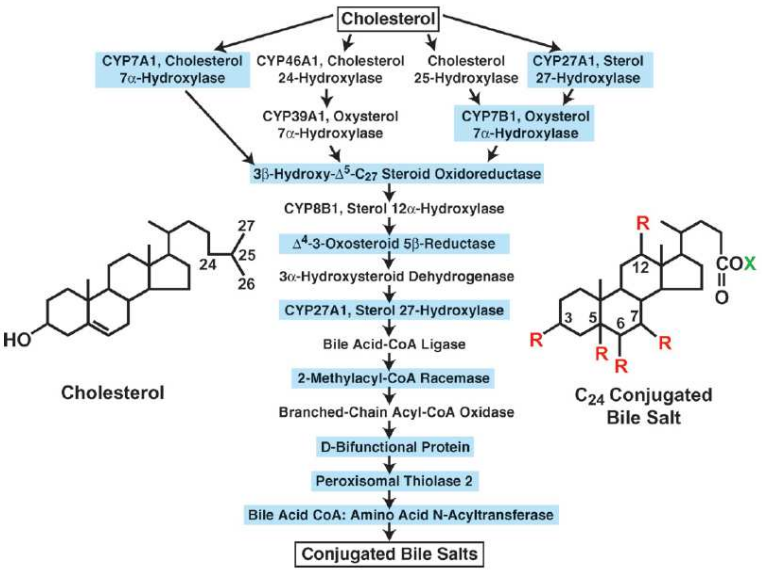


Fig.1.4. Enzymes of bile acids synthesis. The 16 enzymes involved and the 17 reactions are shown. Mutations causing human disease have been identified in the nine enzymes that are boxed in blue. The structure of cholesterol and a generic conjugated bile salt are shown. R= OH or H; X= glycine or taurine (Russell, 2009).

After the ring modifications, the sterol side chain is oxidized, activated and shortened by cleaving off propionyl-CoA. The first steps of this process are catalyzed by CYP27A1, which is the same mitochondrial enzyme that can initiate bile acid biosynthesis through formation of 27-hydroxycholesterol. The enzyme introduces a hydroxyl-group at carbon 27 and then oxidizes this group first to an aldehyde and then to a carboxylic acid. The products of this reaction are the bile acid intermediates 3α,7α-dihydroxycholestanoic acid (DHCA) and 3α,7α,12α-trihydroxycholestanoic acid (THCA), which subsequently have to be activated to their CoA-ester. Two enzymes have been identified which can perform this reaction. The first is the very long-chain acyl-CoA synthetase (VLCS, encoded by SLC27A2) localized at the endoplasmic reticulum and in the peroxisome, and the second is bile acyl-CoA synthetase (BACS, also called very long-chain acyl-CoA

synthetase homolog 2, encoded by SLC27A5) exclusively localized at the endoplasmic reticulum. BACS, which is only present in liver, also displays activity towards C_{24}-bile acids and is thought to be involved in the activation of bile acids which return to the liver via the enterohepatic circulation after deconjugation in the small intestine. It is unclear which of the two synthetases capable of activating C_{27}-bile acids is responsible *in vivo* for the CoA-ester formation of DHCA and THCA, and whether this occurs at the endoplasmic reticulum or in the peroxisome (Russell, 2003; Russell, 2009).

A.1.1.1.1. Peroxisomal steps in bile acid biosynthesis

The first peroxisomal enzyme required for the conversion of DHC-CoA and THC-CoA to chenodeoxycholoyl-CoA (CDCCoA) and choloyl-CoA (CA-CoA), respectively, is α-methylacyl-CoA racemase (AMACR). Its activity is necessary for this conversion, because only (25R)-stereoisomers are formed by CYP27A1, whereas the peroxisomal β-oxidation system can only handle (25S)-isomers. Thus, AMACR racemizes (25R)-DHC-CoA and (25R)-THC-CoA to their respective (S)-isomers, allowing subsequent β-oxidation of these substrates. In mouse and rat, THC-CoA oxidase (encoded by Acox2) oxidizes the CoA-esters of the C_{27}-bile acid intermediates. Its ortholog in humans is the branched chain acyl-CoA oxidase (BCOX, encoded by ACOX2), which is involved in the oxidation of the C_{27}-bile acid intermediates and of pristanoyl-CoA that has a methyl-group at the 2 position just like the side-chains of DHCA and THCA. The acyl-CoA oxidase desaturates (25S)-THC-CoA to 3α,7α,12α- trihydroxy-5β-cholest-24-en-26-oyl-CoA (24(E)-ene-THCCoA). The two subsequent reactions are catalyzed by D-bifunctional protein (DBP, also called multifunctional protein 2, multifunctional enzyme II, or D-peroxisomal bifunctional enzyme, encoded by HSD17B4). DBP consists of multiple functional units and is a stereospecific enzyme. Its hydratase unit converts 24(E)-ene-THC-CoA exclusively into (24R,25R)-3α,7α,12α,24-tetrahydroxy-5β-cholestanoyl-CoA (24-OH-THC-CoA), which is converted into 3α,7α,12α-trihydroxy-24-keto-5β-cholestanoyl-CoA (24-keto-THC-CoA) by the dehydrogenase unit of DBP. Finally, sterol carrier protein X (SCPx) cleaves 24-keto-THC-CoA to CA-CoA and propionyl-CoA. The β-oxidation of the side-chain of DHC-CoA is analogous to that of THC-CoA.

The last step in bile acid biosynthesis involves conjugation of CDC-CoA and CA-CoA with an amino acid. This reaction is catalyzed by the bile acyl-CoA: amino acid N-acyltransferase (BAAT), which is predominantly expressed in liver and is localized in peroxisomes and the cytosol. This dual localization suggests that the peroxisomal BAAT is responsible for conjugation of the newly formed primary bile acids within the peroxisome, whereas the cytosolic BAAT will be involved in reconjugation of the recycled primary and secondary bile acids that have been deconjugated in the gut. After conjugation, bile acids have to be transported from the peroxisome by an as yet unidentified transporter and are then excreted from the hepatocyte into the bile (Russell, 2003; Russell, 2009; Ferdinandusse, 2006).

A.1.1.1.2. Bile acid conjugation

C_{24} bile acids are conjugated with either glycine or taurine. Conjugation with taurine results in the formation of a molecule that is fully ionized and highly soluble at the small intestine pH during digestion (pH 6-7). In addition, conjugation results in a molecule that is always negatively charged and therefore towards which cell membranes are impermeable. Moreover, the bile salt molecule is too large to diffuse through the paracellular junctions of the biliary tract and small intestine. Impermeability in the apical membrane of cholangiocytes and enterocytes and the paracellular junctions between these cells is a key factor in promoting the high intraluminal concentration of conjugated bile acids in the biliary tract and small intestine. Conjugation with glycine converts a weak acid (pKa of 5) to a slightly stronger acid (pKa of 4). The result is increased solubility and ionization at the pH conditions prevailing during digestion. In man, intubation studies have shown that most glycine conjugated bile acids are in solution during digestion (Hofmann and Hagey, 2008).

A.1.2. Bile acid coenzyme A:amino acid N-acyltransferase

Bile acid coenzyme A:amino acid N-acytransferase (BAAT) is the enzyme involved in the addition of an amino acid, usually glycine or taurine, in amide linkage to carbon 24 (Fig.1.5). The N-acyltransferase is a remarkably efficient enzyme as more than 98% of bile acids excreted from the liver are amidated.

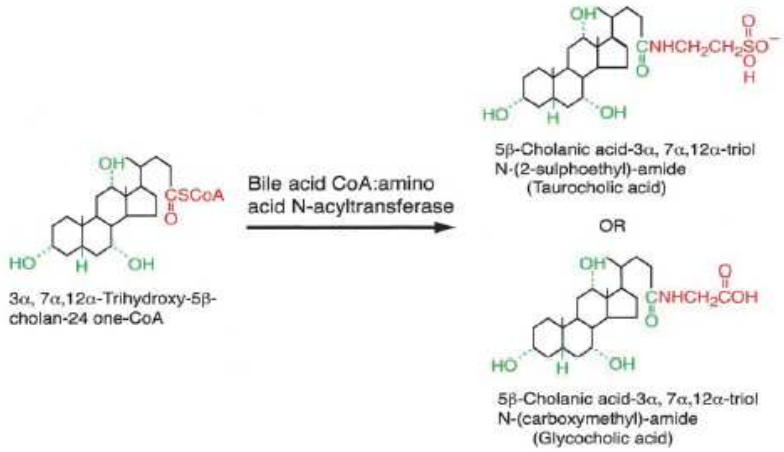


Fig.1.5. Conjugation of bile acids. The chemical modifications introduced by the enzyme are indicated in red on the products of the reaction. Cumulative changes to sterol intermediates are indicated in green (Russell, 2003).

BAAT activities have been identified in numerous species, including fish, seabream, dog, bovine, domestic fowl, rat, mouse and human. BAAT activity has been characterized not only in the liver of these animals, but activity has also been observed in the rat kidney (Shonsey *et al*., 2005).

Sequence similarity searches have detected significant similarity of BAAT to dienelactone hydrolase and other α/β hydrolases of known structure. The α/β hydrolase fold is shared by several enzymes that apparently have diverged from a common ancestor. Despite their different catalytic functions, these enzymes all contain a conserved nucleophile- histidine-acid catalytic triad with the histidine being a completely conserved amino acid and the nucleophile and acid loops accommodating more than one type of amino acid. Amino acid sequence alignments also revealed that hBAAT and other BAATs have high homology (above or equal to 40%) with peroxisomal, mitochondrial, and cytosolic long chain acyl-CoA thioesterases or acyl-CoA thioesterases.

Using mutation analysis it was established that hBAAT utilizes a catalytic triad composed of Cys-235, His-362, and Asp-328 with Cys-235 as the nucleophile. Catalysis in hBAAT may follow the classical charge-relay mechanism shared by serine proteases and thioesterases proposed by Pazirandeh *et al.*,1991. The reaction involves five steps: 1) nucleophilic attack of Cys-235 on the carbonyl carbon of bile acid-CoA substrate to form a tetrahedral intermediate; 2) decomposition of the tetrahedral intermediate to the acyl-enzyme intermediate

(CoA is cleaved from bile acid-CoA); 3) amino acid (glycine or taurine) attack on the carbonyl carbon of the acyl-enzyme intermediate; 4) formation of the second tetrahedral intermediate; and 5) generation of the bile acid-amino acid conjugate and recovery of the active enzyme (Sfakianos *et al*., 2002)(Fig.1.6).

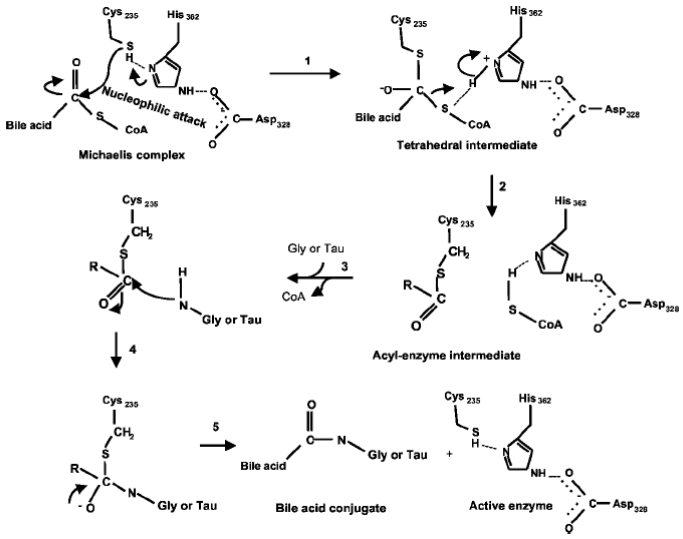


Fig.1.6. The catalytic mechanism of hBAAT (Sfakianos *et al*., 2002).

In recent evolutionary history (last 25 million years), conjugation with glycine has appeared in several branches of the mammalian kingdom. Interestingly, most monkeys in the New World are exclusively taurine conjugators, whereas Old World primates conjugate with both amino acids. Some mammals (pigs and rabbits) are principally glycine conjugators. Mouse BAAT is only active with taurine (Falany *et al*., 1997), but humans can conjugate bile acids with both taurine and glycine. The ratio of glycine to taurine conjugated bile acids in humans is solely dependent on the relative abundance of the two amino acids.

Several procedures have been described for the purification of BAAT from rat liver, human liver, domestic fowl liver and bovine liver. Kinetic studies about hBAAT revealed apparent K_m values for taurine and glycine of 1.1 and 5.8 mM, respectively, with corresponding V_{max}, values of 0.33, 0.19, and 0.77 pmol/min/mg protein (Johnson *et al*., 1991).

The cDNA encoding human BAAT was isolated and characterized by Falany *et al*., 1994. Sequencing and expression of the hBAAT cDNA validated the identity

of the purified BAAT. The K_m values for glycine (5.6±0.1 mM) and taurine (1.8±0.1 mM) in the formation of cholic acid conjugates catalyzed by the expressed purified enzyme were very similar to the values (5.8 and 1.1 mM, respectively) obtained with pure liver hBAAT.

Expression of the BAAT enzyme is induced by the farnesoid X receptor (FXR, NR1H4). FXR regulates bile acid and lipid homeostasis by acting as an intracellular bile acid-sensing transcription factor. Analysis of the human BAAT gene revealed the presence of functional response elements in the intronic region between exon 1 and 2. The response elements resemble the consensus FXR binding site consisting of two nuclear receptor half-sites organized as an inverted repeat and separated by a single nucleotide (IR-1).

These response elements directly bind FXR/retinoid X receptor (RXR) heterodimers and confer the activity of FXR ligands in transient transfection experiments. Further mutational analysis confirms that the IR-1 sequence of the BAAT gene mediates transactivation by FXR/RXR heterodimers. In addition, rats treated with the synthetic FXR ligand GW4064 clearly show increased transcript levels of BAAT mRNA (Pircher *et al*., 2003).

hBAAT and rBAAT have been reported to reside in peroxisomes with variable amounts present in the cytosol. Consequently, it has been proposed that peroxisomal BAAT is required for conjugation of *de novo* synthesized bile salts, whereas the cytosolic pool of BAAT is required for reconjugation of deconjugated bile salts returning from the intestine (Solaas *et al*., 2000; He *et al*., 2003). Some studies affirm that the majority of BAAT activity resides in the cytosol and approximately 15% of BAAT activity is present in the peroxisomal matrix (Styles *et al*., 2007); others suggest that the enzyme localizes specifically to peroxisomes (Pellicoro *et al*., 2007). Recently using photobleaching techniques it was found that GFP tagged BAAT was exclusively localized in the cytosol (Buch *et al*., 2009). hBAAT and rBAAT contain an established variant of PTS-1 (C-terminal SLK) at their C-termini, SQL. This signal is recognized by the cytosolic PTS-1 receptor Pex5p, which is responsible for sorting to the peroxisome. Probably the cytosolic localization is due to its carboxy terminal non-consensus peroxisomal targeting signal since mutation of the SQL to SKL resulted in BAAT being efficiently imported into peroxisomes.

A.1.2.1. BAAT deficiency

Four members of two Amish families have been reported with a homozygous mutation in the BAAT gene. The patients had increased serum bile acid levels and the bile acids in serum were unconjugated. Because unconjugated bile acids are not efficiently excreted in the bile, the bile acid concentrations in the intestine were hypothesized to be very low resulting in fat malabsorption, failure to thrive and a vitamin-K dependent coagulopathy. These clinical features correspond to a genetic syndrome known as Familial hypercholanemia (FHCA). Genomic sequencing of exons and exon-intron boundaries of BAAT identified a Met76-to-Val (M76V) mutation (Carlton *et al*., 2003).

Another boy had been described previously with no glyco- or tauro-conjugated C_{24}-bile acids, but at that time a specific BAAT deficiency had not been established with certainty. This patient had a similar clinical presentation with cholestasis associated with high serum and urine bile acid levels and low levels of fat-soluble vitamins.

B. Two membrane enzymes: Farnesyl Cysteine-Carboxyl Methyltransferase and Stearoyl-CoA 9-Desaturase

B.1.1. Membrane proteins

Membrane proteins continue to be among the most challenging targets in structural biology. All cells and organelles are contained within a hydrophobic lipid bilayer. Membrane proteins are involved in a wide variety of biological processes including photosynthesis, respiration, signal transduction, molecular transport and catalysis. These proteins represent between 20 and 30% of the proteomes of most organisms (Krogh *et al*., 2001) and more than 40% of drug targets (Overington *et al*., 2006) and yet very few structures of these molecules have been solved by X-ray crystallography or NMR. The first membrane protein structure was published in 1985 (Deisenhofer *et al*., 1985) and since then the number has increased slowly but steadily (White, 2009) (Fig.1.7).

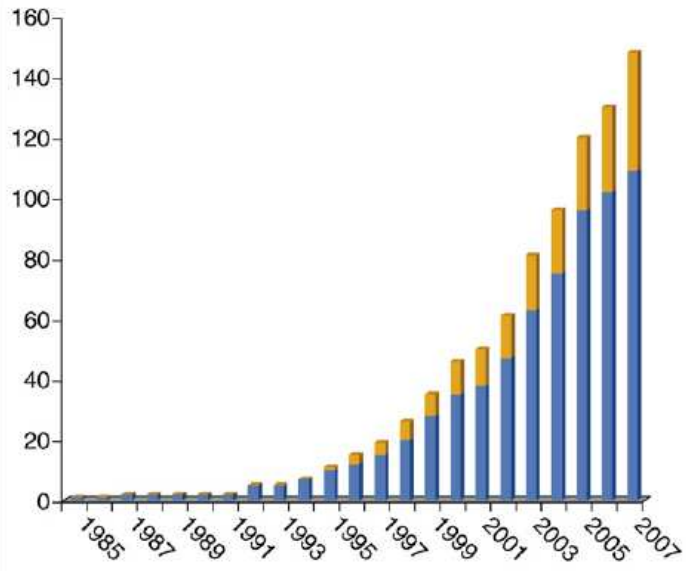


Fig.1.7. Growth of unique membrane protein structures deposited in the PDB. In blue, the number of structures of prokaryotic membrane proteins, and in yellow, the total number of eukaryotic structures. Both monotopic and multispanning proteins are included. For this study proteins are regarded as unique if they come from the same family but for different species (Carpenter *et al*., 2008).

To date there are over 50000 entries in the Protein Data Bank (PDB) repository of protein structures, but less than 1% of these entries represent membrane proteins. Of the 368 membrane protein structures in the 'Membrane proteins of known 3D structure' database (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html), 148 belong to unique proteins. Eukaryotic membrane proteins are particularly under-represented, with only 39 examples (Carpenter *et al.*, 2008).

In general membrane proteins occur in a wide distribution of sizes and shapes and can exist as monomers or in homo- hetero-oligomeric complexes. These proteins may also be completely buried into the bilayer or have large extra membranous domains.

Membrane proteins are typically classified in two ways: according to their interaction with the lipid bilayer (peripheral/integral membranes proteins) or according to the mechanism by which the proteins assemble into the membranes (costitutive/non-costitutive membrane proteins).

Proteins that bind or hardly penetrate the membrane surface without spanning it are referred to as monotopic or peripheral membrane proteins. These proteins can often be released from the membrane by an increase in ionic strength. Peripheral membrane proteins associate with lipid bilayers using different structural motifs such as amphipathic α-helices that orient parallel to the membrane plane, hydrophobic loops, or by the formation of covalent bonds to membrane lipids (lipidation). Integral membrane proteins span the bilayer once or several times. These proteins can only be released from the membrane by bilayer disruption with detergents. Integral transmembrane proteins can be divided into two classes according to the secondary structure adopted by the bilayer-embedded regions: α-helical bundles (e.g., rhodopsin) and β-barrels (e.g., porins).

Costitutive membrane proteins are assembled through a complex translocation/insertion process that involves synthesis by ribosomes that transiently attach to translocon proteins embedded in the membrane. Non-costitutive membrane proteins, which are generally pore forming toxins, exist in at least three different forms: water-soluble, bound to the membrane surface and with membrane embedded regions. In contrast to costitutive integral proteins, non-costitutive proteins spontaneously partition into the membrane without translocon complexes.

B.1.1.1. Structural biology of membrane proteins

Determining the atomic structure of a membrane protein is a very challenging task. The main difficulty is that large quantities of homogeneous, aqueous-soluble samples are required for most structural biology techniques. Thus, membrane proteins must be isolated from the lipid environment by solubilization with detergents. The use of detergents often causes protein instability, loss of activity, and denaturation.

In addition, membrane proteins are difficult to express and purify, often precipitate upon concentration, and are difficult to crystallize since detergents may interfere with the formation of protein-protein contacts. A variety of approaches have been used to crystallize membrane proteins. Diffraction quality crystals have been mainly obtained by the vapour diffusion method from detergent solubilised proteins (Iwata, 2003). In some cases these preparations have been aided by co-crystallization with Fv or Fab fragments, which provide additional hydrophilic regions for the formation of crystal contacts (Hunte and Michel, 2002). A more recent, but less popular approach to crystallizing membrane proteins is the meso crystallization method, which is based on the lipidic cubic phase (LCP) medium or mesophase (Caffrey, 2000; Caffrey, 2003).

B.1.1.2. Large-scale production of membrane proteins

Membrane proteins of known structure have been purified from natural sources, produced recombinantly or, in the case of short peptides, synthesised chemically. They have been successfully expressed in the bacteria *Escherichia coli* and *Lactococcus lactis*, the yeasts *Pichia pastoris* and *Saccharomyces cerevisiae*, in insect cells and in mammalian cell lines (Junge *et al*., 2008).

Despite major efforts, the expression of eukaryotic membrane proteins still represents a major bottleneck in the production of proteins for structural studies. Although there are several examples in the literature of successful expression of eukaryotic membrane proteins in *E. coli* (Grisshammer *et al*., 2005; Sarramegna *et al*., 2003) there is to date only one high resolution structure of an eukaryotic membrane protein expressed in *E. coli* (human 5-lipoxygenase-activating protein, Ferguson *et al*., 2007).

Thus, eukaryotic proteins are likely to require the use of eukaryotic systems for expression. Firstly, membrane proteins have to be targeted to the host cell membrane before they can fold correctly. Specific systems are required in the host cell such as the SRP-Sec61 system that inserts membrane proteins into the endoplasmic reticulum of eukaryotic cells. Secondly, membrane proteins are embedded in lipid, and the composition of these lipids varies among the systems. The nature of the lipids can affect the stability of the protein and consequently its likelihood of crystallisation. Thirdly eukaryotic proteins may undergo posttranslational modifications, such as glycosylation, and only higher eukaryotic cell lines provide the necessary machinery (Carpenter *et al*., 2008).

A number of membrane proteins expressed in *P.pastoris* have yielded high resolution structures including the mammalian voltage-dependent potassium channel (Long *et al*., 2005), the plant aquaporin (Törnroth-Horsefield *et al*., 2006) and the human LTC4 synthase (Molina *et al*., 2007) and most recently the human ABC multidrug transporter, P-glycoprotein (Aller *et al*., 2009). *S. cerevisae* has been used to produce protein which ultimately yielded the high resolution structure of P-type proton pump from *Arabidopsis thaliana* (Pedersen *et al*., 2007) and diffracting crystals of rabbit Ca^{2+}ATPase SERCA1a (Jidenko *et al*., 2005). Insect cell based expression systems have had spectacular successes for the production of GPCRs for structural studies, yielding three independent structures, the human adenosine A_{2A} receptor (Jaakola *et al*., 2008), the turkey β_1 adrenergic receptor (Warne *et al*., 2008) and the human β_2-adrenergic receptor (Cherezov *et al*., 2007). In addition, insect cells were also used to produce the chicken voltage dependent acid sensing ion channel which subsequently was crystallized and the 3D structure solved to 1.9 Å (Jasti *et al*., 2007).

The majority of researchers working in the area of eukaryotic membrane protein structural determination use one of these three systems for protein production. Other systems are available including the Semliki Forest Virus mammalian cell expression system (Lundstrom, 2003) and an alternative bacterial system using the gram positive bacterium *Lactococcus lactis* (Kunji *et al*., 2005).

Neither system has yet resulted in a high resolution structure but it is unclear whether this is due to inherent unsuitability of the systems for large scale production of eukaryotic membrane proteins for structural studies or to the fact

that these systems are not as extensively utilized as the yeast and insect cell systems.

Not many major advances have been made in terms of expression of eukaryotic membrane proteins in the last 5-10 years or so, although some researchers are performing careful, rational optimization of the existing systems with a view to maximizing quality and quantity of the target proteins. One exception to this is the development of cell-free expression systems for the production of membrane proteins (Schwarz *et al*., 2008; Liguori *et al*., 2007). This type of approach involves in-vitro production of proteins outside intact cells from a DNA or mRNA template using a basic set of biological building blocks obtained from cell lysate. A wide range of cell lysates are available and it has been shown that those from eukaryotic sources, eg. rabbit reticulocytes, are capable of post-translational modifications suggesting these may be suitable for production of eukaryotic membrane proteins. By its very nature the system removes the problem of cytotoxicity and also simplifies protein isolation as the number of contaminant proteins is markedly reduced. A further advantage of this system is the high level of control a user has, in terms of modifying media components to both increase the stability of the expressed proteins (eg protease inhibitors, lipids, co-factors) and to efficiently label for analysis by NMR or X-ray crystallography. A major advantage for membrane proteins is that the expressed membrane proteins are maintained in a soluble state in detergent micelles post-translationally since there are no native membrane environments for insertion. There are many examples of functionally active prokaryotic and eukaryotic membrane proteins being produced in this system (Schwarz *et al*., 2008).

As ever, work on eukaryotic membrane proteins lags behind but efforts are currently underway to exploit this system effectively for structural studies.

B.1.1.3. Membrane protein solubilisation

Membrane proteins are extracted from the host cell membranes by the addition of detergents, which disrupt the membrane and cover the hydrophobic surface of the protein allowing their solubilization. The choice of the detergent is a crucial part of the purification process. A number of different detergents are usually tested and

the one that extracts the largest amount of soluble, active, homogeneous, stable protein is used in all the purification steps (Carpenter *et al*., 2008).

Water solubility is provided by the hydrophilic portion of a detergent molecule. Based on the hydrophilic group, the detergents can be divided in ionic (cationic or anionic), nonionic, or zwitterionic. Ionic detergents, including sodium dodecyl sulfate (SDS), N-lauryl sarcosine, cetyltrimethylammoniumbromide (CTAB), and sodium cholate are effective in extracting proteins from the membrane. However, these detergents are harsh and tend to be denaturing because they efficiently disrupt both inter- and intra-molecular protein-protein interactions. SDS, for example, is commonly used as a membrane protein denaturant in quantitative protein unfolding/folding studies. Bile acid salts (i.e., sodium cholate, deoxycholic acid) are also ionic detergents; however, they tend to be milder than straight chain ionic detergents. Nonionic detergents, including maltosides, glucosides, and polyoxyethylene glycols are characterized by uncharged hydrophilic head groups. These detergents are mild and nondenaturing because they disrupt protein-lipid and lipid-lipid interactions rather than protein-protein interactions. Short chain (i.e., C7-C10) nonionic detergents are typically more deactivating than longer chain (i.e., C12-C14) nonionic detergents. A majority of detergents used in the purification and structural studies of membrane proteins (i.e., lauryl maltoside, octyl glucoside) are nonionic detergents. Zwitterionic detergents, including the Zwittergents®, Fos-Cholines®, CHAPS/CHAPSO, and amine oxides contain both a positive and negative charge in their hydrophilic head group. These compounds are electrically neutral like the nonionic detergents, but can often disrupt protein-protein interactions like the ionic detergents (Schimerlik, 2002; Iwata, 2003; Prive, 2007; Lin and Guidotti, 2009).

B.1.2. Farnesyl cysteine-carboxyl methyltransferase

In eukaryotic cells, a specific set of proteins are modified by C-terminal attachment of 15-carbon farnesyl groups or 20-carbon geranylgeranyl groups that function both as anchors for fixing proteins to membranes and as molecular handles for facilitating binding of these lipidated proteins to other proteins (Parish and Rando, 1996). Additional modification of these prenylated proteins includes

C-terminal proteolysis and methylation, and attachment of a 16-carbon palmitoyl group (Casey 1995; Casey and Seabra, 1996). Blocking protein prenylation is proving to be therapeutically useful for the treatment of certain cancers, infection by protozoan parasites and the rare genetic disease Hutchinson-Gilford progeria syndrome.

B.1.2.1. Prenylation process

Isoprenoids are lipids made up of five-carbon blocks, and they constitute one of the main classes of natural products. A subset of isoprenoids called prenyl groups are found on a variety of biological substances, including proteins. Prenylated proteins and peptides are found in most eukaryotes. They arise from the post-translational attachment of 15-carbon farnesyl or 20-carbon geranylgeranyl groups to the C-terminal segment of proteins.

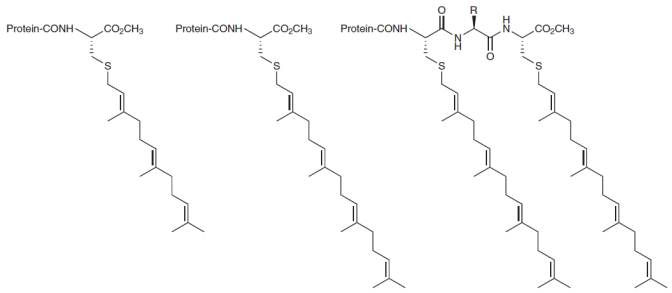


Fig.1.8. Structures of the C termini of prenylated proteins. Farnesylated protein (left); geranylgeranylated protein (middle); doubly geranylgeranylated protein (right)(Gelb *et al.*, 2006).

Fig.1.8 shows the structures of the C termini of prenylated proteins. Known farnesylated proteins include nuclear lamins A and B, which form a filamentous structural layer on the inner side of the nuclear membrane; the γ subunit of the heterotrimeric G protein transducin, which functions in visual signal transduction in the retina; all three isoforms of Ras GTP-binding proteins, which function in cellular growth regulation; the large-antigen component of the hepatitis δ virus; and yeast a-factor (Gelb *et al.*, 2006).

All farnesylated proteins contain a C-terminal sequence CAAX, where C is cysteine, A is usually but not necessarily an aliphatic amino acid and X is one of a

variety of amino acids. After attachment of the farnesyl group via a thioether linkage to the cysteine residue, the last three amino acids, AAX, are removed by a prenyl protein-specific endoprotease, and the α-carboxyl group of the newly exposed farnesylated cysteine is methylated by a prenyl protein–specific methyltransferase. Known geranylgeranylated proteins include the γ subunit of heterotrimeric G proteins that function at the plasma membrane, and many small GTP-binding proteins. The Rab subfamily of GTP-binding proteins contain a pair of geranylgeranyl groups on adjacent or nearly adjacent cysteines at the C terminus of the protein (Gelb *et al*., 2006).

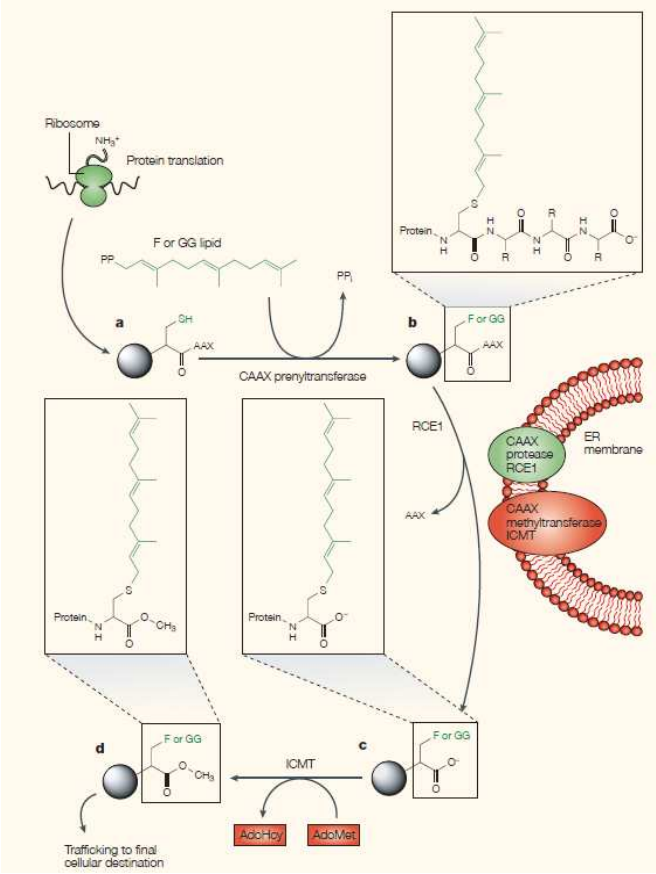


Fig.1.9. An overview of CAAX-protein processing. a) Proteins that contain a carboxy-terminal CAAX motif are initially modified by the cytosolic CAAX prenyltranferases farnesyltransferase or geranylgeranyltransfersase-I. b) Following prenylation, the CAAX protein travels to the endoplasmic reticulum (ER), where the C-terminal three amino acids (AAX) are proteolytically removed by RCE1. c) The protein is then carboxyl methylated by isoprenylcysteine carboxyl methyltransferase (Icmt). d) Following methylation, fully processed CAAX proteins are directed to their appropriate cellular location, which is often the cytoplasmic surface of cell membranes (Winter-Vann and Casey, 2005).

Three distinct protein prenyltransferases that attach prenyl groups to proteins have been identified. Protein farnesyltransferase (PFT) transfers the farnesyl group from farnesyl diphosphate to the cysteine SH of the CAAX motif at the C terminus of proteins. Protein geranylgeranyltransferase type I (PGGT-I) catalyzes a similar reaction using geranylgeranyl diphosphate as the prenyl donor. Protein geranylgeranyltransferase type II (also known as Rab geranylgeranyltransferase) catalyzes the transfer of both geranylgeranyl groups from geranylgeranyl diphosphate to two cysteine SH groups at the C terminus of Rab proteins. Much is known about the structure and catalytic properties of these enzymes. All three are heterodimeric proteins that use active site-bound Zn^{2+} as a catalytic cofactor (Zhang *et al*., 2000; Long *et al*., 2002). The specificity rules that dictate which CAAX motifs are acted on by PFT, which are acted on by PGGT-I and which are not prenylated at all are not completely understood The X residue of CAAX is probably the most important element for differential recognition by PFT versus PGGT-I, but the AA portion of CAAX is also important (Reid et al., 2004). The postprenylation CAAX processing machinery includes the endoprotease Rce1 (Ras converting enzyme I), which releases an intact AAX tripeptide from the newly prenylated CAAX protein, and isoprenylcysteine carboxylmethyltransferase (Icmt), which transfers a methyl group from *S*-adenosylmethionine to the α-carboxyl group of the prenyl cysteine. The RCE1 CAAX protease, is an integral membrane protein that resides in the endoplasmic reticulum. The catalytic mechanism of RCE1 does not allow this protein to be neatly categorized into any previously known enzyme class, but sequence and motif analysis has been used to argue that RCE1 is a metalloprotease. The final CAAX-directed processing step also takes place at the endoplasmic reticulum. Icmt was first identified in yeast, on the basis of its requirement for processing the a-factor mating pheromone and, like the endoproteases, is an integral membrane protein, which is unusual for a methyltransferase. Following modification by Icmt, the fully processed C terminus of a CAAX protein consists of a methyl-esterified farnesyl or geranylgeranylcysteine. Similar to RCE1-mediated proteolysis, the methylation of CAAX proteins is essential for mouse development; mice lacking Icmt die by embryonic day (Winter-Vann and Casey, 2005; Gelb *et al*., 2006) (Fig.1.9).

B.1.2.1.1. Therapeutic intervention based on protein prenylation

Regardless of how the protein uses its modified cysteine residue, one aspect is very clear - the modifications are crucial for the biological activities of the proteins. For this reason, the CAAX proteins prenyltransferases have been the focus of drug-discovery programs for over a decade. For example, interest in understanding protein farnesylation for the purpose of cancer therapy grew out of the finding that one of the most common cancer-promoting elements in cells, the Ras proteins, require farnesylation to transform cells into tumors progenitors. Early preclinical studies showed that farnesyltransferase inhibitors (FTIs) had high efficacy and surprisingly low toxicity, and clinical trials of two such compounds have reached Phase III trials. Although these experimental agents have shown significant activity in several clinical trials, particularly in haematological cancers, the overall response rates in patients with solid tumours have been less successful than was initially hoped. One frequently cited possibility for this is the phenomenon known as 'alternate prenylation', which allows some FTase substrates to be modified by GGTase-I and thereby escape inhibition by FTIs (Winter-Vann and Casey, 2005). Following the discovery that RAS proteins are mislocalized in cells that lack the RCE1 protease or Icmt methyltransferase (Kim *et al*., 1999; Bergo *et al*., 2000), these post-prenylation-processing steps have attracted attention as potential targets in oncogenesis, because even alternately prenylated forms of CAAX proteins would probably be sensitive to inhibition of both enzymes.

B.1.2.2. Isoprenylcysteine carboxyl methyltransferase family

Prenylation allows otherwise hydrophilic proteins to associate with cellular membranes, presumably by the insertion of the prenyl lipid into the phospholipid bilayer. Such an insertion would bring the α-carboxyl group of the prenylcysteine into proximity with the negatively charged head groups that are predominant among the phospholipids of the cytosolic leaflet of the plasma membrane. Methyl esterification removes the negative charge on the cysteine and thereby eliminates the electrostatic repulsion that would otherwise occur. Thus, carboxyl methylation of prenylproteins is believed to be a mechanism to increase their affinity for the

plasma membrane. Methyltransferases catalyze the transfer of a methyl group from methyl donors such as *S*-adenosylmethionine to various methyl acceptor substrates:

S-adenosyl-L-methionine + protein C-terminal S-farnesyl-L-cysteine

=

S-adenosyl-L-homocysteine + protein C-terminal S-farnesyl-L-cysteine methyl ester.

Ste14 from *S.cerevisiae* was the first member of the prenylcysteine protein carboxyl methyltransferase family to be cloned and sequenced (Sapperstein *et al.*, 1994). Subsequently, several homologues of Ste14 have been identified. These include *S. pombe* mam4p, *X. laevis* mam4p, and the human homologue Icmt (Imai *et al.*, 1997; Dai *et al.*, 1998). In addition, database searches identified a mouse EST and two *C. elegans* open reading frames that have significant amino acid sequence homology to Ste14. There is strong conservation among these family members, with mam4p having the highest (44%) amino acid sequence identity to Ste14. The members of this family have similar hydropathy profiles and constitute one of the few subgroups of methyltransferases that contain multiple membrane spans. Most of the amino acid conservation between these family members occurs at their C-termini, suggesting that residues important for CAAX methyltransferase activity are located at the C-terminus. Like *S. cerevisiae* Ste14, the homologues *S. pombe* mam4p, *Xenopus* mam4p, and human Icmt have been shown to possess in vitro CAAX methyltransferase activity (Imai *et al.*, 1997; Dai *et al.*, 1998). These CAAX methyltransferases are able to transcomplement other CAAX methyltransferase family members in vivo (Romano *et al*., 1998). Together these proteins form a novel family of protein prenylcysteine carboxyl methyltransferases. To date the only kinetic studies of an isoprenylcysteine methyltransferase activity were performed on the bovine methyltransferase in retinal rod outer segment membranes (Shi and Rando, 1992) and, most recently, on recombinant human Icmt (Baron and Casey, 2004). They have demonstrated that the kinetic mechanism of CAAX protein methylation by Icmt is an ordered sequential mechanism in which the S-adenosyl-L-methionine substrate associates first with the enzyme and the S-adenosyl-L-homocysteine product dissociates at the end.

B.1.2.1.1. Farnesyl cysteine-carboxyl methyltransferase from *S.cerevisiae*

Ste14 is a 26 KDa, ER-multispanning enzyme. Ste14 contains no known protein motifs, and the residues involved in its function and localization have not been well identified. Romano and Michaelis, 2001 proposed that the protein contains six membrane spans, two of which forming a helical hairpin and has the majority of its hydrophylic regions located in the cytosol.

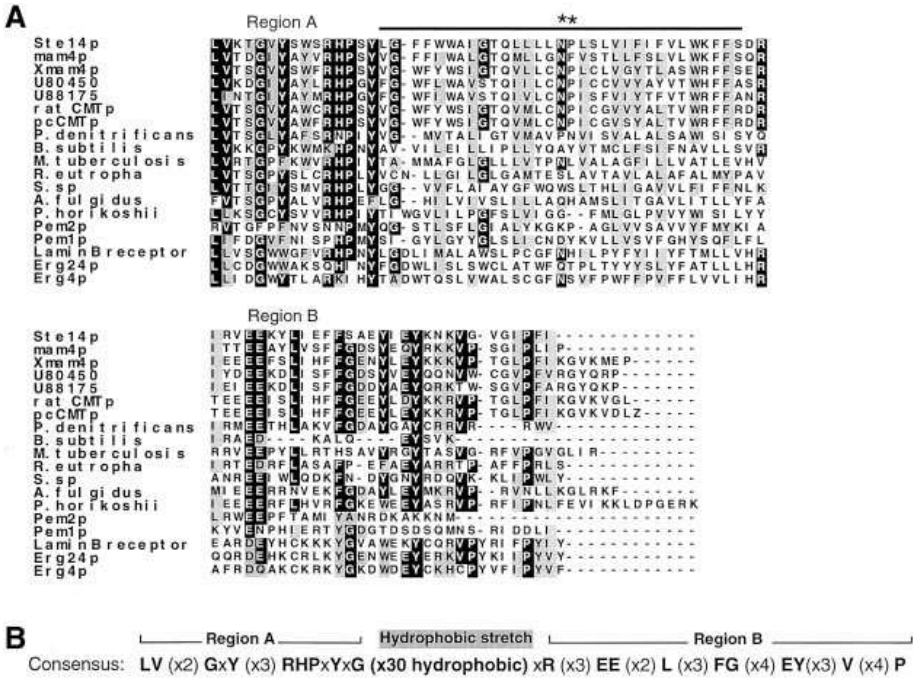


Fig.1.10. A) Alignment of the C-terminal region of *S. cerevisiae* Ste14 and its homologues (*S. pombe* mam4p, *X.laevis* Xmam4p, 2 *C. elegans* open reading frames, a rat open reading frame, and human Icmt), open reading frames from *Pseudomonas denitrificans*, *Bacillus subtilis*, *Mycobacterium tuberculosis*, *Ralstonia eutropha*, *Synechocystis sp*, *Archaeoglobus fulgidus*, and *Pyrococcus horikoshii*, the *S. cerevisiae* lipid methyltransferases Pem1p and Pem2p, the sterol biosynthesis enzymes from *S. cerevisiae* Erg4p and Erg24p, and the human lamin B receptor, which is an Erg24p homologue. Black boxes denote amino acid identity and gray boxes denote amino acid similarity. The bar above the sequence denotes the putative membrane span of Ste14. B) Consensus sequence is based on the alignment shown in A. Only residues that are identical in ≥11 of the proteins in the alignment were scored as consensus residues; other residues are shown as an x. The hydrophobic region was determined by a hydropathy plot (Romano and Michealis, 2001).

It was also shown that five out of six loss-of-function mutations obtained by random mutagenesis lie in residues that are conserved between Ste14 and its Icmt

homologues. In addition, by database search and sequence comparisons a C-terminal region of Ste14 was identified that is conserved between the Icmt family of protein methyltransferases, two phospholipid methyltransferases and three sterol reductases. This region defines a consensus sequence that was designated the RHPxY-hyd-EE motif (Fig.1.10). Site-directed and random mutations indicate that residues in this region play a critical role in Ste14 function.

Recently the topology of human Icmt was also mapped; Icmt traverses the ER membrane eight times (Wright *et al*., 2009) (Fig.1.11). Comparison of the structure of human Icmt with that of Ste14 reveals several striking features.

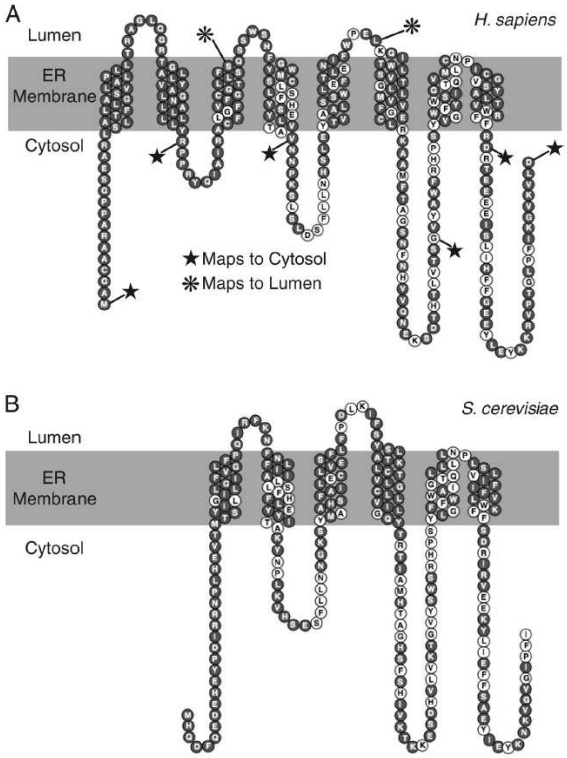


Fig.1.11. A) Topology of Icmt deduced from data derived from the fluorescent probe and glycosylation reporter (Wright *et al.*, 2009). Stars indicate sites that mapped to the cytosol, and asterisks indicate sites that mapped to the lumen of the ER. B) Topology of Ste14 of *S. cerevisiae* as described by Romano and Michaelis, 2001.

First, the luminal loops of both enzymes are minimal, consistent with the fact that the catalytic face of the enzyme must be cytosolic since that is where the substrates are found. Second, the structure of human Icmt TM3-8 corresponds closely to that of Ste14, suggesting that all of the catalytic activity residues are in

the C-terminal portion of mammalian Icmt. Third, residues conserved between human Icmt and Ste14 occur both in the TM segments as well as in the cytoplasmic loops and the C-terminal cytoplasmic tail. Therefore, the conserved residues in the cytosolic loops and tail of the enzyme are likely to be involved in substrate binding and/or catalysis. The fourth, seventh, and eighth TM segments have clusters of conserved amino acids, suggesting that they also participate in substrate binding and/or catalysis.

B.1.2.1.2. Farnesyl cysteine-carboxyl methyltransferase from *A.thaliana*

The plant CAAX processing machinery is evolutionarily conserved. The Arabidopsis CAAX proteases AtSte24 and AtRce1 can complement corresponding yeast mutants promoting a-factor processing. Proteolysis of plant prenylated proteins by AtRce I was sufficient to promote their plasma membrane localization in yeast cells. All CAAX processing enzymes are localized in the ER, like their homologs in yeast and mammalian cells (Bracha-Drori *et al*., 2008). Plant protein extracts have been shown to contain Icmt activity, and two genes encoding Icmt have been cloned. The two genes were designated AtIcmtA/AtSte14A and AtIcmtB/AtSte14B (Narasimha Chary *et al*., 2002). It was shown that AtIcmtA and AtIcmtB are differentially expressed and that in vitro AtIcmtB had higher catalytic activity. Based on a topology model, Ste14 contains six membrane spans, with two forming a helical hairpin. AtIcmtB and AtIcmtA have similar topology and contain between six to eight transmembrane helices.

B.1.3. Stearoyl-CoA 9-desaturase

Fatty acid desaturases introduce a double bond in a specific position of long-chain fatty acids. Unsaturation of a fatty acid chain is a major determinant of the melting temperature of triglycerides as well as of the fluidity of biological membranes that are made up of a bilayer of phospholipids. Thus, fatty acid desaturases are conserved across kingdoms. In addition to this fundamental function of maintaining the physical property of phospholipids and triglycerides, another class of unsaturated fatty acids such as arachidonic acid (20:4 n-6) and

docosahexaenoic acid (22:6 n-3) is essential for many physiological functions in animals including humans.

B.1.3.1.Classification and characteristic features of desaturases

Fatty acid desaturases are nonheme iron-containing enzymes. Delta desaturases create a double bond at a fixed position counted from the carboxyl end of fatty acids, whereas omega desaturases act on a specific position counted from the methyl end of a fatty acid. This reaction requires molecular oxygen, NAD(P)H, an electron transport system (ferredoxin-NADPH reductase and ferredoxin, or cytochrome *b*5 reductase and cytochrome *b*5), and a terminal desaturase.

Desaturases can be classified into two groups: soluble desaturases and membrane-bound desaturases. Acyl-acyl carrier protein (ACP) desaturases are soluble desaturases, which are exclusively localized in the plant plastid. These enzymes require NADPH and oxygen, and are associated with an electron transport sequence comprising ferredoxin-NADPH reductase and ferredoxin. Acyl-ACP desaturases contain two atoms of iron and the two D/EXXH motifs of amino acid sequence involved in binding the di-iron complex (Shanklin and Cahoon, 1998; Sperling *et al*., 2003).

Membrane-bound desaturases can be divided into two subgroups. One is the acyl-lipid desaturases. This group of enzymes is localized in the membranes of cyanobacterial thylakoid, plant endoplasmic reticulum, and plastid. Acyl-lipid desaturases use either ferredoxin (in cyanobacteria and plant plastid) or cytochrome *b*5 (in plant ER) as an electron donor. Ferredoxin, a soluble protein, acts as an electron donor for both acyl-ACP desaturases and acyl-lipid desaturases in plant plastid. Acyl-lipid desaturases in cyanobacteria and plant plastid can desaturate stearic (18:0) and oleic (18:1 n-9) acyl groups in monogalactosyl diacylglycerol (cyanobacteria and plant plastid) and in phosphatidylglycerol (plant plastid), whereas plant ER desaturases mostly use fatty acid in phosphatidylcholine (Tocher *et al*., 1998).

The other subgroup of membrane-bound desaturases is the acyl-coenzyme A (CoA) desaturases. These desaturases are present in ER membrane and use fatty acyl-CoAs as substrates. Like ER-bound acyl-lipid desaturases, acyl-CoA desaturases require cytochrome *b*5 as an electron donor. Acyl-CoA desaturases

are present in animals including insects and nematodes as well as in fungi. All mammalian desaturases identified are acyl-CoA desaturases. Analysis of the predicted amino acid sequences of membrane-bound desaturases indicates that these enzymes contain two long hydrophobic domains that would be capable of spanning the membrane bilayer twice. Sequence comparison has also revealed the existence of three regions of conserved His-box motifs that contain eight histidine residues: HX3-4H, HX2-3HH, and H/QX2-3HH. These histidine residues are potential ligands of iron atoms and act at the catalytic center of desaturases (Nakamura and Nara, 2004).

There are many similarities between the chemistry performed by soluble and membrane-bound desaturases but there are also notable differences (Shankin *et al*., 2009). The crystal structure of acyl-ACP desaturase shows that these desaturases are homodimeric proteins, with each monomer folded into a compact single domain composed of nine helices. The diiron active site of these enzymes is buried within a core four-helix bundle and is positioned alongside a deep, bent, narrow hydrophobic cavity in which the substrate is bound during catalysis. It is a textbook example of a lock-and-key type of binding site in which the bound fatty acid moiety is poised for formation of the *cis*-fatty acid product (Moche *et al*., 2003; Whittle *et al*., 2005; Guy *et al*., 2007).

The membrane desaturase active site, on the other hand, seems to have a fundamentally different architecture from that of the soluble desaturases, perhaps having a cleft into which substrates enter laterally rather than a deep binding cavity into which substrates enter in extended conformation as in the soluble desaturases. The two classes of desaturases exhibit also different rate-limiting steps. For the membrane desaturases, the initial C–H bond-breaking step is rate-limiting. The limiting step for the soluble desaturase may be the release of acyl-ACP product, which involves the energetically unfavorable solvation of the hydrophobic product as it leaves the hydrophobic cavity and enters the aqueous phase. This is in contrast to what is observed in membrane desaturases, in which the lipid-linked product is released into the lipids of the membrane.

Three desaturases are known in humans. Stearoyl CoA desaturases (SCDs, also called Δ9 desaturases) catalyze the synthesis of monounsaturated fatty acids, whereas Δ6 desaturase (D6D) and Δ5 desaturase (D5D) are required for the synthesis of highly unsaturated fatty acids.

B.1.3.1.1.Mammalian stearoyl-CoA desaturase

In mammals, the SCD reaction is an aerobic process requiring molecular oxygen, NAD(P)-cytochrome *b*5 reductase, and the electron acceptor cytochrome *b*5. The electrons flow from NAD(P)H via cytochrome *b*5 reductase, to cytochrome *b*5, to SCD, and finally to O_2, which is reduced to H_2O. The enzyme complex introduces a single double bond at the Δ9,10 position of long-chain acyl-CoAs (Fig.1.12) either from de novo synthesis or from the diet. Based on kinetic isotope data of plant desaturase, the current hypothesis for the desaturation reaction is that the enzyme removes hydrogen atoms starting with that at the C-9 position, followed by the removal of the second hydrogen atom from the C-10 position.

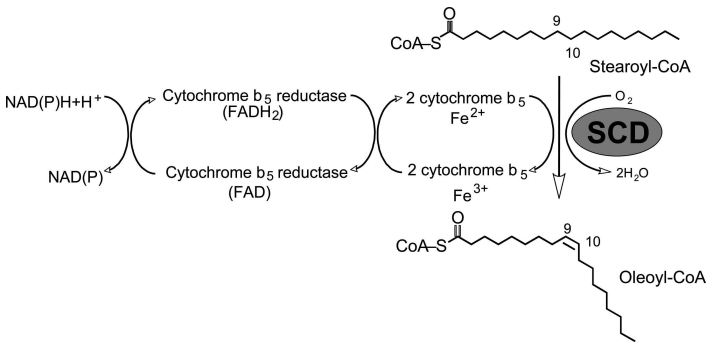


Fig.1.12. The pathway of electron transfer in the desaturation of fatty acids by stearoyl-CoA desaturase (Paton and Ntambi, 2008).

This stepwise mechanism is highly specific for the position at which the double bond is introduced, implying that the C-9 and C-10 bond is accurately positioned with respect to the diiron center of the enzyme (Paton and Ntambi, 2008).

The genes for SCD have been cloned from different species, including yeast, *D.melanogaster*, *C.elegans*, sheep, hamster, rat, mice, and human. In the mouse, four isoforms (SCD-1, SCD-2, SCD-3, and SCD-4) have been identified, whereas in the rat, two isoforms have been characterized. Humans express two Δ-9 desaturase genes designated SCD-1 and SCD-5, with the former exhibiting nearly ubiquitous expression and the latter being restricted primarily to the brain and pancreas.

SCD-1 contains four transmembrane domains with both the NH_2 and COOH termini oriented toward the cytosol. The single cytoplasmic loop and the COOH

terminus contain the eight histidine residues known to form the His box, which binds iron within the catalytic center of the desaturase (Fig.1.13). The two ER luminal loops are relatively small compared with the cytosolic loop, which houses two of the three conserved His motifs. Purified SCD-1 migrates as a 37-kDa band on SDS-PAGE with the His segments located at positions 119, 156, and 296. These regions are expected to be necessary to provide ligands for nonheme iron within the catalytic site of the enzyme.

Studies with tissue cultures (Gomez *et al*., 2002), of mouse, and rat livers (Heinemann *et al*., 2003) have shown that mammalian SCD activity is tightly regulated at the transcriptional, translational, and post-translational levels. Transcription is under control of the PPARγ signaling pathway (Miller and Ntambi, 1996), and may also respond to various secondary messengers such as polyunsaturated fatty acids. Translational regulation is mediated through mRNA stability by endoplasmic reticulum membrane proteins, like Mga2p (Martin *et al*., 2007), and, as recently suggested, the 15-lipoxygenase-differentiation control element, a conserved feature in mammalian 3′-untranslated regions (Ren *et al*., 2004). Other studies in murine microsomes (Heinemann *et al*., 2003) and differentiated mouse 3L3-L1 adipocytes (Kato et al., 2006) showed post-translational proteolytic processing of the N-terminus of the enzyme, subsequent inactivation, and proteosome-directed degradation.

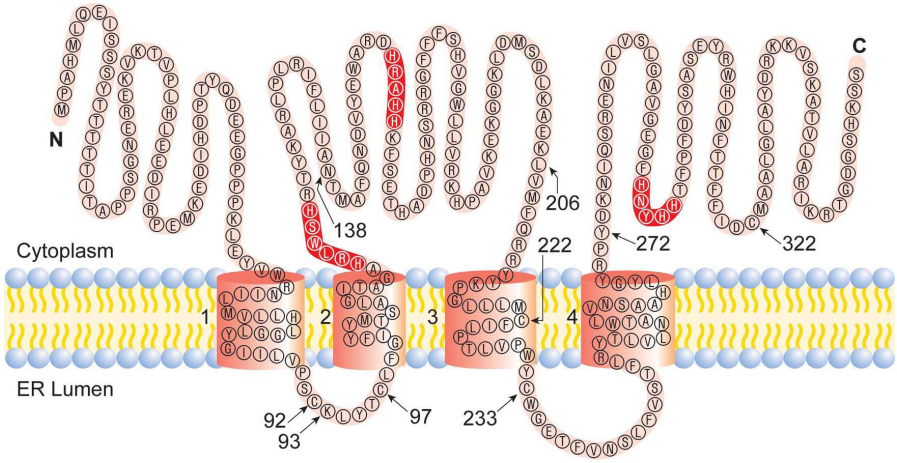


Fig.1.13. Proposed model for the membrane topology of mouse SCD-1. The highlighted residues represented the conserved histidine regions, which are catalytically essential(Paton and Ntambi, 2008).

Recent studies on SCD1 suggest that the enzyme is a critical control point of lipid homeostasis and body-weight regulation. As several manifestations of the metabolic syndrome and type 2 diabetes mellitus – including insulin resistance, hypertension, dysfunction of pancreatic b-cells and liver steatosis – are associated with intracellular lipid accumulation, it was proposed that SCD may be a potential therapeutic target in the treatment of obesity and the metabolic syndrome. Increased SCD activity was found in humans and animals with hepatic steatosis, whereas SCD1 deficiency ameliorates both high-fat diet and genetically induced liver steatosis. SCD1-/- mice are resistant to diet-induced obesity, have increased energy expenditure, reduced adiposity and increased insulin sensitivity. In addition, SCD1 is involved in adipocyte differentiation and could be important in the synthesis of *N*- oleoylethanolamide, which is known to inhibit food intake and stimulate lipolysis. Taken together, the findings reveal SCD to be a promising therapeutic target for the treatment of obesity, diabetes, liver steatosis and other metabolic diseases. However, the potential use of an SCD inhibitor as a human therapeutic agent awaits a more complete understanding of the mechanism underlying the effects of SCD deficiency and indication that the inhibition of this enzyme is both safe and efficacious (Dobrzyn and Ntambi, 2005; Flowers and Ntambi, 2009).

C. Aim of the project

The aim of this thesis work is the heterologous expression, purification and crystallization of three enzymes in order to determine their three-dimensional structure by means of X-ray protein crystallography.

Firstly, the human bile acid coenzyme A:amino acid N-acytransferase (BAAT) was chosen for a large-scale production. BAAT is a key enzyme in bile acid biosynthesis and catalizes the conjugation of bile acids to glycine and taurine for excretion into bile. Bile acid amidation is important for the absorption of fat soluble substances, such as fat soluble vitamins, in humans. Mutations in the enzyme due to genetic defects are correlated to several pathologies linked to the lack of these essential substances. BAAT is believed to contain an α/β hydrolase catalytic domain in the C-terminal region, with a Cys-His-Asp catalytic triad. However, until now no structural information is available for this enzyme. Analysis of its biochemical and structural features may help to understand the catalytic mechanism and its proposed but not yet confirmed role in the protection against the accumulation of free bile acids..

Secondly, a new GFP-based scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae* was adopted to quickly identify those targets that could be selected as good targets for crystallization studies (Newstead *et al*., 2007; Drew *et al*., 2008). Using this protocol, many genes can be rapidly cloned by homologous recombination into a GFP-fusion vector and their overexpression potential determined using whole-cell and in-gel fluorescence. The quality of the overproduced eukaryotic membrane protein-GFP fusions can then be evaluated using confocal microscopy and fluorescence size-exclusion chromatography (FSEC). The purification of targets that pass the quality criteria can be scaled up to obtain the sufficient amount of pure membrane protein for structural studies.

The most promising targets were selected among 96 *C.glabrata* membrane proteins. This part of the work was carried out in the MPC group, Imperial College, London under the supervision of the Prof. B. Byrne. The project started from a collaboration with Prof. K.Haynes, Imperial College, London. The Haynes

group is interested in understanding the molecular basis of fungal virulence and their research focuses on the pathogen *C.glabrata*, the major fungal pathogen in humans. The application of the new GFP- based method was useful to sort out interesting *C.glabrata* protein targets for functional and structural studies.

Thirdly, it was decided to study the human membrane enzyme stearoyl-CoA desaturase (SCD), a key and highly regulated enzyme that is required for the biosynthesis of monounsaturated fatty acids. SCD catalyzes the Δ9-*cis* desaturation of a range of fatty acyl-CoA substrates; the preferred substrates are palmitoyl- and stearoyl-CoA, which are converted into palmitoleoyl- and oleoyl-CoA respectively. SCD appears to be an important metabolic control point, and inhibition of its expression could be of benefit for the treatment of obesity, diabetes, and other metabolic diseases.

The multiple layers of regulation of desaturase activity, in diverse organisms, emphasize the importance of this enzyme in cellular function, but also complicate recombinant expression using living hosts. Until now there are no reports on a successful purification and no structural information about this class of proteins.

The structure determination of a member of the membrane desaturases would drastically improve our understanding of the structure-function relationships of this protein family. In this thesis work, preliminary attempts to identify a protocol for the expression and purification of functional desaturases from different organisms were carried out.

Chapter 2

Materials and Methods

2.1.Cell strains

2.1.1. *Escherichia coli* strains

XL1-Blue were used for routine cloning applications.

Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*qZΔ*M15* Tn*10* (Tetr)]

BL21 (DE3) cells were used for protein expression.

They are an all purpose strain for high-level protein expression and easy induction.
Genotype: *E. coli* B F⁻ *dcm ompT hsdS*(rB– mB–) *gal* λ(DE3)

BL21 (DE3) C41 cells were used for protein expression.

They are effective in overexpressing toxic proteins from all classes of organisms, eubacteria, archaea, yeast, plant, mammals. The strain C41(DE3) was derived from BL21(DE3) [*E. coli* B F⁻ *dcm ompT hsdS*(rB– mB–) *gal* λ(DE3)]. This strain has at least one uncharacterized mutation, which prevents cell death associated with the expression of toxic recombinant proteins.

BL21 CodonPlus(DE3)-RIPL cells were used for membrane protein expression.

BL21-CodonPlus strains are engineered to contain extra copies of genes that encode the tRNAs that most frequently limit translation of heterologous proteins in *E. coli*. Availability of tRNAs allows high-level expression of many heterologous recombinant genes in BL21-CodonPlus cells that are poorly expressed in conventional BL21 strains.
Genotype: *E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte [*argU proL*Camr] [*argU ileY leuW* Strep/Specr]

Like the parental BL21(DE3) strain, C41(DE3) and CodonPlus(DE3) carry the lambda DE3 lysogen which expresses T7 RNA polymerase from the l*ac*UV5

promoter by IPTG induction. As *E. coli* B strains, these are general protein expression strains that lack both the *lon* protease and the *ompT* outer membrane protease, which can degrade proteins during the purification steps.

2.1.2. *Saccharomyces cerevisiae* strains

FGY217 were used for membrane protein expression.

In this cell line the vacuolar endopeptidase Pep4p is deleted. Deletion of the *pep4* gene not only inhibits Pep4p protease activity but also reduces the levels of other vacuolar hydrolases.
Genotype: *MATa, ura3-52, lys2Δ201, pep4Δ*

2.1.3. Insect cell lines

SF9 (Invitrogen™).

This line originated from the IPLBSF-21 cell line, derived from the pupal ovarian tissue of the fall army worm, *Spodoptera frugiperda*. Sf9 cells grow both in monolayer and suspension culture and they are adaptable to a serum-free medium. The Sf9 cell line is a clonal isolate of IPLBSF21-AE (Sf21). They are suitable for transfection, plaque purification, generating high titer stocks, plaque formation and expression of recombinant proteins.

High Five ™ (Invitrogen™).

This line is originated from the ovarian cells of the cabbage looper, *Trichoplusia ni*. They double in less than 24 hours and grow well in adherent cultures, but form irregular monolayers. They are adaptable to suspension culture and serum-free medium and provide 5-10 fold higher secreted expression than Sf9 cells. High Five cells are excellent for expressing recombinant proteins.

2.2. Vectors

2.2.1. Cloning in pET15 vector

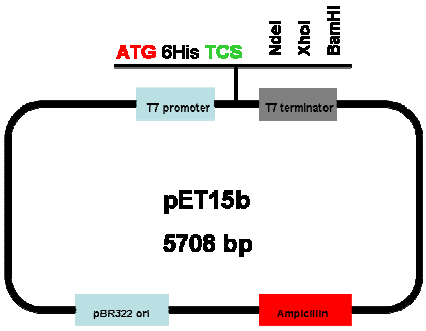


Fig.2.1. Schematic representation of pET15 plasmid.

The gene of interest is cloned in pET15 vector for heterologous expression in *E.coli*. This vector adds a 6Histag to the N-terminus of the protein and holds a T7 promoter and ampicillin resistance (see Fig.2.1).

2.2.2. Cloning in pDDGFP-2 vector

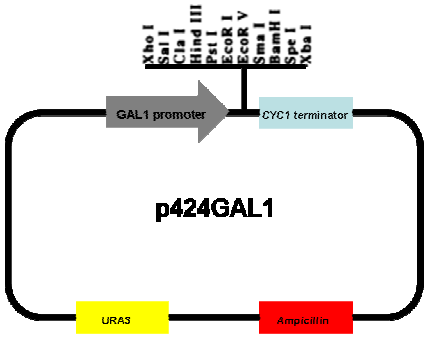


Fig.2.2. Schematic representation of p424GAL1 plasmid.

The genes of interest are cloned by homologue recombination in pDDGFP-2 vector for the heterologous expression in *Saccharomyces cerevisiae*. This plasmid harbors a *GAL1* promoter and URA selection marker; pDDGFP-2 vector derived from p424GAL1 vector (see Fig 2.2). It was modified by D. Drew (Newstead *et al*., 2007) by cloning the sequence for yEGFP with a C-terminal octa-His tag and

a TEV protease site. In addition, a *SmaI* site was inserted before the TEVp-GFP-8His sequence. Sequences of interest are amplified with primers that include 5' overhangs complementing the upstream and downstream sequences to either side of the SmaI site in the GFP-8His vector. Digestion of vector with SmaI gives rise to blunt ends that eliminate background colonies caused by re-annealing of the linearized vector during transformation by homologue recombination (see Fig.2.3).

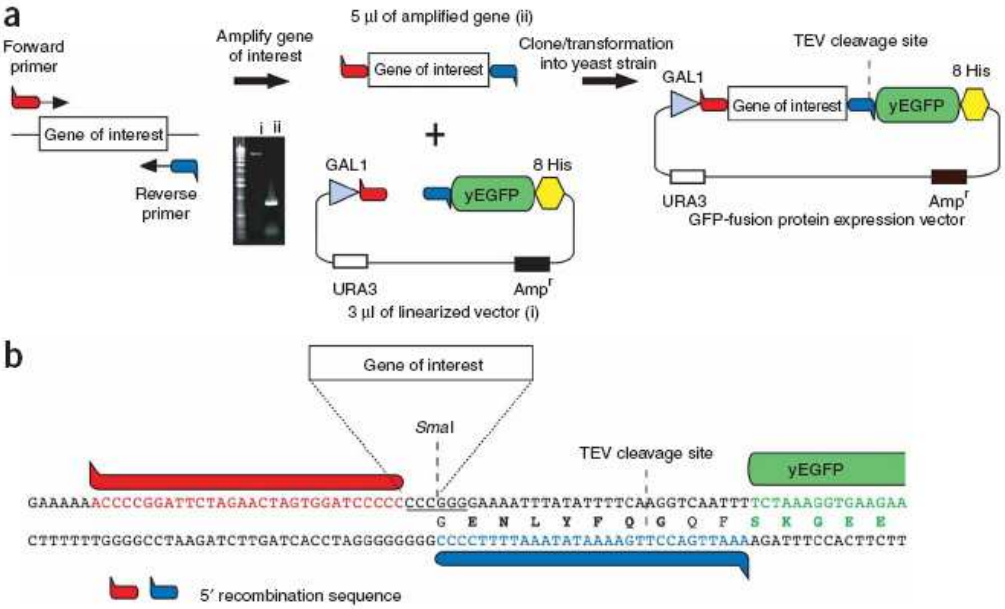


Fig.2.3. Cloning by homologous recombination into *S.cerevisiae* GFP-fusion vector. (a) The amount of PCR and vector used in cloning is depicted in the UV-exposed agarose-gel inset: i, SmaI linearilized vector; ii, amplified gene with overhangs required for homologous recombination. (b) Cloning site used GFP-fusion vector pDDGFP-2. TEV, tobacco etch virus; yEGFP, yeast-enhanced green fluorescent protein (Drew *et al*.,2008).

2.2.3. Cloning in pENTR-D-TOPO and BaculoDirect™Baculovirus Expression Systems

Cloning cDNA in this plamid (see Fig.2.4) allows to generate an entry clone with *att*L sites; in this case the gene of interest is then transferred into the BaculoDirect™ Linear DNA by using LR reaction for the heterologous expression in insect cells. The LR reaction facilitates recombination of an *att*L site

substrate (entry clone) with an *att*R substrate (BaculoDirect™ C-Term Linear DNA) to create an *att*B-containing expression virus.

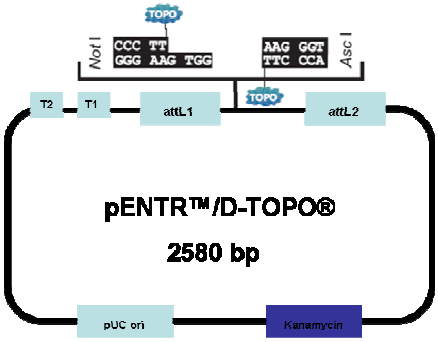


Fig.2.4. Schematic representation of pENTR™/D-TOPO® plasmid.

2.3. Experimental techniques

2.3.1. Plasmid DNA preparation

All plasmid purifications were prepared by using the GenElute™ Plasmid Miniprep Kit (Sigma).

2.3.2. Agarose gel electrophoresis

DNA samples were diluted in agarose gel loading buffer and resolved on agarose gel (0.8 % agarose, SYBR® SAFE DNA gel stain dissolved in TBE). Electrophoresis was performed at room temperature (100 volts, 1 hours) in the agarose gel running buffer.

2.3.3. Transformation of competent cells

Transformations of the gene-vector construct into chemically competent for cells of all bacterial strains were performed following a standard protocol.

For chemical transformation, cells are grown to mid-log phase, harvested and treated with divalent cations (in this case $CaCl_2$). Cells treated in such a way are said to be competent. To chemically transform competent cells, they are mixed

with the DNA, in ice, followed by a brief heat shock. Then, the cells are incubated with LB medium and allowed to express the antibiotic resistant gene for 30-60 minutes before plating.

Transformation of *S.cerevisiae* is similar to that applied to *E.coli*. Intact yeast cells, which have been harvested during the logarithmic phase of growth, are treated with a lithium acetate solution, followed by a polyalcohol treatment (in this case PEG 3350), that facilitates the DNA uptake. Addition of carrier DNA (single-stranded carrier DNA, salmon sperm) also promotes the uptake of vector DNA (Drew *et al*.,2008).

2.3.4. Insect cell transfection (BaculoDirect™Baculovirus Expression System, Invitrogen)

Cellfectin® Reagent is provided with the BaculoDirect™ Expression and Transfection Kits for lipid-mediated transfection of insect cells (Sf9 cell line) with LB recombination reaction. Cellfectin® Reagent is a 1:1.5 (M/M) liposome formulation of the cationic lipid N, NI, NII, NIII-Tetramethyl-N, NI, NII, NIII-tetrapalmitylspermine (TM-TPS) and dioleoyl phosphatidylethanolamine (DOPE) in membrane-filtered water. Budded virus should be released into the medium 96 hours after transfection. Once the cells appear infected, the virus has to be harvested from the cell culture medium (V0 viral stock). The V0 viral stock is a low-titer stock ($1x10^5$-$1x10^6$ pfu/ml). Cells were infected with the V0 stock to generate a high-titer V1 stock of approximately $5x10^7$-$1x10^8$ pfu/ml. This V1 viral stock can then be used to generate a large-scale, high-titer viral stock (V2 viral stock) suitable for expression studies.

2.3.5. Recombinant protein induction

All bacterial large volume cultures were performed in a 5 l Erlenmeyer flask in liquid LB medium supplemented with the appropriate antibiotics. An overnight culture was diluted to an OD_{600} of 0.15. Cells were grown at 37°C in an orbital shaker to an OD_{600} of 1.0. Protein expression was induced by adding 0.5 mM

IPTG and leaving the culture overnight, depending on the protein, at the optimal temperature for expression.

All yeast large volume cultures were performed in Tunair 2.5 l full baffled shaker flask (VWR) in liquid URA medium (with 0.1% glucose) diluting an overnight culture (liquid URA medium with 2% glucose) to an OD_{600} of 0.12. Cells were grown in an orbital shaker at 30°C; at an OD_{600} of 0.6 (after ~7 hours) expression of the GFP-fused protein was induced with 2% galactose for 22 hours.

To identify the best expression conditions for each recombinant protein, several small volume culture expression trials were performed in 50 ml tubes. After induction cells were harvested by centrifugation at 8.000g for 10 minutes at 4°C and the cell pellet was stored at -80°C.

All large suspension insect cell cultures were prepared in Erlenmeyer flasks after growing enough log phase adherent cells to start a culture of the desired size with $1x10^6$ cells/ml. Cells were infected using an appropriate volume of V2 viral stock. Maximum expression of proteins (non secreted) is generally observed between 48-96 hours post-infection.

For recombinant protein expression in insect cells small expression adherent cultures were prepared testing different volumes of V2 viral stock and different infection times. After infection cells were harvested by centrifugation at 5.000g for 5 minutes at 4°C and cell pellets were stored at -80°C. For the Sf9 cell line, the cell pellet was washed with PBS buffer to remove any serum protein contamination.

2.3.6. Screening of membrane protein overexpression using whole-cell and in gel-fluorescence

Cloning in pDDGFP-2 vector (see 2.4.2) allows to follow the protein expression measuring the GFP fluorescence. After induction, 200 µl of cell suspension can be transferred to a black Nunc 96-well optical bottom plate; GFP fluorescence emission was measured at 512 nm by excitation at 488 nm. The membrane protein overexpression level (in mg/l) was estimated as describe in Drew *et al*. (2008) was used. The cell suspension was then transferred in a 1.5 ml capped tube and disrupted by vortexing or using a mixer-mill disruptor set at 30 Hz with glass

beads. Unbroken cells were removed by centrifugation at 22000g for 5 seconds at 4°C. To pellet crude membranes, the supernatant was centrifuged at 22000g at 4°C for 1 hour. Crude membranes were resuspended and loaded on a SDS-PAGE gel to check for fluorescent bands (see 2.5.8.2).

2.3.7. SDS-PAGE for protein analysis

2.3.7.1. Standard SDS-PAGE

Protein samples were diluted in SDS-PAGE loading buffer, denatured at 95ºC for 5 min and resolved on denaturing polyacryamide gels. Electrophoresis was performed at room temperature (200 volts, 1 hours) in SDS-PAGE running buffer.

2.3.7.2. In gel-fluorescence

To visualize membrane proteins fused to GFP on a SDS-PAGE, a Novex 12% Tris-Gly gel (Invitrogen) was used. Protein samples were diluted with in gel-fluorescence loading buffer. After running (see 2.5.8.1), the gel was rinsed with H_2O and the fluorescent bands were detected with a CCD camera system. The gel was exposed to blue light (EPI source) set at 460 nm with a cut-off filter at 515 nm.

2.3.8. Western blotting

After SDS-PAGE proteins are transferred to a PVDF membrane (100V for 1hour on ice) where they are probed using Monoclonal Anti-polyHistidine-Peroxidase antibody produced in mouse (Sigma).

2.3.9. Recombinant proteins extraction

Bacterial cell pellets were resuspended in 25 ml of CRB(1) or CRB(2) per liter of original cell culture and sonicated on ice. Lysate were subsequently centrifuged at 8000g for 10 minutes at 4°C. Yeast cell pellets were resuspended in 25 ml of CRB(2) per liter of original cell culture and broken using a heavy-duty cell

disruptor with four passes at incremental pressures of 25, 30, 32 and 35 kpsi (1.7-2.4 x10^3 atm) at 4-15 °C. The lysate was subsequently centrifuged at 8000g for 10 minutes at 4°C. To remove the unbroken cells and debris the lysate was centrifuged at 10000g at 4 °C for 10 minutes. Insect cell pellets were resuspended in 1/10 of the original cell culture volume. Cell were disrupted using CBR(2) buffer with 1% detergent and leaving 1 hour at 4°C with mild agitation.
To purify soluble proteins the cleared supernatant was directly loaded onto an affinity column. For membrane proteins, cell membranes were collected from the cleared supernatant by centrifugation at 100000g at 4°C for 120 minutes.

2.3.9.1. Membrane protein solubilization

In this study detergent screening was carried out using N,N-Dimethyldodecylamine N-oxide (LDAO; Anatrace, cat. no. D360), n-Decyl-b-D-maltopyranoside (DM; Anatrace, cat. no. D322), n-Dodecyl-b-D-maltopyranoside (DDM; Anatrace, cat. no. D310), Fos-choline 12 (Anatrace, cat. no. F308) and Polyethylene(9)dodecyl ether ($C_{12}E_9$; Anatrace, cat. no. AP0129) (see Tab.2.1).

Detergent	Molecular weight	Aggregation number	CMC (%)
LDAO	229.41	76	0.023
DM	482.6	69	0.87
DDM	510.6	78-149	0.0087
Fos-choline 12	351.5	54	0.047
$C_{12}E_9$	avg. 583	-	0.003

Tab.2.1. Physical properties of commonly used detergents.

For membrane proteins the use of a cleavable green fluorescent protein (GFP) with a histag fused to the C-terminus of the protein has proved to be very effective as a way of following the protein during purification, a technique pioneered by Prof. Jan-Willem De Gier and Eric Gouaux. To be successful with this expression system, the target membrane protein must have a cytoplasmic C-terminus, since the GFP can only be correctly folded and become fluorescent in the cytoplasm. GFP fusion proteins expressed *S. cerevisiae* allow rapid selection of targets with the highest expression yields for large-scale purification. In-gel fluorescence

analysis and fluorescence size-exclusion chromatography (FSEC), GFP fusion proteins clearly show whether a protein is monodisperse in particular detergents without the need of any purification step.

2.3.10. Confocal microscopy (Leica TCS SP2 upright confocal microscope, Leica)

Harvested cells from a 10-ml culture overexpressing the membrane protein fused to GFP were resuspended in 1 ml of URA media containing 50% glycerol.
One microliter of the cell suspension was add to a microscope slide and a coverslip was placed on top. The sample was focused on using Köhler illumination with an X10 magnification lens and the focal plane was set to zero. A drop of lens oil was added changing to the higher magnification oil-immersion lens. The brightfield lamp was turned off. The blue light was then turned on, and the gross localization of the membrane protein-GFP fusion was estimated. The blue light was turned off and using the argon laser and exciting at 488 nm, a detailed localization image of the GFP-fusion was captured (emission detection at 505–535 nm).

2.3.11. Immobilized metal ion affinity chromatography (IMAC)

For both soluble and membrane proteins the first purification step was carried out using an appropriate quantity of Ni Sepharose 6 Fast Flow (GE Healthcare) packed into a plastic column. Cell lysates or solubilized membranes were slowly loaded onto the column. The resin was then washed with CRB(1) or EB with increasing concentration of imidazole. The protein of interest was eluted from the column with the same buffer containing 250 mM imidazole.

2.3.12. Size exclusion chromatography (SEC)

SEC was carried out using Superdex 200 10/300 GL Tricorn gel filtration column (GE Healthcare). The column was equilibrated in TSB or DB and operated at a flow rate of 0.5 ml/min connected to an ÅKTA FPLC system (GE Healthcare) .

2.3.13. Fluorescence size exclusion chromatography (FSEC)

Detergent-solubilized yeast membranes were subjected to fluorescence size exclusion chromatography (FSEC) to measure the 'monodispersity' of the sample. FSEC was performed injecting the detergent-solubilized sample onto a Superose 6 10/300 column (GE Healthcare) equilibrated with DB (0.03% DDM) at a flow rate of 0.4 ml/min. After a 6 ml flow, the 0.2 ml fractions were collected row-by-row into a 96-well plate. The plate was read by a plate reader (SpectraMax M2e microplate reader). To improve the signal-to-noise ratio GFP fluorescence emission was measured at 512 nm with excitation at 470 nm (this wavelength is used instead of 488 nm, as it produces a lower background fluorescent signal). GFP fluorescence in each well was plotted against the fraction number.

2.3.14. Hydrophobic interaction chromatography

Hydrophobic interaction chromatography was performed using a column packed with the hydrophobic resin Lipidex 1000 (PerkinElmer).

2.3.15. Mass spectrometry

Mass spectrometry was performed on an ion trap Esquire 6000 mass spectrometer (Bruker-Daltonik, Germany). For protein mass determination (see 3.1.4), spectra were acquired in a positive mode. For protein activity analysis (see 3.1.4.1), mass spectra were acquired in a negative ion mode over a mass scan range of *m*/*z* 50–1200 for 5 min at a scan rate of 13000 m/z per second. Capillary voltage was 4 kV and a dry gas flow rate of 6.5 L/min was used with a temperature of 230 °C.

2.3.16. NMR spectroscopy

One-dimensional (1D) NMR experiments were run on a Bruker 800 US^2 equipped with a cryoprobe. The experiment was carried out by collecting 512 scans. The spectra width was 13 ppm and the frequency offset was 4.7 ppm. The experimental temperature was set to 298 K.

2.3.17. Dynamic light scattering (DLS)

Measurements were carried out on a Zetasizer Nano (Malvern Instruments). The instrument uses light scattering techniques to measure hydrodynamic size, zeta potential and molecular weight of proteins. Typical applications of DLS are the measurement of the size and size distribution of particles, emulsions and molecules dispersed or dissolved in a solution. In this case it was used to determine the presence of aggregates before setting crystallisation screenings.

2.3.18. Protein crystallization

Several techniques exist for setting up crystallisation experiments, which include vapour diffusion (hanging and sitting drop), microbatch, microdialysis,…In this work all crystal screenings were performed manually or by the Mosquito Nanodrop Crystallisation robot using the vapour diffusion method.

2.3.18.1. Crystallization by the vapour diffusion method

In a vapour diffusion experiment, a small droplet (from 200 nl to several µl) of the protein solution is mixed with a precipitant solution (usually in a 1:1 ratio) and allowed to equilibrate in a sealed chamber against a reservoir containing a larger volume of the precipitant solution. Since the initial concentration of precipitants is lower in the droplet than in the reservoir, the droplet dries while gradually increasing supersaturation, where nucleation followed by crystal grows occurs.

For crystal sceening performed manually by the hanging drop method, BD Falcon™ 24-well Multiwell Plates (Becton Dikinson) were used whereas the new Innovadyne/Wilden plates (designed by Jan Lowe of the LMB in Cambridge) were adopted for the sitting drop method carried out by the robot.

2.3.18.2. Crystallization screening

In this study the following crystallization screens were used (see Tab.2.2).

Screen	Company
Mem-Start & Mem-Sys	Molecular Dimensions Ltd
MemGold	Molecular Dimensions Ltd
Structure screen (I + II)	Molecular Dimensions Ltd
PEGRx1	Hampton Research

Tab.2.2. Crystallization screenings.

Chapter 3

Results and Discussion

A.3.1. Bile Acid Coenzyme A:Amino Acid N-Acyltransferase

A.3.1.1. Generation of the human BAAT prokaryotic expression vector

For the expression of enzymatically active hBAAT in *E.coli*, the hBAAT cDNA was subcloned into the *NdeI-BamHI* sites of the bacterial expression vector pET15 (see paragraph 2.2.1). The PCR was performed using as template the plasmid pOTB7 containing the complete cDNA sequence (IMAGE ID 4071788, obtained from RZPD, Deutsches Ressourcenzentrum fuer Genomforschung GmbH). The gene was amplified by PCR using the following primers (the coding sequence is shown in bold):

BAAT-ET-for: 5'- GGTGGTCATATG**ATCCAGTTGACAGCTACCCCT** - 3'
BAAT-ET-rev: 5'- GGTGGTGGATCC**TTAGAGTTGACTGGTCACATC** - 3'

CATATG: *NdeI* restriction site
GGATCC: *BamHI* restriction site

Template DNA and primers were subjected to the program A (see appendix, paragraph 3). PCR reactions were extracted from agarose gel using the GenElute™Gel Extraction Kit (Sigma) and digested with *NdeI-BamHI*. The resultant PCR product was cloned into pET15 giving the expression vector hBAAT-pET15; XL1Blue cells were transformed using the ligation reaction and positive colonies were selected by colony PCR (see appendix, paragraph 3-program C). The resulting constructs were confirmed by restriction digestion and DNA sequencing.
At the same time the human BAAT sequence was also cloned into the bacterial vector pQE50 that harbors a C-terminal His-tag.

A.3.1.2. Bacterial expression of human BAAT

The constructs hBAAT-pET15 and hBAAT-pQE50 were transformed into several *E.coli* strains (for the transformation protocol see paragraph 2.3.3). Numerous small-scale expression trials were set up using different induction temperatures.

All cell strains transformed with hBAAT-pQE50 did not show the presence of soluble protein (data not shown). On the other hand, the presence of soluble BAAT protein was detected in the following strains: SG13009 (pREP4), Origami, ER2566, BL21 (DE3), Rosetta (DE3) and BL21 C41 (DE3)) transformed with hBAAT-pET15. The best overexpression level was demonstrated in BL21 C41 (DE3) strain.

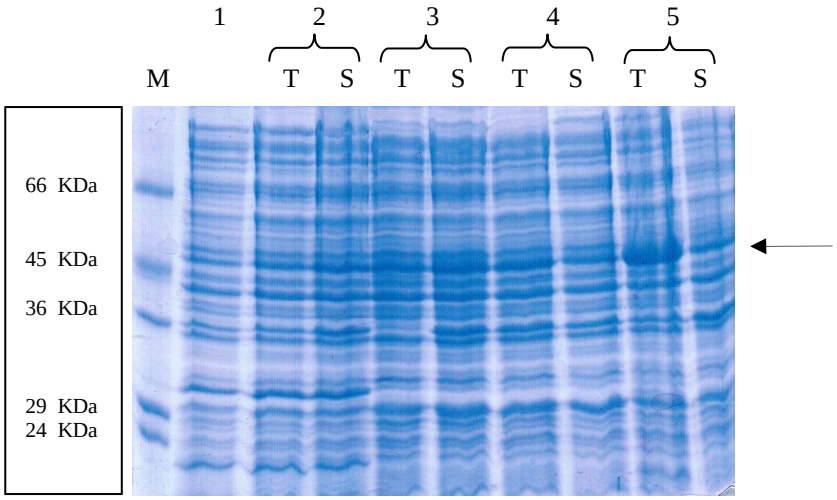


Fig. 3.1. 12% standard SDS-PAGE: expression trials of hBAAT in Bl21 C41 (DE3). Samples harvested 1) before induction, 2) after induction overnight at 16°C, 3) after induction overnight at 20°C, 4) after induction overnight at 28°C and 5) after induction overnight at 37°C. T: total cell fraction, S: soluble cell fraction, M: protein ladder.

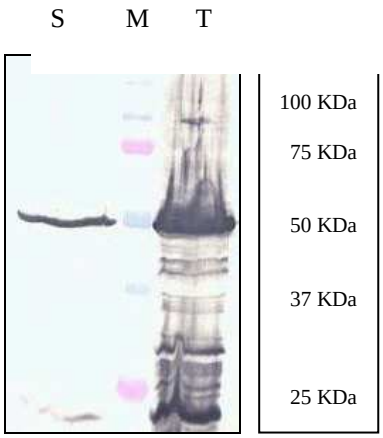


Fig.3.2. Western blot analysis: overexpression of hBAAT in Bl21 C41 (DE3). Samples were harvested after induction overnight at 20°C. T: total cell fraction, S: soluble cell fraction, M: protein ladder.

As shown in the Fig. 3.1 the most attractive condition was the induction with 0.5 mM IPTG at 20°C overnight. Protein expression was also confirmed by Western blot analysis using anti-histidine tag monoclonal antibodies (see Fig. 3.2).

A.3.1.3. Large-scale expression and purification of human BAAT

The selected condition was used to prepare a large-scale culture. Two liters of bacterial culture were prepared to purify the protein of interest. The cells were resuspended in a Tris-HCl buffer at pH 7.5 containing 0.5 M NaCl, 40 mM imidazole and 10 mM β-mercaptoethanol and sonicated. The soluble fraction was loaded onto a column packed with 2 ml of Nichel Sepharose 6 Fast Flow resin equilibrated with the same buffer. The column was then washed with 10 column volumes of CRB(1) with 40 mM imidazole. A second washing step was performed with 10 column volumes of CRB(1) with 80 mM imidazole. The protein was then eluted with 2-3 column volumes of CRB(1) with imidazole at the final concentration of 300 mM.

This protocol was established after a preliminary purification using a 5 ml column of Chelating Sepharose™ Fast Flow (Amersham Biosciences) connected to an ÅKTA FPLC system. Elution was performed with a 0-0.5 M imidazole gradient. Fig. 3.3 it is shows the absorbance profile at 280 nm.

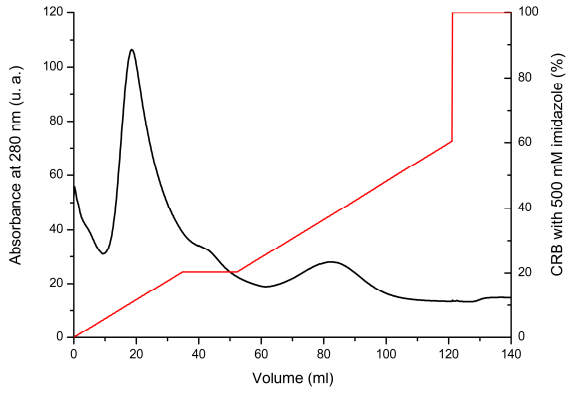


Fig. 3.3. Absorbance profile at 280 nm: purification of hBAAT by IMAC applying a manual modified gradient of imidazole. BAAT protein starts to be eluted with about 150 mM imidazole.

The eluted fractions were dialysed against CRB(1) without imidazole; the his-tag was removed by proteolytic cleavage by treating the partially purified protein with thrombin (1 U of enzyme per mg of protein at 20°C overnight) after the addition of 2.5 mM $CaCl_2$. The complete removal of the tag was assessed by SDS-PAGE as shown in Fig.3.4.

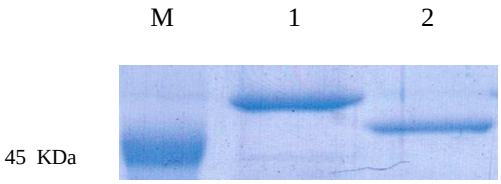


Fig. 3.4. 12% standard SDS-PAGE: 1) protein sample not subjected to the enzymatic cleavage, 2) protein sample harvested after digestion overnight at 20°C with thrombin, M: protein ladder.

A reverse IMAC was adopted to remove thrombin and some other contaminants still present after the first chromatographic step. The protein sample in CRB(1) without imidazole was loaded on a column filled with 2 ml of equilibrated resin Nichel Sepharose 6 Fast Flow; the column was then washed using CRB(1) with 10 mM imidazole, 40 mM imidazole, 80 mM imidazole and finally 300 mM imidazole to remove the bound material.

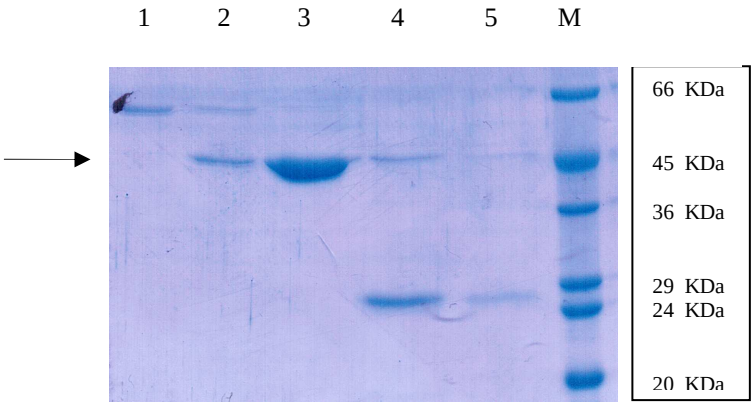


Fig.3.5. 12% standard SDS-PAGE: samples collected during a manual reverse IMAC. 1) Flow through with 0 mM imidazole, elution with 2) 10 mM imidazole, 3) 40 mM imidazole, 4) 80 mM imidazole and 5) 300 mM imidazole. M: protein ladder.

hBAAT was eluted mainly with 40 mM imidazole (see Fig. 3.5); the sample was concentrated using a centrifugal concentrator (cut off 10 KDa) to 0.5 ml and

subsequently injected onto a size exclusion Superdex 200 column pre-equilibrated with TBS without any reducing agent.

The Fig. 3.6 shows the SEC trace; the peak indicates that hBAAT is eluted as a monodisperse protein. Fractions were collected and analyzed by SDS-PAGE. The pure protein was concentrated to 10 mg/ml using a centrifugal concentrator (cut off 10 KDa).

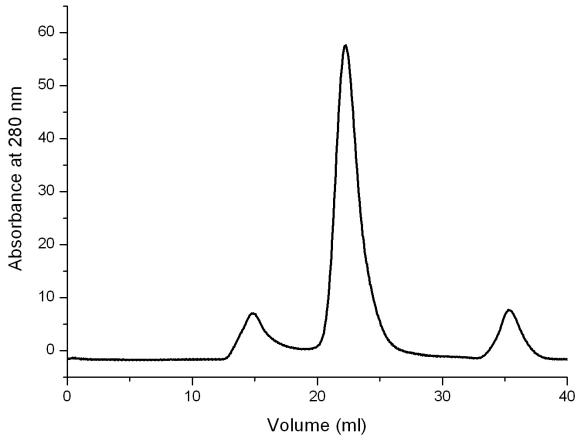


Fig.3.6. Absorbance profile at 280 nm: additional purification step of hBAAT by G200 chromatography.

The final yield was 1 mg of pure protein per liter of culture. To overcome the problem of such a poor yield, other heterologous expression systems were taken into consideration. Attempts to express human BAAT in insect cells and plants were not effective because the protein was present only in the insoluble fraction of cell lysate (data not shown).

A.3.1.4. Human BAAT sample control

To check the purity and the quality of the sample, 1 µl of concentrated BAAT was loaded on a SDS-PAGE and, at the same time, on a Tris-glycine native gel (see Fig. 3.8). The native gel shows the absence of aggregates on the prepared sample (Fig.3.8, right).

In addition, the presence of secondary and tertiary structures was confirmed by a NMR ^{1}H mono dimensional spectrum; in Fig. 3.9 the presence of peaks around 0 ppm can be noticed suggesting that the protein is folded.

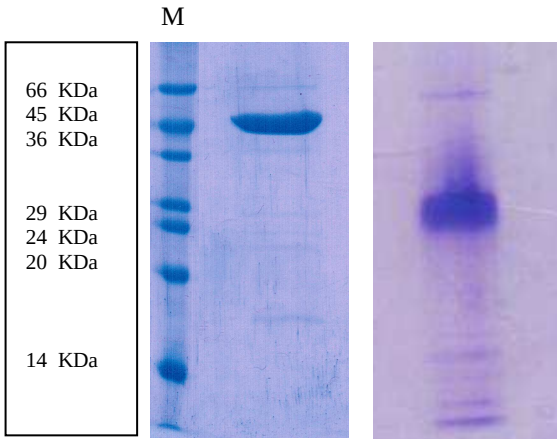


Fig.3.8. 15% standard SDS-PAGE (left) and native Novex 12% Tris-Gly gel (right): BAAT purified sample after concentration to 10 mg/ml. M: protein ladder.

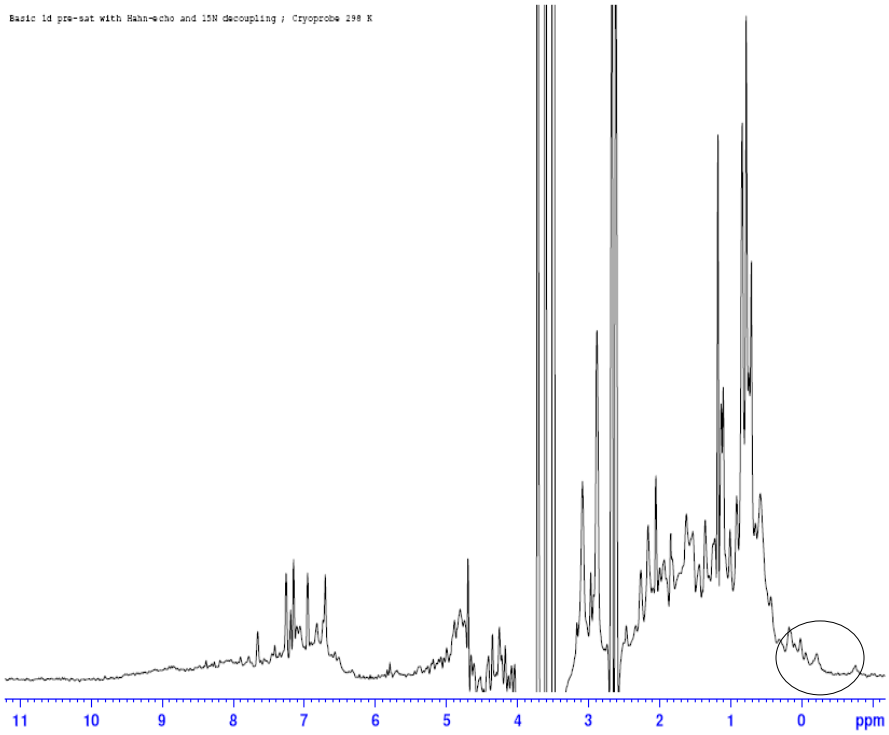


Fig.3.9. hBAAT NMR ^{1}H mono dimensional spectrum.

The exact protein mass was determined by electrospray mass spectrometry (see Fig. 3.10); the calculated mass agrees very well with the mass deduced from the cDNA sequence (46552 Da, see appendix, paragraph 5.1.3).

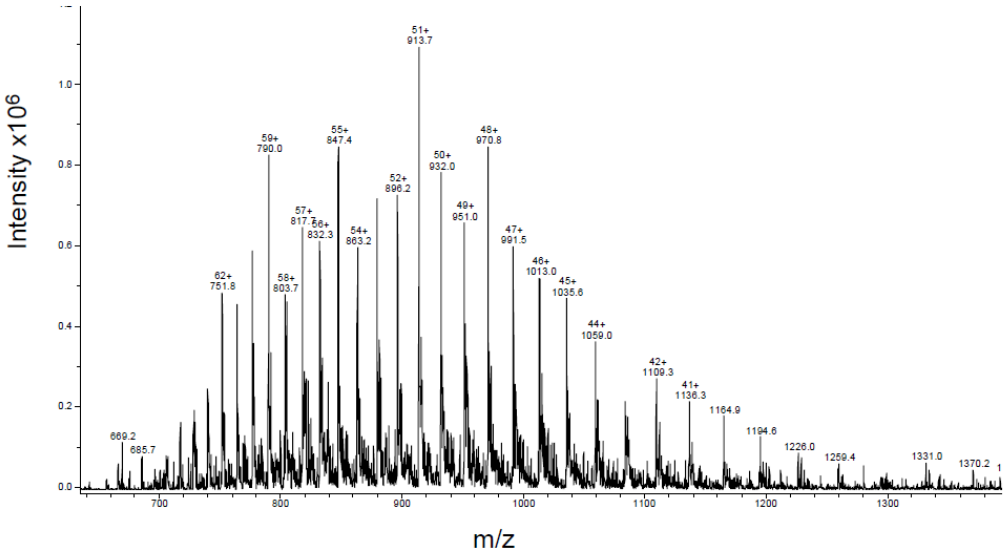


Fig.3.10. MS spectrum of recombinant BAAT.

A.3.1.4.1. Electrospray mass spectrometry assay of human BAAT activity

In addition to bile acids-CoA derivatives, hBAAT hydrolyzes also fatty acyl-CoAs and/or fatty acids conjugate to glycine (O'Byrne *et al*., 2003). Since the natural substrate of hBAAT is not commercially available and difficult to synthesize it was decided to verify the activity of recombinant hBAAT following the protocol proposed by O'Byrne *et al*., (2003). The reaction in this case was performed using stearoyl-CoA (Sigma), a CoA-derivative of stearic acid (a saturated fatty acid).

Incubation mixtures were set up containing stearoyl-CoA (100 µM), hBAAT (10 µg), and taurine (25 mM) in 50 mM potassium phosphate buffer, pH 8. The mixtures were incubated for 48 h at room temperature after which they were purified using columns packed with the Lipidex 1000 resin. The Lipidex columns were equilibrated with 5 ml volumes of chloroform:methanol (2:1), followed by 95% methanol in water, and finally by water. The incubation mixtures were then loaded onto the column followed by a further washing step with 5 ml of water. The column was then eluted using 5 ml of methanol, and the eluate was dried under nitrogen. The samples containing the stearoyl-CoA reaction products were reconstituted in tetrahydrofurane-water (1:1 v/v) and 4-5 µl were analyzed by mass spectrometry (see paragraph 2.3.15). Recombinant hBAAT was incubated

with fatty acyl-CoA in both the absence and presence of 25 mM taurine (Fig.3.11).

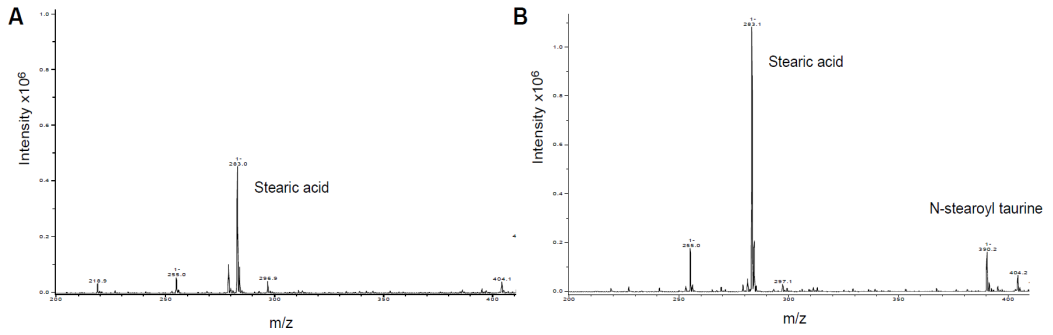


Fig.3.11. Negative ion electrospray mass spectrometry analysis of BAAT fatty acid conjugating activity. The spectrum in A) corresponds to the incubation containing 2-5 µg of BAAT and 40 µM stearoyl-CoA and the spectrum in B) corresponds to the incubation containing 2-5 µg of BAAT, 40 µM stearoyl-CoA and 50 mM taurine.

The enzyme shows both the N-acyltransferase activity (N-stearoyl taurine peak) and the thioesterase activity (stearic acid peak) due to its high degree of homology with the type I acyl-CoA thioesterases.

A.3.1.5. Crystallization of human BAAT

Several crystallization trials were set up at different protein concentrations (8 mg/ml, 10 mg/ml, 15 mg/ml, 20 mg/ml). The protein sample was prepared in the absence or in the presence of ligands. A batch of holo-enzyme was obtained adding 5mM cholate/5 mM glycine, 5 mM cholate/5 Mm taurine or 5 mM glycocholate. In fact, it was noticed that BAAT binds to a cholate and a glycocholate affinity column (Vessey, 1978). Although the recombinant hBAAT could be purified at an appreciable level (1 mg/L), up to now no crystals grew from the trials prepared.

A.3.1.6.Generation and expression of human BAAT mutants

To abolish thioesterase activity the catalytic mutant BAAT C235A was prepared. Furthermore, examination of the protein sequence shows also the presence of a

couple of cysteines in positions 372-373 (see appendix paragraph 5.1.3). Since cysteines could form non physiological inter- and/or intramolecular disulfide bridges and therefore be a source of sample microheterogenity, a second double mutant was prepared, C372-373A. A third triple mutant C235A-C372/3A with no cys residues was also expressed. The original pET15-BAAT plasmid was used as a template and four primers were ordered. The primer sequences were the following:

BAAT-C235A-for: 5'- GTAGTCTCTGTAGCTCAAGGAGTACAG - 3'
BAAT-C235A-rev: 5'- CTGTACTCCTTGAGCTACAGAGACTAC - 3'

BAAT-C372/3A-for: 5'- TATTCTCCTCTGGCGGCCGCCTCAACGACC - 3'
BAAT-C372/3A-rev: 5'- GGTCGTTGAGGCGGCCGCCAGAGGAGAATA - 3'

PCR was performed as described in appendix, paragraph 3 (program B); PCR products were then digested with the *DpnI* restriction enzyme to eliminate the original template. XL1Blue cells were transformed with the reaction and several colonies were sequenced. Positive mutant plasmids were used to transform BL21 (DE3) C41 cells.

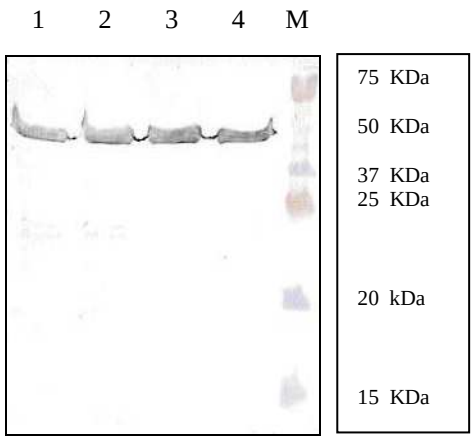


Fig.3.12. Western blot analysis: soluble cell fractions obtained from small-scale expression trials of the proteins 1) BAAT wild type, 2) BAAT catalytic mutant form, 3) C372/373A and 4) C235A-C372/373A. Cells were induced with 0.5 mM IPTG at 20°C overnight. M: protein ladder.

The mutants were expressed and purified adopting the same conditions used for the expression and purification of wild-type hBAAT (see Fig.3.12).

Fig.3.13 shows SEC traces of wild type BAAT and the three mutants. The purity of the four samples was comparable but the yield and the behavior were completely different. Yield of the double and the triple mutant was 0.5 mg/l or less; in particular the triple mutant did not seem to be monodisperse and it was not stable at high concentration.

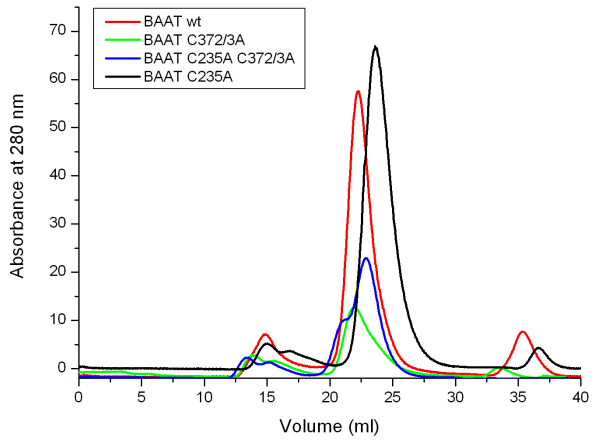
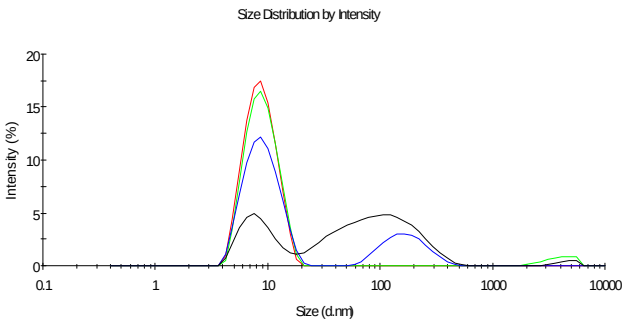
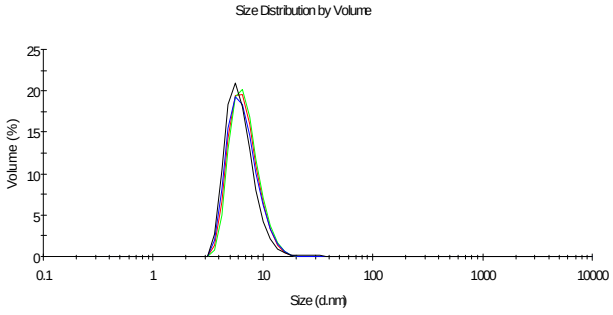


Fig.3.13. Absorbance profile at 280 nm: G200 chromatography showing the elution peaks of wild type BAAT and the three mutant forms.

A



B



C

Sample	Z-average (d.nm)	Pdl
BAAT wild type	8.09	0.088
BAAT C235A	8.75	0.204
BAAT C372/373A	10.90	0.357
BAAT C235A C372/373A	23.57	0.825

Fig.3.14. Dynamic light scattering experiments. Measurements were made on a Malvern Zeta sizer instrument at 20°C: A) intensity and B) volume size distributions obtained for the wild type enzyme and the three mutants, C) z-average diameters and polydispersity index (Pdl) of the samples. Red: BAAT wild type, green:BAAT C2372/373A, blue: BAAT C235A-C372/373A, black: BAAT C235A.

The purification protocol, the storage conditions, the choice of the buffers (pH and ionic strenght) and the modification on the amino acid chain (mutagenesis) may be crucial to obtain a crystallization quality sample. For this reason, the monodispersity of the samples was assessed using the DLS technique. Fig. 3.14 shows DLS data for the wild type enzyme and the three mutant forms. Samples were analyzed in TBS at the concentration of 3 mg/ml and measurements were conducted at 20 °C. The samples are clearly different. Visually, the peak shape of the double and triple mutant forms suggests the presence of oligomeric assemblies and/or aggregates, appreciable even if considering only the size distribution by intensity. The associated polydispersity index is greater than 0.2, indicating the presence of more than one species. These results preclude their use for crystallization trials. Only the catalytic mutant (C235A) appears to be adequate for crystallization experiments.

A.3.1.6.1. Crystallization of the catalytic mutant of BAAT C235A

Several crystallization trials were set up at different protein concentrations (10 mg/ml, 15 mg/ml). The protein sample was prepared in absence or in presence of different ligands. A batch of the holo-enzyme was obtained by adding 5 mM stearoyl-CoA. Micro crystals appeared only with the holo-enzyme in 0.1 M Hepes pH 7.5, 10% PEG 8000, 8% ethylene glycol as precipitant (Fig. 3.15).

Several optimization trials were set up to improve the quality of these microcrystals but up to now no suitable crystals for X-ray diffraction experiments were obtained.

Fig.3.15. BAAT C235A microcrystlas obtained with 0.1M Hepes pH 7.5, 10% PEG 8000, 8% ethyleneglycol as precipitants.

A.3.1. Discussion

No protocols for the expression and purification of recombinant human BAAT had been described in the literature. Here, we describe how the human enzyme BAAT was cloned and expressed in E.coli. The purification protocol includes three steps: IMAC, reverse IMAC and SEC. Sufficient pure material was obtained to set up several crystallization trials. Although the purity and the quality of the wild type protein samples were excellent, no crystal was obtained. One of the problems might be the presence of large mobile loops that could interfere with the formation of molecular contacts inside the crystal. In fact, the use of secondary structure prediction programs suggests that large portions on the N-terminal part of the protein are unstructured. It was then decided to create a new construct that lacks a portion of the amino acid sequence on the N-terminal part but maintains the integrity of the catalytic domain with the Cys-His-Asp catalytic triad in the C-terminal portion. The major problem was to identify stable deletion mutants without any information about the secondary structure of the protein. Some constructs were cloned but no soluble protein was obtained (data not shown). Recently, the crystallographic structure of human mitochondrial acyl-CoA thioesterase (ACOT2) was published (Mandel *et al*., 2009). The enzyme contains an α/β hydrolase catalytic domain in the C-terminal region, with a Ser-His-Asp catalytic triad. The 3D structure of ACOT2 contains two domains, N and C domains.

Fig.3.16. Human BAAT 3D model builds using the program HHpred and as a template the pdb file of human ACOT2.

The N domain contains a seven-stranded β-sandwich; the active site is located in a large pocket at the interface between the two domains. BAAT shares significant amino acid sequence conservation with this type of enzyme. Using the program HHpred (Homology detection & structure prediction by HMM-HMM comparison) and the ACOT2 3D structure as template, a model of BAAT 3D structure was built (see Fig.3.16). This model suggests a hypothetic site where to cut the protein sequence. Therefore a new construct was made starting from amino acid 135. At the moment preliminary expression and purification studies are in progress. Obtaining enough pure material of the catalytic domain could be useful to set up crystallization trials but also to conduct activity studies that could help in understanding the role of the two domains.

On the other hand, the generation and purification of the inactive C235A mutant allowed the preparation of reproducible crystals that did not grow in any of the crystallization trials carried out with the wild type protein. The ongoing crystallization experiments are focused on obtaining larger crystals suitable for X-ray diffraction studies.

B.3.1. Two membrane enzymes: Farnesyl Cysteine-Carboxyl Methyltransferase and Stearoyl-CoA 9-Desaturase

B.3.1.1. Farnesyl cysteine-carboxyl methyltransferase from *C.glabrata*

B.3.1.1.1. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *S.cerevisiae*

Obtaining pure membrane proteins in large quantities is fundamental for analyzing their biochemical and structural features. In this case a new high-throughput GFP-based method (Newstead *et al*., 2007; Drew *et al*., 2008) was used to select some promising targets among 96 *C.glabrata* membrane proteins.

B.3.1.1.1.1. Controlling membrane protein overexpression using whole-cell and in gel-fluorescence

96 cDNAs coding for membrane proteins from *C.glabrata* were cloned in the pDDGFP-2 vector by homologue recombination (see 2.2.3). PCR products and linearized vectors were transformed in FGY217 cells as described in paragraph 2.3.3. For each clone a 10 ml culture was prepared to check the expression level and to control the protein integrity (see 2.3.6 and 2.3.7.2). Expression levels of the well expressing targets are summarized in Tab.3.1.

Name/Gene	Protein	Molecular weight (kDa)	Position in 96 well plate	Fluorescence counts	Expression level (mg/L)
CAGL0K04367g-F	L-methionine transporter Mup1 (YGR055w)	63.1	A5	25172 25358 26610	1.99
CAGL0B02079g-F	Azole transporter AZR (YGR224w)	68.1	A8	49095 43876 49011	4.21
CAGL0M08272g-F	Polyamine transporter (YKL174c)	69.7	B1	32461 27303 24767	2.66
CAGL0H00847g-F	UDP-galactose transporter HUT1 (YPL244c)	40.4	B2	61314 60609 66892	3.33
CAGL0F02805 g-F	Farnesyl cysteine carboxyl-methyltransferase Ste14 (YDR410c)	29	C4	26885 30794 37964	1.06
CAGL0H09196 g-F	Nucleotide sugar transporter GOG5 (YGL225w)	35.8	C6	28232 41188	1.61
CAGL0K04609 g-F	H+/biotin symporter Alg2 (YGR065c)	71.7	C7	16523 23655	1.38

CAGL0G02453 g-F	Peptide transporter PTR2 (YKR093w)	67.8	D1	22062 24861 38666	1.97
CAGL0C02123 g-F	Telomere maintenance ERY5 (YFR041c)	34.4	E11	39427 54928 18816	2.16
CAGL0G04851 g-F	Fatty acid elongase activity SUR4 (YLR372w)	41.2	F7	15873 15830	0.75
CAGL0J04378 g-F	Oligosaccharide transporter RFT1 (YBL020w)	63.1	H8	20216 31455	1.54
CAGL0M00748 g-F	Cell organision and biogenesis ECM7 (YLR443w)	49.6	H10	23488 42411	1.45

Tab.3.1. *C. glabrata* well expressing targets.

Crude membranes extracted from each of the 10 ml cultures were loaded for in-gel fluorescence detection (see Fig.3.16) to check if a visible band was present.

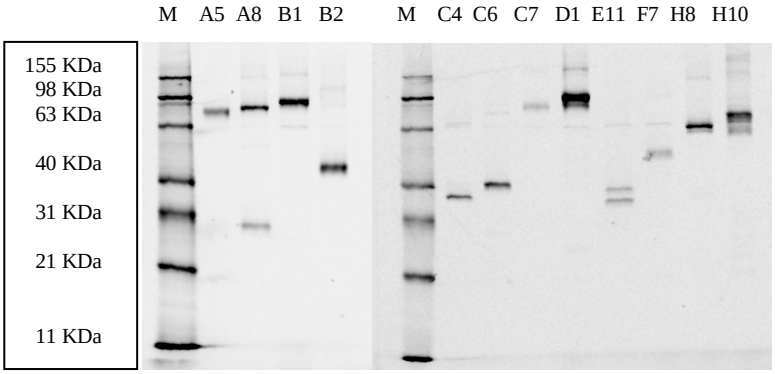


Fig.3.16. In-gel fluorescence detection of crude yeast membrane samples isolated after the overexpression of 12 different membrane proteins from *C.glabrata*. M: fluorescent proteins ladder.

B.3.1.1.1.2. Optimizing expression with temperature and chemical chaperones

In the literature, it has been reported that the addition of chemical chaperones such as DMSO (2.5% vol/vol) or histidine (0.04%) to yeast cultures can improve membrane protein folding. The effectiveness of these chemical chaperones was tested in the expression of these proteins. In Tab.3.2 the expression levels of the well expression targets were reported. Because a lower temperature can also increase the yield, expression at 20°C for 40 h after GAL1 induction was also tested (see Tab.3.3).

Clone number	Expression level (mg/L)		
	without additive	with DMSO 2.5%	with L-HIS 0.04%
A5	1.8	1.9	3.1
B1	2.3	1.5	3.3
B2	4.4	5.3	4.4
C7	1.9	1.8	2.2
F7	0.8	0.7	1.2
H8	2.5	2.4	2.8
A8	3.5	2.5	5.3
C4	1.2	1.3	2
C6	1.9	2	2.3
D1	1.5	1.2	1.2
E11	3.6	4.4	3.6
H10	2.6	2.9	2.8

Tab.3.2. Expression levels of the 12 well expressing targets after 22 hours of induction at 30°C with or without additives.

Clone number	Expression level (mg/L)		
	without additive	with DMSO 2.5 %	with L-HIS 0.04%
A5	1.6	1.2	0.4
B1	1.2	0.5	0.7
B2	1.7	1.5	0.6
C7	0.3	0.4	0.2
F7	0.1	0.1	0.1
H8	0.8	0.7	0.3
A8	0.7	2.1	0.8
C4	1.2	1.2	1.3
C6	1.4	2.6	1.2
D1	0.8	0.8	0.4
E11	3.2	1.8	3.7
H10	0.9	0.5	0.6

Tab.3.3. Expression levels of the 12 well expressing targets after 40 hours of induction at 20°C with or without additives.

B.3.1.1.1.3. Assessing the quality of overproduced fusions by FSEC

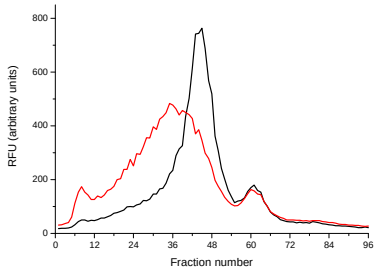
The best overexpression conditions were chosen for the isolation of the membranes from 1 liter of yeast culture. Membrane pellets from 1 liter were resuspended in 6 ml of MRB. Membrane suspensions were used for the detergent screen (see 2.3.9.1). Initially, solubilization efficiency was tested adjusting membrane resuspention to a protein concentration of 3.5 mg/ml in 1 % of the detergent of choice, PBS 1X and glycerol 10%. The mixture was incubated at 4°C for 1 hour with mild agitation; GFP fluorescence measure was made before and after centrifugation at maximum speed for one hour (see Fig.3.17, left).

FSEC was used to determine whether the detergent extraction efficiency can be related to the protein stability (see 2.3.13). Recently, Kawate and Gouaux (Kawate and Gouaux, 2006) showed that detergent-solubilized fusion proteins that are monodisperse, as judged by FSEC, are typically stable after purification. This

correlation removes the need to test each protein-detergent sample separately by purification. DDM and LDAO solubilized samples of each fusion construct (500 µl) were injected onto a size-exclusion column. The criterion for 'monodispersity' was that the fusion peak was expected to be symmetric and equal to or larger than either the aggregate or free-GFP peaks (Alguel *et al*,.2010).

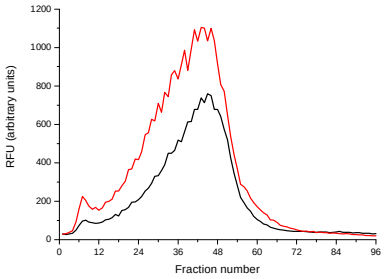
A5: L-methionine transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	5503	4624	84
DM	5810	4281	73.6
LDAO	6067	4500	74
Fos-choline 12	5603	5396	96
$C_{12}E_9$	5684	4055	71



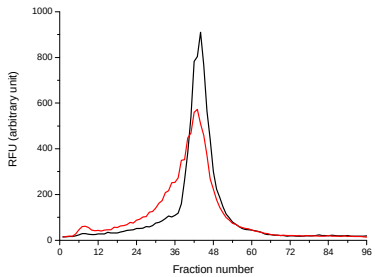
A8: Azole transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	9294	7144	77
DM	9178	6419	70
LDAO	9563	8010	83
Fos-choline 12	9388	9158	97
$C_{12}E_9$	9508	7182	75



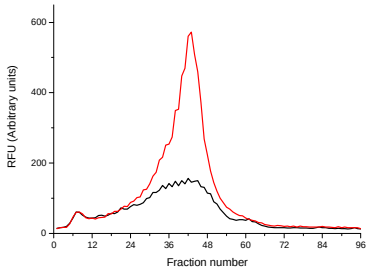
B1: Polyamine transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	3143	2271	72
DM	3016	2196	72.8
LDAO	3158	2382	75.4
Fos-choline 12	3021	2486	82
$C_{12}E_9$	3023	2074	68.6



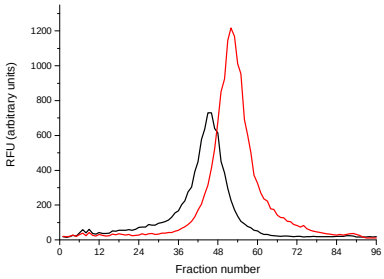
B2: UDP-galactose transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspension	After centrifugation	
DDM	6958	3522	50
DM	7148	2593	36.2
LDAO	6953	3253	46.7
Fos-choline 12	6248	6200	100
$C_{12}E_9$	6772	2550	37.6



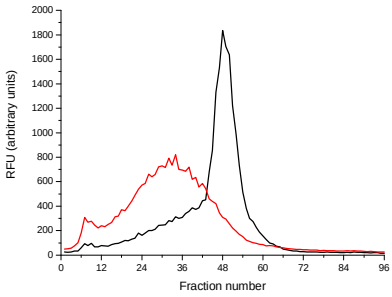
C4: Farnesyl cysteine carboxyl-methyltransferase

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	5330	3897	73
DM	5478	3495	63.8
LDAO	5389	4670	86.6
Fos-choline 12	5326	4884	91.7
$C_{12}E_9$	5299	3306	62



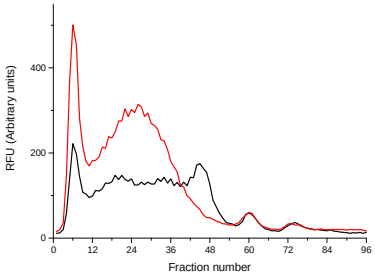
C6: Nucleotide sugar transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	10228	7754	75.8
DM	11276	6349	56.3
LDAO	11007	7628	69.3
Fos-choline 12	10746	9891	92
$C_{12}E_9$	10472	5488	52.4



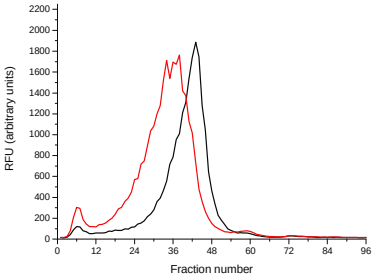
C7: H^+/biotin symporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	6738	3484	51
DM	6876	1845	26
LDAO	6986	3462	49
Fos-choline 12	6680	5257	78
$C_{12}E_9$	7048	4361	61



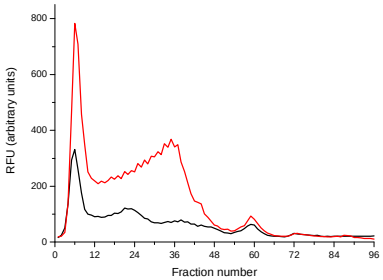
D1: Peptide transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	8857	5570	62.8
DM	8624	5435	63
LDAO	8249	5869	71
Fos-choline 12	8258	5919	71.6
$C_{12}E$	8384	5188	61.8



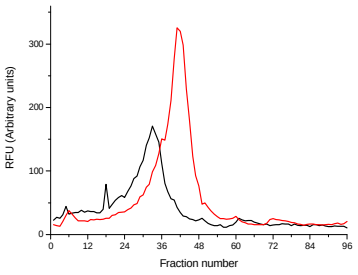
E11: Protein involved in telomere maintenance

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	10940	2462	22
DM	11359	2619	22.5
LDAO	15370	4348	28
Fos-choline 12	10632	8813	82
$C_{12}E_9$	11038	1927	17



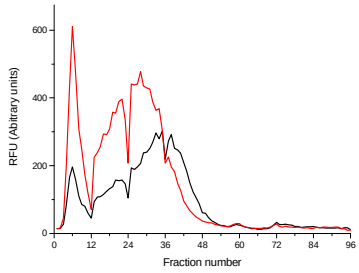
F7: Protein with fatty acid elongase activity

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	2775	1742	62.7
DM	2609	1471	56
LDAO	2413	1742	72
Fos-choline 12	2565	2016	78.5
$C_{12}E_9$	2433	1695	69.6



H8: Oligosaccharide transporter

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	6712	4319	64
DM	7212	3566	49.4
LDAO	6601	4259	64.5
Fos-choline 12	6326	5827	92
$C_{12}E_9$	7427	3836	51.6



H10: Protein involved in cell organision and biogenesis

Detergents (1%)	Fluorescence counts		Percentage (%)
	Membrane resuspention	After centrifugation	
DDM	8784	2075	23.6
DM	8395	1790	21
LDAO	8336	2882	34.5
Fos-choline 12	7888	6480	82
$C_{12}E_9$	8902	2304	25.8

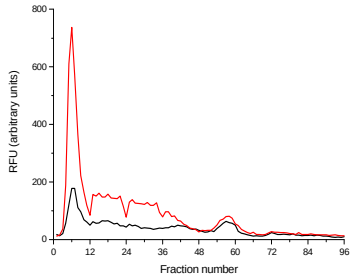


Fig.3.17. (left) Fluorescence counts before and after centrifugation of solubilized membranes obtained from 1 liter of yeast culture overexpressing the selected targets, (right) FSEC profile for DDM-(black) and LDAO-(red) solubilized membranes.

Only two or three out of the tested membrane proteins seemed to be promising for a successful overproduction. Considering the expression level, the stability after solubilization and the medical relevance it was decided to scale up the purification of the GFP-fusion C4, Farnesyl cysteine-carboxyl methyltransferase (Ste14), a transmembrane enzyme of 29 kDa. Because of its small size, well-characterized enzymatic activity, highly hydrophobic nature and pharmaceutical interest, Ste14 may represent an ideal ‘test case’ for developing methodologies to crystallize multispanning integral membrane enzymes, a critically neglected areas of current biology.

B.3.1.1.1.4. Confirming quality of membrane-integrated expression by confocal microscopy

S.cerevisiae has a highly regulated quality control system in the endoplasmic reticulum (ER), ER-associated degradation, and only correctly folded proteins exit the ER. Misfolded proteins are either retained in the ER until they are folded correctly or dislocated from the ER and degraded by the proteasome in the cytosol. Proper subcellular targeting of overexpressed membrane proteins therefore provides a good indication of correct folding. Global fluorescence-microscopy localization studies on C-terminally GFP tagged proteins have shown that for the majority of proteins, GFP does not interfere with trafficking (Newstead *et al*., 2007).

Harvested cells from 10-ml culture overexpressing Ste14-GFP, were tested by confocal microscopy (see paragraph 2.3.10). Fig.3.18 shows Ste14 GFP-fusion localized in membrane organelles, probably endoplasmic reticulum.

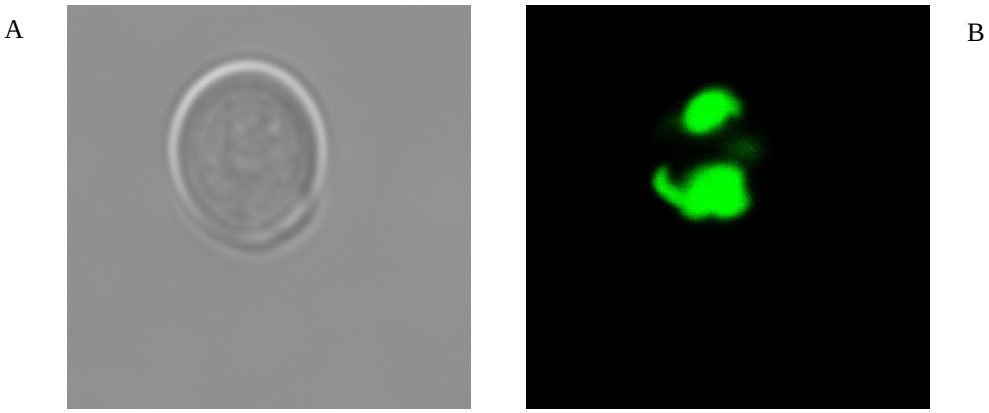


Fig.3.18. Ste14 GFP-fusion localization in *S. cerevisiae* monitored by confocal microscopy. A) Widefield and B) confocal image of an yeast cell.

B.3.1.1.2. Large-scale expression and purification of the Ste14-GFP fusion protein

A 10-liter culture was prepared as described in 2.5.6; induction was conducted using the parameters established in the overexpression optimization screen (at 30°C for 22 hours in the presence of galactose 2% and L-His 0.04%).

Membranes were isolated (see 2.3.9) and resuspended in 60 ml of MRB. The membrane resuspension was diluted at a final protein concentration of 3 mg/ml using EB. The detergent powder (1%) that produced the most homogeneous sample was added. In this case two different purifications were done, the first using DDM and the second using LDAO. The purification scheme shown in Fig.3.19 was followed.

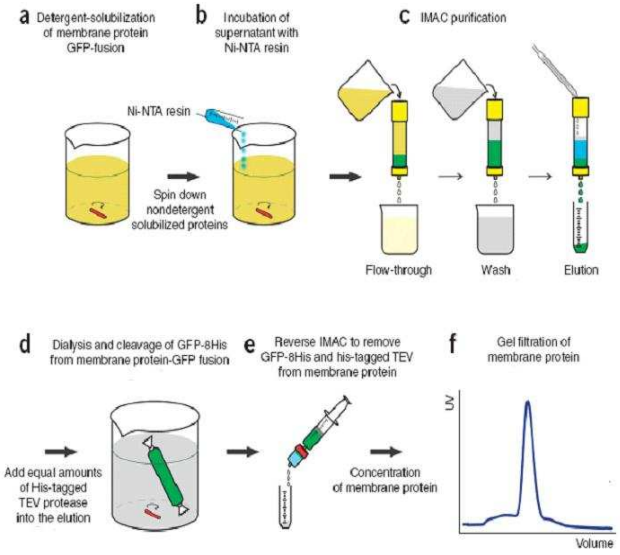


Fig.3.19. Flowchart illustrating the purification of Ste 14 from GFP-fusion. (a) Detergent solubilization of Ste14-GFP fusion, (b) incubation of supernatant with Ni Sepharose resin, (c) IMAC purification, (d) dialysis and cleavage of GFP from Ste14-GFP fusion, (e) reverse IMAC to remove GFP and tobacco etch virus (TEV) from Ste14, (f) gel filtration (Drew *et al.*, 2008).

The mixture was incubated for one hour at 4°C with mild agitation and then centrifuged at 100.000g for 45 minutes to pellet the insoluble material. 1 ml of Ni Sepharose 6 Fast Flow per 1 mg of GFP-fusion was added to the supernatant. The solution was mixed using a magnetic stirrer at 4°C for 3 hours. Ste14-GFP fusion was purified by IMAC: the supernatant was loaded in the presence of 40 mM imidazole, the column was washed with EB in the presence of 80 mM imidazole and the GFP-fused protein eluted in the presence of 250 mM imidazole. A reverse IMAC in the presence of 40 mM imidazole was then performed after cleavage of the GFP-His8 tag (during overnight dialysis to remove imidazole) with his-tagged tobacco etch virus (TEV) protease. The eluted protein was concentrated using a

centrifugal concentrator (cut off 30KDa) and then loaded on a Superdex 200 10/300 GL Tricorn gel filtration column equilibrated in DB. Each purification step could be followed by measuring GFP counts. It was also possible to estimate the membrane protein quantity using the methodology detailed in Drew *et al*. (2008). All chromathographic steps were performed in the presence of a reducing agent (10 mM β-mercaptoethanol during IMAC and 1 mM DTT during SEC); preliminary purifications suggested that the protein dimerized in the absence of a reducing agent (see Fig.3.20).

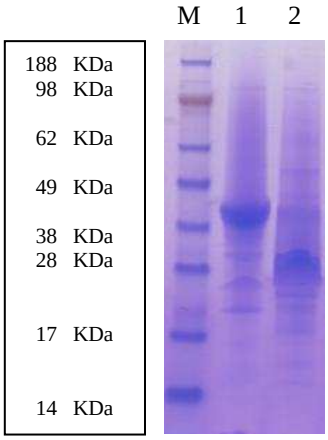


Fig.3.20. Nu PAGE gel (Invitrogen): Ste 14 protein 2) in the presence and 1) in the absence of 1 mM DTT.

B.3.1.1.2.1. Purification of Ste14 in DDM

Tab. 3.4. summarizes the information recorded throughout the purification steps of Ste14 in DDM (0.03%).

Membrane preparation	
Cells breakage efficiency (%)	67
Membrane protein in membrane resuspention (mg)	6
Purification	
Solubility (%)	75
Binding (%)	80
Amount of GFP-fusion in elution (mg)	3.6

Tab.3.4. Data-sheet showing information recorded throughout this protocol.

Fig.3.21 shows the SEC profile; the membrane protein peak of the UV-trace was collected and controlled by SDS-PAGE.

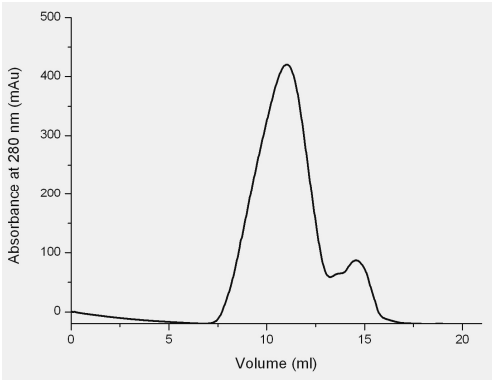


Fig.3.21. Absorbance profile at 280 nm: G200 chromatography of Ste 14 in DDM 0.03%.

The protein sample was then concentrated to 10 mg/ml using a centrifugal concentrator (cut off 30KDa). To check the purity and the quality of the sample, 1 µl was loaded on a SDS-PAGE (see Fig.3.22).

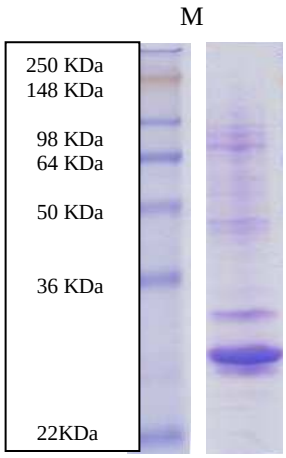


Fig.3.22. 12% standard SDS-PAGE: Ste14 purified sample after concentration to 10 mg/ml. M: protein ladder.

B.3.1.1.2.2. Purification of Ste14 in LDAO

Tab.3.4 reports the information recorded throughout the purification of Ste14 in LDAO (0.06%).

Membrane preparation	
Cells breakage efficiency (%)	70
Membrane protein in membrane resuspention (mg)	6
Purification	
Solubility (%)	95
Binding (%)	70%
Amount of GFP-fusion in elution (mg)	4

Tab.3.5. Data-sheet showing information recorded throughout this protocol.

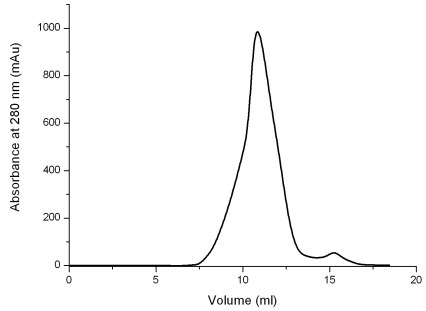


Fig.3.23. Absorbance profile at 280 nm: G200 chromatography of Ste 14 in LDAO 0.06%.

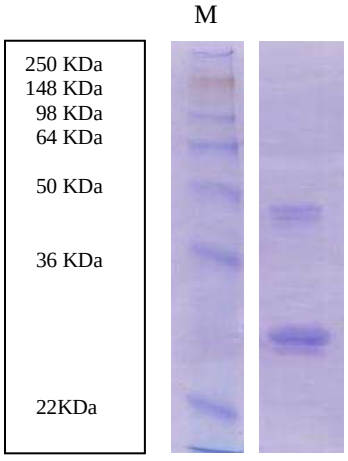


Fig.3.24. 12% standard SDS-PAGE: Ste14 purified sample after concentration to 10 mg/ml. M: protein ladder.

Fig.3.23 shows the SEC profile; the UV-trace pack was collected and analysed by SDS-PAGE. The protein sample was then concentrated to 10 mg/ml using a centrifugal concentrator (cut off 30KDa). To check the purity and the quality of the sample, 1 µl was loaded on a SDS-PAGE (see Fig.3.24).

B.3.1.1.3. Crystallization trials of Ste14

Several crystal trials were set up at various concentrations of the protein (10 mg/ml, 15 mg/ml, 20 mg/ml) both with DDM- and LDAO-solubilized Ste14. Although the recombinant Ste14 could be purified at an appreciable level, up to now no crystals grew from the trials prepared.

B.3.1.2. Farnesyl cysteine-carboxyl methyltransferase from *A. thaliana*

Since no appreciable result was obtained with these crystallization trials, it was decided to express the Ste14 homologue from *A.thaliana* (*At*Ste14A) (for a brief description see B.1.2.1.2.).
The *At*Ste14 amino acid sequence shows 36% identity with the yeast Ste14 and 45% identity with the human Icmt (see Appendix 2.7). The *At*Ste14 topology model is similar to the topology model of the yeast protein: they have both six membrane spans. However, the protein from *A.thaliana* lacks an N-terminal cytoplasmic portion of about 50 amino acids. Due to its small size and the absence of disordered portions, this protein could be a better target for crystallization experiments.

B.3.1.2.1. Cloning, expression and solubilization attempts

The Ste14A cDNA from *A.thaliana* was subcloned into the *NdeI-BamHI* sites of the bacterial expression vector pET15 (see paragraph 2.2.1). The PCR was performed using as template the plasmid pGEM-T Easy (Promega) containing the complete cDNA sequence. The *At*Ste14 cDNA sequence was amplified from the total *A.thaliana* cDNA (kindly provided by Prof. Furini, University of Verona), inserted in pGEM-T Easy and then sequenced to confirm the sequence identity.
In order to insert the gene in pET15, it was amplified by PCR using the following primers (the coding sequence is shown in bold):

STEara-ET-for: 5'- GGTCATATG**ACAGAGATCTTCAGTGACACC** - 3'

STEara-ET-rev: 5'- GGTGGATCC**TTAGTTCACAAATGGAACACCAGA**-3'

CATATG: *NdeI* restriction site

GGATCC: *BamHI* restriction site

Template DNA and primers were subjected to program A (see appendix paragraph 3). PCR reactions were extracted from agarose gels using the GenElute™Gel Extraction Kit (Sigma) and digested with *NdeI-BamHI*. The resultant PCR product was cloned into pET15 giving the expression vector Ste14ara-pET15; XL1Blue cells were transformed using the ligation reaction and positive colonies were selected by colony PCR (see appendix paragraph 3, program C). The constructs obtained were confirmed by restriction digestion and DNA sequencing. The Ste14ara-pET15 expression vector was transformed into *E.coli* BL21 CodonPlus (DE3)-RIPL cells. Several expression trials were carried out at different temperatures in order to improve the expression level (see 2.3.5). The presence of Ste14 protein was confirmed by Western blot analysis using anti-histidine tag monoclonal antibodies (see Fig.3.25).

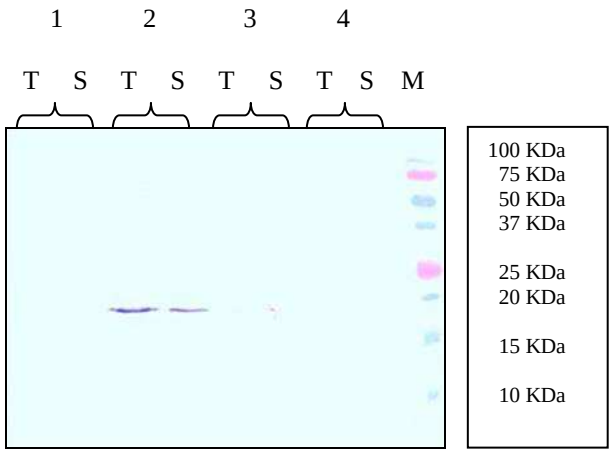


Fig.3.25. Western blot analysis: overexpression of *At*Ste14 in Bl21 CodonPlus(DE3)-RIPL. Samples harvested 1) after induction overnight at 16°C, 2) after induction overnight at 20°C, 3) after induction overnight at 28°C and 4) after induction at 37°C for 3 hours. T: total cell fraction, S: soluble cell fraction, M: protein ladder.

The best overexpression condition was chosen (induction at 20°C overnight with 0.5 mM IPTG) for the isolation of the membranes from 1 liter of culture (see 2.3.9). Membranes were resuspended in 2 ml of MRB and solubilized in 1% DDM and 1% LDAO like described in 3.1.1.2. Ten microliters of the solubilized

material before and after ultracentifugation were loaded on an SDS-PAGE and Ste14 was visualized by immunodetection using anti-histidine tag monoclonal antibodies (see Fig.3.26). Almost the total amount of Ste14 was solubilized in LDAO as demonstrated in Fig.3.26; no other detergent is effective in solubilizing this protein (data not shown).

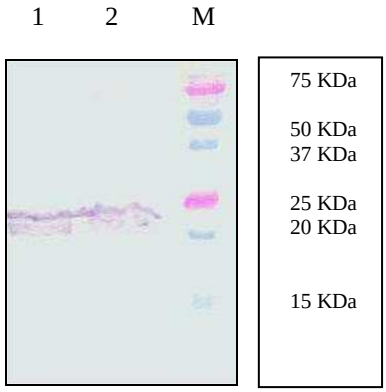


Fig.3.26. Western blot analysis: solubilized material in LDAO 1% harvested 1)before and 2) after centrifugation. M:protein ladder.

B.3.1.2.2. Large-scale purification of *At*Ste14

The selected condition was applied in a large-scale expression experiment. Four liters of bacterial culture were prepared to purify the protein of interest.

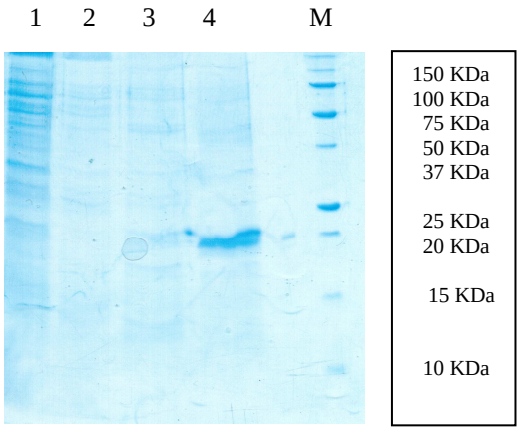


Fig.3.27. 15% standard SDS-PAGE: samples collected during IMAC. 1) Flow through with 0 mM imidazole, washing with 2) 10 mM imidazole, 3) 40 mM imidazole and 4) 300 mM imidazole. M: protein ladder.

Membranes were resuspended in 8 ml of MRB and then diluted in EB in the presence of 1% LDAO. Detergent-solubilized Ste14 was purified to homogeneity using IMAC; the sample was loaded on a column filled with 1 ml of the resin Ni Sepharose 6 Fast Flow in the presence of 10 mM imidazole. The column was then washed with EB with the addition of 40 mM imidazole and the protein eluted in 300 mM imidazole (see Fig.3.27)..

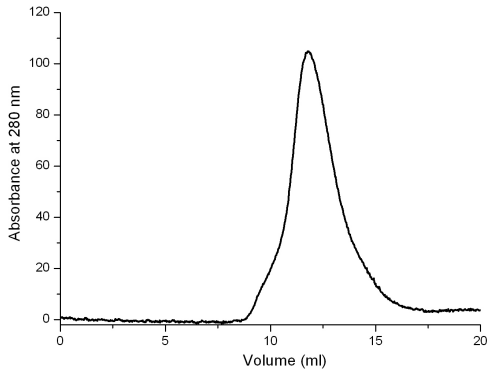


Fig.3.28. Absorbance profile at 280 nm: G200 chromatography of AtSte 14 in LDAO 0.06%.

The his-tag was not removed and the sample was concentrated using a centrifugal concentrator (cut off 10 KDa) to be injected onto a Superdex 200 pre-equilibrated with DB. The SEC profile shown in Fig.3.28 suggests that the protein is stable and monodisperse in LDAO; the aggregate peak is absent and the protein peak is symmetric.

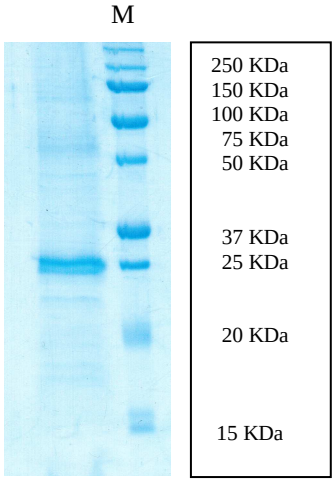


Fig.3.29. 15% standard SDS-PAGE: *At*Ste14 purified sample after concentration to 8 mg/ml. M: protein ladder.

The fractions containing Ste14 protein were collected and then concentrated to 8 mg/ml using a centrifugal concentrator (cut off 30 KDa). To check the purity and the quality of the sample, 1 µl was loaded on a SDS-PAGE (see Fig.3.29). The protein sample seemed to be sufficiently pure to proceed with structural analysis.

B.3.1.2.3. Crystallization trials of *At*Ste14

Several crystallization trials were prepared at the concentration of 8 mg/ml in presence or in absence of ligands. Two batches of the holo-enzyme were obtained by adding either 2 mM S-(5′-adenosyl)-L-methionine chloride (SIGMA) or 2 mM N-acetyl-S-farnesyl-L-cysteine (Cayman chemical), a selective inhibitor of isoprenylated protein methyltransferases. Up to now no crystals were obtained in these trials.

B.3.1.3. Stearoyl-CoA 9-desaturase

B.3.1.3.1. Cloning and expression of human SCD in SF9 insect cells

Due to the problems encountered in the large quantity expression of full-lenght hSCD in several prokaryotic and eukaryotic hosts (*E.coli*, *P.pastoris*, *S.cerevisiae*), the BaculoDirect™Baculovirus Expression System was utilized to obtain hSCD protein from Sf9 insect cells.

According to the published hSCD sequence (the complete cDNA sequence, IMAGE ID 3844850, was obtained from RZPD, Deutsches Ressourcenzentrum fuer Genomforschung GmbH), primers were designed to amplify full-length SCD cDNA using as a template the plasmid pOTB7.

The following forward primer incorporates the recombination sequence (in green) and the Kozak sequence (in red); the coding sequence is shown in bold:

SCDbac-for: 5’- CACCGGTACCGAA**ATGCCGGCCCACTTGCTGCAG** - 3’

GGTACC: *KpnI* restriction site

The following reverse primer adds the thrombin cleavage site (in pink). This could be used to remove the his-tag that is introduced on the C-terminal cloning into the linear baculovirus DNA.

SCDbac-rev: 5'- AATAATAAGCTTGGAACCACGCGGAACCAG**GTTTCCA TCTCCGGTTCT** - 3'

AAGCTT: *HindIII* restriction site

Template DNA and primers were subjected to program A (see appendix paragraph 3) using PFU enzyme. PCR reactions were extracted from agarose gels using the GenElute™Gel Extraction Kit (Sigma). The resultant PCR product was cloned into pENTR-D-TOPO (see 2.2.3); XL1Blue cells were transformed using the reaction and positive colonies were selected by colony PCR (see appendix paragraph 3, program C). The constructs obtained were confirmed by restriction digestion and DNA sequencing. By LB reaction using LR clonase II, the SCD sequence was inserted into linear baculovirus DNA. The recombinant baculovirus containing the SCD sequence was confirmed by PCR.

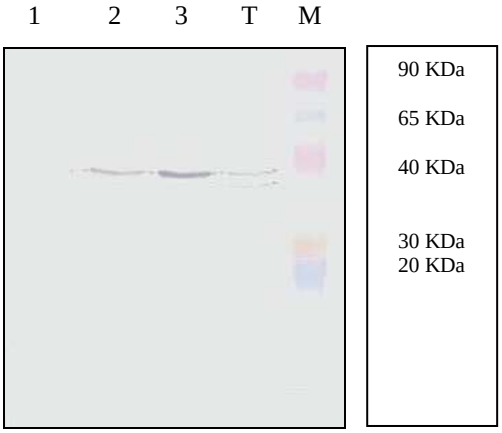


Fig.3.30. Western blot analysis: samples harvested after 96 hours of infection with 1) 15 µl V2, 2) 7 µl V2 and 3) 3 µl V2. T: total fraction, before centrifugation; M: protein ladder.

To generate recombinant baculovirus particles, $8x10^5$ Sf9 cells were transfected with the LB recombination reaction (see 2.3.4). V0 and V1 viral stocks were prepared; different quantities of V1 were tested infecting $1x10^6$ Sf9 and High Five

cells seeded in a 6-well plate in 2 ml/well of medium. After 96 hours of infection cells were harvested and suspended for one hour in mild agitation in PBS 1X and 1% DDM. The suspended material was centrifuged at the maximum speed for one hour at 4°C. Samples were analyzed by SDS-PAGE and Western blot before and after centrifugation. The protein was expressed by both cell lines but only in the Sf9 cells the protein was present in the soluble fraction (data not shown). Therefore the V1 viral stock was used to generate a large-scale, high-titer viral stock (V2 viral stock) for expression studies in Sf9 cells. Different quantities of V2 were tested as described previously, the solubilized material was loaded on a gel after centrifugation and the protein of interest was detected by Western blot analysis (see Fig.3.30).

B.3.1.3.2. Large-scale expression and purification of human SCD

To obtain an adequate amount of pure hSCD, a 300 ml-culture was prepared; Sf9 cells in mid-log phase in suspension culture were infected with recombinant baculovirus (3 µl of V2/1x10^6 cells) and incubated in a 100 rpm orbital shaker at 27°C for 96 hours. Infected cells were harvested by centrifugation, cell pellets were resuspended in 30 ml of PBS 1X, 1% DDM and 50 µM Fe^{2+} solution ($Fe(NH_3)_2(SO_4)_2$ dissolved in 10-fold molar excess of ascorbate in degassed 60 mM Hepes pH 7.4)(Goren and Fox, 2008). The sample was left for 1 hour in mild agitation at 4°C. The soluble fraction was obtained by centrifugation at 100000g for 45 minutes.

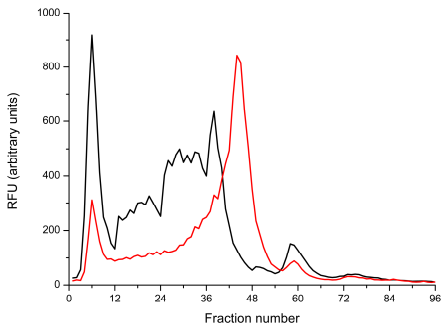


Fig.3.31. Determining the monodispersity of hSCD-GFP fusion in DDM (red) and LDAO (black) using FSEC.

The best solubilizing detergent was previously identified in small scale purification trials with the GFP-fusion protein in *S. cerevisiae*: the hSCD sequence was cloned in pDDGFP-2 and overexpressed. As suggested by FSEC, only the DDM solubilized protein seems to be monodisperse (see Fig.3.31).
Detergent-solubilized hSCD was purified to homogeneity using IMAC; the sample was loaded on a column filled with 1 ml of the resin Ni Sepharose 6 Fast Flow in the presence of 10 mM imidazole. The column was then washed with EB with the addition of 60 mM imidazole and the protein was eluted in 300 mM imidazole (see Fig.3.32A). At the same time the protein identity was confirmed by Western blot analysis (see Fig.3.32B).

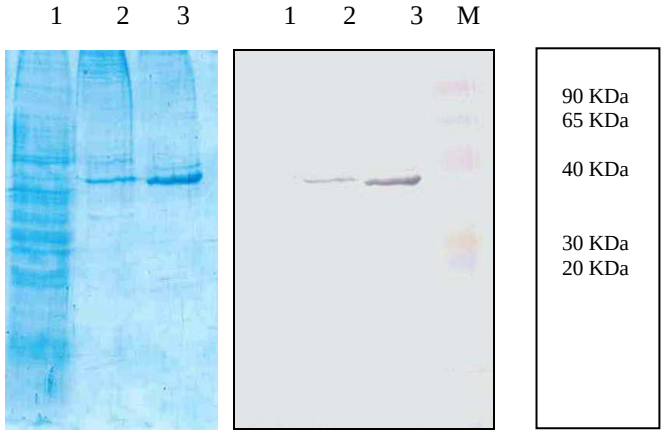


Fig.3.32. Samples collected during IMAC: A) 12% standard SDS-PAGE and B) Western blot analysis. 1) Flow through with 10 mM imidazole, washing with 2) 60 mM imidazole and 3) 300 mM imidazole. M: protein ladder.

The his-tag was not removed and the sample was concentrated using a centrifugal concentrator (cut off 50 KDa). Initial crystal trials were set but up to now no crystals were obtained.

B.3.1.Discussion

B.3.1.1. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *S.cerevisiae*

One way to improve the chances of success in obtaining pure membrane proteins for biochemical and structural analysis is to use GFP-based fusion technology (Newstead *et al*., 2007; Drew *et al*., 2008). As the C-terminal GFP folds and becomes fluorescent only if the upstream membrane protein integrates into the membrane, the resultant fluorescence is a fast and accurate measure of membrane-integrated expression. Fluorescence is easy to measure directly in liquid culture, standard SDS-gels and detergent-solubilized membranes. Detergent-solubilized membranes can also be subjected to fluorescence size-exclusion chromatography (FSEC) to measure the 'monodispersity' of the sample. To date, 96 membrane proteins from *C.glabrata* were cloned into a GFP-fusion vector. After obtaining colonies, the expression of one colony from each transformation was tested. Analyzing the overexpression levels, 12 clones were selected as the best overexpressing targets. These data emphasize the importance of initially screening the overexpression potential of many membrane proteins. Modest improvement could be obtained for a third of the proteins tested by adding DMSO or His to the culture medium. Detergent screening and FSEC analysis was carried out for 12 membrane proteins. The criterion for monodispersity was that a protein peak should be symmetric and is equal to or larger than any 'free GFP' or aggregation peak(s). There was no correlation between the efficiency of the detergent to extract the fusion protein and the quality of the FSEC traces. Seven of twelve yeast membrane proteins tested were monodisperse in at least one detergent. Considering the expression level, the stability after solubilization and the medical relevance the most promising clone, C4, was chosen to scale up the purification.

B.3.1.1.1. Farnesyl cysteine-carboxyl methyltransferase from *C.glabrata*

The clone C4 (Ste14, farnesyl cysteine-carboxyl methyltransferase) was expressed in a large-scale culture (10 l). Isolated membranes were solubilised and Ste14

purified with both the detergents tested by FSEC (LDAO and DDM). The two purifications gave about the same yield and the same final protein purity (2.5-3 mg). The SEC profile suggests that the protein is monodisperse in both detergents. Although several crystallization trials were set up, no crystals were obtained. In the near future, it is planned to improve the purification protocol increasing the purity up to >95% and, at the same time, to asses the correct fold and catalytic properties of the enzyme by measuring the *in vitro* activity (Hrycyna and Clarke, 1990).

B.3.1.1.2. Farnesyl cysteine-carboxyl methyltransferase from *A.thaliana*

The Ste14 homologue from *A.thaliana* (*At*Ste14A) was also cloned in a plasmid for expression in *E.coli*. The protein is smaller than Ste14 from *C.glabrata* and lacks a 50-amino acid cytoplasmic portion on the N-terminal end. *At*Ste14 can be easily extracted from the membranes in LDAO and the SEC profile suggests that the protein is monodisperse. Several crystallization trials were set up in the presence of 2 mM S-(5′-adenosyl)-L-methionine chloride (a substrate of the enzyme) or 2 mM of the selective inhibitor N-acetyl-S-farnesyl-L-cysteine, but up to now no crystals were obtained. This is probably due to a low protein concentration (8 mg/ml) in the sample. In collaboration with prof. B. Byrne (Imperial College, London), it is now planned to clone the cDNA in the pDDGFP-2 vector for expression in *S.cereviesiae*. The availability of a large quantity of pure protein will allow to prepare several crystal screenings at different concentrations of the apo protein or in presence of different ligands.

B.3.1.2. Stearoyl-CoA 9-desaturase

There is no published protocol for a large-scale expression and purification of human SCD. Recently, the functional complex of human stearoyl-CoA desaturase was produced using a wheat germ cell-free method but no reasonable enzyme quantity for structural studies was obtained with this system (Goren and Fox, 2008). Here, the baculovirus expression system was selected as a method for large scale production. The protein can be efficiently solubilised from the insect cell membranes with DDM. A purification protocol was established starting from 300

ml of insect cell culture. Up to now 0.5 mg of 90% pure protein was obtained from 300 ml of culture. We intend to conduct activity studies to verify if the enzyme is active and, subsequently, to scale up the protein production for structural studies.

Appendix

1. Media compositions

1.1. *Escherichia coli* growth media

Liquid Luria-Bertani (LB) medium
1% tryptone
0.5% yeast extract
0.5% NaCl
For plates 1.5% bacteriological agar was added.

1.2. *Saccharomyces cerevisiae* growth media

Liquid URA medium
2 g yeast synthetic drop-out medium without Ura
6.7 g yeast nitrogen base without aminoacid
2% glucose (for pre-culture) or 0.1% glucose (expression culture)
For plates 1.5% bacteriological agar and 2% glucose was added.

Liquid YPD medium
2% tryptone
1% yeast extract
2% glucose

1.3. Insect cells growth media

Sf-900 II SFM with the addition of 10% FSB (fetal bovine serum) was used for Sf9 cell line.

Express Five® SFM with the addition of 2 mM L-glutamine was used for High Five cell line.

To all this media antibiotics and antimycotics mix (penicillin G 10 units/ml, streptomycin 10 μg/ml, amphotericin B 0.025 μg/ml) was added before using.

2. Buffer compositions

2.1. Agarose gel running buffer

TBE
45 mM Tris-borate
1 mM EDTA

2.2. Agarose gel loading buffer (6X)

30% glycerol
0.25% bromophenol blue

2.3. SDS-PAGE running buffer

25 mM Tris-HCl
200 mM Glycine
0.1% (w/v) SDS

2.4. SDS-PAGE loading buffer (4X)

10% glycerol
62.5 mM Tris-HCl pH 6.8
2% SDS
0.01 mg/ml bromophenol blue
0.02 5% β-mercaptoethanol

2.5. In gel-fluorescence loading buffer (2X)

50 mM Tris–HCl (pH 7.6)
5% glycerol, 5 mM EDTA (pH 8.0)
0.02% bromophenol blue
Before use 200 ml of 20% SDS and 100 ml of 0.5M DTT was added.

2.6. SDS-PAGE staining solution

10% (v/v) acetic acid
0.1% (w/v) Coomassie Blue dye
40% (v/v) ethanol

2.7. SDS-PAGE destaining solution

10% (v/v) acetic acid
10% isopropanol

2.8. Western blotting transfer buffer

25 mM Tris-HCl
192 mM Glycine
10% Methanol

2.9. Soluble proteins purification

CRB (1)
20 mM Tris-HCl pH 7.5
500 mM NaCl

TBS
20 mM Tris-HCl pH 7.5
150 mM NaCl

2.10. Membrane proteins purification

CRB (2)
50 mM Tris-HCl pH 7.5
1 mM EDTA
0.6 M sorbitol

EB

1X PBS

150 mM NaCl

10% glycerol

3x CMC detergent of choice

MRB

20 mM Tris-HCl pH 7.5

0.3 M sucrose

DB

20 mM Tris-HCl pH 7.5

150 mM NaCl

3x CMC detergent of choice

PBS

137 mM NaCl

2.7 mM KCl

10 mM Na_2HPO_4

2 mM KH_2PO_4

The pH is adjusted to 7.5.

3. PCR programs

A. Sequence amplification from cDNA

Plasmidic DNA (5-50 ng)
Primer FOR (25µM)
Primer REV (25µM)
dNTPs (2.5mM)
Buffer Taq Polimerase (with $MgCl_2$)
High Fidelity Taq Polimerase (2-5 units)

Step	Temperature	Time
Initial denaturation	95°C	10 minutes
Amplification (35 cycles)	95°C (denaturation)	1 minute

	It depends on primer Tm (annealing)	1 minute
	72°C (extension)	It depends on sequence length and polymerase speed
Final extension	72°C	10 minutes

B. Site-directed mutagenesis

Point mutations were introduced by PCR using the QuikChange™ site-directed mutagenesis kit (Stratagene).

Plasmidic DNA (5-50 ng)
Primer FOR (25µM)
Primer REV (25µM)
dNTPs (2.5 mM)
Buffer PFU (with $MgCl_2$)
PFU (2-5 units)

Step	Temperature	Time
Initial denaturation	95°C	1 minute
Amplification (18 cycles)	95°C (denaturation)	1 minute
	55°C (annealing)	1 minute
	68°C (extension)	It depends on sequence length (plasmid and protein sequence) and polymerase speed
Final extension	68°C	10 minutes

C. Colony PCR

Bacterial colony diluted in water
Primer FOR (25µM)
Primer REV (25µM)
dNTPs (2.5 mM)
Buffer Home made TAQ polymerase (with $MgCl_2$)
Home made TAQ polymerase (2-5 units)

Step	Temperature	Time
Initial denaturation	95°C	10 minutes
Amplification (30 cycles)	95°C (denaturation)	1 minute

	55°C (annealing)	1 minute
	72°C (extension)	It depends on sequence length and polymerase speed
Final extension	72°C	10 minutes

4. Standard SDS-PAGE compositions

Running gel

	12% acrylamide	15% acrylamide
H_20	1.32 ml	0.92 ml
30% acrylamide mix	1.6 ml	2 ml
1.5 M Tris pH 8.8	1 ml	1 ml
10% SDS	0.04 ml	0.04 ml
10% APS	0.04 ml	0.04 ml
TEMED	0.004 ml	0.004 ml

Stacking gel

	5% acrylamide
H_20	1.4 ml
30% acrylamide mix	0.33 ml
1 M Tris pH 6.8	0.25 ml
10% SDS	0.02 ml
10% APS	0.02 ml
TEMED	0.002 ml

5. Sequences

5.1. Bile acid coenzyme A:amino acid N-acyltransferase

5.1.1. Human BAAT cDNA sequence

ATGATCCAGTTGACAGCTACCCCTGTGAGTGCACTTGTTGATGAGCCAGTGCATA
TCCAAGCTACAGGCCTGATTCCCTTTCAGATGGTGAGTTTTCAGGCATCACTGGA
AGATGAAAACGGAGACATGTTTTATTCTCAAGCCCACTATAGGGCCAATGAATTC

GGTGAGGTGGACCTGAATCATGCTTCTTCACTTGGAGGGGATTATATGGGAGTCC
ACCCCATGGGTCTCTTCTGGTCTCTGAAACCTGAAAAGCTATTAACAAGACTGTT
GAAAAGAGATGTGATGAATAGGCCTTTCCAGGTCCAAGTAAAACTTTATGACTTA
GAGTTAATAGTGAACAATAAAGTTGCCAGTGCTCCAAAGGCCAGCCTGACTTTGG
AGAGGTGGTATGTGGCACCTGGTGTCACACGAATTAAGGTTCGAGAAGGCCGCCT
TCGAGGAGCTCTCTTTCTCCCTCCAGGAGAGGGTCTCTTCCCAGGGGTAATTGAT
TTGTTTGGTGGTTTGGGTGGGCTGCTTGAATTTCGGGCCAGCCTCCTAGCCAGTC
GTGGCTTCGCCTCCTTGGCCTTGGCTTACCATAACTATGAAGACCTGCCCCGCAA
ACCAGAAGTAACAGATTTGGAATATTTTGAGGAGGCTGCCAACTTTCTCCTGAGA
CATCCAAAGGTCTTTGGCTCAGGCGTTGGGGTAGTCTCTGTATGTCAAGGAGTAC
AGATTGGACTATCTATGGCTATTTACCTAAAGCAAGTCACAGCCACGGTACTTAT
TAATGGGACCAACTTTCCTTTTGGCATTCCACAGGTATATCATGGTCAGATCCAT
CAGCCCCTTCCCCATTCTGCACAATTAATATCCACCAATGCCTTGGGGTTACTAG
AGCTCTATCGCACTTTTGAGACAACTCAAGTTGGGGCCAGTCAATATTTGTTTCC
TATTGAAGAGGCCCAGGGGCAATTCCTCTTCATTGTAGGAGAAGGTGATAAGACT
ATCAACAGCAAAGCACACGCTGAACAAGCCATAGGACAGCTGAAGAGACATGGGA
AGAACAACTGGACCCTGCTATCTTACCCTGGGGCAGGCCACCTGATAGAACCTCC
CTATTCTCCTCTGTGCTGTGCCTCAACGACCCACGATTTGAGGTTACACTGGGGA
GGAGAGGTGATCCCACACGCAGCTGCACAGGAACATGCTTGGAAGGAGATCCAGA
GATTTCTCAGGAAGCACCTCATTCCAGATGTGACCAGTCAACTCTAA

5.1.2. Human BAAT protein sequence

MIQLTATPVSALVDEPVHIQATGLIPFQMVSFQASLEDENGDMFYSQAHYRANEF
GEVDLNHASSLGGDYMGVHPMGLFWSLKPEKLLTRLLKRDVMNRPFQVQVKLYDL
ELIVNNKVASAPKASLTLERWYVAPGVTRIKVREGRLRGALFLPPGEGLFPGVID
LFGGLGGLLEFRASLLASRGFASLALAYHNYEDLPRKPEVTDLEYFEEAANFLLR
HPKVFGSGVGVVSVCQGVQIGLSMAIYLKQVTATVLINGTNFPFGIPQVYHGQIH
QPLPHSAQLISTNALGLLELYRTFETTQVGASQYLFPIEEAQGQFLFIVGEGDKT
INSKAHAEQAIGQLKRHGKNNWTLLSYPGAGHLIEPPYSPLCCASTTHDLRLHWG
GEVIPHAAAQEHAWKEIQRFLRKHLIPDVTSQL

Number of amino acids: 418

Molecular weight: 46271 Da

5.1.3. Human BAAT protein sequence in pET15

Thrombin cleavage site

MGSSHHHHHHSSGLVPRGSHMIQLTATPVSALVDEPVHIQATGLIPFQMVSFQAS
LEDENGDMFYSQAHYRANEFGEVDLNHASSLGGDYMGVHPMGLFWSLKPEKLLTR
LLKRDVMNRPFQVQVKLYDLELIVNNKVASAPKASLTLERWYVAPGVTRIKVREG
RLRGALFLPPGEGLFPGVIDLFGGLGGLLEFRASLLASRGFASLALAYHNYEDLP
RKPEVTDLEYFEEAANFLLRHPKVFGSGVGVVSVCQGVQIGLSMAIYLKQVTATV
LINGTNFPFGIPQVYHGQIHQPLPHSAQLISTNALGLLELYRTFETTQVGASQYL
FPIEEAQGQFLFIVGEGDKTINSKAHAEQAIGQLKRHGKNNWTLLSYPGAGHLIE
PPYSPLCCASTTHDLRLHWGGEVIPHAAAQEHAWKEIQRFLRKHLIPDVTSQL

Number of amino acids (__): 421

Molecular weight (__): 46552.4

C = cysteines

CDH = catalytic triade

SQL = peroxisome sorting signal

5.2. Farnesyl cysteine-carboxyl methyltransferase

5.2.1.Ste14 cDNA sequence from *C.glabrata*

ATGACTGGATCAAACAAATCTGAGAAATTTGTGCCTGATGCGGAAGGCCTGCCTG
TAATAATAAATGGTAAAGCATACCCAGACATTCGTCGCAATCCATTAGATGACAT
TGCTGTAACCTCATTTGCACTGGGAATCCTTCTCGGTATATTCGTTGGTCTCCTT
ACATTCACACAGTTCAGAAATTTCAATGTTTATATCATTGCCCTGTCAATATTCC
ATTTCCTTGAGTTTTATGTAACTGCCAAGGTGAACCCAGGTAAAGTCAACTCAGA
CTCCTTCCTTTTGAACAATGGTATCGGTTATCTTGGTGCACACATGTTCGGTGTC
CTTGAGTGTCTTGTGGAAAGTATATTCTTTCCTAATATGAAGAAGATTAGCAACT
CTTGGGTAGCAATATTCTTAACAATCATCGGGGTAATATTGATACTTTTAGGTCA
AGTGTCTAGATCAATGGCCATGTATCAAGCTGGTAAGTCATTCTCTCATATCTTG
AAGACGGAGAAACTCGACGATCACACATTAGTGACCACAGGCATTTATGGATATT
TGAGGCATCCAAGTTATTTTGGTTTTTTCTGGTGGGCTATAGGGACTCAATTGCT
ACTCTTAAATCCAATTGCATTGATTGTCTTTGTAGGCGTGCTTTGGAGCTTCTTT
AATAAGCGCATATTAATTGAGGAGAAATACCTTATACAGTTCTTTGGTGAGCAGT
ACATAAACTACAAAAAAAGGGTCAAGGTATGGATACCGTTAATTGATTAG

5.2.2. Ste14 protein sequence from *C.glabrata*

MTGSNKSEKFVPDAEGLPVIINGKAYPDIRRNPLDDIAVTSFALGILLGIFVGLL
TFTQFRNFNVYIIALSIFHFLEFYVTAKVNPGKVNSDSFLLNNGIGYLGAHMFGV
LECLVESIFFPNMKKISNSWVAIFLTIIGVILILLGQVSRSMAMYQAGKSFSHIL
KTEKLDDHTLVTTGIYGYLRHPSYFGFFWWAIGTQLLLLNPIALIVFVGVLWSFF
NKRILIEEKYLIQFFGEQYINYKKRVKVWIPLID

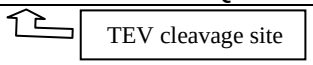
Number of amino acids: 254

Molecular weight: 28885 Da

5.2.3. Ste14 protein sequence from *C.glabrata* in pDDGFP-2

MTGSNKSEKFVPDAEGLPVIINGKAYPDIRRNPLDDIAVTSFALGILLGIFVGLL
TFTQFRNFNVYIIALSIFHFLEFYVTAKVNPGKVNSDSFLLNNGIGYLGAHMFGV
LECLVESIFFPNMKKISNSWVAIFLTIIGVILILLGQVSRSMAMYQAGKSFSHIL

KTEKLDDHTLVTTGIYGYLRHPSYFGFFWWAIGTQLLLLNPIALIVFVGVLWSFF
NKRILIEEKYLIQFFGEQYINYKKRVKVWIPLIDGENLYFQGQPSKGEE...



Number of amino acids (__): 261

Molecular weight (__): 29737 Da

C = cysteine

SKGEE… = GFP sequence

5.2.4. Ste14 cDNA sequence from *A.thaliana*

ATGACAGAGATCTTCAGTGACACCAGCATCAGACAGTTATCTCAAATGCTACTAT
CACTAATCTTCTTCCACATATCCGAATACATTCTAGCCATCACCATTCACGGAGC
ATCAAACGTAACTCTTAGTTCGCTTTTAATCACCAAGCATTACGCTTTAGCAATG
CTTCTGTCGCTTCTCGAATACCTAACGGAGATTATCCTCTTCCCGGGGCTGAAAC
AACACTGGTGGGTCAGCAACTTTGGACTCATAATGATCATCGTTGGGGAAATCAT
CAGGAAGGCAGCGATAATAACAGCGGGAAGATCGTTCACTCACCTCATAAAGATC
AACTACGAAGAGCATCACGGGCTTGTGACTCATGGTGTGTATAGACTAATGAGGC
ATCCAAGTTACTGCGGTTTTCTCATCTGGTCGGTCGGGACACAAGTTATGCTCTG
TAACCCCGTTTCAGCAGTTGCGTTCGCGGTTGTCGTGTGGCGGTTTTTTGCTCAG
AGAATACCGTACGAGGAGTATTTTCTGAATCAGTTTTTTGGGGTACAGTATCTAG
AGTATGCAGAGAGTGTTGCCTCTGGTGTTCCATTTGTGAACTGA

5.2.5. Ste14 protein sequence from *A.thaliana*

MTEIFSDTSIRQLSQMLLSLIFFHISEYILAITIHGASNVTLSSLLITKHYALAM
LLSLLEYLTEIILFPGLKQHWWVSNFGLIMIIVGEIIRKAAIITAGRSFTHLIKI
NYEEHHGLVTHGVYRLMRHPSYCGFLIWSVGTQVMLCNPVSAVAFAVVVWRFFAQ
RIPYEEYFLNQFFGVQYLEYAESVASGVPFVN

Number of amino acids: 197

Molecular weight: 22525 Da

5.2.6. Ste14 protein sequence from *A.thaliana* in pET15

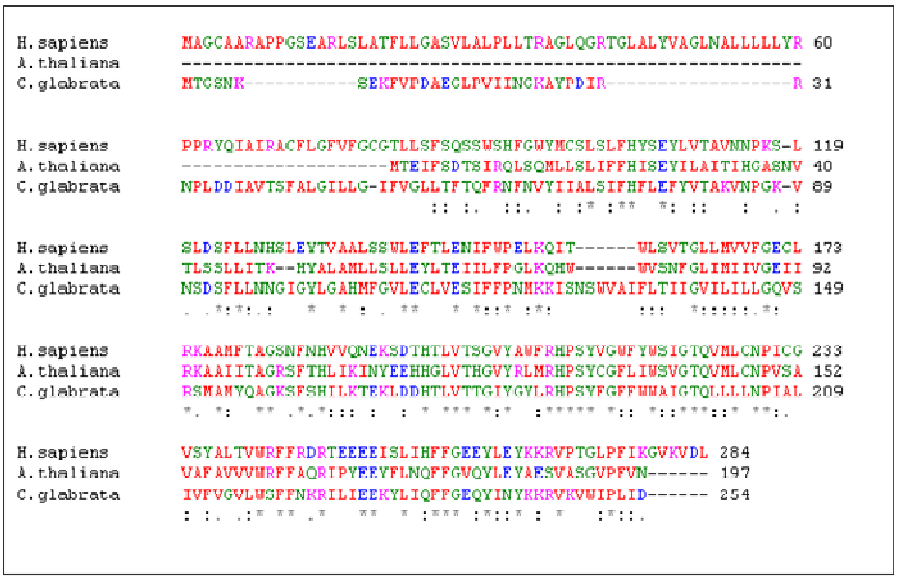
MGSSHHHHHHSSGLVPRGSHMTEIFSDTSIRQLSQMLLSLIFFHISEYILAITIH
GASNVTLSSLLITKHYALAMLLSLLEYLTEIILFPGLKQHWWVSNFGLIMIIVGE
IIRKAAIITAGRSFTHLIKINYEEHHGLVTHGVYRLMRHPSYCGFLIWSVGTQVM
LCNPVSAVAFAVVVWRFFAQRIPYEEYFLNQFFGVQYLEYAESVASGVPFVN

Number of amino acids: 217

Molecular weight: 24688 Da

LVPRGS = thrombin cleavage site

5.2.7. Sequences alignment of human Icmt, AtSte14 and yeast Ste14



5.3. Stearoyl-CoA 9-desaturase

5.3.1. Human SCD cDNA sequence

ATGCCGGCCCACTTGCTGCAGGACGATATCTCTAGCTCCTATACCACCACCACCA
CCATTACAGCGCCTCCCTCCAGGGTCCTGCAGAATGGAGGAGATAAGTTGGAGAC
GATGCCCCTCTACTTGGAAGACGACATTCGCCCTGATATAAAAGATGATATATAT
GACCCCACCTACAAGGATAAGGAAGGCCCAAGCCCCAAGGTTGAATATGTCTGGA
GAAACATCATCCTTATGTCTCTGCTACACTTGGGAGCCCTGTATGGGATCACTTT
GATTCCTACCTGCAAGTTCTACACCTGGCTTTGGGGGGTATTCTACTATTTTGTC
AGTGCCCTGGGCATAACAGCAGGAGCTCATCGTCTGTGGAGCCACCGCTCTTACA
AAGCTCGGCTGCCCCTACGGCTCTTTCTGATCATTGCCAACACAATGGCATTCCA
GAATGATGTCTATGAATGGGCTCGTGACCACCGTGCCCACCACAAGTTTTCAGAA
ACACATGCTGATCCTCATAATTCCCGACGTGGCTTTTTCTTCTCTCACGTGGGTT
GGCTGCTTGTGCGCAAACACCCAGCTGTCAAAGAGAAGGGGAGTACGCTAGACTT
GTCTGACCTAGAAGCTGAGAAACTGGTGATGTTCCAGAGGAGGTACTACAAACCT
GGCTTGCTGATGATGTGCTTCATCCTGCCCACGCTTGTGCCCTGGTATTTCTGGG
GTGAAACTTTTCAAAACAGTGTGTTCGTTGCCACTTTCTTGCGATATGCTGTGGT
GCTTAATGCCACCTGGCTGGTGAACAGTGCTGCCCACCTCTTCGGATATCGTCCT
TATGACAAGAACATTAGCCCCCGGGAGAATATCCTGGTTTCACTTGGAGCTGTGG
GTGAGGGCTTCCACAACTACCACCACTCCTTTCCCTATGACTACTCTGCCAGTGA
GTACCGCTGGCACATCAACTTCACCACATTCTTCATTGATTGCATGGCCGCCCTC
GGTCTGGCCTATGACCGGAAGAAAGTCTCCAAGGCCGCCATCTTGGCCAGGATTA
AAAGAACCGGAGATGGAAACTAA

5.3.2. Human SCD protein sequence

MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIY
DPTYKDKEGPSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFV
SALGITAGAHRLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHHKFSE
THADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKP
GLLMMCFILPTLVPWYFWGETFQNSVFVATFLRYAVVLNATWLVNSAAHLFGYRP
YDKNISPRENILVSLGAVGEGFHNYHHSFPYDYSASEYRWHINFTTFFIDCMAAL
GLAYDRKKVSKAAILARIKRTGDGN

Number of amino acids: 355

Molecular weight: 41087 Da

5.3.3. Human SCD protein sequence in linear baculovirus DNA

MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIY
DPTYKDKEGPSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFV
SALGITAGAHRLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHHKFSE
THADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKP
GLLMMCFILPTLVPWYFWGETFQNSVFVATFLRYAVVLNATWLVNSAAHLFGYRP
YDKNISPRENILVSLGAVGEGFHNYHHSFPYDYSASEYRWHINFTTFFIDCMAAL
GLAYDRKKVSKAAILARIKRTGDGNLVPRGSKLIIKGGRADPAFLYKVVRMNEDL
GKPIPNPLLGLDSTRTGHHHHHH

Number of amino acids: 408

Molecular weight: 46967 Da

LVPRGS = thrombin cleavage site

References

Alguel Y, Leung J, Sing S, Rana R, Civiero L, Alves C, Byrne B. New tools for membrane protein research. *Curr Protein Pept Sci*. 2010 Jan 21(Epub ahead of print);

Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science.* 2009 Mar 27;323(5922):1718-22;

Bergo MO, Leung GK, Ambroziak P, Otto JC, Casey PJ, Young SG. Targeted mislocalization of K-Ras in mammalian cells. *J Biol Chem*. 2000 Jun 9;275(23):17605-10;

Buch C, Hunt MC, Alexson SE, Hallberg E. Localization of peroxisomal matrix proteins by photobleaching. *Biochem Biophys Res Commun.* 2009 Oct 16;388(2):355-9;

Caffrey M. *A* lipid's eye view of membrane protein crystallization in mesophases. *Curr Opin Struct Biol.* 2000 Aug;10(4):486-97;

Caffrey M. Membrane protein crystallization. *J Struct Biol.* 2003 Apr;142(1):108-32;

Carlton VE, Harris BZ, Puffenberger EG, Batta AK, Knisely AS, Robinson DL, Strauss KA, Shneider BL, Lim WA, Salen G, Morton DH, Bull LN. Complex inheritance of familial hypercholanemia with associated mutations in TJP2 and BAAT. *Nat Genet.* 2003 May;34(1):91-6;

Casey PJ, Seabra MC. Protein prenyltransferases. *J Biol Chem*. 1996 Mar 8;271(10):5289-92;
Casey PJ. Protein lipidation in cell signaling. *Science*. 1995 Apr 14;268(5208):221-5;

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*. 2007 Nov 23;318(5854):1258-65;

Chiang JY. Bile acid regulation of gene expression: roles of nuclear hormone receptors. *Endocr Rev*. 2002 Aug;23(4):443-63;

Dai Q, Choy E, Chiu V, Romano J, Slivka SR, Steitz SA, Michaelis S, Philips MR. Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum. *J Biol Chem.* 1998 Jun 12;273(24):15030-4;

Deisenhofer J, Epp O, Miki K, Huber R, Michel H. X-ray structure analysis of a membrane protein complex. Electron density map at 3 A resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodopseudomonas viridis*. *J Mol Biol*. 1984 Dec 5;180(2):385-98;

Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat Protoc*. 2008;3(5):784-98;

Falany CN, Fortinberry H, Leiter EH, Barnes S. Cloning, expression, and chromosomal localization of mouse liver bile acid CoA:amino acid N-acyltransferase. *J Lipid Res*. 1997 Jun;38(6):1139-48;

Falany CN, Johnson MR, Barnes S, Diasio RB. Glycine and taurine conjugation of bile acids by a single enzyme. Molecular cloning and expression of human liver bile acid CoA:amino acid N-acyltransferase. *J Biol Chem*. 1994 Jul 29;269(30):19375-9;

Ferdinandusse S, Houten SM. Peroxisomes and bile acid biosynthesis. *Biochim Biophys Acta*. 2006 Dec;1763(12):1427-40;

Ferguson AD, McKeever BM, Xu S, Wisniewski D, Miller DK, Yamin TT, Spencer RH, Chu L, Ujjainwalla F, Cunningham BR, Evans JF, Becker JW. Crystal structure of inhibitor-bound human 5-lipoxygenase-activating protein. *Science*. 2007 Jul 27;317(5837):510-2;

Gomez FE, Miyazaki M, Kim YC, Marwah P, Lardy HA, Ntambi JM, Fox BG. Molecular differences caused by differentiation of 3T3-L1 preadipocytes in the

presence of either dehydroepiandrosterone (DHEA) or 7-oxo-DHEA. *Biochemistry*. 2002 Apr 30;41(17):5473-82;

Grisshammer R, White JF, Trinh LB, Shiloach J. Large-scale expression and purification of a G-protein-coupled receptor for structure determination - an overview. *J Struct Funct Genomics*. 2005;6(2-3):159-63;

Guy JE, Whittle E, Kumaran D, Lindqvist Y, Shanklin J. The crystal structure of the ivy Delta4-16:0-ACP desaturase reveals structural details of the oxidized active site and potential determinants of regioselectivity. *J Biol Chem*. 2007 Jul 6;282(27):19863-71;

He D, Barnes S, Falany CN. Rat liver bile acid CoA:amino acid N-acyltransferase: expression, characterization, and peroxisomal localization. *J Lipid Res*. 2003 Dec;44(12):2242-9;

Heinemann FS, Korza G, Ozols J. A plasminogen-like protein selectively degrades stearoyl-CoA desaturase in liver microsomes. *J Biol Chem*. 2003 Oct 31;278(44):42966-75;

Hofmann AF, Hagey LR. Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cell Mol Life Sci*. 2008 Aug;65(16):2461-83;

Houten SM, Watanabe M, Auwerx J. Endocrine functions of bile acids. *EMBO J*. 2006 Apr 5;25(7):1419-25;

Hunte C, Michel H. Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol*. 2002 Aug;12(4):503-8;

Imai Y, Davey J, Kawagishi-Kobayashi M, Yamamoto M. Genes encoding farnesyl cysteine carboxyl methyltransferase in *Schizosaccharomyces pombe* and *Xenopus laevis*. *Mol Cell Biol*. 1997 Mar;17(3):1543-51;

Iwata S. In Crystallization informatics of membrane proteins. In Methods and results in crystallization of membrane proteins. Edited by Iwata. International University Line; 2003. 281-294;

Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*. 2008 Nov 21;322(5905):1211-7;

Jasti J, Furukawa H, Gonzales EB, Gouaux E. Structure of acid-sensing ion channel 1 at 1.9 A resolution and low pH. *Nature*. 2007 Sep 20;449(7160):316-23;

Jidenko M, Nielsen RC, Sørensen TL, Møller JV, le Maire M, Nissen P, Jaxel C. Crystallization of a mammalian membrane protein overexpressed in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 2005 Aug 16;102(33):11687-91;

Johnson MR, Barnes S, Kwakye JB, Diasio RB. Purification and characterization of bile acid-CoA:amino acid N-acyltransferase from human liver. *J Biol Chem*. 1991 Jun 5;266(16):10227-33;

Junge F, Schneider B, Reckel S, Schwarz D, Dötsch V, Bernhard F. Large-scale production of functional membrane proteins. *Cell Mol Life Sci*. 2008 Jun;65(11):1729-55;

Kato H, Sakaki K, Mihara K. Ubiquitin-proteasome-dependent degradation of mammalian ER stearoyl-CoA desaturase. *J Cell Sci*. 2006 Jun 1;119(Pt 11):2342-53;

Kim E, Ambroziak P, Otto JC, Taylor B, Ashby M, Shannon K, Casey PJ, Young SG. Disruption of the mouse Rce1 gene results in defective Ras processing and mislocalization of Ras within cells. *J Biol Chem*. 1999 Mar 26;274(13):8383-90;

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.

Kunji ER, Chan KW, Slotboom DJ, Floyd S, O'Connor R, Monné M. Eukaryotic membrane protein overproduction in *Lactococcus lactis*. *Curr Opin Biotechnol*. 2005 Oct;16(5):546-51;

Liguori L, Marques B, Villegas-Méndez A, Rothe R, Lenormand JL. Production of membrane proteins using cell-free expression systems. *Expert Rev Proteomics*. 2007 Feb;4(1):79-90;

Lin SH, Guidotti G. Purification of membrane proteins. *Methods Enzymol*. 2009;463:619-29;

Long SB, Campbell EB, Mackinnon R. Crystal structure of a mammalian voltage-dependent Shaker family K^+ channel. *Science*. 2005 Aug 5;309(5736):897-903;

Long SB, Casey PJ, Beese LS. Reaction path of protein farnesyltransferase at atomic resolution. *Nature*. 2002 Oct 10;419(6907):645-50;

Lundstrom K. Semliki Forest virus vectors for rapid and high-level expression of integral membrane proteins. *Biochim Biophys Acta*. 2003 Feb 17;1610(1):90-6;

Martinez Molina D, Wetterholm A, Kohl A, McCarthy AA, Niegowski D, Ohlson E, Hammarberg T, Eshaghi S, Haeggström JZ, Nordlund P. Structural basis for synthesis of inflammatory mediators by human leukotriene C4 synthase. *Nature*. 2007 Aug 2;448(7153):613-6;

McPherson A. Current approaches to macromolecular crystallization. *Eur. J. Biochem*. 1990;189: 1-23;

Miller CW, Ntambi JM. Peroxisome proliferators induce mouse liver stearoyl-CoA desaturase 1 gene expression. *Proc Natl Acad Sci U S A*. 1996 Sep 3;93(18):9443-8;

Moche M, Shanklin J, Ghoshal A, Lindqvist Y. Azide and acetate complexes plus two iron-depleted crystal structures of the di-iron enzyme delta9 stearoyl-acyl carrier protein desaturase. Implications for oxygen activation and catalytic intermediates. *J Biol Chem*. 2003 Jul 4;278(27):25072-80;

Newstead S, Kim H, von Heijne G, Iwata S, Drew D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression

and purification in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 2007 Aug 28;104(35):13936-41;

Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov*. 2006 Dec;5(12):993-6;

Parish CA, Rando RR. Isoprenylation/methylation of proteins enhances membrane association by a hydrophobic mechanism. *Biochemistry*. 1996 Jul 2;35(26):8473-7;

Pazirandeh M, Chirala SS, Wakil SJ. Site-directed mutagenesis studies on the recombinant thioesterase domain of chicken fatty acid synthase expressed in *Escherichia coli*. *J Biol Chem*. 1991 Nov 5;266(31):20946-52;

Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, Nissen P. Crystal structure of the plasma membrane proton pump. *Nature*. 2007 Dec 13;450(7172):1111-4;

Pellicoro A, van den Heuvel FA, Geuken M, Moshage H, Jansen PL, Faber KN. Human and rat bile acid-CoA:amino acid N-acyltransferase are liver-specific peroxisomal enzymes: implications for intracellular bile salt transport. *Hepatology*. 2007 Feb;45(2):340-8;

Pircher PC, Kitto JL, Petrowski ML, Tangirala RK, Bischoff ED, Schulman IG, Westin SK. Farnesoid X receptor regulates bile acid-amino acid conjugation. *J Biol Chem*. 2003 Jul 25;278(30):27703-11;

Privé GG. Detergents for the stabilization and crystallization of membrane proteins. *Methods*. 2007 Apr;41(4):388-97;

Reid TS, Terry KL, Casey PJ, Beese LS. Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol*. 2004 Oct 15;343(2):417-33.

Ren J, Knorr C, Huang L, Brenig B. Isolation and molecular characterization of the porcine stearoyl-CoA desaturase gene. *Gene*. 2004 Sep 29;340(1):19-30;

Russell DW. Fifty years of advances in bile acid synthesis and metabolism. *J Lipid Res*. 2009 Apr;50 Suppl:S120-5;

Russell DW. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem*. 2003;72:137-74;

Sapperstein S, Berkower C, Michaelis S. Nucleotide sequence of the yeast STE14 gene, which encodes farnesylcysteine carboxyl methyltransferase, and demonstration of its essential role in a-factor export. *Mol Cell Biol*. 1994 Feb;14(2):1438-49;

Sarramegna V, Talmont F, Demange P, Milon A. Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification. *Cell Mol Life Sci*. 2003 Aug;60(8):1529-46;

Schimerlik MI. Overview of membrane protein solubilization. *Curr Protoc Neurosci*. 2001 May;Chapter 5:Unit 5.9;

Schwarz D, Dötsch V, Bernhard F. Production of membrane proteins using cell-free expression systems. *Proteomics*. 2008 Oct;8(19):3933-46;

Sfakianos MK, Wilson L, Sakalian M, Falany CN, Barnes S. Conserved residues in the putative catalytic triad of human bile acid coenzyme A:amino acid N-acyltransferase. *J Biol Chem*. 2002 Dec 6;277(49):47270-5;

Shanklin J, Cahoon EB. Desaturation and related modifications of fatty acids. *Annu Rev Plant Physiol Plant Mol Biol*. 1998 Jun;49:611-641;

Shonsey EM, Sfakianos M, Johnson M, He D, Falany CN, Falany J, Merkler DJ, Barnes S. Bile acid coenzyme A: amino acid N-acyltransferase in the amino acid conjugation of bile acids. *Methods Enzymol*. 2005;400:374-94;

Solaas K, Ulvestad A, Söreide O, Kase BF. Subcellular organization of bile acid amidation in human liver: a key issue in regulating the biosynthesis of bile salts. *J Lipid Res*. 2000 Jul;41(7):1154-62;

Sperling P, Ternes P, Zank TK, Heinz E. The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids.* 2003 Feb;68(2):73-95;

Styles NA, Falany JL, Barnes S, Falany CN. Quantification and regulation of the subcellular distribution of bile acid coenzyme A:amino acid N-acyltransferase activity in rat liver. *J Lipid Res*. 2007 Jun;48(6):1305-15;

Tocher DR, Leaver MJ, Hodgson PA. Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. *Prog Lipid Res*. 1998 Jul-Aug;37(2-3):73-117;

Törnroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P. Structural mechanism of plant aquaporin gating. *Nature*. 2006 Feb 9;439(7077):688-94;

Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*. 2008 Jul 24;454(7203):486-91;

White SH. Biophysical dissection of membrane proteins. *Nature*. 2009 May 21;459(7245):344-6;

Whittle E, Cahoon EB, Subrahmanyam S, Shanklin J. A multifunctional acyl-acyl carrier protein desaturase from *Hedera helix L*. (English ivy) can synthesize 16- and 18-carbon monoene and diene products. *J Biol Chem*. 2005 Aug 5;280(31):28169-76;

Zhang H, Seabra MC, Deisenhofer J. Crystal structure of Rab geranylgeranyltransferase at 2.0 A resolution. *Structure*. 2000 Mar 15;8(3):241-51.

Papers and Meeting Communications

Curr Protein Pept Sci. 2010 Jan 21. [Epub ahead of print]

New Tools for Membrane Protein Research.

Alguel Y, Leung J, Singh S, Rana R, Civiero L, Alves C, Byrne B.

Division of Molecular Biosciences, Imperial College London, South Kensington, London
b.byrne@imperial.ac.uk.

The last five years have seen a dramatic increase in the number of membrane protein structures. The vast majority of these 191 unique structures are of membrane proteins from prokaryotic sources. Whilst these have provided unprecedented insight into the mechanism of action of these important molecules our understanding of many clinically important eukaryotic membrane proteins remains limited by a lack of high resolution structural data. It is clear that novel approaches are required to facilitate the structural characterization of eukaryotic membrane proteins. Here we review some of the techniques developed recently which are having a major impact on the way in which structural studies of eukaryotic membrane proteins are being approached. Several different high throughput approaches have been designed to identify membrane proteins most suitable for structural studies. One approach is to screen large numbers of related or non-related membrane proteins using GFP fusion proteins. An alternative involves generating large numbers of mutants of a single protein with a view to obtaining a fully functional but highly stable membrane protein. These, and other novel techniques that aim to facilitate the production of protein likely to yield well-diffracting crystals are described.

Current Protein and Peptide Science, 2010, *11,* 000-000

New Tools for Membrane Protein Research

Yilmaz Alguel, James Leung, Shweta Singh, Rohini Rana, Laura Civiero, Claudia Alves and Bernadette Byrne*

Division of Molecular Biosciences, Imperial College London, South Kensington, London, SW7 2AZ, UK

Abstract: The last five years have seen a dramatic increase in the number of membrane protein structures. The vast majority of these 191 unique structures are of membrane proteins from prokaryotic sources. Whilst these have provided unprecedented insight into the mechanism of action of these important molecules our understanding of many clinically important eukaryotic membrane proteins remains limited by a lack of high resolution structural data. It is clear that novel approaches are required to facilitate the structural characterization of eukaryotic membrane proteins. Here we review some of the techniques developed recently which are having a major impact on the way in which structural studies of eukaryotic membrane proteins are being approached. Several different high throughput approaches have been designed to identify membrane proteins most suitable for structural studies. One approach is to screen large numbers of related or non-related membrane proteins using GFP fusion proteins. An alternative involves generating large numbers of mutants of a single protein with a view to obtaining a fully functional but highly stable membrane protein. These, and other novel techniques that aim to facilitate the production of protein likely to yield well-diffracting crystals are described.

Keywords: Eukaryotic membrane proteins, high resolution structure determination, high throughput pipelines, conformational thermostabilization, aggregation, amphiphiles.

INTRODUCTION

The lipid bilayer of biological cells and organelles constitutes the dynamic barrier between the interior and exterior environments. Integral membrane proteins housed within the lipid bilayer are involved in a wide range of biological functions including respiration, photosynthesis, uptake of nutrients, efflux of waste products and toxins as well as mediating cellular responses to a wide range of biologically active molecules. In addition, it is estimated that 20-30% of eukaryotic open reading frames (ORFs) encode for α-helical membrane proteins [1,2]. The importance of these proteins to cellular and whole organism physiology has made them the subject of intensive study. Until relatively recently, understanding of almost all groups of membrane proteins was severely hampered by the lack of high-resolution structural information. However the last 10 years have seen a mini explosion in the numbers of membrane protein structures [3]. This has provided unprecedented insight into the mechanism of action of several groups of membrane proteins. However of the 191 unique membrane protein structures available (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html) only 47 are of eukaryotic origin with the majority of these obtained using protein from naturally abundant native sources and the remaining obtained using protein from recombinantly expressed sources. These numbers reflect the difficulty of obtaining high quality eukaryotic integral membrane protein (eMP) samples suitable for structural studies.

Whether the relatively high number of prokaryotic membrane protein structures can be used to fill the gaps in knowledge about eukaryotic membrane proteins was addressed recently by Granseth and colleagues [4]. Somewhat unsurprisingly, the conclusion of their analysis was that to an extent, prokaryotic membrane protein structures where available can be used to model eukaryotic homologues. In the absence of a prokaryotic homologue or where more definitive data is required, i.e., in almost all cases, then a high resolution structure of the eukaryotic protein is required.

The pressing need for these structures means that novel methods are required to address the problems associated with structural studies of eukaryotic integral membrane proteins (eMPs). This review outlines some of these problems and summarises some of the key methodological advances made in the last few years, which have been successfully applied to membrane protein structure determination.

EXPRESSION

Despite major effort, the expression of eMPs still represents a major bottleneck in the production of proteins for structural studies [5,6]. There are examples of successful expression of eMPs in *E. coli* [7,8] however to date there is only one high resolution structure of an eMP expressed in *E. coli* (human 5-lipoxygenase-activating protein [9]). More successful alternatives are the eukaryotic expression systems, particularly the yeasts, *Pichia pastoris* and *Saccharomyces cerevisae,* and insect cell based systems. These have a number of advantages over *E. coli* in that they are capable of post-translational modifications and have lipid contents closer to higher eukaryotes [10]. A number of membrane proteins expressed in *P. pastoris* have yielded high resolution structures including the mammalian voltage-dependent potassium channel [11], the plant aquaporin [12] and the human LTC4 synthase [13] and most recently the mouse

*Address correspondence to this author at the Division of Molecular Biosciences, Imperial College London, South Kensington, London, SW7 2AZ, UK; Fax: +44 20 7594 3022; E-mail: b.byrne@imperial.ac.uk

1389-2037/10 $55.00+.00

ABC multidrug transporter, P-glycoprotein [14]. *S. cerevisae* has been used to produce protein which ultimately yielded the high resolution structure of P-type proton pump from *Arabidopsis thaliana* [15] and diffracting crystals of rabbit Ca^{2+}ATPase SERCA1a [16]. Insect cell based expression systems have had spectacular successes for the production of G-protein coupled receptors (GPCRs) for structural studies, so far yielding structures of 3 independent proteins, the human b_2-adrenergic receptor [17,18], the turkey b_1 adrenergic receptor [19] and the human adenosine A_{2A} receptor [20]. In addition, insect cells were also used to produce the chicken voltage dependent acid sensing ion channel which was solved to 1.9 Å resolution [21].

The majority of researchers working in the area of eukaryotic membrane protein structural determination use one of these three systems for protein production. Other systems are available including the Semliki Forest Virus mammalian cell expression system [22], which has had particular success in high level expression of GPCRs [23] and an alternative bacterial system using the Gram positive bacterium *Lactococcus lactis* [24]. Neither system has yet resulted in a high resolution structure but it is unclear whether this is due to inherent unsuitability of the systems for large scale production of eMPs for structural studies or the fact that these systems are not as extensively utilized as the yeast and insect cell systems. It will be interesting to see if in the future, these systems prove viable alternatives to yeast and insect cell cultures. An alternative approach is to attempt expression in as many different systems as possible as described for the human serotonin receptor [25] .

Although some researchers are performing careful, rational optimization of the existing systems with a view to maximizing quality and quantity of the target proteins [26], not many major advances have been made in terms of expression of eMPs in the last 5-10 years. One exception to this is the development of cell-free expression systems for the production of membrane proteins [27,28]. This type of approach involves in-vitro production of proteins outside intact cells from a DNA or mRNA template using a basic set of biological building blocks prepared from cell lysate [29]. A wide range of cell lysates are available and it has been shown that those from eukaryotic sources, e.g. rabbit reticulocytes, are capable of post-translational modifications [30], suggesting these may be suitable for production of eukaryotic membrane proteins. By its very nature, the system removes the problem of cytotoxicity and also simplifies protein isolation as the number of contaminant proteins is markedly reduced. A further advantage of this system is the high level of control a user has, in terms of modifying media components to both increase the stability of the expressed proteins (e.g., addition of protease inhibitors, lipids, co-factors) and to effectively, efficiently label for analysis by NMR [31] or X-ray crystallography. A major advantage for membrane proteins is that the expressed MPs are maintained in a soluble state in detergent micelles post-translationally since there are no native membrane environments for insertion [32]. There are many examples of the production of functionally active prokaryotic and eukaryotic MPs in cell free systems [27]. In addition, and most promisingly, it has proved possible to prepare selenomethionine labeled sample of a bacterial multidrug resistance transporter, EmrE. This protein was successfully crystallized and these crystals contributed to structure determination [33]. As ever, work on eMPs lags behind but efforts are currently underway to exploit this system effectively for structural studies.

PIPELINE APPROACHES FOR THE RAPID IDENTIFICATION OF EMPS SUITABLE FOR STRUCTURAL STUDIES

Homologue screening as a means to identify membrane proteins suitable for structural studies has been utilized effectively a number of times, for example in the case of the bacterial Sec translocon [34]. In this case, researchers started with 10 bacterial homologues and focused their efforts on the protein which expressed to the highest level and was most stable in a range of detergents [34].

Larger scale operations for membrane proteins based on the soluble protein structural genomics efforts have been much less successful as it is much more difficult to streamline operations using single expression constructs, expression systems and purification strategies as has been described for soluble protein work Parallelisation approaches attempting high-throughput screening in a number of different expression systems and using a wide range of expression constructs have been attempted and produced large quantities of data but the overall process has proved somewhat cumbersome and extremely expensive. The development of so-called pipelines approaches, analogous to the drug discovery pipelines utilized extensively by the pharmaceutical industry may have gone some way to rationalizing the identification of eMPs for structural studies. Like the drug discovery pipelines these approaches have a series of built in quality control checkpoints. Any protein which fails to reach a set of criteria is excluded from the target list. As in the case of drug discovery, the overall aim is to remove less suitable targets as early on in the process as possible, diverting more time and resources to the most promising ones. Adequate numbers of homologues must be identified and robust checkpoints established.

One heavily used approach is the *S. cerevisae* GFP pipeline [35,36]. The expression vector incorporates a galactose inducible promoter and a tobacco etch virus (TEV) cleavage site followed by the gene coding for a C-terminally His tagged green fluorescent protein (GFP). Homologous recombination is utilized to rapidly generate expression constructs where the gene of interest is inserted upstream of the TEV cleavage site. This results in a construct whereby the recombinantly expressed target protein is fused to a C-terminally located GFP (Fig. **1A**). The His tag on the GFP facilitates purification while the TEV cleavage site allows efficient removal of the GFP resulting in an almost native protein. Protein expression is performed in a protease deficient *S. cerevisae* cell line [37] using small scale cultures (10 ml). The key to this system is the presence of the GFP, which allows rapid assessment of the expression levels by simple fluorescence measurements. These measurements represent the first checkpoint on the pipeline-any protein expressing at less than 1 mg/L can be subjected to simple optimization. If the required expression level is not reached however, the target is removed from the pipeline. Integrity of the expressed protein can be assessed using in-gel fluores-

cence, a convenient alternative to Western blot analysis (Fig. **1B**) which is much quicker and does not require the use of expensive antibodies. A protein which shows multiple bands on a gel may be subject to either degradation and/or aggregation and may warrant either further investigation or removal from the pipeline. Further preliminary analysis, requiring larger scale culture (500 mL), can be performed using fluorescent size exclusion chromatography (FSEC [38]), a technique which allows assessment of the aggregation status of solubilised GFP tagged proteins (Fig. **1C**). Size exclusion chromatography (SEC) is carried out as normal but the samples can be either collected in an adapted fraction collector suitable for analysis by a 96-well plate fluorimeter or an in-line fluorescence detector. Since only the fluorescent proteins are detected it is possible to analyse crude solubilised samples in this way (Fig. **1C**). Several different detergents can be assessed to give an indication of the stability of the target protein in both long chain detergents e.g., *n*-dodecyl-β-D-maltopyranoside (DDM) and shorter chain detergents potentially more suitable for crystallization trials e.g., *n*-nonyl-β-D-maltopyranoside (NM). Proteins which aggregate heavily (detected by an abnormally large void peak) or degrade (detected by an abnormally large GFP peak) can be excluded from the pipeline or subjected to further screening. A standard purification protocol is used for isolation of the proteins which has been tried and tested for a range of eMPs [36] prior to crystallization trials. The proteins are submitted to Ni^{2+} immobilized metal affinity chromatography (IMAC) followed by TEV protease cleavage to remove the GFP. The target protein is then separated from both the His-tagged GFP and the His tagged TEV protease by reverse IMAC. A final polishing step of SEC is also performed, which gives an indication of the aggregation status of the protein. The approach has so far yielded >30 pure eMPs in a form suitable for structural studies [36].

One clear potential criticism of such a method is that the protein assessed in the early stages is a GFP fusion protein and therefore not truly representative of the final protein sample. However the careful analysis carried out by Newstead and colleagues [35], demonstrates that the preliminary analysis is a reasonable indicator of the suitability of the target protein for further studies. A DDM-solubilised crude protein sample that exhibited aggregation during FSEC, also showed aggregation after purification and removal of the GFP tag. The same trend was observed for all the successfully purified examples where the absence of aggregation during FSEC analysis correlated well with the production of pure, monodispersed protein. However even using this system, there are instances where proteins aggregate heavily upon concentration to 10 mg/ml. It is very difficult to predict how any protein behaves at high concentrations without first isolating large quantities of that protein. In this case alternative strategies can be utilized in order to screen for buffer conditions which maintain the protein in a monodispersed state (See screening for aggregation below).

An alternative to the GFP pipeline has been developed by Robert Stroud's group [39]. In this case 384 *S. cerevisae* membrane proteins predicted to have three or more transmembrane domains were used as targets. As with the GFP system, ligation independent cloning techniques are used to clone the gene of interest into a vector incorporating the galactose inducible GAL1 promoter, an N-terminal FLAG tag and a PreScission 3C protease cleavage site, upstream of the gene of interest and a C-terminal thrombin protease cleavage site followed by a His tag. The gene of interest is cloned between the protease cleavage sites. This method proceeds directly to medium scale culture (500 ml) and measurement of both expression level and solubilisation efficiency in parallel. Both are assessed by semi-quantitative Western blot analysis using antibodies against both the FLAG and His tags. The checkpoints in this case are the presence of a Western blot signal indicating an approximate expression level and at least 50% recovery of DDM solubilised protein following high-speed centrifugation. Targets not meeting these criteria are removed from the pipeline. A straightforward purification protocol utilizing IMAC followed by SEC is used to both isolate and further characterize the proteins. Proteins yielding less than 0.5 mg/L after IMAC and those exhibiting high levels of aggregation during SEC are also removed from the pipeline. This approach rapidly identified a sub-population of targets (~25%) suitable for further, more intensive study. The approach was developed using eMPs from S. *cerevisae*, however application of the method to human MPs confirms the validity of the method to production of MPs from higher eukaryotic sources.

The possibility for fully quantitative analysis is one of the major advantages of the GFP pipeline however the Stroud approach is faster, simpler and requires very little sophisticated equipment. It is possible to easily obtain more information about the target proteins using the GFP pipeline, for example localization of the expressed protein; however it is not clear at this point whether that information is required to identify those targets suitable for further studies. In addition, the Stroud approach is not constrained by the requirement that GFP must be localized in the cytoplasm thus limiting its use to proteins with an intracellular C-terminus. Although it has been demonstrated that the large majority of membrane proteins are likely to have an intracellular C-terminus [40]. Overall both approaches are valid ways of rapidly identifying targets in the earlier stages of the process where the functional significance of a given protein only becomes a major consideration when that protein is shown to be suitable for structural studies. Both have high success rates in terms of producing crystallization quality protein (at least 90% pure and in mg quantities). However the identification and purification of these targets is not the end of the screening process since painstaking optimization of sample buffer conditions and crystallization conditions is still required for each individual target. The two approaches are based on the same principle that some eMPs are inherently more suitable for structural studies than others, and it is just necessary to sift through many unsuitable targets before finding the true gems.

ANTIBODY APPROACHES FOR GPCRS

In contrast to the high-throughput pipelines designed to screen a wide range of related targets, approaches focused much more on individual target proteins have been successfully applied to the structure determination of GPCRs. The GPCRs are key mediators of cellular responses to a wide range of biologically active molecules including hormones, neurotransmitters and approximately 50% of all available

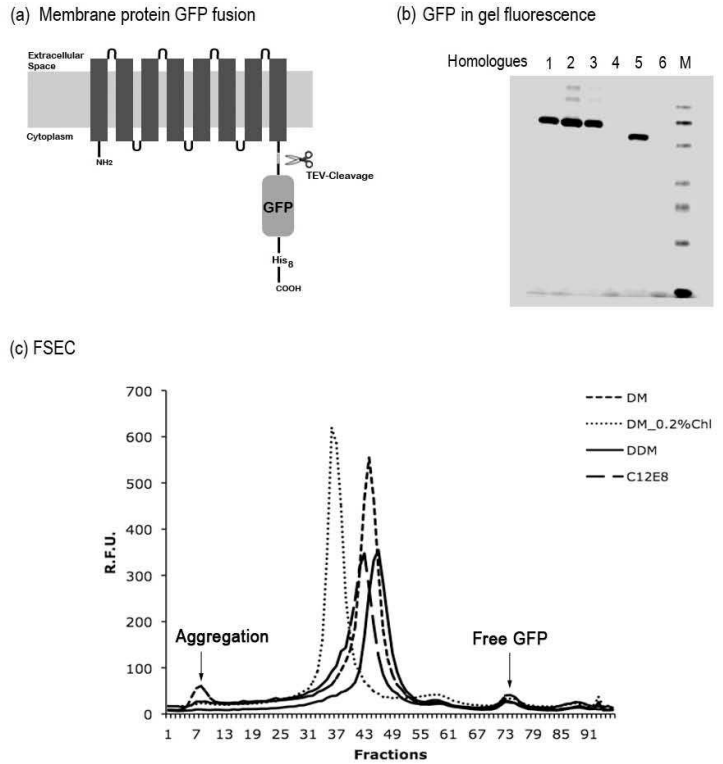


Fig. (1). (**a**) A topographical representation of a MP-GFP fusion protein. Purification of the fusion protein using IMAC is followed by cleavage with TEV protease to remove the GFP. The untagged target protein is separated from His-tagged GFP and His-tagged TEV protease by a further IMAC step. (**b**) In-gel fluorescence analysis of eukaryotic membrane proteins expressed as fusions with GFP. Crude membrane preparations containing the GFP-fusion proteins were separated on 10% Tris-Glycine SDS-PAGE. Visualization by in-gel fluorescence confirms the presence of full-length MP-GFP fusion proteins for 4 of the 6 tested proteins. (**c**) FSEC profile of a eukaryotic MP-GFP fusion protein solubilised in DDM, DM, DM +0.2% Cholesterol, and $C_{12}E_8$. All detergents yield a monodispersed solubilisation profile although it is possible that there is higher oligomer formation in the presence of cholesterol. The expected positions of peaks corresponding to aggregated protein and free GFP are indicated on the FSEC profile.

drugs [41]. For many years structural studies on these molecules had lagged behind those on other classes of membrane proteins including transporters and ion channels. The first major success came in 2000 with the high resolution structure of bovine rhodopsin [42], a non-typical GPCR with an intrinsic ligand which changes conformation upon interaction with a photon of light. This induces a further conformational change in the GPCR which activates the G-protein and initiates an intracellular signaling cascade. This structure provided a template for other studies and indeed drug design [43]. However the very features of rhodopsin that made it amenable to structural studies made it of limited use for understanding the precise details of ligand-receptor interactions for all other GPCRs. High resolution structures are needed for a wide range of such receptors.

The major problem of low yields of functional recombinant expression of GPCRs had been overcome in a number of cases using a variety of expression constructs and systems [7,8,44,45]. In addition, there were examples of GPCRs being purified to high levels of homogeneity and exhibiting long term functional stability [45-47]. The research also

benefitted from the ready availability of high affinity ligands to produce a more conformationally homogeneous sample. However despite the production of protein apparently suitable for structural studies no well-diffracting crystals were obtained. Almost simultaneously, three different approaches produced breakthroughs resulting in high-resolution protein structures of two different but related receptors.

The first structure, a partial structure of the β_2-adrenergic receptor [18], was solved in complex with an antibody fragment (Fig. **2A**). The antibody bound to the third intracellular loop, stabilising the loop and providing the principal surface for crystal contacts within the lattice. The structure only revealed details of about two thirds of the protein, mainly the intracellular loops and the cytoplasmic side of the transmembrane (TM) helices. The extracellular ends of the TM helices and the extracellular loops were too disordered to allow the structure to be resolved in these regions. Nevertheless, this first structure gave insights into the overall architecture of the β_2-adrenergic receptor and allowed comparison with the structure of rhodopsin [18]. The antibody approach, to extend the hydrophilic domain of membrane proteins, has been used successfully for a number of both prokaryotic and eukaryotic membrane proteins [48-51]. Despite these successes, the traditionally long and expensive preparation and screening process for the antibodies with no guarantee of success has made this approach unattractive to many working in the field. Novel approaches using designed ankyrin repeat proteins (DARPins) as alternative co-crystallisation agents have been successfully applied to the high resolution structure determination of the bacterial multidrug resistance protein, AcrB [52]. Ankyrin repeat motifs are present in a wide range of naturally occurring proteins with diverse cellular roles, and are best characterized in terms of their importance in protein-protein interactions. The sequence motif has a defined architecture with a relatively small number of residues responsible for protein-protein interaction. Protein libraries have been generated using phage display and ribosome display technology which allow the rapid generation of a large number of potentially high affinity binding proteins. The proteins vary in the amino acid composition of the protein-protein interaction sites but have a maintained structural fold. These libraries can be screened for high affinity binding to a target protein with the best binders being used as co-crystallisation agents [53]. The relative ease of generation and screening of the DARPins makes them an attractive alternative to monoclonal antibody fragments. Currently there are no examples of eMPs crystallized in complex with a DARPin, however these binders have significant potential for the future.

THE T4-LYSOZYME FUSION APPROACH

The second approach to structural determination of a GPCR involved removal of disordered regions of the human β_2-adrenergic receptor (intracellular C-terminal domain and third intracellular loop) and insertion of a highly ordered protein, T4 lysozyme, into the third intracellular loop [54]. This chimeric protein expressed well in insect cells, was shown to be highly functional and ultimately yielded well diffracting crystals and a high resolution structure [17]. The presence of the T4 lysozyme did not significantly alter the conformation of the receptor as confirmed by a comparison with the structure of the partial protein in complex with the antibody fragment [18]. The major flaw in the approach is the removal of key domains important for G-protein coupling and downstream signaling events. However the structure obtained provided the first detailed insight into the ligand binding pocket of the b_2-adrenergic receptor. Subsequent studies using the same receptor construct have also provided insight into the role of cholesterol in the structural integrity of the receptor [55]. This approach has since been successfully applied to the structure determination of another GPCR, the human adenosine A_2 receptor [20] suggesting that this method may be suitable for a wide range of GPCRs.

ALANINE SCANNING MUTAGENESIS

The third and final approach involved mutagenesis of the target receptor, in this case the b_1 adrenergic receptor from turkey, with a view to generating a mutant receptor with increased thermostability in solution and thus more suitable for structural studies. A similar approach had previously been used for the diacylgycerol kinase (DGK), an integral membrane enzyme from *E.coli*. Random mutagenesis was used to generate DGK mutants, which were assessed for enzyme activity before and after heat treatment. The most thermostable substitutions were combined to produce a quadruple mutant which had a half-life of 35 min at 80°C when solubilised in *n*-octyl-β-D-glucopyranoside (OG) compared to a half-life of less than 1 min for the wild-type protein [56].

Researchers began the study with a turkey β_1 adrenergic receptor construct which had already been significantly modified to improve expression in insect cells, facilitate purification and eliminate proteolytic sensitivity [45]. The resultant construct (residues 34-424) was submitted to alanine scanning mutagenesis [57] whereby 318 amino acid residues of the truncated turkey β_1-adrenergic receptor were mutated to an alanine, except when an alanine was present in the native protein, in which case a leucine substitution was made. Each of the mutants was expressed in *E. coli* for screening purposes. The thermostability of each mutant protein was assessed by radioligand binding analysis before and after heating to 32°C (T_m of the wild-type receptor) and compared with the wild-type protein. Whilst most of the mutations had either no effect or a negative effect on thermostability some resulted in increased thermostability. Certain combinations of the thermostabilising mutants were shown to further increase thermostability and the best construct, m23, containing a combination of six point mutations, had an apparent T_m 21°C higher than the wild-type receptor. Interestingly, unlike the wild-type form which was only stable in DDM, the m23 mutant was stable in short chain detergents and in the absence of ligand was preferentially in the antagonist bound conformation leading to greater homogeneity of the protein sample. The overall improved stability of m23 had a major effect on the ability of the protein to crystallise and this construct yielded well-diffracting crystals in the C8 detergent octylthioglucoside and ultimately a high resolution structure ([19]; Fig. **2C**). It should be mentioned that for structural studies, the m23 mutant was expressed in insect cells and not in *E. coli*.

This conformational thermostabilization process has been performed on two further GPCRs; the human adenosine A_{2A}

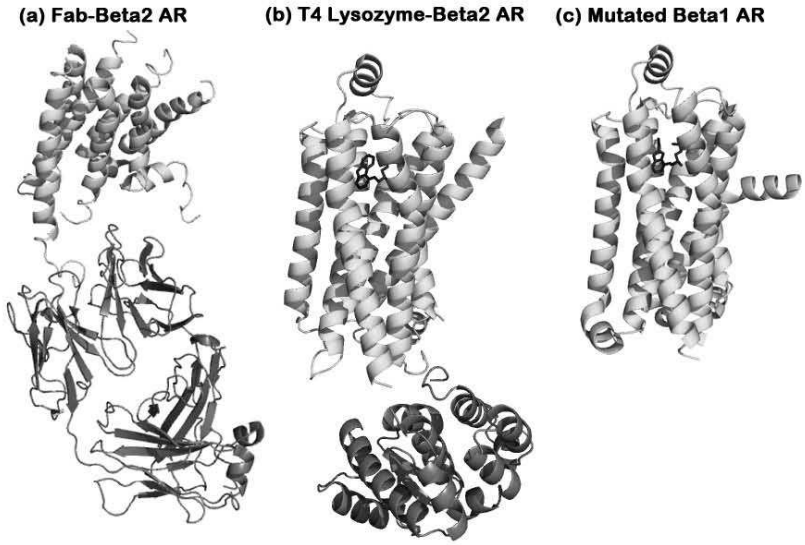


Fig. (2). Ribbon representations of three GPCR structures obtained using proteins stabilised in three different ways. **(a)** Human β_2-adrenergic receptor (light grey) fused with a T4-lysozyme (dark grey), ([17], 2RH1.pdb). The inverse agonist carazolol, bound in the ligand binding site of the receptor, is indicated as a stick model. **(b)** Structure of the human β_2-adrenergic receptor crystallized in complex with a Fab-antibody fragment (dark grey ribbons) fragment ([18], 2R4R.pdb). **(c)** Structure of the thermostabilised mutant β_1-adrenergic receptor from turkey ([19], 2VT4.pdb). The high affinity antagonist, cyanopindolol, bound in the ligand binding site of the receptor, is shown as a stick model.

receptor [58] and the rat neurotensin receptor [59]. As yet these studies have not resulted in high resolution structures but it will be interesting to see how widely applicable this approach is both to other GPCRs and other classes of eukaryotic membrane proteins. One interesting issue raised by this work is that it is not clear why mutating particular residues makes the receptors more stable. After completing the process with three separate receptors no pattern has emerged allowing prediction of sites where mutation would increase thermostability. Currently, at least, future work using this technique still requires extensive alanine-scanning mutagenesis together with a means for reliable assessment and comparison of thermostability and functionality of the mutants.

ASSESSMENT OF MEMBRANE PROTEIN STABILITY

As mentioned above, improved stability has proved critical to the successful crystallization of a number of eMPs. The recognition of this has led to the development of generic methods to assess the thermostability of a given membrane protein and to allow rapid and efficient screening of conditions which give stable, protein suitable for structural studies. The recent report of a fluorescent thermal stability assay specific for membrane proteins [60] has followed on from the successful application of such methods to soluble proteins. The basis of these assays is the use of dye molecules which bind to proteins as they unfold. Upon heating unfolding is detected as an increase in fluorescence as a result of the reaction between the dye and the unfolding protein. The dyes, e.g. Sypro orange, used for soluble proteins are not suitable for use with membrane proteins as they give very high background levels thought to be due to non-specific binding to detergent molecules and the hydrophobic regions of membrane protein molecules. The recently described approach [60] utilizes an alternative dye, N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM)[60], a fluorochrome which reacts readily with the free thiols of cysteine residues. As described for the soluble protein work, as the membrane protein unfolds the normally buried cysteine residues become available to interact with the dye and the level of fluorescence gives an indication of the rate of unfolding. The CPM is virtually non-fluorescent in the unbound form and exhibits low non-specific binding to detergent and folded membrane protein. The assay procedure is very straightforward involving dilution of small amounts of protein (1-20 μg) in an appropriate test buffer, addition of the CPM dye and slowly heating the sample from low to high temperature in a fluorimeter, taking fluorescence measurements at regular intervals. The rate of unfolding in any

given assay condition can be easily compared with standard test conditions, e.g., NM compared to DDM.

The very specific interaction between the CPM and the thiol groups of cysteine residues is the major strength of the method however it is only possible to detect unfolding in those proteins containing cysteine residues. For proteins lacking cysteines alternative methods need to be used to assess thermostability or cysteine residues need to be mutated into the protein [60]. In addition the dye is not compatible with buffer conditions lower than pH 5.0 and higher than pH 8.0 again potentially requiring alternative methods. Despite these minor drawbacks the thermal stability assay is a very efficient method for comparing the stability of a given membrane protein in a range of different detergents, buffer and ionic strength conditions, additives etc. which requires very small amounts of protein and can be performed in a very short time [55]. These features make the assay an ideal method for high-throughput analysis of sample conditions more suitable for crystallization. Other techniques for analyzing the thermostability of integral membrane proteins include differential scanning calorimetry (DSC), circular dichroism (CD), UV/VIS spectroscopy and nuclear magnetic resonance (NMR). These methods can provide complementary assessment of thermostability however they usually require much larger amounts of proteins and are currently not well suited to high-throughput analysis.

SCREENING FOR AGGREGATION

Detergent-solubilised membrane proteins are very prone to non-specifically aggregate, particularly at the high concentrations required for crystallization trials. This is highly disadvantageous to downstream structural studies since non-specific aggregation of the protein reduces the likelihood of the formation of specific protein-protein interactions essential for crystal lattice formation. There are a number of methods to assess aggregation status but many of these are either low throughput or unsuitable for use with detergent solubilised samples. One recent development which allows both detection of aggregation and rapid screening of conditions to inhibit aggregation is the ultracentrifugation dispersity sedimentation (UDS) assay where ultracentrifugation of small (5 µL) volumes of purified, soluble membrane protein is combined with SDS-PAGE analysis to rapidly assess the degree of protein aggregation in high concentration (10-20 mg/ml) sample [61]. The novelty of the assay lies in the combination of these two standard laboratory techniques to assess the aggregation status of a protein in a number of different conditions simultaneously. The assay is based on the assumption that protein aggregates are orders of magnitude heavier than dispersed protein particles and can therefore be removed by sedimentation at high *g* forces. Protein samples taken before and after high-speed ultracentrifugation (~ 350, 000 *g*) are visualised on SDS-PAGE gels. Those samples which exhibit aggregation will show lower levels of protein in solution after ultracentrifugation [61]. The small volumes required mean that the UDS assay is economical in terms of both protein and detergent. It also allows assessment of the aggregation status of the protein sample at high concentrations, not requiring dilution of the sample or modification of the sample buffer as is necessary for SEC. This method has been used to both identify conditions which maintain a membrane protein in a monodispersed state at high concentration and to screen for detergents and buffer conditions suitable for optimisation of preliminary crystals.

DEVELOPMENT OF NEW DETERGENTS

The extraction of membrane proteins from the bilayer into detergent solubilized conditions is a prerequisite for structural studies. The best classical detergents for membrane protein research are usually nonionic and have the key feature of being able to extract membrane proteins with high efficiency whilst maintaining structural and functional integrity [62]. However protein-detergent micelles are particularly unstable, prone to aggregation and often exhibit a loss of protein function. The detergent molecules themselves are also disordered [63] and likely to inhibit the formation of highly ordered crystal lattice. In order to address these issues much effort has been directed towards the production of novel detergents that form more stable protein-detergent micelles, more suitable for protein isolation and crystallization [64-66]. The aim is to reduce the disorder introduced by the intrinsically flexible alkyl chains of commercial detergents and thus increase the chance of an ordered crystal lattice in all three dimensions. There are a range of such molecules, including amphipols [64] and tripod amphiphiles [66]. The amphipols (amphipathic polymers) were designed to form a tighter interaction with the membrane protein in order to reduce instability incurred through rapid association and dissociation of the detergent molecules. These have been shown to increase stability of some proteins. The major use of the amphipols may be as additives since their stabilizing effects seem to be greater when used in combination with detergents and lipids [65,67]. The so called "tripod amphiphiles'" (TPAs) are comprised of a tetrasubstituted carbon atom with one hydrophilic substituent and three hydrophobic substituents [68]. The tetrasubstituted carbon is thought to introduce extra rigidity through the reduction of conformational flexibility. This feature means that TPAs should be more stable than conventional detergents and this leads to the hypothesis that these molecules are more likely to contribute to an ordered crystal lattice [66]. Studies have demonstrated that TPA (Fig. **3**) gives greater solubilisation efficiency of bacteriorhodopsin (bR) from purple membranes than OG, Triton X-100 or DDM and more importantly TPA–solubilised bR was more stable [66]. Small crystals of TPA-solubilised bR have been obtained which diffract to 2.5 Å resolution [69]. This represents a small increase in resolution

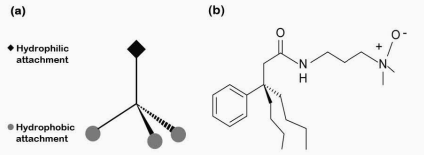


Fig. (3). Schematic representation of a tripod amphiphile. **(a)** The basic shape of a tripod amphiphile consisting of a tetra-substituted carbon atom with one hydrophilic substituent and three hydrophobic substituents. **(b)** The chemical structure of the tripod amphiphile that was successfully used to crystallise bacteriorhodopsin (bR).

compared to larger crystals of OG-solubilised bR. The increase in resolution is suggestive that TPAs may facilitate crystallization of membrane proteins either as alternatives to conventional detergents or as additives to crystallization trials. More research is required to further characterize the TPAs and to explore their true potential as alternatives to standard detergents.

CONCLUSION AND FUTURE PERSPECTIVES

Here we have summarized some of the key developments in the methodologies available to researchers working with membrane proteins. These advances have led to some spectacular success in the structure determination of eMPs, particularly the GPCRs. As described, the process of obtaining a high resolution structure is often as much an issue of identifying the unsuitable constructs, protein homologues and buffer conditions and excluding these from further studies as it is about identifying the optimal ones. Much effort has been put into developing techniques to make this process faster and more effective. While the number of eMP structures remains currently low it is anticipated that the long term effects of these advances will soon be felt as more eMPs suitable for structural studies are identified and studied in detail. Until then, it remains difficult to state which approaches will prove the most useful, although the production of T4 lysozyme fusion proteins holds great potential for the comparatively routine structure determination of GPCRs. One key issue remains the structure determination of multiple conformational states of membrane proteins. In the case of GPCRs, it is possible that such an issue may be addressed by for example, the use of specific ligands together with G-proteins as was described for the recent structure of squid opsin [70] or a combination of mutagenesis and inhibitors as has been used successfully used for prokaryotic transporters [71]. Alternatively other combinations, e.g., alanine scanning mutagenesis together with use of a T4 lysozyme fusion protein may be advantageous. In addition, the use of antibody fragments and DARPins are likely to be of use in improving the resolution of MP structures and thus the molecular details.

Whilst X-ray crystallography remains the technique of choice for structure determination of membrane proteins, other techniques are emerging. The very recent report of the low resolution structure of the human BK potassium channel using cryoelectron microsocopy is the first structure of a membrane protein in a lipid environment [72]. Clearly such a technique has great potential for providing insight into the specific roles of protein-lipid interactions and for elucidation of multiple conformational states.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Biotechnology and Biological Research Council, the Defense, Science and Technology Laboratories and GlaxoSmithKline for support. This work was also supported by the European Commission under the FP7 EDICT project (Grant No. 201924). CA is supported by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Brazilian fellowship.

ABBREVIATIONS

eMP	=	Eukaryotic integral membrane protein
ORF	=	Open reading frame
DDM	=	*n*-Dodecyl-β-D-maltopyranoside
NM	=	*n*-Nonyl-β-D-maltopyranoside
OG	=	*n*-Octyl-β-D-glucopyranoside
IMAC	=	Immobilized metal affinity chromatography
TEV	=	Tobacco etch virus
SEC	=	Size exclusion chromatography
FSEC	=	Fluorescent size exclusion chromatography
GPCR	=	G-protein coupled receptor
UDS	=	Ultracentrifugation dispersity sedimentation
CPM	=	N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide
TPA	=	Tripod amphiphile.

REFERENCES

[1] Wallin, E.; Von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Prot. Sci.*, **1998**, *7*(4), 1029-1038.

[2] Granseth, E.; Daley, D.O.; Rapp, M.; Melen, K.; von Heijne, G. Experimentally constrained topology models for 51,208 bacterial inner membrane proteins. *J. Mol. Biol.*, **2005**, *352*(3), 489-94.

[3] Carpenter, E.P.; Beis, K.; Cameron, A.D.; Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.*, **2008**, *18*(5), 581-6.

[4] Granseth, E.; Seppälä, S.; Rapp, M.; Daley, D.O.; Von Heijne, G. Membrane protein structural biology: How far can the bugs take us? *Mol. Memb. Biol.*, **2007**, *24*(5), 329 - 332.

[5] Midgett, C.R.; Madden, D.R. Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies. *J. Struct. Biol.*, **2007**, *160*(3), 265-74.

[6] Junge, F.; Schneider, B.; Reckel, S.; Schwarz, D.; Dotsch, V.; Bernhard, F. Large-scale production of functional membrane proteins. *Cell Mol. Life Sci.*, **2008**, *65*(11), 1729-55.

[7] Sarramegna, V.; Talmont, F.; Demange, P.; Milon, A. Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification. *Cell Mol. Life Sci.*, **2003**, *60*(8), 1529-46.

[8] Grisshammer, R.; White, J.F.; Trinh, L.B.; Shiloach, J. Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview. *J. Struct. Funct. Genomics,* **2005**, *6*(2-3), 159-63.

[9] Ferguson, A.D.; McKeever, B.M.; Xu, S.; Wisniewski, D.; Miller, D.K.; Yamin, T.T.; Spencer, R.H.; Chu, L.; Ujjainwalla, F.; Cunningham, B.R.; Evans, J.F.; Becker, J.W. Crystal structure of inhibitor-bound human 5-lipoxygenase-activating protein. *Science,* **2007**, *317*(5837), 510-2.

[10] Opekarova, M.; Tanner, W. Specific lipid requirements of membrane proteins--a putative bottleneck in heterologous expression. *Biochim. Biophys. Acta,* **2003**, *1610*(1), 11-22.

[11] Long, S.B.; Campbell, E.B.; MacKinnon, R. Crystal Structure of a Mammalian Voltage-Dependent Shaker Family K+ Channel. *Science,* **2005**, *309*(5736), 897-903.

[12] Tornroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P. Structural mechanism of plant aquaporin gating. *Nature,* **2006**, *439*(7077), 688-694.

[13] Martinez Molina, D.; Wetterholm, A.; Kohl, A.; McCarthy, A.A.; Niegowski, D.; Ohlson, E.; Hammarberg, T.; Eshaghi, S.; Haeggstrom, J.Z.; Nordlund, P. Structural basis for synthesis of inflammatory mediators by human leukotriene C4 synthase. *Nature,* **2007**, *448*(7153), 613-6.

[14] Aller, S.G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P. M.; Trinh, Y. T.; Zhang, Q.; Urbatsch, I.L.; Chang, G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science,* **2009**, *323*(5922), 1718-22.

[15] Pedersen, B.P.; Buch-Pedersen, M.J.; Morth, J.P.; Palmgren, M.G.; Nissen, P. Crystal structure of the plasma membrane proton pump. *Nature,* **2007**, *450* (7172), 1111-4.

[16] Jidenko, M.; Nielsen, R.C.; Sorensen, T.L.; Moller, J.V.; le Maire, M.; Nissen, P.; Jaxel, C. Crystallization of a mammalian membrane protein overexpressed in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA,* **2005**, *102*(33), 11687-91.

[17] Cherezov, V.; Rosenbaum, D.M.; Hanson, M.A.; Rasmussen, S.G.; Thian, F.S.; Kobilka, T.S.; Choi, H.J.; Kuhn, P.; Weis, W.I.; Kobilka, B.K.; Stevens, R.C. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science,* **2007**, *318*(5854), 1258-65.

[18] Rasmussen, S.G.; Choi, H.J.; Rosenbaum, D.M.; Kobilka, T.S.; Thian, F.S.; Edwards, P.C.; Burghammer, M.; Ratnala, V.R.; Sanishvili, R.; Fischetti, R.F.; Schertler, G.F.; Weis, W.I.; Kobilka, B.K. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature,* **2007**, *450*(7168), 383-7.

[19] Warne, T.; Serrano-Vega, M.J.; Baker, J.G.; Moukhametzianov, R.; Edwards, P.C.; Henderson, R.; Leslie, A. G.; Tate, C.G.; Schertler, G.F. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature,* **2008**, *454*(7203), 486-91.

[20] Jaakola, V.P.; Griffith, M.T.; Hanson, M.A.; Cherezov, V.; Chien, E.Y.; Lane, J.R.; Ijzerman, A.P.; Stevens, R.C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science,* **2008**, *322*(5905), 1211-7.

[21] Jasti, J.; Furukawa, H.; Gonzales, E.B.; Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature,* **2007**, *449*(7160), 316-323.

[22] Lundstrom, K. Semliki Forest virus vectors for rapid and high-level expression of integral membrane proteins. *Biochim. Biophys. Acta,* **2003**, *1610*(1), 90-6.

[23] Hassaine, G.; Wagner, R.; Kempf, J.; Cherouati, N.; Hassaine, N.; Prual, C.; Andre, N.; Reinhart, C.; Pattus, F.; Lundstrom, K. Semliki Forest virus vectors for overexpression of 101 G protein-coupled receptors in mammalian host cells. *Protein Expr. Purif.,* **2006**, *45*(2), 343-51.

[24] Kunji, E.R.; Chan, K.W.; Slotboom, D.J.; Floyd, S.; O'Connor, R.; Monne, M. Eukaryotic membrane protein overproduction in Lactococcus lactis. *Curr. Opin. Biotechnol.,* **2005**, *16*(5), 546-51.

[25] Tate, C.G.; Haase, J.; Baker, C.; Boorsma, M.; Magnani, F.; Vallis, Y.; Williams, D.C. Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. *Biochim. Biophys. Acta,* **2003**, *1610*(1), 141-153.

[26] Bonander, N.; Hedfalk, K.; Larsson, C.; Mostad, P.; Chang, C.; Gustafsson, L.; Bill, R.M. Design of improved membrane protein production experiments: quantitation of the host response. *Prot. Sci.,* **2005**, *14*(7), 1729-40.

[27] Daniel Schwarz, V.D. Frank Bernhard,, Production of membrane proteins using cell-free expression systems. *Proteomics,* **2008**, *8*(19), 3933-3946.

[28] Liguori, L.; Marques, B.; Villegas-Méndez, A.; Rothe, R.; Lenormand, J.-L. Production of membrane proteins using cell free expression systems. *Exp. Rev. Proteomics,* **2007**, *4*(1), 79-90.

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

[30] Jackson, A.M.; Boutell, J.; Cooley, N.; He, M. Cell-free protein synthesis for proteomics. *Brief Funct. Genomic Proteomic,* **2004**, *2*(4), 308-19.

[31] Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. Optimal isotope labelling for NMR protein structure determinations. *Nature,* **2006**, *440*(7080), 52-57.

[32] Klammt, C.; Schwarz, D.; Fendler, K.; Haase, W.; Dötsch, V.; Bernhard, F.; Evaluation of detergents for the soluble expression of α-helical and β-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J.,* **2005**, *272*(23), 6024-6038.

[33] Chen, Y.J.; Pornillos, O.; Lieu, S.; Ma, C.; Chen, A.P.; Chang, G. X-ray structure of EmrE supports dual topology model. *Proc. Natl. Acad. Sci. USA,* **2007**, *104*(48), 18999-9004.

[34] Van den Berg, B.; Clemons, W.M.Jr.; Collinson, I.; Modis, Y.; Hartmann, E.; Harrison, S.C.; Rapoport, T.A. X-ray structure of a protein-conducting channel. *Nature,* **2004**, *427*(6969), 36-44.

[35] Newstead, S.; Kim, H.; von Heijne, G.; Iwata, S.; Drew, D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA,* **2007**, *104*(35), 13936-41.

[36] Drew, D.; Newstead, S.; Sonoda, Y.; Kim, H.; von Heijne, G.; Iwata, S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat. Protoc.* **2008**, *3*(5), 784-98.

[37] Kota, J.; Gilstring, C.F.; Ljungdahl, P.O. Membrane chaperone Shr3 assists in folding amino acid permeases preventing precocious ERAD. *J. Cell Biol.,* **2007**, *176*(5), 617-628.

[38] Kawate, T.; Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure,* **2006**, *14*(4), 673-81.

[39] Li, M.; Hays, F.A.; Roe-Zurz, Z.; Vuong, L.; Kelly, L.; Ho, C.M.; Robbins, R.M.; Pieper, U.; O'Connell, J.D.3rd; Miercke, L.J.; Giacomini, K.M.; Sali, A.; Stroud, R.M. Selecting optimum eukaryotic integral membrane proteins for structure determination by rapid expression and solubilization screening. *J. Mol. Biol.,* **2009**, *385*(3), 820-30.

[40] Kim, H.; Melén, K.; Osterberg, M.; von Heijne, G. A global topology map of the Saccharomyces cerevisiae membrane proteome. In *Proc. Natl. Acad. Sci. USA,* **2006**, *103*(30) 11142-7.

[41] Overington, J.P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.,* **2006**, 5(12), 993-996.

[42] Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C.A.; Motoshima, H.; Fox, B.A.; Le Trong, I.; Teller, D.C.; Okada, T.; Stenkamp, R.E.; Yamamoto, M.; Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science,* **2000**, *289*(5480), 739-45.

[43] Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem.,* **2002**, *3*(10), 928-44.

[44] McCusker, E.C.; Bane, S.E.; O'Malley, M.A.; Robinson, A.S. Heterologous GPCR expression: a bottleneck to obtaining crystal structures. *Biotechnol. Prog.,* **2007**, *23*(3), 540-7.

[45] Warne, T.; Chirnside, J.; Schertler, G.F. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *Biochim. Biophys. Acta,* **2003**, *1610*(1), 133-40.

[46] Grisshammer, R.; Tucker, J. Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein Expr. Purif.,* **1997**, *11*(1), 53-60.

[47] Weiss, H.M.; Grisshammer, R. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Eur. J. Biochem.,* **2002**, *269*(1), 82-92.

[48] Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H. Structure at 2.8 A resolution of cytochrome c oxidase from Paracoccus denitrificans. *Nature,* **1995**, *376*(6542), 660-669.

[49] Hunte, C.; Koepke, J.; Lange, C.; Rossmanith, T.; Michel, H. Structure at 2.3 Å resolution of the cytochrome bc1 complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Structure,* **2000**, *8*(6), 669-684.

[50] Zhou, Y.; Morais-Cabral, J.H.; Kaufman, A.; MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution. *Nature,* **2001**, *414*(6859), 43-48.

[51] Dutzler, R.; Campbell, E.B.; MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science,* **2003**, *300*(5616), 108-12.

[52] Sennhauser, G.; Amstutz, P.; Briand, C.; Storchenegger, O.; Grutter, M.G. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol.,* **2007**, *5*(1), e7.

[53] Sennhauser, G.; Grutter, M.G. Chaperone-assisted crystallography with DARPins. *Structure,* **2008**, *16*(10), 1443-53.

[54] Rosenbaum, D.M.; Cherezov, V.; Hanson, M.A.; Rasmussen, S.G.F.; Thian, F.S.; Kobilka, T.S.; Choi, H.-J.; Yao, X.-J.; Weis, W.I.; Stevens, R.C.; Kobilka, B.K. GPCR Engineering Yields High-Resolution Structural Insights into 2-Adrenergic Receptor Function. *Science,* **2007**, *318*(5854), 1266-1273.

[55] Hanson, M.A.; Cherezov, V.; Griffith, M.T.; Roth, C.B.; Jaakola, V.-P.; Chien, E.Y.T.; Velasquez, J.; Kuhn, P.; Stevens, R.C. A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human [beta]2-Adrenergic Receptor. *Structure,* **2008**, *16*(6), 897-905.

[56] Zhou, Y.; Bowie, J.U. Building a thermostable membrane protein. *J. Biol. Chem.,* **2000**, *275*(10), 6975-9.

[57] Serrano-Vega, M.J.; Magnani, F.; Shibata, Y.; Tate, C.G. Conformational thermostabilization of the beta1-adrenergic

receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. USA,* **2008**, *105*(3), 877-82.
[58] Magnani, F.; Shibata, Y.; Serrano-Vega, M.J.; Tate, C.G. Co-evolving stability and conformational homogeneity of the human adenosine A2a receptor. *Proc. Natl. Acad. Sci. USA,* **2008**, *105*(31), 10744-9.
[59] Shibata, Y.; White, J.F.; Serrano-Vega, M.J.; Magnani, F.; Aloia, A.L.; Grisshammer, R.; Tate, C.G. Thermostabilization of the Neurotensin Receptor NTS1. *J. Mol. Biol.,* **2009**, *390*(2), 262-77.
[60] Alexandrov, A.I.; Mileni, M.; Chien, E.Y.; Hanson, M.A.; Stevens, R.C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure,* **2008**, *16*(3), 351-9.
[61] Gutmann, D.A.; Mizohata, E.; Newstead, S.; Ferrandon, S.; Postis, V.; Xia, X.; Henderson, P.J.; van Veen, H.W.; Byrne, B. A high-throughput method for membrane protein solubility screening: the ultracentrifugation dispersity sedimentation assay. *Prot. Sci.,* **2007**, *16*(7), 1422-8.
[62] Privé, G.G. Detergents for the stabilization and crystallization of membrane proteins. *Methods,* **2007**, *41*(4), 388-397.
[63] Timmins, P.; Pebay-Peyroula, E.; Welte, W. Detergent organisation in solutions and in crystals of membrane proteins. *Biophys. Chem.,* **1994**, *53*(1-2), 27-36.
[64] Popot, J.L.; Berry, E.A.; Charvolin, D.; Creuzenet, C.; Ebel, C.; Engelman, D.M.; Flotenmeyer, M.; Giusti, F.; Gohon, Y.; Hong, Q.; Lakey, J.H.; Leonard, K.; Shuman, H.A.; Timmins, P.; Warschawski, D.E.; Zito, F.; Zoonens, M.; Pucci, B.; Tribet, C. Amphipols: polymeric surfactants for membrane biology research. *Cell Mol. Life Sci.,* **2003**, *60*(8), 1559-74.
[65] Tribet, C.; Audebert, R.; Popot, J.L. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci USA,* **1996**, *93*(26), 15047-50.
[66] Yu, S.M.; McQuade, D.T.; Quinn, M.A.; Hackenberger, C.P.; Krebs, M.P.; Polans, A.S.; Gellman, S.H. An improved tripod amphiphile for membrane protein solubilization. *Prot. Sci.,* **2000**, *9*(12), 2518-27.
[67] Champeil, P.; Menguy, T.; Tribet, C.; Popot, J.L.; le Maire, M. Interaction of amphipols with sarcoplasmic reticulum Ca^{2+}-ATPase. *J. Biol. Chem.,* **2000**, *275*(25), 18623-37.
[68] McQuade, D.T.; Seungju, M.A.Q.; Yu, S.M.; Polans, A.S.; Krebs M.P.; Gellman, S.H. Rigid Amphiphiles for Membrane Protein Manipulation. *Angewandte Chemie,* **2000**, *39*(4), 758-761.
[69] Theisen, M.J.; Potocky, T.B.; McQuade, D.T.; Gellman, S.H.; Chiu, M.L. Crystallization of bacteriorhodopsin solubilized by a tripod amphiphile. *Biochim. Biophy.s Acta,* **2005**, *1751*(2), 213-6.
[70] Scheerer, P.; Park, J.H.; Hildebrand, P.W.; Kim, Y.J.; Krausz, N.; Choe, H.-W.; Hofmann, K.P.; Ernst, O.P. Crystal structure of opsin in its G-protein-interacting conformation. *Nature,* **2008**, *455*(7212), 497-502.
[71] Singh, S.K.; Piscitelli, C.L.; Yamashita, A.; Gouaux, E. A Competitive Inhibitor Traps LeuT in an Open-to-Out Conformation. *Science,* **2008**, *322*(5908), 1655-1661.
[72] Wang, L.; Sigworth, F.J. Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy. *Nature,* **2009**, *461*(7261), 292-295.

Received: June 09, 2009 Revised: September 22, 2009 Accepted: October 01, 2009

EXPRESSION, PURIFICATION AND CRYSTALLIZATION ATTEMPTS OF HUMAN BILE ACID-COA:AMINO ACID N-ACYLTRANSFERASE (BAAT)

Laura Civiero, Stefano Capaldi, Massimiliano Perduca and Hugo L. Monaco
Biocrystallography Lab, Department of Biotechnology, Strada le Grazie 15, 37134 Verona, Italy.
E-mail: laura.civiero@univr.it

Bile acid-coenzyme A: Amino Acid N-Acyltransferase (BAAT) is the sole enzyme responsible for conjugation of primary and secondary bile acids to taurine and glycine (Figure 1) [1]. The human cDNA BAAT sequence was cloned into the pET15b vector and its expression was tested in several *E.coli* strains. Because of the toxicity associated with overexpression of BAAT in several strains, BL21 (DE3) C41 was selected to express the recombinant protein. Only with this specific strain, a reasonable amount of the protein was obtained in the soluble fraction. The purification steps involved Nickel-sepharose chromatography followed by reverse IMAC after histidine-tag cleavage. Size-exclusion and hydrophobic interaction chromatography steps were also included to increase the purity of the sample. After these steps the protein shows a single band in both SDS and native PAGE. The purified enzyme was assessed for activity and was found to be able to catalyze the conjugation of taurine to CoA-fatty acids [2]. The presence of non physiological inter- and/or intramolecular disulfide bridges between the three cysteine residues present in the protein sequence as a source of sample microheterogenity was also examined. Three cysteine mutants (C235A, C372A and C373A, and the triple mutant without cysteines) were constructed and tested following the same protocols used for the wild type protein; significant differences in the behaviour of the four species were found. Crystallization trials with the apo- and holo-wild type enzyme, as well as with the three mutants, are in progress but up to now only microcrystals have been obtained.

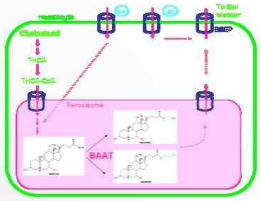


Fig.1 Bile salts serve as detergents to keep fat-soluble compounds in solution and, in addition, have been recognized as important signaling molecules which are involved in cholesterol homeostasis.

They are synthesized from cholesterol in the liver, and are secreted into the bile via the bile salt export pump. Before secretion, bile salts must be conjugated with either taurine or glicine which are linked to the side chain of C_{24}.

Conjugation of bile salts with either taurine or glycine is catalyzed by the enzyme BAAT [3].

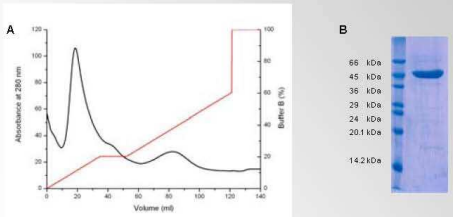


Fig.2 A) Nickel-sepharose chromatography: the soluble fraction from 2 L of culture was loaded into a His-trap column and eluted with an imidazole gradient; B) SDS-PAGE of the enzyme at the end of the purification protocol.

Cloning and expression details:

- ✓ BAAT was cloned in the pET15b vector (that adds an N-terminal His-tag)
- ✓ the culture was induced at 1.0 OD.ml with 0.25 mM IPTG, overnight at 20°C
- ✓ the pure protein yield is 1.5 mg/L of culture

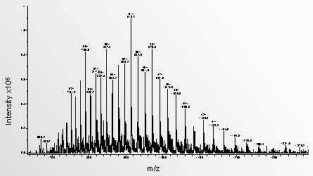


Fig.3 The integrity of the enzyme was checked by mass spectrometry, the mass is 46552 Da, in agreement with the calculated value.

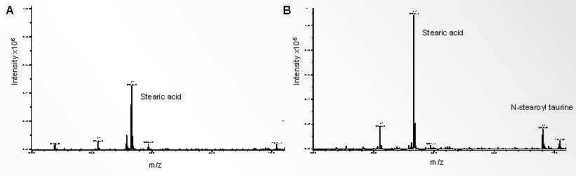


Fig.4 Negative ion electrospray mass spectrometry analysis of BAAT fatty acid conjugating activity was carried out on incubations containing A) 2-5 µg of BAAT and 40 µM stearoyl-CoA and B) 2-5 µg of BAAT, 40 µM stearoyl-CoA and 50 mM taurine.

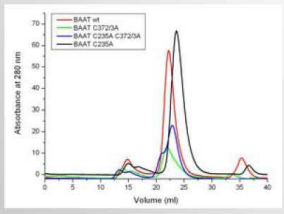
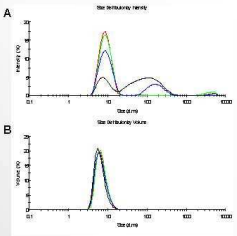


Fig.5 Size exclusion chromatography of the wild type enzyme and the three mutants.



C

Sample	Z-average (d.nm)	PdI
BAAT wild type (3 mg/ml)	8.09	0.088
BAAT C235A (6 mg/ml)	8.75	0.204
BAAT C372/3A (1mg/ml)	10.90	0.357
BAAT C235A C372/3A (1mg/ml)	23.57	0.825

Fig.6 Dynamic light scattering experiments. Measurements were made on a Malvern Zetasizer instrument at 20°C: A) intensity and B) volume size distributions obtained for the wild type enzyme and the three mutants; C) z-average diameters and polydispersity index (PdI) of the samples.

Conclusions

- ✓ The purity and quality of the wild type protein is suitable for crystallization trials
- ✓ The enzyme is active
- ✓ The examined mutants are more difficult to purify. Yield and stability are lower
- ✓ The polydispersity of some mutants precludes their use for crystallization trials
- ✓ Only the catalytic mutant (C235A) appears to be adequate for crystal growth

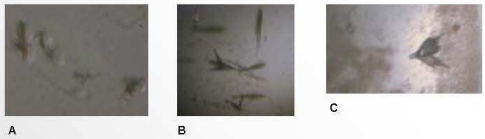


Fig.7 BAAT C235A microcrystals obtained with 0.1M Hepes pH 7.5, 10% PEG 8000, 8% ethylene glycol as precipitant. A) apo-protein, B) in presence of 5 mM stearoyl-CoA. C) Crystals of wild type BAAT in presence of 1 mM sodium cholate, 1 mM CoA obtained with 0.2M calcium acetate hydrate, 0.1M sodium cacodylate pH 6.5, 18% PEG 8000.

References

[1] Stiakiancos MK, Wilson L, Sakellan M, Falany CN, Barnes S. *J Biol Chem* 2002, 277, 47270.
[2] O'Byrne J, Hunt MC, Rai DK, Sadel M, Alexson SE. *J Biol Chem* 2003, 278, 34237.
[3] Pellicoro A, van den Heuvel F, Geuken M, Moshage H, Jansen P, Faber K. *Hepatology* 2007, 45, 340.